Mechanisms of transcriptional repression by DNA methylation

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Sebastian Kaluscha

Basel, 2022

Genehmigt von der

Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Dirk Schübeler
Dr. Michael Stadler
Prof. Dr. Pierre-Antoine Defossez

Basel, den 22.06.2021

Dekan
Prof. Dr. Marcel Mayor

Acknowledgements

This thesis would not have been possible without the support and contribution of many people.
First and foremost, I would like to thank my PhD advisor, Dirk Schübeler, for the opportunity to work in his group and for excellent guidance throughout my PhD. I am grateful for the many interesting discussions and the freedom and flexibility in exploring new ideas.

I want to thank the co-authors of the prepared manuscript, in particular Silvia Domcke for the great collaboration, scientific discussions and feedback. In addition, I want to thank Christiane Wirbelauer for establishing the Ngn2 cell lines that has enabled us to generate neurons without DNA methylation. I want to thank Lukas Burger for his essential bioinformatics support and for teaching me analysis of sequencing data with great patients and fun.

Special thanks go to the FMI Genomics Facility, particularly to Sebastien Smallwood for his support with high-throughput sequencing.
My sincere appreciation goes to all members of the Schübeler group I have been working with for a great working atmosphere and an amazing scientific environment. Particularly, I want to thank Paul Ginno and Mario Iurlaro for constant discussions, friendship and fun outside the lab.

I would like to thank the members of my PhD thesis committee, Pierre-Antoine Defossez, Luca Giorgetti, Tuncay Baubec and Piera Cicchetti for valuable scientific input, interesting discussions and support throughout my PhD.
A special thanks goes to my friends, my family and Marion Frank for continuous support throughout this exciting time.

Table of Contents

List of Abbreviations

5-Aza	5-Aza-2′-deoxycytidine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ATAC-seq	Assay for Transposase-Accessible Chromatin followed by sequencing
BAZ2A	Bromodomain Adjacent To Zinc Finger Domain 2A
BAZ2B	Bromodomain Adjacent To Zinc Finger Domain 2B
bp	Base pair
cAMP	Cyclic adenosine monophosphate
Cas9	CRISPR-associated protein 9
CGI	CpG island
ChIP-seq	Chromatin immunoprecipitation sequencing
CpG	Cytosine-phosphate-Guanine
CRE	cAMP response element
CREB1	cAMP responsive element binding protein 1
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
DAP-seq	DNA affinity purification sequencing
DAZL	Deleted in azoospermia-like
DBD	DNA binding domain
DNA	Desoxyribonucleic acid
DNMT	DNA methyltransferases
DNMT-TKO	DNA methyltransferases triple knockout
ES	Embryonic stem
FDR	False discovery Rate
H3K27ac	Histone 3 lysine 27 acetylation
H3KXme	Histone 3 lysine X (e.g., 4, 9, 27, 36) methylation
HDAC	Histone deacetylation complex
HEK293	Human embryonic kindey 293
HNF6	Hepatocyte nuclear factors
HP1	Heterochromatin Protein 1
HT-SELEX	High-throughput systematic evaluation of ligands by exponential enrichment
ICR	Imprinted control region
ING	Inhibitor of Growth
KAP1	KRAB interacting protein 1
kb	Kilo base pair
KZFP	Krueppel-associated box domain (KRAB) zinc finger proteins (ZFP)
MBD	Methyl-CpG binding domain
MBD-QKO	Methyl-CpG binding domain quadruple knockout
MCAF	MBD1-containing chromatin-associated factor
MeCP2	Methyl-CpG-binding protein 2

NCoR1/2	Nuclear Receptor Corepressor 1/2
NID	NCoR1/2 interaction domain
NRF1	Nuclear respiratory factor 1
NuRD	Nucleosome Remodeling Deacetylase
PBM	Protein binding microarrays
PHD	Plant Homeodomain
piRNA	PIWI-interacting RNAs
QKO	Quadruple knockout
RING	Really Interesting New Gene
RNA	Ribonucleic acid
RTT	Rett syndrome
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SETDB1	SET Domain Bifurcated Histone Lysine Methyltransferase 1
SRA	SET- and RING-associated
TE	Transposable element
TET	Ten-eleven translocation methylcytosine dioxygenase
TF	Transcription factor
TKO	Triple knockout
TRD	Transcriptional repressor domain
UHRF1	Ubiquitin Like With PHD And Ring Finger Domains 1
WGBS	Whole-genome bisulfite sequencing
WT	Wild type
ZFP	Zingc finger protein

1. Abstract

Modification at the fifth carbon of cytosines (5mC) in the context of CpG (cytosine-phosphate-guanin) dinucleotides is a wide-spread DNA modification that is essential for mammalian development. DNA methylation is associated with transcriptional repression and required for silencing of evolutionary young repetitive elements and some CpG rich promoters in somatic cells. However, how DNA methylation translates into transcriptional repression remains enigmatic. Accumulating *in vitro* evidence suggests that DNA methylation can directly repel TF binding by motif methylation, although *in vivo* evidence remains scarce. An indirect repression model suggests that methyl-CpG binding domain (MBD) proteins recognize densely methylated DNA and cause transcriptional repression by the interaction with co-repressors. Importantly, a direct and indirect repression model are not mutually exclusive and it is likely that both mechanisms are at play. However, individual or combinatorial deletions of some MBD proteins during mouse embryogenesis, do not result in a phenotype that is associated with the loss of DNA methylation. While it has been argued that this could be explained by functional redundancy between MBD proteins, genetic evidence is missing, as a simultaneous deletion of all MBD proteins has not been reported to date. Here, we test the indirect repression model by deleting all 5mC binding MBD proteins in different mammalian cell lines. We show that upon deletion, we do not detect upregulation of methylated CpG-rich promoters found for instance in germline-specific genes or repetitive elements in mouse embryonic stem cells or derived neurons. In contrast, neurons that lack DNA methylation die after several days in culture, which is associated with strong de-repression of repetitive elements. Mouse ES cells tolerate the loss of DNA methylation and do not show strong upregulation of TEs, which is attributed to an alternative repression mechanism involving H3K9me3. We suggest that absence of this mark in neurons explains the essentiality of DNA methylation for repeat repression in somatic cells.

In order to further dissect a DNA methylation mediated repression model, we explore the TF repertoire of young transposable elements that are highly de-repressed in methylation deficient neurons. By systematically investigating CREB1, we provide evidence that this regulation entails the direct repulsion of methylation-sensitive TFs.

In the second part of this thesis, we investigate MBD proteins outside the murine lineage. Therefore, we comprehensively delete all MBD proteins in a human cancer-derived cell line and contrast this to cells with a hypomethylated genome. This reveals, in line with the observations in mouse cells, a minor role of MBD proteins in translating DNA methylation into transcription repression. Taken together, this work provides evidence that the dominant mode of gene and repeat repression is the direct repulsion of transcription factors by DNA methylation.

2. General introduction

Eukaryotic transcription factors and chromatin

The genome consists of four nucleotides that encodes the information required for an organism's identity. Organisms such as prokaryotes or eukaryotes alike face a common challenge: the correct spatiotemporal read out of genomic information in form of RNA. While the genome is stable throughout our life (except for a subset of our immune cells), the information that is read from it is highly variable, depending on physiological stimuli, environmental factors or developmental stages. At the core of these regulatory processes are transcription factors (TFs), proteins that bind DNA motifs in a sequence-specific manner and regulate transcription (Vaquerizas et al. 2009; Fulton et al. 2009). Such motif specificity allows TFs to interpret genomic information and exert specific gene regulatory programs. Indeed, many TF have critical functions as regulators of cell fate (Vierbuchen and Wernig 2012) and compelling examples include TF-induced reprogramming of fibroblasts into pluripotent stem cells or many TF mediated-differentiations regimes (Takahashi and Yamanaka 2006, 2016). In humans, about eight percent of all genes are TFs (Lambert et al. 2018) and are overrepresented in disease phenotypes (Köhler et al. 2014), thus highlighting their importance.

Early research in bacteria pioneered by investigation of the *lac operon* in *Escherichia coli* (Jacob and Monod 1961), established basic concepts of gene regulation and a principle understanding that TFs bind DNA in a sequence specific manner (Gilbert and Maxam 1973; Ptashne 1967). Since then, a major effort has been put into the characterization of binding principles of TFs. Recently, *in vitro* high-throughput binding assays have been used to identify DNA binding specificities of TFs. For instance, protein binding microarrays (PBM) (Weirauch et al. 2014), high-throughput systematic evaluation of ligands by exponential enrichment (HT-SELEX) (Jolma et al. 2010) and DAP-seq (Bartlett et al. 2017) have expanded the collection of TF motifs annotated in large online libraries such as JASPAR (Khan et al., 2018). While TFs have similar motif binding preferences *in vitro* and *in vivo* (X. Liu et al. 2006), in eukaryotes motif presence appears to be a poor predictor for TF occupancy *in vivo*, and this underscores a fundamental differences to prokaryotes (Wunderlich and Mirny 2009). For instance, in prokaryotes the DNA binding domain (DBD) of TFs is large enough to

recognize a unique region within the genome that typically results in one binding event (Wunderlich and Mirny 2009). Paradoxically, DBDs of eukaryotes share relatively similar protein domains, yet encounter a much larger genome and therefore do not encode sufficient information to provide unique and site specific binding (Wunderlich and Mirny 2009). Instead, TF binding sites are small - 6-8 bp for homeodomain TFs (Berger et al. 2008), often have degenerate nucleotides and are predicted to appear by chance every four thousand base pairs in most mammalian genomes (Wunderlich and Mirny 2009).

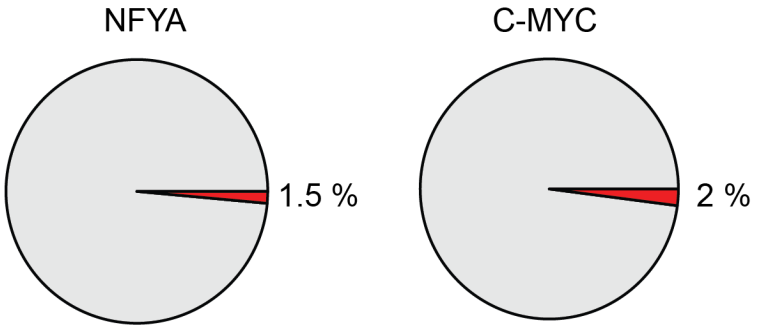


Figure 1. Eukaryotic transcription factors only bind a fraction of their motifs. Here illustrated for NFYA and C-MYC that only occupy (red) a few percent of their predicted motifs (grey, data from mouse embryonic stem cells NFYA (Tiwari et al. 2011), C-MYC (Chronis et al. 2017)).

Profiling TF binding using *in vivo* techniques such as chromatin immunoprecipitation followed by sequencing (ChIP-seq) reveals that only a minority of predicted binding sites are occupied in eukaryotic organisms **(Figure 1)** (Biggin 2011). This may be partially explained by additional sequence factors such as DNA shape (Slattery et al. 2014) or binding site syntax (Farley et al. 2016), as well as TF cooperativity (Jolma et al. 2015) and the non-uniform 3D distribution of TF molecules within the nucleus (Kribelbauer et al. 2019). However, a major barrier for DNA accessibility is attributed to chromatin proteins (i.e., histones) and DNA methylation. Yet, to what extent these impact and regulate TF binding and what principle mechanism are underlying this process remains enigmatic.

TF binding in the context of nucleosomes

Eukaryotic DNA is wrapped around histones in a repeating fashion that appears as "beads on a string". Two copies of the four canonical histones H3, H4 and H2A and H2B create an octamer that is wrapped with about 147 bp of DNA to form a complex termed the nucleosome (Kouzarides 2007). This allows for compaction of DNA and is referred to as chromatin, though the role of chromatin far extends its initially suggested role of condensing DNA in the nucleus. Nucleosomes have a key function in gene regulation as they create a barrier for accessing DNA for DNA binding proteins (Makowski, Gaullier, and Luger 2020; G. Li and Widom 2004) **(Figure 2a)**. It was demonstrated in reconstituted biochemical systems that nucleosomes can impede TF binding to DNA *in vitro* (Knezetic and Luse 1986; Lorch, LaPointe, and Kornberg 1987) and indeed most occupied TF binding sites are devoid of nucleosomes *in vivo* (Yuan et al. 2005).

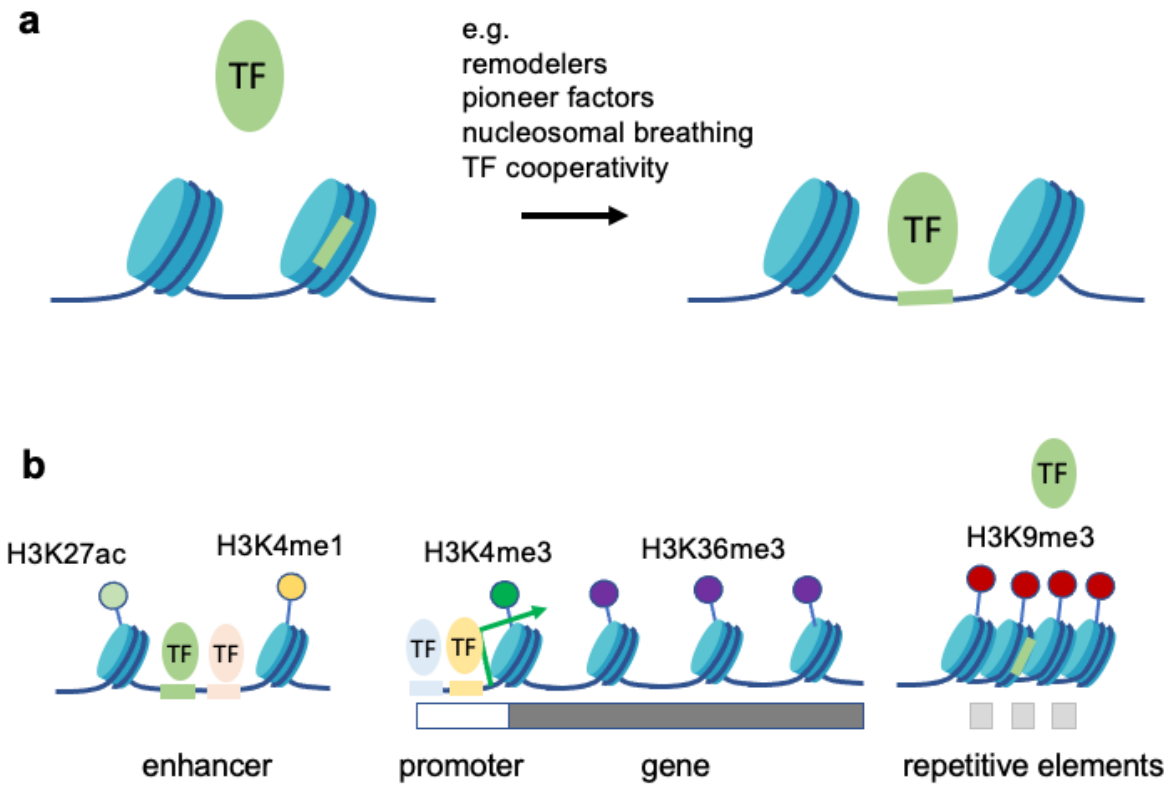


Figure 2. TF binding and nucleosomes a) TF binding is restricted by nucleosomes. Different mechanisms are proposed on how TF can bind their cognate motif in the presence of nucleosomes. b) Histone modifications correlate with different genomic features such as active regulatory regions, actively transcribed gene bodies or repetitive elements. How and if these modifications influence TF binding is unclear.

However, how nucleosomes influence TF binding and gene regulation in the cellular context remains elusive. A small subset of TFs have been proposed to bind motifs located within nucleosomes, termed ‘pioneer TFs’, and have been associated with lineage specification (Iwafuchi-Doi and Zaret 2014). It is suggested that upon binding these factors initiate the unwrapping of nucleosomal DNA in order to allow binding of further TFs to initiate transcription (Makowski, Gaullier, and Luger 2020). However, many pioneer TFs also do not occupy all motifs and show differential binding between cell types, suggesting additional complexity must exists to explain TF occupancy (Makowski, Gaullier, and Luger 2020). Other means of TFs to engage with nucleosomal DNA is the involvement of ATP-dependent chromatin remodelers that are able to actively slide, eject and re-assemble nucleosomes (Clapier et al. 2017). For instance, the yeast pioneer TF Rap1 can invade compact chromatin, cooperating with a chromatin remodeler to shift promoter nucleosomes (Mivelaz et al. 2020). In mammalian cells, different remodelers have been demonstrated to mediate binding of distinct TF (Barisic et al., 2019). Furthermore, cooperative binding of TFs has been suggested as a mechanism to compete with nucleosomes for DNA binding (Miller and Widom 2003). In addition, nucleosomes - influenced by post-translational modification and histone variants - are highly dynamic and a temporal and partial unwrapping of nucleosomal DNA (nucleosomal ‘breathing’) could provide a window of opportunity for TF binding (G. Li et al. 2005). Currently, it remains elusive how these proposed mechanisms integrate to regulate TF binding by nucleosomes in the cellular context.

TF binding in the context of histone modifications

Histones can be post-translationally modified, as established by the pioneering work of Vincent Allfreys in 1964 (Allfrey, Faulkner, and Mirsky 1964), which now includes several modifications such as acetylation, methylation, phosphorylation or ubiquitination etc. (Bannister and Kouzarides 2011). While modifications can occur in the core globular structure of the histone octamer, predominantly the histone tails that protrude from the nucleosome are modified (Strahl and Allis 2000).

Post-translational histone modifications associate with different activity states of chromatin. In mammals, transcriptionally inactive and closed chromatin (heterochromatin) associates with low acetylation levels and methylation of certain histones (e.g., H3K9, H3K27 and H420) (Kouzarides 2007). In contrast, actively transcribed euchromatin is characterized with high levels of acetylation and methylation of different histones such H3K4 or H3K36 (Kouzarides 2007).

Post-translational modification of histones often correlates with activity states of regulatory elements, however, their mechanistic impact on TF binding remains elusive **(Figure 2b)**. For instance, in heterochromatin trimethylation of H3K9 is associated with constitutive repression of mainly repetitive and centromeric elements, while trimethylation of H3K27me3 is implicated in cell type and lineage specific repression (Nicetto and Zaret 2019). Both modifications can block transcriptional initiation, yet two different mechanisms appear to be at play (Margueron and Reinberg 2011; Matsui et al. 2010; Dellino et al. 2004). While repressed and H3K27me3 marked promoters remain accessible to specific TFs and the transcriptional machinery, H3K9me3 domains are inaccessible and suggested to even impede binding of TFs (Becker, Nicetto, and Zaret 2016; Breiling et al. 2001).

Two non-mutually exclusive modes of actions have been suggested of how post translational histone modifications can affect the chromatin state (Bannister and Kouzarides 2011). The first is a direct structural perturbation of nucleosomes by changing the net histone affinity for negatively charged DNA (e.g., acetylation or phosphorylation). This changes the electrostatic interaction properties between histones and DNA that in turn impacts chromatin compaction and possibly accessibility for DNA binding factors (Bannister and Kouzarides 2011). Another scenario describes the interaction of the histone tails with chromatin associated factors. Indeed, numerous proteins have been characterized to either write, read or erase histone modifications

via conserved protein domains. Many domains have been identified to facilitate binding of different proteins to distinct histone modifications. For instance, the bromodomain mediates binding to acetyl-lysine and is found in numerous co-factors including co-activators and chromatin remodeling complexes (Mujtaba, Zeng, and Zhou 2007). An abundance of factors recognize methyl-lysine, potentially reflecting the relative importance of this modification (Kouzarides 2007). These include chromodomains or Plant Homeodomain (PHD) fingers that are found in many co-factors. For instance, Inhibitor of Growth (ING) proteins recognize via there PHD finger H3K4me3 and are suggested to recruit co-repressors such as histone deacetylation complexes (HDACs) (Guérillon, Larrieu, and Pedeux 2013).

Another prominent example is Heterochromatin Protein 1 (HP1), which binds histone tails via its chromodomain that recognizes to H3K9me3 (Loyola et al. 2001). Through self-oligomerization and recruitment of other repressive histone modifiers, HP1 is thought to promote chromatin compaction and transcriptional repression (Nicetto and Zaret 2019). Recently, it has been suggested that H3K9me3, via HP1, is involved in the establishment of non-membrane nuclear condensates that have distinct biophysical properties with potentially regulatory implications for TFs (Sabari, Dall'Agnese, and Young 2020; Larson et al. 2017).

Taken together, how and if post-translational histone modifications affect TF binding *in vivo* remains unclear and needs to be tested.

DNA methylation

Eukaryotic methylation patterns and genome defense

Methylation of the fifth carbon of cytosines (5-methylcytosine – 5mC) can be found in many eukaryotes such as plants, animals and fungi but is also present in many bacterial species (Zemach and Zilberman 2010). DNA methyltransferases (DNMTs) are enzymes that facilitate this modification. They comprise a catalytic methyltransferase domain that is highly conserved in eukaryotes and share sequence motifs with bacteria (Goll and Bestor 2005).

5mC is highly variable in its abundance and distribution in eukaryotic genomes or even absent is some organisms. For instance, the fruit fly *D.melanogaster* or the nematode *C.elegans* are virtually devoid of DNA methylation, highlighting that it is not essential

for eukaryotic life (Zemach and Zilberman 2010). Vertebrate genomes display blanket methylation almost throughout the entire genome ('hypermethylation'). Invertebrates and plants primarily have scarce DNA methylation with a 'mosaic' pattern that most likely resembles the ancestral eukaryotic state (Zemach et al. 2010; J. A. Law and Jacobsen 2010). Interestingly, a recent study identified that the invertebrate *Amphimedon* has a hypermethylated genome similar to that of vertebrates (de Mendoza et al. 2019). While this supports the hypothesis that hypermethylated genomes evolved gradually from a mosaic pattern, it challenges the idea that hypermethylation only occurred once in vertebrates (de Mendoza et al. 2019).

Over 30 years ago Timothy Bestor proposed that DNA methylation evolved from a prokaryotic immune system to a silencing mechanism in eukaryotes to prevent the activity of transposable elements (TE) (Bestor, 1990). In line with this model, mosaic DNA methylation occurs predominantly in TEs and actively transcribed gene bodies of plants and fungi examined so far (Zemach and Zilberman 2010). Indeed, loss of DNA methylation in *Arabidopsis* causes reactivation of TEs (Kato et al. 2003). In contrast, DNA methylation in most invertebrates investigated appears to have lost its repressive function, as most intragenic TEs lack DNA methylation (Zemach and Zilberman 2010; de Mendoza, Lister, and Bogdanovic 2020). The role of gene body methylation remains enigmatic as it is highly conserved across eukaryotes (Zemach et al. 2010). It has been suggested to be involved in protecting gene bodies of harmful TE insertions (Dirk Schübeler 2015) or suppressing intragenic promoters (Maunakea et al. 2010). In vertebrates, where high DNA methylation levels is the default state, TEs are highly methylated. Loss of DNA methylation results in de-repression of TEs during mouse embryogenesis throughout almost every tissue and early embryonic death (Okano et al., 1999; Walsh et al., 1998). Eukaryotic genomes dramatically vary in size, in a way that does not necessarily correlate with protein coding and non-coding genes (Hidalgo et al. 2017), yet scales with the amount of repetitive DNA, in particular TEs. In fact, it is argued that DNA methylation enabled the accommodation and co-existence of TEs within the host DNA that resulted in eukaryotic genome expansion (W. Zhou et al. 2020). In mouse and humans transposable elements populate about half of the genome, although this is likely an underestimate as most TEs are unrecognizable anymore due to genetic drift (Friedli and Trono 2015).

Targeted repression of transposable elements by DNA methylation and H3K9me3

Organisms that harbor TEs need to counteract their activity to prevent harmful consequences that can arise by jumping into genic or regulatory elements. Furthermore, hypomethylation of repeats is associated with erroneous homologous recombination of non-allelic elements that can cause chromosomal rearrangements and meiotic defects (Zamudio and Bourc'his 2010).

In mammals, global DNA methylation levels are temporarily reduced during pre-implantation development (from 70 % methylated CpGs to about 20 %) and almost completely erased in primordial germ cells (Lee, Hore, and Reik 2014). However, remaining residual methylation can be found especially at young TEs during both waves of genomic hypomethylation together with H3K9me3 (Greenberg and Bourc'his 2019). Deletion of pathway members important for setting DNA methylation at TEs or H3K9me3 in early primordial germ cells results in de-repression of TEs (S. Liu et al. 2014; Greenberg and Bourc'his 2019).
H3K9me3 is a DNA methylation independent silencing mechanism that is sufficient to repress TEs in absence of DNA methylation in mouse embryonic stem cells (Matsui et al. 2010). While DNA methylation in eukaryotes initially evolved as a mechanism to silence repeats this function presumably got lost in early animals (Zemach and Zilberman 2010). Typically invertebrates use alternative pathways for silencing TEs such as H3K9me3 (Gasser 2016). It is hypothesized that vertebrates re-evolved DNA methylation as a silencing mechanism for TEs, yet based on the invertebrate system (Zemach and Zilberman 2010). Therefore, H3K9me3 appears to provide vertebrates with an additional regulatory layer to silence TEs, especially when global DNA methylation levels are low. Two major classes of mediators are indicated to underly the sequence specific targeting of DNA methylation and H3K9me3 to TEs (Friedli and Trono 2015).

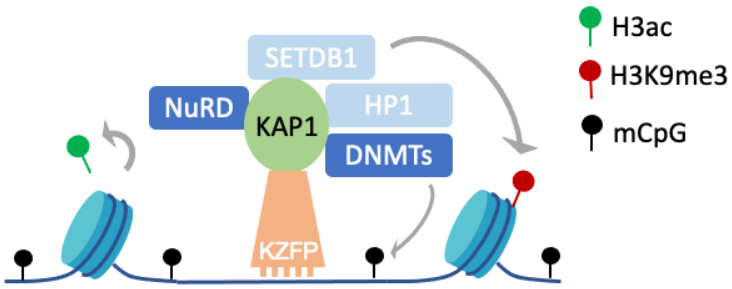


Figure 3. KRAP-ZFP/ KAP1 repressor complex with individual components. KZFP recognizes with its zinc fingers specific genomic regions such as transposable elements or imprinted regions and recruits KAP1. Multiple co-repressors are recruited to KAP1 that then establish heterochromatin. H3ac, acylated histone H3. Figure adapted from Ecco et al 2017.

A first class are Krueppel-associated box domain zinc finger proteins (KRAB-ZFPs) that make up the largest family of TFs in higher vertebrates (Imbeault, Helleboid, and Trono 2017). KRAB-ZFPs (KZFP) recognize TEs primarily with an array of sequence-specific zinc fingers and recruit KRAB associated protein 1 (KAP1) (Ecco, Imbeault, and Trono 2017). In turn, KAP1 functions as a scaffold for proteins involved in establishing heterochromatin such as the methyltransferase SETDB1 (David C. Schultz et al. 2002) that deposits H3K9me3, histone deacetylases (D. C. Schultz, Friedman, and Rauscher 2001) or DNMTs (Quenneville et al. 2012) during embryogenesis to enable long-term repression (Ecco, Imbeault, and Trono 2017) **(Figure 3)**. The KZFP gene family is currently thought to have appeared in the common ancestor of coelacanths, lungfish and tetrapods (Imbeault, Helleboid, and Trono 2017) and expanded in waves parallel with TEs and are under strong positive selection at ZFP encoding regions (Emerson and Thomas 2009; Ecco, Imbeault, and Trono 2017). These observations led to a model of an evolutionary arms race between KZFPs and retrotransposons driving their co-evolution (Ecco, Imbeault, and Trono 2017).

A second class are small RNAs that are considered a primary and ancient line of the host's defense mechanism against TEs that is prevalent in plants (Friedli and Trono 2015). A prominent example are PIWI-interacting RNAs (piRNA) that are conserved in many metazoans where they have a pivotal role in counteracting TE activity in the germline (Aravin, Hannon, and Brennecke 2007). They derive from transcripts of TEs that interact with evolutionary conserved and germline-restricted PIWI proteins (Aravin, Hannon, and Brennecke 2007). In mammals, via homology recognition, piRNAs guide DNA methylation to the promoter regions of TEs where they establish

life-long silencing during spermatogenesis (Zamudio and Bourc'his 2010; Aravin et al. 2008). Recently, a specialized DNA methyltransferase was identified, namely DNMT3C, that is specifically expressed in male germ cells of rodents where it methylates promoters of evolutionary young TEs (Barau et al. 2016). DNMT3C is suggested to be guided by small RNAs including piRNAs, although a direct interaction has not been proven yet (Greenberg and Bourc'his 2019). The absence of DNMT3C or of PIWI-related proteins results in repeat de-repression and male infertility (Barau et al. 2016; Zamudio and Bourc'his 2010). piRNAs are considered the first line of defense against new TEs, which arose by new insertion or mutation events. In a second step, that could have spanned up to 7 million years, KZFP paralogs might have evolved that recognize new TEs and hamper their activity further (Ecco, Imbeault, and Trono 2017).

DNA methylation in mammals

The DNA methylation machinery

In mammals, the most prevalent DNA modification is the methylation of cytosines in the context of CpG dinucleotides. About 70% of CpGs are symmetrically methylated in somatic tissue and maintained during replication by the maintenance DNA methyltransferase DNMT1, which shows preference towards hemi-methylated DNA as a substrate (Smith and Meissner 2013). DNMT1 is recruited to the replication fork by UHRF1 (also known as NP95), where it methylates the newly synthesized strand (Bostick et al. 2007). UHRF1 has the ability to recognize H3K9me2/3 (Rothbart et al. 2012) and non-symmetrically methylated CpGs (Avvakumov et al. 2008; Arita et al. 2008). Absence of DNMT1 or its adaptor protein UHRF1 during murine development results in early embryonic death (Li et al., 1992; Sharif et al., 2007)

DNMT3A and DNMT3B catalyze *de novo* DNA methylation with redundant and distinct functions (Okano et al., 1999). For instance, DNMT3B appears more important for embryonic development, while DNMT3A is critical for germline DNA methylation (Greenberg and Bourc'his 2019).

In contrast, Ten-Eleven Translocation (TET) enzymes can actively demethylate 5mC to 5-hydroxymethylcytosine (5hmC) and further oxidized derivatives that can be removed passively through replication or actively by repair enzymes (Tahiliani et al.

2009; Ambrosi, Manzo, and Baubec 2017). 5mC can also be passively lost by preventing its maintenance throughout cell division (T. Chen et al. 2003).

CpG islands

Methylated cytosines spontaneously undergo deamination that can result in a mutagenic transition from cytosine to thymine (Cohen, Kenigsberg, and Tanay 2011). While mutagenic deamination has been argued to promote the inactivation of TEs, it is also attributed to have caused the genome-wide CpG depletion that is observed in mammals (one fifth of its expected frequency) and other eukaryotes (Bird 1986; W. Zhou et al. 2020). However, some regions resisted depletion and resulted in an uneven distribution of CpGs throughout the genome. So called CpG islands (CGI) are genomic regions (about 1kb) with relatively high levels of CpG dinucleotides compared to the rest of the genome (Gardiner-Garden and Frommer 1987; Bird 1986). Over two-thirds of all promoters overlap with CpG islands and are mainly associated with housekeeping and some developmental genes (Mohn and Schübeler 2009). As in many other tissues, CGI promoters are typically unmethylated in germline cells, which might explain why these regions resisted CpG depletion throughout evolution (Smallwood et al., 2011; Weber et al., 2007). Exons also have elevated CpG levels compared to introns resulting from selection pressure to maintain codon sequence (Ambrosi, Manzo, and Baubec 2017).

DNA methylation is associated with transcriptional repression

In the 1980s is was observed that DNA methylation negatively correlates with promoter activity and causes stable transcriptional inhibition of exogenous DNA (Busslinger et al., 1983). This pushed forward the idea that DNA methylation is important for transcriptional repression.
Since then technological advances have enabled precise DNA methylation maps of mammalian genomes, revealing that high methylation levels are indeed absent at CGI promoters or TF occupied active regulatory regions (Stadler et al. 2011).

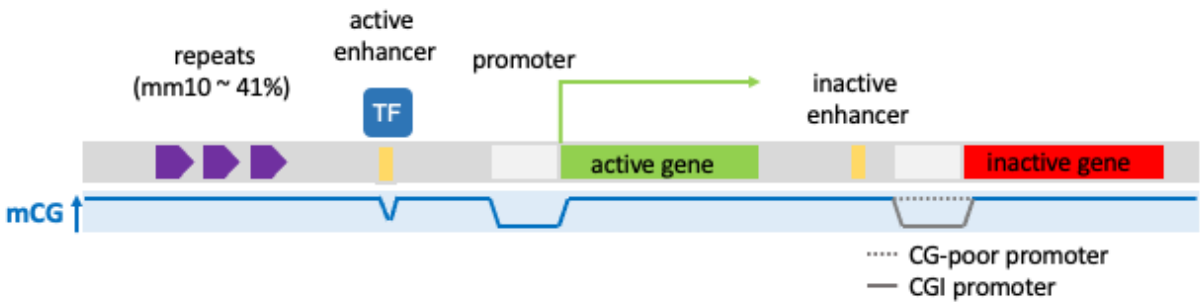


Figure 4. Genomic distribution of DNA methylation in mammals. Genome illustrated with different features (upper grey bar). The graph below (blue and grey lines) depicts methylation levels of CpG dinucleotides. DNA methylation is high genome-wide with exceptions at regulatory regions such as TF bound, active enhancers (low levels of methylations) or CpG island promoters (unmethylated). Adapted from Schübeler 2015.

This implies an intriguing mechanism of DNA methylation mediated gene regulation especially at CGI island promoters that indeed can be repressed by DNA methylation (Schübeler et al., 2000). However, only a fraction of inactive CGI promoters are actually methylated or change their methylation status during development or disease (Mohn et al., 2008; Weber et al., 2007). In human somatic cells about 4% of all CGI promoters are methylated (Weber et al., 2007). Instead, the majority of inactive promoters are marked by Polycomb group proteins mediated H3K27me3 (Lynch et al. 2012; Tanay et al. 2007; Mohn et al. 2008). Both modifications are mutually exclusive (Brinkman et al. 2012) and thought to resemble two modes of repression. H3K27me3 is attributed to be a temporary mode of repression, where genes depending on development or environmental signals can by dynamically activated (Greenberg and Bourc'his 2019). In contrast, DNA methylation is associated with stable and long-term repression of genes (Deaton and Bird 2011). Prominent examples are X chromosome inactivation (Jaenisch and Bird, 2003), genomic imprinting (E. Li, Beard, and Jaenisch 1994) and germline-specific gene silencing. Germline-specific genes represent a small fraction of genes that are tightly controlled by DNA methylation (Mohn et al., 2008; Weber et al., 2007). Their CGI promoters are highly methylated upon differentiation, which is indeed required for their stable repression in somatic tissue (Dahlet et al., 2020; Karimi et al., 2011). In a recent work TET1 was artificially recruited to germline-specific promoters in murine fibroblasts resulting in de-methylation and transcriptional reactivation, providing evidence of a causal link between DNA methylation and repression of endogenous germline-specific genes (Dahlet et al., 2020).

Mechanistic principles of transcriptional repression by DNA methylation

The regulatory role of DNA methylation is unclear

High levels of DNA methylation are absent at virtually all active regulatory region such as enhancers and promoters, suggesting a repressive role for DNA methylation. However, different observations have questioned the generality of this putative repressive function. For instance, while it is established that many genes are repressed by DNA methylation, not all of them reactivate upon loss of this modification, potentially due to the lack of activators (Fouse et al. 2008; Ambrosi, Manzo, and Baubec 2017). Along these lines, CpG-poor active distal regulatory elements (enhancers) display low levels of DNA methylation that result from binding of tissue-specific TFs that cause local hypomethylation (Hodges et al. 2011; Stadler et al. 2011) **(Figure 4)**. It is therefore unclear if DNA methylation in specific genomic contexts is the cause or the consequence of an inactive state. Central to understanding the regulatory role of DNA methylation is the underlying molecular mechanisms by which cytosine methylation translates into transcriptional repression.

A direct vs. an indirect repression model

Two prominent mechanisms are associated with DNA methylation mediated transcriptional repression (Tate and Bird 1993). One model suggest that repression occurs directly via the repulsion of TFs by methylation of the binding motif **(Figure 5a)**. 5mC has similar chemical properties as thymine ("thymine mimicry") and by steric or hydrophobic interference could change the sensitivity of a TF towards its motif (Kribelbauer et al., 2020) . Theoretically, however, such a mechanism would only affect TFs that contain a CpG within their motif. An analysis of database for vertebrate motifs (JASPAR CORE 2018) suggests that 70% contain no prominent CpG and therefore would be unaffected by this mechanism (Héberlé and Bardet 2019).

An additional proposed pathway is that protein repressors recognize methylated DNA and suppress transcription **(Figure 5b)**. Evidence for this model was provided by the identification of methyl-CpG binding domain (MBD) proteins that bind methylated DNA largely in a sequence-unspecific fashion and interact with co-repressors (Tate and Bird 1993). Importantly, both mechanisms are not mutually exclusive and could be at play individually or simultaneously.

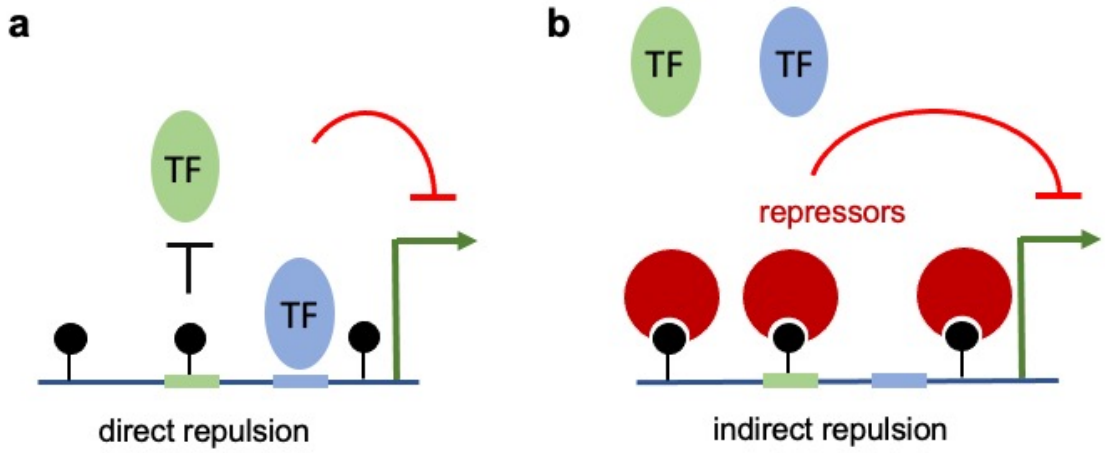


Figure 5. Mechanisms of transcriptional repression by DNA methylation involving a direct and indirect repression model at a CpG island promoter. Black circles represent methylated CpGs. A) While the green transcription factor (TF) is directly repelled by DNA methylation, the blue TF is unaffected as its motif does not contain a CpG that can be methylated. B) In an indirect repression model that involves factors that recognize methylated DNA (e.g., methyl-CpG domain containing proteins), repression occurs indirectly by recruiting co-repressor. Thereby, DNA methylation could also affect TFs that do not bind CpG motifs. Importantly, both models are not mutually exclusive.

Impact of DNA methylation on TF binding

First evidence that DNA methylation impedes TF binding came from *in vitro* experiments that involved gel retardation assays. For instance, Watt and Molloy could show in 1988 that cytosine methylation of a TF binding site prevented complex formation with DNA, but methylation outside of the motif had no such effect (Watt and Molloy, 1988). Since then many *in vitro* experiments provided evidence that DNA methylation can repel TF binding, but also revealed opposite effects namely that 5mC can in some cases promote TF binding even outside the canonical motif (Hu et al., 2013; Iguchi-Ariga and Schaffner, 1989; Mann et al., 2013; Spruijt et al., 2013). Recent high throughput *in vitro* characterizations of TF interactions with methylated DNA ligands, suggest a much more widespread and complex role of DNA methylation in modulating TF binding. For instance, Yin et al. characterized the binding specificities of 519 purified human TFs in full length or only their DBD to methylated and unmethylated DNA ligands using methyl-SELEX and bisulfite-SELEX (Yin et al., 2017). This revealed that 23% of TFs were negatively affected by DNA methylation in their binding, of which 82% contained a CpG in their primary motif. 34 % of TFs showed a binding preference for their methylated ligand and 39 % were not or very little affected by DNA methylation. Using a similar approach, another study analyzed fewer TFs, but investigated position effects of cytosine methylation (Kribelbauer et al., 2017). This revealed that 5mC can have opposing effects on binding of the same TF, depending on which cytosine is methylated.
While these recent *in vitro* findings support the notion that CpG methylation within a motif can generally repel TF binding, they also suggest a much more complex and nuanced influence of DNA methylation. However, how these observations translate into the *in vivo* situation - especially in context of chromatin - and thereby impact gene regulation remains to be tested.

Initial evidence for a methylation sensitive TF *in vivo* came from studying CTCF at the imprinted control region (ICR) *Igf2/H19*. Experimental data suggested that CTCF is methylation sensitive *in vitro* and only binds the unmethylated ICR thereby influencing nearby *Igf2* activity in an allele-specific manner (A. C. Bell and Felsenfeld 2000). However, later experiments revealed that CTCF sensitivity to DNA methylation at this

locus is rather an exception, as genome-wide profiling revealed largely no effect on CTCF binding upon abrogation of 5mC (Maurano et al. 2015; Stadler et al. 2011).
A recent study by Domcke, Bardet et al. compared TF binding events in mouse ES cells that lack DNA methylation by deletion of all three DNA methyltransferases (DNMT TKO) to their isogenic WT counterpart by mapping DNaseI hypersensitive sites (Domcke et al., 2015). This identified multiple TF candidates to be methylation sensitive *in vivo*. Profiling binding of the candidate NRF1 using ChIP-seq in both cell lines revealed that this TF occupies thousands of additional new sites in the unmethylated genome. This validated NRF1 as a TF that is methylation-sensitive genome wide. Of note, the primary motif associated with NRF1 contains two prominent CpGs, suggesting that a direct repulsion mechanism could be at play. Furthermore, the authors could show that binding to a methylated reporter cassette required local demethylation by methylation-insensitive CTCF, establishing a TF cooperativity mediated by DNA methylation. Importantly however, the majority of TFs were unaffected in their binding by the absence of DNA methylation in ES cells. This raises the question if, in differentiated cells, DNA methylation could regulate a much larger set of TFs given the fact that 5mC is essential outside the pluripotent cell state. However, investigating the genome-wide impact of cytosine methylation on TF binding *in vivo* remains challenging, as this would require the removal of DNA methylation in differentiated cells, which at the same time is essential for viability.

