
**Mechanistic and functional insights
into the recognition and regulation of DNA modifications
by UHRF1, DNMT1 and TET proteins**

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

Carina Trummer
München, 2020



Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. Heinrich Leonhardt
im Bereich Humanbiologie und BioImaging
an der Ludwig-Maximilians-Universität München

Erstgutachter: Prof. Dr. Heinrich Leonhardt
Zweitgutachter: Prof. Dr. Peter Geigenberger

Tag der Abgabe: 16.03.2020

Tag der mündlichen Prüfung: 17.07.2020

Erklärung

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist. Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt. Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 16.03.2020

Carina Trummer

TABLE OF CONTENTS

SUMMARY	1
ZUSAMMENFASSUNG	2
1 INTRODUCTION	4
1.1 Introducing Epigenetics	4
1.2 The Epigenetic Tool Box	5
1.2.1 DNA modifications.....	5
1.2.2 Histone modifications.....	6
1.2.3 Histone variants.....	8
1.2.4 Nucleosome remodeling.....	9
1.2.5 Non-coding RNA.....	10
1.3 Molecular Basis of DNA methylation	11
1.3.1 De novo methylation	11
1.3.2 Maintenance methylation	12
1.3.2.1 DNMT1	12
1.3.2.2 UHRF1.....	13
1.4 Molecular Basis of DNA demethylation	16
1.4.1 Active demethylation	16
1.4.2 Passive demethylation.....	18
1.5 Cellular Functions of DNA methylation & demethylation	18
1.5.1 Pre-implantation development.....	18
1.5.2 Primordial germ cells	19
1.5.3 Transcriptional control & transposon silencing	20
1.5.4 Formation of cancer and other human diseases.....	21
1.5.4.1 Dysfunction of DNMTs.....	22
1.5.4.2 Dysfunction of UHRF1.....	22
1.5.4.3 Dysfunction of TET and IDH enzymes.....	23
2 RESULTS.....	27
2.1 Systematic analysis of the binding behavior of UHRF1 towards different methyl- and carboxylecytosine modification patterns at CpG dyads.....	27
2.2 Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation	73
2.3 Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive demethylation in mammals	107
2.4 Metabolic regulation of TET enzymes in mESCs	177
2.5 Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise	199
3 DISCUSSION.....	241
3.1 Binding of UHRF1 towards different mC- and caC- modification patterns.....	241
3.2 Two distinct modes of DNMT1 recruitment via UHRF1.....	244
3.3 TET-controlled and DPPA3/STELLA-driven passive demethylation	249
3.4 Metabolic regulation of TET enzymes	251
3.5 Binding of MBD proteins to DNA blocks TET1 function	255

TABLE OF CONTENTS

4	APPENDIX	257
4.1	References.....	257
4.2	Abbreviations	281
4.3	List of publications	284
4.4	Statutory declaration and statement.....	285
4.5	Declaration of contribution	286
4.6	Acknowledgement	288
4.7	Curriculum vitae	289
4.8	List of generated plasmids.....	290

SUMMARY

The regulatory epigenome is essential in the development of organisms as it greatly contributes to the establishment and maintenance of cellular identity. Different layers of epigenetic control, for instance the chemical modification of histones and DNA, are closely interconnected and determine the accessibility of chromatin and how genetic information is utilized in different cell types. These layers stably protect genome integrity on the one hand and enable a certain degree of phenotypic plasticity on the other as they dynamically respond to external stimuli and environmental changes. This thesis aimed to further examine how DNA methylation patterns are regulated within the epigenetic landscape and to dissect the precise function of proteins directly involved in controlling DNA methylation levels, especially UHRF1, DNMT1 and TET proteins. In contrast to other epigenetic marks, the inheritance of DNA methylation patterns is well-studied and relies mainly on the activity of the maintenance methyltransferase DNMT1 and its co-factor UHRF1.

Within this thesis, a systematic *in vitro* analysis of the binding properties of UHRF1 towards different DNA modifications is described, revealing that UHRF1 exhibits a preference for carboxylated cytosine (caC) besides hemi-mC. This is based on specific binding modes and the highly flexible NKR finger region of UHRF1 as investigated in complementary MD simulations. Furthermore, UHRF1 is shown to generate a second recruitment signal for DNMT1, namely ubiquitylated PAF15 (PAF15ub2), which is similarly bound by DNMT1 as H3K9Ub2. Whereas maintenance methylation through DNMT1 in early S-phase is demonstrated to mainly depend on PAF15Ub2, H3Ub2 is important for the methylation of late-replicating chromatin. Additionally, the investigation of naïve pluripotent mESCs uncovered that the hypomethylated genome, characteristic for these cells, is largely promoted by the inhibition of the maintenance methylation machinery through DPPA3-mediated abrogation of UHRF1 binding to chromatin. It is further described that the expression of DPPA3 is directly regulated by TET1 and TET2, two α -ketoglutarate-dependent dioxygenases, which actively remove methylation marks, and that this DPPA3-mediated passive demethylation represents an evolutionary new concept of boreoeutherian mammals. Another section of this thesis addresses the metabolic regulation of TET proteins in mESCs and demonstrates that α -ketoglutarate constitutes a rate-limiting factor for the activity of these enzymes with consequences on pluripotency. Moreover, the inhibitory effect of 2-HG, an oncometabolite produced by mutant IDH enzymes, is also examined in mESCs, offering the possibility to precisely study the basis of epigenetic alterations observed in tumors harboring IDH mutations. Lastly, this thesis includes the examination of cross-regulating functions of TET1 and mC-binders, in particular MeCP2 and MBD2. As evident *in vitro* and *in vivo*, mC-binding proteins restrict the access of TET1 to DNA thereby protecting methylated cytosines from TET1-mediated oxidation. This in turn is discussed to be a critical mechanism lacking in patients with Rett syndrome, a neurological disorder caused by MeCP2 mutations.

In conclusion, this work provides mechanistic and functional insights into the role of UHRF1, DNMT1 and TET enzymes in recognizing and regulating DNA modifications and highlights new aspects of these factors during mammalian development and disease.

ZUSAMMENFASSUNG

Epigenetische Regulationsmechanismen sind essentiell für die Entwicklung von Lebewesen, da sie die Identität von Zellen etablieren und diese aufrechterhalten. Verschiedene Ebenen der epigenetischen Kontrolle, wie beispielsweise chemische Modifikationen an Histonproteinen und der DNA, sind dabei eng miteinander verknüpft und entscheiden über die Zugänglichkeit bestimmter Chromatinbereiche und wie die genetische Information von unterschiedlichen Zelltypen genutzt werden kann. Diese epigenetischen Kontrollebenen schützen zum einen die Integrität des Genoms und tragen andererseits zur phänotypischen Plastizität bei, da sie das Epigenom anhand externer Reize und veränderten Umweltfaktoren entsprechend anpassen.

Ziel dieser Arbeit war es genauer zu untersuchen, wie epigenetische DNA Methylierungsmuster reguliert werden und die exakte Funktionsweise der Faktoren zu entschlüsseln, die hauptsächlich an der Steuerung der DNA Methylierung beteiligt sind; allen voran UHRF1, DNMT1 und TET Proteine. Im Gegensatz zu anderen epigenetischen Parametern, ist die Weitergabe von DNA Methylierungsmustern an Tochterzellen weitestgehend bekannt und beruht vornehmlich auf der Aktivität von DNMT1, einer DNA Methyltransferase, in Zusammenarbeit mit ihrem Kofaktor UHRF1.

Im ersten Teil dieser Arbeit wird zunächst eine systematische *in vitro* Analyse beschrieben, in der die Bindungseigenschaften von UHRF1 gegenüber unterschiedlichen DNA Modifikationen untersucht wurden. Diese konnte zeigen, dass UHRF1 neben hemi-methyliertem Cytosin (hemi-mC) eine Bindungspräferenz für carboxyliertes Cytosin (caC) aufweist. Durch ergänzende MD Simulationen konnte ferner gezeigt werden, dass die Präferenz für caC durch spezifische Bindungskonformationen und die sehr anpassungsfähige NKR-Finger-Region von UHRF1 zustande kommt. Es wird außerdem dargelegt, dass UHRF1 ein weiteres Rekrutierungssignal für DNMT1 generiert, ubiquityliertes PAF15 (PAF15Ub₂), das von DNMT1 sehr ähnlich gebunden wird wie H3Ub₂. Während PAF15Ub₂ innerhalb der frühen S-Phase eine entscheidende Rolle für die Methylierungsaktivität von DNMT1 spielt, ist H3Ub₂ primär für Chromatin wichtig, das in der späten S-Phase repliziert wird. Darüber hinaus wird in einer Analyse von naiven embryonalen Stammzellen untersucht, was ihr charakteristisches, hypomethyliertes Genom bedingt. In diesem Zusammenhang kann gezeigt werden, dass die beiden α -Ketoglutarat-abhängigen Dioxygenasen, TET1 und TET2, die Expression von DPPA3 regulieren und DPPA3 die Bindung von UHRF1 an Chromatin verhindert. Somit wird die Aufrechterhaltung der DNA Methylierung inhibiert, was zur Hypomethylierung des Genoms führt. Es wird weiterhin dargelegt, dass diese DPPA3-vermittelte passive Demethylierung einen evolutionär neuen Mechanismus höherer Säugetiere darstellt.

Ein weiterer Abschnitt dieser Arbeit beschäftigt sich mit der Regulation von TET Proteinen durch den Stoffwechsel und beweist, dass α -Ketoglutarat ein limitierender Faktor für die Aktivität dieser Enzyme ist und den Pluripotenz-Zustand von Stammzellen beeinflusst. Des Weiteren wird der inhibierende Effekt des Onko-Metaboliten 2-HG, der durch mutierte IDH Enzyme produziert wird, in murinen Stammzellen untersucht, um grundlegende epigenetische Veränderungen, die in Tumoren mit IDH Mutationen beobachtet wurden, zielgerichtet untersuchen zu können. Zuletzt wird in dieser Arbeit die gegenseitige Regulation von TET1 und mC-bindenden Proteinen, vor allem MeCP2 und MBD2, betrachtet. Hierbei kann sowohl *in vitro* als auch *in vivo* gezeigt werden, dass mC-bindende Proteine den Zugang zu methylierter DNA beschränken und somit methylierte Cytosine vor der Oxidation durch TET1 schützen. Darüber hinaus wird dieser entscheidende

Mechanismus als potentieller Auslöser des Rett-Syndroms diskutiert, eine neurologische Erkrankung, die durch Mutationen im *Mecp2* Gen hervorgerufen wird. Zusammenfassend beschreibt diese Arbeit neue mechanistische und funktionelle Erkenntnisse über die Rolle von UHRF1, DNMT1 und TET Proteinen in der Erkennung und Regulation von DNA Modifikationen.

1 INTRODUCTION

1.1 Introducing Epigenetics

Genetics, as one of the largest branches of biology, was for a long time the only discipline studying molecular mechanisms of inheritance, i.e. how traits and genetic information is handed down from parents to offspring (Pearson, 2006). Already in the 19th century, Mendel established new laws of inheritance and paved the way for the modern science of genetics (William Bateson Mendel and Bateson, 2009). Shortly after, the macromolecule deoxyribonucleic acid (DNA) was discovered and a landmark study of Watson and Crick in 1953 revealed that the genetic information of every organism is stored as a four-letter code in a three-dimensional double helix (Dahm, 2005; Watson and Crick, 1953). Since the inherited information, i.e. DNA, is identical within every cell and the DNA sequence serves as a general template for producing cellular components, it raises the question how unique cell types emerge (Alberts et al., 2002). In other words, how do different phenotypes occur without any changes in the underlying genotypes?

A question that can be answered based on epigenetics; a term coined by Conrad H. Waddington that describes the inheritance of defined gene expression patterns without any alterations of the DNA sequence (Holliday, 1990; Waddington, 1956). Epigenetic mechanisms rather add an additional layer of information to the actual DNA template by adapting chromatin, the “packaging unit” of our genetic information (Li and Reinberg, 2011). Chromatin consists of the DNA itself and nuclear proteins and is maintained in different compaction states. Identical genetic material can therefore be differentially packaged, which eventually allows for different cell-type identities (Li and Reinberg, 2011).

Epigenetic regulation is not only crucial for the terminal differentiation of cells into kidney, brain or liver cells etc., but is equally important for developmental phenomena like X-inactivation and imprinting (Allis and Jenuwein, 2016). Here, one entire X-chromosome in female cells or specific genes in the germline are silenced by epigenetic means to compensate for dosage differences between sexes or to express genes in a parent-of-origin-specific manner, respectively (Augui et al., 2011; Ferguson-Smith, 2011). Epigenetic mechanisms are evolutionary conserved and do not only exist in mammals but also in plants or single-celled microorganisms like yeast (Skvortsova et al., 2018). Especially plants are often exposed to unpredictable environments or unfavorable growth conditions and therefore require a high degree of phenotypic plasticity to rapidly respond to environmental changes. A process enabled by over 130 plant-based epigenetic modulators (Pikaard and Mittelsten Scheid, 2014).

As described above, epigenetic regulation plays a pivotal role in the development of organisms and the adaptation to environmental cues, thus it is not surprising that deregulation of these epigenetic processes is found in a multitude of diseases, most notably in cancer (Portela and Esteller, 2010). Continual innovations in genome sequencing allowed the sampling of thousands of cancer patients and demonstrated that about 50% of human cancers harbor mutations that are involved in modulating epigenetic networks (Jones et al., 2016). Moreover, epigenetic alterations not only support the initial tumor growth but also help tumor cells to escape chemotherapy and the immune surveillance of the patient (Jones et al., 2016). Hence, drugging the epigenome became a promising therapeutic approach and first epigenetic drugs have recently been approved by the US Food and Drug Administration (FDA) (Qi et al., 2016).

1.2 The Epigenetic Tool Box

DNA molecules need to be stored in a highly compact fashion within the nucleus of a cell. DNA is therefore not present as a naked and loose strand, but is neatly wrapped around the so-called nucleosome, an octameric complex comprised of histone proteins (Kornberg and Thomas, 1974). These DNA/protein units form the centerpiece of chromatin and several of them in a row are known as the “beads on a string” structure that in turn represents the first layer of chromatin organization (Olins and Olins, 1974). Chromatin in this compaction state is generally referred to as “euchromatin” with transcriptional activity (Kouzarides, 2007) (Figure 2a). If multiple nucleosomes assemble together the chromatin gets more condensed, a state defined as “heterochromatic” where transcription of genes is mainly repressed (Kouzarides, 2007) (Figure 2b). Thus, chromatin organization influences gene transcription predominantly by affecting the accessibility to certain genomic regions. However, the precise regulation of transcription is accomplished by covalent epigenetic marks on chromatin that are dynamically coordinated by writers, readers and erasers of the epigenetic machinery (Soshnev et al., 2016).

1.2.1 DNA modifications

Chemical modifications of DNA were observed long before the era of epigenetic research begun to flourish (Hotchkiss, 1948). Nowadays methylation of the 5th carbon atom of cytosine, yielding 5mC on DNA, is presumably the most studied epigenetic mark in mammals. Set by so-called DNA methyltransferases (DNMTs) (Lyko, 2018), 5mC is mainly found in the context of CpG dinucleotides (Feng et al., 2010) and its presence was early linked to transcriptional repression (Razin and Riggs, 1980). However, recent large-scale analyses of methylomes showed that gene bodies of transcribed genes can be heavily methylated (Ball et al., 2009), therefore indicating that DNA methylation is versatile and goes beyond simple repression of gene expression (Zhu et al., 2016). Methylation of DNA is found across many species, however, the methylation density of the genome varies greatly in different organisms (Zhu et al., 2016). Whereas invertebrates feature rather low levels of methylation, somatic tissues of mammals are known to be methylated at 70-80% of all CpG sites (Li and Zhang, 2014), with the key exception of CpG islands (Suzuki and Bird, 2008).

5mC, often termed the 5th base of DNA due to its high incidence and importance in our genome, was long considered to be the only influential modification (Kumar et al., 2018). In 2009 however a new group of enzymes was discovered and with them three more DNA modifications, namely 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Wu and Zhang, 2017) (Figure 1). Each of these modifications represents an oxidized derivative of 5mC and they were initially discussed as simple intermediates in an active DNA demethylation process (for details see chapter 1.4.1). Even though their levels across the genome are magnitudes lower compared to 5mC (e.g. 5hmC is 14x lower than 5mC in mouse embryonic stem cells (mESCs) (Tahiliani et al., 2009)), mapping of the oxidized modifications revealed a distinct distribution in certain genomic regions (Neri et al., 2015; Shen et al., 2013; Wen et al., 2014). 5hmC is found at genic regions and distal regulatory elements (Wen et al., 2014), whereas fC and caC are enriched at active promoters, enhancers and repetitive elements (Neri et al., 2015; Shen et al., 2013). These studies conclusively demonstrate that hmC, fC and caC occur in non-random patterns, which points to distinct biological roles of these modifications.

To better understand the function of those modifications including 5mC, researchers made huge attempts to identify readers of methylated and oxidized cytosines (Song and Pfeifer, 2016). For instance, the MBD protein family in mammals harbors a methyl-CpG-binding domain (MBD) and

binds methylated CpGs in a non-sequence specific manner (Hendrich and Tweedie, 2003). How important it is to properly recognize and process methylation patterns is demonstrated by mutations in the gene encoding MeCP2, a member of the MBD family (Lewis et al., 1992), that leads to the neurodevelopmental disorder Rett syndrome (Amir et al., 1999). Besides the classical MBD-containing methyl-binders, some transcription factors (TFs) like KAP1 were shown to interact with methylated cytosines in a sequence-dependent manner opening the opportunity of transcriptional activation at methylated sites (Quenneville et al., 2011). Unique interactors were also identified for hmC, fC and caC that seem to bind their substrates in a context-specific manner (Spruijt et al., 2013). In a mass-spectrometry-based proteomic approach, Spruijt et al. (2013) classified dynamic readers during the neural differentiation of mESCs and described for example UHRF2 as a high-affinity binder of hmC in neural progenitor cells (NPCs) and TDG as a specific binder for fC and caC. Determining specific interactors of the respective modifications tremendously helps to understand how this information is converted within the complex epigenetic network and how signals are ultimately transferred to other epigenetic layers (Zhu et al., 2016). Moreover, 5fC was reported to cause helical unwinding by facilitating the conformational transition of DNA from its B to F form (Raiber et al., 2015) and modification of DNA bases generally influences the thermostability of the DNA double-helix (Zhang et al., 2017a). Hence, such modifications can affect the shape and intrinsic properties of DNA, which in turn could also alter DNA-protein interactions.

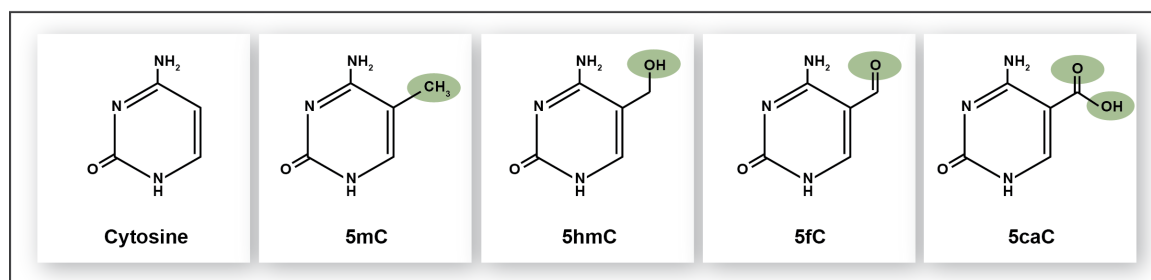


Figure 1: DNA cytosine modifications

Unmodified cytosine represents one of four major DNA bases and can be methylated at its 5th carbon atom of the pyrimidine ring through the activity of DNA methyltransferases, yielding 5-methylcytosine (5mC). Successive oxidation of the methyl group mediated by the enzymatic activity of TET proteins generates 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).

1.2.2 Histone modifications

Modification of histones, the protein units of nucleosomes, represents another layer of epigenetic control that was heavily investigated over the past twenty years (Allis and Jenuwein, 2016). Histones are highly alkaline proteins sub-divided into five major families: H1/H5, H2A, H2B, H3, and H4. Two copies each of the four latter ones build the actual core nucleosome, whereas H1/H5 known as the “linker histones” sit at the sites where the DNA enters and exits the nucleosome (Arents et al., 1991; Luger et al., 1997). The N-termini or “tails” of the core histones are broadly post-translationally modified and the repertoire of histone marks range from acetylation, methylation, phosphorylation and ubiquitylation to sumoylation, ADP ribosylation and propionylation (Bannister and Kouzarides, 2011). Each of these chemical moieties is established by a distinct set of “writer” enzymes harboring specific domains for their catalytic activity. Methylation marks for instance are established by lysine methyltransferases (KMTs) containing a SET domain (Ng et al., 2009) and histone acetylation is accomplished by a huge family of histone acetyltransferases (HATs) (Yang and

Seto, 2007). Interestingly, as for DNA methylation, all histone modifications are reversible, which implies that other enzymes are involved in the removal of covalent histone marks. Indeed, several groups discovered epigenetic modifiers acting as “erasers”, such as lysine demethylases (KDMs) (Shi et al., 2004; Tsukada et al., 2006) and histone deacetylases (HDACs) (Yang and Seto, 2007). Pioneering work from Allfrey in 1964 already suggested a critical role for modified histone residues in the regulation of gene activity (Allfrey et al., 1964). Later, Jenuwein and Allis introduced the histone code hypothesis that describes how these histone modifications together serve as a combinatorial pattern or code for the binding of effector molecules (or “readers”) to modulate chromatin structure and gene transcription (Jenuwein and Allis, 2001). Important for transcriptional regulation is not only the type of chemical modification but also its position in the tail’s amino acid sequence. Trimethylation of lysine K9 or K27 on H3 tails (H3K9me₃/H3K27me₃) both constitute classic marks of transcriptional silencing, whereas the same modification at other lysines, e.g. H3K4me₃, marks active euchromatin (Greer and Shi, 2012) (Figure 2). More recent findings demonstrated that histone modification exceeds the “simple” alteration of histone tails and several marks have been described within the histone core regions (Lawrence et al., 2016). Especially the core modifications at the lateral surface are of great interest regarding their direct contact with the DNA. Albeit most of the “core readers” remain to be identified, first analyses revealed distinct functions for the modifications themselves (Lawrence et al., 2016). For example, H3K122ac was shown to evict histones from DNA *in vitro*, thereby increasing transcriptional activity (Tropberger et al., 2013). As with the modified histone tails, many modification sites within the globular domains correspond to opposing gene states depending on the type of modification. Methylation of H3K56 for instance leads to a repressed state (Jack et al., 2013), acetylation of the same site on the contrary allows active transcription (Tjeertes et al., 2009) (Figure 2).

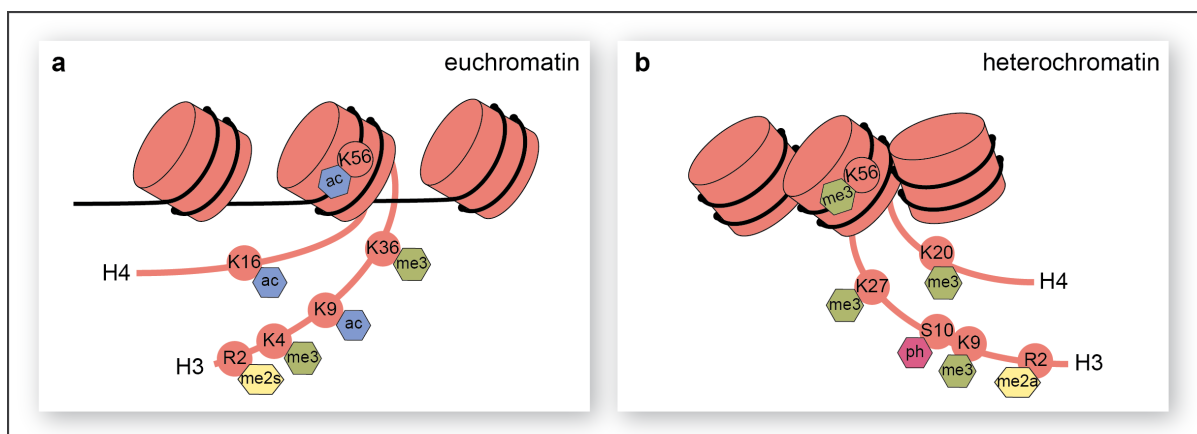


Figure 2: Histone modifications in euchromatin and heterochromatin

Histone (H) proteins can post-translationally be modified at residues of their protruding N-terminal tails and at core residues located on the lateral surface. Post translational modifications (PTMs) are mainly found on lysine (K), arginine (R) and serine (S) residues and include amongst others: acetylation (ac), phosphorylation (ph), di- and tri-methylation (me₂ and me₃). Different PTMs at the same residue, e.g. K56, are often specifically linked to open euchromatin or condensed heterochromatin, s: symmetric; a: asymmetric.

Besides such well-known modifications, many novel ones, e.g. crotonylation (Tan et al., 2011), have been discovered in recent years, which need to be characterized to fully understand the functional network of histone modifications in epigenetic control (Lawrence et al., 2016).

Evolutionarily, eukaryotic histones evolved from archaeal histones evidenced by astonishing structural similarities of their nucleosomes (Henikoff and Smith, 2015). More precisely,

superimposing archaeal histone tetramers with the tetrameric H3-H4 unit of eukaryotes leads to a nearly perfect fit. However, eukaryotes doubled the number of histone subunits, which in turn allows for two DNA wraps instead of only one in archaea (Ammar et al., 2012). On top, histone tails do not exist in archaeal ancestors and were just acquired during eukaryotic evolution. How these packaging units and their alterations evolved in detail remains uncertain, but the fact that histones evolved already million years ago suggests that DNA packaging is an indispensable requirement of cellular organisms (Henikoff and Smith, 2015).

1.2.3 Histone variants

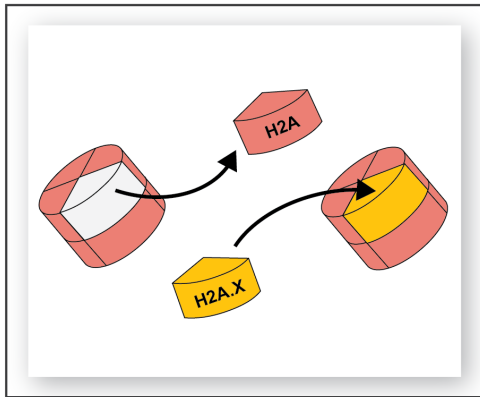


Figure 3: Histone variants

The octameric complex of nucleosomes is formed by eight histone proteins. Whereas canonical nucleosomes comprise two copies of each of the four core histone proteins (H2A, H2B, H3 and H4), epigenetic regulation can also substitute single histones by histone variants with specific functions like H2A.X.

As described before, the core of nucleosomes comprises four different types of histones: H2A, H2B, H3 and H4 (Arents et al., 1991; Luger et al., 1997). These proteins make up the bulk and represent the canonical types of histones, but other so called “histone variants” are also found to be incorporated into nucleosomes (Henikoff and Smith, 2015) (Figure 3). Surprisingly, the diversification among the four groups of histones is not even and variants only exist for H2A and H3. H2B and H4 lack different paralogs, possibly due to different evolutionary forces acting at different positions within the nucleosome (Henikoff and Smith, 2015). One major discrepancy between canonical histones and histone variants is the temporal deposition during the cell cycle. Whereas canonical histone production and deposition is coupled to DNA synthesis during S-phase (Marzluff et al., 2002), histone variants are

deposited independently from replication (Alabert et al., 2015). For instance, a histone chaperone complex that assists the deposition of canonical H3 and H4 is known to directly interact with the proliferating cell nuclear antigen (PCNA), a replication processivity clamp (Smith and Stillman, 1989; Zhang et al., 2016b). During replication the bulk histones are incorporated into nucleosomal gaps that occur behind the replication fork. For the incorporation of histone variants on the other hand existing nucleosomes or subunits get replaced, which can take place throughout the entire course of the cell cycle (Henikoff and Smith, 2015).

One prominent H3 variant in mammals is CENP-A that is specifically found in the nucleosomes of centromeres (Palmer et al., 1990) and many studies have proven its essentiality at these genomic loci (Talbert and Henikoff, 2010). CENP-A or respective counterparts (cenH3s) in other eukaryotes are a prerequisite for assembling the kinetochore and for segregating chromosomes during mitose and meiose (Amor et al., 2004). Another H3 variant differing from its canonical form by only four amino acids, H3.3, is mainly present in active chromatin illustrating that variants can mark specific transcriptional states (Filipescu et al., 2013). Additionally, H3.3 is necessary for proper germline function by resetting the chromatin of gametes to a totipotent-like state (Santenard et al., 2010). The H2A.X variant belonging to the H2A histone family possesses a sequence motif that undergoes rapid phosphorylation when DNA double-strand breaks (DSBs) occur. This in turn could either stabilize the chromatin in the vicinity of DSBs or it represents a recruiting signal for the DSB repair machinery to restore the DNA helix (Lowndes and Toh, 2005; Morrison and Shen, 2005). Another

family member, H2A.Z, shows only 60% sequence identity to canonical H2A and its deletion in mouse is lethal pointing towards a non-redundant and essential role for H2A.Z (Faast et al., 2001). Accordingly, H2A.Z's function in regulating chromatin is extremely diverse and partially even contradictory and further work is needed to pinpoint which of these functions confers its essentiality (Gaiimo et al., 2019). Other variants like H2A.B and macroH2A participate in the X inactivation process (Chadwick et al., 2001), highlighting a role for histone variants in very specific epigenetic phenomena (Brockdorff and Turner, 2015). Generally, histone variants may form the most elementary level of chromatin differentiation and allow for the establishment of distinct epigenetic states and processes like chromosome segregation (Henikoff and Smith, 2015).

1.2.4 Nucleosome remodeling

Considering that chromatin is mainly a compact structure ensuring the tight packaging of our genome, the accessibility to most genomic regions is low. As this is generally desirable for silenced genes in heterochromatic regions, DNA in active loci requires to be accessible (Becker and Workman, 2013). Otherwise regulatory factors like TFs would not be able to bind their target sequences and transcriptional initiation would be impeded (Yin et al., 2017). Other activities like DNA damage repair also need direct access to DNA and nucleosomes constitute a hindrance in this context (Soria et al., 2012). Correspondingly, enhancers, promoters and other active regulatory elements were shown to be free of canonical nucleosomes as evident by their considerably high sensitivity to DNase treatment (Boyle et al., 2008; Reik et al., 1991). Since these accessibility requirements are locus-specific, eukaryotic chromatin seems to possess a mechanism by which it dynamically regulates its accessibility. In fact, a group of ATP-dependent enzymes is responsible for the “remodeling” of nucleosomes (Cairns, 2007). These nucleosome remodeling factors all contain an ATPase domain that is evolutionarily related to nucleic acid helicases and the remodeling factors usually associate with other proteins into multisubunit complexes (Flaus et al., 2006). These complexes can range from a few subunits to more than a dozen as seen for the large INO80 remodeler (Becker and Workman, 2013). Depending on specific sequence features within the ATPase domain, nucleosome remodelers are classified into six subfamilies (Flaus et al., 2006). Even though the molecular mechanism is not fully understood, these enzyme complexes are thought to act as DNA translocases moving along the DNA, thereby making contact to particular histones and linker DNA. Once correctly positioned, a conformation change enabled by its ATPase activity pulls on the DNA wrapped around the nucleosome, which results in the detachment of a DNA section from the histone octamer and the formation of a DNA bulge (Gangaraju and Bartholomew, 2007; Backer and Workman, 2013). The displacement of DNA in turn could serve as the first step for the delocalization of the respective nucleosome, also termed “sliding” of nucleosomes. However, detaching the DNA from the histone core could possibly also initiate the replacement of a histone, the incorporation of a specific variant or the eviction of an entire nucleosome. Securing the correct distance between individual nucleosomes, known as nucleosome “spacing”, is also thought to be a function of nucleosome remodeling (Becker and Workman, 2013) (Figure 4). Some groups additionally presented the cooperation of ATP-dependent remodelers with histone chaperones to assist in nucleosome assembly (Burgess and Zhang, 2013).

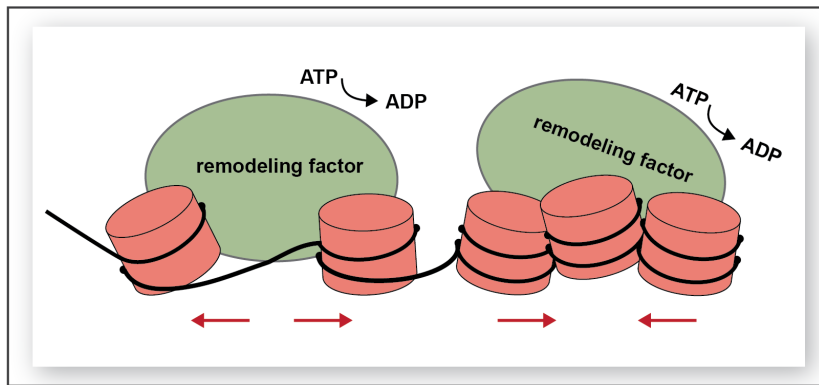


Figure 4:
Nucleosome remodeling
 To modulate the accessibility to certain genomic loci, ATP-dependent nucleosome remodeling factors enable the precise positioning of single nucleosomes, thereby defining their distance to each other, a mechanism known as nucleosome “spacing”.

One major conception of nucleosome remodeling research is that different types of remodelers can comprise the same ATPase but differ in the composition of associating subunits (Yadon and Tsukiyama, 2011). These associated proteins are often equipped with specialized domains recognizing methylated DNA, post-translationally modified histones and other proteins. This in turn allows the same ATPase to carry out diverse nuclear functions (Li et al., 2006; Syntichaki et al., 2000). For the correct recruitment of remodeling factors to the sites of transcriptional initiation for example, they often interact with sequence-specific transcription factors of the respective target gene (Bowman and McKnight, 2017). Taken together, the activity of nucleosome remodelers plays a pivotal role in different aspects of genome organization. It does not only contribute to transcriptional control, but also facilitates histone variant exchange and ensures the accurate folding and integrity of chromatin fibers (Becker and Workman, 2013).

1.2.5 Non-coding RNA

Non-coding ribonucleic acids (ncRNAs) are defined as functional RNA molecules that are transcribed from DNA but not further translated into protein (Palazzo and Lee, 2015). They gained great attention when the RNA interference (RNAi) pathway was discovered in 1998 (Fire et al., 1998). Extensively studying this silencing mechanism revealed that either endogenous micro RNAs (miRNAs) or exogenous short interfering RNAs (siRNAs) are processed into short fragments by a collection of key RNAi enzymes. These small and single-stranded RNA moieties in turn are able to base-pair with homologous mRNA sequences, which leads to the cleavage or degradation of the mRNA molecules and finally prevents their translation (Wilson and Doudna, 2013). However, miRNAs and siRNAs are not only involved in the regulation of target mRNAs but a growing body of evidence points towards a critical role in regulating chromatin. Based on their size, epigenetic-related ncRNAs can be split into two categories: long ncRNAs and short chain ncRNAs, including miRNAs and siRNAs but also PIWI-interacting RNAs (piRNAs) (Wei et al., 2017).

If one considers the specificity with which certain genomic loci are silenced (or re-activated), one recurring question is how chromatin marks are guided to the desired loci. There is growing evidence now that ncRNAs can confer the sequence specificity for chromatin modifying enzymes to find their respective targets (Aufsatz et al., 2002; Mochizuki et al., 2002; Volpe et al., 2002) (Figure 5). Studies in *Arabidopsis thaliana* for example revealed that its de novo DNA methyltransferase is targeted to specific sequences by siRNAs (Chan and -L. Chan, 2004). In mice on the other hand, a cluster of miRNAs was shown to control differentiation of embryonic stem cells by degrading a repressor of de novo methyltransferases, thereby securing methylation at the Oct4 promoter, a key transcription factor of pluripotent mESCs (Sinkkonen et al., 2008). piRNAs were mainly studied in *Drosophila melanogaster* and have been established as important guardians of transposon activity, i.e.

preventing transposon activation (Malone et al., 2009). In mammals, piRNAs are required for the silencing of transposons in germ cells (Kuramochi-Miyagawa et al., 2004).

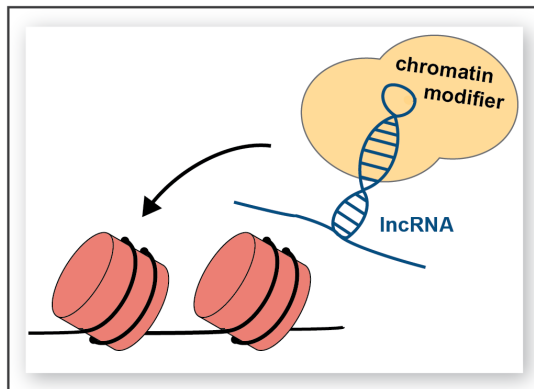


Figure 5: Non-coding RNAs

Short chain and long non-coding RNAs (lncRNAs) can serve as recruiting signals or guiding molecules for chromatin-modifying enzymes to be precisely targeted to specific sequences in the genome.

In contrast to short chain ncRNAs with an average size of 20-30 bp, long ncRNAs are composed of > 200 bp (Wei et al., 2017). The most prominent member of the long ncRNA family is Xist (17 kb), a long ncRNA that is involved in X chromosome inactivation. In a defined interplay with its antisense transcript Tsix (40 kb), Xist expression gets upregulated upon differentiation and Xist molecules start coating the future inactive X chromosome (Xi). This in turn recruits a chromatin repressive complex that establishes extensive histone methylation for silencing the Xi (Heard et al., 2001). Another famous contribution of long ncRNAs has been observed in the regulation of

HOX genes that encode proteins essential for embryonic development. The physical interaction of the HOTAIR RNA with the Polycomb Repressive Complex 2 (PCR2) directs the formation of H3K27me3 to the HoxD gene locus that ultimately leads to its transcriptional repression (Tsai et al., 2010). Besides that, several other studies also demonstrated a role for ncRNAs in allele-specific silencing of epigenetically imprinted genes (O'Neill, 2005).

1.3 Molecular Basis of DNA methylation

As mentioned earlier, methylation of the fifth carbon atom of cytosine (5mC) represents the prime example of DNA modifications (Kumar et al., 2018) (cp. chapter 1.2.1 and figure 1). However, DNA methylation was also discovered at nitrogen atoms of adenosines (6mA) in eukaryotes, albeit very low levels compared to prokaryotic DNA (Koziol et al., 2016). The enzymes responsible for methylating DNA are so called DNA methyltransferases (DNMTs) with five major cytosine-DNMTs found in human and mice: DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. They share one collective feature, namely a catalytic domain composed of ten signature motifs, but exhibit limited structural similarity otherwise (cp. figure 6) (Lyko, 2018). Counterintuitively, DNMT2 and DNMT3L are catalytically inactive regarding the methylation of DNA and execute their functions on other substrates and as a co-factor, respectively (Bourc'his et al., 2001; Goll et al., 2006). During the enzymatic reaction of DNMTs a covalent intermediate between enzyme and cytosine is formed and s-adenosylmethionine (SAM) is used as the methylgroup donor (Wu and Santi, 1985). To gain full access to their substrate, DNMTs utilize a base-flip mechanism that positions the cytosine base directly in their catalytic pocket (Klimasauskas et al., 1994). Historically, DNMTs are divided into two groups based on their role in de novo and maintenance methylation.

1.3.1 De novo methylation

Taking “de novo” literally means “starting from the beginning”. Hence, de novo methylation describes the establishment of methylation. Responsible for setting these new methylation marks in mammals are DNMT3A and DNMT3B (Okano et al., 1999). Both enzymes contain three main

structure domains: a) the MTase domain located in the carboxy-terminus, which secures the enzyme's catalytic activity, b) a Pro-Trp-Trp-Pro (PWWP) domain and c) the ATRX-DNMT3-DNMT3L (ADD) domain (Figure 6). The two latter ones are necessary for binding chromatin and both play an important role in regulating the recruitment of DNMT3A and B (Lyko, 2018). The PWWP domain binds to methylated histone tails including H3K36me₃, a mark that is often found in the bodies of actively transcribed genes (Baubec et al., 2015; Dhayalan et al., 2010). Mutations within the PWWP domain not only impedes the recruitment of DNMT3B to these H3K36me₃-marked gene bodies in mESCs, but also abolishes the methylation reaction at major satellite repeats in humans (Baubec et al., 2015; Shirohzu et al., 2002). In contrast to gene bodies, the promoters of active genes are predominantly free of DNA methylation and exhibit high levels of H3K4me₃ (Piunti and Shilatifard, 2016). H3K4 sites are sensed by the ADD domain of DNMT3 enzymes and trimethylation of K4 inhibits their binding (Otani et al., 2009). As a consequence, the unengaged ADD domain folds back and binds to the MTase domain thereby leading to the auto-inhibition of the catalytic domain and thus abolishes methylation activity (Guo et al., 2015). Albeit its catalytic inactivity, DNMT3L also possesses an ADD domain that mediates its targeting to unmodified H3K4 sites. Through its complex formation with DNMT3A and B, DNMT3L increases the affinity of these enzymes to chromatin (Jia et al., 2007; Ooi et al., 2007). In fact, the co-factor role of DNMT3L towards DNMT3A and B is especially important in the development of the germline (Bourc'his et al., 2001).

Moreover, *in vitro* studies revealed the ability of DNMT3A and B to form complexes with RNA that pointed towards a novel recruitment mechanism facilitated by sequence-specific RNA molecules (Jeffery and Nakielny, 2004). Recent reports confirmed this hypothesis by illustrating that the recruitment of DNMT3 enzymes and the establishment of methylation marks at specific genomic loci is indeed partially controlled by ncRNAs (Denis et al., 2011). De novo methyltransferases can further be regulated by PTMs. Phosphorylation of residues within the PWWP domain of DNMT3A for example targets the enzyme to heterochromatic repeats (Deplus et al., 2014). Sumoylation of DNMT3A and B on the other hand is thought to affect the protein-protein interaction with other chromatin modifiers thereby altering the methylation capacity (Kang et al., 2001; Ling et al., 2004). In summary, the molecular mechanisms regulating the methylation activity of DNMT3A and B are versatile to guarantee the precise establishment of methylation patterns. Interestingly, recent studies discovered a new de novo methyltransferase, DNMT3C (Figure 6). As yet, this homolog seems to be specific to the Muroidea superfamily (containing mice and rats) and its expression is restricted to male germ cells (Barau et al., 2016).

1.3.2 Maintenance methylation

1.3.2.1 DNMT1

Maintenance methylation refers to every methylation activity that preserves methylation marks where they have previously been established. Such methylation activities are mandatory in each DNA replication cycle where two hemi-methylated DNA strands are generated from one symmetrically-methylated strand, i.e. harboring the methyl group at CpG sites on both strands (Holliday and Pugh, 1975; Riggs, 1975). The methyltransferase that accomplishes this “replenishing reaction” and therefore ensures the inheritance of symmetric methylation patterns to daughter cells is DNMT1 (Li et al., 1992). Its structural composition, mainly in the N-terminal part, is substantially different to other DNMT family members (Figure 6). Dnmt1 comprises a CXXC-type zinc finger that specifically binds to unmethylated CpG dinucleotides and two bromo-adjacent homology

(BAH) domains, which are largely uncharacterized (Lyko, 2018). The replication foci targeting sequence (RFTS) determines the nuclear localization of DNMT1 in a cell-cycle dependent manner (Easwaran et al., 2004; Leonhardt et al., 1992). During S-phase, this targeting mechanism is further supported through DNMT1's interaction with PCNA, an essential component of replication forks, that is mediated by its PCNA interacting motif, the (PIP)-box (Chuang et al., 1997). Interestingly, the RFTS domain is not only involved in targeting, but also in regulating the enzymatic activity of DNMT1 by intramolecularly blocking the catalytic domain in the absence of DNA (Takeshita et al., 2011). Another autoinhibitory mechanism is provided by a small linker located between the CXXC and the BAH domains. Crystal structures of mouse and human DNMT1 revealed that the linker is positioned between the catalytic domain and the DNA if the CXXC-type zinc finger is bound to unmethylated DNA (Song et al., 2011). Whereas the targeting and regulation of DNMT1 is attributed to its N-terminal part, the catalytic activity of DNMT1 is mediated by its C-terminus, analogous to all the other DNMTs (Fatemi et al., 2001).

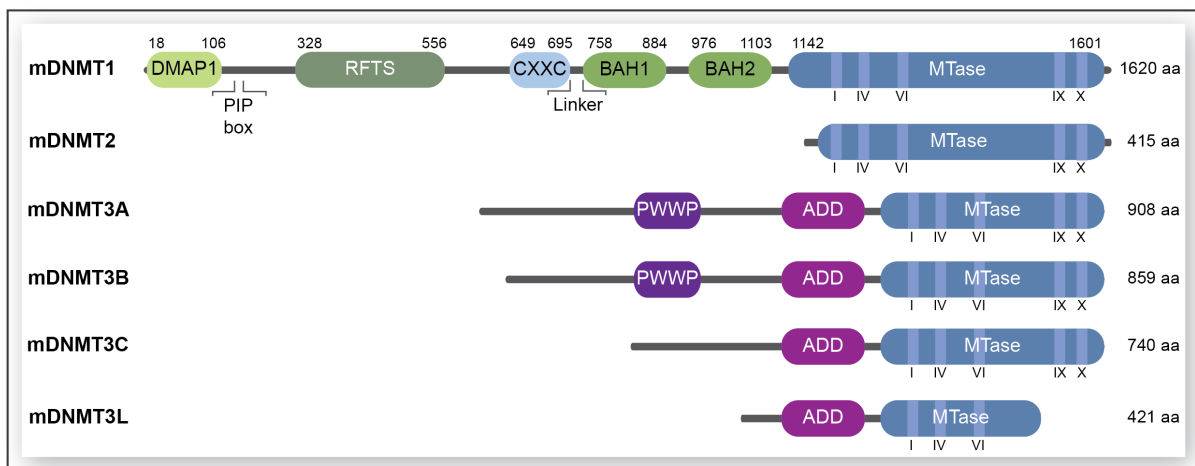


Figure 6: Domain structure of the DNMT family

All members of the murine DNA methyltransferase (DNMT) family comprise a C-terminal methyltransferase (MTase) domain. DNMT3 homologs harbor an ADD domain, DNMT3A/B additionally possess a PWWP domain. The N-terminus of DNMT1 features various regulatory domains, incl. DMAP1: DMAP1 interacting domain; PIP-box: PCNA interacting peptide box; RFTS: replication foci targeting sequence; a CXXC zinc finger and two BAH: bromo-adjacent homology domains. ADD: ATRX-Dnmt3-Dnmt3L domain, PWWP: Pro-Trp-Trp-Pro motif domain. DNMT1 domains are labeled with respective residue numbers, aa: amino acids.

1.3.2.2 UHRF1

To faithfully propagate methylation patterns, DNMT1 works in a tandem together with the ubiquitin-like containing plant homeodomain and really interesting new gene finger domains 1 (UHRF1) (Bostick et al., 2007; Sharif et al., 2007). The tissue-specific expression of UHRF1 is closely related to the expression of DNMT1 and knocking-out UHRF1 phenotypically resembles the knock-out of DNMT1 (Fagerberg et al., 2014; Sharif et al., 2007). This in turn emphasizes the close relationship between the two proteins and hints at an interdependent mode of action. Uhrf1, also known as NP95 in mice, is a multi-domain protein comprising five major domains: a ubiquitin-like (UBL) domain, a tandem tudor (TTD) domain, a plant homeodomain (PHD), a SET- and RING-associated (SRA) domain and one enzymatically active domain, the really interesting new gene (RING) domain with ubiquitin ligase activity (Xie and Qian, 2018) (Figure 7a). Whereas the TTD and PHD domain are required for binding histones (H3K9me2/3 and unmodified H3R2, respectively), the SRA domain facilitates UHRF1's binding to hemi-methylated DNA that is

generated during DNA replication (Avvakumov et al., 2008). In the absence of hemi-methylated DNA however, UHRF1 adopts a closed conformation with its spacer region (aa 642-659 in human) contacting the TTD domain (Fang et al., 2016). This intramolecular rearrangement impedes not only the binding of TTD to H3K9me2/3, but also the appropriate positioning of the PHD and SRA domain towards their targets and as such constitutes an auto-inhibitory mechanism (Fang et al., 2016) (Figure 7c).

But how exactly do these features of UHRF1 coordinate the methylation reaction of DNMT1? Numerous studies proved UHRF1 to be an indispensable co-factor of maintenance methylation as it recruits DNMT1 to the appropriate genomic loci (Bostick et al., 2007; von Meyenn et al., 2016; Sharif et al., 2007). UHRF1 performs this recruitment by two means: firstly, by providing recruitment signals on chromatin that are recognized by DNMT1. Secondly, through directly interacting with DNMT1 (Bronner et al., 2019). To set recruiting signals for DNMT1, UHRF1 exploits its ubiquitin ligase activity and ubiquitinates specific lysine residues on histone H3, namely K14, K18 and K23 in mammals (Ishiyama et al., 2017; Nishiyama et al., 2013; Qin et al., 2015). These ubiquitination moieties are bound by a ubiquitin interacting motif (UIM) within the RFTS domain of DNMT1 (Misaki et al. 2016; Qin et al., 2015). Subsequent studies observed that the enzymatic activity of UHRF1 is dependent on the stimulating effect of its UBL domain and the activity is further enhanced through binding to hemi-methylated DNA (DaRosa et al., 2018; Foster et al., 2018; Harrison et al., 2016).

Interestingly, the ubiquitin ligase activity of UHRF1 is not restricted to histones. Earlier studies described DNMT1 and the promyelocytic leukemia (PML) protein as ubiquitination targets (Guan et al., 2013; Qin et al., 2011) and UHRF1 was also shown to exhibit auto-ubiquitination activity (Vaughan et al., 2018). Ubiquitination of these proteins decreases their stability and may end up in proteasomal degradation if no compensating mechanisms are present. DNMT1's degradation for example is antagonized by the ubiquitin specific peptidase 7 (USP7), a deubiquitinating enzyme that mediates not only deubiquitination of DNMT1 (Qin et al., 2011), but also of UHRF1 and H3 residues (Yamaguchi et al., 2017; Zhang et al., 2015). More recently, a ubiquitinome analysis in UHRF1^{-/-} mESCs comprehensively screened for further targets and identified PCNA-associated factor of 15 kDa (PAF15) as a major ubiquitination target of UHRF1 (Karg et al., 2017). PAF15 is an intrinsically disordered protein with a histone-like N-terminal tail that is proposed to assist PCNA in sliding along the DNA (De Biasio et al., 2015). Moreover, ubiquitination of PAF15 seems to play a crucial role in replication-fork-blocking lesions by regulating the recruitment of translesion DNA synthesis polymerases, thereby safeguarding genome integrity (Povlsen et al., 2012).

To secure accurate recruitment of DNMT1, UHRF1 not only sets ubiquitination marks for DNMT1 but also directly interacts with the protein, mediated by the UHRF1-SRA and the DNMT1-RFTS domain (Berkyurek et al., 2014). The SRA domain itself specifically recognizes hemi-methylated CpG dinucleotides and flips the methylated cytosine out of the DNA helix (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). This mechanism is enabled by the "hand-like" structure of the SRA domain with a thumb- and finger-like region, which together tightly grab the DNA helix to position the methylated cytosine appropriately (Avvakumov et al., 2008) (Figure 7b). Through its direct interaction with UHRF1, DNMT1 is situated in close proximity to its substrate and UHRF1's base-flipping is thought to make room for DNMT1 to methylate the cytosine on the opposite strand (Bronner et al., 2019). Besides actively binding hemi-methylated sites, UHRF1's SRA domain was shown to bind 5hmC with similar affinity, presumably in a context-specific manner (Frauer et al., 2011). Also, UHRF1 interacts not only with DNMT1 but also with the regulatory domains of DNMT3A and B in mESCs (Meilinger et al., 2009).

In a recent review article, C. Bronner and colleagues (2019) made great effort in chronologically order the different steps of UHRF1-assisted maintenance of DNA methylation and introduced the following model: First, UHRF1 interacts with hemi-methylated DNA via its SRA domain and binds H3K9me3 via the TTD domain. This leads to conformational rearrangements of the protein and allows the RING domain to dual ubiquitinate histone H3 and/or itself. The ubiquitin moieties on H3 and/or UHRF1 serve as anchorage signals for DNMT1 and DNMT1 directly interacts with UHRF1 via its RFTS domain. This in turn abrogates the inhibitory function of the RFTS towards the methylation activity of the enzyme and DNMT1 can finally set the missing methylation mark on the newly-synthesized DNA strand. Noteworthy, UHRF1 was lately reported to be recruited to replication sites by a second means, namely the direct and avid binding to methylated DNA ligase 1 (LIG1), a member of the replication machinery (Ferry et al., 2017). Intriguingly, LIG1 possesses an H3K9-like mimic within its protein sequence, which is methylated by the G9a/GLP methyltransferase (Ferry et al., 2017), the same enzyme responsible for methylating H3K9 (Tachibana et al., 2001, 2005).

There is another UHRF family member, UHRF2, that shares a highly similar domain architecture with UHRF1 (Bronner et al., 2007) and for instance was shown to be a reader with increased affinity for hmC in neuronal progenitor cells (Spruijt et al., 2013). However, albeit its related composition, UHRF2 cannot compensate for the loss of UHRF1 in mESCs and is generally not considered to be as critical for epigenetic regulation as UHRF1 (Pichler et al., 2011; Bronner et al., 2019). If at all, UHRF2 seems to play a role in the epigenetic control of differentiated cells (Pichler et al., 2011).

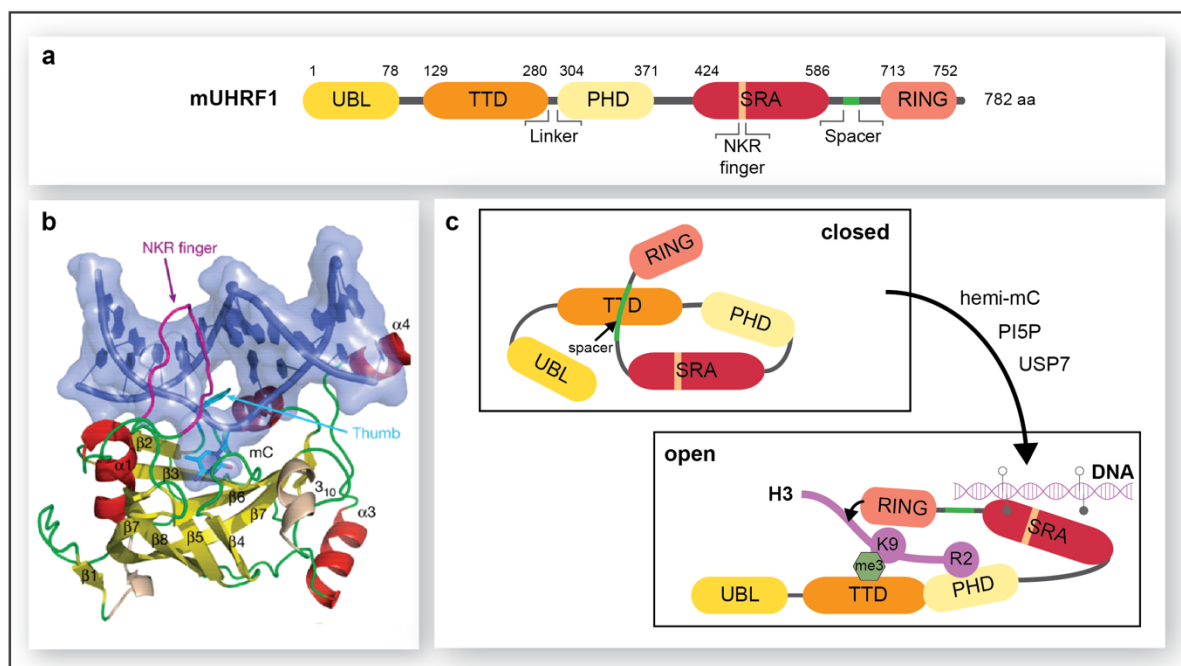


Figure 7: Domain architecture and conformational states of UHRF1

(a) Functional domains of mouse UHRF1 with individual domains labeled with respective residue numbers. UBL: Ubiquitin-Like domain; TTD: Tandem Tudor Domain; PHD: Plant Homeodomain; SRA: SET and Ring-Associated domain; RING: Really Interesting New Gene domain with E3 ligase activity, aa: amino acids. (b) Crystal structure of the hand-like SRA domain bound to DNA (from Avvakumov et al. 2008, PDB:3CLZ) (c) Potential conformations of UHRF1 as a function of the enzyme's activity status, PI5P: phosphatidylinositol 5-phosphate

1.4 Molecular Basis of DNA demethylation

1.4.1 Active demethylation

Almost equally important to methylation of DNA is the removal of methyl groups from DNA, a process known as DNA demethylation. Astonishingly, the molecular mechanism for active demethylation was not revealed until 2009 when two groundbreaking papers discovered 5hmC and linked TET1 to its formation (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Originally, the naming of TET1 stems from its identification as a fusion partner of MLL in acute myeloid leukemia containing a ten-eleven translocation (TET) (Lorsbach et al., 2003). Besides TET1, two other family members, TET2 and TET3, catalyze the iterative oxidation of 5mC to 5hmC (Ito et al., 2010) and further to 5fC and 5caC (He et al., 2011; Ito et al., 2011). These oxidized cytosine variants in turn can result in DNA demethylation by two different means: The unmodified cytosine can either be restored through the replication-dependent dilution of oxidized 5mC or through the so called “TDG-BER pathway” (Kohli and Zhang, 2013). The former process is also known as active modification-passive dilution (AM-DP) and is justified by the observation that DNMT1 is less active at hemi-hydroxymethylated, -formylated or -carboxylated CpGs compared to hemi-methylated ones (Hashimoto et al., 2012; Ji et al., 2014; Wu and Zhang, 2017). Accordingly, the maintenance machinery is impaired which ultimately leads to demethylation after some rounds of DNA replication.

The TDG-BER pathway is also referred to as active modification-active removal (AM-AR) and includes the excision of fC or caC, but not hmC, by the thymine DNA glycosylase (TDG) (Maiti and Drohat, 2011; Wu and Zhang, 2017). The resulting abasic site is further processed by the base excision repair (BER) machinery. In fact, *in vitro* reconstruction of the TDG-BER pathway showed that AP endonuclease 1 (APE1) leads to a single-strand break at the abasic site, followed by the insertion of a deoxycytidine by DNA polymerase β (Pol β) and restoration of the single-strand break via XRCC1 and DNA ligase 3 (LIG3) (Weber et al., 2016). A study in mESCs revealed that NEIL glycosylases can substitute for TDG and similarly facilitate the restoration of unmodified cytosines (Müller et al., 2014). To date, the combined action of TETs, TDG and the BER enzymes has gained the most support and is therefore considered the main mechanism of active DNA demethylation (Wu and Zhang, 2017).

TET proteins belong to the family of iron(II)/ α -ketoglutarate (Fe(II)/ α -KG)-dependent dioxygenases and are structurally divided into a large, mainly uncharacterized, regulatory N-terminal part and the C-terminal catalytic domain (Figure 8a). The catalytic domain is comprised of a cysteine-rich and a double-stranded beta-helix (DSBH) (also jelly roll) domain, both of which assist in co-factor binding and stabilizing the TET-DNA interaction (Rasmussen and Helin, 2016). Whereas TET1 and TET3 feature a CXXC domain within their N-termini, the CXXC domain of TET2 evolutionary separated from its ancient protein through genomic inversion and became an independent protein named IDAX (Iyer et al., 2009).

The DNA demethylation activity of TET proteins can be regulated on different levels, including the availability of substrates and co-factors for the enzymatic reaction (Wu and Zhang, 2017). To perform successive oxidations, TET proteins require oxygen as a substrate and α -KG and Fe(II) as co-substrates/co-factors. During the two-electron oxidation reaction, TETs transfer one oxygen atom to the respective cytosine derivative supported by Fe(II) and one oxygen atom is incorporated into α -KG resulting in the formation of succinate and CO₂ (Loenarz and Schofield, 2011) (Figure 8b). The major portion of α -KG is produced within the citric acid cycle by so called isocitrate

dehydrogenases (IDHs), metabolic enzymes that catalyze the oxidative decarboxylation of isocitrate, generating α -KG. Thus, perturbing the enzymatic activity of IDH enzymes will influence α -KG levels and ultimately the reaction kinetics of TET enzymes (Dang and Su, 2017). This has been shown in mice administered with glucose, glutamate and glutamine, which resulted in increased α -KG levels and enhanced 5hmC levels in their liver tissue (Yang et al., 2014). Another study uncovered an important role for the phosphoserine aminotransferase 1 (PSAT1) in governing self-renewal and differentiation in mESCs by producing α -KG, thereby again enhancing 5hmC levels (Hwang et al., 2016). Also Fe(II) and even Vitamin C were shown to influence the catalytic activity of TETs as evidenced by iron-binding TET mutants that exhibit reduced catalytic activity (Laukka et al., 2016) and reversely, by increased TET activity after exogenously adding Vitamin C (Blaschke et al., 2013; Minor et al., 2013). The effect of oxygen on TET catalytic activity however is controversial. Whereas the hypoxia-inducible factor (HIF) was demonstrated to upregulate TET and 5hmC levels on the one hand (Mariani et al., 2014), hypoxia is also known to reduce 5hmC levels on the other hand (Thienpont et al., 2016).

Regulating TET-mediated DNA demethylation can further be achieved by post-transcriptional modulation of TET mRNA, mainly via microRNAs like miR-22, miR-29a or an entire network of miRNAs as reported for human TET2 (Cheng et al., 2013; Song et al., 2013; Zhang et al., 2013). Furthermore, TET proteins can covalently be modified after translation, for instance through GlcNAcylation, phosphorylation or acetylation (Bauer et al., 2015; Zhang et al., 2014, 2017c). These PTMs can influence the binding affinity to chromatin, enhance or weaken the enzymatic activity or change the subcellular localization of TET proteins (Wu and Zhang, 2017). Interaction with other proteins, e.g. IDAX in case of TET2, can alter the protein levels of TET by inducing proteolysis (Ko et al., 2013). Other interacting partners, like NANOG, LIN28A, WT1 and other transcription factors can additionally regulate the demethylation activity of TET proteins by the selective recruitment of TETs to specific genomic loci (Costa et al., 2013; Wang et al., 2015c; Zeng et al., 2016).

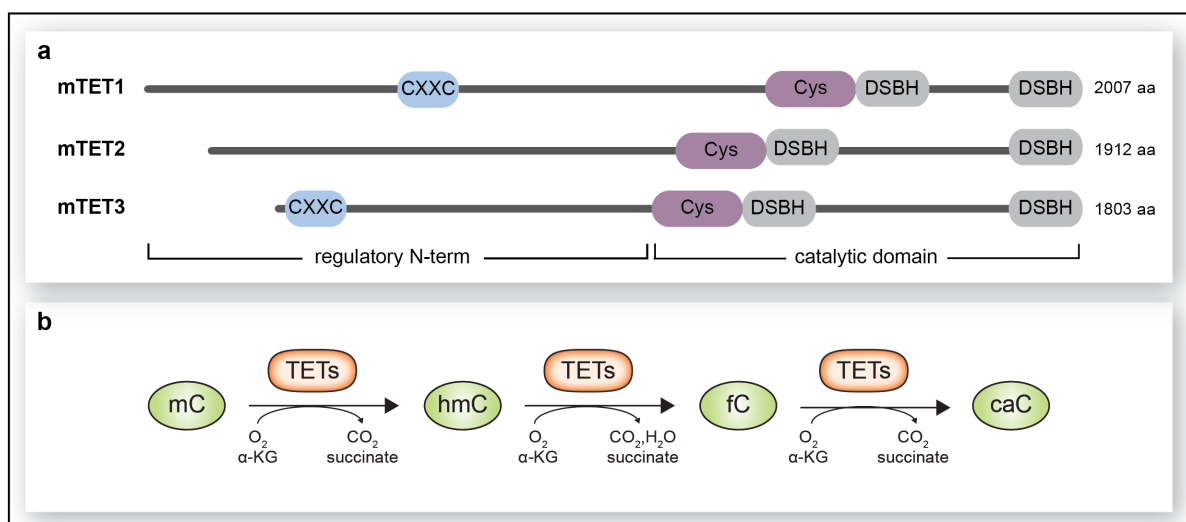


Figure 8: Domain structure of the TET family and its enzymatic activity

(a) The C-terminal catalytic domain of the three murine TET proteins comprise a cysteine (Cys)-rich region and two DSBHs. The regulatory N-terminus is mainly uncharacterized, except of the CXXC zinc finger domains in TET1 and TET3, DSBH: double-stranded beta helix; aa: amino acids. (b) TET enzymes can successively oxidize methylated cytosine (mC) to hydroxymethylcytosine (hmC), formylcytosine (fC) and carboxylcytosine (caC) while simultaneously converting alpha-ketoglutarate (α -KG) to succinate.

1.4.2 Passive demethylation

As elaborated on in the previous paragraph, active DNA demethylation is defined as removing the methyl group by means of an enzymatic reaction. On the contrary, passive demethylation is referred to a demethylation process that takes place when maintenance methylation is perturbed (Wu and Zhang, 2017). For instance, 5-azacytidine (5-aza) is a pharmaceutical demethylating agent that forms a covalent complex with DNMTs and thereby inhibits DNMT activity (Christman, 2002; Schermelleh et al., 2005). Accordingly, treating cells with 5-aza over several rounds of replication will globally lead to the “dilution” of methylation marks, i.e. global passive demethylation (Mund et al., 2005). Since DNMT1 accomplishes maintenance methylation in a tandem with UHRF1 (Bostick et al., 2007), passive demethylation can also occur by interfering with UHRF1. For example, a study in primordial germ cells (PGCs) of mice, which are known to undergo massive erasure of DNA methylation (Sasaki and Matsui, 2008), analyzed the expression levels of genes involved in DNA methylation/demethylation (Kagiwada et al., 2013). Having found DNMT1 to be highly expressed, but UHRF1 to be massively downregulated, the authors suggest that the absence of UHRF1 prevents DNMT1 localization to replication foci, thereby impeding normal maintenance methylation (Kagiwada et al., 2013). Funaki et al. (2014) further demonstrated that UHRF1 delocalization can equally hinder maintenance methylation. By overexpressing the developmental pluripotency-associated protein 3 (DPPA3), also known as Stella or PGC7, UHRF1 exhibited aberrant localization patterns and subsequently failed to recruit DNMT1. In accordance with this, global DNA methylation levels passively decreased upon DPPA3 overexpression in NIH3T3 cells (Funaki et al., 2014).

1.5 Cellular Functions of DNA methylation & demethylation

DNA methylation and demethylation are of tremendous importance in various processes of mammalian development. Massive epigenetic reprogramming is exceptionally pivotal to the pre-implantation development of embryos and the generation of primordial germ cells (PGCs). In both cases, dynamic and genome-wide changes in the methylation landscape confer epigenetic plasticity required to successfully accomplish the respective developmental stage (Greenberg and Bourc'his, 2019). DNA methylation is additionally known to play a fundamental role in repressing transcription especially of transposons and other repetitive elements to preclude genomic instability (Deniz et al., 2019).

1.5.1 Pre-implantation development

Upon fertilization of sperm and oocyte in mammals, DNA methylation patterns of the parental genomes need to get erased to assure the formation of totipotent cells and the removal of acquired epimutations, both mandatory for embryonic development (Greenberg and Bourc'his, 2019). Whereas the maternal genome within the zygote is thought to be mainly demethylated through DNA replication-dependent passive dilution, the paternal genome undergoes two steps of demethylation (Guo et al., 2014; Shen et al., 2014). First, TET3 leads to the active conversion of 5mC into 5hmC (or 5fC and 5caC) (Gu et al., 2011). Second, these oxidized variants get passively diluted as active modification in combination with active removal might not occur due to vanishingly low expression levels of TDG in zygotes (Tang et al., 2011). The absence or rather extremely low incidence of TET3-mediated active oxidation in the maternal genome (Peat et al., 2014) is due to the protective property of DPPA3 in fertilized zygotes, a maternal effect-protein specifically expressed in pre-implantation embryos and germ cells (Sato et al., 2002). By binding H3K9me2, DPPA3 is thought

to prevent these sites from TET3-mediated demethylation (Nakamura et al., 2007, 2012). Finally, at embryonic day E3.5 in mice when the blastocyst is formed, DNA methylation reaches its lowest point. The genome at this stage is globally demethylated, albeit leaving a few genes like imprints and IAP retrotransposons specifically methylated (Smith et al., 2012) (Figure 8a).

But how do these genic regions escape such global DNA demethylation events during pre-implantation? Imprinted genes, defined to be monoallelically expressed in a parent-of-origin dependent manner, are mainly arranged in chromosomal clusters containing a single germline differentially methylated region (gDMR) that controls transcriptional activity of these genes (Voon and Gibbons, 2016). Common to all gDMRs of mice is a consensus sequence that enables, if methylated, the binding of ZFP57, a KRAB-zinc finger protein recruiting KRAB-associated protein-1 (KAP1) (Quenneville et al., 2011). KAP1 in turn acts as a scaffold for the recruitment of heterochromatin proteins, like DNMTs or heterochromatin protein 1 (HP1) (Messerschmidt science 2012). Thus, the combination of ZFP57 and KAP1 is of great relevance in maintaining DNA methylation at imprinted DMRs during early development (Messerschmidt et al., 2012). Notably, genomic imprinting does not rely exclusively on DNA methylation, but histone modifications do play equally important roles at these gene clusters (Singh et al., 2011).

With the implantation of the blastocyst in turn, DNMT3A and DNMT3B start to methylate the embryonic DNA again and proceed until methylation levels in the epiblast resemble levels of differentiated somatic tissue with about 80% CpG methylation (Wang et al., 2014). Recent studies further propose that the demethylation processes in early embryonic development are not linear and indeed, some de novo methylation has been observed in co-occurrence with the demethylation waves stated above (Amouroux et al., 2016).

1.5.2 Primordial germ cells

Post implantation at embryonic day E6.5 in mice, a specific subset of stem cells within the epiblast separate to form the primary cells of the germline lineage, namely primordial germ cells (PGCs) (Sasaki and Matsui, 2008). Similar to the paternal genome in zygotes, also PGCs undergo two stages of demethylation (Wu and Zhang, 2017). In a first round, DNA gets globally demethylated by passive dilution due to a repression of UHRF1, which is furthermore accompanied by low levels of DNMT3A and B (Kagiyada et al., 2013). In a second step, TET1 and TET2 mediate active demethylation by locus-specific means to selectively demethylate meiotic and imprinted genes (Yamaguchi et al., 2013), which is confirmed by the generation of 5hmC in PGCs (Hackett et al., 2013) (Figure 9b). The removal of former imprinted regions is indispensable for the specification of PGCs, i.e. re-establishing imprints based on the respective gender (Hayashi et al., 2007). In male germ cells, the re-establishment of methylation marks takes place during fetal gonocyte development and reaches almost 80-90% of CpG methylation with a patterning as in somatic cells. In contrast, re-methylation in female germ cells does not occur before oocytes mature in adulthood. Interestingly, when reprogramming in oocytes is finished the level of CpG methylation does not exceed 50% and methylation is mainly found in gene bodies (Sasaki and Matsui, 2008). The discrepancy to the male germline methylation level is attributed to DPPA3 that sequesters UHRF1 to the cytoplasm in oocytes. This consequently leads to the retention of DNMT1 and the observed “hypomethylation” of female germ cells (Li et al., 2018b). Surprisingly, methylation levels in Dppa3-mutant oocytes were twice as high compared to wildtype due to the nuclear localization of DNMT1, indicating an unexpected de novo methylation activity of DNMT1 (Li et al., 2018b). Although the remethylation dynamics differ between male and female germ cells, both require DNMT3A and DNMT3L for the establishment of their sex-specific imprints (Bourc’his et al., 2001). Besides that,

a newly discovered methyltransferase, DNMT3C, in rodent genomes was shown to be essential for male PGC development (Barau et al., 2016).

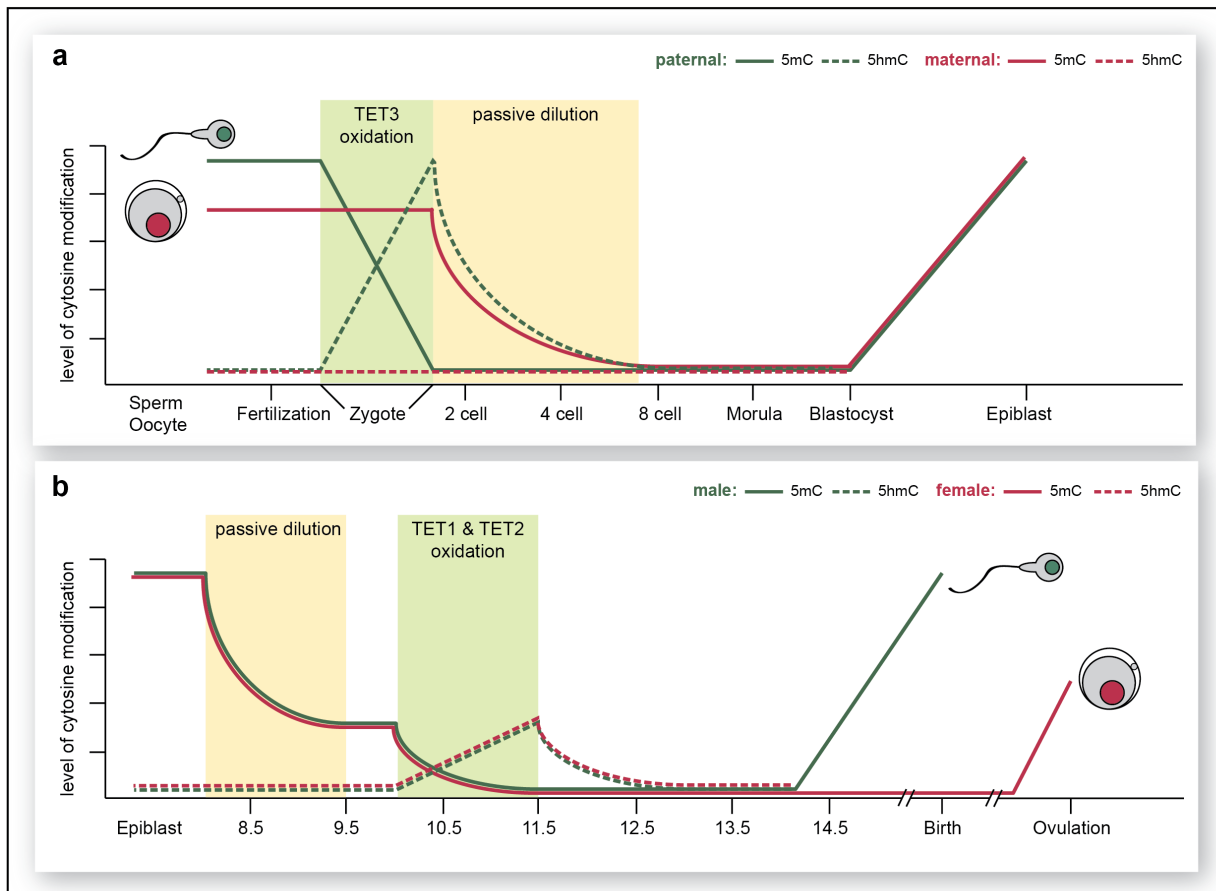


Figure 9: DNA (de)methylation during pre-implantation development and PGC specification

(a) 5mC and 5hmC dynamics in pre-implantation embryos. The paternal genome is demethylated through a combination of TET3-mediated oxidation of 5mC and replication-dependent passive dilution of the oxidized products (i.e. 5hmC, but also 5fC and 5caC). The maternal genome is only demethylated through passive dilution. At the blastocyst stage, methylation of both genomes is re-established. (b) 5mC and 5hmC dynamics during primordial germ cell (PGC) specification. The majority of DNA methylation patterns in PGCs is removed upon passive dilution, whereas remaining 5mC is subsequently removed through the oxidative activity of TET1 and TET2. Re-methylation of germ cells occurs later in development. X-axis indicates days postfertilization, adapted from Kohli and Zhang 2013.

1.5.3 Transcriptional control & transposon silencing

Historically, DNA methylation has early been linked to silencing of gene expression (Razin and Riggs, 1980). Since not all genes are active at all times, DNA methylation was initially considered to be “the” epigenetic tool turning respective genes off (McGhee and Ginder, 1979). Re-activation of the inactivated X chromosome and imprinted genes in mice lacking DNMTs further confirmed this observation (Suzuki and Bird, 2008). Methylated gene promoters lead to decreased binding of almost $\frac{1}{4}$ of all human transcription factors (Yin et al., 2017) and DNMTs are known to interact with proteins of the heterochromatin machinery, like H3K9 methyltransferases (Esteve et al., 2006). 5mC-binding proteins (MBD family) were furthermore shown to interact with nucleosome remodelers and histone deacetylases (Baubec et al., 2013; Nan et al., 1998). Taken together, these findings point towards a DNA methylation-based mechanism of gene silencing.

However, the evolution of DNA methylation was probably not driven by the necessity to control gene transcription but rather by means of defending the genome against transposable elements (Greenberg and Bourc'his, 2019). Transposable elements are DNA sequences that are able to jump out of the genome and integrate somewhere else, thereby changing their position (Bourque et al., 2018). They are classified into two major groups according to the “copy and paste” or “cut and paste” mechanism of transposition: retrotransposons and DNA transposons, respectively (Kapitonov and Jurka, 2008). Retrotransposons account for about 50% of our genome and their potential to produce insertions and rearrangements bears a high risk for mutations and genomic instability. Hence, uncontrolled transcriptional activity of these sequences must be avoided at all cost (Payer and Burns, 2019). In mouse embryos lacking DNMT1 an evolutionary young group of retrotransposons, intracisternal A particle (IAP) retrotransposons, are massively upregulated indicating that DNA methylation plays an integral part in silencing transposons (Walsh et al., 1998). Accordingly, also transposons acquire resistance to demethylation events in early embryonic development due to ZFP57-mediated recruitment of KAP1 and the maintenance of methylation, similar to imprinted genes (Rowe et al., 2010). Compared to the erasure of methylation marks in early embryos, the erasure in PGCs is even more extensive including the demethylation of imprinted genes and residual CpG methylation of just 6-8% (Wang et al., 2014). Intriguingly, the residual methylation in PGCs is mainly found at young and potentially deleterious retrotransposons (Guibert et al., 2012). In male germ cells of rodents, the selective repression of these evolutionarily young transposons seems to even have yielded a new DNA methyltransferase, DNMT3C, that is supposed to perform the repression in concert with DNMT3L and a piwi-interacting RNA (Barau et al., 2016). Notably, a growing body of evidence started to challenge the conventional view of DNA methylation as an exclusive tool of gene silencing. Firstly, methylated DNA motifs were recently reported to be specifically read by transcription activators, like OCT4, which potentially facilitates the activation of genes in otherwise inert chromatin regions (Yin et al., 2017). Secondly, DNA methylation is enriched in gene bodies and this has been linked to enhanced transcriptional activity (Lister et al., 2009). Functionally, it has been proposed that these methylation marks either repress intragenic cryptic promoters or help in elongating transcription (Greenberg and Bourc'his, 2019). Moreover, UHRF1 was lately shown to regulate active transcriptional marks at bivalent domains in ESCs despite its widely known function in heterochromatin formation together with DNMT1, the G9a methyltransferase and HDACs (Kim et al., 2018).

1.5.4 Formation of cancer and other human diseases

Epigenetic regulatory networks including the methylation and demethylation of DNA is indispensable for mammalian development and the ability of genomes to adapt to environmental factors (Flores et al., 2013; Greenberg and Bourc'his, 2019). Conrad Waddington's concept of the epigenetic landscape for instance figuratively illustrates how cell fate decisions, i.e. the terminal differentiation of cells, are based on epigenetic means (Waddington, 1957). Abrogating epigenetic regulatory mechanisms therefore alters these cell fates and represents a chance for abnormal cell growth (Kanwal and Gupta, 2012). In fact, 50% of human cancers harbor mutations that affect epigenetic regulation and many studies revealed abnormal epigenetic activities during tumorigenesis and in other diseases. Aberrant methylation patterns are observed in various cancer types, mainly defined by global hypomethylation and small islands of hypermethylation especially at promoters of tumor suppressor genes (Jones, 2012).

1.5.4.1 Dysfunction of DNMTs

Germline mutations in the *Dnmt3a* gene are linked to growth disorders (Heyn et al., 2019; Tatton-Brown et al., 2014). Interestingly, the growth phenotype depends on the type of mutation with gain-of-function mutations in the PWWP-domain leading to reduced body size and microcephaly (Heyn et al., 2019). Haploinsufficiency mutations on the contrary induce macrocephalic overgrowth (Tatton-Brown et al., 2014). The occurrence of somatic DNMT3A mutations plays an active role in hematological disorders and is found in 15-35% of acute myeloid leukemia (AML) (Greenberg and Bourc'his, 2019). Recurrent mutations that reduce the enzymatic activity of DNMT3A affect amino acid R882 and have been shown to significantly reduce DNA methylation at a subset of genomic locations in AML patients (Ley et al., 2010).

Germline mutations in the *Dnmt3b* gene are associated with the immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome (Xu et al., 1999). Albeit ICF syndromes are normally defined by hypomethylated pericentromeric satellite repeats (Jeanpierre et al., 1993), DNMT3B mutations do not seem to trigger these methylation defects. Hypomethylation due to reduced DNMT3B activities rather targets the promoters of germline and Xi-linked genes (Jin et al., 2008). This is in line with studies in mice where DNMT3B is reported to regulate these promoters during early embryogenesis (Auclair et al., 2014; Gendrel et al., 2012). Remarkably, the frequency of genetic mutations in the *Dnmt3b* gene in human cancer is extremely low and DNMT3B influences tumor progression predominantly by altered expression levels (Gagliardi et al., 2018).

The incidence of *Dnmt1* mutations was initially described in connection with neurodegenerative disorders like hereditary sensory and autonomic neuropathy type IE (HSANIE) and autosomal dominant cerebellar ataxia deafness and narcolepsy (ADCA-DN) (Klein et al., 2011; Winkelmann et al., 2012). Both diseases involve sensory impairment and dementia and patients exhibit DNA hypomethylation (Baets et al., 2015). Strikingly, all DNMT1 mutations identified so far accumulate within the RFTS domain of the protein and a study in mESCs revealed that these mutations abrogate the binding of DNMT1 to UHRF1 and impair the differentiation of ESCs into the neuronal lineage (Smets et al., 2017). Deregulation of or mutations in DNMT1 have also been attributed to malignant transformation, e.g. in breast and colon cancer (Agoston et al., 2005; Kanai et al., 2003). In a recent publication, deletion of the RFTS domain of DNMT1 in healthy epithelial cells resulted in inverse changes of DNA methylation, namely focal hypermethylation and global hypomethylation (Wu et al., 2014a). This reflects the situation in cancer cells and the findings of Wu et al. provide a coherent and DNMT1-based mechanism that could explain the opposing methylation levels of cancer cells (Bashtrykov and Jeltsch, 2015).

Besides mutations in the writer enzymes of the methylation machinery, also mutant 5mC reader proteins are found in some diseases, e.g. mutant MECP2 in Rett syndrome (Amir et al., 1999). Normally, MECP2 is highly expressed in neuronal tissue and facilitates silencing of methylated genes by intensively interacting with various repressive complexes (Ebert et al., 2013; Jones et al., 1998; Lyst et al., 2013; Nan et al., 1997). Mutations within MECP2 abolish its recruiting function of repressive mediators and entails derepression of repetitive elements as observed in the neurological disorder, Rett syndrome (Muotri et al., 2010; Skene et al., 2010).

1.5.4.2 Dysfunction of UHRF1

Surprisingly, there is no human disease known that implicates any mutation within the *Uhrf1* gene (Bronner et al., 2007). However, a growing number of publications indicate that overexpression of UHRF1 plays a role in various cancer types such as colorectal (Wang et al., 2012), prostate (Jazirehi et al., 2012) and lung cancer (Unoki et al., 2010). Deregulation of UHRF1 was reported to

transcriptionally inhibit a selection of tumor suppressor genes (Xue et al., 2019), but the underlying molecular mechanism remains elusive. In healthy individuals, UHRF1 expression varies between different tissues, but the expression level is mainly positively correlated to the proliferative capability of the cells (Mousli et al., 2003). Hence, mRNA of UHRF1 is not detected in differentiated tissues. Upregulating UHRF1 could therefore help cancer cells to keep their proliferative potential up and evade terminal differentiation (Bronner et al., 2007). For this reason, UHRF1 is considered a universal biomarker for many cancer types and bears the potential to serve as an anticancer drug target (Sidhu and Capalash, 2017).

1.5.4.3 Dysfunction of TET and IDH enzymes

Besides deregulating the methylation machinery in form of mutations and altered expression of DNA methyltransferases or UHRF1, also DNA demethylation processes can be perturbed in favor of a disease state (Pfister and Ashworth, 2017). Somatic mutations in the *Tet2* locus are among the most frequent causes of hematopoietic malignancies, including AML (Weissmann et al., 2012), chronic myelomonocytic leukemia (CMML) (Kosmider et al., 2009) and myelodysplastic syndrome (MDS) (Langemeijer et al., 2009). The spectrum of mutations ranges from single amino acid substitutions to frame shifts and the generation of truncated TET2 proteins due to stop codon insertions (Feng et al., 2019). Haploinsufficiency of TET2 is considered an early driver of leukemogenesis, although heterozygous loss of TET2 has rarely been observed, too (Feng et al., 2019). Hematopoietic cells with a loss of TET2 were reported to exhibit a hypermethylation phenotype mainly at gene promoters (Figuroa et al., 2010) whereas a more recent study stated that increased methylation levels are predominantly captured at active enhancers (Rasmussen et al., 2015). Concomitantly, 5hmC levels in TET2 mutant cells are significantly reduced compared to healthy wildtype cells (Ko et al., 2010; Rasmussen et al., 2015). Genetic inactivation of TET2 in the hematopoietic system of mice leads to increased proliferation of hematopoietic stem cells (HSCs) and impairs their terminal differentiation potential (Li et al., 2011; Moran-Crusio et al., 2011). However, even though the genetic ablation of TET2 results in a disease-like phenotype, the penetrance of the malignancy in mouse models remains low, indicating that cooperating mutations are required for a full-blown leukemic transformation (Rasmussen and Helin, 2016). Besides the effect on DNA methylation and the consequential deregulation of gene expression, loss of TET2 is also thought to impact other continuative mechanisms like DNA damage repair or immune regulation of hematopoietic cells (Feng et al., 2019; Inoue et al., 2016). It is also reasonable that the malignant function of TET2 is not a consequence of impaired catalytic activity but stems from a non-catalytic role of TET2, e.g. acting as a scaffold protein to recruit other proteins or chromatin modifiers (Feng et al., 2019).

Interestingly, mutations in proteins that regulate TETs, e.g. IDH1, IDH2 and WT1, are also found in hematological disorders and occur in a mutually exclusive manner with TET mutations (Figuroa et al., 2010; Gaidzik et al., 2012; Rampal et al., 2014). Albeit there are three IDH paralogs in mammals that perform the oxidative decarboxylation of isocitrate to alpha-ketoglutarate (IDH1, IDH2, and IDH3), IDH3 employs a differing catalytic mechanism, catalyzes an irreversible reaction and no somatic mutations of the gene have been found so far (Dang and Su, 2017). On the contrary, somatic mutations in IDH1 and IDH2 occur frequently in a wide spectrum of cancers, most prominently in AML (20%) and secondary glioblastoma (80%) (Montalban-Bravo and DiNardo, 2018; Yan et al., 2009). Intriguingly, nearly all mutations map to specific arginine residues responsible for isocitrate binding within the catalytic core (R132 in IDH1, R140 and R172 in IDH2) that confers a neomorphic gain-of-function to the enzymes (Dang and Su, 2017). The reduced affinity to

isocitrate is accompanied by increased NADPH binding and favors parts of the reverse reaction, namely the reduction of α -KG without an additional carboxylation step, yielding (R)-2-hydroxyglutarate ((R)-2HG) instead of isocitrate (Dang et al., 2009). Although altered metabolic states is a hallmark of cancer, which ensures sufficient supply of energy for unrestrained proliferation, mutant IDH enzymes gained strong interest due to its direct link to the epigenetic regulatory network (Ward and Thompson, 2012). In fact, the structural similarity of (R)-2HG and α -KG allows (R)-2HG to perfectly bind to the catalytic center of α -KG-dependent dioxygenases, thereby inhibiting their enzymatic reactions (Chowdhury et al., 2011; Xu et al., 2011) (Figure 10). As TET proteins and jumonji domain containing histone demethylases are members of this dioxygenase family, numerous studies reported altered DNA and histone methylation in mutant IDH cancer cells (Figuroa et al., 2010; Lu et al., 2012; Turcan et al., 2012). Overexpression of cancer-derived mutant IDH enzymes *in vitro* in various cell types like erythroleukemic cells, fibroblasts, astrocytes and murine hematopoietic progenitor cells stimulated their proliferation and simultaneously blocked differentiation (Figuroa et al., 2010; Koivunen et al., 2012; Losman et al., 2013; Lu et al., 2012). Notably, the effects of mutant IDH expression *in vitro* can be recapitulated by treating cells with cell-permeable (R)-2HG in concentrations intracellularly detected in tumors (Chaturvedi et al., 2016; Losman et al., 2013; Lu et al., 2012). On these grounds, (R)-2HG has been defined as a potent oncometabolite. Mutant IDH enzymes however are not enough to transform primary cells *in vivo* as demonstrated by Sasaki et al. in both, brain- and hematopoietic-specific IDH1 R132H knock-in mice (Sasaki et al., 2012a, 2012b). Comparable to TET2 mutations, other cooperating genetic alterations, e.g. HoxA9 overexpression or FLT3 mutations in case of AML, are required for the formation of tumors (Kats et al., 2014).

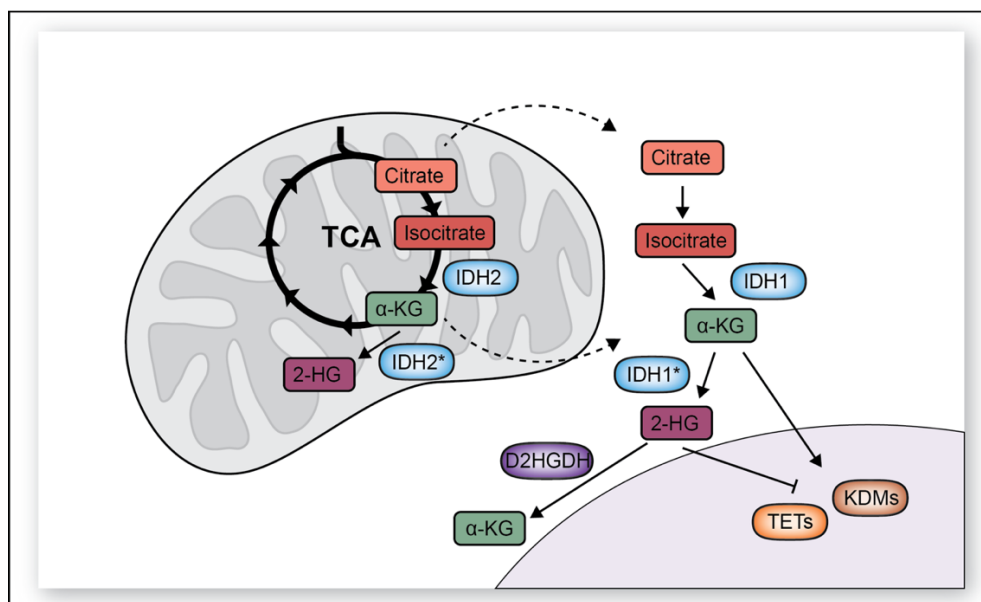


Figure 10: Metabolic function of wildtype and mutant IDH enzymes

Cytosolic IDH1 and mitochondrial IDH2 produce alpha-ketoglutarate (α -KG) from isocitrate within the TCA cycle. Mutant IDH enzymes (marked with *) gain neomorphic activities and generate hydroxyglutarate (2-HG) instead. This metabolite can be reconverted into α -KG through the activity of D-2-HG dehydrogenases (D2HGDHs). α -KG and 2-HG in turn can modulate the activity of α -KG-dependent enzymes like TET proteins and lysine-specific demethylases (KDMs), adapted from Trummer et al. (unpublished).

Since the mammalian genome codes for about 70-80 different kinds of α -KG-dependent dioxygenases (i.a. prolyl hydroxylases and AlkB family enzymes) (Loenarz and Schofield, 2011), the impact of mutant IDH enzymes goes far beyond the deregulation of epigenetic reactions. Other cellular alterations include impaired collagen maturation in the basement membrane of the brain by IDHmut-mediated inhibition of collagen prolyl hydroxylases (Sasaki et al., 2012a) and perturbed function of cytochrome-c oxidase (COX) in the mitochondrial respiratory chain (Chan et al., 2015). Furthermore, (R)-2HG hinders the repair of nucleic acid alkylation damage by allosterically inhibiting the activity of ALKBH2 and ALKBH3 (Wang et al., 2015b). The relationship between IDHmut-produced (R)-2HG and the functional regulation of HIF α , a master transcription factor under hypoxic conditions, remains controversial. Different studies reported both, accumulation and degradation of HIF α through the inhibition or activation of prolyl hydroxylases, respectively (Koivunen et al., 2012; Sasaki et al., 2012a; Xu et al., 2011). In the past, the majority of studies focused on the cellular effects of (R)-2HG, however, mutated *Idh* alleles influence the cellular state also independently of the oncometabolite (Dang and Su, 2017). For instance, the forward reactions of IDH1 and IDH2 represent an important source of NADPH production and the wildtype enzymes are therefore crucial for the redox state of cells. Besides reduced NADPH levels, cells with mutant IDH also exhibit lower glutamine and glutamate levels and are thought to generally slow down the TCA cycle (Dang and Su, 2017).




Surprisingly, 2-HG is not solely produced by mutant IDH enzymes, but also results as an unwanted by-product from naturally occurring metabolic reactions (Rzem et al., 2007; Struys et al., 2005a). Cellular levels of these metabolites however are normally kept to a minimum based on the activity of 2-hydroxyglutarate dehydrogenases (2HGDHs), which recycle 2HG back to α -KG (Steenweg et al., 2010; Struys et al., 2005b) (Figure 10). The extremely elevated (R)-2HG levels detected in tumors may arise because mutated IDH enzymes catalytically overwhelm the capacity of these 2HGDHs (Losman and Kaelin, 2013) or because 2HGDHs are not sufficiently expressed in the respective tumor. Accordingly, humans lack effective defense mechanisms against (R)-2HG and much effort has gone into the development of mutant IDH inhibitors. A large number of small molecules has been tested to date, which already led to the market release of two FDA approved drugs for the treatment of AML (Montalban-Bravo and DiNardo, 2018). However, whereas hematological malignancies are typically very sensitive to IDH inhibitors, patients with gliomas lack obvious responses to the treatment and further work is required to successfully medicate IDH mutant tumors in the near future (Tommasini-Ghelfi et al., 2019).

2 RESULTS


2.1 Systematic analysis of the binding behavior of UHRF1 towards different methyl- and carboxylcytosine modification patterns at CpG dyads


RESEARCH ARTICLE

Systematic analysis of the binding behaviour of UHRF1 towards different methyl- and carboxylcytosine modification patterns at CpG dyads

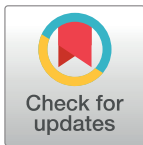
Markus Schneider¹ , Carina Trummer² , Andreas Stengl², Peng Zhang^{2,3}, Aleksandra Szwagierczak², M. Cristina Cardoso³, Heinrich Leonhardt², Christina Bauer² , Iris Antes¹ *

1 Center for Integrated Protein Science Munich at the TUM School of Life Sciences, Technische Universität München, Freising, Germany, **2** Center for Integrated Protein Science Munich at the Department of Biology II, Ludwig Maximilians University Munich, Planegg-Martinsried, Germany, **3** Cell Biology and Epigenetics at the Department of Biology, Technische Universität Darmstadt, Darmstadt, Germany

 These authors contributed equally to this work.

 Current address: Department of Biomedicine, University of Basel, Basel, Switzerland

* antes@tum.de



OPEN ACCESS

Citation: Schneider M, Trummer C, Stengl A, Zhang P, Szwagierczak A, Cardoso MC, et al. (2020) Systematic analysis of the binding behaviour of UHRF1 towards different methyl- and carboxylcytosine modification patterns at CpG dyads. *PLoS ONE* 15(2): e0229144. <https://doi.org/10.1371/journal.pone.0229144>

Editor: Albert Jeltsch, Universität Stuttgart, GERMANY

Received: September 4, 2019

Accepted: January 30, 2020

Published: February 21, 2020

Copyright: © 2020 Schneider et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by the Deutsche Forschungsgemeinschaft (SFB 1035/A10 and SFB749/C08 to IA, SFB 1064/A17 and SFB 1243/A01 to HL, CIPSM to IA and HL, CA198/10-1 to MCC). MS was supported by the TUM International Graduate School of Science and Engineering (IGSSE). CB and CT were supported

Abstract

The multi-domain protein UHRF1 is essential for DNA methylation maintenance and binds DNA via a base-flipping mechanism with a preference for hemi-methylated CpG sites. We investigated its binding to hemi- and symmetrically modified DNA containing either 5-methylcytosine (mC), 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), or 5-carboxylcytosine (caC). Our experimental results indicate that UHRF1 binds symmetrically carboxylated and hybrid methylated/carboxylated CpG dyads in addition to its previously reported substrates. Complementary molecular dynamics simulations provide a possible mechanistic explanation of how the protein could differentiate between modification patterns. First, we observe different local binding modes in the nucleotide binding pocket as well as the protein's NKR finger. Second, both DNA modification sites are coupled through key residues within the NKR finger, suggesting a communication pathway affecting protein-DNA binding for carboxylcytosine modifications. Our results suggest a possible additional function of the hemi-methylation reader UHRF1 through binding of carboxylated CpG sites. This opens the possibility of new biological roles of UHRF1 beyond DNA methylation maintenance and of oxidised methylcytosine derivatives in epigenetic regulation.

Introduction

UHRF1 (also referred to as Np95) is an essential protein for DNA methylation maintenance in mammals. It consists of 5 domains: A ubiquitin-like domain, a Tandem-Tudor domain, a PHD domain, a DNA-binding SRA domain, and a RING domain with E3 ubiquitin ligase activity ([Fig 1a](#)) [1–3]. UHRF1 was originally reported to preferentially bind to hemi-

by the International Max Planck Research School for Molecular Life Sciences (IMPRS-LS). AS was trained and supported by the graduate school GRK1721 of the Deutsche Forschungsgemeinschaft. PZ received support from the "Vereinigung von Freunden der Technischen Universität Darmstadt". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

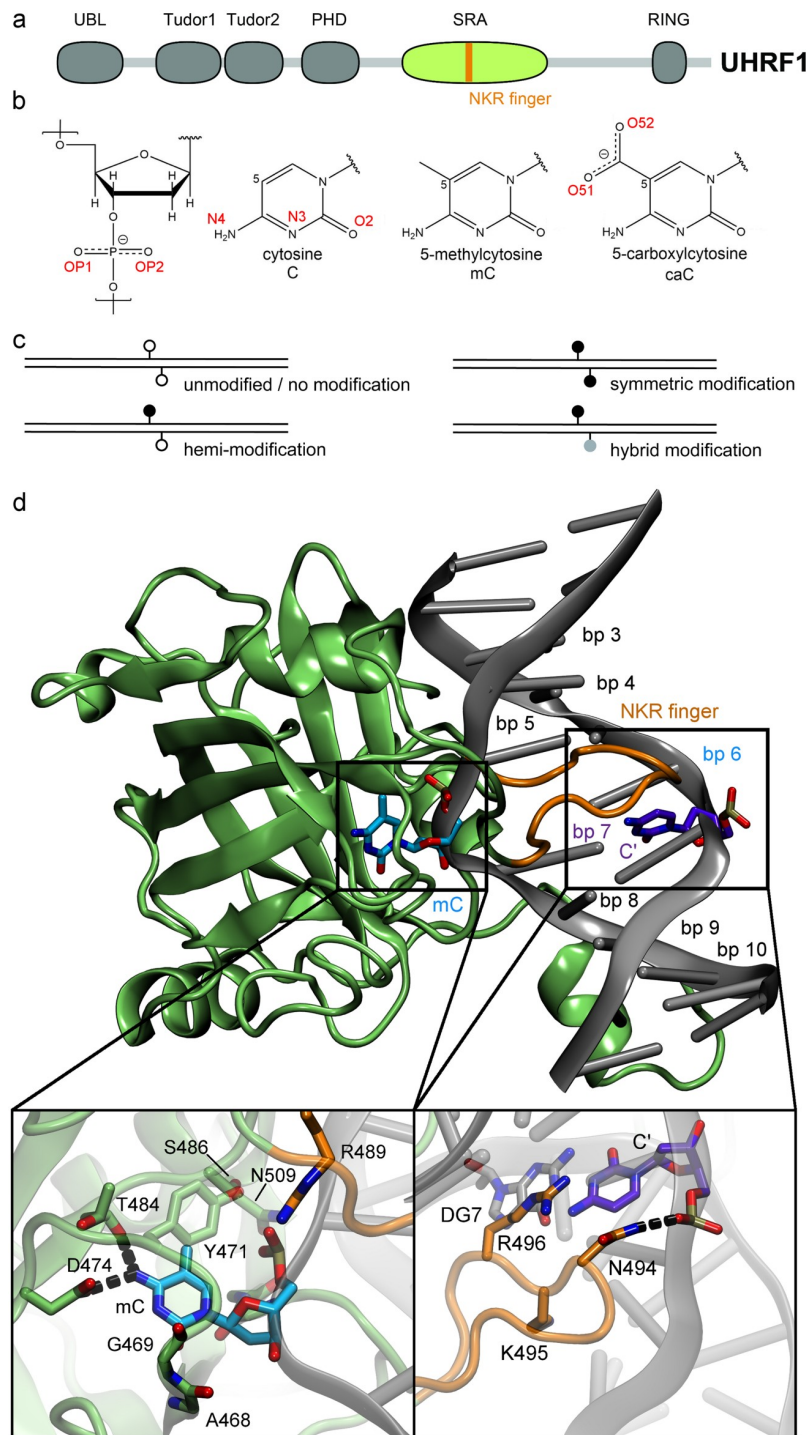


Fig 1. Structure of the UHRF1—DNA complex. (a) Schematic structure of UHRF1. The Tudor-like domains and the PHD-type zinc finger recognize the histone marks H3K9me2/3 and H3R2me0, respectively, while the SRA domain (in green, also referred to as YDG domain) is important for DNA binding. (b) Chemical structure and atom names of the modified DNA bases methylcytosine (mC) and carboxylcytosine (caC). (c) Schematic illustration of possible cytosine modification configurations on CpG dyads. (d) Representative molecular dynamics structure of the SRA domain of UHRF1 bound to hemi-methylated DNA. Insets show a magnification of the nucleotide binding pocket and NKR finger regions. DNA base pairs (bp) are numbered based on the strand binding the flipped-out base.

<https://doi.org/10.1371/journal.pone.0229144.g001>

methylated DNA, i.e. DNA harbouring 5-methylcytosine (mC) only on one strand. Upon binding of the methylated strand, UHRF1 recruits DNA methyltransferase 1 (DNMT1) for additional methylation of the second strand, yielding a symmetrically methylated CpG site [1–3]. This recruitment depends on specific histone ubiquitination, set by the RING domain of UHRF1 and recognized by a ubiquitin interaction motif of DNMT1 [4–6].

Besides mC, three other cytosine (C) modifications exist in mammalian cells, i.e. 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxylcytosine (caC) [7–9]. These variants are generated by the family of TET proteins through step-wise oxidation of mC and are discussed to be either intermediates in active DNA demethylation or independent epigenetic marks [10]. Their overall abundance *in vivo* is normally magnitudes lower than that of methylated sites [11], but the ratio increases under certain conditions. Higher hmC concentrations were observed in neuronal cells [12], while a study investigating breast and glioma tumour tissues found that a substantial portion of the samples exhibited increased caC levels [13]. Efforts to map mC, hmC, fC, and caC modifications in the genome showed that they accumulate at functionally distinct regions of transcription regulation [14–16]. One common conclusion of these studies was that methylation/demethylation of CpG sites is a highly dynamic and genome-wide process. In this light, low concentrations of some DNA modifications could represent a transient state in a high turnover process, while the accumulation at functionally diverse sites suggests that some variants might have a biological role beyond being demethylation intermediates. It has been demonstrated that several proteins recognize some oxidised variants with similar or even greater affinity than mC. The UHRF family member UHRF2, which features a highly similar domain architecture to UHRF1 [17, 18], is a reader with increased affinity for hmC in neuronal progenitor cells [19]. Other examples include SUVH5, which binds both mC and hmC with similar strength [20], while POL II, WT1 and TET3 specifically recognize caC [21–23]. It is currently unclear how frequent certain CpG modification patterns occur *in vivo*. DNA replication during S-phase will generally result in hemi-modified CpG sites. In case of mC, the subsequent restoration of the DNA modification to symmetry is well studied and described [24]. Nevertheless, the degree of persistent hemimethylation varies between cell types and genomic elements [25]. For hmC, fC, and caC, no maintenance pathways have been described so far. *In vitro*, TET proteins predominantly generate symmetric fC sites [26], whereas genomic mapping approaches suggest the existence of hmC and fC/caC in hemi-modified form [15, 27]. The occurrence of hybrid modifications with mC on one and an oxidised cytosine derivative on the other strand is also likely (Fig 1c).

Structural analysis revealed that the SRA domain of UHRF1 flips the methylated cytosine out of the DNA strand and envelopes it within its binding pocket. In addition, the protein binds to the DNA by inserting its thumb region into the minor groove and its NKR finger region into the major groove [2, 28, 29]. In a previous work, our groups showed by a combination of *in vitro* experiments and molecular dynamics (MD) simulations that UHRF1 binds hemi-modified hmC with similar affinity as hemi-mC [30]. Although subsequent studies revealed that UHRF1 binds hmC with lower affinity than mC, it still binds hmC with 1.3 to 3-fold higher affinity than unmodified C [19, 31, 32]. These results are in line with an unbiased mass spectrometry screen for epigenetic readers in embryonic stem cells, which demonstrated UHRF1 binds to all modified cytosines, but in particular to mC and hmC [19]. Experiments with UHRF1 and symmetrically modified mC sites, i.e. CpG sites in which both DNA strands feature methylcytosine, consistently show reduced binding affinity [1, 2, 28, 29]. This selectivity is commonly explained by a hydrogen bond between N494 at the tip of the NKR finger and the C' cytosine, i.e. the base that potentially carries the symmetric modification (Fig 1d) [29]. Throughout the manuscript we use a terminal apostrophe to mark bases on the distal DNA strand (e.g. C'). Bianchi et al. observed in a computational study that the presence of mC on

both strands sterically impairs binding of the NKR finger of UHRF1 to the major groove [33]. In contrast to mC and hmC, the structural effects of fC and caC variants on UHRF1-DNA binding are still not well elucidated. Investigations of several SRA domains by Rajakumara et al. suggest a reduced affinity of UHRF1 towards hemi-hmC, -fC and -caC containing DNA [20]. Crystal structures of POL II and TDG, which exhibit specific activity towards caC, show that the caC carboxyl group participates in specific hydrogen bond networks, which are crucial for binding key recognition residues in the protein [21, 34].

It was recently shown that UHRF1 allosterically regulates its activity and binding properties through intramolecular conformational changes [35–38]. The formation of these extensive inter-domain interactions illustrates an inherent flexibility of UHRF1 and allows the protein to adapt to different substrates. As we already observed solid binding of UHRF1 to hemi-hmC, we sought to systematically analyse the binding behaviour of UHRF1 towards CpG sites containing C, mC, hmC, fC, and caC either in a hemi-, hybrid or symmetrically modified state. The highest binding affinities are observed for hemi-mC, symmetric caC, and the caC-mC' hybrid. To understand the differences in recognition of these modifications, we performed molecular dynamics simulations of mC- and caC-modified DNA in complex with the SRA domain of UHRF1 (see Fig 1d).

Materials & methods

Electrophoretic mobility shift assays (EMSAs)

Expression constructs for GFP-mUHRF1 and mUHRF2-GFP have been described previously [18, 39]. In general, protein purification and EMSAs were performed as reported in Spruijt et al. [19]. Briefly, a 2-fold serial dilution of protein (300 nM to 4.69 nM) in binding buffer (including 100 ng/μl BSA final concentration) was incubated with a 1:1 mixture of two fluorescently labelled 42 bp oligonucleotides (Eurofins Genomics) at a stable concentration of 250 nM each. After 30 min of incubation on ice, reactions were run over a 6% native PAGE in 0.5x TBE buffer (45 mM Tris-borate, 1 mM EDTA). ATTO647N-labelled DNA (“C⁶⁴⁷”) served as internal control and reference whereas ATTO550-labelled DNA carried one of the following cytosine variants at the central CG site: canonical C, mC, hmC, fC, or caC (“xC⁵⁵⁰”). Fluorescent signal was detected with a Typhoon Trio+ scanner (GE Healthcare Life Sciences). Signal of bound and unbound fractions were quantified with ImageJ by plotting the mean grey values per lane and measuring the area under the selected peaks. Before quantitation, gel pictures were assigned random names to blind the experimenter during analysis. Box plots show $\frac{\text{ATTO550 bound fraction}}{\text{ATTO647 bound fraction}} \times \frac{\text{ATTO647 total signal}}{\text{ATTO 550 total signal}}$ with the C⁵⁵⁰/C⁶⁴⁷ experiment as control. All raw gel image scans with annotations are provided as S1 Fig.

