The regulation of Tet proteins (5-methylcytosine modifiers) by Mbd proteins (5-methylcytosine readers)

Peng Zhang
The regulation of Ten-eleven translocation proteins (methylcytosine modifiers) by methyl-CpG binding domain proteins (methylcytosine readers)
# Index

**Summary/Zusammenfassung** ....................................................................................................................................... IV

**General introduction** ........................................................................................................................................................ 1

- DNA modifications and modifiers ........................................................................................................................................ 3
- DNA modification readers .............................................................................................................................................. 8
- Role of 5mC writers, readers and modifiers in disease ................................................................................................ 15
- Concluding remarks ..................................................................................................................................................... 18

**Aim of this work** ............................................................................................................................................................. 27

**Chapters**

- Chapter 1: Poly(ADP-ribosyl)ation of methyl CpG binding protein 2 regulates chromatin structure ......................... 28
- Conclusion and outlook ................................................................................................................................................. 39

**Annex** .............................................................................................................................................................................. 41

- Abbreviations ............................................................................................................................................................... 41
- List of contributions ..................................................................................................................................................... 42
- Acknowledgements ...................................................................................................................................................... 43
- Declaration - Ehrenwörtliche Erklärung ...................................................................................................................... 44
- Curriculum vitae .......................................................................................................................................................... 45
- List of publications ...................................................................................................................................................... 46
List of figures

General introduction: Modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in disease .................................................................................................................................. 1

Figure 1 Schematic representation of the Dnmt protein family. .................................................................................................................. 3
Figure 2 DNA base modifications with respective enzymes. .................................................................................................................. 5
Figure 3 Schematic representation of the Tet protein family. .................................................................................................................. 6
Figure 4 5-(Hydroxy)methylcytosine readers. .................................................................................................................................. 8
Figure 5 Schematic representation of the MBD protein family. .......................................................................................................... 9
Figure 6 Schematic representation of the Kaiso-like protein family. .......................................................................................... 13
Figure 7 Schematic representation of the SRA domain protein family ...................................................................................... 14
Figure 8 Writing, reading and translating DNA modifications. .......................................................................................... 18

Chapter 1: Poly(ADP-ribosyl)ation of methyl CpG binding protein 2 regulates chromatin structure ................. 28

Figure 1 Endogenous Mecp2 is poly(ADP-ribosyl)ated in mouse brain. .................................................................................. 31
Figure 2 Poly(ADP-ribosyl)ation counteracts clustering of pericentric heterochromatin. .................................................. 32
Figure 3 The chromatin binding ability of Mecp2 is elevated in Parp-1−/− cells. ........................................................................ 33
Figure 4 Summary of factors influencing Mecp2 mediated heterochromatin binding and clustering. ..................... 34

Conclusion and outlook .......................................................................................................................................................... 39

Figure 1 Summary of how Tet oxidize 5mC and how this process is regulated by Mbd proteins. ......................... 39

List of tables

General introduction: Modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in Disease .................................................................................................................................. 1

Table 1. Phenotypes of initial Dnmt knockout mouse models. ......................................................................................... 4
Table 2. Phenotypes of initial Tet knockout mouse models .......................................................................................... 6
Table 3: Phenotype of initial MBP deficient mouse models .......................................................................................... 11
Table 4: Comparison of human proteins and their mouse orthologues ........................................................................ 15
Table 5: Summary of disease related DNMT and TET mutations. .................................................................................. 16
Table 6: Summary of disease related MBP alterations .................................................................................................. 17
Summary

Cytosine modifications diversify the genome and allow cell differentiation by the action of cytosine modification readers and modifiers. The epigenetic information of 5-methylcytosine can be translated by cytosine modification readers, such as the methyl-CpG binding domain (MBD) proteins. The aberrant interactions of MBD proteins with 5-methylcytosine cause diseases like Rett syndrome and also decrease genome stability, thus, the levels of both MBD protein and its substrate 5mC must be precisely regulated. Although, 5mC can be modified by Ten-eleven translocation (Tet) protein to 5-hydroxymethylcytosine (5hmC), which affects the binding ability of MBD proteins to DNA, the interplay of MBD proteins, Tet1 proteins and their substrate is still unknown.

Since post translational modifications of Mecp2, the founding member of MBD protein family have been described before, we initially focused on the effect of poly(ADP-ribosyl)ation of Mecp2 on chromatin structure and its DNA binding ability. We show that in mouse brain endogenous Mecp2 is poly(ADP-ribosyl)ated in vivo. Furthermore, we find that poly(ADP-ribosyl)ation of Mecp2 decreases its ability to cluster pericentric heterochromatin. Finally, we demonstrate that poly(ADP-ribosyl)ated Mecp2 decrease binding ability to heterochromatin DNA.

To understand the regulation of Tet mediated 5mC oxidation, we focused on how Tet proteins convert 5mC to 5hmC. We developed and optimized methods to step by step detect processes involved in Tet oxidation, including Tet-DNA binding, 5mC flipping and 5mC oxidation. We show that the catalytic domain of Tet1 (Tet1CD) binds to DNA in a non-sequence specific manner. Furthermore, we were able to detect DNA base flipping induced by Tet1CD. Finally, our methods can be used to easily and sensitively detect Tet oxidation products.

By using these methods, we next tested whether MBD proteins affect Tet mediated 5mC oxidation. We focused on the five best studied MBD proteins including Mbd1, Mbd2, Mbd3, Mbd4 and Mecp2. We show that Mbd1 enhances Tet1 mediated 5hmC formation by facilitating its localization to methylated DNA. Moreover, the CXXC3 domain of Mbd1 is necessary for this enhancement. Compared with Mbd1, we find that Mbd3 and Mbd4 do not affect Tet1 mediated 5mC oxidation. In contrast to Mbd1, we show that Mbd2 and Mecp2 as well as its subdomains MBD and IDTRD block Tet mediated 5hmC formation in a concentration dependent manner in vivo and in vitro. Moreover, binding of Mecp2 to DNA impairs the DNA binding ability of Tet1CD in vitro and thus, direct binding of Mecp2 to DNA is sufficient to effectively prevent Tet1 mediated 5mC oxidation. These results indicate that the binding ability of MBD proteins and Tet proteins to DNA is important for 5mC conservation and conversion, respectively.

Finally, we focused on the biological consequences of MBD proteins and Tet proteins mediated 5mC conservation and conversion. In mouse cells, we find that the Tet oxidation product 5hmC is enriched in neurons of mouse model for Rett syndrome (Mecp2 knockout mice). Moreover, we find that Tet1 reactivates expression of major satellite repeats in the absence of Mecp2. In human cells, we show that Tet1 activates endogenous and ectopic long interspersed nuclear elements 1 expression and transposition and this activation can be repressed by Mbd2 and Mecp2 as well as its subdomains MBD and IDTRD. These results indicate that the fine balance between methylcytosine readers” and “erasers/writers” regulates transcriptional noise and genome stability.
Zusammenfassung

Modifikationen der DNA-Base Cytosin haben eine Erweiterung des genetischen Alphabets zur Folge und regulieren mit Hilfe von Proteinen, die Cytosin und dessen Modifikationen binden beziehungsweise weiter modifizieren können, die Differenzierung der Zelle. Die epigenetische Information von 5-methylcytosin (5mC) wird von Cytosinmodifikations-Lesern, wie beispielsweise der Methyl-CpG-Bindedomäne (MBD) Proteinfamilie verarbeitet. Gestörte Interaktionen von MBD Proteinen mit 5mC haben Krankheiten, wie das Rett Syndrom zur Folge und beeinträchtigen die genomische Stabilität. Folglich muss die Menge an MBD Proteinen, sowie deren Substrat 5mC präzise reguliert werden. Obwohl bekannt ist, dass die Oxidation von 5mC zu 5-hydroxymethylcytosin (5hmC) durch Ten-eleven translocation (Tet) Proteine die DNA-Bindung der MBD Proteinen beeinflusst, weiß man erst wenig über die Interaktion von MBD Proteinen, Tet Proteinen, sowie deren Substrat 5mC.

Da Mecp2, das zuerst ausführlich erforschte Mitglied der MBD Proteinfamilie, posttranslational modifiziert wird, haben wir zunächst den Effekt der poly(ADP-ribosyl)ierung von Mecp2 auf die Chromatin Struktur, sowie auf die Bindung von DNA untersucht. Wir zeigen, dass endogenes Mecp2 im Maushirn poly(ADP-ribosyl)iert wird. Des Weiteren beobachten wir, dass poly(ADP-ribosyl)ierung von Mecp2 dessen Fähigkeit zur Kondensierung von perizentrischem Heterochromatin und die Bindung an heterochromatische DNA reduziert.

Um die Regulation der von Tet katalysierten 5mC Oxidation besser zu verstehen, untersuchten wir wie Tet Proteine 5mC modifizieren. Hierfür entwickelten beziehungsweise optimierten wir Methoden um die bei der Tet katalysierten 5mC Oxidation beteiligten Schritte (Bindung von Tet an DNA, 5mC-flipping und 5mC Oxidation) zu detektieren. Wir zeigen, dass die katalytische Domäne von Tet1 (Tet1CD) Sequenz-unabhängig DNA bindet. Des Weiteren zeigen wir, dass Tet1CD einen Base-flipping Mechanismus benutzt, um Zugang zu 5mC zu erhalten. Letztlich ermöglichen die von uns entwickelten Methoden die einfache und sensitive Detektion von Tet Oxidationsprodukten.

Durch Anwendung der oben beschriebenen Methoden untersuchten wir weiterhin, ob MBD Proteine die von Tet katalysierte 5mC Oxidation beeinträchtigen. Hierfür legten wir den Fokus auf die fünf am besten beschriebenen MBD Proteine Mbd1, Mbd2, Mbd3, Mbd4 und Mecp2. Wir zeigen, dass Mbd1 die Ausbildung von 5hmC begünstigt, indem es die Binding von Tet1 an methylierte DNA fördert, wofür die CXXC3 Domäne von Mbd1 unbedingt notwendig ist. Mbd3 und Mbd4 haben im Vergleich zu Mbd1 keine Auswirkungen auf die von Tet katalysierte 5mC Oxidation. Mbd2 und Mecp2, sowie die beiden Mecp2 Subdomänen MBD und IDTRD hingegen, verhindern das Entstehen von 5hmC in Abhängigkeit ihrer Konzentration in vitro als auch in vivo. Darüber hinaus beeinträchtigt die Bindung von Mecp2 an DNA die Interaktion von Tet1CD mit DNA in vitro. Folglich ist die Bindung von Mecp2 an DNA ausreichend, um die von Tet1 katalysierte 5mC Oxidation zu unterbinden. Dies zeigt, dass die Bindung von MBD und Tet Proteinen sowohl für den Schutz, als auch für die Umsetzung von 5mC entscheidend ist.

Zuletzt konzentrierten wir uns auf die biologischen Konsequenzen der MBD- und Tet Protein vermittelten 5mC Konservierung und Konvertierung. Wir zeigen erhöhte Level des Tet Oxidationsprodukts 5hmC in Neuronen eines Maus-Modells für Rett Syndrom (Mecp2 knockout Mäuse). Des Weiteren beobachten wir, dass Tet1 in der Abwesenheit von Mecp2 die Expression von Major Satellite DNA in Mauszellen reaktiviert. Für humane Zellen zeigen wir, dass Tet1 die Expression und Retrotransposition von endogenen und exogenen L1 Elementen (long interspersed nuclear elements 1) begünstigt. Dies kann wiederum durch Mbd2 und Mecp2, sowie den beiden Mecp2 Subdomänen MBD und IDTRD unterdrückt werden. Dies deutet darauf hin, dass ein feines Gleichgewicht zwischen Proteinen, die 5mC erstellen, Proteinen die 5mC binden und Proteinen die 5mC entfernen, Transkriptionslevel reguliert und darüber hinaus die genomische Stabilität gewährleistet.
Modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in disease

Anne K. Ludwig*, Peng Zhang*, M. Cristina Cardoso

Cell Biology and Epigenetics, Department of Biology, Technische Universität Darmstadt, Germany

* Equal contribution


Author contributions:
P.Z. wrote the text and prepared figures (Figure 1, Figure 2, Figure 3) and tables (Table 1, Table 2, Table 5) for cytosine modifiers (section 1.1), DNA base modifications (section 1.2) and methylcytosine modifiers (section 1.3), and corresponding disease (section 3.1 and 3.2). P.Z. prepared Table 4 together with A.K.L.
Abstract

Cytosine base modifications in mammals underwent a recent expansion with the addition of several naturally occurring further modifications of methylcytosine in the last years. This expansion was accompanied by the identification of the respective enzymes and proteins reading and translating the different modifications into chromatin higher order organization as well as genome activity and stability, leading to the hypothesis of a cytosine code. Here, we summarize the current state-of-the-art on DNA modifications, the enzyme families setting the cytosine modifications and the protein families reading and translating the different modifications with emphasis on the mouse protein homologs. Throughout this review, we focus on functional and mechanistic studies performed on mammalian cells, corresponding mouse models and associated human diseases.

Keywords
cytosine modifications; Dnmt; epigenetics; hydroxymethylcytosine; MBD; methylcytosine; mouse models; Tet
1 DNA modifications and modifiers

1.1 Cytosine modifiers: Dnmts

In mammals, the modified cytosine was initially described in 1948 by R. D. Hotchkiss (Hotchkiss, 1948) and was further extensively studied since the 1970s (Razin & Cedar, 1977). Recently, evidence for methylation of adenine has been also reported in mammals (Koziol, Bradshaw et al., 2016). Here, we will focus on cytosine modifications in mammals.

DNA cytosine methylation is catalyzed by DNA methyltransferases (Dnmts) that transfer a methyl group from S-adenosyl methionine (SAM) to the fifth carbon of a cytosine residue to form 5mC. The majority of 5mC bases are present in CpG dinucleotides, however non-CpG methylation was also observed especially in mouse embryonic stem cells (mESCs) and brain tissue (Guo, Su et al., 2014). DNA methylation plays a major role in gene expression, cellular differentiation, genomic imprinting, X-inactivation, inactivation of transposable elements and embryogenesis.

Cytosine methylation patterns are mainly established by de novo methyltransferases Dnmt3a, Dnmt3b and their regulatory unit Dnmt3l during early embryonic and germ cell development. Once the patterns are established, they are maintained throughout cell generations by Dnmt1 (Bestor, Laudano et al., 1988, Li, Bestor et al., 1992). Unlike Dnmt1 and Dnmt3a/3b, Dnmt2 is a RNA methyltransferase rather than a DNA methyltransferase (Goll, Kirpekar et al., 2006, Okano, Xie et al., 1998, Yoder & Bestor, 1998). A summary of the mouse Dnmt protein family and their domains is shown in Figure 1 and a summary of the respective knockout mice phenotypes is shown in Table 1.

Figure 1 Schematic representation of the Dnmt protein family.
Shown are domain structures of mouse Dnmt proteins and the initial references. Numbers represent amino acid positions. DMAP: Dnmt1-associated protein binding domain; PBD: proliferating cell nuclear antigen (PCNA)-binding domain; TS: targeting sequence; CXXC: CXXC zinc finger domain; PBHD: polybromo-1 protein homologous domain; MTase: methyltransferase; PWWP: proline-tryptophan-tryptophan-proline motif; Cys: cysteine rich domain.

1.1.1 De novo DNA methylation

Overexpression of Dnmt3a and Dnmt3b is capable of methylating both native and synthetic DNA with no preference for hemimethylated DNA (Okano, Bell et al., 1999). The domain structure for de novo methyltransferases Dnmt3a and Dnmt3b is similar, including a DNA binding domain PWWP domain (Qiu, Sawada et al., 2002) and a C-terminal catalytic domain (Okano et al., 1999) (Figure 1). However, several studies showed that the distribution and expression of Dnmt3a and Dnmt3b varies among cell types. Dnmt3a is expressed relatively ubiquitously and two isoforms of Dnmt3a have been identified. One binds to euchromatic and the other to heterochromatic regions (Chen, Ueda et al., 2002, Okano et al., 1998). Dnmt3a knockout mice developed to term and appeared to be normal at birth but most of the homozygous mutant mice became runted and died at about 4 weeks of age (Table 1). Dnmt3b is highly expressed in embryonic implantation stages, as well as in stem cells and progenitor cells and is the major methyltransferase in early embryogenesis (Watanabe, Suetake et al., 2002, Watanabe, Suetake et al., 2004). Several isoforms were identified and among all isoforms only Dnmt3b1 and Dnmt3b2 possess DNA methyltransferase activity (Aoki, Suetake et al., 2001). No viable Dnmt3b knockout mice were recovered at birth, further highlighting its functions in early embryogenesis (Table 1). The major substrates of Dnmt3a/3b are CpGs, but non-CpG methylation activity of Dnmt3a/3b was also detected (Aoki et al., 2001).
Although Dnmt3l does not possess DNA methylation activity (Bourc’his, Xu et al., 2001), it strongly interacts with Dnmt3a/3b and enhances their methylation activity (Aapola, Kawasaki et al., 2000, Hu, Hirasawa et al., 2008, Suetake, Shinozaki et al., 2004). However, high expression levels of Dnmt3l are found only in germ cells and early stage embryos but not in somatic cells (Watanabe et al., 2004) indicating that the methylation activity enhancement is cell type and developmental stage dependent. Disruption of Dnmt3l caused azoospermia in homozygous males and heterozygous progeny of homozygous female died before midgestation (Table 1).