Methyl-CpG binding proteins

Three main protein families have been identified to bind methylated DNA: SET and RING (SRA) domain, zinc finger and methyl-CpG (mCpG) binding domain proteins (Buck-Koehntop and Defossez 2013). UHRF1/2 both harbor a SRA domains but, in line with its role for DNA methylation maintenance, UHRF1 only recognizes hemi-methylated DNA, while UHRF2 is suggested to preferentially bind 5hmC (Avvakumov et al. 2008; T. Zhou et al. 2014). Zinc-finger proteins such as ZBTB33 (KAISO), ZBTB38 and ZBTB4 (Sasai, Nakao, and Defossez 2010; Prokhortchouk et al. 2001) or ZFP57 can bind methylated DNA in a sequence-specific context (X. Li et al. 2008; Quenneville et al. 2011). While ZBTB38 and ZBTB34 are linked to oxidative stress response or genomic stability (Roussel-Gervais et al. 2017; Miotto et al. 2018), respectively, ZFP57 binding to methylated DNA is associated with the maintenance of allele-specific methylation at ICRs (Greenberg and Bourc'his 2019). In contrast, MBD proteins are described to recognize mCpGs in an sequence-unspecific context, making them the most interesting candidates to be universal readers of DNA methylation and therefore put them at the center of an indirect repression model (Baubec and Schübeler 2014).

MBD proteins bind methylated DNA in vitro and in vivo

Methyl-CpG-binding protein 2 (MeCP2) was the first protein discovered to bind single and symmetrically methylated CpG dinucleotides via its MBD domain (Lewis et al. 1992). A homology search for MBD-like sequences identified MBD1-6, BAZ2A, BAZ2B, SETDB1 and SETDB2 (Hendrich and Bird, 1998; Roloff et al., 2003).
However, MBD1, MBD2, MBD4 and MeCP2 (henceforth MBD proteins) are the only MBD-containing proteins reported to bind mCpGs *in vitro* and *in vivo* and are therefore the most likely candidates to translate DNA methylation into transcriptional repression (Baubec et al., 2013; Hendrich and Bird, 1998; Laget et al., 2010; Lewis et al., 1992) **(Figure 6)**. The MBD3 protein shares 70% sequence similarity with MBD2 and is most likely derived by a gene duplication event, but lacks affinity towards mCpGs due to amino acid alterations within the MBD (Baubec et al., 2013; Hendrich and Tweedie, 2003; Saito and Ishikawa, 2002). An affinity of MBD3 towards 5hmC has been

proposed yet this is still highly debated (Rausch, Hastert, and Cardoso 2019). Structural investigations of functional MBDs determined that MBD proteins share similar modes of mCpG recognition through base specific contacts, which further supports the notion that methylated CpGs alone determine binding specificity (Buck-Koehntop and Defossez 2013). ChIP-seq experiments of ectopically expressed and biotin-tagged MBD proteins in mouse ES cells and derived neurons indeed revealed that the major determinant for genomic binding *in vivo* is the density of mCpGs and requires a functional MBD (Baubec et al., 2013). MBD1 was proposed to bind unmodified CpGs via its CXXC domains (Jørgensen, Ben-Porath, and Bird 2004). However, the beforementioned *in vivo* study could only detect such binding behavior by artificially removing the MBD. Taken together, density of mCpGs is the best predictor for MBD proteins binding *in vitro* and *in vivo*.

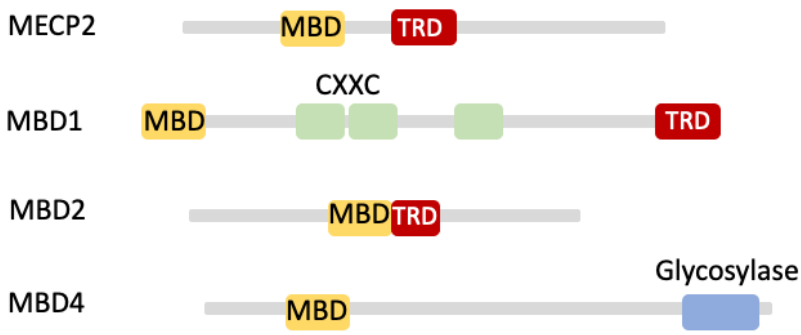


Figure 6. Schematic representation of methyl-CpG binding (MBD) proteins. TRD, transcriptional repressor domain is associates with the interaction of co-repressors. A region within the TRD of MECP2 is also referred to as NCoR1/2 interaction domain (NID). MBD1 additionally contains CXXC domains, that presumably confers MBD1 with binding to unmodified CpGs. The glycosylase domain is unique to MBD4. Adapted from Sasai et al. 2009.

MBD proteins are mainly associated with transcriptional repression in vitro

Early-on biochemical interaction studies revealed that MBD proteins mainly interact with histone deacetylase (HDAC) containing co-repressors. For instance, Nan et al. identified a transcriptional repressor domain (TRD) within MECP2 that interacts with the HDAC containing SIN3A co-repressor complex (Nan et al., 1998) **(Figure 6)**. Recruitment of MECP2's TRD to an episomal reporter gene decreased its activity,

which was partially restored by treatment with an HDAC inhibitor, supporting a model where co-repressors are tethered to methylated DNA via MBD proteins. Further experiments revealed that different MBD proteins interact with both shared and distinct co-repressor complexes. For instance MBD2 was shown to be part of the HDAC1/2 containing nucleosome remodeling and deacetylase (NuRD) and SIN3 complex (Le Guezennec et al. 2006; Y. Zhang et al. 1999). Later it was shown that MECP2 also interacts with the HDAC3-containing NCOR1/2 co-repressor complexes (Lyst et al. 2013). MBD1's repressive properties are associated with histone deacetylation and methylation that are linked to its interaction with the Suv39h1-HP1 complex (Fujita, Watanabe, Ichimura, Tsuruzoe, et al. 2003; Ng, Jeppesen, and Bird 2000). In addition, a deacetylation independent repressive mechanism through interaction with MBD1-containing chromatin-associated factor (MCAF) was also described (Fujita, Watanabe, Ichimura, Ohkuma, et al. 2003). Although it is reported that MBD4s repressive activity involves the recruitment of HDACs (Kondo et al. 2005), its primary function is associated with the repairment of U to G or T to G mismatches in CpG dinucleotide that arise from cytosine or 5mC deamination, respectively (B. Hendrich et al. 1999). This repair function is attributed to the C-terminal glycosylase domain that among the MBD proteins is unique to MBD4 (B. Hendrich et al. 1999). Mice lacking MBD4 activity indeed show a 2-3 fold increase of C to T transitions in the context of CpGs in the small intestine and spleen (Wong et al. 2002). This suggests that MBD4 most likely evolved in order to counteract the mutational load of cytosine methylation.
Taken together, these *in vitro* observations established MBD proteins as bridging molecules that can read methylated DNA and create a heterochromatic environment by the recruitment of co-repressors such as HDACs **(Figure 7)**.

Are MBD proteins functionally redundant in vivo?

Mutations in X-linked *MECP2* are causal for the severe neurological disorder Rett syndrome (RTT) (Tillotson and Bird 2019). Mice lacking *Mecp2* die shortly after birth or in case of heterozygous females display a RTT-like phenotype (Chen et al., 2001; Guy et al., 2001). Except for mice with mutated *Mecp2*, deletion of other MBD proteins that bind 5mC results in viable mice with only mild phenotypes (Hendrich et al., 2001; Millar et al., 2002; Zhao et al., 2003). This is in stark contrast to the phenotype observed when members of the DNA methylation pathway are deleted, that typically

results in embryonic death (Li et al., 1992; Okano et al., 1999; Sharif et al., 2007). It is hypothesized that MBD proteins are functionally redundant which could explain the absence of a strong phenotype when MBD genes are individually deleted. Indeed, MBD proteins are expressed in all major tissues (Hendrich and Bird, 1998). Caballero et al. attempted to address this question by generating mice that simultaneously lack *Mbd2*, *Mecp2* and *Kaiso* (termed 3KO) (Martín Caballero et al. 2009). 3KO mice went through embryogenesis and were born with the expected RTT phenotype and died shortly after birth. No upregulation of other mCpG binding protein transcripts were observed in neuronal stem cells derived from 3KO embryos, arguing against a compensatory mechanism. Although the authors did note that any essential function could be carried out by the remaining methyl-CpG biding proteins, including MBD1. Therefore, it remains unclear if MBD proteins are functionally redundant or potentially only play a limited role in DNA methylation mediated repression *in vivo*.

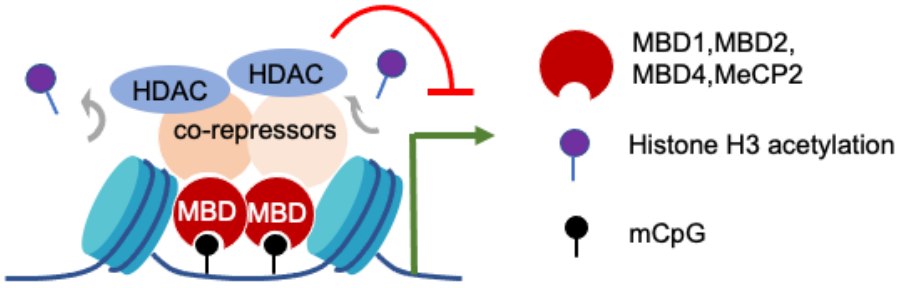


Figure 7. MBD proteins are at the center of an indirect repression model. All four mCpG binding MBD proteins (MBD1, MBD2, MBD4 and MeCP2) bind densely methylated DNA and are suggested to translate DNA methylation into transcriptional repression. The main mechanism is associated with the recruitment of co-repressors including histone deacetylases (HDACs) that in turn promote heterochromatin formation.

Are MBD proteins effectors of DNA methylation in vivo?

Despite the strong biochemical evidence that supports the bridging function of MBD proteins, the limited phenotype of mice with a reduced number of MBD proteins questions their function *in vivo.* Therefore, it is currently unclear how critical these proteins are for DNA methylation mediated repression in the cellular context. While MBD protein mediated repression is especially conceivable to be at play at methylation dense elements such as evolutionary young and CpG-rich transposons or CGI promoters, different observations have challenged this view.

For instance, conditional knockout experiments in mice revealed that RTT can be primarily attributed to the absence of MeCP2 in the brain (Tillotson and Bird 2019). MeCP2 is highly expressed in neurons where it reaches a molecule number similar to histone levels (Skene et al. 2010). Recently, it was shown that the MBD and the co-repressor interaction domain are not only sufficient to prevent a RTT-like phenotype in mice, but can also reverse the phenotype when reactivated (Tillotson et al. 2017). While this supports the bridging theory of MBD proteins, further investigations revealed that MeCP2 function is not attributed to mCpG binding, but rather owed to the dual binding specificity of MeCP2's MBD towards methylated CAC (mCAC) (Tillotson et al., 2021). Indeed, mCAC levels are particularly high in neurons and derive from an unspecific *de novo* activity of DNMT3A (Ramsahoye et al. 2000; Tillotson and Bird 2019). How MeCP2 affects neuronal function remains unclear on a molecular level, but emerging studies imply a limited role of mCpGs.

Another observation was made in the beforementioned study from Baubec et al. where the authors also investigated the genomic co-occupancy of MBD2 and NuRD (Baubec et al., 2013). This revealed that these two factors only co-localize at unmethylated regions of primarily active promoters and enhancers. Strikingly, they could show that MBD2 binding to these unmethylated regions depends on interaction with NuRD. These findings challenge a bridging model where MBD proteins recruit co-repressors to methylated DNA.
In conclusion, the *in vivo* role of MBD proteins translating DNA methylation into transcriptional repression remains elusive.

3. Scope of this thesis

DNA methylation is associated with transcriptional repression, yet the underlying mechanism remains elusive. Accumulating evidence suggests that motif methylation directly influences binding of many TFs *in vitro*. However, functional *in vivo* evidence remains scarce, especially in somatic cells. An alternative mechanism of DNA methylation mediated repression proposes that factors read methylated DNA, recruit co-repressors that in turn establish a repressive chromatin environment. This indirect repressive mechanism is appealing, as it would also affect binding of TFs that are devoid of CpG dinucleotides within their motif. Importantly, both mechanisms are not mutually exclusive. At the center of the indirect repression model are methyl-CpG binding domain (MBD) proteins as they bind methylated DNA and biochemically interact with co-repressors. However, as a comprehensive deletion of all functional MBD proteins is missing we lack genetic proof that this protein family indeed translates DNA methylation into transcriptional repression *in vivo*.

In the first experimental part of this thesis, I addressed this question by generating a mouse ES cell line that lacks all 5mC binding MBD proteins (MBD1, MBD2, MBD4 and MeCP2). In addition, I generated a methylation deficient ES cell line, since an indirect repression model would predict similar or overlapping phenotypes. Using a forced differentiation protocol, I then derived viable neurons from both cell lines. By comparative genomics I comprehensively studied the resulting phenotypes in stem and somatic cells and provide insights into the role of MBD proteins in DNA methylation mediated repression.

I then explored the impact of DNA methylation on TF binding in methylation deficient neurons and search for evidence of a direct repulsion mechanism.

As part of these efforts, I systematically dissect the contribution of a methylation-sensitive TF in regulating the activity of TEs in somatic cells, that highly depend on DNA methylation for their suppression.

Lastly, I further investigate the role of MBD proteins outside of the murine lineage. To this end, I simultaneously delete 5mC binding MBD proteins in human somatic cells and characterize consequences on transcription, the regulatory landscape and DNA methylation.

4. Results and Discussions

Direct inhibition of transcription factor binding is the dominant mode of gene and repeat repression by DNA methylation (prepared manuscript)

Direct inhibition of transcription factor binding is the dominant mode of gene and repeat repression by DNA methylation

Sebastian Kaluscha[1,2]*, Silvia Domcke[3]*, Christiane Wirbelauer[1], Lukas Burger[1,4] and Dirk Schübeler[1,2]

1: Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH 4058 Basel
2: University of Basel, Faculty of Sciences, Petersplatz 1, CH 4003 Basel,
3: Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA.
4: Swiss Institute of Bioinformatics, 4058 Basel, Switzerland

*These authors contributed equally to this work

Correspondence should be addressed to: dirk@fmi.ch

Methylation of cytosines (5mC) efficiently silences CpG rich regulatory regions in mammalian genomes and is required for the repression of several classes of repetitive elements in somatic cells. It remains open to what extent this epigenetic phenomenon relies on either direct inhibition of binding of transcription factors (TFs) or on indirect inhibition via recruitment of Methyl-CpG Binding domain proteins (MBD) to 5mC. Here we show that combinatorial genetic deletions of all proteins with functional MBDs does not reactivate genes or repeats silenced by DNA methylation in either mouse embryonic stem cells or derived neurons. In contrast, absence of DNA methylation activates a set of genes in both cell types and causes rampant transcription of distinct subfamilies of repeats in neurons. Loss of H3K9me3 at repeats during differentiation could explain this differential reliance on DNA methylation for repeat silencing across cell types. Comparative analysis of genome accessibility nominates several TFs as causal for the upregulation of genes and repeats, which we validate by showing methylation dependent binding for several TFs including HNF6. We further identify the CRE motif as a characteristic feature of upregulated repetitive elements and show that it enhances LTR-driven transcription in the absence of DNA methylation. Among several investigated TFs binding this motif, we show that CREB1 causally contributes to upregulation. Collectively this expands the list of TFs that are sensitive to DNA methylation *in vivo*. They put forth a model where direct interference with TF binding is the dominant mechanism of repression by DNA methylation at regulatory regions and repeats likely accounting for the essentiality of DNA methylation in differentiated cells.

Introduction

Mammalian genomes are characterized by high levels of DNA methylation, with more than 80% of cytosines in the context of CpG dinucleotides carrying this modification. It has been suggested that DNA methylation evolved as a sort of 'genomic immune system', to defend the host against the invasion of virus DNA and transposable elements (Bestor, 1990). Recognizing and methylating these foreign repetitive DNA sequences, which make up 50-70% of our genome (Lander et al. 2001; Padeken, Zeller, and Gasser 2015), enables their transcriptional repression and prevents their further expansion within the host genome. This silencing mechanism appears to have been co-opted by vertebrates for other means of transcriptional repression: Methylation of CpG-dense gene promoters has been shown to cause robust transcriptional repression (Busslinger et al., 1983; Schübeler et al., 2000) and is at the basis of the two established incidents of long-term mono-allelic silencing (Illingworth and Bird 2009): X chromosome inactivation (Jaenisch and Bird, 2003; Panning and Jaenisch, 1996) and genomic imprinting (Li et al. 1993; Bourc'his et al. 2001), and is also associated with silencing of tumor suppressor genes in cancer (Jones 2012).

Although the ability of methylation to repress transcription at CpG-rich regulatory regions and repeats is undisputed, the mechanism of how this is achieved remains unclear in light of two pathways, which are not mutually exclusive: On the one hand, methylation could block transcription in an indirect manner through methyl-CpG-binding domain proteins (MBDs) recognizing dense arrays of methylated CpGs and recruiting histone deacetylases (Klose and Bird, 2006; Nan et al., 1998). This in turn would lead to chromatin compaction and thus exclusion of transcription factors (TFs) independent of their sequence motifs. There are four core MBD family proteins in mammals, which contain a methyl-CpG-binding domain and have been shown to bind methylated DNA in vitro and in vivo (Hendrich and Tweedie, 2003; Klose and Bird, 2006): MeCP2, MBD1, MBD2, and MBD4. To date, no combined genetic deletion of all four MBDs has been reported. Loss of individual MBDs only results in mild phenotypes in mice (Hendrich et al., 2001; Millar et al., 2002; Zhao et al., 2003), with

the exception of MeCP2 which is the causal gene for Rett syndrome (Amir et al., 1999; Chen et al., 2001; Guy et al., 2001). It has been proposed that functional redundancy between the MBDs explains the absence of severe transcriptional upregulation in the single or combinatorial knockouts generated thus far (Fatemi and Wade, 2006; Hendrich and Tweedie, 2003) in line with the observation that the extent of genomic methylation generally correlates with the number of MBD proteins in a species (Hendrich and Tweedie, 2003).

On the other hand, methylation of cytosines within a sequence motif could directly obstruct TF binding by affecting the shape and base readout of the matching TF (Dantas Machado et al. 2014). While sensitivity of some TFs to methylation of their binding site was indeed observed *in vitro* (Bednarik et al., 1991; Campanero et al., 2000; Iguchi-Ariga and Schaffner, 1989; Prendergast et al., 1991; Watt and Molloy, 1988) we still lack structural evidence to support a direct disruptive effect of the methyl-group on these protein-DNA interactions and it remains unclear if methylation-sensitivity of TFs *in vivo* is the rule or the exception (Domcke et al., 2015; Yin et al., 2017).

Defining the contribution of different pathways to methylation-mediated silencing has been hampered by the essentiality and/or redundancy of various players. Firstly, complete removal of DNA methylation by deletion of the enzymes setting this mark, DNA methyltransferases (Dnmt1, Dnmt3a, Dnmt3b), leads to rapid cell death in differentiated vertebrate cells or human embryonic stem (ES) cells (Chen et al., 2007; Li et al., 1992; Liao et al., 2015). Being cellular essential has been attributed to misregulation of critical genes (Jackson-Grusby et al. 2001), activation of repeats (Yoder, Walsh, and Bestor 1997), or was linked to the induction of DNA damage (Shaknovich et al. 2011) and mitotic catastrophe (T. Chen et al. 2007). The only vertebrate cell type known to be viable in culture without DNA methylation are mouse ES cells, which represent an earlier stage of development than their human counterparts (Nichols and Smith 2009; Tsumura et al. 2006; Liao et al. 2015): They are isolated from preimplantation blastocysts, whose genomes are globally

demethylated (Auclair and Weber 2012), so mechanisms need to remain in place at this developmental stage to ensure cellular function in spite of low DNA methylation levels. These DNMT-TKO mouse ES cells, while viable, have been reported to be unable to differentiate (Tsumura et al. 2006). Secondly, even the availability of a methylation-free model system does not enable distinguishing between direct and indirect modes of repression. Teasing apart the contribution of these modes requires contrasting repression abilities in wildtype cells, cells without DNA methylation (both modes affected) and cells without MBDs (only indirect mode affected).

Here we define in both pluripotent and terminally differentiated cells the extent to which DNA methylation-mediated transcriptional repression functions through an direct or indirect mode. We generated cells that lack all functional MBD proteins by performing four consecutive deletions using CRISPR. We then monitored the resulting effect on gene and repeat repression as well as cell viability in both stem cells and derived postmitotic neurons using a rapid differentiation system that enables to generate neuronal cells that lack DNA methylation. This reveals that deleting all functional MBD proteins has only a very minor impact on gene expression and chromatin accessibility not only in stem cells but also in neurons. Absence of DNA methylation in neurons however causes activation of genes controlled by methylated CpG islands but also rampant transcription of repeats. This entails reorganization of the accessibility landscape driven by TFs that are methylation sensitive. Experimental validation identifies novel candidates of epigenetically restricted TFs and reveals a causal role for the methylation sensitive TF CREB in repeat upregulation.

Results

ES cells are viable in absence of all proteins with a functional MBD

To study the role of MBD-proteins in DNA methylation-mediated transcriptional repression we sought to generate cells that lack all 5mC-binding MBD proteins. Since mouse embryonic stem (ES) cells tolerate the loss of DNA methylation and enable measurement of differential gene expression (Karimi et al. 2011; Tsumura et al. 2006), we reasoned mouse ES cells should likewise be amenable to a comprehensive deletion of readers of this epigenetic mark.

More specifically we focused on MBD1, MBD2, MBD4 and MECP2 (henceforth MBD proteins) and excluded MBD3 as it does not recognize 5mC (Baubec et al., 2013; Hendrich and Bird, 1998; Hendrich et al., 2001; Saito and Ishikawa, 2002). By sequential CRISPR targeting we generated two independent mouse stem cell clones that are quadruple knockout of all four MBD protein genes (MBD-QKO), each derived with a different set of gRNAs. Sequencing confirmed that these introduced frameshift mutations into exons encoding the MBD (**SuppFig 1a**). Absence of MBD proteins was validated by western blot (**Fig. 1a**).

MBD-QKO ES cells are viable in culture and show no difference in regards to proliferation morphological appearance (**Fig. 1b**). When we tested the effect of MBD protein loss on the transcriptome using RNA-seq we were surprised to observe that the transcriptome of MBD-QKO ES resembles that of WT ES cells (**Fig. 1c, SuppFig 1b**). Only two genes are reproducibly up-regulated compared to the WT between clones, while only slightly more genes are down-regulated (n= 33, including *Mbd1*, *Mbd2* and *Mbd4,* **SuppFig 1c**). The latter are enriched in processes such chromatin silencing and negative regulation of gene expression **(SuppFig 1d upper panel)**. To ask if absence of the MBD proteins affects the accessibility of chromatin we performed ATAC-seq in both MBD-QKO and WT ES cells **(Fig.1d and SuppFig 1e)**. Again this revealed very limited changes, in line with the lack of transcriptional response (**Fig. 1d**). We conclude that deleting all functional MBD proteins in mouse stem cells has a very limited effect on transcription and local chromatin accessibility.

In order to compare MBD-QKO to ES cells lacking DNA methylation we similarly deleted *Dnmt1/3a/3b* in ES cells using CRISPR/Cas9 (henceforth DNMT-TKO), rendering ES cells free of CpG methylation. In contrast to the MBD-QKO ES cells, DNMT-TKO ES cells show downregulation of 504 genes and upregulation of 849 genes by RNA-seq (**Fig 1c**). Upregulated genes are enriched for being gamete-specific (**SuppFig 1d, lower panel**) in line with previous observations that many of these are controlled by CpG-rich promoters that are methylated in soma (Weber et al., 2007) but activated in DNMT-TKO ES cells (Domcke et al., 2015; Karimi et al., 2011).

Analysis of the chromatin accessibility landscape by ATAC-seq in DNMT-TKO ES cells identifies several thousand regions that are methylated in wildtype and located distally from promoter distal but gain accessibility (**Fig 1d and SuppFig 1f**), in line with our previous study using DNase-seq (Domcke et al., 2015). This can be at least in part attributed to the binding of methylation sensitive TFs such as NRF1 (Domcke et al., 2015) or BANP (Grand et al., submitted), both of which contain CpGs within their motif. Of note, while we observe an increase in accessibility around NRF1 and BANP sites we do not see this gain in MBD protein depleted cells (**Fig 1e**).

In summary, while ES cells tolerate the loss of both MBD proteins or DNA methylation writers, only absence of DNA methylation affects significantly the accessibility of the regulatory landscape and the transcriptome.

Viable neurons with distinct transcriptional phenotypes can be generated in the absence of DNA methyltransferases or MBD proteins

Mouse ES cells are unique as they proliferate in culture in absence of DNA methylation while in differentiated cells DNA methylation becomes essential at the cellular level (Chen et al., 2007; Egger et al., 2006; Fan et al., 2001; Li et al., 1992; Okano et al., 1999; Sen et al., 2010). This is likely due to the presence of DNA methylation independent pathways in mES that repress repeats in the germline and early embryogenesis via SETDB1 dependent histone methylation of lysine 9 of H3 (Karimi et al., 2011; Leung and Lorincz, 2012; Rowe et al., 2010). The presence of such alternative pathways of repression might also account for the lack of phenotype upon removal of MBD proteins. To test this hypothesis requires generating differentiated

cells from both knockout lines. This is hindered by the fact that DNMT-TKO cells do not differentiate according to reports in the literature as well as our attempts (data not shown) to generate neurons using retinoic acid based protocols that span multiple weeks of differentiation (Jackson et al. 2004; Tsumura et al. 2006). This is in line with the observation that DNA hypmethylation in the adult brain causes lethality in neurons (Ramesh et al. 2016; Hutnick et al. 2009). We hypothesized in the planning of this project that a rapid differentiation regime might enable us to generate methylation free neuronal cells. Towards this goal we employed neuronal differentiation via ectopic expression of a neurotrophic TF (*Ngn2*), previously shown to generate functional glutamatergic neurons within five days following induction (Thoma et al. 2012; Yingsha Zhang et al. 2013). This approach utilizes a dox-inducible *Ngn2* expression cassette, which we integrated into the parental ESC line from which we generated all subsequent DNMT-TKO and MBD-QKO ESC lines **(Fig. 2a)**. This enabled us to subject both mutants to identical differentiation conditions and contrast this to the parental WT clone.

Upon induction of *Ngn2* expression, both DNMT-TKO and MBD-QKO cells exited the cell cycle, adopted neuronal morphology, and formed axonal networks similar to the WT within about 3 days (**Fig 2b**). We conclude that a rapid neurogenesis paradigm allows us to generate neuronal cells *in vitro* in absence of DNA methylation or MBD proteins. This enables phenotyping the effect of the knockouts on transcriptome and epigenome in a differentiated cell state. While the absence of MBD proteins did not affect long-term cell viability of the neurons, DNMT-TKO neurons only survived for around ten days (**SuppFig 2a**).
MBD-QKO neurons are remarkably similar to WT neurons at the level of the transcriptome, with minor but reproducible changes (56 genes up, 177 down, FDR <= 0.01 and foldchange >= 2) across both MBD-QKO clones (**Fig. 2c and SuppFig. 2b**). Importantly, genes differentially expressed in MBD-QKO cells tend to be unmethylated at their promoters in wildtype cells and already expressed in WT neurons (**Fig. 2d and SuppFig 2e**). A gene ontology analysis within the downregulated set of genes reveals an enrichment in different pathways for tissue development (**SuppFig 2d**).

The transcriptome of DNMT-TKO neurons globally resembles that of WT neurons arguing that they, in line with their morphology, acquired a neuronal identity **(Figure**

2c and SuppFig 2b). At the same time, they are more dissimilar to WT than the MBD-QKO neurons (**SuppFig 2b**) since they display a roughly ten times larger set of differentially expressed genes **(Fig 2c)**. Genes upregulated in DNMT-TKO neurons tend to harbor promoters that are methylated and inactive in WT (**Fig 2d and SuppFig 2e**) and are again enriched for gamete-specific genes (**SuppFig 2c**). Prominent examples are *Dazl* or *Asz1* which are established to rely on promoter methylation for repression in somatic cells (Dahlet et al., 2020; Weber et al., 2007). Importantly these genes do not respond to absence of MBD proteins **(Figure 2c)**.

Among the small set of genes that change expression in MBD-QKO neurons 66 % (116/177) of down- but only 43 % (24/56) of upregulated ones are also dysregulated in DNMT-TKO. However, these genes tend to be unmethylated (**Fig 2d and SuppFig 2e**), expressed in WT neurons (**Fig 2d**) and together only account for 7 % of all genes differentially expressed in DNMT-TKO neurons. Collectively, this distinct transcriptional phenotype argues against a prominent role of MBD proteins in maintaining or establishing repression of genes with methylated promoters that become activated in absence of DNA methylation in neurons.

The chromatin accessibility landscape in neurons changes in response to loss of methylation but not MBD proteins

In analogy to our experiments in stem cells we next determined changes of the chromatin landscape in neurons using ATAC-seq in absence of MBD proteins or DNA methylation.

This revealed that MBD-QKO neurons are highly similar to wildtype cells with surprisingly few yet reproducible changes **(Figure 3a and SuppFig 3a)**. This is in line with the limited transcriptional phenotype and suggests that absence of MBDs has limited effects on genic expression and that it does not activate regulatory regions.

In contrast, DNMT-TKO neurons display several thousand differentially accessible regions **(Figure 3a and SuppFig 3a)**. In line with our observation in ES cells, more regions gain than lose accessibility (7128 up, 5627 down). DNMT-TKO specific sites tend to be further away from TSS (**SuppFig 3b**), are less frequently overlapping CpG

islands (**SuppFig 3c**) and less broad than shared sites (**SuppFig 3d**). This is in line with most changes occurring at distal regulatory regions or non-functional sites which are methylated (**SuppFig 3e)** and unoccupied in WT conditions (**Figure 3a**).

The observed changes in regulatory regions relate to transcriptional changes since genes closest to differentially accessible neuronal sites show a significant increase in expression in DNMT-TKO neurons compared to genes next to constitutively open regions (**SuppFig 3f**). This is in line the repressive effect of DNA methylation and is in agreement with the tendency towards upregulation in differential gene expression. We conclude that even in somatic cells absence of MBDs has no global impact on transcription and accessibility at regulatory regions. In contrast, absence of DNA methylation causes widespread changes suggesting that the contribution of MBD proteins to methylation dependent silencing of regulatory regions is minor.

Nomination of candidate methylation-sensitive TFs

Having shown that MBDs do not account for methylation dependent changes in regulatory activity argues that upregulation in DNMT-TKO happens by relieving direct inhibition of TFs. If indeed the case, we should be able to identify some of these. To this end, we performed motif discovery in the top DNMT-TKO specific ATAC-seq peaks. This identified 36 known TF motifs (**Fig 3c)**, many with high sequence similarity (**Fig 3d and SuppFig 3g**). Among the top enriched motifs is the one for NRF1 similar to stem cells in line with this TF being methylation sensitive and also expressed in neurons. Other prominent motifs are only distinct for neurons and include HNF6 (aka ONECUT1). Several enriched motifs do not contain a CpG, which is unexpected as as we do not expect this for a methylation sensitive TF. This illustrates that motif occurresence with differential accessible regions, while useful to nominate TFs, does require subsequent experimental validation.

Hnf6 is a methylation-sensitive TF

We first focused on HNF6 as its motif is nominated to be methylation sensitive in neurons but not in stem cells, where this particular factor is not expressed. HNF6 is strongly upregulated upon differentiation in several tissues and is considered a key regulator in liver, pancreas and the nervous system (Audouard et al. 2013). The canonical HNF6 motif does not contain a CpG, while a motif variant has been reported with a prominent CpG (Ballester et al. 2014; Wang et al. 2014). Indeed, performing an unbiased enrichment analysis of HNF6 motif hexamers reveals that only the CpG-containing motif is enriched in DNMT-TKO specific ATAC peaks suggesting that this particular motif variation is methylation sensitive (**SuppFig 3g**). To test this we performed genomic location analysis by ChIP-seq in wild-type and DNMT-TKO neurons **(Fig 4a and SupFig 4a)**. Comprehensive analysis of the resulting datasets revealed ~1300 sites that showed binding only in absence of DNA methylation, while only ~380 sites displayed reduced binding (**Fig 4a and SupFig. 4b**). Generally DNMT-TKO enriched sites are promoter distal, similar to WT peaks, and in agreement with the differential ATAC-seq analysis also gain accessibility in DNMT-TKO neurons (**SupFig 4c,d**). Importantly, the top identified motif in the DNMT-TKO enriched peaks matches the CpG-variant (**Fig. 4b**). The motif identified across all DNMT-TKO peaks appears to be a mix of both canonical and CpG-variant and resembles a motif identified by HT-SELEX suggesting that HNF6 binds both motifs equally well in absence of DNA methylation (Jolma et al. 2013) (**Fig. 4b)**.

To further explore motif occurrences we calculated the frequencies of the canonical or variant HNF6 motif sequence (ATTGAT or ATCGAT) in all peak regions. As expected, most wildtype peaks were enriched for the canonical HNF6 motif, whereas a lower percentage contained the CpG-variant (**Fig. 4c**). In DNMT-TKO peaks the CpG-variant frequency increases by more than two-fold; 80% of peaks enriched in DNMT-TKO harbor the CpG-variant (**Fig. 4c, SupFig 4e)**.

To ask whether CpG motif variant methylation is instructive for HNF6 binding we calculated the methylation frequencies of all CpG-variants bound in any condition. This revealed that most variants are fully methylated in WT neurons, and show the largest

HNF6 binding increase in DNMT-TKO neurons, whereas HNF6 binding is largely unaffected at unmethylated motifs (**Fig 4d,e**). This establishes that HNF6 is methylation sensitive but restricted to the CpG-containing motif variant, readily supporting a model of direct TF repulsion by DNA methylation within the binding site. This model of direct inhibition is further supported by the fact that the presence of neighboring CpGs is not predictive of binding in absence of DNA methylation **(SuppFig 4f)**. We conclude that similar to our previous observation in stem cells (Domcke et al., 2015), methylation sensitive TFs bind to additional motif occurrences in postmitotic cells that lack DNA methylation revealing novel TFs that are restricted by genomic methylation.

DNA methylation dependent and MBD independent repeat derepression in neurons

Having focused on genic transcription we next asked how expression of repetitive sequences are affected in neurons upon absence of DNA methylation or MBD protein affects. We first mapped all RNA reads to the repeat annotation of RepeatMasker, and assigned non-unique hits randomly to one location in the reference genome. This revealed no upregulation in repeat expression upon deletion of the MBD proteins (**Fig 5a,c SupFig 5a**) in line with their limited roles in affecting transcription of genes that are silenced by DNA methylation. Absence of the DNMTs and DNA methylation on the other hand revealed a dramatic increase of repeat derived RNA **(Fig 5a)**. Closer analysis revealed that especially certain members of the ERVK family, namely Intracisternal A-type particle (IAP) elements, are highly derepressed in DNMT-TKO neurons (**Fig 5b**), now making up a substantial part of repeat derived RNA (**Fig 5a**). Considering only uniquely mapped reads confirms this upregulation of individual IAP retrotransposons (**SupFig 5b,c**). In agreement with our observation in differentiated ES cells, similar de-repression of repeats was observed *in vivo* in murine *Dnmt1*$^{-/-}$ embryos (Dahlet et al., 2020; Walsh et al., 1998) or in conditional UHRF1 depleted postnatal mouse cortex (Ramesh et al. 2016) (**Fig 5c**).

In contrast to the striking expression levels of IAP elements observed in DNMT-TKO neurons, depletion of DNA methylation in ES cells only causes limited de-repression of IAP elements (**Fig 5c**). Genetic deletion experiments have shown that this DNA methylation independent repression requires tri-methylation of lysine 9 at histone H3

set (H3K9me3) by KAP1-SETDB1 at repeats since absence of this pathway results in IAP upregulation and cell death (**Fig 5c**) (Karimi et al., 2011; Leung and Lorincz, 2012; Rowe et al., 2010; Sharif et al., 2016). Therefore, we asked if the change to methylation dependent repression during differentiation coincides with reduced function of H3K9me3.

Profiling of H3K9me3 levels in neurons by ChIP-seq indeed reveals reduced H3K9me3 levels at IAP elements compared to ES cells **(Fig 5d and SuppFig 5d)**. This suggests that down-regulation of H3K9me3 likely accounts for the fact that DNA methylation becomes necessary for repeat repression upon cellular differentiations.

CRE is important for IAPLTR1/1a activity

Having established that silencing of endogenous retroviruses occurs independently of MBD proteins suggests that repeat repression in neurons could similarly involve the blocking of methylation sensitive TFs. To address this question, we asked if repeat copies that do respond to absence of methylation are enriched for particular TF motifs.

Given their strong depression upon removal of DNA methylation we focused on IAP elements. ERVK class II IAP proviruses are characterized by a 5' and 3' Long Terminal Repeat (LTR) (300-1000bp) enclosing an internal sequence (6 – 9kb), that encodes viral proteins (Gag, Pol, Env) (Mager and Stoye 2015). RepeatMasker lists a fragmented annotation of IAP proviruses, where the internal region, and its flanking LTRs are individually annotated. We curated the RepeatMasker annotation by combining LTR and internal fragments belonging to the same element (**suppFig 6a and see methods**). This allowed us to count uniquely mapping RNA reads in annotation-curated IAP elements and to assign the transcriptional activity to the corresponding 5' LTR, as the promoter of an IAP provirus. This assignment revealed that almost all IAPLTR1/1a elements gain activation (FDR =< 0.05 and foldchange >= 2, **SuppFig 6b**), while other IAP element types with lower LTR sequence similarity are activated to a lower extent in DNMT-TKO neurons (**SuppFig 6c**). Having linked the transcriptional activity of a IAPLTR1/1a containing IAP elements to the associated LTRs, we searched for known JASPAR motifs in the 5' LTR and asked whether motif presence was associated with differences in RNA expression levels.

This nominated multiple candidates with the cyclic AMP response element (CRE) being the top significant motifs that separated high from low expressing repeats in both IAPLTR1/1a elements (**SuppFig 6d**). Categorizing repeats based on the presence of the perfect CRE sequence (TGACGTCA) in the 5' LTR revealed that IAPLTR1/1a elements with this sequence are significantly higher expressed in DNMT-TKO neurons (**Fig 6a**).

To functionally test the contribution of CRE for repeat activity at an individual LTR we generated reporter constructs containing the fully conserved IAPLTR1a sequence with or without the CRE upstream of a Luciferase gene (**Fig 6b)**. For optimal comparison we placed both constructs as single-copy integrants into both WT and DNMT-TKO ES cells at a defined genomic site by recombinase-mediated cassette exchange (RMCE) and in parallel as a control the same reporter driven by a promoter of a housekeeping gene (PGK).
In WT and DNMT-TKO the PGK promoter is equally active, while the IAP is silent in WT and only weakly expressed in DNMT-TKO (**SuppFig 6e)** recapitulating the behavior of its endogenous counterparts. Upon differentiation into neurons this repression is preserved in WT cells, while we observe strong IAP reporter upregulation in neurons derived from DNMT-TKO ES cells **(SuppFig 6b)**. Thus the reporter constructs also mimic the endogenous IAPLTR1a upon terminal differentiation. Importantly loss of the CRE reduces activity by half, in both ES cells and neurons. We conclude that the CRE motif is critical for full IAPLTR1/1a activity in absence of DNA methylation.

Methylation-sensitive CREB1 binds CRE in IAP elements

It has previously been reported that *in vitro* binding to the CRE element by an at the time unknown TF is impeded by DNA methylation of its single CpG (Iguchi-Ariga and Schaffner, 1989). Importantly, CRE resembling motifs are among the most strongly enriched motifs in all DNMT-TKO specific ATAC-seq sites (**Fig 3d,e and SuppFig 3e**). Selectively profiling accessibility around CRE genome-wide showed an unidirectional gain in accessibility in DNMT-TKO neurons for hundreds of sites (**SuppFig 6f**), implying this motif is indeed bound preferentially in absence of DNA methylation.

Multiple TFs of the basic leucine zipper TF family are predicted to bind CRE as homo- or heterodimers (Hai and Hartman 2001). While some mostly bind CRE related sequences (Steven et al., 2020), the Cyclic AMP (cAMP)-response element-binding protein 1 (CREB1) was shown to preferentially bind CRE as a homodimer or its half-site motif as a monomer in genic promoters (X. Zhang et al. 2005; Montminy and Bilezikjian 1987) and viral genes (Tierney et al. 2000; Millhouse et al. 1998; Kirby, Rickinson, and Bell 2000). CREB1 is ubiquitously expressed throughout mouse tissues (Steven et al., 2020; Yamamoto et al., 1990) as well as in ES cells and Ngn2 neurons.

In order to determine the influence of DNA methylation on CREB1 genomic binding we compared CREB1 occupancy by ChIP-seq in WT and DNMT-TKO neurons. This revealed that binding sites across both cell lines are highly correlated (**SuppFig 6g**), enriched for CRE or CRE half-sites (**SuppFig h,i)** and located almost exclusively in CpG rich and unmethylated promoters of active genes (**SuppFig 6k,l).** Gene set enrichment analysis determined that these genes are associated with general cellular functions (e.g. mRNA processing, chromatin modifications) **(SuppFig 6n)**, in line with CREB1 being ubiquitously expressed across different tissues (Steven et al., 2020; Yamamoto et al., 1990). While we only identify three sites that are bound exclusively in WT, we identify 116 newly bound sites in DNMT-TKO neurons (**SuppFig 6j**), that are mainly located in or proximal to promoters (**SuppFig 6k)**.
Scoring methylation of CRE full or half-sites bound in WT versus DNMT-TKO revealed that CREB1 binding is inversely correlated with motif methylation in WT (**SuppFig 6m**) and DNMT-TKO specific binding appears at sites that are methylated in WT (**Fig 6d**). Therefore, we conclude that CREB1 is indeed methylation sensitive *in vivo*, in agreement with previous reports that determined a negative effect of CRE methylation on CREB1 binding *in vitro* (Mancini et al., 1999; Spruijt et al., 2013; Tierney et al., 2000; Yin et al., 2017).

