STUDIES OF MECHANISMS REGULATING TET PROTEINS

MENGXI WANG

MÜNCHEN 2014

STUDIES OF MECHANISMS REGULATING TET PROTEINS

Dissertation

an der Fakultät für Biologie

der Ludwig-Maximilians-Universität

München

vorgelegt von

Mengxi Wang

aus Beijing

München, den 2014-02-10

Erstgutachter: Prof. Dr. Heinrich Leonhardt

Zweitgutachter: Prof. Dr.Bettina Kempkes

Tag der mündlichen Prüfung: 2014-03-14

CONTENT

SUMMARY

DNA methylation plays a central role in the epigenetic system of mammalian cells and in controlling gene expression. DNA methyltransferase 1 (Dnmt1) maintains the DNA methylation patterns after DNA replication. Removal of methylation can occur passively due to replication in the absence of DNA methyltransferases, with consequent dilution of this methylation modification. In addition, there is evidence supporting the occurrence of active demethylation in mammals. Ten-eleven translocation (Tet) dioxygenases, promote DNA demethylation by converting 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). However, it remains elusive how Tet proteins and their enzymatic activity are regulated.

The aim of this study is to elucidate the multifaceted regulation mechanisms of Tet proteins. It was known that Tet1 has an ancestral intramolecular CXXC domain, and another gene encodes a standalone CXXC domain protein near *tet2* in the same chromosome. Our research provided the first experimental evidence that Tet3 also harbors a CXXC. The goal was to achieve a better understanding of the complex functions of CXXC modules on Tet proteins. For this purpose, a variety of *in vivo* and *in vitro* assays were applied. CXXC motifs were reported to mediate chromatin binding in many proteins, and in our research, interaction between CXXC and Tet proteins were detected. Given that, it is assumed that the CXXC interactions have an impact on the gene targeting and chromatin binding dynamics of Tet proteins. To address this question, *in vitro* DNA competition affinity assays were performed. Among the unmodified, methylated or hydroxymethylated DNA substrates, CXXC domains share a similar binding preference of cytosine modification states. More specifically, 5mC and unmodified cytosine bind to CXXC motifs, but 5hmC shows a significant inhibition effect on DNA binding. Furthermore, live cell photobleaching analysis implied a slightly lower mobility and a weak increase in the immobile fraction of the CXXC-encoding Tet3 isoform. Although genomic 5hmC abundance was similar between cells expressing Tet3 isoforms with or without CXXC domain, significant transcriptional difference between the two Tet3 isoforms was detected in various mouse tissues. Our data suggest that variable association with CXXC modules may contribute to the regulation of Tet proteins.

To shed light on the regulation pathways in the endogenous context, we developed tools and techniques to monitor the Tet proteins in their natural, cellular environment. A set of monoclonal antibodies against Tet proteins were generated, to study Tet proteins with different cell biological and biochemical assays. Tet antibodies were utilized to perform co-immunoprecipitation assays followed by mass spectrometry analyses. Interactome identification of Tet proteins revealed that O-linked N-GlcNAc transferase (Ogt) is a major partner of all three Tet proteins. GlcNAcylation is a highly abundant post-translational modification in cells. Ogt and O-GlcNAcase (Oga) are the main enzymes catalyzing addition and removal of O-linked N-GlcNAcylation. Furthermore, antibody staining suggested a direct correlation between endogenous Tet, Ogt and histone marks associated with active transcription. We could confirm a regulatory function of Ogt/Oga on the glycosylation status of Tet proteins. To resolve further details of the interaction, a fine

mapping of Tet interaction domains was performed. I generated a library of single domains and deletion mutants of Tet proteins. The *in vivo* fluorescence three-hybrid assay demonstrated that the catalytic domain is the major interaction platform.

In conclusion, this work contributes to the elucidation of mechanisms regulating Tet proteins and active DNA demethylation.

1 INTRODUCTION

1.1 DNA methylation mediated epigenetic regulation

1.1.1 Overview of epigenetics

The terminology “epigenetics” describes heritable functionally relevant modifications of the genome that do not involve a change in the DNA sequence. Historically, the term “epigenetics” has been under discussion for a long time (Bird, 2007). At a Cold Spring Harbor meeting in 2008, epigenetics was defined as: “stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" (Berger et al., 2009).

For cell division and differentiation undergoing in multicellular organisms, epigenetic modifications play a crucial role in defining cell fate as well as restricting the cell developmental potential. Specific epigenetic processes result in gene activation and silencing, gene imprinting, X chromosome inactivation, reprogramming, and are involved in the progress of carcinogenesis. Thus, cells could maintain a long-term cellular status or differentiate into new cell types without changing their genetic identity.

Chromatin is a complex of DNA and histone proteins, and localized in the cell nucleus. The repeating unit of chromatin is the nucleosome, which is a 146 bp-long DNA sequence assembled with a histone octamer, including H2A, H2B, H3 and H4 (Luger et al., 1997). The eukaryotic genome is broadly organized into higher structural chromatin domains, called euchromatin and heterochromatin (Tremethick, 2007; Woodcock and Dimitrov, 2001). DNA staining dyes, like DAPI, normally stain heterochromatin more intensely. This indicates that heterochromatin is tighter packed with less gene activation, while euchromatin is less condensed and actively transcribed. Histone and DNA modifications differ between heterochromatin and euchromatin (Huisinga et al., 2006; Noma et al., 2001).