### Table 1. Phenotypes of initial Dnmt knockout mouse models

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dnmt1 null</td>
<td>Homozygous knockout Dnmt1 were stunted, delayed in development and did not survive past midgestation</td>
<td>(Li et al., 1992)</td>
</tr>
<tr>
<td>Dnmt3a null</td>
<td>Knockout mice developed to term and appeared to be normal at birth but most of homozygous mutant mice became runted and died at about 4 weeks of age</td>
<td>(Okano et al., 1999)</td>
</tr>
<tr>
<td>Dnmt3b null</td>
<td>No viable Dnmt3b knockout mice were recovered at birth</td>
<td>(Okano et al., 1999)</td>
</tr>
<tr>
<td>Dnmt2 null</td>
<td>Mice homozygous for this Trdm1 (formerly Dnmt2) knock-out have abnormal RNA methylation while genomic DNA methylation patterns are not detectably altered</td>
<td>(Goll et al., 2006)</td>
</tr>
<tr>
<td>Dnmt3l null</td>
<td>Disruption of Dnmt3l caused azoospermia in homozygous males and heterozygous progeny of homozygous female died before midgestation</td>
<td>(Bourc’his et al., 2001)</td>
</tr>
</tbody>
</table>

#### 1.1.2 Maintenance DNA methylation

Dnmt1 has a preference for hemi-methylated DNA substrates (Song, Rechkoblit et al., 2011) and is the enzyme responsible for the maintenance of DNA methylation after DNA replication (Leonhardt, Page et al., 1992). Homozygous knockout Dnmt1 mice were runted, delayed in development and did not survive past midgestation (Table 1). The major isoform of Dnmt1 in mouse contains 1620 amino acids and includes an N-terminal regulatory domain and a C-terminal catalytic domain (Tucker, Talbot et al., 1996, Yoder, Yen et al., 1996). However, one isoform lacking the most N-terminal 118 amino acids was shown to accumulate in mouse oocytes (Mertineit, Yoder et al., 1998).

The Dnmt1-associated protein (DMAP) binding domain is located at the beginning of the N-terminus of Dnmt1 and it recruits DMAP1 to further maintain the heterochromatin state (Rountree, Bachman et al., 2000). With the contribution of Uhrf1 (ubiquitin-like with PHD and ring finger domains 1), Dnmt1 methylates hemi-methylated DNA generated upon DNA replication by a mechanism encompassing base flipping (Song et al., 2011, Song, Teplova et al., 2012).

In most mouse cells, Dnmt1 localizes to the cell nucleus. In fact, Dnmt1 contains several functional nuclear localization sequences within its N-terminal regulatory domain (Cardoso & Leonhardt, 1999). In early embryos (Cardoso & Leonhardt, 1999) and in post-mitotic neurons (Inano, Suetake et al., 2000) though, it is retained in the cytoplasm. Although highly expressed in mouse embryos, the exclusion of Dnmt1 from nuclei might inhibit DNA methylation conservation after DNA replication (Grohmann, Spada et al., 2005), implying that localization of Dnmt1 also regulates its methylation activity. Within the cell nucleus, the distribution of Dnmt1 is cell cycle dependent (Leonhardt et al., 1992). In G1-phase, it is diffusely distributed throughout the nucleoplasm. In early S-phase, its proliferating cell nuclear antigen (PCNA)-binding domain (PBD) targets Dnmt1 to replication sites and in late S-phase, the targeting sequence (TS) further enhances Dnmt1 binding to replicating pericentromeric heterochromatin (Schermelleh, Haemmer et al., 2007, Schneider, Fuchs et al., 2013). In G2-phase, Dnmt1 is de novo loaded onto pericentromeric heterochromatin via a replication independent mechanism (Easwaran, Schermelleh et al., 2004). Besides its PBD and TS domains, the polybromol-1 protein homologous domain (PBHD) is also involved in targeting Dnmt1 to replication foci (Liu, Oakeley et al., 1998). Between the TS and PBHD domains, a CXXC domain can be found in Dnmt1. The CXXC domain of Dnmt1 occludes access of Dnmt1 catalytic site to non-methylated CpGs and allows Dnmt1 to bind and specifically methylate hemi-methylated CpGs (Song et al., 2011).

#### 1.2 DNA base modifications

The stable covalent C-C bond formed between the methyl group and the cytosine is difficult to be directly removed and, therefore, 5mC is thought to be a long-lived epigenetic mark. After DNA replication, Dnmt1 association with the replication machinery ensures the maintenance of the methylation pattern onto the newly synthesized strand. Failure to do so, e.g., by retention in the cytoplasm as mentioned above, leads to gradual passive loss of DNA methylation over cell generations. DNA replication independent (active) loss of global DNA methylation was also observed in some biological processes such as reprogramming of the paternal genome after fertilization (Mayer, Niveleau et al., 2000) and...
modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in disease

Development of primordial germ cells (PGC) (Hajkova, Erhardt et al., 2002). The active loss of DNA methylation allows rapid reprogramming of the genome in a short time. Similar observations were made in postmitotic neurons indicating that active loss of DNA methylation also occurs in somatic cells and might have important roles in the regulation of gene expression (Martinowich, Hattori et al., 2003).

For several decades, scientists have been interested in identifying pathways or proteins involved in the active loss of DNA methylation. Lacking the evidence to show that C-C bonds can be directly broken in mammals, multi-step processes have been proposed to be involved in the active removal of DNA methylation marks. In 1972, several additional modifications of cytosines were described in rat, mouse and frog brain tissue including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxylcytosine (5caC) and 5-hydroxymethyluracil (5hmU) (Penn, Suwalski et al., 1972) (Figure 2). However, these modifications were considered to be oxidative damage products of DNA (de Rojas-Walker, Tamir et al., 1995, Tardy-Planechaud, Fujimoto et al., 1997). Three decades later, 5hmC was re-discovered in mouse brain tissue (Kriaucionis & Heintz, 2009) and ESCs (Tahiliani, Koh et al., 2009). Furthermore, a family of proteins (TET) was identified that oxidize 5mC to 5hmC both in humans (Tahiliani et al., 2009) and mice (Ito, D'Alessio et al., 2010). TET1 was first described in 2003 as a fusion partner of the mixed lineage leukemia (MLL) gene in acute myeloid leukemia (Lorsbach, Moore et al., 2003) and six years later it was re-discovered as an oxygenase, which can convert 5mC to 5hmC (Tahiliani et al., 2009). Further studies showed that Tet proteins also convert 5hmC to 5fC and 5caC (Ito, Shen et al., 2011, Pfaffeneder, Hackner et al., 2011).

Figure 2 DNA base modifications with respective enzymes.

Dnmts catalyze the addition of a methyl group to cytosine bases. Tet proteins oxidize methylated cytosines to 5hmC, 5fC and 5caC in an iterative manner. 5mC and 5hmC can be further deaminated by Aid/Apobec to T and 5hmU. T, 5hmU, 5fC and 5caC can be removed by the indicated glycosylases. Initial references are indicated.

Deaminases such as Aid and Apobec can recognize 5mC and 5hmC and further convert 5mC to T and 5hmC to 5hmU. Although the deaminase activity is quite low, it is still a possible pathway for DNA demethylation (Guo, Su et al., 2011). In addition, Tets were also shown to oxidize thymine (T) to 5hmU in mouse ESCs (Pfaffeneder, Spada et al., 2014), which additionally leads to loss of DNA methylation. The oxidation products like 5fC, 5caC and 5hmU can be recognized and excised by the glycosylases Tdg (Maiti & Drohat, 2011) and Neil (Muller, Bauer et al., 2014) to create an abasic site on DNA, which is further repaired by enzymes of the base excision repair (BER) pathway. In addition to Tdg, 5hmU can also be recognized by other glycosylases like Mbd4 (Hashimoto, Zhang et al., 2012b) and Smug1 (Kennemer, Dingler et al., 2012). Accordingly, a combination of oxidation, deamination and BER might contribute to the active removal of DNA methylation. In mouse zygotes the decrease of 5mC and increase of 5hmC suggests that 5hmC might be an intermediate of DNA methylation removal. However, recent studies showed that loss of 5mC mainly happens before S-phase, whereas gain of 5hmC occurred after DNA replication (Amouroux, Nashun et al., 2016), indicating that besides the conversion of 5mC to 5hmC, other pathways might...
contribute to methylation removal before DNA replication in mouse zygotes.

### 1.3 Methylcytosine modifiers

Until now three members of the Tet protein family named Tet1 (mouse homolog of human TET1), Tet2 (mouse homolog of human TET2) and Tet3 (mouse homolog of human TET3) have been identified in mice and humans. All three Tets share a conserved C-terminal catalytic domain including a Cys-rich and a double-stranded β helix (DSBH) domain, which belong to the cupin-like dioxygenase superfamily; and exhibit iterative iron- and oxoglutarate dependent oxidation activity (Figure 3).

#### 1.3.1 Tissue and genome-wide distribution of Tet

During mouse embryo development, Tet3 is highly expressed in oocytes and zygotes. Female mice depleted of Tet3 in the germ line showed severely reduced fecundity and their heterozygous mutant offspring lacking maternal Tet3 suffer an increased incidence of developmental failure. Since Tet1, Tet2 as well as Tet1 & Tet2 double knockout mice are viable, this suggests that Tet1 and Tet2 are not essential for mouse development (Table 2).

![Figure 3 Schematic representation of the Tet protein family.](image)

**Table 2. Phenotypes of initial Tet knockout mouse models**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1 null</td>
<td>Mice are viable, fertile and grossly normal though some mutant mice have a slightly smaller body size at birth</td>
<td>(Dawlaty, Ganz et al., 2011)</td>
</tr>
<tr>
<td>Tet1 null</td>
<td>Animals exhibited abnormal hippocampal long-term depression and impaired memory extinction</td>
<td>(Rudenko, Dawlaty et al., 2013, Zhang, Cui et al., 2013)</td>
</tr>
<tr>
<td>Tet2 null</td>
<td>Approximately 1/3 of Tet2−/− and 8% of Tet2+− mice died within 1 year of age because of the development of myeloid malignancies resembling characteristics of chronic myelomonocytic leukemia (CMML), myeloproliferative disorder (MPD)-like leukemia, and myelodysplastic syndrome (MDS)</td>
<td>(Li, Cai et al., 2011)</td>
</tr>
<tr>
<td>Tet3 null</td>
<td>Female mice depleted of Tet3 in the germ line show severely reduced fecundity and their heterozygous mutant offspring lacking maternal Tet3 suffer an increased incidence of developmental failure. Oocytes lacking Tet3 also seem to have a reduced ability to reprogram the injected nuclei from somatic cells</td>
<td>(Gu, Guo et al., 2011)</td>
</tr>
<tr>
<td>Tet1 &amp; Tet2 null</td>
<td>Double deficient mice had reduced 5mC and increase 5hmC levels and abnormal methylation at various imprinted loci. Animals of both sexes were fertile with females having smaller ovaries and reduced fertility</td>
<td>(Dawlaty, Breiling et al., 2014)</td>
</tr>
</tbody>
</table>

Tet mediated 5mC to 5hmC conversion is thought involved in reprogramming the paternal genome (Gu et al., 2011, Iqbal, Jin et al., 2011, Wossidlo, Nakamura et al., 2011, Zhang, Su et al., 2012) and also in reprogramming donor cell DNA during somatic cell nuclear transfer (Gu et al., 2011). In addition, HIV-1 Vpr binding protein (VprBP) mediated monoubiquitylation promotes Tet binding to chromatin and enhances 5hmC formation (Nakagawa, Lv et al., 2015) in mouse embryos. This process is involved in female germ cell development and genome reprogramming in zygotes (Yu, Zhang et al., 2013). During PGC reprogramming, Tet1 and Tet2 are highly expressed (Hackett, Sengupta et al., 2013). However, genome wide DNA methylation removal is unaffected by the absence of Tet1 and Tet2 and, thus, 5mC, indicating that the first comprehensive 5mC loss does not involve 5hmC formation. Instead Tet1 and Tet2 have a locus...
specific role in shaping the PGC epigenome during subsequent development (Vincent, Huang et al., 2013). Further studies showed that Tet1 has a critical role in the erasure of genomic imprinting (Yamaguchi, Shen et al., 2013) and it controls meiosis by regulating meiotic gene expression (Yamaguchi, Hong et al., 2012). In mouse ESCs, both Tet1 and Tet2, as well as their oxidation product 5hmC are highly abundant (Ito et al., 2010). While Tet2 preferentially acts on gene bodies, Tet1 preferentially acts on promoters and transcription start sites (Huang, Chavez et al., 2014). Tet1 and Tet2 double knockout ESCs remained pluripotent, but were depleted of 5hmC and caused developmental defects in chimeric embryos (Dawlaty et al., 2014). During somatic reprogramming, Tet2 is required for 5hmC formation at the Nanog locus (Doerge, Inoue et al., 2012). Further studies showed that the recruitment of Tet1 by Nanog facilitates the expression of a subset of reprogramming target genes, such as Oct4 (Costa, Ding et al., 2013). Accordingly, Tet1 can replace Oct4 during somatic cell reprogramming in conjunction with Sox2, Klf4 and e-Myc (Gao, Chen et al., 2013). The data above indicate that Tet mediated 5hmC formation is not only important for ESCs differentiation but also for somatic reprogramming. In mouse brain, 5hmC is a constituent of nuclear DNA (Kriaucionis & Heintz, 2009). Tet1 plays an important role in regulating neural progenitor cell (NPC) proliferation in adult mouse brain (Zhang et al., 2013) and is critical for neuronal activity-regulated gene expression and memory extinction (Table 2) (Rudenko et al., 2013).

1.3.2 Regulation of Tet activity
Similar to Dnmt1, Tet proteins use a base flipping mechanism to oxidize 5mC, which includes binding of DNA by a Watson-Crick polar hydrogen and van der Waals interactions, flipping out 5mC (Hashimoto, Pais et al., 2014, Hu, Li et al., 2013) and oxidation of 5mC to 5hmC (Hashimoto, Pais et al., 2015, Hu, Lu et al., 2015). Although Tet proteins successively oxidize 5mC to 5caC, recent experimental data showed that, in comparison with 5hmC and 5CaC, 5mC is the preferential substrate for Tet2 (Hu et al., 2015). This preference was further confirmed by computer simulations (Lu, Hu et al., 2016). In cultured cells, the majority of genomic 5hmC nucleotides are stable (Bachman, Uribe-Lewis et al., 2014), indicating that 5mC is not only involved in loss of DNA methylation, but represents an additional stable epigenetic mark. The global content of 5hmC varies in mouse tissues, as the cells adapt to cell culture conditions (Nestor, Ottaviano et al., 2012). The cell-, tissue- and developmental stage specific distribution of 5hmC indicates that the conversion of 5mC to 5hmC is highly regulated. Although the N-terminal domain of Tet proteins was shown to be dispensable for their catalytic activity, it was shown to possess regulatory functions. A CXXC domain, which usually binds specifically to unmethylated CpGs can be found in the N-terminus of Tet1 and Tet3 (Liu, Wang et al., 2013). While the CXXC domain of Tet1 cannot bind to DNA in vitro (Frauer, Rottach et al., 2011b), it binds to unmodified C, 5mC- or 5hmC-modified CpGs in vivo (Xu, Wu et al., 2011b, Zhang, Zhang et al., 2010). Moreover, binding of the CXXC domain to DNA was shown to control DNA methylation levels by preventing unwanted DNA methyltransferase activity in ESCs (Xu et al., 2011b) or aberrant methylation spreading into CpG islands (CGIs) in differentiated cells (Jin, Lu et al., 2014). The CXXC domain of Xenopus Tet3 recognizes non-methylated cytosines in either CpG or non-CpG context, and it is critical for specific Tet3 targeting (Xu, Xu et al., 2012). Although Tet2 proteins do not have a CXXC domain, recent studies showed that the ancestral CXXC domain of Tet2 is encoded by a distinct gene named Idax. Unlike the CXXC domain of Tet1 and Tet3, the CXXC domain of Idax binds unmethylated CpGs. Through direct protein-protein interactions of Tet2 and Idax, Tet2 is recruited to DNA. Furthermore, Tet2 is degraded by caspase activation, which is triggered by the CXXC of Idax (Ko, An et al., 2013).

Two parts of the DSBH domain are connected by a potential regulatory spacer region. Although the spacer region was shown to be dispensable for 5mC catalytic activity (Hu et al., 2013), post-translational modifications, such as phosphorylation and O-GlcNAcylation were observed in the spacer region (Bauer, Göbel et al., 2015) indicating that it might exhibit regulatory functions. O-GlcNAc transferase (Ogt) directly interacts with Tet proteins and consequently Tet proteins are GlcNAcylated. The GlcNAcylation does not affect the hydroxylase activity of Tet2 and Tet3, rather Tet2 and Tet3 were shown to promote Ogt activity (Deplus, Delatte et al., 2013) by enhancing the localization of Ogt to chromatin (Chen, Chen et al., 2013, Ito, Katsura et al., 2014). However, it was shown that Ogt drives Tet3 out of the nucleus further affecting its activity on DNA (Zhang, Liu et al., 2014). In mouse ESCs, Ogt is recruited to unmethylated CpG promoters in a Tet1 dependent manner (Vella, Scelfo et al., 2013). In addition to posttranslational modifications, mutations within the spacer region of Tet2 were observed in myelodysplastic syndrome (MDS), thus further highlighting the importance of this region (Ko, Huang et al., 2010). In vivo, besides posttranslational modifications, Tet activity is regulated by protein-protein interactions, such as with Sin3a. In mouse ESCs, the interaction between Sin3a and Tet1 allows Sin3a to repress a subset of Tet1 target genes (Williams, Christensen et al., 2011). In mouse zygotes, Tet3 mediated 5mC to 5hmC conversion is involved in reprogramming of the paternal but not the maternal genome although they share the same cytoplasm (Mayer et al., 2000). The resistance of the maternal genome to reprogramming is achieved by a protein named developmental pluripotency associated 3 (Dppa3, or PGC7). Dppa3 binds to histone H3K9me2 (Nakamura, Liu et al., 2012) and interacts with Tet3 further blocking the activity of Tet3 (Bian & Yu, 2014). Dazl, an RNA-binding protein known to play a key role in germ cell development, was shown to enhance Tet1 mediated 5mC
to 5hmC conversion by enhancing Tet1 protein translation (Welling, Chen et al., 2015). In addition, growth arrest and DNA damage inducible protein 45 (Gadd45) interacts with Tet1 and Tdg and promotes loss of DNA methylation by enhancing 5fC/5caC removal (Kienhöfer, Musheev et al., 2015, Li, Gu et al., 2015). Finally, Tet mediated 5mC to 5hmC conversion was shown to be regulated by Tet cofactors. 2-ketoglutarate (2-KG), one of the cofactors for Tet oxidation is produced by isocitrate dehydrogenase 1/2 (Idh1/2) in vivo. However, mutated Idh1/2 produce 2-hydroxyglutarate, a competitive inhibitor of 2-KG, which can further inhibit 5mC to 5hmC conversion (Konstandin, Bultmann et al., 2011). Vitamin C is a potential cofactor for Tet mediated oxidation and was shown to enhance Tet activity, which leads to increased global 5hmC in ESCs (Blaschke, Ebata et al., 2013). ATP was also shown to be involved in regulating Tet activity. In vitro, the reaction of Tet mediated 5mC to 5caC can be enhanced by addition of ATP (He, Li et al., 2011).

