

The role of DNA modifications in development and disease

Sebastian Bultmann

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

vorgelegt von
Sebastian Bultmann
aus München
München, den 21. Mai 2012

Erstgutachter: Prof. Dr. Heinrich Leonhardt
Zweitgutachter: Prof. Dr. Thomas Lahaye

Tag der mündlichen Prüfung: 25.6.2012

Contents

Summary	iii
1 Introduction	2
1.1 Epigenetic information	2
1.2 Histone modifications and Histone variants	2
1.3 DNA methylation	4
1.3.1 DNA methylation in development and disease	5
1.3.2 Mammalian DNA methyltransferases	6
1.3.3 Factors that bind methylated CpGs	12
1.4 Mouse embryonic stem cells as a model of early mammalian development	16
1.4.1 Self-renewal and pluripotency of ESCs	17
1.4.2 Role of Oct4, Sox2, and Nanog in pluripotency	17
1.4.3 Embryonic stem cell differentiation	19
1.5 DNA hydroxymethylation	21
1.5.1 Ten-eleven-translocation protein family	21
1.5.2 5-hmC and DNA demethylation	22
1.5.3 Biological function of 5-hmC	28
1.6 Transcription activator-like effectors	29
1.6.1 Biology of TAL effectors	29
1.6.2 Designer TAL effectors as a tool for genome editing	31
1.7 Aims of this work	33
2 Results	35
2.1 Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA	35
2.2 Characterization of PvuRtsII endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine	42

2.3	Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1	64
2.4	Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia	87
2.5	Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain . . .	98
2.6	Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells	117
2.7	Targeted transcriptional activation of silent <i>oct4</i> pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers	130
3	Discussion	148
3.1	The role and function of 5-hmC and Tet enzymes in development and disease	148
3.1.1	Methods for quantification and mapping of 5-hmC	148
3.1.2	Recognition of 5-hmC by the epigenetic machinery	150
3.1.3	The role of the CXXC domain for Tet1 function	152
3.1.4	Tet2 and 5-hmC in myeloid leukemia	153
3.1.5	5-hmC a stable epigenetic mark or a demethylation intermediate?	155
3.2	The role of DNA methylation in transgene silencing and its mechanism . .	156
3.3	Designer TALEs as a tool for targeted transcriptional activation	158
3.3.1	FairyTALE a simple web tool for dTALE target optimisation	161
4	Annex	163
4.1	References	163
4.2	Contributions	191
4.3	Declaration	192
4.4	Acknowledgement	193
	Curriculum vitae	196

Summary

DNA methylation plays a central role in epigenetic regulation of mammalian gene expression. The overall pattern of DNA methylation is established during early development and is essential for lineage specification and maintenance of differentiated states. These observations led to the notion that DNA methylation represents a stable epigenetic mark which is subjected to little changes after the initial setting of cell type specific patterns. In the last decade this idea evolved into a more dynamic view where DNA methylation levels can rapidly change in response to internal and external signalling. However, it remained highly debated what mechanisms are used to remove DNA methylation marks, a process that is required in order to explain the observable DNA methylation dynamics. Recently, the Tet family of enzymes were described to catalyse the oxidation of 5-methylcytosine (5-mC) to 5-hydroxymethyl-cytosine (5-hmC), a potential intermediate in DNA demethylation.

To gain first insights into the function of 5-hmC in epigenetic gene regulation we developed a sensitive enzymatic assay for quantification of 5-hmC content in genomic DNA. Using this assay we discovered that 5-hmC is most abundant in the central nervous system and changes dramatically during development. In addition, we showed that Tet3 is the major Tet protein in adult tissues and its expression correlates well with 5-hmC content during early development. Furthermore, we used the assay to quantify genomic 5-hmC levels in patients with myeloid leukemia uncovering a correlation between global gene expression profile, Tet2 mutational status and 5-hmC content. Additionally, we characterised a novel restriction enzyme PvuRts1I that specifically cleaves 5-hmC containing DNA and can be used to analyse localisation of 5-hmC in genomic DNA. Next, we addressed the question how 5-hmC is integrated in the epigenetic network. Interestingly, we found that 5-hmC is not recognized by the 5-mC binding protein MeCP2 representing a mechanism by which oxidation of 5-mC could lead to gene activation. Moreover, we discovered that the CXXC domain of Tet1 represents a specific subtype of CXXC zinc-finger domains found in a small group of proteins involved in the wnt signalling pathway.

Using a fluorescence microscopy based high-throughput assay we systematically analysed the involvement of epigenetic factors in transgene silencing and discovered a so far unknown pathway by which silencing is achieved. At last, we studied the function of designer transcription activator-like effectors (dTALEs) in activating the *oct4* promoter and found that in combination with low dosages of epigenetic inhibitors *oct4* transcription could efficiently be activated.

In conclusion, we developed a new set of methods for quantification and mapping of 5-hmC and used these to gain first insights into the function of 5-hmC and Tet enzymes in development and disease. Moreover, we further elucidated the function of epigenetic factors in transgene silencing and developed a technique for targeted transcriptional activation using dTALEs in combination with epigenetic inhibitors.

1 Introduction

1.1 Epigenetic information

“One can say. . . that the elucidation of the genetic code is indeed a great achievement. It is, in a sense, the key to molecular biology because it shows how the great polymer languages, the nucleic acid language and the protein language, are linked together.”[Crick, 1958].

Francis Crick was right to predict a breakthrough in molecular biology by the discovery of the genetic code. However, the “key to molecular biology” has proven to open a door into a room (or rather a hall) filled with many answers but even more questions. The direct connection between “nucleic acid language” and “protein language” which seemed so obvious and straight forward in the late 1950’s turned out to be an extremely complex, inter-dependent relationship.

Considering the enormous variety of cell types, their different functions and morphologies it became clear that knowing the nucleotide sequence alone is only a small part of the puzzle. Although all cells of a given multicellular organism contain the same genetic information, they differ in their function and gene expression profiles; hence the distinct properties of the cells do not reside in the nucleotide sequence but in how the cells make use of their common genomic background. This level of information was termed epigenetic (epi (Greek): over, above). Epigenetic mechanisms control cell-, tissue-, and development- specific gene expression and are therefore responsible for the identity of different cell types. Moreover, epigenetic information is heritable and thus can be passed on from one cell to its progeny.

1.2 Histone modifications and Histone variants

Eukaryotic DNA is organized into a higher order structure called chromatin. The basic unit of chromatin is the nucleosome, which consists of 147 base pairs (bp) of DNA wrapped around an octamer of core histones. This histone octamer is composed of two

heterodimers of histone H3 and H4 associated with two heterodimers of histones H2A and H2B [Finch et al., 1977; Dubochet and Noll, 1978]. These proteins share related globular domains that mediate histone-histone interaction and DNA binding. Besides, each histone also harbors a 20-35 amino acid long N-terminal peptide that extends from the surface of the nucleosome. The histone “tails” and to a lower extent the core of the protein are subject to a large number and variety of posttranslational modifications, including methylation and acetylation of lysines and arginines, phosphorylation of serines and threonines, ubiquitinylation and sumoylation of lysines, as well as ribosylation. It is believed that many of these modifications play important roles in the regulation of transcription. In principle, this can be achieved in two ways. Some modifications may lead to alterations in structure and charge of the nucleosome which cause changes in DNA binding and nucleosome packaging. For example acetylation of lysine residues neutralizes their positive charge which may weaken the interaction with the negatively charged DNA backbone and thus lead to an open chromatin state where transcription factors can access DNA more efficiently. Although this might be a possibility by which histone modifications can regulate transcriptional activity, it is likely that most act by controlling the recruitment of regulatory factors. For example, the chromodomain of heterochromatin protein 1 (HP1) binds to histone H3 when lysine (K) 9 is methylated [Lachner et al., 2001] and this can lead to repression of transcription [Danzer and Wallrath, 2004]. Furthermore, the bromodomains of several proteins involved in transcriptional activation bind to acetylated lysines of histone H3 and H4 [Jacobson et al., 2000]. Histone modifications can also lead to the recruitment of DNA methyltransferases and thereby to DNA methylation and transcriptional repression [Tachibana et al., 2008].

Another mechanism by which histones modulate chromatin is via histone variants. While the major histone proteins are encoded by multiple copies of histone genes, histone variants are usually present as single-copy genes. Furthermore, histone variants exhibit significant differences in the primary sequence compared to the major histones. Some variants have distinct biophysical characteristics that are thought to alter the properties of nucleosomes, while others localize to specific regions of the genome. Some histone variants are exchanged with the pre-existing histones during development and differentiation leading to tissue-specific expression patterns. These observations have led to the suggestion that the histone variants have specialized functions in regulation of chromatin dynamics. Several histone variants have been shown to function in transcription, particularly in repression. One example is the H2A variant MacroH2A which localizes to the inactive X-chromosome and some models suggest that the C-terminal tail of

MacroH2A can repress transcription enzymatically. Other variants have been shown to aid in transcriptional activation like H2A-Bbd which facilitates nucleosome displacement by destabilizing the nucleosome [Kamakaka and Biggins, 2005].

While nucleosomes have long been viewed as stable complexes, there is strong evidence that they are highly dynamic, being constantly altered in their composition, structure, and location along the DNA. Chromatin-remodeling complexes contain ATPase subunits and are known to slide nucleosomes, replace histones, or alter the histone-DNA interactions [Kamakaka and Biggins, 2005; Längst and Becker, 2004].

1.3 DNA methylation

DNA methylation in mammals refers to the addition of a methyl-group to the 5' carbon atom of cytosine which leads to the formation of 5-methyl cytosine (5-mC). It occurs predominantly at CpG dinucleotides but is also found at non-CpG sites albeit to a lesser degree [Lister et al., 2009; Laurent et al., 2010]. In mammalian somatic cells 4% of cytosines are methylated, which accounts for 70%-80% of all CpG dinucleotides in the genome [Ehrlich et al., 1982]. The remaining 20%-30% mainly comprise CpG islands, regions with a high CpG density which are associated with most promoters of constitutively expressed genes and 40% of genes that display a tissue-specific expression profile [Larsen et al., 1992]. While methylation in promoter regions is thought to be associated with gene silencing [Colot and Rossignol, 1999], there is emerging evidence that highly transcribed genes carry methylation marks in the gene body, but the functional consequences of this are unknown so far [Ball et al., 2009; Laurent et al., 2010].

Several mechanisms have been proposed on how global DNA methylation patterns are established. While there is evidence that DNA-binding factors are involved in creating and keeping regions from being methylated especially in the context of CpG islands [Brandeis et al., 1994; Macleod et al., 1994; Dickson et al., 2010], other studies could show that certain chromatin marks and DNA methylation occur in relation to each other [Weber et al., 2007; Meissner et al., 2008; Hawkins et al., 2010]. A number of factors have been identified that could mediate the functional interplay between DNA methylation and chromatin modifications. These factors have been shown to bind to histone modifications and CpG sites thereby connecting both epigenetic mechanisms [Zhao et al., 2009; Hashimoto et al., 2010; Rottach et al., 2010; Pichler et al., 2011]. A recent study revealed that in addition to trans acting mechanisms mentioned above, promoter sequences contain methylation-determining regions (MDRs) that are sufficient

to mediate both hypomethylation and *de novo* methylation in cis [Lienert et al., 2011].

1.3.1 DNA methylation in development and disease

DNA methylation has several important biological functions. During embryonic development the genome experiences large changes in methylation levels. While the genomes of egg and sperm cells are highly methylated [Sanford et al., 1987], the methylation is rapidly lost after fertilization by passive [Rougier et al., 1998] and active [Mayer et al., 2000; Gu et al., 2011; Wossidlo et al., 2011] mechanisms. After implantation embryonic DNA methylation patterns are re-established through lineage-specific *de novo* methylation [Kafri et al., 1992; Santos et al., 2002]. The importance of DNA methylation during embryonic development is supported by the discovery that embryos, which have defects in DNA methylation show severe developmental deficiencies and die before birth [Okano et al., 1999]. Another important function of CpG methylation is the maintenance of mono-allelic expression of imprinted genes [Li et al., 1993]. In Embryos lacking the maintenance DNA methyltransferase Dnmt1, alleles of both *Igf2* and *Igf2r*, which are normally paternally and maternally expressed, respectively, are silenced. Furthermore, the *H19* gene, which is normally maternally transcribed, is bi-allelically expressed. In addition to its crucial role in imprinting, DNA methylation is also important for the X-inactivation in female mammals as the expression of *Xist* is controlled by methylation [Norris et al., 1994]. Moreover, DNA methylation is crucial for chromosomal stability. Patients with the ICF syndrome (Immunodeficiency, Centromere Instability, Facial Anomalies syndrome) carry a hypomorphic germline mutation in the gene coding for the *de novo* methyltransferase DNMT3B and exhibit, besides other defects, a loss of DNA methylation in centromeric and pericentromeric repeat regions [Miniou et al., 1997]. This leads to pericentromeric decondensation and chromosomal instability. Consistent with these findings, mouse embryonic stem cells (ESCs) lacking the two *de novo* methyltransferases *dnmt3a/dnmt3b*, exhibit elevated rates of centromeric sister chromatid exchange [Jaco et al., 2008]. In addition, CpG methylation is crucial for the silencing of retroviruses and transposon inactivation [Cherry et al., 2000].

Aberrant changes in global DNA methylation patterns are characteristic for many cancer types. In many cases a combination of global hypomethylation and promoter-localized hypermethylation is observed. However, the hypermethylation is not always confined to promoter regions but can be spread over large gene “neighborhoods” up to whole chromosome bands resulting in severe changes of gene expression patterns [Miremedi et al., 2007].

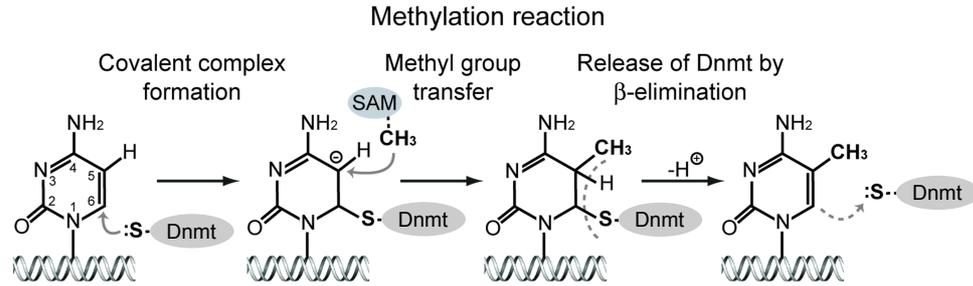


Figure 1: Schematic representation of the catalytic mechanism of Dnmts adapted from [Schermelleh et al., 2005]. The low reactivity of the C5 is overcome by a covalent complex formation at the C6 by a conserved prolylcysteine dipeptide. After methyl-group transfer from S-adenosyl-L-methionine (SAM) the covalent bond is resolved by β -elimination.

1.3.2 Mammalian DNA methyltransferases

There are two types of DNA methyltransferase (Dnmt) activities in mammals, *de novo* methylation and maintenance methylation. *de novo* methylation is by definition the creation of new methylation marks at previously unmethylated CpG sites. It is important during differentiation and embryonic development [Okano et al., 1999] as well as for silencing of retroviral expression. The methylation pattern established by *de novo* methylation has to be passed on from one cell to its progeny. This is achieved by maintenance DNA methylation. The difference to *de novo* methylation is that methylation marks are not set on unmethylated but on hemi-methylated DNA which is produced during DNA replication. The catalytic mechanism by which cytosine-C5 DNA methyltransferases catalyze the generation of 5-mC is conserved from prokaryotes to mammals. These enzymes overcome the low reactivity of the cytosine C5 atom by covalent complex formation at the C6 position. After flipping the target base out of the DNA helix, a process called “base flipping”, the thiolate of a conserved prolylcysteine (PC) dipeptide forms a covalent bond with the C6 of the cytosine. This leads to the activation of the C5 atom for the methyl-group transfer from S-adenosyl-L-methionine (SAM), the common methyl group donor for all methyltransferases (Figure 1). Subsequently, the covalent bond is resolved by β -elimination. In mammals there are three families of DNA methyltransferases which will be described in more detail in the next paragraphs.

In mammals there are three types of DNA methyltransferases (Dnmts). Dnmt1, Dnmt2 and Dnmt3 of which the first and the latter are known to catalyze the formation of 5-mC. In their catalytic region they display a high homology to bacterial DNA (cytosine-5) methyltransferases (Figure 2). Moreover, the basic enzymatic mechanism is conserved

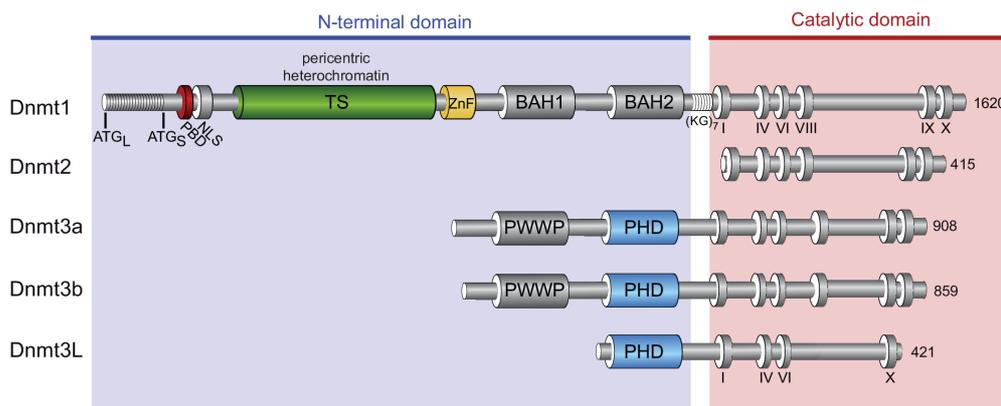


Figure 2: Schematic representation of the mammalian DNA methyltransferase family adapted from [Rottach et al., 2009]. All Dnmts have a similar C-terminal catalytic domain characterized by the highly conserved motifs (I-X) also found in prokaryotic DNA (cytosine-5) methyltransferases. The Dnmts differ, however, in their regulatory region. Dnmt1 contains the PCNA binding domain (PBD), the pericentric heterochromatin targeting sequence (TS), a CXXC-type zinc finger motif (ZnF), and two bromo adjacent homology domains (BAH). The start codon of the long (ATGL) and short (ATGS) isoforms, as well as the seven lysine-glycine repeat linker (KG7) are indicated. The regulatory domains of Dnmt3a and 3b comprise a PWWP domain named after a conserved Pro-Trp-Trp-Pro motif of the plant homeodomain (PHD)

between mammals and bacteria. In contrast to their prokaryotic ancestors, mammalian Dnmts do not exhibit strong sequence specificity. However, the spatial distribution of 5-mC in mammalian genomes is vital. The necessary specificity is mediated by a large N-terminal regulatory region that contains several different domains which mediate the interaction with proteins that regulate the localization and activity of the methyltransferases. (Figure 2).

In this context it is important to take into account that DNA methylation is part of a complex epigenetic network. DNA methylation, chromatin modifications and chromatin remodeling pathways mutually affect each other in multiple ways. The N-terminal region of the Dnmts is responsible for the correct integration of DNA methylation in this network. Through its direct cross-talk with chromatin or indirectly through interaction with other proteins the N-terminal region regulates the localization and activity of the Dnmts. For example, Dnmt3L binds directly to unmethylated H3K4 via its plant homeodomain (PHD domain) [Jia et al., 2007]. However, in most cases the interpretation of chromatin signals is achieved indirectly and Dnmts have been shown to interact with a variety of proteins that are involved in epigenetic processes. Interactions have been described with histone 3 lysine 9 (H3K9) methyltransferases (HMTs), Suv39h1, SetDB1 and G9a,

components of the Polycomb repressive complex 2, histone deacetylases (HDACs) and the heterochromatin protein 1 (HP1) [Cedar and Bergman, 2009].

Dnmt1

Dnmt1 was the first eukaryotic DNA methyltransferase to be discovered [Bestor, 1988]. It has been shown that it methylates hemimethylated DNA much more efficiently than unmethylated substrates [Bestor and Ingram, 1983] which led to the assignment of Dnmt1 as a maintenance DNA methyltransferase. Although there is evidence that Dnmt1 has *de novo* methylation activity *in vitro* [Pradhan et al., 1997], it is most likely that its main biological role *in vivo* is maintaining genomic methylation patterns. This is supported by the finding that Dnmt1 colocalizes with the replication machinery [Leonhardt et al., 1992]. At replication sites hemimethylated DNA is formed when the newly synthesized unmethylated strand pairs with the methylated template strand. Although the (transient) association with the replication machinery makes it possible that Dnmt1 could directly methylate newly forming hemimethylated CpG sites, it seems not to be essential for maintaining postreplicative methylation levels [Schermelleh et al., 2007; Spada et al., 2007]. Intriguingly, Dnmt1 alone is not sufficient to stably maintain DNA methylation as in ESCs lacking both *de novo* methyltransferases, Dnmt3a and Dnmt3b, global methylation levels slowly decrease during long term culture, although they still express Dnmt1 [Chen et al., 2003a]. Furthermore, Dnmt1 is ubiquitously expressed and its presence is essential for the survival of somatic cells where apoptosis is induced via a p53 mediated pathway, when Dnmt1 is depleted [Jackson-Grusby et al., 2001]. The importance of Dnmt1 during development is shown by the fact that mice lacking Dnmt1 do not develop correctly and exhibit a growth arrest prior the 8-somite stage [Li et al., 1992; Lei et al., 1996]. Moreover, Dnmt1 plays a crucial role in the maintenance of chromosomal stability as mice expressing Dnmt1 at strongly reduced levels are viable at birth but soon develop aggressive T cell lymphomas with a high frequency of chromosome 15 trisomy [Gaudet et al., 2003]. The murine somatic form of Dnmt1 consist of an 1100 amino acid long N-terminal regulatory region and a 500 amino acid long C-terminal catalytic domain (Figure 2). The latter is common to all eukaryotic Dnmts and consists basically of ten conserved motifs which are crucial for the catalytic activity. The N-terminal region is build up by a number of functional domains that have regulatory functions. The first 125 amino acids mediate the interaction with the DMAP1 transcriptional repressor [Rountree et al., 2000]. Dnmt1, except in early development, exhibits a nuclear localization and has several nuclear localization signals (NLof

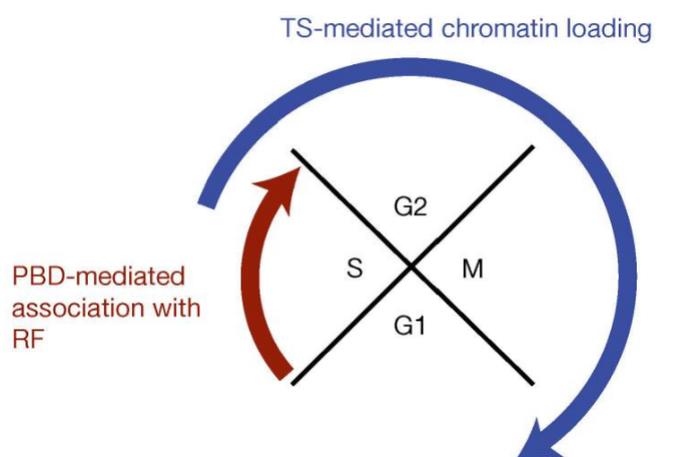


Figure 3: Scheme of the cell cycle dependent localization of Dnmt1 and its involved domains [Easwaran et al., 2004]. Dnmt1 associates with replication foci (RF) during early to mid S-phase via its PCNA binding domain (PBD). From late S-phase to M-phase targeting sequence (TS) mediated heterochromatin loading is observable.

which the major one is formed by amino acids 178-202 [Cardoso and Leonhardt, 1999]. Furthermore, the N-terminal region contains the PCNA binding domain (PBD) which mediates the interaction with the Proliferating Cell Nuclear Antigen (PCNA) [Chuang et al., 1997]. PCNA is a trimeric protein that serves as a loading platform for many proteins involved in DNA replication and repair. Together with the targeting sequence (TS) domain, located at amino acids 310-629, it is responsible for the cell cycle dependent localization pattern of Dnmt1. While PCNA mediates the accumulation of Dnmt1 at replication foci throughout S phase, the TS domain mediates the accumulation at pericentric heterochromatin from late S-phase till early G1 (Figure 3) [Easwaran et al., 2004].

In addition to these two domains, Dnmt1 also contains a zinc (Zn) binding domain which is located between amino acids 649-696 and mediates DNA binding [Fatemi et al., 2001]. Another domain of the N-terminal region is the Polybromo- homology domain (PBHD) which resides between amino acid 762-964 and contains two BAH (bromo adjacent homology) domains that are likely involved in protein-protein interactions [Nicolas and Goodwin, 1996]. There are several isoforms of Dnmt1 of which two are expressed in somatic cells, namely Dnmt1s and Dnmt1b. Dnmt1s is the predominant form and is expressed in all dividing cells. Dnmt1b which contains additional 48 nucleotides between exon 4 and 5 [Hsu et al., 1999] is expressed at much lower levels than Dnmt1s and has been described only in human. Both have comparable enzymatic properties but

the function of the Dnmt1b isoform is so far not clear. In oocytes and preimplantation embryos another isoform, Dnmt1o, is expressed. It is transcribed from an oocyte specific promoter and its first exon differs from the one of Dnmt1s and Dnmt1b [Gaudet et al., 1998].

Dnmt2

Dnmt2 shows a high homology to other DNA methyltransferases. The inferred protein sequence contains all 10 catalytic motifs in the canonical order (Figure 2). In contrast to other eukaryotic DNA methyltransferases, Dnmt2 homologues do not possess a regulatory N-terminal region and in this respect resemble more closely bacterial cytosine methyltransferases. In addition, the crystal structure of human DNMT2 revealed that the structures of DNMT2 and the bacterial restriction methyltransferase M.HhaI are essentially superimposable [Dong et al., 2001]. In fact, the Dnmt2 family is the most strongly conserved and most widely distributed family of eukaryotic cytosine methyltransferase homologues [Goll and Bestor, 2005]. Despite all similarities to other cytosine methyltransferases, the functional role of the Dnmt2 family remained enigmatic as Dnmt2 homologues could not be shown to possess considerable DNA methyltransferase activity and mice lacking Dnmt2 do not exhibit DNA methylation abnormalities [Okano et al., 1998b]. In 2006 it was shown that DNMT2 methylates the aspartic acid transfer RNA (tRNA^{Asp}) at cytosine 38 in the anticodon loop [Goll et al., 2006].

Interestingly, analysis of tRNA^{Asp} sequences showed complete conservation of the anticodon loop in species whose genomes encode Dnmt2 homologues, whereas in *C. elegans* and *S. cerevisiae*, which lack a Dnmt2 homologue, tRNA^{Asp} anticodon loops have diverged. These findings indicate coevolution of Dnmt2 and the anticodon loop of tRNA^{Asp}. However, the functional consequence of the tRNA^{Asp} methylation remains unclear. The only phenotypic effect of Dnmt2 depletion reported so far was found in zebrafish, where a knockdown in embryos results in differentiation defects in particular organs, including retina, liver, and brain. In agreement with its role in tRNA^{Asp} methylation, cytoplasmatically located Dnmt2 could rescue this phenotype [Rai et al., 2007]. Dnmt2 seems to have an additional role in some organisms. In *Drosophila* very low levels of cytosine methylation are present [Gowher et al., 2000]. Its genome encodes only a Dnmt2 methyltransferase and lacks any of the canonical *de novo* or maintenance Dnmts. In this organism, as well as in *Dictyostelium* [Kuhlmann et al., 2005], Dnmt2-dependent DNA methylation was shown to be necessary for retrotransposon silencing and telomere integrity [Phalke et al., 2009].

Dnmt3

The mammalian genome encodes two functional Dnmt3 methyltransferases, namely Dnmt3a and Dnmt3b, and a third homologue, Dnmt3L which lacks cytosine methyltransferase activity. Dnmt3a and Dnmt3b are closely related proteins that, similar to Dnmt1, possess an N-terminal regulatory region and a C-terminal catalytic domain (Figure 2). Both were found to methylate CpG dinucleotides *in vitro* without preference for hemimethylated DNA and thereby assigning their possible role as *de novo* Dnmts [Okano et al., 1998a]. This was confirmed *in vivo*, by using a stable episomal system that employs plasmids as targets for *de novo* DNA methylation [Hsieh, 1999] and by the finding that *dnmt3a/dnmt3b* double knockout ESCs exhibit an inability to *de novo* methylate newly introduced retroviral elements while the maintenance of imprinted methylation pattern is not affected [Okano et al., 1999].

The N-terminal regions of Dnmt3a and Dnmt3b harbor a PWWP domain which is found in many chromatin-associated proteins. By mutagenesis analysis this domain was shown to be required for pericentric heterochromatin association. Furthermore, disruption of the PWWP domain abolishes the ability of Dnmt3a and Dnmt3b to methylate major satellite repeats at pericentric heterochromatin [Chen et al., 2004]. Both proteins also contain an ATRX-homology domain, a cysteine rich zinc-binding domain mainly found in proteins involved in eukaryotic transcription regulation. It has been shown that the ATRX-homology domain of Dnmt3a is sufficient to repress transcription, independently of the methyltransferase activity, by associating with the histone deacetylase HDAC1 [Fuks et al., 2001]. In addition, this domain has been shown to mediate the binding to symmetrically di-methylated arginine 3 at Histone 4 (H4R3) [Zhao et al., 2009]. Dnmt3a and Dnmt3b have both been shown to be important in mouse embryonic development and differentiation. Both genes are expressed in ESCs and form a complex *in vivo*. Single knockout of either Dnmt3a or Dnmt3b in ESC results in reduction of promoter methylation of the pluripotency markers Oct-4 and Nanog upon differentiation via retinoic acid treatment. Simultaneous knockout of Dnmt3a and 3b completely abolishes *de novo* methylation at these loci [Li et al., 2007]. Besides their similarities and synergistic function, Dnmt3a and Dnmt3b also have some non-overlapping functions which become obvious by the phenotypic differences of Dnmt3a and Dnmt3b single knockout embryos. While the latter die at around E9.5, *dnmt3a*^{-/-} appear normal at birth and die not before 4 weeks of age. Global methylation patterns seem to be normal in *dnmt3a* deficient mice [Okano et al., 1999]. Deletion of Dnmt3a in the female germ line leads to hypomethylation at differentially methylated regions (DMR) of all maternally imprinted

genes examined so far. In contrast, *dnmt3b* knockout in germ cells does not result in such a phenotype [Kaneda et al., 2004]. Inactivation of *dnmt3b*, but not *dnmt3a*, in mouse embryonic fibroblasts (MEF), results in partial loss of genome wide DNA methylation. This suggests that, in addition to the major maintenance methyltransferase Dnmt1, Dnmt3b is required for maintaining DNA methylation in somatic cells [Dodge et al., 2005]. In ESCs however, both *de novo* methyltransferases need to be absent in order to achieve a gradual loss of global DNA methylation [Li et al., 2007]. Dnmt3L shows high homology to Dnmt3a and Dnmt3b in its N- and C-terminal domains but lacks the PWWP domain (Figure 2). The catalytic motifs have been subject to nonconservative substitutions and Dnmt3L is not able to catalyze cytosine methylation.

Dnmt3L is mainly expressed in the germ line where it is essential for the establishment of a subset of methylation patterns [Bourc'his et al., 2001]. Interestingly, it seems that Dnmt3L has different functions in male and female germ cells. In male mice targeted disruption of *dnmt3L* causes azoospermia with germ line cells displaying nonhomologous synapsis, asynapsis, and the accumulation of highly abnormal synaptonemal complexes. Abnormal synapsis is likely to be a secondary effect of the observed hypomethylation of transposable elements [Bourc'his and Bestor, 2004]. In contrast, *dnmt3L* deficiency does not interfere with oogenesis and oocytes are methylated normally at transposons. However, female germ cells exhibit a methylation defect in single copy sequences associated with maternal imprinting instead [Bourc'his et al., 2001]. As Dnmt3L is not catalytically active, the methylation defects in Dnmt3L-deficient mice are thought to be caused by the missing activation of Dnmt3a as Dnmt3L stimulates the *de novo* methylation activity of Dnmt3a *in vivo* [Chedin et al., 2002]. In addition, targeted disruption of *dnmt3a* results in phenotypes similar to Dnmt3L knockout [Kaneda et al., 2004].

1.3.3 Factors that bind methylated CpGs

One main function of DNA methylation is transcriptional silencing and there are two models of how this is achieved. The first model suggests that CpG methylation interferes with the binding of transcription factors that require contact with cytosine in the major groove of the double helix [Hark et al., 2000]. While the second model proposes that DNA methylation is translated into a repressive chromatin state. This is mediated by factors that recognize and bind to methylated CpGs which repress transcription indirectly by the recruitment of corepressors. So far there are three protein families known that bind methyl-CpG.

Methyl-CpG-binding-domain (MBD) family proteins

The first methyl-CpG binding protein discovered was MeCP2. It was shown that it exhibits a strong preference for methylated DNA which is mediated by its methyl-CpG-binding domain (MBD). Additionally to the MBD it also contains a transcriptional repression domain (TRD) [Meehan et al., 1989; Nan et al., 1993]. Homology searches identified four additional proteins containing an MBD which were termed MBD1-4 (Figure 4). All MBD proteins, except MBD3, specifically recognize and bind methylated CpGs [Hendrich and Bird, 1998]. Furthermore, MBDs have been shown to be involved in transcriptional repression, except MBD4 which is involved in DNA damage response [Hendrich and Bird, 1998], by cooperating with histone deacetylases and histone methylases [Clouaire and Stancheva, 2008]. For example, MeCP2 binds methylated CpGs and recruits histone deacetylases via its interaction with Sin3A [Nan et al., 1998]. In addition, MeCP2 recruits histone methyltransferases like Suv39 which methylate lysine 9 on histone 3 and thereby creates a binding site for HP1, a major constituent of heterochromatin [Lachner et al., 2001; Fuks et al., 2003]. MeCP2 is not the only connection between DNA methylation and transcriptional repression as MeCP2-deficient cells nevertheless repress methylated constructs [Ng and Bird, 1999]. In these cells alternative pathways involving different MBDs, like MBD3, might be used. MBD3 is a component of the Mi2/NuRD deacetylase complex [Wade et al., 1999] which is the most abundant macromolecular form of histone deacetylase complexes found in mammalian cells, as well as in *Xenopus* eggs and embryos.

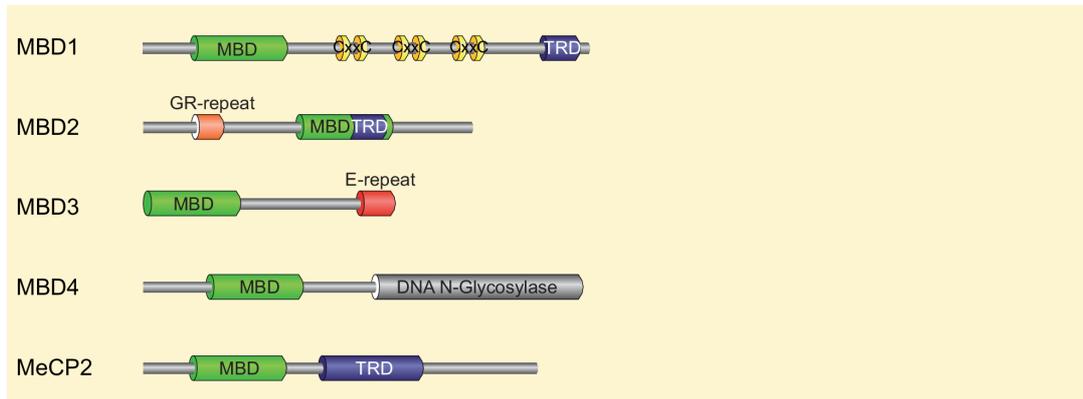


Figure 4: Schematic diagram of the MBD protein family [Rottach et al., 2009]. All family members contain a methyl-CpG-binding domain (MBD). MBD1, MBD2 and MeCP2 additionally harbour a transcriptional repression domain (TRD). MBD4 the only family member with catalytic activity which is mediated by its DNA N-glycosylase domain. In addition to its MBD and TRD domain, MBD1 also contains three CXXC-type zinc-finger domains (CxxC).

SRA domain proteins

It was recently discovered that Dnmt1 needs the presence of another protein, called Np95, to stably maintain genomic methylation levels. In ESCs lacking Np95, also known as Uhrf1, global and local DNA methylation levels are defective and almost identical to that of *dnmt1* knockout cells [Sharif et al., 2007]. Furthermore, Np95 colocalizes with Dnmt1 *in vivo* at replication forks during mid-to-late-S-phase when pericentromeric heterochromatin is replicated [Papait et al., 2007]. Co-immunoprecipitation experiments showed a direct interaction of Np95 and Dnmt1 [Sharif et al., 2007]. In addition, it has been shown that Np95 can bind directly to methyl-CpG via its SRA (SET and RING associated) domain [Unoki et al., 2004]. All these data suggest that Np95 recruits Dnmt1 to hemimethylated CpG sites at replication forks so that Dnmt1 can copy the methylation mark onto the newly synthesized strand. Cococrystallization of the SRA domain with a hemimethylated DNA substrate showed that upon binding, the 5-mC is flipped out of the DNA helix and positioned in a binding pocket [Hashimoto et al., 2008]. Interestingly, a similar mechanism for DNA binding has been described for DNA methyltransferases [Klimasauskas et al., 1994] and this base flipping is thought to be involved in the coordinated transfer of the hemi-methylated CpG site from Np95 to Dnmt1.

In general SRA domain proteins fall in two distinct families. The first is characterized by the association of the SRA domain with PHD and RING domains. The only known mammalian homologues discovered so far are Np95 and the closely related Np97, also



Figure 5: Schematic diagram of the SRA domain proteins Np95 and Np97 [Rottach et al., 2009]. Np95 and Np97 consist of an N-terminal ubiquitin-like domain (Ubl) followed by a Tudor, a plant homeodomain (PHD) and a SET and RING associated domain (SRA). At the C-terminus both proteins harbour a really interesting new gene domain (RING).

known as Uhrf2. However in *Arabidopsis thaliana* at least five members have been identified. The second family of SRA domain proteins is thought to be plant-specific and includes members of the SUVH of SET domain histone methyltransferases.

Besides the SRA domain, Np95 harbors at least four other functional domains, an ubiquitin-like domain (UBL), followed by a tandem tudor domain, a plant homeodomain (PHD), and a really interesting new gene (RING) domain (Figure 5). All of the domains are somehow connected to chromatin formation which makes Np95 a possible mediator between DNA methylation and chromatin modulation. For example, the RING domain is thought to confer E3 ubiquitin ligase activity on Np95 and shows specific activity for histone H3 [Citterio et al., 2004] and histone tail ubiquitylation is an important determinant in the regulation of chromatin structure and gene transcription [Jason et al., 2002]. Furthermore, the PHD domain has been shown to bind to methylated histone 3 lysine 9 (H3K9), a repressive chromatin marker [Karagianni et al., 2008]. In addition, the PHD domain of Np95 has been proposed to be involved in large-scale reorganization of pericentromeric heterochromatin by recruiting chromatin modifying enzymes and Dnmts [Papait et al., 2007]. Another domain of Np95 that binds to chromatin is the tandem tudor domain which was crystallized with trimethylated H3K9 bound [Hashimoto et al., 2009]. Moreover, Np95 interacts with the histone methyltransferase HDAC1 and the histone methyltransferase G9a, both of which are involved in heterochromatin formation [Unoki et al., 2004; Kim et al., 2009]. In summary, the fact that Np95 contains domains that recognize DNA methylation as well as repressive histone marks and the ability to recruit histone modifying enzymes, suggests that it may functionally link DNA methylation and chromatin modifications.

Kaiso family proteins

The third protein family that is known to bind methylated CpGs is the Kaiso family. These proteins harbor a three-zinc-finger motif that mediates methyl-CpG binding

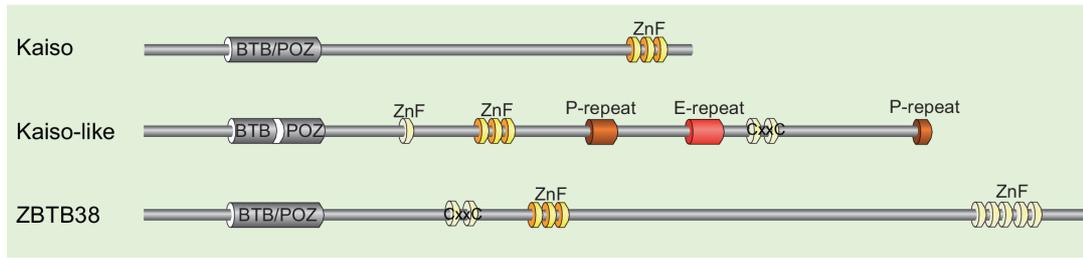


Figure 6: Schematic diagram of the Kaiso family proteins [Rottach et al., 2009]. Kaiso, ZBTB4, and ZBTB38 are characterized by several zinc finger motifs. Binding to methylated DNA is mediated by a C2H2 zinc finger motif (yellow). The broad complex, tramtrack, and bric à brac (BTB/POZ) domain is depicted in gray.

[Prokhortchouk et al., 2001]. The founding member kaiso and the recently identified kaiso-like proteins ZBTB4 and ZBTB38 also contain a poxvirus and zinc finger (POZ) domain that is involved in protein-protein interaction (Figure 6). This domain is thought to be involved in transcriptional repression as kaiso lacking the POZ domain is not capable of silencing methylated reporters. However, transcriptional silencing was not diminished in the presence of the histone deacetylase inhibitor TSA as shown for MBDs [Prokhortchouk et al., 2001]. Inconsistent with these findings, kaiso seems to be part of the repressive N-CoR complex which contains HDAC and histone remodeling activities [Yoon et al., 2003]. Moreover, ZBTB4 and ZBTB38 have also been shown to mediate repression of transcription [Filion et al., 2006]. In summary, Kaiso family proteins might mediate methylation dependent transcriptional repression in a way similar to that of MBD proteins.

1.4 Mouse embryonic stem cells as a model of early mammalian development

ESCs are pluripotent cells derived from the inner cell mass (ICM) of the blastocyst [Evans and Kaufman, 1981]. Murine ESCs retain the full developmental potential of the ICM as they can contribute to all tissues of the embryo and adult *in vivo* after reintroduction into mouse blastocysts [Bradley et al., 1984]. This feature makes it possible to create knockout mice and cell lines which are an important tool in elucidating the function of proteins *in vivo*. Furthermore, deletion of several genes, including Dnmts, are lethal for somatic cells while their ESC counterparts are viable. *In vitro*, ESC can

be differentiated into a broad range of cell type representative for all three germ layers of the mouse embryo [Lake et al., 2000]. Moreover, during *in vitro* differentiation ESC undergo developmental changes and processes similar to that seen in the ICM during early embryonic development. These properties make mouse ESCs an optimal tool for studying early processes in embryonic development.

1.4.1 Self-renewal and pluripotency of ESCs

Self-renewal and pluripotency are key features of ESCs. They can be expanded indefinitely in culture and retain their full developmental potential without exhibiting a bias in the generation of different somatic lineages or germline cells upon reintroduction to the embryo. Although ESCs exhibit a characteristic pattern of epigenetic modifications they seem not to be crucial for self-renewal and pluripotency. This is supported by the finding that ESCs deficient for important epigenetic regulators, such as DNA methyltransferases or histone methyltransferases, are viable without compromising self-renewal or genomic integrity [Tsumura et al., 2006]. However, perturbation of DNA methylation and the chromatin modifying machinery often results in increased cell death during differentiation [Jackson et al., 2004]. Nevertheless, rescue of the epigenetic machinery by reintroduction of the missing components results in full restoration of the developmental potential which suggests a role in successful lineage commitment rather than in retaining pluripotency. In contrast, depletion of ESCs of one of the three transcriptional organizers, Oct4, Sox2, and Nanog results in unscheduled differentiation into trophoblast and hypoblast cells which cannot be rescued by reintroduction of these factors [Nichols et al., 1998; Niwa, 2007]. These fate choices are considered abnormal as they resemble lineages that ICM cells have already passed beyond their segregation points prior to ESC establishment. This indicates that naive pluripotency of ESCs is critically dependent on the action of Oct4, Sox2, and Nanog rather than on the epigenetic machinery.

1.4.2 Role of Oct4, Sox2, and Nanog in pluripotency

ESCs are in a constant struggle between differentiation and self-renewal. Interestingly, in both decisions the three transcriptional organizers Oct4, Sox2, and Nanog seem to be involved. All three are expressed in cells of the ICM and ESCs. They appear to be responsible for the ongoing repression of the expression and activity of lineage specification factors and thereby in retaining pluripotency [Smith, 2005]. However, Oct4 and Sox2 seem also to be key regulators in the extinction of pluripotency by directing the

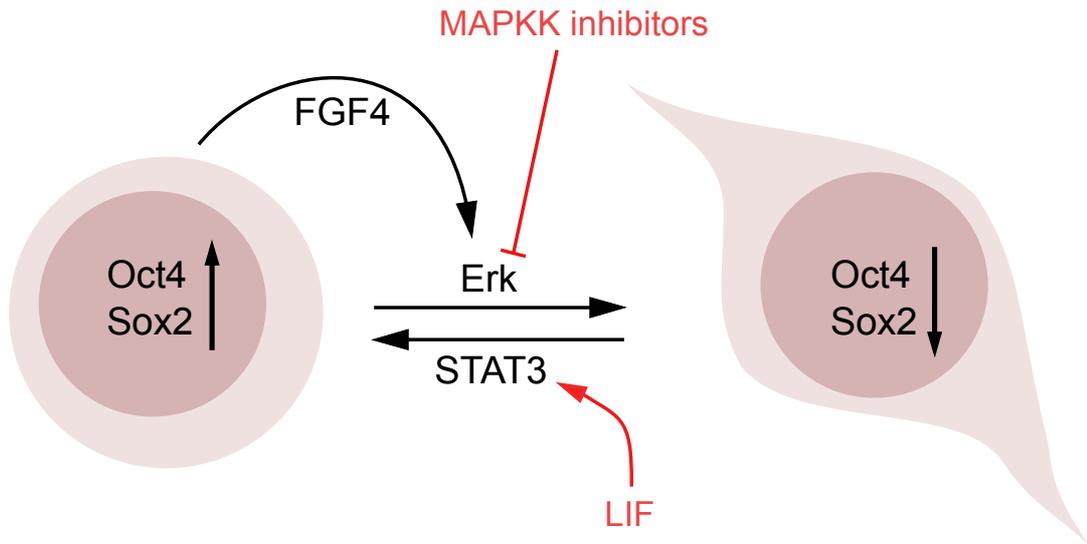


Figure 7: Self-renewal of the pluripotent ESC state requires overcoming the FGF4/Erk signal. Inhibition of FGF4/Erk signalling by small molecules prevents spontaneous differentiation of ESCs in culture. Leukemia inhibiting factor (LIF) stimulates STAT3 signalling which promotes ESC self-renewal and proliferation.

expression of fibroblast growth factor 4 (FGF4). FGF4 propels ESCs towards lineage specification via the mitogen activated protein (MAP) kinase Erk1/2 pathway. Importantly the FGF4/Erk signal does not lead to the differentiation of a certain lineage but results in a general susceptibility for further lineage specific signaling. Consistent with this, blocking of this signaling pathway leads to a general impairment of differentiation [Kunath et al., 2007; Silva and Smith, 2008](Figure 7). In summary, by promoting the expression of FGF4, Oct4 and Sox2 synergistically drive ESCs into differentiation. Consequently, to maintain the naive undifferentiated ESCs in culture the signaling by FGF4 needs to be inhibited. This can be achieved by the addition of the cytokine leukemia inhibiting factor (LIF). LIF stimulates the Stat3 transcription factor signaling which acts downstream of Erk and also promotes ESC growth and viability. In addition small-molecule inhibitors of the MAP kinase kinase (MAPKK or MEK) can be used to inhibit Erk signaling (Figure 7).

In contrast to the homogenous expression Oct4 and Sox2, Nanog expression is subjected to great fluctuations (Figure 8A). Interestingly, constitutive expression of Nanog is sufficient to prevent ESC differentiation even in the presence of active FGF4/Erk signaling. Moreover, Nanog-deficient ESCs can remain undifferentiated and pluripotent in culture, but exhibit a greatly increased tendency to differentiate [Chambers et al.,

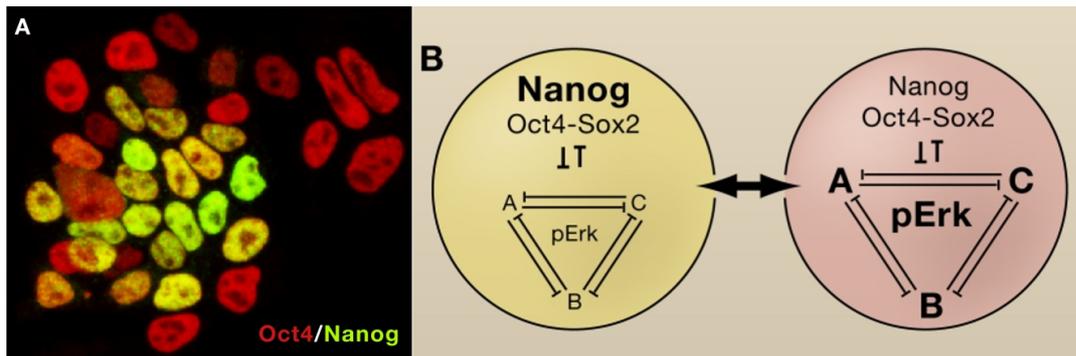


Figure 8: (A) Immunostaining shows highly variable levels of Nanog protein (green) in Oct4 (red) positive undifferentiated ESCs. (B) Nanog prevents differentiation of ESCs. Coincidence of low Nanog expression and elevated FGF4/Erk signaling (pErk) result in susceptibility for further lineage specific signals (symbolized by A,B and C expression circuits) [Silva and Smith, 2008].

2007]. This suggests that Nanog counteracts the FGF4/Erk signaling and thereby retains pluripotency. Moreover, the cell-to-cell heterogeneity of expression levels creates differences in the resistance to differentiation. This means that those ESCs in a culture that express low levels of Nanog are prone to exit self-renewal and will start to differentiate if intrinsic and/or extrinsic lineage specific signals are present above a certain threshold. Although it seems that ESCs can repeatedly switch between high and low levels of Nanog expression, it is clear that many cells that express little Nanog will start to differentiate.

In summary, the three transcriptional organizers Oct4, Sox2, and Nanog together repress lineage-associated transcriptional activity. In addition, Oct4 and Sox2 activate the expression of FGF4 which signals via the MAP kinase Erk1/2 promoting lineage specification of ESCs. This is antagonized by Nanog via an unknown mechanism. As Nanog expression is highly variable among ESCs, coincidence of low Nanog levels with elevated FGF4/Erk signaling leads to activation of intrinsic lineage-specific transcriptional activity resulting in differentiation and loss of self-renewal (Figure 8B).

1.4.3 Embryonic stem cell differentiation

Mouse ESCs are routinely cultivated in the presence of feeder layers to maintain pluripotency and self-renewal. These feeder cells can also be substituted by LIF and with cells grown on gelatinized culture dishes. When feeder cells or LIF is removed ESCs spontaneously differentiate into derivatives of the three embryonic germ layers, mesoderm, endoderm, and ectoderm [Keller, 2005]. In principle, there are three important methods

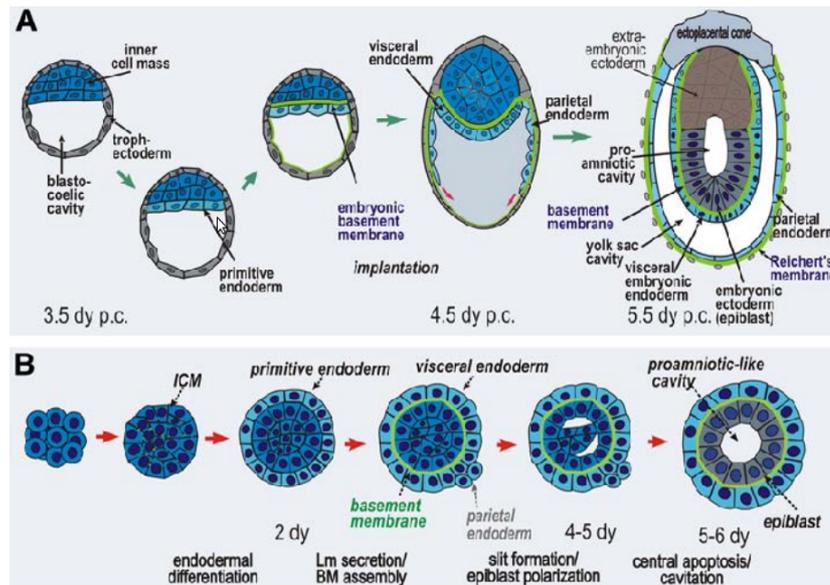


Figure 9: Embryonic (A) and embryoid body (B) development [Li and Yurchenco, 2006]. Embryoid bodies closely resemble the embryonic transition from undifferentiated inner cell mass (ICM) to a two germ-layer structure corresponding to the early egg cylinder stage embryo just before gastrulation.

that have been developed to promote the efficient and reproducible differentiation of ESCs; the culture of ESC as monolayers on extracellular matrix proteins, the culture of ESCs directly on stromal layers, and the formation of three-dimensional aggregates known as embryoid bodies (EBs). This part of the introduction will focus on the differentiation via EBs and its parallels to early embryonic development. As already mentioned, ESCs are derived from the inner cell mass of blastocysts. When differentiated as EBs they closely resemble the embryonic transition from undifferentiated ICM to a two germ-layer structure corresponding to the early egg cylinder stage embryo just before gastrulation (Figure 9). Like the ICM, embryoid bodies will first develop a primitive endoderm which forms on the surface of the EB. This is followed by the assembly of the embryonic basement membrane between the visceral endoderm, which derives from the primitive endoderm, and the ICM-like cells. At this stage also a parietal endoderm is differentiating but due to the lack of trophoectoderm it only forms small peripheral aggregates. Subsequently, the ICM-like cells of EBs produce a central cavitation, mainly by apoptosis, with differentiation of the surviving cells into epiblast cells. Until this point the differentiation of EBs very closely resembles the developmental events of the embryos ICM [Li and Yurchenco, 2006]. Moreover, the epiblast cells of both will differentiate into cells of the different germ layers. However, while in embryonic develop-

ment this follows directed and well coordinated pathways, in EBs epiblast cells exhibit a more random differentiation pattern though there are several methods that make more directed differentiation possible at this stage.

1.5 DNA hydroxymethylation

Until recently the only known covalent epigenetic modification on DNA was methylation at position 5 of cytosine. In 2009, however, it was discovered that 5-mC is further oxidized by the enzyme ten-eleven translocation 1 (TET1) to 5-hydroxymethylcytosine (5-hmC) [Tahiliani et al., 2009] a base that was already detected in mammalian DNA in 1972 but at this time mainly considered as a by-product of oxidative DNA damage [Penn et al., 1972]. In the short time since the discovery of the oxidation from 5-mC to 5-hmC by Tet1 an impressive amount of publications has accumulated (Figure 10). Many possible biological functions for 5-hmC in epigenetic gene regulation have been proposed and it has become increasingly clear that oxidation of 5-mC plays an important role in DNA demethylation.

1.5.1 Ten-eleven-translocation protein family

TET1 was the first protein of the Ten-eleven-translocation protein family to be discovered by genetic analysis of several cases of acute myeloid leukemia (AML) harbouring a $t(10;11)(q22;q23)$. This chromosomal rearrangement was shown to result in an N-terminal fusion of the mixed lineage leukemia (MLL) H3K4 methyltransferase to the C-terminus of TET1. TET1 was found to be a member of a novel and well conserved protein family of, at this time, unknown biological function [Ono et al., 2002; Lorschach et al., 2003]. The function of Tet proteins was discovered by their homology to the DNA modifying enzymes JBP1 and JBP2 in Trypanosomes. JBP1 and JBP2 belong to the 2-oxoglutarate- and Fe(II)-dependent dioxygenase (2OGFeDO) superfamily and catalyse the oxidation of the methyl-group of thymine leading to the formation of 5-hydroxymethyluracil (hmU) the first step in the biosynthesis of base J (β -D-glucosyl-hydroxymethyluracil). Like the JBP proteins, Tet proteins contain a 2OGFeDO domain characterized by a double-stranded β helix (DSBH) fold which in the case of Tet proteins has been shown to catalyse the oxidation of 5-mC to 5-hmC [Tahiliani et al., 2009]. Moreover, it has been discovered that Tet proteins can further oxidise 5-hmC to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) [He et al., 2011; Pfaffeneder

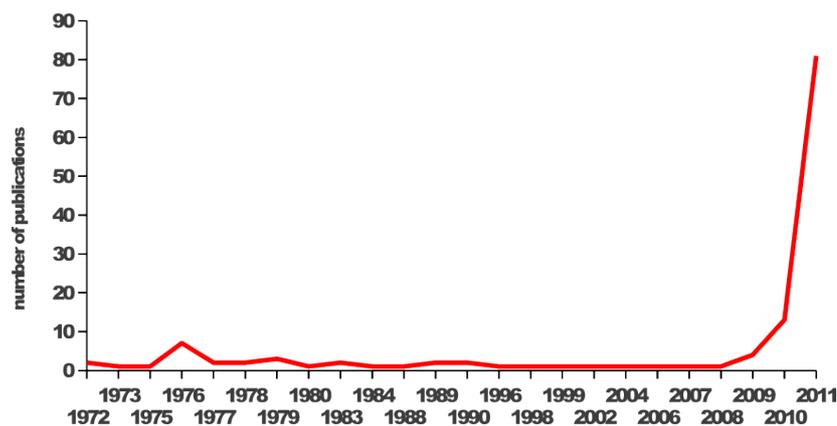


Figure 10: Numbers of publication found by a PubMed search with “5-hydroxymethylcytosine” in title or abstract since 1972

et al., 2011].

In addition to the DSBH domain a directly adjacent cystein-rich region (Cys) was found to be essential for the catalytic activity, at least in the case of Tet1 (Figure 11) [Tahiliani et al., 2009]. Furthermore, Tet1 harbours a functional CXXC type zinc finger domain. The CXXC domain can also be found in several other chromatin proteins and plays a role in DNA binding and possible protein-protein interactions. The CXXC domain seems to be absent in Tet2 and Tet3. However, in the direct chromosomal vicinity to the Tet2 genomic locus the CXXC4 gene can be found which is believed to have been segregated from an ancestral Tet2-CXXC4 fusion gene suggesting a possible functional link of the two proteins [Iyer et al., 2009]. In the case of Tet3 confusing and contradicting reports have been made about whether or not a CXXC is present in the protein. Upstream of the annotated Tet3 locus open reading frame (ORF) has been predicted *in silico* that contains a CXXC domain [Kato and Kato, 2004]. The relatively short distance of this putative ORF to the Tet3 locus makes it possible that a splicing isoform of Tet3 exists which contains this CXXC domain. However, experimental data supporting this hypothesis are still missing.

1.5.2 5-hmC and DNA demethylation

First indications that Tet proteins and 5-hmC may be involved in DNA demethylation came from the observation that overexpression of TET1 in cultured cells leads to a decrease in 5-mC levels [Tahiliani et al., 2009]. In turn, depletion of TET1 in ESCs leads to an increase of DNA methylation levels globally as well as at specific genomic regions,



Figure 11: Schematic representation of the murine Ten-eleven translocation protein family. All three Tet proteins harbour a cystein-rich region (Cys) followed by the catalytic domain characterised by a double stranded β helix (DSBH) fold. Tet1 contains a CXXC-type zinc finger domain (CXXC) in its N-terminal part. In the direct chromosomal vicinity of Tet2 the CXXC4 gene can be found which might be functionally linked to Tet2. A putative ORF close upstream of Tet3 codes for a CXXC-type zinc-finger which might be spliced to Tet3.

such as LINE1 retrotransposons and transcription factor binding sites [Ficz et al., 2011; Xu et al., 2011]. Tet1 has been shown to be important for the demethylation of brain-derived neurotrophic factor (Bdnf) and fibroblast growth factor 1 (Fgf1) promoters in the adult mouse brain [Guo et al., 2011]. Finally, loss of 5-mC in the male pronucleus of zygotes correlates with an increase of 5-hmC staining and Tet3-depletion results in failure to demethylate the paternal genome as well as promoters of pluripotency genes, such as *oct4* and *nanog* [Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011]. The close functional link of Tet proteins to DNA methylation in animals is also evident in the fact that Tets only occur in species that express a DNA methyltransferase (Figure 12). However, Dnmts can be found in species that do not express Tet proteins suggesting that other mechanisms for DNA demethylation exist. In fact, in plants, which do not have genes coding for Tet proteins, DNA demethylation is well studied. Several mechanisms by which 5-hmC contributes to DNA demethylation have been proposed which include DNA repair pathways and prevention of DNA methylation maintenance.

DNA demethylation by DNA repair pathways

DNA demethylation has long been a highly debated field. The only widely accepted mechanism, in mammals, has been passive demethylation where DNA methylation is not maintained through replication and thereby diluted during each cell cycle. However, it has become increasingly clear that DNA demethylation can also occur in a replication independent context e.g. in post-mitotic tissues such as neurons. Moreover, at several stages during development DNA methylation levels decrease with kinetics that can not be accounted for by passive DNA demethylation alone. Active DNA demethylation is

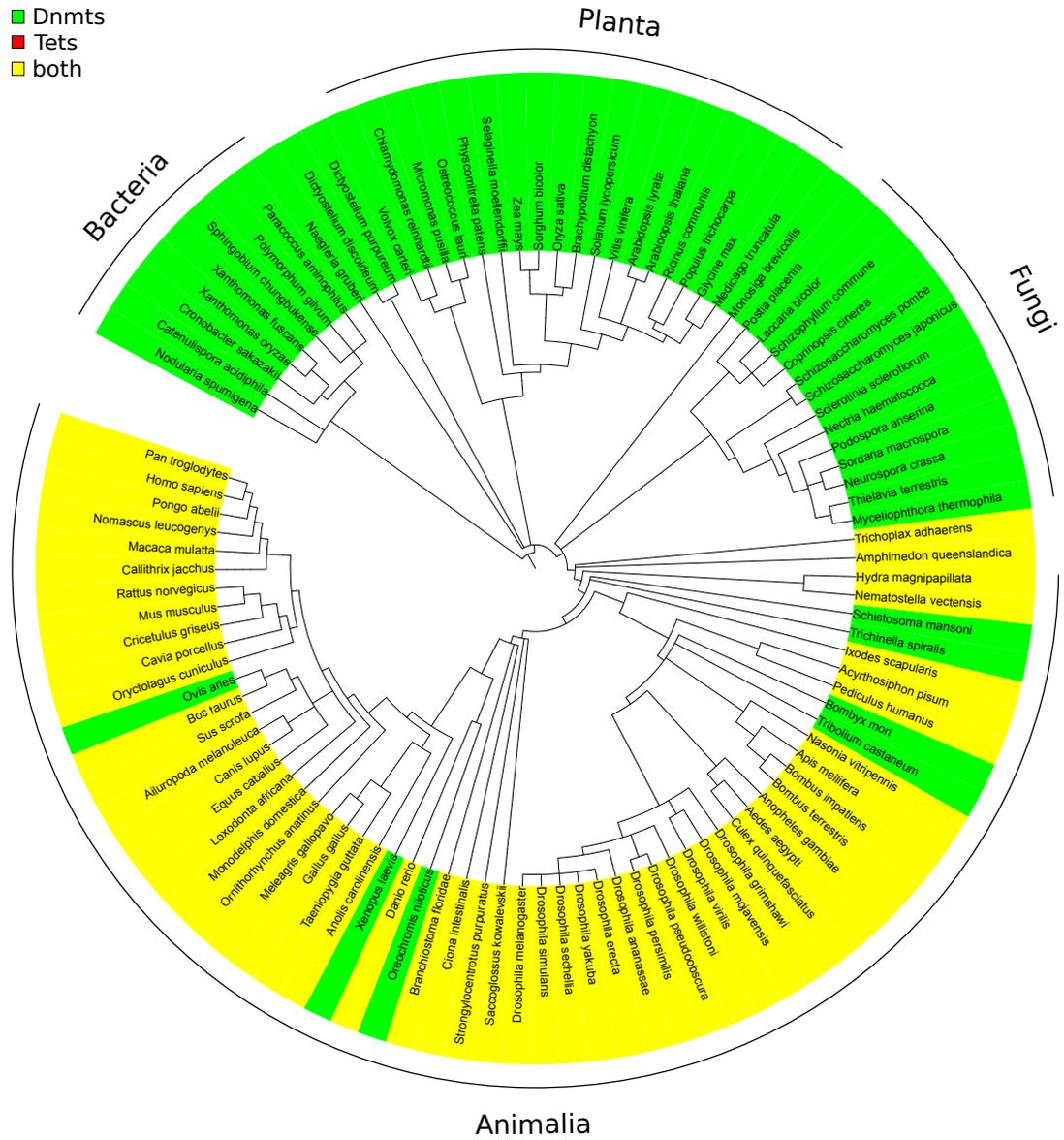


Figure 12: Evolutionary conservation of Tet and Dnmt proteins. Tet proteins only occur in species that also express a Dnmt. The figure was generated using iTOL coupled to a PSI-blast (e-value = $3e-24$) using the mouse sequences of Dnmts (Dnmt1, Dnmt2, Dnmt3a and Dnmt3b) and Tets (Tet1, Tet2 and Tet3) [Letunic and Bork, 2007]

best studied in plants where a family of DNA glycosylases is responsible for the removal of 5-mC. DNA glycosylases cleave the glycosidic bond between 5-mC and the deoxyribose, creating an abasic site (AP). An AP endonuclease removes the deoxyribose at the AP site and the gap is filled by DNA polymerase and DNA ligase. The result of this base excision repair (BER) pathway is the replacement of the methylated cytosine by an unmethylated cytosine [Zhu, 2009].

In mammals similar mechanisms have been proposed. In contrast to the glycosylases in plants, known mammalian glycosylases like thymine DNA glycosylase (TDG) and methyl-CpG-binding domain protein 4 (MBD4) only show weak activity on 5mC *in vitro*. However, these enzymes have strong activities against T:G mismatches which can be created through deamination of 5mC by cytidine deaminases [Zhu et al., 2000]. In fact, cytidine deaminases of the apolipoprotein B mRNA editing catalytic polypeptide (APOBEC) family have been shown to be involved in active DNA demethylation by deaminating 5-mC.

One member of the APOBEC family, the activation-induced cytidine deaminase (AID), has been studied in great detail over the last decade because of its critical role in generating antibody diversity in lymphocytes [Chaudhuri et al., 2007; Delker et al., 2009]. In B-lymphocytes AID takes part in somatic hyper-mutation and class-switch recombination by deaminating cytosines to uracils which in turn are processed by error-prone BER or mismatch repair (MMR) pathways. This mechanism results in mutations essential for the vast diversity of antibodies present in mammals [Liu and Schatz, 2009; Maul and Gearhart, 2010]. For a long time AID was thought to preferentially target the immunoglobulin locus in B-lymphocytes by an unknown mechanism. However, studies in B-lymphocytes of mice deficient in BER and MMR revealed that AID acts extensively on non-immunoglobulin loci and that these regions are protected by error-free repair mechanisms. At that time these findings were interpreted as a protection mechanism against miss targeted AID activity [Liu et al., 2008]. Only recently, AID has been implicated in active DNA demethylation.

First findings suggesting a role for AID in DNA demethylation came from studies done in zebrafish embryos. Overexpression of AID or zebrafish APOBEC deaminases and the DNA glycosylase MBD4 led to DNA demethylation of the genome and of injected methylated DNA [Rai et al., 2008]. Evidence for a role of AID in DNA demethylation in mammals was found in mice completely lacking AID. In the primordial germ cells of these animals an increase in genome-wide methylation was observable. However, AID null mice are viable and fertile suggesting that other redundant pathways

may exist which can compensate for the loss of AID [Popp et al., 2010]. Studies of nuclear reprogramming provided the first evidence that AID plays a role in active DNA demethylation [Bhutani et al., 2010]. Fusion of mouse ESCs with human fibroblast into non-dividing heterokaryons leads to rapid loss of DNA methylation at the promoters of the pluripotency genes OCT4 and NANOG in the somatic genome. This process was shown to be AID-dependent as knock-down of AID using siRNA resulted in complete loss of pluripotency promoter demethylation and transcriptional induction. Moreover, the *AID* gene is located in a cluster of pluripotency genes together with *nanog* and *stella* and is coexpressed with these genes in oocytes, embryonic germ cells and tissues where DNA demethylation has been shown to occur [Morgan et al., 2004; Bhutani et al., 2010]. Besides the APOBEC family, DNA methyltransferases have been proposed to play an important role in active DNA demethylation by deaminating 5-mC. In human breast cancer cells the *de novo* methyltransferases Dnmt3a and 3b can convert 5-mC to T through deamination during the activation of oestradiol-estrogen receptor target gene pS2 by E2. The resulting T:G mismatch is then removed by BER [Mtivier et al., 2008]. Similar observations were made during the activation of the vitronectin gene by the nuclear receptor chicken ovalbumin upstream promoter-transcription factor I (COUP-TFI). Moreover, Dnmt3a was found to interact with the glycosylase TDG and could enhance COUP-TF1-mediated activation of a methylated reporter gene [Gallais et al., 2007]. The involvement of DNA methyltransferases in setting and removing DNA methylation raises the question how these counteracting functions are separated and controlled.