Microscale Thermophoresis (MST)

For MST, the SRA domain of mouse UHRF1 (residues 419–628) was cloned into a hexahistidine-tagged construct and protein was expressed in *Escherichia coli* BL21(DE3)-Gold cells (Stratagene). The purified SRA domain was labelled with a NT-647 dye using the Monolith NT[™] His-Tag Labelling Kit RED-tris-NTA (NanoTemper Technologies) according to the manufacturer's instructions and 50 nM of the labelled protein was incubated for 20 min at room temperature with increasing concentrations of the corresponding DNA oligonucleotide (C-C', mC-C', caC-C', caC-caC', mC-caC') in PBS-T (0.05% Tween-20). The solutions were then aspirated into NT.115 Standard Treated Capillaries (NanoTemper Technologies) and placed into the Monolith NT.115 instrument (NanoTemper Technologies). Experiments were conducted with 60% LED power and 80% MST power. Obtained fluorescence signals were

normalized (F_{norm}) and the change in F_{norm} was plotted as a function of the concentration of the titrated binding partner using the MO. Affinity Analysis software version 2.3 (NanoTemper Technologies). For fluorescence normalization ($F_{\text{norm}} = F_{\text{Hot}}/F_{\text{Cold}}$), the manual analysis mode was selected and cursors were set as follows: $F_{\text{cold}} = -1$ to 0, $F_{\text{hot}} = 9$ to 10 (see [S2 Fig](#)). Data of four to five independent measurements were analysed and means were fitted to obtain the respective K_D values. More detailed information and additional experimental procedures can be found in [S1 Text](#).

Force field parameterization of modified cytosine bases

We generated parameters for the parmbsc1 force field [40] for both deoxy-5-methylcytosine (mC) and deoxy-5-carboxylcytosine (caC) using the mC structure and bonded parameters template from Lankas et al. [41], which was originally derived for parmbsc0 [42]. The atom type of the C3' atom was changed from CT to CE to adjust the template to parmbsc1. Fixed point atom charges were derived for both mC and caC following the procedure in ref. [43] using the R.E.D Dev webserver [44–48]. Atom types were assigned and final parameter files prepared using the programs antechamber and prepgen of the AmberTools17 package [49]. The final parameter files are provided in [S1 File](#).

Molecular dynamics simulations

Molecular dynamics simulations were performed with the Amber16/AmberTools17 software suite [49] using the Amber14SB force field for protein and parmbsc1 for nucleic acid parameters [40, 50]. All systems were based on the crystal structure of a mouse UHRF1 SRA domain bound to DNA featuring a single mC (PDB-ID: 3FDE). The same structure had been used in our previous work analysing the binding of 5-hydroxymethylcytosine [30] and featured the best resolution (1.41 Å) of published UHRF1 structures at the time of this study. Cytosine modifications were modelled and topologies prepared using leap (AmberTools). Each system was solvated in a box of TIP3P water [51] with a minimum face distance of 15 Å and 150 mM NaCl. A direct space cutoff of 12 Å was used for nonbonded potentials and PME summation was applied for electrostatic interactions. Energy minimization was performed until convergence to $0.01 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-1}$ using the XMIN minimizer. Then, the volume of the solvent box was modified such that the density increased in $0.02 \text{ kg} \cdot \text{m}^{-3}$ steps and energy minimization was repeated for each step until a target density of $1.00 \text{ kg} \cdot \text{m}^{-3}$ was reached. For all molecular dynamics simulations hereafter, a time step of 1 fs and SHAKE [52] for bonds connected to hydrogens were used. The system was gradually heated from 0 to 300 K over 1.7 ns, applying a variation of the step-wise heatup protocol established within our group [53]. Within these steps, restraints of $2.39 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ were applied to all heavy atoms until 20 K and on protein/DNA backbone atoms until 200 K. For heatup, a Langevin thermostat was used with a collision frequency of 4 ps^{-1} , and for the last 0.5 ns a Berendsen barostat was employed with a relaxation time of 2 ps. During the following simulations at 300 K, a slow coupling Berendsen thermostat with a coupling time of 10 ps was used in combination with a Berendsen barostat and a respective relaxation time of 5 ps. Backbone phosphates and oxygens of terminal DNA residues were harmonically restrained with a constant of $2.39 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ while resetting target coordinates in 500 ps intervals. For all replicas, different initial velocities and random seeds for the Langevin thermostat were generated at the beginning of each step of the heatup protocol (i.e. for each temperature simulated). Each replicon was simulated for 200 ns, yielding a total simulation time of 1 μs per system (5 replicas). In two out of thirty simulations (caC-caC'_r2 and mC-caC'_r2), the DNA structure diverged notably from the others (RMSD > 4 Å; see [S3](#) and [S4 Figs](#)). In the case of caC-caC'_r2, the distortion correlates with an interaction

between the protein's free C-terminal helix and the DNA strand, bending it out of position, which is clearly an artefact due to the use of the isolated SRA domain. Therefore, and as it is in general difficult to determine whether such diverging trajectories show a rare but physically relevant conformational change or a simulation artefact, we excluded these two replicas from our analysis. The remaining simulations showed stable RMSD curves after about 20 ns. To allow for proper equilibration and to minimize any bias towards the initial structure, we extracted only the last 100 ns of each trajectory and afterwards merged the trajectories of all five replicas into a single system-specific trajectory that was used for all computational analyses.

Trajectory post-processing was performed with CPPTRAJ [54] version 17.00 unless otherwise indicated. Salt bridges were calculated using the "nativecontacts" command and a cutoff of 5 Å, saving both native and non-native time series and selecting interactions with opposite formal charges involving Arg, Lys, Glu, Asp and nucleotide residues. Hydrogen bonds were extracted using the "hbond" command, a cutoff distance of 4 Å and an angle cutoff of 120°. CPPTRAJ outputs were merged and converted into networks using our analysis tools AIFGen and CONAN (manuscript in preparation). Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) calculations were performed for non-hydrogen atoms using the CPPTRAJ "rmsd" and "atomicfluct" commands after aligning each simulation frame to the protein's C α atoms without the terminal regions (residues 432 to 586). For RMSD, the reference frame was the simulation's initial structure, while for RMSF the protein was aligned to its simulation average. DNA major and minor groove widths were calculated using the method of El Hassan and Calladine [55] as implemented in the "nastruct" command in CPPTRAJ (version 18.01). Figures of protein and DNA structures were prepared using VMD 1.9.3 [56]. Plots and supporting calculations (e.g. gaussian kernel estimates) were generated with matplotlib 2.0.0 [57].

Results

Experimental investigation of the binding behaviour of UHRF1 towards different cytosine variants

For systematic analysis of the binding specificities of UHRF1 towards the five known cytosine variants, we performed EMSA experiments with full-length UHRF1 in complex with 42 bp oligonucleotides harbouring C, mC, hmC, fC, or caC at a central CpG site (Fig 2a). To correct for general DNA binding affinity, two DNA fragments were used in direct competition in each EMSA experiment: A 647-labeled unmodified oligonucleotide and a 550-labeled oligonucleotide carrying the modification of interest in either hemi-modified (xC-C') or symmetric (xC-xC') state. 647-labeled unmodified DNA is used as internal control and reference for quantification. This allows direct comparison of UHRF1 binding affinity to all modifications without the need for pair-wise competition assays. Generally, EMSAs showed binding of UHRF1 to all studied DNA variants (example gel pictures are shown in Fig 2b). However, quantitation of the shifted fractions reveals a 1.5-fold preference for hemi-mC and a statistically significant 2-fold preference for symmetric caC (Fig 2c). All other modification variants, including hemi-caC, were bound with comparable strength to unmodified DNA. Similarly, we observed a 2-fold preference of UHRF2 for symmetric caC (S5 Fig).

Upon UHRF1 binding, the melting temperature of CpG-containing DNA is slightly reduced compared to its unbound state or a non-CpG-control, indicating a destabilization of the DNA duplex (S6a Fig). Complementary to our EMSA results, the SRA domain of UHRF1 substantially shifted the melting temperature of symmetrically carboxylated DNA to lower temperatures, whereas a weaker shift was observed for unmodified and hemi-methylated DNA (S6 Fig). To rule out that the thermal shift observed for symmetrically carboxylated DNA is due to different binding stoichiometries, we examined DNA-protein complex formation by

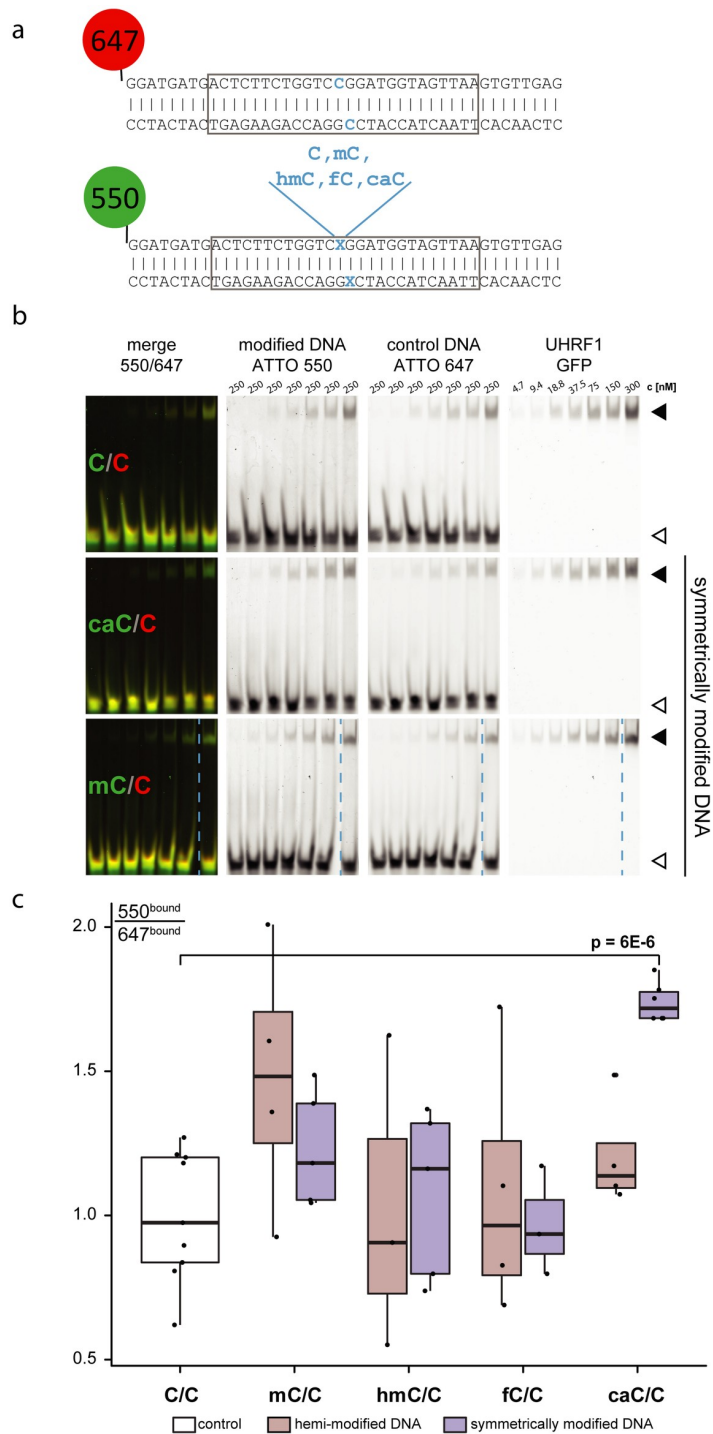


Fig 2. Binding of UHRF1 to differentially modified CpG sites. (a) DNA used in EMSA experiments. The 550-labelled DNA contains a central CG site harbouring different cytosine modifications: Unmodified C, mC, hmC, fC, or caC. The modification resides either on one strand (hemi-modification) or on both strands (symmetric modification). The 647-labelled oligonucleotide is always unmodified and serves as an internal control and reference. Grey boxes indicate sequences of the shorter DNA fragments used in Fig 3. (b) Representative images of EMSAs. Fluorescently labelled DNA oligonucleotides of 42 bp are incubated with GFP-UHRF1 at increasing protein concentrations. Black arrowheads indicate the DNA-protein complex (bound fraction); white arrowheads show free DNA. Dashed blue lines indicate empty gel lanes that have been removed for presentation purposes. (c) Quantitation of the bound fraction of symmetric and hemi-modified DNA incubated with wild type UHRF1, p value of two-tailed student's t-test.

<https://doi.org/10.1371/journal.pone.0229144.g002>

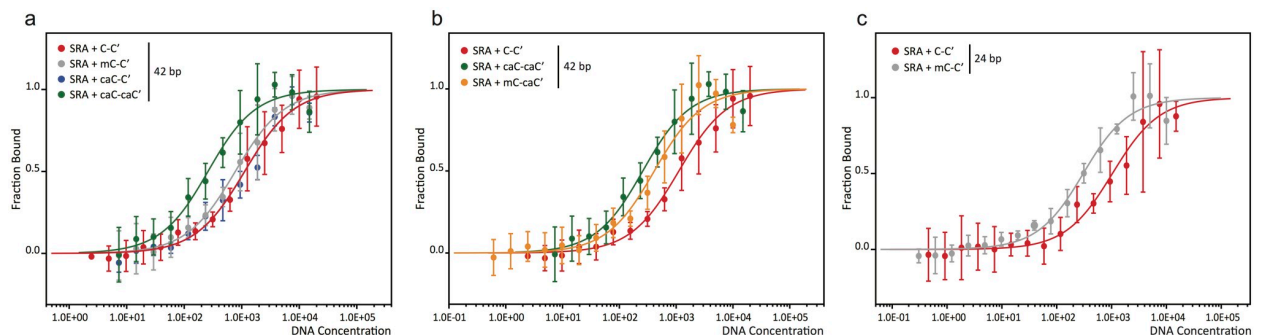


Fig 3. Microscale Thermophoresis experiments of UHRF1-SRA bound to DNA with modified CpG sites. (a,b) Dissociation constants of UHRF1 bound to a 42 bp DNA oligonucleotide: $1.10 \pm 0.15 \mu\text{M}$ for C-C', $0.75 \pm 0.11 \mu\text{M}$ for mC-C', $1.10 \pm 0.29 \mu\text{M}$ for caC-C', $0.23 \pm 0.05 \mu\text{M}$ for caC-caC', and $0.39 \pm 0.11 \mu\text{M}$ for mC-caC'. (c) Dissociation constants of UHRF1 bound to a 24 bp oligonucleotide; $1.01 \pm 0.20 \mu\text{M}$ for C-C' and $0.28 \pm 0.06 \mu\text{M}$ for mC-C'. Curves show the fitted average values of 4–5 independent experiments.

<https://doi.org/10.1371/journal.pone.0229144.g003>

size-exclusion chromatography. Binding of the SRA domain to the modified DNA oligonucleotides led to a comparable shift in retention time for all modifications tested (S7 Fig), indicating a uniform binding stoichiometry for UHRF1 independent of the DNA's modification state.

To better characterize the binding of UHRF1 to hemi-mC, hemi-caC and symmetric caC, we determined the respective dissociation constants (K_D) with Microscale Thermophoresis [58] (MST) experiments (Fig 3a). We observed slightly stronger binding of hemi-mC ($K_D = 0.75 \pm 0.11 \mu\text{M}$ vs. $1.10 \pm 0.15 \mu\text{M}$ for unmodified DNA) and considerably enhanced binding of symmetric caC ($K_D = 0.23 \pm 0.05 \mu\text{M}$). In agreement with the EMSA results, hemi-carboxylated DNA ($K_D = 1.10 \pm 0.29 \mu\text{M}$) is bound with similar affinity as unmodified DNA. Taken together, we performed three independent experimental assays, i.e. EMSAs, melting temperature analysis and MST, which consistently confirm a binding preference of UHRF1 towards symmetric caC.

Additionally, as the enzymatic reactions involved in generation of mC and caC modifications suggest the potential existence of hybrid mC-caC' sites, we determined the K_D of the SRA domain of UHRF1 and a mC-caC' oligonucleotide and observed binding comparable to symmetric caC ($K_D = 0.39 \pm 0.11 \mu\text{M}$ vs. $0.23 \pm 0.05 \mu\text{M}$). In summary, UHRF1 exhibits a binding preference for caC modifications opposite of mC or caC, but not C.

Since the difference in K_D between unmodified and hemi-methylated DNA was smaller than expected from the literature [1, 32, 36, 59, 60], we repeated the MST experiments with shorter DNA oligonucleotides of 24 bp to reduce the number of unspecific binding sites (Fig 3c). With this new setup we observed a 3.6-fold preference of the SRA domain of UHRF1 towards hemi-methylated CpG sites ($K_D = 0.28 \pm 0.06 \mu\text{M}$ for mC-C' vs. $1.01 \pm 0.20 \mu\text{M}$ for C-C'). This ratio is in very good agreement with data by Greiner et al. [60] and Zhou et al. [32] (Table 1), who reported a 3.5 or 3.4-fold smaller K_D for hemi-methylated CpGs for a 12 bp oligonucleotide, respectively, compared to unmodified DNA. Generally, caution is advised when published K_D values of UHRF1 and differentially modified DNA are compared, since applied methods, DNA substrates and protein constructs used vary greatly among studies, resulting in a broad range of K_D values from 1.8 nM to 9.23 μM (Table 1). Nonetheless, previous studies and our results not only demonstrate the sensitivity of UHRF1 to different types of cytosine modification, but also the dependency of measured binding affinities on modification density, i.e. the number of DNA modifications compared to unmodified DNA stretches.

Table 1. Published K_D values for UHRF1 and DNA with differentially modified CpG sites.

Citation	Method	Affinity	DNA substrate	protein construct
Bostick, M. et al., 2007, 10.1126/science.1147939	EMSA	$K_D(\text{mC-C}') = 1.8 \text{ nM}$ $K_D(\text{mC-mC}') = 12.1 \text{ nM}$	39mer, 13 modification sites	murine SRA
Fang, J., 2016, 10.1038/ncomms11197	Fluorescence Polarization	$K_D(\text{UHRF1}) = 0.35 \text{ }\mu\text{M}$ $K_D(\text{SRA}) = 9.23 \text{ }\mu\text{M}$ $K_D(\text{SRA+Spacer}^a) = 0.49 \text{ }\mu\text{M}$	12mer, 1 modification site	human UHRF1, different constructs with mC-C'
Greiner, V. J., 2015, 10.1021/acs.biochem.5b00419	FRET	$K_D(\text{mC-C}') = 0.08 \text{ }\mu\text{M}$ $K_D(\text{mC-mC}') = 0.25 \text{ }\mu\text{M}$ $K_D(\text{C-C}') = 0.28 \text{ }\mu\text{M}$ $K_D(\text{T-C}') = 0.55 \text{ }\mu\text{M}$	12mer, 1 modification site	human SRA
Qian, C., 2008, 10.1074/jbc.C800169200	Fluorescence Polarization	$K_D(\text{mC-C}') = 0.2 \text{ }\mu\text{M}$	13mer, 1 modification site	human SRA
Zhou, T., 2014, 10.1016/j.molcel.2014.04.003	Fluorescence Polarization	$K_D(\text{C-C}') = 8.61 \text{ }\mu\text{M}$ $K_D(\text{mC-C}') = 2.56 \text{ }\mu\text{M}$ $K_D(\text{hmC-hmC}') = 7.97 \text{ }\mu\text{M}$	12mer, 1 modification site	human SRA
Schneider, Trummer et al., 2019	MST	$K_D(\text{C-C}') = 1.01 \text{ }\mu\text{M}$ $K_D(\text{mC-C}') = 0.28 \text{ }\mu\text{M}$	24mer, 1 modification site	murine SRA
Schneider, Trummer et al., 2019	MST	$K_D(\text{C-C}') = 1.10 \text{ }\mu\text{M}$ $K_D(\text{mC-C}') = 0.75 \text{ }\mu\text{M}$ $K_D(\text{caC-C}') = 1.10 \text{ }\mu\text{M}$ $K_D(\text{caC-caC}') = 0.23 \text{ }\mu\text{M}$ $K_D(\text{mC-caC}') = 0.39 \text{ }\mu\text{M}$	42mer, 1 modification site	murine SRA

^a Spacer: amino acid stretch C-terminal of SRA domain

<https://doi.org/10.1371/journal.pone.0229144.t001>

Molecular dynamics simulations of the UHRF1-SRA domain bound to CpG sites with mC and caC modifications

For methylated CpG sites, UHRF1 binds stronger to mC-C' modified DNA than to the symmetric modification variant mC-mC' (Table 1) [1, 60]. As discussed above, in our experiments the opposite was observed for caC modifications, as caC-caC' DNA was preferred over caC-C'. To understand this behaviour, we performed MD simulations of UHRF1-DNA complexes with different nucleotide modifications, i.e. hemi-modified and symmetrically modified mC and caC as well as the hybrid modification variants mC-caC' and caC-mC'. As simulation of the full binding process for all variants was not feasible due to the high complexity and computational cost of such simulations, we focused on studying the complex with the flipped-out modified base bound in the protein's binding pocket, based on the experimental structure of mC-C' bound to UHRF1 (PDB-ID: 3FDE). Various experimental data indicate that this is the most relevant state for recognition: Fluorescence kinetics experiments [61] showed that the stability of the DNA flipped state is correlated to the lifetime of the flipped state bound to protein. Regarding flipping propensity, previous simulation studies showed no substantial intrinsic difference between mC and caC [62] and furthermore, NMR experiments of Dickerson-Drew dodecamers showed that both mC and caC bases were slightly less likely to flip compared to unmodified cytosines [63]. Finally, in a study of another base-flipping protein, bacterial cytosine-5-methyltransferase, it was found that specific protein-base interactions were responsible for facilitating and stabilizing the flipped out state [64]. We chose to simulate the second potentially modified base on the distal strand in the flipped-in state, motivated by

the following observations: First, stable flipping of the distal base has only been observed for proteins which can bind in a 2:1 protein-DNA ratio to the same CpG site, like UHRF2 or SUVH5, but not UHRF1 [2, 28, 29, 32, 65]. Second, the NKR finger can recognize modifications on the distal strand directly, as demonstrated by the crystal structure contacts of N494 [2, 29, 66] and third, it was observed that a single mutation of this residue abolishes the selectivity of UHRF1 between mC-C' and mC-mC' [29]. Finally, computational studies reported that the first stable intermediate in the flipping process requires a flip angle of at least 50° [62, 67]. It is difficult to imagine how direct interactions of the NKR finger could be sustained with the modified base in this position. For these reasons, we consider the complex conformation with a flipped-out pocket bound base and a flipped-in base on the distal DNA strand as the most relevant for explaining the selectivity of UHRF1.

Therefore, we did not aim at the simulation and analysis of the binding process itself and its related binding affinities, but rather at identifying similarities and differences in the binding modes of the different DNA modifications, i.e. which regions of the protein are likely to sense the chemical differences of these modification types and how this influences their interaction patterns. In contrast to mC, the caC modification contains an additional carboxyl group, which can form additional salt bridges and hydrogen bonds. Thus, we analysed whether this difference in interaction capacity could affect the polar interaction network and the local conformations of the binding pocket and NKR finger regions, which are in direct contact with the two modification sites.

Analysis of mC and caC recognition in the UHRF1-SRA nucleotide binding pocket. In Fig 4 we provide the interaction networks of the flipped base in the nucleotide binding pocket as derived from our MD simulations. Nodes represent residues of the protein and atoms of the modified DNA bases (see naming conventions in Fig 1b), while edges show the average number of hydrogen bonds (black lines) and salt bridges (red lines) between two nodes during the simulation. The canonical binding mode of mC-C' (Fig 4a) is characterized by strong hydrogen bonds between the mC atom N4 to T484 and D474 (1.84 and 1.04 hydrogen bonds on average per analysed simulation frame, respectively) and between the pyrimidine oxygen O2 and G470 and A468 (1.0 and 0.98 hydrogen bonds on average). Thus, the base is effectively locked at these two positions with the N4 and O2 atoms acting as handles. In addition, the mC backbone atom OP1 (phosphate oxygen 1) forms one stable hydrogen bond with G453 and the adjacent OP2 forms approximately two (1.86) salt bridges with R489, the latter being located at the beginning of the NKR finger. Overall, the binding pocket of the mC-C' simulation shows a regular and stable polar interaction pattern. This pattern is nearly identical to the one observed in the mC-mC' and mC-caC' simulations (Fig 4c and 4e), indicating that modifications on the distal strand have little effect on the conformation and interactions of the nucleotide binding pocket containing flipped mC.

Analysis of the binding mode of the hemi-modified caC-C' system (Fig 4b) shows that this modification leads to a very different interaction pattern: The previously observed hydrogen bonds of the nucleotide N4 atom are substantially weakened (-1.87 hydrogen bonds), while interactions of O2 are dispersed from two to three amino acids (-0.2 hydrogen bonds total). Although several hydrogen bond donors such as S486, N509, and the backbone atoms of I454 and G453 are available in the binding pocket, the carboxyl atoms O51 and O52 of caC predominantly interact with R489, forming very strong interactions (1.92 salt bridges on average) with this residue. This interaction pattern is unexpected, since the caC modification is located within the binding site, whereas R489 is located at its edge, usually interacting only with the DNA backbone. This may cause a force pulling the base out of position and could explain the weaker hydrogen bonds formed by the base's N4 nitrogen. The NKR finger region consisting of residues 488 to 502 is a flexible loop important for DNA binding with residues N494, K495,

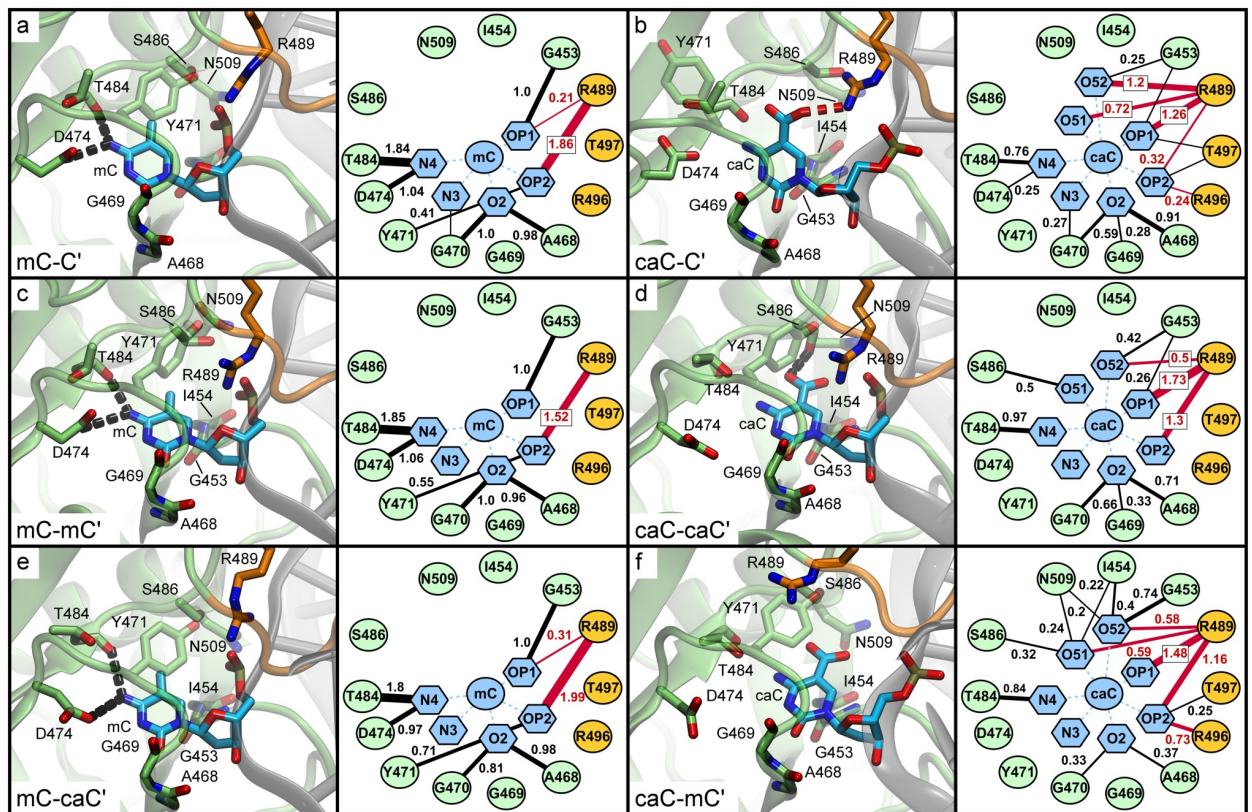


Fig 4. Interaction networks of the nucleotide binding pocket based on molecular dynamics simulations of UHRF1-SRA. Structures show representative conformations of the flipped-out modified DNA base within the binding pocket as observed during MD simulations. To the right of each structure a corresponding network of hydrogen bonds (black lines) and salt bridges (red lines) averaged over the course of the simulation is shown. Numbers next to edges show the average number of interactions per time frame. Edges representing interactions occurring in $\leq 15\%$ of simulation time are omitted for clarity. For node pairs featuring both hydrogen bonds and salt bridges, only salt bridges are displayed.

<https://doi.org/10.1371/journal.pone.0229144.g004>

and R496 at its tip. Observing that R489 is involved directly in interactions with the carboxyl oxygens establishes a direct link between the flipped-out base and the NKR finger, which predominantly interacts with the distal DNA strand. The interaction pattern of the caC-caC' system (Fig 4d) is consistent with this observation. In this system, the caC N4 and O2 atoms show an overall similar interaction pattern to the hemi-modified variant. However, distinct differences are seen in the interaction with R489: The salt bridges between the carboxyl oxygens and R489 are much weaker (only 0.5), whereas the residue forms very strong interactions (3.03) with the backbone atoms OP1 and OP2 (+ 0.96 compared to mC-C'). To compensate for the weaker R489 interactions, O51 and O52 form fluctuating weak (≤ 0.5) hydrogen bonds with S486 and G453 in the binding pocket. The caC-mC' system (Fig 4f) shows a mixture between these patterns, as R489 establishes 1.17 salt bridges to O51 and O52 of caC and 2.64 salt bridges to the caC backbone. The hydrogen bonds of the carboxyl oxygens are more dispersed compared to the caC-caC' system, interacting weakly (< 0.5) with S486, N509, and I454 and moderately strong (0.74) with G453. In turn, O2 establishes only 0.7 hydrogen bonds to G470, G469, and A468, which is 1 less than in caC-caC'. The differences we observed in the binding modes of caC-C', caC-caC' and caC-mC' indicate that the caC carboxyl oxygens have several possible interaction partners in the nucleotide binding pocket and the interaction networks are more heterogeneous compared to bound mC. In addition to interactions within the binding

pocket (S486, N509, I454, G453), caC oxygens O51/O52 can establish alternative interactions outside of the main pocket, particularly with the NKR finger residue R489. In combination with our observation that the overall interaction pattern of R489 is strongly dependent on the xC' modification on the distal strand, this suggests that the binding mode is influenced by the NKR finger, which senses that modification.

Another notable difference between the interaction networks is the hydrogen bond of the Y471 hydroxyl atom to the OP2 atom of the modified base, which is absent in the carboxylated variants (Fig 4). As Y471 has been described previously to form a hydrophobic cage, closing like a lid over the modified base [2], we analysed whether the distances between the tyrosine and pyrimidine rings were influenced by the nucleotide modification. S8 Fig shows that for both mC-C' and mC-mC' the distances cluster in two close narrow peaks with tyrosine being stabilized in its position, while for the carboxylated variants the distances fluctuate between multiple distinct conformations due to changes in the nucleotide binding mode. The distance histograms tend to differ more between replicas than during a single simulation, indicating that Y471 flips between distinct conformations with characteristic transition times roughly in the ~ 10–100 ns range or longer. Interestingly, the distribution of mC-caC' shows a similar pattern to the other methylated variants, but an additional small peak at 8–9 Å, indicating a partial destabilization of the Y471 lid. In summary, carboxylation of the flipped base leads to a different local conformation of the binding pocket compared to methylation. While during the simulations of complexes featuring a flipped mC base very similar binding modes were observed, strong differences were found in the binding modes of complexes containing a flipped caC depending on the xC' modification on the distal strand. These differences suggest potential conformational long-range correlations between the binding pocket and the NKR finger, in particular R489, which can interact directly with the carboxyl modification of the flipped-out base.

Analysis of mC and caC recognition on the distal DNA strand by the UHRF1-SRA NKR finger. Our observations so far indicated that the NKR finger could play an important role for UHRF1 to differentiate between carboxylated and methylated CpG sites. As for the binding pocket, we analysed the interaction networks between the finger residues and the second modification site on the distal DNA strand (Fig 5). In the native binding conformation represented by the mC-C' simulation (Fig 5a), N494 forms 0.76 hydrogen bonds with the OP2 atom of the unmodified DNA base backbone. This interaction has been described previously as one of the key features for differentiating between hemi-methylated and symmetrically methylated DNA [29, 33]. This is in line with our simulation of mC-mC' in which this interaction is not observed (Fig 5c), as N494 is pushed away from its native position by steric repulsion of the additional methyl group. Interestingly, a similar trend is observed for caC-C' (Fig 5b), for which the N494-OP2 hydrogen bond is also much weaker (0.13) compared to mC-C' despite the lack of any modification on the distal DNA strand. This indicates a shift in the conformation of the NKR finger similar to the mC-mC' system, only that in this case the cause is not the modified base on the distal strand, but it appears that the shift might be mediated by the conformations of R489 as described above. Investigating the interaction pattern of the caC-caC' system (Fig 5d), we observed additional strong salt bridges (3.32) between R496 and the caC' O51/O52 atoms. No interactions are formed between the modified base and N494, likely related to steric repulsion similar to the methyl group as in mC-mC'. The interaction pattern of mC-caC' (Fig 5e) is similar to caC-caC', but with slightly weaker individual interactions as R496 forms only 1.74 salt bridges to the carboxyl oxygens (- 1.58), albeit with support from spurious interactions of K495 (0.61). In contrast, the interaction pattern of caC-mC' (Fig 5f) resembles mC-mC' with an additional loss of 0.51 hydrogen bonds between N494 and the N4 base atom of mC', with nearly no polar interactions remaining between the NKR finger and the modified base.

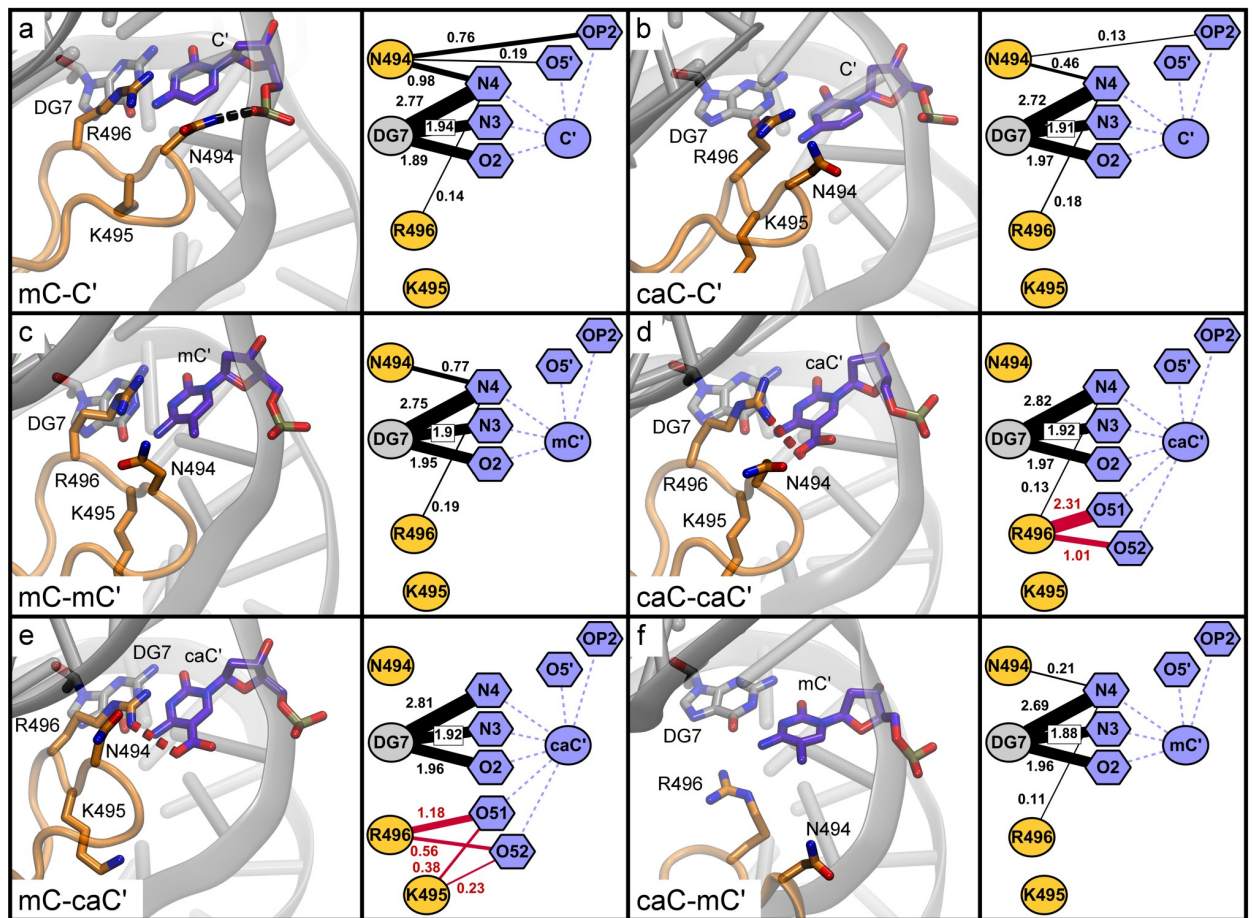


Fig 5. Interaction networks of the NKR finger based on molecular dynamics simulations of UHRF1-SRA. Structures show representative conformations of the NKR finger close to the distal (symmetrical) DNA modification site as observed during the MD simulations. To the right of each structure a corresponding network of hydrogen bonds (black lines) and salt bridges (red lines) over the course of the simulation is shown. Numbers next to edges show the average number of interactions per time frame. Edges representing interactions occurring in $\leq 10\%$ of simulation time are omitted for clarity. For node pairs featuring both hydrogen bonds and salt bridges, only salt bridges are displayed.

<https://doi.org/10.1371/journal.pone.0229144.g005>

R496 is generally a strong interaction partner for the DNA in all simulated systems, partaking in hydrogen bonds with adjacent bases and stacking interactions with the modified base. The interactions of the carboxyl group seem to modulate this role, either directly through salt bridges or by influencing stacking, although stacking effects are not quantifiable using classical force fields. As our analyses showed that only mC-C' retained the native interaction pattern of the NKR finger, we were interested in whether there was any effect on the flexibility of the finger. To quantify this, we compared the Root Mean Square Fluctuation (RMSF) for all protein residues (Fig 6a). Overall, very similar residue flexibility is observed for most regions of the protein independent of DNA modifications. Only two regions show substantial differences: The first is located in the region between residues 468 and 475, which corresponds to the conformational flexibility of Y471 discussed above. The second region featuring pronounced differences is located between residues 488 and 502 forming the NKR finger (Fig 6b). Although the NKR finger shows a different conformation in the mC-mC' simulation, the flexibility of the finger is comparable to the mC-C' reference system. In contrast, for the caC-C', caC-caC', and mC-caC' systems, the finger shows increased flexibility with a slightly different pattern:

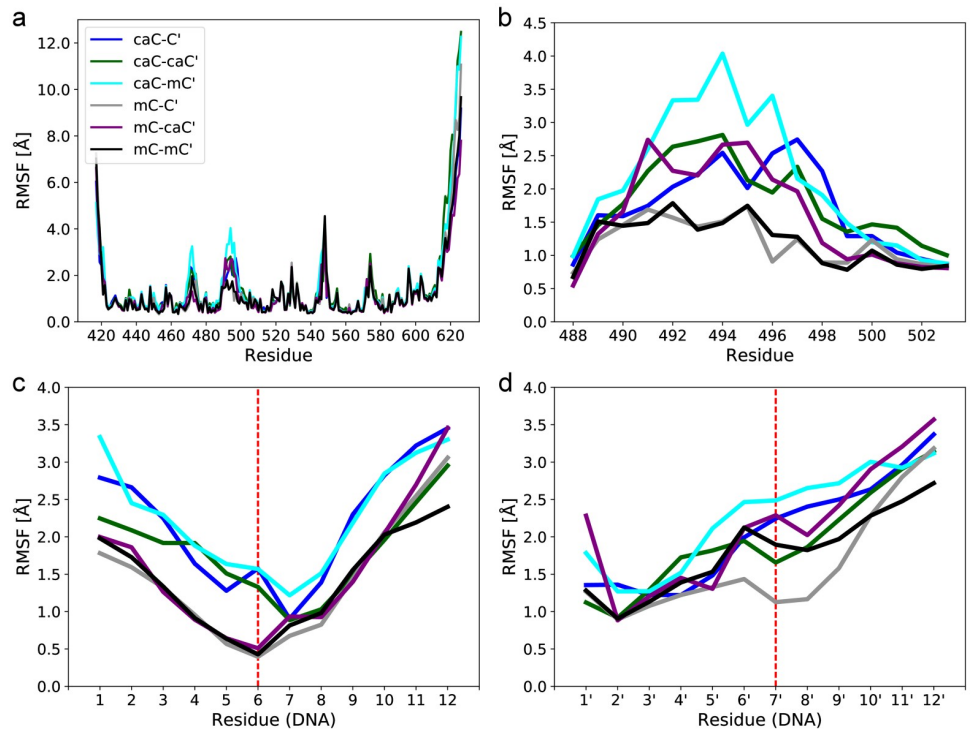


Fig 6. Root Mean Square Fluctuation (RMSF) of protein and DNA regions in molecular dynamics trajectories of UHRF1-SRA. (a) Full protein. (b) NKR finger. (c) DNA strand containing the flipped xC base bound by the protein. (d) Distal DNA strand containing the modified xC' base. Red dashed lines show the xC/xC' modification sites.

<https://doi.org/10.1371/journal.pone.0229144.g006>

The hemi-modified variant being more flexible in the 495–499 region and both the caC-caC' and mC-caC' variants more flexible between residues 490 and 494. Finally, the largest finger flexibility of all systems is observed for caC-mC', in line with the previously observed loss of interactions of the NKR finger.

UHRF1 encloses the flipped base by inserting a thumb into the minor groove and the NKR finger into the major groove of the DNA strands. Having observed differences in interaction pattern and flexibility of the NKR finger depending on the CpG modification pattern, we asked how the DNA structure around the modified sites was affected. Fig 6c shows that overall flexibility of the bound strand increases if caC is in the binding pocket, including particularly strong differences at the flipped xC base in position 6. For the distal strand, flexibility compared to mC-C' increases in all systems around the modified base 7' (Fig 6d), likely reflecting the loss of the stabilizing hydrogen bond between N494 and the DNA backbone. For a more detailed analysis, we examined how the modified bases affected the minor and major grooves, as they are strongly influenced by shifts in the DNA backbone. A small but consistent increase of minor groove width by about 1–2 Å is observed between base pairs 3 to 5 in all simulations containing caC in the binding pocket, while widths decrease by roughly the same amount between base pairs 7 and 9 (S9 Fig; locations of base pairs are shown in Fig 1d). The major groove follows a similar but weaker trend due to the large variances within replicas (S10 Fig). Although individual effects are small, their consistency and anti-symmetry with respect to the modified bases 6 and 7' is notable. Therefore, the flipped base appears to be important for the local flexibility of the DNA backbone, which is more rigid for mC and more flexible for caC. This could potentially contribute to the increased flexibility of NKR finger residues,

particularly R489, which is in a prime position to sense distortions due to its strong salt bridges with the phosphate backbone of the flipped base. These observations agree with our interaction network analyses, showing that binding of a flipped caC base leads to conformational rearrangements including the DNA strands in locations close to the modification sites.

In summary, our simulations reveal that all DNA modifications investigated lead to differences in the conformation and binding pattern of the nucleotide binding pocket and NKR finger compared to the native conformation of the mC-C' system. Interestingly, in the hemi-carboxylated variant caC-C', local conformational changes in the binding pocket are transmitted to the NKR finger via R489, which in turn becomes more flexible and thus compromises the essential N494 hydrogen bond to the C' backbone on the distal strand [29]. The symmetrically carboxylated variant caC-caC' also shows increased NKR finger flexibility, but different interaction patterns, particularly for R489 and R496. The latter forms strong salt bridges with the caC' modified base, possibly compensating for the loss of the N494 hydrogen bond. This is in strong contrast to the recognition of hemi- and symmetrically methylated CpG sites, which show much smaller differences. Our additional analysis of the hybrid modification variant mC-caC' suggests that the NKR finger can recognize and interact with the caC' modification without large changes in the binding pocket containing a flipped mC. In the opposite case of caC-mC', a heterogeneous binding pocket conformation is met with an almost complete loss of NKR finger interactions with the mC' base. Based on this simulation data, we formulate the hypothesis that UHRF1 binding of a flipped-out caC base leads to conformational changes in the protein, which can propagate to and induce shifts in the protein's NKR finger and the DNA backbone. In turn, modification of the distal DNA strand can influence the overall binding mode via steric repulsion or attractive interactions with the NKR finger, coupling recognition of both modification sites.

Discussion

The role of UHRF1 as a specific hemi-mC reader is well established [1, 3]. Reported dissociation constants range from 1.8 nM to 9.23 μ M depending on the protein construct and DNA substrate [1, 32, 36, 59, 60] (Table 1). Here, we use a relatively long DNA fragment (42 bp) with a single modified CpG site, whereas other studies have used either oligonucleotides with multiple methylated sites [1] or shorter DNA fragments with one modification site [29, 32]. We observe a relatively low preference of hemi-methylated over unmodified DNA compared to published data [1, 19, 29, 32], which we explain by the lower density of methylated sites in our experiments. To verify this relation, we also measured binding of a shorter DNA fragment which increased the affinity of UHRF1 for hemi-mC to the order of what has been reported in literature [32, 60]. A possible explanation can be given by the proposed "sliding" mechanism of UHRF1 [60, 61, 68, 69]: In this model, fast unspecific binding occurs between the protein and DNA, followed by a sliding "scan" for a modified base. Thus, the relative differences in apparent binding affinities would decrease with the length of the DNA fragments, which corresponds to our observations. In three independent assays, we observe that UHRF1 prefers binding symmetrically carboxylated CpG sites over the hemi-carboxylated variant, which is the opposite behaviour as observed for methylcytosine. Interestingly, we also measure increased affinity of UHRF1 towards hybrid mC-caC' sites. To find a possible explanation for the underlying molecular mechanisms of these differences, we performed MD simulations of the UHRF1-SRA domain in complex with hemi-, hybrid, and symmetrically modified DNA based on the crystal structure of mC-C', which features the flipped-out base in the protein's binding pocket and the second potentially modified base on the distal strand in the flipped-in state. As discussed in the results section, we preferred this approach over simulating the entire flipping process.

Our simulations revealed substantial differences in the conformations and binding patterns of the nucleotide binding pocket and the NKR finger between caC and mC modifications. If caC is bound in the binding pocket, these two regions appear to be coupled and able to influence each other in a more pronounced manner than for mC. In the caC-C' system, this coupling leads to reduced hydrogen bonding between N494 and the DNA backbone, which is an essential interaction for binding [29]. The same interaction is interrupted by steric repulsion when mC' and caC' modifications are present on the distal strand, sterically pushing the NKR finger out of its native binding position. The simulations provide no indication that the mC' modification could be beneficial to overall binding, but the caC' modification forms stable salt bridges to the NKR finger, which might compensate for the loss of the N494-DNA hydrogen bond. Thus, the caC' oxygens push the NKR finger away from its hydrogen bond with the DNA backbone and at the same time offer salt bridges to bind the finger in its new position. In this light, we propose that the carboxyl group of both, the caC and caC' bases, has a strong influence on their local interaction network partners in UHRF1, leading to conformational changes in which R489, N494, and R496 play key roles in differentiating DNA modifications. Other proteins are already known to recognize caC' modifications using finger regions: TET3, one of the three dioxygenases that generate hmC, fC, and caC, was also shown to specifically bind symmetrically carboxylated CpG sites with a finger-like structure containing a NRRT sequence [23]. Comparing the NKRT sequence of UHRF1 to the NRRT sequence of TET3, it is intriguing to speculate that such a flexible stretch of basic amino acids facilitates the binding of distant carboxyl groups.

The biological role of UHRF1 binding to symmetrically carboxylated DNA remains to be determined, considering the low abundance of this modification in cells. For this reason, it is likely that the majority of UHRF1 in a proliferating cell population interacts with hemi-methylated CpG sites, but a certain fraction may encounter and bind mC-caC' and caC-caC' depending on the cell type and cell cycle phase. Carboxylcytosine has been suggested to be an intermediate of active DNA demethylation and is detected at gene regulatory elements and promoters of actively transcribed genes, indicating dynamic DNA methylation turnover [14–16]. Several DNA repair mechanisms have been associated with this demethylation [70–72], most prominently removal of fC and caC by TDG and the base excision repair pathway [8, 73–75]. Interestingly, both UHRF1 and UHRF2 have been shown to play a role in DNA damage response [76–78]. Additionally, the bona fide UHRF1 interaction partner DNMT1 has been described to change its genomic localization upon oxidative stress [79, 80]. Furthermore, besides being demethylation intermediates, fC and caC are thought to influence DNA replication and genome stability [81, 82]. By transiently pausing RNA polymerases, fC and caC may lead to precise fine-tuning of gene expression [21]. Accordingly, the binding of UHRF1 to caC as demonstrated in our study could also represent a way of locus-specific gene expression regulation in addition to its well-established role in recognizing hemi-mC sites and initiating DNA maintenance methylation. Last but not least, UHRF1 has recently been described as a regulator of bivalent promoters and an interactor of SETD1A [83]. Interestingly, both functions have been attributed to TET proteins as well [84, 85]. This raises the intriguing possibility that UHRF1 integrates several epigenetic marks at bivalent domains and that caC, generated by TET proteins, is one of these marks involved in maintenance of the bivalent state. However, further work is needed to determine whether and where exactly UHRF1 binds caC sites *in vivo* and what implications this might have on epigenetic gene regulation.

Supporting information

S1 Fig. Raw gel images of EMSA experiments. All raw gel scans that have been used to generate the EMSA results presented in Fig 2b/2c and S5 Fig. An overview of all individual

quantitative values and the corresponding statistics is provided on page 1.

(PDF)

S2 Fig. Normalized MST traces of UHRF1 bound to C-C', mC-C', mC-caC' and caC-caC'.

Fluorescence traces that have been used to generate the binding curves in Fig 3. Traces are shown individually for all modifications and are coloured by experimental replicate. Blue and red bars indicate the time points that were used for the analysis; blue: t_{cold} (pre infra-red laser), red: t_{hot} (post infra-red laser).

(TIF)

S3 Fig. Root Mean Squared Deviation (RMSD) of DNA atoms in molecular dynamics trajectories of UHRF1-SRA. Coordinates were fitted to the initial crystal structure using the $C\alpha$ atoms of protein residues 432 to 586. Only the last 1000 frames of each trajectory were used for analysis (vertical lines). Horizontal lines were added at 4 Å to highlight trajectories with strong structural distortions.

(TIF)

S4 Fig. Root Mean Squared Deviation (RMSD) of protein atoms in molecular dynamics trajectories of UHRF1-SRA. Coordinates were fitted to the initial crystal structure using the $C\alpha$ atoms of protein residues 432 to 586. Only the last 1000 frames of each trajectory were used for analysis (vertical lines). Horizontal lines were added at 4 Å to highlight trajectories with strong structural distortions.

(TIF)

S5 Fig. EMSAs of UHRF2 with differentially modified DNA. Quantitation of the bound fraction of EMSAs of wild type UHRF2-GFP with 42 bp DNA oligonucleotides carrying different cytosine modifications. Experiments and analyses have been performed as in Fig 2.

(TIF)

S6 Fig. Melting temperatures of modified DNA in presence of UHRF1-SRA. (a) The melting temperature of double-stranded DNA containing C-C' in a CpG context (red) or no CpG site (black) with (solid lines) or without (dotted lines) a 5-fold excess of the SRA domain of UHRF1, measured using high resolution melting temperature (HRM) analysis. As control, proteins were digested by proteinase K before HRM analysis (right panel). Experiments were performed independently three times; one representative experiment is depicted as average of three technical replicates. (b) Melting temperatures as in (a) with DNA harbouring symmetric caC (green) or hemi-mC (gray) at the central CpG site.

(TIF)

S7 Fig. Size exclusion chromatograms of differentially modified DNA in the presence or absence of UHRF1-SRA. To test for different binding stoichiometries of the SRA domain towards differentially modified DNA, ATTO550-labeled DNA oligonucleotides were incubated with a 10-fold excess of SRA. Size exclusion chromatograms of analyzed DNA oligonucleotides at an absorbance of 554 nm (a) and 260nm/280nm (b) show a clear and comparable shift in retention time for the SRA-bound DNA (left peaks) compared to free DNA (right peaks).

(TIF)

S8 Fig. Histograms of distances between Y471 and the flipped-out DNA base in molecular dynamics trajectories of UHRF1-SRA. Individual replicas are shown as separate bars stacked on top of each other. Distances were measured between the geometric centres of the phenyl and pyrimidine rings. Red lines show a gaussian kernel estimate of the probability density

function (pdf). The estimated pdf of the mC-C' system is shown as black dashed lines.
(TIF)

S9 Fig. Distribution of DNA minor groove widths in molecular dynamics trajectories of UHRF1-SRA. Blue faces represent gaussian kernel estimates of the underlying values. Black bars show distribution means and standard deviations.
(TIF)

S10 Fig. Distribution of DNA major groove widths in molecular dynamics trajectories of UHRF1-SRA. Blue faces represent gaussian kernel estimates of the underlying values. Black bars show distribution means and standard deviations.
(TIF)

S1 Text. Additional experimental procedures.
(DOCX)

S1 File. Parameter files for mC/caC used during molecular dynamics simulations.
(ZIP)

Author Contributions

Conceptualization: Markus Schneider, Carina Trummer, Heinrich Leonhardt, Christina Bauer, Iris Antes.

Data curation: Markus Schneider, Carina Trummer, Christina Bauer.

Formal analysis: Markus Schneider, Carina Trummer, Andreas Stengl, Peng Zhang, Christina Bauer.

Funding acquisition: M. Cristina Cardoso, Heinrich Leonhardt, Iris Antes.

Investigation: Carina Trummer, Andreas Stengl, Peng Zhang, Aleksandra Szwagierczak, Christina Bauer.

Methodology: Markus Schneider, Carina Trummer, Heinrich Leonhardt, Christina Bauer, Iris Antes.

Project administration: Heinrich Leonhardt, Iris Antes.

Resources: Heinrich Leonhardt, Iris Antes.

Software: Markus Schneider.

Supervision: Heinrich Leonhardt, Christina Bauer, Iris Antes.

Validation: Markus Schneider, Carina Trummer, Christina Bauer.

Visualization: Markus Schneider, Carina Trummer, Christina Bauer.

Writing – original draft: Markus Schneider, Christina Bauer, Iris Antes.

Writing – review & editing: Markus Schneider, Carina Trummer, Christina Bauer, Iris Antes.

References

1. Bostick M, Kim JK, Esteve P-O, Clark A, Pradhan S, Jacobsen SE. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*. 2007; 317:1760–4. <https://doi.org/10.1126/science.1147939> PMID: 17673620.
2. Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*. 2008; 455:826–9. <https://doi.org/10.1038/nature07280> PMID: 18772888.

3. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*. 2007; 450(7171):908–12. <https://doi.org/10.1038/nature06397> PMID: 17994007.
4. Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature*. 2013; 502(7470):249–53. <https://doi.org/10.1038/nature12488> PMID: 24013172.
5. Qin W, Wolf P, Liu N, Link S, Smets M, La Mastra F, et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res*. 2015; 25(8):911–29. <https://doi.org/10.1038/cr.2015.72> PMID: 26065575.
6. Ishiyama S, Nishiyama A, Saeki Y, Moritsugu K, Morimoto D, Yamaguchi L, et al. Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. *Mol Cell*. 2017; 68(2):350–60 e7. <https://doi.org/10.1016/j.molcel.2017.09.037> PMID: 29053958.
7. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009; 324(5929):930–5. <https://doi.org/10.1126/science.1170116> PMID: 19372391.
8. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011; 333(6047):1303–7. <https://doi.org/10.1126/science.1210944> PMID: 21817016.
9. Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl*. 2011; 50(31):7008–12. <https://doi.org/10.1002/anie.201103899> PMID: 21721093.
10. Nabel CS, Kohli RM. Molecular biology. Demystifying DNA demethylation. *Science*. 2011; 333(6047):1229–30. <https://doi.org/10.1126/science.1211917> PMID: 21885763.
11. Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol*. 2014; 10(7):574–81. <https://doi.org/10.1038/nchembio.1532> PMID: 24838012.
12. Globisch D, Münzel M, Müller M, Michalakakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*. 2010; 5:1–9. <https://doi.org/10.1371/journal.pone.0015367> PMID: 21203455.
13. Eleftheriou M, Pascual AJ, Wheldon LM, Perry C, Abakir A, Arora A, et al. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clinical epigenetics*. 2015; 7. <https://doi.org/10.1186/s13148-015-0117-x> PMID: 26300993.
14. Lu X, Han D, Boxuan Simen Z, Song C-X, Zhang L-S, Doré LC, et al. Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Research*. 2015; 25:386–9. <https://doi.org/10.1038/cr.2015.5> PMID: 25591929.
15. Neri F, Incarnato D, Krepelova A, Rapelli S, Anselmi F, Parlato C, et al. Single-Base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA Methylation Dynamics. *Cell Reports*. 2015; 10:674–83. <https://doi.org/10.1016/j.celrep.2015.01.008> PMID: 25660018.
16. Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*. 2013; 153:692–706. <https://doi.org/10.1016/j.cell.2013.04.002> PMID: 23602152.
17. Bronner C, Achour M, Arima Y, Chataigneau T, Saya H, Schini-Kerth VB. The UHRF family: oncogenes that are drugable targets for cancer therapy in the near future? *Pharmacol Ther*. 2007; 115(3):419–34. <https://doi.org/10.1016/j.pharmthera.2007.06.003> PMID: 17658611.
18. Pichler G, Wolf P, Schmidt CS, Meilinger D, Schneider K, Frauer C, et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem*. 2011; 112(9):2585–93. <https://doi.org/10.1002/jcb.23185> PMID: 21598301.
19. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, et al. Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*. 2013; 152:1146–59. <https://doi.org/10.1016/j.cell.2013.02.004> PMID: 23434322.
20. Rajakumara E, Nakarakanti NK, Nivya MA, Satish M. Mechanistic insights into the recognition of 5-methylcytosine oxidation derivatives by the SUVH5 SRA domain. *Scientific Reports*. 2016; 6:20161. <https://doi.org/10.1038/srep20161> PMID: 26841909
21. Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, et al. Molecular basis for 5-carboxylcytosine recognition by RNA polymerase II elongation complex. *Nature*. 2015; 523:621–5. <https://doi.org/10.1038/nature14482> PMID: 26123024.

22. Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.* 2014; 28(20):2304–13. <https://doi.org/10.1101/gad.250746.114> PMID: 25258363.
23. Jin S-G, Zhang Z-M, Dunwell TL, Harter MR, Wu X, Johnson J, et al. Tet3 reads 5-carboxylcytosine through its CXXC domain and is a potential guardian against neurodegeneration. *Cell Rep.* 2016; 14:493–505. <https://doi.org/10.1016/j.celrep.2015.12.044> PMID: 26774490.
24. Gowher H, Jeltsch A. Mammalian DNA methyltransferases: new discoveries and open questions. *Biochemical Society Transactions.* 2018; 46(5):1191–202. <https://doi.org/10.1042/BST20170574> PMID: 30154093
25. Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* 2012; 8(6):e1002750. <https://doi.org/10.1371/journal.pgen.1002750> PMID: 22761581.
26. Xu L, Chen YC, Chong J, Fin A, McCoy LS, Xu J, et al. Pyrene-based quantitative detection of the 5-formylcytosine loci symmetry in the CpG duplex content during TET-dependent demethylation. *Angew Chem Int Ed Engl.* 2014; 53(42):11223–7. <https://doi.org/10.1002/anie.201406220> PMID: 25159856.
27. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012; 149(6):1368–80. <https://doi.org/10.1016/j.cell.2012.04.027> PMID: 22608086.
28. Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature.* 2008; 455(7214):818–21. <https://doi.org/10.1038/nature07249> PMID: 18772891.
29. Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature.* 2008; 455(7214):822–5. <https://doi.org/10.1038/nature07273> PMID: 18772889.
30. Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, Antes I, et al. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE.* 2011; 6:1–8. <https://doi.org/10.1371/journal.pone.0021306> PMID: 21731699.
31. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research.* 2012; 40(11):4841–9. <https://doi.org/10.1093/nar/gks155> PMID: 22362737
32. Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, et al. Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2. *Molecular Cell.* 2014; 54:879–86. <https://doi.org/10.1016/j.molcel.2014.04.003> PMID: 24813944.
33. Bianchi C, Zangi R. UHRF1 discriminates against binding to fully-methylated CpG-Sites by steric repulsion. *Biophysical Chemistry.* 2013; 171:38–45. <https://doi.org/10.1016/j.bpc.2012.10.002> PMID: 23245651.
34. Hashimoto H, Zhang X, Cheng X. Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH. *DNA Repair.* 2013; 12:535–40. <https://doi.org/10.1016/j.dnarep.2013.04.003> PMID: 23680598.
35. Gelato KA, Tauber M, Ong MS, Winter S, Hiragami-Hamada K, Sindlinger J, et al. Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. *Mol Cell.* 2014; 54(6):905–19. <https://doi.org/10.1016/j.molcel.2014.04.004> PMID: 24813945.
36. Fang J, Cheng J, Wang J, Zhang Q, Liu M, Gong R, et al. Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. *Nat Commun.* 2016; 7:11197. <https://doi.org/10.1038/ncomms11197> PMID: 27045799.
37. Harrison JS, Cornett EM, Goldfarb D, DaRosa PA, Li ZM, Yan F, et al. Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. *Elife.* 2016; 5. <https://doi.org/10.7554/eLife.17101> PMID: 27595565.
38. Vaughan RM, Dickson BM, Whelihan MF, Johnstone AL, Cornett EM, Cheek MA, et al. Chromatin structure and its chemical modifications regulate the ubiquitin ligase substrate selectivity of UHRF1. *Proc Natl Acad Sci U S A.* 2018; 115(35):8775–80. <https://doi.org/10.1073/pnas.1806373115> PMID: 30104358.
39. Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, Leonhardt H. The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.* 2010; 38(6):1796–804. <https://doi.org/10.1093/nar/gkp1152> PMID: 20026581.
40. Ivani I, Dans PD, Noy A, Perez A, Faustino I, Hospital A, et al. Parmbsc1: a refined force field for DNA simulations. *Nat Methods.* 2016; 13(1):55–8. <https://doi.org/10.1038/nmeth.3658> PMID: 26569599.
41. Lankaš F, Cheatham TE, Špačáková Na, Hobza P, Langowski J, Šponer J. Critical Effect of the N2 Amino Group on Structure, Dynamics, and Elasticity of DNA Polypurine Tracts. *Biophysical Journal.* 2002; 82(5):2592–609. [https://doi.org/10.1016/s0006-3495\(02\)75601-4](https://doi.org/10.1016/s0006-3495(02)75601-4) PMID: 11964246

42. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE 3rd, Laughton CA, et al. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*. 2007; 92(11):3817–29. <https://doi.org/10.1529/biophysj.106.097782> PMID: 17351000.
43. Cieplak P, Cornell Wendy D, Bayly C, Kollman Peter A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry*. 1995; 16(11):1357–77. <https://doi.org/10.1002/jcc.540161106>
44. Vanquuelef E, Simon S, Marquant G, Garcia E, Klimerek G, Delepine JC, et al. R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res*. 2011; 39(Web Server issue):W511–7. <https://doi.org/10.1093/nar/gkr288> PMID: 21609950.
45. Wang F, Becker J-P, Cieplak P, Dupradeau F-Y. R.E.D. Python: Object oriented programming for Amber force fields. Université de Picardie—Jules Verne, Sanford Burnham Prebys Medical Discovery Institute. 2013.
46. Dupradeau FY, Pigache A, Zaffran T, Savineau C, Lelong R, Grivel N, et al. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys*. 2010; 12(28):7821–39. <https://doi.org/10.1039/c0cp00111b> PMID: 20574571.
47. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*. 1993; 97(40):10269–80. <https://doi.org/10.1021/j100142a004>
48. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09 Revision A.2. 2009.
49. Case DA, Cerutti DS, T.E. Cheatham I, Darden TA, Duke RE, Giese TJ, et al. AMBER 2017. University of California, San Francisco. 2017.
50. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 2015; 11(8):3696–713. <https://doi.org/10.1021/acs.jctc.5b00255> PMID: 26574453.
51. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983; 79(2):926. <https://doi.org/10.1063/1.445869>
52. Miyamoto S, Kollman PA. Settle—an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of Computational Chemistry*. 1992; 13(8):952–62.
53. Duell ER, Glaser M, Le Chapelain C, Antes I, Groll M, Huber EM. Sequential Inactivation of Gliotoxin by the S-Methyltransferase TmtA. *ACS chemical biology*. 2016; 11(4):1082–9. <https://doi.org/10.1021/acscchembio.5b00905> PMID: 26808594.
54. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput*. 2013; 9(7):3084–95. <https://doi.org/10.1021/ct400341p> PMID: 26583988.
55. El Hassan MA, Calladine CR. Two distinct modes of protein-induced bending in DNA. *J Mol Biol*. 1998; 282(2):331–43. <https://doi.org/10.1006/jmbi.1998.1994> PMID: 9735291
56. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996; 14(1):33–8, 27–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570.
57. Hunter JD. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. 2007; 9(3):90–5. <https://doi.org/10.1109/MCSE.2007.55>
58. Seidel SA, Dijkman PM, Lea WA, van den Bogaart G, Jerabek-Willemsen M, Lazić A, et al. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*. 2013; 59(3):301–15. <https://doi.org/10.1016/j.ymeth.2012.12.005> PMID: 23270813.
59. Qian C, Li S, Jakoncic J, Zeng L, Walsh MJ, Zhou MM. Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J Biol Chem*. 2008; 283(50):34490–4. <https://doi.org/10.1074/jbc.C800169200> PMID: 18945682.
60. Greiner VJ, Kovalenko L, Humbert N, Richert L, Birck C, Ruff M, et al. Site-Selective Monitoring of the Interaction of the SRA Domain of UHRF1 with Target DNA Sequences Labeled with 2-Aminopurine. *Biochemistry*. 2015; 54(39):6012–20. <https://doi.org/10.1021/acs.biochem.5b00419> PMID: 26368281.
61. Kilin V, Gavvala K, Barthes NPF, Michel BY, Shin D, Boudier C, et al. Dynamics of Methylated Cytosine Flipping by UHRF1. *Journal of the American Chemical Society*. 2017; 139(6):2520–8. <https://doi.org/10.1021/jacs.7b00154> PMID: 28112929
62. Helabad MB, Kanaan N, Imhof P. Base Flip in DNA Studied by Molecular Dynamics Simulations of Differently-Oxidized Forms of Methyl-Cytosine. *International Journal of Molecular Sciences*. 2014; 15(7):11799–816. <https://doi.org/10.3390/ijms150711799> PMID: 24995694

63. Szulik MW, Pallan PS, Nocek B, Voehler M, Banerjee S, Brooks S, et al. Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson–Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine. *Biochemistry*. 2015; 54(5):1294–305. <https://doi.org/10.1021/bi501534x> PMID: 25632825
64. Huang N, Banavali NK, MacKerell AD. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proceedings of the National Academy of Sciences*. 2003; 100(1):68.
65. Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D, et al. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes Dev*. 2011; 25(2):137–52. <https://doi.org/10.1101/gad.1980311> PMID: 21245167.
66. Arita K, Isogai S, Oda T, Unoki M, Sugita K, Sekiyama N, et al. 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*. 2012; 109(32):12950–5. <https://doi.org/10.1073/pnas.1203701109> PMID: 22837395.
67. Bianchi C, Zangi R. Dual base-flipping of cytosines in a CpG dinucleotide sequence. *Biophysical Chemistry*. 2014; 187–188:14–22. <https://doi.org/10.1016/j.bpc.2013.12.005> PMID: 24469333.
68. Hashimoto H, Horton JR, Zhang X, Cheng X. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics*. 2009; 4:8–14. <https://doi.org/10.4161/epi.4.1.7370> PMID: 19077538.
69. Bronner C, Fuhrmann G, Chédin FL, Macaluso M, Dhe-Paganon S. UHRF1 Links the Histone code and DNA Methylation to ensure Faithful Epigenetic Memory Inheritance. *Genetics & epigenetics*. 2010; 2009(2):29–36.
70. Grin I, Ishchenko AA. An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Res*. 2016; 44(8):3713–27. <https://doi.org/10.1093/nar/gkw059> PMID: 26843430.
71. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*. 2013; 502:472–9. <https://doi.org/10.1038/nature12750> PMID: 24153300.
72. Bochtler M, Kolano A, Xu GL. DNA demethylation pathways: Additional players and regulators. *Bioessays*. 2017; 39(1):1–13. <https://doi.org/10.1002/bies.201600178> PMID: 27859411.
73. Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*. 2011; 286(41):35334–8. <https://doi.org/10.1074/jbc.C111.284620> PMID: 21862836.
74. Muller U, Bauer C, Siegl M, Rottach A, Leonhardt H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res*. 2014; 42(13):8592–604. <https://doi.org/10.1093/nar/gku552> PMID: 24948610.
75. Weber AR, Krawczyk C, Robertson AB, Kusnierczyk A, Vagbo CB, Schuermann D, et al. Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat Commun*. 2016; 7:10806. <https://doi.org/10.1038/ncomms10806> PMID: 26932196.
76. Luo T, Cui S, Bian C, Yu X. Uhrf2 is important for DNA damage response in vascular smooth muscle cells. *Biochem Biophys Res Commun*. 2013; 441(1):65–70. <https://doi.org/10.1016/j.bbrc.2013.10.018> PMID: 24134842.
77. Mistry H, Tamblin L, Butt H, Sigoreo D, Gracias A, Larin M, et al. UHRF1 is a genome caretaker that facilitates the DNA damage response to gamma-irradiation. *Genome Integr*. 2010; 1(1):7. <https://doi.org/10.1186/2041-9414-1-7> PMID: 20678257.
78. Tian Y, Paramasivam M, Ghosal G, Chen D, Shen X, Huang Y, et al. UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold. *Cell Reports*. 2015; 10:1957–66. <https://doi.org/10.1016/j.celrep.2015.03.038> PMID: 25818288.
79. Laget S, Miotto B, Chin HG, Esteve PO, Roberts RJ, Pradhan S, et al. MBD4 cooperates with DNMT1 to mediate methyl-DNA repression and protects mammalian cells from oxidative stress. *Epigenetics*. 2014; 9(4):546–56. <https://doi.org/10.4161/epi.27695> PMID: 24434851.
80. O'Hagan HM, Wang W, Sen S, Destefano Shields C, Lee SS, Zhang YW, et al. Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell*. 2011; 20(5):606–19. <https://doi.org/10.1016/j.ccr.2011.09.012> PMID: 22094255.
81. Kamiya H, Tsuchiya H, Karino N, Ueno Y, Matsuda A, Harashima H. Mutagenicity of 5-Formylcytosine, an Oxidation Product of 5-Methylcytosine, in DNA in Mammalian Cells1. *The Journal of Biochemistry*. 2002; 132(4):551–5. <https://doi.org/10.1093/oxfordjournals.jbchem.a003256> PMID: 12359069
82. Shibutani T, Ito S, Toda M, Kanao R, Collins LB, Shibata M, et al. Guanine- 5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci Rep*. 2014; 4:5220. <https://doi.org/10.1038/srep05220> PMID: 24910358.

83. Kim KY, Tanaka Y, Su J, Cakir B, Xiang Y, Patterson B, et al. Uhrf1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells through Setd1a. *Nat Commun.* 2018; 9(1):2583. <https://doi.org/10.1038/s41467-018-04818-0> PMID: 29968706.
84. Verma N, Pan H, Dore LC, Shukla A, Li QV, Pelham-Webb B, et al. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat Genet.* 2018; 50(1):83–95. <https://doi.org/10.1038/s41588-017-0002-y> PMID: 29203910.
85. Deplus R, Delatte B, Schwinn MK, Defrance M, Mendez J, Murphy N, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* 2013; 32(5):645–55. <https://doi.org/10.1038/emboj.2012.357> PMID: 23353889.

Supporting information

S1 Fig. Raw gel images of EMSA experiments.

All raw gel scans that have been used to generate the EMSA results presented in [Fig 2b/2c](#) and [S5 Fig](#). An overview of all individual quantitative values and the corresponding statistics is provided on page 1.

Information about raw images in Fig S1:

Page 1 shows an overview of all performed EMSA experiments for UHRF1 and UHRF2. Quantitation has been performed as described in the Materials and Methods section. Boxplots in Fig 2c and S3 Fig are based on these numbers.

Pages 2 to 10 show minimally cropped original scans of all gels used for the study; neither brightness nor contrast has been altered. Gels have been marked with empty lanes to avoid confusion during handling. These empty lanes have been ignored during analysis and have been removed in the main figure (Fig 2b).





Each scan has 3 channels (shown individually):

- the 488-channel represents the GFP-coupled protein
- the 550-channel depicts the signal of the modified DNA oligonucleotide (modifications are indicated in the figure)
- the 647-channel shows the signal of the unmodified control DNA

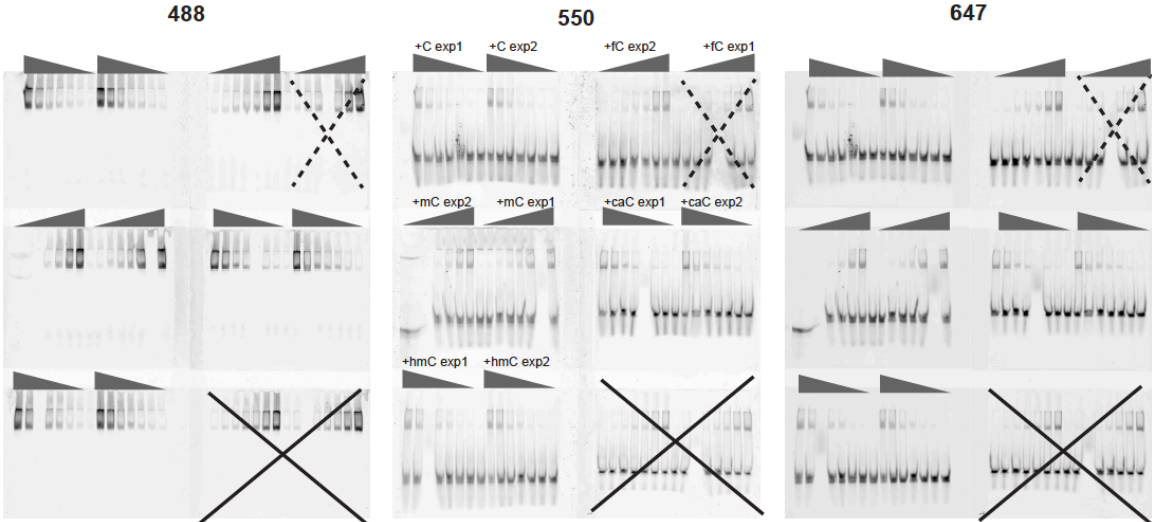
UHRF1	C/C	hemi-mC/C	hemi-hmC/C	hemi-fC/C	hemi-caC/C	sym. mC/C	sym. hmC/C	sym. fC/C	sym. caC/C
Exp1	0.85					1.06	1.34		1.71
Exp2	1.20					1.07	1.39	1.19	1.71
Exp3									1.81
Exp4	0.91					1.51	1.18		1.78
Exp5	0.99					1.41	0.75	0.95	1.88
Exp6	0.63					1.20	0.81	0.81	1.71
Exp7	1.23	1.63		1.12	1.12				
Exp8	0.82	0.94	0.92	0.84	1.09				
Exp9	1.29	2.04	1.65	1.75	1.51				
Exp10	1.22	1.38	0.56	0.70	1.19				
mean	1.02	1.50	1.04	1.10	1.23	1.25	1.09	0.98	1.77
stddev	0.23	0.46	0.55	0.47	0.19	0.20	0.30	0.19	0.07
norm.	1.00	1.47	1.02	1.08	1.21	1.23	1.08	0.97	1.74
stddev	0.32	0.56	0.59	0.52	0.33	0.34	0.38	0.28	0.39
p-value	1.00E+00	2.45E-02	9.09E-01	6.64E-01	1.38E-01	8.23E-02	5.94E-01	8.20E-01	3.17E-06
#Rep	9	4	3	4	4	5	5	3	6

UHRF2	C/C	hemi-mC/C	hemi-hmC/C	hemi-fC/C	hemi-caC/C	sym. mC/C	sym. hmC/C	sym. fC/C	sym. caC/C
Exp1	1.51					0.99	1.78	2.03	2.48
Exp2	1.42					1.17	1.43	1.42	2.23
Exp3	1.32					1.22	1.35	1.32	2.35
Exp4	1.29	1.25	1.32	0.94					
Exp5	0.87	0.78	0.78	0.77	0.90				
Exp6	0.71	0.86		0.77	0.87				
Exp7	1.17	0.90	0.88	1.12	1.01				
mean	1.18	0.94	0.99	0.90	0.92	1.13	1.52	1.59	2.35
stddev	0.29	0.21	0.29	0.17	0.08	0.12	0.23	0.38	0.12
norm.	1.00	0.80	0.84	0.76	0.78	0.95	1.28	1.34	1.99
stddev	0.35	0.26	0.32	0.23	0.20	0.26	0.37	0.46	0.50
p-value	1.00E+00	1.86E-01	3.72E-01	1.11E-01	1.79E-01	7.64E-01	1.17E-01	1.02E-01	1.89E-04
#Rep	7	4	3	4	3	3	3	3	3

Legend

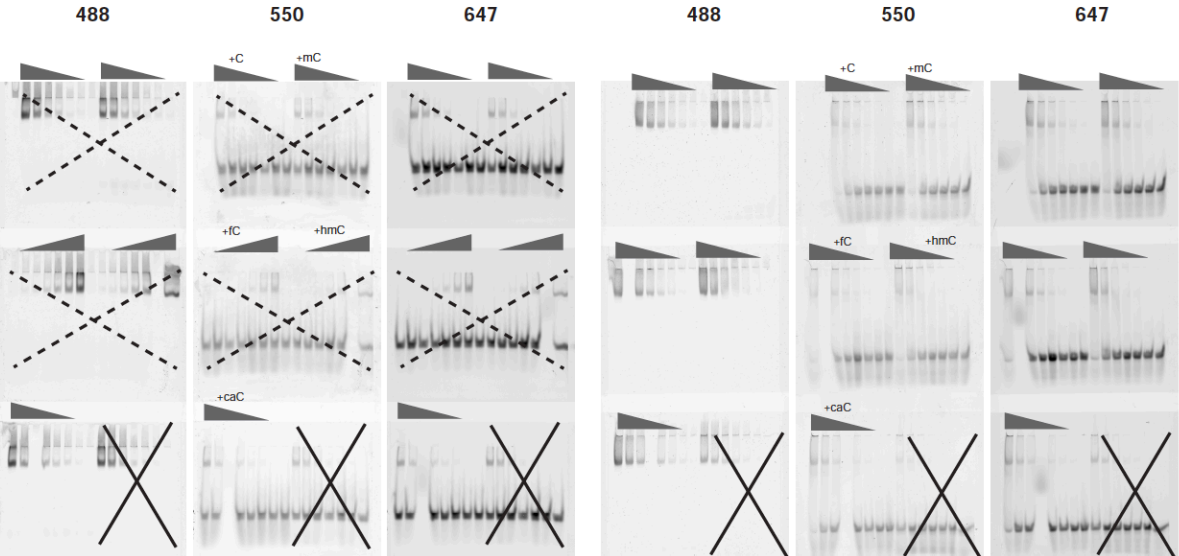
-  protein serial dilution
-  excluded due to lack of relevance
-  excluded due to quality concerns or technical reasons
-  gel belongs to different experiment and is labeled there

UHRF1 symmetric modifications, Experiments 1 + 2



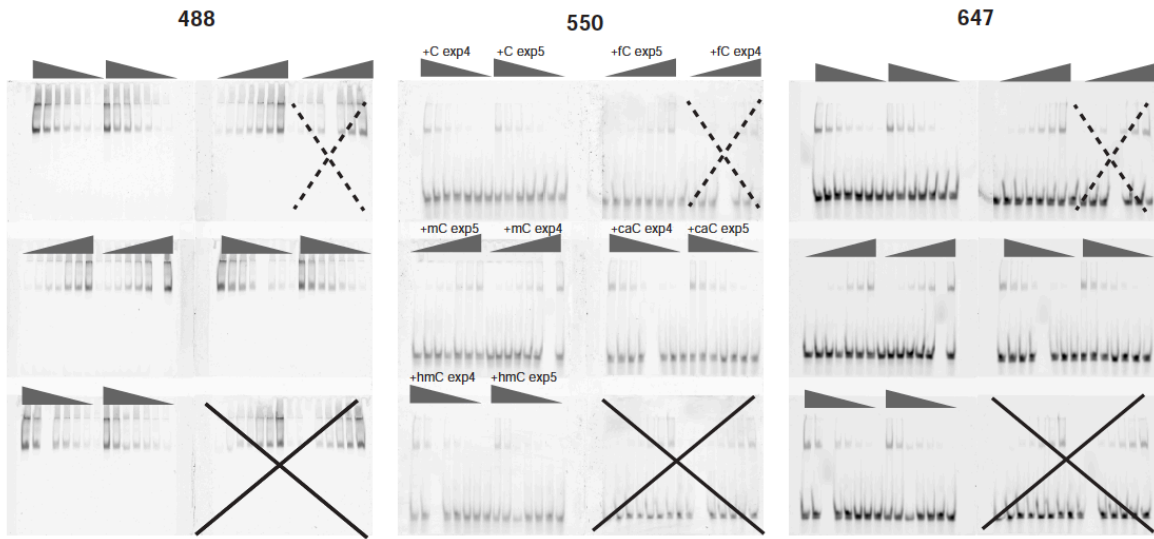
page 2

UHRF1 symmetric modifications, Experiments 3 + 6



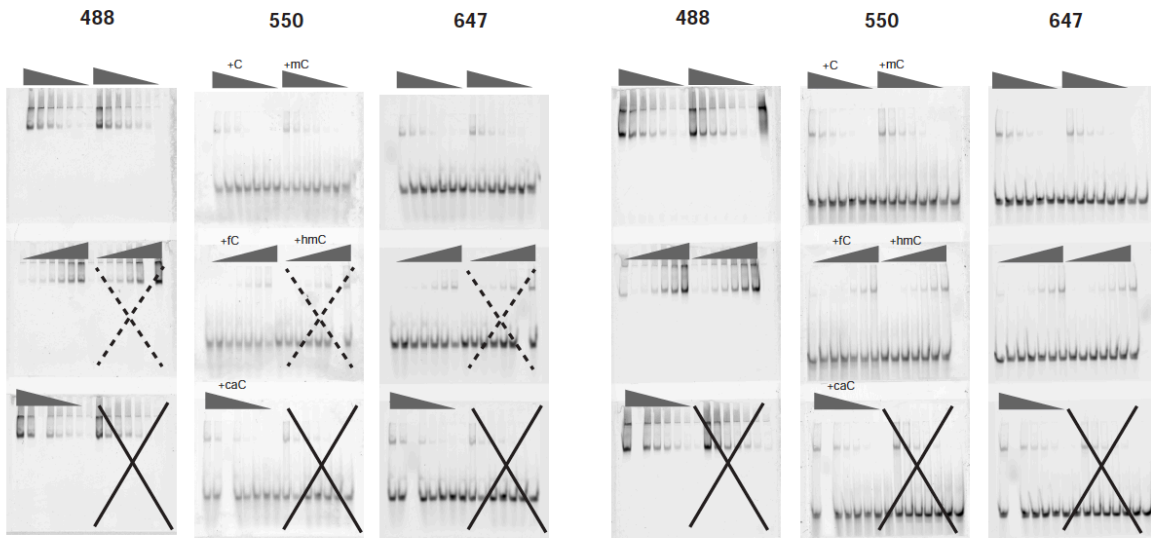
page 3

UHRF1 symmetric modifications, Experiments 4 + 5



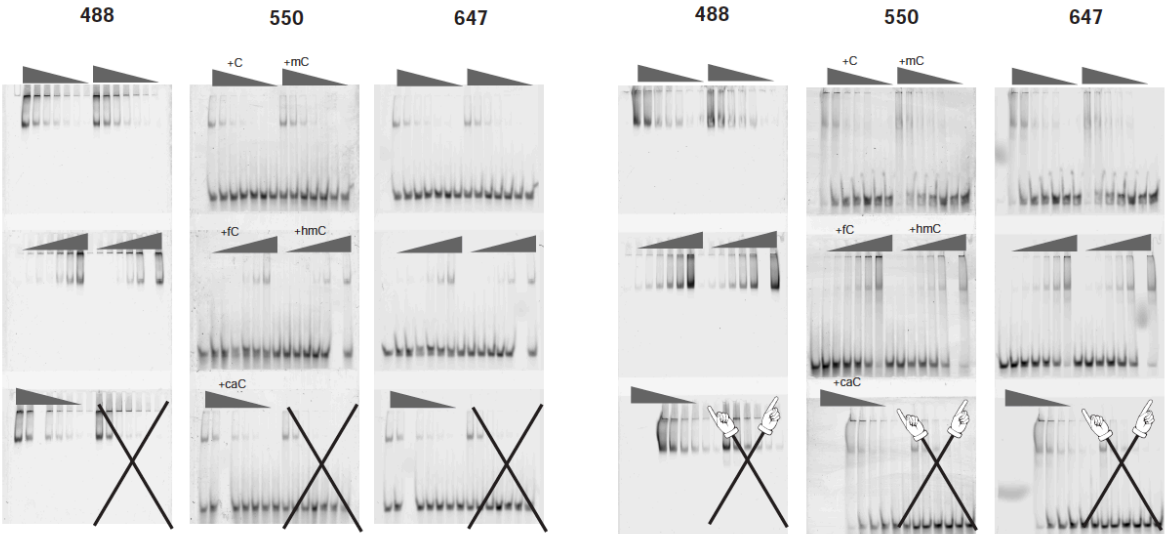
page 4

UHRF1 hemi-modifications, Experiments 7 + 8



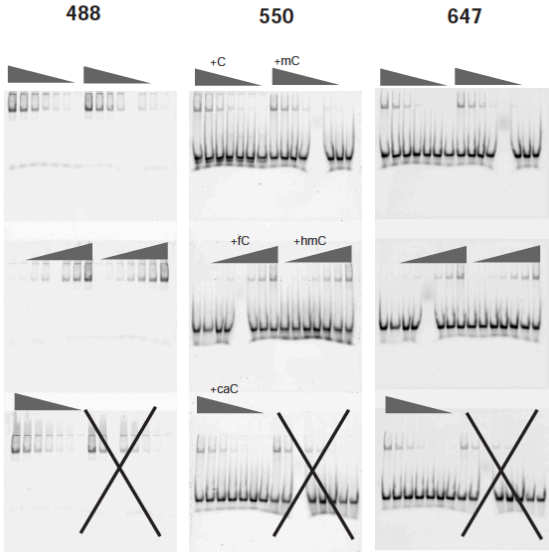
page 5

UHRF1 hemi-modifications, Experiments 9 + 10



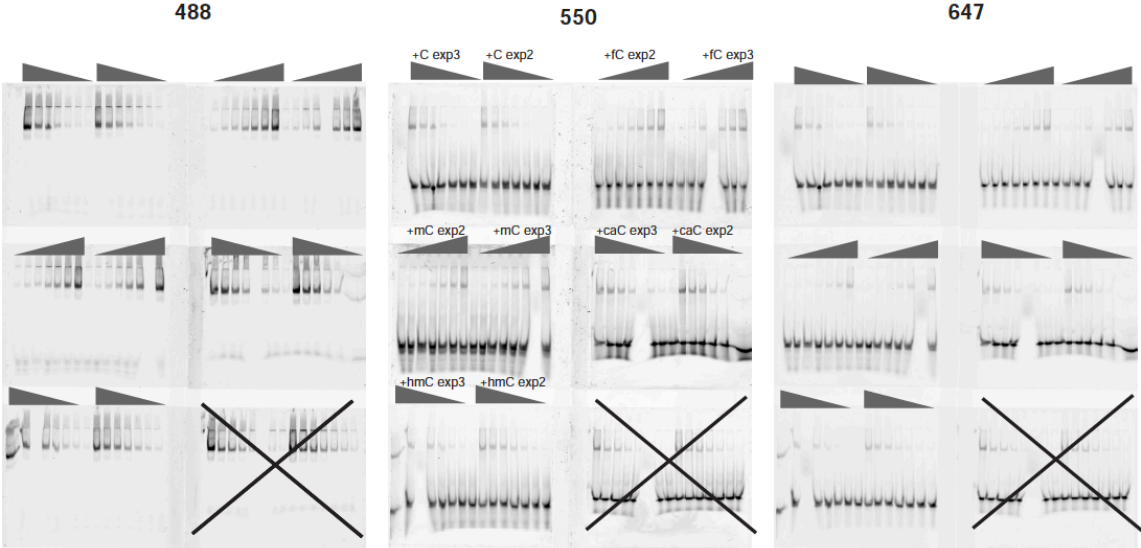
page 6

UHRF2 symmetric modifications, Experiment 1



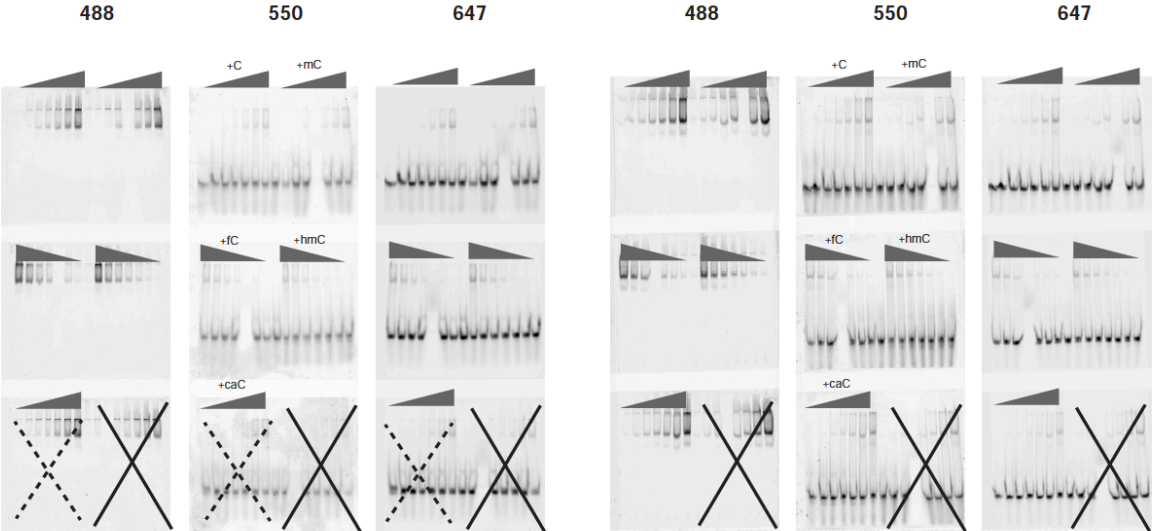
page 7

UHRF2 symmetric modifications, Experiments 2 + 3



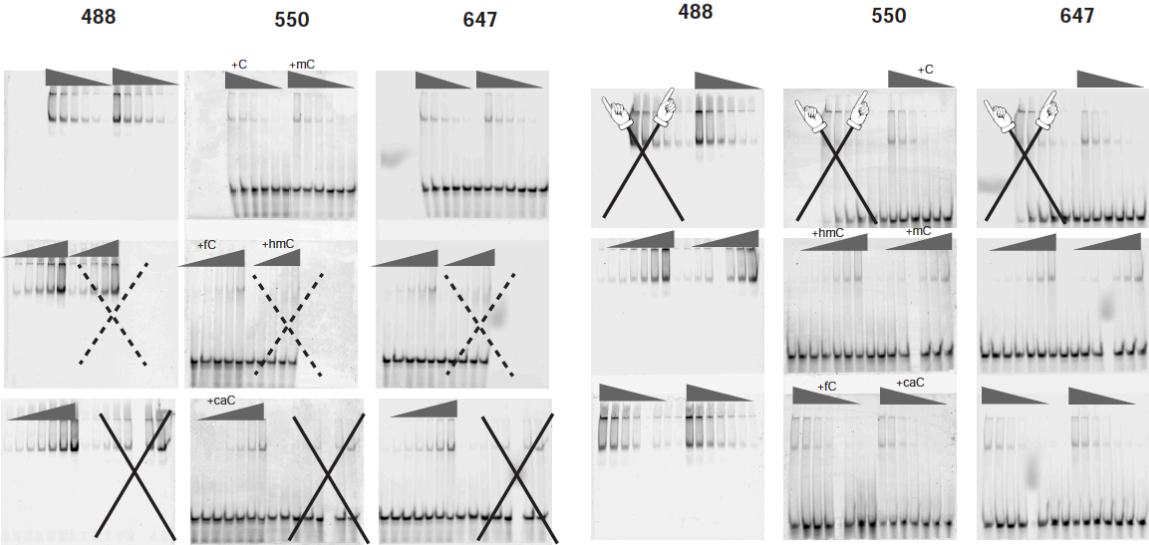
page 8

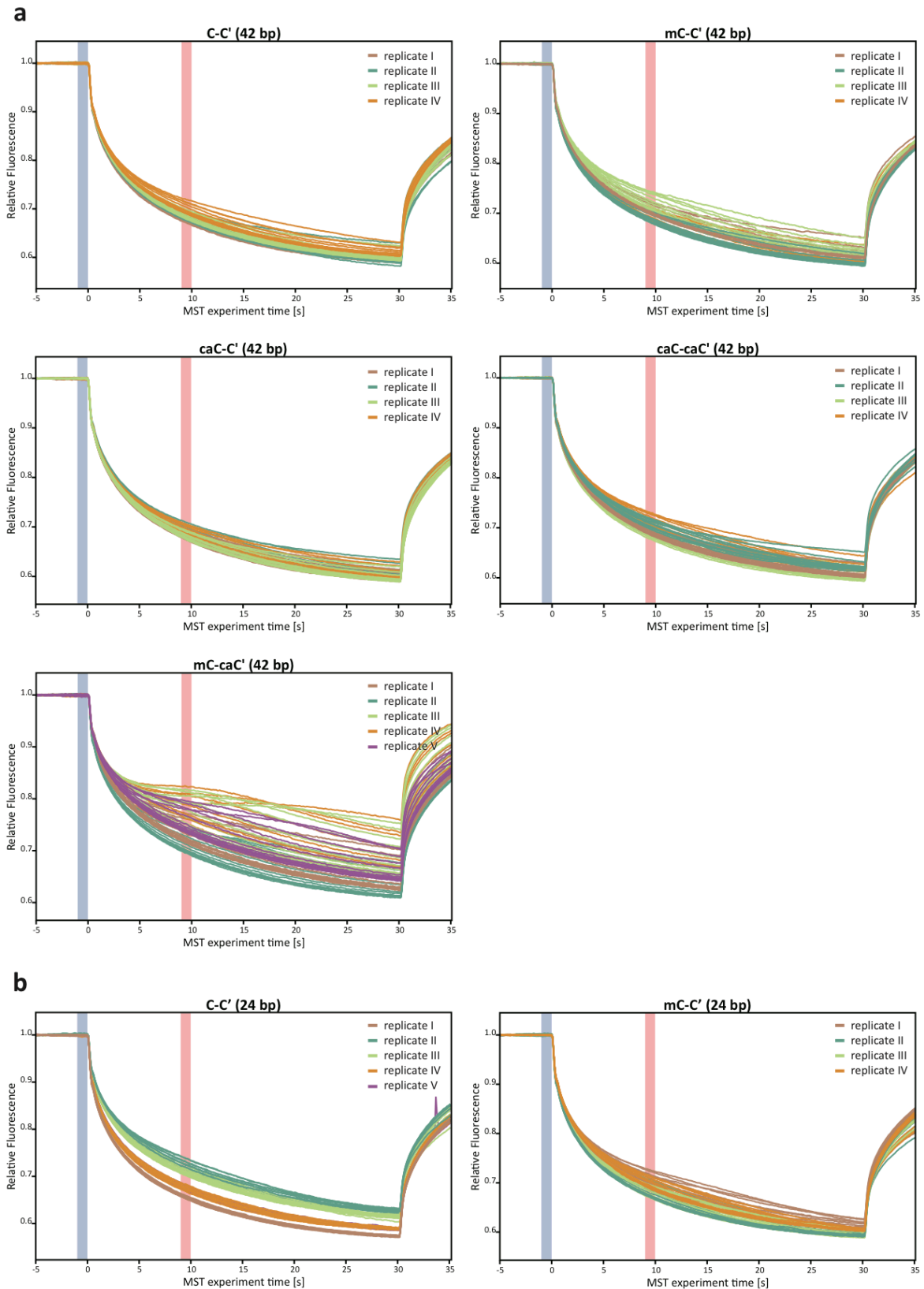
UHRF2 hemi-modifications, Experiments 4 + 5



page 9

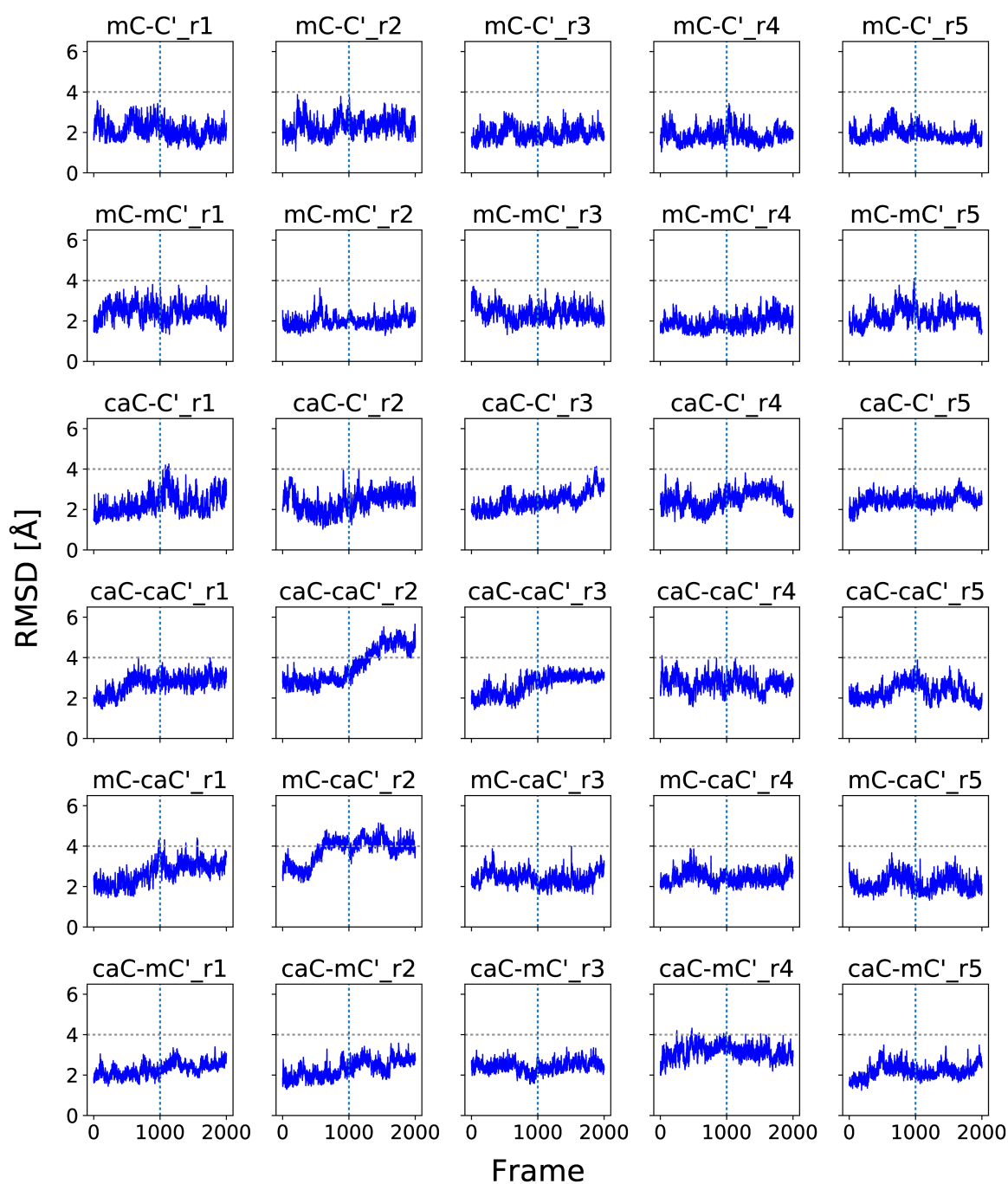
UHRF2 hemi-modifications, Experiments 6 + 7





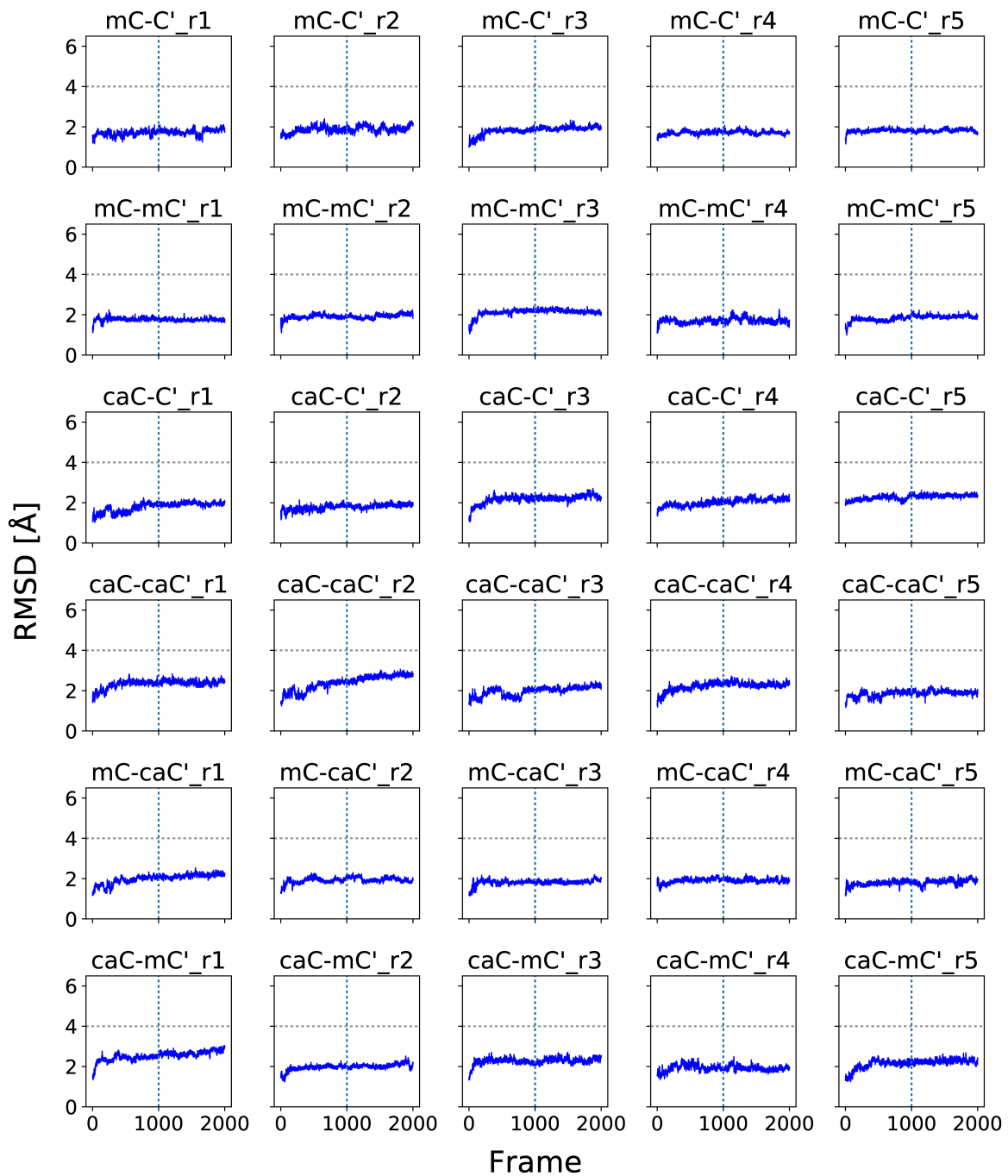
S2 Fig. Normalized MST traces of UHRF1 bound to C-C', mC-C', mC-caC' and caC-caC'.

Fluorescence traces that have been used to generate the binding curves in Fig 3. Traces are shown individually for all modifications and are coloured by experimental replicate. Blue and red bars indicate the time points that were used for the analysis; blue: t_{cold} (pre-infra-red laser), red: t_{hot} (post-infra-red laser).



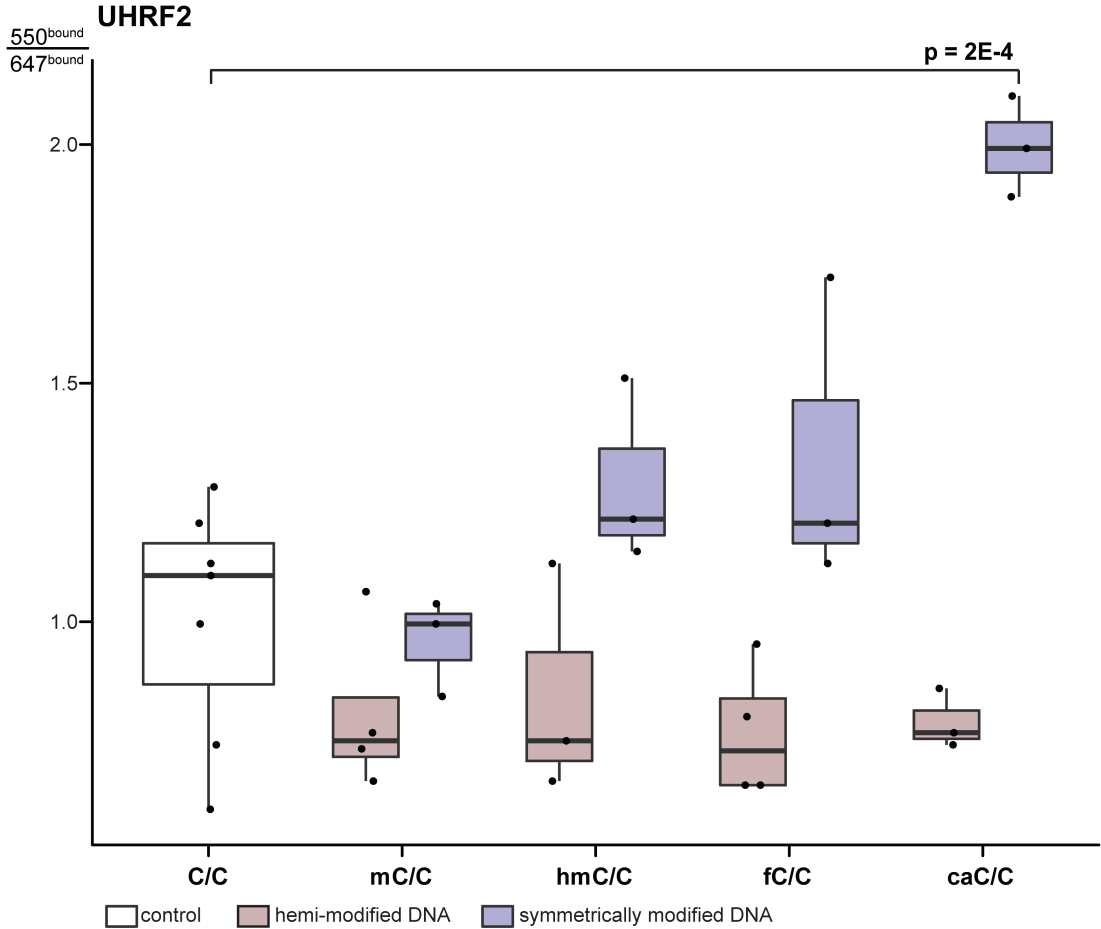
S3 Fig. Root Mean Squared Deviation (RMSD) of DNA atoms in molecular dynamics trajectories of UHRF1-SRA.

Coordinates were fitted to the initial crystal structure using the Ca atoms of protein residues 432 to 586. Only the last 1000 frames of each trajectory were used for analysis (vertical lines). Horizontal lines were added at 4 Å to highlight trajectories with strong structural distortions.



