ROLE AND REGULATION OF TET-MEDIATED DNA MODIFICATIONS IN GENE EXPRESSION

Udo Müller

München 2014

ROLE AND REGULATION OF TET-MEDIATED DNA MODIFICATIONS IN GENE EXPRESSION

Dissertation
an der Fakultät für Biologie
der Ludwigs-Maximilians-Universität München

vorgelegt von
Udo Müller
München, den 28.10.2014

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

Tag der mündlichen Prüfung: 19.12.2014

TABLE OF CONTENTS

SUMMARY

In the mammalian genome, cytosine methylation (5mC) plays a central role in the epigenetic regulation of gene expression and has been implicated in a variety of biological processes, including genome stability, imprinting or differentiation. Compared to other epigenetic marks, DNA methylation has been thought to be relatively stable. However, genome-wide loss of 5mC, or DNA demethylation, has been observed in specific developmental stages and in various types of cancer. The discovery of the TET family of enzymes in 2009 was a watershed moment in comprehending the mechanisms of DNA demethylation. TET proteins oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formlycytosine (5fC) and 5-carboxylcytosine (5caC), which not only serve as key intermediates in active DNA demethylation pathways, but can also act as independent epigenetic marks. In this study, various aspects of TET-mediated DNA demethylation have been intensively investigated.

Using quantitative mass-spectrometry-based proteomics readers for the different cytosine derivatives in mouse embryonic stem cells (ESCs), neuronal progenitor cells, and adult mouse brain tissue were identified. Readers for these modifications are only partially overlapping and are dynamic during differentiation. Moreover, the oxidized derivatives of 5mC recruit distinct transcription regulators as well as a large number of DNA repair proteins, implicating DNA damage response as the main pathway contributing to active DNA demethylation.

To identify additional non-canonical DNA bases, highly sensitive quantitative mass-spectrometry led to the discovery of 5-hydroxymethyluracil (5hmU) in ESCs. Genomic 5hmU is not generated via deamination of 5hmC, as widely suggested, but through direct oxidation of thymine by TET proteins. In addition, screening for specific 5hmU readers identified different transcriptional and epigenetic factors, implicating that this mark has a specific function in ESCs.

So far, only little is known how TET enzymes are regulated and how they are modified by posttranslational modifications (PTMs). Mapping TET phosphorylation and glycosylation sites at amino acid resolution revealed that these PTMs are interdependent and mostly occur at regulatory protein regions.

Finally, a reporter gene based assay could demonstrate that *in vitro* methylation causes gene silencing while subsequent oxidation, resulting in DNA demethylation, leads to gene reactivation *in vivo*. Different knockout and rescue experiments clearly show that oxidation of methylcytosine by TET proteins and subsequent removal by TDG or NEIL glycosylases and the base excision repair pathway results in reactivation of epigenetically silenced genes.

In conclusion, this work provides new insights how TET proteins can set DNA modifications, how these oxidized bases are read by various factors and how TET proteins can be posttranslationally modified. Furthermore, removal of 5mC is achieved through TET-mediated oxidation and depends on the activity of specific glycosylases, which leads to gene reactivation.

1 INTRODUCTION

1.1 Epigenetic regulation

DNA is the essential macromolecule that stores all genetic information and it propagates this information to the next generation through the germ line [Avery et al., 1944]. Determining the structural details of the DNA double helix, which is considered as one of the landmark discoveries in biology and other findings gave rise to the "central dogma" of modern molecular biology [Watson and Crick, 1953]. This dogma describes the processes and mechanisms involved in maintaining and translating the genetic template required for life. The principal steps are the self-propagation of DNA by semiconservative replication, unidirectional transcription of the DNA into an intermediary messenger RNA (mRNA) and the translation of mRNA into chains of amino acids, the building blocks of proteins [Crick, 1970].

Cells of multicellular organisms contain the same genetic information, however, they differ dramatically in morphology and function. These differences are based on complex and dynamic changes of gene expression patterns that cannot be explained by alterations in the DNA sequence [Jaenisch and Bird, 2003]. This additional layer of information is termed epigenetic and responsible for the identity of every single cell [Waddington, 1957]. The structure and composition of chromatin are generally modulated by heritable epigenetic modifications, which result in the establishment and maintenance of stable chromatin signatures. These so called "epigenetic landscapes" are highly organized systems of information storage on the genome and mediate a response to environmental signals that lead to changes in gene expression [Goldberg et al., 2007].

Epigenetic mechanisms include DNA methylation, posttranslational histone modifications, histone variants, nucleosome remodeling, as well as regulation by non-coding RNAs and ensure long-term stability of transcriptional states (Figure 1). In recent years, it has become apparent that epigenetic modifications do not stand on their own, but are connected in various combinations to cross talk with each other in a sense that manifolds their functions [Bannister and Kouzarides, 2011; Ponting et al., 2009; Reik, 2007].

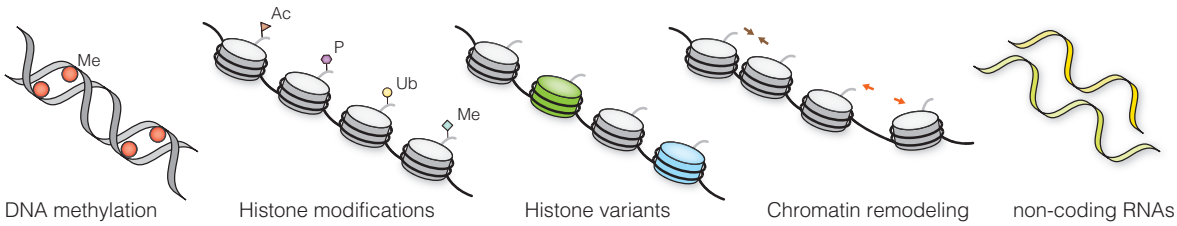


Figure 1 Epigenetic mechanisms

Epigenetic regulation of gene expression is mediated by the crosstalk between DNA methylation and various posttranslational histone modifications. The replacement of canonical histone proteins by histone variants and nucleosome remodeling can additionally alter the structural composition and accessibility of chromatin. Further factors contributing to epigenetic regulation are long non-coding RNAs.

The establishment, the reading and the removal of these epigenetic marks control the pluripotency of embryonic stem cells and other key developmental processes such as imprinting, neurogenesis as well as X inactivation in female cells [Augui et al., 2011; Kacem and Feil, 2009; Tee and Reinberg, 2014].

1.2 Chromatin structure and histone modifications

In eukaryotic cells, DNA is not present as a naked double helix, but rather organized into complex higher order chromatin structures. 146 base-pair long DNA strands are wrapped around an octamer of histone proteins to form nucleosomes, representing the basic level of chromatin compaction [Richmond and Davey, 2003]. One histone octamer contains the four core histone proteins H2A, H2B, H3 and H4, which pairwise interact to form the barrel-shaped histone particle. Moreover, each histone also harbors a 20-35 amino acid long N-terminal peptide that extends from the surface of the nucleosome [Luger et al., 1997]. These so called histone tails and to a lower extent the core histone itself are subject to a large number and variety of posttranslational modifications. The modifications include methylation and acetylation of lysines and arginines, phosphorylation of serines and threonines, ubiquitination and sumoylation of lysines, as well as ribosylation and proline isomerization [Kouzarides, 2007]. Additionally, methylated and unmethylated arginines undergo deimination, resulting in their conversion to citrulline [Cuthbert et al., 2004].

In general, histone modifications can act via two different mechanisms to control chromatin structure and thus gene expression. First, the modifications can alter the electrostatic charge of the histone resulting in a structural change and disturbed binding to DNA. Of all known modifications, acetylation shows the highest potential to decondense chromatin, since it neutralizes the basic charge of histone proteins. Second, the composition of histone modifications can promote or inhibit the binding of non-histone proteins to chromatin. One example is HP1 (heterochromatin protein 1) that specifically binds to histone H3 trimethylated on lysine 9 (H3K9me3), a predominant mark for heterochromatin and transcriptionally inactive regions [Jacobs et al., 2001]. The release of HP1 from its binding sites during mitosis is regulated by a transient phosphorylation of H3 on serine 10 [Fischle et al., 2005].

Moreover, these histone marks can regulate each other, providing dynamic epigenetic crosstalk. One modification can influence the occurrence of one or more subsequent modifications on the same histone molecule, or between different histone molecules and across nucleosomes. Histone modifications may even form a code that is read by the binding of specific proteins to regulate distinct downstream functions [Latham and Dent, 2007; Strahl and Allis, 2000].

1.3 DNA methylation

DNA methylation was the first epigenetic mechanism known to directly correlate with gene expression [Razin and Riggs, 1980]. In eukaryotes, DNA methylation implies the addition of a methyl group to the C5 position of cytosine, which results in the formation of 5-methylcytosine (5mC). This modification has been intensively studied and is well conserved among most plant, animal and fungal species. 5mC is almost exclusively found in a symmetric CpG dinucleotide context, however, small amounts of non-CpG methylation were observed [Patil et al., 2014; Ramsahoye et al., 2000]. Mammalian genomic landscapes are overall CpG-poor and global methylation occurs at 70-80% of all CpG sites [Ehrlich et al., 1982]. The remaining unmethylated CpGs were first detected as a DNA fraction that was cleaved frequently by DNA methylation sensitive restriction enzymes [Cooper et al., 1983]. These dense clusters termed CpG islands (CGIs) are on average 1000 base pairs in length and show an elevated CpG content [Suzuki and Bird, 2008]. As 5mC can be converted to thymine by enzymatic or spontaneous deamination, the evolutionary loss of genomic CpGs is thought to have occurred due to deamination of methylated sequences in the germline. Therefore, the existence of CGIs can be explained as they are never or probably only transiently methylated in the germline [Smallwood et al., 2011]. Approximately 70% of gene promoters are associated with a CGI, making this the most common promoter type in the mammalian genome [Saxonov et al., 2006]. Almost all housekeeping genes, as well as a large number of tissue-specific and developmental genes are characterized by CGI promoters, which outline the strong correlation between CGIs and transcription initiation [Larsen et al., 1992; Zhu et al., 2008]. However, a large class of CGIs that are remote from annotated transcription start sites were recently identified, but they nevertheless fulfill promoter function [Zhu et al., 2008]. The majority of CGIs is found in a hypomethylated state, but a small percentage acquires methylation during normal development leading to a stable transcriptional repression. Some of these examples are known to play a key role in X inactivation and genomic imprinting [Edwards and Ferguson-Smith, 2007; Payer and Lee, 2008].

DNA methylation is not only restricted to promoters where it is associated with transcriptional repression. The generally CpG-poor gene bodies are highly methylated and contain multiple repetitive and transposable elements. Methylation of the CpG sites in gene exons is a major reason for cytosine to thymine transition mutations, resulting in disease-causing mutations in the germline and cancer-causing mutations in somatic cells [Rideout et al., 1990]. It has been known from the early days that gene body methylation is a feature of transcribed genes [Wolf et al., 1984]. In line with that, extensive positive correlations between gene body methylation and active transcription have been confirmed on the active X chromosome, as well as on whole animal and plant genomes [Feng et al., 2010; Hellman and Chess, 2007; Lister et al., 2009]. Furthermore, exons show higher

5mC levels than introns and transitions in the degree of methylation are predominantly found at exon–intron boundaries, possibly suggesting a role for methylation in regulating cotranscriptional splicing [Laurent et al., 2010; Maunakea et al., 2013].

1.4 Setting the methylation mark

The addition of the methyl group to cytosine is catalyzed by the protein family of DNA methyltransferases (DNMTs). In vertebrates four different DNMT members have been described, which all apart from DNMT3L comprise an N-terminal regulatory and a catalytically active C-terminal domain (Figure 2). In the first step of the methylation reaction, DNMTs bind to the DNA and flip out the target base. Subsequently, they form a covalent complex by a conserved cysteine nucleophile with the C6 position of cytosine. This results in the activation of the C5 atom and the methyl group is transferred from the cofactor S-adenosyl-L-methionine (SAM), which serves as the common cellular methyl group donor. In the last reaction step, the covalent bond is resolved by β-elimination and the enzyme released from the DNA [Cheng and Blumenthal, 2008].

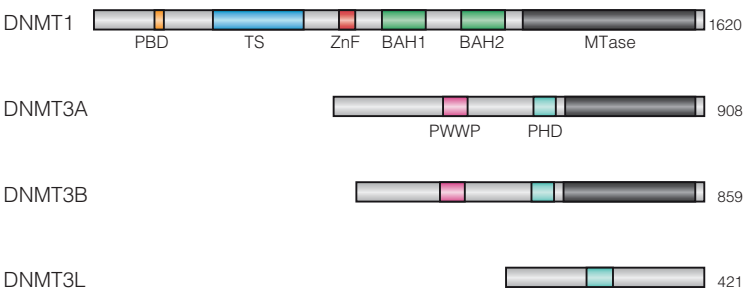


Figure 2 Domain structure of the mammalian DNMT family

All DNMTs except DNMT3L contain a regulatory N-terminal and a highly conserved catalytically active C-terminal domain. PBD: PCNA binding domain; TS: Targeting sequence; ZnF: zinc finger domain; BAH: bromo-adjacent homology domain; PWWP: Pro-Trp-Trp-Pro motif containing domain; PHD: plant homeo-domain; MTase: methyltransferase domain. Numbers indicate the length of the murine proteins in amino acids.

DNMT1

DNMT1 was the first mammalian DNA methyltransferase to be characterized and numerous genetic and biochemical studies demonstrated its role in maintaining methylation patterns during semiconservative replication [Bestor et al., 1988]. Initial studies showed that targeted mutations in the *Dnmt1* gene result in global hypomethylation in murine embryonic stem cells (ESCs) and homozygous embryos deficient in DNMT1 are delayed in development and died during mid-gastrulation [Li et al., 1992]. *Dnmt1* compound heterozygous mice display chromosomal instability and develop severe T cell lymphomas with a high frequency of chromosome 15 trisomy [Gaudet et al., 2003]. DNMT1 was also shown to be essential for X chromosome inactivation and maintenance of genomic imprints [Howell et al., 2001]. Furthermore, mouse embryos lacking DNMT1 exhibit increased transcription of intracisternal A particle (IAP) retrotransposons,

suggesting that DNMT1 also contributes to suppression of retroviral and transposable elements [Gaudet et al., 2004].

Based on bioinformatic studies, mammalian DNMT1 evolved by fusion of at least three ancestral genes [Margot et al., 2000]. Its large N-terminal regulatory part contains a PCNA (Proliferating cell nuclear antigen) binding domain (PBD), a targeting sequence (TS domain), a zinc finger domain (ZnF), two bromo adjacent homology domains (BAH1 and BAH2) and is linked to the C-terminal catalytic domain by several lysyl-glycyl dipeptide repeats [Goll and Bestor, 2005; Rottach et al., 2009]. The PBD is responsible for the interaction with PCNA, which serves as a loading platform and processivity factor for proteins contributing to DNA replication and repair [Chuang et al., 1997; Maga and Hubscher, 2003; Mortusewicz et al., 2005; Sporbert et al., 2005]. The highly conserved TS domain triggers the accumulation at pericentric heterochromatin from late S phase until early G1 phase [Easwaran et al., 2004; Leonhardt et al., 1992; Schneider et al., 2013]. Moreover, structural insights proposed a potential autoinhibitory effect of the TS domain on the catalytic activity of DNMT1 [Syeda et al., 2011]. The two BAH domains are likely involved in protein-protein interactions and the CXXC zinc finger mediates binding to DNA [Nicolas and Goodwin, 1996; Oliver et al., 2005]. On the one hand, it has been shown that the ZnF of DNMT1 prefers unmethylated CpG sites [Fatemi et al., 2001; Frauer et al., 2011b; Pradhan et al., 2008]. On the other hand, new structural details support a preference of the ZnF for hemimethylated DNA and an inhibitory role of binding to unmethylated CpG sites during maintenance methylation [Song et al., 2011; Song et al., 2012]. The catalytic domain of DNMT1 contains all conserved motifs for the methyl group transfer but additional intramolecular interactions with the N-terminal regulatory domain are required for allosteric activation [Fatemi et al., 2001; Margot et al., 2000].

Numerous posttranslational modifications can additionally modulate DNMT1: the abundance and stability of DNMT1 during cell cycle is controlled by ubiquitination and acetylation [Du et al., 2010; Felle et al., 2011; Qin et al., 2011]. Phosphorylation with subsequent methylation of DNMT1 regulates stability and sumoylation was shown to enhance the catalytic activity *in vivo* [Esteve et al., 2011; Lee and Muller, 2009].

DNMT3 family

DNMT3A and DNMT3B are closely related and they show activity on unmethylated DNA and are responsible for establishing *de novo* methylation during embryogenesis and gametogenesis. While *Dnmt3a/Dnmt3b* double-knockout ESCs are unable to methylate newly introduced retroviral elements, the maintenance methylation of imprinted regions is not affected [Kaneda et al., 2004; Okano et al., 1999]. Accordingly, both DNMT3A and DNMT3B are highly expressed in ESCs as

well as germ cells and down-regulated in differentiated cells and tissues [Watanabe et al., 2002]. Knockout mice lacking either DNMT3A or DNMT3B die at different stages during embryonic development. *Dnmt3b*$^{-/-}$ mice are not viable and display several developmental defects, whereas *Dnmt3a*$^{-/-}$ mice die four weeks after birth. This indicates an essential role for DNMT3B during early developmental stages, while DNMT3A is important for methylation in later development [Okano et al., 1999]. Human patients with mutations in *DNMT3B* suffer from the ICF (immunodeficiency, centromere instability and facial abnormalities) syndrome and display methylation defects at pericentric heterochromatin and at CpG islands on the inactive X chromosome [Ehrlich et al., 1982; Miniou et al., 1994]. Recent studies revealed that mutations in the functional domains of the human *DNMT3A* gene cause overgrowth syndromes, intellectual disabilities and facial dysmorphism. The associated mutations are likely to disrupt histone binding and intramolecular interactions within DNMT3A [Tatton-Brown et al., 2014]. In addition, DNMT3A was shown to catalyze non-CpG methylation especially during neuronal maturation. First studies proposed a potential role in transcriptional repression for this special case of DNA methylation, but the exact mechanisms and functions are not fully understood [Aoki et al., 2001; Guo et al., 2014].

Both DNMT3A and DNMT3B contain an N-terminal regulatory domain linked to a C-terminal catalytic domain. The PWWP (Pro-Trp-Trp-Pro) domain was shown to bind H3K36me3, thus, enhancing methyltransferase activity and heterochromatin targeting [Dhayalan et al., 2010; Fuks et al., 2001; Ge et al., 2004]. The PHD (plant homeodomain) is essential for multiple interactions with chromatin-associated proteins including HP1, HDAC1 (histone deacetylase 1) and the histone methyltransferase SUV39H1 [Fuks et al., 2001; Fuks et al., 2003].

The third member of the DNMT family, DNMT3L, lacks some essential catalytic motifs and has no methyltransferase activity. However, DNMT3L serves as a cofactor for DNMT3A and DNMT3B and stimulates their activity via direct interaction [Chedin et al., 2002; Gowher et al., 2005; Jia et al., 2007].

1.5 New aspects of DNA methylation

DNMT1 is involved in *de novo* methylation

The classical model implies that 5mC is introduced *de novo* at both DNA strands at CpG sites by DNMT3A and DNMT3B, which results in a pattern of either fully methylated or unmethylated sites. After each round of DNA replication, methylation is still present in the parental strands generating a pattern of hemimethylated and unmethylated CpG sites. For this reason, methylation can be maintained by DNMT1 that specifically remethylates hemimethylated CpG sites (Figure 3).

During the past decade compelling experimental evidence indicated that the classical site-specific methylation model needs to be revised and expanded [Jeltsch and Jurkowska, 2014; Jones and Liang, 2009].

The *de novo* methylation only by the DNMT3 enzymes during embryogenesis is challenged by the fact that both methyltransferases show distinct and selective flanking sequence preferences, which surrounds the targeted CpG site. This sequence-specific activity leads to the generation of hemimethylated sites, which represent ideal substrates for DNMT1 [Handa and Jeltsch, 2005; Lin et al., 2002; Wienholz et al., 2010]. Additionally, DNMT3A binds to DNA in a tilted manner and, thus, cannot methylate both strands of one CpG site during one binding event [Jia et al., 2007]. As DNMT1 preferentially methylates hemimethylated CpGs, *de novo* methylation of unmethylated DNA can be accomplished more efficiently by a cooperation of DNMT3 enzymes with DNMT1 [Kim et al., 2002]. Furthermore, *de novo* methylation activity of DNMT1 was confirmed *in vitro* and *in vivo* as residual DNA methylation was observed in DNMT3A/DNMT3B double-knockout embryos [Goyal et al., 2006; Okano et al., 1999]. Recent genome-wide DNA methylation analysis compared wildtype to *Dnmt1*$^{-/-}$ cells and provided evidence that DNMT1 has considerable *de novo* methylation activity at single copy sequences and certain repetitive elements [Arand et al., 2012].

Maintenance methylation is not site-specific

The classical maintenance methylation model is based on two important requirements: DNMT1 has to work as a perfect copy machine and the methylation state of each CpG site has to be stably inherited, which would imply that all cells of the same tissue exhibit identical 5mC patterns. Indeed, genome wide methylation analysis at single base pair resolution revealed that the average methylation levels of DNA regions are maintained. However, no exact CpG site-specific methylation patterns could be observed, thus, changes in methylation densities occur through stochastic processes [Landan et al., 2012; Zhang et al., 2009]. Various biochemical studies determined that DNMT1 has a 10-40-fold preference for hemimethylated substrates [Bashtrykov et al., 2012; Bestor, 1992; Fatemi et al., 2001; Song et al., 2012]. Nevertheless, this preference is not nearly sufficient to guarantee accurate and site-specific copying of methylation patterns at all CpG sites in the genome during DNA replication.

DNMT3A and DNMT3B are involved in maintenance methylation

Knockout studies imply that DNMT1 alone is only capable of maintaining DNA methylation at CpG-poor regions but DNMT3A or DNMT3B are additionally required for methylation at repetitive elements such as the LINE1 promoter [Liang et al., 2002]. Furthermore, depletion of

DNMT3A and DNMT3B leads to a loss of methylation at repetitive elements, despite the presence of DNMT1 [Chen et al., 2003]. Together with recent genome wide studies [Arand et al., 2012], these findings clearly indicate that cooperation between DNMTs is crucial for the maintenance methylation during replication.

Complex regulation of DNMTs controls DNA methylation

The activity of DNMTs is affected by targeting to and regulation of these enzymes at particular genomic regions rather than to individual CpG sites. Thus, DNA methylation is more complex than previously assumed by the classical site-specific methylation model (Figure 3).

DNMT3 enzymes are predominantly enriched at heterochromatic regions and this recruitment is mediated by specific histone tail modifications. The PHD domain of DNMT3A and DNMT3B and their cofactor DNMT3L prevents binding of the complex to H3K4me2/3, a mark for active chromatin [Otani et al., 2009; Zhang et al., 2010c]. In addition, DNMT3A is targeted to H3K36me2/3 found inside gene bodies and heterochromatic regions via the PWWP domain [Dhayalan et al., 2010]. Lacking the catalytic domain, DNMT3L also has the potential to antagonize DNMT3A and DNMT3B activity and reduce DNA methylation at promoters of bivalent genes [Neri et al., 2013]. Since DNMT1 is associated with the replication fork and with hemimethylated sites, this targeting reduces its activity on other regions of the genome [Sharif et al., 2007]. Recently, the inhibition of DNMT1 activity by long non-coding RNAs has been shown as well [Di Ruscio et al., 2013].

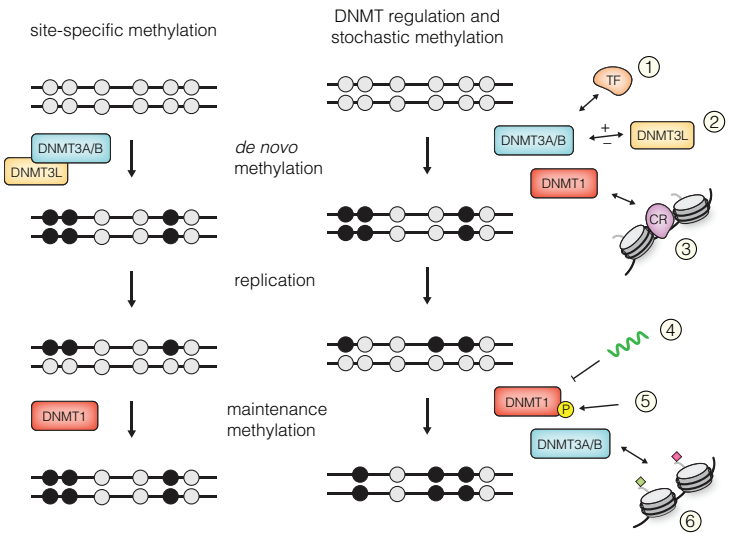


Figure 3 Classic versus stochastic methylation model

Left: In the classic methylation model fully methylated CpGs are generated by the *de novo* methyltransferases DNMT3A/B and their cofactor DNMT3L. This methylation pattern is maintained throughout DNA replication by DNMT1 in a site-specific manner.
Right: No strict separation into *de novo* and maintenance DNMTs. DNMT activity and recruitment is controlled by the interaction with other proteins such as transcription factors (1). DNMT3L can activate or repress DNMT3A/B activity (2) and the access to DNA is achieved by chromatin remodeling complexes (3). Long non-coding RNA represses DNMT1 function (4) and posttranslational modifications alter the enzymatic activity (5). Targeting of DNMTs is regulated by modifications on histone tails (6). The complex control of DNMT activity and epigenetic crosstalk leads to a more stochastic setting of methylation patterns with instructions of other epigenetic pathways.

Moreover, structural and *in vitro* methylation studies revealed that DNA wrapped around nucleosomes does not provide an ideal substrate for DNMTs [Jia et al., 2007; Takeshima et al., 2006]. This suggests that in addition to DNMT recruiting, nucleosomes have to be removed or shifted by chromatin remodeling complexes in order to allow DNA methylation in the nucleosomal core region. Maintenance methylation occurs rapidly after replication, when nucleosomes are not yet assembled and not in later stages and during *de novo* methylation. Indeed the LSH (lymphoid specific helicase) protein, which belongs to the family of SWI/SNF (switch sucrose non fermentable) chromatin remodelers, plays an important role in *de novo* DNA methylation in mice [Dennis et al., 2001; Zhu et al., 2006], supporting the cooperation between chromatin remodeling and DNA methylation.

In summary, DNA methylation is controlled by regulation and targeting of DNMTs, which include interaction with chromatin marks, posttranslational modifications, long non-coding RNAs and other factors. Furthermore, chromatin remodeling processes influence the accessibility of DNA for DNMTs. Together with the complex epigenetic crosstalk and non-CpG methylation, a more stochastic methylation model can be established, meaning that only the combined methylation density levels of larger DNA regions rather than site-specific methylation events is stably inherited.

1.6 Reading the methylation mark

How can DNA methylation in promoter regions lead to gene silencing? One possibility is that the presence of 5mC in the major groove of the DNA interferes with the binding of transcription factors that usually regulate gene expression [Choy et al., 2010; Watt and Molloy, 1988]. The second mode of repression involves proteins that are recruited to methylated CpG sites.

Methyl-CpG-binding domain (MBD) proteins

The MBD family has five known members in mammals: MBD1, MBD2, MBD3, MBD4 and MECP2 (Figure 4). MECP2 was the first characterized methyl-CpG binding protein and genomic sequencing projects and homology searches led to the identification of the other members [Becker, 2006; Hendrich and Bird, 1998; Lewis et al., 1992].

They all contain a conserved MBD, which, except for MBD3, preferentially binds to methylated DNA substrates. MECP2 is highly expressed in the nervous system and sporadic mutations of the X chromosome linked *MECP2* gene cause the neurodevelopmental Rett syndrome [Amir et al., 1999]. Recent genome wide studies revealed that MBD2 is mainly recruited to CpG island promoters that are highly methylated, whereas MBD3 localizes at promoters, gene bodies and

enhancers of active genes [Menafra et al., 2014; Shimbo et al., 2013]. Interestingly, MBD1 localization to certain chromatin regions is controlled by long non-coding RNAs [Monnier et al., 2013].

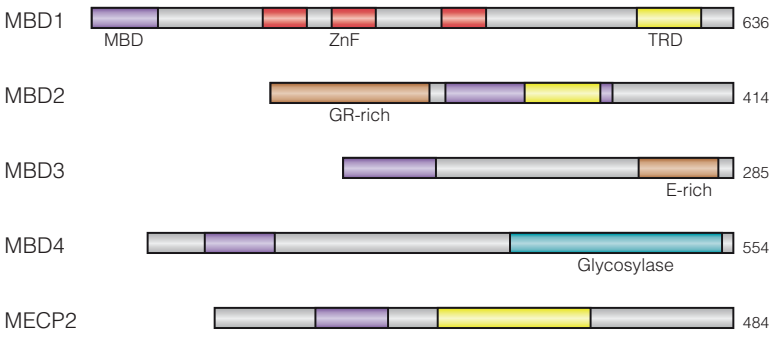


Figure 4 Domain structure of the methyl-CpG-binding domain (MBD) proteins

All MBD family members share a conserved MBD domain, which is essential for DNA binding. MBD1, MBD2 and MBD3 contain a transcriptional repression domain (TRD). MBD2 harbors glycine-arginine-rich (GR-rich) and MBD3 glutamine-rich (E-rich) repeats. MBD4 contains a glycosylase domain at the C-terminal end. Numbers indicate the length of the murine proteins in amino acids.

All members regulate transcriptional repression by indirect HDAC interactions and all except MBD4 associate with nucleosome remodeling complexes such as the NuRD complex, which generates repressive chromatin states [Jones et al., 1998; Morey et al., 2008; Nan et al., 1998]. Furthermore, MBD1 and MECP2 were shown to directly recruit histone methyltransferases like SUV39H1, which methylates lysine 9 on histone H3, thereby creating a binding site for HP1. In this manner, a dynamic connection between DNA methylation and repressive histone modifications is established by MBD proteins [Agarwal et al., 2007; Lachner et al., 2001; Sarraf and Stancheva, 2004].

Additionally, MBD1, MBD2 and MECP2 comprise a non-conserved transcription repressor domain (TRD), which in the case of MECP2 was described to mediate interaction with DNMT1 [Kimura and Shiota, 2003]. MBD2 and MBD3 were also identified to associate with DNMT1 and proposed to contribute to maintaining DNA methylation during DNA replication [Tatematsu et al., 2000]. Notably, MBD1 and MECP2 also bind to unmethylated DNA and were shown to generate condensed secondary and tertiary chromatin structures, which act as a physical barrier preventing the binding of activating transcription factors at these sites [Georgel et al., 2003; Jorgensen et al., 2004; Nikitina et al., 2007]. Interestingly, MBD4 is the only member harboring a thymine DNA glycosylase domain at the C-terminal end, which has been shown to be involved in the repair of T·G mismatches generated by deamination of 5mC [Hendrich et al., 1999].

UHRF protein family

The UHRF family, comprising the multi-domain proteins UHRF1 and UHRF2, contain an N-terminal ubiquitin-like domain (Ubl), a tandem Tudor domain (TTD), a plant homeodomain (PHD), a SET-and-RING-finger associated domain (SRA) and a C-terminal RING domain (Figure 5, top).

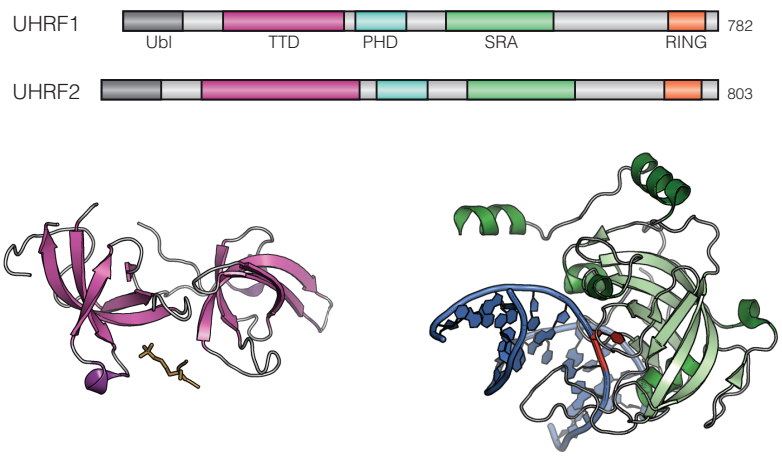


Figure 5 Domain structure of the UHRF family proteins

Top: The multi-domain proteins UHRF1 and UHRF2 contain a terminal ubiquitin-like domain (Ubl), a tandem Tudor domain (TTD), a plant homeodomain (PHD), a SET-and-RING-finger associated domain (SRA) and a RING domain. Numbers indicate the length of the murine proteins in amino acids.
Bottom: UHRF1 links histone modifications with DNA methylation. Crystal structure of the UHRF1-TTD (pink, PDB 3DB3) bound to an H3K9me3 histone peptide (brown) and SRA domain (green) bound to hemimethylated DNA (blue, PDB 3FI). 5mC (red) is flipped out from the DNA helix and inserted into a conserved pocket on the inner concave surface of the SRA domain.

UHRF1 colocalizes with PCNA during S phase, indicating a function during cell cycle progression, DNA replication or DNA damage repair [Fujimori et al., 1998; Muto et al., 2002; Uemura et al., 2000]. Furthermore, UHRF1 was shown to interact with DNMT3A, DNMT3B and several histone-modifying enzymes like HDAC1 or the histone methyltransferase G9A [Achour et al., 2009; Kim et al., 2009; Meilinger et al., 2009]. UHRF1 binds and flips out hemimethylated DNA via the SRA domain and targets DNMT1 to its substrate for maintenance DNA methylation (Figure 5, bottom) [Arita et al., 2008; Avvakumov et al., 2008; Bostick et al., 2007; Hashimoto et al., 2008]. The SRA domain of UHRF1 was described to also mediate direct interaction with DNMT1, thus, stimulating the methyltransferase activity by an allosteric mechanism. Consistently, UHRF1 not only increases the accessibility for DNA inside the catalytic center of DNMT1 but also the specificity for hemimethylated CpG sites [Bashtrykov et al., 2014; Berkyurek et al., 2014]. Additionally, UHRF1 deficient embryos display a phenotype very similar to *Dnmt1*$^{-/-}$ embryos, including genome wide DNA hypomethylation and early embryonic lethality [Sharif et al., 2007].

The involvement of UHRF1 in maintenance DNA methylation is also based on its ability to recognize H3K9me2/3 through the TTD (Figure 5, bottom), thus, connecting DNA methylation with repressive histone marks [Nady et al., 2011; Rothbart et al., 2012; Rottach et al., 2010]. The adjacent PHD of UHRF1, previously implicated in transcriptional regulation and heterochromatin organization, was shown to bind the unmodified N-terminus of histone H3 and also target DNMT1 to hemimethylated sites [Hu et al., 2011; Papait et al., 2007; Rajakumara et al., 2011; Wang et al., 2011a]. Crystal structures of the linked TTD–PHD bound to H3K9me3 peptides revealed that UHRF1 can simultaneously bind the unmodified H3 N-terminus and H3K9me3 on a single histone H3 tail through this connected recognition module [Arita et al., 2012; Cheng et al., 2013]. However, the UHRF1 TTD and PHD are also able to bind histone peptides independently

of one another *in vitro* [Hu et al., 2011; Rajakumara et al., 2011; Wang et al., 2011a]. Moreover, the UHRF1 dependent ubiquitination at H3K27 in *Xenopus* egg extracts was proposed to serve as an additional platform for the recruitment of DNMT1 to DNA replication sites [Nishiyama et al., 2013].

The second member of the UHRF family, UHRF2, shares a high structural similarity to UHRF1 [Bronner et al., 2007]. However, both members show opposite expression patterns: while UHRF1 is mainly expressed in ESCs, UHRF2 is upregulated during differentiation and highly abundant in differentiated tissues. UHRF2 displays a preference for hemimethylated DNA only in combination with binding to H3K9me3 heterochromatin marks mediated by the TTD. Interestingly, the localization and *in vivo* binding characteristics of UHRF2 were described to require an intact TTD and depend on H3K9me3 recognition but not on DNA methylation. It was therefore postulated, that the cooperative interplay of UHRF2 domains might regulate gene expression in differentiated cells. Notably, ectopic expression of UHRF2 in *Uhrf1*$^{-/-}$ ESCs was not able to restore DNA methylation at major satellites arguing that the proteins are not functionally redundant [Pichler et al., 2011].

UHRF2 was also identified to take part in the intranuclear degradation of polyglutamine aggregates and recently it was found to interact with cell cycle proteins including cyclins, cyclin-dependent kinases, retinoblastoma 1 protein (RB1), tumor protein p53 (P53), and PCNA [Iwata et al., 2009; Mori et al., 2012]. Therefore, UHRF2 might play an essential role in connecting the cell cycle with the epigenetic network.

1.7 Erasing the methylation mark

Compared with the highly dynamic histone modifications, DNA methylation is a relatively stable epigenetic mark. However, reversed DNA methylation has been observed in different biological contexts and this so-called demethylation process can occur actively or passively. Active DNA demethylation refers to the enzymatic removal or modification of the methyl group from 5mC. In contrast, passive DNA demethylation describes the loss or dilution of 5mC after repeated rounds of DNA replication in the absence or inhibition of the functional maintenance DNA methylation machinery. Although passive DNA demethylation is generally understood, the evidence for active DNA demethylation and how it is regulated has been controversial and incomplete. However, a number of recent discoveries shed new light onto the understanding of these complex processes and the underlying mechanisms.

Suggested DNA demethylation mechanisms independent of oxidation

Direct removal of the methyl group

The simplest mechanism leading to DNA demethylation is the direct enzymatic removal of the methyl group from 5mC. MBD2 was the first reported enzyme to catalyze this reaction. It was proposed that no specific cofactors were necessary and elimination of the methyl group resulted in the release of methanol [Bhattacharya et al., 1999]. To date, no compelling evidence supports that this thermodynamically unfavorable reaction could take place *in vivo*. As MBD2 can stably bind to methylated CpGs, it is unlikely that binding could occur if MBD2 was so efficient at removing the methyl group. Furthermore, MBD2 knockout mice show normal genomic methylation patterns and the paternal pronucleus of *Mbd2*$^{-/-}$ zygotes still undergoes normal demethylation [Hendrich et al., 2001; Santos et al., 2002]. Since no other laboratories were able to reproduce this specific MDB2 activity, serious doubts have been raised whether this protein is really involved in active demethylation.

Radical SAM mechanism

ELP3 (elongator complex protein 3) was described to be involved during active demethylation of the paternal genome in mouse zygotes [Okada et al., 2010]. Immunostaining and bisulphite sequencing of selected retrotransposons showed a potential role for ELP3 in demethylation. It was suggested that the radical SAM domain of ELP3 may be involved in the direct removal of the methyl group but clear biochemical evidence confirming this activity is still missing.

Nucleotide excision of 5mC

The repair of short genomic regions that contain methylated CpG sites can indirectly lead to active removal of 5mC. Nucleotide excision repair (NER) is mainly involved in repairing DNA containing bulky lesions, which result from exposure to radiation or chemicals. After the damaged DNA site is recognized, dual incisions flanking the lesion are made by specific glycosylases and a 24–32 oligonucleotide is excised. The resulting gap on the single-stranded DNA is then filled in by polymerases and sealed by a ligase. The GADD45 (growth arrest and DNA-damage-inducible protein 45) family of proteins was reported to stimulate active DNA demethylation via NER. Overexpression of GADD45A in mammalian cell lines resulted in loci-specific as well as global demethylation, whereas knockdown led to DNA hypermethylation [Barreto et al., 2007]. Since GADD45A has previously been implicated in NER, loss of DNA methylation could be accompanied by DNA synthesis and requires the NER endonuclease XPG (xeroderma pigmentosum group G-complementing protein), which directly interacts with GADD45A

[Barreto et al., 2007; Zhan, 2005]. However, it is not fully understood how this NER-based demethylation pathway is initiated and whether GADD45A is directly involved during this process. Furthermore, two independent studies have raised doubts on the role of GADD45A since a more detailed analysis of the *Gadd45a*$^{-/-}$ mice showed neither loci-specific nor global effects on DNA methylation levels [Engel et al., 2009; Jin et al., 2008].

Base excision of 5mC or T·G mismatches

Active DNA demethylation could also be accomplished by direct removal of 5mC by base excision repair (BER). Strong genetic and biochemical evidence support the use of this pathway in plants [Bauer and Fischer, 2011; Zhu, 2009]. This type of repair is initiated by a DNA glycosylase that recognizes and excises the target base resulting in an abasic (apurinic and apyrimidinic (AP)) site. The DNA backbone is subsequently nicked by an AP endonuclease, which removes the 3′ sugar group leaving a single nucleotide gap that is ultimately filled in by DNA polymerase β and DNA-Ligase III [Fortini and Dogliotti, 2007]. Although it is clear that plants use BER to directly remove 5mC, evidence for a similar mechanism in vertebrates has been less compelling. The first indication that BER could contribute to DNA demethylation came from chicken embryo extracts, showing 5mC glycosylase activity against hemimethylated DNA [Jost, 1993; Jost et al., 1995]. Subsequent purification of this activity revealed that a homologue of mammalian thymine DNA glycosylase (TDG) was important for this process [Jost et al., 1999; Zhu et al., 2000b]. However, its excision activity towards 5mC is about 40-fold lower compared with that towards T and although TDG can flip C and T analogues into its active site, it is not able to lyse the N-glycosidic bond [Bennett et al., 2006]. In addition to TDG, MBD4 was also proposed to show glycosylase activity against 5mC, but again this activity is about 40-fold lower than towards T·G mismatches [Zhu et al., 2000a]. In accordance, *Mbd4*$^{-/-}$ zygotes display normal demethylation of the zygotic paternal pronucleus and *Mbd4* knockout mice are viable and show normal DNA methylation pattern during development [Millar et al., 2002]. In contrast, TDG knockout mice die during embryonic development and exhibit modestly increased 5mC levels at some CpG-rich gene promoters [Cortazar et al., 2011; Cortellino et al., 2011].

Spontaneous deamination of cytosine generates uracil and the resulting mismatch can also be repaired by BER. The APOBEC family of cytidine deaminases are known to directly participate in generating mutations in RNA, which contribute to antibody diversification in B cells [Conticello, 2008]. Previous *in vitro* studies showed that the APOBEC member AID (activation-induced deaminase) additionally converts 5mC to T in DNA [Morgan et al., 2004]. However, AID is only active on single-stranded DNA and displays up to 20-fold lower activity on 5mC relative to

cytosine, its canonical substrate [Bransteitter et al., 2003; Nabel et al., 2012]. AID knockout mice exhibit the expected B cell and immunological defects, but are viable and fertile [Morrison et al., 1996]. Nevertheless, studies in mouse primordial germ cells (PGC), mouse ESC/human fibroblast fused heterokaryons and induced pluripotent stem cells (iPSCs), support a potential role of AID-mediated 5mC deamination in DNA demethylation [Bhutani et al., 2010; Kumar et al., 2013; Popp et al., 2010].

Notably, DNMT3A and DNMT3B have been proposed to deaminate 5mC *in vitro* in the absence of the cofactor SAM [Metivier et al., 2008]. As SAM is abundant *in vivo* the DNMT-mediated deamination of 5mC remains elusive.

TET proteins and oxidized cytosines

In 1972, the presence of 5-hydroxymethylcytosine (5hmC), an oxidized derivative of 5mC, was detected in the mammalian genome [Penn et al., 1972]. However it took almost 40 years until the Ten-eleven-translocation (TET) family of Fe(II)-2-oxoglutarate dependent dioxygenases were identified as the enzymes that convert 5mC to 5hmC [Kriaucionis and Heintz, 2009; Tahiliani et al., 2009]. Subsequent studies showed that TET proteins can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC, Figure 6 and Figure 8) [Ito et al., 2011; Pfaffeneder et al., 2011]. These groundbreaking discoveries have highly contributed to the understanding of the molecular mechanisms leading to active DNA demethylation in mammals.

TET1 was initially discovered as a fusion partner of the histone H3 lysine 4 methyltransferase MLL1 (Mixed lineage leukemia 1) in acute myeloid leukemia [Lorsbach et al., 2003; Ono et al., 2002]. First experiments revealed that overexpression of TET1 leads to decreased genomic 5mC levels and recombinant TET1 protein can oxidize 5mC *in vitro* generating 5hmC [Tahiliani et al., 2009]. Similar enzymatic activity was also demonstrated for TET2 and TET3, the two other members of the TET protein family [Ito et al., 2010]. The three TET proteins are expressed differently in a developmental-stage- and cell-type-specific manner. Whereas TET1 is specific for ESCs, the inner cell mass of blastocysts, and developing PGCs, TET2 and TET3 are broadly expressed in various adult tissues [Ito et al., 2010; Szwagierczak et al., 2010; Yamaguchi et al., 2012]. Interestingly, TET3 is the only member present in mouse oocytes and zygotes at one-cell stage [Iqbal et al., 2011; Wossidlo et al., 2011].

All TET family members contain a C-terminal catalytic domain, which includes a cysteine-rich (Cys-rich) domain and the double-stranded beta helix (DSBH) with the binding sites for the cofactors Fe(II) and 2-oxoglutarate (2-OG). An unstructured insert that shows great variation between TET family members separates the DSBH domain, but its exact function remains elusive

(Figure 6, top). A recent crystal structure of the TET2 catalytic domain in complex with DNA shows that the Cys-rich domain stabilizes the DNA above the DSBH core. Furthermore, TET2 specifically recognizes CpG dinucleotides with 5mC inserted into the catalytic cavity using a base-flipping mechanism (Figure, 6 bottom). As the methyl group is not involved in the TET2-DNA interaction, the composition of the catalytic cavity allows TET2 to accommodate 5mC-oxidized derivatives for further oxidation [Hu et al., 2013]. In addition, *Tet1* and a long splicing variant of *Tet3* encode a CXXC-zinc finger domain at their N-termini [Iyer et al., 2009; Liu et al., 2013b]. The CXXC domain is present in many chromatin-associated proteins and has a strong preference for unmethylated CpGs [Long et al., 2013]. Although TET2 lacks a CXXC domain, a neighboring gene *Idax* (inhibition of the Dvl and axin complex) encodes for a CXXC domain similar to those in TET1 and TET3. IDAX is thought to have been originally part of an ancestral *Tet2* gene that underwent a chromosomal inversion, which separated the CXXC domain from the catalytic domain. It was shown that IDAX directly interacts with TET2 and is enriched at unmethylated CpG sites [Iyer et al., 2009; Ko et al., 2013].

The isolated CXXC domain of TET1 was first described to have no DNA binding activity on its own and dispensable for catalytic activity *in vivo* [Frauer et al., 2011b]. However, further biochemical and structural analyses indicate that an extended TET1 CXXC domain may also recognize methylated cytosines and the TET3 CXXC domain can target unmethylated cytosines within both CpG and non-CpG contexts [Xu et al., 2011b; Xu et al., 2012; Zhang et al., 2010a]. Thus, unlike other CXXC domains that only bind to unmethylated CpG sites, CXXC domains of TET1 or TET3 and IDAX are more flexible in sequence selectivity and might recruit TET enzymes to their specific genomic targets.

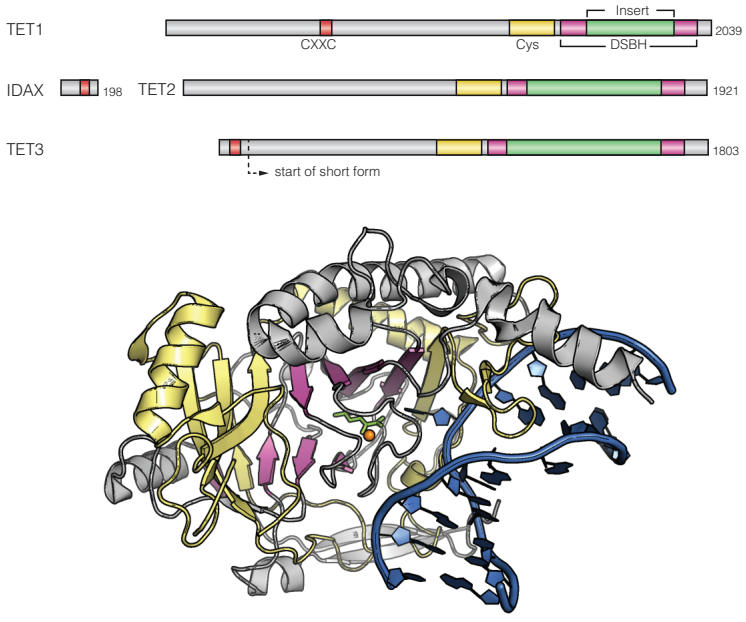


Figure 6 Domain composition of TET proteins and crystal structure of human TET2 catalytic domain bound to DNA

Top: The C-terminal catalytic domain of the three TET proteins contains a cysteine-rich domain (Cys) and a double-stranded beta-helix (DSBH) domain, which includes a large insert. TET1 and a long splicing variant of TET3 also inherit a N-terminal CXXC domain. During evolution a chromosomal inversion detached the catalytic domain of TET2 from its CXXC domain, which became a separate gene encoding IDAX. Numbers indicate the length of the murine proteins in amino acids.

Bottom: Crystal structure of human TET2 catalytic domain bound to a methylated CpG containing substrate (PDB 4NM6). The target cytosine is flipped out of the DNA helix (blue) inside the core of the DSBH consisting of eight antiparallel beta-sheets (purple). The Cys-rich region (yellow) wraps around the DSBH. Crystallization was preformed with the 2-OG analog N-oxalylglycine (green) and iron (orange). Due to its low-complexity the insert was not crystallized and the two parts of the DSBH domain directly connected by a linker.

5hmC can be detected in most cell-types (1%-5% of 5mC) and is particularly enriched in the genome of adult neurons (up to 40% of 5mC) [Globisch et al., 2010; Kriaucionis and Heintz, 2009; Szwagierczak et al., 2010]. With improved detection methods, TET proteins have been shown to be capable of further oxidizing 5hmC to 5fC and 5caC. Mass spectrometry analysis revealed that 5fC is present in various tissue types and 5caC is detectable in mouse ESCs, although their levels are at least an order of magnitude less than that of 5hmC [Ito et al., 2011; Pfaffeneder et al., 2011].

Passive demethylation by replication-dependent dilution of oxidized cytosines

The presence of oxidized 5mC bases at CpG sites may contribute to passive replication-dependent loss of 5mC. While there are conflicting results whether UHRF1 can specifically recognize hemihydroxymethylated CpG sites, DNMT1 has been shown to be less efficient in methylating hemihydroxymethylated CpGs than hemimethylated sites *in vitro* [Frauer et al., 2011b; Hashimoto et al., 2012b]. Thus, TET proteins may initiate a two-step demethylation process in dividing cells: The initial oxidation of 5mC could be subsequently followed by a replication-dependent passive dilution of 5hmC or potentially 5fC and 5caC. This mode of active DNA demethylation is distinct from simple passive dilution of 5mC, as this oxidation-based loss may be effective even in the presence of a functional methylation maintenance machinery.

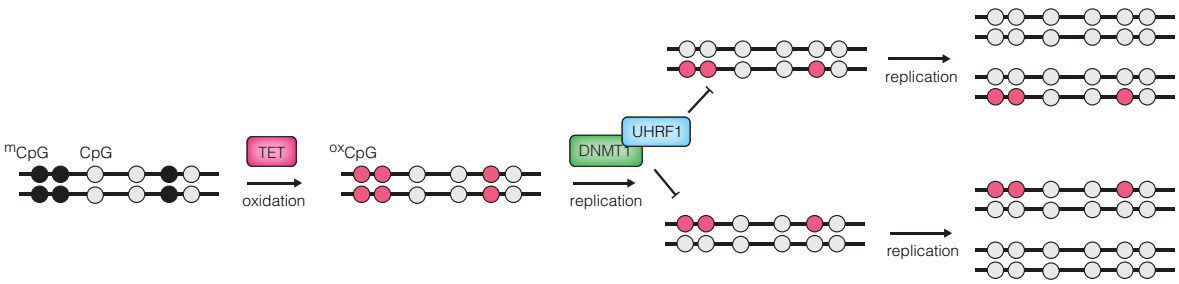


Figure 7 Passive DNA demethylation after oxidation by TET proteins

TET proteins oxidize methylated CpG sites (^{m}CpG) to generate symmetrically oxidized CpGs (^{ox}CpG). During DNA replication, the oxidized cytosines may interfere with maintenance methylation by inhibiting UHRF1 binding or DNMT1 activity. Thus, the CpG sites progressively lose methylation through successive DNA replication cycles.

Active demethylation by base excision repair

Two replication-independent demethylation mechanisms have been proposed, which couple iterative oxidation of methylcytosine with BER (Figure 8). In the first mechanism, TET proteins further oxidize 5hmC to generate 5fC or 5caC, which can be excised by TDG and their

subsequent replacement with unmodified cytosine results in demethylation [He et al., 2011; Maiti and Drohat, 2011]. TDG is not able to efficiently remove 5mC or 5hmC and structural analyses indicate that TDG binds 5fC·G or 5caC·G mismatches with higher affinity than T·G mismatches. Depletion of TDG causes 2–10-fold increased 5fC and 5caC levels in ESCs, consistent with the fact that these bases are demethylation intermediates that can be excised by TDG [Shen et al., 2013; Song et al., 2013; Zhang et al., 2012]. However, even in TDG-deficient cells, 5fC is rare compared with 5mC (0.2% - 0.3% of 5mC), and 5caC is even less abundant indicating that there might be additional active demethylation mechanisms using other glycosylases.

In the second proposed mechanism, 5hmC is deaminated to 5-hydroxyuracil (5hmU) by AID. Subsequently, 5hmU could be removed by single-strand-selective monofunctional uracil DNA glycosylase 1 (SMUG1), TDG or MBD4 and ultimately replaced by cytosine (Figure 8) [Guo et al., 2011a]. In support of such mechanisms, TDG and MBD4 can excise 5hmU·G mismatches *in vitro*. Furthermore, PGCs derived from AID- or TDG-deficient mice exhibit modestly increased levels of methylation at some CpG island promoters [Cortazar et al., 2011; Cortellino et al., 2011; Hashimoto et al., 2012c; Nabel et al., 2012]. Against a deamination based mechanism argues that the AID enzyme primarily acts on single-stranded DNA and APOBEC enzymes display no detectable activity on 5hmC [Nabel et al., 2012; Rangam et al., 2012]. Therefore, it seems unlikely that AID and APOBEC enzymes play a role in 5hmC-dependent demethylation pathways, although their involvement under specific conditions cannot be ruled out.

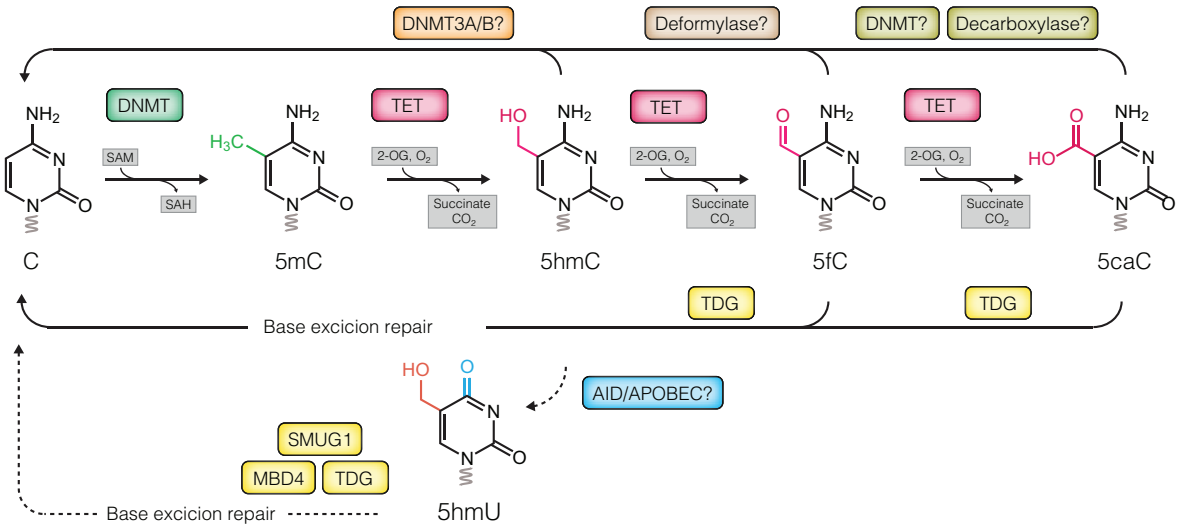


Figure 8 Oxidized cytosines and potential active demethylation pathways

The cytosine modification pathway starts with DNMTs, which use SAM as a methyl donor to catalyze methylation at the 5-position of cytosine, yielding S-adenosylhomocysteine (SAH). TET proteins subsequently oxidize 5mC to 5hmC, 5fC and 5caC using Fe(II), 2-OG and O_2. 5fC and 5caC can be removed by TDG and replaced by cytosine via BER. Deamination of 5hmC leads to the formation of hydroxymethyluracil (5hmU). 5hmU:G mismatches are proposed to be excised by SMUG1, TDG or MBD4 glycosylases. Other direct mechanisms are less well established, including dehydroxymethylation by DNMT3A and DNMT3B or decarboxylation.

Enzymatic removal of the oxidized methyl group

Decarboxylation of 5caC

Direct decarboxylation of 5caC to cytosine without BER was described in ESC lysates. This observation is based on oligonucleotides containing 5caC, isotopically labeled with ^{15}N at both positions of the pyrimidine ring. After incubation with the lysates and tracing of the labeled oligonucleotide, a small but detectable amount of [$^{15}N_2$]-dC was measured [Schiesser et al., 2012]. However, the factor(s) that catalyze this decarboxylation reaction have not been identified. Recently, DNMT enzymes have been shown to decarboxylate 5caC containing DNA substrates *in vitro*, but this specific activity is unlikely under reducing cellular conditions *in vivo* [Liutkeviciute et al., 2014].

Dehydroxymethylation by DNMT3 enzymes

DNMT3A and DNMT3B enzymes were shown to directly remove the hydroxymethyl group of 5hmC *in vitro,* whereas this catalytic activity was not observed for DNMT1 [Chen et al., 2012]. Reducing conditions favored the methyltransferase activity of DNMT3A and oxidizing conditions enhanced dehydroxymethylation of double-stranded DNA oligonucleotides. Whether this redox-state dependent reaction occurs in living mammalian cells, which contain a reducing environment is still unknown. Furthermore, additional structural and functional data of DNMT3A/B catalyzing this specific dehydroxymethylation reaction were not obtained so far.

Biological functions of DNA demethylation

Global erasure of 5mC occurs at specific stages of mammalian development, including early preimplantation embryos and developing germ cells. Uncovering the biochemical mechanisms of DNA demethylation and the emergence of highly sensitive technologies for mapping cytosine modifications allows studying the dynamic DNA demethylation pathways in biological processes.

DNA demethylation dynamics during preimplantation development

After the sperm fertilizes the oocyte and before the two pronuclei merge, the paternal genome undergoes an epigenetic remodeling process, which includes global DNA demethylation [Mayer et al., 2000]. In contrast, the maternal genome is not affected in one-cell zygotes, but gradually loses 5mC during subsequent cell divisions as DNMT1 is actively excluded from the nucleus (Figure 9) [Cardoso and Leonhardt, 1999]. Loss of genome-wide paternal DNA methylation concurs with a rapid increase in 5hmC, 5fC and 5caC, suggesting that TET-mediated 5mC oxidation contributes to demethylation [Inoue et al., 2011; Inoue and Zhang, 2011; Iqbal et al., 2011].

Deletion of TET3, the only TET member present at this stage, abolishes the loss of 5mC in the male pronucleus [Gu et al., 2011; Wossidlo et al., 2011]. After the paternal and the maternal pronuclei fuse, bulk 5hmC, 5fC and 5caC in sperm-derived chromosomes and 5mC in oocyte-derived chromosomes are lost by passive demethylation [Inoue et al., 2011; Inoue and Zhang, 2011]. However, several maternally derived methylated CpG-rich regions, including maternal imprinting control regions (ICRs) remain fully or partially methylated [Smallwood et al., 2011; Smith et al., 2012]. Recent evidence shows that persistence of 5mC at ICRs is mediated by recruitment of DNMT1 and DNMT3A to these regions by zinc finger transcription factor ZFP57 [Li et al., 2008].

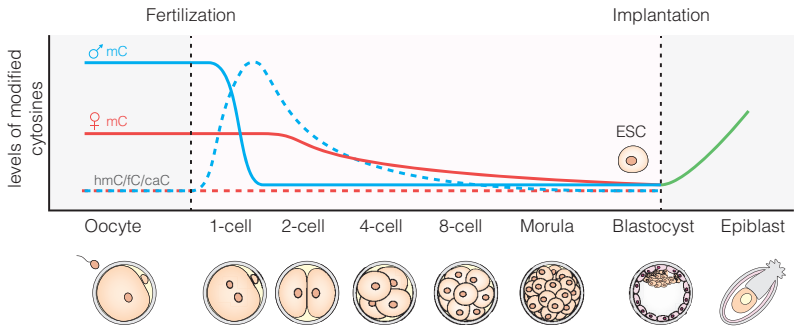


Figure 9 DNA methylation and demethylation dynamics in pre-implantation embryos

Immediately after fertilization, paternal 5mC is rapidly oxidized by TET3. Oxidized 5mC levels in the paternal genome and 5mC in the maternal genome are diluted through passive demethylation. Blastocysts and ESCs show lowest mC levels. After implantation, DNA methylation is rapidly re-established. Modified from [Wu and Zhang, 2014]

TET proteins in pluripotency and reprogramming

ESCs, which are obtained from the inner cell mass of the blastocyst, are characterized by their pluripotency and ability for self-renewal. Under defined conditions, they can be expanded indefinitely in culture and maintain their full developmental potential.

Current data indicate that, despite the presence of relatively high expression levels of TET1 and TET2 proteins in ESCs, both enzymes are largely dispensable for ESC maintenance but may play a role in guiding ESCs to properly differentiate into defined lineages [Koh et al., 2011]. Indeed, while *Tet1*$^{-/-}$ and *Tet2*$^{-/-}$ single-mutant mice are viable, roughly half of *Tet1*/*Tet2* double- knockout mice die perinatally with severe developmental defects, indicating a key function of TET1/2 proteins in regulating embryonic development. Interestingly, some double-mutant embryos survived to normal and fertile adult mice, which could be explained by partial compensation through TET3 [Dawlaty et al., 2013; Dawlaty et al., 2011; Li et al., 2011]. In line, *Tet3*$^{-/-}$ single mutant mice die perinatally [Gu et al., 2011].

Although the exact functions of TET proteins in ESC self-renewal needs to be further studied, several recent publications indicate that they are involved in reprogramming of somatic cells to generate induced pluripotent stem cells (iPSCs). The transfection of differentiated cells with

specific combinations of transcription factors can reprogram them into iPSCs that show phenotypically similar features as pluripotent ESCs. The classic set of transcription factors (TFs) used in initial experiments was SOX2, OCT4, KLF4 and MYC, but other combinations have also been applied successfully [Takahashi and Yamanaka, 2006]. Initial studies showed that TET2 is recruited to the *Nanog* and *Esrrb* loci, which both are pluripotency-related TFs, to enhance their transcription in the early stage of reprogramming [Doege et al., 2012]. Furthermore, TET1 and TET2 are proposed to directly interact with NANOG and promote iPSC generation based on the enzymatic activity of TET [Costa et al., 2013]. Interestingly, TET1 overexpression not only promotes transcriptional reactivation of *Oct4*, but also can replace OCT4 as a reprogramming factor [Gao et al., 2013]. Recent reprogramming experiments also highlight the role of ascorbic acid (vitamin C) in modulating TET functions during iPSC generation and improvement of iPSC quality [Chen et al., 2013a; Stadtfeld et al., 2012]. In summary, the ability of TET proteins to generate oxidized methylcytosines and demethylate selected regions may help to establish the pluripotent state and also support the reprogramming of somatic cells to iPSCs.

TET proteins and cancer

Aberrant DNA methylation is one of the hallmarks associated with cancer cells [Baylin and Jones, 2011]. Initially, human *TET1* was identified as a rare fusion partner of MLL in patients with acute myeloid leukemia (AML) [Ono et al., 2002]. Subsequent studies clearly pointed out a connection between mutations in the *TET2* gene and various myelodysplastic disorders [Ko et al., 2010; Konstandin et al., 2011; Langemeijer et al., 2009]. Loss of TET2 in mice causes a dysregulation and enhanced self-renewal of hematopoietic stem cells, indicating a potential role as a tumor suppressor for leukemia [Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011]. Interestingly, most TET2 mutations are mutually exclusive with neomorphic mutations in IDH1/2 (isocitrate deydrogenase 1/2) in AML. Wild-type IDH1/2 converts isocitrate to 2-OG during the citric acid cycle, whereas mutant IDH produce 2-hydroxyglutarate (2-HG), an oncometabolite that competitively inhibits 2-OG-dependent dioxygenases, including TET proteins [Dang et al., 2010; Figueroa et al., 2010; Ward et al., 2010; Xu et al., 2011a]. Thus, mutations in citric acid cycle enzymes could result in inactivation of TET proteins and alteration of methylation patterns in cancer cells. In line with that, transcriptional downregulation of TET proteins and *IDH1/2* mutations can be identified in many solid cancers and global loss of oxidized cytosine derivatives could be used as a diagnostic biomarker for human melanoma, breast, liver, and lung cancer [Lian et al., 2012; Yang et al., 2013]. Multiple mechanisms may affect TET-mediated 5mC oxidation, but most molecular targets of TET proteins that are causally linked to cancer development and progression still remain elusive.

1.8 Aims of this work

In mammals, DNA methylation plays a central role in the epigenetic regulation of gene expression and it is crucial for maintaining genomic stability. DNA methylation was initially thought to be a relatively stable mark, however, the idea evolved that this modification is subject to dynamic changes in response to different stimuli. In 2009, the TET family of proteins was discovered, which catalyze the oxidation of 5mC to 5hmC, 5fC and 5caC, potential intermediates of DNA demethylation.

The first objective of this thesis was to screen for reader proteins of 5mC and its oxidized derivatives in ESCs, NPCs and adult brain tissue. For this aim, quantitative mass spectrometry based proteomics was applied and the functions of specific readers were further characterized with structural and biochemical methods.

After the discovery of the TET proteins, several BER-based active DNA demethylation mechanisms have been proposed. One controversially discussed pathway also involves the conversion of 5hmC to 5hmU by specific deaminases. To elucidate whether 5hmU is really present in genomic DNA, high sensitive mass-spectrometry was performed. Furthermore, isotope-tracing experiments were used to address the question, which enzyme is responsible for the generation of 5hmU.

Although TET enzymes have been intensively investigated over the last five years only little is known how they are post-translationally modified. To investigate the phosphorylation and O-GlcNAcylation status, mass-spectrometry was applied to obtain the modification status of all three TET proteins at amino acid resolution.

Moreover, the function of the oxidized cytosine derivatives with regard to gene expression was analyzed. For this, an *in vitro* oxidation of a reporter gene was combined with the high-throughput expression analysis *in vivo*. Finally, a new glycosylase family was identified to be capable of serving as a backup for TDG, which was the main glycosylase involved in excision of the oxidized cytosines.

2 RESULTS

2.1 Dynamic readers for 5-(Hydroxy)Methylcytosine and its oxidized derivatives

Cell

Resource

Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives

Cornelia G. Spruijt,[1,9] Felix Gnerlich,[2,9] Arne H. Smits,[1] Toni Pfaffeneder,[2] Pascal W.T.C. Jansen,[1] Christina Bauer,[3] Martin Münzel,[2] Mirko Wagner,[2] Markus Müller,[2] Fariha Khan,[4,5] H. Christian Eberl,[6] Anneloes Mensinga,[1] Arie B. Brinkman,[7] Konstantin Lephikov,[8] Udo Müller,[3] Jörn Walter,[8] Rolf Boelens,[5] Hugo van Ingen,[5] Heinrich Leonhardt,[3] Thomas Carell,[2,*] and Michiel Vermeulen[1,*]

[1]Department of Molecular Cancer Research, Proteomics and Chromatin Biology, UMC Utrecht, 3584 CG Utrecht, the Netherlands
[2]Center for Integrated Protein Science at the Fakultät für Chemie und Pharmazie, Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[3]Center for Integrated Protein Science at the Fakultät für Biologie, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany
[4]University Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Rawalpindi, Pakistan
[5]NMR Spectroscopy Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands
[6]Proteomics and Signal Transduction, Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany
[7]Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands
[8]Genetik/Epigenetik, Universität des Saarlandes, 66123 Saarbrücken, Germany
[9]These authors contributed equally to this work
*Correspondence: thomas.carell@lmu.de (T.C.), m.vermeulen-3@umcutrecht.nl (M.V.)
http://dx.doi.org/10.1016/j.cell.2013.02.004

SUMMARY

Tet proteins oxidize 5-methylcytosine (mC) to generate 5-hydroxymethyl (hmC), 5-formyl (fC), and 5-carboxylcytosine (caC). The exact function of these oxidative cytosine bases remains elusive. We applied quantitative mass-spectrometry-based proteomics to identify readers for mC and hmC in mouse embryonic stem cells (mESC), neuronal progenitor cells (NPC), and adult mouse brain tissue. Readers for these modifications are only partially overlapping, and some readers, such as Rfx proteins, display strong specificity. Interactions are dynamic during differentiation, as for example evidenced by the mESC-specific binding of Klf4 to mC and the NPC-specific binding of Uhrf2 to hmC, suggesting specific biological roles for mC and hmC. Oxidized derivatives of mC recruit distinct transcription regulators as well as a large number of DNA repair proteins in mouse ES cells, implicating the DNA damage response as a major player in active DNA demethylation.

INTRODUCTION

Methylation of cytosine residues at carbon atom 5 of the base (mC) represents a major mechanism via which cells can silence genes. Cytosine methylation mostly occurs in a CpG dinucleotide context. However, CpG islands (CGIs), which are characterized by a very high CpG density and are often found in promoter regions of genes, are typically hypomethylated. Methylation of these CGIs results in transcriptional silencing. The molecular mechanisms underlying the association between DNA methylation and repression of transcription have proven difficult to decipher. The classic view is that methylation of DNA results in the recruitment of methyl-CpG-binding proteins (MBPs) that possess transcriptionally repressive enzymatic activities (Defossez and Stancheva, 2011). However, in vivo validation for this model on a genome-wide level is still lacking. In contrast, recent in vivo data have revealed that CXXC-domain-containing proteins specifically bind to nonmethylated cytosines. In this case, hypomethylated CGIs serve as a recruitment signal for CXXC-domain-containing activators that establish a transcriptionally active chromatin state (Thomson et al., 2010).

It was discovered 4 years ago that Tet enzymes convert mC to 5-hydroxymethylcytosine (hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This modification is particularly abundant in the brain and in embryonic stem cells but is detectable in all tissues tested (Globisch et al., 2010; Szwagierczak et al., 2010). Tet enzymes can catalyze further oxidation of hmC to 5-formylcytosine (fC) and 5-carboxylcytosine (caC) (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). fC and caC can subsequently serve as substrates for thymine-DNA glycosylase (Tdg), which eventually results in the generation of a nonmethylated cytosine (He et al., 2011; Maiti and Drohat, 2011). Therefore, this Tet-Tdg pathway represents an active DNA demethylation pathway. It is not clear whether hmC, fC, and caC have additional DNA-demethylation-independent

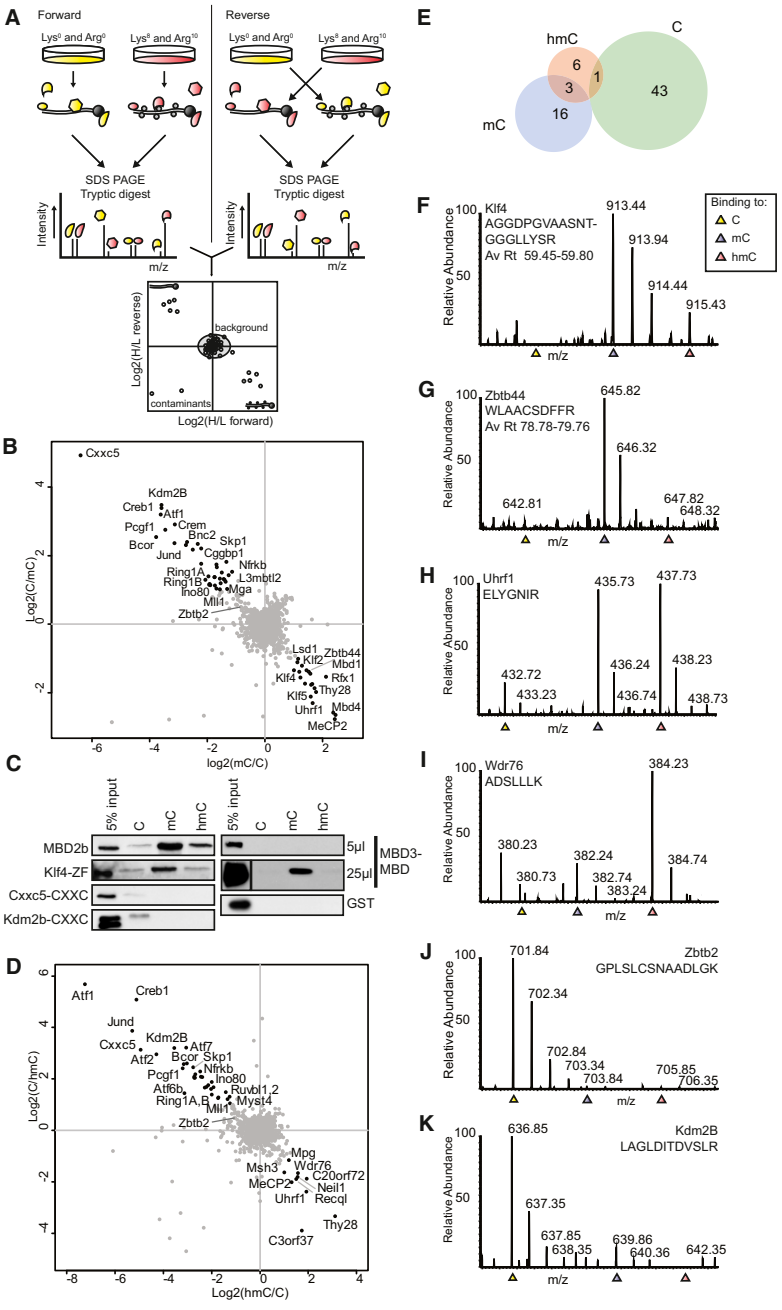


Figure 1. Identification of mC- and hmC-Specific Readers in Mouse Embryonic Stem Cells

(A) Schematic overview of the workflow.

(B) Scatterplot of a SILAC-based mC DNA pull-down in mESC nuclear extracts.

(C) Validation of the mC-specific binding of Klf4 and nonmethyl-C-specific binding of Cxxc5 and Kdm2b. DNA pull-downs were performed with recombinant GST-fusion proteins followed by western blotting. For MBD3_25, an empty lane was cut out.

(D) Scatterplot of a SILAC-based hmC DNA pull-down in mESC nuclear extract.

(E) Venn diagram showing overlap of readers for C, mC, and hmC.

(F–L) Representative mass spectra obtained in the triple-SILAC DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for nonmethylated (yellow), methylated (blue), and hydroxymethylated (red) DNA.

See also Figure S1 and Table S1.

functions, as very few specific binders, or "readers," for these oxidized versions of mC have been described thus far.

We applied quantitative mass spectrometry (MS)-based proteomics to identify a large number of readers for mC and its oxidized derivatives in mouse embryonic stem cells (mESCs). Furthermore, we also identified readers for mC and hmC in neuronal progenitor cells (NPCs) and in adult mouse brain. Our data reveal that each cytosine modification recruits a distinct and dynamic set of proteins. The known biology of these interacting proteins suggests a role for hmC, fC, and caC in active DNA demethylation pathways via base excision repair (BER), as well as an epigenetic recruitment function in certain cell types.

RESULTS

Identification of mC and hmC Readers in mESCs

To identify readers for methylcytosine and its oxidized derivatives, we made use of a DNA pull-down approach combined with quantitative MS. In brief, nuclear extracts from mESCs grown in "light" or "heavy" SILAC medium were incubated with a nonmodified or modified double-stranded DNA sequence (5′-AAG.ATG.ATG.AXG.AXG.AXG.AXG.ATG.ATG-3′, with X representing C, mC, or hmC; "forward" pull-down; Figure 1A). As a control, a label-swap, or "reverse," experiment was performed. Following incubation and washes, beads were combined and bound proteins were in-gel digested with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Raw MS data were analyzed using MaxQuant (Cox and Mann, 2008). Specific interactors are distinguishable from background proteins by their H/L ratio. Proteins binding selectively to the modified DNA have a high ratio in the forward pull-down and a low ratio in the reverse pull-down, whereas readers for the nonmodified DNA show opposite

binding (low forward ratio, high reverse ratio). Background proteins will have a ~1:1 ratio in both pull-downs (Figure 1A).

As shown in Figure 1B and Table S1 available online, we identified 19 proteins enriched for mC compared to C in mESC nuclear extracts ($p < 0.05$ and ratio >2 in both pull-downs). Among these are the methyl-CpG-binding proteins MeCP2, Mbd1, Mbd4, and Uhrf1 (Defossez and Stancheva, 2011). Other interactors include Rfx1 and Zfhx3, which were previously identified as mC readers (Bartke et al., 2010; Sengupta et al., 1999). Interestingly, three Klf proteins were identified as mC readers: Klf2, -4, and -5. These proteins carry three Krüppel-like zinc fingers, just like the Kaiso family of mC-binding proteins. Klf4 is one of the four Yamanaka reprogramming factors and has not been previously identified as a mC-binding protein in HeLa or U937 cells (Bartels et al., 2011; Bartke et al., 2010). This may be due to the low expression of Klf4 in differentiated cells relative to mESCs. We confirmed the direct binding of the Klf4 Krüppel-like zinc fingers to mC using recombinant protein and two different DNA sequences (Figure 1C and S1A). A motif bearing similarities to a recently published consensus binding site for Klf4, as determined by ChIP-seq (GGGXGTG) (Chen et al., 2008), revealed that Klf4 binds this motif with the highest affinity when "X" is mC (Figure S1A). These results establish Klf4 as a sequence-specific mC binding protein.

