

The Guardians of the Epigenome

by

Anne K. Ludwig



...what is essential is invisible to the eye.

Antoine de Saint-Exupéry

The Guardians of the Epigenome

Regulation and Role of Nucleotide Modifications



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vom Fachbereich Biologie der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines
Doctor rerum naturalium

genehmigte Dissertation von
Dipl. Biol. Anne Kathrin Ludwig
aus Bensheim

Referent: Prof. Dr. M. Cristina Cardoso

Koreferent: Prof. Dr. Beatrix Süß

Tag der Einreichung: 16.3.2017

Tag der mündlichen Prüfung: 11.5.2017

Darmstadt 2017

D17

Contents

1 Summary/Zusammenfassung	1
2 General introduction	6
2.1 Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease	6
3 Aims of this study	33
4 Results	36
4.1 Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise	36
4.2 L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins	80
4.3 Principles of protein targeting to the nucleolus	118
4.4 Visualization of the Nucleolus in Living Cells with Cell-Penetrating Fluorescent Peptides	138
5 General Conclusions and Perspectives	153
6 Annex	163
6.1 Abbreviations	163
6.2 List of contributions	169
6.2.1 General introduction	169
6.2.2 Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise	169
6.2.3 L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins	169
6.2.4 Principles of protein targeting to the nucleolus	170
6.2.5 Visualization of the Nucleolus in Living Cells with Cell-Penetrating Fluorescent Peptides	170
6.2.6 General Conclusions and Perspectives	170
6.3 Acknowledgments	172
6.4 Declaration - Ehrenwörtliche Erklärung	174
6.5 Curriculum vitæ	175

List of Figures

General introduction	6
1 Schematic representation of the Dnmt protein family	9
2 DNA base modifications with respective enzymes	11
3 Schematic representation of the Tet protein family	12
4 5-(Hydroxy)methylcytosine readers	14
5 Schematic representation of the MBD protein family	15
6 Schematic representation of the Kaiso-like protein family	19
7 Schematic representation of the SRA domain protein family	19
8 Writing, reading, and translating DNA modifications	22
Aims of this study	33
1 Scheme summarizing the main aims of Chapter 1 and Chapter 2	33
2 Scheme summarizing the main aims of Chapter 3 and Chapter 4	34
Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise	36
1 Impact of methyl-CpG binding domain proteins on the efficiency of Tet1 mediated 5mC oxidation . . .	45
2 Influence of the chronological DNA binding order on the protecting ability of MBD proteins	47
3 Effect of different DNA binding modes on Tet1 activity	48
4 Impact of 5mC-specific and sequence-unspecific DNA binding proteins on the DNA binding ability of Tet1CD proteins	50
5 Epigenetic composition of pericentric heterochromatin in a mouse model for Rett syndrome	51
6 Correlation of subnuclear 5hmC distribution and major satellite expression	52
7 Mecp2 and Mbd2 preserve chromatin composition and thus genomic integrity by insulating 5mC from Tet1 activity	54
L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins	80
1 Tet1 reactivates retrotransposition of endogenous L1	89
2 Tet1 reactivates retrotransposition of engineered L1	92
3 MBD proteins prevent Tet-mediated reactivation of endogenous L1	94
Principles of protein targeting to the nucleolus	118
1 Distribution of poly-(R) and poly-(K) peptides in living cells	124
2 Intracellular distribution of peptide tagged GFP tracer proteins in living cells	125
3 Intracellular pH landscape by ratiometric fluorescence microscopy	126
4 Analyses of RNA binding to poly-(R) peptides and GFP fusions	127
5 Chart of peptide and protein pI values, intracellular distribution and sequence composition	128

Visualization of the Nucleolus in Living Cells with Cell- Penetrating Fluorescent Peptides	138
1 Application of a fluorescent cell-penetrating-peptide-based nucleolar marker	147
2 Intracellular distribution of nucleolar markers in living cells of various species and origin	148
3 Image analysis steps for segmentation of nucleoli	149
General Conclusions and Perspectives	153
1 Scheme summarizing the main conclusions drawn from our studies in Chapter 1 and 2	153
2 Immunostaining and colocalization analysis of Tet proteins with nucleoli	154
3 Subnuclear distribution of endogenous, GFP-tagged Tet1 and Tet2 proteins in stable J1 mESC cell lines	155
4 <i>In silico</i> analysis of Tet protein sequences	156

List of Tables

General introduction	6
1 Phenotypes of initial <i>Dnmt</i> knockout mouse models	9
2 Phenotypes of initial <i>Tet</i> knockout mouse models	12
3 Phenotypes of initial MBD-deficient mouse models	16
4 Comparison of human proteins and their mouse orthologs	20
5 Summary of disease-related DNMT and TET mutations	21
6 Summary of disease-related MBP alterations	22

1 Summary

While from a genetic perspective all cells of an organism are identical, they vary greatly in type and function. Major determinants of cellular diversity are epigenetic alterations, including post-synthetic modifications of nucleic acids. In DNA, the best-studied chemical modification is the methylation of cytosine at carbon C5. 5-methylcytosine (5mC) plays a central role in the regulation of gene expression and has been implicated in a variety of biological processes and diseases. While initially considered as a relatively stable epigenetic mark, a family of proteins named Ten eleven translocation (Tet), were recently shown to catalyze the conversion of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in an iterative oxidation reaction. These oxidized derivatives have been shown to modulate gene expression and control cellular metabolism. Hence, to avert the formation of aberrant, pathogenic DNA modifications, Tet activity must be precisely regulated.

Here, we analyzed the potential of methyl-CpG binding domain (MBD) proteins to control Tet dioxygenase activity *in vitro* and *in vivo*. We demonstrate that prior binding of Mecp2 and Mbd2 to DNA protects 5mC from Tet1 mediated oxidation in a concentration dependent manner. The mechanism is based on competitive, sequence unspecific binding to DNA and correlates with nucleic acid coverage and retention time of MBD proteins on DNA. Accordingly, we find increased levels of the Tet oxidation product 5hmC at pericentric heterochromatin in neurons of Mecp2 deficient mice with concomitant reactivation of highly methylated major satellite DNA repeats. Moreover, we find increased expression and retrotransposition of endogenous and engineered long interspersed nuclear elements (LINE1) as potential consequence of unconfined Tet1 activity in human cells.

Similar to DNA, RNA contains a variety of post-synthetic modifications that extend their chemical information and properties. One out of < 100 noncanonical ribonucleobases is C5-methylated cytosine, which is present in various types of RNA, including ribosomal RNA (rRNA). Recent studies have shown that Tet proteins do not exclusively hydroxylate methylcytosine in DNA, but also act on 5mC in single-stranded ribonucleic acids. Since modification of rRNA potentially requires the localization of Tet proteins to the nucleolus, the place of rRNA synthesis and ribosome assembly, it is important to understand the general principles of protein targeting to this subnuclear compartment. Therefore, we determined the molecular requirements that are necessary and sufficient for the localization and accumulation of peptides and proteins inside nucleoli of living cells. Our data indicate that peptide units composed of consecutive, positively charged arginines with an isoelectric point ≥ 12.6 meet the chemical conditions for nucleolar localization. Consistent with this, we mapped an arginine-rich region within the low complexity insert of Tet2, which accumulated in nucleoli upon deletion of its regulatory N-terminal domain. Using a pH sensitive dye, we revealed that the nucleolus is relatively acidic. Accordingly, we show that arginine-rich peptides, which carry a net positive charge under these electrochemical conditions, interact with negatively charged RNA *in vitro*. This interaction in turn, is consistent with the conservation of this nucleolar targeting principle from insects to man. Finally we developed a detailed protocol for the visualization of nucleoli in living cells using fluorescently tagged cell-penetrating peptides (CPP), which allows precise nucleolar localization analysis of various proteins, such as Tet. In summary, our data contribute to understanding the regulation of Tet activity outside and (Tet) protein localization inside of nucleoli.

1 Zusammenfassung

Aus genetischer Sicht sind alle Zellen eines Organismus identisch, jedoch unterscheiden sie sich deutlich in ihrer Art und Funktionsweise. Wichtige Determinanten zellulärer Diversität sind epigenetische Änderungen, darunter post-synthetische Modifikationen von Nucleinsäuren. Die am besten untersuchte chemische Modifikation von DNA ist die Methylierung von Cytosin am fünften Kohlenstoffatom. 5-Methylcytosin (5mC) spielt eine wichtige Rolle bei der Regulation der Genexpression und steht in Verbindung mit einer Vielzahl biologischer Prozesse und Krankheiten. Obwohl zunächst angenommen wurde, dass 5mC eine relativ stabile epigenetische Modifikation ist, wurde kürzlich gezeigt, dass eine Familie von Proteinen, Ten eleven translocation (Tet), die Umsetzung von 5mC zu 5-Hydroxymethylcytosin (5hmC), 5-Formylcytosin (5fC) und 5-Carboxylcytosin (5caC) katalysiert. Es wurde gezeigt, dass diese oxidativen Derivate an der Modulierung der Genexpression und Steuerung des Zellstoffwechsels beteiligt sind. Um der Bildung aberranter, krankheitsverursachender DNA Modifikationen vorzubeugen, muss daher die Aktivität von Tet Proteinen präzise reguliert werden.

In dieser Arbeit wurde das Potenzial von Proteinen mit Methyl-CpG-Bindedomäne untersucht, die Dioxygenase-Aktivität von Tet Proteinen *in vitro* und *in vivo* zu kontrollieren. Wir zeigen, dass vorzeitiges Binden von Mecp2 und Mbd2 an DNA 5mC vor Tet1-katalysierter Oxidation in Abhängigkeit ihrer Konzentration schützt. Der Mechanismus basiert auf kompetitiver, sequenzspezifischer Bindung an DNA und korreliert mit der Abdeckung von Nucleinsäuren und der Verweildauer von MBD Proteinen an DNA. Folglich finden wir erhöhte Level des Tet Oxidationsprodukts 5hmC an perizentrischem Heterochromatin in Neuronen von Mecp2 defizienten Mäusen mit

einhergehender Reaktivierung stark methylierter "major satellite DNA repeats". Darüber hinaus messen wir eine erhöhte Expression und Retrotransposition von endogenen und konstruierten "long interspersed nuclear elements" (LINE1) als mögliche Folge uneingeschränkter Tet Aktivität in humanen Zellen. RNA enthält, ähnlich wie DNA, eine Vielzahl post-synthetischer Modifikationen, die deren chemische Information und Eigenschaften erweitern. Eine von mehr als 100 nicht-kanonischen Ribonucleobasen ist C5-methyliertes Cytosin, das in unterschiedlichen RNA Typen, darunter ribosomaler RNA, vorhanden ist. Neueste Studien haben gezeigt, dass Tet Proteine nicht ausschließlich Methylcytosin von DNA, sondern auch 5mC einzelsträngiger RNA hydroxylieren. Da eine Modifikation ribosomaler RNA möglicherweise eine Lokalisation von Tet Proteinen im Nucleolus, dem Ort der rRNA Synthese und Zusammensetzung von Ribosomen, voraussetzt, ist es wichtig, die allgemeinen Prinzipien der gezielten nucleolaren Lokalisation von Proteinen und Peptiden zu verstehen. Hierfür haben wir die molekularen Anforderungen, die für eine Lokalisation und Akkumulierung im Nucleolus lebender Zellen notwendig und ausreichend sind, untersucht. Unsere Daten zeigen, dass Peptideinheiten aus aneinandergereihten, positiv geladenen Argininen mit einem isoelektrischen Punkt ≥ 12.6 , die chemischen Bedingungen für eine nucleolare Lokalisation erfüllen. Diesem Prinzip entsprechend, identifizierten wir eine Arginin-reiche Region innerhalb des "low complexity inserts" von Tet2 Proteinen, die nach Deletion des regulatorischen N-terms im Nucleolus akkumulieren. Mit Hilfe eines pH-sensitiven Farbstoffs konnten wir visualisieren, dass der Nucleolus relativ sauer ist. Folglich zeigen wir, dass Peptide und Proteine mit hohem isoelektrischen Punkt, die unter diesen elektrochemischen Bedingungen eine positive Ladung tragen, mit negativ geladener RNA *in vitro* interagieren. Diese Interaktion erklärt

wiederum die evolutionäre Konservierung dieses Lokalisationsprinzips von Insekten bis zum Mensch. Letztlich entwickelten wir ein Protokoll zur Visualisierung von Nukleoli in lebenden Zellen, wodurch die Lokalisation unterschiedlicher Proteine, wie beispielsweise Tet, genau analysiert werden kann. Zusammenfassend tragen unsere Daten dem Verständnis der Tet Regulation außerhalb und der Lokalisation von (Tet) Proteinen innerhalb von Nukleoli bei.

2 General introduction

2.1 Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease

A.K. Ludwig wrote the text, prepared figures (4, 5, 6, 7, 8) and tables (3, 6) for sections 2 (DNA modification readers), 2.1 (MBD protein family), 2.1.1 (Mecp2), 2.1.2 (Mbd1), 2.1.3 (Mbd2), 2.1.4 (Mbd3), 2.1.5 (Mbd4), 2.2 (Kaiso), 2.3 (SRA domain protein family), 3.3 (MBP proteins in disease), 4 (Concluding remarks). A.K. Ludwig prepared table 4 together with P. Zhang.

Content of figures:

Figure 4: Overview of 5-(Hydroxy)methylcytosine readers. **Figure 5:** Schematic representation of the MBD protein family. **Figure 6:** Schematic representation of the Kaiso-like protein family. **Figure 7:** Schematic representation of the SRA domain protein family. **Figure 8:** Graphical summary of how DNA modification writers, readers, and translators can impact on chromatin composition, structure, as well as genome expression and stability.

Content of tables:

Table 3: Phenotype of initial MBD-deficient mouse models. **Table 6:** Summary of disease-related MBP alterations. **Table 4:** Comparison of human proteins and their mouse orthologs.



Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease

Anne K. Ludwig[†], Peng Zhang[†] and M. C. Cardoso^{*}

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

OPEN ACCESS

Edited by:

Joanna Mary Bridger,
Brunel University London, UK

Reviewed by:

Abhijit Shukla,
Harvard Medical School, USA
Jorg Tost,
Commissariat à l'Énergie
Atomique-Institut de Génomique,
France

*Correspondence:

M. C. Cardoso
cardoso@bio.tu-darmstadt.de

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Genetics

Received: 31 March 2016

Accepted: 06 June 2016

Published: 21 June 2016

Citation:

Ludwig AK, Zhang P and
Cardoso MC (2016) Modifiers
and Readers of DNA Modifications
and Their Impact on Genome
Structure, Expression, and Stability
in Disease. *Front. Genet.* 7:115.
doi: 10.3389/fgene.2016.00115

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

DNA MODIFICATIONS AND MODIFIERS

Cytosine Modifiers: Dnmts

In mammals, the modified cytosine was initially described by Hotchkiss (1948) and was further extensively studied since the 1970s (Razin and Cedar, 1977). Recently, evidence for methylation of adenine has been also reported in mammals (Kozioł 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 to the fifth carbon of a cytosine residue to form 5-methylcytosine (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 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 et al., 1988; Li et al., 1992). Unlike Dnmt1 and Dnmt3a/3b, Dnmt2 is a RNA methyltransferase rather than a DNA methyltransferase (Okano et al., 1998; Yoder and Bestor, 1998; Goll et al., 2006). 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**.

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 et al., 1999). The domain structure for *de novo* methyltransferases Dnmt3a and Dnmt3b is similar, including a DNA binding domain PWWP domain (Qiu 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 (Okano et al., 1998; Chen et al., 2002). *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 et al., 2002, 2004). Several isoforms were identified and among all isoforms only Dnmt3b1 and Dnmt3b2 possess DNA methyltransferase activity (Aoki 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).

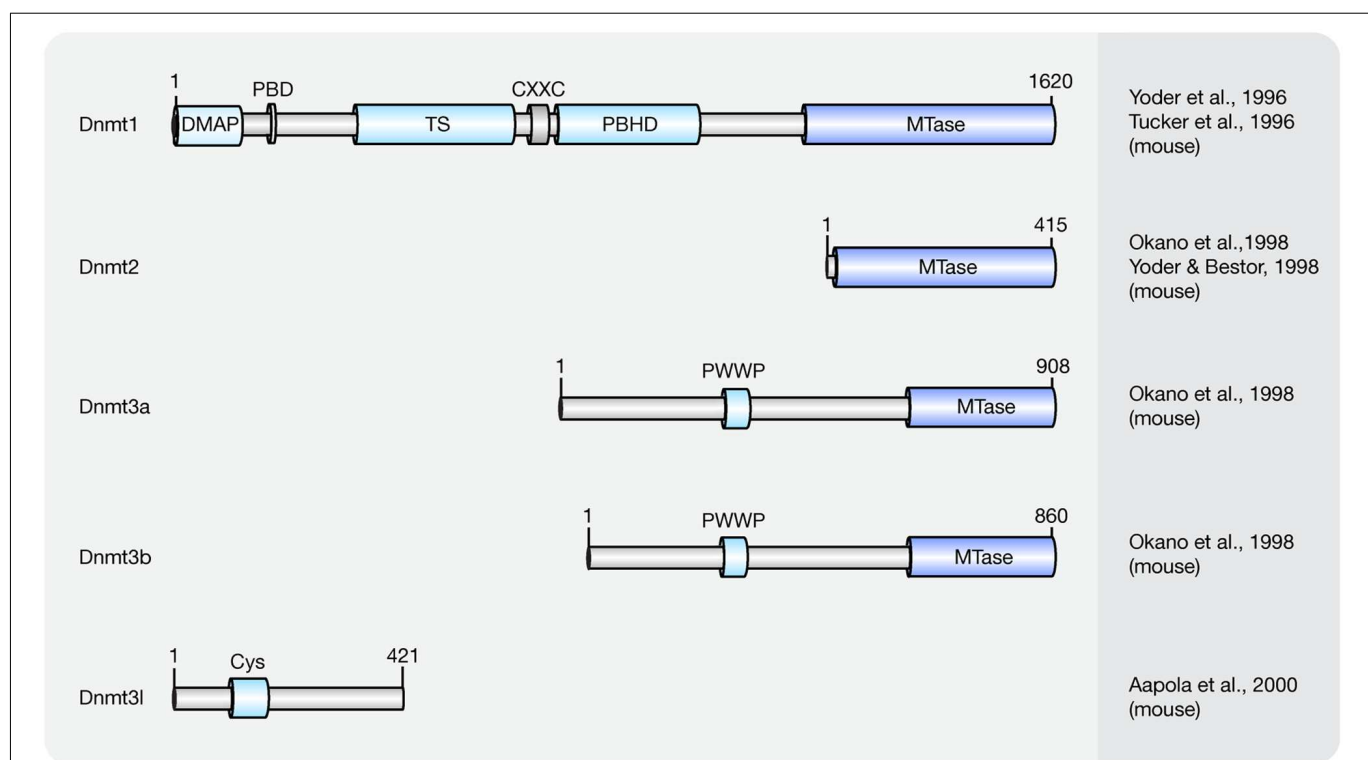


FIGURE 1 | Schematic representation of the Dnmt protein family. Shown are domain structures of mouse Dnmt proteins and the initial references. Numbers represent amino acid positions. DMAP, Dnmt1-associated protein binding domain; PBD, proliferating cell nuclear antigen (PCNA)-binding domain; TS, targeting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

TABLE 1 | Phenotypes of initial Dnmt knockout mouse models.

Genotype	Phenotype	Reference
<i>Dnmt1</i> null	Homozygous knockout <i>Dnmt1</i> were stunted, delayed in development, and did not survive past midgestation	Li et al., 1992
<i>Dnmt3a</i> null	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	Okano et al., 1999
<i>Dnmt3b</i> null	No viable <i>Dnmt3b</i> knockout mice were recovered at birth	Okano et al., 1999
<i>Dnmt2</i> null	Mice homozygous for this <i>Trdm1</i> (formerly <i>Dnmt2</i>) knock-out have abnormal RNA methylation while genomic DNA methylation patterns are not detectably altered	Goll et al., 2006
<i>Dnmt3l</i> null	Disruption of <i>Dnmt3l</i> caused azoospermia in homozygous males and heterozygous progeny of homozygous female died before midgestation	Bourc'his et al., 2001

Although Dnmt3l does not possess DNA methylation activity (Bourc'his et al., 2001), it strongly interacts with Dnmt3a/3b and enhances their methylation activity (Aapola et al., 2000; Suetake et al., 2004; Hu et al., 2008). 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).

Maintenance DNA Methylation

Dnmt1 has a preference for hemi-methylated DNA substrates (Song et al., 2011) and is the enzyme responsible for the maintenance of DNA methylation after DNA replication (Leonhardt 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 mice contains 1620 amino acids and includes an N-terminal regulatory domain and a C-terminal catalytic domain (Tucker et al., 1996; Yoder et al., 1996). However, one isoform lacking the most N-terminal 118 amino acids was shown to accumulate in mouse oocytes (Mertineit 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 et al., 2000). With the contribution of Uhrf1 [ubiquitin-like with plant homeodomain (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, 2012).

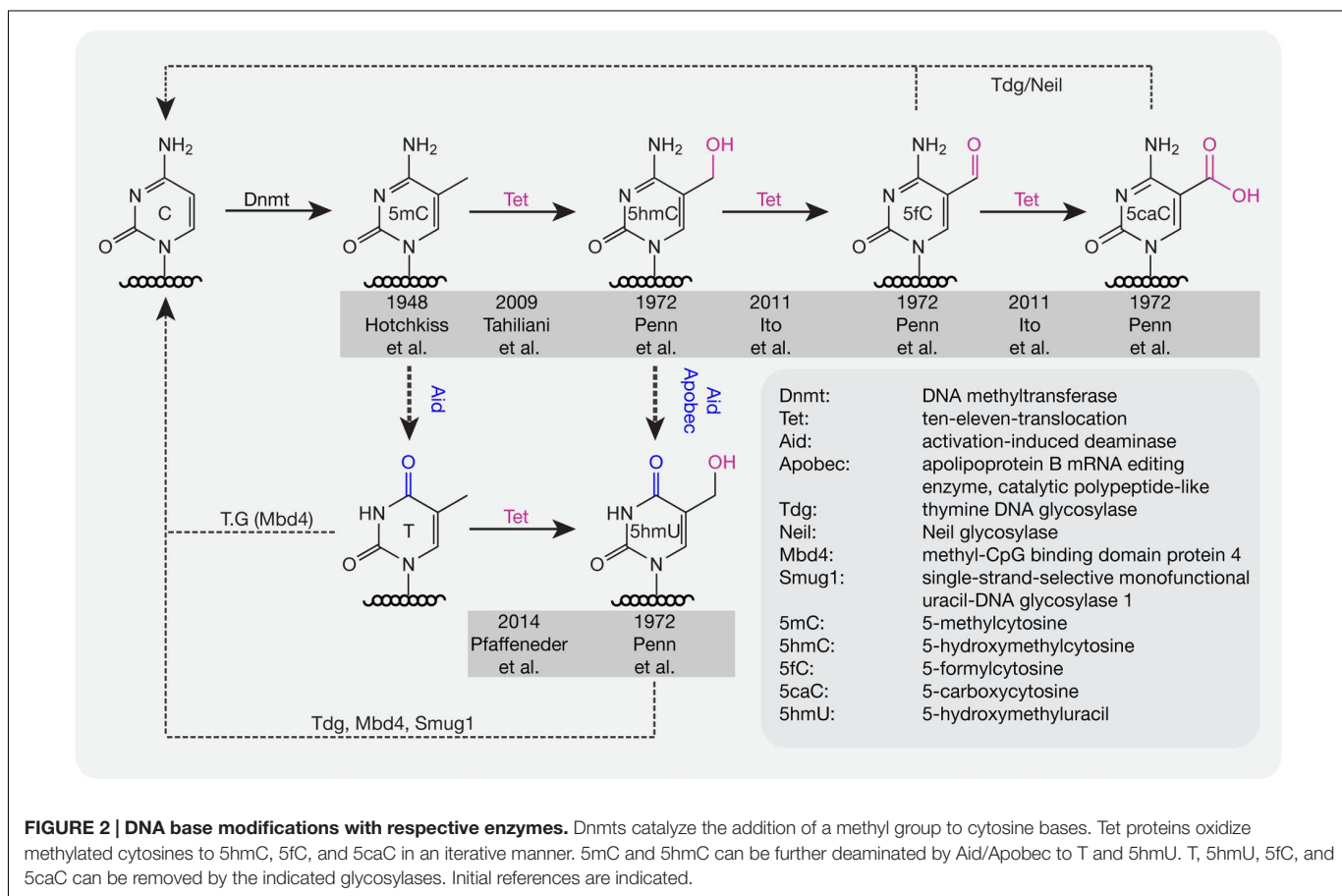
In most mouse cells, Dnmt1 localizes to the cell nucleus. In fact, Dnmt1 contains several functional nuclear localization sequences within its N-terminal regulatory domain (Cardoso and Leonhardt, 1999). In early embryos (Cardoso and Leonhardt, 1999) and in post-mitotic neurons (Inano et al., 2000) though, 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 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 et al., 2007; Schneider et al., 2013). In G2-phase, Dnmt1 is *de novo* loaded onto pericentromeric heterochromatin via a replication independent mechanism (Easwaran et al., 2004). Besides its PBD and TS domains, the polybromo-1 protein homologous domain (PBHD) is also involved in targeting Dnmt1 to replication foci (Liu 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).

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 et al., 2000) and development of primordial germ cells (PGC; Hajkova et al., 2002). The active loss of DNA methylation allows rapid reprogramming of the genome in a short time. Similar observations were made in post-mitotic 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 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 et al., 1972; Figure 2). However, these modifications were considered to be oxidative damage products of DNA (de Rojas-Walker et al., 1995; Tardy-Planechaud et al., 1997). Three decades later, 5hmC was re-discovered in mouse brain tissue (Kriaucionis and Heintz, 2009) and embryonic stem cells (ESCs; Tahiliani et al., 2009). Furthermore, a family of proteins (ten-eleven translocation, TET) was identified that oxidize 5mC to 5hmC both in humans (Tahiliani et al., 2009) and mice (Ito et al., 2010). TET1 was first described in 2003 as a fusion partner of the mixed lineage leukemia (MLL) gene in acute myeloid leukemia (AML; Lorschbach et al., 2003) and 6 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 et al., 2011; Pfaffeneder et al., 2011).

Deaminases such as Aid and Apobec can recognize 5mC and 5hmC and further convert 5mC to thymine (T) and 5hmC to 5hmU. Although the deaminase activity is quite low, it is still a possible pathway for DNA demethylation (Guo et al., 2011). In addition, Tets were also shown to oxidize T to 5hmU in mESCs (Pfaffeneder 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 and Drohat, 2011) and Neil (Muller 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 et al., 2012b) and Smug1 (Kemmerich 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 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.

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 cysteine-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).

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 and *Tet2* double knockout mice are viable, this suggests that *Tet1* and *Tet2* are not essential for mouse development (Table 2).

Tet-mediated 5mC to 5hmC conversion is though involved in reprogramming the paternal genome (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011; Zhang 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 et al., 2015) in mouse embryos. This process is involved in female germ cell development and genome reprogramming in zygotes (Yu et al., 2013).

During PGC reprogramming, Tet1 and Tet2 are highly expressed (Hackett et al., 2013). However, genome-wide DNA methylation removal is unaffected by the absence of Tet1 and Tet2 and, thus, 5hmC, 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 et al., 2013). Further studies showed that Tet1 has a critical role in the erasure of genomic imprinting (Yamaguchi et al., 2013) and it controls meiosis by regulating meiotic gene expression (Yamaguchi et al., 2012).

In mESCs, 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

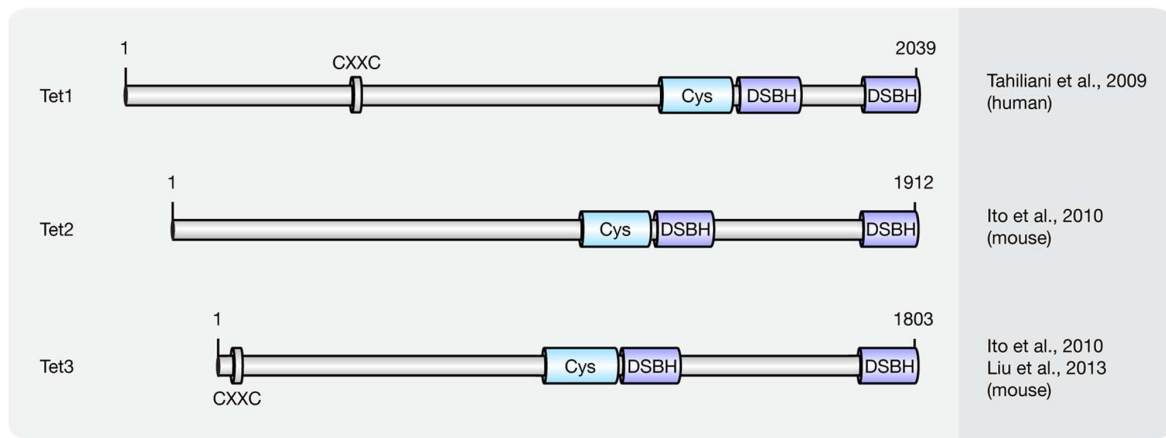


FIGURE 3 | Schematic representation of the Tet protein family. Shown are domain structures of mouse Tet proteins and the initial references. Numbers represent amino acid positions. CXXC, CXXC zinc finger domain; Cys, cysteine-rich domain; DSBH, double-stranded β -helix.

TABLE 2 | Phenotypes of initial Tet knockout mouse models.

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

promoters and transcription start sites (TSS; Huang 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 (Doege 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 et al., 2013). Accordingly, *Tet1* can replace *Oct4* during somatic cell reprogramming in conjunction with *Sox2*, *Klf4*, and *c-Myc* (Gao 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 and 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).

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

(Hu et al., 2013; Hashimoto et al., 2014) and oxidation of 5mC to 5hmC (Hashimoto et al., 2015; Hu et al., 2015). Although Tet proteins successively oxidize 5mC to 5caC, recent experimental data showed that, in comparison with 5hmC and 5fC, 5mC is the preferential substrate for *Tet2* (Hu et al., 2015). This preference was further confirmed by computer simulations (Lu et al., 2016). In cultured cells, the majority of genomic 5hmC nucleotides are stable (Bachman et al., 2014), indicating that 5hmC 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, does not correlate with 5mC content and rapidly decreases as the cells adapt to cell culture conditions (Nestor 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 (NTD) 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 et al., 2013). While the CXXC domain of *Tet1* cannot bind to DNA *in vitro* (Frauer et al., 2011b), it binds to unmodified C, 5mC- or 5hmC-modified CpGs *in vivo* (Zhang et al., 2010; Xu Y. et al., 2011). Moreover, binding of the CXXC domain to DNA was shown to control DNA methylation

levels by preventing unwanted DNA methyltransferase activity in ESCs (Xu Y. et al., 2011) or aberrant methylation spreading into CpG islands (CGIs) in differentiated cells (Jin 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 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 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 (PTMs), such as phosphorylation and O-GlcNAcylation were observed in the spacer region (Bauer 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 hydroxylation activity of Tet2 and Tet3, rather Tet2 and Tet3 were shown to promote Ogt activity (Deplus et al., 2013) by enhancing the localization of Ogt to chromatin (Chen et al., 2013; Ito et al., 2014). However, it was shown that Ogt drives Tet3 out of the nucleus further affecting its activity on DNA (Zhang et al., 2014). In mESCs, Ogt is recruited to unmethylated CpG promoters in a Tet1-dependent manner (Vella et al., 2013). In addition to PTMs, mutations within the spacer region of Tet2 were observed in myelodysplastic syndrome (MDS), thus further highlighting the importance of this region (Ko et al., 2010).

In vivo, besides PTMs, Tet activity is regulated by protein–protein interactions, such as with Sin3a. In mESCs, the interaction between Sin3a and Tet1 allows Sin3a to repress a subset of Tet1 target genes (Williams 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 et al., 2012) and interacts with Tet3 further blocking the activity of Tet3 (Bian and 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 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 et al., 2015; Li 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

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 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 et al., 2011).

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 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 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 et al., 2003) suggesting Dnmt3a/3b might play a role in maintaining 5hmC after DNA replication in repeat elements.

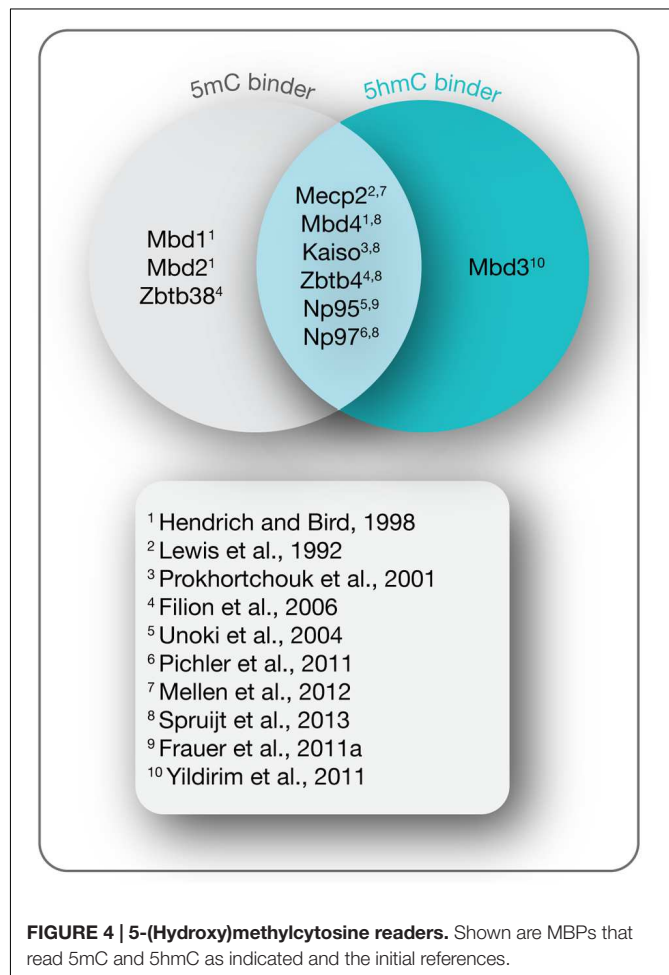
In vivo, the majority of 5hmC is present in CpG dinucleotides. However, 5hmC has also been observed in non-CpG context, especially in gene bodies (Pastor et al., 2011; Xu Y. et al., 2011). One important role of CpG methylation in gene promoter regions is the repression of gene expression by directly or indirectly preventing interactions between promoter and transcription factors. Hydroxymethylated CpGs might affect binding of transcription factors and/or 5mC readers to DNA.

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 (Lewis et al., 1992; Cross et al., 1997; Hendrich and Bird, 1998; Hendrich and Tweedie, 2003; Laget et al., 2010; Baymaz et al., 2014), the Kaiso protein family (Daniel and Reynolds, 1999; Filion et al., 2006), and the SET and RING (really interesting new gene) finger associated (SRA) domain protein family (Hopfner et al., 2000; Mori et al., 2002). While initially identified as 5mC binding proteins, recent studies indicate that a distinct and dynamic set of MBPs binds the Tet oxidation product 5hmC during differentiation (Figure 4; Frauer et al., 2011a; Mellen et al., 2012; Spruijt 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.

MBD Protein Family

Presently, the MBD protein family consists of eleven members (Mecp2, Mbd1–6, SETDB1, SETDB2, TIP5/BAZ2A, and BAZ2B; Lewis et al., 1992; Cross et al., 1997; Hendrich and Bird, 1998;



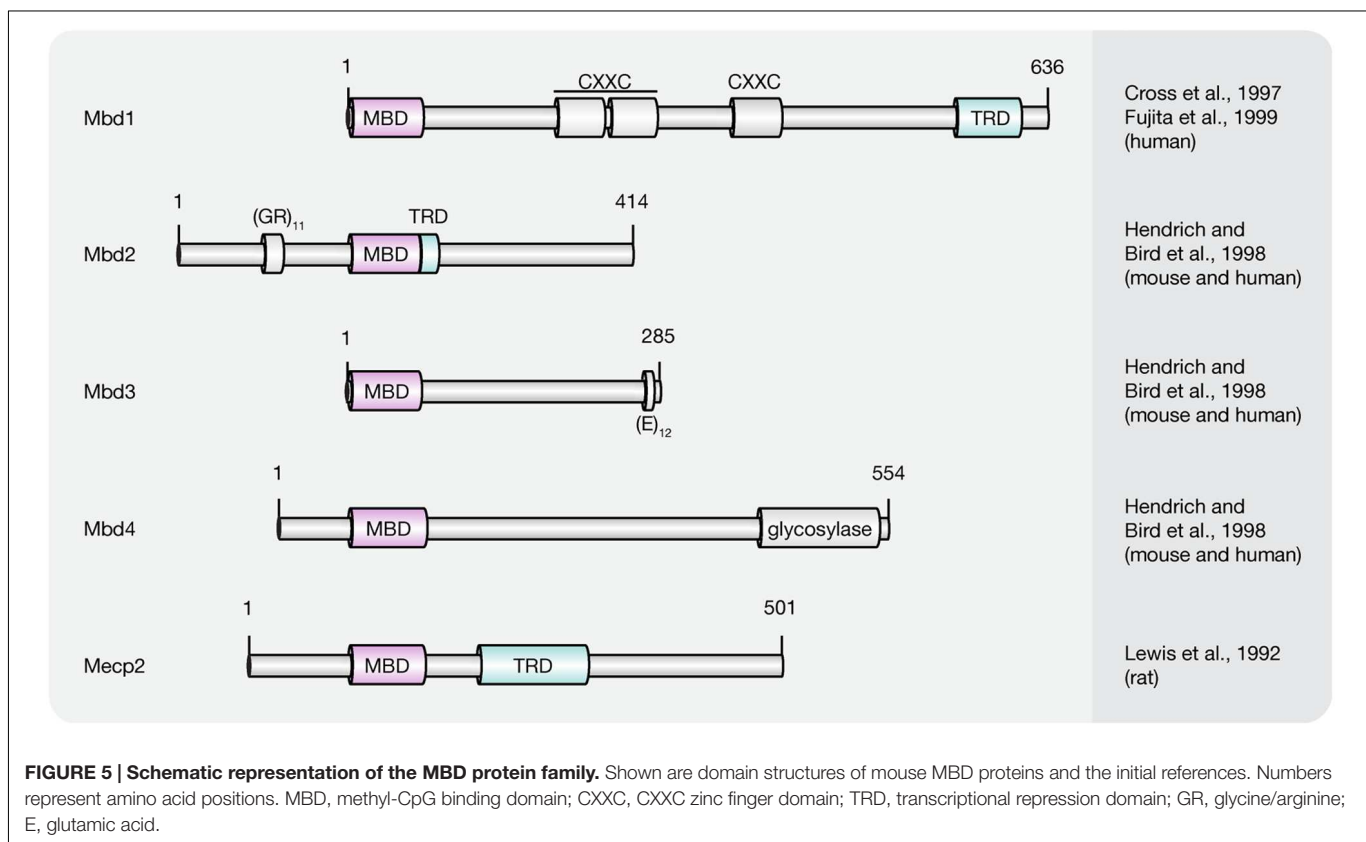
Hendrich and Tweedie, 2003; Laget et al., 2010; Baymaz et al., 2014). All of them share a common protein motif, the 70–85 amino acids long MBD, which enables some, but not all family members, to selectively bind to single methylated CpG dinucleotides. With the exception of Mbd2 and Mbd3, MBD proteins bear little resemblance outside their MBD (Hendrich and 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 dinucleotides, i.e., methyl-CpG binding protein 2 (Mecp2) and methyl-CpG binding domain proteins 1–4 (Mbd1–4).

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 et al., 2001; Traynor et al., 2002; Jung et al., 2003). 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 and

Bird, 2004; Mnatzakanian et al., 2004). Although both isoforms distribute differently in developing and post-natal mouse brains, no functional differences have been identified so far (Dragich et al., 2007).

Both Mecp2 variants include two functionally characterized domains, the MBD and the transcriptional repression domain (TRD). While the MBD proved sufficient to direct specific binding to methylated cytosines (Nan et al., 1993), the TRD was originally identified as the region required for transcriptional repression *in vitro* and *in vivo* (Lewis et al., 1992; Nan et al., 1997; Jones et al., 1998; Kaludov and Wolffe, 2000). Circular dichroism and protease digestion analysis revealed that outside these functional domains the full-length protein is largely devoid of secondary structure (Adams 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 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 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, as well as the C-terminal domain alpha (Ghosh et al., 2010). The 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 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 et al., 2007a), which bears strong resemblance to structures induced by H1 (Hamiche et al., 1996; Bednar et al., 1998). 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 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 (Brero et al., 2005; Agarwal et al., 2011). More recently, Szulwach et al. (2011) provided evidence that binding of Mecp2 to methylated CpG dinucleotides may protect 5mC against Tet-mediated oxidation thereby preventing reactivation of silenced genes. 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 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 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 (Skene et al., 2010; Cohen et al., 2011). A number of other repressive protein partners of Mecp2 have been identified including the co-repressors c-Ski (Kokura et al., 2001), CoREST (Lunyak et al., 2002), and NCoR/SMRT (Stancheva et al., 2003), as well as DNA methyltransferase Dnmt1 (Kimura and Shiota, 2003) and H3K9 methyltransferase (Fuks et al., 2003).

Both, binding of Mecp2 to DNA, as well as interactions with protein partners are affected by 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 and 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 et al., 2016). Furthermore, ubiquitylation (Gonzales et al., 2012), SUMOylation (Cheng et al., 2014), acetylation (Zocchi and Sassone-Corsi, 2012), and methylation (Jung 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 et al. (2013) who identified Mecp2 as reader of 5hmC in mESC by quantitative mass-spectrometry-based proteomics. Moreover, independent studies of Mellen et al. (2012), revealed Mecp2 as the major 5hmC-binding protein in mouse brain, which moreover turned out to bind both, 5hmC- and 5mC-containing substrates with similar affinity.

Finally, chip-chip analysis using antibodies against MECP2 in a human neuronal cell line demonstrated that around two-third of strongly MECP2 bound promoters were transcriptionally active (Yasui 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 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 3 and 8 weeks of age and most died between 6 and 12 weeks (Chen et al., 2001; Guy et al., 2001). Furthermore, most animals developed hind limb clasping, 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*^{lox-Stop/y} mice proved sufficient to reverse the neurological symptoms of Rett syndrome (RTT), indicating that *Mecp2*-deficient neurons develop normally and are not irreversibly damaged (Guy et al., 2007). Further microarray analyses revealed that knockout of *Mecp2* implicates only minor changes in gene expression (Tudor et al., 2002). Subsequent studies demonstrating increased expression restricted to non-coding RNA in brain of *Mecp2*-deficient mice (Muotri 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).

Mbd1

Mbd1, initially termed PCM1, is expressed in somatic cells and represents the largest member of the MBD family (Cross et al., 1997; Hendrich and Bird, 1998). Similar to *Mecp2*, *Mbd1* contains a MBD and a TRD, which have analog functions to that of *Mecp2* (Ng et al., 2000). In addition, depending on the isoform, *Mbd1* contains two or three CXXC zinc finger motifs (Fujita et al., 1999; Jorgensen et al., 2004). The most C-terminal one, referred to as CXXC3, is homolog to zinc fingers found in Dnmt1, CpG binding protein CGBP, histone H3K4 methyltransferase MLL and histone H3K36 deacetylases of the Jumonji family JHDM1A and JHDM1B (Jorgensen et al., 2004; Lee and Skalnik, 2005; Tsukada et al., 2006). While CXXC3 was shown to bind unmethylated CpG dinucleotides *in vitro* (Birke et al., 2002; Lee and Skalnik, 2002, 2005; Jorgensen et al., 2004), 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 (Jorgensen et al., 2004; Baubec et al., 2013).

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 non-methylated 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).

Instead, MBD1 has been found associated with histone H3K9 methyltransferases SETDB1 (Sarraf and Stancheva, 2004) and Suv39h1 (Fujita et al., 2003). Association to SETDB1 mediates transcriptional repression throughout the cell cycle (Sarraf and 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 and Stancheva, 2004). Accordingly, H3K9 methylation is lost in the absence of MBD1 and results in activation of specific genes, such as p53BP2 (Sarraf and Stancheva, 2004).

MBD1-mediated transcriptional repression and heterochromatin maintenance was shown to be regulated by SUMOylation (Lyst et al., 2006; Uchimura et al., 2006). In human cells, two E3 SUMO-ligases (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

TABLE 3 | Phenotype of initial MBP-deficient mouse models.

Genotype	Phenotype	Reference
<i>Mecp2</i> null	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.	Chen et al., 2001; Guy et al., 2001
<i>Mbd1</i> null	Viable and fertile. Impaired spatial learning, decreased neurogenesis, reduced long-term potentiation, decreased genomic stability.	Zhao et al., 2003
<i>Mbd2</i> null	Viable, fertile. Maternal nurturing defects: reduced litter size and weight of pups.	Hendrich et al., 2001
<i>Mbd3</i> null	Early embryonic lethality	Hendrich et al., 2001
<i>Mbd4</i> null	Viable and fertile. Increased number of C:G to T:A transitions at CpG sites.	Millar et al., 2002
<i>Kaiso</i> null	Viable and fertile. Reduced tumorigenesis	Prokhorchouk et al., 2006
<i>Np95</i> null	Early gestational lethality. Developmental arrest shortly after gastrulation.	Sharif et al., 2007
<i>Np97</i> null	Phenotype not described.	Li et al., 2013

(Zhao et al., 2003). Moreover, Mbd1-deficient neural stem cells differentiated less and had decreased genomic stability (Zhao et al., 2003).

Mbd2

Mbd2 and Mbd3 are the only known members of the MBD protein family with significant sequence similarity beyond the MBD (Hendrich and Bird, 1998) and, thus, are believed to have arisen from an ancient duplication during evolution of the vertebrate lineage (Hendrich and Tweedie, 2003). Consistent with this, a homolog Mbd2/3 like protein was identified in invertebrates, including *Drosophila* (Lyko et al., 2000; Marhold 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 and 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 (Hendrich and Bird, 1998; Gnanapragasam et al., 2011; Becker et al., 2013).

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 and Bird, 1998). *In vivo*, however, only Mbd2a, but not Mbd2b, has been detected (Ng 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 and 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 et al., 2000). Different from other MBD family members, the sequence required for 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 HDAC1 and HDAC2 (Ng et al., 1999; Wade et al., 1999; Zhang et al., 1999; Mahajan et al., 2005; Le Guezennec et al., 2006). EMSAs 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 MeCP1 complex (Meehan et al., 1989; Ng et al., 1999).

Mbd2 was shown to bind 5mC in a manner similar to the isolated MBD of Mecp2 (Hendrich and Bird, 1998; Wade et al., 1999). Binding of oxidative 5mC derivatives, however, has not been observed (Hashimoto et al., 2012a; Mellen et al., 2012; Spruijt et al., 2013).

Mbd2b has also been reported to have DNA demethylase activity (Bhattacharya et al., 1999), but this finding has been questioned (Ng et al., 1999; Wade et al., 1999).

Mbd2-deficient mice (**Table 3**) are viable and fertile, but exhibit a maternal nurturing defect resulting in reduced litter size and weight of pups (Hendrich et al., 2001).

Mbd3

The smallest member of the MBD family, coding for a protein of approximately 30 kDa is Mbd3 (Hendrich and Bird, 1998). It appears in a rich diversity of splice variants and is expressed in ESCs as well as somatic tissues (Hendrich and Bird, 1998; Roloff 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 and Bird, 1998; Wade et al., 1999; Saito and Ishikawa, 2002). 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 et al., 1999; Wakefield et al., 1999; Saito and Ishikawa, 2002).

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 (Zhang et al., 1999; Kaji et al., 2006). ESCs lacking Mbd3-NuRD displayed a severe defect in differentiation that lead to persistent self-renewal even in the absence of leukemia inhibitory factor (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, and Myc; Luo et al., 2013; Rais et al., 2013). Accordingly, Mbd3 was proposed to play a key role in lineage commitment and pluripotency (Yildirim et al., 2011; Reynolds et al., 2012; Whyte et al., 2012). Contradictory studies using neural and epiblast-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 et al., 2014).

Binding sites of Mbd3 have been mapped genome-wide in mouse and human cells (Yildirim et al., 2011; Baubec et al., 2013; Gunther et al., 2013; Shimbo et al., 2013). While Yildirim et al. (2011) identified Mbd3 bound to TSS of CpG-rich, hydroxymethylation marked promoters, Baubec et al. (2013) found Mbd3 bound to enhancers independent of CpG density and (hydroxy)methylation status. Further data questioning the interaction of Mbd3 with hydroxymethylated DNA was provided by Spruijt et al. (2013), who did not detect Mbd3 among hydroxymethylation-specific readers.

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**).

Mbd4

Mbd4, also referred to as MED1 (Bellacosa et al., 1999), is the only known member of the MBD protein family not associated with HDAC activity (Hendrich and Bird, 1998). Instead, several lines of evidence suggest that Mbd4 plays a role in DNA repair

(Bader et al., 1999; Bellacosa et al., 1999; Hendrich et al., 1999; Riccio et al., 1999; Petronzelli et al., 2000; Millar et al., 2002). 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 BER (Michaels et al., 1990; Hendrich and Bird, 1998). 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 (Hendrich et al., 1999; Petronzelli et al., 2000; Hashimoto et al., 2012b). 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 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.

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 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 and Reynolds, 1999; Kiefer et al., 2005; Park et al., 2005; Filion et al., 2006).

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 (Prokhortchouk et al.,

2001; Filion et al., 2006). 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 and 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 and 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 et al., 2004), as it competes with DNA for the access to the Kaiso zinc finger domains (Daniel et al., 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.

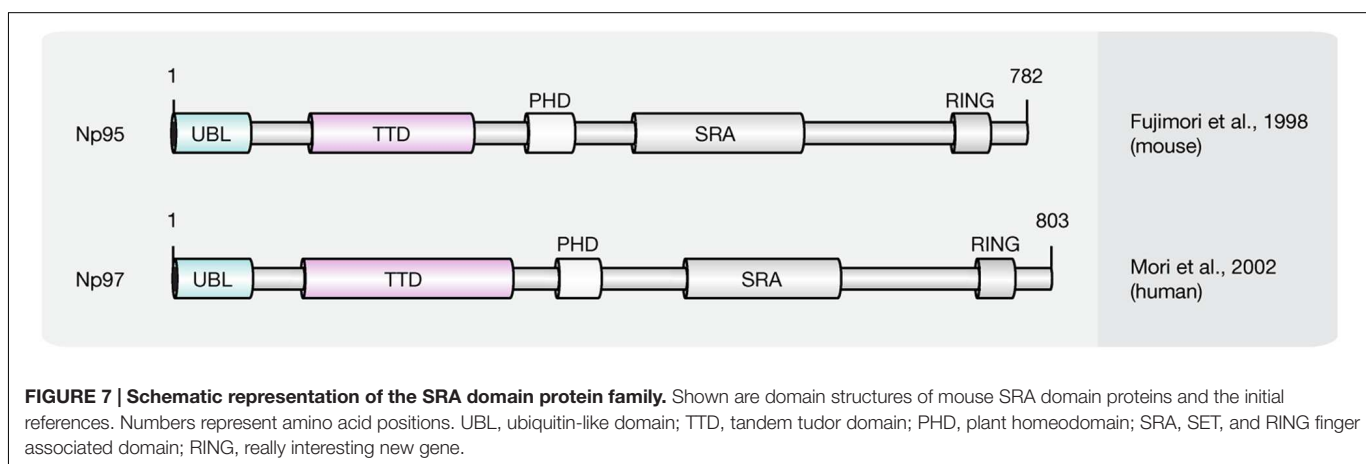
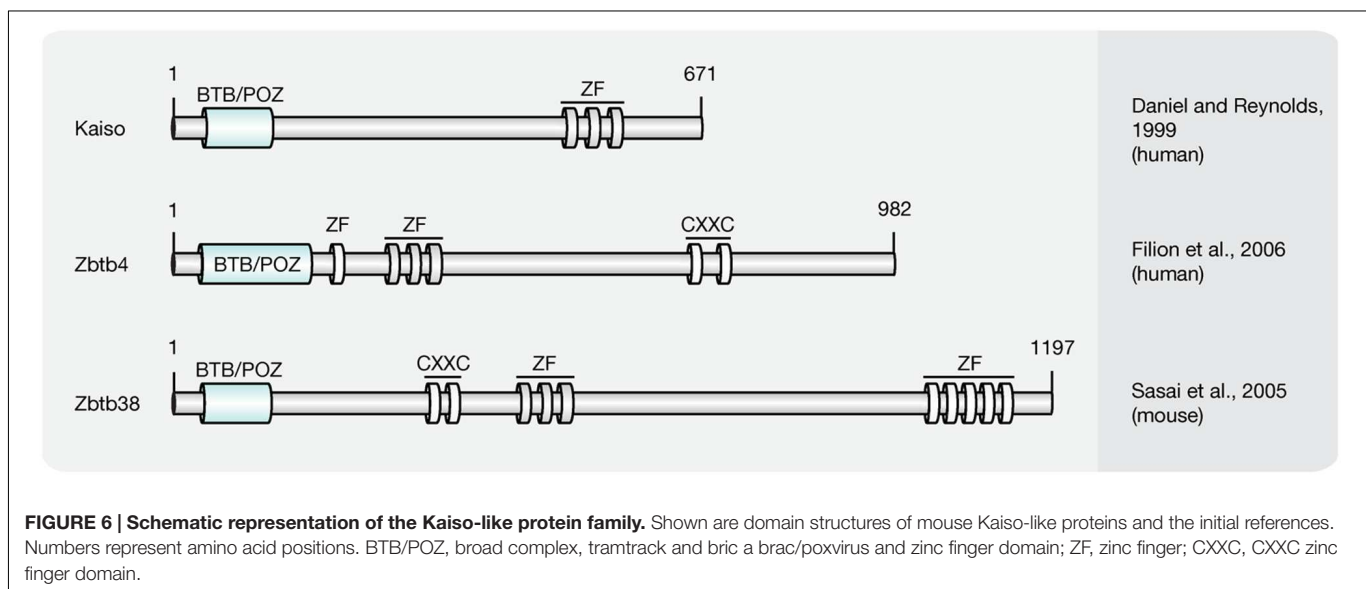
Kaiso was shown to recruit the NCoR complex to promoters of target genes to introduce histone hypoacetylation, as well as H3K9 methylation (Yoon et al., 2003). Moreover, Kaiso was identified as component of an alternative MeCP1 complex in NIH3T3 cells (Prokhortchouk et al., 2001). Zbtb38 was found to interact with the co-repressors CtBPs (C-terminal binding proteins), which include HDAC, methyltransferase, and demethylase activities (Sasai et al., 2005; Zocchi and Sassone-Corsi, 2012). Zbtb4 was shown to associate with the Sin3A/HDAC complex to repress expression of p21^{CIP1} in response to stimuli that activate p53 (Weber et al., 2008).

Kaiso-like proteins exhibit diverging expression patterns. While Kaiso is ubiquitously expressed, *Zenon*, the rat homolog 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).

SRA Domain Protein Family

Recent studies implicate that yet another protein fold, the SRA domain could read DNA (hydroxy)methylation marks *in vitro* and *in vivo* (Unoki et al., 2004; Johnson et al., 2007; Woo et al., 2007; Frauer et al., 2011a; Spruijt et al., 2013). In mammals, two SRA domain-containing proteins (Figure 7), Np95 (mouse homolog of human ICBP90, gene name *UHRF1*) and Np97 (mouse homolog of human NIRE, gene name *UHRF2*), have been characterized (Unoki et al., 2004; Woo et al., 2007; Zhang et al., 2011). While Np95 was first discovered during the generation process of antibodies against murine



thymic lymphoma (Fujimori 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 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 et al., 2008; Papait et al., 2008; Rottach et al., 2010); a PHD, which binds (un)modified histones; and a C-terminal RING, which exhibits ubiquitin E3 ligase activity.

ICBP90 and Np95 play a critical role in epigenetic inheritance and maintenance of DNA methylation (Bostick et al., 2007; Sharif et al., 2007). Accordingly, ICBP90/Np95 was shown to colocalize with PCNA during S phase and to interact with Dnmt3a, Dnmt3b and several histone-modifying enzymes like HDAC1, as well as histone methyltransferase G9a (Achour et al., 2009; Kim et al., 2009; Meilinger et al., 2009). Moreover, besides its ability to bind and flip out hemi-methylated DNA, the SRA domain of ICBP90 was shown to target Dnmt1 to replicating pericentric

heterochromatin for maintenance methylation (Bostick et al., 2007; Arita et al., 2008; Avvakumov et al., 2008; Hashimoto 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 (Rottach et al., 2010; Nady et al., 2011; Rothbart et al., 2012). The PHD of ICBP90, on the other hand, was found associated with the N-terminal tail of histone H3 (Papait et al., 2007; Hu et al., 2011; Rajakumara et al., 2011; Wang et al., 2011; Arita et al., 2012; Cheng et al., 2013). 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 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. (2013), 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.

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 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**).

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 (Kumar et al., 1994; Hendrich and Bird, 1998; Mori et al., 2002; Filion et al., 2006; Bostick et al., 2007; Ito et al., 2010; Qin et al., 2011), we, hereafter, summarize the state-of-the-art regarding the role of the human orthologs of the aforementioned mouse Dnmts, Tets, and MBPs in human diseases.

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 et al., 2011) and autosomal

TABLE 4 | Comparison of human proteins and their mouse orthologs.

Mouse protein	Human protein	Amino acid similarity (%)
Dnmt1 (1620 aa)	DNMT1 (1632 aa)	76
Dnmt2 (415 aa)	DNMT2 (391 aa)	77
Dnmt3a (908 aa)	DNMT3A (912 aa)	96
Dnmt3b (860 aa)	DNMT3B (853 aa)	80
Dnmt3l (421 aa)	DNMT3L (387 aa)	56
Tet1 (2039 aa)	TET1 (2136 aa)	50
Tet2 (1912 aa)	TET2 (2002 aa)	55
Tet3 (1803 aa)	TET3 (1795 aa)	89
Mecp2 (501 aa)	MECP2 (498 aa)	94
Mbd1 (636 aa)	MBD1 (605 aa)	68
Mbd2 (414 aa)	MBD2 (411 aa)	94
Mbd3 (285 aa)	MBD3 (291 aa)	92
Mbd4 (554 aa)	MBD4 (580 aa)	58
Kaiso (671 aa)	KAISO (672 aa)	84
Zbtb4 (982 aa)	ZBTB4 (1013 aa)	85
Zbtb38 (1197 aa)	ZBTB38 (1195 aa)	81
Np95 (782 aa)	ICBP90 (806 aa)	72
Np97 (803 aa)	NIRF (802 aa)	90

dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN; Winkelmann et al., 2012). Mutations of Y495C, Y495H, D490E-P491Y (Klein et al., 2011, 2013) in exon 20 cause HSAN1E. Those mutations caused premature degradation of mutant proteins, reduced methyltransferase activity and impaired heterochromatin binding during G2 phase leading to global hypomethylation and site-specific hypermethylation (Klein et al., 2011). ADCA-DN is a polymorphic disorder first described in 1995 in a Swedish pedigree. Unlike mutations in HSAN1E located in exon 20, mutations in ADCA-DN including A570V, G605A, and V606F were found in exon 21 of the *DNMT1* gene.

Mutations in DNMT3A were found in *de novo* AML and are associated with poor survival (**Table 5**; Ley et al., 2011). The most frequent mutation occurred in amino acid R882, however, frameshift, nonsense and splice site mutations were also reported (Ley et al., 2011). Mutations of DNMT3A are not only observed in AML patients, but also in MDS. Similar to mutations leading to AML, amino acid R882 located in the methyltransferase domain of DNMT3A is the most common mutation site (Walter 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 et al., 2014).

ICF syndrome (immunodeficiency, chromosomal instability, and facial anomalies), a human genetic disorder is caused by DNMT3B mutations (**Table 5**; Hansen et al., 1999; Xu et al., 1999). 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

TABLE 5 | Summary of disease-related DNMT and TET mutations.

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

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 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 CGIs, which are silenced in somatic tissues by DNA methylation (Strichman-almashanu 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 et al., 2004). Although the cancer genome is hypomethylated, several studies showed that Dnmts are upregulated in cancer cells (Ahluwalia et al., 2001; Lin et al., 2007; Roll et al., 2008), suggesting that demethylation enzymes might be additionally involved in loss of DNA methylation in cancer.

TET Proteins in Disease

MLL gene is located in 11q23 and is the most frequent cytogenetic finding in 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 5hmC, thus playing an oncogenic role (Huang et al., 2013).

In myeloproliferative neoplasms, 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 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 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 et al., 2011; Jin et al., 2011; Li and Liu, 2011; Xu W. et al., 2011; Kraus et al., 2012).

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 and Lovell, 2013). This indicates that DNA methylation might play an important role in memory-related disease.

MBPs in Disease

As readers and translators of epigenetic information, alterations in MBP 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 RTT (Table 6), a late onset (6–18 months post-birth) debilitating neurological disease that affects 1 in 10,000–15,000 female live births (Hagberg et al., 1983; Amir et al., 1999). 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 (Rett, 1966; Hagberg et al., 1983).

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

TABLE 6 | Summary of disease-related MBP alterations.

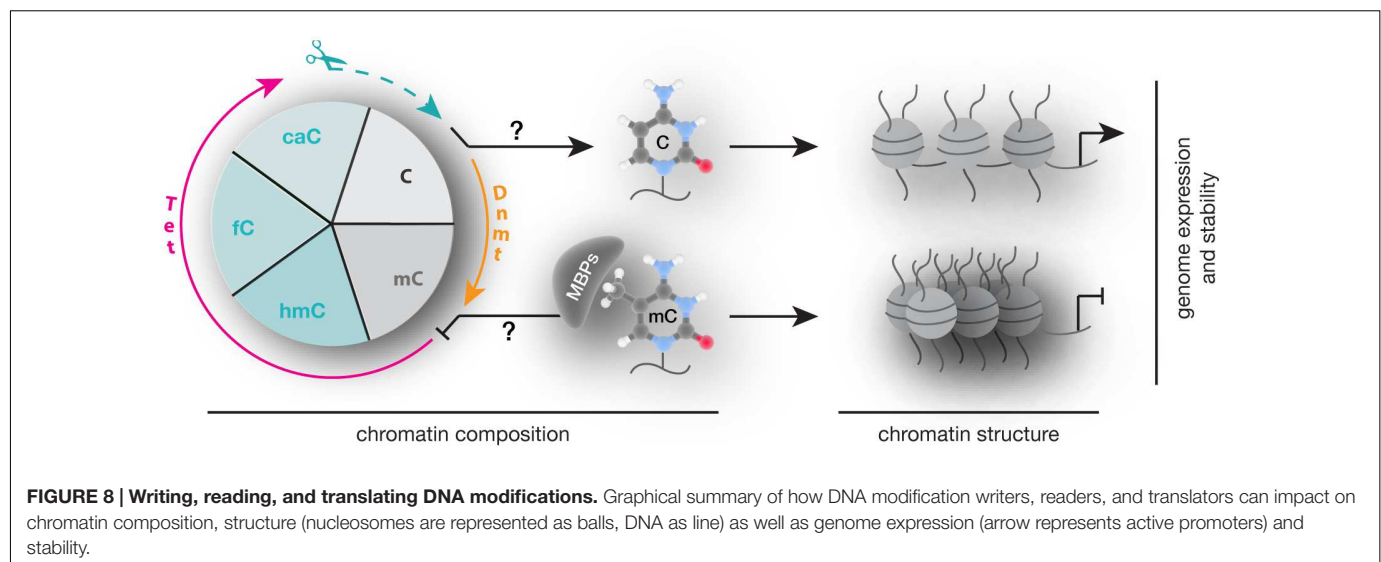
Protein	Disease	Alteration	Reference
MECP2	Rett syndrome	Causal MECP2 mutations of Rett syndrome are summarized in: http://mecp2.chw.edu.au/mecp2/index.php	Amir et al., 1999
MBD1	Prostate cancer	Upregulated	Patra et al., 2003
MBD2	Breast cancer	Upregulated	Billard et al., 2002
MBD3	Glioblastoma	Upregulated	Schlegel et al., 2002
MBD4	Colorectal cancer Endometrial cancer Pancreas cancer	Frameshift mutation Frameshift mutation Frameshift mutation	Riccio et al., 1999
KAISO	Colorectal cancer	Upregulated	Lopes et al., 2008
ZBTB4	Neuroblastoma	Downregulated	Weber et al., 2008
ICBP90	Non-small-cell lung cancer	Upregulated	Daskalos et al., 2011
NIRF	Lung cancer	Upregulated	He et al., 2009

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 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 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. (2012) revealed that some of these mutations exclusively affect the chromatin binding but not linking ability. 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 PTM 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 et al., 2013; Lyst 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 et al., 2002). Histone H3, however, was hyperacetylated indicating abnormal chromatin architecture and misregulated gene expression (Shahbazian et al., 2002). Moreover, Muotri et al. (2010) identified increased susceptibility for L1 transposition and genome instability in RTT patients with truncating mutations.

In addition to RTT, Mecp2 was implicated in other neurological diseases, including Hirschsprung's disease, autism spectrum disorder, schizophrenia, Prader-Willi, and Angelman syndromes (Carney et al., 2003; Shibayama et al., 2004; Nagarajan et al., 2006; Loat et al., 2008; Ramocki et al., 2009; Zhou et al., 2013).

More recently MBP have been associated with several types of human cancers (Table 6). While Mecp2 was overexpressed in estrogen receptor positive human breast cancer (Muller et al., 2003), MBD1 mRNA and protein levels were increased in



prostate cancer (Patra et al., 2003). Accordingly, Patra et al. (2003) proposed MBD1 as the major cause of hypermethylated chromatin regions in prostate cancer through the recruitment of HDAC1/2 and subsequent histone deacetylation. MBD2 mRNA level were 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 (Ricchio 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 (Ricchio et al., 1999). As a regulator of target genes of the canonical and non-canonical 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 et al. (2008) suggested Kaiso as a methylation-dependent oncogene that represses hypermethylated tumor suppressor genes. 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^{CIP1}* 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 CGIs 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 NIFK and reduced P21^{CIP1} protein level, thereby most likely contributing to lung carcinogenesis (He et al., 2009).

CONCLUDING REMARKS

In summary, alterations in 5mC writers, readers, and modifiers that affect their level, PTMs, 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.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Our research has been supported by grants of the German Research Foundation (DFG), the Volkswagen Foundation, and the German Ministry for Education and Research (BMBF). PZ was funded by a fellowship of the Chinese Scholarship Council.

ACKNOWLEDGMENTS

We apologize to the colleagues whose work was not cited due to space constraints. We thank all the past and present members of our laboratory for their many contributions along the years. Last but not least, we thank our collaborators over the years, which have made our work so much more enjoyable.