Next we sought to determine if CREB1 binds 5' LTR regions of IAPLTR1/1a elements in absence of DNA methylation. The repetitive nature of retroviruses and the short fragment lengths of ChIP-seq libraries makes it inherently challenging to uniquely map reads to accurately quantify factor occupancy. Therefore, we first profiled POL2 binding in WT and DNMT-TKO neurons in order to benchmark our ability to measure

changes in factor occupancy in repetitive regions by ChIP-seq (**SuppFig 7a**). Counting uniquely mapped reads in 5' LTR regions of IAP elements that are derepressed in absence of DNA methylation (measured by RNA-seq; FDR <= 0.05 and foldchange >= 2, **SuppFig 7b**) revealed only minor POL2 changes at most elements in absence of DNA methylation (**SuppFig 7d),** indicative of the limited sensitivity to detect changes in factor occupancy at repeats by ChIP-seq. However, a small subset of elements show larger and reproducible POL2 changes that are consistent with RNA changes (**SuppFig 7d,f)**. These include IAPLTR1/1a and IAPEY elements, which are of high copy number amongst de-repressed repeats (**SuppFig 6c**), displaying a reproducible and more than two-fold enrichment in POL2 binding (**Fig 6e and SuppFig 7f**).

This is accompanied by local increased accessibility (**Fig 6e**). As expected, we do not detect POL2 binding in WT neurons (**Fig 6e**). This shows our ability to measure changes in factor occupancy in 5' LTRs by ChIP-seq, such as in IAPLTR1/1a and IAPEY elements.

To ask whether CREB1 binds in IAPLTR1/1a elements we repeated this analysis with uniquely mapping reads derived from profiling CREB1 binding by ChIP-seq. This revealed reproducible counts between replicates (**SuppFig 7c**) and a positive correlation (R=0.66) between CREB1 binding and the fraction of CRE or half-site motifs present in 5' LTRs (**SuppFig 7e**). IAPLTR1/1a and IAPEY elements with 5' LTRs, which nearly all comprise CRE or half-site motifs, indeed show binding of CREB1 only in absence of DNA methylation (**Fig 6e and SuppFig 7f)**. We conclude that CREB1 binds to IAP repeats upon genomic loss of DNA methylation.

Deletion of CREB1 results in reduced activity at genes and IAP repeats

Having shown that the CRE motif is required for strong activity of an LTR and that CREB1 binds to the CRE motif in absence of DNA methylation we next wanted to test directly the contribution of CREB1 to repeat activity. Using CRISPR/Cas9 we deleted *Creb1* in DNMT-TKO ES cells and generated neurons by NGN2 induction (**SuppFig 8a**). The resulting transcriptional profile revealed a high correlation in genic expression between CREB1 depleted and WT DNMT-TKO neurons (**Fig Supp 8b**), with 109 down- and 79 upregulated genes (FDR <= 0.01 and foldchange >= 2, **SuppFig 8c**).

To distinguish primary from secondary targets we focused our analysis on genes with CREB1 bound in their promoter. Among these the majority of responding genes are down- (n=46) rather than upregulated (n=8) (FDR < 0.01 and fold change >=2) (**SuppFig 8c)**, in line with CREB1 being described primarily as an transcriptional activator (Steven et al., 2020).

Downregulated genes include *Ddx4* a germline-specific gene that is under control of a CGI that normally is methylated in soma. It is upregulated and bound by CREB1 only upon removal of DNA methylation but again downregulated when CREB1 is deleted (**SuppFig 8c**). This provides a genic example of CREB1 mediated upregulation upon loss of DNA methylation.

Chromatin accessibility (ATAC-seq) in CREB1 depleted neurons and matching control cell lines (**SuppFig 8d**) reveals that sites that are newly bound by CREB1 and increase in accessibility upon removal of DNA methylation (**SuppFig 8e, left panel**), decrease in accessibility upon CREB1 deletion (**SuppFig 8e, right panel**). Thus CREB1 responds to genome demethylation by binding to new sites leading to increased chromatin accessibility and transcriptional activation.

Decreased accessibility is similarly observed in 5' LTRs of IAPLTR1/1a in DNMT-TKO upon removal of CREB1 (**Fig 6f left panel**), in line with reduced transcriptional activity (**Fig 6f, right panel**). Interestingly, further analysis at all de-repressed IAP elements likewise revealed a reduction of accessibility over the 5' LTR regions (**SuppFig8f left panel)** accompanied by a loss of expression **(SuppFig 8f, right panel)**, that is inversely correlated with CREB1 binding.

Taken together, this shows that motif methylation of CRE abrogates binding of CREB1 in IAP repeats that in absence of DNA methylation significantly contributes to activity providing a prominent example of repeat repression by blocking TF binding through motif methylation.

Discussion

In this study we asked to what extend the repression of regulatory regions by DNA methylation depends on direct inhibition of binding of transcription factors (TFs) or on indirect inhibition via recruitment of Methyl-CpG Binding domain proteins (MBD) to 5mC. We show that deletion of all functional MBDs in both stem cells and differentiating cells leads to a very small transcriptional response that seems not linked to sites of methylation. Conversely, complete loss of DNA methylation results in upregulation of a subset of genes controlled by otherwise methylated CGI promoters in stem and differentiated cells where it also causes rampant transcription of endogenous retroviruses. We proceed to identify novel candidate TFs that are blocked from binding by DNA methylation and activate retroviruses in its absence. These results argue that direct impediment of TF binding is a dominant mechanism of methylation-mediated transcriptional repression of regulatory regions in the cell states studied.

Concurrent deletion of all MBDs does not cause a transcriptional phenotype or obvious changes in the accessibility of the chromatin landscape in mouse stem cells and derived neurons. While this strongly suggests that in the cell systems tested MBD proteins are largely dispensable for methylation mediated repression of regulatory regions it does not preclude that MBD proteins could be involved in other aspects of gene regulation. For example, MeCP2 has been suggested to impact alternative splicing (Maunakea et al., 2013; Young et al., 2005) and transcriptional elongation involving methyl-CA binding (Cholewa-Waclaw et al., 2019; Gabel et al., 2015; Lagger et al., 2017; Tillotson et al., 2021) or influence micro RNA processing (Cheng et al. 2014). While we can rule out functional redundancy between the four MBD proteins as a reason for absence of transcriptional and chromatin accessibility phenotypes, it is possible that other to date uncharacterized sequence agnostic methyl-CpG binders exist and are able to mediate indirect repression, thus we cannot fully exclude this mechanism of action. It remains further conceivable that MBD proteins participate in stabilizing aspects of transcriptional repression in a way that is redundant in the cell systems we have employed.

In contrast, we do observe methylation of CpGs within specific motifs interfering with TF binding, as evidenced by gains in chromatin accessibility, TF binding and

transcription upon removal of DNA methylation, both genome-wide and in reporter assays testing the role of individual TF motifs. In addition to constitutively expressed TFs shown to be methylation-sensitive *in vivo* at their canonical motif (NRF1, BANP, CREB1), with HNF6 we find an interesting case of a developmental TF that is only methylation-sensitive at a CpG-containing motif variant, not its canonical motif. While the methylation sensitive nature of HNF6 was previously reported in an *in vitro* methyl-SELEX study (Yin et al., 2017), our approach provides quantitative insights into the relative contributions of these variants to the *in vivo* HNF6 binding landscape in the presence and absence of DNA methylation.

Structural analysis of CREB1 (Schumacher, Goodman, and Brennan 2000) and HNF6 (Iyaguchi et al. 2007) exist in complexes with unmethylated DNA. Although from different TF families, both proteins interact mainly with the major groove of the DNA, where the methyl-group of the cytosine is positioned (Dantas Machado et al. 2014), leading to groove widening (Kribelbauer et al., 2020; Tippin and Sundaralingam, 1997). Indeed CREB1 is unable to bind a motif variant where the central cytosine is replaced with a thymidine, which structurally resembles methyl-cytosine, in contrast to other closely related bZIP proteins capable of binding this CRE variant (Benbrook and Jones 1994). Of note, methylation can also change the DNA shape at neighboring base pairs, thus affecting binding for motifs that do not contain central CpGs (Kribelbauer et al., 2017, 2020). Although comparison with ancestral genomes reveals ongoing depletion of CpG-containing TF motifs (Żemojtel et al. 2011; Smith et al. 2007), it is tempting to speculate that methylation directly restricting binding at select TF motifs can function to mediate TF hierarchies (Domcke et al., 2015) or specifically regulate different motif variants in a cell type-specific manner, thus expanding the gene regulatory toolkit at a subset of sites.

It is unclear if the essential nature of DNA methylation in differentiated cells (Kribelbauer et al., 2020) is driven by aberrant gene expression (Jackson-Grusby et al. 2001) or repeat activation in its absence (Yoder, Walsh, and Bestor 1997). Both might be linked to mitotic catastrophe, which has been suggested to be at the center of rapid cell death in the absence of DNA methylation (T. Chen et al. 2007). Of note, our methylation-devoid neuronal cells are in fact postmitotic for several days before cell death, making such a direct link to a cell cycle checkpoint unlikely. Repeat

activation is the key feature that distinguishes DNMT-TKO neurons, which are unable to survive for many days, from DNMT-TKO ES cells, which do not show a phenotype. Activation of transposable elements can potentially induce cell death in several ways, e.g. by sheer transcriptional load or insertion of active ERVs into genes or promoter regions, thus producing mutants or high levels of chimeric transcripts (Timothy H. Bestor 2003).

We provide evidence that release of direct inhibition of TF binding at key motifs, such as CREB1 binding at the CRE, is the predominant mechanism of repeat activation. The differences in repeat activation we observe across differentiation stages could be explained as follows: In pluripotent cells, ERVs are recognized by KRAB zinc-finger proteins or the piRNA pathway and are both H3K9 as well as DNA methylated (Molaro and Malik 2016). Even when undergoing a period of low methylation, as occurs naturally after fertilization and in primordial germ cells, or upon complete loss in DNMT-TKO ES cells, these repeats remain mostly silent. Since IAPLTRs are activated upon simultaneous loss of both DNA methylation and H3K9me3 in ES cells (Sharif et al., 2016), it is likely that the TFs responsible for driving the expression are already expressed at this stage, but largely prevented from binding through H3K9me3. During differentiation, H3K9me3 is depleted at repeats, but these are still silent since DNA methylation prevents binding of TFs. However, if DNA methylation is now removed at these elements, due to genetic manipulation or disease, TFs can bind and induce transcription.

The presence of other repeat silencing pathways such as H3K9me3 or m6A RNA methylation (Chelmicki et al. 2021) enables vertebrates to undergo periods of global low methylation in the germline, allowing them to reset the (epi)genome to a basic, totipotent state before establishing sex-specific and germ cell-specific epigenetic signatures and transcription profiles (Messerschmidt, Knowles, and Solter 2014). At the same time, however, transcription and transposition in the germline is required for genomic expansion of TFs and thus for their evolutionary “success”, since their transcriptional activity in somatic cells would reduce fitness of the host without germline fixation of the TE expansions. Indeed there is measurable transcriptional activity of ERVs in both the mouse and human germline (Brûlet et al. 1983; Dupressoir and Heidmann 1996; Göke et al. 2015; Grow et al. 2015; Seisenberger et al. 2012;

Tang et al. 2015). It is thus tempting to speculate that transcriptional control by a methylation-sensitive TF provides evolutionary benefit for the TE as it ensures expression in the germline but repression in somatic cells. It also enables use of a ubiquitously expressed strong activator such as CREB1 without having deleterious effects in somatic cells and might potentially contributing to the fact that IAP elements are the most successful TEs in the mouse genome (Maksakova et al. 2006).

While IAP elements do not exist in humans, the larger family of ERV-K elements has a human counterpart, the HERVK LTR retrotransposons. This is indeed the only ERV family member that has continued to replicate in the human population (Marchi et al. 2014). It is lowly but detectably expressed during normal human embryogenesis as well as in many cancers, some autoimmune/ inflammatory diseases and HIV-infected cells (Grow et al. 2015; Wildschutte et al. 2016). Interestingly, several human LTR retrotransposons contain CRE motifs, and CREB or ATF/AP-1 factors have been implicated in driving expression of human ERVs, Human T cell leukemia virus type 1 and HIV (Caselli et al. 2012; Grant et al. 2006; Toufaily et al. 2015). CRE methylation has also been linked to promoter silencing of the Epstein-Barr virus genome (Tierney et al. 2000).

Taken together, our study provides insights into the general mechanisms of transcriptional repression through DNA methylation at genes and repeats. DNA methylation likely evolved as a means to repress repetitive elements, therefore insight into this process can educate us on how and whether methylation-mediated silencing has been co-opted at other regulatory regions as well as its misregulation in disease.

Figures 1-6

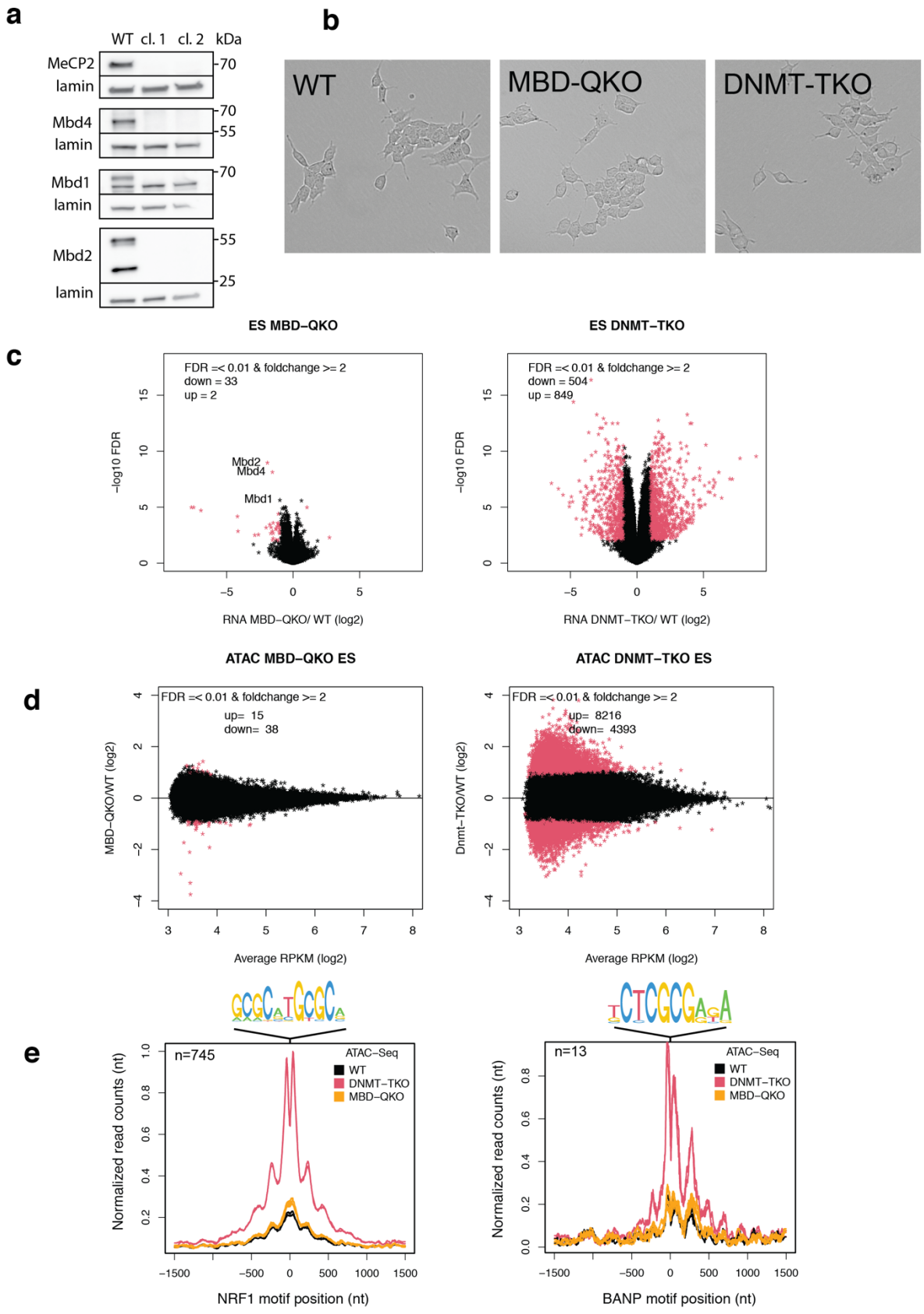

a
WT cl. 1 cl. 2 kDa
MeCP2
lamin
70
Mbd4
lamin
70
55
Mbd1
lamin
70
Mbd2
lamin
55
25
b
WT
MBD-QKO
DNMT-TKO
c
ES MBD–QKO
FDR =< 0.01 & foldchange >= 2
down = 33
up = 2
Mbd2
Mbd4
Mbd1
–log10 FDR
RNA MBD–QKO/ WT (log2)
ES DNMT–TKO
FDR =< 0.01 & foldchange >= 2
down = 504
up = 849
–log10 FDR
RNA DNMT–TKO/ WT (log2)
d
ATAC MBD–QKO ES
FDR =< 0.01 & foldchange >= 2
up= 15
down= 38
MBD–QKO/WT (log2)
Average RPKM (log2)
ATAC DNMT–TKO ES
FDR =< 0.01 & foldchange >= 2
up= 8216
down= 4393
Dnmt–TKO/WT (log2)
Average RPKM (log2)
e
n=745
ATAC–Seq
WT
DNMT–TKO
MBD–QKO
Normalized read counts (nt)
NRF1 motif position (nt)
n=13
ATAC–Seq
WT
DNMT–TKO
MBD–QKO
Normalized read counts (nt)
BANP motif position (nt)

Figure 1. ES cells are viable without 5mC-binding MBD proteins with limited changes in transcription and chromatin accessibility

a) Western blot detecting different MBD proteins in WT or two MBD-QKO clones derived from two different sets of gRNAs. Lamin serves as loading control. Nuclear extracts from mES cells or, for MECP2, from ES derived neurons.
b) Images for wild-type, MBD-QKO or DNMT-TKO mES cells.
c) Gene expression changes in mES cells deleted for MBD proteins or DNA methylation shown as volcano plots.
d) Mean chromatin accessibility (ATAC-seq) versus accessibility change for mES cells deleted for MBD proteins or DNA methylation compared to WT.
e) Chromatin accessibility changes in different cell lines at NRF1 (745 motifs) or BANP (13 motifs) motifs that gain accessibility in DNMT-TKO mES cells. Two replicates for each cell line are depicted in black (WT) or red (DNMT-TKO), while replicates of both MBD-QKO clones are shown in the same color (orange).

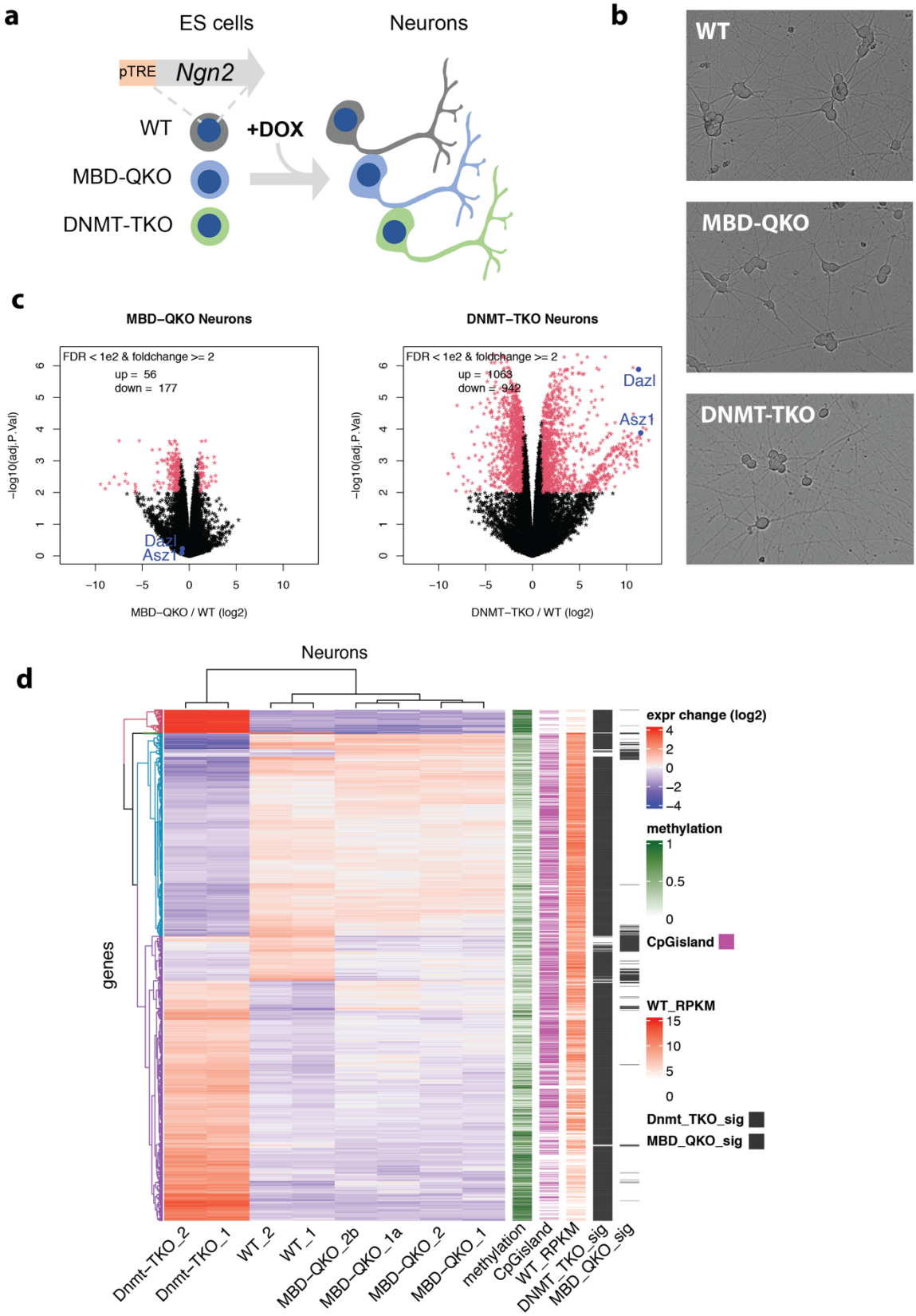

a
ES cells
Neurons
pTRE
Ngn2
WT
+DOX
MBD-QKO
DNMT-TKO
b
WT
MBD-QKO
DNMT-TKO
c
MBD−QKO Neurons
DNMT−TKO Neurons
FDR < 1e2 & foldchange >= 2
up = 56
down = 177
up = 1063
down = 942
Dazl
Asz1
−log10(adj.P.Val)
MBD−QKO / WT (log2)
DNMT−TKO / WT (log2)
d
Neurons
genes
expr change (log2)
methylation
CpGisland
WT_RPKM
Dnmt_TKO_sig
MBD_QKO_sig
Dnmt−TKO_2
Dnmt−TKO_1
WT_2
WT_1
MBD−QKO_2b
MBD−QKO_1a
MBD−QKO_2
MBD−QKO_1

Figure 2. MBD or DNMT deleted neurons show distinct transcriptional phenotypes

a) ES cells carrying a *Ngn2* expression cassette under the control of a pTRE-tight can be rapidly differentiated towards neurons by doxycycline (DOX) induction.
b) Wildtype, MBD-QKO or DNMT-TKO neurons eight days following dox induction.
c) Gene expression changes in neuron cells deleted for MBD proteins or DNA methylation compared to WT. Germline-specific genes are highlighted in blue.
d) Hierarchical clustering of genes differentially expressed (FDR <1e2 & foldchange >=2) in either MBD-QKO or DNMT-TKO neurons (n = 2089). Each row in the heatmap depicts expression fold changes (log2) relative to the mean expression across all samples. Methylation bar, promoter methylation of each gene in WT neurons. Green shading represents fraction of methylated CpGs. CpG island bar, promoters that overlap with a CpG island are depicted in purple. WT_RPKM bar, genes expression in WT neurons. Red shading indicates expression levels in log2 (RPKM). DNMT_TKO_sig or MBD_QKO_sig bars, a gene is significantly down- or upregulated (in black) in either DNMT- or MBD knockout neurons or both compared to WT.

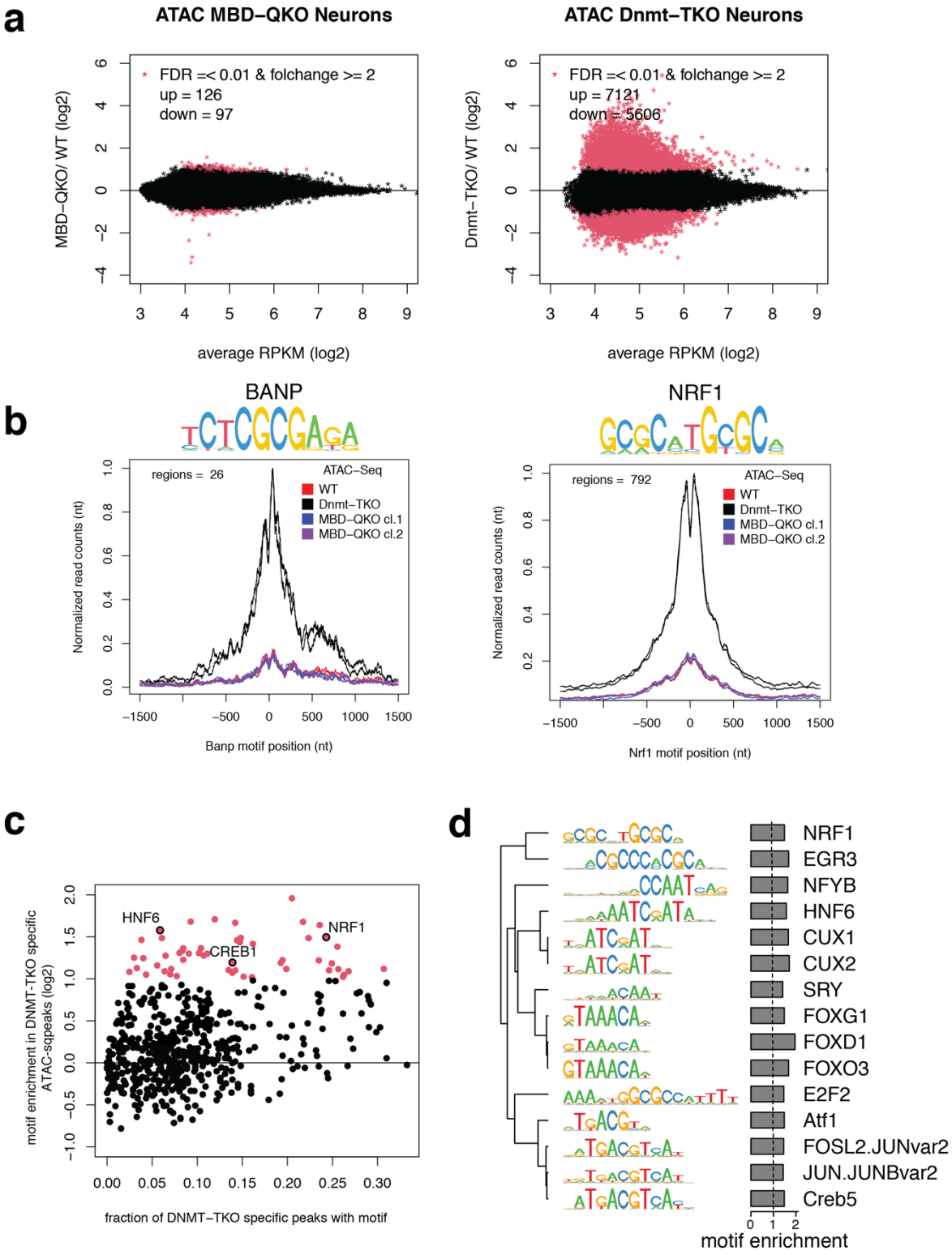

a
ATAC MBD−QKO Neurons
ATAC Dnmt−TKO Neurons
FDR =< 0.01 & folchange >= 2
up = 126
down = 97
up = 7121
down = 5606
MBD−QKO/ WT (log2)
Dnmt−TKO/ WT (log2)
average RPKM (log2)
b
BANP
NRF1
regions = 26
regions = 792
ATAC−Seq
WT
Dnmt−TKO
MBD−QKO cl.1
MBD−QKO cl.2
Normalized read counts (nt)
Banp motif position (nt)
Nrf1 motif position (nt)
c
HNF6
CREB1
NRF1
motif enrichment in DNMT−TKO specific ATAC−sqpeaks (log2)
fraction of DNMT−TKO specific peaks with motif
d
NRF1
EGR3
NFYB
HNF6
CUX1
CUX2
SRY
FOXG1
FOXD1
FOXO3
E2F2
Atf1
FOSL2.JUNvar2
JUN.JUNBvar2
Creb5
motif enrichment

Figure 3. The chromatin accessibility landscape in neurons changes in response to loss of DNA methylation but not MBD proteins.

a) Mean chromatin accessibility (ATAC-seq) versus accessibility change for neurons lacking MBD proteins or DNA methylation compared to WT. For differential accessibility analysis all replicates from both MBD-QKO clones were considered as equal.
b) Chromatin accessibility changes in different cell lines at motifs of NRF1 (792 motifs) or BANP (26 motifs) that gain accessibility in DNMT-TKO mES cells. Two replicates for each cell line are depicted in red (WT) or black (DNMT-TKO), while replicates of both MBD-QKO clones are shown in blue or purple.
c) Motif enrichment (see methods) in DNMT-TKO specific ATAC-seq peaks vs. the fraction of DNMT-TKO specific peaks with a motif. Strong DNMT-TKO specific ATAC-seq peaks (FDR < 0.01 & foldchange > 8) outside of CpG islands were selected (n= 972). Red points indicate motifs that are enriched (FDR < 0.01 and foldchange > 2) in DNMT-TKO specific peaks (n=36).
d) Unbiased clustering of the top 15 motifs enriched in DNMT-TKO specific ATAC-seq peaks from c) by weight matrix similarity.

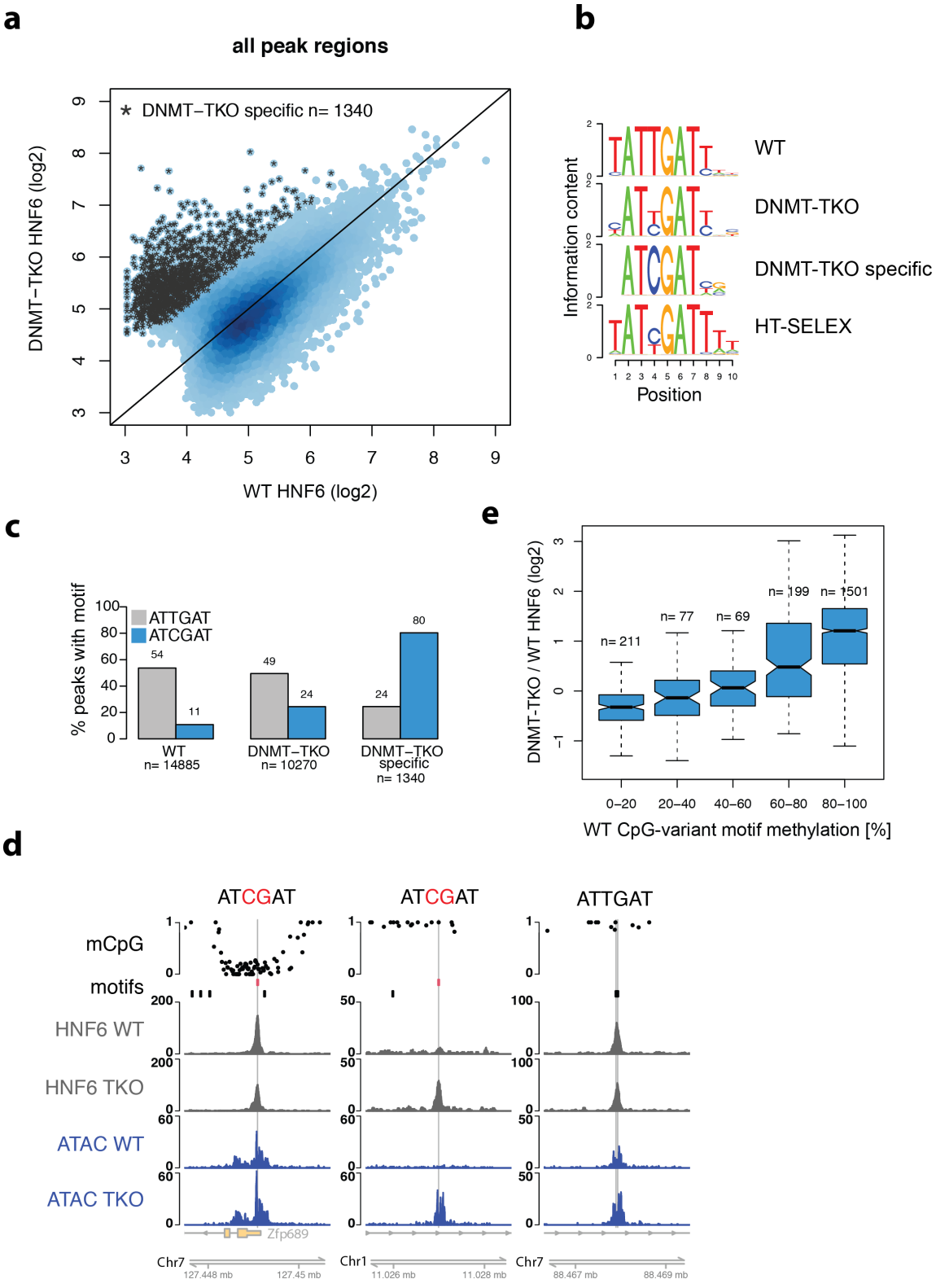

a
all peak regions
* DNMT−TKO specific n= 1340
DNMT−TKO HNF6 (log2)
WT HNF6 (log2)
b
Information content
WT
DNMT-TKO
DNMT-TKO specific
HT-SELEX
Position
c
% peaks with motif
ATTGAT
ATCGAT
54
11
49
24
24
80
WT
n= 14885
DNMT−TKO
n= 10270
DNMT−TKO
specific
n= 1340
e
DNMT-TKO / WT HNF6 (log2)
n= 211
n= 77
n= 69
n= 199
n= 1501
0–20
20–40
40–60
60–80
80–100
WT CpG-variant motif methylation [%]
d
ATCGAT
ATCGAT
ATTGAT
mCpG
motifs
HNF6 WT
HNF6 TKO
ATAC WT
ATAC TKO
Zfp689
Chr7
127.448 mb
127.45 mb
Chr1
11.026 mb
11.028 mb
Chr7
88.467 mb
88.469 mb

Figure 4. HNF6 is methylation-sensitive at a CpG-containing motif variant

a) Wild-type and DNMT-TKO HNF6 ChIP-seq signal at all peak regions. Black asterisks represent regions that are DNMT-TKO specific in neurons.
b) Motifs enriched in the top 500 WT, DNMT-TKO or DNMT-TKO specific HNF6 peak regions. *In vitro* identified human HNF6 motif by HT-SELEX (Jolma et al. 2013).
c) Bar plot of HNF6 peaks with canonical or CpG-containing motif variant, indicating that the variant is enriched in DNMT-TKO specific peaks.
d) Single locus examples of HNF6 binding (grey tracks) in wild-type and DNMT-TKO (TKO) neurons at canonical motifs (black) or CpG-containing motif variants (red). Top track indicates methylation frequencies of individual CpGs (black dot). ATAC-seq tracks in blue for WT and DNMT-TKO neurons. HNF6 only binds its motif variant in absence of DNA methylation. Read counts in running windows of 51nt (ChIP-seq) or 11nt (ATAC-seq), replicate data combined.
e) Change in HNF6 binding between DNMT-TKO and wild-type at all peak regions grouped according to their average motif methylation.

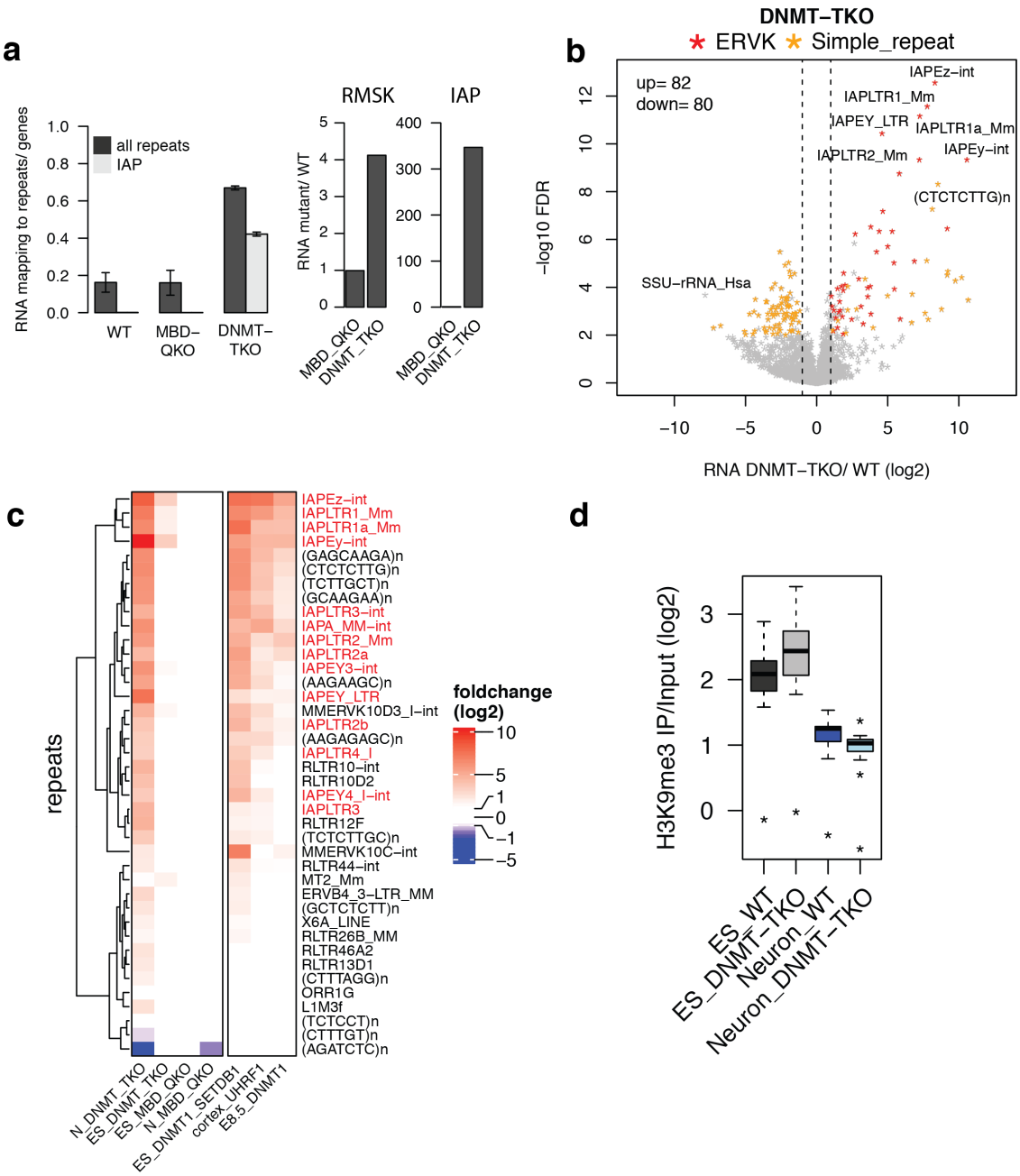

a
RNA mapping to repeats/ genes
all repeats
IAP
WT
MBD–QKO
DNMT–TKO
RMSK
IAP
RNA mutant/ WT
MBD_QKO
DNMT_TKO
b
DNMT–TKO
ERVK
Simple_repeat
up= 82
down= 80
IAPEz–int
IAPLTR1_Mm
IAPEY_LTR
IAPLTR1a_Mm
IAPLTR2_Mm
IAPEy–int
(CTCTCTTG)n
SSU–rRNA_Hsa
–log10 FDR
RNA DNMT–TKO/ WT (log2)
c
repeats
IAPEz–int
IAPLTR1_Mm
IAPLTR1a_Mm
IAPEy–int
(GAGCAAGA)n
(CTCTCTTG)n
(TCTTGCT)n
(GCAAGAA)n
IAPLTR3–int
IAPA_MM–int
IAPLTR2_Mm
IAPLTR2a
IAPEY3–int
(AAGAAGC)n
IAPEY_LTR
MMERVK10D3_I–int
IAPLTR2b
(AAGAGAGC)n
IAPLTR4_I
RLTR10–int
RLTR10D2
IAPEY4_I–int
IAPLTR3
RLTR12F
(TCTCTTGC)n
MMERVK10C–int
RLTR44–int
MT2_Mm
ERVB4_3–LTR_MM
(GCTCTCTT)n
X6A_LINE
RLTR26B_MM
RLTR46A2
RLTR13D1
(CTTTAGG)n
ORR1G
L1M3f
(TCTCCT)n
(CTTTGT)n
(AGATCTC)n
foldchange (log2)
N_DNMT_TKO
ES_DNMT_TKO
ES_MBD_QKO
N_MBD_QKO
ES_DNMT1_SETDB1
cortex_UHRF1
E8.5_DNMT1
d
H3K9me3 IP/Input (log2)
ES_WT
ES_DNMT–TKO
Neuron_WT
Neuron_DNMT–TKO

Figure 5 Repeats are de-repressed in neurons in absence of DNA methylation but not in absence of MBD proteins

a) Ratio of 75 bp paired-end reads mapping to regions from RepeatMasker or genic regions (overlapping regions were excluded) in neurons (left barplot). If a read mapped to multiple regions (max. 100) a region was chosen by random. Standard deviation indicated of WT (n=2), MBD-QKO (n=4, both clones combined), DNMT-TKO (n=2). Change of repeat expression in wild-type versus mutant neurons for all repeats from RepeatMasker (middle) or IAP elements (right).
b) Differential expressed repeat subfamilies in DNMT-TKO neurons using multimapping reads. Dashed lines indicate twofold expression change. Two most abundant repeat subfamilies differentially expressed (FDR =< 0.01 and foldchange >=2) are colored.
c) Heatmap of RNA expression fold changes of mutant/perturbation over WT. Rows depict repeat subfamilies that are differentially expressed in DNMT-TKO neurons (FDR < 1e-06). Left box includes Ngn2 cells where N denotes neuron and ES embryonic stem cell. Right box includes published RNA expression data. ESC_DNMT1_SETDB1, conditional deletion of Setdb1 and DNMT1 in ESC from Sharif et al. 2016. Cortex_UHRF1, conditional deletion of Uhrf1 in cerebral cortex of postnatal day 5 mice from Ramesh et al 2016. E8.5_DNMT1, *Dnmt1* mutant embryos at day 8.5 from Dahlet et al, 2020. For counting multimapping RNA-seq reads (single end, 50bp) were considered.
d) Box plot of H3K9me3 ChIP-seq signal at IAP elements in different neuronal cell lines. Multimapping reads were counted in IAP elements annotated by RepeatMasker and summed by RepName (n=25). ES WT (n=2), ES DNMT-TKO (n=2), Neuron WT (n=2), Neuron DNMT-TKO (n=1).