Mining published bisulfite sequencing data of mESCs and NPCs (Stadler et al., 2012) and overlapping this data with the Klf4 ChIP-seq profile in mESCs (Chen et al., 2008) revealed a substantial number of methylated Klf4-binding sites in this cell type (Figure S1B), which are mainly intronic and intergenic (Figure S1C). Out of the 7,321 Klf4-binding sites in mESCs that were covered in the bisulfite sequencing data set, 1,356 show high levels of DNA methylation in mESCs (18.5%). Many of these Klf4-binding sites contain a methylated Klf4-binding motif, such as GGCGTG (Figures S1D and S1E). Interestingly, many Klf4-binding sites that are nonmethylated in ES cells become hypermethylated in NPC cells (Stadler et al., 2012) (Figures S1B and S1D). This finding may be highly relevant in the context of Klf4-mediated cellular reprogramming. During reprogramming, Klf4 may be able to bind these methylated loci in differentiated cells to initiate stem-cell-specific gene expression patterns. Enrichment analyses for functional domains among the mC interactors revealed DNA-binding zinc fingers to be significantly enriched (Benj.Hoch.FDR = $10^{-2.45}$; Figure S3A). These zinc fingers may also interact with the methylated DNA in a sequence-specific manner.

In addition to the cluster of mC-binding proteins, a large number of proteins displayed preferential binding to nonmethylated DNA (Figure 1B, upper-left quadrant). Consistent with previous observations, this cluster of proteins contains a number of CXXC-domain-containing proteins that are known to preferentially bind to nonmethylated CpGs (Blackledge et al., 2010; Thomson et al., 2010). Examples include Cxxc5, Kdm2b, and Mll1 (see Figure 1C). We also identified other subunits of the Mll1 and PRC1.1 (Bcor) complexes, which most likely bind to the nonmethylated DNA indirectly via Mll1 and Kdm2b, respectively. Other interactors include the Ino80 chromatin-remodeling complex and zinc-finger-containing transcription factors such as Zbtb2, as well as basic leucine zipper-containing proteins (enriched Benj.Hoch.FDR = $10^{-5.57}$; Figure S3A) such as JunD, Creb1, and Atf7, for which sequence-specific DNA binding is most likely abolished by DNA methylation.

Readers for hmC showed partial overlap with proteins observed to interact with mC (Figure 1D, lower-right quadrant, and Figure 1E), as only three proteins interacted with both modified baits: MeCP2, Uhrf1, and Thy28. Uhrf1 and MeCP2 are known to bind both mC and hmC, although MeCP2 clearly binds with a higher affinity to mC compared to hmC (Frauer et al., 2011; Hashimoto et al., 2012; Mellén et al., 2012). Thy28 is an uncharacterized protein that is associated with apoptosis (Toyota et al., 2012) and contains an EVE domain, which is possibly involved in (ds)RNA binding (Bertonati et al., 2009). Interestingly, two DNA glycosylases (Mpg and Neil3) and a helicase (Recql) were identified as hmC readers in mESCs. These proteins might be involved in active DNA demethylation pathways to convert hmC back to cytosine via base excision repair mechanisms, as has been suggested previously (Hajkova et al., 2010; Wossidlo et al., 2010). In addition, a number of previously uncharacterized proteins, Wdr76 and C3orf37, preferentially bound to hmC compared to C. We purified WDR76 as a GFP fusion protein from HeLa cells and found interactions with OCR, HELLS, and GAN (Figure S1F). The mouse protein Hells, or Lsh, is a DNA helicase that has previously been implicated in regulating DNA methylation levels in cells (Dennis et al., 2001). Interestingly, OCR, or Spindlin-1, is a protein that is known to bind trimethylated H3 lysine 4 (H3K4me3) (Bartke et al., 2010). A large number of proteins preferentially bound to the nonmodified DNA, as was observed for the mC pull-down (Figure S1G). We validated some of these findings using western blotting for endogenous proteins (Figure S1H).

To further investigate the relative affinity of proteins for C versus mC versus hmC in a single experiment, we made use of a triple pull-down approach (Vermeulen et al., 2010), in which mESCs are grown in three different SILAC media. "Light," "medium," and "heavy" nuclear extracts derived from these cells are incubated with C-, mC-, and hmC-containing DNA, respectively (Table S1). Quantitative MS is used to visualize the relative abundance of a protein in each of the three different pull-downs. This experiment confirmed most of the observations made in Figures 1B and 1D, although for some proteins, the ratios in the triple pull-down are lower. As shown in Figures 1F and 1G, Klf4 and Zbtb44 preferentially bind to the methylated DNA. Other proteins bind to both modified baits, such as Uhrf1 (Figure 1H). Kdm2b preferentially binds to the nonmodified DNA (Figure 1K). Contrary to a previous report (Yildirim et al., 2011), we did not observe a specific interaction between MBD3 and hmC (forward ratio, 0.448; reverse ratio, 1.823). We validated these observations using recombinant protein (Figure 1C). At higher concentrations of recombinant MBD3 protein, we observed a specific interaction with mC (Figure 1C), which is in agreement with a recent study that revealed that MBD3 has the highest affinity for mC compared to hmC and C (Hashimoto et al., 2012).

Taken together, these experiments reveal that mC and hmC both recruit distinct proteins in mESCs with little overlap. Furthermore, a large number of proteins preferentially bind to nonmodified DNA. The number of observed interactions with

hmC is moderate, and some of these suggest that hmC acts as an intermediate in active DNA demethylation pathways in mESCs.

fC and caC Recruit a Large Number of Proteins in Mouse Embryonic Stem Cells, Including DNA Glycosylases and Transcription Regulators

We also applied our SILAC-based DNA pull-down approach to identify readers for fC and caC in mESCs. Colloidal blue analysis revealed that the total amount of protein binding to each bait is similar (Figure S2A). Ratios of the forward and reverse pull-downs with hmC, fC, or caC were individually averaged, and these average ratios were then plotted against each other in two-dimensional graphs (Figures 2A–2C and Table S1). From these plots, it is clear that both fC (blue, purple, and green) and caC (yellow and green) recruit many more proteins than hmC does (red and purple). Strikingly, there is only limited overlap between fC and caC binders (green) (Figure 2D). One of the proteins that binds to fC and caC, but not to hmC, is Tdg, which is consistent with its reported substrate specificity (Maiti and Drohat, 2011). We validated this binding behavior using recombinant protein in electromobility shift assays (EMSA) (Figures 2E and 2F). We also purified GFP-Tdg from ES cells to identify Tdg interaction partners (Figure S2B and Table S1). None of the Tdg interactors were identified as specific readers in the fC and caC pull-down, indicating that these fC and caC interactions are Tdg independent. Another fC-specific reader is the p53 protein, which plays an important role in DNA damage response (Kastan et al., 1991). Interestingly, Dnmt1 specifically interacted with caC. This interaction was confirmed by EMSA as well as western blotting using an antibody against endogenous protein (Figures 2F and S2C). We also identified subunits of the Swi/Snf chromatin-remodeling complex, such as BAF170, as readers for caC. Three proteins bind to all oxidized derivatives of mC: Thy28, C3orf37, and Neil1. GO term enrichment for biological processes shows that fC significantly enriches for proteins that are related to DNA repair (Benj.Hoch.FDR = $10^{-2.71}$) (Figure S3A), whereas caC interactors are not enriched for any biological process. RNA-binding proteins, mitochondrial proteins, and other proteins that are less likely to be associated with regulation of gene expression or DNA repair binding were identified as binders for fC and caC (Table S3). Some of these may have a basic affinity for the formyl and carboxyl groups on the DNA strands, which are more reactive than methyl or hydroxymethyl. To exclude the possibility that many fC and caC interactors are binding to damaged or abasic DNA, we validated the homogeneity of the DNA strands using HPLC (Figure S2D). Furthermore, we analyzed the DNA before (blue) and after incubation (red) with mESC nuclear extract by MALDI-TOF-MS (Figure S2E). Quantification of the modified residues by LC-MS/MS shows that there is no significant loss of the modified bases after incubation with nuclear extract (Figure S2F). Figures 2A–2C also show that the group of proteins that bind preferentially to nonmodified cytosine (black, lower-left quadrant) shows a large overlap between the three pull-downs and contains the PRC1.1, Mll1, and Ino80 complexes. To compare the relative affinity of proteins for these three modifications in a single experiment, we performed a triple pull-down. Analyses of the triple pull-down ratios for the identified fC and caC readers show similar trends, although some of the observed ratios are less prominent. As shown in Figures 2G–2L (and Table S1), the representative spectra of the indicated peptides of Tdg, Neil3, Mpg, Dnmt1, MeCP2, and Uhrf1 show relative ratios that are in agreement with ratios obtained in the independent experiments shown in Figures 2A–2C.

In summary, our data suggest that oxidized cytosine bases may induce a DNA damage response and trigger base excision repair pathways, which may finally result in DNA demethylation. In addition, each of these modifications recruits transcription regulators and other proteins that are not likely to be related to active DNA demethylation.

NPCs Contain a Distinct Set of mC and hmC Readers, Including Uhrf2, which Has the Highest Affinity for hmC

To investigate whether interactions with mC and hmC are dynamic during differentiation, we differentiated mESCs to NPCs. Nuclear extracts were generated from these cells followed by DNA pull-downs. Because no SILAC-compatible neurobasal medium is available, these experiments were performed using label-free quantification (LFQ) (Eberl et al., 2013; Hubner and Mann, 2011). Each DNA pull-down is analyzed separately and in triplicate. For all of the identified proteins (Table S1), we used ANOVA statistics ($p = 0.025$ and $S_0 = 2$) to compare the relative enrichment of proteins for each of the three baits. All significant outliers (192) were hierarchically clustered based on correlation after normalization by row mean subtraction (Figure 3A). Protein enrichment is indicated in red, whereas lack of enrichment is shown in blue. A large number of proteins bind to C or mC, whereas fewer proteins are specifically enriched in the pull-downs with hmC. Three smaller groups of proteins bind specifically to two of the baits (C/hmC, C/mC, or mC/hmC). As was observed in the DNA pull-downs from mESC nuclear extracts, CXXC-domain-containing proteins (Kdm2b and Mll, indicated in black) and their associated factors Bcor/Ring1a/b (blue) and Rbbp5/Ash2l (black) are enriched in the DNA pull-downs with nonmodified DNA relative to mC- and hmC-containing DNA. We identified Mbd2 and associated Mi-2/NuRD complex subunits as mC readers (indicated in yellow). Other identified MBD proteins include Mbd4, MeCP2, and Mbd1. Furthermore, a number of winged-helix (WH)-domain-containing proteins bound specifically to mC, including Rfx5 and its associated factors Rfxap and Rfxank (orange), which have previously been identified as methyl CpG interactors (Bartke et al., 2010).

Strikingly, these proteins bind more strongly to C compared to hmC. We further substantiated these observations by using recombinant protein (Figure 3B). This result indicates that, for some readers, oxidation of mC not only weakens the interaction, but also repels the mC interactor. The homeobox domain is significantly enriched in the cluster of mC-specific readers (Benj.Hoch.FDR = $10^{-1.8}$, Figure S3A), which is consistent with a previous study (Bartke et al., 2010). In addition, several known mC readers, such as Kaiso, Uhrf1, and Mbd4, bind both modified forms of cytosine. A number of DNA glycosylases bind specifically to hmC (Neil1, Neil3), as well as some helicases (Hells, Harp, Recql, and its homolog Bloom), which again suggests a

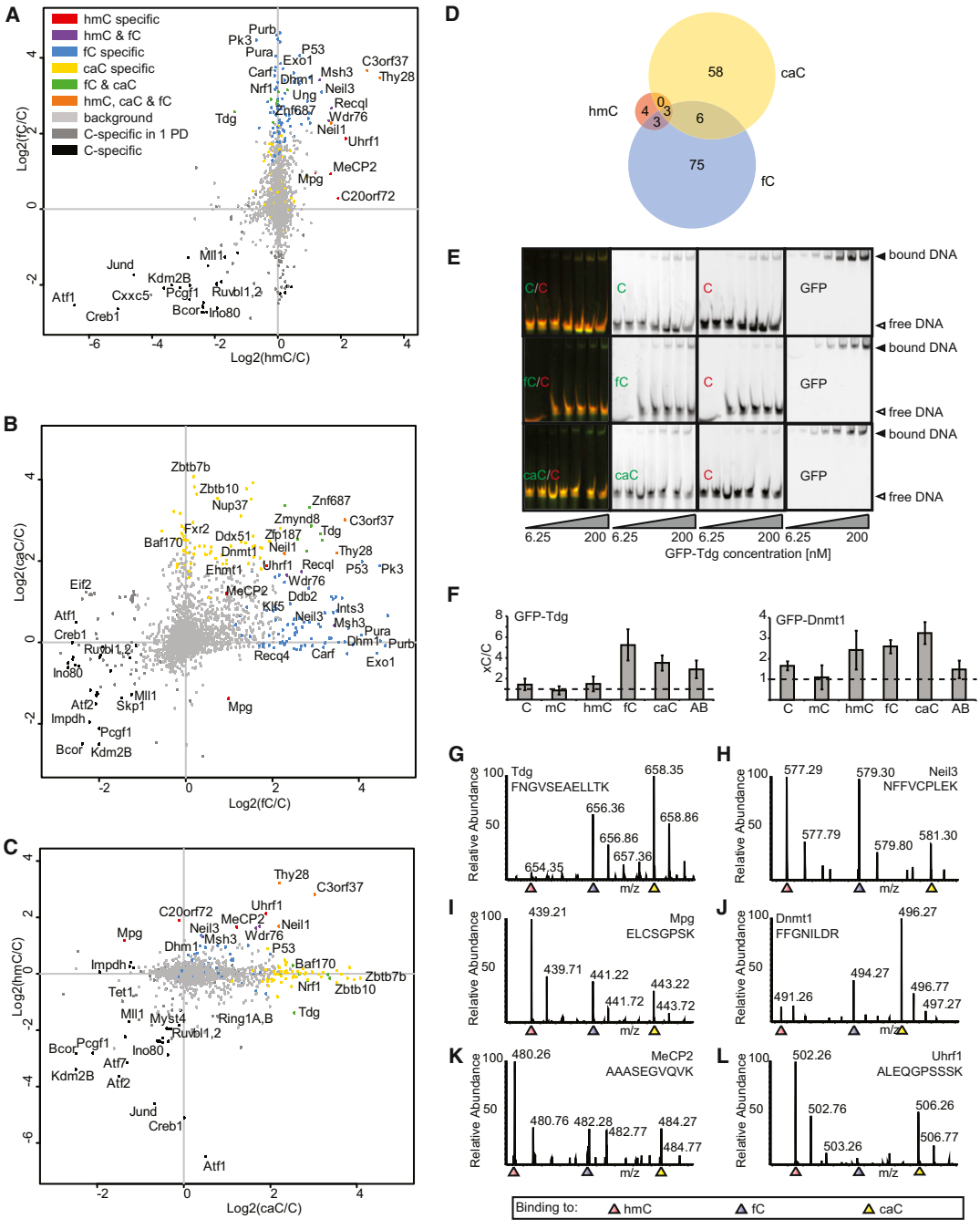


Figure 2. fC and caC Recruit a Large Number of Nonoverlapping Proteins in Mouse Embryonic Stem Cells

(A–C) Scatterplots of SILAC-based hmC, fC, and caC DNA pull-downs in mESC nuclear extract. The average ratio of all the identified and quantified proteins in the forward and reverse experiment for each of the three modifications is plotted on the X, Y, and Z axes of a three-dimensional cube. Shown in (A–C) are different side views of the cube. Colors indicate in which of the three pull-downs a protein was significantly enriched.

(D) Venn diagram showing the number of significantly enriched proteins for each of the baits.

(E) EMSA with GFP-Tdg at increasing protein concentrations (6.25–200 nM) incubated with dsDNA (250 nM of differentially labeled xC- and C-containing oligonucleotide, each).

(F) Electrophoretic mobility shift assays as shown in (E) performed with GFP-Tdg and GFP-Dnmt1 for all six residue variants (C, mC, hmC, fC, caC, and abasic site [AB]) in direct comparison to unmodified DNA. The binding preference was determined as the ratio of fluorescence signals of the different DNA substrates in the shifted bands. Shown are the means of three experiments; error bars represent SD.

(legend continued on next page)

DNA-repair-involved DNA demethylation pathway (GO DNA repair; Benj.Hoch.FDR = $10^{-3.91}$; Figure S3A). Although homeobox proteins are known to bind specifically to mC, a number of homeobox proteins show preferential binding to hmC in NPC extracts (examples include Zhx1 and -2). Finally, Uhrf2 was identified as a specific hmC-binding protein in NPCs, which we confirmed using recombinant protein (Figure 3B). Uhrf2 is not expressed in mESCs, and its levels increase upon differentiation (Pichler et al., 2011). This explains why Uhrf2 was not identified as an hmC-specific reader in mESC DNA pull-downs.

Taken together, these experiments reveal that interactions with mC and hmC are highly dynamic during differentiation. Furthermore, the observations made in NPCs strengthen our hypothesis that oxidation of mC serves as a trigger for active DNA demethylation. Nevertheless, some hmC-specific readers in NPCs do not appear to be linked to DNA repair mechanisms, indicating that, in these cells, hmC may also serve a role as a "classical" epigenetic mark that recruits transcriptional regulators.

NMR-Based Analysis of the Rfx5 WH Domain Bound to mC DNA

The specific interaction between the Rfx5 WH domain and mC DNA was studied in detail using solution nuclear magnetic resonance (NMR) spectroscopy in order to derive binding affinity and identify the mC-binding site. Addition of a singly methylated 18bp DNA fragment to the Rfx5-WH domain results in large changes in the ^{1}H-^{15}N HSQC "fingerprint" spectrum (Figure 3C). After addition of a slight molar excess of DNA, the spectrum does not show any further changes, indicating that Rfx5-WH strongly binds mC DNA and preferentially at only one of the two mC sites (Figure 3C). The affinity of Rfx5 for mC DNA was derived from the observed peak displacement for residues in the fast exchange regime, such as T104 and E102, assuming that the two mC are independent and equivalent, which resulted in an apparent dissociation constant $K_{D,app}$ of ~3 μM (with 95% probability limits 10 nM<K_D<16 μM) (Figure 3D and Supplemental Information). Based on DNA pull-downs done with recombinant protein, which revealed a quantitative depletion of the WH domain from the lysate, we anticipate the K_D to be in the nM range (Figure 3B) To identify the residues that are responsible for specific mC binding, we used the DNA-bound Rfx1 WH domain crystal structure (PDB ID 1DP7; sequence identity 35%; Avvakumov et al., 2008; Gajiwala et al., 2000) to construct a homology model structure of Rfx5-WH and validated it against the experimental chemical shifts (data not shown). The homology model contains a hydrophobic pocket that includes residues with the largest chemical shift changes and is well aligned with an extended basic surface that is responsible for DNA binding in Rfx1. This binding pocket, formed by the side chains of K110, V113, Y114, T132, F135, L139, and Y169, is appropriately shaped to capture the mC base via a flip-out mechanism, as seen in the case of UHRF1 (Figure 3E). Steric clashes introduced by the presence of an additional hydroxyl group could cause the observed specificity for mC. Given the apparent high affinity and DNA-sequence-independent binding to mC, we propose that the WH domain that is present in Rfx proteins is a bona fide mCpG-binding domain.

Brain-Specific Readers for mC and hmC Include Dlx Proteins

The adult brain is the organ with the highest levels of hmC (Globisch et al., 2010). Tet enzymes and hmC have been shown to play a role in active DNA demethylation of certain genes in this organ (Guo et al., 2011). To identify readers for C, mC, and hmC in the adult brain, nuclear extracts were prepared from this tissue, and these extracts were used for DNA pull-downs. LFQ was used to determine differential binders (Table S1). In brain extracts, we identified fewer specific readers compared to NPCs (108, p = 0.025 and S_0 = 0; Figure 4), most likely due to the presence of highly abundant structural proteins derived from connective tissue and extracellular matrix in these nuclear extracts. Interestingly, more proteins specifically bind to hmC compared to mC in brain extracts. This is in contrast to NPCs and mESCs, in which more interactions with mC relative to hmC are observed, which may imply a specific role for hmC in brain tissue.

The nonmodified DNA pull-down enriched for the same factors as those observed in mESCs and NPCs, including Cxxc5, Kdm2b, and Bcor (CXXC-domains indicated in black, PRC1 complex in blue, and Ino80 in red). In this case, mC DNA was bound by the Mbd2/NuRD complex, which contains the brain-specific ATPase Chd5 (Eberl et al., 2013; Potts et al., 2011) (indicated in yellow). Interestingly, we identified three distal-less homeobox proteins (Dlx1, -5, and -6) as specific mC interactors. Dlx proteins play a role in the development of the brain and are also expressed in specific regions of the adult brain (Jones et al., 2011; Wang et al., 2011). Wdr76 and Thy28 are hmC specific, as was also observed in NPCs. Thap11 (or Ronin) is identified as a brain-specific hmC reader. Interestingly, this protein is highly expressed in certain regions of the brain, including Purkinje cells (Dejosez et al., 2008). Finally, we identified all four subunits of replication factor C (Rfc2-5) and the associated factor Rfc1 as hmC-specific readers (indicated in green).

Altogether, these experiments further emphasize the dynamic nature of the mC and hmC interactomes during development.

Global Absolute Quantification of Protein Levels in mESCs, NPCs, and Adult Mouse Brain Extracts Reveals Expression-Level-Dependent and -Independent Interaction Dynamics

Our screening for mC- and hmC-specific readers in mESCs, NPCs, and adult mouse brain revealed a large number of cell-type- or organ-specific interactors (Figure S3B). The most

(G–L) Representative spectra of the indicated peptides obtained in the triple-labeled DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for hmC-containing (red), fC-containing (blue), and caC-containing (yellow) DNA. Spectra are shown for Tdg (G), Neil3 (H), Mpg (I), Dnmt1 (J), MeCP2 (K), and Uhrf1 (L).
See also Figure S2 and Table S1.

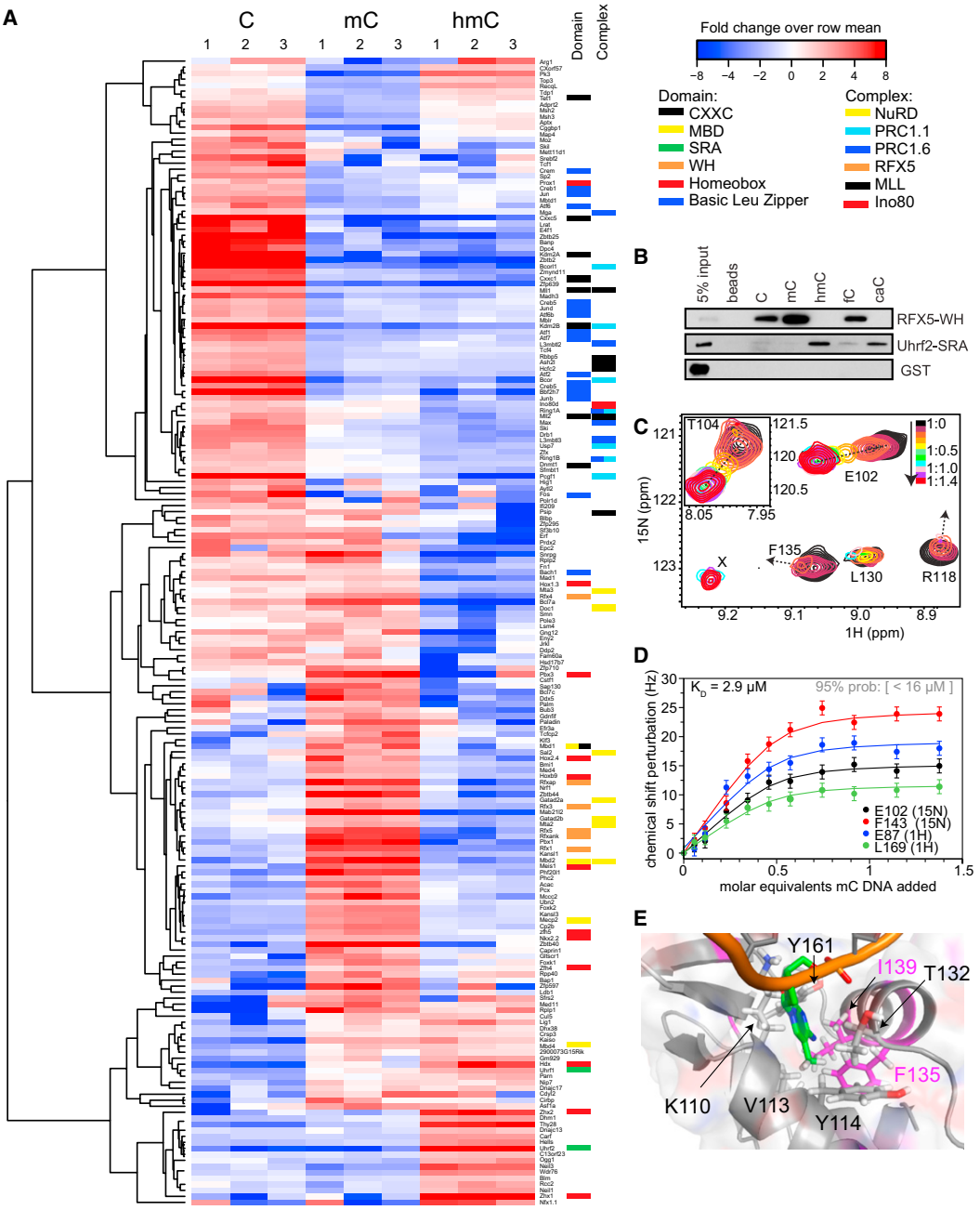


Figure 3. Hierarchical Clustering of NPC-Specific C, mC, and hmC Readers

(A) Correlation-based clustering of the LFQ intensities after log2 transformation and normalization by row mean subtraction. Included in the clustering are proteins that are significantly binding to at least one of the baits as determined by an ANOVA test. Blue indicates lack of enrichment, whereas enrichment is indicated in red. Domain and Complex columns indicate the DNA-binding domain(s) that may be responsible for direct binding to the bait and the complexes that readers are part of, respectively.

(B) Biochemical validation experiments using DNA pull-downs with recombinant DNA-binding domains.

(C) Overlay of Rfx5-WH HSQC spectra with increasing amounts of mC DNA added and color-coded on the indicated scale listing the WH domain:DNA ratio. Some residues, such as F135 and R118, cannot be unambiguously tracked to their bound states because their chemical shift changes are very large. Peaks corresponding to their bound state, such as "X," appear only after addition of a full molar equivalent of DNA.

(D) Selected binding curves and fits for resonances that are in the fast exchange regime throughout the titration. Error bars (SD) for the peak positions are set to 1.2 Hz.

(E) Close-up of the putative mC-binding pocket in the RFX5 WH domain. The methylated cytosine is indicated in green.

See also Table S1.

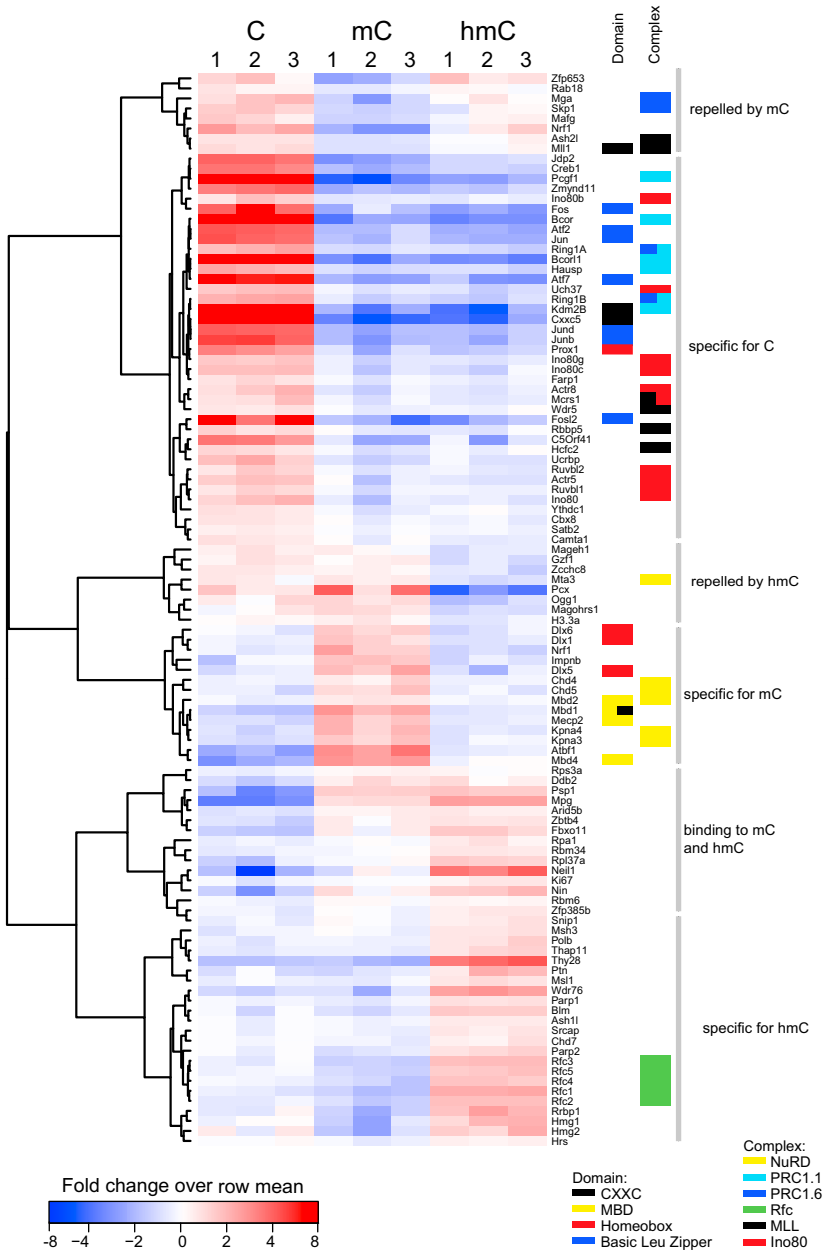


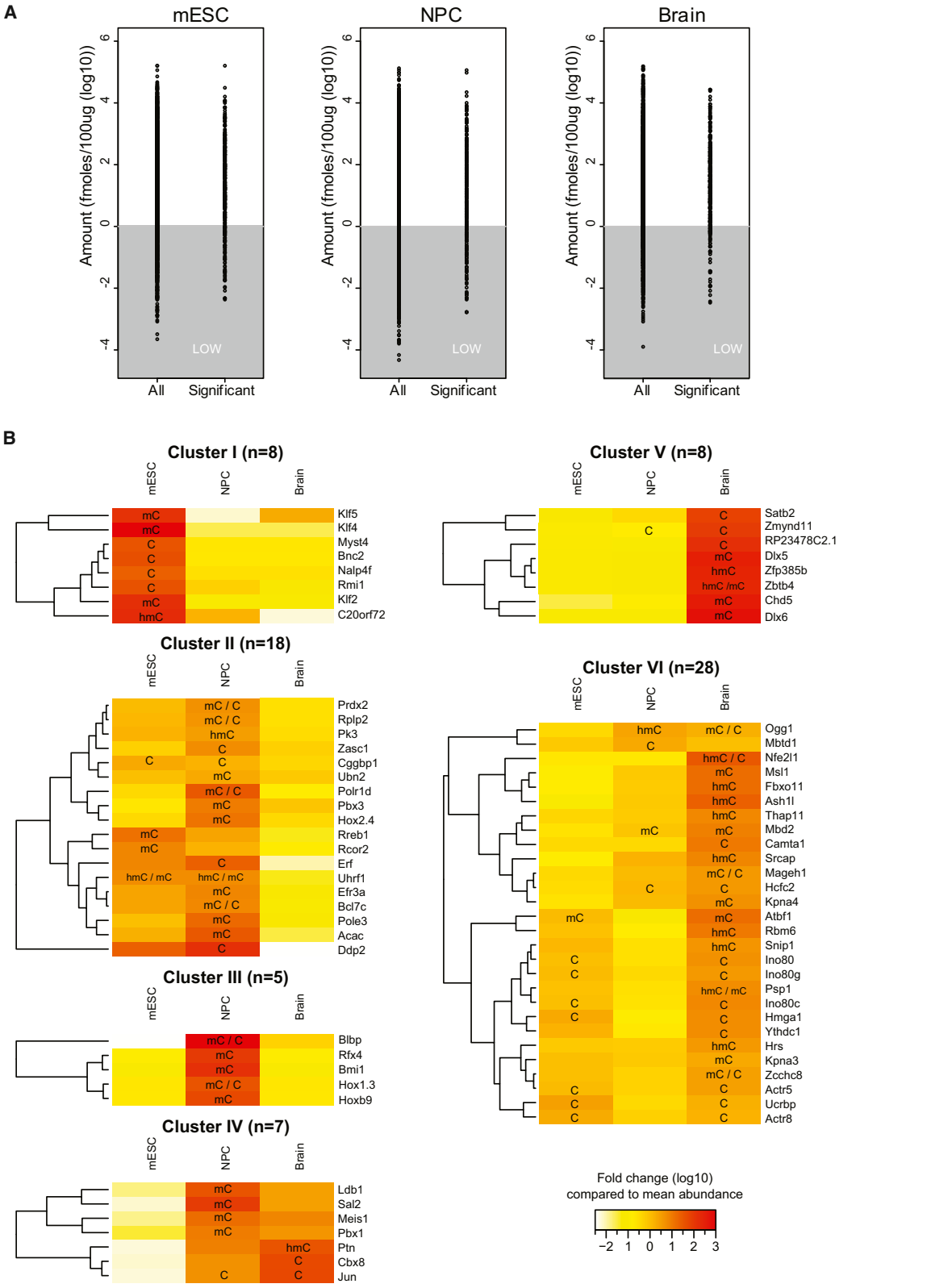
Figure 4. Hierarchical Clustering of Brain-Specific C, mC, and hmC Readers

Correlation-based clustering of the row-mean-subtracted LFQ intensities of proteins in C, mC, and hmC DNA pull-downs in adult mouse brain nuclear extracts.

See also Table S1.

obvious explanation for these observed differential interactions is regulation of reader abundance at the protein level. Alternatively, the interaction between a reader and (modified) DNA may be affected by posttranslational modifications (PTMs). To investigate global absolute protein levels in the different nuclear extracts that were used for the pull-downs, we made use of a method called intensity-based absolute quantification (iBAQ) (Schwanhäusser et al., 2011). Approximately 8,000 proteins were quantified in at least one of the extracts (Table S2). All proteins with at least a 10-fold change in concentration were clustered based on their expression pattern (Figure S4B).

The cluster of mESC-specific proteins is enriched for anchoring junction (Benj.Hoch.FDR = $10^{-2.96}$) and cell adhesion (Benj.Hoch.FDR = $10^{-2.14}$), whereas proteins in brain enriched GO terms such as synaptic transmission (Benj.Hoch.FDR = $10^{-3.77}$) and cognition (Benj.Hoch.FDR = $10^{-2.75}$), as expected (Figure S4C). The molar concentrations of proteins that are significantly enriched in one of the DNA pull-downs are spread over several orders of magnitude, indicating that our screening is not biased toward high-abundant proteins (Figure 5A). Of the 259 proteins that showed dynamic interactions through development (Table S3), 20 proteins were not quantified in the iBAQ measurements. The 74 proteins (~31%) that do show a correlation between interaction pattern and protein abundance in the different extracts can be divided into six clusters (Figure 5B). A correlation was defined as gaining or losing an interaction accompanied by at least a 2-fold change in protein abundance. An example of a protein that was identified as a specific (mC) reader only in mESCs was Klf4. As shown in Figure 5B, this protein is highly expressed in mESCs but is less abundant in NPCs or in the adult mouse brain. Another example is represented by the Dlx5 and Dlx6 proteins, which are highly abundant in brain nuclear extract and exclusively bind to mC in pull-downs from these extracts. For about 185 proteins, no correlation is observed between expression levels (at least 2-fold change) and binding behavior. For these proteins, the cause of differential binding may be explained through PTMs that affect the interaction between a reader and DNA or a differentially expressed cofactor. A good example of the latter is the Mi-2/NuRD complex. Although most of its subunits display equal expression levels in mESCs, NPCs, and brain, mC-specific interactions are not observed in mESCs. This can be explained by the fact that Mbd2, which is the direct reader of mC within the NuRD complex, is low in abundance in mESCs and is upregulated during differentiation (Figure 5B). Thereby, it controls the mC-specific binding of the entire complex. In mESCs, the majority of the Mi-2/NuRD complex contains Mbd3, which is



(legend on next page)

the MBD-containing protein that has lost its high-affinity mC binding ability. Furthermore, technical reasons for not identifying an interactor could be the presence of highly abundant structural proteins in the brain lysate or binding competition among different readers in the extracts. Altogether, the absolute quantification of protein abundance in the different nuclear extracts revealed large differences in protein levels between mESC, NPCs, and adult mouse brain. This data set serves as a rich resource on its own but also enables us to explain many of the differential interactions that we identified using quantitative MS-based interactomics.

Uhrf2 Stimulates the Sequential Activity of the Tet1 Enzyme

The first protein that was identified as an hmC binder was Uhrf1 (Frauer et al., 2011), a protein that is involved in maintenance of DNA methylation (Bostick et al., 2007). Our data revealed that Uhrf1 binds with a similar affinity to mC and hmC, which is consistent with previously published data (Frauer et al., 2011). This is in contrast to Uhrf2, which we identified as a high-affinity hmC-binding protein in NPC cells that shows a lower affinity for mC. The function of Uhrf2 is not well understood. It is clear, however, that Uhrf2 cannot rescue the phenotype of Uhrf1 knockout cells, which lose DNA methylation (Pichler et al., 2011; Zhang et al., 2011). Uhrf1 is highly expressed in mESCs, whereas Uhrf2 levels increase during differentiation (Table S3 and Pichler et al., 2011). Altogether, this prompted us to investigate whether Uhrf2 expression affects the levels of mC and its oxidized derivatives. The Tet1-catalytic domain was transfected into HEK293T cells with and without coexpression of Uhrf2. Total genomic DNA modification levels were determined using LC-MS/MS (Figure 6 and Supplementary Information). As shown in Figure 6D, Uhrf2 overexpression increases the level of hmC. More striking is the increase of fC and caC levels upon Uhrf2 coexpression together with the Tet1 catalytic domain. Because fC and caC serve as substrates for Tdg and BER, the detected increase in the levels of fC and caC following Uhrf2 expression may be an underestimation of the actual production of these bases. It therefore seems that Uhrf2 promotes repetitive oxidation of mC by the Tet proteins. We hypothesize that flipping the modified cytosine base out of the DNA double helix, as has been described for Uhrf1 binding to methylated and hydroxymethylated DNA (Arita et al., 2008; Frauer et al., 2011), may enhance accessibility of the hydroxymethylated base to the Tet enzymes, thereby promoting further oxidation.

DISCUSSION

In this study, we have used quantitative MS-based proteomics to identify readers for mC and its oxidized derivatives in mESCs, as well as readers for mC and hmC in NPCs and adult mouse brain. Readers for individual modifications were found to be highly dynamic throughout the three cell types and tissues that we investigated (Figure 7). This is in contrast to interactions with histone modifications, such as trimethylated lysines on histone H3. For these modifications, the majority of interactors are constant between different cell types or developmental stages (Eberl et al., 2013 and M.V., unpublished data). Readers for distinct cytosine modifications show limited overlap. This indicates that, at least from a biochemical perspective, mC, hmC, fC, and caC behave quite differently. Although little overlap was observed with regard to proteins that interact with each of the epigenetic marks, they all repelled a common set of proteins, such as several CXXC-domain-containing proteins and their interactors. It remains to be determined which of the consequences of DNA (hydroxy)methylation is functionally most relevant: recruitment of transcriptionally repressive complexes or preventing the binding of certain (activating) proteins to unmodified DNA. A detailed biochemical characterization of the interactions and their dissociation constants will be important to answer this question.

Our experiments revealed a number of DNA glycosylases and DNA repair proteins that bind to hmC, fC, and caC, whereas we identified few such proteins binding to mC. The enriched binding of DNA-repair-associated proteins was most pronounced for fC. From this observation, one can conclude that the conversion of hmC to fC is a signal that is likely to result in repair-associated removal of the modified base by proteins that are rather ubiquitously expressed. It is therefore surprising that, in different cell types and tissues, rather constant levels of hmC, fC, and caC are found. The maintenance of such constant levels of these bases in mESCs may indicate a high turnover of DNA methylation, probably involving a constant ''correction'' by de novo methylation. Regardless, it will be important to investigate which mechanisms control Tet enzyme conversion of mC to hmC and further oxidation to fC and caC. Our data reveal that coexpression of Uhrf2 with the catalytic domain of Tet1 results in a (transient) upregulation of hmC, fC, and caC, indicating that Uhrf2 promotes the sequential oxidation of mC by Tet1. One of the other factors influencing the catalytic activity of the Tet enzymes is the concentration of cellular metabolites. It has been shown that oncometabolites such as 2-hydroxyglutarate can competitively inhibit the activity of 2-oxo-glutarate-dependent enzymes, such as the Tet proteins (Chowdhury et al., 2011; Xu et al., 2011). Furthermore, mutations in IDH1 and -2, which generate 2-oxo-glutarate, are phenocopied by mutations in the TET enzymes and result in cancer (Figueroa et al., 2010). Mutations in the IDH2 and TET2 genes were also linked to lower genomic hmC levels and altered gene expression patterns in myeloid cancers (Ko et al.,

Figure 5. Global Absolute Protein Quantification in mESCs, NPCs, and Adult Mouse Brain

(A) Graphs indicating the concentration of all proteins identified in the nuclear extract (all) and the identified readers (significant) in each of the cell types. The gray area indicates the concentration at which protein quantification is inaccurate.

(B) Readers for which protein expression levels correlate with DNA binding patterns were clustered into six groups based on their expression in the three different nuclear extracts. The color indicates protein levels (white, low; red, high), whereas binding preference is indicated by C, mC, hmC, or combinations thereof.

See also Figure S4 and Table S2 and Table S3.

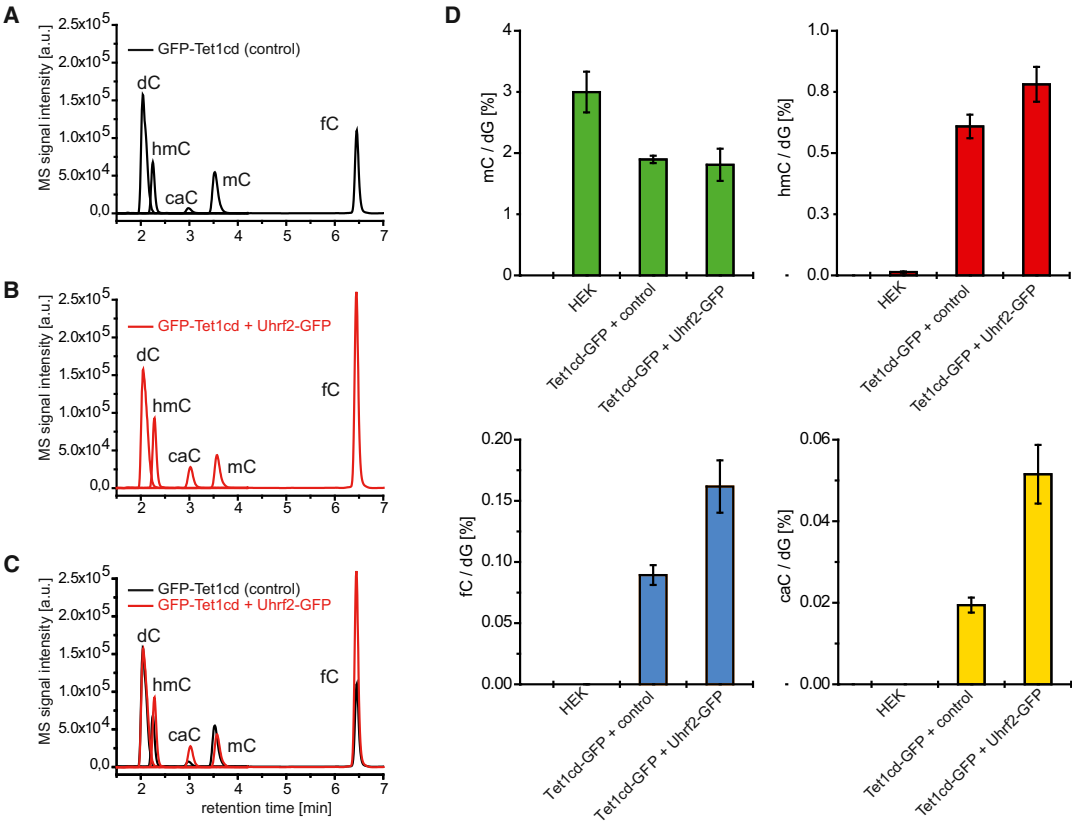


Figure 6. Stable Isotope-Dilution-Based LC-ESI-MS/MS Quantification of Cytosine Derivatives in HEK293T Cells

(A) Nonquantitative LC-MS/MS chromatogram of digested genomic DNA from HEK293T cells cotransfected with Tet1-catalytic domain-GFP (GFP-Tet1cd) and an unrelated expression construct (control). Depicted are the overlaid ion chromatograms of the MS/MS transitions for dC and the cytosine derivatives (black curves). dC, mC, and hmC were measured by a factor of $\sim 10^2$–10^3 less sensitive in comparison to caC and fC.

(B) Same as (A) except that Uhrf2-GFP was coexpressed together with GFP-Tet1cd. The MS signal intensities were normalized to the dC content of (A).

(C) Superposition of (A) and (B).

(D) Levels of cytosine derivatives relative to the total cytosine content (dG) as determined by quantitative LC-MS/MS mass spectrometry.

Shown are the means of technical triplicates; error bars reflect SD.

2010; Konstandin et al., 2011). In support of these observations, which clearly link hmC to cancer, we noticed that many hmC, fC, and caC readers are implicated in cancer, including UHRF2, CARF, p53, and HELLS (Lee et al., 2000). Interestingly, mutations in the Hells helicase, which we identified as an hmC reader in NPCs, result in a decrease of DNA methylation levels in cells (Myant et al., 2011). It seems clear that regulating the levels of mC and its oxidized derivatives is essential for normal cell homeostasis and that deregulation of the readers, writers, and erasers of these marks results in a disturbance of the balance between cell proliferation and differentiation during development.

EXPERIMENTAL PROCEDURES

Cell Culture

IB10 mESCs were cultured in light (R^0K^0) or heavy ($R^{10}K^8$) SILAC medium in the presence of 2i compounds. For triple labeling, a third type of medium was used containing medium-labeled L-lysine (K^4) and L-arginine (R^6). mESCs were differentiated to NPCs in N2B27 medium and cultured in NSA medium, consisting of NSA MEM, 1% glutamine, 1× N2 supplement, 10 ng/ml bFGF, and 10 ng/ml EGF.

DNA Pull-Downs

Nuclear extracts were generated as described previously (Eberl et al., 2013; Vermeulen et al., 2010). DNA (see Table S4) immobilized on Dynabeads My-One C1 was incubated with nuclear extract in 50 mM Tris-HCl (pH 8), 150 mM NaCl, 1 mM DTT, 0.25% NP40, and complete protease inhibitors (Roche, EDTA-free) in the presence of poly-dAdT. After extensive washes (using incubation buffer without poly-dAdT), bound proteins were in-gel digested using trypsin. After sample preparation, peptides were desalted on Stage-tips (Rappsilber et al., 2003).

Mass Spectrometry

Peptides were separated using an EASY-nLC (Proxeon) connected online to an LTQ-Orbitrap Velos mass spectrometer (Thermo) as described (Smits et al., 2013). Raw data were analyzed using MaxQuant version 1.2.2.5 and searched against protein database ipi.MOUSE.v3.68.fasta. Using Perseus, data were filtered and scatter plots were made using R. The raw mass spectrometry data have been deposited to the ProteomeXchange Consortium

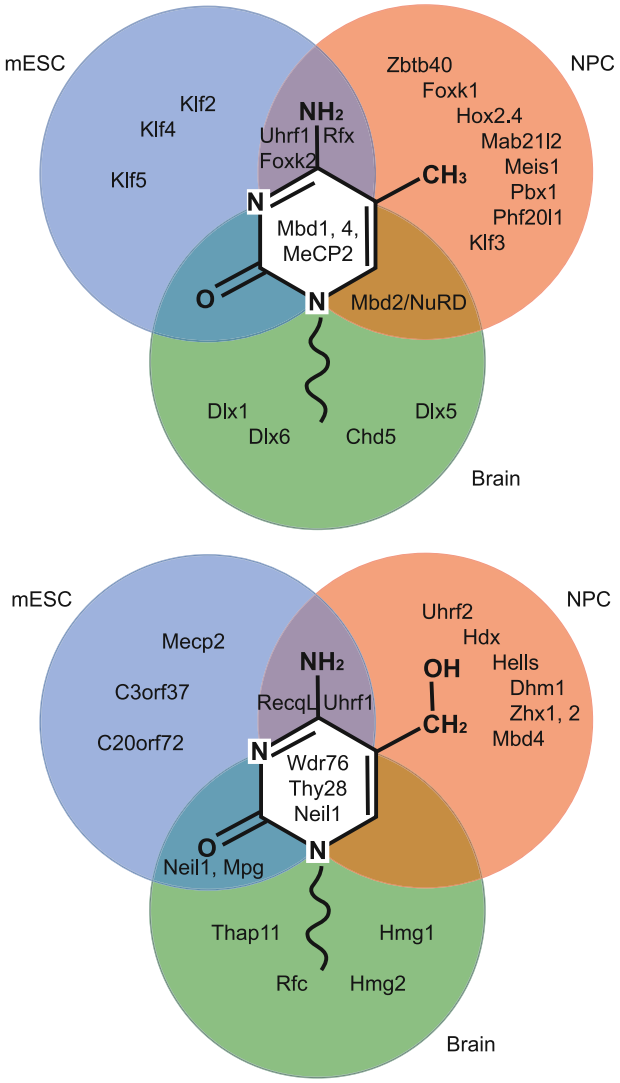


Figure 7. Visualization of Cell-Type-Specific and General mC and hmC Readers

Venn diagram showing examples of mC (A) and hmC (B) readers that were identified in mESCs (blue), NPCs (orange), and adult mouse brain (green). See also Figure S3 and Table S3.

(http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaíno et al., 2013) with the dataset identifier PXD000143.

Recombinant Protein Expression and DNA Pull-Downs

DNA-binding domains were cloned into the GST-containing PRP256NB vector. The Uhrf2(aa416–626) GST-fusion construct was kindly provided by Dr. Jiemin Wong. Protein expression was performed in *E. coli* BL21 codon$^+$ cells. Bacterial lysate was cleared by ultracentrifugation. DNA pull-downs were performed as described above with the addition of 10 μM $ZnCl_2$ to the incubation buffer.

iBAQ

iBAQ was performed essentially as described in Schwanhäusser et al. (2011). A more detailed description can be found in the Extended Experimental Procedures.

LC-MS/MS Analysis of Genomic DNA

Cotransfections were performed in HEK293T cells, and genomic DNA was purified according to Münzel et al. (2010). Quantification of DNA nucleosides from genomic DNA is based on a further development of our isotope dilution method (Pfaffeneder et al., 2011 and data not shown). LC-MS/MS analysis was performed on an Agilent 6490 triple quadrupole mass spectrometer coupled to an Agilent 1290 UHPLC system. For general source- and compound-dependent parameters, see the Extended Experimental Procedures and Tables S5 and S6. The transitions of the nucleosides were analyzed in the positive-ion-selected reaction monitoring mode (SRM) operating MS1 and MS2 under unit mass resolution conditions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.02.004.

ACKNOWLEDGMENTS

We would like to thank Dr. Jiemin Wong for providing constructs. We are very thankful to Dr. Nikolay Outchkourov for providing NPCs, and we thank Tony Hyman and Ina Poser for providing the BAC-GFP WDR76 cell line. We would also like to thank Deepani Poramba Liyanage for support. Furthermore, we would like to thank all group members that contributed to fruitful discussions and Susan Kloet for critical reading of the manuscript. We thank the PRIDE team for their help uploading the raw mass spec data. Work in the Vermeulen lab is supported by a grant from the Netherlands Organization for Scientific Research (NWO-VIDI) and an ERC Starting Grant. The Carell group thanks the DFG Normalverfahren CA275/8-4, SFB 749, SFB 646, and the Volkswagen Foundation for financial support. Work in the Leonhardt group is supported by the SFB 646 and SPP 1463 grants from The Deutsche Forschungsgemeinschaft (DFG). M. Münzel and T.P. thank the Fonds der Chemischen Industrie for predoctoral fellowships. F.K. is supported through the International Research Support Initiative Program from the Higher Education Commission, Pakistan. H.v.I. is supported by a Veni fellowship from the Dutch Science Foundation Chemical Sciences, NWO-CW. K.L. and J.W. thank the DFG (WA1029/6) for support through SPP1356.

Received: October 9, 2012
Revised: January 11, 2013
Accepted: February 5, 2013
Published: February 21, 2013

REFERENCES

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

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y.J., Duan, S.L., 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.

Bartels, S.J.J., Spruijt, C.G., Brinkman, A.B., Jansen, P.W.T.C., Vermeulen, M., and Stunnenberg, H.G. (2011). A SILAC-based screen for Methyl-CpG binding proteins identifies RBP-J as a DNA methylation and sequence-specific binding protein. PLoS ONE *6*, e25884.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470–484.

Bertonati, C., Punta, M., Fischer, M., Yachdav, G., Forouhar, F., Zhou, W.H., Kuzin, A.P., Seetharaman, J., Abashidze, M., Ramelot, T.A., et al. (2009). Structural genomics reveals EVE as a new ASCH/PUA-related domain. Proteins *75*, 760–773.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Mol. Cell *38*, 179–190.

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.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W.W., Jiang, J.M., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

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.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Defossez, P.A., and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. Prog. Mol. Biol. Transl. Sci. *101*, 377–398.

Dejosez, M., Krumenacker, J.S., Zitur, L.J., Passeri, M., Chu, L.F., Zhou, S.Y., Thomson, J.A., and Zwaka, T.P. (2008). Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. Cell *133*, 1162–1174.

Dennis, K., Fan, T., Geiman, T., Yan, Q.S., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev. *15*, 2940–2944.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Mol. Cell *49*, 368–378, Published onlne November 29, 2012. http://dx.doi.org/10.1016/j.molcel.2012.10.026.

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.

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

Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature *403*, 916–921.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE *5*, e15367.

Guo, J.U., Su, Y.J., Zhong, C., Ming, G.L., and Song, H.J. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423–434.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science *329*, 78–82.

Hashimoto, H., Liu, Y.W., Upadhyay, A.K., Chang, Y.Q., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X.D. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q.Y., Ding, J.P., Jia, Y.Y., Chen, Z.C., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Hubner, N.C., and Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). Methods *53*, 453–459.

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.

Jones, D.L., Howard, M.A., Stanco, A., Rubenstein, J.L.R., and Baraban, S.C. (2011). Deletion of Dlx1 results in reduced glutamatergic input to hippocampal interneurons. J. Neurophysiol. *105*, 1984–1991.

Kastan, M.B., Onyekwere, O., Sidransky, D., Vogelstein, B., and Craig, R.W. (1991). Participation of p53 protein in the cellular response to DNA damage. Cancer Res. *51*, 6304–6311.

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.

Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S., et al. (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. Leukemia *25*, 1649–1652.

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

Lee, D.W., Zhang, K.J., Ning, Z.Q., Raabe, E.H., Tintner, S., Wieland, R., Wilkins, B.J., Kim, J.M., Blough, R.I., and Arceci, R.J. (2000). Proliferation-associated SNF2-like gene (PASG): a SNF2 family member altered in leukemia. Cancer Res. *60*, 3612–3622.

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.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. Cell *151*, 1417–1430.

Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmiller, V., Michalakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew. Chem. Int. Ed. Engl. *49*, 5375–5377.

Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Res. *21*, 83–94.

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

Pichler, G., Wolf, P., Schmidt, C.S., Meilinger, D., Schneider, K., Frauer, C., Fellinger, K., Rottach, A., and Leonhardt, H. (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. J. Cell. Biochem. *112*, 2585–2593.

Potts, R.C., Zhang, P.S., Wurster, A.L., Precht, P., Mughal, M.R., Wood, W.H., 3rd, Zhang, Y.Q., Becker, K.G., Mattson, M.P., and Pazin, M.J. (2011). CHD5, a brain-specific paralog of Mi2 chromatin remodeling enzymes, regulates expression of neuronal genes. PLoS ONE *6*, e24515.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*, 663–670.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Sengupta, P.K., Ehrlich, M., and Smith, B.D. (1999). A methylation-responsive MDBP/RFX site is in the first exon of the collagen alpha2(I) promoter. J. Biol. Chem. *274*, 36649–36655.

Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res. *41*, e28.

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.

Tahiliani, M., Koh, K.P., Shen, Y.H., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

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

Toyota, H., Jiang, X.Z., Asakura, H., and Mizuguchi, J. (2012). Thy28 partially prevents apoptosis induction following engagement of membrane immunoglobulin in WEHI-231 B lymphoma cells. Cell. Mol. Biol. Lett. *17*, 36–48.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980.

Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al. (2013). The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. *41*(Database issue), D1063–D1069.

Wang, B., Lufkin, T., and Rubenstein, J.L.R. (2011). Dlx6 regulates molecular properties of the striatum and central nucleus of the amygdala. J. Comp. Neurol. *519*, 2320–2334.

Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., Schöler, H., and Walter, J. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. EMBO J. *29*, 1877–1888.

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.

Yildirim, O., Li, R.W., Hung, J.H., Chen, P.B., Dong, X.J., Ee, L.S., Weng, Z.P., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell *147*, 1498–1510.

Zhang, J.Q., Gao, Q.Q., Li, P.S., Liu, X.L., Jia, Y.H., Wu, W.C., Li, J.W., Dong, S., Koseki, H., and Wong, J.M. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. Cell Res. *21*, 1723–1739.

Cell

EXTENDED EXPERIMENTAL PROCEDURES

SILAC Labeling of ES Cells

IB10 murine Embryonic stem cells were cultured feeder-free on gelatin coated dishes in medium consisting of 500 ml SILAC Dulbecco's Modified Eagle Medium without arginine, lysine and glutamine (PAA, E15-086), supplemented with 15% MESC serum substitute (Thermo Scientific), Glutamine, Penicillin/Streptomycin, 1x Non-essential amino acids, sodium pyruvate, 73 μg/ml L-Lysine (light/K^0 (Sigma, A6969), medium/K^4 (Sigma, 616192 or Silantes, 211103912) or heavy/K^8 (Sigma, 608041 or Silantes, 211603902)) and 29.4 μg/ml arginine (light/R^0 (Sigma, A6969), medium/R^6 (Sigma, 643440 or Silantes, 201203902) or heavy/R^{10} (Sigma, 608033 or Silantes, 201603902)), LIF (1000 U/ml), β-mercaptoethanol and 2i compounds (CHIR99021 and PD0325901, 3 and 1 μM respectively). Cells were cultured in SILAC medium until labeling efficiency exceeded 95% after which cells were expanded and harvested to generate nuclear extracts.

NPC Culturing

Neuronal progenitor cells were kindly provided by Dr. N. S. Outchkourov. They were cultured in medium consisting of NSA MEM (Euromed EVM0883LD), 1% glutamine, 1x N2 supplement, 10 ng/mL bFGF (RD systems 233-F3) and 10 ng/mL EGF (235-E9) on gelatin-coated dishes. Cells were detached from culture plates using accutase. Nuclear extracts were made as described below.

Mice Brain Nuclear Extracts

Nuclei from adult mouse brain were purified by centrifugation through a sucrose cushion following homogenization, modified from (Lavery and Schibler, 1993). Then nuclei were lysed as described below.

Nuclear Extract Preparation

This protocol is based essentially on Dignam et al. (Dignam et al., 1983). Briefly, cells were trypsinized and washed two times with PBS. Using a hypotonic buffer, the cells were swollen, after which the cells were lysed by dounce homogenizing in the presence of 0.15% NP40 and complete protease inhibitors. After centrifugation, the pellet consisting of nuclei was lysed by 90 min incubation in 2 volumes of nuclear lysis buffer (420 mM NaCl, 20 mM HEPES pH 7.9, 20% v/v glycerol, 2 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP40, complete protease inhibitor w/o EDTA (Roche) and 0.5 mM DTT). After centrifugation, the supernatant containing the soluble nuclear extract was aliquoted and snap frozen until further usage. Protein concentrations of the nuclear extracts were determined using the Biorad Protein assay.

DNA Synthesis

The synthesis of the oligonucleotides was performed on an ABI 394 DNA/RNA Synthesizer (Applied Biosystems) using typical reagent concentrations (activator: 0.25 M benzylthiotetrazole in MeCN (10 ppm H_2O), detritylation: 3% dichloroacetic acid in CH_2Cl_2, oxidation: 25 mM I_2 in MeCN/H_2O/2,6-lutidine (11/5/1), capping: Ac_2O/2,6-lutidine/MeCN (30 ppm H_2O) (20/30/50) and 20% *N*-methylmidazole in MeCN (10 ppm H_2O). The oligonucleotide syntheses were performed on 200 nmol low-volume polystyrene carriers using 0.1 M DNA CE-phosphoramidites: A (Bz-dA), C (Bz-dC), G (*i*Bu-dG), T, mC (Bz-mC) obtained from *Glen Research* or *Link Technologies*. hmC, fC and caC phosphoramidites were synthesized according to literature (Münzel et al., 2010b) and incorporated into DNA using the standard protocol. Benzylthiotetrazole was prepared according to literature (Welz and Muller, 2002). The coupling times for the modified bases were increased to 3 min to ensure maximum coupling efficiency.

The mC and the unmodified strands were treated with ethanolic ammonia for cleavage of the carrier and removal of the permanent protecting groups. hmC, fC and caC containing DNA was cleaved and deprotected using 0.4 M NaOH in MeOH/H_2O 4:1 for 18 hr at room temperature. After addition of 600 μl triethylammonium acetate (1 M) and centrifugation, the supernatant was concentrated to 30% of the original volume in a speedvac. Analysis and purification was performed on a Waters HPLC system (Waters Alliance 2695 with PDA 2996, preparative HPLC: 1525EF with 2482 UV detector) with VP 250/10 Nucleosil 100-7 C 18 columns from *Macherey Nagel* using a gradient of 0.1 M triethylamine/acetic acid in water and 80% acetonitrile. The quality of the strands was determined by MALDI-MS. The forward and reverse oligos were combined and annealed in 10mM Tris pH8; 50 mM NaCl and 1 mM EDTA. Biotin-14-ATP was used to fill in the TT-overhang using Klenow exo-, followed by purification of the DNA on sephadex-G50 columns.

DNA Pull-Downs

For each DNA pull-down, 10 μg of DNA (see Table S4) was immobilized on 75 μl of Dynabeads MyOne C1 (Invitrogen) by incubating for 1 hr at room temperature in a total volume of 350 μl of DNA binding buffer (1M NaCl, 10mM Tris-HCl pH8, 1mM EDTA pH 8 and 0.05% NP40. Coupling of the DNA to the beads was always verified by agarose gel electrophoresis. Beads containing immobilized DNA were then incubated with 400 μg of nuclear extract in a total volume of 600 μl of protein binding buffer (50mM Tris-HCl pH8, 150 mM NaCl, 1mM DTT, 0.25% NP40 and complete protease inhibitors (Roche, EDTA-free)) in the presence of 10 μg poly-dAdT for 2 hr at 4°C. Baits were then washed three times with 0.5 ml of protein binding buffer after which beads containing different DNA modifications and different SILAC labels were combined and loaded on 4%–12% NuPage gradient gels (Invitrogen) (for example, C-beads with light extract were combined with mC beads that were incubated with heavy extract; forward pull-down).

For the label-free analysis, three separate DNA pull-downs with every bait were performed and each of these was loaded on gel separately. For Western blot validation using endogenous antibodies, protein amounts were scaled down by a factor of four.

In-Gel Digestion

Samples were analyzed on 4%–12% precast NuPage gels (Invitrogen) and subsequently stained using colloidal blue staining (Invitrogen). Each lane was cut into 8-12 gel slices and each of these slices was subjected to in-gel trypsin digestion overnight. Tryptic peptides were desalted on Stage-tips (Rappsilber et al., 2003).

Mass Spectrometry

Peptides were separated on an EASY-nLC (Proxeon) connected online to an LTQ-Orbitrap-Velos mass spectrometer. Spectra were recorded in CID mode. A gradient of organic solvent (5%–30% acetonitrile) was applied (120 min) and the top 15 most abundant peptides were fragmented for MS/MS, using an exclusion list of 500 proteins for 45 s.

Data Analysis

Raw data were analyzed using Maxquant version 1.2.2.5 and the integrated Andromeda search engine against protein database ipi.MOUSE.v3.68. Using Perseus, data was filtered for contaminants, reverse hits, number of peptides (>1) and unique peptides (>0). Ratios were logarithmitized (log2) and groups (consisting of forward and reverse) were defined. Proteins were filtered to have at least 2 valid values in one of the groups and missing values were imputed based on a normal distribution (width = 0.2 and shift = 0), after which Significance B was calculated (Benj.Hoch.FDR = 0.05). Scatterplots were made using R. Proteins were defined to be significant when both forward and reverse significance $p < 0.05$ and minimal ratios were > 2 in both experiments. The H/L ratios shown in Figure 2A–C were calculated using the formula (log(forward ratio) – log(reverse ratio))/2.

Label-Free Quantification

LFQ values, based on the summed measured intensities of all tryptic peptides of a single protein, allows for comparing the relative abundance of a protein in different pull-downs. Changes in the LFQ intensity of a protein between pull-downs with different DNA modifications indicate preferential binding of that protein to one modification over another. Raw data were analyzed using Maxquant version 1.2.2.5 and protein database ipi.MOUSE.v3.68.fasta. Settings that were different from SILAC analyses were: multiplicity set at 1 and the options for 'label-free quantification' and 'match between runs' were selected Using Perseus, data were filtered for contaminants, reverse hits, number of peptides (>1) and unique peptides (>0). LFQ intensities were logarithmitized (log2). After defining each triplicate as a group, proteins were filtered to have at least 3 values in a single group, assuming that when a protein binds specifically to one modification, it may only be identified in the three pull-downs with that modification. The missing values were imputed using a normal distribution (with = 0.3, shift = 1.8). Groups were defined and the significant outliers were calculated using ANOVA (FDR = 0.025, $S_0 = 2$ for NPC and $S_0 = 0$ for brain). Correlation based clustering was done in R for the ANOVA-outliers only, using LFQ-values which had been normalized by row-mean-subtraction.

Purification of GFP-Fusion Proteins for EMSA

HEK293T cells were transfected with expression constructs encoding for GFP-Tdg or GFP-Dnmt1. 48 hr after transfection, cells were lysed 30 min on ice in Lysis-Buffer (50 mM NaH_2PO_4, 150 mM NaCl, 10 mM Imidazole, 0.5 mM EDTA, 0.5% Tween, 1 g/l DNaseI, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$, 1 mM PMSF, 1x Protease-Inhibitor-Mix M (SERVA Electrophoresis GmbH)). The lysate was cleared by centrifugation (14 000 rpm, 10 min, 4°C) and incubation of the supernatant with equilibrated Ni-NTA beads (QIAGEN) in IP-buffer (50 mM NaH_2PO_4, 150 mM NaCl, 10 mM Imidazole, 0.5 mM EDTA, 0.05% Tween). After centrifugation (2200 rpm, 2 min) the supernatant was added to equilibrated GBP-Ni-NTA beads (Chromotek) in IP-buffer and rotated for 2 hr at 4°C. After washing three times with Washing-Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, 0.1% Tween), the GFP-fusion proteins were eluted with 50 mM NaH_2PO_4, 150 mM NaCl, 250 mM Imidazole, 0.05% Tween. The elution buffer was exchanged to 20 mM TrisHCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT for EMSA reactions. The glycosylase activity of the purified Tdg was tested on T/G mismatch containing DNA (data not shown).

Electrophoretic Mobility Shift Assays of Fluorescent DNA Oligonucleotides with GFP-Fusion Proteins

GFP-Tdg and GFP-Dnmt1 at decreasing concentrations (200 nM, 150 nM, 100 nM, 50 nM, 25 nM 12.5 nM and 6.25 nM) were incubated for 30 min on ice with a 1:1 mixture of two distinctly labeled fluorescent 42mers (see Table S4, MWG-Eurofins, 250 nM each) containing a central CG site. The ATTO647N-labeled oligonucleotide contains only canonical bases whereas the ATTO550-labeled DNA bears different cytosine modifications (C, mC, hmC, fC and caC) or an abasic site at the CG position on both strands. Samples were run on a 6% non-denaturating polyacrylamide gel (pre-run 1 hr with 0.5x TBE) at 4°C. Oligonucleotide- and GFP-fluorescence was detected by the Typhoon Scanner (GE Healthcare). Quantifications were done with ImageJ.

DNA Purification and Analysis after NE Incubation

DNA pull-downs were performed as described above, but all amounts were scaled up 3 times. As a control, all baits were also incubated in buffer plus poly-dAdT without nuclear extract for 2 hr at 4°C. The beads were washed 3x using 1 ml of incubation buffer

and 1x using 1M NaCl, 10mM Tris-HCl pH8, 1mM EDTA and 0.05% NP40, to reduce contamination with DNA from the nuclear extracts. Beads were then resuspended in 200 uL incubation buffer and DNA was purified using phenol/chloroform extraction from the beads. The DNA-strands were finally dissolved in milliQ, enzymatically hydrolyzed to nucleosides and analyzed in triplicate (15 pmol each) by MALDI-MS or LC-MS/MS.

GFP Pull-Downs

HeLa wild-type cells and a BAC-GFP transgenic cell line (WDR76) were cultured in SILAC medium for eight cell doublings, after which cells were expanded and nuclear extracts were made. For each pull-down 20 μl of GFP-trap slurry (50% v/v; Chromotek) was washed and incubated for 90 min at 4°C with 1 mg of nuclear extract of (WT L, WT H, GFP L and GFP H) in a total volume of 400 μl incubation buffer (300 mM NaCl, 20 mM HEPES KOH pH 7.9, 20% v/v glycerol, 2 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP40, complete protease inhibitor w/o EDTA (Roche) and 0.5 mM DTT) in the presence of 2 μl ethidium bromide (10mg/ml, final concentration 50 μg/ml). Beads were then washed two times with this incubation buffer, twice with PBS + 0.5% NP40 and two times with PBS only. During the last wash, beads of light control and heavy GFP pull-down were mixed and vice versa. Bound proteins were then subjected to on-bead trypsin digestion (Hubner and Mann, 2011) and significant proteins were determined as described for the SILAC DNA pull-downs. For the GFP-Tdg pull-down, mESC were cultured in normal mESC medium and a transient transfection with the GFP-Tdg plasmid (15 μg/15cm dish) using PEI (ratio DNA:PEI = 1:3) was performed. GFP-Tdg was purified in a label-free method, thus 3 pull-downs were performed using GFP-trap beads and as a control the same extract was incubated in triplicate with control blocked agarose beads (Chromotek). For each pull-down 20 μl of bead slurry (50% v/v) was washed and incubated for 90 min at 4°C with 1 mg of the nuclear extract in a total volume of 400 μl incubation buffer (150 mM NaCl, 50 mM Tris-HCl pH8, 1mM DTT, 0.25% NP40 and complete protease inhibitor w/o EDTA (Roche)) to mimic the conditions of the DNA pull-downs as close as possible in the presence of 50 μg/ml of ethidium bromide. Beads were then washed two times with 0.5 ml of incubation buffer, twice with PBS + 0.5% NP40 and two times with PBS only, after which bound proteins were on-bead digested. The Tdg-GFP purification was analyzed using a permutation-based t test (FDR = 0.05 & S0 = 3) to determine significant interactors.

Recombinant Protein Expression/DNA Pull-Downs

Klf4(aa396-483), KDM2B(aa606-647), Cxxc5(aa234-293), MBD3(aa1-77) and Rfx5(aa85-173) were cloned into PRP256NB vector, containing a GST with a C-terminal multiple cloning site. Uhrf2 (aa416-626) GST fusion was kindly provided by Dr. Jiemin Wong. hMBD2b-GST was provided by Stefanie Bartels.

Protein expression was performed in *E. coli* BL21-DE3 Codon+ by growing them at 37°C until OD_{600} of 0.5, after which expression was induced using 1mM IPTG and culturing for 3 additional hours at 25°C. Cells were lysed in 50 mM Tris-HCl pH 8.0/ 20% sucrose/ 1 mM EDTA/ 0.5 mM PMSF/ 1 mM DTT/ 1 μg/ml aprotinin using lysozyme and Triton X-100 and repeated freeze-thawing. Bacterial debris was removed by ultracentrifugation.

DNA pull-downs were performed using 2.5 μg DNA coupled to 16.75 μl MyOne beads and 5 μl of bacterial lysate/ nuclear extract in 250 μl total volume (50mM Tris-HCl pH8, 150 mM NaCl, 1mM DTT, 0.25% NP40 and complete protease inhibitors (Roche, EDTA-free)) in the presence of 2.5 μg polydAdT. After 3 times of washing with 0.5 ml of this buffer, beads were boiled in sample buffer. 5% of the input material and 100% of the bound material was loaded on gel for Western blot analyses.

Western Blot

Gels were blotted onto nitrocellulose membranes. Blots were blocked using 5%-skimmed milk in TBST. Used antibodies are: MouseαMBD3 (IBL, 3A3), GoatαMBD2 (Everest Biotech, EB07538), RabbitαRBBP5 (Bethyl, BL766), GoatαJun-C (SantaCruz), RabbitαDNMT1 (Abcam, ab13537), RabbitαCarf (Abcam, ab140519), RabbitαGST (Santa Cruz, SC-138), RabbitαGFP (home made), DonkeyαmouseHRP and DonkeyαRabbitHRP.

NMR-Spectroscopy-Based Interaction Study of Rfx5 and mC DNA

The winged-helix (WH) domain of human Rfx5 (residues 85-173, plus 18 additional residues at the N-terminus) was expressed as a GST-fusion in BL21-DE3 Codon+ bacterial strains at 25°C in M9 minimal medium with $^{15}NH_4Cl$ and/or ^{13}C-glucose. The protein was purified by binding to a Glutathione agarose (GA) column (Sigma) and eluted with 50 mM reduced glutathione (Sigma). After thrombin digestion, Rfx5-WH was purified over a Sephadex-75 (HiLoad 16/60) column in buffer A (50mM KPi pH 7, 100 mM KCl, 5 mM DTT, 0.5 mM PMSF and protease inhibitors). NMR samples used for backbone assignment contained ca. 0.3 mM WH domain in 90/10% H_2O/D_2O in buffer A. NMR spectra (HNCACB, CBCACONH, HNCA, and HNCO) were recorded at 298K on a 600 or 750 MHz Bruker Avance II spectrometer, processed using the NMRPipe package (Delaglio et al., 1995), and analyzed using CcpNmr Analysis (Vranken et al., 2005). Backbone assignments were obtained for 90 out of 106 residues in the Rfx5-WH construct.

Interaction study with mC DNA was done using an 18bp DNA fragment (see Table S4; (Biolegio)) carrying a single mC on each strand. Annealed DNA oligos were lyophilized and dissolved in buffer A to a stock concentration of 620 μM. The Rfx5-WH domain (103 μM) was titrated with mC DNA, and after each addition (11 points in total) the ^{1}H-^{15}N HSQC spectrum of Rfx5-WH as recorded (298K / 600 MHz Bruker Avance II). Since the DNA sequence used is not palindromic, the two mC may be inequivalent in their capability to bind Rfx5. At high DNA:Rfx ratios, several peaks appear split in two in a roughly 1:1 ratio, suggesting that although the Rfx5-WH domain senses the distinct DNA sequence context of the two mC sites, it recognizes both with similar affinities

(data not shown). Although a few residues showed non-linear titration profiles, most peak displacements were linear. For further analysis, the binding sites were treated as being independent, resulting an apparent dissociation constant for the Rfx5-WH – mC interaction.

Titration data were fitted using using MatLAB scripts (MATLAB version 7.13.0, The MathWorks Inc., 2011) using the fast-exchange assumption for residues with observed chemical shift perturbations between 10 and 30 Hz (fast-exchange regime; 15 residues) in a global fit. The error bars for the observed peak position was set to 1.2 Hz. The overall reduced chi-square for the fit was 2.17. The error in the fitted K_D was estimated using 1000 MonteCarlo simulations resulting in an average of 3.2 ± 0.9 μM. The range of acceptable fits was examined using F-statistics from a grid search, resulting in 95% probability limits of 10 nM < K_D < 16 μM.

A homology model of Rfx5-WH domain was constructed on the basis of the DNA-bound crystal structure of the Rfx1 winged helix domain (PDB-id: 1DP7; 35% sequence identity) using the SwissModel server (Schwede et al., 2003). The model was validated against the predicted backbone dihedral angles from the observed backbone chemical shifts using TALOS+ (Shen et al., 2009). The model of mC bound to the putative binding pocket was constructed in PyMol by superimposing the mC DNA from the UHRF1-mDNA crystal structure (PDB-id 3CLZ) onto the Rfx1-bound DNA, such that the binding pocket and mC are aligned. To achieve a proper fit, the mC base was set to a *syn*-conformation. The side chains orientations of K110 and Y161 were adjusted manually to minimize clashes.

In Silico Analysis of Klf4 ChIP-Seq Profile and Bisulfite Sequencing Data in mESCs Cells and NPCs

Klf4 binding data (ChIP-seq) was taken from (Chen et al., 2008) (GSM288354), and DNA methylation data (whole-genome bisulfite sequencing) was taken from (Stadler et al., 2012) (GSE30202). Annotated Klf4 peak centers (mESC) were extended with 50 bp on both sides to obtain 100-bp Klf4 binding regions. The mean CpG methylation of each 100-bp region was calculated for mESCs and NPCs and plotted as a scatterplot (Suppl. Figure S1B). For each quadrant of this scatterplot, the genomic distribution of the 100-bp Klf4 binding regions was calculated and plotted as a pie chart (Suppl. Figure S1C). Promoters were defined as -/+ 1 kb upstream and downstream from transcription start sites of the RefSeq mm9 annotation. The DNA sequences of the 100-bp Klf4 binding regions were used to search for the GGCGTG motif, and the CpG methylation within these motifs was calculated. The obtained distribution was plotted as a histogram (Suppl. Figure S1D). Analyses were done using Python, Perl and R.

iBAQ

iBAQ was performed essentially as described in (Schwanhäusser et al., 2011). 3.3 μg of UPS2 standard (Sigma) was added to 10 μg of nuclear extract, which was digested using the FASP protocol (Wiśniewski et al., 2009). In addition, 100 μg of NE was digested using FASP after which the peptides were separated into 8 fractions using SAX. Each of these samples was measured during a 4 hr gradient of LC-MSMS. A linear fit was made for the known amounts of the UPS2 standard and the measured iBAQ intensities in the 10 μg sample. Using this curve, iBAQ values of all other identified proteins in the 10 μg sample were converted to amounts. A linear fit was again made using these amounts and the iBAQ values in the eight SAX fractions, which were used to extrapolate absolute protein amounts of all identified proteins in these samples.