REFERENCES

- Aapola, U., Kawasaki, K., Scott, H. S., Ollila, J., Vihinen, M., Heino, M., et al. (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* 65, 293–298. doi: 10.1006/geno.2000.6168
- Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., et al. (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* 114, 144–147. doi: 10.1182/blood-2009-03-210039
- Achour, M., Fuhrmann, G., Alhosin, M., Ronde, P., Chataigneau, T., Mousli, M., et al. (2009). UHRF1 recruits the histone acetyltransferase Tip60 and controls its expression and activity. *Biochem. Biophys. Res. Commun.* 390, 523–528. doi: 10.1016/j.bbrc.2009.09.131
- Adams, V. H., McBryant, S. J., Wade, P. A., Woodcock, C. L., and Hansen, J. C. (2007). Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem.* 282, 15057–15064. doi: 10.1074/jbc.M700855200
- Agarwal, N., Becker, A., Jost, K. L., Haase, S., Thakur, B. K., Brero, A., et al. (2011). MeCP2 Rett mutations affect large scale chromatin organization. *Hum. Mol. Genet.* 20, 4187–4195. doi: 10.1093/hmg/ddr346
- Ahluwalia, A., Hurteau, J. A., Bigsby, R. M., and Nephew, K. P. (2001). DNA methylation in ovarian cancer. II. Expression of DNA methyltransferases in ovarian cancer cell lines and normal ovarian epithelial cells. *Gynecol. Oncol.* 82, 299–304. doi: 10.1006/gyno.2001.6284
- Akbarian, S., Chen, R. Z., Gribnau, J., Rasmussen, T. P., Fong, H., Jaenisch, R., et al. (2001). Expression pattern of the Rett syndrome gene MeCP2 in primate prefrontal cortex. *Neurobiol. Dis.* 8, 784–791. doi: 10.1006/nbdi.2001.0420
- Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188. doi: 10.1038/13810

- Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W. S., D'Souza, Z., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* 18, 225–233. doi: 10.1038/ncb3296
- Aoki, A., Suetake, I., Miyagawa, J., Fujio, T., Chijiwa, T., Sasaki, H., et al. (2001). Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* 29, 3506–3512. doi: 10.1093/nar/29.17.3506
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455, 818–821. doi: 10.1038/nature07249
- Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., et al. (2012). Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12950–12955. doi: 10.1073/pnas.1203701109
- Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C., et al. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455, 822–825. doi: 10.1038/nature07273
- Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* 6, 1049–1055. doi: 10.1038/nchem.2064
- Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M., et al. (1999). Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* 18, 8044–8047. doi: 10.1038/sj.onc.1203229
- Ballestar, E., Yusufzai, T. M., and Wolffe, A. P. (2000). Effects of Rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry* 39, 7100–7106. doi: 10.1021/bi0001271
- Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* 153, 480–492. doi: 10.1016/j.cell.2013.03.011
- Bauer, C., Göbel, K., Nagaraj, N., Colantuoni, C., Wang, M., Müller, U., et al. (2015). Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-Linked N-Acetylglucosamine transferase (OGT). *J. Biol. Chem.* 290, 4801–4812. doi: 10.1074/jbc.M114.605881
- Baymaz, H. I., Fournier, A., Laget, S., Ji, Z., Jansen, P. W., Smits, A. H., et al. (2014). MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics* 14, 2179–2189. doi: 10.1002/pmic.201400013
- Becker, A., Allmann, L., Hofstätter, M., Casa, V., Weber, P., Lehmkuhl, A., et al. (2013). Direct homo- and hetero-interactions of MeCP2 and MBD2. *PLoS ONE* 8:e53730. doi: 10.1371/journal.pone.0053730
- Becker, A., Zhang, P., Allmann, L., Meilinger, D., Bertulat, B., Eck, D., et al. (2016). Poly(ADP-ribosyl)ation of Methyl CpG Binding Domain Protein 2 regulates chromatin structure. *J. Biol. Chem.* 291, 4873–4881. doi: 10.1074/jbc.M115.698357
- Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., et al. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14173–14178. doi: 10.1073/pnas.95.24.14173
- Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., et al. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3969–3974. doi: 10.1073/pnas.96.7.3969
- Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. *J. Mol. Biol.* 203, 971–983. doi: 10.1016/0022-2836(88)90122-2
- Bhattacharya, S. K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* 397, 579–583. doi: 10.1038/17533
- Bian, C., and Yu, X. (2014). PGC7 suppresses TET3 for protecting DNA methylation. *Nucleic Acids Res.* 42, 2893–2905. doi: 10.1093/nar/gkt1261
- Billard, L. M., Magdinier, F., Lenoir, G. M., Frappart, L., and Dante, R. (2002). MeCP2 and MBD2 expression during normal and pathological growth of the human mammary gland. *Oncogene* 21, 2704–2712. doi: 10.1038/sj/onc/1205357
- Birke, M., Schreiner, S., Garcia-Cuellar, M. P., Mahr, K., Titgemeyer, F., and Slany, R. K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res.* 30, 958–965. doi: 10.1093/nar/30.4.958
- Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martinez, J. A., Goyal, P., Mahapatra, S., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500, 222–226. doi: 10.1038/nature12362
- Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U., and Renkawitz, R. (2000). The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. *J. Biol. Chem.* 275, 34963–34967. doi: 10.1074/jbc.M005929200
- Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., and Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764. doi: 10.1126/science.1147939
- Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., and Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science (New York)* 294, 2536–2539. doi: 10.1126/science.1065848
- Bradley-Whitman, M. A., and Lovell, M. A. (2013). Epigenetic changes in the progression of Alzheimer's disease. *Mech. Ageing Dev.* 134, 486–495. doi: 10.1016/j.mad.2013.08.005
- Brero, A., Easwaran, H. P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H., et al. (2005). Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.* 169, 733–743. doi: 10.1083/jcb.20002062
- Cardoso, M. C., and Leonhardt, H. (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. *J. Cell Biol.* 147, 25–32. doi: 10.1083/jcb.147.1.25
- Carney, R. M., Wolpert, C. M., Ravan, S. A., Shahbazian, M., Ashley-Koch, A., Cuccaro, M. L., et al. (2003). Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr. Neurol.* 28, 205–211. doi: 10.1016/S0887-8994(02)00624-0
- Casas-Delucchi, C. S., Becker, A., Bolius, J. J., and Cardoso, M. C. (2012). Targeted manipulation of heterochromatin rescues MeCP2 Rett mutants and re-establishes higher order chromatin organization. *Nucleic Acids Res.* 40, e176. doi: 10.1093/nar/gks784
- Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T., Qin, J., et al. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320, 1224–1229. doi: 10.1126/science.1153252
- Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561–564. doi: 10.1038/nature11742
- Chen, R. Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* 27, 327–331. doi: 10.1038/85906
- Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell Biol.* 23, 5594–5605. doi: 10.1128/MCB.23.16.5594
- Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J. Biol. Chem.* 277, 38746–38754. doi: 10.1074/jbc.M205312200
- Cheng, J., Huang, M., Zhu, Y., Xin, Y. J., Zhao, Y. K., Huang, J., et al. (2014). SUMOylation of MeCP2 is essential for transcriptional repression and hippocampal synapse development. *J. Neurochem.* 128, 798–806. doi: 10.1111/jnc.12523
- Cheng, J., Yang, Y., Fang, J., Xiao, J., Zhu, T., Chen, F., et al. (2013). Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. *J. Biol. Chem.* 288, 1329–1339. doi: 10.1074/jbc.M112.415398
- Christodoulou, J., Grimm, A., Maher, T., and Bennets, B. (2003). RettBASE: the IRSA MECP2 variation database—a new mutation database in evolution. *Hum. Mutat.* 21, 466–472. doi: 10.1002/humu.10194
- Cohen, S., Gabel, H. W., Hemberg, M., Hutchinson, A. N., Sadacca, L. A., Ebert, D. H., et al. (2011). Genome-wide activity-dependent MeCP2 phosphorylation

- regulates nervous system development and function. *Neuron* 72, 72–85. doi: 10.1016/j.neuron.2011.08.022
- Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliha, P. V., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495, 370–374. doi: 10.1038/nature11925
- Cross, S. H., Meehan, R. R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat. Genet.* 16, 256–259. doi: 10.1038/ng0797-256
- Daniel, J. M., and Reynolds, A. B. (1999). The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol. Cell. Biol.* 19, 3614–3623. doi: 10.1128/MCB.19.5.3614
- Daniel, J. M., Spring, C. M., Crawford, H. C., Reynolds, A. B., and Baig, A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res.* 30, 2911–2919. doi: 10.1093/nar/gkf398
- Daskalos, A., Oleksiewicz, U., Filia, A., Nikolaidis, G., Xinarianos, G., Gosney, J. R., et al. (2011). UHRF1-mediated tumor suppressor gene inactivation in nonsmall cell lung cancer. *Cancer* 117, 1027–1037. doi: 10.1002/cncr.25531
- Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Inmaculada, M., Cheng, A. W., et al. (2014). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* 24, 310–323. doi: 10.1016/j.devcel.2012.12.015
- Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y. C., Markoulaki, S., Cheng, A. W., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9, 166–175. doi: 10.1016/j.stem.2011.07.010
- de Rojas-Walker, T., Tamir, S., Ji, H., Wishnok, J. S., and Tannenbaum, S. R. (1995). Nitric oxide induces oxidative damage in addition to deamination in macrophage DNA. *Chem. Res. Toxicol.* 8, 473–477. doi: 10.1021/tx00045a020
- Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Mendez, J., Murphy, N., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* 32, 645–655. doi: 10.1038/emboj.2012.357
- Doerge, C. A., Inoue, K., Yamashita, T., Rhee, D. B., Travis, S., Fujita, R., et al. (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* 488, 652–655. doi: 10.1038/nature11333
- dos Santos, R. L., Tosti, L., Radziszewska, A., Caballero, I. M., Kaji, K., Hendrich, B., et al. (2014). MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. *Cell Stem Cell* 15, 102–110. doi: 10.1016/j.stem.2014.04.019
- Dragich, J. M., Kim, Y. H., Arnold, A. P., and Schanen, N. C. (2007). Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. *J. Comp. Neurol.* 501, 526–542. doi: 10.1002/cne.21264
- Easwaran, H. P., Schermelleh, L., Leonhardt, H., and Cardoso, M. C. (2004). Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.* 5, 1181–1186. doi: 10.1038/sj.embor.7400295
- Ebert, D. H., Gabel, H. W., Robinson, N. D., Kastan, N. R., Hu, L. S., Cohen, S., et al. (2013). Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature* 499, 341–345. doi: 10.1038/nature12348
- Filion, G. J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P. A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol. Cell. Biol.* 26, 169–181. doi: 10.1128/MCB.26.1.169-181.2006
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., et al. (2011a). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* 6:e21306. doi: 10.1371/journal.pone.0021306
- Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., et al. (2011b). Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE* 6:e16627. doi: 10.1371/journal.pone.0016627
- Fujimori, A., Matsuda, Y., Takemoto, Y., Hashimoto, Y., Kubo, E., Araki, R., et al. (1998). Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell proliferation. *Mamm. Genome* 9, 1032–1035. doi: 10.1007/s003359900920
- Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., et al. (1999). Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. *Mol. Cell. Biol.* 19, 6415–6426. doi: 10.1128/MCB.19.9.6415
- Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., et al. (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *J. Biol. Chem.* 278, 24132–24138. doi: 10.1074/jbc.M302283200
- Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., and Kouzarides, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* 278, 4035–4040. doi: 10.1074/jbc.M210256200
- Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., et al. (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* 12, 453–469. doi: 10.1016/j.stem.2013.02.005
- Georgel, P. T., Horowitz-Scherer, R. A., Adkins, N., Woodcock, C. L., Wade, P. A., and Hansen, J. C. (2003). Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.* 278, 32181–32188. doi: 10.1074/jbc.M305308200
- Ghosh, R. P., Nikitina, T., Horowitz-Scherer, R. A., Gierasch, L. M., Uversky, V. N., Hite, K., et al. (2010). Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry* 49, 4395–4410. doi: 10.1021/bi9019753
- Gnanapragasam, M. N., Scarsdale, J. N., Amaya, M. L., Webb, H. D., Desai, M. A., Walavalkar, N. M., et al. (2011). p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7487–7492. doi: 10.1073/pnas.1015341108
- Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., et al. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398. doi: 10.1126/science.1120976
- Gonzales, M. L., Adams, S., Dunaway, K. W., and LaSalle, J. M. (2012). Phosphorylation of distinct sites in MeCP2 modifies cofactor associations and the dynamics of transcriptional regulation. *Mol. Cell. Biol.* 32, 2894–2903. doi: 10.1128/MCB.06728-11
- Grohmann, M., Spada, F., Schermelleh, L., Alenina, N., Bader, M., Cardoso, M. C., et al. (2005). Restricted mobility of Dnmt1 in preimplantation embryos: implications for epigenetic reprogramming. *BMC Dev. Biol.* 5:18. doi: 10.1186/1471-213X-5-18
- Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610. doi: 10.1038/nature10443
- Gunther, K., Rust, M., Leers, J., Boettger, T., Scharfe, M., Jarek, M., et al. (2013). Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res.* 41, 3010–3021. doi: 10.1093/nar/gkt035
- Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., et al. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* 17, 215–222. doi: 10.1038/nn.3607
- Guo, J. U., Su, Y., Zhong, C., Ming, G. L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423–434. doi: 10.1016/j.cell.2011.03.022
- Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315, 1143–1147. doi: 10.1126/science.1138389
- Guy, J., Hendrich, B., Holmes, M., Martin, J. E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* 27, 322–326. doi: 10.1038/85899
- Hackett, J. A., Sengupta, R., Zyllicz, J. J., Murakami, K., Lee, C., Down, T. A., et al. (2013). Germ-line DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339, 448–452. doi: 10.1126/science.1229277
- Haffner, M. C., Chaux, A., Meeker, A. K., Esopi, D. M., Gerber, J., Pellakuru, L. G., et al. (2011). Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* 2, 627–637. doi: 10.18632/oncotarget.316
- Hagberg, B., Aicardi, J., Dias, K., and Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: rett's syndrome: report of 35 cases. *Ann. Neurol.* 14, 471–479. doi: 10.1002/ana.410140412
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., et al. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* 117, 15–23. doi: 10.1016/S0925-4773(02)00181-8

- Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P., and Prunell, A. (1996). Linker histone-dependent DNA structure in linear mononucleosomes. *J. Mol. Biol.* 257, 30–42. doi: 10.1006/jmbi.1996.0144
- Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M., et al. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14412–14417. doi: 10.1073/pnas.96.25.14412
- Hashimoto, H., Horton, J. R., Zhang, X., Bostick, M., Jacobsen, S. E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 455, 826–829. doi: 10.1038/nature07280
- Hashimoto, H., Horton, J. R., Zhang, X., and Cheng, X. (2009). UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics* 4, 8–14. doi: 10.4161/epi.4.1.7370
- Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., et al. (2012a). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 40, 4841–4849. doi: 10.1093/nar/gks155
- Hashimoto, H., Pais, J. E., Dai, N., Corrêa, I. R., Zhang, X., Zheng, Y., et al. (2015). Structure of *Naegleria* Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. *Nucleic Acids Res.* 43, 10713–10721. doi: 10.1093/nar/gkv870
- Hashimoto, H., Pais, J. E., Zhang, X., Saleh, L., Fu, Z. Q., Dai, N., et al. (2014). Structure of a *Naegleria* Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* 506, 391–395. doi: 10.1038/nature12905
- Hashimoto, H., Zhang, X., and Cheng, X. (2012b). Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* 40, 8276–8284. doi: 10.1093/nar/gks268
- He, X., Duan, C., Chen, J., Ou-Yang, X., Zhang, Z., Li, C., et al. (2009). Let-7a elevates p21(WAF1) levels by targeting of NIRF and suppresses the growth of A549 lung cancer cells. *FEBS Lett.* 583, 3501–3507. doi: 10.1016/j.febslet.2009.10.007
- He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307. doi: 10.1126/science.1210944
- Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* 18, 6538–6547. doi: 10.1128/MCB.18.11.6538
- Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* 15, 710–723. doi: 10.1101/gad.194101
- Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401, 301–304. doi: 10.1038/45843
- Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* 19, 269–277. doi: 10.1016/S0168-9525(03)00080-5
- Hopfner, R., Mousli, M., Jeltsch, J. M., Voulgaris, A., Lutz, Y., Marin, C., et al. (2000). ICBP90, a novel human CCAAT binding protein, involved in the regulation of topoisomerase II α expression. *Cancer Res.* 60, 121–128.
- Hotchkiss, D. (1948). THE QUANTITATIVE SEPARATION OF PURINES, and NUCLEOSIDES BY paper chromatography. *J. Biol. Chem.* 175, 315–332.
- Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., et al. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* 155, 1545–1555. doi: 10.1016/j.cell.2013.11.020
- Hu, L., Li, Z., Wang, P., Lin, Y., and Xu, Y. (2011). Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res.* 21, 1374–1378. doi: 10.1038/cr.2011.124
- Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., et al. (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature* 527, 118–122. doi: 10.1038/nature15713
- Hu, Y. G., Hirasawa, R., Hu, J. L., Hata, K., Li, C. L., Jin, Y., et al. (2008). Regulation of DNA methylation activity through Dnmt3L promoter methylation by Dnmt3 enzymes in embryonic development. *Hum. Mol. Genet.* 17, 2654–2664. doi: 10.1093/hmg/ddn165
- Huang, H., Jiang, X., Li, Z., Li, Y., Song, C.-X., He, C., et al. (2013). TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11994–11999. doi: 10.1073/pnas.1310656110
- Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., et al. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1361–1366. doi: 10.1073/pnas.1322921111
- Inano, K., Suetake, I., Ueda, T., Miyake, Y., Nakamura, M., Okada, M., et al. (2000). Maintenance-type DNA methyltransferase is highly expressed in post-mitotic neurons and localized in the cytoplasmic compartment. *J. Biochem.* 128, 315–321. doi: 10.1093/oxfordjournals.jbchem.a022755
- Iqbal, K., Jin, S. G., Pfeifer, G. P., and Szabo, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3642–3647. doi: 10.1073/pnas.1014033108
- Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., et al. (2014). TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* 19, 52–65. doi: 10.1111/gtc.12107
- Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. (2010). Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129–1133. doi: 10.1038/nature09303
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303. doi: 10.1126/science.1210597
- Jin, C., Lu, Y., Jelinek, J., Liang, S., Estecio, M. R. H., Barton, M. C., et al. (2014). TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.* 42, 6956–6971. doi: 10.1093/nar/gku372
- Jin, S. G., Wu, X., Li, A. X., and Pfeifer, G. P. (2011). Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res.* 39, 5015–5024. doi: 10.1093/nar/gkr120
- Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., et al. (2007). The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* 17, 379–384. doi: 10.1016/j.cub.2007.01.009
- Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., et al. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* 19, 187–191. doi: 10.1038/561
- Jorgensen, H. F., Ben-Porath, I., and Bird, A. P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol. Cell. Biol.* 24, 3387–3395. doi: 10.1128/MCB.24.8.3387-3395.2004
- Jung, B. P., Jugloff, D. G., Zhang, G., Logan, R., Brown, S., and Eubanks, J. H. (2003). The expression of methyl CpG binding factor MeCP2 correlates with cellular differentiation in the developing rat brain and in cultured cells. *J. Neurobiol.* 55, 86–96. doi: 10.1002/neu.10201
- Jung, S. Y., Li, Y., Wang, Y., Chen, Y., Zhao, Y., and Qin, J. (2008). Complications in the assignment of 14 and 28 Da mass shift detected by mass spectrometry as in vivo methylation from endogenous proteins. *Anal. Chem.* 80, 1721–1729. doi: 10.1021/ac7021025
- Kaji, K., Caballero, I. M., MacLeod, R., Nichols, J., Wilson, V. A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat. Cell Biol.* 8, 285–292. doi: 10.1038/ncb1372
- Kaludov, N. K., and Wolffe, A. P. (2000). MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. *Nucleic Acids Res.* 28, 1921–1928. doi: 10.1093/nar/28.9.1921
- Karagianni, P., Amazit, L., Qin, J., and Wong, J. (2008). ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol. Cell. Biol.* 28, 705–717. doi: 10.1128/MCB.01598-07
- Kemmerich, K., Dingler, F. A., Rada, C., and Neuberger, M. S. (2012). Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res.* 40, 6016–6025. doi: 10.1093/nar/gks259
- Kiefer, H., Chatail-Hermitte, F., Ravassard, P., Bayard, E., Brunet, L., and Mallet, J. (2005). ZENON, a novel POZ Kruppel-like DNA binding protein associated with differentiation and/or survival of late postmitotic neurons. *Mol. Cell. Biol.* 25, 1713–1729. doi: 10.1128/MCB.25.5.1713-1729.2005

- Kienhöfer, S., Musheev, M. U., Stapf, U., Helm, M., Schomacher, L., Niehrs, C., et al. (2015). GADD45a physically and functionally interacts with TET1. *Differentiation* 90, 59–68. doi: 10.1016/j.diff.2015.10.003
- Kim, J. K., Esteve, P. O., Jacobsen, S. E., and Pradhan, S. (2009). UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res.* 37, 493–505. doi: 10.1093/nar/gkn961
- Kim, S. W., Park, J. I., Spring, C. M., Sater, A. K., Ji, H., Otchere, A. A., et al. (2004). Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nat. Cell Biol.* 6, 1212–1220. doi: 10.1038/ncb1191
- Kimura, H., and Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.* 278, 4806–4812. doi: 10.1074/jbc.M209923200
- Klein, C. J., Bird, T., Ertekin-Taner, N., Lincoln, S., Hjorth, R., Wu, Y., et al. (2013). DNMT1 mutation hot spot causes varied phenotypes of HSN1 with dementia and hearing loss. *Neurology* 80, 824–828. doi: 10.1212/WNL.0b013e31828
- Klein, C. J., Botuyan, M. V., Wu, Y., Ward, C. J., Nicholson, G. A., Hammans, S., et al. (2011). Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* 43, 595–600. doi: 10.1038/ng.830
- Ko, M., An, J., Bandukwala, H. S., Chavez, L., Aijö, T., Pastor, W. A., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122–126. doi: 10.1038/nature12052
- Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839–843. doi: 10.1038/nature09586
- Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., et al. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. *J. Biol. Chem.* 276, 34115–34121. doi: 10.1074/jbc.M105747200
- Konstantin, N., Bultmann, S., Schwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., et al. (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* 25, 1649–1652. doi: 10.1038/leu.2011.134
- Kozioł, M. J., Bradshaw, C. R., Allen, G. E., Costa, A. S., Frezza, C., and Gurdon, J. B. (2016). Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* 23, 24–30. doi: 10.1038/nsmb.3145
- Kraus, T. F. J., Globisch, D., Wagner, M., Eigenbrod, S., Widmann, D., Münzel, M., et al. (2012). Low values of 5-hydroxymethylcytosine (5hmC), the “sixth base,” are associated with anaplasia in human brain tumors. *Int. J. Cancer* 131, 1577–1590. doi: 10.1002/ijc.27429
- Kriaucionis, S., and Bird, A. (2004). The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic Acids Res.* 32, 1818–1823. doi: 10.1093/nar/gkh349
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930. doi: 10.1126/science.1169786
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J., et al. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* 22, 1–10. doi: 10.1093/nar/22.1.1
- Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., et al. (2010). The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS ONE* 5:e11982. doi: 10.1371/journal.pone.0011982
- Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., et al. (2011). DNMT3A Mutations in Acute Myeloid Leukemia. *N. Engl. J. Med.* 363, 2424–2433. doi: 10.1056/NEJMoa1005143
- Langemeijer, S., Kuiper, R., Berends, M., Knops, R., Aslanyan, M., Massop, M., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. *Leuk. Res.* 33, 838–842. doi: 10.1016/S0145-2126(09)70132-8
- Le Guezennec, X., Vermeulen, M., Brinkman, A. B., Hoeijmakers, W. A., Cohen, A., Lasonder, E., et al. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell Biol.* 26, 843–851. doi: 10.1128/MCB.26.3.843-851.2006
- Lee, J. H., and Skalnik, D. G. (2002). CpG-binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax. Identification of nuclear matrix targeting signals. *J. Biol. Chem.* 277, 42259–42267. doi: 10.1074/jbc.M205054200
- Lee, J. H., and Skalnik, D. G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* 280, 41725–41731. doi: 10.1074/jbc.M508312200
- Leonhardt, H., Page, A. W., Weier, H. U., and Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71, 865–873. doi: 10.1016/0092-8674(92)90561-P
- Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., et al. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69, 905–914. doi: 10.1016/0092-8674(92)90610-O
- Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., et al. (2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 681–683. doi: 10.1038/nbt.2661
- Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926. doi: 10.1016/0092-8674(92)90611-F
- Li, H., and Chang, Q. (2014). Regulation and function of stimulus-induced phosphorylation of MeCP2. *Front. Biol. (Beijing)* 9:367–375. doi: 10.1007/s11515-014-1330-2
- Li, W., and Liu, M. (2011). Distribution of 5-hydroxymethylcytosine in different human tissues. *J. Nucleic Acids* 2011, 870726. doi: 10.4061/2011/870726
- Li, Z., Cai, X., Cai, C. L., Wang, J., Zhang, W., Petersen, B. E., et al. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* 118, 4509–4518. doi: 10.1182/blood-2010-12-325241
- Li, Z., Gu, T. P., Weber, A. R., Shen, J. Z., Li, B. Z., Xie, Z. G., et al. (2015). Gadd45a promotes DNA demethylation through TDG. *Nucleic Acids Res.* 43, 3986–3997. doi: 10.1093/nar/gkv283
- Lin, R. K., Hsu, H. S., Chang, J. W., Chen, C. Y., Chen, J. T., and Wang, Y. C. (2007). Alteration of DNA methyltransferases contributes to 5′CpG methylation and poor prognosis in lung cancer. *Lung Cancer* 55, 205–213. doi: 10.1016/j.lungcan.2006.10.022
- Liu, N., Wang, M., Deng, W., Schmidt, C. S., Qin, W., Leonhardt, H., et al. (2013). Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS ONE* 8:e62755. doi: 10.1371/journal.pone.0062755
- Liu, Y., Oakeley, E. J., Sun, L., and Jost, J. P. (1998). Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. *Nucleic Acids Res.* 26, 1038–1045. doi: 10.1093/nar/26.4.1038
- Loat, C. S., Curran, S., Lewis, C. M., Duvall, J., Geschwind, D., Bolton, P., et al. (2008). Methyl-CpG-binding protein 2 polymorphisms and vulnerability to autism. *Genes Brain Behav.* 7, 754–760. doi: 10.1111/j.1601-183X.2008.00414.x
- Lopes, E. C., Valls, E., Figueroa, M. E., Mazur, A., Meng, F. G., Chiosio, G., et al. (2008). Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. *Cancer Res.* 68, 7258–7263. doi: 10.1158/0008-5472.CAN-08-0344
- Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T., and Downing, J. R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 17, 637–641. doi: 10.1038/sj.leu.2402834
- Lu, J., Hu, L., Cheng, J., Dong, F., Wang, C., Yu, K., et al. (2016). A computational investigation on the substrate preference of TET2. *Phys. Chem. Chem. Phys.* 18, 4728–4738. doi: 10.1039/C5CP07266B
- Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., et al. (2002). Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* 298, 1747–1752. doi: 10.1126/science.1076469
- Luo, M., Ling, T., Xie, W., Sun, H., Zhou, Y., Zhu, Q., et al. (2013). NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells* 31, 1278–1286. doi: 10.1002/stem.1374
- Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538–540. doi: 10.1038/35046205
- Lyst, M. J., Ekiert, R., Ebert, D. H., Merusi, C., Nowak, J., Selfridge, J., et al. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* 16, 898–902. doi: 10.1038/nn.3434
- Lyst, M. J., Nan, X., and Stancheva, I. (2006). Regulation of MBD1-mediated transcriptional 1478 repression by SUMO and PIAS proteins. *EMBO J.* 25, 5317–5328. doi: 10.1038/sj.emboj.7601404

- Mahajan, M. C., Narlikar, G. J., Boyapaty, G., Kingston, R. E., and Weissman, S. M. (2005). Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the beta-globin locus control region. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15012–15017. doi: 10.1073/pnas.0507596102
- Maiti, A., and Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334–35338. doi: 10.1074/jbc.C111.284620
- Marhold, J., Kramer, K., Kremmer, E., and Lyko, F. (2004). The *Drosophila* MBD2/3 protein mediates interactions between the MI-2 chromatin complex and CpT/A-methylated DNA. *Development* 131, 6033–6039. doi: 10.1242/dev.01531
- Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., et al. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* 302, 890–893. doi: 10.1126/science.1090842
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* 403, 501–502. doi: 10.1038/35000654
- Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* 20, 5085–5092. doi: 10.1093/nar/20.19.5085
- Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., and Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58, 499–507. doi: 10.1016/0092-8674(89)90430-3
- Mellinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I. M., Klinkert, W. E., et al. (2009). Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep.* 10, 1259–1264. doi: 10.1038/embor.2009.201
- Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430. doi: 10.1016/j.cell.2012.11.022
- Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M., and Bestor, T. H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development* 125, 889–897.
- Michaels, M. L., Pham, L., Nghiem, Y., Cruz, C., and Miller, J. H. (1990). MutY, an adenine glycosylase active on G-A mispairs, has homology to endonuclease III. *Nucleic Acids Res.* 18, 3841–3845. doi: 10.1093/nar/18.13.3841
- Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., et al. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297, 403–405. doi: 10.1126/science.1073354
- Mnatzakanian, G. N., Lohi, H., Munteanu, I., Alfred, S. E., Yamada, T., MacLeod, P. J., et al. (2004). A previously unidentified MEC2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.* 36, 339–341. doi: 10.1038/ng1327
- Mori, T., Ikeda, D. D., Yamaguchi, Y., Unoki, M., and Project, N. (2012). NIRF/UHRF2 occupies a central position in the cell cycle network and allows coupling with the epigenetic landscape. *FEBS Lett.* 586, 1570–1583. doi: 10.1016/j.febslet.2012.04.038
- Mori, T., Li, Y., Hata, H., Ono, K., and Kochi, H. (2002). NIRF, a novel RING finger protein, is involved in cell-cycle regulation. *Biochem. Biophys. Res. Commun.* 296, 530–536. doi: 10.1016/S0006-291X(02)00890-2
- Muller, H. M., Fiegl, H., Goebel, G., Hubalek, M. M., Widschwendter, A., Muller-Holzner, E., et al. (2003). MeCP2 and MBD2 expression in human neoplastic and non-neoplastic breast tissue and its association with oestrogen receptor status. *Br. J. Cancer* 89, 1934–1939. doi: 10.1038/sj.bjc.6601392
- Muller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* 42, 8592–8604. doi: 10.1093/nar/gku552
- Muotri, A. R., Marchetto, M. C., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., et al. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446. doi: 10.1038/nature09544
- Nady, N., Lemak, A., Walker, J. R., Avvakumov, G. V., Kareta, M. S., Achour, M., et al. (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J. Biol. Chem.* 286, 24300–24311. doi: 10.1074/jbc.M111.234104
- Nagarajan, R. P., Hogart, A. R., Gwyne, Y., Martin, M. R., and LaSalle, J. M. (2006). Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MEC2 promoter methylation. *Epigenetics* 1, e1–e11. doi: 10.4161/epi.1.4.3514
- Nakagawa, T., Lv, L., Nakagawa, M., Yu, Y., Yu, C., D'Alessio, A. C., et al. (2015). CRL4VprBP E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases. *Mol. Cell* 57, 247–260. doi: 10.1016/j.molcel.2014.12.002
- Nakamura, T., Liu, Y. J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., et al. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 486, 415–419. doi: 10.1038/nature11093
- Nan, X., Campoy, F. J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 88, 471–481. doi: 10.1016/S0092-8674(00)81887-5
- Nan, X., Meehan, R. R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* 21, 4886–4892. doi: 10.1093/nar/21.21.4886
- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., et al. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386–389. doi: 10.1038/30764
- Nestor, C. E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., et al. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* 22, 467–477. doi: 10.1101/gr.126417.111
- Ng, H. H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. *Mol. Cell Biol.* 20, 1394–1406. doi: 10.1128/MCB.20.4.1394-1406.2000
- Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., et al. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* 23, 58–61. doi: 10.1038/12659
- Nikitina, T., Ghosh, R. P., Horowitz-Scherer, R. A., Hansen, J. C., Grigoryev, S. A., and Woodcock, C. L. (2007a). MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. *J. Biol. Chem.* 282, 28237–28245. doi: 10.1074/jbc.M704304200
- Nikitina, T., Shi, X., Ghosh, R. P., Horowitz-Scherer, R. A., Hansen, J. C., and Woodcock, C. L. (2007b). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Mol. Cell Biol.* 27, 864–877. doi: 10.1128/MCB.01593-06
- Ohki, I., Shimotake, N., Fujita, N., Nakao, M., and Shirakawa, M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J.* 18, 6653–6661. doi: 10.1093/emboj/18.23.6653
- Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257. doi: 10.1016/S0092-8674(00)81656-6
- Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Am. Inc.* 19, 219–220. doi: 10.1038/890
- Papaït, R., Pistore, C., Grazini, U., Babbio, F., Cogliati, S., Pecoraro, D., et al. (2008). The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin. *Mol. Biol. Cell* 19, 3554–3563. doi: 10.1091/mbc.E07-10-1059
- Papaït, R., Pistore, C., Negri, D., Pecoraro, D., Cantarini, L., and Bonapace, I. M. (2007). Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Mol. Biol. Cell* 18, 1098–1106. doi: 10.1091/mbc.E06-09-0874
- Park, J. I., Kim, S. W., Lyons, J. P., Ji, H., Nguyen, T. T., Cho, K., et al. (2005). Kaiso/p120-catenin and TCF/beta-catenin complexes coordinately regulate canonical Wnt gene targets. *Dev. Cell* 8, 843–854. doi: 10.1016/j.devcel.2005.04.010
- Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394–397. doi: 10.1038/nature10102
- Patra, S. K., Patra, A., Zhao, H., Carroll, P., and Dahiya, R. (2003). Methyl-CpG-DNA binding proteins in human prostate cancer: expression of CXXC sequence containing MBD1 and repression of MBD2 and MeCP2. *Biochem. Biophys. Res. Commun.* 302, 759–766. doi: 10.1016/S0006-291X(03)00253-5

- Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* 126, 781–790. doi: 10.1042/bj1260781
- Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Stoerker, J., Genuardi, M., et al. (2000). Biphasic kinetics of the human DNA repair protein MBD1 (MBD4), a mismatch-specific DNA N-glycosylase. *J. Biol. Chem.* 275, 32422–32429. doi: 10.1074/jbc.M004535200
- Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C. A., et al. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Edn.* 50, 7008–7012. doi: 10.1002/anie.201103899
- Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., et al. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* 10, 574–581. doi: 10.1038/nchembio.1532
- Pichler, G., Wolf, P., Schmidt, C. S., Meilinger, D., Schneider, K., Frauer, C., et al. (2011). Cooperative DNA and histone binding by Uhrf1 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* 112, 2585–2593. doi: 10.1002/jcb.23185
- Prokhorchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., et al. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* 15, 1613–1618. doi: 10.1101/gad.198501
- Prokhorchouk, A., Sansom, O., Selfridge, J., Caballero, I. M., Salozhin, S., Aithozhina, D., et al. (2006). Kaiso-deficient mice show resistance to intestinal cancer. *Mol. Cell. Biol.* 26, 199–208. doi: 10.1128/MCB.26.1.199-208.2006
- Qin, W., Leonhardt, H., and Pichler, G. (2011). Regulation of DNA methyltransferase 1 by interactions and modifications. *Nucleus* 2, 392–402. doi: 10.4161/nucl.2.5.17928
- Qiu, C., Sawada, K., Zhang, X., and Cheng, X. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat. Struct. Biol.* 9, 217–224. doi: 10.1038/nsb759
- Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502, 65–70. doi: 10.1038/nature12587
- Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., et al. (2011). PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol. Cell.* 43, 275–284. doi: 10.1016/j.molcel.2011.07.006
- Ramocki, M. B., Peters, S. U., Tavayev, Y. J., Zhang, F., Carvalho, C. M., Schaaf, C. P., et al. (2009). Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome. *Ann. Neurol.* 66, 771–782. doi: 10.1002/ana.21715
- Razin, A., and Cedar, H. (1977). Distribution of 5-methylcytosine in chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 74, 2725–2728. doi: 10.1073/pnas.74.7.2725
- Rett, A. (1966). [On an until now unknown disease of a congenital metabolic disorder]. *Krankenschwester* 19, 121–122.
- Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* 10, 583–594. doi: 10.1016/j.stem.2012.02.020
- Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., et al. (1999). The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nat. Genet.* 23, 266–268. doi: 10.1038/15443
- Roll, J. D., Rivenbark, A. G., Jones, W. D., and Coleman, W. B. (2008). DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines. *Mol. Cancer* 7, 15. doi: 10.1186/1476-4598-7-15
- Roloff, T. C., Ropers, H. H., and Nuber, U. A. (2003). Comparative study of methyl-CpG-binding domain proteins. *BMC Genomics* 4:1. doi: 10.1186/1471-2164-4-1
- Rothbart, S. B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A. I., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat. Struct. Mol. Biol.* 19, 1155–1160. doi: 10.1038/nsmb.2391
- Rottach, A., Frauer, C., Pichler, G., Bonapace, I. M., Spada, F., and Leonhardt, H. (2010). The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.* 38, 1796–1804. doi: 10.1093/nar/gkp1152
- Rountree, M. R., Bachman, K. E., and Baylin, S. B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.* 25, 269–277. doi: 10.1038/77023
- Rudenko, A., Dawlaty, M. M., Seo, J., Cheng, A. W., Meng, J., Le, T., et al. (2013). Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* 79, 1109–1122. doi: 10.1016/j.neuron.2013.08.003
- Saito, M., and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.* 277, 35434–35439. doi: 10.1074/jbc.M203455200
- Sarraf, S. A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell.* 15, 595–605. doi: 10.1016/j.molcel.2004.06.043
- Sasai, N., Matsuda, E., Sarashina, E., Ishida, Y., and Kawaichi, M. (2005). Identification of a novel BTB-zinc finger transcriptional repressor, CIBZ, that interacts with CtBP corepressor. *Genes Cells* 10, 871–885. doi: 10.1111/j.1365-2443.2005.00885.x
- Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., et al. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* 35, 4301–4312. doi: 10.1093/nar/gkm432
- Schlegel, J., Guneyesu, S., and Mennel, H. D. (2002). Expression of the genes of methyl-binding domain proteins in human gliomas. *Oncol. Rep.* 9, 393–395.
- Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., et al. (2013). Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res.* 41, 4860–4876. doi: 10.1093/nar/gkt191
- Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., et al. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* 35, 243–254. doi: 10.1016/S0896-6273(02)00768-7
- Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908–912. doi: 10.1038/nature06397
- Shibayama, A., Cook, E. H. Jr., Feng, J., Glanzmann, C., Yan, J., Craddock, N., et al. (2004). MECP2 structural and 3'-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 128B, 50–53. doi: 10.1002/ajmg.b.30016
- Shimbo, T., Du, Y., Grimm, S. A., Dhasarathy, A., Mav, D., Shah, R. R., et al. (2013). MBD3 localizes at promoters, gene bodies and enhancers of active genes. *PLoS Genet.* 9:e1004028. doi: 10.1371/journal.pgen.1004028
- Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., et al. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell.* 37, 457–468. doi: 10.1016/j.molcel.2010.01.030
- Song, J., Rechkoblit, O., Bestor, T. H., and Patel, D. J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331, 1036–1040. doi: 10.1126/science.1195380
- Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D. J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* 335, 709–712. doi: 10.1126/science.1214453
- Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159. doi: 10.1016/j.cell.2013.02.004
- Stancheva, I., Collins, A. L., Van den Veyver, I. B., Zoghbi, H., and Meehan, R. R. (2003). A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by notch in *Xenopus* embryos. *Mol. Cell.* 12, 425–435. doi: 10.1016/S1097-2765(03)00276-4
- Strichman-almashanu, L. Z., Lee, R. S., Onyango, P. O., Perlman, E., Flam, F., Frieman, M. B., et al. (2002). A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes a genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res* 12, 543–554. doi: 10.1101/gr.224102
- Su, L. K., Kinzler, K. W., Vogelstein, B., Preisinger, A. C., Moser, A. R., Luongo, C., et al. (1992). Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* 256, 668–670. doi: 10.1126/science.256.5060.1114-c
- Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and

- Dnmt3b through a direct interaction. *J. Biol. Chem.* 279, 27816–27823. doi: 10.1074/jbc.M400181200
- Suter, C. M., Martin, D. I., and Ward, R. L. (2004). Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. *Int. J. Colorectal. Dis.* 19, 95–101. doi: 10.1007/s00384-003-0539-3
- Szulwach, K. E., Li, X., Li, Y., Song, C. X., Wu, H., Dai, Q., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* 14, 1607–1616. doi: 10.1038/nn.2959
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi: 10.1126/science.1170116
- Tardy-Planechaud, S., Fujimoto, J., Lin, S. S., and Sowers, L. C. (1997). Solid phase synthesis and restriction endonuclease cleavage of oligodeoxynucleotides containing 5-(hydroxymethyl)-cytosine. *Nucleic Acids Res.* 25, 553–559. doi: 10.1093/nar/25.3.553
- Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., et al. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* 46, 385–388. doi: 10.1038/ng.2917
- Traynor, J., Agarwal, P., Lazzaroni, L., and Francke, U. (2002). Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. *BMC Med. Genet.* 3:12. doi: 10.1186/1471-2350-3-12
- Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., et al. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439, 811–816. doi: 10.1038/nature04433
- Tucker, K. L., Talbot, D., Lee, M. A., Leonhardt, H., and Jaenisch, R. (1996). Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12920–12925. doi: 10.1073/pnas.93.23.12920
- Tudor, M., Akbarian, S., Chen, R. Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15536–15541. doi: 10.1073/pnas.242566899
- Uchimura, Y., Ichimura, T., Uwada, J., Tachibana, T., Sugahara, S., Nakao, M., et al. (2006). Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. *J. Biol. Chem.* 281, 23180–23190. doi: 10.1074/jbc.M602280200
- Unoki, M., Nishidate, T., and Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* 23, 7601–7610. doi: 10.1038/sj.onc.1208053
- Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., et al. (2013). Tet proteins connect the O-Linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* 49, 645–656. doi: 10.1016/j.molcel.2012.12.019
- Vincent, J. J., Huang, Y., Chen, P.-Y., Feng, S., Calvopiña, J. H., Nee, K., et al. (2013). Stage-specific roles for Tet1 and Tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell* 12, 470–478. doi: 10.1016/j.stem.2013.01.016
- Wade, P. A., Geronne, A., Jones, P. L., Ballestar, E., Aubry, F., and Wolffe, A. P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat. Genet.* 23, 62–66. doi: 10.1038/12664
- Wakefield, R. I., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., et al. (1999). The solution structure of the domain from MeCP2 that binds to methylated DNA. *J. Mol. Biol.* 291, 1055–1065. doi: 10.1006/jmbi.1999.3023
- Walter, M. J., Ding, L., Shen, D., Shao, J., Grillot, M., McLellan, M., et al. (2011). Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia* 25, 1153–1158. doi: 10.1038/leu.2011.44
- Wang, C., Shen, J., Yang, Z., Chen, P., Zhao, B., Hu, W., et al. (2011). Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. *Cell Res.* 21, 1379–1382. doi: 10.1038/cr.2011.123
- Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech. Dev.* 118, 187–190. doi: 10.1016/S0925-4773(02)00242-3
- Watanabe, D., Suetake, I., Tajima, S., and Hanaoka, K. (2004). Expression of Dnmt3b in mouse hematopoietic progenitor cells and spermatogonia at specific stages. *Gene Expr. Patterns* 5, 43–49. doi: 10.1016/j.modgep.2004.06.008
- Weber, A., Marquardt, J., Elzi, D., Forster, N., Starke, S., Glaum, A., et al. (2008). Zbtb4 represses transcription of P21CIP1 and controls the cellular response to p53 activation. *EMBO J.* 27, 1563–1574. doi: 10.1038/emboj.2008.85
- Welling, M., Chen, H.-H., Muñoz, J., Musheev, M. U., Kester, L., Junker, J. P., et al. (2015). DAZL regulates Tet 1 translation in murine embryonic stem cells. *EMBO Rep.* 16, 791–802. doi: 10.15252/embr.201540538
- Whyte, W. A., Bilodeau, S., Orlando, D. A., Hoke, H. A., Frampton, G. M., Foster, C. T., et al. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225. doi: 10.1038/nature10805
- Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A., Rappasilber, J., et al. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348. doi: 10.1038/nature10066
- Winkelmann, J., Lin, L., Schormair, B., Kornum, B. R., Faraco, J., Plazzi, G., et al. (2012). Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum. Mol. Genet.* 21, 2205–2210. doi: 10.1093/hmg/dd5035
- Woo, H. R., Pontes, O., Pikaard, C. S., and Richards, E. J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev.* 21, 267–277. doi: 10.1101/gad.1512007
- Woodcock, C. L. (2006). Chromatin architecture. *Curr. Opin. Struct. Biol.* 16, 213–220. doi: 10.1016/j.sbi.2006.02.005
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., et al. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* 2:241. doi: 10.1038/ncomms1240
- Xie, Z. H., Huang, Y. N., Chen, Z. X., Riggs, A. D., Ding, J. P., Gowher, H., et al. (2006). Mutations in DNA methyltransferase DNMT3B in ICF syndrome affect its regulation by DNMT3L. *Hum. Mol. Genet.* 15, 1375–1385. doi: 10.1093/hmg/ddl059
- Xu, G. L., Bestor, T. H., Bourc'His, D., Hsieh, C. L., Tommerup, N., Bugge, M., et al. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402, 187–191. doi: 10.1038/46052
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* 19, 17–30. doi: 10.1016/j.ccr.2010.12.014
- Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* 42, 451–464. doi: 10.1016/j.molcel.2011.04.005
- Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for xenopus eye and neural development. *Cell* 151, 1200–1213. doi: 10.1016/j.cell.2012.11.014
- Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., et al. (2012). Tet1 controls meiosis by regulating meiotic gene expression. *Nature* 492, 443–447. doi: 10.1038/nature11709
- Yamaguchi, S., Shen, L., Liu, Y., Sandler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature* 504, 460–464. doi: 10.1038/nature12805
- Yasui, D. H., Peddada, S., Bieda, M. C., Vallero, R. O., Hogart, A., Nagarajan, R. P., et al. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19416–19421. doi: 10.1073/pnas.0707442104
- Yildirim, O., Li, R., Hung, J. H., Chen, P. B., Dong, X., Ee, L. S., et al. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510. doi: 10.1016/j.cell.2011.11.054
- Yoder, J. A., and Bestor, T. H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.* 7, 279–284. doi: 10.1093/hmg/7.2.279
- Yoder, J. A., Yen, R. W., Vertino, P. M., Bestor, T. H., and Baylin, S. B. (1996). New 5' regions of the murine and human genes for DNA (cytosine-5)-methyltransferase. *J. Biol. Chem.* 271, 31092–31097. doi: 10.1074/jbc.271.49.31092

- Yoon, H. G., Chan, D. W., Reynolds, A. B., Qin, J., and Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol. Cell* 12, 723–734. doi: 10.1016/j.molcel.2003.08.008
- Yu, C., Zhang, Y. L., Pan, W. W., Li, X. M., Wang, Z. W., Ge, Z. J., et al. (2013). CRL4 complex regulates mammalian oocyte survival and reprogramming by activation of TET proteins. *Science* 342, 1518–1521. doi: 10.1126/science.1244587
- Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y. G. (2010). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* 20, 1390–1393. doi: 10.1038/cr.2010.156
- Zhang, J., Gao, Q., Li, P., Liu, X., Jia, Y., Wu, W., et al. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res.* 21, 1723–1739. doi: 10.1038/cr.2011.176
- Zhang, P., Su, L., Wang, Z., Zhang, S., Guan, J., Chen, Y., et al. (2012). The involvement of 5-hydroxymethylcytosine in Active DNA demethylation in Mice. *Biol. Reprod.* 86, 104. doi: 10.1095/biolreprod.111.096073
- Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J., et al. (2014). Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT). *J. Biol. Chem.* 289, 5986–5996. doi: 10.1074/jbc.M113.524140
- Zhang, R. R., Cui, Q. Y., Murai, K., Lim, Y. C., Smith, Z. D., Jin, S., et al. (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* 13, 237–245. doi: 10.1016/j.stem.2013.05.006
- Zhang, Y., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* 13, 1924–1935.
- Zhao, X., Ueba, T., Christie, B. R., Barkho, B., McConnell, M. J., Nakashima, K., et al. (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6777–6782. doi: 10.1073/pnas.1131928100
- Zhou, Z., Qin, J., Tang, J., Li, B., Geng, Q., Jiang, W., et al. (2013). Down-regulation of MeCP2 in Hirschsprung's disease. *J. Pediatr. Surg.* 48, 2099–2105. doi: 10.1016/j.jpedsurg.2013.07.011
- Zocchi, L., and Sassone-Corsi, P. (2012). SIRT1-mediated deacetylation of MeCP2 contributes to BDNF expression. *Epigenetics* 7, 695–700. doi: 10.4161/epi.20733

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ludwig, Zhang and Cardoso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

3 Aims of this study

Aberrant DNA methylation has been implicated in various severe human diseases, such as cancer and neurological disorders. To understand how 5-methylcytosine is regulated, we analyzed the interplay of proteins reading, translating and erasing the repressive epigenetic mark.

In Chapter 1 (Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise) we analyzed:

- the impact of MBD proteins on the efficiency of Tet1 mediated 5mC oxidation
- the influence of the chronological DNA binding order of Tet and MBD proteins on the ability of MBDs to protect 5mC
- the effect of different MBD-DNA interaction modes on Tet1 activity and DNA binding ability
- the epigenetic composition of pericentric heterochromatin in a mouse model for Rett syndrome
- the correlation between pericentric 5hmC accumulation and major satellite expression

In Chapter 2 (L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins) we studied:

- the effect of Tet1 activity on promoter methylation levels, expression and retrotransposition of endogenous L1 elements
- the influence of Tet1 activity on the retrotransposition efficiency of an engineered L1 element
- the impact of MBD proteins on Tet1-mediated demethylation and reactivation of endogenous L1 elements

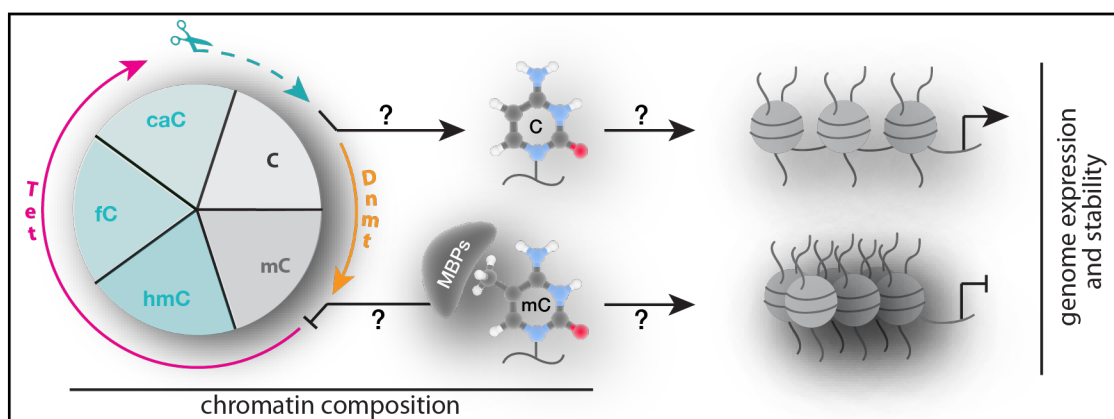


Figure 1: Scheme summarizing the main aims of Chapter 1 and Chapter 2

Similar to DNA, ribonucleic acids contain next to the four canonical bases a variety of post-synthetic nucleoside modifications that affect their structural and functional properties. Recent data indicate that Tet enzymes catalyze the formation of 5-hydroxymethylcytidine in RNA. To assess whether Tet proteins accumulate in the nucleolus to modify ribosomal RNA directly at the place of synthesis, the general principles of protein targeting to this subnuclear compartment need to be determined.

In Chapter 3 (Principles of protein targeting to the nucleolus) we investigated:

- the molecular requirements, including charge, lengths and composition of peptides and proteins that are necessary and sufficient to reach and accumulate inside of nucleoli
- the relative pH of the nucleolus compared to the surrounding nucleoplasm
- whether peptides and proteins that meet the sequence requirements for accumulation in nucleoli bind RNA *in vitro*

In Chapter 4 (Visualization of the Nucleolus in Living Cells with Cell-Penetrating Fluorescent Peptides) we aimed to

- develop a protocol for the selective labelling of nucleoli in living cells using fluorescently tagged cell-penetrating peptides

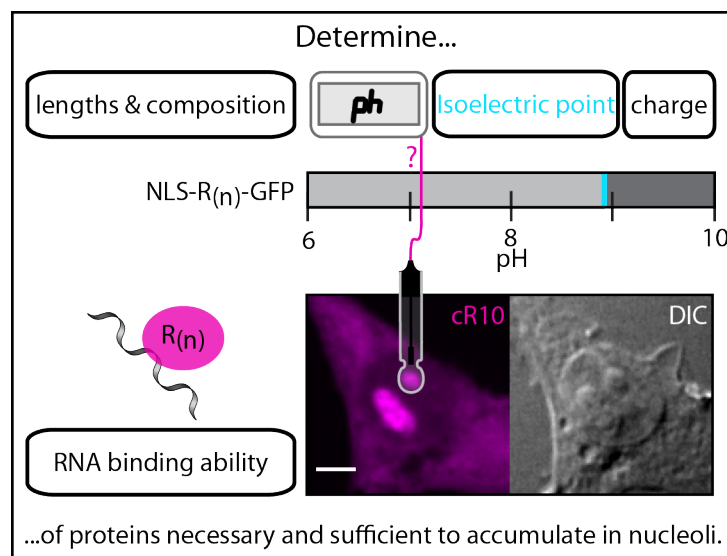


Figure 2: Scheme summarizing the main aims of Chapter 3 and Chapter 4

4 Results

4.1 Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise

A.K. Ludwig wrote the manuscript and prepared all of the figures (except for figures S4, S6, S10a (right) S10b, S10c and S12a).

A.K. Ludwig performed and analyzed the experiments of figures 1a, 3a, S10b, S10c and S11 together with P. Zhang.

A.K. Ludwig performed and analyzed the experiments of figures S7 and S10a (left) together with F. Hastert.

A.K. Ludwig performed and analyzed the experiments for figures 1b, 3b, 6e, S2, S16c.

A.K. Ludwig imaged and analyzed the data of figures 5, 6b, S14 and S16d

A.K. Ludwig analyzed the data of figures 6c, 6d, S16a and S16b.

Content of figures:

Figures 1a and 3a: Radioactive and immunological assay to determine 5hmC levels in genomic DNA of transiently transfected HEK cells. **Figures S10b and S10c:** Accumulation kinetics of Mbd2, Mecp2, MBD and IDTRD to pericentric heterochromatin in C2C12 mouse myoblasts. **Figure S11:** Analysis of Tet1CD and MBD expression levels in HEK versus C2C12 cells by flow cytometry analysis. **Figure S7:** Immunological assay to determine 5hmC levels in genomic DNA of wild type and Mecp2 knockout MTF cells. **Figures S10a (left), 1b and 3b:** *In situ* staining and quantification of genomic 5hmC levels in transiently transfected C2C12 mouse myoblasts. **Figure 6e:** RT-qPCR analysis of major satellite RNA transcript levels in transiently transfected C2C12 mouse myoblasts and brain of wild type and Mecp2 *y*^{-/-} mice, respectively. **Figure S2:** Western blot analysis of Tet1CD and MBD protein levels of transiently transfected, FACS sorted HEK cells. **Figure S16c:** Immunostaining of transiently transfected C2C12 mouse myoblasts for Mecp2 and 5hmC, respectively. **Figure 5:** Immunostaining and quantification of Mecp2, 5mC and 5hmC levels in NeuN positive cells of wild type and Mecp2 knockout mouse pontes, respectively. **Figure 6b:** Line intensity profiles of DNA and 5hmC distribution through pericentric heterochromatin in neurons of wild type and Mecp2 *y*^{-/-} mouse pontes, respectively. **Figure S14:** Imaging and analysis of coronal cross sections of mouse brain immunostained for Tet1, NeuN and Mecp2. **Figure S16d:** Accumulation analysis of 5hmC at chromocenters in neurons of wild type and Mecp2 *y*^{-/-} mouse pontes, respectively. **Figures 6c, 6d, S16a and S16b:** Detection and quantification of major satellite transcripts by Fluorescent In Situ Hybridization (FISH).

Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise

Anne K. Ludwig^{1,†}, Peng Zhang^{1,†}, Florian D. Hastert¹, Stephanie Meyer¹, Cathia Rausch¹, Henry D. Herce¹, Udo Müller², Anne Lehmkuhl¹, Ines Hellmann³, Carina Trummer², Christian Storm⁴, Heinrich Leonhardt² and M. Cristina Cardoso^{1,*}

¹Cell Biology and Epigenetics, Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany, ²Human Biology and Biomedicine, Department of Biology II, LMU Munich, 82152 Martinsried, Germany, ³Anthropology and Human Genomics, Department Biology II, LMU Munich, 82152 Martinsried, Germany and ⁴Chemical Plant Ecology, Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany

Received June 16, 2016; Revised November 03, 2016; Editorial Decision November 16, 2016; Accepted November 20, 2016

ABSTRACT

Aberrant DNA methylation is a hallmark of various human disorders, indicating that the spatial and temporal regulation of methylation readers and modifiers is imperative for development and differentiation. In particular, the cross-regulation between 5-methylcytosine binders (MBD) and modifiers (Tet) has not been investigated. Here, we show that binding of Mecp2 and Mbd2 to DNA protects 5-methylcytosine from Tet1-mediated oxidation. The mechanism is not based on competition for 5-methylcytosine binding but on Mecp2 and Mbd2 directly restricting Tet1 access to DNA. We demonstrate that the efficiency of this process depends on the number of bound MBDs per DNA molecule. Accordingly, we find 5-hydroxymethylcytosine enriched at heterochromatin of Mecp2-deficient neurons of a mouse model for Rett syndrome and Tet1-induced repression of silenced major satellite repeats. These data unveil fundamental regulatory mechanisms of Tet enzymes and their potential pathophysiological role in Rett syndrome. Importantly, it suggests that Mecp2 and Mbd2 have an essential physiological role as guardians of the epigenome.

INTRODUCTION

Methylation of DNA is generally accepted to be decisively involved in regulating gene expression (1). In mammals, 5-methylcytosine (5mC) accounts for 1% of all DNA bases and is primarily found as symmetrical methylation of CpG dinucleotides (2). A minor proportion of 5mC is localized within so-called CpG islands at the 5' ends of many genes, including those, responsible for genomic imprinting and X-

inactivation (3). The vast majority of methylated cytosines, however, are found in repetitive, endoparasitic sequences (4), whose transcriptional activity must be repressed to prevent translocations, gene disruption and chromosomal instability (5,6). The methylome is read and translated by conserved families of proteins, such as the methyl-CpG binding domain proteins (7). All members (of which the five best studied ones are Mecp2, Mbd1, Mbd2, Mbd3 and Mbd4) share a common protein motif, the methyl-CpG-binding domain (MBD) (8), which enables all family members except for Mbd3 to selectively bind to single methylated CpG dinucleotides (9). Moreover, all MBD proteins with the exception of Mbd4 have been described to function in transcriptional repression in part by recruiting silencing complexes such as histone deacetylases (HDACs) (1,10).

Mecp2, the founding member of the MBD protein family, is highly expressed in brain and was shown to mediate silencing of neuronal genes by the recruitment of the Sin3a–HDAC chromatin remodeling complex via its transcriptional repression domain, abbreviated TRD (10,11). In addition, Mecp2 was described to link methylated DNA with the nuclear receptor corepressor (NCoR), as well as the silencing mediator of retinoic acid and thyroid receptor (SMRT) in a neuronal activity dependent manner (12,13). Unlike its name suggests, Mecp2 binds preferentially, but not exclusively to methylated DNA (9,14,15). In addition to its core methyl-CpG binding domain (MBD), Mecp2 contains various non-sequence specific interaction sites for double-stranded DNA, including the TRD domain 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) (14). Upon binding to DNA, the ID and TRD domains of Mecp2, which constitute a large proportion of the extensively disordered protein, acquire secondary structure and stabilize Mecp2-chromatin complexes. Accordingly, deletion of these DNA binding do-

*To whom correspondence should be addressed. Tel: +49 6151 16 21882; Fax: +49 6151 16 21880; Email: cardoso@bio.tu-darmstadt.de

†These authors contributed equally to this work as first author.

mains were shown to considerably increase the fraction of unbound Mecp2 molecules within the cell nucleus (14,16). Besides this, MBD-based binding affinity was described to highly depend on the density of methylated CpG sites (15) and, thus, might vary extensively among different cell types. In mouse cells, Mecp2 was described to highly accumulate at densely methylated pericentric heterochromatin (17). As a consequence of homo- and hetero-interactions with itself and Mbd2 (18), as well as its multivalent DNA and 5mC binding ability, Mecp2 induces large-scale chromatin reorganization (19) accompanied by dampening transcriptional noise of highly methylated repetitive elements (20).

More recently, three mammalian enzymes (TET1-3) named after the ten-eleven translocation (t(10;11)(q22;23)) identified in a few cases of acute myeloid and lymphocytic leukemia (21–23), were shown to catalyze the conversion of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) in an iterative, Fe(II)- and oxoglutarate dependent oxidation reaction (23–25). This may either result in the erasure of the repressing methylcytosine mark with the aid of deaminases and enzymes of the base excision repair system (26), or the stable genomic integration of the oxidized cytosine derivatives as additional epigenetic information (27). Consequently, TET proteins have been proposed to play a key role in the long sought mechanism of active DNA demethylation (23), as well as in diversifying the epigenetic landscape, whose composition is dynamically regulated during development and in disease (27).

DNA hypo- as well as hypermethylation as a consequence of miss- or nonfunctioning 5mC writers, readers and modifiers, have been implicated in many malignancies including neurological and autoimmune disorders and cancer (28). Mutations in the X-linked *MECP2* gene cause Rett syndrome (29,30), a debilitating neurological disease that, at a molecular level, is characterized by increased expression and retrotransposition of repetitive elements (20,31).

By dissecting the interplay of 5mC readers and modifiers, we test the hypothesis of whether the anomalous transcriptional response observed in Rett patients is due to unconfined access of TET proteins to their substrate 5mC. In accordance with this, our data unveil a molecular mechanism by which Mecp2 and Mbd2 protect 5mC from Tet1 mediated oxidation *in vivo* and *in vitro* and provide definite indications of aberrant Tet activity in a mouse model for Rett syndrome, which lacks the aforementioned MBD-based defense system.

MATERIALS AND METHODS

Plasmids

Mammalian expression constructs coding for GFP-tagged mouse Mbd2-, mouse Mbd3- and rat Mecp2 full length proteins, rat Mecp2 deletion mutants (Mecp2G.9: aa 163–310 and Mecp2Y.5: aa 77–162), as well as for human Mecp2 deletion mutant R111G were previously described (8,18,19,32,33).

For construction of the mCherry-tagged catalytic active (Tet1CD: aa 1365–2057) and inactive (Tet1CDmut: aa 1365–2057, H1652Y, D1654A) domain of mouse Tet1,

Np95 was replaced from the mammalian expression vector pCAG-mCherry-Np95-IB (34) by Tet1CD (27) and Tet1CDmut (35), respectively using AsiSI and NotI sites.

For the expression of GFP, the commercial vector pEGFP-N1 (Clontech; Mountain View, CA, USA) was used.

Insect expression constructs coding for GFP-tagged mouse Mbd2- and rat Mecp2 full length proteins, as well as for rat Mecp2 deletion mutants Mecp2G.9 (aa 163–310) and Mecp2Y.5 (aa 77–162) were previously described (18,32,36).

For construction of the His-tagged catalytic domain of mouse Tet1, an N-terminal Histag, enterokinase- and AsiSI cutting site were amplified (fwd primer: 5'-gcc cga att cat gag cca tc-3', rev primer: 5'-ccc ggc ggc cgc tta-3') from an oligo (gcc cga att cat gag cca tca tca tca tca tca tga cga cga cga caa gag cga tcg cat gtc aac cag gag gga agc tta agc ggc cgc egg g) and inserted into the commercial transfer vector pFBDM of the MultiBac Expression System (37) using EcoRI and NotI sites. The catalytic active domain of mouse Tet1 (aa 1365–2057) was then cut from the mammalian expression vector (27) and inserted into the modified pFBDM transfer vector using AsiSI and NotI sites.

For the generation of the GFP- and mCherry-tagged catalytic active domain of mouse Tet1, GFP- and cherry-tagged Tet1CD, were cut from the mammalian expression vector pCAG-GFP-Tet1CD (27) and pCAG-cherry-Tet1CD (described above), respectively and inserted into the modified pFBDM transfer vector (described above) using BamHI and NotI sites.

Cell culture and transfection

C2C12 mouse myoblasts (38) were cultured using standard conditions described previously (39).

C2C12 cells were grown to 70% confluence on 16 mm glass cover slips in 6-well plates and transfected 3 hours post seeding using poly-ethylenimine (PEI, 1 mg/ml in ddH₂O, pH 10; Sigma-Aldrich, St. Louis, MO, USA) (40).

HEK 293-EBNA cells (Invitrogen; Paisley PA4 9RF, UK) were grown as previously described (41).

For the isolation of genomic DNA, HEK cells were seeded in 6-well plates at a target density of 5×10^5 cells per well and transfected 24 hours post seeding using poly-ethylenimine (PEI, 1mg/mL in ddH₂O, pH 7; Sigma-Aldrich, St. Louis, MO, USA).

For the production of GFP-tagged Tet1CD (27) and Mecp2 (19) proteins, HEK cells were grown to 70% confluence and transiently transfected using poly-ethylenimine (PEI, 1mg/ml in ddH₂O, pH 7; Sigma-Aldrich, St. Louis, MO, USA).

Sf9 insect cells (Invitrogen, Paisley PA4 9RF, UK) used for protein production were cultivated and transfected as previously described (18).

V6.5 wild type and triple Tet-knockout mouse embryonic stem cells (42) were maintained under serum-free and feeder-free conditions on Geltrex-coated flasks in N2B27 medium (50% neurobasal (Life Technologies, Carlsbad, California, USA) and 50% DMEM/F12 medium (Life Technologies, Carlsbad, California, USA) containing 2 mM L-glutamine (Life Technologies, Carlsbad, CA, USA), 0.1 mM β -mercaptoethanol, N2 supplement (Life Technolo-

gies, Carlsbad, CA, USA), B27 serum-free supplement (Life Technologies, Carlsbad, CA, USA), 100 U/ml Penicillin-Streptomycin, 1000 U/ml LIF and 2i (1 μ M PD032591 and 3 μ M CHIR99021 (Axon Medchem, Groningen, Netherlands)).

Mouse tail fibroblast (MTF) *lox/y* and MTF *-/y* (*Mecp2* knockout) (43) cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum. These cells are also heterozygous for *Mbd2*.

Primary neurons were isolated from brain of adult C57BL/6 mice. Whole brains were removed from mice under sterile conditions, cut into small pieces, put into 10 ml HBSS (Hank's Balanced Salt Solution) and washed by centrifugation (200 \times g, 1 min). After centrifugation, HBSS was discarded and the brain pellet resuspended in 5 ml 0.25% trypsin solution supplemented with 150 units DNaseI. After incubation for 20 min at 37°C in a waterbath, cells were centrifuged (200 \times g, 1 min) again. Trypsin was discarded and the pellet resuspended in 5 ml FBS for 2 min at 37°C. After centrifugation (200 \times g, 1 min), FBS was removed and the pellet resuspended in 5 ml Neurobasal medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with B-27 (Thermo Fisher Scientific, Waltham, MA, USA). Next, the brain suspension was triturated by passing it through a 10 ml serological plastic pipette for 15 times and subsequently through a flamed-tip glass Pasteur pipette for 15 times. Following centrifugation (400 \times g, 1 min), the pellet was resuspended in 10 ml fresh neurobasal medium. Big parts were allowed to settle for around 30 s and the supernatant was distributed in dishes, set up with laminin (Sigma, St. Louis, MO, USA) coated glass coverslips. Cells were cultured for 10 days by replacing 50% of the culture medium every two to three days before they were used for immunostaining.

Mice

Dissected brains of male *Mecp2* knockout mice (43) (~ postnatal day 40) (WL.B6.129P2(C)-*Mecp2*tml.1Bird>/J; Charles River Laboratories International, Inc., Wilmington, MA 01887, USA) used for immunostainings were kindly provided by AM Bischoff, lab of D. Richter (Universitätsmedizin Göttingen, Germany).

Dissected brains of male *Mecp2* knockout mice (43) (~ postnatal day 40) (WL.B6.129P2(C)-*Mecp2*tml.1Bird>/J; Charles River Laboratories International, Inc., Wilmington, MA 01887, USA) used for RNA isolation were kindly provided by the laboratory of Adrian Bird.

Brains of wild type mice (~ postnatal day 40) (C57BL/6N; Charles River Laboratories International, Inc., Wilmington, MA 01887, USA) were used as control.

Protein preparation

GFP-tagged Tet1CD proteins (used in Supplementary Figure S9) were prepared from whole cell lysates of HEK cells 36 h post-transfection using 50 mM NaH₂PO₄; pH 7.5, 150 mM NaCl, 10 mM imidazole, 0.5% Tween-20, 0.5 mM EDTA, 2 mM MgCl₂, 0.5 mM CaCl₂, 1 mM PMSF, 1 μ g/ μ l DNaseI and 1 \times mammalian protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA). Following centrifugation, supernatant was added to pre-equilibrated (50

mM NaH₂PO₄; pH 7.5, 150 mM NaCl, 10 mM imidazole, 0.5 mM EDTA and 0.05% Tween-20) Ni-NTA beads that were coupled to the GFP-binding protein (GFP-Trap, ChromoTek, Planegg-Martinsried, Germany) (44,45) and incubated for 2 h at 4°C on a rotary shaker. To remove unbound proteins, beads were centrifuged and washed with wash buffer (50 mM NaH₂PO₄; pH 7.5, 300 mM NaCl, 10 mM imidazole and 0.1% Tween-20). Elution was performed using wash buffer containing 250 mM imidazole. Elution buffer was exchanged to 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT and 100 ng/ μ l BSA using PD-10 desalting columns (GE Healthcare, Freiburg, Germany).

GFP-tagged *Mecp2* R111G proteins were prepared from whole cell lysates of HEK cells 24 h post-transfection using re-suspension buffer (32) containing 1 M NaCl and protease inhibitors in following concentrations: AEBSF 1 mM (AppliChem, Darmstadt, Germany), E64 10 μ M (AppliChem, Darmstadt, Germany), Pepstatin A 1 μ M (Sigma-Aldrich, St. Louis, MO, USA) and Aprotinin 2 ng/ml (Sigma-Aldrich, St. Louis, MO, USA). Cells were disrupted by syringe treatment (3 \times 20 gauge, 3 \times 21 gauge) followed by incubation on ice for 10 min.

Proteins were eluted by the addition of 4 M MgCl₂, pH 4.4 and subsequent incubation on ice for 10 min. Elution buffer was exchanged to 1 \times PBS using Amicon Ultra centrifugal filter units (Sigma-Aldrich, St. Louis, MO, USA).

All of the other GFP-, YFP- and mCherry-tagged proteins were purified from Sf9 insect cells as previously described (18,32) with following exceptions: The re-suspension buffer (32) was supplied with protease inhibitors in concentrations as described above. For the purification of Tet proteins, the sodium chloride (NaCl) concentration of the re-suspension buffer was decreased to 0.5 M. Cells were disrupted by syringe treatment (3 \times 20 gauge, 3 \times 21 gauge) followed by incubation on ice for 10 min. Proteins were eluted by the addition of 4 M MgCl₂, pH 4.4 and subsequent incubation on ice for 10 minutes. Elution buffer was exchanged to 1 \times PBS using Amicon Ultra centrifugal filter units (Sigma-Aldrich, St. Louis, MO, USA).

His-tagged proteins were purified from Sf9 insect cells using TALON ion metal affinity chromatography (Clontech Laboratories, Inc., CA, USA) according to the manufacturer's instructions with following changes. The re-suspension buffer contained 50 mM NaH₂PO₄, 300 mM NaCl and 10 mM imidazole, pH 8.0 and was supplied with protease inhibitors as described above. The elution buffer contained 50 mM NaH₂PO₄, 300 mM NaCl and 150 mM imidazole, pH 8.0. Elution buffer was exchanged to 1 \times PBS using Amicon Ultra centrifugal filter units (Sigma-Aldrich, St. Louis, MO, USA).

Western blot analysis

HEK cells were lysed in re-suspension buffer (32) containing 1 M NaCl for 20 min on ice and whole protein lysates were blotted as described before (46) on a nitrocellulose membrane (GE Healthcare, München, Germany). Visualization of the immunoreactive bands was achieved by ECL plus Western Blot Detection reagent (GE Healthcare, München, Germany). The following antibodies were used:

4 Nucleic Acids Research, 2016

monoclonal mouse anti GFP (Roche, Mannheim, Germany), polyclonal rabbit anti PCNA (Santa Cruz Biotechnology, Heidelberg, Germany), monoclonal rat anti RFP (47), Alexa488 conjugated goat anti mouse IgG (The Jackson Laboratory, Bar Harbor, USA), cy5 conjugated donkey anti rabbit IgG (The Jackson Laboratory, Bar Harbor, USA) and TexasRed conjugated donkey anti rat IgG (The Jackson Laboratory, Bar Harbor, USA).

Genomic DNA preparation

For the preparation of genomic DNA (gDNA), sorted HEK-EBNA, as well as MTF *lox/y* and *-/y* cells were pelleted (10 min, 2000 rpm, 4°C) and incubated overnight at 50°C in TNES buffer (10 mM Tris; pH 7.5, 400 mM NaCl, 10 mM EDTA, 0.6% SDS) supplemented with 1 mg/ml Proteinase K (Carl Roth, Karlsruhe, Germany) (48). RNA was removed by the addition of 0.6 mg/ml RNase A (Qiagen, Hilden, Germany) for 30 min at 37°C. gDNA was extracted by the addition of 6 M NaCl at a final concentration of 1.25 M and vigorous shaking (48). After centrifugation (15 min, 13200 rpm, RT), gDNA was precipitated from the supernatant by the addition of 100% ice cold ethanol followed by incubation at -20°C for 1 h and subsequent centrifugation (10 min, 13 200 rpm, 4°C). After a washing step in 70% ethanol, gDNA was air dried and solved in ddH₂O. Isolated gDNA from HEK cells was fragmented (<2000 bp) by sonication using the Biorupter TM UCD-200 (Diagenode, Seraing Ougrée, Belgium). The concentration of gDNA was measured on a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland).

RNA preparation

For sorted mouse C2C12 myoblasts, total RNA was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instruction. To remove traces of genomic DNA, RNA was treated with RNase-free recombinant DNaseI (Macherey Nagel, Dueren, Germany) for 90 min at 37°C and further purified with the Qiagen RNeasy Mini Kit. To assess the concentration and purity of RNA, the ratio of absorbance at 260 and 280 nm was measured on a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland).

For *Mecp2 y/-* and wild type mouse brain, total RNA was isolated and treated with RNase-free DNaseI as previously described (49).

Flow cytometry

C2C12 mouse myoblasts transiently expressing mCherry-tagged Tet1CD/Tet1CDmut and GFP-tagged *Mecp2* proteins, as well as HEK-EBNA cells transiently co-expressing high protein levels of mCherry-Tet1CD and GFP-tagged MBD proteins (*Mecp2*, *Mbd2*, *IDTRD* and *MBD*, respectively) were respectively trypsinized, re-suspended in PBS and separated from untransfected cells by fluorescent-activated cell sorting (FACS) on a S3e Cell Sorter (Bio-Rad Laboratories, Hercules, CA, USA) equipped with 488 and 561 nm excitation lasers and 525 ± 30 and 586 ± 25 nm emission filters, respectively. Sorted populations were either

processed for RNA- (C2C12 cells) or gDNA preparation (HEK-EBNA cells), respectively.

Real-time quantitative reverse transcription-polymerase chain reaction of major satellite repeats

For C2C12 mouse myoblasts, 20–200 ng of total RNA were used for cDNA synthesis using 200 units M-MuLV reverse transcriptase (NEB, Frankfurt, Germany), 0.01 OD units random primer of the Prime-It II Random Primer Labeling Kit (Stratagene, La Jolla, CA, USA), 0.5 mM dNTPs (Carl Roth, Karlsruhe, Germany) and 40 units recombinant ribonuclease inhibitor RNaseOUT (Invitrogen, Paisley PA4 9RF, UK) in a total reaction volume of 20 µl. Cycles were set to 10 min at 25°C, 60 min at 42°C and 20 min at 65°C.

For *Mecp2 y/-* and wt mouse brain, cDNA was synthesized as described previously (49) and kindly provided by Congdi Song.

Equal amounts of cDNA (0.5 ng) were used for real-time PCR with Platinum SYBR Green qPCR SuperMix-UDG w/ROX (Invitrogen, Paisley PA4 9RF, UK) on a StepOne-Plus Real-Time PCR System (Applied Biosystems, Darmstadt, Germany) according to the manufacturer's instruction. UDG was inactivated for 2 min at 50°C and cDNA was denatured for 10 min at 95°C. Cycle parameters were set to 40 cycles of 15 s at 95°C and 45 s at 60°C. Specificity of amplification products was confirmed by melting curve analysis.

Gene expression level were normalized to *Gapdh* and calculated using the comparative CT method ($\Delta\Delta CT$ method).

Primers for quantitative real-time PCR contained the following sequences: *Gapdh* forward: 5'-CCA TAC ATA CAG GTT TCT CCA G-3', *Gapdh* reverse: 5'-CTG GAA AGC TGT GGC GTG ATG G-3', *MajSat* forward (20): 5'-GGC GAG AAA ACT GAA AAT CAC G-3', *MajSat* reverse (20): 5'-AGG TCC TTC AGT GTG CAT TTC-3'.