Several epigenetic phenotypes play roles in controlling the formation and condensation of chromatin structures. And the best-studied epigenetic mechanisms are DNA methylation and post-translational histone modifications. Those epigenetic marks can be inherited through many cell generations, and thus stabilize the status of cells, which is also known as cell memory (Bryan et al., 2002; Bird, 2002). The epigenetic system is under regulation of different pathways, which allows adaptation to molecular stimuli, and thus response to the environmental or developmental signaling.

1.1.2 5-Methylcytosine in DNA

Modifications of nucleic acid bases are observed in both prokaryotic and eukaryotic cells. The best-studied modification is the methylation of cytosine residues in mammalian cells, describing a stably addition of a methyl group to 5-carbon atom. This process occurs post-replicative and is catalyzed by a family of DNA methyltransferases.

In 1950, Wyatt was the first that clearly confirmed 5-methylcytosine (5mC) from DNA by chromatographic and spectral analyses (Wyatt, 1950; Wyatt, 1951). Today we know that DNA methylation is present in all kingdoms of life, and the level of DNA methylation is biologically and functionally relevant. DNA methylation precisely controls the gene expression pattern in different cell types and is strongly associated with transcriptional repression (Busslinger et al., 1983; Jaenisch and Bird, 2003). Promoters of housekeeping genes are unmethylated in most cell types, whereas tissue-specific genes are found to be unmethylated in corresponding tissues while highly methylated in others (Naveh-Many and Cedar, 1981). This indicates that DNA methylation undergoes dynamic changes during the whole development, and finally leads to distinct cell types that make up organisms. CpG di-nucleotide is the most well-studied and frequent DNA context that methylation modification occurs. CpG-enriched regions, which are also called CpG islands, are often found in the 5' regulatory region and gene body of many genes (Maunakea et al., 2010). The methylation status of promoter regions is important for controlling gene transcription (Gardiner-Garden and Frommer, 1987; Saxonov et al., 2006). In addition, gene body methylation of high-level expression genes was also reported recently (Ball et al., 2009), as well as non-CpG methylation (cytosines that do not precede guanines) in embryonic stem cells (Dodge et al., 2002; Lister et al., 2009).

Cell biology and molecular biology studies provide an understanding of how DNA methylation regulates gene transcription. An important mechanism is that DNA methylation changes the accessibility of transcription complexes to their responsive DNA binding sites at gene promoters. More specifically, methylated DNA can be bound by proteins known as methyl-CpG-binding domain proteins (Mbds). These proteins could recruit additional proteins, such as histone deacetylases (Hdacs), lysine-residues methyltransferases (HMTs) and other chromatin associated proteins (Nan et al., 1998; Fuks et al., 2003), to the locus and thereby control chromatin compaction. DNA methylation can also prevent transcription activation proteins from binding to CpG sites. This mechanism is supported by many locus-specific examples (Bell and Felsenfeld, 2000). Evidence is that the CXXC finger protein 1 (Cfp1), a protein recruiting transcription activation mark H3K4me3, only specifically localizes to unmethylated cytosine (Thomson et al., 2012). Moreover, there is also a hypothesis that the methylation of DNA itself may physically impede the binding of transcription complexes to the gene promoters (Choy et al., 2010).

Correct DNA methylation is of great importance for normal cellular functions, and aberrant patterns are sometimes associated with diseases as well as cancer pathology. Genome-wide hypomethylation of DNA during progression of malignancy has been found in many tumor cell lines (Esteller, 2008; Feinberg et al., 2004; Hansen et al., 2011). Global

hypomethylation might have several effects, including chromatin instability (Eden et al., 2003; Karpf and Matsui, 2005), reactivation of transposon (Bestor, 2005), and loss of imprinting genes (Cui et al., 2003; Holm et al., 2005). Hypermethylation of tumor-suppressor genes is another mechanism of carcinogenesis (Esteller, 2008; Baylin et al., 1986). The understanding of epigenetic oncogenesis pathways provides basis for drug development. 5-azacytidine is an inhibitor of C5 DNA methyltransferase and it is used in the treatment of myelodysplastic syndrome, a disease with the potential to progress into acute myelogenous leukemia (Voso et al., 2009; Christman, 2002).

Both DNA methylation and histone modification are involved in establishing patterns of gene repression during development. Different amino acid residues of histones are substrates for several types of modification, including acetylation, methylation, phosphorylation and ubiquitination. There is a functional relationship between DNA methylation and chemical modifications of histone tails. Recent studies suggest that the establishment of histone modification pattern mediates DNA methylation (Ooi et al., 2007). At the post-implantation stage, histone modification plays a role in controlling the establishment of *de novo* DNA methylation patterns and thus together silence pluripotent genes, such as octamer-binding transcription factor 4 (Oct4), in early development. The lysine residue H3K9 of histone 3 is firstly deacetylated, and subsequently H3K9 can be methylated by a complex containing the G9a protein, which is a histone methyltransferase (Tachibana et al., 2002; Freitag and Selker 2005). This modification further mediates heterochromatinization, and these combined processes finally stabilize the gene silencing by recruiting DNA methyltransferase 3a and 3b (Dnmt3a and 3b) to establish the DNA methylation pattern (Feldman et al., 2006; Epsztejn-Litman et al., 2008). Conversely, DNA methylation is also important in maintaining histone modification patterns. Researches demonstrated that DNA methylation mediates histone H3 lysine 9 (H3K9) methylation and inhibits histone H3 lysine 4 (H3K4) methylation, both of which are evidence for chromatin compaction (Hashimshony et al., 2003; Lande-Diner et al., 2007).