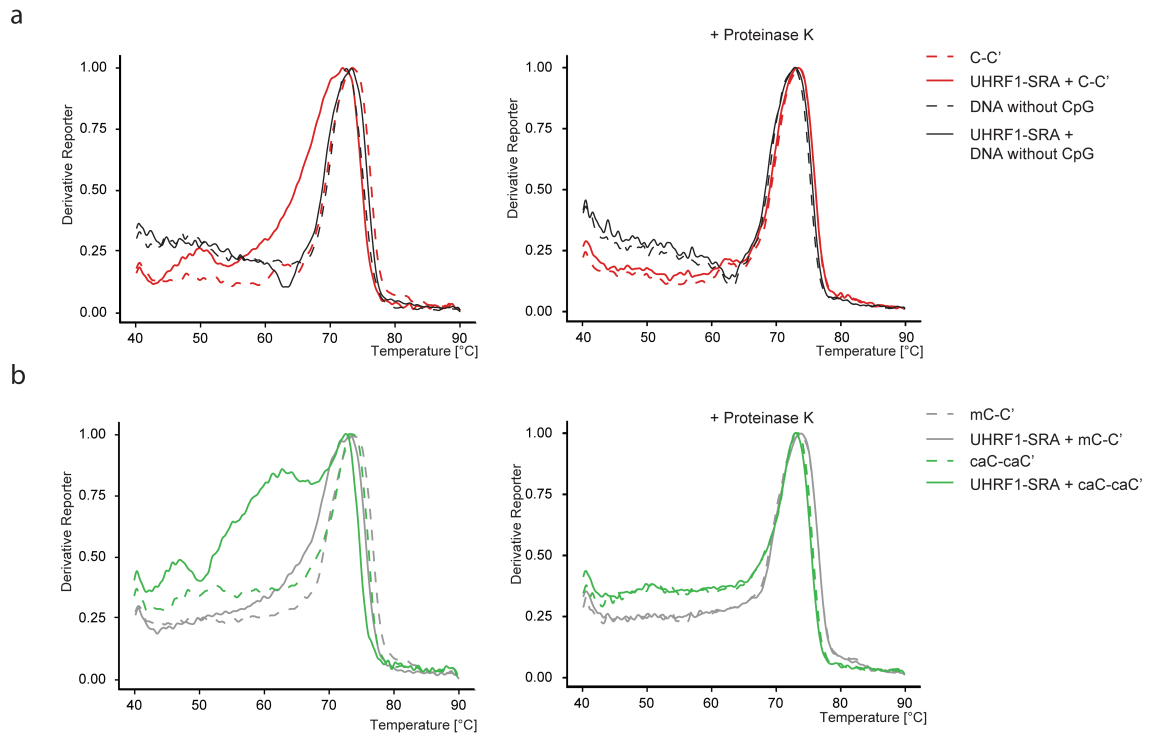
S4 Fig. Root Mean Squared Deviation (RMSD) of protein atoms in molecular dynamics trajectories of UHRF1-SRA.

Coordinates were fitted to the initial crystal structure using the Ca atoms of protein residues 432 to 586. Only the last 1000 frames of each trajectory were used for analysis (vertical lines). Horizontal lines were added at 4 Å to highlight trajectories with strong structural distortions.



S5 Fig. EMSAs of UHRF2 with differentially modified DNA.

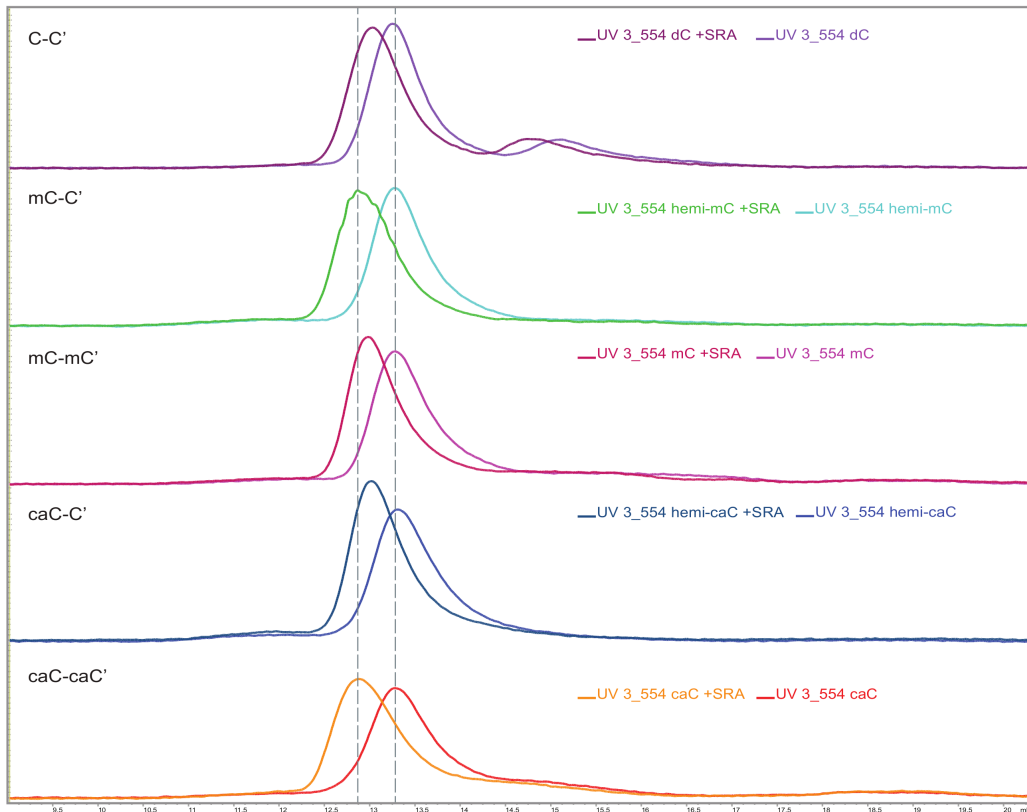
Quantitation of the bound fraction of EMSAs of wild type UHRF2-GFP with 42 bp DNA oligonucleotides carrying different cytosine modifications. Experiments and analyses have been performed as in Fig 2.



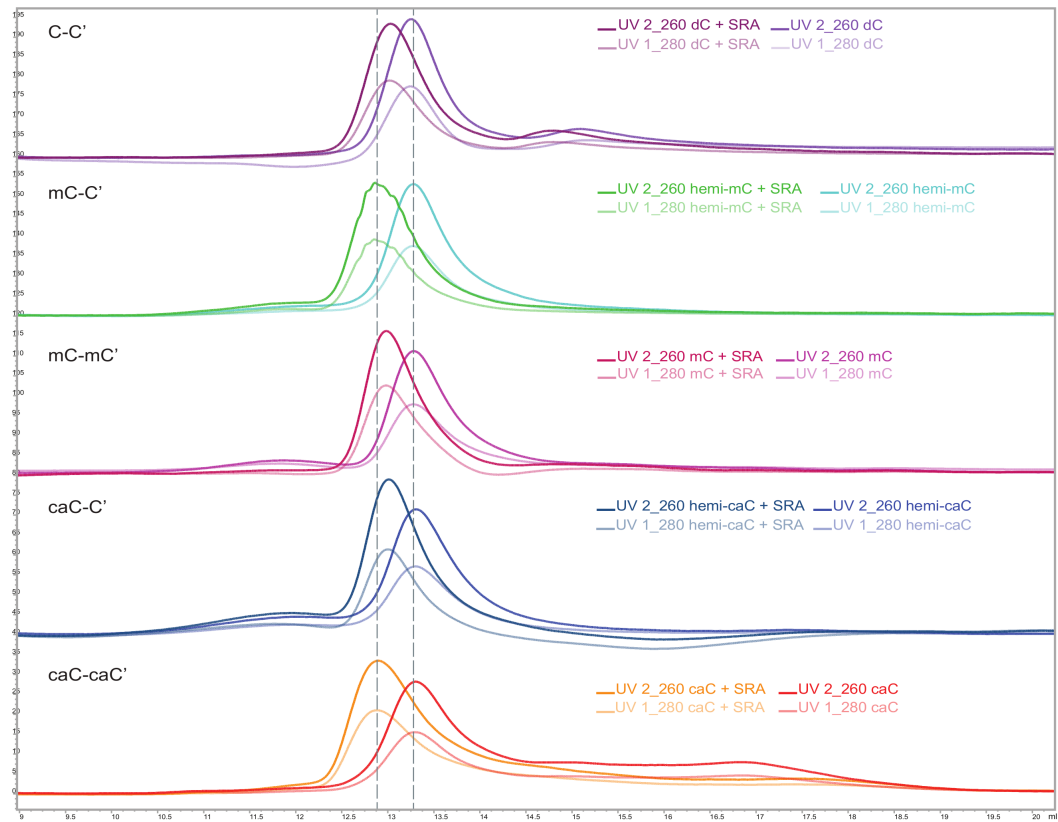
S6 Fig. Melting temperatures of modified DNA in presence of UHRF1-SRA.

(a) The melting temperature of double-stranded DNA containing C-C' in a CpG context (red) or no CpG site (black) with (solid lines) or without (dotted lines) a 5-fold excess of the SRA domain of UHRF1, measured using high resolution melting temperature (HRM) analysis. As control, proteins were digested by proteinase K before HRM analysis (right panel). Experiments were performed independently three times; one representative experiment is depicted as average of three technical replicates. (b) Melting temperatures as in (a) with DNA harbouring symmetric caC (green) or hemi-mC (gray) at the central CpG site.

a

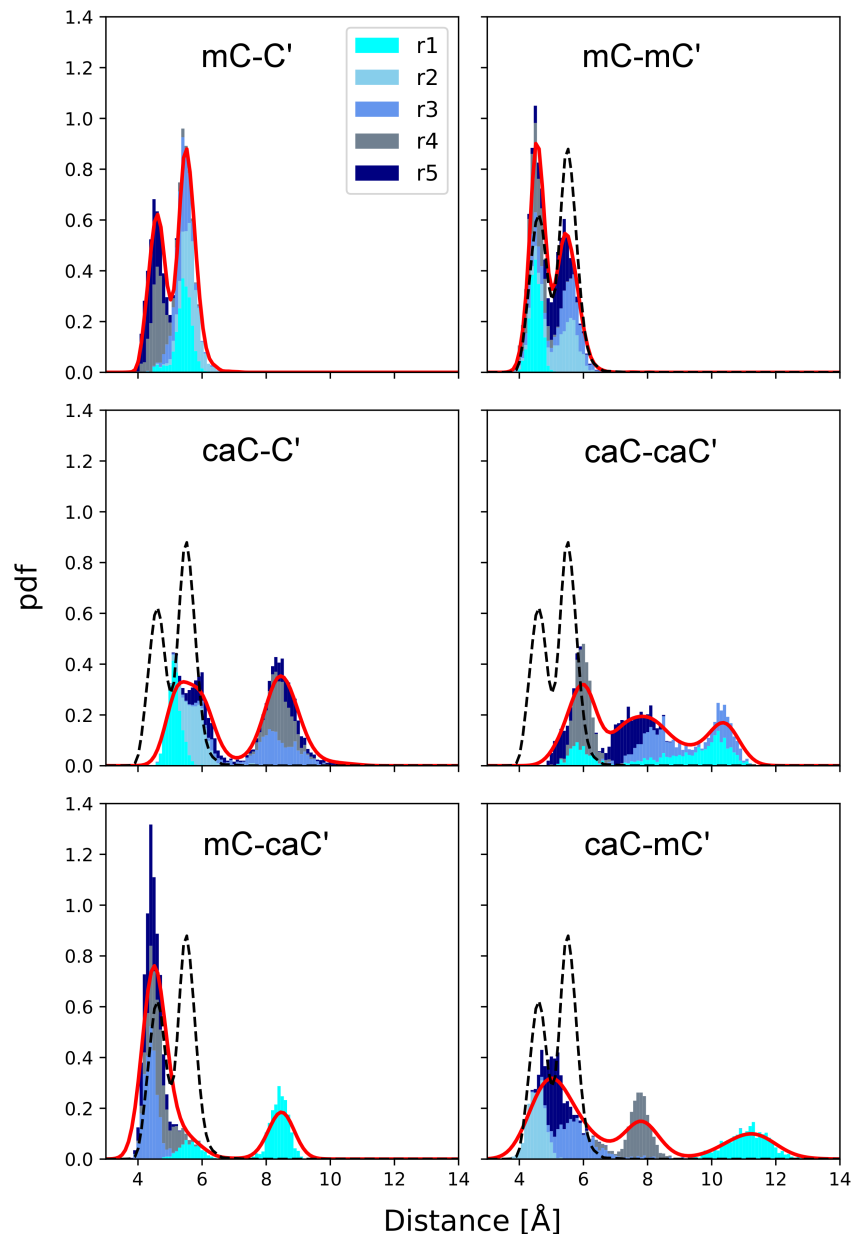


b



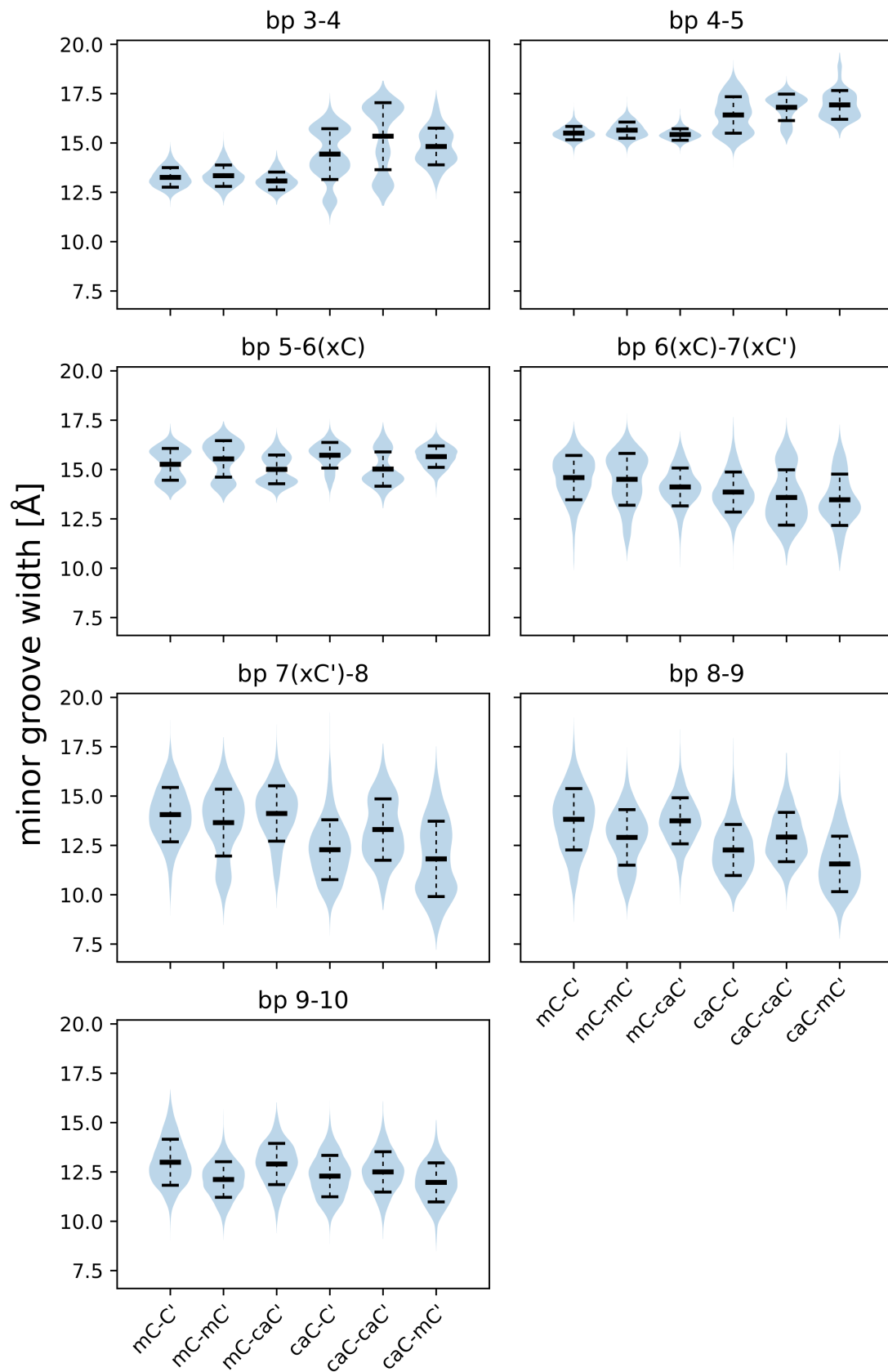
S7 Fig. Size exclusion chromatograms of differentially modified DNA in the presence or absence of UHRF1-SRA.

To test for different binding stoichiometries of the SRA domain towards differentially modified DNA, ATTO550-labeled DNA oligonucleotides were incubated with a 10-fold excess of SRA. Size exclusion chromatograms of analyzed DNA oligonucleotides at an absorbance of 554 nm (a) and 260nm/280nm (b) show a clear and comparable shift in retention time for the SRA-bound DNA (left peaks) compared to free DNA (right peaks).



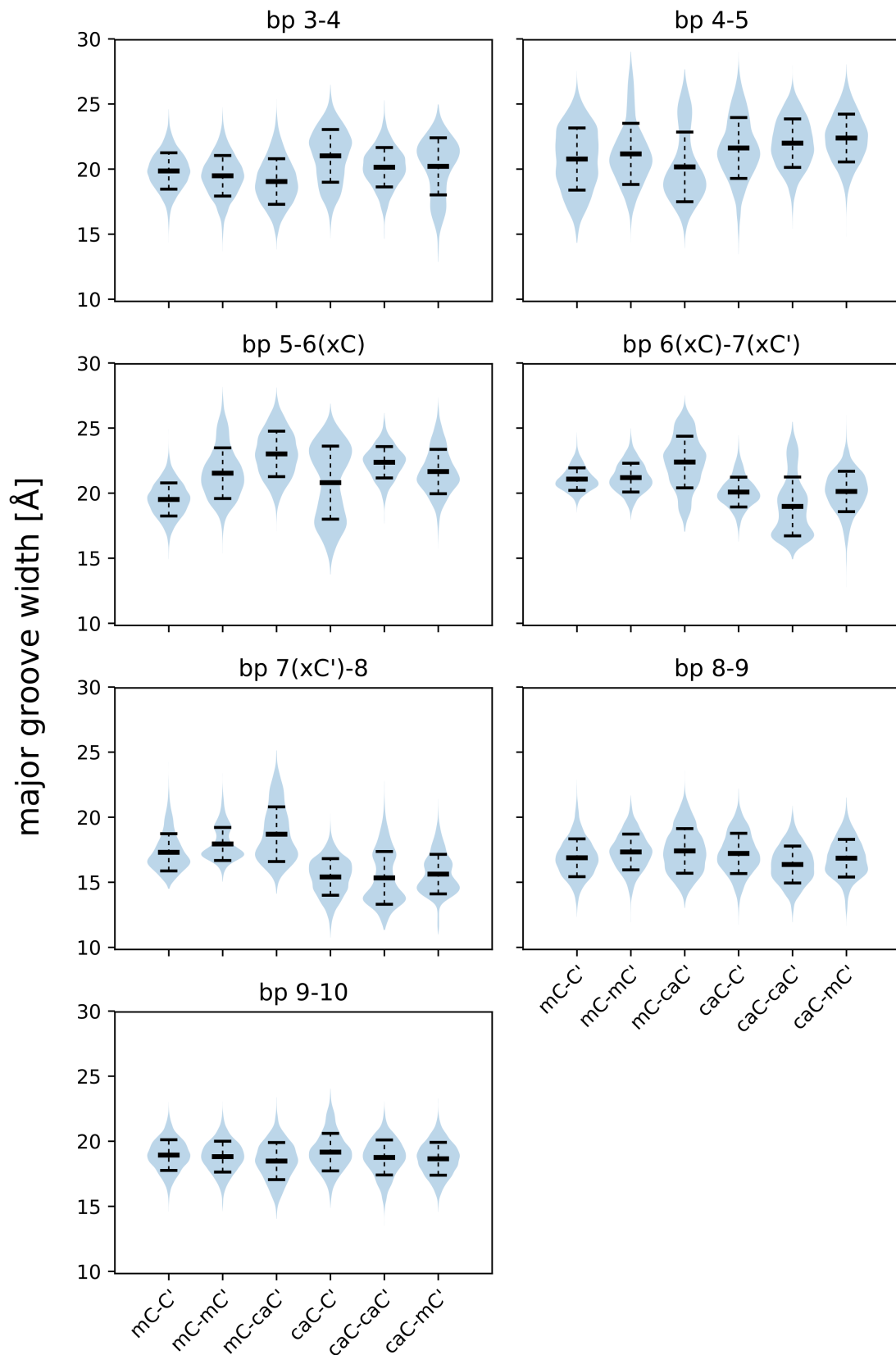
S8 Fig. Histograms of distances between Y471 and the flipped-out DNA base in molecular dynamics trajectories of UHRF1-SRA.

Individual replicas are shown as separate bars stacked on top of each other. Distances were measured between the geometric centres of the phenyl and pyrimidine rings. Red lines show a gaussian kernel estimate of the probability density function (pdf). The estimated pdf of the mC-C' system is shown as black dashed lines.



S9 Fig. Distribution of DNA minor groove widths in molecular dynamics trajectories of UHRF1-SRA.

Blue faces represent gaussian kernel estimates of the underlying values. Black bars show distribution means and standard deviations.



S10 Fig. Distribution of DNA major groove widths in molecular dynamics trajectories of UHRF1-SRA.

Blue faces represent gaussian kernel estimates of the underlying values. Black bars show distribution means and standard deviations.

1 **S1 Text. Additional Experimental Procedures.**

2 **Protein purification**

3 10 p150 dishes of HEK293T cells were transfected with the expression construct,
4 grown for 2 - 3 days in DMEM including 10 % FCS, and harvested by physical
5 detachment. Cells were lysed for 30 min on ice in lysis buffer (50 mM NaH₂PO₄, pH =
6 7.5, 150 mM NaCl, 10 mM imidazole, 0.5 % Tween-20, 0.5 mM EDTA, 1 g/l DNase I,
7 2 mM MgCl₂, 0.5 mM CaCl₂, 1 mM PMSF, 1x mammalian protease inhibitor (SERVA)).
8 Cleared lysates were incubated with NiNTA-coupled GFP-binder [1] for 1 - 2 hours at
9 4°C. After washing, proteins were eluted with 250 mM imidazole (in 50 mM NaH₂PO₄,
10 pH = 7.5, 300 mM NaCl, 0.05 % Tween-20). Buffers were exchanged to binding buffer
11 (20 mM TrisHCl, pH = 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT) via PD-10
12 columns (GE Healthcare Life Sciences). Protein quantification was performed based
13 on GFP intensities with an Infinite M1000 plate reader (Tecan).
14 For purification of the SRA domain from *E. coli*, expression cultures were grown at 37
15 °C and protein expression was induced with 0.5 mM IPTG when OD₆₀₀ reached 0.6-
16 0.8. After 3 hours, cells were lysed in binding buffer B (30 mM Tris-HCl pH 7.5, 300
17 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol) supplemented with 1 mM
18 PMSF, 100 µg/ml lysozyme and 25 µg/ml DNase under constant rotation. The lysate
19 was sonicated and centrifuged at 20,000 g for 20 min. Inclusion bodies were cleared
20 from cell debris by resuspending pelleted matter in wash buffer W (Buffer B + 0.5 %
21 Triton X 100) and subsequent centrifugation at 20,000 g for 20 min. All purification
22 steps were carried out on an Äkta Purifier system (GE Healthcare). Highly pure
23 inclusion bodies were dissolved in denaturation buffer D (Buffer B + 8 M Urea) and
24 loaded onto a HisTrap FF crude 1 mL column (GE Healthcare). Immobilized protein
25 was refolded on column by applying a linear gradient from 100 % buffer D to 100 %
26 Buffer B over 20 column volumes. Refolded soluble protein was eluted from the column

27 with a linear gradient from 100 % buffer B to 100 % elution buffer E (30 mM Tris-HCl
28 pH 7.5, 300 mM NaCl, 500 mM imidazole, 5 mM β -mercaptoethanol) over five column
29 volumes. Peak fractions were collected and analysed by SDS-PAGE for purity and
30 protein integrity.

31 **Melting Temperature Analysis**

32 To test the effect of SRA proteins on the stability of double-stranded DNA, we
33 incubated 2.5 μ M of the purified SRA domain, 0.5 μ M of the 42 bp double-stranded
34 DNA oligonucleotide and 1x reaction buffer (50 mM Tris-HCl, 0.5 mM β -ME, 10 mM
35 EDTA, pH 7.5) in a total volume of 10 μ L at 37°C for 30 minutes. Then, 10 μ L of SYBR
36 Green I Mix (Platinum SYBR Green qPCR Super Mix, 1:4 dilution, Invitrogen) was
37 added to each reaction and high-resolution melting temperature analysis was
38 performed in the StepOnePlus™ Real-Time PCR system (Applied Biosystems) by
39 increasing temperature from 40°C to 90°C with 0.1°C steps. As a control, SRA proteins
40 were digested with 20 μ g proteinase K (Sigma-Aldrich) at 50°C for 1 hour before adding
41 SYBR Green I Mix. To visualize the melting temperature of DNA, the derivative values
42 (SYBR Green I fluorescence against the temperature) were exported from StepOne
43 software 2.1 (Applied Biosystems) and further plotted using RStudio (0.98.1087).

44 **Analytic size exclusion chromatography**

45 To test for different binding stoichiometries of the SRA domain towards differentially
46 modified DNA, 26 μ M purified SRA domain in binding buffer (including 100 ng/ μ l BSA
47 final concentration) was mixed with 20 μ M of the corresponding 42bp DNA
48 oligonucleotide in a ratio of 10:1 and incubated on ice for 30 min. The formation of
49 complexes was assessed by size exclusion chromatography on an Aekta Pure system
50 equipped with a Superdex 200 Increase 10/300 GL column. Absorption at 260nm,
51 280nm and 554nm was monitored to detect DNA, protein and the fluorescence of the
52 oligos' ATTO550-label.









53 References

- 54 1. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H. A versatile nanotrap for
55 biochemical and functional studies with fluorescent fusion proteins. *Molecular & cellular proteomics : MCP*.
56 2008;7(2):282-9. doi: 10.1074/mcp.M700342-MCP200. PubMed PMID: 17951627.

57

2.2 Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation

Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation

Atsuya Nishiyama ^{1✉}, Christopher B. Mulholland², Sebastian Bultmann², Satomi Kori³, Akinori Endo⁴, Yasushi Saeki ⁴, Weihua Qin², Carina Trummer², Yoshie Chiba ¹, Haruka Yokoyama³, Soichiro Kumamoto ¹, Toru Kawakami ⁵, Hironobu Hojo⁵, Genta Nagae⁶, Hiroyuki Aburatani ⁶, Keiji Tanaka⁴, Kyohei Arita^{3✉}, Heinrich Leonhardt ^{2✉} & Makoto Nakanishi ^{1✉}

Stable inheritance of DNA methylation is critical for maintaining differentiated phenotypes in multicellular organisms. We have recently identified dual mono-ubiquitylation of histone H3 (H3Ub2) by UHRF1 as an essential mechanism to recruit DNMT1 to chromatin. Here, we show that PCNA-associated factor 15 (PAF15) undergoes UHRF1-dependent dual mono-ubiquitylation (PAF15Ub2) on chromatin in a DNA replication-coupled manner. This event will, in turn, recruit DNMT1. During early S-phase, UHRF1 preferentially ubiquitylates PAF15, whereas H3Ub2 predominates during late S-phase. H3Ub2 is enhanced under PAF15 compromised conditions, suggesting that H3Ub2 serves as a backup for PAF15Ub2. In mouse ES cells, loss of PAF15Ub2 results in DNA hypomethylation at early replicating domains. Together, our results suggest that there are two distinct mechanisms underlying replication timing-dependent recruitment of DNMT1 through PAF15Ub2 and H3Ub2, both of which are prerequisite for high fidelity DNA methylation inheritance.

¹Division of Cancer Cell Biology, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, Japan. ²Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Human Biology and Biomedicine, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany. ³Structure Biology Laboratory, Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, Japan. ⁴Laboratory of Protein Metabolism, Tokyo Metropolitan Institute of Medical Science, 2-1-6 Kamikitazawa, Setagaya-ku, Tokyo, Japan. ⁵Laboratory of Protein Organic Chemistry, Institute for Protein Research, Osaka University, Suita, Osaka, Japan. ⁶The Research Center for Advanced Science and Technology, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan. ✉email: anishiya@ims.u-tokyo.ac.jp; arita@yokohama-cu.ac.jp; h.leonhardt@lmu.de; mkt-naka@ims.u-tokyo.ac.jp

DNA cytosine methylation is a conserved epigenetic modification essential for embryonic development, transcriptional regulation, and genome stability¹. In higher eukaryotes, individual differentiated cells possess unique DNA methylation patterns that determine their cellular phenotypes. Therefore, the DNA methylation pattern must be precisely maintained in coordination with DNA replication during S phase². DNA methyltransferase 1 (DNMT1) contains multiple functional domains, including a replication foci targeting sequence (RFTS), an unmethylated CpG DNA-binding CXXC domain, two bromo-adjacent homology domains, and a C-terminal catalytic domain³. The RFTS domain is not only critical for DNMT1 recruitment to DNA methylation sites⁴ but also functions as an auto-inhibitory domain of DNMT1^{5,6}.

The recruitment of DNMT1 to DNA methylation sites requires UHRF1, an E3 ubiquitin ligase^{7,8}. UHRF1 recognizes specific epigenetic modifications on DNA strands and histone H3 tails through its SET- and RING-associated (SRA) domain and tandem Tudor domain (TTD)–plant homeodomain (PHD), respectively^{9–14}. The former binds to hemi-methylated DNA, while the latter recognizes N-terminal ¹ARTK⁴ residues and tri-methylated Lys9 of H3 (H3K9me3). The TTD domain also contributes to the interaction between UHRF1 and DNA ligase I^{15,16}. Furthermore, the E3 ubiquitin ligase activity of UHRF1 plays an essential role in DNMT1 recruitment to DNA methylation sites^{17,18}, and is enhanced by association with hemi-methylated DNA and H3K9me3^{19,20}. The UBL domain of UHRF1 also stimulates the E3 ligase activity of UHRF1 through its interaction with E2 ubiquitin-conjugating enzyme UbcH5a/UBE2D1^{21,22}. We and others have recently reported that UHRF1-mediated dual mono-ubiquitylation of histone H3 (H3Ub2) on lysine residues 14, 18, and 23 plays a role in the RFTS-dependent recruitment of DNMT1 and its enzymatic activation, ensuring the high fidelity of DNA maintenance methylation^{18,23,24}. DNMT1-bound USP7 also accumulates at DNA methylation sites^{25,26} and contributes to efficient DNA methylation by deubiquitylation of histone H3 and DNMT1^{25–28}.

However, the existence of two distinct modes of DNMT1 recruitment to hemi-methylation sites, one coupled with and the other uncoupled from DNA replication machinery, has previously been suggested by the finding that DNMT1 co-localizes with LIG1 foci in early and mid-S phase but not in late S phase²⁹. While H3Ub2 serves as one mark of DNMT1 recruitment, how this mark is coordinated with S-phase progression remains unknown.

In this report, we identify dual mono-ubiquitylation of PAF15 (PAF15Ub2) as a molecular mark coupling DNMT1 recruitment with DNA replication. During DNA replication, DNMT1 predominantly utilizes PAF15Ub2. When the PAF15-dependent mechanism is perturbed, DNMT1 utilizes H3Ub2, suggesting that H3Ub2 functions as a backup system for the maintenance of DNA methylation.

Results

Ubiquitylated PAF15 specifically binds replicating chromatin.

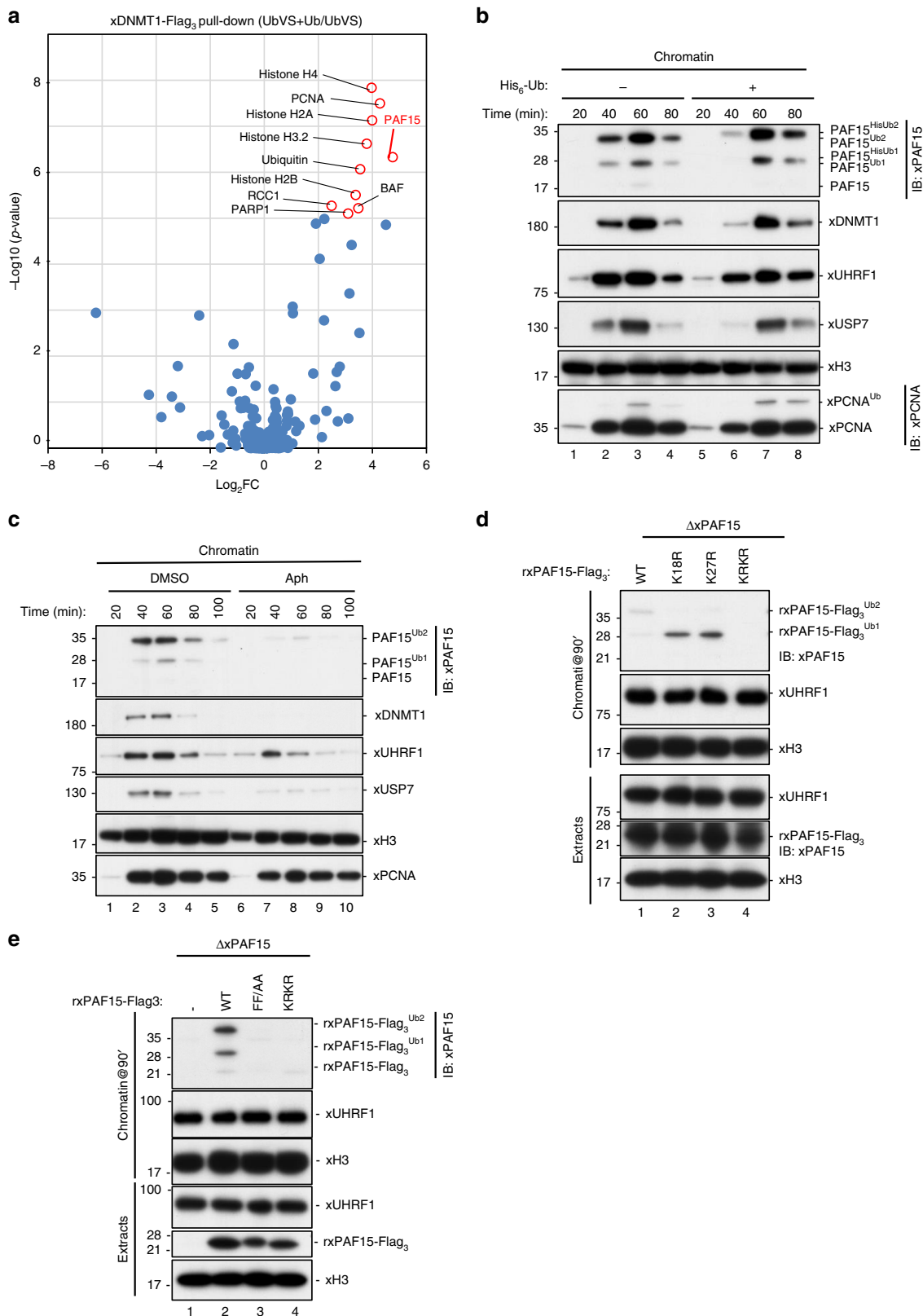
Given that the ubiquitin ligase activity of UHRF1 and the ubiquitin binding activity of DNMT1 are essential for the recruitment of DNMT1 to hemi-methylated DNA sites, we speculated that factors associated with the DNA replication machinery also utilize ubiquitin signals to recruit DNMT1. To identify factors capable of binding DNMT1 in a ubiquitin signal-dependent manner, we used ubiquitin vinyl sulfone (UbVS) treatment, a pan-deubiquitylation enzyme inhibitor³⁰, to specifically enrich for ubiquitylated proteins in cell-free *Xenopus* extracts. In brief, pretreatment of egg extracts with UbVS inhibits ubiquitin turnover and results in an almost complete

depletion of free ubiquitin, leading to the inhibition of both ubiquitylation and deubiquitylation pathways³¹. Thus, the addition of recombinant ubiquitin to UbVS-treated extracts specifically enhanced ubiquitin signals, including UHRF1-mediated histone H3 ubiquitylation^{23,25}. Chromatin lysates from UbVS-treated extracts in the presence (UbVS+Ub) or absence (UbVS) of free ubiquitin were subjected to a pull-down experiment using recombinant Flag-tagged wild-type *Xenopus* DNMT1 (rxDNMT1^{WT}) purified from insect cells (Supplementary Fig. 1a–c). As reported previously^{17,23}, rxDNMT1^{WT} specifically interacted with H3Ub2 in denatured chromatin lysates (Supplementary Fig. 1d, +sodium dodecyl sulfate (+SDS)). In native chromatin lysates, rxDNMT1^{WT} interacted with H3Ub2 as well as with unmodified and mono-ubiquitylated histone H3 (Supplementary Fig. 1d, –SDS), suggesting that indirect binding is also preserved under this condition.

We next subjected the pull-downs of rxDNMT1^{WT} or endogenous xDNMT1 from native chromatin lysates to mass spectrometric analysis. We identified 2840 unique peptides (including 26 ubiquitylated and 17 phosphorylated peptides), which mapped to 303 protein groups in chromatin lysates from UbVS-treated extracts in the presence (UbVS+Ub) or absence (UbVS) of free ubiquitin (Supplementary Data 1). Of these xDNMT1-interacting chromatin proteins, 24 were highly enriched in the xDNMT1 pull-downs in response to the addition of ubiquitin to UbVS-treated extracts (log₂fold-change >2, *p* value < 0.05; Fig. 1a, Supplementary Data 1). We also found an enrichment of eight ubiquitylated and two phosphorylated peptides in the data set (Supplementary Data 2 and 3). Histone H3 variants were identified, together with other histone proteins, validating the interactors (Fig. 1a and Supplementary Data 1). Among the identified proteins, we focused on PAF15, one of the most highly enriched proteins (log₂fold-change = 4.75), because it was reported to be associated with both proliferating cell nuclear antigen (PCNA) and DNMT1³², and was targeted for dual mono-ubiquitylation at its H3-like N-terminal sequence during S phase in human cells (see also Supplementary Fig. 1e)³³, suggesting that this interaction is conserved among vertebrates and is regulated in a ubiquitin signal-dependent manner.

The addition of sperm chromatin to cell-free *Xenopus* interphase egg extracts induces the assembly of replication-competent nuclei and a single round of DNA replication. Under these conditions, DNA replication typically begins approximately 40 min after sperm addition. After the completion of DNA replication and maintenance of DNA methylation, many chromosomal replication regulators including DNMT1 and UHRF1 dissociate from the chromatin^{17,34}. Using interphase egg extracts, we first examined the S-phase chromatin binding and ubiquitylation of xPAF15 along with the proteins involved in maintenance of DNA methylation. We found that slow migrating forms of xPAF15 bound to chromatin (Fig. 1b) in line with results using human cells³³. When an excess amount of recombinant His₆-tagged ubiquitin was added to the egg extracts (Supplementary Fig. 1f), the slowly migrating xPAF15 bands were further upshifted (Fig. 1b), suggesting that these slow forms correspond to ubiquitylated xPAF15 (Fig. 1b). The binding kinetics of xPAF15 were generally similar to those of xDNMT1 or xUSP7 over the same time course (Fig. 1b). The chromatin binding of xPAF15, as well as that of xDNMT1 and xUSP7, were lost in the presence of aphidicolin (Fig. 1c), a DNA polymerase inhibitor, suggesting that the chromatin binding of these proteins requires ongoing DNA synthesis.

In order to explore the role of xPAF15 ubiquitylation in xDNMT1 recruitment, we first identified ubiquitylation sites in xPAF15. We determined that the highly conserved lysine residues, K15 and K24 of human PAF15 (hPAF15), correspond to K18 and K27 of xPAF15 (Supplementary Fig. 1e)³³. Interphase egg extracts depleted of endogenous xPAF15 were supplemented with recombinant xPAF15-Flag₃ purified from insect cells, then sperm chromatin



was added. Wild-type recombinant xPAF15 (rxPAF15^{WT}), as well as the endogenous xPAF15, underwent ubiquitylation and bound to chromatin during DNA replication (Fig. 1d). In contrast, mutant xPAF15 with a substitution of lysine to arginine at both K18 and K27 (KRKR) failed to do so (Fig. 1d). Single xPAF15 mutants with the substitution at either site (K18R or K27R) underwent mono-

ubiquitylation and retained the chromatin-binding activity (Fig. 1d). Next, we examined how depletion of free ubiquitin from egg extracts affects xPAF15 chromatin binding. Pretreatment of egg extracts with UbVS completely suppressed the chromatin loading of both xPAF15 and xDNMT1, whereas xUHRF1 chromatin binding was maintained (Supplementary Fig. 1g, h). Addition of free

Fig. 1 Dual mono-ubiquitylated PAF15 (PAF15Ub2) specifically binds to replicating chromatin. **a** xDNMT1 pull-downs from native chromatin extracts were analyzed by LC-MS/MS. The volcano plot summarizes the quantitative results and highlights the interacting proteins enriched upon addition of ubiquitin to UbVS-treated extracts. **b** *Xenopus* interphase egg extracts were added with sperm chromatin and incubated in the absence or presence of His₆-ubiquitin (58 μM final). Chromatin-bound proteins were isolated and analyzed by immunoblotting using the indicated antibodies. For PAF15 levels in the extracts, see Supplementary Fig. 1f. **c** Interphase egg extracts were added with sperm chromatin and incubated in the presence of 15 μM aphidicolin (Aph) or in its absence (DMSO). Chromatin-bound proteins were isolated and analyzed by immunoblotting using the indicated antibodies. **d** PAF15-depleted extracts were supplemented with wild-type xPAF15-Flag₃ and its variants (K18R, K27R, and K18R/K27R). After the addition of sperm chromatin, chromatin-bound proteins were isolated and analyzed by immunoblotting using the indicated antibodies. The extracts were also analyzed by immunoblotting. **e** PAF15-depleted extracts were supplemented with wild-type xPAF15-Flag₃, its PIP-box mutant (FF/AA), or K18R/K27R mutant (KRKR). After the addition of sperm chromatin, chromatin-bound proteins were isolated and analyzed by immunoblotting using the indicated antibodies. The extracts were also analyzed by immunoblotting. Source data are provided as a Source Data file.

ubiquitin to UbVS-treated extracts efficiently restored PAF15 chromatin binding (Supplementary Fig. 1h). These results demonstrate that the mono-ubiquitylation of xPAF15 at K18 and/or K27 is important for stable xPAF15 chromatin association.

We then examined the role of xPCNA binding in xPAF15 chromatin loading. xPAF15 formed a stable complex with xPCNA in the egg extracts (Supplementary Fig. 1i) as it did in human cells^{32,33}. Substitution of phenylalanine with alanine at two conserved residues within the PCNA interacting peptide motif (PIP-box) of glutathione S-transferase (GST)-xPAF15 (FF/AA) abolished its interaction with xPCNA (Supplementary Fig. 1j). Although the WT rxPAF15 bound to the chromatin (Fig. 1e), the rxPAF15^{FF/AA} mutant failed to do so, as did the rxPAF15^{KRKR} mutant (Fig. 1e). These results suggest that xPAF15 chromatin loading requires interaction with xPCNA.

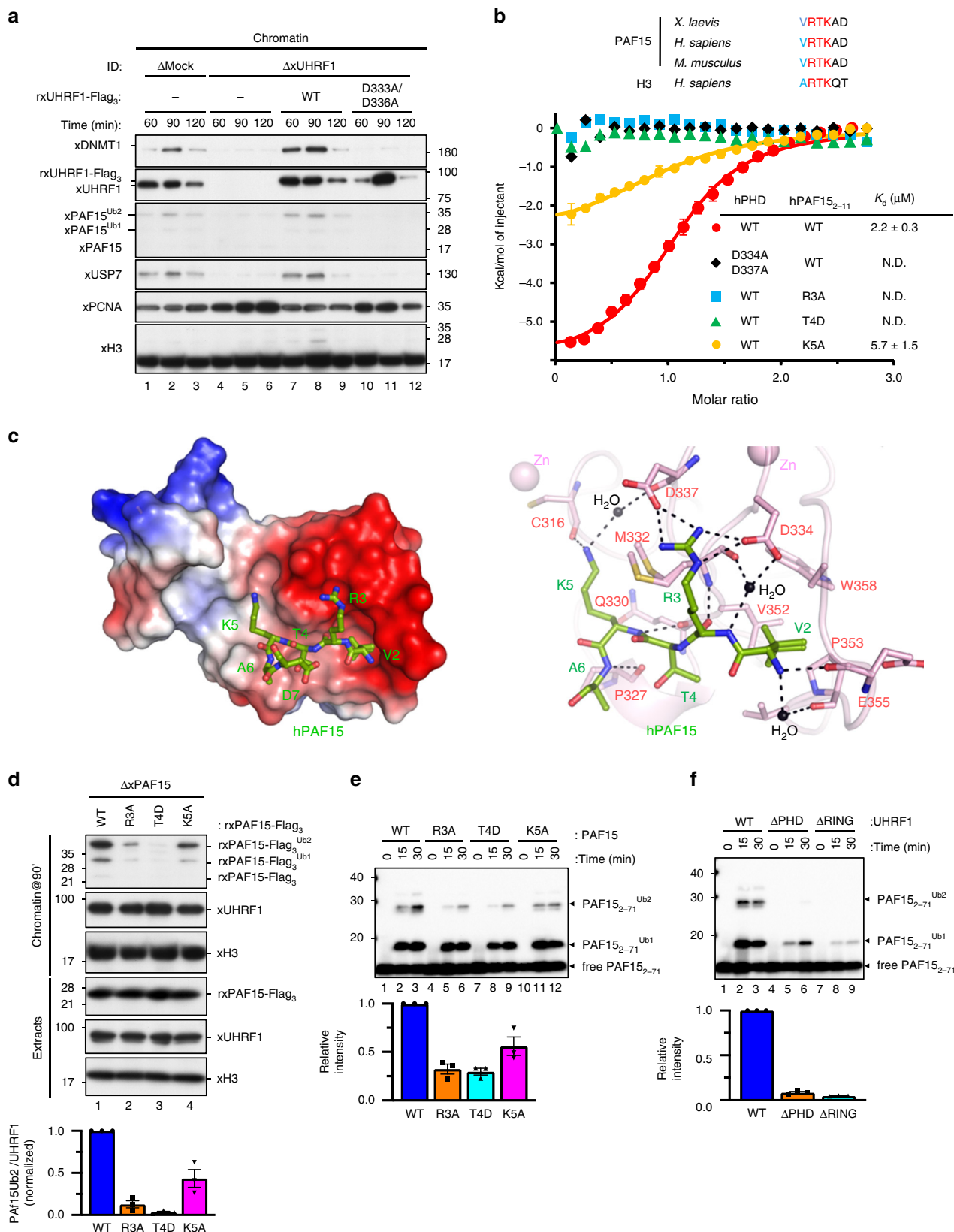
UHRF1 recognizes the N-terminal H3-like sequence of PAF15.

We next examined the requirement of the E3 ubiquitin ligase xUHRF1 for xPAF15 ubiquitylation and chromatin loading. As we demonstrated previously, immunodepletion of UHRF1 from egg extracts inhibited DNMT1 recruitment and chromatin association of xPAF15 (Fig. 2a, Supplementary Fig. 2a). Addition of recombinant WT xUHRF1 (rxUHRF1^{WT}) purified from insect cells to UHRF1-depleted extracts rescued the chromatin loading of xPAF15 (Fig. 2a, Supplementary Fig. 2a). We also tested the effect of recombinant xUHRF1 containing D333A/D336A, point mutations in the PHD finger that are expected to cause a loss of interaction with the histone H3 tail^{13,14}. Strikingly, rxUHRF1^{D333A/D336A} did not support xPAF15 ubiquitylation and chromatin loading (Fig. 2a, Supplementary Fig. 2a), suggesting that UHRF1-PHD has a crucial role in the regulation of PAF15. In contrast, xDNMT1 depletion resulted in the accumulation of xUHRF1 and ubiquitylation of histone H3 on the chromatin (Supplementary Fig. 2b, c). Dual mono-ubiquitylated xPAF15 (xPAF15Ub2) also accumulated on the chromatin (Supplementary Fig. 2b, c). These effects were restored by the addition of rxDNMT1 (Supplementary Fig. 2b, c). Our results indicate that both xPAF15 ubiquitylation and its chromatin recruitment are xUHRF1 dependent.

The PHD finger of hUHRF1 (hPHD) has been shown to bind to N-terminal ¹ARTK⁴ residues of histone H3^{13,14}. Given that the N-terminal portion of PAF15 shares significant homology with the N-terminal tail of histone H3 (Fig. 2b), we reasoned that hPHD likely to bind the N-terminal portion of PAF15. Isothermal titration calorimetry (ITC) demonstrated that hPHD bound to human PAF15₂₋₁₁ with $K_d = 2.2 \pm 0.3 \mu\text{M}$. This value is comparable to that for the N-terminal histone H3 peptide ($K_d = 1.5 \pm 0.1 \mu\text{M}$) (Fig. 2b, Supplementary Fig. 2d). In order to determine the binding mode of hPHD to PAF15, the crystal structure of hPHD bound to hPAF15₂₋₁₁ was determined at 1.7 Å resolution (Table 1). The structure showed that the hPAF15₂₋₁₁ peptide bound to the acidic surface of hPHD (Fig. 2c left), with the

²VRTK⁵ sequence of hPAF15₂₋₁₁ being recognized in a manner similar to that of the ¹ARTK⁴ of histone H3 (Supplementary Fig. 2e, f). The N-terminus of hPAF15₂₋₁₁ formed a hydrogen bond with hPHD-E355, and hPAF15₂₋₁₁-V2 was surrounded by the hydrophobic residues V352, P353, and W358 of the hPHD (Fig. 2c right). hPAF15₂₋₁₁-R3 and -K5 formed an electrostatic interaction with hPHD-D334 and -D337 and a hydrogen bond with hPHD-C316, respectively (Fig. 2c right). The importance of the above interactions in the complex formation was further validated by mutation analysis. ITC data demonstrated that hPHD-D334A/D337A failed to bind to WT hPAF15₂₋₁₁ while the WT hPHD was unable to bind to hPAF15₂₋₁₁-R3A or -T4D (Fig. 2b). Consistently, rxPAF15^{R3A} and rxPAF15^{T4D} failed to bind to chromatin during S phase in xPAF15-depleted extracts (Fig. 2d). Highlighting the importance of the hPHD recognition of hPAF15, *in vitro* ubiquitylation assays revealed that UHRF1 D334A/D337A and PAF15 R3A or T4D mutations significantly decreased the ubiquitylation of hPAF15, as well as the UHRF1 H741A mutation that disrupted E3 activity (Fig. 2e, f). The PAF15 K5A mutation had only a small effect on chromatin binding, interaction with hPHD, and ubiquitylation (Fig. 2b, d, e). Together, these findings suggest that the PHD finger of UHRF1 is responsible for the association with the N-terminal end of PAF15 with a binding mode similar to that of histone H3 and plays a critical role in PAF15 ubiquitylation by UHRF1.

PAF15Ub2 forms a complex with DNMT1. We recently reported that DNMT1 specifically binds to H3Ub2 via the RFTS domain²³. Given the similarity of PAF15 to the H3 tail and its ability to be dual mono-ubiquitylated, we asked whether PAF15 is also specifically recognized by the RFTS domain of DNMT1. Immunoprecipitation (IP)-western blotting analysis using solubilized chromatin revealed that the majority of xPAF15Ub2 bound to xDNMT1 (Fig. 3a, Supplementary Fig. 3a). Although xDNMT1 bound to H3Ub2, xPAF15Ub2 failed to do so (Fig. 3a), suggesting that xDNMT1-H3Ub2 and xDNMT1-xPAF15Ub2 complexes are mutually exclusive. Similar results were obtained using UbVS+Ub-treated egg extracts (Supplementary Fig. 3b). Next, we examined the DNMT1 binding of ubiquitylation-deficient xPAF15 mutants on chromatin. rxPAF15^{K18RK27R} failed to bind to chromatin as described above (Fig. 3b, see also Fig. 1d). Although rxPAF15^{K18R} or rxPAF15^{K27R} mutants bound to xPCNA on the chromatin as effectively as had rxPAF15^{WT}, they failed to bind to xDNMT1 (Fig. 3b). These results suggest that the binding of xPAF15 to xDNMT1 requires dual mono-ubiquitylation of PAF15 and that single mono-ubiquitylation of xPAF15 is not sufficient for the complex formation. This may also explain the apparently strong chromatin interaction of single mono-ubiquitylation of PAF15 K18R or K27R compared to dual mono-ubiquitylation of PAF15^{WT}, likely due to defective recruitment of DNMT1/USP7 complex.



To further analyze the interaction between the RFTS of human DNMT1 (hRFTS) and PAF15Ub₂, we prepared ubiquitylated hPAF15 (residues 2–30) analogs, in which G76C Ub was linked to K15C and/or K24C of hPAF15 by disulfide bonds (hPAF15₂₋₃₀Ub₂,

hPAF15K15ub, and hPAF15K24Ub, Supplementary Fig. 3c, see “Methods”). The ITC experiment using hRFTS and hPAF15₂₋₃₀Ub₂ was performed under a condition with higher *c* value ($c = n[\text{titrand}]/K_d$: 10,000) than that with an optimal value ($1 < c < 1000$)

Fig. 2 UHRF1 recognizes and ubiquitylates the N-terminal H3-like sequence of PAF15. **a** Mock-depleted or UHRF1-depleted extracts were supplemented with the indicated recombinant proteins (wt/D333A/D336A xUHRF1; see “Methods”) and chromatin was isolated. Chromatin-bound proteins were analyzed by immunoblotting using the indicated antibodies. For the protein levels of each protein in the extracts, see Supplementary Fig. 2a. **b** Comparison of the N-terminal sequence of PAF15 and histone H3 across different species. Residues mutated in the PAF15 mutants used in this study are shaded. Superimposition of plots of enthalpy changes in the interaction between hPHD and hPAF15₂₋₁₁ peptides by ITC measurement. **c** Recognition of the N-terminus of hPAF15 by hPHD. The left panel shows the crystal structure of PHD in complex with hPAF15. hPHD as a surface model with electrostatic potential (red, negative; blue, positive). The right panel shows recognition of PAF15 N-terminus (green stick model) by hPHD (pink stick model). Hydrogen bonds and water molecules are shown as black lines and balls, respectively. **d** PAF15-deleted extracts were supplemented with wild-type PAF15-Flag₃ and its variants (R3A, T4D, and K5A). After the addition of sperm chromatin, chromatin-bound proteins were isolated after 90 min and analyzed by immunoblotting using the indicated antibodies. The level of PAF15Ub₂ on chromatin was quantified for each set of conditions as explained in the “Methods” section. **e, f** In vitro ubiquitylation assay using the indicated hUHRF1 E3-ligases and hPAF15 substrates. Lower panels show the relative intensity of the band corresponding to dual mono-ubiquitylated PAF15. Bars represent the means of three independent experiments with SEM. Source data are provided as a Source Data file.

Table 1 Data collection and refinement statistics.

	PHD:PAF15 (PDB: 6IIW)
Data collection	
Beam line	PF-BL17A
Wavelength (Å)	0.98
Space group	P6 ₁ 22
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	36.7, 37.6, 220.2
Resolution (Å)	44.03–1.70 (1.73–1.70) ^a
<i>R</i> _{sym} or <i>R</i> _{merge} (%)	6.5 (49.7) ^a
<i>I</i> / σ (<i>I</i>)	27.5 (5.6) ^a
CC _{1/2}	99.9 (97.6) ^a
Completeness (%)	100 (99.9) ^a
Redundancy	17.2 (17.6) ^a
Unique reflections	10,762 (538)
Refinement	
Resolution (Å)	36.69–1.70
No. of reflections	10,653
<i>R</i> _{work} / <i>R</i> _{free} (%)	17.6/18.9
No. of atoms	
PHD	522
PAF15	51
Ion	4
Water	96
Average B factors (Å ²)	
PHD	26.6
PAF15	26.4
Ion	19.4
Water	38.4
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.928

^aValues in parentheses are for the highest-resolution shell.

because the measurement using lower concentrations of proteins (even 1/4 of the original) resulted in an insufficient calorimetric reaction for the reliable detection. Nevertheless, the results indicated that hRFTS binds to the hPAF15₂₋₃₀Ub₂ with high affinity ($K_d = 1.4 \pm 0.7$ nM) in a 1:1 stoichiometric complex, which is comparable to that of hRFTS bound to H3Ub₂²³. In contrast, the binding affinity of hRFTS to hPAF15₂₋₃₀K15Ub was much lower ($K_d = 1.2 \pm 0.8$ μ M) than that of hRFTS to hPAF15Ub₂. Interaction of hRFTS with hPAF15₂₋₃₀K24Ub resulted in a complex thermodynamic curve showing both exothermic and endothermic responses, which makes it difficult to determine its precise binding constant ($K_d =$ n.d.). In addition, stoichiometric binding of 1:1 was abrogated in hRFTS:PAF15Ub₁ at K15 or K24. These results indicate that dual mono-ubiquitylation of PAF15 is important for specific interaction with hRFTS. We then

performed size-exclusion chromatography in line with small-angle X-ray scattering (SEC-SAXS) of hRFTS, hRFTS-hPAF15₂₋₃₀Ub₂, or hRFTS-H3_{1-37W}Ub₂ (dual mono-ubiquitylated at K18 and K23; Supplementary Table 1, see “Methods”). The molecular weight estimation based on $I(0)/c$ (*c*: the concentration of protein) of Ovalbumin as a standard confirmed that the hRFTS-hPAF15₂₋₃₀Ub₂ or hRFTS-H3_{1-37W}Ub₂ formed the complex structure (Supplementary Fig. 3d, e and Supplementary Table 1). SAXS demonstrated that the radius of gyration (R_g), the shape of distance distribution function $P(r)$, and the maximum dimension D_{max} of hRFTS-hPAF15₂₋₃₀Ub₂ were almost identical to those of hRFTS-H3_{1-37W}Ub₂ (Supplementary Fig. 3f). We then confirmed whether xDNMT1 binds to xPAF15Ub₂ with a binding mode similar to that used for binding xH3Ub₂. Although rxDNMT1^{WT} bound to xPAF15Ub₂ in denatured UbVS/Ub-treated chromatin lysates, xDNMT1 mutants harboring substitutions of amino acids within the RFTS essential for xH3Ub₂ binding (P253AL256A or I317AI362A)²³ failed to do so (Fig. 3d), indicating that the interaction of xDNMT1 with xPAF15Ub₂ requires two mono-ubiquitin molecules that are conjugated on PAF15.

xPAF15Ub₂ is predominantly utilized for xDNMT1 recruitment. We next examined the role of xPAF15Ub₂ in the recruitment of xDNMT1 and subsequent maintenance of DNA methylation in *Xenopus* egg extracts. When xPAF15 was almost completely depleted from the extracts (Supplementary Fig. 4a), DNA replication-dependent DNA methylation of sperm DNA was partially suppressed compared to the control (Fig. 4a). Very importantly, although xH3Ub₂ was hardly detected in the mock-depleted chromatin in clear contrast to xPAF15Ub₂ (Fig. 4b, lanes 1–3), xH3Ub₂ was drastically enhanced when xPAF15 was depleted (Fig. 4b, lanes 4–6). Addition of rxPAF15^{WT} to the endogenous xPAF15-depleted extracts suppressed the enhanced xH3Ub₂, whereas that of rxPAF15^{K18R/K27R} failed to do so (Fig. 4b, Supplementary Fig. 4b), suggesting that depletion of xPAF15 was complemented by xH3Ub₂. Interestingly, the kinetics of xDNMT1 chromatin loading appeared to correlate with the dual mono-ubiquitylation of either xPAF15 or xH3. Consistent with this, DNMT1 predominantly interacted with PAF15Ub₂, not with H3Ub₂, on chromatin in mock-depleted extracts, whereas the level of H3Ub₂ in the DNMT1 complex significantly increased in the absence of PAF15 (Fig. 4c). Taken together, our results reveal an essential role for PAF15Ub₂ in maintenance of DNA methylation, which can only be partially compensated for by H3Ub₂.

Complementation of xDNMT1 recruitment by xH3Ub₂ in the PAF15-depleted extracts suggests that residual DNA methylation activity in these extracts originated from the enhanced xH3Ub₂. To address this, we aimed to suppress the xH3Ub₂-dependent pathway. As we previously observed that the deletion of the

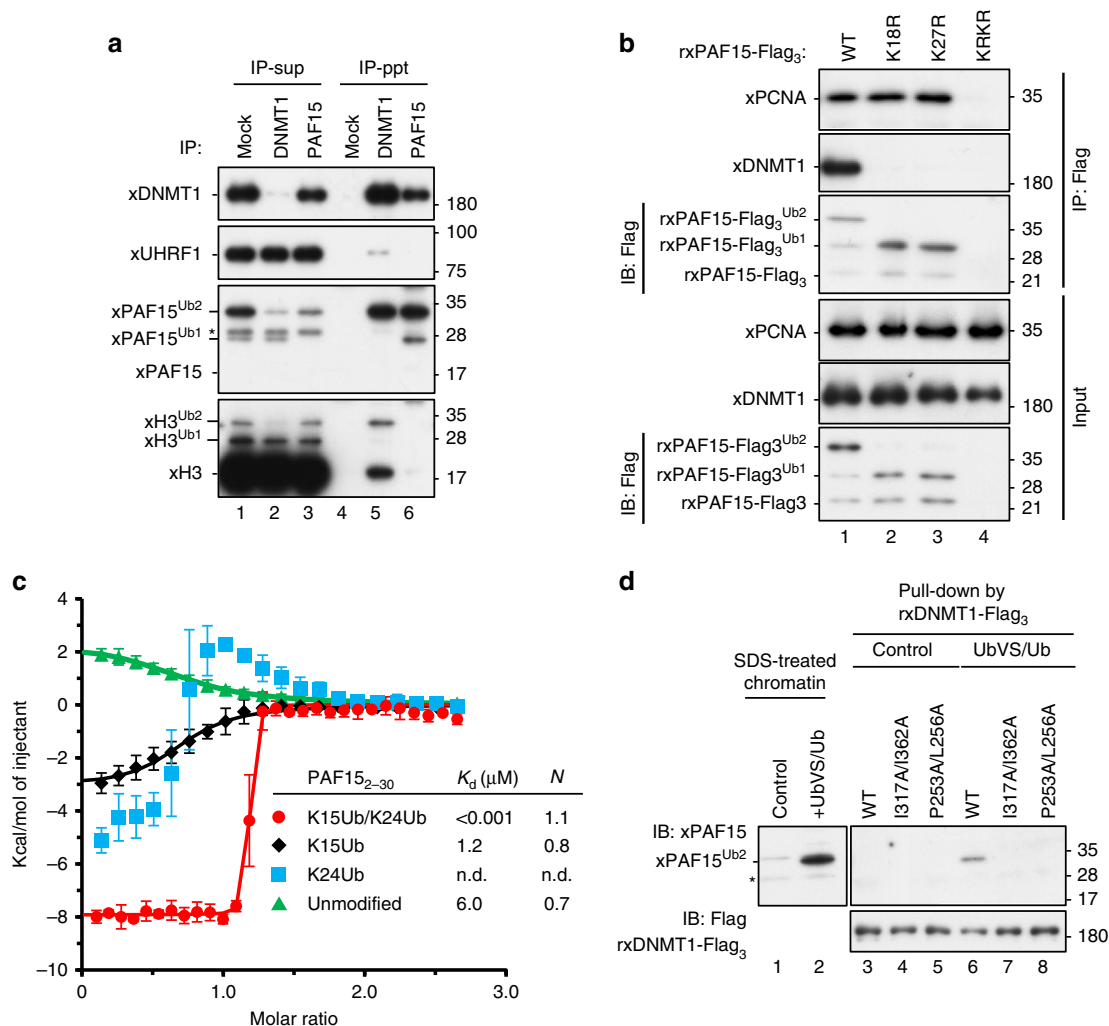


Fig. 3 PAF15Ub2 forms a complex with DNMT1. **a** Reciprocal immunoprecipitation of PAF15 and DNMT1 from chromatin lysates. IP was performed with control (Mock), anti-xDNMT1 (DNMT1), or anti-xPAF15 (PAF15) antibody from chromatin lysates. Supernatants after immunoprecipitation (IP-sup) or immunoprecipitates (IP-ppt) were analyzed by immunoblotting using the indicated antibodies. **b** Sperm chromatin was replicated in interphase egg extracts containing xPAF15-Flag₃ [wild-type, K18R, K27R, or K18RK27R (KRKR)]. Isolated and solubilized chromatin proteins were subjected to immunoprecipitation using anti-Flag antibodies. The resultant immunoprecipitates were analyzed by immunoblotting using the indicated antibodies. **c** Superimposition of plots of enthalpy changes in the interaction between hRFTS and hPAF15₂₋₃₀ or its ubiquitylated analogs by ITC measurement. **d** Pull-down of ubiquitylated PAF15 from denatured chromatin extracts using recombinant wild-type xDNMT1-Flag₃ and its ubiquitin-binding mutants (P253A/L256A or I317A/I362A). Source data are provided as a Source Data file.

C-terminal region of DNMT1 largely increased the binding of the RFTS domain to unmodified H3 and H3Ub2³⁵, it might be possible to preferentially suppress the xH3Ub2 pathway in extracts by supplying an optimal amount of recombinant hRFTS. We estimated that *Xenopus* interphase egg extracts contained ~0.1 μM of xDNMT1. Addition of 0.6 μM of hRFTS to the extracts resulted in the persistent presence of H3Ub2 on chromatin over the duration of S phase (Supplementary Fig. 4c), presumably due to suppression of xDNMT1 binding to xH3Ub2 and its deubiquitylation by xDNMT1-bound USP7²⁵. Under these conditions, xPAF15Ub2 was also upregulated, but the effect appeared to be transient and much weaker than in the case of xH3Ub2 (Supplementary Fig. 4c). These results suggest that the addition of an optimal amount of RFTS selectively inhibited the xH3Ub2 pathway in terms of its ability to recruit xDNMT1 (Supplementary Fig. 4c), apparently without affecting DNA replication and maintenance of DNA methylation. Importantly, the addition of 0.6 μM of hRFTS and the concomitant depletion of PAF15 dramatically reduced DNA methylation in the egg

extracts (Fig. 4d). Consistent with this, chromatin loading of both xPAF15Ub2 and xDNMT1 was readily detectable in extracts in the presence of hRFTS, whereas that of xDNMT1 was not when xPAF15 was depleted (Fig. 4e). Taken together, our results indicate that UHRF1 primarily ubiquitylates PAF15 during S phase, promoting DNMT1 recruitment and subsequent maintenance of DNA methylation in *Xenopus* egg extracts.

We next addressed whether UHRF1 differentially ubiquitylates PAF15 and histone H3 during S-phase progression. To this end, we depleted endogenous xUHRF1 from interphase extracts and supplemented the reaction mixture with rxUHRF1^{WT} at various time points after the addition of sperm chromatin. Subsequently, chromatin fractions were isolated at the indicated time points, and the ubiquitylation of both PAF15 and H3 on the chromatin was assessed (Fig. 4f). rxUHRF1^{WT} added to UHRF1-depleted extracts before the start of DNA replication (*t* = 0 min, Supplementary Fig. 4d, e) did not effectively ubiquitylate PAF15 or histone H3 (Fig. 4g, lanes 1–2). In contrast, when rxUHRF1 was added back in early S phase (*t* = 30 or 60 min, Fig. 4g, lanes 3–6, Supplementary

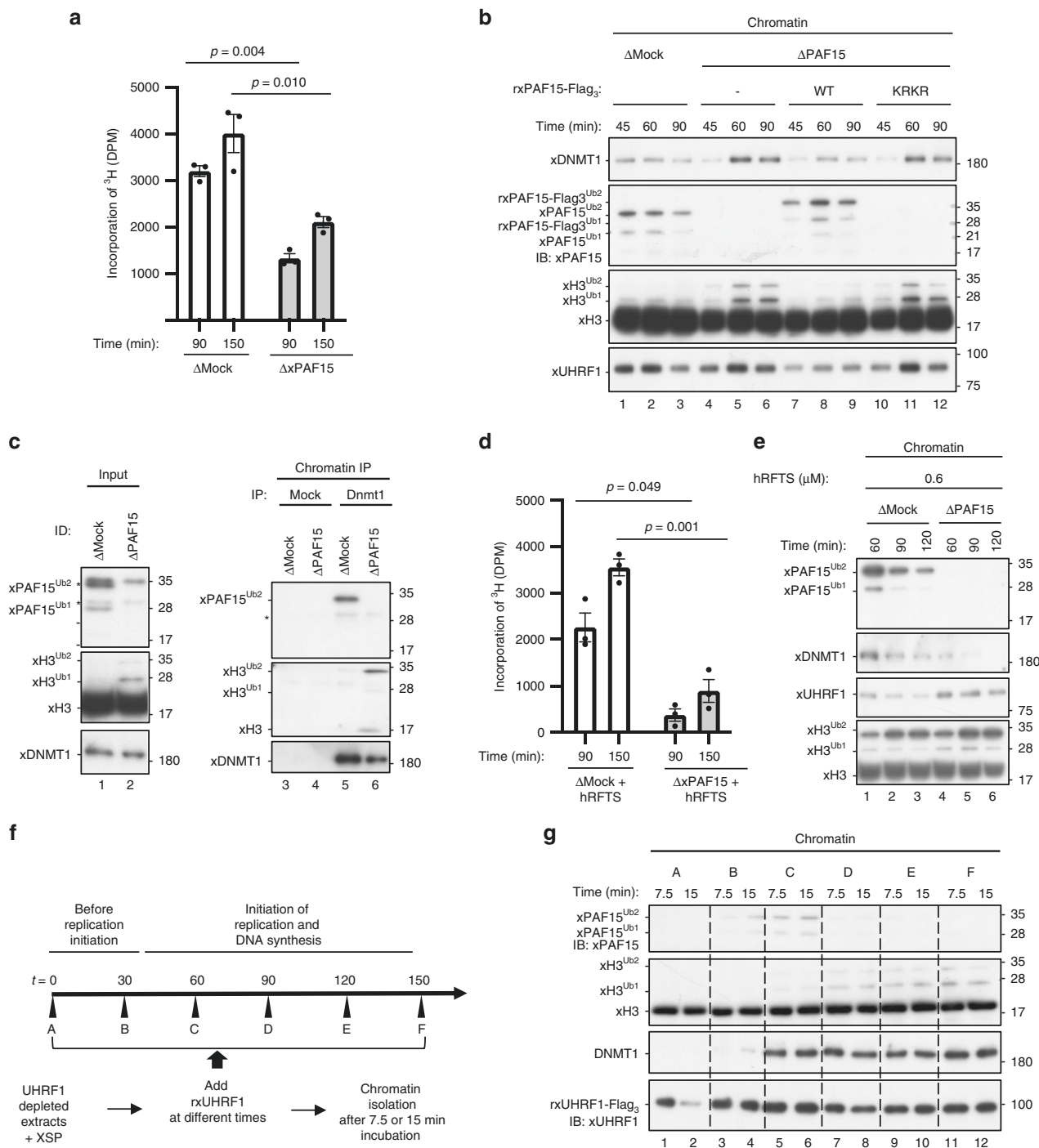


Fig. 4 xPAF15Ub2 promotes recruitment of xDNMT1 and maintenance of DNA methylation. **a, d** Sperm chromatin was added to either mock- or xPAF15-depleted extracts containing radiolabeled S-[methyl-³H]-adenosyl-L-methionine in the absence (**a**) or presence of 0.6 μM hRFTS (**d**). The efficiency of DNA methylation was measured at the time points indicated. Bar graphs depict the quantification of incorporated SAM into genomic DNA with mean and SEM from three independent experiments. Statistical significance was determined using Student’s *t* test. **b, e** Sperm chromatin was added to mock- or xPAF15-depleted interphase extracts in the absence (**b**) or presence (**e**) of hRFTS. PAF15-depleted extracts were supplemented with either buffer alone (lanes 4–6), purified wild-type xPAF15-Flag₃ or K18R/K27R(KRKR) mutant xPAF15-Flag₃ (320 nM final concentration, lanes 7–9 or 10–12, respectively) in the experiment described in **b**. At the indicated time points, chromatin fractions were isolated and subjected to immunoblotting using the antibodies indicated. For the PAF15 levels in extracts, see Supplementary Fig. 2a. **c** Sperm chromatin was replicated in mock- or PAF15-depleted interphase egg extracts. Isolated and solubilized chromatin proteins were subjected to immunoprecipitation using an anti-xDNMT1 antibody. The resultant immunoprecipitates were analyzed by immunoblotting using the indicated antibodies. Asterisks, non-specifically detected proteins. **f** Schematic of experimental approach to test the differential regulation through UHRF1 during the progression of S phase. **g** Sperm chromatin was added to xUHRF1-depleted extracts and incubated for 0, 30, 60, 90, 120, or 150 min. Extracts were then supplemented with recombinant xUHRF1-Flag₃ and further incubated for 7.5 or 15 min. Chromatin fractions were isolated and chromatin-bound proteins were analyzed by immunoblotting using the antibodies indicated. Source data are provided as a Source Data file.

Fig. 4d, e), it restored PAF15 ubiquitylation and chromatin recruitment. However, the addition of rUHRF1^{WT} at later time points ($t = 90\text{--}150$ min, lanes 7–12, Supplementary Fig. 4d, e) failed to restore PAF15 ubiquitylation and instead induced significant histone H3 ubiquitylation. Notably, unlike rUHRF1^{WT}, the addition of rUHRF1^{D333A/D336A} failed to induce histone H3 ubiquitylation under these conditions (Supplementary Fig. 4f). Consistent with the recruitment of DNMT1 via both PAF15 and histone H3 ubiquitylation, we found that DNMT1 loading was restored by the addition of rUHRF1 in both early and late S phase (Fig. 4g). UHRF1 therefore efficiently promotes PAF15 ubiquitylation during early S phase but prefers histone H3 as its substrate in late S phase for DNMT1 chromatin recruitment.

PAF15 is important for maintenance of DNA methylation in mouse embryonic stem cells (mESCs). We had previously shown that murine UHRF1 (mUHRF1) ubiquitylates two neighboring lysines at the N-terminus of mPAF15 (K15 and K24)³⁶ with a similar spacing as in histone H3, suggesting a similar role in the recruitment of DNMT1 and the maintenance of DNA methylation in murine cells. To investigate the interaction between murine DNMT1 (mDNMT1) and mPAF15 and the role of ubiquitylation, we used CRISPR/Cas9-based gene editing to introduce K15R, K24R, or both K15R/K24R (KRKR) mutations into the endogenous *Paf15* gene in mESCs (Supplementary Fig. 5a–g). Co-IP experiments with antibodies against mDNMT1 yielded a faint band of ~34 kDa corresponding to mPAF15Ub2 in WT ESCs, but no co-precipitation was detected with mPAF15 lacking either a single (K15R or K24R) or both ubiquitylation sites (KRKR) or in cells without mUHRF1 (U1KO) (Fig. 5a). As we could detect mPAF15 only in precipitates but not in the less concentrated input controls, we performed the reciprocal experiment. Upon IP with antibodies against mPAF15, we detected mPAF15Ub2 co-precipitating mDNMT1 in WT ESCs (Fig. 5b). In the KRKR mutant line, however, we detected a weaker band at ~15 kDa corresponding to the unmodified mPAF15, which did not co-precipitate mDNMT1 (Fig. 5b). The weaker signal obtained for mPAF15 KRKR was mostly due to losses during nuclear extract preparation as half of the unmodified mPAF15 was in the cytosol while the ubiquitylated mPAF15Ub2 was bound in the nucleus (Supplementary Fig. 6a). This diffuse distribution of the unmodified mPAF15 may in part be caused by reduced interactions with nuclear proteins. To be able to better compare the interaction and precipitation efficiency of WT mPAF15 and mutant mPAF15 KRKR, we titrated precipitates to comparable levels and could show that the modified WT mPAF15Ub2 clearly binds and precipitates mDNMT1 more efficiently (Supplementary Fig. 6b).

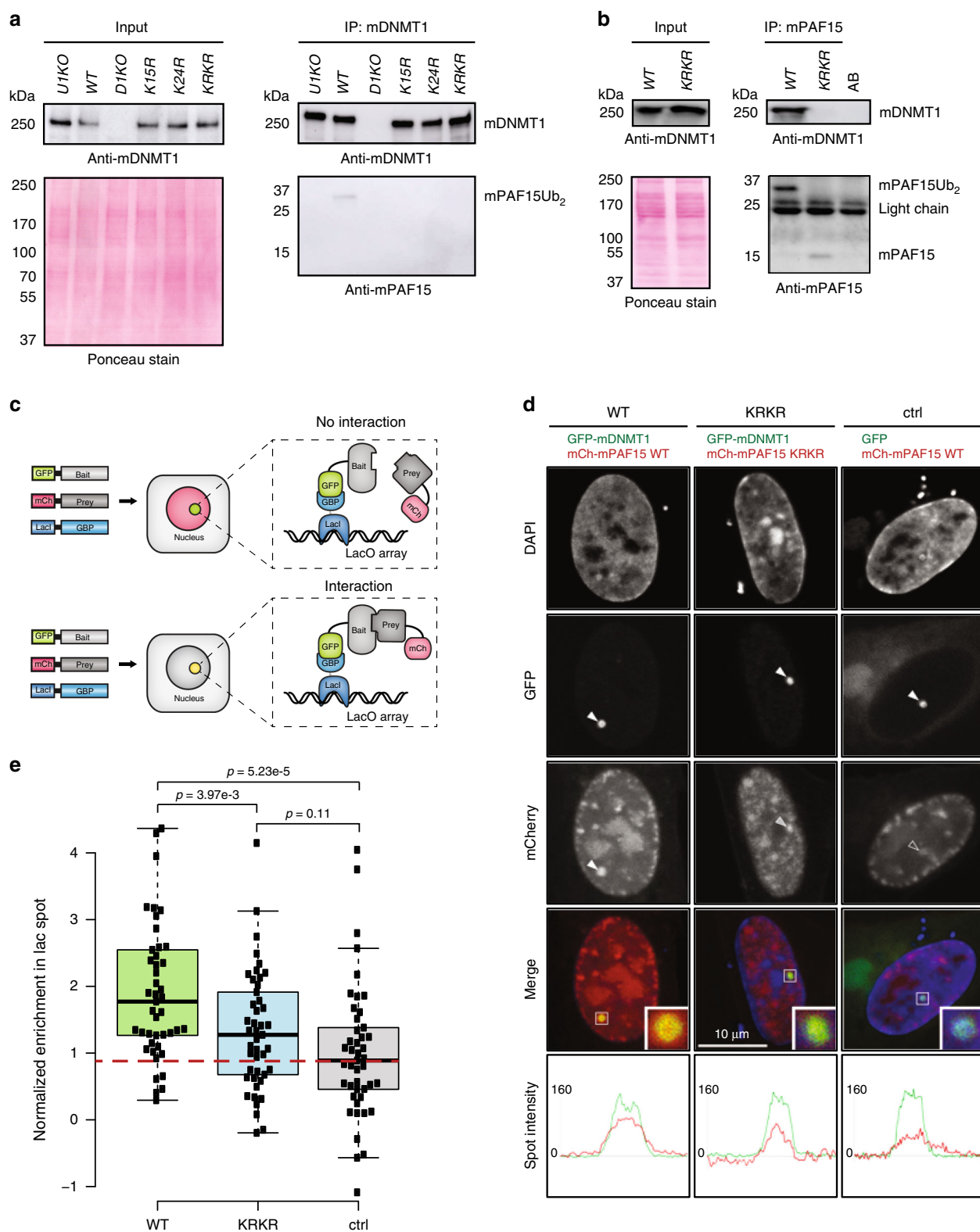
To validate the above findings with an independent approach, we applied a fluorescent-3-hybrid (F3H) assay³⁷ to assess the interaction of mDNMT1 and mPAF15 in vivo. In brief, the F3H assay provides a means of quantifying protein–protein interactions in living cells by measuring the efficiency with which a green fluorescent protein (GFP)-tagged “bait” protein immobilized at a nuclear spot (lacO array) is able to recruit a different, fluorescently labeled “prey” protein (schematic in Fig. 5c). For the negative control, cells were transfected with monomeric GFP and mCherry-tagged WT mPAF15. As expected, GFP was effectively immobilized at the lacO spot yet failed to efficiently recruit mCherry-tagged WT PAF15 (Fig. 5d, e). We next co-expressed GFP-mDNMT1 in addition to either mCherry-tagged WT or the ubiquitylation-deficient mPAF15 (KRKR) harboring both K15R and K24R substitutions. In contrast to monomeric GFP, immobilized GFP-mDNMT1 readily recruited a significant fraction of WT mPAF15 to the lacO spot, but not mPAF15-KRKR (Fig. 5d, e). These results

demonstrate that mammalian DNMT1 and PAF15 interact in vivo and show this interaction to be dependent, as in *Xenopus* (Fig. 3b), on dual mono-ubiquitylation of PAF15.

We next sought to determine whether the endogenous interaction of mDNMT1 and mPAF15Ub2 serves a role in the maintenance of DNA methylation in mESCs. To this end, we first used high-content immunofluorescence-based detection of 5-methylcytosine (5-mC) to measure DNA methylation levels in our *Paf15* mutant ESC lines as well as in control cell lines, *Uhrf1* knockout (*Uhrf1* KO) and *Dnmt1* knockout (*Dnmt1* KO) ESCs, lacking maintenance of DNA methylation activity (Supplementary Fig. 6c). As expected, *Dnmt1* KO and *Uhrf1* KO ESCs exhibited a near complete loss of DNA methylation (Supplementary Fig. 6c). Strikingly, *Paf15* single (K15R and K24R) and double mutant (KRKR) ESCs also displayed a substantial global reduction in DNA methylation when compared with WT ESCs (Supplementary Fig. 6c). However, global DNA methylation levels in *Uhrf1* KO ESCs, in which both mH3Ub2 and mPAF15Ub2 are absent, were lower than those in *Paf15* mutant ESCs, suggesting that the H3Ub2 pathway can partially compensate for the loss of mPAF15Ub2 (Supplementary Fig. 6c). Taken together, these results revealed that mPAF15Ub2 has an essential and largely non-redundant role in ensuring proper maintenance of DNA methylation in mESCs.

To assess how mPAF15Ub2 shapes the methylome of mESCs at single-nucleotide resolution, we performed reduced representation bisulfite sequencing (RRBS) on WT and *Paf15* KRKR mESCs (Supplementary Table 2). Consistent with our immunofluorescence measurements, RRBS analysis revealed a significant loss of global DNA methylation in *Paf15* KRKR ESCs compared to WT ESCs (Fig. 6a, c, Supplementary Fig. 6d). Furthermore, we observed a significant decrease in DNA methylation levels at all genomic regions examined, including repetitive elements, gene bodies, promoters, and CpG islands in *Paf15* KRKR mESCs (Fig. 6b). To determine whether mPAF15Ub2-dependent methylation is associated with particular chromatin features, we analyzed the levels of several histone modifications (H3K9me2³⁸, H3K9me3³⁹, and H3K14ac⁴⁰) at regions differentially methylated in *Paf15* KRKR ESCs ($p < 0.05$ and methylation difference $> 25\%$). However, we found neither active nor repressive histone modifications to be enriched at hypomethylated regions resulting from mPAF15Ub2 loss (Supplementary Fig. 6f). We then analyzed how the loss of mPAF15Ub2 affects DNA methylation levels of lamina-associated, late-replicating regions found to be hypomethylated in a multitude of cancer types⁴¹. These hypomethylated regions, referred to as partially methylated domains (PMDs), differ from the heavily methylated domains (HMDs) comprising the bulk of the remaining genome⁴². Our RRBS analysis demonstrated a stark reduction in DNA methylation at both PMDs and HMDs in *Paf15* KRKR ESCs (Supplementary Fig. 6e), suggesting that mPAF15Ub2 contributes to the maintenance of DNA methylation at both PMDs and HMDs in mESCs.

To investigate the relationship between replication timing and mPAF15Ub2-dependent maintenance of methylation, we compared our *Paf15* KRKR methylome data with Repli-seq maps from ESCs⁴³. Remarkably, regions hypomethylated in *Paf15* KRKR ESCs were associated with a significantly earlier replication timing than regions of unchanged DNA methylation, which on average tended to replicate later in S phase (Fig. 6d). These results indicate that mPAF15Ub2 has an essential role in the maintenance of DNA methylation, especially at early replicating sequences, and imply that mH3Ub2 is sufficient to sustain DNA methylation at late replicating regions in the absence of mPAF15Ub2. In contrast to *Paf15* KRKR ESCs, the average replication timing of hypomethylated regions in *Uhrf1* KO and *Dnmt1* KO ESCs was essentially identical to that of regions of



unchanged DNA methylation (Fig. 6e, f). These results indicate that the complete disruption of maintenance of DNA methylation by genetic ablation of DNMT1 or UHRF1, which abolishes both mPAF15Ub₂ and mH3Ub₂, leads to genome-wide hypomethylation irrespective of replication timing. Together, these data show that mPAF15Ub₂ and mH3Ub₂ constitute two distinct pathways of mDNMT1 recruitment that together accomplish complete

maintenance of DNA methylation throughout every cell cycle (Fig. 6g).

Discussion

Our current study provides clear evidence that PAF15 within DNA replication machinery complexes plays a pivotal role in the maintenance of DNA methylation. We have recently reported that

Fig. 5 Dual mono-ubiquitylation of mPAF15 is required for the mPAF15–mDNMT1 interaction in mouse ESCs. **a** Immunoprecipitation of endogenous DNMT1 from whole-cell lysates of wild-type J1 (WT), *Dnmt1* KO (D1KO), *Uhrf1* KO (U1KO), *Paf15* K15R (K15R), *Paf15* K24R (K24R), and *Paf15* K15/24R (KRKR) mESCs using an anti-mDNMT1 nanobody. Bound fractions were subjected to immunoblotting with anti-mDNMT1 and anti-mPAF15 antibodies. The anti-mDNMT1 blot and Ponceau staining are shown as loading controls. **b** Immunoprecipitation of endogenous mPAF15 from WT and KRKR mESC nuclear extracts using an anti-mPAF15 antibody. Bound fractions were subjected to immunoblotting with anti-mDNMT1 and anti-mPAF15 antibodies. The anti-mDNMT1 blot and Ponceau staining are shown as loading controls. **c** Schematic of the fluorescent-3-hybrid (F3H) assay for the in vivo determination of protein–protein interactions. GFP-tagged bait protein is immobilized at an array of Lac operator (LacO) sequences by a GFP-binding protein (GBP) coupled to the lac repressor (Lacl). When the GFP-tagged bait protein does not interact with the prey protein, only a GFP signal is visible at the LacO locus, whereas a yellow spot (combination of GFP and mCherry signal) is visible at the LacO locus in the case of a positive interaction. **d, e** F3H assay for a BHK cell-based analysis of ubiquitylation-mediated recruitment of mPAF15 to mDNMT1. **d** Cells containing a stably integrated lacO array were transfected with the GBP–Lacl, a GFP-tagged bait (GFP–mDNMT1 or GFP), and an mCherry-tagged prey (mCherry–mPAF15 wild-type (WT) or mCherry–mPAF15 K15R/K24R double mutant (KRKR)). Line intensity profiles for GFP and mCherry in the respective spots are shown below the confocal images. Scale bar, 10 μ m. **e** Quantification of the F3H assay. Background subtracted mCherry/GFP ratios within the spots were normalized to the control and plotted with $n = 45$ from 3 independent replicates (per replicate, $n = 15$). In the boxplots, horizontal black lines within boxes represent median values, boxes indicate the upper and lower quartiles, and whiskers indicate the 1.5 \times interquartile range. Statistical significance was determined using Student's *t* test. Source data are provided as a Source Data file.

UHRF1-mediated H3Ub2 recruits DNMT1 to DNA methylation sites, which likely functions independently of DNA replication fork progression. PAF15 in a complex with PCNA also undergoes UHRF1-mediated dual mono-ubiquitylation, which is essential for DNMT1 recruitment, and subsequent maintenance of DNA methylation. Thus, our results suggest that dual mono-ubiquitylation at two lysine residues spaced by 4–9 amino acids (mH3 K14~K18~K23 and mPAF15 K15~K24) in the flexible region of the proteins serves as a specific code for the maintenance of DNA methylation. This notion is supported by the finding that the recognition of PAF15Ub2 by RFTS was very similar to that of H3Ub2.

The fact that UHRF1 targets two distinct proteins, histone H3 and PAF15, for generating a specific code is consistent with the previous report that there are two modes of maintenance of DNA methylation²⁹. As to why UHRF1 would have two modes of usage, our results strongly suggest that PAF15Ub2 and H3Ub2 function in different contexts depending on the replication timing, as PAF15 ubiquitylation occurs only during early S phase, whereas histone H3 ubiquitylation can be induced in late S phase. It has been reported that DNA methylation levels are different between early and late replicating domains, with the former containing a much higher degree of DNA methylation than the latter⁴⁴. The enrichment of DNA methylation sites in early replicating domains would explain why cells expressing mPAF15-K15R/K24R have a substantial loss of DNA methylation as observed in mESCs, as well as in *Xenopus* egg extracts. Although H3Ub2 is markedly increased when PAF15Ub2 is perturbed, it might occur with less efficiency in early replicating domains.

H3Ub2 might also play a dominant role in the recruitment of DNMT1 under particular conditions in which PAF15 is not functional. For example, the replication block induced by ultraviolet (UV) irradiation leads to PAF15 poly-ubiquitylation and subsequent proteasomal degradation³³. Therefore, replication fork stalling across heterochromatin at late replicating domains might induce PAF15 degradation⁴⁵, which might then be compensated for by H3 ubiquitylation. Alternatively, H3 ubiquitylation could function as a proofreader for the failure of DNA methylation by PAF15Ub2-dependent DNMT1 recruitment. Consistent with this idea, the level of xH3Ub2 on chromatin as well as in complex with xDNMT1 increased upon xPAF15 depletion and the masking of H3Ub2 by RFTS in the absence of PAF15 resulted in an almost complete loss of DNA methylation. Whereas deletion of *Dnmt1* or *Uhrf1* causes embryonic lethality^{7,8,46}, it is noteworthy that *Paf15* knockout mice remain viable despite abnormal hematopoietic stem cell function⁴⁷. These observations suggest that loss of PAF15 function in the

recruitment of DNMT1 could partly be compensated for by histone H3, ensuring the stable inheritance of DNA methylation.

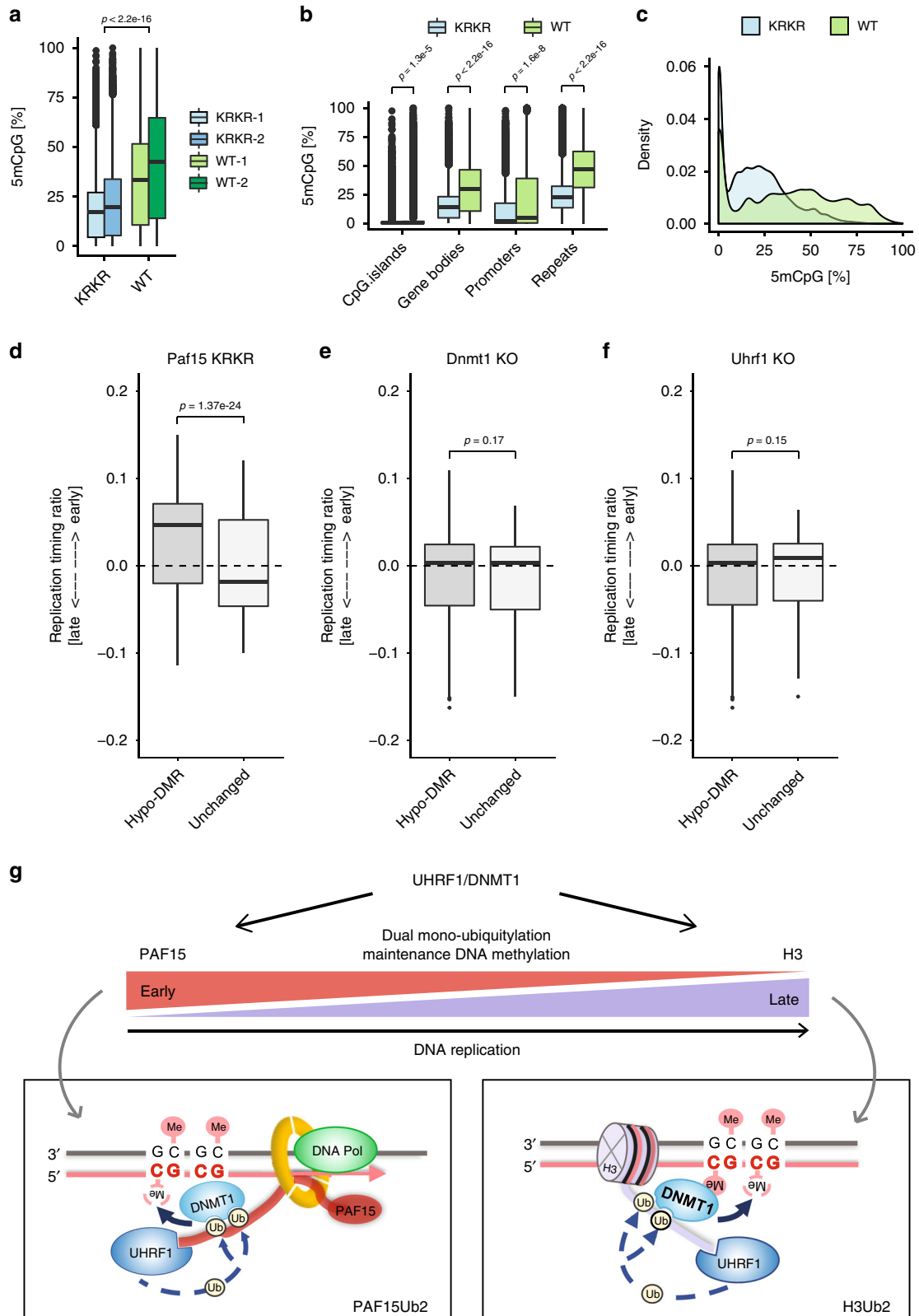
We found that both the interaction with PCNA and dual mono-ubiquitylation by UHRF1 are essential for PAF15 function in the maintenance of DNA methylation. PAF15 is an intrinsically disordered protein and binds to trimeric PCNA via the PIP-box motif at the front face and its N-terminus interacts with the inner ring of PCNA and exits the clamp from the back face⁴⁸, suggesting that the ubiquitylation sites of PAF15 could locate near the nascent strand where a methyl group does not yet exist. Thus, PAF15Ub2 could directly recruit DNMT1 to the back face of PCNA, facilitating the processivity of DNMT1-mediated DNA methylation on the nascent DNA (Fig. 6g). These structural features are also consistent with the fact that early replicating domains contain a much higher degree of DNA methylation at which time PAF15Ub2 is predominantly recruiting DNMT1. We also note that, during the revision of our manuscript, a recent study has also shown that full-length hPAF15Ub2 binds DNMT1 in vitro⁴⁹.

In conclusion, we propose that maintenance of DNA methylation is coordinated with S-phase progression via UHRF1-dependent dual mono-ubiquitylation of two distinct proteins, PAF15 and histone H3, which may contribute to the robustness of DNA maintenance methylation by ensuring the recruitment and activation of DNMT1 (Fig. 6g). Further research is required to clarify how the different modes of DNMT1 recruitment are chosen and to identify potential additional factors contributing to the dual mono-ubiquitylation signaling of DNA maintenance methylation.

Methods

Primers. All oligonucleotide sequences are listed in Supplementary Table 3.

***Xenopus* egg extracts.** *Xenopus laevis* was purchased from Kato-S Kagaku and handled according to the animal care regulations at the University of Tokyo. Preparation of interphase egg extracts, chromatin isolations, immunodepletions, and UbVS reactions was performed as described previously²³ with minor modifications. Briefly, all extracts were supplemented with energy regeneration mix (2 mM ATP, 20 mM phosphocreatine, and 5 μ g/ml creatine kinase). Demembrated sperm nuclei (3000–4000 sperm/ μ l in the final reaction) were added to egg extracts and incubated at 22 °C. For chromatin spin-down from the egg extracts, sperm nuclei were incubated in 15–25 μ l of the extract preparation. The extracts were diluted with ten volumes of ice-cold chromatin purification buffer (CPB; 50 mM KCl, 5 mM MgCl₂, 20 mM HEPES-KOH, pH 7.7) containing 2% sucrose, 0.1% NP-40, and 2 mM *N*-ethylenmaleimide (NEM) and kept on ice for 5 min. Diluted extracts were underlayered with 1.5 ml of a 30% sucrose cushion in CPB and centrifuged at 15,000 \times *g* for 10 min at 4 °C using a swing-bucket rotor. The pellets were resuspended in Laemmli sample buffer. For xPAF15 depletion, 250 μ l of antiserum were coupled to 50 μ l of recombinant protein A-sepharose (rPAS, GE Healthcare). Antibody beads were washed three times in phosphate-buffered saline (PBS) and added with 5 μ l fresh rPAS. Beads were washed twice in



CPB, split into three portions, and 100 μ l extracts were depleted in three rounds at 4 $^{\circ}$ C, each for 1 h. For xUHRF1 depletion, 170 μ l of antiserum were coupled to 35 μ l of rPAS. Antibody beads were washed three times in PBS and added with 4 μ l fresh rPAS. Beads were washed twice in CPB, split into two portions, and 100 μ l extracts were depleted in two rounds at 4 $^{\circ}$ C, each for 1 h. For xDNMT1 depletion, 250 μ l of antiserum were coupled to 50 μ l of rPAS. Antibody beads were washed three times in PBS and added with 5 μ l fresh rPAS. Beads were washed twice in CPB, split into three portions, and 100 μ l extracts were depleted in three rounds at 4 $^{\circ}$ C, each

for 1 h. For add-back experiments, recombinant xPAF15 was added to xPAF15-depleted extracts at 320 nM, recombinant xUHRF1 was added to xUHRF1-depleted extracts at 110 nM, and recombinant xDNMT1 was added to xDNMT1-depleted extracts at 85 nM.

For UbVS reactions, egg extracts were incubated with 20 μ M UbVS (Boston Biochem, Cambridge, MA, USA) for 30 min at 22 $^{\circ}$ C. Sperm nuclei were then added to egg extracts with or without 58 μ M ubiquitin (Boston Biochem). For quantification of PAF15Ub2 on chromatin, immunoblot films from three

Fig. 6 **mPAF15Ub2 is required for the proper maintenance of DNA methylation in mouse ESCs.** **a, b** DNA methylation levels (%) as measured by RRBS in wild-type (WT) and *Paf15* K15R/K24R (KRKR) double mutant ESCs. **a** Global DNA methylation levels and **b** CpG methylation levels at CpG islands, promoters, gene bodies, and repeats in wt and KRKR ESCs. *p* Values based on ANOVA with post hoc Tukey's test. **c** Density plot depicting the distribution of DNA methylation levels of individual CpG sites in wt and KRKR ESCs. **d–f** Replication timing of hypomethylated vs. unchanged tiles in **d** *Paf15* KRKR ESCs, **e** *Dnmt1* KO ESCs, and **f** *Uhrf1* KO ESCs. For comparisons between hypomethylated and unchanged tiles, Welch's two-sided *t* test was used for calculating *p* values. Differentially methylated tiles losing DNA methylation (hypomethylated tiles) were defined as those with *p* < 0.05 and a methylation loss >25%; *p* values were derived from a methylKit package (see "Methods"). **g** Model of the two pathways of dual mono-ubiquitylation facilitating maintenance of DNA methylation. Both requiring UHRF1, PAF15Ub2 and H3Ub2 preferentially contribute to the DNMT1-mediated maintenance of DNA methylation of early and late replicating regions, respectively. For the boxplots in **a, b, d–f**, the horizontal black lines within boxes represent median values, boxes indicate the upper and lower quartiles, and whiskers indicate the 1.5× interquartile range. Source data are provided as a Source Data file.

independent experiments were scanned. The pixel intensity of protein bands was then quantified with Image J, and the average intensity normalized to UHRF1 was calculated for each set of conditions. Antibodies against xPAF15 were raised in rabbits by immunization with a GST-tagged recombinant full-length xPAF15. Antisera were further affinity-purified with the recombinant protein immobilized on a nitrocellulose membrane (1:500 dilution for western blots). Rabbit polyclonal antibodies raised against *Xenopus* DNMT1 and UHRF1 have been previously described¹⁷ (1:500 dilution for western blots). Rabbit polyclonal USP7 antibody A300-033A was purchased from Bethyl Laboratories (1:1000 dilution for western blots). Mouse monoclonal antibody against PCNA PC-10 was purchased from Santa Cruz Biotechnology and used for immunoblotting (1:1000 dilution for western blots). Rabbit polyclonal histone H3 antibody ab1791 was purchased from Abcam (1:3000 dilution for western blots). The following antibodies were generous gifts: xPCNA antibody used for IPs (TS. Takahashi, Kyusyu Univ.), xORC2 (J. Maller, University of Colorado, 1:1000 dilution for western blots), and xCdt1 (Marcel Mechali, CNRS, 1/2000 dilution for western blots). For IP, 10 µl of Protein A agarose (GE Healthcare) was coupled with 2 µg of purified antibodies or 5 µl of antiserum. The agarose beads were washed twice with CPB buffer containing 2% sucrose. The antibody beads were incubated with egg extracts for 1 h at 4 °C. The beads were washed four times with CPB buffer containing 2% sucrose and 0.1% Triton X-100 and resuspended in 20 µl of 2× Laemmli sample buffer.

Pull-down of DNMT1-interacting proteins from chromatin. MNase-digested chromatin fractions were prepared as described previously^{23,35}. Briefly, the chromatin pellet was resuspended and digested in 100 µl of digestion buffer (10 mM HEPES-KOH, 50 mM KCl, 2.5 mM MgCl₂, 0.1 mM CaCl₂, 0.1% Triton X-100, pH 7.5, and 10 µM PR-619) containing 4 U/ml micrococcal nuclease (MNase) at 22 °C for 20 min. The reaction was stopped by the addition of 10 mM EDTA, and the mixture was centrifuged at 17,700 × *g* for 10 min. To prepare denatured chromatin lysates, the supernatant was treated with 1% SDS and then immediately diluted with lysis buffer (150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 15 mM Tris-HCl, pH 8.0). For the pull-down experiment, 5 µg purified recombinant xDNMT1 or its mutant were coupled with 20 µl of anti-FLAG M2 affinity resin at 4 °C for 1 h. The beads were collected and washed with CPB buffer containing 2% sucrose and then incubated with MNase-digested chromatin. After incubation at 4 °C for 1 h, beads were washed with CPB buffer containing 0.1% Triton X-100. Bound proteins were analyzed by immunoblotting.

Mass spectrometry. Liquid chromatography–tandem mass spectrometry (LC-MS/MS) analyses were performed essentially as previously described⁵⁰ with some modifications. Immunoprecipitated proteins were separated by SDS-polyacrylamide gel electrophoresis (PAGE) and stained with Bio-Safe Coomassie (Bio-Rad). Gels were washed in Milli-Q water, and excised. The gel pieces were washed in 50 mM ammonium bicarbonate (AMBC)/30% acetonitrile (ACN) for >2 h, and subsequently with 50 mM AMBC/50% ACN for >1 h. The gels were then dehydrated in 100% ACN for 15 min. Trypsin digestion was performed by incubation at 37 °C for 12–15 h with 20 ng/µl modified sequence-grade trypsin (Promega) in 50 mM AMBC and 5% ACN, pH 8.0. After digestion, the peptides were extracted four times with 0.1% trifluoroacetic acid/70% ACN. Extracted peptides were concentrated by vacuum centrifugation. For the LC-MS/MS analyses, a Nanoflow UHPLC, Easy nLC 1000 (Thermo Fisher Scientific) was connected online to a quadrupole-equipped Orbitrap MS instrument (Q Exactive, Thermo Fisher Scientific) with a nanoelectrospray ion source (Thermo Fisher Scientific). Peptides were separated on C18 analytical columns (Reprosil-Pur 3 µm, 75 µm id × 12 cm packed tip column, Nikyo Technos Co., Ltd.) in a 90-min three-step gradient (0–40% Solvent B for 72 min, 40–100% for 12 min, and 100% for 6 min) at a constant flow rate of 300 nl/min. The Q Exactive was operated using the Xcalibur software (Thermo Fisher Scientific) in data-dependent MS/MS mode, and the top 10 most intense ions with a charge state of +2 to +5 were selected with an isolation window of 2.0 *m/z* and fragmented by higher-energy collisional dissociation with a normalized collision energy of 28. Resolution and automatic gain control targets were set to 70,000 and 3E6, respectively. The data were analyzed using the Sequest HT search program in Proteome Discoverer 2.2 (Thermo Fisher Scientific). The maximum of missed cleavage sites of trypsin was set to three. Acetylation (Protein N-term and Lys), oxidation (Met), GlyGly modification (Lys), phosphorylation

(Ser, Thr, Tyr), and pyroglutamate conversion (N-term Gln) were selected as variable modifications. Peptide identification was filtered at a false discovery rate <0.01. Non-label protein quantification was performed using the Precursor Ions Quantifier node in Proteome Discoverer 2.2. The RAW files have been deposited to the ProteomeXchange Consortium^{51,52}.

Recombinant *Xenopus* proteins. The *X. laevis* *Paf15* cDNA was amplified by PCR from a *X. laevis* cDNA library using primers 3621 and 3622 and ligated into pTA2 vector. For GST-xPAF15 expression, the amplified *xPaf15* genes with primers 3667 and 3636 were gel-isolated and ligated into linearized pGEX4T-3 using In-Fusion (Clontech) according to the manufacturer's instructions. Protein expression in *Escherichia coli* (BL21-CodonPlus) was induced by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside to media followed by incubation for 12 h at 20 °C. For purification of GST-tagged proteins, cells were collected and resuspended in lysis buffer (20 mM HEPES-KOH pH 7.6, 0.5 M NaCl, 0.5 mM EDTA, 10% glycerol, 1 mM dithiothreitol (DTT)) supplemented with 0.5% NP40 and protease inhibitors and were then disrupted by sonication on ice. After centrifugation, the supernatant was applied to glutathione Sepharose beads (GE Healthcare) and rotated at 4 °C for 2 h. Beads were then washed three times with wash buffer 1 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% TritonX-100, 1 mM DTT) and once with wash buffer 2 (100 mM Tris-HCl (pH 7.5), 100 mM NaCl). Bound proteins were released in elution buffer (100 mM Tris-HCl pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM DTT) containing 42 mM reduced glutathione, and the purified protein was loaded on a PD10 desalting column equilibrated with EB buffer (10 mM HEPES/KOH at pH 7.7, 100 mM KCl, 0.1 mM CaCl₂, 1 mM MgCl₂) containing 1 mM DTT and then concentrated with Vivaspin (Millipore).

For protein expression in insect cells, C-terminally 3× Flag-tagged *xPaf15* genes were transferred from pKS104 vector into pVL1392 vector. The amplified *xPaf15* genes with primers 3720 and 3581 were gel-isolated and ligated into linearized pVL1392 using In-Fusion. R3A, T4D, K5A, K18R, K27R, K18R/K27R, and F72AF73A mutations in pKS104-xPAF15 or pVL1392-xPAF15 constructs were introduced using a KOD-Plus Mutagenesis Kit (Toyobo). All mutations were confirmed by Sanger sequencing. Baculoviruses were produced using a BD BaculoGold Transfection Kit and a BestBac Transfection Kit (BD Biosciences), following the manufacturer's protocol. Proteins were expressed in Sf9 insect cells by infection with viruses expressing xPAF15 WT-3× Flag or its mutant for 72 h. Sf9 cells from a 500 ml culture were collected and lysed by resuspending them in 20 ml lysis buffer (20 mM Tris-HCl, pH 8.0, 100 mM KCl, 5 mM MgCl₂, 10% glycerol, 1% Nonidet P40 (NP-40), 1 mM DTT, 10 µg/ml leupeptin, and 10 µg/ml aprotinin), followed by incubation on ice for 10 min. A soluble fraction was obtained after centrifugation of the lysate at 15,000 × *g* for 15 min at 4 °C. The soluble fraction was incubated for 4 h at 4 °C with 250 µl of anti-FLAG M2 affinity resin (Sigma-Aldrich) equilibrated with lysis buffer. The beads were collected and washed with 10 ml wash buffer (20 mM Tris-HCl, pH 8.0, 100 mM KCl, 5 mM MgCl₂, 10% glycerol, 0.1% NP-40, 1 mM DTT) and then with 5 ml EB (20 mM HEPES-KOH, pH 7.5, 100 mM KCl, 5 mM MgCl₂) containing 1 mM DTT. The recombinant xPAF15 was eluted twice in 250 µl EB containing 1 mM DTT and 250 µg/ml 3× FLAG peptide (Sigma-Aldrich). Eluates were pooled and concentrated using a Vivaspin 500 (GE Healthcare Biosciences). We note that all PAF15 mutant proteins was purified as efficiently as the WT protein from insect cells. For expression of C-terminally 3× Flag-tagged xUHRF1 (Wt and D333A/D336A), C-terminally 3× Flag-tagged *xUhrf1* genes were transferred from pKS104 vector into pVL1392 vector. The amplified *xUhrf1* genes with primers 4620 and 3581 were gel-isolated and ligated into linearized pVL1392 using In-Fusion. The D333A/D336A substitution was introduced using a KOD-Plus Mutagenesis Kit and confirmed by Sanger sequencing. Recombinant xUHRF1 proteins were also purified as described above.

DNA methylation and replication in *Xenopus* egg extracts. DNA methylation was monitored by the incorporation of S-[methyl-³H]-adenosyl-L-methionine, incubated at room temperature, and the reaction was stopped by the addition of CPB containing 2% sucrose up to 300 µl. Genomic DNA was purified using a Wizard Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. Incorporation of radioactivity was quantified with a liquid scintillation counter. DNA replication was assayed by adding [α-³²P]-dCTP to egg extracts

containing sperm chromatin. The reaction was stopped by adding 1% SDS, 40 mM EDTA and spotted onto Whatman glass microfibre filters followed by trichloroacetic acid (TCA) precipitation with 5% TCA containing 2% pyrophosphate. Filters were washed in ethanol, dried, and TCA-precipitated radioactivity was counted in scintillation liquid.