Radioactive beta-glucosyltransferase (BGT) assay

The radioactive BGT assay was performed as described previously with following exceptions (50):

Reference DNA fragments (375 bp) containing 100% hmC (except primer sites) were prepared by PCR, using a 5-hydroxymethylcytosine dNTP Mix (Zymo Research, Freiburg, Germany), and Taq DNA polymerase (Cardoso Lab, Darmstadt, Germany). As template, gDNA isolated from HEK-EBNA cells was used. Primers for PCR contained the following sequences: 5'-ATC CCA CAC CTG GCT CAG AGG G-3' and 5'-GTC AGG GGT CAG GGA CCC ACT TGA GGA-3'. Cycles were set to: 94°C for 2 min, 40× (94°C for 15 s, 62°C for 30 s, 72°C for 40 s), 72°C for 10 min.

PCR products were purified by gel electrophoresis followed by silica column purification using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

Reactions contained 50 mM potassium acetate, 20 mM Tris acetate (pH 7.9), 10 mM magnesium acetate, 1 mM DTT, 2.8 µM 'cold' UDP-glucose (Sigma Aldrich, St. Louis, MO, USA), 86 nM UDP-[3H]glucose (glucose-6-3H; 60 Ci/mmol; Hartmann Analytic GmbH), 1 µg DNA substrate and 75 nM recombinant β-glucosyltransferase in a

total volume of 50 μ l. Reactions were incubated for 1 h at 37°C and terminated by heating at 65°C for 10 min. DNA was purified from the reaction mixture using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Remaining radioactivity was measured using a Liquid Scintillation Analyzer Tri-Carb 2800TR (PerkinElmer, Waltham, Massachusetts, USA) with quench indicating parameter set on tSIE/AEC (transformed spectral index of the external standard/automatic efficiency control) in 2 ml of Rotiszint Eco Plus scintillation liquid (Carl Roth, Karlsruhe, Germany) in Snaptwist vials (Zinsser Analytic, Frankfurt, Germany). Samples were measured for 1 min or until the 2 σ value reached 2%.

***In vitro* oxidation and protection assay**

Reference DNA fragments (375 bp) containing 100% 5mC (except primer sites) were prepared by PCR, using 5-methyl-dCTP (NEB, Frankfurt, Germany). Genomic DNA isolated from HEK-EBNA cells was used as template with primers: 5'-ATC CCA CAC CTG GCT CAG AGG G-3' and 5'-GTC AGG GGT CAG GGA CCC ACT TGA GGA-3', Q5[®] High-Fidelity DNA Polymerase (NEB, Frankfurt, Germany) and the following cycling profile: 98°C for 2 min, 40 \times (98°C for 15 s, 62°C for 30 s, 72°C for 60 s), 72°C for 2 min. PCR products were purified by gel electrophoresis followed by silica column purification using the QIAquick PCR purification kit (Qiagen, Hilden, Germany). For *in vitro* oxidation and protection assays, DNA fragments were incubated with MBD- and Tet1 proteins at 37°C in Tet oxidation buffer (10 μ M Fe(NH₄)₂(SO₄)₂·6H₂O, 100 mM NaCl, 50 mM HEPES (pH 8), 1.2 mM adenosine triphosphate (ATP), 2.5 mM dithiothreitol (DTT), 1 mM α -ketoglutarate (aKG) and 2 mM L-ascorbic acid). Following 120 min of Tet1 incubation, the reaction was stopped by the addition of 20 μ g of proteinase K at 50°C for 2 h.

Slot blotting

gDNA samples and *in vitro* oxidation products were respectively denatured at 99°C for 10 min and placed quickly on ice for 5 min. Denatured gDNA was mixed with ice-cold 20 \times saline-sodium citrate (SSC) buffer at a final concentration of 4.8 \times SSC and blotted on a nitrocellulose membrane (Bio-Rad Laboratories, Hercules, CA, USA), which was pre-equilibrated in 20 \times SSC. After air-drying, the membrane was blocked with 3% milk in PBST (PBS containing 0.1% Tween) for 30 min at room temperature (RT), followed by incubation with either mouse anti 5mC (1:1000, Eurogentec, Seraing, Belgium) or rabbit anti 5hmC (1:5000, Active Motif, La Hulpe, Belgium) antibodies for 1 h at RT. The membrane was washed 3 \times for 10 min with PBST, before it was incubated with horseradish peroxidase (HRP)-conjugated anti mouse IgG (1:5000, GE Healthcare, Freiburg, Germany) or anti rabbit IgG (1:5000, Sigma Aldrich, St. Louis, MO, USA) antibody for 1 h at RT. After three washing steps, remaining signals were detected using Amersham ECL detection reagent (GE Healthcare, Freiburg, Germany) and imaged on a Fuji LAS-1000 imager (FUJI Film, Minato, Tokyo, Japan).

Quantification of 5hmC using a methyl sensitive restriction assay

ATTO550-labeled 42 bp-long, double-stranded oligonucleotides (GGA TGA TGA CTC TTC TGG TCmC GGA TGG TAG TTA AGT GTT GAG) (Eurofins MWG Operon, Ebersberg, Germany) containing a central methylated CpG site were diluted in Tet reaction buffer (50 mM Tris-HCl; pH 7.5, 75 μ M Fe(II), 2 mM sodium ascorbate, 1 mM di-sodium-ketoglutarate) (24,35). Following incubation with purified GFP-Tet1CD and Mecp2-GFP, the reaction was heat-inactivated for 2 min at 95°C. Subsequently, oligonucleotides were digested using MspI at 37°C for 30 min. DNA was separated on a denaturing 17% polyacrylamide gel and imaged using the Typhoon TRIO Imager (GE Healthcare, Freiburg, Germany). Quantification was performed with ImageJ.

Competitive DNA binding assay

Gel mobility shift assays (EMSA) were performed as described previously (<http://www.nature.com/nmeth/journal/v2/n7/abs/nmeth0705-557.html>) with following modifications. GFP-tagged MBD and cherry-tagged Tet1CD proteins were incubated with ATTO647N labeled 42 bp-long, double-stranded oligonucleotides containing a single methylated CpG dinucleotide (5'-CTC AAC AAC TAA CTA CCA TmCGG ACC AGA AGC GTC ATC ATGG -3') in binding buffer composed of 20 mM HEPES pH 7.9, 1 mM EDTA, 3 mM MgCl₂, 2 mM DTT, 4% glycerol and 0.1% Triton X100 for 1.5 h at 37°C. Samples were separated on a non-denaturing 4.5% polyacrylamide gel (30%, 29:1 acrylamide:bisacrylamide), which was pre-run for 2 h at 4°C. Fluorescent signals were detected using a Storm 860 Molecular Imager (GMI, Ramsey, Minnesota, USA) and a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland), respectively.

Immunofluorescence staining of cells

Cells were fixed for 10 min in 4% formaldehyde and permeabilized for 20 min with 0.5% Triton X-100. For detection of genomic 5hmC, endogenous Tet1 proteins, as well as NeuN, cells were incubated following formaldehyde fixation with ice-cold methanol for 5 min. After RNaseA treatment (10 μ g/ml) for 30 min at 37°C, cells were washed and blocked for 30 min in 0.2% fish skin gelatin (Sigma Aldrich, St. Louis, MO, USA) at 37°C. Genomic 5hmC was detected using a rabbit anti-5hmC antibody (1:250; Active Motif, La Hulpe, Belgium) in conjunction with 25 U/ml DNaseI (Sigma Aldrich, St. Louis, MO, USA) for 70 min at 37°C. Endogenous Tet1 and NeuN proteins were detected using a rat anti Tet1 5D8 antibody (51) (1:4) and a mouse anti NeuN (1:50, Merck Millipore, Darmstadt, Germany) in conjunction with 25 U/ml DNaseI (Sigma Aldrich, St. Louis, MO, USA) for 70 min at 37°C. To stop DNaseI digestion, cells were washed with PBS containing 1 mM EDTA and 0.01% Tween. Following incubation with the secondary AMCA conjugated donkey anti rabbit IgG antibody (1:100; The Jackson Laboratory, Bar Harbor, USA), or the cy3 conjugated anti mouse IgG (1:500; The Jackson Laboratory, Bar Harbor, USA) and Alexa488 conjugated donkey

anti rat IgG antibody (1:500; The Jackson Laboratory, Bar Harbor, USA) for 45 min at RT, cells were mounted in Vectashield Medium (Vector Labs, Burlingame, CA, USA).

For immunofluorescence staining of Mecp2, fixed and permeabilized cells were blocked for 30 min in 0.2% fish skin gelatin (Sigma Aldrich, St. Louis, MO, USA). The primary rabbit anti Mecp2 antibody (32) (1:2) was applied for 1 h at RT. After three washing steps using PBST containing 0.01% Tween, the secondary donkey anti rabbit IgG cy3 (1:500, The Jackson Laboratory, Bar Harbor, USA) was applied for 45 min at RT. Following three washing steps in PBST containing 0.01% Tween, DNA was counterstained for 10 min with 1 µg/ml DAPI (Sigma, St. Louis, MO, USA), washed in PBS and mounted in Vectashield Medium (Vector Labs, Burlingame, CA, USA).

Immunofluorescence staining of tissues

Brains of Mecp2 wild type and knockout mice were fixed for 24 h in 10% neutral-buffered formalin (Sigma, St. Louis, MO, USA) at 4°C. Fixed tissues were dehydrated (30 min 70% ethanol (Sigma, St. Louis, MO, USA), 45 min 70% ethanol, 60 min 96% ethanol, 45 min 96% ethanol, 45 min 100% ethanol, 45 min 100% ethanol, 60 min xylol (Sigma, St. Louis, MO, USA), 30 min xylol), embedded in paraffin (60 min paraffin (Carl Roth, Karlsruhe, Germany), 45 min paraffin, 60 min paraffin) and sectioned at a thickness of 6 µm. Following dewaxing in xylol (3 × 5 min) and rehydration (5 min 100% ethanol, 5 min 96% ethanol, 5 min 90% ethanol, 5 min 80% ethanol, 5 min 70% ethanol, 5 min ddH₂O), brain sections were incubated for 30 min at 100°C/1 bar overpressure in 10 mM sodium citrate buffer, pH 6 (Carl Roth, Karlsruhe, Germany). Sections were encircled using a hydrophobic immuno-pen (Merck Millipore, Darmstadt, Germany) and blocked for 30 min in PBS containing 4% BSA. Primary antibodies (rabbit anti 5hmC (1:1000, Active Motif, La Hulpe, Belgium), mouse anti Mecp2 8D11 (32) (1:2), mouse anti 5mC (1:100, Eurogentec, Seraing, Belgium), mouse anti NeuN (1:100, Merck Millipore, Darmstadt, Germany) and rat anti Tet1 5D8 (51) (1:2), respectively) were applied overnight in PBS supplemented with 1% BSA at 4°C. After three washing steps using PBS containing 0.1% Tween, secondary antibodies (donkey anti rabbit IgG cy3 (1:500, The Jackson Laboratory, Bar Harbor, USA), donkey anti rat IgG cy3 (1:500, The Jackson Laboratory, Bar Harbor, USA) and donkey anti mouse IgG cy3 (1:500, The Jackson Laboratory, Bar Harbor, USA)) were respectively applied for 1 h at RT. Following three washing steps in PBS containing 0.1% Tween, DNA was counterstained for 10 min with 1 µg/ml DAPI (Sigma, St. Louis, MO, USA), washed in PBS and mounted in Vectashield Medium (Vector Labs, Burlingame, CA, USA).

Major satellite RNA FISH

For the detection of major satellite RNA transcripts, cDNA probes were amplified and labeled from genomic DNA of mouse myoblasts by PCR (major satellite fwd: 5' AAAATGAGAAACATCCACTTG 3', major satellite rev: 5' CCATGATTTTCAGTTTTCTT 3') and biotin dUTP.

Brain sections were prepared as described for immunofluorescence staining of tissues. Following rehydration in water, sections were hybridized for 1 h at RT and 12 h at 4°C. After three washing steps in water, slides were incubated for 1 h at RT with Alexa-488 conjugated streptavidin (1:500, Invitrogen, Paisley PA4 9RF, UK). To remove unbound Streptavidin, slides were washed in water before DNA was counterstained for 10 min with 1 µg/ml DAPI (Sigma, St. Louis, MO, USA). Brain sections were rinsed in PBS and mounted in Vectashield Medium (Vector Labs, Burlingame, CA, USA). All reagents used for RNA FISH were supplemented with 1x ProtectRNA RNase inhibitor (Sigma, St. Louis, MO, USA). As control, equivalent slides were treated in parallel with RNase A before signal detection.

Microscopy

Images of transiently transfected, anti 5hmC stained C2C12 mouse myoblasts were acquired using the Operetta automated imaging system with a 20× long/0.45 NA objective (PerkinElmer, UK), a xenon fiber-optic as light source, 360–400, 460–490 and 560–580 nm excitation- and 410–480, 500–550 and 590–640 emission filters, respectively. Representative images of the same cells were acquired using a Leica TCS SP5 II confocal laser scanning microscope (Leica Microsystems, Wetzlar, Germany) with a Plan-Apochromatic 100×/1.44 NA oil objective and 405, 488 and 561 nm excitation lasers.

For the analysis of Mecp2-, 5mC-, 5hmC-, NeuN- and Tet1 levels, colocalization studies of Mecp2 and 5hmC, as well as for the detection of major satellite RNA FISH signals in Mecp2 y/- and wild-type mouse brain, 3D z-stacks were acquired using the Operetta automated imaging system with a 2× long/0.08 and 20× long/0.45 NA objective (PerkinElmer, UK), a xenon fiber optic as light source, 360–400, 460–490 and 560–580 nm excitation- and 410–480, 500–550 and 590–640 emission filters, respectively.

Image analysis and quantification

Fluorescence intensity histogram quantification of images acquired on the Operetta automated imaging system (Figures 1B and 3B) was performed using the Harmony 3.5.1 software (PerkinElmer, UK). Nuclei were detected based on Tet1CD signals and further selected pursuant to morphology properties (area and roundness). For each nucleus, mean intensities for Tet1CD- and MBD proteins (Mecp2, IDTRD, MBD, Mbd2 and Mbd3, respectively), as well as for 5hmC were calculated. After background subtraction, nuclei were binned according to (i) Tet1CD signal (sub-group) and (ii) to MBD signal (sub-subgroup). For each independent experiment, mean 5hmC level were averaged per sub-subgroup and normalized to highest 5hmC level of Tet1CD + GFP transfected cells. To automate this procedure, a routine was written in the programming language R.

For calculation of mean Mecp2-, 5mC- and 5hmC level at pericentric heterochromatin of Mecp2 y/- and wild-type mouse brain (Figure 5), mid-optical sections of the 5mC channel were used to generate chromocenter (CC) masks. Therefore, images were processed using a median filter and

thresholded in three successive steps using the basic algorithm. For the generation of binary chromocenter masks, all pixels below the final threshold were set to 0 and all pixels above the final threshold were set to 1. Total Mecp2-, 5mC- and 5hmC- signals overlapping with the chromocenter mask were calculated and divided by the total number of pixels corresponding to the area of chromocenters. To automate this procedure, a routine was written in the programming language python (<http://code.google.com/p/priithon/>).

To determine the accumulation of 5hmC at pericentric heterochromatin in neurons of Mecp2 *y/-* and wild type mouse pontes (Supplementary Figure S16D), mean 5hmC signals at chromocenters were divided by mean 5hmC signals within the nucleoplasm. Therefore, binary chromocenter masks were generated as described above. Nucleoplasm masks were prepared by subtracting a second chromocenter mask with larger surface area from a nuclear mask. For this purpose, chromocenter masks were generated as described above, except that mid-optical sections of the 5mC channel were filtered only once. Binary nuclear masks were prepared by filtering and thresholding the DAPI channel as described earlier. To further improve the nuclear mask, holes were filled using the fill holes algorithm and background was removed via the watershed algorithm. To calculate mean 5hmC signals at both, chromocenters and nucleoplasm, total 5hmC signals overlapping with either of the two masks, were divided by the total number of pixels corresponding to the area of chromocenter- and nucleoplasm masks, respectively. To automate this procedure, a routine was written in the programming language python.

Colocalization of 5hmC and chromocenters in neurons of Mecp2 *y/-* and wild type mouse brain (Figure 6A) was assessed by the H-coefficient (52) as previously described (53) and by line profiles generated with ImageJ (<http://rsb.info.nih.gov/ij/>).

RNA level of major satellite DNA in neurons of Mecp2 *y/-* and wild type mouse brain (Figure 6B) were calculated manually by measuring nuclear RNA FISH signals along a line through pericentric heterochromatin (50 pixel in length) using ImageJ (<http://rsb.info.nih.gov/ij/>) (Supplementary Figure S16A).

For quantification of mean nuclear Tet1 level in wild type and Mecp2 *y/-* brain (Supplementary Figure S14), binary nuclear masks were generated as described above. Total Tet1 signals overlapping with the nuclear mask were calculated and divided by the total number of pixels corresponding to the nuclear area. To automate this procedure, a routine was written in the programming language python.

RNA-seq library preparation

Total RNA was isolated from wild type and triple Tet1 knockout mouse embryonic stem cells (V6.5) in biological quadruplicates using the nucleospin triprep kit (Macherey Nagel, Düren, Germany). 50 ng RNA was reverse transcribed. cDNA was pre-amplified as described elsewhere (54). One ng of cDNA was used as input for tagmentation by the Nextera XT Sample Preparation Kit (Illumina, San Diego, CA, USA), where a second amplification round was

performed for 12 cycles. For each sample, 2.5 ng of final library were pooled.

RNA-seq and data analysis

One hundred base pairs single end reads were sequenced on an Illumina HighSeq 1500. Libraries were barcoded and mixed before sequencing. The resulting reads were mapped to the Mouse genome build mm10 using STAR version STAR 2.5.1b (55) with the specific settings:

```

-outFilterMultimapNmax 100
-outFilterMismatchNmax 4
-winAnchorMultimapNmax 100
→ 100. The junction annotation was taken from ensembl GRCm38.75 and the index was created as recommended using the option -sjdbOverhang → 99.

```

The resulting bam-files were then processed using TETranscript (56) to obtain read count tables for transcripts and transposons, using the TE annotation as provided by the authors of TETranscript (http://labshare.cshl.edu/shares/mhammelllab/www-data/TEToolkit/TE.GTF/mm10rmskT_E.gtf.gz). Normalization and differential expression analysis was done using DESeq2 (57).

Statistical analyses

For Figures 1A and 3A, Tet1CDmut + GFP and mock were excluded from statistical tests as the mean values were at the background level. Homogeneity of variance was tested beforehand with Levene's test (using the median). The Levene's test did not indicate heterogeneous variances ($P > 0.1$) between the groups. Hence, we conducted repeated measures ANOVA for the three replicates in both experiments, which showed highly significant results ($F_{4,8} = 46.1$, $P < 0.001$ and $F_{4,8} = 22.7$, $P < 0.001$). Therefore, post-hoc pairwise *t* tests with false discovery rate correction (58) were performed.

For Figures 1B and 3B and Supplementary Figure S4, we performed Welch's ANOVA which gave a highly significant result ($F_{6,97} = 331.4$, $P < 0.001$), since the variances were heterogeneous (Levene's test: $P < 0.001$). Then, we performed post-hoc pairwise *t* tests with non-pooled standard deviations and false discovery rate correction.

For Figure 3B (lower row), since the variances were not significantly different (Levene's test: $P = 0.19$), we compared the means with a *t* test.

For Figures 2, 3C and 6E, we performed independent two-sample Student's *t*-test.

For Figures 5 and 6A and 6C, as well as Supplementary Figures S14 and S16, we performed a Wilcoxon signed-rank test.

All statistical tests were conducted with R (<https://www.r-project.org/>).

RESULTS AND DISCUSSION

Mecp2 and Mbd2 protect 5-methylcytosine from Tet1 mediated oxidation in a concentration dependent manner

Considering that MBD and Tet proteins share a common substrate, we aimed at clarifying whether binding of MBD proteins to methylated DNA protects epigenetically

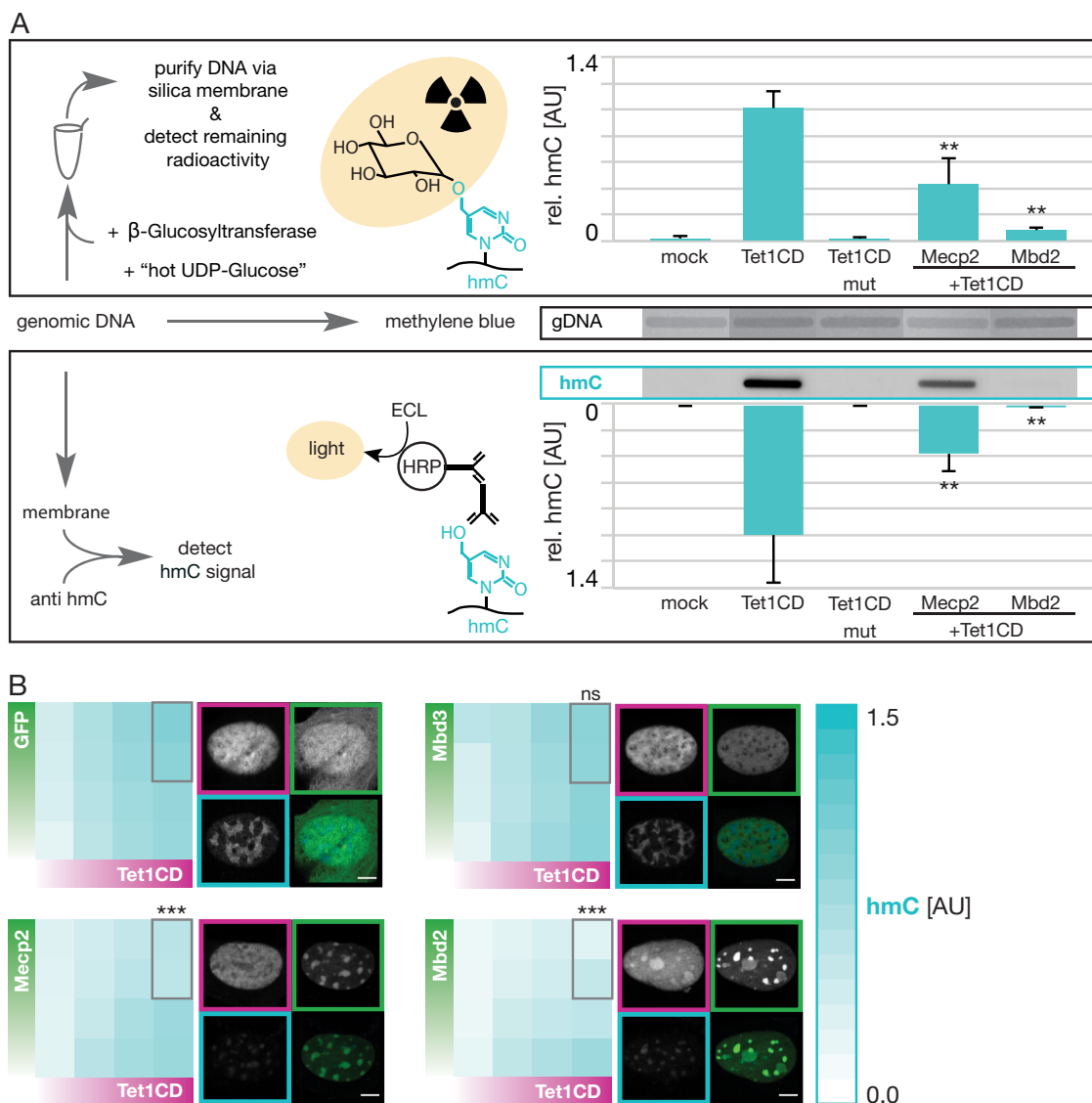


Figure 1. Impact of methyl-CpG binding domain proteins on the efficiency of Tet1 mediated 5mC oxidation. **(A)** Radioactive (top) and immunological (bottom) assay to determine 5hmC levels in genomic DNA (gDNA) of transiently transfected HEK cells. Schemes (left) illustrate the workflow and mode of 5hmC detection. Histograms (right) show relative 5hmC levels of three independent experiments \pm SD. Tet1CD + Mecp2 and Tet1CD + Mbd2 differed significantly ($**P < 0.01$; post-hoc pairwise *t* test) from Tet1CD (see methods for details). gDNA quantities were monitored by methylene blue staining. Tet1CD corresponds to Tet1 catalytic domain and Tet1CDmut is the catalytic domain of Tet1 containing two point mutations that abolish binding of the co-factor Fe^{2+} (see also Supplementary Figure S2). Full blots are shown in Supplementary Figure S1. **(B)** *In situ* staining and quantification of genomic 5hmC levels in transiently transfected C2C12 mouse myoblasts. Images were acquired on an automated high throughput imaging system with a $20\times$, 0.45 NA objective. Gradient heat maps show relative 5hmC (= cyan) signals as a function of increasing Tet1CD (= magenta) and MBD (= green) protein expression levels depicted by the green and magenta gradient bars. Shown are mean values of five (Tet1CD + GFP, $n = 12\,798$), four (Tet1CD + Mbd2, $n = 2598$) or two (Tet1CD + Mecp2, $n = 4760$; Tet1CD + Mbd3, $n = 6449$) independent experiments, respectively. For statistical tests, 5hmC signals of cells with high Tet1CD and high Mecp2/Mbd2/Mbd3 protein levels (framed in grey) were used. Tet1CD + Mecp2 and Tet1CD + Mbd2 differed highly significantly ($***P < 0.001$; post-hoc pairwise *t* test) from Tet1CD + GFP. No significant difference was detected for Tet1CD + Mbd3 ($P = 0.34$). Confocal images of mid optical sections of the same samples represent transiently transfected C2C12 mouse myoblasts (Tet1CD = magenta; GFP/MBD proteins = green) immunostained for 5hmC (5hmC = cyan). Scale bar, 5 μm .

silenced regions from Tet-mediated DNA demethylation. To this end, we either radioactively (Figure 1A, top), or immunologically (Figure 1A, bottom and Supplementary Figure S1) labeled and subsequently quantified global 5hmC levels in genomic DNA of FACS sorted HEK cells expressing comparable levels of the catalytic active (Tet1CD) and inactive domain (Tet1CDmut) of Tet1 alone, or in combination with Mecp2 and Mbd2, respectively (Figure 1A and Supplementary Figure S2). When compared to mock and Tet1CDmut transfected cells, both, the radioactive (Figure 1A, top), as well as the immunological (Figure 1A, bottom) assay revealed increased 5hmC levels in genomic DNA of cells expressing the catalytic active domain of Tet1 alone. Coexpression of Mecp2 or Mbd2, however, significantly decreased global 5hmC levels by at least 50%, demonstrating reduced Tet1 effectiveness in the presence of substrate-competitive proteins.

Further single-cell analysis (Supplementary Figure S3) of transiently transfected mouse myoblasts (Figure 1B) and HEK cells (Supplementary Figure S4) immunostained for 5hmC revealed a correlation between Tet1CD protein and 5hmC levels in a subpopulation of cells containing low Mecp2 (Figure 1B, bottom, left and Supplementary Figure S4, bottom, left) and Mbd2 (Figure 1B, bottom, right and Supplementary Figure S4, top, right) protein amounts, respectively. The remainder cells of the population, characterized by high expression levels of the Mbd2 and Mecp2 proteins, in contrast, showed no longer any correlation between Tet1CD protein levels and the occurrence of its oxidation product. Instead, 5hmC levels anti-correlated with increasing levels of Mecp2 (Figure 1B, bottom, left and Supplementary Figure S4, bottom, left) and Mbd2 (Figure 1B, bottom, right and Supplementary Figure S4, top, right), respectively, indicating that protection of 5mC from Tet1 catalyzed oxidation highly depends on the MBD protein concentration. In contrast to Mecp2 and Mbd2, even the highest expression levels of GFP (Figure 1B, top, left and Supplementary Figure S4, top, left) and Mbd3 (Figure 1B, top, right) proved insufficient to repress Tet1 activity. As both proteins are not capable of binding to (methylated) DNA (for Mbd3, see Supplementary Figure S5A), this suggests that direct interaction with (methylated) DNA is a prerequisite for the effective conservation of 5mC.

To determine, whether the levels of mCherry-Tet1CD obtained through overexpression in mouse myoblasts and human embryonic kidney cells are within the physiological range of endogenous Tet1 in primary mouse neurons and mouse embryonic stem cells (ESCs), we stained all of the four cell types for Tet1 and quantified the resulting immunofluorescent signals (Supplementary Figure S6). To allow a direct comparison to the Tet1 expression levels plotted in Figure 1B and Supplementary Figure S4, transiently transfected mouse myoblasts and HEK cells were binned according to the ectopic Tet1CD signal (e.g. group 1 of Supplementary Figure S6 corresponds to the first column of all heatmaps in Figure 1B and Supplementary Figure S4, respectively). We found that bins 1 and 2 of mouse myoblasts, as well as bin 1 of HEK cells express (combined ectopic+endogenous) Tet1 levels comparable to mouse ESCs. As shown previously, the level of overexpressed Mecp2 in mouse myoblasts is in the range of endogenous physiolog-

ical Mecp2 levels per mouse neuronal cell nucleus (36,59). Accordingly, cells expressing low Tet1CD protein levels do not create an artificial phenotype and, thus, reflect the situation *in vivo*.

Finally, we immunologically quantified 5hmC levels in genomic DNA of Mecp2 lox/y and -/y mouse tail fibroblasts (MTF). Compared to the floxed control group (lox/y), we detected increased 5hmC levels in the corresponding Mecp2 knockout cell line (-/y), indicating that Mecp2 represses the formation of 5hmC *in vivo* (Supplementary Figure S7).

Prior binding of Mecp2 and Mbd2 to 5-methylcytosine enhances blocking of Tet1 catalyzed 5-hydroxymethylcytosine formation *in vitro*

As described above, Tet1 mediated 5hmC formation is impaired by Mecp2 and Mbd2 *in vivo*. To gain a deeper understanding of the protective mechanism, we next sought to determine, whether the chronological order of DNA binding by MBD and Tet1 proteins would influence the extent of 5mC protection. Since at the cellular level the chronological access of MBD- and Tet1CD proteins to DNA is difficult to control, we further investigated its influence on 5mC protection in *in vitro* experiments. Therefore, various conceivable binding scenarios were systematically mimicked on a molecular level. Briefly, same molar ratios of Tet1CD and MBD proteins were incubated simultaneously or consecutively with a PCR fragment containing multiple methylated cytosines. Following 2 h of Tet1CD incubation, DNA was blotted on a membrane to then immunologically detect the amount of remaining unoxidized 5mC, as a measure of 5mC protection by MBD proteins (Figure 2A and Supplementary Figure S8).

Altogether, 5mC levels were comparatively high in any of the samples containing, in addition to Tet1CD, Mecp2 and Mbd2 (Figure 2B and C), respectively, indicating restricted Tet1CD activity in the presence of 5mC binding proteins, which is in accordance with our *in vivo* data (Figure 1). In more detail, when compared to the fully unprotected control group (DNA + Tet1CD) (Figure 2B and C, first column), 5mC levels were increased by a factor of 1.9 (Mecp2) and 1.6 (Mbd2) in samples allowing simultaneous access of Tet1CD and MBD proteins to their common substrate 5mC (Figure 2B and C, second column). Incubation of MBD proteins with the methylated DNA prior to the addition of Tet1CD enzymes, yielded 5mC signals 3.9 (Mecp2) and 2.1 (Mbd2) fold higher relative to control samples without MBD proteins (Figure 2B and C, third column). Delayed addition of methylated DNA to pre-incubated Tet1CD and MBD proteins resulted in relative 5mC levels of 2.1 (Mecp2) and 1.3 (Mbd2) (Figure 2B and C, fourth column). Accordingly, among all tested conditions, early incubation of MBD proteins with methylated DNA before the addition of Tet1CD enzymes revealed the highest 5mC signals and, thus, the best possible protection against Tet1CD catalyzed oxidation (Figure 2B and C, third column).

As Mecp2 can bind to a single methylated CpG site (mCpG), we further tested whether the protection against Tet oxidation could take place on single mCpG containing substrates. Therefore, we measured the degree of protection

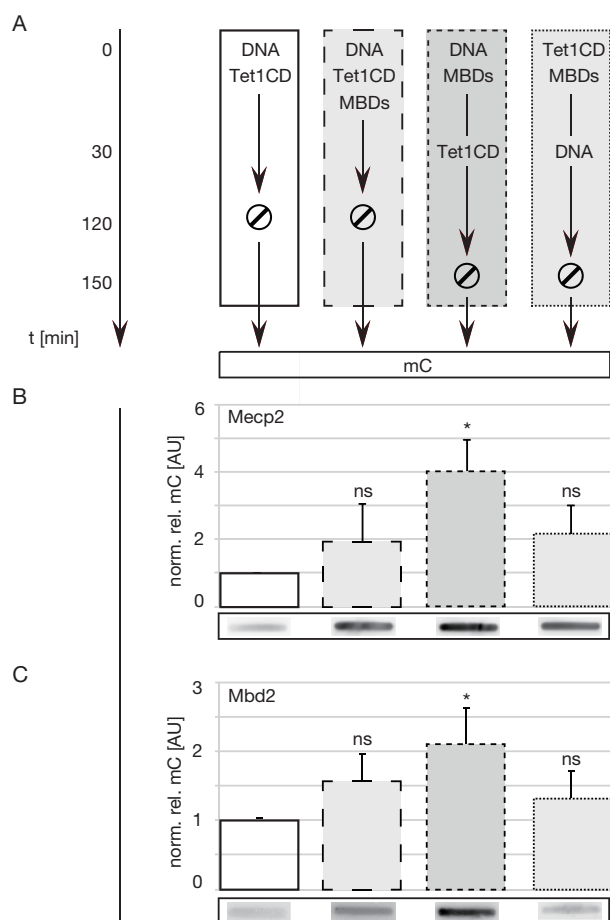


Figure 2. Influence of the chronological DNA binding order on the protecting ability of MBD proteins. Quantification of remaining 5mC levels in 20 pmol of double-stranded DNA containing multiple 5mC nucleotides after simultaneous or successive incubation with 20 pmol of Tet1CD (catalytic domain) and 20 pmol of MBD proteins by slot blot. (A) Experimental setup illustrating the incubation order and time of proteins and oligos prior slot blotting. (B and C) Histograms show relative 5mC levels of 5mC containing PCR product after incubation with (B) Tet1CD and Mecp2 ($n = 4$) and (C) Tet1CD and Mbd2 ($n = 4$). Shown are mean values + SD. Significant differences were detected for pre-incubation of Mecp2 and Mbd2 with DNA ($*P < 0.05$; ns = non significant; post-hoc pairwise t tests). Full blots are shown in Supplementary Figure S8.

over time using a methyl sensitive restriction endonuclease assay (35). Consistent with our previous results, single 5mC containing, double-stranded oligonucleotides that were pre-incubated with Mecp2 lost the least amount of 5mC. Even after 2.5-h treatment with Tet1CD, the presence of Mecp2 (previously incubated with DNA) protected 80% of 5mC from oxidation versus only 20% of 5mC surviving in the absence of Mecp2 (Supplementary Figure S9).

On this basis, we conclude that binding of MBD proteins to DNA, especially prior binding provides the greatest contribution towards preserving the methylation status of CpG dinucleotides. Alternatively, protein-protein interactions, which could have formed most effectively by pre-

incubation of Tet1CD and MBD proteins, do not play a role in protecting DNA from Tet1CD driven oxidation.

Direct binding to DNA is sufficient to effectively prevent 5mC oxidation by Tet1

As indicated earlier, Mecp2 contains various different interaction sites for DNA. While the IDTRD domain of Mecp2 was shown to bind both, methylated and unmethylated DNA with similar affinity (9,10,15), the MBD domain of Mecp2 has a preference for methylated CpG dinucleotides (14,60).

To test whether and which of the above-mentioned binding mode is responsible for the conservation of 5mC, we tested the protecting ability of both Mecp2 subdomains using the battery of assays employed before (Figure 1 and Figure 2). Quantification of 5hmC levels in genomic DNA of human HEK cells revealed that both subdomains of Mecp2 avert Tet1CD mediated 5mC oxidation to a similar extent, indicating that the adverse impact on Tet1CD activity does not directly correlate with 5mC affinity (Figure 3A). To further test this conclusion, we measured the protective effect of full length Mecp2 proteins carrying an R111G mutation in their MBD domain. While mutation of arginine 111 abolishes binding to 5mC *in vitro* and to methylated heterochromatin *in vivo* (61), the mutant protein is still able to interact with unmethylated DNA in a sequence unspecific manner. Accordingly, it shifts 5mC containing DNA in the absence, but less efficiently in the presence of poly dI:dC competitor DNA (Supplementary Figure S5B). As cells expressing this mutant Mecp2 variant also exhibited low 5hmC levels (Supplementary Figure S10A left), we deduce that sequence unspecific DNA interactions, considerably contribute to defending 5mC from Tet1CD mediated oxidation. Additional 5mC recognition by a functional MBD, as demonstrated by wild type Mecp2, improved the protecting ability only marginally (Supplementary Figure S10A left). Indeed, we found that two proteins specifically binding major satellite DNA sequences independent of methylation, the polydactyl zinc finger MaSat (62), as well as the transcription activator-like effector protein msTALE (63), repressed the formation of 5hmC *in situ* (Supplementary Figure S10A, right). Thus, 5mC recognition by the MBD is unlikely to *per se* play a major role in the protective mechanism.

To validate and extend these results, we immunostained and quantified genomic 5hmC levels in single C2C12 mouse myoblasts (Figure 3B) and HEK cells (Supplementary Figure S4). Similar to C2C12 and HEK cells coexpressing Mecp2 and Mbd2 (Figure 1B and Supplementary Figure S4), genomic 5hmC content correlated with Tet1CD protein levels in a subpopulation of cells containing low IDTRD protein amounts (Figure 3B, top, left and Supplementary Figure S4, bottom, middle; rows 3 and 4 from top to bottom). Cells, characterized by high expression levels of IDTRD proteins, in contrast, showed no longer any correlation between Tet1CD protein levels and its oxidation product. Instead, 5hmC levels anti-correlated with increasing levels of IDTRD (Figure 3B, top, left and Supplementary Figure S4, bottom, middle; rows 1 and 2 from top to bottom). In C2C12 cells, the MBD domain of Mecp2 repressed merely the catalytic activity of a small number of

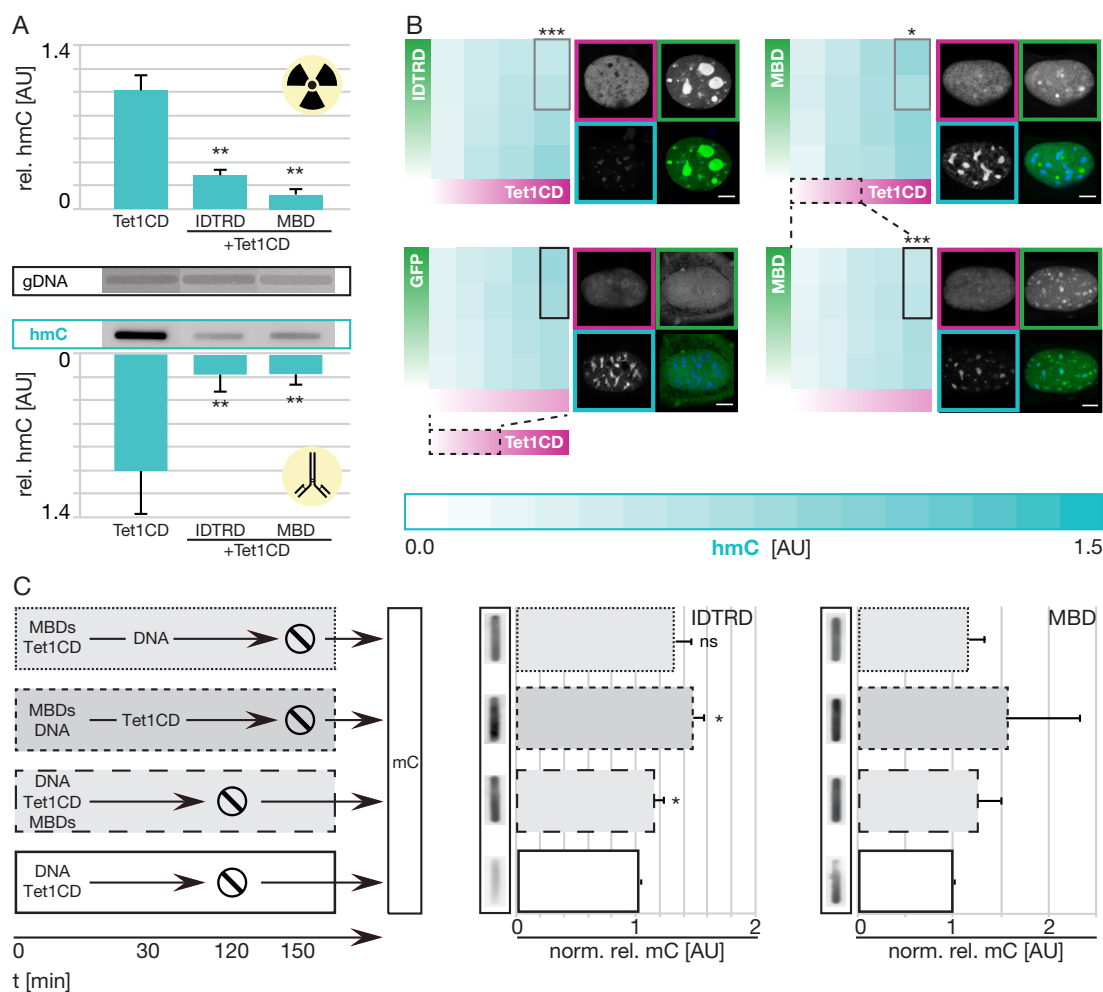


Figure 3. Effect of different DNA binding modes on Tet1 activity. **(A)** Radioactive (top) and immunological (bottom) assay to determine 5hmC levels in genomic DNA (gDNA) of transiently transfected HEK cells. Histograms show relative 5hmC levels of three independent experiments + SD. Tet1CD + IDTRD and Tet1CD + MBD differed significantly (** $P < 0.01$; post-hoc pairwise t tests) from Tet1CD + GFP. gDNA quantities were monitored by methylene blue staining. Tet1CD corresponds to Tet1 catalytic domain and IDTRD and MBD correspond to the subdomains of Mecp2 (Supplementary Figure S2). Full blots are shown in Supplementary Figure S8. **(B)** *In situ* staining and quantification of genomic 5hmC levels in transiently transfected C2C12 mouse myoblasts. Images were acquired on an automated high throughput imaging system with a 20 \times , 0.45 NA objective. Gradient heat maps show relative 5hmC (= cyan) signals as a function of Tet1CD (= magenta) and Mecp2 subdomain (= green) levels depicted by the green and magenta gradient bars. Shown are mean values of two (Tet1CD+IDTRD, $n = 6495$; Tet1CD + MBD, $n = 1800$) independent experiments. **(B, upper row)** For statistical tests, 5hmC signals of cells with high Tet1CD and high IDTRD/MBD protein levels (framed in gray) were used. Tet1CD + IDTRD differed highly significantly (** $P < 0.001$; post-hoc pairwise t tests) from Tet1CD + GFP (see Figure 1B). Weakly significant difference was detected for Tet1CD + MBD ($*P < 0.05$; post-hoc pairwise t tests). **(B, lower row)**. For statistical tests, 5hmC signals of cells with low Tet1CD and high MBD protein levels were used. From these, we selected the values framed in black. Highly significant differences were detected between both groups (** $P < 0.001$; t test). Confocal images of mid optical sections of the same samples represent transiently transfected C2C12 mouse myoblasts (Tet1CD = magenta; GFP/MBD proteins = green) immunostained for 5hmC (5hmC = cyan). Scale bar, 5 μ m. **(C)** Quantification of remaining 5mC levels in double-stranded methylated DNA after simultaneous or successive incubation with Tet1CD proteins and Mecp2 subdomains by slot blot. Scheme illustrates the incubation order and time of proteins and methylated DNA prior slot blotting. Histograms show relative 5mC signals after incubation with Tet1CD and IDTRD ($n = 4$), as well as Tet1CD and MBD ($n = 5$), respectively. Shown are mean values + SD. Significant differences were detected for pre-incubation of IDTRD with DNA and for simultaneous incubation of IDTRD, Tet1CD and DNA ($*P < 0.05$; ns = non significant; post-hoc pairwise t tests). For MBD, a similar trend was detected, however, without statistical significance.

Tet1CD molecules (Figure 3B, top, right (columns 1 and 2 from left to right) and Figure 3B, bottom). In the presence of high Tet1CD protein amounts, even the highest MBD protein concentrations failed to repress the formation of 5mC by Tet1CD (Figure 3B, top, right (columns 3 and 4 from left to right)). In HEK cells, which expressed ectopic proteins at a substantially higher level per cell than the previously analyzed C2C12 cells (Supplementary Figure S11 and Supplementary Figure S6), however, MBD protein levels were sufficient to avert the catalytic activity of low to high Tet1CD protein levels (Supplementary Figure S4, bottom, right). Accordingly, we conclude that the extent of 5mC protection substantially depends on the concentration of MBD and IDTRD molecules per cell as it determines the coverage of DNA in a sequence-unspecific manner (see also Figure 4).

Similar to the MBD domain of Mecp2, Mbd2 has been shown to preferentially bind 5mC (64,65). In contrast, though, Mbd2 was more efficient than the MBD in protecting 5mC from oxidation (Figures 1 and 3). To test whether binding kinetics *in vivo* may contribute to 5mC protection, we performed fluorescence recovery after photobleaching (FRAP) experiments. Whereas the MBD showed fast recovery at pericentric heterochromatin with halftimes of 2 s, Mbd2 recovered 15-fold slower after photobleaching (30 s) (Supplementary Figure S10B and S10C). Hence, we propose that long retention times of Mbd2 at methylated cytosines improve the efficiency of 5mC protection. However, since the IDTRD subdomain depicted similar fast recovery kinetics (2 s) like the MBD domain (Supplementary Figure S10B and 10C) but was more efficient in protecting 5mC, we deduce that additional sequence-unspecific DNA binding parameters (e.g. stoichiometry) must play a role.

Finally, *in vitro* 5mC oxidation studies using a PCR fragment containing multiple methylated cytosines showed that similar to Mecp2 and Mbd2, prior binding of MBD and IDTRD to DNA additionally strengthens the conservation of 5mC (Figure 3C and Supplementary Figure S8).

In summary, these data highlight the complexity of the MBD based 5mC protection mechanism, which achieves best performance through prior and long lasting coverage of DNA in a sequence-unspecific manner. It also differs from previous reports (66) suggesting that 5mC binding *per se* is critical to protect from oxidation.

Binding of Mecp2 to DNA impairs the DNA binding ability of Tet1CD *in vitro*

Tet-mediated oxidation of 5mC was recently described as a complex, multistep process, initiated by the binding of Tet proteins to DNA via hydrophobic interactions, followed by recognition of 5mC in CpG context, base flipping and oxidation (67,68).

To investigate how prior binding of MBD proteins to DNA protects 5mC from Tet catalyzed oxidation (Figures 1-3), we next analyzed which of the above-mentioned step is affected. To this end, we tested the DNA binding ability of Tet proteins, considered as the first step towards 5mC oxidation, in the presence of low and high concentrations of 5mC specific (MBD, Figure 4B and Supplementary Figure S12B), as well as sequence-unspecific (IDTRD, Figure

4C and Supplementary Figure S12B) Mecp2 DNA binding domains by electrophoretic mobility shift assays (EMSA) (Figure 4 and Supplementary Figure S12A).

To validate that both Mecp2 subdomains bind to DNA under the given reaction conditions, we initially verified their efficiency to form complexes with short double-stranded (ds), single 5mC-containing DNA in the absence of Tet1CD proteins (Figure 4A-C, first and second row). While incubation of DNA with low substoichiometric MBD (Figure 4B, first row) and IDTRD (Figure 4C, first row) protein concentrations, resulted in a single, slow migrating band, increasing protein amounts (Figure 4B-C, second row) gave rise to an additional high molecular super-shift, originating from additive accumulation of proteins to an already bound DNA molecule. Hence, our data prove the suitability of the present reaction conditions. Furthermore, it indicates that at high protein levels, where most of the unbound DNA substrate is depleted, multiple binding of either Mecp2 subdomain to the same DNA fragment is promoted (Figure 4A-C, compare first and second rows). As the ratio of high molecular weight versus low molecular weight shifted DNA in the IDTRD is higher than with the MBD, we conclude that the IDTRD of Mecp2 is more efficient in fully covering DNA molecules than the MBD (Figure 4B and C, second row).

Addition of Tet1CD molecules (Supplementary Figure S12B) to single 5mC containing dsDNA, which was pre-incubated with low protein amounts of the MBD or IDTRD, respectively, resulted in two discrete prominent DNA shifts (Figure 4A-C, third row). While the fast migrating DNA co-localized with MBD (Figure 4B, third row) and IDTRD (Figure 4C, third row) protein signals, respectively, the high molecular DNA band coincided with protein signals for the catalytic domain of Tet1 (Figure 4B and C, third row). Consequently, our data indicate that in the presence of low competitive protein concentrations and excess availability of uncovered DNA substrate, Tet1CD binds to DNA without compromising efficiency (Figure 4B and C, third row). Pre-incubation of dsDNA with a higher number of IDTRD molecules, in contrast, greatly diminished Tet1CD signals (arrowhead), which instead strongly colocalized with signals derived from IDTRD proteins (arrow, Figure 4C, fourth row). Thus, we conclude that under the present reaction conditions, under which most of the DNA substrate is covered by IDTRD molecules (Figure 4C, second row), binding of Tet1 proteins to their common substrate is almost entirely averted. Similar to the IDTRD, however, as a result of lower DNA coverage, less significant, higher protein level of MBD (arrow) reduced the formation of Tet1-DNA complexes (arrowhead) (Figure 4B, fourth row). Similar results were obtained with equimolar amounts of MBD/IDTRD and Tet1CD proteins (Supplementary Figure S13).

In summary, these data demonstrate that the amount of free Tet1CD enzyme highly correlates with the number of bound methyl-CpG binding domain proteins per DNA molecule. Hence, we conclude that binding of MBD proteins to DNA protects 5mC from oxidation by restricting access of Tet1CD enzymes to DNA, whereby any further steps of the oxidation procedure are inhibited. Besides this, we deduce that the efficiency of Tet DNA binding inhibi-

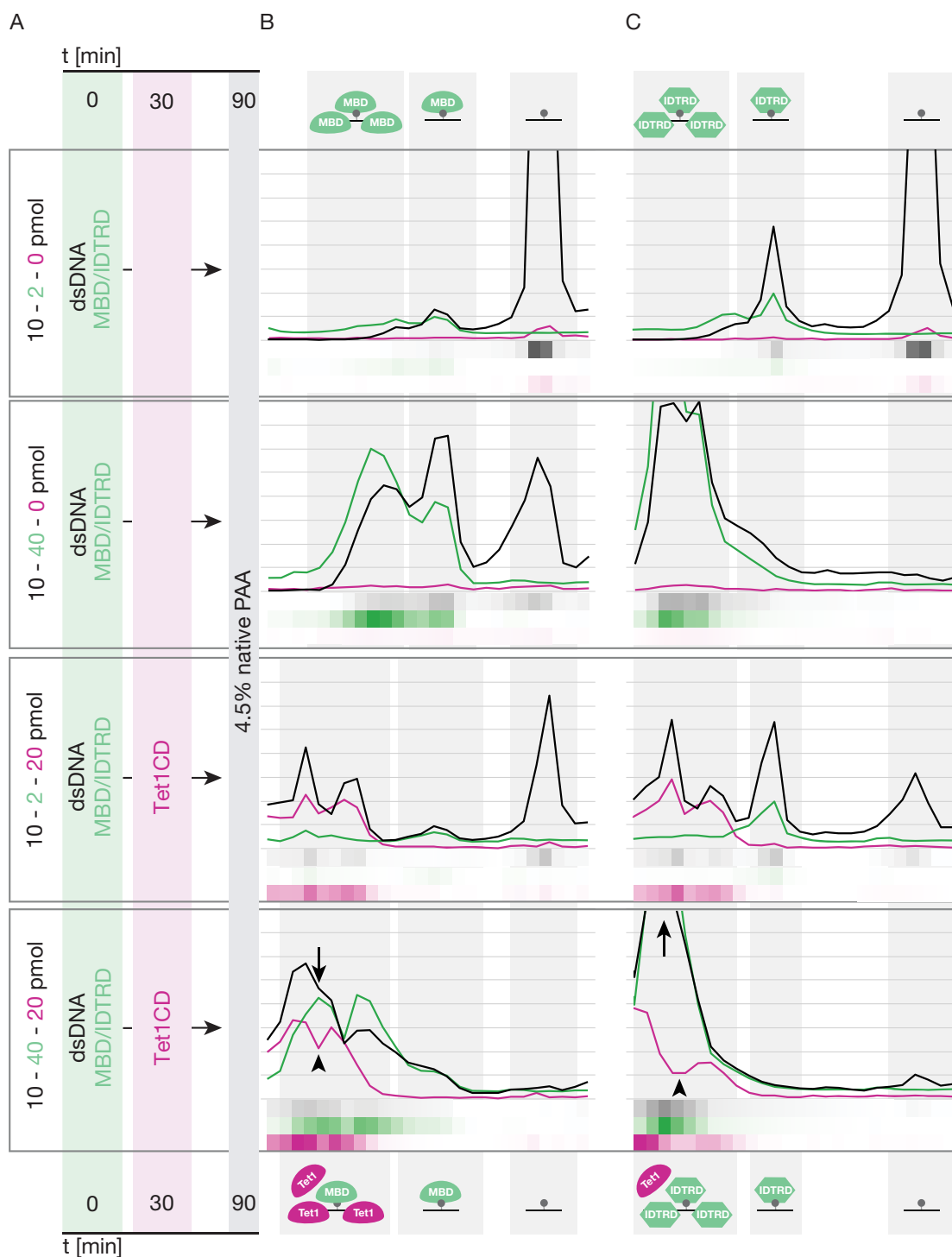


Figure 4. Impact of 5mC-specific and sequence-unspecific DNA binding proteins on the DNA binding ability of Tet1CD proteins. Electrophoretic mobility shift assay (EMSA) to determine the binding ability of fluorescently tagged Tet1CD to double-stranded, single mC containing DNA (ATTO647 labeled) in the presence of low and high amounts of (B) 5mC specific (fluorescently tagged MBD) and (C) sequence-unspecific (fluorescently tagged IDTRD) DNA binding domain proteins, respectively. (A) Experimental setup illustrating the amount, as well as the incubation order and time of proteins and DNA prior EMSA. (B) Separation of MBD-dsDNA (top, $n = 3$), MBD-Tet1CD-dsDNA (bottom, $n = 3$), as well as (C) IDTRD-dsDNA (top, $n = 3$) and IDTRD-Tet1CD-dsDNA (bottom, $n = 3$) complexes via electrophoresis through a native polyacrylamide gel visualized using a fluorescent plate reader (see also Supplementary Figure S12). Running direction is from left (– pole) to right (+ pole).

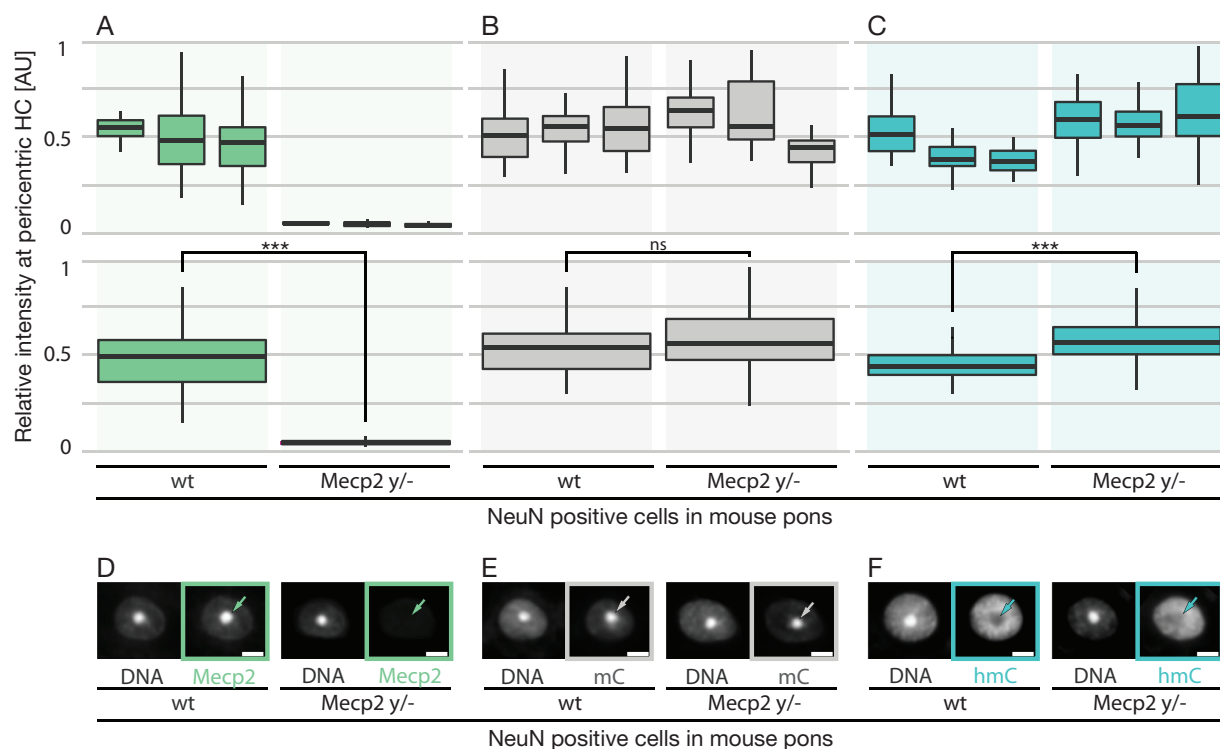


Figure 5. Epigenetic composition of pericentric heterochromatin in a mouse model for Rett syndrome. Immunostaining and quantification of *Mecp2*, 5mC and 5hmC levels in NeuN positive cells of wild type and *Mecp2* knockout (*y/-*) mouse pons, respectively. (A–C) Box plots represent the distribution of (A) *Mecp2* (wt1 *n* = 18, wt2 *n* = 38, wt3 *n* = 47, KO1 *n* = 18, KO2 *n* = 39, KO3 *n* = 31; from left to right), (B) 5mC (wt1 *n* = 42, wt2 *n* = 39, wt3 *n* = 44, KO1 *n* = 34, KO2 *n* = 29, KO3 *n* = 20; from left to right) and (C) 5hmC (wt1 *n* = 42, wt2 *n* = 39, wt3 *n* = 44, KO1 *n* = 34, KO2 *n* = 29, KO3 *n* = 20; from left to right) levels at pericentric heterochromatin (HC) in neurons of three individual (top) and combined (bottom) wild type and *Mecp2* *y/-* mouse pons, respectively (*n* = number of cells). Plotted is the median, as well as the first and third quartiles. Whiskers extend to 1.5 times the interquartile range. *P* values were calculated by Wilcoxon signed-rank test (***) *P* < 0.001, ns = non significant). (D–F) Mid-confocal optical sections of NeuN positive cells of wild type and *Mecp2* *y/-* mouse pons immunostained for (D) *Mecp2*, (E) 5mC and (F) 5hmC, respectively. Arrows point to pericentric heterochromatin. DNA was counterstained with DAPI. Scale bar, 5 μm.

tion, which is proportional to DNA coverage by MBD proteins, may differ from cell type to cell type, as the binding mode of MBD proteins is strongly affected by DNA methylation density and binding partners (14).

The Tet oxidation product 5hmC is enriched in neurons of a mouse model for Rett syndrome

As we found that *Mecp2* represses Tet1-mediated 5mC oxidation *in vivo* (Figure 1) and *in vitro* (Figure 2), we next tested whether the previously reported transcriptional increase of repetitive elements in *Mecp2* knockout brain (20), may be considered as pathophysiological consequence of unconfined Tet activity. To address this hypothesis, we analyzed genomic 5mC- and 5hmC levels in the pons (Supplementary Figures S14A and S15) of a mouse model for Rett syndrome (*Mecp2*^{-/y^{tm1.1}Bird}), which was previously identified as brain region partially responsible for the devastating breathing disturbances observed in Rett patients (69). Since in wild type brain *Mecp2* was primarily found at pericentric heterochromatin of neurons (Figure 5A and Supplementary Figure S14B and C), we consequently focused our analysis to these chromatin regions, which in mouse cells assemble

into higher order aggregates known as chromocenters (CC) (Supplementary Figure S15).

While knockout of *Mecp2* had little effect on the distribution of pericentric 5mC levels (Figure 5B), the amount of the Tet oxidation product 5hmC was significantly increased at chromocenters of *Mecp2* deficient neurons of the pons (Figure 5C), which is in agreement with previous data (66). Using LC–MS it has been shown in different brain regions, that 4.5% of all cytosines are methylated and 5mC levels do not change between regions. Moreover, 5hmC levels were shown to vary between 0.3% and 0.6%, with an average 0.45%, i.e. 10 times lower than 5mC (70). Using similar methods, Wu and colleagues showed that the distribution of 5hmC, but not of 5mC, varies between tissues. They, furthermore, showed that in Tet1 knockdown ES cells, although the 5hmC decreased to less than half of the control cells, 5mC did not change (71), which is similar to our results. As we measured a change of 40% for 5hmC in *Mecp2* knockout relative to wild type neurons, we would expect maximally 4% change of 5mC levels in a pure population of neuronal cells. According to Münzel *et al.* (70) the maximally expected change of 4% lies within the experimental error rate ($\pm 5\%$ for 5mC and 5hmC) and is, therefore, most

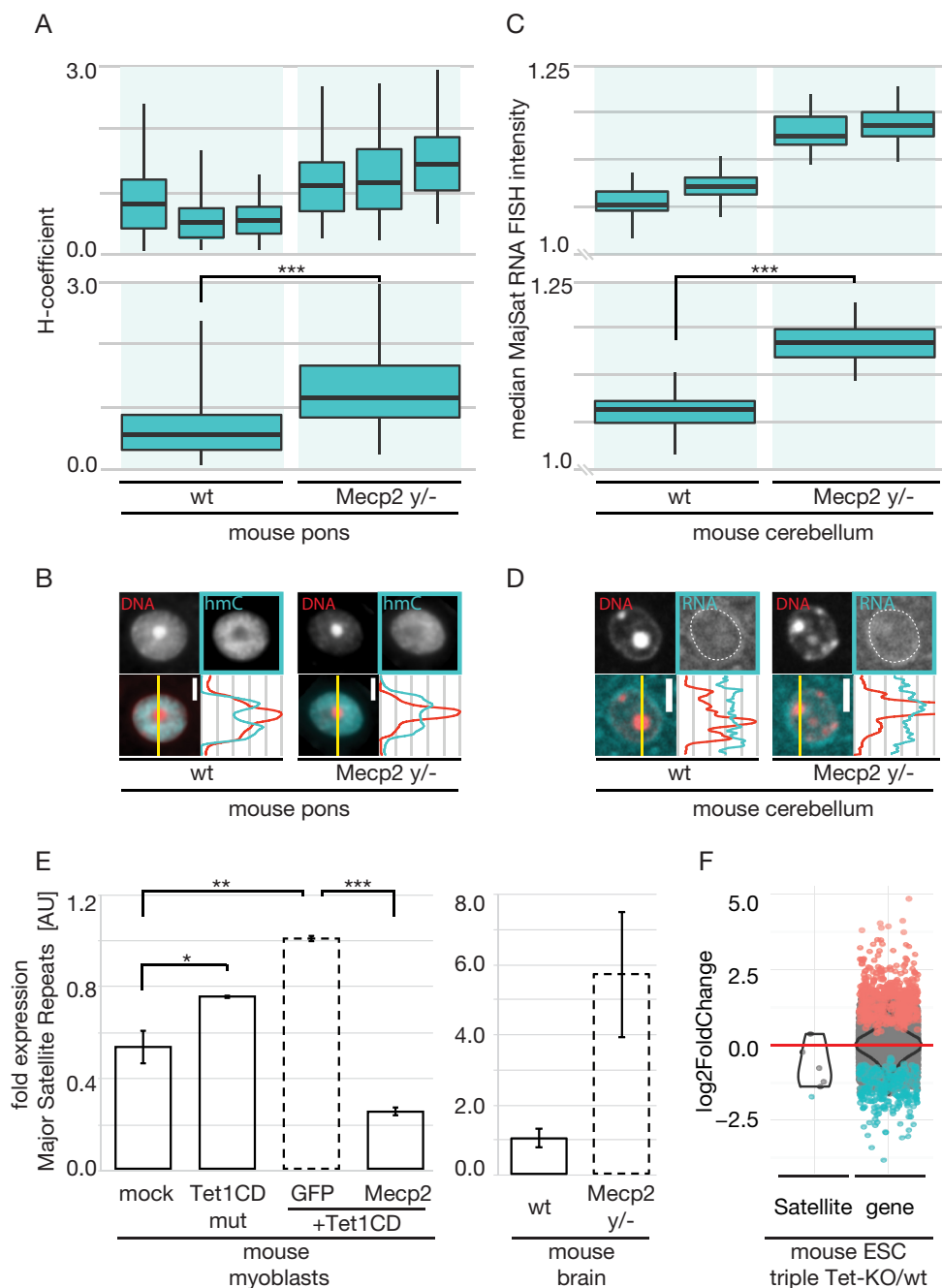


Figure 6. Correlation of subnuclear 5hmC distribution and major satellite expression. **(A and B)** Immunostaining and colocalization analysis of 5hmC with pericentric heterochromatin. **(A)** Box plots show the median 5hmC colocalization with pericentric heterochromatin in neurons of three individual (top) and combined (bottom) wild type and *Mecp2* *y/-* mouse ponses (wt1 *n* = 42, wt2 *n* = 39, wt3 *n* = 44, KO1 *n* = 34, KO2 *n* = 29, KO3 *n* = 20; from left to right; *n* = number of cells), as well as the first and third quartiles. Whiskers extend to 1.5 times the interquartile range. *P* values were calculated by Wilcoxon signed-rank test ($***P < 0.001$). **(B)** Line intensity plots of DNA (red) and 5hmC (cyan) distribution through pericentric heterochromatin in neurons of wild type and *Mecp2* *y/-* mouse ponses, respectively. DNA was counterstained with DAPI. Scale bar, 5 μ m. **(C and D)** Detection and quantification of major satellite transcripts by Fluorescent In Situ Hybridization (FISH). **(C)** Box plots show the median major satellite RNA FISH signal in neurons of two individual (top) and combined (bottom) wild type and *Mecp2* *y/-* mouse cerebella (wt1 *n* = 29, wt2 *n* = 26, KO1 *n* = 30, KO2 *n* = 30; from left to right; *n* = number of cells), as well as the first and third quartiles. Whiskers extend to 1.5 times the interquartile range. *P* values were calculated by Wilcoxon signed-rank test ($***P < 0.001$). **(D)** Line intensity plots of DNA (red) and major satellite RNA (cyan) distribution through pericentric heterochromatin in neurons of wild type and *Mecp2* *y/-* mouse ponses, respectively. DNA was counterstained with DAPI. Scale bar, 5 μ m. **(E)** RT-qPCR analysis of major satellite RNA transcript levels in transiently transfected C2C12 mouse myoblasts (left) and brain of wild type and *Mecp2* *y/-* mice (right), respectively.

likely not detectable as our results indicate. Importantly, it should be noted that the increase in 5hmC was not due to enhanced Tet1 expression as wild type and Mecp2 deficient neurons had comparable Tet1 levels (Supplementary Figure S14D and E).

In conclusion, our data demonstrate on the subcellular level that knockout of Mecp2 results in increased neuronal 5hmC levels. We cannot exclude that other secondary effects may contribute to the increase of 5hmC at chromocenters *in vivo*. However, expression of Mecp2 in cells that do not normally express endogenous Mecp2 (mouse myoblasts, Figure 1B), decreases Tet-mediated oxidation of 5mC. The only difference between both sets of cells is the presence or not of Mecp2. Hence, according to our cell data (Figure 1A and B), unconfined access of Tet1 proteins to pericentric heterochromatin, which is occupied when Mecp2 is present, is very likely the dominant mechanism that allows 5hmC accumulation at pericentric heterochromatin.

Previous data showed that both, Mecp2 protein and 5hmC levels are high in neurons. To address this apparent contradictory coexistence, we furthermore analyzed the expression of Tet1 in different cell types and found high levels of the 5mC oxygenase Tet1 in NeuN positive compared to surrounding glial cells (Supplementary Figure S14A–C). Furthermore, we found the degree of colocalization between pericentric heterochromatin and 5hmC considerably increased as a consequence of Mecp2 depletion. While in wild-type brain, 5hmC is anti-correlated with DNA dense chromocenters, this is not the case for Mecp2 deficient neurons (Figure 6A, top). Similar results were obtained from line intensity plots of 5hmC distribution through pericentric heterochromatin (Figure 6B), as well as accumulation studies of 5hmC at chromocenters (Supplementary Figure S16D). Accordingly, 5hmC is indeed abundant in neurons of wild-type mice, however, only at sites of low Mecp2 accumulation. Therefore, Mecp2 has a local protective effect at pericentric heterochromatin.

In the absence of Mecp2, Tet1 reactivates expression of major satellite repeats

Next, we tested whether hypomethylation of chromocenters (Figures 5 and 6A, B), which were previously described to be rich in major satellite repeats (72), leads to reactivation of these epigenetically silenced elements. Hence, we labeled and subsequently quantified (Supplementary Figure S16A and B) their RNA transcript levels by fluorescence in situ hybridization (FISH) in single cells of Mecp2 knockout mouse cerebella (Figure 6C). Compared to wild type, mean major satellite RNA FISH signals were significantly increased in nuclei of Mecp2 deficient cells (Figure 6C). Moreover, line intensity profiles of RNA FISH levels across chromocenters of the same nuclei, showed accumulation of major satellite transcripts directly at and in

close proximity to pericentric heterochromatin (Figure 6D). To ensure that the observed transcriptional increase of major satellite DNA is not limited to the analyzed brain region, we additionally confirmed its elevated expression levels in whole Mecp2 *y/-* mouse brain by reverse transcription quantitative polymerase chain reaction (RT-qPCR) (Figure 6E, right). Moreover, we made use of C2C12 mouse myoblasts, which show no detectable levels of Mecp2 and Mbd2 (Supplementary Figure S16C, top; (19)) and, thus, allowed us to directly test the effect of Tet proteins on the expression of DNA repeats. Hence, mouse myoblasts ectopically expressing the catalytic active domain of Tet1 were sorted by flow cytometry and the transcriptional levels of major satellite repeats were quantified by RT-qPCR. When compared to mock treated cells, major satellite RNA transcripts were increased in mouse myoblasts, congenitally lacking Mecp2 (Supplementary Figure S16C, top) and ectopically expressing Tet1CD (Figure 6E, left). Coexpression of Mecp2, however, abolished Tet1CD-mediated reactivation of major satellite repeats and reduced major satellite transcription by half when compared to mock treated cells (Figure 6E, left). While transcription level of major satellite repeats almost doubled upon ectopic expression of the catalytically active Tet1 domain, overexpression of the inactive variant resulted in an increase of only 40% (Figure 6E, left). Accordingly, we conclude that the induction of major satellite expression requires at least in part the catalytic activity of Tet1 and, thus, results from increased 5hmC levels. As overexpression of both, the catalytically active and inactive domain of Tet1 leads to decondensation of pericentric heterochromatin (Zhang *et al.*, submitted), we furthermore deduce that the 40% increase of major satellite expression in cells expressing mutant Tet1CD, might be partially caused by reorganization of chromatin to a more open and, thus, accessible state.

Finally, we analyzed expression of satellite elements in triple Tet-knockout (KO) and wild type (wt) mouse embryonic stem cells (ESC) by RNA-seq. The median log₂-fold change was -0.99, indicating that the expression of most genomic satellite sequences is down-regulated upon Tet1/2/3 depletion (Figure 6F). Taken together, our data demonstrate that in the absence of Mecp2, Tet1 reactivates the expression of epigenetically silenced (major) satellite repeats, which in turn might compromise genome stability (73,74). Therefore, we suggest that unrestricted Tet activity may be part of a pathogenic cascade in Rett syndrome, which is initiated by Mecp2 gene mutations that reduce or abolish DNA binding.