Cell Culture and Transfection Experiments

The mammalian GFP-Tet1cd expression vector was generated by PCR amplification of mouse (E14) cDNA encoding the catalytic domain of Tet1 (amino acids 1365 to 2057) and N-terminal GFP fusion. HEK293T cells were grown at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, *Invitrogen 41966-029*) supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin. Cells were passaged at 80% confluency. All transfections were performed using the *jetPRIME* system (*PEQLAB Biotechnologie GmbH*) according to the manufacturer's instructions. HEK293T cells were seeded 24 hr prior to transfection at a density of $3x10^6$ cells per 75 cm^2-flask and incubated in 10 ml of medium at 37°C and 5% CO_2 for 24 hr. Cotransfection of GFP-Tet1cd plasmid (6 μg) either with mouse Uhrf2-GFP plasmid DNA (6 μg) (Pichler et al., 2011) or 6 μg of pCMV6-Cdk5Rap1-v2 (*Origene RG216600*) as an unrelated control was carried out in a 75 cm^2 flask containing 10 ml of fresh medium. The transfection solution (500 μl of *jetPRIME* buffer, 12 μg of plasmid DNA and 24 μl of *jetPRIME* reagent) was added to the medium and the cells were incubated at 37°C and 5% CO_2 for 48 hr. After removal of the medium the cells were washed once with PBS and then lysed for DNA extraction according to (Münzel et al., 2010). The DNA was enzymatically digested to the nucleosides and subsequently analyzed by LC-ESI-MS/MS.

LC-MS/MS Analysis of Genomic DNA and Synthetic DNA

The following LC-MS/MS method for the quantification of DNA-nucleosides is based on a further development of our precise and sensitive isotope dilution method ((Pfaffeneder et al., 2011) and manuscript in preparation). In the following we shortly summarize the parameters of the method. Genomic or synthetic DNA was enzymatically digested to the nucleoside level. A specific amount of internal standards with a stable isotope label were spiked to the digestion mixture for precise quantification. The following labeled nucleosides were used as internal standards: [$^{15}N_2$]-dC, [D_3]-mC, [D_2,$^{15}N_2$]-hmC, [$^{15}N_2$]-fC, [$^{15}N_2$]-caC and [D_3]-dT. In case of genomic DNA the dC- or dG-content was determined by LC-UV-Detection.

LC-MS/MS analysis was performed on an Agilent 6490 triple quadrupol mass spectrometer coupled to an Agilent 1290 UHPLC system. The general source-dependent parameters were as follows: Gas Temp 50°C, Gas Flow 15 L/min, Nebulizer 30 psi, Sheath

Gas Heater 300°C, Sheath Gas Flow 11 L/min, Capillary Voltage 2500 V and Nozzle Voltage 500 V. For compound-dependent parameters used for genomic DNA see Table S5, for compound-dependent parameters used for synthetic DNA see Table S6. The transitions of the nucleosides were analyzed in the positive ion selected reaction monitoring mode (SRM) operating MS1 and MS2 under unit mass resolution conditions.

For the analysis a C8 column from Agilent was used (1.8 μm, 2.1 mm x 150 mm). The compounds were separated by a gradient using water and acetonitril with 0.0075% formic acid. The column temperature was maintained at 30°C. The flow rate was 400 μl min^{-1}, and the injection volume amounted to 29 μL. The effluent up to 1.5 min (total run time of 12 min) was diverted to waste by a Valco valve in order to protect the mass spectrometer.

Validation of Quantification Method for Genomic DNA Modifications

In accordance with the FDA guidance for bioanalytical method validation, linearity, precision, and accuracy (i.e., recovery determined from spiked matrix samples) of the established method were investigated. Validation for the established LC-MS/MS quantification method was based on five different series (i.e., calibration functions and quality control samples) accomplished on different days. Calibration standards were analyzed at least in triplicates. Quality control samples to evaluate accuracy, intra- and inter-batch (see intra- and inter-assay) precision were determined using a biological sample with internal standards. Furthermore, each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Additionally, acceptable accuracy (80%–120%) as well as precision (<20% RSD) was required. Linear regression was applied to obtain calibration curves. Therefore, the peak area ratio (y) of the unlabeled nucleoside to the internal standard versus the concentration ratio of the unlabeled nucleoside to the internal standard (x) was plotted. Calibration functions were calculated without weighting. Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of −20°C was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch. Short-term stability at room temperature was studied in overnight experiments. In this process, the results of quantification by LC-ESI-MS/MS directly after preparing the samples were compared with those obtained from samples kept overnight at room temperature.

SUPPLEMENTAL REFERENCES

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475–1489.

Lavery, D.J., and Schibler, U. (1993). Circadian transcription of the cholesterol 7 alpha hydroxylase gene may involve the liver-enriched bZIP protein DBP. Genes Dev. *7*, 1871–1884.

Münzel, M., Globisch, D., Trindler, C., and Carell, T. (2010b). Efficient synthesis of 5-hydroxymethylcytosine containing DNA. Org. Lett. *12*, 5671–5673.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. *31*, 3381–3385.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2012). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

Welz, R., and Muller, S. (2002). 5-(benzylmercapto)-1H-tetrazole as activator for 2 '-O-TBDMS phosphoramidite building blocks in RNA synthesis. Tetrahedron Lett. *43*, 795–797.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

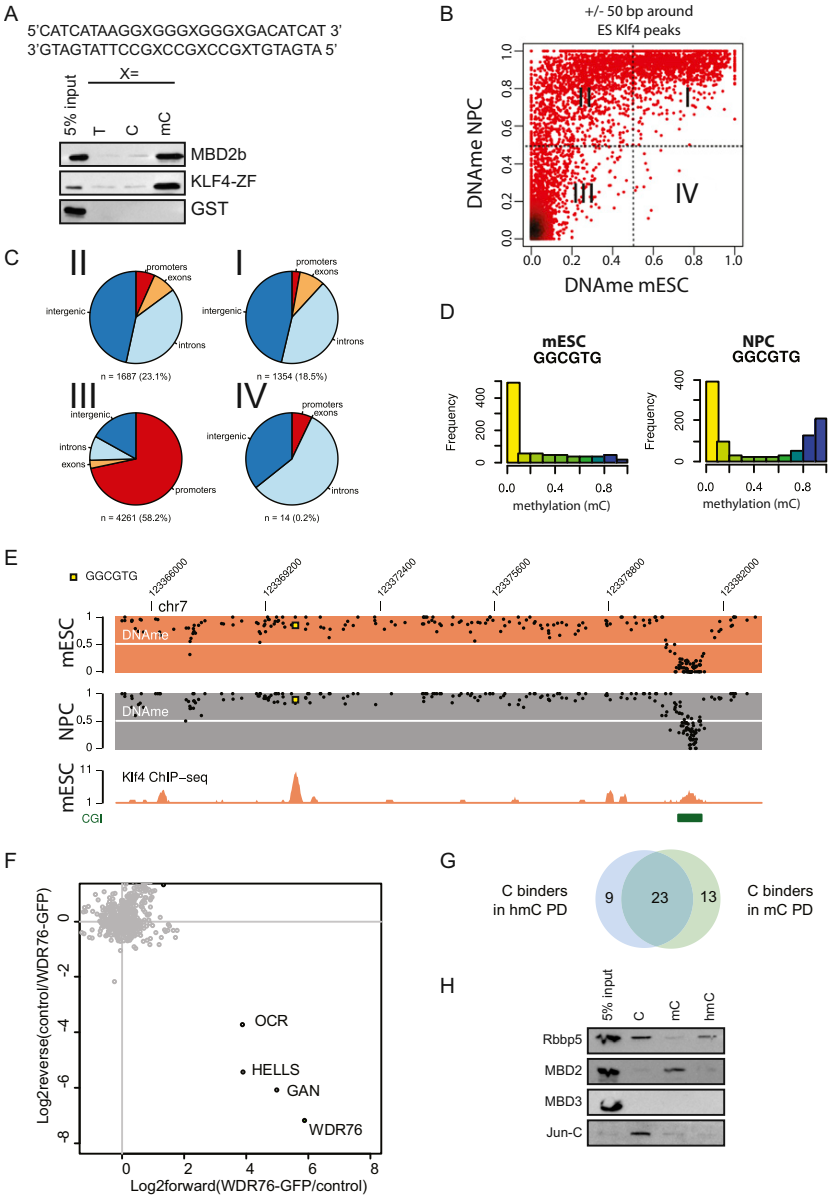


Figure S1. Genome-wide Localization of Klf4 Partially Correlates with DNA Methylation, Related to Figure 1

(A) DNA pull-downs with recombinant GST-fusion proteins of DNA binding domains and Western blotting analysis.

(B) DNA methylation of Klf4 sites in ES cells and NPCs. Whole-genome bisulfite sequencing was used to determine DNA methylation within a window of +/− 50 bp around Klf4 peak centers. Darker coloring indicates high density of datapoints.

(C) Pie charts showing the genomic distribution of Klf4 sites as presented in the different quadrants of (B).

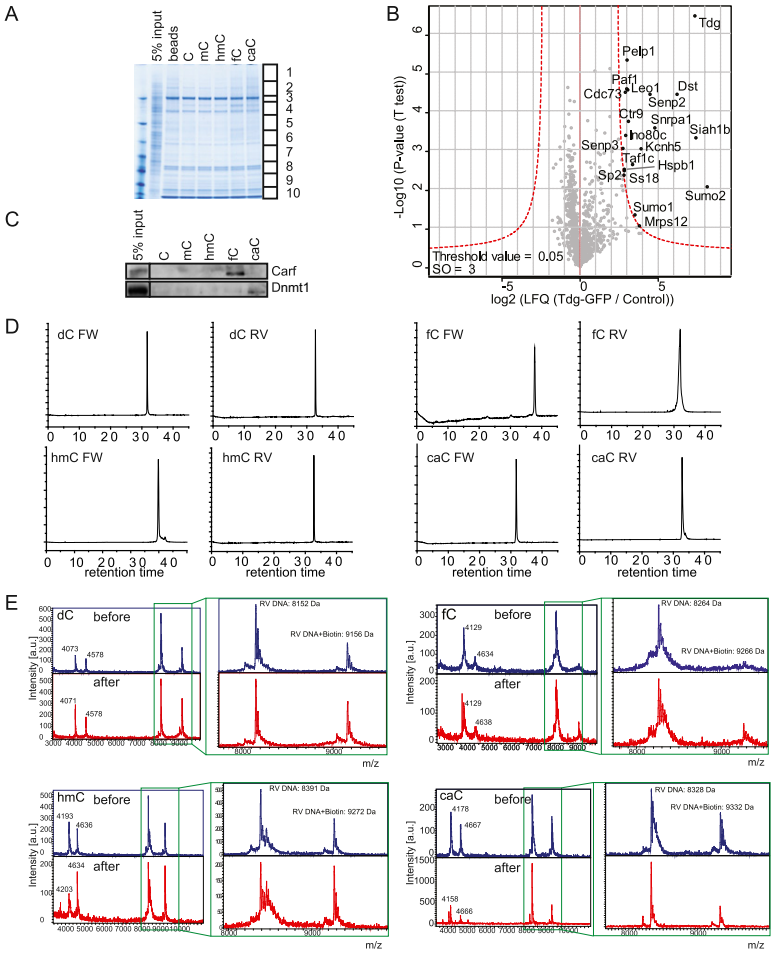
(D) Distribution of DNA methylation specifically within the GGCGTG sequence present underneath Klf4 sites.

(E) Example of DNA methylation profiles and Klf4 binding (ChIP-seq), showing binding of Klf4 to both methylated and unmethylated sites. Yellow squares indicate the presence of the GGCGTG sequence underneath Klf4 sites.

(F) SILAC-based GFP-purification from HeLa cells stably expressing WDR76-GFP. Significant interactors are indicated in black (high forward WDR76-GFP/control ratio, low reverse control/WDR76-GFP ratio).

(G) Venn diagram showing the overlap of C-specific readers in the mC and hmC DNA pull-downs from mESC nuclear extracts.

(H) Validation of C and mC specific binders by DNA pull-downs in HeLa nuclear extract and Western blotting for the endogenous proteins.



F

	NE incubation	dC/dT	hmC/dT	β-fC/dT	α-fC/dT	caC/dT
dC	w/o	0.9340				
dC	2 hr	0.9172				
hmC	w/o	0.0738	0.7258	0.0009		
hmC	2 hr	0.0745	0.7102	0.0009		
fC	w/o	0.4251		0.2818	0.1794	0.0035
fC	2 hr	0.4172		0.2950	0.1811	0.0034
caC	w/o	0.3780				0.5729
caC	2 hr	0.3755				0.5414

Figure S2. Identification of TDG Interactors, Western Blot Verification of fC and caC Interactors, and Validation of Bait DNA Quality, Related to Figure 2

(A) The indicated immobilized DNA baits were incubated with mouse nuclear extract. Following washes, bound proteins were analyzed by colloidal blue staining. Note that the elution profile of all these baits looks similar, indicating that specific interactors are masked by a large number of high abundant background binders.

(B) Volcano plot of a label-free GFP-Tdg pull-down in mESC nuclear extract. Significant interactors of GFP-Tdg are identified by permutation-based t test (FDR = 0.05 & S0 = 3). The LFQ intensity of the GFP pull-down over the control is plotted against the –Log10 (p-value). The red line indicates the permutation-based FDR. Also see Table SI.

(C) Western blot validation of the fC-specific binding of Carf and caC-specific binding of Dnmt1 in mESC nuclear extract. A single empty lane was removed from the blot.

(D) HPL-Chromatograms of the purified FW and RV DNA obtained from solid phase DNA synthesis showing the purity of the employed strands.

(E) The mass spectra of the DNA before (blue) and after (red) NE incubation as determined by MALDI MS showing the expected m/z before and after NE incubation. Major alterations of the DNA like degradation or strand breaks can be excluded.

(F) Synthetic DNA-strands which were used for DNA pull-downs were compared without (w/o) and with nuclear extract (NE) treatment (2 hr, 4°C) to proof the stability of the indicated modifications. The quantification of the nucleoside content was carried out by LC-MS/MS. For this, the DNA was digested to the nucleoside level and spiked with a specific amount of the following internal standards for precise quantification: [$^{15}N_2$]-dC, [D_2,$^{15}N_2$]-hmC, [$^{15}N_2$]-fC, [$^{15}N_2$]-caC and [D_3]-dT. The absolute amount (pmol) of each nucleoside was calculated by calibration curves (not shown). Depicted are ratios of the modified nucleoside (pmol) to deoxy-thymidine (dT; pmol), which were obtained from three independent measurements. The relative standard deviation was between 0.3%–6.2%. No or only marginal loss of the modified nucleosides was observed.

A

mESC

	GO biological process	-log(Benj.Hoch.FDR)
C	RNA biosynthetic process	7.61
	cellular biosynthetic process	6.65
	biosynthetic process	6.48
	DNA recombination	5.14
	DNA metabolic process	5.06
	nucleobase-containing compound metabolic process	4.49
	transcription, DNA-dependent	4.48
	nucleic acid metabolic process	4.42
	cellular macromolecule biosynthetic process	4.30
	macromolecule biosynthetic process	4.29
	cellular nitrogen compound metabolic process	4.28
	nitrogen compound metabolic process	4.21
	regulation of macromolecule biosynthetic process	3.85
	regulation of cellular macromolecule biosynthetic process	3.85
	transcription initiation, DNA-dependent	3.82
	regulation of transcription, DNA-dependent	3.80
	regulation of cellular biosynthetic process	3.77
	regulation of biosynthetic process	3.75
	DNA repair	3.71
	regulation of gene expression	3.69
	regulation of nucleobase-containing compound metabolic process	3.53
	regulation of nitrogen compound metabolic process	3.53
	regulation of RNA metabolic process	3.33
	regulation of primary metabolic process	3.32
	regulation of macromolecule metabolic process	3.31
	response to DNA damage stimulus	3.26
	regulation of cellular metabolic process	3.12
	regulation of metabolic process	3.05
	cellular response to stress	2.97
	RNA metabolic process	2.72
	cellular macromolecule metabolic process	2.58
	response to stress	2.57
	macromolecule metabolic process	2.37
	eye morphogenesis	2.05
	cellular metabolic process	2.04
	primary metabolic process	1.86
	metabolic process	1.72
mC	transcription, DNA-dependent	2.84
	RNA biosynthetic process	2.81
	regulation of cellular biosynthetic process	1.76
	macromolecule biosynthetic process	1.76
	one-carbon metabolic process	1.74
	regulation of biosynthetic process	1.73
	regulation of macromolecule biosynthetic process	1.73
	regulation of nitrogen compound metabolic process	1.72
	regulation of RNA metabolic process	1.71
	nucleic acid metabolic process	1.70
	regulation of macromolecule metabolic process	1.70
fC	DNA repair	2.71
	response to DNA damage stimulus	2.60
	DNA metabolic process	2.59
	cellular response to stress	1.79

	Cellular Compartment	-log(Benj.Hoch.FDR)
C	transcription factor TFIID complex	2.17
	TFIID complex	7.65
	MLL1 complex	6.06
	histone methyltransferase complex	4.20
	methyltransferase complex	4.37
	transcription factor complex	3.24
	nucleoplasm part	1.73

	Domains	-log(Benj.Hoch.FDR)
C	BRLZ	7.40
	bZIP_1	5.57
mC	ZnF_C2H2	2.45
	MBD	2.24

NPC

	GO biological process	-log(Benj.Hoch.FDR)
C	regulation of nitrogen compound metabolic process	5.04
	regulation of biosynthetic process	5.03
	regulation of nucleobase-containing compound metabolic process	5.01
	regulation of RNA metabolic process	5.01
	transcription, DNA-dependent	5.00
	regulation of cellular biosynthetic process	4.94
	RNA biosynthetic process	4.93
	regulation of macromolecule biosynthetic process	4.90
	regulation of cellular macromolecule biosynthetic process	4.88
	regulation of transcription, DNA-dependent	4.85
	regulation of gene expression	4.81
	cell differentiation	2.60
	positive regulation of transcription from RNA polymerase II promoter	2.51
	cellular response to metal ion	2.49
	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	2.48
	cellular response to inorganic substance	2.47
	positive regulation of biosynthetic process	2.46
	cellular response to chemical stimulus	2.45
	cellular response to calcium ion	2.45
	regulation of transcription from RNA polymerase II promoter	2.44
	gland development	2.39
	cellular developmental process	2.37
	positive regulation of RNA metabolic process	2.35
	adipose tissue development	2.33
	positive regulation of metabolic process	2.26
	positive regulation of nucleobase-containing compound metabolic process	2.21
	positive regulation of cellular biosynthetic process	2.16
	positive regulation of nitrogen compound metabolic process	2.14
	positive regulation of transcription, DNA-dependent	1.99
	positive regulation of macromolecule metabolic process	1.98
	positive regulation of gene expression	1.92
	response to calcium ion	1.88
	positive regulation of macromolecule biosynthetic process	1.85
	regulation of developmental process	1.83
	response to inorganic substance	1.79
	anatomical structure development	1.78
	regulation of cell differentiation	1.74
	response to metal ion	1.73
mC	positive regulation of macromolecule biosynthetic process	2.25
	positive regulation of cellular biosynthetic process	2.20
	positive regulation of biosynthetic process	2.17
	positive regulation of RNA metabolic process	2.14
	regulation of transcription from RNA polymerase II promoter	2.14
	positive regulation of gene expression	2.11
	positive regulation of cellular metabolic process	2.10
	dorsal spinal cord development	2.07
	positive regulation of metabolic process	2.04
	positive regulation of nucleobase-containing compound metabolic process	2.00
	positive regulation of cellular process	1.95
	positive regulation of nitrogen compound metabolic process	1.94
	positive regulation of biological process	1.86
hmC	DNA metabolic process	4.97
	DNA repair	3.91
	response to DNA damage stimulus	3.61
	base-excision repair	3.32
	cellular response to stress	2.96

	Cellular Compartment	-log(Benj.Hoch.FDR)
C	histone methyltransferase complex	2.02
	intracellular membrane-bounded organelle	1.86

	Domains	-log(Benj.Hoch.FDR)
C	bZIP_1	14.42
	zf-CXXC	7.91
	MBT	1.97
	Jun	1.77
mC	RFX_DNA_binding	1.98
	MBD	1.81
	Homeobox	1.80

Brain

	GO biological process	-log(Benj.Hoch.FDR)
C	transcription, DNA-dependent	5.32
	RNA biosynthetic process	5.31
	regulation of biosynthetic process	5.24
	regulation of transcription, DNA-dependent	5.18
	regulation of macromolecule biosynthetic process	5.02
	regulation of nucleobase-containing compound metabolic process	5.00
	regulation of nitrogen compound metabolic process	4.99
	regulation of cellular macromolecule biosynthetic process	4.99
	regulation of gene expression	4.95
	regulation of cellular biosynthetic process	4.95
	regulation of RNA metabolic process	4.94
	regulation of primary metabolic process	4.70
	regulation of macromolecule metabolic process	4.65
	regulation of metabolic process	4.33
	DNA recombination	4.18
	regulation of cellular metabolic process	4.15
	cellular macromolecule biosynthetic process	4.08
	biosynthetic process	3.59
	response to DNA damage stimulus	2.75
	cellular response to stimulus	2.61
	cellular response to stress	2.58
	regulation of transcription from RNA polymerase II promoter	2.58
	DNA repair	2.47
	anatomical structure morphogenesis	2.46
	cellular response to calcium ion	2.15
	histone H4 acetylation	2.13
	eye morphogenesis	2.06
	response to calcium ion	2.03
	cellular response to metal ion	2.01
	cellular response to inorganic substance	2.00
	response to metal ion	1.71
	DNA metabolic process	1.70
hmC	DNA metabolic process	4.57
	DNA replication	3.54
	base-excision repair	1.78

	Cellular Compartment	-log(Benj.Hoch.FDR)
C	histone methyltransferase complex	5.42
hmC	Brd4-Rfc complex	6.11
	DNA replication factor C complex	3.93

	Domains	-log(Benj.Hoch.FDR)
C	bZIP_1	8.60
	Jun	1.93
mC	MBD	2.70
hmC	Rep_fac_C	3.61
	AAA	2.11

B

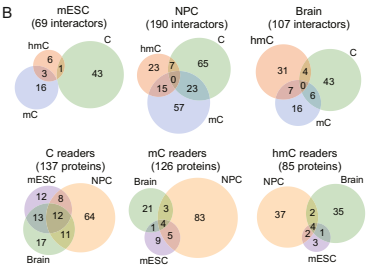


Figure S3. Modification and Cell-Type-Specific GO Term Enrichment Analysis, Related to Figures 1, 2, 3, and 4

(A) Shows GO term enrichment and enriched domains for the different baits (C, mC, hmC, fC and caC) in mESC, NPCs and adult mouse brain.

(B) Venn diagrams showing the overlap between C, mC and hmC readers within each cell type and the overlap between C, mC and hmC readers between mESCs, NPCs and adult mouse brain.

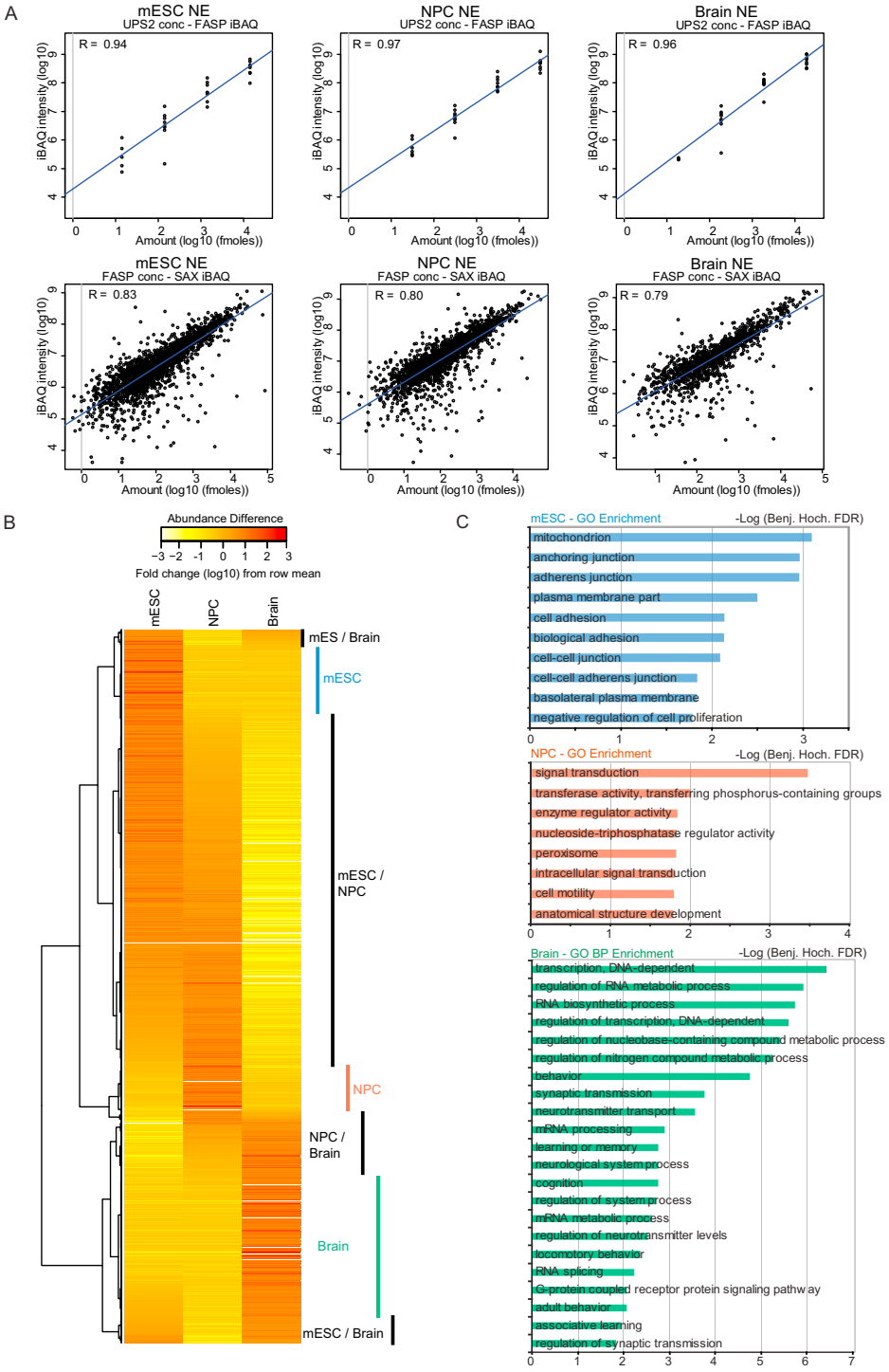


Figure S4. iBAQ Analyses of mESC, NPC, and Adult Mouse Brain Nuclear Extracts, Related to Figure 5

(A) Standard and linear regression curves for the iBAQ of protein abundance in the different nuclear extracts that were used for the DNA pull-downs.

(B) Correlation based clustering of proteins that show at least a 10-fold change in protein levels. Yellow is low abundance, red is high.

(C) GO term enrichment for mESC (indicated in blue in fig S4B), NPC (indicated in red in fig S4B) and adult mouse brain (indicated in green in fig S4) specific proteins.

2.2 Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

ARTICLE
PUBLISHED ONLINE: 18 MAY 2014 | DOI: 10.1038/NCHEMBIO.1532
nature chemical biology

Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

Toni Pfaffeneder[1,8], Fabio Spada[1,8], Mirko Wagner[1,8], Caterina Brandmayr[1], Silvia K Laube[1], David Eisen[1], Matthias Truss[2], Jessica Steinbacher[1], Benjamin Hackner[1], Olga Kotljarova[1], David Schuermann[3], Stylianos Michalakis[4], Olesea Kosmatchev[1], Stefan Schiesser[1], Barbara Steigenberger[1], Nada Raddaoui[1], Gengo Kashiwazaki[1], Udo Müller[5], Cornelia G Spruijt[6], Michiel Vermeulen[6,7], Heinrich Leonhardt[5], Primo Schär[3], Markus Müller[1]* & Thomas Carell[1]*

Ten eleven translocation (Tet) enzymes oxidize the epigenetically important DNA base 5-methylcytosine (mC) stepwise to 5-hydroxymethylcytosine (hmC), 5-formylcytosine and 5-carboxycytosine. It is currently unknown whether Tet-induced oxidation is limited to cytosine-derived nucleobases or whether other nucleobases are oxidized as well. We synthesized isotopologs of all major oxidized pyrimidine and purine bases and performed quantitative MS to show that Tet-induced oxidation is not limited to mC but that thymine is also a substrate that gives 5-hydroxymethyluracil (hmU) in mouse embryonic stem cells (mESCs). Using MS-based isotope tracing, we show that deamination of hmC does not contribute to the steady-state levels of hmU in mESCs. Protein pull-down experiments in combination with peptide tracing identifies hmU as a base that influences binding of chromatin remodeling proteins and transcription factors, suggesting that hmU has a specific function in stem cells besides triggering DNA repair.

Methylcytosine is an epigenetically important nucleobase associated with the control of transcriptional activity, genomic imprinting, X-chromosome inactivation and suppression of transposable elements[1]. Controlled formation and removal of mC at specific genomic loci is critical for correct genome programming or reprogramming during cellular differentiation[2]. Recently, it was discovered that Tet proteins (Tet1–3) oxidize mC to give the oxidized C-derived nucleobases hmC[3], 5-formylcytosine (fC)[4,5] and 5-carboxycytosine (caC)[5,6], whose biological functions are still yet unclear (**Fig. 1a**)[7]. As fC and caC are both removed by thymine DNA glycosylase (Tdg)[6,8], it is currently assumed that they serve as intermediates of an active DNA demethylation process involving base excision repair. In addition to these oxidized C derivatives, cells also contain oxidized T nucleobases such as hmU and fU. These compounds are currently known as oxidative lesions that are thought to form upon the reaction of T with reactive oxygen species (ROS)[9,10]. It was recently suggested that hmU might also be produced by deamination of hmC, a hypothesis that remains controversial[11–14]. Deamination of hmC, situated in a base pair with G (hmC:G), would give rise to hmU:G mismatches, which are known substrates for the DNA glycosylases Tdg, Smug1, Mbd4, Ung2 (ref. 15), Neil1 and Nthl1 (ref. 16). Deamination of hmC:G to hmU:G followed by mismatch repair would therefore establish an alternative pathway to active demethylation (**Fig. 1a**).

To unravel the origin of oxidized nucleobases, and of hmU in particular, in DNA from mESCs, we performed isotope tracing and quantitative MS studies using the chemically synthesized isotopologs of mC, hmC, fC, caC, hmU and fU as internal standards (**Fig. 1b** and **Supplementary Results**, **Supplementary Fig. 1**). For the assessment of oxidation products that are formed by the action of ROS, we additionally quantified 8-oxo-G because 8-oxo-G is a well-established ROS reaction product formed from G[17,18]. We show here that hmU is generated enzymatically from thymidine during stem cell differentiation by the action of the Tet enzymes. A proteomic analysis provides new insight into how genomic hmU can influence the binding of chromatin remodeling proteins and transcription factors.

RESULTS

hmU is present at elevated levels in mESCs

We first created an inventory of the named nucleosides (**Supplementary Fig. 1**) in mESCs (**Fig. 1c**) and adult cortex tissue (**Fig. 1d**). In mESCs, we observed that hmC is, as expected, the most abundant oxidized pyrimidine (10% relative to mC), followed by fC (~1–2% of hmC). The ROS marker 8-oxo-G was detected at similar levels (~45% of fC), showing that nonenzymatic, ROS-induced oxidations of nucleobases are important processes, as expected. We also found relatively high levels of fU (22% of fC). Clearly detectable were also caC and hmU. hmU, which is at the center of this study, was unequivocally detected, as shown by its retention time and its specific fragmentation pattern, which were found to be identical with that of the internal standard [D_2]hmU (**Fig. 1e**). Notably, both caC and hmU were present in comparable amounts (5% of fC). In adult mouse cortex DNA (**Fig. 1d**; for other tissues see

[1]Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany. [2]Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Berlin, Germany. [3]Department of Biomedicine, University of Basel, Basel, Switzerland. [4]Center for Integrated Protein Science at the Department of Pharmacy—Center for Drug Research, Ludwig-Maximilians-Universität München, München, Germany. [5]Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. [6]Department of Molecular Cancer Research, Cancer Genomics Netherlands, Utrecht, The Netherlands. [7]Present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. [8]These authors contributed equally to this work.
*e-mail: markus.mueller@cup.uni-muenchen.de or thomas.carell@cup.uni-muenchen.de

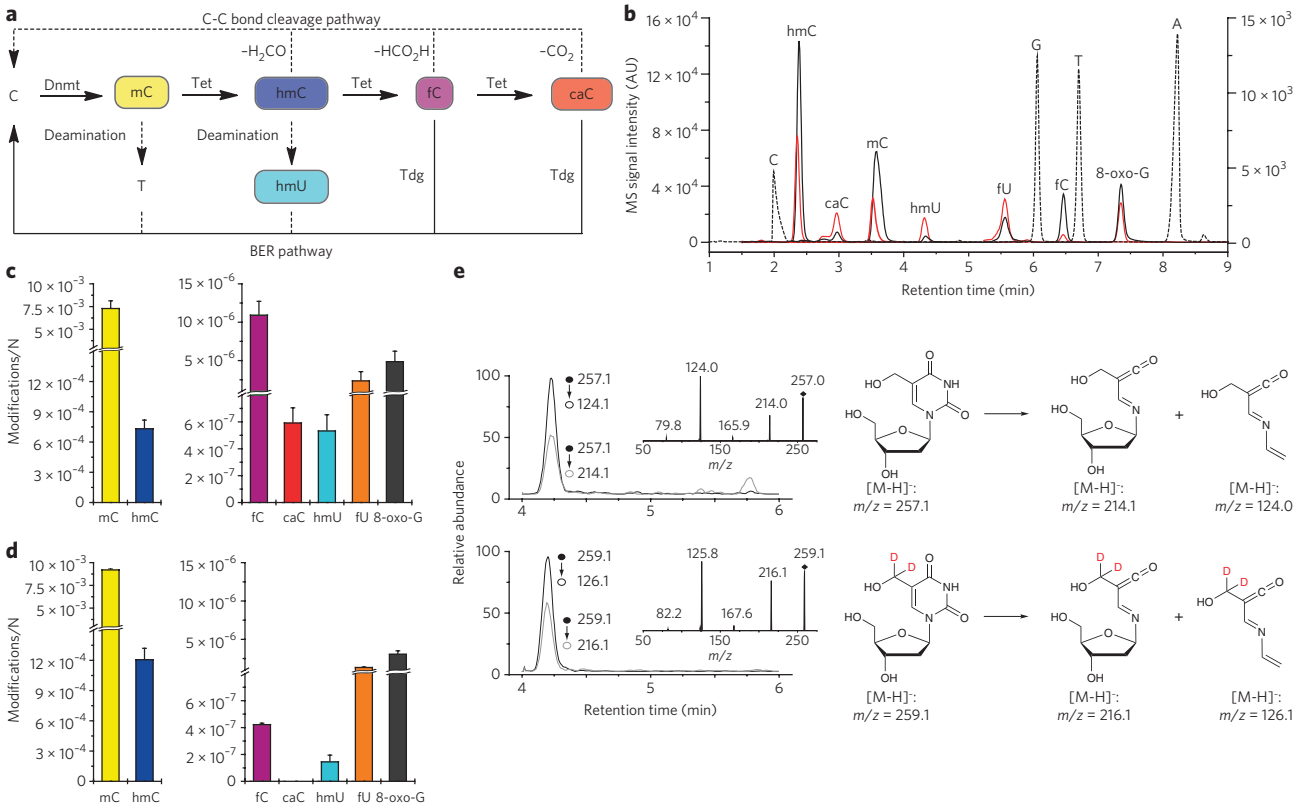


Figure 1 | Metabolism of cytosine derivatives, their detection by LC-UV-ESI-MS/MS and levels in mESCs and mouse cortex. (**a**) Potential active demethylation pathways. (**b**) Overlaid LC/UV and LC/MS/MS chromatograms of a representative DNA sample from mESCs. The dotted LC/UV chromatogram of C, G, T and A is scaled arbitrarily; the overlaid LC/MS/MS chromatograms of hmC, mC, fC and 8-oxo-G are scaled to the left *y* axis; the LC/MS/MS chromatograms of caC, hmU and fU are scaled to the right *y* axis. Red chromatograms refer to the corresponding labeled internal standards depicted in **Supplementary Figure 1**. AU, arbitrary units. (**c**,**d**) DNA modification levels per nucleoside (N) in mESCs (**c**; WT01, $n = 7$) and 3-month-old mouse cortex tissue (**d**; $n = 3$). Depicted are mean values ± s.d. (**e**) Representative LC/MS/MS chromatograms for identification and quantification of hmU. Shown are the overlaid chromatograms for the two characteristic fragment ions of hmU (top trace) and the [D_2]hmU internal standard (bottom trace) derived from a mESC DNA sample. The insets show the MS/MS full-scan spectra of synthetic hmU and [D_2]hmU matching the proposed fragmentation pathway.

Supplementary Fig. 2), the hmC levels are very high (13% relative to mC), as previously reported[19–21]. In contrast, fC was detected only in very small amounts (0.03% relative to hmC), and caC was not observed at all[22]. The detected amounts of 8-oxo-G and fU are comparable. Notably, the detected levels of hmU in the cortex were significantly lower compared to that in mESCs (27%; $P = 2.2 \times 10^{-4}$ by Student's *t*-test). Because the detected levels of 8-oxo-G were similar in the cortex (**Fig. 1d**) and in mESCs (**Fig. 1c**), the elevated hmU levels in mESCs cannot be explained by greater ROS-induced damage in mESCs and must have a different origin.

We next investigated this in more detail and quantified the levels of hmU in three different mESC lines (WT01, J1 and R1) and in a variety of tissues (**Fig. 2a**). Clearly, the hmU levels were higher in mESCs than in somatic tissues by factors of 2–15. The detected hmU levels correspond roughly to 500–1,700 hmU bases per genome in tissue and 2,900–7,800 in mESCs (**Fig. 2a**) depending on the cell type and growth conditions. We cannot explain these differences by elevated oxidative stress levels in mESCs (**Supplementary Fig. 3**). We next analyzed the levels of C, mC and the oxidized pyrimidines hmC, fC, hmU and fU as well as 8-oxo-G in somatic tissues (**Supplementary Fig. 2**) and performed a correlation and cluster analysis to reveal relationships of the modification levels (**Fig. 2b**). The data analysis confirmed that the low levels of hmU in somatic tissues correlate strongly (significant on a level <0.001) with the levels of fU and 8-oxo-G within a separated cluster. This result showed that in somatic cells, hmU and fU are ROS-induced reaction products. If we assume that in somatic tissues hmU is exclusively formed by ROS, the data showed that in mESCs, 70–80% of the detected hmU is produced by ROS-independent processes (**Supplementary Fig. 3**).

hmU formation by oxidation of T

To analyze the origin of the oxidized nucleobases in mESCs, we performed isotope tracing experiments (**Fig. 3a–c** and **Supplementary Fig. 4**). Substitution of L-methionine with [*methyl*-$^{13}CD_3$]L-methionine ([$^{13}CD_3$]L-Met) in the growth medium is known to give the labeled *S*-adenosylmethionine cofactor, which is needed for the conversion of C to mC. Replacement of L-Met by [$^{13}CD_3$]L-Met for 5 d (2 passages) in the medium at a concentration of 0.2 mM furnished 89% labeled [$^{13}CD_3$]mC, 88% labeled [*hydroxymethyl*-$^{13}CD_2$]hmC and 93% labeled [*formyl*-^{13}CD]fC (**Fig. 3b** and **Supplementary Fig. 4**). Within the detection limit (7 and 50 molecules per 10^8 nucleosides, respectively), no incorporation of the isotopes ^{13}C and D into hmU and fU was observed, showing that hmC is not the precursor of hmU. We next added isotope-labeled [$^{13}C,^{15}N_2$]thymidine ([$^{13}C,^{15}N_2$]T) to the growth medium and observed ~76% of label incorporation into T, hmU and fU (**Fig. 3a**, **Supplementary Fig. 4** and **Supplementary Table 1**). The combined data showed that hmU is not generated by deamination of hmC but by oxidation of T (T → hmU; **Fig. 3c**). Consequently, all of the detected hmU

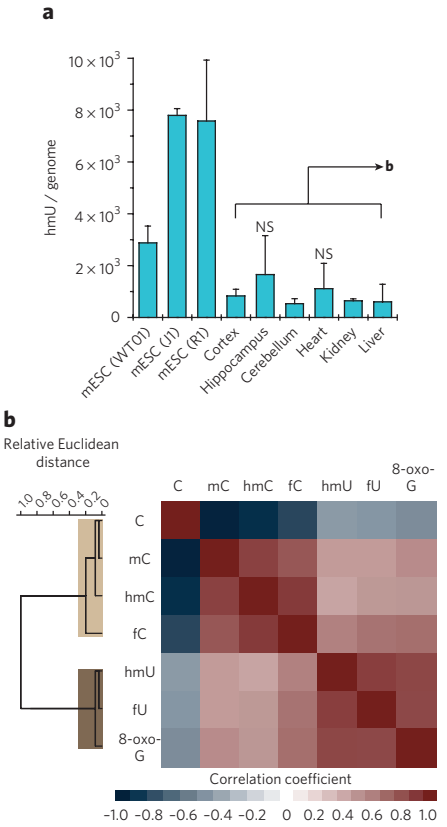


Figure 2 | hmU is present at elevated levels in mESCs compared to tissue. (**a**) hmU levels per genome in mESCs (n_{WT01} = 7 replicates, n_{J1} = 2, n_{R1} = 3) and mouse tissue (3-month-old individuals, n = 3). Levels per genome were obtained considering a mouse genome size of 2.7 × 10^9 base pairs. Depicted are mean values ± s.d. The differences between mESCs and mouse tissues are significant (P = 2.9 × 10^{-5} to 3.6 × 10^{-2}; unpaired two-tailed t-test) except for WT01 and hippocampus (P = 0.292) or heart (P = 0.069). These exceptions are due to higher hmU levels in the hippocampus and heart caused by higher background oxidation (higher 8-oxo-G levels). hmU levels normalized to oxidative background (8-oxo-G) levels are in **Supplementary Figure 3**. NS, not significant. (**b**) Unsupervised clustering analysis of Pearson correlation coefficients of 24 data sets of selected mouse organs at a defined time point (3-month-old individuals: cortex, hippocampus, cerebellum, heart, liver and kidney). Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001.

(and also fU) resides in an A base pair context (hmU:A). In this base pair, hmU is repaired by Smug1 but not Tdg[12]. This was confirmed by siRNA-mediated knockdown of Smug1 in mESC cells and HEK-293T cells. Indeed, the hmU level increased in these cells (**Supplementary Fig. 5**).

The fact that we were unable to detect hmU derived from hmC deamination, which would be situated in a base pair with G (hmU:G), could also be explained by very fast repair. If repair of the hmU:G base pair is extremely efficient, for example, because deamination and glycosylase-based repair occurs in a tight complex of the involved enzymes, we would be unable to detect this type of hmU because of low steady state levels. Indeed, it was proposed that hmC deamination and hmU excision requires a complex of the cytidine deaminase Aid and the glycosylase Tdg[12]. To assess this possibility, we performed isotope tracing experiments using [$^{13}CD_3$]L-Met in *Tdg*$^{-/-}$ mESCs stably complemented with either empty vector (control) or a minigene expressing a catalytically incompetent Tdg at near-endogenous levels. These cells are able to form the Aid–Tdg complex, but the Tdg is inactive, which should give elevated hmU levels if deamination occurs. In both cell lines, we detected high levels of labeled hmC. In the control cells, [*hydroxymethyl*-$^{13}CD_2$] hmU was not detected. However, in the cell line complemented with inactive Tdg, some labeled [*hydroxymethyl*-$^{13}CD_2$]hmU was indeed observed, albeit only in small amounts (~7% of total hmU and ~0.06% of total hmC; **Supplementary Fig. 6**). Thus, the Tdg protein is required for deamination of hmC to hmU. In wild-type (WT) mESCs, the hmU:G mismatches are obviously repaired so quickly that they do not contribute to steady state levels of hmU. All of the detected hmU was derived from T oxidation and resided in hmU:A base pairs.

Tet enzymes form hmU in correlation with mC oxidation

We next investigated whether enzymatic oxidation or ROS-dependent processes are responsible for the observed oxidation of T to hmU. To this end, we studied mESCs with genetic or functional depletions of Tet or DNA methyltransferase (Dnmt) enzymes (**Fig. 4**). Although hmU levels were maintained at normal levels in the severely hypomethylated, Dnmt-depleted cell lines, they were substantially reduced upon knockdown of Tet1 and Tet2. As the levels of fU and 8-oxo-G remained unaffected, the data establish that, though both hmU and fU are generated by T oxidation, their formation must occur by two independent processes in mESCs. The data supported the idea that fU is a ROS-created lesion similar to 8-oxo-G[9,10,17,18], whereas most hmU is generated by Tet-induced oxidation of T.

In support of enzymatic T oxidation by Tet proteins, ectopic expression of the catalytic domain of Tet1 (Tet1cd) in HEK-293T cells led to a 65-fold increase for both hmC and hmU levels. This was not observed when a catalytic mutant of Tet1 (Tet1cm) was expressed (**Supplementary Fig. 7a**). To determine whether under these conditions hmU is generated by deamination, as previously suggested[11,12], we again replaced natural L-Met with [$^{13}CD_3$]L-Met in the medium, but we did not detect incorporation of heavy isotopes into hmU, providing evidence that the elevated hmU levels do not originate from hmC deamination (**Supplementary Table 1**). Finally, we confirmed enzymatic generation of hmU *in vitro* by incubating recombinant Tet1cd with a plasmid that was premethylated by bacterial methyltransferase M.SssI. In addition to oxidation of mC to hmC, fC and caC, we detected hmU at a prominent level of 9% relative to hmC (**Supplementary Fig. 7b**), showing that the catalytic center of the Tet enzymes clearly has the capacity to oxidize T to hmU.

To further confirm that hmU is formed in mESCs in the process of epigenetic reprogramming, we analyzed the dynamic changes of mC and all of the oxidized pyrimidines plus 8-oxo-G during differentiation. It was recently shown that mC and hmC levels sharply increase when mESCs maintained in the naive state are shifted to a primed state in serum-containing medium[23–25]. To investigate global kinetics of all of the Tet-generated oxidation products under more physiologically relevant priming conditions, we used established protocols based on serum-free N2B27 medium for differentiation of naive mESCs into states resembling that of post-implantation epiblasts[26,27]. mESCs were first grown for several passages in the presence of MEK and GSK3 inhibitors (dual inhibition or 2i conditions) and LIF to induce a hypomethylated state resembling that of the naive epiblast[23–25]. The data in **Figure 5a** are averaged from three independent differentiation experiments, each performed with two cell lines in the absence of growth factors. First, we observed that the fU and 8-oxo-G levels stayed constant (**Fig. 5a**), in line with the idea that both are ROS-derived products. In contrast, fC and caC levels peaked at about 8 h. Both hmU and hmC also peak between 8 h and 16 h. Isotope tracing with [$^{13}CD_3$]L-Met under these conditions provided no evidence for switched-on deamination of hmC to hmU (**Supplementary Table 1**). At these peak levels, we estimated that mESCs contain roughly 110,000 fC bases, 4,400 caC bases and

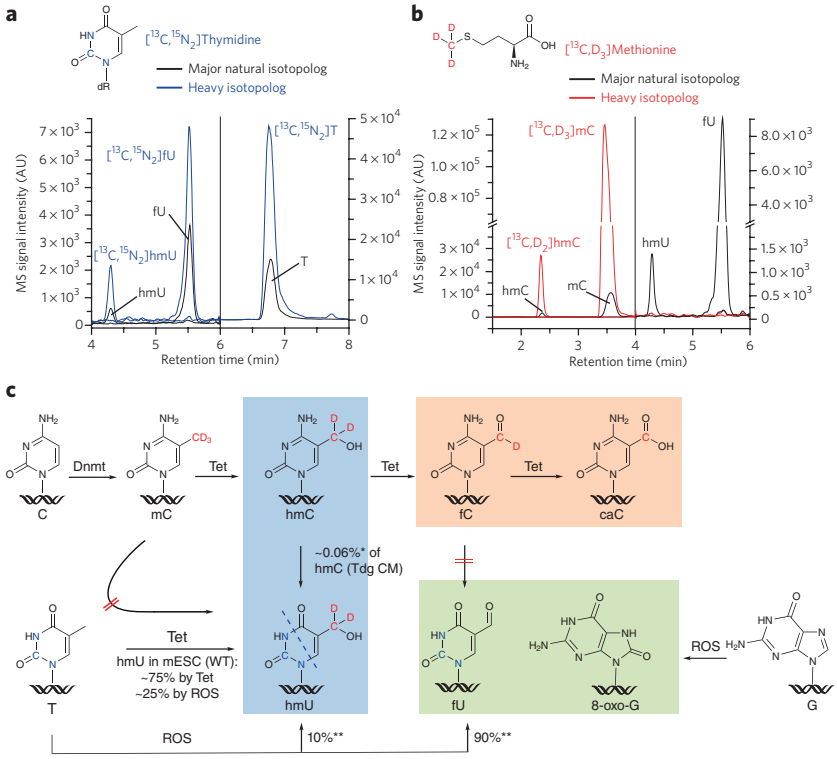


Figure 3 | hmU and fU are thymine oxidation products in WT mESCs with no detectable contribution from hmC or fC deamination. (**a**) Overlaid LC/MS/MS chromatograms of heavy ([^{13}C,^{15}N$_2$]hmU, [^{13}C,^{15}N$_2$]fU and [^{13}C,^{15}N$_2$]T; blue) and major natural (black) isotopologs of hmU, fU and T. AU, arbitrary units. (**b**) Overlaid chromatograms of heavy ([^{13}C,D$_3$]mC, [^{13}C,D$_2$]hmC, [^{13}C,D$_2$]hmU and [^{13}C,D]fU; red) and the major natural (black) isotopologs of mC, hmC, hmU and fU. (**c**) Enzymatic and ROS-dependent pathways leading to the formation of hmC, fC, caC, hmU and fU. Single asterisks denote labeled hmU generated by deamination of labeled hmC, which was observed only in Tdg catalytic mutant (CM) cells (**Supplementary Fig. 6**), representing ~7% of the total hmU content and corresponding to deamination of ~0.06% hmC. Double asterisks denote basal rates of ROS-dependent T oxidation, which were determined in HEK-293T cells, where Tet activity is lowest (3.6 hmC per 10^5 nucleosides; **Supplementary Table 2**). Here, hmU and fU roughly represent 10% and 90% of T oxidation products relative to the sum of each other.

14,000 hmU bases per genome, showing that, at its peak level, hmU is three times more abundant than caC. The dynamic peaking data allowed us to estimate half-life times for fC (7 h), caC (5 h) and hmU (4 h) during the differentiation process (**Supplementary Fig. 8**). Using the dynamic quantitative data, we performed a correlation and cluster analysis of the DNA modification levels (**Fig. 5b**). Three independent clusters were obtained. One cluster involves members of cytosine methylation (and demethylation) dynamics, including C, mC, fC and caC, but, to our surprise, not hmC. Instead, hmC groups with hmU. We therefore concluded that the formation of these modifications is tightly coupled by the action of the Tet enzymes. The ROS-induced lesions fU and 8-oxo-G form the third, well-separated cluster. Notably, hmU does not correlate with the ROS lesions.

Recently, it was shown that exposure of naive mESCs to fibroblast growth factor 2 (FGF-2) and activin A (ActA) for 48 h under similar conditions as those described above gives a homogeneous cell population whose transcriptome closely resembles that of the post-implantation epiblast[26]. Under these conditions, mC levels increased more rapidly, approaching somatic levels within a time frame closely reflecting that observed during embryonic development (**Supplementary Fig. 9a**). Levels of hmC rose steadily throughout the 48-h time course. Whereas fC peaked at 36 h, hmU reached its maximum at 24 h (**Supplementary Fig. 9b**). qPCR data showed a robust induction of Dnmt3b that is most likely responsible for the increasing mC levels (**Supplementary Fig. 9a**). A transient peak of Tet1 expression at 24 h went in hand with rising hmU and fC levels, whereas Tet3 expression started rising slowly from the same time point and may drive further accumulation of hmC and fC (**Supplementary Fig. 9b**). Tdg was progressively upregulated together with the onset of methylation activity, whereas Smug1 showed little, if any, change (**Supplementary Fig. 10**). These data supported our view that hmU peaking is not caused by downregulation of the repair pathway.

Thus, the kinetic data showed a complex interplay between methylation and two oxidation reactions during differentiation, which depends on the exact conditions. Most important is the fact that hmU showed a time-dependent occurrence similar to the other oxidized bases hmC, fC and caC independently of the conditions investigated, confirming its formation during epigenetic reprogramming.

hmU attracts specific readers

To obtain initial insight into a potential biological function of hmU in comparison to hmC, we screened for specific readers associating with hmU:A as well as with hmC:G and hmU:G using protein pulldown and relative quantification by LC/MS/MS. In previous stable isotope labeling by amino acids in cell culture (SILAC)-based proteomics studies with hmC, fC and caC containing oligonucleotides, we and others observed a high number of specific protein readers, arguing that the new bases influence a variety of different processes[28,29]. For this hmU study, we further developed this approach for the detection of proteins that directly interact with the modified bases hmU (as well as hmC for comparison) to get a more direct insight into their function (**Fig. 6a**). For the study, we not only included into the biotinylated DNA duplexes (24mers) an hmU (hmC) base but also equipped the counter strand with a polyethylene glycol–based linker carrying a reactive *N*-hydroxysuccinimide (NHS) ester moiety and a reductively cleavable disulfide bond in the middle. This reactive linker cannot be inserted into the DNA strand using solid phase synthesis. In addition, the NHS ester does not survive hybridization conditions. We therefore attached the linker as its azide derivative, using Cu(I)-catalyzed click chemistry, to an alkyne-bearing base present in the DNA duplex[30,31]. These DNA duplexes were subsequently incubated with nuclear extracts from mESCs. The DNA-bound protein complexes were isolated using streptavidin-coated magnetic beads (**Fig. 6a**)[32]. DNA duplexes with the canonical base pairs A:T and C:G at the respective positions served as reference strands. The NHS linker has two functions in the experiment: First, it will covalently trap the reader proteins specifically at the ε-amino groups of lysines, which allows the identification of transiently binding proteins as well. Second, because the linker is cleavable, it will tag the trapped lysine residues, leaving a defined label on those proteins that bind in close proximity to the modified bases hmC and hmU. This allowed us to distinguish protein readers that bound close to hmC and hmU from proteins that are secondary members of the complexes. The covalently trapped proteins were next tryptically digested and labeled with tandem mass tagging (TMT) isobaric tags to allow protein

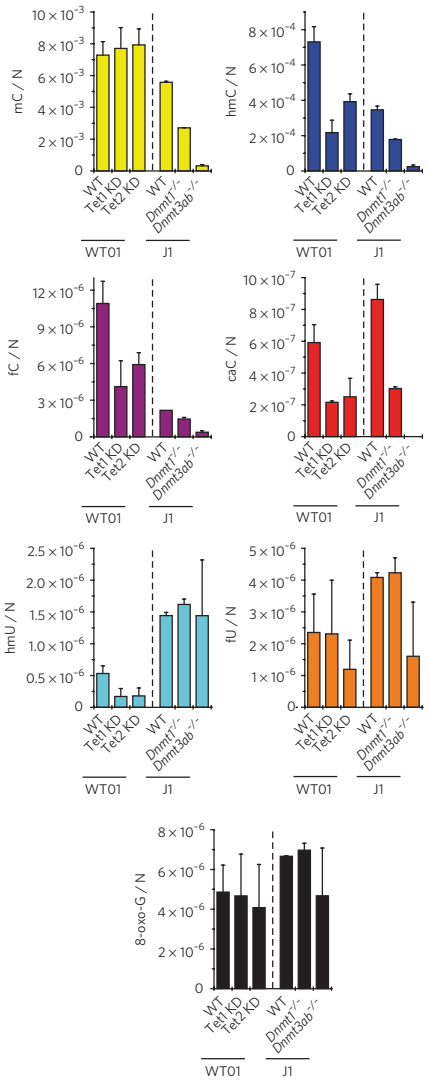


Figure 4 | Tet1 and Tet2 generate hmU in mESCs. Effect of Tet and Dnmt depletion on DNA modification levels in mESCs. J1 cell lines bearing homozygous *Dnmt*-null mutations and WT01 mESCs expressing Tet1 or Tet2 shRNAs (knockdown (KD)) were analyzed. Depicted are mean values per nucleoside (N) ± s.d. of biological replicates as follows: wild type (WT01, $n = 7$), Tet1 KD (WT01, $n = 3$) and Tet2 KD (WT01, $n = 3$); WT (J1, $n = 2$), *Dnmt1*$^{-/-}$ (J1, $n = 2$) and *Dnmt3a/b*$^{-/-}$ (J1, $n = 6$).

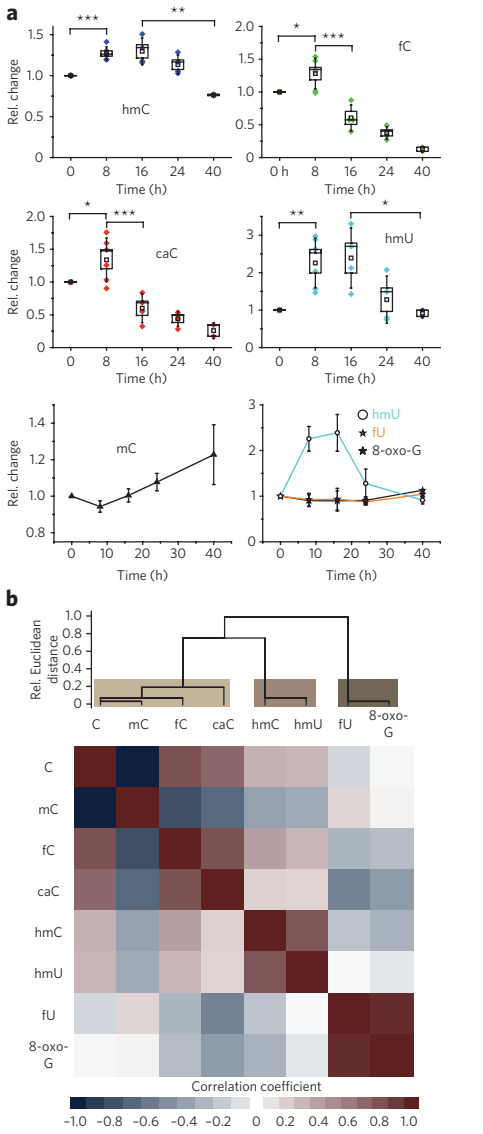


Figure 5 | hmU is produced during mESC differentiation. (**a**) Time course of DNA modification levels during early differentiation (0–40 h) of naive mESCs in the absence of growth factors. Box plot of relative modification changes for hmC, fC, caC and hmU averaged from six independent experiments using two different mESC lines. In the four upper plots, colored diamonds reflect mean values of technical triplicate measurements, open squares reflect mean values of biological replicates, boxes represent the s.e.m., and whiskers represent the the s.d. of the biological replicates; $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.005$ by unpaired two-tailed *t*-test. In the two lower plots, the relative modification changes of mC, hmU, fU and 8-oxo-G are depicted as biological mean values ± s.e.m. (**b**) Unsupervised clustering analysis of Pearson correlation coefficients of 22 data sets obtained from the differentiation experiments depicted in **a**. Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001.

identification and quantification by MS[32,33]. Only proteins that were enriched in both the forward and the reverse TMT experiment were considered to be specific hmU (hmC) readers (**Fig. 6**). All of the readers were subsequently divided into two groups. Proteins that were enriched relative to the control strands are termed specific readers (sRs). Of those, the proteins that were identified with a peptide containing the tag are termed direct-specific readers (dsRs).

The scatter plot (**Fig. 6b** and **Supplementary Fig. 11a**) revealed over 250 sRs recognizing the hmU:A base pair (relative to T:A). Of those proteins, 99 were identified with tagged peptides. These proteins were therefore designated dsRs. A gene ontology analysis showed that 64 proteins of the 250 sRs are nucleotide binders, and 25 proteins are involved in chromatin organization (**Supplementary Fig. 11b**).

Among the dsRs we identified were the regulatory proteins Uhrf2, transcription factor HIVEP3 or the poly [ADP-ribose] polymerase 14 (Parp14), which showed that these proteins interacted directly with hmU. We also detected several H2A and H2B histone proteins among the dsRs and the sRs. Notably, in the group of the sRs, we observed also Dnmt3a and Dnmt3b, which are involved in gene expression regulation. Methyl CpG-binding protein 2 (MeCP2) was identified as well, but it showed only low enrichment. We also detected several transcriptional regulators, such as bromodomain PHD finger

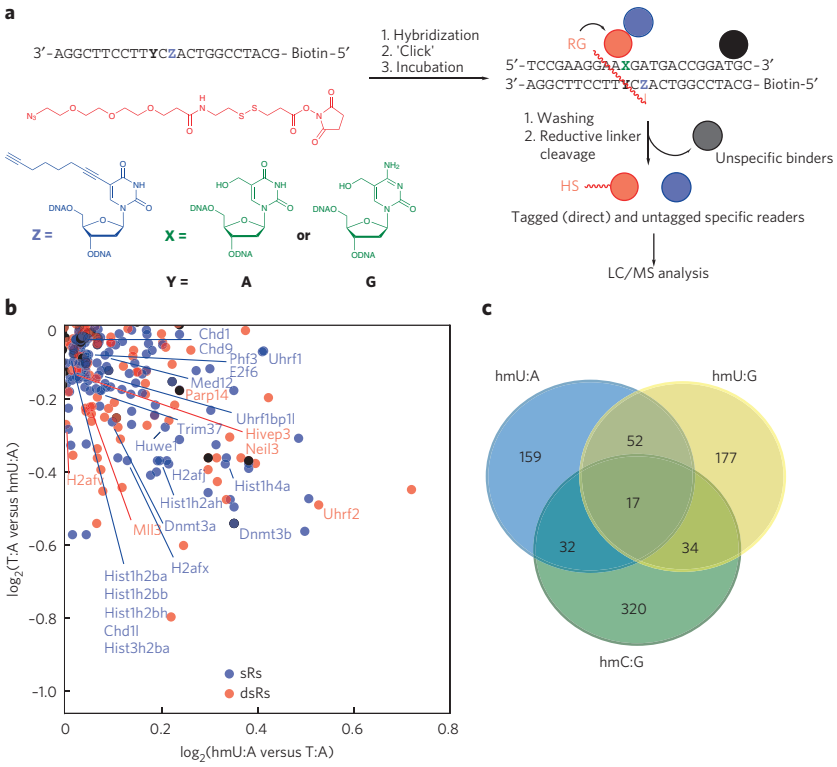


Figure 6 | Identification of hmU:A readers. (**a**) Workflow of pulldown experiments with hmU:A-containing oligonucleotides. DNA oligomers with the modified base Z containing an alkyne group for click reaction and biotin for the pulldown are hybridized with DNA oligomers containing hmU (X = hmU and Y = A). After the click reaction with the DNA-protein cross-linker (red), which has an NHS ester as a lysine-specific reactive group (RG), the dsDNA oligomers are incubated with a nuclear extract from mESCs, and specific binders are covalently linked to the DNA. After three washing steps, the linker is reductively cleaved, and the proteins are enzymatically digested and labeled with TMT2plex reagents. The same dsDNA oligomer without hmU modification (X = T and Y = A) was used as a control sample. Specifically enriched proteins are mixed with the control sample before LC/MS analysis. The same workflow was used for experiments with hmU:G or hmC:G and C:G as a control. (**b**) Magnified view of proteins enriched with hmU:A-containing oligomer. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross-linker and marked in red. Black dots are considered unspecific binders. A full-scale representation is shown in **Supplementary Figure 11**. (**c**) Venn diagram showing overlap of specific binders for hmU:A, hmU:G and hmC:G.

transcription factor (Bptf), transcription factor E2F6 and mediator of RNA polymerase II transcription subunit 12 (Med12). Moreover, we identified hmU-binding E3 ubiquitin-protein ligases such as Uhrf1, Trim37 and Huwe1 and finally the chromodomain helicase DNA-binding proteins 1 and 9 (Chd1 and Chd9) among the hmU readers, which are chromatin remodeling factors and regulate polymerase I and II transcription. To exclude bias by sequence-specific binding, we repeated the experiment with hmU in a different sequence context. Indeed, 78% of the identified protein readers are also observed in this control experiment, and included among them were all of the proteins discussed above, with the notable exception of Uhrf1, which was not observed in this experiment (**Supplementary Data Set 1**). The data show that hmU:A recruits specific proteins that are involved in chromatin biochemistry in a broad sense.

As a first validation of the MS-identified readers, we focused on Uhrf1 and Uhrf2. Uhrf1 is a known mC and hmC binder[34], whereas we recently identified the homolog Uhrf2 as a specific binder of hmC[29]. The MS data for hmU obtained in this study showed an enrichment of Uhrf1 and Uhrf2. In addition, Uhrf2 was identified as a direct binder. To validate the MS findings, we overexpressed Uhrf1 or Uhrf2 together with Tet1cd in HEK-293T[29] and measured the levels of mC, hmC, fC, caC and also hmU. Indeed, we observed increased levels of hmU, proving a functional relation between hmU levels and Uhrf1 and Uhrf2 (**Supplementary Fig. 12**). As Uhrf2 is known to interact with Dnmt3a and Dnmt3b[35] and these two proteins are also identified as sRs of hmU, our initial data provide a consistent picture.

Analysis of the specific hmU readers showed that 49 of the identified proteins were also enriched with an hmC-modified strand. Among these proteins were, for instance, bromodomain PHD finger transcription factor (Bptf), MeCP2 and 19 other nucleotide-binding proteins. The obtained list of specifically binding proteins is in good agreement with our recently published data set[29], which further supports the validity of the covalent trapping method. Finally, we compared the results of the hmU:A pulldown with a pulldown of hmU:G, which would be the product of hmC deamination. We found 69 of 280 sRs in both experiments. We identified transcription regulators as sRs as well in the hmU:G pulldown. However, we also found several DNA repair proteins, such as DNA repair protein complementing XP-C cells homolog (XPC), DNA repair protein Rad50 and five helicases, among others (**Supplementary Data Set 1**).

DISCUSSION

This work aimed to analyze the origin of hmU in mESCs. Although hmU as well as fU are both well-characterized oxidation products of thymidine[9,10], hmU in particular was recently discussed as an intermediate in active demethylation by deamination of hmC[11,12]. We used stable isotopologs of the main oxidation products hmC, fC, caC, hmU, fU and 8-oxo-G in combination with quantitative MS to show that the levels of hmU are strongly elevated in stem cells in comparison to somatic tissue. Notably, we did not observe equally elevated levels of fU and 8-oxo-G, both of which are formed by the reaction of DNA with ROS. This analysis suggests that hmU in stem cells is not only formed by ROS. The quantitative data allowed us to estimate that in WT stem cells, about 70–80% of the found hmU is not formed by ROS but is derived from a different process. Isotope tracing studies with labeled [$^{13}CD_3$]L-Met and [^{13}C,$^{15}N_2$] thymidine showed that the detected hmU originates exclusively from T. Previous studies already proposed a connection between Tet activity and hmU levels, but these studies did not address the origin of hmU, which led to the hypothesis that hmU could form by deamination of hmC[11,22]. To investigate the involvement of deamination[11–14], we studied *Tdg*$^{-/-}$ stem cells reconstituted with a catalytically inactive Tdg. In these cells, the putative deamination–repair complex[12] of Tdg and Aid could form, and we expected to see elevated hmU levels. Indeed, in this experiment, we were able to detect labeled hmU, showing that deamination of hmC does occur, but the levels were low. Most notably, special conditions were required to detect this deamination. In WT cells, all of the detected hmU is clearly derived from T oxidation. As such, the detected hmU is situated in a base pair with A (hmU:A). In conclusion, the majority of hmU in mESCs is produced independently of both ROS and deamination.

Using a combination of functional knockdown in mESCs, ectopic expression of Tet1 catalytic domain in HEK-293T cells and *in vitro* studies with recombinant Tet protein, we showed that the Tet enzymes, known to oxidize mC to hmC, are also responsible for oxidation from T to hmU. A kinetic study in which the temporal changes of all of the oxidized pyrimidines were observed under two different differentiation kinetics showed finally that hmU is formed in processes that also generate hmC and fC. Together with the observed peak levels of hmU of about 14,000 bases per genome, this suggests that hmU could have an epigenetic function similar to hmC. It is conceivable that the Tet enzymes introduce a small amount of hmU during the hmC manufacturing process to trigger faster demethylation by induction of DNA repair. It is known that hmC is not a substrate for repair[8,12], whereas hmU is efficiently recognized and repaired[36], for example, by the repair glycosylase Smug1 (refs. 37,38). Sporadic introduction of hmU could therefore allow recruitment of repair factors, for example, for long patch repair, as recently suggested[39]. Alternatively, hmU might influence transcription factor binding, which is known to be a hallmark of epigenetic reprogramming[40]. To study the proteins that closely interact with hmU, a new pulldown experiment was devised in which we inserted a trapping linker into the oligonucleotide. This linker is able to react covalently with proteins that assemble on the DNA duplex close to the hmU base. Because the linker is reductively cleavable, it leaves a tag on the respective lysine residues, which is detected in subsequent MS-based proteomics measurements. Using this technology, we observed that hmU:A recruits transcription factors and proteins that are involved in chromatin biochemistry, including Uhrf1 and Uhrf2. This result is in line with the recent observation that oxidative lesions in DNA can change the binding affinity of the transcription factor CREB when they are present in the respective cognate sequence[41]. A further analysis in HEK-293T revealed that the hmU readers Uhrf1 and Uhrf2 are able to modulate the levels of all of the oxidized mC bases and of hmU. It is interesting to note that the observed level increases of hmC were small, whereas they were substantial for hmU and also for the further oxidized bases fC and caC. As Uhrf1 and Uhrf2 are proteins known to be involved in numerous epigenetic processes, our data supported the hypothesis that Tet-induced oxidation of T to hmU may have an epigenetic function. Even in the case that hmU is formed just as a side product of Tet-induced hmC formation, it is now clear that the oxidation chemistry performed by the Tet enzymes has to go in hand with effective DNA repair.

Received 15 December 2013; accepted 17 April 2014; published online 18 May 2014

METHODS

Methods and any associated references are available in the online version of the paper.

References

1. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
2. Franchini, D.M., Schmitz, K.M. & Petersen-Mahrt, S.K. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu. Rev. Genet.* **46**, 419–441 (2012).
3. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
4. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
5. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
6. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
7. Tan, L. & Shi, Y.G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895–1902 (2012).
8. Maiti, A. & Drohat, A.C. 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 (2011).
9. Bjelland, S. *et al.* Oxidation of thymine to 5-formyluracil in DNA: mechanisms of formation, structural implications, and base excision by human cell free extracts. *Biochemistry* **34**, 14758–14764 (1995).
10. Mouret, J.F., Polverelli, M., Sarrazini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chem. Biol. Interact.* **77**, 187–201 (1991).
11. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
12. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
13. Nabel, C.S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
14. Rangam, G., Schmitz, K.M., Cobb, A.J. & Petersen-Mahrt, S.K. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS ONE* **7**, e43279 (2012).
15. Jacobs, A.L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
16. Zhang, Q.M. *et al.* DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA Repair (Amst.)* **4**, 71–79 (2005).
17. Burrows, C.J. Surviving an oxygen atmosphere: DNA damage and repair. *ACS Symp. Ser. Am. Chem. Soc.* **2009**, 147–156 (2009).
18. Taghizadeh, K. *et al.* Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat. Protoc.* **3**, 1287–1298 (2008).
19. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
20. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
21. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
22. Liu, S. *et al.* Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res.* **41**, 6421–6429 (2013).
23. Ficz, G. *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
24. Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
25. Leitch, H.G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
26. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
27. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186 (2003).
28. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
29. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
30. Burley, G.A. *et al.* Directed DNA metallization. *J. Am. Chem. Soc.* **128**, 1398–1399 (2006).
31. Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
32. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003); erratum **75**, 4942 (2003); erratum **78**, 4235 (2006).
33. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
34. Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
35. Pichler, G. *et al.* Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* **112**, 2585–2593 (2011).
36. Lewis, H.L., Muhleman, D.R. & Ward, J.F. Serologic assay of DNA base damage. I. 5-Hydroxymethyldeoxyuridine, a radiation product of thymidine. *Radiat. Res.* **75**, 305–316 (1978).
37. Boorstein, R.J. *et al.* Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J. Biol. Chem.* **276**, 41991–41997 (2001).

38. Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H.E. Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst.)* **6**, 505–516 (2007).
39. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* **6**, 39 (2013).
40. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).
41. Moore, S.P.G., Toomire, K.J. & Strauss, P.R. DNA modifications repaired by base excision repair are epigenetic. *DNA Repair (Amst.)* **12**, 1152–1158 (2013).

Acknowledgments

We thank the Excellence Cluster Center for Integrated Protein Science Munich (CiPS^{M}) and the collaborative research centers SFB749, SFB646 and SFB1032 as well as German Research Foundation (DFG) grant CA275/8-4, the Volkswagen foundation, NGFNplus (01GS0870) and the Netherlands Organization for Scientific Research (NWO-VIDI) for financial support. T.P. and S.S. thank the Fonds der Chemischen Industrie for predoctoral fellowships. C.B. thanks the Boehringer Ingelheim Fonds for a predoctoral fellowship. G.K. thanks the Japan Society for the Promotion of Science (JSPS) for a postdoctoral fellowship for research abroad. We thank M. Moser (Max Planck Institute for Biochemistry) for providing R1- and C57Bl6/129–derived mESCs, G. Höfner and K.T. Wanner for their initial help with MS as well as M. Wirsing, L. Belzner and P. Laube for providing bioinformatic tools for data processing.

Author contributions

T.P. synthesized MS standards, performed the sample preparation and ultra high-performance LC/MS/MS method development, did LC/MS analysis, interpreted data and performed statistical analysis. F.S. and N.R. performed mESC differentiation and isotope tracing experiments. M.W. and C.B. performed the HEK-293T experiments, did LC/MS analysis and interpreted data. C.B. performed qPCR and analyzed the data. S.K.L. and D.E. performed the protein pulldown studies and interpreted data. M.T. performed mESC knockdown and knockout experiments. J.S. and O. Kosmatchev did sample preparation and LC/MS analysis. B.H., S.S. and J.S. prepared MS standards. O. Kotljarova performed *in vitro* assays. B.S. synthesized oligonucleotides for protein capture. G.K. synthesized tandem mass tags, and S.M. provided mouse tissue samples. U.M. and H.L. constructed Tet expression plasmids. C.G.S. and M.V. performed Uhrf1/2 over-expression in HEK-293T cells. P.S. and D.S. provided plasmids and cell lines. M.M. and T.C. conceived and supervised the project, interpreted data and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to M.M. and T.C.

ONLINE METHODS

General materials and methods. Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR or Acros Organics and used without further purification. Acetonitrile of LC/MS grade was purchased from Carl Roth GmbH + Co., KG. Formic acid, p.a. for MS, was purchased from Fluka, and water was purified with a Milli-Q Plus system from Merck Millipore.

The MS standards 5-methyl-2′-deoxycytidine (mC), 5-trideuteromethyl-2′-deoxycytidine ([D_3]mC), 5-hydroxymethyl-2′-deoxycytidine (hmC), 5-dideuterohydroxymethyl-2′-deoxy-(N^1,N^3–$^{15}N_2$)-cytidine ([$^{15}N_2$,D_2]hmC), 5-formyl-2′-deoxycytidine (fC), 5-formyl-2′-deoxy-(N^1,N^3–$^{15}N_2$)-cytidine ([$^{15}N_2$]fC), 5-carboxy-2′-deoxycytidine (caC), 5-carboxy-2′-deoxy-(N^1,N^3–$^{15}N_2$)-cytidine ([$^{15}N_2$]caC), 5-hydroxymethyl-2′-deoxyuridine (hmU), 5-(dideuterohydroxymethyl)-2′-deoxyuridine ([D_2]hmU), 5-formyl-2′-deoxyuridine (fU) and 5-formyl-2′-deoxy-(N^1,N^3–$^{15}N_2$)-uridine ([$^{15}N_2$]fU) were synthesized according to earlier published work[19,21,42]. All of the synthesized compounds were characterized and purity confirmed by ^{1}H-NMR, ^{13}C-NMR and ESI-MS, and some were additionally validated by ^{15}N-NMR. 8-hydroxy-2′-deoxy-($^{15}N_5$)-guanosine ([$^{15}N_5$]8-oxo-G) (99 atom% ^{15}N) was purchased from Cambridge Isotope Laboratories; 8-hydroxy-2′-deoxyguanosine (8-oxo-G) was from BIOLOG; 2′-deoxyguanosine (G) and 2′-deoxycytidine (C) were from ChemGenes. (*methyl*-^{13}C,D_3)-L-Met (99 atom% D and ^{13}C) was purchased from Sigma-Aldrich, and 2′-deoxy-(C^2–^{13}C, N^1,N^3–$^{15}N_2$)-thymidine (99 atom% ^{13}C and ^{15}N) from Hartmann Analytic. Aqueous stock solutions of these compounds were stored at −20 °C and warmed up to RT before usage.

Oligonucleotide synthesis. Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions (DMT off) and acetyl-protected dC. Phosphoramidites (including the 5-hydroxymethyl-dU-CE phosphoramidite and the 5′-biotin phosphoramidite) and polystyrene carriers were obtained from Glen Research. 5-octadinyl-dU (Z) phosphoramidite was synthesized according to literature and characterized by ^{1}H-NMR, ^{13}C-NMR, ^{31}P-NMR and ESI-MS[43]. The crude oligonucleotide was cleaved from the resin and deprotected in 30% (v/v) ammonium hydroxide and 40% (v/v) methylamine (1:1) at 65 °C for 10 min and purified by preparative and analytical HPLC (Waters Breeze and Alliance, respectively). Separation was performed by applying a VP 250/10 Nucleosil 100-7 C18 column (flow: 5 ml/min) from Macherey-Nagel with a gradient of buffer A (0.1 M $NHEt_3OAc$ in water) and buffer B (0.1 M $NHEt_3OAc$ in 80% MeCN). DNA-containing fractions were characterized by MALDI-TOF (Bruker Autoflex II) and analytical HPLC, combined, and desalted by C18-Sep-Pak cartridges (Waters). For analytical HPLC, separation was performed by applying a CC 250/4 Nucleosil 120-3 C18 column from Macherey-Nagel (flow: 0.5 ml/min) with the aforementioned buffer system.