Structure of the UHRF1 PHD finger bound to PAF15₂₋₁₁. The PHD finger of human UHRF1 (residues 299–366) was expressed as a fusion protein with GST and small ubiquitin-like modifier-1 (SUMO-1) at its N-terminus. Cell culture and purification were performed according to our previous report^{14,23}. Briefly, hPHD was purified using GST affinity column of glutathione Sepharose 4B (GS4B; GE Healthcare). The GST-SUMO-1 fused hPHD was eluted with reduced glutathione and then GST-SUMO-1 tag was removed by the SUMO-specific protease GST-SEN2. The protein was further purified by anion-exchange chromatography using a HiTrap Q HP column and by SEC using a HiLoad 26/60 Superdex75 column (GE Healthcare). PAF15₂₋₁₁ was synthesized at Toray Research Center (Tokyo, Japan). The PHD finger:PAF15₂₋₁₁ complex was prepared by adding a 1.5-molar excess of the PAF15₂₋₁₁ peptide to the protein before its concentration using an Amicon concentrator with a 3000 Da cutoff (Millipore). The crystal was obtained using a 30 mg/ml concentration of the complex at 20 °C and the hanging drop vapor diffusion method with a reservoir solution containing 0.1 M HEPES-NaOH (pH 7.5) and 70% (v/v) 2-methyl-2,4-pentanediol. The crystal was directly frozen in liquid nitrogen. The X-ray diffraction data were collected at a wavelength of 0.9800 Å on a Pilatus3 6M detector in beam line BL-17A at Photon Factory (Tsukuba, Japan) and scaled at 1.70 Å resolution with the program XDS package⁵³ and Aimless⁵⁴. This was followed by molecular replacement by PHASER⁵⁵ and several cycles of model refinement by PHENIX⁵⁶. The final model converged at 1.70 Å resolution with a crystallographic *R*-factor of 17.6% and a free *R*-factor of 18.7%. The crystallographic data and refinement statistics are given in Table 1. The figures were generated using PyMOL (<http://www.pymol.org>).

ITC measurements. Preparation and purification of the disulfide linked K15 and K24 mono-ubiquitylated analog of human PAF15₂₋₃₀ and the human DNMT1 RFTS domain, residues 351–600, were performed according to our previous report²³. hRFTS was purified using GST-affinity column of GS4B. After removing the GST-SUMO1 tag by GST-SEN2, the protein was further purified by anion-exchange chromatography of HiTrap Q HP column and HiLoad 26/60 Superdex75 column. hPAF15₂₋₃₀ K15C/K24C mutant was synthesized at Toray Research Center (Tokyo, Japan). Ubiquitin G76C mutant activated by 5,5'-dithiobis-(2-nitrobenzoic acid) (Wako) and hPAF15₂₋₃₀ K15C/K24C were incubated for 1 h at room temperature and then purified by cation-exchange chromatography of Mono-S (GE Healthcare) to separate from the by-products. The UHRF1 PHD finger was buffer-exchanged using Superdex 200 Increase 10/300 GL (GE Healthcare) equilibrated with 10 mM HEPES-NaOH (pH 7.5), 150 mM NaCl, and 0.25 mM tris(2-carboxyethyl)phosphine hydrochloride and lyophilized PAF15 (residues 2–11) peptide was dissolved in the same buffer. hRFTS and PAF15₂₋₃₀Ub2 were equilibrated with 10 mM HEPES-NaOH (pH 7.5), 150 mM NaCl, and 10 μM zinc acetate. A MicroCal LLC calorimeter, VP-ITC (MicroCal), was used for the ITC measurements. The data were analyzed with the software ORIGIN (MicroCal) using a one-site model.

In vitro ubiquitylation assay. Protein expression in *E. coli* and purification of mouse UBA1 (E1), human UHRF1 (WT and its mutants), and ubiquitin were performed according to the previous reports¹⁷. E1 enzyme was expressed in *E. coli* Rosetta 2 (DE3) (Novagen) as a six histidine-tag fusion protein. The protein was purified using TALON[®] (Clontech), HiTrap Q HP, and SEC using HiLoad 26/60 Superdex200 column (GE Healthcare). hUHRF1 WT and mutants, D334A/D337A and H741A, were expressed in *E. coli* Rosetta 2 (DE3) as a GST-fusion protein. The protein was purified GST-affinity chromatography of GS4B column. After removal of GST-tag by HRV-3C protease, the protein was further purified by HiTrap Heparin HP column (GE Healthcare) and HiLoad 26/60 Superdex200 column. Purification procedure of ubiquitin was as follows: after cell lysis and centrifugation, the supernatant was boiled at 85 °C for 15 min. After removing the debris by centrifugation, ubiquitin was further purified using cation-exchange chromatography of HiTrap SP HP (GE Healthcare) and HiLoad 26/60 Superdex75 column. UBCh5 (E2) was purified using TALON[®] and SEC of HiLoad 26/60 Superdex75 column. The cDNA encoding amino acids 2–71 of human PAF15 harboring HA-tag at the C-terminus was cloned into the modified pET21b vector, pET-N^{Pro} vector⁵⁷. The N^{Pro}-fused PAF15 was purified from the pellet fraction. The inclusion body was then solubilized in buffer containing 8 M urea, 50 mM Tris-HCl (pH 7.5), and 25 mM DTT by stirring overnight at 4 °C. Then the denatured fusion proteins were purified by Ni Sepharose 6 Fast Flow (GE Healthcare). The eluents were dialyzed in a step-wise manner to gradually remove the urea. The solution was additionally incubated with a buffer containing 100 mM Tris-HCl (pH 7.5), 200 mM NaCl, and 2 mM DTT for 12–24 h at room temperature for completing autocleavage of N^{Pro}. The protein was further purified using HiTrap SP HP and HiLoad 16/60 Superdex 30 (GE Healthcare).

Standard ubiquitylation reaction mixtures contained 116 μM ubiquitin, 200 nM E1, 6 μM E2, 3 μM E3, 5 mM ATP, and 50 μM PAF15-HA as a substrate in

ubiquitylation reaction buffer (50 mM HEPES [pH 8.0], 150 mM NaCl, 5 mM MgCl₂, 0.1% Triton X-100, 2 mM DTT). The mixture was incubated at 30 °C for 30 min, and the reaction was stopped by adding 3× SDS loading buffer. The reaction was analyzed by SDS-PAGE, followed by western blotting using 1/20,000 diluted anti-HA antibody (MBL, #M180-3).

SEC-SAXS data collection, processing, and interpretation. SAXS data were collected on Photon Factory BL-10C using a UPLC[®] ACQUITY (Waters) integrated SAXS set-up. Fifty μl of a 6 mg/ml sample were loaded onto a Superdex 200 Increase 5/150 GL (GE Healthcare Science) pre-equilibrated with 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 5% glycerol at a flow rate of 0.25 ml/min at 4 °C. The flow rate was reduced to 0.025 ml/min at an elution volume of 1.63–2.30 ml. X-ray scattering was collected every 20 s on a PILATUS3 2M detector over an angular range of $q_{\min} = 0.00690 \text{ \AA}^{-1}$ to $q_{\max} = 0.27815 \text{ \AA}^{-1}$. UV spectra at a range of 200–450 nm were recorded every 10 s. Circular averaging and buffer subtraction were carried out using the program SAnGLer⁵⁸ to obtain one-dimensional scattering data $I(q)$ as a function of q ($q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength 1.5 Å). The scattering intensity was normalized on an absolute scale using the scattering intensity of water⁵⁹. The multiple concentrations of the scattering data around the peak at A_{280} and $I(0)$ were extrapolated to zero-concentration using a Serial Analyzer⁶⁰. The molecular weights of samples were calculated from the $I(q)$ data of Ovalbumin (Sigma) at the highest values of A_{280} and $I(0)$. The radius of gyration R_g and the forward scattering intensity $I(0)$ were estimated from the Guinier plot of $I(q)$ in the smaller angle region of $qR_g < 1.3$. The distance distribution function $P(r)$ of the sample at the highest peak of A_{280} and $I(0)$ was calculated using the program GNOM⁶¹, where the experimental $I(q)$ data were used in a q -range of 0.00885–0.17670 Å^{-1} . The maximum particle dimension D_{\max} was estimated from the $P(r)$ function as the distance r for which $P(r) = 0$. The molecular weight of the sample was estimated by comparing the $I(0)/c$ (where c is the protein concentration) of the sample to that of Ovalbumin.

Cell culture. The mESC line J1 was originally provided by the laboratory of Dr. Rudolf Jaenisch (Whitehead Institute). *Dnmt1* KO mESCs were described in ref. ⁶² and *Uhrf1* KO mESCs were described in ref. ³⁶. All mESC lines were maintained on 0.2% gelatin-coated dishes in Dulbecco's modified Eagle's medium (Sigma) supplemented with 16% fetal bovine serum (FBS, Sigma), 0.1 mM β-mercaptoethanol (Invitrogen), 2 mM L-glutamine (Sigma), 1× Minimum Essential Medium non-essential amino acids (Sigma), 100 U/ml penicillin, 100 mg/ml streptomycin (Sigma), recombinant LIF (ESGRO, Millipore), and 2i (1 mM PD032591 and 3 mM CHIR99021 (Axon Medchem, Netherlands)). Baby hamster kidney (BHK) cells containing a stably integrated lac operator (lacO) array used for the F3H assay were kindly provided by the laboratory of Dr. David L. Spector⁶³. BHK cells were grown in a humidified atmosphere at 37 °C and 5% CO₂ in Dulbecco's modified Eagle's medium supplemented with 1 mM Gentamycin (Serva GmbH) and 10% FBS (Sigma). All cell lines were regularly tested for mycoplasma contamination.

CRISPR/Cas9 gene editing and excision. For generation of *Paf15* K15R and K24R mutant mESCs, specific gRNAs for each mutation were cloned into a modified version of the SpCas9-T2A-Puromycin/gRNA vector (px459;⁶⁴ Addgene plasmid #62988), in which SpCas9 is fused to truncated human Geminin (hGem) to preferentially generate double-strand breaks when homology-directed repair is active⁶⁵. To generate targeting donors for each desired mutation, single-stranded oligonucleotides harboring either the K15R or K24R substitution and ~100 bp homologous to the respective genomic locus were synthesized (IDT, Coralville, IA, USA). Cells were transfected with a 4:1 ratio of donor oligo and Cas9/gRNA construct. RNA vector was obtained via cut-ligation. Two days after transfection, cells were plated at clonal density and subjected to a transient puromycin selection (1 mg/ml) for 40 h. Colonies were picked out 6 days after transfection. Cell lysis in 96-well plates, PCRs of lysates, and restriction digestion were performed as previously described⁶². Successful insertion of *Paf15* K15R and K24R mutations was confirmed by Sanger sequencing. For generation of the *Paf15* K15R/K24R double-mutant ESC lines, three characterized *Paf15* K24R single mutants were subjected to a second round of gene editing to achieve the K15R substitution as described above.

Quantitative real-time PCR (qRT-PCR) analysis. Total RNA was isolated using a NucleoSpin Triprep Kit (Macherey-Nagel) according to the manufacturer's instructions. cDNA synthesis was performed with a High-Capacity cDNA Reverse Transcription Kit (with RNase Inhibitor; Applied Biosystems) using 2 μg of total RNA as input. qRT-PCR assays with the oligonucleotides listed in Supplementary Table 3 were performed in 8 μl reaction volumes with 5 ng of cDNA used as input. For SYBR green detection, FastStart Universal SYBR Green Master Mix (Roche) was used. The reactions were run on a LightCycler480 (Roche).

Co-IP and western blotting of mouse samples. For co-IP of DNMT1, 1.5×10^7 of mESCs were lysed in 250 μl of lysis buffer (10 mM Tris/Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP40, 1.5 mM MgCl₂, 0.5 μg/ml Benzamide (Sigma-Aldrich), 1 mM PMSF, 1× mammalian Protease Inhibitor Cocktail (e.g., Serva[®]), 5 mM NEM (Sigma)) at 4 °C for 30 min. Lysates were cleared by centrifugation at 20,000 × *g* for 15 min at 4 °C, and the protein concentration was measured using Pierce[™] 660 nm

Protein Assay Reagent according to the manufacturer's instructions. For DNMT1 IP, we used an anti-DNMT1 nanobody (commercial name: DNMT1-Trap, ChromoTek), which is an antigen-binding domain (V_{H1H}) derived from the heavy chain of an alpaca antibody raised against DNMT1. Equal amounts of protein extracts were incubated with 25 μ l of DNMT1-Trap (undiluted) for 2 h at 4 °C under constant rotation. Beads were washed three times with washing buffer (10 mM Tris/HCl pH7.5, 150 mM NaCl, 0.5 mM EDTA) and boiled in Laemmli buffer at 95 °C for 10 min. Bound fractions were separated and visualized as a western blot.

To isolate cytoplasmic and nuclear fractions, 2×10^7 of mESCs were treated with 400 μ l of hypotonic buffer (10 mM Tris-HCl pH 8, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 1 \times Protease Inhibitor, 2 mM PMSF, 5 mM NEM, and 0.1% Triton X-100) at 4 °C for 5 min. The cytoplasmic fraction was separated from nuclei by centrifugation at 1300 \times g for 10 min at 4 °C, then supplemented with 150 mM NaCl and clarified by centrifugation at 20,000 \times g for 15 min at 4 °C. Nuclei were lysed as described above. Anti-mPAF15 antibody (Santa Cruz, sc-390515, 2 μ g) was added to the cytoplasmic and nuclear lysate and incubated for 2 h at 4 °C under constant rotation. To precipitate mPAF15-bound proteins, 20 μ l of protein G beads (GE17-0618-06) were added to the lysate for an overnight incubation at 4 °C under constant rotation.

Western blots for mDNMT1 were performed as described previously¹⁸ using a monoclonal antibody (rat anti-DNMT1, 14F6, 1:10 dilution) and a polyclonal antibody (rabbit anti-DNMT1, Abcam, ab87654, 1:2500 dilution). Other antibodies used for detection were mouse anti-PAF15 antibody (Santa Cruz, sc-390515, 1:1500 dilution), polyclonal rabbit-anti-H3 (Abcam, ab1791, 1:5000 dilution), and a monoclonal mouse-anti-tubulin (Sigma, T9026, 1:2000 dilution). The following secondary antibodies conjugated to horseradish peroxidase were used: goat polyclonal anti-rat IgG (Dianova, 112-035-003, 1:5000), goat polyclonal anti-rabbit IgG (Bio-rad), and rabbit polyclonal anti-mouse IgG (Sigma, A9044, 1:5000). For detection of horseradish peroxidase-conjugated antibodies, an ECL Plus reagent (GE Healthcare, Thermo Scientific) was used.

Reduced representation bisulfite sequencing. For RRBS, genomic DNA was isolated using a QIAamp DNA Mini Kit (QIAGEN), after an overnight lysis and proteinase K treatment. Preparation of the RRBS library was carried out as described previously⁶⁶, with the following modifications: bisulfite treatment was performed using an EZ DNA Methylation-Gold™ Kit (Zymo Research Corporation) according to the manufacturer's protocol except that libraries were eluted in 2×20 mL M-elution buffer. RRBS libraries were sequenced on an Illumina HiSeq 1500 in 50 bp paired-end mode.

RRBS alignment and analysis. Raw RRBS reads were first trimmed using Trim Galore (v.0.3.1) with the “-rbs” parameter. Alignments were carried out with the mouse genome (mm10) using bsmap (v.2.90) and the parameters “-s 12 -v 10 -r 2 -i 1.” CpG-methylation calls were extracted from the mapping output using bsmaps methratio.py. Analysis was restricted to CpG with a coverage >10. A methylKit⁶⁷ was used to identify differentially methylated regions between the respective contrasts for the following genomic features: (1) repeats (defined by Replibase), (2) gene promoters (defined as gene start sites -2 kb/+2 kb), and (3) gene bodies (defined as the longest isoform per gene) and CpG islands (as defined by ref.⁶⁸). Differentially methylated regions were identified as regions with $p < 0.05$ and a difference in methylation means between two groups >25%.

Data processing and analysis. Chromatin IP-sequencing reads for H3K9me2³⁸, H3K9me3³⁹ and H3K14ac⁴⁰ in ESCs and EpiLCs were downloaded from GSE6020467, GSE2394368, and GSE3128469, respectively. Reads were aligned to the mouse genome (mm10) with Bowtie (v.1.2.2) with parameters “-a -m 3 -n 3 -best -strata.” Peak calling and signal pile-up was performed using MACS2 callpeak⁶⁹ with the parameters “-extsize 150 -nomodel -B -nolambda” for all samples. Tag densities for 1 kb Tiles detected in RBBS were calculated using custom R scripts. Replication domain data for mouse ESCs (mm10) for replication timing analysis was taken from <http://www.replicationdomain.org/>⁶⁹. The average replication timing ratio was calculated over 1 kb Tiles detected in RBBS using custom R scripts. Data of partially methylated domains and highly methylated domains (mm10) was downloaded from <https://zwdzwd.github.io/pmd42> and used to calculate average DNA methylation levels (RRBS) over these regions.

High-throughput immunofluorescence and image analysis. ESCs were grown in 96-well microplates (μ Clear, Greiner Bio-One), washed with PBS, and fixed with 3.7% formaldehyde. After three washing steps with PBST, cells were permeabilized (0.5 % Triton-X100), treated with denaturing solution (2 N HCl) for 40 min, and incubated with renaturing solution (150 mM Tris-HCl, pH 8.5) for 20 min. Cells were then blocked in 2% bovine serum albumin for 1 h and incubated with primary antibody (mouse-anti 5mC, Diagenode 33D3) for 1 h at 37 °C. After washing three times with PBST, cells were incubated with secondary antibody (goat-anti-mouse coupled to Alexa647, Thermo Fisher) for 1 h at 37 °C. Cells were washed three times with PBST, counterstained with 200 ng/ml 4,6-diamidino-2-phenylindole (DAPI), and finally covered with PBS. Images were acquired by automation with an Operetta High-Content Image Analysis System (PerkinElmer, $\times 40$ high NA objective) followed by analysis with the Harmony

software (PerkinElmer). DAPI was used for the detection of single nuclei and 5mC modifications were measured in selected nuclei based on the antibody signal intensity.

F3H assay. The F3H assay was performed as described previously³⁷. In brief, BHK cells containing multiple lac operator repeats were transiently transfected on coverslips using polyethyleneimine and fixed with 3.7% formaldehyde 24 h after transfection. For DNA counterstaining, coverslips were incubated in a solution of DAPI (200 ng/ml) in PBST and mounted in Vectashield. Cell images were collected using a Leica TCS SP5 confocal microscope. To quantify the interactions within the lac spot, the following intensity ratio was calculated for each cell: $(mCherry_{spot} - mCherry_{background}) / (GFP_{spot} - GFP_{background})$ in order to account for different expression levels. The following constructs used in the F3H assay have been described previously: pCAG-eGFP-IB⁷⁰, pCAG-eGFP-mDNMT1⁷⁰, and pGBP-LacI³⁷. To generate the mCherry-mPAF15 WT and KRKR expression constructs, the coding sequences of mPAF15 WT and KRKR were excised via AsiSI and NotI restriction digest from the GFP-PAF15 WT and KRKR constructs³⁶ and ligated into the pCAG-Cherry-IB vector⁷⁰.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding authors upon reasonable request. The crystal structures of the human UHRF1 PHD in complex with PAF15(2-11) has been deposited in the Protein Data Bank under accession code 6IIW. Sequencing data reported in this paper (wt and PAF15KRKR RRBS) are available at ArrayExpress (EMBL-EBI) under accession number E-MTAB-7930. The mass spectrometric proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository with dataset identifier PXD015282. The source data underlying Figs. 1b–e, 2a–f, 3a–d, 4a–e, f, and 5a, b, e and Supplementary Figs. 1b–d, f–g, 2a–c, 3a, b, d, e, 4a–f, 5c, d, g, and 6a–c are provided as a Source Data file.

Received: 24 April 2019; Accepted: 10 February 2020;
Published online: 06 March 2020

References

- Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
- Jones, P. A. & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* **10**, 805–811 (2009).
- Edwards, J. R., Yarychivska, O., Bouliard, M. & Bestor, T. H. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* **10**, 23 (2017).
- Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**, 865–873 (1992).
- Takeshita, K. et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *PNAS* **108**, 9055–9059 (2011).
- Syeda, F. et al. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J. Biol. Chem.* **286**, 15344–15351 (2011).
- Bostick, M. et al. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760–1764 (2007).
- Sharif, J. et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908–912 (2007).
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**, 818–821 (2008).
- Hashimoto, H. et al. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455**, 826–829 (2008).
- Avvakumov, G. V. et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455**, 822–825 (2008).
- Rothbart, S. B. et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat. Struct. Mol. Biol.* **19**, 1155–1160 (2012).
- Rajakumara, E. et al. PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol. Cell* **43**, 275–284 (2011).
- Arita, K. et al. Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *PNAS* **109**, 12950–12955 (2012).

15. Ferry, L. et al. Methylation of DNA ligase 1 by G9a/GLP recruits UHRF1 to replicating DNA and regulates DNA methylation. *Mol. Cell* **67**, 550–565.e5 (2017).
16. Kori, S. et al. Structure of the UHRF1 tandem Tudor domain bound to a methylated non-histone protein, LIG1, reveals rules for binding and regulation. *Structure* **27**, 485–496.e7 (2018).
17. Nishiyama, A. et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* **502**, 249–253 (2013).
18. Qin, W. et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res.* **25**, 911–929 (2015).
19. Harrison, J. S. et al. Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. *Elife* **5**, 818 (2016).
20. Vaughan, R. M. et al. Chromatin structure and its chemical modifications regulate the ubiquitin ligase substrate selectivity of UHRF1. *PNAS* **115**, 8775–8780 (2018).
21. DaRosa, P. A. et al. A bifunctional role for the UHRF1 UBL domain in the control of hemi-methylated DNA-dependent histone ubiquitylation. *Mol. Cell* **72**, 753–765.e6 (2018).
22. Foster, B. M. et al. Critical role of the UBL domain in stimulating the E3 ubiquitin ligase activity of UHRF1 toward chromatin. *Mol. Cell* **72**, 739–752.e9 (2018).
23. Ishiyama, S. et al. Structure of the Dnmt1 reader module complexed with a unique two-mono-ubiquitin mark on histone H3 reveals the basis for DNA methylation maintenance. *Mol. Cell* **68**, 350–360.e7 (2017).
24. Li, T. et al. Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic Acids Res.* **14**, 204–3231 (2018).
25. Yamaguchi, L. et al. Usp7-dependent histone H3 deubiquitylation regulates maintenance of DNA methylation. *Sci. Rep.* **7**, 55 (2017).
26. Cheng, J. et al. Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. *Nat. Commun.* **6**, 7023 (2015).
27. Felle, M. et al. The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Res.* **39**, 8355–8365 (2011).
28. Du, Z. et al. DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Sci. Signal.* **3**, ra80–ra80 (2010).
29. Easwaran, H. P., Schermelleh, L., Leonhardt, H. & Cardoso, M. C. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.* **5**, 1181–1186 (2004).
30. Borodovsky, A. et al. A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. *EMBO J.* **20**, 5187–5196 (2001).
31. Dimova, N. V. et al. APC/C-mediated multiple monoubiquitylation provides an alternative degradation signal for cyclin B1. *Nat. Cell Biol.* **14**, 168–176 (2012).
32. Emanuele, M. J., Ciccio, A., Elia, A. E. H. & Elledge, S. J. Proliferating cell nuclear antigen (PCNA)-associated KIAA0101/PAF15 protein is a cell cycle-regulated anaphase-promoting complex/cyclosome substrate. *PNAS* **108**, 9845–9850 (2011).
33. Povlsen, L. K. et al. Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nat. Cell Biol.* **14**, 1089–1098 (2012).
34. Blow, J. J. & Laskey, R. A. Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of *Xenopus* eggs. *Cell* **47**, 577–587 (1986).
35. Miki, T. et al. The replication foci targeting sequence (RFTS) of DNMT1 functions as a potent histone H3 binding domain regulated by autoinhibition. *Biochem. Biophys. Res. Commun.* **470**, 741–747 (2016).
36. Karg, E. et al. Ubiquitome analysis reveals PCNA-associated factor 15 (PAF15) as a specific ubiquitination target of UHRF1 in embryonic stem cells. *J. Mol. Biol.* **429**, 3814–3824 (2017).
37. Herce, H. D., Deng, W., Helma, J., Leonhardt, H. & Cardoso, M. C. Visualization and targeted disruption of protein interactions in living cells. *Nat. Commun.* **4**, 2660 (2013).
38. Kurimoto, K. et al. Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. *Cell Stem Cell* **16**, 517–532 (2015).
39. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
40. Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H. & Tora, L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13**, 424 (2012).
41. Timp, W. et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med.* **6**, 61–11 (2014).
42. Zhou, W. et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018).
43. Hiratani, I. et al. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Res.* **20**, 155–169 (2010).
44. Aran, D., Toperoff, G., Rosenberg, M. & Hellman, A. Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* **20**, 670–680 (2011).
45. Branzei, D. & Foiani, M. Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol.* **11**, 208–219 (2010).
46. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
47. Amrani, Y. M. et al. The Paf oncogene is essential for hematopoietic stem cell function and development. *J. Exp. Med.* **208**, 1757–1765 (2011).
48. De March, M. et al. p15PAF binding to PCNA modulates the DNA sliding surface. *Nucleic Acids Res.* **79**, 1233–9828 (2018).
49. Gonzalez-Magaña, A. et al. Double monoubiquitination modifies the molecular recognition properties of p15PAF promoting binding to the reader module of Dnmt1. *ACS Chem. Biol.* **14**, 2315–2326 (2019).
50. Tsuchiya, H. et al. In vivo ubiquitin linkage-type analysis reveals that the Cdc48-Rad23/Dsk2 axis contributes to K48-linked chain specificity of the proteasome. *Mol. Cell* **66**, 488.e7–502.e7 (2017).
51. Deutsch, E. W. et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45**, D1100–D1106 (2017).
52. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
53. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
54. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
55. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
56. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
57. Goda, N. et al. An optimized Npro-based method for the expression and purification of intrinsically disordered proteins for an NMR study. *Intrinsically Disord. Proteins* **3**, e1011004 (2015).
58. Shimizu, N. et al. Software development for analysis of small-angle x-ray scattering data. *AIP Conf. Proc.* **1741**, 050017 (2016).
59. Orthaber, D., Bergmann, A. & Glatter, O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J. Appl. Crystallogr.* **33**, 218–225 (2000).
60. Yonezawa, K., Takahashi, M., Yatabe, K., Nagatani, Y. & Shimizu, N. Software for serial data analysis measured by SEC-SAXS/UV-Vis spectroscopy. *AIP Conf. Proc.* **2054**, 060082 (2019).
61. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
62. Mulholland, C. B. et al. A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res.* **43**, e112–e112 (2015).
63. Tsukamoto, T. et al. Visualization of gene activity in living cells. *Nat. Cell Biol.* **2**, 871–878 (2000).
64. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
65. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F. & Chin, L. Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell Rep.* **14**, 1555–1566 (2016).
66. Boyle, P. et al. Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol.* **13**, R92 (2012).
67. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
68. Illingworth, R. S. et al. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* **6**, e1001134 (2010).
69. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137–R139 (2008).
70. Meilinger, D. et al. 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 (2009).

Acknowledgements

We acknowledge the kind support of the beam line staff at the Photon Factory for X-ray data collection. We are grateful to En Li for providing wt J1 mESCs. We thank Martha Smets for the preparation of Dnmt1 KO and Uhrf1 KO samples. This study was supported by a PRESTO from JST (to K.A.), by MEXT/JSPS KAKENHI (JP26250027, JP22118003, JP16K15239, and JP19H05740 to M.N., JP18H02392 and JP19H05294 to K.A., JP15H01188, JP16H04818, and JP19H05285 to A.N., JP18H05498 to Y.S., JP26000014 to K.T.), by the Research Program on Hepatitis of the Japan Agency for Medical Research and

Development, and by an AMED grant (19cm0106134h0002) to M.N. H.L. acknowledges the support of the Deutsche Forschungsgemeinschaft (SFB1064/A17/A22 and SFB1243/A01). A.N. was supported in part by grants from the Naito Foundation and Mochida Memorial Foundation for Medical and Pharmaceutical Research.

Author contributions

A.N., K.A., H.L., and M.N. conceived the study and experimental design, analyzed the experimental results, and co-wrote the manuscript. A.N., Y.C., S.K., T.K., H.H., G.N., and H.A. performed most of the *Xenopus* studies. S.K., H.Y., and K.A. analyzed the structural basis of the hPHD:PAF15₂₋₁₁ and hRFTS:PAF15-K15Ub/K24Ub complex. A.E., Y.S., and K.T. performed the LC-MS/MS analysis. C.B.M., S.B., W.Q., C.T., and H.L. analyzed mPAF15 and mDNMT1 interaction and DNA methylation using mouse embryonic stem cells.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-15006-4>.

Correspondence and requests for materials should be addressed to A.N., K.A., H.L. or M.N.

Peer review information *Nature Communications* thanks Pierre-Antoine Defossez, Vytautas Iesmantavicius, and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



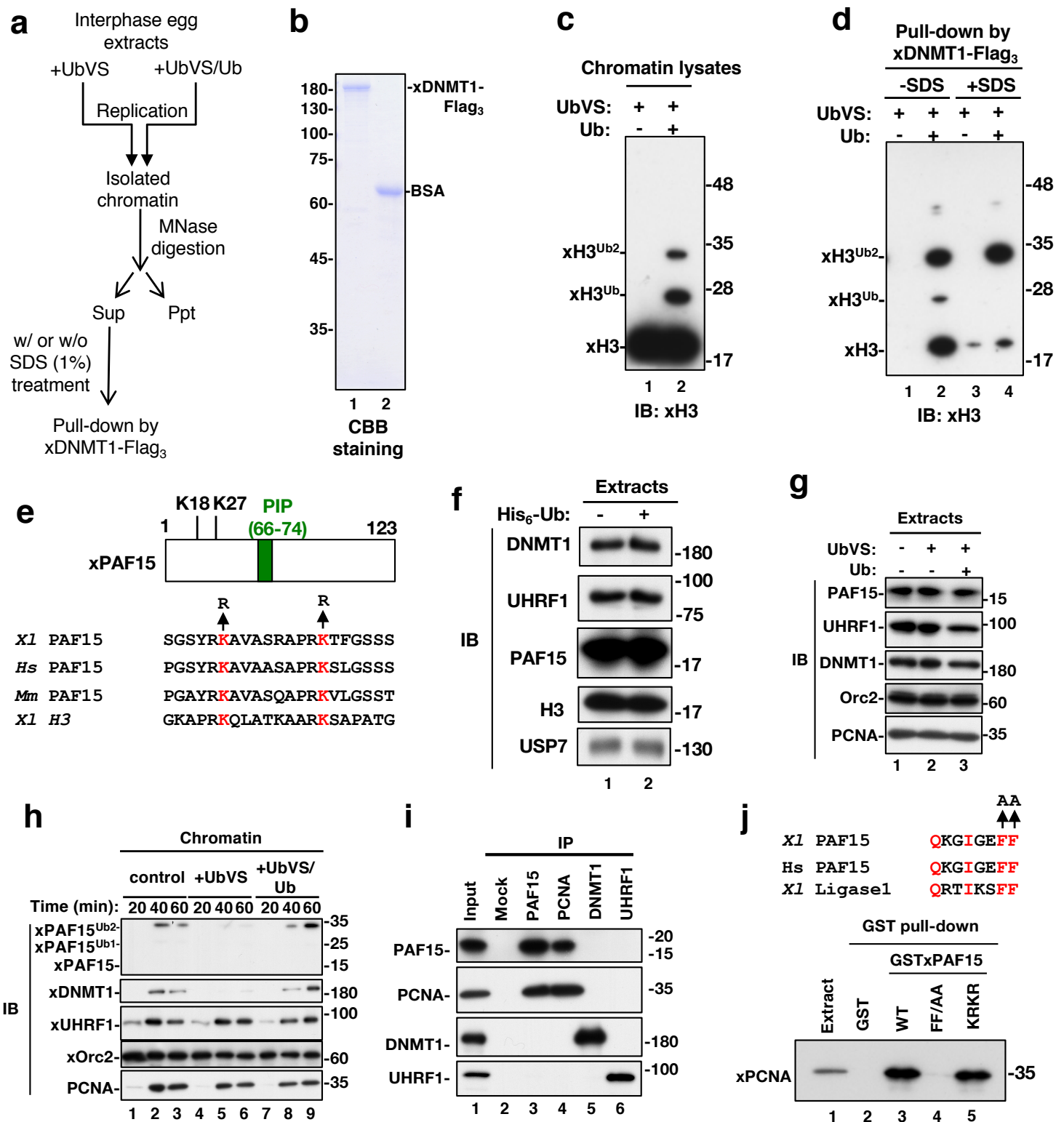
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

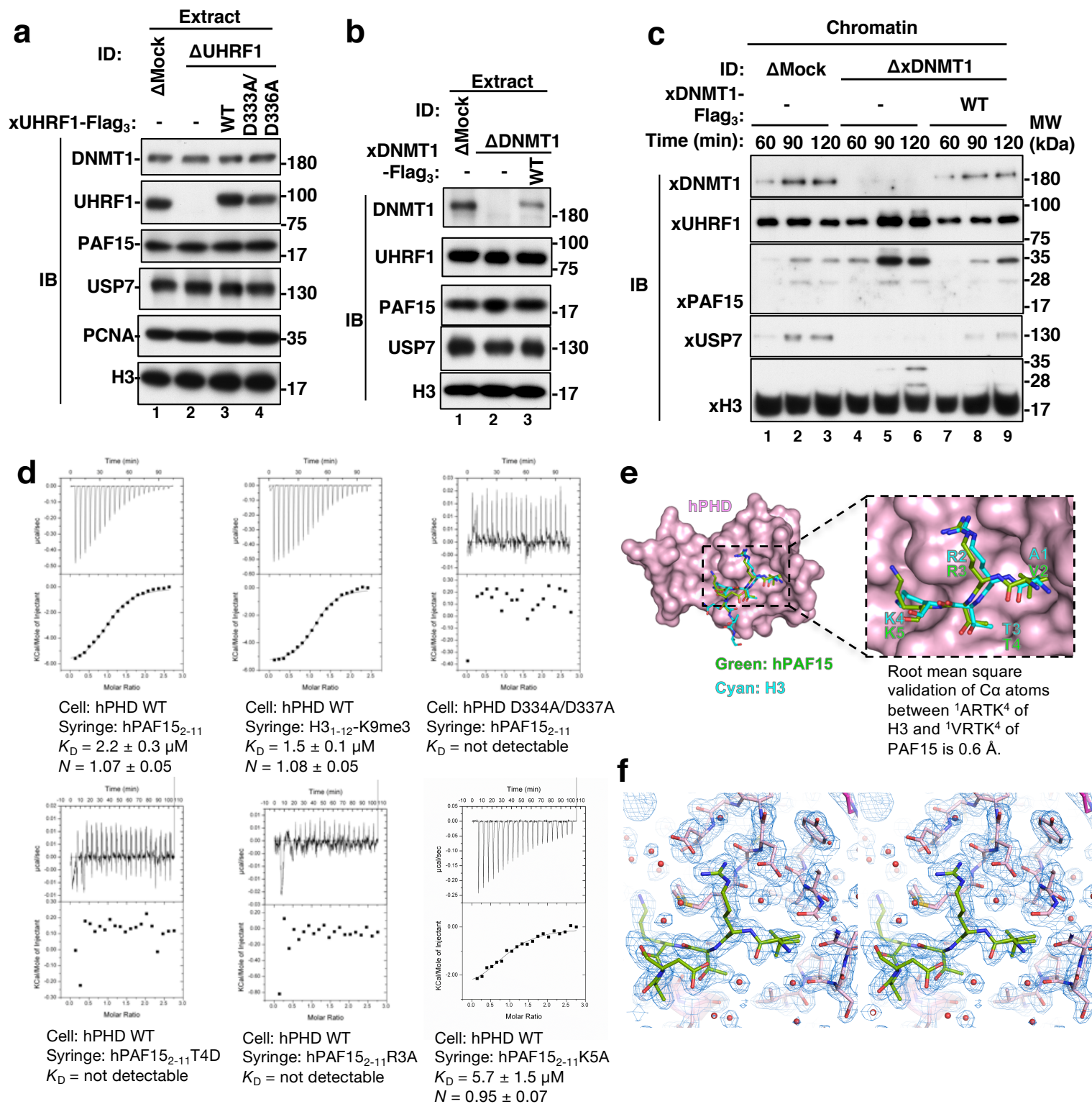
Supplementary Information

**Two distinct modes of DNMT1 recruitment ensure
stable maintenance DNA methylation**

Nishiyama et al.

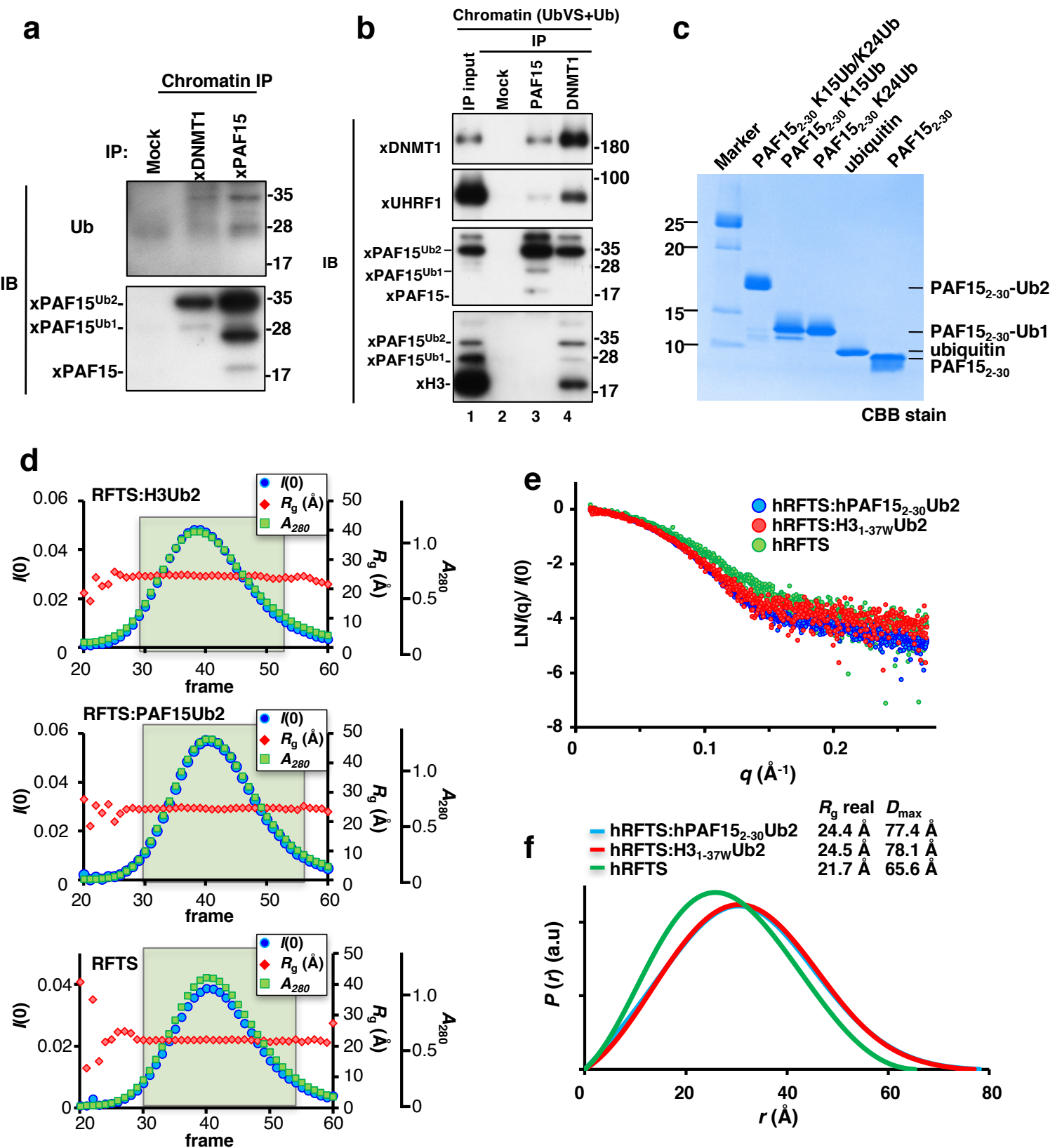


Supplementary Fig. 1 | Isolation of ubiquitin signaling-dependent DNMT1-interacting proteins from chromatin lysates, and characterization of xPAF15. **a**, Schematic of assay to isolate DNMT1-interacting proteins from chromatin lysates. **b**, 1 μ g of recombinant xDNMT1-Flag₃ (lane 1), and BSA were separated by SDS-PAGE, and gel was stained with CBB. **c**, Sperm chromatin was added to interphase egg extracts pretreated with UbVS (14 μ M) in the presence or absence of free ubiquitin (58 μ M). Isolated chromatin fractions were subjected to MNase digestion and solubilized proteins were analyzed by immunoblotting using anti-histone H3 antibody. **d**, Chromatin lysates were subjected to a pull-down experiment using Flag-tagged recombinant wild-type xDNMT1 coupled with anti-Flag M2 beads. The resultant immunoprecipitates were analyzed by immunoblotting using anti-H3 antibody. **e**, The domain structure of xPAF15 and sequence alignment of the conserved ubiquitylation sites across different species. Ubiquitylation sites are shown in red. **f** and **g**, Extracts used in Fig. 1b and Supplementary Fig. 1h were analyzed by immunoblotting using indicated antibodies. **h**, Sperm chromatin was replicated in interphase egg extracts containing buffer (lanes 1-3) or 14 μ M UbVS in the absence (lanes 4-6) or presence of 58 μ M recombinant ubiquitin (lanes 7-9), and chromatin-associated proteins were analyzed by immunoblotting using the indicated antibodies. **i**, Immunoprecipitates from *Xenopus* interphase egg extracts using anti-xPAF15 (lane 3), anti-xPCNA (lane 4), anti-xUHRF1 (lane 5), and anti-xDNMT1 (lane 6) antibodies, or control IgG (lane 2) as well as egg extracts (lane 1) were subjected to immunoblotting using the antibodies indicated. **j**, Sequence alignment of the PIP box of PAF15. A PIP box of *Xenopus* DNA ligase 1 (Lig1) is also aligned. Red residues in the PIP boxes are conserved. GST-tagged full-length xPAF15 wild-type (WT, lane 3), F72AF73A(FF/AA, lane 4), or K18RK27R(KRK R, lane 5) were immobilized on GSH beads and incubated with interphase egg extracts. Bound proteins were analyzed by immunoblotting with PCNA antibodies. Source data are provided as a Source Data file.



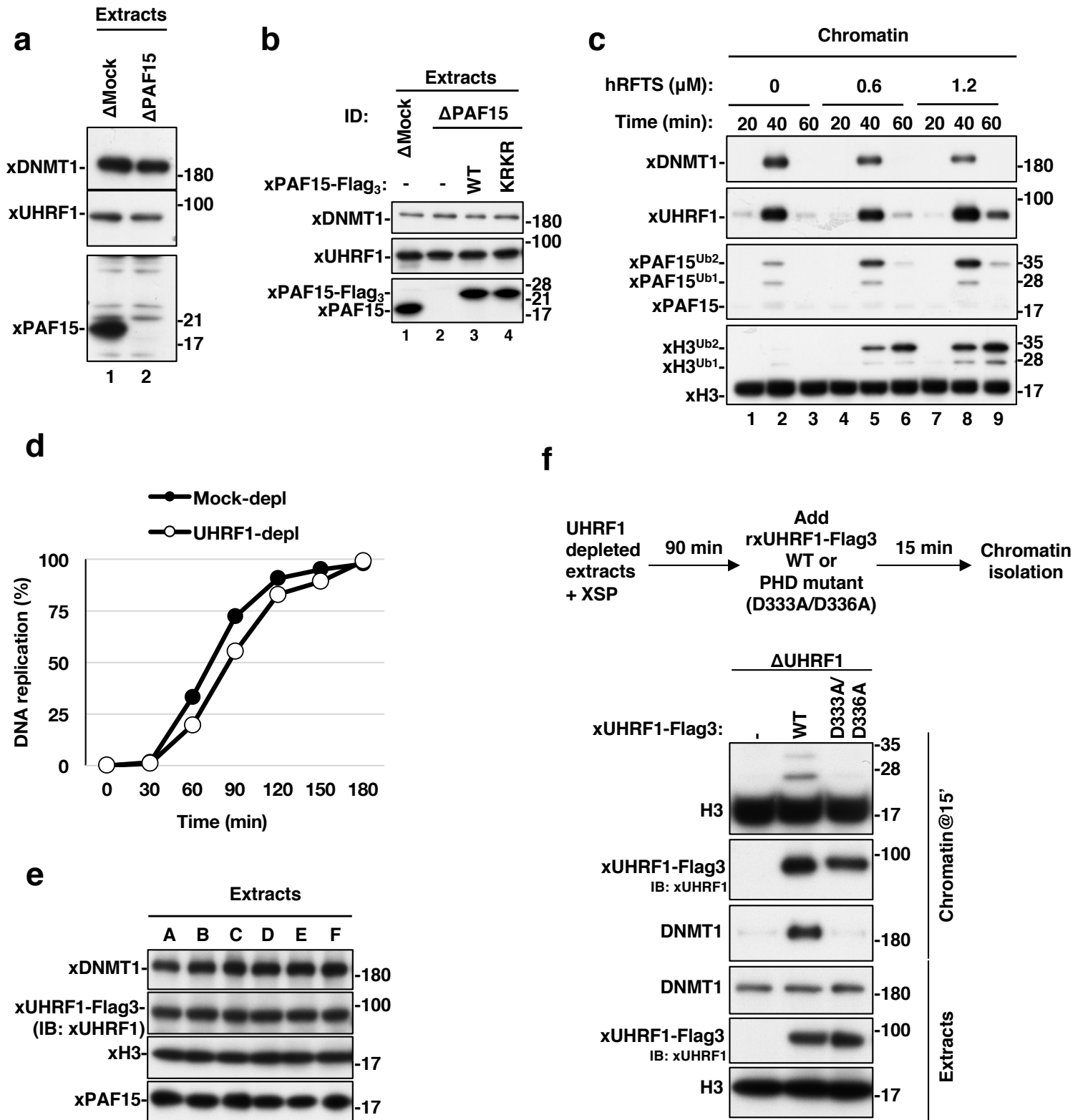
Supplementary Fig. 2 | UHRF1-dependent regulation of PAF15 in egg extracts, ITC thermograms and a structural comparison

a, Mock- and UHRF1-depleted extracts used as shown in Fig. 2a were analyzed by immunoblotting using the indicated antibodies. **b**, Mock- and DNMT1-depleted extracts used in **c** were analyzed by immunoblotting using the indicated antibodies. **c**, Mock-depleted or DNMT1-depleted extracts were supplemented with the indicated recombinant proteins (wt xDNMT1) and chromatin was isolated. Chromatin-bound proteins were analyzed by immunoblotting using the indicated antibodies. **d**, Representative ITC thermograms (upper) and plots of corrected heat values (lower) for the indicated binding experiments. The first data point of each measurement was omitted from the plots in the lower panels and parameter fittings. **e**, Crystal structure of hPHD (pink surface model) in complex with PAF15₂₋₁₁ is superposed on that in complex with H3 peptide (PDB: 3ASL [<https://www.rcsb.org/structure/3ASL>]). Close-up view shows a structural comparison of the 1-4 N-terminal residues of H3 (cyan) and PAF15 (green). **f**, Stereo view of the PAF15 recognition site of hPHD. hPHD and PAF15 are depicted as light-pink and green stick models, respectively. water molecules are shown as red sphere. $2|F_o| - |F_c|$ map contoured at 1.0σ (light-blue) is superimposed on the models. Source data are provided as a Source Data file.



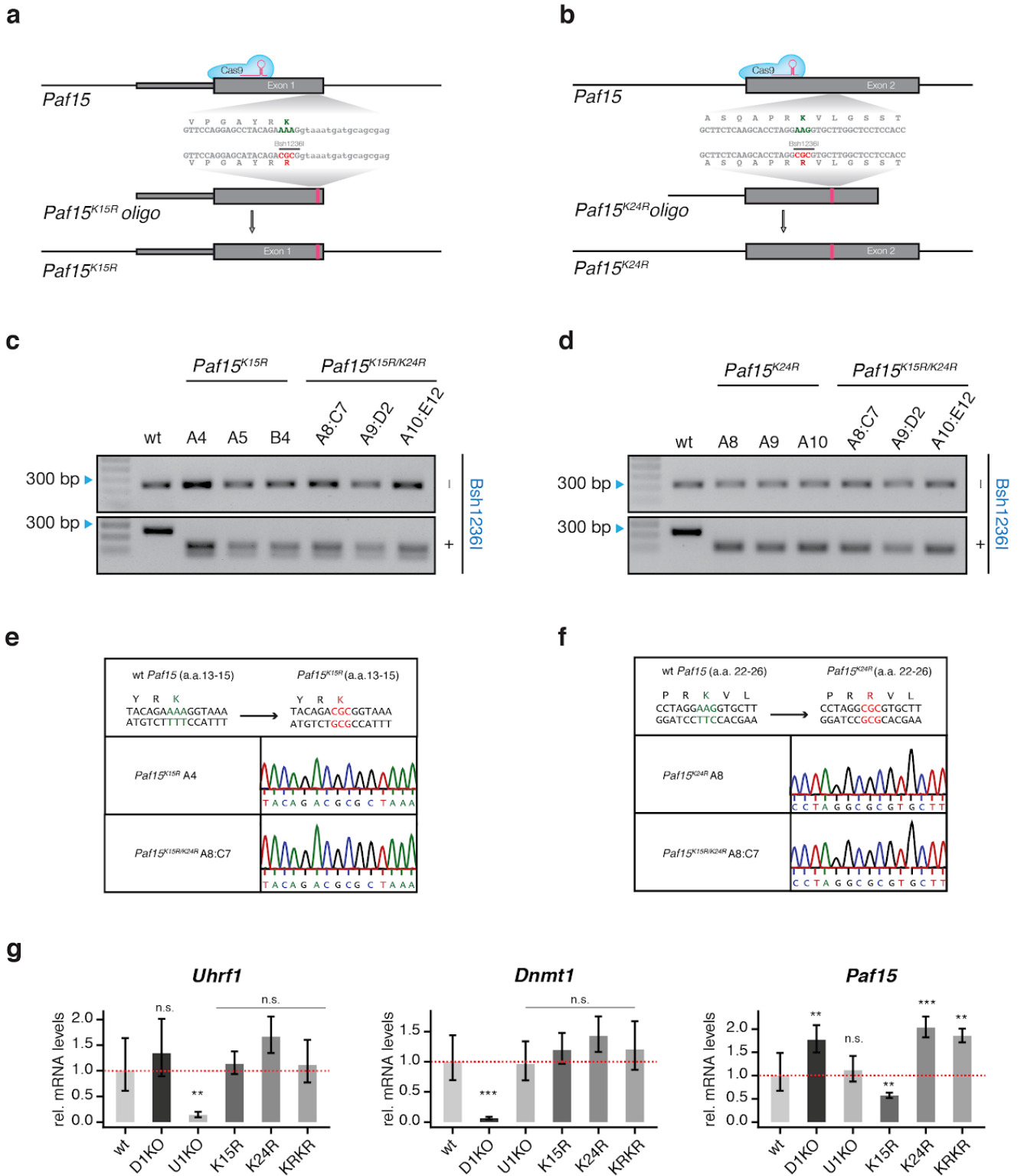
Supplementary Fig. 3 | DNMT1 specifically interacts with two mono-ubiquitylated PAF15

a, Immunoprecipitates from chromatin lysates (Chromatin-IP) with control (Mock), anti-xDNMT1 (DNMT1), or anti-xPAF15 (PAF15) antibody were analyzed by immunoblotting. **b**, Sperm chromatin was replicated in interphase egg extracts containing 14 μ M UbVS and 0.2 mg/ml ubiquitin. Isolated and solubilized chromatin proteins were subjected to immunoprecipitation using anti-PAF15 or DNMT1 antibodies. The resultant immunoprecipitates were analyzed by immunoblotting using the indicated antibodies. **c**, Purity check of disulfide-mediated ubiquitylated PAF15 analyzed by SDS-PAGE under oxidative condition. **d**, R_g (blue) and $I(0)$ (red) plots for SEC-SAXS of hRFTS:H3_{1-37W}Ub2 (top), hRFTS:PAF15₂₋₃₀Ub2 (middle), and RFTS (bottom). X-ray scattering frames highlighted as green were used for extrapolation to zero-concentrations. **e**, Experimental X-ray scattering curves of hRFTS:hPAF15₂₋₃₀Ub2 (cyan circle), hRFTS:H3_{1-37W}Ub2 (red circle) and apo-hRFTS (green circle). The q -range is scattering curves are collected from 0.0113 to 0.2729 \AA^{-1} . Vertical and horizontal axes indicate absolute intensity $\ln(I(q)/I(0))$ and scattering angle $q = 4\pi\sin\theta/\lambda$, respectively. **f**, Pair distance distribution functions $P(r)$ of hRFTS:PAF15₂₋₃₀Ub2 (cyan), hRFTS:H3_{1-37W}Ub2 (red) and hRFTS alone (green) determined from SAXS data. The $P(r)$ functions were normalized by $I(0)$ calculated from each scatter plot. Source data are provided as a Source Data file.



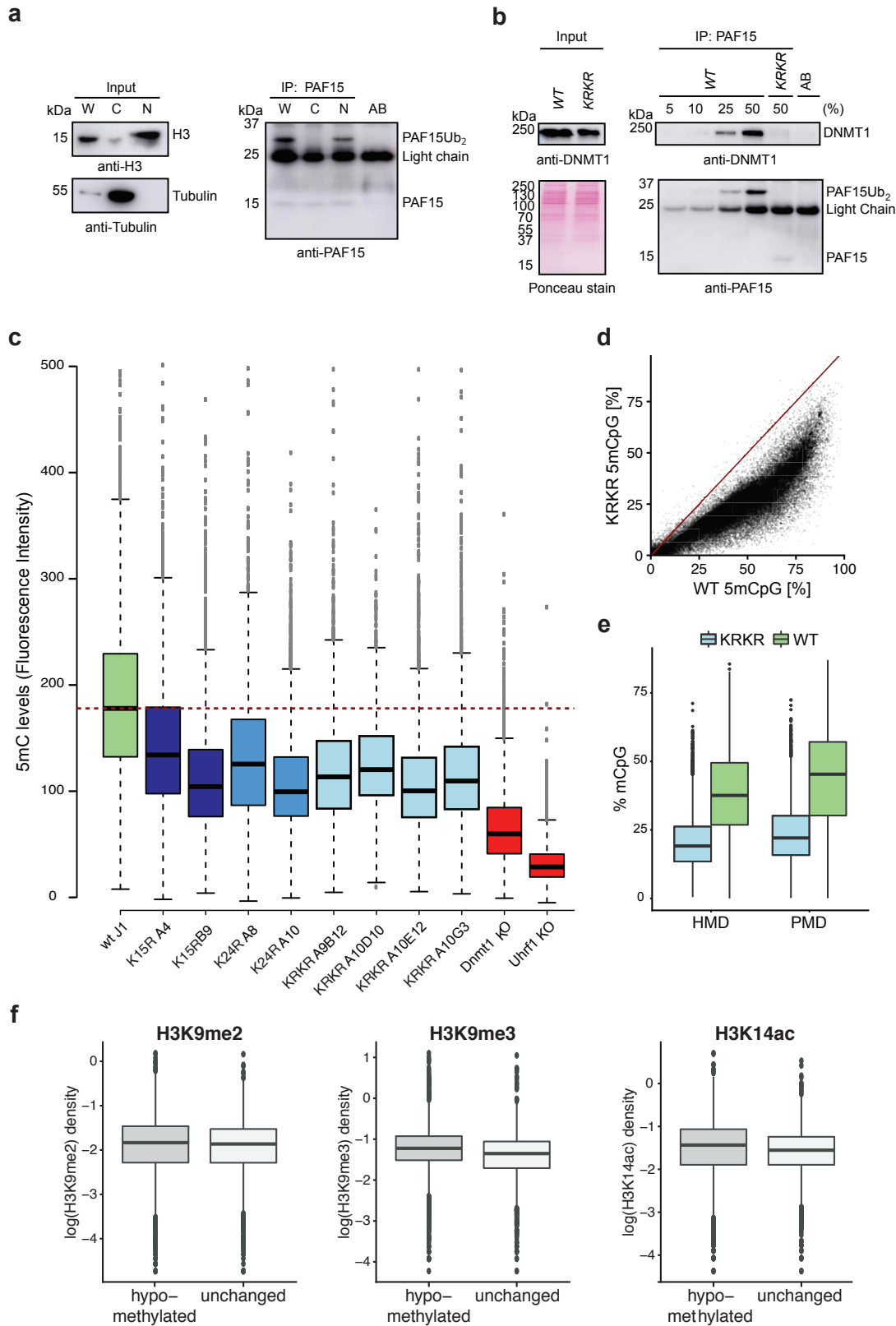
Supplementary Fig. 4 | Immunodepletion of xPAF15 or UHRF1 from *Xenopus* interphase egg extracts and effect of recombinant hRFTS addition to *Xenopus* egg extracts

a, 0.25 μl of mock-depleted or xPAF15-depleted interphase egg extracts were immunoblotted with the indicated antibodies. **b**, Mock- and PAF15-depleted extracts used in Figure 4b were analyzed by immunoblotting using the indicated antibodies. **c**, Interphase egg extracts containing 0.6 or 1.2 μM hRFTS were treated with sperm chromatin. Chromatin fractions were isolated at the indicated times and analyzed by immunoblotting. **d**, Sperm chromatin was incubated in the indicated interphase egg extracts in the presence of [α -³²P]dCTP. The percentage of input DNA replicated at various times is plotted. **e**, Extracts used in Fig. 4f and g were analyzed by immunoblotting using the indicated antibodies. **f**, Schematic of experimental approach to test the role of UHRF1 in histone H3 ubiquitylation. Sperm chromatin was added to xUHRF1-depleted extracts which were incubated for 90 min. The extracts were then supplemented with recombinant xUHRF1-Flag₃-WT or -D333A/D336A and further incubated for 15 min. Chromatin fractions were isolated and chromatin-bound proteins were analyzed by immunoblotting using the antibodies indicated. Extracts were also analyzed by immunoblotting. Source data are provided as a Source Data file.



Supplementary Fig. 5 | Generation and characterization of *Paf15*^{K15R}, *Paf15*^{K24R}, and *Paf15*^{K15R/K24R} (KRKR) mutant mESCs.

a,b. Schematic representation of the CRISPR/Cas9 gene editing strategy used for generating *Paf15* (a) K15R and (b) K24R substitutions in mouse embryonic stem cells (mESCs). Restriction enzyme recognition sites generated by gene editing for restriction fragment length polymorphism (RFLP) screening and amino acid substitutions are shown. **c,d.** Genotyping of (c) *Paf15*^{K15R} and (d) *Paf15*^{K24R} mutant clones via RFLP analysis. **e,f.** Confirmation of successful insertion of (e) K15R and (f) K24R substitutions in the indicated *Paf15* mutants as assessed by Sanger sequencing. **g.** Expression of *Uhrf1*, *Dnmt1*, and *Paf15* in wt, *Dnmt1* KO (D1KO), *Uhrf1* KO (U1KO), *Paf15*^{K15R} (K15R), *Paf15*^{K24R} (K24R) and *Paf15*^{K15R/K24R} (KRKR) mESCs as assessed by qRT-PCR. Error bars represent means \pm SD from n=3 biological replicates. **P < 0.01, ***P < 0.001, P-values calculated by two-tailed Student's t-test between wild-type and the indicated mutant. n.s., not significant. Source data are provided as a Source Data file.



Supplementary Figure 6: PAF15 dual-monoubiquitylation is critical for interacting with DNMT1 and maintaining global DNA methylation in mouse ESCs

a, Whole-cell (W), cytoplasmic (C), and nuclear (N) extracts from wild-type mESCs were subjected to immunoprecipitation of endogenous mPAF15 using an anti-mPAF15 antibody. mPAF15 was detected in the bound fraction with an anti-mPAF15 antibody. mH3 and mTubulin blots were used as indicators of successful nuclear and cytoplasmic fractionation. AB, antibody. **b**, Immunoprecipitation of endogenous mPAF15 from wild-type (WT) and *Paf15* KRKR (KRKR) mESC nuclear extracts using an anti-mPAF15 antibody. Bound fractions were subjected to immunoblotting with anti-mDNMT1 and anti-mPAF15 antibodies. The anti-mDNMT1 blot and Ponceau staining are shown as loading controls. Prior to loading, anti-mPAF15 immunoprecipitated material from wild-type ESCs was titrated (percentage indicated) to achieve levels of mPAF15 comparable to those from *Paf15* KRKR ESCs. AB, antibody. **c**, Quantification of anti-5mC staining in wild-type, *Dnmt1* KO, *Uhrf1* KO ESCs and two independent clones of *Paf15* K15R (K15R), *Paf15* K24R (K24R) and *Paf15* K15R/K24R (KRKR) mutant ESCs with $n = 500$ -2500 cells per replicate. **d**, Comparison of the DNA methylation levels of individual CpG sites in wild-type and *Paf15* KRKR ESCs. **e**, DNA methylation levels of highly methylated domains (HMDs) and partially methylated domains (PMDs) in *Paf15* KRKR and wild-type (WT) ESCs. **f**, H3K9me2, H3K9me3 and H3K14ac density (\log_{10}) at hypomethylated and unchanged tiles of *Paf15* KRKR mESCs. Differentially methylated tiles losing DNA methylation (hypomethylated tiles) were defined as those with $P < 0.05$ and a methylation loss $> 25\%$; P-values were derived from methylKit package (see Materials and Methods). For the boxplots in c, e and f, the horizontal black lines within boxes represent median values, boxes indicate the upper and lower quartiles, and whiskers indicate the 1.5x interquartile range. Source data are provided as a Source Data file.

Supplementary Table 1. SAXS data collection parameters

	hRFTS: hPAF15 ₂₋₃₀ Ub2	hRFTS: H3 _{1-37W} -Ub2	hRFTS	Ovalbumin
Data collection parameters				
Instrument	Photon Factory BL-10C			
Wavelength (Å)	1.5			
q range (Å ⁻¹)	0.0069–0.2781			
Detector	Pilatus3 2 M			
Detector distance (mm)	2,027			
Exposure (s per image)	20			
SEC Column	5/150GL INCREASE Superdex200			
Flow rate (mL.min ⁻¹)	0.025			
Injected sample conc. (mg.mL ⁻¹)	6.0			
Injection volume (µL)	50			
Temperature (K)	277			
Structural parameters (data of extrapolated to zero concentration)				
R_g (Å) [from $P(r)$]	24.4 ± 0.2	24.5 ± 0.2	21.7 ± 0.1	24.3 ± 0.1
R_g (Å) [from Guinier]	24.4 ± 0.2	24.3 ± 0.2	21.7 ± 0.2	24.1 ± 0.2
D_{max} (Å)	77.4	78.1	65.6	75.1
Porod volume estimate (Å ³)	59,300	57,700	36700	63,300
Molecular-mass determination				
¹ Molecular mass M_r [$I(0)/c$ of standard]	43.9 ± 1	43.9 ± 3	29.6 ± 1	-
² $I(0)/c$ [from Guinier]	0.046204 ± 0.001351	0.046238 ± 0.003432	0.031144 ± 0.001024	0.04679
Calculated M_r from sequence	48.0	48.9	28.0	44.4
Software employed				
Primary data reduction	SAnGLAR			
Data processing	PRIMUS			
<i>Ab initio</i> analysis	DAMMIF			
Validation and averaging	DAMAVER & DAMMIN			
Computation of model intensities	CRYSOL			

¹average and deviation are derived from the indicated frames in Supplementary Fig. 3d.

²sample concentration (mg/ml)

Supplementary Table 2: RRBS Information				
Sample_ID	Genotype	# total reads	#mapped reads	% of mapped reads
KRKR_1	PAF15_KRKR_ESC	28,140,727.00	19135694	68
KRKR_2	PAF15_KRKR_ESC	35,670,848.00	26753136	75
WT-1	wt_J1_ESC	33,012,500.00	24429250	74
WT-2	wt_J1_ESC	21,910,914.00	15118530	69
D1KO_r1	DNMT1_KO_ESC	3,686,943.00	2617730	71
D1KO_r2	DNMT1_KO_ESC	3,846,150.00	2499998	65
D1KO_r3	DNMT1_KO_ESC	3,688,377.00	2508097	68
U1KO_r1	UHRF1_KO_ESC	3,248,968.00	2436726	75
U1KO_r2	UHRF1_KO_ESC	3,581,465.00	2399582	67
U1KO_r3	UHRF1_KO_ESC	3,680,511.00	2539553	69
wt_r1	wt_J1_ESC	3,229,351.00	2615775	81
wt_r2	wt_J1_ESC	2,956,850.00	2247206	76
wt_r3	wt_J1_ESC	3,081,518.00	2280324	74

combined coverage of all samples over genomic elements	
Genomic Element	coverage [fraction of total]
Repeats	0.013915533
CpG islands	0.610142975
Promoters	0.430718894
Gene bodies	0.445765294

Supplementary Table 3: Oligonucleotides used in this study

Name	Sequence 5'-3'	Description
3621	ATAGCGCTGGAGGGAATTCAGTGTAAACGCA	xPAF15 amplification
3622	AAAAGCAGCATGAATGCTCTAGTCCAGGCTT	xPAF15 amplification
3667	CGTGGATCCCCGAATTCATGGTGGGACTAAGGCAGA	GST-xPAF15
3636	GGCCGCTCGAGTCGATTATTTACAAATATACAAAGC	GST-xPAF15
3720	ggcgcggatcagatctcATGGTGGGACTAAGGCAGACT	pVL1392-xPAF15-Flag3
3581	GGGCCCTCTAGAATTCTACTTGTATCGTCATCCT	pVL1392-xPAF15-Flag3
3707	AgGGCTGTTGCTGCCAGAGCACCAAGGA	xPAF15 K18R mutation
3708	CCTGTAGCTCCCCGATGAAGAGCCC	xPAF15 K18R mutation
3709	AgAACATTTGGGAGTAGTTCAGTGGTT	xPAF15 K27R mutation
3710	CCTTGGTGTCTGGCAGCAACAGCC	xPAF15 K27R mutation
3711	gcCgcTGGATCACCATCCACAAGTCAGCCTG	xPAF15F72AF73A mutation
3712	GTCTCCTATACCTTTTCTGCCAGGTA	xPAF15F72AF73A mutation
3896	gcGACTAAGGCAGACTGCGCGGGCTCTT	pKS104-xPAF15R3A mutation
3897	CACCATGAATTCTCGAGTGCAAAAA	xPAF15R3A mutation
3898	CGGgaTAAGGCAGACTGCGCGGGCTCTT	xPAF15T3D mutation
3899	CGGACTgcGGCAGACTGCGCGGGCTCTT	xPAF15K4A mutation
4449	CACCATGAGATCTGATCCGCGCCCG	pVL1392-xPAF15R3A mutation
4620	GGCGCGGATCAGATCTCATGTGGATACAGGTGCGTAC	pVL1392-xUHRF1-Flag3
4271	GCCATGGCGTTTCACATTTATGCGCTTA	xUHRF1D333A mutation
4272	ACACTCATCACAAGAAGTTGTTTC	xUHRF1D333A mutation
4273	GCTGAGTGTGCCATGGCGTTTCACATTT	xUHRF1D333AD336A mutation
4274	ACAAAGAAGTTGTTTCTCTGGGTCC	xUHRF1D333AD336A mutation
Dnmt1_F	GGCGGAAATCAAAGGAGGAT	RT-qPCR
Dnmt1_R	CCTGGGCTGGAACCTCTTTTATC	RT-qPCR
Uhrf1_F	GGCAGCTGAAGCGGATGA	RT-qPCR
Uhrf1_R	CCATGCACCGAAGATATTGTCA	RT-qPCR

PAF15_F	CAAGTTCGTCGAGAAAAGCTGA	RT-qPCR
PAF15_R	ACAGCCTGAAGAATTCCCCG	RT-qPCR
Gapdh_F	CATGGCCTTCCGTGTTCTTA	RT-qPCR
Gapdh_R	CTTACCACCTTCTTGATGTCATC	RT-qPCR
PAF15_K15_gRNA_F	CACCGCATCATTTACCTTTTCTGT	Cloning gRNAs
PAF15_K15_gRNA_R	aaacACAGAAAAGTAAATGATGC	Cloning gRNAs
PAF15_K24_gRNA_F	CACCGAGCCAAGCACCTTCCTAG	Cloning gRNAs
PAF15_K24_gRNA_R	aaacCTAGGAAGGTGCTTGGCTCC	Cloning gRNAs
PAF15_K15_scrF	CGGGAAGAGACCCATTTAAAC	PCR and RFLP Analysis
PAF15_K15_scrR	GCCTTCTAGTGCTCAATGG	PCR and RFLP Analysis
PAF15_K24_scrF	CTGGCCTGGGACTGTTGTAG	PCR and RFLP Analysis
PAF15_K24_scrR	CAGGTTAGTACTGCCTTGCC	PCR and RFLP Analysis
Paf15_K15R_Donor	CCTGCCTTCTAGTGCTCAATGGGAGGACCCATGGGCGTCTCCACCCCT GGACAGGCTGCCTAGGGAACCCCTGCCACCTCGCTGCATCATTTACCgc gTCTGTATGCTCCTGGAACGTAGTTTGTCTTGGTCCGCACCATGTTTACA CAAGAAGAGACAACCTTTCACCGTCACCCCAACTGCAGATGTCTCAATTAG	ssDNA Donor Oligonucleotides
Paf15_K24R_Donor	ATGCTCTCGGGGTGTTACTTTCAGAAGCTTCCACGACCTTCTTACCTTTT CTCGACGAACTTGAAGAATTGGTGACAAAGGTGGAGGAGCCAAGCACgcg CCTAGGTGCTTGAGAAGCCACCGCTGCAGAGAGAGATAAATAGGGCGTT CAGAAAAGGCAGGAGGTTTCGGATCCCCGAGCTTTGTTCTACAACAGTC	ssDNA Donor Oligonucleotides
hPAF15_F	ATGGTGCGGACTAAAGCAGACAGTGTCCAGGCACCTACAGAAAA	hPAF15 amplification
hPAF15_R	CTATTCTTTTTCATCATTTGTGTGATCAGGTTGCAAAGGACATGC	hPAF15 amplification
hPAF15_72stop_F	TAGGTTGTAACCTAAAGATTCTGAAAAAGA	hPAF15 ₂₋₇₁
hPAF15_72stop_R	CTTTAGGTTTACAACCTAAAGAATTCTCCAA	hPAF15 ₂₋₇₁
N ^{PfO} -hPAF15_F	CATTATGGGTAAGTGTGCGTGCAGCTAAAGCAGACAGTGTTC	pET21b-N ^{PfO} -hPAF15 ₂₋₇₁
N ^{PfO} -hPAF15_R	TCCGACATTTGGTCTTATTACAACCTAAAGAATTCTCCAATTCCT	pET21b-N ^{PfO} -hPAF15 ₂₋₇₁
hPAF15_K15C/K24C-F	TACAGATGTGTGGTGGCTGCTCGAGCCCCAGATGTGTGCTTGGTCTTC	hPAF15 K15C/K24C
hPAF15_K15C/K24C-R	AAGCACACATCTGGGGCTCGAGCAGCCACCACACATCTGTAAGTGCCTG	mutation
hPAF15_C54S-F	CCCCGTTTCCGTGCGCCCAACTCCCAAGTG	hPAF15 C54S mutation
hPAF15_C54S-R	GGCGCACGGAACGGGGTTCCCTCCTGCAT	hPAF15 C54S mutation
PAF15_HA_F	TACCCATACGATGTCCAGATTACGCTTAATAAGACCAAATGTCGGATCCACTAGTG	hPAF15 ₂₋₇₁ -HA
PAF15_HA_R	AGCGTAATCTGGAACATCGTATGGGTACAACCTAAAGAATTCTCCAATTCCTTTTGG	hPAF15 ₂₋₇₁ -HA

PAF15_R2A_F	TTGCGTGGCCACTAAAGCAGACAGTGTTC	hPAF15 R2A mutation
PAF15_R2A_R	CTTTAGTGGCCACGCAACTAGTTACCCATA	hPAF15 R2A mutation
PAF15_T3D_F	CGTGCGGGACAAAGCAGACAGTGTCCAGG	hPAF15 T3D mutation
PAF15_T3D_R	CTGCTTGTCCCGCAGCAACTAGTTACCC	hPAF15 T3D mutation
PAF15_K4A_F	GCGGACTGCCGCAGACAGTGTCCAGGCAC	hPAF15 K4A mutation
PAF15_K4A_R	TGTCTGCGCAGTCCGCACGCAACTAGTTA	hPAF15 K4A mutation
UBCh5a-F	ATGGCGCTGAAGAGGATTCAGAAAGAATTGAGTGATCTACAGCGC	hUBCh5 amplification
UBCh5a-R	TTACATTGCATATTTCTGAGTCCATTCTCTGCATGTCTGTTGTA	hUBCh5 amplification
UHRF1_D334A/D337A_F	TGTGCGCTGAGTGCGCCATGGCCTTCCACA	hUHRF1 D334A and D337A mutation
UHRF1_D334A/D337A_R	CCATGGCGCACTCAGCGCACATGAGCTGCT	
UHRF1_H741A_F	GTGCCAGGCCAACGTGTGCAAGGACTGCCT	hUHRF1 H741A mutation
UHRF1_H741A_R	ACACGTTGGCCTGGCACACGGTCGTGATGG	

Description of Additional Supplementary Files

File Name: Supplementary Data 1 (41467_2020_15006_MOESM5_ESM.xlsx)

Description: This file contains the MS-based quantification of DNMT1 interacting chromatin proteins.

File Name: Supplementary Data 2 (41467_2020_15006_MOESM6_ESM.xlsx)

Description: This file contains the list of identified ubiquitylated peptides in xDNMT1 pull-down.

File Name: Supplementary Data 3 (41467_2020_15006_MOESM7_ESM.xlsx)

Description: This file contains the list of identified phosphorylated peptides in xDNMT1 pull-down.

2.3 Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive demethylation in mammals

– manuscript in preparation –

Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive demethylation in mammals

Authors:

Christopher B. Mulholland¹, Atsuya Nishiyama^{2†}, Joel Ryan^{1†}, Ryohei Nakamura³, Merve Yiğit¹, Ivo M. Glück⁴, Carina Trummer¹, Weihua Qin¹, Michael D. Bartoschek¹, Franziska R. Traube⁵, Edris Parsa⁵, Enes Ugur^{1,8}, Miha Modic⁷, Aishwarya Acharya¹, Paul Stolz¹, Christoph Ziegenhain⁶, Michael Wierer⁸, Wolfgang Enard⁶, Thomas Carell⁵, Don C. Lamb⁴, Hiroyuki Takeda³, Makoto Nakanashi², Sebastian Bultmann^{*1§} and Heinrich Leonhardt^{*1§}

Affiliations:

¹ Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Human Biology and BiImaging, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany

² Division of Cancer Cell Biology, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

³ Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

⁴ Physical Chemistry, Department of Chemistry, Center for Nanoscience, Nanosystems Initiative Munich and Center for Integrated Protein Science Munich, Ludwig-Maximilians-Universität München, Munich, Germany

⁵ Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany

⁶ Department of Biology II, Anthropology and Human Genomics, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany

⁷ The Francis Crick Institute and University College London, Institute of Neurology, London, United Kingdom

⁸ Department of Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, Am Klopferspitz 18, 82152, Martinsried, Germany.

[†]These authors contributed equally

[§]These authors jointly supervised this work

*Correspondence and requests for material should be addressed to S.B. (bultmann@bio.lmu.de) or H.L. (h.leonhardt@lmu.de)

ABSTRACT

Genome-wide DNA demethylation is a unique feature of mammalian development and naïve pluripotent stem cells. So far, it was unclear how mammals specifically achieve global DNA hypomethylation, given the high conservation of the DNA (de-)methylation machinery among vertebrates. We found that DNA demethylation requires TET activity but mostly occurs at sites where TET proteins are not bound suggesting a rather indirect mechanism. Among the few specific genes bound and activated by TET proteins was the naïve pluripotency and germline marker *Dppa3* (*Pgc7*, *Stella*), which undergoes TDG dependent demethylation. The requirement of TET proteins for genome-wide DNA demethylation could be bypassed by ectopic expression of *Dppa3*. We show that DPPA3 binds and displaces UHRF1 from chromatin and thereby prevents the recruitment and activation of the maintenance DNA methyltransferase DNMT1. We demonstrate that DPPA3 alone can drive global DNA demethylation when transferred to amphibians (*Xenopus*) and fish (medaka), both species that naturally do not have a *Dppa3* gene and exhibit no post-fertilization DNA demethylation. Our results show that TET proteins are responsible for active and - indirectly also for - passive DNA demethylation; while TET proteins initiate local and gene-specific demethylation in vertebrates, the recent emergence of DPPA3 introduced a unique means of genome-wide passive demethylation in mammals and contributed to the evolution of epigenetic regulation during early mammalian development.

INTRODUCTION

During early embryonic development the epigenome undergoes massive changes. Upon fertilization the genomes of highly specialized cell types - sperm and oocyte - need to be reprogrammed in order to obtain totipotency. This process entails decompaction of the highly condensed gametic genomes and global resetting of chromatin states to confer the necessary epigenetic plasticity required for the development of a new organism (Ladstätter and Tachibana, 2019). At the same time the genome needs to be protected from the activation of transposable elements (TEs) abundantly present in vertebrate genomes (Warren et al., 2015). Activation and subsequent transposition of TEs results in mutations that can have deleterious effects and are passed onto offspring if they occur in the germline during early development (Arkhipova, 2018; Warren et al., 2015). The defense against these genomic parasites has shaped genomes substantially (Friedli and Trono, 2015; Jacobs et al., 2014).

DNA methylation in vertebrates refers to the addition of a methyl group at the C5 position of cytosine to form 5-methylcytosine (5mC). Besides its important role in gene regulation, the most basic function of DNA methylation is the repression and stabilization of TEs and other repetitive sequences (Rowe and Trono, 2011). Accordingly, the majority of genomic 5mC is located within these highly abundant repetitive elements. Global DNA methylation loss triggers the derepression of transposable and repetitive elements, which leads to genomic instability and cell death, highlighting the crucial function of vertebrate DNA methylation (Chernyavskaya et al., 2017; Chiappinelli et al., 2017; Iida et al., 2006; Jackson-Grusby et al., 2001; Roulois et al., 2015; Walsh et al., 1998). Hence, to ensure constant protection against TE reactivation, global DNA methylation levels remain constant throughout the lifetime of non-mammalian vertebrates (Ortega-Recalde et al., 2019; Skvortsova et al., 2019; Stancheva et al., 2002; Veenstra and Wolffe, 2001). Paradoxically, mammals specifically erase DNA methylation during preimplantation development (Monk et al., 1987; Sanford et al., 1987), a process that would seemingly expose the developing organism to the risk of genomic instability through the activation of TEs. DNA methylation also acts as an epigenetic barrier to restrict and stabilize cell fate decisions and thus constitutes a form of epigenetic memory. The establishment of pluripotency in mammals requires the erasure of epigenetic memory and as such, global hypomethylation is a defining characteristic of pluripotent cell types including naive embryonic stem cells (ESCs), primordial germ cells (PGCs), and induced pluripotent stem cells (iPSCs)(Lee et al., 2014).

In vertebrates, DNA methylation can be reversed to unmodified cytosine by two mechanisms; either actively by Ten-eleven translocation (TET) dioxygenase-mediated oxidation of 5mC in concert with the base excision repair machinery (Cortellino et al., 2011; He et al., 2011; Ito et al., 2011; Tahiliani et al., 2009) or passively by a lack of functional DNA methylation maintenance during the DNA replication cycle (Howlett and Reik, 1991; Rougier et al., 1998). In mammals, both active and passive demethylation pathways have been suggested to be involved in the establishment of global demethylation during preimplantation development (Amouroux et al., 2016; Carlson et al., 1992; Gu et al., 2011; Guo et al.,

2014; Shen et al., 2014; Wang et al., 2014; Wossidlo et al., 2011). Intriguingly however, whereas hypomethylated states during development are mammal-specific, TET-proteins as well as the DNA methylation machinery are highly conserved among all vertebrates (Iyer et al., 2009; Zemach et al., 2010). This discrepancy implies the existence of so far unknown mammalian-specific pathways and factors controlling the establishment and maintenance of genomic hypomethylation.

Here, we use mouse embryonic stem cells (ESCs) cultured in conditions promoting naïve pluripotency (Hackett et al., 2013; Leitch et al., 2013; Marks et al., 2012) as a model to study global DNA demethylation in mammals. By dissecting the contribution of the catalytic activity of TET1 and TET2 we show that TET-mediated active demethylation drives the expression of the Developmental pluripotency-associated protein 3 (DPPA3/PGC7/STELLA). We show that DPPA3 directly binds UHRF1 prompting its release from chromatin, resulting in the inhibition of maintenance methylation and global passive demethylation. Although only found in mammals, DPPA3 can also induce global demethylation in non-mammalian vertebrates. In summary, we described a novel TET-controlled and DPPA3-driven pathway for passive demethylation in naïve pluripotency in mammals.

RESULTS

TET1 and TET2 indirectly protect the naïve genome from hypermethylation.

Mammalian TET proteins, TET1, TET2, and TET3, share a conserved catalytic domain and the ability to oxidize 5mC but exhibit distinct expression profiles during development (Rasmussen and Helin, 2016). Naïve ESCs and the inner cell mass (ICM) of the blastocyst from which they are derived feature high expression of *Tet1* and *Tet2* but not *Tet3* (Boroviak et al., 2015; Ficz et al., 2013; Ito et al., 2010; Wossidlo et al., 2011). To dissect the precise contribution of TET-mediated active DNA demethylation to global DNA hypomethylation in naïve pluripotency we generated isogenic *Tet1* (T1CM) and *Tet2* (T2CM) single as well as *Tet1/Tet2* (T12CM) double catalytic mutant mouse ESC lines using CRISPR/Cas-assisted gene editing (Figure S1). We confirmed the inactivation of TET1 and TET2 activity by measuring the levels of 5-hydroxymethylcytosine (5hmC), the product of TET-mediated oxidation of 5mC (Tahiliani et al., 2009)(Figure S1I). While the loss of either *Tet1* or *Tet2* catalytic activity significantly reduced 5hmC levels, inactivation of both TET1 and TET2 resulted in the near total loss of 5hmC in naïve ESCs (Figure S1I) indicating that TET1 and TET2 account for the overwhelming majority of cytosine oxidation in naïve ESCs. We then used reduced representation bisulfite sequencing (RRBS) to determine the DNA methylation state of T1CM, T2CM, and T12CM ESCs as well as wildtype (wt) ESCs. All *Tet* catalytic mutant (T1CM, T2CM and T12CM) cell lines exhibited severe DNA hypermethylation throughout the genome including promoters, gene bodies, and repetitive elements (Figure 1A and 1B). The increase in DNA methylation was particularly pronounced at LINE-1 (L1) elements of which 97%, 98%, and 99%

were significantly hypermethylated in T1CM, T2CM, and T12CM ESCs, respectively (Figure S2A). This widespread DNA hypermethylation was reminiscent of the global increase in DNA methylation accompanying the transition of naïve ESCs to primed epiblast-like cells (EpiLCs) (Ficz et al., 2013; Habibi et al., 2013; Pfaffeneder et al., 2014), which prompted us to investigate whether the hypermethylation present in T1CM, T2CM, and T12CM ESCs represents a premature acquisition of a more differentiated DNA methylation signature. In line with this hypothesis, *Tet* catalytic mutant ESCs displayed DNA methylation levels similar or higher than those of wt EpiLCs (Figure S2B). Moreover, hierarchical clustering and principal component analyses (PCA) of the RRBS data revealed that ESCs from *Tet* catalytic mutants clustered closer to wt EpiLCs than wt ESCs (Figure 1C and S2C). In fact, the vast majority of significantly hypermethylated CpGs in *Tet* catalytic mutant ESCs overlapped with those normally gaining DNA methylation during the exit from naïve pluripotency (Figure 1D). However, T1CM, T2CM, and T12CM transcriptomes clearly clustered by differentiation stage indicating that the acquisition of an EpiLC-like methylome was not due to premature differentiation (Figure S2D). Intriguingly, we found that the majority of sites hypermethylated in *Tet* catalytic mutant ESCs are not bound by TET1 or TET2 (Figure 1E and 1F) suggesting that the catalytic activity of TET1 and TET2 maintains the hypomethylated state of the naïve methylome by indirect means.

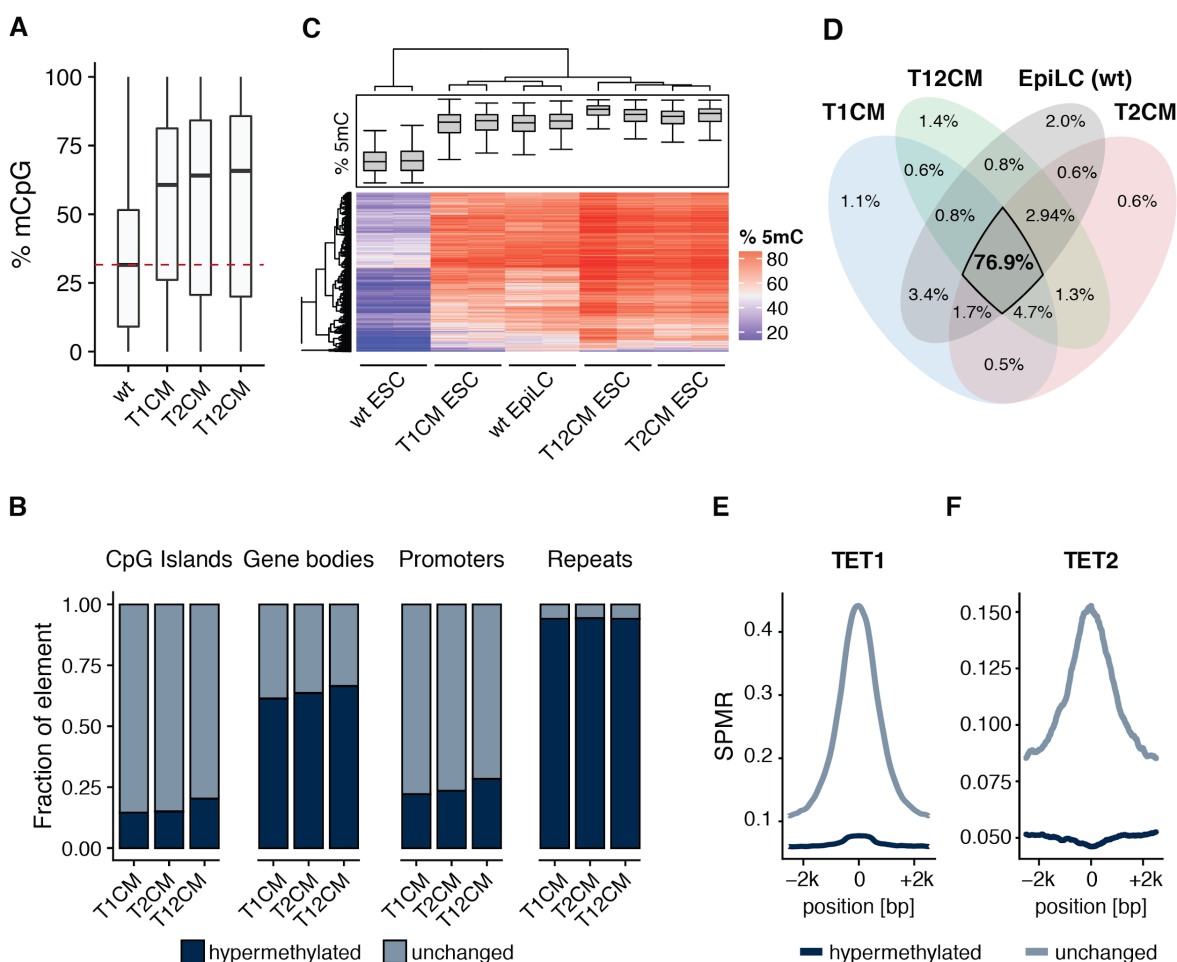


Figure 1: TET1 and TET2 prevent hypermethylation of the naïve genome.

(A) Loss of TET catalytic activity leads to global DNA hypermethylation. Percentage of total 5mC as measured by RRBS. (B) Loss of TET catalytic activity leads to widespread DNA hypermethylation especially at repetitive elements. Relative proportion of DNA hypermethylation (q value < 0.05; absolute methylation difference > 20%) at each genomic element in T1CM, T2CM, and T12CM ESCs compared to wt ESCs. (C) Heat map of the hierarchical clustering of the RRBS data depicting the top 2000 most variable 1kb tiles during differentiation of wt ESCs to EpiLCs. (D) Venn diagram depicting the overlap of hypermethylated (compared to wt ESCs; q value < 0.05; absolute methylation difference > 20%) sites among T1CM, T2CM, and T12CM ESCs and wt EpiLCs. (E and F) TET binding is not associated with DNA hypermethylation in TET mutant ESCs. Occupancy of (E) TET1 (Khoueiry et al., 2017) and (F) TET2 (Xiong et al., 2016) over 1 kb tiles hypermethylated (dark blue) or unchanged (light blue) in T1CM and T2CM ESCs, respectively. In the boxplots in (A and C) horizontal black lines within boxes represent median values, boxes indicate the upper and lower quartiles, and whiskers indicate the 1.5 interquartile range.

TET1 and TET2 control *Dppa3* expression in a catalytically dependent manner.

To understand how TET1 and TET2 indirectly promote demethylation of the naïve genome, we examined the expression of the enzymes involved in DNA methylation. Loss of TET catalytic activity was not associated with changes in the expression of *Dnmt1*, *Uhrf1*, *Dnmt3a*, and *Dnmt3b*, indicating the hypermethylation in *Tet* catalytic mutant ESCs is not a result of the DNA methylation machinery being upregulated (Figure 2A). To identify candidate factors involved in promoting global hypomethylation, we compared the transcriptome of hypomethylated wild-type ESCs with those of hypermethylated cells, which included wt EpiLCs as well as T1CM, T2CM, and T12CM ESCs (Figure 2B). Among the 14 genes differentially expressed in hypermethylated cell lines, the naïve pluripotency factor, *Dppa3* (also known as *Stella* and *Pgc7*), was an interesting candidate due to its reported involvement in the regulation of global DNA methylation in germ cell development and oocyte maturation (Funaki et al., 2014; Li et al., 2018; Nakamura et al., 2007; Nakashima et al., 2013). In addition, *Dppa3* is also a direct target of PRDM14, a PR domain-containing transcriptional regulator known to promote the DNA hypomethylation associated with naïve pluripotency (Leitch et al., 2013; Magnúsdóttir et al., 2013; Okashita et al., 2014; Yamaji et al., 2013) (Figure 2E). While normally highly expressed in naïve ESCs and only downregulated upon differentiation (Hayashi et al., 2008; Kalkan et al., 2017), *Dppa3* was prematurely repressed in T1CM, T2CM, and T12CM ESCs (Figure 2D). The significantly reduced expression of *Dppa3* in TET mutant ESCs was accompanied by significant hypermethylation of the *Dppa3* promoter (Figure 2E), consistent with reports demonstrating *Dppa3* to be one of the few pluripotency factors regulated by promoter methylation *in vitro* and *in vivo* (Auclair et al., 2014; Hackett et al., 2013; Hayashi et al., 2008; Kalkan et al., 2017). In contrast to the majority of genomic sites gaining methylation in TET mutant ESCs (Figure 1E and 1F), hypermethylation at the *Dppa3* locus occurred at sites bound by both TET1 and TET2 (Figure 2E) (Khoueiry et al., 2017; Xiong et al., 2016). This hypermethylation overlapped with regions at which the TET oxidation product 5-carboxylcytosine (5caC) accumulates in Thymine DNA glycosylase (TDG)-knockdown ESCs (Figure 2E) (Shen et al., 2013), indicating that the *Dppa3* locus is a direct target of TET/TDG-mediated active DNA demethylation in ESCs.

PRDM14 has been shown to recruit TET1 and TET2 to sites of active demethylation and establish global hypomethylation in naïve pluripotency (Ficz et al., 2013; Grabloe et al., 2013; Leitch et al., 2013; Okashita et al., 2014; Yamaji et al., 2013). As the expression of *Prdm14* was not altered in *Tet* catalytic mutant ESCs (Figure 2A), we analysed PRDM14 occupancy at the *Dppa3* locus using publicly available ChIP-seq data (Yamaji et al., 2013). This analysis revealed that PRDM14 binds the same upstream region of *Dppa3* occupied by TET1 and TET2 (Figure 2E). Taken together, these data suggest that TET1 and TET2 are recruited by PRDM14 to maintain the expression of *Dppa3* by active DNA demethylation. Strikingly, systematic analysis of public databases revealed that while the DNA (de)methylation (DNMTs, UHRF1, TETs) is conserved among vertebrates *Dppa3* is only present in mammals potentially representing a novel pathway that regulates mammalian-specific global hypomethylation in naïve pluripotency (Figure 2C).

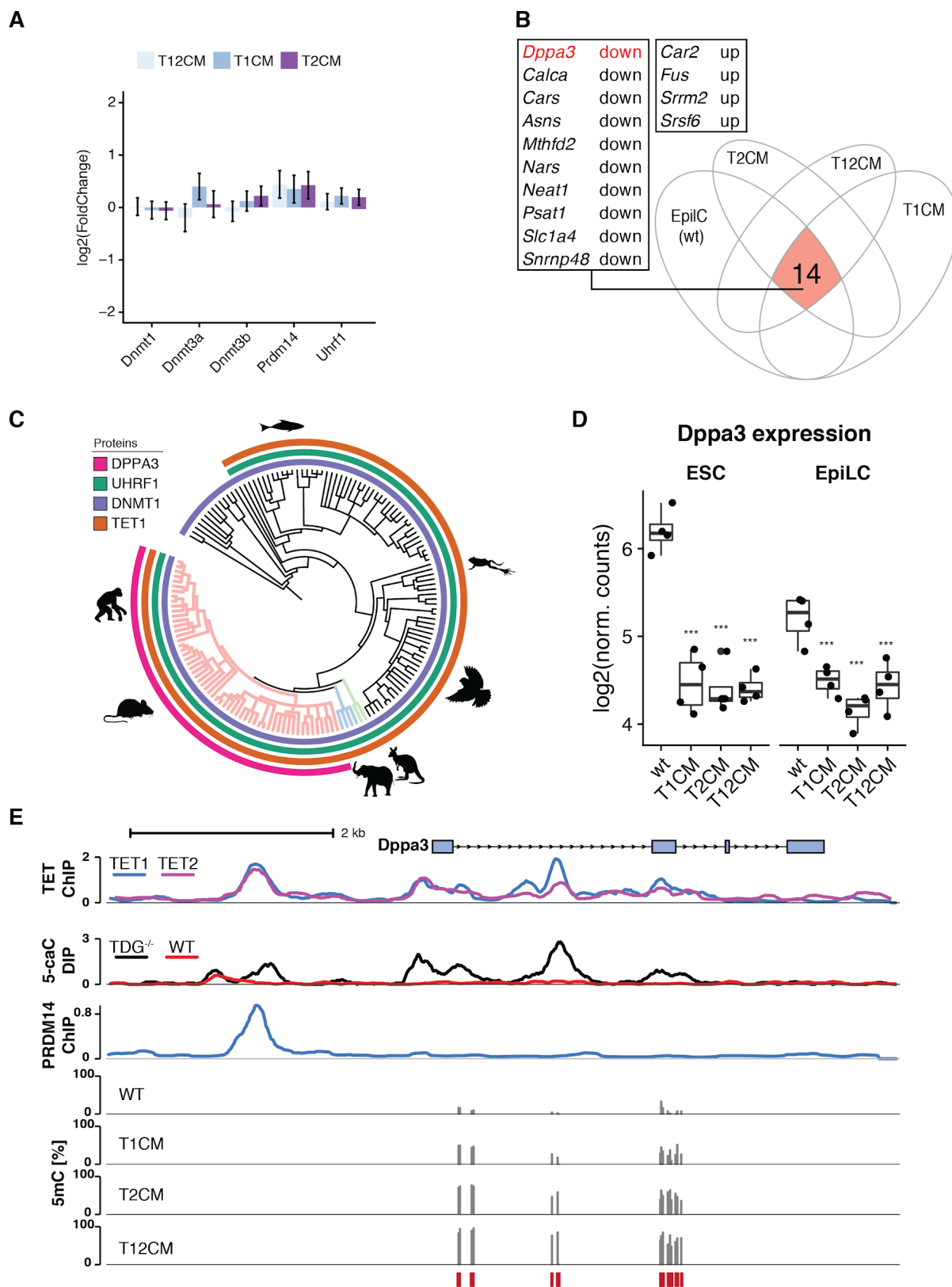


Figure 2: TET1 and TET2 catalytic activity is necessary for *Dppa3* expression.

(A) Expression of genes involved in regulating DNA methylation levels in T1CM, T2CM, and T12CM ESCs as assessed by RNA-seq. Expression is given as the \log_2 fold-change compared to wt ESCs. Error bars indicate mean \pm SD, $n=4$ biological replicates. No significant changes observable (Likelihood ratio test). (B) *Dppa3* is downregulated upon loss of TET activity and during differentiation. Venn diagram depicting the overlap (red) of genes differentially expressed (compared to wt ESCs; adjusted $p < 0.05$) in T1CM, T2CM, T12CM ESCs, and wt EpiLCs. (C) Phylogenetic tree of TET1, DNMT1, UHRF1, and DPPA3 in metazoa. (D) *Dppa3* expression levels as determined by RNA-seq in the indicated ESC and EpiLC lines ($n=4$

biological replicates). **(E)** TET proteins bind and actively demethylate the *Dppa3* locus. Genome browser view of the *Dppa3* locus with tracks of the occupancy (Signal pileup per million reads; (SPMR)) of TET1 (Khoueiry et al., 2017), TET2 (Xiong et al., 2016), and PRDM14 (Yamaji et al., 2013) in wt ESCs, 5caC enrichment in wt vs. TDG^{-/-} ESCs (Shen et al., 2013), and 5mC (%) levels in wt, T1CM, T2CM, and T12CM ESCs (RRBS). Red bars indicate CpGs covered by RRBS. In **(D)** boxplots horizontal black lines within boxes represent median values, boxes indicate the upper and lower quartiles, and whiskers indicate the 1.5 interquartile range. P-values were calculated using Welch's two-sided t-test comparing *Tet* catalytic mutants to their corresponding wt: ** P<0.01; *** P<0.001.

DPPA3 acts downstream of TET1 and TET2 and is required to safeguard the naïve methylome

DPPA3 has been reported to both prevent or promote DNA demethylation depending on the developmental time points (Han et al., 2018; Huang et al., 2017; Li et al., 2018; Nakamura et al., 2007, 2012; Nakashima et al., 2013; Shin et al., 2017; Xu et al., 2015). However, the function of DPPA3 in the context of naïve pluripotency, for which it is a well-established marker gene (Hayashi et al., 2008), has yet to be explored. We first sought to characterize the relationship between the high expression of *Dppa3* and DNA hypomethylation both accompanying naïve pluripotency. To this end, we established isogenic *Dppa3* knockout (Dppa3KO) mouse ESCs using CRISPR/Cas (Figure S3A-S3C) and profiled their methylome by RRBS. Loss of DPPA3 led to severe global hypermethylation (Figure 3A) with substantial increases in DNA methylation observed across all genomic elements (Figure S3D). Repetitive sequences and TEs, in particular, were severely hypermethylated including 98% of L1 elements (Figure S3D). A principal component analysis of the RRBS data revealed that Dppa3KO ESCs clustered closer to wt EpiLCs and *Tet* catalytic mutant ESCs rather than wt ESCs (Figure 3B). Furthermore, we observed a striking overlap of hypermethylated CpGs between *Tet* catalytic mutant and Dppa3KO ESCs (Figure 3C), suggesting that DPPA3 and TETs promote demethylation at largely the same targets. A closer examination of the genomic distribution of overlapping hypermethylation in *Tet* catalytic mutant and Dppa3KO ESCs revealed that the majority (~85%) of common targets reside within repetitive elements (Figure 3D, S3E, and S3F) and are globally correlated with heterochromatic histone modifications (Figure 3H). In contrast, only half of the observed promoter hypermethylation among all cell lines was dependent on DPPA3 (classified as “common”, Figure 3D, S3F, and S3G). This allowed us to identify a set of strictly TET-dependent promoters (N=1573) (Figure 3D and S3F; Table S1), which were enriched for developmental genes (Fig 3E and S3F; Table S2). Intriguingly, these TET-specific promoters contained genes (such as *Pax6*, *Foxa1* and *Otx2*) that have recently been shown to be conserved targets of TET-mediated demethylation during *Xenopus*, zebrafish and mouse development (Bogdanović et al., 2016).

DPPA3 appeared to act downstream of TETs as the global increase in DNA methylation in Dppa3KO ESCs was not associated with a reduction in 5hmC levels nor with a downregulation of TET family members (Figure 3F and S3I). In support of this notion, inducible overexpression of *Dppa3* (Figure S3J-S3L) completely rescued the observed hypermethylation phenotype at LINE-1 elements in T1CM as well as T2CM ESCs and resulted in a significant reduction of hypermethylation in T12CM cells (Figure 3G). Strikingly, prolonged induction of *Dppa3* even resulted in hypomethylation in wild-type as well as T1CM ESCs (Figure 3G). Collectively, these results show that TET1 and TET2 activity contributes to genomic

RESULTS

hypomethylation in naïve pluripotency by direct and indirect pathways, the active demethylation of developmental promoters and the passive, DPPA3-mediated global demethylation.

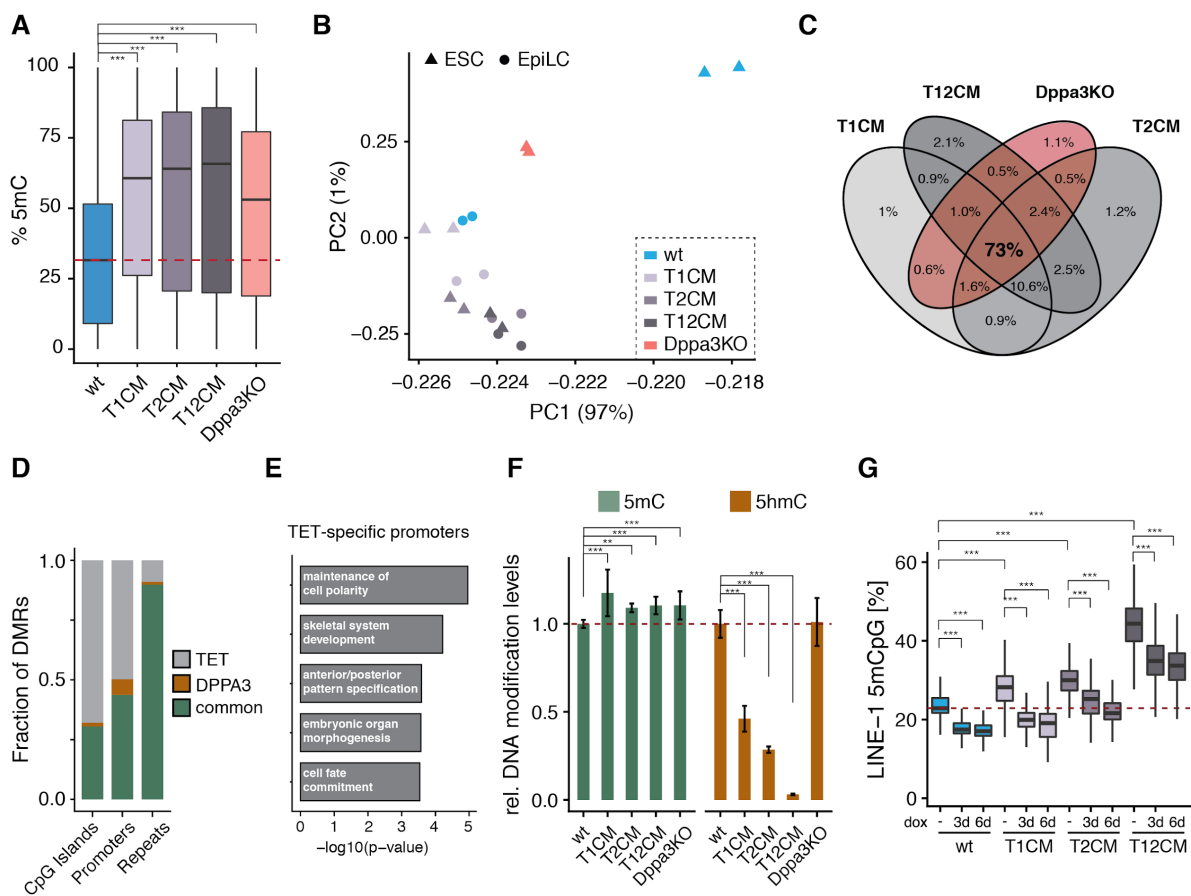


Figure 3: DPPA3 acts downstream of TET1 and TET2 to establish and preserve global hypomethylation.

(A) *Dppa3* loss results in global hypermethylation. Percentage of total 5mC as measured by RRBS. **(B)** *Dppa3* prevents the premature acquisition of a primed methylome. Principal component (PC) analysis of RRBS data from wt, T1CM, T2CM and T12CM ESCs, wt EpiLCs and *Dppa3*KO ESCs. **(C)** *Dppa3* and TET proteins promote demethylation of largely similar targets. Venn Diagrams depicting the overlap of hypermethylated sites among T1CM, T2CM, T12CM, and *Dppa3*KO ESCs. **(D)** *Dppa3* protects mostly repeats from hypermethylation. Fraction of hypermethylated genomic elements classified as TET-specific (only hypermethylated in TET mutant ESCs), DPPA3-specific (only hypermethylated in *Dppa3*KO ESCs), or common (hypermethylated in TET mutant and *Dppa3*KO ESCs). **(E)** Gene ontology (GO) terms associated with promoters specifically dependent on TET activity; adjusted p-values calculated using Fisher's exact test followed by Benjamini-Hochberg correction for multiple testing. **(F)** TET activity remains unaffected in *Dppa3*KO ESCs. Relative DNA modification levels for 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) as measured by mass spectrometry (LC-MS/MS). Error bars indicate mean \pm SD calculated from $n > 3$ biological replicates. **(G)** *Dppa3* expression can rescue the hypermethylation in TET mutant ESCs. DNA methylation levels at LINE-1 elements (%) as measured by bisulfite sequencing 0, 3, or 6 days after doxycycline (dox) induction of *Dppa3* expression. In **(A and G)** boxplots horizontal black lines within boxes represent median values, boxes indicate the upper and lower quartiles, and whiskers indicate the 1.5 interquartile range. The dashed red line indicates the median methylation level of wt ESCs. In **(A, F, and G)**, P-values were calculated using Welch's two-sided t-test: ** $P < 0.01$; *** $P < 0.001$.

TET-dependent expression of DPPA3 regulates UHRF1 subcellular distribution and controls DNA methylation maintenance in embryonic stem cells

To investigate the mechanism underlying the regulation of global DNA methylation patterns by DPPA3 we first generated an endogenous DPPA3-HALO fusion ESC line to monitor the localization of DPPA3 throughout the cell cycle (Figure S4A and S4C). Recent studies have shown that DPPA3 binds H3K9me2 (Nakamura et al., 2012) and that in oocytes its nuclear localization is critical to inhibit the activity of UHRF1 (Li et al., 2018), a key factor for maintaining methylation. Expecting a related mechanism to be present in ESCs, we were surprised to find a strong cytoplasmic localization of DPPA3 in ESCs (Figure 4A). Furthermore, DPPA3 did not bind to mitotic chromosomes indicating a low or absent chromatin association of DPPA3 in ESCs (Figure 4A). To further understand the mechanistic basis of DPPA3-dependent DNA demethylation in ESCs, we performed FLAG-DPPA3 pulldowns followed by liquid chromatography tandem mass spectrometry (LC MS-MS) to profile the DPPA3 interactome in naïve ESCs. Strikingly, among the 303 significantly enriched DPPA3 interaction partners identified by mass spectrometry we found UHRF1 and DNMT1 (Figure 4B; Table S3), the core components of the DNA maintenance methylation machinery (Hashimoto et al., 2008; Sharif et al., 2007). A reciprocal immunoprecipitation of UHRF1 confirmed its interaction with DPPA3 in ESCs (Figure S4F). Furthermore, GO analysis using the top 131 interactors of DPPA3 showed the two most enriched GO terms to be related to DNA methylation (Table S4). These findings were consistent with previous studies implicating DPPA3 in the regulation of maintenance methylation (Funaki et al., 2014; Li et al., 2018). In addition, we also detected multiple members of the nuclear transport machinery indicating that DPPA3 might undergo nuclear shuttling in ESCs (highlighted in purple, Figure 4B; Table S3) which prompted us to investigate whether DPPA3 influences the subcellular localization of UHRF1. Surprisingly, biochemical fractionation experiments revealed UHRF1 to be present in both the nucleus and cytoplasm of naive wt ESCs (Figure S4E). Despite comparable total UHRF1 protein levels in wt and Dppa3KO ESCs (Figure S4G), loss of DPPA3 completely abolished the cytoplasmic fraction of UHRF1 (Figure S4E).