The accumulating evidence for the involvement of deamination-coupled DNA repair in active DNA demethylation led to the identification of several DNA glycosylases involved in this process. The family of glycosylases implicated in the deamination-coupled BER pathway are members of the uracil DNA glycosylase (UDG) family that include TDG, MBD4 and single-stranded-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) [Zhu et al., 2000; Cortellino et al., 2011; Guo et al., 2011]. The DNA glycosylases TDG and SMUG1 have been shown to convert 5-hmU to cytosine suggesting that they act in concert with Tet and AID/APOBEC proteins. Interestingly, knock-out of TDG results in early embryonic lethality underscoring the importance of BER glycosylases during development and DNA demethylation. TDG has been shown to directly interact with AID by immunoprecipitation experiments [Cortellino et al., 2011; Guo et al., 2011]. Recent studies revealed that Tet proteins can further oxidize 5-hmC to 5-fC and 5-caC. Both cytosine derivatives are present in mouse organs and cultured cells although in a much lower abundance than 5-hmC. Interestingly, 5-caC and 5-fC are specifically

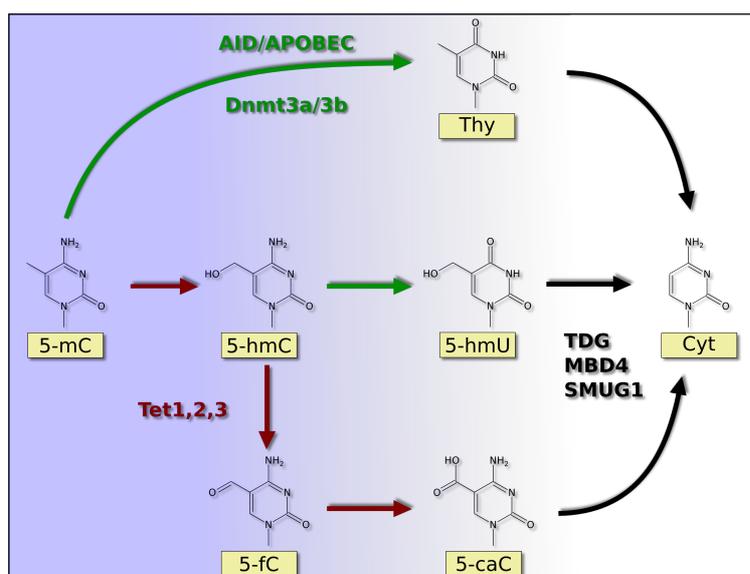


Figure 13: Active DNA demethylation pathways in mammals. 5-mC can be directly deaminated by AID/APOBEC family or Dnmt3 proteins producing a T:G mismatch which is repaired by TDG/MBD4 glycosylases. Alternatively, 5-mC is oxidized to 5-hmC by Tet proteins and either deaminated to 5-hmU by AID/APOBEC proteins or further oxidized to 5-caC. 5-hmU and 5-caC are then recognized by TDG/MBD4/SMUG glycosylases.

recognized by TDG and siRNA mediated depletion of TDG leads to accumulation of 5-caC [He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011].

Taken together these data suggest that in mammals, in contrast to the one-step process in plants, active DNA demethylation may occur in a two-step process. However, the first step of the demethylation pathway either involves deamination of 5-mC/5-hmC by APOBEC family deaminases or further oxidation of 5-hmC by Tet proteins. In either case the modified cytosine is then recognized by glycosylases of the UDG family and 5-mC is replaced by an unmodified cytosine via the BER/MMR pathway (Figure 13).

DNA demethylation by prevention of DNA methylation maintenance

A different mechanism by which 5-hmC may contribute to DNA demethylation is passive DNA demethylation. Hemi-modified 5-hmCpGs are not recognized by Dnmt1 and thereby DNA methylation is not maintained at these sites [Valinluck and Sowers, 2007]. Recently, *in vivo* evidence for this potential mechanism has been found in pre-implantation embryos where 5-hmC is passively lost through replication [Inoue and Zhang, 2011]. However, it is not clear whether 5-hmC mediated passive DNA demethylation is a general mechanism and further studies are required to clarify this question.

1.5.3 Biological function of 5-hmC

5-hmC in pluripotency

Many studies have looked at the roles of 5-hmC and Tet1 in ESCs where Tet1 is predominantly expressed. Immunofluorescence staining for 5-hmC showed that while 5-mC is enriched in pericentromeric heterochromatin, 5-hmC seems to mainly occur in euchromatic regions suggesting that it might be associated with gene activity (Ficz et al., 2011; Szulwach et al., 2011a). Extensive genome-wide mapping of 5-hmC and Tet1 revealed that both are enriched at transcriptional start sites (TSSs) and within gene bodies [Ficz et al., 2011; Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011a; Xu et al., 2011]. Tet1 depletion results in an increase of 5-mC at Tet1-targeted CpG islands suggesting that Tet1 and 5-hmC play a role in maintaining CpG island hypomethylation either by blocking access to the DNA methylation machinery or by demethylating 5-mC [Wu et al., 2011a]. Moreover, 5-hmC has been found to be enriched at binding sites for pluripotency-associated transcription factors suggesting a function in maintaining pluripotency [Ficz et al., 2011; Wu et al., 2011a]. In gene bodies presence of 5-hmC can be consistently correlated with gene expression and several genes are down-regulated upon Tet1 knock-down in ESCs [Ficz et al., 2011; Pastor et al., 2011; Wu et al., 2011b; Xu et al., 2011]. However, a large number of genes are derepressed by Tet1-depletion many of which are Polycomb repressive complex 2 (PRC2) targets [Wu et al., 2011b]. In ESCs, PRC2 targets are often bivalently marked by active (H3K4me3) and repressive (H3K27me3) histone marks and enriched for genes important during development [Bernstein et al., 2006]. These genes are referred to as poised and can rapidly be activated or silenced upon differentiation. Tet1 and 5-hmC have been shown to be enriched at bivalent genes and Tet1 knock-down results in impaired binding of PRC2 to these targets suggesting that maintenance of a hypomethylated state of these genes is required for PRC2 function. [Ficz et al., 2011; Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011a,b]. In addition, Tet1 has been found to interact with the Sin3a co-repressor complex, which is involved in histone deacetylation implying an additional non-catalytic function for Tet1 in gene repression [Williams et al., 2011].

The findings that Tet1 expression as well as 5-hmC levels rapidly decrease upon differentiation suggests a function of Tet1 and 5-hmC in maintaining pluripotency. This hypothesis is supported by the observation that during the reprogramming of fibroblasts into induced pluripotent stem cells (iPSCs), Tet1 is reactivated and 5-hmC levels increase [Tahiliani et al., 2009; Koh et al., 2011]. Tet1 and Tet2 were shown to be reg-

ulated by pluripotency related transcription factors Oct4 and Sox2 and Tet1 depletion in ESCs leads to down-regulation of several pluripotency related genes accompanied by an increase in methylation of their promoters [Ficz et al., 2011; Williams et al., 2011; Wu et al., 2011a]. Furthermore, extra-embryonic lineage markers are derepressed in Tet1- depleted ESCs increasing the transdifferentiation potential of these cells into extra-embryonic tissues upon differentiation into embryoid bodies and in teratoma formation [Ito et al., 2011; Koh et al., 2011]. However, both *tet1* and *tet2* knock-out mice are viable and fertile, although *tet1*^{-/-} mice display an overall reduced body size [Dawlaty et al., 2011; Ko et al., 2011; Li et al., 2011b; Moran-Crusio et al., 2011]. In summary, it is not clear to what extent Tet1 and 5-hmC contribute to pluripotency. Although Tet1 seems to regulate developmentally important genes, it is not crucial for embryonic development. To elucidate whether 5-hmC itself is required for embryogenesis, generation of triple Tet-knockout mice would be necessary as only a 35% reduction in 5-hmC levels in *tet1* knockout ESCs is observable [Dawlaty et al., 2011].

1.6 Transcription activator-like effectors

1.6.1 Biology of TAL effectors

Transcription activator-like effectors (TALEs) were discovered as key proteins in the pathogenicity of xanthomonads. *Xanthomonas* pathogens are Gram-negative phytopathogenic bacteria of considerable agricultural impact that infect a wide range of important crops. The bacteria enter their host through natural openings, such as stomata and hydathodes, or wounds and grow in the intercellular spaces of the plant tissue. The pathogenicity of *Xanthomonas* is dependent on a type III secretion (T3S) system by which the bacteria inject TALEs into the plant cells. TALEs are then imported into the nucleus where they activate genes important for colonization and spreading of the pathogen (Figure 14) [Kay and Bonas, 2009]. TAL effectors are a family of proteins found in many but not all species of *Xanthomonas*. Individual *Xanthomonas* strains can have up to several dozen different TALEs. All TALEs share a common N-terminal domain required for the T3S mediated secretion and a C-terminus containing nuclear localization signals and an acidic transcriptional activation domain (AD). However, TALEs differ in their central domain which consists of nearly perfect, 33-34 amino acid (aa) long repeats. In the beginning of the central repeat domain a non-canonical repeat is found which seems to be conserved among different TALEs. The domain ends with a 20 aa

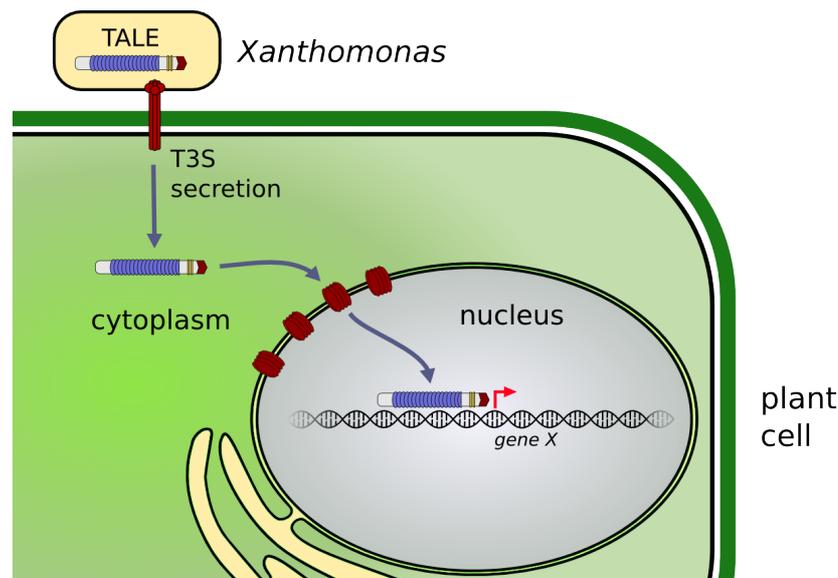


Figure 14: The function of TALEs during *Xanthomonas* infection. TALEs are injected into the plant cell cytoplasm by Type 3 secretion (T3S). Subsequently, the protein is imported into the nucleus where it activates the expression of genes important for pathogen survival.

long truncated repeat. TALEs differ in the number of repeats and the single repeats differ within and across TALEs mainly at residue 12 and 13. These two aa long polymorphisms are referred to as repeat-variable diresidues (RVDs) [Bogdanove et al., 2010] (Figure 15A).

By identifying the sequences that are targeted by different TALEs it was discovered that different RVDs associate preferentially with certain nucleotides, with the four most common RVDs (HD, NG, NI and NN) mediating the binding to the four nucleotides (Figure 15). A non-canonical repeat in the beginning of the central repeat domain seems to bind thymine. With this information it is possible to generate designer TALEs (dTALEs) by arranging the repeats and the string of RVDs to target a defined DNA sequence [Boch et al., 2009; Moscou and Bogdanove, 2009]. The length of the sequences that have been successfully targeted ranged up to 19 base pairs. Target sites are only limited by the requirement for the presence of a thymine as the first nucleotide. Intriguingly, the length of the targeted sequence implies that for binding such a long sequence the TALE has to follow the helical turns of the DNA. This model is supported by the recently published crystal structures of two different TALEs binding to DNA. TALEs seem to bind DNA by following the helical turns along the major groove [Deng et al., 2012; Mak et al., 2012] (Figure 15B and C).

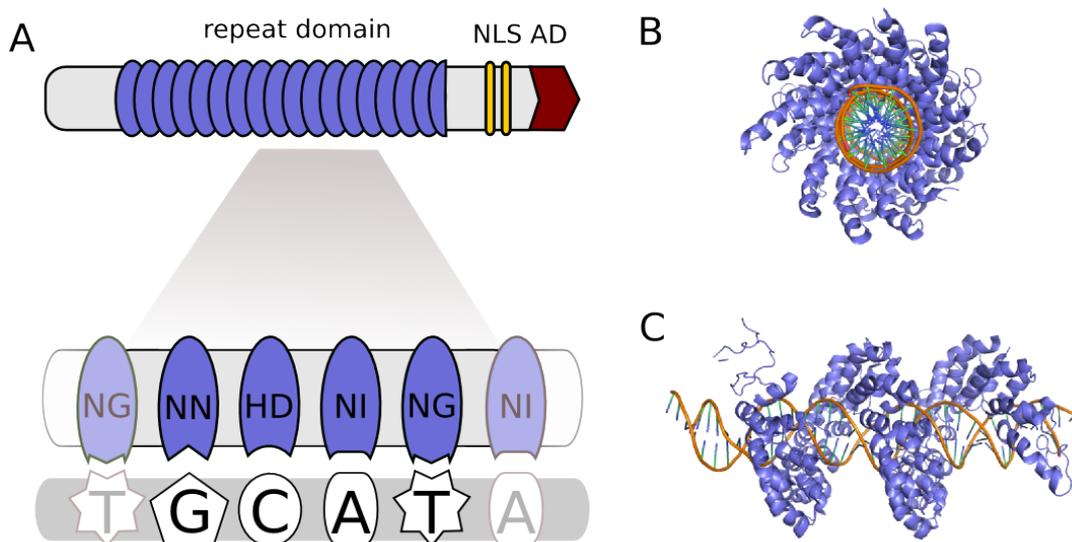


Figure 15: DNA recognition of TALEs. (A) Schematic representation of a TALE with its central repeat domain, the C-terminal NLS and activation domain (AD). Each repeat is represented by a blue ellipse. The RVDs responsible for the binding to the four nucleotides are NN, HD NI and NG. (B-C) Crystal structure of the TAL effector PthXo1 bound to its DNA target (PBD 3ugm). (B) Bottom view. (C) Side view.

1.6.2 Designer TAL effectors as a tool for genome editing

The discovery of the TALE code has led to a broad application of TAL effectors in biotechnology. The straightforward relationship between the type of repeat and DNA sequence it recognizes allows a simple designing of dTALEs to target any desired DNA sequence. Several strategies have been developed that allow cheap, reliable and quick assemble of dTALEs with desired specificities. All strategies are based on Golden Gate cloning a method that relies on type II restriction enzymes which cleave outside their recognition sequence and create unique 4-bp overhangs. Well planned design of the overhangs produced by these enzymes allows ligation of multiple DNA fragments in an ordered fashion in a single reaction [Engler et al., 2008, 2009].

The main application is the use of dTALEs for targeted editing of genomic sequences. For this purpose dTALEs are used to target the catalytic domain of the FokI nuclease to specific sites in the genome to introduce DNA double-strand breaks (DSBs). Because FokI nucleases work as dimers, the TALE nucleases (TALENs) are designed as pairs that bind opposing DNA target sites which are separated by a spacer (Figure 16). This layout allows the FokI monomers to dimerize only at the desired location and create a DSB.

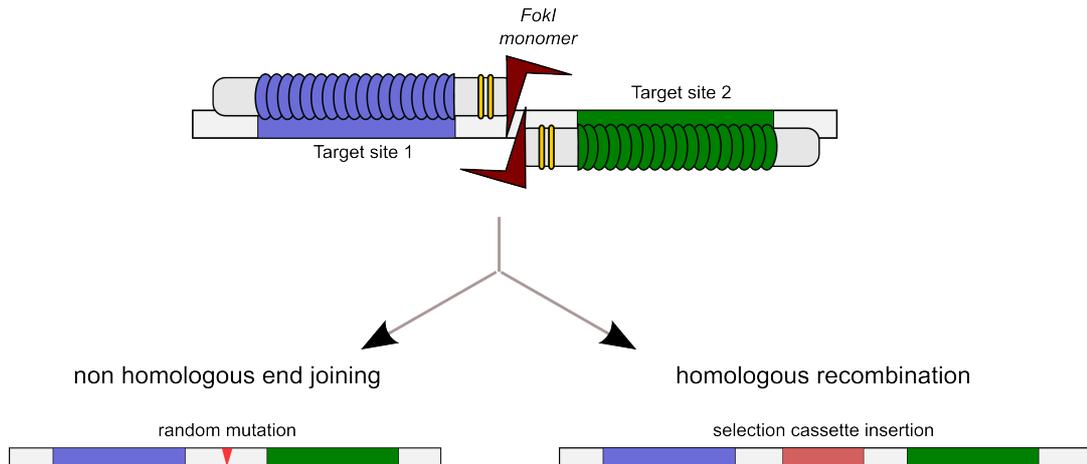


Figure 16: designer TALEs as tool for genomic targeting. dTALEs are fused to FokI monomers to generate TALE nucleases (TALENs). TALENs are designed as pairs binding opposing DNA target sites which are separated by a spacer. When both TALENs bind their target sites the FokI nucleases dimerize and generate a double-strand break (DSB) within the spacer sequence. The DSB is repaired by either non homologous end joining or homologous recombination. The latter is used to insert selection cassettes for further selection of clones containing mutated alleles.

The created DSBs are then repaired by one of two highly conserved pathways which can be exploited to introduce specific DNA sequence modifications [Carroll, 2011]. In non homologous end joining (NHEJ), the broken chromosome is repaired by imprecise rejoining. NHEJ therefore introduces random mutations in the form of small deletions or insertions at the break site which can disrupt gene function. The second pathway that can be induced by DSBs is homologous recombination (HR) (Figure 16). HR is an error-free DNA repair pathway in which homologous sequences of the sister chromatid, or any other homologous template in close proximity, are used to repair the damaged site. This pathway can be exploited to introduce specific mutations or additional sequences into the targeted locus.

First findings that TALENs can be used to introduce site-specific mutations came from reporter assays performed in yeast (Christian et al., 2010). Subsequent studies showed the potential of TALENs to mediate site-directed mutagenesis in human embryonic kidney cells with high efficiency [Miller et al., 2011]. Since then TALENs have been applied in a variety of experimental systems. Human embryonic stem cells and induced pluripotent stem cells were successfully modified at five different loci by TALEN-mediated targeting [Hockemeyer et al., 2011]. In plants TALENs were shown to cleave an episomal target in leaves of tobacco as well as to introduce NHEJ-mediated mutagenesis of an endogenous locus in *Arabidopsis thaliana* [Cermak et al., 2011; Mahfouz et al.,

2011]. In *Caenorhabditis elegans* and *Caenorhabditis briggsae* TALENs were used to introduce mutations in germ line cells of adults which successfully produced off-spring with mutations in the targeted genomic loci [Wood et al., 2011]. Somatic and heritable gene knock-outs have been made in zebrafish using TALENs [Huang et al., 2010; Sander et al., 2011]. Furthermore, IgM knock-out rats were generated by embryo microinjection of TALEN DNA or mRNA constructs [Tesson et al., 2011].

A general concern in using targeted nucleases for genome editing is the possibility of mutagenesis at unintended sites. This is of particular importance for therapeutic applications of TALENs in humans. Although specificity is expected to be relatively high due to the required pairing of two TALENs to generate an active nuclease, it is not clear how many unintended mutations are introduced by TALENs. Complete genome sequencing before and after TALEN treatment would need to be performed in order to answer this question.

1.7 Aims of this work

DNA methylation plays a central role in epigenetic regulation of mammalian gene expression. The overall pattern of DNA methylation is established during early development and is essential for lineage specification and maintenance of differentiated states. These observations led to the notion that DNA methylation represents a stable epigenetic mark which is subjected to little changes after the initial setting of cell type specific patterns. In the last decade this idea evolved into a more dynamic view where DNA methylation levels can rapidly change in response to internal and external signalling. However, it remained highly debated what mechanisms are used to remove DNA methylation marks, a process that is required in order to explain the observable DNA methylation dynamics. Recently, the Tet family of enzymes were described to catalyse the oxidation of 5-methyl-cytosine (5-mC) to 5-hydroxymethyl-cytosine (5-hmC), a potential intermediate in DNA demethylation.

At first, I aimed at developing a new set methods to quantify and map 5-hmC in genomic DNA. Using these new techniques I wanted to gain first insights into the function of 5-hmC and Tet enzymes by analysing the abundance of this modification in mouse tissues, during differentiation and in myeloid leukemia. Moreover, I tried to analyse the functional links between 5-hmC and other epigenetic networks by studying the 5-hmC binding characteristics of known 5-mC binding proteins as well as by analysing the function of the Tet1 CXXC domain.

Furthermore, I tried to identify new epigenetic factors and networks involved in transgene silencing. For this I wanted to develop an assay based on the different silencing characteristics of two commonly used promoter systems. At last, I aimed at studying the function of designer transcription activator-like effectors (dTALEs) in activating transcription of the pluripotency gene *oct4*.

2 Results

2.1 Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA

Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA

Aleksandra Szwagierczak, Sebastian Bultmann, Christine S. Schmidt, Fabio Spada* and Heinrich Leonhardt*

Department of Biology, Center for Integrated Protein Science Munich, Ludwig Maximilians University Munich, 82152 Planegg-Martinsried, Germany

Received June 14, 2010; Revised July 19, 2010; Accepted July 21, 2010

ABSTRACT

The recent discovery of genomic 5-hydroxymethylcytosine (hmC) and mutations affecting the respective Tet hydroxylases in leukemia raises fundamental questions about this epigenetic modification. We present a sensitive method for fast quantification of genomic hmC based on specific transfer of radiolabeled glucose to hmC by a purified glucosyltransferase. We determined hmC levels in various adult tissues and differentiating embryonic stem cells and show a correlation with differential expression of *tet* genes.

INTRODUCTION

DNA methylation plays a crucial role in the epigenetic regulation of gene expression during development and disease (1). The post-replicative addition of a methyl group to the carbon-5 of cytosine has long been the only known enzymatic modification to bases in mammalian genomic DNA, and due to its crucial role as an epigenetic mark it is often referred to as the fifth base. Recently, the Ten-Eleven Translocation 1 gene (*tet1*) was shown to encode a 2-ketoglutarate- and Fe(II)-dependant hydroxylase that converts genomic 5-methylcytosine (mC) to 5-hydroxymethylcytosine (hmC) and on the basis of sequence homology the closely related Tet2 and 3 proteins are expected to catalyze the same reaction (2,3). To date hmC has been detected in genomic DNA isolated from embryonic stem cells (ESCs) and some adult tissues and it appears to be relatively abundant in the central nervous system (2,4–6). The functional relevance of hmC in these tissues is unknown and roles as an epigenetic mark and/or an intermediate of oxidative demethylation are intriguing possibilities (2,7). In addition,

translocations as well as nonsense and missense mutations of *tet* genes have been identified in myelodysplastic syndromes including several forms of myeloid leukemia (8–10), raising the possibility that aberrant genomic hmC patterns may be involved in these pathologies. These observations grant sustained efforts to define the role(s) of hmC in mammalian genomes.

Quantification and selective detection of genomic hmC is technically challenging due to the relatively low abundance and similarity of hmC to the more abundant mC, not only in structural terms but also with respect to lack of deamination by bisulfite treatment (11–12). We sought to exploit enzymes involved in hmC modification that evolved as part of the struggle between prokaryotes and their viruses.

The three methods used so far to quantify global hmC content in mammalian genomes are designed to detect hmC in hydrolyzed DNA globally (HPLC/esi-ms/ms) or at subsets of CpG sites (2,4,5). As hmC may also occur at non-CpG sites, the latter type of methods may underestimate its abundance. In addition, none of these procedures is easily applicable to large sample numbers. We therefore sought to establish a highly sensitive and accurate method to detect hmC independently of sequence context and with higher throughput capacity. To this aim, we turned our attention to glucosyltransferases of T-even bacteriophages that transfer glucose from UDP-glucose donor to genomic hmC. Notably, all cytosines in the T4 genome are replaced by hmC residues that are invariably modified by α - and β -glucosyltransferases (α - and β -gt; Figure 1A). We reasoned that by using UDP-[³H]glucose the incorporation of radiolabeled glucose in DNA should reflect the abundance of hmC. We focused on β -gt rather than α -gt, as it was shown to glucosylate to completion all tested hmC-containing DNA substrates both *in vivo* and *in vitro*, including the non-glucosylated T4 genome

*To whom correspondence should be addressed. Tel: +49 89 2180 74230; Fax: +49 89 2180 74236; Email: f.spada@lmu.de
Correspondence may also be addressed to Heinrich Leonhardt. Tel: +49 89 2180 74232; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2010. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

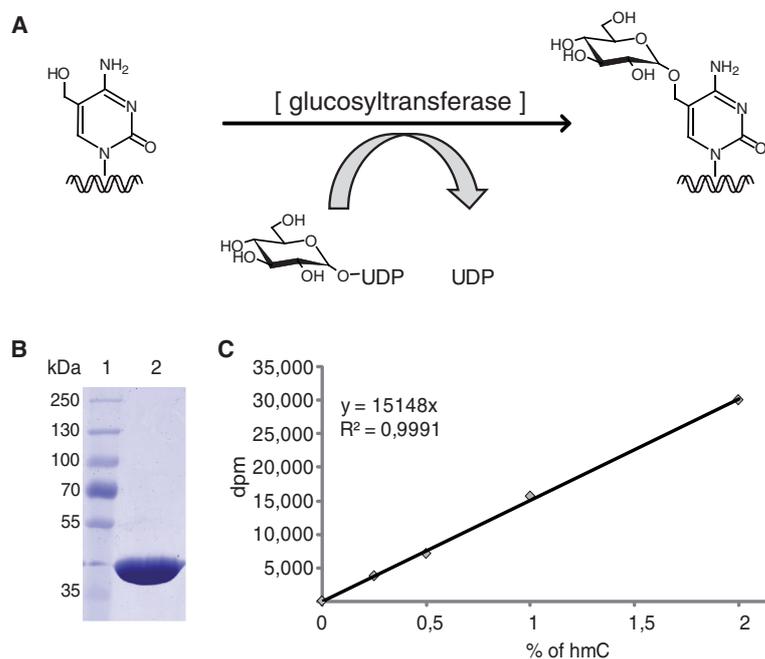


Figure 1. Elements of the hmC glucosylation assay. (A) Schematic representation of the hmC glucosylation reaction catalyzed by β -gt. (B) Coomassie blue stained gel showing the purified β -gt preparation. (C) Example of calibration curve using mixtures of hmC-containing and unmodified reference fragment (equal total DNA amounts). Note the linear relationship between [3 H]glucose incorporation and percentage of hmC.

(13,14). This indicates that β -gt can glucosylate hmC independently of DNA sequence and structural context and therefore is ideally suited for this assay.

MATERIALS AND METHODS

Cell culture and differentiation of ESCs

Undifferentiated J1 and E14 ESCs were maintained on gelatin-coated dishes in Dulbecco's modified Eagle's medium containing 16% fetal bovine serum (PAA Laboratories GmbH), 0.1 mM β -mercaptoethanol (Invitrogen), 2 mM L-glutamine, 1 \times MEM Non-essential Amino Acid Solution, 100 U/ml penicillin, 100 μ g/ml streptomycin (PAA Laboratories GmbH) and 1000 U/ml recombinant mouse LIF (Millipore). To induce embryoid body (EB) formation, ESCs were resuspended in the same medium as above but without LIF and cultured in hanging drops (600 cells/20 μ l drop) for 4 days. Subsequently, EBs were cultured in bacterial culture dishes and the medium was replaced every 4 days.

DNA and RNA isolation

All tissue samples were prepared from 6-week-old 129sv mice. Genomic DNA and total RNA were isolated from tissue samples using the NucleoSpin Triprep Kit (Macherey-Nagel). Genomic DNA and total RNA were isolated from ESCs and EBs using the Blood & Cell culture DNA mini kit (QIAGEN) and TRIzol reagent

(Invitrogen), respectively. To avoid genomic DNA contamination, isolated RNA was digested with recombinant RNase-free DNase I (Roche) and further purified with the QIAGEN RNeasy kit. Genomic DNA samples were sheared to 500–1500 bp fragments by sonication to reduce the viscosity and improve homogeneity. The concentration of genomic DNA samples was measured by fluorometry. Fifty microliters of diluted sample were mixed with 50 μ l of 2 \times TNE (Tris 20 mM, pH 7.4; NaCl 400 mM and EDTA 2 mM) containing 200 ng/ μ l of Hoechst 33258. Fluorescence was measured in a TECAN infinite M1000 plate reader (Ex: 350/10; Em: 455/10). Serial dilutions (20–2000 ng/ml) of the hmC containing reference DNA fragment (see below) were used as standard for quantification.

Protein expression and purification

The sequence encoding bacteriophage T4 β -gt was synthesized at Mr. Gene GmbH (Regensburg) and cloned into pET28b vector (Novagen). BL21(DE3) *E. coli* cells carrying the expression construct were grown at 37°C until $A_{600} = 0.6$ – 0.7 and induced with 1 mM isopropyl β -D-thiogalactopyranoside for 16 h at 20°C. Lysates were prepared by sonication in 300 mM NaCl, 50 mM Na_2HPO_4 , 10 mM imidazole, 1 mM β -mercaptoethanol, cleared by centrifugation and applied to nickel-nitrilotriacetic acid column (QIAGEN) pre-equilibrated with lysis buffer. Washing and elution were performed with lysis buffer containing 20 and 250 mM imidazole,

respectively. Eluted proteins were applied to a Superdex S-200 preparative gel filtration column (GE Healthcare) in 150 mM NaCl, 20 mM Tris, pH 8.0, 1 mM DTT. Fractions containing the β -gt peak were pooled and applied to a ResourceQ anion exchange column (GE Healthcare) in order to eliminate residual contaminants, resulting in pure β -gt in the flowthrough.

Preparation of reference DNA fragments

Reference DNA fragments (1139 bp) containing 0 and 100% hmC were prepared by polymerase chain reaction (PCR), using dCTP and 5-hydroxymethyl-dCTP (Bioline GmbH), respectively. T4 phage DNA was used as template with primers: 5'-TGG AGA AGG AGA ATG AAG AAT AAT-3' and 5'-GTG AAG TAA GTA ATA AAT GGA TTG-3', Phusion HF DNA Polymerase (Finnzymes) and the following cycling profile: one cycle of 98°C for 2 min and 35 cycles of 98°C for 10 s; 58°C for 10 s; and 72°C for 30 s. Primer sequences were selected that do not contain cytosine residues. PCR products were purified by gel electrophoresis followed by silica column purification (Nucleospin, Macherey-Nagel).

Quantitative hmC glucosylation assay

Reactions contained 150 mM NaCl, 20 mM Tris, pH 8.0, 25 mM CaCl₂, 1 mM DTT, 2.8 μ M 'cold' UDP-glucose (Sigma-Aldrich), 0.86 nM UDP-[³H]glucose (glucose-6-³H; 60 Ci/mmol; Hartmann Analytic GmbH), 1 μ g of DNA substrate and 36 nM recombinant β -gt in a total volume of 50 μ l. Reactions were incubated for 20 min at room temperature and terminated by heating at 65°C for 10 min. Twenty microliters of each reaction were spotted in duplicate on paper filters (Whatmann) and DNA was precipitated by incubation in 5% TCA for 15 min at room temperature. Filters were washed twice with 5% TCA and once with 70% ethanol. Remaining radioactivity was measured using a Liquid Scintillation Analyzer Tri-Carb 2100TR (Packard) with quench indicating parameter set on tSIE/AEC (transformed spectral index of the external standard/automatic efficiency control) in 4 ml of Rotiszint Eco Plus scintillation liquid (Roth GmbH) in Snaptwist vials (Simport). Samples were measured for 30 min or until the 2 σ value reached 2%. The percentage of hmC per total cytosine was calculated from the incorporation of [³H]glucose using a calibration curve measured with the reference fragment series for every experiment. The percentage of hmC was then corrected for the difference in C abundance between reference fragment (35%) and mouse genome (42%).

cDNA synthesis and real-time PCR

Five hundred nanograms of total RNA were used for cDNA synthesis with the High-Capacity cDNA Reverse Transcription Kit (with RNase Inhibitor; Applied Biosystems). Equal amounts of cDNA were used for real-time PCR with Power SYBR Green PCR Master Mix (Applied Biosystems) on a 7500 Fast Real-Time PCR System (Applied Biosystems) according to the manufacturer's instructions. Gene expression levels were

normalized to Gapdh and calculated using the comparative C_T method ($\Delta\Delta C_T$ method).

Primers for quantitative real-time PCR were designed with the Primer Express software (Applied Biosystems) and contained the following sequences: Gapdh forward 5'-CAT GGC CTT CCG TGT TCC TA-3', Gapdh reverse 5'-CTT CAC CAC CTT CTT GAT GTC ATC-3', Tet1 forward 5'-CCA GGA AGA GGC GAC TAC GTT-3', Tet1 reverse 5'-TTA GTG TTG TGT GAA CCT GAT TTA TTG T-3', Tet2 forward 5'-ACT TCT CTG CTC ATT CCC ACA GA-3', Tet2 reverse 5'-TTA GCT CCG ACT TCT CGA TTG TC-3', Tet3 forward 5'-GAG CAC GCC AGA GAA GAT CAA-3' and Tet3 reverse 5'-CAG GCT TTG CTG GGA CAA TC-3'.

RESULTS AND DISCUSSION

T4 β -gt was expressed in bacteria as a 6 \times His tag fusion and purified to homogeneity by sequential nickel-NTA, size exclusion and ion exchange chromatography (Figure 1B). To assess whether transfer of [³H]glucose to DNA is proportional to the hmC content within the range previously reported for mammalian tissues, we prepared a series of standard DNA substrate samples with global hmC content ranging from 0.25 to 2% of total cytosine by mixing corresponding proportions of two preparations of the same 1.2 kb DNA fragment, one having all cytosine residues replaced by hmC and the other containing no hmC (Figure 1C). Using a 325-fold excess of unlabelled UDP-glucose, the incorporation of radiolabeled glucose in 1 μ g of total DNA substrate was strictly linear in this range. This standard sample series was measured in every assay to generate a calibration curve for the calculation of hmC content in genomic DNA samples. We first measured genomic hmC levels in wild-type and Dnmt1, 3a and 3b triple knockout (TKO) J1 ESCs (15) (Figure 2A and B). Due to the absence of all three major DNA methyltransferases, genomic DNA from TKO ESCs is expected to contain very little, if any, cytosine methylation. Indeed, the measured level of genomic hmC in TKO ESCs was at the detection limit (0.025%) of our assay, while genomic DNA from wild-type ESCs contained 0.3% hmC relative to total cytosine. Real-time reverse transcription (RT) PCR analysis showed that Tet1–3 mRNA levels are similar in wild-type and TKO ESCs, with Tet1 transcripts largely preponderant and Tet3 mRNA the least abundant (more than 40-fold lower than Tet1). It was previously shown that differentiation of mouse ESCs by withdrawal of LIF from monolayer cultures for 5 days results in a reduction of genomic hmC and concomitant decrease in Tet1 transcripts (2). We followed genomic hmC and Tet1–3 transcript dynamics during EB differentiation of two commonly used wild-type ESC lines (Figure 2A and B and Supplementary Figure S1). In both cases, a sharp decrease of hmC content was evident after 4 days of EB culture, but a substantial recovery was observed after additional 4 days of culture (Day 8). Interestingly, the *tet* genes showed distinct expression dynamics during ESCs

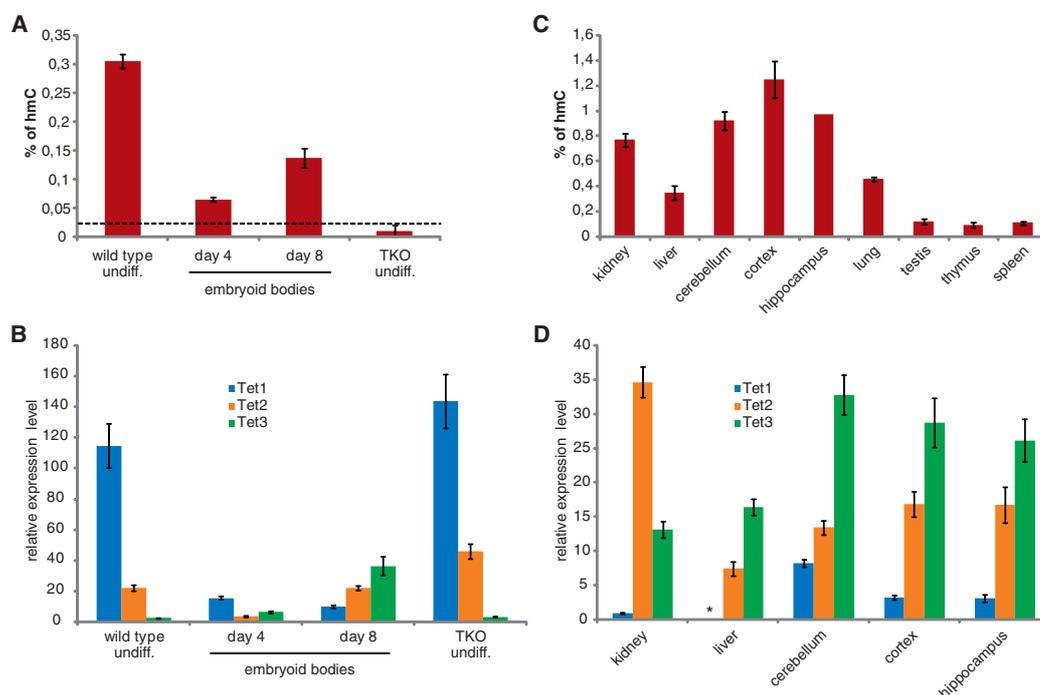


Figure 2. Quantification of genomic hmC and Tet transcripts in mouse tissues, undifferentiated ESCs and EBs. (A and C) hmC glucosylation assays. The percentage of hmC per total cytosine was calculated from the incorporation of [3 H]glucose using a calibration curve from the reference fragment (see Figure 1C). Shown are average values and error bars from two (A) or one (C) biological replicates, each measured in two independent assays, with the exception of hippocampus that was measured only once. In every assay, each sample was measured in duplicate. The dashed line in (A) indicates the estimated limit of detection. (B and D) Real-time RT-PCR analysis for Tet transcript levels. Expression levels are all relative to Tet1 in kidney (set to 1), so that values in b and d are directly comparable. Shown are average values and error bars from two (B) and one (D) biological replicates, each measured from two independent cDNA synthesis reactions. In every real-time PCR reaction, each sample was measured in triplicate. Genomic DNA and RNA samples used in A/C and B/D, respectively, were isolated from the very same cell and tissue lysates.

differentiation. Tet1 transcripts drastically decreased in the first 4 days and further dropped by Day 8 of EB culture. Tet2 mRNA levels also decreased in the first 4 days, but were fully restored in Day 8 EBs. In contrast, Tet3 transcripts doubled at Day 4 and increased about 20 times by Day 8 of EB culture. Thus, the relatively high hmC content in undifferentiated ESCs correlates with high levels of Tet1 and, to a lower extent, Tet2 transcripts, while the partial recovery of genomic hmC in Day 8 EBs correlates with increased Tet2 and Tet3 mRNA levels.

We then analyzed several adult mouse tissues (Figure 2C and D). As reported earlier, the highest levels of genomic hmC were found in brain regions, although kidney also showed relatively high levels. In all cases, the hmC content was at least four times higher than the detection limit, while in a previous report using a different assay the same non-neural tissues were either marginally above or right at the detection limit (5). Abundant hmC in brain tissues correlated with high levels of Tet3 and to a lower extent Tet2, a pattern similar to Day 8 EBs. Thus, most differentiated tissues are characterized by very low levels of Tet1 and high levels of Tet3, while undifferentiated ESCs show the opposite pattern. It will be

interesting to determine whether all pluripotent cell types have this pattern and at which stages along the specification of the various somatic lineages the relative expression levels of *tet* genes change. Interestingly, kidney represents an exception among the adult tissues analyzed as it shows relatively high hmC content and a prevalence of Tet2 transcripts. This is consistent with a cellular defect in proximal convoluted tubules of the kidney as the only phenotype described for Tet2 null mice (16). These observations suggest that Tet proteins have redundant roles and that the lack of a specific Tet family member may result in phenotypic alterations only in tissues where high levels of that Tet enzyme cannot be compensated by the other family members. In this context, it should be noted that the assay described here could also be employed to measure the enzymatic activity of Tet proteins and their mutants identified in leukemia patients by using mC-containing DNA substrates.

In conclusion, we have established an accurate assay for the quantification of genomic hmC that: (i) is more sensitive than previously described methods; (ii) is not subject to sequence bias; (iii) allows simultaneous processing of large sample numbers; and (iv) does not require specialized and expensive equipment. It should be noted

that lower concentrations of 'cold' UDP-glucose should allow scaling down the amount of substrate DNA without loss of signal. This assay will be highly useful to determine the global abundance of hmC in genomic DNA, especially in situations where limited amounts of tissue are available such as isolates of rare primary cell types and clinical samples.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Daniela Meilinger, Andrea Steiner and Andreas Brachmann (Ludwig Maximilians University Munich) for help with EB culture, dissection of mouse tissues and scintillation counting, respectively. TKO ESCs were provided by Masaki Okano (RIKEN Center for Developmental Biology, Kobe, Japan).

FUNDING

Deutsche Forschungsgemeinschaft (Grants DFG SPP 1356, SFB 646 and 684 to H.L.); financial support by the Elite Network of Bavaria (International Doctorate Program NanoBioTechnology) and the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS to C.S.S.). S.B. is a fellow of the graduate school Life Science Munich (LSM). Funding for open access charge: Deutsche Forschungsgemeinschaft (DFG SPP 1356, SFB 646 and 684 to H.L.).

Conflict of interest statement. None declared.

REFERENCES

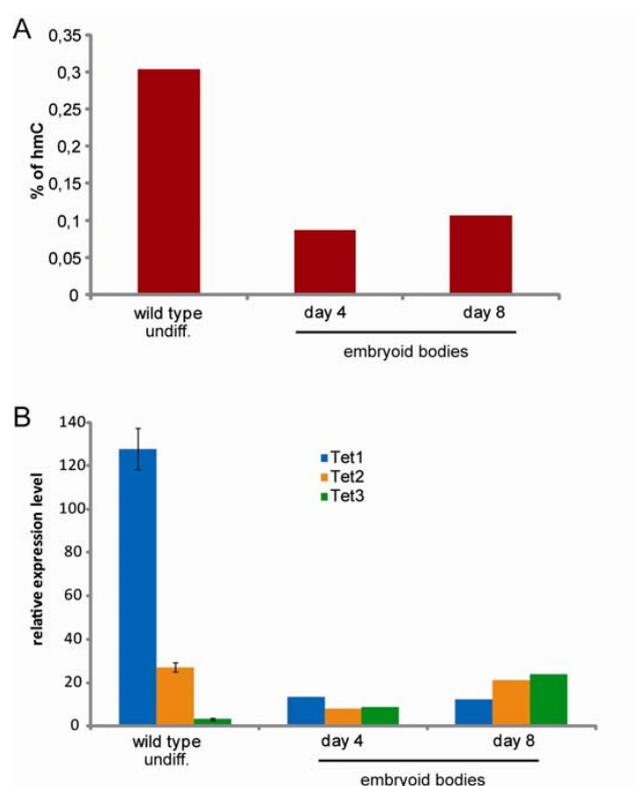
- Rottach,A., Leonhardt,H. and Spada,F. (2009) DNA methylation-mediated epigenetic control. *J. Cell. Biochem.*, **108**, 43–51.
- Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
- 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.
- Feng,J., Zhou,Y., Campbell,S.L., Le,T., Li,E., Sweatt,J.D., Silva,A.J. and Fan,G. (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.*, **13**, 423–430.
- Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- Penn,N.W., Suwalski,R., O'Riley,C., Bojanowski,K. and Yura,R. (1972) The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.*, **126**, 781–790.
- Loenarz,C. and Schofield,C.J. (2009) Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chem. Biol.*, **16**, 580–583.
- 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.
- Ono,R., Taki,T., Taktani,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 Res.*, **62**, 4075–4080.
- Langemeijer,S.M.C., Kuiper,R.P., Berends,M., Knops,R., Aslanyan,M.G., Massop,M., Stevens-Linders,E., van Hoogen,P., van Kessel,A.G., Raymakers,R.A.P. *et al.* (2009) Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.*, **41**, 838–842.
- Hayatsu,H. and Shiragami,M. (1979) Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. *Biochemistry*, **18**, 632–637.
- Huang,Y., Pastor,W.A., Shen,Y., Tahiliani,M., Liu,D.R. and Rao,A. (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE*, **5**, e8888.
- Georgopoulos,C.P. and Revel,H.R. (1971) Studies with glucosyl transferase mutants of the T-even bacteriophages. *Virology*, **44**, 271–285.
- Kornberg,S.R., Zimmerman,S.B. and Kornberg,A. (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. *J. Biol. Chem.*, **236**, 1487–1493.
- Tsumura,A., Hayakawa,T., Kumaki,Y., Takebayashi,S., Sakaue,M., Matsuoka,C., Shimotohno,K., Ishikawa,F., Li,E., Ueda,H.R. *et al.* (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells*, **11**, 805–814.
- Tang,H., Araki,K., Li,Z. and Yamamura,K. (2008) Characterization of Ayu17-449 gene expression and resultant kidney pathology in a knockout mouse model. *Transgenic Res.*, **17**, 599–608.

Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA

Aleksandra Szwagierczak, Sebastian Bultmann, Christine Schmidt, Fabio Spada and Heinrich Leonhardt

Ludwig Maximilians University Munich, Department of Biology, Center for Integrated Protein Science Munich (CIPS^M), 82152 Planegg-Martinsried, Germany.

SUPPLEMENTARY FIGURE S1



Quantification of genomic hmC and Tet transcripts in undifferentiated E14 embryonic stem cells (ESCs) and embryoid bodies (EBs). **(A)** hmC glucosylation assay. The percentage of hmC over total cytosine is calculated from the incorporation of [³H]glucose using a calibration curve from the reference fragment (see Fig. 1C) and corrected for the difference in cytosine abundance between the latter (35%) and mouse genome (42%). Shown are values from one assay, where each sample was measured in duplicate. **(B)** Real time RT-PCR analysis for Tet transcript levels. Shown are average values and error bars from two independent cDNA synthesis reactions. In every PCR reaction each sample was measured in triplicate. Genomic DNA and RNA samples used in A and B, respectively, were isolated from the very same ESC and EB lysates.

2.2 Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine

Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine

Aleksandra Szwagierczak, Andreas Brachmann, Christine S. Schmidt, Sebastian Bultmann, Heinrich Leonhardt* and Fabio Spada*

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

Received October 18, 2010; Revised February 14, 2011; Accepted February 15, 2011

ABSTRACT

In mammalian genomes a sixth base, 5-hydroxymethylcytosine (^{hm}C), is generated by enzymatic oxidation of 5-methylcytosine (^mC). This discovery has raised fundamental questions about the functional relevance of ^{hm}C in mammalian genomes. Due to their very similar chemical structure, discrimination of the rare ^{hm}C against the far more abundant ^mC is technically challenging and to date no methods for direct sequencing of ^{hm}C have been reported. Here, we report on a purified recombinant endonuclease, PvuRts1I, which selectively cleaves ^{hm}C-containing sequences. We determined the consensus cleavage site of PvuRts1I as ^{hm}CN_{11–12}/N_{9–10}G and show first data on its potential to interrogate ^{hm}C patterns in mammalian genomes.

INTRODUCTION

In higher eukaryotes, only the C5 position of genomic cytosine is subject to enzymatically catalyzed post-replicative modification. Methylation at this position has long been known to play major roles in epigenetic control of transcriptional activity and, as a consequence, to affect fundamental processes such as development (including natural reprogramming of cell fate), imprinting, X chromosome inactivation, genome stability and predisposition to neoplastic transformation (1,2). The recent discovery of the further modification of 5-methylcytosine (^mC) to 5-hydroxymethylcytosine (^{hm}C) by the family of Tet dioxygenases has raised major questions on the functional relevance of this sixth base in mammalian genomes (3,4). While recent evidence supports a role for ^{hm}C as an intermediate in the erasure of cytosine methylation (5), other roles in controlling

genomic functions cannot be excluded. The definition of these roles will require profiling of genomic ^{hm}C patterns, which presents a major technical challenge as ^{hm}C is structurally and chemically very similar to ^mC but in general far less abundant in mammalian genomes (3,4,6–9). The gold standard methodology for profiling of genomic ^mC sites, bisulfite conversion, cannot discriminate ^{hm}C from ^mC and all available restriction endonucleases are either equally sensitive to ^mC and ^{hm}C or not sensitive to either (10–12). While antibodies raised against ^{hm}C are commercially available, their use to probe ^{hm}C frequency by DNA immunoprecipitation has yet to be reported and the accuracy of this method will depend on the relative affinity of these antibodies for ^{hm}C versus ^mC as the latter is present in large excess in mammalian genomes. Very recently enzymatic methods for selective labeling and identification of ^{hm}C have been reported (7,13).

Interestingly, ^{hm}C is also present in the genomes of viruses that infect bacteria and unicellular algae, where it serves as protection against the restriction systems of the host. In particular, ^{hm}C accounts for up to 100% of the cytosine residues in the genomes of T-even coliphages. In these phages the hydroxymethyl group is added at the level of the dCMP precursor and further linked to glucose (in both α - and β -configurations) or gentiobiose after incorporation of the nucleotide in the genome (14–16). We sought to exploit enzymatic activities that evolved as part of the struggle between bacteria and these viruses to selectively detect ^{hm}C in mammalian genomes. Recently, we described an assay for quantification of global genomic ^{hm}C levels based on the transfer of tritiated glucose to ^{hm}C by T4 β -glucosyltransferase (7). Interestingly, restriction systems have evolved in bacteria that address the phage counter defense measures by specifically recognizing modified cytosine. Among these the McrBC system and the recently described MspJI endonuclease recognize sequences containing both ^mC and ^{hm}C (17,18) and

*To whom correspondence should be addressed. Tel: +49 89 2180 74230; Fax: +49 89 2180 74236; Email: f.spada@lmu.de
Correspondence may also be addressed to Heinrich Leonhardt. Tel: +49 89 2180 74232; Fax +49 89 2180 74236; Email: h.leonhardt@lmu.de

© The Author(s) 2011. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

therefore *per se* are not useful to discriminate these modified cytosines. At least two endonucleases, PvuRtsI and GmrSD, were shown to restrict DNA containing glucosylated ^{hm}C (19,20). However, GmrSD does not cleave non-glucosylated (^{hm}C-containing) T4 DNA, has the additional disadvantages of being a heterodimer and of co-purifying with the GroEL chaperonin (19). PvuRtsI is encoded by a single gene present on the kanamycin resistance plasmid RtsI originally isolated from *Proteus vulgaris* and its restriction activity *in vivo* was shown to be modulated by ^{hm}C glucosylation in a complex fashion (20). However, as PvuRtsII was not purified, its activity has not been characterized *in vitro*.

Here, we show that purified recombinant PvuRtsII selectively cleaves ^{hm}C-containing DNA and determine its cleavage site. In addition, we present initial data on the use of ^{hm}C as a tool to investigate ^{hm}C patterns in mammalian genomes.

MATERIALS AND METHODS

Cloning and purification of PvuRtsII

The sequence encoding PvuRtsII was synthesized at Mr Gene GmbH (Regensburg) and cloned into the pET28a vector (Novagen). BL21(DE3) *Escherichia coli* cells carrying the expression vector were grown in LB medium at 37°C until $A_{600} = 0.6-0.7$ and induced with 1 mM isopropyl β -D-thiogalactopyranoside for 16 h at 18°C. Lysates were prepared by sonication in 300 mM NaCl, 50 mM Na₂HPO₄ pH 8.0, 10 mM imidazole, 10% glycerol and 1 mM β -mercaptoethanol, cleared by centrifugation and applied to a nickel-nitrilotriacetic acid column (QIAGEN) pre-equilibrated with lysis buffer. Washing and elution were performed with lysis buffer containing 20 and 250 mM imidazole, respectively. Eluted proteins were applied to a Superdex S-200 preparative gel filtration column (GE Healthcare) in 150 mM NaCl, 20 mM Tris pH 8.0, 10% glycerol, 1 mM DTT and peak fractions were pooled. The stability of PvuRtsII upon storage was improved by supplementation with 10% glycerol.

Preparation of DNA substrates

In vivo α/β -glucosylated and non-glucosylated T4 phage DNA was isolated essentially as described (4). Briefly, T4 stocks were propagated on *E. coli* strain CR63, which was also used for the isolation of glucosylated T4 DNA. To isolate non-glucosylated T4 DNA, wild-type T4 phage was amplified on an ER1565 *galU* mutant strain. β -glucosylated T4 DNA was generated *in vitro* by treatment of non-glucosylated T4 DNA with purified T4 β -glucosyltransferase (7). Genomic DNA was isolated from mouse cerebellum and triple knockout (TKO) embryonic stem cells (ESCs) (21) as described (7).

Reference DNA fragments containing exclusively ^{hm}C, ^mC or unmodified cytosine residues were prepared by PCR using 5-hydroxymethyl-dCTP (Bioline GmbH), 5-methyl-dCTP (Jena Bioscience GmbH) and dCTP, respectively. T4 phage DNA template, Phusion HF DNA

Polymerase (Finnzymes) and primer 5'-GTG AAG TAA GTA ATA AAT GGA TTG-3', which does not contain cytosine residues, were used for amplification of all reference DNA fragments. To generate the reference 1139 bp fragment with 100% ^{hm}C for restriction with PvuRtsII the second primer was 5'-TGG AGA AGG AGA ATG AAG AAT AAT-3', which also does not contain cytosine residues. To generate the 800 and 500 bp control substrates containing only ^mC and only unmodified cytosine for restriction with PvuRtsII the second primer was 5'-GCC ATA TTG ATA ATG AAA TTA AAT GTA-3' and 5'-TCA GCA ATT TTA ATA TTT CCA TCT TC-3', respectively. PCR products were purified by gel electrophoresis followed by silica column purification (Nucleospin, Macherey-Nagel). The 140 bp fragment used to determine the orientation of the PvuRtsII cleavage overhang was amplified with primers 5'-TAT ACT GAA GTA CTT CAT CA-3' and 5'-CTT TGC GTG ATT TAT ATG TA-3'.

For the preparation of substrates with a single PvuRtsII consensus containing ^{hm}C or ^mC in symmetrical or asymmetrical configuration a 94 bp fragment was amplified from the T4 genome with primers 5'-CTC GTA GAC TGC GTA CCA ATC TAA CTC AGG ATA GTT GAT-3' and 5'-TAT GAT AAG TAT GTA GGT TAT T-3'. This fragment contains a single site corresponding to the identified PvuRtsII consensus ^{hm}CN₁₁₋₁₂/N₉₋₁₀G and was used as a template according to the strategy depicted in Figure 3. To generate substrates with symmetric cytosine modifications or unmodified cytosine the fragment was amplified with forward primer 5'-CTC GTA GAC TGC GTA CCA-3' and reverse primer 1 5'-TAT GAT AAG TAT GTA GGT TAT T-3' in the presence of the respective modified or unmodified dCTP. To generate substrates with asymmetric cytosine modifications the same forward primer was paired with reverse primer 2 5'-TAT GAT AAG TAT GTA GGT TAT TCA A-3'.

DNA restriction with PvuRtsII and identification of cleavage and recognition site

Unless otherwise stated the reaction conditions contained 150 mM NaCl, 20 mM Tris pH 8.0, 5 mM MgCl₂, 1 mM DTT. One unit of PvuRtsII was defined as amount of enzyme required to digest 1 μ g of ^{hm}C-containing T4 DNA in 15 min at 22°C. For assessment of enzyme specificity, 100 ng of each control fragment were digested separately or together with 200 ng of genomic DNA in 30 μ l reactions containing standard buffer and 1 U of purified PvuRtsII at 22°C for 15 min.

For identification of the cleavage and recognition site, the 1139 bp fully hydroxymethylated fragment amplified from the T4 genome or whole non-glucosylated T4 DNA were digested under standard conditions. Fragment ends were blunted with Klenow polymerase (NEB) and cloned using the Zero Blunt[®] PCR Cloning Kit (Invitrogen). Randomly selected clones were sequenced and the data were analyzed using WebLogo (22).

RESULTS

^{hm}C-specific endonuclease activity of PvuRtsII

His-tagged PvuRtsII was expressed in *E. coli* and purified to homogeneity by sequential Ni²⁺ affinity and size exclusion chromatography (Figure 1A). As bacteria carrying the RtsI plasmid were shown to restrict the ^{hm}C-containing T-even phages, but not ^mC-containing T-odd phages or λ phage, which does not contain modified cytosine (20), we initially used T4 genomic DNA as a substrate to test the activity of purified PvuRtsII. T4 genomic DNA was isolated from both *galU*⁺ and *galU*⁻ strains, the latter being UDP-glucose deficient and thus containing only non-glucosylated ^{hm}C. Under the same digestion conditions non-glucosylated T4 DNA was digested more efficiently than both naturally α- and β-glucosylated and *in vitro* β-glucosylated counterparts (Figure 1B). Non-glucosylated T4 DNA was cleaved into fragments with an apparent size of about 200 bp, indicating that PvuRtsII recognizes a frequently occurring sequence (Figure 1B and Supplementary Figures S1 and S2). We then used non-glucosylated T4 DNA to test the activity of the enzyme under various conditions. PvuRtsII was strictly dependant on Mg²⁺ ions, which could not be substituted with Ca²⁺, and endonuclease activity was maximal in the presence of 100–200 mM NaCl (Supplementary Figure S1A and B). However, during purification we observed that the enzyme is unstable in solutions of ionic strength lower than 150 mM NaCl. The activity of PvuRtsII was found highest at pH 7.5–8.0 and was unaffected by the presence of Tween 20 or Triton X-100 (Supplementary Figure S2A and B). We also observed that after prolonged incubation PvuRtsII precipitates even at room temperature, consistent with the reported temperature sensitivity of the phage restriction activity in cells carrying the RtsI plasmid (20). Upon short incubation times maximal activity was observed at 22°C (Supplementary Figure 2C). Thus, the relative amounts of enzyme and DNA substrate were standardized so that digestion was complete in 15 min at 22°C in the presence of 150 mM NaCl (Supplementary Figures S1C and S2C).

The specificity of PvuRtsII with respect to cytosine modification was further tested by digesting reference fragments containing exclusively unmodified cytosine

(500 bp), ^mC (800 bp) or ^{hm}C (1139 bp; Figure 1C). Under standard digestion conditions purified PvuRtsII selectively cleaved the ^{hm}C-containing fragment, consistent with the relative restriction efficiency of bacteriophages with distinct cytosine modifications by bacteria carrying the RtsI plasmid (20).

Determination of PvuRtsII cleavage sites

To identify the cleavage pattern of PvuRtsII we generated libraries of restriction fragments from either the whole T4 genome (Supplementary Figure S3) or an 1139 bp fragment amplified from the same genome containing exclusively hydroxymethylated cytosines (Figure 2). Random sequencing of 161 and 133 fragment ends from the whole T4 genome and 1139 bp fragment libraries revealed that 85 and 89%, respectively, matched the consensus sequence ^{hm}CN_{11–12}/N_{9–10}G. Among these 78 and 87%, respectively, showed one of three similar sequence patterns, ^{hm}CN₁₂/N₁₀G, ^{hm}CN₁₂/N₉G and ^{hm}CN₁₁/N₉G, while for the remaining fragment ends the exact number of nucleotides between the modified cytosine and the cleavage site could not be determined unambiguously due to the occurrence of multiple ^{hm}C residues upstream of the cleavage site. Of the sequenced fragment ends, 14 and 11% from the whole T4 genome and 1139 bp fragment libraries, respectively, did not match the ^{hm}CN_{11–12}/N_{9–10}G consensus. However, 100 and 80% of these ends, respectively, contained at least one ^{hm}C residue 10–13 nt upstream of the cleavage site, while no guanine was present in the T4 genomic sequence 10–11 nt downstream the cleavage site (Supplementary Figure S4). The sequenced clones from the 1139 bp T4 genomic fragment library corresponded to an 81% coverage of the fragment, with some PvuRtsII fragments occurring multiple times, while other fragments that were predicted on the basis of the ^{hm}CN_{11–12}/N_{9–10}G consensus were not found (Figure 2 and Supplementary Figure S5). Examination of the missing fragments did not show any common sequence feature beyond the ^{hm}CN_{11–12}/N_{9–10}G consensus (Supplementary Figure S6), suggesting that their absence from the sequenced fragments was due to limited sampling. Alignment of sequenced fragment ends from the T4 genomic fragment library showed that 2 nt around the cleavage site were missing from all clones, suggesting a 2 nt 3'-overhang cleavage pattern

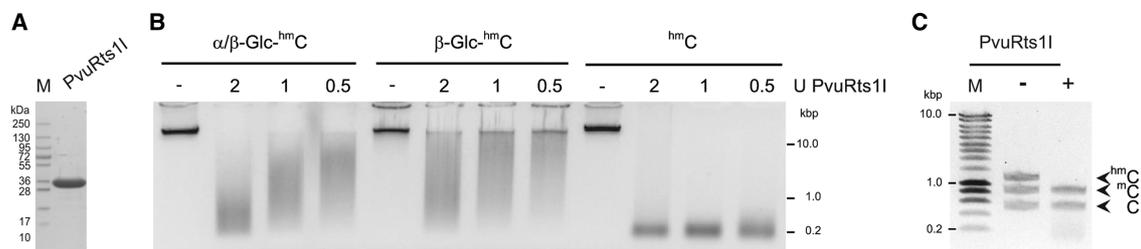


Figure 1. Selective restriction of ^{hm}C-containing DNA by PvuRtsII. (A) Purified PvuRtsII was resolved on a SDS-polyacrylamide gel and stained with coomassie blue. (B) T4 genomic DNA with the naturally occurring pattern of α- and β-glucosylated ^{hm}C, only β-glucosylated ^{hm}C or non-glucosylated ^{hm}C was incubated without or with decreasing amounts of PvuRtsII as indicated. (C) Reference PCR fragments of 1139, 800 and 500 bp containing ^{hm}C, ^mC and unmodified cytosine at all cytosine residues, respectively, were incubated with or without PvuRtsII as indicated.

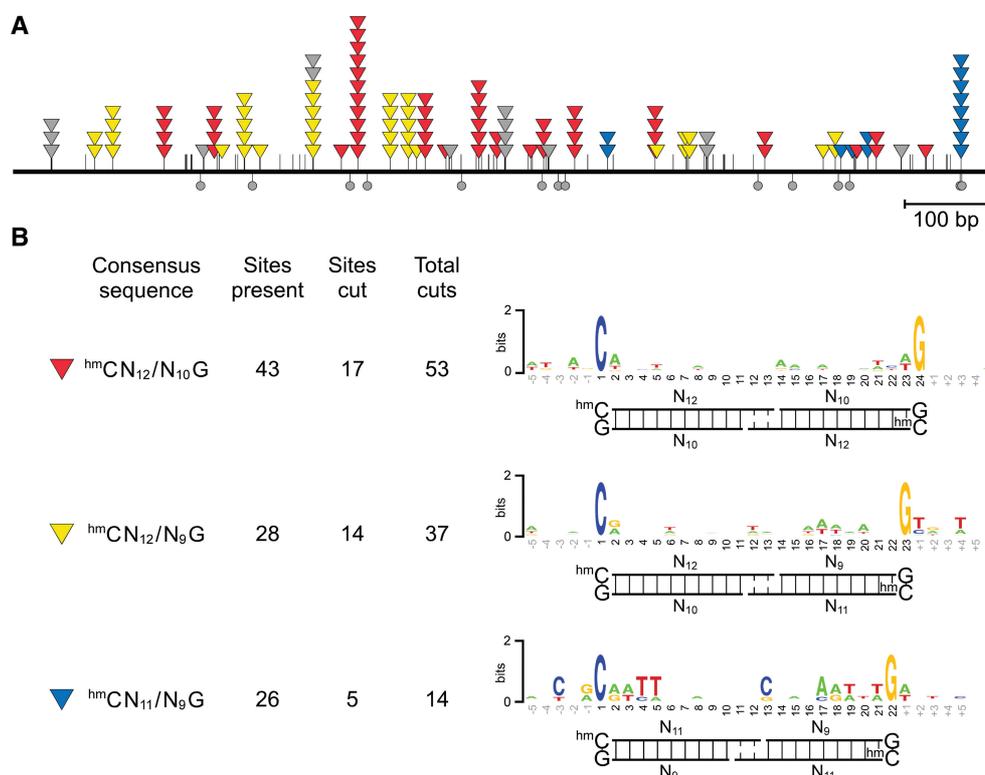


Figure 2. Cleavage site of PvuRtsII. A library of PvuRtsII restriction fragments was generated from an 1139 bp PCR fragment containing only hydroxymethylated cytosine residues and the sequence of 133 restriction fragment ends from randomly chosen clones was determined. (A) Graphical map of the fragment ends. A total of 119 analyzed fragment ends (triangles) matched the consensus sequence ^{hm}CN₁₁₋₁₂/N₉₋₁₀G, which was present at 97 sites (thin vertical lines) in the 1139 bp PCR fragment (thick horizontal line). Fifty three fragment ends related to the sequence motif ^{hm}CN₁₂/N₁₀G (dark green triangles), 37 to ^{hm}CN₁₁/N₁₀G (bright green triangles) and 14 to ^{hm}CN₁₁/N₉G (light green triangles), while 15 fragment ends matching the consensus sequence ^{hm}CN₁₁₋₁₂/N₉₋₁₀G could not be assigned unambiguously to any of these subsets (gray triangles). Fourteen fragment ends did not match the prevalent consensus sequence (gray circles, see Supplementary Figure S3). (B) Occurrence of the three subsets of cleavage sites and LOGO representation of the corresponding consensus sequence. The absolute height of each position reflects its overall conservation, while the relative height of nucleotide letters represents their relative frequency. The slash in the three cleavage sequence subtypes indicates the exact cleavage site.

(Supplementary Figure S5). This was confirmed by direct sequencing of the two fragments generated by digestion of a 140 bp amplicon containing a single PvuRtsII site (Supplementary Figure S7).

The results above reveal a symmetric nature of the preferred cleavage sites and raise the issue of PvuRtsII activity on sites with modified cytosine in symmetric and asymmetric configuration. To clarify this issue, we used a PCR strategy to generate DNA substrates with identical sequence and containing a single PvuRtsII consensus site with ^{hm}C or ^mC in symmetrical and asymmetrical configurations or no modified cytosine (Figure 3A). In the presence of enzyme amounts that did not cleave substrates with unmodified and ^mC sites, digestion of substrates with asymmetric ^{hm}C at the PvuRtsII site was reduced with respect to substrates with symmetric ^{hm}C, but still appreciable. Residual undigested substrate with symmetric ^{hm}C at the PvuRtsII site in these reaction conditions was typically observed with such short substrates, but not with longer ones.

Digestion of mammalian genomic DNA with PvuRtsII

To investigate cleavage site preference and efficiency of PvuRtsII digestion for mammalian genomic DNA, we initially selected the upstream regulatory region III of the mouse *nanog* gene (23). As this region was shown to be bound by Tet1 and to acquire CpG methylation upon knockdown of Tet1 in ESCs (5), it represents a potential candidate as a mammalian genomic sequence containing ^{hm}C. Real time amplification of this region from ESC genomic DNA did not show a significant decrease of product after PvuRtsII digestion (data not shown). We then devised a strategy to positively identify rare PvuRtsII digestion products. After PvuRtsII digestion genomic fragments were ligated to a linker with a random 2 nt 3'-overhang. Ligation products were then amplified using *nanog* specific primers paired with a linker specific primer, but no amplification product could be obtained (data not shown). This result may be explained by an extremely seldom occurrence of ^{hm}C at

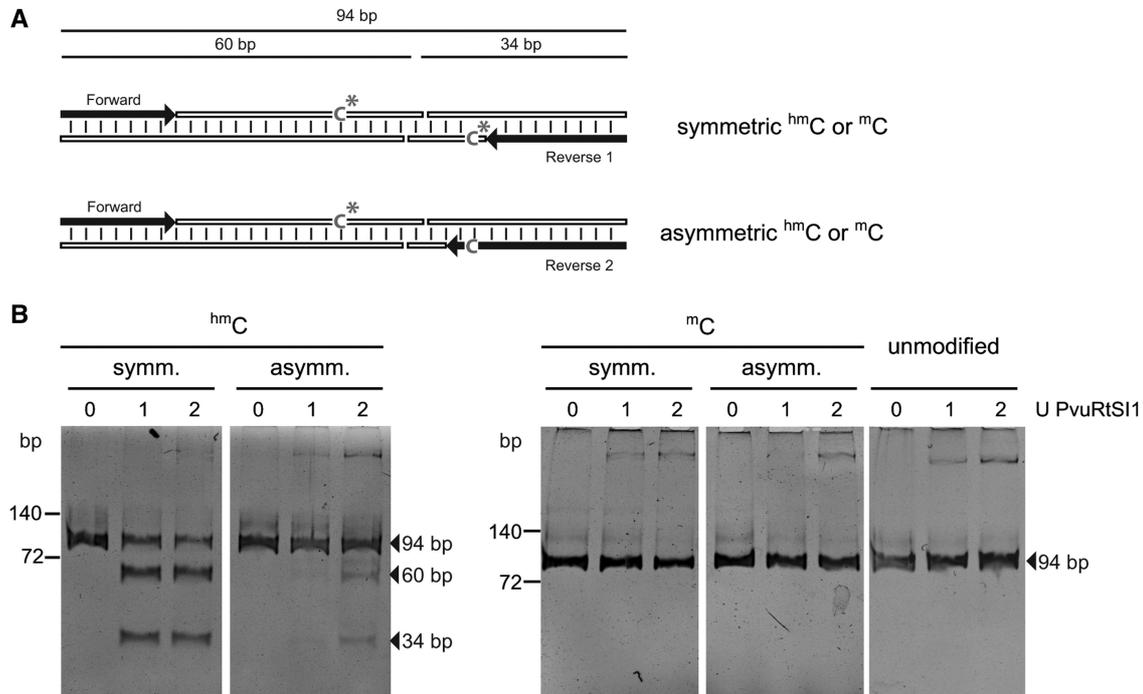


Figure 3. Differential activity of PvuRtsII on sites with symmetric and asymmetric ^{hm}C. Ninety-four bp long substrates with identical sequence were generated that contain a single PvuRtsII consensus site (CN₁₂/N₁₀G) with ^{hm}C or ^mC in symmetrical and asymmetrical configurations or no modified cytosine. (A) Strategy for generation of the substrates by PCR amplification in the presence of modified nucleotides. The size of the PvuRtsII digestion products is indicated. (B) The variously modified substrates were digested with the indicated amounts of PvuRtsII and digestion products were resolved on polyacrylamide gels. Note the reduced but tangible digestion of the substrate containing asymmetric ^{hm}C.

cleavage sites of this locus (especially in symmetric configuration), inefficiency of PvuRtsII digestion or both. In this regard, it is important to consider that positive identification of ^{hm}C sites in this region of the *nanog* locus has actually not been reported for ESCs. In addition, during the revision of the present work a manuscript was published (24) that could not confirm the reduced *nanog* expression and ESC differentiation previously reported upon Tet1 knockdown (5), raising uncertainty about the actual occurrence of ^{hm}C at the *nanog* promoter in ESCs.