Although DNA methylation is chemically stable, there are two waves of global demethylation during development. When fertilization, DNA methylation marks can be removed by an epigenetic “reprogramming”. The erasure of gametic marks during embryo formation is very important for establishing totipotency (Reik et al., 2010). Widespread reprogramming also takes place in primordial germ cells (PGCs). At this stage, chromatin undergoes global alterations and parental imprints are erased (Surani, 2001). The reprogramming processes will be discussed further in later chapter (chapter 1.2).

1.1.3 DNA methyltransferases

As illustrated above, DNA methylation plays a crucial role in many cellular processes. The mechanisms of the initial establishment (*de novo* methylation) and maintenance of DNA methylation during replication are of great research interest. DNA methyltransferases are the key enzymes to catalyze these reactions, and contain several members: Dnmt1, Dnmt2, Dnmt3a, Dnmt3b and Dnmt3L.

1.1.3.1 Dnmt1

Dnmt1 is a DNA methyltransferase that is most abundant and ubiquitously expressed in mammalian cells. It was first identified and cloned from mouse cells (Bestor et al., 1988). On one hand, it plays a central role in maintenance of DNA methylation patterns during DNA replication (Li et al., 1992) and DNA repair (Mortusewicz et al., 2005). During replication, the synthesis of a new DNA strand generates hemi-methylated DNA strands. Using the template strand with an established methylation pattern, Dnmt1 modifies the newly synthesized strand to generate symmetric methylation patterns (Razin and Riggs, 1980). Biochemical research indicates that Dnmt1 has a 5- to 30-fold preferentially binding of hemi-methylated DNA substrates over unmethylated substrates, which is consistent with its maintenance function (Yoder et al., 1997; Frauer and Leonhardt, 2009). On the other hand, Dnmt1 is also reported to contribute to *de novo* methylation in cancer cells (Jair et al., 2006; Ting et al., 2006).

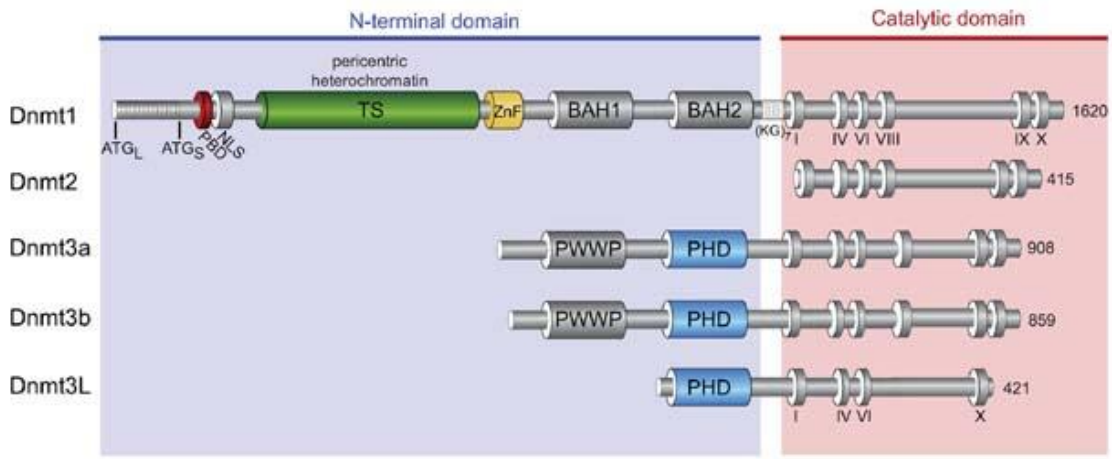


Figure 1: Structure of the mammalian DNA methyltransferase family. The Dnmts differ in their N-terminal domains, which are their regulatory regions. Abbreviations indicate: PBD: PCNA binding domain; TS: pericentric heterochromatin targeting sequence; ZnF: CXXC-type zinc finger motif; BAH: bromo adjacent homology domains; PWWP: domain with a conserved Pro-Trp-Trp-Pro motif; PHD: plant homeodomain. The C-terminal parts of all Dnmts share a similar catalytic domain with conserved motifs (I–X) (Rottach et al., 2009).

Overall, Dnmt1 is composed of two parts, the C-terminal catalytic domain and the remaining N-terminal regulatory domain, which are connected by a flexible Lys-Gly (KG)-repeat linker (Figure 1). The N-terminal domain of Dnmt1 harbors a proliferating cell nuclear antigen (PCNA) binding domain (PBD), targeting sequence (TS domain), a CXXC-type zinc finger domain (ZnF), and two bromo adjacent homology domains (BAH1/2) (Figure 1). It was reported that in absence of the regulatory domain, the isolated C-terminal domain of Dnmt1 is catalytically inactive, suggesting that the

regulatory domain may play a role in enzymatic activity of Dnmt1 (Fatemi et al., 2001; Pradhan et al., 2008; Margot et al., 2000).

In the N-terminus of Dnmt1, the targeting sequence (TS) domain presents very important regulatory function. Protein interaction assay demonstrated that the N-terminal domain of Dnmt1 can form stable dimers by hydrophobic interaction between two TS-domains (Fellinger et al., 2009). Furthermore, in late S to G2 phase of cell cycle, the TS domain and PBD domain have been found to mediate association of Dnmt1 to heterochromatin (Leonhardt et al., 1992; Easwaran et al., 2004). The TS domain of Dnmt1 shows necessity in interacting with the Uhrf1 protein (ubiquitin-like, containing PHD and RING finger domains, 1), which is an important regulator of Dnmt1 (Achour et al., 2008; Frauer and Leonhardt, 2011; Sharif et al., 2007). Another important domain is embedded between TS and BAH regions: the cysteine-rich CXXC (C is cysteine; X is any amino acid) domain that is also found in many other CpG dinucleotides binding proteins.