As maintenance DNA methylation critically depends on the correct targeting and localization of UHRF1 within the nucleus (Nishiyama et al., 2013; Qin et al., 2015; Rothbart et al., 2012; Zhao et al., 2016), we asked whether TET-dependent regulation of DPPA3 might affect the subnuclear distribution of UHRF1. To this end, we tagged endogenous UHRF1 with GFP in wild-type (U1G/wt) as well as Dppa3KO and T12CM ESCs (U1G/Dppa3KO and U1G/T12CM, respectively) enabling us to monitor UHRF1 localization dynamics in living cells (Figure S4B and S4D). Whereas UHRF1-GFP localized to both the nucleus and cytoplasm of wt ESCs, UHRF1-GFP localization was solely nuclear in Dppa3KO and T12CM ESCs (Figure S4H and S4I). In addition, UHRF1 appeared to display a more diffuse localization in wt ESCs compared to Dppa3KO and T12CM ESCs, in which we observed more focal patterning of UHRF1 particularly at heterochromatic foci (Figure S4H). To quantify this observation, we calculated the coefficient of variation (CV) of nuclear UHRF1-GFP among wt, Dppa3KO, and T12CM ESCs. The

CV of a fluorescent signal correlates with its distribution, where low CV values reflect more homogenous distributions and high CV values correspond to more heterogeneous distributions (Osswald et al., 2019; Weihs et al., 2018). Indeed, the pronounced focal accumulation of UHRF1-GFP observed in *Dppa3*KO and T12CM ESCs corresponded with a highly significant increase in the CV values of nuclear UHRF1-GFP compared with wt ESCs (Figure S4H and S4I).

To assess whether these differences in nuclear UHRF1 distribution reflected altered chromatin binding, we used fluorescence recovery after photobleaching (FRAP) to study the dynamics of nuclear UHRF1-GFP in wt, *Dppa3*KO, and T12CM ESCs. Our FRAP analysis revealed markedly increased UHRF1 chromatin binding in both *Dppa3*KO and T12CM ESCs as evidenced by the significantly slower recovery of UHRF1-GFP in these cell lines compared to wt ESCs (Figure 4C, S4J, and S4K). Additionally, these data demonstrated increased UHRF1 chromatin binding to underlie the more heterogenous nuclear UHRF1 distributions in *Dppa3*KO and T12CM ESCs. Interestingly, although strongly reduced compared to wt ESCs, UHRF1 mobility was slightly higher in T12CM ESCs than *Dppa3*KO ESCs, consistent with a severe but not total loss of DPPA3 in the absence of TET activity (Figure S4L). Induction of *Dppa3* rescued the cytoplasmic fraction of UHRF1 (N/C Ratio: Figure 4D) as well as the diffuse localization of nuclear UHRF1 in *Dppa3*KO ESCs (CV: Figure 4D), which reflected a striking increase in the mobility of residual nuclear UHRF1-GFP as assessed by FRAP (Figure S4M, S5A, and S5B). This analysis also revealed that UHRF1's nucleocytoplasmic translocation and the inhibition of chromatin binding followed almost identical kinetics (N/C $t_{1/2}$ =84.4 min; CV $t_{1/2}$ =82.8) (Figure 4D). UHRF1 is required for the proper targeting of DNMT1 to DNA replication sites and therefore essential for DNA methylation maintenance (Bostick et al., 2007; Sharif et al., 2007). We observed a marked reduction of both UHRF1 and DNMT1 at replication foci upon induction of *Dppa3*, indicating that DPPA3 promotes hypomethylation in naive ESCs by impairing DNA methylation maintenance (Figure S5C and S5D). Ectopic expression of DPPA3 not only altered the subcellular distribution of endogenous UHRF1 in mouse ESCs (Figure 4D and S5E) but also in human ESCs suggesting evolutionary conservation of this mechanism among mammals (Figure S5F and S5G). Collectively our results demonstrate that TET-proteins control both the subcellular localization and chromatin binding of UHRF1 via the regulation of DPPA3 levels in naïve ESCs. Furthermore, these data show that DPPA3 is both necessary and sufficient for ensuring the nucleocytoplasmic translocation, diffuse nuclear localization, and attenuated chromatin binding of UHRF1 in ESCs.

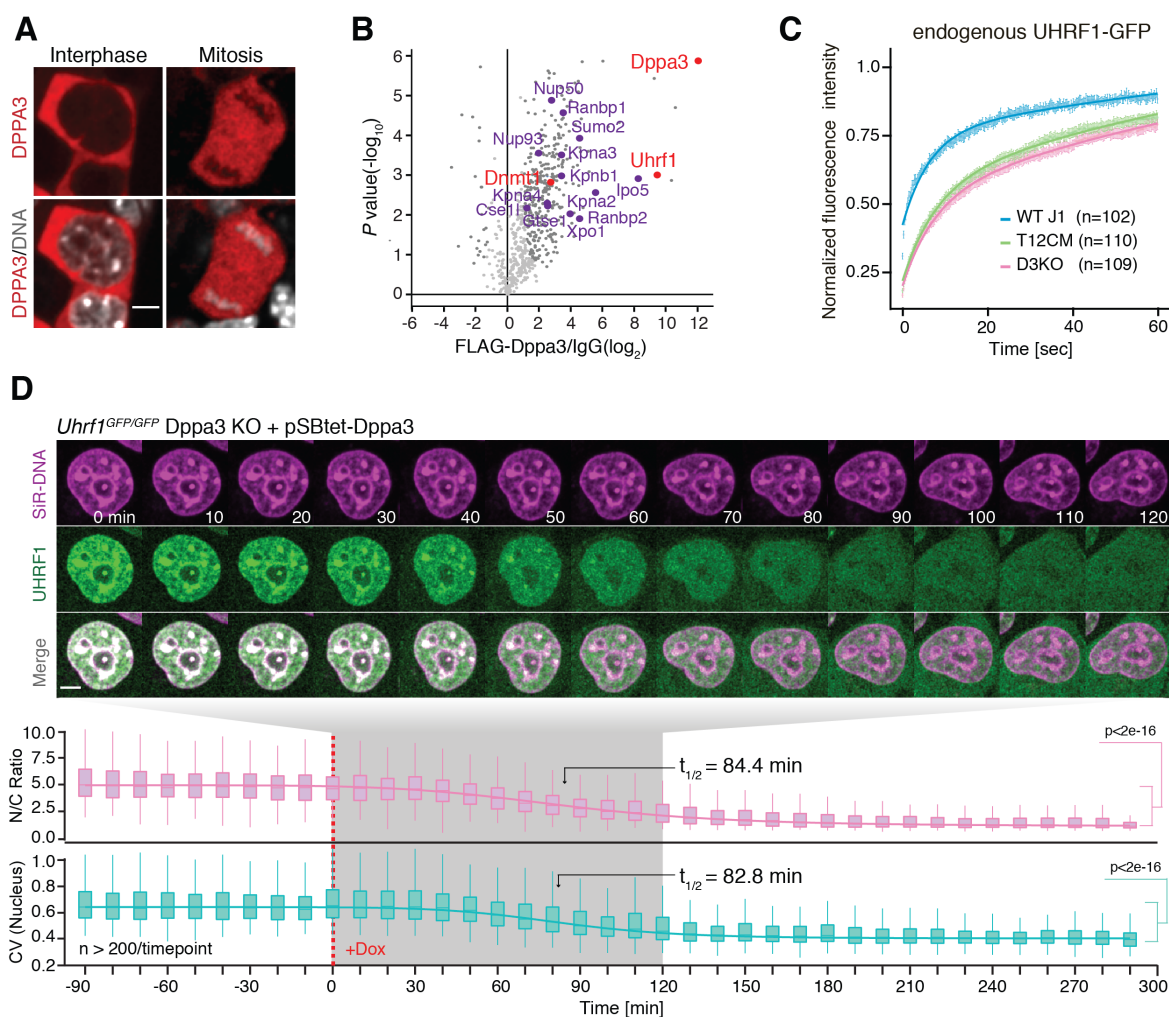


Figure 4: DPPA3 prevents UHRF1 chromatin binding to impede maintenance methylation in embryonic stem cells.

(A) DPPA3 is primarily localized to the cytoplasm of ESCs. Images illustrating the localization of endogenous DPPA3-HALO in live ESCs counterstained with SiR-Hoechst (DNA). Scale bar: 5 μ m. (B) DPPA3 interacts with endogenous UHRF1 in ESCs. Volcano plot from DPPA3-FLAG pulldowns in ESCs. Dark grey dots: significantly enriched proteins (FDR < 0.05). Red dots: proteins involved in DNA methylation regulation. Purple dots: proteins involved in nuclear transport. anti-FLAG antibody: n=3 biological replicates, IgG control antibody: n=3 biological replicates. Statistical significance determined by performing a Student's t-test with a permutation-based FDR of 0.05 and an additional constant S0=1. (C) Loss of DPPA3 leads to increased UHRF1 chromatin binding. FRAP analysis of endogenous UHRF1-GFP. Each genotype comprises the combined single-cell data from two independent clones. (D) Normal UHRF1 localization can be rescued by ectopic Dppa3 expression. Localization dynamics of endogenous UHRF1-GFP in response to *Dppa3* induction in U1G/D3KO + pSBtet-D3 ESCs with confocal time-lapse imaging over 8 h (10 min intervals). $t=0$ corresponds to start of *Dppa3* induction with doxycycline (+Dox). (top panel) Representative images of UHRF1-GFP and DNA (SiR-Hoechst stain) throughout confocal time-lapse imaging. Scale bar: 5 μ m. (middle panel) Nucleus to cytoplasm ratio (N/C Ratio) of endogenous UHRF1-GFP signal. (bottom panel) Coefficient of variance (CV) of endogenous UHRF1-GFP intensity in the nucleus. (middle and bottom panel) N/C Ratio and CV values: measurements in n>200 single cells per time point, acquired at n=16 separate positions. Curves represent fits of four parameter logistic (4PL) functions to the N/C Ratio (pink line) and CV (green line) data. In (C) the mean fluorescence intensity of n cells (indicated in the plots) at each timepoint are depicted as shaded dots. Error bars indicate mean \pm SEM. Curves (solid lines) indicate double-exponential functions fitted to the FRAP data. In the boxplots in (D) darker horizontal lines within boxes represent median values. The limits of the boxes indicate upper and lower quartiles, and whiskers indicate the 1.5-fold interquartile range. P-values based on Welch's two-sided t-test.

DPPA3-mediated inhibition of UHRF1 chromatin binding causes hypomethylation and is attenuated by nuclear export

Our results demonstrated cytoplasmic accumulation of UHRF1 and the disruption of its focal nuclear patterning to occur with almost identical kinetics upon induction of *Dppa3* expression (Figure 4D). In principle, either a decrease in nuclear UHRF1 concentration or the impaired chromatin loading of UHRF1 would on their own be sufficient to impair maintenance DNA methylation (Rothbart et al., 2012, 2013). To dissect these two modes and their contribution to the inhibition of maintenance methylation in naïve ESCs, we generated inducible *Dppa3*-mScarlet expression cassettes (Figure S6A) harboring mutations to residues described to be critical for its nuclear export (Δ NES) (Nakamura et al., 2007) and the interaction with UHRF1 (KRR and R107E) (Li et al., 2018) as well as truncated forms of DPPA3 found in zygotes, 1-60 and 61-150 (Shin et al., 2017) (Figure 5A). After introducing these *Dppa3* expression cassettes into U1GFP/*Dppa3*KO ESCs, we used live-cell imaging to track each DPPA3 mutant's localization and ability to rescue the loss of DPPA3 (Figure 5B). Whereas DPPA3- Δ NES and DPPA3 61-150, both lacking a functional nuclear export signal, were sequestered to the nucleus (Figure 5B), DPPA3-WT, DPPA3-KRR, DPPA3-R107E, and DPPA3 1-60 mutants localized primarily to the cytoplasm and recapitulated the localization of endogenous DPPA3 in naïve ESCs (Figure 5B and 4A). Regardless, all tested DPPA3 mutants failed to efficiently reestablish nucleocytoplasmic translocation of UHRF1 (Figure 5B and S6B), indicating that DPPA3 requires both the capacity to interact with UHRF1 as well as a functional nuclear export signal to promote nucleocytoplasmic shuttling of UHRF1 in naïve ESCs.

Nevertheless, DPPA3- Δ NES and DPPA3 61-150 managed to significantly disrupt the focal patterning and association with chromocenters of UHRF1 within the nucleus itself, with DPPA3- Δ NES causing an even greater reduction in the CV of nuclear UHRF1 than DPPA3-WT (Figure 5B and Figure S6C). In contrast, the loss or mutation of residues critical for its interaction with UHRF1 compromised DPPA3's ability to effectively restore the diffuse localization of nuclear UHRF1 (Figure 5B and S6C). On the one hand, FRAP analysis revealed that the disruption or deletion of the UHRF1 interaction interface (DPPA3-KRR, DPPA3-R107E, DPPA3 1-60) severely diminished the ability of DPPA3 to release UHRF1 from chromatin (Figure 5C and S6F-S6K). On the other hand, the C-terminal half of DPPA3 lacking nuclear export signal but retaining UHRF1 interaction came close to fully restoring the mobility of UHRF1 (Figure 5C, S6I, S6J, and S6K). Remarkably, DPPA3- Δ NES mobilized UHRF1 to an even greater extent than DPPA3-WT (Figure 5C, S6D, S6E, S6J, and S6K), suggesting that operative nuclear export might even antagonize DPPA3-mediated inhibition of UHRF1 chromatin binding. Supporting this notion, chemical inhibition of nuclear export using leptomycin-B (LMB) significantly enhanced the inhibition of UHRF1 chromatin binding in U1G/D3KO ESCs expressing DPPA3-WT (Figure S5H-S5K). Taken together our data shows that the efficiency of DPPA3-dependent release of UHRF1 from chromatin requires its interaction with UHRF1 but not its nuclear export.

To further address the question whether the nucleocytoplasmic translocation of UHRF1 and impaired UHRF1 chromatin binding both contribute to DPPA3-mediated inhibition of DNA methylation

maintenance, we assessed the ability of each DPPA3 mutant to rescue the hypermethylation of LINE-1 elements in *Dppa3*KO ESCs (Figure 5C). Strikingly, DPPA3- Δ NES fully rescued the hypermethylation and achieved a greater loss of DNA methylation than DPPA3-WT, whereas DPPA3 mutants lacking the residues important for UHRF1 binding failed to restore low methylation levels (Figure 5D). Overall, the ability of each DPPA3 mutant to reduce DNA methylation levels closely mirrored the extent to which each mutant impaired UHRF1 chromatin binding (Figure 5C and S6D-S6K). In line with the increased mobility of UHRF1 occurring in the absence of DPPA3 nuclear export (Figure 5C, S5H-S5K, S6D, S6E, S6J, and S6K), the nucleocytoplasmic translocation of UHRF1 is not only dispensable but rather attenuates DPPA3-mediated inhibition of maintenance methylation (Figure 5D). Collectively, our findings demonstrate the inhibition of UHRF1 chromatin binding, as opposed to nucleocytoplasmic translocation of UHRF1, to be the primary mechanism by which DPPA3 drives hypomethylation in naïve ESCs.

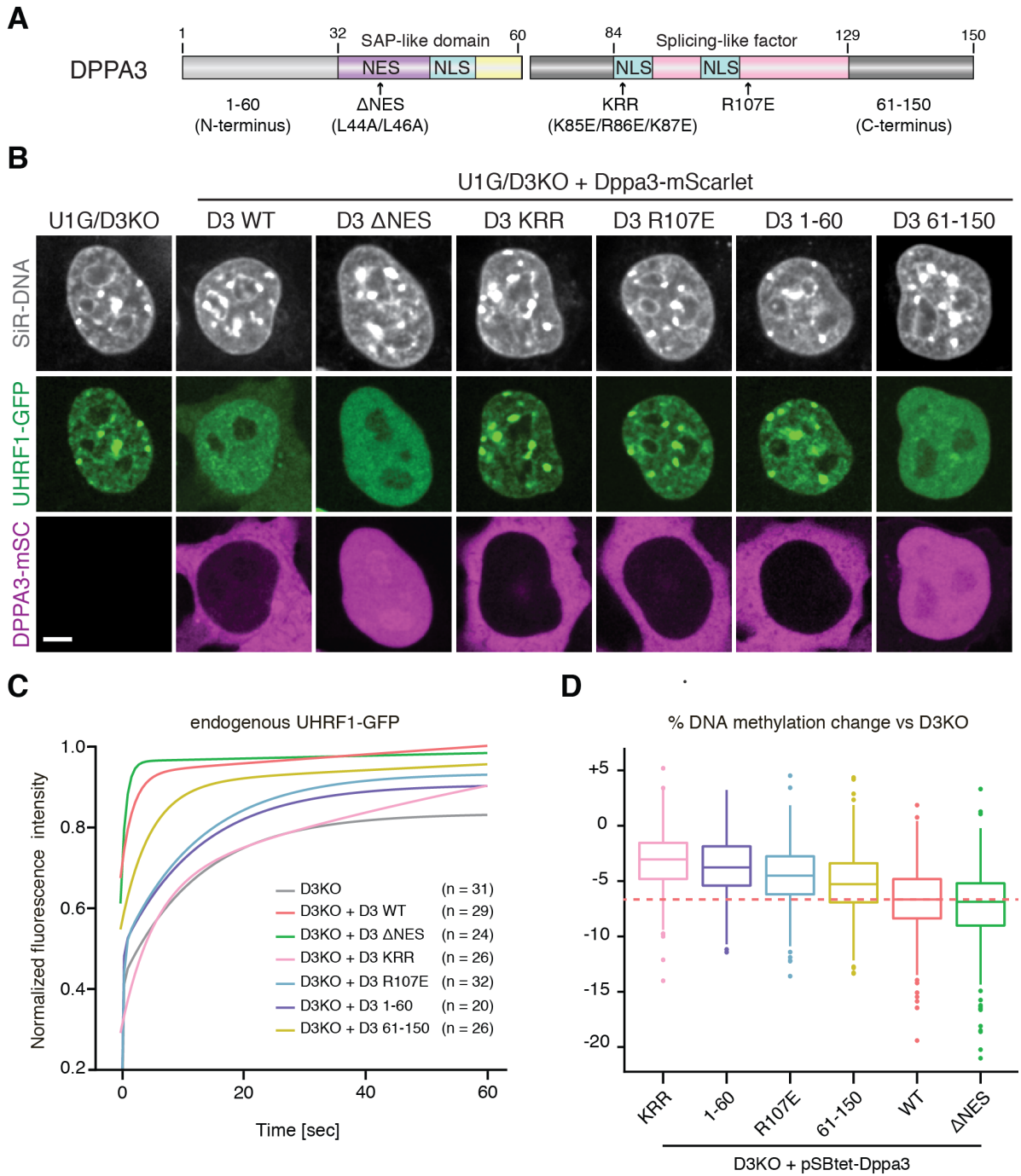


Figure 5: DPPA3 inhibits maintenance DNA methylation by impairing UHRF1 chromatin binding

(A) Schematic illustration of murine DPPA3 with the nuclear localization signals (NLS), nuclear export signal (NES), and predicted domains (SAP-like and splicing factor-like (Payer et al., 2003) annotated. For the DPPA3 mutant forms used in this study, point mutations are indicated with arrows (Δ NES, KRR, R107E) and the two truncations are denoted by the middle break (1-60, left half; 61-150, right half). (B and C) Nuclear export and the C-terminus of DPPA3 are dispensable for disrupting focal UHRF1 patterning and chromatin binding in ESCs. (B) Representative confocal images illustrating the localization of endogenous UHRF1-GFP and the indicated mDPPA3-mScarlet fusions in live U1G/D3KO + pSB-D3-mSC ESCs. DNA counterstain: SiR-Hoechst. Scale bar: 5 μ m. (C) FRAP analysis of endogenous UHRF1-GFP in U1G/D3KO ESCs expressing the indicated mutant forms of Dppa3. FRAP Curves (solid lines) indicate double-exponential functions fitted to the FRAP data acquired from *n* cells (shown in the plots). For single-cell FRAP data and additional quantification, see Figure S6D-S6K. (D) DPPA3-mediated inhibition of UHRF1 chromatin binding is necessary and sufficient to promote DNA demethylation. Percentage of DNA methylation change at LINE-1 elements (%) in D3KO ESCs after induction of the indicated mutant forms of Dppa3 as measured by bisulfite sequencing of *n*=4 biological replicates. In the boxplot horizontal

black lines within boxes represent median values, boxes indicate the upper and lower quartiles, whiskers indicate the 1.5 interquartile range, and dots indicate outliers. P-values based on Welch's two-sided t-test

DPPA3 binds nuclear UHRF1 with high affinity prompting its release from chromatin in ESCs

We next set out to investigate the mechanistic basis of DPPA3's ability to inhibit UHRF1 chromatin binding in naïve ESCs. DPPA3 has been reported to specifically bind H3K9me2 (Nakamura et al., 2012), a histone modification critical for UHRF1 targeting (Citterio et al., 2004; Karagianni et al., 2008; Rothbart et al., 2012). These prior findings led us to consider two possible mechanistic explanations for DPPA3-mediated UHRF1 inhibition in naïve ESCs: i) DPPA3 blocks access of UHRF1 to chromatin by competing in binding to H3K9me2, ii) DPPA3 directly or indirectly binds to UHRF1 and thereby prevents it from accessing chromatin.

To simultaneously assess the dynamics of both UHRF1 and DPPA3 under physiological conditions in live ES cells, we employed raster image correlation spectroscopy with pulsed interleaved excitation (PIE-RICS) (Figure 6A). RICS is a confocal imaging method which enables the measurement of binding and diffusive properties in living cells. Using images acquired on a laser scanning confocal microscope, spatiotemporal information of fluorescently labeled proteins can be extracted from the shape of the spatial autocorrelation function (ACF). A diffusive model is fitted to the ACF which yields the average diffusion coefficient, the concentration, and the fraction of bound molecules (Digman et al., 2005). If two proteins are labeled with distinct fluorophores and imaged simultaneously with separate detectors, the extent of their interaction can be extracted from the cross-correlation of their fluctuations using cross-correlation RICS (ccRICS) (Figure 6A)(Digman et al., 2009).

We first measured the mobility of DPPA3-mScarlet variants expressed in U1GFP/D3KOs (Figure S7A and S7B). RICS analysis revealed that over the timescale of the measurements, nuclear DPPA3-WT was predominantly unbound from chromatin and freely diffusing through the nucleus at a rate of $7.18 \pm 1.87 \mu\text{m}^2/\text{s}$ (Figure S7F). The fraction of mobile DPPA3-mScarlet molecules was measured to be $88.4 \pm 5.2\%$ (Figure 6F), validating globally weak binding inferred from ChIP-Seq profiles(Huang et al., 2017). These mobility parameters were largely unaffected by disruption of the UHRF1 interaction, with the DPPA3-KRR mutant behaving similarly to wild-type DPPA3 (Figure 6F and S7F). To rule out a potential competition between UHRF1 and DPPA3 for H3K9me2 binding, we next used RICS to determine if DPPA3 dynamics are altered in the absence of UHRF1. For this purpose, we introduced the DPPA3-WT-mScarlet cassette into Uhrf1KO (U1KO) ESCs(Karg et al., 2017), in which free eGFP is expressed from the endogenous *Uhrf1* promoter (Figure S7C). However, neither the diffusion rate nor the mobile fraction of DPPA3 were appreciably altered in cells devoid of UHRF1, suggesting the high fraction of unbound DPPA3 to be unrelated to the presence of UHRF1 (Figure 6F and S7F). Overall, our RICS data demonstrate that, in contrast to zygotes (Nakamura et al., 2012), DPPA3 in ESCs lacks a strong capacity for chromatin binding, and as such is not engaged in competition with UHRF1 for chromatin binding.

We next used RICS to analyse the dynamics of UHRF1-GFP in response to DPPA3 induction (Figure 6A). In cells expressing DPPA3-KRR, RICS measurements revealed that only $32.4 \pm 10\%$ of UHRF1 is mobile, indicating that the majority of UHRF1 is chromatin-bound (Figure 6G). In contrast, expression of wild-type DPPA3 lead to a dramatic increase in the mobile fraction UHRF1 ($60.6 \pm 13.7\%$ mobile fraction for UHRF1) (Figure 6G, S7G, and S7H). Furthermore, the mobile fraction of UHRF1 increased as a function of the relative abundance of nuclear DPPA3 to UHRF1 (Figure S7I), thereby indicating a stoichiometric effect of DPPA3 on UHRF1 chromatin binding, consistent with a physical interaction. Thus, these results demonstrate that DPPA3 potently disrupts UHRF1 chromatin binding in live ESCs and suggest its interaction with UHRF1 to be critical to do so.

To determine whether such an interaction is indeed present in the nuclei of live ESCs, we performed cross-correlation RICS (ccRICS) (Figure 6A). We first validated ccRICS in ESCs by analyzing live cells expressing a tandem eGFP-mScarlet fusion (Figure 6E and S7D), or expressing both freely diffusing eGFP and mScarlet (Figure 6D and S7E). For the tandem eGFP-mScarlet fusion, we observed a clear positive cross-correlation indicative of eGFP and mScarlet existing in the same complex (Figure 6E and 6H), as would be expected for an eGFP-mScarlet fusion. On the other hand, freely diffusing eGFP and mScarlet yielded no visible cross-correlation (Figure 6D and 6H), consistent with two independent proteins which do not interact. Upon applying ccRICS to nuclear UHRF1-GFP and DPPA3-mScarlet, we observed a prominent cross-correlation between wild-type DPPA3 and the primarily unbound fraction of UHRF1 (Figure 6B and 6H), indicating that mobilized UHRF1 exists in a high affinity complex with DPPA3 in live ESCs. In marked contrast, DPPA3-KRR and UHRF1-GFP failed to exhibit detectable cross-correlation (Figure 6C and 6H), consistent with the DPPA3-KRR mutant's diminished capacity to bind (Li et al., 2018) and mobilize UHRF1 (Figure 5C, S6F, S6J, and S6K). Overall, these findings demonstrate that nuclear DPPA3 interacts with UHRF1 to form a highly mobile complex in naïve ESCs which precludes UHRF1 chromatin binding.

To determine whether the DPPA3-UHRF1 complex identified *in vivo* (Figure 6H) corresponds to a high affinity direct interaction, we performed microscale thermophoresis (MST) measurements using recombinant UHRF1-GFP and DPPA3 proteins. MST analysis revealed a direct and high affinity (K_D : $0.44 \mu\text{M}$) interaction between the DPPA3 WT and UHRF1 (Figure 6I). No binding was observed for DPPA3 1-60, lacking the residues essential for interaction with UHRF1 (Figure 6I). In line with the results obtained by ccRICS, these data support the notion that DPPA3 directly binds UHRF1 *in vivo*. Interestingly, the affinity of the UHRF1-DPPA3 interaction was comparable or even far greater than that reported for the binding of UHRF1 to H3K9me3 or unmodified H3 peptides, respectively (Fang et al., 2016; Harrison et al., 2016).

To better understand how UHRF1 chromatin loading is impaired by its direct interaction with DPPA3, we applied a fluorescent-three-hybrid (F3H) assay to identify the UHRF1 domain bound by DPPA3 *in vivo* (Figure S7J and S7K). In short, this method relies on a cell line harboring an array of lac operator binding sites in the nucleus at which a GFP-tagged “bait” protein can be immobilized and visualized as a

spot. Thus, the extent of recruitment of an mScarlet-tagged “prey” protein to the nuclear GFP-spot offers a quantifiable measure of the interaction propensity of the “bait” and “prey” proteins *in vivo* (Figure S7K) (Herce et al., 2013). Using UHRF1-GFP domain deletions as the immobilized bait (Figure S7J), we assessed how the loss of each domain affected the recruitment of mDPPA3-mScarlet to the GFP spot. In contrast to the other UHRF1 domain deletions, removal of the PHD domain essentially abolished recruitment of DPPA3 to the lac spot, demonstrating DPPA3 binds UHRF1 via its PHD domain *in vivo* (Figure S7L and S7M). The PHD of UHRF1 is essential for its recruitment to chromatin (Arita et al., 2012; Harrison et al., 2016; Rothbart et al., 2013), ubiquitination of H3 and recruitment of DNMT1 to replication foci (Nishiyama et al., 2013; Qin et al., 2015). Thus, our *in vivo* results suggest that the high affinity interaction of DPPA3 with UHRF1’s PHD domain precludes UHRF1 from binding chromatin in ESCs, which is also supported by a recent report demonstrating that DPPA3 specifically binds the PHD domain of UHRF1 to competitively inhibit H3 tail binding *in vitro* (Du et al., 2019).

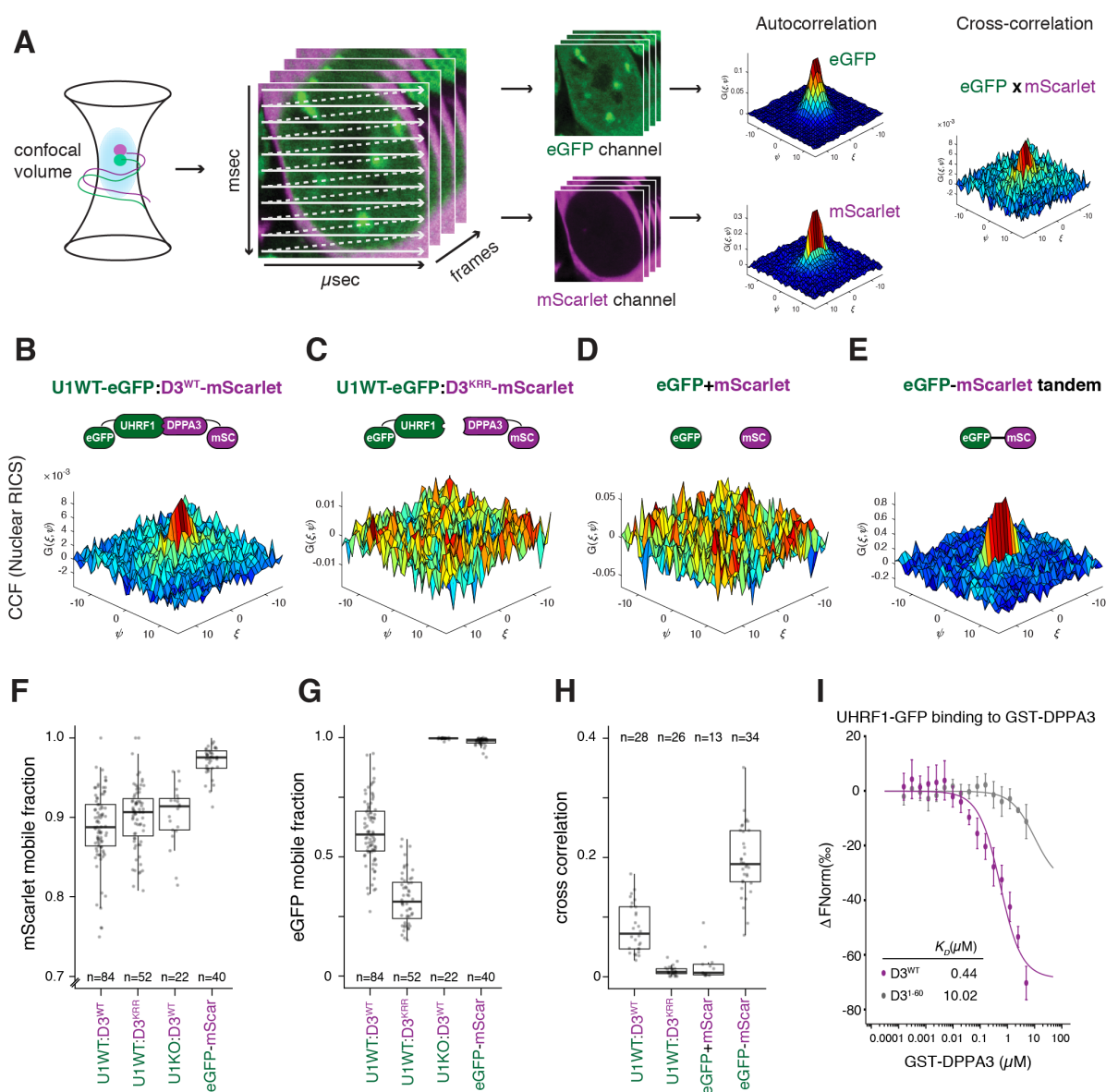


Figure 6: DPPA3 binds nuclear UHRF1 with high affinity prompting its release from chromatin in ESCs

(A) Overview of RICS and ccRICS. Confocal image series are acquired using a calibrated point-scanning laser, generating spatio-temporal fluorescence information on the microsecond and millisecond timescales. An autocorrelation function (ACF) is calculated from the fluorescence fluctuations and used to fit a diffusive model (Digman et al., 2005). The cross-correlation of fluctuations between two channels is used to estimate the co-occurrence of two fluorescent molecules within live cells (Digman et al., 2009). The mean cross-correlation of fluctuations is calculated and shown in the 3D plot, color-coded according to the correlation value. (B-E) Representative plots of the cross-correlation function (CCF) between the depicted fluorescent molecules in cells from each cell line measured, including U1G/D3KO + pSBtet-D3 ESCs expressing the following forms of DPPA3-mScarlet: (B) wild-type (U1WT:D3^{WT}) and (C) K85E/R85E/K87E mutant (U1WT:D3^{KRR}), and control ESCs expressing (D) free eGFP and free mScarlet (eGFP + mScarlet) and (E) an eGFP-mScarlet tandem fusion (eGFP-mScarlet). See Figure S7 for the images and ACF plots of the cells used to make the representative CCF plots. (F and G) Mobile fraction of (F) mScarlet and (G) eGFP species in the cell lines depicted in (B, C, and E) as well as in Uhrf1KO ESCs expressing free eGFP and wild-type Dppa3-mScarlet (U1KO:D3^{WT}) (Figure S7C). The mobile fraction was derived from a two-component model fit of the autocorrelation function. (H) Mean cross-correlation values of mobile eGFP and mScarlet measured in the cell lines depicted in (B-E). The fast timescale axis is indicated by ξ , and the slow timescale axis is indicated by ψ . (I) Microscale thermophoresis measurements of UHRF1-eGFP binding to GST-DPPA3 WT (D3^{WT}) or GST-DPPA3 1-60 (D3¹⁻⁶⁰). Error bars indicate mean \pm SEM of $n=2$ technical replicates from $n=4$ independent experiments. In (F-H), each data point represents the measured and fit values from a single cell where n = number of cells measured (indicated in the plots). In the box plots, darker horizontal lines within boxes represent median values. The limits of the boxes indicate upper and lower quartiles, and whiskers indicate the 1.5-fold interquartile range.

DPPA3 can inhibit UHRF1 function and drive global DNA demethylation in distantly related, non-mammalian species

Whereas UHRF1 and TET proteins are widely conserved throughout plants and vertebrates (Feng et al., 2010; Iyer et al., 2009), both early embryonic global hypomethylation (Wu and Zhang, 2010) and the *Dppa3* gene are unique to mammals. Consistent with UHRF1's conserved role in maintenance DNA methylation, a multiple sequence alignment of UHRF1's PHD domain showed that the residues critical for the recognition of histone H3 are completely conserved from mammals to invertebrates (Figure 7A). This prompted us to consider the possibility that DPPA3 might be capable of modulating the function of distantly related UHRF1 homologs outside of mammals. To test this hypothesis, we used amphibian (*Xenopus laevis*) egg extracts to assess the ability of mouse DPPA3 (mDPPA3) to interact with a non-mammalian form of UHRF1. Despite the 360 million year evolutionary distance between mouse and *Xenopus* (Kumar and Hedges, 1998), mDPPA3 not only bound *Xenopus* UHRF1 (xUHRF1) with high affinity (Figure 7B, 7C, S8A, and S8B) it also interacted with xUHRF1 specifically via its PHD domain (Figure S8C, S8D, and S8E). Moreover, the first 60 amino acids of DPPA3 were dispensable for its interaction with UHRF1 (Figure S8A and S8B). Interestingly, mutation to R107, reported to be critical for DPPA3's binding with mouse UHRF1 (Li et al., 2018), diminished but did not fully disrupt the interaction (Figure 7C, S8B, and S8E). The R107E mutant retained the ability to bind the xUHRF1-PHD domain but exhibited decreased binding to xUHRF1-PHD-SRA under high-salt conditions (Figure S8E), suggesting that R107E changes the binding mode of mDPPA3 to xUHRF1, rather than inhibiting the complex formation. Considering the remarkable similarity between DPPA3's interaction with mouse and *Xenopus* UHRF1, we reasoned that the ability of DPPA3 to inhibit UHRF1 chromatin binding and maintenance DNA methylation might be transferable to *Xenopus*. To address this, we took advantage of a cell-free system derived from interphase *Xenopus* egg extracts to reconstitute DNA maintenance

methylation (Nishiyama et al., 2013). Remarkably, recombinant mDPPA3 completely disrupted chromatin binding of both *Xenopus* UHRF1 and DNMT1 without affecting the loading of replication factors such as xCDC45, xRPA2, and xPCNA (Figure 7D). We determined that the inhibition of xUHRF1 and xDNMT1 chromatin loading only requires DPPA3's C-terminus (61-150 a.a.) and is no longer possible upon mutation of R107 (R107E) (Figure S8H), in line with our results in mouse ESCs (Figure 5D). Moreover, DPPA3-mediated inhibition of xUHRF1 chromatin loading resulted in the severe perturbation of histone H3 dual-monoubiquitylation (H3Ub2), which is necessary for the recruitment of DNMT1) (Ishiyama et al., 2017; Nishiyama et al., 2013; Qin et al., 2015)(Figure S8F). To determine whether mDPPA3 can displace xUHRF1 already bound to chromatin, we first depleted *Xenopus* egg extracts of xDNMT1 to stimulate the hyper-accumulation of xUHRF1 on chromatin (Nishiyama et al., 2013; Yamaguchi et al., 2017) and then added recombinant mDPPA3 after S-phase had commenced (Figure S8G). Under these conditions, both wild-type mDPPA3 and the 61-150 fragment potently displaced xUHRF1 from chromatin, leading to suppressed H3 ubiquitylation (Figure S8G). We next assessed the effect of DPPA3 on *Xenopus* maintenance DNA methylation. Consistent with the severe disruption of xDNMT1 chromatin loading, both DPPA3 wild-type and 61-150 effectively abolished replication-dependent DNA methylation in *Xenopus* egg extracts (Figure 7E). In contrast, DPPA3 1-60 and DPPA3 R107E, which both failed to suppress xUHRF1 and xDNMT1 binding, did not significantly alter maintenance DNA methylation activity (Figure 7E, S8D, and S8E). Taken together, our data demonstrate DPPA3 to be capable to potently inhibit maintenance DNA methylation in a non-mammalian system.

These findings raised the question whether a single protein capable of inhibiting UHRF1 function like DPPA3 could establish a mammalian-like global hypomethylation during the early embryonic development of a non-mammalian organism. To explore this possibility we turned to the biomedical model fish, medaka (*Oryzias latipes*), which does not exhibit genome-wide erasure of DNA methylation (Walter et al., 2002) and diverged from mammals 450 million years ago (Kumar and Hedges, 1998). We injected medaka embryos with Dppa3 mRNA at the one-cell stage and then tracked their developmental progression. Remarkably, medaka embryos injected with Dppa3 failed to develop beyond the blastula stage (Figure 7F) and exhibited a near-complete elimination of global DNA methylation as assessed by immunofluorescence (Figure 7G). Dppa3-mediated DNA methylation loss was both dose dependent and sensitive to the R107E mutation, which induced only partial demethylation (Figure S8H). Interestingly, medaka embryos injected with DPPA3 R107E showed far fewer developmental defects than those injected with wild-type DPPA3 (Figure 7G), suggesting that the embryonic arrest resulting from DPPA3 expression is a consequence of the global loss of DNA methylation. Taken together, these results demonstrate that mammalian DPPA3 can inhibit UHRF1 to drive passive demethylation in distant, non-mammalian contexts.

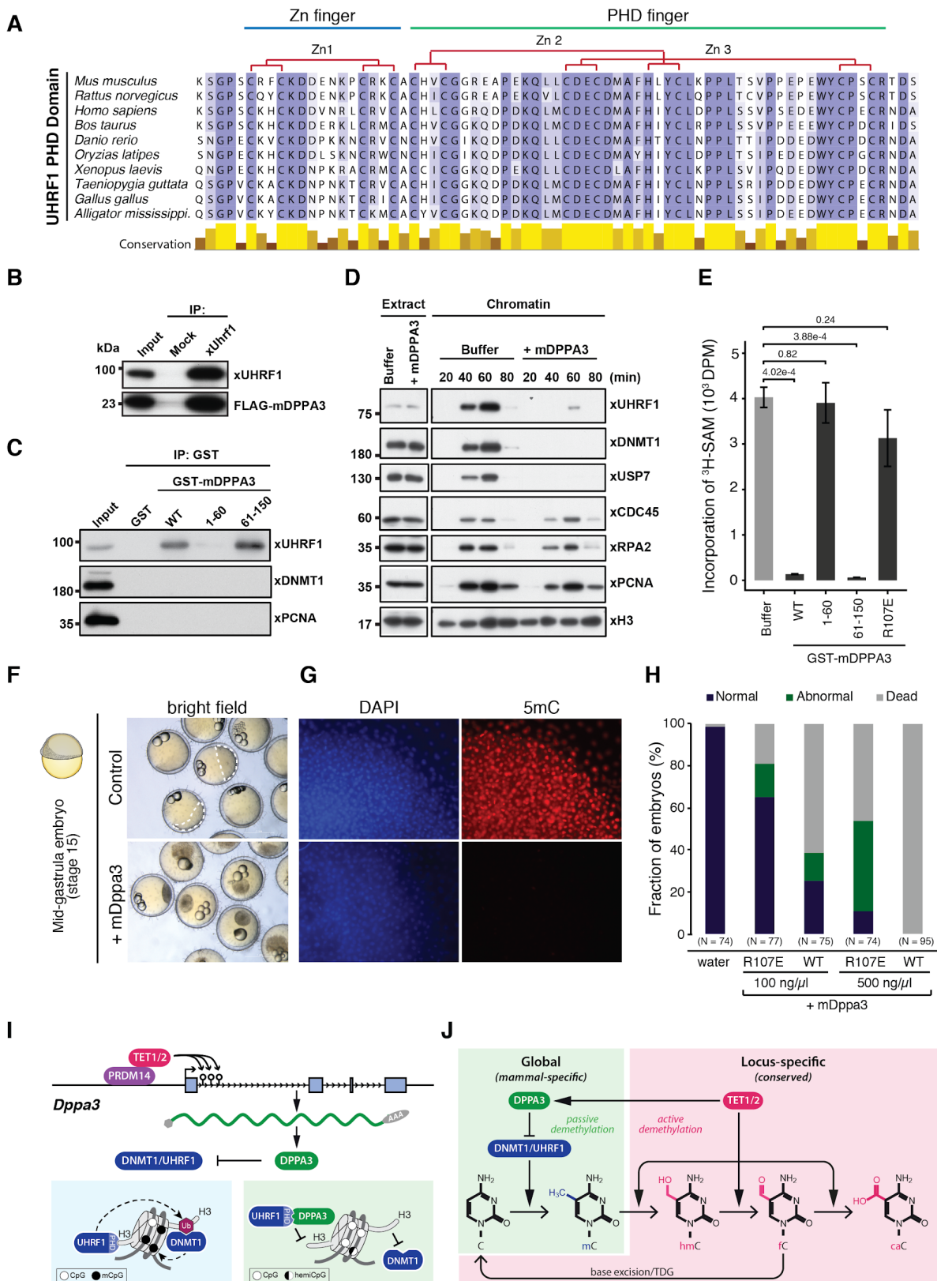


Figure 7: DPPA3 evolved in boreoeutherian mammals but also functions in lower vertebrates.

(A) Protein sequence alignment of the PHD domain of the UHRF1 family. The UHRF1 PHD domain shows high amino acid conservation throughout vertebrates, especially the residues involved in Zinc coordination (indicated above). (B) Endogenous xUHRF1 binds mDPPA3. IPs were performed on *Xenopus* egg extracts incubated with FLAG-mDPPA3 using either a control (Mock) or anti-xUHRF antibody and then analyzed by immunoblotting using the indicated antibodies. (C) xUHRF1 binds the C-terminus of mDPPA3. GST-tagged mDPPA3 wild-type (WT), point mutant R107E, and truncations (1-60 and 61-150) were immobilized on GSH beads and incubated with *Xenopus* egg extracts. Bound proteins were analyzed using the indicated

antibodies. **(D)** mDPPA3 inhibits xUHRF1 and xDNMT1 chromatin binding. Sperm chromatin was incubated with interphase *Xenopus* egg extracts supplemented with buffer (+buffer) or GST-mDPPA3 (+mDPPA3). Chromatin fractions were isolated and subjected to immunoblotting using the antibodies indicated. **(E)** mDPPA3 inhibits maintenance DNA methylation in *Xenopus*. The efficiency of maintenance DNA methylation was assessed by the incorporation of radiolabelled methyl groups from S-[methyl-³H]-adenosyl-L-methionine (³H-SAM) into DNA purified from egg extracts. Disintegrations per minute (DPM). Depicted p-values based on a Student's two-sided t-test. **(F)** mDPPA3 compromises the embryonic development of medaka. Representative images of developing mid-gastrula stage embryos (control injection) and arrested, blastula stage embryos injected with mDppa3. Injections were performed on one-cell stage embryos and images were acquired ~18 h after fertilization. **(G)** mDPPA3 drives global DNA demethylation in medaka embryos. Representative 5mC immunostainings in control and mDppa3-injected medaka embryos at the late blastula stage (~8 h after fertilization). Scale bars represent 50 μ m. DNA counterstain: DAPI,4',6-diamidino-2-phenylindole. **(H)** Global hypomethylation causes developmental arrest in medaka. Percentage of normal, abnormal, or dead. Embryos were injected with wild-type mDppa3 (WT) or mDppa3 R107E (R107E) at two different concentrations (100 ng/ μ l or 500 ng/ μ l) or water at the one-cell stage and analyzed ~18 h after fertilization. N = number of embryos from n= 3 independent injection experiments. **(I)** TET1 and TET2 are recruited by PRDM14 to the promoter of *Dppa3* where they promoter active DNA demethylation and transcription of *Dppa3*. DPPA3 is expressed and inhibits maintenance DNA methylation by directly binding UHRF1 and releasing it from chromatin. **(J)** TET1 and TET2 control DNA methylation levels by two evolutionary and mechanistically distinct pathways.

DISCUSSION

While the appearance of genome-wide DNA demethylation in mammals represents a momentous change with far-reaching consequences to epigenetic gene regulation during early development, the key enzymes involved in DNA modification are highly conserved in vertebrates. As the role of TET enzymes in active demethylation is well documented (Wu and Zhang, 2017), we investigated their contribution to the hypomethylated state of naïve ESCs. Mutation of the catalytic core of TET enzymes caused - as expected - a genome-wide increase in DNA methylation but mostly at sites where TET proteins do not bind suggesting a rather indirect mechanism. Among the few genes depending on TET activity for expression in naïve ESCs and downregulated at the transition to EpiLCs was *Dppa3*. Demethylation at the *Dppa3* locus coincides with TET1 and TET2 binding and TDG dependent removal of oxidized cytosine residues via base excision repair. DPPA3 in turn binds and displaces UHRF1 from chromatin and thereby prevents the recruitment of DNMT1 and the maintenance of DNA methylation in ESCs (see graphic summary in Figure 7).

Despite long recognized as a marker of naïve ESCs resembling the inner cell mass (Hayashi et al., 2008; Singer et al., 2014), we provide, to our knowledge, the first evidence that DPPA3 directly promotes the genome-wide DNA hypomethylation characteristic of mammalian naïve pluripotency. This unique pathway, in which TET proteins indirectly cause passive demethylation, is based upon two uniquely mammalian innovations: the expression of TET genes in pluripotent cell types (Almeida et al., 2012; Bogdanović et al., 2016; Ito et al., 2010) and the evolution of the novel *Dppa3* gene, positioned within a pluripotency gene cluster and dependent on TET activity for expression. In support of this novel pathway for passive demethylation, we found that TET mutant ESCs show a similar phenotype as *Dppa3*KO cells with respect to UHRF1 inhibition and hypomethylation and can be rescued by ectopic expression of *Dppa3*.

Our findings also provide the missing link to reconcile previous, apparently conflicting reports. To date,

three distinct mechanisms have been proposed for global hypomethylation accompanying naive pluripotency: TET-mediated active demethylation (Ficz et al., 2013; Hackett et al., 2013; von Meyenn et al., 2016), impaired maintenance DNA methylation (von Meyenn et al., 2016), and PRDM14-dependent suppression of methylation (Hackett et al., 2013; Leitch et al., 2013; Yamaji et al., 2013). As both, a downstream target of TETs and PRDM14 as well as a direct inhibitor of maintenance DNA methylation, DPPA3 mechanistically links and integrates these three proposed pathways of demethylation (see graphic summary in Figure 7).

Our mechanistic data showing DPPA3 to displace UHRF1 and DNMT1 from chromatin provide a conclusive explanation for the previous observation that global hypomethylation in naive ESCs was accompanied by reduced levels of UHRF1 at replication foci (von Meyenn et al., 2016). The hypomethylated state of naive ESCs has also been reported to be dependent on PRDM14 (Leitch et al., 2013; Yamaji et al., 2013), which has been suggested to promote demethylation by repressing *de novo* DNA methyltransferases (Ficz et al., 2013; Grabole et al., 2013; Leitch et al., 2013; Yamaji et al., 2013). However, recent studies have demonstrated that the loss of *de novo* methylation only marginally affects DNA methylation levels in mouse and human ESCs (Liao et al., 2015; von Meyenn et al., 2016). Interestingly, the loss of *Prdm14* lead to global hypermethylation and, to our surprise, also the downregulation of *Dppa3* (Grabole et al., 2013; Ma et al., 2011; Yamaji et al., 2013). Our results suggest that the reported ability of PRDM14 to promote hypomethylation in naive ESCs largely relies on its activation of the *Dppa3* gene ultimately leading to an inhibition of maintenance methylation.

The comparison of TET catalytic mutants and *Dppa3* KO ESCs allows us to distinguish TET-dependent passive DNA demethylation mediated by DPPA3 from *bona fide* active demethylation. We show that TET activity is indispensable for the active demethylation of a subset of promoters in naïve ESCs, especially those of developmental genes. These findings uncover two evolutionary and mechanistically distinct functions of TET catalytic activity. Whereas TET-mediated active demethylation of developmental genes is evolutionarily conserved among vertebrates (Bogdanović et al., 2016; Dai et al., 2016; Li et al., 2016; Verma et al., 2018), the use of TET proteins to promote global demethylation appears to be specific to mammalian pluripotency (Ficz et al., 2013; Hackett et al., 2013; von Meyenn et al., 2016) and mediated by the recently evolved *Dppa3* (Figures 2C and 7J).

To date, our understanding of DPPA3's function in the regulation of DNA methylation had been clouded by seemingly conflicting reports from different developmental stages and cell types. DPPA3's ability to modulate DNA methylation was first described in the context of zygotes (Nakamura et al., 2007), where it was subsequently demonstrated to specifically protect the maternal genome from TET3-dependent demethylation (Han et al., 2018; Nakamura et al., 2012; Wossidlo et al., 2011). In contrast, DPPA3 was later shown to facilitate DNA demethylation during PGC specification (Nakashima et al., 2013), iPSC reprogramming (Xu et al., 2015) and oocyte maturation (Han et al., 2019; Li et al., 2018). Whereas DPPA3 was shown to disrupt UHRF1 function by sequestering it to the cytoplasm in oocytes (Li et al., 2018), we demonstrate that DPPA3-mediated nucleocytoplasmic translocation of UHRF1 is not only

dispensable but actually attenuates DPPA3's promotion of hypomethylation in ESCs. In light of our data from naïve ESCs, *Xenopus*, and medaka, DPPA3's capacity to directly bind UHRF1's PHD domain and thereby inhibit UHRF1 chromatin binding appears to be its most basal function. Considering that DPPA3 localization is highly dynamic during the different developmental time periods at which it is expressed (Nakashima et al., 2013; Payer et al., 2006; Shin et al., 2017), it stands to reason that its role in modulating DNA methylation might also be dynamically regulated by yet-to-be determined regulatory mechanisms. For example, immediately following fertilization, full length DPPA3 is cleaved and its C-terminal domain is specifically degraded (Shin et al., 2017). Interestingly, we identified this exact C-terminal stretch of DPPA3 to be necessary and sufficient for DPPA3's inhibition of maintenance DNA methylation. Thus, the precisely timed destruction of this crucial domain might offer an explanation for the differing roles of DPPA3 in regulating DNA methylation between oocytes and zygotes (Han et al., 2018, 2019; Li et al., 2018; Nakamura et al., 2012).

As the most basic and evolutionarily conserved function of DNA methylation is the repression of transposable elements (Schmitz et al., 2019), the emergence of genome-wide DNA demethylation in mammals raises several fundamental questions. While the DPPA3 mediated erasure of parental DNA methylation might facilitate the establishment of new epigenetic patterns during development, it should be noted that non-mammalian vertebrates manage to undergo normal development without genome-wide demethylation. Moreover, the global loss of DNA methylation even poses a severe threat as excessive demethylation triggers derepression of TEs leading to genomic instability and ultimately cell death in most cell types (Chernyavskaya et al., 2017; Chiappinelli et al., 2017; Iida et al., 2006; Jackson-Grusby et al., 2001; Roulois et al., 2015; Walsh et al., 1998). Remarkably, mammalian naive pluripotent cell types seem to have acquired the ability to tolerate global hypomethylation, suggesting that the evolutionary emergence of DPPA3 was likely accompanied by measures to control and productively integrate this new factor in epigenetic regulation in ESCs. In fact, many TEs in mammals are not only expressed during development but appear to have been co-opted to drive transcriptional networks critical for the establishment of pluripotency and progression through pre-implantation development (Cosby et al., 2019; Robbez-Masson and Rowe, 2015).

A good example for the functional integration of TEs in regulatory networks is the reactivation of endogenous retroviruses (ERVs) that is critical for the maternal-to-zygotic transition (MZT), the first major developmental step when maternal mRNAs are degraded and zygotic transcription begins (De Iaco et al., 2017; Hendrickson et al., 2017; Huang et al., 2017; Jachowicz et al., 2017). Strikingly, the activation of ERVs is severely impaired in *Dppa3* knockout embryos resulting in MZT failure (Huang et al., 2017). It is tempting to speculate that mammal-specific demethylation originates from an arms race between TE and host. DPPA3 may have arisen as a means to overcome the host defence system and was then co-opted by the host and gradually integrated into regulatory networks during evolution. Such a scenario is compatible with the unique occurrence of *Dppa3* in mammals and with our finding that DPPA3 alone is sufficient to inhibit DNA methylation maintenance in *Xenopus* and medaka, species that

harbor no *Dppa3* gene and exhibits constant DNA methylation levels at all stages of development (Stancheva et al., 2002; Veenstra and Wolffe, 2001). Follow-up studies that investigate the origin of *Dppa3* and whether a similar rewiring of early development may have occurred in other, not yet studied branches of vertebrates, are needed to further explore this evolutionary scenario.

REFERENCES

- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* *13*, R87.
- Almeida, R.D., Loose, M., Sottile, V., Matsa, E., Denning, C., Young, L., Johnson, A.D., Gering, M., and Ruzov, A. (2012). 5-hydroxymethyl-cytosine enrichment of non-committed cells is not a universal feature of vertebrate development. *Epigenetics* *7*, 383–389.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* *25*, 3389–3402.
- Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* *18*, 225–233.
- Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., Kuwata, K., Hamamoto, R., Tochio, H., Sato, M., 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.
- Arkhipova, I.R. (2018). Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. *Mol. Biol. Evol.* *35*, 1332–1337.
- Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* *15*, 545.
- Bauer, C., Göbel, K., Nagaraj, N., Colantuoni, C., Wang, M., Müller, 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.
- Bogdanović, O., Smits, A.H., de la Calle Mustienes, E., Tena, J.J., Ford, E., Williams, R., Senanayake, U., Schultz, M.D., Hontelez, S., van Kruijsbergen, I., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat. Genet.* *48*, 417–426.
- Borodovsky, A., Kessler, B.M., Casagrande, R., Overkleeft, H.S., Wilkinson, K.D., and Ploegh, H.L. (2001). A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome

association of USP14. *EMBO J.* 20, 5187–5196.

Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev. Cell* 35, 366–382.

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

Boyle, P., Clement, K., Gu, H., Smith, Z.D., Ziller, M., Fostel, J.L., Holmes, L., Meldrim, J., Kelley, F., Gnirke, A., et al. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol.* 13, R92.

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

Chernyavskaya, Y., Mudbhary, R., Zhang, C., Tokarz, D., Jacob, V., Gopinath, S., Sun, X., Wang, S., Magnani, E., Madakashira, B.P., et al. (2017). Loss of DNA methylation in zebrafish embryos activates retrotransposons to trigger antiviral signaling. *Development* 144, 2925–2939.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., et al. (2017). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* 169, 361.

Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P., and Bonapace, I.M. (2004). Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell. Biol.* 24, 2526–2535.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146, 67–79.

Cosby, R.L., Chang, N.-C., and Feschotte, C. (2019). Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* 33, 1098–1116.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805.

Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., and Nagaraj, N. (2014). MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Molecular & Cellular.*

Dai, H.-Q., Wang, B.-A., Yang, L., Chen, J.-J., Zhu, G.-C., Sun, M.-L., Ge, H., Wang, R., Chapman,

D.L., Tang, F., et al. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. *Nature* 538, 528.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945.

Digman, M.A., Brown, C.M., Sengupta, P., Wiseman, P.W., Horwitz, A.R., and Gratton, E. (2005). Measuring fast dynamics in solutions and cells with a laser scanning microscope. *Biophys. J.* 89, 1317–1327.

Digman, M.A., Wiseman, P.W., Horwitz, A.R., and Gratton, E. (2009). Detecting protein complexes in living cells from laser scanning confocal image sequences by the cross correlation raster image spectroscopy method. *Biophys. J.* 96, 707–716.

Du, W., Dong, Q., Zhang, Z., Liu, B., Zhou, T., Xu, R.-M., Wang, H., Zhu, B., and Li, Y. (2019). Stella protein facilitates DNA demethylation by disrupting the chromatin association of the RING finger-type E3 ubiquitin ligase UHRF1. *J. Biol. Chem.* 294, 8907–8917.

Fang, J., Cheng, J., Wang, J., Zhang, Q., Liu, M., Gong, R., Wang, P., Zhang, X., Feng, Y., Lan, W., et al. (2016). Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. *Nat. Commun.* 7, 11197.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8689–8694.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* 13, 351–359.

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

Funaki, S., Nakamura, T., Nakatani, T., Umehara, H., Nakashima, H., and Nakano, T. (2014). Inhibition of maintenance DNA methylation by Stella. *Biochem. Biophys. Res. Commun.* 453, 455–460.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.

Grabole, N., Tischler, J., Hackett, J.A., Kim, S., Tang, F., Leitch, H.G., Magnúsdóttir, E., and Surani, M.A. (2013). Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep.* 14, 629–637.

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

- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.
- Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., et al. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* 15, 447–459.
- Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F., and Chin, L. (2016). Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep.* 14, 1555–1566.
- Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* 13, 360–369.
- Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013). Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem Cell Reports* 1, 518–531.
- Han, L., Ren, C., Li, L., Li, X., Ge, J., Wang, H., Miao, Y.-L., Guo, X., Moley, K.H., Shu, W., et al. (2018). Embryonic defects induced by maternal obesity in mice derive from Stella insufficiency in oocytes. *Nat. Genet.* 50, 432–442.
- Han, L., Ren, C., Zhang, J., Shu, W., and Wang, Q. (2019). Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development. *Cell Discov* 5, 9.
- Harrison, J.S., Cornett, E.M., Goldfarb, D., DaRosa, P.A., Li, Z.M., Yan, F., Dickson, B.M., Guo, A.H., Cantu, D.V., Kaustov, L., et al. (2016). Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. *Elife* 5.
- 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.
- Hayashi, K., and Saitou, M. (2013). Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat. Protoc.* 8, 1513–1524.
- Hayashi, K., Lopes, S.M.C. de S., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* 3, 391–401.
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934.

Hendrix, J., Baumgärtel, V., Schrimpf, W., Ivanchenko, S., Digman, M.A., Gratton, E., Kräusslich, H.-G., Müller, B., and Lamb, D.C. (2015). Live-cell observation of cytosolic HIV-1 assembly onset reveals RNA-interacting Gag oligomers. *J. Cell Biol.* *210*, 629–646.

Hendrix, J., Dekens, T., Schrimpf, W., and Lamb, D.C. (2016). Arbitrary-Region Raster Image Correlation Spectroscopy. *Biophys. J.* *111*, 1785–1796.

Herce, H.D., Deng, W., Helma, J., Leonhardt, H., and Cardoso, M.C. (2013). Visualization and targeted disruption of protein interactions in living cells. *Nat. Commun.* *4*, 2660.

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

Huang, Y., Kim, J.K., Do, D.V., Lee, C., Penfold, C.A., Zylitz, J.J., Marioni, J.C., Hackett, J.A., and Surani, M.A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *Elife* *6*.

Iida, A., Shimada, A., Shima, A., Takamatsu, N., Hori, H., Takeuchi, K., and Koga, A. (2006). Targeted reduction of the DNA methylation level with 5-azacytidine promotes excision of the medaka fish Tol2 transposable element. *Genet. Res.* *87*, 187–193.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* *6*, e1001134.

Ishiyama, S., Nishiyama, A., Saeki, Y., Moritsugu, K., Morimoto, D., Yamaguchi, L., Arai, N., Matsumura, R., Kawakami, T., Mishima, Y., et al. (2017). Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. *Mol. Cell* *68*, 350–360.e7.

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.

Iwamatsu, T. (2004). Stages of normal development in the medaka *Oryzias latipes*. *Mech. Dev.* *121*, 605–618.

Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* *8*, 1698–1710.

Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.-E. (2017).

LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* *49*, 1502–1510.

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

Jacobs, F.M.J., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* *516*, 242–245.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., Stunnenberg, H.G., Lilley, K.S., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. *Development* *144*, 1221–1234.

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.

Karg, E., Smets, M., Ryan, J., Forné, I., Qin, W., Mulholland, C.B., Kalideris, G., Imhof, A., Bultmann, S., and Leonhardt, H. (2017). Ubiquitome Analysis Reveals PCNA-Associated Factor 15 (PAF15) as a Specific Ubiquitination Target of UHRF1 in Embryonic Stem Cells. *J. Mol. Biol.* *429*, 3814–3824.

Khoueiry, R., Sohni, A., Thienpont, B., Luo, X., Velde, J.V., Bartocetti, M., Boeckx, B., Zwijsen, A., Rao, A., Lambrechts, D., et al. (2017). Lineage-specific functions of TET1 in the postimplantation mouse embryo. *Nat. Genet.*

Kowarz, E., Löscher, D., and Marschalek, R. (2015). Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol. J.* *10*, 647–653.

Kumar, S., and Hedges, S.B. (1998). A molecular timescale for vertebrate evolution. *Nature* *392*, 917–920.

Ladstätter, S., and Tachibana, K. (2019). Genomic insights into chromatin reprogramming to totipotency in embryos. *J. Cell Biol.* *218*, 70–82.

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

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* *20*, 311–316.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* *23*, 127–128.

Li, X., Yue, X., Pastor, W.A., Lin, L., Georges, R., Chavez, L., Evans, S.M., and Rao, A. (2016). Tet

proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proc. Natl. Acad. Sci. U. S. A.* *113*, E8267–E8276.

Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., et al. (2018). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* *564*, 136–140.

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

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.

Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. (2011). Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat. Struct. Mol. Biol.* *18*, 120–127.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* *15*, 905–915.

Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* *149*, 590–604.

Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., et al. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* *41*, 753–761.

McQuin, C., Goodman, A., Chernyshev, V., Kametsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* *16*, e2005970.

Méndez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.* *20*, 8602–8612.

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N.Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., et al. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Mol. Cell.*

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

- Mulholland, C.B., Smets, M., Schmidtman, E., Leidescher, S., Markaki, Y., Hofweber, M., Qin, W., Manzo, M., Kremmer, E., Thanisch, K., et al. (2015). A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res.* *43*, e112.
- Müller, B.K., Zaychikov, E., Bräuchle, C., and Lamb, D.C. (2005). Pulsed interleaved excitation. *Biophys. J.* *89*, 3508–3522.
- Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., et al. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* *9*, 64–71.
- Nakamura, T., Liu, Y.-J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* *486*, 415–419.
- Nakashima, H., Kimura, T., Kaga, Y., Nakatani, T., Seki, Y., Nakamura, T., and Nakano, T. (2013). Effects of *dppa3* on DNA methylation dynamics during primordial germ cell development in mice. *Biol. Reprod.* *88*, 125.
- Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F., et al. (2013). Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* *502*, 249–253.
- Okashita, N., Kumaki, Y., Ebi, K., Nishi, M., Okamoto, Y., Nakayama, M., Hashimoto, S., Nakamura, T., Sugawara, K., Kojima, N., et al. (2014). PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development* *141*, 269–280.
- Ortega-Recalde, O., Day, R.C., Gemmell, N.J., and Hore, T.A. (2019). Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. *Nat. Commun.* *10*, 3053.
- Osswald, M., Santos, A.F., and Morais-de-Sá, E. (2019). Light-Induced Protein Clustering for Optogenetic Interference and Protein Interaction Analysis in *Drosophila* S2 Cells. *Biomolecules* *9*.
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* *7*.
- Payer, B., Saitou, M., Barton, S.C., Thresher, R., Dixon, J.P.C., Zahn, D., Colledge, W.H., Carlton, M.B.L., Nakano, T., and Surani, M.A. (2003). Stella is a maternal effect gene required for normal early development in mice. *Curr. Biol.* *13*, 2110–2117.
- Payer, B., Chuva de Sousa Lopes, S.M., Barton, S.C., Lee, C., Saitou, M., and Surani, M.A. (2006). Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. *Genesis* *44*, 75–83.

Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S.K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., et al. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* *10*, 574–581.

Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., La Mastra, F., Forné, I., Pichler, G., Hörl, D., Fellingner, K., et al. (2015). DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res.* *25*, 911–929.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* *44*, W160–W165.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* *8*, 2281–2308.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* *2*, 1896–1906.

Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* *30*, 733–750.

Rau, A., Gallopin, M., Celeux, G., and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* *29*, 2146–2152.

Robbez-Masson, L., and Rowe, H.M. (2015). Retrotransposons shape species-specific embryonic stem cell gene expression. *Retrovirology* *12*, 45.

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

Rothbart, S.B., Dickson, B.M., Ong, M.S., Krajewski, K., Houlston, S., Kireev, D.B., Arrowsmith, C.H., and Strahl, B.D. (2013). Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. *Genes Dev.* *27*, 1288–1298.

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

Roulois, D., Loo Yau, H., Singhanian, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Jones, P.A., Pugh, T.J., et al. (2015). DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* *162*, 961–973.

Rowe, H.M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. *Virology* *411*, 273–287.

- Sanford, J.P., Clark, H.J., Chapman, V.M., and Rossant, J. (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes Dev.* *1*, 1039–1046.
- Scheltema, R.A., and Mann, M. (2012). SprayQc: A Real-Time LC–MS/MS Quality Monitoring System To Maximize Uptime Using Off the Shelf Components. *J. Proteome Res.* *11*, 3458–3466.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* *9*, 676–682.
- Schmitz, R.J., Lewis, Z.A., and Goll, M.G. (2019). DNA Methylation: Shared and Divergent Features across Eukaryotes. *Trends Genet.* *35*, 818–827.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* *9*, 671–675.
- Schrimpf, W., Barth, A., Hendrix, J., and Lamb, D.C. (2018). PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data. *Biophys. J.* *114*, 1518–1528.
- Sharif, J., Muto, M., Takebayashi, S.-I., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* *450*, 908–912.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D’Alessio, A.C., Fung, H.-L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* *153*, 692–706.
- Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell* *15*, 459–471.
- Shin, S.-W., John Vogt, E., Jimenez-Movilla, M., Baibakov, B., and Dean, J. (2017). Cytoplasmic cleavage of DPPA3 is required for intracellular trafficking and cleavage-stage development in mice. *Nat. Commun.* *8*, 1643.
- Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* *55*, 319–331.
- Skvortsova, K., Tarbashevich, K., Stehling, M., Lister, R., Irimia, M., Raz, E., and Bogdanovic, O. (2019). Retention of paternal DNA methylome in the developing zebrafish germline. *Nat. Commun.* *10*, 3054.
- Sporbert, A., Domaing, P., Leonhardt, H., and Cardoso, M.C. (2005). PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic Acids Res.* *33*, 3521–3528.

Stancheva, I., El-Maarri, O., Walter, J., Niveleau, A., and Meehan, R.R. (2002). DNA methylation at promoter regions regulates the timing of gene activation in *Xenopus laevis* embryos. *Dev. Biol.* *243*, 155–165.

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.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* *13*, 731.

Veenstra, G.J., and Wolffe, A.P. (2001). Constitutive genomic methylation during embryonic development of *Xenopus*. *Biochim. Biophys. Acta* *1521*, 39–44.

Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q.V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.-J., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* *50*, 83–95.

Wagner, M., Steinbacher, J., Kraus, T.F.J., Michalakis, S., Hackner, B., Pfaffeneder, T., Perera, A., Müller, M., Giese, A., Kretzschmar, H.A., et al. (2015). Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angew. Chem. Int. Ed Engl.* *54*, 12511–12514.

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

Walter, R.B., Li, H.-Y., Intano, G.W., Kazianis, S., and Walter, C.A. (2002). Absence of global genomic cytosine methylation pattern erasure during medaka (*Oryzias latipes*) early embryo development. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* *133*, 597–607.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell* *157*, 979–991.

Warren, I.A., Naville, M., Chalopin, D., Levin, P., Berger, C.S., Galiana, D., and Volff, J.-N. (2015). Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosome Res.* *23*, 505–531.

Weihs, F., Wacnik, K., Turner, R.D., Culley, S., Henriques, R., and Foster, S.J. (2018). Heterogeneous localisation of membrane proteins in *Staphylococcus aureus*. *Sci. Rep.* *8*, 3657.

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

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

Cell Biol. *11*, 607–620.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu, G.-L., Wang, H., et al. (2016). Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. *Mol. Cell* *64*, 913–925.

Xu, X., Smorag, L., Nakamura, T., Kimura, T., Dressel, R., Fitzner, A., Tan, X., Linke, M., Zechner, U., Engel, W., et al. (2015). Dppa3 expression is critical for generation of fully reprogrammed iPS cells and maintenance of Dlk1-Dio3 imprinting. *Nat. Commun.* *6*, 6008.

Yamaguchi, L., Nishiyama, A., Misaki, T., Johmura, Y., Ueda, J., Arita, K., Nagao, K., Obuse, C., and Nakanishi, M. (2017). Usp7-dependent histone H3 deubiquitylation regulates maintenance of DNA methylation. *Sci. Rep.* *7*, 55.

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., and Saitou, M. (2013). PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* *12*, 368–382.

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

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

Zhao, Q., Zhang, J., Chen, R., Wang, L., Li, B., Cheng, H., Duan, X., Zhu, H., Wei, W., Li, J., et al. (2016). Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nat. Commun.* *7*, 12464.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* *65*, 631–643.e4.

MATERIALS AND METHODS

Cell culture

Naïve J1 mouse ESCs were cultured and differentiated into EpiLCs as described previously (Hayashi and Saitou, 2013; Mulholland et al., 2015). In brief, for both naïve ESCs and EpiLCs defined media was used, consisting of N2B27: 50% neurobasal medium (Life Technologies), 50% DMEM/F12 (Life Technologies), 2 mM L-glutamine (Life Technologies), 0.1 mM β -mercaptoethanol (Life Technologies), N2 supplement (Life Technologies), B27 serum-free supplement (Life Technologies), 100 U/mL penicillin, and 100 μ g/mL streptomycin (Sigma). Naïve ESCs were maintained on flasks treated with 0.2% gelatin in defined media containing 2i (1 μ M PD032591 and 3 μ M CHIR99021 (Axon Medchem, Netherlands)), 1000 U/mL recombinant leukemia inhibitory factor (LIF, Millipore), and 0.3% BSA (Gibco) for at least three passages before commencing differentiation. For reprogramming experiments, naive media was supplemented with freshly prepared 100 μ M Vitamin C (L-ascorbic acid 2-phosphate, Sigma).

To differentiate naïve ESCs into Epiblast-like cells (EpiLCs), flasks were first pre-treated with Geltrex (Life Technologies) diluted 1:100 in DMEM/F12 (Life Technologies) and incubated at 37 °C overnight. Naïve ESCs were plated on Geltrex-treated flasks in defined medium containing 10 ng/mL Fgf2 (R&D Systems), 20 ng/mL Activin A (R&D Systems) and 0.1 \times Knockout Serum Replacement (KSR) (Life Technologies). Media was changed after 24 h and EpiLCs were harvested for RRBS and RNA-seq experiments after 48 h.

For CRISPR-assisted cell line generation, mouse ESCs were maintained on 0.2% gelatin-coated dishes in Dulbecco's modified Eagle's medium (Sigma) supplemented with 16% fetal bovine serum (FBS, Sigma), 0.1 mM β -mercaptoethanol (Invitrogen), 2 mM L-glutamine (Sigma), 1 \times MEM Non-essential amino acids (Sigma), 100 U/mL penicillin, 100 μ g/mL streptomycin (Sigma), homemade recombinant LIF tested for efficient self-renewal maintenance, and 2i (1 μ M PD032591 and 3 μ M CHIR99021 (Axon Medchem, Netherlands)).

For experiments in which cells were propagated in “serum LIF” conditions, the cells were maintained on 0.2% gelatin-coated dishes in Dulbecco's modified Eagle's medium (Sigma) supplemented with 16% fetal bovine serum (FBS, Sigma), 0.1 mM β -mercaptoethanol (Invitrogen), 2 mM L-glutamine (Sigma), 1 \times MEM Non-essential amino acids (Sigma), 100 U/mL penicillin, 100 μ g/mL streptomycin (Sigma), LIF (ESGRO, Millipore).

HESCs (line H9) were maintained in mTeSR1 medium (05850, STEMCELL Technologies) on Matrigel-coated plates (356234, Corning) prepared by 1:100 dilution, and 5 ml coating of 10 cm plates for 1 h at 37 °C. Colonies were passaged using the gentle cell dissociation reagent (07174, StemCell Technologies).

All cell lines were regularly tested for Mycoplasma contamination by PCR.

Sleeping Beauty Constructs

To generate the sleeping beauty donor vector with an N-terminal 3xFLAG tag and a fluorescent readout of doxycycline induction, we first used primers with overhangs harboring SfiI sites to amplify the IRES-DsRed-Express from pIRES2-DsRed-Express (Clontech). This fragment was then cloned into the NruI site in pUC57-GentR via cut-ligation to generate an intermediate cloning vector pUC57-SfiI-IRES-DsRed-Express-SfiI. A synthesized gBlock (IDT, Coralville, IA, USA) containing Kozak-BIO-3XFLAG-AsiSI-NotI-V5 was cloned into the Eco47III site of the intermediate cloning vector via cut-ligation. The luciferase insert from pSBtet-Pur (Kowarz et al., 2015), Addgene plasmid #60507) was excised using SfiI. The SfiI-flanked Kozak-BIO-3XFLAG-AsiSI-NotI-V5-IRES-DsRed-Express cassette was digested out of the intermediate cloning vector using SfiI and ligated into the pSBtet-Pur vector backbone linearized by SfiI. The end result was the parental vector, pSBtet-3xFLAG-IRES-DsRed-Express-PuroR. The pSBtet-3x-FLAG-mScarlet-PuroR vector was constructed by inserting a synthesized gBlock (IDT, Coralville, IA, USA) containing the SfiI-BIO-3XFLAG-AsiSI-NotI-mScarlet sequence into the SfiI-linearized pSBtet-Pur vector backbone using Gibson assembly (Gibson et al., 2009). For *Dppa3* expression constructs, the coding sequence of wild-type and mutant forms of *Dppa3* were synthesized as gBlocks (IDT, Coralville, IA, USA) and inserted into the pSBtet-3xFLAG-IRES-DsRed-Express-PuroR vector (linearized by AsiSI and NotI) using Gibson assembly. To produce the *Dppa3*-mScarlet fusion expression constructs, wild-type and mutant forms of *Dppa3* were amplified from pSBtet-3xFLAG-*Dppa3*-IRES-DsRed-Express-PuroR constructs using primers with overhangs homologous to the AsiSI and NotI restriction sites of the pSBtet-3x-FLAG-mScarlet-PuroR vector. Wild-type and mutant *Dppa3* amplicons were subcloned into the pSBtet-3x-FLAG-mScarlet-PuroR vector (linearized with AsiSI and NotI) using Gibson assembly.

For experiments involving the SBtet-3xFLAG-*Dppa3* cassette, all inductions were performed using 1 μ g/mL doxycycline (Sigma-Aldrich). The DPPA3-WT construct was able to rescue the cytoplasmic localization and chromatin association of UHRF1 indicating that C-terminally tagged DPPA3 remains functional (Figure 5B-D).

CRISPR/Cas9 genome engineering

For the generation of Tet1, Tet2, and Tet1/Tet2 catalytic mutants, specific gRNAs targeting the catalytic center of *Tet1* and *Tet2* were cloned into a modified version of the SpCas9-T2A-GFP/gRNA (px458; (Ran et al., 2013), Addgene plasmid #48138), to which we fused a truncated form of human Geminin (hGem) to SpCas9 in order to increase homology-directed repair efficiency (Gutschner et al., 2016).

A 200 bp ssDNA oligonucleotide harboring the H1652Y and D1654A mutations and ~100 bp of homology to the genomic locus was synthesized (IDT, Coralville, IA, USA). For targetings in wild-type J1 ESCs, cells were transfected with a 4:1 ratio of donor oligo and Cas9/gRNA construct. Positively transfected cells were isolated based on GFP expression using fluorescence-activated cell sorting (FACS) and plated at clonal density in ESC media 2 days after transfection. Cell lysis in 96-well plates, PCR on

RESULTS

lysates, and restriction digests were performed as previously described (Mulholland et al., 2015). *TetI* catalytic mutation was confirmed by Sanger sequencing.

As C-terminally tagged GFP labeled UHRF1 transgenes were shown to be able to rescue UIKO (Qin et al., 2015), the tagging of endogenous Uhrf1 was also performed at the C-terminus. For insertion of the HALO or eGFP coding sequence into the endogenous *Dppa3* and *Uhrf1* loci, respectively, *Dppa3* and *Uhrf1* specific gRNAs were cloned into SpCas9-hGem-T2A-Puromycin/gRNA vector, which is a modified version of SpCas9-T2A-Puromycin/gRNA vector (px459;(Ran et al., 2013), Addgene plasmid #62988) similar to that described above. To construct the homology donors plasmids, gBlocks (IDT, Coralville, IA, USA) were synthesized containing either the HALO or eGFP coding sequence flanked by homology arms with ~200-400 bp homology upstream and downstream of the gRNA target sequence at the *Dppa3* or *Uhrf1* locus, respectively, and then cloned into the *NruI* site of pUC57-GentR via cut-ligation. ESCs were transfected with equimolar amounts of gRNA and homology donor vectors. Two days after transfection, cells were plated at clonal density and subjected to a transient puromycin selection (1 ug/mL) for 40 h. After 5-6 days, ESCs positive for HALO or eGFP integration were isolated via fluorescence-activated cell sorting (FACS) and plated again at clonal density in ESC media. After 4-5 days, colonies were picked and plated on Optical bottom μ Clear 96-well plates and re-screened for the correct expression and localization of eGFP or HALO using live-cell spinning-disk confocal imaging. Cells were subsequently genotyped using the aforementioned cell lysis strategy and further validated by Sanger sequencing (Mulholland et al., 2015).

To generate *Dppa3* knockout cells, the targeting strategy entailed the use of two gRNAs with target sites flanking the *Dppa3* locus to excise the entire locus on both alleles. gRNA oligos were cloned into the SpCas9-T2A-PuroR/gRNA vector via cut-ligation. ESCs were transfected with an equimolar amount of each gRNA vector. Two days after transfection, cells were plated at clonal density and subjected to a transient puromycin selection (1 ug/mL) for 40 h. Colonies were picked 6 days after transfection. The triple PCR strategy used for screening is depicted in Figure S3A. Briefly, PCR primers 1F and 4R were used to identify clones in which the *Dppa3* locus had been removed, resulting in the appearance of a ~350 bp amplicon. To identify whether the *Dppa3* locus had been removed from both alleles, PCRs were performed with primers 1F and 2R or 3F and 4R to amplify upstream or downstream ends of the *Dppa3* locus, which would only be left intact in the event of mono-allelic locus excision. Removal of the *Dppa3* locus was confirmed with Sanger sequencing and loss of *Dppa3* expression was assessed by qRT-PCR.

For CRISPR/Cas gene editing, all transfections were performed using Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. All DNA oligos used for gene editing and screening are listed in Table S5.

Bxb1-mediated recombination and Sleeping Beauty Transposition

To generate stable mESC lines carrying doxycycline inducible forms of *Dppa3* or *Dppa3-mScarlet*, mESC cells were first transfected with equimolar amounts of the pSBtet-3xFLAG-Dppa3-IRES-DsRed-PuroR

or pSBtet-3xFLAG-Dppa3-mScarlet-PuroR and the Sleeping Beauty transposase, pCMV(CAT)T7-SB100(Mátés et al., 2009), Addgene plasmid #34879) vector using Lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's instructions. Two days after transfection, cells were plated at clonal density and subjected to puromycin selection (1 ug/mL) for 5-6 days. To ensure comparable levels of *Dppa3* induction, cells were first treated for 18 h with doxycycline (1 µg/mL) and then sorted with FACS based on thresholded levels of DsRed or mScarlet expression, the fluorescent readouts of successful induction. Post sorting, cells were plated back into media without doxycycline for 7 days before commencing experiments.

To generate stable doxycycline-inducible *Dppa3* hESC lines, hES cells were first transfected with equimolar amounts of the pSBtet-3xFLAG-Dppa3-IRES-DsRed-PuroR and Sleeping Beauty transposase pCMV(CAT)T7-SB100(Mátés et al., 2009), Addgene plasmid #34879) vector using using the P3 Primary Cell 4D-Nucleofector™ Kit (V4XP-3012 Lonza) and the 4D-Nucleofector™ Platform (Lonza), program CB-156. Two days after nucleofection, cells were subjected to puromycin selection (1 ug/mL) for subsequent two days, followed by an outgrowth phase of 4days. At this stage, cells were sorted with FACS based on thresholded levels of DsRed expression to obtain two bulk populations of positive stable hESC lines with inducible *Dppa3*.

For the generation of the *Uhrfl*^{GFP/GFP} cell line, we used our previously described ESC line with a C-terminal MIN-tag (*Uhrfl*^{attP/attP}; Bxb1 *attP* site) and inserted the GFP coding sequence as described previously (Mulholland et al., 2015). Briefly, attB-GFP-Stop-PolyA (Addgene plasmid #65526) was inserted into the C-terminal of the endogenous *Uhrfl*^{attP/attP} locus by transfection with equimolar amounts of Bxb1 and attB-GFP-Stop-PolyA construct, followed by collection of GFP-positive cells with FACS after 6 days.

Cellular fractionation and Western Blot

Western blot for T1CM ESCs were performed as described previously (Mulholland et al., 2015) using monoclonal antibody rat anti-TET1 5D6 (1:10) (Bauer et al., 2015) and polyclonal rabbit anti-H3 (1:5,000; ab1791, Abcam) as loading control. Blots were probed with secondary antibodies goat anti-rat (1:5,000; 112-035-068, Jackson ImmunoResearch) and goat anti-rabbit (1:5,000; 170-6515, Bio-Rad) conjugated to horseradish peroxidase (HRP) and visualized using an ECL detection kit (Thermo Scientific Pierce).

Cell fractionation was performed as described previously with minor modifications (Méndez and Stillman, 2000). 1×10^7 ESCs were resuspended in 250 µL of buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sucrose, 10% glycerol, 0.1% Triton X-100, 1 mM DTT, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1x mammalian protease inhibitor cocktail (PI; Roche)) and incubated for 5 min on ice. Nuclei were collected by centrifugation (4 min, $1,300 \times g$, 4 °C) and the cytoplasmic fraction (supernatant) was cleared again by centrifugation (15 min, $20,000 \times g$, 4 °C). Nuclei were washed once with buffer A, and then lysed in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT,

RESULTS

1 mM PMSF, 1x PI). Insoluble chromatin was collected by centrifugation (4 min, $1,700 \times g$, 4 °C) and washed once with buffer B. Chromatin fraction was lysed with 1x Laemmli buffer and boiled (10 min, 95°C).

As markers of cytoplasmic and chromatin fractions, alpha-tubulin and histone H3 were detected using monoclonal antibody (mouse anti-alpha-Tubulin, Sigma T9026 or rat anti-Tubulin, Abcam ab6160) and polyclonal antibody (rabbit anti-H3, Abcam ab1791). UHRF1 was visualized by rabbit anti-UHRF1 antibody (Citterio et al., 2004). Western blots for DNMT1 were performed as described previously using a monoclonal antibody (rat anti-DNMT1, 14F6) or a polyclonal antibody (rabbit anti-DNMT1, Abcam ab87654)(Mulholland et al., 2015). GFP and FLAG tagged proteins were visualized by mouse anti-GFP (Roche) and anti-FLAG M2 antibodies (Sigma, F3165), respectively.

Quantitative real-time PCR (qRT-PCR) Analysis

Total RNA was isolated using the NucleoSpin Triprep Kit (Macherey-Nagel) according to the manufacturer's instructions. cDNA synthesis was performed with the High-Capacity cDNA Reverse Transcription Kit (with RNase Inhibitor; Applied Biosystems) using 500 ng of total RNA as input. qRT-PCR assays with oligonucleotides listed in Table S5 were performed in 8 μ L reactions with 1.5 ng of cDNA used as input. FastStart Universal SYBR Green Master Mix (Roche) was used for SYBR green detection. The reactions were run on a LightCycler480 (Roche).

LC-MS/MS analysis of DNA samples

Isolation of genomic DNA was performed according to earlier published work ((Pfaffeneder et al., 2014).

1.0–5 μ g of genomic DNA in 35 μ L H₂O were digested as follows: 1) An aqueous solution (7.5 μ L) of 480 μ M ZnSO₄, containing 18.4 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and labeled internal standards were added ([¹⁵N₂]-cadC 0.04301 pmol, [¹⁵N₂,D₂]-hmdC 7.7 pmol, [D₃]-mdC 51.0 pmol, [¹⁵N₅]-8-oxo-dG 0.109 pmol, [¹⁵N₂]-fdC 0.04557 pmol) and the mixture was incubated at 37 °C for 3 h. After addition of 7.5 μ l of a 520 μ M [Na]₂-EDTA solution, containing 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for 3 h at 37 °C and then stored at –20 °C. Prior to LC/MS/MS analysis, samples were filtered by using an AcroPrep Advance 96 filter plate 0.2 μ m Supor (Pall Life Sciences).

Quantitative UHPLC-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer. Natural nucleosides were quantified with the stable isotope dilution technique. An improved method, based on earlier published work (Pfaffeneder et al., 2014; Wagner et al., 2015) was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 L/min (N₂), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 15 L/min (N₂), capillary voltage 2,500 V in the positive ion mode, capillary voltage –2,250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380

V/ 250 V. Delta EMV was set to 500 V for the positive mode. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 μm , 2.1 mm \times 150 mm) at 35 $^{\circ}\text{C}$ using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min: 0 \rightarrow 4 min; 0 \rightarrow 3.5% (v/v) MeCN; 4 \rightarrow 6.9 min; 3.5 \rightarrow 5% MeCN; 6.9 \rightarrow 7.2 min; 5 \rightarrow 80% MeCN; 7.2 \rightarrow 10.5 min; 80% MeCN; 10.5 \rightarrow 11.3 min; 80 \rightarrow 0% MeCN; 11.3 \rightarrow 14 min; 0% MeCN. The effluent up to 1.5 min and after 9 min was diverted to waste by a Valco valve. The autosampler was cooled to 4 $^{\circ}\text{C}$. The injection volume was amounted to 39 μL . Data were processed according to earlier published work (Pfaffeneder et al., 2014).

RNA-seq and Reduced Representation Bisulfite Sequencing (RRBS)

For RNA-seq, RNA was isolated using the NucleoSpin Triprep Kit (Machery-Nagel) according to the manufacturer's instructions. Digital gene expression libraries for RNA-seq were produced using a modified version of single-cell RNA barcoding sequencing (SCRB-seq) optimized to accommodate bulk cells (Ziegenhain et al., 2017) in which a total of 70 ng of input RNA was used for the reverse-transcription of individual samples. For RRBS, genomic DNA was isolated using the QIAamp DNA Mini Kit (QIAGEN), after an overnight lysis and proteinase K treatment. RRBS library preparation was performed as described previously (Boyle et al., 2012), with the following modifications: bisulfite treatment was performed using the EZ DNA Methylation-Gold™ Kit (Zymo Research Corporation) according to the manufacturer's protocol except libraries were eluted in 2 x 20 μL M-elution buffer. RNA-seq and RRBS libraries were sequenced on an Illumina HiSeq 1500.

Targeted Bisulfite Amplicon (TaBA) Sequencing

Genomic DNA was isolated from 10^6 cells using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The EZ DNA Methylation-Gold Kit (Zymo Research) was used for bisulfite conversion according to the manufacturer's instructions with 500 ng of genomic DNA used as input and the modification that bisulfite converted DNA was eluted in 2 x 20 μL Elution Buffer.

The sequences of the locus specific primers (Table S5) were appended with Illumina TruSeq and Nextera compatible overhangs. The amplification of bisulfite converted DNA was performed in 25 μL PCR reaction volumes containing 0.4 μM each of forward and reverse primers, 2 mM Betainitalne (Sigma-Aldrich, B0300-1VL), 10 mM Tetramethylammonium chloride solution (Sigma-Aldrich T3411-500ML), 1x MyTaq Reaction Buffer, 0.5 units of MyTaq HS (Bioline, BIO-21112), and 1 μL of the eluted bisulfite converted DNA (~12.5 ng). The following cycling parameters were used: 5 min for 95 $^{\circ}\text{C}$ for initial denaturation and activation of the polymerase, 40 cycles (95 $^{\circ}\text{C}$ for 20 s, 58 $^{\circ}\text{C}$ for 30 s, 72 $^{\circ}\text{C}$ for 25 s) and a final elongation at 72 $^{\circ}\text{C}$ for 3 min. Agarose gel electrophoresis was used to determine the quality and yield of the PCR.

For purifying amplicon DNA, PCR reactions were incubated with 1.8x volume of CleanPCR beads (CleanNA, CPCR-0005) for 10 min. Beads were immobilized on a DynaMag™-96 Side Magnet (Thermo

Fisher, 12331D) for 5 min, the supernatant was removed, and the beads washed 2x with 150 μ L 70% ethanol. After air drying the beads for 5 min, DNA was eluted in 15 μ L of 10 mM Tris-HCl pH 8.0. Amplicon DNA concentration was determined using the Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Fisher, P7589) and then diluted to 0.7 ng/ μ L. Thereafter, indexing PCRs were performed in 25 μ L PCR reaction volumes containing 0.08 μ M (1 μ L of a 2 μ M stock) each of i5 and i7 Indexing Primers, 1x MyTaq Reaction Buffer, 0.5 units of MyTaq HS (Bioline, BIO-21112), and 1 μ L of the purified PCR product from the previous step. The following cycling parameters were used: 5 min for 95 °C for initial denaturation and activation of the polymerase, 40 cycles (95 °C for 10 s, 55 °C for 30 s, 72 °C for 40 s) and a final elongation at 72 °C for 5 min. Agarose gel electrophoresis was used to determine the quality and yield of the PCR. An aliquot from each indexing reaction (5 μ L of each reaction) was then pooled and purified with CleanPCR magnetic beads as described above and eluted in 1 μ L x Number of pooled reactions. Concentration of the final library was determined using PicoGreen and the quality and size distribution of the library was assessed with a Bioanalyzer. Dual indexed TaBA-seq libraries were sequenced on an Illumina MiSeq in 2x300 bp output mode.

RNA-seq processing and analysis

RNA-seq libraries were processed and mapped to the mouse genome (mm10) using the zUMIs pipeline (Parekh et al., 2018). UMI count tables were filtered for low counts using HTSFilter (Rau et al., 2013). Differential expression analysis was performed in R using DESeq2 (Love et al., 2014) and genes with an adjusted $P < 0.05$ were considered to be differentially expressed. Hierarchical clustering was performed on genes differentially expressed in T1KO and T1CM at ESC and EpiLC stage, respectively, using k-means clustering ($k=4$) in combination with the ComplexHeatmap R-package (Gu et al., 2016). Principal component analysis was restricted to genes differentially expressed during wild-type differentiation and performed using all replicates of wild-type, T1KO and T1CM ESCs and EpiLCs.

RRBS alignment and analysis

Raw RRBS reads were first trimmed using Trim Galore (v.0.3.1) with the '--rrbs' parameter. Alignments were carried out to the mouse genome (mm10) using bsmmap (v.2.90) using the parameters '-s 12 -v 10 -r 2 -I 1'. CpG-methylation calls were extracted from the mapping output using bsmmaps methratio.py. Analysis was restricted to CpG with a coverage > 10 . methylKit (Akalin et al., 2012) was used to identify differentially methylated regions between the respective contrasts for the following genomic features: 1) all 1-kb tiles (containing a minimum of three CpGs) detected by RRBS; 2) Repeats (defined by RepeatMasker); 3) gene promoters (defined as gene start sites $-2\text{kb}/+2\text{kb}$); and 3) gene bodies (defined as longest isoform per gene) and CpG islands (as defined by (Illingworth et al., 2010)). Differentially methylated regions were identified as regions with $P < 0.05$ and a difference in methylation means between two groups greater than 20%. Principal component analysis of global DNA methylation profiles was performed on single CpGs using all replicates of wild-type, T1KO and T1CM ESCs and EpiLCs.

Chromatin immunoprecipitation (ChIP) and Hydroxymethylated-DNA immunoprecipitation (hMeDIP) alignment and analysis.

ChIP-seq reads for TET1 binding in ESCs and EpiLCs were downloaded from GSE57700 (Xiong et al., 2016) and PRJEB19897 (Khoueiry et al., 2017), respectively. hMeDIP reads for wild-type ESCs and T1KO ESCs were download from PRJEB13096 (Khoueiry et al., 2017). Reads were aligned to the mouse genome (mm10) with Bowtie (v.1.2.2) with parameters ‘-a -m 3 -n 3 --best --strata’. Subsequent ChIP-seq analysis was carried out on data of merged replicates. Peak calling and signal pile up was performed using MACS2 callpeak (Zhang et al., 2008) with the parameters ‘--extsize 150’ for ChIP, ‘--extsize 220’ for hMeDIP, and ‘--nomodel -B --nolambda’ for all samples. Tag densities for promoters and 1kb Tiles were calculated using the deepTools2 computeMatrix module (Ramírez et al., 2016). TET1 bound genes were defined by harboring a TET1 peak in the promoter region (defined as gene start sites -2kb/+2kb).

Immunofluorescence staining

For immunostaining, naïve ESCs were grown on coverslips coated with Geltrex (Life Technologies) diluted 1:100 in DMEM/F12 (Life Technologies), thereby allowing better visualization of the cytoplasm during microscopic analysis. All steps during immunostaining were performed at room temperature. Coverslips were rinsed two times with PBS (pH 7.4; 140 mM NaCl, 2.7 mM KCl, 6.5 mM Na₂HPO₄, 1.5 mM KH₂PO₄) prewarmed to 37°C, cells fixed for 10 min with 4% paraformaldehyde (pH 7.0; prepared from paraformaldehyde powder (Merck) by heating in PBS up to 60°C; store at -20°C), washed three times for 10 min with PBST (PBS, 0.01% Tween20), permeabilized for 5 min in PBS supplemented with 0.5% Triton X-100, and washed two times for 10 min with PBS. Primary and secondary antibodies were diluted in blocking solution (PBST, 4% BSA). Coverslips were incubated with primary and secondary antibody solutions in dark humid chambers for 1 h and washed three times for 10 min with PBST after primary and secondary antibodies. For DNA counterstaining, coverslips were incubated 6 min in PBST containing a final concentration of 2 µg/mL DAPI (Sigma-Aldrich) and washed three times for 10 min with PBST. Coverslips were mounted in antifade medium (Vectashield, Vector Laboratories) and sealed with colorless nail polish.

Following primary antibodies were used: polyclonal rabbit anti-DPPA3 (1:200; ab19878, Abcam) and monoclonal rabbit anti-UHRF1 (1:250; D6G8E, Cell Signaling). Following secondary antibodies were used: polyclonal donkey anti-rabbit conjugated to Alexa 488 (1:500; 711-547-003, Dianova), polyclonal donkey anti-rabbit conjugated to Alexa 555 (1:500; A31572, Invitrogen), and polyclonal donkey anti-rabbit conjugated to DyLight fluorophore 594 (1:500; 711-516-152, Dianova).

Immunofluorescence and Live-cell imaging

For immunofluorescence, stacks of optical sections were collected on a Nikon TiE microscope equipped with a Yokogawa CSU-W1 spinning disk confocal unit (50 µm pinhole size), an Andor Borealis illumination unit, Andor ALC600 laser beam combiner (405nm/488nm/561nm/640nm), Andor IXON 888 Ultra EMCCD camera, and a Nikon 100x/1.45 NA oil immersion objective. The microscope was

controlled by software from Nikon (NIS Elements, ver. 5.02.00). DAPI or fluorophores were excited with 405 nm, 488 nm, or 561 nm laser lines and bright-field images acquired using Nikon differential interference contrast optics. Confocal image z-stacks were recorded with a step size of 200 nm, 16-bit image depth, 1x1 binning, a frame size of 1024×1024 pixels, and a pixel size of 130 nm. Within each experiment, cells were imaged using the same settings on the microscope (camera exposure time, laser power, and gain) to compare signal intensities between cell lines.

For live-cell imaging, cells were plated on Geltrex-coated glass bottom 2-well imaging slides (Ibidi). Both still and timelapse images were acquired on the Nikon spinning disk system described above equipped with an environmental chamber maintained at 37°C with 5% CO₂ (Oko Labs), using a Nikon 100x/1.45 NA oil immersion objective and a Perfect Focus System (Nikon). Images were acquired with the 488, 561, and 640 nm laser lines, full-frame (1024x1024) with 1x1 binning, and with a pixel size of 130 nm. Transfection of a RFP-PCNA vector (Sporbert et al., 2005) was used to identify cells in S-phase. For DNA staining in live cells, cells were exposed to media containing 200 nM SiR-DNA (Spirochrome) for at least 1 h before imaging. For imaging endogenous DPPA3-HALO in live cells, cells were treated with media containing 50 nM HaloTag-TMR fluorescent ligand (Promega) for 1 h. After incubation, cells were washed 3x with PBS before adding back normal media. Nuclear export inhibition was carried out using media containing 20 nM leptomycin-B (Sigma-Aldrich).

Image analysis

For immunofluorescence images, Fiji software (ImageJ 1.51j) (Schindelin et al., 2012; Schneider et al., 2012) was used to analyze images and create RGB stacks. For analysis of live-cell imaging data, CellProfiler Software (version 3.0)(McQuin et al., 2018) was used to quantify fluorescence intensity in cells stained with SiR-DNA. CellProfiler pipelines used in this study are available upon request. In brief, the SiR-DNA signal was used to segment ESC nuclei. Mean fluorescence intensity of GFP was measured both inside the segmented area (nucleus) and in the area extending 4-5 pixels beyond the segmented nucleus (cytoplasm). GFP fluorescence intensity was normalized by subtracting the experimentally-determined mean background intensity and background-subtracted GFP intensities were then used for all subsequent quantifications shown in Figure 4D, S4H, S5H, S6B, and S6C.

Fluorescence Recovery After Photobleaching (FRAP)

For FRAP assays, cells cultivated on Geltrex-coated glass bottom 2-well imaging slides (Ibidi) were imaged in an environmental chamber maintained at 37°C with 5% CO₂ either using the Nikon system mentioned above equipped with a FRAPPA photobleaching module (Andor) or on an Ultraview-Vox spinning disk system (Perkin-Elmer) including a FRAP Photokinesis device mounted to an inverted Axio Observer D1 microscope (Zeiss) equipped with an EMCCD camera (Hamamatsu) and a 63x/1.4 NA oil immersion objective, as well as 405, 488 and 561 nm laser lines.

For endogenous UHRF1-GFP FRAP, eight pre-bleach images were acquired with the 488 nm laser, after which an area of 4 x 4 pixels was irradiated for a total of 16 ms with a focused 488 nm laser (leading to

a bleached spot of ~1 μm) to bleach a fraction of GFP-tagged molecules within cells, and then recovery images were acquired every 250 ms for 1 - 2 min. Recovery analysis was performed in Fiji. Briefly, fluorescence intensity at the bleached spot was measured in background-subtracted images, then normalized to pre-bleach intensity of the bleached spot, and normalized again to the total nuclear intensity in order to account for acquisition photobleaching. Images of cells with visible drift were discarded.

Xenopus egg extracts

The interphase extracts (low-speed supernatants (LSS) were prepared as described previously (Nishiyama et al., 2013). After thawing, LSS were supplemented with an energy regeneration system (5 $\mu\text{g}/\text{ml}$ creatine kinase, 20 mM creatine phosphate, 2 mM ATP) and incubated with sperm nuclei at 3,000-4,000 nuclei per μl . Extracts were diluted 5-fold with ice-cold CPB (50 mM KCl, 2.5 mM MgCl_2 , 20 mM HEPES-KOH, pH 7.7) containing 2% sucrose, 0.1% NP-40 and 2 mM NEM, overlaid onto a 30% sucrose/CPB cushion, and centrifuged at 15,000g for 10 min. The chromatin pellet was resuspended in SDS sample buffer and analyzed by SDS-PAGE. GST-mDPPA3 was added to egg extracts at 50 $\text{ng}/\mu\text{l}$ at final concentration.

Monitoring DNA methylation in Xenopus egg extracts

DNA methylation was monitored by the incorporation of S-[methyl- ^3H]-adenosyl-L-methionine, incubated at room temperature, and the reaction was stopped by the addition of CPB containing 2% sucrose up to 300 μl . Genomic DNA was purified using a Wizard Genomic DNA purification kit (Promega) according to the manufacturer's instructions. Incorporation of radioactivity was quantified by liquid scintillation counter.

Plasmid construction for Recombinant mDPPA3

To generate GST-tagged mDPPA3 expression plasmids, mDPPA3 fragment corresponding to full-length protein was amplified by PCR using mouse DPPA3 cDNA and primers 5'-TTAGCAGCCGGATCCCTAATTCTTCCCGATTTTCGCA-3' and 5'-CGTGGATCCCCGAATTCATGGAGGAACCATCAGAGAAAGTC'. The resulting DNA fragment was cloned into pGEX4T-3 vector digested with EcoRI and SalI using an In-Fusion HD Cloning Kit.

Protein expression and purification

For protein expression in *Escherichia coli* (BL21-CodonPlus), the mDPPA3 genes were transferred to pGEX4T-3 vector as described above. Protein expression was induced by the addition of 0.1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) to media followed by incubation for 12 hour at 20°C. For purification of Glutathione S transferase (GST) tagged proteins, cells were collected and re-suspended in Lysis buffer (20 mM HEPES-KOH (pH 7.6), 0.5M NaCl, 0.5 mM EDTA, 10% glycerol, 1 mM DTT) supplemented with 0.5% NP40 and protease inhibitors, and were then disrupted by sonication on ice. After centrifugation, the supernatant was applied to Glutathione Sepharose (GSH) beads (GE Healthcare) and rotated at 4°C for 2 hour. Beads were then washed three times with Wash buffer 1 (20 mM Tris-HCl

(pH 8.0), 150 mM NaCl, 1% TritonX-100, 1 mM DTT) three times and with Wash buffer 2 (100 mM Tris-HCl (pH 7.5), 100 mM NaCl) once. Bound proteins were eluted in Elution buffer (100 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% glycerol, 1 mM DTT) containing 42 mM reduced Glutathione and purified protein was loaded on PD10 desalting column equilibrated with EB buffer (10 mM HEPES/KOH at pH 7.7, 100 mM KCl, 0.1 mM CaCl₂, 1 mM MgCl₂) containing 1 mM DTT, and then concentrated by Vivaspin (Millipore).

Data collection for the presence of TET1, UHRF1, DNMT1 and DPPA3 throughout metazoa

Megabat (*Pteropus vampyrus*) protein sequences for TET1 (ENSVPAP00000010999), UHRF1 (ENSPVAP00000002809), DNMT1 (ENSPVAP00000003477) and DPPA3 (ENSVPAP00000004109) were subjected to three iterations of PSI-blast (Altschul et al., 1997), respectively. Subsequently, result tables were filtered to contain correct gene names using a custom python script. Presence of the proteins throughout metazoa was visualized using iTOL (Letunic and Bork, 2007).

Chromatin immunoprecipitation coupled to Mass Spectrometry and Proteomics data analysis

For Chromatin immunoprecipitation coupled to Mass Spectrometry (ChIP-MS), whole cell lysates of the doxycycline inducible Dppa3-FLAG mES cells were used by performing three separate immunoprecipitations with an anti-FLAG antibody and three samples with a control IgG. Proteins were digested on the beads after the pulldown and desalted subsequently on StageTips with three layers of C18 (Rappsilber et al., 2007). Here, peptides were separated by liquid chromatography on an Easy-nLC 1200 (Thermo Fisher Scientific) on in-house packed 50 cm columns of ReproSil-Pur C18-AQ 1.9- μ m resin (Dr. Maisch GmbH). Peptides were then eluted successively in an ACN gradient for 120 min at a flow rate of around 300 nL/min and were injected through a nanoelectrospray source into a Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific). After measuring triplicates of a certain condition, an additional washing step was scheduled. During the measurements, the column temperature was constantly kept at 60°C while after each measurement, the column was washed with 95% buffer B and subsequently with buffer A. Real time monitoring of the operational parameters was established by SprayQc (Scheltema and Mann, 2012) software. Data acquisition was based on a top10 shotgun proteomics method and data-dependent MS/MS scans. Within a range of 400-1650 m/z and a max. injection time of 20 ms, the target value for the full scan MS spectra was 3×10^6 and the resolution at 60,000.

The raw MS data was then analyzed with the MaxQuant software package (version 1.6.0.7)(Cox and Mann, 2008). The underlying FASTA files for peak list searches were derived from Uniprot (UP000000589_10090.fasta and UP000000589_10090_additional.fasta, version June 2015) and an additional modified FASTA file for the FLAG-tagged Dppa3 in combination with a contaminants database provided by the Andromeda search engine (Cox et al., 2011) with 245 entries. During the MaxQuant-based analysis the “Match between runs” option was enabled and the false discovery rate was set to 1% for both peptides (minimum length of 7 amino acids) and proteins. Relative protein amounts

were determined by the MaxLFQ algorithm (Cox et al., 2014), with a minimum ratio count of two peptides.

For the downstream analysis of the MaxQuant output, the software Perseus (Tyanova et al., 2016) (version 1.6.0.9) was used to perform Student's *t*-test with a permutation-based FDR of 0.05 and an additional constant $S_0 = 1$ in order to calculate fold enrichments of proteins between triplicate chromatin immunoprecipitations of anti-FLAG antibody and control IgG. The result was visualized in a scatter plot.

For GO analysis of biological processes the Panther classification system was used (Tyanova et al., 2016). For the analysis, 131 interactors of Dppa3 were considered after filtering the whole amount of 303 significant interactors for a p-value of at least 0.0015 and 3 or more identified peptides. The resulting GO groups were additionally filtered for a fold enrichment of observed over expected amounts of proteins of at least 4 and a p-value of 5.30 E-08.

Dppa3 overexpression in medaka embryos and immunostaining

Medaka d-rR strain was used. Medaka fish were maintained and raised according to standard protocols. Developmental stages were determined based on a previous study (Iwamatsu, 2004). Dppa3 and mutant Dppa3 (R107E) mRNA were synthesized using HiScribe T7 ARCA mRNA kit (NEB, E2060S), and purified using RNeasy mini kit (QIAGEN, 74104). Dppa3 or mutant Dppa3 (R107E) mRNA was injected into the one-cell stage (stage 2) medaka embryos. After 7 hours of incubation at 28°C, the late blastula (stage 11) embryos were fixed with 4% PFA in PBS for 2 hours at room temperature, and then at 4°C overnight. Embryos were dechorionated, washed with PBS, and permeabilized with 0.5% Triton X-100 in PBS for 30 minutes at room temperature. DNA was denatured in 4 M HCl for 15 minutes at room temperature, followed by neutralization in 100 mM Tris-HCl (pH 8.0) for 20 minutes. After washing with PBS, embryos were blocked in blocking solution (2% BSA, 1% DMSO, 0.2% Triton X-100 in PBS) for 1 hour at room temperature, and then incubated with 5-methylcytosine antibody (1:200; Active Motif #39649) at 4 °C overnight. The embryos were washed with PBSDT (1% DMSO, 0.1% Triton X-100 in PBS), blocked in blocking solution for 1 hours at room temperature, and incubated with Alexa Fluor 555 goat anti-mouse 2nd antibody (1:500; ThermoFisher Scientific #A21422) at 4 °C overnight. After washing with PBSDT, cells were mounted on slides and examined under a fluorescence microscope.

Fluorescence Three Hybrid (F3H) assay

The F3H assay was performed as described previously (Herce et al., 2013). In brief, BHK cells containing multiple lac operator repeats were transiently transfected with the respective GFP- and dsRed-constructs on coverslips using PEI and fixed with 3.0 % formaldehyde 24 hrs after transfection. For DNA counterstaining, coverslips were incubated in a solution of DAPI (200 ng/ml) in PBS-T and mounted in Vectashield. Images were collected using a Leica TCS SP5 confocal microscope. To quantify the interactions within the lac spot, the following intensity ratio was calculated for each cell in order to account for different expression levels: $(mCherry_{spot} - mCherry_{background}) / (GFP_{spot} - GFP_{background})$.