As there are no clear and quantitative data on the levels and density of ^{hm}C at specific genomic sites available yet we generated defined substrates to validate the PvuRtsII cut-ligation amplification protocol for the identification of ^{hm}C sites. We PCR amplified region III of the *nanog* promoter in the presence of increasing concentrations of 5-hydroxymethyl-dCTP and confirmed the incorporation of proportional levels of ^{hm}C using the recently reported β-glucosylation assay (7) (data not shown). Fragment samples with increasing ^{hm}C content were then digested with PvuRtsII and the same ligation/PCR strategy for the identification of digestion products was applied as described above (Supplementary Figure S8A). Detection of fragments with ends corresponding to the PvuRtsII cleavage pattern raised with increasing ^{hm}C content.

We previously quantified global ^{hm}C levels in genomic DNA from ESCs and adult somatic tissues using *in vitro*

^{hm}C glucosylation (7). Consistent with other studies (3,6,8,9), this analysis revealed that genomic DNA from adult brain regions has a high ^{hm}C content. In addition, we showed that in ESCs that are TKO for all three major DNA methyltransferases Dnmt1, 3a and 3b (21) genomic ^{hm}C levels were around the estimated limit of detection, although reproducibly above background. Therefore, we compared the PvuRtsII restriction pattern of genomic DNA from cerebellum and TKO ESCs as representative of samples with high and very low ^{hm}C levels, respectively. As internal controls, we co-digested each of the two genomic DNA samples with the same reference fragments as used to test the specificity of PvuRtsII with respect to cytosine modification (Figure 1C). As expected from the relative low abundance of ^{hm}C in mammalian genomic DNA, there was a limited reduction of high molecular weight fragments and appearance of lower molecular weight smear (Figure 4). However, DNA from cerebellum was clearly digested to a higher extent than DNA from TKO ESCs as evident from the line scans across the respective gel lanes (Figure 4). The low but appreciable degree of digestion observed for genomic DNA from TKO ESCs does not seem to result from relaxed specificity or contaminating nuclease activities, as only control substrates containing ^{hm}C, but not ^mC or unmodified cytosine, were digested when incubated either separately or together with genomic DNA (Figure 1C and Figure 4).

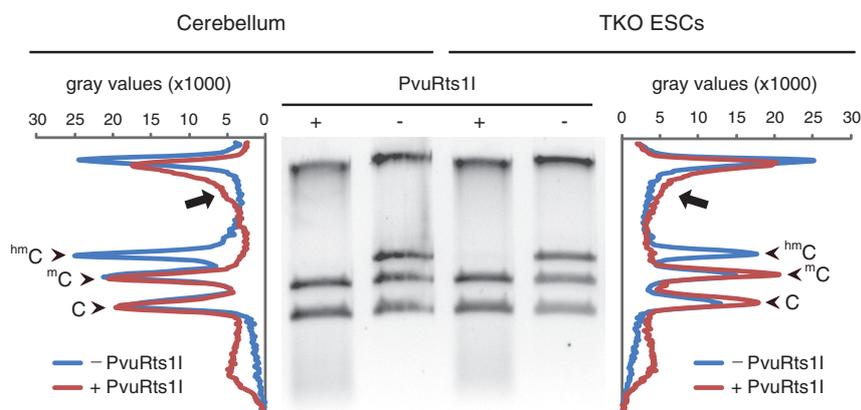


Figure 4. Restriction of mouse genomic DNA by PvuRts1I reflects ^{hm}C content. Genomic DNA from mouse cerebellum or TKO ESCs was mixed with three reference PCR fragments of 1139, 800 and 500 bp containing ^{hm}C , ^{m}C and unmodified cytosine at all cytosine residues, respectively, and incubated with or without PvuRts1I as indicated. Digests were resolved on a 0.8% agarose gel stained with ethidium bromide. Line scans of the gel lanes are aligned to the image of the gel. Red and blue lines correspond to samples incubated with and without enzyme, respectively. Arrows point to the main difference in the profiles from cerebellum and TKO ESC DNA digested with PvuRts1I (red lines).

Absence of digestion of control substrates containing ^{m}C and unmodified cytosine was evident from the unaltered ratio of their respective signals in the presence and absence of enzyme. This result shows that the extent of digestion by PvuRts1I reflects the relative ^{hm}C content in mammalian genomic DNA.

DISCUSSION

Several modification and restriction systems have evolved as defense and counter defense strategies in the struggle between unicellular microorganisms and their viruses. Here, we show that, in contrast to previously characterized endonucleases which cleave ^{hm}C -containing sequences, PvuRts1I has a preference for the non-glucosylated form of this base and discriminates against ^{m}C . This specificity makes PvuRts1I an attractive tool to investigate genomic ^{hm}C patterns in higher eukaryotes and complements the very recently published methods for enzymatic labeling of this sixth base (7,13).

Importantly, we show that the extent of PvuRts1I digestion reflects the relative abundance of ^{hm}C in genomic DNA from cerebellum and TKO ESCs. The limited extent of digestion even for samples with relatively high ^{hm}C content is in line with the cleavage site preference and dependence on cytosine modification that we determined. We calculate that the statistical probability of the PvuRts1I consensus site $CN_{11-12}/N_{9-10}G$ in the mouse genome is 0.126. Combined with the global ^{hm}C occurrence in mouse tissues (up to 0.13% of all bases or 0.65% of Cs) (3,7–9) this translates into a PvuRts1I cleavage site every 1.9×10^5 bases. As this is in the size range of fragments typically obtained with standard procedures for isolation of genomic DNA, more careful isolation methods should be used and/or PvuRts1I specific ends could be enriched by ligating biotinylated PvuRts1I compatible linkers. Alternatively, digestion conditions could be optimized or DNA could be denatured and a

second strand synthesized with ^{hm}C nucleotides to cut and reveal the likely more abundant hemimodified PvuRts1I sites.

Notably, while cerebellum has been previously reported among the tissues with the highest levels of genomic ^{hm}C (3,7,9), complete absence of ^{m}C and therefore ^{hm}C would be expected in TKO ESCs due to the lack of all three major Dnmts (21). However, we previously detected ^{hm}C levels slightly above background in TKO ESCs (7) and here we show minimal but appreciable digestion by PvuRts1I. In this context, it is interesting to note that ESCs express the highly conserved Dnmt2 (25,26), the only Dnmt family member with an intact catalytic domain that has not been genetically inactivated in TKO ESCs. Although Dnmt2 has a major role as a tRNA methyltransferase and its function as a DNA methyltransferase is still debated (27–32), it was recently shown to methylate genomic sequences in *Drosophila* (32,33). Future work should clarify whether the genome of TKO ESCs harbors any residual ^{m}C and ^{hm}C .

Restriction of genomic DNA with PvuRts1I may be combined with PCR amplification for analysis of specific loci or with massive parallel sequencing or microarray hybridization for genome-wide mapping. The calculations reported above for the frequency of PvuRts1I cleavage sites based on a random ^{hm}C distribution bring up the argument that the extent of random breaks in genomic DNA preparations would contribute very significant noise in deep sequencing and microarray applications. This drawback may at least be partially overcome if specific PvuRts1I ends are enriched by ligating linkers with a random 2 nt 3'-overhang as described here and discussed above, a strategy that can be integrated with procedures for generation of sequencing libraries. Also, our simulation of genomic fragments containing known levels of randomly distributed ^{hm}C clearly shows that relatively high local concentrations of ^{hm}C sites are required for efficient detection by PvuRts1I. The first genome-wide ^{hm}C

profiles from mammalian tissues have just been reported (13). From these first data sets, it is apparent that genomic ^{hm}C is not randomly distributed and that its accumulation in gene bodies is proportional to transcriptional activity. Thus, PvuRtsII may prove a valuable tool to probe ^{hm}C accumulation at defined genomic regions. In addition, the selectivity of PvuRtsII for ^{hm}C-containing sites may constitute an advantage with respect to endonucleases such as McrBC and MspJI as these enzymes do not discriminate between ^mC and ^{hm}C and require *in vitro* enzymatic ^{hm}C glucosylation to specifically protect ^{hm}C-containing sites from digestion and thus distinguish them from ^mC sites.

In conclusion, we show that PvuRtsII is an ^{hm}C specific endonuclease and provide a biochemical characterization of its enzymatic properties for future applications as diagnostic tool in the analysis of ^{hm}C distribution at genomic loci in development and disease.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Ken Kreutzer (Duke University, Durham, NC) for providing the T4 phage stock and *E. coli* strains CR63 and ER1565 *galU* and Masaki Okano (RIKEN Center for Developmental Biology, Kobe, Japan) for the TKO ESCs.

FUNDING

Deutsche Forschungsgemeinschaft (DFG SPP 1356, SFB 646 and 684 to H.L.); Life Science Munich (LSM to S.B.). Financial support by the Elite Network of Bavaria (International Doctorate Program NanoBioTechnology) and International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS to C.S.S.). Funding for open access charge: Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Rottach, A., Leonhardt, H. and Spada, F. (2009) DNA methylation-mediated epigenetic control. *J. Cell. Biochem.*, **108**, 43–51.
- Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL Partner TET1. *Science*, **324**, 930–935.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. and Zhang, Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
- Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J. and Fan, G. (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.*, **13**, 423–430.
- 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.
- Münzel, M., Globisch, D., Bruckl, T., Wagner, M., Welzmler, 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.
- Globisch, D., Münzel, M., Müller, 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.
- Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R. and Rao, A. (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE*, **5**, e8888.
- Jin, S.-G., Kadam, S. and Pfeifer, G.P. (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **38**, e125.
- Nestor, C., Ruzov, A., Meehan, R. and Dunican, D. (2010) Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *Biotechniques*, **48**, 317–319.
- Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X. *et al.* (2010) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.*, **29**, 68–72.
- Flaks, J.G. and Cohen, S.S. (1957) The enzymic synthesis of 5-hydroxymethyldeoxycytidylic acid. *Biochim. Biophys. Acta*, **25**, 667–668.
- Wiberg, J.S. and Buchanan, J.M. (1964) Studies on labile deoxycytidylate hydroxymethylases from *Escherichia coli* B infected with temperature-sensitive mutants of bacteriophage T4. *Proc. Natl Acad. Sci. USA*, **51**, 421–428.
- Lehman, I.R. and Pratt, E.A. (1960) On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4 and T6. *J. Biol. Chem.*, **235**, 3254–3259.
- Raleigh, E.A. (1992) Organization and function of the mcrBC genes of *Escherichia coli* K-12. *Mol. Microbiol.*, **6**, 1079–1086.
- Zheng, Y., Cohen-Karni, D., Xu, D., Chin, H.G., Wilson, G., Pradhan, S. and Roberts, R.J. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res.*, **38**, 5527–5534.
- Bair, C.L. and Black, L.W. (2007) A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J. Mol. Biol.*, **366**, 768–778.
- Janosi, L., Yonemitsu, H., Hong, H. and Kaji, A. (1994) Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRtsII) modulated by glucosylation of DNA. *J. Mol. Biol.*, **242**, 45–61.
- Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R. *et al.* (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells*, **11**, 805–814.
- Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
- Hattori, N., Imao, Y., Nishino, K., Hattori, N., Ohgane, J., Yagi, S., Tanaka, S. and Shiota, K. (2007) Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes Cells*, **12**, 387–396.
- 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.
- Okano, M., Xie, S. and Li, E. (1998) Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.*, **26**, 2536–2540.

26. Yoder, J.A. and Bestor, T.H. (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.*, **7**, 279–284.
27. Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E. and Bestor, T.H. (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science*, **311**, 395–398.
28. Rai, K., Chidester, S., Zavala, C.V., Manos, E.J., James, S.R., Karpf, A.R., Jones, D.A. and Cairns, B.R. (2007) Dnmt2 functions in the cytoplasm to promote liver, brain and retina development in zebrafish. *Genes Dev.*, **21**, 261–266.
29. Schaefer, M. and Lyko, F. (2010) Lack of evidence for DNA methylation of Invader4 retroelements in *Drosophila* and implications for Dnmt2-mediated epigenetic regulation. *Nat. Genet.*, **42**, 920–921.
30. Schaefer, M. and Lyko, F. (2010) Solving the Dnmt2 enigma. *Chromosoma*, **119**, 35–40.
31. Zemach, A., McDaniel, I.E., Silva, P. and Zilberman, D. (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, **328**, 916–919.
32. Gou, D., Rubalcava, M., Sauer, S., Mora-Bermúdez, F., Erdjument-Bromage, H., Tempst, P., Kremmer, E. and Sauer, F. (2010) SETDB1 is involved in postembryonic DNA methylation and gene silencing in *Drosophila*. *PLoS ONE*, **5**, e10581.
33. Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M.C. and Reuter, G. (2009) Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.*, **41**, 696–702.

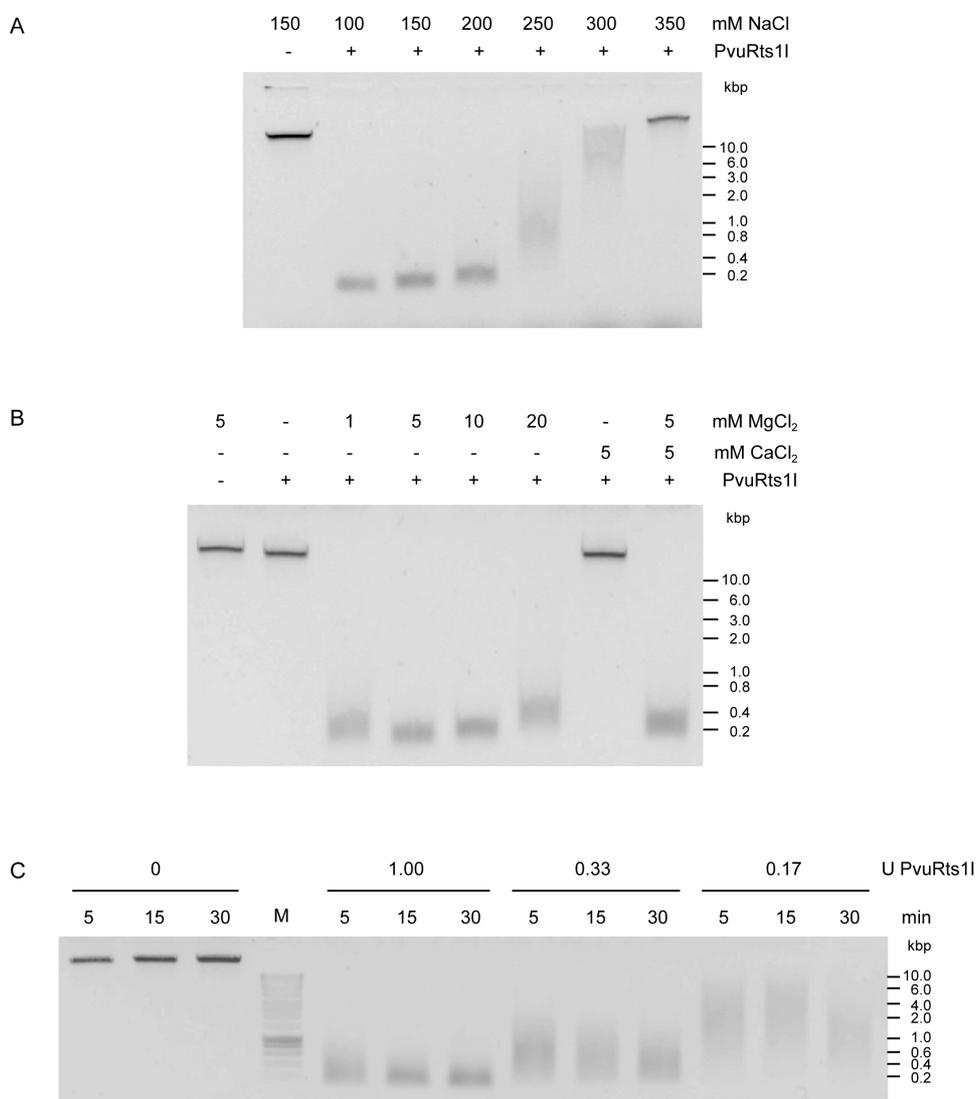
**Characterization of PvuRtsII endonuclease as a tool to investigate genomic
5-hydroxymethylcytosine**

Aleksandra Szwagierczak, Andreas Brachmann, Christine S. Schmidt, Sebastian Bultmann,
Heinrich Leonhardt and Fabio Spada

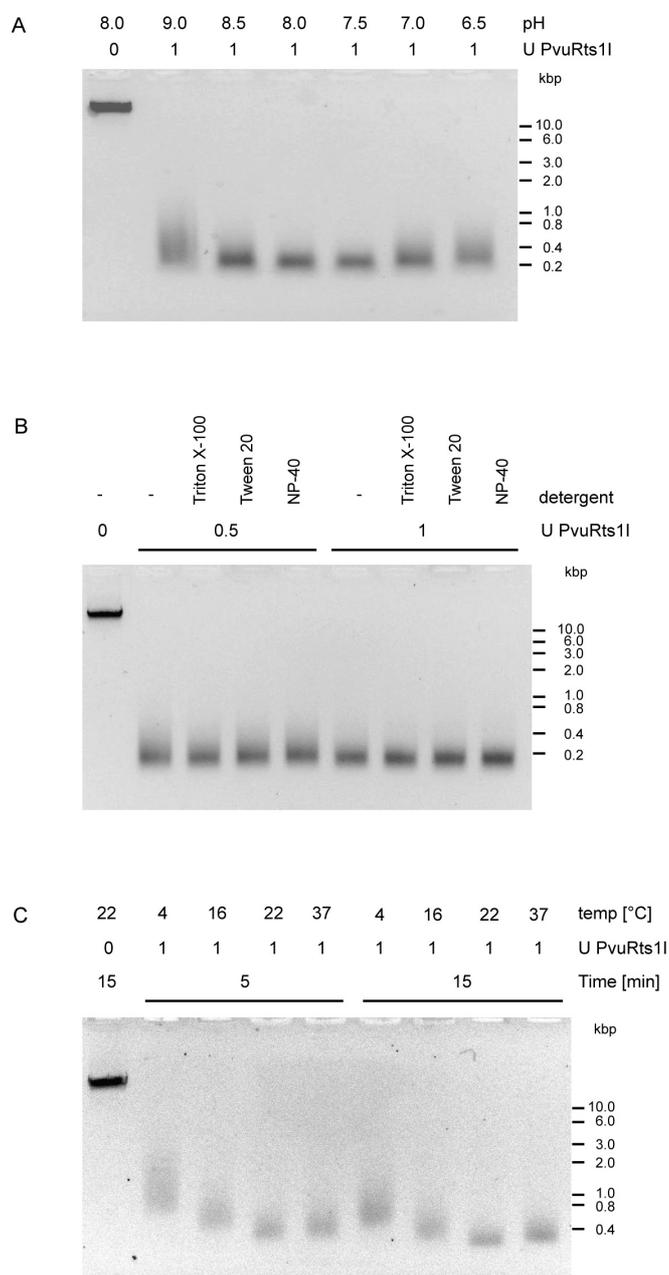
Ludwig Maximilians University Munich, Department of Biology and Center for Integrated
Protein Science Munich (CIPS^M), 82152 Planegg-Martinsried, Germany.

SUPPLEMENTARY INFORMATION

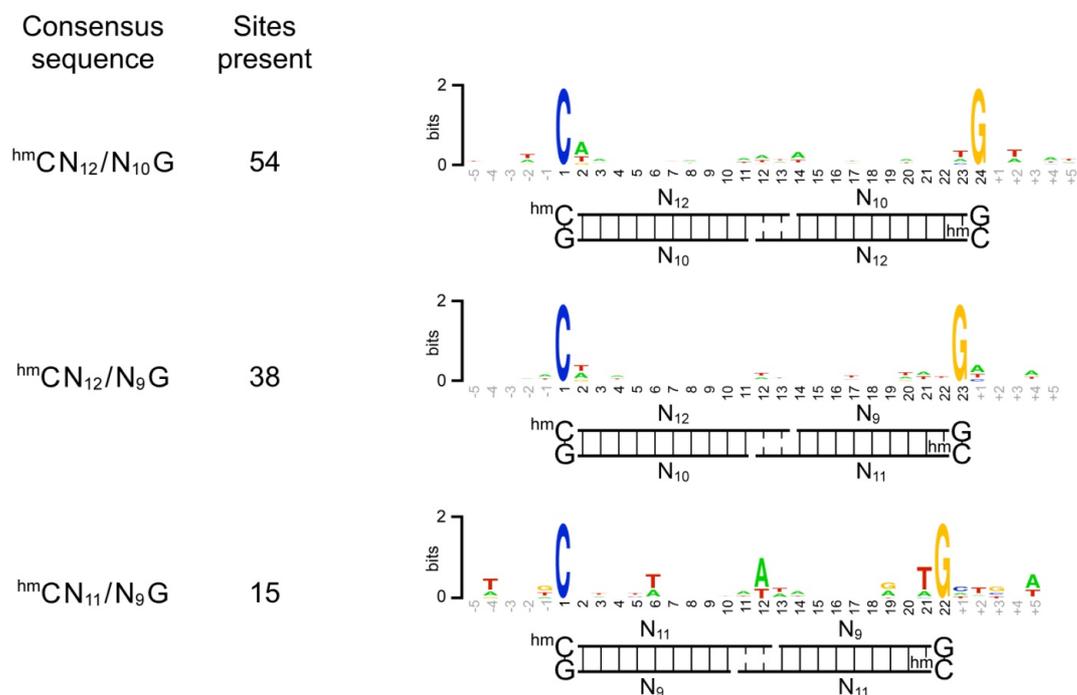
Supplementary Figures S1-8
Supplementary Methods



Supplementary Figure S1. Optimization of PvuRts1I restriction conditions using non-glucosylated T4 genomic DNA as substrate. **(A-B)** Comparison of cleavage rates in the presence different ionic strength conditions and types and concentrations of bivalent ions. One μg of DNA was digested with 1U of enzyme in buffer containing 20 mM Tris pH 8.0 and **(A)** 5 mM MgCl₂ and the indicated concentrations of NaCl or **(B)** 150 mM NaCl and the indicated concentrations of MgCl₂ or CaCl₂. **(C)** Combined time course and enzyme titration in buffer containing 20 mM Tris pH 8.0, 150 mM NaCl and 5 mM MgCl₂.



Supplementary Figure S2. Characterization of PvuRts1I activity under different pH (A), detergent conditions (B) and temperature (C). Non-glucosylated T4 genomic DNA was used as substrate. In A and C incubation was for 15 min at 22°C.



Supplementary Figure S3. Cleavage site of PvuRts1I as deduced from a restriction fragment library from the whole non-glucosylated T4 genome. A total of 161 fragment ends were sequenced. 137 fragment ends matched the consensus sequence ${}^{\text{hm}}\text{CN}_{11-12}/\text{N}_{9-10}\text{G}$, of which 54 related to the sequence motif ${}^{\text{hm}}\text{CN}_{12}/\text{N}_{10}\text{G}$, 38 to ${}^{\text{hm}}\text{CN}_{11}/\text{N}_{10}\text{G}$, 15 to ${}^{\text{hm}}\text{CN}_{11}/\text{N}_9\text{G}$, while 30 could not be assigned unambiguously to any of these subsets due to the occurrence of multiple ${}^{\text{hm}}\text{C}$ residues upstream of the cleavage site. 24 fragment ends had at least one ${}^{\text{hm}}\text{C}$ residue at a distance 10-13 nucleotides from the cutting site, but no guanine was present in the T4 genomic sequence 10-11 nucleotides downstream the cleavage site. Shown is the occurrence (left) and LOGO graphic representation (right) of the three consensus sequence subtypes. In the graphic representations the absolute height of each position and the relative height of each nucleotide letter reflect overall conservation and relative nucleotide frequency, respectively (Crooks *et al.*, 2004).

Restriction site ^{hm}CN₁₃↓ (5 fragment ends)

```

AATTGCGGCTGATTTTAAT CGTTAATTGCTTTATT
AGATACTACTGCTCTGCTT AAAAAATTCCCTACTA
ACCTTCAAATTTAGCAGCA CCTTGTTACCTTTTG
AGATTCTGCTCTGACCCGT GTTAAATATTCTTTGA
TGTTGCGGAAACAATTGCG GCTGATTTTAATCGTT
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 ↓ +1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11 +12 +13 +14 +15 +16

```

Restriction site ^{hm}CN₁₂↓ (5 fragment ends)

```

GCCCTCTATTAAAAATCG TGGGTAGAATCAGCTT
GCTGGCCAAAAAATTGTT GAGCGGGCATCAGCAA
AGTTTCATGTAATTTCCCT TCAATTCAAATGAGAT
TAATGCGCAAACATACTA CAGAATCCCTAAACGC
CTGTACTCGGATCGGCTG CTGGCCAAAAAATTGT
  1  2  3  4  5  6  7  8  9 10 11 12 13 ↓ +1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11 +12 +13 +14 +15 +16

```

Restriction site ^{hm}CN₁₁↓ (1 fragment end)

```

AGATACACGCAACAGCT GTTGAAGGTTCTTCAGC
  1  2  3  4  5  6  7  8  9 10 11 12 ↓ +1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11 +12 +13 +14 +15 +16

```

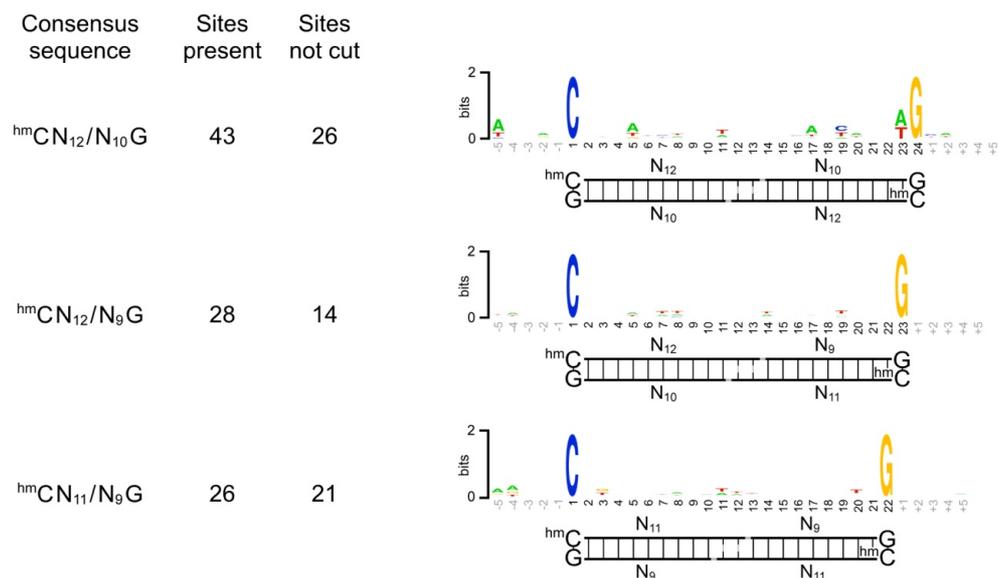
No consensus sequence (3 fragment ends)

```

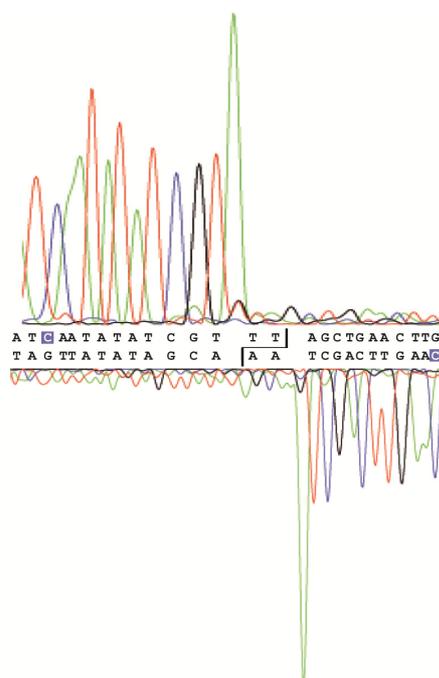
ACTAATGCGCAAACATAC TACAGAATCCCTAAAC
TGTTAAATATTCTTTGAC TCTTGGTGATTATGAT
AATATTTAAAATTGCTGAT GCCCGCTCAACAATTT
  1  2  3  4  5  6  7  8  9 10 11 12 13 ↓ +1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11 +12 +13 +14 +15 +16

```

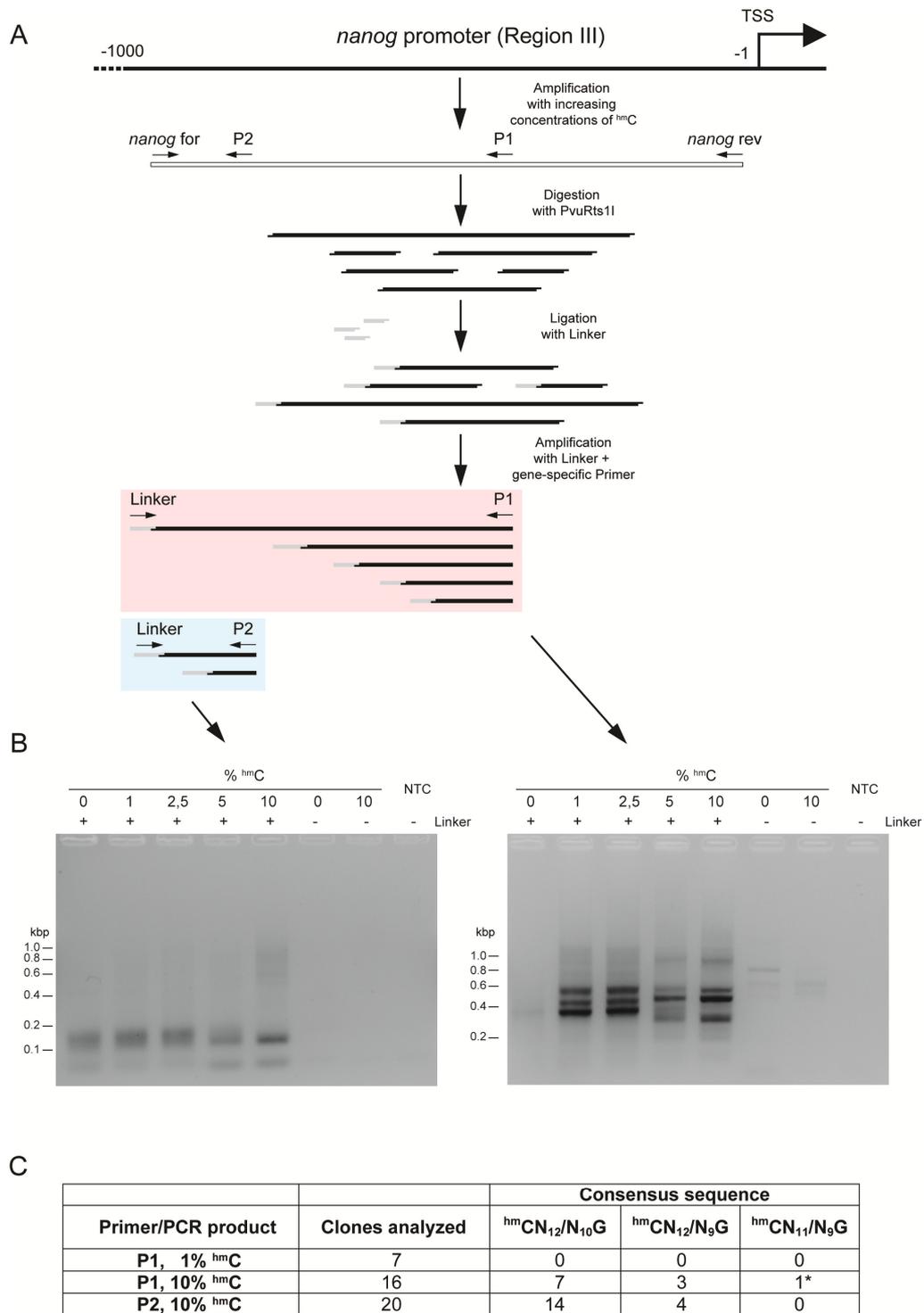
Supplementary Figure S4. Sequences from the T4 genomic 1139 bp fragment cut by PvuRtsII that deviate from the predicted consensus sequence ^{hm}C N₁₁₋₁₂/N₉₋₁₀G. All cytosine residues are hydroxymethylated but for simplicity they are here indicated as Cs. ^{hm}C and guanine residues 11-13 nucleotides upstream of and 9-10 nucleotides downstream to the cleavage site, respectively, are highlighted in red. Residues 21-23 nucleotides downstream of a ^{hm}C are shaded in light red.



Supplemental Figure S6. Analysis of sequences from the T4 genomic 1139 bp fragment matching the PvuRtsII consensus cleavage site ${}^{\text{hm}}\text{CN}_{11-12}/\text{N}_{9-10}\text{G}$ that were not found among the sequenced fragments. In the LOGO graphic representations on the right the absolute height of each position and the relative height of each nucleotide letter reflect overall conservation and relative nucleotide frequency, respectively (Crooks *et al.*, 2004).



Supplementary Figure S7. Confirmation of a two nucleotide 3' overhang cleavage pattern by PvuRts1I. A 140 bp fragment containing only hydroxymethylated cytosine residues and a single PvuRts1I site was amplified from the T4 genome and digested with PvuRts1I. The two ensuing PvuRts1I restriction fragments were directly sequenced from their respective 5' ends employing the same primers used for amplifying the original 140 bp fragment. Alignment of the two sequence tracks to the original sequence revealed a two nucleotide gap consistent with a 3' overhang configuration of these nucleotides at PvuRts1I ends. Only the ends of the sequence tracks corresponding to the PvuRts1I site are shown. The appropriately spaced ^{hm}C residues on either side of the cleavage site and opposite strands that constitute the PvuRts1I site are highlighted. The large adenine peaks (green) present at the end of each sequence track but not in the original sequence are due to addition of a 3' overhanging adenine residue by the DNA polymerase used for the sequencing reaction.



Supplementary Figure S8. Identification of PvuRts1I fragments from substrates with increasing ^{hm}C content. **(A)** The proximal upstream regulatory region of the *nanog* locus (region III) was amplified in the presence of increasing concentrations of 5-hydroxymethyl-dCTP, yielding fragments with randomly distributed ^{hm}C sites in the respective proportions (not shown). These fragments were digested with PvuRts1I and ligated to linkers with random two nucleotide overhangs to match PvuRts1I ends. Ligation products were amplified with two distinct *nanog* specific primers (*nanog* P1 and P2) each paired with a linker specific primer. The PCR products obtained are shown in **(B)**. The percentage of ^{hm}C in the original substrate fragments and the presence of the linker in the ligation reaction are indicated. NTC: no template control. **(C)** Products from PCR reactions shown in **(B)** were randomly cloned and sequenced. The numbers of sequences containing ends corresponding to the PvuRts1I consensus and site subtype are reported. The asterisk demarks a sequence that could not be univocally assigned to ^{hm}CN₁₂/N₉G or ^{hm}CN₁₁/N₉G due to the presence of consecutive C residues and is reported under both categories. In the case of substrates containing 10% ^{hm}C both primer sets yielded fragments with specific PvuRts1I digestion products that mapped to several predicted cleavage sites (not shown). We note that 1% ^{hm}C is in the same range as measured only in mouse tissues with the highest global ^{hm}C content (3,4,6-9,23). It follows that high local concentrations of ^{hm}C sites facilitate detection by PvuRts1I with this procedure.

SUPPLEMENTARY METHODS

Generation of fragments from the nanog upstream region III containing known levels of^{hmC}, PvuRtsII digestion and identification of digestion products.

Genomic DNA from JM8A3.N1 ESCs (EUComm, Helmholtz Center Munich, Neuherberg, Germany) was isolated using the NucleoSpin Triprep Kit (Macherey-Nagel). To prepare substrates containing different hmC levels (0%, 1%, 2,5%, 5%, 10%), genomic DNA from JM8A3.N1 cells was used as a template to amplify a 867 bp fragment from region III of the *nanog* promoter (Hattori et al, Genes to cell, 2007) using corresponding ratios of 5-hydroxymethyl-dCTP (Bioline GmbH) and dCTP, Phusion HF DNA Polymerase (Finnzymes) and the following primers: *nanog* for 5'-TCA GGA GTT TGG GAC CAG CTA-3' and *nanog* rev 5'-CCC CCC TCA AGC CTC CTA-3'. After purification of the PCR fragments using the NucleoSpin Extract II kit (Macherey-Nagel), 250ng of each fragment was digested with 2U of PvuRtsII for 15min at 22°C and the enzyme was heat inactivated at 60°C for 20 min. Twentyfive nanograms of digested fragment were ligated to a linker containing random two nucleotide 3' overhangs, generated by annealing the following primers: For 5'-CTC GTA GAC TGC GTA CCA TG NN-3' and Rev 5'-CA TGG TAC GCA GTC TAC CAG-3'. The ligation reaction was carried out using T4 DNA Ligase (NEB) overnight at 16°C. As a control for ligation specificity, each fragment was ligated in the absence of the linker. To selectively amplify fragments cut by PvuRtsII, the ligated products were amplified by PCR with Phusion HF DNA Polymerase (Finnzymes) using a linker specific forward primer (For 5'-CTC GTA GAC TGC GTA CCA TG-3') and *nanog* specific reverse primers (P2: 5'-GAG TCA GAC CTT GCT GCC AAA-3' and P1: 5'-GCC GTC TAA GCA ATG GAA GAA-3'). Libraries of digested and ligated fragments containing 1 and 10% hmC were generated using the Zero Blunt® PCR Cloning Kit (Invitrogen). Randomly selected clones were sequenced and analyzed for the presence of PvuRtsII ends.

2.3 Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1

Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1

Carina Frauer¹[§], Andrea Rottach¹[§], Daniela Meilinger¹, Sebastian Bultmann¹, Karin Fellinger¹[¶], Stefan Hasenöder¹[¶], Mengxi Wang¹, Weihua Qin¹, Johannes Söding², Fabio Spada¹^{*}, Heinrich Leonhardt¹^{*}

¹ Department of Biology II and Center for Integrated Protein Science Munich (CIPS^M), Ludwig Maximilians University Munich, Planegg, Germany, ² Gene Center Munich, Ludwig Maximilians University Munich, Munich, Germany.

Abstract

Several mammalian proteins involved in chromatin and DNA modification contain CXXC zinc finger domains. We compared the structure and function of the CXXC domains in the DNA methyltransferase Dnmt1 and the methylcytosine dioxygenase Tet1. Sequence alignment showed that both CXXC domains have a very similar framework but differ in the central tip region. Based on the known structure of a similar MLL1 domain we developed homology models and designed expression constructs for the isolated CXXC domains of Dnmt1 and Tet1 accordingly. We show that the CXXC domain of Tet1 has no DNA binding activity and is dispensable for catalytic activity *in vivo*. In contrast, the CXXC domain of Dnmt1 selectively binds DNA substrates containing unmethylated CpG sites. Surprisingly, a Dnmt1 mutant construct lacking the CXXC domain formed covalent complexes with cytosine bases both *in vitro* and *in vivo* and rescued DNA methylation patterns in *dnmt1*^{-/-} embryonic stem cells (ESCs) just as efficiently as wild type Dnmt1. Interestingly, neither wild type nor Δ CXXC Dnmt1 re-methylated imprinted CpG sites of the *H19a* promoter in *dnmt1*^{-/-} ESCs, arguing against a role of the CXXC domain in restraining Dnmt1 methyltransferase activity on unmethylated CpG sites.

Citation: Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, et al. (2011) Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1. PLoS ONE 6(2): e16627. doi:10.1371/journal.pone.0016627

Editor: Anton Wutz, Wellcome Trust Centre for Stem Cell Research, United Kingdom

Received: September 14, 2010; **Accepted:** January 5, 2011; **Published:** February 2, 2011

Copyright: © 2011 Frauer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB646 and TR5; <http://www.dfg.de/en/index.jsp>), the Nanosystem Initiative Munich (NIM; <http://www.nano-initiative-munich.de/>) and the Center for NanoScience (CeNS; <http://www.cens.de/>) to HL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: f.spada@lmu.de (FS); h.leonhardt@lmu.de (HL)

§ These authors contributed equally to this work.

¶ Current address: Intervet International GmbH, Unterschleißheim, Germany

¶ Current address: Institute of Stem Cell Research, Helmholtz Zentrum München, Neuherberg, Germany

Introduction

In mammals DNA methylation is restricted to cytosine residues and mainly involves CpG dinucleotides. CpG methylation is widespread across mammalian genomes, including gene bodies regardless of their transcriptional activity [1–4]. However, highly CpG-rich regions (CpG islands) are refractory to methylation and mostly coincide with promoters of constitutively active genes. The methylation state of other regulatory sequences with moderate to low CpG density, including promoters and enhancers, shows developmental and/or tissue-specific variations and positively correlates with a transcriptionally silent state [1,3–8]. Dense methylation of repetitive sequences is also thought to maintain these elements in a silent state and thus contribute to genome stability [9–11]. In mammals cytosine methylation is catalyzed by a family of DNA methyltransferases (Dnmts) [12]. Dnmt3a and Dnmt3b establish methylation patterns during embryonic development of somatic as well as germ cell lineages and, consistently, show developmental stage and tissue specific expression patterns. In contrast, Dnmt1 is ubiquitous and generally the most abundant DNA methyltransferase in mammalian tissues, where it associates with the replication machinery and restores symmetrical methylation at hemimethylated CpG sites generated by the semi-

conservative DNA replication process [13]. Thus, Dnmt1 maintains methylation patterns with high fidelity and is essential for embryonic development and genome integrity [9,14,15].

Dnmt1 is a large enzyme with a complex domain structure that likely evolved by fusion of at least three genes [16]. It comprises a regulatory N-terminal region and a C-terminal catalytic domain connected by a linker of seven glycine-lysine repeats (Figure 1A) [17]. The N-terminal part contains a PCNA binding domain (PBD), a heterochromatin targeting sequence (TS), a CXXC-type zinc finger domain and two Bromo-Adjacent Homology domains (BAH1 and BAH2). The C-terminal domains of mammalian Dnmts contain all ten catalytic motifs identified in bacterial DNA (cytosine-5) methyltransferases [12]. Thus, prokaryotic and mammalian cytosine methyltransferases are thought to adopt the same catalytic mechanism. However, the C-terminal domain of Dnmt1 is the only DNA methyltransferase domain in Dnmts that is not catalytically active when expressed separately. Indeed, interaction with the N-terminal part is required for allosteric activation of the enzyme [18]. Remarkably, the first 580 amino acids (aa) of human DNMT1 are dispensable for both enzymatic activity and substrate recognition, whereas deletion of the first 672 aa results in an inactive enzyme [19]. Interestingly, this truncation eliminates part of the CXXC domain, suggesting an involvement of this domain in allosteric activation. However, addition

Figure 1. Sequence and predicted structural homology of CXXC domains. (A) Schematic representation of the domain structure in Dnmt1 and Tet1. The catalytic domain and the N-terminal region of Dnmt1 are connected by seven lysine-glycine repeats [(KG)₇]. PBD: PCNA binding domain; TS: targeting sequence; CXXC: CXXC-type zinc finger domain; BAH1 and 2: bromo-adjacent homology domain; NLS: nuclear localization signal; Cys-rich: cysteine rich region. (B) Alignment of mammalian CXXC domains. Numbers on the right side indicate the position of the last amino acid in the corresponding protein. The Mbd1a isoform contains three CXXC motifs (Mbd1_1-3). Absolutely conserved residues, including the eight cysteines involved in zinc ion coordination are highlighted in red and the conserved KFGG motif is in red bold face. Positions with residues in red face share 70% similarity as calculated with the Risler algorithm [66]. At the top residues of MLL1 involved in β sheets β 1 and β 2 (black arrows), α helices α 1 and α 2 and strict α turns (TTT) are indicated. All sequences are from *M. musculus*. Accession numbers (for GenBank unless otherwise stated): Dnmt1, NP_034196; Mll1, NP_001074518; Mll4, O08550 (SwissProt); CGBP, NP_083144; Kdm2a, NP_001001984; Kdm2b, NP_001003953; Fbxl19, NP_766336; Mbd1, NP_038622; CXXC4/Idax, NP_001004367; CXXC5, NP_598448; CXXC10 (see Materials and Methods). (C) A homology tree was generated from the alignment in (B). The three subgroups of CXXC domains identified are in different colors. Average distances between the sequences are indicated. (D–E) Homology models of the mouse Dnmt1 (D; red) and Tet1 (E; blue) CXXC domains superimposed to the CXXC domain of MLL1 (green; [35]). MLL1 residues that were described to contact DNA according to chemical shift measurements [35] are cyan in (E), while cysteines involved in coordination of the two zinc ions are yellow. Arrows point to the KFGG motif in MLL1 and Dnmt1. The locations of α helices and β sheets are indicated as in (B). doi:10.1371/journal.pone.0016627.g001

of an N-terminal fragment containing the isolated CXXC domain to the catalytic domain was not sufficient for catalytic activation [20].

CXXC-type zinc finger domains are found in several other proteins with functions related to DNA or chromatin modification, including the histone H3 lysine 4 (H3K4) methyltransferases mixed-lineage leukaemia (MLL) proteins 1 and 4, the CpG-binding protein (CGBP, also known as Cfp1 or CXXC1), the methyl-CpG binding domain protein 1 (MBD1), the H3 lysine 36 (H3K36) demethylases KDM2A and B (also known as JHD1A/FBXL11 and JHD1B/FBXL10) and the MLL1 fusion partner TET1 (Figure 1A) [21–28]. The CXXC domains of some of these proteins were shown to mediate specific binding to double stranded DNA templates containing unmethylated CpG sites [21,22,29,30]. A region of Dnmt1 which mainly includes the CXXC domain (aa 628–753) was also shown to bind Zn ions and DNA [20,31,32]. However, available data on the selectivity of this DNA binding activity are conflicting. Whereas a fragment including aa 613–748 of mouse Dnmt1 was shown to bind DNA with a slight preference for hemimethylated CpG sites [20], aa 645–737 of human DNMT1 were shown to selectively bind unmethylated DNA [32]. As these studies used different constructs and species, the selectivity of DNA binding by the CXXC domain of Dnmt1 with regard to CpG methylation state and the role of the CXXC domain in allosteric activation and substrate discrimination remain to be firmly established.

Notably, not all CXXC domains show DNA binding activity, as exemplified by the fact that only one out of three CXXC domains in MBD1 binds DNA [29]. Interestingly, TET1 was recently shown to be a 2 oxoglutarate- and Fe(II)-dependent dioxygenase responsible for converting genomic 5-methylcytosine (mC) to 5-hydroxymethylcytosine (hmC) [33,34]. However, it is not known whether the CXXC domain of TET1 is involved in recognition of methylated DNA substrates.

Here we report a functional study and characterization of the DNA binding activity for the CXXC domains of mouse Dnmt1 and Tet1 proteins. We generated isolated CXXC domain and deletion constructs based on structural homology models to minimize structural alterations. We show that the CXXC domain of Dnmt1 preferentially binds DNA substrates containing unmethylated CpG sites, but does not contribute significantly to the DNA binding properties of the full length enzyme and is dispensable for its catalytic activity *in vitro* and *in vivo*. In addition, we found that the CXXC domain of Tet1 does not bind DNA *in vitro* and is also dispensable for catalytic activity of Tet1 *in vivo*.

Results

Sequence homology and structural modeling identify distinct CXXC domain subtypes

Dnmt1 contains a zinc finger domain of the CXXC type, which is present in several mammalian proteins including MLL1

(Figure 1A–C) and is highly conserved among Dnmt1 sequences from various animal species (Figure S1 in File S1). The primary structure of CXXC domains spans two clusters of 6 and 2 cysteine residues separated by a stretch of variable sequence and length. Sequence alignment and homology tree construction identified three distinct groups of CXXC domains (Figure 1B and C). The sequence between the two cysteine clusters in the CXXC domains of Dnmt1, CGBP/Cfp1, Fbxl19, Mll1, Mll2 and Kdm2 proteins and CXXC domain 3 of Mbd1 is highly conserved and contains a KFGG motif. The two other homology groups, including the CXXC domains 1 and 2 of Mbd1 on one side and those of Tet1, Cxxc4/Idax, Cxxc5/RINF and Cxxc10 on the other side, lack the KFGG motif and diverge from the first group and from each other in the sequence between the cysteine clusters. We generated structural homology models for the CXXC domains of mouse Dnmt1 and Tet1 using the NMR structure of the MLL1 CXXC domain as a template (Figure 1D and E)[35]. The CXXC domains of these proteins adopt an extended crescent-like structure that incorporates two Zn²⁺ ions each coordinated by four cysteine residues. The peptide of the MLL1 CXXC domain predicted to insert into the major groove of the DNA double helix (cyan in Fig. 1E) is located on one face of the structure and is contiguous to the KFGG motif [35]. The predicted structure of the Tet1 CXXC domain lacks the short β 1 helix (η 1 in Figure 1E) formed by residues PKF and partially overlapping the KFGG motif, but is similar to the MLL1 CXXC domain in the region of the DNA-contacting peptide. However, each of the two predicted β -strands in Tet1 carries three positive charges, whereas there is only one or no charged residue in the C-terminal strands of the CXXC domains in MLL1 and Dnmt1. Depending on the orientation of the positively charged side chains, it cannot be excluded that the charge density prevents strand pairing in the Tet1 CXXC domain.

The Dnmt1 CXXC domain binds unmethylated DNA

To investigate the binding properties of the Dnmt1 CXXC domain, we generated a GFP fusion construct including aa 652–699 (GFP-CXXC^{Dnmt1}). According to our homology model the ends of this fragment form an antiparallel β -sheet that structurally delimits the domain as in MLL1. We first compared the localization and mobility of GFP-CXXC^{Dnmt1} and GFP in mouse C2C12 myoblasts. While GFP was diffusely distributed in both nucleus and cytoplasm, GFP-CXXC^{Dnmt1} was exclusively nuclear with a punctuated pattern throughout the nucleoplasm and was enriched in nucleoli, a pattern independent of cell cycle stage (Figure 2A and Figure S2 in File S1). Enrichment in the nucleus and nucleoli is frequently observed with constructs containing stretches with high density of basic residues. After photobleaching half of the nuclear volume we observed a slower fluorescence recovery rate for GFP-CXXC^{Dnmt1} than for GFP (Figure 2B). To rule out a contribution of nucleolar interactions to the slower kinetics of GFP-CXXC^{Dnmt1},

we separately bleached nucleoplasmic and nucleolar regions and found that GFP-CXXC^{Dnmt1} has even faster kinetics within the nucleolus (Figure S3 in File S1). These results are consistent with a binding activity of GFP-CXXC^{Dnmt1} in the nucleus and very transient, unspecific binding in the nucleolus. To investigate whether the CXXC domain of Dnmt1 binds DNA and its possible selectivity with respect to CpG methylation we used a recently developed fluorescent DNA binding assay [36,37]. GFP-CXXC^{Dnmt1} was transiently expressed in HEK293T cells, immunopurified with the GFP-trap (Figure S4 in File S1) and incubated with fluorescent DNA substrates containing either no CpG site or one central un-, hemi- or fully methylated CpG site in direct competition. As shown in Figure 2C, GFP-CXXC^{Dnmt1} displayed a significant preference for the substrate containing one unmethylated CpG site, which increased substantially with a five-fold higher concentration of the DNA substrates (Figure S5 in File S1). These results are consistent with the reported binding preference of the CXXC domains in human DNMT1 and other factors belonging to the same CXXC homology group [21,22,29,32]. Notably, the CXXC domains 1 and 2 of Mbd1 lack the KFGG motif and do not bind DNA, while mutation of this motif prevented DNA binding by the CXXC domain of MLL1 [29,38]. Therefore,

we generated a GFP-CXXC^{Dnmt1} construct where the KFGG motif was mutated to AAGG (GFP-CXXC^{Dnmt1KF/AA}, Figure S4 in File S1) to test the requirement of the KFGG motif for binding by the CXXC domain of Dnmt1. The mutant domain showed significantly decreased binding to all DNA substrates and complete loss of preferential binding to the unmethylated substrate *in vitro* (Figure 2B). In addition, GFP-CXXC^{Dnmt1KE/AA} showed faster recovery after photobleaching (FRAP) *in vivo* compared to the corresponding wild type construct (Figure 2C). These results further support the importance of the KFGG motif for DNA binding by CXXC domains.

The CXXC domain of Tet1 shows no specific DNA binding activity and is dispensable for enzymatic activity *in vivo*

It was recently shown that Tet1 oxidizes genomic mC to hmC. However, the mechanism by which Tet1 is targeted to genomic mC is not known. Our model for the structure of the Tet1 CXXC domain diverged from the structure of the MLL1 CXXC domain with respect to the KFGG motif but not to the DNA-contacting peptide, suggesting that the Tet1 CXXC domain may still bind DNA. To test this we generated a GFP-tagged Tet1 CXXC

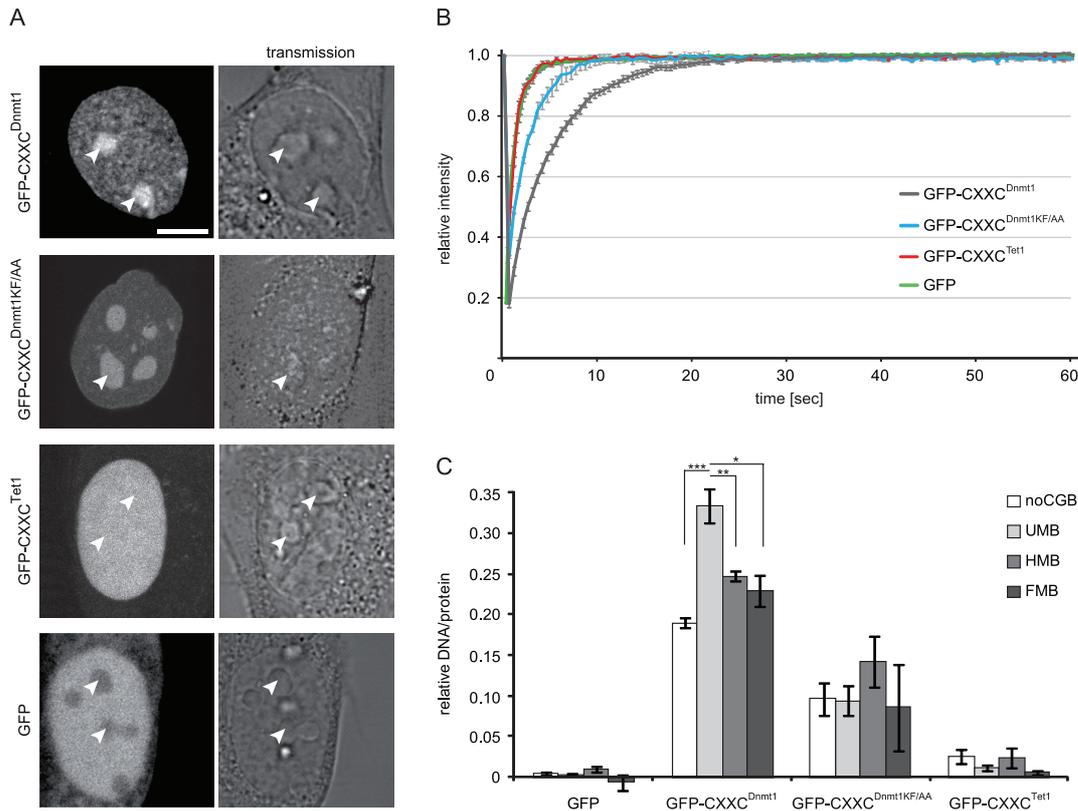


Figure 2. Properties of isolated Dnmt1 and Tet1 CXXC domains. (A–B) Subcellular localization (A) and binding kinetics (B) of GFP-CXXC^{Dnmt1}, GFP-CXXC^{Dnmt1KF/AA}, GFP-CXXC^{Tet1} and GFP in mouse C2C12 myoblasts. Localization and binding kinetics were independent from the cell cycle stage (Figures S2 and S5 in File S1). Arrowheads in (A) point to nucleoli. Scale bar: 5 μm. Binding kinetics were analyzed by FRAP. (C) DNA binding specificity of the Dnmt1 and Tet1 CXXC domains. GFP, GFP-CXXC^{Dnmt1}, GFP-CXXC^{Dnmt1KF/AA} and GFP-CXXC^{Tet1} were pulled down from extracts of transiently transfected HEK293T cells and incubated with fluorescent DNA substrates containing no CpG site or one central un-, hemi- or fully methylated CpG site in direct competition (noCGB, UMB, HMB, FMB, respectively). Shown are the mean DNA/protein ratios and corresponding standard errors from 5 (GFP), 4 (GFP-CXXC^{Dnmt1} and GFP-CXXC^{Dnmt1KF/AA}) and 2 (GFP-CXXC^{Tet1}) independent experiments. * $P = 0.01$; ** $P = 0.007$; *** $P = 0.001$. doi:10.1371/journal.pone.0016627.g002

construct (GFP-CXXC^{Tet1}) following the same criteria as for GFP-CXXC^{Dnmt1} and investigated its cellular localization, *in vivo* binding kinetics and *in vitro* DNA binding activity. GFP-CXXC^{Tet1} was prevalently nuclear with a homogeneous distribution including nucleoli that was independent of cell cycle stage (Figure 2A and Figure S6 in File S1). After photobleaching GFP-CXXC^{Tet1} showed very fast recovery kinetics similar to GFP (Figure 2B) and its DNA binding activity *in vitro* was also similar to the background levels of the GFP control (Figure 2C). We conclude that the isolated CXXC domain of Tet1 has no specific DNA binding activity. Together with the observation that the CXXC domains 1 and 2 of Mbd1 also lack the KFGG motif and do not bind DNA [29] and that mutation of this motif reduced DNA binding by the CXXC domains of both Dnmt1 (Figure 2C) and MLL1 [38], this result indicates that the KFGG motif is a major determinant for DNA binding by CXXC domains.

To assess whether the CXXC domain is required for catalytic activity of Tet1 we generated a GFP-Tet1 fusion construct and a corresponding mutant lacking the CXXC domain (GFP-Tet1^{ΔCXXC}). In C2C12 myoblasts GFP-Tet1 and GFP-Tet1^{ΔCXXC} showed punctuated nuclear patterns that did not depend on the cell cycle stage (Figure 3A and data not shown). The same constructs were transfected in HEK293T cells and global levels of genomic hmC were measured using a recently described hmC glucosylation assay [39]. Overexpression of GFP-Tet1 and GFP-Tet1^{ΔCXXC} determined a similar 5-fold increase of genomic hmC levels relative to control samples overexpressing GFP (Figure 3B), indicating that the CXXC domain is not required for enzymatic activity of Tet1 *in vivo*.

Deletion of the CXXC domain does not affect the activity of Dnmt1 *in vitro*

To explore the role of the CXXC domain in Dnmt1 function we generated GFP-Dnmt1 fusion constructs where the CXXC domain, as defined by our homology model, was deleted. We reasoned that precise deletion of the entire structure delimited by the antiparallel β -sheet (Figure 1D) would have the highest chances to preserve native folding of the rest of the protein. We introduced this deletion in GFP fusion constructs encoding either the full length Dnmt1 or the isolated N-terminal region (GFP-Dnmt1^{ΔCXXC} and GFP-NTR^{ΔCXXC}, respectively; Figure 4A and

Figure S4 in File S1). We then compared DNA binding properties, catalytic activity and interaction between N-terminal region and C-terminal catalytic domain of Δ CXXC and corresponding wild type constructs. Competitive DNA binding assays with the same set of substrates as used for the experiments with GFP-CXXC^{Dnmt1} and GFP-CXXC^{Tet1} reported above (Figure 2C) showed that both GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} bind DNA independently of the presence and methylation state of a CpG site (Figure 4B). As the isolated CXXC domain preferentially bound the substrate containing an unmethylated CpG site, the result with GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} indicates that the CXXC domain contributes negligibly to the DNA binding specificity of the full-length enzyme.

Several groups reported that interaction between the N-terminal region and the C-terminal catalytic domain of Dnmt1 leads to allosteric activation of Dnmt1 [16,18–20,40]. To test whether the CXXC domain is involved in this intramolecular interaction, we co-expressed either GFP-tagged N-terminal region (GFP-NTR) or GFP-NTR^{ΔCXXC} constructs with a Cherry- and His-tagged C-terminal domain (Ch-CTD-His) in HEK293T cells and performed co-immunoprecipitation experiments. Ch-CTD-His co-precipitated both GFP-NTR and GFP-NTR^{ΔCXXC}, indicating that the CXXC domain is dispensable for the interaction between the N-terminal region and the C-terminal domain of Dnmt1 (Figure 4C).

To investigate whether the CXXC domain is needed for enzymatic activity or substrate recognition, we tested formation of the covalent complex with cytosine and transfer of the methyl group for GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC}. We first employed an assay to monitor covalent complex formation that exploits the formation of an irreversible covalent bond between the enzyme and the mechanism-based inhibitor 5-aza-2-deoxycytosine (5-aza-dC). This results in permanent trapping of the enzyme by DNA substrates containing 5-aza-dC, as opposed to the reversible complex formed with substrates containing the natural substrate 2-deoxycytosine (dC) [36]. GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} were incubated with fluorescent DNA substrates containing either dC (binding) or 5-aza-dC (trapping) at a single CpG site in direct competition. DNA-protein complexes were then isolated by GFP pulldown and molar DNA/protein ratios were calculated from fluorescence measurements (Figure 4D). Covalent complex

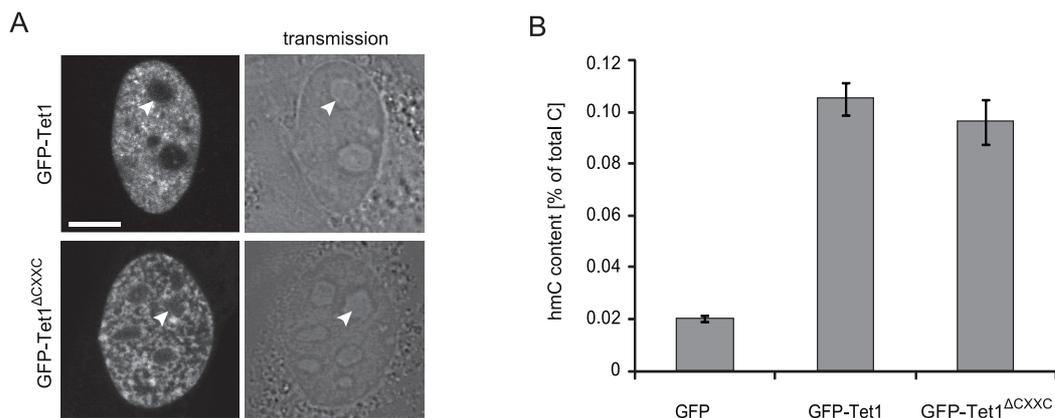


Figure 3. Cellular localization and *in vivo* catalytic activity of GFP-Tet1 and GFP-Tet1^{ΔCXXC}. (A) Live images of C2C12 myoblasts expressing GFP-Tet1. Scale bar: 5 μ m. (B) Genomic hmC content in HEK293T cells overexpressing GFP, GFP-Tet1 and GFP-Tet1^{ΔCXXC}. Shown are mean values and standard deviation of hmC percentage over total cytosine for three measurements from one transfection. doi:10.1371/journal.pone.0016627.g003

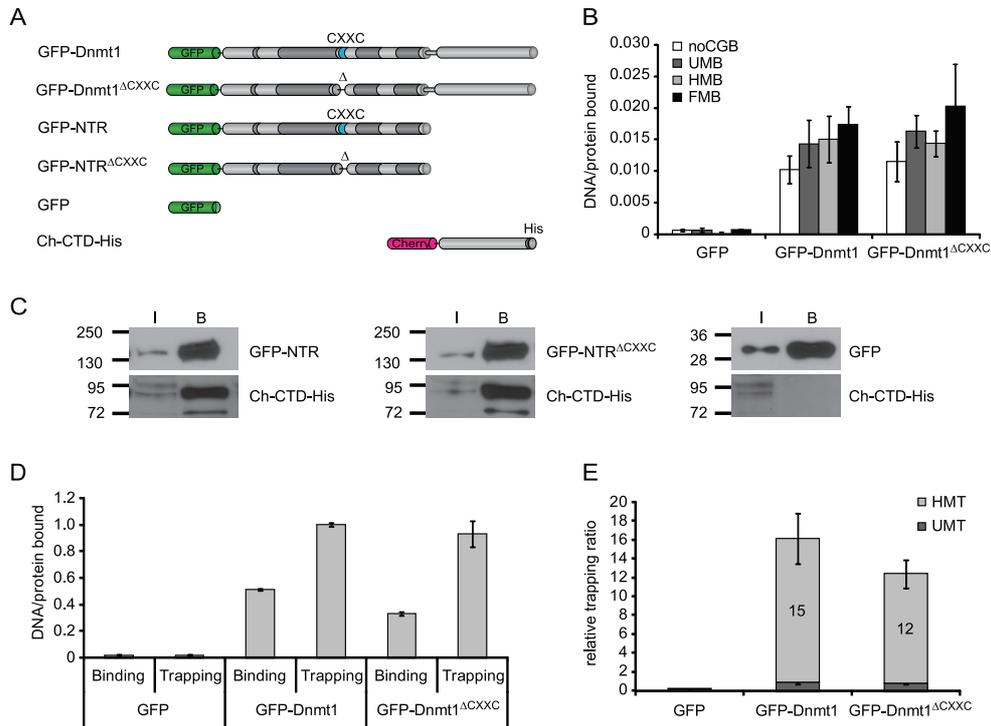


Figure 4. DNA binding specificity, intramolecular interaction and trapping of wild-type Dnmt1 and CXXC deletion constructs *in vitro*. (A) Schematic representation of Dnmt1 expression constructs. (B) DNA binding specificity of GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} were assayed as described in Figure 2C. (C) Co-immunoprecipitation of the C-terminal domain of Dnmt1 (Ch-CTD-His) and the N-terminal region with and without deletion of the CXXC domain (GFP-NTR and GFP-NTR^{ΔCXXC}, respectively). GFP fusions were detected using an anti-GFP antibody, while the C-terminal domain construct was detected using an anti-His antibody. GFP was used as negative control. I = input, B = bound. (D) Comparison of binding and trapping activities for GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} to monitor irreversible covalent complex formation with hemimethylated substrates. (E) Relative covalent complex formation rate of GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} on substrates containing one un- (UMT) or hemi-methylated CpG site (HMT) in direct competition. The trapping ratio for GFP-Dnmt1 on unmethylated substrate was set to 1. In (D) and (E) the means and corresponding standard deviations of triplicate samples from three independent experiments are shown. GFP was used as negative control.
doi:10.1371/journal.pone.0016627.g004

formation was then estimated by comparing trapping and binding activities. GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} showed comparable covalent complex formation rates (trapping/binding ratios), which were about 15- and 12-fold higher for hemi- than unmethylated substrates, respectively (Figure 4E). Together with the data from binding experiments (Fig. 4B), this result indicates that the preference of Dnmt1 for hemimethylated substrates is determined at the covalent complex formation step rather than upon DNA binding. Furthermore, the CXXC domain clearly does not play a major role in determining either the efficiency or the methylation state-specificity of covalent complex formation.

Next, we tested whether deletion of the CXXC domain affects the ability of Dnmt1 to transfer [³H]methyl groups from the donor S-adenosylmethionine (SAM) to a poly(dI·dC)-poly(dI·dC) substrate, a standard DNA methyltransferase activity assay. This showed that *in vitro* GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} are equally active methyltransferases (Figure S7 in File S1). This result is in contrast with a previous report showing that deletion of aa 647–690 in human DNMT1 encompassing the CXXC domain resulted in a drastic loss of catalytic activity [32]. However, according to our homology model the deletion by Pradhan *et al.* would eliminate the predicted N-terminal β-strand (β1 in Figure 1) preventing the formation of the antiparallel β-sheet and potentially distort the

folding of the rest of the protein. This is in contrast with our GFP-Dnmt1^{ΔCXXC} mutant that was designed to retain the β-sheet structure. To test whether this may account for the observed discrepancy, we generated GFP fusion constructs of wild type human DNMT1 and the same deletion as reported by Pradhan *et al.* and tested covalent complex formation with 5-aza-dC containing DNA substrates as described above. While the human wild type construct showed the same preference for hemimethylated over unmethylated trapping substrates as the mouse constructs, this preference was clearly reduced for the human CXXC deletion mutant (Figure S8 in File S1). This result is consistent with the loss of enzymatic activity shown by Pradhan *et al.* for this mutant and together with the retention of trapping and catalytic activity by the different deletion in GFP-Dnmt1^{ΔCXXC} suggests that disruption of the antiparallel β-sheet delimiting the CXXC domain results in further distortion and loss of activity of the enzyme.

In conclusion, we showed that, *in vitro*, deletion of the CXXC domain does not affect the interaction between N-terminal region and C-terminal domain, DNA binding, the preference for hemimethylated substrates upon covalent complex formation and the methyltransferase activity of Dnmt1. Together, these data strongly argue against an involvement of the CXXC domain in allosteric activation of Dnmt1.