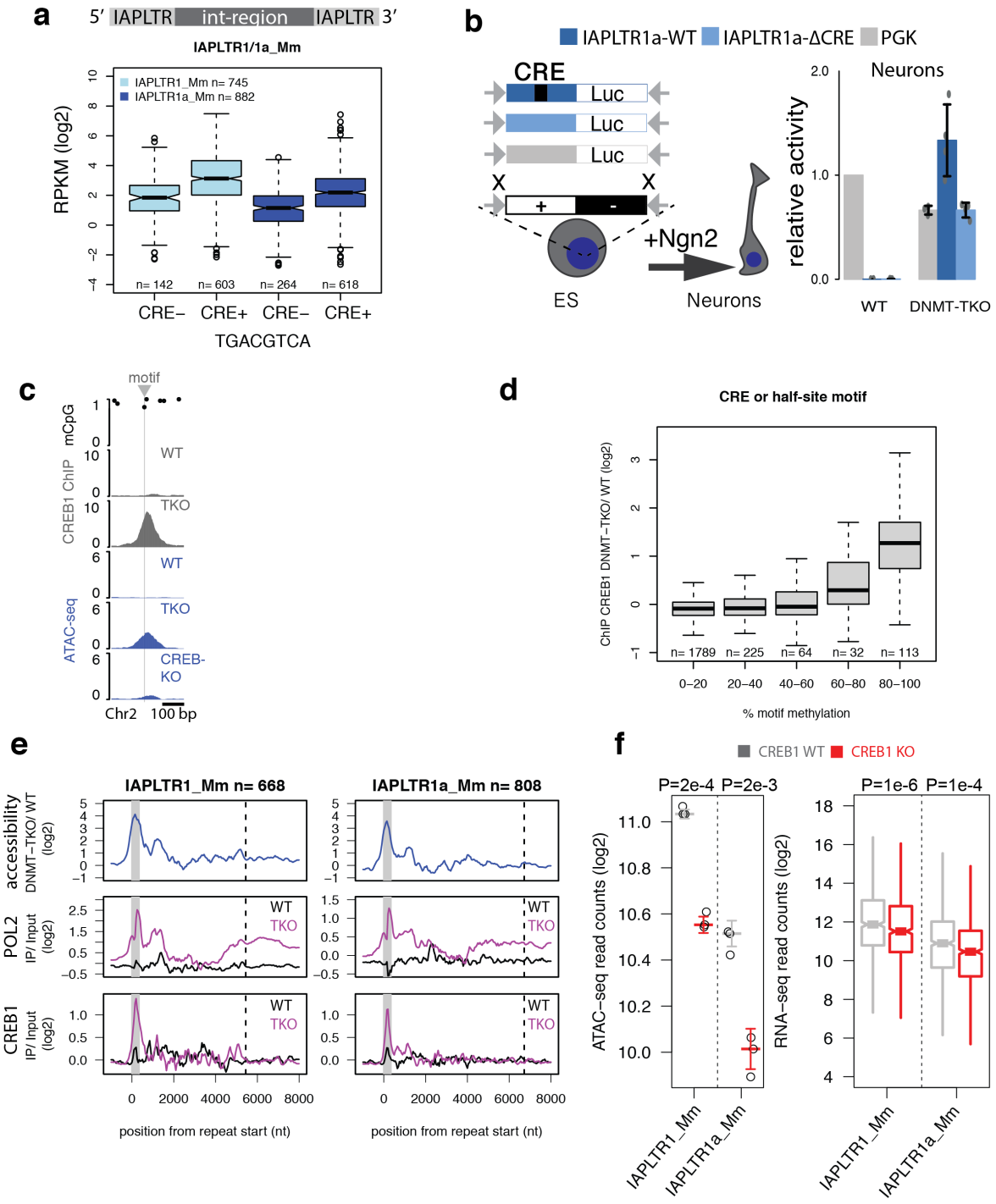

a
5′ IAPLTR int-region IAPLTR 3′
IAPLTR1/1a_Mm
IAPLTR1_Mm n= 745
IAPLTR1a_Mm n= 882
RPKM (log2)
n= 142
n= 603
n= 264
n= 618
CRE−
CRE+
CRE−
CRE+
TGACGTCA
b
IAPLTR1a-WT
IAPLTR1a-ΔCRE
PGK
CRE
Luc
Luc
Luc
X
X
+
-
+Ngn2
ES
Neurons
Neurons
relative activity
WT
DNMT-TKO
c
motif
mCpG
CREB1 ChIP
WT
TKO
ATAC-seq
WT
TKO
CREB-KO
Chr2
100 bp
d
CRE or half-site motif
ChIP CREB1 DNMT−TKO/ WT (log2)
n= 1789
n= 225
n= 64
n= 32
n= 113
0–20
20–40
40–60
60–80
80–100
% motif methylation
e
IAPLTR1_Mm n= 668
IAPLTR1a_Mm n= 808
accessibility DNMT−TKO/ WT (log2)
POL2 IP/ Input (log2)
CREB1 IP/ Input (log2)
WT
TKO
position from repeat start (nt)
position from repeat start (nt)
f
CREB1 WT
CREB1 KO
P=2e-4
P=2e-3
P=1e-6
P=1e-4
ATAC−seq read counts (log2)
RNA−seq read counts (log2)
IAPLTR1_Mm
IAPLTR1a_Mm
IAPLTR1_Mm
IAPLTR1a_Mm

Figure 6. CREB1 binds methylation-sensitive to IAPLTR1/1a elements and contributes to repeat activity

a) Illustration (top) of an IAPLTR1 or 1a element with 5' and 3' LTR flanking the internal region (int-region). Box plots (bottom) showing expression of IAPLTR1/1a elements grouped by having a perfect match with the CRE motif (TGACGTCA) in the 5' LTR. Number of elements indicated by n.

b) Scheme illustrating single-copy and stable integration of different IAPLTR1a and PGK reporter construct by recombinase-mediated cassette exchange (RMCE) into ES cells and differentiated to neurons. Bar plot on the right showing reporter activity in neurons, indicating that IAPLTR1a reporters are silent in WT but active in DNMT-TKO neurons. Importantly, absence of CRE reduces the reporter activity by half.
WT_PGK (n=5), WT_LTR (n=3), WT_LTR_∆CRE (n=3), DNMT-TKO_PGK (n=3), DNMT-TKO_LTR (n=4), DNMT-TKO_LTR_∆CRE (n=4)

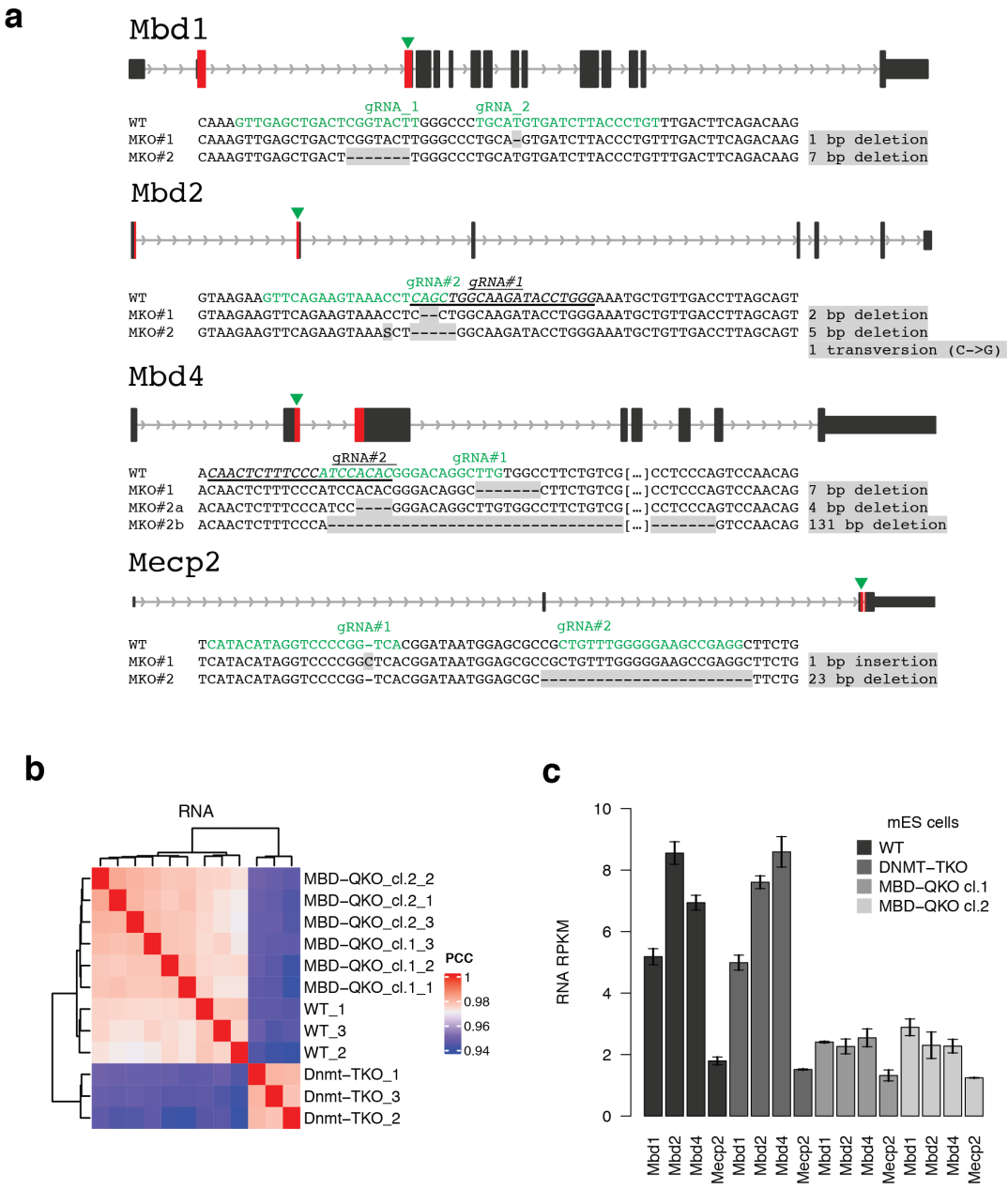
c) Single locus examples of CREB1 binding (grey tracks) in wild-type and DNMT-TKO (TKO) neurons. Accompanied chromatin accessibility changes (ATAC-seq) for WT, DNMT-TKO and CREB1-KO in DNMT-TKO neurons shown in blue tracks. Grey triangle indicated CREB1 motif. Top track indicates methylation frequencies for each CpG (black dots). ChIP-seq or ATAC-seq read counts in running windows of 51nt, replicate data combined.

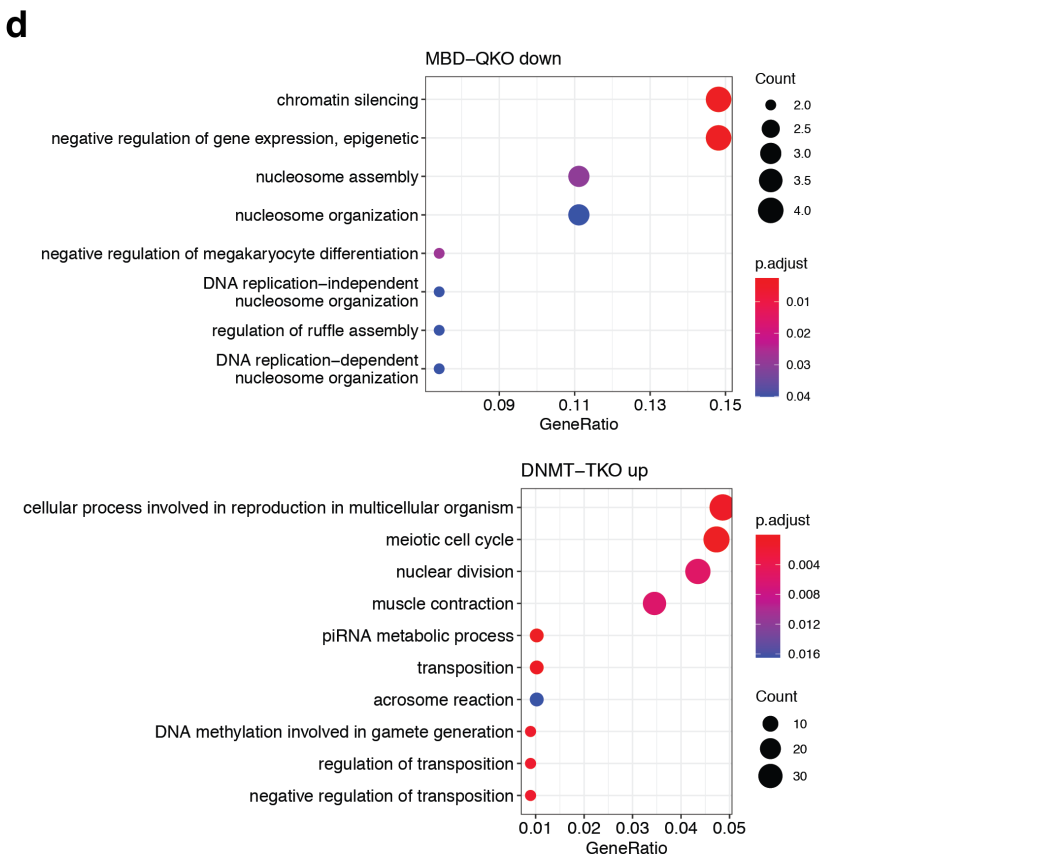
d) Change in CREB1 binding between DNMT-TKO and wild-type neurons grouped according to their average motif methylation (CRE full-length, TGACGTCA or half-site motifs, TGACG/ CGTCA).

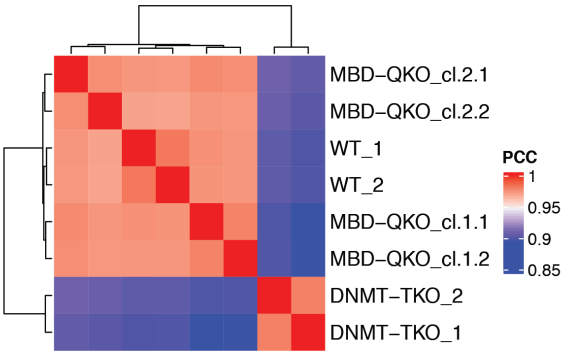
e) Changes in chromatin accessibility (top tracks, ATAC-seq), POL2 binding (middle tracks, ChIP-seq) or CREB1 binding (bottom tracks, ChIP-seq) in WT and DNMT-TKO neurons at IAPLTR1/1a elements that gain expression in absence of DNA methylation (FDR <= 0.05 & foldchange >2). Signal is centered at the start site of 5' LTR. Grey bars depict average width of the 5' LTR and dashed lines display the average length of an entire element including the 5' and 3' LTR regions. Only unique mapped reads are considered. Minimum of two replicates per condition are combined in each composite plot. Number of elements are indicated by n.

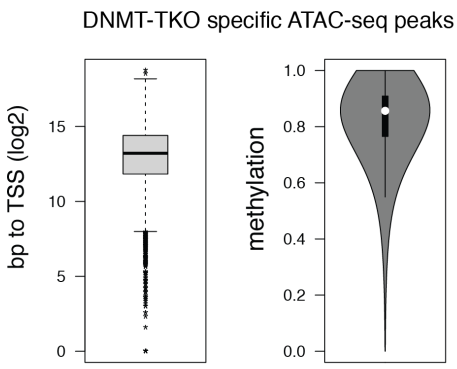
f) ATAC-seq signal at 5'LTRs of IAPLTR1 (n=746) or 1a (n=884) elements (left plot) or RNA expression levels (right plot) in DNMT-TKO neurons with (grey) or without CREB1 (red) at elements de-repressed in absence of DNA methylation (FDR <= 0.05 & foldchange >2). Only uniquely mapped reads to the reference genome were considered. ATAC-seq reads counted in 5' LTRs were summed in a replicate (n=3, black circles). Replicates (n=2) of RNA-seq reads counted in IAPLTR1/1a element were combined by condition. P, P-value derived from a t-test.

Supplementary Figures 1-8


a
Mbd1
gRNA_1
gRNA_2
WT CAAAGTTGAGCTGACTCGGTACTTGGGCCCTGCATGTGATCTTACCCTGTTTGACTTCAGACAAG
MKO#1 CAAAGTTGAGCTGACTCGGTACTTGGGCCCTGCA-GTGATCTTACCCTGTTTGACTTCAGACAAG 1 bp deletion
MKO#2 CAAAGTTGAGCTGACT-------TGGGCCCTGCATGTGATCTTACCCTGTTTGACTTCAGACAAG 7 bp deletion
Mbd2
gRNA#2 gRNA#1
WT GTAAGAAGTTCAGAAGTAAACCTCAGCTGGCAAGATACCTGGGAAATGCTGTTGACCTTAGCAGT
MKO#1 GTAAGAAGTTCAGAAGTAAACCTC--CTGGCAAGATACCTGGGAAATGCTGTTGACCTTAGCAGT 2 bp deletion
MKO#2 GTAAGAAGTTCAGAAGTAAAGCT-----GGCAAGATACCTGGGAAATGCTGTTGACCTTAGCAGT 5 bp deletion
1 transversion (C->G)
Mbd4
gRNA#2 gRNA#1
WT ACAACTCTTTCCCATCCACACGGGACAGGCTTGTGGCCTTCTGTCG[…]CCTCCCAGTCCAACAG
MKO#1 ACAACTCTTTCCCATCCACACGGGACAGGC-------CTTCTGTCG[…]CCTCCCAGTCCAACAG 7 bp deletion
MKO#2a ACAACTCTTTCCCATCC----GGGACAGGCTTGTGGCCTTCTGTCG[…]CCTCCCAGTCCAACAG 4 bp deletion
MKO#2b ACAACTCTTTCCCA----------[…]-------GTCCAACAG 131 bp deletion
Mecp2
gRNA#1 gRNA#2
WT TCATACATAGGTCCCCGG-TCACGGATAATGGAGCGCCGCTGTTTGGGGGAAGCCGAGGCTTCTG
MKO#1 TCATACATAGGTCCCCGGCTCACGGATAATGGAGCGCCGCTGTTTGGGGGAAGCCGAGGCTTCTG 1 bp insertion
MKO#2 TCATACATAGGTCCCCGG-TCACGGATAATGGAGCGC----------TTCTG 23 bp deletion
b
RNA
MBD-QKO_cl.2_2
MBD-QKO_cl.2_1
MBD-QKO_cl.2_3
MBD-QKO_cl.1_3
MBD-QKO_cl.1_2
MBD-QKO_cl.1_1
WT_1
WT_3
WT_2
Dnmt-TKO_1
Dnmt-TKO_3
Dnmt-TKO_2
PCC
1
0.98
0.96
0.94
c
RNA RPKM
mES cells
WT
DNMT-TKO
MBD-QKO cl.1
MBD-QKO cl.2
Mbd1 Mbd2 Mbd4 Mecp2 Mbd1 Mbd2 Mbd4 Mecp2 Mbd1 Mbd2 Mbd4 Mecp2 Mbd1 Mbd2 Mbd4 Mecp2

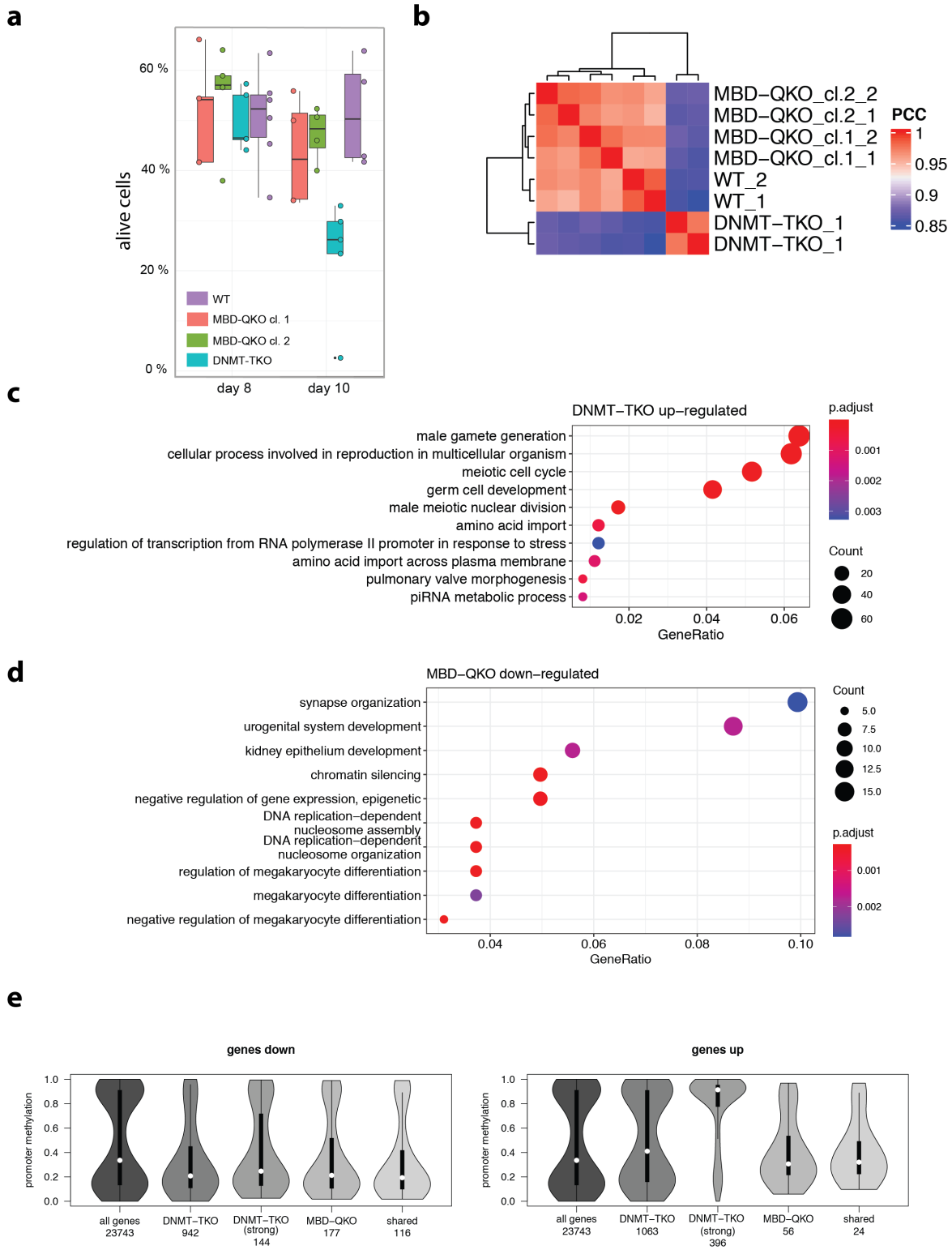

d
MBD−QKO down
chromatin silencing
negative regulation of gene expression, epigenetic
nucleosome assembly
nucleosome organization
negative regulation of megakaryocyte differentiation
DNA replication−independent nucleosome organization
regulation of ruffle assembly
DNA replication−dependent nucleosome organization
0.09
0.11
0.13
0.15
GeneRatio
Count
2.0
2.5
3.0
3.5
4.0
p.adjust
0.01
0.02
0.03
0.04
DNMT−TKO up
cellular process involved in reproduction in multicellular organism
meiotic cell cycle
nuclear division
muscle contraction
piRNA metabolic process
transposition
acrosome reaction
DNA methylation involved in gamete generation
regulation of transposition
negative regulation of transposition
0.01
0.02
0.03
0.04
0.05
GeneRatio
p.adjust
0.004
0.008
0.012
0.016
Count
10
20
30


e
MBD−QKO_cl.2.1
MBD−QKO_cl.2.2
WT_1
WT_2
MBD−QKO_cl.1.1
MBD−QKO_cl.1.2
DNMT−TKO_2
DNMT−TKO_1
PCC
1
0.95
0.9
0.85


f
DNMT-TKO specific ATAC-seq peaks
bp to TSS (log2)
0
5
10
15
methylation
0.0
0.2
0.4
0.6
0.8
1.0

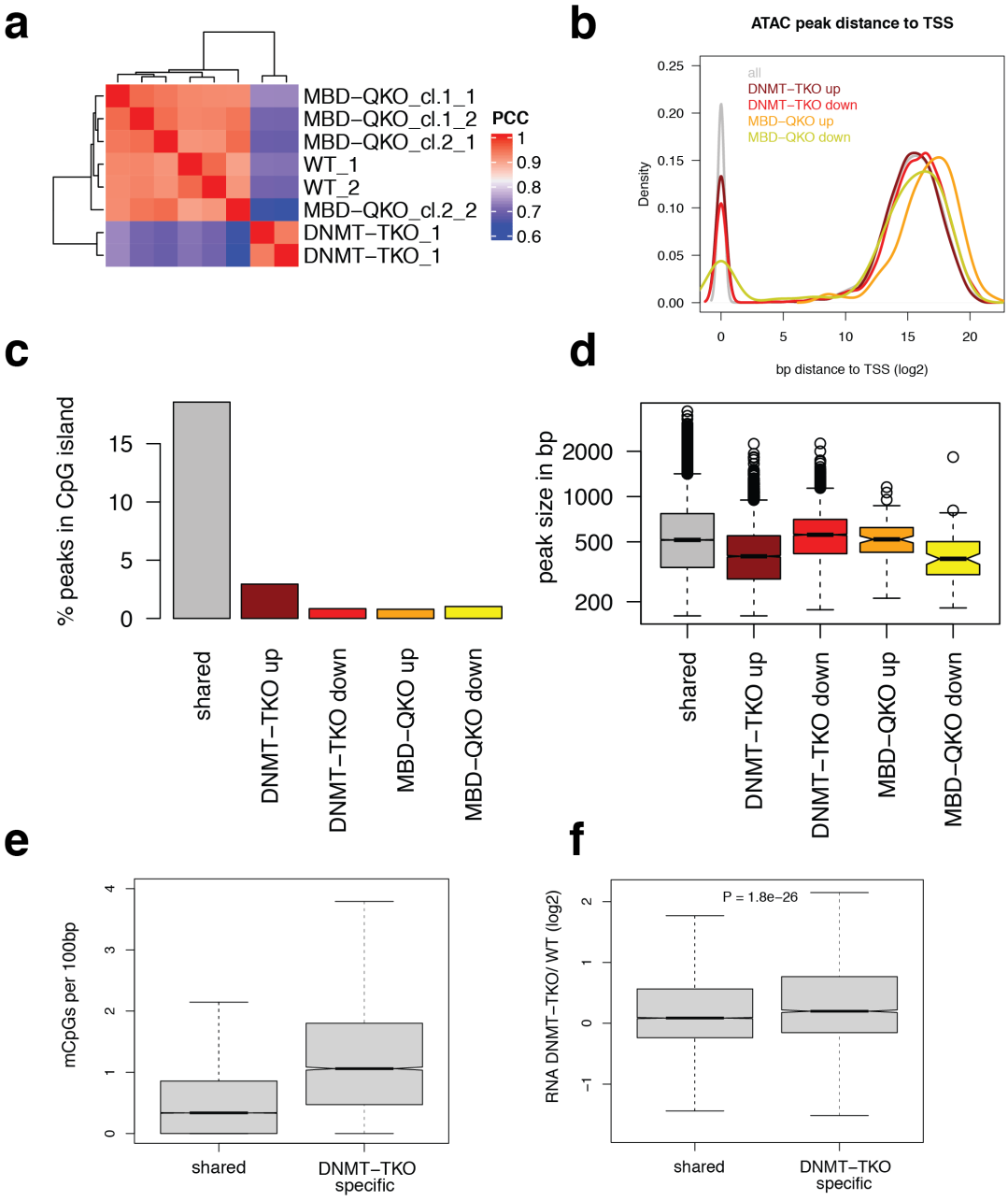
Supplementary Figure 1. ES cells are viable without 5mC-binding MBD proteins with limited changes in transcription and chromatin accessibility

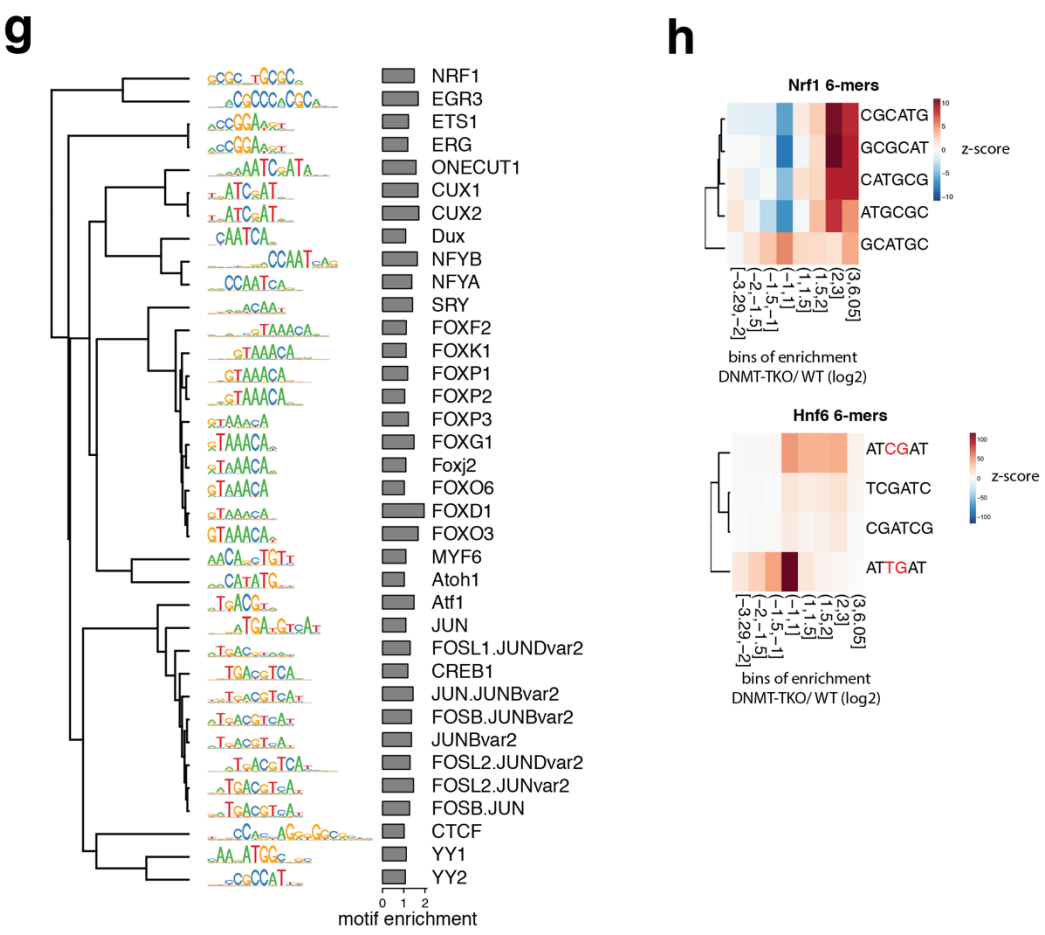
a) Sequencing reads of MBD-QKO clones (MKO) nr.1 and nr2. Shema representing genes with exons (black), gRNA target sites (green triangle) and methyl-CpG binding domain coding region (red). gRNA target sequence is highlighted in green or underlined.
b) Unsupervised clustering of RNA-seq samples from WT and mutant mES cells counted in genes (RPKM). PCC, Pearson's correlation coefficient.
c) Expression levels (RPKM) of MBD genes in different cell lines (mean ± standard deviation, n=3 replicates), indicating reduces expression levels in MBD-QKO mES cells.
d) Gene Ontology (GO) terms enriched in the set of significantly changing genes from figure 1c for MBD-QKO (downregulated, excluding *Mbd* genes) or DNMT-TKO (upregulated) compared to WT mES cells. The dots represent the top eight (MBD-QKO) or top ten (DNMT-TKO) terms with highest gene ratio (fraction of genes represented in the given GO term) with dot size and color representing gene counts and the adjusted p-value, respectively.
e) Unsupervised clustering of ATAC-seq samples from WT and mutant mES cells. Colors indicate pairwise Pearson's correlation coefficients (PCC) of log-transformed normalized read counts in ATAC-seq peaks, indicating clear separation of DNMT-TKO from WT and MBD-QKO mES cells.
f) DNMT-TKO specific ATAC-seq peak distance to promoters (boxplot, left) or DNA methylation levels (violin plot, right)


a
60 %
40 %
20 %
0 %
alive cells
day 8
day 10
WT
MBD-QKO cl. 1
MBD-QKO cl. 2
DNMT-TKO
b
MBD−QKO_cl.2_2
MBD−QKO_cl.2_1
MBD−QKO_cl.1_2
MBD−QKO_cl.1_1
WT_2
WT_1
DNMT−TKO_1
DNMT−TKO_1
PCC
1
0.95
0.9
0.85
c
DNMT−TKO up−regulated
male gamete generation
cellular process involved in reproduction in multicellular organism
meiotic cell cycle
germ cell development
male meiotic nuclear division
amino acid import
regulation of transcription from RNA polymerase II promoter in response to stress
amino acid import across plasma membrane
pulmonary valve morphogenesis
piRNA metabolic process
0.02
0.04
0.06
GeneRatio
p.adjust
0.001
0.002
0.003
Count
20
40
60
d
MBD−QKO down−regulated
synapse organization
urogenital system development
kidney epithelium development
chromatin silencing
negative regulation of gene expression, epigenetic
DNA replication−dependent nucleosome assembly
DNA replication−dependent nucleosome organization
regulation of megakaryocyte differentiation
megakaryocyte differentiation
negative regulation of megakaryocyte differentiation
0.04
0.06
0.08
0.10
GeneRatio
Count
5.0
7.5
10.0
12.5
15.0
p.adjust
0.001
0.002
e
genes down
promoter methylation
1.0
0.8
0.6
0.4
0.2
0.0
all genes 23743
DNMT−TKO 942
DNMT−TKO (strong) 144
MBD−QKO 177
shared 116
genes up
promoter methylation
all genes 23743
DNMT−TKO 1063
DNMT−TKO (strong) 396
MBD−QKO 56
shared 24

Supplementary Figure 2. MBD or DNMT deleted neurons show distinct phenotypes

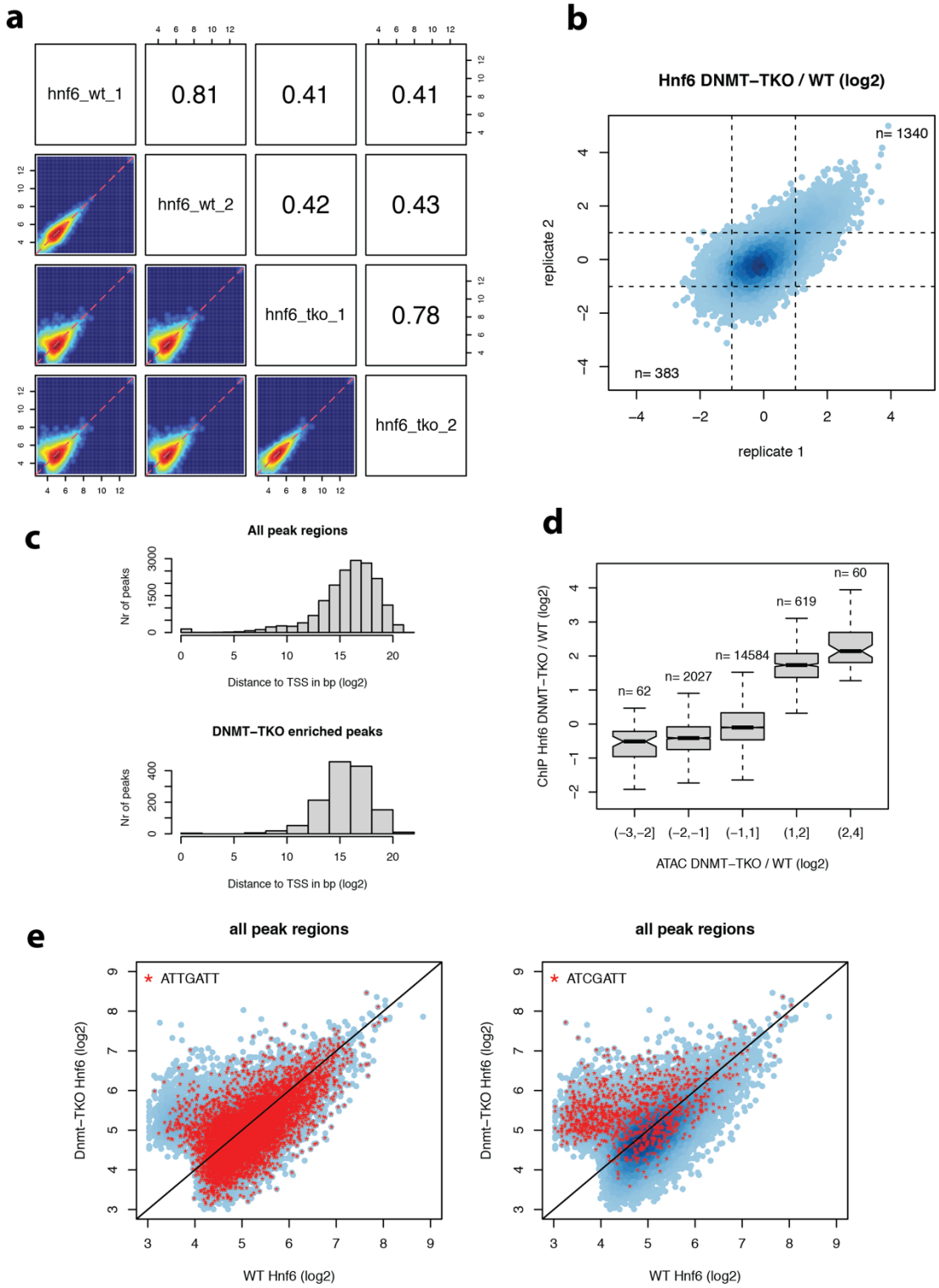
a) Cell viability eight days after induction of neuronal differentiation (see methods).
b) Unsupervised clustering of RNA-seq samples from WT and mutant neuron cells counted in genes. Colors indicate pairwise Pearson's correlation coefficients (PCC) of log-transformed RPKM, indicating clear separation of DNMT-TKO from WT and MBD-QKO neurons.
c) Gene Ontology (GO) terms enriched in the set of upregulated (FDR < 1e2 & foldchange >=2) DNMT-TKO compared to WT neurons. The dots represent the top ten terms with highest gene ratio (fraction of genes represented in the given GO term) with dot size and color representing gene counts and the adjusted p-value, respectively.
d) Same as in c) but for the downregulated set of genes in MBD-QKO neurons compared to WT.
e) Violin plots of promoter methylation levels in neurons from all genes or genes that decrease in expression (FDR < 1e2 & foldchange > 2) in absence of DNA methylation or of MBD proteins or both (shared). Genes with a strong decrease in expression levels (FDR < 1e2 & foldchange > 8) in absence of DNA methylation are separately labeled. Number below labels indicates gene number.
f) Same as in e) for all genes or genes that increase in expression in absence of DNA methylation of MBD proteins or both (shared). Genes with a strong increase in expression levels (FDR < 1e2 & foldchange > 8) in absence of DNA methylation are separately labeled.


a
MBD−QKO_cl.1_1
MBD−QKO_cl.1_2
MBD−QKO_cl.2_1
WT_1
WT_2
MBD−QKO_cl.2_2
DNMT−TKO_1
DNMT−TKO_1
PCC
1
0.9
0.8
0.7
0.6
b
ATAC peak distance to TSS
all
DNMT−TKO up
DNMT−TKO down
MBD−QKO up
MBD−QKO down
Density
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
bp distance to TSS (log2)
c
% peaks in CpG island
15
10
5
0
shared
DNMT−TKO up
DNMT−TKO down
MBD−QKO up
MBD−QKO down
d
peak size in bp
2000
1000
500
200
shared
DNMT−TKO up
DNMT−TKO down
MBD−QKO up
MBD−QKO down
e
mCpGs per 100bp
4
3
2
1
0
shared
DNMT−TKO
specific
f
RNA DNMT−TKO/ WT (log2)
P = 1.8e−26
2
1
0
−1
shared
DNMT−TKO
specific

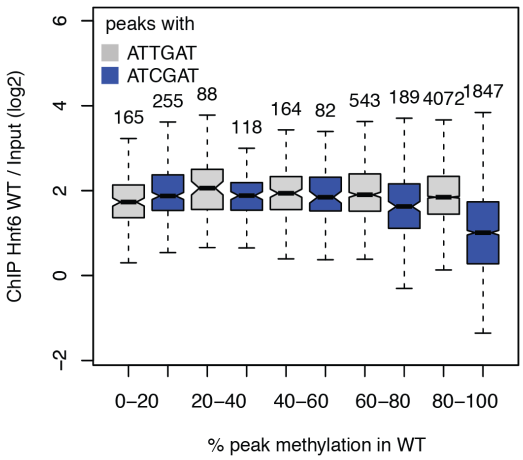

g
NRF1
EGR3
ETS1
ERG
ONECUT1
CUX1
CUX2
Dux
NFYB
NFYA
SRY
FOXF2
FOXK1
FOXP1
FOXP2
FOXP3
FOXG1
Foxj2
FOXO6
FOXD1
FOXO3
MYF6
Atoh1
Atf1
JUN
FOSL1.JUNDvar2
CREB1
JUN.JUNBvar2
FOSB.JUNBvar2
JUNBvar2
FOSL2.JUNDvar2
FOSL2.JUNvar2
FOSB.JUN
CTCF
YY1
YY2
0 1 2
motif enrichment
h
Nrf1 6-mers
CGCATG
GCGCAT
CATGCG
ATGCGC
GCATGC
z-score
[−3.29,−2]
(−2,−1.5]
(−1.5,−1]
(−1,1]
(1,1.5]
(1.5,2]
(2,3]
(3.6,05]
bins of enrichment
DNMT-TKO/ WT (log2)
Hnf6 6-mers
ATCGAT
TCGATC
CGATCG
ATTGAT
z-score
[−3.29,−2]
(−2,−1.5]
(−1.5,−1]
(−1,1]
(1,1.5]
(1.5,2]
(2,3]
(3.6,05]
bins of enrichment
DNMT-TKO/ WT (log2)