Microscale Thermophoresis (MST)

For MST measurements, mUHRF1 C-terminally tagged with GFP- and 6xHis-tag was expressed in HEK 293T cells and then purified using Qiagen Ni-NTA beads (Qiagen #30230). Recombinant mDPPA3 WT and 1-60 were purified as described above. Purified Uhrf1 (200 nM) was mixed with different concentrations of purified Dppa3 (0.15 nM to 5 μ M) followed by a 30 min incubation on ice. The samples were then aspirated into NT.115 Standard Treated Capillaries (NanoTemper Technologies) and placed into the Monolith NT.115 instrument (NanoTemper Technologies). Experiments were conducted with 80% LED and 80% MST power. Obtained fluorescence signals were normalized (F_{norm}) and the change in F_{norm} was plotted as a function of the concentration of the titrated binding partner using the MO. Affinity Analysis software version 2.1 (NanoTemper Technologies). For fluorescence normalization ($F_{\text{norm}}=F_{\text{hot}}/F_{\text{cold}}$), the manual analysis mode was selected and cursors were set as follows: $F_{\text{cold}}=-1$ to 0 s, $F_{\text{hot}}=10$ to 15 s. The K_d was obtained by fitting the mean F_{norm} of eight data points (four independent replicates, each measured as a technical duplicate).

RICS

Data for Raster Image Correlation Spectroscopy (RICS) was acquired on a home-built laser scanning confocal setup equipped with a pulsed interleaved excitation (PIE) system as used elsewhere (Hendrix et al., 2015). Samples were excited using pulsed lasers at 470 (Picoquant) and 561 nm (Toptica Photonics), synchronized to a master clock, and then delayed ~ 20 ns relative to one another to achieve PIE. Laser excitation was separated from descanned fluorescence emission by a Di01-R405/488/561/635 polychroic mirror (Semrock, AHF Analysentechnik) and eGFP and mScarlet fluorescence emission was separated by a 565 DCXR dichroic mirror (AHF Analysentechnik) and collected on avalanche photodiodes, a Count Blue (Laser Components) and a SPCM-AQR-14 (Perkin-Elmer) with 520/40 and a 630/75 emission filters (Chroma, AHF Analysentechnik). Detected photons were recorded by time-correlated single-photon counting.

The alignment of the system was verified prior to each measurement session by performing FCS with PIE of a mixture of Atto-488 and Atto565 dyes, excited with pulsed 470 and 561 nm lasers set to 10 μ W (measured in the collimated space before entering the galvo-scanning mirror system), 1 μ m above the surface of the coverslip (Müller et al., 2005). Cells were plated on Ibidi two-well glass bottom slides, and induced with doxycycline overnight prior to measurements. Scanning was performed in cells at maintained at 37C with a stage top incubator, with a total field of view of 12 x 12 μ m, composed of 300 pixels x 300 lines a pixel size of 40 nm, a pixel dwell time of 11 μ s, a line time 3.33 ms. , at one frame per second, for 100-200 seconds. Pulsed 470 and 561 nm lasers were adjusted to 4 and 5 μ W respectively.

Image analysis was done using the Pulsed Interleaved Excitation Analysis with Matlab (PAM) software (Schrimpf et al., 2018). Briefly, time gating of the raw photon stream was performed by selecting only photons emitted on the appropriate detector after pulsed excitation, thereby allowing to generate cross-talk free images for each channel. Then, using Microtime Image Analysis (MIA), slow fluctuations were

removed by subtracting a moving average of 3 frames, and a region of interest corresponding to the nucleus was selected, excluding nucleoli and dense aggregates. The spatial autocorrelation and cross-correlation functions (SACF and SCCF) were calculated as done previously (Hendrix et al., 2016) using arbitrary region RICS (Schrimpf et al. 2016):

$$G(\xi, \psi) = \frac{\langle I_{RICS,1}(x, y) I_{RICS,2}(x + \xi, y + \psi) \rangle_{XY}}{\langle I_{RICS,1} \rangle_{XY} \langle I_{RICS,2} \rangle_{XY}}$$

where ξ and ψ are the correlation lags in pixel units along the x- and y-axis scan directions. The correlation function was then fitted to a two-component model (one mobile and one immobile component) in MIAfit:

$$G_{fit}(\xi, \psi) = A_{mob} G_{fit, mob}(\xi, \psi) + A_{imm} \exp(-\delta r^2 \omega_{imm}^{-2} (\xi^2 + \psi^2)) + y_0, \text{ where}$$

$$G_{fit, mob}(\xi, \psi) = \left(1 + \frac{4D(\tau_p \xi + \tau_i \psi)}{\omega_r^2}\right)^{-1} \left(1 + \frac{4D(\tau_p \xi + \tau_i \psi)}{\omega_z^2}\right)^{-1/2} \cdot \exp\left(-\frac{\delta r^2 (\xi^2 + \psi^2)}{\omega_r^2 + 4D(\tau_p \xi + \tau_i \psi)}\right)$$

,

which yields parameters such as the diffusion coefficient (D) and the amplitudes of the mobile and immobile fractions (A_{mob} and A_{imm}). The average number of mobile molecules per excitation volume on the RICS timescale was determined by:

$$N_{mob} = \frac{\gamma}{A_{mob}} \frac{2\Delta F}{2\Delta F + 1}$$

where γ is a factor pertaining to the 3D Gaussian shape of the PSF, and $2\Delta F/(2\Delta F+1)$ is a correction factor when using a moving average subtraction prior to calculating the SACF. The bound fraction is the contribution of particles which remain visible during the acquisition of 5-10 lines of the raster scan, corresponding to ~30 ms. The cross-correlation model was fitted to the cross-correlation function, and the extent of cross-correlation was calculated from the amplitude of the mobile fraction of the cross-correlation fit divided by the amplitude of the mobile fraction of the autocorrelation fit of DPPA3-mScarlet.

Data and Code Availability

Sequencing data reported in this paper are available at ArrayExpress (EMBL-EBI) under accessions E-MTAB-6785 (wild-type and Tet catalytic mutants RRBS), E-MTAB-6797 (RNA-seq), E-MTAB-6800 (Dppa3KO RRBS). The raw mass spectrometry proteomics data from the FLAG-DPPA3 pulldown will be deposited at the ProteomeXchange Consortium via the PRIDE partner repository. Significant interactors of FLAG-DPPA3 in ESCs can be found in Table S3.

Statistics and reproducibility

No statistical methods were used to predetermine sample size, the experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment. All statistical tests are clearly described in the figure legends and/or in the Methods section, and P values or adjusted P values are given where possible. For all RRBS experiments, data are derived from $n = 2$ biological replicates. For bisulfite sequencing of LINE-1 elements $n = 2$ biological replicates were analyzed from 2 independent clones for T1CM, T2CM, T12CM, and Dppa3KO ESCs or 2 independent cultures for wt ESCs. For all RNA-seq experiments $n = 4$ biological replicates corresponding to $n = 2$ independently cultured samples from 2 clones (T1KO and T1CM) or 4 independently cultured samples (wild-type) were used. LC-MS/MS quantification (Figure 3e) was performed on $n = 4$ biological replicates corresponding to $n = 2$ independently cultured samples from 2 independent clones (T1CM, T2CM, T12CM, and Dppa3KO ESCs) or 4 independently cultured samples (wild-type ESCs)

ACKNOWLEDGEMENTS

We would like to thank Dr. H. Blum, Dr. S. Krebs (Laboratory for Functional Genome Analysis, LMU Munich), and Dr. A. Brachmann (Faculty of Biology, Department of Genetics, LMU Munich) for next generation sequencing services. We further thank Dr. Ralf Heermann (Bioanalytics core facility, LMU Munich) for help with the MST measurements. We thank V. Laban, C. Kirschner, T.J. Fischer, D. Nixdorf and V.B. Ozan for help with experiments, J. Koch for technical assistance, Dr. M. Mann and Dr. I. Paron for mass spectrometry and Dr. I. Hellmann for advice on data analysis and providing computational infrastructure. We would like to thank Dr. S. Stricker and Dr. I. Solovei for helpful discussions and constructive criticism on the manuscript. We thank Dr. Feng Zhang for providing the pSpCas9(BB)-2A-Puro (PX459) and SpCas9-T2A-GFP (PX458) (Addgene plasmids #62988 and 48138) plasmids, Dr. Eric Kowarz for the gift of the pSB-tetPur (Addgene plasmid #60507) construct, and Dr. Zsuzsanna Izsvak for the gift of the pCMV(CAT)T7-SB100 (Addgene plasmid # 34879). J.R., C.T. and M.D.B. are fellows of the International Max Planck Research School for Molecular Life Sciences (IMPRS-LS). C.B.M. gratefully acknowledges the support of the Fulbright Commission and the late Dr. Glenn Cuomo. F.R.T. thanks the Boehringer Ingelheim Fonds and J.R. the Fonds de Recherche du Québec en Santé for a PhD fellowship.

FUNDING

The work was funded by the Deutsche Forschungsgemeinschaft (DFG grants SFB1064/A22 to S.B., SFB1064/A17 and SFB1243/A01 to H.L., and SFB1243/A14 to W.E.).

AUTHOR CONTRIBUTIONS

C.B.M. and S.B. designed and conceived the study. S.B. and H.L. supervised the study. C.M., S.B., and H.L. prepared the manuscript with the help of M.D.B. C.B.M. performed cellular and molecular experiments. C.B.M. generated cell lines with help from M.Y.. C.B.M. performed RRBS and RNA-Seq with help and supervision from C.Z., S.B., and W.E.. J.R. and C.B.M. performed live-cell microscopy and photobleaching analyses. I.G., J.R. and C.B.M. performed RICS experiments under the supervision of D.L.. C.T. performed MST and F3H assays. A.N. performed *Xenopus* experiments under the supervision of M.N.. R.N. performed the experiments in medaka embryos under the supervision of H.T.. M.D.B. and P.S. helped with cell line validation and performed fluorescence microscopy analysis. W.Q. performed the biochemical analyses with assistance from A.A. M.M performed hESC experiments. E.U. conducted proteomics experiments and analyses under the guidance of M.W. F.R.T. and E.P. quantified modified cytosines by LC-MS/MS with the supervision by T.C. S.B. performed data analysis. All authors read, discussed, and approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

SUPPLEMENTARY FIGURES

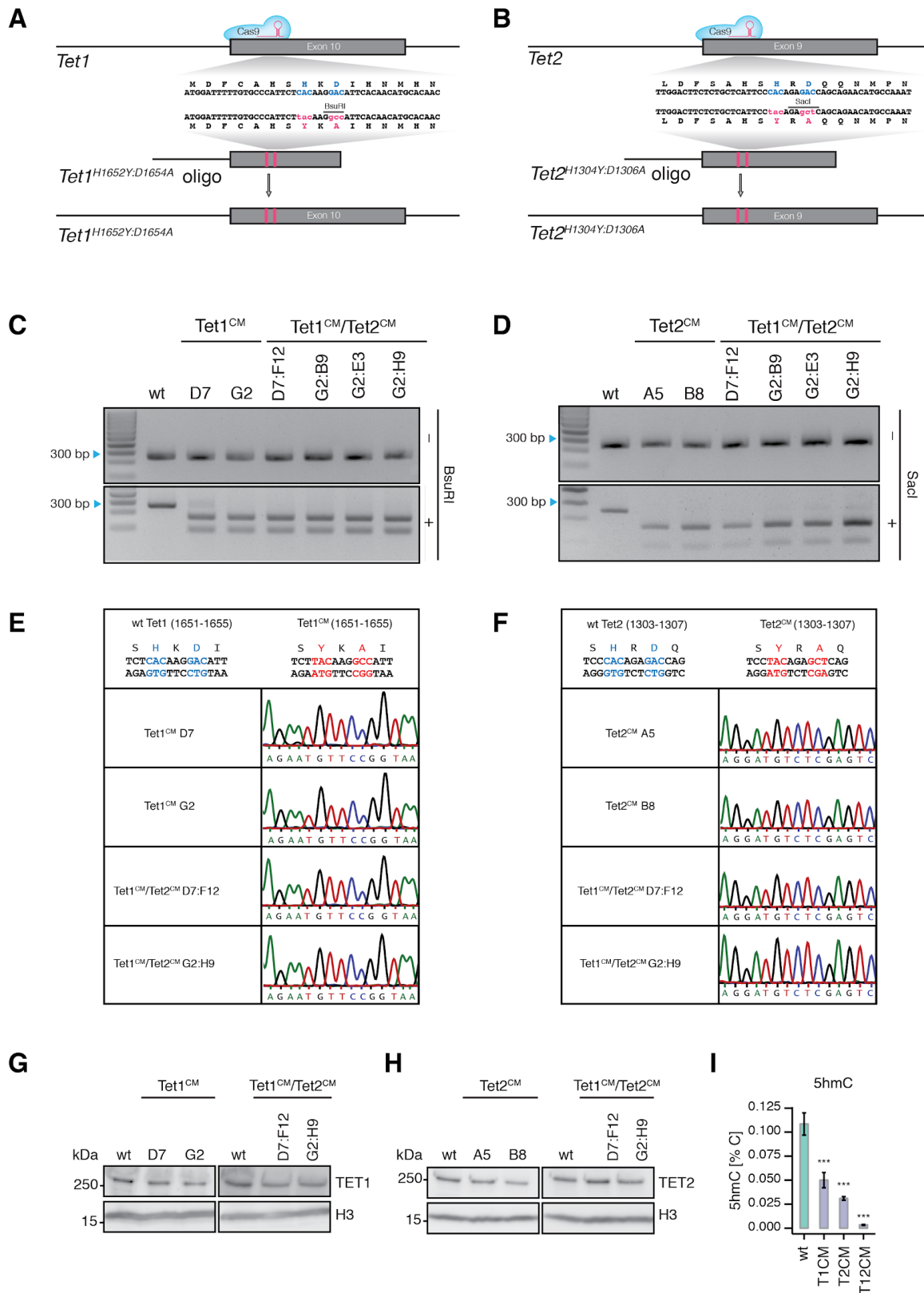


Figure S1: Generation and characterization of T1CM, T2CM, and T12CM mESCs

(**A and B**) Schematic representation of CRISPR/Cas9 gene editing strategy used to mutate the catalytic center (HxD) of *Tet1* (**A**) and *Tet2* (**B**). gRNA target sequences and restriction enzyme recognition sites for restriction fragment length polymorphism (RFLP) screening are shown (See Table S5). (**C and D**) Genotyping using RFLP analysis of the *Tet1* (**C**) and *Tet2* (**D**) locus. (**E and F**) Sanger sequencing results confirming the successful insertion of point mutations, YxA, at the *Tet1* (**G**) and *Tet2* (**H**) locus, HxD. (**G and H**) Immunoblot detection of endogenous TET1 (**G**) and TET2 (**H**) protein levels in

T1CM, T2CM, and T12CM mESCs. **(I)** DNA modification levels as percentage of total cytosines in wt (n = 24), T1CM (n = 8), T2CM (n = 12), and T12CM (n = 11) mESCs. Depicted are mean values \pm standard deviation; *P < 0.05, **P < 0.01, ***P < 0.005 to wt as determined using a one-way ANOVA followed by a post-hoc Tukey HSD test.

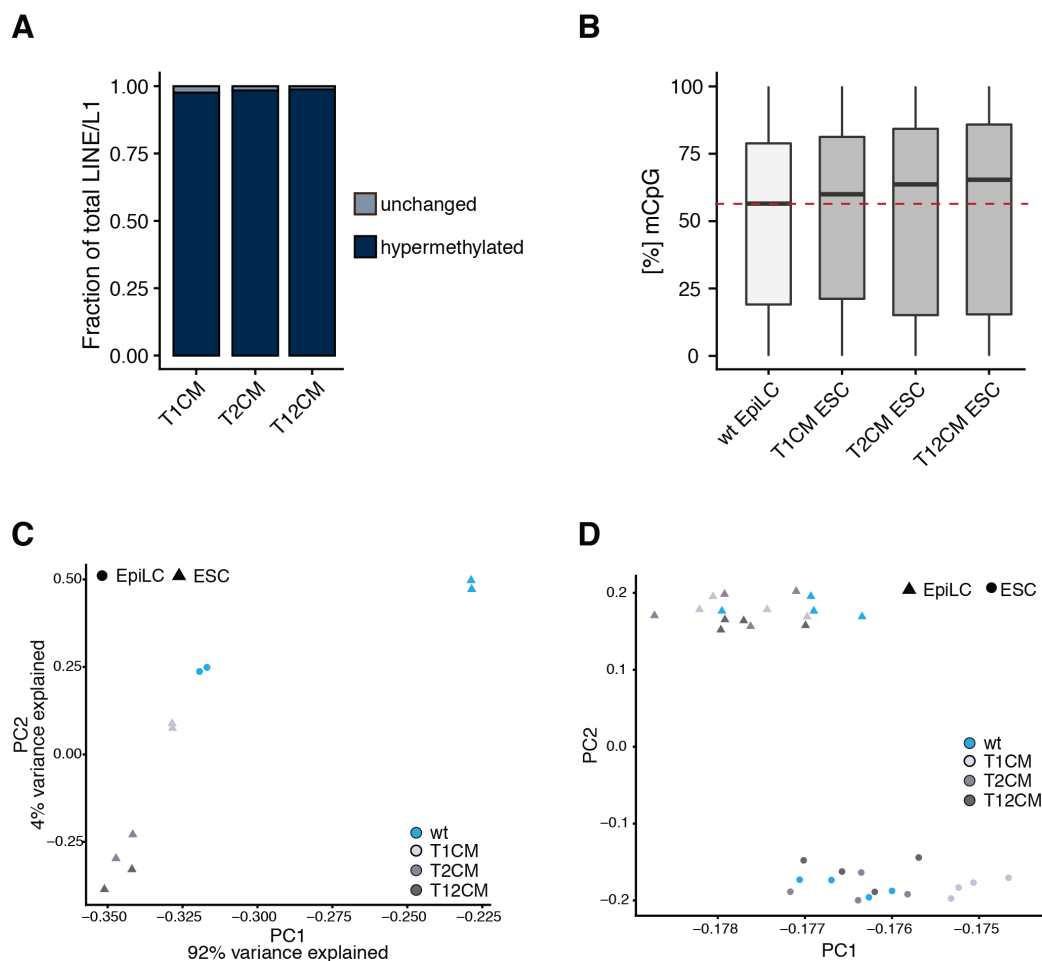


Figure S2: Methyloome and transcriptome analysis of T1CM, T2CM, and T12CM ESCs

(A) Relative proportion of DNA hypermethylation (q value < 0.05; absolute methylation difference > 20%) at LINE1/L1 elements in T1CM, T2CM, and T12CM ESCs compared to wt ESCs. **(B)** Loss of TET catalytic activity in ESCs results in similar or higher DNA methylation levels than in wt EpiLCs. Percentage of total CpG DNA methylation (5mCpG) as measured by RRBS. Horizontal black lines within boxes represent median values, boxes indicate the upper and lower quartiles, and whiskers indicate the 1.5 interquartile range. Dashed red line indicates the median mCpG methylation in wt EpiLCs. **(C)** Principal component (PC) analysis of RRBS data from wt, T1CM, T2CM, and T12CM ESCs and wt EpiLCs. **(D)** PC analysis of RNA-seq data from wt, T1CM, T2CM, and T12CM ESCs during EpiLC differentiation.

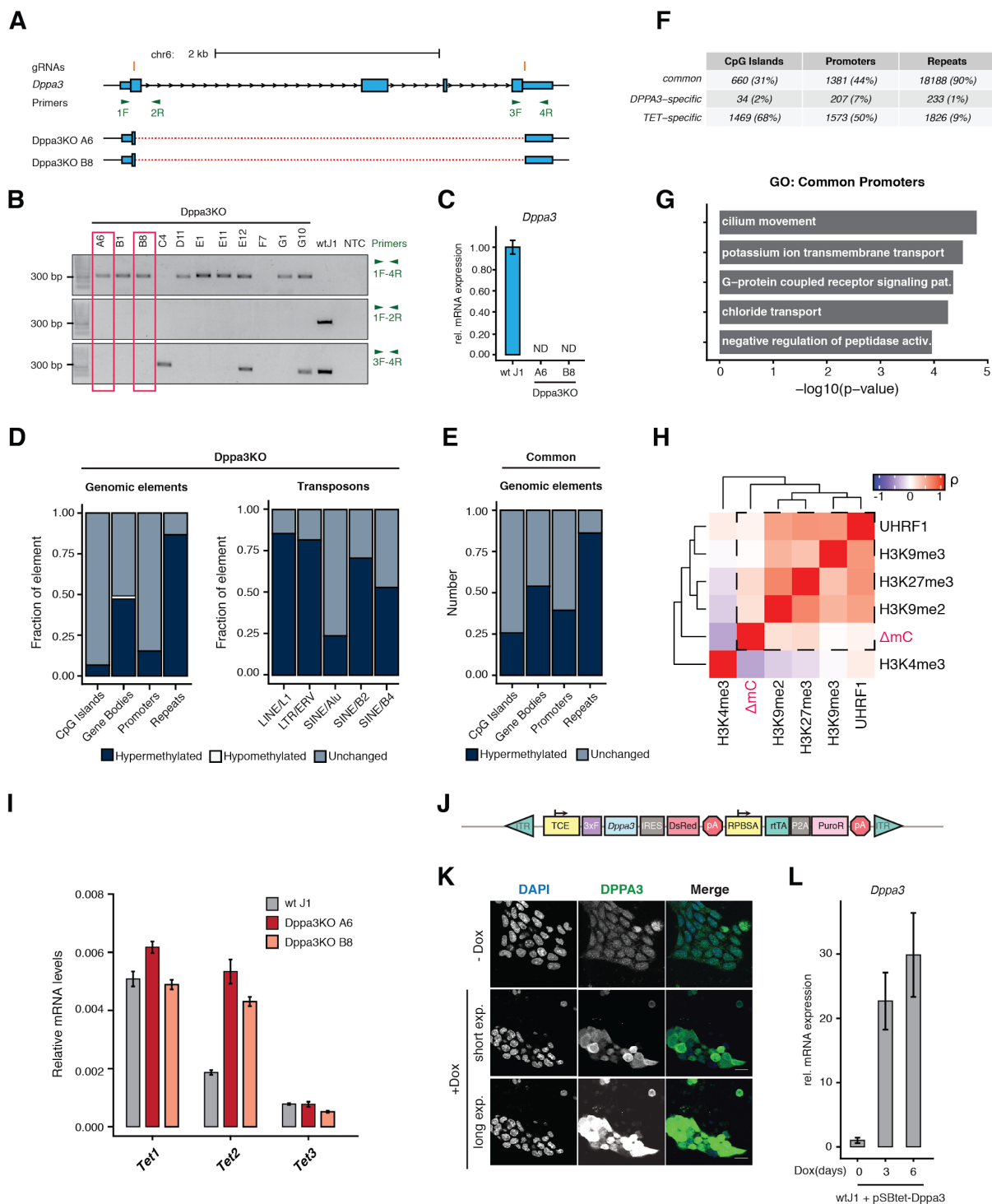


Figure S3: Generation, characterization, and methylome profiling of Dppa3KO ESCs

(A) Schematic representation of the CRISPR/Cas9 gene editing strategy used to excise the entire *Dppa3* locus. The position of the two locus-flanking gRNAs are shown in orange. PCR primers for determination of locus removal and zygosity are indicated in green. (B) Results of PCR using the primers indicated in (A). *Dppa3* knockout (*Dppa3*KO) clones, A6 and B8, chosen for further experiments are indicated by the red boxes. The wild-type (wt) and no template control (NTC) reactions are depicted on the right. The sequence alignment of the amplicon generated using primers 1 and 4 for *Dppa3*KO clones A6 and B8 is provided in the lower portion of (A) with solid boxes and the dashed lines representing successfully aligned sequences and genomic sequences not found in the sequenced amplicons respectively. (C) *Dppa3* transcript levels of the two *Dppa3*KO clones A6 and B8 are depicted relative to mRNA levels in wtJ1 ESCs after normalization to *Gapdh*. Error bars indicate mean \pm SD calculated from technical triplicate reactions from n=4 biological replicates. N.D., expression not detectable. (D) Relative proportion of DNA hypermethylation occurring at each genomic element or retrotransposon class in *Dppa3*KO ESCs. (E) Relative proportion of DNA hypermethylation common to T1CM, T2CM, and T12CM ESCs and *Dppa3*KO ESCs at each genomic element. (F) Summary of differentially methylated regions either unique to TET mutants

(TET-specific) and *Dppa3*KO ESCs (DPPA3-specific) or shared among TET mutants and *Dppa3*KO ESCs (common). **(G)** Gene ontology (GO) terms associated with promoters specifically dependent on TET activity; adjusted p-values calculated using Fisher's exact test. **(H)** Cross-correlation analysis between DNA hypermethylation (Δ mC) in *Dppa3*KO ESCs and genomic occupancy of histone modifications and UHRF1 binding. Correlations are calculated over 1kb tiles. Positive correlations with Δ mC are bounded by a dashed-line square. **(I)** Hypermethylation in *Dppa3* KO ESCs is not associated with *Tet* downregulation. Expression of *Tet* genes in wt *Dppa3*KO ESC clones depicted as mRNA levels relative to *Gapdh*. Error bars indicate mean \pm SD calculated from technical triplicate reactions from n=2 biological replicates. **(J)** Schematic representation of the pSBtet-D3 (pSBtet-3xFLAG-*Dppa3*-IRES-DsRed) cassette for the Sleeping Beauty transposition-mediated generation of doxycycline (Dox) inducible *Dppa3* ESC lines. Abbreviations: inverted terminal repeat (ITR), tetracycline response element plus minimal CMV (TCE), 3xFLAG tag (3xF), internal ribosomal entry site (IRES), polyA signal (pA), constitutive RPBSA promoter (RPBSA), reverse tetracycline-controlled transactivator (rtTA), self-cleaving peptide P2A (P2A), puromycin resistance (PuroR). **(K)** Confirmation of DPPA3 protein induction as assessed by immunofluorescence in uninduced ESCs (-Dox) or after 24 h of doxycycline treatment (+Dox). To illustrate the increase in DPPA3 protein levels, the same acquisition settings used to detect DPPA3, including a long exposure, in uninduced cells were applied for detection of DPPA3 after induction (+Dox (long exp.); bottom panel) leading to a saturated signal. Shorter exposure settings were also applied to the induced cells (+Dox (short exp.); middle panel) to better resolve the localization of DPPA3 after induction. Scale bar: 10 μ m. **(L)** *Dppa3* expression before induction and after 3 or 6 days of doxycycline treatment in wt J1 + pSBtet-D3 ESCs. mRNA levels of *Dppa3* are shown relative to those in uninduced wtJ1 ESCs after normalization to *Gapdh*. Error bars indicate mean \pm SD calculated from technical triplicate reactions from n=2 biological replicates. In **(D-H)** hypermethylation is a gain in 5mC compared to wt ESCs (q value < 0.05; absolute methylation difference > 20%).

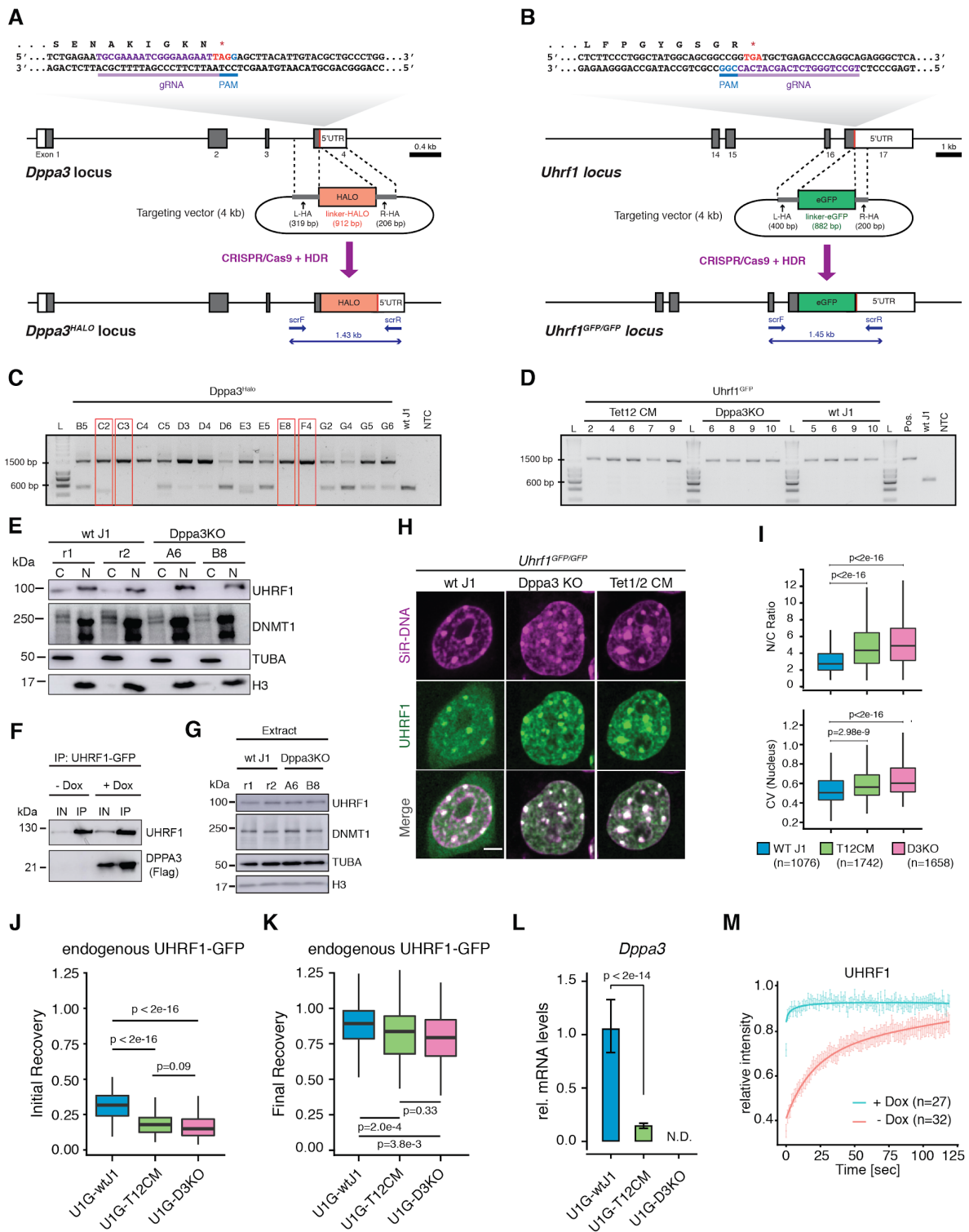


Figure S4: DPPA3 hinders UHRF1 chromatin association in mouse ESCs

(**A and B**) Schematic of the CRISPR/Cas9 targeting strategy used to insert fluorescent protein tags at the (**A**) *Dppa3* and (**B**) *Uhrf1* loci. Exons are depicted as boxes, with the coding sequences shaded in grey. Stop codons are indicated with a red line. The nucleotide and amino acid sequence encompassing the gRNA target site are enlarged for detail. gRNAs target sequences are colored purple and their respective PAMs blue. Donor constructs harboring the coding sequence for (**A**) HALO or (**B**) eGFP are flanked by homology arms (L-HA, left homology arm and R-HA, right homology arm) and indicated as grey rectangles. Genotyping PCR primers used in (**C and D**) are shown as arrows along with the size of the amplicon confirming successful integration of the donor sequence. (**C and D**) Results of genotyping PCRs for confirming the successful integration of (**C**) HALO into the *Dppa3* locus of wt J1 ESCs and (**D**) eGFP into the *Uhrf1* locus of wt J1, *Dppa3*KO, and T12CM ESCs.

Wild-type (wt) and negative control without DNA template (NTC) are depicted on the right. **(C)** Clones with Sanger-sequence validated homozygous insertion of the HALO coding sequence are indicated with red boxes. **(D)** PCR results after pre-screening via microscopy. All depicted clones have a correct homozygous insertion of eGFP as validated by Sanger-sequencing. PCRs were performed using the primers depicted in **(A and B)**. **(E)** DPPA3 regulates subcellular UHRF1 distribution. Cytoplasmic, “C”, and nuclear, “N”, fractions from wt and *Dppa3*KO ESCs analyzed by immunoblot detection using the indicated antibodies. **(F)** Immunoblot detection of anti-GFP immunoprecipitated material (IP) and the corresponding input (IN) extracts from U1G/wt + pSBtet-D3 ESCs treated before (-Dox) and after (+Dox) 24 h induction of *Dppa3* with doxycycline. **(G)** UHRF1 and DNMT1 protein levels are unaffected by DPPA3 loss. Whole-cell extracts from wt and *Dppa3*KO ESCs analyzed by immunoblot detection using the indicated antibodies. **(H and I)** DPPA3 alters the localization and nuclear patterning of endogenous UHRF1-GFP. **(H)** Representative confocal images of UHRF1-GFP in live U1G-wt, U1G-T12CM, U1G-*Dppa3*KO ESCs with DNA counterstain (SiR-Hoechst). Scale bar: 5 μ m. **(I)** Quantification of endogenous UHRF1-GFP (top panel) Nucleus to cytoplasm ratio (N/C Ratio) and (bottom panel) coefficient of variance (CV) within the nucleus of cells (number indicated in the plot) from n=3 biological replicates per genotype. **(J and K)** Endogenous DPPA3 prevents excessive UHRF1 chromatin binding in ESCs. Further analysis of the single cell FRAP data presented in Figure 4C. showing the **(J)** initial and **(K)** final relative recovery of endogenous UHRF1-GFP. Intensity measurements 1 s and 60 s after photobleaching were used for calculating **(J)** initial recovery and **(K)** final recovery, respectively. **(L)** *Dppa3* is significantly downregulated in U1G-T12CM ESCs. *Dppa3* transcript levels in U1G-wt, U1G-T12CM, U1G-*Dppa3*KO ESCs are depicted relative to mRNA levels in U1G-wt1 ESCs after normalization to Gapdh. Error bars indicate mean \pm SD calculated from technical triplicate reactions from n=4 biological replicates (independent clones). N.D., expression not detectable. In the boxplots in **(I-K)**, darker horizontal lines within boxes represent median values. The limits of the boxes indicate upper and lower quartiles, and whiskers indicate the 1.5-fold interquartile range. P-values in **(I-L)**; Welch’s two-sided t-test. **(M)** DPPA3 expression abolishes UHRF1 chromatin binding. FRAP analysis of endogenous UHRF1-GFP before (-Dox) and after 48 h of *Dppa3* induction (+Dox) in U1G-wt + pSBtet-D3 ESCs. In **(M)** the mean fluorescence intensity of *n* cells (indicated in the plots) at each timepoint depicted as a shaded dot. Error bars indicate mean \pm SEM. Curves (solid lines) indicate double-exponential functions fitted to the FRAP data.

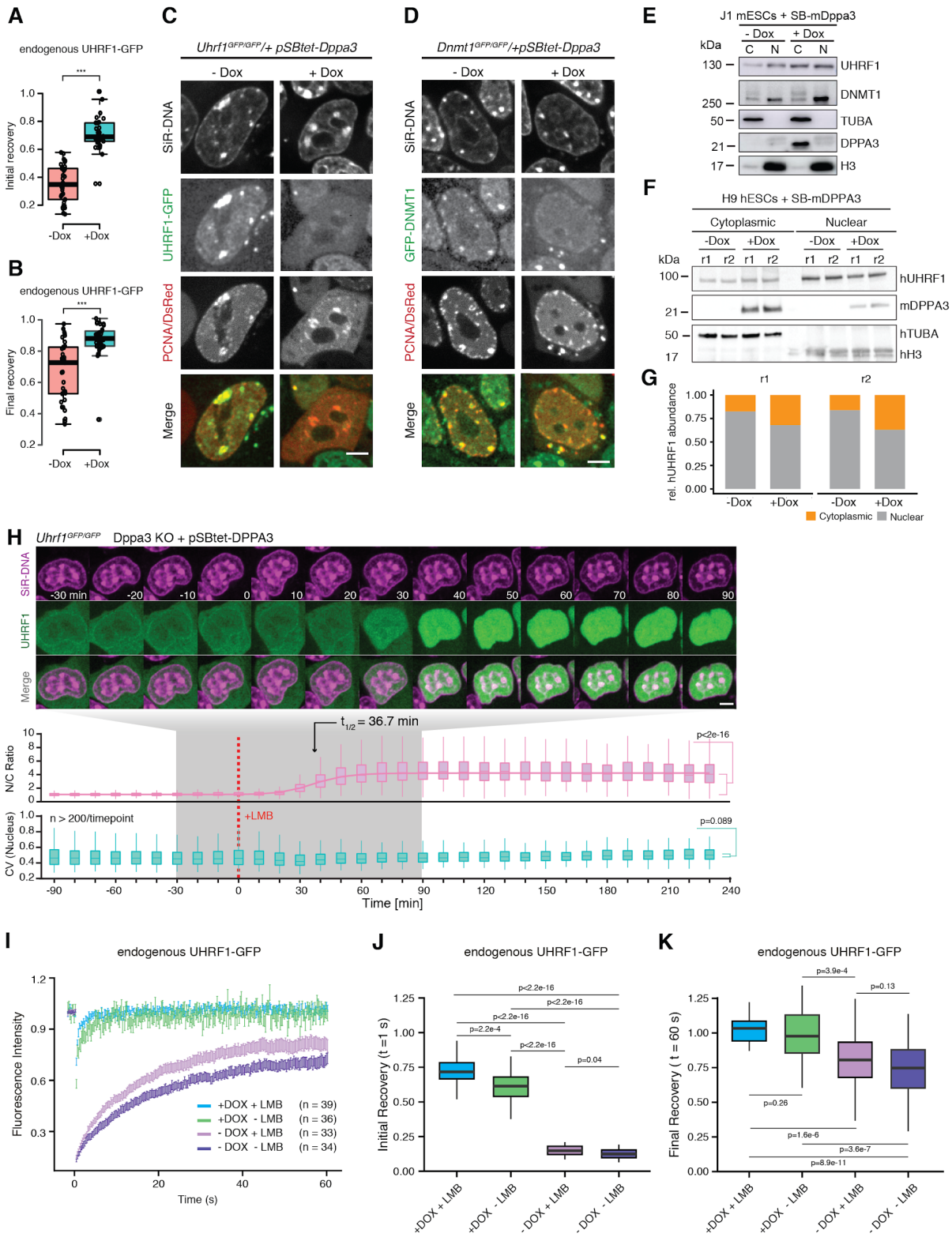


Figure S5: DPPA3 alters the localization and chromatin binding of endogenous UHRF1 and DNMT1 in ESCs

(**A and B**) DPPA3 releases UHRF1 from chromatin. Initial (**A**) and final (**B**) relative recovery of UHRF1-GFP in *Uhrf1^{GFP/GFP}/SBtet-3xFLAG-Dppa3* ESCs before (-Dox) and after 48 h of *Dppa3* induction (+Dox) (+Dox: n=27; -Dox: n=32) (data from FRAP experiment in Figure S4M). (**C and D**) DPPA3 disrupts the recruitment of UHRF1 and DNMT1 to replication foci. Live cell imaging illustrating the late S-phase localization of (**C**) UHRF1-GFP in U1G/wt + pSBtet-D3 ESCs or (**D**) GFP-DNMT1 in live *Dnmt1^{GFP/GFP} + pSBtet-D3* ESCs before (-Dox) and after 48 h of *Dppa3* induction. Transfected RFP-PCNA marks sites of active DNA replication within the nucleus. Expressed free DsRed is a marker of doxycycline induction (see cytoplasmic signal in +Dox PCNA/DsRed panels and Figure S3J). DNA was stained in live cells using a 30

min treatment of SiR-DNA (SiR-Hoechst). Scale bar: 5 μm . **(E-G)**, DPPA3 alters the subcellular distribution of endogenous UHRF1 in mouse ESCs and human ESCs. **(E and F)** Immunoblots of nuclear, “N”, and cytoplasmic, “C”, fractions from **(E)** U1G/wt + pSBtet-D3 ESCs and **(F)** Human H9 ESCs +SBtet-3xFLAG-Dppa3 before (-Dox) and after 24 hours of *Dppa3* induction (+Dox) using the indicated antibodies. An anti-FLAG antibody was used for detection of FLAG-DPPA3. **(G)** Quantification of the relative abundance of hUHRF1 in the cytoplasmic and nuclear fractions shown in **(F)** with the results of two independent biological replicates (r1 and r2) displayed. **(H)** Nuclear export is dispensable for DPPA3-mediated inhibition of UHRF1 chromatin association. Localization dynamics of endogenous UHRF1-GFP in response to inhibition of nuclear export using leptomycin-B (LMB) after *Dppa3* induction in U1G/D3KO + pSBtet-D3 ESCs with confocal time-lapse imaging over 5.5 h (10 min intervals). $t=0$ corresponds to start of nuclear export inhibition (+LMB). *(top panel)* Representative images of UHRF1-GFP and DNA (SiR-Hoechst stain) throughout confocal time-lapse imaging. Scale bar: 5 μm . *(middle panel)* Nucleus to cytoplasm ratio (N/C Ratio) of endogenous UHRF1-GFP signal. *(bottom panel)* Coefficient of variance (CV) of endogenous UHRF1-GFP intensity in the nucleus. *(middle and bottom panel)* N/C Ratio and CV values: measurements in $n>200$ single cells per time point, acquired at $n=20$ separate positions. Curves represent fits of four parameter logistic (4PL) functions to the N/C Ratio (pink line) and CV (green line) data. **(I-K)** Nuclear export is not only dispensable for but attenuates DPPA3-mediated inhibition of UHRF1-GFP chromatin binding. **(I)** FRAP analysis of endogenous UHRF1-GFP within the nucleus of U1G/D3KO + pSBtet-D3 ESCs before (-Dox) and after 48 h of *Dppa3* induction and before (-LMB) and after (+LMB) 3 h of leptomycin-B (LMB) treatment. The mean fluorescence intensity of n cells (indicated in the plots) at each timepoint depicted as a shaded dot. Error bars indicate mean \pm SEM. **(J and K)** Further analysis of the data from **(I)** showing the **(J)** initial and **(K)** final relative recovery of endogenous UHRF1-GFP. For **(A, B, J, and K)** initial and final recoveries were calculated using the intensities measured 1 s or 60 s after photobleaching, respectively. In the boxplots, darker horizontal lines within boxes represent median values. The limits of the boxes indicate upper and lower quartiles, and whiskers indicate the 1.5-fold interquartile range. Welch’s two-sided t-test. *** $P<0.001$; ** $P=0.002$; * $P=0.03$; n.s.: not significant.

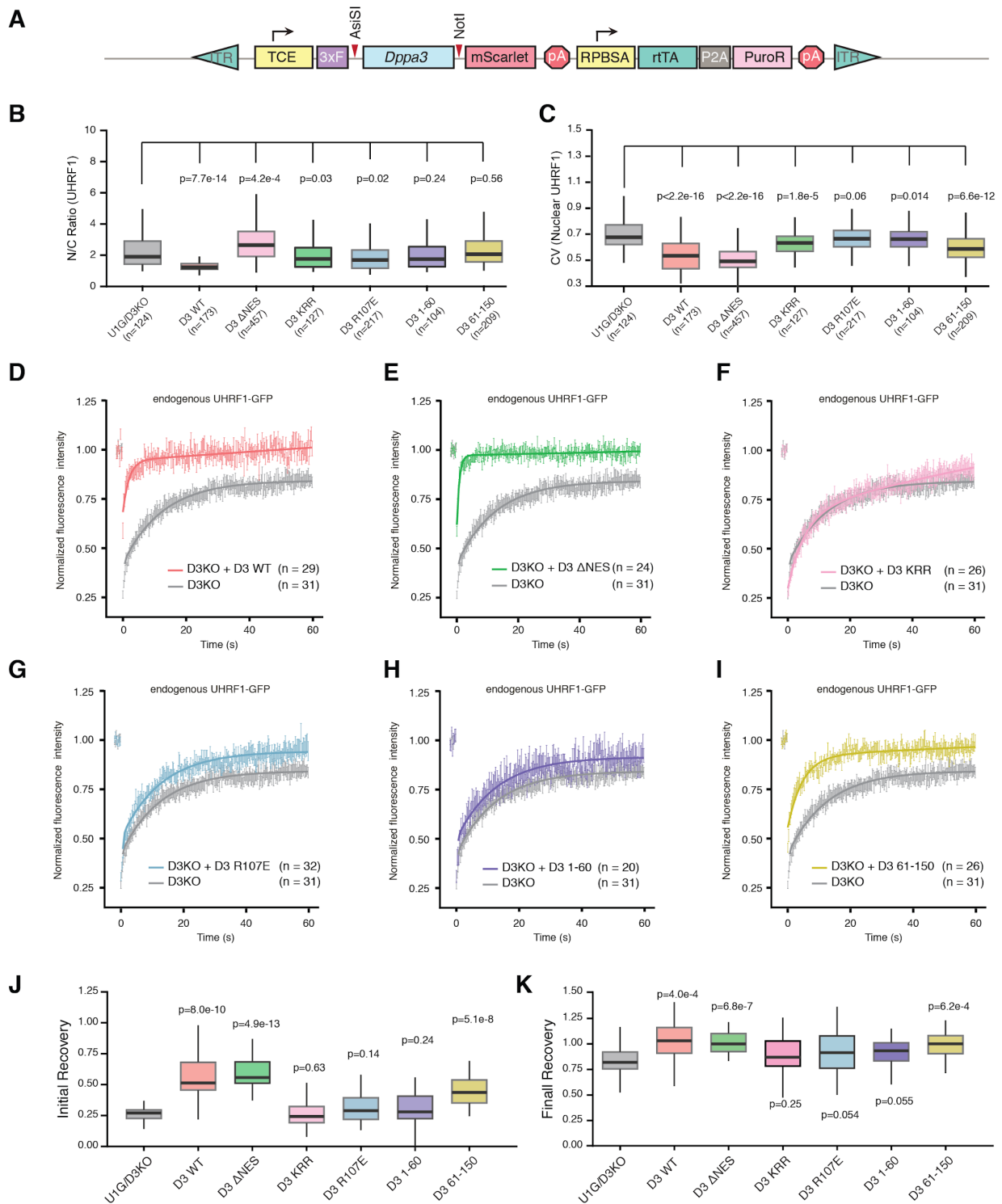


Figure S6: Further analysis of endogenous UHRF1-GFP localization and chromatin binding in response to mutant forms of mDPPA3

(A) Schematic representation of the pSBtet-D3-mSC (pSBtet-3xFLAG-Dppa3-mScarlet) cassette for the Sleeping Beauty transposition-mediated generation of ESC lines with doxycycline (Dox) inducible expression of DPPA3-mScarlet fusions. Abbreviations: inverted terminal repeat (ITR), tetracycline response element plus minimal CMV (TCE), 3xFLAG tag (3xF), internal ribosomal entry site (IRES), polyA signal (pA), constitutive RPBSA promoter (RPBSA), reverse tetracycline-controlled transactivator (rtTA), self-cleaving peptide P2A (P2A), puromycin resistance (PuroR). Restriction sites used for exchanging the Dppa3 coding sequences are indicated with red arrows. (B and C) DPPA3 alters the nuclear localization of UHRF1 independently of promoting UHRF1 nucleocytoplasmic translocation. Quantification of endogenous UHRF1-GFP (B) Nucleus to cytoplasm ratio (N/C Ratio) and (C) coefficient of variance (CV) within the nucleus of U1G/D3KO + pSBtet-

D3 ESCs expressing the indicated mutant forms of Dppa3. n = number of cells analyzed (indicated in the plot). **(D-I)** Nuclear export and the N-terminus of DPPA3 are dispensable for its inhibition of UHRF1 chromatin binding. FRAP analysis of endogenous UHRF1-GFP within the nucleus of U1G/D3KO + pSBtet-D3 ESCs expressing the following forms of DPPA3: **(D)** wild-type (D3 WT), **(E)** nuclear export mutant L44A/L46A(D3- Δ NES), **(F)** K85E/R85E/K87E mutant (D3-KRR), **(G)** R107E mutant (D3-R107E), **(H)** N-terminal 1-60 fragment (D3 1-60), **(I)** C-terminal 61-150 fragment (D3 61-150). In **(D-I)** the mean fluorescence intensity of n cells (indicated in the plots) at each timepoint is depicted as a shaded dot. Error bars indicate mean \pm SEM. Curves (solid lines) indicate double-exponential functions fitted to the FRAP data. Individual same fits correspond to those in Figure 5C. **(J and K)** Further analysis of the data from **(D-I)** showing the **(J)** initial and **(K)** final relative recovery of endogenous UHRF1-GFP. Intensity measurements 1 s and 60 s after photobleaching were used for calculating **(J)** initial recovery and **(K)** final recovery, respectively. In the boxplots in (B,C,J, and K) darker horizontal lines within boxes represent median values. The limits of the boxes indicate upper and lower quartiles, and whiskers indicate the 1.5-fold interquartile range. P-values based on Welch's two-sided t-test of the difference between U1/D3KO and each Dppa3 mutant.

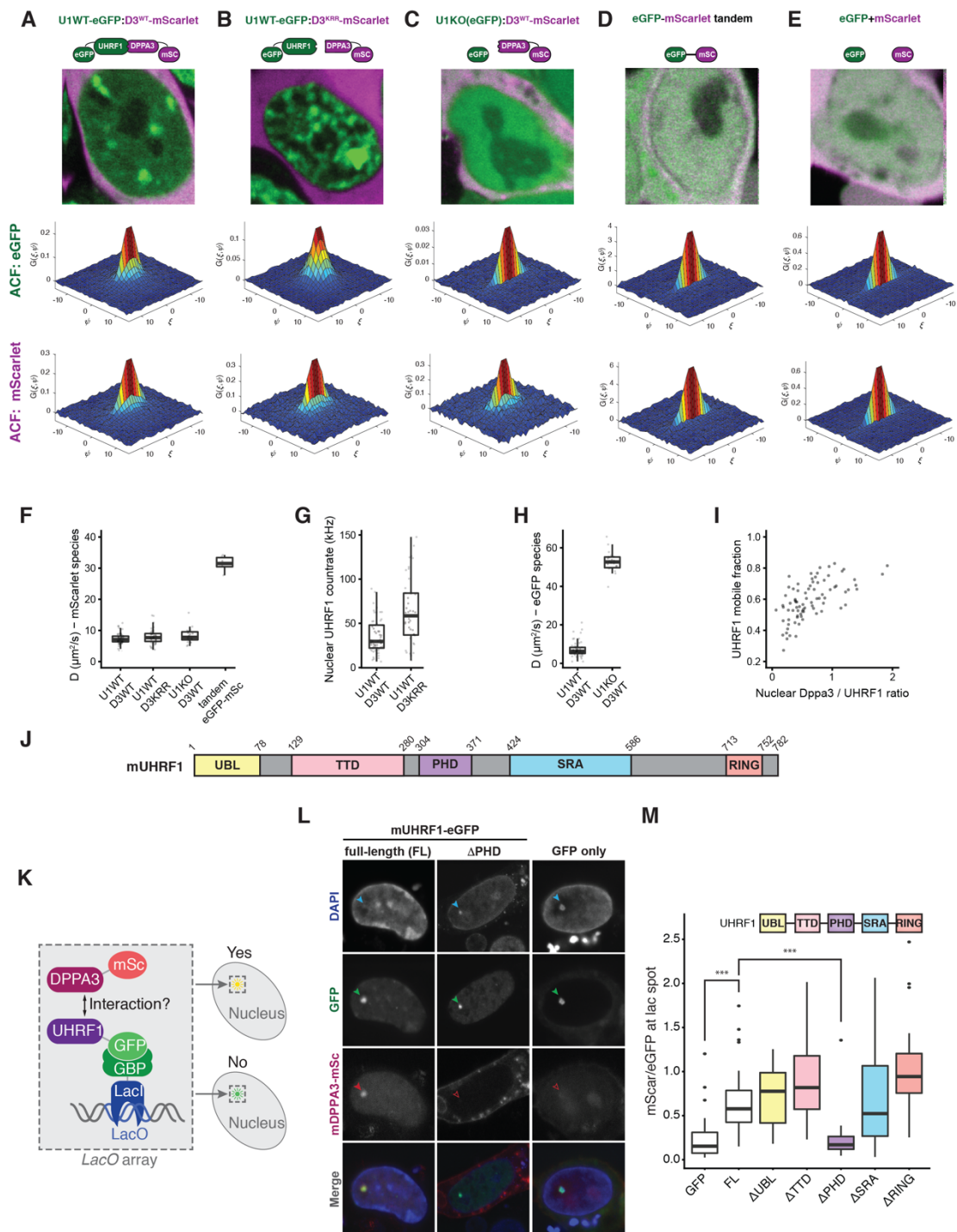


Figure S7: DPPA3 binds UHRF1 via its PHD domain to form a mobile complex in ESCs

(A-E) Representative dual-color live confocal images (top row) and associated 3D plots of mean autocorrelation (ACF) values of eGFP (middle row) and mScarlet (bottom row) species in U1G/D3KO + pSBtet-D3 ESCs expressing the following forms of DPPA3-mScarlet: (A) wild-type (U1WT:D3^{WT}) and (B) K85E/R85E/K87E mutant (U1WT:D3^{KRR}), in (C) Uhrf1KO ESCs expressing free eGFP and wild-type Dppa3-mScarlet (U1KO:D3^{WT}), and in control ESCs expressing (D) an eGFP-mScarlet tandem fusion (eGFP-mScarlet) and (E) free eGFP and free mScarlet (eGFP + mScarlet). In (A-E), each image is the result of merging the sum projections of 250-frame image series acquired simultaneously from eGFP (green) and mScarlet (magenta) channels. Autocorrelation plots are color-coded to indicate mean correlation value. The fast timescale axis is indicated by ξ , and the slow timescale axis is indicated by ψ . (F) Calculated diffusion coefficient of mScarlet species in

different cell types (those shown in **(A-D)**), derived from a two-component fit of mScarlet autocorrelation functions. **(G)** Photon count rates of UHRF1-eGFP in nuclei of U1G/D3KOs expressing either DPPA3-WT (U1WT:D3^{WT}) or DPPA3-KRR mutant (U1WT:D3^{KRR}). **(H)** Calculated diffusion coefficients of UHRF1-GFP (in U1WT:D3^{WT}) or free eGFP (in U1KO:D3^{WT}), derived as those in C. **(I)** Scatter plot showing the relationship between the mobile fraction of UHRF1-eGFP and the ratio of DPPA3-mScarlet:UHRF1-eGFP photon count rates in the nucleus. **(J)** A schematic diagram illustrating the domains of mUHRF1: ubiquitin-like (UBL), tandem tudor (TTD), plant homeodomain (PHD), SET-and-RING-associated (SRA), and really interesting new gene (RING). **(K)** Overview of the F3H assay used to find the domain of UHRF1 which binds to DPPA3. **(L-M)** The PHD domain of UHRF1 is necessary for mediating the interaction with DPPA3. Representative confocal images of free GFP or the Δ PHD UHRF1-GFP constructs **(J)** immobilized at the lacO array (indicated with green arrows). Efficient or failed recruitment of DPPA3-mScarlet to the lacO spot are indicated by solid or unfilled red arrows respectively. Quantification and statistics in **(M)**. **(M)** The efficiency of DPPA3-mScarlet recruitment to different UHRF1-eGFP deletion constructs immobilized at the lacO array is given as the fluorescence intensity ratio of mScarlet (DPPA3) to eGFP (UHRF1 constructs) at the nuclear LacO spot. In **(F-H)**, each data point represents the measured and fit values from a single cell where n = number of cells measured (indicated in the plots). In the box plots **(F-H and M)**, darker horizontal lines within boxes represent median values. The limits of the boxes indicate upper and lower quartiles, and whiskers indicate the 1.5-fold interquartile range. **(M)** P-values were calculated using Welch's two-sided t-test: ** $P < 0.01$; *** $P < 0.001$.

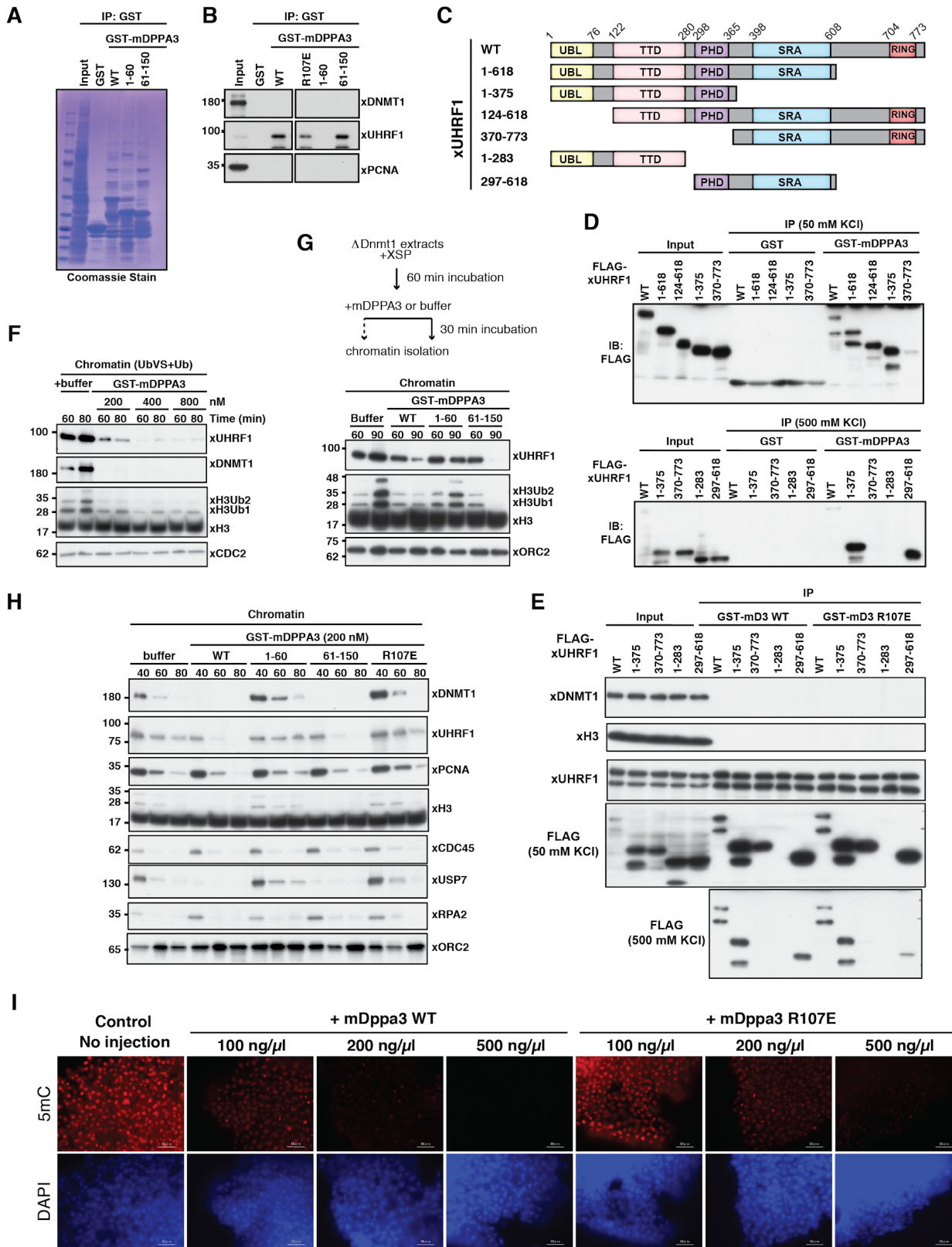


Figure S8: DPPA3 evolved in boreoeutherian mammals but its function is conserved in lower vertebrates.

(A) mDPPA3 C-terminus is sufficient for xUHRF1 binding. GST-tagged mDPPA3 wild-type (WT) and truncations (1-60 and 61-150) were immobilized on GSH beads and incubated with *Xenopus* egg extracts. The samples shown in Figure 7c were subjected to SDS-PAGE and stained with Coomassie Blue. (B) High affinity interaction between mDPPA3 and xUHRF1 is weakened by R107E. GST-tagged mDPPA3 wild-type (WT), the point mutant R107E, and truncations (1-60 and 61-150) were immobilized on GSH beads and incubated with *Xenopus* egg extracts. Pull-downs were subjected to stringent 500 mM

KCl washing. Bound proteins were analyzed using the indicated antibodies. **(C)** Schematic diagram illustrating the xUHRF1 deletion constructs used in the pull-downs in **(D and E)**. **(D and E)** xUHRF1 binds mDPPA3 via its PHD domain. In vitro translated xUHRF1 fragments were added to interphase *Xenopus* egg extracts. **(D)** GST and GST-mDPPA3 wild-type (WT) or **(E)** GST-tagged mDPPA3 wild-type (WT) and GST-tagged mDPPA3 R107E were immobilized on GSH beads and then used for GST-pulldowns on egg extracts containing the indicated recombinant xUHRF1 fragments. Bound proteins were analyzed using the denoted antibodies. Pull-downs were subjected to either 50 mM KCl or more stringent 500 mM KCl washing as indicated. **(F)** mDPPA3 disrupts xUHRF1-dependent ubiquitylation of H3. As dual-monoubiquitylation of H3 (H3Ub2) is hard to detect given its quick turnover (Yamaguchi et al., 2017), we specifically enhanced ubiquitylation by simultaneous treatment of extracts with ubiquitin vinyl sulfone (UbVS), a pan-deubiquitylation enzyme inhibitor (Borodovsky et al., 2001) and free ubiquitin (+Ub) as described previously (Yamaguchi et al., 2017). Buffer or the displayed concentration of recombinant mDPPA3 were then added to the extracts. After the indicated times of incubation, chromatin fractions were isolated and subjected to immunoblotting using the antibodies indicated. **(G)** mDPPA3 displaces chromatin-bound xUHRF1. To stimulate xUHRF1 accumulation on hemi-methylated chromatin, *Xenopus* extracts were first immuno-depleted of xDNMT1 as described previously (Nishiyama et al., 2013). After addition of sperm chromatin, extracts were incubated for 60 min to allow the accumulation of xUHRF1 on chromatin during S-phase and then the indicated form of recombinant mDPPA3 (or buffer) was added. Chromatin fractions were isolated either immediately (60 min) or after an additional 30 min incubation (90 min) and subjected to immunoblotting using the antibodies indicated. **(H)** The region 61-150 of mDPPA3 is sufficient but requires R107 to inhibit xUHRF1 chromatin binding. Sperm chromatin was incubated with interphase *Xenopus* egg extracts supplemented with buffer (+buffer) and GST-mDPPA3 wild-type (WT), the point mutants R107E, truncations (indicated). Chromatin fractions were isolated after the indicated incubation time and subjected to immunoblotting using the antibodies indicated. **(I)** mDPPA3-mediated DNA demethylation in medaka. Representative 5mC immunostainings in late blastula stage (~8 h after fertilization) medaka embryos injected with wild-type mDppa3 (WT) or mDppa3 R107E (R107E) at three different concentrations (100 ng/ μ l, 200 ng/ μ l, or 500 ng/ μ l). Scale bars represent 50 μ m. DNA counterstain: DAPI, 4',6-diamidino-2-phenylindole.

ADDITIONAL SUPPLEMENTARY FILES

File Name: Supplementary_Table_S1

Description: Tet_specific_and_common_promoters

File Name: Supplementary_Table_S2

Description: Gene ontology (GO)_Tet_specific

File Name: Supplementary_Table_S3

Description: DPPA3_interactors

File Name: Supplementary_Table_S4

Description: GO analysis_DPPA3_interactome

File Name: Supplementary_Table_S5

Description: Oligonucleotides

2.4 Metabolic regulation of TET enzymes in mESCs

– manuscript in preparation –

Metabolic regulation of TET enzymes in mESCs

Carina Trummer¹, Udo Müller¹, Martin Lehmann², Laura Kraus¹, Sebastian Bultmann¹, Martha Smets¹, Peter Geigenberger², Heinrich Leonhardt^{1*}

¹ Human Biology and BioImaging, Center for Integrated Protein Science (CIPS), Department of Biology II, Ludwig-Maximilians-Universität München, Germany

² Plant Sciences, Department of Biology I, Ludwig-Maximilians-Universität München, Germany

*Corresponding Author, contact: h.leonhardt@lmu.de, phone: +49 (0)89 / 2180 - 74 232

Abstract

Epigenetic mechanisms play a pivotal role in regulating gene expression during embryonic development and differentiation. To add and remove epigenetic modifications, a vast number of enzymes are involved with many of those enzymes consuming metabolites as co-factors for their activity, thereby directly linking metabolism to epigenetic regulation. Perturbation of these metabolic reactions and their underlying epigenetic pathways were shown to not only influence embryonic/early development but also mediate tumorigenesis. Hence, modulating the activity of epigenetic enzymes solely by the availability of corresponding metabolic co-factors and/or inhibitors represents an interesting approach for tuning epigenetic activity.

Here, we focused on the feasibility to modulate the activity of ten-eleven-translocation (TET) proteins – enzymes involved in locus-specific removal of DNA methylation. We treated mouse embryonic stem cells (mESCs) with esterified metabolites and monitored their uptake kinetics and cytosolic conversion by GC-MS. We show that the activity of TET proteins is strongly influenced by both, its natural co-factor alpha-ketoglutarate (α -KG) and the oncometabolite 2-hydroxyglutarate (2-HG), which is produced by isocitrate dehydrogenase (IDH) mutants in several cancers, e.g. in AML and gliomas. RNAseq analysis additionally revealed that exogenously applied α -KG leads to the downregulation of pluripotency genes, therefore illustrating a differentiation-promoting role for α -KG. Furthermore, inducible wildtype and mutant IDH mESC cell lines were generated to stably overexpress the respective IDH variant in culture. This allows both, to further study the effects of α -KG and 2-HG that are successfully produced upon induction and to investigate the cellular impact of mutant IDH proteins besides the production of 2-HG, e.g. mutant IDH-induced reduction of NADPH levels.

Taken together, our results suggest that metabolic co-factors can act as rate-limiting components in epigenetic reactions, which is medically relevant since most tumor cells harboring TET mutations retain one intact allele. Thus, methods to boost the activity of the remaining TET enzyme could possibly lead to novel therapeutic approaches restoring proper epigenetic regulation.

Introduction

Epigenetic regulation is based on covalent modifications on chromatin and modulates gene expression within every cell, especially during early embryonic development (Bird, 2002; Strahl and Allis, 2000). A vast number of enzymes catalyze the addition and removal of these epigenetic modifications and many of the proteins require metabolites as co-factors for their activity (Janke et al., 2015). DNA-methylation patterns for example are established by DNA methyltransferases, which depend on S-adenosyl methionine (SAM) as a methyl-donor for their reactions (Jeltsch and Jurkowska, 2016). The removal of these methylation marks is implemented by so-called ten-eleven-translocation (TET) family proteins (Kohli and Zhang, 2013) - enzymes that need oxygen, iron and alpha-ketoglutarate (α -KG) for their activity (fig. 1b) (Loenarz and Schofield, 2011). To generate unmodified cytosine, TET proteins successively oxidize methylated cytosine (5mC) to hydroxymethyl cytosine (5hmC), formyl cytosine (5fC) and carboxyl cytosine (5caC) (Ito et al., 2011; Tahiliani et al., 2009). In a second step, the two latter ones can be excised by the thymine DNA glycosylase (TDG) and unmodified C is subsequently incorporated by base excision repair (BER) enzymes (Weber et al., 2016). Another layer of epigenetic regulation is the modification of histone cores and tails ranging from acetylation, methylation and phosphorylation to ubiquitination (Lawrence et al., 2016) that requires a whole assortment of different metabolites (Suganuma and Workman, 2018). In fact, there is an apparent link between metabolism and epigenetics and recent studies begun to understand how metabolic reactions coordinate epigenetic reactions and that alterations in these pathways contribute to impaired differentiation of cells and organs and the development of various diseases (Tzika et al., 2018).

For instance, embryonic stem cells (ESCs) undergo a complete metabolic switch from a bivalent to a more glycolytic state in their naïve-to-primed transition (Zhou et al., 2012), accompanied by the remodeling of the whole epigenetic landscape (Sperber et al., 2015). Vitamin C in turn was shown to promote an ICM-like DNA methylation state in ESCs (Blaschke et al., 2013), thereby preventing such a transition, whereas the amino acid L-proline induces differentiation of ESCs (Washington et al., 2010). The TCA-cycle intermediate α -KG was reported to control pluripotency of ESCs in a stage-dependent manner: While Carey et al. (2015) claim that intracellular α -KG under naive culturing conditions maintains the pluripotency of ESCs by demethylation of repressive chromatin marks, Teslaa et al. (2016) argue that α -KG promotes early differentiation of Epiblast-like stem cells (EpiSCs). All together these findings demonstrate how diversely metabolism can impact the epigenetic regulation of embryonic development.

In cancer, metabolic reprogramming influences cellular function and appears to mediate tumorigenesis (Tzika et al., 2018). Several mutations in metabolic enzymes like succinate dehydrogenase (SDH), fumarate hydratase (FH) and isocitrate dehydrogenase (IDH) have been found (King et al., 2006; Parsons et al., 2008), which lead to the accumulation of TCA cycle intermediates (Pollard et al., 2003) and epigenetic deregulation (Xiao et al., 2012). In contrast to mutations in FH and SDH, mutant IDH1 and IDH2 not only lose their normal catalytic activity to produce α -KG but gain the neomorphic function of

reducing α -KG to 2-hydroxyglutarate (2-HG) (fig. 1a) (Dang and Su, 2017). Due to its structural similarity to α -KG, 2-HG is capable of binding the active sites of α -KG-dependent dioxygenases acting as a competitive inhibitor (Xu et al., 2011). Among these dioxygenases, epigenetic modifiers such as histone demethylases and TET proteins are found (fig. 1b), whose inhibition by 2-HG results in histone and DNA hypermethylation *in vitro* and *in vivo* (Figueroa et al., 2010; Lu et al., 2012; Turcan et al., 2012; Xu et al., 2011).

Hence, modulating the activity of epigenetic enzymes solely by the availability of corresponding metabolic co-factors and/or inhibitors provides the opportunity to gain more detailed knowledge about the regulation of epigenetic networks. In this study, we focused on modifying the activity of TET proteins in mESCs. This represents not only an intriguing approach to address basic epigenetic questions of DNA demethylation processes, but is particularly relevant in the medical context of mutant TET2 tumors. Since TET2 mutations are predominantly heterozygous, one intact TET2 allele is retained within cancer cells (Gaidzik et al., 2012). Boosting the activity of TET2 enzymes from the remaining allele might counterbalance the reduced protein levels and thus, could possibly restore proper epigenetic regulation.

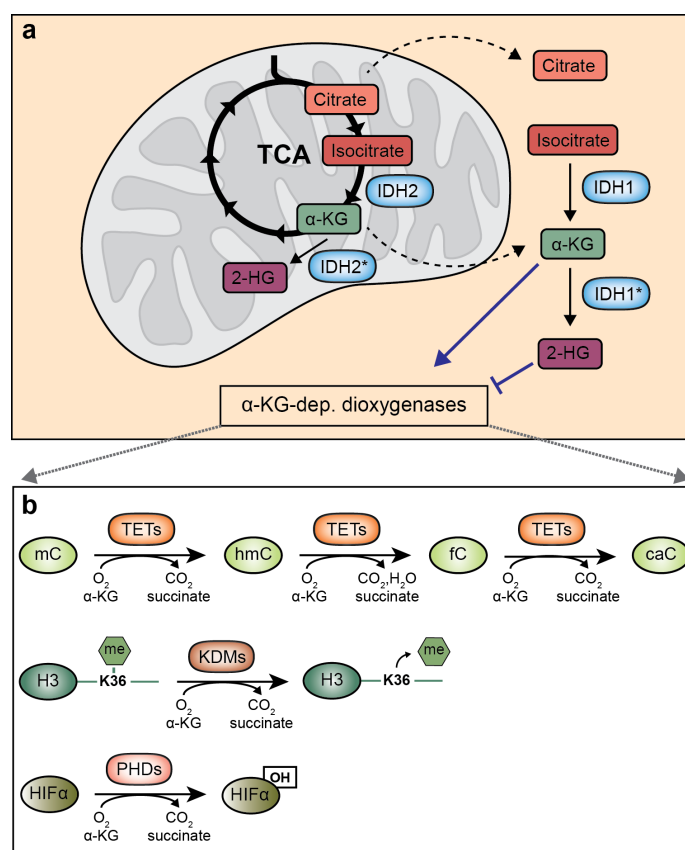


Figure 1: Metabolic regulation of α -KG-dependent dioxygenases. (a) As part of the TCA cycle, wildtype IDH produces α -KG within the cytosol (IDH1) or the mitochondria (IDH2). Mutated IDH (marked with asterisk) generates 2-HG. Both metabolites, α -KG and 2-HG, bind to and regulate the activity of α -KG-dependent dioxygenases. (b) Enzymatic reaction of three exemplary α -KG-dependent dioxygenase families: Demethylation of cytosines via TET proteins, demethylation of histones (H3) via histone demethylases and hydroxylation of HIF α via prolyl hydroxylases. TCA: tricarboxylic acid, IDH: isocitrate dehydrogenase, α -KG: alpha-ketoglutarate, 2-HG: 2-hydroxyglutarate, KDMs: lysine demethylases, PHDs: prolyl hydroxylase domain proteins, HIF α : hypoxia-inducible factor alpha

Results and Discussion

To test whether TET proteins in mESCs respond to exogenously-applied metabolites, we treated cells with ester derivatives of 2-HG and α -KG; diethyl-2-HG (DE-2-HG) and dimethyl- α -KG (DM- α -KG), respectively. While 2-HG and α -KG in their native forms are too hydrophilic to pass the cell membrane, the more lipophilic esters are able to enter the cell (fig. 2a). In the cytosol the ester groups are removed by hydrolysis releasing the active metabolite similar to prodrug-systems (Yang et al., 2011). To verify this mechanism and to determine the kinetics behind the enzymatic conversion, mESCs were cultured in the presence of DE-2HG for different time periods and intracellular metabolite levels were measured by GC-MS (fig. 2b+c). As depicted in figure 2b, DE-2-HG levels oscillated around zero indicating that the ester is immediately hydrolyzed when entering the cell. The non-esterified 2-HG, however, accumulated over time illustrating that the uptake and conversion of DE-2-HG exceeds the degradation of the active metabolite (2-HG) within the cell. This resembles the situation observed in IDH mutant tumors where the production of the oncometabolite 2-HG exceeds its removal, thereby leading to 2-HG levels of up to 30 mM (Dang et al., 2009; Gross et al., 2010). In contrast, if fresh medium was applied after 6h to eliminate DE-2HG and the cells were cultured for another 6h, only minor traces of 2-HG were detected, demonstrating that 2-HG is in fact degradable (dotted line in fig. 2b). This is in line with studies of Achouri et al. (2004) demonstrating that so called 2-HG-dehydrogenases (2HGDHs) are able to metabolize 2-HG, which is infrequently generated as an unwanted by-product during normal metabolic reactions (Rzem et al., 2007; Struys et al., 2005). In this process, 2-HG is converted into α -KG (Achouri et al., 2004) raising the question whether the observed accumulation of 2-HG ultimately causes a concomitant increase in α -KG. In fact, α -KG levels did increase over time, albeit only slightly (fig. 2c), suggesting that 2-HG gets at least partly converted into α -KG.

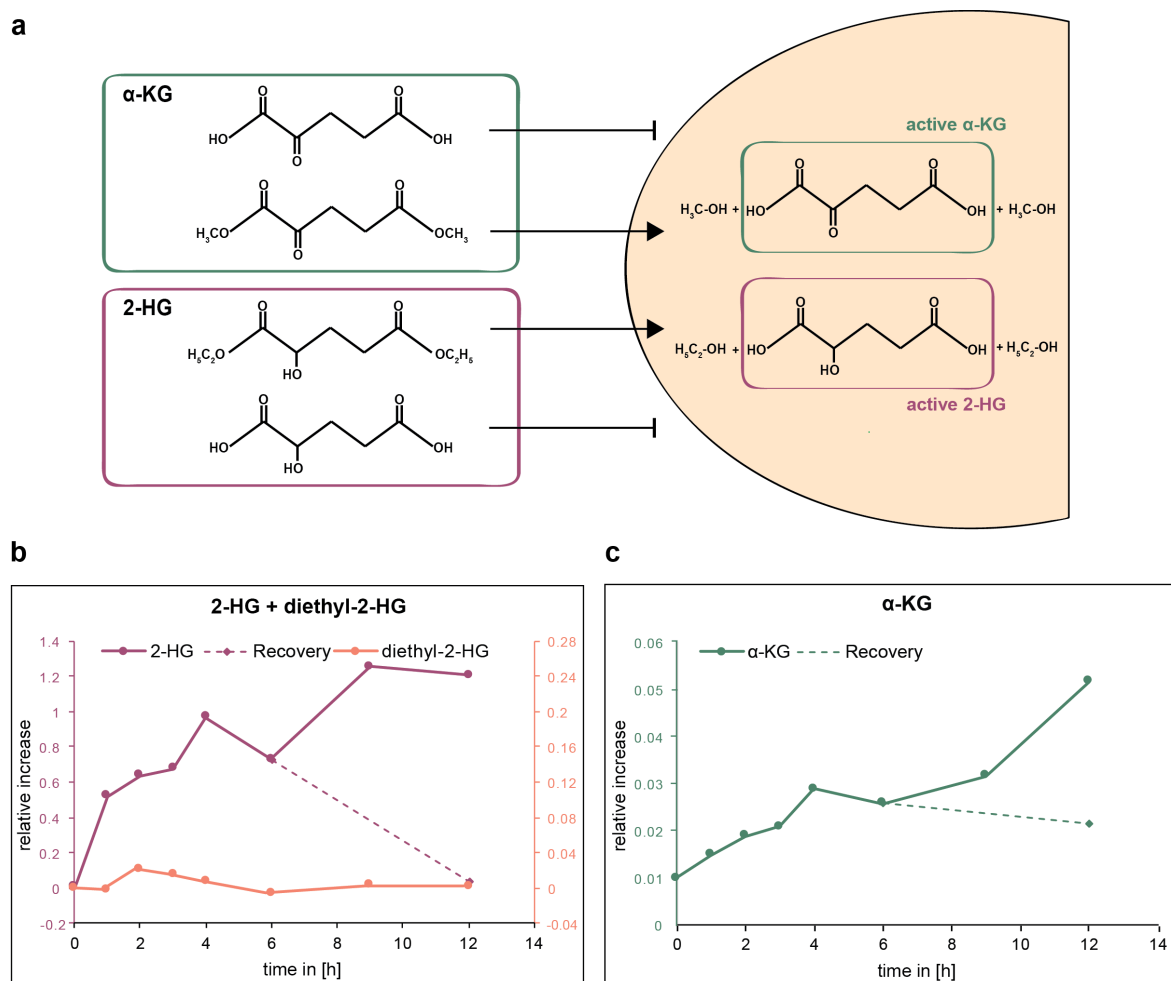


Figure 2: Conversion of metabolic ester derivatives in mESCs. (a) The active, native forms of α -KG and 2-HG do not pass the hydrophobic cell membrane. Esterification of α -KG and 2-HG, in form of dimethylation and diethylation, respectively, allows the metabolites to penetrate the cell membrane. (b+c) Accumulation of 2-HG and α -KG in mESCs after exogenously adding diethyl-2HG over 12 hours. GC-MS/MS detected intracellular levels of 2-HG and diethyl-2HG (b) and α -KG (c). For the recovery sample, diethyl-2HG-containing medium was exchanged to normal medium after 6 h and cells were further cultured for additional 6 h.

After successfully verifying the conversion of ester derivatives in mESCs, we sought to test the responsiveness of α -KG-dependent dioxygenases to exogenously applied metabolite esters. As TET proteins belong to this family and hmC, fC and caC are specifically catalyzed by TET enzymes (Ito et al., 2011; Tahiliani et al., 2009) (fig. 1b), these modifications represent a direct readout of TET activity and allow to determine the effect of exogenously applied metabolites. Accordingly, we performed immunostainings against hmC, fC and caC in combination with high-throughput image analysis quantification. Both esterified metabolites, DM- α -KG and DE-2HG, impacted TET-mediated modifications in a concentration-dependent manner compared to non-esterified controls (fig. 3a+b). α -KG as the active co-factor of TET proteins led to a significant increase of the oxidized cytosine variants in the μ M range, with an 2.5-fold increase for hmC and an 1.5-fold increase for fC and caC at 500 μ M α -KG (fig. 3a). 1- 10 mM 2-HG caused opposing effects and reduced the levels of the respective modifications (fig. 3b), as previously demonstrated for hmC in other cell types (Chaturvedi et al., 2016;

Losman et al., 2013; Xu et al., 2011). Co-administration of α -KG and 2-HG rescued the inhibiting effect of 2-HG on TET proteins albeit α -KG's lower concentration (1 mM α -KG vs. 5 mM 2-HG, fig. 3c). This suggests that 2-HG only acts as a weak antagonist of α -KG, probably because of its inferior binding to the catalytic core of α -KG-dependent dioxygenases as previously proposed by Xu et al. (2011). Taken together, these results indicate that the enzymatic activity of TET proteins in mESCs is adjustable solely upon administration of the respective metabolites. To elucidate whether the metabolically-induced alterations of TET activity impact gene expression in mESCs, we aimed for an RNA-seq analysis. Due to technical reasons (degradation of the chemical DE-2-HG), the expressional analysis of the RNA-seq library was only feasible for the samples treated with DM- α -KG. As depicted in figure 3d, deregulated genes were grouped into four different clusters depending on their expressional change over the treatment time of 4h and 8h. Numerous genes were repressed upon α -KG administration with the most prominent reduction of genes in cluster 3. Interestingly, cluster 3 comprises many pluripotency markers, indicating that α -KG levels may contribute to the pluripotent state of mESCs. It is important to note here that cells were cultured in LIF-only medium since culturing conditions and related differentiation stages of cells seem to be decisive for the cellular effect of α -KG (Carey et al., 2015; TeSlaa et al., 2016). Whereas α -KG maintains pluripotency in the naive state (Carey et al., 2015), it promotes differentiation of Epiblast-like stem cells (EpiSCs) (TeSlaa et al., 2016). Our results would equally argue for a differentiation-promoting role of α -KG as pluripotency markers are downregulated. However, the assignment to an epiblast-like stage would be incorrect as LIF-only cultured cells constitute a mixture of naive and primed ESCs which in turn prevents the direct comparison to the aforementioned studies. Nonetheless, α -KG shows an opposing effect to what is generally reported for 2-HG, namely the prevention of cellular differentiation of precursor cells like human hematopoietic stem cells (HSCs) (Ye et al., 2018). Keeping in mind that human ESCs are more similar to the epiblast than the inner cell mass (ICM) of mice (Harvey et al., 2016), comparing the effect of α -KG and 2-HG in mouse EpiSCs and human ESCs is reasonable. Secondly, assuming that LIF cells represent a differentiated stage, although less pronounced than EpiLCs, our result and other studies suggest that α -KG and 2-HG affect the differentiation of stem cells in mice and human in an antagonistic way (Losman et al., 2013; Lu et al., 2012; Suijker et al., 2015). Importantly, the effect of α -KG that we observed on pluripotency marker gene expression is not necessarily based on the altered activity of TET enzymes, albeit α -KG clearly increased the levels of their catalytic products (fig. 3a). In fact, 2-HG and α -KG affect over 60 different α -KG-dependent dioxygenases in humans including histone demethylases (fig. 1b) (Loenarz and Schofield, 2011). Hence, it is difficult to define which of the dioxygenases are effectively involved in the observed gene expression changes after exogenous administration of DM- α -KG.

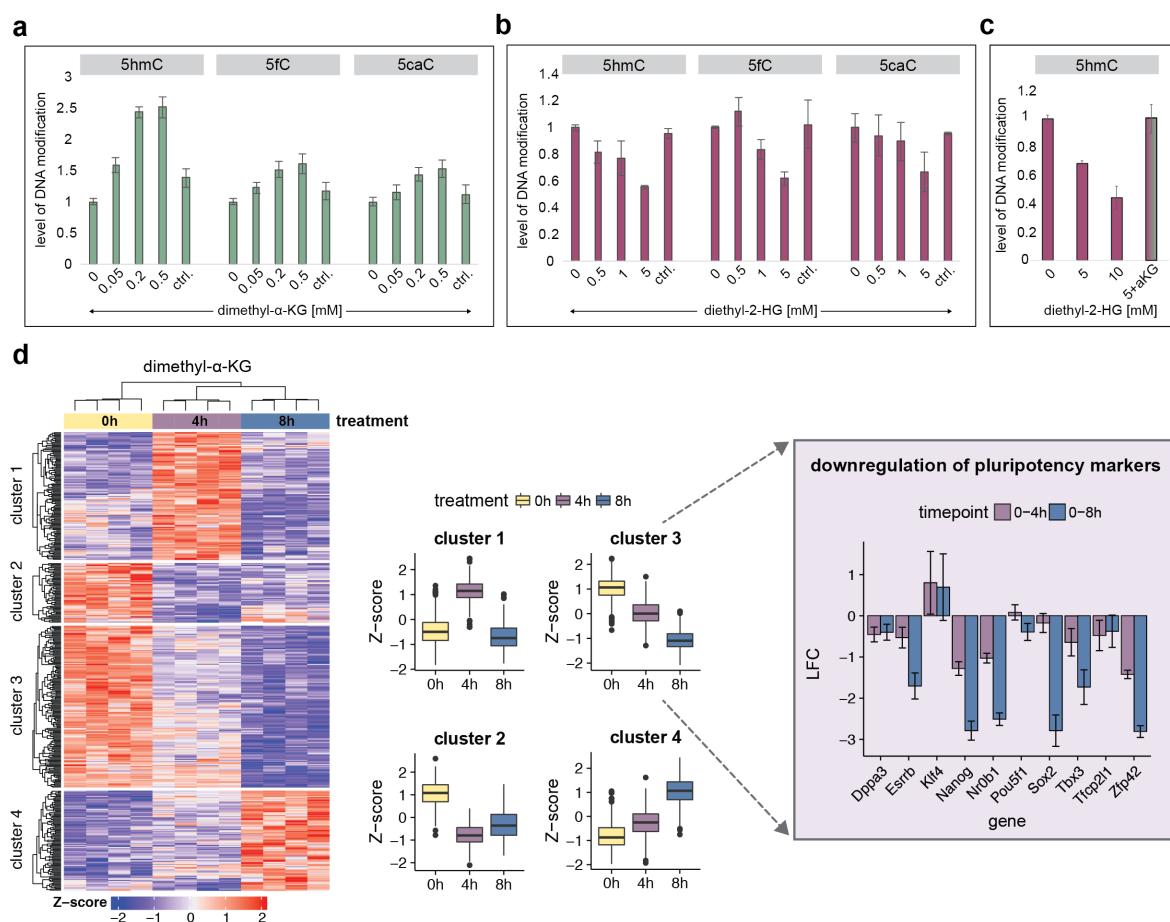


Figure 3: α -KG and 2-HG modulate TET activity in mESCs and influence pluripotency. (a+b) Quantification of IF-stained nuclear 5hmC, 5fC and 5caC levels after the administration of increasing concentrations of dimethyl- α -KG (a) and diethyl-2-HG (b), ctrl: respective non-esterified metabolite. (c) Rescue effect of α -KG towards 5hmC levels in 2-HG treated mESCs. (d) RNA-seq analysis of mESCs cultured with dimethyl- α -KG for 0h, 4h or 8h. Genes were grouped into four clusters depending on their temporal expression patterns. The bar plot on the right displays the expressional deregulation of pluripotency genes corresponding to cluster 3.

To reproduce the initial effects observed on DNA modifications and pluripotency after the treatment with esterified metabolites, we generated mESC cell lines harbouring doxycycline (DOX)-inducible IDH cassettes: wildtype IDH1, mutant IDH1 R132H, wildtype IDH2, mutant IDH2 R140Q and mutant IDH2 R172K (fig. S1a). Notably, this enables two things; further investigating the effect of the metabolites and simultaneously elucidating the cellular impact of mutant IDH proteins beyond the production of 2-HG, e.g. the reduction of NADPH levels (Dang and Su, 2017). After validating the induced expression of IDH through DOX addition in the respective cell lines (fig. S1 b-d), we checked hmC levels upon IDH overexpression. However, neither IF stainings within the nucleus nor quantification of hmC on genomic DNA produced meaningful results (fig. S1 e-f). We therefore re-checked the efficiency of the DOX induction and concluded that the low percentage of positively-induced cells within the batch/pool of cells (cp. “batch” column in fig. S2a) is not sufficient and potential effects might be masked by uninduced cells. To overcome this, we individually picked induced (meaning dsRed-positive) clones from the batch, let them grow into single colonies and FACS-sorted these cells after induction, which greatly enhanced

the fraction of dsRed-positive cells (fig. S2 a+b). Next, we screened all clones with a dsRed-positive fraction >50% for successful expression of the respective IDH variant on mRNA and protein level (fig. S2 c+d). Following the validation of IDH inducibility, we performed an IF-staining against 5hmC in wildtype and mutant IDH clones, which have been induced for 48h (fig. 4a). Albeit the strong induction with 3 μ g/ml DOX, consistent effects were only observed for IDH1wt and IDH1 R132H clones, but not for clones of the IDH2 variants (fig. 4a). This could be explained by uneven integration efficiencies of the sleeping beauty transposase in different IDH variants ultimately leading to different IDH levels after induction. However, strong expressional discrepancies between the group of IDH1 and the group of IDH2 clones was not observed (fig. S2 c+d). It is therefore more likely that the experimental results are based on functional disparities of IDH1 and IDH2 that naturally occur in mESCs. In fact, mitochondria in ESCs are poorly developed with spherical, cristae-few mitochondria in human ESCs (Bukowiecki et al., 2014; Lees et al., 2017). As IDH2 but not IDH1 localizes to mitochondria (Dang and Su, 2017), limited mitochondrial capacity might impede any effects of overexpressed IDH2 proteins. On the contrary, the function of IDH1 might be independent from the morphology of mitochondria due to its cytoplasmic localization (Dang and Su, 2017). Hence, we confined further experiments on the analysis of IDH1 clones, i.e. IDH1wt and IDH1 R132H.

To confirm that the observed changes in 5hmC after IDH1wt and IDH1 R132H induction (fig. 4a) are in fact due to altered α -KG and 2-HG levels, we quantified the respective metabolite amounts using LC-MS/MS. Whereas the induction of IDH1wt clones yielded diminutive changes in α -KG, induction of IDH1 R132H clones led to a massive increase of 2-HG (fig. 4b). α -KG levels in IDH1wt clones raised about 25% on average. 2-HG levels in mutant IDH1 clones #9 and #12 however escalated with a 13- and 29-fold increase of the oncometabolite, respectively (fig. 4b+c). Levels of 2-HG in wildtype and α -KG in mutant clones did not change substantially as expected (fig. 4b). Next, we sought to re-investigate the effect of α -KG and 2-HG on mESC pluripotency. To this end, mRNA expression of established pluripotency markers (Nanog, Zfp42 and Sox2) was examined in induced IDH1 clones using RT-qPCR (fig. 4d). Based on the RNAseq results of mESCs treated with exogenous DM- α -KG (fig. 3d), we anticipated a decrease of pluripotency marker expression for IDH1wt clones and reversely, an increase for 2-HG producing mutant IDH1 clones. However, overexpressing IDH1 in none of the clones changed the expression of these genes significantly compared to corresponding uninduced samples. As *Idh1* expression itself (measured in parallel) did clearly increase for clones #4, #11 and #12, deficient induction of samples can be excluded (except R132H #9, fig. 4d). Surprisingly, *Zfp42* and *Sox2* expression was markedly enhanced in clone R132H #12 irrespective of DOX administration. This however, could be an artefact of the applied genome engineering technique using random integration through transposases or might be a consequence of slight variations during cell culture handling. Taken together, endogenous production of 2-HG by overexpressing R132H mutants seems to clearly outperform the production of α -KG through overexpressed IDH1wt. This is probably not due to profound differences in the enzymatic activity of the two proteins or the number of randomly-integrated expression cassettes, but most likely

stems from the fact that α -KG is a frequent and naturally-occurring metabolite in mESCs (Harvey et al., 2016). In fact, multiple metabolic pathways are present in mESCs that allow the metabolization of α -KG but not 2-HG, which finally leads to its accumulation (Zdzisińska et al., 2017). This is further supported by a 5hmC staining of cells induced with 1 μ g/ml DOX for 4 or 10 days (data not shown), where IDH1wt expressing clones lost their 5hmC-increasing effects upon prolonged culturing, IDH1 R132H expressing clones however intensified the reduction of 5hmC. Moreover, endogenously-produced metabolites had no impact on pluripotency of mESCs whereas exogenous DM- α -KG repressed pluripotency markers. Notably, overexpressed IDH proteins dependent on the availability of intracellular precursor metabolites and catalyze their reactions based on the energetic state of the cell (Al-Khallaf, 2017). Exogenously administered DM- α -KG in turn is available in excess. Therefore, it should be tested whether endogenously produced metabolites do not reach the intracellular concentration of DM- α -KG after its uptake and whether these endogenous metabolite levels are too low to affect the expressional control of pluripotent genes.

Conclusion

As there is accumulating evidence for a direct linkage between metabolic and epigenetic regulation of cells, we sought to investigate the metabolite-driven tunability of TET proteins in mESCs and demonstrate that α -KG levels constitute a rate-limiting factor of TET activity. Increasing the amount of intracellular α -KG not only increases TET-mediated DNA modification levels, but also reduces the pluripotent features of mESCs. Additionally, mutant IDH produced 2-HG appears to exhibit comparable characteristics in mESCs as reported for the oncometabolite in tumors like AML or gliomas. In conclusion, deciphering metabolic effects on pluripotency of stem cells will provide insights into both, the intricate process of early embryogenesis and the evolution of certain cancer entities that are initiated through metabolically-induced perturbation of epigenetic regulation.

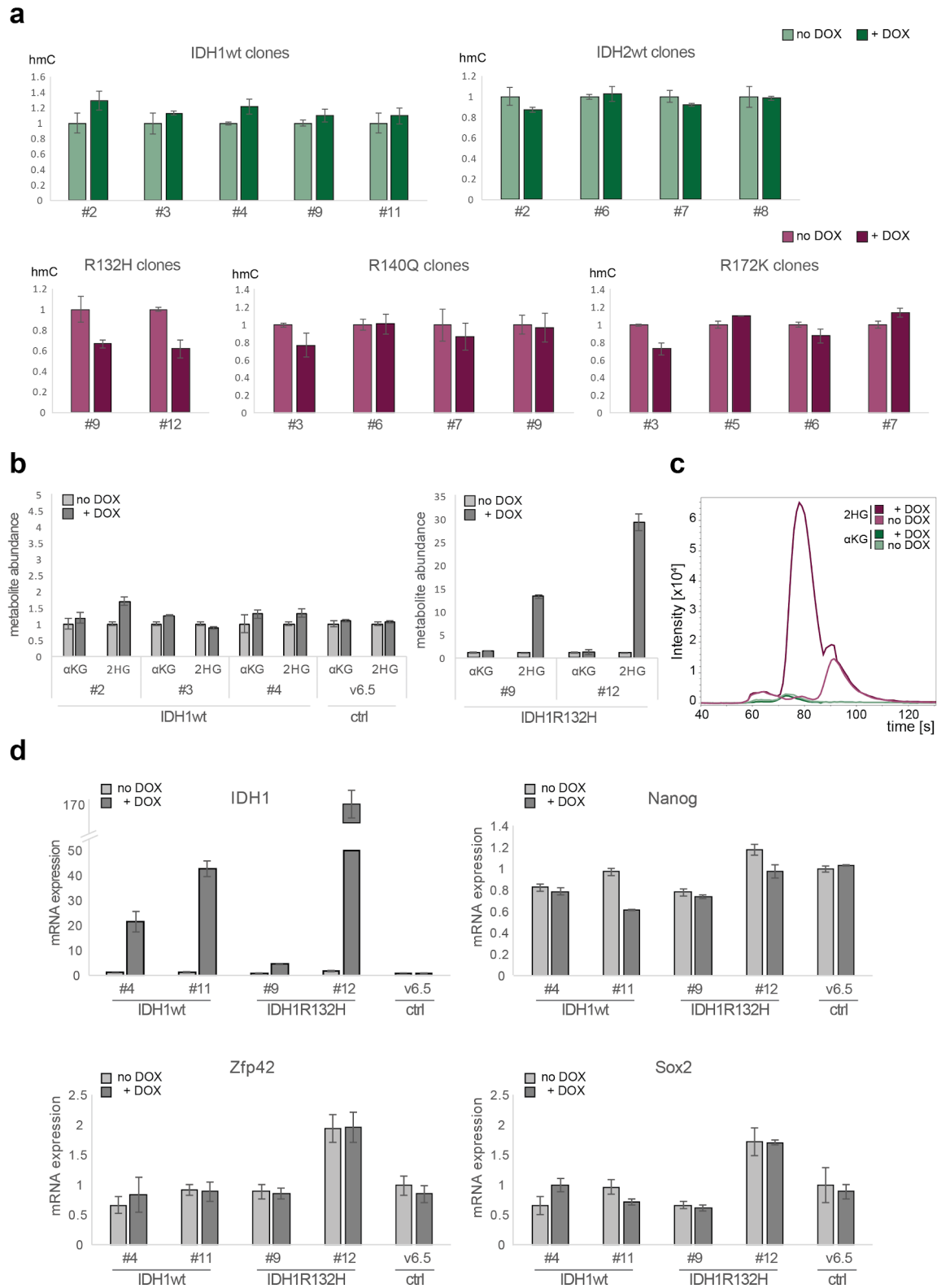


Figure 4: Endogenous production and effects of α -KG and 2-HG in mESCs with dox-inducible IDH. (a) Quantification of IF-stained nuclear 5hmC in mutant and wildtype IDH clones (indicated with #) after 48 h induction with 3 μ g/ml doxycycline (+DOX). (b+c) LC-MS/MS analysis of cell extracts of IDH1 clones after 48 h induction with 1 μ g/ml doxycycline (+DOX). Quantification of metabolites in (b) is based on triplicates. (c) depicts the chromatogram of replicate 1 of IDH1R132H clone 12 for the metabolites measured: α -KG and 2-HG. (d) Expression levels of IDH1, Nanog, Zfp42 and Sox2 in dox-induced IDH1 clones. RNA was isolated after 48h induction with 1 μ g/ml doxycycline and mRNA expression was determined via RT-qPCR.

Material & Methods

Cell culture

Mouse V6.5 ESCs were cultured on gelatin-coated culture flasks or 96-well microplates (μ Clear, Greiner Bio-One) in Dulbecco's modified Eagle's medium (Sigma) supplemented with 16% fetal bovine serum (FBS) (Sigma), 1x MEM Non-essential amino acids (Sigma), 2 mM L-glutamine (Sigma), 1% Pen/Strep, 0.1 mM β -mercaptoethanol, 0.2% recombinant leukemia inhibitory factor (LIF) (Millipore or homemade), 1 μ M PD032591 and 3 μ M CHIR99021 (2i, Axon Medchem). For experiments in which cells were cultured under "LIF only" conditions, the same medium was used without adding the inhibitors PD and CHIR (2i).

For the treatment of mESCs with cell-permeable α -KG and 2-HG, dimethyl- α -KG (dimethyl 2-oxoglutarate, Sigma-Aldrich 349631) or diethyl-2-HG (Pentanedioic acid, 2-hydroxy-, 1,5-diethyl ester, fluorochem 387349) was directly added to the cell culture medium with various final concentrations and for different time periods as indicated in the experiments. [Cautionary note: Based on extreme batch-to-batch variations that were observed during this study, new batches of esterified metabolites were routinely tested in GC-MS/MS analyses.] Non-esterified α -KG (α -Ketoglutaric acid disodium salt dihydrate, 75892 Sigma-Aldrich) and 2-HG (L- α -Hydroxyglutaric acid disodium salt, 90790 Sigma-Aldrich) were used as negative controls.

Cloning

For IDH expression constructs, the coding sequence of wildtype IDH1 and IDH2 was amplified from V6.5 ESC cDNA using respective primers with AsiSI and NotI sites. For the integration of IDH1 and IDH2 inserts into a pCAG-GFP-AsiSI-xxx-NotI vector, inserts and vector were digested with AsiSI/NotI, purified and ligated in a T4 DNA ligase reaction mix. DNA was transformed into JM109 competent bacteria and plasmids were purified using the PureYield Plasmid Midiprep System (Promega) according to the manufacturer's instruction. For the generation of mutant IDH constructs (IDH1R132H, IDH2R140Q, IDH2R172K), site-directed mutagenesis was applied. In brief, the respective mutation was generated in a two-step PCR. First, two separate PCR reactions were run using the aforementioned IDH plasmids as template with reaction mix I comprising the normal IDH-AsiSI-forward and a mutation-containing reverse primer. Reaction mix II comprised a mutation-containing forward and the normal IDH-NotI-reverse primer. These two PCR amplicons (overlapping at the mutation site) together served as the template for the second PCR reaction, which was run with the IDH-AsiSI-forward and IDH-NotI-reverse primer to amplify full-length IDH1 or IDH2 harboring the respective mutation.

For the generation of doxycycline-inducible IDH expression constructs, a previously-established sleeping beauty system with a pSBtet-3xFLAG-IRES-Ds-Red-Express-PuroR vector was used (Mulholland et al., 2018). First, wildtype and mutant forms of IDH1 and IDH2 were amplified from the pCAG-GFP-IDH constructs using primers with overhangs homologous to the AsiSI and NotI restriction sites of the pSBtet-

3xFLAG-IRES-Ds-Red-Express-PuroR vector. Second, wildtype and mutant IDH amplicons were cloned into the pSBtet-3xFLAG-IRES-Ds-Red-Express-PuroR vector (linearized with AsiSI and NotI) using Gibson assembly (Gibson et al., 2009).

Sleeping Beauty Transposition

To generate doxycycline-inducible, stable V6.5 mESCs with IDH1wt, IDH1R132H, IDH2wt, IDH2R140Q or IDH2R172K expression constructs, cells were transfected with equimolar amounts of the pSBtet-3xFLAG-IDH-IRES-Ds-Red-Express-PuroR and the Sleeping Beauty transposase, pCMV(CAT)T7-SB100 (Mátés et al., 2009, Addgene plasmid #34879) vector using Lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's instructions. Two days after transfection, cells were transferred into selection media with 1 µg/ml puromycin and propagated for 5-6 days. Afterwards, cells were induced with 1 µg/ml doxycycline overnight and sorted with FACS based on DsRed expression to select for cells that successfully integrated the sleeping beauty (SB) cassette.

To pick individual clones, SB-IDH cells were induced with 1 µg/ml doxycycline and plated with low density into p-100 dishes. Using an EVOS cell imaging system, DsRed-positive single colonies were absorbed with a pipette, transferred into a 96-well plate, separated into a single-cell suspension with trypsin and expanded before DsRed expression of individual clones was again checked by FACS analysis. Clones with DsRed expression >50% were kept for commencing experiments.

Immunofluorescence (IF) staining

For high-throughput imaging, mESCs were grown on 96-well microplates (µClear, Greiner Bio-One), washed with PBS and fixed with 4% formaldehyde in PBS for 10 min at room temperature (RT). The fixative was gradually exchanged to PBST (0.02 % Tween/PBS) and cells were washed twice with PBST. Cells were permeabilized (0.5 % Triton-X100/PBST) for 10 min at RT and washed twice with PBST again. Cells were then treated with denaturation solution (2 N HCl) for 40 min at RT, followed by the incubation in renaturation solution (150 mM Tris-HCl, pH 8.5) for 20 min at RT. Cells were subsequently blocked (2% BSA/PBST) for 60 min at RT and incubated with primary antibodies (mouse-anti mC, Diagenode 33D3; rabbit anti-hmC, active-motif 39769; rabbit anti-fC, active-motif 61223; rabbit anti-caC, active-motif 61225) for 1 h at 37 °C. After washing three times with PBST, cells were incubated with secondary antibodies (goat-anti-mouse or goat-anti-rabbit coupled to Alexa fluorophores) for 1 h at 37 °C. Cells were washed three times with PBST, counterstained with 200 ng/ml DAPI, washed again and finally covered with PBS. Images were acquired with the Operetta high-content image analysis system (PerkinElmer, 40x high NA objective) followed by analysis with the Harmony software (PerkinElmer). DAPI was used for the detection of single nuclei and cytosine modifications were measured in the selected nuclei based on the antibody signal intensity. Per modification and experimental replicate, between 100-3000 nuclei were analysed.

For higher resolution images of mESCs with DOX-inducible FLAG-tagged IDH cassettes, cells were grown on coverslips, stained with a mouse-anti-FLAG M2 primary antibody (Sigma, F3165) and a goat-anti-mouse secondary antibody coupled to Alexa 647. Coverslips were further incubated in a solution of DAPI (200 ng/ml) in PBS-T and mounted in Vectashield. Images were collected using a Leica TCS SP5 confocal microscope.

Western Blot

For western blot analysis of mESCs with inducible FLAG-IDH cassettes, protein fractions were isolated from cell samples using the NucleoSpin TriPrep Kit (Macherey-Nagel) according to the manufacturer's protocol. Protein fractions were subsequently separated in SDS-PAGE and blotted onto nitrocellulose membranes. A mouse-anti-FLAG M2 antibody (1:500, Sigma, F3165) was used for FLAG-IDH detection and a polyclonal rabbit anti-H3 antibody (1:5,000; ab1791, Abcam) served as loading control. Blots were probed with secondary antibodies goat anti-mouse (1:5,000; A9044 Sigma Aldrich) and goat anti-rabbit (1:5,000; 170–6515, Bio-Rad) conjugated to horseradish peroxidase (HRP) and visualized using an ECL detection kit (Thermo Scientific Pierce).

Slot Blot

For detection of 5hmC on genomic DNA, DNA fractions were isolated from cell samples using the NucleoSpin TriPrep Kit (Macherey-Nagel) according to the manufacturer's protocol. DNA was diluted in 2x SSC buffer, denatured at 95 °C for 10 min and loaded onto an activated nylon membrane (Amersham Hybond N+, GE Healthcare) with the Bio-Rad slot blot system. After denaturing and neutralizing the membrane, DNA was crosslinked to the membrane using a Stratagene StrataLinker UV1800. The membrane was then blocked in 5% milk and 5hmC was detected using a rabbit-anti-hmC primary antibody (active-motif 39769) and a fluorescence-coupled anti-rabbit secondary antibody. Fluorescence signal was visualized using a Amersham Typhoon TRIO Gel and Blot Imaging System (GE healthcare).

Quantitative real-time PCR (qRT-PCR) Analysis

RNA fractions were isolated from cell samples using the NucleoSpin Triprep Kit (Macherey-Nagel) according to the manufacturer's protocol. cDNA synthesis was performed with the High-Capacity cDNA Reverse Transcription Kit (Thermo Scientific) using 1000 ng of total RNA as input. qRT-PCRs were conducted with oligonucleotides listed in Table S1 using 0.5 ng cDNA as input in 8 µl reactions. LightCycler 480 SYBR Green I Master Mix (Roche) was used for SYBR green detection and the reactions were run on a LightCycler480 (Roche).

RNA-seq preparation, processing and analysis

RNA was isolated using the NucleoSpin Triprep Kit (Macherey-Nagel) according to the manufacturer's protocol and reversely transcribed into cDNA. Digital gene expression libraries for RNA-seq were produced using a modified version of single-cell RNA barcoding sequencing (SCRB-seq) optimized to accommodate bulk cells (Ziegenhain et al., 2017). Libraries were barcoded and mixed and 100 bp single end reads were sequenced on an Illumina HighSeq 1500. Resulting reads were mapped to the mouse genome build mm10 using STAR version STAR 2.5.1b. Differential expression analysis was performed in R using DESeq2 (Love et al., 2014) and genes with an adjusted $P < 0.05$ were considered to be differentially expressed. Hierarchical clustering was performed on genes differentially expressed upon α -KG treatment using k-means clustering ($k=4$) in combination with the ComplexHeatmap R-package (Gu et al., 2016).

GC-MS/MS and LC-MS/MS analysis of cell extracts

To analyze metabolic turnover rates of esterified 2-HG in wildtype mESCs and the endogenous production of α -KG and 2-HG in mESCs with DOX-inducible IDH cassettes, respective cell samples were seeded in triplicates. 5mM diethyl-2-HG was administered for 0-12 h to wildtype V6.5 mESCs and IDH-inducible cells were cultured in 1 μ g/ml DOX-containing medium for 48 h before samples were harvested. Therefor, cells were washed 2x in DMEM and incubated in ice-cold 80% MeOH supplemented with ribitol/sorbitol standard for 10 min at 4°C on a plate shaker. MeOH was subsequently transferred into an eppendorf tube and centrifuged at 10.000x g for 10 min at 4 °C. Supernatant was collected and stored at -80 °C. For both experiments, “blank” samples were included representing gelatine-coated and medium-filled cell culture vessels without cells to subtract non-cellular backgrounds.

For the analysis of DE-2-HG turnover rates, samples were evaporated in a SpeedVac concentrator and processed in a Pegasus HT GC-TOF-MS (Leco). Derivatization and injection of samples was completed automatically using the autosampler application. Chromatography columns included a 10 m pre-column and a 30 m VF-5 ms column. Quantitative analysis of α -KG and 2-HG production in DOX-induced samples was performed without derivatization using an impact high resolution hybrid quadrupole-time-of-flight (HD QTOF) mass spectrometer (Bruker). For the examination of both data sets, analytical software of the respective devices and the TagFinder software were used.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) (SFB1243 Cancer Evolution, Project A01). CT was further supported by the International Max Planck Research School for Molecular Life Sciences (IMPRS-LS). HL is a member of the Center for NanoScience (CeNS).