In the present study, we demonstrate that prior binding of methyl-CpG binding domain proteins Mecp2 and Mbd2 to DNA protects 5mC from Tet1CD mediated oxidation in a concentration dependent manner, thereby regulating chromatin composition (Figure 7A).

← Expression levels are relative to Tet1CD transfected cells (left), or wild type mouse brain (right). Shown are average values from \geq two biological replicates each measured from one (left), or two (right) independent cDNA synthesis reactions, respectively. Error bars represent \pm SD. *P* values were calculated by an independent two-sample student's *t*-test (**P* < 0.05, ***P* = 0.01, ****P* < 0.001). (F) Violin-plot of the log₂-fold changes of the triple Tet-knockout (KO) relative to wild type (wt) mouse embryonic stem cells (ESC (v6.5)) for all genes and all satellites. Negative values indicate a down-regulation in the knockout cells relative to the wild type, positive values an up-regulation. Significant elements are marked in color. The red line is at zero, i.e. the expected value if expression were identical in the wild type and knockout. Triple Tet-knockout: *P* = 4.84×10^{-2} ; genes: *P* = 5×10^{-15} .

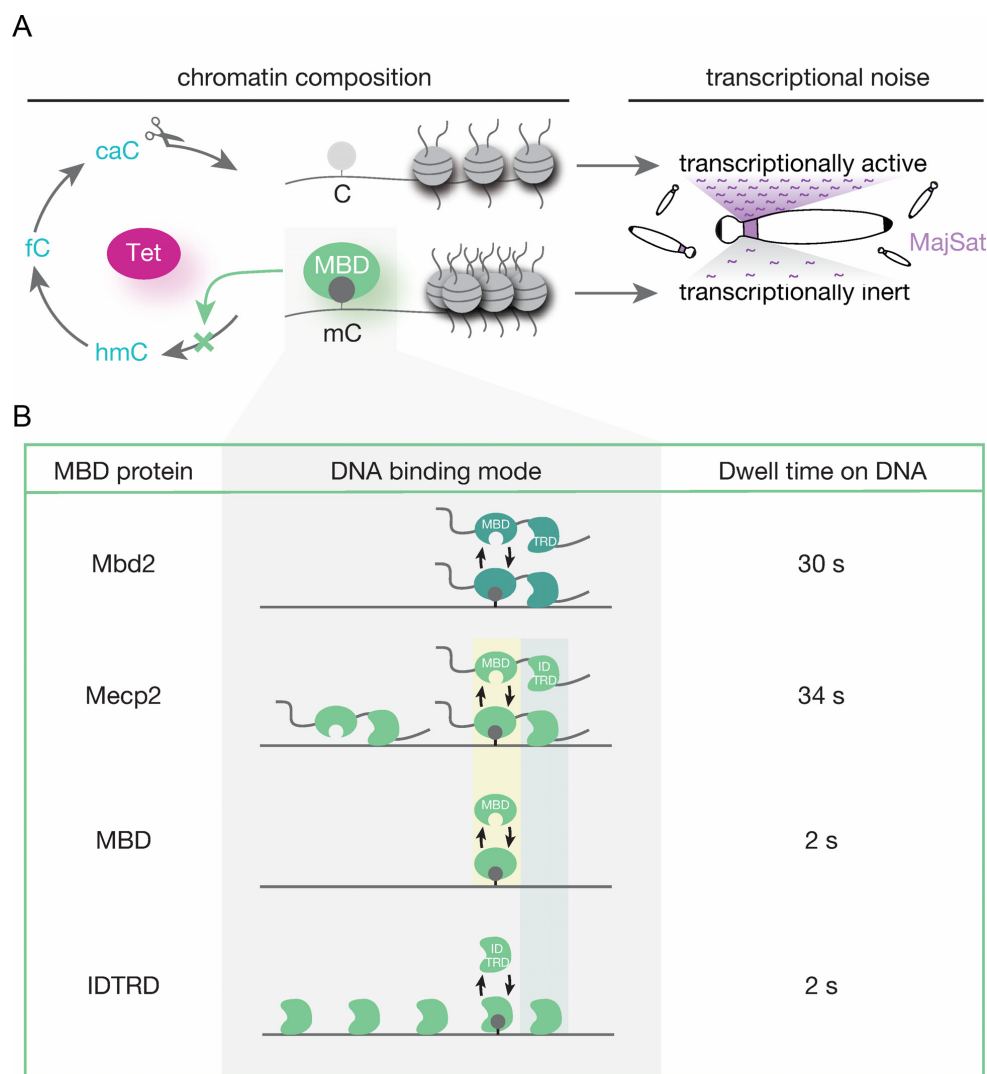


Figure 7. Mecp2 and Mbd2 preserve chromatin composition and thus genomic integrity by insulating 5mC from Tet1 activity. Scheme summarizing the main conclusions drawn from our studies. (A) We show that Mecp2 and Mbd2 protect 5mC from Tet1-mediated oxidation in a concentration dependent manner *in vivo* and *in vitro*. (B) The protection mechanism is not based on competition for 5mC *per se* but rather on sequence unspecific coverage of DNA and correlates with the respective MBD protein dwell time on DNA. (A) As a biological consequence, we measured increased 5hmC level in neurons of a mouse model for Rett syndrome with concomitant reactivation of epigenetically silenced pericentric DNA repeats.

The underlying molecular mechanism relies on competitive, sequence unspecific coverage of DNA and is affected by the respective MBD protein dwell time on DNA (Figure 7B). Accordingly, Tet binding to its substrate and, consequently, 5mC modification are inhibited and chromatin composition maintained. Hence, we infer that Tet1 activity is likely to vary according to tissue and cell specific distribution of methylated CpG sites, as it influences the binding affinity of MBD proteins (15). Furthermore, we propose that the quantity of methyl-CpG binding domain molecules per cell must be precisely regulated to accurately control Tet1 activity. Indeed, either duplication of the *MECP2* gene with increased respective protein level or mutant *MECP2*

proteins with impaired DNA binding, are both observed in Rett patients (61,75).

As a biological consequence, we measured increased 5hmC at pericentric heterochromatin in neurons of Mecp2 deficient mice with concomitant reactivation of epigenetically silenced major satellite repeats (Figure 7A). Compensatory effects by Mbd2 cannot come into play as its expression levels are significantly reduced in Mecp2 knockout brain (49). As Tet1 reactivates transcription of major satellite repeats in the absence of Mecp2 and Mbd2 proteins, we conclude that the transcriptional noise increase in Rett animal models (20) is likely to result, at least in part, from unconfined Tet activity and, thus, provide a potential Tet-induced pathophysiological pathway in Rett syndrome.

Since almost all mature, postmitotic neurons were shown to express abundant levels of methyltransferases Dnmt1 and Dnmt3a (76,77), we propose that stabilization or reversion of Rett symptoms upon delivery of MeCP2 (78) results from re-methylation and subsequent binding and protection of 5mC by the exogenous wild type MeCP2 protein.

In summary, these data provide mechanistic insights into the regulation of Tet1 activity by methyl-CpG binding domain proteins and argue for a role of the MBD proteins as guardians of the epigenome.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are indebted to Anna M. Bischoff (University of Goettingen, Germany) for providing MeCP2 y/- brain, Congdi Song (LMU Munich, Germany) for providing cDNA from MeCP2 y/- brain, Aleksandra Szwagierczak (LMU Munich, Germany) for providing plasmids, Adrian Bird (Wellcome Trust Centre for Cell Biology, Edinburgh, UK) and Rudolf Jaenisch (Whitehead Institute for Biomedical Research, Cambridge, USA) for providing cell lines, Bodo Laube (Technische Universität Darmstadt) for help with the primary neuron cultures, Corella S. Casas-Delucchi, Annette Becker, Alexander Rapp, Bianca Bertulat and Manuela Milden for experimental advice and discussions.

Author contributions: A.K.L., P.Z., S.M., U.M., F.D.H., C.R., C.T. and A.L. performed experiments. A.K.L., P.Z., H.D.H., U.M. and I.H. analyzed the data. C.S. performed statistical tests for Figures 1 and 3. H.L. and M.C.C. conceived the project, provided expertise and feedback. A.K.L., P.Z. and M.C.C. conceived and designed the experiments and wrote the manuscript.

FUNDING

China Scholarship Council (to P.Z; Deutsche Forschungsgemeinschaft [DFG] CA 198/7 to M.C.C. and SFB1064/A17 to H.L.). Funding for open access charge: DFG.

Conflict of interest statement. None declared.

REFERENCES

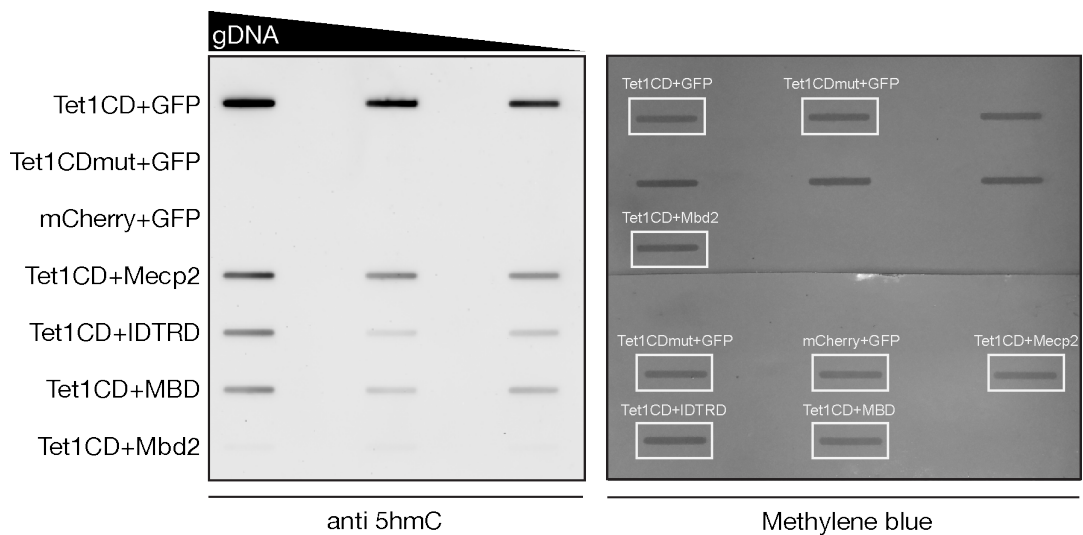
- Bird, A.P. and Wolffe, A.P. (1999) Methylation-induced repression—belts, braces, and chromatin. *Cell*, **99**, 451–454.
- Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A. and Gehrke, C. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.*, **10**, 2709–2721.
- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
- Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.*, **8**, 286–298.
- Peng, J.C. and Karpen, G.H. (2007) H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat. Cell Biol.*, **9**, 25–35.
- Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33**(Suppl), 245–254.
- Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell Biol.*, **18**, 6538–6547.
- Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A. and Esteller, M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.*, **31**, 1765–1774.
- Nan, X., Campoy, F.J. and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471–481.
- Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, **19**, 187–191.
- Ebert, D.H., Gabel, H.W., Robinson, N.D., Kastan, N.R., Hu, L.S., Cohen, S., Navarro, A.J., Lyst, M.J., Ekiert, R., Bird, A.P. *et al.* (2013) Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature*, **499**, 341–345.
- Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N.R., Robinson, N.D., de Lima Alves, F. *et al.* (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.*, **16**, 898–902.
- Ghosh, R.P., Nikitina, T., Horowitz-Scherer, R.A., Gierach, L.M., Uversky, V.N., Hite, K., Hansen, J.C. and Woodcock, C.L. (2010) Unique physical properties and interactions of the domain of methylated DNA binding protein 2. *Biochemistry-US*, **49**, 4395–4410.
- Hansen, J.C., Ghosh, R.P. and Woodcock, C.L. (2010) Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. *Iubmb Life*, **62**, 732–738.
- Kumar, A., Kamboj, S., Malone, B.M., Kudo, S., Twiss, J.L., Czymmek, K.J., LaSalle, J.M. and Schanen, N.C. (2008) Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J. Cell Sci.*, **121**, 1128–1137.
- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, **69**, 905–914.
- Becker, A., Allmann, L., Hofstatter, M., Casa, V., Weber, P., Lehmkuhl, A., Herce, H.D. and Cardoso, M.C. (2013) Direct homo- and hetero-interactions of MeCP2 and MBD2. *PLoS One*, **8**, e53730.
- Brero, A., Easwaran, H.P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H. and Cardoso, M.C. (2005) Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.*, **169**, 733–743.
- Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R. and Bird, A.P. (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell*, **37**, 457–468.
- Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T. and Downing, J.R. (2003) TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, **17**, 637–641.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. and Zhang, Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L. and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, **145**, 423–434.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F. *et al.* (2013)

- Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
28. Portela, A. and Esteller, M. (2010) Epigenetic modifications and human disease. *Nat. Biotechnol.*, **28**, 1057–1068.
 29. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.*, **23**, 185–188.
 30. Kriaucionis, S. and Bird, A. (2003) DNA methylation and Rett syndrome. *Hum. Mol. Genet.*, **12**(Spec No 2), R221–R227.
 31. Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K. and Gage, F.H. (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, **468**, 443–446.
 32. Jost, K.L., Rottach, A., Mildner, M., Bertulat, B., Becker, A., Wolf, P., Sandoval, J., Petazzi, P., Huertas, D., Esteller, M. *et al.* (2011) Generation and characterization of rat and mouse monoclonal antibodies specific for MeCP2 and their use in X-inactivation studies. *PLoS One*, **6**, e26499.
 33. Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Schanen, C. and Tamura, M. (2003) Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J. Med. Genet.*, **40**, 487–493.
 34. Qin, W., Leonhardt, H. and Spada, F. (2011) Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J. Cell. Biochem.*, **112**, 439–444.
 35. Muller, U., Bauer, C., Siegl, M., Rottach, A. and Leonhardt, H. (2014) TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.*, **42**, 8592–8604.
 36. Becker, A., Zhang, P., Allmann, L., Meilinger, D., Bertulat, B., Eck, D., Hofstaetter, M., Bartolomei, G., Hottiger, M.O., Schreiber, V. *et al.* (2016) Poly(ADP-ribosyl)ation of methyl CpG binding domain protein 2 regulates chromatin structure. *J. Biol. Chem.*, **291**, 4873–4881.
 37. Berger, I., Fitzgerald, D.J. and Richmond, T.J. (2004) Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.*, **22**, 1583–1587.
 38. Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725–727.
 39. Cardoso, M.C., Leonhardt, H. and Nadal-Ginard, B. (1993) Reversal of terminal differentiation and control of DNA replication: cyclin A and Cdk2 specifically localize at subnuclear sites of DNA replication. *Cell*, **74**, 979–992.
 40. Casas-Delucchi, C.S., Becker, A., Bolius, J.J. and Cardoso, M.C. (2012) Targeted manipulation of heterochromatin rescues MeCP2 Rett mutants and re-establishes higher order chromatin organization. *Nucleic Acids Res.*, **40**, e176.
 41. Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H. and Cardoso, M.C. (2007) MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res.*, **35**, 5402–5408.
 42. Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., Lyko, F. *et al.* (2014) Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell*, **29**, 102–111.
 43. Guy, J., Hendrich, B., Holmes, M., Martin, J.E. and Bird, A. (2001) A mouse MeCP2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.*, **27**, 322–326.
 44. Pichler, G., Leonhardt, H. and Rothbauer, U. (2012) Fluorescent protein specific Nanotraps to study protein-protein interactions and histone-tail peptide binding. *Methods Mol. Biol.*, **911**, 475–483.
 45. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C. and Leonhardt, H. (2008) A versatile nanotrapp for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell. Proteomics: MCP*, **7**, 282–289.
 46. Mortusewicz, O., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2006) Differential recruitment of DNA Ligase I and III to DNA repair sites. *Nucleic Acids Res.*, **34**, 3523–3532.
 47. Rottach, A., Kremmer, E., Nowak, D., Leonhardt, H. and Cardoso, M.C. (2008) Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma (Larchmt)*, **27**, 337–343.
 48. Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.*, **16**, 1215.
 49. Song, C., Feodorova, Y., Guy, J., Peichl, L., Jost, K.L., Kimura, H., Cardoso, M.C., Bird, A., Leonhardt, H., Joffe, B. *et al.* (2014) DNA methylation reader MECP2: cell type- and differentiation stage-specific protein distribution. *Epigenet. Chromatin*, **7**, 17.
 50. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. and Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.
 51. Bauer, C., Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Muller, U., Kremmer, E., Rottach, A. and Leonhardt, H. (2015) Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J. Biol. Chem.*, **290**, 4801–4812.
 52. Herce, H.D., Casas-Delucchi, C.S. and Cardoso, M.C. (2013) New image colocalization coefficient for fluorescence microscopy to quantify (bio-)molecular interactions. *J. Microsc.*, **249**, 184–194.
 53. Casas-Delucchi, C.S., van Bommel, J.G., Haase, S., Herce, H.D., Nowak, D., Meilinger, D., Stear, J.H., Leonhardt, H. and Cardoso, M.C. (2012) Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. *Nucleic Acids Res.*, **40**, 159–169.
 54. Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G. and Sandberg, R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
 55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
 56. Jin, Y., Tam, O.H., Paniagua, E. and Hammell, M. (2015) TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, **31**, 3593–3599.
 57. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
 58. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate — a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.*, **57**, 289–300.
 59. Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R. and Bird, A.P. (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell*, **37**, 457–468.
 60. Khrapunov, S., Warren, C., Cheng, H., Berko, E.R., Grealley, J.M. and Brenowitz, M. (2014) Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity. *Biochemistry-US*, **53**, 3379–3391.
 61. Agarwal, N., Becker, A., Jost, K.L., Haase, S., Thakur, B.K., Brero, A., Hardt, T., Kudo, S., Leonhardt, H. and Cardoso, M.C. (2011) MeCP2 Rett mutations affect large scale chromatin organization. *Human Mol. Genet.*, **20**, 4187–4195.
 62. Lindhout, B.I., Fransz, P., Tessadori, F., Meckel, T., Hooykaas, P.J. and van der Zaal, B.J. (2007) Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic Acids Res.*, **35**, e107.
 63. Thanisch, K., Schneider, K., Morbitzer, R., Solovei, I., Lahaye, T., Bultmann, S. and Leonhardt, H. (2014) Targeting and tracing of specific DNA sequences with dTALES in living cells. *Nucleic Acids Res.*, **42**, e38.
 64. Menafra, R., Brinkman, A.B., Matarese, F., Franci, G., Bartels, S.J., Nguyen, L., Shimbo, T., Wade, P.A., Hubner, N.C. and Stunnenberg, H.G. (2014) Genome-wide binding of MBD2 reveals strong preference for highly methylated loci. *PLoS One*, **9**, e99603.
 65. Desai, M.A., Webb, H.D., Sinanan, L.M., Scarsdale, J.N., Walavalkar, N.M., Ginder, G.D. and Williams, D.C. Jr (2015) An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res.*, **43**, 3100–3113.
 66. Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I. *et al.* (2011) 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.*, **14**, 1607–1616.

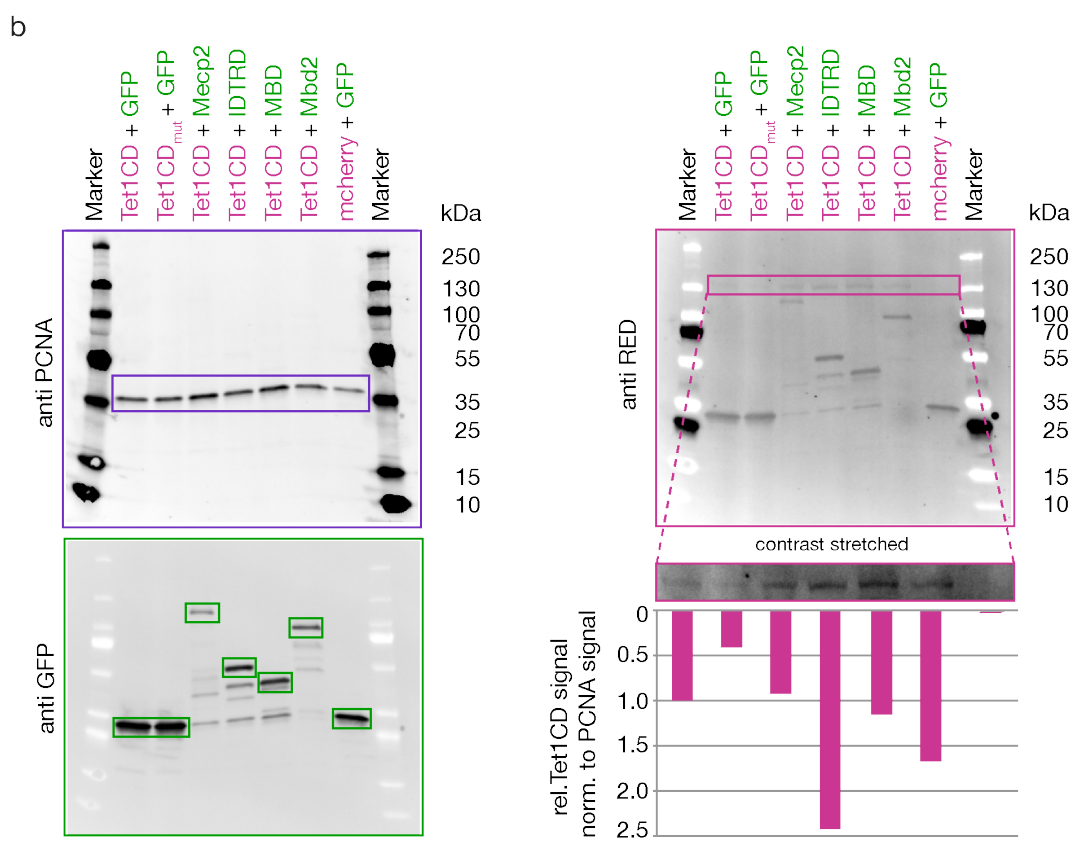
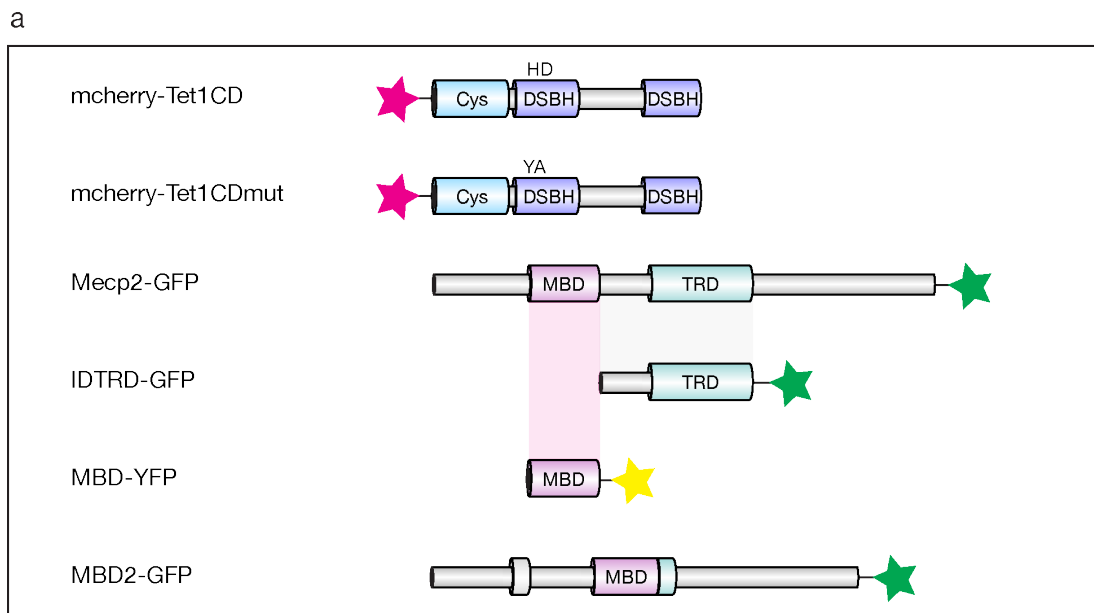
20 *Nucleic Acids Research*, 2016

67. Hashimoto, H., Pais, J.E., Zhang, X., Saleh, L., Fu, Z.Q., Dai, N., Correa, I.R. Jr, Zheng, Y. and Cheng, X. (2014) Structure of a *Naegleria* Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature*, **506**, 391–395.
68. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P. and Xu, Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, **155**, 1545–1555.
69. Ramirez, J.M., Ward, C.S. and Neul, J.L. (2013) Breathing challenges in Rett syndrome: lessons learned from humans and animal models. *Respir. Physiol. Neurobiol.*, **189**, 280–287.
70. Munzel, M., Globisch, D., Bruckl, T., Wagner, M., Welzmler, V., Michalak, S., Muller, M., Biel, M. and Carell, T. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem.*, **49**, 5375–5377.
71. Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E. and Zhang, Y. (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.*, **25**, 679–684.
72. Jones, K.W. (1970) Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature*, **225**, 912–915.
73. Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M. and Gartler, S.M. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14412–14417.
74. Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A. *et al.* (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, **107**, 323–337.
75. Ramocki, M.B., Peters, S.U., Tavyev, Y.J., Zhang, F., Carvalho, C.M., Schaaf, C.P., Richman, R., Fang, P., Glaze, D.G., Lupski, J.R. *et al.* (2009) Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome. *Ann. Neurol.*, **66**, 771–782.
76. Watanabe, D., Uchiyama, K. and Hanaoka, K. (2006) Transition of mouse de novo methyltransferases expression from Dnmt3b to Dnmt3a during neural progenitor cell development. *Neuroscience*, **142**, 727–737.
77. Goto, K., Numata, M., Komura, J.I., Ono, T., Bestor, T.H. and Kondo, H. (1994) Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. *Differentiation*, **56**, 39–44.
78. Garg, S.K., Li, D.T., Cheval, H., McGann, J.C., Bissonnette, J.M., Murtha, M.J., Foust, K.D., Kaspar, B.K., Bird, A. and Mandel, G. (2013) Systemic delivery of MeCP2 rescues behavioral and cellular deficits in female mouse models of Rett syndrome. *J. Neurosci.*, **33**, 13612–13620.

Supplementary Figures

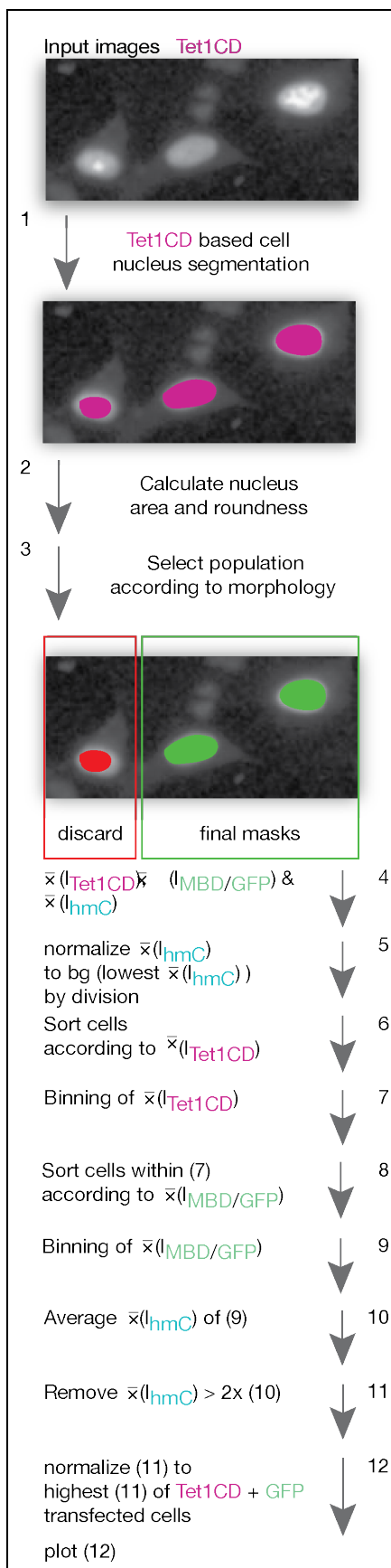


Supplementary Figure 1. Full slot blots used for the quantification of 5hmC in genomic DNA (gDNA) of transiently transfected HEK cells (Figure 1a, bottom). (left) anti 5hmC. (right) Methylene blue staining.



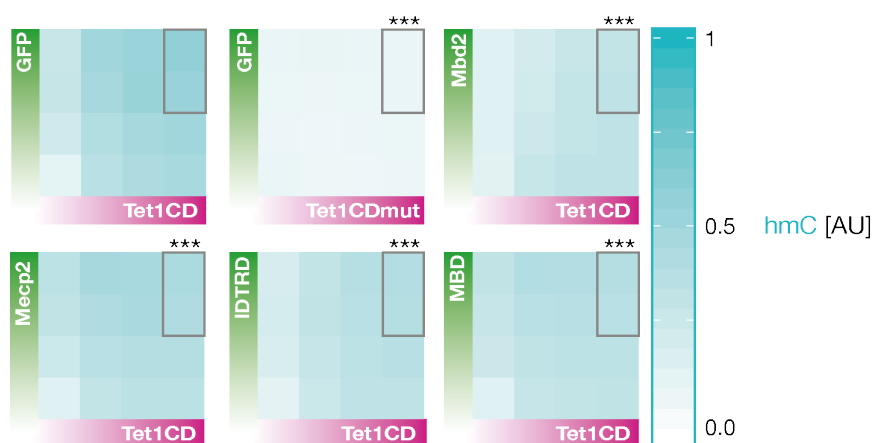
Supplementary Figure 2. Western blot analysis of Tet1CD and MBD protein levels of transiently transfected, FACS sorted HEK cells

(a) Representative schemes of proteins that were used for *in vitro* and *in vivo* studies. Tet1CD corresponds to the Tet1 catalytic domain (aa 1365-2057) and Tet1CDmut is the catalytically inactive domain of Tet1 (aa 1365-2057) containing two point mutations (H1652Y, D1654A) that abolish binding of the co-factor Fe^{2+} . Magenta colored star=mCherry; green colored star=GFP; yellow colored star=YFP. Cys=cysteine rich region; DSBH=double stranded beta helix; MBD=methyl CpG binding domain; TRD=transcriptional repression domain. (b) Shown are GFP or YFP-tagged Mecp2, IDTRD, MBD and Mbd2 (green, anti GFP), as well as, mCherry-tagged Tet1CD and Tet1CDmut (magenta, anti RED) proteins of transiently transfected HEK cells after enrichment for double-transfected cells by Fluorescent Activated Cell Sorting (FACS). Proliferating cell nuclear antigen (PCNA) was used as a loading control. Tet1CD expression levels are similar in cells coexpressing MBD proteins and comparatively low in cells coexpressing GFP. Accordingly, decreased genomic 5hmC levels observed in cells coexpressing MBD proteins (**Fig. 1 and Fig. 3**) are not due to lower Tet1CD expression levels.



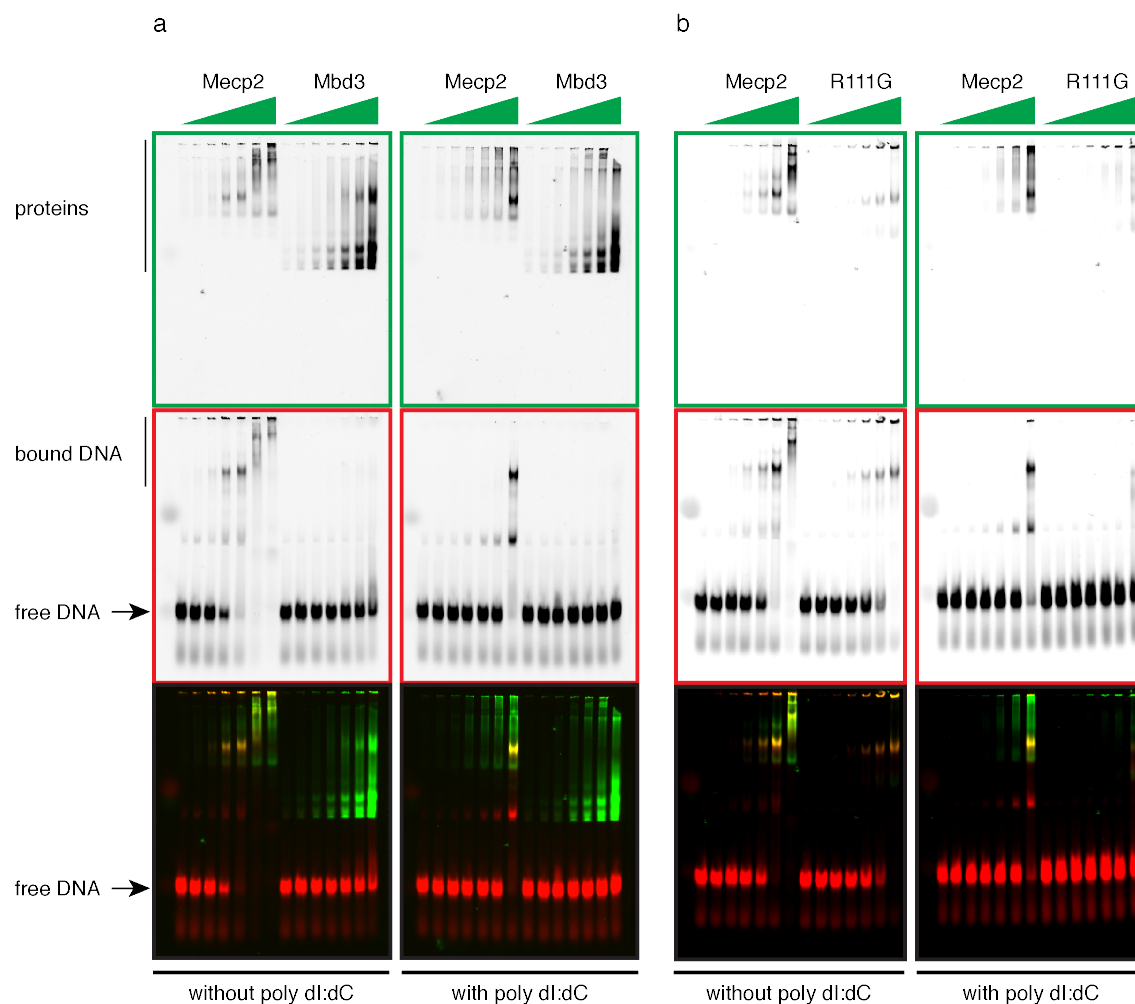
Supplementary Figure 3. Summary of the different steps used for quantification and normalization of genomic 5hmC levels in transiently transfected C2C12 mouse myoblasts

(1) Cell nuclei segmentation according to Tet1 signals. (2) Nuclei area and roundness calculation. (3) Selection of cell populations based on morphology properties. (4) Calculation of mean Tet1CD-, MBD- and 5hmC intensities. (5) Normalization of mean hmC signals to background. (6) Sort cells according to Tet1 signals. (7) Bin cells based on mean Tet1 intensities. (8) Sort cells within subgroups (7) according to mean MBD intensities. (9) Bin cells within subgroups (7) based on MBD intensities. (10) Average mean 5hmC intensities of each sub-subgroup (9). (11) Remove outliers. (12) Normalize mean 5hmC signals of each sub-subgroup (10) to highest 5hmC level of Tet1CD + GFP transfected cells and plot in form of a heat map as a function of Tet1CD- and MBD protein expression levels.

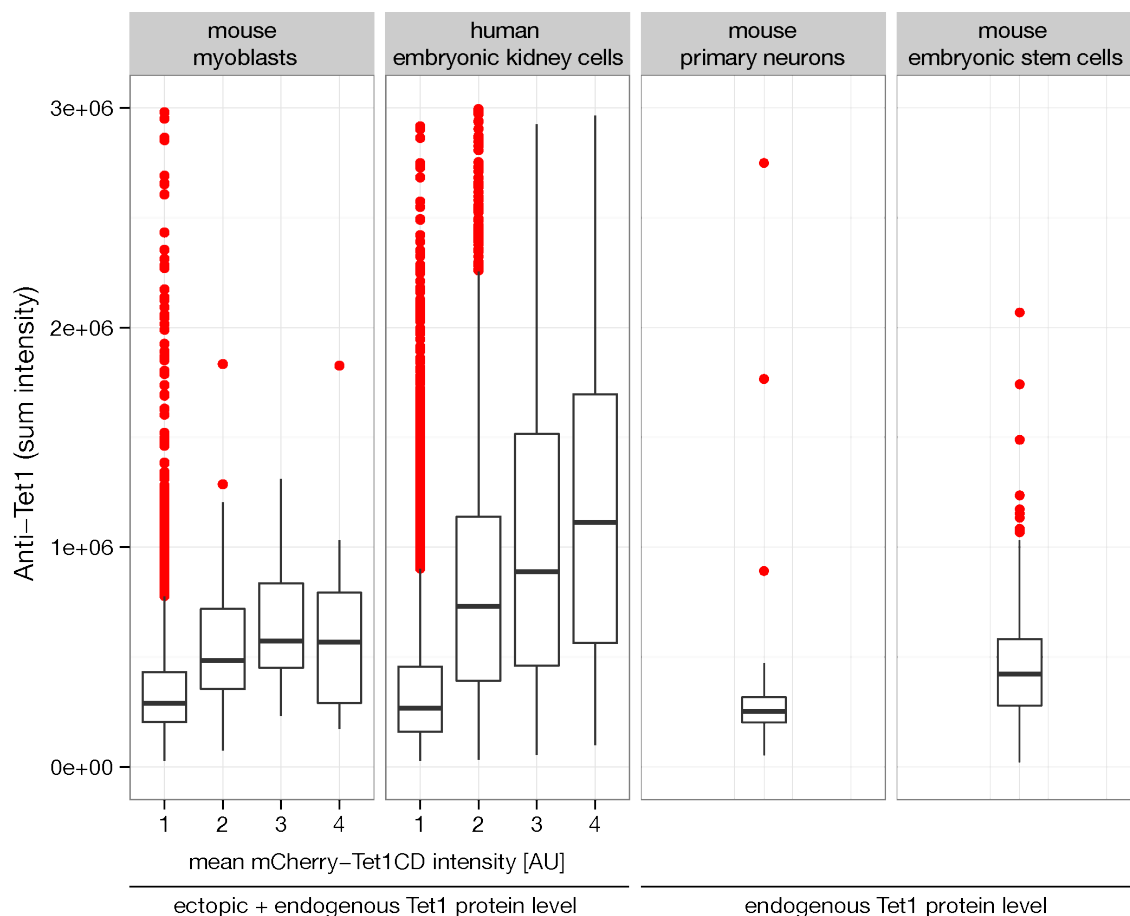


Supplementary Figure 4. Effect of methyl-CpG binding domain proteins on Tet1CD activity in human embryonic kidney cells

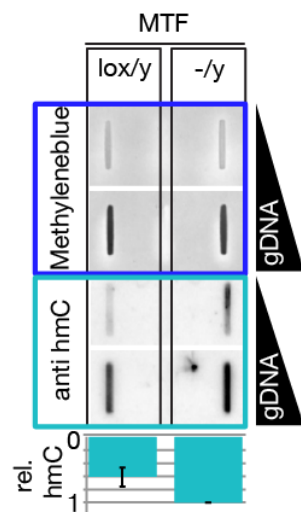
In situ staining and quantification of genomic 5hmC levels in transiently transfected human embryonic kidney (HEK) cells. Images were acquired on an automated high throughput imaging system with a 20x, 0.45 NA objective. Gradient heat maps show relative 5hmC (=cyan) signals as a function of increasing Tet1CD (=magenta) and MBD (=green) protein expression levels depicted by the green and magenta gradient bars. Shown are mean values of three (Tet1CD+GFP, n=38840; Tet1CDmut+GFP, n=53761; Tet1CD+Mbd2, n=39572; Tet1CD+Mecp2, n=41568; Tet1CD+IDTRD, n=30919; Tet1CD+MBD, n=32957) independent experiments, respectively. For statistical tests, 5hmC signals of cells with high Tet1CD and high Mecp2/Mbd2/Mbd3 protein levels (framed in grey) were used. All samples differed highly significant (***, $p < 0.001$; post-hoc pairwise t test) from Tet1CD+GFP.



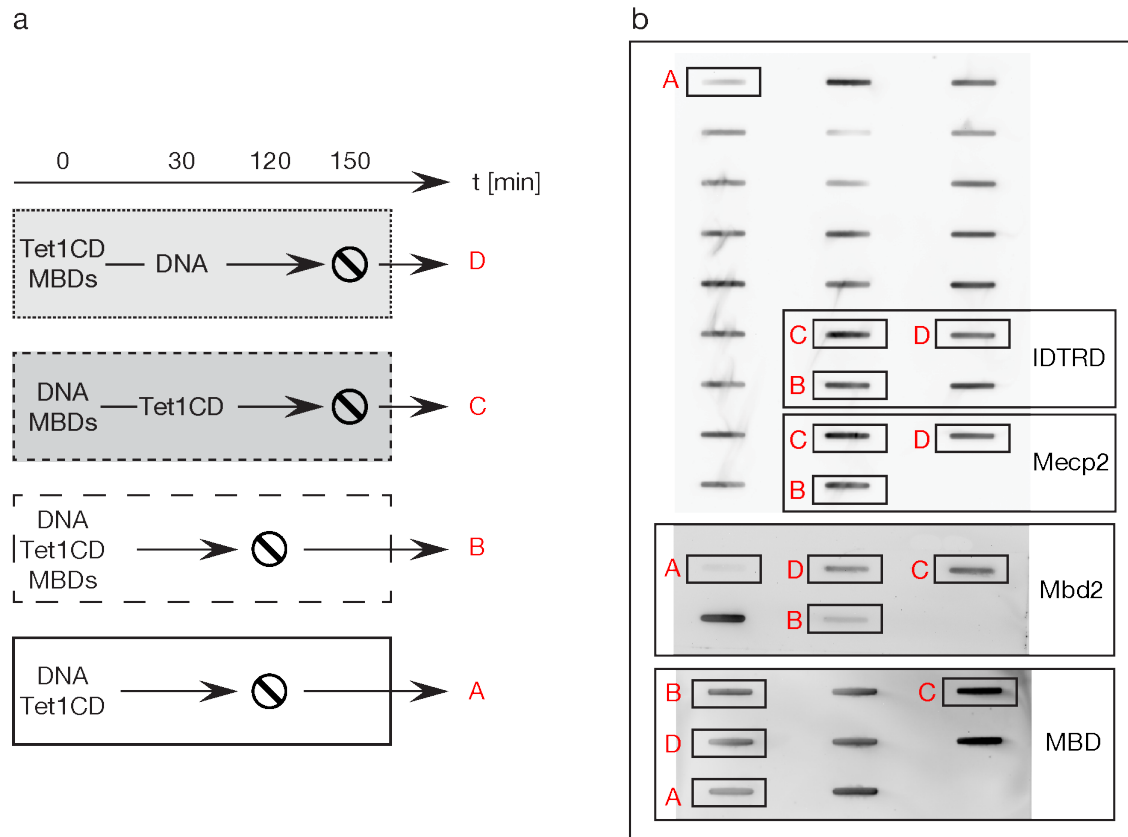
Supplementary Figure 5. Binding properties of MeCP2, Mbd3 and MeCP2_{R111G} to methylated DNA. (a) Electrophoretic mobility shift assay (EMSA) of methylated DNA (red) with MeCP2 and Mbd3 (green) in the absence (left) and presence (right) of poly di:dC. (b) Electrophoretic mobility shift assay (EMSA) of methylated DNA (red) with MeCP2 and MeCP2_{R111G} (green) in the absence (left) and presence (right) of poly di:dC.



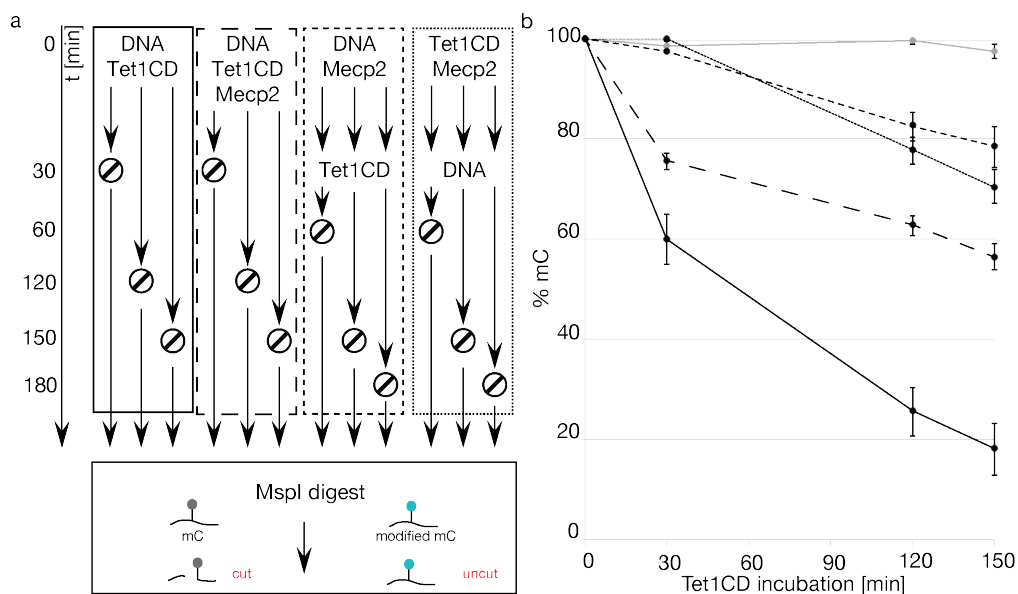
Supplementary Figure 6. Estimation of the level of mCherry-Tet1CD in transfected mouse myoblasts and human embryonic kidney cells. *In situ* staining and quantification of Tet1 in mCherry-Tet1CD transfected mouse myoblasts ($n=2704$) and human embryonic kidney cells ($n=11025$), as well as untransfected primary mouse neurons ($n=102$) and mouse embryonic stem cells ($n=373$). Images were acquired on an automated high throughput imaging system with a 20x, 0.45 NA objective. Mouse myoblast and human embryonic kidney cells were binned (as in Figure 1b and 3b) according to the ectopic mCherry-Tet1CD expression levels. Box plots represent the total nuclear Tet1 signal (endogenous + ectopic), as well as the first and third quartiles. Whiskers extend to 1.5 times the interquartile range.



Supplementary Figure 7. Genomic 5hmC level in the presence and absence of Mecp2 in mouse tail fibroblast (MTF) cells. Immunological assay to determine 5hmC levels in genomic DNA (gDNA) of wild type (lox/y) and Mecp2 knockout (-/y) MTF cells. gDNA quantities were monitored by methylene blue staining.

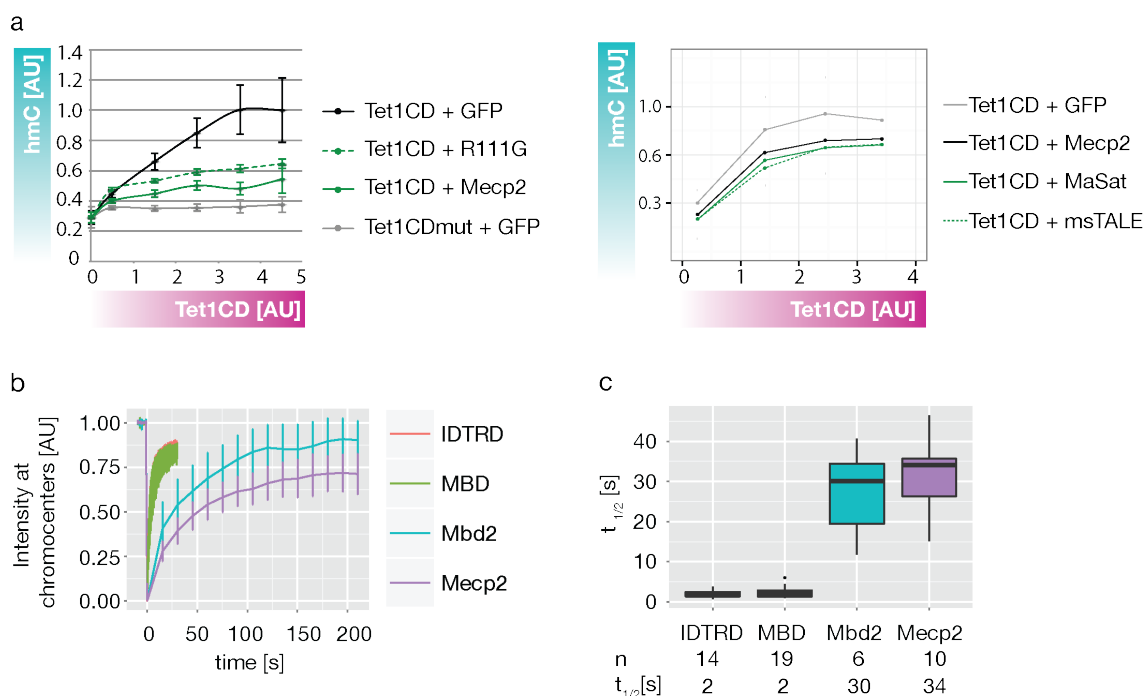


Supplementary Figure 8. Full slot blots used for the quantification of remaining 5mC in double-stranded DNA after simultaneous or successive incubation with Tet1CD and MBP proteins (see Fig. 2 and Fig. 3c). (a) Experimental setup illustrating the incubation order and time of proteins and methylated PCR product prior slot blotting. To be able to assign each of the four groups to the individual slots of the slot blots, they were labeled as A, B, C and D, respectively. **(b)** One representative full slot blot for IDTRD, Mecp2, Mbd2 and MBP. Unmarked slots are unrelated to the figure.



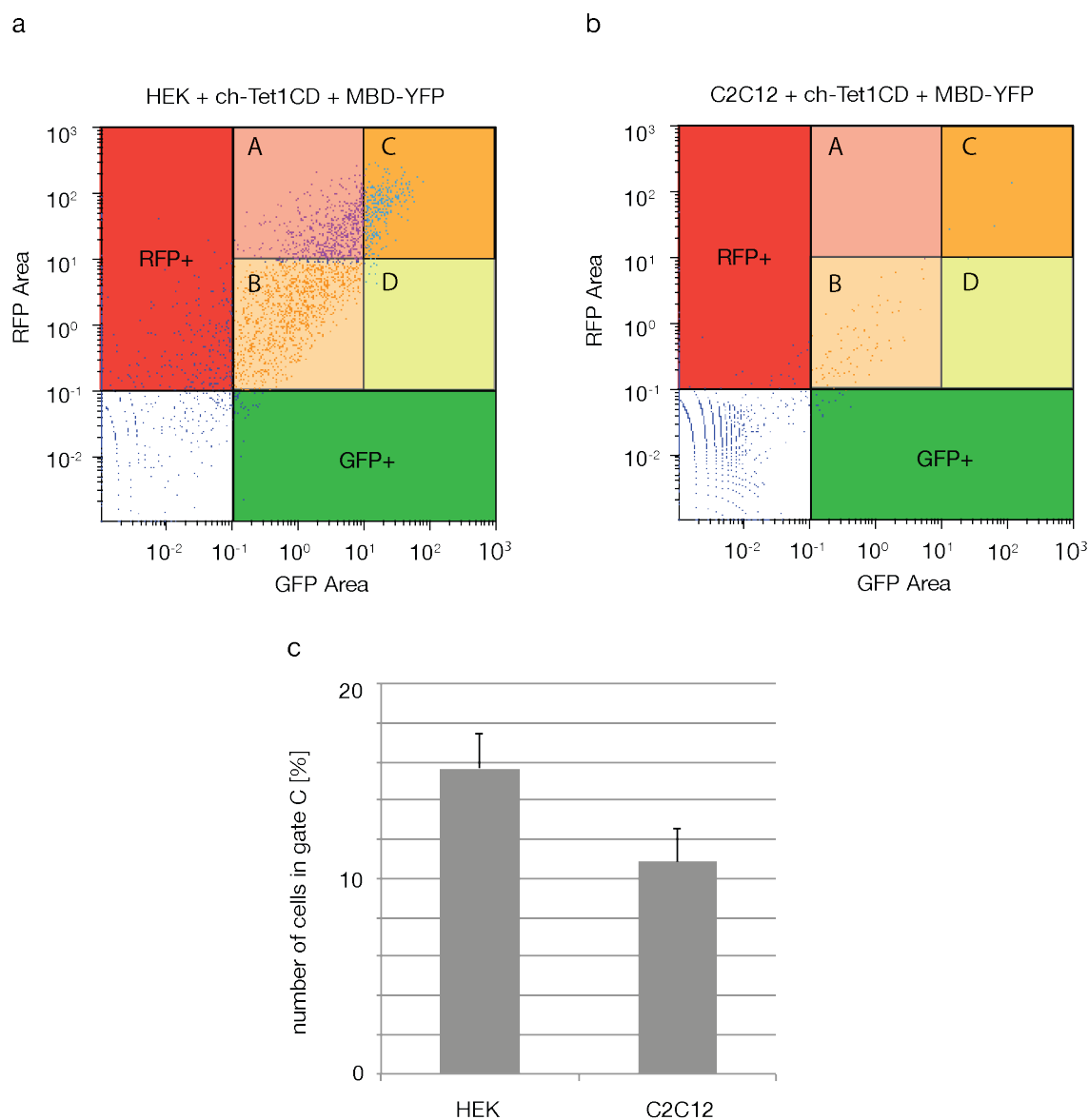
Supplementary Figure 9. Influence of the chronological DNA binding order on the protecting ability of MBD proteins over time

Quantification of remaining 5mC levels in single 5mC-containing oligos after simultaneous and successive incubation with Tet1CD- and MeCP2 proteins by MspI restriction digest. **(a)** Experimental setup illustrating the incubation order and time of proteins and oligos prior MspI restriction digest. When Tet1 oxidizes 5mC in the context of CCGG, the cleavable MspI site becomes uncleavable. **(b)** Diagram shows relative 5mC levels of single 5mC containing oligos after incubation with Tet1CD and MeCP2 (n=2). Incubation of Tet1CD with methylated oligos in the absence of its cofactor α -ketoglutarat was used as a negative control (grey line). Shown are mean values +/-SD.



Supplementary Figure 10. Contribution of direct and long-lasting 5mC binding to the protection of 5mC from Tet1CD catalyzed oxidation

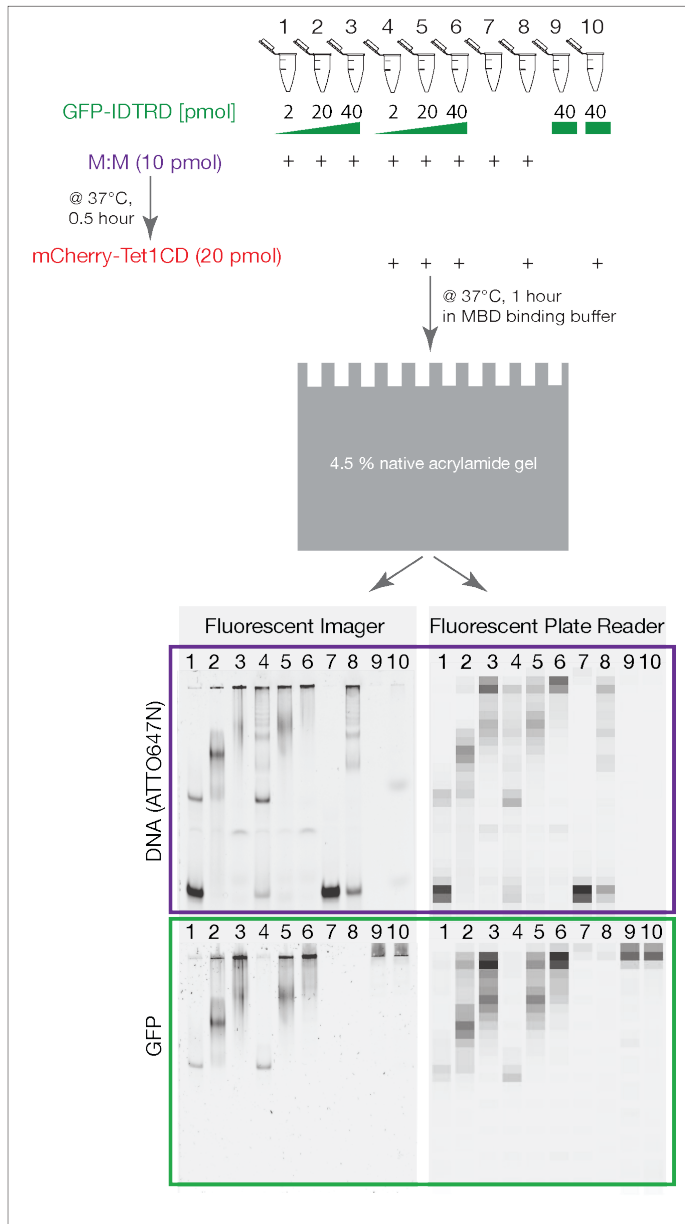
(a) *In situ* staining and quantification of genomic 5hmC levels in transiently transfected C2C12 mouse myoblasts. Shown are relative 5hmC (=cyan) signals as a function of Tet1CD (=magenta) level represented as gradient bars. (a, left) Error bars represent standard deviation. Mean values of two independent experiments are plotted (Tet1CD+GFP, n=2255; Tet1CD+Mecp2, n=2351; Tet1CD+R111G, n=3244; n=number of cells). (a, right) Mean values of three independent experiments are plotted (Tet1CD+GFP, n=17741; Tet1CD+Mecp2, n=27130; Tet1CD+MaSat, n=15149; Tet1CD+msTALE, n=18593) (b) Accumulation kinetics of Mbd2, Mecp2, MBD and IDTRD to pericentric heterochromatin in C2C12 mouse myoblasts. (c) Plateau level of accumulation curves. Box plots show the median accumulation of MBD proteins to pericentric heterochromatin (IDTRD, n=14; MBD, n=19; Mbd2, n=6; Mecp2, n=10; n=number of cells), as well as the first and third quartiles. Whiskers extend to 1.5 times the interquartile range.



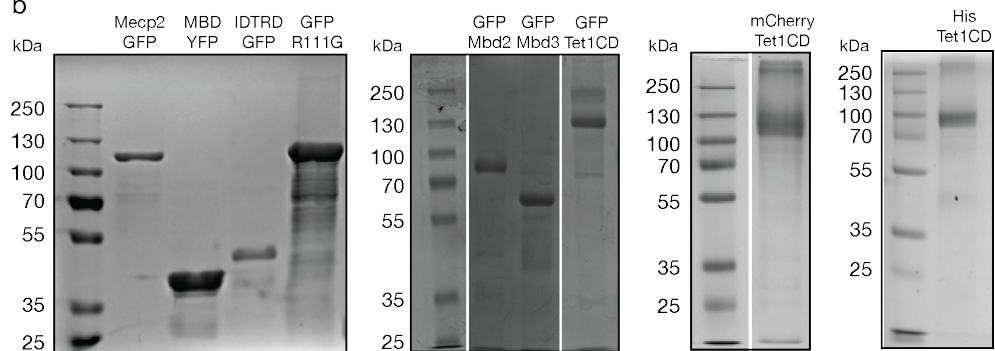
Supplementary Figure 11. Expression levels of Tet1CD and MBD in HEK versus C2C12 cells

Flow cytometry analysis of (a) HEK and (b) C2C12 cells ectopically coexpressing mCherry-tagged Tet1CD and the YFP-tagged MBD domain of Mecp2. Gate A: cells expressing high Tet1CD and low MBD protein levels. Gate B: cells expressing low Tet1CD and low MBD protein levels. Gate C: cells expressing high Tet1CD and high MBD protein levels. Gate D: Cells expressing low Tet1CD and high MBD protein levels. Two independent experiments were performed. Data of one representative experiment are shown. (c) Bar diagrams represent the number of cells expressing high Tet1CD and high MBD protein levels (Gate C). Shown are mean values and standard deviation of two independent experiments.

a

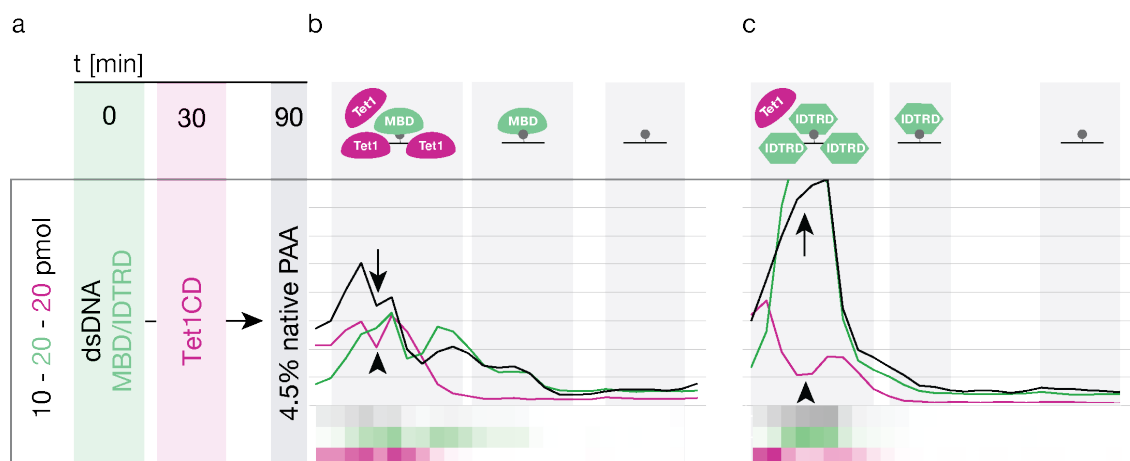


b



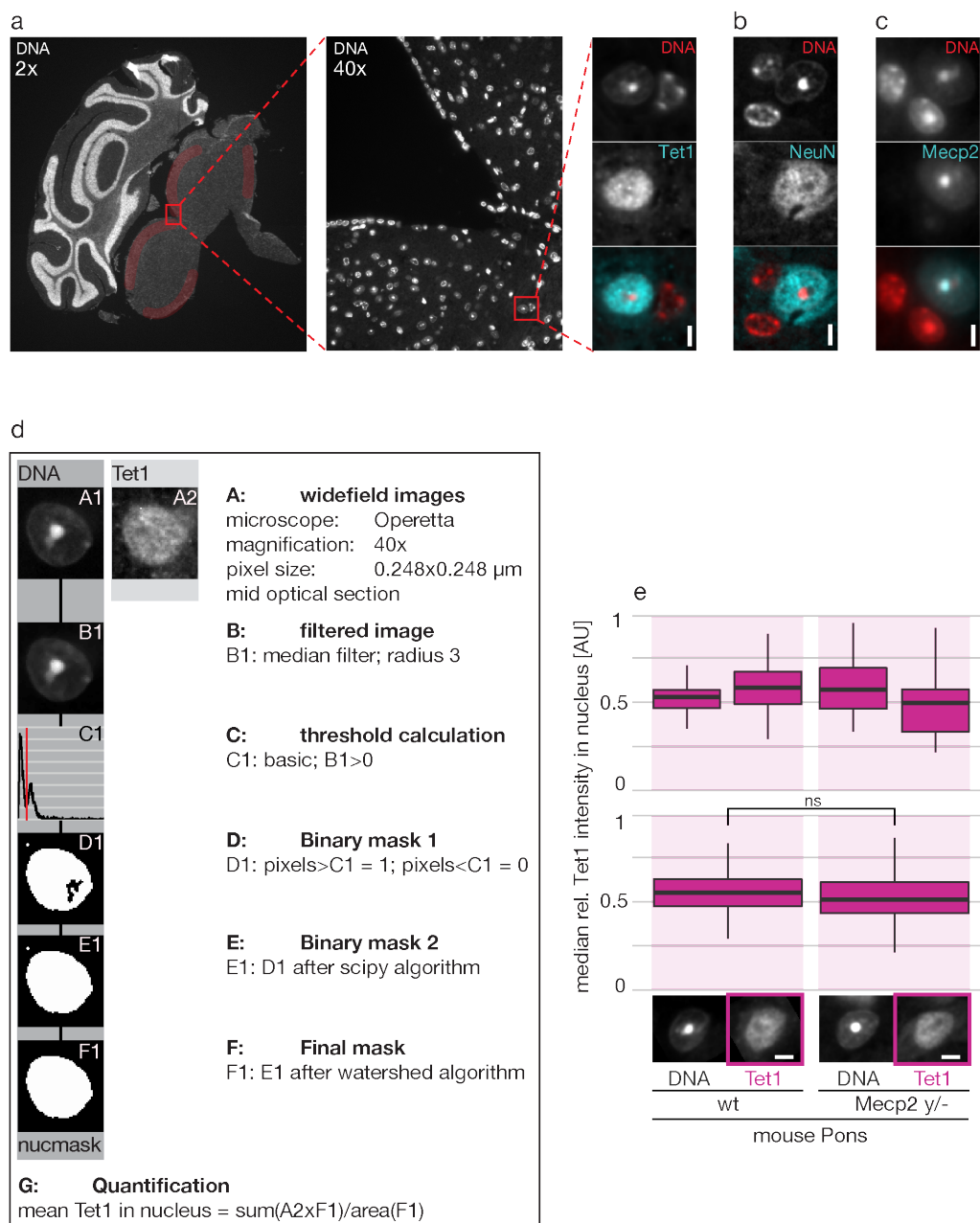
Supplementary Figure 12. Preparation of various controls required for the implementation and interpretation of Electrophoretic Mobility Shift Assays

(a) Schemes (top) illustrate the workflow of Electrophoretic Mobility Shift Assays (EMSA) fluorescent imager compared to a fluorescent plate reader. Both detection methods lead to the same result and are thus equally suitable for the analysis of EMSA. (b) Separation of purified, fluorescently tagged proteins via electrophoresis through a denaturing polyacrylamide gel visualized by Coomassie Brilliant Blue staining.



Supplementary Figure 13. Impact of 5mC-specific and sequence-unspecific DNA binding proteins on the DNA binding ability of Tet1CD proteins

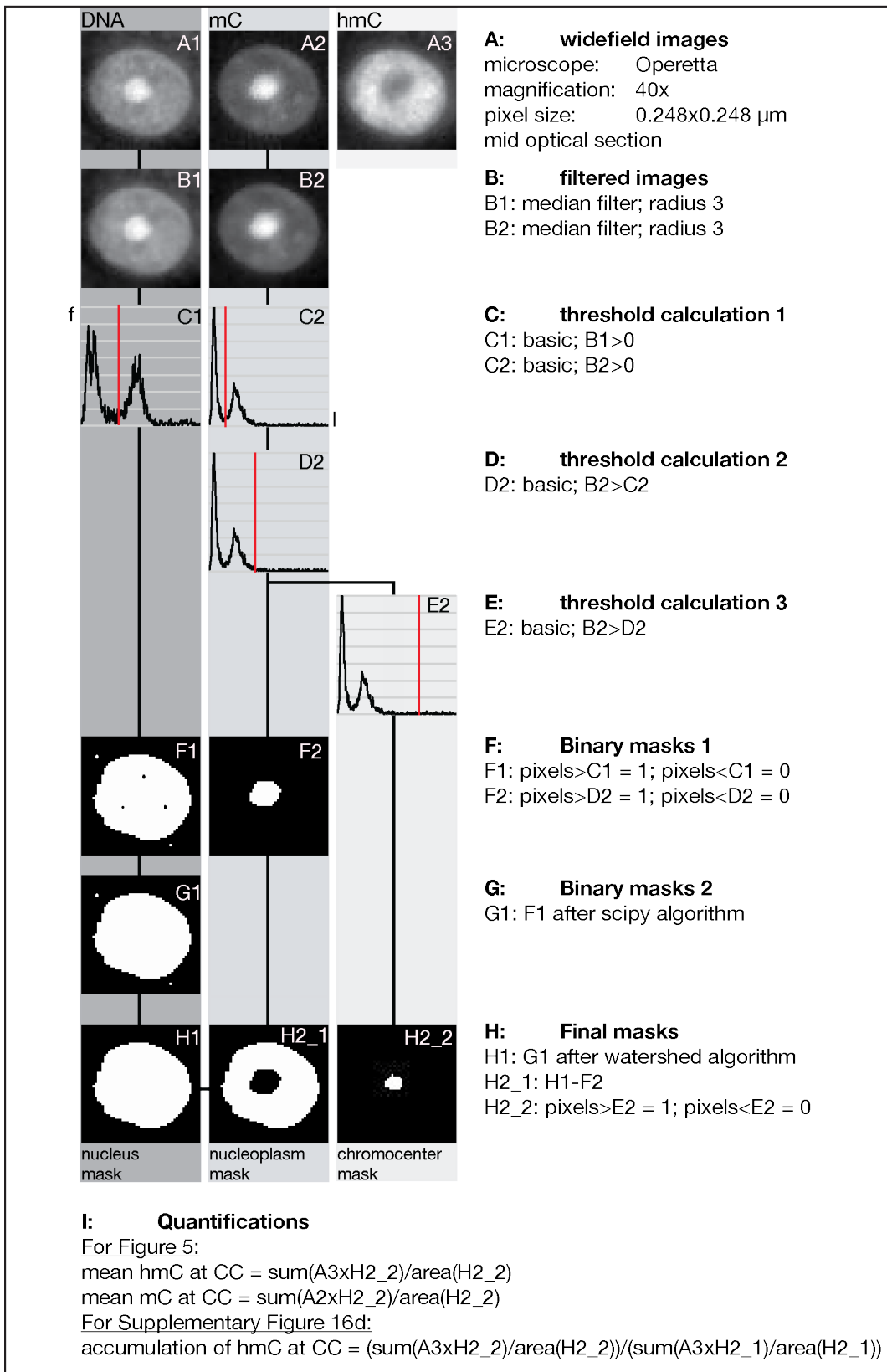
Electrophoretic mobility shift assay (EMSA) to determine the binding ability of fluorescently tagged Tet1CD to double-stranded, single mC containing DNA (ATTO647 labeled) in the presence of equimolar amounts of (b) 5mC specific (fluorescently tagged MBD) and (c) sequence-unspecific (fluorescently tagged IDTRD) DNA binding domain proteins, respectively. (a) Experimental setup illustrating the amount, as well as the incubation order and time of proteins and DNA prior to EMSA. (b) Separation of MBD-Tet1CD-dsDNA (n=3), as well as (c) IDTRD-Tet1CD-dsDNA (n=3) complexes via electrophoresis through a native polyacrylamide gel visualized using a fluorescent plate reader (see also **Suppl. Fig. 12**). (b and c) Running direction is from left (- pole) to right (+ pole). Arrow points to MBD and IDTRD protein/DNA complexes. Arrowhead points to Tet1CD/DNA complexes.