For the protein pulldown studies, an hmU-containing oligonucleotide was hybridized with a DNA strand modified with a 5′-biotin and a 5-octadienyl-dU. The sequences of the DNA strands are summarized in **Supplementary Table 3**. To 10 nmol of the dsDNA, 0.5 µl of a 200 mM solution of the cross-linking azide-PEG_3-S-S-NHS ester ($C_{18}H_{29}N_5O_8S_2$, Jena Bioscience, Jena, Germany) in DMSO was added. In a separate tube, CuBr was dissolved in a TBTA solution (DMSO/*t*BuOH 3:1, 100 mM) resulting in a 1:1 Cu(I):TBTA ratio. This solution was immediately added to the DNA/azide mixture. Furthermore, 45 µl of DMSO/*t*BuOH (3:1) were added, the mixture was shaken at 37 °C for 3 h, and the resulting product was purified by ethanol precipitation.

LC/MS analysis of DNA samples. Quantitative LC/UV-ESI-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 coupled with the stable isotope dilution technique. An improved method, based on earlier published work[29,42,44,45], was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature 50 °C, gas flow 15 l/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 l/min (N_2), 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. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 4** and **5**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 30 °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 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 7.2 min; 5 → 80% MeCN; 7.2 → 10.5 min; 80% MeCN; 10.5 → 11.3 min; 80 → 0% MeCN; 11.3 → 13 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 10 °C. The injection volume was amounted to 29 µl. Calibration curves, method validation and data processing are in **Supplementary Note 2**. A complete compilation of LC/MS quantifications results see **Supplementary Note 3**.

DNA digestion. 5–25 µg of genomic DNA in 25 µl H_2O were digested as follows: An aqueous solution (7.5 µl) of 480 µM $ZnSO_4$, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards (**Supplementary Note 2**) were added, and the mixture was incubated at 37 °C for 3 h. After addition of 7.5 µl of a 520 µM $[Na]_2$-EDTA solution, containing 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h at 37 °C and then stored at −20 °C. Prior to LC/MS/MS analysis, samples with up to 15 µg DNA, for which the quantification of low amounts of caC was aspired, were filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences). In contrast, samples with 15–25 µg DNA (isotope-tracing experiments) were filtered by using an AcroPrep Advance 96 filter plate 10K Omega (Pall Life Sciences).

Genomic DNA isolation. Tissues of female WT mice (C57-BL6/N) were dissected at postnatal day 90 and prepared as earlier described[19,21]. Genomic DNA was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit except for mESC samples differentiated in the presence of growth factors (see below). Extraction was performed following the manufacturer's instructions for genomic DNA isolation from cell culture samples or tissue samples, respectively. All buffers until loading of the sample on Genomic-tip 100/G were additionally supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene. (BHT, 200 µM) and deferoxamine mesylate salt (desferal, 200 µM) as well as the deaminase inhibitor tetrahydrouridine (THU, 200 µM), according to published methods, to reduce background oxidation or deamination[18]. Elution buffer QF was supplemented with 200 µM BHT. Following elution, all steps were performed on ice. DNA was then precipitated with NaOAc (0.3 M final) and 0.7 volumes *i*PrOH. DNA pellets from cultured cells were washed twice with ice-cold 70% EtOH and resuspended in H_2O containing 20 µM BHT using a Qiagen TissueLyser (30 Hz, 2 min). DNA pellets from mouse tissues were resuspended in PBS buffer and additionally extracted with phenol/$CHCl_3$, precipitated, washed and resuspended as described above.

R1 mESC samples differentiated in the presence of growth factors or transfected with Smug1 esiRNAs were lysed directly in the plates with RLT buffer (Qiagen) supplemented with BHT and desferal as described above. DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction, except that elution was repeated four times with 100 µl of elution buffer supplemented with BHT (200 µM). Eluted DNA was precipitated with 2 M ammonium acetate and two volumes of absolute ethanol and finally resuspended in H_2O containing 20 µM BHT. The flow-through from the spin columns was used to isolate RNA (see real-time PCR analysis).

mESC cell culture. Feeder independent WT01 mESCs (C57BL/6 strain)[46] were cultured in the presence of serum and LIF as previously described[4]. *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ mESCs were described previously[47]. *Tdg*$^{-/-}$ mESCs were complemented by random integration of either empty vector (hereafter referred to as *Tdg*$^{-/-}$ mESCs) or a minigene expressing catalytically inactive Tdg (N151A)[47]. Clonal mESC lines with targeted Tdg alleles, R1 cells (strain 129/Sv)[48], J1 cell lines (strain 129S4/SvJae)[49] and a mESC line derived from C57Bl/6/129 mixed background[50] were routinely maintained on gelatinized plates in DMEM (PAA or Sigma) supplemented with 10% FBS, 1× MEM-nonessential amino acids (NEAA), 0.2 mM L-alanyl-L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin (all from PAA), 0.1 mM β-mercaptoethanol, 20 ng/ml (≥1 ×10^3 U/ml) mouse recombinant LIF (ORF Genetics), 1 µM PD 0325901 and 3 µM CHIR 99021 (2i; both from Axon Medchem). In these conditions, the global levels of genomic mC were very low (and, as a consequence, the levels of its oxidized derivatives were even lower; data not shown). Before DNA isolation, 2i cultures were passaged twice (over 5 d) in DMEM supplemented with FBS and LIF as above but lacking 2i. With this strategy, primed mESC cultures were obtained with no sign of overt differentiation and modified genomic cytosines reached reproducibly higher and stable levels. For isotope tracing with heavy

doi:10.1038/nchembio.1532

thymidine in serum-primed mESCs 2i cultures of R1, cells were passaged twice (5 d) in the same serum-containing medium lacking 2i and simultaneously supplemented with 100 μM [^{13}C,$^{15}N_2$]T. For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of R1 cells and mESC lines with targeted Tdg alleles were passaged twice (over 5 d) without 2i in L-Met-free DMEM (Life Technology) supplemented as above and with 0.2 mM of either [*methyl*-^{13}C,D_3]L-Met or natural L-Met. For mESC differentiation without growth factors, R1 cells and the C57Bl/6/129 mixed background cell line were first plated at 1×10^5 cells/cm^2 on gelatin-coated plates in N2B27 medium containing 1,000 U/ml LIF to favor attachment and initial survival[51]. After 12 h, the medium was replaced without addition of LIF (defined as time point 0 h). The medium was replaced once more at 24 h. For isotope tracing with [*methyl*-^{13}C,D_3]L-Met during mESC differentiation, R1 cells were cultured for two passages in L-Met-free N2B27 medium supplemented with LIF, 2i and 0.2 mM of either [*methyl*-^{13}C,D_3]L-Met or natural L-Met. Differentiation of R1 cells in the presence of FGF-2 and ActA was as described[26], with minor modifications. Briefly, mESCs were cultured in N2B27 medium containing 2i and 1,000 U/ml LIF for several passages and then seeded at 2.2×10^5 cells/cm^2 in N2B27 medium containing 1% KnockOut Serum Replacement (Life Technologies), 12 ng/ml FGF-2 (PeproTech) and 20 ng/ml ActA (ORF Genetics) on plates coated with a thin layer of Geltrex extracellular matrix preparation (Life Technologies). The medium was exchanged after 24 h.

Knockdown experiments in mESCs. shRNA expression vectors targeting Tet1 and Tet2 were generated by cloning synthetic oligonucleotides in pLKO.1 (ref. 52). Recombinant lentiviruses were produced by cotransfecting pLKO.1 shRNA expression vectors and packaging plasmids in HEK-293 cells. 48 h after transduction in the presence of 8 μg/ml polybrene, shRNA-expressing mESCs were selected with 4 μg/ml puromycin. Cell pools were continuously cultured in the presence of puromycin. shRNA target sequences were as follows: SCR (control), 5′-CCT AAG GTT AAG TCG CCC TCG-3′ (ref. 52); Tet1, 5′-TGT AGA CCA TCA CTG TTC GAC-3′ (see ref. 52), Tet2: 5′-TTC GGA GGA GAA GGG TCA TAA-3′. esiRNAs for Smug1 knockdown were generated as described[53]. The cDNA template for *in vitro* transcription was generated by PCR using following primers: forward, 5′- CGT AAT ACG ACT CAC TAT AGG GAG CCC GTG GGT G-3′, and reverse, 5′-CGT AAT ACG ACT CAC TAT AGG GGT TTC GTC CAC TGG G-3′. R1 mESCs were weaned from 2i for two passages in FBS- and LIF-containing medium as described above. Upon plating the second passage, the cells were transfected in a p60 plate with 6 μg of Smug1 esiRNAs (34.5 nM) and 20 μl of Lipofectamine RNAi MAX (Life Technologies) according to the manufacturer's instructions and were lysed 72 h after transfection.

Culture and transfection procedures for HEK-293T cells. All transfections were performed using jetPRIME transfection reagent (PEQLAB Biotechnologie GmbH) according to the manufacturer's instructions. HEK-293T cells were seeded 24 h before transfection at a density of 2.5×10^6 cells per 75 cm^2 flask and incubated in 10 ml of medium. The transfection solution (500 μl of jetPRIME buffer, a specific amount of plasmid DNA (**Supplementary Note 4**) and 20 μl of jetPRIME reagent) was added to the medium, and the cells were incubated for 48 h, with an additional medium exchange 24 h after transfection. When cotransfection of esiRNA was performed, a second transfection step (500 μl of jetPRIME buffer, 5 μg of esiRNA and 20 μl of jetPRIME reagent) was carried out 4 h after transfection of plasmid DNA. esiRNAs were purchased from Sigma (human TDG esiRNA EHU038971; human SMUG1 esiRNA EHU098861; human CDK5RAP1 esiRNAEHU079221). **Supplementary Note 4** summarizes the overexpression and knockdown procedures.

Isotope tracing with [^{13}C,$^{15}N_2$]thymidine or [^{13}C,D_3]L-Met in HEK-293T cells transfected with Tet1cd. 24 h before transfection, 2.5×10^6 cells were seeded in a 75-cm^2 flask containing 10 ml either of (for [^{13}C,$^{15}N_2$]thymidine) DMEM medium supplemented with 50 μM [^{13}C,$^{15}N_2$]T or (for [^{13}C,D_3]L-Met) DMEM medium lacking L-Met, L-cystine and pyruvate, which was supplemented with 10% dialyzed FBS, 2 mM [^{13}C,D_3]L-Met and 0.2 mM L-cystine. Transfection was performed as described above using labeled medium.

Real-time PCR analysis of mRNA expression. For analysis of Tet1 and Tet2 knockdown in mESC total RNA was prepared with Trizol (Invitrogen), cDNA synthesis was performed with Quantitect reverse transcription kit from Qiagen, and real-time PCR was performed with the Power Sybr Green PCR master mix from Applied Biosystems on an Applied Biosystems 7500 Fast system. Knockdown efficiencies relative to control samples transfected with SCR esiRNAs were 79% and 70% for Tet1 and Tet2, respectively. The primers used to estimate them are listed in **Supplementary Note 4**. For analysis of Smug1 knockdown in mESCs and EpiLC differentiation samples, total RNA was prepared with RNeasy spin columns (Qiagen), followed by DNase treatment using TURBO DNA-free (Ambion, Life Technologies); cDNA synthesis was carried out using iScript cDNA Synthesis kit (Bio-Rad); real-time PCR was performed with SsoFast EvaGreen Supermix (Bio-Rad). Smug1 knockdown efficiency relative to control samples transfected with esiRNAs targeting GFP was estimated to be 60%. Quantification of Tet, Dnmt, Tdg and Smug1 transcripts during EpiLC differentiation and Smug1 knockdown samples was performed using the primers listed in **Supplementary Note 4**. Expression levels were quantified with respect to the housekeeping gene *Gapdh* and normalized to time point 0 h.

Tet *in vitro* assay. A plasmid was prepared from *dam*$^-$/*dcm*$^-$ competent *E. coli* strain (New England BioLabs) and methylated with M.SssI (New England BioLabs). 1.5 μg of plasmid DNA were then treated with recombinant Tet1 from the 5hmC TAB-Seq Kit (Wisegene) corresponding to ref. 54. After 3 h incubation at 37 °C and proteinase K treatment, the oxidized plasmids were purified with GeneJET PCR Purification Kit from Thermo Scientific and eluted in 25 μl water. Samples were then subjected to LC/MS/MS analysis as described[42]. The results are compiled in **Supplementary Note 3**.

Correlation and cluster analyses. Statistical data analysis was performed using IBM SPSS Statistics 19. Results of bivariate correlation analyses are summarized in **Supplementary Note 5**. Unsupervised clustering of species with respect to its correlation coefficients was applied by average linkage hierarchical clustering using a squared Euclidean distance measure.

Pulldown assay. For the pulldown assay, 250 μg (50 μl) of the crude nuclear protein extracts were filled up to 500 μl with 50 mM TEAB and 1 mM $MgCl_2$. The binding conditions were 45 mM TEAB, 1.1 mM $MgCl_2$, 2 mM HEPES, 42 mM NaCl and 20 μM EDTA. Complete Protease Inhibitor Cocktail Tablets were used from Roche Diagnostics (Indianapolis, IN, USA). The DNA oligomers with DNA-protein cross-linker were dissolved in neat DMSO, and 1 nmol was added to the protein lysate and incubated for 20 min at room temperature. Streptavidin-coated magnetic particles (Roche Diagnostics, Indianapolis, IN, USA) were washed three times with binding buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 7.4) before 200 μl of the bead slurry (equal to 2 mg beads) were added to the sample. Following 2 h incubation at room temperature under constant rotating, the beads were washed three times with 50 mM TEAB and 1 mM $MgCl_2$. The beads were reconstituted in 50 mM TEAB and 1 mM $MgCl_2$. Disulfide bonds of the cross linker were cleaved and alkylated in the process of enzymatic digestion, and the magnetic particles were removed before adding trypsin (described below).

Protein sample preparation. Cell lysis of mouse embryonic stem cells was performed as described in ref. 29. For each lysis, approximately 7.5×10^7 cells were used. Protein concentration was determined by Bradford assay. For each experiment 250 μg (50 μl) of the crude nuclear protein extract were used. Protein samples for MS analysis were reduced by adding 100 mM TCEP and by incubating on a shaker at 650 r.p.m. for 1 h at 60 °C and subsequently alkylated with 200 mM iodoacetamide in the dark for 30 min at 25 °C. Protein samples were digested with 0.5 μg trypsin (Promega, Madison, MA, USA) for 16 h at 37 °C. The reaction was stopped using 1 mM phenylmethylsulphonylfluoride. After tryptic digestion, peptide labeling with the TMT2plex reagents (Thermo Fisher Scientific, Waltham, MA, USA) was performed according to the manufacturer's instructions. TMT2plex reagents 126 and 127 were used to label the samples. When the sample (proteins enriched with hmU- or hmC-containing DNA strands) was labeled with TMT126, the control sample (proteins enriched with no modified DNA-bases) was labeled with the TMT127 reagent and vice versa. Subsequent to the labeling, both sample and control, were combined. This way, each experiment was performed twice as a so-called label swap experiment. Organic solvent was removed by vacuum centrifugation, and the sample was finally reconstituted in 1% (v/v) formic acid for MS analysis.

LC/MS analysis of protein samples. The samples were analyzed using an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific, Waltham, MA, USA) coupled to an LTQ-Orbitrap XL (Thermo Fisher

Scientific, Waltham, MA, USA). Of each eluate, 15 µl were injected for the analysis. The samples were desalted and concentrated on a µ-precolumn cartridge (PepMap100, C18, 5 µM, 100 Å, size 300 µm i.d. x 5 mm) and further processed on a custom-made analytical column (ReproSil-Pur, C18, 3 µM, 120 Å, packed into a 75 µm i.d. x 150 mm and 8 µm picotip emitter). A 57-min multistep analytical separation was performed at a flow rate of 300 nl/min. In the first 50 min, a linear gradient was ramped up from 5% (v/v) solvent B (acetonitrile containing 0.1% formic acid and 5% DMSO) and 95% solvent A (water containing 0.1% formic acid and 5% DMSO) to 95% solvent B. This level was held for 5 min and then ramped down again to 5% solvent B within 2 min. Mass spectrometric analyses were performed starting with a full mass scan in the mass range between *m*/*z* 300 and *m*/*z* 1,650. This survey scan was followed by three MS/MS scans using the FTMS mass analyzer and high normalized collision energy of 70 in the HCD cell and three additional scans using the ion trap mass analyzer and a normalized collision energy of 35.

Protein identification and relative quantification method. The Thermo Proteome Discoverer 1.1 software (Thermo Fisher Scientific, Waltham, MA, USA) was used for protein identification and for relative quantification. The Sequest (Thermo Fisher Scientific, Waltham, MA, USA) search engine was used in combination with a Uniprot database (*Mus musculus*; date of download, 04/2013). As a limit of detection, a ratio of threefold signal over the noise filter was applied. A maximum of two missed cleavage sites was allowed. The mass tolerances were 10 p.p.m. for the precursor mass and 0.5 Da for the fragment ion mass. Carbamidocysteine was set as static modification. Dynamic modifications were: cation, Na (D, E); the residue of the DNA-protein crosslinker (+146.028 Da; K, Y); Oxidation (M) as well as TMT2plex (N-term. and K). Identified, nonredundant peptides, which were labeled with the TMT2 reagent, were used for relative quantification. The integration window tolerance was 20 p.p.m., and the integration method was set to 'most confident centroid'. The signals of the TMT2 reporter ions 126 and 127 were used to calculate ratios and monitor either preferred or nonpreferred binding of the identified proteins to the modified DNA bases in comparison to the control strand. From the identified proteins, the only proteins considered as 'specific readers' were enriched in both the forward and the reverse experiment.

42. Schiesser, S. *et al.* Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
43. Gierlich, J. *et al.* Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639–3642 (2006).
44. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2′-deoxycytidine, 2′-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
45. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
46. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
47. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
48. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**, 8424–8428 (1993).
49. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
50. Montanez, E. *et al.* Kindlin-2 controls bidirectional signaling of integrins. *Genes Dev.* **22**, 1325–1330 (2008).
51. Ying, Q.L. & Smith, A.G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
52. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
53. Kittler, R., Heninger, A.K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).
54. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

doi:10.1038/nchembio.1532

Supplementary Information

Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

Toni Pfaffeneder[1#], Fabio Spada[1#], Mirko Wagner[1#], Caterina Brandmayr[1], Silvia Laube[1], David Eisen[1], Matthias Truss[2], Jessica Steinbacher[1], Benjamin Hackner[1], Olga Kotljarova[1], David Schuermann[5], Stylianos Michalakis[3], Olesea Kosmatchev[1], Stefan Schiesser[1], Barbara Steigenberger[1], Nada Raddaoui[1], Gengo Kashiwazaki[1], Udo Müller[4], Cornelia G. Spruijt[6], Michiel Vermeulen[6≠], Heinrich Leonhardt[4], Primo Schär[5], Markus Müller[1]* and Thomas Carell[1]*

[1] Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany.

[2] Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Ziegelstr. 5-9, 10098 Berlin, Germany.

[3] Center for Integrated Protein Science at the Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany.

[4] Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany.

[5] Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland.

[6] Department of Molecular Cancer Research, Cancer Genomics Netherlands, UMC Utrecht, 3584 CG Utrecht, Netherlands.

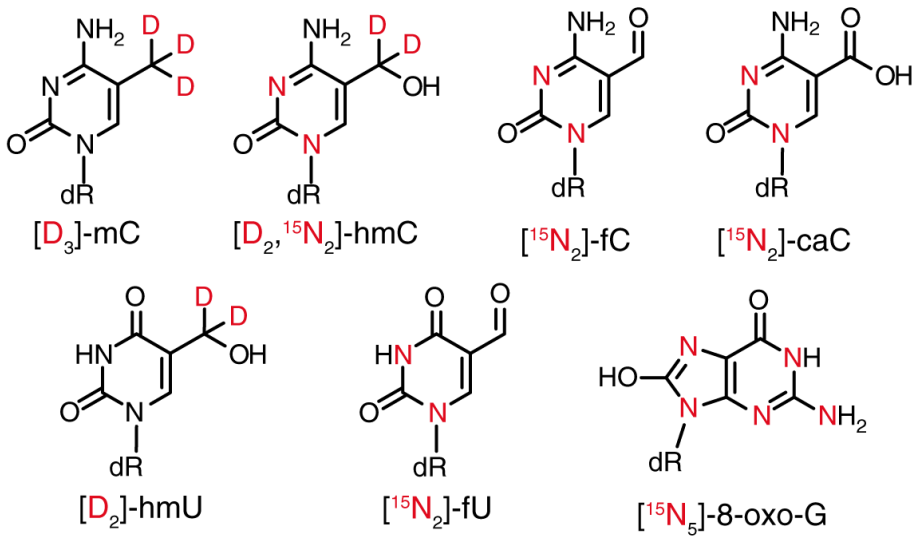
[≠] Present Address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA, Nijmegen, The Netherlands

* Correspondence to: markus.mueller@cup.uni-muenchen.de and thomas.carell@cup.uni-muenchen.de

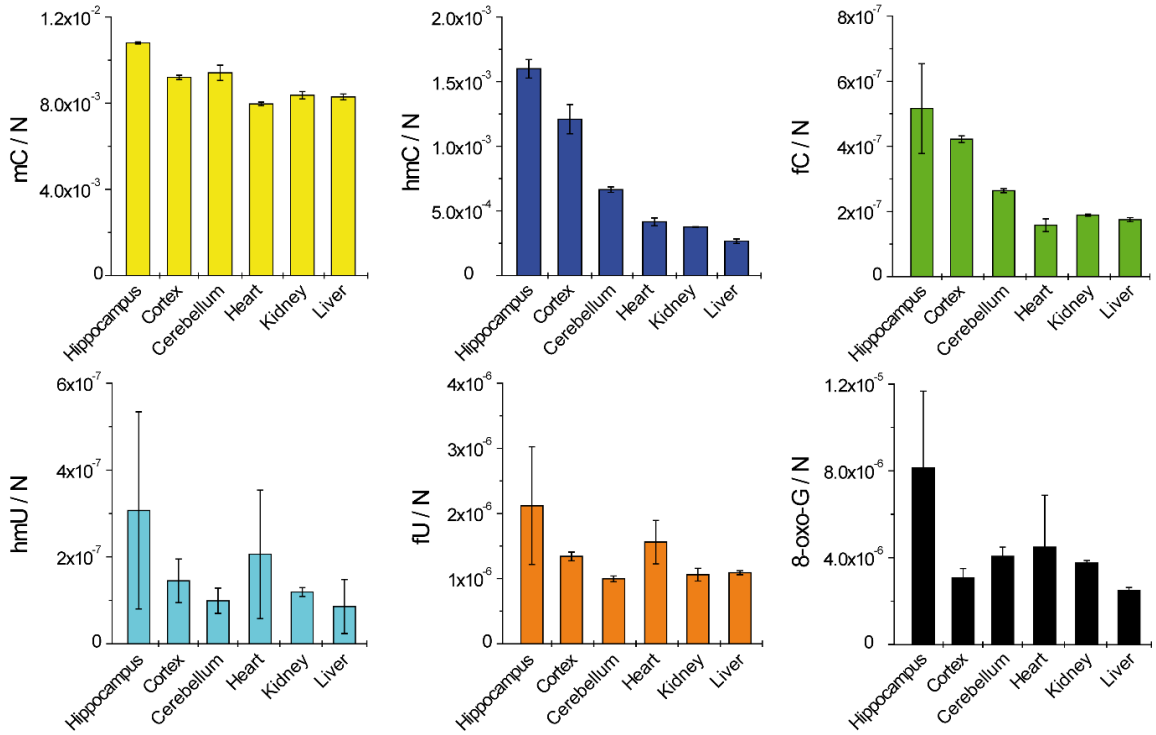
[#] These authors contributed equally

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Results

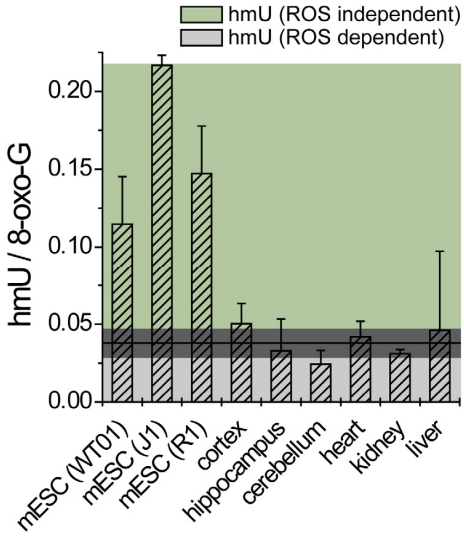


Supplementary Figure 1. Isotopically labeled nucleosides used as internal standards for quantitative LC-MS/MS analysis (dR = -2´-deoxyribose).



Supplementary Figure 2. DNA modification levels per nucleoside (N) of different murine tissues from 3 month old individuals (n = 3). Depicted are biological mean values ±SD.

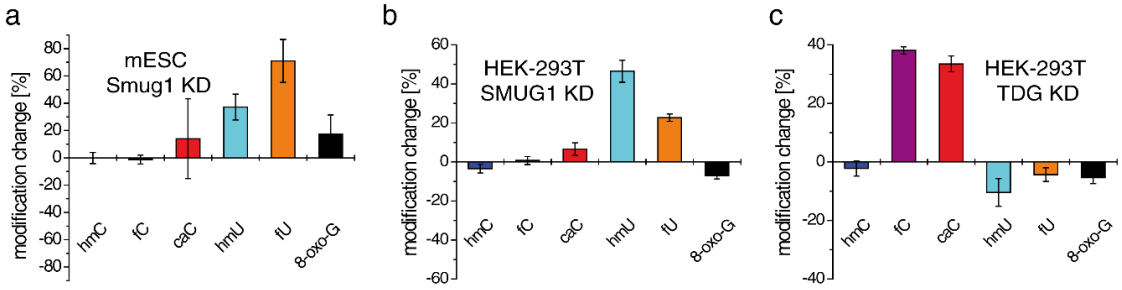
Nature Chemical Biology: doi:10.1038/nchembio.1532



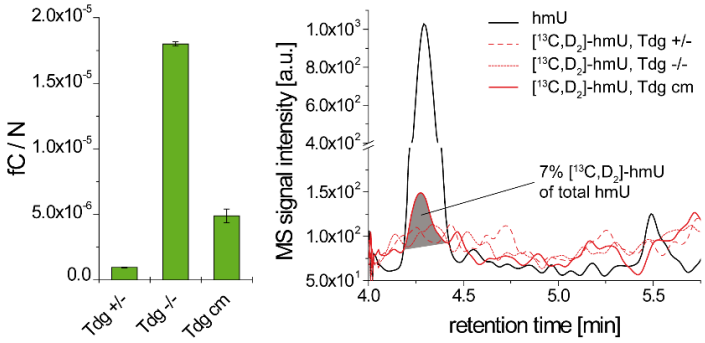
Supplementary Figure 3. hmU-levels normalized to the oxidative background marker 8-oxo-G in mESCs (n_{WT01} = 7; n_{J1} = 2, n_{R1} = 2) and murine tissues (n = 3, 3 month old individuals) in order to dissect ROS dependent and ROS independent processes. Normalization was necessary to take deviating background oxidation of DNA sample preparation into account. The light grey area reflect the hmU-level fractions, which are generated by ROS dependent processes. The green area reflect the hmU-level fractions, which are generated by ROS independent processes. The assignment is based on the assumption, that hmU-levels in somatic tissue are exclusively ROS created lesions (derived from the cluster analysis in **Fig. 2b**). The dark grey area reflect the mean value ±SD of hmU/8-oxo-G ratios of the tissue data. The difference between the height of the hmU/8-oxo-G ratios of mESCs and the mean of the tissue data give the hmU-fraction which is formed by ROS-independent processes. In WT01, J1 and R1 cells about 67%, 83% and 74%, respectively, of the global hmU-levels are estimated to be created by ROS independent processes. Bars reflect biological mean values ±SD. The differences between mESCs (WT01, J1) and murine tissues are significant ($P = 9.3\text{x}10^{-5}$–$5.0\text{x}10^{-3}$; unpaired two-tailed t-test) except for mESCs (WT01, J1) and liver (P = 0.081, 0.15).

Supplementary Figure 4. Schematic representation of isotope tracing experiments with [^{13}C,^{15}N$_2$]-T (left; blue) and [^{13}C,D$_3$]-methionine (right; red) and exchange rates of derived genomic isotopologues. Small negligible deviations in the exchange rates are due to differential noise sources. LOD = limit of detection. wt = wild type mESCs; Tdg cm = Tdg / mESCs complemented with catalytic inactive Tdg (see Supplementary Fig. 6).

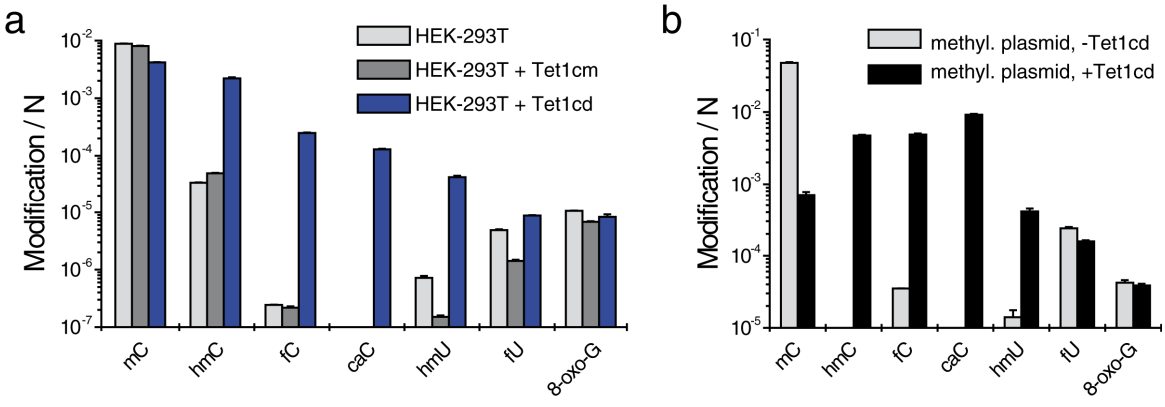
Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 5. (**a**) Effect of Smug1 depletion on modification levels in mESCs (R1). Effect of SMUG1 (**b**) and TDG (**c**) depletion on modification levels in HEK-293T cells overexpressing Tet1cd. Shown is the percent change in modification content per nucleoside in cells co-transfected with esiRNAs targeting TDG or SMUG1 relative to co-transfection with control esiRNA. Depicted are technical mean values ± SD.

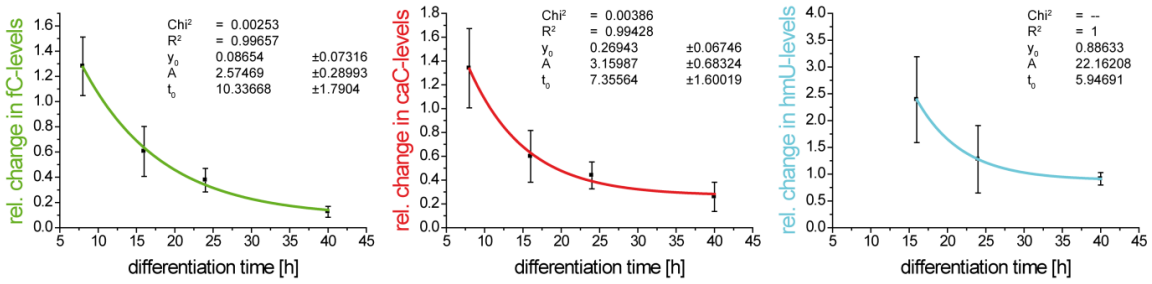


Supplementary Figure 6. Isotope tracing experiments with *Tdg*$^{+/-}$, *Tdg*$^{-/-}$ mESCs as well as *Tdg*$^{-/-}$ mESCs complemented with a catalytic mutant of Tdg (Tdg cm) grown in the presence of [^{13}C,D$_3$]-methionine (200 µM). The catalytic mutant of Tdg is not completely inactive (fC levels are between Tdg +/- and Tdg -/- cells, left). Only in case of *Tdg-/-* cells complemented with a catalytic inactive Tdg (Tdg cm) labeled hmU was detected, which originated from the deamination of labeled hmC. ~7% [^{13}C,D$_2$]-hmU over total hmU was observed. This corresponds to ~0.06% deamination of hmC to hmU under these conditions (2.5x10^{-4} total hmC / N; 2.2x10^{-6} total hmU / N). Labeled fU was not observed. Depicted fC-levels represent technical mean values ±SD.

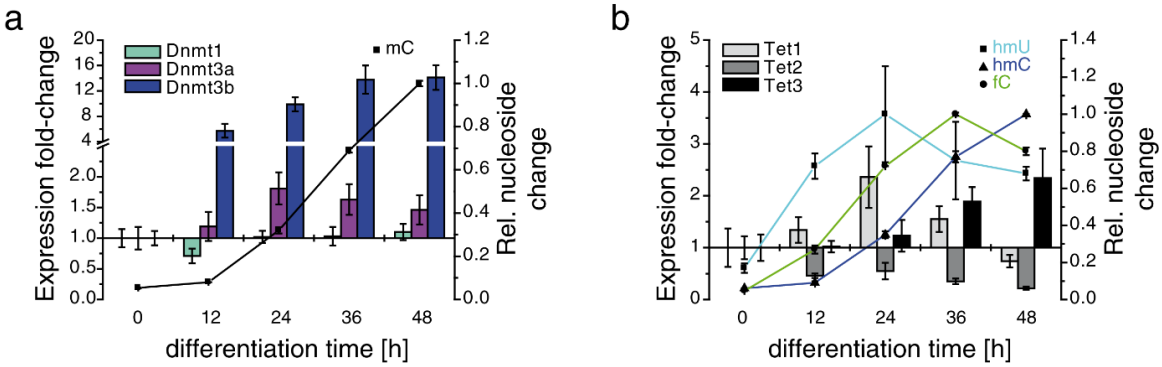


Supplementary Figure 7. Tet1 and Tet2 generate hmU in HEK-293T and in vitro (a) Effect of Tet1 overexpression on modified pyrimidines in HEK-293T cells. Modification levels in cells overexpressing wt and catalytic mutant versions of Tet1 catalytic domain (Tet1cd, blue bars and Tet1cm, gray bars, respectively), or a control construct (white bars). Depicted are mean values ±SD of technical triplicates on a logarithmic scale. **(b)** Pyrimidine modification levels in methylated plasmid DNA after treatment in vitro with Tet1cd. Depicted are mean values ±SD of technical duplicates. Note the logarithmic scale.

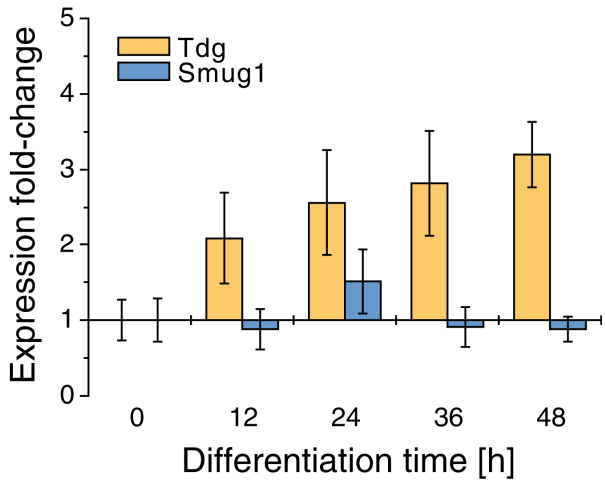
Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 8. Exponential models for fitting the decay curve of fC, caC and hmU in combined data sets from differentiation of R1 and C57Bl6/129-derived mESCs (6 biological independent experiments). In a simplified approach a single exponential decay model ($y = y_0 + A*\exp(-x/t_0)$) was plotted using ORIGIN®. The parameters y_0 (offset), t_0 (time constant) and A (amplitude) of each decay function were iteratively optimized until the minimum of the Chi^2 value of the fitting was reached. Half-life times ($t_{1/2} = t_0*\ln 2$) for fC, caC and hmU were 7.2±1.2, 5.1±1.1 and 4.1 h, respectively.

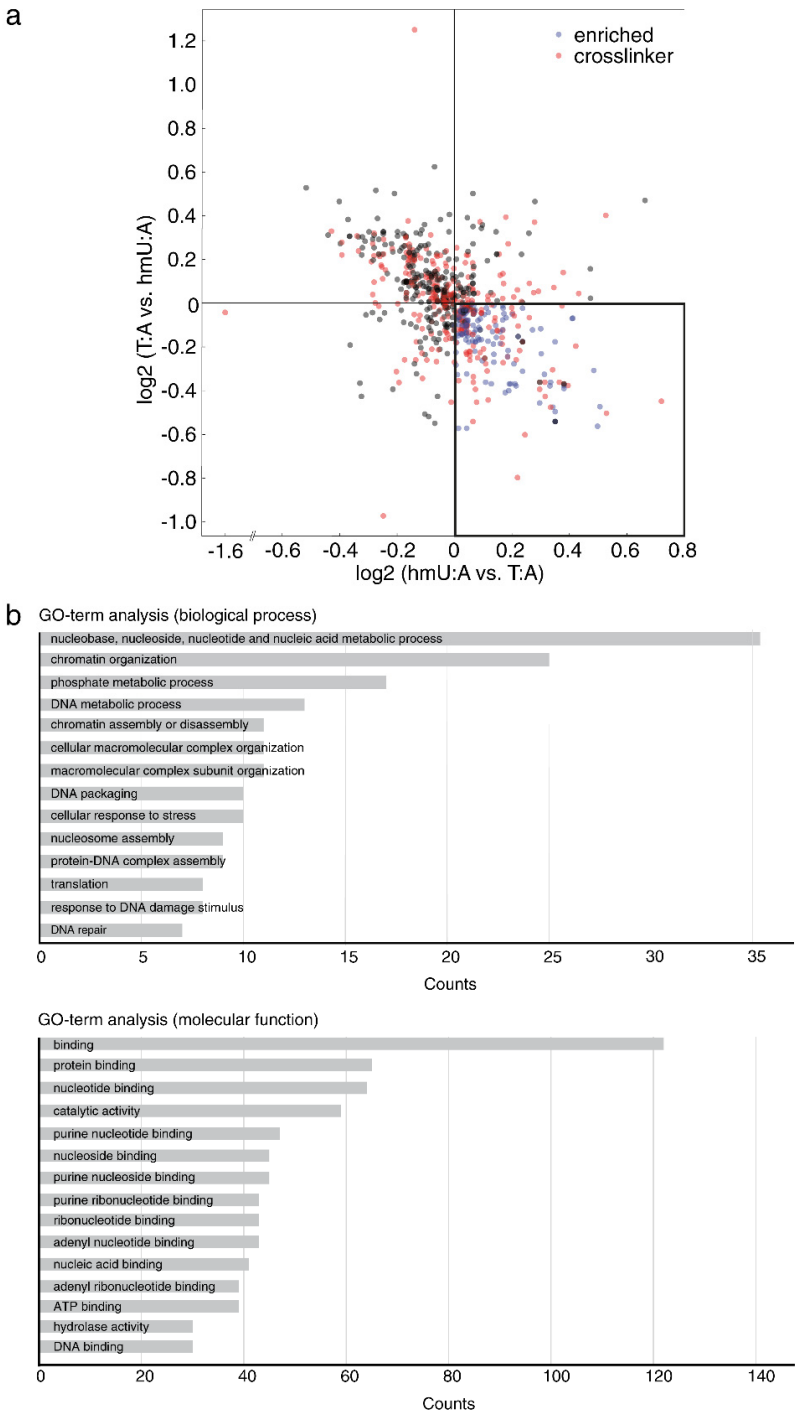


Supplementary Figure 9. Normalized transcript levels of Dnmts (c), Tet1–3 (d) and normalized modification levels of mC (c), hmC, fC and hmU (d) during differentiation of naïve mESCs in the presence of FGF-2 and ActA.



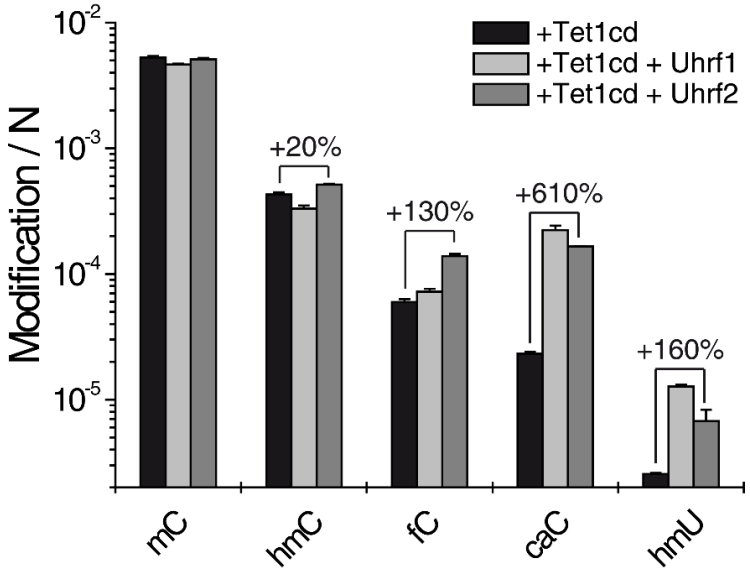
Supplementary Figure 10. Expression level analysis of Tdg and Smug1 during differentiation of mESCs in the presence of FGF-2 and ActA. Expression levels were quantified with respect to the housekeeping gene Gapdh and normalized to time point 0 h. Depicted are technical mean values ± SD.

Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 11. (a) Scatterplot of proteins enriched with the hmU:A containing oligomer. Ratios of a forward and a reverse experiment are plotted. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross linker and marked in red. Gray dots are considered unspecific binders. See **Fig. 6** for detailed view. **(b)** Gene Ontology Analysis performed with DAVID Bioinformatics Resources 6.7[4]

Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 12. Effect on modified pyrimidines in HEK-293T cells upon Tet1cd and Uhrf1 (light gray bars) or Uhrf2 (gray bars) co-overexpression. Depicted are mean values ±SD of technical triplicates on a logarithmic scale.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 1. Isotope tracing experiments by supplementing the growth medium of mES cells (LIF), differentiating mESCs (R1, without growth factors) and HEK-293T cells with either [^{13}C,$^{15}N_2$]-T (50 or 100 μM) or [^{13}C,D_3]-methionine (0.2 mM). Small deviations in the exchange yields are due to differential noise sources and are negligible. LOD = Limit of detection. In case of [^{13}C,D_1]-fU no difference was observed compared to the natural control.

cell type	growth medium	[^{13}C,$^{15}N_2$]-T / T [%]	[^{13}C,$^{15}N_2$]-hmU / hmU [%]	[^{13}C,$^{15}N_2$]-fU / fU [%]
mESC (2i)	100 μM natural T	0.1	< LOD	< LOD
mESC (2i)	100 μM [^{13}C,$^{15}N_2$]-T	76.0	78.2	74.6
HEK + Tet1cd (72h)	50 μM [^{13}C,$^{15}N_2$]-T	73.8	74.2	71.0

cell type	growth medium	[^{13}C,D_3]-mC / mC [%]	[^{13}C,D_2]-hmC / hmC [%]	[^{13}C,D_2]-hmU / hmU [%]	[^{13}C,D_1]-fU / fU [%]
mESC (LIF)	natural methionine	0.1	< LOD	< LOD	3.3
mESC (LIF; 0 h)	[^{13}C,D_3]-methionine	88.9	87.6	< LOD	3.0
diff. mESC (12 h)	[^{13}C,D_3]-methionine	89.3	88.4	< LOD	3.1
diff. mESC (24 h)	[^{13}C,D_3]-methionine	90.1	89.3	< LOD	3.2
diff. mESC (48 h)	[^{13}C,D_3]-methionine	90.4	90.5	< LOD	3.5
mESC *Tdg*$^{+/-}$	[^{13}C,D_3]-methionine	88.0	87.6	< LOD	3.7
mESC *Tdg*$^{-/-}$	[^{13}C,D_3]-methionine	87.4	87.2	< LOD	< LOD
mESC *Tdg*$^{-/-}$ + Tdg cm	[^{13}C,D_3]-methionine	86.9	86.7	7.4	2.2
HEK + Tet1cd (72h)	[^{13}C,D_3]-methionine	87.4	83.4	< LOD	< LOD

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 2. Assessment of ROS dependent hmU and fU product ratio of T-oxidation in HEK-293T wild type cells where TET activity is lowest (related to Fig. 3c). Modified nucleosides / N are given as mean values plus SD of three independent technical replicates. When T is oxidized by ROS about 9.8% hmU and 90.2% fU is generated.

n	mC / N		hmC / N		fC / N		caC / N
	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean
1	6.14E-03	1.97E-04	2.92E-05	9.90E-08	2.56E-07	2.59E-09	n.d.
2	6.21E-03	1.32E-05	2.95E-05	2.21E-09	3.10E-07	2.06E-08	n.d.
3	5.76E-03	4.36E-05	3.32E-05	6.55E-07	2.86E-07	2.95E-09	n.d.
4	9.01E-03	2.87E-04	5.23E-05	5.01E-07	3.50E-07	5.00E-09	n.d.
5	8.80E-03	9.62E-05	3.39E-05	5.44E-07	2.43E-07	2.41E-09	n.d.
6	8.55E-03	7.69E-05	3.70E-05	2.17E-07	2.14E-07	1.30E-08	n.d.
biol. mean	7.41E-03		3.59E-05		2.76E-07		
biol. SD	1.52E-03		8.56E-06		4.90E-08		

n	hmU / N		fU / N		hmU/ (hmU+fU)	fU/ (hmU+fU)	8-oxo-G / N	
	techn. mean	SD	techn. mean	SD	[%]	[%]	techn. mean	SD
1	2.66E-07	5.54E-08	5.15E-06	1.59E-07	4.9	95.1	8.49E-06	1.30E-07
2	1.21E-06	1.62E-07	8.02E-06	6.66E-07	13.1	86.9	1.00E-05	1.98E-07
3	3.65E-08	7.35E-09	9.76E-07	8.75E-09	3.6	96.4	3.02E-06	7.02E-08
4	6.31E-07	9.10E-10	3.51E-06	3.28E-08	15.3	84.7	7.43E-06	1.34E-07
5	7.21E-07	5.93E-08	4.89E-06	1.59E-07	12.8	87.2	1.06E-05	1.14E-07
6	3.58E-07	4.95E-08	3.56E-06	9.64E-09	9.1	90.9	7.22E-06	1.15E-08
biol. mean	5.37E-07		4.35E-06		9.8	90.2	7.80E-06	
biol. SD	4.13E-07		2.33E-06		4.8	4.8	2.70E-06	

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 1: oligonucleotide sequences for protein pull-down assays

Supplementary Table 3. DNA oligonucleotides used in protein pull-down studies.

ODN	Sequence (5′→ 3′)	Modifications
1	Biotin-GCA-TCC-GGT-CAY-CGT-TCC-TTC-GGA	Y = 5-octadienyl-U
2	Biotin-GCA-TCC-GGT-CAY-CAT-TCC-TTC-GGA	Y = 5-octadienyl-U
3	TCC-GAA-GGA-AXG-ATG-ACC-GGA-TGC	X= T
4		X= hmU
5		X= C
6		X= hmC
7	Biotin-GCT-CAC-GCT-AGY-CGA-CTC-CGT-GCA	Y = 5-octadienyl-U
8	TGC-ACG-GAG-TXG-ACT-AGC-GTG-AGC	X = T
9		Y = hmU

Hybridization scheme:

Pull-down 1: hmU:A vs. T:A : ODN4/2 vs. ODN3/2
Pull-down 2: hmU:G vs. C:G = ODN4/1 vs. ODN5/1
Pull-down 3: hmC:G vs. C:G = ODN6/1 vs. ODN5/1
Pull-down 4 (scrambled sequence): hmU:A vs. T:A = ODN9/7 vs. ODN8/7

Supplementary Note 2: LC-UV-ESI-MS/MS analysis of DNA

Supplementary Table 4. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA. CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as in the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
$[^{15}N_2]$-caC	274.08	Wide	158.03	Wide	170	5	5	Positive
caC	272.09	Wide	156.04	Wide	170	5	5	Positive
$[^{15}N_2,D_2]$-hmC	262.12	enhanced	146.07	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
$[D_3]$-mC	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C	228.1	enhanced	112.05	enhanced	1	1	0	Positive
time segment 4.0–6.0 min								
$[D_2]$-hmU	259.09	Wide	216.08	Wide	48	7	5	Negative
$[D_2]$-hmU	259.09	Wide	126.05	Wide	48	7	5	Negative
hmU	257.08	Wide	214.07	Wide	48	7	5	Negative
hmU	257.08	Wide	124.04	Wide	48	7	5	Negative
$[^{15}N_2]$-fU	257.06	Wide	213.05	Wide	48	6	5	Negative
fU	255.06	Wide	212.06	Wide	48	6	5	Negative
time segment 6.0–9.0 min								
$[^{15}N_5]$-8-oxo-G	289.08	Wide	173.04	Wide	120	9	7	Positive
8-oxo-G	284.1	Wide	168.05	Wide	120	9	7	Positive
$[^{15}N_2]$-fC	258.09	Wide	142.04	Wide	120	5	5	Positive
fC	256.09	Wide	140.05	Wide	120	5	5	Positive

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 5. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled thymidine ([^{13}C,$^{15}N_2$]-T). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ([M+H]$^+$ species) as well as in the negative ([M-H]$^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
[^{13}C,$^{15}N_2$]-caC	275.09	wide	159.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
[^{13}C,$^{15}N_2$]-hmC	261.11	enhanced	145.06	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
[^{13}C,$^{15}N_2$]-mC	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
[^{13}C,$^{15}N_2$]-C	231.1	enhanced	115.05	enhanced	40	1	3	Positive
C	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
[^{13}C,$^{15}N_2$]-hmU	260.08	wide	215.07	wide	50	7	5	Negative
hmU	257.08	wide	214.07	wide	50	7	5	Negative
[^{13}C,$^{15}N_2$]-fU	258.06	wide	213.05	wide	50	6	5	Negative
fU	255.06	wide	212.06	wide	50	6	5	Negative
time segment 6.0–9.0 min								
[$^{15}N_5$]-8-oxo-G	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
[^{13}C,$^{15}N_2$]-fC	259.09	wide	143.04	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive
[^{13}C,$^{15}N_2$]-T	246.1	enhanced	130.05	enhanced	30	40	3	Positive
T	243.1	enhanced	127.05	enhanced	30	40	3	Positive

Supplementary Table 6. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled (*methyl*-^{13}C,D_3)-methionine.CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ([M+H]$^+$ species) as well as in the negative ([M-H]$^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
[^{13}C]-caC	273.09	wide	157.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
[^{13}C,D_2]-hmC	261.12	enhanced	145.08	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
[^{13}C,D_3]-mC	246.14	enhanced	130.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C-dN	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
[^{13}C,D_2]-hmU	260.09	wide	217.09	wide	60	7	5	Negative
hmU	257.08	wide	214.07	wide	60	7	5	Negative
[^{13}C,D]-fU	257.07	wide	214.07	wide	60	6	5	Negative
fU	255.06	wide	212.06	wide	60	6	5	Negative
time segment 6.0–9.0 min								
[$^{15}N_5$]-8-oxo-G	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
[^{13}C,D]-fC	258.1	wide	142.06	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive

Nature Chemical Biology: doi:10.1038/nchembio.1532

Spiking amounts of labeled internal standards for quantitative LC-MS/MS analysis

The quantification of nucleosides of genomic DNA isolated from mESC or mouse tissue was carried out with the following amounts of internal standards: 51.03 pmol [D_3]-mC, 7.655 pmol [$^{15}N_2$,D_2]-hmC, 45.6 fmol [$^{15}N_2$]-fC, 43.0 fmol [$^{15}N_2$]-caC, 108.9 fmol [$^{15}N_5$]-8-oxo-G; 160.1 fmol [D_2]-hmU and 180.0 fmol [$^{15}N_2$]-fU. The quantification of nucleosides of genomic DNA isolated from HEK293 cells overexpressing Tet was carried out with the following amounts of internal standards: 34.02 pmol [D_3]-mC, 5.103 pmol [$^{15}N_2$,D_2]-hmC, 303.8 fmol [$^{15}N_2$]-fC, 215.1 fmol [$^{15}N_2$]-caC, 108.9 fmol [$^{15}N_5$]-8-oxo-G; 160.1 fmol [D_2]-hmU and 180.0 fmol [$^{15}N_2$]-fU. Genomic DNA samples isolated from cells grown in media supplemented with either [^{13}C,$^{15}N_2$]-T or [^{13}C,D_3]-methionine were not spiked with internal standards except [$^{15}N_5$]-8-oxo-G.

Validation of the LC-UV-MS/MS quantification method:

Method validation, in particular linearity, precision, and accuracy (i.e. determined from matrix samples spiked with isotopically labeled internal standards) of the established method were investigated. Validation for the established LC-UV-ESI-MS/MS quantification method was based on three different series (i.e., calibration functions and quality control samples) accomplished on different days. Each calibration standard (5-8 standard concentrations) was analyzed five times. Each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Linear regression was applied by Origin® 6.0 (*Microcal*™) to obtain calibration curves. Therefore, the ratio of the area under the curve (A/A*) of the unlabeled nucleoside to the internal standard (*) was plotted against the ratio of the amount of substance (n/n*) of the unlabeled nucleoside to the internal standard (*) (see **Supplementary Fig. 13**). Calibration functions were calculated without weighting. Additionally, acceptable accuracy (80–120%) as well as precision (<20% RSD) was required. Accuracy was proven by computing the amount of substance n from the obtained A/A* ratios of the calibration standards using the respective calibration function. Here, accuracy was defined as the ratio of the used amount of substance to the calculated amount of substance in percent and had to be between 80–120% for each standard concentration. Precision was defined as follows: technical replicates of A/A* ratios for each calibration standard had to have relative standard deviations (RSD) smaller than 20%. The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving a response of at least three times the response compared with the blank response. A compilation of absolute and relative LLOQs is shown in **Supplementary Table 7**.

Quality control samples to evaluate intra-batch precision (see below) were investigated using a biological sample spiked with internal standards. Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of −20 °C was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 7. Compilation of absolute lower limits of quantification [fmol] (LLOQ; see Supplementary Fig. 13) and relative LLOQs [per N] depending on the amount of DNA, which is digested. The relative LLOQs were computed by generating ratios of the absolute LLOQ [pmol] to the total amount of nucleosides (N; [pmol]) in the respective amount of DNA [µg]. The total amount of nucleosides was obtained by using the average molar mass of 308.91 g mol^{-1} for the monomeric DNA entity by taking the GC-content (21% C or G) in mouse into account.

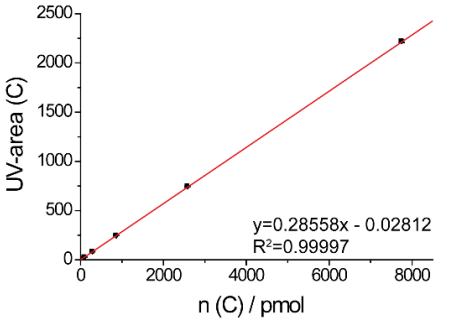
	absolute LLOQ [fmol]	**relative LLOQ [per N]**	**relative LLOQ [per N]**	**relative LLOQ [per N]**
DNA amount		**5 µg**	**10 µg**	**25 µg**
mC	104.5	6.5E-06	3.2E-06	1.3E-06
hmC	100.8	6.2E-06	3.1E-06	1.2E-06
fC	2.1	1.3E-07	6.5E-08	2.6E-08
caC	2.0	1.2E-07	6.2E-08	2.5E-08
8-oxo-G	2.0	1.2E-07	6.2E-08	2.5E-08
dU	14.1	8.7E-07	4.4E-07	1.7E-07
hmU	6.4	4.0E-07	2.0E-07	7.9E-08
fU	5.7	3.5E-07	1.8E-07	7.0E-08

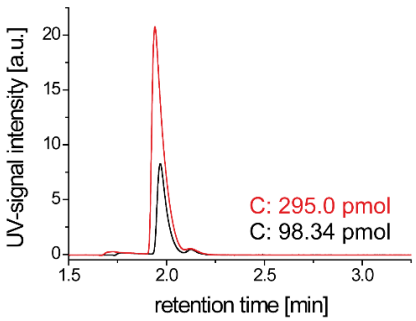
Intra-batch assay and quantification data processing:

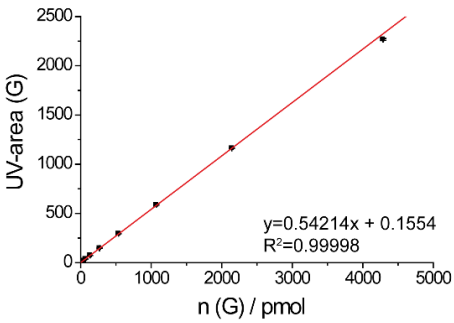
In order to evaluate intra-batch precision (see below) quality samples were investigated using a biological sample spiked with internal standards. The intra-batch-assay was performed for the LC-ESI-MS/MS analysis of the nucleosides G, C, mC, hmC, fC, caC, hmU, fU and 8-oxo-G. For this, a representative mESC DNA sample was analyzed. Technical replicates (n=5; each 4 µg DNA) were independently prepared using the below described digestion protocol. For data processing MassHunter Quantitative Analysis from *Agilent* was used. The area under the curve (A) was determined by LC-MS/MS for mC, hmC, fC, caC, hmU, fU, 8-oxo-G and for the corresponding labeled internal standards (A*); the area under the curve (A_{UV}) for G and C was determined by LC-UV. The amount of substance (n; pmol) of each nucleoside was computed by using the calibration curves (see **Supplementary Fig. 13**). The total sample volume was 40 µL, the injection volume after sample filtration was 29 µL. Therefore, the obtained values of G and C by LC-UV quantification were corrected by the factor given by the ratio of 40 µL/29 µL. Careful monitoring of the exact pipetting and injection volumes was therefore necessary. The obtained absolute amounts (pmol) of the DNA modifications (X= C, mC, fC, caC, hmU, fU, 8-oxo-G) were then related to the amount of G (pmol) giving ratios of X / G in %. The sum of X / G was defined as 100%. These values were then transferred in X / N values, considering that the G content is 21% in mouse.

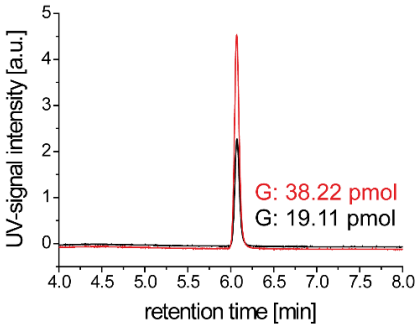
The determined A/A* ratios of the DNA nucleosides to the labeled internal standards (see **Supplementary Table 8**) showed a high precision (RSD = 3.9-18%) for each nucleoside. The necessity in using labeled internal standards for quantification is shown by comparing these results with the relative standard deviation (RSD = 7.6-49.5%) of the uncorrected mass signal (A) of the respective DNA modification. Moreover, in order to gain precision between sample batches measured on different days (data not shown), it is even more important to use internal standards. No memory effect was observed during blank experiments performed after several measurements of a sample. The blank analyses were not contaminated by carry-over.

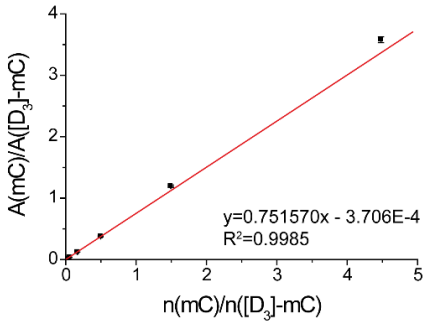
Nature Chemical Biology: doi:10.1038/nchembio.1532

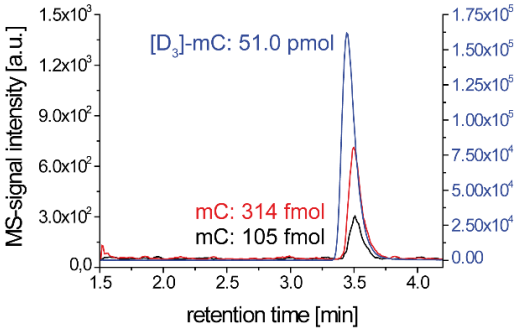


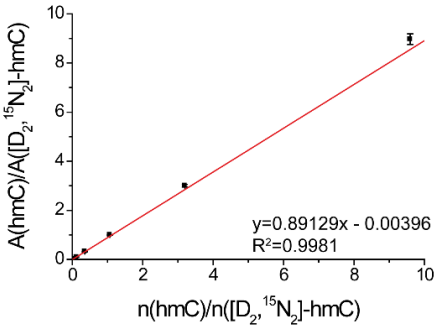


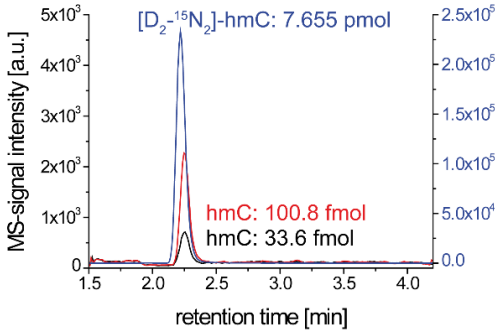


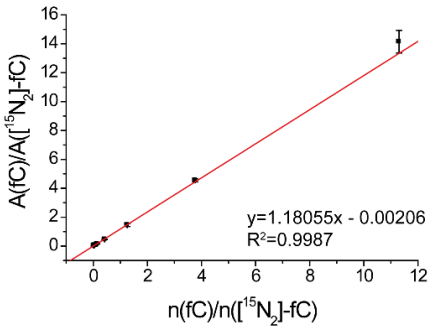


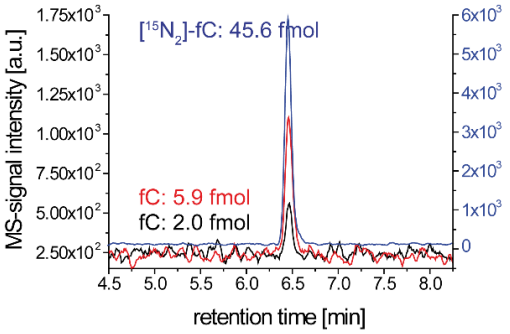




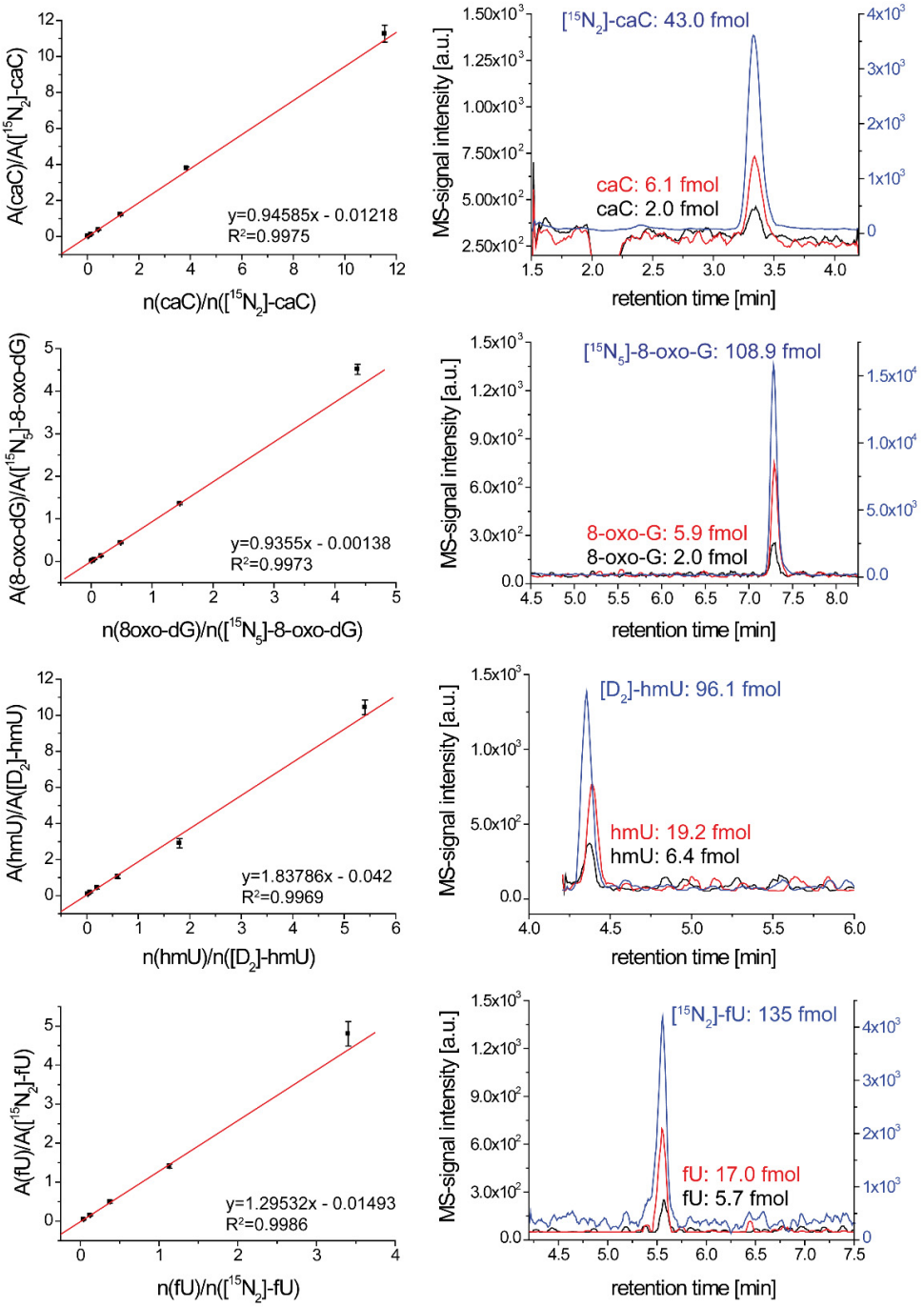








Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 13. UV and LC-MS/MS calibration curves and representative chromatograms of C (UV), G (UV), mC/[D_3]-mC, hmC/[$^{15}N_2$,D_2]-hmC, fC/[$^{15}N_2$]-fC, caC/[$^{15}N_2$]-caC, 8-oxo-G/[$^{15}N_5$]8-oxo-G, U/[$^{15}N_2$]-U, hmU/[D_2]-hmU, fU/[$^{15}N_2$]-fU. These were obtained by applying the compound-dependent parameters summarized in Supplementary Table 4. For hmU the MS/MS transition 257→214 was used. Depicted are the means of five technical replicates of one sample batch. Error bars reflect SD. Linearity was given across the following compound amounts in 29 µL injection volume: 10.63–7751 pmol C; 19.11–4892 pmol G; 104.5 fmol–228.6 pmol mC; 100.8 fmol–73.45 pmol hmC; 2.1–515.0 fmol fC; 2.0–496.6 fmol caC; 2.0–475.7 fmol 8-oxo-G; 6.4–519.3 fmol hmU; 5.7–459.6 fmol fU. The amounts of the labeled internal standards in 29 µL injection volume were as follows: 51.03 pmol [D_3]-mC; 7.655 pmol [$^{15}N_2$,D_2]-hmC; 45.6 fmol [$^{15}N_2$]-fC; 43.0 fmol [$^{15}N_2$]-caC; 108.9 fmol [$^{15}N_5$]-8-oxo-G; 96.1 fmol [D_2]-hmU; 135.0 fmol [$^{15}N_2$]-fU.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 8. Intra-batch-assay and quantification data processing.

	A_{UV}(G)	n(G) [pmol]	A_{UV}(C)	n(C) [pmol]	C / G [%]	C / N
techn. replicate 1	956	2431	485	2341	96.3	2.02E-01
techn. replicate 2	980	2493	495	2391	95.9	2.01E-01
techn. replicate 3	1043	2654	531	2566	96.7	2.03E-01
techn. replicate 4	979	2492	498	2406	96.5	2.03E-01
techn. replicate 5	982	2498	498	2406	96.3	2.02E-01
techn. mean value	988	2513	501	2422	96.4	2.02E-01
SD	33	83	18	85	0.3	6.29E-04
RSD [%]	3.3	3.3	3.5	3.5	0.3	0.3
	A(mC)	**A([D_3]-mC)**	**A(mC)/ A([D_3]-mC**	**n(mC) [pmol]**	**mC / G [%]**	**mC / N**
techn. replicate 1	636358	510501	1.247	91.9	3.78	7.94E-03
techn. replicate 2	678284	512180	1.324	97.7	3.92	8.23E-03
techn. replicate 3	565889	447103	1.266	93.4	3.52	7.39E-03
techn. replicate 4	664238	546597	1.215	89.6	3.60	7.56E-03
techn. replicate 5	719777	570445	1.262	93.1	3.73	7.82E-03
techn. mean value	652909	517365	1.263	93.1	3.71	7.79E-03
SD	57206	46591	0.040	2.9	0.16	3.30E-04
RSD [%]	8.8	9.0	3.1	3.1	4.2	4.2
	A(hmC)	**A([D_2,$^{15}N_2$]-hmC)**	**A(hmC)/ A([D_2,$^{15}N_2$]-hmC)**	**n(hmC) [pmol]**	**hmC / G [%]**	**hmC / N**
techn. replicate 1	80754	84338	0.958	8.51	0.350	7.35E-04
techn. replicate 2	115869	108774	1.065	9.46	0.380	7.97E-04
techn. replicate 3	264594	269496	0.982	8.72	0.329	6.90E-04
techn. replicate 4	139093	133830	1.039	9.23	0.371	7.78E-04
techn. replicate 5	116163	106074	1.095	9.73	0.389	8.18E-04
techn. mean value	143295	140502	1.028	9.13	0.364	7.64E-04
SD	70941	74213	0.057	0.51	0.024	5.12E-05
RSD [%]	49.5	52.8	5.6	5.6	6.7	6.7
	A(fC)	**A([$^{15}N_2$]-fC)**	**A(fC)/ A([$^{15}N_2$]-fC)**	**n(fC) [pmol]**	**fC / G [%]**	**fC / N**
techn. replicate 1	163427	31747	5.148	0.199	8.18E-03	1.72E-05
techn. replicate 2	178366	32585	5.474	0.211	8.48E-03	1.78E-05
techn. replicate 3	196827	35504	5.544	0.214	8.07E-03	1.69E-05
techn. replicate 4	193392	33755	5.729	0.221	8.88E-03	1.86E-05
techn. replicate 5	193493	35959	5.381	0.208	8.32E-03	1.75E-05
techn. mean value	185101	33910	5.455	0.211	8.38E-03	1.76E-05
SD	14070	1816	0.214	0.008	3.17E-04	6.66E-07
RSD [%]	7.6	5.4	3.9	3.9	3.8	3.8
	A(caC)	**A([$^{15}N_2$]-caC)**	**A(caC)/ A([$^{15}N_2$]-caC)**	**n(caC) [pmol]**	**caC / G [%]**	**caC / N**
techn. replicate 1	1444	6666	0.217	0.0104	4.28E-04	8.99E-07
techn. replicate 2	1735	8205	0.211	0.0102	4.08E-04	8.57E-07
techn. replicate 3	2111	9709	0.217	0.0104	3.93E-04	8.26E-07
techn. replicate 4	1985	8301	0.239	0.0114	4.59E-04	9.63E-07
techn. replicate 5	1927	8581	0.225	0.0108	4.31E-04	9.05E-07
techn. mean value	1840	8292	0.222	0.0106	4.24E-04	8.90E-07
SD	260	1089	0.011	0.0005	2.48E-05	5.20E-08
RSD [%]	14.1	13.1	4.8	4.6	5.8	5.8
	A(hmU)	**A([D_2]-hmU)**	**A(hmU)/ A([D_2]-hmU)**	**n(hmU) [pmol]**	**hmU / G [%]**	**hmU / N**
techn. replicate 1	894	5799	0.154	0.0090	3.69E-04	7.74E-07
techn. replicate 2	1278	6368	0.201	0.0130	5.23E-04	1.10E-06
techn. replicate 3	1561	7679	0.203	0.0133	5.00E-04	1.05E-06

Nature Chemical Biology: doi:10.1038/nchembio.1532

techn. replicate 4	1695	8148	0.208	0.0137	5.49E-04	1.15E-06
techn. replicate 5	1236	9010	0.137	0.0075	3.00E-04	6.29E-07
techn. mean value	1333	7401	0.181	0.0113	4.48E-04	9.41E-07
SD	312	1309	0.033	0.0029	1.08E-04	2.27E-07
RSD [%]	23.4	17.7	18.1	25.3	24.2	24.2
	A(fU)	**A([$^{15}N_2$]-fU)**	**A(fU)/ A([$^{15}N_2$]-fU)**	**n(fU) [pmol]**	**fU / G [%]**	**fU / N**
techn. replicate 1	7527	19661	0.383	0.0553	2.27E-03	4.78E-06
techn. replicate 2	8672	22769	0.381	0.0550	2.21E-03	4.63E-06
techn. replicate 3	11884	23930	0.497	0.0711	2.68E-03	5.63E-06
techn. replicate 4	10143	27784	0.365	0.0528	2.12E-03	4.45E-06
techn. replicate 5	14349	30152	0.476	0.0682	2.73E-03	5.73E-06
techn. mean value	10515	24859	0.420	0.0605	2.40E-03	5.04E-06
SD	2695	4148	0.061	0.0085	2.82E-04	5.93E-07
RSD [%]	25.6	16.7	14.5	14.0	11.7	11.7
	A(8oxo-G)	**A([$^{15}N_5$]-8oxo-G)**	**A(8oxo-G)/ A([$^{15}N_5$]-8oxo-G)**	**n(8oxo-G) [pmol]**	**8oxo-G / G [%]**	**8oxoG / N**
techn. replicate 1	237655	245919	0.966	0.113	4.63E-03	9.73E-06
techn. replicate 2	256991	266182	0.965	0.113	4.51E-03	9.48E-06
techn. replicate 3	310924	294412	1.056	0.123	4.64E-03	9.74E-06
techn. replicate 4	282245	299479	0.942	0.110	4.41E-03	9.26E-06
techn. replicate 5	327930	305105	1.075	0.125	5.01E-03	1.05E-05
techn. mean value	283149	282219	1.001	0.117	4.64E-03	9.75E-06
SD	37187	25225	0.060	0.007	2.28E-04	4.79E-07
RSD [%]	13.1	8.9	6.0	6.0	4.9	4.9

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 3: LC-MS/MS quantification results of genomic DNA

Supplementary Table 9. LC-MS/MS quantification results of mESC (WT01, J1 and R1) wild type, knock down (KD) and knock out cells (related to Fig. 1c,, 2a and 4a). The results of independent biological replicates are shown, the biological mean values / N and the biological standard deviation (SD). n.d. = not detected.