References

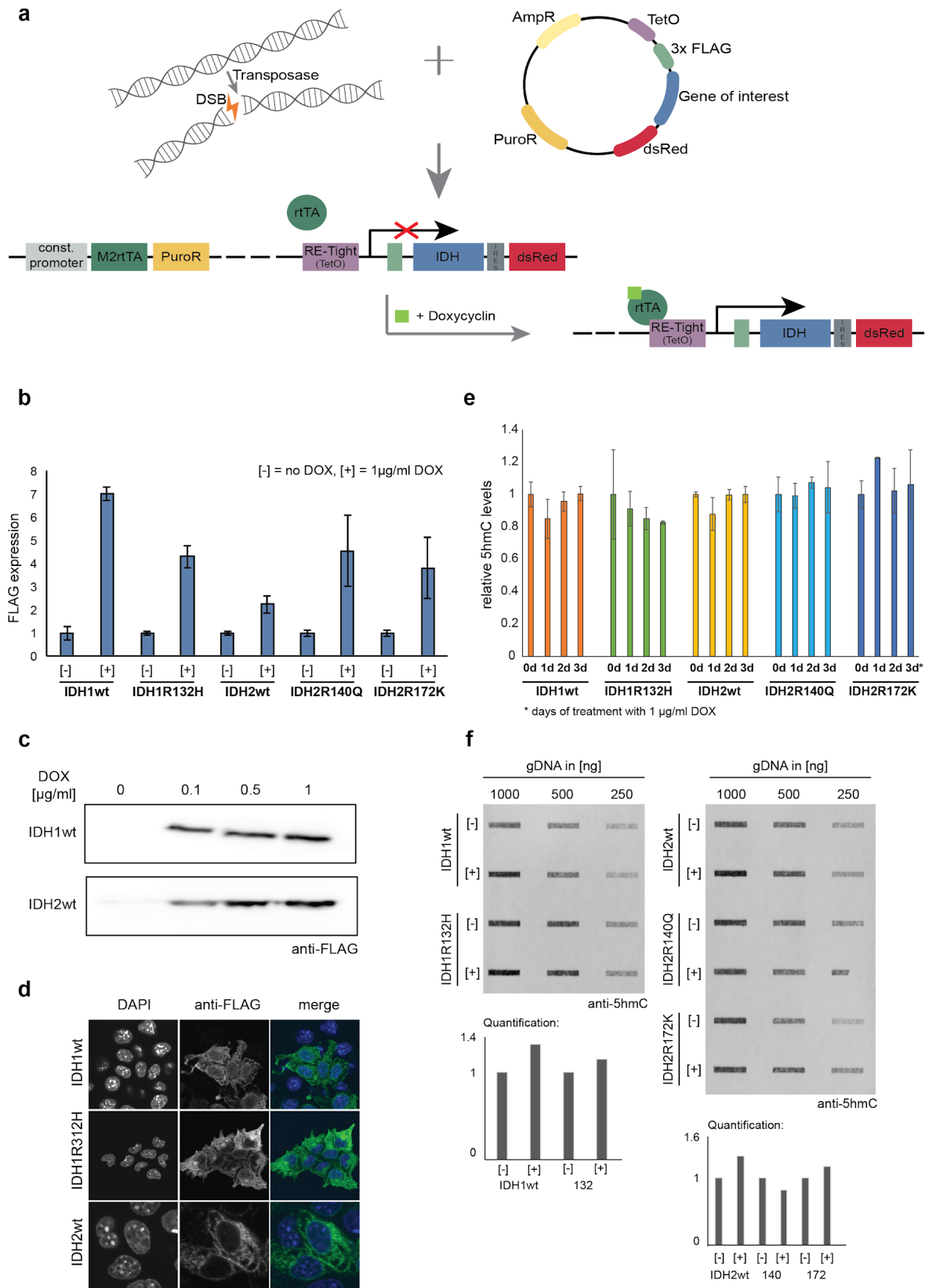
- Achouri, Y., Noël, G., Vertommen, D., Rider, M.H., Veiga-Da-Cunha, M., and Van Schaftingen, E. (2004). Identification of a dehydrogenase acting on D-2-hydroxyglutarate. *Biochem. J* 381, 35–42.
- Al-Khallaf, H. (2017). Isocitrate dehydrogenases in physiology and cancer: biochemical and molecular insight. *Cell Biosci.* 7, 37.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.
- Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500, 222–226.
- Bukowiecki, R., Adjaye, J., and Prigione, A. (2014). Mitochondrial function in pluripotent stem cells and cellular reprogramming. *Gerontology* 60, 174–182.
- Carey, B.W., Finley, L.W.S., Cross, J.R., Allis, C.D., and Thompson, C.B. (2015). Intracellular α -ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* 518, 413–416.
- Chaturvedi, A., Araujo Cruz, M.M., Jyotsana, N., Sharma, A., Goparaju, R., Schwarzer, A., Görlich, K., Schottmann, R., Struys, E.A., Jansen, E.E., et al. (2016). Enantiomer-specific and paracrine leukemogenicity of mutant IDH metabolite 2-hydroxyglutarate. *Leukemia* 30, 1708–1715.
- Dang, L., and Su, S.-S.M. (2017). Isocitrate Dehydrogenase Mutation and (R)-2-Hydroxyglutarate: From Basic Discovery to Therapeutics Development. *Annu. Rev. Biochem.* 86, 305–331.
- Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744.
- Figuroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567.
- Gaidzik, V.I., Paschka, P., Späth, D., Habdank, M., Köhne, C.-H., Germing, U., von Lilienfeld-Toal, M., Held, G., Horst, H.-A., Haase, D., et al. (2012). TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. *J. Clin. Oncol.* 30, 1350–1357.
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.
- Gross, S., Cairns, R.A., Minden, M.D., Driggers, E.M., Bittinger, M.A., Jang, H.G., Sasaki, M., Jin, S., Schenkein, D.P., Su, S.M., et al. (2010). Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J. Exp. Med.* 207, 339–344.
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.
- Harvey, A.J., Rathjen, J., and Gardner, D.K. (2016). Metaboloepigenetic Regulation of Pluripotent Stem Cells. *Stem Cells Int.* 2016, 1816525.
- 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.

RESULTS

- Janke, R., Dodson, A.E., and Rine, J. (2015). Metabolism and epigenetics. *Annu. Rev. Cell Dev. Biol.* *31*, 473–496.
- Jeltsch, A., and Jurkowska, R.Z. (2016). *DNA Methyltransferases - Role and Function* (Springer).
- King, A., Selak, M.A., and Gottlieb, E. (2006). Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. *Oncogene* *25*, 4675–4682.
- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* *502*, 472–479.
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics* *32*, 42–56.
- Lees, J.G., Gardner, D.K., and Harvey, A.J. (2017). Pluripotent Stem Cell Metabolism and Mitochondria: Beyond ATP. *Stem Cells Int.* *2017*, 2874283.
- Loenarz, C., and Schofield, C.J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends in Biochem. Sciences* *36*, 7–18.
- Losman, J.-A., Looper, R.E., Koivunen, P., Lee, S., Schneider, R.K., McMahon, C., Cowley, G.S., Root, D.E., Ebert, B.L., and Kaelin, W.G., Jr (2013). (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* *339*, 1621–1625.
- 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.
- Lu, C., Ward, P.S., Kapoor, G.S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C.R., Khanin, R., Figueroa, M.E., Melnick, A., et al. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* *483*, 474–478.
- Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., et al. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* *41*, 753–761.
- Mulholland, C.B., Ryan, J., Qin, W., Bartoschek, M.D., Traube, F.R., Parsa, E., Modic, M., Nixdorf, D., Ziegenhain, C., Carell, T., et al. (2018). TET1 drives global DNA demethylation via DPPA3-mediated inhibition of maintenance methylation.
- Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., Gallia, G.L., et al. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science* *321*, 1807–1812.
- Pollard, P.J., Wortham, N.C., and Tomlinson, I.P.M. (2003). The TCA cycle and tumorigenesis: the examples of fumarate hydratase and succinate dehydrogenase. *Ann. Med.* *35*, 632–639.
- Rzem, R., Vincent, M.-F., Van Schaftingen, E., and Veiga-da-Cunha, M. (2007). L-2-hydroxyglutaric aciduria, a defect of metabolite repair. *J. Inherit. Metab. Dis.* *30*, 681–689.
- Sperber, H., Mathieu, J., Wang, Y., Ferreccio, A., Hesson, J., Xu, Z., Fischer, K.A., Devi, A., Detraux, D., Gu, H., et al. (2015). The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nat. Cell Biol.* *17*, 1523–1535.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41–45.
- Struys, E.A., Verhoeven, N.M., Ten Brink, H.J., Wickenhagen, W.V., Gibson, K.M., and Jakobs, C. (2005). Kinetic characterization of human hydroxyacid-oxoacid transhydrogenase: relevance to D-2-

- hydroxyglutaric and gamma-hydroxybutyric acidurias. *J. Inherit. Metab. Dis.* 28, 921–930.
- Suganuma, T., and Workman, J.L. (2018). Chromatin and Metabolism. *Annu. Rev. Biochem.* 87, 27–49.
- Suijker, J., Baelde, H.J., Roelofs, H., Cleton-Jansen, A.-M., and Bovée, J.V.M.G. (2015). The oncometabolite D-2-hydroxyglutarate induced by mutant IDH1 or -2 blocks osteoblast differentiation in vitro and in vivo. *Oncotarget* 6, 14832–14842.
- 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.
- TeSlaa, T., Chaikovsky, A.C., Lipchina, I., Escobar, S.L., Hochedlinger, K., Huang, J., Graeber, T.G., Braas, D., and Teitell, M.A. (2016). α -Ketoglutarate Accelerates the Initial Differentiation of Primed Human Pluripotent Stem Cells. *Cell Metab.* 24, 485–493.
- Turcan, S., Rohle, D., Goenka, A., Walsh, L.A., Fang, F., Yilmaz, E., Campos, C., Fabius, A.W.M., Lu, C., Ward, P.S., et al. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483, 479–483.
- Tzika, E., Dreker, T., and Imhof, A. (2018). Epigenetics and Metabolism in Health and Disease. *Front. Genet.* 9, 361.
- Washington, J.M., Rathjen, J., Felquer, F., Lonic, A., Bettess, M.D., Hamra, N., Semendric, L., Tan, B.S.N., Lake, J.-A., Keough, R.A., et al. (2010). L-Proline induces differentiation of ES cells: a novel role for an amino acid in the regulation of pluripotent cells in culture. *Am. J. Physiol. Cell Physiol.* 298, C982–C992.
- Weber, A.R., Krawczyk, C., Robertson, A.B., Kuśnierczyk, A., Vågbø, C.B., Schuermann, D., Klungland, A., and Schär, P. (2016). Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications* 7.
- Xiao, M., Yang, H., Xu, W., Ma, S., Lin, H., Zhu, H., Liu, L., Liu, Y., Yang, C., Xu, Y., et al. (2012). Inhibition of α -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* 26, 1326–1338.
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., Ito, S., Yang, C., Wang, P., Xiao, M.-T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* 19, 17–30.
- Yang, Y.-H., Aloysius, H., Inoyama, D., Chen, Y., and Hu, L.-Q. (2011). Enzyme-mediated hydrolytic activation of prodrugs. *Yao Xue Xue Bao* 1, 143–159.
- Ye, D., Guan, K.-L., and Xiong, Y. (2018). Metabolism, Activity, and Targeting of D- and L-2-Hydroxyglutarates. *Trends Cancer Res.* 4, 151–165.
- Zdzisińska, B., Żurek, A., and Kandefler-Szerszeń, M. (2017). Alpha-Ketoglutarate as a Molecule with Pleiotropic Activity: Well-Known and Novel Possibilities of Therapeutic Use. *Arch. Immunol. Ther. Exp.* 65, 21–36.
- Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C., et al. (2012). HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* 31, 2103–2116.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* 65, 631–643.e4.

Supplementary Material



FLAG-IDH-dsRed expression cassette in turn is controlled by a TetO (RE-Tight) operator that is activated through doxycycline-bound transactivator (TA) binding. Upstream of the dox-inducible cassette, the vector carries a puromycin resistance under a constitutive promoter. To stably integrate the vector, the sleeping beauty transposase is utilized that recognizes its consensus sequence flanking the genetic cargo within the vector and inserts the cargo into genomic TA dinucleotides through DNA double strand breaks (DSBs). (b-d) Validation of IDH-inducibility in respective mESC lines after 24 h of 0-1 $\mu\text{g/ml}$ doxycycline (DOX) administration. (b) Quantification of IF-stained FLAG expression. (c) FLAG expression upon different DOX concentrations by western blot analysis. (d) Confocal images of mESCs stained against DAPI and FLAG. (e) Quantification of IF-stained nuclear 5hmC in cells treated with 1 $\mu\text{g/ml}$ DOX for 0-3 days. (f) Detection and quantification of 5hmC on genomic DNA, isolated from cells treated with [+] or without 1 $\mu\text{g/ml}$ DOX [-] for 7 days.

table S1: qRT-PCR primer		
gene	primer	sequence
<i>Idh1</i>	fwd	AGACTCAGTCGCCCAAGGT
	rev	GCGGTAGTGACGTGTGACAG
<i>Idh2</i>	fwd	CGCCACTATGCTGAGAAGAGG
	rev	ATCCACGTGAGGCAGGATGA
<i>Nanog</i>	fwd	ATTCTGGGAACGCCTCATCAA
	rev	TTCAGAGGAAGGGCGAGGA
<i>Zfp42</i>	fwd	CTGGGACACGTGGCAAAAGAA
	rev	GGGACAACACTTGGAGGCAG
<i>Sox2</i>	fwd	ACAGATGCAACCGATGCACC
	rev	TGGAGTTGTACTGCAGGGCG
<i>Gapdh</i>	fwd	CATGGCCTTCCGTGTTTCCTA
	rev	CTTACCACCTTCTTGATGTCATC

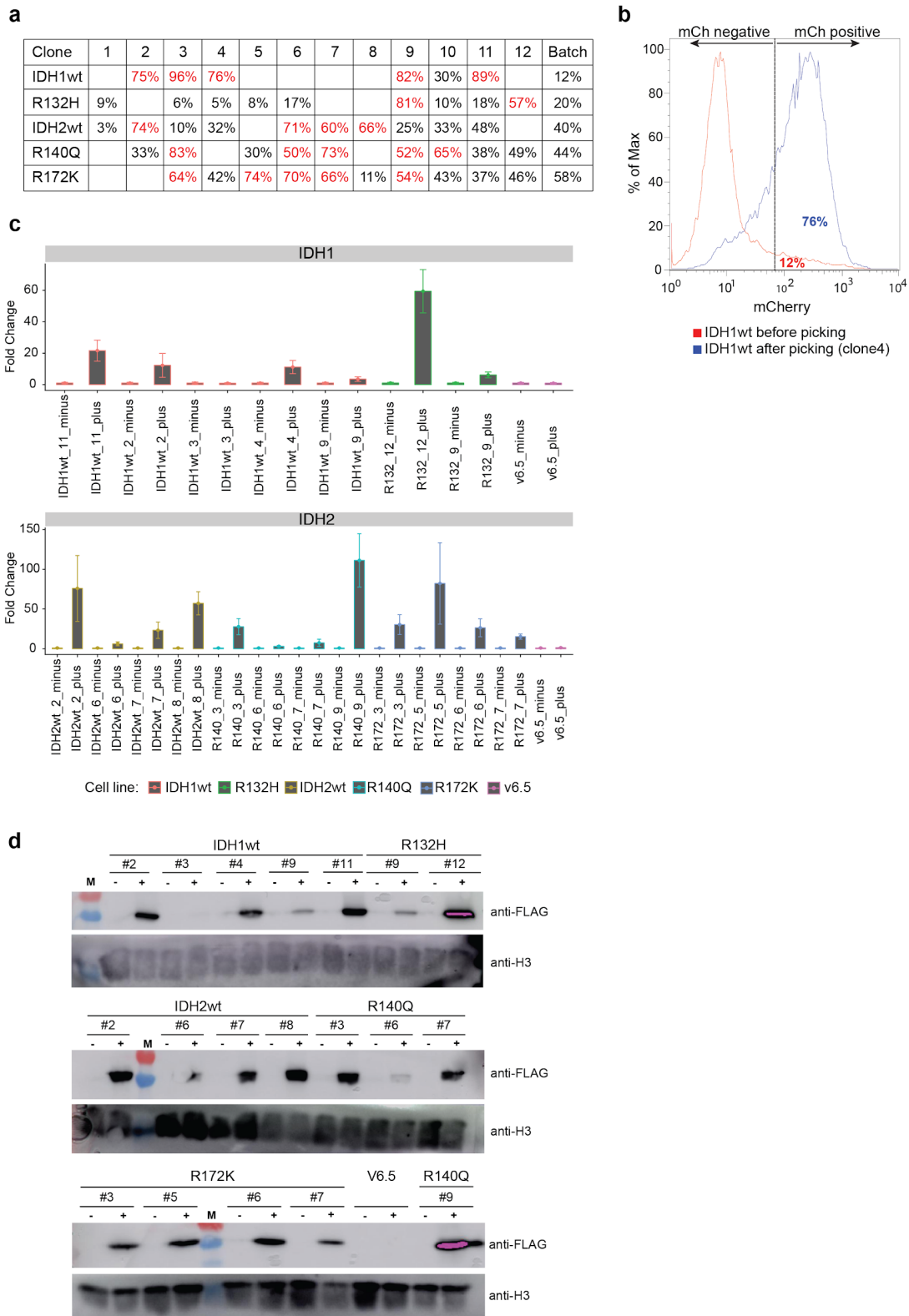


Figure S2: Characterization of inducible IDH clones. (a) overview of individual clones. For each IDH cell line single colonies/clones with moderate dsRed-intensity were picked and propagated. Percentages represent amount of dsRed-positive cells in FACS analysis of propagated clones or initial batch. (b) population of dsRed-negative and -positive cells in FACS analysis of IDH1wt batch (red) and IDH1wt clone 4 (blue). (c+d) RT-qPCR analysis of IDH1 and IDH2 mRNA expression (c) and Western Blot analysis of FLAG-IDH1 and -IDH2 protein expression (d) in the respective clones. RNA and protein was isolated from cells treated with (plus) or without (minus) 1 μ g/ml doxycycline for 48 h. control: wildtype V6.5 ESCs.

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

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 BiImaging, 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 re-expression 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 tgc cat gtc aac cag gag gga agc tta agc ggc cgc cgg 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:

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 (α KG) 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 μ g/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 μ m. Following dewaxing in xylol (3 \times 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 μ g/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 μ g/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 \times 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 \times /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 \times long/0.08 and 20 \times 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 fragmentation 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. 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

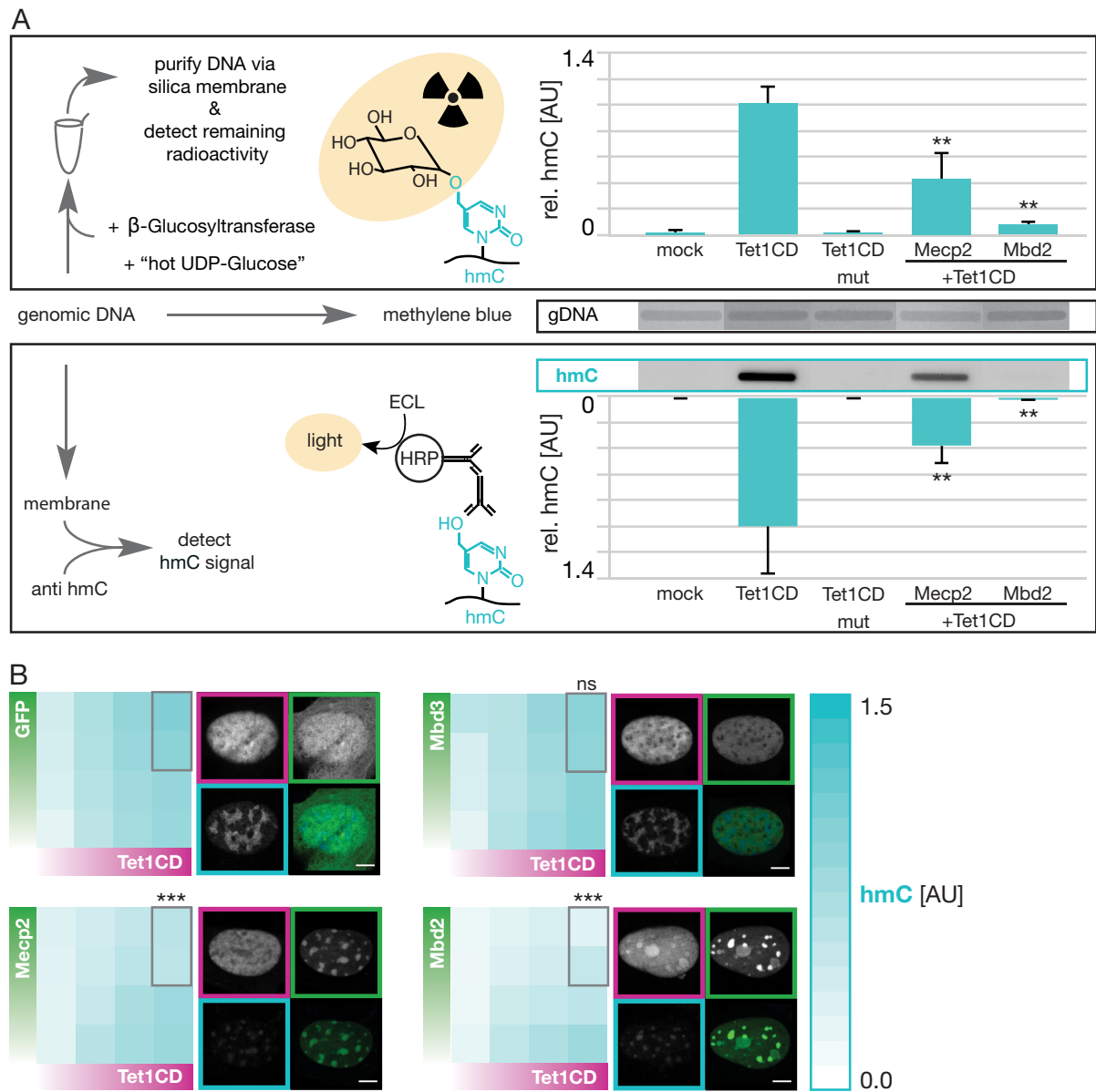


Figure 1. Impact of methyl-CpG binding domain proteins on the efficiency of Tet1 mediated 5mC oxidation. **(A)** Radioactive (top) and immunological (bottom) assay to determine 5hmC levels in genomic DNA (gDNA) of transiently transfected HEK cells. Schemes (left) illustrate the workflow and mode of 5hmC detection. Histograms (right) show relative 5hmC levels of three independent experiments +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 μ m.

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

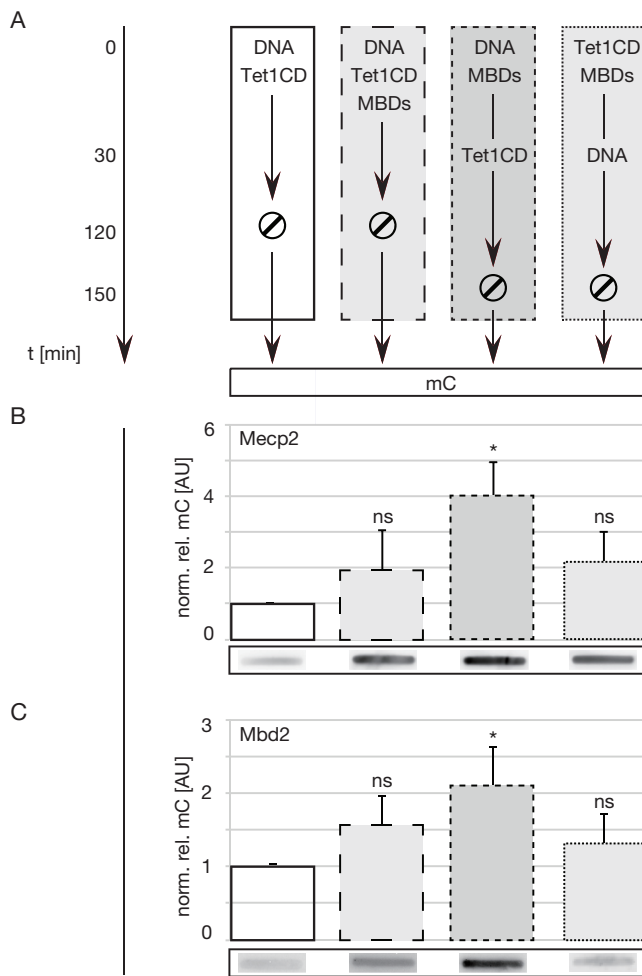


Figure 2. Influence of the chronological DNA binding order on the protecting ability of MBD proteins. Quantification of remaining 5mC levels in 20 pmol of double-stranded DNA containing multiple 5mC nucleotides after simultaneous or successive incubation with 20 pmol of Tet1CD (catalytic domain) and 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

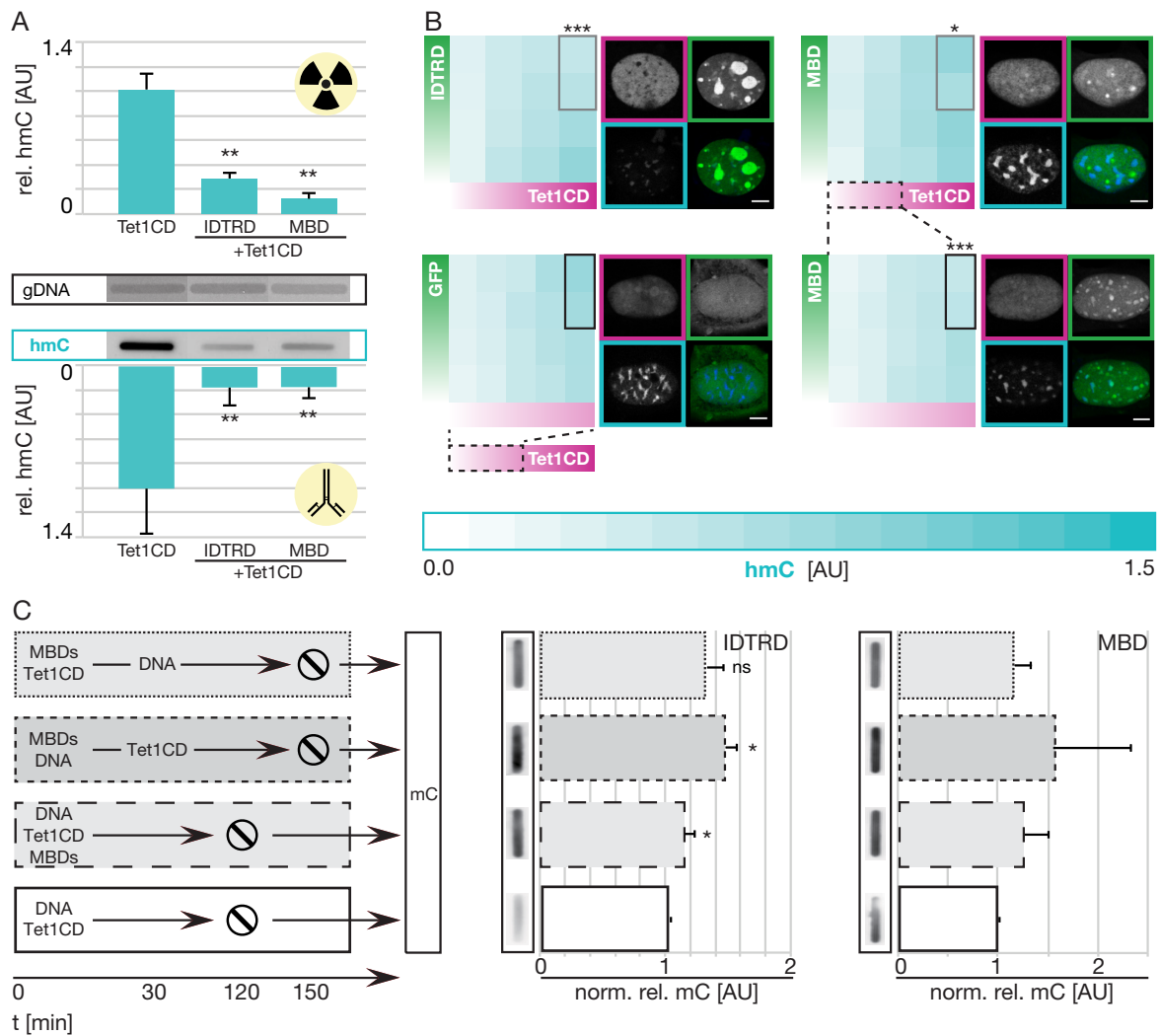


Figure 3. Effect of different DNA binding modes on Tet1 activity. **(A)** Radioactive (top) and immunological (bottom) assay to determine 5hmC levels in genomic DNA (gDNA) of transiently transfected HEK cells. Histograms show relative 5hmC levels of three independent experiments + SD. Tet1CD + IDTRD 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 5hmC 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 S10C) 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-

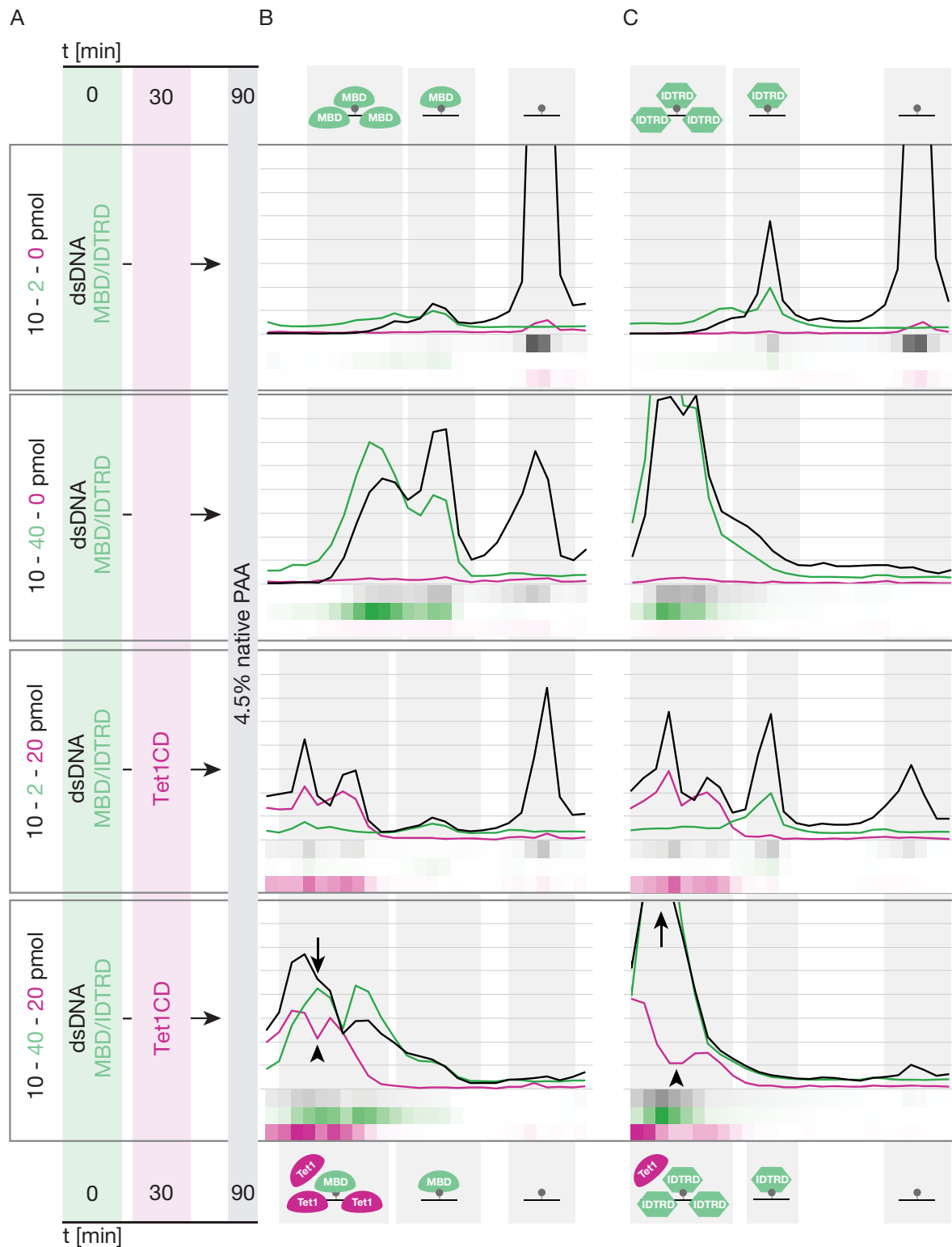


Figure 4. Impact of 5mC-specific and sequence-unspecific DNA binding proteins on the DNA binding ability of Tet1CD proteins. Electrophoretic mobility shift assay (EMSA) to determine the binding ability of fluorescently tagged Tet1CD to double-stranded, single mC containing DNA (ATTO647 labeled) in the 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).

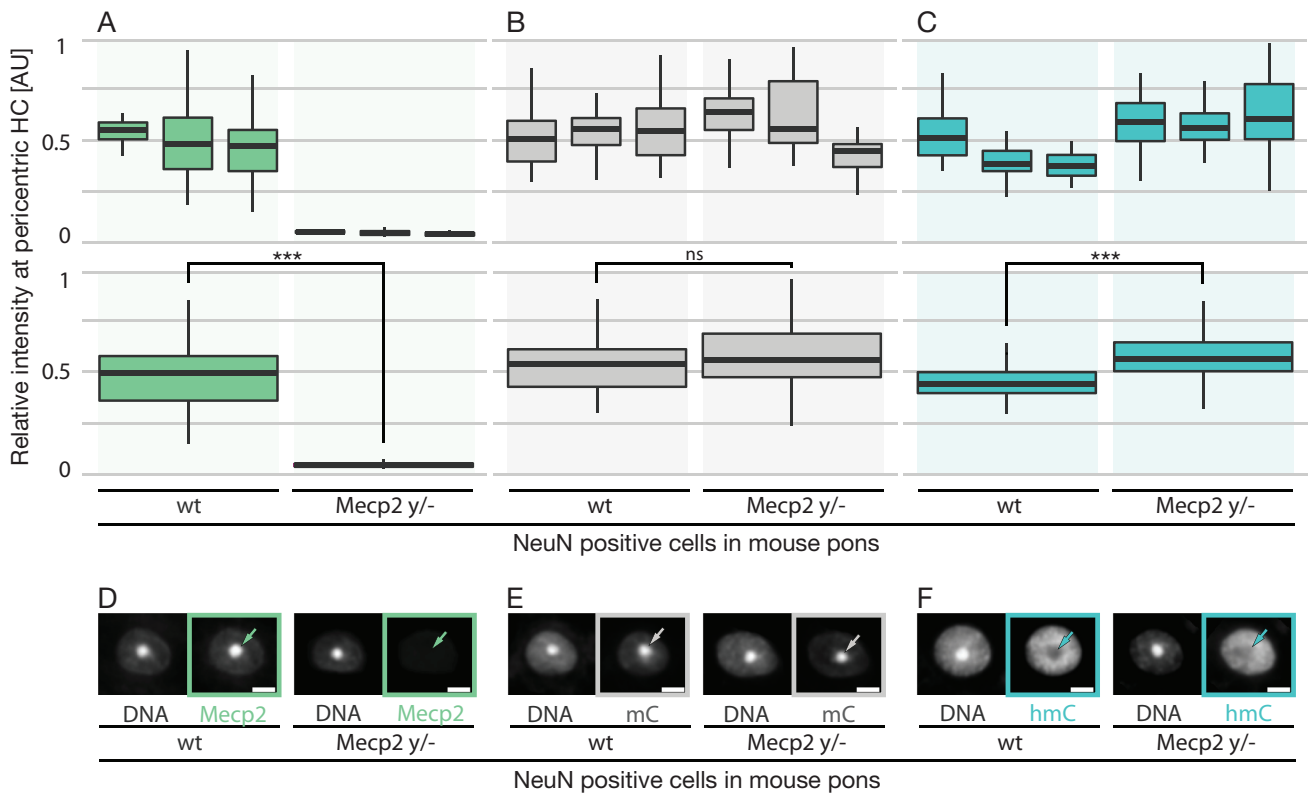


Figure 5. Epigenetic composition of pericentric heterochromatin in a mouse model for Rett syndrome. Immunostaining and quantification of MeCP2, 5mC and 5hmC levels in NeuN positive cells of wild type and MeCP2 knockout ($y^{-/-}$) mouse pons, respectively. (A–C) Box plots represent the distribution of (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 μ m.

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^{-/ym1.1Bird}), 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

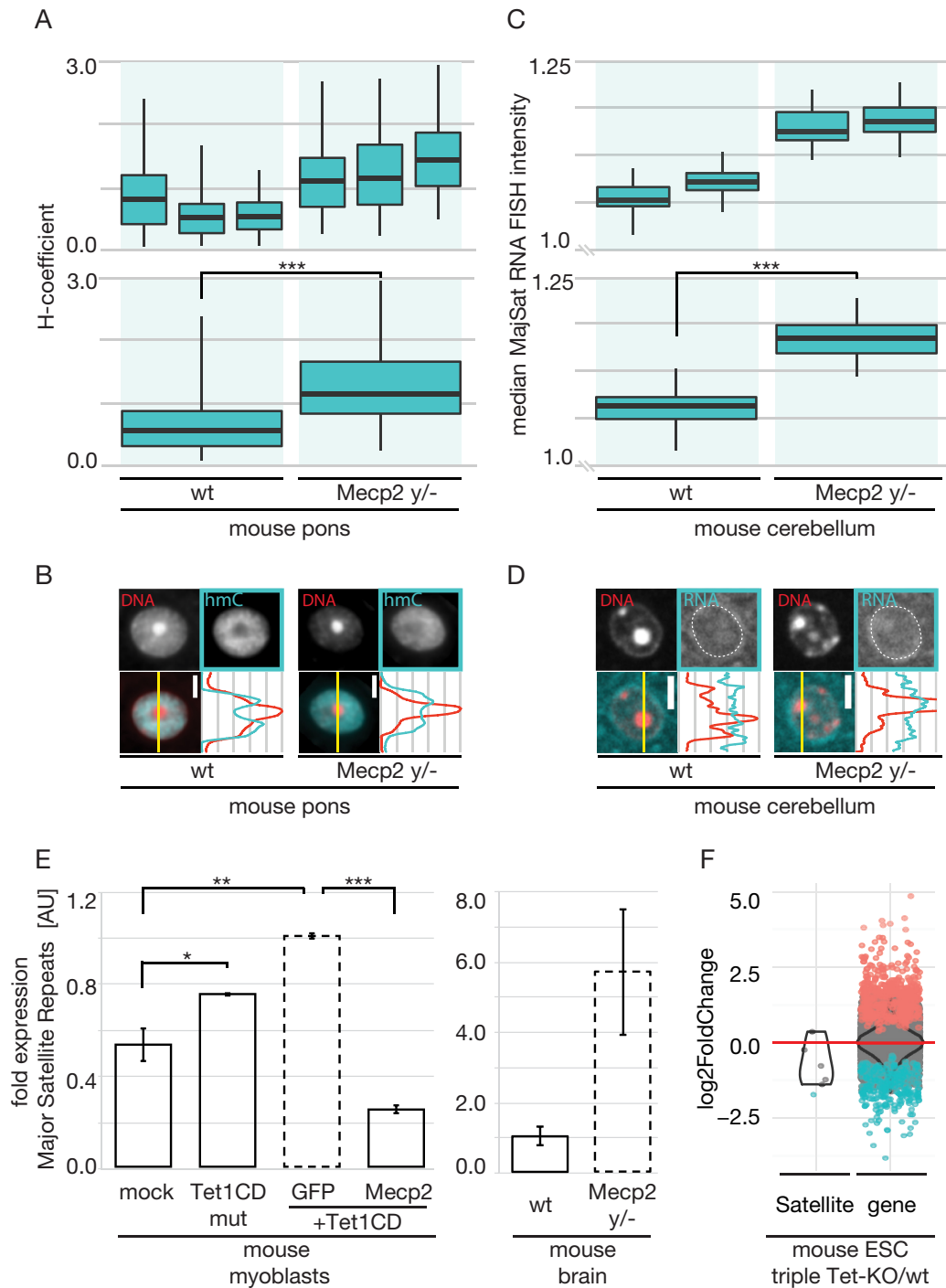


Figure 6. Correlation of subnuclear 5hmC distribution and major satellite expression. **(A and B)** Immunostaining and colocalization analysis of 5hmC with pericentric heterochromatin. **(A)** Box plots show the median 5hmC colocalization with pericentric heterochromatin in neurons of three individual (top) and combined (bottom) wild type and *Mecp2* y/- mouse pons (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 pons, respectively. DNA was counterstained with DAPI. Scale bar, 5 μm. **(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 pons, respectively. DNA was counterstained with DAPI. Scale bar, 5 μm. **(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} .

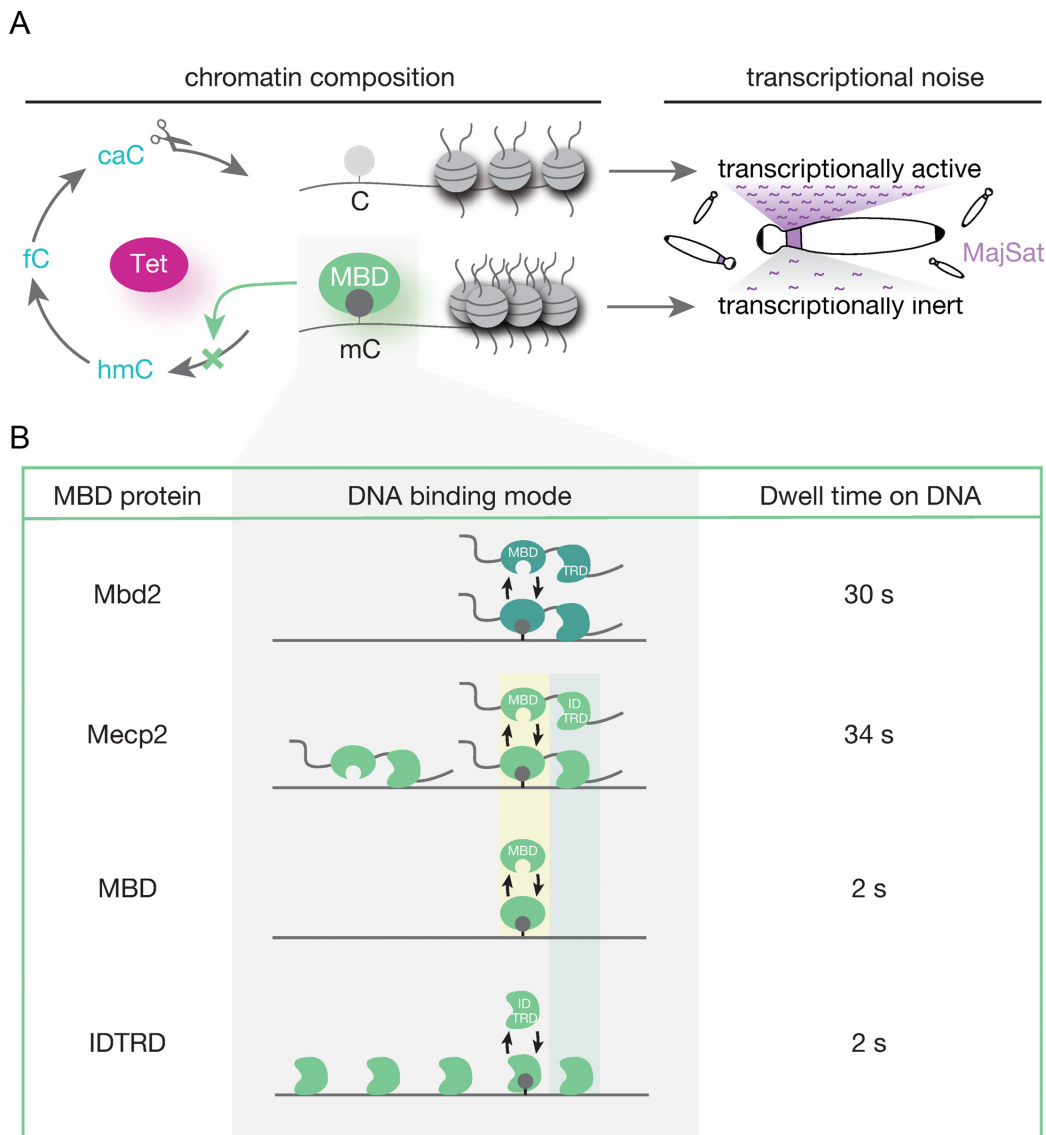


Figure 7. Mecp2 and Mbd2 preserve chromatin composition and thus genomic integrity by insulating 5mC from Tet1 activity. Scheme summarizing the main conclusions drawn from our studies. (A) We show that Mecp2 and Mbd2 protect 5mC from Tet1-mediated oxidation in a concentration dependent manner *in vivo* 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 Mildner 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.

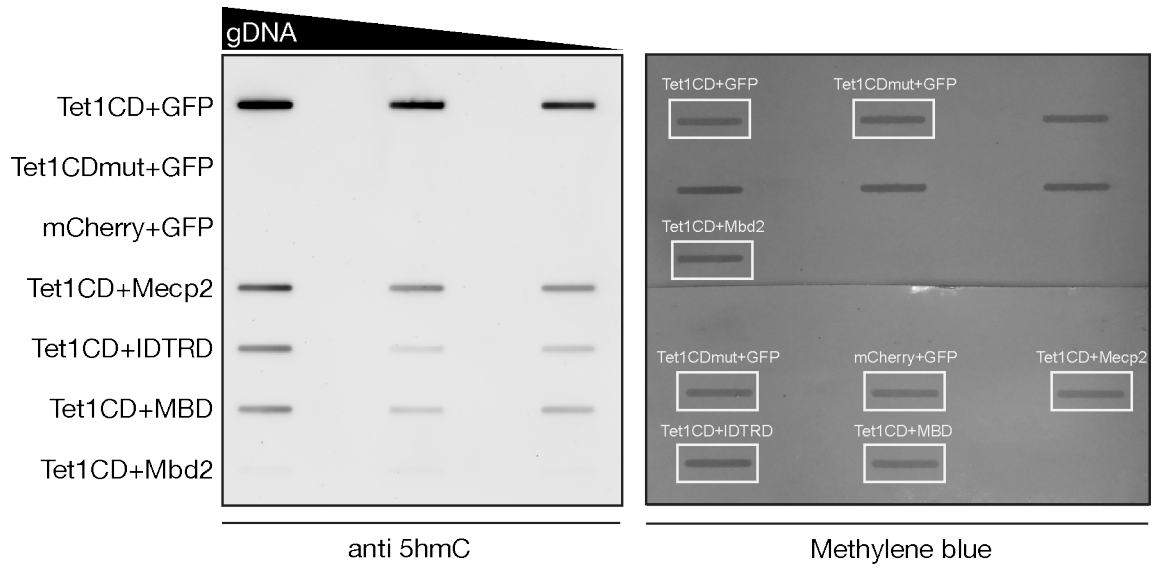
REFERENCES

- 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., Gierasch, L.M., Uversky, V.N., Hite, K., Hansen, J.C. and Woodcock, C.L. (2010) Unique physical properties and interactions of the domains 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. *Tubmb 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-ribosylation) 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. Moortusewicz, 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., Franz, 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. Menafrá, 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.

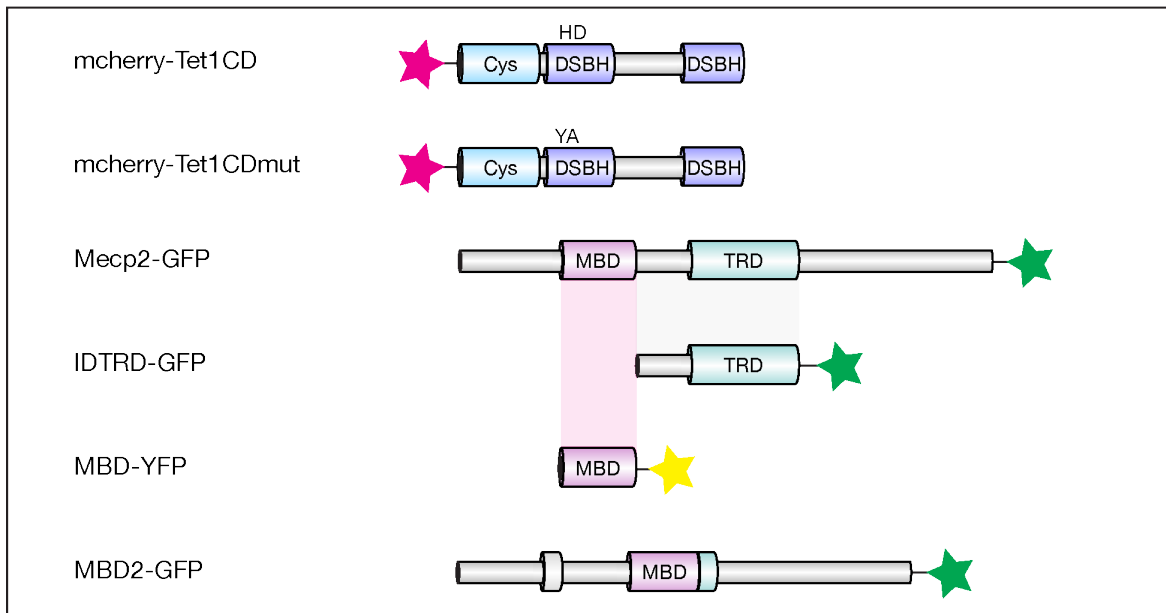
67. Hashimoto, H., Pais, J.E., Zhang, X., Saleh, L., Fu, Z.Q., Dai, N., Correa, I.R. Jr, Zheng, Y. and Cheng, X. (2014) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature*, **506**, 391–395.
68. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., 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., Michalakakis, 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., Liyo, 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

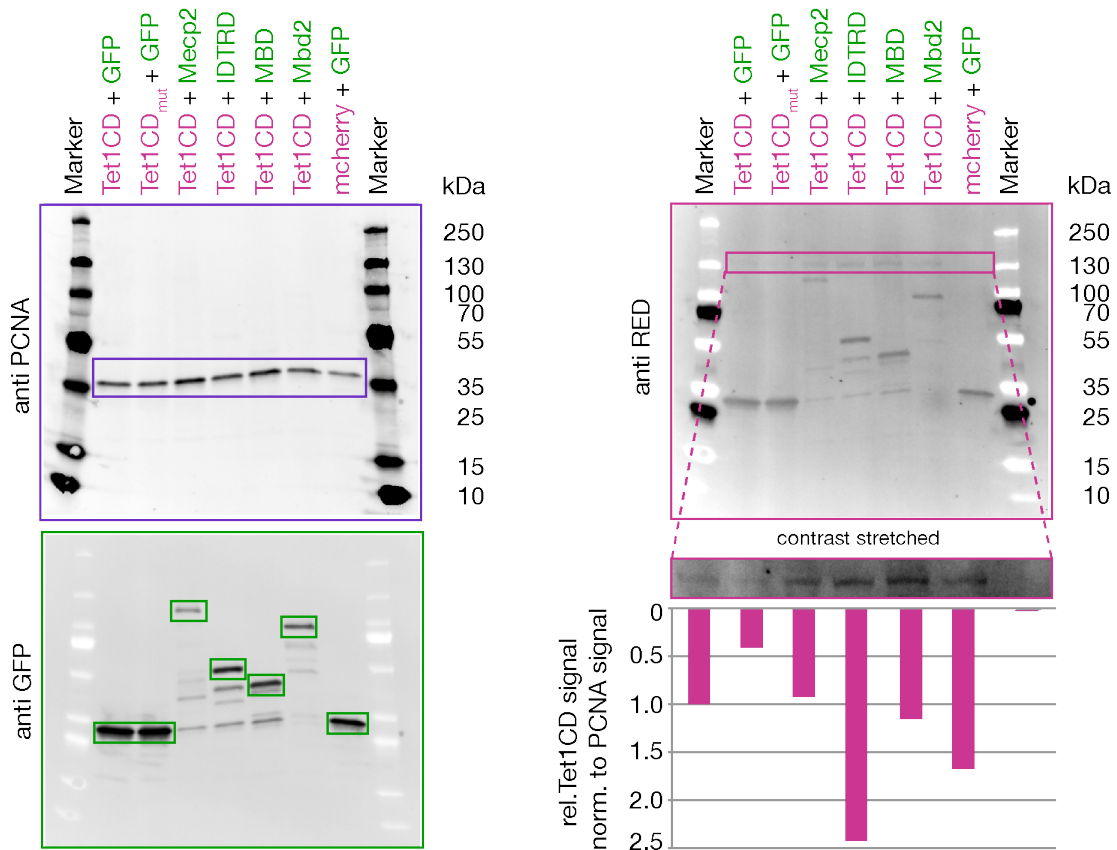


Supplementary Figure 1. Full slot blots used for the quantification of 5hmC in genomic DNA (gDNA) of transiently transfected HEK cells (Figure 1a, bottom). (left) anti 5hmC. (right) Methylene blue staining.

a



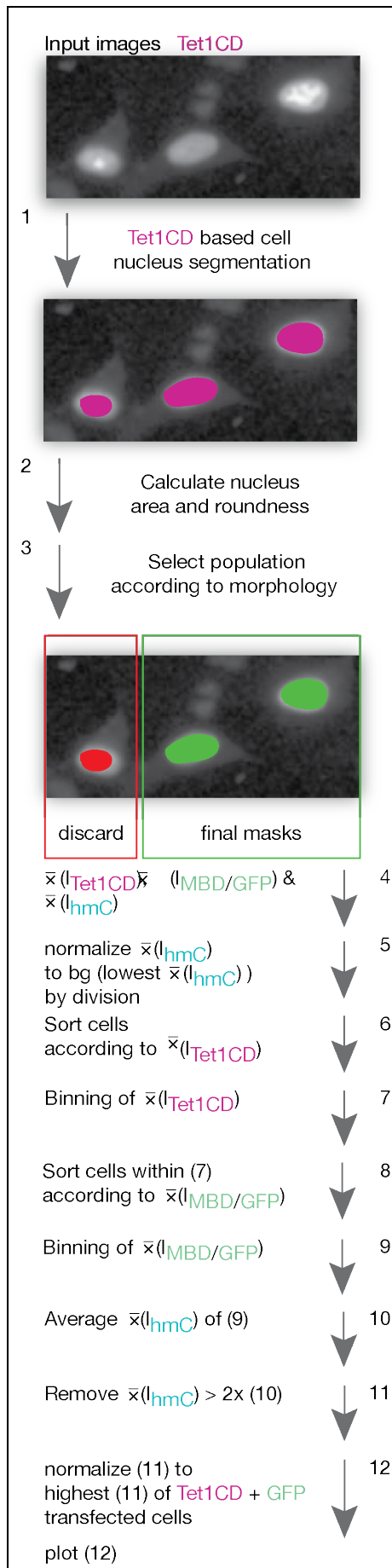
b



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

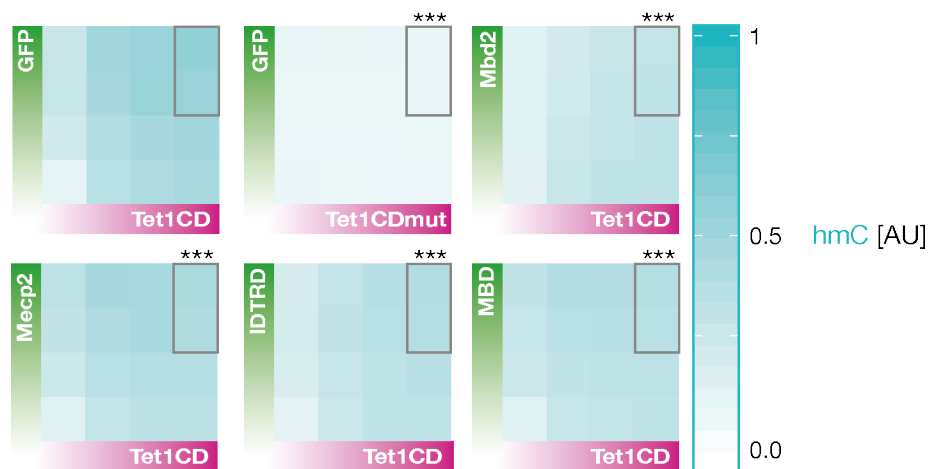
(a) Representative schemes of proteins that were used for *in vitro* and *in vivo* studies. Tet1CD corresponds to the Tet1 catalytic domain (aa 1365-2057) and Tet1CDmut is the catalytically inactive

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.



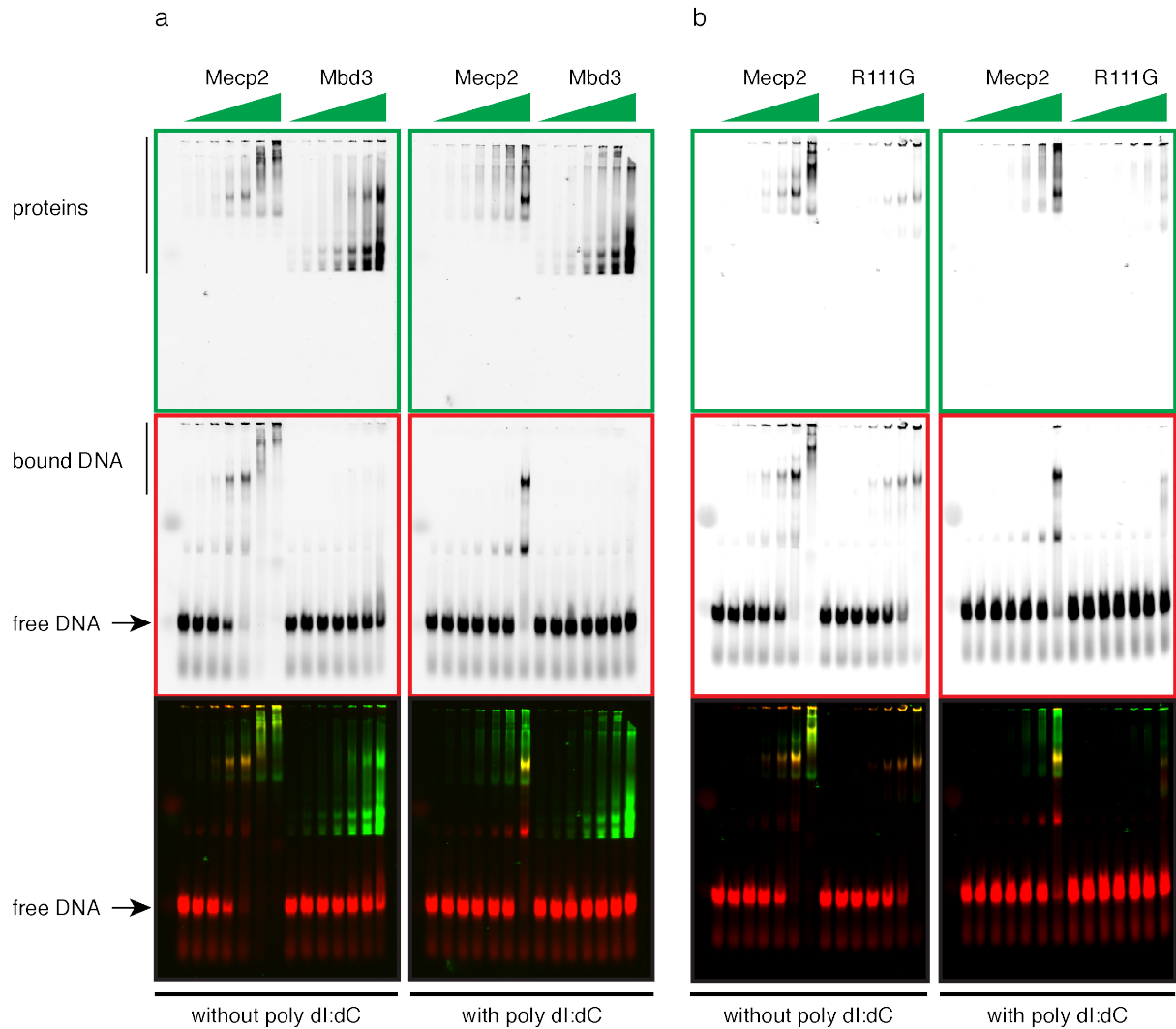
Supplementary Figure 3. Summary of the different steps used for quantification and normalization of genomic 5hmC levels in transiently transfected C2C12 mouse myoblasts

(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.

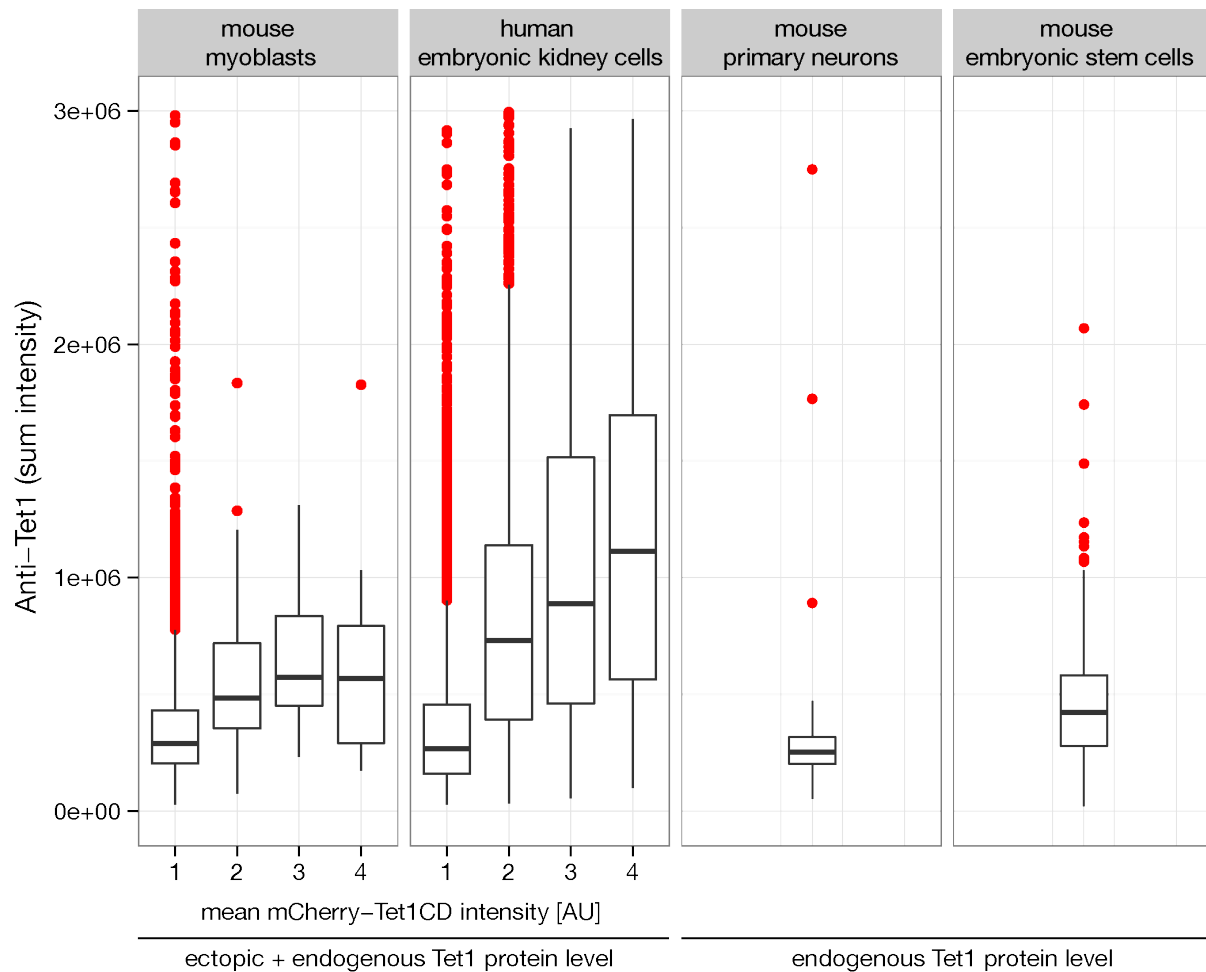


Supplementary Figure 4. Effect of methyl-CpG binding domain proteins on Tet1CD activity in human embryonic kidney cells

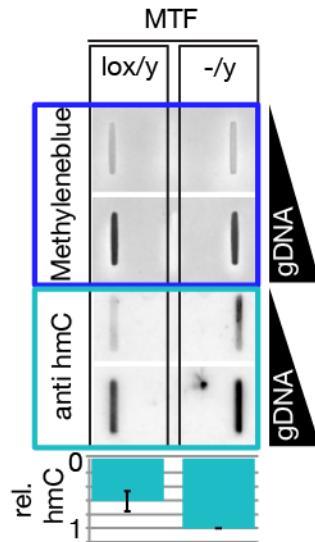
In situ staining and quantification of genomic 5hmC levels in transiently transfected human embryonic kidney (HEK) cells. Images were acquired on an automated high throughput imaging system with a 20x, 0.45 NA objective. Gradient heat maps show relative 5hmC (=cyan) signals as a function of increasing 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.



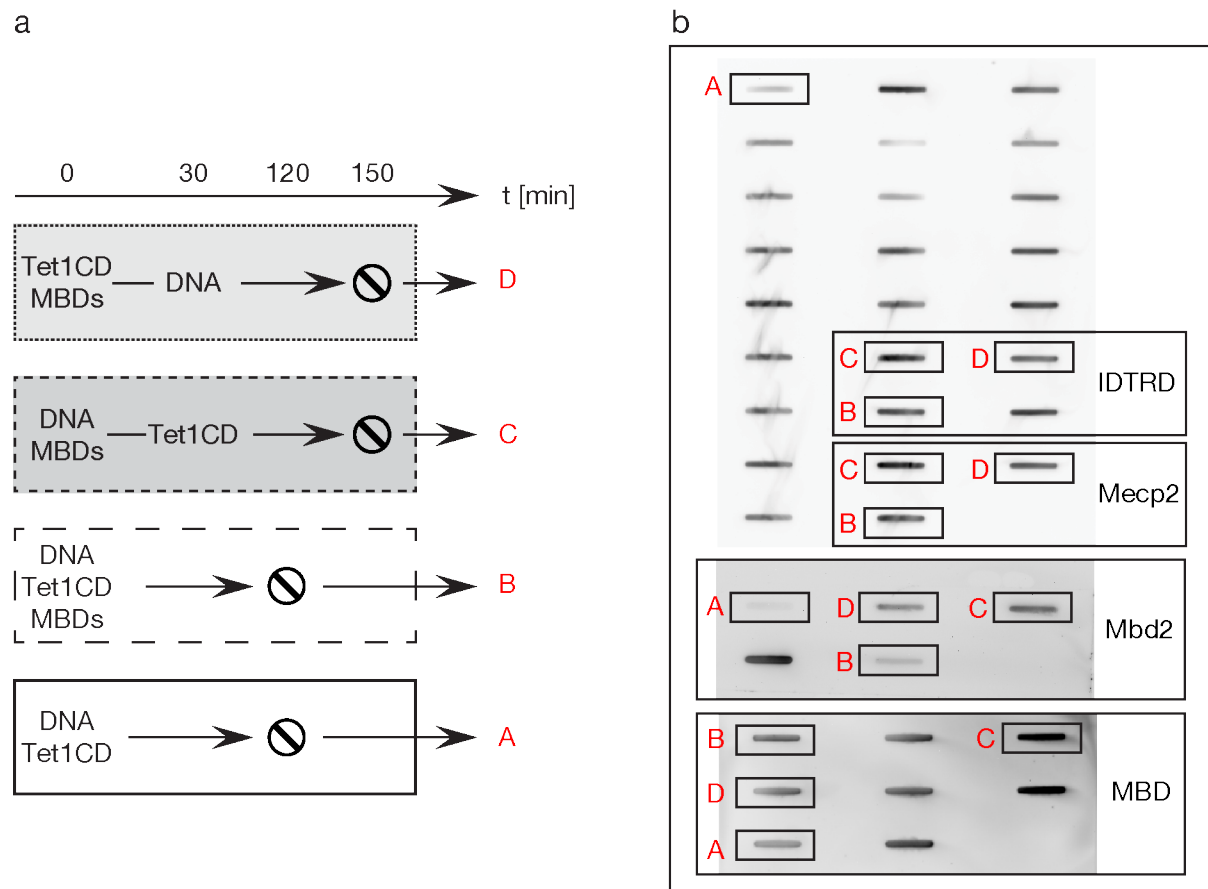
Supplementary Figure 5. Binding properties of MeCP2, Mbd3 and MeCP2_{R111G} to methylated DNA. (a) Electrophoretic mobility shift assay (EMSA) of methylated DNA (red) with MeCP2 and Mbd3 (green) in the absence (left) and presence (right) of poly dI:dC. (b) Electrophoretic mobility shift assay (EMSA)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

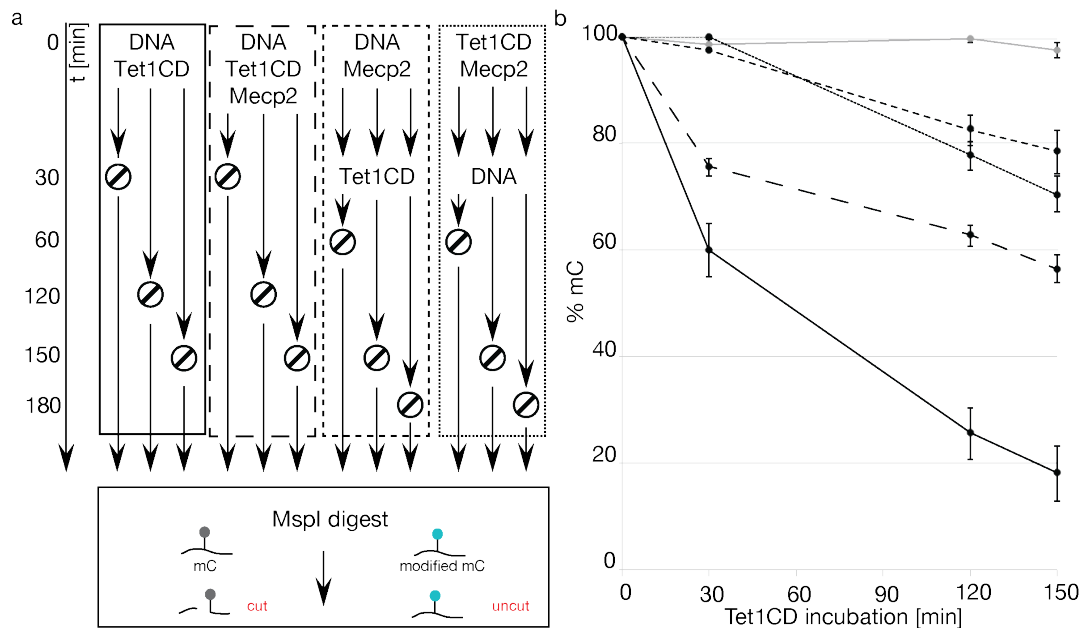


Supplementary Figure 7. Genomic 5hmC level in the presence and absence of Mecp2 in mouse tail fibroblast (MTF) cells. Immunological assay to determine 5hmC levels in genomic DNA (gDNA) of wild type (lox/y) and Mecp2 knockout (-/y) MTF cells. gDNA quantities were monitored by methylene blue staining.



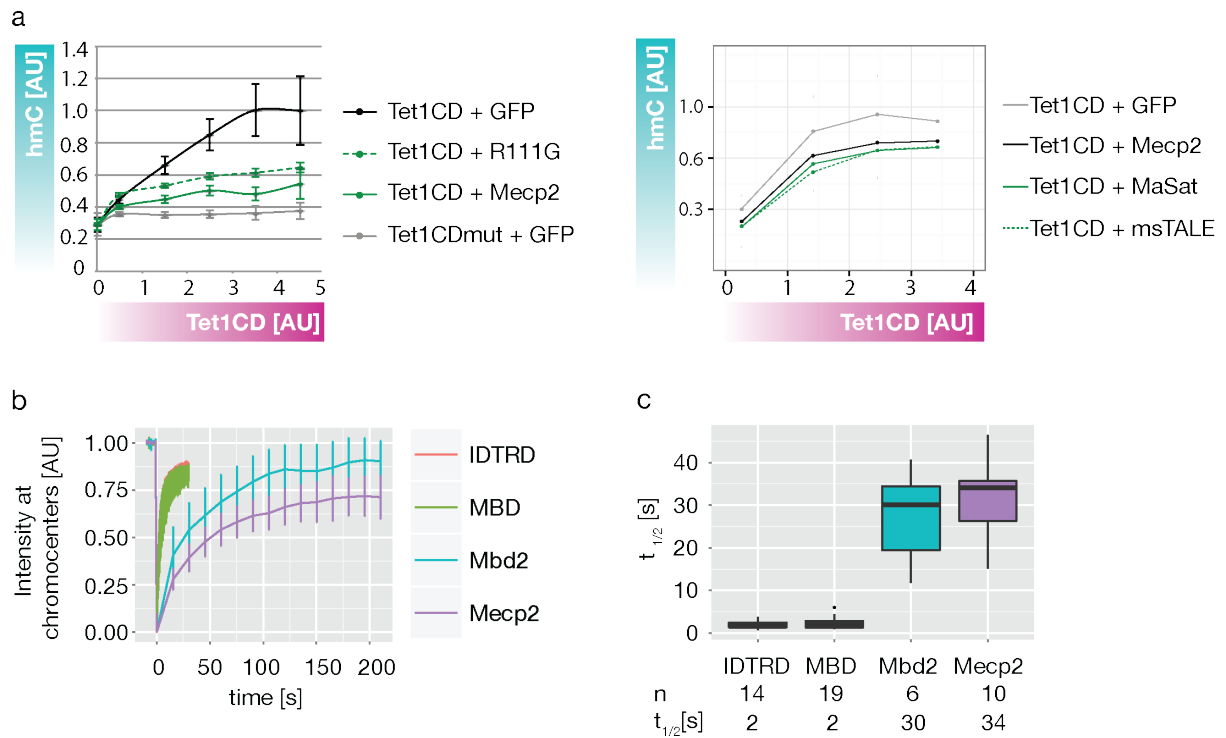
Supplementary Figure 8. Full slot blots used for the quantification of remaining 5mC in double-stranded DNA after simultaneous or successive incubation with Tet1CD and MBD proteins

(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 MBD. Unmarked slots are unrelated to the figure.



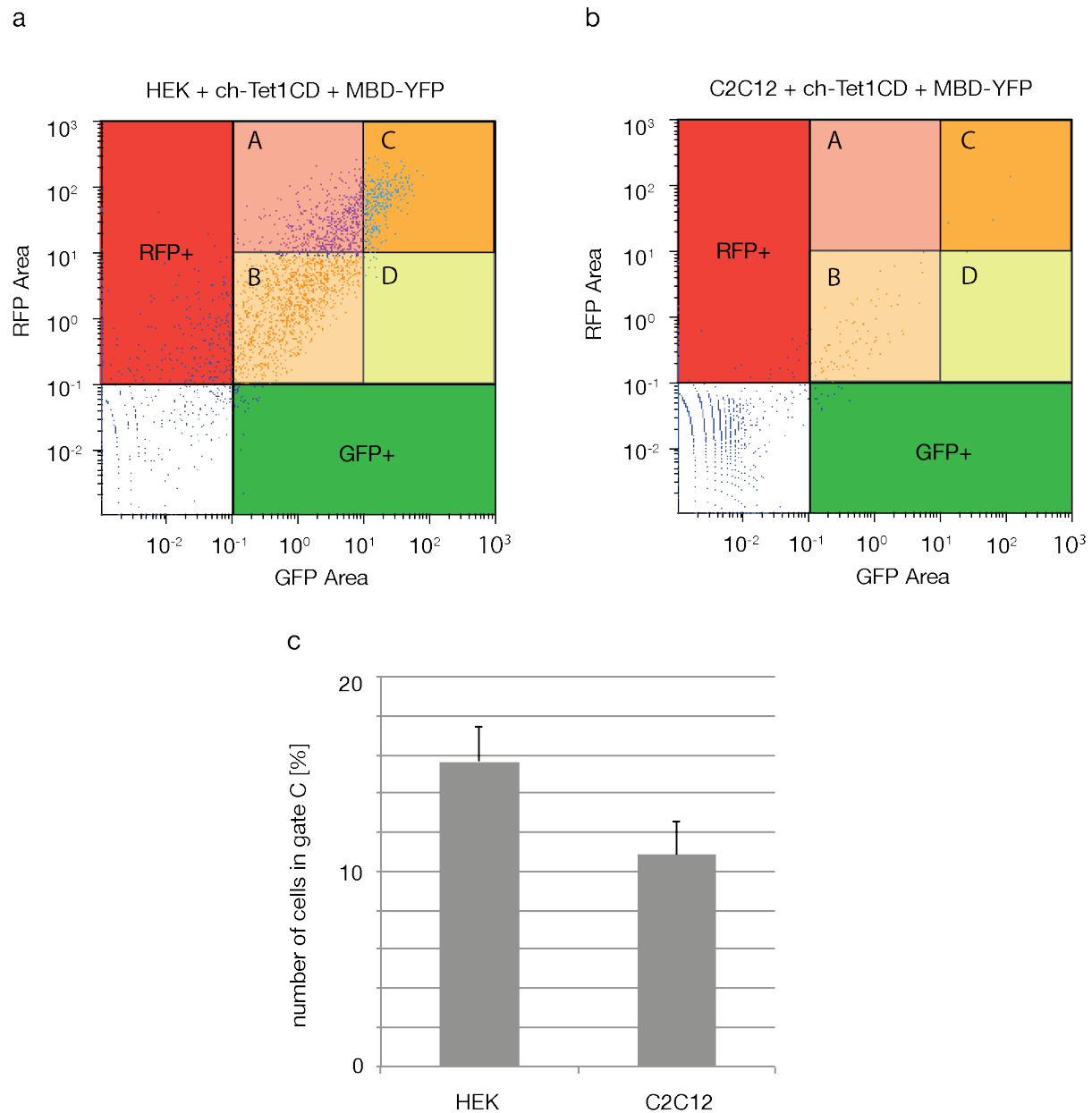
Supplementary Figure 9. Influence of the chronological DNA binding order on the protecting ability of MBD proteins over time

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 uncuttable. (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 \pm SD.



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

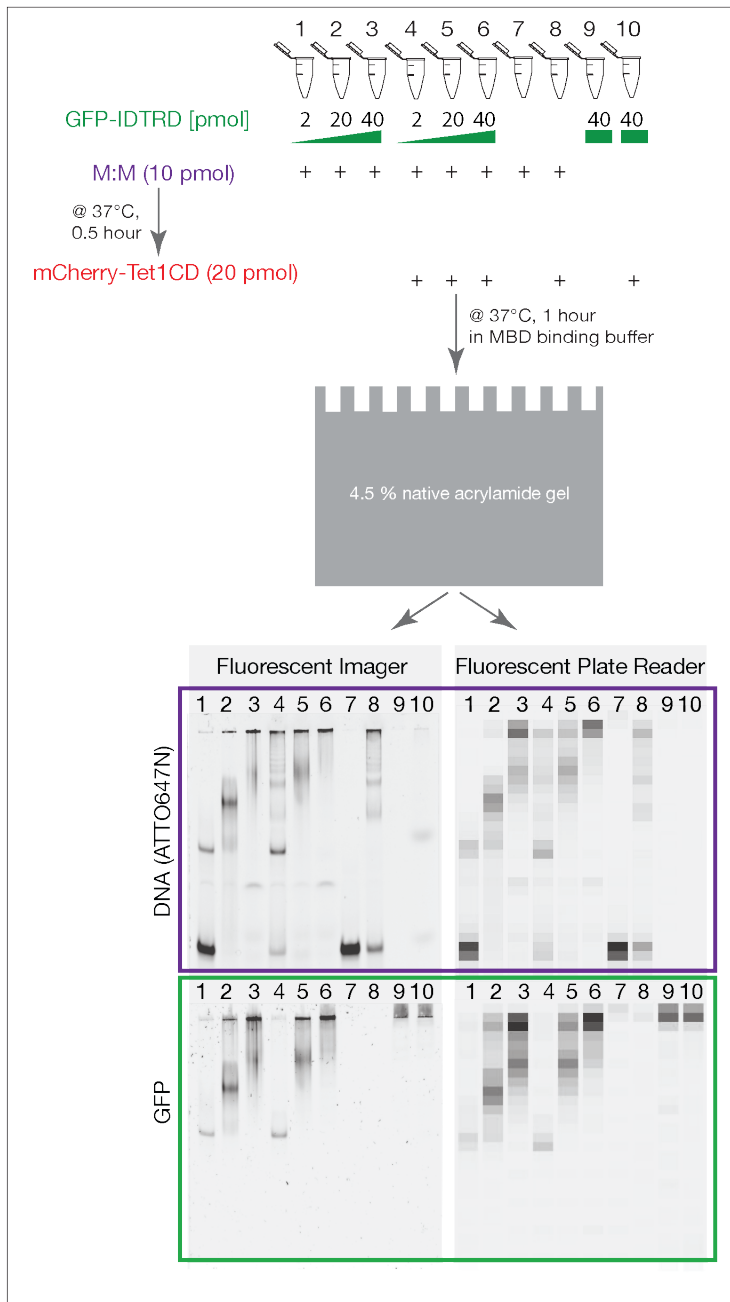
(a) *In situ* staining and quantification of genomic 5hmC levels in transiently transfected C2C12 mouse myoblasts. Shown are relative 5hmC (=cyan) signals as a function of Tet1CD (=magenta) level represented as gradient bars. (a, left) Error bars represent standard deviation. Mean values of two independent 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.



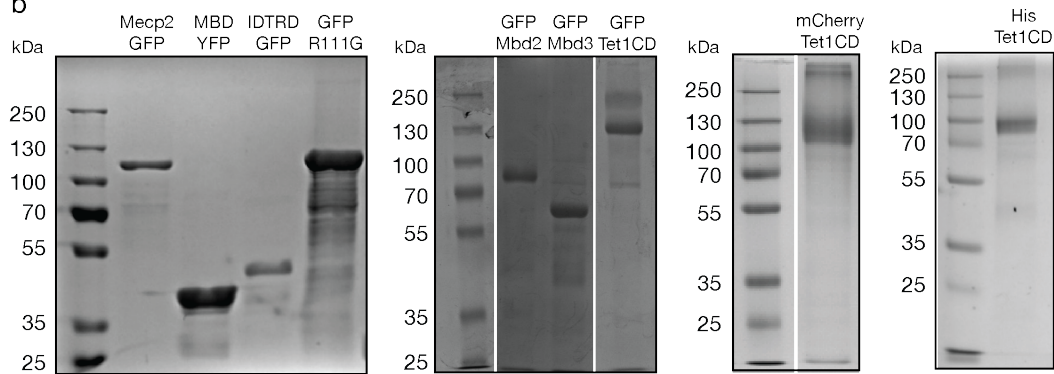
Supplementary Figure 11. Expression levels of Tet1CD and MBD in HEK versus C2C12 cells

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

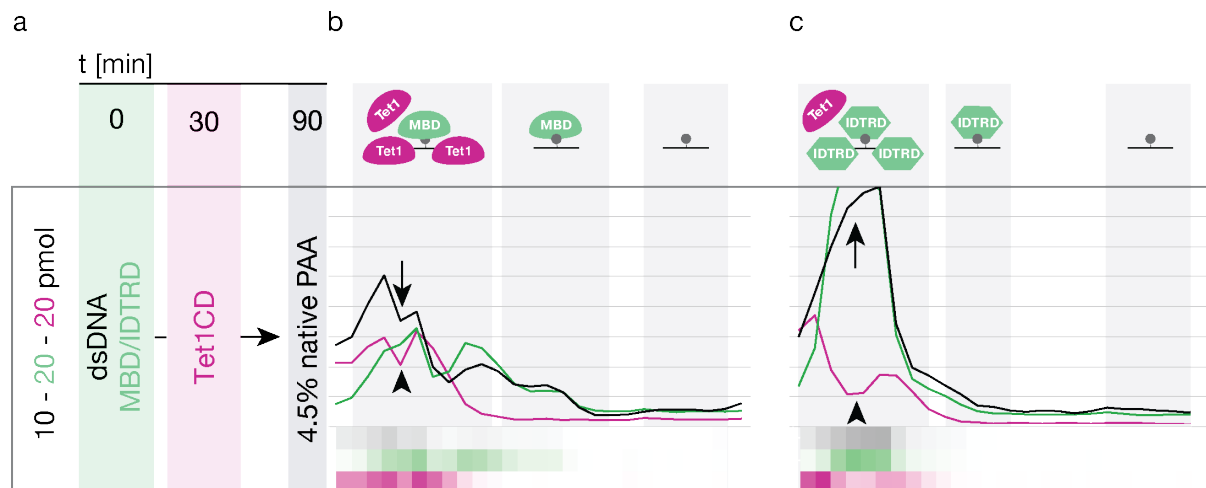


b



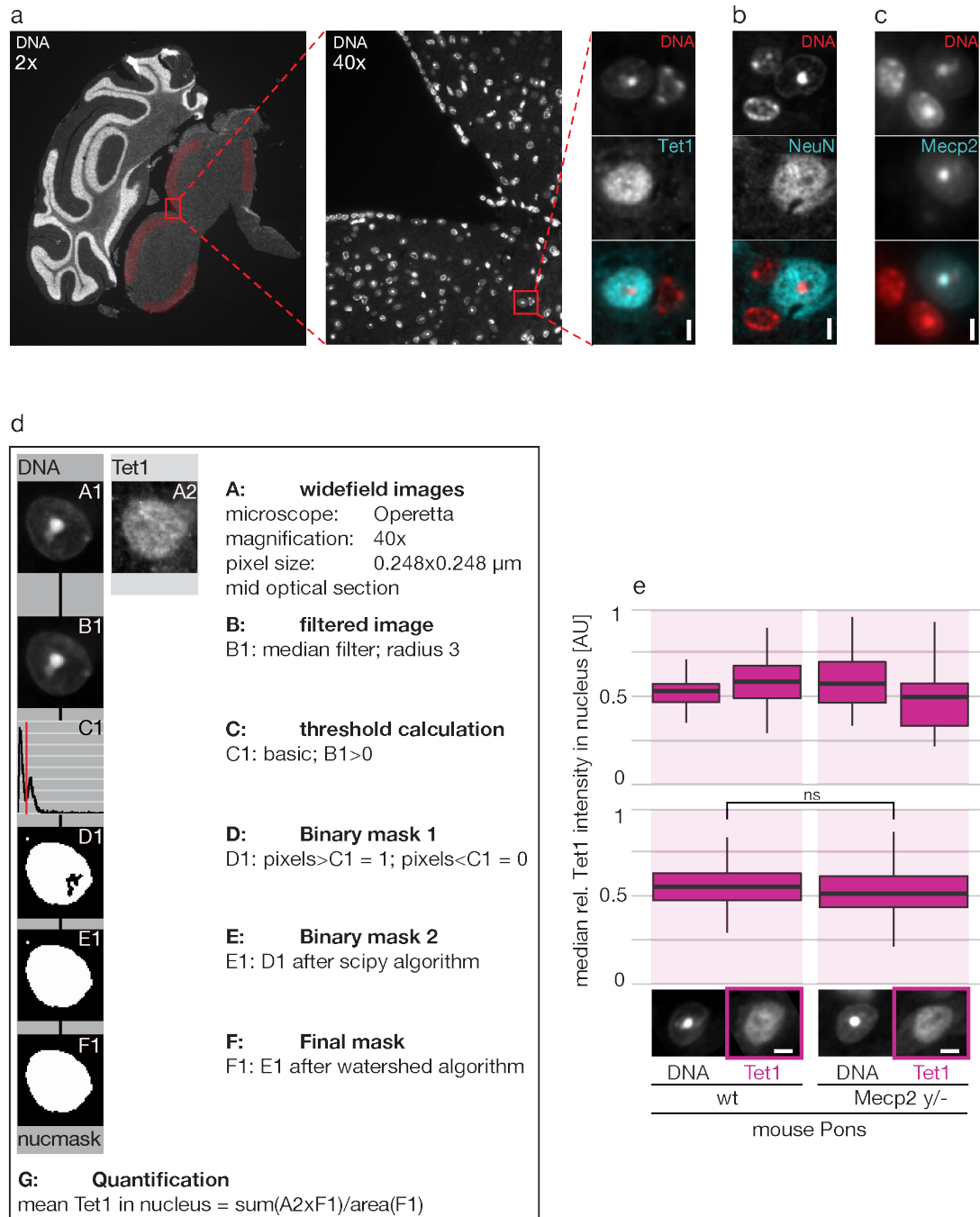
Supplementary Figure 12. Preparation of various controls required for the implementation and interpretation of Electrophoretic Mobility Shift Assays

(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

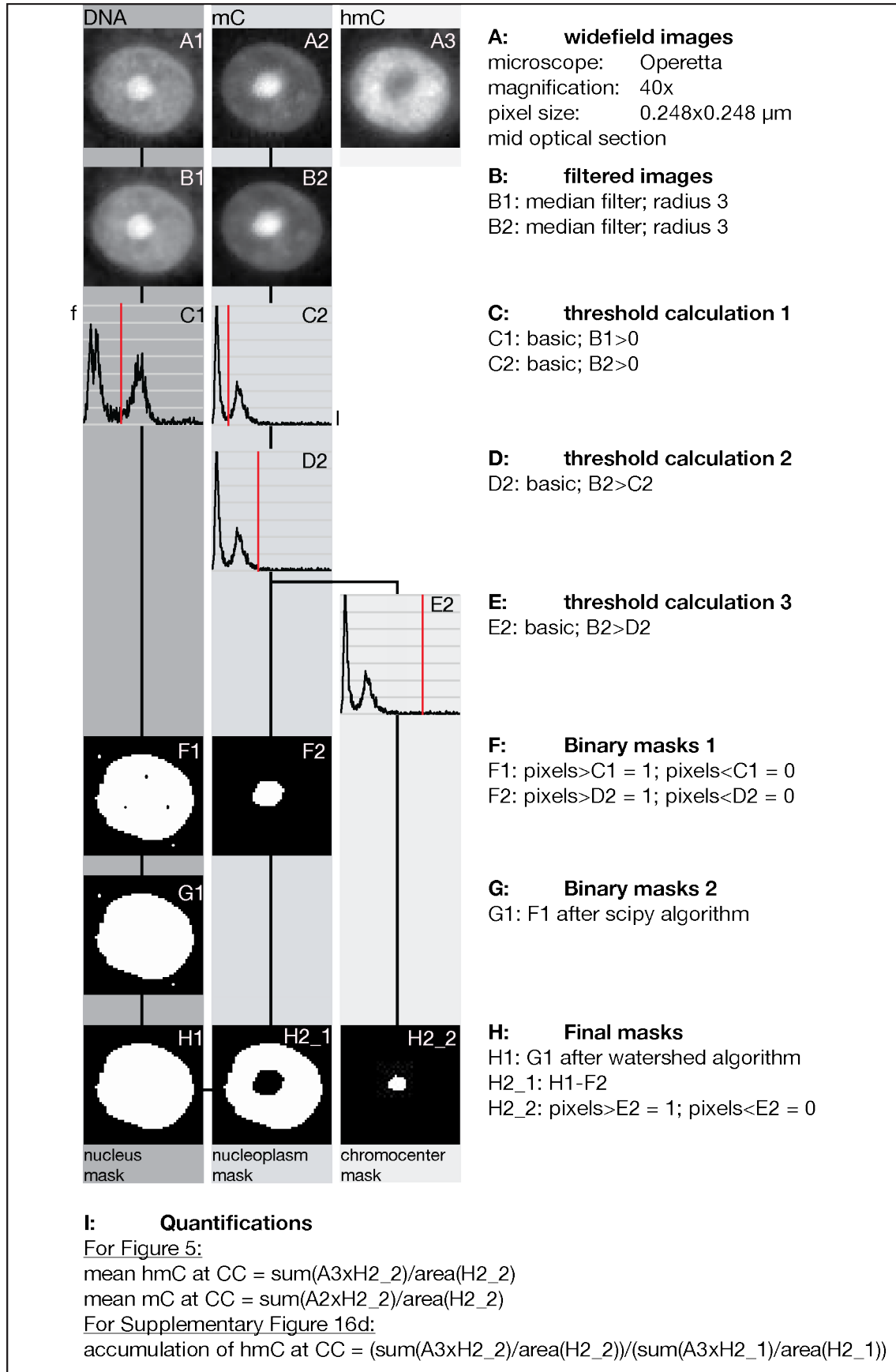
Electrophoretic mobility shift assay (EMSA) to determine the binding ability of fluorescently tagged Tet1CD to double-stranded, single mC containing DNA (ATTO647 labeled) in the presence of equimolar amounts of (b) 5mC specific (fluorescently tagged MBD) and (c) sequence-unspecific (fluorescently tagged 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.



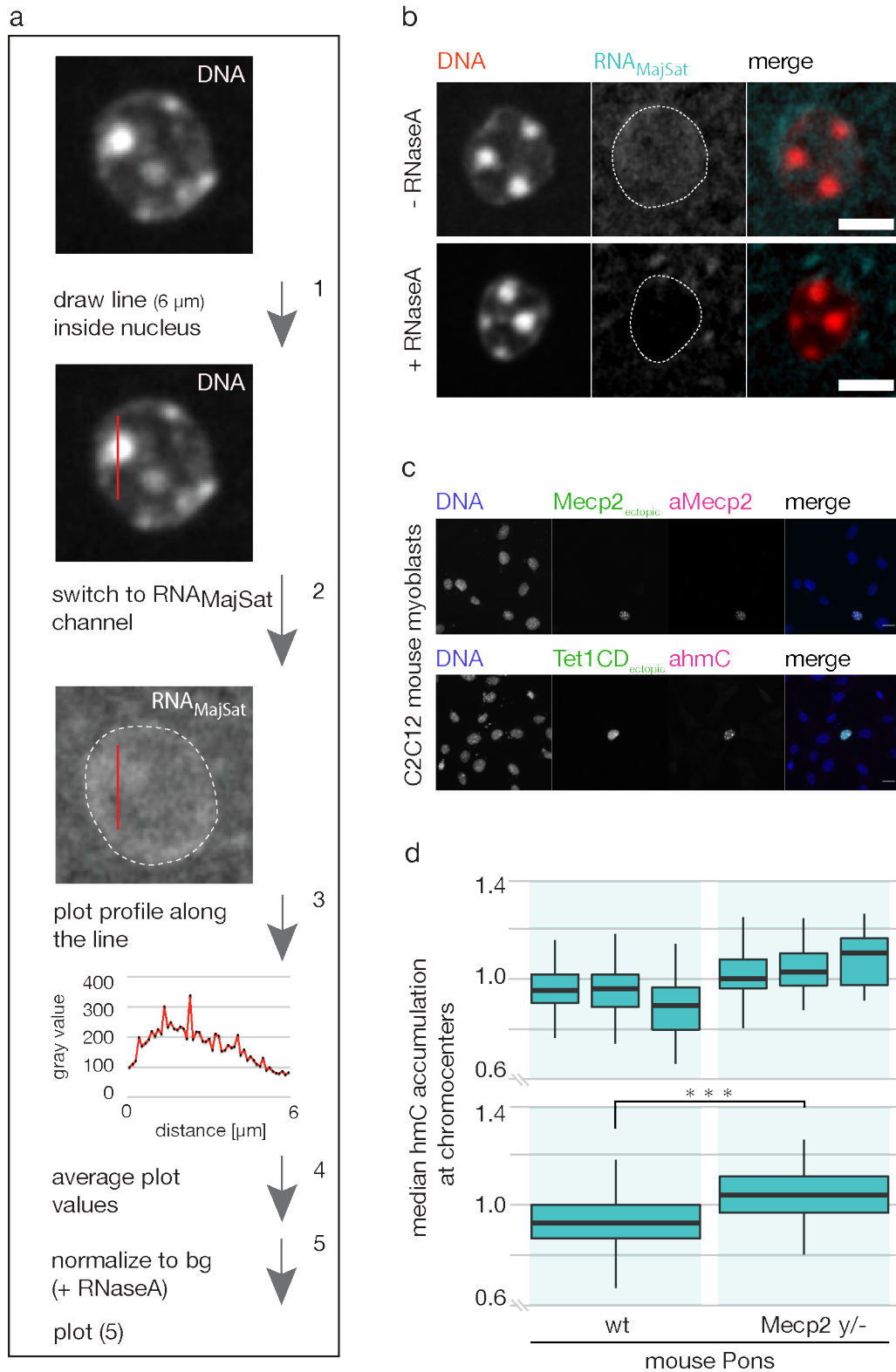
Supplementary Figure 14. Selection of neural cells for the quantification of Tet1, Mecp2, 5mC and 5hmC levels in wild type and Mecp2 knockout mouse pontes. (a) Representative low (2x) and high (40x) magnification scan of a coronal cross section through mouse brain used for quantification of nuclear Tet1 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).

3 DISCUSSION

3.1 Binding of UHRF1 towards different mC- and caC- modification patterns

Methylation of cytosines (mC) in CpG dinucleotides is by far the most abundant modification found in DNA and its role in epigenetic regulation has extensively been studied (Bird, 2002; Greenberg and Bourc'his, 2019; Holliday and Pugh, 1975; Riggs, 1975). However, recent innovations in detection approaches allowed the identification of almost 40 novel DNA modifications (Sood et al., 2019) mainly associated with DNA damage repair (Ito and Kuraoka, 2015). Among these modifications, 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) gained considerable attention and all three modifications are considered to have distinct biological functions (Zhu et al., 2018). Whereas the binding of UHRF1 towards 5mC is well established due to its central task in maintenance methylation (Bostick et al., 2007; Sharif et al., 2007) and some studies also investigated its binding behavior regarding hmC (Frauer et al., 2011), the relationship of UHRF1 and 5fC/5caC is largely unknown. Considering the general flexibility of the entire UHRF1 protein and the adaptability of UHRF1's SRA domain towards hemi-hmC (Fang et al., 2016; Frauer et al., 2011; Gelato et al., 2014), we sought to systematically examine the binding of UHRF1 towards all cytosine modifications, including TET-mediated 5fC and 5caC. In three independent experimental assays (electrophoretic mobility shift assay, microscale thermophoresis and melting temperature analysis), we observed enhanced binding of UHRF1 towards symmetrically carboxylated CpG dyads. To our knowledge, this is the first report demonstrating a binding preference of UHRF1 towards symmetric caC. To better understand the experimentally observed binding preference of UHRF1 towards symmetrically carboxylated cytosine, we performed molecular dynamics (MD) simulations based on the crystal structure of the mouse UHRF1 SRA domain in complex with hemi-methylated DNA (Hashimoto et al., 2008) and computationally integrated the respective cytosine variants (C, mC, caC). We monitored which protein residues are involved in sensing the chemical differences of the cytosine variants and thereby determined different binding modes of the SRA domain towards differentially modified CpG dyads: We observed that systems containing a flipped caC within the binding pocket of UHRF1 (caC-C', caC-caC', caC-mC') possess more heterogeneity in their polar interaction networks than systems with a flipped mC (mC-C', mC-mC', mC-caC'), most likely due to the carboxyl oxygens of caC that allow the formation of additional hydrogen bonds and salt bridges. Interestingly and unexpectedly, we noticed that these carboxyl oxygens form interactions outside of the binding pocket, in particular with R489, a residue of the NKR finger region that is normally involved in sensing the modification on the distal strand, not the one in the binding pocket. While analyzing this NKR finger region more closely, we monitored a remarkable increase in the flexibility of the NKR finger when caC was present at the CpG dyad and therefore explain the observed binding differences as follows: if caC is bound in the binding pocket, the conformational rearrangements within the pocket are propagated to the flexible NKR finger, which ultimately determines the overall binding mode. It either forms stable interactions with the modification on the distal strand, as observed for residue R496 and the carboxyl oxygens in the caC-caC' system, or it encounters steric repulsion that weakens the entire binding, as monitored for caC-C' and caC-mC'.

Due to the extremely low occurrence of 5caC *in vivo* (0.01% of 5mC in mammals, (Zhu et al., 2018)) one could challenge the significance of this modification in general. However, the existence of distinct caC binders argue for a specific role of this modification besides being a demethylation intermediate (Spruijt et al., 2013). Moreover, genome-wide mapping of 5caC indicates that the modification is not randomly distributed, but occurs in specific genomic regions: 5caC is detected in active enhancers as well as in bivalent promoters (Shen et al., 2013) and was shown to overlap with H3K4me1 marks (Shen et al., 2013; Wu et al., 2014b). Notably, profiling methods for DNA modifications at single-base resolution are steadily improved to increase their precision and thus, more in-depth analyses of 5caC distribution will be available in the near future (Zhu et al., 2018). This will probably also shed some light on the distribution of modified cytosines regarding their symmetry. The maintenance of symmetric mC patterns during replication is guaranteed through the activity of DNMT1 and UHRF1 (Bronner et al., 2019). For the other modifications, however, no such maintenance mechanisms are known and it is currently unclear in which combinations they occur at CpG dyads; symmetrically on both strands, only on one strand or in a “hybrid fashion”, i.e. two different modifications at one dyad. As caC is produced through TET-mediated active oxidation on methylated DNA strands (Wu and Zhang, 2017), CpG dyads still harboring mC on one but caC already on the other strand is a physiological state that is likely to occur within a cell. We therefore determined the binding of UHRF1 towards a hybrid oligonucleotide with a central mC-caC’ site and obtained a dissociation constant (KD) comparable to the one for symmetric caC ($0.39 \pm 0.11 \mu\text{M}$ vs. $0.23 \pm 0.05 \mu\text{M}$, respectively). This suggests that UHRF1 preferentially binds caC modifications opposite to mC or caC, but not unmodified C, which was further supported by the polar interaction networks that we obtained in our MD simulations.

Although little is known about the functional role of 5caC due to its low abundance in the genome (Zhu et al., 2018), two recent studies demonstrated that 5caC influences the processivity of RNA polymerase II (Kellinger et al., 2012; Wang et al., 2015a). Acting like “speed bumps”, 5fC and 5caC can transiently slow down the elongation activity of RNA Pol II thereby directly affecting gene expression. The precise deposition of 5fC and 5caC within gene bodies might not constitute a global regulatory mechanism, but rather enables locus-specific fine-tuning of gene expression. Similarly, binding of UHRF1 to 5caC could also represent a mechanism that offers the chance to regulate the expression of specific genes. One could imagine that UHRF1 binding to 5caC blocks the binding of TDG to the respective site, thereby preventing the excision of 5caC and the subsequent incorporation of unmodified cytosine, which is often required for gene (re)-activation (Kohli and Zhang, 2013). UHRF1 at 5caC sites could also act as a scaffold for the recruitment of other epigenetic factors that induce alterations on gene expression. Instead of protecting 5caC from TDG-mediated removal, UHRF1 might also assist in its elimination. Interestingly, isotope tracing in stem cells suggested that direct decarboxylation of 5caC constitutes an alternative in active genome demethylation but the molecular mechanism for this remained elusive (Schiesser et al., 2012). Another group later observed that both, bacterial and mammalian C5-methyltransferases, catalyze the removal of the carboxyl group from 5caC *in vitro* (Liutkevičiūtė et al., 2014). This in turn might offer an explanation for the strong interaction between DNMT1 and 5caC as measured in a SILAC-based mass spectrometry screen in mESCs (Spruijt et al., 2013). Given the fact that DNMT1 performs maintenance methylation in a cooperative tandem with UHRF1 (Bostick et al., 2007; Sharif et al., 2007), it is conceivable that UHRF1 also facilitates potential decarboxylation activities of DNMT1 by recruiting it to the respective target, namely 5caC. Consequently, this would mean that UHRF1/DNMT1 feature some “demethylating capacity”, which is often linked to gene (re)-activation (Greenberg and Bourc’his, 2019), and hence, would represent a contradiction to UHRF1’s

well-known function in regulating repressive chromatin marks (Bronner et al., 2019). In line with such an activating role for UHRF1, it was recently demonstrated that UHRF1 regulates the active histone mark H3K4me3 in bivalent domains during lineage specification in mESCs through an SRA-domain-mediated interaction with the Setd1a/COMPASS complex (Kim et al., 2018). Interestingly, also TET proteins have been shown to interact with the same complex (Deplus et al., 2013) and to safeguard bivalent promoters (Verma et al., 2018). Since TET proteins catalyze the formation of 5caC, it is intriguing to speculate that UHRF1 might sense these 5caC sites at bivalent loci to support the expressional regulation of specific lineage marker during the differentiation of mESCs in a Setd1a/COMPASS-dependent manner.

Moreover, UHRF1 and 5caC have both been implicated in DNA damage repair. As an intermediate in active DNA demethylation, 5caC was early linked to the TDG-BER pathway (Maiti and Drohat, 2011). However, other studies claim that 5fC and 5caC resemble damaged DNA bases with mutagenic properties (Kamiya et al., 2002; Shibutani et al., 2014). 5caC-G pairs were shown to mimic T-G mismatches, which eventually triggers mismatch repair mechanisms (Shibutani et al., 2014). UHRF1 in turn was recently introduced as a DNA lesion recognition factor and presumably acts as a scaffold at DNA damage sites (Tian et al., 2015). Accordingly, UHRF1 binding to 5caC might not necessarily impact gene expression but could also represent a process to protect the genome from damage, thereby maintaining genomic integrity. Admittedly, the potential implications of UHRF1's binding preference to 5caC described above are rather speculative and further research is required to identify its actual *in vivo* function.

Furthermore, we investigated the impact of UHRF1-bound caC on the physiological properties of the DNA helix and its stability. Previous studies already revealed that modified cytosines can alter the structure of the DNA double helix and change its flexibility (Fu et al., 2019; Ngo et al., 2016; Raiber et al., 2015). Within our MD simulations, we discovered that pocket-bound caC affects the width of DNA minor and major groove when compared to bound mC. Especially the minor groove displayed geometric alterations with an opening of the groove of about 1-2 Å at the base pairs preceding the carboxylated cytosine and decreased widths at the base pairs behind. This is in line with another MD simulation where caC particularly induced alterations of minor groove geometry (Fu et al., 2019). In this light, it should be emphasized here that the enhanced binding of the UHRF1-SRA domain towards specific carboxyl cytosine patterns might not only be determined by the modified base per se. 5caC-induced conformational changes in the DNA helix might concomitantly contribute to increased binding by providing a 3D-surface that is more favorable to bind. Additionally, we conducted melting temperature analysis to test the effect of modified cytosines on the thermostability of DNA. Previous studies reported significantly increased thermodynamic stability of DNA double helices in the presence of 5mC (Lercher et al., 2014; Raiber et al., 2015; Thalhammer et al., 2011), whereas the stabilizing effect of mC compared to unmodified C was rather small in our analysis. For DNA oligonucleotides containing a carboxylated CpG site, we observed no substantial difference regarding its melting temperature in comparison to unmodified C, which has been published before (Raiber et al., 2015). Astonishingly however, pre-incubation with the SRA domain of UHRF1 drastically decreased the thermostability of CpG-containing DNA, especially in case of symmetric caC. Another protein well-known for massively destabilizing double stranded DNA helices is the bacteriophage T4 gene 32 protein, albeit no comparability to UHRF1 is given (Alberts and Frey, 1970; Pant et al., 2018). Considering the insertion of a thumb region of UHRF1 into the minor groove and its finger region into the major groove, such a tight binding mode may lead to the destabilization of surrounding base pairs and hence to decreased thermostability of the bound DNA oligonucleotide. The caC-specific geometric alterations of the DNA minor and major

grooves, which are absent in 5mC or 5hmC containing DNA (Renciuk et al., 2013), might further reinforce the destabilization and thus, could explain the stability difference observed for caC- and mC-containing DNA.

In conclusion, our experimental findings together with the performed MD simulation demonstrate that UHRF1 exhibits an unknown binding preference for CpG dyads with specific caC patterns and we further propose that this binding is enabled by the flexible nature of the NKR finger region of UHRF1. Intriguingly, the NKR finger loop was lately reported to mediate specific interactions with DNA that subsequently impact downstream functions of UHRF1 (Vaughan et al., 2019). Moreover, another protein preferentially binding symmetric caC, TET3, was shown to harbor an NRRT sequence (similar to UHRF1's NKRT sequence) with which it recognizes caC' (Jin et al., 2016). Thus, a flexible peptide loop comprising such an amino acid motif might constitute a suitable tool for binding distant carboxylated cytosines. Since the binding preference of UHRF1 towards caC was only determined *in vitro* in this study, the biological function of our finding requires further investigation. However, it already indicates that UHRF1 might have additional functions besides its well-known role in maintenance methylation, presumably in a cell type- or cell cycle-dependent context.

3.2 Two distinct modes of DNMT1 recruitment via UHRF1

DNA methylation is a key player in epigenetic regulation and specific methylation patterns must be maintained once they are established during embryogenesis and gametogenesis as they determine the phenotypic expression of genes in specific cell types, required throughout an individual's lifetime (Bird, 2002). This implies that methylation patterns are maintained within every round of DNA replication, a process coordinated by DNMT1 and UHRF1 (Bostick et al., 2007; Sharif et al., 2007). Both proteins comprise multiple domains, which are jointly involved in regulating maintenance methylation (Bronner et al., 2019). UHRF1 is generally accepted to mediate the recruitment of DNMT1 to chromatin by directly interacting with the protein (Berkyurek et al., 2014) and by specifically binding hemi-methylated DNA, H3K9me3 and unmodified H3R2 via its SRA, TTD and PHD domain, respectively (Arita et al., 2008; Hashimoto et al., 2008; Rajakumara et al., 2011; Rothbart et al., 2012). The RING domain in turn confers E3 ligase activity to UHRF1 and our groups lately demonstrated that the ubiquitylation of lysine residues 14, 18 and 23 on histone H3 constitutes an essential recruiting signal for DNMT1 (Ishiyama et al., 2017; Nishiyama et al., 2013; Qin et al., 2015). DNMT1 itself harbors a PIP-box to interact with PCNA (a DNA sliding clamp essential for DNA replication and repair) (Chuang et al., 1997) and an RFIS domain targeting DNMT1 to replication foci during S-phase (Leonhardt et al., 1992; Margot et al., 2001). Together with the preference of UHRF1's SRA domain for hemi-methylated DNA, which predominantly occurs in the semi-conservative process of DNA replication (Bronner et al., 2019), PCNA-interaction and replication foci targeting strongly suggests that DNMT1-mediated maintenance methylation is coupled to DNA replication. This is further supported by the direct interaction of UHRF1 and LIG1, a ligase essential for DNA replication (Ferry et al., 2017). However, despite deciphering many more functional, mechanistic and structural details about the DNMT1/UHRF1 tandem, some open questions and inconsistencies remain, for instance: DNMT1 was reported to associate with chromatin during G2 and M phase where the replication machinery is no longer active (Easwaran et al., 2004). Furthermore, a recent cell-cycle based proteomic study in mESCs revealed that protein levels of DNMT1 indeed peak in G2/M phase, not in S-phase (Kasvandik 2019). Accordingly, both studies question a DNMT1 recruiting mechanism that is exclusively based on

replication and argue for a more complex scenario of DNMT1 recruitment during cell-cycle progression.