The carboxyl-terminal part of Dnmt1 shows sequence similarity with prokaryotic DNA methyltransferases. These C-terminal parts of Dnmt1/2/3 are highly conserved with 10 motifs, which have specific structural roles and are involved in catalysis (Cheng et al., 1993; Kumar et al., 1994). More specifically, the motifs I and X together create the binding site for reaction cofactor S-adenosyl-L-methionine (SAM). Motif VI transiently protonates the N3 position of target cytosine by a glutamyl residue and will stabilize the DNA-protein complex. Between motif VIII and IX, there is a sequence-specific contact region, named as target recognition domain (Bestor, 2000; Hermann et al., 2004; Jeltsch, 2002). In motif IV, there is a conserved prolylcysteinyl (PC) dipeptide center. The cysteine thiolate in PC center could form a covalent bond with the C6 of cytosine, and this formation activates the C5 atom for electrophilic attack (Bestor and Verdine, 1994; Kumar et al., 1994). The PC center is necessary for the enzymatic function of Dnmts, and mutation of the cysteine in PC center will cause the loss of methyltransferase activity (Hsieh, 1999).

In 2011 and 2012, crystal structures of multi-domain Dnmt1 fragments were resolved (Takeshita et al., 2011; Song et al., 2012b). Takeshita et al. reported the crystal structure of mouse Dnmt1 AA291–1620, under the condition that cofactor SAM and its product S-adenosyl-L-homocysteine (SAH) presenting. It was shown that the TS domain inserts deeply into the DNA binding pocket formed by catalytic domain of Dnmt1. Thus position changing of TS domain and the CXXC motif is required in order to allow DNA binding and succeeding reactions. The binding of SAM flips the position of a cysteine residue in the PC dipeptide center towards the target cytosine (Takeshita et al., 2011). Taking together, maintenance DNA methylation is a multi-step process and structural changes of Dnmt1 domains occur during the process.

The biological function of Dnmt1 is not only controlled by its sequence and structure, but also by its interacting proteins. Many interaction partners of Dnmt1 assist during DNA methylation process (Figure 2, Qin et al., 2011a). For example when DNA is replicated, Dnmt1 is recruited to replication forks by interaction with proliferating cell nuclear antigen (PCNA) via PBD domain of Dnmt1 (Chuang et al., 1997; Leonhardt et al., 1992;

Easwaran et al., 2004). This interaction between Dnmt1 and PCNA is highly dynamic during cell cycle, and increases the methylation activity of approximately 2-fold (Schermelleh et al., 2007; Spada et al., 2007).