1.3.3 Hydroxymethylcytosine maintenance

Dnmt1 recognizes hemi-mC DNA and methylates the nascent DNA strand after replication during the S-phase of the cell cycle. However, in vitro studies showed a 60-fold decreased binding ability of Dnmt1 to hemi-hmC DNA compared to hemi-mC DNA (Hashimoto, Liu et al., 2012a), indicating that hemi-hmC DNA might not be a substrate for Dnmt1. Previous studies showed that Np95 can recognize 5hmC and bind to hemi-hmC DNA (Frauer, Hoffmann et al., 2011a), indicating that Np95 might target Dnmt1 to hemi-hmC containing replication forks to maintain hmC after DNA replication. In addition, Dnmt3a and Dnmt3b recognize hemi-hmC DNA (Hashimoto et al., 2012a) and are necessary for methylation maintenance at repeat genomic elements (Chen, Ueda et al., 2003) suggesting Dnmt3a/3b might play a role in maintaining 5hmC after DNA replication in repeat elements. In vivo, the majority of 5hmC is present in CpG dinucleotides. However, 5hmC has also been observed in non-CpG context, especially in gene bodies (Pastor, Pape et al., 2011, Xu et al., 2011b). One important role of CpG methylation in gene promoter regions is the repression of gene expression by directly or indirectly preventing interactions between promoter and transcription factors. Hydroxymethylated CpGs might affect binding of transcription factors and/or 5mC readers to DNA.

2 DNA modification readers

In mammals, the methylome is specifically read by a variety of proteins known as methyl-CpG binding proteins (MBPs), which based on structural features are further classified into three main families: the methyl-CpG binding domain (MBD) protein family (Baymaz, Fournier et al., 2014, Cross, Meehan et al., 1997, Hendrich & Bird, 1998, Hendrich & Tweedie, 2003, Laget, Joulie et al., 2010, Lewis, Meehan et al., 1992), the Kaiso protein family (Daniel & Reynolds, 1999, Filion, Zhenilo et al., 2006) and the SET and RING finger associated (SRA) domain protein family (Hopfner, Mousli et al., 2000, Mori, Li et al., 2002). While initially identified as 5-methylcytosine (5mC) binding proteins, recent studies indicate that a distinct and dynamic set of MBPs binds the Tet oxidation product 5-hydroxymethylcytosine (5hmC) during differentiation (Figure 4) (Frauer et al., 2011a, Mellen, Ayata et al., 2012, Spruijt, Gnerlich et al., 2013). Through further interactions with multiple protein partners, MBPs provide a link between cytosine derivatives and functional chromatin states in a temporally and spatially regulated fashion.

![Figure 4](image_url)

**Figure 4** (Hydroxy)methylcytosine readers.
Shown are MBP proteins that read 5mC and 5hmC as indicated and the initial references.

2.1 MBD protein family

Presently, the MBD protein family consists of eleven members (Mecp2, Mbd1-6, SETDB1, SETDB2, TIP5/BAZ2A, BAZ2B) (Baymaz et al., 2014, Cross et al., 1997, Hendrich & Bird, 1998, Hendrich & Tweedie, 2003, Laget et al., 2010, Lewis et al., 1992). All of them share a common protein motif, the 70-85 amino acids long methyl-CpG binding domain (MBD), which enables some, but not all family members, to selectively bind to single methylated CpG dinucleotides. With the exception of Mbd2 and Mbd3, MBD proteins bear little resemblance outside their MBD (Hendrich & Bird, 1998). Instead, MBD proteins comprise several distinct domains that confer unique DNA binding, as well as other functional features. Since this review covers DNA (hydroxy)methylation dependent processes, we will thereafter focus on MBD family members (Figure 5) capable of binding to (hydroxy)methylated CpG.
Modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in disease

2.1.1 Mecp2:

The first protein described to selectively recognize and bind single, symmetrically methylated CpG dinucleotides was Mecp2 (Lewis et al., 1992). It is abundantly expressed in the central nervous system with the highest protein levels in post-mitotic neurons (Akbarian, Chen et al., 2001, Jung, Jugloff et al., 2003, Traynor, Agarwal et al., 2002). Of the two alternatively spliced isoforms (Mecp2 e1 and e2), which differ in their N-terminus, Mecp2 e2 was first identified and is, therefore, best characterized (Kriaucionis & Bird, 2004, Mnatzakanian, Lohi et al., 2004). Although both isoforms distribute differently in developing and postnatal mouse brains, no functional differences have been identified so far (Dragich, Kim et al., 2007).

Both Mecp2 variants include two functionally characterized domains, the MBD and the transcriptional repression domain, abbreviated TRD. While the MBD proved sufficient to direct specific binding to methylated cytosines (Nan, Meehan et al., 1993), the TRD was originally identified as the region required for transcriptional repression in vitro and in vivo (Jones, Veenstra et al., 1998, Kaludov & Wolffe, 2000, Lewis et al., 1992, Nan, Campoy et al., 1997). Circular dichroism and protease digestion analysis revealed that outside these functional domains the full-length protein is largely devoid of secondary structure (Adams, McBryant et al., 2007). With almost 60% unstructured regions, Mecp2 is reckoned among the intrinsically disordered proteins, which often undergo a disorder-to-order transition upon binding to other macromolecules (Adams et al., 2007). Indeed, recent studies demonstrate that Mecp2 gains secondary structure and acquires substantial thermal stabilization upon binding to DNA (Ghosh, Nikitina et al., 2010). Unlike its name implies DNA binding is, however, not solely mediated via its 5mC specific MBD. Instead, as indicated by the release of Mecp2 upon salt extraction, regions outside the MBD contribute to the overall binding energy through electrostatic interactions (Meehan, Lewis et al., 1992). As shown by electrophoretic mobility shift assays (EMSA), these sequence-unspecific DNA binding motifs include the TRD and, based on their relative location to the MBD and TRD, the so-called intervening domain (ID), as well as the C-terminal domain alpha (CTD alpha) (Ghosh et al., 2010). The N-terminal domain (NTD) of Mecp2 in contrast, contributes indirectly to the overall binding affinity by enhancing the methylation specificity of the MBD through conformational coupling (Ghosh et al., 2010). An analog synergistic increase in DNA binding efficiency was observed through interdomain interactions between the TRD and the C-terminal part of the protein (Ghosh et al., 2010). Similar to the NTD, the C-terminal domain beta (CTD beta) does not directly interact with DNA (Ghosh et al., 2010). Nevertheless, the overall chromatin binding efficiency was lost upon its deletion (Nikitina, Shi et al., 2007b). Consistent with this, the CTD beta induced moderate and reproducible shifts with nucleosomal arrays, but not with naked DNA (Ghosh et al., 2010), suggesting that the most C-terminal 192 residues of Mecp2 harbor a chromatin interaction surface (Nikitina et al., 2007b). Indeed, Mecp2 has been shown to interact with histone H3 and, similar to the linker histone H1, binds to nucleosomes close to the linker DNA entry-exit site (Nikitina et al., 2007b). As a result, the entering and exiting linker DNA segments are brought in close proximity to form a stem-like motif (Nikitina, Ghosh et al., 2007a), which bears strong

Figure 5 Schematic representation of the MBD protein family.

Shown are domain structures of mouse MBD proteins and the initial references. Numbers represent amino acid positions. MBD: methyl-CpG binding domain; CXXC: CXXC zinc finger domain; TRD: transcriptional repression domain; GR: glycine/arginine; E: glutamic acid.
resemblance to structures induced by H1 (Bednar, Horowitz et al., 1998, Hamiche, Schultz et al., 1996). The modes of chromatin compaction, however, differ significantly from each other. While histone H1 arranges nucleosomes and linker DNA into regular zigzag-shaped chromatin fibers (Woodcock, 2006), Mecp2 forms highly compacted globular structures in vitro due to its multiple DNA and chromatin binding domains (Georgel, Horowitz-Scherer et al., 2003). Accordingly, Mecp2 was shown to induce clustering of pericentric heterochromatin in a dose-dependent manner in vivo to establish a locally repressive chromatin environment (Agarwal, Becker et al., 2011, Brero, Easwaran et al., 2005). More recently, Szulwach and colleagues provided evidence that binding of Mecp2 to methylated CpG dinucleotides may protect 5mC against Tet mediated oxidation thereby preventing reactivation of silenced genes (Szulwach, Li et al., 2011). The underlying mechanism, however, has so far not been described. An additional level of regulation is achieved through various protein-protein interactions. While direct homo- and hetero-interactions of Mecp2 and Mbd2 were shown to cross-link chromatin fibers (Becker, Allmann et al., 2013), physical associations of Mecp2 with the transcriptional co-repressor Sin3a and histone deacetylase 2 (HDAC2) via its TRD contribute to the global heterochromatin architecture through histone hypoacetylation (Jones et al., 1998, Nan, Ng et al., 1998). Consequently, MeCP2 deficiency was demonstrated to result in global changes in neuronal chromatin architecture, elevated histone acetylation levels and increased transcriptional noise in a DNA methylation-dependent manner (Cohen, Gabel et al., 2011, Skene, Illingworth et al., 2010). A number of other repressive protein partners of Mecp2 have been identified including the co-repressors c-Ski (Kokura, Kaul et al., 2001), CoREST (Lunyak, Burgess et al., 2002) and NCoR/SMRT (Stancheva, Collins et al., 2003), as well as DNA methyltransferase Dnmt1 (Kimura & Shiota, 2003) and H3K9 methyltransferase (Fuks, Hurd et al., 2003). Both, binding of Mecp2 to DNA, as well as interactions with protein partners are affected by post-translational modifications (PTMs). Neuronal activity induced phosphorylation and dephosphorylation of Mecp2 was shown to modulate its association with promoters of specific genes, as well as with interaction partners (reviewed in (Li & Chang, 2014)). More recently, poly(ADP-ribosylation) of Mecp2 in mouse brain tissue was reported, which anticorrelated with its chromatin binding affinity and clustering ability (Becker, Zhang et al., 2016). Furthermore, ubiquitylation (Gonzales, Adams et al., 2012), SUMOylation (Cheng, Huang et al., 2014), acetylation (Zocchi & Sassone-Corsi, 2012) and methylation (Jung, Li et al., 2008) were shown to substantially contribute to the functional versatility of Mecp2. Another unanticipated level of functional complexity was demonstrated by recent work of Spruijt and colleagues who identified Mecp2 as reader of 5hmC in mouse embryonic stem cells (mESC) by quantitative mass-spectrometry-based proteomics (Spruijt et al., 2013). Moreover, independent studies of Mellen et al., revealed Mecp2 as the major 5hmC-binding protein in mouse brain, which moreover turned out to bind both, 5mC- and 5mC-containing substrates with similar affinity (Mellen et al., 2012). Finally, chip-chip analysis using antibodies against MECP2 in a human neuronal cell line demonstrated that around 2/3 of strongly MECP2 bound promoters were transcriptionally active (Yasui, Peddada et al., 2007). Subsequent analysis of gene expression patterns in Mecp2 knockout and overexpressing mice concurred that Mecp2 functions as an activator as well as a repressor of transcription (Chahrour, Jung et al., 2008). Hence, the traditional view of Mecp2 as a 5mC-dependent transcriptional silencer may be incomplete and its biology appears far more complicated than previously assumed. Both, male and female mice lacking Mecp2 (Table 3) developed an uncoordinated gait and reduced spontaneous movement between three and eight weeks of age and most died between 6 and 12 weeks (Chen, Akbarian et al., 2001, Guy, Hendrich et al., 2001). Furthermore, most animals developed hind limb claspings, irregular breathing, misaligned jaws and uneven wearing of teeth. Mutant brains were reduced in weight, however no structural abnormalities or signs of neurodegeneration were detected, suggesting that stability of brain function, not brain development per se, is impaired in the absence of Mecp2. Consistent with this hypothesis, re-expression of the Mecp2 gene in Mecp2 lowerscipy mice proved sufficient to reverse the neurological symptoms of RTT, indicating that Mecp2-deficient neurons develop normally and are not irreversibly damaged (Guy, Gan et al., 2007). Further microarray analyses revealed that knockout of Mecp2 implicates only minor changes in gene expression (Tudor, Akbarian et al., 2002). Subsequent studies demonstrating increased expression restricted to non-coding RNA in brain of Mecp2 deficient mice (Muotri, Marchetto et al., 2010, Skene et al., 2010), indicated that Mecp2 may not act as a gene-specific transcriptional repressor, but might instead dampen transcriptional noise genome-wide in a DNA methylation-dependent manner (Skene et al., 2010). Accordingly, expression of repetitive elements (Muotri et al., 2010, Skene et al., 2010) as well as retrotransposition of LINE1 was increased in brain of Mecp2 deficient mice (Muotri et al., 2010).

### 2.1.2 Mbd1:

Mbd1, initially termed PCM1, is expressed in somatic cells and represents the largest member of the MBD family (Cross et al., 1997, Hendrich & Bird, 1998). Similar to Mecp2, Mbd1 contains a MBD and a TRD, which have analogue functions to that of Mecp2 (Ng, Jeppesen et al., 2000). In addition, depending on the isoform, Mbd1 contains two or three CXXC zinc finger motifs (Fujita, Takebayashi et al., 1999, Jorgensen, Ben-Porath et al., 2004). The most C-terminal one, referred to as CXXC3, is homologue to zinc fingers found in Dnmt1, CpG binding protein CGBP, histone H3K4
Modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in disease

11

methyltransferase MLL and histone H3K36 deacetylases of the Jumonji family JHDMA1 and JHDMA1B (Jorgensen et al., 2004, Lee & Skalnik, 2005, Tsukada, Fang et al., 2006). While CXXC3 was shown to bind unmethylated CpG dinucleotides in vitro (Birke, Schreiner et al., 2002, Jorgensen et al., 2004, Lee & Skalnik, 2002, Lee & Skalnik, 2005), the remaining zinc finger motifs of Mbd1 lack a conserved glutamine residue and the characteristic KFFG motif necessary for binding to DNA (Jorgensen et al., 2004). Accordingly, Mbd1 isoforms containing the first two CXXC domains preferentially bind methylated DNA via their MBD, whereas isoforms comprising a complete set of zinc fingers have the ability to bind both, methylated and unmethylated substrates (Baubec, Ivanek et al., 2013, Jorgensen et al., 2004).

As a transcriptional repressor, Mbd1 was thus shown to inhibit transcription from both, methylated and unmethylated promoters in reporter gene assays (Fujita et al., 1999, Jorgensen et al., 2004). While methylation dependent silencing is mediated by the MBD and TRD, suppression of nonmethylated reporter constructs required the presence of the CXXC3 domain (Jorgensen et al., 2004). Although, a precise association between Mbd1 and HDACs has not been described, transcriptional repression was partially sensitive to trichostatin A (TSA), an HDAC inhibitor (Ng et al., 2000). In most assays, however, Mbd1 behaved as an HDAC-independent repressor (Ng et al., 2000).

Table 3: Phenotype of initial MBP deficient mouse models

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mecp2 null</td>
<td>Rett syndrome-like phenotype. Between 3 and 5 weeks: uncoordinated gait, reduced spontaneous movement, hind limb clasping, irregular breathing, misaligned jaws, uneven wearing of teeth, reduced brain weight and neuronal cell size. Between 6 and 12 weeks: rapid weight loss and death.</td>
<td>(Chen et al., 2001, Guy et al., 2001)</td>
</tr>
<tr>
<td>Mbd1 null</td>
<td>Viable and fertile. Impaired spatial learning, decreased neurogenesis, reduced long-term potentiation, decreased genomic stability.</td>
<td>(Zhao, Ueba et al., 2003)</td>
</tr>
<tr>
<td>Mbd2 null</td>
<td>Viable, fertile. Maternal nurturing defects: reduced litter size and weight of pups.</td>
<td>(Hendrich, Guy et al., 2001)</td>
</tr>
<tr>
<td>Mbd3 null</td>
<td>Early embryonic lethality</td>
<td>(Hendrich et al., 2001)</td>
</tr>
<tr>
<td>Mbd4 null</td>
<td>Viable and fertile. Increased number of C:G to T:A transitions at CpG sites.</td>
<td>(Millar, Guy et al., 2002)</td>
</tr>
<tr>
<td>Kaiso null</td>
<td>Viable and fertile. Reduced tumorigenesis</td>
<td>(Prokhortchouk, Sansom et al., 2006)</td>
</tr>
<tr>
<td>Np95 null</td>
<td>Early gestational lethality. Developmental arrest shortly after gastrulation.</td>
<td>(Sharif, Muto et al., 2007)</td>
</tr>
<tr>
<td>Np97 null</td>
<td>Phenotype not described.</td>
<td>(Li, Qiu et al., 2013)</td>
</tr>
</tbody>
</table>

Instead, MBD1 has been found associated with histone H3K9 methyltransferases SETDB1 (Sarraf & Stancheva, 2004) and Suv39h1 (Fujita, Watanabe et al., 2003). Association to SETDB1 mediates transcriptional repression throughout the cell cycle (Sarraf & Stancheva, 2004). During S-phase, however, MBD1 was shown to recruit SETDB1 to the large subunit of chromatin assembly factor CAF-1 to form an S-phase specific complex that mediates methylation of H3K9 in a post-replicative manner (Sarraf & Stancheva, 2004). Accordingly, H3K9 methylation is lost in the absence of MBD1 and results in activation of specific genes, such as p53BP2 (Sarraf & Stancheva, 2004).