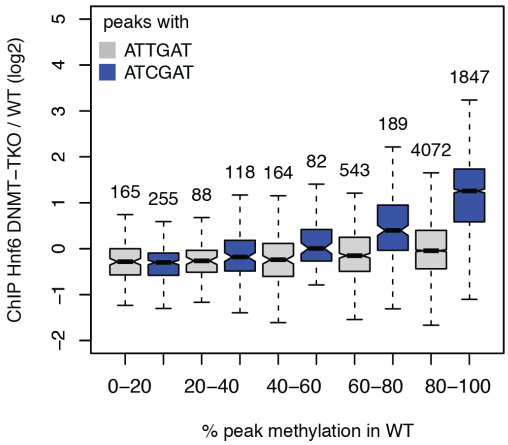
Supplementary Figure 3. The chromatin accessibility landscape in neurons changes in response to loss of DNA methylation but not MBD proteins.

a) Unsupervised clustering of ATAC-seq samples from WT and mutant neuron cells. Colors indicate pairwise Pearson's correlation coefficients (PCC) of log-transformed normalized read counts in ATAC-seq peaks, indicating clear separation of DNMT-TKO from WT and MBD-QKO ES cells.
b) Distance of ATAC-seq peaks identified in any condition (all, n=84,400), DNMT-TKO specific (up, n= 7121 or down, n = 5606) or MBD-QKO specific (up, n=126 or down, n=97) to the nearest transcriptional start site (TSS).
c) Barplot of ATAC-seq peaks overlapping with CpG islands. Shared peaks between MBD-QKO and DNMT-TKO (n= 13264), DNMT-TKO specific (down, n=47 or up, n= 210) and MBD-QKO specific (down, n= 1 or up, n= 1).
d) Boxplot of ATAC-seq peak sizes for shared peaks between WT, MBD-QKO and DNMT-TKO (n= 71432), DNMT-TKO specific (down, n=5606 or up, n= 7121) and MBD-QKO specific (down, n= 97 or up, n= 126).
e) Boxplot of number of methylated CpG dinucleotides (methylation frequency > 0.8) in shared peaks between WT and DNMT-TKO neurons or DNMT-TKO specific peaks (P value < 1e-16 from a t-test).
f) Expression change of genes closest to shared and DNMT-TKO specific peaks (P value from a t-test).
g) Unbiased clustering of motif similarities (position weight matrices) of motifs enriched (FDR < 9e-13 and foldchange > 1, n=36) in DNMT-TKO specific ATAC-seq peaks from Figure 1c.
h) Enrichment of NRF1 or HNF6 hexamers in bins of differentially accessible ATAC-seq peaks between DNMT-TKO and WT neurons. Colors indicate pearson residuals with positive values indicating hexamer enrichments. Similar to NRF, CpG containing HNF6 hexamers are enriched in bins of peaks that gain accessibility in DNA methylation deficient cells, opposed to the canonical (CpG-free) HNF6 6-mers.


a
hnf6_wt_1
0.81
0.41
0.41
hnf6_wt_2
0.42
0.43
hnf6_tko_1
0.78
hnf6_tko_2
b
Hnf6 DNMT−TKO / WT (log2)
n= 1340
n= 383
replicate 2
replicate 1
c
All peak regions
Nr of peaks
Distance to TSS in bp (log2)
DNMT−TKO enriched peaks
Nr of peaks
Distance to TSS in bp (log2)
d
ChIP Hnf6 DNMT−TKO / WT (log2)
n= 62
n= 2027
n= 14584
n= 619
n= 60
(−3,−2]
(−2,−1]
(−1,1]
(1,2]
(2,4]
ATAC DNMT−TKO / WT (log2)
e
all peak regions
ATTGATT
Dnmt−TKO Hnf6 (log2)
WT Hnf6 (log2)
all peak regions
ATCGATT
Dnmt−TKO Hnf6 (log2)
WT Hnf6 (log2)

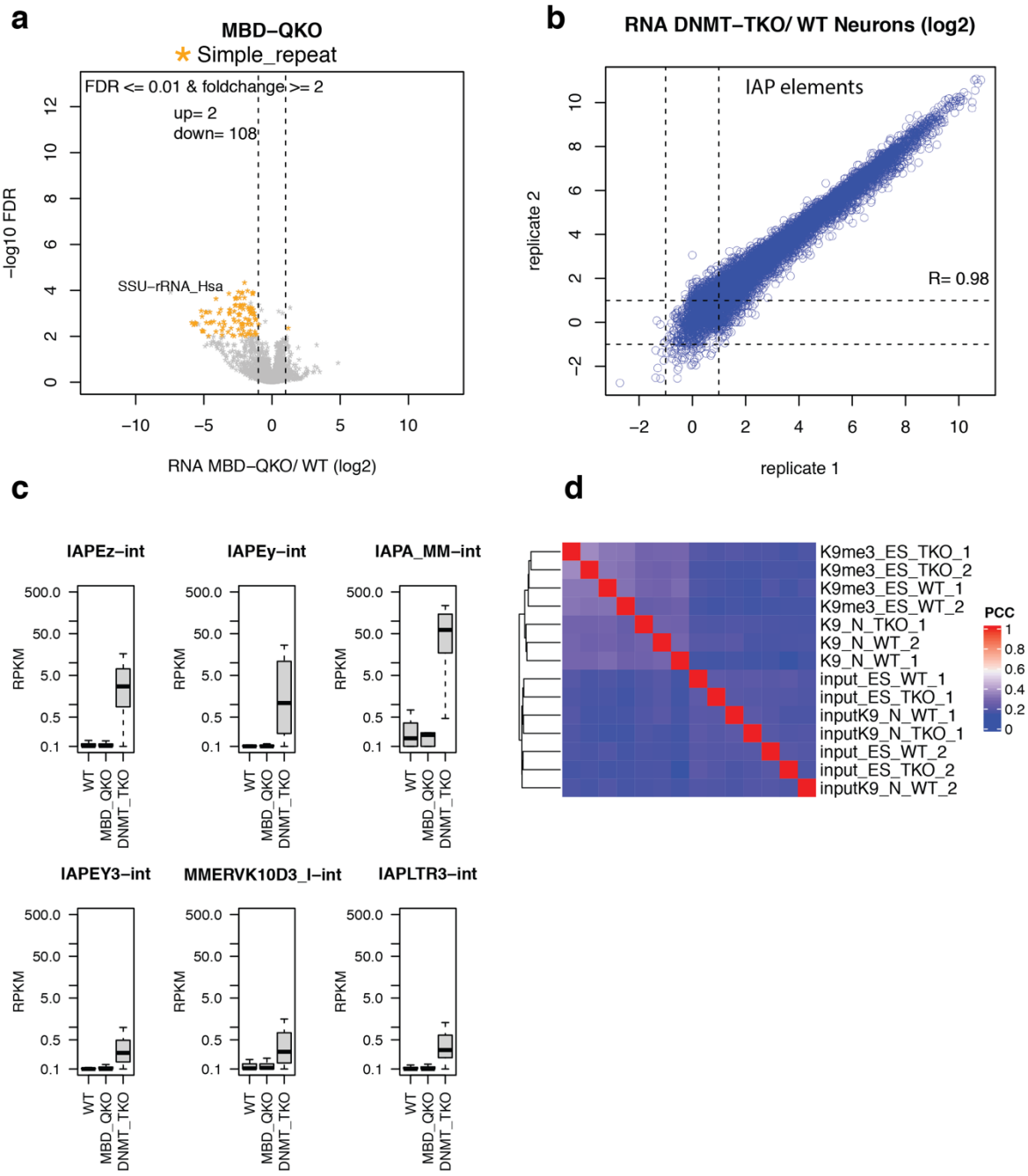
f


peaks with
ATTGAT
ATCGAT
ChIP Hnf6 WT / Input (log2)
6
4
2
0
−2
165
255
88
118
164
82
543
189
4072
1847
0–20
20–40
40–60
60–80
80–100
% peak methylation in WT


peaks with
ATTGAT
ATCGAT
ChIP Hnf6 DNMT−TKO / WT (log2)
5
4
3
2
1
0
−1
−2
165
255
88
118
164
82
543
189
4072
1847
0–20
20–40
40–60
60–80
80–100
% peak methylation in WT

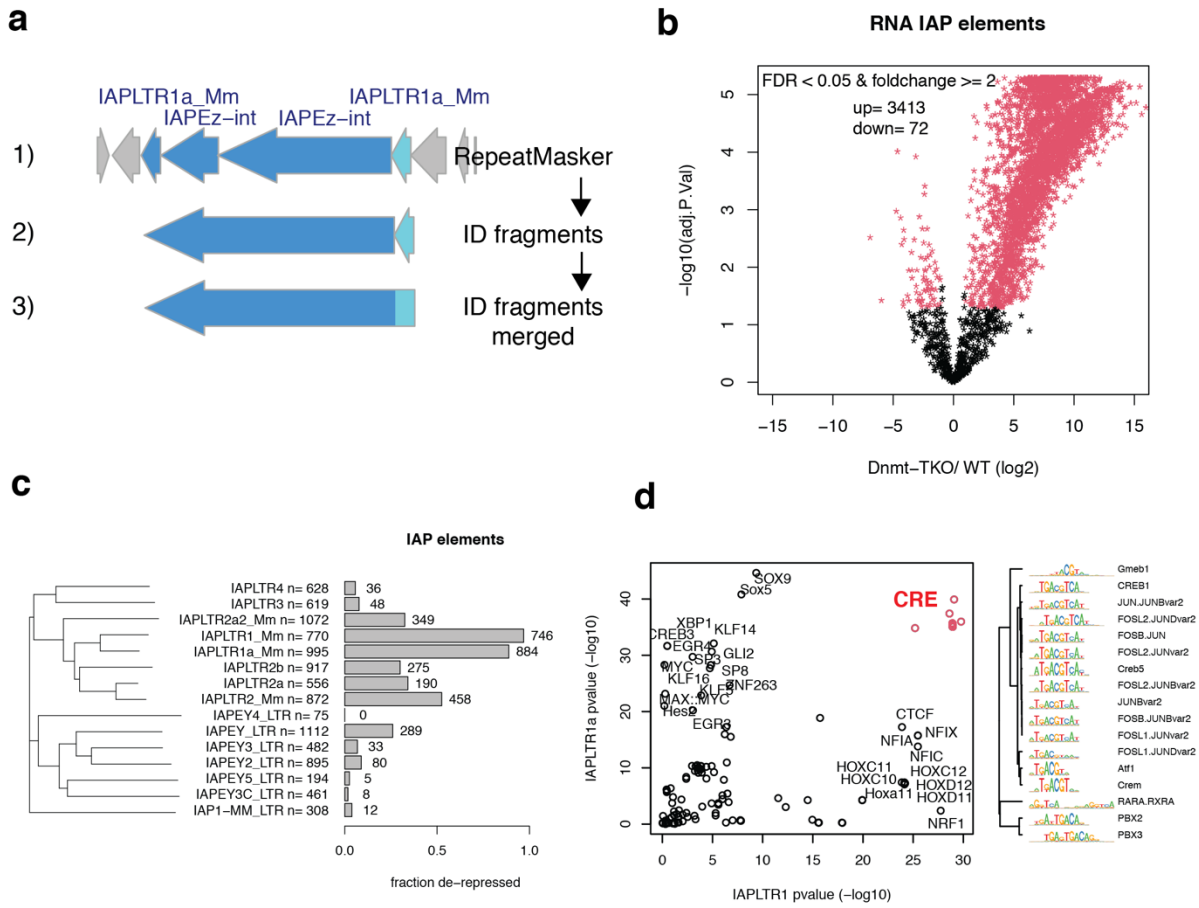
Supplementary Figure 4. HNF6 is methylation-sensitive at its CpG-containing motif variant

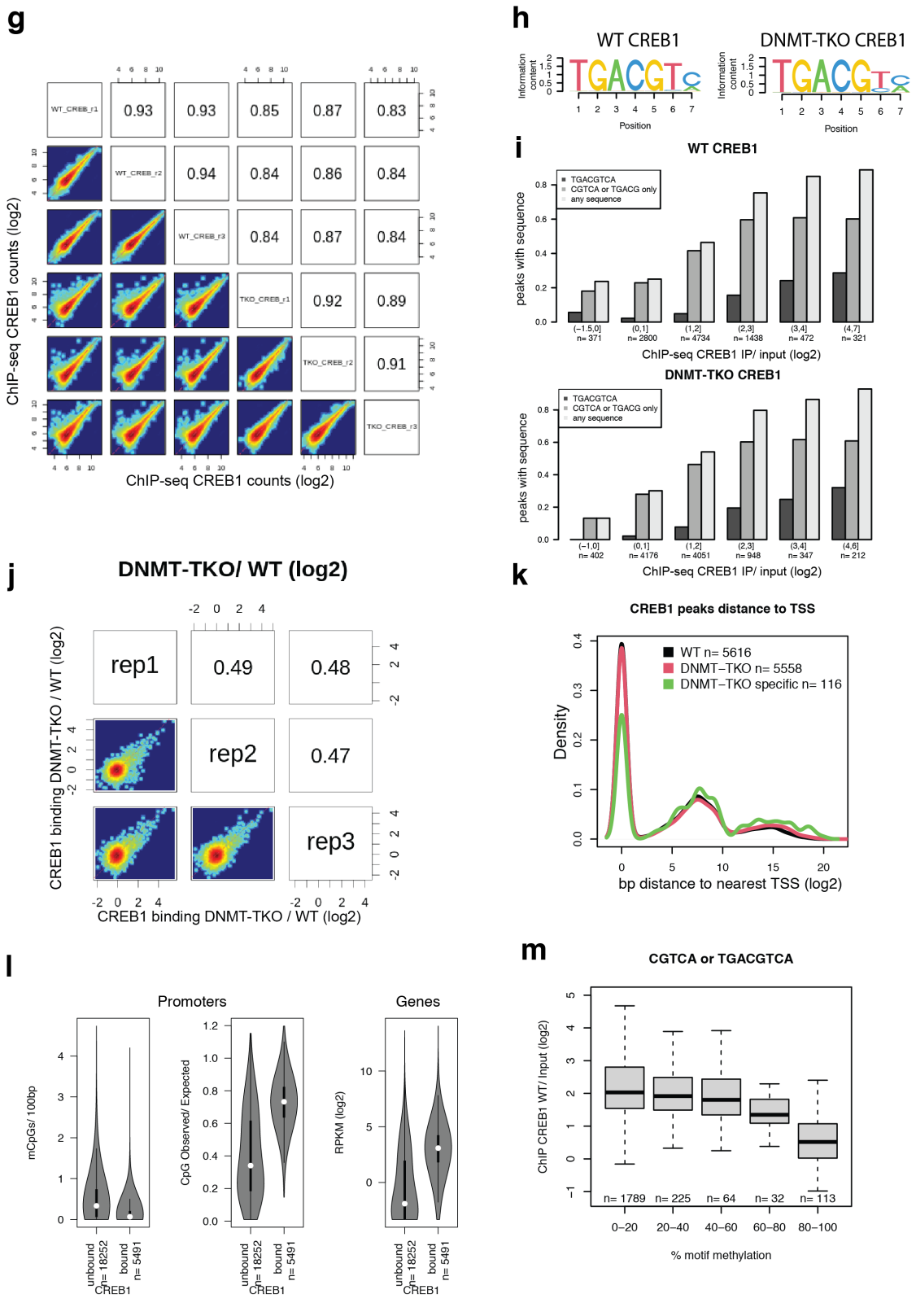
a) Reproducibility of read counts for two independent HNF6 ChIP-seq replicates from wild-type and DNMT-TKO neurons. Reproducible WT and DNMT-TKO peak regions were merged. Pearson correlation coefficients are indicated.
b) Reproducibility of changes in HNF6 binding in DNMT-TKO versus wild-type neurons in all peaks regions. Pearson correlation coefficient is indicated
c) Distance of HNF6 peak regions to the nearest transcriptional start site (TSS). Top, peak regions identified in any condition (n=17,352). Bottom, DNMT-TKO specific peak regions (n=1340).
d) Changes of HNF6 binding versus changes in chromatin accessibility between WT and DNMT-TKO neurons in HNF6 peak regions identified in any condition (n=17,352).
e) WT and DNMT-TKO HNF6 ChIP-seq signal at peak regions identified in any condition. Red asterisks mark peak regions that at least once contain the canonical HNF6 motif (left) or the CpG-containing variant (right).
f) HNF6 enrichment over input in WT neurons (left) or changes in HNF6 binding between DNMT-TKO and WT neurons (right) versus bins of percentage methylation of peak regions. Grey box plots are peak regions containing the canonical motif at least once. Blue box plots are peak regions containing the CpG-containing motif variant at least once. Number above box plots indicates number of peaks.


a
MBD−QKO
Simple_repeat
FDR <= 0.01 & foldchange >= 2
up= 2
down= 108
SSU−rRNA_Hsa
−log10 FDR
RNA MBD−QKO/ WT (log2)
b
RNA DNMT−TKO/ WT Neurons (log2)
IAP elements
R= 0.98
replicate 2
replicate 1
c
IAPEz−int
IAPEy−int
IAPA_MM−int
IAPEY3−int
MMERVK10D3_I−int
IAPLTR3−int
RPKM
WT
MBD_QKO
DNMT_TKO
d
K9me3_ES_TKO_1
K9me3_ES_TKO_2
K9me3_ES_WT_1
K9me3_ES_WT_2
K9_N_TKO_1
K9_N_WT_2
K9_N_WT_1
input_ES_WT_1
input_ES_TKO_1
inputK9_N_WT_1
inputK9_N_TKO_1
input_ES_WT_2
input_ES_TKO_2
inputK9_N_WT_2
PCC

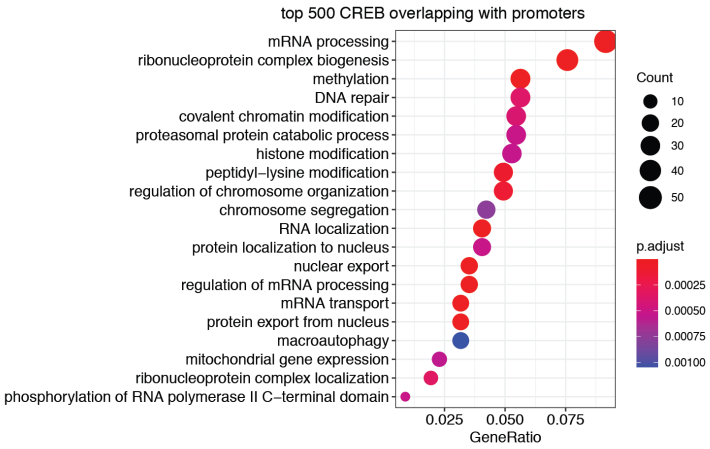
Supplementary Figure 5. Repeats are de-repressed absence of DNA methylation but not MBD proteins in neurons

a) Volcano plot showing differential expressed repeat subfamilies in MBD-QKO neurons using multimapping reads. Dashed lines indicate twofold expression change. The most abundant repeat subfamily differentially expressed (FDR =< 0.01 and foldchange >=2) is colored.
b) Reproducibility of IAP expression change in DNMT-TKO versus wild-type neurons in all IAP elements annotated by RepeatMasker. Only uniquely mapped RNA-seq reads are considered.
c) Boxplots of RNA expression of top six most significant differentially expressed internal sequence of class-2 endogenous retroviruses from figure 5b). Only RNA-seq reads were considered that uniquely mapped to the reference genome. Internal IAP elements selected to have more than 8 counts in at least one condition.
d) Unsupervised clustering of H3K9me3 ChIP-seq samples from WT and DNMT-TKO mES cells and derived neurons. Colors indicate pairwise Pearson's correlation coefficients (PCC) of log-transformed RPKM of repeats annotated by RepeatMasker. Multimapping RNA-seq reads were considered for counting.


a
IAPLTR1a_Mm
IAPEz–int
IAPLTR1a_Mm
IAPEz–int
1)
RepeatMasker
2)
ID fragments
3)
ID fragments
merged
b
RNA IAP elements
FDR < 0.05 & foldchange >= 2
up= 3413
down= 72
−log10(adj.P.Val)
Dnmt–TKO/ WT (log2)
c
IAP elements
IAPLTR4 n= 628
IAPLTR3 n= 619
IAPLTR2a2_Mm n= 1072
IAPLTR1_Mm n= 770
IAPLTR1a_Mm n= 995
IAPLTR2b n= 917
IAPLTR2a n= 556
IAPLTR2_Mm n= 872
IAPEY4_LTR n= 75
IAPEY_LTR n= 1112
IAPEY3_LTR n= 482
IAPEY2_LTR n= 895
IAPEY5_LTR n= 194
IAPEY3C_LTR n= 461
IAP1–MM_LTR n= 308
fraction de–repressed
d
IAPLTR1a pvalue (–log10)
IAPLTR1 pvalue (–log10)
CRE
SOX9
Sox5
XBP1
KLF14
GLI2
SP8
ZNF263
CTCF
NFIX
NFIA
NFIC
HOXC12
HOXD12
HOXD11
NRF1
HOXC11
HOXC10
Hoxa11
EGR3
Gmeb1
CREB1
JUN.JUNBvar2
FOSL2.JUNDvar2
FOSB.JUN
FOSL2.JUNvar2
Creb5
FOSL2.JUNBvar2
JUNBvar2
FOSB.JUNBvar2
FOSL1.JUNvar2
FOSL1.JUNDvar2
Atf1
Crem
RARA.RXRA
PBX2
PBX3
e
ES cells
relative activity
PGK
IAPLTR1a–WT
IAPLTR1a–ΔCRE
WT
DNMT-TKO
f
ATAC–seq unique mappers
ATAC–seq multimappers
DNMT–TKO (log2 counts)
WT (log2)


g
WT_CREB_r1
WT_CREB_r2
WT_CREB_r3
TKO_CREB_r1
TKO_CREB_r2
TKO_CREB_r3
0.93
0.93
0.85
0.87
0.83
0.94
0.84
0.86
0.84
0.84
0.87
0.84
0.92
0.89
0.91
ChIP-seq CREB1 counts (log2)
ChIP-seq CREB1 counts (log2)
h
WT CREB1
DNMT-TKO CREB1
Information content
Position
i
WT CREB1
TGACGTCA
CGTCA or TGACG only
any sequence
peaks with sequence
(−1.5,0] n= 371
(0,1] n= 2800
(1,2] n= 4734
(2,3] n= 1438
(3,4] n= 472
(4,7] n= 321
ChIP-seq CREB1 IP/ input (log2)
DNMT-TKO CREB1
(−1,0] n= 402
(0,1] n= 4176
(1,2] n= 4051
(2,3] n= 948
(3,4] n= 347
(4,6] n= 212
j
DNMT-TKO/ WT (log2)
rep1
rep2
rep3
0.49
0.48
0.47
CREB1 binding DNMT-TKO / WT (log2)
k
CREB1 peaks distance to TSS
WT n= 5616
DNMT−TKO n= 5558
DNMT−TKO specific n= 116
Density
bp distance to nearest TSS (log2)
l
Promoters
Genes
mCpGs/ 100bp
CpG Observed/ Expected
RPKM (log2)
unbound n= 18252
bound n= 5491
CREB1
m
CGTCA or TGACGTCA
ChIP CREB1 WT/ Input (log2)
n= 1789
n= 225
n= 64
n= 32
n= 113
0−20
20−40
40−60
60−80
80−100
% motif methylation

n


top 500 CREB overlapping with promoters
mRNA processing
ribonucleoprotein complex biogenesis
methylation
DNA repair
covalent chromatin modification
proteasomal protein catabolic process
histone modification
peptidyl−lysine modification
regulation of chromosome organization
chromosome segregation
RNA localization
protein localization to nucleus
nuclear export
regulation of mRNA processing
mRNA transport
protein export from nucleus
macroautophagy
mitochondrial gene expression
ribonucleoprotein complex localization
phosphorylation of RNA polymerase II C−terminal domain
0.025
0.050
0.075
GeneRatio
Count
10
20
30
40
50
p.adjust
0.00025
0.00050
0.00075
0.00100

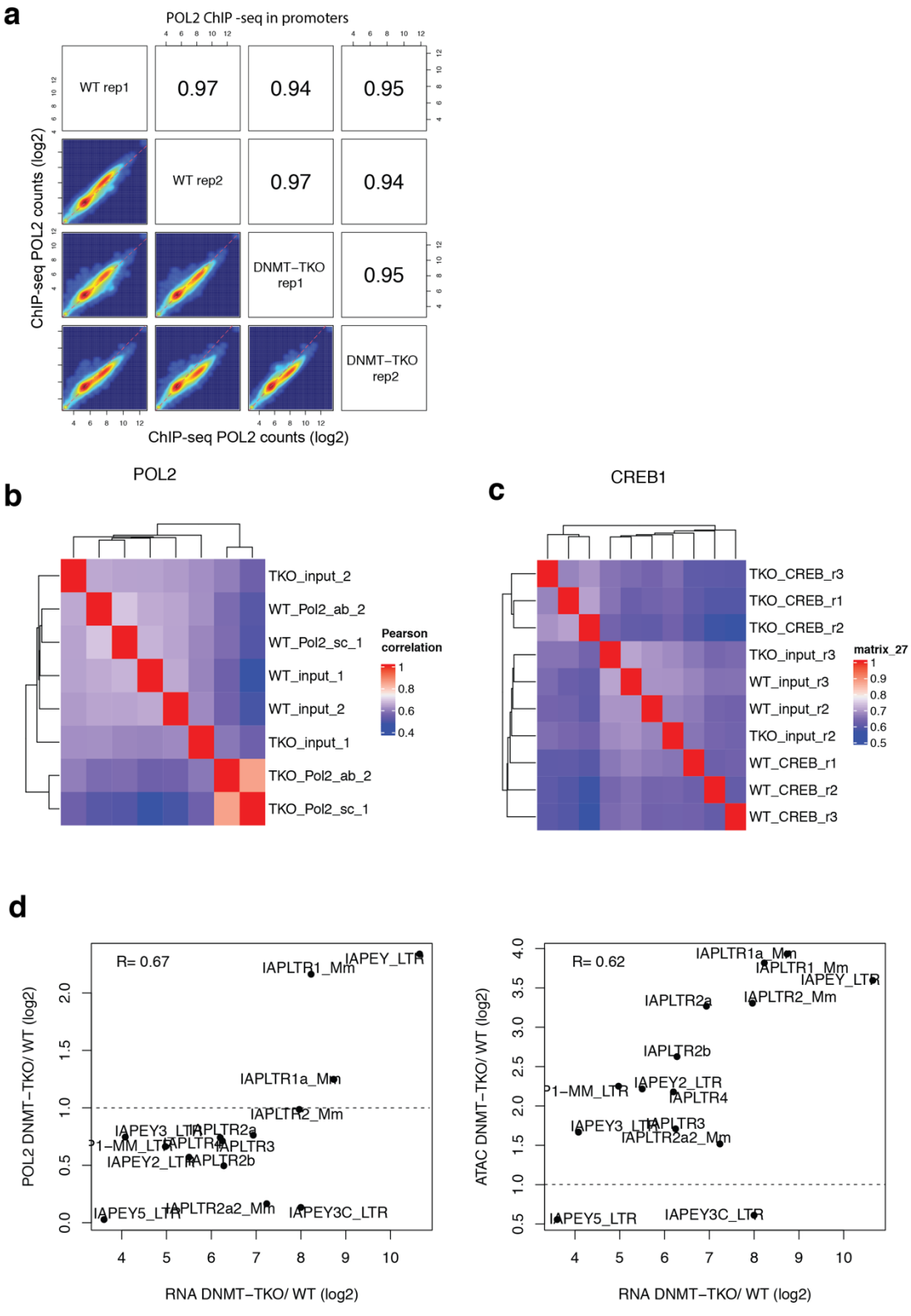
Supplementary Figure 6. CRE is important for IAP activity and bound by methylation-sensitive CREB1

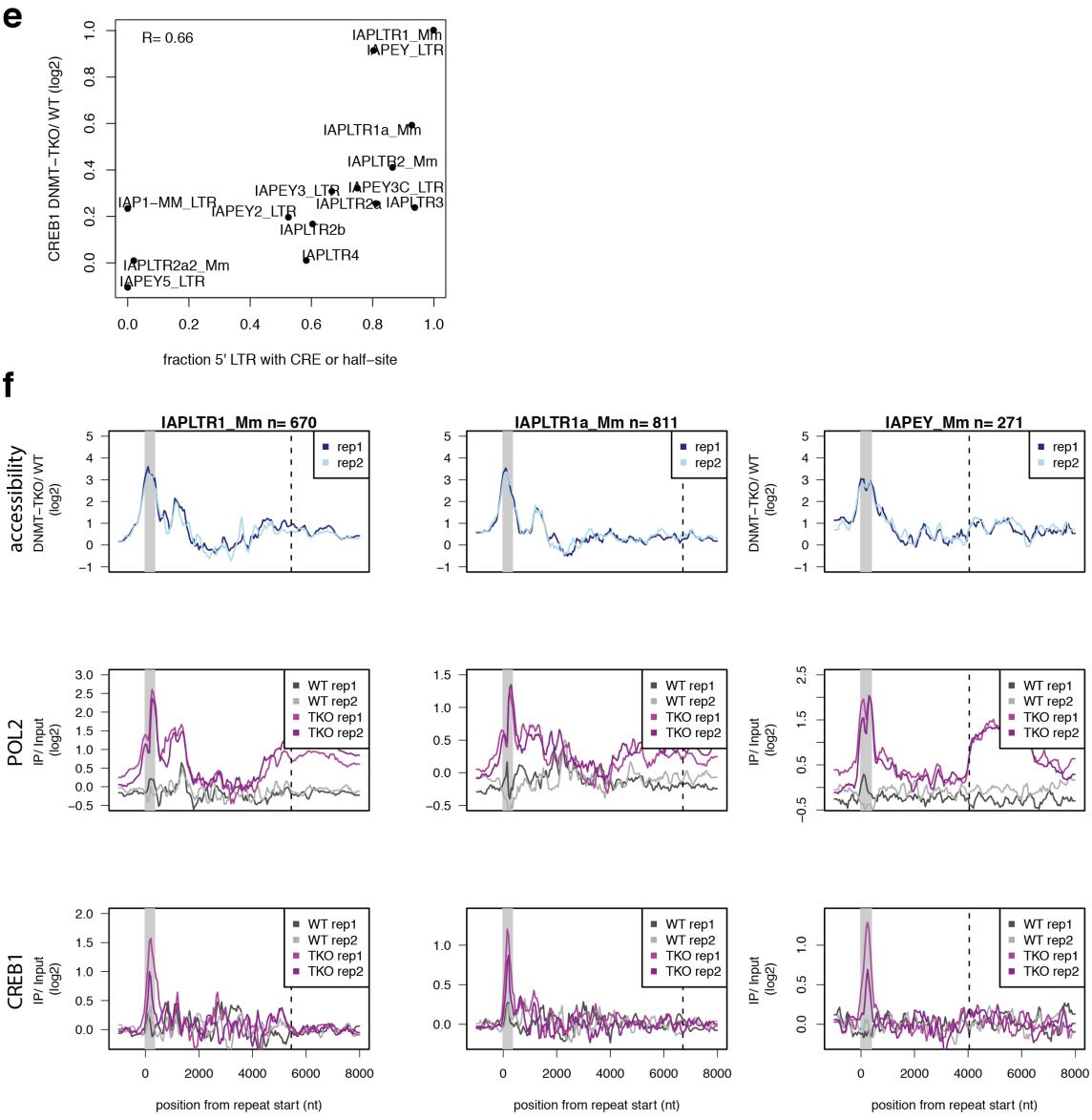
a) Curation of RepeatMasker annotation for IAP elements. 1) RepeatMasker annotation of repeats (arrows) with repName (text above arrows) and ID (in blue and light blue, same color indicates same ID). ID provided by RepeatMasker and indicates related fragments. Grey coloring indicates non-IAP repeat fragments. 2) IAP fragments of same ID merged into one fragment (ID fragments). 3) ID fragments of containing same subfamily and in close proximity (< 1024 bp) are merged. This curation results in an full length IAP element, including the 5' and 3' LTRs.
b) Volcano plots showing differentially expressed IAP elements between DNMT-TKO and WT neurons based on the curated IAP annotation.
c) Phylogenetic tree of sequence similarities between 5' LTR reference sequences from the dfam database (Storer et al. 2021) and fraction of associated elements de-repressed in neurons without DNA methylation. This indicates that related IAP elements show a similar fraction of de-repressed elements.
d) Plot shows Jaspar motifs (black circles) that occur at least 10 times in 5' LTRs of either IAPLTR1 or 1a elements. Axis indicate P values from a Wilcoxon test for the expression difference between groups of IAPLTR1/1a elements separated by motif presence. A cluster of motifs (red circle) separates high and low expressing repeats with high significance in both IAPLTR1 and 1a elements. Their motif logos are depicted next to the plot. Unbiased clustering of motif logos indicates that most motifs resemble the CRE motif (TGACGTCA).
e) Bar plot showing reporter activity in mES cells from figure 6b, indicating that IAPLTR1a reporters are silent in WT and only moderately active in DNMT-TKO mES cells. Importantly, absence of CRE reduces the reporter activity by half. WT_PGK (n=7), WT_LTR (n=7), WT_LTR_∆CRE (n=7), DNMT-TKO_PGK (n=7), DNMT-TKO_LTR (n=7), DNMT-TKO_LTR_∆CRE (n=7)
f) ATAC-seq signal around CRE motifs (TGACGTCA, +- 100 bp) in WT and DNMT-TKO neurons for unique (left) or multimapping reads (right)
g) Reproducibility of read counts for three independent CREB1 ChIP-seq replicates from wild-type and DNMT-TKO neurons in all peak regions. Reproducible WT and DNMT-TKO peak regions were merged. Pearson correlation coefficients are indicated.
h) Top motif found by *de novo* motif search in the top 500 peaks of CREB1 ChIP in WT or DNMT-TKO neurons using HOMER
i) Bar plot showing fraction of peaks with a CRE sequence for different bins of CREB1 enrichment in WT (top) or DNMT-TKO (bottom) neurons. Replicates with matching inputs (n=2) combined for each condition. Number of peaks per enrichment bin indicated by n.
j) Reproducibility of changes in CREB1 binding in DNMT-TKO versus wild-type cells at peak regions identified across cell lines. Pearson correlation coefficient is indicated. r1-r3 indicate biological replicates.
k) CREB1 peak distance to the nearest TSS
l) Comparison of CREB1 bound and unbound promoters (-1500 bp and +500 bp around TSS) for methylated CpGs (left), CpG density

(observed/expected, see methods) and associated gene activity (right) in WT neurons

m) CREB1 binding in wild-type neurons at all peak regions identified in WT and DNMT-TKO cells binned by motif methylation.

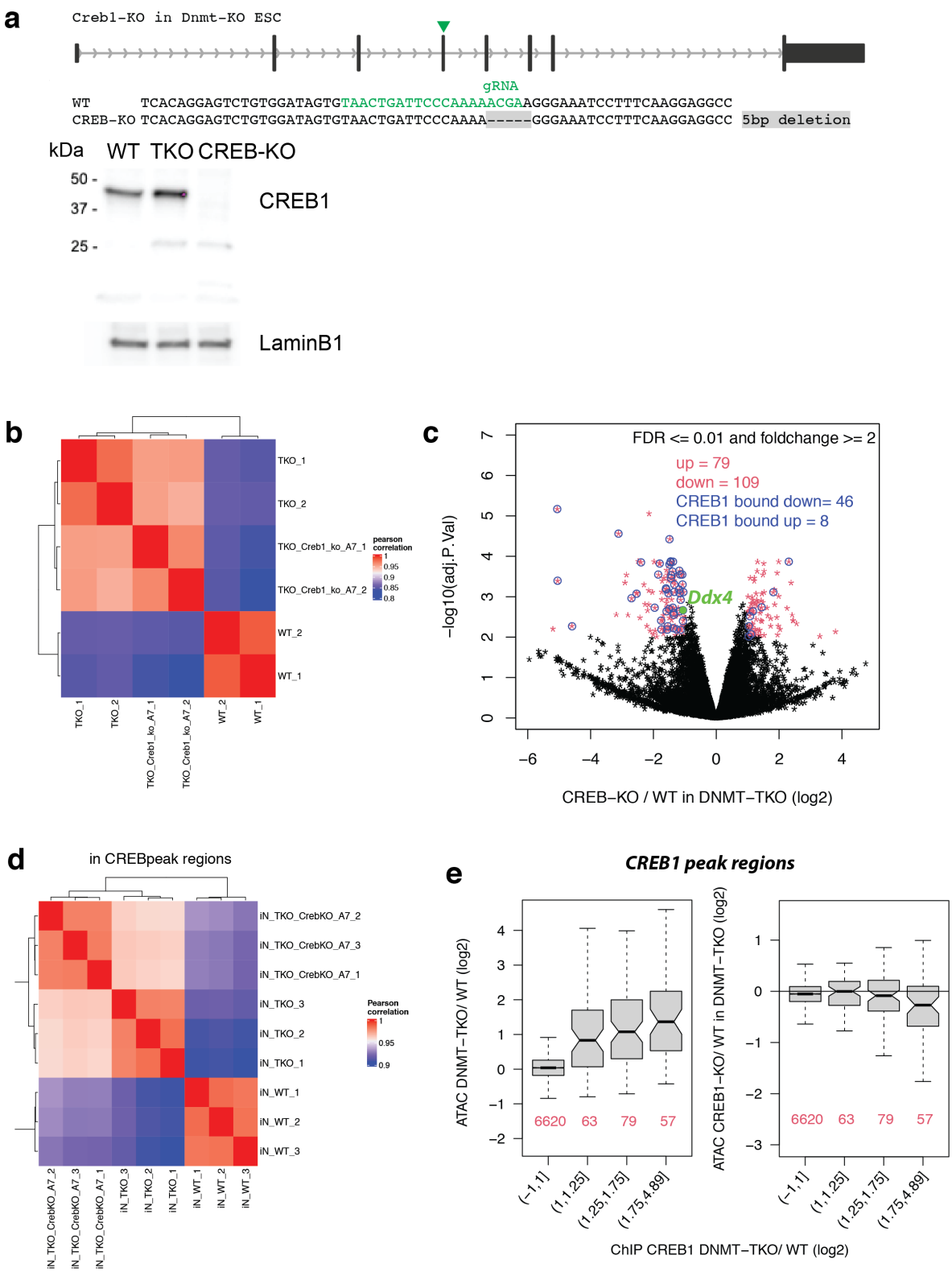
n) Gene Ontology (GO) terms enriched in the set of CREB1 bound promoters (top 500). The dots represent the top 20 terms with highest gene ratio (fraction of genes represented in the given GO term) with dot size and color representing gene counts and the adjusted p-value, respectively.


a
POL2 ChIP -seq in promoters
WT rep1
0.97
0.94
0.95
WT rep2
0.97
0.94
DNMT−TKO rep1
0.95
DNMT−TKO rep2
ChIP-seq POL2 counts (log2)
ChIP-seq POL2 counts (log2)
b
POL2
TKO_input_2
WT_Pol2_ab_2
WT_Pol2_sc_1
WT_input_1
WT_input_2
TKO_input_1
TKO_Pol2_ab_2
TKO_Pol2_sc_1
Pearson correlation
1
0.8
0.6
0.4
c
CREB1
TKO_CREB_r3
TKO_CREB_r1
TKO_CREB_r2
TKO_input_r3
WT_input_r3
WT_input_r2
TKO_input_r2
WT_CREB_r1
WT_CREB_r2
WT_CREB_r3
matrix_27
1
0.9
0.8
0.7
0.6
0.5
d
R= 0.67
IAPEY_LTR
IAPLTR1_Mm
IAPLTR1a_Mm
IAPLTR2_Mm
IAPEY3_LTR
IAPLTR2a
P1−MM_LTR
IAPLTR4
IAPLTR3
IAPEY2_LTR
IAPLTR2b
IAPEY5_LTR
IAPLTR2a2_Mm
IAPEY3C_LTR
POL2 DNMT−TKO/ WT (log2)
RNA DNMT−TKO/ WT (log2)
R= 0.62
IAPLTR1a_Mm
IAPLTR1_Mm
IAPEY_LTR
IAPLTR2a
IAPLTR2_Mm
IAPLTR2b
P1−MM_LTR
IAPEY2_LTR
IAPLTR4
IAPEY3_LTR
IAPLTR3
IAPLTR2a2_Mm
IAPEY5_LTR
IAPEY3C_LTR
ATAC DNMT−TKO/ WT (log2)
RNA DNMT−TKO/ WT (log2)

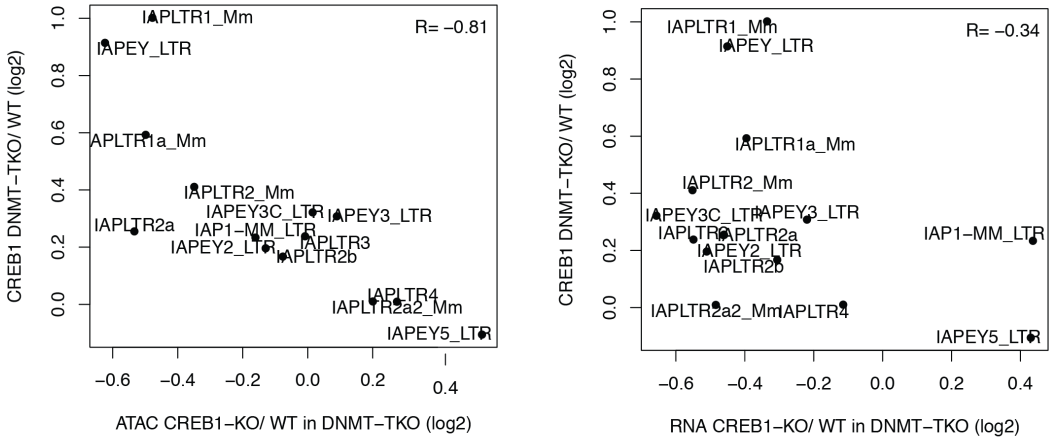

e
R= 0.66
IAPLTR1_Mm
IAPEY_LTR
IAPLTR1a_Mm
IAPLTR2_Mm
IAPEY3_LTR
IAPEY3C_LTR
IAP1−MM_LTR
IAPEY2_LTR
IAPLTR2a
IAPLTR3
IAPLTR2b
IAPLTR4
IAPLTR2a2_Mm
IAPEY5_LTR
CREB1 DNMT−TKO/ WT (log2)
fraction 5' LTR with CRE or half−site
f
IAPLTR1_Mm n= 670
IAPLTR1a_Mm n= 811
IAPEY_Mm n= 271
accessibility
DNMT−TKO/ WT (log2)
rep1
rep2
POL2
CREB1
IP/ Input (log2)
WT rep1
WT rep2
TKO rep1
TKO rep2
position from repeat start (nt)