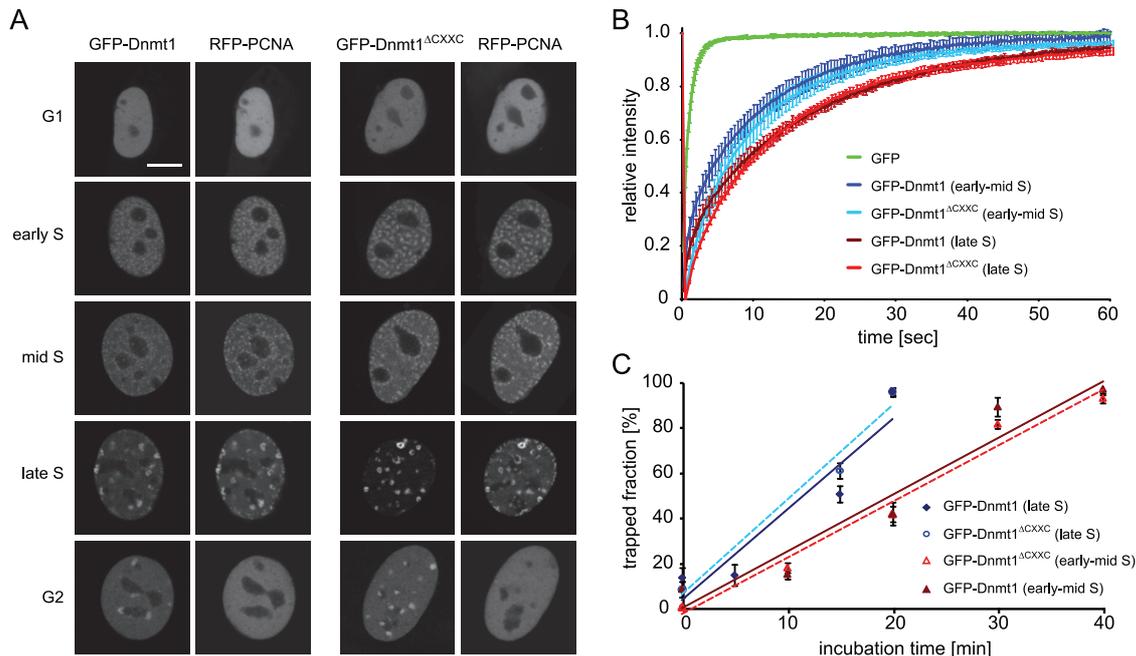


Figure 5. Cell cycle dependent cellular localization, protein mobility and trapping of wild-type Dnmt1 and CXXC deletion constructs in mouse C2C12 myoblasts. (A) Cell cycle dependent localization of GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} constructs. Scale bar: 5 μm. (B) Analysis of binding kinetics of GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} in early and late S-phase cells by FRAP. The recovery curve for GFP is shown for comparison. (C) *In vivo* trapping by FRAP analysis in cells treated with 5-aza-dC. The trapped enzyme fraction is plotted over time for early and late S-phase cells. For each construct three to six cells in early-mid and late S phase were analysed per time point. Shown are mean values ± SEM. In (A–C) RFP-PCNA was cotransfected to identify cell cycle stages in living cells. doi:10.1371/journal.pone.0016627.g005

Deletion of the CXXC domain does not affect Dnmt1 activity *in vivo*

We then undertook a functional characterization of the GFP-Dnmt1^{ΔCXXC} construct *in vivo*. We first compared localization and binding kinetics of GFP-Dnmt1 or GFP-Dnmt1^{ΔCXXC} in mouse C2C12 myoblasts co-transfected with RFP-PCNA, which served as S-phase marker [41]. GFP-Dnmt1^{ΔCXXC} showed the same cell-cycle dependent nuclear localization pattern as previously shown for GFP-Dnmt1 and endogenous Dnmt1 (Figure 5A)[42,43]. Interaction with PCNA via the PBD directs Dnmt1 to replication foci throughout S-phase. In addition, in late S-phase and G2 Dnmt1 is enriched at chromocenters, clusters of pericentric heterochromatin (PH) that are observed as discrete domains densely stained by DNA dyes in mouse interphase cells. Association of Dnmt1 with PH at these stages is mediated by the TS domain [42]. Thus, the CXXC domain clearly does not contribute to the subnuclear localization of Dnmt1 at this level of resolution.

We also compared the mobility of GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} in living C2C12 myoblasts by FRAP analysis (Figure 5B). These experiments revealed that the kinetics of Dnmt1 is not significantly affected by deletion of the CXXC domain in early-mid as well as late S-phase.

To test covalent complex formation in living cells, we used a previously established trapping assay [44]. Mouse C2C12 myoblasts were co-transfected with RFP-PCNA and either GFP-Dnmt1 or GFP-Dnmt1^{ΔCXXC} and treated with 5-aza-dC. Immobilization of the Dnmt1 constructs at the site of action was then measured by FRAP analysis (Figure 5C). GFP-Dnmt1 and

GFP-Dnmt1^{ΔCXXC} showed very similar trapping kinetics, the immobile enzyme fraction reaching nearly 100% after 20 and 40 minutes in early-mid and late S-phase, respectively. This result clearly shows that the CXXC domain is dispensable for covalent complex formation also *in vivo*.

Finally, we compared the ability of GFP-Dnmt1 and GFP-Dnmt1^{ΔCXXC} to restore DNA methylation patterns in mouse *dnmt1*^{-/-} ESCs. Cells transiently expressing either GFP-Dnmt1 or GFP-Dnmt1^{ΔCXXC} were FACS sorted 48 h after transfection. Isolated genomic DNA was then bisulfite treated and fragments corresponding to major satellite repeats, intracisternal type A particle (IAP) interspersed repeats, *skeletal α-actin* and *H19a* promoters were amplified and subjected to pyrosequencing (Figure 6). As shown previously [43], under these conditions GFP-Dnmt1 partially restored methylation of major satellite and IAP repeats and the *skeletal α-actin* promoter, but not of the imprinted *H19a* promoter, as establishment of the methylation imprint requires passage through the germ line [45]. Methylation patterns of all these sequences in cells expressing GFP-Dnmt1^{ΔCXXC} were very similar to those in GFP-Dnmt1 expressing cells, including the lack of (re-) methylation at the *H19a* promoter. These results suggest that the CXXC domain is not required for maintenance of DNA methylation patterns by Dnmt1 and does not restrain the DNA methyltransferase activity of Dnmt1 on unmethylated CpG sites. Thus, the CXXC domain does not play a major role in subcellular localization, it does not contribute to the global binding kinetics of Dnmt1 and, consistent with the *in vitro* data reported above, is dispensable for maintaining DNA methylation patterns in living cells.

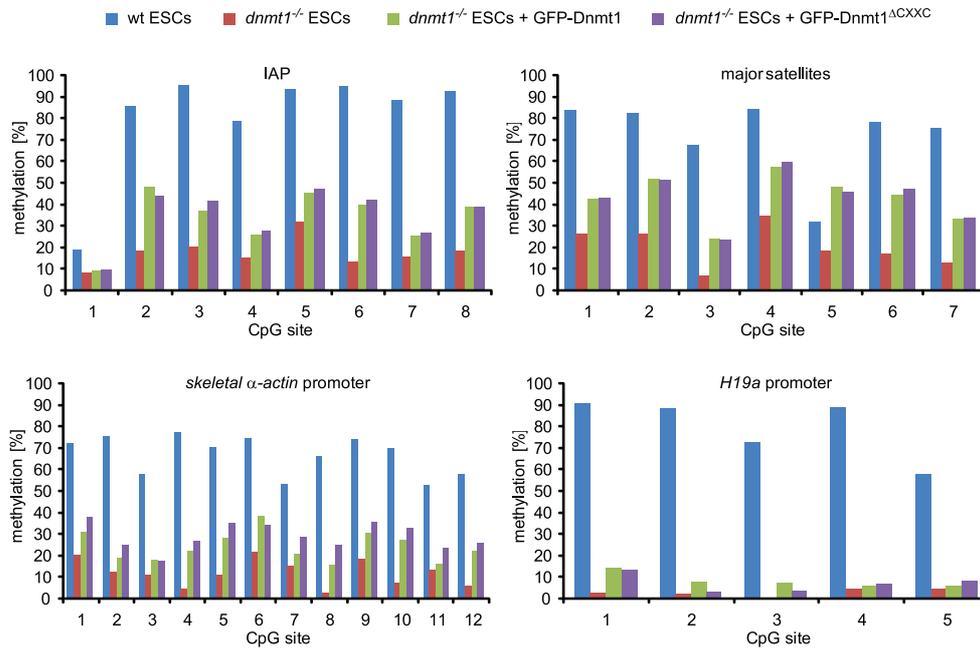


Figure 6. The CXXC deletion construct of Dnmt1 restores methylation in *dnmt1* null cells. Mouse *dnmt1*^{-/-} ESCs transiently expressing GFP-Dnmt1 or GFP-Dnmt1^{ΔCXXC} were isolated by FACS-sorting 48 h after transfection and CpG methylation levels within the indicated sequences were analyzed by bisulfite treatment, PCR amplification and direct pyrosequencing. Methylation levels of untransfected wild type and *dnmt1*^{-/-} ESCs are shown for comparison.
doi:10.1371/journal.pone.0016627.g006

Discussion

We generated homology models based on the reported structure of the MLL1 CXXC domain to design isolated CXXC domain constructs and CXXC domain deletion mutants for Dnmt1 and Tet1 with minimal probability of structural alteration. According to these models CXXC domains are delimited by an antiparallel β -sheet, a discrete structural element. Our data show that the CXXC domain of mouse Dnmt1 preferentially binds DNA substrates containing unmethylated CpG sites as previously shown for CXXC domains of human DNMT1 and other mammalian proteins. We note that sequences C-terminal to the corresponding peptide in CGBP/Cfp1 were reported to be required for DNA binding *in vitro* [22] and that only a significantly larger peptide spanning the CXXC-3 domain of Mbd1a was tested for DNA binding. However, sequences C-terminal to CXXC domains are not conserved (Figure 1B) and our data show that they are not required for DNA binding by the CXXC domain of Dnmt1. Nevertheless, all the CXXC domains reported to selectively bind unmethylated CpG sites cluster in a distinct homology group and contain the KFGG motif. The latter was shown to be crucial for DNA binding by the CXXC domain of MLL1 [38] and here we extend this observation to the CXXC domain of Dnmt1. Sequence alignment reveals two distinct CXXC domain homology groups that lack the KFGG motif (Figure 1A). Consistent with a role of this motif in DNA binding, members of these groups such as CXXC-1/2 of Mbd1 [29] and the CXXC domain of Tet1 (this study) show no DNA binding activity. While no specific function is known for CXXC-1/2 of Mbd1, the CXXC domain of Tet1 is closely related to those in CXXC4/Idax and CXXC5/RINF that were shown to mediate protein-protein interactions [46–48]. This

suggests that the CXXC domain of Tet1, rather than mediating DNA binding, may function as a protein-protein interaction domain. However, our data do not rule out the possibility that the DNA binding properties of the CXXC domain within the context of full length Tet1 may be different from those of the isolated domain. Nevertheless, we show that the CXXC domain is not required for enzymatic activity of Tet1 *in vivo*.

Although we observed a clear DNA binding activity by the isolated CXXC domain of Dnmt1, we found that, within the context of the full length enzyme, this domain is dispensable for overall DNA binding properties, preference for hemimethylated substrates upon covalent complex formation, methyltransferase activity and allosteric activation as well as for the ability to restore methylation of representative sequences in *dnmt1* null ESCs. Consistent with our data, a recent report showed a preference of the CXXC domain of human DNMT1 for substrates containing unmethylated CpG sites [32]. However, the same report showed that deletion of the CXXC domain from the human enzyme results in a significant decrease in methyltransferase activity on hemimethylated substrates *in vitro* and 25% lower methylation at rDNA repeats upon overexpression in HEK293 cells, suggesting a dominant negative effect of the deletion construct. These discrepancies may be due to the fact that the fragment deleted by Pradhan *et al.* includes the N-terminal strand of the predicted antiparallel β -sheet, potentially leading to disruption of native folding, to species-specific differences and/or to the analysis of non-physiological expression levels in HEK293 cells. In our trapping assay the same human deletion mutant showed reduced covalent complex formation, consistent with loss of enzymatic activity. The report from Pradhan *et al.* also showed that mutation of cysteine 667 to glycine within the CXXC domain of human

DNMT1 disrupts DNA binding and enzymatic activity. However, as this point mutation involves one of the zinc coordinating residues it is not unlikely to alter peptide folding with negative consequences potentially extending beyond the CXXC domain and including reduced enzymatic activity. In this respect the dominant negative effect observed upon overexpression of this mutant may be explained by the prevalent occurrence of Dnmt1 as a dimer [49]. These observations, together with preserved ability for covalent complex formation and catalytic activity of our CXXC domain deletion, support the validity of our homology model-driven approach for functional characterization of the CXXC domain. In addition, our genetic complementation approach constitutes a rather physiologic functional assay. However, due to the transient approach and the analysis of genomic methylation at only a few representative sequences, subtle or highly sequence specific effects of deletion of the CXXC domain cannot be excluded.

It was recently shown that binding of Cfp1/CGBP and KDM2A to CpG islands through their CXXC domains leads to local enrichment and depletion of H3K4 and H3K36 methylation, respectively [26,30]. Analogously, Dnmt1 may bind CpG islands through its CXXC domain. However, this interaction would not lead to a straightforward functional interpretation as CpG islands with high CpG density are generally refractive to DNA methylation and a function of Dnmt1 as a *de novo* DNA methyltransferase is not well established. It could be envisaged that binding to unmethylated CpG sites/islands by the CXXC domain may have a negative effect on the enzymatic activity of Dnmt1 and restrain its function as a *de novo* DNA methyltransferase. However, we show that in *dnmt1* null ESCs methylation of the imprinted *H19a* promoter is not restored upon expression of either wild type or Δ CXXC Dnmt1 constructs, arguing against a negative regulatory function of the CXXC domain.

Notably, binding of unmethylated CpG sites by KFGG motif-containing CXXC domains does not exclude a role in protein-protein interaction as the CXXC domain of MLL1 was reported to interact with both DNA and Polycomb Repressive Complex 1 components HPC2/CBX4 and BMI-1 [21,50]. Therefore, it is possible that the CXXC domain of Dnmt1 has regulatory functions in specific cell types or developmental stages that may involve DNA binding and/or interaction with other proteins. The generation of dedicated animal models may be instrumental for testing these possibilities.

Materials and Methods

Bioinformatic methods

Alignments were performed using the ClustalW2 software [51]. The CXXC domain homology tree (Figure 1C) was generated from the alignment in Figure 1B with Jalview 2.4 by unweighted pair group method with arithmetic mean (UPGMA). The neighbor-joining method gave the same result. Average distances between the sequences were calculated using the BLOSSUM62 matrix. The human CXXC10 coding sequence [52] was determined by assembling ESTs AI438961, BX114363, BX492895, BU633058.1, AW207644.1 and the genomic sequence AC073046.7. The putative translational start site is located 16308 bp upstream of the annotated transcriptional start site of *TET3*. A partial coding sequence of murine *Cxxc10* containing the CXXC domain was identified by aligning the human CXXC10 protein sequence to the ORFs present in NT_039353.7 upstream of the *tet3* gene from position 35663306 to 35808487). A very high match was found 13266 nt upstream of *tet3* at positions 35676374-35676572 of NT_039353.7. To build

homology models for the CXXC domains of Dnmt1 (aa 645–696) and Tet1 (aa 561–614), we submitted the respective sequences to the HHpred server [53]. The best template was the CXXC domain of MLL1 (PDB-ID: 2J2S). The 49 residues of the CXXC domain in Dnmt1 can be aligned to this domain with 45% sequence identity and only a single amino acid gap after residue 661 (Figure 1B). 3D models were calculated with the homology modeling software MODELLER [54] (version 9.5) using this alignment. Distance restraints were given to MODELLER to enforce a distance of 2.3 ± 0.1 Å between the eight sulphurs in the Zn-coordinating cysteines and the Zn²⁺ ions. TM-align [55] was used to superpose the model structure with the template domain. Images were generated using the PyMol Molecular Graphics System (Version 1.3, Schrödinger, LLC). The quality of the models and the underlying alignments were checked with DOPE [56] and Verify3D [57] and results for both models were found to be comparable to the MLL1 template structure (2J2S).

Expression constructs

Fusion constructs were generated using enhanced green fluorescent protein, monomeric red fluorescent protein or monomeric cherry and are here referred to as GFP, RFP and Cherry fusions, respectively. Mammalian expression constructs for GFP, mouse GFP-Dnmt1, GFP-NTR and human RFP-PCNA were described previously [42,44,49,58]. The deletion construct GFP-Dnmt1^{ΔCXXC} was obtained by replacing the sequence coding for aa 655–696 with three alanine codons in the GFP-Dnmt1 construct as described [59]. The GFP-DNMT1^{ΔCXXC} construct was generated by subcloning the sequence coding for human DNMT1^{ΔCXXC} from the homonymous construct by Pradhan *et al.* [32] in the pEGFP-C2 vector (Clontech). To generate GFP-Tet1 three partially overlapping fragments spanning the Tet1 coding sequence were amplified using E14 ESCs cDNA as template. The fragments were then joined by overlap extension PCR and inserted into the pCAG-GFP-IB vector [43]. To generate GFP-Tet1^{ΔCXXC} aa 569-621 of murine Tet1 were deleted from GFP-Tet1 using a type II restriction endonuclease approach as described [60]. To generate GFP-CXXC^{Dnmt1} and GFP-CXXC^{Tet1} sequences coding for the respective CXXC domains (aa 643-700 for Dnmt1 and 561-614 for Tet1) were amplified by PCR using the GFP-Dnmt1 expression construct and cDNA from E14 ESCs as templates, respectively. PCR fragments were then inserted into the pCAG-GFP-IB vector. GFP-NTR^{ΔCXXC} was obtained by replacing the BglII-XhoI fragment of GFP-NTR with the same fragment of GFP-Dnmt1^{ΔCXXC}. Ch-CTD-His was generated by replacing the GFP coding sequence in a GFP-CTD construct [49] with the Cherry coding sequence. All constructs were confirmed by sequencing.

Cell culture, transfection and cell sorting

HEK293T cells [61] and mouse C2C12 myoblasts [62] were cultured in DMEM supplemented with 50 μg/ml gentamicin and 10% and 20% fetal calf serum, respectively. For expression of fusion proteins HEK293T cells were transfected with polyethylenimine (Sigma). For live cell imaging, C2C12 cells were grown to 40% confluence on Lab-Tek chambers (Nunc) or μ-slides (Ibidi) and transfected with TransFectin transfection reagent (BioRad) according to the manufacturer's instructions. Mouse ESCs were cultured as described [63] and transfected with FuGENE HD (Roche) according to the manufacturer's instructions. ESCs were sorted with a FACS Aria II instrument (Becton Dickinson). The *dnmt1*^{-/-} J1 ESCs used in this study are homozygous for the c allele [14].

In vitro DNA binding and trapping assays

In vitro DNA binding and trapping assays were performed as described previously [36,37] with the following modifications. DNA substrates labeled with four different ATTO fluorophores (Tables S1 and S2 in File S1) were used at a final concentration of 125 nM each in the pull-down assay with immobilized GFP fusions. After removal of unbound substrate, the amounts of protein and DNA were determined by fluorescence intensity measurements with a Tecan Infinite M1000 plate reader using calibration curves from purified GFP or DNA coupled ATTO fluorophores, respectively. The following excitation/emission \pm detection bandwidth settings were used: 490/511 \pm 10 nm for GFP, 550/580 \pm 15 nm for ATTO550, 600/630 \pm 15 nm for ATTO590, 650/670 \pm 10 nm for ATTO647N and 700/720 \pm 10 nm for ATTO700. Cross detection of GFP and different ATTO dyes was negligible with these settings. Binding and trapping ratios were calculated dividing the concentration of bound DNA substrate by the concentration of GFP fusion on the beads.

In vivo mC hydroxylation assay

Genomic DNA was isolated from HEK293T cells 24 h after transfection with the GFP-Tet1 and GFP-Tet1^{ACXXC} constructs and global hmC levels were measured using the *in vitro* glucosylation assay as previously described [63], except that 100 nM β -glucosyltransferase and only UDP-[³H]glucose donor (0.43 μ M) were used.

Co-immunoprecipitation

Co-immunoprecipitation was performed as described previously [49,64]. Shortly, HEK293T cells were transiently co-transfected with expression plasmids for GFP fusions and the Ch-CTD-His construct, harvested and lysed. GFP fusions were pulled down using the GFP-Trap [65] (Chromotek) and subjected to western blotting using anti-GFP (Roche or Chromotek) and anti-His (Invitrogen) monoclonal antibodies.

Live cell microscopy, FRAP analysis and live cell trapping assay

Live cell imaging and FRAP experiments were performed as described previously [43]. For each construct 6–15 nuclei were averaged and the mean values as well as the standard errors were calculated. For presentation, we used linear contrast enhancement on entire images. The DNA methyltransferase trapping assay was described previously [44]. Briefly, transfected cells were incubated with 30 μ M 5-aza-dC (Sigma) for the indicated periods of time before photobleaching experiments. FRAP analysis was performed with a confocal laser scanning microscope (TCS SP5, Leica)

equipped with a 63 \times /1.4 NA Plan-Apochromat oil immersion objective. Microscope settings were as described except that a smaller region of interest (3 μ m \times 3 μ m) was selected for photobleaching. Mean fluorescence intensities of the bleached region were corrected for background and for total loss of nuclear fluorescence over the time course, and normalized by the mean of the last 10 prebleach values.

DNA Methylation Analysis

Genomic DNA was isolated with the QIAmp DNA Mini Kit (Qiagen) and 1.5 μ g were bisulfite converted using the EZ DNA Methylation-Gold Kit (Zymo research) according to the manufacturer's instructions. Primer sets and PCR conditions for IAP-LTR, *skeletal α -actin* and *H19* promoters were as described [43]. Primer sequences for major satellites were AAAATGAGAA-CATCCACTTG (forward primer) and CCATGATTTT-CAGTTTCTT (reverse primer). For amplification we used Qiagen Hot Start Polymerase in 1 \times Qiagen Hot Start Polymerase buffer supplemented with 0.2 mM dNTPs, 0.2 μ M forward primer, 0.2 μ M reverse primer, 1.3 mM betaine (Sigma) and 60 mM tetramethylammonium-chloride (TMAC, Sigma). Promoter regions and IAP-LTR were amplified with two subsequent (nested) PCR reactions and major satellite repeats were amplified with a single amplification reaction. Pyrosequencing reactions were carried out by Varionostic GmbH (Ulm, Germany). Pyrosequencing primers are listed in Table S3 in File S1.

Supporting Information

File S1 Tables S1–S3, Figures S1–S8 and Supplemental methods. (PDF)

Acknowledgments

The authors thank Sabine Brunner and Lucia Puchbauer for technical assistance in the generation of the homology models. We also thank Taiping Chen and En Li (Novartis Institutes for Biomedical Research, Boston, MA) for providing *dmt1* null ESCs and Shriharsa Pradhan (New England BioLabs, Ipswich, MA) for the human DNMT1^{ACXXC} construct and constructive discussion.

Author Contributions

Conceived and designed the experiments: FS HL. Performed the experiments: CF AR SB DM KF SH MW WQ. Analyzed the data: CF AR SB DM KF SH JS. Contributed reagents/materials/analysis tools: JS. Wrote the paper: FS HL. Generated homology models: JS.

References

- Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, et al. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 27: 361–368.
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9: 465–476.
- Laurent L, Wong E, Li G, Huynh T, Tsirigos A, et al. (2010) Dynamic changes in the human methylome during differentiation. *Genome Res* 20: 320–331.
- Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315–322.
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, et al. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30: 755–766.
- Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, et al. (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol* 27: 353–360.
- Schmidl C, Klug M, Boeld TJ, Andreesen R, Hoffmann P, et al. (2009) Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res* 19: 1165–1174.
- Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, et al. (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Res* 20: 972–980.
- Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, et al. (2003) Induction of tumors in mice by genomic hypomethylation. *Science* 300: 489–492.
- Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20: 116–117.
- Xu G-L, Bestor TH, Bourc'his D, Hsieh C-L, Tommerup N, et al. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402: 187–191.
- Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74: 481–514.

13. Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71: 865–873.
14. Lei H, Oh SP, Okano M, Juttermann R, Goss KA, et al. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122: 3195–3205.
15. Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69: 915–926.
16. Margot JB, Aguirre-Arteaga AM, Di Giacomo BV, Pradhan S, Roberts RJ, et al. (2000) Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J Mol Biol* 297: 293–300.
17. Spada F, Rothbauer U, Zolghadr K, Schermelleh L, Leonhardt H (2006) Regulation of DNA methyltransferase 1. *Adv Enzyme Regul* 46: 224–234.
18. Zimmermann C, Guhl E, Graessmann A (1997) Mouse DNA methyltransferase (MTase) deletion mutants that retain the catalytic domain display neither de novo nor maintenance methylation activity in vivo. *Biol Chem* 378: 393–405.
19. Pradhan S, Esteve PO (2003) Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. *Biochemistry* 42: 5321–5332.
20. Fatemi M, Hermann A, Pradhan S, Jeltsch A (2001) The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J Mol Biol* 309: 1189–1199.
21. Birke M, Schreiner S, Garcia-Cuellar MP, Mahr K, Titgemeyer F, et al. (2002) The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucl Acids Res* 30: 958–965.
22. Lee JH, Voo KS, Skalik DG (2001) Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem* 276: 44669–44676.
23. Cross SH, Meehan RR, Nan X, Bird A (1997) A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet* 16: 256–259.
24. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, et al. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439: 811–816.
25. Frescas D, Guardavaccaro D, Bassermann F, Koyama-Nasu R, Pagano M (2007) JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. *Nature* 450: 309–313.
26. Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, et al. (2010) CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. *Mol Cell* 38: 179–190.
27. Lorschach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, et al. (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.
28. Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, et al. (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 Res* 62: 4075–4080.
29. Jorgensen HF, Ben-Porath I, Bird AP (2004) Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol Cell Biol* 24: 3387–3395.
30. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, et al. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464: 1082–1086.
31. Bestor TH (1992) Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J* 11: 2611–2617.
32. Pradhan M, Esteve PO, Chin HG, Samaranyake M, Kim GD, et al. (2008) CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry* 47: 10000–10009.
33. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324: 930–935.
34. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129–1133.
35. Allen MD, Grummitt CG, Hilcenko C, Min SY, Tonkin LM, et al. (2006) Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *EMBO J* 25: 4503–4512.
36. Frauer C, Leonhardt H (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucl Acids Res* 37: e22.
37. Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, et al. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucl Acids Res* 38: 1796–1805.
38. Ayton PM, Chen EH, Cleary ML (2004) Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol Cell Biol* 24: 10470–10478.
39. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Research* 38: e181.
40. Bacolla A, Pradhan S, Larson JE, Roberts RJ, Wells RD (2001) Recombinant human DNA (cytosine-5) methyltransferase. III. Allosteric control, reaction order, and influence of plasmid topology and triplet repeat length on methylation of the fragile X CGG.CCG sequence. *J Biol Chem* 276: 18605–18613.
41. Easwaran HP, Leonhardt H, Cardoso MC (2005) Cell Cycle Markers for Live Cell Analyses. *Cell Cycle* 4: 453–455.
42. Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep* 5: 1181–1186.
43. Schermelleh L, Haemmer A, Spada F, Rosing N, Meilinger D, et al. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucl Acids Res* 35: 4301–4312.
44. Schermelleh L, Spada F, Easwaran HP, Zolghadr K, Margot JB, et al. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat Methods* 2: 751–756.
45. Tucker KL, Beard C, Dausmann J, Jackson-Grusby L, Laird PW, et al. (1996) Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev* 10: 1008–1020.
46. Andersson T, Södersten E, Duckworth JK, Cascaete A, Fritz N, et al. (2009) CXXC5 Is a Novel BMP4-regulated Modulator of Wnt Signaling in Neural Stem Cells. *J Biol Chem* 284: 3672–3681.
47. Hino S-i, Kishida S, Michiue T, Fukui A, Sakamoto I, et al. (2001) Inhibition of the Wnt Signaling Pathway by Idax, a Novel Dvl-Binding Protein. *Mol Cell Biol* 21: 330–342.
48. London TBC, Lee H-J, Shao Y, Zheng J (2004) Interaction between the internal motif KTXXXI of Idax and mDvl PDZ domain. *Biochem Biophys Res Commun* 322: 326–332.
49. Fellingner K, Rothbauer U, Felle M, Langst G, Leonhardt H (2009) Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. *J Cell Biochem* 106: 521–528.
50. Xia ZB, Anderson M, Diaz MO, Zeleznik-Le NJ (2003) MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. *Proc Natl Acad Sci U S A* 100: 8342–8347.
51. Thomson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res* 22: 4673–4680.
52. Katoh M (2004) Identification and characterization of human CXXC10 gene in silico. *Int J Oncol* 25: 1193–1199.
53. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucl Acids Res* 33: W244–248.
54. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins* 23: 318–326.
55. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucl Acids Res* 33: 2302–2309.
56. Shen M-y, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15: 2507–2524.
57. Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: Assessment of protein models with three-dimensional profiles. In: Carter CWJ, Sweet RM, eds. *Methods Enzymol*: Academic Press. pp 396–404.
58. Sporbert A, Domaing P, Leonhardt H, Cardoso MC (2005) PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucl Acids Res* 33: 3521–3528.
59. Fellingner K, Leonhardt H, Spada F (2008) A mutagenesis strategy combining systematic alanine scanning with larger mutations to study protein interactions. *Anal Biochem* 373: 176–178.
60. Ko J-K, Ma J (2005) A rapid and efficient PCR-based mutagenesis method applicable to cell physiology study. *Am J Physiol Cell Physiol* 288: C1273–1278.
61. DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, et al. (1987) Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol* 7: 379–387.
62. Blau HM, Pavlath GK, Hardeman EC, Chiu CP, Silberstein L, et al. (1985) Plasticity of the differentiated state. *Science* 230: 758–766.
63. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucl Acids Res* 38: e181.
64. Meilinger D, Fellingner K, Bultmann S, Rothbauer U, Bonapace IM, et al. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* 10: 1259–1264.
65. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, et al. (2007) A versatile nanotrapp for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* 7: 282–289.
66. Mohseni-Zadeh S, Brezellec P, Risler JL (2004) Cluster-C, an algorithm for the large-scale clustering of protein sequences based on the extraction of maximal cliques. *Comput Biol Chem* 28: 211–218.

Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1

Carina Frauer[#], Andrea Rottach[#], Daniela Meilinger, Sebastian Bultmann, Karin Fellingner, Stefan Hasenöder, Mengxi Wang, Weihua Qin, Johannes Söding[§], Fabio Spada* and Heinrich Leonhardt*

Department of Biology and Center for Integrated Protein Science Munich (CIPS^M), Ludwig Maximilians University Munich, Planegg-Martinsried and [§]Gene Center Munich, Ludwig Maximilians University Munich, Munich, Germany.

*Corresponding authors:

FS: f.spada@lmu.de; Fax +49 89 2180-74236; Tel. +49 89 2180-74230

HL: h.leonhardt@lmu.de; Fax +49 89 2180-74236; Tel. +49 89 2180-74232

[#]These authors contributed equally to this work

SUPPLEMENTAL INFORMATION FILE

Tables S1-3

Figures S1-8

Supplemental Methods

Table S1. Sequences of DNA oligonucleotides used for preparation of double stranded DNA substrates. M: 5-methylcytosine.

Name	Sequence
CG-up	5'- CTCAACAACCTAACTACCATCCGGACCAGAAGAGTCATCATGG -3'
MG-up	5'- CTCAACAACCTAACTACCATCMGGACCAGAAGAGTCATCATGG -3'
noCG-up	5'- CTCAACAACCTAACTACCATCTGGACCAGAAGAGTCATCATGG -3'
Fill-In-550	5'- ATTO550-CCATGATGACTCTTCTGGTC -3'
Fill-In-590	5'- ATTO590-CCATGATGACTCTTCTGGTC -3'
Fill-In-647N	5'- ATTO647N-CCATGATGACTCTTCTGGTC -3'
Fill-In-700	5'- ATTO700-CCATGATGACTCTTCTGGTC -3'

Table S2. DNA substrates used for the *in vitro* DNA binding and trapping assays.

Name	CpG site	Label	Oligo I	Oligo II	dCTP reaction	Purpose
noCGB 700	no CpG site	700	noCG-up	Fill-In-700	dCTP	Binding
UMB 550	unmethylated	550	CG-up	Fill-In-550	dCTP	Binding
UMB 590		590		Fill-In-590		
UMB 647N		647N		Fill-In-647N		
UMB 700		700		Fill-In-700		
UMT 550	hemimethylated	550	MG-up	Fill-In-550	5-aza-dCTP	Trapping
HMB 590		590		Fill-In-590	dCTP	Binding
HMB 647N		647N		Fill-In-647N		
HMT 550		550		Fill-In-550	5-aza-dCTP	Trapping
HMT 647N	fully methylated	647N	MG-up	Fill-In-647N	5methyl dCTP	Binding
FMB 647N		647N		Fill-In-647N		

Table S3. Primers used for pyrosequencing. Each primer is biotinylated at the 5' end.

Name	Sequence
skeletal α-actin-1	5'- AGTTGGGGATATTTTTTATA -3'
skeletal α-actin-1b	5'- TTTTGGTTAGTGTAGGAGAT -3'
skeletal α-actin-2	5'- TGGGAAGGGTAGTAATATTT -3'
H19-1	5'- ATAGTTATTGTTTATAGTTT -3'
H19-2	5'- AGGAATATGTTATATTTAT -3'
IAP LTR-1	5'- CCCTAATTAACCTACAACCCA -3'
IAP LTR-2	5'- TGTAGTTAATTAGGGAGTGA -3'
Major Satellite-1	5'- AAAATGAGAAATATTTATTTG -3'
Major Satellite-2	5'- GAGAAATATATACTTTAGGA -3'

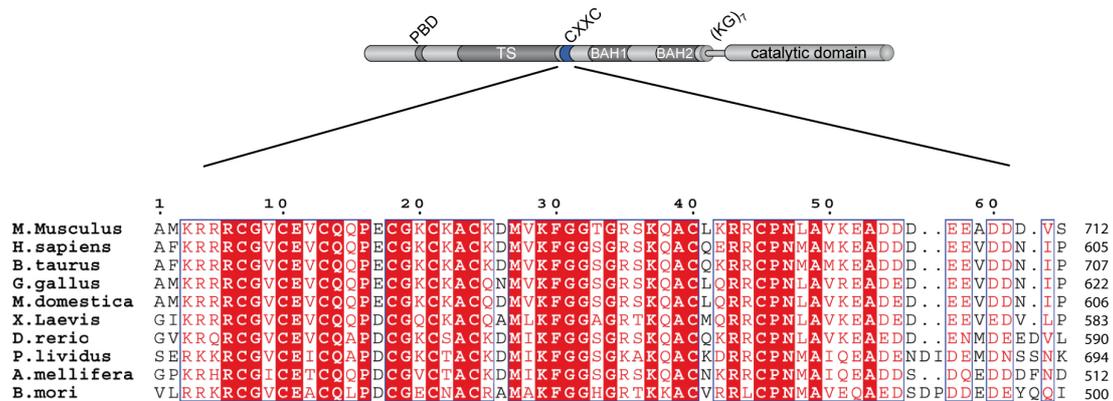


Figure S1. Dnmt1 domain structure and alignment of Dnmt1 CXXC domains from different species. Numbers on the right side indicate the position of the last amino acid in each sequence. PBD: PCNA binding domain; TS: targeting sequence; CXXC: CXXC-type zinc finger domain; BAH1 and 2: bromo-adjacent homology domain; (KG)₇: seven lysine-glycine repeats. Absolutely conserved residues are highlighted in red. Positions with residues in red face share 70% similarity as calculated with the Risler algorithm {Mohseni-Zadeh, 2004 #133}. The alignment was generated with ClustalW2 and displayed with ESPript 2.2. GenBank accession numbers are: *Mus musculus*: NP_034196; *Homo sapiens*: NP_001124295; *Bos taurus*: NP_872592; *Monodelphis domestica*: NP_001028141; *Gallus gallus*: NP_996835; *Xenopus laevis*: NP_001084021; *Danio rerio*: NP_571264; *Paracentrotus lividus*: Q27746 (Swiss Prot); *Apis mellifera*: NP_001164522 (Dnmt1a); *Bombyx mori*: NP_001036980.

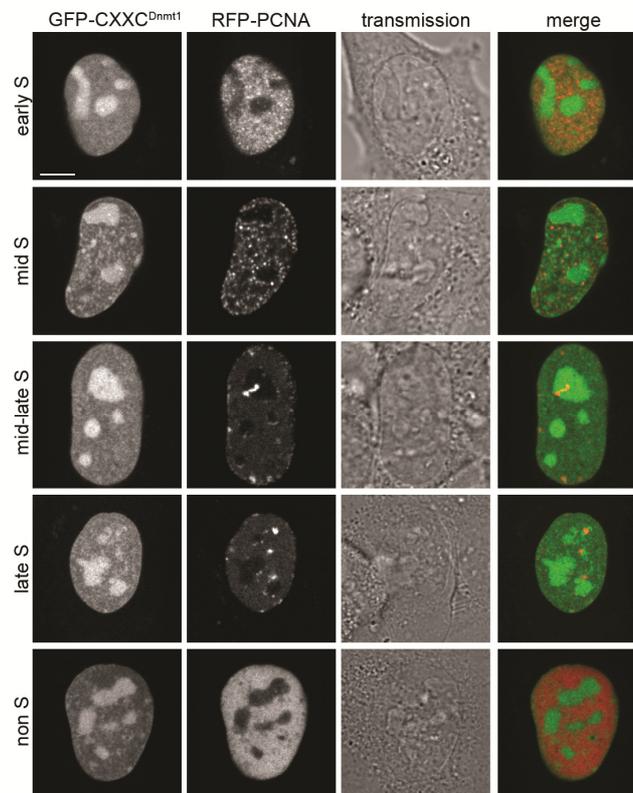


Figure S2. The cellular localization of GFP-CXXC^{Dnmt1} is independent of cell cycle stage. Live images of C2C12 mouse myoblasts cotransfected with expression constructs for GFP-CXXC^{Dnmt1} and RFP PCNA. The latter served for identification of the cell cycle stage. Scale bar: 5 μ m.

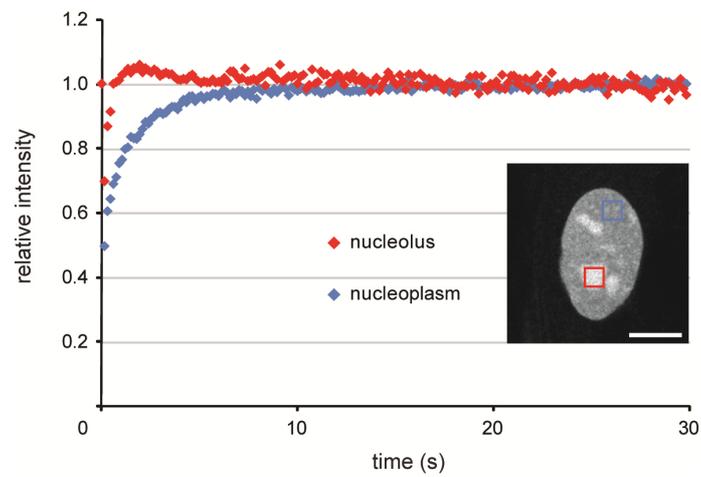


Figure S3. Differential mobility of GFP-CXXC^{Dnmt1} in nucleoli and nucleoplasm of mouse C2C12 myoblasts measured by FRAP analysis. Identical regions of interest over the nucleoplasm or nucleoli (as exemplified in the inset) were bleached and recovery curves were recorded over 30 seconds. GFP-CXXC^{Dnmt1} kinetics is faster in nucleoli than in the nucleus, which indicates more transient (possibly unspecific) binding in the former than in the latter. Scale bar: 5 μ m.

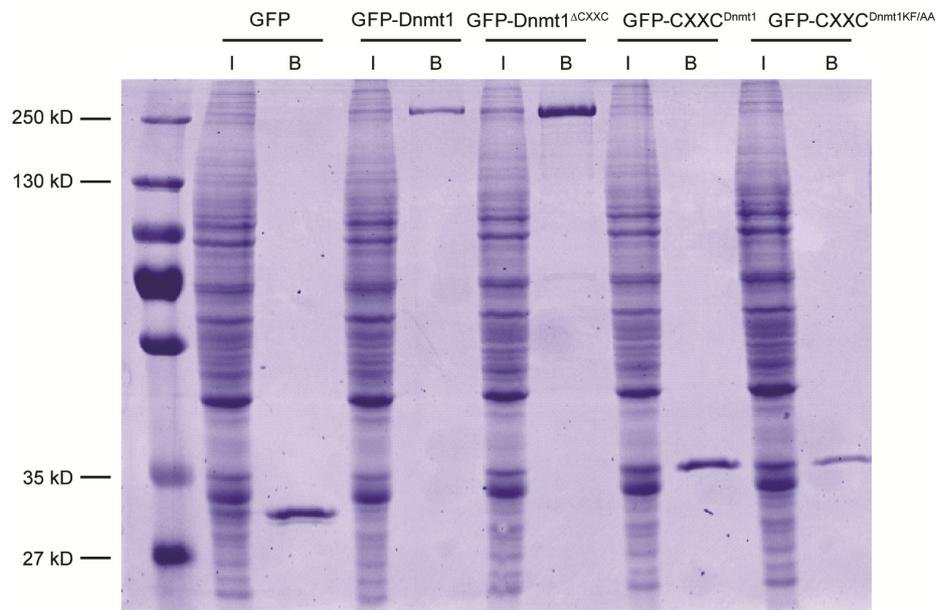


Figure S4. GFP fusion pulldowns from transiently transfected HEK293T cells using the GFP-trap. Shown is a SDS polyacrylamide gel stained with coomassie blue. I = input (1%); B = bound (10%).

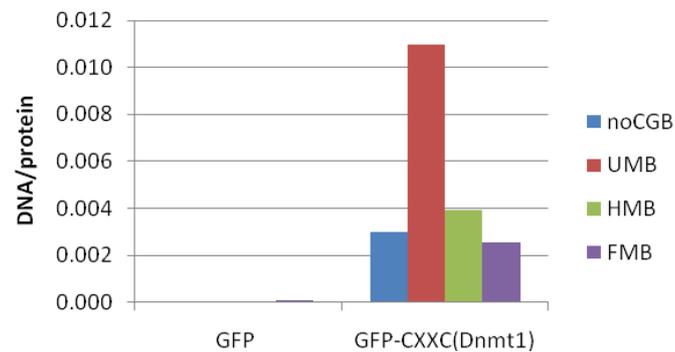


Figure S5. The CXXC domain of Dnmt1 preferentially binds unmethylated CpG sites. GFP and GFP-CXXC^{Dnmt1} purified from transiently transfected HEK293T cells with the GFP trap were challenged with fluorescent DNA substrates containing no CpG site or one central un-, hemi- or fully methylated CpG site in direct competition (noCGB, UMB, HMB and FMB, respectively) as in Figure 2C, except that a five-fold higher concentration (625 nM) of each substrate was used.

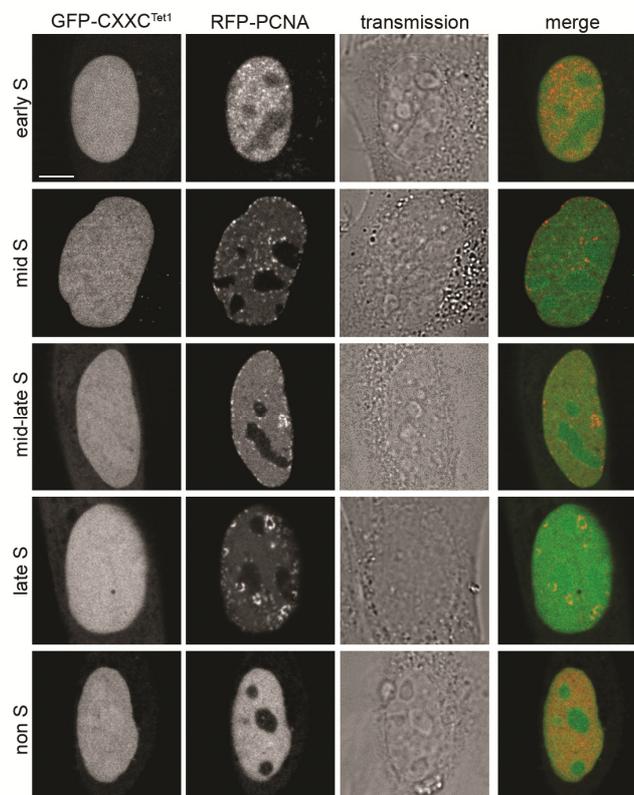


Figure S6. The cellular localization of GFP-CXXC^{Tet1} is independent of cell cycle stage. Live images of C2C12 mouse myoblasts cotransfected with expression constructs for GFP-CXXC^{Tet1} and RFP PCNA. The latter served for identification of the cell cycle stage. Scale bar: 5 μ m.

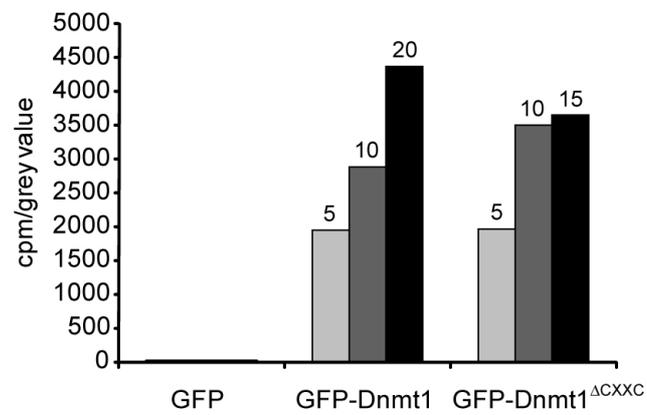


Figure S7. Radioactive methyltransferase activity assay for GFP Dnmt1 and GFP-Dnmt1^{ΔCXXC}. The transfer of [3H]-methyl groups to poly(dI•dC)-poly(dI•dC) substrate was measured for increasing volumes of GFP fusion proteins immunopurified from transiently transfected HEK293T cells. Counts per minute (cpm) were normalized to the relative protein concentration as determined by SDS-PAGE analysis. GFP was used as negative control. Numbers above the bars indicate the volume (μl) of protein solution added.

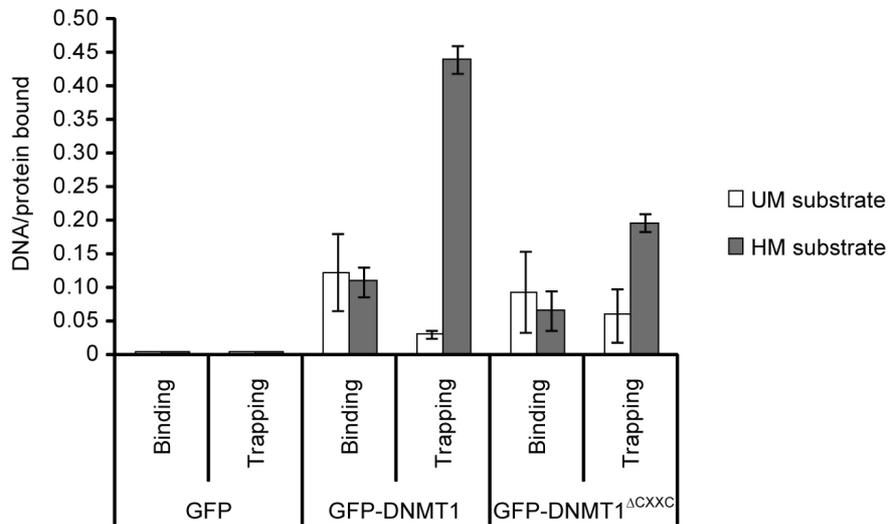


Figure S8. Competitive DNA binding and trapping assays for human GFP-DNMT1 and GFP-DNMT1^{ΔCXXC}. GFP, GFP-DNMT1 and GFP-DNMT1^{ΔCXXC} were purified from transfected HEK293T cells using the GFP-trap and incubated with fluorescent DNA substrates containing one central unmethylated (UM) or hemimethylated (HM) CpG site in direct competition. Both substrates contained either dC (binding) or 5 aza dC (trapping) on the strand opposite to the differentially methylated one. The comparison of binding and trapping ratios reflects irreversible covalent complex formation. Note the reduction in trapping of GFP-DNMT1^{ΔCXXC} relative to GFP-DNMT1 by the hemimethylated substrate. Shown are mean values and standard deviation of DNA/protein ratios from two independent experiments.

SUPPLEMENTARY METHODS***In vitro* methyltransferase activity assay**

Eight milligrams of His-tagged GFP-binding protein (GBP; Chromotek) were coupled to 1ml Ni-NTA agarose beads (Qiagen) by incubating for 2 h at 4°C in PBS and unbound protein was washed out twice with PBS. Extracts of HEK293T cells expressing GFP or a GFP fusions were prepared in 200 µl lysis buffer II (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.5 % Tween-20, 2 mM MgCl₂, 1 mg/ml DNaseI, 2 mM PMSF, 1X mammalian protease inhibitor mix). After centrifugation, supernatants were diluted to 500 µl with immunoprecipitation buffer II (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.05 % Tween-20) and precleared by incubation with 25 µl of equilibrated Ni-NTA agarose beads for 30 min at 4°C followed by centrifugation. Precleared extracts were then incubated with 40 µg of His-tagged GFP-trap coupled to Ni-NTA beads for 2 h at 4°C with constant mixing. GFP or GFP fusions were pulled down by centrifugation at 540 g. After washing twice with wash buffer II (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05 % Tween-20), complexes were eluted with 60 µl of elution buffer (10 mM Tris pH 7.5, 100 mM KCl, 1 mM EDTA, 1 mM DTT, 250 mM imidazole) for 10 min at 25°C with constant mixing. 10 µl aliquots of all eluates were subjected to western blot analysis using mouse or rat monoclonal antibodies to GFP (Roche and Chromotek, respectively) and quantified by densitometry. Indicated volumes of eluate were incubated with 1 µg of poly(dI·dC)-poly(dI·dC) substrate (Sigma), 0.5 µg/µl of BSA and 1 µCi of S-adenosyl-[³H-methyl]-methionine in 50 µl of trapping buffer (10 mM Tris pH 7.5, 100 mM KCl, 1 mM EDTA, 1 mM DTT) for 60 min at 37°C. 15 µl of each sample were spotted onto blotting paper and the DNA was precipitated with ice cold 5 % TCA. After washing twice with 5% TCA and once with cold 70 % ethanol, paper filters were air dried and analyzed by scintillation in 4 ml scintillation cocktail (Rotiszint[®] eco plus, Roth) for 5 min.

2.4 Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia

- 2 He Y, Wertheim JA, Xu L, Miller JP, Karnell FG, Choi JK *et al.* The coiled-coil domain and Tyr177 of bcr are required to induce a murine chronic myelogenous leukemia-like disease by bcr/abl. *Blood* 2002; **99**: 2957–2968.
- 3 Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* 1985; **315**: 550–554.
- 4 Golub TR, Barker GF, Bohlander SK, Hiebert SW, Ward DC, Bray-Ward P *et al.* Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci USA* 1995; **92**: 4917–4921.
- 5 De Braekeleer E, Douet-Guilbert N, Le Bris MJ, Berthou C, Morel F, De Braekeleer M. A new partner gene fused to ABL1 in a t(1;9)(q24;q34)-associated B-cell acute lymphoblastic leukemia. *Leukemia* 2007; **21**: 2220–2221.
- 6 De Keersmaecker K, Graux C, Odero MD, Mentens N, Somers R, Maertens J *et al.* Fusion of EML1 to ABL1 in T-cell acute lymphoblastic leukemia with cryptic t(9;14)(q34;q32). *Blood* 2005; **105**: 4849–4852.
- 7 Graux C, Cools J, Melotte C, Quentmeier H, Ferrando A, Levine R *et al.* Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. *Nat Genet* 2004; **36**: 1084–1089.
- 8 Soler G, Radford-Weiss I, Ben-Abdelali R, Mahlaoui N, Ponceau JF, Macintyre EA *et al.* Fusion of ZMIZ1 to ABL1 in a B-cell acute lymphoblastic leukaemia with a t(9;10)(q34;q22.3) translocation. *Leukemia* 2008; **22**: 1278–1280.
- 9 Hidalgo-Curtis C, Chase A, Drachenberg M, Roberts MW, Finkelstein JZ, Mould S *et al.* The t(1;9)(p34;q34) and t(8;12)(p11;q15) fuse pre-mRNA processing proteins SFPQ (PSF) and CPSF6 to ABL and FGFR1. *Genes Chromosomes Cancer* 2008; **47**: 379–385.
- 10 De Braekeleer E, Douet-Guilbert N, Rowe D, Bown N, Morel F, Berthou C *et al.* ABL1 fusion genes in hematological malignancies: a review. *Eur J Haematol* 2011; **86**: 361–371.
- 11 Damen JE, Liu L, Rosten P, Humphries RK, Jefferson AB, Majerus PW *et al.* The 145-kDa protein induced to associate with Shc by multiple cytokines is an inositol tetrakisphosphate and phosphatidylinositol 3,4,5-triphosphate 5-phosphatase. *Proc Natl Acad Sci USA* 1996; **93**: 1689–1693.
- 12 Helgason CD, Damen JE, Rosten P, Grewal R, Sorensen P, Chappel SM *et al.* Targeted disruption of SHIP leads to hemopoietic perturbations, lung pathology, and a shortened life span. *Genes Dev* 1998; **12**: 1610–1620.
- 13 Luo JM, Yoshida H, Komura S, Ohishi N, Pan L, Shigeno K *et al.* Possible dominant-negative mutation of the SHIP gene in acute myeloid leukemia. *Leukemia* 2003; **17**: 1–8.
- 14 Gilby DC, Goodeve AC, Winship PR, Valk PJ, Delwel R, Reilly JT. Gene structure, expression profiling and mutation analysis of the tumour suppressor SHIP1 in Caucasian acute myeloid leukaemia. *Leukemia* 2007; **21**: 2390–2393.
- 15 Lo TC, Barnhill LM, Kim Y, Nakae EA, Yu AL, Diccianni MB. Inactivation of SHIP1 in T-cell acute lymphoblastic leukemia due to mutation and extensive alternative splicing. *Leuk Res* 2009; **33**: 1562–1566.

Supplementary Information accompanies the paper on the Leukemia website (<http://www.nature.com/leu>)

Genomic 5-hydroxymethylcytosine levels correlate with *TET2* mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia

Leukemia (2011) **25**, 1649–1652; doi:10.1038/leu.2011.134; published online 31 May 2011

The TET proteins are 2-oxoglutarate- and Fe(II)-dependent oxygenase catalyzing the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC).¹ The *TET1* (ten–eleven translocation 1) gene was originally identified as an *MLL* fusion partner in rare cases of acute myeloid leukemia (AML) with a t(10;11)(q22;q23).^{2,3} The definite function of 5hmC still remains elusive, but hydroxylation of 5mC has been suggested to be involved in the process of DNA demethylation. This suggests a possible role of 5hmC in epigenetic gene regulation. Recently, hemizygous deletions and mutations of *TET2* were found in a wide range of myeloid malignancies, including myelodysplastic syndrome (MDS), myeloproliferative disorders such as chronic myelomonocytic leukemia (CMML) and in secondary AML (sAML).^{4–6} Interestingly, very recently, myeloid neoplasias harboring heterozygous *TET2* mutations were shown to have decreased levels of 5hmC.⁷

To explore the relationship among *TET2* mutations, global gene expression profiles (GEPs) and 5hmC levels, we measured 5hmC levels in the genomic DNA in a series of 30 sAML patients using a novel assay method employing β -glucosyltransferase from bacteriophage T4.⁸ In addition to the *TET2* mutational status, we screened for *IDH1/2* mutations (see Supplementary Material).

All patients had developed AML after a preceding MDS, refractory anemia with excess blast or CMML phase. The average age at diagnosis was 70.8 years. Eight patients had a normal karyotype (nk), eight patients had a complex aberrant

karyotype (ak) with more than three chromosomal aberrations and the remainder of the patients had one or two chromosomal aberrations, which are typical of MDS, that is, del(5)(q) (two patients), +8 (five patients), –7 or del(7)(q) (six patients; see Supplementary Table 1). We sequenced the complete coding region of *TET2* in all 30 patients. In all, 7 of the 30 patients (23.3%) had *TET2* mutations. One patient (no. 16) had single-nucleotide deletions in both alleles of *TET2* at amino-acid positions 218 and 519, which caused truncation of the protein after 250 and 533 amino acids, respectively. Two patients (nos. 15 and 26) had nonsense mutations at positions 1216 and 1274, and four patients (nos. 7, 14, 20 and 30) had missense mutations (see Table 1). All the *TET2* mutations (except for patient no. 16) were heterozygous. We did not detect any deletions in patients with *TET2* mutations using a commercially available fluorescence in situ hybridization probe for the *TET2* locus. There was no significant association between *TET2* mutational status and any particular chromosomal abnormality. Although there was a trend toward a higher frequency of *TET2* mutations in patients with a nk in comparison with patients with an ak (50% (4 out of 8 nk patients) versus 13.6% (3 out of 22 ak patients), χ^2 -test: $P=0.17$).

The analysis of the 5hmC levels of the patients' DNA using the β -glucosyltransferase assay revealed a 5hmC content of the DNA, ranging from 0.006 to 0.054%. This range of 5hmC levels, about 9- to 14-fold difference between the lowest and highest measurements, agrees well with the measurements reported by Ko *et al.*,⁷ although very different patient groups were assayed. In contrast to the

Table 1 Overview of the clinical diagnosis, *TET2* and *IDH2* mutational status, as well as 5hmC levels of the 30 sAML patients analyzed

No.	Diagnosis	<i>TET2</i> mutations	<i>IDH2</i> mutations	5hmC (%)	Standard deviation of 5hmC levels
1	sAML after MDS	Wt	Wt	0.02550	0.00444
2	sAML M2 after MDS RA	Wt	Wt	0.02920	0.00185
3	sAML M4 after MDS	Wt	Wt	0.01804	0.00315
4	sAML M4 after MDS	Wt	Wt	0.01762	0.00321
5	sAML M2 after MDS	Wt	Wt	0.05458	0.00174
6	sAML M2 after MDS	Wt	Wt	0.03486	0.00186
7	sAML M4 after MDS	p.Asn1753_Tyr1766dup c.5256_5297dup42	p.Arg140Gln c.419G>A	0.01208	0.00412
8	sAML M2 after MDS	Wt	Wt	0.01633	0.00384
9	sAML M4 after 5q syndrome	Wt	Wt	0.03204	0.00189
10	sAML M2 after MDS	Wt	Wt	0.01636	0.00312
11	sAML M2 after MDS	Wt	Wt	0.02774	0.00128
12	sAML M2 after MDS	Wt	Wt	0.04758	0.00087
13	sAML M4 after MDS	Wt	Wt	0.03065	0.00442
14	sAML M0 after CMML	p.Glu1144Lys c.3430G>A	Wt	0.01062	0.00162
15	sAML M1 after CMML	p.Arg1216* c.3646C>T	Wt	0.01010	0.00239
16	sAML M2 after CMML	p.Val218Trpfs*32 c.651delC p.Phe519Leufs*14 c.1557delT	Wt	0.00698	0.00224
17	sAML M2 after MDS	Wt	p.Arg140Gln c.419G>A	0.00630	0.00056
18	sAML M2 after MPS	Wt	p.Arg140Gln c.419G>A	0.02075	0.00385
19	MDS RAEB-2 transformation to AML M2	Wt	Wt	0.01505	0.00325
20	sAML M2 after MDS	p.Thr1270Pro c.3808A>C	Wt	0.01655	0.00255
21	MDS RAEB-2, borderline AML M6	Wt	Wt	0.04156	0.00334
22	sAML M4 after MDS	Wt	Wt	0.03973	0.00386
23	sAML M2 after MDS	Wt	Wt	0.03396	0.00538
24	sAML after 5q syndrome	Wt	Wt	0.02885	0.00230
25	sAML M0 after MDS	Wt	Wt	0.02051	0.00132
26	sAML M0 after OMF	p.Gln1274* c.3820C>T	Wt	0.02618	0.00432
27	MDS RAEB-2, borderline CMML-2 or AML M4	Wt	Wt	0.00648	0.00075
28	sAML after MDS	Wt	Wt	0.00973	0.00202
29	MDS RAEB-2 transformation to AML	Wt	Wt	0.00758	0.00108
30	MDS RAEB-2, borderline AML M2	p.Ile1873Thr c.5618T>C	Wt	0.00699	0.00081

Abbreviations: AML, acute myeloid leukemia; cDNA, complementary DNA; CMML, chronic myelomonocytic leukemia; 5hmC, 5-hydroxymethylcytosine; MDS, myelodysplastic syndrome; MPS, myeloproliferative disorder; OMF, osteomyelofibrosis; RA, refractory anemia; RAEB, RA with excess blast; sAML, secondary AML; Wt, wild type.
Reference sequences: *TET2* protein: NP_001120680.1; *TET2* cDNA: NM_001127208.1; *IDH2* protein: NP_002159.2; and *IDH2* cDNA: NM_002168.2.

results presented by Ko *et al.*,⁷ we did not observe a clear bimodal distribution of 5hmC values. This could be due to the smaller sample number in our series (30 versus 88) and to different patient characteristics in the two studies (mainly sAML in our study versus a broader range of myeloid malignancies in the study of Ko *et al.*⁷).

When we compared the presence of *TET2* mutation with the 5hmC levels, we found a significant clustering of patients with *TET2* mutations in the lower half of 5hmC levels (Figure 1 a). All but one patient with a *TET2* mutation belonged to the group of the 15 patients with the lowest 5hmC levels (χ^2 -test: $P=0.03$). This corresponded to 5hmC levels of <0.02%. Only one patient with a *TET2* mutation, no. 26) had 5hmC levels of >0.020%. These results agree well with the observation of Ko *et al.*⁷ Interestingly, there were several patients with very low 5hmC levels that did not have a *TET2* mutation. As it was reported recently that *IDH1/2* mutations can impair *TET2* function, which might also correlate with low 5hmC levels,^{9,10} we determined the mutational status of the *IDH1* and *IDH2* genes in our patients. Only three patients (nos. 7, 17 and 18)

had mutations at amino acid R140 in *IDH2* (Table 1, Figure 1 a). No mutations in *IDH1* were detected. Interestingly, one patient (no. 7) had both mutations in *IDH2* and in *TET2*. In a much larger series of patients reported by Figueroa *et al.*,¹⁰ no patient with both a *TET2* and an *IDH1/2* mutation was discovered. Patient no. 17 who had an *IDH2* mutation had the lowest 5hmC levels in our series. Patient no. 18 had intermediate 5hmC levels. However, there are still eight patients in the lower half of the 5hmC level range who have neither a *TET2* nor an *IDH1/2* mutation. There was no correlation between *TET2* expression levels and 5hmC levels in our patients (data not shown).

To determine the impact of *TET2* mutations and 5hmC levels on cellular function, we obtained GEPs from 28 patients (all except patient nos. 9 and 22) and performed two comparisons for differential gene expression: (1) patients with *TET2* mutations (7 patients) versus patients without *TET2* mutations (21 patients) and (2) the 7 patients with the lowest versus the 7 patients with the highest 5hmC levels. The top differentially expressed genes in the high versus low 5hmC level

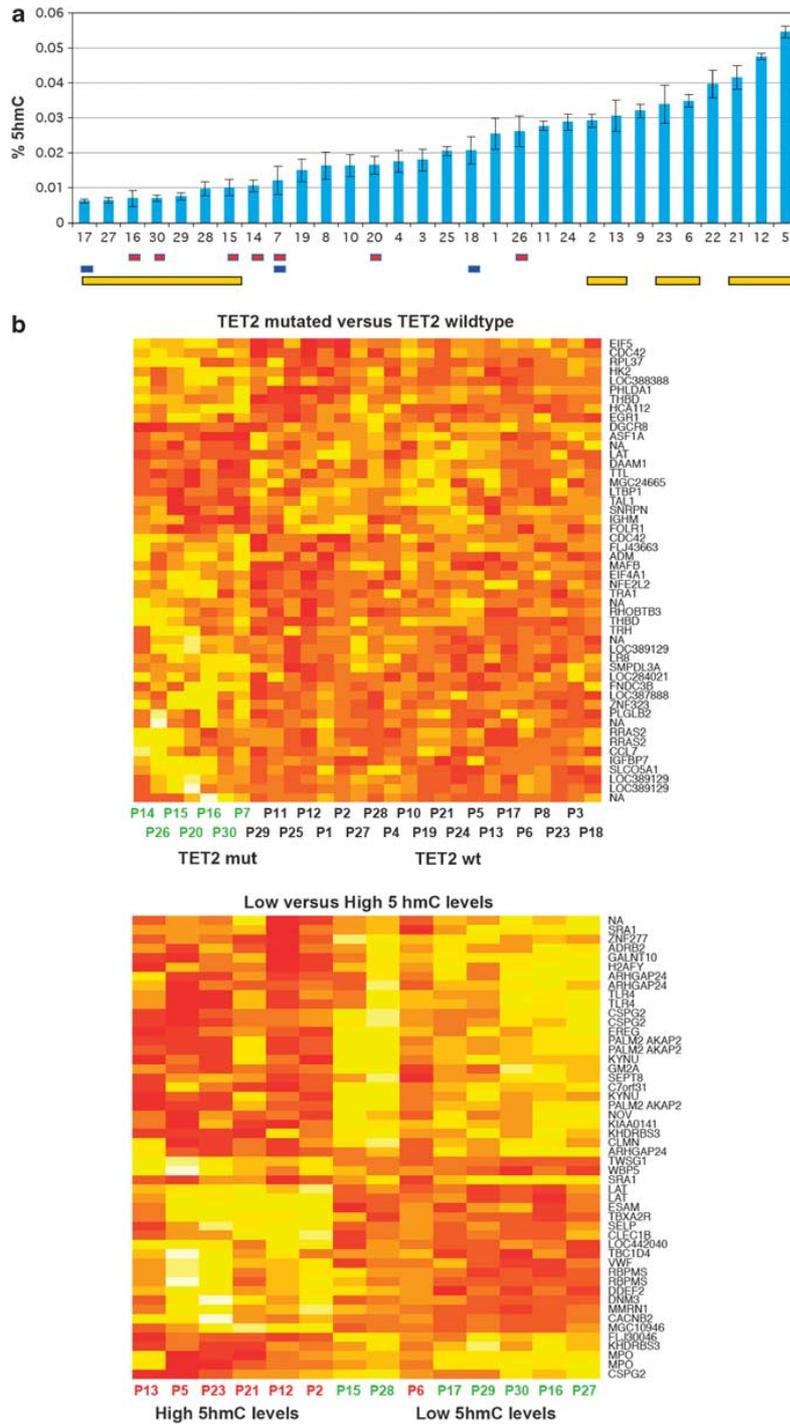


Figure 1 (a) Bar graph of 5hmC (%) levels of the 30 patients of this study. Patients with a TET2 mutation or an IDH2 mutation are indicated with a red or a blue rectangle, respectively. The yellow bars indicate the samples that were used for the differential gene expression analysis in the comparison of low versus high 5hmC levels. Note that the patients 9 and 22 did not have gene expression data of sufficient quality to be included in this analysis. (b) Heatmaps of the 50 most significantly differentially expressed probe sets comparing samples with TET2 mutations ($n=7$) versus TET2 wild type ($N=21$; top) and samples with low ($n=7$) and high ($n=7$) 5hmC levels (bottom).

comparison had a lower *P*-value and had a higher degree of deregulation than the differentially expressed genes from the comparison *TET2* mutated versus wild type (Figure 1 b; Supplementary Figure 1 and Supplementary Tables 2 and 3). These results indicate that 5hmC levels are most likely a more relevant measurement to define biologically distinct secondary leukemia subtypes than the *TET2* (or *IDH1/2*) mutational status. The fact that in some patient samples with low 5hmC levels neither *TET2* nor *IDH1/2* mutations could be identified suggests that additional genes might be directly or indirectly involved in the regulation of 5hmC levels. To further elucidate the regulation of 5hmC levels and their role in leukemogenesis, larger groups of sAML as well as *de novo* AML patients need to be studied.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

SKB, KS and HL are supported by grants from the Deutsche Forschungsgemeinschaft (SFB 684 and SPP 1463). SKB is supported by a grant from the German Ministry of Education and Research (BMBF) National genome research network (NGFNplus; PKL-01GS0876-6) and by institutional funding from the Helmholtz Zentrum Munich, German Research Center for Environmental Health. We thank Natalia Huk for technical assistance and Alexander Kohlmann, Torsten Haferlach and Claudia Haferlach for primary gene expression data and cytogenetic data.

Author Contributions

NK and FS designed and performed the mutation screening with the help of AD and BK, and wrote the manuscript. SB, AS and HL designed and performed the 5hmC measurements and wrote the manuscript. HL supervised the project. PMK and SS performed cytogenetics and fluorescence in situ hybridization analysis. TH and MM analyzed the GEPs. KS designed experiments and wrote the manuscript. SKB designed experiments, analyzed the data, supervised the project and wrote the manuscript.

N Konstandin^{1,2,4}, S Bultmann^{3,4}, A Szwagierczak³, A Dufour¹, B Ksienzyk^{1,2}, F Schneider^{1,2}, T Herold¹, M Mulaw², PM Kakadia^{1,2}, S Schneider¹, K Spiekermann^{1,2}, H Leonhardt^{3,4} and SK Bohlander^{1,2,4}

Supplementary Information accompanies the paper on the Leukemia website (<http://www.nature.com/leu>)

Immunophenotype-defined sub-populations are common at diagnosis in childhood B-cell precursor acute lymphoblastic leukemia

Leukemia (2011) **25**, 1652–1657; doi:10.1038/leu.2011.136; published online 17 June 2011

Neoplasms often display significant heterogeneity in morphology, gene expression (including cell surface markers), genetic

¹Laboratory of Leukemia Diagnostics, Department of Medicine III, University of Munich Hospital, Munich, Germany;

²Helmholtz Zentrum Munich, National Research Center for Environmental Health, Clinical Cooperative Group 'Leukemia', Munich, Germany and

³Department of Biology II, Center for Integrated Protein Science, Ludwig Maximilians University Munich, Planegg-Martinsried, Germany
E-mail: stefan.bohlander@med.uni-muenchen.de or H.Leonhardt@lmu.de

⁴These authors contributed equally to this work.