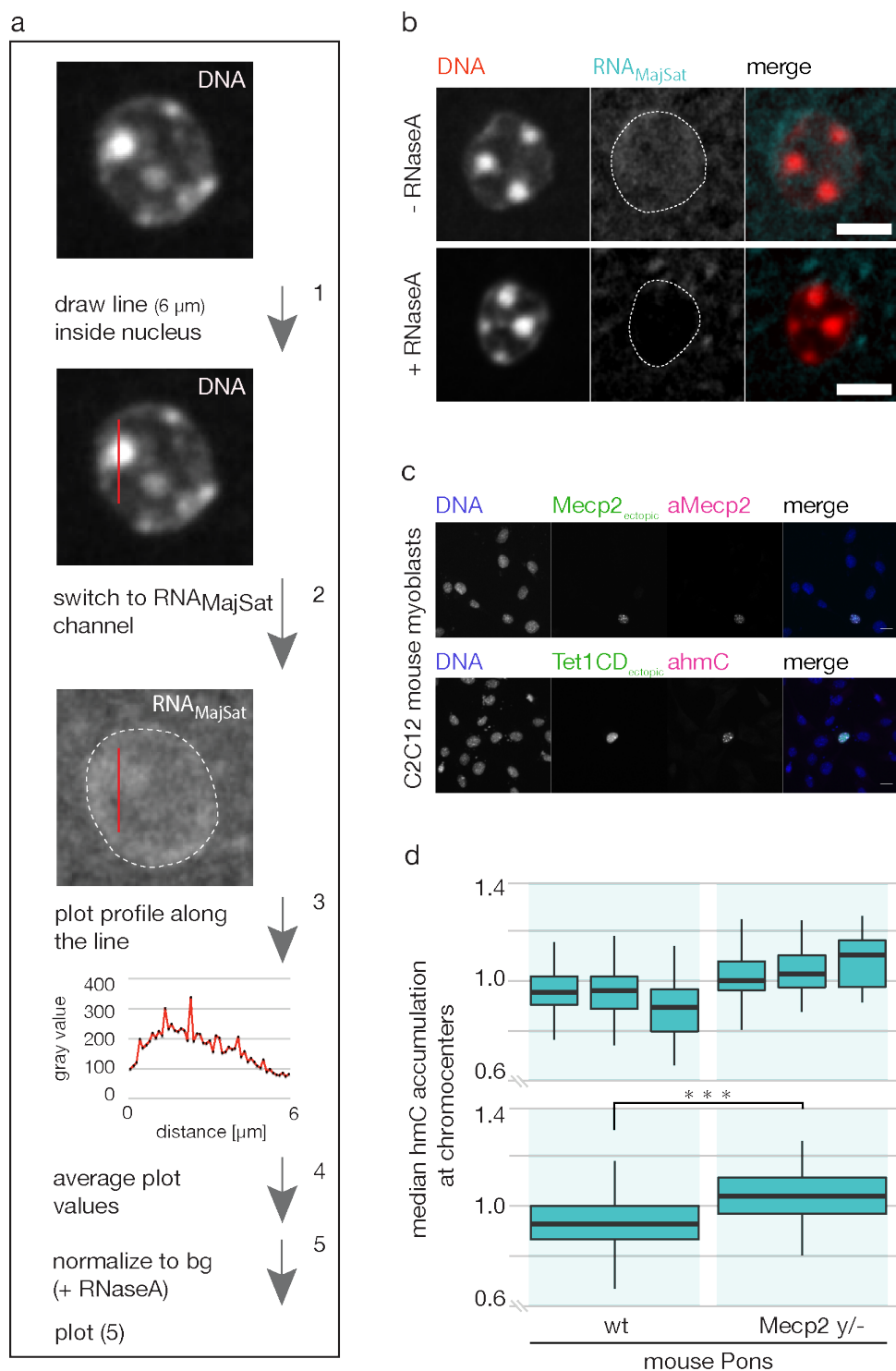
Supplementary Figure 14. Selection of neural cells for the quantification of Tet1, Mecp2, 5mC and 5hmC levels in wild type and Mecp2 knockout mouse pontes. (a) Representative low (2x) and high (40x) magnification scan of a coronal cross section through mouse brain used for quantification of nuclear Tet1 protein levels in single neural cells within the red-shaded regions of wild type and Mecp2 y⁻ pons. Tet1 proteins were detected immunologically using a Tet1-specific antibody. Neuronal cells were selected according to morphology properties, which were determined on (b) equivalent cross sections stained for Neuronal nuclear antigen (NeuN). 5mC, 5hmC and (c) MeCP2

levels were stained and quantified in a similar manner. **(a-c)** DNA was counterstained with DAPI. Scale bar, 5 μm . **(d)** Visual summary of the steps used for quantification of nuclear Tet1 protein levels in neural cells of wild type and *Mecp2* knockout mouse pontes. **(e)** Quantification result of **(d)**. Box plots represent the distribution of nuclear Tet1 levels in neurons of two individual (top) and combined (bottom) wild type and *Mecp2* $y^{-/}$ mouse pontes (wt1 n=29, wt2 n=64, KO1 n=56, KO2 n=43; from left to right; n=number of cells), respectively. Plotted is the median, as well as the first and third quartiles. Whiskers extend to 1.5 times the interquartile range. *P* values were calculated by Wilcoxon signed-rank test (ns = non significant). Mid-confocal optical sections represent NeuN positive cells of wild type and *Mecp2* $y^{-/}$ mouse pontes immunostained for Tet1. DNA was counterstained with DAPI. Scale bar, 5 μm .

a



Supplementary Figure 15. Visual summary of the steps used for quantification of Mecp2, 5mC and 5hmC levels at various different nuclear regions in neural cells of wild type and Mecp2 knockout mouse pontes. (a) Representative summary illustrating the quantification of 5mC and 5hmC levels at pericentric heterochromatin (chromocenter, CC), as well as the accumulation of 5hmC at chromocenters. DNA was counterstained with DAPI. For the quantification of Mecp2 levels, mouse pontes were immunostained for 5mC and Mecp2.



Supplementary Figure 16. Detection and quantification of major satellite transcripts via RNA FISH and RT qPCR, respectively. (a) Visual summary of the different steps used for quantification of Major Satellite RNA transcripts detected by RNA FISH. (b) Mid-optical sections of RNaseA treated and untreated neural cell nuclei hybridized with a probe specific for major satellite repeats. The lack of signal in RNaseA treated cells proves specificity of the probe for major satellite RNA. (c) Mid-optical

sections of transiently transfected (Mecp2 and Tet1CD, shown in green) C2C12 mouse myoblasts immunostained for Mecp2 and 5hmC (shown in magenta), respectively. The lack of signal in untransfected cells verifies the absence of Mecp2 and 5hmC in C2C12 mouse myoblasts. **(b,c)** DNA was counterstained with DAPI. Scale bar, 5 μm . **(d)** Accumulation of 5hmC at chromocenters in neurons of two individual (top) and combined (bottom) wild type and Mecp2 $y^{-/-}$ mouse pontes (wt1 $n=29$, wt2 $n=26$, KO1 $n=30$, KO2 $n=30$; from left to right; n =number of cells), respectively. Plotted is the median, as well as the first and third quartiles. Whiskers extend to 1.5 times the interquartile range. P values were calculated by Wilcoxon signed-rank test (***) $P < 0.001$.

4.2 L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins

A.K. Ludwig revised the manuscript.

A.K. Ludwig prepared figures S1, S2, S7 and S10 and wrote the respective figure legends.

A.K. Ludwig performed and analyzed the experiments of figures 1d, 3b, S1, S2, S7 and S10.

A.K. Ludwig performed and analyzed the experiments of figures 1e and 3c together with P. Zhang.

Content of figures:

Figure S1: Flow cytometry analysis of HEK cells ectopically coexpressing mcherry-Tet1CD/CDmut and GFP-tagged MBD proteins **Figure S2:** Quantification of chromatin condensation states in Tet1CD, Tet1CDmut and GFP expressing C2C12 cells. **Figure S7:** PCR analysis to validate the presence of pCEP4-LRE3-EGFP (EGFP with γ -globin intron) and the absence of genomically integrated LRE3-EGFP (EGFP without γ -globin intron) in the retrotransposition reporter cell line. **Figure S10:** Luciferase reporter assay to analyze whether transcription from a methylated L1 promoter is activated by Tet1CD and repressed by Mecp2. **Figure 1d and 3b:** RT-qPCR analysis of L1 transcript levels in transiently transfected HEK cells. **Figure 1e and 3c:** L1 copy number analysis of transiently transfected HEK cells.

This chapter is accepted for publication in *Nucleus*.

L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins

Peng Zhang^{1*}, Anne K. Ludwig^{1*}, Florian D. Hastert¹, Cathia Rausch¹, Anne Lehmkuhl¹, Ines Hellmann², Martha Smets³, Heinrich Leonhardt³ and M. Cristina Cardoso¹

¹Department of Biology, Technical University Darmstadt, 64287 Darmstadt, Germany

²Anthropology and Human Genomics, Department Biology II, LMU Munich, Germany

³Human Biology and Bioluminescence, Department of Biology II, LMU Munich, Germany

*Equally contributing authors

One of the major functions of DNA methylation is the repression of transposable elements, such as the long-interspersed nuclear element 1 (L1). The underlying mechanism(s), however, are unclear. Here, we addressed how retrotransposon activation and mobilization are regulated by methyl-cytosine modifying ten-eleven-translocation (Tet) proteins and how this is modulated by methyl-CpG binding domain (MBD) proteins. We show that Tet1 activates both, endogenous and engineered L1 retrotransposons. Furthermore, we found that Mecp2 and Mbd2 repress Tet1-mediated activation of L1 by preventing 5hmC formation at the L1 promoter. Finally, we demonstrate that the methyl-CpG binding domain, as well as the adjacent non-sequence specific DNA binding domain of Mecp2 are each sufficient to mediate repression of Tet1-induced L1 mobilization. Our study reveals a mechanism how L1 elements get activated in the absence of Mecp2 and suggests that Tet1 may contribute to Mecp2/Mbd2-deficiency phenotypes, such as the Rett syndrome. We propose that the balance between methylation "reader" and "eraser/writer" controls L1 retrotransposition.

Introduction

In humans, 17% of nuclear DNA consists of long interspersed nuclear element 1 (LINE1 or L1). The majority of L1s, however, are retrotransposition defective (RD-L1) due to 5' truncations, internal rearrangements or mutations. Only 80-100 copies of the half a million human L1s are retrotransposition competent (RC-L1) [1]. Full length L1 has a total length of about six kilobase pairs and contains a 5' untranslated region (5'UTR) with promoter activity in both, sense and antisense directions, three open reading frames (ORFs) and a 3'UTR that ends in an AATAAA polyadenylation signal. ORF1 encodes for a p40 protein with RNA-binding and chaperone activities [2], whereas ORF2 encodes for a protein of 150 kDa in size with endonuclease and reverse transcriptase activities [3]. In contrast to ORF1 and ORF2, ORF0 is primate-specific and lies downstream the 5'UTR in antisense direction. It has two splice donor sites that can react with splice acceptors of downstream genomic

sequences to generate fusion proteins [4]. During the L1 retrotransposition procedure, ORF1 and ORF2 proteins bind to their own RNA in the cytosol to form a ribonucleoprotein particle (RNP), which facilitates the re-import of L1 RNA to the nucleus. The majority of genomic L1 integrations follow a mechanism termed target primed reverse transcription (TPRT) [5], which involves endonuclease and reverse transcriptase activity of ORF2p. However, endonuclease independent L1 integration is also observed in non-homologous end joining (NHEJ) and p53 double deficient cells [6, 7]. Previous studies using engineered L1 elements showed that retrotransposition occurs not only in brain cells like neural progenitor cells of rat hippocampus [8] and human fetal brain [9], but also in non-brain cells such as embryonic stem cells (ESC) [10]. Moreover, transgenic mice, harboring a fluorescently tagged human L1 under the control of its endogenous promoter showed only detectable L1 retrotransposition activity in mouse germ cells and brain. Furthermore, methylation level of L1 elements differed in brain and skin [8], indicating that L1s are differently regulated in tissues and cell types. Altogether, these studies demonstrate that retrotransposition of L1 elements can occur in embryonic and importantly, also in somatic cells and correlates with the L1 promoter methylation status. Although host cells have multiple mechanisms to restrict L1 retrotransposition [11–14], sporadic insertions of a small number of RC-L1s accompanied by large chromosomal rearrangements can occur that lead to genomic instability [15]. Insertions of L1 sequences into protein coding regions of the genome can decrease RNA levels by inhibiting transcriptional elongation [16]. In addition to its ability to propagate itself, 3' transductions of L1 occur in germ and cancer cells, whereby unique sequences downstream of L1 elements can also be retrotransposed, if transcription continues beyond the L1 sequence [17]. Moreover, other studies showed that L1 expression leads to high levels of double strand breaks, as evidenced by the formation of γ H2AX foci and the recruitment of repair proteins involved in the L1 retrotransposition process [18]. One of the most important mechanisms repressing L1 retrotransposition depends on DNA methylation. Methylation of cytosines has been shown to recruit 5-methylcytosine (5mC) binding domain (MBD) proteins. Mecp2, the founding member of the MBD protein family, was subsequently shown to modulate L1 retrotransposition in a DNA methylation dependent manner [19, 20]. While the transcriptional repression domain (TRD) of Mecp2 was shown to be sufficient for repressing L1 retrotransposition in a reporter assay system [20], the mechanism(s) of L1 repression mediated by Mecp2 remain unknown. Mecp2 and Mbd2, another family member of the MBD protein family, specifically bind to methylated CpG dinucleotides in DNA [21]. Hypomethylation of the L1 promoter is associated with overexpression of L1 transcripts [11], suggesting that a decrease in DNA methylation might play a role in L1 retrotransposition. Ten-eleven-translocation (Tet) proteins convert 5-methylcytosine to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in an iterative iron- and oxoglutarate dependent oxidation reaction. These further modifications of methylated cytosines are proposed to be one of the long sought mechanisms leading to loss of DNA methylation [22]. Tet1 and Tet2 depletion in ESCs has been shown to cause loss of 5hmC in the 5' region of L1 [23], but a connection to L1 regulation is lacking. In this study, we investigate whether and how Mbd and Tet proteins affect L1 expression and mobilization in human cells. We detected increased transcription and transposition of human endogenous L1 in the presence of Tet1 proteins and showed that activation of L1 transposition depends on the catalytic activity of Tet1. By the use of an L1 retro-

transposition reporter assay, we additionally showed that Tet1 proteins activate retrotransposition of engineered L1. Moreover, we found that Mbd2, Mecp2, as well as its methylcytosine binding and transcriptional repression subdomains counteract Tet1-mediated reactivation of L1 retrotransposons.

Materials and Methods

Plasmids, cell culture and transfection

Plasmids coding EGFP tagged MBDs were described in previous publications [21, 24–26].

For construction of the mcherry-tagged catalytic active (mcherry-Tet1CD: aa 1367-2007) and inactive (mcherry-Tet1CDmut: aa 1367-2007, H1652Y, D1654A) domain of mouse Tet1, Np95 was replaced from the mammalian expression vector pCAG-mcherry-Np95-IB54 by Tet1CD and Tet1CDmut28, respectively using AsiSI and NotI sites.

The reporter plasmid pGL3-L1.3-Luc [20] coding for a firefly luciferase under the control of the L1.3 promoter was a generous gift of G. Schumann (Paul Ehrlich Institute, Langen, Germany).

HEK-EBNA cell line was purchased from Invitrogen (catalog #R620-07; Paisley PA4 9RF, UK). Cells were cultured and transfected as previously described [27].

For genomic DNA (gDNA) extraction, HEK-EBNA cells were transfected with mcherry-Tet1CD/CDmut and EGFP-MBDs and flow cytometry sorted (Biorad S3 sorter, Bio-Rad Laboratories, USA) according to EGFP and mcherry expression levels. 488 and 561 nm excitation lasers and 525 ± 30 and 586 ± 25 nm emission filters are used for EGFP and cherry detection.

For RNA preparation, HEK-EBNA cells were transfected with mcherry-Tet1CD/CDmut and EGFP-MBDs encoding plasmids, and the expression of mcherry and EGFP was analyzed by flow cytometry. Cells with similar mcherry-

Tet1CD/CDmut expression were used for RNA preparation.

C2C12 mouse myoblast cell line [28] were cultured using the conditions described previously [29]. C2C12 cells were grown to 70% confluency on glass cover slips and transfected with EGFP-Tet1CD/CDmut and EGFP expression constructs 24 hours post seeding using Lipofectamine (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions.

Human AG01522D foreskin fibroblasts cells (obtained from Coriell Cell Repository, Camden, NJ, USA) were cultured in DMEM medium supplemented with 15% FCS. The cells were transfected by electroporation as previously described [30].

Mouse tail fibroblasts (MTF) cells were a gift from A. Bird (Wellcome Trust Centre for Cell Biology, Edinburgh, UK) and were cultured using the conditions as previously described [31].

Mouse neural stem cells were a gift from B. Hendrich (Cambridge Stem Cell Institute, Cambridge, UK) and cultured using the conditions as previously described [32].

Human fibroblast cells (BJ-hTERT; ATCC CRL-4001) 60 were cultured in DMEM medium supplemented with 15% FCS. The cells were transfected by electroporation as previously described [30].

V6.5 wt and Tet-TKO mouse embryonic stem cells [33] were a gift from R. Jaenisch (Whitehead Institute for Biomedical Research, Cambridge, USA) and were maintained under serum-free and feeder-free conditions on Geltrex-coated flasks in N2B27 (50% neurobasal medium (Life Technologies) and 50% DMEM/F12 (Life Technologies) containing 2 mM L-glutamine (Life Technologies), 0.1 mM β -mercaptoethanol, N2 supplement (Life Technologies), B27 serum-free supplement (Life Technologies), 100 U/ml Penicillin-Streptomycin, 1000 U/ml LIF and 2i (1 μ M PD032591 and 3 μ M CHIR99021

(Axon Medchem, Netherlands)).

L1 retrotransposition reporter assay

HEK-EBNA cells were transfected with the pLRE3-EGFP plasmid [9] (gift from J. V. Moran, U. Michigan Medical School, USA). Two days after transfection, cells were cultured in DMEM medium containing 2 $\mu\text{g}/\text{mL}$ puromycin (Invitrogen; Paisley PA4 9RF, UK) for resistance selection. 18 days later, single colonies were picked and passaged in DMEM medium containing 2 $\mu\text{g}/\text{mL}$ puromycin. For retrotransposition reporter assays, cells were cultured in DMEM medium without puromycin selection.

For fixed-cell analysis, the reporter cell line was grown on glass cover slips and transfected with mcherry-Tet1CD/CDmut and only mcherry using PEI. 24 hours and 48 hours after transfection, cells were fixed with 3.7% formaldehyde (Carl Roth GmbH, Karlsruhe, Germany), DNA was counterstained with DAPI (Invitrogen; Paisley PA4 9RF, UK) and cells were mounted in Vectashield antifade medium (Vector Laboratories, Burlingame, CA, USA).

DAPI, EGFP and mcherry were imaged by high content microscopy with 20x long/0.45 NA objective (Operetta, PerkinElmer, UK), a xenon fiber optic as light source, 360-400, 460-490 and 560-580 nm excitation- and 410-480, 500-550 and 590-640 emission filters, respectively. And the intensities were analyzed using the Harmony software (PerkinElmer, UK).

For flow cytometry analysis of the aforementioned reporter cell line, cells either ectopically expressing mcherry-Tet1CD/CDmut or mcherry or treated with 5-azacytidine (5-aza-C, Sigma-Aldrich, St. Louis, MO, USA) were grown in 6-well plates. 48 hours later, cells were harvested and EGFP positive cells were counted by flow cytometry.

RNA-Seq library preparation and data analysis

Total RNA was isolated from mouse embryonic stem

cells (V6.5) in biological quadruplicates using the nucleospin triprep kit from Macherey-Nagel. 50 ng RNA was reverse transcribed. cDNA was pre-amplified as described elsewhere [34]. 1 ng of cDNA was used as input for tagmentation by the Nextera XT Sample Preparation Kit (Illumina), where a second amplification round was performed for 12 cycles. For each sample, 2.5 ng of final library was pooled. The library pool was sequenced 1 x 100 bases on a Illumina HiSeq1500. The average sequencing depth was 1.2 million reads per replicate. Sequencing reads were demultiplexed from the Nextera (i5 and i7) indices. Demultiplexed reads were mapped to the Mouse genome build mm10 using STAR version STAR 2.5.163 with the specific settings: outFilterMultimapNmax 100, outFilterMismatchNmax 4, winAnchorMultimapNmax 100. The junction annotation was taken from ensembl GRCm38.75 and the index was created as recommended using the option sjdbOverhang 99. The resulting bam-files were then processed using TETranscript [35] to obtain read count tables for transcripts and transposons, using the TE annotation as provided by the authors of TETranscript (http://labshare.cshl.edu/shares/mhammelllab/www-data/TEToolkit/TE_GTF/mm10_rmsk_TE.gtf.gz). Normalization and differential expression analysis was done using DESeq265.

Quantitative PCR (q-PCR)

q-PCR was performed on a StepOnePlus Real-Time PCR System (Applied Biosystems, Darmstadt, Germany) with Platinum SYBR Green qPCR SuperMix-UDG w/ROX (Invitrogen, Paisley PA4 9RF, UK) according to the manufacturer's instruction. The program used for the amplification of all fragments consisted of 1) inactivation of UDG for 2 min at 50°C, 2) denaturation of DNA for 10 min at 98°C, 3) 40 cycles of PCR (98°C for 15 sec, 60°C for 1 min), followed by 4) dissociation (melting) curve

analysis to confirm the specificity of the amplicon.

DNA glucosylation, MspI digestion and quantitative PCR based 5hmC and 5mC detection (GluMs-qPCR)

To detect 5hmC after Tet1 transfection, gDNA was extracted from HEK-EBNA cells as described previously [36]. Concentration and purity of DNA was measured on a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland) by the absorbance at 260 nm and 280 nm. 1 μ g of gDNA was treated with or without 0.18 μ M of T4 phage β -glucosyltransferase (T4-BGT) [37] in a final volume of 50 μ l supplemented with 1x NEB cut smart buffer (NEB, Frankfurt, Germany) and 1 mM of UDP-Glucose (Sigma-Aldrich, St. Louis, MO, USA) for 18 hours at 37°C. Then 0.5 μ g of glucosylated or mock treated DNA was used for digestion with 100 units of MspI (NEB, Frankfurt, Germany) at 37°C for 18 hours in a final volume of 20 μ L, which was followed by treatment with 20 μ g of proteinase K (PK, Carl Roth GmbH, Karlsruhe, Germany) for 30 min at 50°C. Following proteolysis, PK enzymatic activity was inactivated for 10 min at 98°C. The MspI-resistant fraction was amplified using qPCR with primers flanking the MspI site (F: 5'- ATCCCACACCTGGCTCAGAGGG -3' and R: 5'- GTCAGGGGTCAGGGACCCACTT -3'). After qPCR, the relative amounts of 5hmC were analyzed as described previously [38]. To detect 5mC at position 482 in L1 5'UTR before Tet1 transfection, the gDNA was treated with or without T4-BGT as described above. Then, the gDNA was further treated with MspI, HpaII (50 units, NEB, Frankfurt, Germany) or mock for 18 hours at 37°C. The qPCR and data analysis were performed as above.

cDNA preparation and reverse transcription quantitative real-time PCR (RT qPCR)

Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's

instruction. To remove traces of genomic DNA, RNA was treated with RNase-free recombinant DNaseI (Macherey Nagel, Dueren, Germany) for 30 min at 37°C and further purified with the Qiagen RNeasy Mini Kit. To assess the concentration and purity of RNA, the ratio of absorbance at 260 nm and 280 nm was measured on a TECAN infinite M200 plate reader. 500 ng of total RNA were used for cDNA synthesis using 200 units M-MuLV reverse transcriptase (NEB, Frankfurt, Germany), 0.01 OD units random primer from the Prime-It II Random Primer Labeling Kit (Stratagene, La Jolla, California, USA), 0.5 mM dNTPs (Carl Roth GmbH, Karlsruhe, Germany) and 40 units recombinant ribonuclease inhibitor RNaseOUT (Invitrogen, Paisley PA4 9RF, UK) in a total reaction volume of 20 μ L. Cycles were set to 5 min at 25°C, 90 min at 50°C and 15 min at 70°C. For qPCR, 0.5 ng of cDNA were used for each reaction. Primers for quantitative real-time PCR contained the following sequences: Gapdh F: 5'- CAT GAG AAG TAT GAC AAC AGC CT-3', Gapdh R: 5'-AGT CCT TCC ACG ATA CCA AAG T-3' [39], hL1 5' UTR F: 5'-GAA TGA TTT TGA CGA GCT GAG AGA A-3', hL1 5' UTR R: 5'-GTC CTC CCG TAG CTC AGA GTA ATT -3' [9]. 5' UTR expression level was normalized to Gapdh and calculated using the comparative CT method ($\Delta\Delta$ CT method) [40].

L1 copy number analysis

To detect newly integrated L1 ORF2 sequences in human embryonic kidney or human fibroblast cells, qPCR was performed as described above. For each reaction, 80 pg of gDNA were used. Primers contained the following sequences: hL1 ORF2 F: 5'-CAAACACCGCATATTCTCACTCA-3', hL1 ORF2 R: 5'-CTTCCTGTGTCCATGTGATCTCA-3' [9]. Computational estimates using the UCSC genome browser in silico PCR function (genome: human, assembly: December 2013 (GRCh38/hg38), target: genome as-

sembly) indicated that at least 2734 endogenous human L1 elements could be detected using the hL1 ORF2 primer set. As control we used primers for hL1 5' UTR. This sequence is usually truncated during the course of retrotransposition. Accordingly, the number of hL1 5' UTR sequences remains constant. Primers contained following sequences: hL1 5' UTR F: 5'-ACAGCTTTGAAGAGAGCAGTGGTT-3', hL1 5' UTR R: 5'-AGTCTGCCCGTTCTCAGATCT-3' [9]. To determine the content of L1 ORF2 sequences in mouse embryonic stem cells (wild type and Tet1/Tet2/Tet3 TKO) and mouse fibroblast cells ectopically expressing mcherry or mcherry-Tet1CD, gDNA was isolated and qPCR was performed as described above. For each reaction, 80 pg of gDNA were used. Primers contained the following sequences: mL1 ORF2 F: 5'-CTGGCGAGGATGTGGAGAA-3', mL1 ORF2 R: 5'-CCTGCAATCCCACCAACAAT-3' [19]. Computational estimates using the UCSC genome browser in silico PCR function (genome: mouse, assembly: December 2011 (GRCm38/mm10), target: genome assembly) indicated that at least 1308 endogenous mouse L1 elements could be detected using the mL1 ORF2 primer set. As control we used non-mobile genomic repetitive sequences: 5SRNA F: 5'-ACGGCCATACCACCCTGAA-3'; 5SRNA R: 5'-GGTCTCCCATCCAAGTACTAACC-3' [19]. To calculate the relative copy number of genomic L1 ORF2, the comparative CT method ($\Delta\Delta$ CT method) was used. For HEK cells, as well as for mouse and human fibroblast cells, relative ORF2 content was further normalized to Tet1CD expressing cells. For mouse ESCs, relative ORF2 content of wild type cells was further normalized to the mean relative ORF2 content of Tet-TKO cells.

Chromatin decondensation analysis

3D structured illumination microscopy images were acquired as previously described [41]. To quantify the

grade of chromatin decondensation, binary nuclear masks were generated. Therefore, images were processed using a 3D median filter. Filtered images of the DAPI channel were then thresholded using the basic algorithm. Next, all DAPI pixels below the threshold were set to 0 and all pixels above the threshold were set to 1. For further improvement of the nuclear masks, binary images were additionally processed using the "fill holes" and "watershed" algorithms. Finally, the standard deviation of all DAPI histograms was calculated automatically. To automate this procedure, a routine was written in the programming language python.

Immunofluorescence staining

Human AG522D fibroblasts cells were fixed for 10 minutes in 4% formaldehyde in PBS and permeabilized for 20 minutes with 0.5% Triton X-100. For detection of genomic 5hmC and γ H2AX, the cells were further fixed with ice-cold methanol for 5 minutes. After RNaseA treatment (10 μ g/mL) for 30 minutes at 37°C, cells were washed and blocked for 60 minutes in 0.2% fish skin gelatin (Sigma Aldrich, St. Louis, MO, USA) at 37°C. Then 5hmC and γ H2AX were detected using a rabbit anti-5hmC (1:250, catalog number: 39769, Active Motif, La Hulpe, Belgium) and a mouse Anti-phospho-Histone H2A.X (1:400, catalog number: 05-636, Merck, Darmstadt, Germany), antibody in conjunction with 25 U/mL DNaseI (Sigma Aldrich, St. Louis, MO, USA) for 70 minutes at 37°C. To stop DNaseI digestion, cells were washed three time with PBS containing 1mM EDTA and 0.01% Tween. Following incubation with secondary antibody mixture of AMCA donkey anti-rabbit IgG (1:100, catalog number: 715-155-151, The Jackson Laboratory, Bar Harbor, USA) and Alexa Fluor 488 goat anti-mouse IgG (1:250, catalog number: A11029, Invitrogen Invitrogen, Paisley PA4 9RF, UK) for 50 minutes at RT. After three time washing with PBS containing 0.01%

Tween the cells were mounted in Vectashield Medium (Vector Labs, Burlingame, CA, USA) and imaged with high content screen microscopy with 20x long/0.45 NA objective (Operetta, PerkinElmer, UK). The intensities were further calculated using the Harmony software (Operetta, PerkinElmer, UK) and blotted by RStudio (<https://www.rstudio.com>).

L1 ORF1p proteins were detected in HEK using the polyclonal rabbit anti-L1 ORF1p antibody [42] as described before [43]. As secondary antibody the AMCA donkey anti-rabbit IgG was used (1:100, catalog number: 715-155-151, The Jackson Laboratory, Bar Harbor, USA).

Luciferase reporter assay

Methylated pGL3-L1.3-Luc [20] reporter plasmids were obtained by incubation with M.HpaI and controlled by digestion with HpaI and MspI. Cells were seeded in 6-well dishes at 7×10^5 cells/well. Three hours post seeding, cells were cotransfected with the unmethylated or methylated reporter plasmid pGL3-L1.3-Luc and effector constructs coding for Tet1CD/Tet1CDmut/Mecp2/Tet1CD+Mecp2, respectively. Luciferase activity was determined using the "Luciferase Assay System" (Promega, Madison, Wisconsin, US) as described by the manufacturer on a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland). To control for consistent transfection of methylated and unmethylated reporter plasmids, the fluorescent signal emanating from the effector proteins was quantified in parallel (GFP: excitation 475 nm, emission 520 nm; RFP: excitation 585 nm, emission 630 nm) and was used for normalization of the luciferase signal. In addition, fluorescent signals were used to control for homogeneous expression of Tet1CD in single (Tet1CD) and double (Tet1CD + Mecp2) transfected cells.

Results

Tet1 activates retrotransposition of endogenous L1

Three steps are involved in L1 retrotransposition, comprising loss of DNA methylation in the L1 5'UTR, L1 transcription and L1 transposition (Fig 1A). 5hmC, which is produced by Tet proteins, is thought to be an intermediate modification during loss of DNA methylation [44, 45], as well as a stable epigenetic mark, which tunes a large number of CpG dinucleotides located at poised enhancers and actively transcribed regions [46, 47]. To test whether L1 retrotransposition is reactivated by Tet-mediated 5mC to 5hmC conversion, we transfected HEK-EBNA cells with the catalytically active domain of Tet1 fused to mcherry (mcherry-Tet1CD), which was previously shown to be sufficient to induce genome-wide hydroxymethylation *in vivo* [48–50]. As control, mcherry-Tet1CDmut, which lacked catalytic activity due to two mutations in the Fe(II) binding sites (H1652Y, D1654A) was used [48]. Since loss of DNA methylation in the L1 5'UTR is the first step that makes L1 retrotransposition possible, we performed GluMs-qPCR (DNA glucosylation, MspI digestion and quantitative PCR based 5hmC detection, described in the methods) to determine the methylation state of the L1 promoter. L1 has 20 mapped CpG sites within its 5'UTR. Methylation of only a subset of CpG dinucleotides, such as nucleotide position 482, were shown to correlate inversely with retrotransposition activity of L1 elements [9]. Therefore, we chose CpG dinucleotide at position 482 to quantify the methylation status of the L1 promoter (Fig 1B) before and after transfection. In untransfected cells, we observed using GluMs-qPCR that around 60% of all CpG dinucleotides at position 482 were methylated upon Tet1CD expression, however, 5hmC levels at position 482 were increased compared

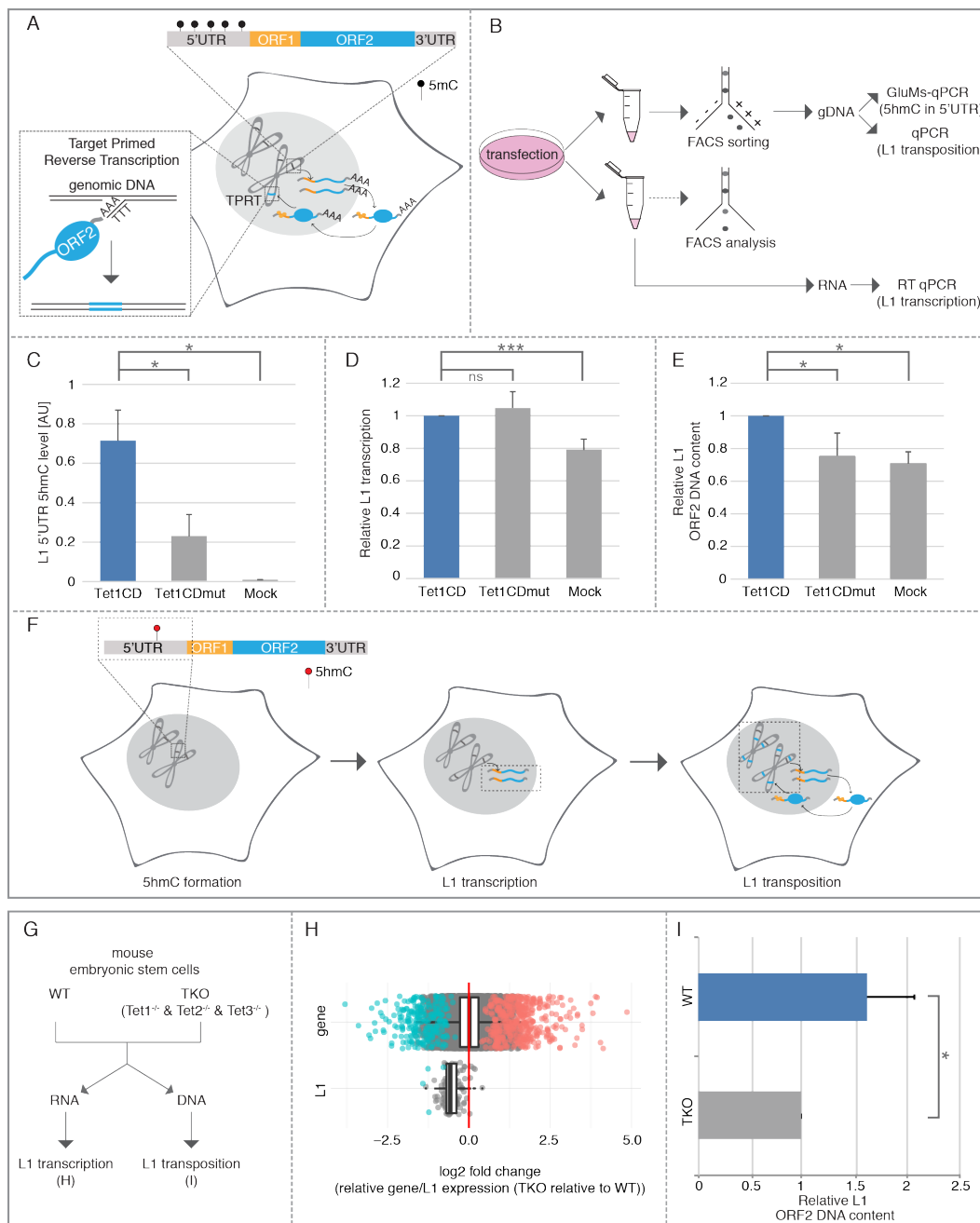


Figure 1: Tet1 reactivates retrotransposition of endogenous L1. (A) Schematic overview of L1 retrotransposition. (B) Experimental rationale of human L1 retrotransposition detection. (C) Relative 5hmC levels in L1 5'UTR ($n=3$, $*p < 0.05$, independent two-sample student's t-test), (D) relative L1 transcription levels ($n \geq 4$; ns= non significant; $***p < 0.001$, independent two-sample student's t-test) and (E) relative L1 ORF2 DNA content was checked 48 hours after Tet1CD, Tet1CDmut and mock transfection ($n=4$, $*p < 0.05$, independent two-sample student's t-test). (F) Scheme summarizing the effect of Tet1 on the three steps of L1 retrotransposition. Black and red circles indicate 5mC and 5hmC nucleotides, respectively. Bars represent the mean + standard deviation (SD). (G) Experimental rationale of mouse L1 retrotransposition detection. (H) Boxplot of the log₂-fold changes of the triple mouse embryonic stem cells Tet-knockout relative to wild-type (V6.5) for all genes and all L1 elements. Negative values indicate a down-regulation in the knock-out relative to the wild-type, positive values an up-regulation. Significant elements are marked in color. The red line is at zero, i.e. the expected value if expression were identical in the wild type and mutant. (I) Relative mouse L1 ORF2 content. Bars represent mean + SD. ($n=3$, $*p < 0.05$, independent two-sample student's t-test).

to Tet1CDmut and mock transfected cells (Fig 1C), indicating that Tet1 oxidizes 5mC in the L1 promoter. Consequently, we aimed to determine, whether Tet1-mediated 5hmC formation of the L1 promoter leads to reactivation of L1 transcription. Since RNA transcript levels might be affected by several exogenous factors during cell sorting [51], cells cotransfected with mcherry-Tet1CD and EGFP, as well as mcherry-Tet1CDmut and EGFP were analysed by flow cytometry and only cell populations with comparable expression levels between samples were chosen for RNA preparation (Fig S1). Reverse transcription quantitative real-time PCR (RT qPCR) showed that besides Tet1CD, ectopic expression of Tet1CDmut leads to increased L1 transcription when compared to mock transfected cells (Fig 1D), indicating that L1 transcription does not depend on the catalytic activity of Tet1 proteins. Previous studies showed that besides DNA methylation, histone acetylation [52] and chromatin structure [53] affect L1 retrotransposition. In that regard, we found chromatin decondensation caused by Tet proteins in an oxygenase-independent manner (Fig S2), which might contribute to the reactivation of L1 transcription. Using high content imaging, we next tested, whether the observed increase in L1 transcription results in elevated L1 ORF1p levels. Therefore, HEK cells ectopically expressing mcherry-Tet1CD, mcherry-Tet1CDmut or mcherry were fixed and immunostained for L1 ORF1p. While the catalytically active Tet1 variant increased mean protein levels of L1 ORF1 by 60%, ORF1p levels were elevated by 20% in Tet1CDmut expressing cells (Fig S3). Accordingly, L1 ORF1p formation requires, at least in part, the catalytic activity of Tet proteins. Based on the knowledge that L1 elements mobilize via an RNA intermediate using a "copy and paste" mechanism and our observation that Tet1 proteins increase L1 transcription, we next analyzed L1 copy numbers in genomic DNA (gDNA) of Tet1CD and Tet1CDmut overexpressing cells, respectively. Since most of the newly inserted L1 elements are 5' truncated, we used primers specific for the 5'UTR as a constant normalization control and ORF2 primers as an indication of newly inserted L1 elements to detect de novo transposition events. Compared to Tet1CDmut and mock transfected cells, genomic L1 ORF2 content was increased upon Tet1CD overexpression (Fig 1E), indicating that L1 transposition depends on the oxygenase activity of Tet1. As we observe that catalytically active and inactive Tet1 proteins generate similar L1 mRNA levels, but have different L1 transposition efficiencies, we suggest that the Tet 1 catalytic activity may enhance L1 transposition through a yet unknown mechanism. We found that the global Tet1-mediated 5hmC increase leads to accumulation of γ H2AX (Fig S4). As L1 retrotransposition has been shown to be proportional to the number of γ H2AX foci [54], we suggest that Tet1 induced formation of γ H2AX might enhance L1 integration. Since, on the other hand, the catalytically inactive Tet1 does not produce 5hmC and, accordingly, γ H2AX was not increased (Fig S4), L1 transposition was thus not elevated. In summary, our data demonstrate that Tet1 proteins induce loss of L1 promoter methylation and further activate L1 transcription and transposition in human HEK cells (Fig 1F). As ESCs depleted of Tet1 and Tet2 showed loss of 5hmC in the 5' region of L1 [23], we further tested, whether loss of 5hmC affects L1 transcription and transposition. To this end, we performed RNA-seq analysis and determined L1 copy numbers using RNA and DNA from Tet1/Tet2/Tet3 triple knockout ESCs (Tet-TKO), as well as the corresponding wild type cells (V6.5 wt) (Fig 1G). Although global gene expression did not change, L1 expression was significantly reduced in Tet-TKO cells (Fig 1H), indicating that Tet proteins regulate L1 transcription. In agreement with this, L1 copy numbers were decreased in Tet-TKO cells, indicating that Tet proteins

modulate L1 transposition. The activity of L1 is tissue and cell type dependent [8]. Low and high activity of L1 retrotransposition was observed in fibroblasts and neural stem cells, respectively, indicating that host cells adopt a protection mechanism to prevent L1 retrotransposition. To test whether Tet proteins could activate L1 in both cell types, we analyzed L1 copy number in cells expressing either mcherry-Tet1CD or mcherry. As shown in Figure S5, genomic L1 ORF2 content was increased in mouse neural stem cells, however, it was not changed in both mouse and human fibroblasts upon Tet1 expression. These results indicate that although Tet proteins can activate L1 retrotransposition, cell type dependent protection mechanisms also play a role in preventing L1 activity.

Tet1 activates retrotransposition of engineered L1

To further validate Tet1 mediated L1 activation, we made use of a cell culture based retrotransposition assay, where the plasmid pLRE3-EGFP [9] was episomally present in HEK-EBNA cells. The L1 cassette of the pLRE3-EGFP construct contains a full-length human L1 element, with a sense oriented γ -globin intron, which interrupts the antisense EGFP cassette in its 3'UTR region. Therefore, EGFP positive cells arise only when the EGFP is transcribed from the L1 promoter, spliced, reverse transcribed and integrated into the genome [9] (Fig S6A). Self-replicating, viral-based vectors are capable of long-term episomal persistence in mammalian cells, in particular in HEK-EBNA cells. Two components are needed for episomal maintenance, the latent origin of replication (oriP) present in the vector and the transactivator protein EBNA-1 stably expressed in HEK-EBNA transgenic cells [55, 56]. The episomal plasmid contains a puromycin resistance gene to facilitate selection of cells containing the episome. HEK-EBNA cells were

transfected with the pLRE3-EGFP construct and treated with 2 μ g/mL puromycin two days later. From three days after transfection, EGFP expression was observed in a small number of cells, indicating that HEK-EBNA cells are suitable to detect engineered L1 retrotransposition from the episome into the genome. Since we observed only a small number of EGFP expressing cells, i.e., where transposition took place, we wanted to test whether the EGFP negative cells still contained the episomal plasmid pLRE3-EGFP. To this end, PCR was performed on whole cell lysates of sorted, EGFP negative cells using primers for EGFP, which amplify both, spliced (genomically integrated copy) and unspliced (episomal copy) EGFP versions. As shown in Figure S7, only unspliced EGFP was amplified from EGFP negative cells, indicating that pLRE3-EGFP is present in these cells. A spliced, genomically integrated EGFP cassette (CMV promoter+EGFP), however, could not be detected indicating that no retrotransposition took place in these cells. After 15 days of antibiotic selection, puromycin-resistant colonies containing few EGFP positive cells formed (Fig S6B). Antibiotic resistance of a large number of EGFP negative cells, however, indicated that the pLRE3-EGFP was episomally present, but its ability to retrotranspose was most likely silenced by DNA methylation (Fig S6B). To test whether Tet1 can activate L1 transposition from the silenced episome in these cells, we transfected the reporter cell line with mcherry-Tet1CD, mcherry-Tet1CDmut or mcherry, fixed them 24 and 48 hours post transfection and further quantified the EGFP positive cells by high content screening microscopy. The ratio of EGFP-positive cells in Tet1/mcherry-positive and -negative cells was used to show L1 retrotransposition events induced by Tet activity (Fig 2A). We observed an increase in the number of EGFP-positive cells upon overexpression of Tet1CD, compared to the overexpression of Tet1CDmut and mcherry alone at 48 hours (Fig 2C).

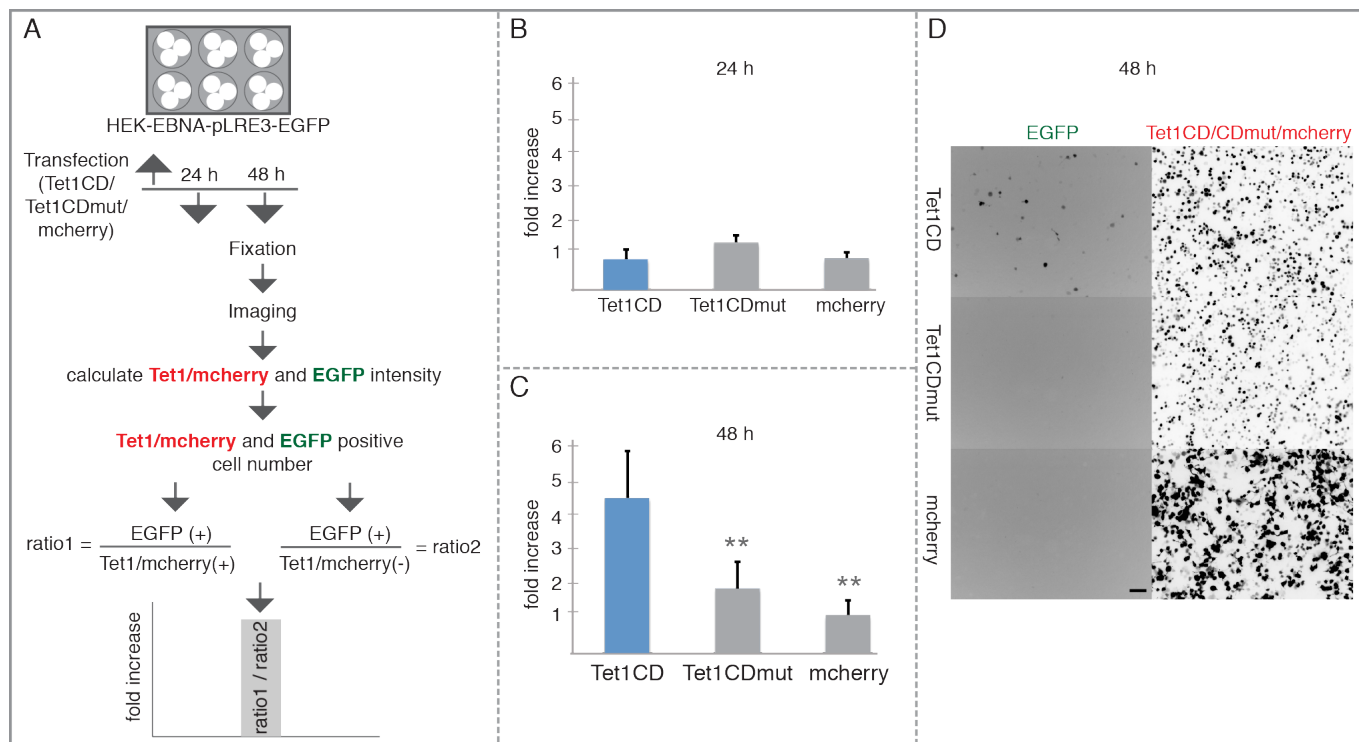


Figure 2: Tet1 reactivates retrotransposition of engineered L1. (A) Experimental rationale. The ratio of EGFP-positive cells in Tet1/mcherry-positive and negative cells, was quantified to detect recent L1 retrotransposition events. (B-C) Relative increases of retrotransposition events (B) 24 hours and (C) 48 hours after Tet1CD, Tet1CDmut and mcherry transfection. At least three independent experiments were performed and more than 250,000 cells for each group were analyzed and the bar represents the mean + SD. Independent two-sample student's t-test was performed between Tet1CD and Tet1CDmut or Tet1CD and mcherry transfected cells. Only significant differences were indicated on the plots as two asterisks ($p < 0.005$). (D) Representative images of the pLRE3-EGFP reporter cell line 48 hours after mcherry-Tet1CD/CDmut and mcherry transfection, respectively. Scale bar: 100 μm .

24 hours after transfection no increase of EGFP positive cells was detected in the presence of Tet1CD (Fig 2B). Consistent with the reporter assay using fixed cells, live-cell time lapse imaging analysis showed a large increase in the amount of EGFP positive cells after 40 hours with Tet1CD transfection, but not with Tet1CDmut and mcherry (Fig S8A, videos 1-3). This fits well the previously reported L1 de novo retrotransposition kinetics [57]. To further validate the increase of L1 retrotransposition induced by Tet1CD, we used flow cytometry analysis. The number of mcherry positive (+) and negative (-), as well as EGFP positive (+) cells was counted by flow cytometry and the ratio between EGFP+/mcherry+

and EGFP+/mcherry- cells was calculated to show recent L1 integrations. The results showed an increase of EGFP positive cells in the presence of Tet1CD 48 hours after transfection (Fig S8B). These data indicate that the L1 promoter of the episomal plasmid was activated by Tet1CD ultimately leading to transposition into the host cell's genome and expression of the EGFP gene. To further verify that the activation of EGFP resulted from L1 retrotransposition and not from silenced genomically integrated copies, we cultured the reporter cell line in the absence of puromycin. Four and 11 days later, the cells were transfected with mcherry-Tet1CD plasmids and two days after transfection the EGFP positive cell

numbers were counted using flow cytometry. As shown in Figure S9, EGFP positive cell numbers were decreased in the absence of puromycin. Previous studies showed that the episomal plasmids are easily lost during cell generations [58]. In the absence of puromycin, the cells without pLRE3-EGFP plasmid are able to survive, but due to the lack of the episome no transposition can take place and, hence, those cells do not express EGFP. These results indicate that the activation of EGFP expression upon Tet expression results from L1 retrotransposition from the episome into the genome rather than from genomically integrated EGFP cassette, since the latter would be activated by Tet1CD in the absence or presence of puromycin. Increase of L1 retrotransposition and the conversion of 5mC to 5hmC induced by Tet1 prompted us to test, whether the observed L1 reactivation arose from DNA methylation changes, so we made use of the cytidine analogue 5-azacytidine (5-aza-C) to induce DNA hypomethylation [59]. The stable cell line was treated with either 5 μ M or 50 μ M of 5-aza-C and 48 hours later the number of EGFP positive cells was quantified by flow cytometry analysis. Compared to untreated cells, the 5-aza-C treatment increased the number of EGFP positive cells in a dose dependent manner (Fig S8C), indicating that decreased DNA methylation is involved in the activation of L1. This effect is in line with results from non-transformed cells that were treated with 5-aza-C [60]. Since 5mC oxidation products such as 5hmC are considered to be intermediates of DNA demethylation, we suggest that the activation of L1 by Tet1 is mainly dependent on its DNA demethylation activity potentially involving subsequent repair processes.

Mecp2 and Mbd2 repress endogenous human L1 retrotransposition

As Mecp2 has been shown to regulate L1 retrotransposition in a methylation dependent manner [19, 20],

we wanted to test whether Tet1 mediated activation of L1 can be counteracted by MBD proteins. To this end, we analyzed the effects of co-expressed Tet1CD, Tet1CDmut, Mecp2 and Tet1CD+Mecp2 on transcription of a luciferase reporter plasmid driven by the internal L1.3 promoter (nucleotides 1-909) in response to its methylation by HpaII methyltransferase [20, 61] (Fig S10A and S10B). While in the absence of any effector protein, methylation of the L1.3 promoter led to a weak transcriptional decrease, overexpression of Tet1CD/Tet1CDmut resulted in similar transcription rates for the methylated and unmethylated promoter (Fig S10C). Ectopic expression of Mecp2, in contrast, reduced transcription from the methylated promoter in the absence and presence of Tet1CD (Fig S10C), indicating that Mecp2 represses Tet1 mediated activation of L1 transcription. Next, we co-expressed Tet1 with Mbd2, Mecp2, or its subdomains MBD and IDTRD, respectively in HEK-EBNA cells and analyzed 5hmC levels at the endogenous L1 promoter by GluMS-qPCR. The results showed decreased 5hmC levels in the 5'UTR of L1 upon co-overexpression of MBDs together with Tet1CD (Fig 3A), indicating that MBDs block Tet mediated 5mC to 5hmC conversion in the 5'UTR of L1. The failure of Tet binding to DNA in the presence of Mbd2, MeCP2, as well as its subdomains MBD and IDTRD [50] might thus be causative for the observed decrease of 5hmC at the L1 5'UTR. Previous studies showed activation of L1 retrotransposition in the absence of Mecp2 [19], and our data indicate that this activation might be due to Tet induced 5hmC formation. In accordance with the decreased 5hmC levels, L1 mRNA levels and copy numbers were lower in cells co-expressing Tet1CD and MBDs (Fig 3B and Fig 3C), indicating that MBDs can prevent L1 transcription and transposition. In summary, the activation of L1 by Tet1 can be prevented by the action of MBD proteins.

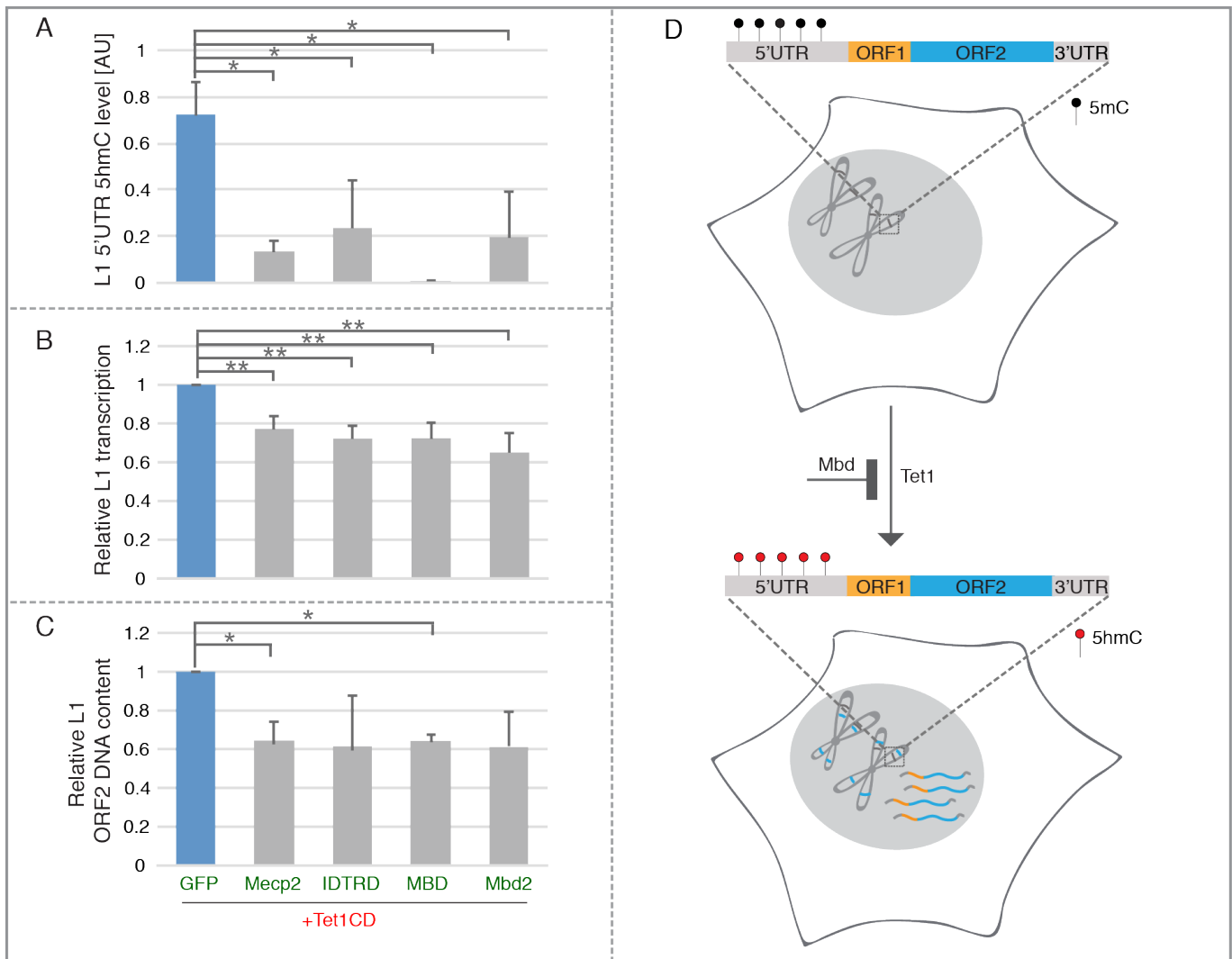


Figure 3: MBD proteins prevent Tet-mediated reactivation of endogenous L1. (A) Relative 5hmC in L1 5'UTR ($n=3$, $*p < 0.05$, independent two-sample student's t-test), (B) relative L1 transcription levels ($n=4$; $**p < 0.01$, independent two-sample student's t-test) and (C) relative L1 ORF2 DNA content were checked 48 hours after cotransfection with plasmids coding for Tet1CD- and MBD proteins ($n=3$, $*p < 0.05$, independent two-sample student's t-test). All of the three independent experiments for IDTRD and Mbd2 showed decreased, but variable L1 copy number when compared to Tet1CD, giving rise to an apparent non-significant difference. (D) Scheme illustrating the effect of Tet1- and MBD proteins on L1 retrotransposition. Black and red circles indicate 5mC and 5hmC nucleotides, respectively. Bars represent the mean + SD.

Here, we show that Tet1 mediated formation of 5hmC activates the L1 retrotransposition. Previous studies have shown that the activation of endogenous L1 transcription does not depend on Tet3 in mouse zygotes [62], but the potential effect of Tet3 on L1 transposition is still unknown. To further check the effect of Tets on L1 transposition, we took advantage of Tet1, Tet2 and Tet3

triple knockout ESCs, which do not express any Tets but their corresponding wild type cells express high levels of Tets, especially Tet1 and Tet2. We showed that the transcription and transposition of mouse L1 is significantly decreased in Tet1, Tet2 and Tet3 triple knockout mouse ESCs as compared to the corresponding wild type cells, indicating that Tets are involved in L1 retrotransposi-

tion activation. Since 5hmC is not only involved in loss of DNA methylation, but also a stable epigenetic mark in mouse embryonic stem cells, we propose that both, loss of DNA methylation and the 5hmC mark itself are involved in L1 retrotransposition activation. Rett syndrome (RTT), a postnatal occurring neurological disorder with an incidence of 1 in 10,000 female births [63], is mostly caused by mutations in the MBD and IDTRD of Mecp2 [27, 64]. The increased L1 retrotransposition events in RTT patients suggest that L1 activity is facilitated upon loss of Mecp2 function in human cells [19], but the mechanism is still unknown. Here, we show that L1 retrotransposition is activated by Tet1 and repressed by MBD proteins. In the absence of Mecp2, Tet proteins oxidize methylated DNA, which is usually bound by Mecp2, leading to L1 element activation. Although previous studies showed that Mecp2 binds to 5hmC enriched within active genes [65], we found, on the other hand, that the MBD of Mecp2 does not bind to 5hmC [66]. Therefore, our data is not consistent with sustained inhibition of L1 activation through binding of Mecp2 to 5hmC. However, Mecp2 binds with high affinity to DNA and this could explain the results of Mellen et al [65]. Previous studies showed that transgenic mice harboring either DNA binding-incompetent MBD or TRD of Mecp2 is sufficient to cause RTT [64]. Here, we show that MBD as well as the IDTRD of Mecp2 are also each sufficient to mediate repression of L1 mobilization. In addition to Mecp2, we show that Mbd2 can also repress human L1 retrotransposition, indicating that the repression of L1 is not Mecp2 specific. Besides 5mC specific binding domains, such as the MBD of Mecp2, the non sequence specific DNA binding domain IDTRD of Mecp2 can also repress the L1 retrotransposition. When compared to *in vitro* methylated L1 5'UTR using HpaII methyltransferase [20], human endogenous L1 5'UTR [60] might provide more binding sites to Mbd2. Thus,

four methylated CpG sites seem to be not enough for the repression of L1 retrotransposition by Mbd2, but it is enough for Mecp2 because of its sequence unspecific strong DNA binding ability [67]. In summary, we show that human L1 can be re-activated by Tet1 proteins and this might lead to decreased genome stability [15] and activation of proto-oncogenes in cancer [68]. Finally, our data indicates that the Tet1 mediated activation of L1 can be repressed by Mecp2 and Mbd2 revealing a role of Mecp2 and Mbd2 as guardians of genome stability by preventing retrotransposition.

References

- [1] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczyk, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Showstee, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx,

- S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11237011>
- [2] S. L. Martin and F. D. Bushman, "Nucleic acid chaperone activity of the orf1 protein from the mouse line-1 retrotransposon," *Mol Cell Biol*, vol. 21, no. 2, pp. 467–75, 2001. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11134335>
- [3] J. Kazazian, H. H. and J. V. Moran, "The impact of L1 retrotransposons on the human genome," *Nat Genet*, vol. 19, no. 1, pp. 19–24, 1998. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9590283>
- [4] A. M. Denli, I. Narvaiza, B. E. Kerman, M. Pena, C. Benner, M. C. Marchetto, J. K. Diedrich, A. Aslanian, J. Ma, J. J. Moresco, L. Moore, T. Hunter, A. Saghatelian, and F. H. Gage, "Primate-specific orf0 contributes to retrotransposon-mediated diversity," *Cell*, vol. 163, no. 3, pp. 583–93, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26496605>
- [5] G. J. Cost, Q. Feng, A. Jacquier, and J. D. Boeke, "Human L1 element target-primed reverse transcription in vitro," *EMBO J*, vol. 21, no. 21, pp. 5899–910, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12411507>
- [6] T. A. Morrish, N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato, G. E. Taccioli, M. A. Batzer, and J. V. Moran, "Dna repair mediated by endonuclease-independent line-1 retrotransposition," *Nat Genet*, vol. 31, no. 2, pp. 159–65, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12006980>
- [7] N. G. Coufal, J. L. Garcia-Perez, G. E. Peng, M. C. Marchetto, A. R. Muotri, Y. Mu, C. T. Carson, A. Marcia, J. V. Moran, and F. H. Gage, "Ataxia telangiectasia mutated (atm) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells," *Proc Natl Acad Sci U S A*, vol. 108, no. 51, pp. 20382–7, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22159035>
- [8] A. R. Muotri, V. T. Chu, M. C. Marchetto, W. Deng, J. V. Moran, and F. H. Gage, "Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition," *Nature*, vol. 435, no. 7044, pp. 903–10, 2005. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15959507>
- [9] N. G. Coufal, J. L. Garcia-Perez, G. E. Peng, G. W. Yeo, Y. Mu, M. T. Lovci, M. Morell, K. S. O'Shea, J. V. Moran, and F. H. Gage, "L1 retrotransposition in human neural progenitor cells," *Nature*, vol. 460, no. 7259, pp. 1127–31, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19657334>
- [10] J. L. Garcia-Perez, M. C. Marchetto, A. R. Muotri, N. G. Coufal, F. H. Gage, K. S. O'Shea, and J. V. Moran, "Line-1 retrotransposition in human embryonic stem cells," *Hum Mol Genet*, vol. 16, no. 13, pp. 1569–77, 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17468180>
- [11] E. M. Wolff, H. M. Byun, H. F. Han, S. Sharma, P. W. Nichols, K. D. Siegmund, A. S. Yang, P. A. Jones, and G. Liang, "Hypomethylation of a line-1 promoter activates an alternate transcript of the met oncogene in bladders with cancer," *PLoS Genet*, vol. 6, no. 4, p. e1000917, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20421991>

- [12] A. V. Horn, S. Klawitter, U. Held, A. Berger, A. A. Vasudevan, A. Bock, H. Hofmann, K. M. Hanschmann, J. H. Trosemeier, E. Flory, R. A. Jabulowsky, J. S. Han, J. Lower, R. Lower, C. Munk, and G. G. Schumann, "Human line-1 restriction by apobec3c is deaminase independent and mediated by an orf1p interaction that affects line reverse transcriptase activity," *Nucleic Acids Res*, vol. 42, no. 1, pp. 396–416, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24101588>
- [13] G. Meister, "Argonaute proteins: functional insights and emerging roles," *Nat Rev Genet*, vol. 14, no. 7, pp. 447–59, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23732335>
- [14] P. R. Cook, C. E. Jones, and A. V. Furano, "Phosphorylation of orf1p is required for l1 retrotransposition," *Proc Natl Acad Sci U S A*, vol. 112, no. 14, pp. 4298–303, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25831499>
- [15] D. E. Symer, C. Connelly, S. T. Szak, E. M. Caputo, G. J. Cost, G. Parmigiani, and J. D. Boeke, "Human l1 retrotransposition is associated with genetic instability in vivo," *Cell*, vol. 110, no. 3, pp. 327–38, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12176320>
- [16] J. S. Han, S. T. Szak, and J. D. Boeke, "Transcriptional disruption by the l1 retrotransposon and implications for mammalian transcriptomes," *Nature*, vol. 429, no. 6989, pp. 268–74, 2004. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15152245>
- [17] J. M. Tubio, Y. Li, Y. S. Ju, I. Martincorena, S. L. Cooke, M. Tojo, G. Gundem, C. P. Pipinikas, J. Zamora, K. Raine, A. Menzies, P. Roman-Garcia, A. Fullam, M. Gerstung, A. Shlien, P. S. Tarpey, E. Papaemmanuil, S. Knappskog, P. Van Loo, M. Ramakrishna, H. R. Davies, J. Marshall, D. C. Wedge, J. W. Teague, A. P. Butler, S. Nik-Zainal, L. Alexandrov, S. Behjati, L. R. Yates, N. Bolli, L. Mudie, C. Hardy, S. Martin, S. McLaren, S. O'Meara, E. Anderson, M. Maddison, S. Gamble, I. B. C. Group, I. B. C. Group, I. P. C. Group, C. Foster, A. Y. Warren, H. Whitaker, D. Brewer, R. Eeles, C. Cooper, D. Neal, A. G. Lynch, T. Visakorpi, W. B. Isaacs, L. van't Veer, C. Caldas, C. Desmedt, C. Sotiriou, S. Aparicio, J. A. Foekens, J. E. Eyfjord, S. R. Lakhani, G. Thomas, O. Myklebost, P. N. Span, A. L. Borresen-Dale, A. L. Richardson, M. Van de Vijver, A. Vincent-Salomon, G. G. Van den Eynden, A. M. Flanagan, P. A. Futreal, S. M. Janes, G. S. Bova, M. R. Stratton, U. McDermott, and P. J. Campbell, "Mobile dna in cancer. extensive transduction of nonrepetitive dna mediated by l1 retrotransposition in cancer genomes," *Science*, vol. 345, no. 6196, p. 1251343, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25082706>
- [18] S. L. Gasior, T. P. Wakeman, B. Xu, and P. L. Deininger, "The human line-1 retrotransposon creates dna double-strand breaks," *J Mol Biol*, vol. 357, no. 5, pp. 1383–93, 2006. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16490214>
- [19] A. R. Muotri, M. C. Marchetto, N. G. Co-ufal, R. Oefner, G. Yeo, K. Nakashima, and F. H. Gage, "L1 retrotransposition in neurons is modulated by mecp2," *Nature*, vol. 468, no. 7322, pp. 443–6, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21085180>
- [20] F. Yu, N. Zingler, G. Schumann, and W. H. Stratling, "Methyl-cpg-binding protein 2 represses line-1 expression and retrotransposition but not alu transcription," *Nucleic Acids Res*, vol. 29,

- no. 21, pp. 4493–501, 2001. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11691937>
- [21] B. Hendrich and A. Bird, “Identification and characterization of a family of mammalian methyl-cpg binding proteins,” *Mol Cell Biol*, vol. 18, no. 11, pp. 6538–47, 1998. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9774669>
- [22] R. M. Kohli and Y. Zhang, “Tet enzymes, tdg and the dynamics of dna demethylation,” *Nature*, vol. 502, no. 7472, pp. 472–9, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24153300>
- [23] G. Ficiz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, and W. Reik, “Dynamic regulation of 5-hydroxymethylcytosine in mouse es cells and during differentiation,” *Nature*, vol. 473, no. 7347, pp. 398–402, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21460836>
- [24] A. Becker, L. Allmann, M. Hofstatter, V. Casa, P. Weber, A. Lehmkühl, H. D. Herce, and M. C. Cardoso, “Direct homo- and hetero-interactions of mecp2 and mbd2,” *PLoS One*, vol. 8, no. 1, p. e53730, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23335972>
- [25] A. Brero, H. P. Easwaran, D. Nowak, I. Grunewald, T. Cremer, H. Leonhardt, and M. C. Cardoso, “Methyl cpg-binding proteins induce large-scale chromatin reorganization during terminal differentiation,” *J Cell Biol*, vol. 169, no. 5, pp. 733–43, 2005. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15939760>
- [26] K. L. Jost, A. Rottach, M. Mildner, B. Bertulat, A. Becker, P. Wolf, J. Sandoval, P. Petazzi, D. Huertas, M. Esteller, E. Kremmer, H. Leonhardt, and M. C. Cardoso, “Generation and characterization of rat and mouse monoclonal antibodies specific for mecp2 and their use in x-inactivation studies,” *PLoS One*, vol. 6, no. 11, p. e26499, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22140431>
- [27] N. Agarwal, T. Hardt, A. Brero, D. Nowak, U. Rothbauer, A. Becker, H. Leonhardt, and M. C. Cardoso, “Mecp2 interacts with hp1 and modulates its heterochromatin association during myogenic differentiation,” *Nucleic Acids Res*, vol. 35, no. 16, pp. 5402–8, 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17698499>
- [28] D. Yaffe and O. Saxel, “Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle,” *Nature*, vol. 270, no. 5639, pp. 725–7, 1977. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/563524>
- [29] M. C. Cardoso, H. Leonhardt, and B. Nadal-Ginard, “Reversal of terminal differentiation and control of dna replication: cyclin a and cdk2 specifically localize at subnuclear sites of dna replication,” *Cell*, vol. 74, no. 6, pp. 979–92, 1993. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8402887>
- [30] A. Becker, P. Zhang, L. Allmann, D. Meilinger, B. Bertulat, D. Eck, M. Hofstaetter, G. Bartolomei, M. O. Hottiger, V. Schreiber, H. Leonhardt, and M. C. Cardoso, “Poly(adp-ribosyl)ation of methyl cpg binding domain protein 2 regulates chromatin structure,” *J Biol Chem*, vol. 291, no. 10, pp. 4873–81, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26772194>
- [31] J. Guy, B. Hendrich, M. Holmes, J. E. Martin, and A. Bird, “A mouse mecp2-null mutation causes neurological symptoms that

- mimic rett syndrome," *Nat Genet*, vol. 27, no. 3, pp. 322–6, 2001. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11242117>
- [32] I. Martin Caballero, J. Hansen, D. Leaford, S. Pollard, and B. D. Hendrich, "The methyl-cpg binding proteins mecp2, mbd2 and kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation," *PLoS One*, vol. 4, no. 1, p. e4315, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19177165>
- [33] M. M. Dawlaty, A. Breiling, T. Le, M. I. Barrasa, G. Raddatz, Q. Gao, B. E. Powell, A. W. Cheng, K. F. Faull, F. Lyko, and R. Jaenisch, "Loss of tet enzymes compromises proper differentiation of embryonic stem cells," *Dev Cell*, vol. 29, no. 1, pp. 102–11, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24735881>
- [34] S. Picelli, A. K. Bjorklund, O. R. Faridani, S. Sagasser, G. Winberg, and R. Sandberg, "Smart-seq2 for sensitive full-length transcriptome profiling in single cells," *Nat Methods*, vol. 10, no. 11, pp. 1096–8, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24056875>
- [35] Y. Jin, O. H. Tam, E. Paniagua, and M. Hammell, "Tetrascripts: a package for including transposable elements in differential expression analysis of rna-seq datasets," *Bioinformatics*, vol. 31, no. 22, pp. 3593–9, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26206304>
- [36] S. M. Aljanabi and I. Martinez, "Universal and rapid salt-extraction of high quality genomic dna for pcr-based techniques," *Nucleic Acids Res*, vol. 25, no. 22, pp. 4692–3, 1997. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9358185>
- [37] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, and H. Leonhardt, "Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic dna," *Nucleic Acids Res*, vol. 38, no. 19, p. e181, 2010. [Online]. Available: <http://nar.oxfordjournals.org/content/38/19/e181.full.pdf>
- [38] T. Davis and R. Vaisvila, "High sensitivity 5-hydroxymethylcytosine detection in balb/c brain tissue," *J Vis Exp*, no. 48, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21307836>
- [39] L. A. Clarke, L. Sousa, C. Barreto, and M. D. Amaral, "Changes in transcriptome of native nasal epithelium expressing f508del-cftr and intersecting data from comparable studies," *Respir Res*, vol. 14, p. 38, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23537407>
- [40] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative pcr and the 2(-delta delta c(t)) method," *Methods*, vol. 25, no. 4, pp. 402–8, 2001. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11846609>
- [41] L. Schermelleh, P. M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M. C. Cardoso, D. A. Agard, M. G. Gustafsson, H. Leonhardt, and J. W. Sedat, "Subdiffraction multicolor imaging of the nuclear periphery with 3d structured illumination microscopy," *Science*, vol. 320, no. 5881, pp. 1332–6, 2008. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18535242>
- [42] J. Raiz, A. Damert, S. Chira, U. Held, S. Klawitter, M. Hamdorf, J. Lower, W. H. Stratling, R. Lower, and G. G. Schumann, "The non-autonomous retrotransposon sva is trans-mobilized by the human line-1 protein machinery," *Nucleic Acids Res*, vol. 40,