Supplementary Figure 7 POL2 and CREB1 binding are detectable at IAPLTR1/1a and IAPEY elements in absence of DNA methylation.

a) Reproducibility of read counts for two independent POL2 ChIP-seq replicates from wild-type and DNMT-TKO neurons in promoter regions. Pearson correlation coefficients are indicated.
b) Unsupervised clustering of POL2 ChIP-seq samples from wild-type and DNMT-TKO (TKO) neurons. Colors indicate pairwise Pearson's correlation coefficients (PCC) of uniquely mapped read counts in 5' LTRs of annotation curated IAP elements that are de-repressed in absence of DNA methylation, indicating a clear separation from DNMT-TKO POL2 samples from WT and input samples.
c) Same as b) but with CREB1 ChIP-seq samples from WT and DNMT-TKO neurons.
d) Changes in POL2 binding in 5' LTR regions of annotation curated IAP elements versus expression change between WT and DNMT-TKO neurons measured by RNA-seq. Only uniquely mapped reads considered, replicates per condition are combined (n=2).
e) Changes ATAC-seq signal in 5' LTR regions of annotation curated IAP elements versus expression change between WT and DNMT-TKO neurons measured by RNA-seq. Only uniquely mapped reads considered, replicates per condition are combined (n=3).
f) Changes in CREB1 binding in 5' LTR regions of annotation curated IAP elements versus the fraction of LTRs containing a CRE or CRE half-site. Only uniquely mapped reads considered, replicates per condition are combined (n=3).
g) Changes in chromatin accessibility (top tracks, ATAC-seq), POL2 binding (middle tracks, ChIP-seq) or CREB1 binding (bottom tracks, ChIP-seq) in WT and DNMT-TKO neurons at IAPLTR1/1a elements that gain expression in absence of DNA methylation (FDR <= 0.05 & foldchange >2). Signal is centered at the start site of 5' LTR. Grey bars depicts average width of the 5' LTR and dashed lines display the average length of an entire element including the 5' and 3' LTR regions. Only unique mapped reads are considered. Individual replicates per condition are depicted. Number of elements are indicated by n.


a
Creb1-KO in Dnmt-KO ESC
gRNA
WT TCACAGGAGTCTGTGGATAGTGTAACTGATTCCCAAAAACGAAGGGAAATCCTTTCAAGGAGGCC
CREB-KO TCACAGGAGTCTGTGGATAGTGTAACTGATTCCCAAAA-----GGGAAATCCTTTCAAGGAGGCC 5bp deletion
kDa WT TKO CREB-KO
50 -
37 -
25 -
CREB1
LaminB1
b
TKO_1
TKO_2
TKO_Creb1_ko_A7_1
TKO_Creb1_ko_A7_2
WT_2
WT_1
pearson correlation
1
0.95
0.9
0.85
0.8
c
FDR <= 0.01 and foldchange >= 2
up = 79
down = 109
CREB1 bound down= 46
CREB1 bound up = 8
Ddx4
−log10(adj.P.Val)
CREB−KO / WT in DNMT−TKO (log2)
d
in CREBpeak regions
iN_TKO_CrebKO_A7_2
iN_TKO_CrebKO_A7_3
iN_TKO_CrebKO_A7_1
iN_TKO_3
iN_TKO_2
iN_TKO_1
iN_WT_1
iN_WT_2
iN_WT_3
Pearson correlation
1
0.95
0.9
e
CREB1 peak regions
ATAC DNMT−TKO/ WT (log2)
ATAC CREB1−KO/ WT in DNMT−TKO (log2)
6620
63
79
57
(−1,1]
(1,1.25]
(1.25,1.75]
(1.75,4.89]
ChIP CREB1 DNMT−TKO/ WT (log2)

f


IAPLTR1_Mm
IAPEY_LTR
IAPLTR1a_Mm
IAPLTR2_Mm
IAPEY3C_LTR
IAPEY3_LTR
IAPLTR2a
IAP1−MM_LTR
IAPEY2_LTR
IAPLTR3
IAPLTR2b
IAPLTR4
IAPLTR2a2_Mm
IAPEY5_LTR
R= −0.81
CREB1 DNMT−TKO/ WT (log2)
ATAC CREB1−KO/ WT in DNMT−TKO (log2)
R= −0.34
RNA CREB1−KO/ WT in DNMT−TKO (log2)

Supplementary Figure 8 CREB1 deletion in DNMT-TKO neurons causes reduced chromatin accessibility and transcription

a) Shema of CREB1 gene with exons in black and CRISPR cutting site indicated by green triangle. Sequence below displays a 5 bp deletion in CREB1 deleted DNMT-TKO neurons. Western blot depicts CREB1 signal in WT and DNMT-TKO neurons, which is absent CREB1-KO cells. Loading control using LaminB1.
b) Unsupervised clustering of RNA-seq samples from WT and mutant neuron cells counted in genes (RPKM). PCC, Pearson's correlation coefficient.
c) Volcano plot showing gene expression changes between WT or CREB1 deleted DNMT-TKO neurons.
d) Unsupervised clustering of ATAC-seq samples from WT and mutant neuron cells. Colors indicate pairwise Pearson's correlation coefficients (PCC) of log-transformed normalized read counts in ATAC-seq peaks, indicating reproducibility between replicates and separation of all three genotypes.
e) Box plots showing changes in chromatin accessibility (ATAC-seq) at CREB1 peak regions that gain binding in absence of DNA methylation. Accessibility changes between WT and DNMT-TKO neurons (left) and between WT and CREB1 deleted DNMT-TKO neurons (right). Replicate samples (n=3) per condition are combined.
f) Changes in CREB1 binding (ChIP-seq) between WT and DNMT-TKO neurons versus accessibility change (ATAC-seq) between CREB1 deleted or WT DNMT-TKO neurons in 5' LTR regions of annotation curated IAP elements. Only uniquely mapped reads considered, replicates (n=3) per condition are combined.
g) Changes in CREB1 binding (ChIP-seq) between WT and DNMT-TKO neurons versus expression change (RNA-seq) between CREB1 deleted or WT DNMT-TKO neurons in annotation curated IAP elements. Only uniquely mapped reads considered, replicates per condition are combined.

Methods

Cell culture

HA36 mouse ES cells were maintained in Dulbecco's Modified Eagle Medium (Invitrogen), supplemented with 15 % Fetal Calf Serum (Invitrogen), 1x GlutaMax (ThermoScientific) and 1x non-essential amino acids (Gibco), 0.001% betamercaptoethanol (Sigma) and leukemia inhibitory factor (LIF; produced in-house). All experiments were performed with cells grown for several passages on plates coated with 0.2% gelatin (Sigma).

Cell line generation

Mouse HA36 ES cells with a stable integration of the *Neurogenin2* gene under control of pTRE-tight were a kind gift from the Jeff Chao lab (FMI). In order to generate MBD-QKO clones these cells were co-transfected (Lipofectamine 3000, Thermo Fisher Scientific) with two vectors encoding CRISPR/Cas9 and gRNAs against either *Mbd2* or *Mecp2* in addition with a puromycin selection marker. Puromycin resistant clones were genotyped for frameshift mutations, expanded and validated by western blot. The same process was repeated to delete *Mbd1* and *Mbd4* individually. This protocol was repeated again with a new set of gRNAs in order to retrieve a total of two biological replicates.

target	Used for clone #	gRNAs
Mbd1	1	GTTGAGCTGACTCGGTACTT
Mbd1	2	ACAGGGTAAGATCACATGCA
Mbd2	1	CCCAGGTATCTTGCCAGCTG
Mbd2	2	GTTCAGAAGTAAACCTCAGC
Mbd4	1	ATCCACACGGGACAGGCTTG
Mbd4	2	CAACTCTTTCCCATCCACAC
Mecp2	1	CATACATAGGTCCCCGGTCA
Mecp2	2	CCTCGGCTTCCCCCAAACAG

In order to generate DNMT-TKO cells the three DNA methyltransferases *Dnmt1*, *Dnmt3a* and *Dnmt3b* were deleted by CRISPR-Cas9 gene editing as previously described (Domcke et al., 2015). *Dnmt* genes were sequenced to confirm successful targeting of all six alleles and residual methylation levels were measured by Zymo Research (www.zymoresearch.com), using high-pressure liquid chromatography coupled to mass spectrometry.

Antibodies

target	Vendor and catalog number	Application
MBD1	Santa Cruz, sc10751 (M254)	WB
MBD2	Abcam, ab188474	WB
MBD4	Santa Cruz, sc365974	WB
MECP2	Millipore, 07-013	WB
CREB1	Santa Cruz, sc377154X	ChIP, WB
POL2	Santa Cruz, sc899X and Abcam, ab8WG16	ChIP
HNF6	R&D systems AF6277	ChIP
laminB1	Abcam, ab188474	WB

Recombinase-mediated cassette exchange (RMCE)

For targeted insertion, the IAPLTR1a_Mm consensus sequence with or without the CRE (downloaded from repbase (Bao, Kojima, and Kohany 2015)) or a PGK promoter were cloned into a plasmid containing a multiple cloning site flanked by two inverted L1 *Lox* sites. Recombinase-mediated cassette exchange was performed in WT or DNMT-TKO HA36 mouse embryonic stem cells as previously described (Lienert et al. 2011).

Neuronal differentiation

For HA36 ES cells containing the pTRE-*Ngn2* construct differentiation was carried out by inducing expression of NGN2 with 1 µg/mL doxycycline as previously described (Thoma et al., 2012). Neurons were harvested 8 or 9 days after induction. For RNA-seq or ATAC-seq experiments the protocol was slightly modified in that differentiating

cells at day three were not re-plated, but proliferation media was exchanged to differentiation media according to protocol. After day three 50 % media was exchanged with fresh differentiation media every second day.

Luciferase Assay

WT or DNMT-TKO HA36 cells carrying pTRE-Ngn2 expression and reporter cassettes were differentiated into neurons. Luciferase activity in ES cells and neurons at day 8 were measured with the Luciferase Assay System (Promega, #E1500) according to manufacturer's instructions. Normalization was carried out by protein concentration of lysed ES cells or neurons in 1x lysis buffer with Protein Assay (Biorad, #500006). Luminescence was measured using a luminometer (Berthold Technologies, Centro XS3 LB 960)

Quantification of cell viability

Media from neuron cell lines cultured in 6 wells was aspirated at day 8 and 10 and a staining mix of 1ul Heochst, 8ul Propidium Iodide (PI) and 10ul AnnexinV in 125 ul AnnexinV binding buffer (ThermoFischer, #V13242) was added. After 15 min of incubation at 37 ºC, images were acquired with ZOE Fluorescent Cell imager (Bio-Rad, #145-0031) and analyzed using ImageJ. In brief, nuclei of cells were segmented based on Hoechst signal and determined as region of interest (ROI). Then AnnexinV and PI signal was calculated withing the ROI and normalized to background signal (outside of ROI). Nuclei containing both AnnexinV and PI enrichments (ROI vs background) with $\log_2$FC < 1.5 were considered healthy.

RNA-seq

RNA was isolated from pellets of mES cell with the RNeasy mini kit (Qiagen) using on-column DNA digestion or of neurons (eight days after dox induction) with Direct-zol RNA Microprep (Zymo research, catalog number R2061) with on-column DNA digestion. Sequencing libraries were prepared from purified RNA for a minimum of two biological replicates per condition using TruSeq stranded total RNA Library Prep (Illumina). ES libraries were sequenced on an a HiSeq platform with 50 cycles single

end. Neuron libraries were sequenced on an Illumina NextSeq platform with paired-end reads with 150 cycles (2x75 bp).

ChIP-seq

ChIP was carried out as previously described (Barisic et al., 2019) with the following modifications: (1) chromatin was sonicated for 20 cycles of 30 s using a Diagenode Bioruptor Pico, with 30 s breaks in between cycles; (2) protein A magnetic Dynabeads Magnetic beads (Invitrogen, catalog number 10008D) were used; and (3) DNA was purified using AMPure XP beads. Immunoprecipitated DNA and input DNA were submitted to library preparation (NEB Next Ultra DNA Library Prep Kit, Illumina). In the library preparation protocol, input samples (200 ng) were amplified using 6 PCR cycles and immunoprecipitation samples using 12 cycles. Libraries were sequenced paired-end 150 cycles (2x75bp) on the Illumina HiSeq 2500.

ATAC-seq

ATAC-seq was performed according to a previously described protocol (Buenrostro et al. 2015). Briefly, about 50,000 cells were washed with cold phosphate-buffered saline and resuspended in lysis buffer to extract the nuclei. The nuclei were cold centrifuged at 500*g* for 10 min. The nuclei pellet was incubated with transposition reaction buffer for 30 min at 37 °C. The DNA was purified using the MinElute PCR Purification Kit (QIAGEN). The eluted transposed DNA was submitted to PCR using Q5 High-Fidelity Polymerase (New England Biolabs). DNA was amplified with 11–12 cycles of PCR. The libraries were sequenced on the Illumina NextSeq platform at 41 bp paired-end. All ATAC-seq experiments were performed in at least two independent biological replicates per condition.

Curation of IAP elements

RepeatMasker annotation was downloaded from the UCSC genome annotation database for the Dec. 2011 (GRCm38/mm10) assembly of the mouse genome

ftp://hgdownload.cse.ucsc.edu/goldenPath/mm10/database/rmskOutBaseline.txt.gz.
In order to curate the fragmented RepeatMasker annotation of IAP proviruses, we grouped all IAP fragments based on repName and ID (describes related fragments, provided by RepeatMasker). Then we selected the start and end position of the first or last fragment, respectively, and generated a new fragment (ID fragment). In case two ID fragments of the same subfamily were in close proximity (< 1024bp) we combined both ID fragments. This gave rise to IAP elements containing either one LTR or 5' and 3' LTRs.

RNA-seq analysis in genes and repeats

For gene expression analysis, reads were aligned to the mm10 genome using Hisat2 in QuasR with splicedAlignment=T, aligner= "Rhisat2" and maxHits=1 for single-end 50 bp or maxHits=100 for 2x75 bp RNA-seq (Gaidatzis et al. 2015; Kim, Langmead, and Salzberg 2015). Alignments overlapping the opposite (single-end) or any (paired-end) strand of any exon of a gene or RepeatMasker annotation were counted using the qCount function from the QuasR Bioconductor package version 1.30.0 with default parameters.
Differential expression analysis (adjusted P value and foldchange cutoffs are indicated in the manuscript) was performed using edgeR's normalization combined with voom transformation from limma R package with default settings for gene analysis (Robinson, McCarthy, and Smyth 2010; Law et al. 2014; Ritchie et al. 2015). For differential expression analysis of repetitive elements library size and normalization factors for samples in edgeR's generated DGEList object were set manually. Library size for each sample was set using the sum of genic counts and the normalization factor for each sample was set using genic normalization factors calculated by $\min(n_s)/n_s$ where n_s are the summed uniquely mapped reads in genes per sample. Overrepresentation of gene ontology categories in selected gene sets was analysed using the clusterProfiler (Yu et al. 2012).
For comparison of published datasets, all of our own and published total RNA-seq samples were trimmed to 50 bp read length and aligned to the mm10 genome using Hisat2 in QuasR with splicedAlignment=T, aligner= "Rhisat2" and considering unique (maxHits=1) or multi mapped reads (maxHits = 100) (Gaidatzis et al. 2015; Kim, Langmead, and Salzberg 2015). Repeat counts in RepeatMasker annotation from

multi mapped aligned reads were collapsed to repName level for further analysis. Repeat counts were normalised to genic library size (Unique aligned reads counted in TxDb.Mmusculus.UCSC.mm10.knownGene using qCount) and replicates merged to the average read counts.

ChIP-seq analysis

Paired-end (2x75bp) ChIP-seq reads were aligned to the genome using qAlign from the QuasR package version 1.30.0 using bowtie with default parameters, only reporting alignments for reads with a unique match to the genome (Gaidatzis et al. 2015). For TF ChIP–seq data, peaks were called using macs2 v.2.2.6 (Zhang et al., 2008) with parameters callpeak -t IP.bam --f BAMPE -g mm -q 0.05. Aligned reads were counted in peak summits resize to 201 bp using the qCount function from the QuasR package v. 1.30.0, whereby reads were shifted by half the average fragment length. In each case counts were scaled down to the smaller library size. For each dataset, enrichment over input in peaks was defined as $\log2(counts_{IP} + 8) - \log2(counts_{nput} + 8)$, using a pseudo-count of 8 to decrease noise levels in case of low read counts. Only peaks with a reproducible log2 enrichment of at least 1 were retained for further analysis. For each TF, a binding motif was identified using the top 500 peaks ordered by enrichment using findMotifsGenome.pl from HOMER v.4.11 (Heinz et al. 2010) with parameters -len 8,10,12,14 -size given -noknown. The best (first) motif identified are depicted in the manuscript.
For repeat analysis, uniquely mapped reads were counted in the curated IAP annotation of 5' LTRs (shift="halfInsert") and normalized to the smallest library size. Library size normalized counts for each replicate per condition were summed, a pseudo-count of 8 added and log-transformed. For each dataset and condition, enrichments were defined as $\log2(norm\_counts_{condition\_A} + 8) - \log2(norm\_counts_{condition\_B} + 8)$.

ATAC-seq analysis

Adapters were first filtered using the cutadapt software (Martin 2011). Both reads were trimmed by one bp at the 3' end to allow for mapping of overlapping reads. Reads were mapped to the mm10 genome using Rbowtie in the QuasR package (v1.30.0)

(Gaidatzis et al. 2015). Mitochondrial reads were subsequently removed using samtools. For ATAC–seq data, peaks of individual replicates were called using macs2 v.2.2.6 (Zhang et al., 2008) with parameters callpeak -t IP.bam -f BAMPE -g mm -q 0.05. Uniquely mapped alignments were counted in peaks or promoters using the qCount function from the QuasR package v.1.30.0 using the default settings. Peaks were normalized according to the sample with the smallest promoter read counts.
Differential expression analysis (adjusted P value and foldchange cutoffs are indicated in the manuscript) was performed using edgeR's normalization combined with voom transformation (normalize="quantile") from limma R package (Robinson, McCarthy, and Smyth 2010; Law et al. 2014; Ritchie et al. 2015). ATAC–seq metaprofiles centered around NRF1 (from JASPAR2018 (Khan et al., 2018), MA0506.1) and BANP (Grand et al., unpublished) motifs were generated using the qProfile function from the QuasR Bioconductor package (v.1.30.0). Read counts in a 1.5-kb window anchored by the TF-binding motif were normalized by multiplying each profile with the minimum library size, then smoothed with a running mean of 51 bp and divided by the maximum read count of any condition.
In order to search for known TF motifs enriched in DNMT-TKO specific ATAC-seq peaks, peak regions were removed overlapping with CpG islands (about 16%, 71179/84400). TF motifs were then searched using findMotifsGenome.pl from HOMER v.4.11 (Heinz et al. 2010) with parameters -size given -mknown using the JASPAR2018 motif database (Khan et al., 2018).

ChIP-seq and ATAC-seq metaprofiles in IAP elements

ChIP-seq or ATAC-seq metaprofiles were generated with qProfile from the QuasR package v. 1.30.0 as followed: uniquely mapped reads were counted -1000 nt and +8000 nt from the start position of 5' LTR of the curated IAP annotation and normalized to the smallest library size. Profiles were then smoothed with a running sum of 201 bp to account for noise resulting from a low number of read counts due to an impaired mappability rate in repetitive regions. Library size normalized counts for each replicate per condition were summed, a pseudo-count of 8 added and log-transformed resulting in norm_counts. For each dataset and condition, enrichments were defined as $\log2(\text{norm\_counts}_{\text{condition\_A}} + 8) - \log2(\text{norm\_counts}_{\text{condition\_B}} + 8)$.

Public data sets

RNA-seq datasets were obtained from GEO: P5 mouse cortex *cUhrf1* KO (GSM2241736/39/40) and matching heterozygote (GSM2241735/37) (Ramesh et al. 2016). ES *cSetdb1 cDnmt1* KO (GSM2059172/3) and matching WT (GSM2059171) (Sharif et al., 2016), E8.5 whole embryos *Dnmt1*-KO (GSM3752651/52/53) and matching WT (GSM3752646/47/48) (Dahlet et al., 2020).

Kmer enrichments

Peaks outside of CpG islands were resized to 401 nucleotides around their midpoints (in order to avoid larger contributions to 6mer counts from longer peaks), split into bins according to changes in ATAC-seq signal ($\log2(n\_counts_{DNMT\text{-}TKO} + 8) - \log2(n\_counts_{WT} + 8)$, where n_counts are counts in peaks normalized to promoter counts) and for each bin separately, the number of occurrences of each 6mer in each peak sequence was determined. To avoid a strong influence of very repetitive sequences, we performed zoops (zero-or-one-occurrence) counting, i.e. a 6mer that occurred multiple times in a peak sequence was counted only once. Finally, we summed all the 6mer counts over all sequences of the peak set resulting in a set of foreground 6mer frequencies. As a background model, we estimated expected 6mer frequencies using a second order Markov model. To be consistent with zoops counting, we fitted the Markov model not on the original peak sequences, but on an artifical set of 6mer sequences that contained each 6mer as many times as its foreground frequency. Finally, we calculated, for each 6mer, its enrichment as (nfg – nbg)/sqrt(nbg) (Pearson residual), where nfg and nbg are the foreground and background frequencies of a particular 6mer.

References

Amir, Ruthie E., Ignatia B. Van den Veyver, Mimi Wan, Charles Q. Tran, Uta Francke, and Huda Y. Zoghbi. 1999. “Rett Syndrome Is Caused by Mutations in X-Linked MECP2 , Encoding Methyl-CpG-Binding Protein 2.” *Nature Genetics* 23 (2): 185–88.

Auclair, Ghislain, and Michael Weber. 2012. “Mechanisms of DNA Methylation and Demethylation in Mammals.” *Biochimie* 94 (11): 2202–11.

Ballester, Benoit, Alejandra Medina-Rivera, Dominic Schmidt, Mar Gonzàlez-Porta, Matthew Carlucci, Xiaoting Chen, Kyle Chessman, et al. 2014. “Multi-Species, Multi-Transcription Factor Binding Highlights Conserved Control of Tissue-Specific Biological Pathways.” *ELife* 3 (October): e02626.

Bao, Weidong, Kenji K. Kojima, and Oleksiy Kohany. 2015. “Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes.” *Mobile DNA* 6 (1): 1–6.

Barisic, Darko, Michael B. Stadler, Mario Iurlaro, and Dirk Schübeler. 2019. “Mammalian ISWI and SWI/SNF Selectively Mediate Binding of Distinct Transcription Factors.” *Nature* 569 (7754): 136–40.

Baubec, Tuncay, Robert Ivánek, Florian Lienert, and Dirk Schübeler. 2013. “Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family.” *Cell* 153 (2): 480–92.

Bednarik, D. P., C. Duckett, S. U. Kim, V. L. Perez, K. Griffis, P. C. Guenthner, and T. M. Folks. 1991. “DNA CpG Methylation Inhibits Binding of NF-Kappa B Proteins to the HIV-1 Long Terminal Repeat Cognate DNA Motifs.” *The New Biologist* 3 (10): 969–76.

Benbrook, D. M., and N. C. Jones. 1994. “Different Binding Specificities and Transactivation of Variant CRE’s by CREB Complexes.” *Nucleic Acids Research* 22 (8): 1463–69.

Bestor, T. H. 1990. “DNA Methylation: Evolution of a Bacterial Immune Function into a Regulator of Gene Expression and Genome Structure in Higher Eukaryotes.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 326 (1235): 179–87.

Bestor, Timothy H. 2003. “Cytosine Methylation Mediates Sexual Conflict.” *Trends in Genetics: TIG* 19 (4): 185–90.

Bourc’his, D., G. L. Xu, C. S. Lin, B. Bollman, and T. H. Bestor. 2001. “Dnmt3L and the Establishment of Maternal Genomic Imprints.” *Science* 294 (5551): 2536–39.

Brûlet, P., M. Kaghad, Y. S. Xu, O. Croissant, and F. Jacob. 1983. “Early Differential Tissue Expression of Transposon-like Repetitive DNA Sequences of the Mouse.” *Proceedings of the National Academy of Sciences of the United States of America* 80 (18): 5641–45.

Buenrostro, Jason D., Beijing Wu, Howard Y. Chang, and William J. Greenleaf. 2015. “ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide.” *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* 109 (1): 21.29.1-21.29.9.

Busslinger, M., J. Hurst, and R. A. Flavell. 1983. “DNA Methylation and the Regulation of Globin Gene Expression.” *Cell*. https://doi.org/10.1016/0092-8674(83)90150-2.

Campanero, M. R., M. I. Armstrong, and E. K. Flemington. 2000. “CpG Methylation as a Mechanism for the Regulation of E2F Activity.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (12): 6481–86.

Caselli, Elisabetta, Sabrina Benedetti, Jessica Grigolato, Arnaldo Caruso, and Dario Di Luca. 2012. “Activating Transcription Factor 4 (ATF4) Is Upregulated by Human Herpesvirus 8 Infection, Increases Virus Replication and Promotes Proangiogenic Properties.” *Archives of Virology* 157 (1): 63–74.

Chelmicki, Tomasz, Emeline Roger, Aurélie Teissandier, Mathilde Dura, Lorraine Bonneville, Sofia Rucli, François Dossin, Camille Fouassier, Sonia Lameiras, and Deborah Bourc'his. 2021. "M6A RNA Methylation Regulates the Fate of Endogenous Retroviruses." *Nature*, January. https://doi.org/10.1038/s41586-020-03135-1.

Chen, R. Z., S. Akbarian, M. Tudor, and R. Jaenisch. 2001. "Deficiency of Methyl-CpG Binding Protein-2 in CNS Neurons Results in a Rett-like Phenotype in Mice." *Nature Genetics* 27 (3): 327–31.

Chen, Taiping, Sarah Hevi, Frédérique Gay, Naomi Tsujimoto, Timothy He, Bailin Zhang, Yoshihide Ueda, and En Li. 2007. "Complete Inactivation of DNMT1 Leads to Mitotic Catastrophe in Human Cancer Cells." *Nature Genetics* 39 (3): 391–96.

Cheng, Tian-Lin, Zhizhi Wang, Qiuming Liao, Ying Zhu, Wen-Hao Zhou, Wenqing Xu, and Zilong Qiu. 2014. "MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8/Drosha Complex." *Developmental Cell* 28 (5): 547–60.

Cholewa-Waclaw, Justyna, Ruth Shah, Shaun Webb, Kashyap Chhatbar, Bernard Ramsahoye, Oliver Pusch, Miao Yu, Philip Greulich, Bartlomiej Waclaw, and Adrian P. Bird. 2019. "Quantitative Modelling Predicts the Impact of DNA Methylation on RNA Polymerase II Traffic." *Proceedings of the National Academy of Sciences of the United States of America* 116 (30): 14995–0.

Dahlet, Thomas, Andrea Argüeso Lleida, Hala Al Adhami, Michael Dumas, Ambre Bender, Richard P. Ngondo, Manon Tanguy, et al. 2020. "Genome-Wide Analysis in the Mouse Embryo Reveals the Importance of DNA Methylation for Transcription Integrity." *Nature Communications* 11 (1): 3153.

Dantas Machado, Ana Carolina, Tianyin Zhou, Satyanarayan Rao, Pragya Goel, Chaitanya Rastogi, Allan Lazarovici, Harmen J. Bussemaker, and Remo Rohs. 2014. "Evolving Insights on How Cytosine Methylation Affects Protein–DNA Binding." *Briefings in Functional Genomics* 14 (1): 61–73.

Domcke, Silvia, Anaïs Flore Bardet, Paul Adrian Ginno, Dominik Hartl, Lukas Burger, and Dirk Schübeler. 2015. "Competition between DNA Methylation and Transcription Factors Determines Binding of NRF1." *Nature* 528 (7583): 575–79.

Dupressoir, A., and T. Heidmann. 1996. "Germ Line-Specific Expression of Intracisternal A-Particle Retrotransposons in Transgenic Mice." *Molecular and Cellular Biology* 16 (8): 4495–4503.

Egger, Gerda, Shinwu Jeong, Sonia G. Escobar, Connie C. Cortez, Tony W. H. Li, Yoshimasa Saito, Christine B. Yoo, Peter A. Jones, and Gangning Liang. 2006. "Identification of DNMT1 (DNA Methyltransferase 1) Hypomorphs in Somatic Knockouts Suggests an Essential Role for DNMT1 in Cell Survival." *Proceedings of the National Academy of Sciences of the United States of America* 103 (38): 14080–85.

Fan, G., C. Beard, R. Z. Chen, G. Csankovszki, Y. Sun, M. Siniaia, D. Biniszkiewicz, et al. 2001. "DNA Hypomethylation Perturbs the Function and Survival of CNS Neurons in Postnatal Animals." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 21 (3): 788–97.

Fatemi, Mehrnaz, and Paul A. Wade. 2006. "MBD Family Proteins: Reading the Epigenetic Code." *Journal of Cell Science* 119 (Pt 15): 3033–37.

Gabel, Harrison W., Benyam Kinde, Hume Stroud, Caitlin S. Gilbert, David A. Harmin, Nathaniel R. Kastan, Martin Hemberg, Daniel H. Ebert, and Michael E. Greenberg. 2015. "Disruption of DNA-Methylation-Dependent Long Gene Repression in Rett Syndrome." *Nature* 522 (7554): 89–93.

Gaidatzis, Dimos, Lukas Burger, Maria Florescu, and Michael B. Stadler. 2015. "Analysis of

Intronic and Exonic Reads in RNA-Seq Data Characterizes Transcriptional and Post-Transcriptional Regulation." *Nature Biotechnology* 33 (7): 722–29.
Göke, Jonathan, Xinyi Lu, Yun-Shen Chan, Huck-Hui Ng, Lam-Ha Ly, Friedrich Sachs, and Iwona Szczerbinska. 2015. "Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells." *Cell Stem Cell* 16 (2): 135–41.
Grant, Christian, Unsong Oh, Kazunori Fugo, Norihiro Takenouchi, Caitlin Griffith, Karen Yao, Timothy E. Newhook, Lee Ratner, and Steven Jacobson. 2006. "Foxp3 Represses Retroviral Transcription by Targeting Both NF-KB and CREB Pathways." *PLoS Pathogens* 2 (4): e33.
Grow, Edward J., Ryan A. Flynn, Shawn L. Chavez, Nicholas L. Bayless, Mark Wossidlo, Daniel J. Wesche, Lance Martin, et al. 2015. "Intrinsic Retroviral Reactivation in Human Preimplantation Embryos and Pluripotent Cells." *Nature* 522 (7555): 221–25.
Guy, J., B. Hendrich, M. Holmes, J. E. Martin, and A. Bird. 2001. "A Mouse Mecp2-Null Mutation Causes Neurological Symptoms That Mimic Rett Syndrome." *Nature Genetics* 27 (3): 322–26.
Hai, Tsonwin, and Matthew G. Hartman. 2001. "The Molecular Biology and Nomenclature of the Activating Transcription Factor/CAMP Responsive Element Binding Family of Transcription Factors: Activating Transcription Factor Proteins and Homeostasis." *Gene*. https://doi.org/10.1016/s0378-1119(01)00551-0.
Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities." *Molecular Cell* 38 (4): 576–89.
Hendrich, B., and A. Bird. 1998. "Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins." *Molecular and Cellular Biology* 18 (11): 6538–47.
Hendrich, B., J. Guy, B. Ramsahoye, V. A. Wilson, and A. Bird. 2001. "Closely Related Proteins MBD2 and MBD3 Play Distinctive but Interacting Roles in Mouse Development." *Genes & Development* 15 (6): 710–23.
Hendrich, Brian, and Susan Tweedie. 2003. "The Methyl-CpG Binding Domain and the Evolving Role of DNA Methylation in Animals." *Trends in Genetics: TIG* 19 (5): 269–77.
Hutnick, Leah K., Peyman Golshani, Masakasu Namihira, Zhigang Xue, Anna Matynia, X. William Yang, Alcino J. Silva, Felix E. Schweizer, and Guoping Fan. 2009. "DNA Hypomethylation Restricted to the Murine Forebrain Induces Cortical Degeneration and Impairs Postnatal Neuronal Maturation." *Human Molecular Genetics* 18 (15): 2875–88.
Iguchi-Ariga, S. M., and W. Schaffner. 1989. "CpG Methylation of the CAMP-Responsive Enhancer/Promoter Sequence TGACGTCA Abolishes Specific Factor Binding as Well as Transcriptional Activation." *Genes & Development* 3 (5): 612–19.
Illingworth, Robert S., and Adrian P. Bird. 2009. "CpG Islands--'a Rough Guide'." *FEBS Letters* 583 (11): 1713–20.
Iyaguchi, Daisuke, Min Yao, Nobuhisa Watanabe, Jun Nishihira, and Isao Tanaka. 2007. "DNA Recognition Mechanism of the ONECUT Homeodomain of Transcription Factor HNF-6." *Structure* 15 (1): 75–83.
Jackson, Melany, Anna Krassowska, Nick Gilbert, Timothy Chevassut, Lesley Forrester, John Ansell, and Bernard Ramsahoye. 2004. "Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem

Cells." *Molecular and Cellular Biology* 24 (20): 8862–71.
Jackson-Grusby, Laurie, Caroline Beard, Richard Possemato, Matthew Tudor, Douglas Fambrough, Györgyi Csankovszki, Jessica Dausman, et al. 2001. "Loss of Genomic Methylation Causes P53-Dependent Apoptosis and Epigenetic Deregulation." *Nature Genetics*. https://doi.org/10.1038/83730.
Jaenisch, Rudolf, and Adrian Bird. 2003. "Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals." *Nature Genetics* 33 Suppl (March): 245–54.
Jolma, Arttu, Jian Yan, Thomas Whitington, Jarkko Toivonen, Kazuhiro R. Nitta, Pasi Rastas, Ekaterina Morgunova, et al. 2013. "DNA-Binding Specificities of Human Transcription Factors." *Cell* 152 (1–2): 327–39.
Jones, Peter A. 2012. "Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond." *Nature Reviews. Genetics* 13 (7): 484–92.
Karimi, Mohammad M., Preeti Goyal, Irina A. Maksakova, Misha Bilenky, Danny Leung, Jie Xin Tang, Yoichi Shinkai, et al. 2011. "DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in MESCs." *Cell Stem Cell* 8 (6): 676–87.
Khan, Aziz, Oriol Fornes, Arnaud Stigliani, Marius Gheorghe, Jaime A. Castro-Mondragon, Robin van der Lee, Adrien Bessy, et al. 2018. "JASPAR 2018: Update of the Open-Access Database of Transcription Factor Binding Profiles and Its Web Framework." *Nucleic Acids Research* 46 (D1): D1284.
Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. 2015. "HISAT: A Fast Spliced Aligner with Low Memory Requirements." *Nature Methods* 12 (4): 357–60.
Kirby, Helen, Alan Rickinson, and Andrew Bell. 2000. "The Activity of the Epstein--Barr Virus BamHI W Promoter in B Cells Is Dependent on the Binding of CREB/ATF Factors." *Microbiology* 81 (4): 1057–66.
Klose, Robert J., and Adrian P. Bird. 2006. "Genomic DNA Methylation: The Mark and Its Mediators." *Trends in Biochemical Sciences* 31 (2): 89–97.
Kribelbauer, Judith F., Oleg Laptenko, Siying Chen, Gabriella D. Martini, William A. Freed-Pastor, Carol Prives, Richard S. Mann, and Harmen J. Bussemaker. 2017. "Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes." *Cell Reports* 19 (11): 2383–95.
Kribelbauer, Judith F., Xiang-Jun Lu, Remo Rohs, Richard S. Mann, and Harmen J. Bussemaker. 2020. "Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors." *Journal of Molecular Biology* 432 (6): 1801–15.
Lagger, Sabine, John C. Connelly, Gabriele Schweikert, Shaun Webb, Jim Selfridge, Bernard H. Ramsahoye, Miao Yu, et al. 2017. "MeCP2 Recognizes Cytosine Methylated Tri-Nucleotide and Di-Nucleotide Sequences to Tune Transcription in the Mammalian Brain." *PLoS Genetics* 13 (5): e1006793.
Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409 (6822): 860–921.
Law, Charity W., Yunshun Chen, Wei Shi, and Gordon K. Smyth. 2014. "Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts." *Genome Biology* 15 (2): R29.
Leung, Danny C., and Matthew C. Lorincz. 2012. "Silencing of Endogenous Retroviruses: When and Why Do Histone Marks Predominate?" *Trends in Biochemical Sciences* 37 (4): 127–33.
Li, E., C. Beard, A. C. Forster, T. H. Bestor, and R. Jaenisch. 1993. "DNA Methylation, Genomic Imprinting, and Mammalian Development." *Cold Spring Harbor Symposia*

on Quantitative Biology 58: 297–305.
Li, E., T. H. Bestor, and R. Jaenisch. 1992. “Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality.” *Cell* 69 (6): 915–26.
Liao, Jing, Rahul Karnik, Hongcang Gu, Michael J. Ziller, Kendell Clement, Alexander M. Tsankov, Veronika Akopian, et al. 2015. “Targeted Disruption of DNMT1, DNMT3A and DNMT3B in Human Embryonic Stem Cells.” *Nature Genetics* 47 (5): 469–78.
Lienert, Florian, Christiane Wirbelauer, Indrani Som, Ann Dean, Fabio Mohn, and Dirk Schübeler. 2011. “Identification of Genetic Elements That Autonomously Determine DNA Methylation States.” *Nature Genetics* 43 (11): 1091–97.
Mager, Dixie L., and Jonathan P. Stoye. 2015. “Mammalian Endogenous Retroviruses.” *Microbiology Spectrum* 3 (1): MDNA3-0009–2014.
Maksakova, Irina A., Mark T. Romanish, Liane Gagnier, Catherine A. Dunn, Louie N. van de Lagemaat, and Dixie L. Mager. 2006. “Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line.” *PLoS Genetics* 2 (1): e2.
Mancini, D. N., S. M. Singh, T. K. Archer, and D. I. Rodenhiser. 1999. “Site-Specific DNA Methylation in the Neurofibromatosis (NF1) Promoter Interferes with Binding of CREB and SP1 Transcription Factors.” *Oncogene* 18 (28): 4108–19.
Marchi, Emanuele, Alex Kanapin, Gkikas Magiorkinis, and Robert Belshaw. 2014. “Unfixed Endogenous Retroviral Insertions in the Human Population.” *Journal of Virology* 88 (17): 9529–37.
Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.Journal* 17 (1): 10–12.
Maunakea, Alika K., Iouri Chepelev, Kairong Cui, and Keji Zhao. 2013. “Intragenic DNA Methylation Modulates Alternative Splicing by Recruiting MeCP2 to Promote Exon Recognition.” *Cell Research* 23 (11): 1256–69.
Messerschmidt, Daniel M., Barbara B. Knowles, and Davor Solter. 2014. “DNA Methylation Dynamics during Epigenetic Reprogramming in the Germline and Preimplantation Embryos.” *Genes & Development* 28 (8): 812–28.
Millar, Catherine B., Jacky Guy, Owen J. Sansom, Jim Selfridge, Eilidh MacDougall, Brian Hendrich, Peter D. Keightley, Stefan M. Bishop, Alan R. Clarke, and Adrian Bird. 2002. “Enhanced CpG Mutability and Tumorigenesis in MBD4-Deficient Mice.” *Science* 297 (5580): 403–5.
Millhouse, Scott, Joseph J. Kenny, Patrick G. Quinn, Vivien Lee, and Brian Wigdahl. 1998. “ATF/CREB Elements in the Herpes Simplex Virus Type 1 Latency-Associated Transcript Promoter Interact with Members of the ATF/CREB and AP-1 Transcription Factor Families.” *Journal of Biomedical Science*. https://doi.org/10.1007/bf02255935.
Molaro, Antoine, and Harmit S. Malik. 2016. “Hide and Seek: How Chromatin-Based Pathways Silence Retroelements in the Mammalian Germline.” *Current Opinion in Genetics & Development* 37 (April): 51–58.
Montminy, M. R., and L. M. Bilezikjian. 1987. “Binding of a Nuclear Protein to the Cyclic-AMP Response Element of the Somatostatin Gene.” *Nature* 328 (6126): 175–78.
Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. 1998. “Transcriptional Repression by the Methyl-CpG-Binding Protein MeCP2 Involves a Histone Deacetylase Complex.” *Nature* 393 (6683): 386–89.
Nichols, Jennifer, and Austin Smith. 2009. “Naive and Primed Pluripotent States.” *Cell Stem Cell* 4 (6): 487–92.
Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. “DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for de Novo Methylation and Mammalian Development.” *Cell* 99 (3): 247–57.

Padeken, Jan, Peter Zeller, and Susan M. Gasser. 2015. “Repeat DNA in Genome Organization and Stability.” *Current Opinion in Genetics & Development* 31 (April): 12–19.

Panning, B., and R. Jaenisch. 1996. “DNA Hypomethylation Can Activate Xist Expression and Silence X-Linked Genes.” *Genes & Development* 10 (16): 1991–2002.

Prendergast, G. C., D. Lawe, and E. B. Ziff. 1991. “Association of Myn, the Murine Homolog of Max, with c-Myc Stimulates Methylation-Sensitive DNA Binding and Ras Cotransformation.” *Cell* 65 (3): 395–407.

Ramesh, Vidya, Efil Bayam, Filippo M. Cernilogar, Ian M. Bonapace, Markus Schulze, Markus J. Riemenschneider, Gunnar Schotta, and Magdalena Götz. 2016. “Loss of Uhrf1 in Neural Stem Cells Leads to Activation of Retroviral Elements and Delayed Neurodegeneration.” *Genes & Development* 30 (19): 2199–2212.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. “Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. “EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26 (1): 139–40.