References

- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009; **324**: 930–935.
- Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. 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 Res* 2002; **62**: 4075–4080.
- Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 2003; **17**: 637–641.
- Mohamedali AM, Smith AE, Gaken J, Lea NC, Mian SA, Westwood NB *et al.* Novel TET2 mutations associated with UPD4q24 in myelodysplastic syndrome. *J Clin Oncol* 2009; **27**: 4002–4006.
- Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* 2009; **41**: 838–842.
- Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Massé A *et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* 2009; **360**: 2289–2301.
- Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 2010; **468**: 839–843.
- Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 2010; **38**: e181.
- Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* 2011; **19**: 17–30.
- Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 2010; **18**: 553–567.

aberrations, cell proliferation kinetics and response to therapy.¹ Heterogeneity in antigen marker expression is well known in acute myeloid leukemia (AML);² however, has rarely been studied in acute lymphoblastic leukemia (ALL). In ALL, shifts in immunophenotypic and genetic profiles can occur between

Supplemental Information:

Methods:

Patient Samples

Patient samples were obtained from the Laboratory for Leukemia Diagnostics, University of Munich Hospital Grosshadern. The samples were received for routine cytogenetics and molecular genetics diagnostics. The institutional review board of the University Hospital Munich has approved the use of left over diagnostic samples for this research project in an anonymous fashion.

Cytogenetics and FISH

Each sample was routinely karyotyped according to standard protocols (GTG banding). All patients were screened for deletions of the *TET2* locus using a fluorescence in situ hybridization (FISH) probe covering the *TET2* gene from marker RH43141 to RH69608 (XL TET2, MetaSystems, Altlußheim, Germany).

Mutation Analysis

Sample Preparation

Cells from diagnostic bone marrow and peripheral blood samples were enriched by Ficoll density gradient centrifugation and thereafter lysed by RLT buffer (Quiagen, Hilden, Germany) and stored at -80 °C.

Genomic DNA was extracted from patient samples with secondary AML using QIAamp DNA mini Kit (Quiagen, Hilden, Germany).

Sequencing of *TET2*

Sanger sequencing analysis was performed on PCR-amplified genomic DNA fragments spanning the entire coding region of *TET2* isoform a (primers are shown below).

For the PCR of *TET2*, 50 ng of genomic DNA was used and amplified with Taq PCR Master Mix Kit (Quiagen, Hilden, Germany) in a 25 µl reaction using the following program: 95°C for 5 min, 35 cycles at 95°C for 30 sec, 56°C for 30 sec and 72°C for 1 min with a terminal elongation at 72 °C for 5 min.

Purified PCR fragments were sequenced using the Big Dye Terminator v1.1 Cycle Sequencing Kit on 3100 Avant Genetic Analyzer (Applied Biosystems, Darmstadt, Germany).

All fragments were sequenced bidirectional and sequence variations were confirmed again by sequencing from the original DNA.

Sequencing of *IDH1/2*

Sanger sequencing was performed on PCR- amplified genomic DNA fragments including the genomic region of *IDH1/2* containing the mutational hotspots R132 and SNP rs 11554137 in the *IDH1* exon 4, and R140 and R172 in the *IDH2* exon 4.

For *IDH1* the following primers were used:

IDH1-F 5'-CATAATGTTGGCGTCAAATGTG-3' , *IDH1-R* 5'-ACATGCAAAATCACATTATTGCC-3'

For *IDH2* the following primers were used:

IDH2-F 5'-GTTCAAGCTGAAGAAGATGTGG-3' , *IDH2-R* 5'-TGTGGCCTTGTACTGCAGAG-3'

For the PCR of *IDH1* and *IDH2*, 50 ng of genomic DNA was used and amplified with Taq PCR Master Mix Kit (Qiagen, Hilden, Germany) in a 25 µl reaction using the following program: 94°C for 5 min, 35 cycles at 94°C for 1min, 56°C for 1min and 72°C for 1 min with a terminal elongation at 72 °C for 10 min.

Sequencing was performed as described above

TET2 primers:

cDNA sequence: uc003hxx.2

Genomic sequence: hg18_chr4:106286391-106421407

NAME	LEFT_PRIMER	RIGHT_PRIMER	PRODUCT_SIZE
Exon3_1	CAGTTTGCTATGTCTAGGTATTCCG	TTTCCCCTCCTGCTCATTTC	740
Exon3_2	ATGTCTCCGATTTGAGTGATAAG	CCCTGGATGTTATTTTCTGC	607
Exon3_3	TGCAATGCTAAATACCTGTTC	TTCTGACATFGGTCTGTTTTCTC	607
Exon3_4	CATCTACACATGTATGCAGCCC	TGGTCTGTTTTGGAGAAGTGC	607
Exon3_5	CCAAGTTGAAATGAATCAAGGG	TGTATTGTTTTGAACAAGAACCTG	613
Exon3_6	CAAAATCAAGCGAGTTCGAG	ATACAGGCATGTGGCTTGC	604
Exon3_7	GAAGCAAGAACAGCAGCAAAC	TCTGAAGATAAATTTGCTAATTTCTGG	600
Exon4	GCCCTTAATGTGTAGTTGGGG	TGCTTTGTGTGTGAAGGCTG	271
Exon5	TGCCCTTTGAATTCATTTGC	GGGTAACCAATTCAGGG	234
Exon6	TGCAAGTGACCCTTGTTTTG	CAACCAAAGATTGGGCTTTC	342
Exon7	CAGCTGCACAGCCTATATAATG	TCACTTCATCTAAGCTAATGAATTC	279
Exon8	GGGATTCAAAATGTAAGGGG	TGCAGTGGTTTCAACAATTAAG	323
Exon9	TGTCATTCCATTTTGTCTGG	TCTGCTCCTCAACATGAGATG	691
Exon10	ACACACACGTTTTCTTTGGG	cagaacttacaaGTTGATGGGG	511
Exon11_1	CCTACATTTAAGTATCCTCACTAGCC	AACTGCTGAAACCATCTCCC	676
Exon11_2	AGGTATCCAAGCCAAGACCC	GGATCCAGAAAGCTCTGCTC	593
Exon11_3	GTCAGGAAAAGCAGCCATTG	CTGACAGGTTGGTTGTGGTC	594

5hmC Assay

For quantification of 5-hydroxymethylcytosine (hmC) in the genomic DNA samples a recently published method was applied (1). This method utilizes the beta -glucosyltransferase (β -gt) from the bacteriophage T4 which specifically transfers the glucose moiety from UDP-glucose to hmC. By measuring the incorporation of radioactively labeled glucose into genomic DNA the hmC content of the DNA can be calculated.

The bacteriophage T4 protein β -gt was expressed and purified as described before (1). In brief, BL21(DE3) *E. coli* cells carrying the expression construct were grown at 37°C until $A_{600} = 0.6-0.7$ and induced with isopropyl β -d-thiogalactopyranoside for 16 h at 20°C. Cells were lysed and the protein was purified using a nickel-nitrilotriacetic acid column (QIAGEN) followed by gel filtration. Fractions containing the β -gt peak were pooled and applied to a

ResourceQ anion exchange column (GE Healthcare) in order to eliminate residual contaminants, resulting in pure β -gt in the flowthrough.

Reference PCR fragments containing a known amount of 5hmC were produced as described before (1).

Genomic DNA samples were sheared to 500–1500_bp fragments by sonication to reduce the viscosity and improve homogeneity.

Reactions contained 150 mM NaCl, 20 mM Tris, pH 8.0, 25 mM CaCl₂, 1 mM DTT, 0.86 nM UDP-[³H]glucose (glucose-6-³H; 60 Ci/mmol; Hartmann Analytic GmbH), 500 ng of DNA substrate and 100 nM recombinant β -gt in a total volume of 50 μ l. Reactions were incubated for 20 min at room temperature and terminated by heating at 65°C for 10 min. Unincorporated nucleotides were removed by using the NucleoSpin® Extract II kit (Macherey-Nagel) according to manufacturers protocol. Remaining radioactivity was measured using a Liquid Scintillation Analyzer Tri-Carb 2100TR (Packard) with quench indicating parameter set on tSIE/AEC (transformed spectral index of the external standard/automatic efficiency control) in 4 ml of Rotiszint Eco Plus scintillation liquid (Roth GmbH) in Snaptwist vials (Simport). Samples were measured for 30 min or until the 2σ value reached 2%. The percentage of hmC per total cytosine was calculated from the incorporation of [³H]glucose using a calibration curve measured with the reference fragment series for every experiment. The percentage of hmC was then corrected for the difference in C abundance between reference fragment (17,5%) and human genome (20,5%).

Gene Expression profiles

For the expression analysis, RNA from patient samples was labeled and hybridized to Affymetrix HG-U133A and HG-U133B or HG-U133plus2 chips as previously described (2). The HG-U133A, HG-U133B and HG-U133plus2 .CEL files were first normalized separately using robust multi-array (RMA) normalization and then combined into one matrix and normalized all together using the empirical Bayes (eB) method as previously described (3). The normalized expression data were analyzed with the twilight program (4) in the R software package (<http://www.r-project.org/>).

Supplemental Figure Legends:

Figure 1:

Top:

Degree of up- and down regulation of the 50 most significantly deregulated probe sets in the comparison TET2 mutated (n = 7) versus TET2 wild type (n = 21). The genes toward the right side are upregulated in TET2 mutated samples. The scale is linear (i.e. 2 denotes a 2 fold higher expression in TET2 mutated samples).

Bottom:

Degree of up- and down regulation of the 50 most significantly deregulated probe sets in the comparison low (n = 7) versus high (n = 7) 5hmC levels. The genes toward the right side are upregulated in samples with low 5hmC levels. The scale is linear (i.e. 2 denotes a 2 fold higher expression in samples with low 5hmC levels).

Supplemental Tables:

Supplemental Table 1:

Additional clinical and karyotype data of the patients.

Supplemental Table 2:

The 50 most significantly deregulated probe sets in the comparison TET2 mutated (n = 7) versus TET2 wild type (n = 21).

Supplemental Table 3:

The 50 most significantly deregulated probe sets in the comparison low (n = 7) versus high (n = 7) 5hmC levels.

References:

1. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* 2010;38:e181.

2. Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC, *et al.* An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood*. 2008;112:4193-201.
3. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118-27.
4. Scheid S, Spang R. twilight; a Bioconductor package for estimating the local false discovery rate. *Bioinformatics*. 2005;21:2921-2.

2.5 Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain

Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain

Carina Frauer^{1,4}*, Thomas Hoffmann^{2,4}*, Sebastian Bultmann^{1,4}, Valentina Casa³, M. Cristina Cardoso³, Iris Antes^{2,4}*, Heinrich Leonhardt^{1,4}*

1 Department of Biology II, Ludwig Maximilians University Munich, Planegg-Martinsried, Germany, **2** Department of Life Sciences, Technical University Munich, Freising-Weihenstephan, Germany, **3** Department of Biology, Technical University Darmstadt, Darmstadt, Germany, **4** Center for Integrated Protein Science Munich (CIPSM), Munich, Germany

Abstract

Recent discovery of 5-hydroxymethylcytosine (5hmC) in genomic DNA raises the question how this sixth base is recognized by cellular proteins. In contrast to the methyl-CpG binding domain (MBD) of MeCP2, we found that the SRA domain of Uhrf1, an essential factor in DNA maintenance methylation, binds 5hmC and 5-methylcytosine containing substrates with similar affinity. Based on the co-crystal structure, we performed molecular dynamics simulations of the SRA:DNA complex with the flipped cytosine base carrying either of these epigenetic modifications. Our data indicate that the SRA binding pocket can accommodate 5hmC and stabilizes the flipped base by hydrogen bond formation with the hydroxyl group.

Citation: Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, et al. (2011) Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. PLoS ONE 6(6): e21306. doi:10.1371/journal.pone.0021306

Editor: Shuang-yong Xu, New England Biolabs, Inc., United States of America

Received: March 18, 2011; **Accepted:** May 25, 2011; **Published:** June 22, 2011

Copyright: © 2011 Frauer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the Nanosystems Initiative Munich (NIM, <http://www.nano-initiative-munich.de>), the Center for NanoSciences (CeNS, <http://www.cens.de>) Munich, and by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB TR5, <http://www.dfg.de>) to HL. CF gratefully acknowledges support by the Elite Network of Bavaria (International Doctorate Program NanoBioTechnology, IDK-NBT, <http://www.cens.de/doctorate-program>) and the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS, <http://www.imprs-ls.de>). SB was supported by the Graduate School for Life Sciences Munich (LSM, <http://www.lsm.bio.lmu.de>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: antes@wzw.tum.de (IA); h.leonhardt@lmu.de (HL)

☯ These authors contributed equally to this work.

Introduction

DNA methylation is an epigenetic modification that is well known to control eukaryotic gene expression [1,2]. In fact, methylation of regulatory sequences often correlates with a transcriptionally silent state. DNA methylation in mammals occurs as 5-methylcytosine (5mC) within CpG dinucleotides and is catalyzed by a family of DNA methyltransferases (Dnmts) [3]. Dnmt members are distinguished by their function; while the *de novo* methyltransferases Dnmt3a and Dnmt3b establish methylation patterns during development and cellular differentiation [4,5], the *maintenance* methyltransferase Dnmt1 copies these patterns during DNA replication [6,7,8]. Although DNA methylation per se can prevent binding of transcriptional regulators [9], the main mechanism by which transcriptional repression is achieved appears to involve 5mC binding proteins (MBPs). MBPs specifically recognize methylation marks and consequently stabilize silent chromatin states by recruitment of histone modifying enzymes and chromatin remodeling factors [10].

There are three families of MBPs known to date: the methyl-CpG binding domain (MBD) family, the Uhrf family and the Kaiso protein family. In contrast to the members of the MBD and Kaiso families that specifically recognize fully methylated CpG sites, Uhrf1, the best characterized member of the Uhrf family, preferentially binds hemimethylated DNA, the substrate of maintenance methylation [11,12,13,14]. Notably, crystal struc-

tures of the DNA binding domains of MeCP2 and Uhrf1 in complex with DNA revealed striking differences: whereas the MeCP2 MBD recognizes methylated CpG sites based on hydration of the DNA major groove, the Uhrf1 (Set and Ring associated) SRA domain uses a base-flipping mechanism to bind DNA containing hemimethylated CpG sites [11,12,14,15]. Interestingly, Uhrf1 recently emerged as essential cofactor for maintenance methylation potentially by recruiting Dnmt1 to its target sites [13,16,17].

In addition to 5mC, genomic DNA has been recently shown to contain 5-hydroxy-methylcytosine (5hmC), which results from oxidation of 5mC catalyzed by Tet proteins [18,19,20]. This new modification has been implicated in DNA demethylation, either passively as 5hmC containing DNA is not a substrate for Dnmt1 [21], or actively by so far unknown mechanisms. The central questions remain which proteins recognize 5hmC modified DNA and whether 5hmC has a direct role in gene regulation similar to its analog 5mC.

In this study, we characterized the 5mC/5hmC DNA binding properties of two representative 5mC binding protein domains, the MBD of MeCP2 and the SRA domain of Uhrf1. We found that in contrast to the MBD, the SRA domain binds hydroxymethylated DNA substrates with similar affinity as methylated substrates. We investigated the binding mode and energies of Uhrf1 to DNA substrates containing 5mC and 5hmC using molecular dynamics simulations of the respective SRA:DNA complexes.

Results

Uhrf1 binds DNA substrates containing hydroxymethylated CpG sites

Using a newly established DNA binding assay [22,23] as well as electrophoretic mobility shift assays, we investigated the DNA binding activity of Uhrf1, its SRA domain (SRA^{Uhrf1}) and the MBD of MeCP2 (MBD^{MeCP2}) to methylated and hydroxymethylated DNA in direct competition (Figure 1, Supplementary Figure S1; note that all supplementary information can also be found in the Combined Supporting Information File S1). We found that the Uhrf1 constructs bind 5mC and 5hmC containing substrates with similar affinities independent of whether one or both cytosine residues of the palindromic CpG site were modified. Control experiments performed with hemimethylated DNA in competition with either unmethylated substrates or substrates containing no CpG site showed that the observed binding activity to methylated and hydroxymethylated DNA is indeed specific (Supplementary Figure S2). In stark contrast to Uhrf1, we found that MBD^{MeCP2} clearly discriminates between methylation and hydroxymethylation, which is in accordance with previous reports [21,24].

Molecular dynamics simulations of SRA:DNA complexes with 5mC and 5hmC

To investigate the binding mode of the SRA domain to DNA containing 5mC or 5hmC, we performed molecular dynamics simulations for both SRA:DNA complexes. Consistent with the *in vitro* DNA binding data, modeling of an additional hydroxyl group into the complex structure of the Uhrf1 SRA domain with DNA containing hemimethylated CpG sites revealed no spatial constraints for accommodation of the flipped 5hmC nucleotide within the binding pocket (Figure 2). Based on these initial models

of the bound conformation, we performed molecular dynamics simulations for a time interval of 57 ns and monitored the RMSD and RMSF values (Supplementary Figures S3 and S4). In both systems equilibrium was reached after 20 to 30 ns. To assure evaluation of equilibrated systems, we continued the equilibrium simulations for another 27 ns and used only the last 10 ns for subsequent interaction energy analysis [25]. To evaluate the stability of the flipped nucleotides within the binding site, we monitored the occurrence and stability of all hydrogen bonds in the vicinity of the binding site with respect to the progress of the simulations (Figure 3).

Before starting the simulations, all water molecules from the X-ray structure were removed and new water molecules were placed by the setup solvation algorithm of NAMD [26]. Therefore, no water molecules were present in the vicinity of the flipped nucleotides at the beginning of the simulations. Interestingly, in both simulations, water molecules from the water-filled simulation box moved into the nucleotide binding site within the first couple of nanoseconds (Figures 3C and 3D, hydrogen bonds 14 to 18). During the remainder of the simulation time, one water molecule was stabilized within the binding site by formation of distinct hydrogen bonds with protein and DNA. Notably, the position of this water molecule in the 5mC complex corresponds to that of a conserved water molecule in the experimental structure (Supplementary Figure S5), confirming the stability and accuracy of our simulations.

Despite the presence of a conserved water molecule in the binding pockets of both complexes, the corresponding hydrogen bond networks showed interesting differences. In the 5mC complex, this water molecule forms hydrogen bonds with the phosphodiester group of the methylated nucleotide as well as with the SRA residues I454 and G453, thereby bridging the DNA

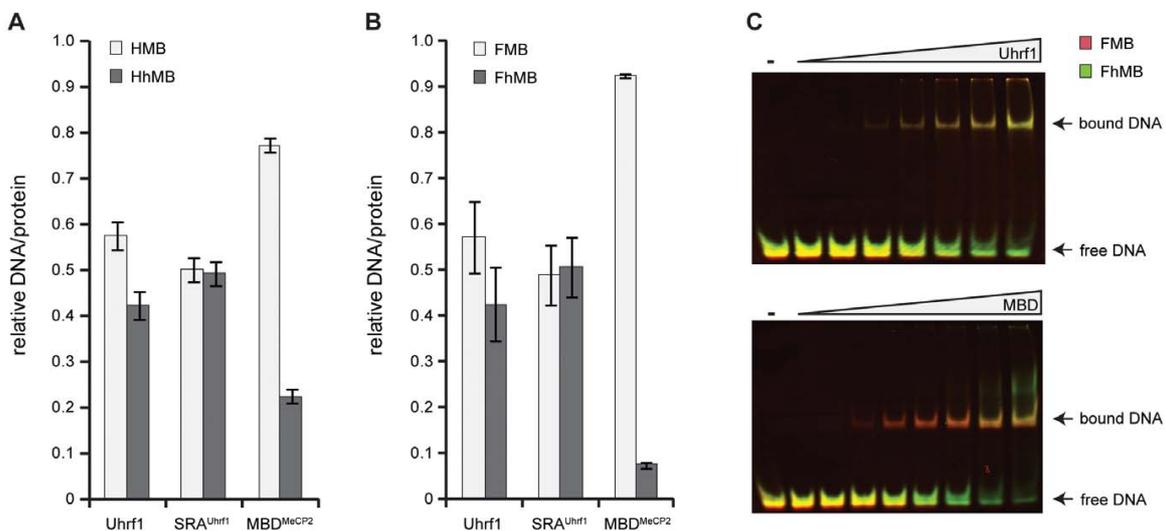


Figure 1. DNA binding specificity of 5-methylcytosine binding proteins. (A+B) Relative DNA/protein ratios of Uhrf1, its SRA domain (SRA^{Uhrf1}) and the MBD of MeCP2 (MBD^{MeCP2}) with two differentially labeled DNA substrates in direct competition. (A) Binding to DNA substrates containing a hemimethylated or hemihydroxymethylated CpG site (HMB versus HhMB, respectively). (B) Binding to DNA substrates containing a fully methylated or fully hydroxymethylated CpG site (FMB versus FhMB, respectively). Results are shown as means of three independent experiments with standard deviation error bars. Note that MBD^{MeCP2} preferentially binds to FMB, whereas the Uhrf1 constructs do not discriminate between FMB and FhMB. (C) Electrophoretic mobility shift assays were performed with Uhrf1 or MBD^{MeCP2} and equimolar amounts of FMB (red) and FhMB (green) in competition. The overlay of the two substrate channels reveals simultaneous shifting of both DNA substrates with Uhrf1, whereas with MBD^{MeCP2} the FMB substrate shifts at a lower protein concentration than the FhMB substrate, confirming differential binding. doi:10.1371/journal.pone.0021306.g001

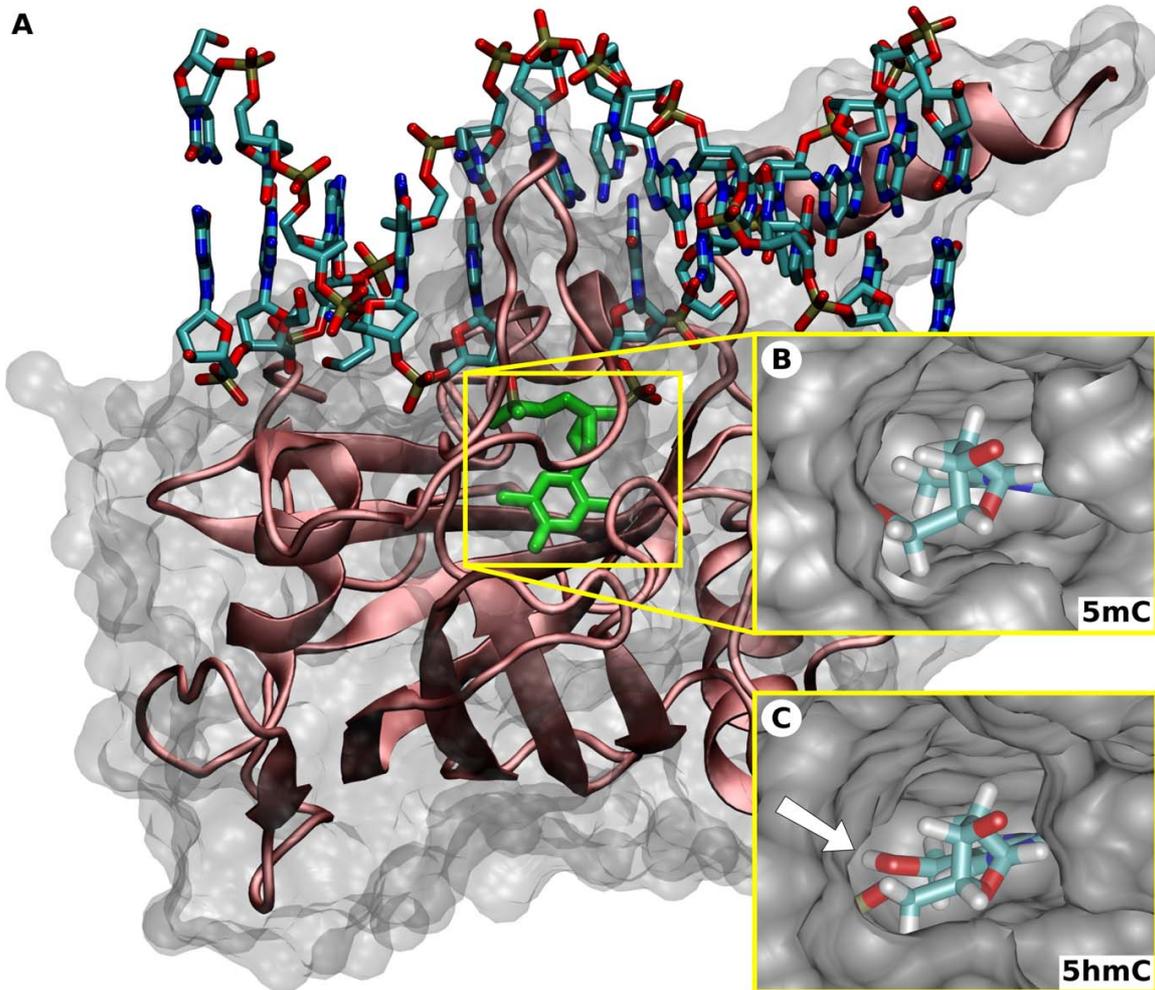


Figure 2. Structure of the Uhrf1 SRA domain in complex with hemimethylated and hemihydroxymethylated DNA. (A) Experimental structure of the Uhrf1 SRA domain in complex with hemimethylated DNA (PDB-ID:3fde, [14]). The protein is shown in cartoon and the DNA in licorice representation. The 5mC nucleotide is highlighted in green. Note that the 5mC residue is flipped out of the DNA double helix. (B+C) Models of the SRA binding pocket with bound 5mC (B) and 5hmC (C) serving as starting points for the molecular dynamics simulations. The location of the hydroxyl group in the 5hmC complex is highlighted by the white arrow. The view is from the top of the binding site (DNA backbone) and rotated by 90 degrees compared to (A).
doi:10.1371/journal.pone.0021306.g002

backbone:protein interaction (Figure 3A–C, hydrogen bonds 14–16, Figure 4A). Furthermore, direct hydrogen bonds between the 5mC DNA backbone and the protein are formed involving residues G453, S486, and R489 (hydrogen bonds 1–4).

The hydrogen bond network of the 5hmC complex is more stable compared to the 5mC complex (Figure 3D, compare with 3C). Most prominently, one additional and very stable hydrogen bond is formed between the conserved water molecule and the hydroxyl group of the 5hmC nucleotide (hydrogen bond 17). This interaction seems to specifically stabilize the hydrogen bonding network between the DNA backbone and the binding pocket residues G453, S486, and R489 (hydrogen bonds 1–4). Interestingly, these hydrogen bonds have been previously identified to be important for DNA binding [14] and possibly stabilize the flipped conformation of the nucleotide within the binding site. In addition,

the hydrogen bond network within the protein involving residues V466 and G453 as well as residues T484 and D474 is stabilized in the 5hmC complex (hydrogen bonds 11–13).

Since water dynamics and to some extent also DNA dynamics can depend on the ion concentration parameters used in the molecular dynamics simulation, we performed a second simulation of the 5hmC complex with a higher ion concentration (Supplementary Figure S6). Consistent to the first simulation with 5hmC, we observed the same overall water dynamics and hydrogen bonding patterns including hydrogen bond formation between the hydroxyl group of the 5hmC nucleotide and the conserved water molecule within the SRA structure. Notably, the stable hydrogen bonding between protein residue S486 and the DNA backbone in the first simulation (hydrogen bonds 2a and 2b) seems to be replaced by a stable hydrogen bond of S486 with the water

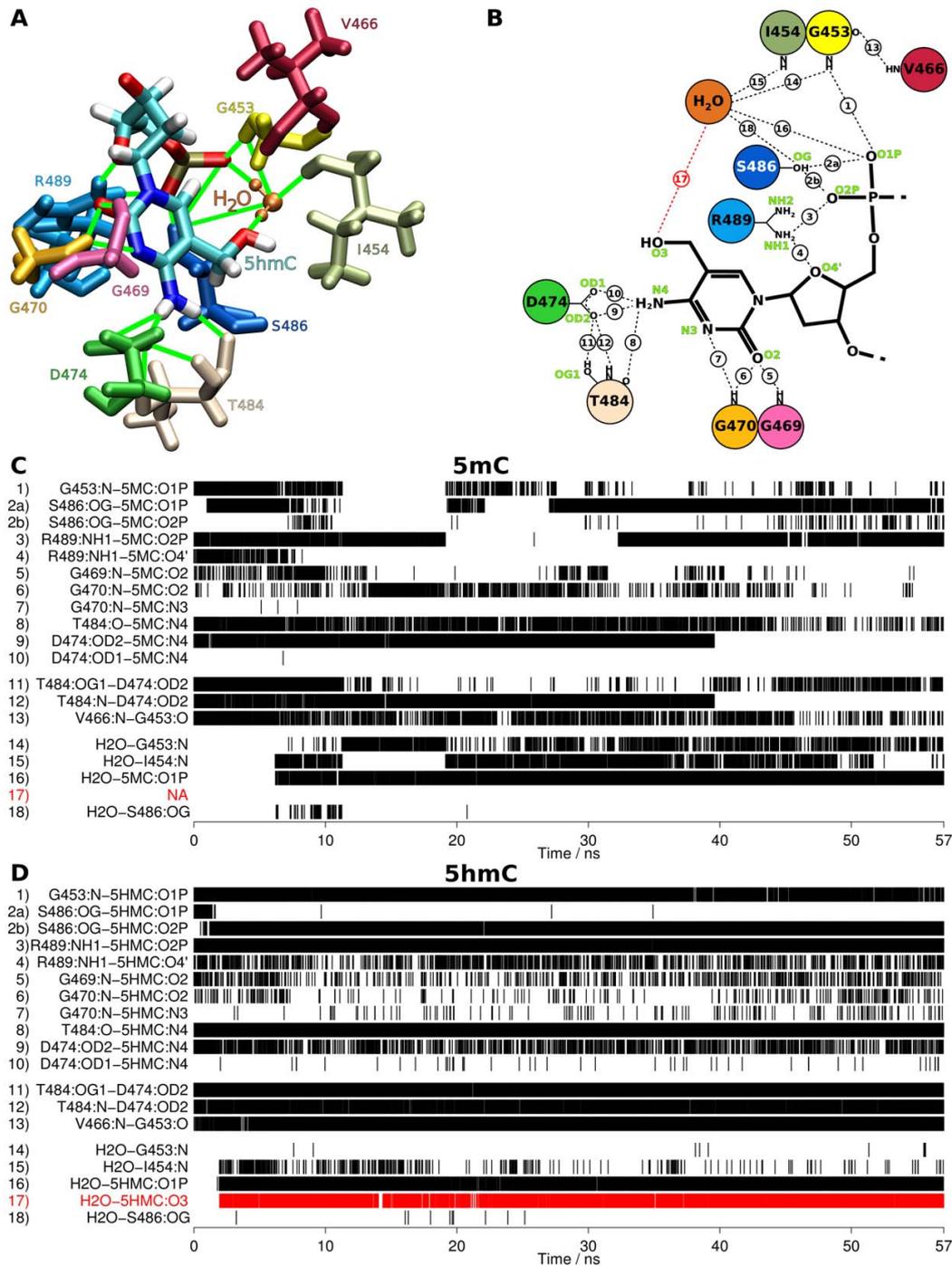


Figure 3. Molecular dynamics simulations of the SRA domain in complex with 5mC and 5hmC containing DNA. (A+B) Three and two-dimensional schematic drawings summarizing the hydrogen bond networks between the nucleotides, the SRA binding pocket, and a conserved water molecule during the simulations. The numbers in (B) correspond to the numbering in (C+D). **(C+D)** Hydrogen bond occurrences during the molecular dynamics simulations of the SRA domain in complex with either 5mC (C) or 5hmC containing DNA (D). Each vertical line represents a single observed hydrogen bond. The hydrogen bond between 5hmC and the conserved water is highlighted in red.
doi:10.1371/journal.pone.0021306.g003

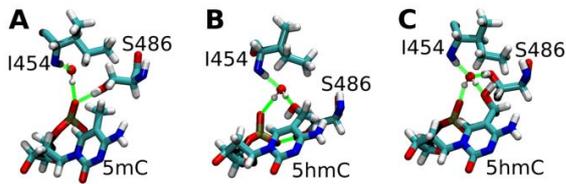


Figure 4. Hydrogen bond networks stabilizing 5mC and 5hmC within the SRA binding pocket. (A) SRA complex with DNA containing 5mC. (B+C) SRA complex with DNA containing 5hmC. In the 5hmC complex, the water molecule stably interacts with the hydroxyl group of the nucleotide, but two alternative conformations of the SRA binding pocket exist depending on the ion concentration. In the absence of salt, binding involves an interaction of the S486 residue with the phosphate group of the flipped nucleotide (B), whereas in the presence of 0.5 M NaCl, residue S486 interacts with the conserved water molecule (C).
doi:10.1371/journal.pone.0021306.g004

molecule in the second simulation (hydrogen bond 18), indicating two alternative interaction patterns for the S486 residue in the 5hmC complex (Figures 4B and 4C, compare Figure 3D and Supplementary Figure S6B). In conclusion, these data suggest that stable, water bridged hydrogen bond formation of the hydroxyl group of the flipped 5hmC nucleotide with its surrounding occurs in and stabilizes this DNA:SRA complex.

Similar interaction energies for SRA complexes with 5mC and 5hmC containing DNA

To estimate the binding affinity between the Uhrf1 SRA domain and DNA containing either 5mC or 5hmC, we calculated the respective interaction energies using the linear interaction energy (LIE) approach [25]. To exclude energy contributions due to base-flipping when comparing the interaction of the DNA with the protein (bound state) or with the solvent (unbound state), we simulated the DNA in a flipped state in both cases. We determined the difference between the binding energies of the two complexes ($\Delta\Delta G = \Delta G_{5mC} - \Delta G_{5hmC}$). We included either i) the whole DNA and SRA structure ($\Delta\Delta G = -7.94$ kcal/mol) or ii) the flipped nucleotide with its five neighboring nucleotides and the binding pocket of the protein, defined as all residues within a distance of 15 Å from the nucleotide in the starting conformation ($\Delta\Delta G = -6.65$ kcal/mol). These values suggest that the slight difference in binding affinity is predominantly due to interaction of the flipped nucleotide with the proximal protein residues that form the binding site. Considering the estimated uncertainty of about 3–4 kcal/mol in our calculations, these values indicate that both 5mC and 5hmC containing DNA substrates bind with very similar affinity to the SRA domain of Uhrf1.

Discussion

In summary, we observed fundamentally different binding specificities for the DNA binding domains of representative 5mC binding proteins. Hydroxylation of 5mC clearly interferes with DNA binding by the MBD of MeCP2 and might prevent subsequent establishment of repressive chromatin structures in a cellular context, thereby changing the cellular interpretation of an epigenetic modification. Notably, MeCP2 expression is highest in brain tissues where also 5hmC levels are highest [18,27,28]. In stark contrast, Uhrf1, a key factor in maintenance methylation, recognizes 5hmC as well as 5mC. The results of our molecular dynamics simulations provide a structural explanation for recognition of 5hmC. Interestingly, the flipped 5hmC base not only fits into the binding pocket of the Uhrf1 SRA domain, but is

specifically stabilized by hydrogen bond formation involving the 5hmC hydroxyl group. This interaction is bridged by a conserved water molecule present within the SRA binding pocket and seems to stabilize the overall hydrogen bond network of the 5hmC complex. Also in the 5mC complex a conserved water molecule is found in the vicinity of the flipped cytosine, which in this case, however, only interacts with the SRA domain and the backbone of the DNA and not with the flipped nucleotide itself.

The specific binding of Uhrf1 to 5hmC containing DNA was clearly unexpected and puts the existing hypothesis on Uhrf1 function into a new perspective. Knock-out studies in mouse embryonic stem cells and embryos revealed that Uhrf1 is essential for maintenance DNA methylation by Dnmt1 [17]. Based on the specific binding of Uhrf1 to hemimethylated CpG sites and its interaction with Dnmt1, Uhrf1 was suggested to operate by recruiting Dnmt1 to its target sites [11,12,13,14,17]. Recent studies suggested a role of hydroxymethylation in passive [21] and/or active [29,30,31] DNA demethylation. The binding of Uhrf1 to hydroxymethylated DNA reported in this study now raises the question how Uhrf1 contributes to change or maintenance of methylation *in vivo*. In this context it should also be noted that the preferential binding of Uhrf1 to hemimethylated DNA is relatively weak, especially if compared to the intrinsic preference of Dnmt1 for methylation of these substrates [22,23]. Moreover, multiple interactions of Uhrf1 with repressive histone tail modifications [23] as well as other heterochromatin associated proteins [32,33] seem to be required for the specific localization and targeting of Uhrf1 *in vivo*. Together, these data strongly argue for a more complex mechanism of Uhrf1 function in living cells and emphasize the need for further studies to understand the pivotal role of Uhrf1 in the establishment, maintenance and change of genome-wide methylation patterns.

Using a combination of *in vitro* and *in silico* studies, we clearly demonstrate that Uhrf1 can bind 5hmC containing DNA. It still remains elusive whether or in which specific context Uhrf1 binds 5hmC modified DNA substrates in living cells. Uhrf1 binding to 5hmC and possible functional consequences *in vivo* are likely to depend on additional interacting factors. Comparison of genome-wide Uhrf1 ChIP profiles with 5mC and 5hmC distribution should help to clarify the interactions and functions of Uhrf1 *in vivo*. Finally, it is interesting to note that Uhrf1 is the only base-flipping protein with so far unknown catalytic function on DNA. The direct interaction of a water molecule with the hydroxyl group of 5hmC within the SRA binding pocket might possibly point towards a role of Uhrf1 in the further modification of this sixth DNA base. In conclusion, our study provides new perspectives on the cellular interpretation and possible further metabolism of this new epigenetic DNA modification.

Materials and Methods

Expression constructs, cell culture and transfection

Mammalian expression constructs for enhanced green fluorescent protein (GFP), Uhrf1 (GFP-Uhrf1), the SRA domain of Uhrf1 (GFP-SRA^{Uhrf1}) and the MBD of MeCP2 (MBD^{MeCP2}-YFP) were described previously [22,23,34]. Note that all constructs encode fusion proteins of either GFP or yellow fluorescent protein (YFP). HEK293T cells [35] were cultured in DMEM supplemented with 50 µg/ml gentamicin and 10% fetal calf serum. For expression of GFP/YFP fusion proteins, HEK293T cells were transfected with the corresponding expression constructs using polyethylenimine (Sigma).

DNA substrate preparation

Fluorescently labeled DNA substrates were prepared by mixing two HPLC-purified DNA oligonucleotides (IBA GmbH, Supplementary Tables S1 and S2) in equimolar amounts, denaturation

for 30 sec at 92°C and slow cool-down to 25°C allowing hybridization. After purification by 15% non-denaturing PAGE, DNA substrates were resuspended in binding buffer (20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT).

Pull-down DNA binding assay

In vitro DNA binding assays were performed as described previously [22,23]. In brief, GFP/YFP fusions were purified from HEK293T extracts using the GFP-Trap® (ChromoTek GmbH) and incubated with two differentially labeled DNA substrates at a final concentration of 200 nM DNA/50–100 nM immobilized protein for 45 min at room temperature in binding buffer. After removal of unbound substrate, the amounts of protein and DNA were determined by fluorescence intensity measurements with a Tecan Infinite M1000 plate reader. Binding ratios were calculated dividing the concentration of bound DNA substrate by the concentration of GFP/YFP fusion on the beads, corrected by values from a control experiment using DNA substrates of the same sequence but with different fluorescent labels, and normalized by the total amount of bound DNA.

Electrophoretic mobility shift assay

For competitive electrophoretic mobility shift assays, equimolar amounts of two differentially labeled DNA substrates (250 nM each) were incubated with increasing amounts of GFP/YFP fusion protein (Supplementary Figure S1), subjected to 6% non-denaturing PAGE and analyzed with a Typhoon scanner (GE Healthcare), which allowed separate detection of DNA substrates and protein by ATTO labels and GFP tag, respectively, using the following laser/filter settings: 532 nm/580 nm (ATTO550), 633 nm/none (ATTO700), 488 nm/520 nm (GFP/YFP).

Molecular dynamics simulations

Molecular dynamics simulations were performed based on the X-ray structure of the Uhrf1 SRA domain with the PDB identifier 3FDE [14], using the program NAMD 2.7b1 [26] and the CHARMM22/27 force field [36,37]. Binding free energies were estimated using the Linear Interaction Energy (LIE) model [25].

After energy minimization of 50,000 steps, one hydrogen atom of the methyl group of the protein-bound 5-methylcytosine (5mC) residue was substituted by a hydroxyl group using the tool psfgen. CHARMM22 force field parameters were available for 5mC (patch: PRES 5MC2), but not for 5-hydroxymethylcytosine (5hmC). Therefore, a new 5hmC residue was created based on the 5mC parameters and topology. For this purpose, one hydrogen atom of the 5mC methyl group was exchanged by a hydroxyl group. The charges of the hydroxyl group were subsequently set to charges of the hydroxyl group of a serine residue according to the CHARMM27; the charges of the CH₂ group were adjusted accordingly (Supplementary Table S3). After solvation, the 5mC and 5hmC structures were further energy minimized for 50,000 steps. For each structure, two simulations were performed, in which the charges were either neutralized or a salt concentration of 0.5 M was used.

Each simulation was performed using periodic boundary conditions and particle-mesh-ewald summation [38] for long range non-bonded interactions. The non-bonded cutoff was set to 14 Å with a switching/shifting distance of 12 Å. A stepsize of 1 fs was chosen. The systems were heated from 0 to 200 K for 160 ps under constant volume. Harmonic restraints (1000 kcal mol⁻¹ nm⁻²) were applied to all atoms of the complex. The heat up was continued without harmonic restraints from 200 to 300 K for 80 ps under constant pressure conditions, using a Nose-Hoover barostat [39,40] with a target pressure of 1.01325 bar, an oscillation time scale of 100 fs, and

a damping time scale of 50 fs. The temperature was maintained by Langevin dynamics using a damping coefficient of 5/ps. The temperature bath was not coupled to hydrogen atoms. After the heat up procedure, the simulations were continued for 57 ns. During the simulations, all bond lengths were constrained to ideal values using the Shake algorithm [41,42].

For analysis of the simulation results, all hydrogen bonds formed by the flipped nucleotides and the binding site were identified and monitored throughout the simulations and the occurrence of water molecules in and around the binding site was monitored every 5 ps. In order to estimate the difference in the binding free energy of the two nucleotides, we performed three further simulations in which the protein and the two DNA molecules were simulated separately using the conditions described above. To keep the DNA in the flipped state, we additionally applied harmonic restraints to the whole DNA backbone (atom names: C4', P, O1P, O2P, O5', C5', C3', O3'). The solvated single protein was simulated for 34 ns and the separated DNA molecules were simulated for 20 ns.

To estimate the binding affinity of the two DNA molecules to the protein, we estimated the binding free energy according to the Linear Interaction Energy (LIE) model [25]:

$$\Delta G_{bind} = \alpha \Delta \langle V_{DNA-s}^{vdw} \rangle + \beta \Delta \langle V_{DNA-s}^{el} \rangle + \gamma \quad (1)$$

$$\Delta \langle V_{DNA-s}^{el/vdw} \rangle = \langle V_{bound}^{el/vdw} \rangle - \langle V_{unbound}^{el/vdw} \rangle \quad (2)$$

In this approach the binding free energy is approximated by the difference between the interaction energies ΔV^{el} and ΔV^{vdw} of the ligand in the protein-ligand complex (bound state) and in solution (unbound state). The $\langle \rangle$ denotes the average values obtained from the simulation trajectories. According to the linear response approximation the weights α and β were set to 1 and 0.5, respectively. We calculated the DNA-(protein+solvent) (bound state) and the DNA-solvent (free state) interaction energies from the trajectories of the DNA/SRA and the DNA/solvent simulations, using the average energy over the last 10 ns.

Supporting Information

Figure S1 Electrophoretic mobility shift assays with methylated and hydroxymethylated DNA substrates. Increasing amounts of Uhrf1, its SRA domain (SRA^{Uhrf1}) or the MBD domain of MeCP2 (MBD^{MeCP2}) were incubated with two differentially ATTO-labeled DNA substrates, which contain either one central fully methylated or fully hydroxymethylated CpG site (FMB-ATTO700 or FhMB-ATTO550, respectively), in direct competition. Samples were subjected to 6% non-denaturing PAGE and analyzed with a Typhoon scanner (GE Healthcare). The first, second and third columns show the scans for GFP/YFP, ATTO700 and ATTO550 fluorescence, respectively. The overlay of the two ATTO channels is shown in the fourth column (FMB: red, FhMB:green). (PDF)

Figure S2 DNA binding specificity of Uhrf1. Relative DNA/Uhrf1 ratios are shown for two differentially labeled fluorescent DNA substrates in direct competition. (A) Binding of Uhrf1 to DNA substrates containing no CpG site or one central hemimethylated CpG site (noCGB versus HMB, respectively). (B) Binding of Uhrf1 to DNA substrates containing one central un- or hemimethylated CpG site (UMB versus HMB, respectively). Results are shown as means of three independent experiments with standard deviation error bars. DNA substrates were prepared

by hybridization as described in the main text, except for noCGB, which was prepared by primer extension as described previously [22]. See Supplementary Tables S1 and S2 for DNA oligonucleotide sequences and purification grade of the used substrates. (PDF)

Figure S3 Atom-positional root-mean-square deviation of the protein and DNA backbone atoms during the simulations. The terminal DNA and protein residues were excluded from the calculations in the “subset” sets (red and black lines). (PDF)

Figure S4 Atom-positional root-mean-square fluctuations of the protein (A, C) and both DNA strands (B, D) during two simulation periods. Note that both structures show the same flexibility pattern during both simulation periods and are overall stable during both periods. This is in agreement with the RMSD data in Figure S3, which shows that equilibration is reached after 30 ns of simulation time. (PDF)

Figure S5 Superposition of the equilibrated 5mC structure after simulation (atom-name specific coloring) and the crystal structure (PDB-ID:3fde [14], green). The 5mC nucleotide, the residue I454 of the SRA binding pocket and the conserved water molecule are shown. Note that the distance between the oxygen atoms of the conserved water molecules in the two structures is only 1.1 Å. (PDF)

References

- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16: 6–21.
- Rottach A, Leonhardt H, Spada F (2009) DNA methylation-mediated epigenetic control. *J Cell Biochem* 108: 43–51.
- Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74: 481–514.
- Lei H, Oh S, Okano M, Juttermann R, Goss K, et al. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122: 3195–3205.
- Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99: 247–257.
- Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71: 865–873.
- Chuang LS-H, Ian H-I, Koh T-W, Ng H-H, Xu G, et al. (1997) Human DNA-Cytosine-5 Methyltransferase-PCNA Complex as a Target for p21WAF1. *Science* 277: 1996–2000.
- Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69: 915–926.
- Becker PB, Ruppert S, Schutz G (1987) Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. *Cell* 51: 435–443.
- Sasai N, Defossez PA (2009) Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int J Dev Biol* 53: 323–334.
- Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455: 818–821.
- Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, et al. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455: 822–825.
- Bostick M, Kim JK, Esteve P-O, Clark A, Pradhan S, et al. (2007) UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* 317: 1760–1764.
- Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 455: 826–829.
- Ho KL, McNaie IW, Schmiedeberg L, Klose RJ, Bird AP, et al. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell* 29: 525–531.
- Achour M, Jacq X, Ronde P, Alhosin M, Charlot C, et al. (2008) The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene* 27: 2187–2197.
- Sharif J, Muto M, Takebayashi SI, Suetake I, Iwamatsu A, et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*.
- Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324: 929–930.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324: 930–935.
- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129–1133.
- Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 67: 946–950.
- Frauer C, Leonhardt H (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res* 37: e22.
- Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, et al. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res* 38: 1796–1804.
- Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, et al. (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 Res* 32: 4100–4108.
- Aqvist J, Luzhkov VB, Brandsdal BO (2002) Ligand binding affinities from MD simulations. *Acc Chem Res* 35: 358–365.
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.
- Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, et al. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5: e15367.
- Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 38: e181.
- Guo JU, Su Y, Zhong C, Ming GL, Song H (2011) Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell* 145: 423–434.
- Iqbal K, Jin SG, Pfeifer GP, Szabo PE (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.
- Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, et al. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2: 241.
- Kim JK, Esteve PO, Jacobsen SE, Pradhan S (2009) UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res* 37: 493–505.
- Meilinger D, Fellingner K, Bultmann S, Rothbauer U, Bonapace IM, et al. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b,

- and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* 10: 1259–1264.
34. Brero A, Easwaran HP, Nowak D, Grunewald I, Cremer T, et al. (2005) Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J Cell Biol* 169: 733–743.
 35. DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, et al. (1987) Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol* 7: 379–387.
 36. MacKerell AD, Jr., Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56: 257–265.
 37. MacKerell AD, Bashford D, Bellott, Dunbrack RL, Evanseck JD, et al. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *The Journal of Physical Chemistry B* 102: 3586–3616.
 38. Darden T, York D, Pedersen L (1993) PARTICLE MESH EWALD - AN $N \cdot \text{LOG}(N)$ METHOD FOR EWALD SUMS IN LARGE SYSTEMS. *Journal of Chemical Physics* 98: 10089–10092.
 39. Martyna GJ, Tobias DJ, Klein ML (1994) CONSTANT-PRESSURE MOLECULAR-DYNAMICS ALGORITHMS. *Journal of Chemical Physics* 101: 4177–4189.
 40. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) CONSTANT-PRESSURE MOLECULAR-DYNAMICS SIMULATION - THE LANGEVIN PISTON METHOD. *Journal of Chemical Physics* 103: 4613–4621.
 41. Miyamoto S, Kollman PA (1992) SETTLE - AN ANALYTICAL VERSION OF THE SHAKE AND RATTLE ALGORITHM FOR RIGID WATER MODELS. *Journal of Computational Chemistry* 13: 952–962.
 42. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) NUMERICAL-INTEGRATION OF CARTESIAN EQUATIONS OF MOTION OF A SYSTEM WITH CONSTRAINTS - MOLECULAR-DYNAMICS OF N-ALKANES. *Journal of Computational Physics* 23: 327–341.

Supplementary Material

Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain

Carina Frauer, Thomas Hoffmann, Sebastian Bultmann, Valentina Casa, M. Cristina Cardoso, Iris Antes and Heinrich Leonhardt

Supplementary Tables

Supplementary Table S1. Sequences of DNA oligonucleotides used for preparation of double stranded fluorescent DNA substrates.

M: 5-methylcytosine. X: 5-hydroxymethylcytosine.

Name	Sequence
CGup	5' - CTCAACAACCTAACTACCATCCGGACCAGAAGAGTCATCATGG -3'
MGup	5' - CTCAACAACCTAACTACCATCMGGACCAGAAGAGTCATCATGG -3'
hmCGup	5' - CTCAACAACCTAACTACCATCXGGACCAGAAGAGTCATCATGG -3'
noCGup	5' - CTCAACAACCTAACTACCATCTGGACCAGAAGAGTCATCATGG -3'
um550	5' - ATTO550-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
um590	5' - ATTO590-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
um647N	5' - ATTO647N-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
um700	5' - ATTO700-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
mC700	5' - ATTO700-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
hmC550	5' - ATTO550-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
550-Fill-In	5' - ATTO550-CCATGATGACTCTTCTGGTC -3'

Supplementary Table S2. DNA substrates used for the DNA binding assays.

Name	CpG site	Label	Oligo I	Oligo II	Purification grade and use
HMB550	hemimethylated	ATTO550	MGup	um550	· hybridization of HPLC-purified oligos
HMB700	hemimethylated	ATTO700	MGup	um700	· gel-purification
HhMB700	hemihydroxymethylated	ATTO700	hmCGup	um700	· used for data in figure 2 and supplementary figure 1
FMB700	fully methylated	ATTO700	MGup	mC700	
FhMB550	fully hydroxymethylated	ATTO550	hmCGup	hmC550	
noCG550	no CpG site	ATTO550	noCGup	550-Fill-In	· primer extension for noCG550
HMB550	hemimethylated	ATTO550	MGup	um550	· hybridization of HPLC-purified oligos
HMB647N	hemimethylated	ATTO647N	MGup	um647N	· gel-purification · used for data in supplementary figure 2A
UMB550	unmethylated	ATTO550	CGup	um550	· hybridization of HPLC-purified oligos
UMB590	unmethylated	ATTO590	CGup	um590	· used for data in supplementary figure 2B, n=2
HMB590	hemimethylated	ATTO590	MGup	um590	
UMB647N	unmethylated	ATTO647N	CGup	um647N	· hybridization of PAGE-purified oligos
UMB700	unmethylated	ATTO700	CGup	um700	· used for data in supplementary figure 2B, n=1
HMB700	hemimethylated	ATTO700	MGup	um700	

Supplementary Table S3. Residue Topology File and parameters used for the 5hmC residue during the simulations.

```

=====
TOPOLOGY (based on 5mC topology from patches: PRES 5MC2 and PRES DEO1)
=====
! 5-hydroxy-methyl cytosine
RESI 5HMC -1.00 !
ATOM P P 1.50 !
ATOM O1P ON3 -0.78 !
ATOM O2P ON3 -0.78 !
ATOM O5' ON2 -0.57 !
ATOM C5' CN8B -0.08 !
ATOM H5' HN8 0.09 !
ATOM H5'' HN8 0.09 !
GROUP !
ATOM C4' CN7 0.16 !
ATOM H4' HN7 0.09 !
ATOM O4' ON6 -0.50 !
ATOM C1' CN7B 0.16 !
ATOM H1' HN7 0.09 !
GROUP !
ATOM N1 NN2 -0.13 !
ATOM C6 CN3 0.05 !
ATOM H6 HN3 0.17 !
ATOM C5 CN3D -0.11 !
ATOM C5M CN9 0.10 !
ATOM H5M1 HN9 0.09 !
ATOM H5M2 HN9 0.09 !
ATOM O3 OH1 -0.66 !
ATOM H3 H 0.43 !
ATOM C2 CN1 0.52 !
ATOM O2 ON1C -0.49 !
ATOM N3 NN3 -0.66 !
ATOM C4 CN2 0.65 !
ATOM N4 NN1 -0.75 !
ATOM H41 HN1 0.37 !
ATOM H42 HN1 0.33 !
GROUP !
ATOM C2' CN8 -0.18 !
ATOM H2'' HN8 0.09 !
ATOM H2' HN8 0.09 !
GROUP !
ATOM C3' CN7 0.01 !
ATOM H3' HN7 0.09 !
ATOM O3' ON2 -0.57 !
BOND P O1P P O2P P O5'
BOND O5' C5' C5' C4' C4' O4' C4' C3' O4' C1'
BOND C1' N1 N1 C1' C2' N1 C2 N1 C6
BOND C2 N3 C4 N4 N4 H41 N4 H42
BOND C4 C5 C2' C3' C3' O3' O3' +P
BOND C1' H1' C2' H2'' C2' H2' C3' H3' C4' H4' C5' H5'
BOND C5' H5'' C6 H6
BOND C5 C5M C5M H5M1 C5M H5M2 C5M O3 O3 H3
ANGL C4 C5 C5M C6 C5 C5M
ANGL C5 C5M H5M1 C5 C5M H5M2 C5 C5M O3 C5M O3 H3
ANGL H5M1 C5M H5M2 H5M1 C5M O3 H5M2 C5M O3
DIHE C5M C5 C4 N3 C5M C5 C4 N4
DIHE C5M C5 C6 H6 C5M C5 C6 N1
DIHE H5M1 C5M C5 C4 H5M1 C5M C5 C6
DIHE H5M2 C5M C5 C4 H5M2 C5M C5 C6
DIHE O3 C5M C5 C4 O3 C5M C5 C6
DIHE H3 O3 C5M C5 H3 O3 C5M H5M2
DIHE H3 O3 C5M H5M1
DOUBLE C2 O2 C5 C6 N3 C4
IMPR C2 N1 N3 O2 C4 N3 C5 N4
IMPR N4 C4 H41 H42
DONO H42 N4
DONO H41 N4
DONO H3 O3
ACCE O2 C2
ACCE N3
ACCE O1P P
ACCE O2P P

```

ACCE O3'
 ACCE O4'
 ACCE O5'
 ACCE O3

BILD -O3'	P	O5'	C5'	1.6001	101.45	-46.90	119.00	1.4401	!alpha
BILD -O3'	O5'	*P	O1P	1.6001	101.45	-115.82	109.74	1.4802	
BILD -O3'	O5'	*P	O2P	1.6001	101.45	115.90	109.80	1.4801	
BILD P	O5'	C5'	C4'	1.5996	119.00	-146.00	110.04	1.5160	!beta
BILD O5'	C5'	C4'	C3'	1.4401	108.83	60.00	116.10	1.5284	!gamma
BILD C5'	C4'	C3'	O3'	1.5160	116.10	140.00	115.12	1.4212	!delta
BILD C4'	C3'	O3'	+P	1.5284	111.92	155.00	119.05	1.6001	!epsilon
BILD C3'	O3'	+P	+O5'	1.4212	119.05	-95.20	101.45	1.5996	!zeta
BILD O4'	C3'	*C4'	C5'	1.4572	104.06	-120.04	116.10	1.5160	
BILD C2'	C4'	*C3'	O3'	1.5284	100.16	-124.08	115.12	1.4212	
BILD C4'	C3'	C2'	C1'	1.5284	100.16	-30.00	102.04	1.5251	
BILD C3'	C2'	C1'	N1	1.5284	101.97	147.89	113.71	1.4896	
BILD O4'	C1'	N1	C2	1.5251	113.71	-97.2	125.59	1.3783	!chi
BILD C1'	C2	*N1	C6	1.4896	117.79	-180.00	120.6	1.364	
BILD C2	N1	C6	C5	1.399	120.6	0.0	121.0	1.337	
BILD C6	N1	C2	N3	1.364	120.6	0.0	118.9	1.356	
BILD N1	N3	*C2	O2	1.399	118.9	180.0	121.9	1.237	
BILD N1	C2	N3	C4	1.399	118.9	0.0	120.0	1.334	
BILD C5	N3	*C4	N4	1.426	121.8	180.00	118.9	1.337	
BILD N3	C4	N4	H41	1.337	117.9	0.00	118.9	1.01	
BILD H41	C4	*N4	H42	1.01	118.9	180.00	120.7	1.01	
BILD N1	C5	*C6	H6	0.0	0.0	180.0	0.0	0.0	
BILD C1'	C3'	*C2'	H2'	1.5284	102.04	-114.67	110.81	1.01	
BILD O4'	C2'	*C1'	H1'	0.0	0.0	-115.0	0.0	0.0	
BILD C1'	C3'	*C2'	H2''	0.0	0.0	115.0	0.0	0.0	
BILD C1'	C3'	*C2'	H2'	0.0	0.0	-115.0	0.0	0.0	
BILD C2'	C4'	*C3'	H3'	0.0	0.0	115.0	0.0	0.0	
BILD C3'	O4'	*C4'	H4'	0.0	0.0	-115.0	0.0	0.0	
BILD C4'	O5'	*C5'	H5'	0.0	0.0	-115.0	0.0	0.0	
BILD C4'	O5'	*C5'	H5''	0.0	0.0	115.0	0.0	0.0	
BILD C6	C4	*C5	C5M	0.0	0.0	180.0	0.0	0.0	
BILD C4	C5	C5M	H5M1	0.0	0.0	180.0	0.0	0.0	
BILD C5	H5M1	*C5M	H5M2	0.0	0.0	-115.0	0.0	0.0	
BILD H5M1	H5M2	*C5M	O3	0.0	0.0	115.0	0.0	0.0	
BILD C4	C5	C5M	O3	0.0	0.0	60.0	0.0	0.0	
BILD C5	C5M	O3	H3	0.0	0.0	180.0	0.0	0.0	

=====

FORCEFIELD PARAMETERS:

=====

...

BONDS

...
 !added for 5HMC TU_TCB TH und ISA
 OH1 CN9 428.000 1.4200 !ACC. TO OH1-CT3

...

ANGLES

...
 !added for 5HMC TU_TCB TH und ISA
 OH1 CN9 CN3D 75.700 110.1000 !ACC. TO OH1-CT2-CT2
 OH1 CN9 HN9 45.900 108.8900 !ACC. TO OH1-CT3-HA
 H OH1 CN9 57.500 106.0000 !ACC. TO H-OH1-CT2

...

DIHEDRALS

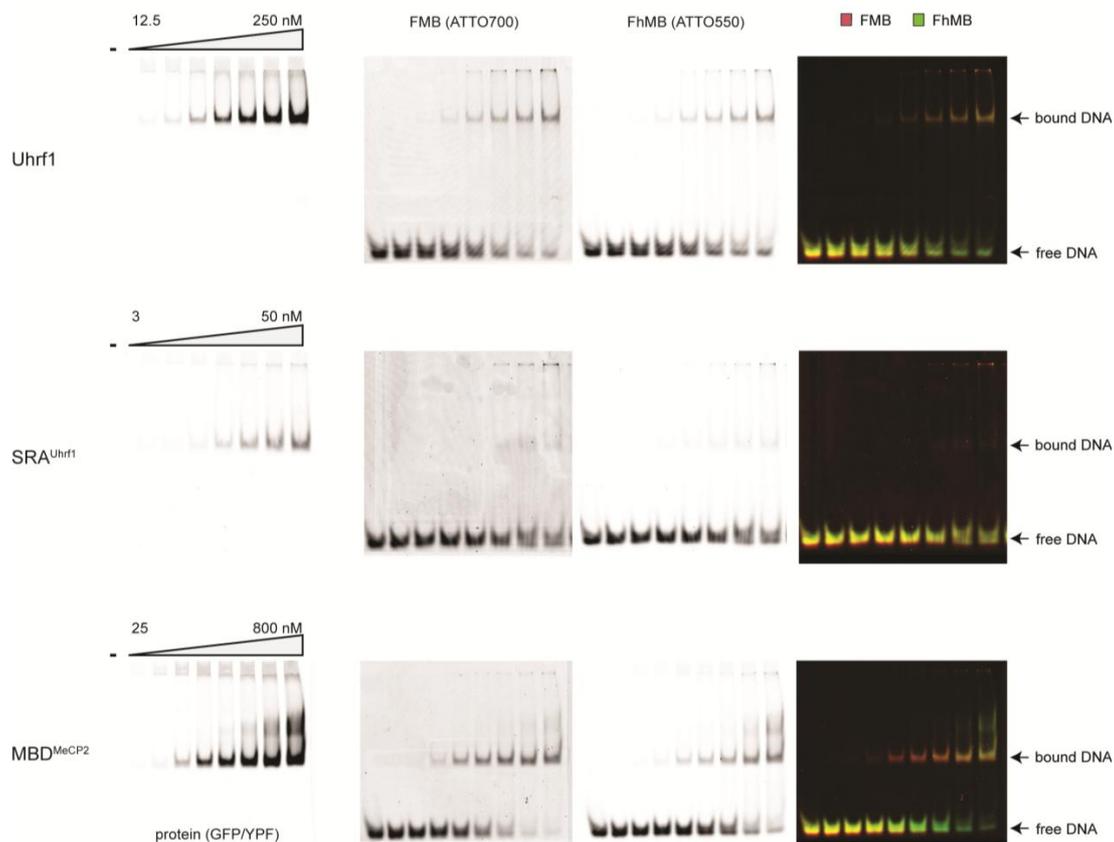
...
 !added for 5HMC TU_TCB TH und ISA
 H OH1 CN9 CN3D 1.3000 1 0.00 !ACC. TO H-OH1-CT2-CT2
 H OH1 CN9 CN3D 0.3000 2 0.00 !ACC. TO H-OH1-CT2-CT2
 H OH1 CN9 CN3D 0.4200 3 0.00 !ACC. TO H-OH1-CT2-CT2
 CN3 CN3D CN9 OH1 0.0 3 0.0 !ACC. TO CN3-CN3D-CN9-HN9
 CN2 CN3D CN9 OH1 0.35 3 0.0 !ACC. TO CN3-CN3D-CN9-HN9
 HN9 CN9 OH1 H 0.1400 3 0.00 !ACC. TO X-CT2-OH1-X

...

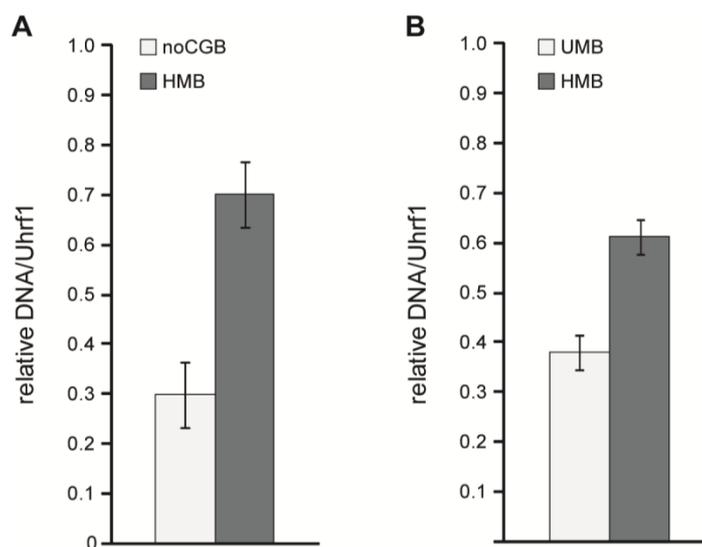
IMPROPER

...

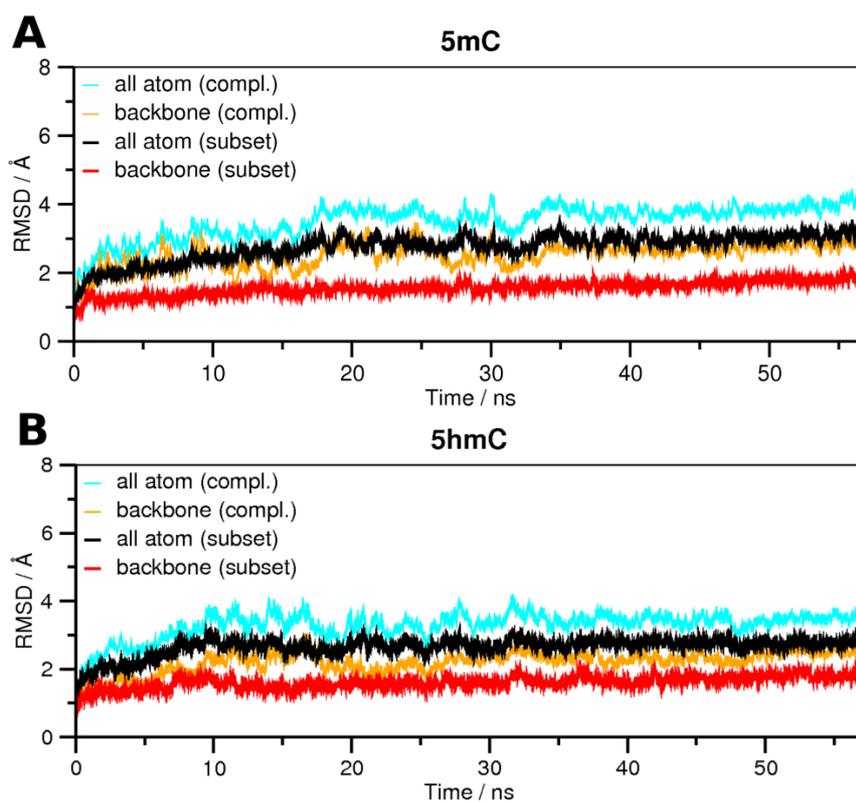
Supplementary Figures



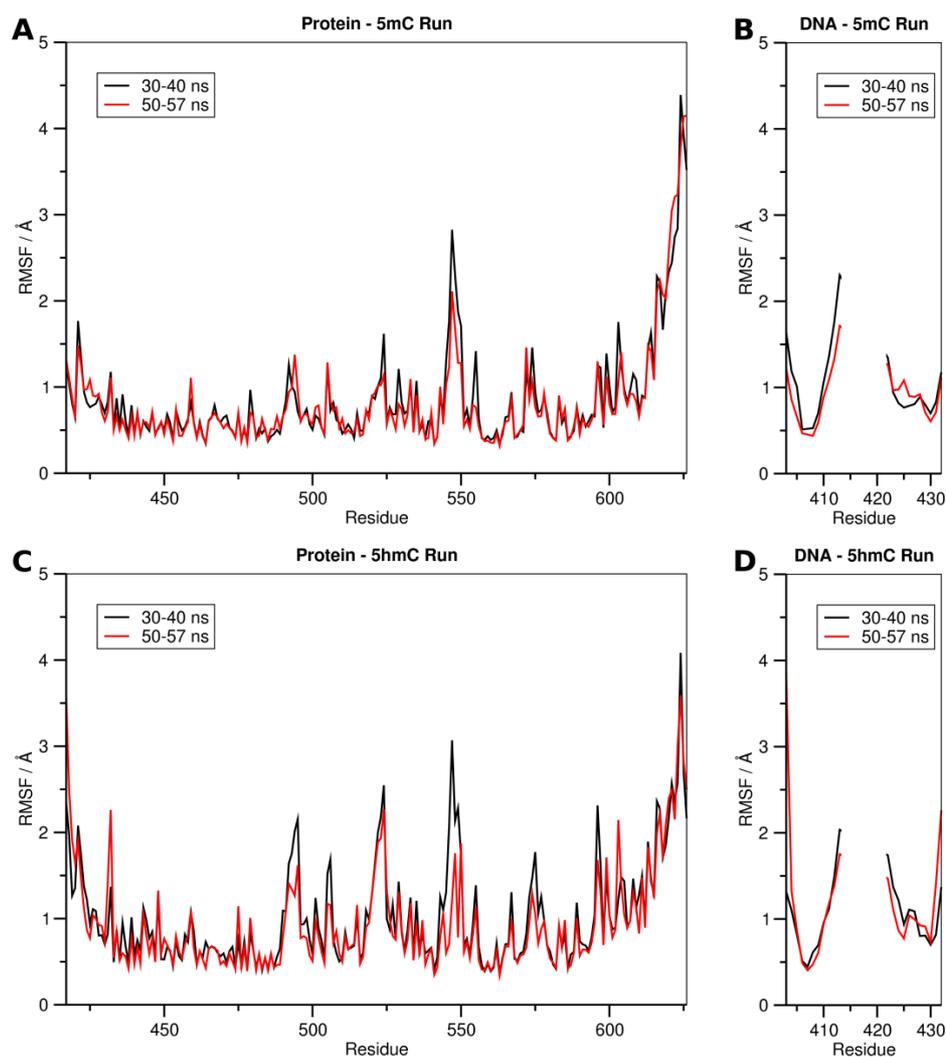
Supplementary Figure S1. Electrophoretic mobility shift assays with methylated and hydroxymethylated DNA substrates. Increasing amounts of Uhrf1, its SRA domain (SRA^{Uhrf1}) or the MBD domain of MeCP2 (MBD^{MeCP2}) were incubated with two differentially ATTO-labeled DNA substrates, which contain either one central fully methylated or fully hydroxymethylated CpG site (FMB-ATTO700 or FhMB-ATTO550, respectively), in direct competition. Samples were subjected to 6 % non-denaturing PAGE and analyzed with a Typhoon scanner (GE Healthcare). The first, second and third columns show the scans for GFP/YFP, ATTO700 and ATTO550 fluorescence, respectively. The overlay of the two ATTO channels is shown in the fourth column (FMB: red, FhMB:green).