- no. 4, pp. 1666–83, 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22053090>
- [43] S. Klawitter, N. V. Fuchs, K. R. Upton, M. Munoz-Lopez, R. Shukla, J. Wang, M. Garcia-Canadas, C. Lopez-Ruiz, D. J. Gerhardt, A. Sebe, I. Grabundzija, S. Merkert, P. Gerdes, J. A. Pulgarin, A. Bock, U. Held, A. Witthuhn, A. Haase, B. Sarkadi, J. Lower, E. J. Wolvetang, U. Martin, Z. Ivics, Z. Izsvak, J. L. Garcia-Perez, G. J. Faulkner, and G. G. Schumann, “Reprogramming triggers endogenous I1 and alu retrotransposition in human induced pluripotent stem cells,” *Nat Commun*, vol. 7, p. 10286, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26743714>
- [44] S. Kriaucionis and N. Heintz, “The nuclear dna base 5-hydroxymethylcytosine is present in purkinje neurons and the brain,” *Science*, vol. 324, no. 5929, pp. 929–30, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19372393>
- [45] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao, “Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian dna by mll partner tet1,” *Science*, vol. 324, no. 5929, pp. 930–5, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19372391>
- [46] L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, Y. Wang, J. Xie, Y. Zhang, C. Song, M. Yu, X. Liu, P. Zhu, X. Li, Y. Hou, H. Guo, X. Wu, C. He, R. Li, F. Tang, and J. Qiao, “Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain,” *Genome Biol*, vol. 15, no. 3, p. R49, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24594098>
- [47] R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghghi, T. J. Sejnowski, M. M. Behrens, and J. R. Ecker, “Global epigenomic reconfiguration during mammalian brain development,” *Science*, vol. 341, no. 6146, p. 1237905, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23828890>
- [48] U. Muller, C. Bauer, M. Siegl, A. Rottach, and H. Leonhardt, “Tet-mediated oxidation of methylcytosine causes tdg or neil glycosylase dependent gene reactivation,” *Nucleic Acids Res*, vol. 42, no. 13, pp. 8592–604, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24948610>
- [49] H. Chen, H. G. Kazemier, M. L. de Groote, M. H. Ruiters, G. L. Xu, and M. G. Rots, “Induced dna demethylation by targeting ten-eleven translocation 2 to the human icam-1 promoter,” *Nucleic Acids Res*, vol. 42, no. 3, pp. 1563–74, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24194590>
- [50] A. K. Ludwig, P. Zhang, F. D. Hastert, S. Meyer, C. Rausch, H. D. Herce, U. Muller, A. Lehmkuhl, I. Hellmann, C. Trummer, C. Storm, H. Leonhardt, and M. C. Cardoso, “Binding of mbd proteins to dna blocks tet1 function thereby modulating transcriptional noise,” *Nucleic Acids Res*, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/27923996>
- [51] K. P. Nishimoto, D. Newkirk, S. Hou, J. Fruehauf, and E. L. Nelson, “Fluorescence activated cell sorting (facs) using rnalater to minimize rna degradation and perturbation of mrna expression from cells involved in initial host mi-

- crobe interactions," *J Microbiol Methods*, vol. 70, no. 1, pp. 205–8, 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17512621>
- [52] J. L. Garcia-Perez, M. Morell, J. O. Scheys, D. A. Kulpa, S. Morell, C. C. Carter, G. D. Hammer, K. L. Collins, K. S. O'Shea, P. Menendez, and J. V. Moran, "Epigenetic silencing of engineered l1 retrotransposition events in human embryonic carcinoma cells," *Nature*, vol. 466, no. 7307, pp. 769–73, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20686575>
- [53] M. Van Meter, M. Kashyap, S. Rezazadeh, A. J. Geneva, T. D. Morello, A. Seluanov, and V. Gorbunova, "Sirt6 represses line1 retrotransposons by ribosylating kap1 but this repression fails with stress and age," *Nat Commun*, vol. 5, p. 5011, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25247314>
- [54] E. A. Farkash, G. D. Kao, S. R. Horman, and E. T. Prak, "Gamma radiation increases endonuclease-dependent l1 retrotransposition in a cultured cell assay," *Nucleic Acids Res*, vol. 34, no. 4, pp. 1196–204, 2006. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16507671>
- [55] J. Black and J. M. Vos, "Establishment of an orip/ebna1-based episomal vector transcribing human genomic beta-globin in cultured murine fibroblasts," *Gene Ther*, vol. 9, no. 21, pp. 1447–54, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12378407>
- [56] Y. Durocher, S. Perret, and A. Kamen, "High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-ebna1 cells," *Nucleic Acids Res*, vol. 30, no. 2, p. E9, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11788735>
- [57] E. M. Ostertag, E. T. Prak, R. J. DeBerardinis, J. V. Moran, and J. Kazazian, H. H., "Determination of l1 retrotransposition kinetics in cultured cells," *Nucleic Acids Res*, vol. 28, no. 6, pp. 1418–23, 2000. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10684937>
- [58] A. Nanbo, A. Sugden, and B. Sugden, "The coupling of synthesis and partitioning of ebv's plasmid replicon is revealed in live cells," *EMBO J*, vol. 26, no. 19, pp. 4252–62, 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17853891>
- [59] L. Schermelleh, F. Spada, H. P. Easwaran, K. Zolghadr, J. B. Margot, M. C. Cardoso, and H. Leonhardt, "Trapped in action: direct visualization of dna methyltransferase activity in living cells," *Nat Methods*, vol. 2, no. 10, pp. 751–6, 2005. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16179921>
- [60] D. M. Woodcock, C. B. Lawler, M. E. Linsenmeyer, J. P. Doherty, and W. D. Warren, "Asymmetric methylation in the hypermethylated cpg promoter region of the human l1 retrotransposon," *J Biol Chem*, vol. 272, no. 12, pp. 7810–6, 1997. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9065445>
- [61] R. E. Thayer, M. F. Singer, and T. G. Fanning, "Undermethylation of specific line-1 sequences in human cells producing a line-1-encoded protein," *Gene*, vol. 133, no. 2, pp. 273–7, 1993. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7693554>
- [62] A. Inoue, S. Matoba, and Y. Zhang, "Transcriptional activation of transposable elements in

- mouse zygotes is independent of tet3-mediated 5-methylcytosine oxidation," *Cell Res*, vol. 22, no. 12, pp. 1640–9, 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23184059>
- [63] R. E. Amir, I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, and H. Y. Zoghbi, "Rett syndrome is caused by mutations in x-linked mecp2, encoding methyl-cpg-binding protein 2," *Nat Genet*, vol. 23, no. 2, pp. 185–8, 1999. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10508514>
- [64] K. Brown, J. Selfridge, S. Lagger, J. Connelly, D. De Sousa, A. Kerr, S. Webb, J. Guy, C. Merusi, M. V. Koerner, and A. Bird, "The molecular basis of variable phenotypic severity among common missense mutations causing rett syndrome," *Hum Mol Genet*, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26647311>
- [65] M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis, and N. Heintz, "Mecp2 binds to 5hmc enriched within active genes and accessible chromatin in the nervous system," *Cell*, vol. 151, no. 7, pp. 1417–30, 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23260135>
- [66] C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes, and H. Leonhardt, "Recognition of 5-hydroxymethylcytosine by the uhrf1 sra domain," *PLoS One*, vol. 6, no. 6, p. e21306, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21731699>
- [67] J. C. Hansen, R. P. Ghosh, and C. L. Woodcock, "Binding of the rett syndrome protein, mecp2, to methylated and unmethylated dna and chromatin," *IUBMB Life*, vol. 62, no. 10, pp. 732–8, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21031501>
- [68] K. Hur, P. Cejas, J. Feliu, J. Moreno-Rubio, E. Burgos, C. R. Boland, and A. Goel, "Hypomethylation of long interspersed nuclear element-1 (line-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis," *Gut*, vol. 63, no. 4, pp. 635–46, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23704319>

Supplementary Figures

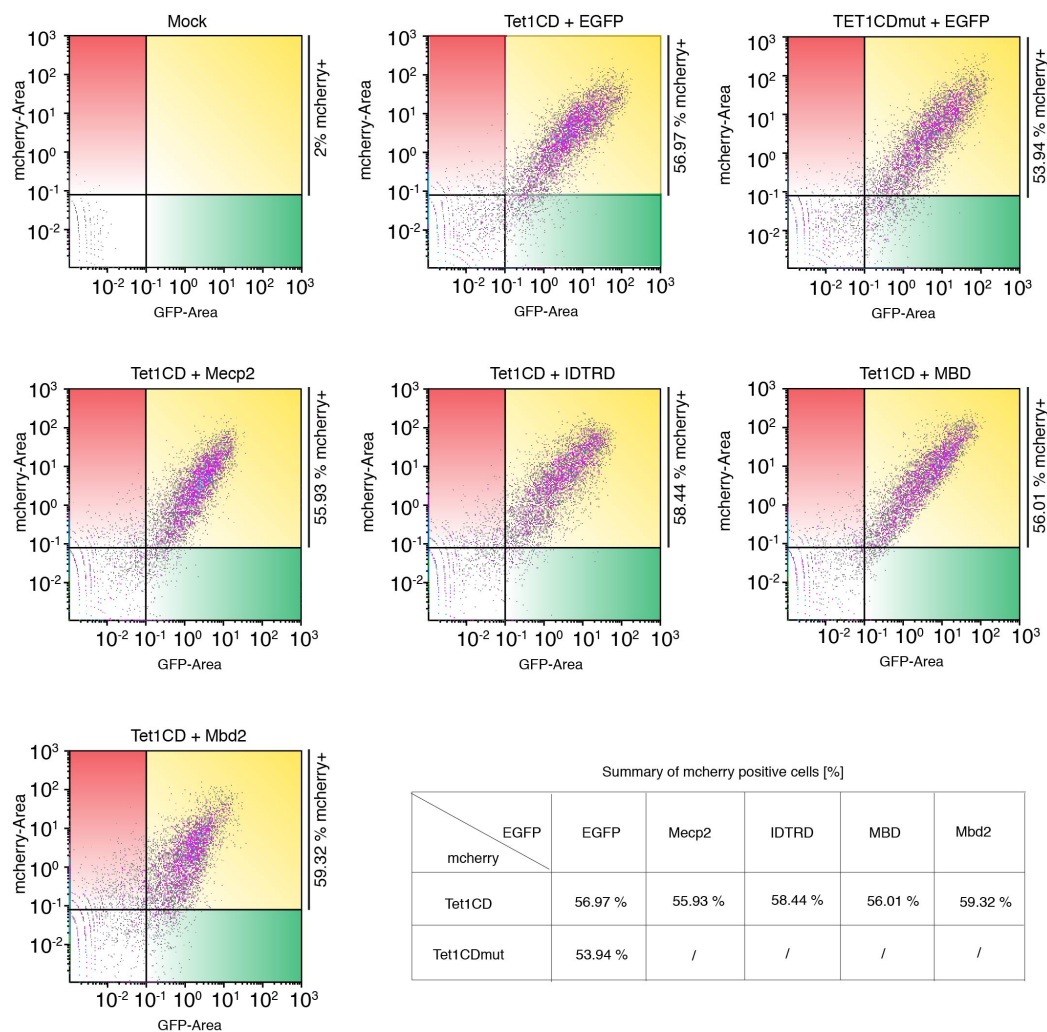


Figure S1 Flow cytometry analysis of cells ectopically coexpressing mcherry-Tet1CD/CDmut and GFP-tagged MBD proteins. Shown are cell populations expressing similar mcherry-Tet1CD/CDmut levels that were used for RNA preparation.

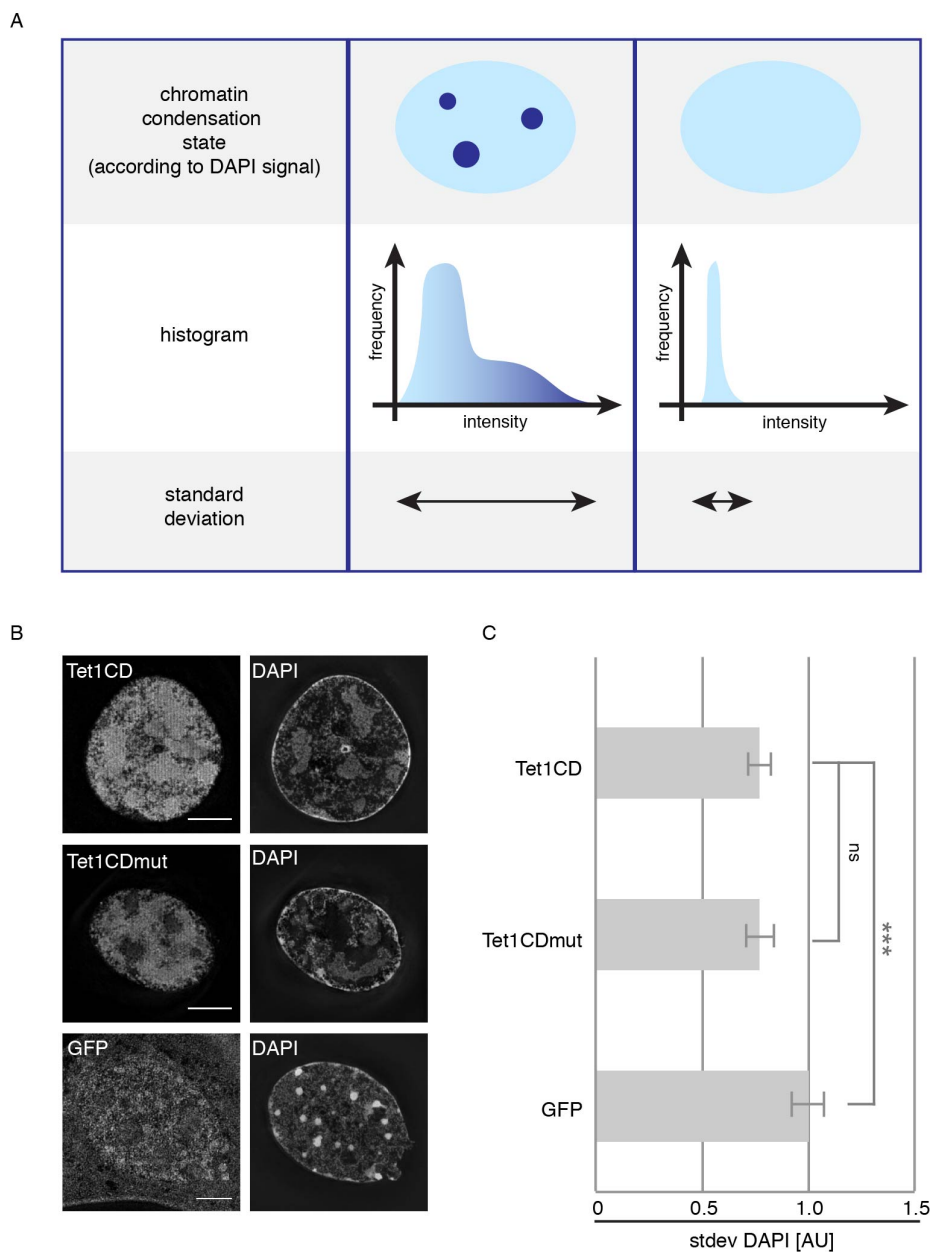


Figure S2 Effect of Tet proteins on chromatin structure. (A) Quantification of chromatin condensation states. On a histogram, condensed chromatin (DAPI stained) is characterized by high DAPI intensities. Accordingly, cell nuclei containing condensed chromatin have a higher DAPI standard deviation than cells containing decondensed chromatin, which in turn is characterized by low DAPI intensities. (B) 3D structured illumination microscopy images of fixed and DAPI stained mouse C2C12 myoblasts 24 hours post transfection with GFP-Tet1CD, GFP-Tet1CDmut and GFP. Scale bar 5 μ m. (C) DAPI standard deviation values of mouse myoblasts 24 hours post transfection with GFP-Tet1CD, GFP-Tet1CDmut and GFP as negative control. Shown are mean values of two independent experiments and the corresponding 95% confidence interval. Independent two-sample student's *t*-test was performed (***: $p < 0.001$, ns: non significant).

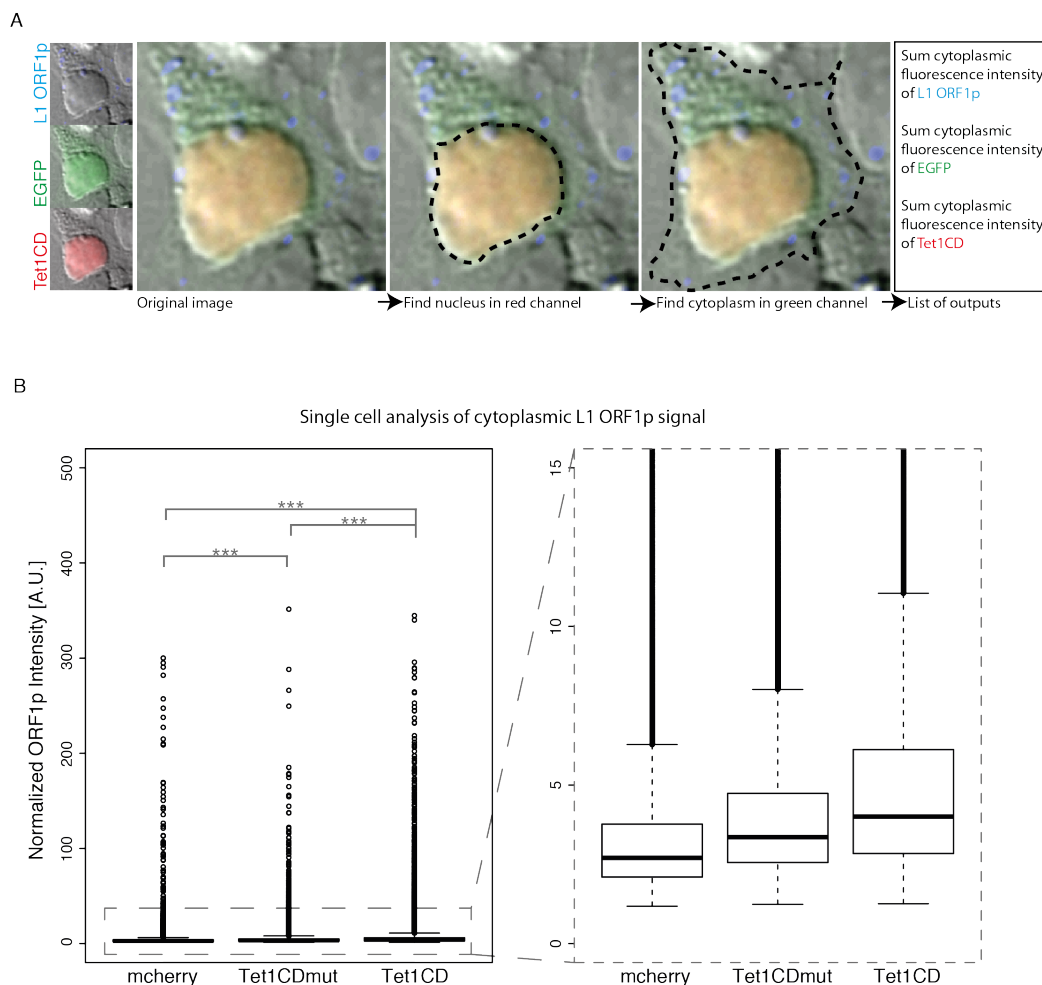


Figure S3 Tet1 increases L1 ORF1p levels. (A) Schematic overview of the automated analysis procedure performed with the Harmony 3.5.1 software (PerkinElmer). Shown are representative confocal images of HEK293T cells cotransfected with EGFP and mcherry-Tet1CD and immunostained for L1 ORF1p. For calculation of the sum fluorescence EGFP, mcherry-Tet1CD and L1 ORF1p intensity, the nucleoplasm was identified according to the mcherry-Tet1CD signal and the cytoplasm was identified (in mcherry-Tet1CD positive cells) according to the EGFP signal. Subsequently, the total fluorescence intensity of the nucleoplasm was subtracted from the total fluorescence intensity of the cytoplasm (= sum cytoplasmic fluorescence intensity). Same analysis was performed for HEK293T cells coexpressing EGFP and mcherry-Tet1CDmut, or mcherry. (B) Boxplots of three independent biological repeats with at least 45000 cells per condition. Tet1CD overexpression increases L1 ORF1p levels significantly more than ectopic expression of Tet1-CDmut or mcherry. Independent two-sample student's *t*-test was performed (***: $p < 0.001$).

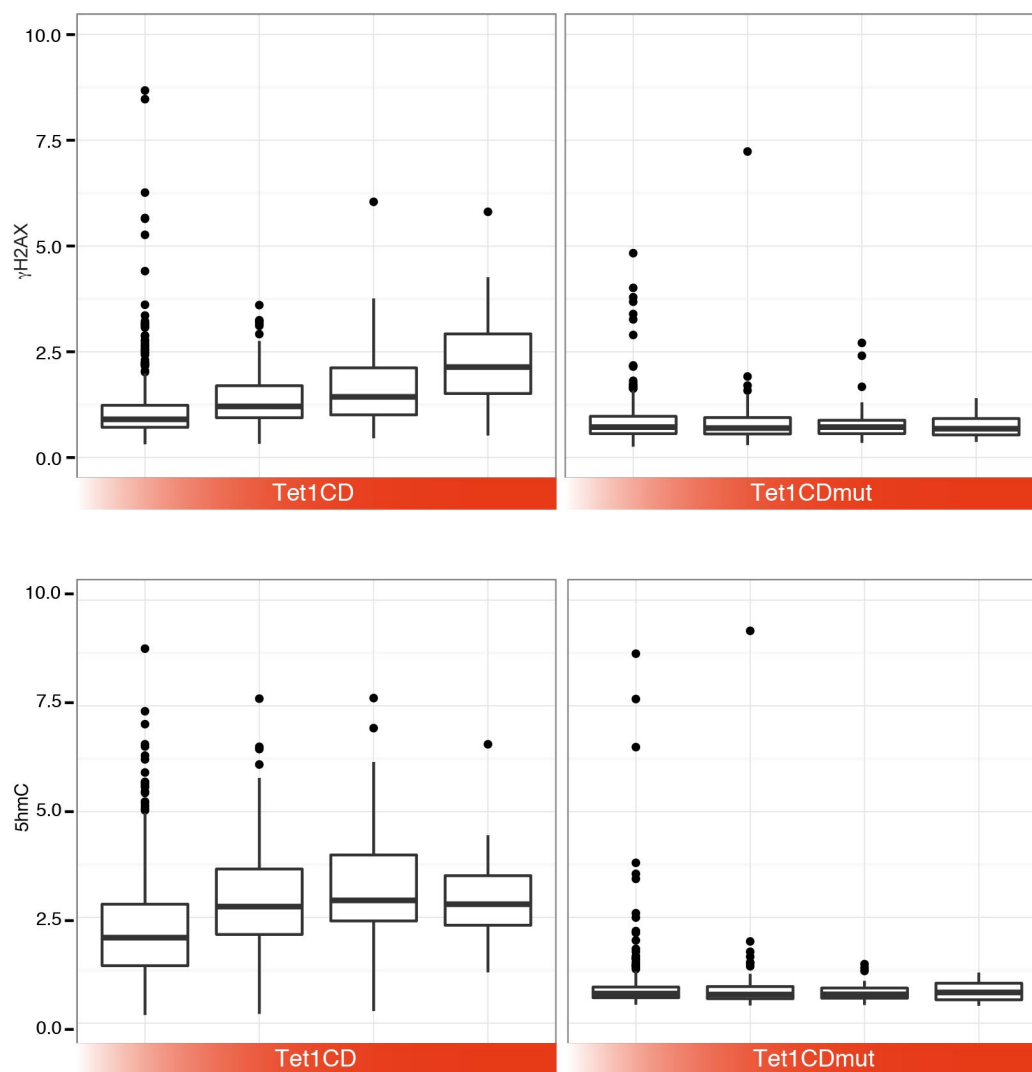


Figure S4 Catalytic activity of Tet1 induces γ H2AX formation. The human AG cells were transfected with mcherry tagged Tet1CD or Tet1CDmut and 24 hours after transfection, the cells were used for immunostaining with γ H2AX and 5hmC antibody. The results showed that 5hmC levels are increased with increase expression of Tet1CD but not Tet1CDmut. Accordingly, the γ H2AX signal is also increased with Tet1CD but not Tet1CDmut. Two independent experiments were performed.

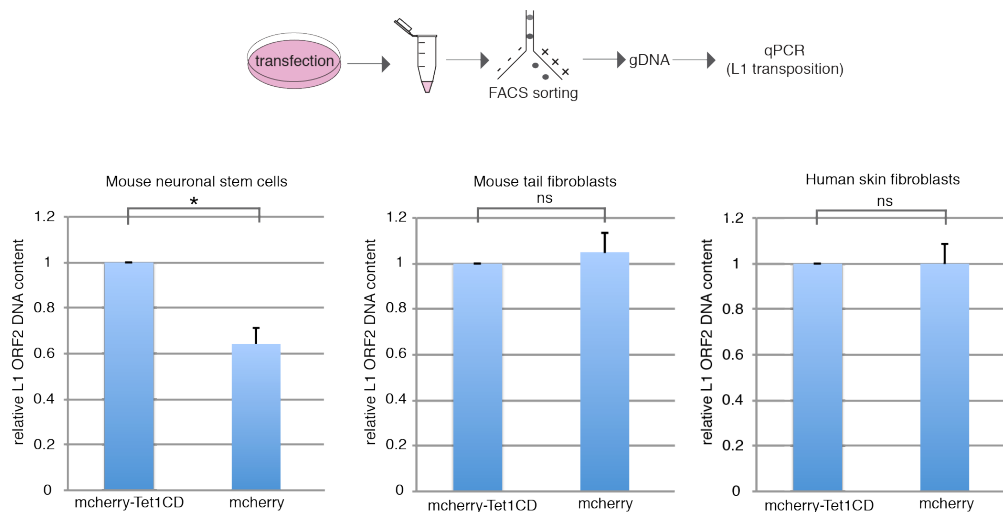


Figure S5 Tet proteins activate L1 retrotransposition in mouse neural stem cells but not mouse and human fibroblast cells. Mouse neural stem cells, mouse tail fibroblast and human skin fibroblast cells were transfected with mcherry-Tet1CD or mcherry. 48 hours after transfection, mcherry positive cells were separated from untransfected cells using flow cytometry. Genomic L1 ORF2 content was checked using quantitative PCR. The relative amount of ORF2 to 5'UTR and ORF2 to 5S RNA was used to show the copy number changes for human and mouse cells, respectively. The results showed that Tet1 proteins increase genomic L1 ORF2 content in mouse neural stem cells, but not in mouse and human fibroblast cells. Shown are mean values of at least three independent experiments and the corresponding standard deviation. Independent two-sample student's *t*-test was performed (* $p < 0.05$, ns: non significant).

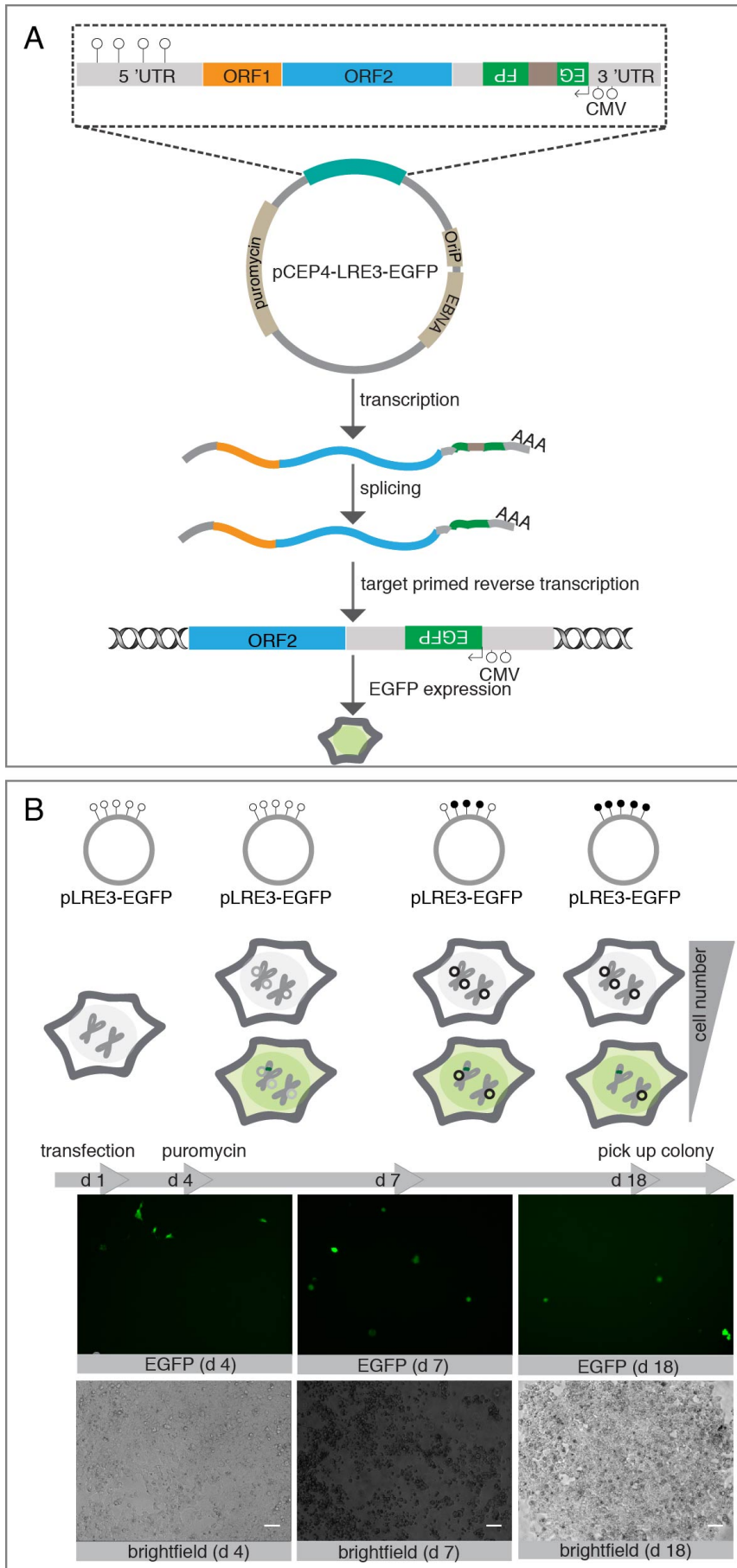


Figure S6 Generation of reporter cell line expressing pLRE3-EGFP (A) The human L1 retrotransposition cassette of the pLRE3-EGFP plasmid. (B) The process of reporter cell generation. The HEK-EBNA cells were transfected with pLRE3-EGFP plasmid. Two days after transfection, cells were treated with 2 µg/mL puromycin for resistance selection. Lower row shows the representative images of EGFP expression and corresponding bright field. Scale bar: 100 µm. Upper row shows the hypothesis that the episomal L1 promoter gets methylated during persisting cultivation. Cartoon cells illustrate expression of EGFP and episomal presence of pLRE3-EGFP. Grey and black circles attached to chromosomes represent unmethylated and methylated pLRE3-EGFP plasmid, respectively. Green dots on chromosomes represent the integrated EGFP cassette, which includes the CMV promoter and the EGFP coding sequence.

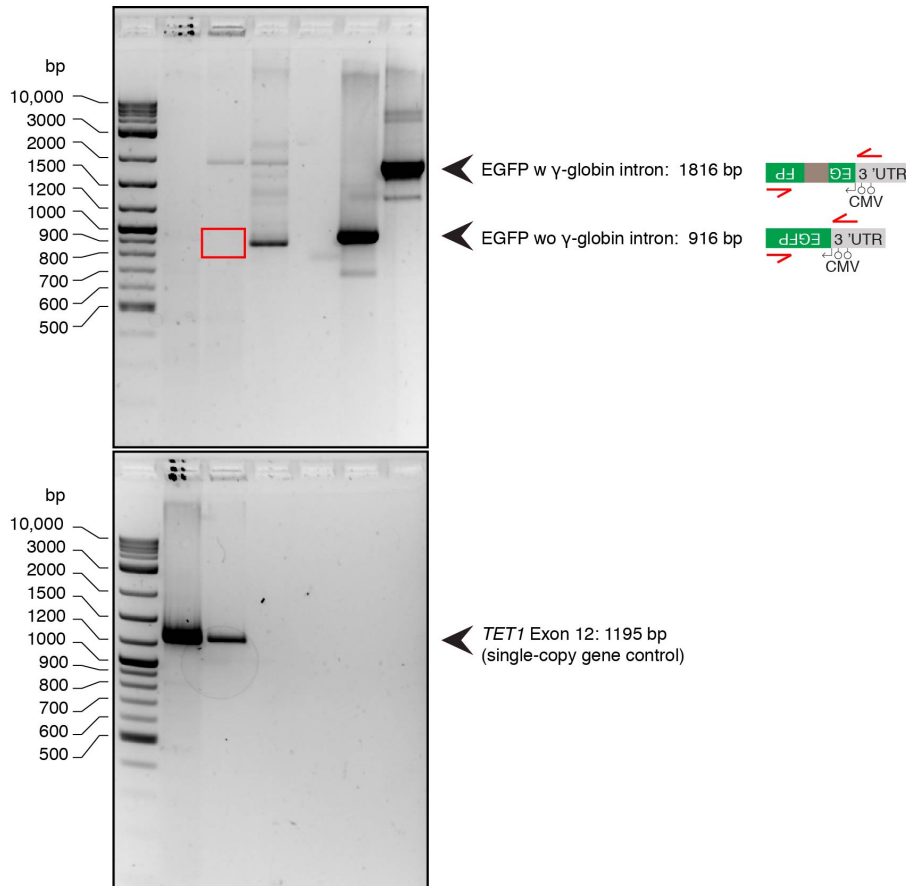
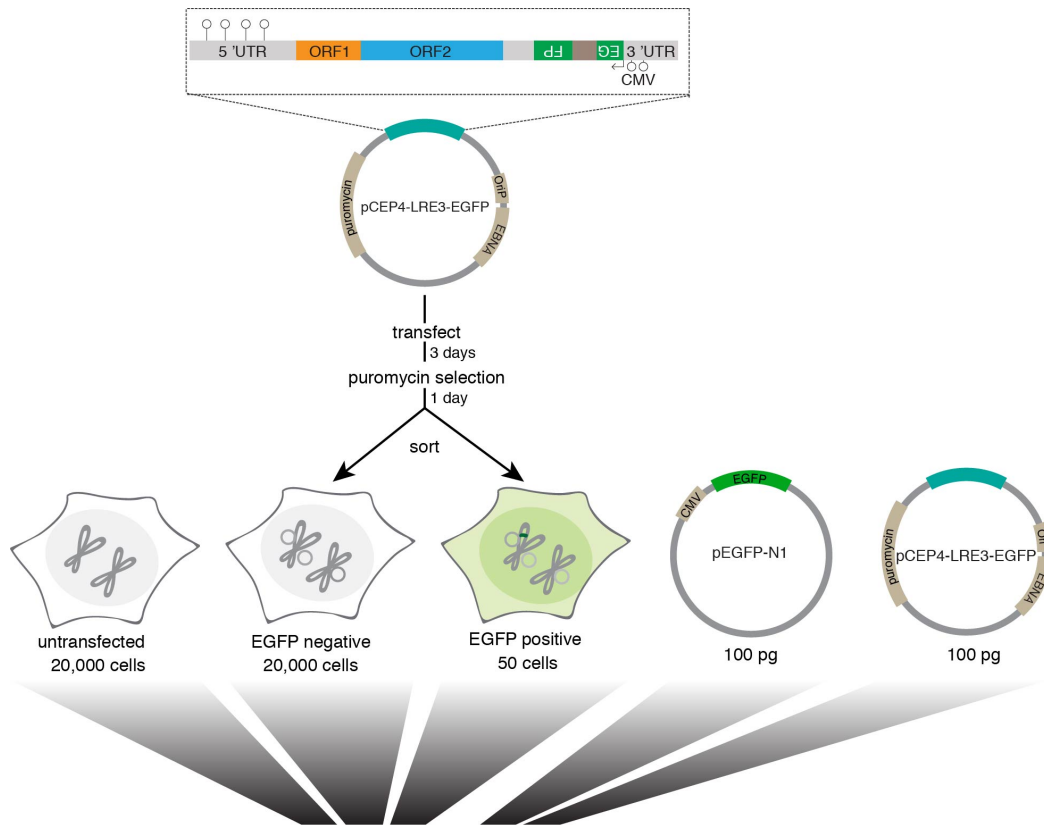


Figure S7 PCR analysis to validate the presence of pCEP4-LRE3-EGFP (EGFP with γ -globin intron) and the absence of genomically integrated LRE3-EGFP (EGFP without γ -globin intron) in the retrotransposition reporter cell line. Three days after transfection with pCEP4-LRE3-EGFP, HEK-EBNA cells were treated with 2 μ g/mL puromycin. One day after puromycin selection, EGFP negative cells were separated from EGFP positive cells using flow cytometry. EGFP negative, EGFP positive and untransfected (negative control) cells were boiled for 20 minutes at 99°C in ddH₂O. Cell lysates containing 20,000 cells were used as template for the amplification of EGFP (CMV fwd: 5' CGTGGATAGCGGTTTGAC 3', EGFP rev: 5' TCTTTGCTCAGGGCGGACTG 3') to determine if the globin intron had been removed by splicing during the retrotransposition process. As positive controls, pEGFP-N1 (EGFP without γ -globin intron) and pCEP4-LRE3-EGFP (EGFP with γ -globin intron) were used. To control for the presence of genomic DNA, we amplified exon 12 of human *TET1* (*TET1* fwd: 5' GAACTCGAGCCAACCAACACAACATCAGC 3', *TET1* rev: 5' GAATTCGAAAATACACTGCACTAGCAAAGG 3'). The spliced EGFP variant could only be detected in EGFP positive, but not in EGFP negative cells (red square).

retrotransposition events in pLRE3-EGFP containing, 5-aza-C treated cells by flow cytometry. Two individual experiments were performed and the bar represents the mean + SD. At least 2,500 cells were analyzed in each time point.

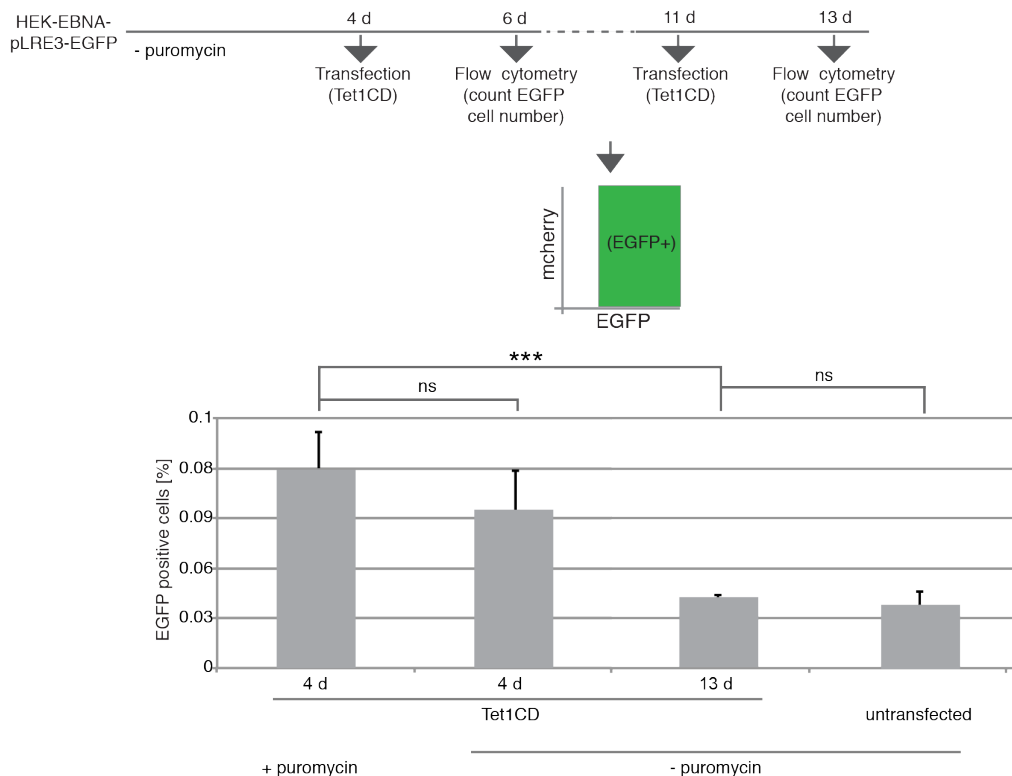


Figure S9 EGFP expression results from *de novo* retrotransposed L1 elements and not from genomically silenced previously integrated EGFP in the reporter cell line. The reporter cell line was cultured in the absence or presence of 2 $\mu\text{g}/\text{uL}$ puromycin. Four and 11 days later, the cells were transfected with mcherry-Tet1CD plasmids and two days after mcherry-Tet1CD transfection, the number of EGFP positive (+) cells was counted by flow cytometry and the percentage of EGFP positive cells was used to determine the quantity of recent L1 integrations. At least three independent experiments were performed and each experiment around 100,000 cells were analyzed. The bar represents the mean + SD. Independent two-sample student's *t*-test was performed (***: $p < 0.001$, ns: non significant).

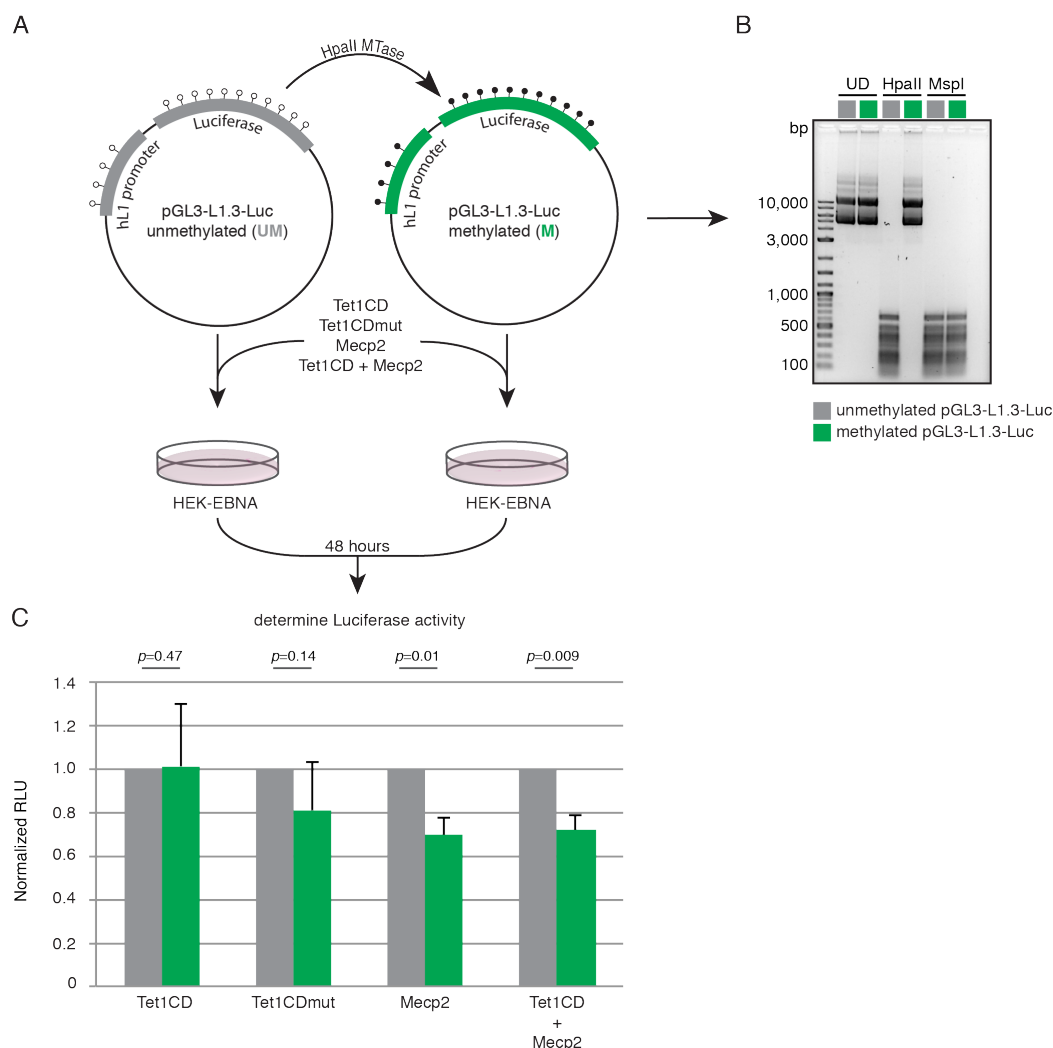


Figure S10 Transcription from a methylated L1 promoter is activated by Tet1CD and repressed by Mecp2. **(A)** Scheme illustrating the workflow and sites (nucleotides 36, 101, 304 and 481) in the L1.3-5'UTR methylated by M.HpaII methyltransferase. **(B)** HpaII and MspI digestion to control for efficient methylation of the pGL3-L1.3-Luc plasmid. While HpaII digestion is impaired by methylation, MspI cuts both, methylated and unmethylated DNA. **(C)** HEK cells were co-transfected with unmethylated (UM) or HpaII-methylated (M) reporter L1.3-Luc and expression constructs coding for Tet1CD, Tet1CDmut, Mecp2 and Tet1CD+Mecp2, respectively. Relative luciferase activity of the unmethylated reporter was set to 1. Bars represent mean luciferase activities of three independent experiments + SD and p values are indicated.

Video 1 Time-lapse fluorescence microscopy of the L1 retrotransposition reporter cell line upon Tet1CD expression. EGFP (left, green) indicates L1 retrotransposition events. mcherry-Tet1CD protein is shown in red (right). Multiple fields of view are shown stitched together and time is indicated at the top. Scale bar: 100 μ m.

Video 2 Time-lapse fluorescence microscopy of the L1 retrotransposition reporter cell line upon Tet1CDmut expression. EGFP (left, green) indicates L1 retrotransposition events. mcherry-Tet1CDmut protein is shown in red (right). Multiple fields of view are shown stitched together and time is indicated at the top. Scale bar: 100 μ m.

Video 3 Time-lapse fluorescence microscopy of the L1 retrotransposition reporter cell line upon mcherry expression. EGFP (left, green) indicates L1 retrotransposition events. mcherry protein is shown in red (right). Multiple fields of view are shown stitched together and time is indicated at the top. Scale bar: 100 μ m.

4.3 Principles of protein targeting to the nucleolus

A.K. Ludwig revised the manuscript.

A.K. Ludwig wrote the sections "Protein preparation", "RNA preparation" and "*In vitro* RNA binding assay" of the Materials and Methods part.

A.K. Ludwig performed protein and RNA preparation.

A.K. Ludwig wrote the results part for figures 4 and S2 together with H. D. Herce.

A.K. Ludwig performed and analyzed the experiments of figures 4 and S2 together with H. D. Herce.

A.K. Ludwig prepared figures 4 and S2 and wrote the respective figure legends together with H. D. Herce.

Content of figures:

Figure 4: Analyses of RNA binding to poly-(R)peptides and GFP fusions. **Figure S2:** Analyses of total RNA binding to arginine-rich cell penetrating peptides.

Principles of protein targeting to the nucleolus

Robert M Martin¹, Gohar Ter-Avetisyan², Henry D Herce³, Anne K Ludwig³, Gisela Lättig-Tünnemann⁴,
and M Cristina Cardoso^{3,*}

¹Instituto de Medicina Molecular; Faculdade de Medicina; Universidade de Lisboa; Lisboa, Portugal; ²Max Delbrueck Center for Molecular Medicine; Berlin, Germany; ³Department of Biology; Technische Universität Darmstadt; Darmstadt, Germany; ⁴Center for Stroke Research Berlin Charité; Universitätsmedizin Berlin; Campus Mitte; Berlin, Germany

Keywords: fluorescence microscopy, GFP, nucleolus, nucleolar localization sequence, protein targeting

Abbreviations: aa, amino acid; CPP, cell penetrating peptide; DDC, dense fibrillar compartment; FC, fibrillar centers; FITC, fluorescein isothiocyanate; GBP, GFP binding protein; GC, granular compartment; GFP, green fluorescent protein; NoLS, nucleolar localization sequence; NLS, nuclear localization sequence; PI, propidium iodide; TAMRA, 5-carboxytetramethylrhodamine; TAT, transactivator of transcription.

The nucleolus is the hallmark of nuclear compartmentalization and has been shown to exert multiple roles in cellular metabolism besides its main function as the place of rRNA synthesis and assembly of ribosomes. Nucleolar proteins dynamically localize and accumulate in this nuclear compartment relative to the surrounding nucleoplasm. In this study, we have assessed the molecular requirements that are necessary and sufficient for the localization and accumulation of peptides and proteins inside the nucleoli of living cells. The data showed that positively charged peptide entities composed of arginines alone and with an isoelectric point at and above 12.6 are necessary and sufficient for mediating significant nucleolar accumulation. A threshold of 6 arginines is necessary for peptides to accumulate in nucleoli, but already 4 arginines are sufficient when fused within 15 amino acid residues of a nuclear localization signal of a protein. Using a pH sensitive dye, we found that the nucleolar compartment is particularly acidic when compared to the surrounding nucleoplasm and, hence, provides the ideal electrochemical environment to bind poly-arginine containing proteins. In fact, we found that oligo-arginine peptides and GFP fusions bind RNA *in vitro*. Consistent with RNA being the main binding partner for arginines in the nucleolus, we found that the same principles apply to cells from insects to man, indicating that this mechanism is highly conserved throughout evolution.

Introduction

Proteins inside the highly compartmentalized cell nucleus often localize in correlation with their function to discrete subnuclear regions also referred to as subnuclear bodies.^{1,2} These nuclear substructures were identified as sites where specific biochemical reactions take place, e.g., DNA replication foci, or as sites of storage and modification for different proteins like in Cajal bodies and nuclear speckles.^{3,4} The localization of proteins in the nucleus is often mediated by specialized signals, targeting or interaction motifs of different length and can change over time, cell cycle and with external stimuli.^{5–8} The most prominent subnuclear structure is the nucleolus, which appears as a dark region in phase contrast microscopy due to its high density.^{9–11} The nucleolus is the place of transcription and storage of the rDNA as well as the assembly and maturation site for ribosomes and, in translationally active cells, filled with pre-ribosomal particles and components. The nucleolar structure is very dynamic

and forms around rDNA loci.^{12,13} It has further substructures, which are connected to the different steps in ribosome biogenesis and can be best identified by antibody labeling or electron microscopy. The fibrillar centers (FC) harbor the rDNA in the nucleolar center and are surrounded by the dense fibrillar components (DFC) containing the nascent rRNA transcripts. The granular component (GC) is filled with pre ribosomal particles composed of rRNA and proteins, extends into the nucleus and is surrounded by chromatin.^{11,14,15}

An early study has shown that the nuclear import of several proteins is mediated by short sequences of 6–10 basic amino acid (aa) termed nuclear localization sequences (NLS), some of which also serve as nucleolar targeting sequence (NoLS).¹⁶ Another nucleolar targeting sequence (NoLS) was described in 1990 and contains a minimum of 5 basic aa, often lysines and arginines.¹⁷ The recent systematic analysis of 46 NoLS sequences for their amino acid composition and sequence features has shown greater sequence diversity

© Robert M Martin, Gohar Ter-Avetisyan, Henry D Herce, Anne K Ludwig, Gisela Lättig-Tünnemann, and M Cristina Cardoso

*Correspondence to: M Cristina Cardoso; Email: cardoso@bio.tu-darmstadt.de

Submitted: 06/24/2015; Revised: 07/29/2015; Accepted: 07/31/2015

<http://dx.doi.org/10.1080/19491034.2015.1079680>

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

including bipartite sequences as well as overlap with NLS. The study determined also a high content of basic amino acids (48%) and a predominant location in the termini of proteins.¹⁸ Proteins smaller than 60 kDa or up to a diameter of 9 nm can passively diffuse into the nucleus¹⁹ but this process becomes more and more inefficient with increasing size (reviewed in^{20,21}). The nucleus is separated from the cytoplasm by a membrane system and involves an active import mechanism.²² In contrast, the localization of proteins to the nucleolus has never been shown to involve active transport mechanisms. Nuclear proteins without nucleolar function are often less concentrated or completely excluded from the nucleoli while nucleolar proteins are highly enriched in this nuclear compartment. The localization or exclusion of proteins from the nucleolus is subject to dramatic changes upon cellular stimuli, stress and complete reorganization during cell cycle progression.²³⁻²⁵

In a previous study, we have shown that a reporter protein traverses nucleoli rapidly with less obstruction compared to nucleoplasmic regions rich in chromatin.²⁶ In that study, we described a structural feature of the nucleolus with channel like regions void of nucleic acids and proteins that facilitates a fast movement of non-nucleolar proteins through this compartment. Contrastingly, the neighboring densely packed structures in the nucleolus reduce the available space for the localization of non-nucleolar proteins.²⁶ Still it is unclear, what are the essential properties or requirements of proteins and peptides to accumulate in the nucleoli and what are the features of these molecules that prevent nucleolar accumulation. The systematic identification of NoLS sequences by an artificial neural network trained with confirmed NoLS sequences has found thousands of potential sequences that could serve to mediate nucleolar targeting. Several previously unknown NoLS have been experimentally confirmed thereafter.^{18,27} However, a considerably false positive prediction rate especially among sequences from nuclear proteins raises the question of how NLS and NoLS can be distinguished. The difference in the composition of a standard NLS and the peptides that accumulate in nucleoli is rather subtle, with the NLS being a few basic charged amino acids shorter than, e.g., the nucleolar marker peptide of 10 arginines described by us.^{16,28} This observation led to the following 2 questions: (a) what are the specific requirements in amino acid composition for peptides and proteins to localize to the nucleoli and (b) what are the key differences between NoLS and NLS sequences?

In this study, we systematically investigated the molecular properties of peptides and proteins that result in different distribution levels between nucleoplasm and nucleolus. For this purpose, we tested and analyzed the distribution of a series of fluorescently labeled short peptides with different charge and isoelectric properties between the cytoplasm, nucleoplasm and their accumulation in the nucleolus. The same peptide sequences were genetically fused to the 5' end of the open reading frame coding for the tracer protein GFP, with no intracellular binding sites, to test the effect of the same short peptide sequences on the localization of proteins.

Materials and Methods

Peptides, plasmids and cell lines

All peptides were synthesized from L-amino acids except decarboxylated arginine (R10), which was made as 2 versions one from L- and one from D-amino acids by Peptide Specialty Laboratories GmbH (Heidelberg, Germany). The peptides were coupled directly to fluorescein at the N-terminus, contain a C-terminal amide-group and were purified by HPLC. The TAT peptide was synthesized with D-amino acids and conjugated with TAMRA as described.²⁹ It can be excluded that different amounts of fluorescent labeling on the different peptides play a role in the intracellular distribution, since the *in vitro* synthesis of peptides coupled to a resin material allows only in line addition of amino acids and the fluorophore. The pI values of the peptides and proteins were calculated using the ProtParam tool on <http://web.expasy.org/protparam/>.

The GFP based mammalian expression constructs were generated by using oligonucleotides encoding the respective poly-D, poly-G and poly-R sequences (supplementary Table 1). The oligos with overhanging 3' and 5' *AgeI* sites were annealed and ligated into the *AgeI* site of an NLS-GFP construct based on the pEVRF vector.³⁰ This vector is identical to the one described before for the CMV-driven expression of GFP-tagged PCNA.³¹ In addition, the pEGFP-C1 vector (Clontech, Heidelberg, Germany) without a NLS sequence was used.

Human HeLa and HEK 293-EBNA cells, as well as mouse C2C12 myoblasts were cultured as described before.^{32,33} Pac2 zebrafish cells were grown at 28°C in Leibovitz L-15 medium supplemented with 15% fetal calf serum in presence of 1% penicillin-streptomycin. Sf9 insect cells were maintained in EX-CELL TM 420 Insect Serum Free (SAFC) medium with L-glutamine (Biosciences) and 10% fetal calf serum and cultured with shaking at 100 rpm and 28°C. The L40ccua yeast strain used for the current study has the following genetic background: (MATa his3_200 trp1-901 leu2-3,112 LYS2::(*lexAop*)4-HIS3 ura3::(*lexAop*)8-lacZ ADE2::(*lexAop*)8-URA3 GAL4 gal80 can1 cyh2).³⁴ Yeast cells were grown at 30°C in yeast extract/peptone/dextrose (YPD) and were supplemented with 2% glucose as a carbon source.

Live cell peptide loading and transfections

All peptide experiments were performed exclusively with living cells plated on 8-well Ibidi chambers. The different polylysine (poly-K) and poly-arginine (poly-R) peptides were scratch loaded into living cells. For the scratch loading of peptides we used syringes and scratched over the chamber bottom in different directions.^{35,36} The size of the peptides permitted passive diffusion into the cells and nuclei after delivery. After the scratch loading the cells were allowed to recover for about half an hour at 37°C before exchanging the medium to remove unloaded peptides. Microscopy and image acquisition was started immediately after the procedures described above to circumvent peptide shortening by proteases and, hence, mislocalization. Peptide concentrations from 10 to 100 μM were initially tested and no apparent changes in cell viability were observed during the course of the

experiments. As low concentrations of 20 μM were sufficient for quantitative imaging, these conditions were used for the subsequent quantification. Scratch loaded cells did not show dramatic differences in the level of intracellular fluorescence, likely because the peptide concentration in the medium was the same in all cases. Only cells with normal cellular physiology comparable to non-scratch loaded cells were imaged. The criteria for the selection of cells that repaired after scratch loading were: i) cells are completely attached and stretched out on the chamber bottom surface; ii) no membrane lesions or leakage of cytoplasmic constituents visible in phase contrast; iii) round or oval shaped nuclei with smooth outline and nucleoli visible in phase contrast.

The transfection of cells with plasmid constructs was performed using the CaPO_4 -DNA co-precipitation protocol as described before,³⁷ except for HEK 293-EBNA cells, which were transfected using poly-ethylenimine (1 mg/mL in ddH_2O , pH 7; Sigma-Aldrich, St. Louis, MO, USA) as described before.³⁸

Cell fixation and antibody staining

Cells grown on glass coverslips for 24 h were incubated with 10 μM FITC-R10 for 1 h. After washing with 1xPBS, cells were fixed in 3.7% formaldehyde in PBS for 10 min and permeabilized with 0.25% Triton X100 in PBS for 10 min. Immunostaining with anti-B23 mouse monoclonal antibody (clone FC82291; Sigma) was followed by detection with donkey anti-mouse IgG antibody conjugated with TexasRed (Jackson). Coverslips were mounted with Moviol (Sigma).

Microscopy, image acquisition and analysis

Live cell microscopy was performed with a Zeiss LSM510Meta focal setup mounted on an Axiovert 200 M inverted microscope using a 63x phase contrast plan-apochromat oil objective NA1.4. The microscope was placed in an incubation chamber heated to 37°C to maintain the cell incubation conditions (Okolab, Italy). For the measurements of plasmid transfected cells, a Leica TCS SP confocal microscope was used. For all acquisition settings the main beam splitter was HFT UV/488/543/633. FITC was excited with the 488 nm line of an argon laser and fluorescence detected with a band-pass filter 500–530 nm. Phase contrast or differential interference contrast images were recorded simultaneously with FITC fluorescence in the transmission channel. The live-cell DNA dye DRAQ5 was used and imaged as described.³²

Cells loaded with peptides were chosen for imaging by having low but easily detectable intracellular fluorescence intensity. The same criteria were applied to cells transfected with plasmid constructs. Imaging of live cells was performed using as much as possible similar settings per instrument.

The SNARF 4F cell-permeant dye (Sigma) was loaded into cells by incubating for 30 min at a final concentration of 5 μM in growth medium. Fluorescence was excited at 561 nm and emission detected at 587 nm and 640 nm.

For image analysis, the nuclei were identified in the corresponding phase contrast images by means of the nuclear membrane delineation and nucleoli correspond to the dark dense structures inside the cell nucleus. Identification of nucleoli was

confirmed in earlier experiments using a nucleolar marker for living cells.²⁸ The selection of regions (cytoplasm, nucleus, nucleolus) and measurements of mean signal intensity were performed in ImageJ. Measurements of mean fluorescence intensities were performed by excluding non-interest areas from the measurement, e.g. the nucleoli areas excluded from nucleoplasm and, correspondingly, for all other regions intensities were determined. In each experiment the average was calculated for 10 cells for each peptide and protein. Data were normalized to 1.0 (100%) with respect to the nucleoplasmic fluorescence (excluding the nucleoli) to directly illustrate the relative accumulation of fluorescence in the nucleoli. Cytoplasmic fluorescence levels were normalized correspondingly (excluding nucleus and nucleoli). Data analysis and statistical tests were performed and displayed with Origin 7 software.

The complete list of predicted nucleolar localization sequences was downloaded from <http://www.compbio.dundee.ac.uk/www-nod/> and analyzed for mean length, mean numbers of amino acids and frequencies in Excel and Origin 7.

Protein preparation

GFP and NLS-R7-GFP were prepared from HEK cells 48 h post transfection using an ice-cold buffer composed of 20 mM Tris-HCl, 1.5 mM MgCl_2 , 0.2 mM EDTA, 1 M NaCl and 0.4% NP40. After homogenization by mechanical shearing through a 23 gauge syringe, lysates were incubated for 10 min on ice. Following centrifugation (14,000 rpm for 12 min at 4°C) supernatants were incubated for one hour with GFP binding protein (GBP) coupled to sepharose beads at 4°C on a rotary shaker as described.³⁹ GBP bound proteins were washed 3 times with ice-cold HEPES buffer (140 mM NaCl, 2.5 mM KCl, 5 mM HEPES, 5 mM glycine, pH 7.4) and used for the *in vitro* RNA pull-down assay. All buffers were supplemented with 1 mM of protease inhibitor PMSF (Carl Roth, Karlsruhe, Germany).

RNA preparation

Total RNA was isolated from HEK cells using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. To remove traces of genomic DNA, RNA was treated with RNase-free, recombinant DNaseI (Macherey Nagel, Dueren, Germany) for 30 min at 37°C and further purified with the Qiagen RNeasy Mini Kit. To assess the concentration and purity of RNA, the ratio of absorbance at 260 nm and 280 nm was measured on a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland). To further verify RNA quality, total RNA was denatured for 5 minutes at 99°C, separated on a 1.5% Tris-acetate-EDTA/agarose gel supplemented with 0.05 $\mu\text{L}/\text{mL}$ Roti-Safe GelStain (Carl Roth, Karlsruhe, Germany) by electrophoresis and imaged on an Amersham Imager 600 (GE Healthcare, Freiburg, Germany).

In vitro RNA binding assays

For RNA slot blots, total RNA was mixed with RNase-free ddH_2O at a final concentration of 500 ng/ μL and blotted on pre-equilibrated (5 min methanol, 5 min 20x saline sodium

citrate) PVDF membranes (0.45 μm pore size; Pall GmbH, Dreieich, Germany). After air-drying, membranes were blocked for 30 min with 5% milk in PBS supplemented with 1x ProtectRNA RNase inhibitor (Sigma-Aldrich, St. Louis, MO, USA) and incubated for 1 h with either TAT, D-R10, or L-R10 (all fluorescently labeled), respectively, on a rotary shaker. Following three washing steps (5 min each) with HEPES buffer, remaining fluorescent signals were detected on an Amersham Imager 600 (GE Healthcare, Freiburg, Germany). Quantities of blotted RNA were monitored in parallel by methylene blue (0.02%) staining (Carl Roth, Karlsruhe, Germany) in 0.3 M sodium acetate (Merck, Darmstadt, Germany) pH 5.5.

For the pulldown assay, immobilized GFP and NLS-R7-GFP, respectively, were incubated with equal amounts (3.4 μg) of total RNA in 20 μL HEPES buffer for 1.5 h at 4°C on a rotary shaker. After three washing steps with HEPES buffer, RNA was labeled with 3.3 $\mu\text{g}/\text{mL}$ propidium iodide (Sigma-Aldrich, St. Louis, MO, USA). Following three washing steps in HEPES buffer, remaining fluorescent signals were measured on a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland) using excitation/emission at 488/555 nm and 515/617 nm for GFP and propidium iodide, respectively. To control for the amount of GFP and NLS-R7-GFP, fluorescent signals of propidium iodide were normalized to GFP signals. In addition, the same beads were imaged in an UltraVIEW VoX spinning disc confocal system (PerkinElmer), mounted on a Nikon TI microscope. Images were taken with a 20x/0.7 NA objective. GFP and propidium iodide or TAMRA were imaged with 488 and 561 nm laser excitation and 527 ± 55 and 612 ± 70 nm emission filters, respectively.

Results

Short specific sequence motifs in proteins are well known to play a key role in localizing proteins to specific cellular and sub-cellular compartments. We have shown earlier that short peptides of 10 arginines with a strong basic charge have the potential to accumulate inside the nucleoli of living cells and can be used to label this subnuclear compartment.²⁸ Now, we wished to investigate the specific molecular requirements of peptides and proteins that are necessary and sufficient to reach and accumulate inside of the nucleolus.

Therefore, our first aim was to determine the properties of charge, length and composition of basic charged peptides necessary for nucleolar targeting and accumulation. We used short synthetic peptides labeled with FITC for visualization, which were scratch loaded into living C2C12 mouse myoblast cells. The peptides, either composed of arginines (R) or lysines (K) with a length varying from 5 to 12 amino acids, were allowed to enter the cells and distribute in the cytoplasm and nucleus. The observation of the cells by confocal microscopy started after a recovery time of 30 min after scratching loading. For each peptide at least 10 cells were selected and imaged with respect to similar intracellular fluorescence intensity levels. The microscopic observation of the cells shows for all peptides a homogeneous

distribution in the cytoplasm, where only vesicular structures were free of labeled peptides. Similarly, in the nucleoplasm the peptides distributed without enriching at any foci. All L-amino acid peptides tested here, K5 to K12 and R5 to R12 accumulated predominantly in the nucleus compared to the cytoplasm, as shown by the fluorescence intensity in mouse C2C12 cells (Fig. 1A and B). The peptides R6 to R12 clearly showed higher fluorescence intensity in areas inside the cell nucleus, which were identified as nucleoli correlating with the dark dense structures visible in the phase contrast images (Fig. 1A). The peptide R5 did not accumulate in nucleoli (Fig. 1A), which was also the case for all poly-K peptides (Fig. 1B). This indicates that charge alone is not the only determinant but rather the isoelectric properties, which differ between K and R polymers. To quantitatively analyze the intensity and distribution patterns, we determined for 10 cells the mean fluorescence intensity in the cytoplasm, nucleus (excluding nucleoli) and in the nucleoli. The regions for fluorescence intensity measurements were chosen with reference to the phase contrast images and overlaid with the fluorescence image as shown in Figure 1C. Data were normalized to the nucleoplasmic levels to reflect the accumulation potential of molecules in the nucleolus over the nucleoplasmic levels (Fig. 1C). The numerical values are given in supplementary Tables 2 and 3.

The comparison of the mean fluorescence intensity shows, for all poly-K peptides, an average cytoplasmic level of 0.74 fold less compared to the nucleoplasm. For poly-R peptides the cytoplasmic level is on average 0.68 fold less than in the nucleoplasm (Fig. 1A, B and supplementary Table 2). These data already demonstrate that poly-R peptides have the potential for stronger accumulation in the nucleus than poly-K peptides. Although all peptides tested are small enough to passively diffuse in and out of the nucleus and should distribute evenly, all of them accumulate in the nucleoplasm.

Strikingly, only the peptides R6 to R12 show a further visible accumulation inside nucleoli with an average of 1.16-fold over the nucleoplasmic level (Fig. 1C). The statistical analysis with a non-parametric Kruskal-Wallis ANOVA test shows significance for the accumulation of R6 to R12 in the nucleolus. The poly-K peptides show only slightly higher level in the nucleoli with values between 1.02 and 1.06 fold over the nucleoplasm. An increase of the peptide length from R6 to R12 and thus the positive charge and isoelectric properties does not linearly increase the nucleolar accumulation level of the peptides. The highest level of nucleolar accumulation was found for the R9 peptide with 1.22 fold over the nucleoplasmic level, while the mean accumulation of R11 and R12 rather decreases slightly to 1.12 and 1.14-fold respectively (Fig. 1A, B and supplementary Table 2). This suggests that there is a saturation/plateau level at and above 9 arginine residues.

From yeast to human cells, main features of the nucleolus structure and function are conserved as it forms around clusters of rRNA genes and is the sites of ribosome biogenesis. We were now interested if the nucleolar localization and accumulation of proteins and peptides is similarly conserved. Therefore, we tested the nucleolar accumulation potential of the R10 peptide in cultured cells of different species. We scratch

loaded the FITC labeled R10 into cells from yeast (*S. cerevisiae*), insects (*D. melanogaster*), fish (*D. rerio*), mouse (*M. musculus*) and human (*H. sapiens*) cells and analyzed the localization of the peptides by confocal microscopy (Fig. 1D). In yeast cells the nucleus and nucleolus are more difficult to identify and clear accumulations in the cells are not as visible as for the cells of higher eukaryotes. Regions with less peptide inside the yeast cells could correspond to the vacuole. In the cells from insect, fish, mouse and human, on the other hand, the nucleoli are visible in the phase contrast images and show clearly an accumulation of the R10 peptide compared to the nucleus and cytoplasm of the cells (Fig. 1D). To clarify the localization of the FITC-R10 in yeast we performed a co-labeling of yeast cells with the live cell DNA dye DRAQ5.³² In this experiment the yeast nucleus is visible as intense labeled round compartment in the cells. In some cells we could identify a crescent shaped region in the periphery of the nucleus with higher FITC-R10 intensity compared to the nuclear interior and the cytoplasm. A corresponding fluorescence intensity linescan analysis confirms an accumulation of the peptide in a crescent shaped region in the nuclear periphery (Fig. 1E). Early studies on the composition of yeast nuclei have shown that, depending on the orientation a round or crescent shaped peripheral nuclear compartment harbors the nucleolus.⁴⁰⁻⁴² All cell types of multicellular organisms tested here show a similar enrichment of the labeled poly-R peptides in the nucleoli. Some yeast cells show a similar accumulation in the nuclear periphery. The accumulation of peptides like R10 inside nucleoli is, thus, evolutionary conserved at least from insects to humans. Notably,

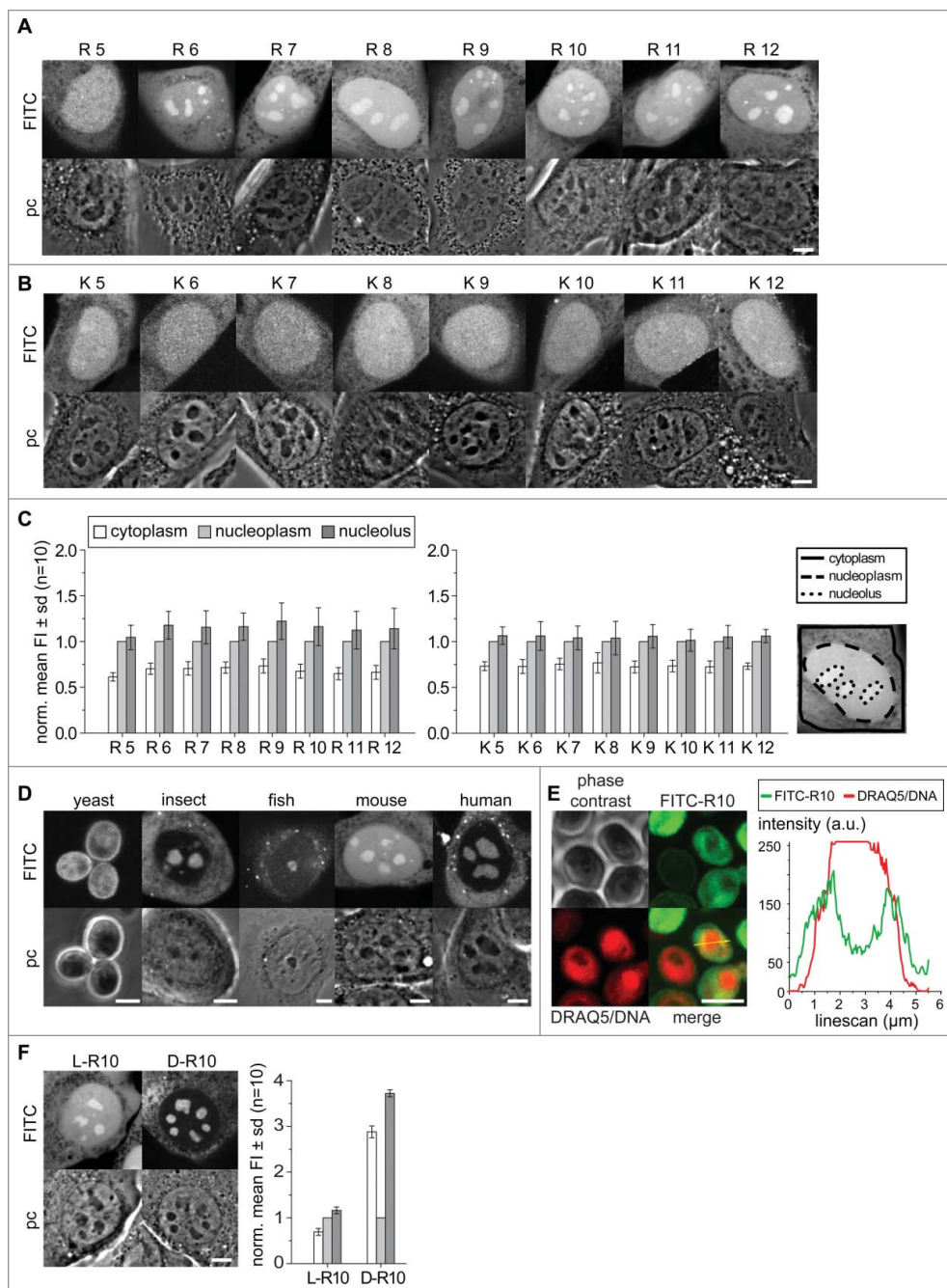


Figure 1. Distribution of poly-(R) and poly-(K) peptides in living cells. Intracellular distribution of poly-R peptides in (A) and poly-K peptides (B) in living C2C12 mouse cells. In each panel, the fluorescence image is on top of the corresponding phase contrast image. Nucleoli are clearly visible as dark round structures within the nuclei in the phase contrast images. The bar diagrams in (C) show the quantification of the poly-K and poly-R peptide mean fluorescence in cytoplasm, nucleoplasm and nucleoli averaged for 10 cells from 2 independent experiments. The nucleoplasmic values were used for normalization. Areas for quantification were defined as described in methods and overlaid with fluorescence images as shown. Intracellular distribution of R10 peptide in living cells of different species as indicated is shown in (D). In (E) living yeast cells were further stained with DRAQ5 for better visualization of the nucleus and a line intensity profile in arbitrary units (a.u.) of both DNA and peptide is shown. The intracellular distribution of (D) and L-R10 peptides in living C2C12 cells and the corresponding quantification of mean fluorescence intensities is shown in (F). Scalebars: 5 μm .

poly-R peptides is directly linked to the structure and function of the nucleolus (supplementary Fig. 1D).