Rowe, Helen M., Johan Jakobsson, Daniel Mesnard, Jacques Rougemont, Séverine Reynard, Tugce Aktas, Pierre V. Maillard, et al. 2010. “KAP1 Controls Endogenous Retroviruses in Embryonic Stem Cells.” *Nature* 463 (7278): 237–40.

Saito, Motoki, and Fuyuki Ishikawa. 2002. “The MCpG-Binding Domain of Human MBD3 Does Not Bind to MCpG but Interacts with NuRD/Mi2 Components HDAC1 and MTA2.” *The Journal of Biological Chemistry* 277 (38): 35434–39.

Schübeler, D., M. C. Lorincz, D. M. Cimbora, A. Telling, Y. Q. Feng, E. E. Bouhassira, and M. Groudine. 2000. “Genomic Targeting of Methylated DNA: Influence of Methylation on Transcription, Replication, Chromatin Structure, and Histone Acetylation.” *Molecular and Cellular Biology* 20 (24): 9103–12.

Schumacher, M. A., R. H. Goodman, and R. G. Brennan. 2000. “The Structure of a CREB BZIP.Somatostatin CRE Complex Reveals the Basis for Selective Dimerization and Divalent Cation-Enhanced DNA Binding.” *The Journal of Biological Chemistry* 275 (45): 35242–47.

Seisenberger, Stefanie, Simon Andrews, Felix Krueger, Julia Arand, Jörn Walter, Fátima Santos, Christian Popp, Bernard Thienpont, Wendy Dean, and Wolf Reik. 2012. “The Dynamics of Genome-Wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells.” *Molecular Cell* 48 (6): 849–62.

Sen, George L., Jason A. Reuter, Daniel E. Webster, Lilly Zhu, and Paul A. Khavari. 2010. “DNMT1 Maintains Progenitor Function in Self-Renewing Somatic Tissue.” *Nature* 463 (7280): 563–67.

Shaknovich, Rita, Leandro Cerchietti, Lucas Tsikitas, Matthias Kormaksson, Subhajyoti De, Maria E. Figueroa, Gianna Ballon, et al. 2011. “DNA Methyltransferase 1 and DNA Methylation Patterning Contribute to Germinal Center B-Cell Differentiation.” *Blood* 118 (13): 3559–69.

Sharif, Jafar, Takaho A. Endo, Manabu Nakayama, Mohammad M. Karimi, Midori Shimada, Kayoko Katsuyama, Preeti Goyal, 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* 19 (1): 81–94.

Smith, Brandon, Hung Fang, Youlian Pan, P. Roy Walker, A. Fazel Famili, and Marianna Sikorska. 2007. “Evolution of Motif Variants and Positional Bias of the Cyclic-AMP

Response Element." *BMC Evolutionary Biology* 7 Suppl 1 (February): S15.
Spruijt, Cornelia G., Felix Gnerlich, Arne H. Smits, Toni Pfaffeneder, Pascal W. T. C. Jansen, Christina Bauer, Martin Münzel, et al. 2013. "Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives." *Cell* 152 (5): 1146–59.
Steven, André, Michael Friedrich, Paul Jank, Nadine Heimer, Jan Budczies, Carsten Denkert, and Barbara Seliger. 2020. "What Turns CREB on? And off? And Why Does It Matter?" *Cellular and Molecular Life Sciences: CMLS* 77 (20): 4049–67.
Storer, Jessica, Robert Hubley, Jeb Rosen, Travis J. Wheeler, and Arian F. Smit. 2021. "The Dfam Community Resource of Transposable Element Families, Sequence Models, and Genome Annotations." *Mobile DNA* 12 (1): 2.
Tang, Walfred W. C., Sabine Dietmann, Naoko Irie, Harry G. Leitch, Vasileios I. Floros, Charles R. Bradshaw, Jamie A. Hackett, Patrick F. Chinnery, and M. Azim Surani. 2015. "A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development." *Cell* 161 (6): 1453–67.
Thoma, Eva C., Erhard Wischmeyer, Nils Offen, Katja Maurus, Anna-Leena Sirén, Manfred Schartl, and Toni U. Wagner. 2012. "Ectopic Expression of Neurogenin 2 Alone Is Sufficient to Induce Differentiation of Embryonic Stem Cells into Mature Neurons." *PloS One* 7 (6): e38651.
Tierney, R. J., H. E. Kirby, J. K. Nagra, J. Desmond, A. I. Bell, and A. B. Rickinson. 2000. "Methylation of Transcription Factor Binding Sites in the Epstein-Barr Virus Latent Cycle Promoter Wp Coincides with Promoter down-Regulation during Virus-Induced B-Cell Transformation." *Journal of Virology* 74 (22): 10468–79.
Tillotson, Rebekah, Justyna Cholewa-Waclaw, Kashyap Chhatbar, John C. Connelly, Sophie A. Kirschner, Shaun Webb, Martha V. Koerner, et al. 2021. "Neuronal Non-CG Methylation Is an Essential Target for MeCP2 Function." *Molecular Cell*, January. https://doi.org/10.1016/j.molcel.2021.01.011.
Tippin, D. B., and M. Sundaralingam. 1997. "Nine Polymorphic Crystal Structures of d (CCGGGCCCGG), d (CCGGGCCm5CGG), d (Cm5CGGGCCm5CGG) and d (CCGGGCC (Br) 5CGG) in Three Different …." *Journal of Molecular Biology*. https://www.sciencedirect.com/science/article/pii/S0022283697909451.
Toufaily, Chirine, Adjimon Gatien Lokossou, Amandine Vargas, Éric Rassart, and Benoit Barbeau. 2015. "A CRE/AP-1-like Motif Is Essential for Induced Syncytin-2 Expression and Fusion in Human Trophoblast-like Model." *PloS One* 10 (3): e0121468.
Tsumura, Akiko, Tomohiro Hayakawa, Yuichi Kumaki, Shin-ichiro Takebayashi, Morito Sakaue, Chisa Matsuoka, Kunitada Shimotohno, et al. 2006. "Maintenance of Self-renewal Ability of Mouse Embryonic Stem Cells in the Absence of DNA Methyltransferases Dnmt1, Dnmt3a and Dnmt3b." *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 11 (7): 805–14.
Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. "Transcription of IAP Endogenous Retroviruses Is Constrained by Cytosine Methylation." *Nature Genetics* 20 (2): 116–17.
Wang, Liguo, Junsheng Chen, Chen Wang, Liis Uusküla-Reimand, Kaifu Chen, Alejandra Medina-Rivera, Edwin J. Young, et al. 2014. "MACE: Model Based Analysis of ChIP-Exo." *Nucleic Acids Research* 42 (20): e156.
Watt, F., and P. L. Molloy. 1988. "Cytosine Methylation Prevents Binding to DNA of a HeLa Cell Transcription Factor Required for Optimal Expression of the Adenovirus Major Late Promoter." *Genes & Development* 2 (9): 1136–43.
Weber, Michael, Ines Hellmann, Michael B. Stadler, Liliana Ramos, Svante Pääbo, Michael Rebhan, and Dirk Schübeler. 2007. "Distribution, Silencing Potential and

Evolutionary Impact of Promoter DNA Methylation in the Human Genome." *Nature Genetics* 39 (4): 457–66.
Wildschutte, Julia Halo, Zachary H. Williams, Meagan Montesion, Ravi P. Subramanian, Jeffrey M. Kidd, and John M. Coffin. 2016. "Discovery of Unfixed Endogenous Retrovirus Insertions in Diverse Human Populations." *Proceedings of the National Academy of Sciences of the United States of America* 113 (16): E2326-34.
Yamamoto, K. K., G. A. Gonzalez, P. Menzel, J. Rivier, and M. R. Montminy. 1990. "Characterization of a Bipartite Activator Domain in Transcription Factor CREB." *Cell* 60 (4): 611–17.
Yin, Yimeng, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, et al. 2017. "Impact of Cytosine Methylation on DNA Binding Specificities of Human Transcription Factors." *Science* 356 (6337). https://doi.org/10.1126/science.aaj2239.
Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. "Cytosine Methylation and the Ecology of Intragenomic Parasites." *Trends in Genetics: TIG* 13 (8): 335–40.
Young, Juan I., Eugene P. Hong, John C. Castle, Juan Crespo-Barreto, Aaron B. Bowman, Matthew F. Rose, Dongcheul Kang, et al. 2005. "Regulation of RNA Splicing by the Methylation-Dependent Transcriptional Repressor Methyl-CpG Binding Protein 2." *Proceedings of the National Academy of Sciences of the United States of America* 102 (49): 17551–58.
Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters." *Omics: A Journal of Integrative Biology* 16 (5): 284–87.
Żemojtel, Tomasz, Szymon M. Kiełbasa, Peter F. Arndt, Sarah Behrens, Guillaume Bourque, and Martin Vingron. 2011. "CpG Deamination Creates Transcription Factor–Binding Sites with High Efficiency." *Genome Biology and Evolution* 3 (October): 1304–11.
Zhang, Xinmin, Duncan T. Odom, Seung-Hoi Koo, Michael D. Conkright, Gianluca Canettieri, Jennifer Best, Huaming Chen, et al. 2005. "Genome-Wide Analysis of CAMP-Response Element Binding Protein Occupancy, Phosphorylation, and Target Gene Activation in Human Tissues." *Proceedings of the National Academy of Sciences of the United States of America* 102 (12): 4459–64.
Zhang, Yingsha, Changhui Pak, Yan Han, Henrik Ahlenius, Zhenjie Zhang, Soham Chanda, Samuele Marro, et al. 2013. "Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells." *Neuron* 78 (5): 785–98.
Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology* 9 (9): R137.
Zhao, Xinyu, Tetsuya Ueba, Brian R. Christie, Basam Barkho, Michael J. McConnell, Kinichi Nakashima, Edward S. Lein, et al. 2003. "Mice Lacking Methyl-CpG Binding Protein 1 Have Deficits in Adult Neurogenesis and Hippocampal Function." *Proceedings of the National Academy of Sciences of the United States of America* 100 (11): 6777–82.

MBD proteins play a limited role in DNA methylation mediated repression in human cells

Introduction

Deletion of 5m-cytosine binding MBD proteins in mouse ES cells and derived neurons only results in limited changes in transcription and in the gene regulatory landscape. This is in stark contrast to the phenotype of methylation deficient ES-derived neurons that display stark upregulation of germline-specific genes and transposable elements. While this argues for a limited role for MBD proteins in an indirect repression model, it remains unclear if these findings are specific to ES cells and derived post-mitotic neurons, and/or to the mouse system. Therefore, we sought to investigate the role of MBD proteins in translating DNA methylation into transcriptional repression outside the murine lineage in a human somatic and proliferating cell line that depends on DNA methylation for cell viability.

Results

HEK293 cells depleted of DNA methylation or deleted for MBD proteins show distinct transcriptional phenotypes.

In order to evaluate the role of MBD proteins in translating DNA methylation mediated repression in human cells, we sought to simultaneously delete *MBD1*, *MBD2*, *MBD4* and *MECP2* in HEK293 cells by introducing frameshift mutations into the MBD encoding exons using CRISPR/Cas9 **(Fig. 1a)**. Double deletions of *MBD2* and *MECP2* in mice have been shown to be compatible with embryogenesis (Martín Caballero et al. 2009), suggesting that these proteins are not essential for viability in somatic cells. Therefore, we first targeted *MBD2* and *MECP2,* confirmed frameshift mutations by genotyping **(Fig. 1b)** and validated the absence of both proteins by western blot **(Fig. 1c)**. Next, we continued to target *MBD1* and *MBD4* in serial and assessed deletions as before. Further validation of the introduces indels and resulting lack of protein confirmed that we successfully generated HEK293 cells with a quadruple knockout of all 5mC binding MBD proteins (henceforth MBD-QKO) **(Fig 1b,c)**. MBD-QKO HEK293 cells are viable, proliferate comparable to WT levels and show no apparent increased cell death determined by visual inspection **(Fig. 1d)**.

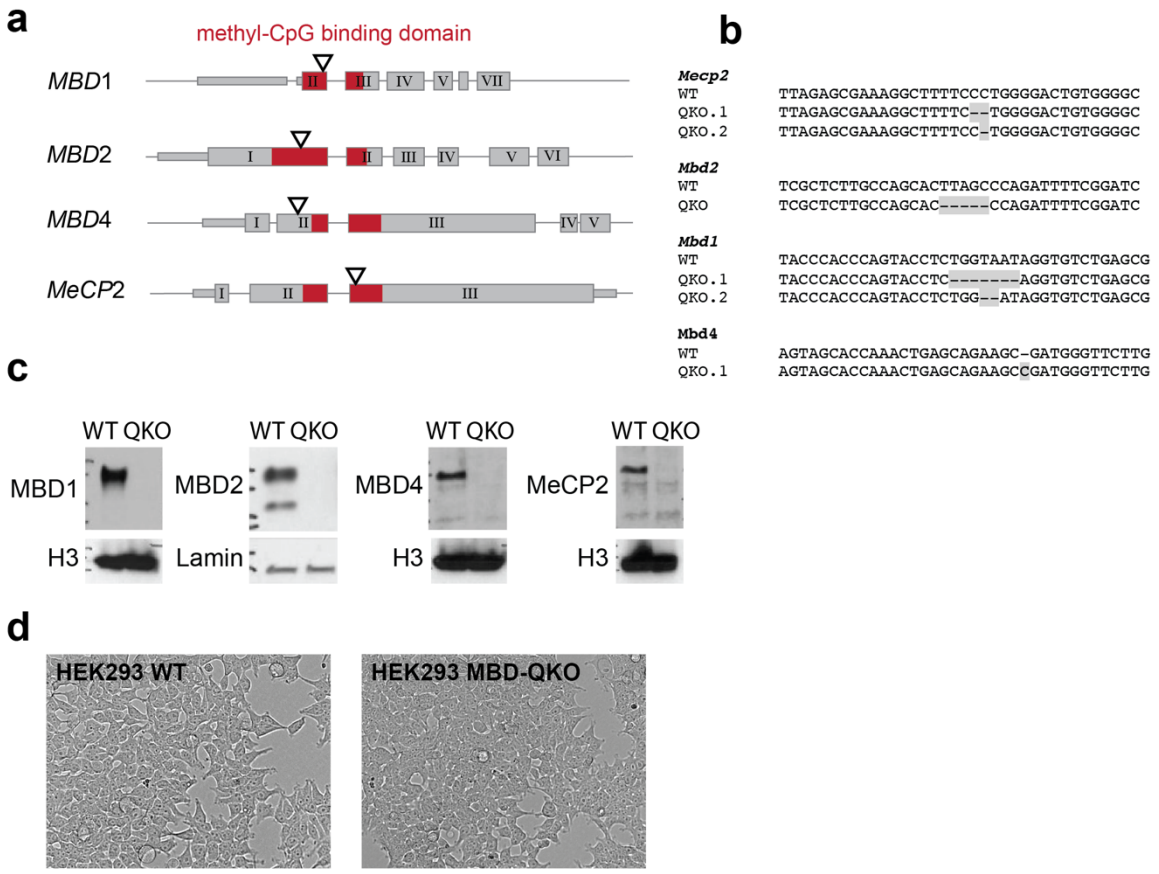


Figure 1. HEK293 cells deleted for MBD proteins are viable.

a) Representation of MBD protein coding genes, roman numerals indicate exons, red indicates MBD and the triangle CRISPR/Cas9 cutting sites.
b) Depiction of genomic DNA sequence around the black triangle depicted in a) for WT and MBD-QKO cells.
c) Western blots detecting MBD protein levels in WT and MBD-QKO HEK293 cells with respective loading control.
d) Pictures of WT and MBD-QKO HEK293 cell culture.

To assess the loss of MBD proteins on transcription we collected WT and MBD-QKO cells in triplicates and performed RNA-sequencing. This revealed a reproducible set of genes being differentially expressed between WT and MBD-QKO HEK293 with 219 genes up- and 281 genes downregulated (FDR <= 0.01 and fold change >=2) **(Fig. 2a)**. Gene ontology analysis in the upregulated set of genes determined weak enrichment of terms associated with different biological processes such as skeletal system development or postsynapse organization **(Fig. 2b)**, while no enrichment was detected in the downregulated set.

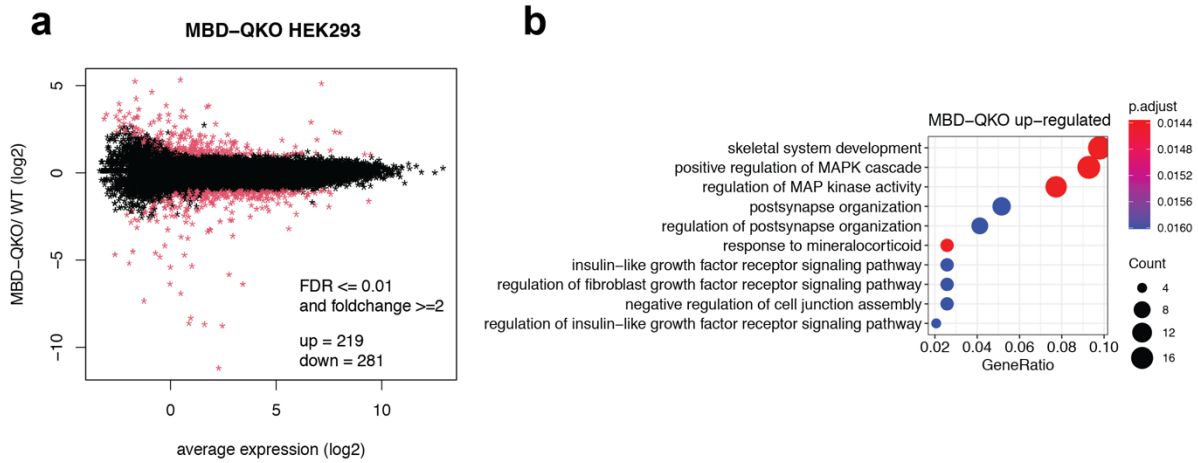


Figure 2. Gene expression changes in MBD-QKO neurons.

a) MA plot (changes vs. average expression) of gene expression changes in MBD-QKO HEK293 compared to their WT counterpart. Red asterisk indicates differentially expressed genes (FDR <= 0.01 and absolute fold change >= 2).
b) Gene Ontology (GO) terms enriched in the set of significantly upregulated genes between MBD-QKO and WT HEK293 cells. The dots represent the ten terms with highest gene ratio (fraction of genes represented in the given GO term) with dot size and color representing gene counts and the adjusted p-value, respectively.

Next, we sought to compare the transcriptome of MBD-QKO cells to HEK293 cells with globally reduced levels of DNA methylation, as an indirect repression model would predict similar or overlapping phenotypes.

Therefore, we treated WT cells with either DMSO (control) or 5-Aza-2′-deoxycytidine (5-Aza) a well-studied inhibitor of DNA methyltransferases that causes genome-wide hypomethylation (Christman 2002).

Treatment of WT cells with 1 µM 5-Aza for 72 hours only lead to minor cell death assessed by visual inspection, despite reported cell toxicity **(Fig. 3a)** (Christman 2002). To asses methylation levels we isolated genomic DNA after treatment and digested it with either the methylation-sensitive restriction endonuclease HpaII or its methylation-insensitive isoschizomer MspI (Waalwijk and Flavell 1978). As expected, this revealed a hypomethylated genome upon 5-Aza treatment **(Fig. 3b)**, in line with a previous report where a similar treatment reduced global methylation levels in HEK293 cells from 70% to 20% of all CpGs. (Ramos et al. 2015).

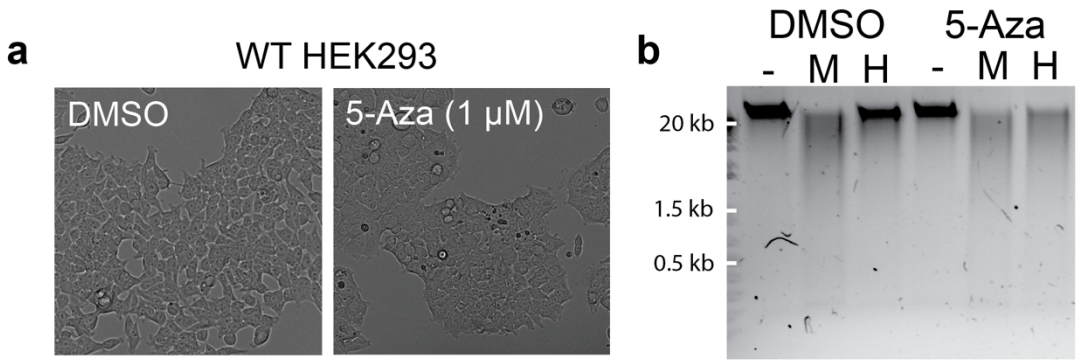


Figure 3. DNA hypomethylated HEK293 cells.

a) WT HEK293 cells treated for 72 h with 5-Aza are morphologically comparable to WT, with only limited cell death observed.
b) Digestion of genomic DNA (gDNA) from DMSO or 5-Aza treated HEK293 cells with restriction endonucleases MspI (M), HpaII (H) or water control (-). Both isoschizomers recognize the sequence CCGG, whereas HpaII is methylation sensitive and MspI insensitive. While gDNA from WT cells is mostly protected by HpaII treatment, as expected by global methylation of the genome, gDNA from 5-Aza treated cells is cleaved by HpaII to similar levels as compared to MspI digestion, indicating genome-wide hypomethylation.

Next, we collected WT cells either treated with DMSO or 5-Aza in triplicates and performed RNA sequencing, which resulted in extremely high correlation among replicates **(Fig. 4a)**. Differential gene expression analysis revealed hundreds of genes misregulated upon 5-Aza treatment, compared to the DMSO control **(Fig. 4b)**. We determined 1479 genes up- but only 349 genes down-regulated (FDR<= 0.01 and absolute fold-change >= 2) in line with DNA methylation being associated with transcriptional repression.

Importantly, genes that are most upregulated tend to be transcriptionally silent and enriched for germline genes **(Fig. 4c,d)**, which are typically methylated at their promoters (Weber et al., 2007). A prominent and classical example is *Deleted in azoospermia-like* (*DAZL)* that is upregulated in DNMT1-deficient human fibroblasts (O'Neill et al. 2018) and is indeed upregulated upon depletion of DNA methylation by 5-Aza treatment **(Fig. 4b,e)**. This argues that 5-Aza treatment of HEK293 cells induces a transcriptional response that particularly affects promoter methylated genes.

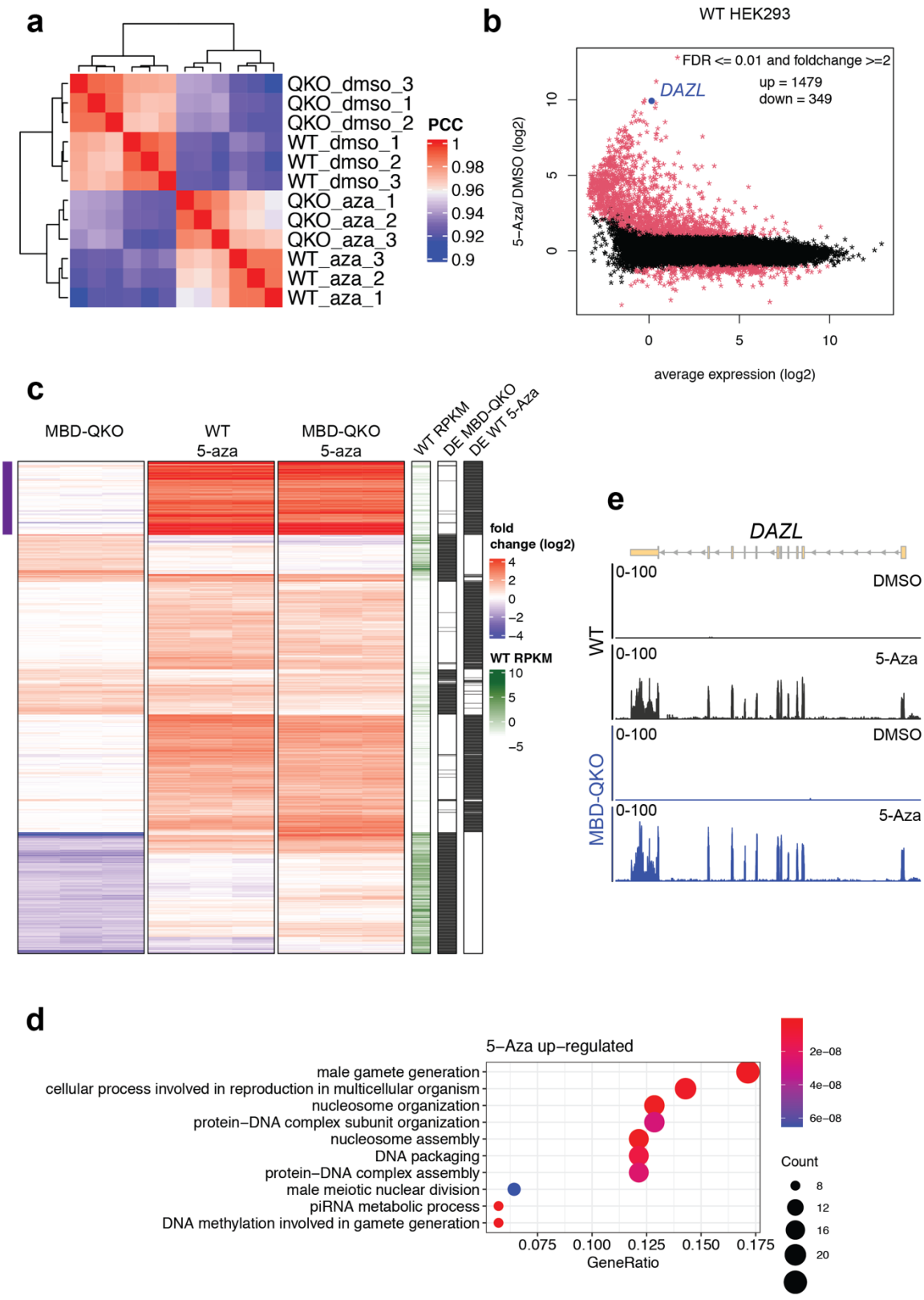


Figure 4. MBD-QKO HEK293 cells show a distinct transcriptional phenotype compared to cells that are depleted of DNA methylation

a) Pearson's correlation of gene expression (low RPKM removed) between WT and MBD-QKO cells treated with either 5-Aza or DMSO as a control.

b) MA plot of differentially expressed genes in 5-Aza treated WT HEK293 cells. Red asterisk indicates significantly changing genes (FDR <= 0.01 and fold change >= 2).
c) Hierarchical clustering of genes differentially expressed in MBD-QKO cells under DMSO conditions or upregulated (FDR < 0.01 and fold change >8) in WT cells treated with 5-Aza. Each row depicts expression fold changes (log2) for selected gene, normalized to the respective genotype under DMSO conditions. Minimum and maximum of coloring is set to the 1st and 3rd quartile of all gene expression changes for better visualization. The gene cluster with most upregulated genes under 5-Aza treatment is depicted with a purple bar. The green column depicts WT gene expression in RPKM under DMSO conditions. The black and white columns highlight differentially expressed genes (in black) for MBD-QKO depicted in Fig. 4a or 5-Aza treated WT cells (FDR < 0.01 and absolute fold change >= 8).
d) Gene Ontology (GO) terms enriched in genes most upregulated in WT cells under 5-Aza treatment, which are annotated with purple bar in c). Dots represent the ten terms with highest gene ratio (fraction of genes represented in the given GO term) with dot size and color representing gene counts and the adjusted p-value, respectively.
e) Single locus gene expression (read counts in 11nt running window, replicate data combined) of WT and MBD-QKO HEK293 cells under DMSO and 5-Aza conditions at *Deleted in azoospermia-like* (*DAZL)*. Indicated by blue dot in b).

Comparing genes that are misregulated in MBD-QKO and DNA hypomethylated HEK293 cells reveals only minor overlap. Of all genes that respond to 5-Aza treatment in WT cells, only 5% (n=17) of down- and 6% (n=95) of upregulated genes are affected in MBD-QKO cells.

Furthermore, the strong asymmetric transcriptional response (~ 4-fold more genes upregulated) driven by depletion of DNA methylation in WT cells is not observed in MBD-QKO cells, where slightly more genes are downregulated **(Fig. 2a)**. Of note, most genes differentially expressed in MBD-QKO cells are already transcriptionally active in WT **(Fig. 4c)**, arguing against a role of MBD proteins in tight transcriptional repression at these genes.

Importantly, genes that are highly de-repressed upon depletion of DNA methylation in WT cells are largely unaffected after deletion of MBD-proteins **(Fig 4c)**. These genes only show strong de-repression in MBD-QKO cells when treated with 5-Aza **(Fig. 4c,e)**. Of note, we do not observe repetitive elements strongly dysregulated in HEK293 cell lacking MBD proteins or 5-Aza treated **(Fig. 5)**.

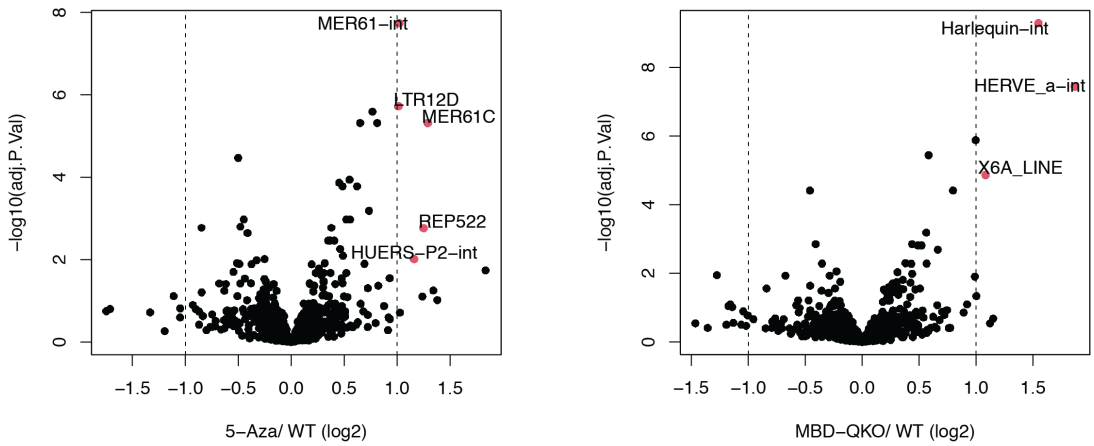


Figure 5. Volcano plots showing expression changes of individual subfamilies of repetitive elements annotated by RepeatMasker for HEK293 cells treated with 5-Aza (left) or lacking MBD proteins (right). Analysis conducted on uniquely mapping 75 bp paired-end RNA reads. Significantly changing subfamilies are colored in red (FDR < 0.01 & fold-change > 2).

Taken together, these distinct phenotypes argue against a critical role of MBD proteins in translating DNA methylation into transcriptional repression at methylated promoters.

Deletion of all MBD proteins causes only minor changes to chromatin accessibility in HEK293

Chromatin accessibility measurements provide information about activity of regulatory element such as promoter distal regulatory regions comprising TF binding events (O. Bell et al. 2011). In order to test if the absence of MBD proteins in HEK293 cells affected the regulatory landscape, we performed ATAC-seq on WT and MBD-QKO cells.

Counting reads in peaks called by MACS2 determined high replicate reproducibility **(Fig. 6a).** Comparing accessibility between WT and MBD-QKO cells revealed only limited yet reproducible changes, with slightly more sites decreasing in accessibility **(Fig 6b)**. This is in good agreement with the limited transcriptional phenotype observed in MBD-QKO cells. Differentially accessible peak regions are generally of smaller peak width than shared sites **(Fig. 6c)**, in line with most differentially bound sites occurring promoter distal **(Fig 6d)**. Of note, differentially accessible peaks are of low CpG density **(Fig 6e)**, which is at odds with MBD proteins binding preferentially to CpG dense and methylated regions (Baubec et al., 2013).

We conclude that the regulatory landscape in HEK293 cells measured by ATAC-seq is largely unaffected in the absence of MBD proteins.

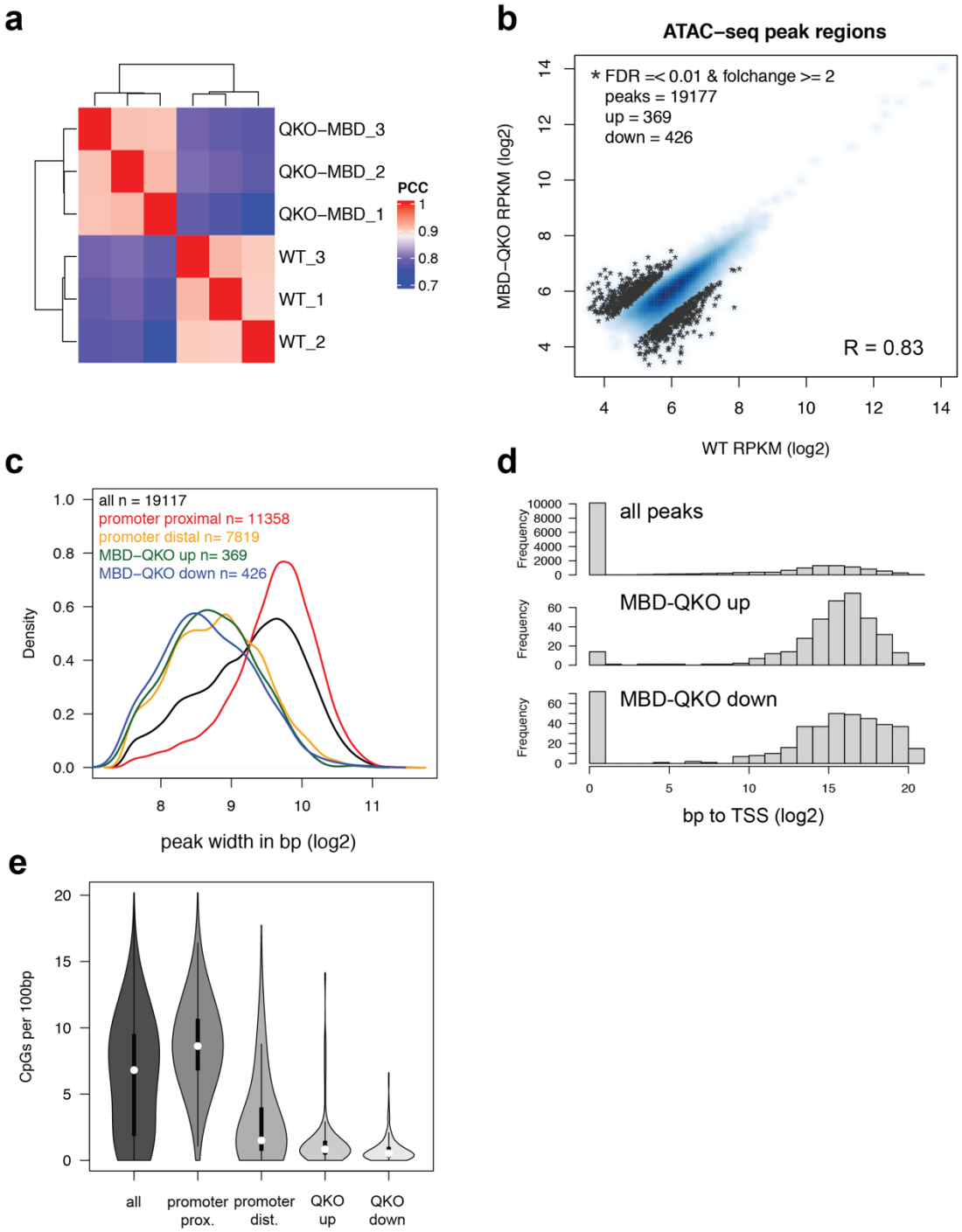


Figure 6. Accessibility changes between WT and MBD-QKO HEK293 cells are limited

a) Pearson's correlation between ATAC-seq read counts (RPKM) in combined peaks reproducibly identified in WT and MBD-QKO cells. PCC = Pearson's correlation coefficient.

b) ATAC-seq signal in WT and MBD-QKO cells at all 19,177 sites reproducibly accessible in at least one of the two cell lines. Differentially accessible sites (absolute fold-change >= 2, adjusted p-value =< 0.01) are indicated with a black asterisk. R = Pearson's correlation coefficient.

c) Comparison of peak size for all identified peaks in at least one of the two cell lines (black, n= 19177), separated by promoter distal (bp from TSS > 1024) or proximal (bp from TSS < 1024) or MBD-QKO peaks differentially accessible (FDR <= 0.01 and absolute fold-change >= 2).

d) Frequency distribution for distance of peaks to TSS for all peaks (upper panel), MBD-QKO enriched (middle panel) and depleted (bottom panel). Same peaks as in b).

e) Comparison of CpGs per peak normalized to peak width. Same peaks as in c).

Genome-wide DNA methylation is similar between wild-type and MBD-QKO HEK293 cells

Previous studies suggest that some MBD proteins interact with components of the DNA methylation maintenance machinery or are linked to the maintenance of DNA methylation at CpG islands during cancer (Kimura and Shiota 2003; Stirzaker et al. 2017). To asses if the absence of MBD proteins impacted genome-wide DNA methylation we performed whole-genome bisulfite sequencing (WGBS) from WT and MBD-QKO HEK293 cells. Resulting datasets display similar CpG read coverage between WT and MBD-QKO HEK293 (2.36 vs 2.56 reads per CpG, respectively) covering about 70% of all CpGs in the genome **(Fig 7a)**. As expected from previous observations in mammalian system, analysis of methylation frequencies in WT HEK293 of individual CpGs determined a bimodal distribution **(Fig. 7b)** of either unmethylated or fully methylated CpGs **(Fig. 7b)**. Calculating methylation levels of different genomic features revealed a pattern generally found in mammalian cells. While most features show high levels of methylation, promoters and regulatory regions (ATAC-seq peaks outside of promoters) depict low methylation levels in WT cells. The same analysis of the MBD-QKO methylome revealed a very similar pattern of methylation. While we do observe slightly higher methylation levels throughout the genome in MBD-QKO cells, it is unclear if this reflects clonal variability or due to the absence of MBD proteins. Taken together, we conclude that the absence of MBD proteins in HEK293 cells did not result in a gross change of DNA methylation patterns or levels. This argues for a limited role of MBD proteins in maintaining or shaping DNA methylation patterns.

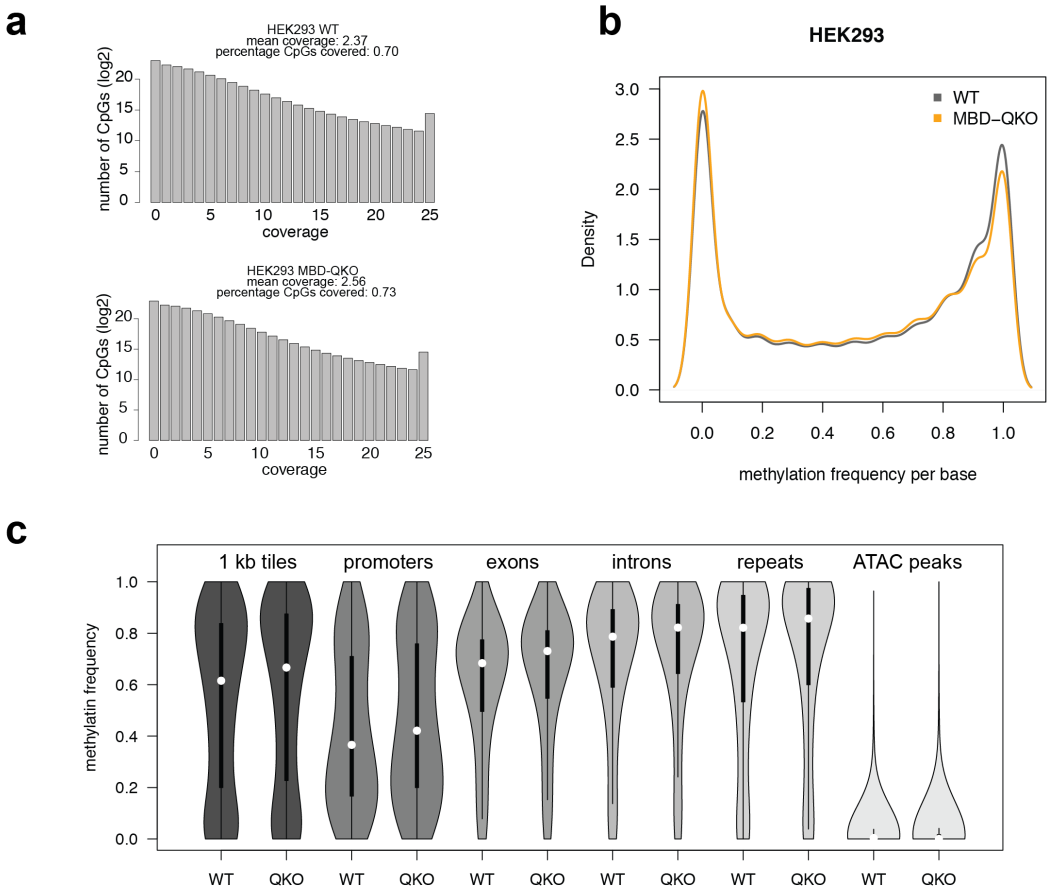


Figure 7. DNA methylation patterns are largely unaffected in absence of MBD proteins.

a) Bar plot of read coverage of all CpGs for each cell line.
b) Density plot of methylation frequencies of CpGs with minimum 10 reads in both cell lines.
c) Violin plot showing distribution of CpG methylation levels in 1 kb tiles of all chromosomes, promoters, exons, introns, repeats (not overlapping genes) and all ATAC-seq peaks identified in any cell line not overlapping genes.

Discussion

In mouse ES cells and derived neurons absence of MBD proteins did not result in a strong phenotype, arguing against a prominent role for MBDs in mediating indirect repression. Here we expanded on this observation by testing the biological function of 5mC binding MBD proteins in DNA methylation mediated repression in a human cell line using genetic deletions.

Using CRISPR/Cas9 we generate viable MBD-QKO HEK293 cells that simultaneously lack MBD1, MBD2, MBD4 and MECP2. This revealed a limited impact on transcription upon removal of MBD proteins, which does not resemble the transcriptional response of HEK293 cells with a 5-Aza induced hypomethylated genome. Importantly, germline-specific genes that are typically methylated at their promoters and reactivate upon DNA hypomethylation, remain silent in MBD-QKO cells. Furthermore, we do not observe a strong upregulation of repetitive elements in MBD protein lacking cells, which surprisingly does not happen also in 5-Aza treated cells. This could be due to the residual levels of DNA methylation that are not removed after 72h of 5-Aza treatment. Taken together, this argues that MBD proteins in HEK293 cells are dispensable for DNA methylation mediated repression.