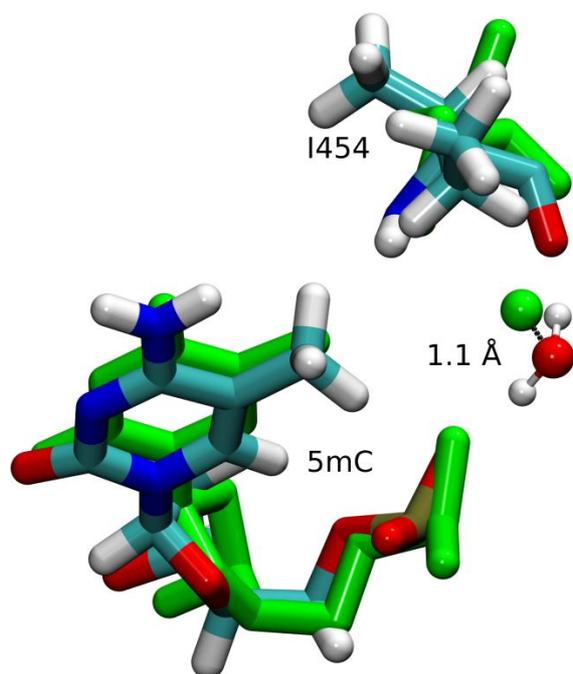
Supplementary Figure S2. DNA binding specificity of Uhrf1. Relative DNA/Uhrf1 ratios are shown for two differentially labeled fluorescent DNA substrates in direct competition. (A) Binding of Uhrf1 to DNA substrates containing no CpG site or one central hemimethylated CpG site (noCGB versus HMB, respectively). (B) Binding of Uhrf1 to DNA substrates containing one central un- or hemimethylated CpG site (UMB versus HMB, respectively). Results are shown as means of three independent experiments with standard deviation error bars. DNA substrates were prepared by hybridization as described in the main text, except for noCGB, which was prepared by primer extension as described previously [1]. See Supplementary Tables 1 and 2 for DNA oligonucleotide sequences and purification grade of the used substrates.



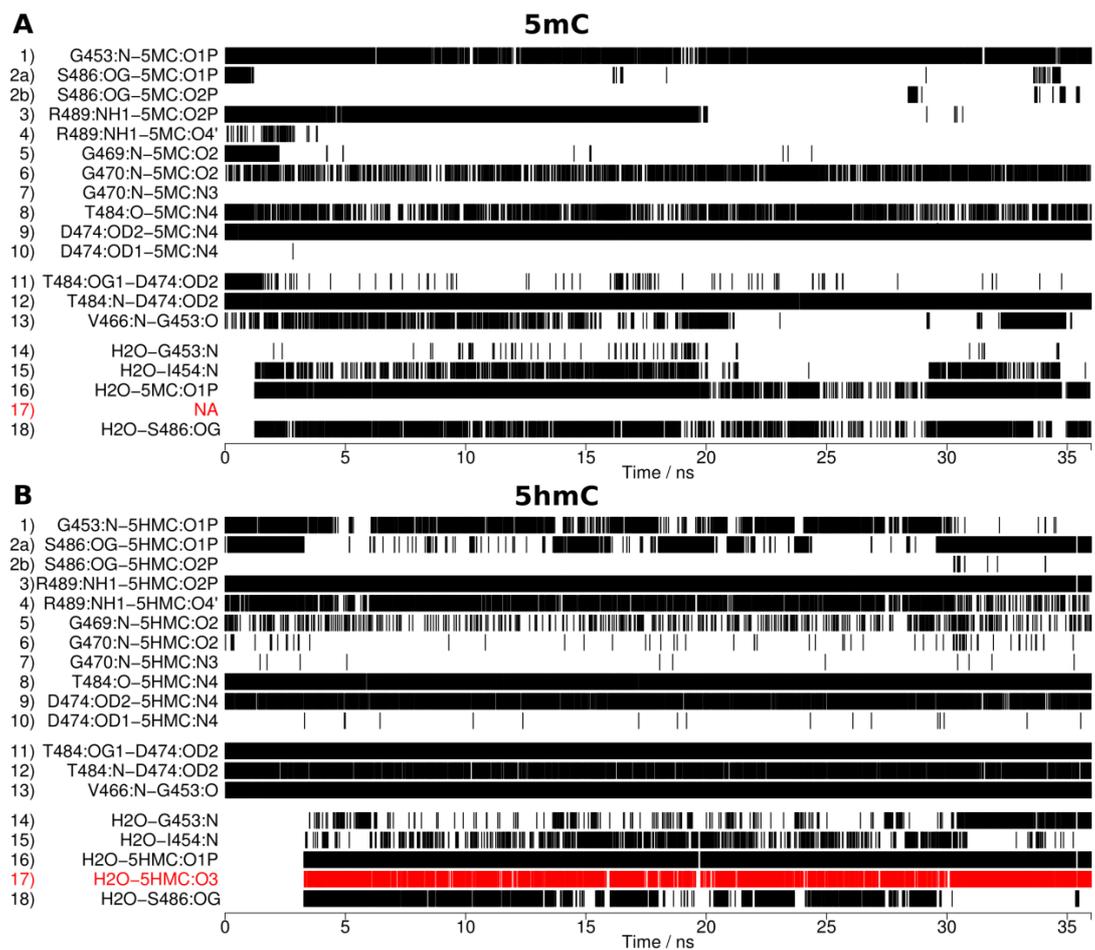
Supplementary Figure S3. Atom-positional root-mean-square deviation of the protein and DNA backbone atoms during the simulations. The terminal DNA and protein residues were excluded from the calculations in the “subset” sets (red and black lines).



Supplementary Figure S4. Atom-positional root-mean-square fluctuations of the protein (A, C) and both DNA strands (B, D) during two simulation periods. Note that both structures show the same flexibility pattern during both simulation periods and are overall stable during both periods. This is in agreement with the RMSD data in Figure S3, which shows that equilibration is reached after 30 ns of simulation time.



Supplementary Figure S5. Superposition of the equilibrated 5mC structure after simulation (atom-name specific coloring) and the crystal structure (PDB-ID:3fde [2], green). The 5mC nucleotide, the residue I454 of the SRA binding pocket and the conserved water molecule are shown. Note that the distance between the oxygen atoms of the conserved water molecules in the two structures is only 1.1 Å.



Supplementary Figure S6. Molecular dynamics simulations of the Uhrf1 SRA domain in complex with 5mC (A) and 5hmC (B) containing DNA in 0.5 M NaCl. Hydrogen bond occurrences during the simulation of the SRA:DNA complex using a concentration of 0.5 M NaCl.

Supplementary References

1. Frauer C, Leonhardt H (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res* 37: e22.
2. Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 455: 826-829.

2.6 Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells

Np95 interacts with *de novo* DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells

Daniela Meilinger^{1*}, Karin Fellingner^{1*}, Sebastian Bultmann¹, Ulrich Rothbauer¹, Ian Marc Bonapace², Wolfgang E.F. Klinkert³, Fabio Spada¹⁺ & Heinrich Leonhardt¹⁺⁺

¹Department of Biology II, Center for Integrated Protein Science Munich, Ludwig Maximilians University Munich, Planegg-Martinsried, Germany, ²Department of Structural and Functional Biology, University of Insubria, Busto Arsizio, Italy, and ³Department of Neuroimmunology, Max Planck Institute of Neurobiology, Am Klopferspitz, Martinsried, Germany

 This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits distribution, and reproduction in any medium, provided the original author and source are credited. This license does not permit commercial exploitation without specific permission.

Recent studies have indicated that nuclear protein of 95 kDa (Np95) is essential for maintaining genomic methylation by recruiting DNA methyltransferase (Dnmt) 1 to hemi-methylated sites. Here, we show that Np95 interacts more strongly with regulatory domains of the *de novo* methyltransferases Dnmt3a and Dnmt3b. To investigate possible functions, we developed an epigenetic silencing assay using fluorescent reporters in embryonic stem cells (ESCs). Interestingly, silencing of the cytomegalovirus promoter in ESCs preceded DNA methylation and was strictly dependent on the presence of either Np95, histone H3 methyltransferase G9a or Dnmt3a and Dnmt3b. Our results indicate a regulatory role for Np95, Dnmt3a and Dnmt3b in mediating epigenetic silencing through histone modification followed by DNA methylation.

Keywords: DNA methylation; histone modification; epigenetics; silencing; Uhrf1

EMBO reports (2009) 10, 1259–1264. doi:10.1038/embo.2009.201

¹Department of Biology II, Center for Integrated Protein Science Munich (CIPSM), Ludwig Maximilians University Munich, Großhaderner Street 2, 82152 Planegg-Martinsried, Germany

²Department of Structural and Functional Biology, University of Insubria, Via da Giussano 12, 21052 Busto Arsizio (VA), Italy

³Department of Neuroimmunology, Max Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany

*These authors contributed equally to this work

+Corresponding author. Tel: +49 89 218074230; Fax: +49 89 218074236; E-mail: f.spada@lmu.de

++Corresponding author. Tel: +49 89 218074232; Fax: +49 89 218074236; E-mail: h.leonhardt@lmu.de

Received 19 December 2008; revised 3 July 2009; accepted 4 August 2009; published online 2 October 2009

INTRODUCTION

In mammals, DNA methylation contributes to the establishment and maintenance of cell-type-specific gene expression programmes, imprinting, X-chromosome inactivation and genome stability (Bird, 2002). The majority of genomic methylation occurs at cytosine residues within CpG dinucleotides and is catalysed by the DNA methyltransferases (Dnmt) 1, 3a and 3b. Dnmt1 is responsible for maintaining genomic methylation, whereas Dnmt3a and Dnmt3b are mainly involved in *de novo* establishment of methylation patterns during cellular differentiation (Leonhardt *et al*, 1992; Li *et al*, 1992; Lei *et al*, 1996; Okano *et al*, 1999; Spada *et al*, 2007). Nuclear protein of 95 kDa (Np95; also known as Uhrf1) has recently been identified as an essential co-factor for maintaining genomic methylation (Bostick *et al*, 2007; Sharif *et al*, 2007; Achour *et al*, 2008). *dnmt1*^{-/-} and *np95*^{-/-} embryonic stem cells (ESCs) and embryos have similar reduced levels of DNA methylation. In addition, Np95 interacts with Dnmt1, binds hemi-methylated CpG sites through its Set and Ring associated (SRA) domain and both Np95 and Dnmt1 accumulate at replication sites (Uemura *et al*, 2000; Bostick *et al*, 2007; Papait *et al*, 2007; Arita *et al*, 2008; Avvakumov *et al*, 2008; Hashimoto *et al*, 2008). Thus, it has been proposed that Np95 mediates maintenance of genomic methylation by recruiting Dnmt1 to hemi-methylated CpG sites generated during replication.

Here, we investigated a possible involvement of Np95 in epigenetic regulation beyond its role in Dnmt1-mediated maintenance of DNA methylation. We found that Np95 interacts with the *de novo* methyltransferases, Dnmt3a and Dnmt3b, and mediates promoter silencing before DNA methylation is detected.

RESULTS AND DISCUSSION

Np95 interacts with Dnmt3a and Dnmt3b

Immunoprecipitation experiments showed that different isoforms of both *de novo* methyltransferases Dnmt3a and Dnmt3b interact

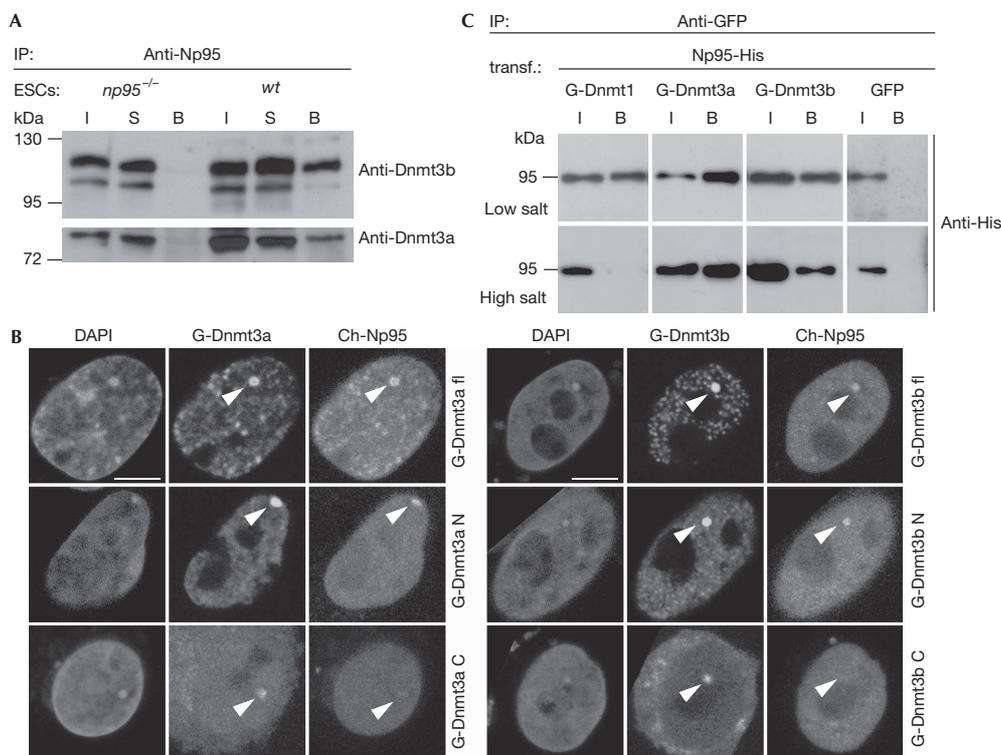


Fig 1 | Np95 interacts with *de novo* methyltransferases Dnmt3a and Dnmt3b. (A) Co-immunoprecipitation of Dnmt3a and Dnmt3b with Np95 in wt and *np95^{-/-}* E14 ESCs. The Dnmt3a2 isoform is shown in the lower panel. (B) F2H shows recruitment of Cherry-Np95 (prey) at the *lac* operator array (indicated by arrowheads) when GFP fusions of full-length Dnmt3a and Dnmt3b1 (G-Dnmt3a/b fl) or their amino-terminal regions (G-Dnmt3a/b N) are used as bait and not with their isolated C-terminal catalytic domains (G-Dnmt3a/b C). Scale bars, 5 μ m. (C) Co-immunoprecipitation of Np95-His with GFP-tagged Dnmt1, Dnmt3a and Dnmt3b1 (G-Dnmt) transiently co-expressed in HEK293T cells. Co-expression of GFP was used as the control. In the upper row, immunoprecipitations carried out in the presence of 150 mM NaCl throughout the procedure are shown, whereas in the lower row, immunoprecipitation and wash buffers were carried out using 300 and 500 mM NaCl, respectively. Two per cent of input and supernatant relative to bound fractions were loaded in (A) and (C). B, bound; Dnmt, DNA methyltransferase; ESCs, embryonic stem cells; F2H, fluorescent two-hybrid assay; GFP, green fluorescent protein; HEK293T, human embryonic kidney 293T; I, input; Np95, nuclear protein of 95 kDa; S, supernatant; wt, wild type.

with Np95 in wild-type (wt) ESCs, including the more abundant Dnmt3a2 and Dnmt3b1 (Fig 1A). Furthermore, using a green fluorescent protein (GFP) trap (Rothbauer et al, 2008), we co-immunoprecipitated endogenous Dnmt1 and isoforms of Dnmt3a and Dnmt3b with a GFP-Np95 fusion construct transiently expressed in *np95^{-/-}* ESCs and, vice versa, endogenous Np95 co-immunoprecipitated with GFP-Dnmt3a or GFP-Dnmt3b1 fusions in *dnmt3a* and *3b* double knockout (DKO) ESCs (supplementary Fig S1A,B online). In addition, we observed co-immunoprecipitation of endogenous DNMT3b and inverted CCAAT box binding protein of 90 kDa—the human homologue of Np95—from human embryonic kidney 293T (HEK293T) cell extracts (supplementary Fig S1C online). We confirmed the interaction of Np95 with Dnmt3a/b by using a recently developed fluorescent two hybrid assay (F2H; Zolghadr et al, 2008). GFP-Dnmt3 fusion constructs were used as bait by tethering them to a *lac* operator array present in baby hamster kidney (BHK) cells, so that the array was visible as a distinct

nuclear spot of enriched GFP fluorescence (Fig 1B). A Cherry-Np95 fusion (prey) accumulated at this spot only when GFP fusions of full-length Dnmt3a and Dnmt3b1 or their amino-terminal regions were used as bait and not when their isolated Carboxy-terminal catalytic domains were used. We further mapped the interaction of Np95 with Dnmt3a/b through co-immunoprecipitation of deletion constructs and isolated domains transiently expressed in HEK293T cells (supplementary Fig S2 online). The results were consistent with those produced by F2H: the N-terminal regions of Dnmt3a and Dnmt3b1, but not their C-terminal catalytic domains, interacted with Np95. Deletion of the PHD or PWWP domains of Dnmt3a and Dnmt3b did not eliminate the interaction with Np95. We then determined the domains of Np95 involved in this interaction. We found that the SRA domain and the N-terminal 298 amino acids of Np95, which include the ubiquitin-like domain, interacted with Dnmt3a and Dnmt3b1, whereas the PHD domain and the C-terminal 132 amino acids, including the Ring domain, did not. Furthermore,

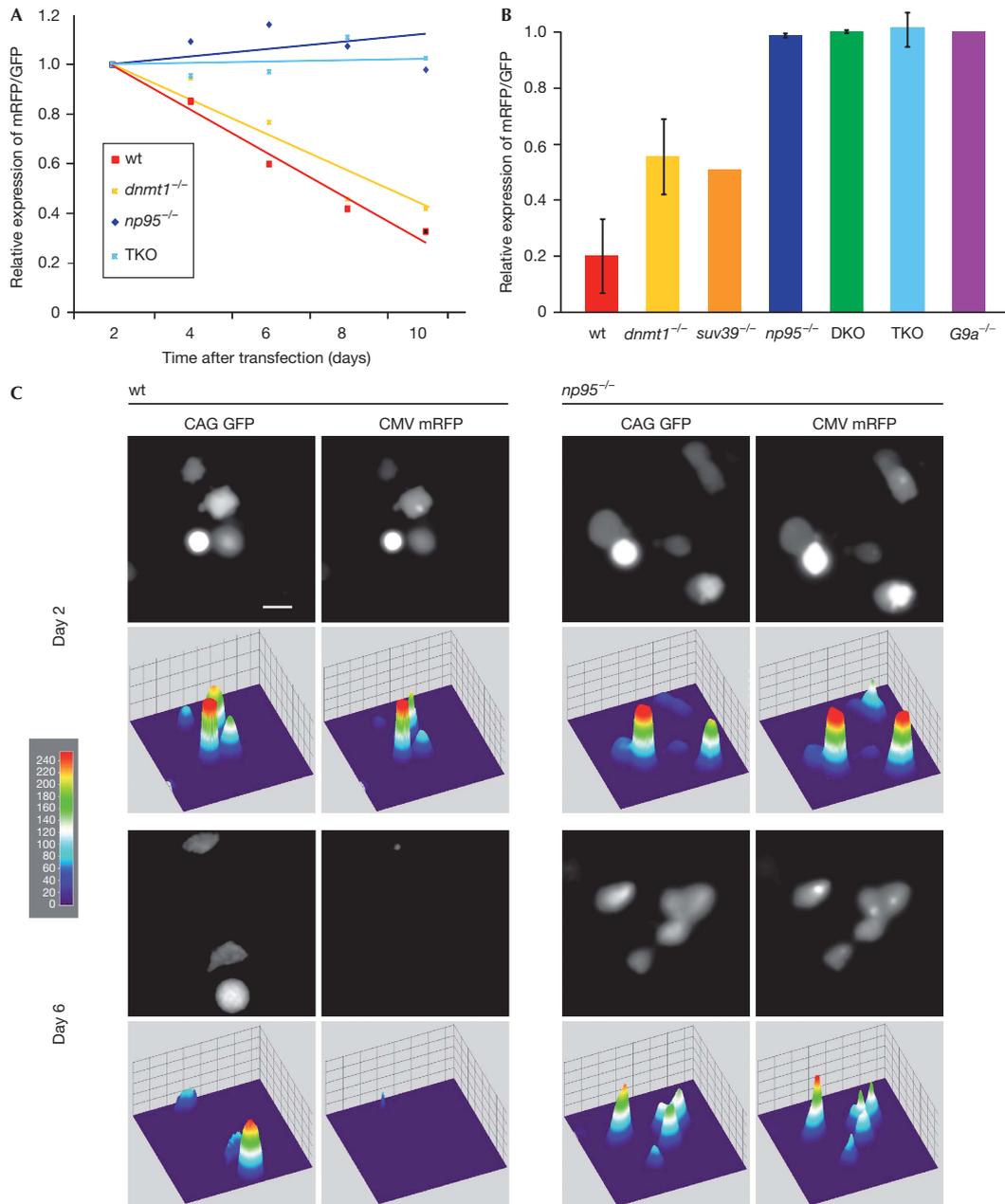


Fig 2 | Promoter silencing activity in wild-type and mutant ESCs. ESCs with the indicated phenotypes were transiently co-transfected with CMV promoter-driven mRFP and CAG promoter-driven GFP reporter constructs. Between 90 and 150 images per sample were acquired either (A) every second day after transfection from a single experiment or (B) only at days 2 and 7–10 after transfection from 3–5 independent experiments. Relative levels of red over green fluorescence are shown with values for day 2 (first day of imaging) set to 1 (A). Wild-type J1 and E14 cells gave similar results (data not shown). *suv39*^{-/-} stands for Suv39h1/2 double null ESCs. (C) Representative images of wt and *np95*^{-/-} E14 ESCs co-transfected as in (A) and (B) (upper panels) and respective heat map intensity plots (lower panels). Scale bar, 15 μm. CAG, CMV early enhancer/chicken β actin promoter; CMV, cytomegalovirus promoter; DKO, double knockout; Dnmt, DNA methyltransferase; ESCs, embryonic stem cells; GFP, green fluorescent protein; mRFP, monomeric red fluorescent protein; Np95, nuclear protein of 95 kDa; TKO, triple knockout; wt, wild type.

we observed co-immunoprecipitation of endogenous Np95 with GFP-Dnmt3a and GFP-Dnmt3b transiently expressed in *dnmt1*^{-/-} ESCs, indicating that Dnmt3a and Dnmt3b interact with Np95 independently of Dnmt1 (supplementary Fig S1D online). To compare the relative association between endogenous Np95 and Dnmts, we re-probed the blot in Fig 1A with a Dnmt1 antibody and observed a substantially weaker signal for the co-immunoprecipitated Dnmt1 relative to the input than in the case of Dnmt3a2 and Dnmt3b1 (supplementary Fig S3A online). To compare further the stability of Np95 interactions with the Dnmts, we transiently co-expressed Np95-His with GFP-Dnmt1, GFP-Dnmt3a or GFP-Dnmt3b1 in HEK293T cells and immunoprecipitated with the GFP trap in the presence of different salt concentrations (Fig 1C; supplementary Fig S3B online). Interestingly, under high salt conditions, the interaction between Np95-His and GFP-Dnmt1 was lost, whereas co-immunoprecipitation of GFP-Dnmt3a and GFP-Dnmt3b1 remained relatively unaffected. These data clearly indicate that Np95 interacts more strongly with the *de novo* methyltransferases, Dnmt3a and Dnmt3b, than with Dnmt1.

Np95, Dnmt3a/3b and G9a mediate epigenetic silencing

As DNA methylation has a central role in epigenetic silencing, we investigated the requirement of DNA methyltransferases and Np95 for promoter silencing in ESCs. We found that, on transient transfection of wt ESCs, constructs driven by the cytomegalovirus (CMV) promoter were rapidly silenced, as opposed to longer-lasting expression of constructs driven by the chimeric CMV early enhancer/chicken β actin (CAG) promoter (Fig 2), which is consistent with the popularity of the CAG promoter for stable transgene expression in ESCs and mice. We then established an epigenetic silencing assay based on this observation. ESCs were co-transfected with two distinct plasmids, one expressing monomeric red fluorescent protein (mRFP) under the CMV promoter, the other expressing GFP driven by the CAG promoter. mRFP and GFP expression was monitored after transfection for up to ten days by using automated image acquisition and quantification of fluorescent signals (supplementary Fig S4A online). The ratio between mRFP and GFP expression declined steadily in wt ESCs, reflecting preferential silencing of the CMV promoter (Fig 2). By contrast, DKO ESCs and ESCs lacking all three major DNA methyltransferases (*dnmt1*, *3a* and *3b* triple knockout) showed no preferential silencing of the CMV promoter. Surprisingly, *np95*^{-/-} ESCs were also unable to silence the CMV promoter, whereas *dnmt1*^{-/-} ESCs showed only partly reduced silencing under these conditions. Similar results were obtained on swapping GFP and mRFP reporter sequences, ruling out potential artefacts due to differences in their coding sequences or stability of the reporter proteins (supplementary Fig S4B online). Thus, despite expressing a full complement of DNA methyltransferases, ESCs lacking Np95 are as deficient in promoter silencing activity as ESCs lacking all three major Dnmts. We next investigated whether silencing of the CMV promoter correlates with CpG methylation. Interestingly, promoter methylation was detected only ten days after transfection and was lower in *np95*^{-/-} than in wt ESCs, whereas none of the *dnmt* mutant ESCs showed appreciable DNA methylation (Fig 3; supplementary Fig S5A online). At the same time no obvious methylation was detected in any of the ESC lines within the CpG island of the CAG promoter construct (supplementary Fig S5B online). Thus, CMV promoter silencing depends on the

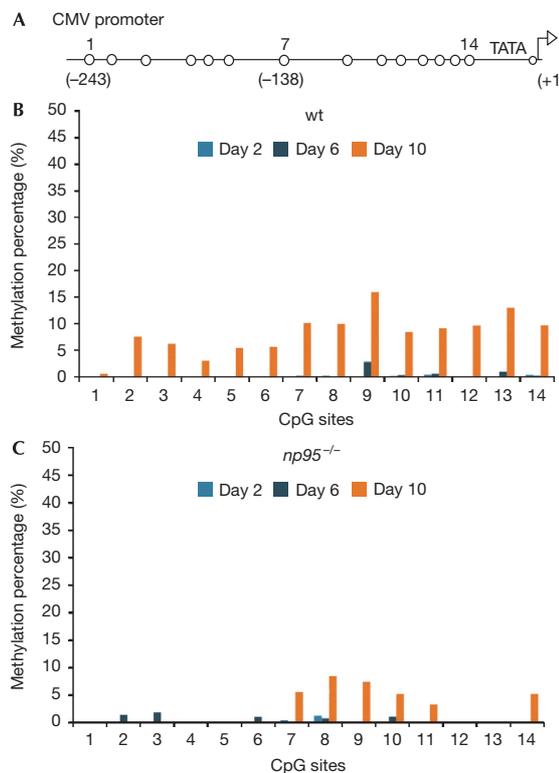


Fig 3 | Methylation of the CMV promoter 2, 6 and 10 days after transfection. Wild-type and *np95*^{-/-} ESCs were transfected as in Fig 2 and GFP-positive cells were sequentially sorted at the indicated days after transfection. Total DNA was isolated from sorted cells and bisulphite-treated. A proximal part of the CMV promoter was amplified and subjected to pyrosequencing. (A) Schematic drawing of the 14 proximal CpG sites analysed (shown as open circles). The numbers above correspond to the CpG sites shown in (B) and (C) and numbers in brackets refer to the position of CpG sites with respect to the transcription start site. (B,C) Methylation percentage at individual CpG sites for (B) wt and (C) *np95*^{-/-} ESCs as measured by pyrosequencing. CMV, cytomegalovirus; ESCs, embryonic stem cells; GFP, green fluorescent protein; wt, wild type.

presence of both Np95 and *de novo* Dnmts, but ensues well before *de novo* methylation of the promoter is detected. This prompted us to investigate the involvement of repressive histone methylation as a possible mechanism for the observed silencing. We found that in the absence of histone H3 lysine 9 methyltransferases (H3K9MTs), G9a or Suv39h1/2, silencing of the CMV promoter was completely abolished or reduced, respectively, indicating that G9a and, in part, Suv39h1/2 are also required for silencing (Fig 2B).

The results shown here indicate that Np95 interacts with Dnmt3a and Dnmt3b and mediates silencing of the CMV promoter by mechanisms that are, at least initially, independent of *de novo* DNA methylation. Importantly, our data also show the involvement of H3K9MTs, G9a and Suv39h1/2 in CMV promoter silencing. H3K9MTs were reported to associate with *de novo*

Dnmts, and major satellite repeats were found to be hypomethylated in ESCs lacking either Suv39h1/2 or Dnmt3 enzymes. However, major satellite transcript levels were altered only in Suv39h1/2-deficient cells and not in Dnmt3-deficient cells (Fuks et al, 2003; Lehnertz et al, 2003; Martens et al, 2005). A recent study showed that G9a, Dnmt1, Dnmt3a and Dnmt3b are required for normal methylation at long terminal repeats of endogenous retrotransposable elements, although transcription of these elements was increased in Dnmt-deficient ESCs, but not G9a-deficient ESCs (Dong et al, 2008). Furthermore, recent studies have shown that G9a interacts with Dnmt3a and Dnmt3b and mediates *de novo* methylation of the *oct4*, *nanog* and *dnmt3l* promoters on retinoic-acid-induced differentiation of ESCs (Feldman et al, 2006; Li et al, 2007; Epsztejn-Litman et al, 2008). However, two of these studies showed that neither G9a nor *de novo* Dnmts are required to silence the *oct4* promoter, and G9a was also found to be dispensable for silencing the *nanog* and *dnmt3l* promoters (Feldman et al, 2006; Epsztejn-Litman et al, 2008). In the third study, *nanog*, but not *oct4*, was shown to be silenced in differentiating ESCs lacking both Dnmt3a and Dnmt3b (Li et al, 2007). We found that silencing of *oct4* during embryoid body differentiation is largely independent from the presence of Np95 as well as all three major Dnmts, and occurs in the absence of DNA methylation (D.M., F.S., S.B., and H.L., unpublished data). These data, together with our findings on silencing of the CMV promoter in ESCs, indicate that Dnmts, Np95 and H3K9MTs mediate silencing through many mechanisms that do not necessarily involve DNA methylation and might depend on the presence of different *cis* elements and an intricate interplay with other epigenetic and transcription factors. Interestingly, Np95 was recently shown to interact with G9a (Kim et al, 2009) and here we show that silencing of the CMV promoter in ESCs strictly depends on Np95 and on *de novo* Dnmts as well as G9a. Taken together these observations suggest that Np95, *de novo* Dnmts and G9a might be involved in a common silencing pathway.

In summary, our data clearly support a crucial role of Np95 in epigenetic silencing mediated by *de novo* DNA and histone methyltransferases, and make Np95 an attractive target for epigenetic reprogramming strategies.

METHODS

Cell culture and transfection. HEK293T cells, BHK cells and ESCs were cultured and transfected as described by Schermelleh et al (2007), except FuGENE HD (Roche, Mannheim, Germany) was used for transfection of ESCs. The *dnmt1*^{-/-} J1 ESCs used in this study were homozygous for the c allele (Lei et al, 1996). BHK cells were co-transfected on glass coverslips with GFP-Dnmt3 and Cherry-Np95 constructs using Transfectin (Bio-Rad, Munich, Germany) according to the manufacturer's instructions. Cell fixation and microscopy were carried out as described by Zolghadr et al (2008).

Co-immunoprecipitation. ESCs and HEK293T cell extracts were prepared in lysis buffer (20 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 2 mM phenylmethyl sulphonyl fluoride and 0.5% NP40) containing 150 or 300 mM NaCl (high-salt condition) and diluted with lysis buffer without NP40. GFP trap (Rothbauer et al, 2008) and a specific rabbit antiserum (Citterio et al, 2004) were used for immunoprecipitation of GFP fusions and endogenous Np95, respectively. GFP trap and protein G beads (Sigma, Taufkirchen,

Germany) were washed with dilution buffer containing increasing salt concentrations (150 and 300 mM, or 300 and 500 mM NaCl for the high-salt condition) and re-suspended in SDS-PAGE sample buffer. The following mouse monoclonal antibodies were used for immunoblotting: anti-His (C-terminal, Invitrogen, Karlsruhe, Germany), anti-Dnmt3a (clone 64B1446, Imgenex, San Diego, CA, USA) and anti-Dnmt3b (clone 52A1018, Abcam, Cambridge, UK). Np95 was detected with the same antiserum used for immunoprecipitation and a rabbit antiserum was used for detection of Dnmt1 (Grohmann et al, 2005). Horseradish peroxidase-conjugated rabbit anti-mouse or goat anti-rabbit secondary antibodies (Sigma) and ECL Plus reagent (GE Healthcare, Munich, Germany) were used for detection.

Silencing assay. ESCs were co-transfected with pCAG-eGFP-IB and pCMV-mRFP as described above and images from live cells were acquired at the indicated time points with an InCell Analyser 1000 (GE Healthcare) using a ×20 air objective (NA=0.45) and standard filter settings for GFP and RFP. A total of 90–150 images were acquired for each channel, using the same exposure time throughout the time course. Cells were passaged every second day and images were taken 4–5 h after seeding. Images were analysed using ImageJ v1.42a software. To calculate fluorescent reporter expression, pictures were processed using a Gaussian blur algorithm (radius (sigma)=2), and a threshold for maximal signal and minimal background coverage was adjusted and applied to each channel (supplementary Fig S4A online). The threshold was converted into area selection and the total size of the selected area was measured.

DNA methylation analysis. ESCs were transfected as in the silencing assay with pCAG-eGFP-IB and pCMV-mRFP, and GFP-positive cells were sequentially sorted with a FACS Vantage or FACS Aria II (Becton Dickinson, Heidelberg, Germany) at days 2, 6 and 10 after transfection. After each sorting, total DNA was isolated using the QIAamp DNA Mini kit (Qiagen, Hilden, Germany) and bisulphite treated with the EZ DNA Methylation-Gold kit (Zymo research, Orange, CA, USA). The following primers were used for PCR amplification: CMV-forward TGGGAT TTTTTATTTGGTAGT; CMV-reverse ATGGGAGTTTGGTTTGG TATTA; CAG-forward GGAGAGGTGAGGAGGTAGTTAATTAGA and CAG-reverse CCCCAAACCCCTCAAACCTT. Pyrosequencing was carried out by Varionostic GmbH (Ulm, Germany).

Supplementary information is available at *EMBO reports* online (<http://www.emboreports.org>).

ACKNOWLEDGEMENTS

We are grateful to the following colleagues for providing ESCs: M. Muto and H. Koseki for wt and *np95*^{-/-} E14; E. Li and T. Chen for wt, *dnmt1*^{-/-} and DKO J1; Masaki Okano for TKO J1; Gunnar Schotta for *Suv39h1/2* dn; Y. Shinkai and T. Jenuwein for wt and *G9a*^{-/-} TT2. We thank L. Schermelleh for help with image processing and K. Zolghadr for assistance with the InCell Analyzer. This study was supported by the Nanosystems Initiative Munich and Biolmaging Network Munich, and by grants from the Deutsche Forschungsgemeinschaft to H.L.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

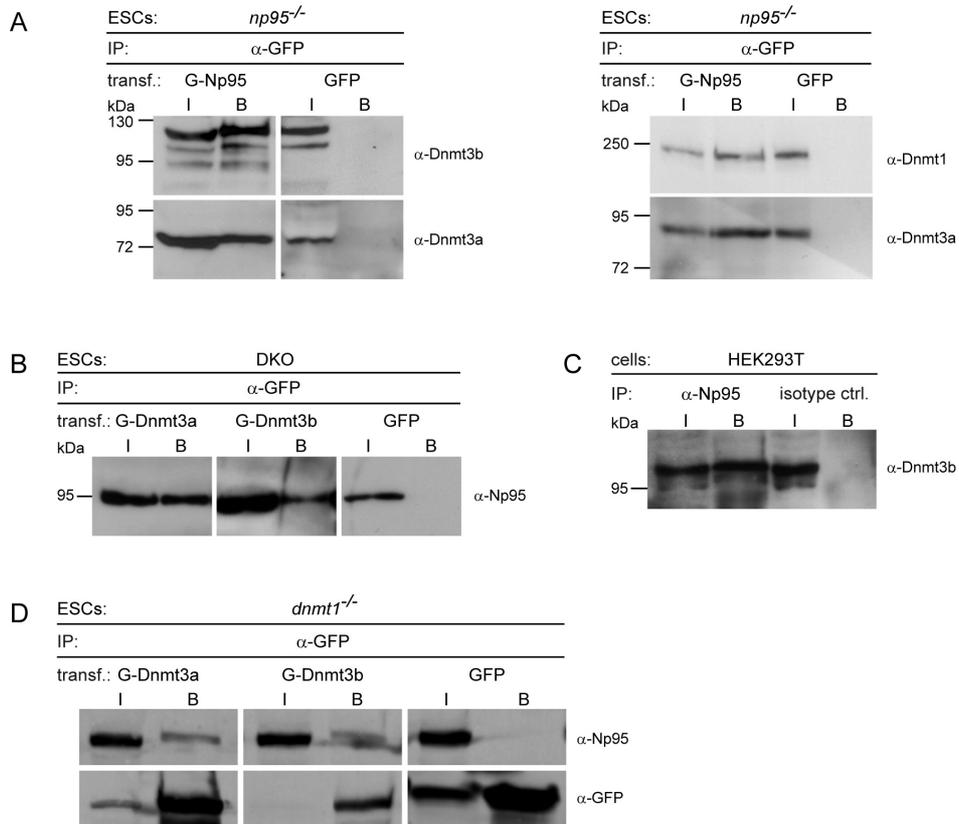
REFERENCES

Achour M et al (2008) The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene* **27**: 2187–2197

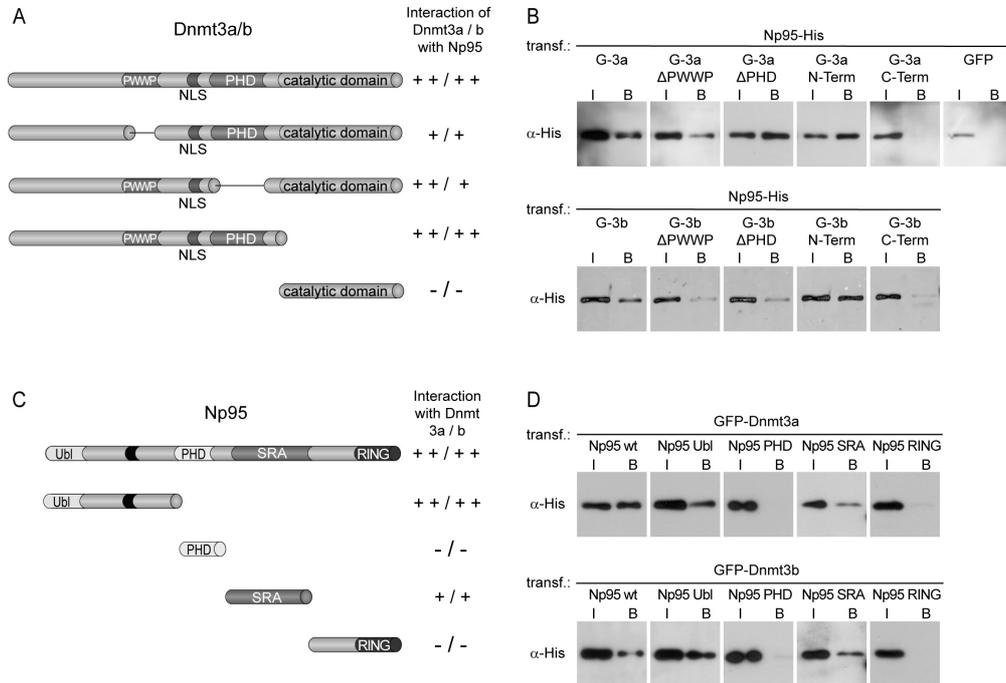
- Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**: 818–821
- Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455**: 822–825
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* **16**: 6–21
- Bostick M, Kim JK, Esteve P-O, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**: 1760–1764
- Citterio E, Papait R, Nicassio F, Vecchi M, Gomiero P, Mantovani R, Di Fiore PP, Bonapace IM (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol Cell Biol* **24**: 2526–2535
- Dong KB et al (2008) DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J* **27**: 2691–2701
- Epsztejn-Litman S et al (2008) *De novo* DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat Struct Mol Biol* **15**: 1176–1183
- Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, Cedar H, Bergman Y (2006) G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* **8**: 188–194
- Fuks F, Hurd PJ, Depluis R, Kouzarides T (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* **31**: 2305–2312
- Grohmann M, Spada F, Schermelleh L, Alenina N, Bader M, Cardoso MC, Leonhardt H (2005) Restricted mobility of Dnmt1 in preimplantation embryos: implications for epigenetic reprogramming. *BMC Dev Biol* **5**: 18
- Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455**: 826–829
- Kim JK, Esteve PO, Jacobsen SE, Pradhan S (2009) UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res* **37**: 493–505
- Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* **13**: 1192–1200
- Lei H, Oh S, Okano M, Juttermann R, Goss K, Jaenisch R, Li E (1996) *De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**: 3195–3205
- Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**: 865–873
- Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915–926
- Li JY et al (2007) Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. *Mol Cell Biol* **27**: 8748–8759
- Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* **24**: 800–812
- Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**: 247–257
- Papait R, Pistore C, Negri D, Pecoraro D, Cantarini L, Bonapace IM (2007) Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Mol Biol Cell* **18**: 1098–1106
- Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H (2008) A versatile nanotrapp for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* **7**: 282–289
- Schermelleh L, Haemmer A, Spada F, Rosing N, Meilinger D, Rothbauer U, Cristina Cardoso M, Leonhardt H (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res* **35**: 4301–4312
- Sharif J et al (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**: 908–912
- Spada F, Haemmer A, Kuch D, Rothbauer U, Schermelleh L, Kremmer E, Carell T, Langst G, Leonhardt H (2007) DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J Cell Biol* **176**: 565–571
- Uemura T, Kubo E, Kanari Y, Ikemura T, Tatsumi K, Muto M (2000) Temporal and spatial localization of novel nuclear protein NP95 in mitotic and meiotic cells. *Cell Struct Funct* **25**: 149–159
- Zolghadr K, Mortusewicz O, Rothbauer U, Kleinhans R, Goehler H, Wanker EE, Cardoso MC, Leonhardt H (2008) A fluorescent two-hybrid (F2H) assay for direct visualization of protein interactions in living cells. *Mol Cell Proteomics* **7**: 2279–2287



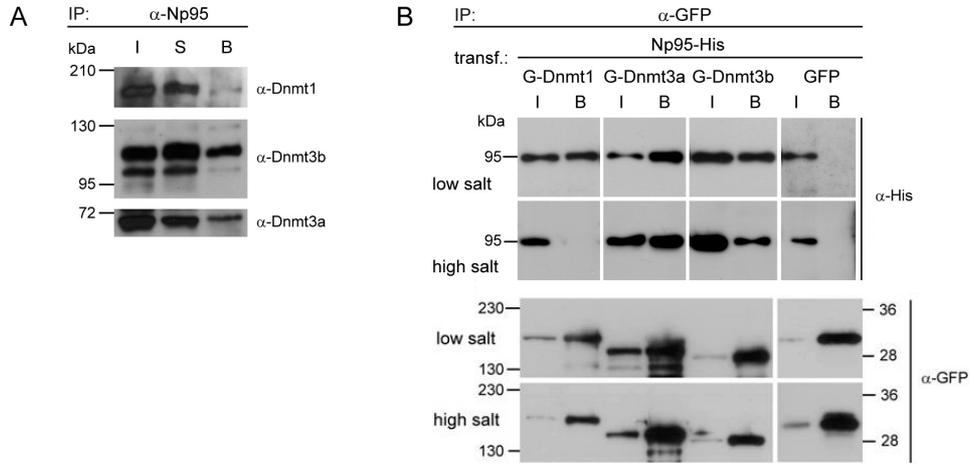
EMBO reports is published by Nature Publishing Group on behalf of European Molecular Biology Organization. This article is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 License. [<http://creativecommons.org/licenses/by-nc-sa/3.0/>]



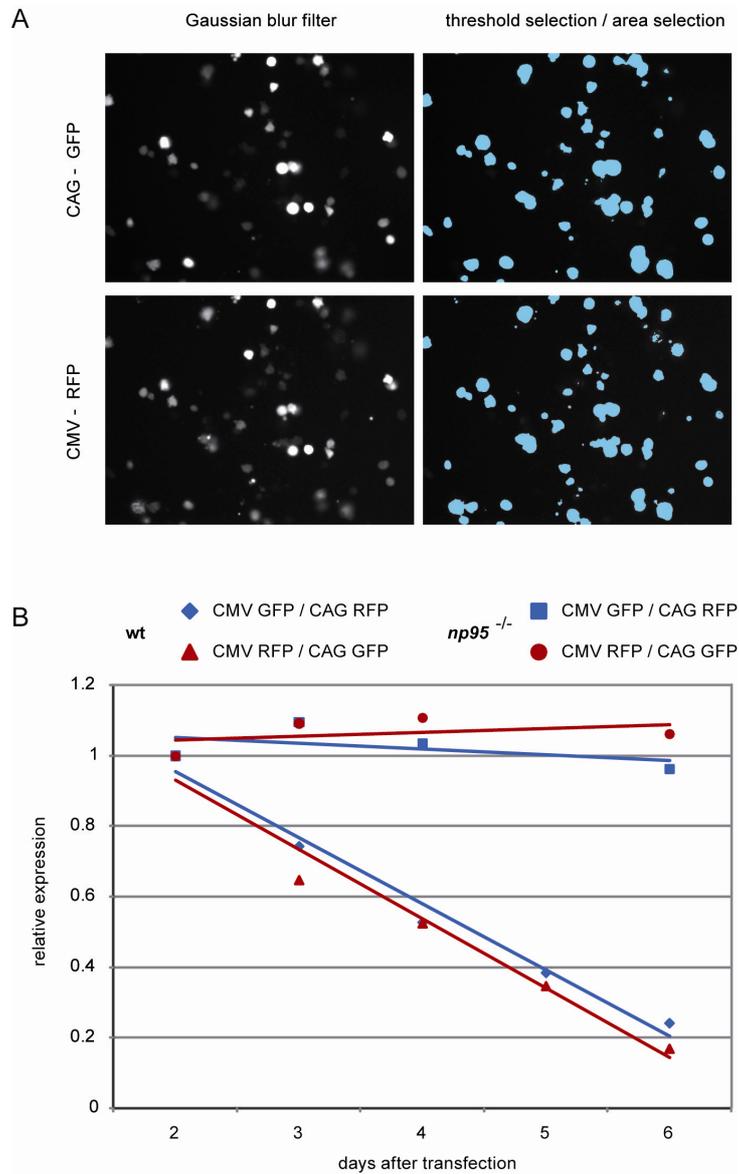
Supplementary Figure S1. Np95 interacts with *de novo* methyltransferases Dnmt3a and 3b. **(A)** Co-immunoprecipitation of endogenous Dnmt3a2 (left and right), Dnmt3b isoforms (left) and Dnmt1 (right) with GFP-Np95 transiently expressed in *np95^{-/-}* ESCs. Left and right panels are from independent experiments where 2 and 4% of input (I) relative to bound (B) fractions was loaded, respectively. **(B)** Co-immunoprecipitation of endogenous Np95 with either GFP-Dnmt3a (left panel) or GFP-Dnmt3b1 (central panel) transiently expressed in DKO ESCs. Transient expression of GFP was used as control (right panel). GFP and GFP fusions were immunoprecipitated with GFP-trap as in experiments shown in Fig. 1B. 2% of input (I) relative to bound (B) fractions was loaded. **(C)** Co-immunoprecipitation of endogenous ICBP90/UHRF1 and DNMT3b in HEK293T cells. Antibodies to mouse proteins cross-react with the respective human homologues. 4% of input (I) relative to bound (B) fractions was loaded. **(D)** Co-immunoprecipitation of endogenous Np95 with either GFP-Dnmt3a or GFP-Dnmt3b1 (lower panel) transiently expressed in *dnmt1^{-/-}* ESCs.



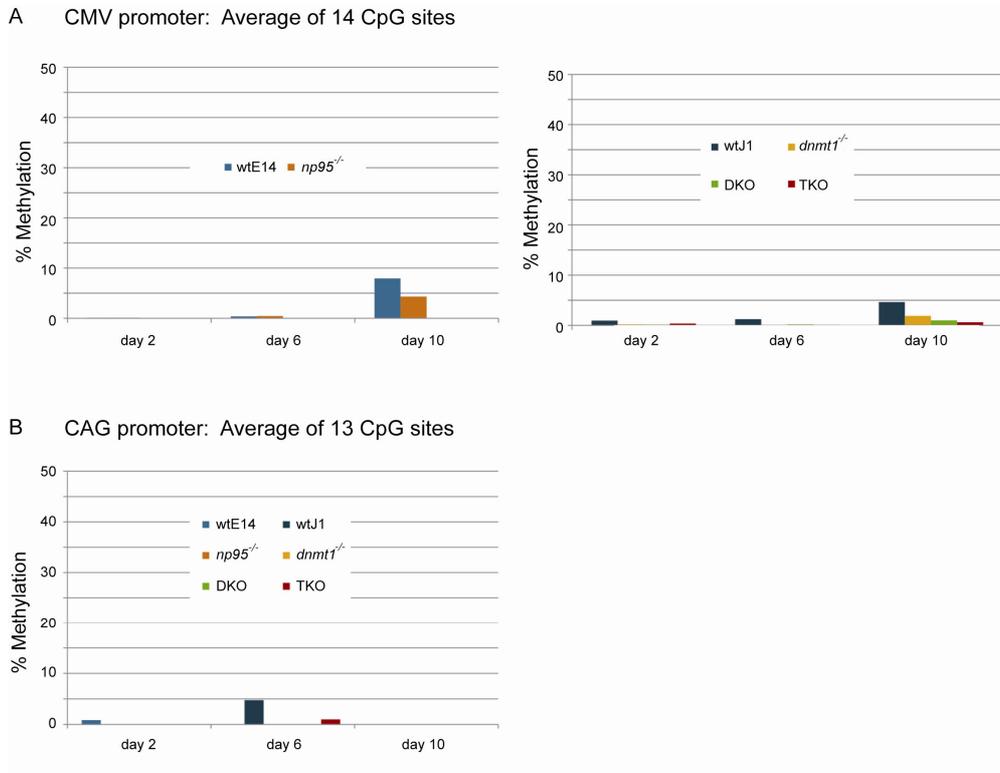
Supplementary Figure S2. Mapping the interaction domains of Dnmt3a/b and Np95. **(A)** Schematic representation of GFP-Dnmt3a/b fusion constructs used for mapping the interaction site with Np95 (N-terminal GFP tag is not shown). **(B)** Co-immunoprecipitation of Np95-His with GFP-Dnmt3 constructs (G-3a/b) from extracts of transiently transfected HEK293T cells. **(C)** Schematic representation of Np95-His constructs used for mapping the interaction site with Dnmt3a/b. **(D)** Co-immunoprecipitation of Np95-His domains shown in c with GFP-Dnmt3a/b constructs from extracts of transiently transfected HEK293T cells. G indicates the GFP fusion. The GFP-trap was used for all the immunoprecipitations in panels B and D. 0.5% of input (I) and 40% of bound (B) fractions were loaded. PWWP, domain with conserved pro-trp-trp-pro motif; NLS, nuclear localization signal; PHD, plant homeodomain; Ubl, ubiquitin-like domain; SRA, set and ring associated domain; RING, really interesting new gene domain. Results of mapping are scored by + or -.



Supplementary Figure S3. Relative stability of Np95 interactions with Dnmt1, 3a and 3b. **(A)** The Dnmt3a/3b blot in Fig. 1A (shown here as middle and lower panels) was reprobbed with an anti-Dnmt1 antibody (upper panel) to compare the relative amounts of endogenous DnmTs associated with Np95. The lower panel shows the Dnmt3a2 isoform. **(B)** The blots in Fig. 1C (shown here in the two upper panels) were reprobbed with an anti-GFP antibody (two lower panels) to reveal that similar amounts of each GFP construct were immunoprecipitated in low and high salt conditions. Both results point to a tighter association of Np95 with Dnmt3a and 3b as compared to Dnmt1.



Supplementary Figure S4. (A) Automated procedure for quantification of fluorescent signals from digital micrographs for the promoter silencing assay. A macro was written for the ImageJ software that applies a Gaussian blur filter (left panel) and signal thresholding (right panel) to raw images (data not shown) and then calculates the total signal area. (B) Silencing assay results are not affected by the choice of fluorescent reporter. wt and *np95*^{-/-} ESC were cotransfected with either CMV-driven mRFP and CAG-driven GFP (red) or CAG-driven mRFP and CMV-driven GFP (blue) expression constructs and the ratio of CMV- over CAG-driven fluorescence was quantified at the indicated time points after transfection as for Figure 2A.



Supplementary Figure S5. Methylation analysis of CMV and CAG promoters 2, 6 and 10 days after transfection. ESCs with the indicated genotypes were transfected, sorted, total DNA was isolated and bisulfite treated as for Fig. 3. **(A)** The same proximal region of the CMV promoter was amplified and pyrosequenced as in Fig. 3. **(B)** A fragment of the CAG construct containing 13 CpG sites was amplified and subjected to pyrosequencing. The analyzed fragment spans across the 3' part of the promoter, the first exon and the 5' part of the first intron of the chicken β -actin gene and is part of a CpG island. Methylation percentages at individual CpG sites within the respective promoter sequences are averaged. The plot on the left of A was derived from the same data presented in Fig. 3C.

Supplementary methods

Plasmid construction. The CMV-driven enhanced GFP construct was from Clontech (pEGFP-C1). To generate the CMV-driven mRFP construct (pCMV-mRFP) the coding sequence for eGFP in pEGFP-C1 was replaced with that for mRFP from pRSETB-mRFP (Campbell et al, 2002; provided by Roger Tsien). To create CAG-driven eGFP, mRFP and mCherry expression constructs (pCAG-eGFP-IB, pCAG-mRFP-IB and pCAG-mCherry-IB, respectively) sequences coding for the respective fluorescent proteins from pEGFP, pRSETB-mRFP and pRSETB-mCherry (Shaner et al, 2004; also provided by R. Tsien) were inserted downstream to the CAG promoter in the pCAG-IRESblast vector (Chen et al, 2003). The expression construct for Np95-His was described previously (Citterio et al, 2004). To generate expression constructs for GFP-Np95, Ch-Np95, GFP-Dnmt3a and GFP-Dnmt3b1 the sequences coding for Np95, Dnmt3a or Dnmt3b1 were then transferred from the respective CMV promoter-driven constructs (Chen et al, 2003; Citterio et al, 2004) to either pCAG-eGFP-IB or pCAG-mCherry-IB downstream to sequences coding for the fluorescent protein. GFP-Dnmt3a and GFP-Dnmt3b1 deletion constructs were generated by overlap extension mutagenesis (Ho et al, 1989) to remove the following amino acids from Dnmt3a and 3b1, respectively: 278-343 and 223-287 (Δ PWWP); 485-582 and 435-532 (Δ PHD). GFP fusion constructs of N-terminal regions (aa 1-629 and 1-580) and C-terminal domains (aa 630-908 and 581-859) of Dnmt3a and 3b, respectively, were generated by PCR cloning using full length constructs as templates. All constructs were characterised by sequencing and immunoblotting.

Supplementary references

Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* 99(12): 7877-7882

Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* 23(16): 5594-5605

Citterio E, Papait R, Nicassio F, Vecchi M, Gomiero P, Mantovani R, Di Fiore PP, Bonapace IM (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol Cell Biol* 24(6): 2526-2535

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77(1): 51-59

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* 22(12): 1567-1572

2.7 Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers

Targeted transcriptional activation of silent *oct4* pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers

Sebastian Bultmann¹, Robert Morbitzer², Christine S. Schmidt¹, Katharina Thanisch¹, Fabio Spada¹, Janett Elsaesser², Thomas Lahaye^{2,*} and Heinrich Leonhardt^{1,*}

¹Department of Biology, Center for Integrated Protein Science Munich (CIPSM) and ²Department of Biology, Institute of Genetics, 82152 Planegg-Martinsried, Ludwig Maximilians University Munich, Germany

Received September 22, 2011; Revised and Accepted February 13, 2012

ABSTRACT

Specific control of gene activity is a valuable tool to study and engineer cellular functions. Recent studies uncovered the potential of transcription activator-like effector (TALE) proteins that can be tailored to activate user-defined target genes. It remains however unclear whether and how epigenetic modifications interfere with TALE-mediated transcriptional activation. We studied the activity of five designer TALEs (dTALEs) targeting the *oct4* pluripotency gene. In vitro assays showed that the five dTALEs that target distinct sites in the *oct4* promoter had the expected DNA specificity and comparable affinities to their corresponding DNA targets. In contrast to their similar in vitro properties, transcriptional activation of *oct4* by these distinct dTALEs varied up to 25-fold. While dTALEs efficiently upregulated transcription of the active *oct4* promoter in embryonic stem cells (ESCs) they failed to activate the silenced *oct4* promoter in ESC-derived neural stem cells (NSCs), indicating that as for endogenous transcription factors also dTALE activity is limited by repressive epigenetic mechanisms. We therefore targeted the activity of epigenetic modulators and found that chemical inhibition of histone deacetylases by valproic acid or DNA methyltransferases by 5-aza-2'-deoxycytidine facilitated dTALE-mediated activation of the epigenetically silenced *oct4* promoter in NSCs. Notably, demethylation of the *oct4* promoter occurred only if chemical inhibitors and dTALEs were applied together but not upon treatment with inhibitors or dTALEs only. These results show that

dTALEs in combination with chemical manipulation of epigenetic modifiers facilitate targeted transcriptional activation of epigenetically silenced target genes.

INTRODUCTION

The ability to specifically manipulate the expression of endogenous genes by engineered designer transcription factors has wide-ranging applications in basic and applied biology (1–4). Availability of suitable DNA-binding scaffolds that can be tailored to bind user-defined target sequences has been the major limitation in the generation and application of designer transcription factors. Recent studies however demonstrated that transcription activator-like effector proteins (TALEs) from the plant pathogenic bacterial genus *Xanthomonas* contain a DNA-binding domain that can be adjusted to bind any desired target sequence with high specificity (5–9). The TALE DNA-binding domain is composed of tandem arranged 33–35 amino acid repeats, with each repeat binding to one base (10,11). Base preferences of repeats are specified by residues 12 and 13, known as the repeat variable diresidues (RVDs), that determine preferential pairing with A (NI), C (HD), G (NK) and T (NG) nucleotides, respectively. The use of this TALE code facilitates the assembly of TALE repeat arrays that bind any desired DNA sequence (12).

A recent study investigated a large number of dTALEs and found that most, but not all, activated the desired target promoters (5). Notably, the epigenetically controlled *oct4* and *c-myc* gene could not be upregulated by their matching dTALEs, suggesting that epigenetic modifications affect dTALE-mediated gene activation.

We systematically investigated the application of dTALEs to the murine pluripotency gene *oct4* to clarify

*To whom correspondence should be addressed. Tel: +49 89 2180 74740; Fax: +49 89 2180 74702; Email: lahaye@biologie.uni-muenchen.de
Correspondence may also be addressed to Heinrich Leonhardt. Tel: +49 89 2180 74232; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2012. Published by Oxford University Press.

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

how epigenetic modifications affect their performance. The inspection of five dTALEs that bind to distinct DNA sequences within the *oct4* promoter revealed similar affinities to their DNA targets but up to 25-fold differences in their efficiency as transcriptional activators. Further studies revealed that dTALE-mediated activation of a silent *oct4* promoter in neural stem cells (NSCs) can be drastically improved by treatment with the histone deacetylase (HDAC) inhibitor valproic acid (VPA) and the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine (5azadC). These data suggest that chromatin modifications that are involved in transcriptional gene silencing, hinder dTALE-mediated gene activation and that simultaneous inhibition of HDACs and DNA methyltransferases may overcome this limitation of dTALE technology.

MATERIALS AND METHODS

Construction of plasmids

A Gateway cassette from pGWB5 (13) was amplified (forward primer: 5'-GGGGCGATCGCACAAGTTTGTACAAAAAAGCTGAACGAG-3'; reverse primer: 5'-GGGCGGCCGCAACCACTTTGTACAAGAAAGCTGAACG-3'), thereby adding *Asi*SI and *Not*I restriction sites. This fragment was cloned via *Asi*SI and *Not*I into pCAG_mCh (14) generating pCAG_mCh_GW. The VP16AD was amplified from RSV

E2F1-VP16 (15) (forward primer: 5'GGGGGTCTCTCACCATGGATCCTGCCCCCGACCGATGTCAGC-3'; reverse primer: 5'-GGGGGTCTCCCTTCTACCCACCGTACTCGTCAATTCCAAGG-3'), thereby adding a *Bam*HI restriction site to the 5' end and cloned into pENTR-D- TOPO (Invitrogen) generating pENTR-D-*Bam*HI_VP16AD. TALE repeat arrays were generated via multi-fragment cut-ligation using golden gate cloning (16) and ligated either into pENTR-D-TALE- Δ rep-*Bpi*I-A or pENTR-D-TALE- Δ rep-*Bpi*I-AC-VP16AD. All entry clones were transferred by LR recombination (Invitrogen) into the expression vector pCAG_mCh_GW.

The *oct4* reporter construct (*poct4*-GFP) was generated by inserting the *Xho*I/*Avr*II fragment of GOF-18 (17) which includes the basepairs -1 to -4716 upstream of the transcriptional start site of *oct4* together with a linker oligo (5'-CCTAGGTGAGCCGTCTTTCCACCA GGCCCCGGCTCGGGGTGCGATCGCCGCCCAT GG-3') into pGL-3 basic (Promega) cut with *Xho*I/*Nco*I. Subsequently, the Luciferase ORF was removed by cutting with *Kas*I/*Fse*I and the eGFP ORF (amplified with: forward primer: 5'-AAAGGCGCCAGTGAGCAA GGGCG-3'; reverse primer: 5'-AAAGGCCGGCCTTACTTGACAGCTCGTCC-3') was inserted.

The promoter mutants *TB83*, *TB68*, *TB60* and *TB31* were generated by site-directed mutagenesis using a *Asi*SI/*Aat*II derived sub-cloned *poct4*-GFP fragment as template with either forward primer: 5'-TCTCCACCC CCACAGCTCTGCTCCTTTGGGGAGGGAGAGGT GAAAC-3', 5'-GCTCTGCTCCTCCACCCACCCAGG GTTGGGGAGGGAGGGTGAAACCG-3', 5'-CC TCCACCCACCCAGGGCGGGCCTTGGGGAG

GGAGAGGTGAAACCG-3' or 5'-GGTCAAGGCTAG AGGGTGGGATTGGGTTGGGGAGGGAGAGGTG AAACCG-3' together with reverse primer: 5'-GAAACTG AGGCGAGCGCTATCTG-3', thereby deleting *TB83*, *TB68*; *TB60* and *TB31* and inserting them individually at the position of *TB31*.

Immunofluorescence staining

For immunostaining, ogNSCs were grown on cover slips and transiently transfected with the T-83VP16 construct for Oct4 stainings or untransfected for Pax6, Nestin and Olig2 stainings. Cells were fixed with 2.0% or 3.7% formaldehyde in phosphate-buffered saline (PBS) and permeabilized in PBS containing 0.2% Triton X-100. The Oct4 staining was performed using a goat primary antibody against the murine Oct4 (goat; 1:1000, Santa Cruz) and a secondary anti-goat antibody coupled to Alexa Fluor 647 (1:2000, Molecular Probes). The neural stem cell markers Pax6 (rabbit; 1:1000, Millipore), Nestin (mouse monoclonal, Rat-401; 1:10, Developmental Studies Hybridoma Bank, University of Iowa) and Olig2 (rabbit; 1:500, Millipore) were detected with secondary antibodies conjugated to Alexa Fluor 488 (Molecular Probes). The antibodies were diluted in PBS containing 0.02% Tween 20 and 2% bovine serum albumin (BSA). Cells were counterstained with DAPI and mounted in Vectashield (Vector Laboratories). Images were acquired with a Zeiss Axioplan 2 fluorescence microscope equipped with a Plan-NEOFLUAR 40 \times /1.3 oil objective (Zeiss).

Cell culture, transfection and fluorescence-activated cell sorting

HEK293T cells (18) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 50 μ g/ml gentamicin and 10% fetal bovine serum (FBS). For expression of fusion proteins, HEK293T cells were transfected with polyethylenimine (Sigma). ogESCs were cultured as described (19). ogNSCs were cultured in N2B27 medium supplemented with 20 ng/ml FGF-2 and EGF. NSCs and ESCs were transfected using Lipofectamin 2000 (Invitrogen) according to the manufacturer's instructions and sorted with a fluorescence-activated cell sorting (FACS) Aria II instrument (Becton Dickinson).

Generation of transgenic cell lines

ogESCs were generated by transfecting wt J1 ESCs (20) with the *poct4*-GFP reporter construct and repeated sorting for eGFP expression. Finally, single cell sorting was used to obtain a clonal transgenic cell line.

Derivation of NSCs from ESCs

ogESCs were differentiated into ogNSCs as previously described (21–23). In brief, 3.5×10^5 cells were plated in a 25 cm² culture flask with N2B27 medium containing 1000 U/ml of LIF (ESGRO, Millipore). The next day the medium was exchanged against N2B27 without LIF to initiate differentiation into the neural lineage. After 7 days cells were plated in Euromed-N (Euroclone)

supplemented with 20 ng/ml EGF and FGF2 (PeproTech). After 5 days, neurospheres were collected and plated in gelatin-coated flasks in N2B27 medium containing 20 ng/ml EGF and FGF2 to allow outgrowth of NSCs.

Treatment of ogNSCs with epigenetic inhibitors

VPA sodium salt (Sigma-Aldrich) was dissolved in PBS at a concentration of 250 mM and sterile filtered. 5-aza-2'-deoxycytidine (Sigma-Aldrich) was dissolved in PBS at a concentration of 30 mM. Trichostatin A (TSA; Sigma-Aldrich) was dissolved in dimethyl sulfoxide (DMSO) at a concentration of 5 mM. Cells were transfected with the T-83 construct as described above. Medium was changed after 12 h against medium containing dilutions of the respective inhibitor or combination thereof as indicated in Figure 4 and respective legend. Cells were cultured for additional 36 h followed by FACS and quantitative real-time-polymerase chain reaction (qRT-PCR).

In vitro methylation and reporter gene assay

In vitro methylation of *poct4*-GFP was performed using M.SssI (New England Biolabs). Forty-five units of enzyme were incubated with 45 µg of plasmid DNA in the presence of 160 µM SAM overnight. After 3 h of incubation, fresh SAM (160 µM) was added to ensure complete methylation. Methylation status of the plasmid after *in vitro* methylation was tested by digestion with MspI and HpaII (Fermentas). For the reporter gene assay HEK293T cells were plated in six-well plates and grown to 70% confluence. Subsequently, cells were co-transfected with the reporter plasmid and the respective dTALE construct. Forty-eight hours after transfection cells were lysed in PBS containing 0.5% NP40 and mammalian protease inhibitors. The lysate was cleared by centrifugation and eGFP and mCherry fluorescence was measured with a Tecan Infinite M1000 plate reader.

RNA Isolation, cDNA synthesis and qRT-PCR

Isolation of RNA and reverse transcription was carried out as described previously (19). Real-time PCR analysis was performed on the 7500 Fast Real-Time PCR System (Applied Biosystems) at standard reaction conditions using either the TaqMan Gene Expression Master Mix (Applied Biosystems) or the Power SYBR Green PCR Master Mix (Applied Biosystems). The following TaqMan Gene expression assays were used: *gapdh* (Assay ID: Mm9999915_g1), *oct4* (Assay ID: Mm00658129_gH) and *nanog* (Assay ID: Mm01617761_g1). Primer sequences for SYBR Green PCRs: *gapdh* (For 5'catgcccttcctgttctcta 3' Rev 5'cttcaccacctcttgatgcac 3'); *tet1* (for 5'ccaggaagagcgactacgtt 3' Rev 5'ttagtgtgtgtgaacctgattttattgt 3') and *hnf4a* (for 5'caagagtcctatggtgtttaagg 3', Rev 5'cgctcatctccgtagct 3'). Relative mRNA levels were normalized to *gapdh* and calculated with the comparative CT Method ($\Delta\Delta CT$ Method).