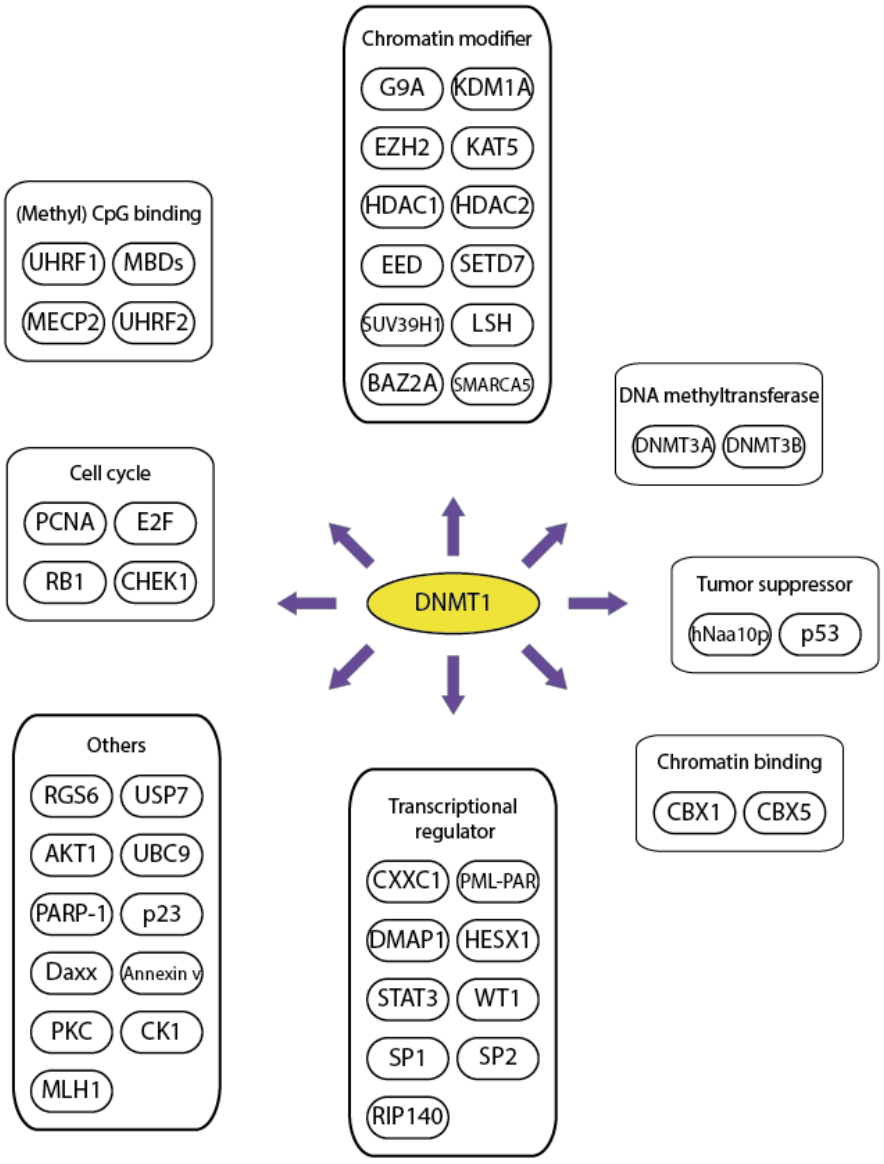


Figure 2: Overview of Dnmt1 interacting proteins. Interaction proteins of Dnmt1 include many different classes, such as DNA methyltransferases, DNA binding proteins, chromatin binding proteins, tumor suppressors and transcriptional regulators. (Modified from Qin et al., 2011a; Ruzov et al., 2009).

Another important interaction partner is the multi-domain protein Uhrf1 (also known as Np95), which is involved in directing Dnmt1 to modification sites (Bostick et al., 2007; Sharif et al., 2007). Uhrf1 protein recognizes hemi-methylated, rather than fully-methylated DNA, via its SET and RING-associated (SRA) domain (Kim et al., 2009; Arita et al., 2008; Sharif et al., 2007; Rottach et al., 2010). Uhrf1 also interacts with many transcription-repressive marks and then mediates gene silencing via Dnmt1 (Kim et al., 2009; Nady et al., 2011).

Dnmt1 undergoes several post-translational modifications, including phosphorylation, methylation, ubiquitination and SUMOylation, which regulate the stability and functions of Dnmt1 (Sugiyama et al., 2010; Wang et al., 2009; Agoston et al., 2005; Lee and Muller 2009). It has been proposed that phosphorylation increases stability of Dnmt1 and regulates its DNA binding activity (Sun et al., 2007). The methylation of several lysine residues of Dnmt1 was reported to destabilize Dnmt1 via proteasomal degradation, which is controlled by histone methyltransferase Set7 and demethylase Lsd1 (also known as Kdm1) (Esteve et al., 2009; Wang et al., 2009).

Ubiquitination-mediated proteasomal degradation is also involved in controlling the stability of Dnmt1 (Agoston et al., 2005). Dnmt1 and Uhrf1 were shown to interact with ubiquitin specific peptidase 7 (Usp7), which protein belongs to the ubiquitin specific peptidase class of deubiquitinating enzymes. Usp7, together with Uhrf1 as E3-ligase, regulates the ubiquitination of Dnmt1, and thus controls stability of Dnmt1 via ubiquitin-mediated proteasomal degradation (Qin et al., 2011b; Bronner, 2011). Research demonstrated that human DNMT1 is a substrate for acetylation by a protein lysine acetyltransferase 5 (KAT5, also known as Tip60) and this reaction triggers ubiquitination of DNMT1 by the E3-ligase activity of UHRF1 (Du et al., 2011). In contrast, DNMT1 is deacetylated by histone deacetylase 1 (HDAC1) and deubiquitinated by USP7. Proteasomal degradation of DNMT1 is promoted by acetylation via KAT5 and UHRF1; while DNMT1 is protected and stabilized by HDAC1 and USP7. These antagonistic reactions regulate the stability of DNMT1 (Du et al., 2011).

1.1.3.2 The Dnmt3 family

In 1998, the group from Okano cloned and characterized the enzymatic function of two members of the Dnmt family, Dnmt3a and 3b (Okano et al., 1998). Dnmt3a and Dnmt3b are known to catalyze *de novo* methylation reactions in early developmental stages (Goll and Bestor, 2005; Hermann et al., 2004). These enzymes mainly participate in establishing methylation patterns during embryogenesis (Hermann et al., 2004) and in site-specific methylation of cancer-associated genes during tumorigenesis (Roll et al., 2008; Rhee et al., 2003). Both enzymes are comprised of a regulatory N-terminal domain and a catalytic C-terminal domain that bears all MTase motifs (Figure 1). Dnmt3a and 3b show 84% identity in their C-terminal regions.

Later, researchers identified a DNA methylation regulator, Dnmt3-like (Dnmt3L) protein, which contains an N-terminal part similar to Dnmt3a and 3b but lacks the catalytic motifs. Knock-out mice of Dnmt3L show an abnormal sex-specific *de novo* methylation in germ cells, indicating the necessity of Dnmt3L in establishing methylation patterns during development, even though Dnmt3L lacks methyltransferase activity (Webster et al., 2005). It was also shown that H3K4-methylation affects the binding of Dnmt3L to H3 histone. This evidence suggests that Dnmt3L is a necessary modulator in *de novo* DNA methylation, and functionally links DNA methylation with histone modifications (Hu et al., 2009, Ooi et al., 2007).

1.1.3.3 Dnmt2

Dnmt2 was discovered by molecular informational manipulation (Okano et al., 1998). It widely distributes among different species and is much conserved from yeast to mammals (Yoder and Bestor, 1998). In contrast to Dnmt1 and 3, Dnmt2 only contains a C-terminal catalytic domain and is located in the cytoplasm. It has been shown that Dnmt2 has methylation activity on $tRNA^{ASP}$, specifically on cytosine 38 in the anticodon loop of $tRNA^{ASP}$, indicating that Dnmt2 is actually a RNA methyltransferase rather than a DNA modifier (Goll et al., 2006; Jurkowski et al., 2008).