In agreement with the limited transcriptional phenotype, genome-wide chromatin accessibility and DNA methylation patterns are largely unaffected in the absence of MBD proteins, which suggest no apparent role for MBD proteins in regulating TF binding or shaping the methylome. These results are in good agreement with the limited phenotype observed in mouse ES cells and derived neurons that lack the same MBD protein family. As in the murine system, we do not observe a strong phenotype in MBD-QKO HEK293 cells that would suggest an alternative molecular function of MBD proteins. Therefore, it will be intriguing to test if our findings also translate into the context of mammalian development by for instance generating a mouse line lacking all four MBD proteins. In conclusion, in both cell systems tested, we do not find evidence for MBD proteins playing a central role in DNA methylation mediated repression, challenging an indirect repression model where DNA methylation is targeted by sequence-unspecific methyl-binding proteins which impair binding of transcription factors and transcriptional repression.

5. Summary

Methylation of the DNA base cytosine (5mC) found in CpG dinucleotides of mammalian genomes is associated with transcriptional repression, primarily of CpG-rich promoters and transposable elements (TEs). While DNA methylation is a conserved and abundant silencing mechanism, the way in which it translates into repression is not yet clear. Two main models explain this regulatory function of DNA methylation: a direct repression model, whereby some transcription factors (TFs) bind DNA differentially according to the methylation status of their recognition motifs; and an indirect repression model, which postulates that methyl-CpG binding domain (MBD) proteins suppress transcription by recruiting repressors.

The goal of this study was to investigate how DNA methylation translates into transcriptional repression, and to determine whether one or both of these models contribute to DNA methylation-dependent silencing. To achieve this, we characterized the effects on gene regulation in absence of all 5mC binding MBD proteins in different mammalian cell lines. To further dissect the mechanism of DNA methylation repression, we explored the influence of DNA methylation on TF binding in somatic cells and systematically probed the role of methylation-sensitive CREB1 in regulating the activity of TEs.

MBD proteins

The advent of easily programmable genomic editing technologies such as CRISPR has extended our ability to perform multiple precise genomic manipulations within the same cell (Adli 2018). This has enabled us to address the question of functional redundancy between MBD proteins by performing a simultaneous deletion of all MBD proteins associated with reading DNA methylation.

Despite previous suppositions, we do not find evidence for a prominent role of these proteins in translating DNA methylation into transcriptional repression **(Figure 1)**. These findings are unexpected given that the biochemical *in vitro* evidence suggests a bridging model, whereby DNA methylation recruits chromatin-modifying MBD proteins in a sequence-independent fashion. However, we cannot formally exclude that unknown methyl-CpG binding proteins functionally compensate for the deleted MBD proteins.

While other proteins have been identified that carry either MBD-like sequences or other methyl-CpG binding domains, to date only MBD proteins were found to bind symmetrically methylated DNA in an sequence-independent fashion (Hendrich and Bird, 1998; Laget et al., 2010; Roloff et al., 2003; Tillotson and Bird, 2019).
A potential way to identify novel readers of DNA methylation could entail a genome-wide loss-of-function CRSIPR screen with a methylation-sensitive reporter. However, multiple simultaneous deletions would be necessary in order to account for genetic redundancy. Nonetheless, such a screen performed in MBD-QKO cells could reduce putative redundancies of 5mC readers and potentially elicit novel candidates.

Our findings thus join additional studies that challenged a model where MBD proteins translate binding to 5mC into transcriptional repression (Baubec et al., 2013; Cholewa-Waclaw et al., 2019; Martín Caballero et al., 2009). It is possible that MBD proteins are involved in other regulatory functions. For instance, MECP2 is described to impact alternative splicing (Maunakea et al., 2013) or impact elongation by binding to methylated CAs (Cholewa-Waclaw et al., 2019; Gabel et al., 2015; Lagger et al., 2017; Tillotson et al., 2021). However, the limited transcriptional response upon MBD deletion in the mouse or human cell lines tested here does not point towards an obvious alternative function. Relatedly, while MECP2 inactivity causes Rett syndrome, its molecular phenotype is associated with limited changes in gene expression (Tillotson and Bird 2019). It would be intriguing to generate a mouse model that lacks all 5mC binding MBD proteins, which would potentially uncover alternative roles of the MBD proteins in gene regulation.

Methylation-sensitive TFs in somatic cells

In parallel we find ample evidence for cases of direct repression as we identify multiple TF candidates that appear to be methylation-sensitive in neurons, and that could therefore directly mediate 5mC-based repression. We experimentally validate this for HNF6 and CREB1 in line with previous *in vitro* experiments (Yin et al., 2017). HNF6 presents an interesting case, as this TF is methylation-sensitive only at its CpG-containing motif variant (AT**CG**AT), in agreement with an *in vitro* methyl-SELEX study (Yin et al., 2017). Importantly, we do not see any influence of surrounding DNA methylation on binding of HNF6 to its canonical motif (AT**TG**AT), in line with the limited

evidence of an indirect repression model. It remains to be seen whether methylation of the HNF6 motif variant encompasses a regulatory role, i.e. whether this motif variant allows to regulate HNF6 binding in different tissues through differential motif methylation.

Recent *in vitro* experiments suggest that a much larger fraction of TFs (including HNF6 and CREB1) is repelled by DNA methylation than the number of TF candidates identified in this study (Kribelbauer et al., 2017; Yin et al., 2017). This could have multiple reasons. For instance, in our approach we search for motifs that are enriched in chromatin accessible sites that are specific to the unmethylated neuronal genome. This requires a TF to create chromatin accessibility footprints in sufficient numbers with enough motif occurrences in order to be detected by our method. This might limit the detection of TFs that only bind a small subset of CpG containing motifs or do not create detectable accessibility footprints in absence of DNA methylation. Another explanation could be that a methylation-sensitive TF is simply not expressed in our cell system. This for instance explains why HNF6 was previously not identified in ES cells (Domcke et al., 2015), where this TF is not expressed. It is therefore likely, that similar screens in additional cell lines will identify more methylation-sensitive TFs, enlarging the repertoire of TF that can be repelled by DNA methylation *in vivo*.

Regulation of TEs by DNA methylation

Mouse ES cells tolerate the loss of DNA methylation, which is owed to the presence of a DNA methylation independent silencing mechanism involving H3K9me3 that represses TEs in these cells **(Figure 1)**(Karimi et al., 2011; Leung and Lorincz, 2012; Rowe et al., 2010; Sharif et al., 2016). We show that this mark is reduced in neurons, potentially explaining the requirement of DNA methylation for the survival of somatic cells, as TEs are dramatically de-repressed in DNMT-TKO neurons. We provide evidence that DNA methylation mediated repression of TEs in neurons entails the repulsion of methylation-sensitive TFs such as CREB1 **(Figure 1)**. Removing this factor indeed results in reduced repeat activity. That this reduction is only partially could be explained by functional redundancy between different CRE binding TFs (Steven et al., 2020) and it remains to be tested whether other bZIP TFs such as ATF1 can functionally compensate for the loss of CREB1 at TEs *in vivo*.

DNA methylation mediated repression might be a highly beneficial mode of transcriptional regulation for TEs as this mark would block the binding of ubiquitously expressed activators such as CREB1 throughout the soma. This might prevent transposition events that are non-inheritable, but potentially harmful to the fitness of the host. In contrast, low methylation levels during gametogenesis or in pre-implantation embryos might result in a less stringent repression of TEs at a developmental time where new copies can be passed on to the next generation.

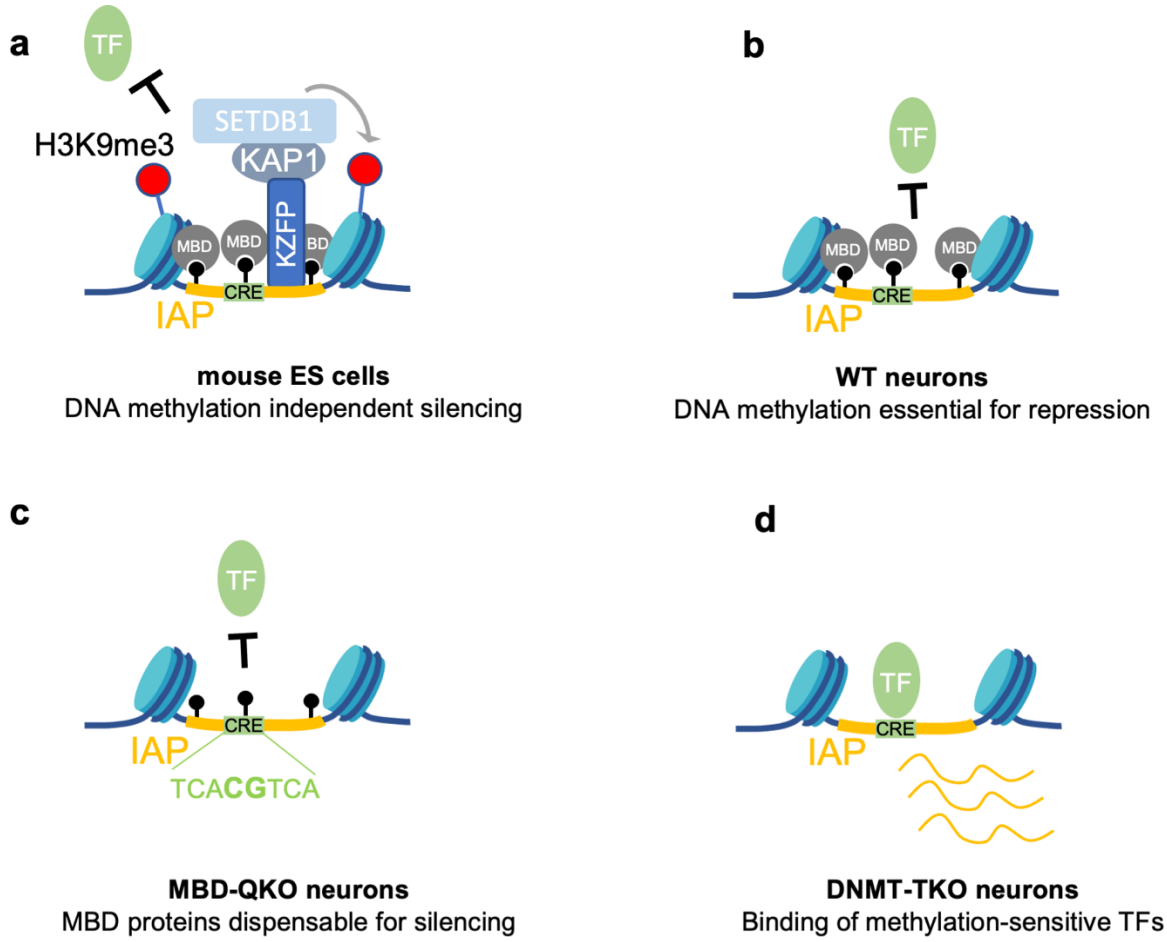


Figure 1 Transcriptional repression of IAP elements by DNA methylation entails the direct repulsion of TFs in somatic cells. a) Evolutionary young TE (i.e., IAP elements) are transcriptionally repressed primarily by H3K9me3, which is deposited by the KZFP/KAP1 complex in mouse ES cells and prevents binding of activating TFs such as CREB1. DNA methylation only plays a limited role in repressing these elements b) In differentiated cells H3K9me3 is depleted and IAP elements are now repressed primarily by DNA methylation, which is bound by MBD-proteins c) However, IAP elements remain silent in absence of all 5mC binding MBD proteins, suggesting a limited role of these factors in DNA methylation mediated repression. Instead, methylation-sensitive TFs such as CREB1 are directly repelled from their cognate motif by CpG methylation. d) Removal of DNA methylation result in binding of CREB1 and de-repression of IAP elements in somatic cells.

Taken together, this work finds limited evidence for a model of indirect repression whereby DNA methylation is translated into silencing of regulatory regions by MBD

proteins **(Figure 1)**. Instead, we provide evidence that the dominant mode of gene and TE repression by DNA methylation is the direct repulsion of TFs, through differential methylation of their binding sites. While *in vitro* evidence suggest that a substantial fraction of TFs is influenced by DNA methylation, here we expand the list of TF that are methylation-sensitive *in vivo*. It is likely that further methylation-sensitive TF will be identified, whose differential binding capabilities contribute to the generation of highly precise gene expression patterns.

6. Material and Methods

Cell culture

HEK293 cells were cultured in Dulbecco's Modified Eagle Medium (Invitrogen), supplemented with 10 % Fetal Calf Serum (Invitrogen) and 2mM L-Glutamine (ThermoScientific). All experiments were performed with cells grown for several passages.

Cell line generation and guide RNAs

Cell lines were generated using CRISPR/Cas9 system (Cong et al. 2013). Briefly, HEK293 cells (carrying a HyTK cassette) were co-transfected (Lipofectamine 3000, Thermo Fisher Scientific) with two vectors encoding a red fluorescence protein (RFP) and a guide RNA either against *MBD2* or *MECP2* together with a third vector encoding Cas9. RFP+ HEK293 cells were sorted into 96 wells and genotyped. Clones carrying a frameshift mutation were expanded and validated by western blot. The same process was repeated to delete *MBD1* and *MBD4*, individually.

target	Guide RNA
MBD1	GCTCAGACACCTATTACCAG
MBD2	CGAAAATCTGGGCTAAGTGC
MBD4	CCAAACTGAGCAGAAGCGAT
MECP2	TAGAGCGAAAGGCTTTTCCC

Antibodies

MBD1 (Abcam, catalog no. ab108510), MBD2 (Abcam, ab188474), MBD4 (Bethyl laboratories, catalog no. A301-634AM) and MeCP2 (Millipore, catalog no. 07-013)

5-Aza treatment of HEK293T cells

HEK293 cells (seeded the day before) were treated with either 1 μM 5-Aza-2′-deoxycytidine (#A3656-10MG, Sigma) or DMSO. The next day media with fresh 5-Aza or DMSO was replaced. After 75 h cells were harvested for RNA isolation.

Assessment of DNA methylation levels with MspI and HpaII digestion

600 ng of DNA isolated with QIAamp DNA Mini Kit (Qiagen, #51306) according to manufacturer's instructions were digested either with 2 μl of MspI (NEB, #R0106S), 2ul of HpaII (NEB, B7204S) or water in CutSmart Buffer (NEB, #N7204S) and 0.07 ul of 10 mg/ml RNase A (Roche, #10109169001) in a total volume of 50 ul for 1h at 37 °C and 20 min at 80 °C, and then loaded on to a 0.8 % agarose gel.

RNA-seq

RNA was isolated from HEK293 cells with the RNeasy mini kit (Qiagen) using on-column DNA digestion. Sequencing libraries were prepared from purified RNA for three biological replicates using TruSeq stranded total RNA Library Prep (Illumina). Libraries were sequenced on an a Illumina NextSeq platform with paired-end reads (2x38bp).

ATAC-seq

ATAC-seq was performed according to a previously described protocol (Buenrostro et al. 2015). Briefly, about 50,000 cells were washed with cold phosphate-buffered saline and resuspended in lysis buffer to extract the nuclei. The nuclei were cold centrifuged at 500*g* for 10 min. The nuclei pellet was incubated with transposition reaction buffer for 30 min at 37 °C. The DNA was purified using the MinElute PCR Purification Kit (QIAGEN). The eluted transposed DNA was submitted to PCR using Q5 High-Fidelity Polymerase (New England Biolabs). DNA was amplified with 5 cycles of PCR. The libraries were sequenced on the Illumina NextSeq platform at 38 bp paired-end. All ATAC-seq experiments were performed in three independent biological replicates per condition.

WGBS-seq

Whole-genome Bisulfite sequencing genomic DNA was isolated using QIAamp DNA mini kit from Qiagen (cat. No 51306) and fragmented using Covaris S220 Focused-ultrasonicator. 500ng of fragmented DNA were then used for library preparation with NEBNext Ultra DNA Library Prep Kit for Illumina (NEB #E7370) and bisulfite treatment with EZ DNA methylation-lightning Kit (Zymo Research #D5046). Final PCR amplification was performed using KAPA HiFi HotStart Uracil+ ReadyMix PCR Kit (KAPA Biosystems #KK2801) and 12 cycles of amplification. Resulting libraries were sequenced on Illumina NextSeq (75 cycles single-end).

Analysis of RNA-seq at genes and repeats

For gene and repeat expression analysis paired-end RNA-seq data was aligned to hg19 genome using STAR (--outFilterMultimapNmax 1) (Dobin et al. 2013). Alignments were counted in genes (TxDb.Hsapiens.UCSC.hg19.knownGene) or in RepeatMasker annotations (hg19 genome, downloaded from the UCSC table browser (http://genome.ucsc.edu/) (Karolchik et al. 2004)) using the qCount function (orientation = "any", useRead = "first") from the QuasR Bioconductor package (version 1.26.0). Counts in RepeatMasker annotations were summed by repName.
Differential expression analysis (adjusted P value and foldchange cutoffs are indicated in the manuscript) was performed using edgeR's normalization combined with voom transformation (default settings) from limma R package (Robinson, McCarthy, and Smyth 2010; C. W. Law et al. 2014; Ritchie et al. 2015). Overrepresentation of gene ontology categories in selected gene sets was analysed using the clusterProfiler (Yu et al. 2012).

Analysis of ATAC-seq

Adapters were first filtered using the cutadapt software (Martin 2011). Both reads were trimmed by one bp at the 3' end to allow for mapping of overlapping reads. Reads were mapped to the hg19 genome using Rbowtie in the QuasR package (v1.26.0). Mitochondrial reads were subsequently removed using samtools. For ATAC–seq data,

peaks of individual replicates were called using macs2 v.2.2.6 (Zhang et al., 2008) with parameters callpeak -t IP.bam -f BAMPE -g mm -q 0.05. Aligned reads were counted in peaks using the qCount function from the QuasR package v.1.30.0. Differential expression analysis (adjusted P value and foldchange cutoffs are indicated in the manuscript) was performed using edgeR's normalization combined with voom transformation (normalize="quantile") from limma R package (Robinson, McCarthy, and Smyth 2010; C. W. Law et al. 2014; Ritchie et al. 2015).

Analysis of WGBS-seq

For DNA methylation analysis WGBS data was aligned to the hg19 genome using the qAlign function (bisulfite="dir") from the QuasR package (v1.26.0). Total and methylated counts for Cs in CpG context (genome-wide) were calculated using the qMeth function from QuasR. To evaluate methylation levels of individual CpGs between WT and MBD-QKO HEK293 cells, only C were considered with minimum 10 reads in both WT and MBD-QKO HEK293 cells. To determine methylation levels in different genomic features, all methylated counts were summed and divided by the total number of counts for a given feature. Genomic features with less than 10 total reads were not considered.

7. References

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat. Commun. *9*, 1911.

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.

Ambrosi, C., Manzo, M., and Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. J. Mol. Biol. *429*, 1459–1475.

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.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761–764.

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

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.

Auclair, G., and Weber, M. (2012). Mechanisms of DNA methylation and demethylation in mammals. Biochimie *94*, 2202–2211.

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.

Ballester, B., Medina-Rivera, A., Schmidt, D., Gonzàlez-Porta, M., Carlucci, M., Chen, X., Chessman, K., Faure, A.J., Funnell, A.P.W., Goncalves, A., et al. (2014). Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. Elife *3*, e02626.

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

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA *6*, 1–6.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909–912.

Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. Nature *569*, 136–140.

Bartlett, A., O'Malley, R.C., Huang, S.-S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nature Protocols *12*, 1659–1672.

Baubec, T., and Schübeler, D. (2014). Genomic patterns and context specific interpretation of DNA methylation. Curr. Opin. Genet. Dev. *25*, 85–92.

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.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. *32*, 29–41.

Bednarik, D.P., Duckett, C., Kim, S.U., Perez, V.L., Griffis, K., Guenthner, P.C., and Folks, T.M. (1991). DNA CpG methylation inhibits binding of NF-kappa B proteins to the HIV-1 long terminal repeat cognate DNA motifs. New Biol. *3*, 969–976.

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

Bell, O., Tiwari, V.K., Thomä, N.H., and Schübeler, D. (2011). Determinants and dynamics of genome accessibility. Nat. Rev. Genet. *12*, 554–564.

Benbrook, D.M., and Jones, N.C. (1994). Different binding specificities and transactivation of variant CRE's by CREB complexes. Nucleic Acids Res. *22*, 1463–1469.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

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

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

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. Dev. Cell *21*, 611–626.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

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.

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

Breiling, A., Turner, B.M., Bianchi, M.E., and Orlando, V. (2001). General transcription factors bind promoters repressed by Polycomb group proteins. Nature *412*, 651–655.

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Research *22*, 1128–1138.

Brûlet, P., Kaghad, M., Xu, Y.S., Croissant, O., and Jacob, F. (1983). Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse. Proc. Natl. Acad. Sci. U. S. A. *80*, 5641–5645.

Buck-Koehntop, B.A., and Defossez, P.-A. (2013). On how mammalian transcription factors recognize methylated DNA. Epigenetics *8*, 131–137.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1-21.29.9.

Busslinger, M., Hurst, J., and Flavell, R.A. (1983). DNA methylation and the regulation of globin gene expression. Cell *34*, 197–206.

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.

Caselli, E., Benedetti, S., Grigolato, J., Caruso, A., and Di Luca, D. (2012). Activating transcription factor 4 (ATF4) is upregulated by human herpesvirus 8 infection, increases virus replication and promotes proangiogenic properties. Arch. Virol. *157*, 63–74.

Chelmicki, T., Roger, E., Teissandier, A., Dura, M., Bonneville, L., Rucli, S., Dossin, F., Fouassier, C., Lameiras, S., and Bourc'his, D. (2021). m6A RNA methylation regulates the fate of endogenous retroviruses. Nature.

Chen, R.Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat. Genet. *27*, 327–331.

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

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y., and Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat. Genet. *39*, 391–396.

Cheng, T.-L., Wang, Z., Liao, Q., Zhu, Y., Zhou, W.-H., Xu, W., and Qiu, Z. (2014). MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. Dev. Cell *28*, 547–560.

Cholewa-Waclaw, J., Shah, R., Webb, S., Chhatbar, K., Ramsahoye, B., Pusch, O., Yu, M., Greulich, P., Waclaw, B., and Bird, A.P. (2019). Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. Proc. Natl. Acad. Sci. U. S. A. *116*, 14995–15000.

Christman, J.K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene *21*, 5483–5495.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e20.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. *18*, 407–422.

Cohen, N.M., Kenigsberg, E., and Tanay, A. (2011). Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell *145*, 773–786.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Dahlet, T., Argüeso Lleida, A., Al Adhami, H., Dumas, M., Bender, A., Ngondo, R.P., Tanguy, M., Vallet, J., Auclair, G., Bardet, A.F., et al. (2020). Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. Nat. Commun. *11*, 3153.

Dantas Machado, A.C., Zhou, T., Rao, S., Goel, P., Rastogi, C., Lazarovici, A., Bussemaker, H.J., and Rohs, R. (2014). Evolving insights on how cytosine methylation affects protein–DNA binding. Brief. Funct. Genomics *14*, 61–73.

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

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C.R., and Pirrotta, V. (2004). Polycomb Silencing Blocks Transcription Initiation. Molecular Cell *13*, 887–893.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

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.

Dupressoir, A., and Heidmann, T. (1996). Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice. Mol. Cell. Biol. *16*, 4495–4503.

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719–2729.

Egger, G., Jeong, S., Escobar, S.G., Cortez, C.C., Li, T.W.H., Saito, Y., Yoo, C.B., Jones, P.A., and Liang, G. (2006). Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. Proc. Natl. Acad. Sci. U. S. A. *103*, 14080–14085.

Emerson, R.O., and Thomas, J.H. (2009). Adaptive evolution in zinc finger transcription factors. PLoS Genet. *5*, e1000325.

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. *21*, 788–797.

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., and Levine, M.S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. U. S. A. *113*, 6508–6513.

Fatemi, M., and Wade, P.A. (2006). MBD family proteins: reading the epigenetic code. J. Cell Sci. *119*, 3033–3037.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell *2*, 160–169.

Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annu. Rev. Cell Dev. Biol. *31*, 429–451.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003a). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J. Biol. Chem. *278*, 24132–24138.

Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., Saya, H., and Nakao, M. (2003b). MCAF mediates MBD1-dependent transcriptional repression. Mol. Cell. Biol. *23*, 2834–2843.

Fulton, D.L., Sundararajan, S., Badis, G., Hughes, T.R., Wasserman, W.W., Roach, J.C., and Sladek, R. (2009). TFCat: the curated catalog of mouse and human transcription factors. Genome Biol. *10*, R29.

Gabel, H.W., Kinde, B., Stroud, H., Gilbert, C.S., Harmin, D.A., Kastan, N.R., Hemberg, M., Ebert, D.H., and Greenberg, M.E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature *522*, 89–93.

Gaidatzis, D., Burger, L., Florescu, M., and Stadler, M.B. (2015). Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat. Biotechnol. *33*, 722–729.

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

Gasser, S.M. (2016). Selfish DNA and Epigenetic Repression Revisited. Genetics *204*, 837–839.

Gilbert, W., and Maxam, A. (1973). The nucleotide sequence of the lac operator. Proc. Natl. Acad. Sci. U. S. A. *70*, 3581–3584.

Göke, J., Lu, X., Chan, Y.-S., Ng, H.-H., Ly, L.-H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell *16*, 135–141.

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

Grant, C., Oh, U., Fugo, K., Takenouchi, N., Griffith, C., Yao, K., Newhook, T.E., Ratner, L., and Jacobson, S. (2006). Foxp3 Represses Retroviral Transcription by Targeting Both NF-κB and CREB Pathways. PLoS Pathog. *2*, e33.

Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. *20*, 590–607.

Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221–225.

Guérillon, C., Larrieu, D., and Pedeux, R. (2013). ING1 and ING2: multifaceted tumor suppressor genes. Cell. Mol. Life Sci. *70*, 3753–3772.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat. Genet. *27*, 322–326.

Hai, T., and Hartman, M.G. (2001). The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis. Gene *273*, 1–11.

Héberlé, É., and Bardet, A.F. (2019). Sensitivity of transcription factors to DNA methylation. Essays Biochem. *63*, 727–741.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

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

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

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301–304.

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.

Hidalgo, O., Pellicer, J., Christenhusz, M., Schneider, H., Leitch, A.R., and Leitch, I.J. (2017). Is There an Upper Limit to Genome Size? Trends Plant Sci. *22*, 567–573.

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.

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*, e00726.

Hutnick, L.K., Golshani, P., Namihira, M., Xue, Z., Matynia, A., Yang, X.W., Silva, A.J., Schweizer, F.E., and Fan, G. (2009). DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. Hum. Mol. Genet. *18*, 2875–2888.

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.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett. *583*, 1713–1720.

Imbeault, M., Helleboid, P.-Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature *543*, 550–554.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. *28*, 2679–2692.

Iyaguchi, D., Yao, M., Watanabe, N., Nishihira, J., and Tanaka, I. (2007). DNA recognition mechanism of the ONECUT homeodomain of transcription factor HNF-6. Structure *15*, 75–83.

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.

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

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

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.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., et al. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. *20*, 861–873.

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.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384–388.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jørgensen, H.F., Ben-Porath, I., and Bird, A.P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol. Cell. Biol. *24*, 3387–3395.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell *8*, 676–687.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493-6.

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

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018:

update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46*, D1284.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kimura, H., and Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J. Biol. Chem. *278*, 4806–4812.

Kirby, H., Rickinson, A., and Bell, A. (2000). The activity of the Epstein--Barr virus BamHI W promoter in B cells is dependent on the binding of CREB/ATF factors. Microbiology *81*, 1057–1066.

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

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

Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C.M., Brown, D.L., Brudno, M., Campbell, J., et al. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. *42*, D966-74.

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Mol. Cell. Biol. *25*, 4388–4396.

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

Kribelbauer, J.F., Laptenko, O., Chen, S., Martini, G.D., Freed-Pastor, W.A., Prives, C., Mann, R.S., and Bussemaker, H.J. (2017). Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. Cell Rep. *19*, 2383–2395.

Kribelbauer, J.F., Rastogi, C., Bussemaker, H.J., and Mann, R.S. (2019). Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. Annu. Rev. Cell Dev. Biol. *35*, 357–379.

Kribelbauer, J.F., Lu, X.-J., Rohs, R., Mann, R.S., and Bussemaker, H.J. (2020). Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. J. Mol. Biol. *432*, 1801–1815.

Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., Roberts, R.J., and Defossez, P.-A. (2010). The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PLoS One *5*, e11982.

Lagger, S., Connelly, J.C., Schweikert, G., Webb, S., Selfridge, J., Ramsahoye, B.H., Yu, M., He, C., Sanguinetti, G., Sowers, L.C., et al. (2017). MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. PLoS Genet. *13*, e1006793.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665.

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

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

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

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A.M., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol. Cell. Biol. *26*, 843–851.

Lee, H.J., Hore, T.A., and Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell *14*, 710–719.

Leung, D.C., and Lorincz, M.C. (2012). Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem. Sci. *37*, 127–133.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914.

Li, G., and Widom, J. (2004). Nucleosomes facilitate their own invasion. Nat. Struct. Mol. Biol. *11*, 763–769.

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

Li, E., Beard, C., Forster, A.C., Bestor, T.H., and Jaenisch, R. (1993). DNA methylation, genomic imprinting, and mammalian development. Cold Spring Harb. Symp. Quant. Biol. *58*, 297–305.

Li, E., Beard, C., and Jaenisch, R. (1994). Role for DNA methylation in genomic imprinting. Trends in Genetics *10*, 78.

Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. Nat. Struct. Mol. Biol. *12*, 46–53.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev. Cell *15*, 547–557.

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. (2011). Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. *43*, 1091–1097.

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev. *28*, 2041–2055.

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

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell *49*, 203–210.

Loyola, A., LeRoy, G., Wang, Y.H., and Reinberg, D. (2001). Reconstitution of recombinant chromatin establishes a requirement for histone-tail modifications during chromatin assembly and transcription. Genes Dev. *15*, 2837–2851.

Lynch, M.D., Smith, A.J.H., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J. *31*, 317–329.

Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N.R., Robinson, N.D., de Lima Alves, F., et al. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat. Neurosci. *16*, 898–902.

Mager, D.L., and Stoye, J.P. (2015). Mammalian Endogenous Retroviruses. Microbiol Spectr *3*, MDNA3-0009–2014.

Makowski, M.M., Gaullier, G., and Luger, K. (2020). Picking a nucleosome lock: Sequence- and structure-specific recognition of the nucleosome. J. Biosci. *45*.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet. *2*, e2.

Mancini, D.N., Singh, S.M., Archer, T.K., and Rodenhiser, D.I. (1999). Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. Oncogene *18*, 4108–4119.

Mann, I.K., Chatterjee, R., Zhao, J., He, X., Weirauch, M.T., Hughes, T.R., and Vinson, C. (2013). CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. Genome Res. *23*, 988–997.

Marchi, E., Kanapin, A., Magiorkinis, G., and Belshaw, R. (2014). Unfixed endogenous retroviral insertions in the human population. J. Virol. *88*, 9529–9537.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Martín Caballero, I., Hansen, J., Leaford, D., Pollard, S., and Hendrich, B.D. (2009). The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. PLoS One *4*, e4315.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927–931.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253–257.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. *23*, 1256–1269.

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.

de Mendoza, A., Hatleberg, W.L., Pang, K., Leininger, S., Bogdanovic, O., Pflueger, J., Buckberry, S., Technau, U., Hejnol, A., Adamska, M., et al. (2019). Convergent evolution of a vertebrate-like methylome in a marine sponge. Nat Ecol Evol *3*, 1464–1473.

de Mendoza, A., Lister, R., and Bogdanovic, O. (2020). Evolution of DNA Methylome Diversity in Eukaryotes. J. Mol. Biol. *432*, 1687–1705.

Messerschmidt, D.M., Knowles, B.B., and Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. *28*, 812–828.

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

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

Millhouse, S., Kenny, J.J., Quinn, P.G., Lee, V., and Wigdahl, B. (1998). ATF/CREB elements in the herpes simplex virus type 1 latency-associated transcript promoter interact with members of the ATF/CREB and AP-1 transcription factor families. Journal of Biomedical Science *5*, 451–464.

Miotto, B., Marchal, C., Adelmant, G., Guinot, N., Xie, P., Marto, J.A., Zhang, L., and Defossez, P.-A. (2018). Stabilization of the methyl-CpG binding protein ZBTB38 by the deubiquitinase USP9X limits the occurrence and toxicity of oxidative stress in human cells. Nucleic Acids Res. *46*, 4392–4404.

Mivelaz, M., Cao, A.-M., Kubik, S., Zencir, S., Hovius, R., Boichenko, I., Stachowicz, A.M., Kurat, C.F., Shore, D., and Fierz, B. (2020). Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. Mol. Cell *77*, 488-500.e9.

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

Molaro, A., and Malik, H.S. (2016). Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline. Curr. Opin. Genet. Dev. *37*, 51–58.

Montminy, M.R., and Bilezikjian, L.M. (1987). Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. Nature *328*, 175–178.

Mujtaba, S., Zeng, L., and Zhou, M.-M. (2007). Structure and acetyl-lysine recognition of the bromodomain. Oncogene *26*, 5521–5527.

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

Ng, H.H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. Mol. Cell. Biol. *20*, 1394–1406.

Nicetto, D., and Zaret, K.S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr. Opin. Genet. Dev. *55*, 1–10.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

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.

O'Neill, K.M., Irwin, R.E., Mackin, S.-J., Thursby, S.-J., Thakur, A., Bertens, C., Masala, L., Loughery, J.E.P., McArt, D.G., and Walsh, C.P. (2018). Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression. Epigenetics Chromatin *11*, 12.

Padeken, J., Zeller, P., and Gasser, S.M. (2015). Repeat DNA in genome organization and stability. Curr. Opin. Genet. Dev. *31*, 12–19.

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

Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell *65*, 395–407.

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. *15*, 1613–1618.

Ptashne, M. (1967). ISOLATION OF THE lambda PHAGE REPRESSOR. Proc. Natl. Acad. Sci. U. S. A. *57*, 306–313.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol. Cell *44*, 361–372.

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., and Trono, D. (2012). The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development. Cell Reports *2*, 766–773.

Ramesh, V., Bayam, E., Cernilogar, F.M., Bonapace, I.M., Schulze, M., Riemenschneider, M.J., Schotta, G., and Götz, M. (2016). Loss of Uhrf1 in neural stem cells leads to activation of retroviral elements and delayed neurodegeneration. Genes Dev. *30*, 2199–2212.

Ramos, M.-P., Wijetunga, N.A., McLellan, A.S., Suzuki, M., and Greally, J.M. (2015). DNA demethylation by 5-aza-2′-deoxycytidine is imprinted, targeted to euchromatin, and has limited transcriptional consequences. Epigenetics Chromatin *8*, 1–18.

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.

Rausch, C., Hastert, F.D., and Cardoso, M.C. (2019). DNA Modification Readers and Writers and Their Interplay. J. Mol. Biol.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

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

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

Roussel-Gervais, A., Naciri, I., Kirsh, O., and Kasprzyk, L. (2017). Loss of the methyl-CpG–binding protein ZBTB4 alters mitotic checkpoint, increases aneuploidy, and promotes tumorigenesis. Cancer Res.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.

Sabari, B.R., Dall'Agnese, A., and Young, R.A. (2020). Biomolecular Condensates in the Nucleus. Trends Biochem. Sci. *45*, 961–977.

Saito, M., and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. J. Biol. Chem. *277*, 35434–35439.

Sasai, N., Nakao, M., and Defossez, P.-A. (2010). Sequence-specific recognition of methylated DNA by human zinc-finger proteins. Nucleic Acids Res. *38*, 5015–5022.

Schübeler, D. (2015). Function and information content of DNA methylation. Nature *517*, 321–326.

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

Schultz, D.C., Friedman, J.R., and Rauscher, F.J., 3rd (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. Genes Dev. *15*, 428–443.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. *16*, 919–932.

Schumacher, M.A., Goodman, R.H., and Brennan, R.G. (2000). The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. J. Biol. Chem. *275*, 35242–35247.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol. Cell *48*, 849–862.

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.

Shaknovich, R., Cerchietti, L., Tsikitas, L., Kormaksson, M., De, S., Figueroa, M.E., Ballon, G., Yang, S.N., Weinhold, N., Reimers, M., et al. (2011). DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. Blood *118*, 3559–3569.

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

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 *19*, 81–94.

Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol. Cell *37*, 457–468.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordân, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends Biochem. Sci. *39*, 381–399.

Smallwood, S.A., Tomizawa, S.-I., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat. Genet. *43*, 811–814.

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

Smith, B., Fang, H., Pan, Y., Walker, P.R., Famili, A.F., and Sikorska, M. (2007). Evolution of motif variants and positional bias of the cyclic-AMP response element. BMC Evol. Biol. *7 Suppl 1*, S15.

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.

Steven, A., Friedrich, M., Jank, P., Heimer, N., Budczies, J., Denkert, C., and Seliger, B. (2020). What turns CREB on? And off? And why does it matter? Cell. Mol. Life Sci. *77*, 4049–4067.

Stirzaker, C., Song, J.Z., Ng, W., Du, Q., Armstrong, N.J., Locke, W.J., Statham, A.L., French, H., Pidsley, R., Valdes-Mora, F., et al. (2017). Methyl-CpG-binding protein MBD2 plays a key role in maintenance and spread of DNA methylation at CpG islands and shores in cancer. Oncogene *36*, 1328–1338.

Storer, J., Hubley, R., Rosen, J., Wheeler, T.J., and Smit, A.F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob. DNA *12*, 2.

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

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.

Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. *17*, 183–193.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc. Natl. Acad. Sci. U. S. A. *104*, 5521–5526.

Tang, W.W.C., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., and Surani, M.A. (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell *161*, 1453–1467.

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

Thoma, E.C., Wischmeyer, E., Offen, N., Maurus, K., Sirén, A.-L., Schartl, M., and Wagner, T.U. (2012). Ectopic expression of neurogenin 2 alone is sufficient to induce differentiation of embryonic stem cells into mature neurons. PLoS One *7*, e38651.

Tierney, R.J., Kirby, H.E., Nagra, J.K., Desmond, J., Bell, A.I., and Rickinson, A.B. (2000). Methylation of transcription factor binding sites in the Epstein-Barr virus latent cycle promoter Wp coincides with promoter down-regulation during virus-induced B-cell transformation. J. Virol. *74*, 10468–10479.

Tillotson, R., and Bird, A. (2019). The Molecular Basis of MeCP2 Function in the Brain. J. Mol. Biol.

Tillotson, R., Selfridge, J., Koerner, M.V., Gadalla, K.K.E., Guy, J., De Sousa, D., Hector, R.D., Cobb, S.R., and Bird, A. (2017). Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. Nature *550*, 398–401.

Tillotson, R., Cholewa-Waclaw, J., Chhatbar, K., Connelly, J.C., Kirschner, S.A., Webb, S., Koerner, M.V., Selfridge, J., Kelly, D.A., De Sousa, D., et al. (2021). Neuronal non-CG methylation is an essential target for MeCP2 function. Mol. Cell.

Tippin, D.B., and Sundaralingam, M. (1997). Nine polymorphic crystal structures of d (CCGGGCCCGG), d (CCGGGCCm5CGG), d (Cm5CGGGCCm5CGG) and d (CCGGGCC (Br) 5CGG) in three different …. J. Mol. Biol.

Tiwari, V.K., Stadler, M.B., Wirbelauer, C., Paro, R., Schübeler, D., and Beisel, C. (2011). A chromatin-modifying function of JNK during stem cell differentiation. Nat. Genet. *44*, 94–100.

Toufaily, C., Lokossou, A.G., Vargas, A., Rassart, É., and Barbeau, B. (2015). A CRE/AP-1-like motif is essential for induced syncytin-2 expression and fusion in human trophoblast-like model. PLoS One *10*, e0121468.

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 *11*, 805–814.

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.

Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. Mol. Cell *47*, 827–838.

Waalwijk, C., and Flavell, R.A. (1978). MspI, an isoschizomer of hpaII which cleaves both unmethylated and methylated hpaII sites. Nucleic Acids Res. *5*, 3231–3236.

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

Wang, L., Chen, J., Wang, C., Uusküla-Reimand, L., Chen, K., Medina-Rivera, A., Young, E.J., Zimmermann, M.T., Yan, H., Sun, Z., et al. (2014). MACE: model based analysis of ChIP-exo. Nucleic Acids Res. *42*, e156.

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.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.

Wildschutte, J.H., Williams, Z.H., Montesion, M., Subramanian, R.P., Kidd, J.M., and Coffin, J.M. (2016). Discovery of unfixed endogenous retrovirus insertions in diverse human populations. Proc. Natl. Acad. Sci. U. S. A. *113*, E2326-34.

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M.C., Lipkin, M., et al. (2002). Mbd4 inactivation increases C->T transition mutations and promotes gastrointestinal tumor formation. Proceedings of the National Academy of Sciences *99*, 14937–14942.

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

Yamamoto, K.K., Gonzalez, G.A., Menzel, P., Rivier, J., and Montminy, M.R. (1990). Characterization of a bipartite activator domain in transcription factor CREB. Cell *60*, 611–617.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. *13*, 335–340.

Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S., et al. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proc. Natl. Acad. Sci. U. S. A. *102*, 17551–17558.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

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

Zamudio, N., and Bourc'his, D. (2010). Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity *105*, 92–104.

Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Curr. Biol. *20*, R780-5.

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

Żemojtel, T., Kiełbasa, S.M., Arndt, P.F., Behrens, S., Bourque, G., and Vingron, M. (2011). CpG Deamination Creates Transcription Factor–Binding Sites with High Efficiency. Genome Biol. Evol. *3*, 1304–1311.

Zhang, X., Odom, D.T., Koo, S.-H., Conkright, M.D., Canettieri, G., Best, J., Chen, H., Jenner, R., Herbolsheimer, E., Jacobsen, E., et al. (2005). Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. Proc. Natl. Acad. Sci. U. S. A. *102*, 4459–4464.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924–1935.

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

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron *78*, 785–798.

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.

Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., and Xu, R.-M. (2014). Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2. Molecular Cell *54*, 879–886.

Zhou, W., Liang, G., Molloy, P.L., and Jones, P.A. (2020). DNA methylation enables transposable element-driven genome expansion. Proc. Natl. Acad. Sci. U. S. A. *117*, 19359–19366.