In vitro DNA-binding assays

In vitro DNA-binding assays were performed as described previously (24,25) with the following modifications. Two differentially fluorescently labeled DNA substrates corresponding to position -39 to +18 (A) and -88 to -31 (B) relative to the TSS of the *oct4* promoter (Figure 3) were used in direct competition. Substrates were prepared by annealing 5' ATTO550 or ATTO647N labeled lower strand with the respective unlabeled upper strand oligonucleotide. For competition assays, 200-nM ATTO647N-labeled substrate A and ATTO550-labeled substrate B were added and incubated at room temperature (RT) for 1 h with constant mixing. Fluorescence intensities were measured with a Tecan Infinite M1000 plate reader using the following excitation/emission wavelengths: 490 ± 10 nm / 511 ± 10 nm for eGFP, 550 ± 15 nm / 580 ± 15 nm for ATTO550 and 650 ± 10 nm / 670 ± 10 nm for ATTO647N. The measurements were normalized using standard curves from purified eGFP and ATTO-dye-labeled oligonucleotides. Moreover, a control set of each substrate with distinct fluorophores was used for normalization.

Fluorescence polarization measurements

DNA affinity was determined by fluorescence polarization measurements. eGFP-dTALE fusion proteins were purified as described above and eluted from the Sepharose beads by addition of 250 mM imidazol. Different concentrations of GFP-dTALE fusion proteins were incubated with their specific ATTO647N-labeled substrates (1 nM). After incubation for 30 min, at RT fluorescence polarization was measured with a Tecan Infinite M1000 plate reader using 635 nm for excitation and 670 ± 10 nm for emission. The data of fluorescence polarization over protein concentration were fitted with $y = \frac{V_{max} \cdot x}{K_d + x}$ using gnuplot (<http://www.gnuplot.info>).

DNA methylation analysis

For the analysis of DNA methylation levels at the *oct4* promoter genomic DNA was isolated using the NucleoSpin Triprep Kit (Macherey-Nagel). Bisulfite treatment was performed using the EZ DNA Methylation-Gold™ Kit (Zymo Research Corporation) according to the manufacturer's protocol. Subsequently, the *oct4* promoter sequence was amplified in a semi-nested PCR using the primers:

F1: 5'-ATGGGTTGAAATATTGGGTTTATTTA-3'
 F2: 5'-GTAAGAATTGAGGAGTGGTTTTAG-3'
 R1: 5'ACCCTCTAACCTTAACCTCTAAC 3'
 R2 = R1 with 5'biotin

The biotinylated PCR products of the second PCR were analyzed by pyrosequencing (Varionostic GmbH, Ulm, Germany). The pyrosequencing covered five CpG sites of which the average methylation level was calculated. DNA methylation levels of major satellite repeats and *H19* promoter was performed as previously described (25).

RESULTS

Design and construction of five dTALEs targeting the murine *oct4* promoter

We generated five dTALEs each targeting a distinct 19-bp sequence of the murine pluripotency gene *oct4* to test whether the position of the target sequence influences the efficiency of dTALE-mediated promoter activation. These five dTALEs targeted sequences upstream or downstream of the Sp1/Sp3/hormone responsive element (HRE) box (Figure 1A, Supplementary Figure S1). DNA-binding TALE repeat arrays were generated by Golden Gate cloning as described previously (6) and transferred to mammalian expression vectors by Gateway recombination (Supplementary Figure S1A). To monitor transfection efficiency and expression levels, mCherry (26) was fused to the N-terminus of the dTALEs. Furthermore, we replaced the transcriptional activation domain (AD) of the *Xanthomonas* wild-type TALE protein (wtTALEs) with the VP16 AD from the herpes simplex virus (VP16dTALEs) and compared the activity of these two distinct dTALE architectures.

dTALEs targeting distinct sites in the *oct4* promoter have similar affinities *in vitro* but differ strongly in their *in vivo* performance as transcriptional activators

The activity of the different dTALEs was first analyzed in a transient reporter gene assay. HEK293T cells were co-transfected with an *oct4* promoter-driven *eGFP* reporter (*poct4-eGFP*) and a constitutively expressed dTALE construct. Expression was analyzed by fluorescence measurement 48 h after transfection. Notably, the VP16dTALEs activated the *oct4* promoter to significantly higher levels than the corresponding wtTALEs (Figure 1C), despite the fact that the latter were expressed at slightly higher levels (Supplementary Figure S2). The most distal dTALE (T-83) yielded the strongest transcriptional activation with both, the wtTALE and VP16dTALE architecture (Figure 1B and C). To test whether the variable efficiency of the dTALEs with distinct repeat arrays is caused by the location of the target sites within the promoter, we replaced base pairs -31 to -102 of the *oct4* reporter construct which contain the target sites of four of the five dTALEs, with a shorter sequence containing one dTALE target site only. The resulting four reporter constructs have the respective dTALE target site at the same position (Figure 2A). Transcriptional activation of these four mutated reporter constructs by the corresponding VP16dTALEs was greatly reduced as compared to the activation level of the wild-type *oct4* reporter. Three of the four VP16dTALEs induced similar eGFP expression levels (Figure 2B) while T-60 exhibited a slightly stronger activation in the mutated promoter. The enhanced activity of T-60 is possibly due to the overlap of its target site with the SP1 site in the wild-type promoter, which also results in a relatively higher background in cells transfected only with the mutated reporter construct containing the T-60 binding site (Figure 2C). The dTALEs used in this study were designed to target different sequences within the

promoter region of *oct4*. The distinct RVD compositions of these dTALE repeat arrays might result in different binding affinity, causing the observed difference in transcriptional activation. We therefore determined the affinity and specificity of our dTALEs *in vitro* using fluorescently labeled DNA substrates and eGFP-dTALE fusion proteins. We found specific binding of all five dTALEs to their respective DNA substrates (Figure 3A and B). Dissociation constants were determined by fluorescence polarization and all dTALEs tested yielded affinities for their specific substrates, with K_d values in the low nanomolar to high picomolar range (Figure 3C). Notably, the dTALE T-83, which was the strongest transcriptional activator *in vivo*, had a comparably low affinity *in vitro*. Together, these data strongly suggest that the observed variations in dTALE-mediated activation of the *oct4* promoter are not due to inherent differences in their binding affinity.

dTALEs activate methylated reporter plasmids

In addition to positional effects, we tested whether the epigenetic state of the promoter might influence the efficiency of dTALE-mediated transcriptional activation. We methylated the *poct4-GFP* plasmid *in vitro* (Supplementary Figure S3A) and determined its inducibility by dTALEs. All dTALEs induced eGFP expression from the methylated *oct4* promoter, yet to lower levels as compared to the unmethylated reporter (Figure 1D). Notably, the relative differences in the activity of the highly potent T-83 and the other dTALEs were up to 25-fold and thus more pronounced with the methylated as compared to the unmethylated reporter construct (Figure 1D). These results indicate that dTALEs can activate heavily methylated promoter sequences, albeit to a reduced extent, and suggest that the lack of correlation between *in vitro* binding affinity and *in vivo* activity of dTALEs may reflect their different ability to overcome other repressive epigenetic marks at the target locus.

dTALEs hyperactivate endogenous *oct4* expression in embryonic stem cells

To test the ability of dTALEs to activate the endogenous *oct4* gene we generated mouse embryonic stem cells (ESCs) stably carrying the *poct4-GFP* reporter construct (ogESCs). ogESCs were tested with T-83 fused to the VP16 AD (VP16 T83), the most efficient dTALE, and compared with mCherry control vector. Transfected mCherry-positive cells were selected using FACS and total RNA was isolated followed by reverse transcription and qRT-PCR. FACS analysis showed that ogESCs transfected with the VP16 T-83 had a 3–4-fold higher mean eGFP fluorescence intensity compared to control transfected cells (Supplementary Figure S4A). Transcription of the endogenous *oct4* was induced about 2-fold as determined by qRT-PCR (Supplementary Figure S4B). The relatively low induction rate is likely due to the high basal level of *oct4* transcription in ESCs and the negative feedback of Oct4 on its own promoter (27).

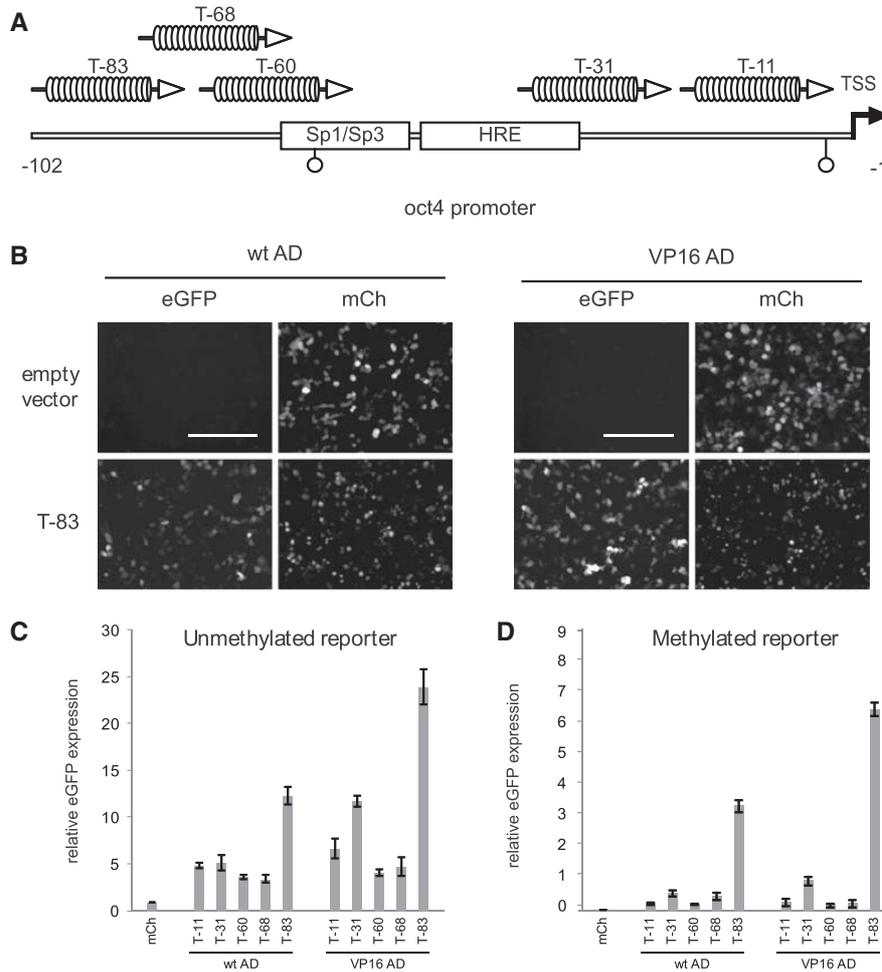


Figure 1. Activation of a transgenic *oct4* reporter construct by dTALEs in HEK293T cells. **(A)** Schematic representation of the 102-bp fragment upstream of the transcriptional start site (TSS) of the *oct4* promoter, including the binding site of the Sp1/Sp3 transcription factors, the hormone responsive element (HRE) and two CpG sites (open circles). *oct4*-specific dTALEs are depicted in correspondence of the location of their target sequence and designated according to the distance between the 5' end of their target sequence and the TSS. **(B)** Fluorescence microscopy images of HEK293T cells co-transfected with the *poct4-GFP* reporter construct and the T-83 dTALE constructs. Left panel shows cells transfected with the T-83 dTALE fused to the wild-type AD (wt AD). Right panel shows cells transfected with the T-83 dTALE fused to the VP16 AD. Scale bar = 200 μ m. **(C)** Transcriptional activation of the unmethylated *poct4-GFP* reporter construct by *oct4*-specific dTALEs. eGFP expression was normalized to cells co-transfected with a control plasmid encoding the fluorescent protein mCherry (mCh) and *poct4-GFP* reporter construct. **(D)** Transcriptional activation of the *in vitro* methylated *poct4-GFP* reporter construct by *oct4*-specific dTALEs. eGFP expression was normalized to cells co-transfected with a control plasmid (mCh) and *poct4-GFP* reporter construct. To allow for a direct comparison of expression levels in (C) and (D) the data observed on the methylated promoter were normalized to the mCherry values observed with the unmethylated promoter (C). Error bars in (C) and (D) represent standard deviation from three independent experiments.

Activation of *oct4* in neural stem cells depends on inhibition of repressive epigenetic mechanisms

To test whether dTALEs can also activate a transcriptionally silent endogenous *oct4* promoter, we differentiated ogESCs into NSCs. During this differentiation process the *oct4* locus is epigenetically silenced and NSCs no longer express *oct4* (28). Analysis by immunofluorescence showed that all cells were positive for the NSC markers (21) Pax6, Nestin and Olig2 (Supplementary Figure S5), indicating successful *in vitro* differentiation from ogESCs

to ogNSCs. The ogNSCs were transfected with the vector encoding the dTALE VP16 T-83 or a control vector encoding mCherry. Forty-eight hours after transfection cells were analyzed by flow cytometry. In contrast to the experiments with ogESCs, the dTALE VP16 T-83 activated neither the transgenic *poct4-eGFP* reporter nor the endogenous *oct4* promoter in ogNSCs (Figure 4A and B). This could be due to the different epigenetic states of the *oct4* promoter in ESCs and NSCs. Whereas the *oct4* promoter in ESCs is active and apparently accessible to dTALEs, *oct4* is not expressed in NSCs and the promoter

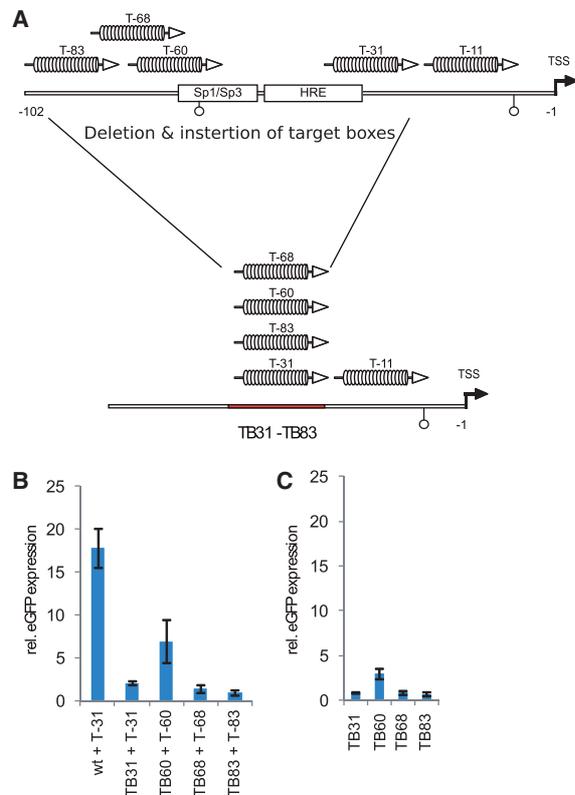


Figure 2. The location of a dTALE target sequence within the *oct4* promoters can affect its functionality. (A) Schematic representation of an *oct4* promoter deletion construct in which base pairs -31 to -102 relative to the TSS were deleted and the target sequences of the four dTALEs were inserted yielding the reporter constructs TB31, TB60, TB68 and TB83. (B) Transcriptional activation of the reporter constructs TB31, TB60, TB68 and TB83 by corresponding dTALEs. (C) Background activity of the mutated reporter constructs in cells co-transfected with respective reporter and mCherry control.

might be less prone to dTALE-mediated activation. Therefore, we envisaged that inhibiting the repressive epigenetic modifiers that prevent activation of the *oct4* promoter in NSCs could allow dTALE-mediated activation of *oct4*. To test this hypothesis, we used the HDAC inhibitors TSA (29) or VPA (30) as well as the DNA methyltransferase (Dnmt) inhibitor 5-aza-2'-deoxycytidine (5azadC) (31) to interfere with two major epigenetic mechanisms by which transcriptional silencing of genes is achieved in mammals. Twelve hours after transfection with VP16 T-83, ogNSCs were treated with the respective inhibitor for additional 36 h. Treatment with 5azadC or VPA but not TSA significantly increased relative eGFP expression in cells transfected with VP16 T-83 (Figure 4A and Supplementary Figure S6A–C). Similarly, endogenous *oct4* transcript levels were induced up to 60% as compared to the levels in ogESCs. However, a combination of 5azadC and VPA did not show additive nor synergistic effects (Figure 4A and B). Treatment with

the inhibitors alone did not result in transcriptional activation of the reporter nor the endogenous *oct4* gene (Figure 4A and B), demonstrating that the observed activation was due to the synergistic action of the dTALE and the inhibitors. Cells transfected with the dTALE VP16 T-83 and treated with VPA, 5azadC or combinations of both showed not only increased *oct4* transcript levels but also Oct4 protein (Figure 4F). Moreover, treatment of VP16 T-83 transfected cells with VPA, 5azadC or combinations of both exhibited up-regulation of the Oct4 target genes *nanog* and *tel1* (Figure 4C and D) (32–34). By contrast, genes that are not part of the Oct4 regulatory network were not influenced by treatment with inhibitors and/or expression of dTALE VP16 T-83 on transcript (Supplementary Figure S7A).

As both, 5azadC and VPA, have been reported to induce DNA demethylation (31,35) we investigated the effects of these inhibitors on the DNA methylation levels of the *oct4* promoter. Interestingly, in all samples that were treated with the inhibitors only and/or transfected with the control plasmid no change in DNA methylation levels was observed. However, expression of the dTALE VP16 T-83 together with VPA and/or 5azadC treatment caused a reduction of DNA methylation at the *oct4* promoter by $\sim 30\%$ (Figure 4E). Treatment with inhibitors alone or in combination with the dTALEs did not influence methylation levels at the *h19* locus and major satellite repeats, showing that the observed effect is specific for the *oct4* promoter (Supplementary Figure S7B and S7C). These results suggest a synergistic effect of dTALEs and epigenetic inhibitors in mammalian cells.

DISCUSSION

Variable efficiency of different dTALEs in transcriptional activation

In eukaryotic cells, transcriptional activation involves the concerted action of multiple factors recognizing target sites at different positions of gene promoters. The possibility to generate dTALEs that bind different sites within the promoter of target genes opens new possibilities to probe and optimize conditions for targeted transcriptional activation. We designed a panel of dTALEs targeting distinct sites in the murine *oct4* promoter and compared their performance *in vitro* and *in vivo*. Expression of dTALE T-83 resulted in a 2-fold increase of the *oct4* mRNA in ESCs. Previous studies reported a number of dTALEs targeting distinct promoters (5–7,36,37). However, none of these studies has systematically investigated whether the relative position of a dTALE target site within a promoter affects its functionality. We engineered five dTALEs, each targeting a distinct 19-bp sequence within the *oct4* promoter. All dTALEs yielded K_d values in the low nanomolar to high picomolar range and were expressed at similar levels but differed largely in their efficiency *in vivo*. Remarkably T-83, the dTALE with one of the lowest binding affinity, showed the highest efficiency in *oct4* promoter activation. Our data obtained with recombinant *oct4* promoter constructs showed that deletions in the native *oct4* promoter severely affected

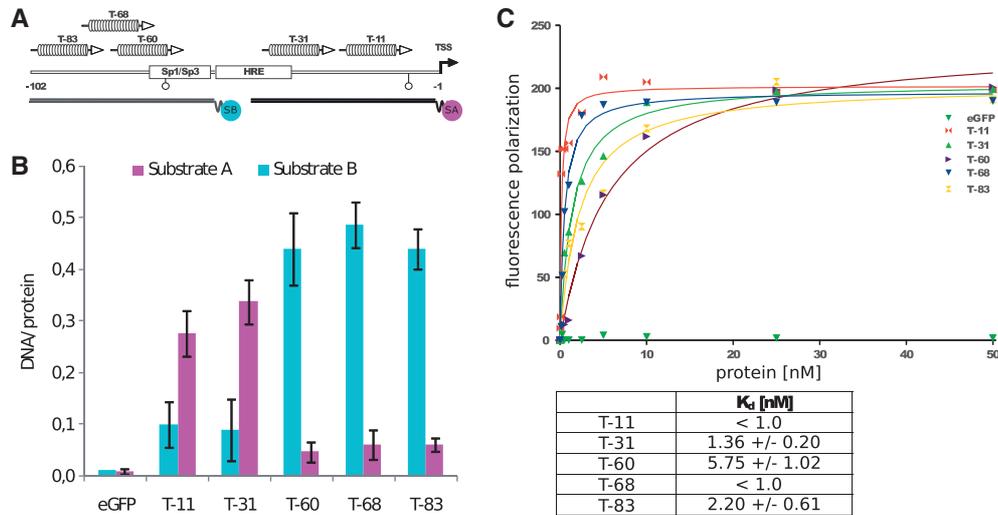


Figure 3. (A) DNA-binding properties of oct4 eGFP-dTALE fusion proteins *in vitro*. Schematic representation of the 102-bp upstream of the transcriptional start site (TSS) of the *oct4* promoter, including the binding site of the Sp1/Sp3 transcription factors, the hormone responsive element (HRE) and two CpG sites (open circles). oct4 eGFP-dTALE fusion proteins are depicted at the position of their target sequence and numbered according to the distance between the 5' end of their target sequence and the TSS of the *oct4* gene. Binding assays were performed using fluorescently labeled double-stranded DNA substrates corresponding to position -39 to +18 [substrate A (SA)] and -88 to -31 [substrate B (SB)] relative to the TSS of the *oct4* gene. Note that substrate A includes the targeting sequences of dTALEs T-11 and T-31 and substrate B includes the targeting sequences of dTALEs T-60, T-68 and T-83. (B) DNA binding of eGFP-dTALE fusions to the specific substrate in competition with the respective unspecific substrate. Shown are fluorescent intensity ratios of bound labeled DNA substrate/eGFP-dTALE fusions. eGFP was used as negative control. Values represent means and \pm SEM from three independent experiments. (C) DNA affinity measurements of the five dTALEs as measured by fluorescence polarization. Upper panel shows the data points acquired for each dTALE and the corresponding fitted curves. The table contains the K_d values for each dTALE calculated from the fittings using gnuplot and the function $f(x) = \frac{Y_{max} * x}{K_d + x}$.

dTALE performance, indicating that the presence of a specific binding site is not sufficient for efficient activation of transcription. These results suggest that the different capacity of dTALEs to activate transcription is less determined by their intrinsic DNA-binding properties but rather by their interactions at target promoters. Studies with the viral transactivator VP16 had previously indicated position-dependent interactions of the VP16 activation domain, possibly with basal transcription factors (38). Therefore, it is likely that also dTALEs are involved in complex interactions at the promoter of target genes that may either hinder or promote transcriptional activation.

As multiple *cis*- and *trans*-acting factors and epigenetic modifications are involved in the regulation of promoter activity, it will be difficult to predict the efficiency of a dTALE *in silico*. Hence, it seems important to construct and test multiple dTALEs for a given target promoter to obtain the most effective transcriptional activator. In the past, the assembly of genes that encode custom-designed repeat arrays was challenging and thus construction of multiple dTALEs targeting one promoter was not a realistic task. However, this is no longer a bottleneck since the recently established hierarchical ligation-based 'Golden Gate' cloning approaches facilitate rapid generation of genes encoding TALE repeat arrays (5,6,9,36,37,39,40).

Another potential bottleneck in the selection of efficient dTALEs is the analysis of promoter activation by

RT-PCR or comparable assays. By contrast, promoter-reporter fusions constructs facilitate rapid quantitative comparison of multiple dTALEs but may not adequately reflect the transcriptional regulation of the corresponding endogenous genes. In this context, it should be noted that the dTALE (T-83), performing best on plasmid reporters, also most efficiently activated the chromosomal *oct4* promoter (Supplementary Figure S6E). Thus, promoter-reporter fusions may greatly facilitate the screening of different dTALE repeat arrays and experimental conditions that can then be verified and optimized in a second step by monitoring transcription of the endogenous genes.

Transcriptional activation by dTALEs is facilitated by epigenetic inhibitors

In a recent study, dTALEs were shown to activate an episomal *oct4* reporter but not the endogenous *oct4* promoter (5). Similarly, we observed a lack of dTALE-mediated *oct4* activation in NSCs, where the promoter is silent. In ESCs, however, where the *oct4* promoter is active, our dTALE clearly increased *oct4* transcription, suggesting that dTALE activity depends on the epigenetic state of the promoter. These results are consistent with the reported multistep inactivation of the *oct4* promoter that occurs during cellular differentiation after implantation and involves H3K9 as well as DNA methylation. This tight epigenetic control apparently safeguards against

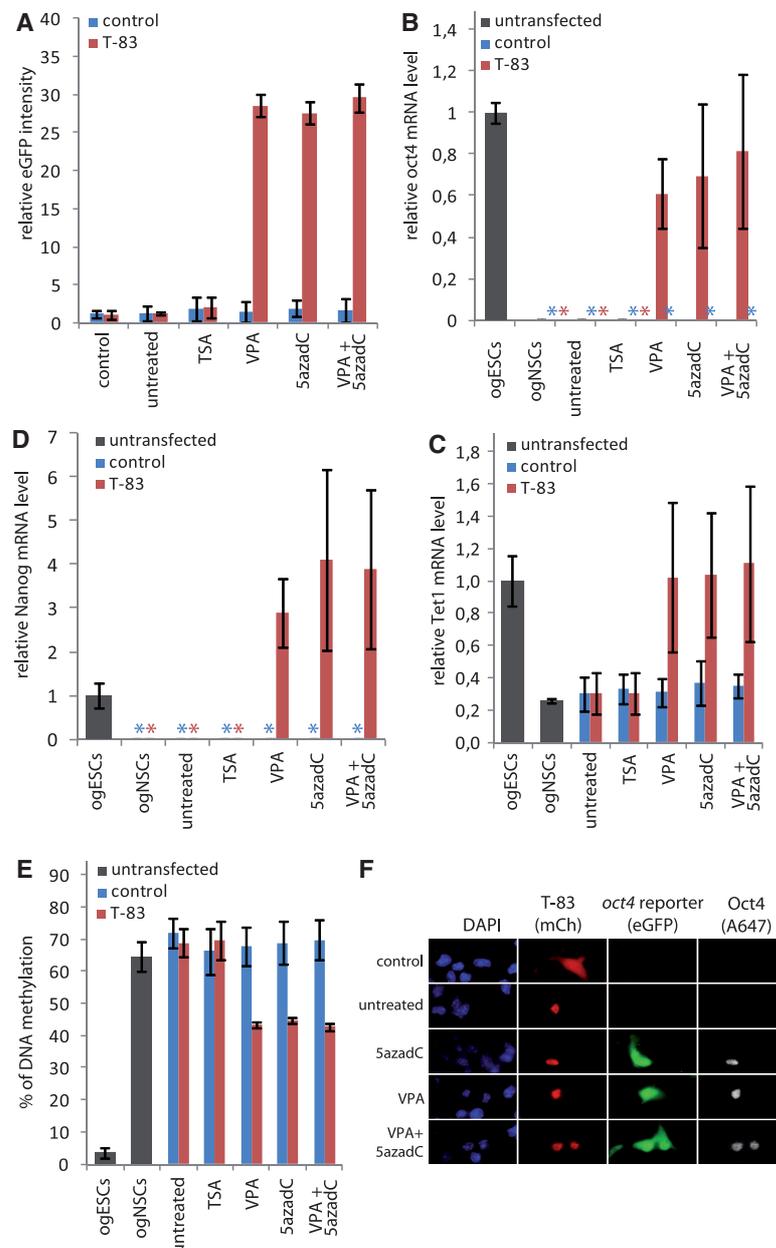


Figure 4. Activation of the endogenous *oct4* gene in NSCs requires inhibition of repressive epigenetic mechanisms. (A) Relative eGFP intensities as measured by flow cytometry of mCherry-positive ogNSCs transfected with the VP16 T-83 *dTALE* construct (T-83). Cells transfected with control plasmid (blue) or T-83 (red) were untreated or treated with TSA (30 nM), VPA (620 μ M), 5azadC (10 nM) or a combination of VPA (310 μ M) and five azadC (5 nM). (B) Relative levels of endogenous *oct4* mRNA measured by quantitative real-time PCR of transfected, mCherry-positive ogNSCs from (A) as well as untransfected ogNSCs and ogESCs as a reference. (C) DNA methylation levels of the *oct4* promoter in samples from (A) and of ogESCs as well as ogNSCs as reference. Percentage of methylation represents the average of five CpG sites in the proximal part of the *oct4* promoter. (D, E) Relative mRNA levels of *tet1* and *nanog* as determined by quantitative real-time PCR of samples from (A) and of ogESCs as well as ogNSCs as reference. (F) Fluorescence microscopy images of ogNSCs transfected with the T-83 construct in combination with 5azadC treatment (10 nM) or no drug. Samples were stained for Oct4 protein (A647) and counterstained with DAPI. mCherry channel shows cells transfected with T-83. eGFP channel shows expression of the *oct4* reporter transgene. Scale bar represents 25 μ m. For images of samples treated with the other inhibitors, see Supplementary Figure S4. Error bars represent standard deviation from two to three independent experiments. Asterisks indicate samples where no mRNA was detectable by quantitative real-time PCR.

inappropriate reactivation of the *oct4* gene and thus prevents uncontrolled proliferation and cancer (41–43).

We found that chemical inhibition of repressive epigenetic modifiers like Dnmts and HDACs enabled dTALE-mediated transcriptional activation of silent *oct4* in NSCs. Interestingly, of the two HDAC inhibitors tested only VPA but not TSA treatment allowed efficient transcriptional activation of *oct4* by dTALEs. A similar difference between the two inhibitors was previously reported for cellular reprogramming and *oct4* promoter activation (44). One possible explanation for the different efficacy of the two inhibitors could be their different target specificities (45). Interestingly, VPA was shown to specifically affect the proximal region of the *oct4* promoter (46) where also the T-83 dTALE binds. This might also explain why inhibitor treatment of cells transfected with T-31, the dTALE with the next greatest activity in the reporter assays, did not facilitate the activation of *oct4* (Supplementary Figure S6E).

Previous studies reported that high concentrations of VPA and/or 5azadC induce demethylation and reactivation of silent genes (35,47,44,46,38). Under our experimental conditions, however, these inhibitors induced DNA demethylation of the *oct4* promoter only in combination with dTALEs, indicating a synergistic effect. A possible explanation could be that binding of the dTALE interferes with maintenance of DNA methylation and, thus, in combination with the epigenetic inhibitors leads to reduction of methylation levels. Such a synergistic effect would be consistent with the recent realization that DNA methylation is rather dynamic and functionally linked to other epigenetic pathways (48). The synergy between low concentrations of epigenetic inhibitors and dTALEs suggests that silent target genes could be activated without genome-wide demethylation and thus avoid unwanted side effects.

In summary, we demonstrated that combining dTALEs with DNA methylation and/or HDAC inhibitors facilitates selective activation of the endogenous *oct4* pluripotency gene. As in turn also Oct4 target genes are reactivated, dTALEs could be used for reprogramming of somatic cells to induced pluripotent stem cells (iPSCs). It remains to be investigated whether single or combinations of several dTALEs are more efficient than present reprogramming strategies involving the Oct4 protein itself. However, in contrast to native transcription factors, dTALEs can be specifically directed against single genes or selected combinations of target genes and thereby allow dissection of complex transcription networks to identify key factors in biological processes like pluripotency and differentiation. The combination with epigenetic inhibitors may, in some cases, facilitate the activation of tightly repressed genes and further expand the utility of dTALEs in basic and applied biosciences.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–7.

ACKNOWLEDGEMENTS

We thank Luca Gentile and Hans R. Schöler (Max Planck Institute for Molecular Biomedicine, Münster) for providing the GOF-18 construct, DG Johnson (Howard Hughes Medical Institute, Durham) for providing the E2F1-VP16 fusion and Kerry Tucker (Ruprecht-Karls-University, Heidelberg) for providing wt J1 ESCs. Furthermore, we thank Fabian Köhler and Tobias Anton for help with the reporter gene assays. We thank Alex Buschle for help with the immunofluorescence staining and Carina Frauer for advice on the DNA binding assays. CSS and KT gratefully acknowledge the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). H.L., T.L., S.B. and F.S. conceived the study. H.L., T.L., F.S., S.B. and R.M. designed the experiments. S.B. performed the cell biological experiments and DNA methylation analysis. K.T. performed the DNA binding assays. R.M. and J.E. established a collection of plasmids for assembly of dTALE genes. RM designed and cloned the dTALEs used in this study. CSS isolated RNA samples, performed the expression analysis and DNA methylation analysis. S.B., F.S., T.L. and H.L. wrote the manuscript.

FUNDING

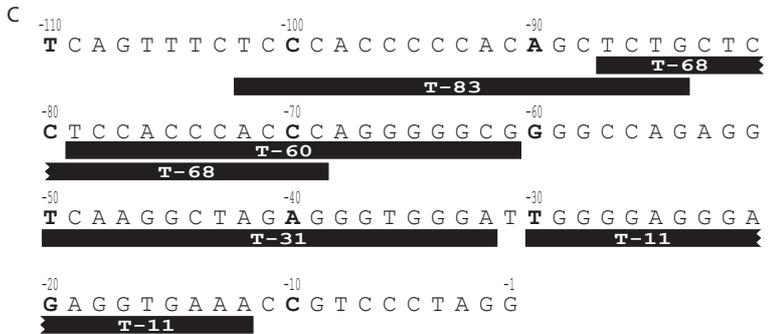
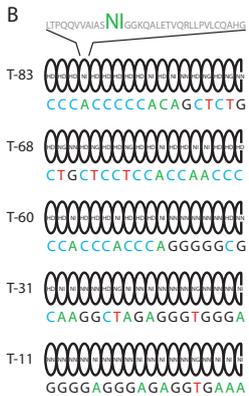
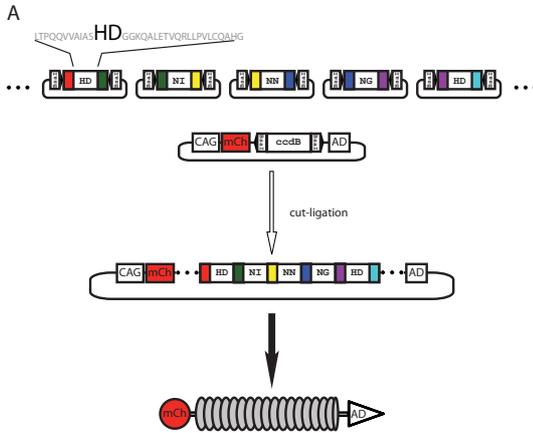
The Deutsche Forschungsgemeinschaft (SPP 1356 and SFBs 646/TR5) to H.L.; grants from the 2Blades foundation to T.L.; The Elite Network of Bavaria (International Doctorate Program NanoBioTechnology) to C.S.S.; S.B. is a fellow of the graduate school Life Science Munich (LSM). Funding for open access charge: The Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

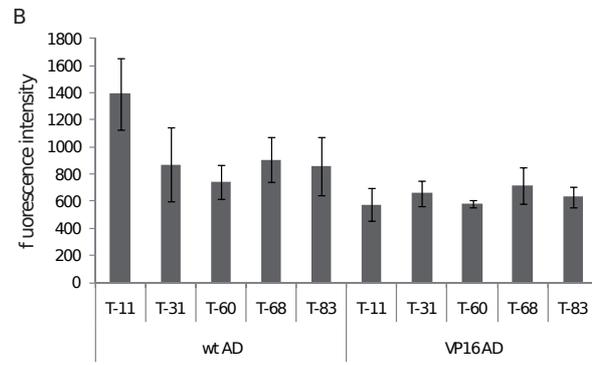
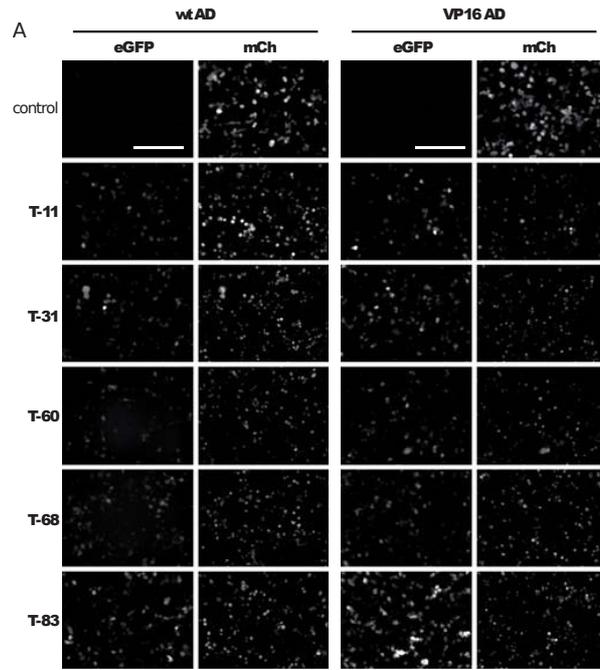
- Klug, A. (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q. Rev. Biophys.*, **43**, 1–21.
- Segal, D.J. (2011) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. USA*, **96**, 2758–2763.
- Cathomen, T. and Jung, J.K. (2008) Zinc-finger nucleases: the next generation emerges. *Mol. Ther.*, **16**, 1200–1207.
- De Francesco, L. (2011) Move over ZFNs. *Nat. Biotechnol.*, **29**, 681–684.
- Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M. and Arlotta, P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.*, **29**, 149–153.
- Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.*, **39**, 5790–5799.
- Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
- Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. and Zhu, J.-K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. USA*, **108**, 2623–2628.

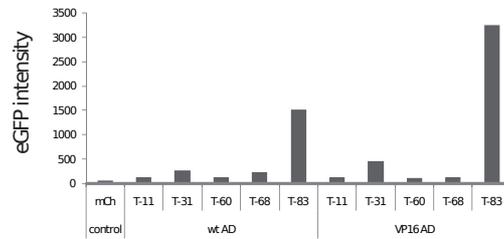
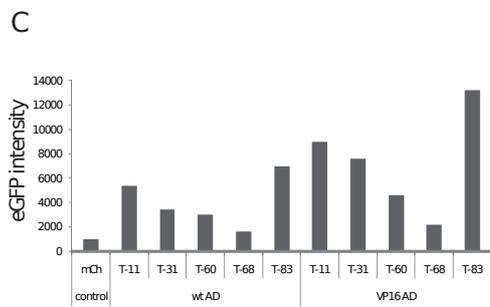
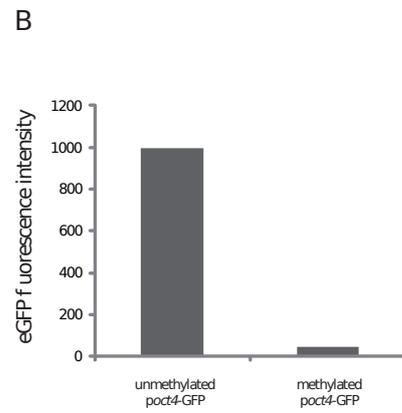
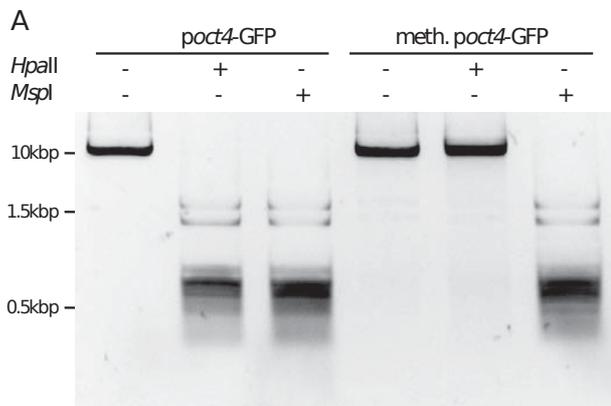
9. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
10. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
11. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
12. Bogdanove, A.J., Schornack, S. and Lahaye, T. (2010) TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.*, **13**, 394–401.
13. Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T. and Kimura, T. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.*, **104**, 34–41.
14. Niwa, H., Yamamura, K. and Miyazaki, J. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene*, **108**, 193–199.
15. Johnson, D.G. (1994) Oncogenic capacity of the E2F1 Gene. *Proc. Natl. Acad. Sci. USA*, **91**, 12823–12827.
16. Engler, C., Gruetzner, R., Kandzia, R. and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIS restriction enzymes. *PLoS One*, **4**, e5553.
17. Yeom, Y., Fuhrmann, G., Ovitt, C., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Scholer, H. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development*, **122**, 881–894.
18. DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H. and Calos, M.P. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol. Cell. Biol.*, **7**, 379–387.
19. Swagierczak, 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.
20. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
21. Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.-L., Cattaneo, E. et al. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.*, **3**, e283.
22. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. and Smith, A. (2003) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.*, **21**, 183–186.
23. Ying, Q.L. and Smith, A.G. (2003) Defined conditions for neural commitment and differentiation. *Methods Enzymol.*, **365**, 327–341.
24. Rottach, A., CFrauer, C., Pichler, G., Bonapace, I.M., Spada, F. and Leonhardt, H. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.*, **38**, 1796–1804.
25. Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., Wang, M., Qin, W., Söding, J., Spada, F. et al. (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One*, **6**, e16627.
26. Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E. and Tsien, R.Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.*, **22**, 1567–1572.
27. Pan, G., Li, J., Zhou, Y., Zheng, H. and Pei, D. (2006) A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *FASEB J.*, **20**, 1730–1732.
28. Kim, J.B., Zaehres, H., Wu, G., Gentile, L., Ko, K., Sebastiano, V., Araúzo-Bravo, M.J., Ruau, D., Han, D.W., Zenke, M. et al. (2008) Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature*, **454**, 646–650.
29. Yoshida, M., Kijima, M., Akita, M. and Beppu, T. (1990) Potent and specific inhibition of mammalian histone deacetylase both *in vivo* and *in vitro* by trichostatin A. *J. Biol. Chem.*, **265**, 17174–17179.
30. Göttlicher, M., Minucci, S., Zhu, P., Krämer, O.H., Schimpf, A., Giavara, S., Sleeman, J.P., Lo Coco, F., Nervi, C., Pelicci, P.G. et al. (2001) Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J.*, **20**, 6969–6978.
31. Santi, D.V., Garrett, C.E. and Barr, P.J. (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell*, **33**, 9–10.
32. Chen, X., Fang, F., Liou, Y.-C. and Ng, H.-H. (2008) Zfp143 regulates Nanog through modulation of Oct4 binding. *Stem Cells*, **26**, 2759–2767.
33. 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.
34. Rodda, D.J., Chew, J.-L., Lim, L.-H., Loh, Y.-H., Wang, B., Ng, H.-H. and Robson, P. (2005) Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.*, **280**, 24731–24737.
35. Dong, E., Chen, Y., Gavin, D.P., Grayson, D.R. and Guidotti, A. (2010) Valproate induces DNA demethylation in nuclear extracts from adult mouse brain. *Epigenetics*, **5**, 730–735.
36. Geißler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.-E. and Boch, J. (2011) Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One*, **6**, e19509.
37. Weber, E., Gruetzner, R., Werner, S., Engler, C. and Marillonnet, S. (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One*, **6**, e19722.
38. Hagmann, M., Georgiev, O. and Schaffner, W. (1997) The VP16 paradox: herpes simplex virus VP16 contains a long-range activation domain but within the natural multiprotein complex activates only from promoter-proximal positions. *J. Virol.*, **71**, 5952–5962.
39. Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.*, **39**, 6315–6325.
40. Scholze, H. and Boch, J. (2011) TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.*, **14**, 47–53.
41. Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H. and Bergman, Y. (2006) G9a-mediated irreversible epigenetic inactivation of *Oct-3/4* during early embryogenesis. *Nat. Cell Biol.*, **8**, 188–194.
42. Gidekel, S., Pizov, G., Bergman, Y. and Pikarsky, E. (2003) Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell*, **4**, 361–370.
43. Looijenga, L.H.J., Stoop, H., de Leeuw, H.P.J.C., de Gouveia Brazao, C.A., Gillis, A.J.M., van Roozendaal, K.E.P., van Zoelen, E.J.J., Weber, R.F.A., Wolffenbuttel, K.P., van Dekken, H. et al. (2003) POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. *Cancer Res.*, **63**, 2244–2250.
44. Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E. and Melton, D.A. (2008) Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.*, **26**, 795–797.
45. Kim, T.-Y., Bang, Y.-J. and Robertson, K. (2006) Histone deacetylase inhibitors for cancer therapy. *Epigenetics*, **1**, 14–23.
46. Teng, H.F., Kuo, Y.-L., Loo, M.R., Li, C.L., Chu, T.W., Suo, H., Liu, H.S., Lin, K.H. and Chen, S.L. (2011) Valproic acid enhances *Oct4* promoter activity in myogenic cells. *J. Cell. Biochem.*, **110**, 995–1004.
47. Al-Salihi, M., Yu, M., Burnett, D.M., Alexander, A., Samlowski, W. and Fitzpatrick, F.A. (2011) The depletion of DNA methyltransferase-1 and the epigenetic effects of 5-aza-2'-deoxycytidine (decitabine) are differentially regulated by cell cycle progression. *Epigenetics*, **6**, 1021–1028.
48. Bhutani, N., Burns, D.M. and Blau, H.M. (2011) DNA demethylation dynamics. *Cell*, **146**, 866–872.

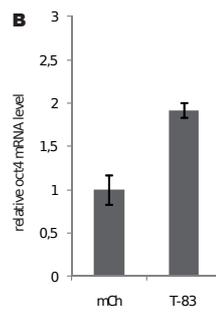
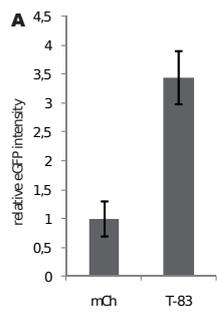


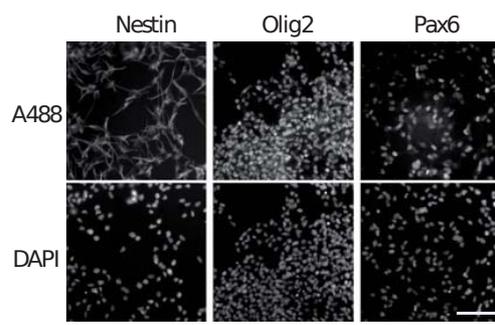
D

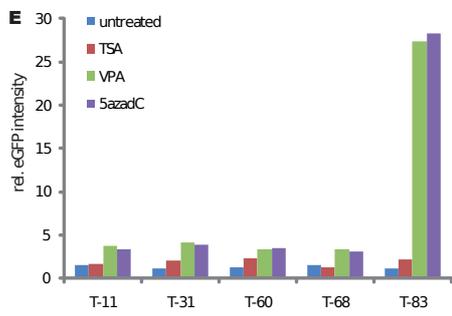
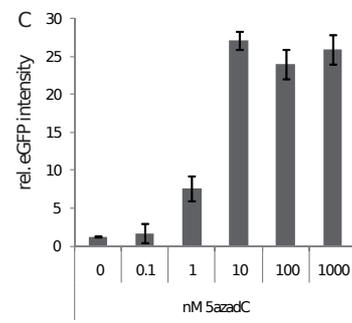
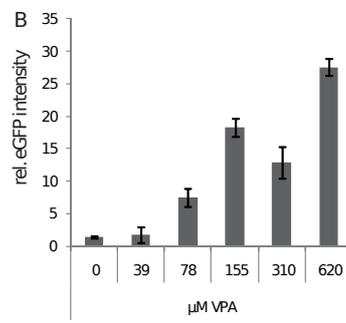
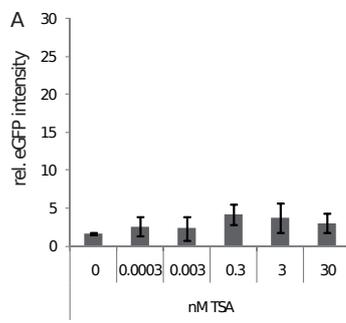
ATGTTGAGCAAGGGGAGGAGGATAAATGGCCATCATCAAGGAGTTCATCGCTTCAAGGTGCACATGGAGGGTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGG
AGGGCCGCCCTACGAGGGCACCAGCCGCAAGCTGAAGGTGACCAAGGGTGGCCCTGCCCTTCGCCTGGGACATCTGTCCCTCAGTTCATGTACGGCTCAAAGGCTAC
GTGAAGACCCCGGGACATCCCGACTCTTGAAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGGTGATGAACTTCGAGGACGGGGCGTGTGACCGTACCCAGGACTC
CTCCCTCGAGGACGGGAGTTCATCTACAAGTGAAGTGTGGCGGCACCACTTCCCTCCGACGGCCCGTAATGACAGAAAGACCATGGCTGGAGGGCTCTCCGAGCGGA
TGTACCCCGAGGACGGGGCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGCCACTACGACGCTGAGGTCAAGACCACTACAAGGCAAGAACCCGTGCA
GTGCCCGGGCTCAACGCTCAACATCAAGTGGACATCACTCCCAACAGGAGGATACACCATCGTGAACAGTACGAAAGCGGGCGGCGCCTCCACCGGGCGATGG
ACGAGCTGTACAAGAGGGCGCTGAGGAGCAGAGCCAGGCAACAGCGCGTGGAGCCACCATGGGCGATCGCAAAATTTGTACAATAAGAGAGGCTCCGGCGCGCC
CTTCAACATGGATGCCATTCGCTGGCGCACCAAGTCTGCGCGGAGCTTTCGCGCGCCCAACCGATGGGGTTCAGCGACTGACGATCGTGGGGTGTCTCGCTGCGGG
CGGGCCCTGGATGGCTTGGCGGCGGAGAGTGTGGGACCGGGTGCATCTCCCTCGCCCTCACCTGGCTTCGCGGGGCGAGCTTCAGTGAGCTTACGTCAAGT
CGATCGCTCACTTTTAAATACATCGCTTTTGTATTCATGCTCCCTTCGCGGCTCACCATACAGAGGCTGCCAAGGGGAGTGGGATGAGGTGCATCGGCTTCGCGGGCAGCG
ACGCCCCCAACCATATGCGCGTGGCTGTCACTGCGGGCGGGCGGGCGGCAAGCGGGCGGGCGACGAGCTGTGGCAACCTCCGACGCTTCGCGGGCGGGCAGGTGG
ATCTACGACGCTCGGGTACAGCCAGCAACAGGAGAAGATCAAAACGAAGTTCGTTGACAGTGGCGCAGCACACGAGGCACTGGTGGCCATGGGTTTACACAGCGCA
CATGTTGGCTCAGCCCAACCCCGCAGCGTTAGGGACCGTCTGCTCAAGTATCAGGACATGATCGCAGCGTTCGACAGGGGACACACAGAAAGCATGTTGGCGTGGGAAAC
AGTGGTCCGGCCAGCGGCTCTGGAGGCTTGTCAAGGTTGGCGGAGAGTGGAGAGTCCACCGTTACAGTTGGACACAGGCCAACTTCTCAAGATTGCAAAACGTTGGCGGCT
GACCGCAGTGGAGGAGTGCATGATGCGCAATGCACTGACGGGTGCCCGCTGAACCTTACGCGCAGCAGGTGGTGGCCATGGCCAGCCAGATGGCGCAAGCAGGGCGTGG
AGACGGTGCAGCGGCTGCTTCGGTGTGTGGCAGGCCATGGCCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGGCGTGGAGACGGTGCACGGA
TTGTTGCCGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGACGCTGGCAAGCAGCGTGGCAAGCAGCGCTGGAGACTGTCCAGCGGCTGTGGCGGCTGT
GTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGGGCTTGAAGCGTGGCAGCGGCTTGAAGCGTGGCAGCGGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAG
GCCTGACCCCGCAGCAGGTGGTGGCCATCGCCAGCCAGCAGGTGGCAAGCAGGCTGTGGAGACGGTGCAGCGGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAG
CAGGTGGTGGCCATCGCCAGCCAGCAGCGGGCAAGCAGCGCTGGAGACGGTGCAGCGGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAG
GTGGCAAGCAGCGCTGGAGACGGTGCAGCGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAGCTGGCAAGCAGCGTGGCAAGCAGCG
GTGGAGACGGTGCAGCGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAGCTGGCAAGCAGCGTGGCAAGCAGCGTGGCAAGCAGCG
AGCGGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAGCTGGCAAGCAGCGTGGCAAGCAGCGTGGCAAGCAGCGTGGCAAGCAGCG
CTGTGCTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGCGCTGGAGACGGTGGCAAGCAGCGTGGCAAGCAGCGTGGCAAGCAGCG
GAGGCTGTGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAGCTGGCAAGCAGCGTGGCAAGCAGCGTGGCAAGCAGCGTGGCAAGCAGCG
GATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCAATATTGGTGGCAAGCAGCGCTGGAGACTTCCAGCGGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAGCTGGCAAGCAGCGTGGCAAGCAGCG
CGGACAGGCTGGCCATGGCCAGCCAGCAGCGTGGCAAGCAGCGCTGGAGACGGTGCAGCGGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCGGAGCAGGTGGTGGCCATCGCCAGCCACGAGCTGGTGGCCATCGCCAGCAATGGCGGGCAGG
AGGGCTGGAGACGGTGCAGCGCTGTGGCGGCTGTGTGCCAGGCCATGGCTGACCCAGCAGCTGGTGGCCATCGCCAGCAATGGCGGGCAGCGGCTGGAGAG
ATTGTTGCCAGTTATCTGGCCATGATCGCGGTTGGCGGTTGGCAAGCAGCACTGGCTGGCTGGCGTGGCTGGCGGAGACTCTCGCTGGAGAGCAAAAGGATT
GGCCAGCGCGGCTGTGATCAAAAGAACCAATCGCGCTATTCCGAAAGCAGCATCCATCGGCTGGCGGACCGGCAAGTGGTGGCGGCTGGGTTTTTTCAGATGCCACT
CCACCCAGCCAAAGATTGTGATGAGCCATGACGCAAGTGGGATGAGCAGGACCGGTTGTACAGCTTTCGAGAGTGGGCGTCAACCACTCGAAGCCGAGTGGAGAGTGGCAAG
GCTCCCGCAGGCTCGCAGGCTGGGACGATCTCCAGGATCAGGGATGAAAAGGGCCAAACCGTCCCTACTTCAACTCAAAAGCCGGATCAGGCGCTTTCGATGCTTGG
CCGATTCGCTGGAGCTGACTTGTGCGCTAGCCCAATGACAGGGAGATCAGACCGGGCAAGCAGCGTAAACGGTCCCGATCGGATCGCTGTCACCGGCTCCCTCCGCA
CAGCAATCGTTGAGGTGGCGTTCGCCAAGCAGCGGATGGCTGCTATTTGCCCTCAGTTGAGGGTAAAAGCCCGGCTACCAAGTTCGGGGGGCGCTCCCGGATCTGCCCC
CATTAGCCAGCTCAGCGTGGAGACGAACTCCGCTGACCGGGAGGAGTGGATATGACCGCCCGGAGGCTGGAGACTTGGAGACTTGGAGACTTGGGGGAGCTGGAGTCC
CTCCCGGGAATGACCCAGCCCGCTGCTGATGGGCTTGGACGTGGACGATTTGAGTTGAACAGATGTTACCAGTGGCATGGCATTGACGACTTGGGGGGTATG

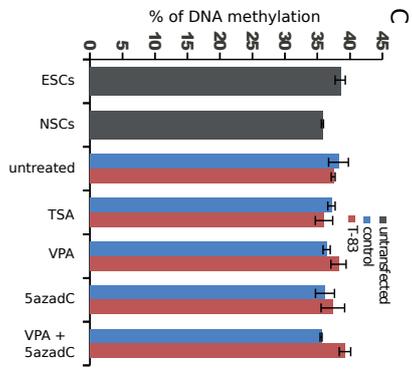
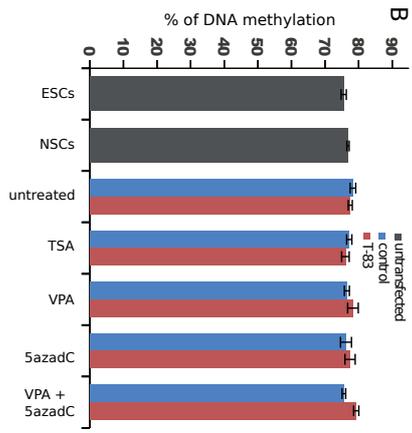
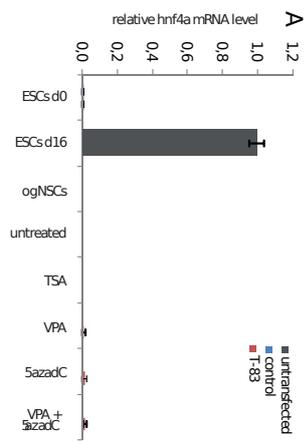












3 Discussion

3.1 The role and function of 5-hmC and Tet enzymes in development and disease

3.1.1 Methods for quantification and mapping of 5-hmC

A first step towards understanding the function of 5-hmC is to analyse in what quantities 5-hmC occurs in different cell types and developmental stages. For this purpose we developed a sensitive enzymatic method which is based on the specific transfer of a radiolabeled glucose to 5-hmC by β -glucosyltransferase (β -gt). This enzyme is used by the T4 phage to protect its genome, which contains exclusively hydroxymethylated cytosines, against bacterial restriction endonucleases. To assess whether transfer of [^3H]glucose by the β -gt to DNA is proportional to the hmC content, we prepared a series of standard DNA substrate samples by mixing corresponding proportions of two DNA fragments of same length, one having all cytosine residues replaced by 5-hmC and the other containing no 5-hmC. With this standard we could show that the assay is linear over a range of several orders of magnitude and extremely sensitive with a detection limit of 0.025% 5-hmC/C [Szwagierczak et al., 2010].

A crucial step for understanding the role of 5-hmC in the epigenetic landscape was to analyse the dynamics of this modification during early embryonic development using embryoid bodies (EBs). Differentiation into EBs is a well established *in vitro* system that recapitulates first steps of the preimplantation development [Li and Yurchenco, 2006]. Using our β -gt assay we analyzed the 5-hmC content of genomic DNA from undifferentiated wildtype ESCs, four and eight day old EBs as well as DNA from ESCs lacking all 3 major DNA methyltransferases (Dnmt1, Dnmt3a and Dnmt3b; triple knockout (TKO)). Interestingly, we could observe that 5-hmC levels are highly dynamic during the first eight days of differentiation. A sharp decrease in 5-hmC levels was evident after four days of EB culture but a substantial recovery was observed after additional 4 days of culture. Similar dynamics can be observed on the transcript levels of Tet1-3. While Tet1

is expressed predominantly in ESCs, its transcript levels decrease drastically after four days of EB culture. Tet3 is expressed at very low levels in ESCs but its mRNA levels start to increase upon differentiation with a dramatic increase between day four and day eight concomitant with the recovery of the 5-hmC level. As expected, due to the lack of DNA methylation no 5-hmC could be detected in TKO ESCs [Szwagierczak et al., 2010]. These findings point to the fact that Tet1 is the main Tet enzyme responsible for generation of 5-hmC in ESCs. In fact other studies could show that a knock-down of Tet1 in ESCs results in increased 5-mC levels at certain CpG islands and deregulation of gene expression of pluripotency-associated genes. Furthermore, Tet1-depletion in preimplantation embryos and ESCs leads to a bias towards differentiation into extra-embryonic tissues supporting a role for Tet1 in ESC maintenance and inner cell mass cell specification [Ito et al., 2010; Ficz et al., 2011; Pastor et al., 2011; Wu et al., 2011b,a; Xu et al., 2011]. Recent findings revealed that Tet1 is directly controlled by the pluripotency factor Oct4, integrating 5-hmC and Tet1 into the pluripotency network (Koh et al., 2011). Moreover, a bioinformatic study which tried to identify a minimal set of pluripotency markers found Tet1 as the best candidate using three different independent methods [Scheubert et al., 2011].

Next we analysed the genomic 5-hmC levels and the expression of Tet enzymes in adult mouse tissues. We found that adult tissues vary greatly in their 5-hmC content with highest levels in the central nervous system. In contrast to ESCs, 5-hmC levels in adult tissues correlate with high levels of Tet3 and to a lower extent Tet2, a pattern similar to day eight EBs. Thus, most differentiated tissues are characterised by very low levels of Tet1 and high levels of Tet3, while undifferentiated ESCs show the opposite pattern [Szwagierczak et al., 2010]. These findings further support the idea that Tet1 plays an important role in pluripotency. Interestingly, kidney represents an exception among the analysed tissues as it exhibits relatively high 5-hmC levels and Tet2 as the predominant transcript. This is in accordance to previous findings in Tet2 null mice where the only observable phenotype is a cellular defect in proximal convoluted tubules of the kidney [Tang et al., 2008].

Most interesting is the finding that 5-hmC levels are highest in the central nervous system (CNS). Cells of the CNS, in particular neurons, have to adapt to a very dynamic environment of inter- and intracellular contacts and signalling pathways. The flexibility to respond to these different signals is achieved by a distinctive epigenetic plasticity. In this context, DNA methylation dynamics have been shown to be involved in activity-dependent gene regulation [Martinowich et al., 2003; Chen et al., 2003b; Ma

et al., 2009], memory and learning [Miller and Sweatt, 2007; Day and Sweatt, 2010], and repeat-associated transcript expression [Muotri et al., 2010; Skene et al., 2010]. Hydroxylation of 5-mC to 5-hmC might present a mechanism by which these DNA methylation dynamics are regulated.

To further understand the function of 5-hmC in gene regulation it is important to be able to map its localisation in the genome. As 5-hmC is chemically and structurally very similar to 5-mC discrimination of these two modification presents a major challenge. The gold standard methodology for profiling of genomic 5-mC sites, bisulfite conversion, cannot discriminate 5-hmC from 5-mC and all available restriction endonucleases are either equally sensitive to mC and hmC or not sensitive to either [Huang et al., 2010; Jin et al., 2010; Nestor et al., 2010].

We found reports of an endonuclease named PvuRts1I which restriction activity *in vivo* was shown to be modulated by 5-hmC glucosylation in a complex fashion [Janosi et al., 1994]. However, as PvuRts1I was not purified, its activity has not been characterized *in vitro*. We could show that recombinant PvuRts1I selectively cleaves 5-hmC containing DNA and determined its cleavage site. Furthermore, we found that the extent of PvuRts1I digestion reflects the relative abundance of 5-hmC in genomic DNA from cerebellum and TKO ESCs [Szwagierczak et al., 2010]. Restriction of genomic DNA with PvuRts1I may be combined with PCR amplification for analysis of specific loci, with massive parallel sequencing or microarray hybridisation for genome-wide mapping. Due to its relatively complex and long recognition sequence cleavage sites occur in large distances from another, raising the argument that the extent of random breaks in genomic DNA preparations would contribute very significant noise in deep sequencing and microarray applications. This drawback can be overcome if PvuRts1I cut fragments are enriched by linkers with specific 3'-overhangs.

3.1.2 Recognition of 5-hmC by the epigenetic machinery

DNA methylation is an epigenetic modification that is involved in the control of eukaryotic gene expression. In fact, methylation of regulatory elements correlated with a transcriptional silent state. The mechanism by which the transcriptional repression is achieved appears to involve 5-mC binding proteins (MBPs). MBPs specifically recognize 5-mC and consequently recruit histone modifying enzymes and chromatin remodelling factors that establish a silent chromatin state [Sasai and Defossez, 2009].

Given the relatively well established mechanisms of how 5-mC is integrated in the epigenetic network, the discovery of 5-hmC in mammalian genomes immediately raised

the question how this new modification is recognized and interpreted by the epigenetic machinery. We therefore, characterised the 5-mC/5-hmC DNA binding properties of two representative 5mC binding proteins, the methyl-CpG binding domain (MBD) of MeCP2 and the SRA domain of Uhrf1. We found that in contrast to the MBD, the SRA domain binds 5-hmC containing substrates with a similar affinity as methylated substrates. To better understand the binding mode and thermodynamics of Uhrf1 to substrates containing 5-mC and 5-hmC we performed molecular dynamics simulations of the respective SRA:DNA complexes. Surprisingly, we found that the flipped 5-hmC base not only fits into the binding pocket of the Uhrf1 SRA domain, but is specifically stabilised by hydrogen bond formation involving the 5-hmC hydroxyl group [Frauer et al., 2011]. The specific binding of Uhrf1 to 5-hmC containing DNA was very surprising as Uhrf1 was shown to be essential for the maintenance of DNA methylation by directing Dnmt1 to hemimethylated CpG sites [Sharif et al., 2007]. The binding of Uhrf1 to hydroxymethylated DNA raises the question of how Uhrf1 contributes to the functions in DNA demethylation or transcriptional activation proposed for 5-hmC. In this context it should be noted that Uhrf1 is the only protein known that uses a base-flipping mechanism for target recognition and has no described catalytic activity on the flipped base. 5-hmC clearly interferes with the binding by the MBD of MeCP2 and might prevent subsequent establishment of repressive chromatin structures in a cellular context. Hydroxylation of 5-mC could thereby represent a mechanism of changing the cellular interpretation of an repressive epigenetic modification. Notably, MeCP2 is expressed highest in the brain where also 5-hmC levels are highest. In fact, a recent study could show that a functional correlation between 5-hmC levels and MeCP2 protein dosage exists [Szulwach et al., 2011].

A key for understanding the function of 5-hmC in epigenetic gene regulation is to uncover the factors that are involved in its recognition and how these proteins integrate 5-hmC into the epigenetic network. While the “writers” of 5-hmC are known, information about the “readers” is still largely missing. Uhrf1 might be just one of a large family of 5-hmC binding proteins to be discovered and its exact function *in vivo* remains to be elucidated. Recently, a member of the MBD protein family has been proposed to be another 5-hmC binding protein. Mbd3 is part of the Mi2/NURD histone deacetylase complex and has been shown to bind preferentially to 5-hmC via its MBD. Mbd3 recruits the Mi2/NURD complex to 5-hmC marked genes and positively regulates their expression by controlling promoter nucleosome occupancy [Yildirim et al., 2011].

3.1.3 The role of the CXXC domain for Tet1 function

While the C-terminal part of Tet1, harbouring the catalytic DSBH and the cystein-rich domain, is relatively well characterised, little is known about the N-terminal region. The only domain described so far is a CXXC zinc-finger, a domain found in several other proteins with functions related to DNA or chromatin modification. The CXXC domain of some of these proteins were shown to mediate binding to double stranded DNA containing unmethylated CpGs sites [Lee et al., 2001; Birke et al., 2002; Jørgensen et al., 2004; Thomson et al., 2010]. We performed sequence alignments and homology tree construction and identified three distinct groups of CXXC domains. The first group was characterised by a KFGG motif between the two cysteine clusters of the CXXC domain and included Dnmt1, CGP, FBx119, Mll1, Mll2 and Kdm2 proteins. The second and third group, including the CXXC domains 1 and 2 of Mbd1 on one side and those of Tet1, Cxxc4/Idax, Cxxc5/RINF and Cxxc10 on the other side, lack the KFGG motif and diverge from the first group and from each other in the sequence between the cysteine clusters.

Based on structural homology models to the crystal structure of the MLL1 CXXC domain we generated GFP fusion constructs of the Dnmt1 and Tet1 CXXC zinc finger ($\text{CXXC}^{\text{Dnmt1}}$, $\text{CXXC}^{\text{Tet1}}$) and tested their DNA binding properties using a fluorescent DNA binding assay [Frauer and Leonhardt, 2009]. $\text{CXXC}^{\text{Dnmt1}}$ exhibited a preference for unmethylated CpG containing substrate as expected for a KFGG containing MLL1/CGP type CXXC domain [Lee et al., 2001; Birke et al., 2002]. Surprisingly, $\text{CXXC}^{\text{Tet1}}$ did not show any DNA binding in these experiments [Frauer et al., 2011]. However, a recent study showed that Tet1 binds to unmodified cytosine or 5-mC or 5-hmC containing CpG-rich DNA through its CXXC domain [Xu et al., 2011]. The authors of this study used a slightly longer and untagged CXXC domain. By changing the N-terminal GFP tag to a C-terminal fusion of $\text{CXXC}^{\text{Tet1}}$ we were able to reproduce these findings (unpublished data). It is unclear what functional significance the strong binding of Tet1 to CpG containing DNA might have. However, binding of Tet1 to CpG rich sequences could provide an additional layer of protection to limit the access of Dnmts.

Next, we analysed the influence of the CXXC domain on localisation and activity of Tet1. For this purpose, we generated a GFP-Tet1 fusion lacking the CXXC domain (Tet1CXXC). Using our β -gt assay we analysed the 5-hmC content of genomic DNA in HEK293T cells overexpressing the wildtype or the deletion construct. The presence of both constructs significantly increased the 5-hmC contents to similar levels suggesting that deletion of the CXXC domain has no effect on the catalytic activity under these

conditions. In addition, wildtype Tet1 and Tet1CXXC exhibited a similar subnuclear localisation *in vivo*.

The N-terminal region of Tet1 comprised almost two thirds of the protein. Homology searches and domain prediction tools could so far not predict any similarities to other domains/proteins except the CXXC domain. It is very likely that, similar to Dnmts which also harbour a N-terminal regulatory domain, The N-terminal region of Tet1 contributes to its localisation and integration in the epigenetic network.

3.1.4 Tet2 and 5-hmC in myeloid leukemia

TET1 was originally identified as an MLL fusion partner in rare cases of acute myeloid leukemia (AML) with a t(10;11)(q22;q23) [Ono et al., 2002; Lorsch et al., 2003]. Recently, heterozygous deletions and mutations of TET2 were found in a wide range of myeloid malignancies, including myelodysplastic syndrome (MDS), myeloproliferative disorders such as chronic myelomonocytic leukemia (CMML) and in secondary AML (sAML) [Delhommeau et al., 2009; Langemeijer et al., 2009; Mohamedali et al., 2009]. To get first insights into the relationship between TET2 mutations, global gene expression profile and 5-hmC levels, we measured the 5-hmC levels in the genomic DNA of 30 sAML patients using our β -gt assay [Szwagierczak et al., 2010] which we optimised to a detection limit of 0.005%. In addition to the TET2 mutational status, we screened for IDH1/2 mutations. IDH1/2 catalyse the oxidative decarboxylation of isocitrate to α -ketoglutarate (α -KG) and have been implicated in a variety of cancer types. The most common mutations in IDH1/2 result in two enzymatic changes. A decreased wildtype IDH function and gain of a new enzymatic activity to reduce α -KG to D-2-hydroxyglutarate (D-2-HG). D-2-HG outcompetes α -KG for binding to several classes of histone demethylases, PHD2 and TET enzymes [Borodovsky et al., 2012]. Hence, IDH1/2 mutations might result in altered 5-hmC levels by inhibiting TET enzyme activity.