1.1.3.4 Catalytic mechanism of Dnmts

A 'base-flipping' is a conserved mechanism (Cheng and Roberts, 2001) that has been studied best in the bacterial 5mC methyltransferase (MTase) M.HhaI (Klimasauskas et al., 1994). Base-flipping is recognized as a necessary strategy for nucleotide access for many enzyme classes, such as Dnmts (Cheng and Blumenthal, 2008), non-catalytic transcription factors, Uhrf1 (Arita et al., 2008) and endonucleases (Horton et al., 2006). Briefly, this mechanism describes a process that the Mtases extract the base from a DNA molecule and insert it into a typically concave catalytic pocket, covalently attack to C6 of cytosine, transfer a methyl group to activate C5, and followed by many consecutive releasing steps.

The reaction cofactor SAM is a really effective donor for methyl moieties (Figure 3). With prokaryotic cytosine methyltransferase M.HhaI as model, research indicated that DNA Mtases have high binding affinity to SAM and buries SAM into a hydrophobic pocket formed by C-terminal motifs of Dnmts (Kumar et al., 1994; Jeltsch, 2002). Some hydrophobic interactions exist between certain amino acids in Dnmts and corresponding atoms in SAM (Roth et al., 1998).

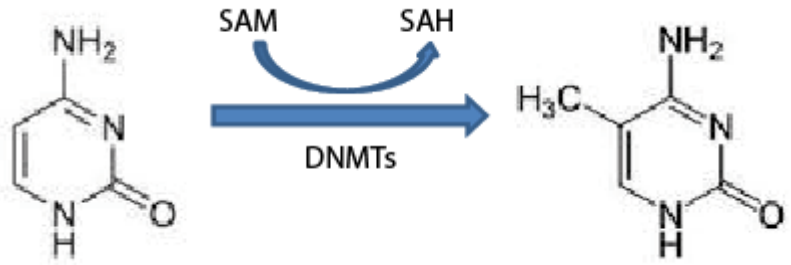


Figure 3: Cytosine is methylated at C5 position, catalyzed by Dnmt proteins. The donor of methyl moiety is a cofactor S-adenosyl-methionine (SAM). The reaction produces S-adenosyl-L-homocysteine (SAH). DNA methylation is a central epigenetic mark in controlling gene expression.

In conclusion, epigenetics is very important research field in cell biology. The epigenetic marks, including DNA methylation and histone modifications, play central roles in regulating many cellular processes. DNA methylation is catalyzed by a family of Dnmt proteins. Dnmt1, which has most abundance and ubiquitously expresses in mammalian cells, is responsible for maintenance and propagation of DNA methylation pattern to the

next generation. Our functional study and characterization on the structure of Dnmt1 showed that a motif in N-terminal part, the CXXC zinc finger, selectively binds DNA substrates containing unmethylated CpG sites. Anyhow, the CXXC-lacking mutation of Dnmt1 can still efficiently rescue DNA methylation patterns in *dnmt1*$^{-/-}$ embryonic stem cells and form covalent complex with cytosines, indicating the CXXC motif of Dnmt1 might participate, but not be indispensible in regulating Dnmt1 function (Frauer et al., 2011, see chapter 2.1).

1.2 Reprogramming and DNA demethylation

During mammalian development, two waves of large-scale reprogramming occur, including a series of chromatin remodeling. DNA demethylation is a landmark event during reprogramming. There are various hypotheses raised to explain the mechanism of active demethylation. Until now, the oxidative removal of methyl moiety catalyzed by Tet family is the most well-studied and credible model. Tet family is a class of 2-oxoglutarate and Fe (II)-dependent dioxygenase, and contains Tet1, 2 and 3. Tet1 and Tet2 participate in epigenetic regulation at the blastocyst pluripotent stage, and Tet2 is also closely associated in hematopoietic cell differentiation and mutagenesis. Tet3 is a key factor in totipotent zygotes and widely exists in many somatic cell lineages.

1.2.1 Cellular reprogramming

1.2.1.1 Reprogramming in early development

The formation of the zygote symbolizes the starting point of development. Fusion of two highly differentiated gametes becomes a zygote and it reacquires totipotency after the fertilization process. The two parental genomes in the zygote are asymmetrically organized (Van der Heijden et al., 2005; Shi and Wu, 2009). Dramatic change of the epigenome takes place and the characteristic epigenetic profile is very crucial for further development. For example, the paternal genome is wrapped by protamines at final stages of spermatogenesis to condensate chromatin (Kimmins and Sassone-Corsi, 2005). After fertilization, histone H3.3 shortly incorporates into the male pronucleus to replace protamines (Torres-Padilla et al., 2006). Also many post-translational modifications of histones of both pronuclei take place during the few cell cycles (Feil, 2009; Figure 4).

Dynamic changes of DNA methylation patterns are significant events occurring after fertilization, particularly an active DNA demethylation wave of paternal pronucleus in a few hours after fusion. Active demethylation describes the process that the paternal genome undergoes a replication-independent, large-scale demethylation in some kinds of mammal. This event is detected by both indirect immunofluorescence (Santos et al., 2002; Mayer et al., 2000) and bisulfite genomic sequencing technology (Oswald et al., 2000). In contrary, the maternal pronucleus is resistant to nuclear reprogramming and preserves DNA methylation, until the passive demethylation begins after the formation of the syngamy (Aoki et al., 1997).

While the first cycle of DNA replication starts, fused diploid genome is passively demethylated (Rougier et al., 1998; Barton et al., 2001). The maintenance DNA methyltransferase Dnmt1 is absent during this time. This mechanism is termed as passive demethylation because its dependence on mitotic cycles, whereas followed by Dnmt3-catalyzed *de novo* methylation in embryos (Figure 4).