All amino acids incorporated into proteins by translation in living organisms are L-enantiomers and only post-translational enzymatic reactions in various organisms convert some to D-amino acids.^{43,44} Besides that, free D-amino acids were discovered only in brain tissue.⁴⁵ We argue that a poly D-arginine peptide may not exhibit the specific binding to the same cellular target components, e.g., proteins or nucleic acids, as the L-enantiomer. In contrast, the peptide charge and isoelectric properties are not altered when composed of D-amino acids. This phenomenon is exploited in mirror image phage displays used in the search for therapeutic peptides with novel targets and improved properties.^{46,47} The latter exploits the stability of the D-amino acid peptides to proteolytic degradation. Hence, we compared the distribution of FITC-R10 composed of D and L amino acids (Fig. 1F). Inside living cells both D- and L-R10 peptides distribute homogeneously in the cytoplasm and accumulated in the nucleoli. The main difference in the distribution is that D-R10 had a much lower concentration in the nucleoplasm compared to the L-R10, therefore the measurement of relative accumulation of D-R10 over the nucleoplasmic level appears much stronger (Fig. 1F). However, the absolute detected fluorescence intensities were comparable between both types of peptides using the same

microscopic detection settings. This finding indicates either a binding of the L-peptide to a nucleoplasmic component, which is not the case for the D-peptide, and/or the lack of degradation of the D-R10. Interestingly though, the mean intensity of fluorescence in the nucleoli is at similar levels for both types of peptides.

Next, we tested the role of various peptides in the targeting of proteins inside the cell nucleus. We used peptides fused to the neutral tracer protein GFP to determine localization patterns. For this purpose, we used GFP already fused to a NLS sequence (SV40 T antigen derived; PKKKRKV⁴⁸) and added downstream further basic, acidic and neutral peptides of different lengths followed by the enhanced GFP coding sequence. The different GFP fusions were expressed in C2C12 mouse myoblast cells (Fig. 2A) and in human HeLa cells (supplementary Fig. 1) and imaged with a confocal microscope. We also tested the distribution of an uncoupled GFP and the NLS-GFP alone as comparison and as controls for cytoplasmic and nucleoplasmic localization with no expected nucleolar localization and accumulation.

The microscopic images of the GFP versions in Figure 2A show for the GFP without NLS a homogeneous distribution throughout the cell with similar levels in cytoplasm and nucleus with significant reduction in the nucleoli ($p = 0.01$ see supplementary Table 2). The slightly lower mean intensity in the cytoplasm is the result of membranous compartments

devoid of GFP protein (Fig. 2A and C; supplementary Table 2). The homogeneous distribution between nucleus and cytoplasm is the result of passive diffusion into the nucleus. Molecules with a diameter of up to ~9 nm are capable of entering the cell nucleus by passive diffusion, which has been measured for fluorescent molecules the size of GFP to occur within a few minutes.^{19,49} The NLS coupled GFP, though, shows more than 2-fold accumulation in the nucleus over the cytoplasmic level but with a similar reduction in the nucleoli, visible as round elevated objects in the DIC images. This distribution pattern is similar for all NLS-GFP variants with additional acidic (aspartate) and neutral (glycine) amino acid series (Fig. 2A and C). The nucleolar level of the different proteins in Figure 2A is significantly reduced in the range of 0.7 to 0.5-fold compared to the nucleoplasmic level as determined by measuring the mean fluorescence intensity (Fig. 2C and supplementary Table 2).

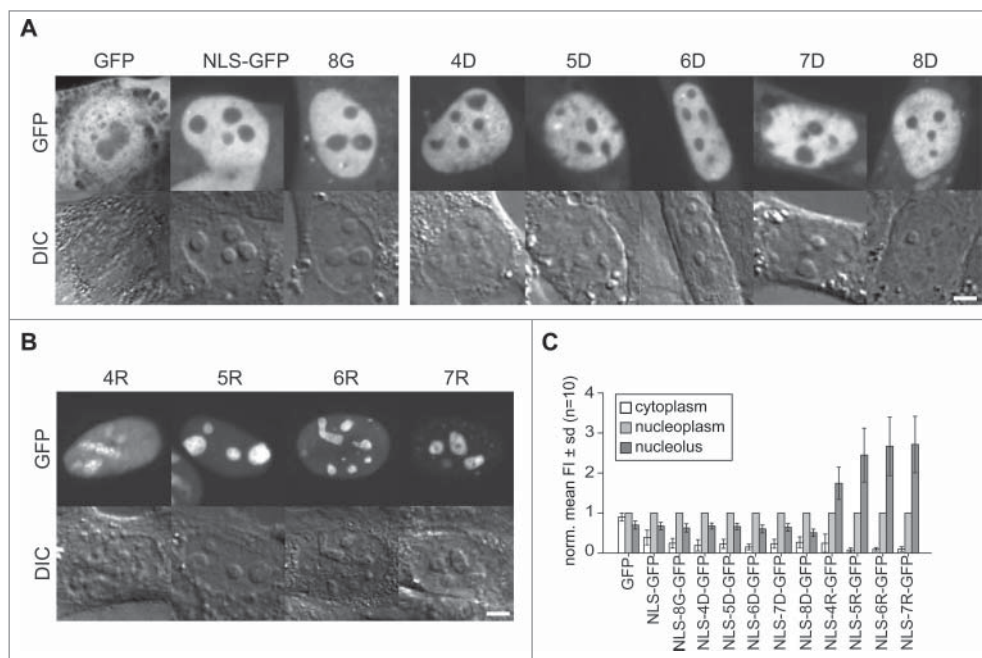


Figure 2. Intracellular distribution of peptide tagged GFP tracer proteins in living cells. (A) Representative microscopic images of live C2C12 cells transfected with GFP, NLS-GFP and a fusion of NLS-GFP with 8 glycines (8G) as example for the addition of neutral amino acids is shown in the first panel. The second panel shows images of cells with fusions of an increasing number of aspartates as examples for acidic amino acid fusions to NLS-GFP. The GFP fluorescence is depicted in the upper row with the corresponding differential interference contrast (DIC) images below. (B) Microscopic Images of the distribution of the NLS-GFP fused to 4 to 7 additional arginines. The graphs in (C) represent the mean distribution of the GFP constructs depicted in (A) and (B) in cytoplasm and nucleoli in relation to the nucleoplasm for an average of 10 cells from 2 independent experiments. Scalebars: 5 μm .

Contrastingly, all NLS-GFP versions coupled to poly-R peptides from R4 to R7 show a clear accumulation in the nucleoli around 2 to 3-fold over the nucleoplasmic level. The nucleolar accumulation is lowest for NLS-R4-GFP at 1.7-fold and increases with the growing number of arginine residues up to 2.7-fold for NLS-R7-GFP (Fig. 2B and C and supplementary Table 2). Interestingly, the comparison of the data for peptides and proteins shows that in contrast to the FITC-R5 peptide, we see already for the R4 fused to NLS-GFP a nucleolar accumulation. This may reflect the proteolytic degradation of the L-amino acids containing peptides versus the more stable corresponding GFP fusion proteins. On the other hand, the additional arginine within the NLS sequence could account for this apparent difference.

Noteworthy, peptides without nucleolar accumulation like K5 and R5 show no reduced concentration inside nucleoli. Already in an earlier study testing functional peptides not accumulating in nucleoli, we observed a similar homogenous nuclear and nucleolar distribution of a fluorophore labeled peptide.⁵⁰ These observations exemplify the dense structure of the nucleoli resulting in a size dependent sieving effect for molecules without potential for accumulation therein. Larger molecules like GFP (with dimensions of around 2.4×4.2 nm)⁵¹ are significantly excluded compared to much smaller peptides. Interestingly, the R5 to R7 coupled NLS-GFP molecules show in addition an even higher efficiency of nuclear import with levels up to 10-fold over the cytoplasm compared to NLS-GFP alone (Fig. 2B, C and supplementary Table 2). This observation shows that the addition of arginine residues to a minimal NLS increases the efficiency of nuclear import largely exceeding the passive leakage efflux of small protein molecules through nuclear pores. Alternatively, their active retention in the nucleus by binding, e.g., to the nucleolus, prevents their exit from the nucleus.

The data presented here so far demonstrate that peptides and proteins with a strong positive charge and high isoelectric point have the potential to accumulate in the nucleolus of live cells. The threshold for strong nucleolar accumulation in peptides is R6, while in proteins bearing an NLS (with one additional R) already 4 additional consecutive arginines are sufficient. In addition, the nuclear import is enhanced by the addition of at least 5 arginine residues in proximity to a minimal NLS. We conclude that the charge and isoelectric properties of the peptides and protein domains responsible for nucleolar accumulation is key in this process because first, the difference between nuclear

import and nucleolar accumulation has never been shown to rely on an active mechanism. Second, even peptides synthesized with D-amino acids show nucleolar accumulation, which may fail to specifically bind to the same cellular targets as their L-enantiomer counterparts.

The evidence for a charge and isoelectric point dependency for peptide and protein accumulation in nucleoli led us to postulate the presence of an electrochemical component. Electrochemical interactions that would accumulate positively charged basic peptides should be mediated by highly abundant or strong negatively charged acidic nucleolar components. These have to be confined to nucleoli as are the accumulated basic peptides and proteins. Such molecules could be nucleic acids, which are highly abundant in the nucleoli in form of rRNA. Therefore, we wanted to determine the pH distribution inside the nucleus in living cells. We used the live cell permeable fluorescent dye SNARF-4F, which has 2 pH dependent ratiometric emission peaks at around 587 and 640 nm. The calculation of the ratio of the fluorescence intensity at the 2 emission peaks gives information on the pH of subcellular structures. The ratiometric images of HeLa cells show in the fluorescence channel view as well as in the intensity ratio view different cellular structures highlighted (Fig. 3A top row). The magnified image detail shows structures at the nuclear periphery and in the center of the nucleus highlighted as more acidic (Fig. 3A bottom row). Although the cytoplasm is generally considered to have a neutral pH, the more acidic regions in the cytoplasm do most likely represent parts of the Golgi network and organelles like peroxisomes.⁵² The central nuclear region with more acidic pH is reminiscent of the nucleoli in HeLa cells shown in the phase contrast images in Figure 1D as well as the DIC images in Figure 2 and in supplementary Figure 1. Almost

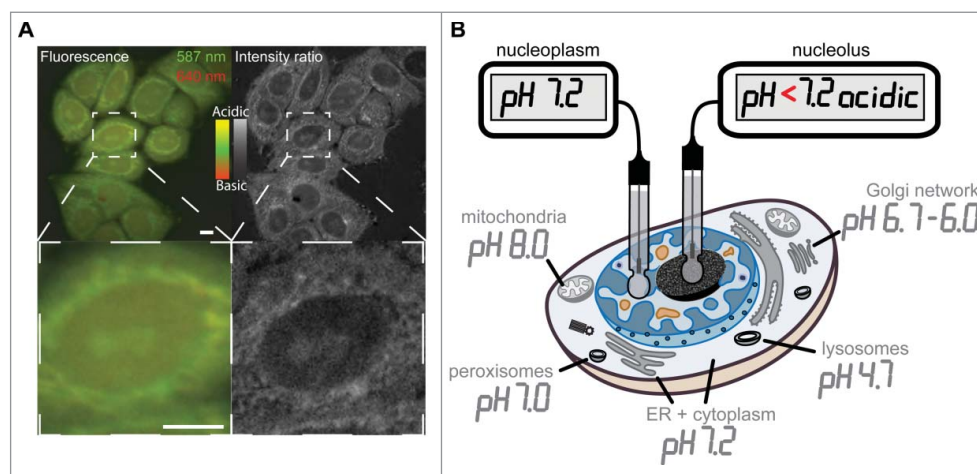


Figure 3. Intracellular pH landscape by ratiometric fluorescence microscopy toward the generation of an intracellular pH landscape. The figure in (A) displays fluorescence microscopy images of the pH sensitive dye SNARF-4F in live HeLa cells (color). The green channel shows the emission at 587 and the red channel the emission of the dye at 640 nm during excitation with the same wavelength. The resulting intensity ratio image (gray scale) shows ratiometric differences between the intensities of the 2 fluorescence emission peaks of the dye due to pH variations in subcellular structures. The fluorescence color scale and the ratio scale in gray indicate the relative range from more acidic to more basic pH. Panel (B) illustrates a map of the intracellular pH landscape by measuring the pH in various cellular compartments and substructures like nucleoplasm and nucleolus (modified from⁵³). Scalebars: 5 μ m.

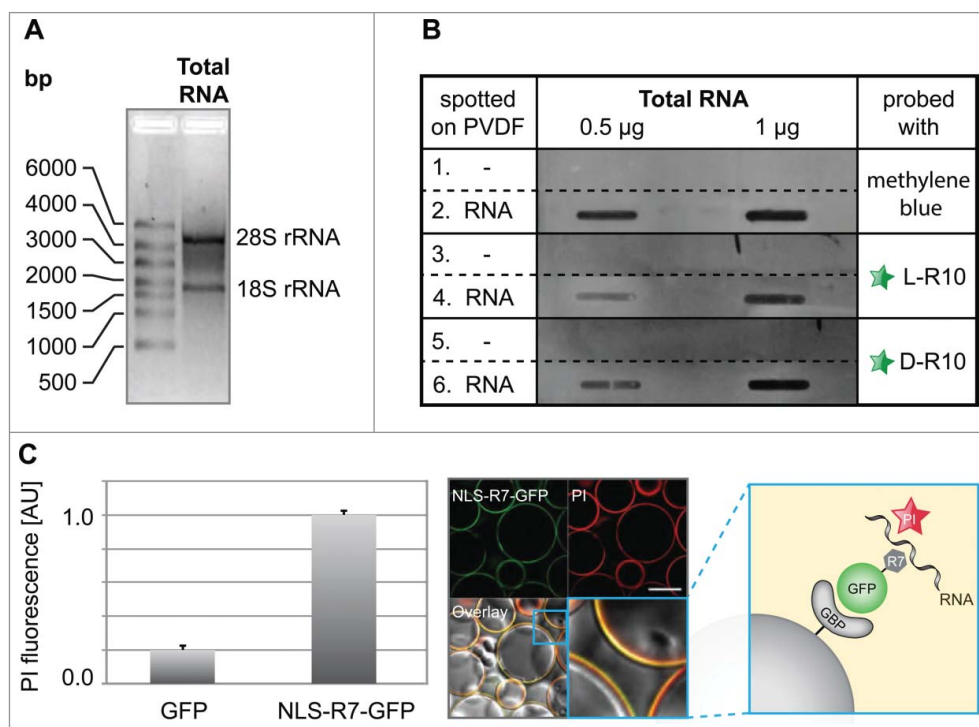


Figure 4. Analyses of RNA binding to poly-(R)peptides and GFP fusions. **(A)** Total RNA used for *in vitro* RNA binding assays separated by size showing the absence of genomic DNA contamination, as well as characteristic bands for the 28S and 18S rRNA, which are primarily synthesized and localized in nucleoli. The slot blot in **(B)** (representative from 2 independent experiments), rows 1 and 2, show a methylene blue stained PVDF membrane in the absence (1) and presence (2) of different amounts of total RNA (0.5 μ g and 1 μ g) and were used as a loading control. The rest of the blot shows the binding of fluorescently tagged L-R10 (3 and 4) and D-R10 (5 and 6) in the absence (3 and 5) and presence (4 and 6) of total RNA. Slots lacking RNA (rows 5 and 6) do not show binding of L- and D-R10, respectively. In contrast, slots probed with increasing amounts of RNA show increasing L- and D-R10 binding. **(C)** *In vitro* RNA pull-down assay using NLS-R7-GFP immobilized to sepharose beads via the GFP binding protein (GBP) (scheme). RNA was stained using propidium iodide (PI) and GFP alone was used as a negative control. RNA binding was measured on a fluorescent plate reader (bar plot) and imaged by confocal microscopy (microscopic images). Plots represent the average plus standard deviation of 2 independent experiments. Scalebar: 100 μ m.

every cell in the overview image has the nucleoli highlighted as more acidic compared to the nucleoplasmic surrounding (Fig. 3A top row). Although precise measurements of subcellular pH values are not possible with the SNARF-4F dye, it allows us to conclude that the nucleolar environment is more acidic than the neutral pH of the nucleoplasm. The subcellular pH landscape was measured in various studies before and is reviewed in.⁵³ Figure 3B summarizes the reviewed data and was adapted to now incorporate the differentiated neutral pH of the nucleoplasm and more acidic pH of the nucleolus obtained from our measurements.

Next, we wanted to test whether the peptides and proteins with the sequence requirements sufficient to accumulate in nucleoli do bind RNA. To test our hypothesis, we purified total RNA (Fig. 4A) and performed slot blot analysis (Fig. 4B). We blotted increasing amounts of total RNA on a PVDF membrane followed by incubation with either fluorescently labeled L-R10, or D-R10, respectively (Fig. 4B). The fluorescently tagged D- and L-R10 were only detected on slots spotted with RNA (Fig. 4B, rows 4 and 6) but not in control slots without

RNA (Fig. 4B, rows 3 and 5). As a loading control, membranes were stained in parallel with methylene blue (Fig. 4B, rows 1 and 2). To test the RNA binding of basic peptides fused to GFP, we performed an *in vitro* RNA pull-down assay using NLS-R7-GFP immobilized to sepharose beads via the GFP binding protein (GBP)³⁹ (Fig. 4C, scheme). This assay allowed us to observe (Fig. 4C, confocal image) and quantify (Fig. 4C, bar plot), the binding of RNA to NLS-R7-GFP. The accumulation of peptides in nucleoli is a common feature of arginine-rich cell penetrating peptides (CPPs).⁵⁴ Consistently, we found that the TAT peptide, one of the best known CPPs derived from HIV-1 TAT protein,⁵⁵ also binds to RNA (supplementary Fig. 2).

Defining experimentally the requirements of nucleolar peptide and protein accumulation has shown that several arginines in sequence are required. At least 6 arginine residues in a peptide and 4 arginines in combination with an NLS provide sufficient localized surface charge for nucleolar accumulation. Importantly, charge alone cannot explain the difference between poly-R and

poly-K, which may additionally rely on the higher number of functional amino groups in arginines vs. lysines. Now we asked, how our findings relate to NoLS sequences found in the human proteome? A previous comprehensive study of nucleolar targeting sequences has systematically evaluated 46 nucleolar targeting sequences curated from the literature. Based on this analysis an algorithm was developed to predict nucleolar targeting based on peptide amino acid composition. The algorithm was then run against the International Protein Index database (IPI 3.4) and predicted >10,000 putative nucleolar targeting sequences. We have used this list to perform further analysis of the amino acid composition of predicted nucleolar targeting sequences. An overall statistics shows that the predicted NoLS have an average length of 27 amino acids with a most frequent length of 20 amino acids. The vast majority (83%) of predicted NoLS sequences has a length of 30 amino acids or less. For these NoLS the mean ratio of the number of basic (lysine, arginine and histidine) versus the total number of amino acids is 0.34. The shortest and most frequently predicted NoLS sequences of 20 amino acids therefore contain on average at least 7 basic amino acids. The 15

amino acids sequence containing NLS-R4 from the GFP coupled construct tested here has a ratio of 0.66 and contains 10 basic amino acids. A statistical analysis of the individual number of the amino acids for the 2001 predicted NoLS of 20 amino acid length most frequently contain at least 3 lysines, 3 arginines, one histidine, 2 cysteines, 2 prolines, and further each one aspartate, glutamate, glycine, alanine, threonine, valine, leucine, asparagine, glutamine. A peptide with the amino acid composition KKKRRRDEGATVLNQ would have a p*K*i value of 11.00, which is comparable to the FITC-K11 peptide tested in this study, which did not show significant nucleolar accumulation. But the combination of 4 arginines with an adjacent but separated NLS within 15 residues, as in NLS-R4-GFP, proved effective.

To gain a systematic overview over the peptide and protein localization dependent on charge, isoelectric properties and amino acid sequence composition, we created gradient charts depicting the pI values along with the respective nuclear and nucleolar accumulation as well as peptide composition maps (Fig. 5). The chart for peptide and total protein pI values shows a total difference between all poly-K peptides (pI 10.6–11.04) exhibiting no nucleolar accumulation and the poly-R peptides with more basic pI values between 12.6 and 13.04 (Fig. 5A). Although the pI values of the R5 and R6 peptides differ only slightly with 12.6 and 12.7 respectively, the

charge driven threshold for nucleolar accumulation of peptides is found between these values (Fig. 5A). When comparing the pI values of proteins it becomes apparent that the total protein pI does not represent specific charged targeting sequences or domains appropriately. All poly-R fused NLS-GFP versions with nucleolar accumulation range between pI 7.77 and 8.83, which

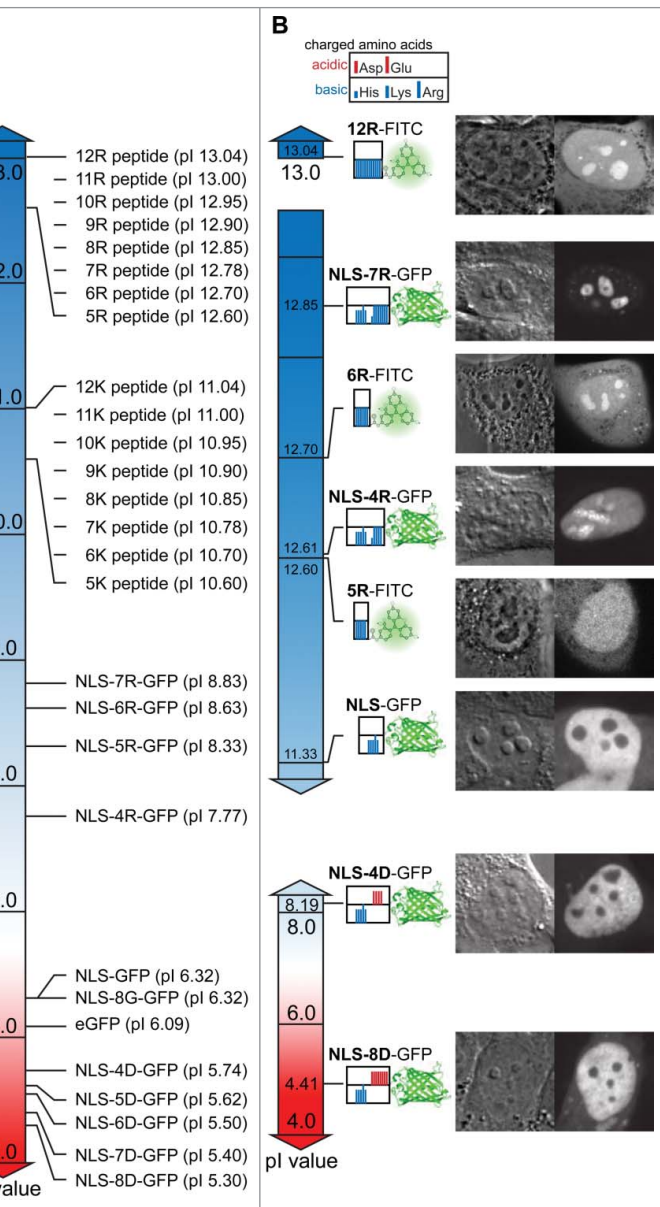


Figure 5. Chart of peptide and protein pI values, intracellular distribution and sequence composition. **(A)** Display of the pI value of FITC labeled peptides and full-length proteins coupled to different peptides along a continuous pI scale. Representative images illustrating the intracellular and intranuclear distribution are indicated along the scale. **(B)** Subsets of the pI scale with the values exclusively for the charged amino acid domains in the proteins and peptides tested combined with an illustrative representation of the domains. For the proteins the domain considered starts with the nuclear localization signal (NLS) and continues through the different motifs of charged amino acids (poly-R and poly-D) introduced into the GFP open reading frame. The peptides tested consist exclusively of basic amino acids. The clustering of the charged amino acids is displayed as color bars in the box adjacent to the respective fluorophore FITC or GFP. A legend for the charged amino acids is given above. Representative images for the distribution inside the cells and nuclei are displayed on the right for the respective peptide or protein. Scalebar: 5 μ m.

is much less basic than any of the poly-R peptides. The pI value of GFP of 6.09 has a strong influence on the calculation of pI values for the fusion proteins with relatively short additions of basic amino acids (Fig. 5A). Therefore, we also plotted the pI of the peptides tested and the protein N-terminal domains containing the NLS and the fused poly-R and poly-D sequences (Fig. 5B).

This representation shows that the pI values of the FITC-5R, which shows homogeneous nuclear distribution, and the NLS-4R, which shows nucleolar accumulation are virtually the same. The main difference between these sequences is the distribution of the charge over just 5 residues (5R) or 15 residues (NLS-4R) resulting in a different surface charge density. Therefore, we reason that in addition to the charge and isoelectric properties of peptides and protein localization sequences, also the charge distribution plays a key role in mediating nucleolar accumulation.

Discussion

In this study, we have analyzed the requirements of peptides and proteins to accumulate in the nucleoli of living cells. It has been already established that specific sequences in proteins are responsible for the nuclear import (NLS) and further for the localization in the nucleoli (NoLS).^{16,56} The systematic analysis of experimentally confirmed NoLS has shown that a high content of basic amino acids is common among nucleolar but also nuclear targeting signals, which makes it difficult to differentiate between NoLS and NLS. This difficulty is further aggravated by the fact that validated and predicted NoLS depict high sequence diversity.^{18,57}

Our results show first that all poly-K and poly-R peptides ranging from 5–12 amino acids accumulate in the cell nucleus compared to the cytoplasmic level and, therefore, match the criteria for nuclear localization sequences. Similarly, all NLS bearing GFP versions tested show a nuclear accumulation, which is generally a feature of nuclear proteins bearing a NLS.⁵⁷ In line with this interpretation, the GFP protein without NLS shows a homogeneous distribution between nucleus and cytoplasm. Interestingly, the mean cytoplasmic fluorescence intensity of all L-peptides tested (0.61 to 0.76 fold) is much higher than for any of the NLS containing GFP versions (0.08 to 0.39) showing that passive nuclear pore leakage is less efficient and slower for larger molecules. The distribution test for the same peptides with different chirality showed that FITC-R10 composed of D-amino acids has a reduced concentration in the nucleoplasm compared to the L-amino acid peptide. We interpret this observation as lack of specific binding to nuclear im/export factors together with a sequestration of the D-peptide in nucleoli. In addition, the potential degradation of L-amino acid peptides during the experiments, would result in freely diffusing FITC molecules not bound to peptides and increase their nucleoplasmic signal altogether.

The nucleolar accumulation of peptides is dependent on the amino acid composition as poly-R peptides, except R5, show strong accumulation whereas poly-K peptides accumulate only slightly and, in most cases, not significantly. A sequence of at least 6 arginines per peptide is required, while even 12 lysines show only very little effect. These findings indicate the existence of an isoelectric threshold to reach accumulation in nucleoli, although the pI difference between R5 and R6 is rather small with 12.6 and 12.7 respectively (Fig. 5).

Surprisingly, already the fusion of 4 arginines residues to NLS-GFP results in a strong protein accumulation in the nucleolus. The level of accumulation is even higher than for all FITC labeled poly-R peptides tested, while the pI of 12.61 for the domain containing the NLS-4R is virtually the same as for the non-accumulating R5 peptide. In this case, a cumulative effect of the poly-R sequence and the adjacent NLS within a short region of 15 amino acids, containing an additional arginine as well as further lysine residues, generates an effective motif that binds to nucleolar constituents. It has been shown a long time ago, that *E. coli* ribosomes contain at least 21 proteins with extremely high pI values >12.0, similar to the poly-R peptides and NLS-poly-R domains fused to GFP.⁵⁸

These data suggest that, in addition to the peptide and protein domain charge alone, the surface charge density is an important determinant of interaction with nucleolar constituents. A protein domain surface with a larger distribution of positive charges below a pI of 12.7 is also effective to promote sufficient interaction to accumulate in nucleoli. On the other hand, sequences with even more lysine residues or shorter examples with few lysines and arginines can also serve as NoLS as found in a previous systematic analysis and predicted by the Nucleolar Localization Sequence Detector webtool.¹⁸ The other tracer proteins GFP, NLS-GFP and those in fusion with neutral or positively charged amino acids show no accumulation in nucleoli. In fact, they are rather largely “apparently” excluded from nucleoli. Nevertheless, it has been shown that even those molecules enter and rapidly traverse nucleoli.²⁶ On the contrary, even the non-accumulating poly-K peptides and others have been found to homogeneously distribute in nucleoplasm and nucleoli. These findings suggest a volume exclusion or molecular sieve effect of the nucleolar compartment similar to the size dependent passive diffusion through nuclear pore channels.

The requirement of a strong positive charge mediated by basic amino acids distributed over a certain molecular interaction surface seems to be a hallmark of the potential to accumulate proteins in the nucleolus. The nucleolar interaction partners are either more abundant than in the surrounding nucleoplasm or lead to a strong retention within this subnuclear compartment.⁵⁹ A charge and isoelectric properties driven accumulation has to be mediated by a countercharged nucleolar component. Highly abundant and integral functional determinant of nucleoli is rRNA, which can be considered as large negatively charged complexes. Several ribosomal proteins possess high pI values, thus, are positively charged and are very likely to electrostatically interact with RNA. We could indeed show the presence of a more acidic pH in the nucleoli by the live-cell pH sensitive dye SNARF-4F. The RNA transcribed in nucleoli is a prime candidate for the locally constrained proton donor that is also an ideal target for protein and peptide nucleolar localization sequences. Our suggestion is supported by a report on protein immobilization in the nucleolus by noncoding RNA via the nucleolar detention sequence (NoDS). In this study, the deletion of arginine motifs in the NoDS of Hsp70 prevents nucleolar accumulation as well as the precipitation of a peptide NoDS failed without arginines in the motif.⁶⁰ Sequence alignment of the NoLS for the

human La protein with several other nucleolar localization domains of human proteins revealed a (R/K) (R/K) × (R/K) motif appearing frequently as single or multiple copies and similar motifs had been known before.^{61,62} The importance of arginines has also been shown for a number of RNA binding proteins with arginine rich motifs.⁶³ In addition, a large-scale analysis of atomic contacts determined that the positively charged amino acid residues arginine and lysine are among the most frequent to interact with nucleic acids with high abundance of interaction with the phosphate backbone.^{64,65} Protein domains with multiple functional arginines are common in RNA recognition motifs, e.g., RGG and RGX often found in a modular composition.^{66,67} Furthermore, the existence of more functional amino groups in arginines relative to lysines explains their more stable electrostatic interaction with the phosphate groups in RNA.⁶⁸ The latter has been exploited as arginine-affinity chromatography to purify RNA.⁶⁹ Our *in vitro* binding analyses demonstrate that poly-R peptides and fusion proteins directly bind RNA, which is mostly abundant in the nucleolus due to massive ribosome production and much less in the nucleoplasm.⁷⁰⁻⁷³ Finally, we could show, using cells from different species, that these principles of nucleolar targeting are evolutionarily conserved.

References

- Cardoso MC, Schneider K, Martin RM, Leonhardt H. Structure, function and dynamics of nuclear subcompartments. *Curr Opin Cell Biol* 2012; 24:79-85; PMID:22227228; <http://dx.doi.org/10.1016/j.ccb.2011.12.009>
- Lamond AI, Earnshaw WC. Structure and function in the nucleus. *Science* 1998; 280:547-53; PMID:9554838; <http://dx.doi.org/10.1126/science.280.5363.547>
- Dundr M, Misteli T. Biogenesis of nuclear bodies. *Cold Spring Harb Perspect Biol* 2010; 2:a000711; PMID:21068152; <http://dx.doi.org/10.1101/cshperspect.a000711>
- Spector DL. Nuclear domains. *J Cell Sci* 2001; 114:2891-3; PMID:11686292
- Bickmore WA, Sutherland HG. Addressing protein localization within the nucleus. *Embo J* 2002; 21:1248-54; PMID:11889031; <http://dx.doi.org/10.1093/emboj/21.6.1248>
- Gorski SA, Dundr M, Misteli T. The road much traveled: trafficking in the cell nucleus. *Curr Opin Cell Biol* 2006; 18:284-90; PMID:16621498; <http://dx.doi.org/10.1016/j.ccb.2006.03.002>
- Mekhail K, Khacho M, Carrigan A, Hache RR, Gunaratnam L, Lee S. Regulation of ubiquitin ligase dynamics by the nucleolus. *J Cell Biol* 2005; 170:733-44; PMID:16129783; <http://dx.doi.org/10.1083/jcb.200506030>
- Eilbracht J, Schmidt-Zachmann MS. Identification of a sequence element directing a protein to nuclear speckles. *Proc Natl Acad Sci U S A* 2001; 98:3849-54; PMID:11274404; <http://dx.doi.org/10.1073/pnas.071042298>
- Lam YW, Trinkle-Mulcahy L, Lamond AI. The nucleolus. *J Cell Sci* 2005; 118:1335-7; PMID:15788650; <http://dx.doi.org/10.1242/jcs.01736>
- Raska I, Shaw PJ, Cmarko D. Structure and function of the nucleolus in the spotlight. *Curr Opin Cell Biol* 2006; 18:325-34; PMID:16687244; <http://dx.doi.org/10.1016/j.ccb.2006.04.008>
- Scheer U, Hock R. Structure and function of the nucleolus. *Curr Opin Cell Biol* 1999; 11:385-90; PMID:10395554; [http://dx.doi.org/10.1016/S0955-0674\(99\)80054-4](http://dx.doi.org/10.1016/S0955-0674(99)80054-4)
- Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI, Mann M. Nucleolar proteome dynamics.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We gratefully acknowledge Jenny Völger, Petra Domaing and Franziska Witzel for their help and support. We are indebted to Dawid Grzela and Zsuzsanna Izsvak for the Pac2 cells and Anup Arumughan for the yeast cells. We thank Heinrich Leonhardt for many fruitful discussions.

Funding

RMM was supported by a post-doctoral grant from Fundação para a Ciência e Tecnologia, Portugal (SFRH-BPD-66611-2009). This work was supported by grants of the German Research Council (DFG CA198/3) to MCC.

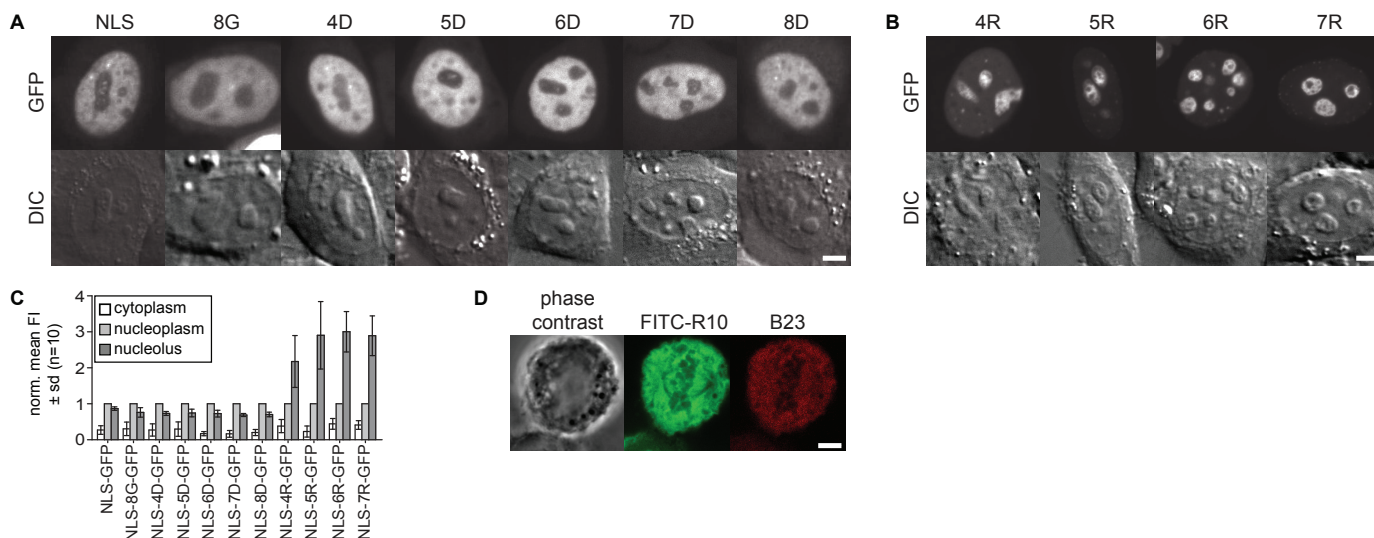
Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

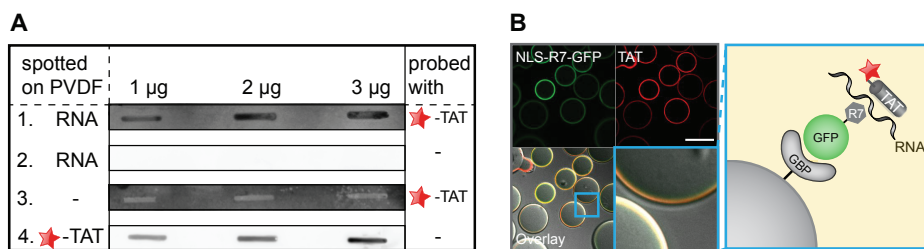
- Nature 2005; 433:77-83; PMID:15635413; <http://dx.doi.org/10.1038/nature03207>
- Grob A, Collieran C, McStay B. Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. *Genes Dev* 2014; 28:220-30; PMID:24449107; <http://dx.doi.org/10.1101/gad.234591.113>
- Cheutin T, O'Donohue MF, Beorchia A, Vandelaar M, Kaplan H, Defever B, Ploton D, Thiry M. Three-dimensional organization of active rRNA genes within the nucleolus. *J Cell Sci* 2002; 115:3297-307; PMID:12140261
- Derenzini M, Pasquini G, O'Donohue MF, Ploton D, Thiry M. Structural and functional organization of ribosomal genes within the mammalian cell nucleolus. *J Histochem Cytochem* 2006; 54:131-45; PMID:16204224; <http://dx.doi.org/10.1369/jhc.5R6780.2005>
- Dang CV, Lee WM. Nuclear and nucleolar targeting sequences of c-erbA, c-myc, N-myc, p53, HSP70, and HIV tat proteins. *J Biol Chem* 1989; 264:18019-23; PMID:2553699
- Hatanaka M. Discovery of the nucleolar targeting signal. *Bioessays* 1990; 12:143-8; PMID:2108666; <http://dx.doi.org/10.1002/bies.950120310>
- Scott MS, Boisvert FM, McDowall MD, Lamond AI, Barton GJ. Characterization and prediction of protein nucleolar localization sequences. *Nucleic Acids Res* 2010; 38:7388-99; PMID:20663773; <http://dx.doi.org/10.1093/nar/gkq653>
- Paine PL, Moore LC, Horowitz SB. Nuclear envelope permeability. *Nature* 1975; 254:109-14; PMID:1117994; <http://dx.doi.org/10.1038/254109a0>
- Lusk CP, Blobel G, King MC. Highway to the inner nuclear membrane: rules for the road. *Nat Rev Mol Cell Biol* 2007; 8:414-20; PMID:17440484; <http://dx.doi.org/10.1038/nrm2165>
- Gorlich D, Kutay U. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* 1999; 15:607-60; PMID:10611974; <http://dx.doi.org/10.1146/annurev.cellbio.15.1.607>
- Breuer M, Goldfarb DS. Facilitated nuclear transport of histone H1 and other small nucleophilic proteins. *Cell* 1990; 60:999-1008; PMID:1690602; [http://dx.doi.org/10.1016/0092-8674\(90\)90348-1](http://dx.doi.org/10.1016/0092-8674(90)90348-1)
- Dundr M, Misteli T, Olson MO. The dynamics of postmitotic reassembly of the nucleolus. *J Cell Biol* 2000; 150:433-46; PMID:10931858; <http://dx.doi.org/10.1083/jcb.150.3.433>
- Mekhail K, Rivero-Lopez L, Al-Masri A, Brandon C, Khacho M, Lee S. Identification of a common subnuclear localization signal. *Mol Biol Cell* 2007; 18:3966-77; PMID:17652456; <http://dx.doi.org/10.1091/mbc.E07-03-0295>
- Leung AK, Lamond AI. The dynamics of the nucleolus. *Crit Rev Eukaryot Gene Expr* 2003; 13:39-54; PMID:12839096; <http://dx.doi.org/10.1615/CritRevEukaryotGeneExpr.v13.i1.40>
- Grunwald D, Martin RM, Buschmann V, Bazett-Jones DP, Leonhardt H, Kubitscheck U, Cardoso MC. Probing Intracellular Environments at the Single Molecule Level. *Biophys J* 2008; 94:2847-58; PMID:18065482; <http://dx.doi.org/10.1529/biophysj.107.115014>
- Scott MS, Troshin PV, Barton GJ. NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC Bioinformatics* 2011; 12:317; PMID:21812952; <http://dx.doi.org/10.1186/1471-2105-12-317>
- Martin RM, Tunnemann G, Leonhardt H, Cardoso MC. Nucleolar marker for living cells. *Histochem Cell Biol* 2007; 127:243-51; PMID:17205309; <http://dx.doi.org/10.1007/s00418-006-0256-4>
- Lattig-Tunnemann G, Prinz M, Hoffmann D, Behlke J, Palm-Apergi C, Morano I, Herce HD, Cardoso MC. Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides. *Nat Commun* 2011; 2:453; PMID:21878907; <http://dx.doi.org/10.1038/ncomms1459>
- Matthias P, Muller MM, Schreiber E, Rusconi S, Schaffner W. Eukaryotic expression vectors for the analysis of mutant proteins. *Nucleic Acids Res* 1989; 17:6418; PMID:2549518; <http://dx.doi.org/10.1093/nar/17.15.6418>
- Leonhardt H, Rahn HP, Weinzierl P, Sporbert A, Cremer T, Zink D, Cardoso MC. Dynamics of DNA replication factories in living cells. *J Cell Biol* 2000; 149:271-80; PMID:10769021; <http://dx.doi.org/10.1083/jcb.149.2.271>
- Martin RM, Leonhardt H, Cardoso MC. DNA labeling in living cells. *Cytometry A* 2005; 67:45-52;

- PMID:16082711; <http://dx.doi.org/10.1002/cyto.a.20172>
33. Agarwal N, Hardt T, Brero A, Nowak D, Rothbauer U, Becker A, Leonhardt H, Cardoso MC. McCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res* 2007; 35:5402-8; PMID:17698499; <http://dx.doi.org/10.1093/nar/gkm599>
 34. Gochler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, Droege A, Lindenberg KS, Knoblich M, Haenig C, et al. A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell* 2004; 15:853-65; PMID:15383276; <http://dx.doi.org/10.1016/j.molcel.2004.09.016>
 35. McNeil PL. Incorporation of macromolecules into living cells. *Methods Cell Biol* 1989; 29:153-73; PMID:2643758; [http://dx.doi.org/10.1016/S0091-679X\(08\)60193-4](http://dx.doi.org/10.1016/S0091-679X(08)60193-4)
 36. Schermelleh L, Solovei I, Zink D, Cremer T. Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Chromosome Res* 2001; 9:77-80; PMID:11272795; <http://dx.doi.org/10.1023/A:1026799818566>
 37. Cardoso MC, Joseph C, Rahn HP, Reusch R, Nadal-Ginard B, Leonhardt H. Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication in vivo. *J Cell Biol* 1997; 139:579-87; PMID:9348276; <http://dx.doi.org/10.1083/jcb.139.3.579>
 38. Herce HD, Deng W, Helma J, Leonhardt H, Cardoso MC. Visualization and targeted disruption of protein interactions in living cells. *Nat Commun* 2013; 4:2660; PMID:24154492; <http://dx.doi.org/10.1038/ncomms3660>
 39. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H. A versatile nanotrapp for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* 2008; 7:282-9; <http://dx.doi.org/10.1074/mcp.M700342-MCP200>
 40. Molenaar I, Sillevius Smitt WW, Rozijn TH, Tonino GJ. Biochemical and electron microscopic study of isolated yeast nuclei. *Exp Cell Res* 1970; 60:148-56; PMID:5424309; [http://dx.doi.org/10.1016/0014-4827\(70\)90500-8](http://dx.doi.org/10.1016/0014-4827(70)90500-8)
 41. Taddei A, Schober H, Gasser SM. The budding yeast nucleus. *Cold Spring Harb Perspect Biol* 2010; 2:a000612; PMID:20554704; <http://dx.doi.org/10.1101/cshperspect.a000612>
 42. Yang CH, Lambie EJ, Hardin J, Craft J, Snyder M. Higher order structure is present in the yeast nucleus: autoantibody probes demonstrate that the nucleolus lies opposite the spindle pole body. *Chromosoma* 1989; 98:123-8; PMID:2673672; <http://dx.doi.org/10.1007/BF00291048>
 43. Blackmond DG. The origin of biological homochirality. *Cold Spring Harb Perspect Biol* 2010; 2:a002147; PMID:20452962; <http://dx.doi.org/10.1101/cshperspect.a002147>
 44. Kreil G. D-amino acids in animal peptides. *Annu Rev Biochem* 1997; 66:337-45; PMID:9242910; <http://dx.doi.org/10.1146/annurev.biochem.66.1.337>
 45. Nagata Y, Yamamoto K, Shimojo T, Konno R, Yasumura Y, Akino T. The presence of free D-alanine, D-proline and D-serine in mice. *Biochim Biophys Acta* 1992; 1115:208-11; PMID:1346751; [http://dx.doi.org/10.1016/0304-4165\(92\)90055-Y](http://dx.doi.org/10.1016/0304-4165(92)90055-Y)
 46. Eckert DM, Malashkevich VN, Hong LH, Carr PA, Kim PS. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* 1999; 99:103-15; PMID:10520998; [http://dx.doi.org/10.1016/S0092-8674\(00\)80066-5](http://dx.doi.org/10.1016/S0092-8674(00)80066-5)
 47. Zhao L, Lu W. Mirror image proteins. *Curr Opin Chem Biol* 2014; 22C:56-61; <http://dx.doi.org/10.1016/j.cbpa.2014.09.019>
 48. Kalderon D, Roberts BL, Richardson WD, Smith AE. A short amino acid sequence able to specify nuclear location. *Cell* 1984; 39:499-509; PMID:6096007; [http://dx.doi.org/10.1016/0092-8674\(84\)90457-4](http://dx.doi.org/10.1016/0092-8674(84)90457-4)
 49. Dultz E, Huet S, Ellenberg J. Formation of the nuclear envelope permeability barrier studied by sequential photoswitching and flux analysis. *Biophys J* 2009; 97:1891-7; PMID:19804719; <http://dx.doi.org/10.1016/j.bpj.2009.07.024>
 50. Tunnemann G, Martin RM, Haupt S, Patsch C, Edenhofer F, Cardoso MC. Cargo-dependent mode of uptake and bioavailability of TAT-containing proteins and peptides in living cells. *Faseb J* 2006; 20:1775-84; PMID:16940149; <http://dx.doi.org/10.1096/fj.05-5523com>
 51. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* 1996; 273:1392-5; PMID:8703075; <http://dx.doi.org/10.1126/science.273.5280.1392>
 52. Wu MM, Llopis J, Adams S, McCaffery JM, Kulomaa MS, Machen TE, Moore HP, Tsien RY. Organellar pH studies using targeted avidin and fluorescein-biotin. *Chem Biol* 2000; 7:197-209; PMID:10712929; [http://dx.doi.org/10.1016/S1074-5521\(00\)00088-0](http://dx.doi.org/10.1016/S1074-5521(00)00088-0)
 53. Casey JR, Grinstein S, Orłowski J. Sensors and regulators of intracellular pH. *Nat Rev Mol Cell Biol* 2010; 11:50-61; PMID:19997129; <http://dx.doi.org/10.1038/nrm2820>
 54. Herce HD, Garcia AE, Cardoso MC. Fundamental molecular mechanism for the cellular uptake of guanidinium-rich molecules. *J Am Chem Soc* 2014; 136:17459-67; PMID:25405895; <http://dx.doi.org/10.1021/ja507790z>
 55. Frankel AD, Pabo CO. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* 1988; 55:1189-93; PMID:2849510; [http://dx.doi.org/10.1016/0092-8674\(88\)90263-2](http://dx.doi.org/10.1016/0092-8674(88)90263-2)
 56. Schmidt-Zachmann MS, Nigg EA. Protein localization to the nucleolus: a search for targeting domains in nucleolin. *J Cell Sci* 1993; 105 (Pt 3):799-806; PMID:8408305
 57. Kuusisto HV, Wagstaff KM, Alvisi G, Roth DM, Jans DA. Global enhancement of nuclear localization-dependent nuclear transport in transformed cells. *Faseb J* 2012; 26:1181-93; PMID:22155563; <http://dx.doi.org/10.1096/fj.11-191585>
 58. Kaltschmidt E. Ribosomal proteins. XIV. Isoelectric points of ribosomal proteins of *E. coli* as determined by two-dimensional polyacrylamide gel electrophoresis. *Anal Biochem* 1971; 43:25-31; PMID:4943259; [http://dx.doi.org/10.1016/0003-2697\(71\)90103-5](http://dx.doi.org/10.1016/0003-2697(71)90103-5)
 59. Audas TE, Jacob MD, Lee S. The nucleolar detention pathway: A cellular strategy for regulating molecular networks. *Cell Cycle* 2012; 11:2059-62; PMID:22580471; <http://dx.doi.org/10.4161/cc.20140>
 60. Audas TE, Jacob MD, Lee S. Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. *Mol Cell* 2012; 45:147-57; PMID:22284675; <http://dx.doi.org/10.1016/j.molcel.2011.12.012>
 61. Emmott E, Hiscox JA. Nucleolar targeting: the hub of the matter. *EMBO Rep* 2009; 10:231-8; PMID:19229283; <http://dx.doi.org/10.1038/embor.2009.14>
 62. Horke S, Reumann K, Schweizer M, Will H, Heise T. Nuclear trafficking of La protein depends on a newly identified nucleolar localization signal and the ability to bind RNA. *J Biol Chem* 2004; 279:26563-70; PMID:15060081; <http://dx.doi.org/10.1074/jbc.M401017200>
 63. Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *Rna* 2005; 11:1848-57; PMID:16314457; <http://dx.doi.org/10.1261/rna.2167605>
 64. Treger M, Westhof E. Statistical analysis of atomic contacts at RNA-protein interfaces. *J Mol Recognit* 2001; 14:199-214; PMID:11500966; <http://dx.doi.org/10.1002/jmr.534>
 65. Lejeune D, Delsaux N, Charlotaux B, Thomas A, Bras-seur R. Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. *Proteins* 2005; 61:258-71; PMID:16121397; <http://dx.doi.org/10.1002/prot.20607>
 66. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 2007; 8:479-90; PMID:17473849; <http://dx.doi.org/10.1038/nrm2178>
 67. Rajyaguru P, Parker R. RGG motif proteins: modulators of mRNA functional states. *Cell Cycle* 2012; 11:2594-9; PMID:22767211; <http://dx.doi.org/10.4161/cc.20716>
 68. Woods AS, Ferre S. Amazing stability of the arginine-phosphate electrostatic interaction. *J Proteome Res* 2005; 4:1397-402; PMID:16083292; <http://dx.doi.org/10.1021/pr050077s>
 69. Martins R, Queiroz JA, Sousa F. New approach in RNA quantification using arginine-affinity chromatography: potential application in eukaryotic and chemically synthesized RNA. *Anal Biochem* 2013; 405:8849-58; PMID:24037617; <http://dx.doi.org/10.1007/s00216-013-7334-y>
 70. Baron-Benhamou J, Gehring NH, Kulozik AE, Hentze MW. Using the lambdaN peptide to tether proteins to RNAs. *Methods Mol Biol* 2004; 257:135-54; PMID:14770003
 71. Chen Y, Varani G. Protein families and RNA recognition. *FEBS J* 2005; 272:2088-97; PMID:15853794; <http://dx.doi.org/10.1111/j.1742-4658.2005.04650.x>
 72. Costa MC, Leitao AL, Enguita FJ. Biogenesis and Mechanism of Action of Small Non-Coding RNAs: Insights from the Point of View of Structural Biology. *Int J Mol Sci* 2012; 13:10268-95; PMID:22949860; <http://dx.doi.org/10.3390/ijms130810268>
 73. Lewis JD, Tollervey D. Like attracts like: getting RNA processing together in the nucleolus. *Science* 2000; 288:1385-9; PMID:10827942; <http://dx.doi.org/10.1126/science.288.5470.1385>

Supplementary Data



Supplementary Figure 1. Intracellular distribution of peptide tagged GFP tracer proteins in living human cells during interphase and metaphase. The images in this figure show live human HeLa cells transfected with NLS-GFP constructs also shown in Figure 2 to validate the same distribution pattern of the charged GFP tracer proteins in a different cell type of a different species. The first panel in A) displays the intracellular distribution of the NLS-GFP control and poly-G coupled NLS-GFP constructs. The second panel shows images of all poly-D containing NLS-GFP constructs. B) The images display the poly-R coupled NLS-GFP constructs. In A) and B) the upper row shows the GFP fluorescence with the corresponding differential interference contrast (DIC) image below. The corresponding quantification of the mean fluorescence intensity for each 10 cells from two independent experiments is plotted in C). Area selection and quantification were done as described for peptides and protein before. The image panel D) shows a fixed HeLa cell in mitosis labeled with FITC-R10 peptide, which highlights the chromosomes as darker regions in the cell by exclusion of the peptide. The cell was further stained with antibodies against the nucleolar antigen B23/nucleophosmin to demonstrate the lack of nucleoli. The FITC-R10 peptide and B23 show a homogeneous distribution throughout the cell with exclusion from the chromosomes aligned in the central plane of the cell. Scalebars: 5 μ m.



Supplementary Figure 2. Analyses of total RNA binding to arginine-rich cell penetrating peptides. The slot blot in A) shows the binding of a standard arginine-rich cell penetrating peptide (TAT) in the absence (3) and presence (1) of total RNA. Slots lacking RNA (3) do not show binding of TAT, while slots probed with increasing amounts of RNA show increasing TAT binding (1). Row 4 with spotted TAT peptide shows, when compared to row 2, that only the presence of labeled peptide results in fluorescence detection but not RNA itself. B) In vitro RNA pulldown assay using NLS-R7-GFP immobilized to sepharose beads via the GFP binding protein (GBP) (scheme). RNA was stained using fluorescently labeled TAT. TAT binding to RNA was imaged by confocal microscopy. Scale bars: 100 μ m.

Supplementary Table 1. List of oligonucleotides used to generate the GFP fusions.

GFP version	primer	sequence (5'-3')
4R-GFP	Fw	CCGGTGCACCGCCGGCGCCGG
	Rev	CCGGCCGGCGCCGGCGGTGCA
5R-GFP	Fw	CCGGTGCACCGCCGGCGCCGG
	Rev	CCGGCCGGCGGGCGCCGGCGGTGCA
6R-GFP	Fw	CCGGTGCACCGCCGGCGCCGGCCGG
	Rev	CCGGCCGGCGGGCGGGCGCCGGCGGTGCA
7R-GFP	Fw	CCGGTGCACCGCCGGCGCCGGCCGGCCGG
	Rev	CCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGTGCA
4D-GFP	Fw	CCGGTGCACGATGACGAT
	Rev	CCGGATCGTCATCGTCGA
5D-GFP	Fw	CCGGTGCACGATGACGATGAC
	Rev	CCGGGTCATCGTCATCGTCGA
6D-GFP	Fw	CCGGTGCACGATGACGATGACGAT
	Rev	CCGGATCGTCATCGTCATCGTCGA
7D-GFP	Fw	CCGGTGCACGATGACGATGACGATGAC
	Rev	CCGGGTCATCGTCATCGTCATCGTCGA
8D-GFP	Fw	CCGGTGCACGACGACGACGACGACGACGAC
	Rev	CCGGGTCGTCGTCGTCGTCGTCGTCGTCGA
8G-GFP	Fw	CCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGTGCA
	Rev	CCGGTGCACCGCCGGCGCCGGCGCCGGCGCCGG

Supplementary Table 2. Lists of normalized mean fluorescence intensity values measured for the fluorescent peptides and proteins in cytoplasm, nucleoplasm and nucleoli and their statistical significance (P-values).

A

	K5	K6	K7	K8	K9	K10	K11	K12
cytoplasm	0.73	0.73	0.76	0.77	0.72	0.73	0.73	0.73
nucleus	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
nucleolus	1.06	1.06	1.04	1.04	1.06	1.02	1.05	1.06
p-value	0.068	0.199	0.290	0.326	0.199	0.762	0.096	0.016*

B

	R5	R6	R7	R8	R9	R10	R11	R12
cytoplasm	0.613	0.702	0.706	0.715	0.732	0.676	0.649	0.663
nucleus	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
nucleolus	1.045	1.178	1.155	1.162	1.223	1.162	1.123	1.141
p-value	0.406	0,002**	0,010**	0,005**	0,004**	0,034*	0,010**	0,041*

C

	L-R10	D-R10
cytoplasm	0.68	2.88
nucleus	1.00	1.00
nucleolus	1.16	3.72
p-value	0,034*	0,0002**

D

	GFP	NLS-GFP	8G-GFP	4D-GFP	5D-GFP	6D-GFP	7D-GFP	8D-GFP	4R-GFP	5R-GFP	6R-GFP	7R-GFP
cytoplasm	0.90	0.39	0.25	0.19	0.23	0.15	0.23	0.27	0.25	0.08	0.09	0.10
nucleus	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
nucleolus	0.70	0.68	0.63	0.67	0.66	0.60	0.64	0.51	1.74	2.45	2.67	2.71
p-value	0,0101*	0,0002**	0,0039**	0,0233*	0,0101*	0,0101*	0,0002**	0,0163*	0,0304*	0,0005**	0,0002**	0,0090**

E

		NLS-GFP	8G-GFP	4D-GFP	5D-GFP	6D-GFP	7D-GFP	8D-GFP	4R-GFP	5R-GFP	6R-GFP	7R-GFP
cytoplasm		0.30	0.30	0.30	0.30	0.20	0.20	0.20	0.40	0.20	0.40	0.40
nucleus		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
nucleolus		0.90	0.80	0.70	0.70	0.70	0.70	0.70	2.20	2.90	3.00	2.90
p-value		0.44 935	0.13 057	0.19 86	0.17 362	0.01 0*	0.05 87	0.09 618	0.01 39*	0.00 25**	0.00 02**	0.00 03**

The p-value was calculated for the nucleoplasm versus the nucleolus mean intensity measurements only with a non-parametric Kruskal-Wallis-ANOVA test. The p-values represent a measure of significance for the mean intensity values from the nucleoplasm and nucleolus belonging to the same or different population of measurements. Significant different distributions are found for nucleolar accumulation but also nucleolar exclusion of proteins. At the level of $p < 0.05$ the measurements of mean fluorescence intensity in nucleoplasm and nucleoli for one peptide or protein are significantly different and indicated by asterisk (*). Highly significant differences ($p < 0.01$) are indicated by double asterisk (**). A) shows the values for the FITC labeled poly-K peptides K5 to K12. B) lists values for the FITC labeled poly-R peptides R5 to R12. C) shows the mean fluorescence intensity and p-values for the comparison of L-R10 versus D-R10 peptide distribution. Note that the high levels of accumulation of the D-R10 peptide in cytoplasm and nucleolus reflect the low intensity measured in the nucleoplasm, which is used as the reference level to normalize for nucleolar accumulation. D) shows the normalized fluorescence intensity and p-values for the mean fluorescence of GFP and all versions of it with NLS as well as additional fusion of poly-G, poly-D and poly-R sequences in mouse C2C12 cells. In E) are the corresponding values for the mean fluorescence in human HeLa cells.

For **Supplementary Table 3**, go to:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4615656/>

4.4 Visualization of the Nucleolus in Living Cells with Cell-Penetrating Fluorescent Peptides

A.K. Ludwig revised the manuscript.

A.K. Ludwig prepared figure 2 and wrote the respective figure legend.

Content of figures:

Figure 2: Intracellular distribution of nucleolar markers in living cells of various species and origin.

Chapter 6

Visualization of the Nucleolus in Living Cells with Cell-Penetrating Fluorescent Peptides

Robert M. Martin, Henry D. Herce, Anne K. Ludwig,
and M. Cristina Cardoso

Abstract

The nucleolus is the hallmark of nuclear compartmentalization and has been shown to exert multiple roles in cellular metabolism besides its main function as the place of ribosomal RNA synthesis and assembly of ribosomes. The nucleolus plays also a major role in nuclear organization as the largest compartment within the nucleus. The prominent structure of the nucleolus can be detected using contrast light microscopy providing an approximate localization of the nucleolus, but this approach does not allow to determine accurately the three-dimensional structure of the nucleolus in cells and tissues. Immunofluorescence staining with antibodies specific to nucleolar proteins albeit very useful is time consuming, normally antibodies recognize their epitopes only within a small range of species and is applicable only in fixed cells. Here, we present a simple method to selectively and accurately label this ubiquitous subnuclear compartment in living cells of a large range of species using a fluorescently labeled cell-penetrating peptide.

Key words Cell permeable nucleolar marker, Cell-penetrating peptide, Fluorescence microscopy, Living cells, Nucleolus, Poly-arginine

1 Introduction

In 1831, a Scottish botanist Robert Brown described for the first time the “nuclei” in plant cells [1]. A few years later, in 1836, Gabriel Gustav Valentin working with neuronal cells described within nuclei a prominent substructure later termed “nucleolus” [2]. It was only one century later, in the 1960s, that findings derived from electron microscopic autoradiography of [³H]uridine incorporation established the major function of the nucleolus in rRNA synthesis, rRNA processing, and ribosome biogenesis [3, 4]. Several other tasks have been more recently added to the functional portfolio of the nucleolus (reviewed in [5, 6]).

The nucleolus is a self-organizing structure not delimited by membranes that gets disassembled and reassembled around the rDNA loci at every cell division cycle ([7]; reviewed in [8]).

Transcription of ribosomal genes, which make up the fibrillar centers, takes place surrounding this structure and creates the so-called dense fibrillar component [9, 10]. Further processing of the precursor RNA and multiple proteins gives rise to ribosome precursor particles, which form the granular component (reviewed in [11]). Over 300 nucleolar proteins have been identified by proteomics [12]. The first nucleolar targeting sequences were identified in viral proteins ([13, 14]; reviewed in [15]) and further shown to be able to cross cellular membranes [16, 17]. Recently, we performed a systematic analysis of peptide sequences sufficient for nucleolar targeting [18]. We found that six or more arginines are sufficient to label the nucleolus by binding to RNA. Hence, these peptide sequences can label the nucleolus in all species tested (insect, fish, mouse, human, and very likely also yeast). Importantly, in particular nona- and deca-arginine peptides are similarly highly active in cell membrane permeation [19].

In this chapter, we describe how adding these peptides conjugated to a fluorescent label directly to the cell culture medium without the need for transient or stable transfections of the cells provides a live cell fluorescent nucleolar marker in a variety of species [20]. The nucleolar marker is membrane permeable and reaches its target within minutes. To prevent proteolytic degradation the peptide can be synthesized with D-amino acids. The cell-penetrating-peptide-based marker provides a method for the noninvasive marking of the nucleolus in living cells, allowing the three-dimensional reconstruction of the nucleolus, which is not possible with contrast microscopy techniques. Besides that, it can easily be used in combination with different fluorophores like fluorescent proteins or DNA dyes to correlate different aspects of nuclear structures with the nucleolus. In addition, nucleolar labeling is also preserved during fixation and staining of the cells. Furthermore, it does not affect cellular viability, proliferation, and rDNA transcription.

2 Materials

Prepare all solutions at room temperature unless otherwise noted. Solutions used to culture and prepare cells should be warmed to 37 °C prior to use. All solutions and materials in contact with live cells must be sterile and have to be handled inside a tissue culture biosafety cabinet. Disposal of reagents, solutions, and cell material has to be carried out following the specific regulations in effect.

2.1 Cells and Materials for Cell Culture

1. Human HeLa cervix epithelium adenocarcinoma cells.
2. Adult rat ventricular cardiomyocytes.
3. Mouse C2C12 myoblast cells.
4. Pac2 zebrafish fibroblast cells (*Danio rerio*).
5. Sf9 insect ovary cells (*Spodoptera frugiperda*).

6. High-glucose (4.5 g/L) Dulbecco's Modified Essential Medium (DMEM).
7. Leibovitz L-15 medium.
8. EX-CELL® 420 Insect Serum Free medium.
9. Glutamine.
10. Gentamicin.
11. Penicillin/Streptomycin 100×.
12. Fetal calf serum (FCS). Aliquots of 50 mL are stored at -20°C .
13. Standard medium for HeLa cell culture: DMEM (**item 6**) was supplemented with 5 mM L-glutamine (**item 9**), 5 $\mu\text{g}/\text{mL}$ gentamycin (**item 10**), and 10% FCS (**item 12**). Prepare a 500 mL bottle under sterile conditions and store at 4°C .
14. Standard medium for adult rat ventricular cardiomyocytes was 10 mM HEPES buffer with 0.3 mM Ca^{2+} and 0.5% bovine serum albumin stored at 4°C .
15. Standard medium for C2C12 cell culture: DMEM (**item 6**) was supplemented with 5 mM L-glutamine (**item 9**), 5 $\mu\text{g}/\text{mL}$ gentamycin (**item 10**), and 20% FCS (**item 12**). Prepare a 500 mL bottle under sterile conditions and store at 4°C .
16. Standard medium for Pac2 cells: Leibovitz L-15 medium (**item 7**) was supplemented with 15% FCS (**item 12**) and 1% penicillin-streptomycin (**item 11**) from the 100× stock solution and stored at 4°C .
17. Standard medium for Sf9 cells: EX-CELL® 420 Insect Serum Free medium (**item 8**) was supplemented with 5 mM L-glutamine (**item 9**) and 10% FCS (**item 12**) and stored at 4°C .
18. Phosphate buffered saline (PBS) supplemented or not with 0.5 mM EDTA (PBS/EDTA).
19. HEPES buffer 1 M, pH 7.2–7.5.
20. 0.25% trypsin in PBS/EDTA prepared from trypsin powder.

2.2 Peptides

Deca-arginine peptides containing amino-terminal fluorescein label (FITC-R₁₀) were synthesized using D-amino acids by Peptide Specialty Laboratories (Heidelberg, Germany). After synthesis the peptides were HPLC purified and their molecular weight was confirmed by mass spectrometry.

Ultrapure water is used to resuspend the lyophilized peptides.

2.3 Chamber Systems and Coverslips for Live Cell Labeling of Nucleoli

1. μ -dish 35 mm low with Ibidi standard bottom (Ibidi®).
2. Chambered coverslip μ -slide 8 well (Ibidi®).
3. Chambered coverglass 4 and 8 well (Thermo Fisher Scientific, Nunc™ Lab-Tek™).
4. Coverslips, 12 mm diameter.
5. Parafilm®.

2.4 Microscopy**2.4.1 Confocal Laser Scanning Microscope**

For live cell microscopy any suitable confocal microscope system can be used. We use a Zeiss LSM510Meta confocal setup mounted on an Axiovert 200 M inverted microscope equipped with a 63× phase contrast plan-apochromat oil objective NA1.4. During all acquisitions, the main beam splitter was a HFT UV/488/543/633 and the parameters for the detection of FITC were as follows:

Excitation wavelength: 488 nm.

Emission filter: BP500-530 nm.

2.4.2 Environmental Chamber and Stage Incubation System

For live cell microscopy any suitable microscope incubation system can be used. The microscope is housed in an environmental cage incubator, which is connected to a heating system providing a constant temperature of 37 °C. In addition, a chambered stage incubation system is used that provides a temperature of 37 °C and humidified 5% CO₂ atm (all incubation components from Okolab, Ottaviano, Italy). Cells grown in chambered coverslip systems are kept inside the stage incubation system during imaging to maintain stable environmental conditions.

2.5 Image Analysis

For the analysis of microscopic images any suitable software package can be applied. We used ImageJ (<http://imagej.nih.gov/ij/>) to measure, e.g., relative fluorescence intensities in nucleoli compared to the nucleoplasm.

3 Methods**3.1 Cell Culture****3.1.1 Maintaining Mammalian Cells in Culture**

1. HeLa and C2C12 cells (**items 1 and 3** in Subheading **2.1**) are grown in p100 dishes in the cell culture incubator at 37 °C and 5% CO₂ atm, until they reach 70–80% confluence. Adult rat ventricular cardiomyocytes primary cells (**item 2** in Subheading **2.1**) cannot be maintained in culture and are used the day they are isolated.
2. Cells are grown in standard medium as indicated in Subheading **2.1**.
3. To split cells, aspirate medium and rinse with 1× PBS/EDTA prewarmed to 37 °C.
4. Directly apply 1.0 mL of Trypsin–EDTA (**item 20** in Subheading **2.1**) solution prewarmed to 37 °C to the cell layer and tilt the dish gently. Leave flask in the incubator for approximately 1–3 min.
5. Hit the side of the dish with the palm of one hand several times to detach the cells, avoid spilling of fluid to the side and lid of the dish.

6. When all cells have detached, add 9.0 mL of growth medium to the flask and resuspend the cells by gentle pipetting up and down several times. Avoid creating bubbles and foam. Transfer between 1.0 (1:10) and 2.0 (1:5) mL to a new cell culture dish.
7. Fill up to a final volume of 10 mL with standard medium.
8. Place the dish in a cell culture incubator and examine the cells daily using a cell culture microscope.

3.1.2 *Maintaining Fish Cells in Culture*

1. Pac2 cells (**item 4** in Subheading 2.1) are grown in p25 flasks in a cell culture incubator at 28 °C without CO₂ enriched atmosphere.
2. Pac2 cells are grown in standard medium (**item 16** in Subheading 2.1).
3. To split cells, aspirate medium and rinse twice with 1× PBS/EDTA (**item 18** in Subheading 2.1).
4. Directly apply sufficient trypsin–EDTA solution (**item 20** in Subheading 2.1) to cover the cell layer, tilt the dish gently, and incubate for 5 min at room temperature.
5. Dilute the detached cells with standard medium and pipette up and down several times to completely detach all cells and disperse cell clots.
6. Transfer 20% of the volume to a new flask and fill up with standard medium and place in the incubator at 28 °C.