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (WT01) 1	2.01E-01	8.08E-03	6.53E-04	9.35E-06	6.01E-07	6.94E-07	3.51E-06	6.75E-06
mESC (WT01) 2	2.02E-01	7.71E-03	6.41E-04	9.81E-06	5.90E-07	7.04E-07	2.52E-06	5.95E-06
mESC (WT01) 3	2.02E-01	7.78E-03	6.82E-04	9.26E-06	5.24E-07	4.81E-07	1.60E-06	4.02E-06
mESC (WT01) 4	2.01E-01	8.17E-03	6.92E-04	1.03E-05	6.44E-07	4.73E-07	2.17E-06	4.90E-06
mESC (WT01) 5	2.02E-01	6.72E-03	8.63E-04	1.36E-05	7.25E-07	3.83E-07	4.34E-06	5.76E-06
mESC (WT01) 6	2.03E-01	6.43E-03	7.70E-04	1.32E-05	3.81E-07	5.14E-07	1.30E-06	3.17E-06
mESC (WT01) 7	2.03E-01	6.10E-03	8.18E-04	1.08E-05	6.77E-07	4.78E-07	1.03E-06	3.47E-06
biol. mean value	2.02E-01	7.28E-03	7.31E-04	1.09E-05	5.92E-07	5.32E-07	2.35E-06	4.86E-06
SD	7.75E-04	8.48E-04	8.63E-05	1.81E-06	1.13E-07	1.21E-07	1.21E-06	1.36E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet1 KD (WT01) 1	2.01E-01	8.99E-03	2.76E-04	3.31E-06	2.21E-07	1.42E-07	7.98E-07	3.21E-06
Tet1 KD (WT01) 2	2.02E-01	7.77E-03	1.39E-04	2.53E-06	2.20E-07	6.44E-08	2.00E-06	3.73E-06
Tet1 KD (WT01) 3	2.03E-01	6.37E-03	2.36E-04	6.51E-06	2.07E-07	3.07E-07	4.13E-06	7.08E-06
biol. mean value	2.02E-01	7.71E-03	2.17E-04	4.11E-06	2.16E-07	1.71E-07	2.31E-06	4.67E-06
SD	1.33E-03	1.31E-03	7.03E-05	2.11E-06	7.87E-09	1.24E-07	1.69E-06	2.10E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet2 KD (WT01) 1	2.00E-01	9.41E-03	4.45E-04	6.12E-06	2.38E-07	9.51E-08	5.18E-07	2.82E-06
Tet2 KD (WT01) 2	2.02E-01	7.25E-03	3.38E-04	4.52E-06	9.02E-08	3.23E-07	6.01E-07	1.78E-06
Tet2 KD (WT01) 3	2.02E-01	7.36E-03	3.98E-04	6.25E-06	3.36E-07	1.21E-07	2.50E-06	5.21E-06
Tet2 KD (WT01) 4	2.02E-01	7.69E-03	3.90E-04	6.78E-06	3.37E-07	n.d.	1.16E-06	6.52E-06
biol. mean value	2.02E-01	7.93E-03	3.93E-04	5.92E-06	2.50E-07	1.80E-07	1.20E-06	4.08E-06
SD	1.04E-03	1.01E-03	4.36E-05	9.73E-07	1.16E-07	1.25E-07	9.16E-07	2.17E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (J1) 1	2.04E-01	5.53E-03	3.61E-04	2.18E-06	7.95E-07	1.48E-06	3.98E-06	6.68E-06
mESC (J1) 2	2.04E-01	5.63E-03	3.32E-04	2.17E-06	9.30E-07	1.41E-06	4.19E-06	6.64E-06
biol. mean value	2.04E-01	5.58E-03	3.47E-04	2.17E-06	8.62E-07	1.44E-06	4.08E-06	6.66E-06
SD	4.96E-05	6.97E-05	2.06E-05	8.94E-09	9.55E-08	4.91E-08	1.51E-07	2.93E-08

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT1 -/- (J1) 1	2.07E-01	2.72E-03	1.80E-04	1.36E-06	3.10E-07	1.68E-06	4.56E-06	7.22E-06
DNMT1 -/- (J1) 2	2.07E-01	2.71E-03	1.77E-04	1.55E-06	2.94E-07	1.56E-06	3.90E-06	6.73E-06
biol. mean value	2.07E-01	2.72E-03	1.79E-04	1.46E-06	3.02E-07	1.62E-06	4.23E-06	6.97E-06
SD	5.80E-06	3.35E-06	2.28E-06	1.31E-07	1.11E-08	8.47E-08	4.73E-07	3.45E-07

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT3ab -/- (J1) 1	2.10E-01	4.12E-04	3.01E-05	2.46E-07	n.d.	8.02E-07	6.25E-07	2.68E-06
DNMT3ab -/- (J1) 2	2.10E-01	3.76E-04	3.54E-05	5.67E-07	n.d.	2.60E-06	7.86E-07	5.73E-06
DNMT3ab -/- (J1) 3	2.10E-01	3.81E-04	3.03E-05	3.71E-07	n.d.	2.27E-06	3.94E-07	2.32E-06
DNMT3ab -/- (J1) 4	2.10E-01	2.67E-04	2.67E-05	4.52E-07	n.d.	1.71E-06	4.50E-07	1.79E-06
DNMT3ab -/- (J1) 5	2.10E-01	2.50E-04	1.07E-05	3.78E-07	n.d.	7.41E-07	4.51E-06	7.68E-06
DNMT3ab -/- (J1) 6	2.10E-01	2.32E-04	1.60E-05	2.64E-07	n.d.	5.28E-07	2.87E-06	5.50E-06
biol. mean value	2.10E-01	3.20E-04	2.48E-05	3.80E-07		1.44E-06	1.61E-06	4.28E-06
SD	9.60E-05	7.83E-05	9.51E-06	1.19E-07		8.76E-07	1.70E-06	2.35E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (R1) 1	2.01E-01	8.37E-03	2.00E-04	1.04E-06	n.d.	1.52E-06	2.56E-06	1.21E-05
mESC (R1) 2	2.03E-01	6.76E-03	3.30E-04	2.00E-06	3.03E-07	9.23E-07	1.20E-06	n.d.
mESC (R1) 3	2.01E-01	8.68E-03	2.70E-04	2.12E-06	n.d.	1.77E-06	4.67E-06	1.05E-05
biol. mean value	2.02E-01	7.94E-03	2.67E-04	1.72E-06	3.03E-07	1.40E-06	2.81E-06	8.42E-06
SD	9.92E-04	1.03E-03	6.52E-05	5.91E-07		4.35E-07	1.75E-06	5.00E-06

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 10. LC-MS/MS quantification results of different murine organs of 3 months old wild type individuals (*n*) (related to Fig. 1d, 2 and Supplementary Fig. 2). Compiled are mean values / N obtained from three independent technical replicates and the standard deviation (SD).

n	organ	DNA isolation	C / N		mC / N		hmC / N		fC / N	
			techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
1	cerebellum		2.01E-01	2.30E-04	8.06E-03	2.40E-04	6.59E-04	1.08E-05	3.45E-07	3.14E-08
1	cerebellum	BHT, + Desf.,	2.01E-01	2.67E-04	8.27E-03	2.59E-04	7.37E-04	1.76E-05	2.86E-07	1.30E-08
1	cerebellum	BHT, + Desf., +THU	2.01E-01	3.66E-04	8.52E-03	3.45E-04	7.29E-04	2.14E-05	2.97E-07	2.03E-08
2	cerebellum	BHT, + Desf., +THU	2.00E-01	3.92E-04	9.81E-03	2.80E-04	6.37E-04	4.00E-05	2.73E-07	1.16E-08
3	cerebellum	BHT, + Desf., +THU	2.00E-01	3.95E-04	8.95E-03	3.45E-04	6.90E-04	2.91E-05	2.61E-07	1.57E-08
4	cerebellum	BHT, + Desf., +THU	2.00E-01	4.55E-04	9.46E-03	3.60E-04	6.69E-04	2.45E-05	2.59E-07	3.89E-09
1	kidney		2.01E-01	1.72E-04	8.13E-03	1.59E-04	4.37E-04	2.79E-05	2.27E-07	1.10E-08
1	kidney	BHT, + Desf.	2.02E-01	3.38E-04	7.79E-03	3.29E-04	4.33E-04	8.84E-06	2.25E-07	2.75E-08
1	kidney	BHT, + Desf., +THU	2.02E-01	1.70E-04	7.71E-03	1.76E-04	4.27E-04	2.15E-05	2.11E-07	1.19E-08
2	kidney	BHT, + Desf., +THU	2.01E-01	4.44E-04	8.59E-03	3.75E-04	3.78E-04	1.35E-05	1.86E-07	1.43E-08
3	kidney	BHT, + Desf., +THU	2.01E-01	1.46E-04	8.20E-03	1.02E-04	3.73E-04	1.67E-05	1.88E-07	8.82E-09
4	kidney	BHT, + Desf., +THU	2.01E-01	3.17E-04	8.30E-03	2.45E-04	3.79E-04	2.06E-05	1.94E-07	2.45E-08
1	cortex	BHT, + Desf., +THU	2.00E-01	4.48E-04	9.06E-03	3.37E-04	1.12E-03	7.21E-05	4.09E-07	4.17E-08
2	cortex	BHT, + Desf., +THU	1.99E-01	2.05E-04	9.29E-03	1.48E-04	1.37E-03	2.05E-05	4.34E-07	1.14E-08
3	cortex	BHT, + Desf., +THU	2.00E-01	4.65E-04	9.23E-03	3.73E-04	1.14E-03	7.48E-06	4.24E-07	1.92E-08
1	hippocampus	BHT, + Desf., +THU	1.98E-01	1.88E-04	1.08E-02	2.11E-04	1.56E-03	6.83E-05	4.49E-07	2.82E-08
2	hippocampus	BHT, + Desf., +THU	1.98E-01	2.13E-04	1.08E-02	1.73E-04	1.55E-03	3.08E-05	7.08E-07	4.67E-08
3	hippocampus	BHT, + Desf., +THU	1.97E-01	1.78E-04	1.09E-02	1.44E-04	1.71E-03	2.66E-05	3.89E-07	2.66E-08
1	heart	BHT, + Desf., +THU	2.02E-01	3.25E-04	7.96E-03	2.87E-04	4.08E-04	3.32E-05	1.85E-07	1.66E-08
2	heart	BHT, + Desf., +THU	2.02E-01	1.68E-04	7.87E-03	1.31E-04	3.84E-04	2.30E-05	1.48E-07	2.64E-09
3	heart	BHT, + Desf., +THU	2.01E-01	4.25E-04	8.07E-03	3.42E-04	4.56E-04	4.39E-05	1.41E-07	1.35E-08
1	liver	BHT, + Desf., +THU	2.01E-01	2.32E-04	8.35E-03	1.92E-04	2.44E-04	3.24E-06	1.66E-07	2.23E-08
2	liver	BHT, + Desf., +THU	2.02E-01	3.16E-04	8.10E-03	2.61E-04	2.73E-04	8.57E-06	1.78E-07	2.26E-08
3	liver	BHT, + Desf., +THU	2.01E-01	3.04E-04	8.42E-03	2.39E-04	2.82E-04	3.44E-05	1.81E-07	2.91E-08

n	organ	DNA isolation	hmU / N		fU / N		8-oxo-G / N	
			techn. mean	SD	techn. mean	SD	techn. mean	SD
1	cerebellum		7.27E-08		2.03E-06	2.29E-07	5.09E-06	1.92E-07
1	cerebellum	BHT, + Desf.,	n.d.		6.89E-07	2.65E-08	2.96E-06	6.04E-08
1	cerebellum	BHT, + Desf., +THU	n.d.		8.05E-07	7.34E-08	3.30E-06	3.79E-08
2	cerebellum	BHT, + Desf., +THU	1.30E-07	7.94E-08	1.02E-06	4.83E-08	3.86E-06	2.70E-07
3	cerebellum	BHT, + Desf., +THU	6.01E-08	3.67E-08	9.33E-07	9.52E-08	3.66E-06	2.26E-07
4	cerebellum	BHT, + Desf., +THU	1.07E-07	7.00E-08	1.03E-06	8.63E-08	4.67E-06	5.69E-07
1	kidney		n.d.		7.58E-07	7.38E-08	3.66E-06	2.63E-07
1	kidney	BHT, + Desf.	n.d.		7.00E-07	5.30E-08	3.24E-06	3.73E-07
1	kidney	BHT, + Desf., +THU	n.d.		7.16E-07	1.73E-08	3.48E-06	2.63E-07
2	kidney	BHT, + Desf., +THU	1.08E-07	7.89E-08	1.08E-06	1.45E-07	3.73E-06	1.42E-07
3	kidney	BHT, + Desf., +THU	1.29E-07	9.06E-08	1.17E-06	1.67E-07	3.91E-06	5.99E-08
4	kidney	BHT, + Desf., +THU	n.d.		9.34E-07	6.36E-08	3.65E-06	1.57E-07
1	cortex	BHT, + Desf., +THU	1.46E-07	6.19E-09	1.28E-06	1.22E-07	2.51E-06	3.74E-08
2	cortex	BHT, + Desf., +THU	8.24E-08	2.81E-08	1.30E-06	7.35E-08	3.13E-06	1.16E-07
3	cortex	BHT, + Desf., +THU	2.05E-07	5.83E-08	1.43E-06	3.72E-08	3.58E-06	2.83E-07
1	hippocampus	BHT, + Desf., +THU	2.74E-07	6.06E-08	2.18E-06	1.29E-07	6.30E-06	2.17E-07
2	hippocampus	BHT, + Desf., +THU	6.00E-07	5.50E-08	3.20E-06	3.46E-07	1.31E-05	7.69E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

3	**hippocampus**	BHT, + Desf., +THU	4.68E-08	1.85E-08	9.85E-07	2.26E-07	5.03E-06	8.96E-08
1	**heart**	BHT, + Desf., +THU	4.14E-07	7.28E-08	1.97E-06	3.46E-07	7.85E-06	1.69E-06
2	**heart**	BHT, + Desf., +THU	9.81E-08	4.23E-09	1.15E-06	7.63E-08	2.48E-06	1.18E-07
3	**heart**	BHT, + Desf., +THU	1.04E-07	3.36E-08	1.55E-06	7.55E-08	3.14E-06	2.27E-07
1	**liver**	BHT, + Desf., +THU	n.d.		1.13E-06	9.97E-08	2.69E-06	3.93E-07
2	**liver**	BHT, + Desf., +THU	1.48E-07	8.67E-08	1.09E-06	1.96E-07	2.43E-06	9.41E-08
3	**liver**	BHT, + Desf., +THU	2.34E-08		1.05E-06	1.18E-07	2.39E-06	4.69E-08

Supplementary Table 11. LC-MS/MS quantification results of n = 6 independent mESC differentiation experiments without growth factors (related to Fig. 5a,b). Compiled are mean values / N obtained from three independent technical measurements and their standard deviation (SD). K = C57Bl6/129 derived mES cell line.

n	**Sample**	**C / N**		**mC / N**		**hmC / N**		**fC / N**		**caC / N**	
	mESC diff. time	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**
1	**K, t= 0h**	2.06E-01	1.47E-04	3.79E-03	1.52E-04	4.25E-04	8.84E-06	1.68E-05	9.20E-07	1.21E-06	5.80E-08
1	**K, t= 8h**	2.06E-01	9.23E-05	3.43E-03	9.96E-05	5.33E-04	1.27E-05	1.71E-05	7.46E-07	1.09E-06	2.97E-08
1	**K, t= 16h**	2.06E-01	1.07E-04	3.88E-03	1.15E-04	4.93E-04	9.98E-06	6.69E-06	2.33E-07	3.92E-07	4.81E-08
2	**R1, t= 0h**	2.08E-01	1.38E-04	1.87E-03	1.05E-04	3.89E-04	3.33E-05	2.17E-05	5.00E-07	1.52E-06	6.10E-08
2	**R1, t= 8h**	2.08E-01	2.55E-05	1.60E-03	1.73E-05	5.50E-04	1.14E-05	2.91E-05	1.12E-06	1.91E-06	3.30E-08
2	**R1, t= 16h**	2.08E-01	7.54E-05	1.74E-03	7.53E-05	5.87E-04	1.28E-05	1.89E-05	4.48E-07	1.04E-06	2.38E-08
3	**K, t= 0h**	2.05E-01	6.83E-05	4.24E-03	6.36E-05	3.84E-04	6.03E-06	9.95E-06	1.46E-07	1.74E-07	1.70E-08
3	**K, t= 8h**	2.05E-01	7.08E-05	4.08E-03	7.14E-05	4.85E-04	5.17E-06	1.47E-05	7.35E-07	2.60E-07	1.84E-08
3	**K, t= 24h**	2.05E-01	5.00E-05	4.65E-03	5.74E-05	4.48E-04	7.86E-06	4.00E-06	1.67E-07	8.74E-08	2.02E-08
4	**R1, t= 0h**	2.07E-01	6.37E-05	2.49E-03	5.86E-05	4.61E-04	1.43E-05	1.71E-05	4.28E-07	3.25E-07	4.80E-08
4	**R1, t= 8h**	2.07E-01	5.45E-05	2.27E-03	4.25E-05	5.53E-04	1.69E-05	2.63E-05	8.05E-07	5.71E-07	2.45E-08
4	**R1, t= 24h**	2.06E-01	5.12E-05	3.00E-03	5.28E-05	5.93E-04	7.13E-06	8.39E-06	2.73E-07	1.74E-07	2.04E-08
5	**K, t=0 h**	2.04E-01	1.02E-04	5.38E-03	1.04E-04	3.89E-04	4.27E-06	7.66E-06	2.13E-07	2.29E-07	1.73E-08
5	**K, t=8 h**	2.04E-01	1.74E-04	5.10E-03	1.72E-04	4.65E-04	4.64E-06	1.01E-05	5.36E-07	2.37E-07	8.38E-10
5	**K, t=16 h**	2.04E-01	2.04E-04	5.88E-03	2.09E-04	5.22E-04	9.13E-06	4.40E-06	2.13E-07	1.92E-07	1.34E-08
5	**K, t=24 h**	2.04E-01	1.84E-04	5.40E-03	1.79E-04	4.01E-04	6.47E-06	2.06E-06	1.87E-07	1.00E-07	1.16E-08
5	**K, t=40 h**	2.04E-01	1.71E-04	5.72E-03	1.71E-04	3.00E-04	1.20E-06	1.20E-06	5.32E-08	7.93E-08	
6	**R1, t=0 h**	2.06E-01	2.16E-05	3.56E-03	3.01E-05	5.85E-04	1.80E-05	2.08E-05	4.72E-07	5.49E-07	5.15E-08
6	**R1, t=8 h**	2.05E-01	9.98E-05	3.83E-03	8.48E-05	7.63E-04	1.51E-05	2.05E-05	5.32E-07	8.74E-07	9.94E-09
6	**R1, t=16 h**	2.06E-01	4.53E-05	3.44E-03	3.95E-05	6.89E-04	7.42E-06	1.20E-05	5.65E-07	3.03E-07	8.59E-09
6	**R1, t=24 h**	2.06E-01	1.19E-04	3.57E-03	8.65E-05	6.16E-04	3.52E-05	7.18E-06	2.03E-07	1.55E-07	2.87E-08
6	**R1, t=40 h**	2.05E-01	6.38E-06	4.95E-03	1.56E-05	4.42E-04	1.11E-05	1.97E-06	4.48E-08	9.46E-08	1.84E-08

n	**Sample**	**hmU / N**		**fU / N**		**8-oxo-G / N**	
	mESC diff. time	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**
1	**K, t= 0h**	3.75E-07	1.67E-09	3.89E-06	2.73E-07	6.66E-06	2.74E-07
1	**K, t= 8h**	1.08E-06	1.57E-07	4.83E-06	3.45E-07	7.79E-06	3.05E-07
1	**K, t= 16h**	1.01E-06	1.76E-07	5.11E-06	5.69E-07	8.07E-06	5.57E-07
2	**R1, t= 0h**	1.14E-06	3.65E-07	3.82E-06	4.10E-07	6.70E-06	1.72E-07
2	**R1, t= 8h**	2.97E-06	4.13E-07	5.23E-06	9.97E-08	9.08E-06	5.89E-07
2	**R1, t= 16h**	3.75E-06	6.55E-07	5.15E-06	5.72E-07	8.81E-06	5.37E-07
3	**K, t= 0h**	8.76E-07	2.02E-08	8.86E-06	9.44E-07	1.35E-05	9.84E-07
3	**K, t= 8h**	1.29E-06	1.76E-07	6.90E-06	9.81E-07	1.00E-05	6.53E-07
3	**K, t= 24h**	6.55E-07	5.89E-08	6.72E-06	7.12E-07	9.73E-06	5.55E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

4	**R1, t= 0h**	1.22E-06	1.58E-07	8.62E-06	7.83E-07	1.26E-05	8.90E-08
4	**R1, t= 8h**	1.91E-06	2.69E-07	6.02E-06	4.56E-07	8.05E-06	2.35E-07
4	**R1, t= 24h**	1.82E-06	3.67E-07	6.60E-06	6.25E-07	1.06E-05	6.27E-07
5	**K, t=0 h**	9.98E-07	2.95E-07	7.04E-06	1.35E-06	9.98E-06	9.08E-07
5	**K, t=8 h**	2.97E-06	2.62E-07	6.89E-06	5.67E-07	9.64E-06	5.85E-07
5	**K, t=16 h**	1.43E-06	3.46E-07	2.71E-06	1.92E-07	3.82E-06	3.04E-07
5	**K, t=24 h**	8.01E-07	2.16E-07	6.60E-06	5.93E-07	9.72E-06	1.33E-07
5	**K, t=40 h**	8.30E-07	3.44E-08	7.78E-06	1.31E-07	1.13E-05	3.07E-07
6	**R1, t=0 h**	1.85E-06	3.00E-07	8.62E-06	1.94E-07	1.12E-05	7.13E-07
6	**R1, t=8 h**	3.77E-06	6.22E-07	4.01E-06	7.44E-07	6.02E-06	1.41E-06
6	**R1, t=16 h**	3.94E-06	8.01E-08	5.93E-06	3.24E-07	7.61E-06	2.01E-07
6	**R1, t=24 h**	3.84E-06	3.90E-07	8.92E-06	1.03E-06	1.23E-05	2.83E-07
6	**R1, t=40 h**	1.84E-06	9.73E-08	8.44E-06	2.94E-07	1.26E-05	1.02E-06

Supplementary Table 12. Relative modification levels of combined data sets from differentiation (0–40 h) of R1 and C57Bl6/129-derived mESCs without growth factors (related to Fig. 5a and Supplementary Fig. 8). In order to obtain these, the absolute modification levels of t = 0 h time points compiled in Supplementary Table 11 were set as 1 and the modification levels of later time points respectively related to these. Summarized are the biological mean values at each differentiation time point and the standard deviation (SD).

	relative C / N		**relative mC / N**		**relative hmC / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.00000E+00		1.000E+00		1.000E+00	
t = 8 h	1.00024E+00	1.2480E-03	9.435E-01	7.579E-02	1.270E+00	8.082E-02
t = 16 h	9.98952E-01	1.4308E-03	1.004E+00	7.157E-02	1.296E+00	1.629E-01
t = 24 h	9.98556E-01	1.4742E-03	1.077E+00	9.513E-02	1.133E+00	1.176E-01
t = 40 h	9.96347E-01	3.4067E-03	1.228E+00	2.314E-01	7.634E-01	1.034E-02

	relative fC / N		**relative caC / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.000E+00		1.000E+00	
t = 8 h	1.280E+00	2.318E-01	1.339E+00	3.329E-01
t = 16 h	6.051E-01	1.980E-01	5.991E-01	2.175E-01
t = 24 h	3.769E-01	9.373E-02	4.393E-01	1.122E-01
t = 40 h	1.255E-01	4.352E-02	2.589E-01	1.225E-01

	relative hmU / N		**relative fU / N**		**relative 8-oxo-G / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.000E+00		1.000E+00		1.000E+00	
t = 8 h	2.258E+00	6.597E-01	9.221E-01	3.417E-01	9.024E-01	3.178E-01
t = 16 h	2.390E+00	8.012E-01	9.337E-01	4.755E-01	8.978E-01	4.410E-01
t = 24 h	1.278E+00	6.283E-01	8.742E-01	1.357E-01	9.086E-01	1.623E-01
t = 40 h	9.129E-01	1.155E-01	1.043E+00	8.856E-02	1.129E+00	3.665E-04

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 13. LC-MS/MS quantification results of mESC differentiation with the growth factors FGF-2 and ActA (related to Supplementary Fig. 9). Modified nucleosides / N are given as mean values plus SD of three independent technical replicates.

Sample	C / N		mC / N		hmC / N		fC / N	
EpiLC diff. time	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
t= 0h	2.096E-01	1.40E-05	3.92E-04	1.35E-05	4.57E-05	6.48E-07	2.88E-07	4.52E-09
t= 12h	2.093E-01	1.01E-05	5.80E-04	4.94E-06	6.85E-05	5.47E-06	1.65E-06	1.34E-07
t= 24h	2.074E-01	1.17E-04	2.29E-03	1.02E-04	2.60E-04	1.41E-05	4.45E-06	6.30E-09
t= 36h	2.044E-01	9.71E-05	4.96E-03	7.38E-05	5.81E-04	2.30E-05	6.16E-06	5.27E-08
t= 48h	2.020E-01	8.85E-05	7.23E-03	8.83E-05	7.53E-04	8.90E-07	4.92E-06	1.31E-07

Sample	hmU / N		fU / N		8-oxo-G / N	
EpiLC diff. time	techn. mean	SD	techn. mean	techn. mean	SD	techn. mean
t= 0h	4.76E-07	7.05E-08	1.87E-06	9.10E-08	4.00E-06	1.10E-07
t= 12h	2.01E-06	1.90E-07	3.91E-06	7.84E-07	9.80E-06	3.07E-07
t= 24h	2.78E-06	7.20E-07	2.42E-06	3.65E-07	5.19E-06	1.45E-07
t= 36h	2.10E-06	5.82E-07	2.20E-06	1.51E-07	5.42E-06	4.12E-07
t= 48h	1.90E-06	1.03E-07	3.74E-06	3.74E-07	6.69E-06	3.08E-07

Supplementary Table 14. HEK-293T wild type vs. HEK + Tet1cm vs. HEK + Tet1cd (related to Supplementary Fig. 7a). Modified nucleosides / N are given as mean values of three independent technical replicates.

	HEK-293T wild type		HEK + Tet1cm		HEK + Tet1cd	
Nucleosides	techn. mean	SD	techn. mean	SD	techn. mean	SD
C / N	2.01E-01	9.67E-05	2.02E-01	1.09E-04	2.03E-01	3.91E-04
mC / N	8.80E-03	9.62E-05	8.04E-03	1.09E-04	4.18E-03	3.93E-05
hmC / N	3.39E-05	5.44E-07	4.95E-05	1.04E-06	2.21E-03	9.42E-05
fC / N	2.43E-07	2.41E-09	2.18E-07	1.17E-08	2.48E-04	5.96E-06
caC / N	n.d.		n.d.		1.29E-04	3.35E-06
hmU / N	7.21E-07	5.93E-08	1.51E-07	9.07E-09	4.24E-05	2.51E-06
fU / N	4.89E-06	1.59E-07	1.42E-06	6.66E-08	8.79E-06	1.07E-07
8-oxo-G / N	1.06E-05	1.14E-07	6.83E-06	1.47E-07	8.31E-06	9.03E-07

Supplementary Table 15. DNA modification levels of Tet1 *in vitro* assay (related to Supplementary Fig. 7b). Plasmid DNA with full CpG methylation was treated with commercially available Tet1.

plasmid	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
untreated	4.8E-02	n.d.	n.d.	n.d.	n.d.	4.4E-06	8.6E-06
untreated	4.6E-02	n.d.	n.d.	n.d.	n.d.	4.0E-06	9.4E-06
+Tet1	7.6E-04	4.8E-03	5.0E-03	9.0E-03	3.8E-04	1.5E-04	4.0E-05
+Tet1	6.5E-04	4.7E-03	4.8E-03	9.3E-03	4.4E-04	1.6E-04	3.7E-05
-Tet1	4.8E-02	n.d.	3.5E-05	n.d.	1.6E-05	2.5E-04	4.0E-05
-Tet1	4.7E-02	n.d.	3.5E-05	n.d.	1.2E-05	2.3E-04	4.5E-05

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 16. Effect of Smug1 depletion on modification levels in mESCs (R1) and effect of TDG and SMUG1 depletion on modification levels in HEK-293T cells overexpressing Tet1cd (related to Supplementary Fig. 5). Percent change values for modified nucleosides of cells treated with esiRNA (targeting Smug1/SMUG1 or TDG) with respect to unrelated control esiRNA. The absolute modification content of HEK-293T cells was normalized based on Tet1cd expression levels (determined by TECAN reading). The percent change is given as a mean value of three independent technical replicates.

	mESC Smug1 KD		**HEK-293T + Tet1cd / TDG KD**		**HEK-293T + Tet1cd / SMUG1 KD**	
Nucleosides	**Percent change**	**SD**	**Percent change**	**SD**	**Percent change**	**SD**
hmC	-0.15	4.07	-2.22	2.61	-3.43	2.24
fC	-1.26	3.29	38.15	1.25	0.77	2.05
caC	14.04	29.24	33.51	2.72	6.59	3.24
hmU	37.14	9.47	-10.41	4.74	46.46	5.65
fU	71.00	15.73	-4.36	2.34	22.66	1.91
8-oxo-G	17.27	14.26	-5.36	2.04	-6.97	1.77

Supplementary Table 17. HEK-293T wild type, HEK with Tet1cd-overexpression, Tet1cd/Uhrf1 co-overexpression or Tet1cd/Uhrf2 co-overexpression (related to Supplementary Fig. 12). Modified nucleosides / N are given as mean values of three independent technical replicates.

	HEK-293T wt		**HEK + Tet1cd**		**HEK + Tet1cd + Uhrf1**		**HEK + Tet1cd + Uhrf2**	
Nucleosides	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**
C / N	2.04E-01	2.97E-05	2.04E-01	1.24E-04	2.05E-01	8.32E-05	2.04E-01	1.15E-04
mC / N	6.01E-03	2.74E-05	5.29E-03	1.22E-04	4.63E-03	8.93E-05	5.11E-03	1.15E-04
hmC / N	2.86E-05	1.57E-06	4.29E-04	1.85E-05	3.31E-04	1.93E-05	5.13E-04	9.85E-06
fC / N	4.51E-07	5.21E-08	5.97E-05	3.43E-06	7.21E-05	4.24E-06	1.39E-04	5.61E-06
caC / N	1.77E-07	2.67E-09	2.32E-05	8.50E-07	2.23E-04	1.84E-05	1.66E-04	1.21E-06
hmU / N	8.18E-07	8.89E-08	2.55E-06	6.26E-08	1.27E-05	4.30E-07	6.74E-06	1.59E-06
fU / N	7.60E-06	3.39E-07	5.40E-06	5.26E-07	5.64E-06	8.42E-07	1.22E-05	2.10E-06
8-oxo-G / N	1.27E-05	7.08E-07	9.23E-06	3.22E-07	6.93E-06	1.09E-06	1.62E-05	8.43E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 4: materials in cell culture

Supplementary Table 18. Overexpression plasmids and esiRNAs used in HEK-293T cell experiments.

Figure	Experiment	Sample	Plasmid DNA	esiRNA
S7a	HEK-293T +/- Tet1xx	HEK + Tet1cd	GFP-Tet1cd (7.5 µg)	x
		HEK + Tet1cm	mCh-Tet1cm (7.5 µg)	x
		Wild type	pCMV6-Cdk5Rap1-v2 (7.5 µg)	x
S5b,S5c	Tet1cd with TDG or SMUG1 KD	HEK + Tet1cd	GFP-Tet1cd (10 µg)	CDK5RAP1 esiRNA (5 µg)
		HEK + Tet1cd with TDG KD	GFP-Tet1cd (10 µg)	TDG esiRNA (5 µg)
		HEK + Tet1cd with SMUG1 KD	GFP-Tet1cd (10 µg)	SMUG1 esiRNA (5 µg)
S12	HEK-293T	Tet1cd	GFP-Tet1cd (6 µg)	x
	HEK-293T	Tet1cd + Uhrf1	GFP-Tet1cd, GFP-Uhrf1[1] (each 6 µg)	x
	HEK-293T	Tet1cd + Uhrf2	GFP-Tet1cd, GFP-Uhrf2[1] (each 6 µg)	x

Supplementary Table 19. Knockdown (KD) efficiencies by Tet relative to SCR shRNAs.

	FWD	REV	Reference
Tet1	GAGCCTGTTCCTCGATGTGG	CAAACCCACCTGAGGCTGTT	*Ito et. al.*[2]
Tet2	TGTTGTTGTCAGGGTGAGAATC	TCTTGCTTCTGGCAAACTTACA	*Ito et. al.*[2]
actin	AAGGCCAACCGTGAAAAGAT	GTGGTACGACCAGAGGCATAC	This work

Supplementary Table 20. Primers for qPCR analysis of Tet, Dnmt, Tdg and Smug1 of EpiLC differentiation and Smug1 knockdown samples.

	FWD	REV	Reference
Gapdh	CATGGCCTTCCGTGTTCCTA	CTTCACCACCTTCTTGATGTCATC	*Szwagierzcak et al.*[3]
Tet1	CCAGGAAGAGGCGACTACGTT	TTAGTGTTGTGTGAACCTGATTTATTGT	*Szwagierzcak et al.*[3]
Tet2	ACTTCTCTGCTCATTCCCACAGA	TTAGCTCCGACTTCTCGATTGTC	*Szwagierzcak et al.*[3]
Tet3	GAGCACGCCAGAGAAGATCAA	CAGGCTTTGCTGGGACAATC	*Szwagierzcak et al.*[3]
Dnmt1	CCTAGTTCCGTGGCTACGAGGAG	TCTCTCTCCTCTGCAGCCGACTC	This work
Dnmt3a	GCTTTCTTCTCAGCCTCCCT	CCATGCCAAGACTCACCTTC	This work
Dnmt3b	CTGGCACCCTCTTCTTCATT	ATCCATAGTGCCTTGGGACC	This work
Tdg	GTCTGTTCATGTCGGGGCTGAGTGAG	CTGCAGTTTCTGCACCAGGATGCGC	This work
Smug1	CACTGGGGCCTACCCATGA	CTCCCAAGCATAATCCACCG	This work

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 5: correlation analysis results of modification levels

Supplementary Table 21. Correlation analysis of DNA modification levels comparing murine tissues from three months old individuals (cortex, hippocampus, cerebellum, heart, liver and kidney). Pearson coefficients (p) and significance values (s) are summarized. n=24 independent DNA samples (see Supplementary Table 10). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, moderate correlations ($0.7 > |p| > 0.6$) with significance level lower than 0.001 are highlighted in light gray.

		C	mC	hmC	fC	8-oxo-G	hmU	fU
C	**p**	1.000	-0.986*	-0.935*	-0.832*	-0.534	-0.466	-0.479
	s		0.000	0.000	0.000	0.007	0.022	0.018
mC	**p**	-0.986*	1.000	0.863*	0.769*	0.534	0.464	0.460
	s	0.000		0.000	0.000	0.007	0.022	0.024
hmC	**p**	-0.935*	0.863*	1.000	0.887*	0.481	0.420	0.472
	s	0.000	0.000		0.000	0.017	0.041	0.020
fC	**p**	-0.832*	0.769*	0.887*	1.000	0.663*	0.586	0.649*
	s	0.000	0.000	0.000		0.000	0.003	0.001
8-oxo-G	**p**	-0.534	0.534	0.481	0.663*	1.000	0.837*	0.835*
	s	0.007	0.007	0.017	0.000		0.000	0.000
hmU	**p**	-0.466	0.464	0.420	0.586	0.837*	1.000	0.871*
	s	0.022	0.022	0.041	0.003	0.000		0.000
fU	**p**	-0.479	0.460	0.472	0.649*	0.835*	0.871*	1.000
	s	0.018	0.024	0.020	0.001	0.000	0.000	

Supplementary Table 22. Correlation analysis of DNA modification levels during early mESC differentiation (0-40 h). Pearson coefficients (p) and significance values (s) are summarized. n=22 independent DNA samples (see Supplementary Table 11). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, weak correlations of hmC/hmU with mC/C are highlighted in pale pink.

		C	mC	hmC	fC	caC	8-oxo-G	hmU	fU
C	**p**	1	-0.997*	0.324	0.806*	0.699*	-0.032	0.299	-0.145
	s		0.000	0.141	0.000	0.000	0.889	0.176	0.518
mC	**p**	-0.997*	1	-0.399	-0.815*	-0.693*	0.057	-0.356	0.160
	s	0.000		0.066	0.000	0.000	0.801	0.104	0.476
hmC	**p**	0.324	-0.399	1	0.404	0.182	-0.312	0.783*	-0.227
	s	0.141	0.066		0.062	0.417	0.157	0.000	0.309
fC	**p**	0.806*	-0.815*	0.404	1	0.797*	-0.266	0.303	-0.299
	s	0.000	0.000	0.062		0.000	0.232	0.170	0.177
caC	**p**	0.699*	-0.693*	0.182	0.797*	1	-0.435	0.165	-0.567
	s	0.000	0.000	0.417	0.000		0.043	0.463	0.006
8-oxo-G	**p**	-0.032	0.057	-0.312	-0.266	-0.435	1	-0.080	0.959*
	s	0.889	0.801	0.157	0.232	0.043		0.723	0.000
hmU	**p**	0.299	-0.356	0.783*	0.303	0.165	-0.080	1	-0.022
	s	0.176	0.104	0.000	0.170	0.463	0.723		0.921
fU	**p**	-0.145	0.160	-0.227	-0.299	-0.567	0.959*	-0.022	1
	s	0.518	0.476	0.309	0.177	0.006	0.000	0.921	

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary References

1. Pichler, G. et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* **112**, 2585-93 (2011).
2. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-3 (2011).
3. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38**, e181 (2010).
4. Huang, D.W., Sherman, B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.***4**(1):44-57 (2009).

Nature Chemical Biology: doi:10.1038/nchembio.1532

2.3 Phosphorylation of TET proteins is suppressed by O-GlcNAcylation through the glycosyltransferase OGT

Manuscript under revision

Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the glycosyltransferase OGT[$ #]

Christina Bauer[1], Klaus Göbel[1], Nagarjuna Nagaraj[2], Christian Colantuoni[1], Mengxi Wang[1], Udo Müller[1], Elisabeth Kremmer[3], Andrea Rottach[1]*, Heinrich Leonhardt[1,4]*

[1] Ludwig-Maximilians University Munich, Biocenter, Planegg-Martinsried, Germany
[2] Max Planck Institute for Biochemistry, Martinsried, Germany
[3] Helmholtz Center Munich, Institute for Molecular Immunology, München-Großhadern, Germany
[4] Center for Integrated Protein Science Munich (CIPSM)

Running title: Phosphorylation and O-GlcNAcylation of TET proteins

To whom correspondence should be addressed: Prof. Dr. Heinrich Leonhardt, Ludwig-Maximilians University Munich, Department Biologie II, Großhadernerstr. 2, 81925 Planegg-Martinsried, Germany
Tel: +49 (0) 89 / 2180 - 74229, Fax: +49 (0) 89 / 2180 - 74236, h.leonhardt@lmu.de

Keywords: epigenetics, 5-hydroxymethylcytosine (5-hmC), post-translational modification (PTM), O-linked N-acetylglucosamine (O-GlcNAc), phosphorylation, dioxygenase, TET proteins, OGT

CAPSULE

Background: TET proteins oxidize 5-methylcytosine and contribute to active DNA demethylation.
Results: OGT modifies TET proteins with N-acetylglucosamine and thereby counteracts TET phosphorylation.
Conclusion: An interplay of post-translational modifications (PTMs) regulates TET proteins at low complexity regions.
Significance: This first map of phosphorylation and O-GlcNAcylation sites of TET proteins at amino acid resolution is the basis for understanding TET regulation.

ABSTRACT

TET proteins oxidize 5-methylcytosine (mC) to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxylcytosine (caC) and thus provide a possible means for active DNA demethylation in mammals. Although their catalytic mechanism is well characterized and the catalytic dioxygenase domain is highly conserved, the function of the regulatory regions — the N-terminus and the low complexity insert between the two parts of the dioxygenase domains — is only poorly understood. Here, we demonstrate that TET proteins are subject to a variety of PTMs that mostly occur at these regulatory regions. We mapped TET modification sites at amino acid resolution and show for the first time that TET1, TET2, and TET3 are highly phosphorylated. The glycosyltransferase OGT, which we identified as a strong interactor of all three TET proteins, catalyzes the addition of an N-acetylglucosamine (GlcNAc) group to serine and threonine residues of TET proteins and thereby decreases both the number of phosphorylation sites as well as the site occupancy. Interestingly, the different TET proteins display unique PTM patterns and some modifications occur in distinct combinations. In summary, our results provide a novel potential mechanism for TET protein regulation based on a dynamic interplay of phosphorylation and O-GlcNAcylation at the N-terminus and the low complexity insert region. Our data suggest strong crosstalk between the modification sites that could allow rapid adaption of TET protein localization, activity, or targeting due to changing environmental conditions as well as in response to external stimuli.

INTRODUCTION

A major epigenetic mechanism of gene regulation in higher eukaryotes is methylation of DNA at the carbon 5 atom of cytosines (1,2). Recently, the family of TET (ten-eleven-translocation)[5] proteins has been shown to successively oxidize mC to hmC, fC and caC (3-6), providing novel insights

into the dynamics of DNA modifications. TET proteins are also active on genomic thymine residues, leading to the generation of 5-hydroxyuracil (hmU) (7). In gnathostomata, there are three TET proteins: TET1, TET2, and TET3 (8), that show distinct expression patterns and functions in different tissues or during development (9-13). TET1 and TET2 are highly expressed in mouse embryonic stem cells (mESCs) and are associated with oxidation of transcription start sites or gene bodies, respectively (14). TET3 is upregulated in the oocyte and oxidizes the silenced paternal pronuclear DNA (10,15). High levels of TET proteins and genomic hmC are described for neuronal tissues (11,16-18). In several patients with myeloid malignancies, mutations of *TET2* correlate with decreased hmC levels and altered gene expression patterns (19-22).

The activity of TET proteins directly depends on two co-factors: Fe(II) and 2-oxoglutarate (2-OG) (3,8). Interestingly, gain-of-function mutations of the enzymes responsible for 2-OG synthesis, IDH1 and IDH2, have been associated with tumorigenesis, in particular glioblastomata and acute myeloid leukemia (AML) (20,23,24). These mutations lead to the synthesis of 2-hydroxyglutarate (2-HG), a potent inhibitor of 2-OG dependent dioxygenases such as TET proteins (24,25). Since IDH1 and IDH2 are enzymes of the Krebs cycle, these findings represent a direct link of TET protein activity to metabolism, especially since low hmC levels are not only found in AML patients with *TET2* loss-of-function mutations, but also with *IDH2* gain-of-function mutations (20). Besides 2-HG, ascorbate has also been shown to influence cytosine oxidation by TET proteins (26-28). In summary, TET protein activity appears to be modulated by several small molecules, either inhibitory such as 2-HG or stimulating like ascorbate.

TET proteins are not only influenced by certain metabolites, but also by interacting proteins. TET1 forms complexes with heterochromatin-associated proteins like HDAC1, HDAC2, SIN3A or EZH2 (29). All three TET proteins interact with a variety of factors of the base-excision repair pathway, including PARP1, LIG3, or XRCC1, and also with several DNA glycosylases such as TDG, NEIL1, or MDB4 (30). Another known interactor of TET proteins is the glycosyltransferase OGT (31-36) that represents an additional interesting connection with metabolism. OGT catalyzes the addition of a GlcNAc group to serine or threonine residues of target proteins (37). Its activity is dependent on the availability of a variety of metabolic molecules like glucose, ATP, glutamine and acetyl-CoA (38). The association of OGT with TET proteins has been reported to influence histone modifications and gene expression (31,36), TET1 protein stability (33) and activity (34) and, for TET3, also on subcellular localization (35).

Taken together, TET protein activity is widely studied in the context of development, tumorigenesis and metabolic conditions. However, only very little is known about the structure and function of the non-catalytic domains of TET proteins. In this study, we show that TET proteins are subject to a large number of PTMs, predominantly occurring at the two low complexity regions, which display only little sequence conservation: the N-terminus and the insert region that separates the two parts of the catalytic dioxygenase domain and is predicted to be unstructured (8). We describe that TET proteins are phosphorylated and that this phosphorylation can be suppressed via O-GlcNAcylation by the glycosyltransferase OGT. Detailed mapping of modification sites to the protein sequence shows that mostly N-terminus and insert region of TET proteins are subjected to PTMs and that their regulation depends on a dynamic interplay of different PTMs.

EXPERIMENTAL PROCEDURES

Antibody generation

A His-tagged protein fragment from the insert region of each TET protein (Supplemental data S1, S2) was expressed in *E. coli* BL21 (DE3) cells (Novagen, Darmstadt, Germany) and purified with the TALON^{TM} Superflow Metal Affinity Resin system (Clontech, Saint Germain, France) under native conditions as described previously (39). Approximately 100 µg of each antigen were injected both intraperitoneally (i.p.) and subcutaneously (s.c.) into Lou/C rats using CPG2006 (TIB MOLBIOL, Berlin, Germany) as adjuvant. After eight weeks, immune response was boosted i.p. and s.c. three days before fusion. Fusion of the myeloma cell line P3X63-Ag8.653 with the rat immune spleen cells was performed

using polyethylene glycol 1500 (PEG 1500, Roche, Mannheim, Germany). After fusion, the cells were cultured in 96 well plates using RPMI1640 with 20 % fetal calf serum, Penicillin/streptomycin, pyruvate, nonessential amino acids (PAA, Linz, Austria) supplemented by aminopterin (Sigma, St. Louis, MO). Hybridoma supernatants were tested in a solid-phase immunoassay. Microtiter plates were coated over night with His-tagged TET antigens at a concentration of 3-5 µg/ml in 0.1 M sodium carbonate buffer, pH = 9.6. After blocking with non-fat milk (Frema, Neuform, Zarrentin, Germany), hybridoma supernatants were added. Bound rat mAbs were detected with a cocktail of biotinylated mouse mAbs against the rat IgG heavy chains, avoiding IgM mAbs (α-IgG1, α-IgG2a, α-IgG2b (ATCC, Manassas, VA), α-IgG2c (Ascenion, Munich, Germany)). The biotinylated mAbs were visualized with peroxidase-labelled avidin (Alexis, San Diego, CA) and o-phenylenediamine as chromogen in the peroxidase reaction. Anti-TET1 5D6 (rat IgG2a), anti-TET1 5D8 (rat IgG2a), anti-TET1 2H9 (rat IgG2a), anti-TET1 4H7 (rat IgG2a), anti-TET2 9F7 (rat IgG2a), anti-TET3 11B6 (rat IgG2a) and anti-TET3 23B9 (rat IgG2a) were stably subcloned and further characterized (Supplemental data S1).

ESC culture, co-IP and MS/MS analysis

mESCs (J1) were cultured as described previously (9). Endogenous TET1 and TET2 proteins were pulled out via monoclonal antibodies (5D6, 5D8, and 9F7) coupled to protein G sepharose beads as described in (39). After co-immunoprecipitation (co-IP), protein samples were digested on beads with trypsin according to standard protocols. Peptide mixtures were analyzed using electrospray tandem mass spectrometry. Experiments were performed with an LTQ Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA). Spectra were analyzed with the Mascot™ Software (Matrix Science, Boston, MA).

Expression constructs

Expression constructs for GFP-TET1, GFP-TET2, GFP-TET3, GFP and mCherry (mCh) were described previously (40-42). To generate the mCh-OGT construct, the coding sequence was amplified using cDNA from mouse E14 ESCs as template and subcloned into the pCAG-Cherry-IB vector. Expression constructs for mCh-OGT^{H508A} (subsequently referred to as OGT^{mut}) were generated by overlap extension PCR. All constructs were verified by DNA sequencing (Eurofins, Ebersberg, Germany).

HEK293T culture, co-IP and Western blot analysis

Co-IP followed by Western blot with GFP- and mCh-tagged proteins expressed in HEK293T cells was performed as described previously (30). O-GlcNAc was detected with a mouse monoclonal antibody (RL2, abcam, Cambridge, UK) and an Alexa647N-conjugated secondary antibody (Sigma, St. Louis, MO).

Sample preparation for mass spectrometric analysis

All experiments were performed in triplicates. GFP-tagged TET proteins and/or mCh-tagged OGT or OGT^{mut} were expressed in HEK293T cells. Cell lysis with RIPA buffer and IP with the GFP-Trap® (Chromotek, Martinsried, Germany) was performed as described previously (30). After IP, samples on beads were rinsed two times with wash buffer (20 mM TrisHCl, pH = 7.5, 300 mM NaCl, 0.5 mM EDTA) and two times with IP buffer (20 mM TrisHCl, pH = 7.5, 150 mM NaCl, 0.5 mM EDTA).

100 µl of denaturation buffer (6 M GdnHCl, 10 mM Tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide in 100 M Tris pH = 8.5) was added to the beads and heated at 70 °C for 5 min. The samples were then subjected to sonication in Diagenode bioruptor plus (UCD-300-TO) at maximum power settings, for 10 cycles consisting of 30 s pulse and 30 s pause. Following sonication, the samples were diluted 1:10 with digestion buffer (25 mM Tris pH = 8.5 containing 10 % acetonitrile) and mixed by vortexing prior to enzyme digestion. Each sample was digested with 1 µg of endoproteinase lysC (Wako Chemicals, Neuss, Germany) for 4 h with subsequent digestion using 1 µg of trypsin (Promega, Madison, WI) under gentle rotation at 37 °C. After digestion, the samples were placed in a speed-vac for 10 minutes to remove acetonitrile from the sample before StageTip purification using SDB-XC material (43). Peptides were then eluted from the StageTip, placed in speed-vac to reduce the sample volume to approximately 6 µl and 5 µl of

the sample was injected on the column for MS/MS analysis.

Liquid chromatography tandem mass spectrometry and data analysis

Samples were loaded on a column (15 cm length and 75 µm inner diameter (New Objective, Woburn, MA)) packed with 3 µm Reprosil C18 beads (Dr Maisch GmbH, Ammerbuch-Entringen, Germany) using the auto sampler of Thermo Easy n-LC (Thermo Scientific, Waltham, MA) coupled via a nano-electrospray source to a LTQ Orbitrap XL mass spectrometer. Each sample was analyzed using a 2 h reversed phase gradient and using a top 5 method for data dependent acquisition. Full scans were acquired in the Orbitrap after accumulating up to $1x10^6$ charges and the MS/MS of the five most abundant precursors were performed using low energy ion trap CID. MS/MS spectra were recorded using the ion trap by radial ejection.

All raw files were analyzed using the MaxQuant (44) computational proteomics platform (version 1.4.1.6). Peak lists were searched with an initial mass deviation of 7 ppm and fragment ion deviation of 0.5 Th. Carbamidomethylation was used as fixed modification and oxidation of methionine, phosphorylation of serine, threonine and tyrosine, O-linked GlcNAc of serine and threonine, ubiquitination (diglycine motif) of lysine and acetylation of the protein N-terminus were used as variable modification. All unmodified and oxidized methionine and N-acetylation containing peptides were used for protein quantification.

MaxQuant output data were further analyzed with the Perseus software 1.4.1.3 (44). Modifications that were detected in only one out of three biological replicates were excluded from analysis. Hierarchical clustering (Figure 3) is based on euclidean distances.

Results

TET proteins interact with OGT

The three TET proteins share a common domain architecture: the C-terminal catalytic dioxygenase domain is split into two parts separated by a low complexity insert region and is preceded by an extension enriched in cysteines (8). All three TET proteins have a large N-terminal part that is mostly uncharacterized so far, except for a CXXC-type zinc finger at the N-terminus of TET1 and TET3 (8,40,45). Murine TET3 exists in two isoforms: one with the zinc finger and one without (41). The cysteine-rich region and the split-dioxygenase domain are conserved among the three murine TET proteins whereas N-terminus and insert region display only little sequence similarity (Figure 1a). The three-dimensional structure of mammalian TET proteins remains unresolved with the exception of the cysteine-rich and dioxygenase domains of TET2 (46), leaving structure and function of the N-terminus and the low complexity insert unknown.

As a first step towards understanding the regulation of TET proteins, we screened for interaction partners in mESCs. Since TET1 and TET2 are highly expressed in mESCs, we generated antibodies against murine TET1 and TET2 using protein fragments derived from the insert region of the catalytic domains as antigens. A detailed description of the obtained antibodies can be found in Supplemental data S1. Clones 5D6-anti-TET1, 5D8-anti-TET1 and 9F7-anti-TET2 proved to be suited for immunoprecipitation of the specific TET protein, respectively, and were used to pull down endogenous TET1 and TET2 from mESCs. Subsequent LC-MS/MS analysis revealed that both TET1 and TET2 interact with the glycosyltransferase OGT (Figure 1b). In accordance with this result, co-IP analysis of GFP-TET1 and GFP-TET2 followed by Western blot analysis shows co-precipitation of mCh-OGT. We also analyzed whether TET3 can interact with OGT and could indeed detect a strong mCh-OGT signal in the bound fraction of the co-IP of GFP-TET3. Taken together, our data indicate that all three TET proteins interact with OGT (Figure 1c).

OGT catalyzes O-GlcNAcylation of TET proteins

Having observed the interaction between TET proteins and OGT, we examined whether TET proteins are modified by OGT and screened for O-GlcNAcylation, the modification that is transferred to the OH-group of serine or threonine residues of target proteins by OGT (38,47). To this end, we specifically enriched GFP-tagged TET proteins co-expressed with either OGT or its catalytically inactive point mutant OGT^{mut} with the GFP-Trap® and probed the subsequent Western blot with an anti-GlcNAc antibody. All three TET proteins

were found to be increasingly O-GlcNAcylated dependent on the coexpression of catalytically active OGT (Figure 2).

O-GlcNAcylation suppresses phosphorylation of TET proteins

To identify OGT-dependent O-GlcNAcylation sites on TET proteins, we performed mass spectrometric analysis of semi-purified proteins. We therefore expressed GFP-tagged TET1, TET2, and TET3 in HEK293T cells, either with OGT, OGT^{mut}, or without interactor. After pull-down with the GFP-Trap® and stringent washing steps, the samples were analyzed by LC-MS/MS. An overall sequence coverage of about 50 % was achieved for TET1, about 60 % for TET2 and about 65 % for TET3 (Supplemental data S3, Supplemental table S4). For data analysis, only sites were considered that were detected in at least two out of three biological replicates. Without co-expression of interactor, only few residues on TET proteins are found to be O-GlcNAcylated at low site occupancy. Coexpression of OGT leads to a strong increase in both number of O-GlcNAcylation sites and site occupancy for TET2 and TET3. The difference in number of O-GlcNAc sites is either due to *de novo* modification by OGT or because the site occupancy without OGT co-expression is below detection limit. For TET1, however, the O-GlcNAc pattern is relatively heterogeneous and only few O-GlcNAc sites can be detected. This heterogeneity is also illustrated by the fact that residues 1327 and 327, which are O-GlcNAcylated in the TET1 samples, are only detected to be modified in one out of three replicates in the TET1/OGT samples and therefore do not occur in the heat map. Although mCh-OGT^{mut} is supposed to be catalytically inactive, co-expression leads to a small increase in O-GlcNAcylation and represents a distinct state from basal levels (Figure 3a).

Since O-GlcNAcylation occurs at serine or threonine residues of the target protein, we also screened for another post-translational modification that can occur at these amino acids, namely phosphorylation. Interestingly, high phosphorylation of TET1 and TET2 and, to a lesser extent, of TET3, was observed. Phosphorylation of all TET proteins decreased dramatically upon co-expression of active OGT, regarding both site occupancy and number of detected phosphorylation sites (Figure 3b).

PTMs occur mostly at the N-terminus and the low complexity insert of TETs

To date, the domains of TET proteins are largely uncharacterized except for the conserved catalytic dioxygenase domain and the CXXC-type zinc finger at the N-terminus of TET1 (8,40,46). Mapping of the detected O-GlcNAc and phosphorylation sites to the TET protein sequence reveals that mostly the N-terminus and the low complexity insert, which separates the two parts of the dioxygenase domain, are subjected to PTMs (Figure 4). Remarkably, O-GlcNAcylation and phosphorylation rarely occur at the exact same residue, although O-GlcNAcylation suppresses phosphorylation. Furthermore, the three TET proteins carry different modification patterns: Whereas TET1 is mostly modified at the N-terminus and the very C-terminal part and is hardly glycosylated, TET2 and TET3 show strong O-GlcNAcylation at the low complexity insert region. The first 350 amino acids of TET3 remain free of PTMs. The observed pattern is not due to differences in sequence coverage as the detected peptides are homogenously distributed over the whole protein sequence (Supplemental data S3).

Interestingly, some of the modifications are detected on the same peptides, indicating that they occur together at the same molecule. For example, TET2-S23 phosphorylation can be found with S15 phosphorylation and phosphorylation of S376 only occurs when S374 is O-GlcNAcylated, but not when it is phosphorylated (Table 2). For TET3, a variety of PTM combinations can be observed for residues 360-368 and 1071-1077. Phosphorylation at S362, for example, exists either alone or in combination with S360 O-GlcNAcylation and S368 phosphorylation. Phospho-S362 also co-occurs with O-GlcNAc-S361. If S362 is O-GlcNAcylated, however, no further modifications on this peptide was observed (Table 3, Figure 4). Apparently, some residues such as TET3-S362 serve as O-GlcNAc/phosphorylation switches that can either promote or suppress neighboring PTMs. These data indicate a strong crosstalk between O-GlcNAcylation and phosphorylation at different residues. Modifications on TET1, on the other hand, appear more isolated and no peptide bearing more than one modification was detected (Table

1). In summary, we detect many interdependent modification sites on TET proteins suggesting that TET1, TET2, and TET3 are dynamically regulated by PTMs.

DISCUSSION

Since oxidation of mC to hmC, fC and caC by TET proteins represents a potential mechanism for active DNA demethylation in higher vertebrates (3-5), these proteins are currently intensively investigated. Here, we provide evidence that all three TET proteins are subject to O-GlcNAcylation through OGT. This finding is in accordance with previous studies, showing that TET1 and TET2 interact with OGT in ESCs and are O-GlcNAcylated (33,34). TET3 has also been described to associate with OGT (32,35) and to alter its subcellular localization dependent on glucose metabolism and O-GlcNAcylation (35). OGT not only directly modifies TET proteins, but the interaction also promotes histone modifications such as H3K4me3 or H2BS112GlcNAc (31,36). TET1 has been shown to associate with the repressive SIN3A complex (48) and TET2 and TET3 with the SET1/COMPASS complex (31).

We show that per default, TET proteins are phosphorylated. Basal O-GlcNAc levels are low, but enhance upon OGT expression. Simultaneously, the phosphorylation levels decrease. This finding indicates regulation of the phosphorylation signal as a novel function for TET O-GlcNAcylation. Interestingly, the underlying mechanism of this observation seems not to be direct competition for the serine or threonine residue that is to be modified, but rather proximal site competition as neighboring residues are interdependent (49). O-GlcNAcylation and phosphorylation occur at distinct amino acids and often, several modifications of the same type appear in close proximity in "modification islands", e.g. O-GlcNAcylation at S1252/S1256/S1263 of TET3, or phosphorylation at S15/S23/S39 of TET2. Furthermore, phosphorylation is not reduced because the residues are masked by the co-expressed OGT as demonstrated by expression of the catalytically inactive OGT^{mut}, where the reduction is far less pronounced. Nevertheless, O-GlcNAcylation of TET proteins is slightly increased by OGT^{mut}, either due to residual activity of the mutant (50) or, more likely, due to recruitment of endogenous active OGT via trimerization of the TPR domain (51).

This observed effect of O-GlcNAcylation on phosphorylation is of particular interest since protein O-GlcNAc levels are influenced by a variety of factors, such as different subcellular localization of OGT and nutrient availability, and seem to be tightly regulated (38). For example, O-GlcNAcylation of TET3 can be enhanced when cells are cultured in high glucose medium, leading to nuclear export of TET3 (35). Therefore, OGT-dependent dephosphorylation represents a novel mechanism on how TET proteins can be regulated in response to changing environmental conditions.

Interestingly, some residues remain stably phosphorylated even at high OGT levels. For TET2 and TET3, they appear in close proximity to each other and just N-terminal of the cysteine-rich region. This persistence of phosphorylation suggests an important OGT-independent regulatory role of these residues that is of interest for future studies. Nevertheless, the majority of phosphorylation sites are reduced in occupancy upon O-GlcNAcylation. We thus observe two different types of phosphorylation, dependent and independent of O-GlcNAcylation.

The hypothesis of interdependence of PTMs on TET proteins is further strengthened by the fact that some modifications are detected on the same peptides in stable combinations whereas others occur as stand-alone modifications. Certain residues appear to be O-GlcNAc/phosphorylation switches that influence the PTM pattern on the neighboring amino acids. The observed crosstalk of modifications enables a variety of potential regulatory mechanisms that could fine-tune TET activity dependent on different environmental conditions such as nutrient availability.

To date, the domain architecture and three-dimensional structure of TET proteins is only poorly understood. The catalytic domain is highly conserved and homologous to other types of Fe(II)- and 2-OG-dependent dioxygenases that act on nucleic acids (8,52). Recently, the crystal structure of the catalytic part of TET2 has provided insights into the reaction mechanism (46). However, the large N-terminus and the low complexity insert, that is characteristic for TET proteins, remain poorly understood — both in

terms of structure and of function. So far, no homologous domains have been described except for the CXXC-type zinc finger at the N-terminus, and the insert region is predicted to be largely unstructured (8). In this study, we show that these two regions are subject to many dynamic PTMs. For TET1 and TET3, few modification sites are also found at the very C-terminus of the proteins, but N-terminus and insert region are the major target of O-GlcNAcylation and phosphorylation. In general, the lower the conservation of one region, the more modification sites are detected. This provides an interesting possible mechanism for regulation of TET protein activity, stability, or targeting. In line, TET1, TET2, and TET3 have been described to colocalize with OGT at transcription start sites and influence gene expression (31,34). The different modifications that we describe in this study might alter binding of TET interaction partners and thus provide a possible explanation of the observed dual role in both transcriptional activation and repression (53).

In summary, we provide the first systematic mapping of O-GlcNAcylation and phosphorylation sites on TET proteins at amino acid resolution. The distribution of these PTMs and the described crosstalk open new perspectives on the regulatory role of the so far poorly characterized non-catalytic domains, the N-terminus and the low complexity insert region. The observed O-GlcNAcylation and phosphorylation are linked to metabolic conditions and thus provide a possible mechanism of TET protein regulation in response to external stimuli.

REFERENCES

1. Goll, M. G., and Bestor, T. H. (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**, 481-514
2. Suzuki, M. M., and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476
3. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935
4. 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
5. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307
6. Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C. A., Hagemeier, C., and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* **50**, 7008-7012
7. Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., Schuermann, D., Michalakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Muller, U., Spruijt, C. G., Vermeulen, M., Leonhardt, H., Schar, P., Muller, M., and Carell, T. (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* **10**, 574-581
8. 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
9. 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. 110.1093/nar/gkq1684

10. 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., Iqbal, K., Shi, Y. G., Deng, Z., Szabo, P. E., Pfeifer, G. P., Li, J., and Xu, G. L. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610
11. Kim, M., Park, Y. K., Kang, T. W., Lee, S. H., Rhee, Y. H., Park, J. L., Kim, H. J., Lee, D., Lee, D., Kim, S. Y., and Kim, Y. S. (2014) Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. *Hum Mol Genet* **23**, 657-667
12. Langlois, T., da Costa Reis Monte Mor, B., Lenglet, G., Droin, N., Marty, C., Le Couedic, J. P., Almire, C., Auger, N., Mercher, T., Delhommeau, F., Christensen, J., Helin, K., Debili, N., Fuks, F., Bernard, O. A., Solary, E., Vainchenker, W., and Plo, I. (2014) TET2 deficiency inhibits mesoderm and hematopoietic differentiation in human embryonic stem cells. *Stem Cells* **32**, 2084-2097
13. Rudenko, A., Dawlaty, M. M., Seo, J., Cheng, A. W., Meng, J., Le, T., Faull, K. F., Jaenisch, R., and Tsai, L. H. (2013) Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109-1122
14. Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., Zepeda-Martinez, J. A., Pape, U. J., Jacobsen, S. E., Peters, B., and Rao, A. (2014) Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **111**, 1361-1366
15. 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. 210.1038/ncomms1240
16. Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., Jui, J., Jin, S. G., Jiang, Y., Pfeifer, G. P., and Lu, Q. (2013) Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep* **3**, 291-300
17. Kriaucionis, S., and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930
18. Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., Diao, J., Wu, F., He, H. H., Cui, Q., Clark, E., Ma, C., Barbara, A., Veenstra, G. J., Xu, G., Kaiser, U. B., Liu, X. S., Sugrue, S. P., He, X., Min, J., Kato, Y., and Shi, Y. G. (2012) Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151**, 1200-1213
19. Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., Huberman, K., Thomas, S., Dolgalev, I., Heguy, A., Paietta, E., Le Beau, M. M., Beran, M., Tallman, M. S., Ebert, B. L., Kantarjian, H. M., Stone, R. M., Gilliland, D. G., Crispino, J. D., and Levine, R. L. (2009) Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144-147
20. Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P. M., Schneider, S., Spiekermann, K., Leonhardt, H., and Bohlander, S. K. (2011) Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* **25**, 1649-1652
21. Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A., Lecluse, Y., Plo, I., Dreyfus, F. J., Marzac, C., Casadevall, N., Lacombe, C., Romana, S. P., Dessen, P., Soulier, J., Viguie, F., Fontenay, M., Vainchenker, W., and Bernard, O. A. (2009) Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-2301
22. 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., Liu, X. S., Aravind, L., Agarwal, S., Maciejewski, J. P., and Rao, A. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843

23. Dang, L., Jin, S., and Su, S. M. (2010) IDH mutations in glioma and acute myeloid leukemia. *Trends Mol Med* **16**, 387-397
24. 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., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liau, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744
25. Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., Ito, S., Yang, C., Wang, P., Xiao, M. T., Liu, L. X., Jiang, W. Q., Liu, J., Zhang, J. Y., Wang, B., Frye, S., Zhang, Y., Xu, Y. H., Lei, Q. Y., Guan, K. L., Zhao, S. M., and Xiong, Y. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30
26. Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martinez, J. A., Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Hirst, M., Rao, A., Lorincz, M. C., and Ramalho-Santos, M. (2013) Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222-226
27. Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., Liu, J., Liu, J., Wu, H., Mao, S. Q., Mo, K., Li, Y., Lai, K., Qi, J., Yao, H., Pan, G., Xu, G. L., and Pei, D. (2013) Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat Genet* **45**, 1504-1509
28. 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
29. Cartron, P. F., Nadaradjane, A., Lepape, F., Lalier, L., Gardie, B., and Vallette, F. M. (2013) Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer* **4**, 235-241
30. 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
31. Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Mendez, J., Murphy, N., Dawson, M. A., Volkmar, M., Putmans, P., Calonne, E., Shih, A. H., Levine, R. L., Bernard, O., Mercher, T., Solary, E., Urh, M., Daniels, D. L., and Fuks, F. (2013) TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* **32**, 645-655
32. Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., and Yokoyama, A. (2014) TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* **19**, 52-65
33. Shi, F. T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M. A., Wan, M., and Songyang, Z. (2013) Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J Biol Chem* **288**, 20776-20784
34. Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., and Pasini, D. (2013) Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* **49**, 645-656
35. 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 beta-N-Acetylglucosamine Transferase (OGT). *J Biol Chem* **289**, 5986-5996
36. Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561-564
37. Hanover, J. A., Yu, S., Lubas, W. B., Shin, S. H., Ragano-Caracciola, M., Kochran, J., and Love, D. C. (2003) Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys* **409**, 287-297
38. Harwood, K. R., and Hanover, J. A. (2014) Nutrient-driven O-GlcNAc cycling - think globally but act locally. *J Cell Sci* **127**, 1857-1867

39. Jost, K. L., Rottach, A., Milden, M., Bertulat, B., Becker, A., Wolf, P., Sandoval, J., Petazzi, P., Huertas, D., Esteller, M., Kremmer, E., Leonhardt, H., and Cardoso, M. C. (2011) Generation and characterization of rat and mouse monoclonal antibodies specific for MeCP2 and their use in X-inactivation studies. *PLoS One* **6**, e26499. 26410.21371/journal.pone.0026499
40. Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F., and Leonhardt, H. (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One* **6**, e16627. 16610.11371/journal.pone.0016627
41. Liu, N., Wang, M., Deng, W., Schmidt, C. S., Qin, W., Leonhardt, H., and Spada, F. (2013) Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS One* **8**, e62755. 62710.61371/journal.pone.0062755
42. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Muller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., and Vermeulen, M. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146-1159
43. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670
44. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372
45. Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y. G. (2010) TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res* **20**, 1390-1393
46. 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
47. Hanover, J. A., Krause, M. W., and Love, D. C. (2012) Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nat Rev Mol Cell Biol* **13**, 312-321
48. Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A., Rappsilber, J., and Helin, K. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348
49. Butkinaree, C., Park, K., and Hart, G. W. (2010) O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim Biophys Acta* **1800**, 96-106
50. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-567
51. Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A., and Conti, E. (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* **11**, 1001-1007
52. Loenarz, C., and Schofield, C. J. (2009) Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chem Biol* **16**, 580-583
53. Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., and Zhang, Y. (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389-393
54. Sievers, F., and Higgins, D. G. (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **1079**, 105-116

FOOTNOTES

[$] This article contains Supplemental data S1, S2, and S3 and Supplemental table S4

[#] This work was funded by the Deutsche Forschungsgemeinschaft (DFG, Collaborative Research Centers SFB 646/B10 and SFB 1064/A17). C. Bauer is supported by the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). H. Leonhardt is a member of the Nanosystems Initiative Munich (NIM).

* corresponding authors: A. Rottach (a.rottach@lmu.de) and H. Leonhardt (h.leonhardt@lmu.de)

[1] Ludwig-Maximilians University Munich, Biocenter, Planegg-Martinsried, Germany

[2] Max Planck Institute for Biochemistry, Martinsried, Germany

[3] Helmholtz Center Munich, Institute for Molecular Immunology, München-Großhadern, Germany

[4] Center for Integrated Protein Science Munich (CIPSM)

[5] abbreviations: TET (ten-eleven-translocation), O-GlcNAc transferase (OGT), post-translational modifications (PTMs), 5-methylcytosine (mC) , 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), 5-carboxylcytosine (caC), 5-hydroxyuracil (hmU), O-linked N-acetylglucosamine (O-GlcNAc); mouse embryonic stem cells (mESCs); 2-oxoglutarate (2-OG); acute myeloid leukemia (AML); 2-hydroxyglutarate (2-HG); intraperitoneally (i.p.); subcutaneously (s.c.); monoclonal antibodies (mAbs); co-immunoprecipitation (co-IP); mCherry fluorescent protein (mCh)

FIGURE LEGENDS

Figure 1: All three TET proteins interact with OGT

a) Schematic representation of the domain architecture of the three murine TET proteins: The catalytic dioxygenase domain (D) is split in two parts, separated by a presumably unstructured low complexity insert (8), and is N-terminally preceded by a cysteine-rich region (Cys). The Fe(II)-binding residues are marked by green asterisks. The N-terminus (NT) of TET1 contains a CXXC-type zinc finger (ZF); TET3 exists in two isoforms, one with zinc finger and one without (41). The mean percent identity of the single domains of TET1, TET2, and TET3 is represented by different shades of grey and was calculated with Clustal2.1 (54).
b) The table depicts the number of detected unique peptides of TET1, TET2, and OGT in IP experiments from mESCs followed by LC-MS/MS.
c) Co-IP of GFP-TET1, GFP-TET2, and GFP-TET3, each coexpressed with mCh-OGT, precipitated with the GFP-Trap® and analyzed by Western blot, shows that all three TET proteins interact with OGT (full arrowheads). GFP as a negative control does not coprecipitate mCh-OGT (empty arrowhead) (I: Input, FT: Flowthrough, B: Bound).

Figure 2: TET proteins are O-GlcNAcylated by OGT

Western blot analysis of TET1, TET2, and TET3 specifically enriched with the GFP-Trap®. Upon co-expression of active OGT, the O-GlcNAcylation signal increases for all three TET proteins (black arrowheads) compared to coexpression of catalytically inactive OGT^{mut}. Interaction between TET proteins and OGT is independent of OGT activity.

Figure 3: TET phosphorylation is suppressed by O-GlcNAcylation

a) Mass spectrometry analysis of TET proteins reveals that the number of O-GlcNAcylated residues and their modification occupancy strongly increases for TET2 and TET3 when OGT is co-expressed. In contrast, TET1 only harbors few O-GlcNAc sites. Heat maps display O-GlcNAc occupancy of single residues (mean of three biological replicates). Residue numbering refers to the murine protein sequences specified in the Supplemental data S3.
b) TET proteins are highly phosphorylated when expressed alone, but the number of phosphorylated residues and the fraction of phosphorylated protein decreases upon coexpression of OGT. Heat maps display phosphorylation occupancy of single residues (mean of three biological replicates). Residue numbering refers to the murine protein sequences specified in the Supplemental data S3.

Figure 4: N-termini and insert regions of TET proteins are densely modified

Schematic and scaled mapping of all TET phosphorylation and O-GlcNAcylation sites to the protein sequence. Modifications are mostly found in the N-terminus and the insert region and rarely occur at the same residue. Residue numbering refers to the murine protein sequences specified in the Supplemental data S3. Green asterisks mark catalytic Fe(II)-binding residues. Basal O-GlcNAc sites occur without any coexpression of OGT or OGT^{mut}; persistent phosphorylation sites show high occupancy despite increase of O-GlcNAcylation. An example of the PTM crosstalk on TET proteins is shown for TET3-S360/361/362/368. empty arrowhead: 2 cooccurring modifications; full arrowheads: 3 cooccurring modifications; blunt arrows: mutual exclusivity.

TABLES

# aa	LP	Modified sequence
160	100%	HSENDSVPSQHATVS(ph)PGTENGEQNR
177	100%	CLVEGES(ph)QEITQSCPVFEER
253	85%	NT(o-)SNQLADLSSQVESIK
270	84%	LS(o-)DPSPNPTGSDHNGFPDSSFR
320	100%	FILAGS(ph)QPDVFDTKPQEK
327	100%	FILAGSQPDVFDT(o-)KPQEK
556	81%	ATAM(ox)SMPVTQASTSS(ph)PPCNSTPPMVER
561	89%	ATAMSMPVTQASTSSPPCNS(ph)TPPM(ox)VER
734	98%	QQTNPS(ph)PTFAQTIR
736	96%	QQTNPSPT(ph)FAQTIR
794	77%	DAM(ox)SVTTS(o-)GGECDHLK
854	100%	DGS(ph)PVQPSLLSLMK
892	70%	LTLEQVVAIEALTQLSEAPSESSS(ph)PSKPEK
950	100%	S(ph)PDSFATNQALIK
969	74%	SQGYPSS(ph)PTAEK
1327	66%	REAQT(o-)SSNGPLGPTTDSAQSEFK
1964	89%	ELHATTSLRS(ph)PK
2016	100%	PADRECPDVS(ph)PEANLSHQIPSR
2016	56%	PADRECPDVS(o-)PEANLSHQIPSR
2042	99%	DNVVTVS(ph)PYSLTHVAGPYNR

Table 1: Detected modified peptides of TET1
aa: residue number of modified amino acid; LP: Localization probability; (ph): phosphorylated; (o-): O-GlcNAcylated; (ox): oxidized; LP was calculated with the MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in the Supplemental data S3. Only peptides that were detected in a minimum of two out of three samples are shown.

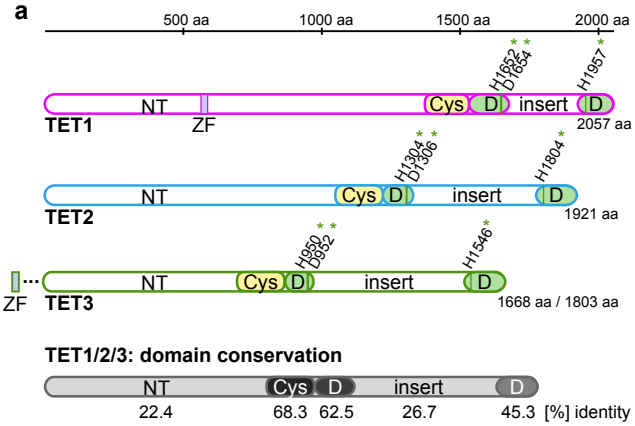
# aa	LP	Modified sequence
15	100%	TTHAEGTRLS(ph)PFLIAPPSPISHTEPLAVK
23	100%	TTHAEGTRLS(ph)PFLIAPPS(ph)PISHTEPLAVK
39	98%	LQNGS(ph)PLAERPHPEVNGDTK
95	98%	RT(o-)VS(o-)EPSLSGLHPNK
97	100%	TVS(ph)EPSLSGLHPNK
97	97%	RT(o-)VS(o-)EPSLSGLHPNK
165	100%	SSRQPNVSGLSDNGEPVTSTTQESSGADAFPT(o-)R
317	98%	SALDIGPS(o-)RAENK
374	82%	DS(ph)ISPTTVTPPSQSLLAPR
374	99%	DS(o-)IS(ph)PTTVTPPSQSLLAPR
376	99%	DS(o-)IS(ph)PTTVTPPSQSLLAPR
464	100%	TSSSQSLNPSVHTPNPPLMLPEQHQNDCGS(ph)PS(ph)PEK
466	100%	TSSSQSLNPSVHTPNPPLMLPEQHQNDCGS(ph)PS(ph)PEK
514	89%	QT(o-)QGSVQAAPGWIELK
545	94%	DIS(o-)LHSVLHSQTGPVNQM(ox)SSK
552	87%	DIS(o-)LHSVLHS(o-)QTGPVNQMSSK
561	95%	DISLHSVLHSQTGPVNQMS(o-)S(o-)K
562	97%	DISLHSVLHSQTGPVNQMS(o-)S(o-)K
565	98%	QS(o-)TGNVNM(ox)PGGFQR
603	100%	AQMYQVQVNQGPS(ph)PGMGDQHLQFQK
625	96%	ALYQECIPRT(o-)DPSSEAHPQAPSVPQYHFQQR
746	98%	VEESFCVGNQYS(o-)K
778	92%	ILT(o-)PNSSNLQILPSNDTHPACER
807	100%	EQALHPVGS(o-)K
889	100%	ALPVPEQGGSQTQT(ph)PPQK
944	100%	YPLS(ph)PPQENMSSR
951	97%	PSSYRYPLSPPQENMS(ph)SR
1437	63%	QM(ox)T(o-)AQPQLSGPVIRQPPTLQR
1443	98%	QMTAQPQLS(o-)GPVIR
1613	87%	DLHRYPNQDHLTNQNLPPIHT(o-)LHQQTFGDSPSK
1622	74%	YPNQDHLTNQNLPPIHTLHQQTFGDS(ph)PSK
1640	76%	DAFT(o-)TNSTLKPNVHHLATFSPYPTPK
1672	100%	M(ox)DSHFM(ox)GAAS(o-)R
1749	100%	TASAQELLYSLTGSS(ph)QEK

Table 2: Detected modified peptides of TET2
aa: residue number of modified amino acid; LP: Localization probability; (ph): phosphorylated; (o-): O-GlcNAcylated; (ox): oxidized; grey: multiple modifications occuring on one peptide; LP was calculated with the MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in the Supplemental data S3. Only peptides that were detected in a minimum of two out of three samples are shown.

# aa	LP	Modified sequence
360	93%	VEAPS(o-)SS(ph)PAPVPS(ph)PISQR
361	79%	VEAPSS(o-)S(ph)PAPVPSPISQR
362	100%	VEAPSSS(ph)PAPVPSPISQR
362	67%	VEAPSSS(o-)PAPVPSPISQR
368	100%	VEAPS(o-)SS(ph)PAPVPS(ph)PISQR
478	100%	S(ph)RDM(ox)QPLFLPVR
557	83%	S(ph)PSPM(ox)VALQSGSTGGPLPPADDKLEELIR
557	76%	S(o-)PSPM(ox)VALQSGSTGGPLPPADDKLEELIR
634	79%	IES(o-)SGAVTVLSTTCFHSEEGGQEATPTK
665	100%	AENPLT(ph)PTLSGFLES(ph)PLK
667	96%	AENPLTPT(ph)LSGFLESPLK
674	100%	AENPLT(ph)PTLSGFLES(ph)PLK
674	99%	AENPLT(ph)PTLSGFLES(o-)PLK
1008	83%	VS(o-)SGAIQVLTAFPR
1071	97%	QEALELAGVT(o-)T(o-)DPGLSLK
1072	89%	QEALELAGVT(o-)T(o-)DPGLSLK
1077	99%	QEALELAGVTT(o-)DPGLS(o-)LK
1105	89%	YS(o-)GNAVVESYSVLGSCRPSDPYSMSSVYSYHSR
1252	94%	VPQLHPAS(o-)RDPSPFAQSSSCYNR
1256	62%	VPQLHPASRDPS(o-)PFAQSSSCYNR
1263	84%	VPQLHPASRDPSPFAQSSS(o-)CYNR
1282	88%	QEPIDPLTQAES(o-)IPR
1293	100%	T(o-)PLPEASQNGGPSHLWGQYSGGPSMSPK
1318	100%	TPLPEASQNGGPSHLWGQYSGGPSM(ox)S(ph)PK
1351	61%	LNSFGAS(ph)CLTPSHFPESQWGLFTGEGQQSAPHAGAR
1404	76%	FGNGTSALTGPSLT(o-)EK
1412	100%	PWGM(ox)GT(o-)GDFNPALK
1442	100%	VEEGRIPT(ph)PGANPLDK
1651	72%	QALAMPTDSAVT(o-)VSSYAYTK
1653	71%	QALAM(ox)PTDSAVTVS(o-)SYAYTK
1654	77%	QALAMPTDSAVTVS(o-)S(o-)YAYTK
1658	99%	GAIPTRQALAMPTDSAVTVSSYAYT(o-)K

Table 3: Detected modified peptides of TET3
aa: residue number of modified amino acid; LP: Localization probability; (ph): phosphorylated; (o-): O-GlcNAcylated; (ox): oxidized; grey: multiple modifications occuring on one peptide; LP was calculated with the MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in the Supplemental data S3. Only peptides that were detected in a minimum of two out of three samples are shown.