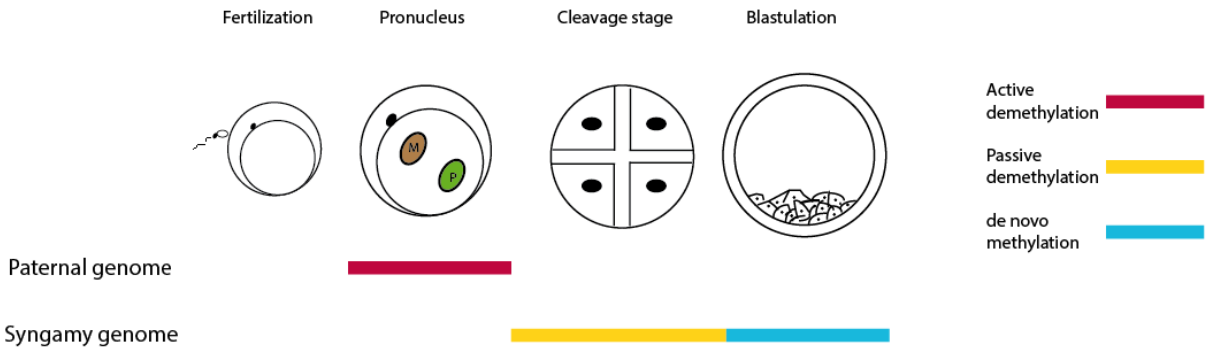


Figure 4: Dynamic changing of genome methylation in mammalian preimplantation embryonic development. Dramatic epigenetic changes occur during the early developmental stages. After fusion of two gametes, paternal genome experiences replacement of protamines, active DNA demethylation and many modifications of histones. After fusion of two haploid genomes, the syngamy takes place passive demethylation and later re-establishes DNA methylation pattern via *de novo* mechanism.

Another large-scale chromatin reprogramming *in vivo* occurs in primordial germ cells (PGCs), the precursors setting up the male and female germ lines. During gastrulation stage, the germ line specification is inducted by signaling system in embryo, and thus actively decides cell fate to form either oocyte or sperm in later development. The commitment of PGCs involves repression of somatic differentiation genes and activation of germ cell lineage genes (Saitou and Yamaji, 2012).

Early PGCs show similar epigenetic profiles compared to epiblast cells, including X chromosome inactivation, DNA methylation and imprinted genes. However, at embryonic (E) day 10.5, PGCs arrive at the genital ridge and begin to mature; from E12.5 onwards, PGCs undergo sex-specific development in gonads. Extensive chromatin remolding and a bulk of DNA demethylation occur during E11.5 to E12.5, including on both normal genes (Hajkova et al., 2002) and imprinted genes (Lee et al., 2002). It is still unclear how the migrating PGCs maintain their methylation pattern, given that an essential component of DNA methylation maintenance machinery, Uhrf1, is specifically down-regulated (Kurimoto et al., 2008). Nevertheless, numerous evidences support the idea that this global DNA demethylation is an active process (Wu and Zhang, 2010). One important supporting point is that the maintenance DNA methyltransferase Dnmt1 is still expressed and localizes in the nucleus at this stage (Sakai et al., 2001). Another cooperative model is also raised where both active and passive demethylation exist as parallel systems in PGCs reprogramming process (Hackett et al., 2012).

There are many questions in the early developmental reprogramming that remain elusive. For example: how maternal genome escapes from active demethylation in zygotic reprogramming? What is the function of the large-scale remodeling of chromatin? And most importantly, what is the mechanism and catalytic enzyme for removal of methyl moiety from 5mC? Some hypotheses have been raised in recent years and will be discussed further in following chapters.

1.2.1.2 Experimental Reprogramming

In 1981, the first embryonic stem cells (ES cells) were isolated from the inner cell mass in blastocyst by Martin (Martin, 1981) and Kaufman (Evans and Kaufman, 1981). These cells can be grown in cell culture condition while still keeping pluripotency, meaning that they can form chimera when transferred into mouse blastocyst and differentiate into all tissues including the germ line (Kuehn et al., 1987). Therefore, ES cells serve as an excellent *in vitro* model for studying molecular and cellular biology at implantation stage.

Experimentally, reprogramming has been achieved by somatic cell nuclear transfer in early decades. In 2006, Yamanaka's group described that a combination of transcription factors (Oct4, Sox2, Klf4 and c-Myc) can dedifferentiate somatic cells back into pluripotent ES-like cells, which then have been named induced pluripotent cells (iPS) (Takahashi and Yamanaka, 2006). It was the first investigation of iPS and this is a landmark report in cell biology. Human iPS cells provide a possibility to generate specialized cell types for individual patients, so it brings a new era for cell therapy and is of great medical interest. Thence Shinya Yamanaka, sharing with John B. Gurdon, was awarded the Nobel Prize in physiology or medicine 2012, for phrasing their finding that mature cells can be reprogrammed to become immature cells capable of developing into all tissues (Rossant and Mummery, 2012). In recent years, many further studies followed after this initial study from Yamanaka for a better understanding of this process and to improve the reprogramming event (Takahashi et al., 2007; Yu et al., 2009; Chang et al., 2009).