The analysis of the 5-hmC levels in the genomic DNA of the 30 sAML patients revealed a 5-hmC content ranging from 0.006 to 0.054% [Konstandin et al., 2011]. Compared to 5hmC levels in other human adult tissues genomic 5-hmC levels of blood samples from patients with sAML were extremely low (60 fold (human lung) to 166 fold (human brain) lower) [Kraus et al., 2012; Terragni et al., 2012]. Interestingly, low 5-hmC levels seem to be a general feature of many cancer types and in the case of glioblastomas can even be correlated with tumour malignancy [Kraus et al., 2012; Kudo et al., 2012; Yang et al., 2012]. These findings suggest that loss of 5-hmC is not just a secondary effect of

global hypomethylation often observed in tumours. Reduced levels of 5-hmC could lead to changes in gene expression patterns that directly affect the expression of oncogenes and thereby promote transformation.

In our study 7 out of 30 patients exhibited mutations in the TET2 gene and these 7 samples significantly clustered in the lower half of 5-hmC levels. Interestingly, several patients with very low 5-hmC levels did not carry a TET2 mutation. Two of these samples harboured an IDH2 mutation possibly explaining the low 5-hmC levels despite the absence of a TET2 mutation. However, 8 patients did not exhibit TET2 nor IDH1/2 mutations but clustered in the lower half of 5-hmC levels [Konstandin et al., 2011]. It is possible that these patients harbour mutations in one of the other TET genes. Another explanation could be that the low 5-hmC levels result from mutations in genes upstream or downstream of the TET enzymes. For example, Dnmts or their regulators could be affected leading to genome-wide hypomethylation and thereby to reduced 5-hmC levels. In addition, BER proteins that are implicated to be involved in the removal of 5-hmC could be deregulated in these patients. Another interesting possibility is that genome-wide loss of 5-hmC might be promoted by active/enhanced growth of cancer cells. 5-hmC might prevent the maintenance of 5-mC by DNMT1 during replication leading to a passive loss of 5-hmC. This hypothesis is supported by *in vitro* data showing a strongly reduced activity of DNMT1 on DNA substrates containing hemi-hydroxylated (5-hmCpG/CpG) CpG sites that could be produced *in vivo* by replication of 5-hmC [Valinluck and Sowers, 2007].

Next we analysed the global gene expression profiles (GEPs) from 28 patient samples. No striking difference could be seen when comparing the GEPs of patients with TET2 mutations against the GEPs of patients without TET2 mutations. However, when we compared the GEPs of the 7 patients with the lowest versus the 7 patients with the highest 5-hmC levels distinctive clustering of the two groups was observable [Konstandin et al., 2011]. Recently, similar observations were made in the context of chronic myelomonocytic leukemia (CMML) (Prez et al., 2012). Taken together these results indicate that 5-hmC levels are most likely more relevant to define biologically distinct secondary leukemia subtypes than the TET2 or IDH1/2 mutational status. Interestingly, a recent study could show that deletion of *tet2* in mice leads to a dramatic reduction of 5-hmC levels in bone marrow cells and to the development of myeloid malignancies [Li et al., 2011b]. These findings suggest that Tet2 is the predominant Tet enzyme in the hematopoietic lineage and functions as a tumour suppressor to maintain cell homeostasis.

3.1.5 5-hmC a stable epigenetic mark or a demethylation intermediate?

With the discovery of 5-hmC the question emerged whether this new covalent DNA modification represents a stable epigenetic mark by itself or an intermediate of DNA demethylation. At first, it had been debated whether active DNA demethylation does at all occur in mammals. But as more and more data accumulated that suggested the active removal of 5-mC in mammalian genomes, the attention turned towards finding the underlying mechanisms. The direct removal of the methyl group was thought to be energetically unfavourable and soon many findings suggested that DNA repair pathways are involved in the process of active DNA demethylation. The critical step in all the proposed mechanisms was the derivatisation of 5-mC by the cellular machinery. Deamination was thought to be the likeliest mechanism by which the 5-mC could be marked for the removal by the DNA repair machinery. However, the data that showed a deamination of 5-mC was mainly indirect by showing the involvement of AID/APOBEC enzymes or came from experiments performed *in vitro*. Therefore, it remained highly debated whether DNA repair is involved in active DNA demethylation despite the accumulating amount of data supporting this hypothesis. The discovery of the hydroxylation of 5-mC by Tet enzymes offered a simple explanation to the question of how a methylated cytosine can be specifically marked for demethylation without contradicting the existing data on deamination coupled DNA repair. In fact, shortly after the discovery of 5-hmC it was shown that AID/APOBEC enzymes, a protein family already implicated in deamination coupled repair of 5-mC, mediate active DNA demethylation by deaminating 5-hmC to 5-hmU [Cortellino et al., 2011; Guo et al., 2011]. In addition to linking active DNA demethylation to deamination coupled DNA repair, the presence of 5-hmC in mammalian genomes also reopens the possibility for a direct removal of the methyl-group from 5-mC. Recent studies have shown that Tet enzymes further oxidise 5-hmC to 5-caC [He et al., 2011; Ito et al., 2011]. Although it was suggested that 5-caC is recognised by the DNA repair machinery [He et al., 2011; Ito et al., 2011], the possibility remains that 5-caC is decarboxylated to cytosine by a so far unknown decarboxylase presenting a pathway for DNA demethylation independent of DNA repair mechanisms. Besides its relatively well established role in active DNA demethylation, 5-hmC might also represent a stable epigenetic mark by itself. Especially in the nervous system 5-hmC levels can account for up to 40% of all modified cytosines suggesting an additional function as an epigenetic regulator [Szwagierczak et al., 2011; Szulwach et al., 2011]. In

this context it is important to note that while Tet1 seems to have a dual function, the presence of 5-hmC is generally correlated with transcriptional activity [Williams et al., 2011; Wu et al., 2011b,a]. One explanation could be that MeCP2, a methylcytosine binding protein especially abundant in the brain, does not bind to 5-hmC and thereby preventing the establishment of a silent chromatin state [Frauer et al., 2011]. Another indication that 5-hmC might be more than a DNA demethylation intermediate comes from the recent finding that 5-hmC is bound by Mbd3. The recruitment of Mbd3 to 5-hmC marked genes positively regulates their expression directly linking 5-hmC to active gene expression [Yildirim et al., 2011]. Although there is first evidence that 5-hmC may have a DNA demethylation independent function in gene regulation, further studies are needed to identify the exact mechanisms.

3.2 The role of DNA methylation in transgene silencing and its mechanism

Transgene expression is a widely used method in cell biology and several vector systems have been developed over the past decades for this purpose. The main differences among the vectors is the promoter which is used for transgene expression. While for some applications transient expression for 1-2 days is sufficient, some experiments require stable long-term expression. We found that upon transient transfection of wt ESCs constructs driven by the cytomegalovirus (CMV) promoter were rapidly silenced, while constructs driven by the chimeric CMV early enhancer/chicken β actin (CAG) promoter yielded long and stable expression. To understand the underlying mechanisms of this observation we developed an epigenetic silencing assay. ESCs were cotransfected with two distinct plasmids, one expressing monomeric red fluorescent protein (mRFP) under the CMV promoter, the other expressing green fluorescent protein (GFP) driven by the CAG promoter. We monitored RFP and GFP expression for up to ten days after transfection by using automated image acquisition and quantification of fluorescent signals. While in wt ESCs CMV driven mRFP expression was rapidly silenced, *dnmt3a/3b* knockout cells exhibited no silencing of the CMV promoter. Surprisingly, *uhrf1*^{-/-} ESCs were also unable to silence the CMV promoter whereas *dnmt1*^{-/-} ESCs showed only partially reduced silencing [Meilinger et al., 2009]. These findings suggest that Dnmt1 might play a minor role in this silencing process. Moreover, it implies a novel function for Uhrf1 that is largely independent of Dnmt1. This was surprising, as Uhrf1 has been reported to be functionally linked to Dnmt1 in maintaining DNA methylation [Sharif

et al., 2007]. Interestingly, the silencing of the CMV promoter showed a similar dependence on the presence of Uhrf1 as for the presence of Dnmt3a/Dnmt3b. In addition, co-immunoprecipitation experiments revealed an interaction between Uhrf1 and both *de novo* Dnmts. This suggests that Np95 might play a role in the *de novo* methylation of the CMV promoter and that the silencing process is DNA methylation dependent. Surprisingly, the CMV promoter exhibited nearly no DNA methylation in the first days after transfection and reached only a very low level on the last day examined in wild type ESCs. However, promoter methylation was less in *uhrf1*^{-/-} cells than in wild type ES and none of the *dnmt* deficient cell lines showed a significant amount of DNA methylation at this time point [Meilinger et al., 2009]. Taken together, all these findings imply that Uhrf1 is cooperating with Dnmt3a and Dnmt3b in mediating the silencing of the CMV promoter and that this process, at least at its onset, is independent of DNA methylation. This is in agreement with the finding that silencing precedes DNA methylation at the promoters of certain genes [Feldman et al., 2006; Epsztejn-Litman et al., 2008].

The dependency of CMV promoter silencing on the presence of Dnmt3a/3b or Uhrf1 well before *de novo* methylation can be detected could be due to the involvement of repressive histone methylation. We therefore tested the involvement of the histone methyltransferases (HMTs) G9a and Suv39h1/2 using the epigenetic silencing assay. Interestingly, both HMTs seem to be involved in this process although to a different extent. While in the absence of G9a no preferential silencing of the CMV promoter was observable, cells lacking Suv39h1/2 did only exhibit a partial silencing defect. This is likely due to the functional differences of the two HMTs. Suv39h1/2 tri-methylate lysine 9 of histone 3 (H3K9) and were shown to be involved in organization of silent chromatin domains such as the pericentric heterochromatin [Jenuwein et al., 1998]. In contrast, G9a is crucial for H3K9 mono- and di-methylation of euchromatin and it is essential for the transcriptional silencing of several genes [Tachibana et al., 2002; Feldman et al., 2006]. This could explain why G9a and not Suv39h1/2 is crucial for the silencing of the CMV promoter. Taken together, the finding that DNA methylation seems to be set at a later stage in the silencing process and the fact that HMTs are involved in the CMV promoter silencing suggests that histone modifications might be the primary mechanism by which the silencing is initiated. This idea is supported by the finding that upon transfection into mammalian cells, plasmids are chromatinized and both their expression and silencing is associated with specific histone modifications [Riu et al., 2007].

3.3 Designer TALEs as a tool for targeted transcriptional activation

In eukaryotic cells, transcriptional activation involves the synergistic action of multiple factors recognizing target sites at different positions of gene promoters. With the discovery of the dTALE DNA binding specificity and its underlying code it became possible to specifically design TALEs that bind to user-defined target sequences. Previous studies reported a number of dTALEs targeting distinct promoters [Geissler et al., 2011; Miller et al., 2011; Morbitzer et al., 2011; Weber et al., 2011; Zhang et al., 2011]. However, none of these studies has systematically investigated whether the relative position of a dTALE target site within a promoter affects its functionality. We designed five dTALEs, each targeting a distinct 19-bp sequence within the *oct4* promoter. All dTALEs yielded Kd values in the low nanomolar to high picomolar range and were expressed at similar levels but differed largely in their efficiency *in vivo*. Intriguingly, T-83, the dTALE with one of the lowest binding affinities, showed the highest efficiency in *oct4* promoter activation. Our data obtained with recombinant *oct4* promoter constructs showed that deletions in the native *oct4* promoter severely affected dTALE performance, indicating that the presence of a specific binding site is not sufficient for efficient activation of transcription [Bultmann et al., 2012]. These results suggest that the different capacity of dTALEs to activate transcription is less determined by their intrinsic DNA-binding properties but rather by their interactions at target promoters. Studies with the viral transactivator VP16 had previously indicated position-dependent interactions of the VP16 activation domain, possibly with basal transcription factors [Hagmann et al., 1997]. Therefore, it is likely that also dTALEs are involved in complex interactions at the promoter of target genes that may either hinder or promote transcriptional activation. As multiple cis- and trans-acting factors and epigenetic modifications are involved in the regulation of promoter activity, it will be difficult to predict the efficiency of a dTALE *in silico*. Hence, it seems important to construct and test multiple dTALEs for a given target promoter to obtain the most effective transcriptional activator. Due to the modular structure of the dTALE DNA binding domain construction of multiple TALEs in a time and cost-efficient manner using recently established hierarchical ligation-based Golden Gate cloning approaches has become a feasible task Cermak et al. [2011]; Geissler et al. [2011]; Li et al. [2011a]; Morbitzer et al. [2011]; Weber et al. [2011]; Zhang et al. [2011]. In a recent study, dTALEs were shown to activate an episomal *oct4* reporter but not the endogenous *oct4* promoter [Zhang et al., 2011]. Similarly, we observed a lack of

dTALE-mediated *oct4* activation in neural stem cells (NSCs), where the promoter is silent [Bultmann et al., 2012]. In ESCs, however, where the *oct4* promoter is active, our dTALE clearly increased *oct4* transcription, suggesting that dTALE activity depends on the epigenetic state of the promoter. These results are consistent with inactivation of the *oct4* promoter that occurs during cellular differentiation after implantation and involves H3K9 as well as DNA methylation. This tight epigenetic control apparently safeguards against inappropriate reactivation of the *oct4* gene and thus prevents uncontrolled proliferation and cancer [Gidekel et al., 2003; Looijenga et al., 2003; Feldman et al., 2006]. Furthermore, the recently solved crystal structures of two TALEs showed that the protein winds around the DNA along the major groove [Deng et al., 2012; Mak et al., 2012]. Due to this conformation dTALE binding is very likely affected by repressive chromatin states as it has to compete with tightly packed nucleosome arrays for DNA binding (Figure 17).

In line with this hypothesis, we found that chemical inhibition of repressive epigenetic modifiers like Dnmts and histone deacetylases (HDACs) enabled dTALE-mediated transcriptional activation of silent *oct4* in NSCs. Interestingly, of the two HDAC inhibitors tested only valproic acid (VPA) but not trichostatin A (TSA) treatment allowed efficient transcriptional activation of *oct4* by dTALEs. A similar difference between the two inhibitors was previously reported for cellular reprogramming and *oct4* promoter activation [Huangfu et al., 2008]. One possible explanation for the different efficacy of the two inhibitors could be their different target specificities [Kim et al., 2006]. Interestingly, VPA was shown to specifically affect the proximal region of the *oct4* promoter [Teng et al., 2010] where also the T-83 dTALE binds. This might also explain why inhibitor treatment of cells transfected with T-31, the dTALE with the next greatest activity in the reporter assays, did not facilitate the activation of *oct4*. Previous studies reported that high concentrations of VPA and/or 5azadC induce demethylation and reactivation of silent genes [Huangfu et al., 2008; Dong et al., 2010; Teng et al., 2010; Al-Salihi et al., 2011]. Under our experimental conditions, however, these inhibitors induced DNA demethylation of the *oct4* promoter only in combination with dTALEs, indicating a synergistic effect. A possible explanation could be that binding of the dTALE interferes with maintenance of DNA methylation and, thus, in combination with the epigenetic inhibitors leads to reduction of methylation levels. The synergy between low concentrations of epigenetic inhibitors and dTALEs suggests that silent target genes could be activated without genome-wide demethylation and thus avoid unwanted side effects. In summary, we demonstrated that combining dTALEs with DNA methylation and/or HDAC in-

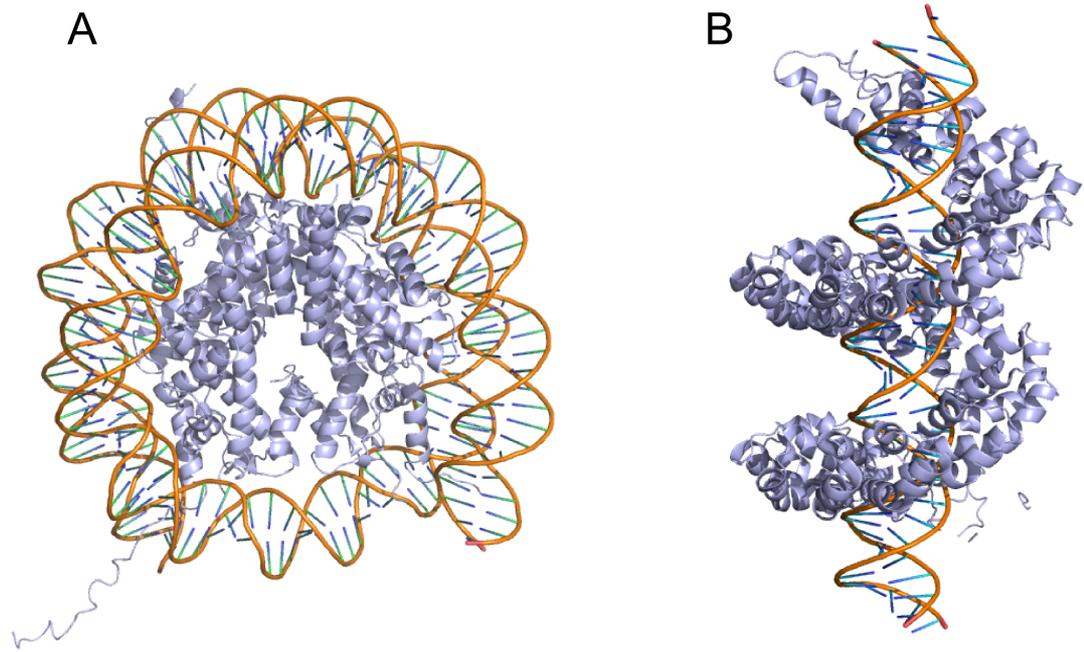


Figure 17: Crystal structures of (A) a nucleosome and (B) the PthXo1 TALE bound to DNA. Binding of the TALE to nucleosomal DNA would probably lead to a sterical clash between the two proteins.

hibitors facilitates selective activation of the endogenous *oct4* pluripotency gene. As in turn also Oct4 target genes are reactivated, dTALEs could be used for reprogramming of somatic cells to induced pluripotent stem cells (iPSCs). It remains to be investigated whether single or combinations of several dTALEs are more efficient than present reprogramming strategies involving the Oct4 protein itself. However, in contrast to native transcription factors, dTALEs can be specifically directed against single genes or selected combinations of target genes and thereby allow dissection of complex transcription networks to identify key factors in biological processes like pluripotency and differentiation. It will be interesting to investigate whether TALEs have to compete with nucleosomes for DNA binding as suggested by the crystal structure. Intriguingly, TALEs evolved to activate transcriptionally silent target genes in plant cells and can do this in a very efficient manner. However, although plants and mammals rely on the same basic epigenetic machinery, transcriptional activation on silent loci in mammals is much lower compared to plants. This leads to the consideration that plant cells might express certain factors that are missing in mammals which are recruited by the TALE proteins to the target site and help to activate the target genes. Identifying these missing factors by comparing the

interactions of dTALEs in mammals versus those in plants, could allow the activation of silenced genes without the need for treatment with epigenetic inhibitors.

3.3.1 FairyTALE a simple web tool for dTALE target optimisation

Since the discovery of the code underlying the DNA binding specificity of the *Xanthomonas* Transcription-activator-like effectors, TALEs have been successfully used in a variety of application and have become a popular and powerful tool in genome editing and engineering. The modular structure of the TALE DNA binding domain allows the construction of dTALEs that bind to any user-defined DNA target sequence. Due to the high degree of flexibility in target choice, optimization of target selection is challenging.

Given that the number of possible target sites in a 1000 base pair long double-stranded DNA fragment is statistically 500, manual search for potential target sites is tedious and time-consuming. The target locus needs to be screened for sequences beginning with a thymidine and to guarantee uniqueness of the found sequences, a BLAST search for every potential target sequence needs to be performed. Due the low specificity of guanine recognition by the NN diresidue the number of guanines in a sequence is preferably to be kept at a minimum [Bogdanove et al., 2010]. In addition, low complexity sequences and repeats should be avoided. To optimize and streamline the process TALE target site definition, we developed a web tool called fairyTALE that allows efficient and automated selection of the best TALE target sites in a user-defined sequence for a variety of applications.

Our web tool allows the definition of optimal target sites for single as well as dTALE pairs. The user chooses the length of the dTALEs along with the targeted species and

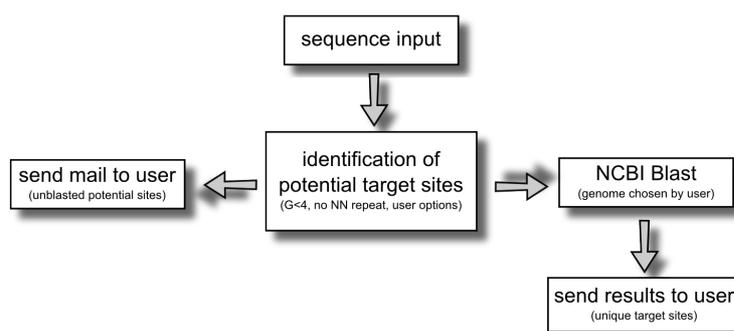


Figure 18: Work-flow of the fairyTALE target finder. Sequence input is screened for potential TALE target sites using the given parameters. Subsequently, potential sites are checked for possible off-targets using NCBI BLAST. Intermediate and final results are send to the user via e-mail.

several other parameters concerning dTALE target site composition (e.g. low complexity filter). The web tool will find TALE target sequences with the given parameters and check the number of possible off-targets via NCBI BLAST. Subsequently, unique TALE target sites in the input sequence are returned to the user (Figure 18). The beta version of the fairyTALE web tool was written exclusively in pythonTM using the Biopython module [Rossum, 2003; Cock et al., 2009] and is freely available under:

`http://bultmann.alwaysdata.net/FT_web.htm`

We believe that the fairyTALE web tool optimizes the target site selection process and increases the likelihood of successful dTALE application by reducing the amount of possible off-targets.

4 Annex

4.1 References

References

Mazin Al-Salihi, Margaret Yu, David M Burnett, Amanda Alexander, Wolfram E Samolowski, and Frank A Fitzpatrick. The depletion of DNA methyltransferase-1 and the epigenetic effects of 5-aza-2'deoxyctidine (decitabine) are differentially regulated by cell cycle progression. *Epigenetics: Official Journal of the DNA Methylation Society*, 6(8):1021–1028, August 2011. PMID: 21725200.

Madeleine P Ball, Jin Billy Li, Yuan Gao, Je-Hyuk Lee, Emily M LeProust, In-Hyun Park, Bin Xie, George Q Daley, and George M Church. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnology*, 27(4):361–368, April 2009. PMID: 19329998.

Bradley E Bernstein, Tarjei S Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J Huebert, James Cuff, Ben Fry, Alex Meissner, Marius Wernig, Kathrin Plath, Rudolf Jaenisch, Alexandre Wagschal, Robert Feil, Stuart L Schreiber, and Eric S Lander. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–326, April 2006. PMID: 16630819.

T H Bestor. Cloning of a mammalian DNA methyltransferase. *Gene*, 74(1):9–12, December 1988. PMID: 3248734.

T H Bestor and V M Ingram. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 80(18):5559–5563, September 1983. PMID: 6577443.

Nidhi Bhutani, Jennifer J Brady, Mara Damian, Alessandra Sacco, Stphane Y Corbel,

- and Helen M Blau. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, 463(7284):1042–1047, February 2010. PMID: 20027182.
- Marco Birke, Silke Schreiner, Mara-Paz Garca-Cullar, Kerstin Mahr, Fritz Titgemeyer, and Robert K Slany. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Research*, 30(4):958–965, February 2002. PMID: 11842107.
- Jens Boch, Heidi Scholze, Sebastian Schornack, Angelika Landgraf, Simone Hahn, Sabine Kay, Thomas Lahaye, Anja Nickstadt, and Ulla Bonas. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(5959):1509–1512, December 2009. PMID: 19933107.
- Adam J Bogdanove, Sebastian Schornack, and Thomas Lahaye. TAL effectors: finding plant genes for disease and defense. *Current Opinion in Plant Biology*, 13(4):394–401, August 2010.
- Alexandra Borodovsky, Meghan J Seltzer, and Gregory J Riggins. Altered cancer cell metabolism in gliomas with mutant IDH1 or IDH2. *Current Opinion in Oncology*, 24(1):83–89, January 2012. PMID: 22080945.
- D Bourc’his, G L Xu, C S Lin, B Bollman, and T H Bestor. Dnmt3L and the establishment of maternal genomic imprints. *Science (New York, N.Y.)*, 294(5551):2536–2539, December 2001. PMID: 11719692.
- Dborah Bourc’his and Timothy H Bestor. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 431(7004):96–99, September 2004. PMID: 15318244.
- A Bradley, M Evans, M H Kaufman, and E Robertson. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309(5965):255–256, May 1984. PMID: 6717601.
- M Brandeis, D Frank, I Keshet, Z Siegfried, M Mendelsohn, A Nemes, V Temper, A Razin, and H Cedar. Sp1 elements protect a CpG island from de novo methylation. *Nature*, 371(6496):435–438, September 1994. PMID: 8090226.
- Sebastian Bultmann, Robert Morbitzer, Christine S Schmidt, Katharina Thanisch, Fabio Spada, Janett Elsaesser, Thomas Lahaye, and Heinrich Leonhardt. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs

- and inhibition of epigenetic modifiers. *Nucleic Acids Research*, March 2012. PMID: 22387464.
- M C Cardoso and H Leonhardt. DNA methyltransferase is actively retained in the cytoplasm during early development. *The Journal of Cell Biology*, 147(1):25–32, October 1999. PMID: 10508852.
- Dana Carroll. Genome engineering with Zinc-Finger nucleases. *Genetics*, 188(4):773–782, 2011.
- Howard Cedar and Yehudit Bergman. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews. Genetics*, 10(5):295–304, May 2009. PMID: 19308066.
- Tomas Cermak, Erin L. Doyle, Michelle Christian, Li Wang, Yong Zhang, Clarice Schmidt, Joshua A. Baller, Nikunj V. Somia, Adam J. Bogdanove, and Daniel F. Voytas. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, 39(12):e82, July 2011.
- Ian Chambers, Jose Silva, Douglas Colby, Jennifer Nichols, Bianca Nijmeijer, Morag Robertson, Jan Vrana, Ken Jones, Lars Grotewold, and Austin Smith. Nanog safeguards pluripotency and mediates germline development. *Nature*, 450(7173):1230–1234, December 2007. PMID: 18097409.
- Jayanta Chaudhuri, Uttiya Basu, Ali Zarrin, Catherine Yan, Sonia Franco, Thomas Perlot, Bao Vuong, Jing Wang, Ryan T Phan, Abhishek Datta, John Manis, and Frederick W Alt. Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Advances in Immunology*, 94:157–214, 2007. PMID: 17560275.
- Frederic Chedin, Michael R Lieber, and Chih-Lin Hsieh. 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(26):16916–16921, December 2002. PMID: 12481029.
- Taiping Chen, Yoshihide Ueda, Jonathan E Dodge, Zhenjuan Wang, and En Li. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by dnmt3a and dnmt3b. *Molecular and Cellular Biology*, 23(16):5594–5605, August 2003a. PMID: 12897133.

- Taiping Chen, Naomi Tsujimoto, and En Li. The PWWP domain of dnmt3a and dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Molecular and Cellular Biology*, 24(20):9048–9058, October 2004. PMID: 15456878.
- Wen G Chen, Qiang Chang, Yingxi Lin, Alexander Meissner, Anne E West, Eric C Griffith, Rudolf Jaenisch, and Michael E Greenberg. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science (New York, N.Y.)*, 302(5646):885–889, October 2003b. PMID: 14593183.
- S R Cherry, D Biniszkiwicz, L van Parijs, D Baltimore, and R Jaenisch. Retroviral expression in embryonic stem cells and hematopoietic stem cells. *Molecular and Cellular Biology*, 20(20):7419–7426, October 2000. PMID: 11003639.
- L S Chuang, H I Ian, T W Koh, H H Ng, G Xu, and B F Li. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science (New York, N.Y.)*, 277(5334):1996–2000, September 1997. PMID: 9302295.
- Elisabetta Citterio, Roberto Papait, Francesco Nicassio, Manuela Vecchi, Paola Gomiero, Roberto Mantovani, Pier Paolo Di Fiore, and Ian Marc Bonapace. Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Molecular and Cellular Biology*, 24(6):2526–2535, March 2004. PMID: 14993289.
- T Clouaire and I Stancheva. Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cellular and Molecular Life Sciences: CMLS*, 65(10):1509–1522, May 2008. PMID: 18322651.
- Peter J A Cock, Tiago Antao, Jeffrey T Chang, Brad A Chapman, Cymon J Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J L de Hoon. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)*, 25(11):1422–1423, June 2009. PMID: 19304878.
- V Colot and J L Rossignol. Eukaryotic DNA methylation as an evolutionary device. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 21(5):402–411, May 1999. PMID: 10376011.
- Salvatore Cortellino, Jinfei Xu, Mara Sannai, Robert Moore, Elena Caretti, Antonio Cigliano, Madeleine Le Coz, Karthik Devarajan, Andy Wessels, Dianne Soprano,

- Lara K Abramowitz, Marisa S Bartolomei, Florian Rambow, Maria Rosaria Bassi, Tiziana Bruno, Maurizio Fanciulli, Catherine Renner, Andres J Klein-Szanto, Yoshihiro Matsumoto, Dominique Kobi, Irwin Davidson, Christophe Alberti, Lionel Larue, and Alfonso Bellacosa. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, 146(1):67–79, July 2011. PMID: 21722948.
- F H Crick. On protein synthesis. *Symposia of the Society for Experimental Biology*, 12: 138–163, 1958. PMID: 13580867.
- John R Danzer and Lori L Wallrath. Mechanisms of HP1-mediated gene silencing in drosophila. *Development (Cambridge, England)*, 131(15):3571–3580, August 2004. PMID: 15215206.
- Meelad M Dawlaty, Kibibi Ganz, Benjamin E Powell, Yueh-Chiang Hu, Styliani Markoulaki, Albert W Cheng, Qing Gao, Jongpil Kim, Sang-Woon Choi, David C Page, and Rudolf Jaenisch. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, 9(2):166–175, August 2011. PMID: 21816367.
- Jeremy J Day and J David Sweatt. DNA methylation and memory formation. *Nature Neuroscience*, 13(11):1319–1323, November 2010. PMID: 20975755.
- Franois Delhommeau, Sabrina Dupont, Vronique Della Valle, Chlo James, Severine Trannoy, Aline Mass, Olivier Kosmider, Jean-Pierre Le Couedic, Fabienne Robert, Antonio Alberdi, Yann Lcluse, Isabelle Plo, Franois J Dreyfus, Christophe Marzac, Nicole Casadevall, Catherine Lacombe, Serge P Romana, Philippe Dessen, Jean Soulier, Franck Vigui, Michaela Fontenay, William Vainchenker, and Olivier A Bernard. Mutation in TET2 in myeloid cancers. *The New England Journal of Medicine*, 360(22): 2289–2301, May 2009. PMID: 19474426.
- Rebecca K Delker, Sebastian D Fugmann, and F Nina Papavasiliou. A coming-of-age story: activation-induced cytidine deaminase turns 10. *Nature Immunology*, 10(11): 1147–1153, November 2009. PMID: 19841648.
- Dong Deng, Chuangye Yan, Xiaojing Pan, Magdy Mahfouz, Jiawei Wang, Jian-Kang Zhu, Yigong Shi, and Nieng Yan. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science (New York, N.Y.)*, 335(6069):720–723, February 2012. PMID: 22223738.

- Jacqueline Dickson, Humaira Gowher, Ruslan Strogantsev, Miklos Gaszner, Alan Hair, Gary Felsenfeld, and Adam G West. VEZF1 elements mediate protection from DNA methylation. *PLoS Genetics*, 6(1):e1000804, January 2010. PMID: 20062523.
- Jonathan E Dodge, Masaki Okano, Fred Dick, Naomi Tsujimoto, Taiping Chen, Shumei Wang, Yoshihide Ueda, Nick Dyson, and En Li. Inactivation of dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *The Journal of Biological Chemistry*, 280(18):17986–17991, May 2005. PMID: 15757890.
- A Dong, J A Yoder, X Zhang, L Zhou, T H Bestor, and X Cheng. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Research*, 29(2):439–448, January 2001. PMID: 11139614.
- Erbo Dong, Ying Chen, David P Gavin, Dennis R Grayson, and Alessandro Guidotti. Valproate induces DNA demethylation in nuclear extracts from adult mouse brain. *Epigenetics: Official Journal of the DNA Methylation Society*, 5(8):730–735, December 2010. PMID: 20716949.
- J Dubochet and M Noll. Nucleosome arcs and helices. *Science (New York, N.Y.)*, 202(4365):280–286, October 1978. PMID: 694532.
- Hariharan P Easwaran, Lothar Schermelleh, Heinrich Leonhardt, and M Cristina Cardoso. Replication-independent chromatin loading of dnmt1 during g2 and m phases. *EMBO Reports*, 5(12):1181–1186, December 2004. PMID: 15550930.
- M Ehrlich, M A Gama-Sosa, L H Huang, R M Midgett, K C Kuo, R A McCune, and C Gehrke. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Research*, 10(8):2709–2721, April 1982. PMID: 7079182.
- Carola Engler, Romy Kandzia, and Sylvestre Marillonnet. A one pot, one step, precision cloning method with high throughput capability. *PloS One*, 3(11):e3647, 2008. PMID: 18985154.
- Carola Engler, Ramona Gruetzner, Romy Kandzia, and Sylvestre Marillonnet. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS One*, 4(5):e5553, 2009. PMID: 19436741.

- Silvina Epsztejn-Litman, Nirit Feldman, Monther Abu-Remaileh, Yoel Shufaro, Ariela Gerson, Jun Ueda, Rachel Deplus, Francois Fuks, Yoichi Shinkai, Howard Cedar, and Yehudit Bergman. De novo DNA methylation promoted by g9a prevents reprogramming of embryonically silenced genes. *Nature Structural & Molecular Biology*, 15(11): 1176–1183, November 2008. PMID: 18953337.
- M J Evans and M H Kaufman. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–156, July 1981. PMID: 7242681.
- M Fatemi, A Hermann, S Pradhan, and A Jeltsch. 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(5):1189–1199, June 2001. PMID: 11399088.
- Nirit Feldman, Ariela Gerson, Jia Fang, En Li, Yi Zhang, Yoichi Shinkai, Howard Cedar, and Yehudit Bergman. G9a-mediated irreversible epigenetic inactivation of oct-3/4 during early embryogenesis. *Nature Cell Biology*, 8(2):188–194, February 2006. PMID: 16415856.
- Gabriella Ficz, Miguel R Branco, Stefanie Seisenberger, Ftima Santos, Felix Krueger, Timothy A Hore, C Joana Marques, Simon Andrews, and Wolf Reik. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, 473(7347):398–402, May 2011. PMID: 21460836.
- Guillaume J P Filion, Svetlana Zhenilo, Sergey Salozhin, Daisuke Yamada, Egor Prokhortchouk, and Pierre-Antoine Defossez. A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and Cellular Biology*, 26(1):169–181, January 2006. PMID: 16354688.
- J T Finch, L C Lutter, D Rhodes, R S Brown, B Rushton, M Levitt, and A Klug. Structure of nucleosome core particles of chromatin. *Nature*, 269(5623):29–36, September 1977. PMID: 895884.
- Carina Frauer and Heinrich Leonhardt. A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Research*, 37(3):e22, February 2009. PMID: 19129216.

- Carina Frauer, Thomas Hoffmann, Sebastian Bultmann, Valentina Casa, M Cristina Cardoso, Iris Antes, and Heinrich Leonhardt. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One*, 6(6):e21306, 2011. PMID: 21731699.
- F Fuks, W A Burgers, N Godin, M Kasai, and T Kouzarides. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *The EMBO Journal*, 20(10):2536–2544, May 2001. PMID: 11350943.
- Francois Fuks, Paul J Hurd, Daniel Wolf, Xinsheng Nan, Adrian P Bird, and Tony Kouzarides. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *The Journal of Biological Chemistry*, 278(6):4035–4040, February 2003. PMID: 12427740.
- Rozenn Gallais, Florence Demay, Peter Barath, Laurence Finot, Renata Jurkowska, Rmy Le Guvel, Frdrique Gay, Albert Jeltsch, Raphal Mtivier, and Gilles Salbert. Deoxyribonucleic acid methyl transferases 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation. *Molecular Endocrinology (Baltimore, Md.)*, 21(9):2085–2098, September 2007. PMID: 17579209.
- F Gaudet, D Talbot, H Leonhardt, and R Jaenisch. A short DNA methyltransferase isoform restores methylation in vivo. *The Journal of Biological Chemistry*, 273(49):32725–32729, December 1998. PMID: 9830015.
- Francois Gaudet, J Graeme Hodgson, Amir Eden, Laurie Jackson-Grusby, Jessica Dausman, Joe W Gray, Heinrich Leonhardt, and Rudolf Jaenisch. Induction of tumors in mice by genomic hypomethylation. *Science (New York, N.Y.)*, 300(5618):489–492, April 2003. PMID: 12702876.
- Ren Geissler, Heidi Scholze, Simone Hahn, Jana Streubel, Ulla Bonas, Sven-Erik Behrens, and Jens Boch. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One*, 6(5):e19509, 2011. PMID: 21625585.
- Sharon Gidekel, Galina Pizov, Yehudit Bergman, and Eli Pikarsky. Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell*, 4(5):361–370, November 2003. PMID: 14667503.
- Mary Grace Goll and Timothy H Bestor. Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, 74:481–514, 2005. PMID: 15952895.

- Mary Grace Goll, Finn Kirpekar, Keith A Maggert, Jeffrey A Yoder, Chih-Lin Hsieh, Xiaoyu Zhang, Kent G Golic, Steven E Jacobsen, and Timothy H Bestor. Methylation of tRNA^{Asp} by the DNA methyltransferase homolog dnmt2. *Science (New York, N.Y.)*, 311(5759):395–398, January 2006. PMID: 16424344.
- H Gowher, O Leismann, and A Jeltsch. DNA of drosophila melanogaster contains 5-methylcytosine. *The EMBO Journal*, 19(24):6918–6923, December 2000. PMID: 11118227.
- Tian-Peng Gu, Fan Guo, Hui Yang, Hai-Ping Wu, Gui-Fang Xu, Wei Liu, Zhi-Guo Xie, Linyu Shi, Xinyi He, Seung-gi Jin, Khurshed Iqbal, Yujiang Geno Shi, Zixin Deng, Piroska E Szab, Gerd P Pfeifer, Jinsong Li, and Guo-Liang Xu. The role of tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 477(7366):606–610, September 2011. PMID: 21892189.
- Junjie U Guo, Yijing Su, Chun Zhong, Guo-li Ming, and Hongjun Song. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, 145(3):423–434, April 2011. PMID: 21496894.
- M Hagmann, O Georgiev, and W Schaffner. The VP16 paradox: herpes simplex virus VP16 contains a long-range activation domain but within the natural multiprotein complex activates only from promoter-proximal positions. *Journal of Virology*, 71(8):5952–5962, August 1997. PMID: 9223485.
- A T Hark, C J Schoenherr, D J Katz, R S Ingram, J M LeVorse, and S M Tilghman. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, 405(6785):486–489, May 2000. PMID: 10839547.
- Hideharu Hashimoto, John R Horton, Xing Zhang, Magnolia Bostick, Steven E Jacobsen, and Xiaodong Cheng. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, 455(7214):826–829, October 2008. PMID: 18772888.
- Hideharu Hashimoto, John R Horton, Xing Zhang, and Xiaodong Cheng. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics: Official Journal of the DNA Methylation Society*, 4(1):8–14, January 2009. PMID: 19077538.
- Hideharu Hashimoto, Paula M Vertino, and Xiaodong Cheng. Molecular coupling of DNA methylation and histone methylation. *Epigenomics*, 2(5):657–669, October 2010. PMID: 21339843.

- R David Hawkins, Gary C Hon, Leonard K Lee, Queminh Ngo, Ryan Lister, Matia Pelizzola, Lee E Edsall, Samantha Kuan, Ying Luu, Sarit Klugman, Jessica Antosiewicz-Bourget, Zhen Ye, Celso Espinoza, Saurabh Agarwahl, Li Shen, Victor Ruotti, Wei Wang, Ron Stewart, James A Thomson, Joseph R Ecker, and Bing Ren. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, 6(5):479–491, May 2010. PMID: 20452322.
- Yu-Fei He, Bin-Zhong Li, Zheng Li, Peng Liu, Yang Wang, Qingyu Tang, Jianping Ding, Yingying Jia, Zhangcheng Chen, Lin Li, Yan Sun, Xiuxue Li, Qing Dai, Chun-Xiao Song, Kangling Zhang, Chuan He, and Guo-Liang Xu. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, N. Y.)*, 333(6047):1303–1307, September 2011. PMID: 21817016.
- B Hendrich and A Bird. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology*, 18(11):6538–6547, November 1998. PMID: 9774669.
- Dirk Hockemeyer, Haoyi Wang, Samira Kiani, Christine S Lai, Qing Gao, John P Casady, Gregory J Cost, Lei Zhang, Yolanda Santiago, Jeffrey C Miller, Bryan Zeitler, Jennifer M Cherone, Xiangdong Meng, Sarah J Hinkley, Edward J Rebar, Philip D Gregory, Fyodor D Urnov, and Rudolf Jaenisch. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotechnology*, 29(8):731–734, August 2011. PMID: 21738127.
- C L Hsieh. In vivo activity of murine de novo methyltransferases, dnmt3a and dnmt3b. *Molecular and Cellular Biology*, 19(12):8211–8218, December 1999. PMID: 10567546.
- D W Hsu, M J Lin, T L Lee, S C Wen, X Chen, and C K Shen. Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17):9751–9756, August 1999. PMID: 10449766.
- Yun Huang, William A Pastor, Yinghua Shen, Mamta Tahiliani, David R Liu, and Anjana Rao. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One*, 5(1):e8888, 2010. PMID: 20126651.
- Danwei Huangfu, Ren Maehr, Wenjun Guo, Astrid Eijkelenboom, Melinda Snitow, Alice E Chen, and Douglas A Melton. Induction of pluripotent stem cells by defined

- factors is greatly improved by small-molecule compounds. *Nature Biotechnology*, 26(7):795–797, July 2008. PMID: 18568017.
- Azusa Inoue and Yi Zhang. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science (New York, N.Y.)*, 334(6053):194, October 2011. PMID: 21940858.
- Khursheed Iqbal, Seung-Gi Jin, Gerd P Pfeifer, and Pirooska E Szab. 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(9):3642–3647, March 2011. PMID: 21321204.
- Shinsuke Ito, Ana C D’Alessio, Olena V Taranova, Kwonho Hong, Lawrence C Sowers, and Yi Zhang. Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310):1129–1133, August 2010. PMID: 20639862.
- Shinsuke Ito, Li Shen, Qing Dai, Susan C Wu, Leonard B Collins, James A Swenberg, Chuan He, and Yi Zhang. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)*, 333(6047):1300–1303, September 2011. PMID: 21778364.
- Lakshminarayan M Iyer, Mamta Tahiliani, Anjana Rao, and L Aravind. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle (Georgetown, Tex.)*, 8(11):1698–1710, June 2009. PMID: 19411852.
- Melany Jackson, Anna Krassowska, Nick Gilbert, Timothy Chevassut, Lesley Forrester, John Ansell, and Bernard Ramsahoye. Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Molecular and Cellular Biology*, 24(20):8862–8871, October 2004. PMID: 15456861.
- L Jackson-Grusby, C Beard, R Possemato, M Tudor, D Fambrough, G Csankovszki, J Dausman, P Lee, C Wilson, E Lander, and R Jaenisch. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genetics*, 27(1):31–39, January 2001. PMID: 11137995.
- Isabel Jaco, Andrs Canela, Elsa Vera, and Maria A Blasco. Centromere mitotic recombination in mammalian cells. *The Journal of Cell Biology*, 181(6):885–892, June 2008. PMID: 18541703.

- R H Jacobson, A G Ladurner, D S King, and R Tjian. Structure and function of a human TAFII250 double bromodomain module. *Science (New York, N.Y.)*, 288(5470):1422–1425, May 2000. PMID: 10827952.
- L Janosi, H Yonemitsu, H Hong, and A Kaji. Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRts1I) modulated by glucosylation of DNA. *Journal of Molecular Biology*, 242(1):45–61, September 1994. PMID: 8078071.
- Laure J M Jason, Susan C Moore, John D Lewis, George Lindsey, and Juan Ausi. Histone ubiquitination: a tagging tail unfolds? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 24(2):166–174, February 2002. PMID: 11835281.
- T Jenuwein, G Laible, R Dorn, and G Reuter. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cellular and Molecular Life Sciences: CMLS*, 54(1):80–93, January 1998. PMID: 9487389.
- Da Jia, Renata Z Jurkowska, Xing Zhang, Albert Jeltsch, and Xiaodong Cheng. Structure of dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, 449(7159):248–251, September 2007. PMID: 17713477.
- Seung-Gi Jin, Swati Kadam, and Gerd P Pfeifer. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Research*, 38(11):e125, June 2010. PMID: 20371518.
- Helle F Jørgensen, Ittai Ben-Porath, and Adrian P Bird. Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Molecular and Cellular Biology*, 24(8):3387–3395, April 2004. PMID: 15060159.
- T Kafri, M Ariel, M Brandeis, R Shemer, L Urven, J McCarrey, H Cedar, and A Razin. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes & Development*, 6(5):705–714, May 1992. PMID: 1577268.
- Rohinton T Kamakaka and Sue Biggins. Histone variants: deviants? *Genes & Development*, 19(3):295–310, February 2005. PMID: 15687254.
- Masahiro Kaneda, Masaki Okano, Kenichiro Hata, Takashi Sado, Naomi Tsujimoto, En Li, and Hiroyuki Sasaki. Essential role for de novo DNA methyltransferase dnmt3a

- in paternal and maternal imprinting. *Nature*, 429(6994):900–903, June 2004. PMID: 15215868.
- Panagiota Karagianni, Larbi Amazit, Jun Qin, and Jiemin Wong. ICBP90, a novel methyl k9 h3 binding protein linking protein ubiquitination with heterochromatin formation. *Molecular and Cellular Biology*, 28(2):705–717, January 2008. PMID: 17967883.
- Masuko Katoh and Masaru Katoh. Identification and characterization of human CXXC10 gene in silico. *International Journal of Oncology*, 25(4):1193–1199, October 2004. PMID: 15375572.
- Sabine Kay and Ulla Bonas. How xanthomonas type III effectors manipulate the host plant. *Current Opinion in Microbiology*, 12(1):37–43, February 2009.
- Ray Keller. Cell migration during gastrulation. *Current Opinion in Cell Biology*, 17(5): 533–541, October 2005. PMID: 16099638.
- Jong Kyong Kim, Pierre-Olivier Estve, Steven E Jacobsen, and Sriharsa Pradhan. UHRF1 binds g9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Research*, 37(2):493–505, February 2009. PMID: 19056828.
- Tae-You Kim, Yung-Jue Bang, and Keith D Robertson. Histone deacetylase inhibitors for cancer therapy. *Epigenetics: Official Journal of the DNA Methylation Society*, 1(1):14–23, March 2006. PMID: 17998811.
- S Klimasauskas, S Kumar, R J Roberts, and X Cheng. HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, 76(2):357–369, January 1994. PMID: 8293469.
- Myunggon Ko, Hozefa S Bandukwala, Jungeun An, Edward D Lamperti, Elizabeth C Thompson, Ryan Hastie, Angeliki Tsangaratou, Klaus Rajewsky, Sergei B Koralov, and Anjana Rao. 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(35):14566–14571, August 2011. PMID: 21873190.
- Kian Peng Koh, Akiko Yabuuchi, Sridhar Rao, Yun Huang, Kerriane Cunniff, Julie Nardone, Asta Laiho, Mamta Tahiliani, Cesar A Sommer, Gustavo Mostoslavsky, Riitta Lahesmaa, Stuart H Orkin, Scott J Rodig, George Q Daley, and Anjana Rao. Tet1

- and tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, 8(2):200–213, February 2011. PMID: 21295276.
- N Konstandin, S Bultmann, A Szwagierczak, A Dufour, B Ksienzyk, F Schneider, T Herold, M Mulaw, P M Kakadia, S Schneider, K Spiekermann, H Leonhardt, and S K Bohlander. Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K.*, 25(10):1649–1652, October 2011. PMID: 21625234.
- Theo F J Kraus, Daniel Globisch, Mirko Wagner, Sabina Eigenbrod, David Widmann, Martin Münzel, Markus Müller, Toni Pfaffeneder, Benjamin Hackner, Wolfgang Feiden, Ulrich Schüller, Thomas Carell, and Hans A Kretzschmar. Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors. *International Journal of Cancer. Journal International Du Cancer*, January 2012. PMID: 22234893.
- Yotaro Kudo, Keisuke Tateishi, Keisuke Yamamoto, Shinzo Yamamoto, Yoshinari Asaoka, Hideaki Ijichi, Genta Nagae, Haruhiko Yoshida, Hiroyuki Aburatani, and Kazuhiko Koike. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Science*, February 2012. PMID: 22320381.
- Markus Kuhlmann, Branimira E Borisova, Markus Kaller, Pontus Larsson, Dirk Stach, Jianbo Na, Ludwig Eichinger, Frank Lyko, Victor Ambros, Fredrik Söderbom, Christian Hammann, and Wolfgang Nellen. Silencing of retrotransposons in dictyostelium by DNA methylation and RNAi. *Nucleic Acids Research*, 33(19):6405–6417, 2005. PMID: 16282589.
- Tilo Kunath, Marc K Saba-El-Leil, Marwa Almousailleakh, Jason Wray, Sylvain Meloche, and Austin Smith. FGF stimulation of the erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development (Cambridge, England)*, 134(16):2895–2902, August 2007. PMID: 17660198.
- M Lachner, D O'Carroll, S Rea, K Mechtler, and T Jenuwein. Methylation of histone h3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410(6824):116–120, March 2001. PMID: 11242053.

- J Lake, J Rathjen, J Remiszewski, and P D Rathjen. Reversible programming of pluripotent cell differentiation. *Journal of Cell Science*, 113 (Pt 3):555–566, February 2000. PMID: 10639341.
- Saskia M C Langemeijer, Roland P Kuiper, Marieke Berends, Ruth Knops, Mariam G Aslanyan, Marion Massop, Ellen Stevens-Linders, Patricia van Hoogen, Ad Geurts van Kessel, Reinier A P Raymakers, Eveline J Kamping, Gregor E Verhoef, Estelle Verburch, Anne Hagemeyer, Peter Vandenbergh, Theo de Witte, Bert A van der Reijden, and Joop H Jansen. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature Genetics*, 41(7):838–842, July 2009. PMID: 19483684.
- Gernot Längst and Peter B Becker. Nucleosome remodeling: one mechanism, many phenomena? *Biochimica Et Biophysica Acta*, 1677(1-3):58–63, March 2004. PMID: 15020046.
- F Larsen, G Gundersen, R Lopez, and H Prydz. CpG islands as gene markers in the human genome. *Genomics*, 13(4):1095–1107, August 1992. PMID: 1505946.
- Louise Laurent, Eleanor Wong, Guoliang Li, Tien Huynh, Aristotelis Tsirigos, Chin Thing Ong, Hwee Meng Low, Ken Wing Kin Sung, Isidore Rigoutsos, Jeanne Loring, and Chia-Lin Wei. Dynamic changes in the human methylome during differentiation. *Genome Research*, 20(3):320–331, March 2010. PMID: 20133333.
- J H Lee, K S Voo, and D G Skalnik. Identification and characterization of the DNA binding domain of CpG-binding protein. *The Journal of Biological Chemistry*, 276 (48):44669–44676, November 2001. PMID: 11572867.
- H Lei, S P Oh, M Okano, R Jüttermann, K A Goss, R Jaenisch, and E Li. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development (Cambridge, England)*, 122(10):3195–3205, October 1996. PMID: 8898232.
- H Leonhardt, A W Page, H U Weier, and T H Bestor. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, 71(5):865–873, November 1992. PMID: 1423634.
- Ivica Letunic and Peer Bork. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)*, 23(1):127–128, January 2007. PMID: 17050570.

- E Li, T H Bestor, and R Jaenisch. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6):915–926, June 1992. PMID: 1606615.
- E Li, C Beard, and R Jaenisch. Role for DNA methylation in genomic imprinting. *Nature*, 366(6453):362–365, November 1993. PMID: 8247133.
- Jing-Yu Li, Min-Tie Pu, Ryutaro Hirasawa, Bin-Zhong Li, Yan-Nv Huang, Rong Zeng, Nai-He Jing, Taiping Chen, En Li, Hiroyuki Sasaki, and Guo-Liang Xu. Synergistic function of DNA methyltransferases dnmt3a and dnmt3b in the methylation of oct4 and nanog. *Molecular and Cellular Biology*, 27(24):8748–8759, December 2007. PMID: 17938196.
- Shaohua Li and Peter D Yurchenco. Matrix assembly, cell polarization, and cell survival: analysis of peri-implantation development with cultured embryonic stem cells. *Methods in Molecular Biology (Clifton, N.J.)*, 329:113–125, 2006. PMID: 16845988.
- Ting Li, Sheng Huang, Xuefeng Zhao, David A. Wright, Susan Carpenter, Martin H. Spalding, Donald P. Weeks, and Bing Yang. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Research*, 39(14):6315–6325, 2011a.
- Zhe Li, Xiaoqiang Cai, Chen-Leng Cai, Jiapeng Wang, Wenyong Zhang, Bruce E Petersen, Feng-Chun Yang, and Mingjiang Xu. Deletion of tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*, 118(17):4509–4518, October 2011b. PMID: 21803851.
- Florian Lienert, Christiane Wirbelauer, Indrani Som, Ann Dean, Fabio Mohn, and Dirk Schübeler. Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genetics*, 43(11):1091–1097, November 2011. PMID: 21964573.
- Ryan Lister, Mattia Pelizzola, Robert H Downen, R David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R Nery, Leonard Lee, Zhen Ye, Que-Minh Ngo, Lee Edsall, Jessica Antosiewicz-Bourget, Ron Stewart, Victor Ruotti, A Harvey Millar, James A Thomson, Bing Ren, and Joseph R Ecker. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–322, November 2009. PMID: 19829295.
- Man Liu and David G Schatz. Balancing AID and DNA repair during somatic hypermutation. *Trends in Immunology*, 30(4):173–181, April 2009. PMID: 19303358.

- Man Liu, Jamie L Duke, Daniel J Richter, Carola G Vinuesa, Christopher C Goodnow, Steven H Kleinstein, and David G Schatz. Two levels of protection for the b cell genome during somatic hypermutation. *Nature*, 451(7180):841–845, February 2008. PMID: 18273020.
- Leendert H J Looijenga, Hans Stoop, Hubert P J C de Leeuw, Carlos A de Gouveia Brazao, Ad J M Gillis, Kees E P van Roozendaal, Everardus J J van Zoelen, Rob F A Weber, Katja P Wolffenbuttel, Herman van Dekken, Friedemann Honecker, Carsten Bokemeyer, Elizabeth J Perlman, Dominik T Schneider, Juha Kononen, Guido Sauter, and J Wolter Oosterhuis. POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. *Cancer Research*, 63(9):2244–2250, May 2003. PMID: 12727846.
- R B Lorschach, J Moore, S Mathew, S C Raimondi, S T Mukatira, and J R Downing. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K.*, 17(3):637–641, March 2003. PMID: 12646957.
- Dengke K Ma, Mi-Hyeon Jang, Junjie U Guo, Yasuji Kitabatake, Min-Lin Chang, Natapol Pow-Anpongkul, Richard A Flavell, Binfeng Lu, Guo-Li Ming, and Hongjun Song. Neuronal activity-induced gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science (New York, N.Y.)*, 323(5917):1074–1077, February 2009. PMID: 19119186.
- D Macleod, J Charlton, J Mullins, and A P Bird. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes & Development*, 8(19):2282–2292, October 1994. PMID: 7958895.
- Magdy M. Mahfouz, Lixin Li, Md. Shamimuzzaman, Anjar Wibowo, Xiaoyun Fang, and Jian-Kang Zhu. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proceedings of the National Academy of Sciences*, 108(6):2623–2628, February 2011.
- Atanu Maiti and Alexander C Drohat. 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(41):35334–35338, October 2011. PMID: 21862836.

- Amanda Nga-Sze Mak, Philip Bradley, Raul A Cernadas, Adam J Bogdanove, and Barry L Stoddard. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (New York, N.Y.)*, 335(6069):716–719, February 2012. PMID: 22223736.
- Keri Martinowich, Daisuke Hattori, Hao Wu, Shaun Fouse, Fei He, Yan Hu, Guoping Fan, and Yi E Sun. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science (New York, N.Y.)*, 302(5646):890–893, October 2003. PMID: 14593184.
- Robert W Maul and Patricia J Gearhart. AID and somatic hypermutation. *Advances in Immunology*, 105:159–191, 2010. PMID: 20510733.
- W Mayer, A Niveleau, J Walter, R Fundele, and T Haaf. Demethylation of the zygotic paternal genome. *Nature*, 403(6769):501–502, February 2000. PMID: 10676950.
- R R Meehan, J D Lewis, S McKay, E L Kleiner, and A P Bird. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, 58(3):499–507, August 1989. PMID: 2758464.
- Daniela Meilinger, Karin Fellingner, Sebastian Bultmann, Ulrich Rothbauer, Ian Marc Bonapace, Wolfgang E F Klinkert, Fabio Spada, and Heinrich Leonhardt. 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(11):1259–1264, November 2009. PMID: 19798101.
- Alexander Meissner, Tarjei S Mikkelsen, Hongchang Gu, Marius Wernig, Jacob Hanna, Andrey Sivachenko, Xiaolan Zhang, Bradley E Bernstein, Chad Nusbaum, David B Jaffe, Andreas Gnirke, Rudolf Jaenisch, and Eric S Lander. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205):766–770, August 2008. PMID: 18600261.
- Courtney A Miller and J David Sweatt. Covalent modification of DNA regulates memory formation. *Neuron*, 53(6):857–869, March 2007. PMID: 17359920.
- Jeffrey C Miller, Siyuan Tan, Guijuan Qiao, Kyle A Barlow, Jianbin Wang, Danny F Xia, Xiangdong Meng, David E Paschon, Elo Leung, Sarah J Hinkley, Gladys P Dulay, Kevin L Hua, Irina Ankoudinova, Gregory J Cost, Fyodor D Urnov, H Steve Zhang, Michael C Holmes, Lei Zhang, Philip D Gregory, and Edward J Rebar. A

- TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*, 29(2): 143–148, February 2011. PMID: 21179091.
- P Miniou, M Jeanpierre, D Bourc'his, A C Coutinho Barbosa, V Blanquet, and E Viegas-Piquignot. alpha-satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. *Human Genetics*, 99(6):738–745, June 1997. PMID: 9187666.
- Ahmad Miremadi, Mikkel Z Oestergaard, Paul D P Pharoah, and Carlos Caldas. Cancer genetics of epigenetic genes. *Human Molecular Genetics*, 16 Spec No 1:R28–49, April 2007. PMID: 17613546.
- Azim M Mohamedali, Alexander E Smith, Joop Gaken, Nicholas C Lea, Syed A Mian, Nigel B Westwood, Corinna Strupp, Norbert Gattermann, Ulrich Germing, and Ghulam J Mufti. Novel TET2 mutations associated with UPD4q24 in myelodysplastic syndrome. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 27(24):4002–4006, August 2009. PMID: 19528370.
- Kelly Moran-Crusio, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E Figueroa, Aparna Vasanthakumar, Jay Patel, Xinyang Zhao, Fabiana Perna, Suveg Pandey, Jozef Madzo, Chunxiao Song, Qing Dai, Chuan He, Sherif Ibrahim, Miloslav Beran, Jiri Zavadil, Stephen D Nimer, Ari Melnick, Lucy A Godley, Iannis Aifantis, and Ross L Levine. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*, 20(1): 11–24, July 2011. PMID: 21723200.
- Robert Morbitzer, Janett Elsaesser, Jens Hausner, and Thomas Lahaye. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Research*, 39(13):5790–5799, July 2011. PMID: 21421566.
- Hugh D Morgan, Wendy Dean, Heather A Coker, Wolf Reik, and Svend K Petersen-Mahrt. 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(50):52353–52360, December 2004. PMID: 15448152.
- Matthew J Moscou and Adam J Bogdanove. A simple cipher governs DNA recognition by TAL effectors. *Science (New York, N. Y.)*, 326(5959):1501, December 2009. PMID: 19933106.

- Alysson R Muotri, Maria C N Marchetto, Nicole G Coufal, Ruth Oefner, Gene Yeo, Kinichi Nakashima, and Fred H Gage. L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, 468(7322):443–446, November 2010. PMID: 21085180.
- Raphal Mtivier, Rozenn Gallais, Christophe Tiffoche, Christine Le Pron, Renata Z Jurkowska, Richard P Carmouche, David Ibberson, Peter Barath, Florence Demay, George Reid, Vladimir Benes, Albert Jeltsch, Frank Gannon, and Gilles Salbert. Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, 452(7183):45–50, March 2008. PMID: 18322525.
- X Nan, R R Meehan, and A Bird. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research*, 21(21):4886–4892, October 1993. PMID: 8177735.
- X Nan, H H Ng, C A Johnson, C D Laherty, B M Turner, R N Eisenman, and A Bird. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683):386–389, May 1998. PMID: 9620804.
- Colm Nestor, Alexey Ruzov, Richard Meehan, and Donncha Dunican. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *BioTechniques*, 48(4):317–319, April 2010. PMID: 20569209.
- H H Ng and A Bird. DNA methylation and chromatin modification. *Current Opinion in Genetics & Development*, 9(2):158–163, April 1999. PMID: 10322130.
- J Nichols, B Zevnik, K Anastassiadis, H Niwa, D Klewe-Nebenius, I Chambers, H Schöler, and A Smith. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor oct4. *Cell*, 95(3):379–391, October 1998. PMID: 9814708.
- R H Nicolas and G H Goodwin. 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(1-2):233–240, October 1996. PMID: 8917104.
- Hitoshi Niwa. Open conformation chromatin and pluripotency. *Genes & Development*, 21(21):2671–2676, November 2007. PMID: 17974911.

- D P Norris, D Patel, G F Kay, G D Penny, N Brockdorff, S A Sheardown, and S Rastan. Evidence that random and imprinted xist expression is controlled by preemptive methylation. *Cell*, 77(1):41–51, April 1994. PMID: 8156596.
- M Okano, S Xie, and E Li. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, 19(3):219–220, July 1998a. PMID: 9662389.
- M Okano, S Xie, and E Li. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Research*, 26(11):2536–2540, June 1998b. PMID: 9592134.
- M Okano, D W Bell, D A Haber, and E Li. DNA methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3):247–257, October 1999. PMID: 10555141.
- Ryoichi Ono, Tomohiko Taki, Takeshi Taketani, Masafumi Taniwaki, Hajime Kobayashi, and Yasuhide Hayashi. 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(14):4075–4080, July 2002. PMID: 12124344.
- Roberto Papait, Christian Pistore, Diego Negri, Daniela Pecoraro, Lisa Cantarini, and Ian Marc Bonapace. Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Molecular Biology of the Cell*, 18(3):1098–1106, March 2007. PMID: 17182844.
- William A Pastor, Utz J Pape, Yun Huang, Hope R Henderson, Ryan Lister, Myunggon Ko, Erin M McLoughlin, Yevgeny Brudno, Sahasransu Mahapatra, Philipp Kapranov, Mamta Tahiliani, George Q Daley, X Shirley Liu, Joseph R Ecker, Patrice M Milos, Suneet Agarwal, and Anjana Rao. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 473(7347):394–397, May 2011. PMID: 21552279.
- N W Penn, R Suwalski, C O’Riley, K Bojanowski, and R Yura. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *The Biochemical Journal*, 126(4):781–790, February 1972. PMID: 4538516.
- Toni Pfaffeneder, Benjamin Hackner, Matthias Truss, Martin Münzel, Markus Müller, Christian A Deiml, Christian Hagemeyer, and Thomas Carell. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie (International Ed. in English)*, 50(31):7008–7012, July 2011. PMID: 21721093.

- Sameer Phalke, Olaf Nickel, Diana Walluscheck, Frank Hortig, Maria Cristina Onorati, and Gunter Reuter. Retrotransposon silencing and telomere integrity in somatic cells of drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nature Genetics*, 41(6):696–702, June 2009. PMID: 19412177.
- Garwin Pichler, Patricia Wolf, Christine S Schmidt, Daniela Meilinger, Katrin Schneider, Carina Frauer, Karin Fellingner, Andrea Rottach, and Heinrich Leonhardt. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *Journal of Cellular Biochemistry*, 112(9):2585–2593, September 2011. PMID: 21598301.
- Christian Popp, Wendy Dean, Suhua Feng, Shawn J Cokus, Simon Andrews, Matteo Pellegrini, Steven E Jacobsen, and Wolf Reik. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, 463(7284):1101–1105, February 2010. PMID: 20098412.
- S Pradhan, D Talbot, M Sha, J Benner, L Hornstra, E Li, R Jaenisch, and R J Roberts. Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase. *Nucleic Acids Research*, 25(22):4666–4673, November 1997. PMID: 9358180.
- A Prokhortchouk, B Hendrich, H Jrgensen, A Ruzov, M Wilm, G Georgiev, A Bird, and E Prokhortchouk. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & Development*, 15(13):1613–1618, July 2001. PMID: 11445535.
- Kunal Rai, Stephanie Chidester, Chad V Zavala, Elizabeth J Manos, Smitha R James, Adam R Karpf, David A Jones, and Bradley R Cairns. Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes & Development*, 21(3):261–266, February 2007. PMID: 17289917.
- Kunal Rai, Ian J Huggins, Smitha R James, Adam R Karpf, David A Jones, and Bradley R Cairns. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*, 135(7):1201–1212, December 2008. PMID: 19109892.
- Efren Riu, Zhi-Ying Chen, Hui Xu, Chen-Yi He, and Mark A Kay. Histone modifications are associated with the persistence or silencing of vector-mediated transgene

- expression in vivo. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 15(7):1348–1355, July 2007. PMID: 17457320.
- Guido Van Rossum. *The Python Language Reference Manual*. Network Theory Limited, September 2003. ISBN 9780954161781.
- Andrea Rottach, Heinrich Leonhardt, and Fabio Spada. DNA methylation-mediated epigenetic control. *Journal of Cellular Biochemistry*, 108(1):43–51, September 2009. PMID: 19565567.
- Andrea Rottach, Carina Frauer, Garwin Pichler, Ian Marc Bonapace, Fabio Spada, and Heinrich Leonhardt. The multi-domain protein np95 connects DNA methylation and histone modification. *Nucleic Acids Research*, 38(6):1796–1804, April 2010. PMID: 20026581.
- N Rougier, D Bourc’his, D M Gomes, A Niveleau, M Plachot, A Pldi, and E Viegas-Piquignot. Chromosome methylation patterns during mammalian preimplantation development. *Genes & Development*, 12(14):2108–2113, July 1998. PMID: 9679055.
- M R Rountree, K E Bachman, and S B Baylin. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature Genetics*, 25(3):269–277, July 2000. PMID: 10888872.
- Jeffrey D Sander, Lindsay Cade, Cyd Khayter, Deepak Reyon, Randall T Peterson, J Keith Joung, and Jing-Ruey J Yeh. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature Biotechnology*, 29(8):697–698, August 2011. PMID: 21822241.
- J P Sanford, H J Clark, V M Chapman, and J Rossant. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes & Development*, 1(10):1039–1046, December 1987. PMID: 3428592.
- Ftima Santos, Brian Hendrich, Wolf Reik, and Wendy Dean. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Developmental Biology*, 241(1):172–182, January 2002. PMID: 11784103.
- Nobuhiro Sasai and Pierre-Antoine Defossez. Many paths to one goal? the proteins that recognize methylated DNA in eukaryotes. *The International Journal of Developmental Biology*, 53(2-3):323–334, 2009. PMID: 19412889.

- Lothar Schermelleh, Fabio Spada, Hariharan P Easwaran, Kouros Zolghadr, Jean B Margot, M Cristina Cardoso, and Heinrich Leonhardt. Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nature Methods*, 2(10):751–756, October 2005. PMID: 16179921.
- Lothar Schermelleh, Andrea Haemmer, Fabio Spada, Nicole Rösing, Daniela Meilinger, Ulrich Rothbauer, M Cristina Cardoso, and Heinrich Leonhardt. Dynamics of dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Research*, 35(13):4301–4312, 2007. PMID: 17576694.
- Lena Scheubert, Rainer Schmidt, Dirk Repsilber, Mitja Lustrek, and Georg Fuellen. Learning biomarkers of pluripotent stem cells in mouse. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 18(4):233–251, 2011. PMID: 21791477.
- Jafar Sharif, Masahiro Muto, Shin-ichiro Takebayashi, Isao Suetake, Akihiro Iwamatsu, Takaho A Endo, Jun Shinga, Yoko Mizutani-Koseki, Tetsuro Toyoda, Kunihiro Okamura, Shoji Tajima, Kohzoh Mitsuya, Masaki Okano, and Haruhiko Koseki. The SRA protein np95 mediates epigenetic inheritance by recruiting dnmt1 to methylated DNA. *Nature*, 450(7171):908–912, December 2007. PMID: 17994007.
- Jose Silva and Austin Smith. Capturing pluripotency. *Cell*, 132(4):532–536, February 2008. PMID: 18295569.
- Peter J Skene, Robert S Illingworth, Shaun Webb, Alastair R W Kerr, Keith D James, Daniel J Turner, Rob Andrews, and Adrian P Bird. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Molecular Cell*, 37(4):457–468, February 2010. PMID: 20188665.
- Austin Smith. The battlefield of pluripotency. *Cell*, 123(5):757–760, December 2005. PMID: 16325569.
- Fabio Spada, Andrea Haemmer, David Kuch, Ulrich Rothbauer, Lothar Schermelleh, Elisabeth Kremmer, Thomas Carell, Gernot Längst, and Heinrich Leonhardt. DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *The Journal of Cell Biology*, 176(5):565–571, February 2007. PMID: 17312023.