Figure 1



b

Protein	unique peptides in respective IP		
	5D6-α-TET1	5D8-α-TET1	9F7-α-TET2
TET1	3	5	0
TET2	0	0	2
OGT	3	7	1

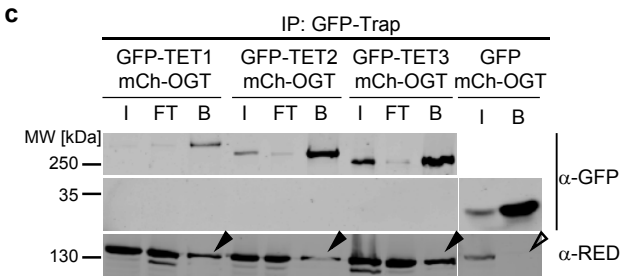


Figure 2

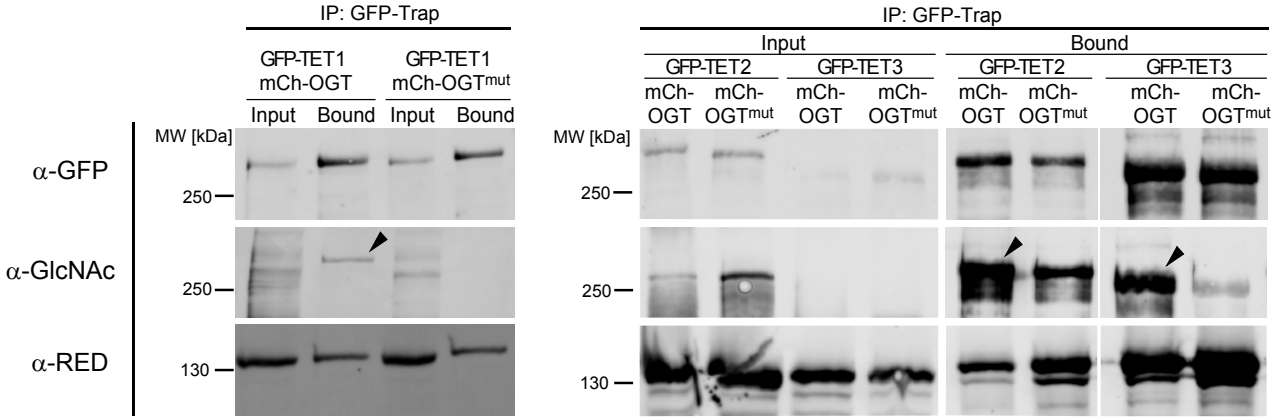

IP: GFP-Trap
GFP-TET1 mCh-OGT
GFP-TET1 mCh-OGTmut
Input
Bound
Input
Bound
MW [kDa]
α-GFP
α-GlcNAc
α-RED
250
250
130
IP: GFP-Trap
Input
Bound
GFP-TET2
GFP-TET3
GFP-TET2
GFP-TET3
mCh-OGT
mCh-OGTmut
mCh-OGT
mCh-OGTmut
mCh-OGT
mCh-OGTmut
mCh-OGT
mCh-OGTmut
MW [kDa]
250
250
130

Figure 3

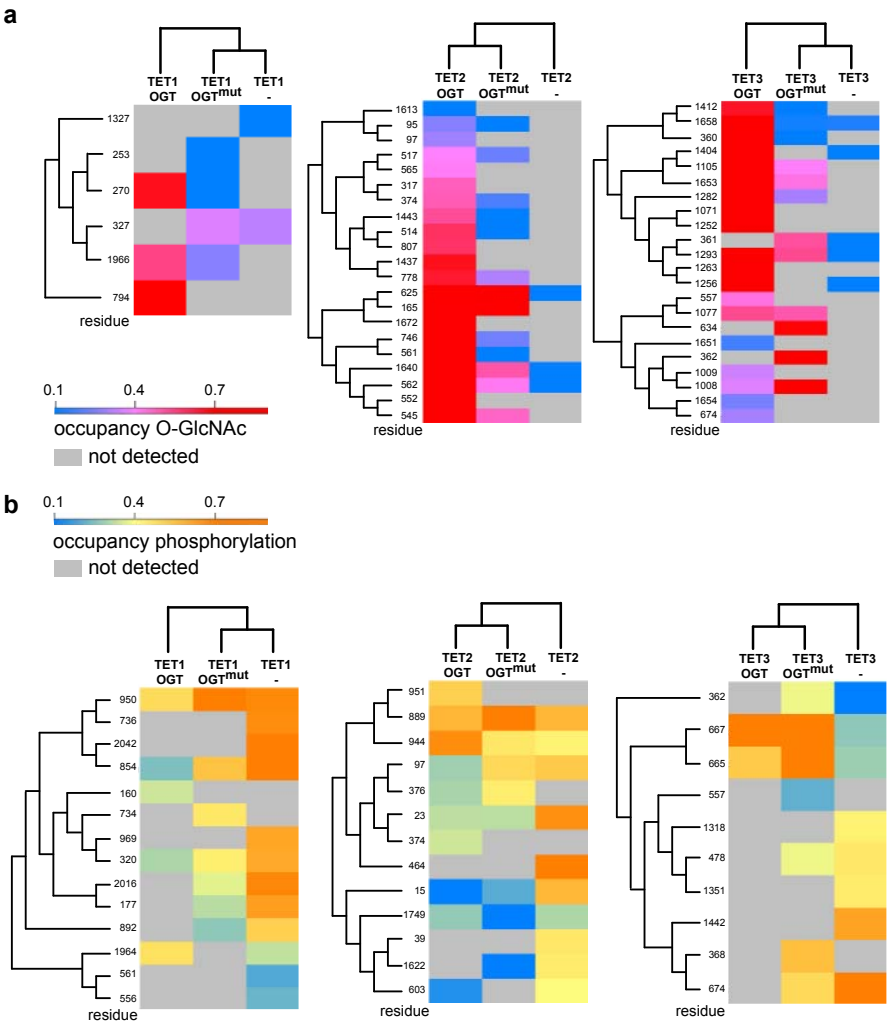
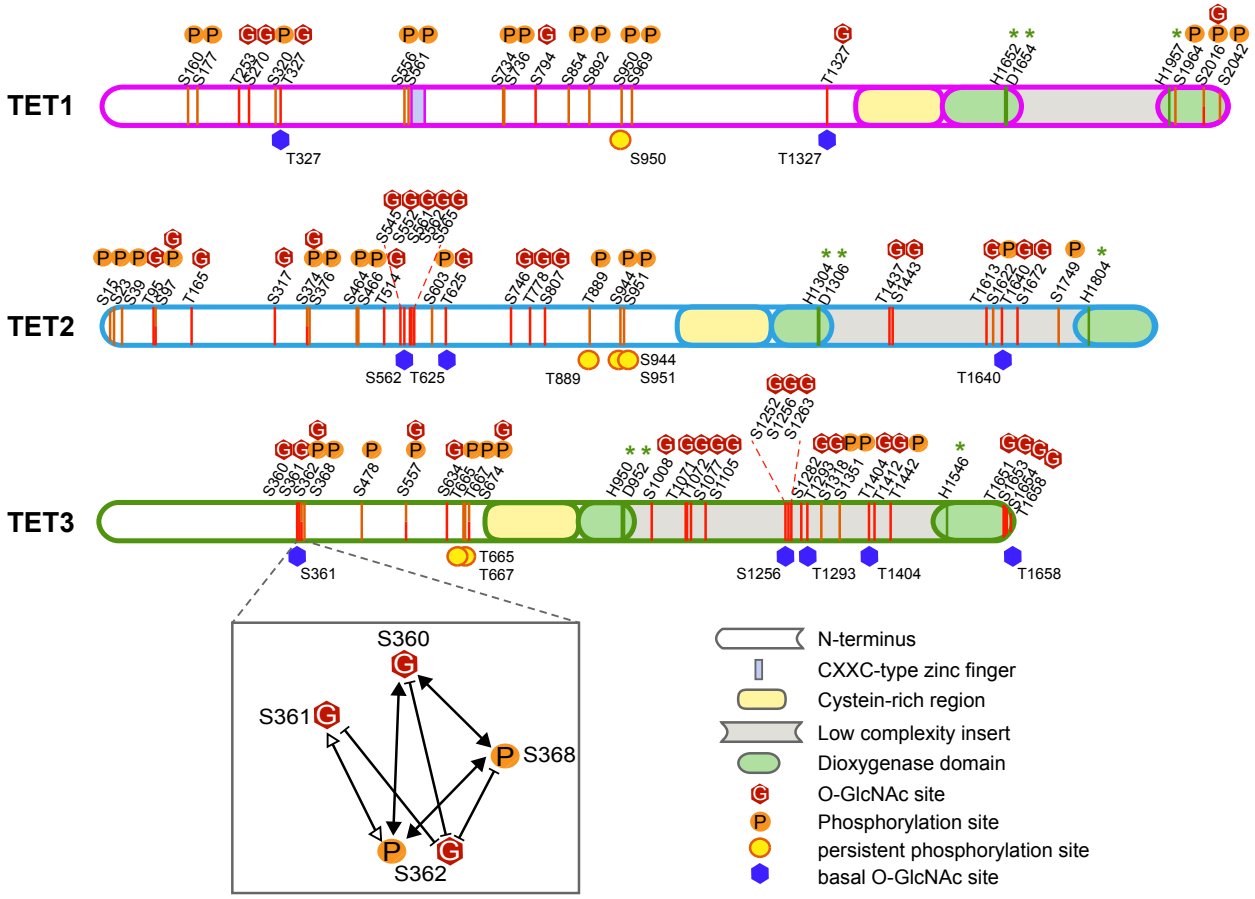

a
TET1 OGT
TET1 OGTmut
TET1 -
1327
253
270
327
1966
794
residue
TET2 OGT
TET2 OGTmut
TET2 -
1613
95
97
517
565
317
374
1443
514
807
1437
778
625
165
1672
746
561
1640
562
552
545
residue
TET3 OGT
TET3 OGTmut
TET3 -
1412
1658
360
1404
1105
1653
1282
1071
1252
361
1293
1263
1256
557
1077
634
1651
362
1009
1008
1654
674
residue
0.1
0.4
0.7
occupancy O-GlcNAc
not detected
b
0.1
0.4
0.7
occupancy phosphorylation
not detected
TET1 OGT
TET1 OGTmut
TET1 -
950
736
2042
854
160
734
969
320
2016
177
892
1964
561
556
residue
TET2 OGT
TET2 OGTmut
TET2 -
951
889
944
97
376
23
374
464
15
1749
39
1622
603
residue
TET3 OGT
TET3 OGTmut
TET3 -
362
667
665
557
1318
478
1351
1442
368
674
residue

Figure 4


TET1
TET2
TET3
T327
S950
T1327
S562 T625
T889
S944
S951
T1640
S361
T665
T667
S1256
T1293
T1404
T1658
S360
S361
S368
S362
N-terminus
CXXC-type zinc finger
Cystein-rich region
Low complexity insert
Dioxygenase domain
O-GlcNAc site
Phosphorylation site
persistent phosphorylation site
basal O-GlcNAc site

SUPPLEMENTAL DATA

Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the glycosyltransferase OGT [$ #]

Christina Bauer[1], Klaus Göbel[1], Nagarjuna Nagaraj[2], Christian Colantuoni[1], Mengxi Wang[1], Udo Müller[1], Elisabeth Kremmer[3], Andrea Rottach[1]*, Heinrich Leonhardt[1,4]*

LEGENDS

Supplemental data S1: Characterization of TET antibodies

a) TET antigens selected for immunization. The Figure shows a schematic outline of the TET proteins. The low complexity insert region of the TET proteins was chosen as antigen, and is highlighted in yellow. Amino acid numbers, sequence and tag of the respective antigens are shown in S2.

b) Generation of monoclonal TET antibodies using the hybridoma technology. Depicted is a typical workflow starting from the immunization of a Lou/C rat with the TET antigen to the final testing of the hybridoma supernatants in different applications. WB: western blot; IP: immunoprecipitation; IF: immunofluorescence; IH: immunohistology.

c) Stably subcloned anti-TET hybridoma clones and their application spectra.

d) Immunofluorescence staining of CGR8 embryonic stem cells using the TET1 5D6 antibody. In addition, cells were stained with an anti-mC antibody (mouse; Eurogentec, Köln, Germany) and a 5-hmC antibody (rabbit; Active Motif, Carlsbad, CA). DAPI was used as DNA counterstain. Merge images are shown in the lower panel. Scale bar represents 5 µm.

Supplemental data S2: Fragments of TET1, TET2, and TET3 used for immunization

Protein fragment sequences comprise a major part of the TET insert region. Amino acid numbers flanking the fragment are indicated. Purification of the antigen was performed via His-Tag (depicted in red).

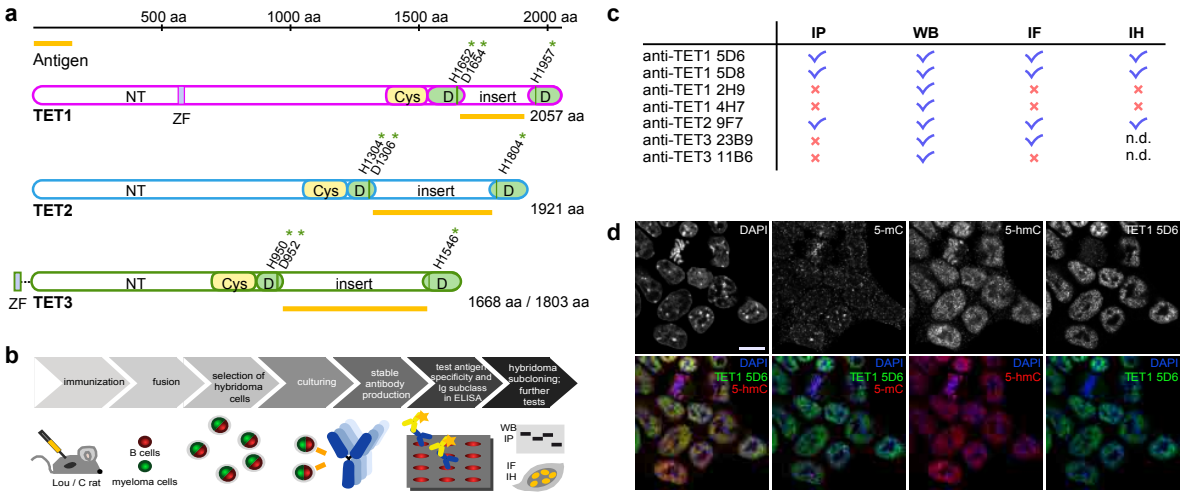
Supplemental data S3: TET protein sequences

Annotated amino acid sequence of the TET constructs used for PTM mapping experiments.
Uniprot entries: Q3URK3 (TET1), Q4JK59 (TET2), Q8BG87 (TET3)
green: catalytic residues (FeII binding)
orange: Phosphorylation
pink: GlcNAcylation
blue: alternative splicing (variation compared to Uniprot sequence)
grey: sequences not covered by MS (no information about PTMs available)

Supplemental table S4 (Excel file)

MaxQuant output tables of all detected TET peptides. Numbering of modified residues (column "positions within proteins") refers toUniprot TET protein sequence and might differ from the positions used in the rest of the paper which refer to the used splicing variants cloned from murine tissues as specified in the experimental procedures.

S1: Characterization of TET antibodies



S2: Fragments of TET1, TET2, and TET3 used for immunization

TET1_1682-1914 (254 aa)

MGHHHHHHHHHHSSGHIEGRHMQLHVLPLYRLADTDEFGSVEGMKAKIKSGAIQVNGPTRKRRLRFTE
PVPRCGKRAKMKQNHNKSGSHNTKSFSSASSTSHLVKDESTDFCPLQASSAETSTCTYSKTASGGFAE
TSSILHCTMPSGAHSGANAAAGECTGTVQPAEVAAHPHQSLPTADSPVHAEPLTSPSEQLTSNQSNQQ
LPLLSNSQKLASCQVEDERHPEADEPQHPEDDNLPQLDEFWSDSEEIYAD

TET2_1332-1779 (500 aa)

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFELRRPPEDEQFHVLPMYIIAPEDEFGSTEGQE
KKIRMGSIEVLQSFRRRRVIRIGELPKSCKKKAEPKKAKTKKAARKHSSLENCSSRTEKGKSSSHTKL
MENASHMKQMTAQPQLSGPVIRQPPTLQRHLQQGQRPQQPQPPQPQPQTTPQPQPQPQHIMPGNSQSV
GSHCSGSTSVYTRQPTPHSPYPSSAHTSDIYGDTNHVNFYPTSSHASGSYLNPSNYMNPYLGLLNQNN
QYAPFPYNGSVPVDNGSPFLGSYSPQAQSRDLHRYPNQDHLTNQNLPPIHTLHQQTFGDSPSKYLSYG
NQNMQRDAFTTNSTLKPNVHHLATFSPYPTPKMDSHFMGAASRSPYSHPHTDYKTSEHHLPSHTIYSY
TAAASGSSSSHAFHNKENDNIANGLSRVLPGFNHDRTASAQELLYSLTGSSQEKQPEVSGQDAAAVQE
IEYWSDSEHNFQDAAALEHHHHHH

TET3_976-1521 (584 aa)

MGHHHHHHHHHHSSGHIEGRHMQIPEDEQLHVLPLYKMASTDEFGSEENQNAKVSSGAIQVLTAFPRE
VRRLPEPAKSCRQRQLEARKAAAEKKKLQKEKLSTPEKIKQEALELAGVTTDPGLSLKGGLSQQSLKP
SLKVEPQNHFSSFKYSGNAVVESYSVLGSCRPSDPYSMSSVYSYHSRYAQPGLASVNGFHSKYTLPSF
GYYGFPSSNPVFPSQFLGPSAWGHGGSGGSFEKKPDLHALHNSLNPAYGGAEFAELPGQAVATDNHHP
IPHHQQPAYPGPKEYLLPKVPQLHPASRDPSPFAQSSSCYNRSIKQEPIDPLTQAESIPRDSAKMSRT
PLPEASQNGGPSHLWGQYSGGPSMSPKRTNSVGGNWGVFPPGESPTIVPDKLNSFGASCLTPSHFPES
QWGLFTGEGQQSAPHAGARLRGKPWSPCKFGNGTSALTGPSLTEKPWGMGTGDFNPALKGGPGFQDKL
WNPVKVEEGRIPTPGANPLDKAWQAFGMPLSSNEKLFGALKSEEKLWDPFSLEEGTAEEPPSKGVVKE
EKSGPTVEEDEEELWSDSEHNFLD

S3: TET protein sequences

TET1

```
MSRSRPAKPSKSVKTKLQKKKDIQMKTKTSKQAVRHGASAKAVNPGKPKQ 50
LIKRRDGKKETEDKTPTPAPSFLTRAGAARMNRDRNQVLFQNPDSLTCNG 100
FTMALRRTSLSWRLSQRPVVTPKPKKVPPSKKQCTHNIQDEPGVKHSEND 150
SVPSQHATVSPGTENGEQNRCLVEGESQEITQSCPVFEERIEDTQSCISA 200
SGNLEAEISWPLEGTHCEELLSHQTSDNECTSPQECAPLPQRSTSEVTSQ 250
KNTSNQLADLSSQVESIKLSDPSPNPTGSDHNGFPDSSFRIVPELDLKTC 300
MPLDESVYPTALIRFILAGSQPDVFDTKPQEKTLITTPEQVGSHPNQVLD 350
ATSVLGQAFSTLPLQWGFSGANLVQVEALGKGSDSPEDLGAITMLNQQET 400
VAMDMDRNATPDLPIFLPKPPNTVATYSSPLLGPEPHSSTSCGLEVQGAT 450
PILTLDSGHTPQLPPNPESSSVPLVIAANGTRAEKQFGTSLFPAVPQGFT 500
VAAENEVQHAPLDLTQGSQAAPSKLEGEISRVSITGSADVKATAMSMPVT 550
QASTSSPPCNSTPPMVERRKRKACGVCEPCQQKANCGECTYCKNRKNSHQ 600
ICKKRKCEVLKKKPEATSQAQVTKENKRPQREKKPKVLKTDFNNKPVNGP 650
KSESMDCSRRGHGEEEQRLDLITHPLENVRKNAGGMTGIEVEKWAPNKKS 700
HLAEGQVKGSCDANLTGVENPQPSEDDKQQTNPSPTFAQTIRNGMKNVHC 750
LPTDTHLPLNKLNHEEFSKALGNNSSKLLTDPSNCKDAMSVTTSGGECDH 800
LKGPRNTLLFQKPGLNCRSGAEPTIFNNHPNTHSAGSRPHPPEKVPNKEP 850
KDGSPVQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSESSSPSKPEKDE 900
EAHQKTASLLNSCKAILHSVRKDLQDPNVQGKGLHHDTVVFNGQNRTFKS 950
PDSFATNQALIKSQGYPSSPTAEKKGAAGGRAPFDGFENSHPLPIESHNL 1000
ENCSQVLSCDQNLSSHDPSCQDAPYSQIEEDVAAQLTQLASTINHINAEV 1050
RNAESTPESLVAKNTKQKHSQEKRMVHQKPPSSTQTKPSVPSAKPKKAQK 1100
KARATPHANKRKKKPPARSSQENDQKKQEQLAIEYSKMHDIWMSSKFQRF 1150
GQSSPRSFPVLLRNIPVFNQILKPVTQSKTPSQHNELFPPINQIKFTRNP 1200
ELAKEKVKVEPSDSLPTCQFKTESGGQTFAEPADNSQGQPMVSVNQEAHP 1250
LPQSPPSNQCANIMAGAAQTQFHLGAQENLVHQIPPPTLPGTSPDTLLPD 1300
PASILRKGKVLHFDGITVVTEKREAQTSSNGPLGPTTDSAQSEFKESIMD 1350
LLSKPAKNLIAGLKEQEAAPCDCDGGTQKEKGPYYTHLGAGPSVAAVREL 1400
METRFGQKGKAIRIEKIVFTGKEGKSSQGCPVAKWVIRRSGPEEKLICLV 1450
RERVDHHCSTAVIVVLILLWEGIPRLMADRLYKELTENLRSYSGHPTDRR 1500
CTLNKKRTCTCQGIDPKTCGASFSFGCSWSMYFNGCKFGRSENPRKFRLA 1550
PNYPLHNYYKRITGMSSEGSDVKTGWIIPDRKTLISREEKQLEKNLQELA 1600
TVLAPLYKQMAPVAYQNQVEYEEVAGDCRLGNEEGRPFSGVTCCMDFCAH 1650
SHKDIHNMHNGSTVVCTLIRADGRDTNCPEDEQLHVLPLYRLADTDEFGS 1700
VEGMKAKIKSGAIQVNGPTRKRRLRFTEPVPRCGKRAKMKQNHNKSGTAG 1750
LRRKRISASPKGAPGSHNTKSFSSASSTSHLVKDESTDFCPLQASSAETS 1800
TCTYSKTASGGFAETSSILHCTMPSGAHSGANAAAGECTGTVQPAEVAAH 1850
PHQSLPTADSPVHAEPLTSPSEQLTSNQSNQQLPLLSNSQKLASCQVEDE 1900
```

```
RHPEADEPQHPEDDNLPQLDEFWSDSEEIYADPSFGGVAIAPIHGSVLIE 1950
CARKELHATTSLRSPKRGVPFRVSLVFYQHKSLNKPNHGFDINKIKCKCK 2000
KVTKKKPADRECPDVSPEANLSHQIPSRVASTLTRDNVVTVSPYSLTHVA 2050
GPYNRWV* 2058
```

TET2

```
MEQDRTTHAEGTRLSPFLIAPPSPISHTEPLAVKLQNGSPLAERPHPEVN   50
GDTKWQSSQSCYGISHMKGSQSSHESPHEDRGYSRCLQNGGIKRTVSEPS  100
LSGLHPNKILKLDQKAKGESNIFEESQERNHGKSSRQPNVSGLSDNGEPV  150
TSTTQESSGADAFPTRNYNGVEIQVLNEQEGEKGRSVTLLKNKIVLMPNG  200
ATVSAHSEENTRGELLEKTQCYPDCVSIAVQSTASHVNTPSSQAAIELSH  250
EIPQPSLTSAQINFSQTSSLQLPPEPAAMVTKACDADNASKPAIVPGTCP  300
FQKAEHQQKSALDIGPSRAENKTIQGSMELFAEEYYPSSDRNLQASHGSS  350
EQYSKQKETNGAYFRQSSKFPKDSISPTTVTPPSQSLLAPRLVLQPPLEG  400
KGALNDVALEEHHDYPNRSNRTLLREGKIDHQPKTSSSQSLNPSVHTPNP  450
PLMLPEQHQNDCGSPSPEKSRKMSEYLMYYLPNHGHSGGLQEHSQYLMGH  500
REQEIPKDANGKQTQGSVQAAPGWIELKAPNLHEALHQTKRKDISLHSVL  550
HSQTGPVNQMSSKQSTGNVNMPGGFQRLPYLQKTAQPEQKAQMYQVQVNQ  600
GPSPGMGDQHLQFQKALYQECIPRTDPSSEAHPQAPSVPQYHFQQRVNPS  650
SDKHLSQQATETQRLSGFLQHTPQTQASQTPASQNSNFPQICQQQQQQQQ  700
QLQRKNKEQMPQTFSHLQGSNDKQREGSCFGQIKVEESFCVGNQYSKSSN  750
FQTHNNTQGGLEQVQNINKNFPYSKILTPNSSNLQILPSNDTHPACEREQ  800
ALHPVGSKTSNLQNMQYFPNNVTPNQDVHRCFQEQAQKPQQASSLQGLKD  850
RSQGESPAPPAEAAQQRYLVHNEAKALPVPEQGGSQTQTPPQKDTQKHAA  900
LRWLLLQKQEQQQTQQSQPGHNQMLRPIKTEPVSKPSSYRYPLSPPQENM  950
SSRIKQEISSPSRDNGQPKSIIETMEQHLKQFQLKSLCDYKALTLKSQKH 1000
VKVPTDIQAAESENHARAAEPQATKSTDCSVLDDVSESDTPGEQSQNGKC 1050
EGCNPDKDEAPYYTHLGAGPDVAAIRTLMEERYGEKGKAIRIEKVIYTGK 1100
EGKSSQGCPIAKWVYRRSSEEEKLLCLVRVRPNHTCETAVMVIAIMLWDG 1150
IPKLLASELYSELTDILGKCGICTNRRCSQNETKKKQSPPRNCCCQGENP 1200
ETCGASFSFGCSWSMYYNGCKFARSKKPRKFRLHGAEPKEEERLGSHLQN 1250
LATVIAPIYKKLAPDAYNNQVEFEHQAPDCCLGLKEGRPFSGVTACLDFS 1300
AHSHRDQQNMPNGSTVVVTLNREDNREVGAKPEDEQFHVLPMYIIAPEDE 1350
FGSTEGQEKKIRMGSIEVLQSFRRRRVIRIGELPKSCKKKAEPKKAKTKK 1400
AARKHSSLENCSSRTEKGKSSSHTKLMENASHMKQMTAQPQLSGPVIRQP 1450
PTLQRHLQQGQRPQQPQPPQPQPQTTPQPQPQPQHIMPGNSQSVGSHCSG 1500
STSVYTRQPTPHSPYPSSAHTSDIYGDTNHVNFYPTSSHASGSYLNPSNY 1550
MNPYLGLLNQNNQYAPFPYNGSVPVDNGSPFLGSYSPQAQSRDLHRYPNQ 1600
DHLTNQNLPPIHTLHQQTFGDSPSKYLSYGNQNMQRDAFTTNSTLKPNVH 1650
HLATFSPYPTPKMDSHFMGAASRSPYSHPHTDYKTSEHHLPSHTIYSYTA 1700
AASGSSSSHAFHNKENDNIANGLSRVLPGFNHDRTASAQELLYSLTGSSQ 1750
EKQPEVSGQDAAAVQEIEYWSDSEHNFQDPCIGGVAIAPTHGSILIECAK 1800
CEVHATTKVNDPDRNHPTRISLVLYRHKNLFLPKHCLALWEAKMAEKARK 1850
EEECGKNGSDHVSQKNHGKQEKREPTGPQEPSYLRFIQSLAENTGSVTTD 1900
STVTTSPYAFTQVTGPYNTFV*
```

TET3

```
MDSGPVYHGDSRQLSTSGAPVNGAREPAGPGLLGAAGPWRVDQKPDWEAA 50
SGPTHAARLEDAHDLVAFSAVAEAVSSYGALSTRLYETFNREMSREAGSN 100
GRGPRPESCSEGSEDLDTLQTALALARHGMKPPNCTCDGPECPDFLEWLE 150
GKIKSMAMEGGQGRPRLPGALPPSEAGLPAPSTRPPLLSSEVPQVPPLEG 200
LPLSQSALSIAKEKNISLQTAIAIEALTQLSSALPQPSHSTSQASCPLPE 250
ALSPSAPFRSPQSYLRAPSWPVVPPEEHPSFAPDSPAFPPATPRPEFSEA 300
WGTDTPPATPRNSWPVPRPSPDPMAELEQLLGSASDYIQSVFKRPEALPT 350
KPKVKVEAPSSSPAPVPSPISQREAPLLSSEPDTHQKAQTALQQHLHHKR 400
NLFLEQAQDASFPTSTEPQAPGWWAPPGSPAPRPPDKPPKEKKKKPPTPA 450
GGPVGAEKTTPGIKTSVRKPIQIKKSRSRDMQPLFLPVRQIVLEGLKPQA 500
SEGQAPLPAQLSVPPPASQGAASQSCATPLTPEPSLALFAPSPSGDSLLP 550
PTQEMRSPSPMVALQSGSTGGPLPPADDKLEELIRQFEAEFGDSFGLPGP 600
PSVPIQEPENQSTCLPAPESPFATRSPKKIKIESSGAVTVLSTTCFHSEE 650
GGQEATPTKAENPLTPTLSGFLESPLKYLDTPTKSLLDTPAKKAQSEFPT 700
CDCVEQIVEKDEGPYYTHLGSGPTVASIRELMEDRYGEKGKAIRIEKVIY 750
TGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVILILA 800
WEGIPRSLGDTLYQELTDTLRKYGNPTSRRCGLNDDRTCACQGKDPNTCG 850
ASFSFGCSWSMYFNGCKYARSKTPRKFRLTGDNPKEEEVLRNSFQDLATE 900
VAPLYKRLAPQAYQNQVTNEDVAIDCRLGLKEGRPFSGVTACMDFCAHAH 950
KDQHNLYNGCTVVCTLTKEDNRCVGQIPEDEQLHVLPLYKMASTDEFGSE 1000
ENQNAKVSSGAIQVLTAFPREVRRLPEPAKSCRQRQLEARKAAAEKKKLQ 1050
KEKLSTPEKIKQEALELAGVTTDPGLSLKGGLSQQSLKPSLKVEPQNHFS 1100
SFKYSGNAVVESYSVLGSCRPSDPYSMSSVYSYHSRYAQPGLASVNGFHS 1150
KYTLPSFGYYGFPSSNPVFPSQFLGPSAWGHGGSGGSFEKKPDLHALHNS 1200
LNPAYGGAEFAELPGQAVATDNHHPIPHHQQPAYPGPKEYLLPKVPQLHP 1250
ASRDPSPFAQSSSCYNRSIKQEPIDPLTQAESIPRDSAKMSRTPLPEASQ 1300
NGGPSHLWGQYSGGPSMSPKRTNSVGGNWGVFPPGESPTIVPDKLNSFGA 1350
SCLTPSHFPESQWGLFTGEGQQSAPHAGARLRGKPWSPCKFGNGTSALTG 1400
PSLTEKPWGMGTGDFNPALKGGPGFQDKLWNPVKVEEGRIPTPGANPLDK 1450
AWQAFGMPLSSNEKLFGALKSEEKLWDPFSLEEGTAEEPPSKGVVKEEKS 1500
GPTVEEDEEELWSDSEHNFLDENIGGVAVAPAHCSILIECARRELHATTP 1550
LKKPNRCHPTRISLVFYQHKNLNQPNHGLALWEAKMKQLAERARQRQEEA 1600
ARLGLGQQEAKLYGKKRKWGGAMVAEPQHKEKKGAIPTRQALAMPTDSAV 1650
TVSSYAYTKVTGPYSRWI* 1669
```

2.4 TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation

8592–8604 *Nucleic Acids Research, 2014, Vol. 42, No. 13* *Published online 19 June 2014*
doi: 10.1093/nar/gku552

TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation

Udo Müller, Christina Bauer, Michael Siegl, Andrea Rottach and Heinrich Leonhardt*

Department of Biology II, Ludwig-Maximilians University Munich and Center for Integrated Protein Science Munich (CIPSM), 82152 Planegg-Martinsried, Germany

Received March 28, 2014; Revised June 05, 2014; Accepted June 10, 2014

ABSTRACT

The discovery of hydroxymethyl-, formyl- and carboxylcytosine, generated through oxidation of methylcytosine by TET dioxygenases, raised the question how these modifications contribute to epigenetic regulation. As they are subjected to complex regulation *in vivo*, we dissected links to gene expression with *in vitro* modified reporter constructs. We used an Oct4 promoter-driven reporter gene and demonstrated that *in vitro* methylation causes gene silencing while subsequent oxidation with purified catalytic domain of TET1 leads to gene reactivation. To identify proteins involved in this pathway we screened for TET interacting factors and identified TDG, PARP1, XRCC1 and LIG3 that are involved in base-excision repair. Knockout and rescue experiments demonstrated that gene reactivation depended on the glycosylase TDG, but not MBD4, while NEIL1, 2 and 3 could partially rescue the loss of TDG. These results clearly show that oxidation of methylcytosine by TET dioxygenases and subsequent removal by TDG or NEIL glycosylases and the BER pathway results in reactivation of epigenetically silenced genes.

INTRODUCTION

DNA methylation at the C5-position of cytosine plays an essential role in a variety of fundamental processes, such as early embryonic development, X-chromosome inactivation, genome stability and imprinting (1,2). In vertebrates, this epigenetic modification is set by the three DNA methyltransferases DNMT1, DNMT3A and DNMT3B and the regulatory subunit DNMT3L (3–5).

Recently, it was discovered that the TET family of Fe(II)- and 2-oxoglutarate-dependent dioxygenases can successively convert 5-methylcytosine to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxylcytosine (caC) *in vitro* and *in vivo* (6–8). Three different TET proteins (TET1, TET2 and TET3), each showing tissue-specific differential expression (9), have been identified in mouse and human (10). Functional studies indicate that they are involved in a variety of cellular processes including epigenetic reprogramming, differentiation, myelopoiesis and imprinting (11–13). Mutations of *TET2* correlating with lower hmC levels and altered gene expression patterns have been linked to various hematopoietic malignancies (14,15).

The discovery of TET proteins and their catalytic products hmC, fC and caC has raised the question about the functions of these oxidized cytosine variants. They might serve as independent epigenetic signals and have been shown to recruit a distinct and dynamic set of 'reader' proteins in embryonic stem cells (ESCs) and differentiated cells (16). It also has been described that cytosine oxidation affects the efficiency of transcription by RNA polymerase II (17). However, the low abundance of fC and caC suggests that these cytosine variants are quickly processed *in vivo* and have been proposed to be intermediates in active DNA demethylation (18,19).

Whereas the mechanism of setting the methylation mark is well understood, the process of its removal has long been elusive. DNA demethylation may either occur by a passive process via the inhibition of DNMT1 maintenance methylation after replication (20–22) or by an active enzymatic reaction. In principle, there are three possibilities: first, the direct removal of the methyl group, second, the excision of either the methylated cytosine or third, of the entire nucleotide. It is currently proposed that the additional oxidized cytosine derivatives hmC, fC and caC are intermediates in active DNA demethylation, thereby contributing to epigenetic plasticity and transcriptional regulation (23,24).

Several biochemical studies revealed that thymine DNA glycosylase (TDG) can specifically bind to and excise fC and caC, resulting in abasic sites, which might be subsequently processed by the base-excision repair (BER) machinery (18,25). In general, the BER pathway repairs damaged DNA sites through recognition and excision of base lesions by substrate-specific glycosylases. The generated abasic site is subsequently cleaved by the AP endonuclease 1 (APEX1), leading to a single-strand break, which is recognized by PARP1 through its N-terminal zinc fingers. PARP1

*To whom correspondence should be addressed: Tel: +49 89 2180 74229; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de

then recruits XRCC1, LIG3 and DNA polymerase beta to complete the BER reactions (26–28). TDG depletion in mice causes embryonic lethality, and TDG deficient ESCs display prominent alterations of CpG modifications at a large number of gene regulatory regions (29,30).

Another discussed alternative for DNA demethylation is based on the initial deamination of hmC to hydroxymethyluracil (hmU) by members of the AID/APOBEC cytidine deaminase family (23). In the following step, hmU might be excised either by TDG, methyl-CpG-binding domain protein 4 (MBD4) or the single-strand-specific monofunctional uracil-DNA glycosylase 1 (SMUG1) (31–33). However, there is evidence that AID/APOBEC members are less active on modified cytosines *in vitro* or *in vivo*, challenging the prominence of the proposed deamination-linked demethylation pathway in living cells (34). Furthermore, a direct decarboxylation of caC to unmodified cytosine has been detected in ESC lysates, but no specific decarboxylase has been identified so far (35).

In addition to TDG, two members of the NEIL family of glycosylases (NEIL1 and NEIL3) have recently been identified as potential binders for oxidized cytosine derivatives (16). However, their function in TET-dependent demethylation has not been investigated to date.

To unravel the effects of TET-mediated cytosine oxidation on gene expression, we generated *in vitro* modified pOct4-reporter plasmids and monitored their *in vivo* expression in ESCs. Whereas methylation of the reporter DNA leads to silencing of gene expression, subsequent oxidation results in gene reactivation. We show that TET proteins interact with BER factors *in vivo* and propose that the observed oxidation-dependent gene reactivation requires the BER machinery. We demonstrate that initiation of this pathway is mainly dependent on TDG activity, but not on MBD4. Our results also indicate that the glycosylases NEIL1, NEIL2 and NEIL3 can contribute to an alternative BER pathway for DNA demethylation and cause gene reactivation.

MATERIALS AND METHODS

Cell culture and transfection

Human embryonic kidney 293T (HEK293T) and baby hamster kidney (BHK) cells containing a stably integrated lac operator array (36) were cultured in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% Fetal Calf Serum (FCS) (Biochrom) and 50 μg/μl gentamycine (PAA). HEK293T and BHK cells were transiently transfected using polyethylenimine pH 7.0 (Sigma) according to the manufacturer's instructions.

Mouse wild-type (wt) E14 as well as *Tdg*-/- and *Mbd4*-/- ESCs (29,37) were cultured on gelatin coated flasks or optical 96-well plates (Greiner) using 1000 U/ml LIF, 1μM PD032591 and 3 μM CHIR99021 (Axon Medchem, (38)). ESCs were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

In vitro methylation and oxidation of plasmid DNA

In vitro methylation of pOct4-GFP plasmid DNA was performed using M.SssI methyltransferase (New England Biolabs) according to the manufacturer's instructions. The methylation status of the plasmid was tested by HpaII and MspI (Fermentas) digestion.

For the *in vitro* oxidation, GFP-TET1CD or GFP-TET1CD^{mut} (H1652Y, D1654A) was purified from mammalian cells. In detail, HEK293T cells were transfected with an expression construct for GFP-TET1CD/TET1CD^{mut} and immunoprecipitation was carried out using GBP-Ni-NTA beads. Proteins were eluted using imidazole. The *in vitro* methylated plasmid was diluted in TET reaction buffer (50 mM HEPES pH 8.0, 75 μM Fe(II), 2 mM Sodium-Ascorbate, 1 mM Di-Sodium-Ketoglutarate (39)) and added to the purified GFP-TET1CD.

Digestion of hydroxymethylated plasmid with PvuRts1I

A total of 200 ng oxidized plasmid DNA and 100 ng of reference DNA fragments containing exclusively unmodified C, mC or hmC were digested with PvuRts1I (150 mM NaCl, 20 mM Tris pH 8.0, 5 mM $MgCl_2$, 1 mM DTT) at 22°C for 20 min (40). The reaction was inactivated at 65°C for 10 min and digestion of the samples was analyzed by agarose-gelelectrophoresis.

Co-immunoprecipitation (Co-IP) using the GFP-Trap

Note that 36 h after transfection, whole cell lysates of HEK293T cells were prepared using RIPA-lysis buffer (50 mM Tris pH 7.0, 150 mM NaCl, 0.1% sodium dodecyl sulphate (SDS), 0,5% sodium deoxycholate, 5 mM ethylenediaminetetraacetic acid (EDTA), 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 2 mM PMSF, 1x Mammalian Protease Inhibitor Cocktail and 1 μg/μl DNaseI). After centrifugation, 10% of the supernatant was collected as input fraction and the remaining supernatant was diluted in IP-buffer (10 mM Tris/Cl pH 7.5, 150 mM NaCl and 0.5 mM EDTA) to 800 μl. Green Fluorescent Protein (GFP)-Trap beads (Chromotek, (41)) pre-equlibrated with IP-buffer were added to the supernatant dilution and rotated for 1.5 h at 4°C. The GFP-beads were centrifuged and 10% of the supernatant was collected as flowthrough fraction. For western blotting, input, flowthrough and bead fractions were boiled with Laemmli buffer at 95°C for 10 min, loaded on an SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane (Biorad). Immunodetection was performed using mouse monoclonal anti-GFP (Roche, 11814460001) or rat monoclonal anti-RFP antibodies (42) and Alexa488/Alexa594 coupled secondary antibodies (Jackson ImmunoResearch).

For mass spectrometry analysis, protein samples were denatured with GdnHCl, precipitated with acetone and digested with trypsin. Peptide mixtures were analyzed using electrospray tandem mass spectrometry in collaboration with the Core Facility of the Max-Planck-Institute for Biochemistry, Martinsried. Experiments were performed with an LTQ Orbitrap mass spectrometer (Thermo Scientific). Spectra were analyzed with MaxQuant (43).

Fluorescent-three-hybrid assay (F3H)

Transgenic BHK cells containing stably integrated lac-operator repeats (36) were grown to 60–70% confluence

on coverslips. The cells were transiently cotransfected with expression constructs encoding for LacI-GBP, murine RFP/mCherry- and GFP-fusion proteins (44). As controls the catalytically inactive mutants GFP-TET1$^{H1652Y\&D1654A}$, GFP-TET2$^{H1304Y\&D1306A}$ and GFP-TET3$^{H950Y\&D952A}$ were used. Note that 24 h after transfection cells were fixed with 4% formaldehyde in phosphate buffered saline (PBS), permeabilized with 0.5% Triton X-100/PBST, counterstained with 4',6-diamidino-2-phenylindole (DAPI) and mounted in Vectashield (Vector Laboratories). Imaging was performed using a Leica TCS SP5 confocal laser scanning microscope with a 63x/1.4 NA Plan-Apochromat oil immersion objective.

The Operetta automated imaging system (PerkinElmer) was used for F3H quantification (Harmony 3.6 software). After imaging, nuclei were detected based on DAPI signal. The lacO-spot was defined in the GFP channel and screened for enrichment at the RFP channel (intensity spot >1.2x mean intensity nucleus; see also Supplementary Figure S4).

High-throughput pOct4-reporter gene expression analysis

Wild type, *Tdg*-/- *and Mbd4*-/- ESCs were transiently transfected with unmodified, M.SssI methylated or *in vitro* oxidized pOct4-GFP plasmid DNA on coated optical 96-well plates (PerkinElmer). Note that 24 h after transfection, the cells were fixed with 4% formaldehyde/PBS, permeabilized with 0.5% Triton X-100/PBST and counterstained with DAPI.

Images were acquired using the Operetta automated imaging system with an 40x high NA objective and expression was quantified using Harmony 3.6 software (PerkinElmer). A total of 16 fields per well were imaged, cells were counted and segmented into nuclei and cytoplasm on the basis of DAPI and reporter mCherry/GFP signal. Prism software (GraphPad) was used for statistical analysis.

Generation of stable transgenic cell lines

Tdg-/- ESCs stably expressing GFP-fusions of wt TDG, TDG^{N151A}, TDG^{N168D} and TDG^{M280H} were generated by transfecting the respective plasmids in the presence of the selection marker blasticidine followed by repeated sorting for GFP expression with the fluorescence-activated cell sorting (FACS) AriaII (Becton-Dickinson) system. Single cell sorting was used to generate clonal transgenic cell lines. GFP-expression of the single cell clones was analyzed using the Operetta system or western blotting.

Activity of GFP-TDG *in vitro*

GFP-TDG and the different mutants were expressed in HEK293T cells and immunoprecipitated as described above. Equal amounts of GFP-tagged protein immobilized on GFP-Trap beads were incubated with 0.4 μM of DNA substrate in TDG reaction buffer (20 mM TrisHCl, pH 8.0 or pH 6.5, 100 mM NaCl, 1 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT, 0.01 mM $ZnCl_2$). In detail, these DNA substrates were fluorescently (ATTO550) labeled 42 bp oligonucleotides (GGA TGA TGA CTC TTC TGG TCC GGA TGG TAG TTA AGT GTT GAG) (Eurofins MWG Operon) with a central modified CpG site: either hmCpG, fCpG or caCpG or harboring a G◦T mismatch at this site. Incubation was performed in the presence of purified GFP-APEX1 for 2 h at room temperature. Following heat-inactivation of TDG (2 min, 95°C), fresh GFP-APEX1 was added and further incubated for 4 h at room temperature. An oligonucleotide containing a deoxyabasic site ('dSpacer', Eurofins MWG Operon) served as a positive control for APEX1 activity. DNA was analyzed on a denaturing 17% polyacrylamide gel with the Typhoon TRIO (GE Healthcare Life Sciences). Quantification was performed with ImageJ.

Genomic DNA extraction and slot blot analysis

Genomic DNA from ESCs was extracted using the Blood & Cell Culture Midi Kit (Qiagen) according to the manufacturer's instruction. Anti-oxidant BHT (200 μM, Sigma) and deaminase inhibitor THU (200 μM, Sigma) were added to the lysis buffer and elution buffer. The Bio-Rad slot blot system was used according to the manufacturer's instruction. Nitrocellulose membranes (Amersham) were crosslinked, blocked with 5% milk and immunostaining was performed using a mouse monoclonal antibody against mC (Eurogentec, 33D3) or rabbit polyclonal antibodies against hmC, fC and caC (Active motif: 39791, 61233, 61224). Alexa488-coupled secondary antibodies were used for detection and the membranes were scanned with the Typhoon TRIO (GE Healthcare Life Sciences). Quantification was performed with ImageJ.

Re-isolation of transfected plasmids

Note that 36 h after transfection, nuclei were extracted from the ESCs using the Blood & Cell Culture Midi Kit (Qiagen) according to the manufacturer's instructions. Plasmid DNA was re-isolated using the Qiaprep Plasmid Midi Kit (Qiagen). A total of 200 ng of re-isolated plasmid DNA was digested with 0.5 μl HpaII (Fermentas).

RESULTS

In vitro oxidation of Oct4 reporter plasmid DNA causes gene activation

Since the discovery of hmC, fC and caC, two major roles for these cytosine modifications have been proposed: first, as intermediates in active DNA demethylation, and second, as independent epigenetic marks. The latter has been investigated by identifying reader proteins for hmC, fC and caC in different murine tissues. These new DNA modifications recruit a dynamic set of readers including DNA repair factors and chromatin remodelers (16). TET-dependent cytosine oxidation has been shown to occur at a large number of gene regulatory elements and repetitive loci (30).

Here, we focus on the effects of hmC, fC and caC on gene expression. We transfected ESCs with Oct4 promoter-driven GFP and mCherry reporter plasmids (pOct4-GFP or pOct4-mCherry) in different modification states and analyzed expression of the reporter by quantitative imaging. To generate the different cytosine modifications *in*

vitro, the unmodified pOct4-reporter construct (CpG) was initially treated with the DNA-methyltransferase M.SssI, thereby creating fully methylated CpG sites (^{m}CpG), and subsequently incubated with the purified catalytic domain of TET1 (TET1CD) to create oxidized cytosine residues (^{ox}CpG; Figure 1a). Specific restriction enzymes were used to monitor the methylation and oxidation state of the plasmid DNA. The *in vitro* methylation of the reporter construct was confirmed by the mC-sensitive restriction endonuclease HpaII and the mC-insensitive enzyme MspI. Both MspI and HpaII fully cleave the unmethylated pOct4-reporter plasmid at CCGG sites, whereas the restriction pattern of the M.SssI-methylated reporter only appeared with MspI digestion (Figure 1b).

To follow the oxidation of mC to hmC by TET1CD, the reporter DNA was treated with the hmC-specific endonuclease PvuRts1I (40) (Supplementary Figure S1a). Treatment with TET1CD resulted in a gradual increase of hmC levels after 15, 45 and 90 min of incubation, visible as progressing fragmentation (Figure 1b). While using a catalytically inactive TET1CD mutant (TET1CD^{mut}) as a control, no hmC levels were detected (^{ox}*CpG), indicating specific enzymatic mC oxidation by TET1CD (Supplementary Figure S1b).

Consistent with the restriction digest results, methylation and oxidation of the reporter plasmid could also be shown by slot blot analysis (Figure 1c). Incubation of methylated plasmid DNA with TET1CD resulted in an increase of not only hmC but also of fC and caC, demonstrating that purified TET1CD did carry out the three oxidation steps *in vitro* (Figure 1c and d). The mC signal decreased over time as hmC, fC and caC appeared in the presence of active TET1CD, while remaining constantly high with TET1CD^{mut} (Supplementary Figure S1d).

Transfection of ESCs with the TET1CD-treated plasmid DNA allows to directly monitor the effect of the oxidized cytosines on gene expression, independent from *in vivo* TET activity. Reporter gene expression from either unmodified, *in vitro* methylated or oxidized pOct4-mCherry was visualized using confocal imaging or automated image acquisition for quantification. Transfection of ESCs with unmodified pOct4-mCherry resulted in a strong nuclear and cytoplasmic expression of the reporter, whereas expression drastically decreased when using the methylated construct. Interestingly, prominent reporter expression could be observed upon transfection of the oxidized plasmid DNA, but not with the ^{ox}*CpG reporter DNA (Figure 1e). This suggests that reactivation of gene expression requires oxidation of methylcytosine by TET proteins.

TET proteins interact with the BER machinery

Currently, three pathways for TET-mediated active DNA demethylation are discussed: TDG-dependent BER, deamination-dependent BER and direct decarboxylation of caC (18,23,35). Since the proteins responsible for gene reactivation in our assay might be physical interaction partners of the TET proteins, we performed an initial unbiased screen for interactors. We expressed GFP-TET1 in HEK293T cells, performed immunoprecipitation and analyzed the co-precipitated proteins by mass spectrometry. Prominently, we found PARP1, XRCC1 and LIG3, a subset of proteins involved in the BER pathway (Supplementary Figure S2a). These results point toward the two BER-dependent demethylation mechanisms.

To further investigate the interplay between TET proteins and BER, we systematically performed Co-IP and a recently established F3H assay of all three TET proteins with the following BER factors: TDG, MBD4, SMUG1, NEIL1, NEIL2, NEIL3, PARP1, LIG3 and XRCC1. F3H allows to directly visualize protein–protein interactions in living cells (44). We therefore co-expressed full-length GFP-TET fusion proteins together with potential mCherry/RFP-tagged interactors (Supplementary Figure S3a) and LacI-GBP in BHK cells containing a stably integrated lac-operator array (45). The GFP-tagged bait is enriched at the lac-operator array via LacI-GBP and is visible as a single spot inside the nucleus. If the mCherry/RFP-tagged prey protein interacts with the bait, it colocalizes at the same spot (Figure 2a).

Consistent with the mass spectrometry results, interactions of TET1, TET2 and TET3 were observed with PARP1, LIG3 and XRCC1 both in Co-IP and F3H. Interestingly, several glycosylases also showed a clear interaction in both assays, among them TDG, MBD4, NEIL1, NEIL2 and NEIL3, but not SMUG1 (Figure 2b and c; Supplementary Figures S2b and S3b, c). Automated high-throughput image analysis was used to quantify the F3H results (Supplementary Figure S4). The interaction of all three TET proteins with TDG was the most robust and detectable in more than 80% of all analyzed cells. For the other factors, numbers vary between 40% and 75% (Figure 2d).

To exclude that the recruitment to the lacO-spot is dependent on the locally enriched cytosine oxidation products generated by TET proteins, we repeated the F3H quantification with catalytically inactive TET mutants (Figure 2e). The percentage of cells showing an interaction did hardly differ compared to the assay with active TET proteins. The only exception is XRCC1 where only half as many cells displayed a colocalization at the spot, suggesting a potential cooperative effect.

Taken together, these results suggest that TET proteins physically interact with the BER machinery and are therefore able to recruit these factors to the site of cytosine oxidation for immediate removal of the modified base *in vivo*.

TDG but not MBD4 mediates oxidation-dependent gene reactivation in ESCs

Since DNA glycosylases catalyze the first step of the BER pathway and are therefore the initiators of TET-dependent cytosine demethylation, we investigated their role in gene reactivation with the reporter gene assay. Besides TDG, also MBD4 has been implicated in DNA demethylation via excision of hmU, the deamination product of hmC (46). To investigate the role of these two glycosylases, we transfected *Tdg*-/- and *Mbd4*-/- ESCs (29,37) with either unmodified, *in vitro* methylated or oxidized pOct4-mCherry plasmid DNA. In contrast to wt E14 ESCs, *Tdg*-/- ESCs showed no reporter gene expression from the oxidized plasmid. However, *Mbd4* knockouts were able to fully reactivate gene expression from the oxidized reporter construct.

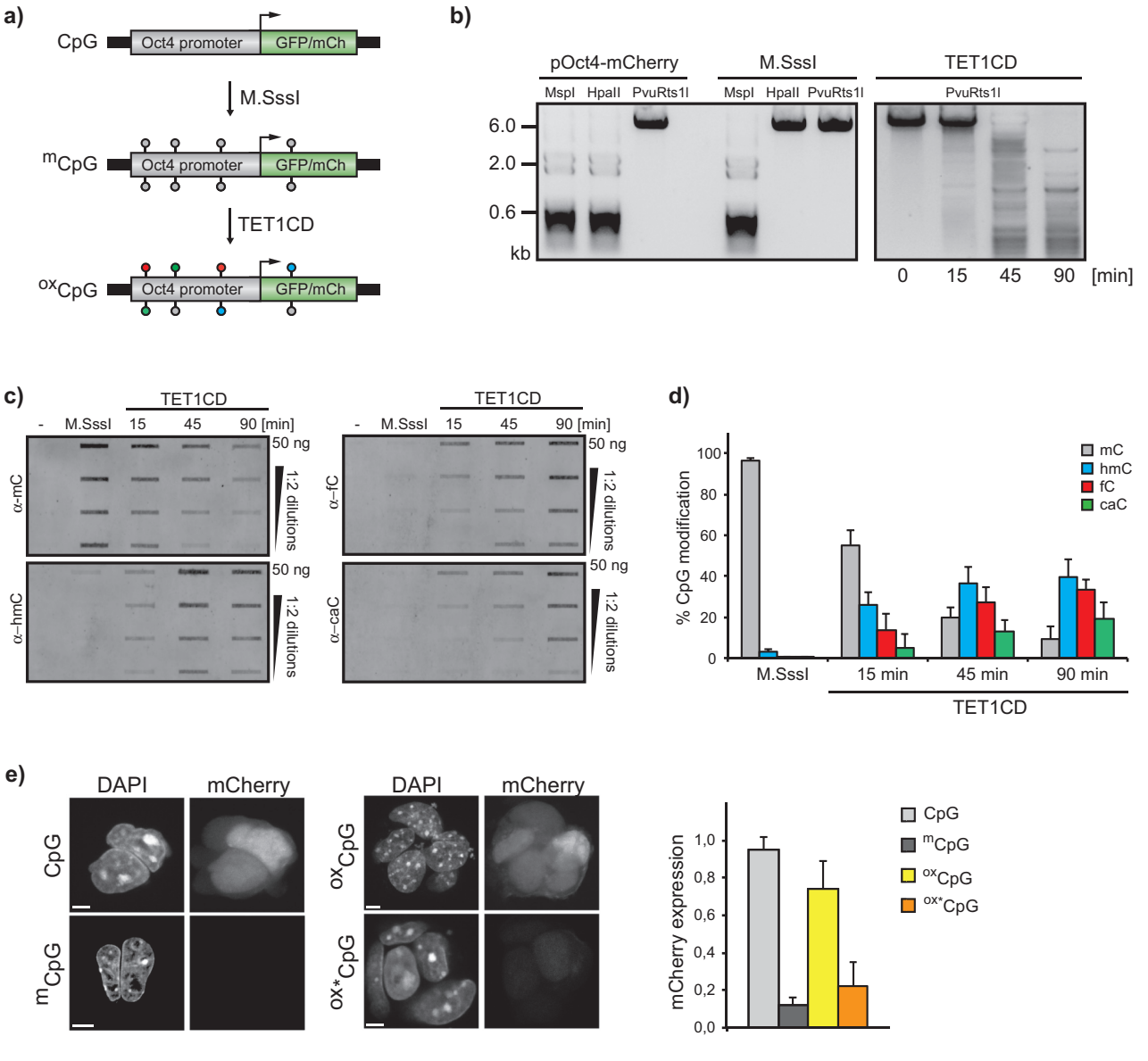


Figure 1. *In vitro* oxidation of mC causes gene reactivation in ESCs. (**a**) Schematic representation of *in vitro* reporter DNA modification: Unmethylated pOct4-reporter DNA was methylated using the CpG methyltransferase M.SssI. Incubation with purified TET1CD results in oxidation of mC sites to hmC, fC and caC. (**b**) M.SssI treatment of pOct4-mCherry results in full methylation as shown after restriction with the methylation sensitive enzyme HpaII. MspI cuts irrespective of the methylation state. The hmC-specific restriction endonuclease PvuRts1I detects increasing hmC levels during incubation of methylated pOct4-mCherry with TET1CD. (**c**) Cytosine modification states of untreated, methylated and TET1CD oxidized pOct4-mCherry plasmid DNA were detected by slot blot. A 2-fold serial dilution of the plasmid DNA was loaded and detected using antibodies against mC, hmC, fC and caC. A gradual increase of hmC, fC and caC signals was obtained with longer incubation time with TET1CD while the mC signal decreases accordingly. (**d**) Quantification of the slot blot signals of pOct4-mCherry after treatment with TET1CD shows increasing oxidation of mC to hmC, fC and caC. The sum of all CpG modification signals was set to 100%. Error bars indicate standard deviation ($n = 3$). (**e**) ESCs were transfected with pOct4-mCherry plasmids containing either unmodified (CpG), methylated (mCpG), TET1CD-oxidized (^{ox}CpG) or TET1CD^{mut}-treated (^{ox}*CpG) cytosines. Confocal imaging and quantification show reporter gene silencing upon methylation and reactivation upon oxidation. Cells were fixed with formaldehyde and counterstained with DAPI. Scale bar: 5 μm. (Right: $n = 200\,000$; error bars indicate standard deviation).

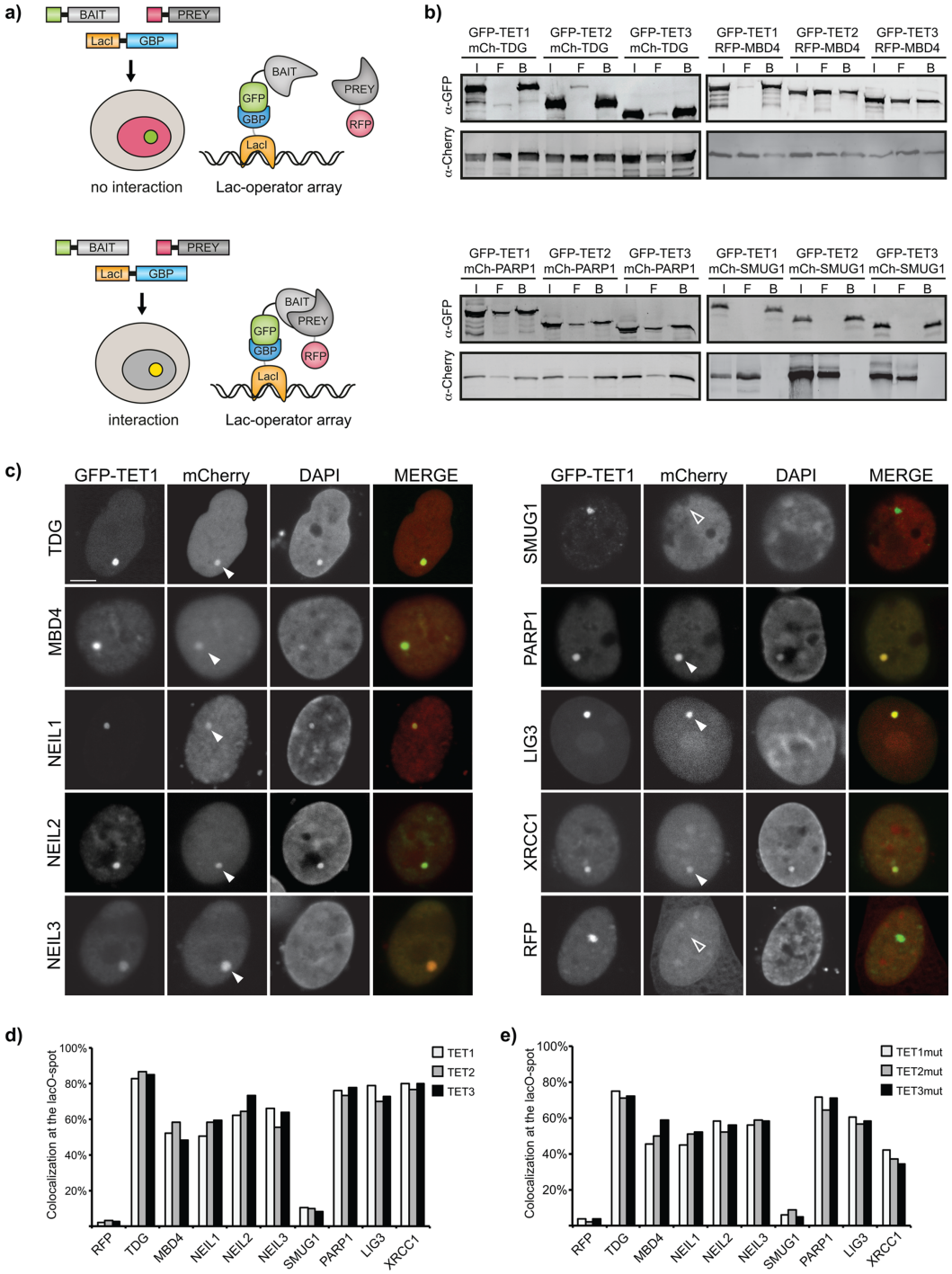


Figure 2. All three TET proteins interact with a variety of BER factors. (**a**) Scheme depicting the F3H assay for *in vivo* visualization of protein interactions: BHK cells containing a stably integrated lac-operator array were transfected with plasmids expressing a lac-repressor-GBP fusion protein, GFP-BAIT and mCherry/RFP-PREY. The Lac-repressor binds to the lac-operator array and recruits the GFP-BAIT through GBP. Proteins interacting with the BAIT are consequently enriched at the lac-operator array. (**b**) Co-IP with subsequent SDS-PAGE and western blotting shows interaction of the three TET proteins with TDG, PARP1, MBD4 but not SMUG1. GFP-tagged TET proteins and the respective mCherry-fusions were expressed in HEK293T cells and immunoprecipitated with the GFP-Trap (I: Input; F: Flowthrough; B: Bound). (**c**) F3H was used to confirm TET1 interactions with different factors involved in BER. GFP-TET1 is enriched at the lac-operator array and mCherry-tagged interacting factors are recruited to the same spot (solid triangle). SMUG1 shows no co-localization at the lacO-spot (empty triangle). Scale bar: 5 μm. (**d**) Quantification of the F3H assay of all three TET proteins with the indicated BER factors. Bars represent percentage of cells that show colocalization at the lacO-spot (n>200). (**e**) As in (**d**) but with catalytically inactive TET mutants.

The rescue experiment with wt GFP-TDG re-established the ability of *Tdg*-/- to express mCherry from the ^{ox}CpG plasmid, whereas the transient rescue of *Mbd4*-/- cells with GFP-MBD4 led to no significant differences with regard to reporter gene expression (Figure 3a).

High-throughput image analysis of 200 000 cells revealed that the signal from the oxidized plasmid in wt E14 ESCs is about 80% of the signal from unmodified reporter DNA. In *Tdg*-/- cells, the ^{ox}CpG signal drops almost to the level of the fully methylated reporter. Thus, the ^{ox}CpG construct remains silent in *Tdg*-/- ESCs. The stable rescue with wt GFP-TDG led to a recovery of mCherry expression to about 70% of the signal from the unmodified reporter. Knockout of *Mbd4* and also the corresponding rescue did not alter expression levels compared to wt E14 ESCs (Figure 3b), indicating that oxidation-dependent gene reactivation requires TDG but not MBD4 expression.

To gain insight into the mechanism by which TDG mediates gene activation, we recovered the transfected reporter plasmid DNA from wt E14 ESCs, *Tdg*-/- and *Mbd4*-/- ESCs. We analyzed the modification status by digestion with HpaII, which specifically cuts unmodified cytosines in a CCGG context (Supplementary Figure S1c). We observed a reappearance of the HpaII restriction pattern on the ^{ox}CpG plasmids isolated from wt E14 and *Mbd4*-/- ESCs indicating that TET-dependent demethylation occurred *in vivo*. HpaII only displayed minor activity on ^{ox}CpG reporter DNA re-isolated from *Tdg*-/- cells (Figure 3c). This provides strong evidence that the substitution of oxidized cytosine by unmodified cytosine is the major mechanism for the observed gene reactivation and that this substitution depends on TDG.

TDG activity is required for reporter gene reactivation

To investigate whether the glycosylase activity of TDG is responsible for the observed gene reactivation, we generated *Tdg*-/- rescue cell lines, which stably express GFP-fusions of either wt, catalytically inactive (N151A), DNA binding deficient (M280H) or caC-specific (N168D) TDG (31,47) at equal levels (Supplementary Figure S5a). The murine caC-specific mutant corresponds to a published human TDG mutant (48). Expression of the oxidized mCherry-reporter plasmid in the stable rescue with wt GFP-TDG increased almost to the levels of wt E14 ESCs in line with the results from the transient rescues. In contrast, the catalytically inactive mutant (N151A) or DNA binding deficient (M280H) TDG was not able to recover reporter expression. Cells stably expressing the caC-specific TDG (N168D) were only capable of partially restoring reporter expression (Figure 4a and b; Supplementary Figure S5b). This demonstrates that base excision by active TDG is essential for gene reactivation.

To further characterize the activity of wt TDG as well as the TDG mutants, we established an *in vitro* assay based on a defined DNA substrate with a single modification site. Base excision by TDG generates an abasic site, which can specifically be converted into a single-strand break by purified APEX1 and can be detected on a denaturing gel. Thus, this assay mimics the first two steps of the BER reactions and is also applicable on fluorescently labeled DNA substrates, in contrast to the previously described 'nicking assay', in which alkaline treatment and subsequent boiling is used to create single-strand breaks (49,50).

TDG has long been known to repair G○T mismatches (51) and was recently found to excise fC and caC (18,25). We could show that wt TDG is more active on fC and caC than on G○T mismatches, its eponymous substrate. No activity was detected on hmC-containing DNA. Although the caC-specific TDG mutant (N168D) was able to partially reactivate reporter gene expression *in vivo*, we could only detect basal activity on the caC substrate *in vitro* (Figure 4c, Supplementary Figure S5d). Since the activity of TDG *in vitro* is pH-dependent (48), we repeated the assay at pH = 6.5. Under these conditions, the preference of the caC-specific TDG mutant (N168D) toward caC could be confirmed (Supplementary Figure S5d). Taken together, these data suggest that the excision of fC and caC by TDG is essential for TET-mediated demethylation and causes reactivation of gene expression.

NEIL1, NEIL2 and NEIL3 glycosylases can partially compensate for loss of TDG

Besides TDG, we identified the family of NEIL glycosylases as interactors of TET proteins (Figure 2c). Interestingly, NEIL1 and NEIL3 have also been described as binders of hmC, fC or caC cytosines in ESCs in a proteome wide screen (16). However, their function in this context has not been investigated so far.

To elucidate whether the NEIL glycosylase family contributes to gene reactivation, we measured the expression of the modified reporter plasmids in *Tdg*-/- cells transiently overexpressing NEIL1, NEIL2 or NEIL3 at similar levels (Figure 5a). Interestingly, we observed a significant increase of pOct4-GFP expression in *Tdg*-/- ESCs rescued with mCherry-NEIL1, NEIL2 or NEIL3 in comparison to *Tdg*-/- ESCs. However, lower expression levels as in rescues with wt TDG were detected. RFP-MBD4 was not able to rescue the *Tdg*-/- phenotype (Figure 5b).

Additionally, we isolated genomic DNA from wt E14 and *Tdg*-/- ESCs as well as from the transient rescues with wt TDG and NEIL1, 2 and 3. Slot blot analyses were carried out for relative hmC, fC and caC quantifications. Genomic hmC was present at comparable levels in all tested cell lines and was not affected by *Tdg* knockout or NEIL overexpression. Since TDG is able to recognize and excise fC and caC, accumulation of these oxidized bases was observed in *Tdg*-/- ESCs, consistent with previous reports (30,52). Rescues of *Tdg*-/- cells with transiently expressed wt TDG, NEIL1, NEIL2 or NEIL3 resulted in decreased genomic fC, and for wt TDG and NEIL1, also in decreased caC levels (Figure 5c). These findings support the role of the glycosylases TDG, NEIL1, NEIL2 and NEIL3 in active DNA demethylation and subsequent reactivation of gene expression via excision of fC and caC followed by BER (Figure 6).

DISCUSSION

In this study, we investigated the effects of the oxidized cytosine variants hmC, fC and caC on gene expression. By carrying out the enzymatic oxidation of a methylated

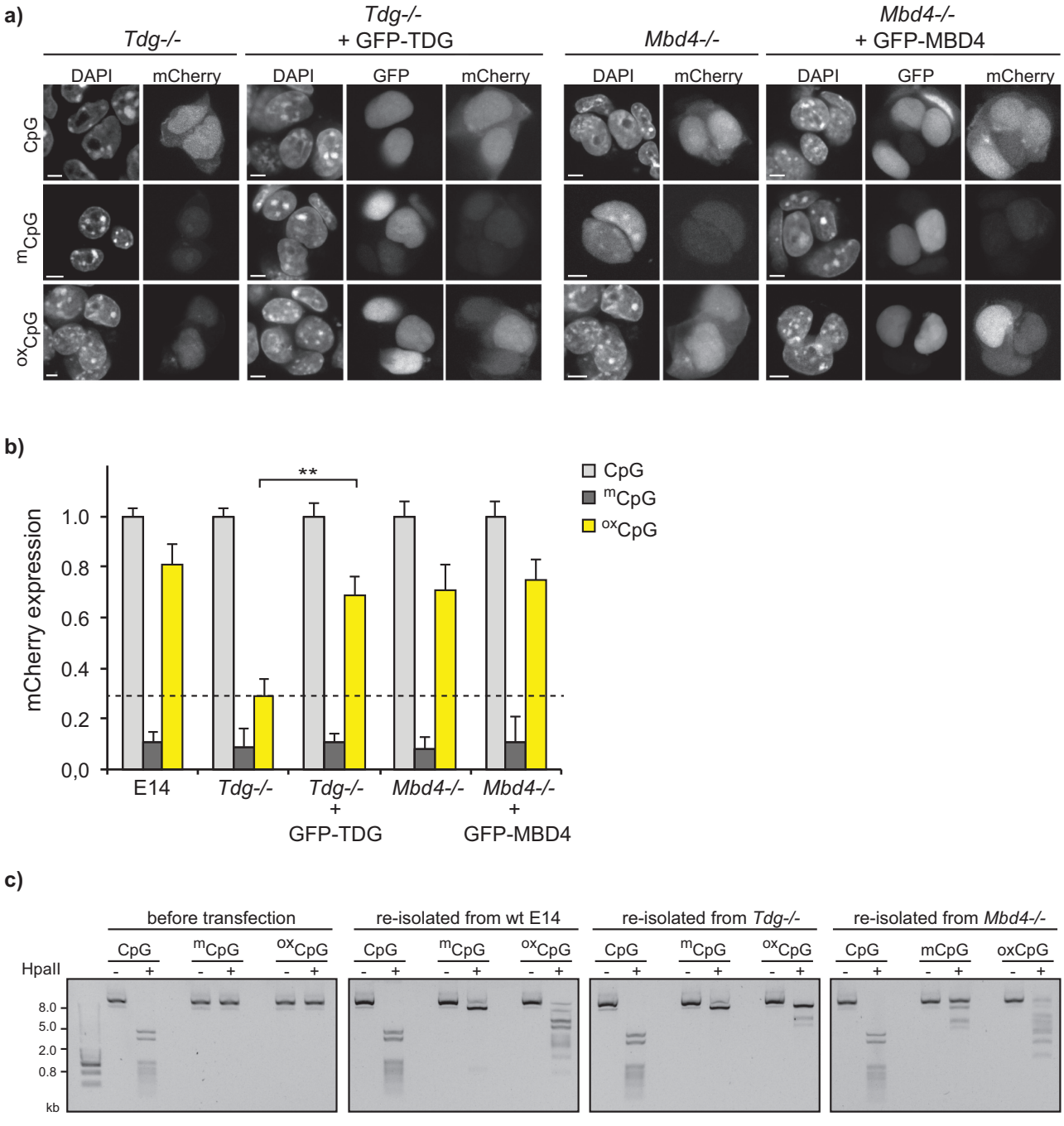


Figure 3. Oxidation of ^{m}CpG plasmid DNA leads to TDG-dependent gene reactivation. (**a**) *Tdg*-/- ESCs were transfected with pOct4-mCherry plasmids containing either unmodified, methylated or oxidized CpGs. Confocal images show a defect of ^{ox}CpG gene reactivation in *Tdg*-/- ESCs but not in *Mbd4*-/- ESCs. Transient rescue of *Tdg*-/- ESCs with GFP-TDG re-establishes ^{ox}CpG reporter gene expression. Cells were fixed with formaldehyde and counterstained with DAPI. Scale bar: 5 μm. (**b**) High-throughput image acquisition and quantification of pOct4-mCherry expression shows that oxidation of ^{m}CpG sites in the pOct4-reporter results in reactivation of mCherry-expression in wt E14 ESCs and *Mbd4*-/- ESCs, but not in *Tdg*-/- ESCs. Expression of GFP-TDG rescues the phenotype (student's *t*-test, $^{**}P < 0.025$, $n = 200\,000$; error bars indicate standard deviation). (**c**) Analytical digest with HpaII of differentially modified reporter plasmid DNA before and after transfection confirms substitution of ^{ox}CpG with CpG in wt E14 and *Mbd4*-/- ESCs.

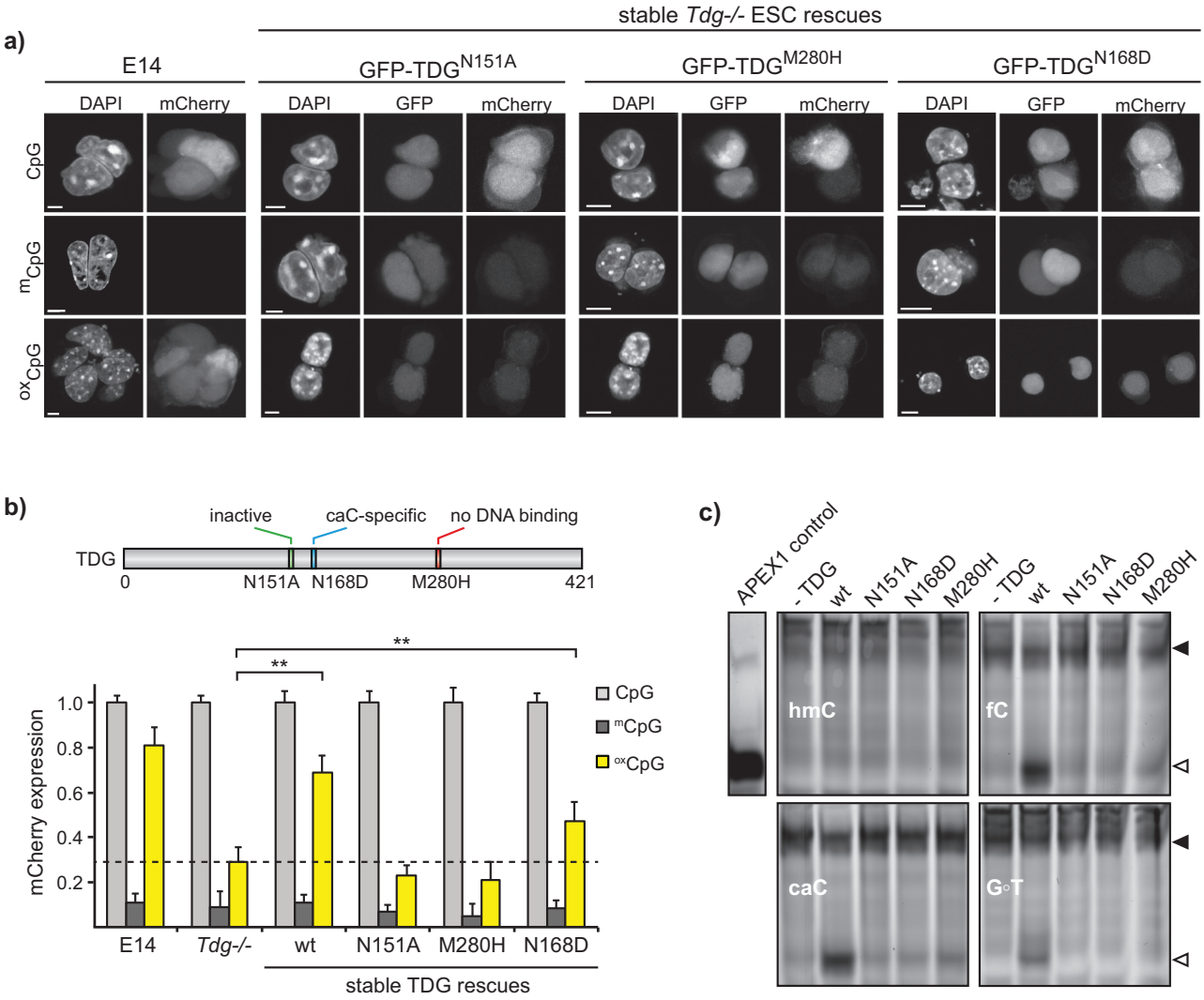


Figure 4. TDG activity is essential for gene reactivation. (**a**) Confocal images depicting expression levels of oxidized pOct4-mCherry expression in *Tdg*-/- ESCs stably rescued with GFP-TDG^{N151A}, GFP-TDG^{M280H} and GFP-TDG^{N168D} in comparison to wt E14 ESCs. Scale bar: 5 μm. (**b**) High-throughput image acquisition and quantification of pOct4-mCherry expression in wt E14, *Tdg*-/- and *Tdg*-/- ESCs stably expressing wt, catalytically inactive, DNA binding deficient and caC-specific TDG mutants. Methylation of the pOct4-mCherry reporter leads to a 5-fold lower expression compared to unmodified plasmid. Oxidation of ^{m}CpG sites in the pOct4-reporter results in reactivation of mCherry-expression in wt ESCs but not in *Tdg*-/- ESCs. This re-increase was also obtained in *Tdg*-/- ESCs rescued with wt or caC-specific TDG, while the latter was not as efficient (student's *t*-test, $^{**}P < 0.025$, $n = 200\,000$; error bars indicate standard deviation). (**c**) TDG activity was monitored using an *in vitro* assay based on the ability of APEX1 to create a single strand break out of an abasic site, detectable as distinct band on the gel (empty triangle). Full length DNA is indicated by a solid triangle. The DNA substrates contain one defined modification site as indicated in the figure. Wt TDG is highly active on fC or caC and to a much lesser extent on a G○T mismatch.

pOct4-reporter construct *in vitro*, we separated the generation of modified cytosines from their further processing *in vivo*. This allowed us to directly investigate the cellular factors responsible for gene reactivation independent of endogenous TET activity and regulation. In wt ESCs, we observed strong reporter expression from oxidized but not from methylated plasmids, suggesting oxidation-dependent gene reactivation.