3.1.3 *Maintaining Insect Cells in Culture*

1. Sf9 cells (**item 5** in Subheading 2.1) are maintained in p25 flasks as suspension culture in an orbital shaker with ambient atmosphere at 28 °C.
2. For labeling experiments with microscopic observation, cells are grown in an incubator with ambient atmosphere at 28 °C.
3. Sf9 cells are grown in standard medium (**item 17** in Subheading 2.1).
4. To subculture cells, resuspend cells gently by using a plastic pipette and transfer 1.0 mL to a new flask, fill up with standard medium and place in the orbital shaker at 28 °C.

3.1.4 *Preparation of Peptides for the Labeling of Live Cells*

1. Peptides are delivered lyophilized and therefore have to be resuspended.
2. Add sterile deionized ultrapure water to the lyophilized peptides to prepare peptide stock solutions at a final concentration of 1 mM, vortex to completely dissolve and homogenize the peptide solution.
3. Spin down peptide solution and prepare aliquots of 50 µL in sterile Eppendorf tubes and store at –20 °C until further use.

3.2 Preparation of Cells and Peptides for Nucleolar Labeling and Imaging

3.2.1 Culturing Cells in Live Cell Chamber Systems

1. Take live cell chambers out of sterile packaging and place into an empty sterile p100 dish (*see Note 1*).
2. Wash, trypsinize, and resuspend HeLa and C2C12 cells as described in Subheading 3.1.1.
3. Wash, trypsinize, and resuspend Pac2 cells as described in Subheading 3.1.2.
4. Use directly the suspension culture of Sf9 cells as described in Subheading 3.1.3, pipette up and down to disperse cells.
5. Seed equal volumes of cell suspension into each well of a sterile live cell chamber system at 60–70% confluence.
6. Typically 100 μL of cell suspension were seeded into one well of an 8-well LabTek or Ibidi chamber and correspondingly 200 μL were seeded into 4 well chambers. These amounts correspond to a split ratio of around 1: 2.5 taking into account the dish and chamber surface areas with cells grown to 70–80% confluence. Amounts of cells seeded have to be adapted to the respective cell type that will be used.
7. Fill up chambers up to half of the volume with growth medium and close the lid. To avoid contamination on the way to the incubator, close additionally the p100 dish containing the live cell chamber.
8. Human HeLa, mouse C2C12, and Pac2 fish cells are grown under the conditions described in Subheadings 3.1.1 and 3.1.2, respectively. Sf9 insect cells, however, are grown in an incubator with ambient atmosphere at 28 °C without shaking to allow attachment of the cells to the chamber surface.
9. Cells are grown over night and used the next day for nucleolar labeling with the transducible peptide marker (*see Note 2*).

3.2.2 Preparation of Coverslips for Fixation of Cells After Labeling the Nucleoli

1. Place 12 mm coverslips in a dedicated holder inside a closed container and immerse in pure 100% ethanol for several hours.
2. Use fine tip forceps to place the desired number of coverslips into an empty sterile cell culture dish.
3. Leave the dish with the lid half open inside the laminar flow cabinet until all traces of ethanol evaporated.
4. Immediately before starting to split cells to seed onto coverslips, use 1 \times PBS and fill carefully into the dish with the coverslips until all coverslips are covered. Avoid coverslips to adhere to each other by applying PBS slowly and carefully at the side of the dish without moving the coverslips.
5. Aspirate PBS and continue with washing, trypsinizing, and resuspending cells as described in Subheadings 3.1.1, 3.1.2, and 3.1.3, respectively.

6. Seed the required volume of resuspended cells to reach 70% confluence and fill up with growth medium. Avoid disturbance to prevent coverslips floating around.
7. Place the dish inside the cell culture incubator at 37 °C and 5% CO₂ until use, typically on the next day.

**3.2.3 Labeling
of the Nucleoli in Living
Cells Using a Cell-
Penetrating Peptide Marker
(See Fig. 1)**

1. Dilute deca-arginine peptides in DMEM without FCS to a final concentration of 10 μM just before the start of the labeling experiments and keep at 4 °C until use (*see Notes 3 and 4*).
2. For imaging experiments place cells grown in chambered coverslips or microscopy dishes inside the stage incubation system, fix using metal clips or similar devices to prevent movements and leave for at least 30 min to adapt to the environment and observe cells by phase contrast microscopy.
3. Remove all growth medium from the first well to be labeled and immediately apply with great care the peptide in DMEM solution (*see Note 5*).
4. Close the chamber lid and the stage incubation system. Adjust the microscope objective to focus the cells. It is possible to observe cells at this stage by microscopy; however, excess labeled peptide and the high concentration of fluorophores in the medium will cause very high intensities when observing the peptide fluorescence.
5. Incubate the cells for 30 min to 1 h with the peptides in DMEM. Aspirate all peptide solution, wash cells one time in DMEM without FCS or alternatively in 1× PBS to remove excess fluorescent peptides, aspirate the washing solution, and apply the standard growth medium to the cells.
6. Close the chambered coverslip and the stage incubation system, focus the cells, and start the experimental observations and image acquisition.
7. For cells grown on coverslips, aspirate medium carefully from the dish and apply the peptides in DMEM at the side of the dish, not directly on the coverslips (*see Note 6*). After removal of the peptide in DMEM solution use forceps to place the coverslips with cells turned upward on Parafilm in a humidified chamber for further treatment.
8. Perform the washing steps with 1× PBS on individual coverslips in the humidified chamber and proceed with a protocol for cell fixation, permeabilization, and immunostaining (*see Note 7*).
9. Alternatively, the application of peptides in DMEM, as well as the washing step can be performed in a laminar flow biosafety cabinet. In this case, cells are kept in the cell culture incubator for the peptide uptake incubation. Microscopic observation can start after labeling or any time later (*see Note 8*).

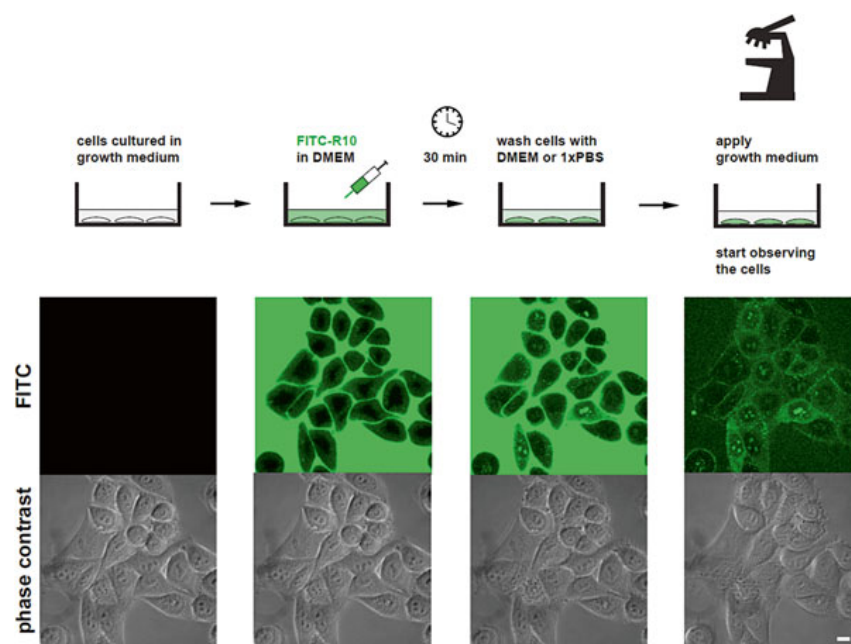


Fig. 1 Application of a fluorescent cell-penetrating-peptide-based nucleolar marker. The *upper row* shows schematically the steps for the labeling of nucleoli in live cells with the cell-penetrating peptide marker. Cells are cultivated in microscopy dishes and after removal of the growth medium incubated with the peptide marker diluted in DMEM. After an incubation time of about 30 min, the peptide solution is removed and cells are washed with fresh DMEM or PBS to remove excess fluorescent peptides. Finally the cells are supplied with standard growth medium followed by immediate microscopic observation of the fluorescently labeled nucleoli. The *lower panel* shows exemplary images of a culture of human HeLa cells at the different stages of the labeling process depicted in the scheme earlier. The *upper image row* shows FITC fluorescence of the cell-penetrating peptide marker with the corresponding phase contrast image below. Note the lack of signals inside the cell after application of the fluorescent peptide in the second step, while after 30 min of incubation in step 3, nucleolar labeling is already visible. After washing away excess peptide, the fluorescently labeled nucleoli are distinctively visible in the vast majority of cells against the darker nucleoplasmic surrounding. Scale bar 10 μm

10. For cells in suspension culture, we suggest mild centrifugation steps to remove growth medium, resuspension of cells in DMEM with diluted peptides followed by mild centrifugation for peptide removal and washing steps.
11. Additional labeling of cells with cell permeable dyes like DRAQ5 [21] can be performed before or after peptide labeling of nucleoli (*see Note 9*).

3.3 Imaging (See Fig. 2)

1. Place the dish containing the cells treated with the nucleolar marker on the microscope stage. Close the incubation chamber and wait 30–60 min to allow the sample to acclimatize to the chamber temperature before starting with image acquisition. This will avoid focus drift while imaging.

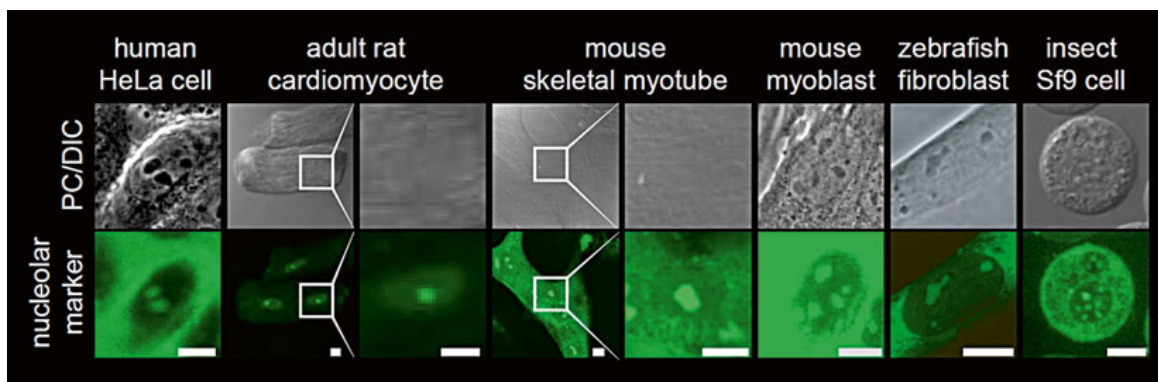


Fig. 2 Intracellular distribution of nucleolar markers in living cells of various species and origin. Shown are mid-optical sections of cultured cell lines and terminally differentiated primary cells as indicated earlier. For visualization of nucleoli in situ, cells were treated for 30 min with medium containing the nucleolar marker, washed in PBS, and incubated in medium for live cell imaging on a confocal microscope equipped with a climatization chamber to maintain a constant temperature of 37 °C, 5% CO₂, and 60% humidity. The *lower image row* shows FITC fluorescence of the cell-penetrating peptide marker with the corresponding phase/differential interference contrast (PC/DIC) image below. Scale bar 10 μm

2. Focus the cells using the transmitted light channel (i.e., differential interference contrast (DIC), phase contrast, etc.).
3. Adjust the microscope settings, in particular, the 488 nm laser wavelength intensity and exposure time to optimize the visualization of the nucleoli (minimizing overexposure of the sample). To capture the entire volume of the nucleolus it is required to image optical sections that span approximately $-4 \mu\text{m}$ to $7 \mu\text{m}$ from the focal plane (for HeLa cells, *see* **Note 10**).
4. Image the cells.
5. Export the images as tiff files for further processing using ImageJ.

3.4 Image Analysis (See Fig. 3)

Segmentation of Nucleoli

1. Open the image stack (or single image) in ImageJ.
2. Select the cell of interest using the “rectangular” selection option in the tools menu bar.
3. Crop (“Image” menu > “Crop”), Fig. 3a.
4. Separate the stack into single images (“Image” menu > “Stacks” > “Stacks to Images”), Fig. 3b.
5. Close without saving the images that do not contain any fraction of the nucleus. The marker distributes strongly in the cytosol and nucleoli and is weakly distributed over the nucleoplasm. This can be used to easily recognize the nucleus without the need of any extra nuclear marker.

For Each Stack Image

6. Select the nucleus using the “Freehand selections” in the tools menu bar, Fig. 3c.

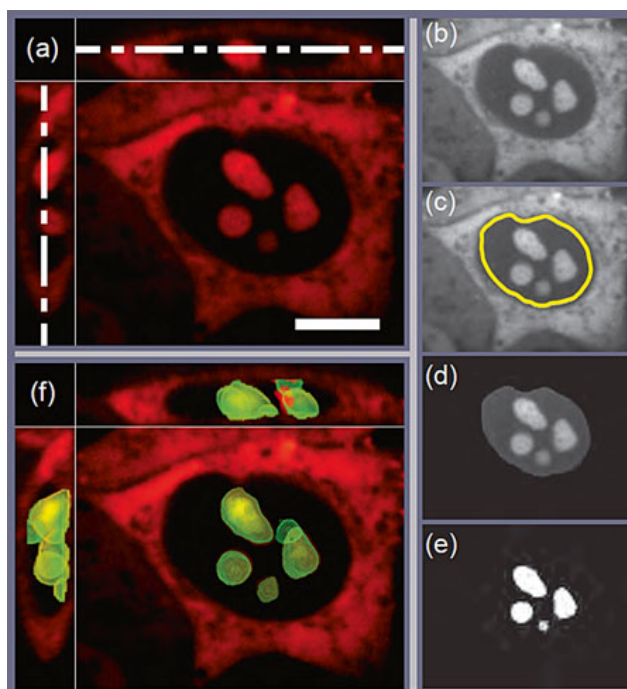


Fig. 3 Image analysis steps for segmentation of nucleoli. **(a)** Image stack of a single HeLa cell visualized using Volocity. *xy* as well as *xz* and *yz* views of the cell are shown. **(b)** Example image from *z*-stack for segmentation. **(c)** Nucleus selection (*yellow line*). **(d)** Clearance of signal outside the nucleus. **(e)** Binary image of the nucleoli. **(f)** Overlay of image stack in **(a)** and a mesh representation of the surface of the 3D rendered nucleoli. Scale bar 10 μm

7. Clear the marker signal outside the nucleus (“Edit” menu > “Clear Outside”), Fig. 3d.
8. Convert the image to a binary form to isolate the nucleoli region (“Process” menu > “Binary” > “Make Binary” and use the “Otsu” filter (*see Note 11*), Fig. 3e).
9. For further processing each image can be used individually as a mask for the nucleolus or the images can be collected into a single stack (“Image” menu > “Stacks” > “Images to Stacks”).
10. Save resulting individual images or image stack as tiff (click selected image and “File” menu > “Save as” > “tiff”).
11. The 3D stack can be visualized using the ImageJ plug-in “3D Viewer” (“Plugins” menu > 3D > “Volume Viewer”), select view angle and options and save as an image “Save View,” Fig. 3f.

4 Notes

1. We place the chambered coverslip systems and microscopy dishes inside p100 cell culture dishes for better handling, easy transport, and sterility between laminar flow biosafety cabinet, incubator, and microscope incubator.

2. For primary cells, e.g., freshly isolated cardiomyocytes it is recommended to perform the peptide labeling and washing steps as soon as possible after preparation of the sample due to their very limited lifespan [20]. In the case of other primary cell cultures with longer survival time, it might be possible to delay the labeling after sample preparation.
3. The efficiency of uptake of cell-penetrating peptide markers is a function of peptide length and concentration but is also cell type specific. A concentration of 10 μM FITC labeled deca-L-arginine was sufficient to label a range of different cell types from different species (*see* Fig. 2 and [18]). However, for certain cell types the concentration of FITC-R₁₀ might be elevated up to 50 μM to achieve efficient nucleolar labeling with a low to moderate rate of toxicity [22]. Higher concentrations of R₁₀ peptide might result in high frequency of damaged or dead cells.
4. For the peptide transduction we prepare a volume of peptides diluted in DMEM sufficient to completely cover all cells in the respective number of chambers to be labeled. Typically in a single well of a 4-well chamber it is sufficient to apply 200 μL of diluted peptide in DMEM to cover all cells completely and provide enough volume to prevent dry spots in the center of the well due to the fluid meniscus. Accordingly, a volume of 100 μL is sufficient in one well of an 8-well chamber.
5. When applying the peptide diluted in DMEM to the cells, place the tip of the pipette into one corner of the well, close to the bottom but not touching it and slowly release the fluid to immerse the cells. Avoid drops of liquid falling on the cells to prevent detaching of cells. When the chamber bottom surface is covered with liquid, add the rest of the volume slightly faster but still with great care. Perform washing steps and the final application of growth medium in the same way.
6. For increased efficiency of peptide uptake into cells to be fixed, we incubated cells with the peptide up to 1 h.
7. Membrane permeable peptides are not recommended for labeling of nucleoli postfixation [20].
8. The persistence of the nucleolar labeling in living cells was successfully observed up to 24 h after the labeling event.
9. For additional labeling of cellular components by transient transfection of cells with, e.g., plasmids encoding proteins tagged with red fluorescent proteins, we suggest to perform the transfection the day before the nucleolar labeling with the peptides.
10. This range depends significantly on the specific cell type and stage of the cell cycle.
11. Most filters (e.g., “MaxEntropy,” “Mean,” etc.) can be used here with very similar results.

Acknowledgements

We thank all present and past members of the laboratory for their contributions over the years. Robert M. Martin is supported by a fellowship of the Fundação para a Ciência e Tecnologia, Portugal (SFRH/BPD/66611/2009). The laboratory of M. Cristina Cardoso is supported by grants of the German Research Foundation (DFG) and the Federal Ministry of Education and Research (BMBF).

References

1. Brown R (1831) Observations on the organs and mode of fecundation in Orchideae and Asclepiadae. *Trans Linnear Soc Lond* 16(3):709–737
2. Valentin G (1836) *Repertorium für Anatomie und Physiologie*, vol 1. Verlag von Veit und Comp, Berlin
3. Brown DD, Gurdon JB (1964) Absence of ribosomal Rna synthesis in the anucleolate mutant of *Xenopus Laevis*. *Proc Natl Acad Sci U S A* 51:139–146
4. Granboulan N, Granboulan P (1965) Cytochimic ultrastructurale du nucleole. II. Etude des sites de synthese du RNA dans le nucleole et le noyau. *Exp Cell Res* 38:604–619
5. Pederson T (1998) The plurifunctional nucleolus. *Nucleic Acids Res* 26(17):3871–3876
6. Pederson T (2011) The nucleolus. *Cold Spring Harb Perspect Biol* 3(3):a000638. doi:10.1101/cshperspect.
7. Dundr M, Misteli T, Olson MO (2000) The dynamics of postmitotic reassembly of the nucleolus. *J Cell Biol* 150(3):433–446
8. Hernandez-Verdun D (2011) Assembly and disassembly of the nucleolus during the cell cycle. *Nucleus* 2(3):189–194. doi:10.4161/nucl.2.3.16246
9. Cheutin T, O'Donohue MF, Beorchia A, Vandelaer M, Kaplan H, Defever B, Ploton D, Thiry M (2002) Three-dimensional organization of active rRNA genes within the nucleolus. *J Cell Sci* 115(Pt 16):3297–3307
10. Derenzini M, Pasquinelli G, O'Donohue MF, Ploton D, Thiry M (2006) Structural and functional organization of ribosomal genes within the mammalian cell nucleolus. *J Histochem Cytochem* 54(2):131–145
11. Scheer U, Hock R (1999) Structure and function of the nucleolus. *Curr Opin Cell Biol* 11(3):385–390
12. Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI, Mann M (2005) Nucleolar proteome dynamics. *Nature* 433(7021):77–83
13. Dang CV, Lee WM (1989) Nuclear and nucleolar targeting sequences of c-erb-A, c-myc, N-myc, p53, HSP70, and HIV tat proteins. *J Biol Chem* 264(30):18019–18023
14. Hatanaka M (1990) Discovery of the nucleolar targeting signal. *Bioessays* 12(3):143–148
15. Leonhardt H, Cardoso MC (1995) Targeting and association of proteins with functional domains in the nucleus: the insoluble solution. In *Structural and Functional Organization of the Nuclear Matrix*. *Int Rev Cytol* 162B:303–335
16. Frankel AD, Pabo CO (1988) Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* 55(6):1189–1193
17. Green M, Loewenstein PM (1988) Autonomous functional domains of chemically synthesized human immunodeficiency virus tat transactivator protein. *Cell* 55(6):1179–1188
18. Martin RM, Ter-Avetisyan G, Herce HD, Ludwig AK, Lattig-Tunnemann G, Cardoso MC (2015) Principles of protein targeting to the nucleolus. *Nucleus* 6(4):314–325. doi:10.1080/19491034.2015.1079680
19. Herce HD, Garcia AE, Cardoso MC (2014) Fundamental molecular mechanism for the cellular uptake of guanidinium-rich molecules. *J Am Chem Soc* 136(50):17459–17467. doi:10.1021/ja507790z
20. Martin RM, Tunnemann G, Leonhardt H, Cardoso MC (2007) Nucleolar marker for living cells. *Histochem Cell Biol* 127(3):243–251
21. Martin RM, Leonhardt H, Cardoso MC (2005) DNA labeling in living cells. *Cytometry A* 67(1):45–52
22. Tunnemann G, Ter-Avetisyan G, Martin RM, Stöckl M, Herrmann A, Cardoso MC (2008) Live-cell analysis of cell penetration ability and toxicity of oligo-arginines. *J Pept Sci* 14(4):469–476. doi:10.1002/psc.968

5 General Conclusions and Perspectives

In chapter 1 of this thesis, we analyzed the effect of methyl-CpG binding domain proteins on the dioxygenase activity and DNA binding ability of Ten eleven translocation proteins. Our data demonstrate an anti-correlation between Tet1 activity/DNA binding capacity and long-lasting DNA coverage by MBD proteins in a non-sequence specific, but concentration-dependent manner. Mutations in the MECP2 (Methyl-CpG binding protein 2) gene give rise to Rett syndrome, a debilitating neurological disease [1], which is characterized by global changes in neuronal chromatin structure and elevated transcriptional noise [2, 3]. Consistent with this, we found increased levels of the Tet oxidation product 5hmC in neurons of Mecp2 deficient mice, with concomitant transcriptional reactivation of epigenetically silenced major satellite repeats (Figure 1). Using triple Tet knockout mouse embryonic stem cells, we were able to attribute the observed increase in tandem repeat expression to unconfined Tet activity. Finally, our data indicate that ectopic expression of Mecp2 is sufficient to rescue the aberrant, Tet1-mediated transcriptional increase of major satellite repeats. In chapter 2,

we expanded our analyses to long interspersed nuclear elements in human cells. Similar to tandem repeats, expression and retrotransposition of interspersed elements was activated by Ten-eleven translocation protein 1 and repressed by methyl-CpG binding domain proteins (Figure 1).

Consistently, previous studies have demonstrated stabilized or reversed symptoms upon systemic administration of functional Mecp2 into female Rett mice [6]. One plausible explanation may be a direct interaction of Mecp2 with the Tet oxidation product 5hmC. However, recent studies analyzing the binding ability of Mecp2 to hmCpG are highly contentious. Mellen and colleagues identified Mecp2 as the major 5hmC-binding protein in the brain with similar binding affinities for 5mC and 5hmC containing substrates [7]. Spruijt et al. determined Mecp2 as a reader of 5hmC in mouse embryonic stem cells, albeit with lower affinity compared to 5mC [8]. Two independent *in vitro* studies demonstrated a greatly reduced binding affinity of Mecp2 to oxidative derivatives of 5mC [9, 10]. Lastly, conversion of 5mC to 5hmC was shown to completely abolish binding of Mecp2 to oligonucleotides *in vitro* [11]. Previous studies have shown that almost all mature, postmitotic neurons express abundant levels of the DNA methyltransferases

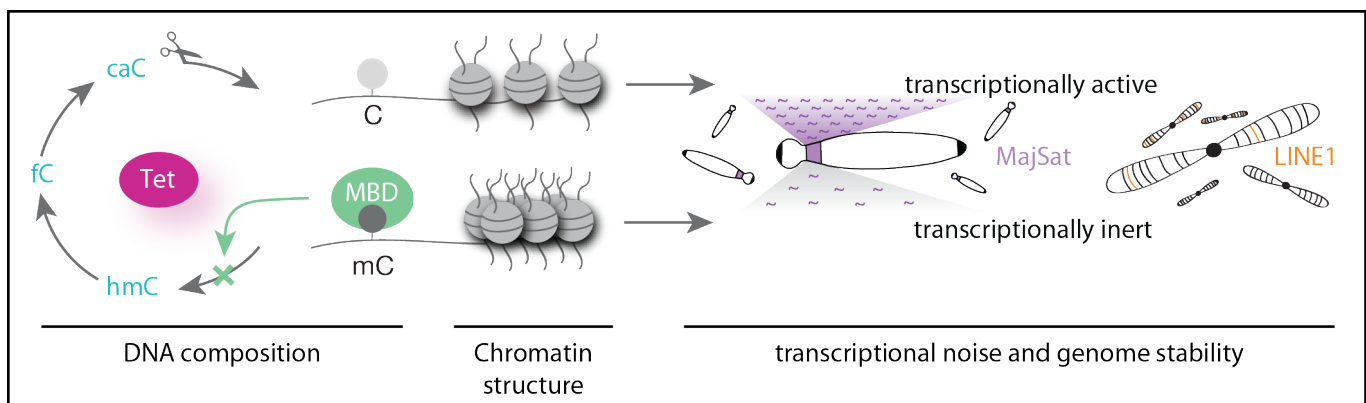


Figure 1: Scheme summarizing the main conclusions drawn from our studies in Chapter 1 and 2. Binding of Mecp2 and Mbd2 to DNA preserves chromatin composition and thus genomic integrity through regulation of Tet1 activity.

Dnmt1 and Dnmt3a [12, 13]. Based on these data, we consider following procedure as another potential Rett symptom reversing mechanism: Oxidation of 5mC by Tet proteins, removal of the oxidized derivatives by glycosylases and enzymes of the base excision repair pathway [14, 15, 16], re-methylation of cytosines by DNA methyltransferases, binding and protection of 5mC by exogenous wild type Mecp2 proteins. However, future work will be necessary to achieve certainty, whether and which of the two mechanisms occurs *in vivo*. In conclusion, our data suggest an additional role for Mecp2 and Mbd2 as guardians of the epigenome through regulation of aberrant Tet activity and argue for a potential pathophysiological function of Tet proteins in Rett syndrome (Figure 1).

Similar to epigenetic marks in DNA, some RNA modifications were recently shown to be reversible. Fu and

coworkers detected 5hmC in RNA of mammalian cells and tissue and demonstrated Tet-catalyzed oxidation of ribonucleic 5mC to 5hmC both *in vitro* and *in vivo* [17]. The identification of 5-formylcytidine (5fC) and 5-carboxylcytosine (5caC) in cellular RNA of mammals [18], as well as the formation of 5fC and 5caC by Tet1 in different RNA contexts [19] provided further evidence that the dioxygenase activity of Tet proteins may not be restricted to DNA, but might be extended to post-transcriptional regulation of RNA function.

In accordance with this, recent studies have shown that hydroxymethylation of mRNA can favor translation [20]. Hence, one could envision a similar role for 5hmC in ribosomal RNA (rRNA), which is synthesized in a relatively acidic subnuclear compartment, the nucleolus [21, 22, 23]. Many processes that take place in the nucleolus, including methylation of rRNA, require re-

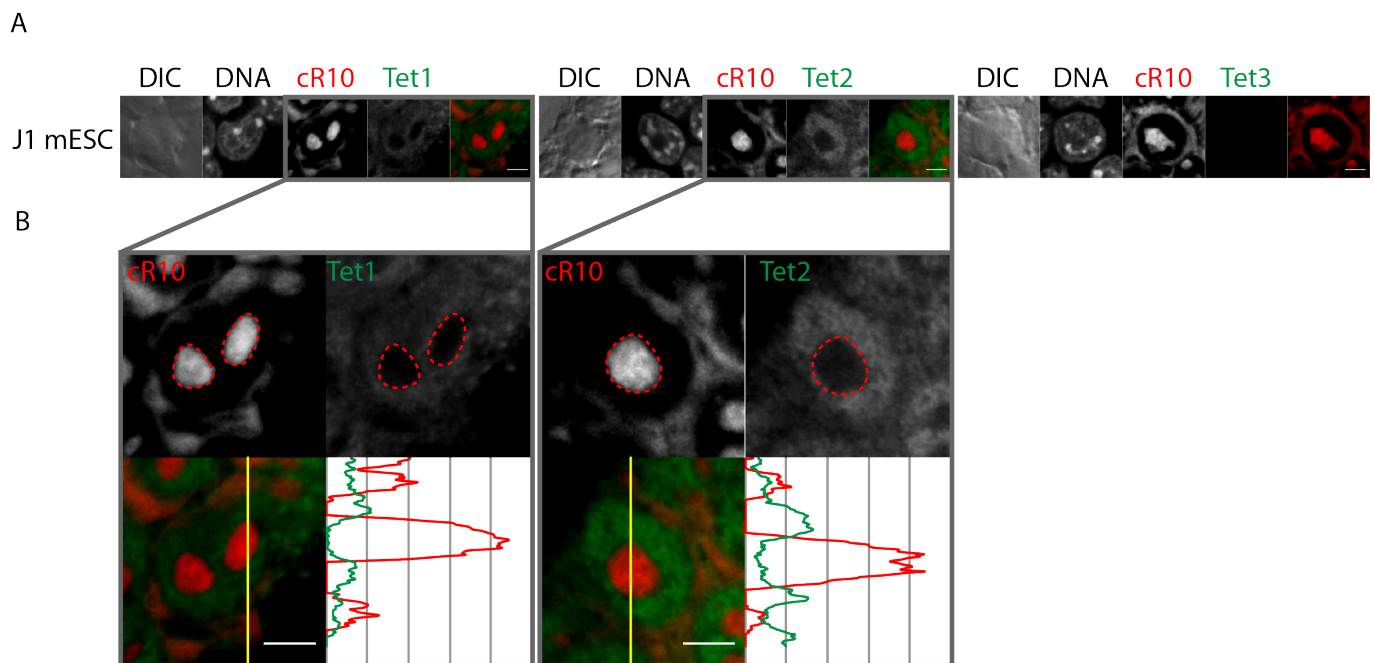


Figure 2: Immunostaining and colocalization analysis of Tet proteins with nucleoli. (A) Fluorescent images of J1 mouse embryonic stem cells (J1 mESC) immunologically stained for Tet1 (left), Tet2 (middle) and Tet3 (right) as previously described [4]. Nucleoli were labeled with the TAMRA-tagged nucleolar marker cR10 (cyclic deca-arginine [5]). DIC: Differential interference contrast. DNA was counterstained with DAPI. Scale bar: 5 μ m. Nucleoli are marked with red dashed lines. (B) Line intensity plots of cR10 (red) and Tet protein (green) distribution through a nucleus of J1 mESC.

location of various functional proteins to this nuclear subcompartment, which is often mediated by specialized amino acid motifs [24, 25].

In chapter 3 of this thesis, we determined experimentally the molecular requirements that are necessary and sufficient for the localization of proteins and peptides in the nucleolus. We demonstrate that besides amino

acid composition, isoelectric properties, as well as the charge distribution of proteins and peptides determine the extent of nucleolar accumulation. Consistent with this, peptide units composed of consecutive arginines with an isoelectric point at and above 12.6 were shown to facilitate localization to the nucleolar subcompartment. Accordingly, we demonstrate that six consecu-

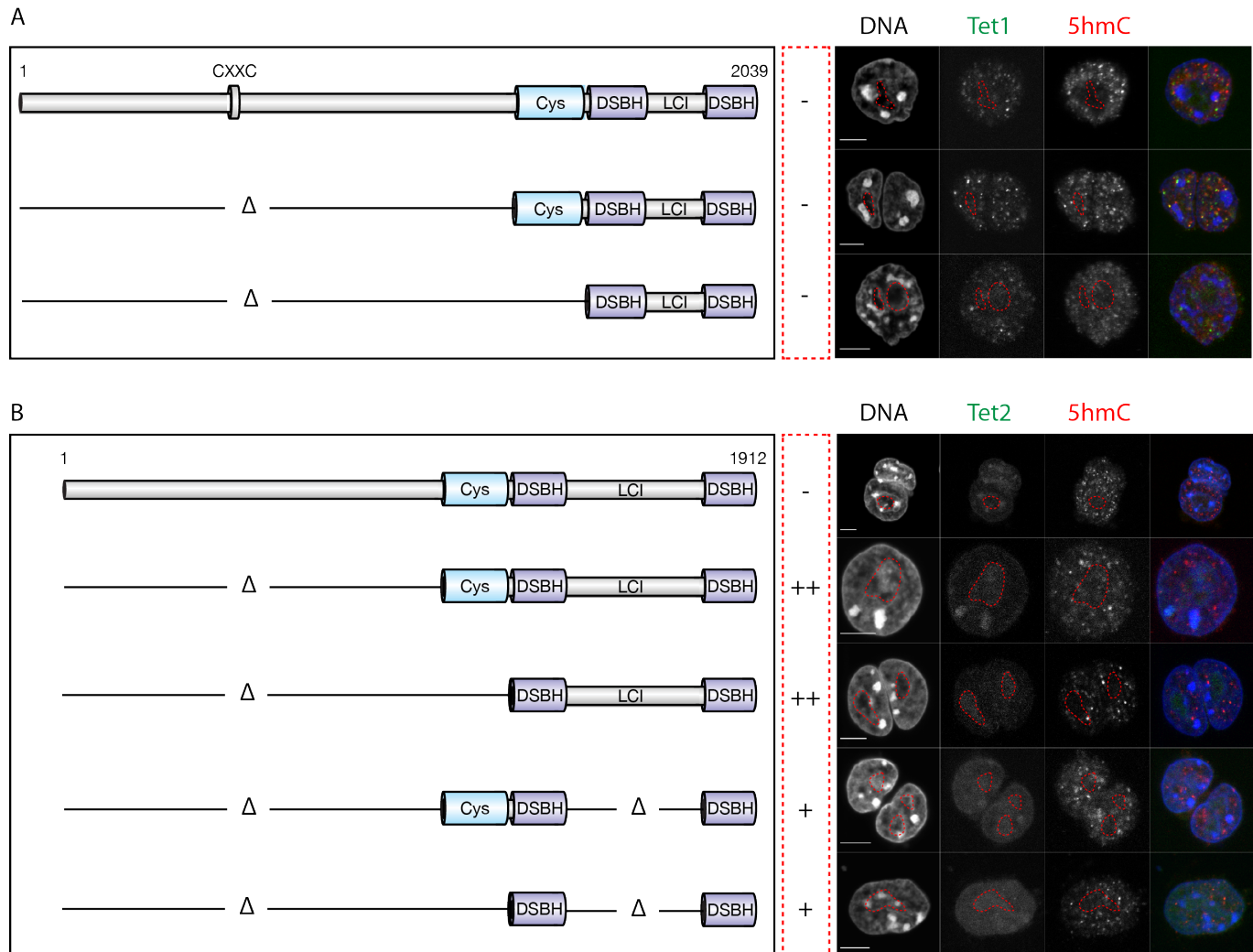


Figure 3: Subnuclear distribution of endogenous, GFP-tagged (A) Tet1 and (B) Tet2 proteins in stable J1 mESC cell lines. (left) Domain structures of wild type Tet proteins and analyzed Tet deletion mutants. LCI: low complexity insert. (right) Mid-confocal optical sections of J1 mESC immunostained for 5hmC (without RNase treatment). The signals of endogenous GFP-tagged Tet proteins were enhanced using an anti-GFP nanobody coupled to an ATTO-488 fluorophore (GFP-booster, ChromoTek GmbH, Munich, Germany). DNA was counterstained with DAPI. Scale bar: 5 μ m. Nucleoli are marked with red dashed lines. The extent of nucleolar accumulation for each deletion mutant is indicated as plus and minus. Immunofluorescent images were taken from the Master Thesis of Mario Hofweber, Leonhardt laboratory, LMU Munich.

tive arginine residues in a peptide and four successive arginines in combination with a nuclear localization signal (NLS) provide sufficient positive surface charge to mediate nucleolar accumulation. Finally, we were able to show that poly-R peptides and fusion proteins directly bind negatively charged RNA molecules *in vitro*, which explains the conservation of this nucleolar targeting principle during evolution. In conclusion, our data unveil effective amino acid motifs, capable of mediating nucleolar accumulation of short peptides and proteins and, as shown in chapter 4, provide an efficient tool for the visualization of the nucleolus in living cells. To investigate a putative role of Tet proteins as rRNA dioxygenases, we made use of this nucleolar visualization tool to study the subnuclear distribution of all three Tet protein family members in mouse embryonic stem cells (mESC). Line intensity plots of the nucleolar

marker cR10 [5] (red) and Tet protein (green) distribution through mESC nuclei revealed an anticorrelation of Tet1 and Tet2 with cR10 (Figure 2), indicating that under the given conditions, Tet proteins do not accumulate in the nucleolus. Consistent with this, several studies have shown that the nucleolar proteome is highly dynamic. Many helicase family proteins, such as WRN, BLM or RECQL4 have been shown to accumulate in the nucleolus only at specific cell cycle stages or in response to specific stimuli, including oxidative stress [26, 27, 28]. Nucleolar accumulation of Apurinic/aprimidinic endonuclease 1 (APE1), the main abasic endonuclease of the base excision repair pathway, depends on positively charged lysine residues within its N-terminal domain that undergo acetylation and thus neutralization upon genotoxic stress [29]. These data suggest that post-translational modifications (PTMs) regulate localization

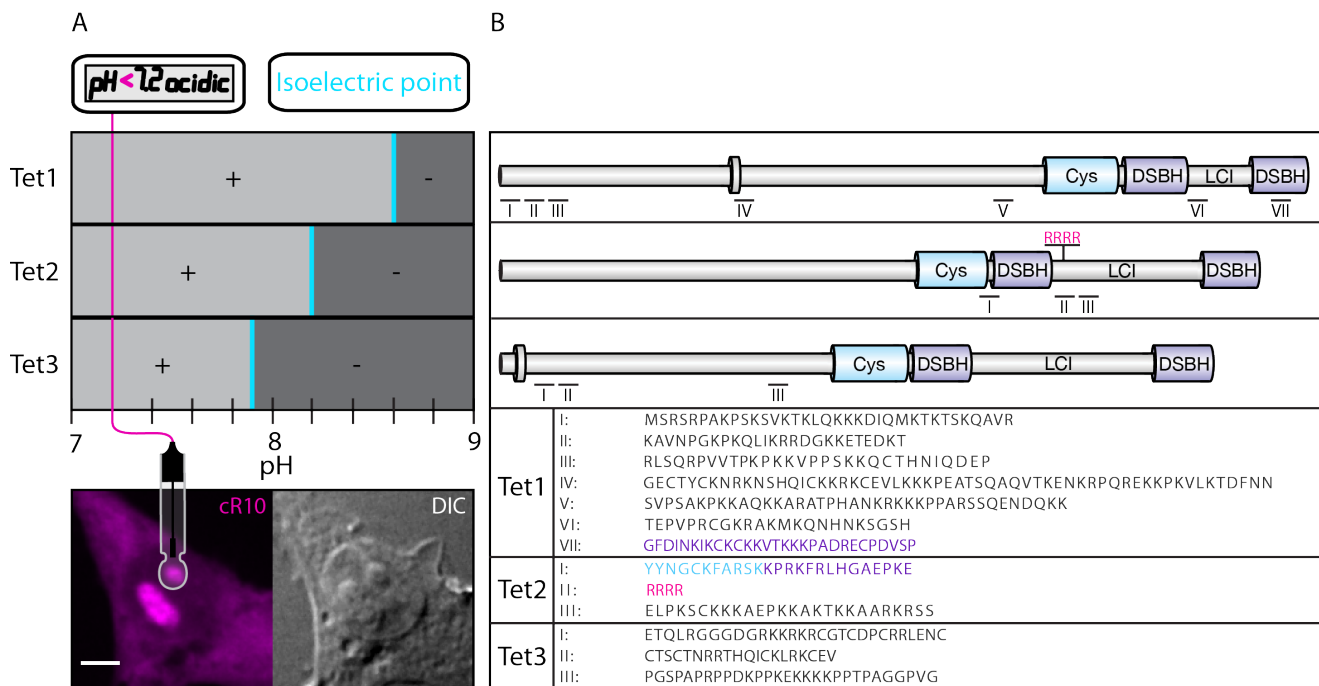


Figure 4: *In silico* analysis of Tet protein sequences (A) pH scale with indicated isoelectric points (pI) of Tet proteins (cyan) and pH value of nucleoli [23] (magenta). Net charges of Tet proteins above and below the respective pI are shown as +/- . A representative phase contrast (DIC) and fluorescent image of a mouse cell incubated with the TAMRA-tagged nucleoli marker cR10 [5] (cyclic deca-arginine peptide) is shown below. Scale bar: 5 μ m. (B) Protein schemes of Tet dioxygenases with putative nucleolar localization signals (NoLS) labeled by roman numerals.

of proteins to the nucleolus. Recent studies have shown that Tet proteins are subjected to a variety of PTMs that mostly occur at the N-terminal domain and the low-complexity insert [30]. Interestingly, N-term deletion of Tet2 (Figure 3B), but not Tet1 (Figure 3A), resulted in strong nucleolar accumulation. Additional removal of the cysteine-rich region, however, did not affect their subnuclear localization. From this we conclude that the N-terminal domain is likely to regulate the nucleolar localization of Tet2. Whether this regulation is mediated through post-translational modifications has to be validated in the future.

To achieve nucleolar accumulation, some proteins, such as the putative rRNA MTase p120, contain arginine-rich regions [25], which, as shown in chapter 3 of this thesis, act as high-affinity binding sites for (ribosomal) RNA. As demonstrated in Figure 4B, we identified four consecutive arginine residues within the low complexity insert of Tet2, which are sufficient for strong nucleolar accumulation of proteins that are fused to a nuclear localization signal (NLS) (see Chapter 3). Interestingly, additional removal of the low complexity insert diminished the previously described nucleolar accumulation of Tet2 N-term deletion mutants (Figure 3B), which led to the conclusion that the tetra-arginine motif within the LCI of Tet2 may indeed serve as nucleolar localization signal (NoLS).

In addition to arginine-rich motifs, numerous nucleolar localization sequences (NoLS) composed of short basic motifs have been identified in unrelated human proteins and were used as a training set to computationally predict NoLS for the entire human proteome [24]. With regard to a putative role of Tet proteins as rRNA dioxygenases, we made use of the resultant "Nucleolar Localization Signal Detector" tool [31] to search whether besides the arginine-rich motif in Tet2, Tet proteins are predicted to harbor additional NoLS. Indeed, several poten-

tial nucleolar localization sequences were detected in all three Tet protein sequences (Figure 4B). While the seven NoLS predicted for mouse Tet1 distributed over the entire length of the protein, the three nucleolar targeting motifs of Tet2 and Tet3 clustered in the catalytic domain and N-terminal region, respectively. To validate these data, further experiments are required to test whether each of the predicted motifs is sufficient to target reporter proteins to the nucleolus. Moreover, point mutations of these sequences should be tested for disruption of nucleolar localization of the respective Tet enzyme, once the conditions for accumulation in nucleoli are identified.

Since the subnuclear localization of proteins does not only depend on their amino acid composition, but also on their isoelectric properties and charge (see Chapter 3), we additionally assessed the nucleolar accumulation potential of Tet proteins based on these parameters. Within the acidic electrochemical environment of the nucleolus ($\text{pH} < 7.2$), all three Tet proteins carry a positive net charge. Tet1, with an isoelectric point (pI) of 8.6, has the highest theoretical net charge, shortly followed by Tet2 ($\text{pI} = 8.2$) and Tet3 ($\text{pI} = 7.9$) (Figure 4A). Hence, Tet proteins possess various features that suggest a potential role in the nucleolus. The function of rRNA hydroxymethylation, however, has to be investigated in future work. To determine, whether oxidation of 5mC in rRNA may influence the translation efficiency, the sizes and amounts of *in vitro* synthesized proteins need to be determined by the use of unmethylated, methylated and hydroxymethylated ribosomes, respectively.

In conclusion, various studies argue for a role of Tet proteins in modifying different classes of RNA. Hypomethylated RNA molecules, however, have been shown to stimulate the innate immune system through activation of Toll-like receptors [32]. Accordingly, Tet activity

has to be precisely regulated on the RNA level. In Chapter 1 and 2 of this thesis, we have demonstrated that binding of Mecp2 to DNA protects 5mC from Tet1 mediated oxidation. Interestingly, previous studies have shown that Mecp2 associates with RNA and regulates mRNA splicing [33]. Whether this interaction serves the protection of 5mC in RNA, however, still has to be determined.

In summary, the functional network of Tet proteins is rapidly growing and the identification of proteins controlling its activity and subnuclear localization will contribute to understanding the dynamic regulation and role of DNA and RNA modifications in various metabolic processes in and outside of the nucleolus.

References

- [1] R. E. Amir, I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, and H. Y. Zoghbi, "Rett syndrome is caused by mutations in x-linked mecp2, encoding methyl-cpg-binding protein 2," *Nat Genet*, vol. 23, no. 2, pp. 185–8, 1999. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/10508514>
- [2] A. P. Bird, "Gene number, noise reduction and biological complexity," *Trends Genet*, vol. 11, no. 3, pp. 94–100, 1995. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/7732579>
- [3] P. J. Skene, R. S. Illingworth, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, R. Andrews, and A. P. Bird, "Neuronal mecp2 is expressed at near histone-octamer levels and globally alters the chromatin state," *Mol Cell*, vol. 37, no. 4, pp. 457–68, 2010. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/20188665>
- [4] A. K. Ludwig, P. Zhang, F. D. Hastert, S. Meyer, C. Rausch, H. D. Herce, U. Muller, A. Lehmkuhl, I. Hellmann, C. Trummer, C. Storm, H. Leonhardt, and M. C. Cardoso, "Binding of mbd proteins to dna blocks tet1 function thereby modulating transcriptional noise," *Nucleic Acids Res*, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/27923996>
- [5] G. Lattig-Tunnemann, M. Prinz, D. Hoffmann, J. Behlke, C. Palm-Apergi, I. Morano, H. D. Herce, and M. C. Cardoso, "Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides," *Nat Commun*, vol. 2, p. 453, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21878907>
- [6] S. K. Garg, D. T. Liroy, H. Cheval, J. C. McGann, J. M. Bissonnette, M. J. Murtha, K. D. Foust, B. K. Kaspar, A. Bird, and G. Mandel, "Systemic delivery of mecp2 rescues behavioral and cellular deficits in female mouse models of rett syndrome," *J Neurosci*, vol. 33, no. 34, pp. 13612–20, 2013. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/23966684>
- [7] M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis, and N. Heintz, "Mecp2 binds to 5hmc enriched within active genes and accessible chromatin in the nervous system," *Cell*, vol. 151, no. 7, pp. 1417–30, 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/23260135>
- [8] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Muller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Muller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, and M. Vermeulen, "Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives," *Cell*, vol. 152,

- no. 5, pp. 1146–59, 2013. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/23434322>
- [9] C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes, and H. Leonhardt, "Recognition of 5-hydroxymethylcytosine by the uhrf1 sra domain," *PLoS One*, vol. 6, no. 6, p. e21306, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21731699>
- [10] S. Khrapunov, C. Warren, H. Cheng, E. R. Berko, J. M. Grealley, and M. Brenowitz, "Unusual characteristics of the dna binding domain of epigenetic regulatory protein mecp2 determine its binding specificity," *Biochemistry*, vol. 53, no. 21, pp. 3379–91, 2014. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/24828757>
- [11] V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, and L. C. Sowers, "Oxidative damage to methyl-cpg sequences inhibits the binding of the methyl-cpg binding domain (mbd) of methyl-cpg binding protein 2 (mecp2)," *Nucleic Acids Res*, vol. 32, no. 14, pp. 4100–8, 2004. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/15302911>
- [12] K. Goto, M. Numata, J. I. Komura, T. Ono, T. H. Bestor, and H. Kondo, "Expression of dna methyltransferase gene in mature and immature neurons as well as proliferating cells in mice," *Differentiation*, vol. 56, no. 1-2, pp. 39–44, 1994. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/8026645>
- [13] D. Watanabe, K. Uchiyama, and K. Hanaoka, "Transition of mouse de novo methyltransferases expression from dnmt3b to dnmt3a during neural progenitor cell development," *Neuroscience*, vol. 142, no. 3, pp. 727–37, 2006. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/16973295>
- [14] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang, "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine," *Science*, vol. 333, no. 6047, pp. 1300–3, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21778364>
- [15] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, and G. L. Xu, "Tet-mediated formation of 5-carboxylcytosine and its excision by tdg in mammalian dna," *Science*, vol. 333, no. 6047, pp. 1303–7, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21817016>
- [16] A. Maiti and A. C. Drohat, "Thymine dna glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of cpg sites," *J Biol Chem*, vol. 286, no. 41, pp. 35334–8, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21862836>
- [17] L. Fu, C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S. G. Jin, L. J. Niedernhofer, G. P. Pfeifer, G. L. Xu, and Y. Wang, "Tet-mediated formation of 5-hydroxymethylcytosine in rna," *J Am Chem Soc*, vol. 136, no. 33, pp. 11582–5, 2014. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/25073028>
- [18] H. Y. Zhang, J. Xiong, B. L. Qi, Y. Q. Feng, and B. F. Yuan, "The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both dna and rna in mammals," *Chem Commun (Camb)*, vol. 52, no. 4, pp. 737–40, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/26562407>

- [19] M. Basanta-Sanchez, R. Wang, Z. Liu, X. Ye, M. Li, X. Shi, P. F. Agris, Y. Zhou, Y. Huang, and J. Sheng, "Tet1-mediated oxidation of 5-formylcytosine (5fc) to 5-carboxycytosine (5cac) in rna," *Chembiochem*, vol. 18, no. 1, pp. 72–76, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/27805801>
- [20] B. Delatte, F. Wang, L. V. Ngoc, E. Collignon, E. Bonvin, R. Deplus, E. Calonne, B. Hassabi, P. Putmans, S. Awe, C. Wetzel, J. Kreher, R. Soin, C. Creppe, P. A. Limbach, C. Gueydan, V. Kruys, A. Brehm, S. Minakhina, M. Defrance, R. Steward, and F. Fuks, "Rna biochemistry. transcriptome-wide distribution and function of rna hydroxymethylcytosine," *Science*, vol. 351, no. 6270, pp. 282–5, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/26816380>
- [21] D. D. Brown and J. B. Gurdon, "Absence of ribosomal rna synthesis in the anucleolate mutant of *xenopus laevis*," *Proc Natl Acad Sci U S A*, vol. 51, pp. 139–46, 1964. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/14106673>
- [22] N. Granboulan and P. Granboulan, "[ultrastructure cytochemistry of the nucleolus. ii. study of the sites of rna synthesis in the nucleolus and the nucleus]," *Exp Cell Res*, vol. 38, pp. 604–19, 1965. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/14330679>
- [23] R. M. Martin, G. Ter-Avetisyan, H. D. Herce, A. K. Ludwig, G. Lattig-Tunnemann, and M. C. Cardoso, "Principles of protein targeting to the nucleolus," *Nucleus*, vol. 6, no. 4, pp. 314–25, 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/26280391>
- [24] M. S. Scott, F. M. Boisvert, M. D. McDowall, A. I. Lamond, and G. J. Barton, "Characterization and prediction of protein nucleolar localization sequences," *Nucleic Acids Res*, vol. 38, no. 21, pp. 7388–99, 2010. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/20663773>
- [25] W. C. Gustafson, C. W. Taylor, B. C. Valdez, D. Henning, A. Phippard, Y. Ren, H. Busch, and E. Durban, "Nucleolar protein p120 contains an arginine-rich domain that binds to ribosomal rna," *Biochem J*, vol. 331 (Pt 2), pp. 387–93, 1998. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/9531475>
- [26] R. A. Marciniak, D. B. Lombard, F. B. Johnson, and L. Guarente, "Nucleolar localization of the werner syndrome protein in human cells," *Proc Natl Acad Sci U S A*, vol. 95, no. 12, pp. 6887–92, 1998. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/9618508>
- [27] L. L. Woo, K. Futami, A. Shimamoto, Y. Furuichi, and K. M. Frank, "The rothmund-thomson gene product recql4 localizes to the nucleolus in response to oxidative stress," *Exp Cell Res*, vol. 312, no. 17, pp. 3443–57, 2006. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/16949575>
- [28] V. Yankiwski, R. A. Marciniak, L. Guarente, and N. F. Neff, "Nuclear structure in normal and bloom syndrome cells," *Proc Natl Acad Sci U S A*, vol. 97, no. 10, pp. 5214–9, 2000. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/10779560>
- [29] L. Lirussi, G. Antoniali, C. Vascotto, C. D'Ambrosio, M. Poletto, M. Romanello, D. Marasco, M. Leone, F. Quadrifoglio, K. K. Bhakat, A. Scaloni, and G. Tell, "Nucleolar accumulation of ape1 depends

- on charged lysine residues that undergo acetylation upon genotoxic stress and modulate its ber activity in cells," *Mol Biol Cell*, vol. 23, no. 20, pp. 4079–96, 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/22918947>
- [30] C. Bauer, K. Gobel, N. Nagaraj, C. Colantuoni, M. Wang, U. Muller, E. Kremmer, A. Rottach, and H. Leonhardt, "Phosphorylation of tet proteins is regulated via o-glcncylation by the o-linked n-acetylglucosamine transferase (ogt)," *J Biol Chem*, vol. 290, no. 8, pp. 4801–12, 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/25568311>
- [31] M. S. Scott, P. V. Troshin, and G. J. Barton, "Nod: a nucleolar localization sequence detector for eukaryotic and viral proteins," *BMC Bioinformatics*, vol. 12, p. 317, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/21812952>
- [32] K. Kariko, M. Buckstein, H. Ni, and D. Weissman, "Suppression of rna recognition by toll-like receptors: the impact of nucleoside modification and the evolutionary origin of rna," *Immunity*, vol. 23, no. 2, pp. 165–75, 2005. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/16111635>
- [33] J. I. Young, E. P. Hong, J. C. Castle, J. Crespo-Barreto, A. B. Bowman, M. F. Rose, D. Kang, R. Richman, J. M. Johnson, S. Berget, and H. Y. Zoghbi, "Regulation of rna splicing by the methylation-dependent transcriptional repressor methyl-cpg binding protein 2," *Proc Natl Acad Sci U S A*, vol. 102, no. 49, pp. 17551–8, 2005. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/16251272>

6 Annex

6.1 Abbreviations

A	Adenine
aa	Amino acid
Abobec	Apolipoprotein B mRNA Editing Enzyme Catalytic Polypeptide-Like
ADP	Adenosine diphosphate
Aid	Activation-induced cytidine
AlkB	Alpha-ketoglutarate-dependent dioxygenase
ALKBH5	AlkB Homolog 5
Alu	Arthrobacter luteus
AML	Acute myeloid leukemia
APCMin	Adenomatous polyposis coli gene
APE1	Apurinic/aprimidinic endonuclease 1
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphate synthetase
BAZ2A/B	Bromodomain Adjacent To Zinc Finger Domain 2A/B
BDEPE	2-bromo-1-(4-diethylaminophenyl)-ethanone
BER	Base excision repair
BGT	β -glucosyltransferase
BLM	Bloom Syndrome RecQ Like Helicase
BTB/POZ	Bric-a-brac, tramtrack, broad complex/poxvirus and zinc finger
c-Myc	cellular Myelocytomatosis
C	Cytosine
CAF-1	Chromatin assembly factor 1
CC	Chromocenter
CDK	cyclin-dependent kinase
CGI	CpG island
CoREST	Co-repressor of RE1 silencing transcription factor
CPP	Cell penetrating peptide
cR10	cyclic deca-arginine
CTD	C-terminal domain
DAPI	4',6-diamidino-2-phenylindole
DFC	Dense fibrillar components
DIC	Differential interference contrast
DMAP	Dnmt1-associated protein

DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
Dnmt	DNA methyltransferase
dNTP	deoxyribonucleoside 5'-triphosphate
Dpp3	Developmental pluripotency associated 3
DSBH	Double-stranded beta helix
DTT	Dithiothreitol
dUTP	2'-deoxyuridine 5'-triphosphate
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
ESC	embryonic stem cell
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FC	Fibrillar centers
Fe	Iron
FISH	Fluorescence in situ hybridization
FRAP	Fluorescence recovery after photobleaching
G-phase	Gap-phase
G	Guanine
Gadd45	Growth arrest and DNA damage inducible protein 45
GAL4	Galactose-responsive transcription factor GAL4
GBP	GFP binding protein
GC	Granular component
gDNA	genomic DNA
GFP	Green fluorescent protein
H3K9me3	Tri-methylation of histone 3 at lysine 9
HBSS	Hank's Balanced Salt Solution
HDAC	Histone deacetylase
HEK	Human embryonic kidney
HPG	Hippocampus/parahippocampal gyrus
HRP	Horseradish peroxidase
IAP	Intracisternal particle A
ICBP90	Inverted CCAAT box binding protein of 90 kDa
ICF	Immunodeficiency, chromosomal instability, and facial anomalies
ID	Intervening domain
Idax	Inhibition of the Dvl and Axin complex
Idh1/2	Isocitrate dehydrogenase 1/2

JHDM1A/B	Jumonji C domain-containing histone demethylases 1A/1B
kDA	Kilo Dalton
Klf4	Kruppel-like factor 4
KO	Knock out
LC-MS	Liquid chromatography mass spectrometry
LIF	Leukemia inhibitory factor
LINE1	Long interspersed nuclear element 1
MaSat	Major Satellite
MBD	Methyl-CpG binding domain
Mbd1-6	Methyl-CpG binding domain proteins 1-6
MBP	Methyl-CpG binding protein
mCpG	methylated CpG dinucleotide
Mecp2	Methyl-CpG binding protein 2
mESC	mouse embryonic stem cells
mNSC	mouse neuronal stem cells
MLL	Mixed lineage leukemia
MTase	Methyltransferase
MTF	Mouse tail fibroblast
NCoR	Nuclear receptor corepressor
Neil	Nei-like
NeuN	Neuronal nuclear antigen
NIRF	Np95/ICBP90-like RING finger protein
NLS	Nuclear localization signal
NoDS	nucleolar detention sequence
NoLS	Nucleolar localization signal
Np95	Nuclear protein (with molecular weight of 95 kilo Daltons)
Np97	Nuclear protein (with molecular weight of 97 kilo Daltons)
NPC	Neural progenitor cell
nt	nucleotides
NTD	N-terminal domain
NuRD	Nucleosome Remodeling Deacetylase
Oct4	Octamer-binding transcription factor 4
OGT	O-linked β -N-acetylglucosamine transferase
p21 ^{CIP1}	Cdk-interacting protein 1 of 21 kDa
p53	Tumor suppressor protein of 53 kDa
PBD	PCNA binding domain
PBHD	Polybromo-1 protein homologous domain

PBS	Phosphate buffered saline
PBST	PBS containing Tween
PC	Phase contrast
PCNA	Proliferating cell nuclear antigen
PCNP	PEST containing nuclear protein
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PGC	Primordial germ cell
PHD	Plant homeodomain
pI	Isoelectric point
Poly-K	Poly-lysine
PolyR	Poly-arginine
pRB	retinoblastoma protein
PTM	Post-translational modification
PWWP	proline-tryptophan-tryptophan-proline motif
R10	Deca-arginine
rDNA	ribosomal DNA
RECQL4	RecQ like helicase 4
rRNA	ribosomal RNA
RING	Really interesting new gene
RNA-seq	RNA sequencing
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT-qPCR	reverse transcription quantitative polymerase chain reaction
RTT	Rett syndrome
S-phase	Synthesis-phase
SETDB1	SET domain, bifurcated 1
Sin3a	Sin3 transcription regulator family member A
SINE	Short interspersed nuclear element
SMRT	Silencing mediator of retinoic acid and thyroid receptor
Sox2	SRY (sex determining region Y)-box 2
SRA	SET and RING finger associated
SSC	Saline-sodium citrate
T	Thymine
TALE	Transcription activator-like effector
TAT	Trans-Activator of Transcription
Tdg	Thymine DNA glycosylase

TET	Ten eleven translocation
Tet1CD	Catalytically active domain of Tet1
Tet1CDmut	Catalytically inactive domain of Tet1
TRD	Transcriptional repression domain
TS	Targeting sequence
TSA	Trichostatin A
TSS	Transcription start site
TTD	Tandem tudor domain
Ubl	Ubiquitin-like domain
Uhrf1	Ubiquitin-like with plant homeodomain and ring finger domains 1
Wnt	fusion of the terms wingless and proto-oncogene integration-1
WRN	Werner syndrome ATP-dependent helicase
wt	Wild type
YFP	Yellow fluorescent protein
Zbtb38	Zinc finger and BTB domain containing 38
Zbtb4	Zinc finger and BTB domain containing 4
2-KG	2-Ketoglutarate
2i	2 inhibitors (Gsk3 β and MEK inhibitors)
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymethyluracil
5mC	5-methylcytosine

6.2 List of contributions

6.2.1 General introduction

A.K. Ludwig wrote the text, prepared figures (4, 5, 6, 7, 8) and tables (3, 4, 6) for sections 2 (DNA modification readers), 2.1 (MBD protein family), 2.1.1 (Mecp2), 2.1.2 (Mbd1), 2.1.3 (Mbd2), 2.1.4 (Mbd3), 2.1.5 (Mbd4), 2.2 (Kaiso), 2.3 (SRA domain protein family), 3.3 (MBP proteins in disease), 4 (Concluding remarks).

P. Zhang wrote the text, prepared figures (1, 2, 3) and tables (1, 2, 4, 5) for sections 1 (DNA modifications and modifiers), 1.1 (Cytosine modifiers: Dnmts), 1.1.1 De novo DNA methylation, 1.1.2 (Maintenance DNA methylation), 1.2 (DNA base modifications), 1.3 (Methylcytosine modifiers), 1.3.1 (Tissue and genome-wide distribution of Tet), 1.3.2 (Regulation of Tet activity), 1.3.3 (Hydroxymethylcytosine maintenance), 3.1 (DNMT proteins in disease), 3.2 (TET proteins in disease). M.C. Cardoso conceived and supervised the project, wrote the abstract and revised the manuscript.

6.2.2 Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise

A.K. Ludwig wrote the manuscript and prepared all of the figures and figure legends (except for figures S4, S6, S10b, S10c and S12a). **A.K. Ludwig** performed and analyzed the experiments of figures 1a, 3a, S10b, S10c and S11 together with P. Zhang. **A.K. Ludwig** performed and analyzed the experiments of figures S7 and S10a (left) together with F.D. Hastert. **A.K. Ludwig** performed and analyzed the experiments of figures 1b, 3b, 6e, S2, S16c. **A.K. Ludwig** imaged and analyzed the data of figures 5, 6b, S14 and S16d. **A.K. Ludwig** analyzed the data of figures 6c, 6d, S16a and S16b.

P. Zhang, F.D. Hastert and M.C. Cardoso revised the manuscript. P. Zhang and F.D. Hastert performed and analyzed experiments for figures 2, 3c, S12b. P. Zhang prepared figures S4, S6, S10a (right) S10b, S10c and S12a, wrote R scripts for analysis and presentation of figures 1b, 3b and S4, performed and analyzed experiments of figures 4, S1, S5, S6, S8, S10a (right), S12a and S13. S.M. prepared and stained wildtype and Mecp2 knockout mouse brains for figures 5, 6b, 6c, 6d, S14, S16a, S16b, S16d. C.R. performed and analyzed the experiments of figure S4. H.D. Herce performed the H-coefficient analysis of figure 6a. U. Müller performed and analyzed the experiment of figure S9. A. Lehmkuhl cultured cells. I. Hellmann and C. Trummer performed and analyzed the experiments of figure 6f. C.S. performed statistical analysis for figures 1 and 3 H. Leonhardt conceived and supervised the project. M.C. Cardoso conceived and supervised the project.

6.2.3 L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins

P. Zhang wrote the manuscript and prepared all of the figures and figure legends (except for figures S1, S2, S7 and S10). P. Zhang performed and analyzed the experiments of figures 1c, 2, 3a, S4, S6, S8 and S9, as well as of videos V1-3.

A.K. Ludwig, F.D. Hastert and M.C. Cardoso revised the manuscript. **A.K. Ludwig** prepared figures S1, S2, S7 and S10 and wrote the respective figure legends. **A.K. Ludwig** performed and analyzed the experiments of figures 1d, 3b, S1, S2, S7 and S10. **A.K. Ludwig** and P. Zhang performed and analyzed the experiments of

figures 1e and 3c.

F.D. Hastert performed and analyzed the experiments of figure S3. F.D. Hastert and C. Rausch performed and analyzed experiments of figures 1i and S5. A. Lehmkuhl cultured cells. I. Hellmann and M. Smets performed and analyzed the experiments of figure 1h. H. Leonhardt conceived and supervised the project. M.C. Cardoso conceived and supervised the project.

6.2.4 Principles of protein targeting to the nucleolus

R.M. Martin prepared and wrote the manuscript, performed and analyzed experiments of figures 1 and S1, analyzed data for figure 2 and supplementary tables S2 and S3, prepared figures 1, 2, 3, 5 and S1, as well as supplementary tables S2 and S3 and the respective figure legends. G. Ter-Avetisyan and G. Lättig-Tünnemann designed and performed experiments for figure 1 and revised the manuscript. H.D. Herce revised the manuscript, performed and analyzed experiments of figure 3, prepared the respective figure and figure legend and wrote the respective Materials and Methods and Results part.

A.K. Ludwig revised the manuscript, wrote the sections "Protein preparation", "RNA preparation" and "*In vitro* RNA binding assay" of the Materials and Methods part and performed protein and RNA preparation. **A.K.**

Ludwig and H.D. Herce performed and analyzed the experiments of figures 4 and S2, wrote the respective Results part and prepared the respective figures and figure legends.

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

6.2.5 Visualization of the Nucleolus in Living Cells with Cell-Penetrating Fluorescent Peptides

R. M. Martin wrote sections 2, 3.1, 3.2 and 3.3, prepared figure 1 and wrote the respective figure legend. H.D. Herce wrote section 3.4, prepared figure 3, wrote the respective figure legend and revised the manuscript.

A.K. Ludwig revised the manuscript, prepared figure 2 and wrote the respective figure legend.

M.C. Cardoso conceived and supervised the project, wrote section 1 and revised the manuscript.

6.2.6 General Conclusions and Perspectives

A.K. Ludwig wrote the text, prepared all figures (except for Figure 3A and B, right) and performed and analyzed all experiments (except for Figure 3A and B, right).

M. Hofweber performed all experiments of Figure 3A and B (right) and prepared the respective figure.

6.3 Acknowledgments

First, I would like to thank Prof. Dr. M. Cristina Cardoso, my doctoral supervisor. I am grateful for your constant support, advice and contagious enthusiasm. Thanks for always having an open ear and for taking care of me at any time.

For agreeing to be the second referee of this thesis, I would like to thank Prof. Dr. B. Süß.

I would like to express my special thanks to all present and past members of the Cardoso lab, particularly...

...Peng and Flori. From the very beginning we built a great team, pulled together in one direction and supported each other whenever we could. I want to say thank you for your support and advice over all the years.

I had a really good time with both of you and I will miss the funny, exciting and thrilling moments in- and outside of the lab.

...Kathrin, Flori and Paddy (affectionately called "Lieblingsbiologen"). Thanks for being more than just colleagues. For listening to all kinds of stories and for all the wonderful moments in the botanical garden where we ascertained, once and for all, that squirrels are certainly no devils.

...Henry. Thanks for your creativity, enthusiasm and interest in my work at any given time. It was a pleasure to work with and learn from you. I will miss our discussions and your honest advice.

...Anne. My warmest thanks for your constant support and for facilitating our daily lab routines.

...Alex for his support, advice and simple, but universal golden rule: "Try to turn it off and on again".

...Cathia who enriched our Tet-Team. Thanks for your support, initiative and reliability, especially during our

revision work.

...Stephanie. Thanks for sharing your knowledge on the preparation and staining of mouse brain, as well as for your support and enthusiasm in our work over all the years.

...Corella und Annette. Although you left Cristinas lab some years ago, you have always supported me in word and deed. Thanks for keeping my spirit up.

I want to thank all of the people who contributed to the success of this work, especially...

...H. Leonhardt (LMU, München), G.G. Schumann (PEI, Langen), M. Hofweber (LMU, München) and A.M. Bischoff (Universität Göttingen).

Vielen Dank auch meinen Hühnern und "Stamm-tischschwwestern" für die Freundschaft, Ablenkung und den Zusammenhalt über all die Jahre.

Herzlichen Dank an Katrin Lindner von "forevermo-ments" für ihre Kreativität und die Gestaltung des Titel-bilds.

Zuletzt möchte ich meiner Familie danken. Meinem Fundament.

Patrik. Danke, dass Du mir immer den Rücken stärkst und ich immer auf Dich zählen kann.

Eva, Matthias und Ida. Danke, dass ihr immer für mich da seid und mein Leben bunter macht.

Mama und Papa. Danke, dass ihr mich immer in meinen Entscheidungen unterstützt und mir diese Zeit ermöglicht habt. Danke, dass ihr mich auf all meinen Wegen begleitet, mit mir Täler durchschreitet und Berge versetzt.

6.4 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 16. März 2017

Anne Kathrin Ludwig

6.5 Curriculum vitæ

Personal Information

Name: Anne Kathrin Ludwig
Adress: Schnittpahnstraße 10, 64287 Darmstadt, Germany
Date of birth: 21.7.1986
Place of birth: Bensheim

Doctoral studies

04/2012 - 04/2017 **PhD thesis**
Technische Universität Darmstadt
Department of Biology
Cell Biology and Epigenetics
Prof. Dr. M. Cristina Cardoso
"Guardians of the Epigenome - Regulation and Role of Nucleotide Modifications"

Academic education

10/2006 - 04/2012 **Diploma studies in biology**
Technische Universität Darmstadt
Department of Biology
Majors: molecular cell- and developmental biology, microbiology, biochemistry
grade: 1.0 (very good) with honors

07/2011 - 04/2012 **Diploma thesis**
Technische Universität Darmstadt
Department of Biology
Cell Biology and Epigenetics
Prof. Dr. M. Cristina Cardoso
"Effect of Ten eleven translocation proteins on chromatin composition, structure and dynamics"

08/2010 - 12/2010 **Research Internship**
Lowy Cancer Research Center, University of New South Wales, Sydney, Australia
Dr. Luke Hesson
"Epigenetic alterations in colorectal cancer - detecting and analyzing LST specific DNA hypermethylation"

Awards	
2016	<p>Poster Award</p> <p>International Meeting of the German Society for Cell Biology</p> <p>March 14-16, Munich</p>
Publications	
2016	<p>A.K. Ludwig*, P. Zhang*, S. Meyer, H.D. Herce, U. Muller, F.D. Hastert, A. Lehmkuhl, C. Storm, H. Leonhardt, M.C. Cardoso</p> <p>*equally contributing authors</p> <p>Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise</p> <p>NAR, doi: 10.1093/nar/gkw1197</p> <p>R.M. Martin, H.D. Herce, A.K. Ludwig, M.C. Cardoso</p> <p>Visualization of the nucleolus in living cells with cell-penetrating fluorescent peptides</p> <p>Methods in Molecular Biology, doi: 10.1007/978-1-4939-3792-9 6</p> <p>S. Srivastava, A.K. Ludwig, J.W.H. Wong, L.B. Hesson</p> <p>An investigation of the potential for epigenetic inactivation by transcription read-through in a sporadic colorectal cancer</p> <p>Gene, doi: 10.1016/j.gene.2016.03.031</p> <p>A.K. Ludwig*, P. Zhang*, M.C. Cardoso</p> <p>Modifiers and Readers of DNA modifications and their impact on genome structure, expression and stability in disease</p> <p>*equally contributing authors</p> <p>Frontiers in Genetics, doi: 10.3389/fgene.2016.00115</p>
2015	<p>R.M. Martin, G. Ter-Avetisyan, H.D. Herce, A.K. Ludwig, M.C. Cardoso</p> <p>Principles of protein targeting to the nucleolus</p> <p>Nucleus, doi: 10.1080/19491034.2015.1079680</p>