Instead, MBD1 has been found associated with histone H3K9 methyltransferases SETDB1 (Sarraf & Stancheva, 2004) and Suv39h1 (Fujita, Watanabe et al., 2003). Association to SETDB1 mediates transcriptional repression throughout the cell cycle (Sarraf & Stancheva, 2004). During S-phase, however, MBD1 was shown to recruit SETDB1 to the large subunit of chromatin assembly factor CAF-1 to form an S-phase specific complex that mediates methylation of H3K9 in a post-replicative manner (Sarraf & Stancheva, 2004). Accordingly, H3K9 methylation is lost in the absence of MBD1 and results in activation of specific genes, such as p53BP2 (Sarraf & Stancheva, 2004).

MBD1-mediated transcriptional repression and heterochromatin maintenance was shown to be regulated by SUMOylation (Lyst, Nan et al., 2006, Uchimura, Ichimura et al., 2006). In human cells, two E3 SUMOligases (PIAS1 and PIAS3) were shown to SUMOylate MBD1 (Lyst et al., 2006). While SUMO1-conjugation blocks the MBD1 and SETDB1 interaction, modification with SUMO2/3 recruits SETDB1 thereby stimulating its repressive function (Uchimura et al., 2006). Although mice lacking Mbd1 (Table 3) developed normally and appeared healthy throughout life, they were impaired in spatial learning, had decreased neurogenesis and reduced long-term potentiation in the dentate gyrus of the hippocampus (Zhao et al., 2003). Moreover, Mbd1 deficient neural stem cells differentiated less and had decreased genomic stability (Zhao et al., 2003).

2.1.3 Mbd2:
Mbd2 and Mbd3 are the only known members of the MBD protein family with significant sequence similarity beyond the MBD (Hendrich & Bird, 1998) and, thus, are believed to have arisen from an ancient duplication during evolution of the vertebrate lineage (Hendrich & Tweedie, 2003). Consistent with this, a homologue Mbd2/3 like protein was identified in invertebrates, including Drosophila (Lyko, Ramsahoye et al., 2000, Marhold, Kramer et al., 2004). Despite the high degree of sequence similarity, Mbd3 lacks the amino-terminal extension of...
Mbd2, which contains a repeat consisting of glycine and arginine residues (Hendrich & Bird, 1998). While both, Mbd2 and Mbd3 contain a C-terminal coiled coil (CC) domain that mediates protein-protein interactions, Mbd3 was shown to comprise an additional glutamic acid repeat at its extreme COOH-terminus (Becker et al., 2013, Gnanapragasam, Scarsdale et al., 2011, Hendrich & Bird, 1998).

Mbd2 contains two in-frame start codons, which give rise to Mbd2a and the truncated version Mbd2b, which lacks the first 140 amino acids (Hendrich & Bird, 1998). In vivo, however, only Mbd2a, but not Mbd2b, has been detected (Ng, Zhang et al., 1999). Inclusion of an alternative third exon gives rise to an additional isoform of Mbd2, named Mbd2c, which lacks the C-terminal TRD and CC domain due to an early stop codon (Hendrich & Bird, 1998). Tethering of Mbd2a near a promoter via a GAL4 DNA binding domain was shown to mediate transcriptional repression that is sensitive to TSA (Ng et al., 1999). Similarly, Mbd2b enhanced transcriptional repression of methylated reporter constructs in co-transfection assays (Boeke, Ammerpohl et al., 2000). Different from other MBD family members, the sequence required for transcriptional repression (TRD) partially overlapped with the MBD (Boeke et al., 2000), indicating a strong interrelation of methylation binding and transcriptional silencing. In line with this, the TRD directly interacts with the transcriptional repressor Sin3A (Boeke et al., 2000). Moreover, Mbd2 co-purified with a large protein complex known as NuRD (Nucleosome remodeling and histone deacetylation), which includes chromatin remodeling ATPase Mi-2, as well as histone deacetylases HDAC1 and HDAC2 (Le Guezennec, Vermeulen et al., 2006, Mahajan, Narlikar et al., 2005, Ng et al., 1999, Wade, Gegonne et al., 1999, Zhang, Ng et al., 1999). Electrophoretic mobility shift assays indicated that Mbd2a directs the NuRD complex, which is implicated in transcriptional silencing, to methylated DNA (Zhang et al., 1999). Finally, immunoprecipitation analysis showed that Mbd2 associates with HDAC1 in mammalian cells and is the long sought methyl-CpG binding component of the 400-800 kDa McCP1 (Methyl-CpG binding protein 1) complex (Meehan, Lewis et al., 1989, Ng et al., 1999).

Mbd2 was shown to bind 5mC in a manner similar to the isolated MBD of Mecp2 (Hendrich & Bird, 1998, Wade et al., 1999). Binding of oxidative 5mC derivatives, however, has not been observed (Hashimoto et al., 2012a, Wakefield, Smith et al., 1999). Despite its inability to recognize 5mC, three different Mbd3 isoforms (Mbd3a-c) that vary in their amino termini were detected within the NuRD repression complex in embryonic stem cells (Kaji, Caballerò et al., 2006, Zhang et al., 1999). ESCs lacking Mbd3-NuRD displayed a severe defect in differentiation that lead to persistent self-renewal even in the absence of leukemia inhibitory factor (LIF) (Kaji et al., 2006). More recently, depletion of Mbd3 in somatic cells was shown to enhance the reprogramming efficiency of the four Yamanaka factors (Oct4, Sox2, Klf4, Myc) (Luo, Ling et al., 2013, Rais, Zviran et al., 2013). Accordingly, Mbd3 was proposed to play a key role in lineage commitment and pluripotency (Reynolds, Latos et al., 2012, Whyte, Bilodeau et al., 2012, Yildirim, Li et al., 2011).

Mbd3 isoforms (Mbd3a-c) that vary in their amino termini were detected within the NuRD repression complex in embryonic stem cells (Kaji, Caballerò et al., 2006, Zhang et al., 1999). ESCs lacking Mbd3-NuRD displayed a severe defect in differentiation that lead to persistent self-renewal even in the absence of leukemia inhibitory factor (LIF) (Kaji et al., 2006). More recently, depletion of Mbd3 in somatic cells was shown to enhance the reprogramming efficiency of the four Yamanaka factors (Oct4, Sox2, Klf4, Myc) (Luo, Ling et al., 2013, Rais, Zviran et al., 2013). Accordingly, Mbd3 was proposed to play a key role in lineage commitment and pluripotency (Reynolds, Latos et al., 2012, Whyte, Bilodeau et al., 2012, Yildirim, Li et al., 2011).

Contradictory studies using neural and epithelial-derived stem cells, however, indicate a role for Mbd3 in facilitating induction of pluripotency and argue that its function may be context specific (dos Santos, Tosti et al., 2014). Binding sites of Mbd3 have been mapped genome-wide in mouse and human cells (Baubec et al., 2013, Gunther, Rust et al., 2013, Shimbo, Du et al., 2013, Yildirim et al., 2011). While Yildirim and colleagues identified Mbd3 bound to transcription start sites (TSS) of CpG-rich, hydroxymethylated marked promoters (Yildirim et al., 2011), Baubec et al. found Mbd3 bound to enhancers independent of CpG density and (hydroxy)methylation status (Baubec et al., 2013). Further data questioning the interaction of Mbd3 with hydroxymethylated DNA was provided by Spruijt and coworkers, who did not detect Mbd3 among hydroxymethylation specific readers (Spruijt et al., 2013).

Although both, Mbd2 and Mbd3 associate with the NuRD complex, the two MBD containing complexes appear to have no functional overlap since knockout of Mbd3 in mice is embryonic lethal, whereas Mbd2 deficient mice are viable and fertile (Hendrich et al., 2001) (Table 3).

2.1.5 Mbd4:
Mbd4, also referred to as MED1 (Bellacosa, Cicchillitti et al., 1999), is the only known member of the MBD protein family not associated with HDAC activity (Hendrich & Bird, 1998). Instead, several lines of evidence suggest that Mbd4 plays a role in DNA repair (Bader, Walker et al., 1999, Bellacosa et al., 1999, Hendrich, Hardeland et al., 1999, Millar et al., 2002, Petronzelli, Riccio et al., 2000, and is expressed in ESCs as well as somatic tissues (Hendrich & Bird, 1998, Roloff, Ropers et al., 2003). DNA binding properties of Mbd3 seem to vary with species. While mammalian Mbd3 is unable to interact with methylated DNA, its amphibian counterpart binds methylated CpG dinucleotides in vitro and in vivo (Hendrich & Bird, 1998, Saito & Ishikawa, 2002, Wade et al., 1999). Sequence comparison of 5mC binding competent MBD domains revealed two highly conserved residues, which are altered in mammalian Mbd3: a largely solvent exposed tyrosine, as well as an amino-terminal lysine or arginine residue (Ohki, Shimotake et al., 1999, Saito & Ishikawa, 2002, Wakefield, Smith et al., 1999).
Riccio, Aaltonen et al., 1999). In addition to its MBD, Mbd4 contains a C-terminal catalytic domain that is highly homologous to bacterial DNA damage specific endonucleases that exhibit glycosylase activity during base excision repair (BER) (Hendrich & Bird, 1998, Michaels, Pham et al., 1990). Accordingly, Mbd4 was shown to remove thymine or uracil from mismatched CpG sites through glycosidic bond cleavage. As genomic G/T mismatches are the expected product of 5mCpG deamination, Mbd4 has been designated a methylation specific DNA repair enzyme (Hashimoto et al., 2012b, Hendrich et al., 1999, Petronzelli et al., 2000). Furthermore, Mbd4 has been implicated in DNA demethylation as it was shown in vitro to excise 5hmU, the deamination product of 5hmC (Hashimoto et al., 2012b).

Knockout and rescue experiments in embryonic stem cells, however, demonstrated that oxidation-dependent reactivation of methylated reporter genes is mediated by the action of thymine DNA glycosylase (TDG), but not by Mbd4 (Muller et al., 2014). Accordingly, deamination of 5hmC to 5hmU and subsequent excision by Mbd4 does not play a major role in ESCs (Hashimoto et al., 2012b). A contribution of Mbd4 to Tet-initiated DNA demethylation in NPCs, however, cannot be excluded, since Mbd4 was shown to bind to 5hmC at this developmental stage (Spruijt et al., 2013).

Mice lacking Mbd4 are viable and fertile (Millar et al., 2002). However, compared to wild type mice, Mbd4 knockout lead to a 3.3-fold higher number of C:G to T:A transitions at CpG sites (Millar et al., 2002). Moreover, Mbd4 −/− mice that were made heterozygous for the Min allele of the adenomatous polyposis coli gene (ApcMin), which pre-disposes mice to develop spontaneous intestinal neoplasia (Su, Kinzler et al., 1992), showed markedly reduced survival compared to Mbd4 +/+ controls. Accordingly, Mbd4 plays an important role in the repair of 5mC deamination at mCpGs. The relatively mild phenotype of Mbd4 knockout mice (Table 3), however, suggests that its absence might be compensated for by other glycosylases, such as TDG.

2.2 Kaiso protein family

Members of the Kaiso-like protein family (Figure 6) present a second class of proteins capable of binding specifically to methylated DNA (Filion et al., 2006). In contrast to members of the MBD protein family, Kaiso, Zbtb4 and Zbtb38 contain a conserved BTB/POZ (Bric-a-brac, tramtrack, broad complex/poxvirus and zinc finger) domain involved in protein-protein interactions and three Kruppel-like C2H2 zinc finger motifs, of which two were found essential for binding to methylated DNA (Filion et al., 2006). Similar to MBD proteins, members of the Kaiso family function as HDAC-dependent transcriptional repressors (Sasai, Matsuda et al., 2005). Several lines of evidence, however, including their variable binding modes, protein partners and expression patterns, suggest that Kaiso-like proteins have different biological functions (Daniel & Reynolds, 1999, Filion et al., 2006, Kiefer, Chatail-Hermitte et al., 2005, Park, Kim et al., 2005).

While Kaiso was shown to require at least two methylated CpG dinucleotides, a single mCpG proved sufficient for efficient binding of the Zbtb4 and Zbtb38 proteins (Filion et al., 2006, Prokhortchouk, Hendrich et al., 2001). Besides its ability to bind methylated DNA, in vitro synthesized Kaiso was shown to interact specifically with an unmethylated consensus sequence, the Kaiso Binding Site (KBS: TCCTGCNA), which can be found at promoters of Wnt target genes (Daniel & Reynolds, 1999, Park et al., 2005). Accordingly, the xWnt11 gene, a target of non-canonical Wnt signaling, was shown to be regulated by Kaiso in Xenopus (Daniel & Reynolds, 1999, Prokhortchouk et al., 2001). Moreover, Kaiso mediated repression of non-canonical and canonical Wnt targets was repressed by interactions with p120-catenin (Kim, Park et al., 2004), 2002). The ability to bind unmethylated KBS sequences is shared by Zbtb4, Zbtb38, however, was shown to interact with the E-box motif (CACCTG) of the rat tyrosine hydroxylase gene promoter (Kiefer et al., 2005), but failed to bind a labeled KBS probe (Filion et al., 2006).
More recently, Kaiso was found to bind 5hmC in NPCs and Zbtb4 was pulled down with hydroxymethylated DNA from brain tissue (Spruijt et al., 2013). The 5hmC binding domains, as well as the biological function, however, remain to be determined.

Kaiso-like proteins contain a BTB/POZ domain, which facilitates interaction with different sets of co-repressors and mediate transcriptional repression. Kaiso was shown to recruit the NCoR complex to promoters of target genes to introduce histone hypoacetylation, as well as H3K9 methylation (Yoon, Chan et al., 2003). Moreover, Kaiso was identified as component of an alternative MeCP1 complex in NIH3T3 cells (Prokhortchouk et al., 2001). Zbtb38 was found to interact with the co-repressors CtBPs (C-terminal Binding Proteins), which include histone deacetylase, methyltransferase and demethylase activities (Sasai et al., 2005, Zocchi & Sassone-Corsi, 2012). Zbtb4 was shown to associate with the Sin3A/histone deacetylase complex to repress expression of p21WAF1 in response to stimuli that activate p53 (Weber, Marquardt et al., 2008).

Kaiso-like proteins exhibit diverging expression patterns. While Kaiso is ubiquitously expressed, Zenon, the rat homologue of ZBTB38, is primarily transcribed in brain and neuroendocrine tissues (Kiefer et al., 2005). For Zbtb4, in contrast, high expression levels were identified in brain, lung, kidney, muscle and heart (Filion et al., 2006).

Kaiso-null mice (Table 3) are viable and fertile, with no detectable changes in gene expression profiles or developmental abnormalities. However, when crossed with tumor-susceptible Apc(Min/+) mice, Kaiso deficient animals showed resistance to intestinal cancer (Prokhortchouk et al., 2006).

2.3 SRA domain protein family

Recent studies implicate that yet another protein fold, the SRA (set and ring associated) domain could read DNA (hydroxy)methylation marks in vitro and in vivo (Frauer et al., 2011a, Johnson, Bostick et al., 2007, Spruijt et al., 2013, Unoki, Nishidate et al., 2004, Woo, Pontes et al., 2007). In mammals, two SRA domain-containing proteins (Figure 7), Np95 (mouse homologue of human ICBP90, gene name UHRF1) and Np97 (mouse homologue of human NIRF, gene name UHRF2), have been characterized (Unoki et al., 2004, Woo et al., 2007, Zhang, Gao et al., 2011). While Np95 was first discovered during the generation process of antibodies against murine thymic lymphoma (Fujimori, Matsuda et al., 1998), NIRF was identified through screenings for PCNP (PEST containing nuclear protein) interaction partners (Mori et al., 2002).