In our current work, we present an updated and two-part model for DNMT1 recruitment including the previously reported UHRF1-mediated ubiquitylation of K14, K18 and/or K23. In fact, we discovered that DNMT1 not only requires ubiquitylated H3 (H3Ub2) but mainly depends on ubiquitylated PAF15 (PAF15Ub2; a PCNA associated factor) for its recruitment to chromatin and its methylation activity. Notably, our group recently identified UHRF1 to be responsible for the ubiquitylation of PAF15 in mESCs (Karg et al., 2017). As the N-terminal parts of PAF15 and H3 comprise highly-similar amino acid sequences (VRTK vs. ARTK), we performed a crystal structure analysis of the PHD domain of hUHRF1 and hPAF15 and observed a comparable binding mode to the one of hPHD and H3 (Rajakumara et al., 2011). Hence, UHRF1 seems to recognize a specific motif via its PHD domain and subsequently dual mono-ubiquitylates PAF15 and/or H3 with a defined spacing of the two ubiquitin moieties. Interestingly, ubiquitylated PAF15 bound chromatin only throughout replication as evidenced by its impaired binding upon inhibition of DNA polymerases. In line with a previous report of Polvsen et al. (2012), chromatin binding of ubiquitylated PAF15 required the interaction with PCNA indicating a close link to the replication machinery. Whereas the interaction of PAF15 and DNMT1 was reported in HEK cells already (Emanuele et al., 2011), we could further reveal that the complex formation is based on dual monoubiquitylation of PAF15 and the recognition of the ubiquitin moieties via the RFTS domain of DNMT1, as it is the case for H3Ub2 binding (Ishiyama et al., 2017; Qin et al., 2015). Strikingly, DNMT1 clearly preferred PAF15Ub2 over H3Ub2 for complex formation suggesting that PAF15Ub2 is the major recruiter for DNMT1. Continulative analysis of this observation uncovered that DNMT1 recruitment through PAF15Ub2 and H3Ub2 is in fact temporally regulated with PAF15 being predominantly ubiquitylated in early S-phase and H3 in late S-phase. Furthermore, S-phase dependent ubiquitylation of UHRF1-targets did not only determine the recruitment of DNMT1 but also appears to differentially impact DNA methylation. First of all, we observed a massive reduction of global DNA methylation levels upon depletion of ubiquitylated PAF15 and subsequently proved that the majority of hypo-methylated regions map to genes being replicated in early S-phase. This in turn greatly strengthens our finding that DNMT1 recruitment through PAF15Ub2 principally occurs in the beginning of S-phase. Interestingly, depletion of PAF15Ub2 strongly enhanced the appearance of H3Ub2 on chromatin, suggesting that H3Ub2 can compensate for diminished PAF15Ub2 levels. Considering the relevance of maintenance methylation in genome integrity and transcriptional regulation (Greenberg and Bourc'his, 2019), establishing compensatory pathways to guarantee DNMT1-mediated propagation of methylation patterns is reasonable. As mentioned earlier, DNMT1 is known to associate with the DNA replication machinery during S-phase and binds heterochromatin in late S- and the G2 phase (Easwaran et al., 2004), which would fit to our proposed model of two different recruiting mechanisms for DNMT1. Whereas the UHRF1/DNMT1 tandem would exclusively rely on factors of the replication machinery in early S-phase, UHRF1 and DNMT1 activity in late S- and presumably G2 phase would rather dependent on H3K9me3 (a prominent mark of heterochromatin) and the ubiquitylation of histone H3 (cp. figure 11). As proposed earlier by our group, this probably implicates that DNMT1 recruitment in early S-phase is mainly dependent on its PCNA-interaction, whereas the binding to H3Ub2 via the RFTS domain becomes more important in late S-phase (Easwaran et al., 2004; Schneider et al., 2013)(Figure 11). Remarkably, the PIP-box-mediated binding of DNMT1 within a cell was shown to be significantly weaker than the RFTS-mediated one (residence times of 10s vs. 22s, respectively, (Schneider et al., 2013)). This indicates that DNMT1 binding during early S-phase is rather transient,

which would match the fast kinetics of the replication machinery (Pray, 2008). Considering that early-replicating euchromatic regions are not that densely packed and readily accessible (Gilbert, 2002; Hansen et al., 2010), rapid progression of the replication machinery together with DNMT1 might still be sufficient to perform the methylation reaction. Late-replicating heterochromatin however, especially constitutive heterochromatic regions containing repetitive elements are known to be heavily methylated and densely packed (Gilbert, 2002; Li and Reinberg, 2011) and thus, DNMT1 might require more time to replenish methylation marks, potentially extending its activity on constitutive heterochromatin into G2 phase. This hypothesis is further supported by FRAP analyses where DNMT1 during early S-phase recovered much faster compared to DNMT1 in late S-phase (35s vs. 86s, (Schermele et al., 2007)). However, taking PAF15Ub2 into account, this would also mean that the RFTS binding to PAF15Ub2 during early S-phase is weaker than the RFTS-mediated binding to H3Ub2 in late S-phase, which awaits further clarification.

Moreover, our study once again illustrates that ubiquitylation seems to be a crucial player of maintenance methylation. Previous studies showed that ubiquitylation in this context is important in two ways: firstly, polyubiquitylation of DNMT1 and UHRF1 regulates the proteins' stability and secondly, mono-ubiquitylated proteins like UHRF1 itself or H3 serve as major recruiting signals for DNMT1 (Foster et al., 2018; Nishiyama et al., 2013; Qin et al., 2011, 2015). Notably, UHRF1-mediated ubiquitylation seems to be tightly controlled as the RING domain of UHRF1 exploits its full ubiquitin ligase activity only if stimulated by its UBL domain and upon binding to hemimethylated DNA (DaRosa et al., 2018; Foster et al., 2018; Harrison et al., 2016). Further supporting the importance of precisely balanced ubiquitylation marks stems from the activity of USP7, a deubiquitinase, reported to be responsible for counteracting UHRF1's ubiquitin ligase activity (Qin et al., 2011; Yamaguchi et al., 2017). For instance, USP7 was shown to remove ubiquitin marks from H3 *in vitro*, probably to facilitate DNMT1's progressive movement along replicating DNA (Yamaguchi et al., 2017) and it was additionally reported to prohibit the polyubiquitin-induced degradation of DNMT1 and UHRF1 (Qin et al., 2011). Another line of evidence for the significance of ubiquitylation in the process of maintenance methylation comes from the observation that DNMT1 binding to ubiquitin is not only essential for its recruitment but also stimulates its methylation activity (Li et al., 2018a). Lastly, K14, K18 and K23 on H3 can be acetylated, which in turn would prohibit the ubiquitylation of these residues (Xie and Qian, 2018). However, DNMT1 is known to form a complex with HDAC1/2, which possibly ensures the removal of potential acetylation moieties and therefore enables the ubiquitylation of the respective sites mandatory for the faithful propagation of methylation marks (Rountree et al., 2000; Xie and Qian, 2018). Collectively, there is accumulating evidence that the ubiquitylation machinery is strongly involved in maintenance methylation and presumably constitutes a mechanism to fine-tune the interaction of certain factors. However, there are still many aspects to be investigated. For example, when H3Ub2 is deubiquitylated by USP7, is this also the case for PAF15Ub2? Does unbound ubiquitin impact methylation and which effect do other UHRF1-ubiquitylated proteins have on maintaining methylation marks (cp. Karg et al., 2017)? Are certain ubiquitin marks controlled in a cell-cycle-dependent fashion to specifically activate pathways required in early-S vs. late-S/G2 phase?

Interestingly, PAF15, UHRF1 and DNMT1 not only play an important role in maintenance methylation but all three factors are also linked to DNA damage repair (Ha et al., 2011; Povlsen et al., 2012; Sidhu and Capalash, 2017; Turchi et al., 2009). PAF15 was early proposed to be critical for the correct function of PCNA during DNA damage repair (Turchi et al., 2009). Later, Povlsen and colleagues (2012) examined this more closely and demonstrated that ubiquitylated PAF15 regulates the access of translesion synthesis (TLS) polymerases to PCNA, which is required to bypass lesions

that would otherwise block replication. DNMT1 is also known to be recruited to sites of DNA damage by interacting with factors of different DNA repair pathways and its recruitment is likely dependent on the interaction with PCNA (Ha et al., 2011; Mortusewicz et al., 2005). UHRF1 in turn was shown to act as a DNA lesion recognition factor (Tian et al., 2015) and to modulate the ubiquitination status of PCNA in response to DNA damage (Hahm et al., 2019). Consequentially, PCNA together with DNMT1 and UHRF1 at DNA damage sites might be required to simply restore epigenetic marks. However, UHRF1-mediated dual mono-ubiquitylation of PAF15 (Karg et al., 2017) and polyubiquitylation of PCNA (Hahm et al., 2019) at DNA lesions might also control translesion synthesis activity to balance the ratio of DNA damage tolerance to actual DNA repair. In summary, we propose a new two-parted, S-phase-dependent recruitment model for DNMT1 that relies on both, ubiquitylated PAF15 and ubiquitylated H3. Remarkably, maintenance methylation appears to utilize two pathways to guarantee its integrity that are mechanistically very similar to each other (Figure 11). In early S-phase, DNMT1 recruitment might exclusively rely on the direct coupling to the replication machinery including PAF15Ub2. In late S-phase however, where densely packed and highly methylated constitutive heterochromatin is replicated, DNMT1 might require more time to access and methylate its substrate and therefore benefits from a second, presumably more stable recruiting pathway; H3Ub2 (Figure 11, (Schneider et al., 2013)). Considering the comparability of these two pathways, this could also be interpreted as two compensatory mechanisms that evolutionary developed to substitute each other in case one pathway is perturbed. This would be in line with studies demonstrating that DNA hypomethylation in UHRF1^{-/-} and DNMT1^{-/-} cell lines can be rescued by H3K9me3-binding deficient UHRF1 and PCNA-binding mutant DNMT1, respectively (Schermelleh et al., 2007; Zhao et al., 2016). Lastly, to ultimately verify that DNMT1 recruitment under normal conditions is split into two pathways, measuring the absolute protein levels of the factors involved, i.e. DNMT1, UHRF1, PAF15Ub2 and H3Ub2, would be highly informative. Knowing the precise distribution of molecules in the course of S-phase could help to assess whether the two mechanism are truly separate or rather merge together.

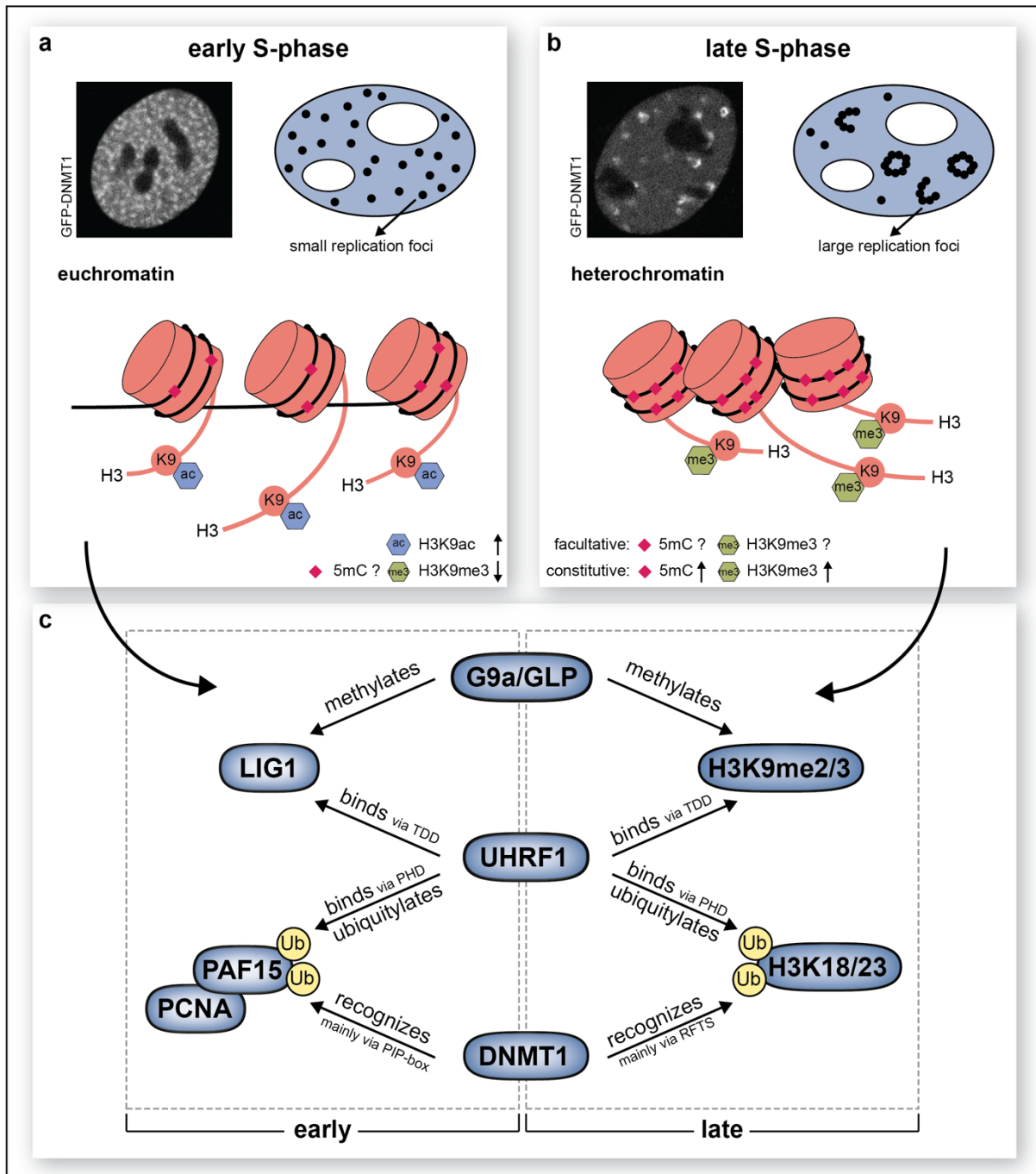


Figure 11: Maintenance methylation in early and late S-phase

(a) Characteristics of early replicating chromatin with high numbers of small replication foci containing DNMT1 (as shown with a confocal image of GFP-tagged DNMT1 in C2C12 myoblasts). During early S-phase, predominantly open euchromatic regions decorated with acetylated H3K9 are replicated. (b) Characteristics of late replicating chromatin with large replication foci containing DNMT1 (as shown with a GFP-tag in (a)). During late S-phase, condensed heterochromatic regions decorated with H3K9me3 and high DNA methylation (in case of constitutive heterochromatin) are replicated. (a) + (b) The levels of 5mC in euchromatin and facultative heterochromatin are currently under debate as indicated with the “?” (cp. Aran et al., 2011; Ball et al., 2009; Suzuki et al., 2011). Confocal images from Schermelleh et al. (2007) (c) Schematics of the factors involved in maintenance methylation. Left side and center depicts factors of the replication machinery, mandatory in early S-phase. Right side and center depicts factors used during late S-phase.

3.3 TET-controlled and DPPA3/STELLA-driven passive demethylation

As elaborated on in previous chapters, a fundamental layer of the epigenetic network is represented by DNA methylation that is dynamically regulated in specific stages of mammalian development (Li and Zhang, 2014). Whereas somatic cells of mammals feature fairly stable and particularly high levels of DNA methylation (approximately 80% of CpG dyads are methylated), DNA methylation patterns of non-somatic cells during development are extensively modified, temporarily or ultimately leading to lower global methylation levels (Li and Zhang, 2014). This includes the methylation levels of pluripotent cells of the preimplantation embryo, primordial germ cells and oocytes (Greenberg and Bourc'his, 2019). One important factor for modulating DNA methylation in these cells, particularly during early embryogenesis and oocyte development, is DPPA3 (also known as PGC7 or Stella) (Han et al., 2019). In fact, DPPA3 was shown to be responsible for the epigenetic asymmetry of the parental genome in zygotes and the protection of methylation marks at imprinted regions (Nakamura et al., 2007). Later, the same group published the underlying mechanism for their initial findings and demonstrated that DPPA3 specifically binds to H3K9me2, predominantly found in the maternal genome, and inhibits TET3-mediated conversion/removal of 5mC (Nakamura et al., 2012). Surprisingly, the protective role of DPPA3 towards methylation seems to be zygote-specific as ensuing studies in somatic cells and oocytes uncovered an opposite function of DPPA3, namely the inhibition of maintenance methylation (Funaki et al., 2014; Li et al., 2018b). However, how these opposing functions of DPPA3 in zygotes and oocytes are regulated remain elusive. Furthermore, despite knowing that DPPA3 expression persists in the preimplantation embryo beyond the zygotic state (Nakamura et al., 2007), the function of DPPA3 prior to implantation has not been explored so far and how DPPA3 expression is generally regulated is only poorly understood (Zhao et al., 2019).

In our present work, we investigated the methylation dynamics of mESCs that have been isolated from the inner cell mass of blastocysts and therefore resemble the low methylation levels of the preimplantation embryo *in vivo* (Ficz et al., 2013). Since numerous studies argued for an important role of TET family enzymes during mouse development (Rasmussen and Helin, 2016; Tan and Shi, 2012), we aimed to dissect the precise involvement of these enzymes in the hypomethylation phenotype of early embryos or naïve mESCs, respectively. As blastocysts and naïve mESCs exhibit high expression levels of TET1 and TET2, but not TET3 (Ito et al., 2010; Koh et al., 2011), we generated CRISPR/Cas-based TET1 and TET2 catalytic mutant (CM) ESC lines (either single or double mutants). To our surprise, catalytic inactivation of TET1 and/or TET2 did indeed lead to genome-wide hypermethylation as expected, but the majority of hypermethylated sites were actually not bound by the enzymes. Interestingly however, one of the few direct targets of TET1/2 was DPPA3, whose expression was repressed upon TET1/2 inactivity as supported by hypermethylation of its promoter region. Previous studies already showed that the *Dppa3* promoter is lowly methylated in naïve ESCs and gains methylation during the differentiation into epiblast-like cells (EpiLCs) with concomitant changes of its expression (Nichols and Smith, 2009; Ying et al., 2008; Zhao et al., 2019). However, the factors responsible for regulating DPPA3 expression in naïve pluripotency are mainly unknown. One protein that is supposedly involved in the regulatory network of DPPA3 expression is MAD2L2 as its deletion in ESCs led to DPPA3 repression and generally caused a reduction of the open chromatin state of naïve ESCs (Rahjouei et al., 2017). To confirm that DPPA3 is a real downstream target of TET1/2, we overexpressed DPPA3 in TET1/2 CMs, which demonstrated a clear rescue effect of DPPA3 towards TET1/2CM-induced hypermethylation. Additionally, the

genome of DPPA3 KO cells featured many hypermethylated sites compared to wildtype cells, which mainly overlapped with the hypermethylated sites of Tet1/2 catalytic mutants. Taken together, our results indicate that DPPA3 safeguards the naïve methylome of ESCs and that DPPA3 expression is controlled by TET1 and TET2 activity.

Notably, the mechanism by which DPPA3 keeps the genome of naïve ESCs hypomethylated seems to be identical to the one described for oocyte maturation, namely the inhibition of the maintenance methylation machinery (Li et al., 2018b). We performed various assays, *in vitro* and *in vivo*, that prove a tight binding of the DPPA3 C-terminus to the PHD domain of UHRF1. We further demonstrate that this high-affinity interaction determines the subcellular localization of UHRF1 as DPPA3 and UHRF1 co-localize to the cytoplasm in wildtype cells, whereas cytoplasmic UHRF1 is absent in DPPA3 KO and TET1/2CM cells. By the cytoplasmic sequestration of UHRF1, its chromatin binding is impeded and thereby the recruitment of DNMT1, which ultimately leads to impaired maintenance methylation (Funaki et al., 2014; Li et al., 2018b). On these grounds, we further deciphered which feature or region of DPPA3 accounts for the inhibition of UHRF1 binding to chromatin. Our results show that the C-terminal part of DPPA3 is mandatory to inhibit UHRF1 binding and maintain genomic hypomethylation and that the nuclear export of UHRF1 into the cytosol is dispensable. DPPA3-mediated hypomethylation in naïve mESCs seems to mainly rely on the nuclear interaction of DPPA3 with the PHD domain of UHRF1 to prevent UHRF1's association with chromatin; a mechanism that was confirmed by another study in somatic cells (Du et al., 2019).

As stated above, it is astonishing how the same protein, DPPA3, performs such different tasks in a narrow time frame ranging from the unfertilized oocyte over the zygote to the preimplantation embryo (Li et al., 2018b; Nakamura et al., 2012). A recent study however reported that the cytosolic fraction of DPPA3 is specifically cleaved in the 2-cell embryo state (Shin et al., 2017). Thus, DPPA3 might be post-translationally “adjusted” to defined developmental time points, for example by cleavage, and this in turn could determine the functionality of the protein, e.g. the protection or inhibition of DNA methylation. How this dynamic regulation of DPPA3 is implemented awaits further research and it is also tempting to ask which factors determine DPPA3 expression in oocytes and zygotes as TET1 and TET2 are not expressed here (Rasmussen and Helin, 2016).

Strikingly, DPPA3 is an evolutionary young player in the regulation of DNA methylation as it is exclusively found in boreoeutherian mammals. Similarly, also the massive DNA demethylation and re-methylation waves observed during preimplantation development are mammal-specific as they do not occur in other vertebrates although UHRF1, DNMTs and TETs are present in these organisms (Almeida et al., 2012; Bogdanović et al., 2016). Thus, we wondered whether mammal-specific DNA hypomethylation could artificially be induced in a non-mammalian organism by simply adding DPPA3. To test this, we used xenopus egg extracts and one-cell embryos of medaka fish as model organisms and introduced purified mouse DPPA3 and *Dppa3* mRNA into the systems, respectively. Intriguingly, replication-dependent methylation of xenopus sperm DNA was abolished due to impaired UHRF1/DNMT1 chromatin binding and DPPA3 in medaka embryos led to intensive genomic hypomethylation and even caused severe developmental defects. Since our methylome data revealed that DPPA3 activity primarily affects DNA repeats, the developmental defects observed in medaka might be a consequence of derepressed transposable elements (TEs) weakening genome stability. However, these defects also indicate that mammals must have developed accompanying factors to the evolution of DPPA3 that enabled them to endure hypomethylated states and derepressed TEs without being harmed. Interestingly, derepression of TEs is also discussed to be functionally integrated into the regulatory networks that drive the

pluripotent states of mammalian development (Cosby et al., 2019). Accordingly, the stage-specific evolution of TET1/2-controlled and DPPA3-mediated passive demethylation might have challenged the stability of the mammalian genome on one hand, but also offered the opportunity to exploit genetic information more holistically and allowed the erasure of epimutations on the other hand.

3.4 Metabolic regulation of TET enzymes

The establishment and regulation of the epigenome is accomplished by a plethora of factors like chromatin-modifying enzymes that alter the chemical properties of both, histone proteins and DNA (Allis and Jenuwein, 2016). To add and remove certain chemical moieties, many of these enzymes require metabolites as co-factors or substrates and hence, are directly influenced by the metabolic state of cells (Harvey et al., 2016). Metabolites that are involved in epigenetic reactions include S-adenosylmethionine (SAM), alpha-ketoglutarate (α -KG), acetyl-coA, nicotinamide adenine dinucleotide (NAD⁺) and uridine diphosphate (UDP) (Etchegaray and Mostoslavsky, 2016). Whereas SAM serves as the universal methyl group donor for histone and DNA methyltransferases, α -KG is an essential co-factor in enzymatic reactions that demethylate histones and DNA, accomplished by jumonji C-domain containing histone demethylases and TET proteins, respectively (Etchegaray and Mostoslavsky, 2016). Acetyl-coA in turn is the key metabolite for the acetylation of histones and NAD⁺ is utilized by sirtuins, a class of histone deacetylases. To O-GlcNAcylate proteins, e.g. histones or TET proteins, the nutrient-sensitive metabolite UDP is required (Etchegaray and Mostoslavsky, 2016). Accordingly, metabolism and epigenetics clearly form a joint axis of cellular regulation. Or to express it more precisely: there is increasing evidence that the epigenome and its chromatin marks are under precise metabolic control, which in turn has causative roles in the development of diseases like cancer (Kinnaird et al., 2016).

In the past, cancer-associated alterations of metabolic and epigenetic mechanisms have intensively been studied, however mainly independent from each other (Janke et al., 2015). Now, as their direct linkage became more apparent, the term “metaboloepigenetics” was introduced (Harvey et al., 2016) and researchers try to understand how cancer cells rewire such metaboloepigenetic regulatory networks to enable carcinogenic growth. Kinnaird and colleagues (2016) recently proposed three potential models of how cancer cells exploit metabolic regulation to modify epigenetic marks. Firstly, by the production of inhibitory metabolites that block the activity of epigenetic enzymes. Secondly, by differently sensing the availability of nutrients, thereby leading to altered expression of responsive genes. Thirdly, by locus-specific recruitment of metabolic enzymes to chromatin sites for the site-specific production of epigenetic substrates and co-factors. One of the best-studied examples for the first model is the neomorphic function of mutant IDH proteins, found in different cancer entities, especially in AML and gliomas (Tommasini-Ghelfi et al., 2019). Due to single amino acid substitutions at arginines within the catalytic core, mutant IDH enzymes produce the oncometabolite 2-HG instead of its natural product α -KG, thereby inhibiting the activity of α -KG-dependent dioxygenases like TET proteins and JmjC-domain histone demethylases (Tommasini-Ghelfi et al., 2019). As demonstrated in our experiments in mESCs, by Wu et al. (2011) in HEK293 cells and by Lu et al. (2012) in NHA cells, 2-HG results in an increase of global 5mC and the reduction of TET-mediated DNA modifications (5hmC, 5fC and 5caC), indicating that the oncometabolite executes its inhibitory function also in healthy cells. We could also show that 2HG, produced endogenously through the activation of inducible IDH1 R132H protein, significantly reduces global 5hmC levels in mESCs, similar to exogenously administered 2-HG. Classically, the

effect of 2-HG has intensively been studied in disease-associated cell and mouse models, collectively showing altered DNA methylation patterns (Figuroa et al., 2010; Losman et al., 2013; Sasaki et al., 2012b). Whereas specific phenotypes like increased white blood cell counts, development of splenomegaly and impaired hematopoietic differentiation have repeatedly been linked to IDH mutations and the production of 2-HG (Chaturvedi et al., 2016; Figuroa et al., 2010; Sasaki et al., 2012b), underlying molecular details await further exploration. In fact, it is still unclear how global alterations of DNA/chromatin modifications lead to a specific disease phenotype and which processes are gradually involved in the evolution of these phenotypes.

Interestingly, our results further showed that the activity of TET proteins in mESCs is not only tunable through 2-HG, but also through the administration of dimethyl- α -KG (DM- α -KG). This illustrates that the availability of this natural co-factor is limited within mESCs and that altered α -KG levels might impact epigenetic regulation by modulating DNA demethylation. This is in line with other studies performed in pluripotent stem cells (PSCs) highlighting α -KG as a crucial and rate-limiting metabolite for demethylation reactions on both, DNA and histones (Carey et al., 2015; TeSlaa et al., 2016; Zhang et al., 2016a). Hence, metabolic regulation of epigenetic networks does not only play decisive roles in carcinogenesis but also in a non-disease context. In fact, many groups studied the impact of energy metabolism on the cellular fate of human and mouse PSCs and revealed that metabolic intermediates affect differentiation and reprogramming to a considerable extent (Harvey et al., 2016). Specific pluripotent stages of ESCs have been linked to distinct patterns of metabolic activity (Sperber et al., 2015; Zhou et al., 2012) and PSCs generally exhibit a high glycolytic flux to support their energy demands during rapid proliferation compared to respirative somatic cells (Zhang et al., 2018) (Figure 11). High rates of glycolysis albeit sufficient oxygen availability is also known from cancer cells to maintain their extreme grow rates (Warburg, 1956). Gardner et al. (1998) could even demonstrate metabolic similarities between blastocysts and cancer cells. Metabolically spoken, cancers seem to recreate an embryonic, pluripotent-like phenotype and employ embryonic pathways for optimal proliferation (Harvey et al., 2016). Thus, trying to understand how metabolism establishes the epigenome and alters gene expression profiles in ESCs to maintain pluripotency or induce differentiation may not only shed light on early embryonic development, but also on the development of cancer.

In this light, we sought to investigate whether 2-HG and α -KG not only modulate TET activity in mESCs but also influence gene expression. Therefore, we performed an RNAseq analysis of mESCs treated with DM- α -KG revealing a strong downregulation of pluripotency marker genes like *Nanog* and *Sox2* upon DM- α -KG administration. This in turn argues for a differentiation-promoting role of α -KG in mESCs. Other studies in mESCs also provide insights on the effect of α -KG levels, however with somewhat contradictory results (Carey et al., 2015; Hwang et al., 2016; TeSlaa et al., 2016; Tischler et al., 2019; Zhang et al., 2016a). Carey et al. propose α -KG to maintain pluripotency in ground state mESCs, Zhang et al. and Hwang et al. show that addition of DM- α -KG impairs the transition of naïve mESCs into EpiLCs and delays the spontaneous differentiation through LIF withdrawal, respectively. Teslaa et al. and Tischler et al. on the contrary claim that α -KG accelerates the initial differentiation of primed mESCs (meaning EpiLCs) indicating that α -KG seems to mediate different effects based on a defined pluripotent state of stem cells. Noteworthy, the variability in these studies elucidating pluripotent metabolism could be subject to differences in the composition of culture medium or might be due to divergent definitions of naïve vs. primed mESCs (Table 1). Furthermore, the aforementioned studies use different approaches to mimic *in vivo* differentiation of cells, ultimately impeding the solid comparison of these results. Hence, to enable the discrimination of true metabolic differences between pluripotency states, culture conditions and

base medium composition need to be maintained within experiments and explicit definitions of pluripotency need to be established. Apart from that, it is challenging to pinpoint which of the α -KG-dependent dioxygenases mediate the effect of α -KG on pluripotency as mammals encode more than 60 of these enzymes (Loenarz and Schofield, 2011). Just because TET activity is influenced by the availability of α -KG does not imply its direct involvement in deregulating gene expression patterns. To further elaborate on this, large-scale knockdown screens of α -KG-dependent dioxygenases could be performed in mESCs cultured with varying α -KG concentrations. Besides α -KG, other metabolic intermediates were also shown to influence the pluripotent state of mESCs. Whereas the amino acids threonine and UDP are required for the self-renewal of pluripotent mESCs (Jang et al., 2012; Wang et al., 2009), L-proline and reduced acetyl-CoA levels are necessary for differentiation of mESCs (Moussaieff et al., 2015; Washington et al., 2010) (Figure 11).

Table 1: Published effects of α -KG and energy metabolism on pluripotency of ESCs

publication	pluripotent ESC culture condition		effect of α -KG
	naïve	primed	
Carey et al. 2015	2i/LIF medium	serum/LIF medium	maintains pluripotency of ground state ESCs
Teslaa et al. 2016	serum/LIF medium	EpiLC medium	induces differentiation of EpiLCs
Zhang et al. 2016a	2i/LIF medium	EpiLC medium	impairs naïve to primed transition
Hwang et al. 2016	2i/LIF medium and serum/LIF medium		delays spontaneous differentiation through LIF withdrawal
Tischler et al. 2019	2i/LIF medium	EpiLC medium	assists transition from EpiLCs into PGCs
Zhou et al. 2012	serum/LIF + 2i medium	EpiLC medium	-
Sperber et al. 2015	serum/LIF + 2i medium	EpiLC medium	-

Studying the cellular effects of IDH mutations predominantly focuses on the neomorphic production of the oncometabolite 2-HG and the inhibition of α -KG-dependent dioxygenases (M. Gagné et al., 2017). However, there is more to be considered when mutant IDH proteins are investigated. Since mutant IDH1 and IDH2 do not generate α -KG anymore, the metabolic homeostasis of subsequent TCA cycle metabolites, e.g. fumarate or malate, is disturbed and the entire TCA cycle possibly downregulated (Reitman et al., 2011). Furthermore, during the forward reaction of wildtype IDH1/2, NADP⁺ is utilized as an electron acceptor and is reduced to NADPH (Lewis et al., 2014). By contrast, mutant IDH1/2 favor the reverse reaction, which consumes NADPH and generates NADP⁺ instead. Thus, the cellular production of NADPH is reduced upon the catalytic activity of mutant IDH proteins, which diminishes the pool of reduced glutathione and concomitantly increases oxidative stress through reactive oxygen species (ROS) (Dang and Su,

2017), as observed in IDH1 mutant glioma cells (Shi et al., 2014). Additionally, glioma tumors with IDH mutations were reported to enhance glutaminolysis to compensate for suppressed glycolytic capacity (Ohka et al., 2014). To enable the investigation of cellular functions of mutant IDH proteins besides the production of 2-HG, we generated stable mESC lines with inducible IDHmut cassettes. Aside from that, also IDHwt cassettes were integrated into wildtype mESCs to circumvent exogenous administration of esterified α -KG. As verified in a LC-MS/MS analysis for mutant and wildtype IDH1 cell lines, dox-induced overexpression of the enzymes successfully produced the respective metabolite (2-HG or α -KG). Surprisingly, the endogenous “over-production” of α -KG and 2-HG showed consistent effects on the activity of TET proteins when IDH1 constructs (wt or R132H) were induced, but not for the overexpression of IDH2 constructs (wt, R140R or R172K). This might be explained by the cellular localization of the two isoforms and the stage of mitochondrial maturation in mESCs. In contrast to elongated mitochondria in terminally differentiated cells with a filamentous network of inner membrane cristae folds, mitochondria in mESCs appear to be round and swollen, lacking ultrastructural features (Folmes et al., 2011) (Figure 11). Hence, oxidative phosphorylation (OXPHOS) in mESCs might be deviant to the one in somatic cells and α -KG production might rather rely on cytosolic IDH1 than mitochondrial IDH2. Our results however would argue against a recent publication in which IDH2-mediated production of α -KG was linked to the metabolic regulation of naïve pluripotency and PGC differentiation (Tischler et al., 2019). As noted before, studies performed to elucidate the impact of metabolic mechanisms on shifting cell identities or maintaining pluripotency delivered variable results so far (Carey et al., 2015; Hwang et al., 2016; TeSlaa et al., 2016; Tischler et al., 2019; Zhang et al., 2016a) and further work is necessary to conclusively explain small but critical metabolic differences in cell fate decisions. Nevertheless, there is increasing evidence that metabolic reactions influence the epigenetic landscape, therefore modulating gene expression patterns and the differentiation potential, in both, PSCs and cancer cells (Teslaa and Teitell, 2015). However, which factors and signaling pathways are exactly involved in the metaboloepigenetic regulation of cells and how shifts in cellular metabolism and accompanying cell fates can be reversed, e.g. in mutant IDH tumors, requires further research.

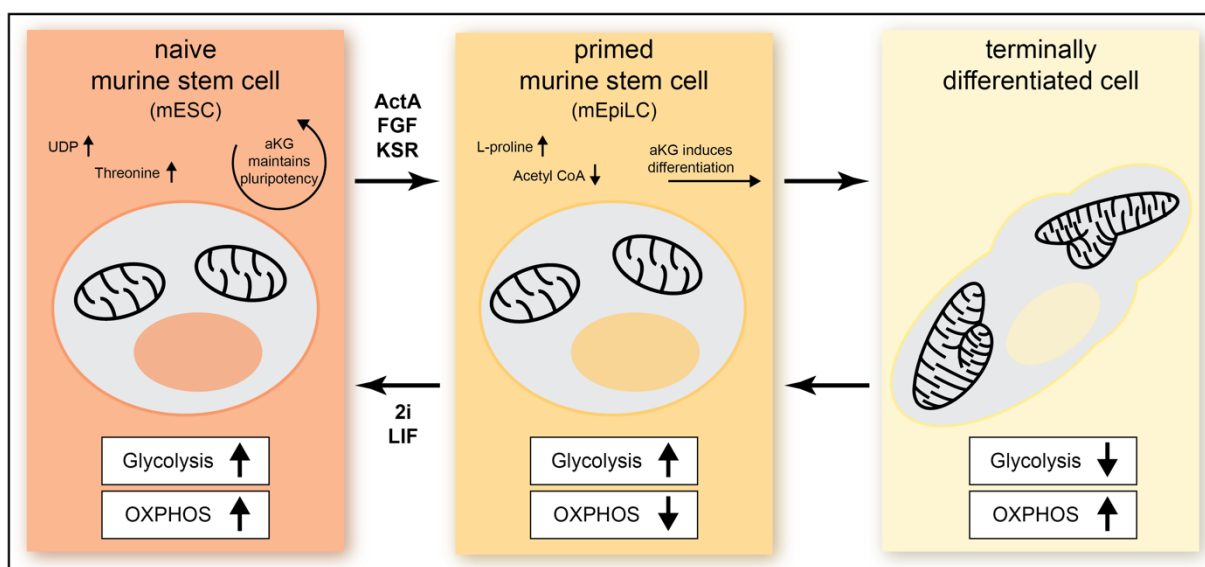


Figure 12: Influence of energy metabolism on pluripotency

Naïve mESCs exhibit a bivalent energy metabolism, whereas primed mESCs are mainly glycolytic. Terminally differentiated cells in turn exclusively use OXPHOS. Important metabolites for the two pluripotent states are indicated in the figure. UDP: uridine diphosphate; α -KG: alpha-ketoglutarate; OXPHOS: oxidative phosphorylation, adapted from Teslaa and Teitell (2015).

3.5 Binding of MBD proteins to DNA blocks TET1 function

As discussed in previous chapters, establishing and maintaining appropriate DNA methylation patterns is indispensable for the development of organisms and the prevention of cancer formation. However, aberrant methylation patterns or impaired recognition of these patterns is also observed in other diseases like the neurological disorder Rett syndrome (Bienvenu and Chelly, 2006). On a molecular level, Rett syndrome is caused by mutations in MECP2, the most prominent member of the methyl-CpG-binding domain (MBD) protein family (Amir et al., 1999). MECP2 preferentially binds to densely methylated regions, like pericentric chromatin in mice, and induces chromatin reorganization and gene repression through the recruitment of various epigenetic factors (Agarwal et al., 2011; Jones et al., 1998; Nan et al., 1996). Besides the transcriptional silencing of specific targets genes, mutational analysis of MECP2 in Rett mouse models revealed that MECP2 represses the activity of retrotransposons in a 5mC-dependent manner (Muotri et al., 2010; Yu et al., 2001). As methylation of the retrotransposon promoter region was reported to reduce transcriptional activity (Muotri et al., 2010), we asked whether active demethylation might account for the opposite behavior (enhanced transcription) and whether MECP2 modulates this demethylation activity. And indeed, we could clearly prove that the turnover of 5mC marks is based on a cross-regulation between mC binders like MECP2 and the 5mC modifier TET1. In fact, we show that not only MECP2 but also MBD2 protects 5mC from TET1-mediated oxidation *in vitro* as well as *in vivo*. Sequence-unspecific binding of the two MBD proteins to DNA effectively restricted the hydrophobic interaction of TET1 with DNA, therefore prohibiting the subsequent steps of oxidation. Interestingly, MECP2 was recently reported to have distinct kinetic features in neuronal cells with very limited nuclear diffusion, presumably to reliably control local chromatin structures and gene expression (Piccolo et al., 2019). Accordingly, a study analyzing Rett-associated MECP2 mutants showed that the majority of mutant proteins failed to cluster heterochromatin and exhibited shorter residence times on chromatin compared to wildtype (Agarwal et al., 2011). Thus, reduced dwell times of MECP2 on DNA might allow TET proteins to gain unrestricted access to their substrate and TET-mediated DNA demethylation is no longer prevented. This is in line with increased 5hmC levels that we observed in neurons of a Rett mouse model. Furthermore, in the absence of MECP2, we could show that ectopically expressed TET1 leads to the expressional reactivation of major satellite transcripts and inversely, these transcripts were downregulated in TET triple knockout mESCs. Taken together, these results indicate that TET1-mediated DNA demethylation causes the expression of repetitive elements when MECP2 protection is missing and suggests that TET1 might also be responsible for uncontrolled LINE-1 (L1) retrotransposition observed in Rett models (Muotri et al., 2010; Yu et al., 2001). In fact, major satellites and transposable elements (TEs) both represent repetitive DNA sequences (de Koning et al., 2011), but the direct association of TET1 and TE derepression was not unambiguously identified. Later, our groups could show that L1 retrotransposition is indeed activated by TET1 and repressed by MBD proteins as verified endogenously and with engineered L1 reporter constructs (Zhang et al., 2017b). These findings are in agreement with an hMeDIP-seq performed in mESCs where depletion of TET1 and TET2 resulted in the loss of 5hmC marks on the 5' region of the L1 element (Ficz et al., 2011), presumably downregulating L1 expression.

Commonly, the evolution of DNA methylation is often considered to be a consequence of the compelling need to maintain deleterious TEs in a repressive state (Yoder et al., 1997). Likewise, the removal of 5mC through the activity of TET enzymes was postulated to reverse silencing of transposons (Gerdes et al., 2016). However, the role of TET enzymes in regulating TE activity

emerges to be more complex than initially thought (Deniz et al., 2019). While various groups report TET activity to be a driver of retrotransposition as for instance evident in TET knockout cells exhibiting decreased IAP expression (Deniz et al., 2018; Kang et al., 2015; Zhang et al., 2017b) and as suggested by our findings regarding major satellite reactivation, de la Rica and colleagues (2016) published some unexpected results about TET-mediated regulation of TEs in ESCs. While the authors show that TET enzymes indeed demethylate LINE-1 elements, they also reveal that TET enzymes recruit the co-repressor SIN3A to ensure LINE-1 silencing. Hence, TET enzymes appear to couple active DNA demethylation to a non-anticipated repressive mechanism (Deniz et al., 2019). This in turn indicates that L1 expression in ESCs is most likely regulated by multiple and 5mC-independent layers as also proposed by Kang et al. (2015) and it further suggests that demethylating TET activity might not be uniformly distributed across all retrotransposon types but is likely restricted to a few of them.

As our study only focused on the effects of TET1 and its potential implication in Rett syndrome associated alterations, it is worth to note that TET2 and TET3 are equally important in mammalian neurogenesis (Wu et al., 2018) and all three TET family members are expressed in brain tissues (Hahn et al., 2013). Accordingly, the three TET isoforms have been reported to be functionally redundant in neurogenesis (Wu et al., 2018) and the protective function of MECP2 and MBD2 towards TET1 oxidation might also apply to TET2 and TET3 *in vivo*. In summary, we show that MBD family proteins take up an important position in protecting the genome from unrestricted active DNA demethylation through TET proteins and propose that the disturbed interplay between these factors promote, at least partially, the development of diseases like the Rett syndrome.

4 APPENDIX

4.1 References

- 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. *Hum. Mol. Genet.* *20*, 4187–4195.
- Agoston, A.T., Argani, P., Yegnasubramanian, S., De Marzo, A.M., Ansari-Lari, M.A., Hicks, J.L., Davidson, N.E., and Nelson, W.G. (2005). Increased protein stability causes DNA methyltransferase 1 dysregulation in breast cancer. *J. Biol. Chem.* *280*, 18302–18310.
- Alabert, C., Barth, T.K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O.N., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev.* *29*, 585–590.
- Alberts, B.M., and Frey, L. (1970). T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. *Nature* *227*, 1313–1318.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Genetic information in eucaryotes (Garland Science).
- Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* *51*, 786–794.
- Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* *17*, 487–500.
- Almeida, R.D., Loose, M., Sottile, V., Matsa, E., Denning, C., Young, L., Johnson, A.D., Gering, M., and Ruzov, A. (2012). 5-hydroxymethyl-cytosine enrichment of non-committed cells is not a universal feature of vertebrate development. *Epigenetics* *7*, 383–389.
- 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.
- Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G.D., Giaever, G., and Nislow, C. (2012). Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* *1*, e00078.
- Amor, D.J., Kalitsis, P., Sumer, H., and Choo, K.H.A. (2004). Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol.* *14*, 359–368.
- Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D’Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* *18*, 225–233.
- Aran, D., Toperoff, G., Rosenberg, M., and Hellman, A. (2011). Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* *20*, 670–680.
- Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 10148–10152.
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemimethylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* *455*, 818–821.

- Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biology* *15*, 545.
- Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J.M. (2002). HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J.* *21*, 6832–6841.
- Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* *12*, 429–442.
- Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* *455*, 822–825.
- Baets, J., Duan, X., Wu, Y., Smith, G., Seeley, W.W., Mademan, I., McGrath, N.M., Beadell, N.C., Khoury, J., Botuyan, M.-V., et al. (2015). Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. *Brain* *138*, 845–861.
- Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* *27*, 361–368.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* *21*, 381–395.
- Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* *354*, 909–912.
- Bashtrykov, P., and Jeltsch, A. (2015). DNMT1-associated DNA methylation changes in cancer. *Cell Cycle* *14*, 5.
- Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* *153*, 480–492.
- Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* *520*, 243–247.
- Bauer, C., Göbel, K., Nagaraj, N., Colantuoni, C., Wang, M., Müller, 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). *Journal of Biological Chemistry* *290*, 4801–4812.
- Becker, P.B., and Workman, J.L. (2013). Nucleosome remodeling and epigenetics. *Cold Spring Harb. Perspect. Biol.* *5*.
- Berkyurek, A.C., Suetake, I., Arita, K., Takeshita, K., Nakagawa, A., Shirakawa, M., and Tajima, S. (2014). The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. *J. Biol. Chem.* *289*, 379–386.
- Bienvenu, T., and Chelly, J. (2006). Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nat. Rev. Genet.* *7*, 415–426.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* *16*, 6–21.
- Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a

blastocyst-like state in ES cells. *Nature* 500, 222–226.

Bogdanović, O., Smits, A.H., de la Calle Mustienes, E., Tena, J.J., Ford, E., Williams, R., Senanayake, U., Schultz, M.D., Hontelez, S., van Kruijsbergen, I., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat. Genet.* 48, 417–426.

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

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

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., et al. (2018). Ten things you should know about transposable elements. *Genome Biol.* 19, 199.

Bowman, G.D., and McKnight, J.N. (2017). Sequence-specific targeting of chromatin remodelers organizes precisely positioned nucleosomes throughout the genome. *Bioessays* 39, 1–8.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311–322.

Brockdorff, N., and Turner, B.M. (2015). Dosage compensation in mammals. *Cold Spring Harb. Perspect. Biol.* 7, a019406.

Bronner, C., Achour, M., Arima, Y., Chataigneau, T., Saya, H., and Schini-Kerth, V.B. (2007). The UHRF family: Oncogenes that are drugable targets for cancer therapy in the near future? *Pharmacology & Therapeutics* 115, 419–434.

Bronner, C., Alhosin, M., Hamiche, A., and Mousli, M. (2019). Coordinated dialogue between UHRF1 and DNMT1 to ensure faithful inheritance of methylated DNA patterns. *Genes* 10.

Burgess, R.J., and Zhang, Z. (2013). Histone chaperones in nucleosome assembly and human disease. *Nat. Struct. Mol. Biol.* 20, 14–22.

Cairns, B.R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. *Nat. Struct. Mol. Biol.* 14, 989–996.

Carey, B.W., Finley, L.W.S., Cross, J.R., Allis, C.D., and Thompson, C.B. (2015). Intracellular α -ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* 518, 413–416.

Chadwick, B.P., Valley, C.M., and Willard, H.F. (2001). Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome. *Nucleic Acids Res.* 29, 2699–2705.

Chan, S.W.-L., and -L. Chan, S.W. (2004). RNA silencing genes control de novo DNA methylation. *Science* 303, 1336–1336.

Chan, S.M., Thomas, D., Corces-Zimmerman, M.R., Xavy, S., Rastogi, S., Hong, W.-J., Zhao, F., Medeiros, B.C., Tyvoll, D.A., and Majeti, R. (2015). Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia. *Nat. Med.* 21, 178–184.

Chaturvedi, A., Araujo Cruz, M.M., Jyotsana, N., Sharma, A., Goparaju, R., Schwarzer, A., Görlich, K., Schottmann, R., Struys, E.A., Jansen, E.E., et al. (2016). Enantiomer-specific and paracrine leukemogenicity of mutant IDH metabolite 2-hydroxyglutarate. *Leukemia* 30, 1708–1715.

Cheng, J., Guo, S., Chen, S., Mastriano, S.J., Liu, C., D'Alessio, A.C., Hysolli, E., Guo, Y., Yao, H., Megyola, C.M., et al. (2013). An extensive network of TET2-targeting MicroRNAs regulates malignant

hematopoiesis. *Cell Rep.* *5*, 471–481.

Chowdhury, R., Yeoh, K.K., Tian, Y.-M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K.H., Li, X.S., Woon, E.C.Y., Yang, M., et al. (2011). The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* *12*, 463–469.

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

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* *277*, 1996–2000.

Cosby, R.L., Chang, N.-C., and Feschotte, C. (2019). Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* *33*, 1098–1116.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* *495*, 370–374.

Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Developmental Biology* *278*, 274–288.

Dang, L., and Su, S.-S.M. (2017). Isocitrate Dehydrogenase Mutation and (R)-2-Hydroxyglutarate: From Basic Discovery to Therapeutics Development. *Annu. Rev. Biochem.* *86*, 305–331.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* *462*, 739–744.

DaRosa, P.A., Harrison, J.S., Zelter, A., Davis, T.N., Brzovic, P., Kuhlman, B., and Klevit, R.E. (2018). A Bifunctional Role for the UHRF1 UBL Domain in the Control of Hemi-methylated DNA-Dependent Histone Ubiquitylation. *Molecular Cell* *72*, 753–765.e6.

De Biasio, A., de Opakua, A.I., Mortuza, G.B., Molina, R., Cordeiro, T.N., Castillo, F., Villate, M., Merino, N., Delgado, S., Gil-Cartón, D., et al. (2015). Structure of p15(PAF)-PCNA complex and implications for clamp sliding during DNA replication and repair. *Nat. Commun.* *6*, 6439.

Denis, H., Ndlovu, M.N., and Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Rep.* *12*, 647–656.

Deniz, Ö., de la Rica, L., Cheng, K.C.L., Spensberger, D., and Branco, M.R. (2018). SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. *Genome Biol.* *19*, 6.

Deniz, Ö., Frost, J.M., and Branco, M.R. (2019). Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* *20*, 417–431.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* *32*, 645–655.

Deplus, R., Blanchon, L., Rajavelu, A., Boukaba, A., Defrance, M., Luciani, J., Rothé, F., Dedeurwaerder, S., Denis, H., Brinkman, A.B., et al. (2014). Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. *Cell Rep.* *8*, 743–753.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* *285*, 26114–26120.

Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores

- distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* *41*, 1350–1353.
- Du, W., Dong, Q., Zhang, Z., Liu, B., Zhou, T., and Xu, R. (2019). Stella protein facilitates DNA demethylation by disrupting the chromatin association of the RING finger-type E3 ubiquitin ligase UHRF1. *The Journal of Biological Chemistry* *294*, 8907–8917.
- Easwaran, H.P., Schermelleh, L., Leonhardt, H., and Cristina Cardoso, M. (2004). Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Reports* *5*, 1181–1186.
- 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.
- Emanuele, M.J., Ciccìa, A., Elia, A.E.H., and Elledge, S.J. (2011). Proliferating cell nuclear antigen (PCNA)-associated KIAA0101/PAF15 protein is a cell cycle-regulated anaphase-promoting complex/cyclosome substrate. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 9845–9850.
- Esteve, P.O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* *20*, 3089–3103.
- Etchegaray, J.-P., and Mostoslavsky, R. (2016). Interplay between Metabolism and Epigenetics: A Nuclear Adaptation to Environmental Changes. *Mol. Cell* *62*, 695–711.
- Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. *Curr. Biol.* *11*, 1183–1187.
- Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* *13*, 397–406.
- Fang, J., Cheng, J., Wang, J., Zhang, Q., Liu, M., Gong, R., Wang, P., Zhang, X., Feng, Y., Lan, W., et al. (2016). Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. *Nature Communications* *7*, 11197.
- Fatemi, M., Hermann, A., Pradhan, S., and Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J. Mol. Biol.* *309*, 1189–1199.
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 8689–8694.
- Feng, Y., Li, X., Cassady, K., Zou, Z., and Zhang, X. (2019). TET2 Function in Hematopoietic Malignancies, Immune Regulation, and DNA Repair. *Front. Oncol.* *9*, 210.
- Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* *12*, 565–575.
- Ferry, L., Fournier, A., Tsusaka, T., Adelmant, G., Shimazu, T., Matano, S., Kirsh, O., Amouroux, R., Dohmae, N., Suzuki, T., et al. (2017). Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation. *Molecular Cell* *67*, 550–565.e5.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Joana Marques, C., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during

differentiation. *Nature* 473, 398–402.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* 13, 351–359.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567.

Filipescu, D., Szenker, E., and Almouzni, G. (2013). Developmental roles of histone H3 variants and their chaperones. *Trends Genet.* 29, 630–640.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.

Flaus, A., Martin, D.M.A., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res.* 34, 2887–2905.

Flores, K.B., Wolschin, F., and Amdam, G.V. (2013). The role of methylation of DNA in environmental adaptation. *Integr. Comp. Biol.* 53, 359–372.

Folmes, C.D.L., Nelson, T.J., Martinez-Fernandez, A., Arrell, D.K., Lindor, J.Z., Dzeja, P.P., Ikeda, Y., Perez-Terzic, C., and Terzic, A. (2011). Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab.* 14, 264–271.

Foster, B.M., Stolz, P., Mulholland, C.B., Montoya, A., Kramer, H., Bultmann, S., and Bartke, T. (2018). Critical Role of the UBL Domain in Stimulating the E3 Ubiquitin Ligase Activity of UHRF1 toward Chromatin. *Molecular Cell* 72, 739–752.e9.

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

Fu, T., Liu, L., Yang, Q.-L., Wang, Y., Xu, P., Zhang, L., Liu, S., Dai, Q., Ji, Q., Xu, G.-L., et al. (2019). Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chem. Sci.* 10, 7407–7417.

Funaki, S., Nakamura, T., Nakatani, T., Umehara, H., Nakashima, H., and Nakano, T. (2014). Inhibition of maintenance DNA methylation by Stella. *Biochem. Biophys. Res. Commun.* 453, 455–460.

Gagliardi, M., Strazzullo, M., and Matarazzo, M.R. (2018). DNMT3B Functions: Novel Insights From Human Disease. *Front Cell Dev Biol* 6, 140.

Gaidzik, V.I., Paschka, P., Späth, D., Habdank, M., Köhne, C.-H., Germing, U., von Lilienfeld-Toal, M., Held, G., Horst, H.-A., Haase, D., et al. (2012). TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. *J. Clin. Oncol.* 30, 1350–1357.

Gangaraju, V.K., and Bartholomew, B. (2007). Mechanisms of ATP dependent chromatin remodeling. *Mutat. Res.* 618, 3–17.

Gardner, D.K. (1998). Changes in requirements and utilization of nutrients during mammalian preimplantation embryo development and their significance in embryo culture. *Theriogenology* 49, 83–102.

Gelato, K.A., Tauber, M., Ong, M.S., Winter, S., Hiragami-Hamada, K., Sindlinger, J., Lemak, A., Bultsma, Y., Houlston, S., Schwarzer, D., et al. (2014). Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. *Mol. Cell* 54, 905–919.

Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D.,

- Montana, G., Taylor, S., et al. (2012). Smcnd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* 23, 265–279.
- Gerdes, P., Richardson, S.R., and Faulkner, G.J. (2016). TET enzymes: double agents in the transposable element-host genome conflict. *Genome Biol.* 17, 259.
- Gaiimo, B.D., Ferrante, F., Herchenröther, A., Hake, S.B., and Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* 12, 37.
- Gilbert, D.M. (2002). Replication timing and transcriptional control: beyond cause and effect. *Curr. Opin. Cell Biol.* 14, 377–383.
- Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398.
- Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* 20, 590–607.
- Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics* 13, 343–357.
- Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.-G., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610.
- Guan, D., Factor, D., Liu, Y., Wang, Z., and Kao, H.-Y. (2013). The epigenetic regulator UHRF1 promotes ubiquitination-mediated degradation of the tumor-suppressor protein promyelocytic leukemia protein. *Oncogene* 32, 3819–3828.
- Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.* 22, 633–641.
- Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., et al. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* 15, 447–459.
- Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517, 640–644.
- Ha, K., Lee, G.E., Pali, S.S., Brown, K.D., Takeda, Y., Liu, K., Bhalla, K.N., and Robertson, K.D. (2011). Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Hum. Mol. Genet.* 20, 126–140.
- Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339, 448–452.
- Hahm, J.Y., Kim, J.-Y., Park, J.W., Kang, J.-Y., Kim, K.-B., Kim, S.-R., Cho, H., and Seo, S.-B. (2019). Methylation of UHRF1 by SET7 is essential for DNA double-strand break repair. *Nucleic Acids Res.* 47, 184–196.
- Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.-G., Jiang, Y., Pfeifer, G.P., et al. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep.* 3, 291–300.
- Han, L., Ren, C., Zhang, J., Shu, W., and Wang, Q. (2019). Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development. *Cell Discov* 5, 9.
- Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O.,

- Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 139–144.
- Harrison, J.S., Cornett, E.M., Goldfarb, D., DaRosa, P.A., Li, Z.M., Yan, F., Dickson, B.M., Guo, A.H., Cantu, D.V., Kaustov, L., et al. (2016). Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. *eLife* *5*.
- Harvey, A.J., Rathjen, J., and Gardner, D.K. (2016). Metaboloepigenetic Regulation of Pluripotent Stem Cells. *Stem Cells Int.* *2016*, 1816525.
- Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* *455*, 826–829.
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* *40*, 4841–4849.
- Hayashi, K., de Sousa Lopes, S.M.C., and Surani, M.A. (2007). Germ cell specification in mice. *Science* *316*, 394–396.
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* *333*, 1303–1307.
- Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D., and Spector, D.L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* *107*, 727–738.
- 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.
- Henikoff, S., and Smith, M.M. (2015). Histone variants and epigenetics. *Cold Spring Harb. Perspect. Biol.* *7*, a019364.
- Heyn, P., Logan, C.V., Fluteau, A., Challis, R.C., Auchynnikava, T., Martin, C.-A., Marsh, J.A., Tagliani, F., Kilanowski, F., Parry, D.A., et al. (2019). Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat. Genet.* *51*, 96–105.
- Holliday, R. (1990). Mechanisms for the control of gene activity during development. *Biol. Rev. Camb. Philos. Soc.* *65*, 431–471.
- Holliday, R., and Pugh, J. (1975). DNA modification mechanisms and gene activity during development. *Science* *187*, 226–232.
- Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* *175*, 315–332.
- Hwang, I.-Y., Kwak, S., Lee, S., Kim, H., Lee, S.E., Kim, J.-H., Kim, Y.A., Jeon, Y.K., Chung, D.H., Jin, X., et al. (2016). Psat1-Dependent Fluctuations in α -Ketoglutarate Affect the Timing of ESC Differentiation. *Cell Metab.* *24*, 494–501.
- Inoue, S., Lemonnier, F., and Mak, T.W. (2016). Roles of IDH1/2 and TET2 mutations in myeloid disorders. *Int. J. Hematol.* *103*, 627–633.
- Ishiyama, S., Nishiyama, A., Saeki, Y., Moritsugu, K., Morimoto, D., Yamaguchi, L., Arai, N., Matsumura, R., Kawakami, T., Mishima, Y., et al. (2017). Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. *Mol. Cell* *68*, 350–360.e7.
- Ito, S., and Kuraoka, I. (2015). Epigenetic modifications in DNA could mimic oxidative DNA damage: A

double-edged sword. *DNA Repair* 32, 52–57.

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.

Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8, 1698–1710.

Jack, A.P.M., Bussemer, S., Hahn, M., and Pünzeler, S. (2013). H3K56me3 is a novel, conserved heterochromatic mark that largely but not completely overlaps with H3K9me3 in both regulation and localization. *PLoS One* 8, e51765.

Jang, H., Kim, T.W., Yoon, S., Choi, S.-Y., Kang, T.-W., Kim, S.-Y., Kwon, Y.-W., Cho, E.-J., and Youn, H.-D. (2012). O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. *Cell Stem Cell* 11, 62–74.

Janke, R., Dodson, A.E., and Rine, J. (2015). Metabolism and epigenetics. *Annu. Rev. Cell Dev. Biol.* 31, 473–496.

Jazirehi, A.R., Arle, D., and Wenn, P.B. (2012). UHRF1: a master regulator in prostate cancer. *Epigenomics* 4, 251–252.

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledest, F., Fischer, A., and Viegas-Pequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum. Mol. Genet.* 2, 731–735.

Jeffery, L., and Nakielny, S. (2004). Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J. Biol. Chem.* 279, 49479–49487.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074–80.

Ji, D., Lin, K., Song, J., and Wang, Y. (2014). Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Molecular BioSystems* 10, 1749.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 449, 248–251.

Jin, B., Tao, Q., Peng, J., Soo, H.M., Wu, W., Ying, J., Fields, C.R., Delmas, A.L., Liu, X., Qiu, J., et al. (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum. Mol. Genet.* 17, 690–709.

Jin, S.-G., Zhang, Z.-M., Dunwell, T.L., Harter, M.R., Wu, X., Johnson, J., Li, Z., Liu, J., Szabó, P.E., Lu, Q., et al. (2016). Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep.* 14, 493–505.

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

Jones, P.A., Issa, J.P., and Baylin, S. (2016). Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* 17, 630–641.

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.

- Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.* *32*, 340–353.
- Kamiya, H., Tsuchiya, H., Karino, N., Ueno, Y., Matsuda, A., and Harashima, H. (2002). Mutagenicity of 5-formylcytosine, an oxidation product of 5-methylcytosine, in DNA in mammalian cells. *J. Biochem.* *132*, 551–555.
- Kanai, Y., Ushijima, S., Nakanishi, Y., Sakamoto, M., and Hirohashi, S. (2003). Mutation of the DNA methyltransferase (DNMT) 1 gene in human colorectal cancers. *Cancer Lett.* *192*, 75–82.
- Kang, E.S., Park, C.W., and Chung, J.H. (2001). Dnmt3b, de novo DNA methyltransferase, interacts with SUMO-1 and Ubc9 through its N-terminal region and is subject to modification by SUMO-1. *Biochem. Biophys. Res. Commun.* *289*, 862–868.
- Kang, J., Lienhard, M., Pastor, W.A., Chawla, A., Novotny, M., Tsagaratou, A., Lasken, R.S., Thompson, E.C., Surani, M.A., Koralov, S.B., et al. (2015). Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* *112*, E4236–E4245.
- Kanwal, R., and Gupta, S. (2012). Epigenetic modifications in cancer. *Clin. Genet.* *81*, 303–311.
- Kapitonov, V.V., and Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* *9*, 411–412; author reply 414.
- Karg, E., Smets, M., Ryan, J., Forné, I., Qin, W., Mulholland, C.B., Kalideris, G., Imhof, A., Bultmann, S., and Leonhardt, H. (2017). Ubiquitome Analysis Reveals PCNA-Associated Factor 15 (PAF15) as a Specific Ubiquitination Target of UHRF1 in Embryonic Stem Cells. *J. Mol. Biol.* *429*, 3814–3824.
- Kats, L.M., Reschke, M., Taulli, R., Pozdnyakova, O., Burgess, K., Bhargava, P., Straley, K., Karnik, R., Meissner, A., Small, D., et al. (2014). Proto-oncogenic role of mutant IDH2 in leukemia initiation and maintenance. *Cell Stem Cell* *14*, 329–341.
- Kellinger, M.W., Song, C.-X., Chong, J., Lu, X.-Y., He, C., and Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* *19*, 831–833.
- Kim, K.-Y., Tanaka, Y., Su, J., Cakir, B., Xiang, Y., Patterson, B., Ding, J., Jung, Y.-W., Kim, J.-H., Hysolli, E., et al. (2018). Uhrf1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells through Setd1a. *Nat. Commun.* *9*, 2583.
- Kinnaird, A., Zhao, S., Wellen, K.E., and Michelakis, E.D. (2016). Metabolic control of epigenetics in cancer. *Nat. Rev. Cancer* *16*, 694–707.
- Klein, C.J., Botuyan, M.-V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C., et al. (2011). Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* *43*, 595–600.
- Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell* *76*, 357–369.
- Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* *468*, 839–843.
- Ko, M., An, J., Bandukwala, H.S., Chavez, L., Åijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* *497*, 122–126.
- Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer,

- C.A., Mostoslavsky, G., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 8, 200–213.
- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479.
- Koivunen, P., Lee, S., Duncan, C.G., Lopez, G., Lu, G., Ramkissoon, S., Losman, J.A., Joensuu, P., Bergmann, U., Gross, S., et al. (2012). Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* 483, 484–488.
- de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7, e1002384.
- Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. *Science* 184, 865–868.
- Kosmider, O., Gelsi-Boyer, V., Ciudad, M., Racœur, C., Jooste, V., Vey, N., Quesnel, B., Fenaux, P., Bastie, J.-N., Beyne-Rauzy, O., et al. (2009). TET2 gene mutation is a frequent and adverse event in chronic myelomonocytic leukemia. *Haematologica* 94, 1676–1681.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.
- Koziol, M.J., Bradshaw, C.R., Allen, G.E., Costa, A.S.H., 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.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Kumar, S., Chinnusamy, V., and Mohapatra, T. (2018). Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* 9, 640.
- Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131, 839–849.
- Langemeijer, S.M.C., Kuiper, R.P., Berends, M., Knops, R., Aslanyan, M.G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A.G., Raymakers, R.A.P., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* 41, 838–842.
- Laukka, T., Mariani, C.J., Ihantola, T., Cao, J.Z., Hokkanen, J., Kaelin, W.G., Jr, Godley, L.A., and Koivunen, P. (2016). Fumarate and Succinate Regulate Expression of Hypoxia-inducible Genes via TET Enzymes. *J. Biol. Chem.* 291, 4256–4265.
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics* 32, 42–56.
- 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.
- Lercher, L., McDonough, M.A., El-Sagheer, A.H., Thalhammer, A., Kriaucionis, S., Brown, T., and Schofield, C.J. (2014). Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* 50, 1794–1796.
- Lewis, C.A., Parker, S.J., Fiske, B.P., McCloskey, D., Gui, D.Y., Green, C.R., Vokes, N.I., Feist, A.M., Vander Heiden, M.G., and Metallo, C.M. (2014). Tracing compartmentalized NADPH metabolism in the cytosol and mitochondria of mammalian cells. *Mol. Cell* 55, 253–263.
- 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.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* 363, 2424–2433.

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.* 6, a019133.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. *Curr. Opin. Genet. Dev.* 21, 175–186.

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

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442, 91–95.

Li, T., Wang, L., Du, Y., Xie, S., Yang, X., Lian, F., Zhou, Z., and Qian, C. (2018a). Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic Acids Res.* 46, 3218–3231.

Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., et al. (2018b). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* 564, 136–140.

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

Ling, Y., Sankpal, U.T., Robertson, A.K., McNally, J.G., Karpova, T., and Robertson, K.D. (2004). Modification of de novo DNA methyltransferase 3a (Dnmt3a) by SUMO-1 modulates its interaction with histone deacetylases (HDACs) and its capacity to repress transcription. *Nucleic Acids Res.* 32, 598–610.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322.

Liutkevičiūtė, Z., Kriukienė, E., Ličytė, J., Rudytė, M., Urbanavičiūtė, G., and Klimašauskas, S. (2014). Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J. Am. Chem. Soc.* 136, 5884–5887.

Loenarz, C., and Schofield, C.J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends in Biochemical Sciences* 36, 7–18.

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.

Losman, J.-A., and Kaelin, W.G., Jr (2013). What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev.* 27, 836–852.

Losman, J.-A., Looper, R.E., Koivunen, P., Lee, S., Schneider, R.K., McMahon, C., Cowley, G.S., Root, D.E., Ebert, B.L., and Kaelin, W.G., Jr (2013). (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* 339, 1621–1625.

Lowndes, N.F., and Toh, G.W.-L. (2005). DNA repair: the importance of phosphorylating histone H2AX. *Curr. Biol.* 15, R99–R102.

- Lu, C., Ward, P.S., Kapoor, G.S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C.R., Khanin, R., Figueroa, M.E., Melnick, A., et al. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* *483*, 474–478.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251–260.
- Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* *19*, 81–92.
- 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.
- 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.
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* *137*, 522–535.
- Margot, J.B., Cardoso, M.C., and Leonhardt, H. (2001). Mammalian DNA methyltransferases show different subnuclear distributions. *J. Cell. Biochem.* *83*, 373–379.
- Mariani, C.J., Vasanthakumar, A., Madzo, J., Yesilkamal, A., Bhagat, T., Yu, Y., Bhattacharyya, S., Wenger, R.H., Cohn, S.L., Nanduri, J., et al. (2014). TET1-mediated hydroxymethylation facilitates hypoxic gene induction in neuroblastoma. *Cell Rep.* *7*, 1343–1352.
- Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. *Genomics* *80*, 487–498.
- McGhee, J.D., and Ginder, G.D. (1979). Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature* *280*, 419–420.
- Meilinger, D., Fellingner, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E.F., Spada, F., and Leonhardt, H. (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 Reports* *10*, 1259–1264.
- Messerschmidt, D.M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science* *335*, 1499–1502.
- Messerschmidt, D.M., Knowles, B.B., and Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* *28*, 812–828.
- von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N.Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., et al. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Mol. Cell* *62*, 983.
- M. Gagné, L., Boulay, K., Topisirovic, I., Huot, M.-É., and Mallette, F.A. (2017). Oncogenic Activities of IDH1/2 Mutations: From Epigenetics to Cellular Signaling. *Trends Cell Biol.* *27*, 738–752.
- Minor, E.A., Court, B.L., Young, J.I., and Wang, G. (2013). Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J. Biol. Chem.* *288*, 13669–13674.
- Misaki, T., Yamaguchi, L., Sun, J., Orii, M., Nishiyama, A., and Nakanishi, M. (2016). The replication foci

- targeting sequence (RFTS) of DNMT1 functions as a potent histone H3 binding domain regulated by autoinhibition. *Biochem. Biophys. Res. Commun.* *470*, 741–747.
- Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* *110*, 689–699.
- Montalban-Bravo, G., and DiNardo, C.D. (2018). The role of IDH mutations in acute myeloid leukemia. *Future Oncol.* *14*, 979–993.
- Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* *20*, 11–24.
- Morrison, A.J., and Shen, X. (2005). DNA repair in the context of chromatin. *Cell Cycle* *4*, 568–571.
- Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M.C., and Leonhardt, H. (2005). Recruitment of DNA methyltransferase I to DNA repair sites. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 8905–8909.
- Mousli, M., Hopfner, R., Abbady, A.-Q., Monté, D., Jeanblanc, M., Oudet, P., Louis, B., and Bronner, C. (2003). ICBP90 belongs to a new family of proteins with an expression that is deregulated in cancer cells. *Br. J. Cancer* *89*, 120–127.
- Moussaieff, A., Rouleau, M., Kitsberg, D., Cohen, M., Levy, G., Barasch, D., Nemirovski, A., Shen-Orr, S., Laevsky, I., Amit, M., et al. (2015). Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab.* *21*, 392–402.
- Müller, 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 Research* *42*, 8592–8604.
- Mund, C., Hackanson, B., Stresemann, C., Lübbert, M., and Lyko, F. (2005). Characterization of DNA demethylation effects induced by 5-aza-2'-deoxycytidine in patients with myelodysplastic syndrome. *Cancer Research* *65*, 7086–7090.
- Muotri, A.R., Marchetto, M.C.N., 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.
- Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., et al. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biology* *9*, 64–71.
- Nakamura, T., Liu, Y.-J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* *486*, 415–419.
- Nan, X., Tate, P., Li, E., and Bird, A. (1996). DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.* *16*, 414–421.
- Nan, X., Javier Campoy, F., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* *88*, 471–481.
- Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* *393*, 386–389.
- Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Dal Bello, F., and Oliviero, S. (2015). Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep.* *10*, 674–683.

- Ng, S.S., Yue, W.W., Oppermann, U., and Klose, R.J. (2009). Dynamic protein methylation in chromatin biology. *Cell. Mol. Life Sci.* *66*, 407–22.
- Ngo, T.T.M., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A., and Ha, T. (2016). Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* *7*, 10813.
- Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* *4*, 487–492.
- Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F., et al. (2013). Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* *502*, 249–253.
- Ohka, F., Ito, M., Ranjit, M., Senga, T., Motomura, A., Motomura, K., Saito, K., Kato, K., Kato, Y., Wakabayashi, T., et al. (2014). Quantitative metabolome analysis profiles activation of glutaminolysis in glioma with IDH1 mutation. *Tumour Biol.* *35*, 5911–5920.
- 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.
- Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). *Science* *183*, 330–332.
- O'Neill, M.J. (2005). The influence of non-coding RNAs on allele-specific gene expression in mammals. *Hum. Mol. Genet.* *14 Spec No 1*, R113–R120.
- Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* *448*, 714–717.
- Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Rep.* *10*, 1235–1241.
- Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? *Front. Genet.* *6*, 2.
- Palmer, D.K., O'Day, K., and Margolis, R.L. (1990). The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. *Chromosoma* *100*, 32–36.
- Pant, K., Anderson, B., Perdana, H., Malinowski, M.A., Win, A.T., Pabst, C., Williams, M.C., and Karpel, R.L. (2018). The role of the C-domain of bacteriophage T4 gene 32 protein in ssDNA binding and dsDNA helix-destabilization: Kinetic, single-molecule, and cross-linking studies. *PLoS One* *13*, e0194357.
- Payer, L.M., and Burns, K.H. (2019). Transposable elements in human genetic disease. *Nat. Rev. Genet.* *20*, 760–772.
- Pearson, H. (2006). Genetics: what is a gene? *Nature* *441*, 398–401.
- Peat, J.R., Dean, W., Clark, S.J., Krueger, F., Smallwood, S.A., Ficz, G., Kim, J.K., Marioni, J.C., Hore, T.A., and Reik, W. (2014). Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. *Cell Rep.* *9*, 1990–2000.
- Pfister, S.X., and Ashworth, A. (2017). Marked for death: targeting epigenetic changes in cancer. *Nat. Rev. Drug Discov.* *16*, 241–263.
- Piccolo, F.M., Liu, Z., Dong, P., Hsu, C.-L., Stoyanova, E.I., Rao, A., Tjian, R., and Heintz, N. (2019). MeCP2 nuclear dynamics in live neurons results from low and high affinity chromatin interactions. *Elife* *8*, e51449.
- Pichler, G., Wolf, P., Schmidt, C.S., Meilinger, D., Schneider, K., Frauer, C., Fellingner, K., Rottach, A., and

- Leonhardt, H. (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *Journal of Cellular Biochemistry* *112*, 2585–2593.
- Pikaard, C.S., and Mittelsten Scheid, O. (2014). Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.* *6*, a019315.
- Piunti, A., and Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* *352*, aad9780.
- Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* *28*, 1057–1068.
- Povlsen, L.K., Beli, P., Wagner, S.A., Poulsen, S.L., Sylvestersen, K.B., Poulsen, J.W., Nielsen, M.L., Bekker-Jensen, S., Mailand, N., and Choudhary, C. (2012). Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nat. Cell Biol.* *14*, 1089–1098.
- Pray, L. (2008). Major molecular events of DNA replication. *Nature Education* *1*, 99.
- Qi, Y., Wang, D., Wang, D., Jin, T., Yang, L., Wu, H., Li, Y., Zhao, J., Du, F., Song, M., et al. (2016). HEDD: the human epigenetic drug database. *Database* *2016*.
- Qin, W., Leonhardt, H., and Spada, F. (2011). Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *Journal of Cellular Biochemistry* *112*, 439–444.
- Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., La Mastra, F., Forné, I., Pichler, G., Hörl, D., Fellinger, K., et al. (2015). DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res.* *25*, 911–929.
- Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* *44*, 361–372.
- Rahjouei, A., Pirouz, M., Di Virgilio, M., Kamin, D., and Kessel, M. (2017). MAD2L2 Promotes Open Chromatin in Embryonic Stem Cells and Derepresses the Dppa3 Locus. *Stem Cell Reports* *8*, 813–821.
- Raiber, E.A., Murat, P., Chirgadzé, D.Y., Beraldi, D., Luisi, B.F., and Balasubramanian, S. (2015). 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* *22*, 44–49.
- Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo, R., Wu, F., Li, H., Lan, F., et al. (2011). PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Molecular Cell* *43*, 275–284.
- Rampal, R., Alkalin, A., Madzo, J., Vasanthakumar, A., Pronier, E., Patel, J., and Others (2014). DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell Rep* *9* (5): 1841--1855 .
- Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* *30*, 733–750.
- Rasmussen, K.D., Jia, G., Johansen, J.V., Pedersen, M.T., Rapin, N., Bagger, F.O., Porse, B.T., Bernard, O.A., Christensen, J., and Helin, K. (2015). Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes & Development* *29*, 910–922.
- Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. *Science* *210*, 604–610.
- Reik, A., Schütz, G., and Stewart, A.F. (1991). Glucocorticoids are required for establishment and maintenance of an alteration in chromatin structure: induction leads to a reversible disruption of

nucleosomes over an enhancer. *EMBO J.* *10*, 2569–2576.

Reitman, Z.J., Jin, G., Karoly, E.D., Spasojevic, I., Yang, J., Kinzler, K.W., He, Y., Bigner, D.D., Vogelstein, B., and Yan, H. (2011). Profiling the effects of isocitrate dehydrogenase 1 and 2 mutations on the cellular metabolome. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 3270–3275.

Renciuk, D., Blacque, O., Vorlickova, M., and Spingler, B. (2013). Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.* *41*, 9891–9900.

de la Rica, L., Deniz, Ö., Cheng, K.C.L., Todd, C.D., Cruz, C., Houseley, J., and Branco, M.R. (2016). TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.* *17*, 234.

Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenetic and Genome Research* *14*, 9–25.

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

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.

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

Rzem, R., Vincent, M.-F., Van Schaftingen, E., and Veiga-da-Cunha, M. (2007). L-2-hydroxyglutaric aciduria, a defect of metabolite repair. *J. Inherit. Metab. Dis.* *30*, 681–689.

Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., and Torres-Padilla, M.-E. (2010). Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nature Cell Biology* *12*, 853–862.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* *9*, 129–140.

Sasaki, M., Knobbe, C.B., Itsumi, M., Elia, A.J., Harris, I.S., Chio, I.I.C., Cairns, R.A., McCracken, S., Wakeham, A., Haight, J., et al. (2012a). D-2-hydroxyglutarate produced by mutant IDH1 perturbs collagen maturation and basement membrane function. *Genes Dev.* *26*, 2038–2049.

Sasaki, M., Knobbe, C.B., Munger, J.C., Lind, E.F., Brenner, D., Brüstle, A., Harris, I.S., Holmes, R., Wakeham, A., Haight, J., et al. (2012b). IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* *488*, 656–659.

Sato, M., Kimura, T., Kurokawa, K., Fujita, Y., Abe, K., Masuhara, M., Yasunaga, T., Ryo, A., Yamamoto, M., and Nakano, T. (2002). Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech. Dev.* *113*, 91–94.

Schermelleh, L., Spada, F., Easwaran, H.P., Zolghadr, K., Margot, J.B., Cardoso, M.C., and Leonhardt, H. (2005). Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods* *2*, 751–756.

Schermelleh, L., Haemmer, A., Spada, F., Rösing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* *35*, 4301–4312.

Schiesser, S., Hackner, B., Pfaffeneder, T., Müller, M., Hagemeyer, C., Truss, M., and Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing.

Angew. Chem. Int. Ed Engl. *51*, 6516–6520.

Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., Álvarez-Castro, J.M., Nalaskowski, M.M., Kremmer, E., Schmid, V., et al. (2013). Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res.* *41*, 4860–4876.

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

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

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell* *15*, 459–471.

Shi, J., Zuo, H., Ni, L., Xia, L., Zhao, L., Gong, M., Nie, D., Gong, P., Cui, D., Shi, W., et al. (2014). An IDH1 mutation inhibits growth of glioma cells via GSH depletion and ROS generation. *Neurological Sciences* *35*, 839–845.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* *119*, 941–953.

Shibutani, T., Ito, S., Toda, M., Kanao, R., Collins, L.B., Shibata, M., Urabe, M., Koseki, H., Masuda, Y., Swenberg, J.A., et al. (2014). Guanine- 5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci. Rep.* *4*, 5220.

Shin, S.-W., Vogt, E.J., Jimenez-Movilla, M., Baibakov, B., and Dean, J. (2017). Cytoplasmic cleavage of DPPA3 is required for intracellular trafficking and cleavage-stage development in mice. *Nat. Commun.* *8*, 1643.

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

Sidhu, H., and Capalash, N. (2017). UHRF1: The key regulator of epigenetics and molecular target for cancer therapeutics. *Tumour Biol.* *39*, 1010428317692205.

Singh, P., Wu, X., Lee, D.-H., Li, A.X., Rauch, T.A., Pfeifer, G.P., Mann, J.R., and Szabó, P.E. (2011). Chromosome-wide analysis of parental allele-specific chromatin and DNA methylation. *Mol. Cell. Biol.* *31*, 1757–1770.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C.G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* *15*, 259–267.

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

Skvortsova, K., Iovino, N., and Bogdanovic, O. (2018). Functions and mechanisms of epigenetic inheritance in animals. *Nat. Rev. Mol. Cell Biol.* *19*, 774–790.

Smets, M., Link, S., Wolf, P., Schneider, K., Solis, V., Ryan, J., Meilinger, D., Qin, W., and Leonhardt, H. (2017). DNMT1 mutations found in HSNIE patients affect interaction with UHRF1 and neuronal differentiation. *Human Molecular Genetics* *26*, 1522–1534.

Smith, S., and Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required

for chromatin assembly during DNA replication in vitro. *Cell* *58*, 15–25.

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

Song, J., and Pfeifer, G.P. (2016). Are there specific readers of oxidized 5-methylcytosine bases? *Bioessays* *38*, 1038–1047.

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

Song, S.J., Ito, K., Ala, U., Kats, L., Webster, K., Sun, S.M., Jongen-Lavrencic, M., Manova-Todorova, K., Teruya-Feldstein, J., Avigan, D.E., et al. (2013). The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation. *Cell Stem Cell* *13*, 87–101.

Sood, A.J., Viner, C., and Hoffman, M.M. (2019). DNAmoD: the DNA modification database. *J. Cheminform.* *11*, 30.

Soria, G., Polo, S.E., and Almouzni, G. (2012). Prime, repair, restore: the active role of chromatin in the DNA damage response. *Mol. Cell* *46*, 722–734.

Soshnev, A.A., Josefowicz, S.Z., and Allis, C.D. (2016). Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. *Mol. Cell* *62*, 681–694.

Sperber, H., Mathieu, J., Wang, Y., Ferreccio, A., Hesson, J., Xu, Z., Fischer, K.A., Devi, A., Detraux, D., Gu, H., et al. (2015). The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nat. Cell Biol.* *17*, 1523–1535.

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.

Steenweg, M.E., Jakobs, C., Errami, A., van Dooren, S.J.M., Adeva Bartolomé, M.T., Aerssens, P., Augoustides-Savvapoulou, P., Baric, I., Baumann, M., Bonafé, L., et al. (2010). An overview of L-2-hydroxyglutarate dehydrogenase gene (L2HGDH) variants: a genotype-phenotype study. *Hum. Mutat.* *31*, 380–390.

Struys, E.A., Verhoeven, N.M., Ten Brink, H.J., Wickenhagen, W.V., Gibson, K.M., and Jakobs, C. (2005a). Kinetic characterization of human hydroxyacid-oxoacid transhydrogenase: relevance to D-2-hydroxyglutaric and gamma-hydroxybutyric acidurias. *J. Inherit. Metab. Dis.* *28*, 921–930.

Struys, E.A., Salomons, G.S., Achouri, Y., Van Schaftingen, E., Grosso, S., Craigen, W.J., Verhoeven, N.M., and Jakobs, C. (2005b). Mutations in the D-2-hydroxyglutarate dehydrogenase gene cause D-2-hydroxyglutaric aciduria. *Am. J. Hum. Genet.* *76*, 358–360.

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

Suzuki, M., Oda, M., Ramos, M.-P., Pascual, M., Lau, K., Stasiak, E., Agyiri, F., Thompson, R.F., Glass, J.L., Jing, Q., et al. (2011). Late-replicating heterochromatin is characterized by decreased cytosine methylation in the human genome. *Genome Res.* *21*, 1833–1840.

Syntichaki, P., Topalidou, I., and Thireos, G. (2000). The Gcn5 bromodomain co-ordinates nucleosome remodelling. *Nature* *404*, 414–417.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.* *276*, 25309–25317.

- Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* *19*, 815–826.
- 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.
- Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A., and Tajima, S. (2011). Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc. Natl. Acad. Sci. U. S. A.* *108*, 9055–9059.
- Talbert, P.B., and Henikoff, S. (2010). Histone variants — ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology* *11*, 264–275.
- Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* *139*, 1895–1902.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., and Montellier, E. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* *146*, 1016–28.
- Tang, F., Barbacioru, C., Nordman, E., Bao, S., Lee, C., Wang, X., Tuch, B.B., Heard, E., Lao, K., and Surani, M.A. (2011). Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* *6*, e21208.
- Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., et al. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* *46*, 385–388.
- Teslaa, T., and Teitell, M.A. (2015). Pluripotent stem cell energy metabolism: an update. *EMBO J.* *34*, 138–153.
- TeSlaa, T., Chaikovsky, A.C., Lipchina, I., Escobar, S.L., Hochedlinger, K., Huang, J., Graeber, T.G., Braas, D., and Teitell, M.A. (2016). α -Ketoglutarate Accelerates the Initial Differentiation of Primed Human Pluripotent Stem Cells. *Cell Metab.* *24*, 485–493.
- Thalhammer, A., Hansen, A.S., El-Sagheer, A.H., Brown, T., and Schofield, C.J. (2011). Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chem. Commun.* *47*, 5325–5327.
- Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., Ghesquière, B., Van Dyck, L., Boeckx, B., Schoonjans, L., et al. (2016). Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* *537*, 63–68.
- Tian, Y., Paramasivam, M., Ghosal, G., Chen, D., Shen, X., Huang, Y., Akhter, S., Legerski, R., Chen, J., Seidman, M.M., et al. (2015). UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold. *Cell Rep.* *10*, 1957–1966.
- Tischler, J., Gruhn, W.H., Reid, J., Allgeyer, E., Buettner, F., Marr, C., Theis, F., Simons, B.D., Wernisch, L., and Surani, M.A. (2019). Metabolic regulation of pluripotency and germ cell fate through α -ketoglutarate. *EMBO J.* *38*, e99518.
- Tjeertes, J.V., Miller, K.M., and Jackson, S.P. (2009). Screen for DNA- damage- responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO J.* *28*, 1878–89.
- Tommasini-Ghelfi, S., Murnan, K., Kouri, F.M., Mahajan, A.S., May, J.L., and Stegh, A.H. (2019). Cancer-associated mutation and beyond: The emerging biology of isocitrate dehydrogenases in human disease. *Sci Adv* *5*, eaaw4543.

- Tropberger, P., Pott, S., Keller, C., and Kamieniarz-Gdula, K. (2013). Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell* *152*, 859–72.
- Tsai, M.-C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* *329*, 689–693.
- Tsukada, Y.-I., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* *439*, 811–816.
- Turcan, S., Rohle, D., Goenka, A., Walsh, L.A., Fang, F., Yilmaz, E., Campos, C., Fabius, A.W.M., Lu, C., Ward, P.S., et al. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* *483*, 479–483.
- Turchi, L., Fareh, M., Aberdam, E., Kitajima, S., Simpson, F., Wicking, C., Aberdam, D., and Virolle, T. (2009). ATF3 and p15 PAF are novel gatekeepers of genomic integrity upon UV stress. *Cell Death Differ.* *16*, 728–737.
- Unoki, M., Daigo, Y., Koinuma, J., Tsuchiya, E., Hamamoto, R., and Nakamura, Y. (2010). UHRF1 is a novel diagnostic marker of lung cancer. *Br. J. Cancer* *103*, 217–222.
- Vaughan, R.M., Dickson, B.M., Whelihan, M.F., Johnstone, A.L., Cornett, E.M., Cheek, M.A., Ausherman, C.A., Cowles, M.W., Sun, Z.-W., and Rothbart, S.B. (2018). Chromatin structure and its chemical modifications regulate the ubiquitin ligase substrate selectivity of UHRF1. *Proc. Natl. Acad. Sci. U. S. A.* *115*, 8775–8780.
- Vaughan, R.M., Rothbart, S.B., and Dickson, B.M. (2019). The finger loop of the SRA domain in the E3 ligase UHRF1 is a regulator of ubiquitin targeting and is required for maintaining DNA methylation. *J. Biol. Chem.* *294*, 15724–15732.
- Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q.V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.-J., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* *50*, 83–95.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* *297*, 1833–1837.
- Voon, H.P.J., and Gibbons, R.J. (2016). Maintaining memory of silencing at imprinted differentially methylated regions. *Cell. Mol. Life Sci.* *73*, 1871–1879.
- Waddington, C.H. (1956). Genetic Assimilation of the Bithorax Phenotype. *Evolution* *10*, 1.
- Waddington, C.H. (1957). *The Strategy of the Genes, a Discussion of Some Aspects of Theoretical Biology*, by C.H. Waddington. With an Appendix by H. Kacser, (G. Allen and Unwin).
- Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* *20*, 116–117.
- Wang, F., Yang, Y.-Z., Shi, C.-Z., Zhang, P., Moyer, M.P., Zhang, H.-Z., Zou, Y., and Qin, H.-L. (2012). UHRF1 promotes cell growth and metastasis through repression of p16(ink4a) in colorectal cancer. *Ann. Surg. Oncol.* *19*, 2753–2762.
- Wang, J., Alexander, P., Wu, L., Hammer, R., Cleaver, O., and McKnight, S.L. (2009). Dependence of mouse embryonic stem cells on threonine catabolism. *Science* *325*, 435–439.
- Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell* *157*, 979–991.
- Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X.-D., et al. (2015a).

Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* 523, 621–625.

Wang, P., Wu, J., Ma, S., Zhang, L., Yao, J., Hoadley, K.A., Wilkerson, M.D., Perou, C.M., Guan, K.-L., Ye, D., et al. (2015b). Oncometabolite D-2-Hydroxyglutarate Inhibits ALKBH DNA Repair Enzymes and Sensitizes IDH Mutant Cells to Alkylating Agents. *Cell Rep.* 13, 2353–2361.

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

Warburg, O. (1956). On the Origin of Cancer Cells. *Science* 123, 309–314.

Ward, P.S., and Thompson, C.B. (2012). Metabolic reprogramming: A cancer hallmark even warburg did not anticipate. *Cancer Cell* 21, 297–308.

Washington, J.M., Rathjen, J., Felquer, F., Lonic, A., Bettess, M.D., Hamra, N., Semendric, L., Tan, B.S.N., Lake, J.-A., Keough, R.A., et al. (2010). L-Proline induces differentiation of ES cells: a novel role for an amino acid in the regulation of pluripotent cells in culture. *Am. J. Physiol. Cell Physiol.* 298, C982–C992.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.

Weber, A.R., Krawczyk, C., Robertson, A.B., Kuśnierczyk, A., Vågbo, C.B., Schuermann, D., Klungland, A., and Schär, P. (2016). Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications* 7, 10806.

Wei, J.-W., Huang, K., Yang, C., and Kang, C.-S. (2017). Non-coding RNAs as regulators in epigenetics (Review). *Oncol. Rep.* 37, 3–9.

Weissmann, S., Alpermann, T., Grossmann, V., Kowarsch, A., Nadarajah, N., Eder, C., Dicker, F., Fasan, A., Haferlach, C., Haferlach, T., et al. (2012). Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia* 26, 934–942.

Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., et al. (2014). Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol.* 15, R49.

William Bateson Mendel, G., and Bateson, W. (2009). Mendel's Principles of Heredity: A Defence, with a Translation of Mendel's Original Papers on Hybridisation (Cambridge Library Collection--Life Sciences).

Wilson, R.C., and Doudna, J.A. (2013). Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* 42, 217–239.

Winkelmann, J., Lin, L., Schormair, B., Kornum, B.R., Faraco, J., Plazzi, G., Melberg, A., Cornelio, F., Urban, A.E., Pizza, F., et al. (2012). Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum. Mol. Genet.* 21, 2205–2210.

Wu, J.C., and Santi, D.V. (1985). On the mechanism and inhibition of DNA cytosine methyltransferases. *Prog. Clin. Biol. Res.* 198, 119–129.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* 18, 517–534.

Wu, B.-K., Mei, S.-C., and Brenner, C. (2014a). RFTS-deleted DNMT1 enhances tumorigenicity with focal hypermethylation and global hypomethylation. *Cell Cycle* 13, 3222–3231.

Wu, H., Wu, X., Shen, L., and Zhang, Y. (2014b). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* 32, 1231–1240.

- Wu, X., Li, G., and Xie, R. (2018). Decoding the role of TET family dioxygenases in lineage specification. *Epigenetics Chromatin* *11*, 58.
- Xie, S., and Qian, C. (2018). The Growing Complexity of UHRF1-Mediated Maintenance DNA Methylation. *Genes* *9*, 600.
- Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* *402*, 187–191.
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., Ito, S., Yang, C., Wang, P., Xiao, M.-T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* *19*, 17–30.
- Xue, B., Zhao, J., Feng, P., Xing, J., Wu, H., and Li, Y. (2019). Epigenetic mechanism and target therapy of UHRF1 protein complex in malignancies. *Onco. Targets. Ther.* *12*, 549–559.
- Yadon, A.N., and Tsukiyama, T. (2011). SnapShot: Chromatin remodeling: ISWI. *Cell* *144*, 453–453.e1.
- Yamaguchi, L., Nishiyama, A., Misaki, T., Johmura, Y., Ueda, J., Arita, K., Nagao, K., Obuse, C., and Nakanishi, M. (2017). Usp7-dependent histone H3 deubiquitylation regulates maintenance of DNA methylation. *Sci. Rep.* *7*, 55.
- Yamaguchi, S., Shen, L., Liu, Y., Sandler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature* *504*, 460–464.
- Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J., et al. (2009). IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* *360*, 765–773.
- Yang, X.-J., and Seto, E. (2007). HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* *26*, 5310–5318.
- Yang, H., Lin, H., Xu, H., Zhang, L., Cheng, L., Wen, B., Shou, J., Guan, K., Xiong, Y., and Ye, D. (2014). TET-catalyzed 5-methylcytosine hydroxylation is dynamically regulated by metabolites. *Cell Res.* *24*, 1017–1020.
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* *356*, eaaj2239.
- Ying, Q.-L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* *453*, 519–523.
- Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* *13*, 335–340.
- Yu, F., Zingler, N., Schumann, G., and Strätling, W.H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* *29*, 4493–4501.
- Zeng, Y., Yao, B., Shin, J., Lin, L., Kim, N., Song, Q., Liu, S., Su, Y., Guo, J.U., Huang, L., et al. (2016). Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Mol. Cell* *61*, 153–160.
- Zhang, J., Ratanasirinrawoot, S., Chandrasekaran, S., Wu, Z., Ficarro, S.B., Yu, C., Ross, C.A., Cacchiarelli, D., Xia, Q., Seligson, M., et al. (2016a). LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell* *19*, 66–80.
- Zhang, J., Zhao, J., Dahan, P., Lu, V., Zhang, C., Li, H., and Teitell, M.A. (2018). Metabolism in Pluripotent Stem Cells and Early Mammalian Development. *Cell Metab.* *27*, 332–338.

Zhang, K., Gao, Y., Li, J., Burgess, R., Han, J., Liang, H., Zhang, Z., and Liu, Y. (2016b). A DNA binding winged helix domain in CAF-1 functions with PCNA to stabilize CAF-1 at replication forks. *Nucleic Acids Res.* *44*, 5083–5094.

Zhang, P., Huang, B., Xu, X., and Sessa, W.C. (2013). Ten-eleven translocation (Tet) and thymine DNA glycosylase (TDG), components of the demethylation pathway, are direct targets of miRNA-29a. *Biochem. Biophys. Res. Commun.* *437*, 368–373.

Zhang, P., Hastert, F.D., Ludwig, A.K., Breitwieser, K., Hofstätter, M., and Cardoso, M.C. (2017a). DNA base flipping analytical pipeline. *Biol Methods Protoc.* *2*, bpx010.

Zhang, P., Ludwig, A.K., Hastert, F.D., Rausch, C., Lehmkuhl, A., Hellmann, I., Smets, M., Leonhardt, H., and Cardoso, M.C. (2017b). L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins. *Nucleus* *8*, 548–562.

Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J., and Wong, J. (2014). Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked β -N-acetylglucosamine transferase (OGT). *J. Biol. Chem.* *289*, 5986–5996.

Zhang, Y.W., Wang, Z., Xie, W., Cai, Y., Xia, L., Easwaran, H., Luo, J., Yen, R.-W.C., Li, Y., and Baylin, S.B. (2017c). Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress. *Mol. Cell* *65*, 323–335.

Zhang, Z.-M., Rothbart, S.B., Allison, D.F., Cai, Q., Harrison, J.S., Li, L., Wang, Y., Strahl, B.D., Wang, G.G., and Song, J. (2015). An Allosteric Interaction Links USP7 to Deubiquitination and Chromatin Targeting of UHRF1. *Cell Rep.* *12*, 1400–1406.

Zhao, Q., Zhang, J., Chen, R., Wang, L., Li, B., Cheng, H., Duan, X., Zhu, H., Wei, W., Li, J., et al. (2016). Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nat. Commun.* *7*, 12464.

Zhao, S., Xu, J., Liu, S., Cui, K., Li, Z., and Liu, N. (2019). Dppa3 in pluripotency maintenance of ES cells and early embryogenesis. *J. Cell. Biochem.* *120*, 4794–4799.

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C., et al. (2012). HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* *31*, 2103–2116.

Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* *17*, 551–565.

Zhu, Q., Stöger, R., and Alberio, R. (2018). A Lexicon of DNA Modifications: Their Roles in Embryo Development and the Germline. *Frontiers in Cell and Developmental Biology* *6*, 24.

4.2 Abbreviations

2-HG	2-hydroxyglutarate
2-OG	2-oxoglutarate (α -KG)
2HGDH	2-HG dehydrogenase
3D	3-dimensional
5-aza	5-azacytidine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
6mA	6-methyladenosine
ADCA-DN	Autosomal dominant cerebellar ataxia deafness and narcolepsy
ADD	ATRX-DNMT3-DNMT3L
ADP	Adenine diphosphate
AM-AR	Active modification-active removal
AM-DP	Active modification-passive dilution
AML	Acute myeloid leukemia
APE1	AP endonuclease 1
ATP	Adenosine triphosphate
BAH	Bromo-adjacent homology
BER	Base excision repair
bp	Base pair
C-terminus	Carboxy-terminus
cenH3	Centromere-specific H3 variant
CENP-A	Centromere protein A
CM	Catalytic mutant
CMML	Chronic myelomonocytic leukemia
CO ₂	Carbon dioxid
COX	Cytochrome-c oxidase
CpG	Cytosine-phosphate bond-guanine
DE-2HG	Diethyl-2-hydroxyglutarate
DM- α -KG	Dimethyl-alpha-ketoglutarate
DMR	Differentially-methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DPPA3	Developmental pluripotency-associated protein 3
DSB	Double-strand break
DSBH	Double-stranded beta helix
EpiLC	Epiblast-like cell
ESC	Embryonic stem cell
FDA	Food and drug administration
Fe(II)	Iron(II)
FLT3	Fms like tyrosine kinase 3
FRAP	Fluorescence recovery after photobleaching
GC-MS	Gas chromatography-mass spectrometry
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H5	Histone 5

HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HIF	Hypoxia-inducible factor
hMeDIP-seq	Hydroxymethylated DNA immunoprecipitation sequencing
HOX	Homeobox
HP1	Heterochromatin protein 1
HSANIE	Hereditary sensory and autonomic neuropathy type IE
HSC	Hematopoietic stem cell
IAP	Intracisternal A-particle
ICF syndrome	Immunodeficiency, centromeric instability & facial anomalies syndrome
ICM	Inner cell mass
IDAX	Inhibition of the Dvl and Axin complex
IDH	Isocitrate dehydrogenase
JmjC-domain	Jumonji C-domain
KAP1	KRAB-associated protein-1
KD	Dissociation constant
KDM	Lysine-specific demethylase
KMT	Lysine-specific methyltransferase
L1	LINE-1
LIF	Leukemia inhibitory factor
LIG1	DNA ligase 1
LINEs	Long interspersed nuclear elements
lncRNA	Long non-coding RNA
MBD	Methyl-CpG-binding domain
MD	Molecular dynamics
MDS	Myelodysplastic syndrome
MECP2	Methyl CpG binding protein 2
miRNA	Micro RNA
MLL	Mixed lineage leukemia
mRNA	Messenger RNA
N-terminus	Amino-terminus
NAD ⁺	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
ncRNA	Non-coding RNA
NEIL	Nei-like
NKR finger	Asparagine-lysine-arginine finger
NP95	Nuclear protein 95 (UHRF1)
NPC	Neural progenitor cell
O-GlcNAc	O-linked N-acetylglucosamine
OXPHOS	Oxidative phosphorylation
PAF15	PCNA-associated factor of 15 kDa
PCNA	Proliferating cell nuclear antigen
PCR2	Polycomb Repressive Complex 2
PGC	Primordial germ cell
PHD	Plant homeodomain
PIP	PCNA interacting protein/peptide
piRNA	PIWI-interacting RNA
PML	Promyelocytic leukemia
Pol II	Polymerase II
Pol β	DNA polymerase β
Pro	Proline
PSAT1	Phosphoserine aminotransferase 1

PSC	Pluripotent stem cell
PTM	Post-translational modification
PWWP	Pro-Trp-Trp-Pro
RFTS	Replication foci targeting sequence
RING	Really interesting new gene
RNA	Ribonucleic acid
RNAi	RNA interference
SAM	S-adenosylmethionine
SILAC	Stable Isotope Labeling with Amino Acids in Cell Culture
siRNA	Short interfering RNA
SRA	SET- and RING-associated
TDG	Thymine DNA glycosylase
TE	Transposable element
TET	Ten-eleven-translocation
TF	Transcription factor
TKO	Triple knockout
TLS	Translesion synthesis
Trp	Tryptophane
TTD	Tandem tudor
UBL	Ubiquitin-like
UDP	Uridine diphosphate
UHRF1	Ubiquitin-like containing PHD and RING finger domains 1
UIM	Ubiquitin interacting motif
USP7	Ubiquitin specific peptidase 7
wt	Wildtype
WT1	Wilms-tumor 1
Xi	Inactive X chromosome
Xist	X inactive specific transcript
α -KG	Alpha-ketoglutarate

4.3 List of publications

Schneider M*, **Trummer C***, Stengl A, Zhang P, Szwagierczak A, Cardoso MC, Leonhardt H, Bauer C and Antes I (2020). Systematic analysis of the binding behavior of UHRF1 towards different methyl- and carboxylcytosine modification patterns at CpG dyads. *PLoS ONE* 15(2): e0229144.

Nishiyama A, Mulholland CB, Bultmann S, Kori S, Endo A, Saeki Y, Qin W, **Trummer C**, Chiba Y, Yokoyama H, Kumamoto S, Kawakami T, Hojo H, Nagae G, Aburatani H, Tanaka K, Arita K, Leonhardt H and Nakanishi M (2020). Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. *Nature Communications* 11(1):1222.

Mulholland CB, Nishiyama A, Ryan J, Nakamura R, Yiğit M, Glück IM, **Trummer C**, Qin W, Bartoschek MD, Traube FR, Parsa E, Ugur E, Modic M, Acharya A, Stolz P, Ziegenhain C, Wierer M, Enard W, Carell T, Lamb DC, Takeda H, Nakanashi M, Bultmann S and Leonhardt H. Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive demethylation in mammals. *Manuscript in preparation*.

Trummer C, Müller U, Lehmann M, Kraus L, Bultmann S, Smets M, Geigenberger P and Leonhardt H. Metabolic regulation of TET enzymes in mESCs. *Manuscript in preparation*.

Ludwig AK*, Zhang P*, Hastert FD, Meyer S, Rausch C, Herce HD, Müller U, Lehmkuhl A, Hellmann I, **Trummer C**, Storm C, Leonhardt H and Cardoso MC (2017). *Nucleic Acid Research*, Vol. 45, No. 5

* these authors contributed equally to this work as first author

4.4 Statutory declaration and statement

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt wurde.

München, den 16.03.2020

Carina Trummer

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig ohne Erfolg einer Doktorprüfung unterzogen habe.

München, den 16.03.2020

Carina Trummer

4.5 Declaration of contribution

Systematic analysis of the binding behavior of UHRF1 towards different methyl- and carboxylcytosine modification patterns at CpG dyads

This study was conceived by Markus Schneider, Carina Trummer, Christina Bauer, Heinrich Leonhardt and Iris Antes. Markus Schneider performed the molecular dynamics simulation, generated interaction networks and analyzed root mean square fluctuations and DNA groove widths in MD trajectories. Carina Trummer performed the experimental part of the study including electrophoretic mobility shift assays, cloning and purification of UHRF1-SRA, microscale thermophoresis, size exclusion chromatography and melting temperature analysis. Christina Bauer and Carina Trummer jointly analyzed experimental data. Markus Schneider and Christina Bauer wrote the original manuscript draft. Carina Trummer assisted in manuscript writing and proofreading. Iris Antes handled the submission process.

Markus Schneider

Carina Trummer

Prof. Dr. Heinrich Leonhardt

Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation

This study was conceived by Atsuya Nishiyama, Kyohei Arita, Heinrich Leonhardt and Makoto Nakanishi. Carina Trummer performed F3H experiments, immunostainings and qPCR analysis in mammalian cells, quantified the data and provided the corresponding figures. Carina Trummer assisted in manuscript writing and proofreading.

Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive demethylation in mammals

Christopher B. Mulholland, Sebastian Bultmann and Heinrich Leonhardt designed and conceived this study. Carina Trummer performed F3H experiments and quantified the data, purified UHRF1 protein, performed microscale thermophoresis experiments, provided the corresponding figures and proofread the manuscript.

Carina Trummer

Prof. Dr. Heinrich Leonhardt

Metabolic regulation of TET enzymes in mESCs

Carina Trummer and Heinrich Leonhardt conceived this study. Carina Trummer designed all experiments and performed the experiments, partially together with Laura Kraus. Martha Smets preped the RNA library and Sebastian Bultmann analyzed the corresponding data. Martin Lehmann performed and analyzed mass spectrometry experiments. Carina Trummer wrote the manuscript with the assistance of Heinrich Leonhardt.

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

This study was designed and conceived by Anne K. Ludwig, Peng Zhang and M. Christina Cardoso. Carina Trummer cultured embryonic stem cells and isolated and purified RNA samples for the corresponding RNAseq analysis. Carina Trummer proofread the manuscript.

Carina Trummer

Prof. Dr. Heinrich Leonhardt

4.6 Acknowledgement

First of all, I would like to thank Prof. Dr. Heinrich Leonhardt for giving me the opportunity to conduct my PhD thesis in his lab. Thank you for your support over the years and the trust you have placed in my work.

I would further like to thank my collaboration partners Markus Schneider, Dr. Christina Bauer, Dr. Martin Lehmann, Dr. Peng Zhang, Prof. Dr. M. Cristina Cardoso and the people around Dr. Atsuya Nishiyama for successful collaborations and your scientific enthusiasm in our joint projects. Christina, you were more than a collaborator and I truly want to thank you for everything. With you on board, even “the never ending story” came to an end.

I am also grateful to my former master student Laura Kraus for generating cell lines and her help with experiments, to the members of my TAC committee, Prof. Dr. Peter Geigenberger and Prof. Dr. Irmela Jeremias, and to the coordination office of my PhD programs IMPRS-LS and IRTG1243.

Most important, I would like to thank all members of the Leonhardt lab for the time I spent with them. Thank you for the friendly and supportive atmosphere you created. I really enjoyed working with you and would like to emphasize the great team spirit we had. My special thanks go to Susi Leidescher, or rather Tzu, for being such a caring colleague and for all the cheerful hours spent outside the lab. Michi Bartolicious Bartoschek, for many IMPRS activities and courses we attended together and for your endless readiness to help others. Chris Mulholland and Dr. Weihua Qin, thanks for always proving to me that true passionate scientists exist and thank you for the scientific but also fun discussions we had. Andreas Stengl, thank you for all the bad jokes while working at the bench and for showing me “special” procedures of protein purification. Katharina Brandstetter, thanks for laid-back coffee breaks and encouraging words. Jeannette Koch, thank you for answering thousands of “silly” questions and for your great stamina and kindness when moving the lab.

Above all, I want to thank Paul, my family and close friends. Writing a doctoral thesis and earning a PhD is one thing, being truly happy at the same time is quite another. Thank you for being who you are and for shaping my personal life the way you did. You are my foundation for everything.