To investigate which pathway is responsible for the observed gene reactivation, we searched for potential TET interaction partners. So far, it has been shown that MBD3 colocalizes with TET1 regulating hmC-marked gene expression (53) and that TET1, TET2 and TET3 interact with OGT controlling protein stability, localization and histone modification (54–57). Also, several chromatin-binding factors, such as HDAC1, EZH2 and MeCP2, have been described to associate with TET1 (58). However, none of these factors is likely to be involved in the process of DNA demethylation. Therefore, we performed a mass spectrometry-based pull-down approach in which several BER factors co-precipitated with TET1. To confirm these results, we used Co-IP and a recently described F3H as-

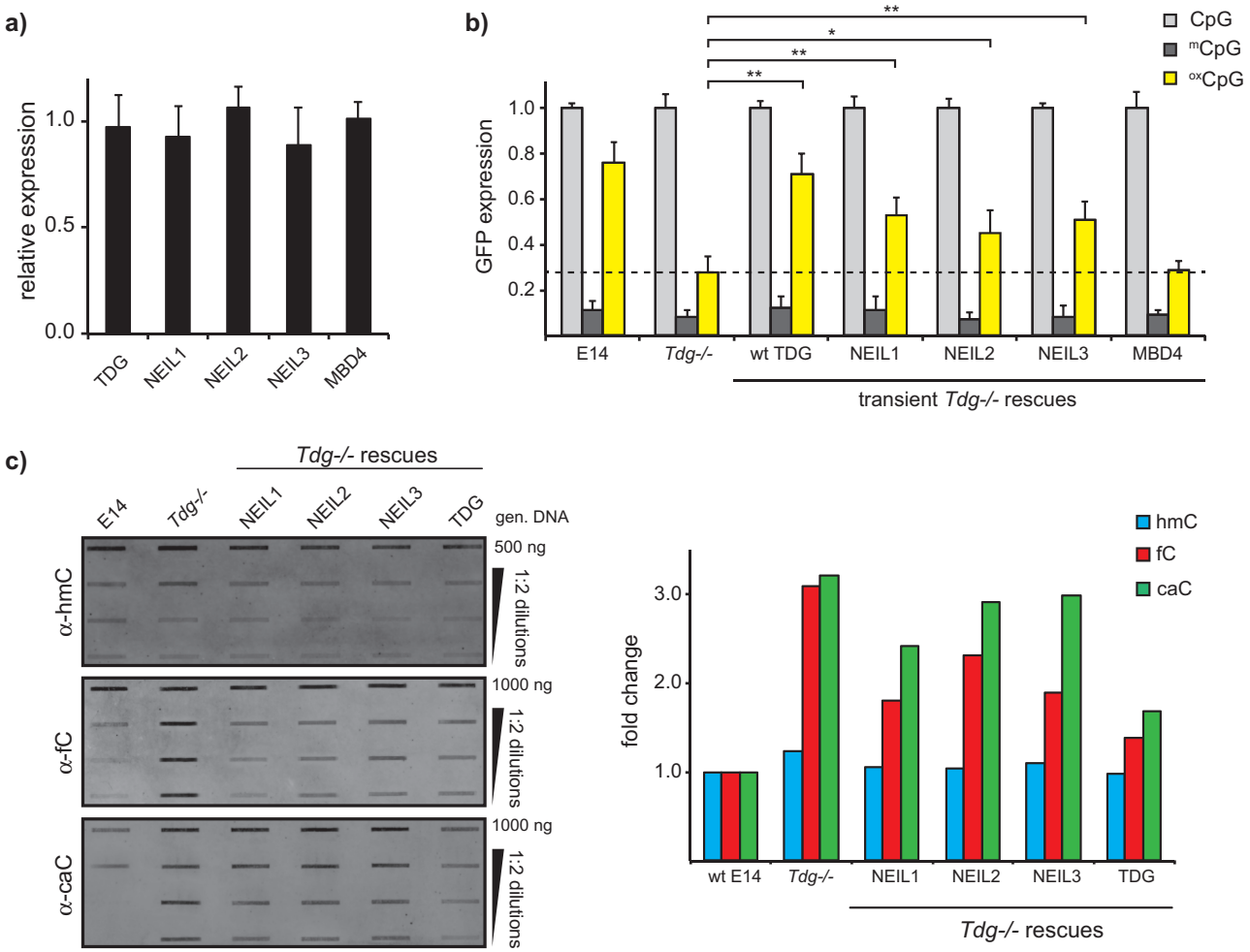


Figure 5. The NEIL glycosylase family can partially compensate for TDG. (**a**) Quantification of mCherry intensities with high-throughput imaging shows that transient *Tdg*-/- rescue ESCs express mCherry-tagged TDG, NEIL1/2/3 and MBD4 glycosylases at comparable levels ($n = 100\,000$; error bars indicate standard deviation). (**b**) The ability of the NEIL family of glycosylases to substitute TDG *in vivo* was monitored by expression of differentially modified pOct4-GFP. With ectopic expression of mCherry-NEIL1, NEIL2 and NEIL3, reporter gene signal was significantly higher than in *Tdg*-/- cells, although not reaching the levels of the wt mCherry-TDG rescue. MBD4 overexpression could not compensate for loss of TDG (student's *t*-test, $*P < 0.05$, $**P<0.025$, $n = 200\,000$; error bars indicate standard deviation). (**c**) Slot blot analysis of genomic DNA isolated from wt E14, *Tdg*-/- and the indicated rescues shows constant levels of hmC and accumulation of fC and caC in *Tdg*-/- cells. Overexpression of wt TDG or NEIL1, 2 or 3 leads to a decrease in genomic fC amounts but does not reach wt E14 levels. Expression of wt TDG or NEIL1 also reduces caC accumulation.

say to test interaction of all three TETs with different BER factors *in vitro* and *in vivo*. We were able to detect interactions of TET proteins with the DNA glycosylases TDG, MBD4 as well as NEIL1, NEIL2 and NEIL3, which excise damaged or oxidized DNA bases (18,25,32,59,60). Furthermore, interactions of all three TET proteins were observed with PARP1, which detects single-strand breaks and modifies repair factors by PolyADP-ribosylation (61). Finally, we showed TET interactions with LIG3 and XRCC1, which are recruited by PARP1 to the site of DNA damage and ligate the DNA strand after the insertion of cytosine (62–64). The observed interactions were largely independent of TET activity, indicating direct protein–protein interactions.

Our findings suggest that both TET-dependent oxidation of mC and subsequent excision of oxidized cytosines by the BER machinery take place in one large protein complex in a spatially and temporally coordinated manner. This close association enables highly efficient replacement of oxidized cytosines. In accordance with these results, fC and caC, in contrast to hmC, are detected at very low genomic levels and are proposed to be immediately removed after generation (18,19).

Initial hypotheses proposed that hmC might be deaminated to hmU by AID/APOBEC deaminases prior to base excision by DNA glycosylases. Suggested candidates were TDG and MBD4, which both have been shown to recognize G○T and G○hmU mismatches (23,65) and have been identified as TET protein interactors in this study. TDG has also been described to be active on fC and caC (18,25). In contrast to wt ESCs, no gene reactivation on the oxidized

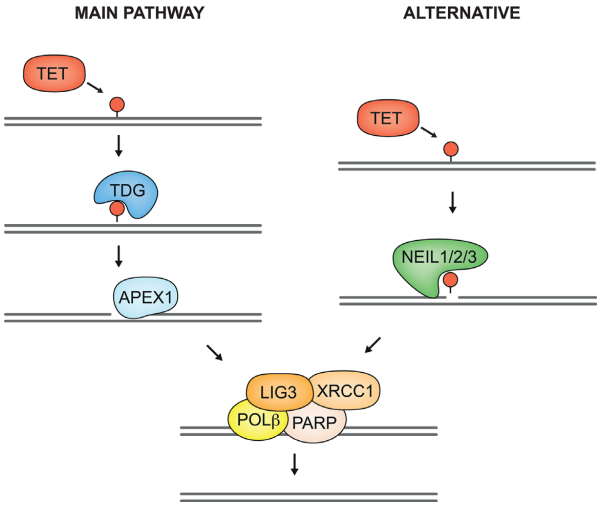


Figure 6. Two alternative pathways for TET-mediated active demethylation. TDG is the major glycosylase that removes fC or caC, generated by TET proteins. Alternatively, the NEIL glycosylases can excise oxidized cytosines, but less efficiently. Both pathways are completed by the BER machinery.

reporter plasmid was observed in *Tdg*-/- cells. Knockout of *Mbd4,* however, had no effect on reporter gene expression, although MBD4 interacts with TET1, TET2 and TET3. Apparently, this association does not contribute to gene activation and may be involved in a different regulatory mechanism.

Our data suggest that conversion of hmC to hmU and subsequent excision by MBD4 does not play a major role in ESCs and are in line with previous studies showing that AID is unable to operate on double-stranded DNA and no detectable deamination of hmC by AID/APOBEC *in vitro* or *in vivo* could be observed (34,66). However, we cannot rule out that AID/APOBEC is involved in the TDG-dependent demethylation pathway or that this pathway contributes to TET3-dependent active demethylation in zygotes (67). Restriction enzyme-based analysis of the oxidized plasmid DNA recovered from wt E14 ESCs provided evidence that conversion of oxidized cytosine to unmodified cytosine led to gene activation. Again, this effect was dependent on TDG but not on MBD4.

Additional *in vivo* experiments with different TDG mutants showed that TDG activity and not the recruitment of unknown factors through TDG is essential for the recovery of gene expression. The specificity of TDG toward fC and caC, but not hmC, was confirmed using a newly established assay based on the ability of APEX1 to recognize glycosylase-generated abasic sites and convert them into single-strand breaks. The results also revealed that TDG activity is much higher on fC or caC than on a G○T mismatch, arguing that oxidized cytosines are the major substrate for TDG and that deamination is not necessary for gene reactivation.

Besides TDG and MBD4, we also investigated the role of the NEIL glycosylase family in TET-mediated demethylation. NEIL1, NEIL2 and NEIL3 have been shown to excise several lesions resulting from DNA oxidations, such as 5-hydroxyuracil or thymine glycol (59,68). NEIL glycosylases are bifunctional, i.e. are also able to convert abasic sites into single-strand breaks (69). *Neil3*-/- mice have been reported to be viable as well as fertile, and the expression of NEIL3 is elevated in hematopoietic tissues, suggesting a function in the immune system or hematopoiesis (70).

Since the NEIL glycosylases have been described as potential binders of oxidized cytosines (16), we tried to compensate for the loss of TDG in *Tdg*-/- cells with each of the three NEIL proteins. Indeed, we could detect a significant reactivation of reporter gene expression, although not reaching the levels of the rescue with wt TDG. We conclude that the NEIL glycosylases can also initiate BER after TET-mediated cytosine oxidation. This hypothesis was further confirmed by the fact that the accumulation of genomic fC and caC in *Tdg*-/- ESCs was far less prominent when NEIL1, 2 or 3 was overexpressed. These data clearly show that the NEIL proteins are not only capable of reactivating the oxidized reporter gene, but also of excising formylated and carboxylated cytosine in its chromatin context. Thus, the NEIL glycosylases may constitute an alternative pathway for active demethylation and reactivation of epigenetically silenced genes (Figure 6).

In summary, we show that the TET proteins interact with a set of factors involved in catalyzing the multiple steps of BER. Furthermore, we demonstrate that TDG is the main glycosylase in TET-mediated reactivation of the epigenetically silenced Oct4 promoter via the BER pathway. It would be of interest whether the activity of the TET–BER machinery differs on other promoter types, such as CpG island containing promoters. Our results also indicate that the NEIL family of glycosylases can functionally replace TDG. It remains to be elucidated to which extent the NEIL glycosylases contribute to TET-mediated demethylation and gene reactivation and how the usage of different glycosylases is regulated *in vivo*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Prof. P. Schär (University of Basel) for providing *Tdg*-/- ESC and N. Nagaraj at the core facility of the MPI for Biochemistry for mass spectrometry analysis. Furthermore, we want to thank E. Schmidtmann for help with establishing the TDG/APEX1 activity assay, C. Trummer for F3H quantifications, D. Meilinger for FACS sorting and C. Mulholland for suggestions on the manuscript.

FUNDING

Deutsche Forschungsgemeinschaft [Collaborative Research Centers SFB 646/B10 and SFB 1064/A17]. C.B. is a member of IMPRS-LS; H.L. is a member of the Nanosystems Initiative Munich (NIM).
Conflict of interest. None declared.

REFERENCES

1. Wutz,A. and Barlow,D.P. (1998) Imprinting of the mouse Igf2r gene depends on an intronic CpG island. *Mol. Cell. Endocrinol.*, **140**, 9–14.
2. Bird,A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
3. Bestor,T., Laudano,A., Mattaliano,R. and Ingram,V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.*, **203**, 971–983.
4. 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.
5. 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.
6. 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.
7. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
8. 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.
9. 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.
10. 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.
11. Costa,Y., Ding,J., Theunissen,T.W., Faiola,F., Hore,T.A., Shliaha,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.
12. 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.
13. Piccolo,F.M., Bagci,H., Brown,K.E., Landeira,D., Soza-Ried,J., Feytout,A., Mooijman,D., Hajkova,P., Leitch,H.G., Tada,T. *et al.* (2013) Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. *Mol. Cell*, **49**, 1023–1033.
14. Konstandin,N., Bultmann,S., Szwagierczak,A., Dufour,A., Ksienzyk,B., Schneider,F., Herold,T., Mulaw,M., Kakadia,P.M., Schneider,S. *et al.* (2011) Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia*, **25**, 1649–1652.
15. 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.
16. 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.
17. 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.
18. 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.
19. Pfaffeneder,T., Hackner,B., Truss,M., Munzel,M., Muller,M., Deiml,C.A., Hagemeier,C. and Carell,T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie*, **50**, 7008–7012.
20. Inoue,A. and Zhang,Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science*, **334**, 194.
21. Iqbal,K., Jin,S.G., Pfeifer,G.P. and Szabo,P.E. (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 3642–3647.
22. Cardoso,M.C. and Leonhardt,H. (1999) DNA methyltransferase is actively retained in the cytoplasm during early development. *J. Cell Biol.*, **147**, 25–32.
23. 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.
24. Wu,S.C. and Zhang,Y. (2010) Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.*, **11**, 607–620.
25. 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.
26. Campalans,A., Kortulewski,T., Amouroux,R., Menoni,H., Vermeulen,W. and Radicella,J.P. (2013) Distinct spatiotemporal patterns and PARP dependence of XRCC1 recruitment to single-strand break and base excision repair. *Nucleic Acids Res.*, **41**, 3115–3129.
27. Dianova,II, Sleeth,K.M., Allinson,S.L., Parsons,J.L., Breslin,C., Caldecott,K.W. and Dianov,G.L. (2004) XRCC1-DNA polymerase beta interaction is required for efficient base excision repair. *Nucleic Acids Res.*, **32**, 2550–2555.
28. Mortusewicz,O., Rothbauer,U., Cardoso,M.C. and Leonhardt,H. (2006) Differential recruitment of DNA Ligase I and III to DNA repair sites. *Nucleic Acids Res.*, **34**, 3523–3532.
29. Cortazar,D., Kunz,C., Selfridge,J., Lettieri,T., Saito,Y., MacDougall,E., Wirz,A., Schuermann,D., Jacobs,A.L., Siegrist,F. *et al.* (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, **470**, 419–423.
30. 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.
31. 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.
32. Hashimoto,H., Zhang,X. and Cheng,X. (2012) Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res.*, **40**, 8276–8284.
33. Kemmerich,K., Dingler,F.A., Rada,C. and Neuberger,M.S. (2012) Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res.*, **40**, 6016–6025.
34. Nabel,C.S., Jia,H., Ye,Y., Shen,L., Goldschmidt,H.L., Stivers,J.T., Zhang,Y. and Kohli,R.M. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.*, **8**, 751–758.
35. Schiesser,S., Hackner,B., Pfaffeneder,T., Muller,M., Hagemeier,C., Truss,M. and Carell,T. (2012) Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angewandte Chemie*, **51**, 6516–6520.
36. Tsukamoto,T., Hashiguchi,N., Janicki,S.M., Tumbar,T., Belmont,A.S. and Spector,D.L. (2000) Visualization of gene activity in living cells. *Nat. Cell Biol.*, **2**, 871–878.
37. Millar,C.B., Guy,J., Sansom,O.J., Selfridge,J., MacDougall,E., Hendrich,B., Keightley,P.D., Bishop,S.M., Clarke,A.R. and Bird,A. (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science*, **297**, 403–405.
38. Ying,Q.L., Wray,J., Nichols,J., Batlle-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.
39. 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.
40. Szwagierczak,A., Brachmann,A., Schmidt,C.S., Bultmann,S., Leonhardt,H. and Spada,F. (2011) Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 5149–5156.
41. Rothbauer,U., Zolghadr,K., Muyldermans,S., Schepers,A., Cardoso,M.C. and Leonhardt,H. (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell. Proteomics*, **7**, 282–289.
42. 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*, **27**, 337–343.
43. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
44. 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.
45. Zolghadr,K., Rothbauer,U. and Leonhardt,H. (2012) The fluorescent two-hybrid (F2H) assay for direct analysis of protein-protein interactions in living cells. *Methods Mol. Biol.*, **812**, 275–282.
46. 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.
47. Hardeland,U., Bentele,M., Jiricny,J. and Schar,P. (2000) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.*, **275**, 33449–33456.
48. Hashimoto,H., Zhang,X. and Cheng,X. (2013) Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant. *J. Mol. Biol.*, **425**, 971–976.
49. Hashimoto,H., Hong,S., Bhagwat,A.S., Zhang,X. and Cheng,X. (2012) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.*, **40**, 10203–10214.
50. Neddermann,P., Gallinari,P., Lettieri,T., Schmid,D., Truong,O., Hsuan,J.J., Wiebauer,K. and Jiricny,J. (1996) Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.*, **271**, 12767–12774.
51. Wiebauer,K. and Jiricny,J. (1990) Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 5842–5845.
52. Raiber,E.A., Beraldi,D., Ficz,G., Burgess,H.E., Branco,M.R., Murat,P., Oxley,D., Booth,M.J., Reik,W. and Balasubramanian,S. (2012) Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.*, **13**, R69.
53. Yildirim,O., Li,R., Hung,J.H., Chen,P.B., Dong,X., Ee,L.S., Weng,Z., Rando,O.J. and Fazzio,T.G. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, **147**, 1498–1510.
54. Deplus,R., Delatte,B., Schwinn,M.K., Defrance,M., Mendez,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.
55. Chen,Q., Chen,Y., Bian,C., Fujiki,R. and Yu,X. (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, **493**, 561–564.
56. Shi,F.T., Kim,H., Lu,W., He,Q., Liu,D., Goodell,M.A., Wan,M. and Songyang,Z. (2013) Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J. Biol. Chem.*, **288**, 20776–20784.
57. 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 beta-N-acetylglucosamine transferase (OGT). *J. Biol. Chem.*, **289**, 5986–5996.
58. Cartron,P.F., Nadaradjane,A., Lepape,F., Lalier,L., Gardie,B. and Vallette,F.M. (2013) Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer*, **4**, 235–241.
59. Dou,H., Mitra,S. and Hazra,T.K. (2003) Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J. Biol. Chem.*, **278**, 49679–49684.
60. Hazra,T.K., Kow,Y.W., Hatahet,Z., Imhoff,B., Boldogh,I., Mokkapati,S.K., Mitra,S. and Izumi,T. (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem.*, **277**, 30417–30420.
61. Gibson,B.A. and Kraus,W.L. (2012) New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat. Rev. Mol. Cell Biol.*, **13**, 411–424.
62. Kubota,Y., Nash,R.A., Klungland,A., Schar,P., Barnes,D.E. and Lindahl,T. (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J.*, **15**, 6662–6670.
63. Rice,P.A. (1999) Holding damaged DNA together. *Nat. Struct. Biol.*, **6**, 805–806.
64. Mortusewicz,O., Ame,J.C., Schreiber,V. and Leonhardt,H. (2007) Feedback-regulated poly(ADP-ribosyl)ation by PARP-1 is required for rapid response to DNA damage in living cells. *Nucleic Acids Res.*, **35**, 7665–7675.
65. Niehrs,C. and Schafer,A. (2012) Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol.*, **22**, 220–227.
66. Bransteitter,R., Pham,P., Scharff,M.D. and Goodman,M.F. (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 4102–4107.
67. Santos,F., Peat,J., Burgess,H., Rada,C., Reik,W. and Dean,W. (2013) Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenet. Chromat.*, **6**, 39.
68. Hazra,T.K., Izumi,T., Boldogh,I., Imhoff,B., Kow,Y.W., Jaruga,P., Dizdaroglu,M. and Mitra,S. (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 3523–3528.
69. Jacobs,A.L. and Schar,P. (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*, **121**, 1–20.
70. Torisu,K., Tsuchimoto,D., Ohnishi,Y. and Nakabeppu,Y. (2005) Hematopoietic tissue-specific expression of mouse Neil3 for endonuclease VIII-like protein. *J. Biochem.*, **138**, 763–772.

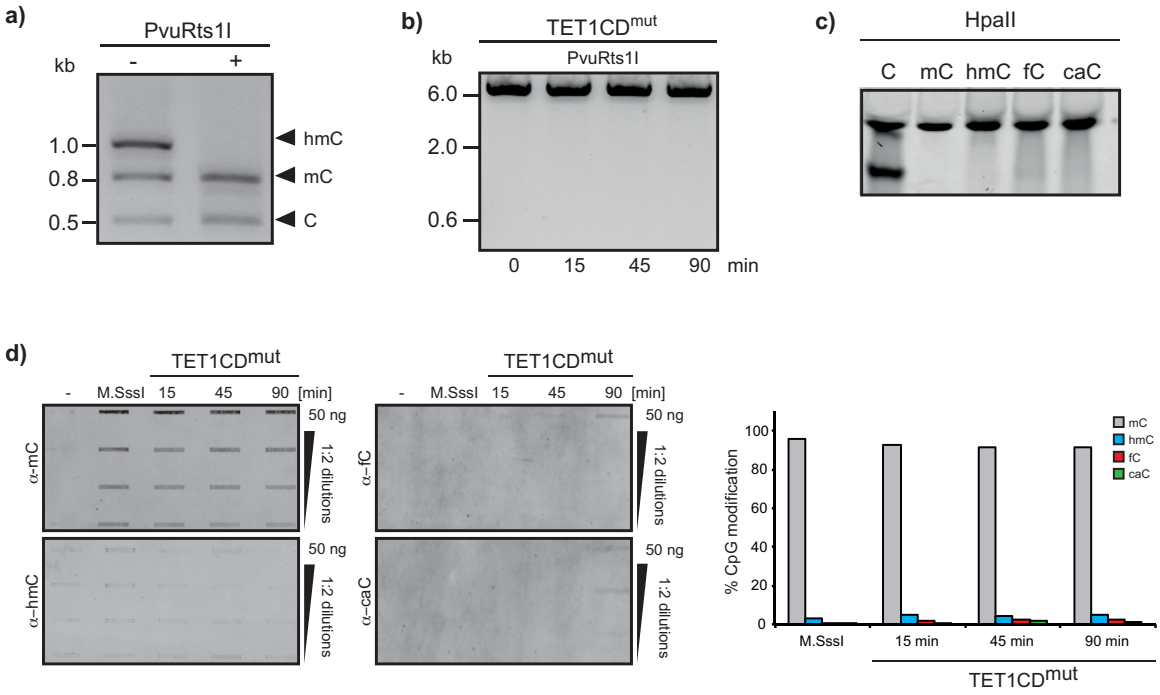
SUPPLEMENTARY FIGURES

TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation

Udo Müller, Christina Bauer, Michael Siegl, Andrea Rottach, Heinrich Leonhardt

Department of Biology II, Ludwig Maximilians University Munich and Center for Integrated Protein Science Munich (CIPSM), 82152 Planegg-Martinsried, Germany

Supplementary Figure S1



Supplementary Figure S1:

a) PvuRts1I specifically cuts hmC-containing DNA. PCR fragments were generated using dATP, dTTP and dGTP with either dCTP, dmCTP or dhmCTP.

b) Incubation of methylated pOct4-GFP with TET1CD^{mut} does not result in any PvuRts1I digestion showing that no DNA oxidation takes place.

c) HpaII specifically cuts unmodified cytosine in a CCGG context. Substrate: 42 bp oligo, containing one defined modification site.

d) Cytosine modifications of untreated, M.SssI-methylated and TET1CD^{mut}-incubated pOct4-mCherry plasmid DNA are detected by slot blot. A 2fold serial dilution of the plasmid DNA was loaded and detected using antibodies against mC, hmC, fC and caC. The mC levels remain high for all samples, except for untreated DNA. Almost no detectable increase in hmC, fC and caC can be observed.

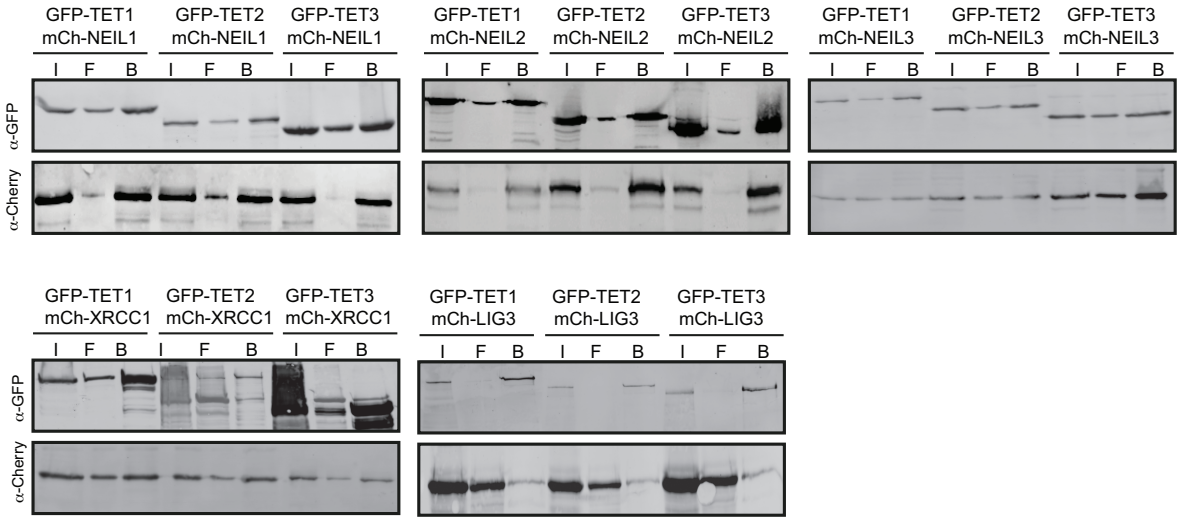
Supplementary Figure S2

a)

Mass spectrometry: IP GFP-TET1

Protein	unique peptides	Sequence coverage	PEP
PARP1	18	26.5 %	0
LIG3	1	1.9 %	9.1×10^{-5}
XRCC1	1.7	8.0 %	3.4×10^{-4}

b)

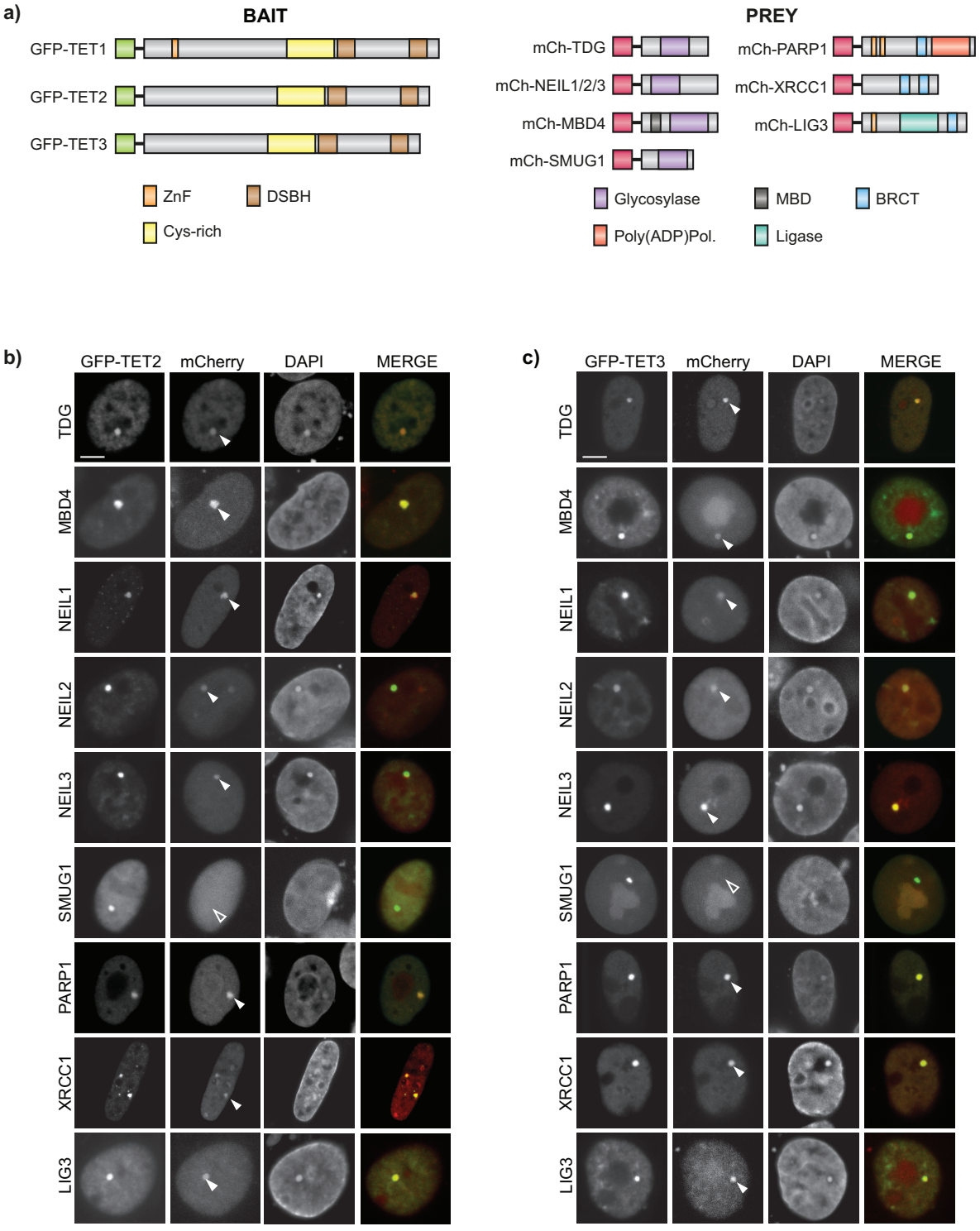


Supplementary Figure S2:

a) PARP1, XRCC1 and LIG3 have been identified as interactors of TET1 by mass spectrometry. Numbers represent the average of three biological replicates. PEP: Posterior error probability.

b) Co-immunoprecipitation with subsequent SDS-PAGE and Western Blotting shows interaction of all three TET proteins with the NEIL1, NEIL2 and NEIL3 glycosylases as well as XRCC1 and LIG3. GFP-tagged TET proteins and the respective mCherry-fusions were expressed in HEK293T cells and immunoprecipitated using the GFP-Trap (I: Input, F: Flowthrough, B: Bound).

Supplementary Figure S3

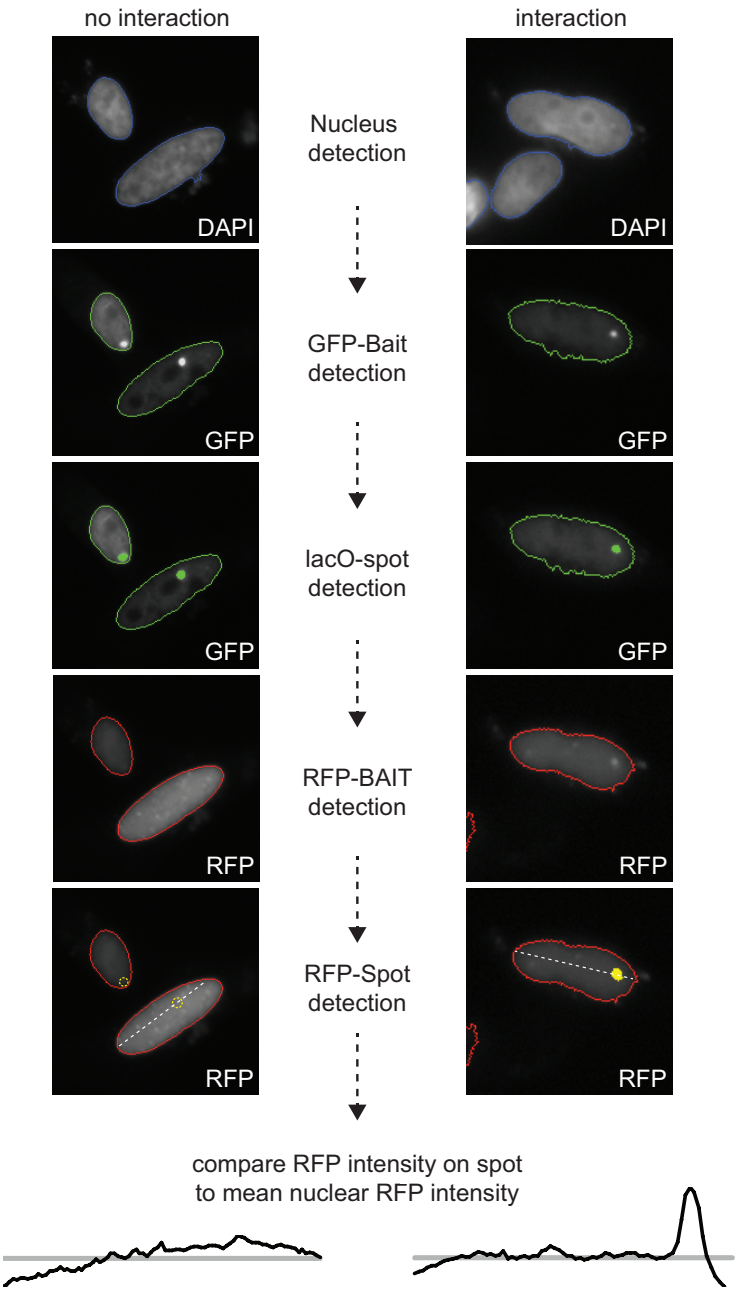


Supplementary Figure S3:

a) Schematic representation of the GFP-BAIT and mCherry-PREY constructs used to analyze TET-BER interactions in F3H and Co-IP.

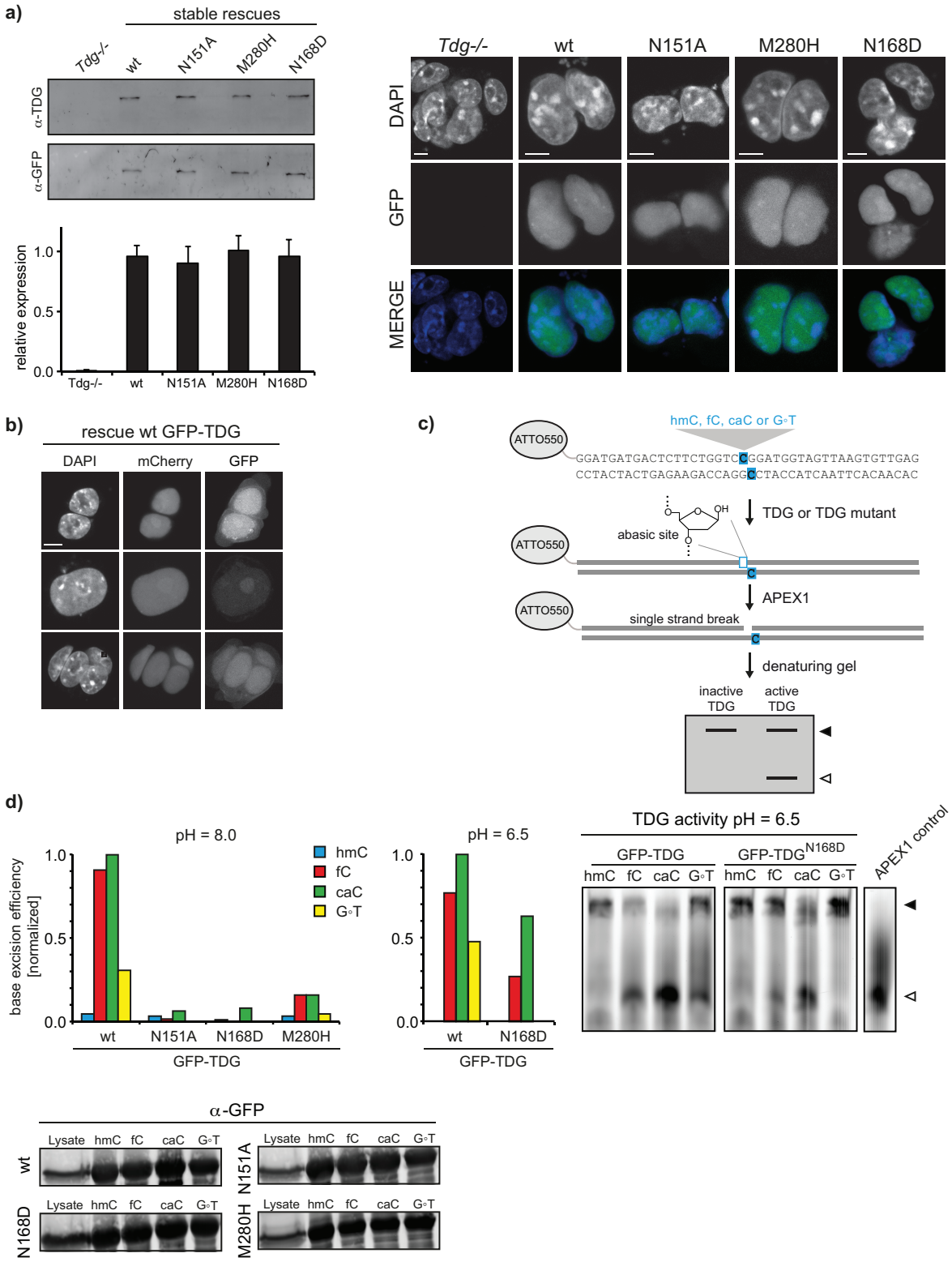
b, c) F3H assay shows *in vivo* interactions of TET2 and TET3 with various glycosylases and BER factors. Positive and negative interactions are marked with a solid or empty triangle, respectively. Scale bar: 5 µm

Supplementary Figure S4



Supplementary Figure S4:
Schematic depiction of the workflow used for automated image acquisition and quantification of the F3H assay.

Supplementary Figure S5



Supplementary Figure S5:

a) Stable *Tdg-/-* rescue ESCs express GFP-TDG and mutants at comparable levels as shown by Western blot analysis (upper left), high-throughput imaging (lower left) and confocal microscopy (right) (n=100,000; error bars indicate standard deviation). Scale bar: 5 µm

b) Rescue of *Tdg-/-* ESCs with wt GFP-TDG re-establishes the capability to express the reporter gene from the ^{ox}CpG plasmid. Scale bar: 5 µm

c) Schematic representation of the APEX1-based TDG activity assay: if TDG is active the formed abasic site is converted into a single strand break by APEX1, resulting in two DNA fragments on a denaturing gel.

d) Quantification of the *in vitro* base-excision activity of purified TDG on defined DNA substrates (left). Equal amounts of GFP-TDG and the mutants were incubated with a 42 bp DNA oligonucleotide harboring a central modified CpG site. At pH=6.5, the caC-specific TDG mutant (N168D) displays a preference for caC and, to a lesser extent, towards fC (right). To confirm equivalent protein levels, protein signal was immunodetected by Western blot using a mouse-anti-GFP antibody (Roche) (bottom).

3 DISCUSSION

3.1 Dynamic readers of cytosine derivatives

Catching specific binders of DNA modifications with quantitative proteomics

Tandem mass spectrometry (MS/MS) is a powerful tool to identify protein-protein interactions, however, the increasing sensitivities of modern mass spectrometers make this approach prone to false positive identifications of interaction partners. SILAC (stable isotope labeling by amino acids in cell culture) based quantitative proteomics can be applied to circumvent these problems. SILAC uses the metabolic incorporation of a given "light" or "heavy" form of stable isotope labeled amino acids into cellular proteins [Ong et al., 2002]. When protein abundance is quantified, the ration of heavy to light enables to distinguish enriched proteins from the background. As this approach also allows purifications to be performed in one step at low stringency, low-affinity but functionally relevant interactions can be retained that would otherwise be lost [Vermeulen et al., 2008].

In the context of epigenetics, interaction mapping followed by classical MS was initially applied with peptide baits that only differ by a single functional group, such as particular histone modifications, to detect "readers" of this mark [Chan et al., 2009; Stucki et al., 2005]. Without a quantitative setting, however, these peptide pull-downs are very challenging as it is difficult to define specific interactors in the presence of a large number of nonspecific binders [Vermeulen et al., 2010]. In addition, using immobilized oligonucleotides, binders of DNA containing 5mC or interactors of single-nucleotide polymorphisms could be identified [Mittler et al., 2009]. Given the increasing importance of long non-coding RNAs in epigenetic regulation, the application of quantitative RNA pull-downs is also of great interest [Butter et al., 2009; Scheibe et al., 2012].

We used SILAC-based quantitative MS to identify readers of unmodified cytosine, 5mC and its oxidized derivatives 5hmC, 5fC and 5caC as well as 5hmU in ESCs. Furthermore, 5mC and 5hmC binders were analyzed during differentiation from murine ESCs to neuronal progenitor cells (NPCs) and adult brain tissue. Different biochemical and structural methods were further applied to characterize selected readers of these DNA modifications [Spruijt et al., 2013].

C, 5mC and 5hmC recruit a distinct set of readers with little overlap in ESCs

We found a large number of factors participating in the epigenetic regulation of gene expression that preferentially bind to unmodified DNA substrates. Among these proteins was INO80, the helicase of the INO80/SWR chromatin-remodeling complex along with its NFRKB (Nuclear factor related to kappa-B-binding protein) subunit, which is responsible for DNA binding [Yao et

al., 2008]. The INO80 complex participates in transcription, replication, cell division, and DNA repair by replacing H2A.Z/H2B histone dimers with the canonical H2A/H2B [Bonisch and Hake, 2012; Papamichos-Chronakis et al., 2011; Shen et al., 2000]. Consequently, higher order processes such as genome stability, pluripotency and differentiation involve INO80 remodeling [Billon and Cote, 2013; Li et al., 2012; Wang et al., 2014].

As another cluster of unmodified DNA-binders we detected CXXC-domain containing proteins, including the transcription factors CXXC5 and ZBTB2 (Zinc finger and BTB domain-containing protein 2). CXXC5 participates in modulating WNT-signaling and serves as a transcriptional activator of FLK1, a receptor for vascular endothelial growth factors [Andersson et al., 2009; Kim et al., 2014]. ZBTB2 was initially described to be located to unmethylated CG-rich sequences, where it acts as a repressor of the ARF-HDM2-P53-p21 pathway, which is crucial for cell cycle regulation [Jeon et al., 2009]. Recently, ZBTB2 was also found to be attracted by 5mC but repelled by 5hmC *in vitro* [Lafaye et al., 2014]. In addition, we identified the CXXC-domain containing lysine demethylase KDM2B (Histone-H3-lysine-36 demethylase 2B), which recruits PRC1 (Polycomb Repressive complex 1) to CpG islands and is important for self-renewal of ESCs as a reader of unmodified DNA [He et al., 2013; Wu et al., 2013].

In line, we found several subunits of the PRC1 itself, such as PCGF1 (Polycomb group RING finger protein 1), the ubiquitin-ligases RING1A/RING1B and BCOR (BCL6-corepressor) in the pull-downs with unmethylated DNA substrates. PRC1 sets the H2AK119ub1 mark and regulates silencing of target genes by interfering with the SWI/SNF chromatin remodeling machinery, blocking transcriptional initiation or recruiting additional silencing factors [Dellino et al., 2004; Lavigne et al., 2004; Shao et al., 1999]. However, the exact function and DNA-binding modes for several readers of unmodified cytosines are still not fully understood yet. Further biochemical and structural studies are needed to elucidate their role in the chromatin context.

Next, we identified 16 proteins enriched for 5mC containing DNA substrates in ESC [Spruijt et al., 2013]. Among these readers were the methyl-CpG-binding group proteins MBD1 and MBD4. Notably, MBD2 was not detected in the pull-downs, which could be explained by the fact that it is not expressed in ESCs [Hendrich and Bird, 1998].

Other 5mC interactors include the transcription factor RFX1 (Regulatory factor X1), which regulates expression of FGF (Fibroblast growth factor) and MHC (Major histocompatability complex) class II genes [Herrero Sanchez et al., 1992; Hsu et al., 2012]. RFX1 is an unusual member of helix-turn-helix DNA-binding proteins because it uses a beta-hairpin (or wing) to recognize DNA instead of the typical recognition helix (Figure 10A) [Gajiwala et al., 2000]. The preference for methylated DNA substrates is surprising, as it was shown that RFX proteins can

prevent methylation of F-box sequences inside promoters of MHC class II genes, the physiological binding sites of RFX proteins [Seguin-Estevez et al., 2009].

Interestingly, we detected the Krüppel-like zinc finger containing transcription factors KLF2, KLF4 and KLF5 among the 5mC binders. KLF4 is one of the four classic Yamanaka reprogramming factors and essential for self-renewal of ESCs as it prevents their differentiation [Takahashi et al., 2007; Zhang et al., 2010b]. Crystal structures of KLF4 bound to a 5mC substrate revealed that methylation had modest, though significant, effects on binding (Figure 10B). Mutating the DNA sequence outside the CpG site, however, abolishes KLF4 binding, regardless of the modification status [Chen et al., 2008]. These data indicate that the interaction between KLF4 and DNA depends largely on the specific sequence context and less profoundly on the cytosine modification state. Recent KLF4 ChIP-seq and genome wide methylation data indicate that CpG sites in the core KLF4 binding motif GGCG are highly methylated in mouse and human ESCs [Liu et al., 2014; Stadler et al., 2011].

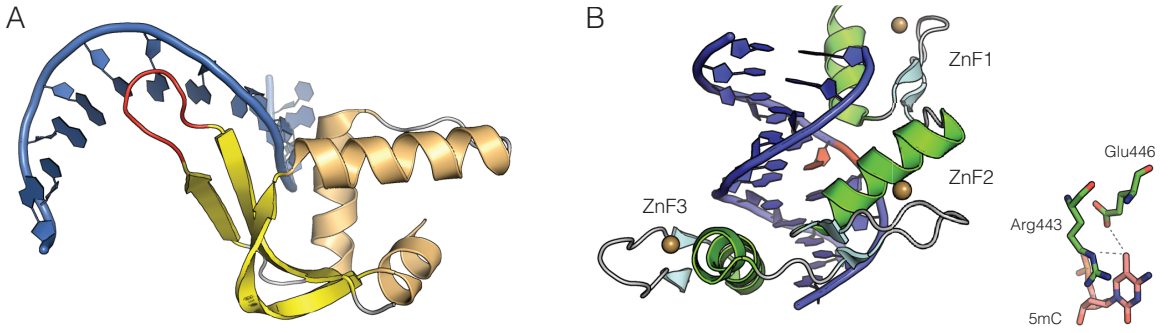


Figure 10: Crystal structures of the 5mC binders RFX1 and KLF4

(A) Binding mode of the human RFX1 transcription factor to a 16-bp DNA oligonucleotide, which resembles the X-box sequence, the physiological RFX1 binding site (PDB 1DP7). The RFX1 DNA-binding domain uses a special mechanism for DNA recognition, which differs from all other known HTH-proteins. Two beta-strands (yellow) and their connecting loop or wing (red), make extensive contacts with the major groove of one half-site DNA (blue). A single side chain from an alpha-helix (light-brown) interacts with the minor groove of the other half-site (not shown).

(B) Murine KLF4 in complex with a methylated CpG containing DNA substrate (PBD 4M9E). KLF4 contains three tandem CXXC-zinc fingers (ZnF; green, zinc ions brown) at its C-terminus, which bind in the major groove of the DNA (blue). The second ZnF (ZnF2) interacts with the central methylated cytosine (red). A spatially conserved Asp-Glu pair is mainly involved in stabilizing the binding to 5mC (right).

Furthermore, five proteins specifically binding to 5hmC containing DNA substrates. Among these factors were the BER glycosylase MPG (N-methylpurine glycosylase) and the helicase RECQ1 (DNA-dependent ATPase Q1).

MPG catalyzes the excision of several mutagenic base substrates, including methylpurines, deaminated adenine and oxidized guanine [Elder et al., 1998; Wyatt et al., 1999]. Crystal structures indicated that MPG intercalates with the minor groove of the DNA, causing base flipping into the active site of the enzyme [Lau et al., 1998]. Interestingly, MBD1 has been identified to interact

with MPG, leading to a synergistic effect on gene silencing. MBD1 and MPG normally bind to methylated promoters. As exogenous and endogenous DNA damage causes the generation of 7-methyl-guanine, the MBD1-MPG complex dissociates from the damaged methylated CpG sites. MPG then spreads along the promoter to remove damaged bases and inhibit transcription at these sites. After completion of repair, MBD1 returns to the repaired methyl-CpG sites together with MPG to reconstruct the repressive chromatin [Hameed et al., 2014; Watanabe et al., 2003]. As will be discussed in section 3.4, MPG binding or activity towards 5hmC has not been described yet and could potentially serve as a pathway for TET-mediated DNA demethylation.

RECQ1 rapidly accumulates at oxidatively damaged DNA sites and at double-strand breaks, thus contributing to the non-homologous end-joining repair pathway [Parvathaneni et al., 2013; Sharma et al., 2012]. Additionally, RECQ1 also plays an essential role in maintenance of telomeres and restoring active replication forks that have regressed as a result of topoisomerase 1 inhibition [Berti et al., 2013; Popuri et al., 2014]. Binding of RECQ1 to 5hmC indicates that this factor might be involved in active DNA demethylation via different repair pathways, however also regulatory functions are conceivable.

We did not detect binding of MBD3 to 5hmC using quantitative MS analysis, which is in line with previous results [Hashimoto et al., 2012b; Iurlaro et al., 2013]. Nevertheless, MBD3 localization to 5hmC has been proposed to be involved in active DNA demethylation and activation of gene expression [Shimbo et al., 2013; Yildirim et al., 2011]. Very recently, binding properties were characterized by the solution structure of the MBD3-MBD in complex with unmethylated, 5mC and 5hmC containing DNA [Cramer et al., 2014]. The overall fold of this domain is highly similar to other MBDs, yet, a key loop involved in DNA targeting is more disordered. It could also be shown that MBD3 preferentially recognizes methylated and to a lesser degree unmethylated CpGs, however, it does not distinguish between hydroxymethylated and unmethylated sites. The homologous MBD2 protein shows much stronger binding to methylated than to unmethylated CpGs. These findings correlate with the observations that both MBD2 and MBD3 are found at unmethylated CGIs, whereas MBD2 binds with much greater affinity to methylated CGIs.

We observed that MECP2 and UHRF1 bind to 5mC as well as 5hmC [Spruijt et al., 2013]. MECP2 is an essential transcriptional repressor that mediates gene silencing through binding to methylated DNA. The additional interaction with 5hmC supports the role of MECP2 as the main 5hmC-binding factor in the brain. In line, MECP2 was shown to bind hemimodified 5hmC- and 5mC-containing DNA with similar affinities [Hashimoto et al., 2012b; Mellen et al., 2012]. Binding specificity has been thought to depend on interactions between 5mC and a hydrophobic patch within the MBD. X-ray analysis of a methylated DNA-MBD crystal revealed, however, that the

methyl group makes contact with a predominantly hydrophilic surface that includes tightly bound water molecules [Ho et al., 2008]. This suggests that MECP2 recognizes hydration of the major groove of methylated DNA rather than cytosine methylation per se. Interestingly, the Rett-syndrome causing mutation R133C inhibits 5hmC binding [Mellen et al., 2012]. This indicates that 5hmC and MECP2 contribute to a cell-specific epigenetic mechanism for regulation of chromatin structure and gene expression. In contrast, *in vitro* studies showed a clear preference for methylated CpGs compared to hydroxymethylated by MECP2 [Frauer et al., 2011a; Valinluck et al., 2004]. The conflicting binding features are also reflected by the findings as MECP2 was shown to serve as a global activator of gene expression in neurons by binding to 5hmC [Li et al., 2013], which contrasts with the notion of MECP2 acting as a transcriptional repressor [Ebert et al., 2013; Jones et al., 1998].

Different opinions prevail regarding the specificity of UHRF1 towards 5hmC. In line with our proteome-wide MS screen, the SRA domain of UHRF1 was shown to bind hemihydroxymethylated DNA with a comparable affinity as hemimethylated substrates. Molecular dynamics simulations indicated that the flipped out 5hmC base fits into the binding pocket of the UHRF1-SRA domain and is additionally stabilized by hydrogen bond formation involving the hydroxyl group [Frauer et al., 2011a]. Notably, UHRF1 and its homologue UHRF2 are the only proteins known so far using a base-flipping mechanism and not showing any catalytic activity on the target base. In contrast, it was proposed that the binding pocket of UHRF1 is too narrow to accommodate 5hmC and the affinity of the SRA domain bound to hemimethylated DNA was significantly higher as for 5mC [Otani et al., 2013]. The difference in the sequences and the lengths of the DNA substrates used in the different approaches may partly be the reason for this discrepancy. Nevertheless, the specific binding of UHRF1 to 5hmC containing DNA is very surprising as UHRF1 was shown to be crucial for the maintenance of DNA methylation by directing DNMT1 to hemimethylated CpG sites [Sharif et al., 2007], which raises the question about the exact function UHRF1 in this context.

5fC and 5caC recruit a large number of binders, including repair proteins and transcription factors

We applied the SILAC-based DNA pull-down approach to also identify readers for 5fC and 5caC in ESCs. Both modifications recruit more proteins than 5hmC and again we only observed limited overlap between their binders [Spruijt et al., 2013]. A number of factors involved in DNA damage response were significantly enriched for 5fC-containing substrates. Among these proteins was the BER glycosylase UNG (see chapter 3.4) as well as DDB2 (DNA damage binding protein 2).

DDB2 recognizes UV-induced DNA damage and recruits proteins of the NER pathway to initiate DNA repair [Li et al., 2006]. Interestingly, the tumor suppressor P53, which was shown to induce upregulation of DDB2 resulting in increased DNA repair, was also able to bind to 5fC [Barckhausen et al., 2014].

Interestingly, among the proteins specifically interacting with 5caC was DNMT1, which was also confirmed using EMSAs as well as western blotting. DNMT1 was shown to decarboxylate 5caC under specific conditions *in vitro* resulting in unmodified cytosine [Liutkeviciute et al., 2014]. Although this enzymatic activity seems unlikely under a reducing cellular environment *in vivo*, these data indicate that DNMT1 binding to 5caC can occur in ESC lysates.

One of the proteins that binds to 5fC and 5caC, but not to 5hmC, is the glycosylase TDG, which is consistent with its reported substrate specificity and its ability to excise 5fC and 5caC [He et al., 2011; Maiti and Drohat, 2011]. Crystal structures of the catalytic domain of human TDG showed that the double-stranded DNA backbone is bend by ~45° and 5caC is flipped out into the active site (Figure 11). Inside the active site pocket 5caC is stabilized via hydrogen bonds and hydrophobic interactions from surrounding residues [Zhang et al., 2012]. These binding properties suggest that TDG is one of the glycosylases involved in the removal of 5fC and 5caC bases via the BER pathway as will be discussed in more detail in chapter 3.4.

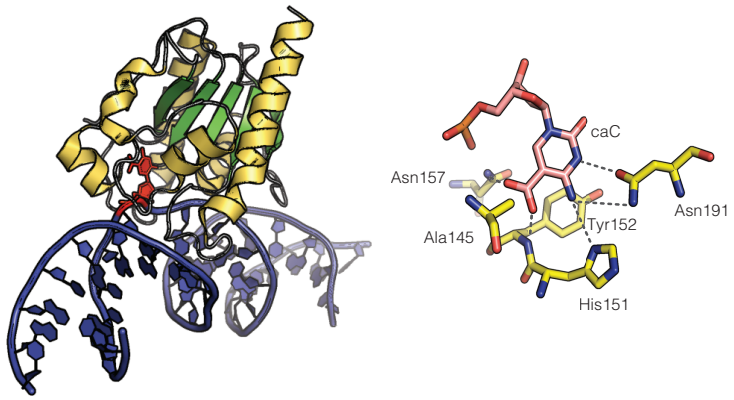


Figure 11: Crystal structure of human TDG bound to a caC- containing substrate.

TDG interacts with the backbone of the 5caC-containing DNA strand via electrostatic interactions and bends the backbone (blue) by ~45° towards the active site. The 5caC base (red) is pushed out of the DNA groove and flipped inside the active site. In addition to several interactions from Asn191 and His151, the carboxyl-group is accommodated in the active site with a binding pocket defined by Ala145, and hydrogen bonds from Asn157 and the backbone amide of Tyr152 (PBD 3UO7).

We identified three proteins being pulled-down with all oxidized derivatives of 5mC: First the uncharacterized factor THY27 (Thymocyte protein 28). Second, C3ORF37, which is a eukaryotic member of the prokaryotic SRAP (SOS response associated peptidase) protein family and is proposed to serve as an autoproteolytic switch that recruits repair enzyme upon DNA damage [Aravind et al., 2013]. Third, NEIL1, a glycosylase with a broad substrate range, which is involved in BER. Thus, C3ORF37 and especially NEIL1 may have a role in removal of the oxidized methylcytosine bases as part of the DNA demethylation pathway.

Readers are highly dynamic during development

To investigate whether interactions with C, 5mC and 5hmC change during development, ESCs were differentiated to NPCs and label-free quantitative MS analysis was applied. In summary, we found a large number of proteins in NPCs, which bound to unmodified cytosine or 5mC, whereas fewer factors were specific for 5hmC. Three smaller groups of proteins were enriched for two different modification states [Spruijt et al., 2013].

As we already observed in ESCs, several CXXC-domain-containing proteins such as the lysine demethylase KDM2B and the histone methyltransferase MLL1 (Myeloid/lymphoid or mixed-lineage leukemia protein 1), together with their associated factors BCOR, RING1A/B and RBBP5 were enriched for unmodified DNA substrates. KDM2B plays an essential role in repressing the *Ink4/Arf* locus and thus suppressing cellular senescence [He et al., 2008; Tzatsos et al., 2009]. Recently, KDM2B was also found to facilitate reprogramming into iPSCs and recruit PRC1/2 to CGIs [Guo et al., 2011b; He et al., 2013]. Interestingly, the reprogramming efficiency of KDM2B increased using Vitamin-C as was also previously reported for TET proteins [Chen et al., 2013a; Wang et al., 2011b]. Both enzymes belong to the family of Fe(II)-2-oxoglutarate dependent dioxygenases and share a similar reaction mechanism. MLL1 is a member of the mixed-lineage leukemia family of histone methyltransferases in mammals [Milne et al., 2002; Nakamura et al., 2002]. It catalyzes mono- di-, and trimethylation of histone H3 on K4 through its conserved SET domain. Both MLL1 and H3K4 methylation were shown to localize across gene promoters, transcription start sites, and 5′ transcribed regions of target genes and facilitate transcription initiation especially controlling the expression of homeobox (*Hox*) genes during differentiation [Guenther et al., 2005; Lauberth et al., 2013].

The MBD proteins MBD1 and MECP2 were specifically enriched for 5mC containing DNA in NPCs. In contrast to ESCs, MBD2 and the associated Mi-2/NuRD complex subunits could be detected as 5mC readers as well. MBD2 was shown to mediate gene repression through recruitment of the Mi-2/NuRD complex to methylated promoters [Barr et al., 2007; Zhang et al., 1999]. It has not only been reported that MBD2 selectively binds methylated DNA *in vitro* [Le Guezennec et al., 2006] but the quantitative MS-based approach supports the very recent findings that this binding also happens *in vivo* [Baubec et al., 2013]. Genome wide mapping of MBD2 binding further showed that binding predominantly occurs at highly methylated, CpG dense regions [Menafra et al., 2014].

Interestingly, several known 5mC readers, such as Kaiso, UHRF1, and MBD4, were also able to bind to 5hmC containing substrates in NPCs, suggesting a potential dual role for these factors previously associated only with transcriptional repression and maintenance of DNA methylation. Again, DNA glycosylases such as NEIL1, NEIL3 and OGG1 (8-oxoguanine DNA glycosylase 1)

as well as several helicases bound specifically to 5hmC, which again suggests DNA repair contributing to DNA demethylation.

Finally, we found that UHRF2 was specific for 5hmC in NPCs. UHRF2 is not expressed in ESCs and is upregulated upon differentiation [Pichler et al., 2011], explaining why UHRF2 was not detected in ESCs. Recently, a crystal structure of the SRA domain of UHRF2 in complex with 5hmC containing DNA [Zhou et al., 2014] revealed that its binding mode is similar to that of the UHRF1-SRA with 5mC DNA [Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008]. The UHRF2-SRA folds into a saddle-shaped structure, which contains seven β-strands and four α-helices. A positively charged surface area surrounding the loop connecting α1 and β2, generally mediates the DNA binding properties. The 5hmC is flipped out of the DNA duplex and is bound inside a narrow pocket adjacent to this positively charged area. Different hydrophobic contacts and hydrogen bonds stabilize the 5hmC base itself. Additionally, UHRF2 has a binding preference for 5hmC over 5mC due to an extra intermolecular hydrogen bond, with greatest affinity for fully hydroxymethylated sites.

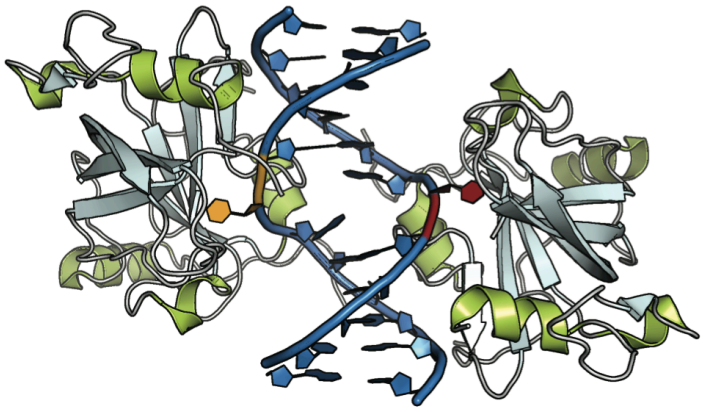


Figure 11: Crystal Structure of UHRF2-SRA bound to hemi-hydroxymethylated DNA.

The UHRF2-SRA domain folds into a saddle-shaped structure consisting of seven β-strands (light blue) and four α-helices (green). UHRF2-SRA contacts the DNA substrate mostly by its phosphate backbone. The SRA-domains bound to a blunt-ended hemi-hydroxymethylated DNA (blue) uses a mechanism, in which both 5hmC (red) and unmethylated cytosine (orange) on the opposite strand are flipped out. The 5hmC base is bound inside a narrow pocket adjacent to a positively charged area. 5hmC is stabilized with hydrophobic contacts by aromatic residues and an additional hydrogen bond (PBD 4PW6).

The binding of UHRF2 to 5hmC could also explain the enhanced activity of TET1, which we observed by co-expression of both factors [Spruijt et al., 2013]. UHRF2 could either target TET proteins to the site of action or more likely by flipping out the base, increase the accessibility of 5hmC for subsequent oxidation. Both cases would lead to elevated 5fC and 5caC levels as we measured in additional experiments [Spruijt et al., 2013].

Readers for the different modified cytosines were also identified in the adult brain, the tissue with highest 5hmC levels. In contrast to NPCs and ESCs, most proteins specifically bound to 5hmC, which may imply a specific role for this cytosine modification in the brain. The analyses with unmodified DNA were enriched for the same proteins as those observed in ESCs and NPCs, including different CXXC containing factors and chromatin remodelers. With exception of

MBD3, all other MBD family members were enriched for 5mC. Notably, the four subunits of replication factor C (RFC2-5) and the associated factor RFC1 were specific 5hmC readers. Replication factor C loads PCNA on the DNA, which organizes various proteins involved in DNA replication, DNA repair, DNA modification, and chromatin modeling [Indiani and O'Donnell, 2006]. Nevertheless, the function of RFC binding to 5hmC remains unclear and has to be further investigated.

Altogether, these results underline the highly dynamic behavior of factors binding to unmodified cytosine, 5mC or 5hmC during differentiation. This study concentrated on the development from ESCs to neuronal cells. It remains to be elucidated, how readers of different DNA modifications act and function in other developmental processes.

3.2 TET proteins oxidize thymine to 5hmU

5hmU is present in ESCs and is bound by specific readers

Naturally occurring DNA damage is responsible for the generation of a variety of base lesions including oxidations, deaminations and base adducts. The presence of 5hmC, 5fC and 5caC in the genome raises the question whether additional base variants can be detected in cells, which are enzymatically generated.

As these oxidized derivatives are present at very low levels compared to 5mC, they can only be quantified using high sensitive MS approaches. These precise quantifications are based on chemically synthesized isotopologues of the different cytosine bases, which are used as internal standards during MS analysis [Pfaffeneder et al., 2011].

With this technique we could detect and quantify for the first time 5hmU at similar levels as 5caC in genomic DNA from ESCs [Pfaffeneder et al., 2014]. Depending on the cell line and growth conditions around 3000-8000 5hmU bases are present in ESCs. Comparable levels of 5hmU were also detected in sperm DNA by a subsequent study [Guz et al., 2014]. In somatic tissues, genomic 5hmU levels were much lower (500-1700 hmU bases) and at the same levels as 5-formlyuracil (5fU) and 8oxoG, which are both formed by reactive-oxygen-species (ROS). Notably, the presence of 5fU in the DNA does not affect replication, however DNA polymerases such as the Klenow-fragment of DNA polymerase I, DNA polymerase alpha and gamma are able to incorporate any of the four bases opposite 5fU allowing transition and transversions. [Bjelland et al., 2001]. 5fU has been described to be repaired by either BER or NER pathways [Belousova et al., 2013; Kino et al., 2004].

The high levels of 5hmU compared to 5fU and 8oxoG in ESCs suggests that 5hmU may be formed by a ROS independent mechanism. Recent studies proposed that deamination of 5hmC by the AID/APOBEC enzymes could lead to the generation of 5hmU [Cortellino et al., 2011]. The AID/APOBEC family of enzymes are mainly involved in converting cytosine to uracil. This reaction plays a key role in adaptive immunity as it seeds the process of antibody diversification and maturation via somatic hypermutation and class-switch recombination [Muramatsu et al., 2000; Pavri and Nussenzweig, 2011]. In innate immunity, deaminases from the APOBEC3 subfamily have been shown to mutate foreign viral genomic intermediates to promote their degradation and prevent viral integration [Marin et al., 2003; Sheehy et al., 2002].

To test whether AID/APOBEC enzymes can also deaminate 5hmC *in vivo*, we performed isotope-tracing experiments by substitution of L-methionine with (*methyl*-$^{13}CD_3$)-L-methionine, which is needed as the methyl-group donor for DNMTs [Pfaffeneder et al., 2014]. We detected the incorporation of the isotope-labeled L-methionine into 5mC, 5hmC and 5fC but not into 5hmU, which indicates that this base is not a product of enzymatic deamination of 5hmC. Using isotope-labeled [^{13}C,$^{15}N_2$]thymidine we observed its incorporation into T and also 5hmU, which clearly demonstrates that genomic 5hmU is generated via oxidation of T.

The presence of 5hmU implies that it may serve as an additional epigenetic mark that is recognized and bound by specific readers as shown for the individual cytosine derivatives. Thus, we also applied SILAC-based quantitative mass spectrometry to identify proteins binding directly or indirectly to 5hmU in ESCs.

Amongst the readers we found the methyltransferases DNMT3A/3B as well as their interacting partners UHRF1 and UHRF2 [Meilinger et al., 2009; Zhang et al., 2011]. It has been shown that DNMT3A interacts with TDG, which mutually affects the enzymatic activity of both proteins: DNMT3A enhances the glycosylase activity, while TDG inhibits the methylation activity *in vitro* [Li et al., 2007]. These data suggest a functional link between DNA repair and DNA methylation at specific sites. Moreover, DNMT3A had no effect on the activity of SMUG1, while it stimulated the excision activity of TDG towards G·U mismatches. TDG has been shown to be active on 5hmU [Hashimoto et al., 2012a], which could be a potential targeting mechanism. These findings, however, were based on the generation of uracil by 5mC deamination and no hmU·A substrate was tested, the result of TET-mediated thymine oxidation. It remains elusive and needs to be validated whether TDG displays *in vivo* activity at hmU·A mismatches.

Furthermore, we found UHRF2 to directly bind to 5hmU and additional biochemical studies revealed that co-expression of UHRF1 or UHRF2 with the catalytic domain of TET1 led to increased 5hmU levels [Pfaffeneder et al., 2014]. Binding to 5hmU could increase epigenetic

crosstalk, as UHRF proteins were shown to link DNA modifications with histone modifications [Pichler et al., 2011], however, their exact mode of action has to be further studied.

In line, other chromatin modifying enzymes were identified to interact directly or indirectly with 5hmU, such as JARID2 (Jumonji/ARID domain-containing protein 2), which contains an AT-rich binding domain and a CXXC zinc finger with a preference for GC rich sequences [Son et al., 2013]. JARID2 mediates localization of PRC2 to the chromatin, inhibiting methylation of H3K27 in ESCs, a mark associated with repressed transcription [Peng et al., 2009; Shen et al., 2009]. JARID2 has also been shown to be involved in gene repression by inhibition of cyclin D1, OCT4 and genes involved in X-inactivation [da Rocha et al., 2014; Li et al., 2010; Shirato et al., 2009]. Thus, 5hmU recognition by JARID2 could provide further possibilities to fine-tune epigenetic regulation.

5hmU also recruits other chromatin-associated proteins such as the CXXC containing histone H3K36me3 demethylase KDM2B and its counterpart SETD2 (SET domain-containing protein 2), which generates the methylation mark leading to transcriptional activation [Edmunds et al., 2008]. Recently, it was shown that H3K36me3 is also required to recruit mismatch repair proteins *in vivo* [Walters et al., 2014]. Further DNA repair proteins were identified as 5hmU binders such as the NEIL3 glycosylase and several helicases indicating a role for these factors in removal of this base. These hmU binders enable a dynamic interplay between DNA modifications, histone marks and DNA repair.

TET proteins convert thymine to 5hmU

As we could show that genomic 5hmU is not generated via enzymatic deamination of 5hmC but by oxidation of thymine [Pfaffeneder et al., 2014] we addressed the question, which enzyme could be responsible for catalyzing this reaction. Given their function as oxidases, we analyzed TET proteins in more detail. Indeed *Tet1/Tet2* knockdowns resulted in decreased 5hmU levels whereas expression of the catalytic domain of TET1 led to a significant increase. Together with the isotope tracing experiments showing no incorporation of (*methyl*-$^{13}CD_3$)-L-methionine into 5hmU and thus no deamination of 5hmC, these data clearly indicate that TET enzymes can directly oxidize thymine to 5hmU. This enzymatic oxidation leads to hmU·A basepairs, which can be recognized by the SMUG1 glycosylase. In line, knockdown of SMUG1 led to increased 5hmU levels whereas knockdown of TDG had no significant effect on these bases. Recently, the activity of SMUG1 has been additionally confirmed *in vivo* using expression constructs containing multiple 5hmU sites [Luhnsdorf et al., 2014]. As will be discussed in chapter 3.4, the glycosylases UNG1, UNG2, MBD4 and TDG have been shown to be active on uracil-derivatives *in vitro* in addition to SMUG1

and could also constitute alternative pathways for removal of 5hmU *in vivo* [Hashimoto et al., 2012a; Hendrich et al., 1999; Kavli et al., 2002; Krokan et al., 2002].

How can mammalian TET enzymes additionally oxidize thymine and not only cytosines? Trypanosomatids, including the human pathogen *Trypanosoma brucei*, possess a unique DNA modification within the trypanosoma genome known as β-d-glucopyranosyloxymethyluracil (base J) [Gommers-Ampt et al., 1993]. Base J is a hypermodified thymine residue that accounts for about 1% of thymines in their genome and is enriched in repetitive sequences, including telomeric repeats [van Leeuwen et al., 1998a; van Leeuwen et al., 1998b]. Base J is generated by a two-step enzymatic process: JBP1 and JBP2 (J-binding proteins 1/2) first oxidize thymine in the DNA to 5hmU [Cliffe et al., 2012]. Second, JGT (J-associated glycosyltransferase) completes the synthesis of base J by adding a glucose group to 5hmU (Figure 13) [Bullard et al., 2014].

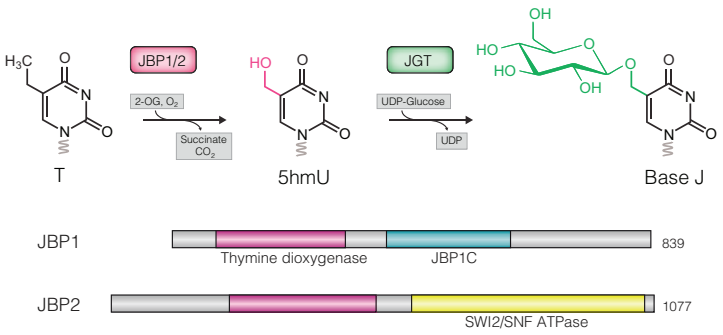


Figure 12: Base J synthesis in trypanosomatids

(Top) Base J synthesis is a two-step reaction in which thymines are first hydroxylated by JBP1 and JBP2 to form 5hmU. The intermediate hmU is subsequently glycosylated by JGT.
(Bottom) Like TET proteins, JBP1 and JBP2 are members of the Fe(II)-2-OG dioxygenase family and both contain a DNA binding motif.

Together with the TET proteins, JBP1/2 are members of the Fe(II)-2OG dependent dioxygenase family [Loenarz and Schofield, 2009; Yu et al., 2007]. Moreover, TET and JBP proteins share some unique features, which are not found in any other dioxygenase. Both families have an extended α-helix just N-terminal to the first core strand inside the DSBH. This long α-helix is kinked due to a conserved proline in the middle of the helix [Hashimoto et al., 2014; Iyer et al., 2009]. Furthermore, JBP and TET proteins show either fusions to DNA-binding or chromatin-associated protein domains, or gene-neighborhood associations with known DNA-binding domains. JBP2 contains a SWI2/SNF2 ATPase module, which is consistent with the role for ATP-dependent chromatin reorganization in synthesis of the J base [DiPaolo et al., 2005]. In contrast, JBP1 binds DNA by a helix-turn-helix domain [Heidebrecht et al., 2011]. TET1 and TET3 proteins either have an intrinsic CXXC DNA-binding domain or in the case of TET2 the neighboring *Idax* gene encodes for a CXXC zinc-finger [Iyer et al., 2009; Ko et al., 2013; Liu et al., 2013b].

A recent evolutionary computational screen identified further JBP/TET members in phages, fungi and algae. Phages encoding TET/JBP genes do not possess any DNA methyltransferases, which

serves as an essential pre-requisite for 5hmC generation [Iyer et al., 2013]. These observations suggest that the nucleotide substrate oxidized by TET/JBP enzymes of phages is very likely a thymine, which is also in line with the TET activity we observed in ESCs. In cyanophages, the TET/JBP gene is further linked to a uracil-5-methyltransferase. In this case, the TET/JBP protein generates 5hmU after synthesis of 5mU (5-methyl-uracil) from thymine. However, in other bacterial species, the TET/JBP gene is fused to a DNA methyltransferase; hence, these versions might oxidize 5mC to generate its oxidized derivatives [Iyer et al., 2013]. This suggests that mammalian TET proteins retained both catalytic activities: oxidation of 5mC and generation of the oxidized cytosine derivatives as well as oxidation of thymine to 5hmU. It remains elusive how the different catalytic activities are regulated in different species and whether they are sequence-context or gene specific.