Besides the eponymous SRA domain, ICBP90 contains at least four additional functional motifs (Hashimoto, Horton et al., 2009): an N-terminal ubiquitin-like domain (Ubl, or NIRF_N); a tandem Tudor domain (TTD) that binds histone H3 tails di/tri-methylated at lysine 9 (H3K9me2/3) (Karagianni, Amazit et al., 2008, Papait, Pistore et al., 2007, Rajakumara, Wang et al., 2011, Wang, Shen et al., 2011). More recently, the SRA domain of ICBP90 was shown to target Dnmt1 to replicating pericentric heterochromatin for maintenance methylation (Arita, Ariyoshi et al., 2008, Avvakumov, Walker et al., 2008, Bostick et al., 2007, Hashimoto, Horton et al., 2008, Papait et al., 2008). In addition, ICBP90 was shown to bind histone H3K9me2/3 via its TTD, thus connecting repressive histone marks with DNA methylation (Nady, Lemak et al., 2011, Rothbart, Krajewski et al., 2012, Rottach, Fast et al., 2010). The PHD of ICBP90, on the other hand, was found associated with the N-terminal tail of histone H3 (Arita, Isogai et al., 2012, Cheng, Yang et al., 2013, Hu, Li et al., 2011, Papait, Pistore et al., 2007, Rajakumara, Wang et al., 2011, Wang, Shen et al., 2011). More recently, the SRA domain of Np95 was demonstrated to bind 5hmC and 5mC containing DNA substrates with similar affinity in vitro (Frauer et al., 2011a). Consistent with this, Np95 was identified as 5hmC reader in mESCs and NPCs. In mouse brain tissue, however, association with 5hmC remained undetected likely due to its low expression.
levels. Although the structure of NIRF, the second member of the SRA domain protein family, is closely related to ICBP90, both proteins possess significantly different expression patterns. While ICBP90 is mainly expressed in proliferating cells (Fujimori et al., 1998), NIRF protein levels increase during differentiation (Pichler, Wolf et al., 2011). NIRF binds hemi-methylated DNA and H3K9me2/3 containing heterochromatin marks in a cooperative manner, whereby localization and in vivo binding dynamics of NIRF, were shown to require an intact TTD and depend on H3K9me3 but not on DNA methylation (Pichler et al., 2011). While Np95 was shown to bind 5hmC in mESCs and NPCs, the interaction of Np97 and 5hmC was specific for NPCs. Furthermore, Np97 exhibited higher binding affinity for 5hmC than for 5mC in NPCs (Spruijt et al., 2013). Finally, Np97 was proposed to promote repetitive oxidation of 5mC by Tet proteins, since the levels of the oxidative cytosine derivatives 5hmC, 5fC and 5caC were increased upon coexpression of Np97 and Tet1 in HEK293T cells (Spruijt et al., 2013). Consequently, Spruijt et al., hypothesized that flipping of the modified base, as previously described for Np95, may enhance the accessibility of Tet enzymes to the hydroxymethylated base, whereby further oxidation is promoted (Spruijt et al., 2013).

Furthermore, ectopic Np97 was unable to rescue DNA methylation defects observed in Np95 -/- ESCs. Neither DNA methylation levels, nor pericentric heterochromatin localization of Dnmt1 in S-phase could be restored upon overexpression of Np97 arguing for functional differences between both proteins (Pichler et al., 2011). NIRF was found to interact with cell cycle proteins including cyclins, cyclin-dependent kinases (CDKs), retinoblastoma protein (pRB), p53, PCNA, HDAC1, DNMTs and G9a (Mori, Ikeda et al., 2012). It was shown to ubiquitinate cyclins D1 and E1, and to induce G1 arrest. Accordingly, NIRF was proposed to link the cell cycle regulatory network with the epigenetic landscape (Mori et al., 2012).

While knockout of Np95 leads to developmental arrest shortly after gastrulation and early gestational lethality (Sharif et al., 2007), the phenotype of Np97 null mice has not been analyzed (Li et al., 2013) (Table 3).

### 3 Role of 5mC writers, readers and modifiers in disease

Mutations in proteins involved in writing, reading and modifying the epigenetic landscape have been implicated in various severe human disorders. Due to their high sequence (Table 4) and functional similarity (Bostick et al., 2007, Filion et al., 2006, Hendrich & Bird, 1998, Ito et al., 2010, Kumar, Cheng et al., 1994, Mori et al., 2002, Qin, Leonhardt et al., 2011), we, hereafter, summarize the state-of-the-art regarding the role of the human orthologues of the aforementioned mouse Dnmts, Tets and MBPs in human diseases.

### Table 4: Comparison of human proteins and their mouse orthologues

<table>
<thead>
<tr>
<th>Mouse protein</th>
<th>Human protein</th>
<th>Amino acid similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dnmt1 (1620 aa)</td>
<td>DNMT1 (1632 aa)</td>
<td>76%</td>
</tr>
<tr>
<td>Dnmt2 (415 aa)</td>
<td>DNMT2 (391 aa)</td>
<td>77%</td>
</tr>
<tr>
<td>Dnmt3a (908 aa)</td>
<td>DNMT3A (912 aa)</td>
<td>96%</td>
</tr>
<tr>
<td>Dnmt3b (860 aa)</td>
<td>DNMT3B (853 aa)</td>
<td>80%</td>
</tr>
<tr>
<td>Dnmt3l (421 aa)</td>
<td>DNMT3L (387 aa)</td>
<td>56%</td>
</tr>
<tr>
<td>Tet1 (2039 aa)</td>
<td>TET1 (2136 aa)</td>
<td>50%</td>
</tr>
<tr>
<td>Tet2 (1912 aa)</td>
<td>TET2 (2002 aa)</td>
<td>55%</td>
</tr>
<tr>
<td>Tet3 (1803 aa)</td>
<td>TET3 (1795 aa)</td>
<td>89%</td>
</tr>
<tr>
<td>Mcf2 (501 aa)</td>
<td>MECP2 (498 aa)</td>
<td>94%</td>
</tr>
<tr>
<td>Mbd1 (636 aa)</td>
<td>MBD1 (605 aa)</td>
<td>68%</td>
</tr>
<tr>
<td>Mbd2 (414 aa)</td>
<td>MBD2 (411 aa)</td>
<td>94%</td>
</tr>
<tr>
<td>Mbd3 (285 aa)</td>
<td>MBD3 (291 aa)</td>
<td>92%</td>
</tr>
<tr>
<td>Mbd4 (554 aa)</td>
<td>MBD4 (580 aa)</td>
<td>58%</td>
</tr>
<tr>
<td>Kaiso (671 aa)</td>
<td>KAISO (672 aa)</td>
<td>84%</td>
</tr>
<tr>
<td>Zbtb4 (982 aa)</td>
<td>ZBTB4 (1013 aa)</td>
<td>85%</td>
</tr>
<tr>
<td>Zbtb3 (1197 aa)</td>
<td>ZBTB38 (1195 aa)</td>
<td>81%</td>
</tr>
<tr>
<td>Np95 (782 aa)</td>
<td>ICBP90 (806 aa)</td>
<td>72%</td>
</tr>
<tr>
<td>Np97 (803 aa)</td>
<td>NIRF (802 aa)</td>
<td>90%</td>
</tr>
</tbody>
</table>

### 3.1 DNMT proteins in disease

Since Dnmt1 knockout is embryonic lethal in mice, it is unlikely to expect a human disease linked to a DNMT1 catalytic domain mutation. But mutations in the regulatory domain of DNMT1 were found (Table 5). Mutations in the TS domain of DNMT1 cause neurodegeneration like hereditary sensory autonomic neuropathy with dementia and hearing loss (HSAN1E) (Klein, Botuyan et al., 2011) and autosomal dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN) (Winkelmann, Lin et al., 2012). Mutations of Y495C, Y495H, D490E-P491Y (Klein, Botuyan et al., 2011) and autosomal dominant neuromyopathy with dementia and hearing loss (HSAN1E) (Klein, Botuyan et al., 2011) were also reported (Lamprecht et al., 2011). Mutations in DNMT3A were found in AML patients, associated with poor survival (Table 5) (Lamprecht, Larson et al., 2011). The most frequent mutation occurred in amino acid R882, however frameshift, nonsense and splice site mutations were also reported (Lamprecht et al., 2011). Mutations of DNMT3A are not only observed in AML patients, but also in myelodysplastic syndrome (MDS). Similar to mutations leading to AML, amino acid R882 located in the methyltransferase domain of DNMT3A is the most frequent mutation occurred in amino acid R882, however frameshift, nonsense and splice site mutations were also reported (Lamprecht et al., 2011). Mutations in DNMT3A are not only observed in AML patients, but also in myelodysplastic syndrome (MDS). Similar to mutations leading to AML, amino acid R882 located in the methyltransferase domain of DNMT3A is the most common mutation site (Walter, Ding et al., 2011). Unlike in AML and MDS, most mutations in overgrowth syndrome do not directly affect the catalytic activity of DNMT3A, but interfere with domain-domain interactions and histone binding, which further affect the activity of DNMT3A (Tatton-Brown, Seal et al., 2014).

ICF syndrome (immunodeficiency, chromosomal instability and facial anomalies), a human genetic disorder is caused by DNMT3B mutations (Table 5)
General introduction

Several mutations were identified and most mutations are located in the catalytic domain of DNMT3B and directly affect the activity of DNMT3B (Xu et al., 1999). However, mutations which do not directly affect its catalytic activity were also observed in ICF syndrome. Two mutations, A766P and R840Q displayed similar methylation activity than the wild-type enzyme but lost the ability to interact with DNMT3L, which further leads to loss of activity in vivo (Xie, Huang et al., 2006). Direct or indirect loss of DNMT3B activity consequently decreased satellite DNA methylation in ICF syndrome patients, indicating that DNMT3B is involved in maintaining genome stability.

5mC, the product of DNMTs is related to tumorigenesis. It was shown that the genome of cancer cells is globally hypomethylated relative to their normal counterparts. Usually, hypomethylation leads to gene activation. In cancer cells, the activation of genes is caused by hypomethylation of nearby CpG islands, which are silenced in somatic tissues by DNA methylation (Strichman-almashanu, Lee et al., 2002). Satellite sequences and repetitive sequences such as LINE1, SINE, IAP and Alu elements are silenced mainly by DNA methylation in normal cells. However, in tumor cells, hypomethylation of L1 promoter was detected and the activation of L1 might promote chromosomal rearrangements and genome instability (Suter, Martin et al., 2004). Although the cancer genome is hypomethylated, several studies showed that Dnmts are upregulated in cancer cells (Ahluwalia, Hurteau et al., 2001, Lin, Hsu et al., 2007, Roll, Rivenbark et al., 2008), suggesting that demethylation enzymes might be additionally involved in loss of DNA methylation in cancer.

3.2 TET proteins in disease

Mixed lineage leukemia (MLL) gene is located in 11q23 and is the most frequent cytogenetic finding in acute myeloid leukemia (AML). In AML, MLL is translocated to chromosome 10 as a fusion with the TET1 gene. The MLL-TET1 fusion protein contains the AT hooks, subnuclear localization domains, and the CXXC domain of MLL and the C-terminus of TET1 (Table 5) (Lorsbach et al., 2003). The function of MLL-TET1 fusion protein is still unknown, but it was showed that TET1 is involved in MLL-rearranged leukemia. TET1 is a direct target of the MLL-fusion protein and is significantly upregulated in MLL-rearranged leukemia, leading to a global increase of 5hmC, thus playing an oncogenic role (Huang, Jiang et al., 2013).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Disease</th>
<th>Alteration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNMT1</td>
<td>Hereditary sensory autonomic neuropathy with dementia and hearing loss (HSAN1E)</td>
<td>Y495C, Y495H, D490E-P491Y</td>
<td>(Klein et al., 2013, Klein et al., 2011)</td>
</tr>
<tr>
<td>DNMT1</td>
<td>Autosomal dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN)</td>
<td>A570V, G605A and V606F</td>
<td>(Winkelmann et al., 2012)</td>
</tr>
<tr>
<td>DNMT3A</td>
<td>Acute myeloid leukemia (AML) myelodysplastic syndrome (MDS)</td>
<td>R882 and frameshift, nonsense and splice site mutations</td>
<td>(Lamprecht et al., 2011, Walter et al., 2011)</td>
</tr>
<tr>
<td>DNMT3A</td>
<td>Overgrowth syndrome</td>
<td>Mutations interfere with domain-domain interactions and histone binding</td>
<td>(Tatton-Brown et al., 2014)</td>
</tr>
<tr>
<td>DNMT3B</td>
<td>Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF) syndrome</td>
<td>Mutations in catalytic domain</td>
<td>(Hansen et al., 1999, Xu et al., 1999)</td>
</tr>
<tr>
<td>TET1</td>
<td>AML</td>
<td>Ten-eleven translocation that gives rise to a MLL-TET1 fusion</td>
<td>(Lorsbach et al., 2003)</td>
</tr>
<tr>
<td>TET2</td>
<td>AML, MDS and myeloproliferative neoplasms (MPN)</td>
<td>Mutations mostly in catalytic domain</td>
<td>(Abdel-Wahab, Mullally et al., 2009)</td>
</tr>
</tbody>
</table>

In myeloproliferative neoplasms (MPNs), mutations of TET2 but not TET1 and TET3 were observed (Table 5) (Abdel-Wahab et al., 2009). Mutations of TET2 were also observed in AML with varied frequency and most of them occurred in the catalytic domain of TET2. In AML, TET2 mutations correlate with genomic 5hmC level (Konstandin et al., 2011). TET2 is one of the most frequently mutated genes in myelodysplastic syndrome (MDS). Mutations of TET2 were detected in most of the bone marrow cells in MDS and these mutations contribute to the malignant transformation of bone marrow cells (Langemeijer, Kuiper et al., 2009), which consequently displayed uniformly low levels of 5hmC in genomic DNA compared to bone marrow samples from healthy controls (Ko et al., 2010). Besides the hematopoietic malignancies, 5hmC levels are also changed in solid tumors. 5hmC level were profoundly reduced in glioma, colon cancer, breast cancer and melanoma compared to normal tissues (Haffner, Chaux et al., 2011, Jin, Wu et al., 2011, Kraus, Globisch et al., 2012, Li & Liu, 2011, Xu, Yang et al., 2011a).
Modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in disease

Unlike in cancer, in the hippocampus/parahippocampal gyrus (HPG) of preclinical and later-stage Alzheimer’s disease patients, significantly increased levels of TET1, 5mC and 5hmC were observed. In contrast, levels of 5fC and 5caC were significantly decreased in the HPG of these patients (Bradley-Whitman & Lovell, 2013). This indicates that DNA methylation might play an important role in memory-related disease.

3.3 MBP proteins in disease

As readers and translators of epigenetic information, alterations in MBP protein sequences affect the precisely coordinated link between DNA methylation, histone modification and higher order chromatin structure.

Mutations in the X-linked *MECP2* gene give rise to Rett syndrome (RTT) (Table 6), a late onset (6-18 months post birth) debilitating neurological disease that affects 1 in 10,000 to 15,000 female live births (Amir, Van den Veyver et al., 1999, Hagberg, Aicardi et al., 1983). After a period of normal development (6-18 months), RTT patients usually lose speech and acquired motor skills (Hagberg et al., 1983). They are afflicted with seizures, autism, loss of motor coordination, abnormal breathing and develop stereotypical, repetitive hand movements (Hagberg et al., 1983). After the initial regression, however, conditions often stabilize and allow viability until adulthood (Hagberg et al., 1983, Rett, 1966).

### Table 6: Summary of disease related MBP alterations

<table>
<thead>
<tr>
<th>Protein</th>
<th>Disease</th>
<th>Alteration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBD1</td>
<td>Prostate cancer</td>
<td>Upregulated</td>
<td>(Patra, Patra et al., 2003)</td>
</tr>
<tr>
<td>MBD2</td>
<td>Breast Cancer</td>
<td>Upregulated</td>
<td>(Billard, Magdinier et al., 2002)</td>
</tr>
<tr>
<td>MBD3</td>
<td>Glioblastoma</td>
<td>Upregulated</td>
<td>(Schlegel, Guneyisu et al., 2002)</td>
</tr>
<tr>
<td>MBD4</td>
<td>Colorectal cancer</td>
<td>Frameshift mutation</td>
<td>(Riccio et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>Endometrial cancer</td>
<td>Frameshift mutation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pancreas cancer</td>
<td>Frameshift mutation</td>
<td></td>
</tr>
<tr>
<td>KAI SO</td>
<td>Colorectal cancer</td>
<td>Upregulated</td>
<td>(Lopes, Valls et al., 2008)</td>
</tr>
<tr>
<td>ZBTB4</td>
<td>Neuroblastoma</td>
<td>Downregulated</td>
<td>(Weber et al., 2008)</td>
</tr>
<tr>
<td>ICBP90</td>
<td>Nonsmall lung cancer</td>
<td>Upregulated</td>
<td>(Daskalos, Oleksiewicz et al., 2011)</td>
</tr>
<tr>
<td>NIRF</td>
<td>Lung cancer</td>
<td>Upregulated</td>
<td>(He, Duan et al., 2009)</td>
</tr>
</tbody>
</table>

Although the first patients were described in 1966 by Andreas Rett (Rett, 1966), more than 30 years passed before mutations within the *MECP2* gene located in Xq28 were identified as the cause of the neurological disorder (Amir et al., 1999). The most frequent mutations observed in patients suffering from RTT are missense mutations that cluster within the MBD (aa 78-162), as well as nonsense mutations primarily found within the TRD (aa 207-310) (Christodoulou, Grimm et al., 2003). In *Xenopus*, missense mutations R106W, R133C, F155S and T158M were shown to reduce the binding ability of Mecp2 to methylated DNA (Ballestar, Yusufzai et al., 2000). Studies in mouse cells showed that the majority of MBD-related missense mutations affected the heterochromatin binding and/or clustering ability of Mecp2 (Agarwal et al., 2011). By artificially targeting chromatin binding deficient Rett mutants (R111G, R133L and F155S) to constitutive heterochromatic regions, however, Casas-Delucchi et al. revealed that some of these mutations exclusively affect the chromatin binding but not linking ability (Casas-Delucchi, Becker et al., 2012). Mutations within the TRD have been shown to influence protein-protein interactions. In knock-in mice bearing the common RTT mutation R306C, neuronal activity fails to induce T308 phosphorylation, a posttranslational modification required to suppress the interaction of Mecp2 with the co-repressor complex NCoR. Accordingly, R306C mutations result in persistent association of both proteins leading to decreased induction of a subset of activity-related genes (Ebert, Gabel et al., 2013, Lyst, Ekiert et al., 2013). In addition to missense and nonsense mutations, reading frame shifts and C-term deletions were shown to give rise to RTT. Mice bearing a truncating mutation similar to those found in RTT patients showed normally localized Mecp2 proteins (Shahbazian, Young et al., 2002). Histone H3, however, was hyperacetylated indicating abnormal chromatin architecture and misregulated gene expression (Shahbazian et al., 2002). Moreover, Muotri and colleagues identified increased susceptibility for L1 transposition and genome instability in RTT patients with truncating mutations (Muotri et al., 2010).
In addition to Rett syndrome, Mecp2 was implicated in other neurological diseases, including Hirschsprung’s disease, autism spectrum disorder, schizophrenia, Prader-Willi and Angelman syndrome (Carney, Wolpert et al., 2003, Loat, Curran et al., 2008, Nagarajan, Hogart et al., 2006, Ramocki, Peters et al., 2009, Shibayama, Cook et al., 2004, Zhou, Qin et al., 2013).

More recently MBP proteins have been associated with several types of human cancers (Table 6). While Mecp2 was overexpressed in oestrogen receptor positive human breast cancer (Muller, Fiegl et al., 2003), MBD1 mRNA and protein levels were increased in prostate cancer (Patra et al., 2003). Accordingly, Patra and co-workers proposed MBD1 as the major cause of hypermethylated chromatin regions in prostate cancer through the recruitment of HDAC1/2 and subsequent histone deacetylation (Patra et al., 2003). MBD2 mRNA level was shown to be significantly elevated in benign tumors of the breast and correlated with tumor size of invasive ductal carcinomas, the most common type of breast cancer (Billard et al., 2002). Accordingly, upregulation of MBD2 was proposed to be associated with breast cell proliferation (Billard et al., 2002). Increased expression of MBD3 and MBD4 were associated with malignant glioma of the brain, and the grade of malignancy correlated with MBD3/4 expression level (Schlegel et al., 2002). Furthermore, frameshift mutations of MBD4 have been identified in colorectal, endometrial and pancreatic cancer with microsatellite instability (Riccio et al., 1999). MBD4 mutations consisted of 1 to 2-bp deletions or 1-bp insertions that caused frameshifts and premature stop codons. The resultant truncated MBD4 proteins were predicted to be non-functional, as they lack the C-terminal catalytic domain, whereby genomic instability was proposed to steadily increase (Riccio et al., 1999). As a regulator of target genes of the canonical and noncanonical Wnt pathway, Kaiso was shown to mediate silencing of tumor suppressor genes CDKN2A and HIC1 in Wnt-driven human colon cancer cell lines (Lopes et al., 2008). Kaiso depletion induced expression of tumor suppressor genes without altering DNA methylation levels (Lopes et al., 2008). As a result, colon cancer cells became susceptible to cell cycle arrest and cell death induced by chemotherapy (Lopes et al., 2008). Accordingly, Lopes and colleagues suggested Kaiso as a methylation-dependent oncogene that represses hypermethylated tumor suppressor genes (Lopes et al., 2008). ZBTB4 expression was shown to be downregulated in advanced stages of human neuroblastoma and multiple human solid tumors (Weber et al., 2008). As a repressor of the $P21^{CDPI}$ gene, an inhibitor of the Cdk2 kinase, ZBTB4 usually blocks cell cycle arrest in response to p53 activation (Weber et al., 2008). Consequently, loss of ZBTB4 inhibits apoptosis and favors long-term survival of affected cells (Weber et al., 2008). In tumors, where many promoter-associated CpG islands are hypermethylated, maintenance of methylation plays a major role. Accordingly, elevated levels of ICBP90 were shown to control cell cycle through maintenance of promoter methylation at CDK2A and RASSF1 in non-small cell lung cancer (Daskalos et al., 2011). Finally, decreased expression of let-7a miRNA in lung cancer was shown to result in elevated NIRF and reduced $P21^{CDPI}$ protein level, thereby most likely contributing to lung carcinogenesis (He et al., 2009).

4 Concluding remarks

In summary, alterations in 5mC writers, readers and modifiers that affect their level, posttranslational modifications, ability to bind and/or modify DNA and protein interactions are each and all potential mechanisms contributing to altered chromatin composition and structure as well as genome activity and stability (Figure 8) and contribute to an overwhelming variety of human diseases. Despite intensive research, genotype-phenotype connections have been generally difficult to establish and subsequent studies are urgently needed to elucidate potential strategies for diagnostic and therapeutic applications.

Figure 8 Writing, reading and translating DNA modifications.
Graphical summary of how DNA modification writers, readers and translators can impact on chromatin composition, structure (nucleosomes are represented as balls, DNA as line) as well as genome expression (arrow represents active promoters) and stability.
References


General introduction


Konecny K, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg

General introduction

Lamprecht T, Larson DE, Ph D, Kandoh C, Jacqueline E, Baty J, Welch H, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koholdt DC, Sch生活在浙江 unfinished.
Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. Epigenetics 1: e1-11


Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99: 247-57


null
Modifiers and readers of DNA modifications and their impact on genome structure, expression and stability in disease


mammalian oocyte survival and reprogramming by activation of TET proteins. Science 342: 1518-21
Zocchi L, Sassone-Corsi P (2012) SIRT1-mediated deacetylation of MeCP2 contributes to BDNF expression. Epigenetics 7: 695-700
Aim of this work

To understand how the 5-methylcytosine (5mC) modifier Ten-eleven translocation 1 (Tet1) oxidizes 5mC to 5-hydroxymethylcytosine (5hmC) and how this process is regulated by 5-methylcytosine readers MBD (Methyl-CpG binding domain) proteins, we designed and performed experiments to answer the following questions:

1. What is the biological consequence of Mecp2 poly(ADP-ribosyl)ation on chromatin structure? (Chapter 1)
2. How does Tet1 oxidize 5-methylcytosine to 5-hydroxymethylation and how to detect this process? (in preparation)
3. Whether and how do MBD proteins regulate Tet1 mediated 5mC to 5hmC conversion? (in revision)
4. Whether and how do Mbd2, Mecp2 and Tet1 regulate transcriptional noise? (in revision)
5. Whether and how do Mbd2, Mecp2 and Tet1 regulate retrotransposition of human long interspersed nuclear elements 1? (in revision)
Poly(ADP-ribosyl)ation of methyl CpG binding protein 2 regulates chromatin structure

Annette Becker¹, Peng Zhang¹#, Lena Allmann¹, Daniela Meilinger², Bianca Bertulat¹, Daniel Eck¹, Maria Hofstaetter³, Giody Bartolomei⁴, Michael Hottiger⁴, Valérie Schreiber⁵, Heinrich Leonhardt², M. Cristina Cardoso¹*¹

¹Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany
²Center for Integrated Protein Science at the Department of Biology, Ludwig Maximilians University Munich, 82152 Planegg-Martinsried, Germany
³Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
⁴Institute of Veterinary Biochemistry and Molecular Biology, University of Zurich, 8057 Zurich, Switzerland
⁵UMR7242 Biotechnology and Cell Signalling, Laboratory of Excellence Medalis, Strasbourg University, CNRS, ESBS, BP10413, 67412 Illkirch Cedex, France


# Author contributions:
P.Z. performed, analyzed and prepared figures for the FRAP and in situ extraction experiments (Fig 3F, Fig 3G) and wrote the corresponding methods and figure legends.
Abstract
The epigenetic information encoded in the genomic DNA methylation pattern is translated by methyl-cytosine binding proteins like Mecp2 into chromatin topology and structure and gene activity states. We have previously shown that Mecp2 level increases during differentiation and causes large-scale chromatin reorganization, which is disturbed by MECP2 Rett syndrome mutations. Phosphorylation and other post-translational modifications of Mecp2 have been recently described to modulate its function. Here, we show poly(ADP-ribosyl)ation of endogenous Mecp2 in mouse brain tissue. Consequently, we find that Mecp2 induced aggregation of pericentric heterochromatin and its chromatin accumulation was enhanced in Parp-1<sup>−/−</sup> compared to wild-type cells. We mapped the poly(ADP-ribosyl)ation domains and engineered Mecp2 mutation constructs to further analyze potential effects on DNA binding affinity and chromatin large scale remodeling. Single or double deletion of the poly(ADP-ribosyl)ated regions as well as PARP inhibition increased the heterochromatin clustering ability of Mecp2. Increased chromatin clustering may reflect increased binding affinity. In agreement with this hypothesis, we found that Parp-1 deficiency significantly increased chromatin binding affinity of Mecp2 in vivo. These data provide novel mechanistic insights into the regulation of Mecp2 mediated higher-order chromatin architecture and suggest therapeutic opportunities to manipulate Mecp2 function.

Keywords: chromatin binding and organization, heterochromatin, DNA methylation, Mecp2, Parp-1, poly(ADP-ribosyl)ation

In mouse brain (background), endogenous Mecp2 (green sphere) is poly(ADP-ribosyl)ated (green sphere with branches) and this reduces Mecp2 ability to bind and cluster heterochromatin.
Introduction

In mammals, methylation of cytosine residues at the dinucleotide CpG is essential for development and is proposed to regulate genome organization and expression. This epigenetic information is recognized and translated by a family of chromatin organizing proteins containing a conserved methyl CpG binding domain (MBD) (Free, Wakefield et al., 2001, Nan, Meehan et al., 1993). MeCP2, the founding member of the MBD protein family, has been described to function as a transcriptional silencer through association with corepressor complexes mediated by its transcriptional repression domain (TRD) (Jones, Veenstra et al., 1998, Nan, Campoy et al., 1997, Nan, Ng et al., 1998). Increased expression of Mecp2 in mouse cells induces aggregation of pericentric heterochromatin in a dose dependent manner and the clustering ability is mostly dependent on the MBD (Agarwal, Hardt et al., 2007, Bertulat, De Bonis et al., 2012, Brero, Easwaran et al., 2005). Besides, purified Mecp2 has also been shown to cause compaction of nucleosomal arrays in vitro (Georgel, Horowitz-Scherer et al., 2003).

Mutations within the X-chromosome located MECP2 gene have been linked to one of the most common human mental retardation disorders in females, termed Rett syndrome (RTT, OMIM #312750) (Amir, Van den Veyver et al., 1999). Whereas missense mutations are mostly accumulated within the MBD (amino acids 78-162), the majority of nonsense mutations occur predominantly within the TRD (amino acids 207-310) (http://mecp2.chw.edu.au/mecp2/). Mecp2 RTT associated mutations have been shown to affect the ability of Mecp2 to bind DNA and to compact poly nucleosomal arrays in vitro (Georgel et al., 2003, Nikitina, Ghosh et al., 2007) as well as MeCP2 chromatin binding kinetics in vivo (Agarwal, Becker et al., 2011, Kumar, Kamboj et al., 2008, Marchi, Guarda et al., 2007). In addition, we recently identified Mecp2 mutants with decreased ability to accumulate at pericentric heterochromatin and / or decreased heterochromatin clustering potential (Agarwal et al., 2011, Casas-Delucchi, Becker et al., 2012). In spite of accumulating evidence in favor of a major role of MeCP2 in controlling large-scale heterochromatin organization, the underlying mechanism and its regulation have so far not been elucidated.

In this study, we found that endogenous Mecp2 from mouse brain tissue is poly(ADP-ribo)sylated in vivo. We identified two distinct Mecp2 domains, relevant for poly(ADP-ribo)sylation and could show that deletion of these modifiable domains increased heterochromatin clustering. Furthermore, we found that Parp-1 deficiency increases Mecp2’s ability to aggregate as well as to bind to pericentric heterochromatin. These findings unravel a novel mechanism modulating Mecp2 dependent chromatin organization.

Results

Endogenous Mecp2 from mouse brain tissue is poly(ADP-ribo)sylated in vivo

Over the last years, several post-translational modifications have been described for Mecp2. Among them, phosphorylation of Mecp2 has been implicated to affect Mecp2 chromatin binding and neurological functions (Bellini, Pavesi et al., 2014, Chen, Chang et al., 2003, Tao, Hu et al., 2009, Zhou, Hong et al., 2006). Prompted by the recently reported poly(ADP-ribo)sylation of MECP2 in U2OS cells (Jungmichel, Rosenthal et al., 2013), we addressed whether endogenous Mecp2 from mouse brain tissue was poly(ADP-ribo)sylated. For that, we incubated boronic acid beads, specifically enriching ribonucleotides, with lysates of brain tissue. SDS-PAGE followed by Western blotting using anti-Mecp2 antibody showed that Mecp2 got enriched by boronic acid from mouse brain extracts, illustrating that Mecp2 is modified by ribonucleotides in mouse brain tissue (Figure 1A). To more specifically investigate poly(ADP-ribo)sylation of endogenous Mecp2 from brain tissue, we performed immunoprecipitation assays of brain extracts with either anti-Mecp2 antibody or control rabbit IgG. Immunoblot analysis with anti-poly(ADP-ribose) antibody showed specific poly(ADP-ribo)sylation of endogenous Mecp2 (Figure 1B). In addition, we observed poly(ADP-ribo)sylation of endogenous Mecp2 enriched from mouse brain extracts using Tris-NTA coupled beads, specifically recognizing the naturally occurring histidine (His) repeat present within the COOH-terminal domain (aa366-aa372) of Mecp2 (Figure 1C). We further observed poly(ADP-ribo)sylation of ectopically expressed Mecp2-GFP but not of GFP alone (Figure 1D).

To determine whether Parp-1 is responsible for Mecp2 poly(ADP-ribo)sylation, we performed in vitro poly(ADP-ribo)sylation analysis using recombinant Mecp2-GFP in the presence of strep-Parp-1 and [α-32P]NAD+ (Figure 1E). Whereas GFP did not get modified, Parp-1 specifically poly(ADP-ribo)sylated Mecp2, which is in agreement with recently published in vitro poly(ADP-ribo)sylation of Mecp2 (Jungmichel et al., 2013).

Subsequent mapping identified the domain spanning ID (inter-domain) and TRD to be strongly poly(ADP-ribo)sylated in vivo (Figure 1F). Interestingly, the NH2-terminus plus MBD as well as the COOH-terminus showed almost no poly(ADP-ribo)sylation. We could further narrow down the modified domain to the ID (amino acids 163 to 206; poly(ADP-ribo)sylated domain 1) and to less extent to amino acids 244 to 275 (poly(ADP-ribo)sylated domain 2) (Figure 1F).

Next, we tested deletion constructs lacking the poly(ADP-ribo)sylated regions (Figure 1G). While deletion of poly(ADP-ribo)sylated domain 2 (Mecp2G.17; deletion of aa 244 to 275) resulted in slightly less poly(ADP-ribo)sylation than the full-length, the construct lacking the poly(ADP-ribo)sylated domain 1 (Mecp2G.16; deletion of aa 163 to 206) showed a strong decrease and the double
deletion (Mecp2G.18) had an even stronger effect (Figure 1G).

Given that MECP2 is modified by PARP-1 in vitro (Figure 1 and (Jungmichel et al., 2013)), we next tested whether the absence of Parp-1 in Parp-1−/− mouse fibroblasts might result in a lack of modification with functional consequence on Mecp2’s ability to reorganize heterochromatin. For that we compared the numbers of heterochromatic centers in Parp-1−/− mouse fibroblasts expressing either Mecp2-GFP or GFP alone to the wild-type cells (Figure 2B). Interestingly, we could measure enhanced aggregation of chromocenters in Mecp2-GFP expressing Parp-1−/− mouse embryonic fibroblasts, concomitant with reduced poly(ADP-ribosyl)ation levels (Figure 2A and 2B). We next compared the median numbers of chromocenters in mouse myoblast cells expressing either Mecp2-GFP or one of the deletion constructs lacking the poly(ADP-ribosyl)ated regions (Figure 2C). These adult stem cells express very low to undetectable levels of endogenous Mecp2 (Brero et al., 2005). First, to rule out that the level of GFP-tagged Mecp2 obtained through ectopic expression in myoblast cells is above the physiological level of Mecp2 in brain neurons, we performed quantitative western blotting in combination with fluorescence imaging. Using recombinant GFP-tagged Mecp2 as a direct calibration standard for the western blot analysis, we could determine the average amount of GFP-tagged Mecp2 in mouse myoblasts to vary between 1.3 – 2 pg per cell (Figure 2F). These amounts are in the range of endogenous physiological Mecp2 level per mouse neuronal cell nucleus (Skene, Illingworth et al., 2010). The numbers of chromocenters in cells expressing the poly(ADP-ribosyl)ated domain 1 deletion (Mecp2G.16) or the double deletion (Mecp2G.18) was significantly reduced compared to cells expressing wild-type Mecp2-GFP (Figure 2C). The deletion of poly(ADP-ribosyl)ated domain 2 (Mecp2G.17) though had a milder effect. These results correlate well with the poly(ADP-ribosyl)ation level of the respective constructs (Figure 1G). We could exclude major conformational changes caused by these deletions since all mutant proteins localized at chromocenters as the wild-type protein (Figure 2D). To further validate that the increase of chromocenter clustering was based on reduced poly(ADP-ribosyl)ation levels and was not simply due to deletion of amino acids within Mecp2, we treated the cells with the PARP inhibitor 3-amino-benzamide (3AB). As the chromocenter numbers of GFP expressing cells treated...
with 3AB were comparable to DMSO treated cells, we concluded that the inhibitors themselves did not have a significant effect on chromocenter aggregation (Figure 2E, left side). In stark contrast, MeCP2-GFP expressing cells incubated with the PARP inhibitor exhibited significantly increased clustering of pericentric heterochromatin relative to the DMSO control (Figure 2E, right side).