3.3 Regulation of TET proteins by posttranslational modifications

TET proteins are subject to glycosylation and phosphorylation

So far only little is known how TET proteins are regulated and whether they are subject to posttranslational modifications. We found that all three TET proteins can interact with OGT (O-linked GlcNAc-transferase), which is in line with previous studies [Deplus et al., 2013; Shi et al., 2013; Vella et al., 2013]. OGT transfers O-GlcNAc from UDP-GlcNAc to the hydroxyl group of serine or threonine residues of nuclear and cytoplasmic proteins [Kreppel et al., 1997]. Using a highly sensitive MS approach we could further map individual glycosylation and phosphorylation sites. Increasing OGT expression led to higher O-GlcNAc levels whereas phosphorylation decreases simultaneously [Bauer et al., 2014].

Interestingly, we did not detect a direct competition for the same serine or threonine residues that are modified, but rather proximal site occupancy. Previous O-GlcNAc and O-phosphate site-mapping studies proposed that there are at least three different types of dynamic interplay between both modifications. First, there is competitive occupancy at the same site, which has been shown for the transcription factor c-MYC and estrogen receptor-beta [Cheng and Hart, 2001; Kamemura et al., 2002]. Second, the competitive and alternative occupancy occur at proximal residues, as observed in the tumour suppressor P53, synapsin I and in our case in TET proteins [Cole and Hart, 1999; Yang et al., 2006]. Third, there is a complex interplay whereby some phosphorylation sites are identical to O-GlcNAc sites, whereas others are adjacent to, or even distant from, each other, such as at the C-terminal domain of RNA polymerase II [Comer and Hart, 2001]. The direct interplay between O-GlcNAc and phosphorylation is also underscored by

the recent finding that OGT forms transient complexes containing the catalytic subunit of protein phosphatase 1 [Wells et al., 2004]. Additional *in vitro* assays depicted that phosphopeptides could be concurrently dephosphorylated and remodified by OGT in this functional unit.

The biological importance of OGT and O-GlcNAcylation is manifested by its essential roles in ESC viability and mouse embryonic development [Shafi et al., 2000]. Although OGT has been shown to interact with all three TET proteins and catalyze O-GlcNAcylation of TET1 and TET2, OGT was not found to significantly affect their activities. In contrast, TET proteins were described to mediate targeting of OGT to chromatin and enhancing its enzymatic activity [Chen et al., 2013b; Deplus et al., 2013; Ito et al., 2014; Vella et al., 2013]. However, a recent study depicted that OGT drives TET3 out of the nucleus in a glucose-dependent manner and, thus, is able to inhibit TET3 activity by altering its subcellular localization [Zhang et al., 2014].

Further evidence for an epigenetic function of OGT has evolved with the discovery that it can glycosylate histone H2B on serine 112, causing its ubiquitinylation and the subsequent trimethylation of histone H3 at lysine 4 (H3K4me3). This modification is associated with active genes and mainly recognized by the SET1/COMPASS complex [Fujiki et al., 2011]. These data propose a model of gene activation, in which TETs associate with OGT on CpG rich promoters and enhance its glycosyltransferase activity.

Protein kinases control a variety of cellular processes, including signal transduction, cell cycle progression, protein localization and activity. Protein phosphorylation also plays a crucial role in intercellular communication during development, in physiological responses and in homeostasis. Mutations or dysregulation of kinases play causal roles in human disease, affording the possibility of developing agonists and antagonists of these enzymes for use in therapy [Blume-Jensen and Hunter, 2001]. Given the large number of more that 500 protein kinases, of which the majority are serine/threonine kinases [Manning et al., 2002] the identification for the TET specific one will be challenging but of great interest.

3.4 DNA demethylation leads to gene expression

The proteins involved in setting the different cytosine modifications, namely DNMTs and TETs, have been intensively studied. We could also show that 5mC and its oxidized derivatives attract specific readers, which are highly dynamic during development. However, it is not fully understood, which mechanisms or pathways can contribute to the removal of the methylation mark, how they are regulated and what effect they have on gene expression.

DNA glycosylases involved in TET mediated gene re-activation

Almost all suggested active DNA demethylation pathways start with the initial recognition and removal of the oxidized bases by specific DNA glycosylases. Eleven DNA glycosylases have been identified in mammals, which show different structural features and are active on a variety of damaged bases. These glycosylases can be divided into two main classes: monofunctional glycosylases, which only display glycosylase activity and bifunctional glycosylases, which also possess AP-lyase activity [Jacobs and Schar, 2012]. This additional activity allows to cleave the phospho-diester bond of the DNA, creating a single-strand break without the need for the AP-Endonuclease (APEX1). It is still not fully understood how DNA glycosylases locate damaged nucleobases among millions of undamaged base pairs. The main challenge is that many damaged bases targeted by glycosylases only slightly differ from their normal counterparts. As glycosylases do not use biochemical energy during the search for DNA damage, they are only thermally driven to translocate and move along the genome. One mechanism for the detection of oxidatively damaged bases suggests the insertion of specific residues into the DNA helix. Thereby, glycosylases diffusely scan the DNA and simultaneously test the strength and flexibility of base pairs and the sugar backbone [Banerjee et al., 2006; Dunn et al., 2011; Nelson et al., 2014]. Whether all glycosylases use the same scanning mechanism or differ depending on their substrates remains elusive.

Applying Co-IP and Fluorescent Three Hybrid (3FH) assays, we demonstrated interactions of all three TET proteins with the monofunctional glycosylases TDG, MBD4 and the bifunctional NEIL1, NEIL2, NEIL3 [Muller et al., 2014]. The recruitment of glycosylases to the site of oxidation could imply that the oxidized products are immediately removed. We also observed interactions of TET proteins with downstream factors of the BER, such as PARP1, LIG3, XRCC1. Thus, not only the removal, but also its replacement with unmodified cytosine takes place in one large complex and enables efficient site-specific DNA demethylation.

In this study we developed a reporter gene assay to further investigate the underlying effects of 5mC and its oxidized derivatives with regard to TDG, MBD4 and the NEIL glycosylase family on gene expression. The enzymatic oxidation of a methylated pOct4-reporter was carried out *in vitro*, thus, the generation of modified cytosines was separated from their further processing *in vivo*. This approach enables a more direct investigation of the biological pathways responsible for gene expression independent of endogenous TET activity and regulation. We applied high-throughput imaging, followed by automated image acquisition and quantification for measurement of reporter-gene expression [Muller et al., 2014]. Using GFP or mCherry expression as readout

allows parallel analysis of a large number of single cells. In contrast to other methods such as the luciferase assay, this approach does not require the addition of the luciferase substrate and thus removes a possible source of error or bias. Furthermore, it is also suited for following expression in living cells, as no cell lysis has to be performed.

In wt and *Mbd4*$^{-/-}$ ESCs, we observed strong reporter expression from unmethylated or oxidized but not from methylated plasmids, suggesting oxidation-dependent gene re-activation. In contrast, *Tdg*$^{-/-}$ ESCs showed no reporter gene expression from the oxidized plasmid. Re-isolating the transfected plasmids clearly showed that DNA demethylation took place in wt ESCs and *Mbd4*$^{-/-}$ ESCs, but only very little in TDG deficient cells [Muller et al., 2014]. This demonstrates that DNA demethylation is the cause of the observed gene re-activation.

To elucidate whether the NEIL glycosylase family contributes to gene reactivation, we measured the expression of the modified reporter plasmids in *Tdg*$^{-/-}$ cells transiently overexpressing NEIL1, NEIL2 or NEIL3. Interestingly, we observed a significant increase of pOct4-GFP expression in these *Tdg*$^{-/-}$ rescue cells using oxidized plasmids. Additionally, we isolated genomic DNA from wt E14 and *Tdg*$^{-/-}$ ESCs as well as from the transient rescues with wt TDG and NEIL1, 2 and 3. The rescues of *Tdg*$^{-/-}$ cells resulted in decreased genomic 5fC, and for wt TDG and NEIL1, also in decreased 5caC levels [Muller et al., 2014]. These data indicate that the NEIL glycosylases can partially compensate for the loss of TDG.

TDG actively processes lesions that result from oxidation, alkylation and deamination of cytosine, 5mC, and thymine [Hardeland et al., 2003]. TDG also has been shown to remove 5′-halogenated derivatives of uracil and cytosine, such as 5-bromouracil [Morgan et al., 2007]. Moreover, TDG exhibits highest processing efficiency on lesions opposite guanine within CpG sites, but can also remove several bases opposite adenine [Morgan et al., 2011]. This broad range of substrates allows TDG to eliminate a variety of mutagenic bases and help to stabilize the genome. Knockout of *Tdg* in mice is embryonic lethal, suggesting that it has an essential function in development [Cortazar et al., 2011; Cortellino et al., 2011].

The rescue with functionally active TDG indicates that this glycosylase is mainly responsible for DNA demethylation and the resulting gene-reactivation. Furthermore our APEX assay demonstrated that TDG has the highest activity toward 5fC and 5caC containing DNA substrates [Muller et al., 2014]. This supports previous finding, which indicated that TDG has robust *in vitro* base excision activity on 5fC and 5caC but not on C, 5mC and 5hmC [He et al., 2011; Maiti and Drohat, 2011]. Together with the fact that we identified TDG as a reader for fC and caC and its interaction with TET proteins, its essential role in removal of the oxidized bases was manifested. Computational studies have further suggested that 5fC and 5caC have destabilized *N*-glycosidic

bonds relative to C, 5mC and 5hmC [Bennett et al., 2006; Williams and Wang, 2012]. TDG also seems to have structural features that mediate recognition of these oxidized C bases, including a binding pocket that can specifically accommodate the 5-carboxyl substituent [Zhang et al., 2012]. Structural data showed that 5caC is flipped out from the double-stranded DNA and is stabilized by several hydrogen bonds in the catalytic pocket of the enzyme (Figure 11) [Hashimoto et al., 2013]. Interestingly, the chemical properties for 5fC recognition seem to differ from 5caC, which raises the question of the relative contribution of 5fC compared to 5caC in demethylation and gene regulation.

MBD4 is unique from other thymine and uracil-processing glycosylases as it contains two DNA binding domains, an MBD and a C-terminal glycosylase domain. MBD4 catalyzes the removal of U and T paired with G within CpG sites, which result from deamination of cytosine or 5mC [Hendrich et al., 1999; Petronzelli et al., 2000]. Additionally, MBD4 is active on several halogenated pyrimidines, including 5-chlorouracil and 5-bromouracil paired with G that are produced during peroxidase-mediated inflammatory processes, as well as chemotherapy-induced 5-Fluorouracil (5FU) [Turner et al., 2006; Valinluck et al., 2005]. In the context of DNA demethylation, MBD4 was shown to be active on 5hmU *in vitro*, initially proposed to be the deamination product of 5hmC [Hashimoto et al., 2012c]. Although MBD4 showed an interaction with all three TET proteins, it did not contribute to gene re-activation in ESCs but may be involved in other regulatory mechanism, which are still unknown. Our data suggest that MBD4 is not active on oxidized cytosines in ESCs, which is generated through TET mediated oxidation. However, a potential involvement in DNA demethylation during special developmental stages such as in zygotes or at specific DNA sequences may still require the presence of MBD4.

The mammalian NEIL glycosylase family consists of three members: NEIL1, NEIL2, and NEIL3. The preferred substrates of NEIL1 and NEIL2 are thymine-glycol (Tg), 5-hydroxycytosine (5hC), FaPyA, and FaPyG [Hazra et al., 2002; Morland et al., 2002], but also 5-hydroxyuracil (5hU) and 8oxoG in DNA bubble structures [Dou et al., 2003]. NEIL3 excises formamidopyrimidines but is inactive on 8oxoG [Liu et al., 2010]. Notably, NEIL1 and NEIL3 show a higher preference for their substrates in telomeric sequences [Zhou et al., 2013].
NEIL family members exhibit a distinct two-domain architecture where the N- and C-terminal domains are connected by a flexible hinge region [Liu et al., 2013a; Prakash et al., 2012]. The N-terminal domain comprises a two-layered β-sandwich flanked by α-helices, while the predominantly α-helical C-terminal domain contains a conserved helix-two-turn-helix (H2TH) motif, as well as two antiparallel β-strands that in the case of NEIL2 and NEIL 3 form a zinc-

finger motif required for DNA binding [Doublie et al., 2004]. Although they differ in their substrate preferences, both NEIL1 and NEIL2 recognize oxidized pyrimidines. They share a common catalytic proline at the N-terminal position, whereas NEIL3 has a valine at this position [Bandaru et al., 2002].

One of the main questions regarding NEIL glycosylases is how these enzymes locate and recognize a damaged base. The structures of a viral ortholog of human NEIL1 glycosylase complexes revealed that the 5hU-lesion is extruded from the DNA helix and inserted in the active site of the enzyme (Figure 14) [Imamura et al., 2012; Prakash et al., 2013]. Unexpectedly, very few hydrogen bond interactions were observed between the enzyme and either lesion, and mutating two of the residues in contact with the lesion did not affect the glycosylase activity. This finding suggests that lesion recognition is likely performed before the lesion is flipped out from the DNA double helix. It needs to be further investigated whether NEIL glycosylases are also active in the presence of TDG and how they are regulated in the context of DNA demethylation.

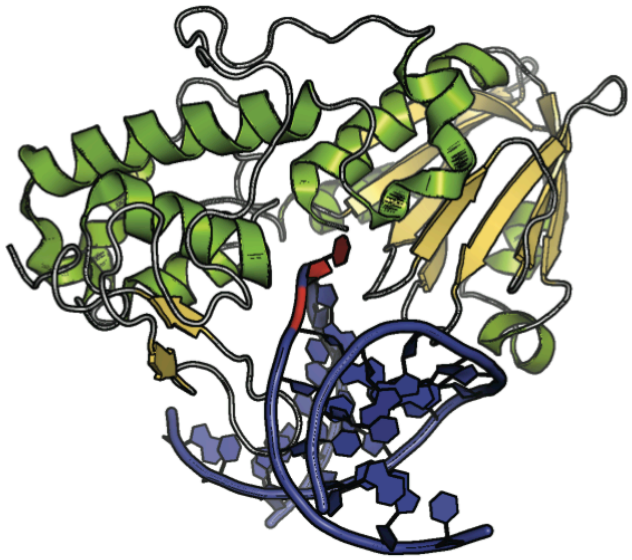


Figure 13: Crystal structure of a viral ortholog of human NEIL1 glycosylase bound to 5hU containing DNA.

NEIL1 is composed of two domains connected by a linker. The N-terminal domain comprises an α-helix followed by a two-layered β-sandwich, with each layer composed of four antiparallel β-strands (yellow). The C-terminal domain comprises several α-helices (green), two of which are involved in the H2TH motif. The DNA is bound in the cleft between the two domains and lies perpendicularly to the long axis of NEIL1. The lesion (red) is flipped out of the DNA double helix and placed in the active site.

Additional human glycosylases and their involvement in DNA demethylation

UNG and SMUG glycosylases are highly specific for processing genomic uracil in mouse and human cells [Haushalter et al., 1999; Nilsen et al., 1997]. Interactions with PCNA target UNG to sites of DNA replication, where its main function is to immediately excise uracil that is incorporated opposite from adenine [Otterlei et al., 1999]. SMUG glycosylases appear to serve as a back-up for UNG in limiting uracil accumulation in genomic DNA and in preventing C → T mutations after cytosine deamination [An et al., 2005]. Notably, SMUG1 has also been shown to contribute to rRNA quality control as it regulates 5hmU levels [Jobert et al., 2013].

Recently, UNG was described to be active on uracil after AID mediated cytosine deamination in mouse zygotes, contributing to active DNA demethylation [Santos et al., 2013]. In line, knockdown of zebrafish *Ung* causes an increase of global DNA methylation levels together with a reduction of overall transcriptional activity [Wu et al., 2014]. However, the AID system is not active in zebrafish embryos before zygotic genome activation [Rai et al., 2008] and also UNG itself did not display any activity towards 5hmU using *in vitro* glycosylase assays. However, we found UNG as a binder of fC in ESCs, which proposes a potential involvement of this glycosylase in DNA demethylation. Interestingly, this finding suggests an additional recognition of oxidized cytosines by UNG besides its already known uracil substrates.

SMUG1 has been implicated in active DNA demethylation as it was shown to excise 5hmU, the deamination product of 5hmC. Although we did not observe a direct interaction of TET proteins with SMUG1, the excision of 5hmU could still occur in *vivo*. However, 5hmU is not generated by deamination of 5hmC but by TET mediated oxidation of thymine (see chapter 3.2). Nevertheless, further biochemical and structural studies are needed to definitely elucidate the role and context of the uracil specific glycosylases in DNA demethylation.

OGG1 serves as the main glycosylase for the excision of 8oxoG opposite cytosine in mammals [Radicella et al., 1997]. In addition, it is also active on other oxidized purines or formamido-guanine (FapyG) [Dherin et al., 1999]. Structural data indicate that the catalytic core of OGG1 is very specific for guanine and it is thus unlikely that cytosine derivatives are processed [Banerjee et al., 2005; Bjoras et al., 2002]. Although, we did not observe any interaction with the TET proteins using F3H [Luitz, 2013], we identified OGG1 as a 5hmC specific reader in brain. Further studies are needed to clarify to what extend OGG1 can process oxidized cytosine bases.

MPG glycosylase excises a variety of alkylated bases, including methylated adenine or guanine, as well as ethylated bases in single- and double-stranded DNA [Lee et al., 2009; O'Connor, 1993]. Chrystal structures show that the base-flipping mechanism used by MPG can only be applied on mutated purine bases [Lau et al., 1998] and its role in removing oxidized cytosines is therefore unlikely. Recently, it was shown that UHRF1 interacts with MPG in cancer cells. However, the functional linkage between MPG-mediated DNA damage repair and UHRF1/2-mediated DNA methylation is not known yet [Liang et al., 2013]. MPG was found as a 5hmC reader in ESCs but we did not observe interactions with the TET proteins [Luitz, 2013]. Thus, its involvement in DNA demethylation remains unclear.

Dynamic regulation of gene expression by TETs and TDG

We could show that TET-dependent oxidation leads to DNA demethylation via BER, with TDG being the main glycosylase involved in this process [Muller et al., 2014].

First studies on DNA demethylation were performed in developing chicken embryos, in which hormone induced gene reactivation was observed independent of DNA replication [Wilks et al., 1984]. Further analysis revealed that TDG was contributing to the observed DNA demethylation effect and a potential activity against 5mC was proposed [Jost, 1993; Jost et al., 1995]. A more detailed screen showed TDG enrichment inside loci at which rapid cycling of C and 5mC is associated with hormonal or cytokine-mediated regulation [Kangaspeska et al., 2008; Metivier et al., 2008; Thillainadesan et al., 2012]. These initial studies imply that active demethylation via TDG may be essential when transcriptional control must be modulated in the absence of DNA replication.

New genome-wide and single-base resolution methods have been developed to discriminate between all known cytosine bases. Several studies have demonstrated a probable regulatory role for 5hmC with significant enrichment for this modification at transcribed gene bodies, bivalent and silent promoters, and distal *cis*-regulatory elements [Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011]. A recent mapping in ESCs has shown the functional relevance of 5fC and 5caC together with TDG. In TDG knockout ESCs, a significant enrichment of 5fC and 5caC was detected in non-repetitive regions, mainly at distal regulatory elements. However, in wild-type ESCs, 5fC or 5caC accumulates at major satellite repeats but not at nonrepetitive loci [Shen et al., 2013]. Furthermore, 5fC and 5caC are transiently accumulated during lineage specification of neural stem cells (NSCs). 5caC is mainly found at the cell-type-specific promoters during differentiation of NSCs, and TDG knockdown leads to increased 5fC/5caC levels in these cells [Wheldon et al., 2014].

Other studies further demonstrated 5fC enrichment in enhancer regions and in CpG islands of promoters and exons [Song et al., 2013]. In line with our results, CGI promoters in which 5fC was relatively more enriched than 5mC or 5hmC corresponded to transcriptionally active genes. Accordingly, 5fC-rich promoters had elevated H3K4me3 levels, associated with active transcription, and were frequently bound by RNA polymerase II. TDG down-regulation led to 5fC accumulation in CGIs in ESCs, which correlates with increased methylation in these genomic regions during differentiation in wild-type and TDG knockout contexts [Raiber et al., 2012]. Thus, 5fC excision in ESCs is necessary for the correct establishment of CGI methylation patterns during differentiation and hence for appropriate patterns of gene expression during development.

Another study connecting TET and TDG functions revealed, that TET-deficient MEFs cannot be reprogrammed because of a block in the mesenchymal-to-epithelial transition step. Interestingly, reprogramming of MEFs deficient in TDG is similarly impaired [Hu et al., 2014]. Thus, oxidative

demethylation to promote gene activation appears to be functionally required for reprogramming of fibroblasts to pluripotent cells. These findings provide new mechanistic insights into the role of epigenetic barriers in cell-lineage conversion.

Recently, it was indicated that TET proteins can be post-transcriptionally regulated by microRNAs. The miR-29-family of microRNAs downregulate TET1 and TDG without affecting DNA methylation, possibly because it can also repress DNMT3A and DNMT3B [Morita et al., 2013]. MiR-26 was shown to inhibit activity of all three TETs and TDG, emphasizing a complex regulation of DNA demethylation enzymes by small RNAs [Fu et al., 2013].

Concerning cytosine oxidation as a step-wise modification process the key question arises, what regulates TET enzymes in stalling at specific intermediates in this pathway or promotes progression through all three catalytic oxidation steps?

As 5hmC levels are higher than 5fC and 5caC and given that TET enzymes can iteratively oxidize these bases, it remains elusive what factors control the modification pathway to stop at 5hmC. Stalling at the 5hmC modification could be regulated through the accessibility of 5hmC by TET proteins or availability of the cofactors 2-OG and Fe(II), which leads to altered enzyme kinetics. Other mechanisms could include either post-translational modifications or interaction with other factors. The basal reactivity of TET with each of its substrates and regulation of its substrate preferences are crucial to investigate. First structural insights into the catalytic domain of TET2 have already started to give first insights into the underlying mechanisms. However, the role of the large N-terminal domain or the insert, which separates the DSBH is unknown yet.

At the next step of the oxidation pathway, it is still not understood whether 5fC and 5caC have different roles, as they can be both excised by TDG. Furthermore, it remains elusive what significance they have beyond serving as intermediates in active DNA demethylation. Although they are very rare in the genome, we could show that they are both bound by different readers, which might indicate additional regulatory roles [Spruijt et al., 2013].

Finally, when the pathway is viewed as a complete cycle of cytosine methylation, oxidation and repair, it immediately begs the question: how does recurrent cycling of this pathway happen? However, evidence of multiple cycling events through a methylation-demethylation pathway at a single locus needs to be investigated.

Taken together, TET mediated oxidation and TDG dependent demethylation seems to be highly dynamic and a complex regulation ensures the control of gene expression and genomic plasticity.

Conclusion and perspectives

Since their discovery in 2009, TET enzymes and their mode of action have dramatically changed the views about epigenetic regulation and chromatin dynamics. Several key studies showed the importance of active or passive TET-mediated demethylation mechanisms in different biological processes such as ESC reprogramming, differentiation but also cancer development and progression. TET proteins are connected with diverse chromatin-related machineries involved in transcriptional regulation or DNA repair. The epigenetic marks they establish can not only act as distinctive signals but also serve as DNA-demethylation intermediates. Besides, TET proteins have additional functions that do not depend on their catalytic activity, such as the recruitment of chromatin associated proteins. Oxidative modifications of 5mC or thymine, generated by TETs and DNA demethylation mechanisms have greatly expanded the possibilities by which the genome can be flexible while maintaining the integrity of its coding information. TET proteins are regulated at various levels including metabolites, miRNAs and post-translational modifications. Furthermore, the increasing numbers of identified interactors could contribute to understanding how these different complex epigenetic processes are interconnected with each other (Figure 15).

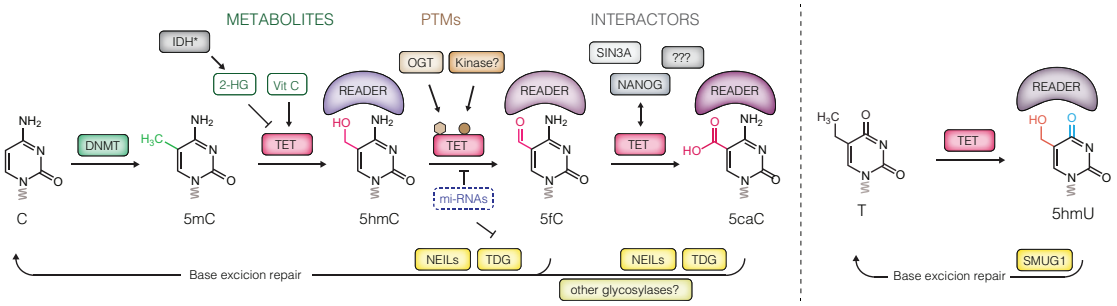


Figure 15: The expanding regulation and functional network of TET proteins and oxidized base derivatives

The activity and localization of TET proteins can be regulated at different levels. Metabolites such as Vitamin C can increase TET activity, whereas 2-HG, generated by mutated IDH can inhibit TET function. Similarly, miRNAs can also affect TET-mediated oxidation by direct downregulation of TET expression or inhibit DNA demethylation by interfering with TDG. TET proteins can be further regulated by PTMs. Glycosylation by OGT has been shown to alter TET localization and thus its activity. Several proteins have been shown to interact with TET enzymes, some controlling reprogramming or differentiation or modulating chromatin states. Some factors also seem to affect TET activity via indirect mechanisms. Readers of oxidized cytosines are highly dynamic and show very little overlap. However, their exact functions still need to be investigated. The ability of TET proteins to oxidize thymine into 5hmU expands the epigenetic network and possibilities to fine-tune gene expression. Active DNA demethylation is mainly achieved by base-excision repair. The main glycosylase initiating this pathway is TDG, however NEIL glycosylases and maybe also others can serve as backup.

Future research will shed new light on the role of dynamic regulation of DNA modifications in pluripotency, development, gene regulation and genome stability.

4 ANNEX

4.1 References

Achour, M., Fuhrmann, G., Alhosin, M., Ronde, P., Chataigneau, T., Mousli, M., Schini-Kerth, V.B., and Bronner, C. 2009. UHRF1 recruits the histone acetyltransferase Tip60 and controls its expression and activity. *Biochemical and biophysical research communications* **390**, 523-528.

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 research* **35**, 5402-5408.

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. *Nature genetics* **23**, 185-188.

An, Q., Robins, P., Lindahl, T., and Barnes, D.E. 2005. C --> T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *The EMBO journal* **24**, 2205-2213.

Andersson, T., Sodersten, E., Duckworth, J.K., Cascante, A., Fritz, N., Sacchetti, P., Cervenka, I., Bryja, V., and Hermanson, O. 2009. CXXC5 is a novel BMP4-regulated modulator of Wnt signaling in neural stem cells. *The Journal of biological chemistry* **284**, 3672-3681.

Aoki, A., Suetake, I., Miyagawa, J., Fujio, T., Chijiwa, T., Sasaki, H., and Tajima, S. 2001. Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic acids research* **29**, 3506-3512.

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V.*, et al.* 2012. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS genetics* **8**, e1002750.

Aravind, L., Anand, S., and Iyer, L.M. 2013. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biology direct* **8**, 20.

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

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. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 12950-12955.

Augui, S., Nora, E.P., and Heard, E. 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature reviews Genetics* **12**, 429-442.

Avery, O.T., Macleod, C.M., and McCarty, M. 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine* **79**, 137-158.

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.

Bandaru, V., Sunkara, S., Wallace, S.S., and Bond, J.P. 2002. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. *DNA repair* **1**, 517-529.

Banerjee, A., Santos, W.L., and Verdine, G.L. 2006. Structure of a DNA glycosylase searching for lesions. *Science* **311**, 1153-1157.

Banerjee, A., Yang, W., Karplus, M., and Verdine, G.L. 2005. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612-618.

Bannister, A.J., and Kouzarides, T. 2011. Regulation of chromatin by histone modifications. *Cell research* **21**, 381-395.

Barckhausen, C., Roos, W.P., Naumann, S.C., and Kaina, B. 2014. Malignant melanoma cells acquire resistance to DNA interstrand cross-linking chemotherapeutics by p53-triggered upregulation of DDB2/XPC-mediated DNA repair. *Oncogene* **33**, 1964-1974.

Barr, H., Hermann, A., Berger, J., Tsai, H.H., Adie, K., Prokhortchouk, A., Hendrich, B., and Bird, A. 2007. Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Molecular and cellular biology* **27**, 3750-3757.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F*., et al.* 2007. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671-675.

Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. 2014. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *The Journal of biological chemistry* **289**, 4106-4115.

Bashtrykov, P., Ragozin, S., and Jeltsch, A. 2012. Mechanistic details of the DNA recognition by the Dnmt1 DNA methyltransferase. *FEBS letters* **586**, 1821-1823.

Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. 2013. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480-492.

Bauer, C., Göbel, K., N., N., Colantuoni, C., Wang, M., Müller, U., Kremmer, E., Rottach, A., and Leonhardt, H. 2014. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the glycosyltransferase OGT. *The Journal of biological chemistry*.

Bauer, M.J., and Fischer, R.L. 2011. Genome demethylation and imprinting in the endosperm. *Current opinion in plant biology* **14**, 162-167.

Baylin, S.B., and Jones, P.A. 2011. A decade of exploring the cancer epigenome - biological and translational implications. *Nature reviews Cancer* **11**, 726-734.

Becker, P.B. 2006. Gene regulation: a finger on the mark. *Nature* **442**, 31-32.

Belousova, E.A., Vasil'eva, I.A., Moor, N.A., Zatsepin, T.S., Oretskaya, T.S., and Lavrik, O.I. 2013. Clustered DNA lesions containing 5-formyluracil and AP site: repair via the BER system. *PloS one* **8**, e68576.

Bennett, M.T., Rodgers, M.T., Hebert, A.S., Ruslander, L.E., Eisele, L., and Drohat, A.C. 2006. Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *Journal of the American Chemical Society* **128**, 12510-12519.

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. *The Journal of biological chemistry* **289**, 379-386.

Berti, M., Ray Chaudhuri, A., Thangavel, S., Gomathinayagam, S., Kenig, S., Vujanovic, M., Odreman, F., Glatter, T., Graziano, S., Mendoza-Maldonado, R*., et al.* 2013. Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. *Nature structural & molecular biology* **20**, 347-354.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of molecular biology* **203**, 971-983.

Bestor, T.H. 1992. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *The EMBO journal* **11**, 2611-2617.

Bhattacharya, S.K., Ramchandani, S., Cervoni, N., and Szyf, M. 1999. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* **397**, 579-583.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042-1047.

Billon, P., and Cote, J. 2013. Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance. *Biochimica et biophysica acta* **1819**, 290-302.

Bjelland, S., Anensen, H., Knaevelsrud, I., and Seeberg, E. 2001. Cellular effects of 5-formyluracil in DNA. *Mutation research* **486**, 147-154.

Bjoras, M., Seeberg, E., Luna, L., Pearl, L.H., and Barrett, T.E. 2002. Reciprocal "flipping" underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase. *Journal of molecular biology* **317**, 171-177.

Blume-Jensen, P., and Hunter, T. 2001. Oncogenic kinase signalling. *Nature* **411**, 355-365.

Bonisch, C., and Hake, S.B. 2012. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic acids research* **40**, 10719-10741.

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

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4102-4107.

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.

Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. 2014. Identification of the glucosyltransferase that converts hydroxymethyluracil to base j in the trypanosomatid genome. *The Journal of biological chemistry* **289**, 20273-20282.

Butter, F., Scheibe, M., Morl, M., and Mann, M. 2009. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10626-10631.

Cardoso, M.C., and Leonhardt, H. 1999. DNA methyltransferase is actively retained in the cytoplasm during early development. *The Journal of cell biology* **147**, 25-32.

Chan, D.W., Wang, Y., Wu, M., Wong, J., Qin, J., and Zhao, Y. 2009. Unbiased proteomic screen for binding proteins to modified lysines on histone H3. *Proteomics* **9**, 2343-2354.

Chedin, F., Lieber, M.R., and Hsieh, C.L. 2002. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16916-16921.

Chen, C.C., Wang, K.Y., and Shen, C.K. 2012. The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *The Journal of biological chemistry* **287**, 33116-33121.

Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z.*, et al.* 2013a. Vitamin C modulates TET1 function during somatic cell reprogramming. *Nature genetics* **45**, 1504-1509.

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

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and cellular biology* **23**, 5594-5605.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117.

Cheng, J., Yang, Y., Fang, J., Xiao, J., Zhu, T., Chen, F., Wang, P., Li, Z., Yang, H., and Xu, Y. 2013. Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. *The Journal of biological chemistry* **288**, 1329-1339.

Cheng, X., and Blumenthal, R.M. 2008. Mammalian DNA methyltransferases: a structural perspective. *Structure* **16**, 341-350.

Cheng, X., and Hart, G.W. 2001. Alternative O-glycosylation/O-phosphorylation of serine-16 in murine estrogen receptor beta: post-translational regulation of turnover and transactivation activity. *The Journal of biological chemistry* **276**, 10570-10575.

Choy, M.K., Movassagh, M., Goh, H.G., Bennett, M.R., Down, T.A., and Foo, R.S. 2010. Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC genomics* **11**, 519.

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.

Cliffe, L.J., Hirsch, G., Wang, J., Ekanayake, D., Bullard, W., Hu, M., Wang, Y., and Sabatini, R. 2012. JBP1 and JBP2 proteins are Fe2+/2-oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. *The Journal of biological chemistry* **287**, 19886-19895.

Cole, R.N., and Hart, G.W. 1999. Glycosylation sites flank phosphorylation sites on synapsin I: O-linked N-acetylglucosamine residues are localized within domains mediating synapsin I interactions. *Journal of neurochemistry* **73**, 418-428.

Comer, F.I., and Hart, G.W. 2001. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* **40**, 7845-7852.

Conticello, S.G. 2008. The AID/APOBEC family of nucleic acid mutators. *Genome biology* **9**, 229.

Cooper, D.N., Taggart, M.H., and Bird, A.P. 1983. Unmethylated domains in vertebrate DNA. *Nucleic acids research* **11**, 647-658.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* 2011. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419-423.

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.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, 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.

Cramer, J.M., Scarsdale, J.N., Walavalkar, N.M., Buchwald, W.A., Ginder, G.D., and Williams, D.C., Jr. 2014. Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *The Journal of biological chemistry* **289**, 1294-1302.

Crick, F. 1970. Central dogma of molecular biology. *Nature* **227**, 561-563.

Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J*., et al.* 2004. Histone deimination antagonizes arginine methylation. *Cell* **118**, 545-553.

da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C*., et al.* 2014. Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Molecular cell* **53**, 301-316.

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.* 2010. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **465**, 966.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M*., et al.* 2013. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Developmental cell* **24**, 310-323.

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

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C., and Pirrotta, V. 2004. Polycomb silencing blocks transcription initiation. *Molecular cell* **13**, 887-893.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. 2001. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes & development* **15**, 2940-2944.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Mendez, 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. *The EMBO journal* **32**, 645-655.

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. *The Journal of biological chemistry* **285**, 26114-26120.

Dherin, C., Radicella, J.P., Dizdaroglu, M., and Boiteux, S. 1999. Excision of oxidatively damaged DNA bases by the human alpha-hOgg1 protein and the polymorphic alpha-hOgg1(Ser326Cys) protein which is frequently found in human populations. *Nucleic acids research* **27**, 4001-4007.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P*., et al.* 2013. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371-376.

DiPaolo, C., Kieft, R., Cross, M., and Sabatini, R. 2005. Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Molecular cell* **17**, 441-451.

Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A*., et al.* 2012. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652-655.

Dou, H., Mitra, S., and Hazra, T.K. 2003. Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *The Journal of biological chemistry* **278**, 49679-49684.

Doublie, S., Bandaru, V., Bond, J.P., and Wallace, S.S. 2004. The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10284-10289.

Du, Z., Song, J., Wang, Y., Zhao, Y., Guda, K., Yang, S., Kao, H.Y., Xu, Y., Willis, J., Markowitz, S.D*., et al.* 2010. DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Science signaling* **3**, ra80.

Dunn, A.R., Kad, N.M., Nelson, S.R., Warshaw, D.M., and Wallace, S.S. 2011. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic acids research* **39**, 7487-7498.

Easwaran, H.P., Schermelleh, L., Leonhardt, H., and Cardoso, M.C. 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.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. 2008. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO journal* **27**, 406-420.

Edwards, C.A., and Ferguson-Smith, A.C. 2007. Mechanisms regulating imprinted genes in clusters. *Current opinion in cell biology* **19**, 281-289.

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 research* **10**, 2709-2721.

Elder, R.H., Jansen, J.G., Weeks, R.J., Willington, M.A., Deans, B., Watson, A.J., Mynett, K.J., Bailey, J.A., Cooper, D.P., Rafferty, J.A.*, et al.* 1998. Alkylpurine-DNA-N-glycosylase knockout mice show increased susceptibility to induction of mutations by methyl methanesulfonate. *Molecular and cellular biology* **18**, 5828-5837.

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D.A. 2009. Conserved DNA methylation in Gadd45a(-/-) mice. *Epigenetics : official journal of the DNA Methylation Society* **4**, 98-99.

Esteve, P.O., Chang, Y., Samaranayake, M., Upadhyay, A.K., Horton, J.R., Feehery, G.R., Cheng, X., and Pradhan, S. 2011. A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. *Nature structural & molecular biology* **18**, 42-48.

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. *Journal of molecular biology* **309**, 1189-1199.

Felle, M., Joppien, S., Nemeth, A., Diermeier, S., Thalhammer, V., Dobner, T., Kremmer, E., Kappler, R., and Langst, G. 2011. The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic acids research* **39**, 8355-8365.

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. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8689-8694.

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.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116-1122.

Fortini, P., and Dogliotti, E. 2007. Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA repair* **6**, 398-409.

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. 2011a. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PloS one* **6**, e21306.

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F., *et al.* 2011b. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PloS one* **6**, e16627.

Fu, X., Jin, L., Wang, X., Luo, A., Hu, J., Zheng, X., Tsark, W.M., Riggs, A.D., Ku, H.T., and Huang, W. 2013. MicroRNA-26a targets ten eleven translocation enzymes and is regulated during pancreatic cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 17892-17897.

Fujiki, R., Hashiba, W., Sekine, H., Yokoyama, A., Chikanishi, T., Ito, S., Imai, Y., Kim, J., He, H.H., Igarashi, K., *et al.* 2011. GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* **480**, 557-560.

Fujimori, A., Matsuda, Y., Takemoto, Y., Hashimoto, Y., Kubo, E., Araki, R., Fukumura, R., Mita, K., Tatsumi, K., and Muto, M. 1998. Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell proliferation. *Mammalian genome : official journal of the International Mammalian Genome Society* **9**, 1032-1035.

Fuks, F., Burgers, W.A., Godin, N., Kasai, M., and Kouzarides, T. 2001. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *The EMBO journal* **20**, 2536-2544.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. 2003. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic acids research* **31**, 2305-2312.

Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. 2000. Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916-921.

Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., *et al.* 2013. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell stem cell* **12**, 453-469.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., and Jaenisch, R. 2003. Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489-492.

Gaudet, F., Rideout, W.M., 3rd, Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. 2004. Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. *Molecular and cellular biology* **24**, 1640-1648.

Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A., and Xu, G.L. 2004. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *The Journal of biological chemistry* **279**, 25447-25454.

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A., and Hansen, J.C. 2003. Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *The Journal of biological chemistry* **278**, 32181-32188.

Globisch, D., Munzel, M., Muller, M., Michalakis, S., Wagner, M., Koch, S., Bruckl, T., Biel, M., and Carell, T. 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PloS one* **5**, e15367.

Goldberg, A.D., Allis, C.D., and Bernstein, E. 2007. Epigenetics: a landscape takes shape. *Cell* **128**, 635-638.

Goll, M.G., and Bestor, T.H. 2005. Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* **74**, 481-514.

Gommers-Ampt, J.H., Van Leeuwen, F., de Beer, A.L., Vliegenthart, J.F., Dizdaroglu, M., Kowalak, J.A., Crain, P.F., and Borst, P. 1993. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. *Cell* **75**, 1129-1136.

Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. 2005. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *The Journal of biological chemistry* **280**, 13341-13348.

Goyal, R., Reinhardt, R., and Jeltsch, A. 2006. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic acids research* **34**, 1182-1188.

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.

Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E., and Young, R.A. 2005. Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8603-8608.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., *et al.* 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature neuroscience* **17**, 215-222.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. 2011a. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423-434.

Guo, X., Shi, M., Sun, L., Wang, Y., Gui, Y., Cai, Z., and Duan, X. 2011b. The expression of histone demethylase JMJD1A in renal cell carcinoma. *Neoplasma* **58**, 153-157.

Guz, J., Gackowski, D., Foksinski, M., Rozalski, R., and Olinski, R. 2014. Comparison of the Absolute Level of Epigenetic Marks--5-Methylcytosine, 5-Hydroxymethylcytosine, and 5-Hydroxymethyluracil Between Human Leukocytes and Sperm. *Biology of reproduction*.

Hameed, U.F., Lim, J., Zhang, Q., Wasik, M.A., Yang, D., and Swaminathan, K. 2014. Transcriptional repressor domain of MBD1 is intrinsically disordered and interacts with its binding partners in a selective manner. *Scientific reports* **4**, 4896.

Handa, V., and Jeltsch, A. 2005. Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *Journal of molecular biology* **348**, 1103-1112.

Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. 2003. The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic acids research* **31**, 2261-2271.

Hashimoto, H., Hong, S., Bhagwat, A.S., Zhang, X., and Cheng, X. 2012a. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic acids research* **40**, 10203-10214.

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. 2012b. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic acids research* **40**, 4841-4849.

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.

Hashimoto, H., Zhang, X., and Cheng, X. 2012c. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic acids research* **40**, 8276-8284.

Hashimoto, H., Zhang, X., and Cheng, X. 2013. Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH. *DNA repair* **12**, 535-540.

Haushalter, K.A., Todd Stukenberg, M.W., Kirschner, M.W., and Verdine, G.L. 1999. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Current biology : CB* **9**, 174-185.

Hazra, T.K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y.W., Jaruga, P., Dizdaroglu, M., and Mitra, S. 2002. Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3523-3528.

He, J., Kallin, E.M., Tsukada, Y., and Zhang, Y. 2008. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). *Nature structural & molecular biology* **15**, 1169-1175.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. 2013. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nature cell biology* **15**, 373-384.

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.

Heidebrecht, T., Christodoulou, E., Chalmers, M.J., Jan, S., Ter Riet, B., Grover, R.K., Joosten, R.P., Littler, D., van Luenen, H., Griffin, P.R.*, et al.* 2011. The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic acids research* **39**, 5715-5728.

Hellman, A., and Chess, A. 2007. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141-1143.

Hendrich, B., and Bird, A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology* **18**, 6538-6547.

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

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

Herrero Sanchez, C., Reith, W., Silacci, P., and Mach, B. 1992. The DNA-binding defect observed in major histocompatibility complex class II regulatory mutants concerns only one member of a family of complexes binding to the X boxes of class II promoters. *Molecular and cellular biology* **12**, 4076-4083.

Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. 2008. MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Molecular cell* **29**, 525-531.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., and Chaillet, J.R. 2001. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **104**, 829-838.

Hsu, Y.C., Kao, C.Y., Chung, Y.F., Chen, M.S., and Chiu, I.M. 2012. Ciliogenic RFX transcription factors regulate FGF1 gene promoter. *Journal of cellular biochemistry* **113**, 2511-2522.

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.

Hu, L., Li, Z., Wang, P., Lin, Y., and Xu, Y. 2011. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell research* **21**, 1374-1378.

Hu, X., Zhang, L., Mao, S.Q., Li, Z., Chen, J., Zhang, R.R., Wu, H.P., Gao, J., Guo, F., Liu, W.*, et al.* 2014. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell stem cell* **14**, 512-522.

Imamura, K., Averill, A., Wallace, S.S., and Doublie, S. 2012. Structural characterization of viral ortholog of human DNA glycosylase NEIL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA. *The Journal of biological chemistry* **287**, 4288-4298.

Indiani, C., and O'Donnell, M. 2006. The replication clamp-loading machine at work in the three domains of life. *Nature reviews Molecular cell biology* **7**, 751-761.

Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. 2011. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell research* **21**, 1670-1676.

Inoue, A., and Zhang, Y. 2011. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194.

Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3642-3647.

Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., and Yokoyama, A. 2014. TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes to cells : devoted to molecular & cellular mechanisms* **19**, 52-65.

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.

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. 2013. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome biology* **14**, R119.

Iwata, A., Nagashima, Y., Matsumoto, L., Suzuki, T., Yamanaka, T., Date, H., Deoka, K., Nukina, N., and Tsuji, S. 2009. Intranuclear degradation of polyglutamine aggregates by the ubiquitin-proteasome system. *The Journal of biological chemistry* **284**, 9796-9803.

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.

Iyer, L.M., Zhang, D., Burroughs, A.M., and Aravind, L. 2013. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic acids research* **41**, 7635-7655.

Jacobs, A.L., and Schar, P. 2012. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1-20.

Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eissenberg, J.C., Allis, C.D., and Khorasanizadeh, S. 2001. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *The EMBO journal* **20**, 5232-5241.

Jaenisch, R., and Bird, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* **33 Suppl**, 245-254.

Jeltsch, A., and Jurkowska, R.Z. 2014. New concepts in DNA methylation. *Trends in biochemical sciences* **39**, 310-318.

Jeon, B.N., Choi, W.I., Yu, M.Y., Yoon, A.R., Kim, M.H., Yun, C.O., and Hur, M.W. 2009. ZBTB2, a novel master regulator of the p53 pathway. *The Journal of biological chemistry* **284**, 17935-17946.

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, S.G., Guo, C., and Pfeifer, G.P. 2008. GADD45A does not promote DNA demethylation. *PLoS genetics* **4**, e1000013.

Jobert, L., Skjeldam, H.K., Dalhus, B., Galashevskaya, A., Vagbo, C.B., Bjoras, M., and Nilsen, H. 2013. The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control. *Molecular cell* **49**, 339-345.

Jones, P.A., and Liang, G. 2009. Rethinking how DNA methylation patterns are maintained. *Nature reviews Genetics* **10**, 805-811.

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. *Nature genetics* **19**, 187-191.

Jorgensen, H.F., Ben-Porath, I., and Bird, A.P. 2004. Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Molecular and cellular biology* **24**, 3387-3395.

Jost, J.P. 1993. Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4684-4688.

Jost, J.P., Schwarz, S., Hess, D., Angliker, H., Fuller-Pace, F.V., Stahl, H., Thiry, S., and Siegmann, M. 1999. A chicken embryo protein related to the mammalian DEAD box protein p68 is tightly associated with the highly purified protein-RNA complex of 5-MeC-DNA glycosylase. *Nucleic acids research* **27**, 3245-3252.

Jost, J.P., Siegmann, M., Sun, L., and Leung, R. 1995. Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. *The Journal of biological chemistry* **270**, 9734-9739.

Kacem, S., and Feil, R. 2009. Chromatin mechanisms in genomic imprinting. *Mammalian genome : official journal of the International Mammalian Genome Society* **20**, 544-556.

Kamemura, K., Hayes, B.K., Comer, F.I., and Hart, G.W. 2002. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *The Journal of biological chemistry* **277**, 19229-19235.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. 2008. Transient cyclical methylation of promoter DNA. *Nature* **452**, 112-115.

Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P.A., Hagen, L., Krokan, H.E., and Slupphaug, G. 2002. hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *The Journal of biological chemistry* **277**, 39926-39936.

Kim, G.D., Ni, J., Kelesoglu, N., Roberts, R.J., and Pradhan, S. 2002. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *The EMBO journal* **21**, 4183-4195.

Kim, H.Y., Yang, D.H., Shin, S.W., Kim, M.Y., Yoon, J.H., Kim, S., Park, H.C., Kang, D.W., Min, D., Hur, M.W*., et al.* 2014. CXXC5 is a transcriptional activator of Flk-1 and mediates bone morphogenic protein-induced endothelial cell differentiation and vessel formation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **28**, 615-626.

Kim, J.K., Esteve, P.O., Jacobsen, S.E., and Pradhan, S. 2009. UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic acids research* **37**, 493-505.

Kimura, H., and Shiota, K. 2003. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *The Journal of biological chemistry* **278**, 4806-4812.

Kino, K., Shimizu, Y., Sugasawa, K., Sugiyama, H., and Hanaoka, F. 2004. Nucleotide excision repair of 5-formyluracil in vitro is enhanced by the presence of mismatched bases. *Biochemistry* **43**, 2682-2687.

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijo, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lahdesmaki, H*., et al.* 2013. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122-126.

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. 2011. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14566-14571.

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.

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.

Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S.*, et al.* 2011. Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* **25**, 1649-1652.

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* **128**, 693-705.

Kreppel, L.K., Blomberg, M.A., and Hart, G.W. 1997. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *The Journal of biological chemistry* **272**, 9308-9315.

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

Krokan, H.E., Drablos, F., and Slupphaug, G. 2002. Uracil in DNA--occurrence, consequences and repair. *Oncogene* **21**, 8935-8948.

Kumar, R., DiMenna, L., Schrode, N., Liu, T.C., Franck, P., Munoz-Descalzo, S., Hadjantonakis, A.K., Zarrin, A.A., Chaudhuri, J., Elemento, O.*, et al.* 2013. AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature* **500**, 89-92.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120.

Lafaye, C., Barbier, E., Miscioscia, A., Saint-Pierre, C., Kraut, A., Coute, Y., Plo, I., Gasparutto, D., Ravanat, J.L., and Breton, J. 2014. DNA binding of the p21 repressor ZBTB2 is inhibited by cytosine hydroxymethylation. *Biochemical and biophysical research communications* **446**, 341-346.

Landan, G., Cohen, N.M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S., Zalcenstein, D.A., Goldfinger, N., Zundelevich, A.*, et al.* 2012. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nature genetics* **44**, 1207-1214.

Langemeijer, S.M., 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.*, et al.* 2009. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature genetics* **41**, 838-842.

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

Latham, J.A., and Dent, S.Y. 2007. Cross-regulation of histone modifications. *Nature structural & molecular biology* **14**, 1017-1024.

Lau, A.Y., Scharer, O.D., Samson, L., Verdine, G.L., and Ellenberger, T. 1998. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249-258.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. 2013. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021-1036.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J.*, et al.* 2010. Dynamic changes in the human methylome during differentiation. *Genome research* **20**, 320-331.

Lavigne, M., Francis, N.J., King, I.F., and Kingston, R.E. 2004. Propagation of silencing; recruitment and repression of naive chromatin in trans by polycomb repressed chromatin. *Molecular cell* **13**, 415-425.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A., Cohen, A., Lasonder, E., and Stunnenberg, H.G. 2006. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and cellular biology* **26**, 843-851.

Lee, B., and Muller, M.T. 2009. SUMOylation enhances DNA methyltransferase 1 activity. *The Biochemical journal* **421**, 449-461.

Lee, C.Y., Delaney, J.C., Kartalou, M., Lingaraju, G.M., Maor-Shoshani, A., Essigmann, J.M., and Samson, L.D. 2009. Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG). *Biochemistry* **48**, 1850-1861.

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.

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

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

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. 2010. Jarid2 and PRC2, partners in regulating gene expression. *Genes & development* **24**, 368-380.

Li, J., Wang, Q.E., Zhu, Q., El-Mahdy, M.A., Wani, G., Praetorius-Ibba, M., and Wani, A.A. 2006. DNA damage binding protein component DDB1 participates in nucleotide excision repair through DDB2 DNA-binding and cullin 4A ubiquitin ligase activity. *Cancer research* **66**, 8590-8597.

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

Li, Y., Wang, H., Muffat, J., Cheng, A.W., Orlando, D.A., Loven, J., Kwok, S.M., Feldman, D.A., Bateup, H.S., Gao, Q.*, et al.* 2013. Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons. *Cell stem cell* **13**, 446-458.

Li, Y.Q., Zhou, P.Z., Zheng, X.D., Walsh, C.P., and Xu, G.L. 2007. Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic acids research* **35**, 390-400.

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.

Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K.H. 2012. Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* **151**, 1608-1616.

Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q.*, et al.* 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* **150**, 1135-1146.

Liang, C., Zhang, X., Song, S., Tian, C., Yin, Y., Xing, G., He, F., and Zhang, L. 2013. Identification of UHRF1/2 as new N-methylpurine DNA glycosylase-interacting proteins. *Biochemical and biophysical research communications* **433**, 415-419.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. 2002. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Molecular and cellular biology* **22**, 480-491.

Lin, I.G., Han, L., Taghva, A., O'Brien, L.E., and Hsieh, C.L. 2002. Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. *Molecular and cellular biology* **22**, 704-723.

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.

Liu, M., Bandaru, V., Bond, J.P., Jaruga, P., Zhao, X., Christov, P.P., Burrows, C.J., Rizzo, C.J., Dizdaroglu, M., and Wallace, S.S. 2010. The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4925-4930.

Liu, M., Imamura, K., Averill, A.M., Wallace, S.S., and Doublie, S. 2013a. Structural characterization of a mouse ortholog of human NEIL3 with a marked preference for single-stranded DNA. *Structure* **21**, 247-256.

Liu, N., Wang, M., Deng, W., Schmidt, C.S., Qin, W., Leonhardt, H., and Spada, F. 2013b. Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PloS one* **8**, e62755.

Liu, Y., Olanrewaju, Y.O., Zheng, Y., Hashimoto, H., Blumenthal, R.M., Zhang, X., and Cheng, X. 2014. Structural basis for Klf4 recognition of methylated DNA. *Nucleic acids research* **42**, 4859-4867.

Liutkeviciute, Z., Kriukiene, E., Licyte, J., Rudyte, M., Urbanaviciute, G., and Klimasauskas, S. 2014. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *Journal of the American Chemical Society* **136**, 5884-5887.

Loenarz, C., and Schofield, C.J. 2009. Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chemistry & biology* **16**, 580-583.

Long, H.K., Blackledge, N.P., and Klose, R.J. 2013. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society transactions* **41**, 727-740.

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.

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 A resolution. *Nature* **389**, 251-260.

Luhnsdorf, B., Epe, B., and Khobta, A. 2014. Excision of uracil from transcribed DNA negatively affects the gene expression. *The Journal of biological chemistry*.

Luitz, J. 2013. Cloning and characterization of the DNA glycosylases OGG1 and MPG.

Maga, G., and Hubscher, U. 2003. Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *Journal of cell science* **116**, 3051-3060.

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. *The Journal of biological chemistry* **286**, 35334-35338.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. 2002. The protein kinase complement of the human genome. *Science* **298**, 1912-1934.

Margot, J.B., Aguirre-Arteta, A.M., Di Giacco, B.V., Pradhan, S., Roberts, R.J., Cardoso, M.C., and Leonhardt, H. 2000. Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *Journal of molecular biology* **297**, 293-300.

Marin, M., Rose, K.M., Kozak, S.L., and Kabat, D. 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature medicine* **9**, 1398-1403.

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

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

Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E., 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.

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417-1430.

Menafra, R., Brinkman, A.B., Matarese, F., Franci, G., Bartels, S.J., Nguyen, L., Shimbo, T., Wade, P.A., Hubner, N.C., and Stunnenberg, H.G. 2014. Genome-Wide Binding of MBD2 Reveals Strong Preference for Highly Methylated Loci. *PloS one* **9**, e99603.

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

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

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. 2002. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Molecular cell* **10**, 1107-1117.

Miniou, P., Jeanpierre, M., Blanquet, V., Sibella, V., Bonneau, D., Herbelin, C., Fischer, A., Niveleau, A., and Viegas-Pequignot, E. 1994. Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. *Human molecular genetics* **3**, 2093-2102.

Mittler, G., Butter, F., and Mann, M. 2009. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome research* **19**, 284-293.

Monnier, P., Martinet, C., Pontis, J., Stancheva, I., Ait-Si-Ali, S., and Dandolo, L. 2013. H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20693-20698.

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.

Morey, L., Brenner, C., Fazi, F., Villa, R., Gutierrez, A., Buschbeck, M., Nervi, C., Minucci, S., Fuks, F., and Di Croce, L. 2008. MBD3, a component of the NuRD complex, facilitates chromatin alteration and deposition of epigenetic marks. *Molecular and cellular biology* **28**, 5912-5923.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *The Journal of biological chemistry* **279**, 52353-52360.

Morgan, M.T., Bennett, M.T., and Drohat, A.C. 2007. Excision of 5-halogenated uracils by human thymine DNA glycosylase. Robust activity for DNA contexts other than CpG. *The Journal of biological chemistry* **282**, 27578-27586.

Morgan, M.T., Maiti, A., Fitzgerald, M.E., and Drohat, A.C. 2011. Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. *Nucleic acids research* **39**, 2319-2329.

Mori, T., Ikeda, D.D., Yamaguchi, Y., Unoki, M., and Project, N. 2012. NIRF/UHRF2 occupies a central position in the cell cycle network and allows coupling with the epigenetic landscape. *FEBS letters* **586**, 1570-1583.

Morita, S., Horii, T., Kimura, M., Ochiya, T., Tajima, S., and Hatada, I. 2013. miR-29 represses the activities of DNA methyltransferases and DNA demethylases. *International journal of molecular sciences* **14**, 14647-14658.

Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjoras, M., and Seeberg, E. 2002. Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic acids research* **30**, 4926-4936.

Morrison, J.R., Paszty, C., Stevens, M.E., Hughes, S.D., Forte, T., Scott, J., and Rubin, E.M. 1996. Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 7154-7159.

Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M.C., and Leonhardt, H. 2005. Recruitment of DNA methyltransferase I to DNA repair sites. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8905-8909.

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 research* **42**, 8592-8604.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553-563.

Muto, M., Kanari, Y., Kubo, E., Takabe, T., Kurihara, T., Fujimori, A., and Tatsumi, K. 2002. Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. *The Journal of biological chemistry* **277**, 34549-34555.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. 2012. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature chemical biology* **8**, 751-758.

Nady, N., Lemak, A., Walker, J.R., Avvakumov, G.V., Kareta, M.S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X.*, et al.* 2011. Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *The Journal of biological chemistry* **286**, 24300-24311.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M., and Canaani, E. 2002. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Molecular cell* **10**, 1119-1128.

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.

Nelson, S.R., Dunn, A.R., Kathe, S.D., Warshaw, D.M., and Wallace, S.S. 2014. Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2091-2099.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H.G., and Oliviero, S. 2013. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. *Cell* **155**, 121-134.

Nicolas, R.H., and Goodwin, G.H. 1996. Molecular cloning of polybromo, a nuclear protein containing multiple domains including five bromodomains, a truncated HMG-box, and two repeats of a novel domain. *Gene* **175**, 233-240.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. 2007. Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Molecular and cellular biology* **27**, 864-877.

Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skorpen, F., and Krokan, H.E. 1997. Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic acids research* **25**, 750-755.

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.

O'Connor, T.R. 1993. Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic acids research* **21**, 5561-5569.

Okada, Y., Yamagata, K., Hong, K., Wakayama, T., and Zhang, Y. 2010. A role for the elongator complex in zygotic paternal genome demethylation. *Nature* **463**, 554-558.

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.

Oliver, A.W., Jones, S.A., Roe, S.M., Matthews, S., Goodwin, G.H., and Pearl, L.H. 2005. Crystal structure of the proximal BAH domain of the polybromo protein. *The Biochemical journal* **389**, 657-664.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* **1**, 376-386.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. 2002. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer research* **62**, 4075-4080.

Otani, J., Kimura, H., Sharif, J., Endo, T.A., Mishima, Y., Kawakami, T., Koseki, H., Shirakawa, M., Suetake, I., and Tajima, S. 2013. Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PloS one* **8**, e82961.

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 reports* **10**, 1235-1241.

Otterlei, M., Warbrick, E., Nagelhus, T.A., Haug, T., Slupphaug, G., Akbari, M., Aas, P.A., Steinsbekk, K., Bakke, O., and Krokan, H.E. 1999. Post-replicative base excision repair in replication foci. *The EMBO journal* **18**, 3834-3844.

Papait, R., Pistore, C., Negri, D., Pecoraro, D., Cantarini, L., and Bonapace, I.M. 2007. Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Molecular biology of the cell* **18**, 1098-1106.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. 2011. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200-213.

Parvathaneni, S., Stortchevoi, A., Sommers, J.A., Brosh, R.M., Jr., and Sharma, S. 2013. Human RECQ1 interacts with Ku70/80 and modulates DNA end-joining of double-strand breaks. *PloS one* **8**, e62481.

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

Patil, V., Ward, R.L., and Hesson, L.B. 2014. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics : official journal of the DNA Methylation Society* **9**.

Pavri, R., and Nussenzweig, M.C. 2011. AID targeting in antibody diversity. *Advances in immunology* **110**, 1-26.

Payer, B., and Lee, J.T. 2008. X chromosome dosage compensation: how mammals keep the balance. *Annual review of genetics* **42**, 733-772.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. 2009. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290-1302.

Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. 1972. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *The Biochemical journal* **126**, 781-790.

Petronzelli, F., Riccio, A., Markham, G.D., Seeholzer, S.H., Genuardi, M., Karbowski, M., Yeung, A.T., Matsumoto, Y., and Bellacosa, A. 2000. Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *Journal of cellular physiology* **185**, 473-480.

Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C.A., Hagemeier, C., and Carell, T. 2011. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie* **50**, 7008-7012.

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. *Nature chemical biology* **10**, 574-581.

Pichler, G., Wolf, P., Schmidt, C.S., Meilinger, D., Schneider, K., Frauer, C., Fellinger, 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.

Ponting, C.P., Oliver, P.L., and Reik, W. 2009. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629-641.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101-1105.

Popuri, V., Hsu, J., Khadka, P., Horvath, K., Liu, Y., Croteau, D.L., and Bohr, V.A. 2014. Human RECQL1 participates in telomere maintenance. *Nucleic acids research* **42**, 5671-5688.

Pradhan, M., Esteve, P.O., Chin, H.G., Samaranayke, M., Kim, G.D., and Pradhan, S. 2008. CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry* **47**, 10000-10009.

Prakash, A., Doublie, S., and Wallace, S.S. 2012. The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage. *Progress in molecular biology and translational science* **110**, 71-91.

Prakash, A., Eckenroth, B.E., Averill, A.M., Imamura, K., Wallace, S.S., and Doublie, S. 2013. Structural investigation of a viral ortholog of human NEIL2/3 DNA glycosylases. *DNA repair* **12**, 1062-1071.

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.

Radicella, J.P., Dherin, C., Desmaze, C., Fox, M.S., and Boiteux, S. 1997. Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 8010-8015.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. 2008. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201-1212.

Raiber, E.A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. 2012. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome biology* **13**, R69.

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.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5237-5242.

Rangam, G., Schmitz, K.M., Cobb, A.J., and Petersen-Mahrt, S.K. 2012. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PloS one* **7**, e43279.

Razin, A., and Riggs, A.D. 1980. DNA methylation and gene function. *Science* **210**, 604-610.

Reik, W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-432.

Richmond, T.J., and Davey, C.A. 2003. The structure of DNA in the nucleosome core. *Nature* **423**, 145-150.

Rideout, W.M., 3rd, Coetzee, G.A., Olumi, A.F., and Jones, P.A. 1990. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* **249**, 1288-1290.

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. *Nature structural & molecular biology* **19**, 1155-1160.

Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F., and Leonhardt, H. 2010. The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic acids research* **38**, 1796-1804.

Rottach, A., Leonhardt, H., and Spada, F. 2009. DNA methylation-mediated epigenetic control. *Journal of cellular biochemistry* **108**, 43-51.

Santos, F., Hendrich, B., Reik, W., and Dean, W. 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Developmental biology* **241**, 172-182.

Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., and Dean, W. 2013. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics & chromatin* **6**, 39.

Sarraf, S.A., and Stancheva, I. 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular cell* **15**, 595-605.

Saxonov, S., Berg, P., and Brutlag, D.L. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1412-1417.

Scheibe, M., Butter, F., Hafner, M., Tuschl, T., and Mann, M. 2012. Quantitative mass spectrometry and PAR-CLIP to identify RNA-protein interactions. *Nucleic acids research* **40**, 9897-9902.

Schiesser, S., Hackner, B., Pfaffeneder, T., Muller, M., Hagemeier, C., Truss, M., and Carell, T. 2012. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angewandte Chemie* **51**, 6516-6520.

Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., Alvarez-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 research* **41**, 4860-4876.

Seguin-Estevez, Q., De Palma, R., Krawczyk, M., Leimgruber, E., Villard, J., Picard, C., Tagliamacco, A., Abbate, G., Gorski, J., Nocera, A., *et al.* 2009. The transcription factor RFX protects MHC class II genes against epigenetic silencing by DNA methylation. *Journal of immunology* **183**, 2545-2553.

Shafi, R., Iyer, S.P., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., and Marth, J.D. 2000. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5735-5739.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37-46.

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

Sharma, S., Phatak, P., Stortchevoi, A., Jasin, M., and Larocque, J.R. 2012. RECQ1 plays a distinct role in cellular response to oxidative DNA damage. *DNA repair* **11**, 537-549.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650.

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, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. 2009. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139**, 1303-1314.

Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. 2000. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541-544.

Shi, F.T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M.A., Wan, M., and Songyang, Z. 2013. Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *The Journal of biological chemistry* **288**, 20776-20784.

Shimbo, T., Du, Y., Grimm, S.A., Dhasarathy, A., Mav, D., Shah, R.R., Shi, H., and Wade, P.A. 2013. MBD3 localizes at promoters, gene bodies and enhancers of active genes. *PLoS genetics* **9**, e1004028.

Shirato, H., Ogawa, S., Nakajima, K., Inagawa, M., Kojima, M., Tachibana, M., Shinkai, Y., and Takeuchi, T. 2009. A jumonji (Jarid2) protein complex represses cyclin D1 expression by methylation of histone H3-K9. *The Journal of biological chemistry* **284**, 733-739.

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

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.

Son, J., Shen, S.S., Margueron, R., and Reinberg, D. 2013. Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes & development* **27**, 2663-2677.

Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H.*, et al.* 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678-691.

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, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* **335**, 709-712.

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 research* **33**, 3521-3528.

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.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., *et al.* 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S., Kim, S.Y., Bestor, T.H., Shioda, T., *et al.* 2012. Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nature genetics* **44**, 398-405, S391-392.

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

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. 2005. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213-1226.

Suzuki, M.M., and Bird, A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews Genetics* **9**, 465-476.

Syeda, F., Fagan, R.L., Wean, M., Avvakumov, G.V., Walker, J.R., Xue, S., Dhe-Paganon, S., and Brenner, C. 2011. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *The Journal of biological chemistry* **286**, 15344-15351.

Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., and Leonhardt, H. 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic acids research* **38**, e181.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872.

Takahashi, K., and Yamanaka, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676.

Takeshima, H., Suetake, I., Shimahara, H., Ura, K., Tate, S., and Tajima, S. 2006. Distinct DNA methylation activity of Dnmt3a and Dnmt3b towards naked and nucleosomal DNA. *Journal of biochemistry* **139**, 503-515.

Tatematsu, K.I., Yamazaki, T., and Ishikawa, F. 2000. MBD2-MBD3 complex binds to hemi-methylated DNA and forms a complex containing DNMT1 at the replication foci in late S phase. *Genes to cells : devoted to molecular & cellular mechanisms* **5**, 677-688.

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. *Nature genetics* **46**, 385-388.

Tee, W.W., and Reinberg, D. 2014. Chromatin features and the epigenetic regulation of pluripotency states in ESCs. *Development* **141**, 2376-2390.

Thillainadesan, G., Chitilian, J.M., Isovic, M., Ablack, J.N., Mymryk, J.S., Tini, M., and Torchia, J. 2012. TGF-beta-dependent active demethylation and expression of the p15ink4b tumor suppressor are impaired by the ZNF217/CoREST complex. *Molecular cell* **46**, 636-649.

Turner, D.P., Cortellino, S., Schupp, J.E., Caretti, E., Loh, T., Kinsella, T.J., and Bellacosa, A. 2006. The DNA N-glycosylase MED1 exhibits preference for halogenated pyrimidines and is involved in the cytotoxicity of 5-iododeoxyuridine. *Cancer research* **66**, 7686-7693.

Tzatsos, A., Pfau, R., Kampranis, S.C., and Tsichlis, P.N. 2009. Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2641-2646.

Uemura, T., Kubo, E., Kanari, Y., Ikemura, T., Tatsumi, K., and Muto, M. 2000. Temporal and spatial localization of novel nuclear protein NP95 in mitotic and meiotic cells. *Cell structure and function* **25**, 149-159.

Valinluck, V., Liu, P., Kang, J.I., Jr., Burdzy, A., and Sowers, L.C. 2005. 5-halogenated pyrimidine lesions within a CpG sequence context mimic 5-methylcytosine by enhancing the binding of the methyl-CpG-binding domain of methyl-CpG-binding protein 2 (MeCP2). *Nucleic acids research* **33**, 3057-3064.

Valinluck, V., Tsai, H.H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. 2004. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic acids research* **32**, 4100-4108.

van Leeuwen, F., de Kort, M., van der Marel, G.A., van Boom, J.H., and Borst, P. 1998a. The modified DNA base beta-D-glucosylhydroxymethyluracil confers resistance to micrococcal nuclease and is incompletely recovered by 32P-postlabeling. *Analytical biochemistry* **258**, 223-229.

van Leeuwen, F., Taylor, M.C., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., and Borst, P. 1998b. beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 2366-2371.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K.*, et al.* 2013. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Molecular cell* **49**, 645-656.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G.*, et al.* 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-980.

Vermeulen, M., Hubner, N.C., and Mann, M. 2008. High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Current opinion in biotechnology* **19**, 331-337.

Waddington, C.H. (1957). The strategy of genes: a discussion of some aspects of theoretical biology.

Walters, Z.S., Villarejo-Balcells, B., Olmos, D., Buist, T.W., Missiaglia, E., Allen, R., Al-Lazikani, B., Garrett, M.D., Blagg, J., and Shipley, J. 2014. JARID2 is a direct target of the PAX3-FOXO1 fusion protein and inhibits myogenic differentiation of rhabdomyosarcoma cells. *Oncogene* **33**, 1148-1157.

Wang, C., Shen, J., Yang, Z., Chen, P., Zhao, B., Hu, W., Lan, W., Tong, X., Wu, H., Li, G.*, et al.* 2011a. Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. *Cell research* **21**, 1379-1382.

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y.L., Zhou, B., Han, L., Fargo, D.C.*, et al.* 2014. INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. *Cell stem cell* **14**, 575-591.

Wang, T., Chen, K., Zeng, X., Yang, J., Wu, Y., Shi, X., Qin, B., Zeng, L., Esteban, M.A., Pan, G.*, et al.* 2011b. The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell stem cell* **9**, 575-587.

Ward, P.S., Patel, J., Wise, D.R., Abdel-Wahab, O., Bennett, B.D., Coller, H.A., Cross, J.R., Fantin, V.R., Hedvat, C.V., Perl, A.E.*, et al.* 2010. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer cell* **17**, 225-234.

Watanabe, D., Suetake, I., Tada, T., and Tajima, S. 2002. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mechanisms of development* **118**, 187-190.

Watanabe, S., Ichimura, T., Fujita, N., Tsuruzoe, S., Ohki, I., Shirakawa, M., Kawasuji, M., and Nakao, M. 2003. Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12859-12864.

Watson, J.D., and Crick, F.H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.

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

Wells, L., Kreppel, L.K., Comer, F.I., Wadzinski, B.E., and Hart, G.W. 2004. O-GlcNAc transferase is in a functional complex with protein phosphatase 1 catalytic subunits. *The Journal of biological chemistry* **279**, 38466-38470.

Wheldon, L.M., Abakir, A., Ferjentsik, Z., Dudnakova, T., Strohbuecker, S., Christie, D., Dai, N., Guan, S., Foster, J.M., Correa, I.R., Jr*., et al.* 2014. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell reports* **7**, 1353-1361.

Wienholz, B.L., Kareta, M.S., Moarefi, A.H., Gordon, C.A., Ginno, P.A., and Chedin, F. 2010. DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. *PLoS genetics* **6**, e1001106.

Wilks, A., Seldran, M., and Jost, J.P. 1984. An estrogen-dependent demethylation at the 5' end of the chicken vitellogenin gene is independent of DNA synthesis. *Nucleic acids research* **12**, 1163-1177.

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

Williams, R.T., and Wang, Y. 2012. A density functional theory study on the kinetics and thermodynamics of N-glycosidic bond cleavage in 5-substituted 2'-deoxycytidines. *Biochemistry* **51**, 6458-6462.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., and Migeon, B.R. 1984. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 2806-2810.

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. *Nature communications* **2**, 241.

Wu, D., Chen, L., Sun, Q., Wu, X., Jia, S., and Meng, A. 2014. Uracil-DNA Glycosylase Is Involved in DNA Demethylation and Required for Embryonic Development in the Zebrafish Embryo. *The Journal of biological chemistry* **289**, 15463-15473.

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 & development* **25**, 679-684.

Wu, H., and Zhang, Y. 2014. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45-68.

Wu, X., Johansen, J.V., and Helin, K. 2013. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Molecular cell* **49**, 1134-1146.

Wyatt, M.D., Allan, J.M., Lau, A.Y., Ellenberger, T.E., and Samson, L.D. 1999. 3-methyladenine DNA glycosylases: structure, function, and biological importance. *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**, 668-676.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T*., et al.* 2011a. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer cell* **19**, 17-30.

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

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., *et al.* 2012. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151**, 1200-1213.

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

Yang, H., Liu, Y., Bai, F., Zhang, J.Y., Ma, S.H., Liu, J., Xu, Z.D., Zhu, H.G., Ling, Z.Q., Ye, D., *et al.* 2013. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**, 663-669.

Yang, W.H., Kim, J.E., Nam, H.W., Ju, J.W., Kim, H.S., Kim, Y.S., and Cho, J.W. 2006. Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nature cell biology* **8**, 1074-1083.

Yao, T., Song, L., Jin, J., Cai, Y., Takahashi, H., Swanson, S.K., Washburn, M.P., Florens, L., Conaway, R.C., Cohen, R.E., *et al.* 2008. Distinct modes of regulation of the Uch37 deubiquitinating enzyme in the proteasome and in the Ino80 chromatin-remodeling complex. *Molecular cell* **31**, 909-917.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147**, 1498-1510.

Yu, Z., Genest, P.A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J.M., Hausinger, R.P., *et al.* 2007. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic acids research* **35**, 2107-2115.

Zhan, Q. 2005. Gadd45a, a p53- and BRCA1-regulated stress protein, in cellular response to DNA damage. *Mutation research* **569**, 133-143.

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. 2010a. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell research* **20**, 1390-1393.

Zhang, J., Gao, Q., Li, P., Liu, X., Jia, Y., Wu, W., Li, J., Dong, S., Koseki, H., and Wong, J. 2011. S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell research* **21**, 1723-1739.

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H., and He, C. 2012. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature chemical biology* **8**, 328-330.

Zhang, P., Andrianakos, R., Yang, Y., Liu, C., and Lu, W. 2010b. Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *The Journal of biological chemistry* **285**, 9180-9189.

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 beta-N-acetylglucosamine transferase (OGT). *The Journal of biological chemistry* **289**, 5986-5996.

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

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

Zhang, Y., Rohde, C., Tierling, S., Jurkowski, T.P., Bock, C., Santacruz, D., Ragozin, S., Reinhardt, R., Groth, M., Walter, J.*, et al.* 2009. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS genetics* **5**, e1000438.

Zhou, J., Liu, M., Fleming, A.M., Burrows, C.J., and Wallace, S.S. 2013. Neil3 and NEIL1 DNA glycosylases remove oxidative damages from quadruplex DNA and exhibit preferences for lesions in the telomeric sequence context. *The Journal of biological chemistry* **288**, 27263-27272.

Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., and Xu, R.M. 2014. Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Molecular cell* **54**, 879-886.

Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., Thiry, S., Siegmann, M., and Jost, J.P. 2000a. 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic acids research* **28**, 4157-4165.

Zhu, B., Zheng, Y., Hess, D., Angliker, H., Schwarz, S., Siegmann, M., Thiry, S., and Jost, J.P. 2000b. 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5135-5139.

Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., and Muegge, K. 2006. Lsh is involved in de novo methylation of DNA. *The EMBO journal* **25**, 335-345.

Zhu, J., He, F., Hu, S., and Yu, J. 2008. On the nature of human housekeeping genes. *Trends in genetics : TIG* **24**, 481-484.

Zhu, J.K. 2009. Active DNA demethylation mediated by DNA glycosylases. *Annual review of genetics* **43**, 143-166.

4.2 Abbreviations

2-HG	2-hydroxyglutarate
2-OG	2-oxoglutarate
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymethyluracil
5fU	5-formyluracil
8oxoG	8-oxoguanine
AID	Activation induced cytidine deaminase
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide
BER	Base excision repair
CGI	CpG island
CpG	Cytosine-phosphatidyl-guanine
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
ESC	Embryonic stem cell
F3H	Fluorescent-three hybrid
Fe	Iron
GADD45	Growth arrest and DNA-damage-inducible protein 45
HDAC	Histone deacetylase
HP1	Heterochromatin binding protein 1
IAP	Intracisternal A particle
IDH	Isocitrate dehydrogenase
IDAX	Inhibition of the Dvl and axin complex protein
JBP	J-binding protein
JGT	J-associated glycosyltransferase
KLF	Krüppel-like factor
LIG3	DNA Ligase 3
MBD	Methyl-CpG binding domain
MeCP2	Methyl-CpG binding protein 2
MPG	N-methylpuringe glycosylase
MLL1	Mixed-lineage leukemia protein 1
MS	Mass spectrometry

NEIL	Nei endonuclease 8-like
NER	Nucleotide excision repair
NPC	Neuronal progenitor cell
OCT4	Octamer binding transcription factor 4
OGG1	8-oxoguanine glycosylase 1
O-GlcNAc	O-linked N-acetylglucosamine
OGT	O-linked N-acetylglucosamine transferase
PARP1	Poly [ADP-ribose] polymerase 1
PCNA	Proliferating cell nuclear antigen
PGC	Primordial germ cells
PRC	Polycomb repressive complex
RNA	Ribonucleic acid
SMUG1	Single-strand selective monofunctional uracil DNA glycosylase
TDG	Thymidine DNA glycosylase
TET	Ten-eleven translocation
UHRF	Ubiquitin-like containing PHD and Ring finger domain proteins
UNG	Uracil-DNA glycosylase
wt	wildtype
ZnF	Zinc Finger
XRCC1	X-ray repair cross-complementing protein 1