Figure 2 Poly(ADP-ribosyl)ation counteracts clustering of pericentric heterochromatin.
(A) MeCP2 exhibits decreased poly(ADP-ribosyl)ation in Parp-1−/− mouse embryonic fibroblast (MEF) cells compared to wild-type (WT) cells. GFP-fused MeCP2 was expressed in WT as well as in Parp-1−/− MEF cells. After immunoprecipitation using the GFP-Trap, poly(ADP-ribosyl)ation of precipitated proteins was checked via Western blotting with anti-PAR followed by anti-GFP antibodies. (B) Parp-1−/− MEF cells exhibit enhanced MeCP2 induced chromocenter aggregation in comparison to WT cells. WT or Parp-1−/− MEF cells were transfected with an expression vector coding for GFP or MeCP2-GFP. Z-stacks of images were recorded from nuclei with similar high expression levels of the protein using constant image acquisition parameters. Experiments were repeated two times with at least 30 cells per construct each time and are shown normalized to the control GFP expressing cells. (C) Pmi28 mouse myoblasts were seeded on coverslips and transfected with an expression vector coding for GFP-fused MeCP2 or GFP-fused MeCP2 deletions lacking the poly(ADP-ribosyl)ated domains. G stands for GFP. Using constant image acquisition parameters, Z-stacks of images were recorded of nuclei with similarly high expression levels of the GFP-tagged protein. Graphs show median numbers of chromocenters of cells expressing the indicated proteins. Error bars represent 95% Confidence Interval (C.I.). Experiments were repeated two times with at least 30 cells per construct analyzed each time. Asterisks represent statistically significant difference: * for p<0.05; ** for p<0.001. Cells expressing MeCP2G or GFP-fused MeCP2 deletions lacking the poly(ADP-ribosyl)ated domains. G stands for GFP. (D) Overview of the subcellular localization of MeCP2 deletions. Pmi28 myoblast cells were transfected with plasmids coding for MeCP2 deletions fused to GFP (G) as indicated. After fixation, DNA was counterstained with DAPI to highlight chromocenters (CC). PC: phase contrast. Scale bar: 5 µm. (E) Cells were transfected with vectors as indicated and treated with the PARP inhibitor 3AB (10 mM) or DMSO control for about 15 hours. (F) Estimation of the amount of GFP-tagged MeCP2 in transfected mouse myoblast cells. Mouse myoblast cells were transfected with plasmids coding for GFP-tagged wild type MeCP2 (MeCP2G). GFP positive cells were isolated and counted using flow cytometry. In parallel, recombinant MeCP2G was purified (using GFP-Trap beads) and its concentration was determined with a BSA calibration standard using CBB staining (left and middle histogram panels). Defined amounts of purified recombinant MeCP2G and MeCP2G expressed in defined numbers of sorted cells were analyzed by SDS-PAGE and quantitative western blotting with an anti-GFP antibody using fluorescence imaging (right histogram panel). The graphs depict the signal intensity (above background) assessed by Image J quantitation and presented as arbitrary units (AU). All signal intensities were within linear detection range.

Mecp2 binding to pericentric heterochromatin is elevated in Parp-1−/− compared to wild-type cells

As the chromatin aggregation potential of MeCP2 was increased in Parp-1−/− mouse fibroblasts in comparison to the wild-type fibroblasts (Figure 2B), we further tested whether the absence of Parp-1 and its mediated post-translational modification in those cells might have a functional consequence on MeCP2’s ability to bind heterochromatin.
Figure 3 The chromatin binding ability of Mecp2 is elevated in Parp-1−/− cells.
(A-D) Image analysis protocol for quantification of chromocenter accumulation. All steps necessary for analysis were performed using Velocity 5.5 built-in functions, such as image processing, object segmentation and measurements. Z-stacks of cells transfected with a GFP fusion protein were cropped to obtain single nuclei per image. Thresholds for segmentation were set for each nucleus individually. Chromocenter and nucleoplasm were segmented utilizing the built-in thresholding function, which uses the percentage of the overall image intensity. (A) Nucleus of a Mecp2-GFP expressing cell before segmentation. Lower right panel shows the nucleus in X-Y axis, upper right and lower left panel shows the nucleus in X-Z and Y-Z axis, respectively. Scale bar = 5 µm. (B) Histogram of the frequency of fluorescence intensities plotted on a logarithmic scale. The threshold for chromocenter (CC) and nucleoplasm (NP) segmentation is displayed in red and blue, respectively. Care was taken to exclude chromocenters from NP segmentation and vice versa. (C) Visualization of chromocenter (red) and nucleoplasm (blue) segmentation according to the thresholds displayed in B. (E) Wild-type (WT) or Parp-1−/− MEF cells were transfected with expression vectors coding GFP- or YFP-tagged Mecp2 constructs as indicated. Z-stacks of images were taken from cells expressing comparable levels of the GFP-fused construct. Experiments were repeated at least two times with as many as 30 cells analyzed per construct each time. The graphs show the accumulation of the Mecp2 constructs at heterochromatin in Parp-1−/− cells normalized to wild-type mouse fibroblasts. Asterisks represent statistically significant difference: * for p<0.05; ** for p<0.001. Error bars represent 95% Confidence Interval (C.I.). (F) FRAP analysis of GFP-tagged Mecp2 full length and mutant proteins. Upper row shows an example of Mecp2G protein recovery after photobleaching. Circles indicate the bleached region. Scale bar = 5 µm. Middle row shows the fluorescence recovery curve for each construct. The experiment was repeated two times with 7-15 cells each time for analysis. Results were averaged and the mean value was plotted. Lower row shows half life (t) of each construct and p value for WT and Parp-1−/− cell line are calculated using t.test. The cell number and median are indicated beside the boxplot. (G) In situ extraction of GFP-tagged Mecp2 full length and mutant proteins. Upper row shows representative images of Mecp2G and Mecp2Y.1 extraction after 0.5% Triton X-100 treatment. Scale bar = 5 µm. Lower row shows the normalized fluorescence intensity before and after treatment. For each construct, the normalized intensity was averaged and the mean value was plotted. The experiment was repeated two times and each time 5-10 cells were used for analysis.
Hence, we transfected wild-type as well as *Parp-1*-/- mouse fibroblasts with a plasmid coding for GFP-labeled Mecp2 and quantified Mecp2 binding to heterochromatin *in vivo* (Figure 3A-D). We found that the heterochromatin accumulation ability of Mecp2-GFP was significantly increased in *Parp-1*-/- compared to wild-type cells (Figure 3E). This is not the result of differences in DNA methylation *per se* at major satellite repeats in *Parp-1*-/- versus wild-type cells, as it was recently reported that cytosine methylation is unchanged (De Vos, El Ramy et al., 2014).

To test whether the poly(ADP-ribosyl)ated domains were needed, we selected Mecp2 deletions including or not the modifiable ID-TRD. To allow scoring of chromatin binding *in vivo*, we further needed to include the MBD domain (Agarwal, 2011, Brero et al., 2005). Comparing all these deletions in *Parp-1*-/- cells it became clear that the ID-TRD was responsible for a significant enhancement in chromatin binding (Figure 3E). These data indicate that the observed effect is not due to methyl-cytosine binding but requires Parp-1 and the ID-TRD of Mecp2.

To further probe Mecp2 heterochromatin binding kinetics, we performed fluorescence recovery after photobleaching (FRAP) analyses of GFP-tagged Mecp2 as well as Mecp2 deletions with or without the modifiable ID-TRD. The results showed that the ability of Mecp2-GFP to accumulate to pericentromeric heterochromatin was accelerated in *Parp-1*-/- compared to wild-type cells as evident by the slower recovery after photobleaching in the *Parp-1*-/- cells (Figure 3F). Strikingly, the Mecp2 deletion construct including the ID-TRD also showed a prominent enhancement in chromatin binding in *Parp-1*-/- compared to wild-type cells, whereas the deletion construct terminating after the MBD domain did not exhibit any altered chromatin accumulation and displayed a very fast exchange at heterochromatin (Figure 3F).

Finally, we performed *in situ* extraction experiments of *Parp-1*-/- and wild-type cells expressing the same full length Mecp2 and deletion mutants. Both, GFP-tagged full length Mecp2 or the deletion containing the ID-TRD domain (Mecp2Y.1), were extracted only after several minutes of Triton X-100 incubation. Importantly, they were faster extracted from wild type cells and were still well detectable at chromocenters after 14 minutes of detergent treatment in *Parp-1*-/- cells (Figure 3G). In stark contrast, the mutant truncated after the MBD, was very fast and equally extracted from both wild type and *Parp-1*-/- cells.

Altogether, the results obtained from these three independent methods addressing chromatin binding clearly indicate that Mecp2 chromatin accumulation depends on Parp-1 and the ID-TRD of Mecp2 and that the ID-TRD alone is sufficient to restore Mecp2 heterochromatin binding ability *in vivo* to comparable degree as the full-length protein.

**Discussion**

In summary, we showed poly(ADP-ribosyl)ation of Mecp2 in mouse brain tissue (Figure 1). In addition, we found that Mecp2 induced pericentric heterochromatin clustering is increased upon absence of Parp-1 (Figure 2) and could show that the chromocenter binding ability of Mecp2 is elevated in *Parp-1*-/- in comparison to wild-type cells (Figure 3). As reduced poly(ADP-ribosyl)ation level of Mecp2 lead to a significant but not too strong increase in chromatin aggregation ability of Mecp2, we suggest a modulatory role of poly(ADP-ribosyl)ation in Mecp2 mediated chromocenter aggregation (Figure 4).
the functional consequences for MeCP2 mediated higher order chromatin organization. We propose that residues within the MBD domain of MeCP2 as well as poly(ADP-ribose)ylation within ID and TRD work in concert to mediate and regulate MeCP2 function in modelling chromatin architecture. As we recently showed that the ID-TRD domain mediates homo-interactions of MeCP2 (Becker et al., 2013), it is tempting to speculate that poly(ADP-ribose)ylation within this domain negatively regulates MeCP2 ability to crosslink chromatin fibers in vivo.

**Experimental procedures**

**Expression plasmids.** Mammalian expression constructs coding for GFP- or YFP-tagged rat MeCP2 full-length (MeCP2G) and deletions (MeCP2Y.3, MeCP2G.8 and MeCP2G.9) were previously described (Brero et al., 2005, Jost, Rottach et al., 2011). Additional mammalian expression constructs were generated in the following way: deletion constructs MeCP2G.11 - 15 designed with flanking Xhol and BamHI sites according to the sequence of MeCP2G were custom synthesized into pCR Script (Sloning Bio Technology; Puchheim, Germany) and subcloned into the Xhol and BamHI sites of MeCP2G.6 (Agarwal et al., 2007). MeCP2G.16 - 18 were generated using site directed mutagenesis as described in detail before (Makarova, Kamberov et al., 2000, Wang & Malcolm, 1999).

For expression in Sf9 insect cells the Bac-to-Bac Baculovirus Expression System (Invitrogen; Paisley PA4 9RF, UK) was used employing rat MeCP2G construct as well as the GFP construct described before (Becker, Allmann et al., 2013, Jost et al., 2011).

To express PARP-1 with a N-terminal strep-tag, a sequence encoding the strep-tactin target peptide strep tag II (Junttila, Saarinen et al., 2005) was synthesized into pPCR-Script-Amp (Teletechn; Bad Abbach, Germany) flanked by BamHI and NotI sites and subcloned into pFastBac1 using the same sites. Human PARP-1 full-length and deletion constructs were generated by PCR amplification using primers with NotI and XhoI sites and subcloned into frame with the strep-tactin target peptide strep tag in the pFastBac1 vector.

**Cell culture and transfection.** Pmi28 diploid mouse myoblasts were cultured as described before (Kaufmann, Kirsch et al., 1999). Cells were grown to 70-80% confluency on 16 mm glass coverslips in 6 well plates and transfected using Transfектин (Bio-Rad; Hercules, CA, USA). For transfection, 3 µg of plasmid DNA together with 3 µl transfectin were incubated in serum free medium for 20 min at RT and added to the cells. After incubation at 37°C for four hours, the medium was changed and the culture was incubated at 37°C overnight. For PARP inhibition assays, cells were treated with 10 mM 3AB (Alexis Biochemicals; Lörrach, Germany) immediately after media change for 12-15 hours. Within this time, medium plus inhibitors were refreshed every three hours. Transfected cells were fixed with 3.7% formaldehyde in 1xPBS for 10 min. In case of PARP inhibition, 10 mM 3AB were also added to the solutions during the fixation. All washing steps after fixation were performed with 1xPBS plus 0.01 % Tween-20. Cells were counterstained with DAPI followed by mounting in Vectashield (Vector Laboratories; Burlingame, CA, USA).

HEK 293-EBNA cells (Invitrogen; Paisley PA4 9RF, UK) were cultured and transfected as described (Agarwal et al., 2007). Wild-type and Parp-1-/- mouse embryonic fibroblast (MEF) (Trucco, Oliver et al., 1998) cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM; 1 g/l glucose) supplemented with 10 % fetal bovine serum, transfected with Transfectin (Bio-Rad; Hercules, CA, USA) or poly-ethylenimine (PEI, 1 mg/ml in ddH2O, neutralized with HCl; Sigma-Aldrich, St. Louis, MO, USA) (Casas-Delucchi et al., 2012) and fixed with formaldehyde as above.

For fluorescence recovery after photobleaching (FRAP) and in situ extraction experiments, wild-type and Parp-1-/+ MEF cells were transfected by electroporation. Briefly, the cell pellet was resuspended in 100 µl Amaza transfection buffer (50 mM KCl, 15mM MgCl2, 120 mM Na2HPO4 and 50 mM mannitol) with 2 µg of plasmid DNA. The mixture was then transfected into an Amaza cuvette and transfected in an Amaza Nucleofector® using B-32 program for wild-type cells and B-16 program for Parp-1-/- cells. Following transfection, the cells were immediately transferred into μ-Dish (ibidi GmbH, Munich, Bavaria, Germany) with 3 ml of pre-warmed and pre-equilibrated DMEM and incubated for 20 hours. Sf9 insect cells (Invitrogen; Paisley PA4 9RF, UK) were maintained in EX-CELL 420 Insect Serum Free (SAFC; Hampshire SP10 3LF, UK) medium supplemented with 10 % fetal bovine serum shaking at 100 rpm and 28°C. Transfection of Sf9 cells to produce recombinant baculovirus, was performed using Cellfectin (Invitrogen; Paisley PA4 9RF, UK) according to the manufacturer’s instructions.

**Microscopy and image analysis.** For chromocenter counting, fixed cells were examined on a Zeiss Axiovert 200 epifluorescence microscope. Image stacks (0.5 µm Z interval) were acquired with 63x Plan-Apochromatic NA 1.4 or 40x Plan-Neofluar NA 1.3 oil immersion phase contrast objectives and a PCO Sensicam QE cooled CCD camera. Images were processed with Adobe Photoshop and ImageJ (http://imagej.nih.gov/ij/). 3D rendering of image stacks was performed using AMIRA (Visage Imaging Inc.; San Diego, CA, USA) software. Image stacks were analyzed for chromocenter numbers as described in detail before (Agarwal et al., 2011).

To evaluate heterochromatin accumulation ability, confocal Z-stacks were acquired using an UltraView VoX spinning disc system (Perkin Elmer, UK) on a Nikon Ti microscope equipped with an oil immersion 60x Plan-Apochromat NA 1.45 objective lens (Nikon; Tokyo, Japan) (voxel size: 0.12 x 0.12 x 0.5 µm) and a cooled 14-bit EMCCD camera (C9100-50, Hamamatsu Photonics K.K.; Hamamatsu City, Japan). Z-stacks were analyzed...
using Volocity 5.5 software (Perkin Elmer, UK). Chromocenter and nucleoplasm were segmented by intensity-based thresholding (Figure 3). Accumulation at chromocenters was calculated from the ratio of the mean grey value at chromocenters to the mean grey value in the nucleoplasm. Accumulation values from both wild-type and Parp-1<sup>−/−</sup> cells were then normalized to the median accumulation in the wild-type cells.

To evaluate the binding kinetics of fluorescently tagged Mecp2 and deletion mutants in wild-type and Parp-1<sup>−/−</sup> cells, a whole chromocenter was photobleached using an UltraVIEW VoX spinning disc system (Perkin Elmer) mounted on a Nikon Ti microscope equipped with an oil immersion 60x Plan-Apochromat NA 1.45 objective lens, as described before (Rajan, Mortusewicz et al., 2015). Quantitative evaluation was performed using ImageJ and fluorescence intensity normalization and curve fitting was performed with the easyFRAP software as described before (Rajan et al., 2015). T half values were extracted from the single exponential fitting and plots generated with RSudio (https://www.rstudio.com).

To evaluate the extractability of fluorescently tagged Mecp2 and deletion mutants in wild-type and Parp-1<sup>−/−</sup> cells, in situ extractions were performed and release of Mecp2 measured in real time. The assay was performed as described before with following exceptions (Agarwal et al., 2011). Live cell imaging was performed on an UltraVIEW VoX spinning disc system (Perkin Elmer) mounted on a Nikon Ti microscope equipped with an oil immersion 60x Plan-Apochromat NA 1.45 objective lens. The cells were washed one time with PBS/EDTA and imaged. Then the solution was changed to PBS containing 0.5% Triton X-100. Confocal Z-stacks were acquired at two min time intervals for Mecp2G and Mecp2Y.1 for 14 min and 40 s intervals for Mecp2Y.3 for 2 min. Quantifications were performed using Velocity (Perkin Elmer, UK). The total fluorescence intensity signal at the chromocenters was calculated for each time point and for each cell the fluorescence intensity was normalized to the total intensity of chromocenter before Triton X-100 treatment.

**In vivo binding assays.** HEK 293-EBNA (Invitrogen; Paisley PA4 9RF, UK) or MEF cells (Trucco et al., 1998), transfected with expression plasmids as indicated, were pelleted after washing with 1xPBS and lysis was performed for 10 min on ice. To obtain disruption of protein-DNA associations and for high protein amount buffer B (25 mM Tris-HCl, pH 8.0; 1 M NaCl; 50 mM glucose; 10 mM EDTA; 0.2 % Tween-20; 0.2 % NP40) was used and supplemented with protease inhibitors (Complete Mini; Roche, Mannheim, Germany).

Mouse whole brain tissue (three months old C57BL/6N; Charles River Laboratories International, Inc., Wilmington, MA 01887, USA) was first fractionated to obtain pure nuclei. Tissue (6 grams) was first homogenized in a 0.25 M sucrose solution (20 mM triethanolamine-HCl pH 7.6, 30 mM KCl, 5 mM MgCl<sub>2</sub>, 0.1 mM PMSF and 1 mM DTT). After centrifugation at 1,000 x g for 10 min, the pellet was resuspended in 2.1 M sucrose solution followed by centrifugation at 50,000 x g for 40 min. The pellet was again dissolved in 0.26 M sucrose solution and centrifugation was done at 1,000 x g for 10 min. The isolated nuclei were incubated in buffer B for 15 min on ice. 500 µl of the extract were diluted 1:4 with buffer C (25 mM Tris-HCl, pH 8.0; 50 mM glucose; 10 mM EDTA; 0.2 % Tween-20; 0.2 % NP40), to obtain a NaCl concentration of 250 mM. After centrifugation (20,000 x g, 15 min, 4°C) rabbit polyclonal anti-Mecp2 antibody (40 µg) (Jost et al., 2011) or chromatographically purified rabbit IgG (40 µg; Organon Teknika Corp; Durham, #55944, NC, USA) were added to the supernatant and incubated for 1.5 hours rotating at 4°C. To pull down the immunocomplexes, 50 µ1 protein A agarose beads (Fast Flow; Upstate, Temecula, CA, USA), equilibrated with the corresponding buffer, were added and incubated for one hour.

For purification of Mecp2 from mouse brain tissue employing TrisNTA (kind gift of R. Tampé, Goethe University Frankfurt, Germany) coupled beads, approximately 1x10<sup>7</sup> of mouse brain nuclei in 1x PBS were subjected to centrifugation (14,000 rpm, 10 min, 4°C). The nuclear pellet was dissolved in 0.2% Triton X-100 in 1 x PBS supplemented with protease inhibitors as described in (Becker et al., 2013), incubated for 10 min on ice and afterwards washed three times with 1 x PBS. 300 µ1 binding buffer (20 mM imidazole, 0.5 M NaCl in 1 x PBS) were added and sonication was performed twice for 20 seconds at 70 % intensity followed by centrifugation (14,000 rpm, 10 min, 4°C). 120 µl of TrisNTA coupled beads (NHS-activated Sepharose 4 Fast Flow; GE Healthcare Europe GmbH) were washed three times with binding buffer and activated for 15-20 min with NiSO<sub>4</sub>-Hexahydrate (20 mM), added to the protein lysate and incubated overnight at 4°C under rotation. To elute the proteins from the beads, the beads were washed with 40 mM imidazole for 15 min.

To purify endogenous Mecp2 from mouse brain tissue using boronic acid beads, mouse brain tissue was lysed in hot lysis buffer (5mM Tris HCl pH 8, 250mM NaCl, 1% SDS, 0.1% NP-40, 5mM MgCl<sub>2</sub> 0.5 mM EDTA), supplemented with protease inhibitors (Roche), PARG inhibitor RBPI-4 (kindly provided by Paul J. Hergenrother; (Finch, Knezevic et al., 2012)) and 100 µM 3AB (Alexis Biochemicals; Lörrach, Germany), followed by centrifugation (14000 rpm, 10 min, RT). 30 µl of slurry boronic acid beads (Chemicell) were washed twice in lysis buffer, added to each lysate and incubated for 30 min at RT followed by three washes in lysis buffer and one in water before proceeding with SDS-PAGE and Western blotting.

For immunoprecipitation using the GFP binder (ChromoTek; Planegg-Martinsried, Germany; (Rothbauer, Zolghadr et al., 2008)), 50 µl protein A agarose beads were incubated with 100 µg GFP binder for one hour, then added to the extract and again incubated for one hour at 4°C rotating. After a short spin, the supernatant was removed and the beads were washed three times with 500 µl of the same buffer used during cell lysis. The beads were resuspended in 1x SDS-containing...
sample buffer, boiled for 10 min at 95°C and analyzed by SDS-PAGE electrophoresis followed by Western blotting.

**Purification of proteins.** Sf9 insect cells (Invitrogen; Paisley PA4 9RF, UK) were infected with the recombinant baculovirus (P3 stock) and incubated at 28°C with shaking for 5 days. The cells were pelleted by centrifugation (200 x g, 5 min, 4°C) and resuspended in either buffer B (25 mM Tris-HCl, pH 8.0; 1 M NaCl; 50 mM glucose; 10 mM EDTA; 0.2 % Tween-20; 0.2 % NP40) or buffer D (PBS containing 300 mM NaCl and 0.05 % NP40). All buffers were supplemented with protease inhibitors (Complete Mini; Roche, Mannheim, Germany). After incubation on ice for 10 min, cells were disrupted with a high-pressure homogenizer (EmulsiFlex-C5, Avestin; Ottawa, Ontario, Canada) followed by centrifugation at 14,000 x g for 30 min.

Strep-tagged recombinant proteins were purified by incubating the supernatant with 500 µl of Strep-Tactin Sepharose (IBA; Göttingen, Germany) beads for three hours at 4°C on a rotary shaker. To elute strep-tagged proteins, the beads were incubated with D-Desthiobiotin (0.5 mg/ml; IBA; Göttingen, Germany), dissolved in 1x PBS, for 30 min at 4°C. After centrifugation (200 x g, 2 min), beads were separated from the eluate containing the purified proteins.

GFP fusion proteins were immobilized using the GFP-Trap (ChromoTek; Planegg-Martinsried, Germany) as described (Rothbauer et al., 2008).

**Western blot analysis.** Western blot analysis was performed as described (Mortusewicz, Rothbauer et al., 2006), using PVDF membrane (BioRad; Hercules, CA, USA). Immunoreactive bands were visualized either by ECL plus or ECL advanced Western Blot Detection Kit (GE Healthcare; München, Germany). The following primary antibodies were used for Western blot analysis: rabbit polyclonal anti-Mecp2 (Upstate, #07-013, Temecula, CA, USA), mouse monoclonal anti-GFP (Roche, #11814460001, Mannheim, Germany), mouse monoclonal anti-Parp-1 (F-2, #sc-8007, Santa Cruz; CA, USA), mouse monoclonal anti-PAR (Trevigen, #4335-MC, Gaithersburg, MD, USA). Secondary antibodies used were: horseradish peroxidase conjugated anti-mouse IgG (GE Healthcare, #NA 931, München, Germany), horseradish peroxidase conjugated anti-rabbit IgG (Sigma, St. Louis, #A-0545, MO, USA).

**In vitro ribosylation assay.** In vitro ribosylation analysis of recombinant GFP, Mecp2G or GFP-tagged Mecp2 deletions immobilized onto GFP-Trap beads (ChromoTek; Planegg-Martinsried, Germany), were performed as described in (Schreiber, Ame et al., 2002) with following modifications: purified st-PARP-1 (50 ng) from Sf9 cells, 20 µM cold NAD⁺ in addition to [α-³²P]NAD⁺ and 100 ng DNase I activated DNA (Alexis Biochemicals; Lörrach, Germany) were used. After the reaction, the proteins were washed three times with buffer B to disrupt binding to st-PARP-1.

**Quantification of overexpressed Mecp2G in mouse cells.** Mouse myoblast cells were transfected with plasmids coding for GFP-tagged Mecp2 using FuGENE HD (promega; Madison, WI, USA) transfection reagent according to the manufacturer’s instructions. Defined amounts of Mecp2G expressing cells were separated from untransfected cells and counted using flow cytometry. Recombinant Mecp2G, extracted from insect cells using buffer B (25 mM Tris-HCl, pH 8.0; 1 M NaCl; 50 mM glucose; 10 mM EDTA; 0.2 % Tween-20; 0.2 % NP40), was immobilized to GFP-Trap beads (ChromoTek; Planegg-Martinsried, Germany). To determine the concentration of immobilized Mecp2G in order to obtain a Mecp2G standard, SDS-PAGE was performed followed by staining of the protein using Coomassie Brilliant Blue (CBB). In parallel, defined amounts of a BSA standard were analyzed by SDS-PAGE with consequent CBB staining and used to determine the amount of Mecp2G. To calculate the amount of Mecp2G overexpressed in mouse myoblast cells, defined amounts of recombinant Mecp2G standard and Mecp2G extracted from defined numbers of sorted mouse myoblasts were analyzed by SDS-PAGE and quantitative western blotting with mouse anti-GFP (Roche; Mannheim, Germany) and anti mouse IgG-Cy5 (Jackson Immuno Research; Suffolk, UK) antibody using fluorescence imaging (Storm 860, Molecular Dynamics).
target in different genotoxic stress responses. Mol Cell 52: 272-85


In this thesis, we focused on how the 5-methylcytosine (5mC) modifier Ten-eleven translocation 1 (Tet1) oxidizes 5mC to 5-hydroxymethylcytosine (5hmC) and how this process is regulated by methylcytosine readers MBD (Methyl-CpG binding domain) proteins.

To understand the Tet1 mediated oxidation mechanism, we developed methods to detect step by step the Tet oxidation process including Tet-DNA binding, 5mC flipping and 5mC oxidation. Moreover, using these methods we were able to detect base flipping induced by the 5mC writer HpaII DNA methyltransferase and the 5mC reader DNA base repair protein Mbd4. Recently, N\textsuperscript{6}-methyldeoxyadenosine (m6dA) was found in the mammalian genome (Koziol, Bradshaw et al., 2016). It would be very interesting to apply our methods to check whether enzymes, such as m6dA methyltransferase (DNA N\textsuperscript{6} adenine methyltransferase) (Greer, Blanco et al., 2015) also use a base flipping mechanism to modify adenine.

Post translational modifications of proteins are very important to regulate protein function and amenable to pharmacological intervention (Haberland, Montgomery et al., 2009). Here, we show that endogenous Mecp2 can be poly(ADP-ribosyl)ated in mouse brain and consequently the binding ability of Mecp2 to DNA is decreased. Previous studies showed that Tet1 proteins can also be poly(ADP-ribosyl)ated (Ciccarone, Valentini et al., 2015). However, the function of this modification of Tet1 is still not clear. Here, we show that the interaction of Tet with DNA is crucial for the 5mC conversion. Future experiments could be performed to test the effect of poly(ADP-ribosyl)ation on the Tet oxidation activity.

Figure 1 Summary of how Tet oxidize 5mc and how this process is regulated by Mbd proteins.
Tet1 proteins use a DNA base flipping mechanism to modify 5-methylcytosine (5mC) and three steps are involved in this process, including 5mC-Tet binding, 5mC flipping and 5mC oxidation. Mbd1 binds to methylated CpGs and also unmethylated CpGs via its CXXC3 and MBD domain, respectively. The binding ability of unmethylated CpGs and interaction to Tet1 allow Mbd1 to recruit Tet1 to methylated DNA. However, preferential binding ability of Mbd2 and Mecp2 to methylated CpGs restricts Tet1 binding to DNA.
Although Mbd1 belongs to the MBD protein family, it regulates Tet1 activity in an isoform dependent manner. The Mbd1 isoform containing CXXC3, which binds unmethylated CpG confers Mbd1 of the unique function to enhance Tet1 mediated 5mC oxidation. Since Mbd1 was reported to be highly expressed in cancer cell lines and tissues (Patra, Patra et al., 2003, Xu, Zhu et al., 2013) and aberrant DNA methylation is a hallmark of cancer, further experiments need to be performed to check the consequences of Mbd1 enhanced-5hmC formation in cancer development. Finally, it would be very interesting to investigate whether other CXXC domain proteins can similarly regulate Tet activity.

References
Ciccarone F, Valentini E, Zampieri M, Caiafa P (2015) 5mC-hydroxylase activity is influenced by the PARylation of TET1 enzyme. Oncotarget 6: 24333-47
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2AP</td>
<td>2-Aminopurine</td>
</tr>
<tr>
<td>5caC</td>
<td>5-carboxylcytosine</td>
</tr>
<tr>
<td>5IC</td>
<td>5-formylcytosine</td>
</tr>
<tr>
<td>5hmC</td>
<td>5-hydroxymethylcytosine</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>CAA</td>
<td>chloroacetaldehyde</td>
</tr>
<tr>
<td>DAPI</td>
<td>4′,6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>Dnmt</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic Mobility Shift Assay</td>
</tr>
<tr>
<td>FRAP</td>
<td>Fluorescence Recovery After Photobleaching</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>HEK</td>
<td>Human Embryonic Kidney</td>
</tr>
<tr>
<td>HRM</td>
<td>High Resolution Melting</td>
</tr>
<tr>
<td>HRP</td>
<td>Horse Radish Peroxidase</td>
</tr>
<tr>
<td>LINE1, L1</td>
<td>Long Interspersed Nuclear Elements 1</td>
</tr>
<tr>
<td>MaSat</td>
<td>Major Satellite</td>
</tr>
<tr>
<td>MBD</td>
<td>Methyl-CpG Binding Domain</td>
</tr>
<tr>
<td>Mecp2</td>
<td>Methyl CpG binding protein 2</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse Embryonic Fibroblasts</td>
</tr>
<tr>
<td>M.HpaII</td>
<td>HpaII Methyltransferase</td>
</tr>
<tr>
<td>MTF</td>
<td>Mouse Tail Fibroblasts</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>PCNA</td>
<td>Proliferating Cell Nuclear Antigen</td>
</tr>
<tr>
<td>S-phase</td>
<td>Synthesis phase</td>
</tr>
<tr>
<td>Tet1</td>
<td>Ten-eleven translocation 1</td>
</tr>
<tr>
<td>Tet11CDmut</td>
<td>Catalytic mutant of Tet1</td>
</tr>
<tr>
<td>Tet1CD</td>
<td>Catalytic Domain of Tet1</td>
</tr>
<tr>
<td>Tm</td>
<td>melting temperature</td>
</tr>
<tr>
<td>TRD</td>
<td>Transcriptional Repression Domain</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
</tbody>
</table>
List of contributions

Introduction

**P.Z.** wrote and prepared figures and tables for DNA modifications, DNA modifiers and corresponding disease (Fig. 1, Fig. 2, Table 1, Table 2, Table 4, Table 5).

**A.K.L.** wrote and prepared figures and tables for DNA modifications readers and corresponding disease (Fig. 4, Fig. 5, Fig. 6, Fig. 7, Table 3, Table 4, Table 6).

**M.C.C.** conceived and supervised the project and revised the manuscript.

Chapter 1

**A.B.** conducted most of the experiments, analyzed the results, and wrote most of the paper.

**P.Z.** performed, analyzed and prepared figures for the FRAP and in situ extraction experiments (Fig 3F, Fig 3G) and wrote the corresponding methods and figure legends.

**L.A.** conducted and, with the help of **B.B.** and **D.E.**, analyzed the chromocenter binding microscopy experiments.

**D.M.** and **H.L.** provided the sorted cell material for Mecp2 quantification.

**M.H.** contributed baculovirus constructs and insect cell expression.

**G.B.** and **M.H.** provided the boronic acid data.

**V.S.** provided materials and protocols.

**M.C.C.** designed the project and wrote the paper with **A.B.**.
Acknowledgements

I would like to thank my supervisor M. Cristina Cardoso to give me an opportunity to perform my Ph.D. work in such a well-equipped and well-organized lab. I would like to express my deepest gratitude to Cristina for her critical comments and useful suggestions on my work.

I also thank Prof. Dr. Beatrix Süß as my second supervisor.

I also would like to thank our methylation-group people (Bianca, Anne, Florian, Cathia, of course also Cristina) for commenting on my work. I learned a lot from you on how to prepare figures and how to perform experiments. Especially many thanks to Anne (Alu) and Florian for helping me not only on experiments but also on my German life and also for reading and reviewing my thesis.

I thank all the members in the Cardoso lab especially:

- Francesco, Kathrin, Patrick and Cathia for reading and commenting on manuscripts.
- Alex for suggestions on fluorescence signal detection on the microscopes and also on the fluorescence microplate reader.
- Henry for providing an idea to detect fluorescence signal on the gel using the fluorescence microplate reader.
- Anne (AL) for preparing and culturing cells.
- Stephanie for preparing and staining mouse brain.
- Manuela for GFP/RFP binders.

I thank all the people who contributed in many ways to the work described in this thesis, especially Heinrich Leonhardt (LMU) and Gerald G. Schumann (Paul Ehrlich Institute).

I thank the China Scholarship Council (CSC) for providing me a Ph.D. scholarship.

Finally, I would like to thank my family for supporting me.

一颗感恩的心，送给所有帮助和支持过我的同学，老师和家人。
非常感谢国家留学基金委提供奖学金资助。
Declaration - Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den 2016
Curriculum vitae

PERSONAL DETAIL

Name: Peng Zhang
Address: Technische Universität Darmstadt
        Schnittspahnstrasse 10
        64287 Darmstadt, Germany
Date of birth: 07, April, 1987
Place of birth: Gansu, China

EDUCATION

Cell biology and epigenetics, Biology department,
Technische Universitaet Darmstadt, Germany.

MS: Sept. 2009—June 2012
Master of Science in Zoology, College of Animal Science and Veterinary Medicine,
Jilin University, P. R. of China. Major: stem cell and animal transgenesis.

BS: Sept. 2005—June 2009
Bachelor of Science in Biotechnology, College of Animal Science and Veterinary Medicine,
Jilin University, P. R. of China. Major: Biotechnology.
List of publications


Manuscripts under revision


Peng Zhang*, Anne K. Ludwig*, Anne Lehmkuhl, M. Cristina Cardoso. Ten-eleven translocation proteins activate L1 retrotransposition and this is repressed by methyl-CpG binding proteins. (* Equal contribution)

Peng Zhang and M. Cristina Cardoso. Methyl-CpG binding domain protein 1 regulates localization and activity of Ten-eleven translocation protein 1 in a CX3C domain dependent manner.

Manuscripts in preparation

Peng Zhang, Florian D. Hastert, Anne K. Ludwig and M. Cristina Cardoso. DNA base flipping analytical pipeline.