Some proteins have been frequently applied in different combinations of ectopic expression factor when generating iPS, e.g. Klf4, Sox2, Oct4, Nanog and c-Myc. Oct4 is a bipartite homeodomain transcription factor that belongs to the Pit-Oct-Unc (POU) family, therefore also named as POU domain, class 5, transcription factor (Pou5f1) (Scholer et al., 1990). Oct4 has long been recognized to be a master factor for inner cell mass formation (Nichols et al., 1998) and maintenance of undifferentiation state in mouse and human ES cells (Niwa et al., 2000; Brandenberger et al., 2004). Researches elegantly showed that the precise relevant level of Oct4 abundance produces different phenotypic effects in ES cells, meaning that obvious decrease in Oct4 expression triggers cell differentiation toward the trophectoderm lineage, proper Oct4 dosage maintains the ES cells pluripotency, and sustained up-regulation in Oct4 expression level promotes cells into mesoderm or endoderm (Niwa et al., 2000; Shimozaki et al., 2003). The expression and silencing of Oct4 is precisely programmed by the epigenetic system and the cell differentiation status. For example, H3K9 methylation and heterochromatinization are involved in Oct4 inactivation, and this progress is mediated by G9a protein, a dominant H3K9 methyltransferase for euchromatic fraction of the genome (Feldman et al., 2006; Tachibana et al., 2002). Nanog, a homeodomain-containing transcription regulator, is another protein undoubtedly linked to mammalian pluripotency (Chambers et al., 2003; Mitsui et al., 2003). Nanog is named after Tir Na Nog, the mythological Celtic land of the ever young, because Nanog confers ability for cell self-renewal and pluripotency in mouse ES cells independently from LIF/Stat3 signaling pathway. Nanog-

deficient mouse ES cells differentiate slowly into extra-embryonic endoderm lineages (Mitsui et al., 2003). The other involved ectopic factor c-Myc is a famous oncogene, as well as a regulator of cell cycle and metabolism (Kim et al., 2010). During generation of iPS cells, c-Myc functions mainly as enhancer of reprogramming efficiency rather than directly inducing pluripotency (Wernig et al., 2008).

These aforesaid transcription factors have been reported to co-occupy and regulate promoter regions of many ES cell-specific genes, and there is also complicated interaction network between these factors. For example, Nanog could be a strong activator of Oct4 expression (Pan et al., 2006). Human and mouse Sox2 often acts as a partner of Oct4 to regulate gene expression (Kuroda et al., 2005; Nakatake et al., 2006). These transcription factor synergy systems govern the self-renewal and undifferentiation of ES cells (Pei, 2009).

Chemical components were also applied to assist cell reprogramming. For example, HDAC inhibitor valproic acid (VPA) improves reprogramming efficiency of fibroblast cells with only Oct4 and Sox2, without Klf4 or c-Myc (Huangfu et al., 2008). With only Oct3/4 and Klf4, G9a inhibitor BIX-01294 enables reprogramming neural progenitor cells, and stem cell pluripotency, global expression profile and epigenetic status are all well re-established in this process (Shi et al., 2008).

1.2.2 Active DNA demethylation

1.2.2.1 Active DNA demethylation is a landmark event of reprogramming

Epigenetics is the key system for reacquiring pluripotency, because the specific gene expression profile in a particular cell type is determined by the epigenetic status. Therefore insight into DNA demethylation is crucial for understanding reprogramming and thus suggests direction for regenerative medicine.

As mentioned above, active demethylation occurs during developmental stages. The global removal of 5mC in the paternal genome during zygotic reprogramming is observed in human, mouse, rat, pig and bovine (Dean et al., 2001; Fulka et al., 2004), but not in sheep (Beaujean et al., 2004). More specifically, some genomic elements escape from this large-scale demethylation, including imprinted genes (Olek and Walter, 1997) and transposable elements (Lane et al., 2003). Experimental dedifferentiated cells show similar DNA methylation level with ES cells, indicating genomic demethylation occurs during reprogramming (Dean et al., 2003; Maherali et al., 2007). PGCs undergo another large-scale wave of active erasure of methyl marks in most genes (Hajkova et al., 2002), which routine to re-establish totipotent epigenetic patterns in the next generation of mammals.

In addition, loci-specific demethylation in somatic cells was also detected (Wu and Zhang, 2010). It was reported that in the presence of VPA, a human estrogen responsive gene trefoil factor 1 (*TFF1*, also named *pS2*) becomes quiescent and its proximal promoter is methylated (Reid et al., 2005). In response to the stimuli of estrogens, the promoter of *pS2/TFF1* gene undergoes local cyclical methylation/demethylation during transcriptional cycle (Metivier et al., 2008; Kangaspeska et al., 2008). Another example of loci-specific demethylation is found in T-cells, where the proximal enhancer of interleukin 2 gene (*il2*) gets rapidly and specifically demethylated after activation signal, which is correlated with enhancement of the *il2* gene expression. This dynamic demethylation is a pathway to control rapid immune response for T-cell (Bruniquel and Schwartz, 2003; Kersh et al., 2006).

1.2.2.2 Hypotheses of active DNA demethylation mechanism

Considering the importance of DNA demethylation in embryogenesis and somatic development, the mechanism is of great interest for researchers. In the last decades, many different hypotheses and catalytic candidates have been raised (Wu and Zhang, 2010; Ooi and Bestor, 2008; Gjerset and Martin, 1982). Among them, some hypotheses describe a direct removal of methyl moiety from cytidine ring (by one or many steps), and some others depict a complete replacement of cytidine (or nucleoside or nucleotide) (Morgan et al., 2005; Okada et al., 2012; Lyer et al., 2009; Lyer et al., 2009; Figure 5).

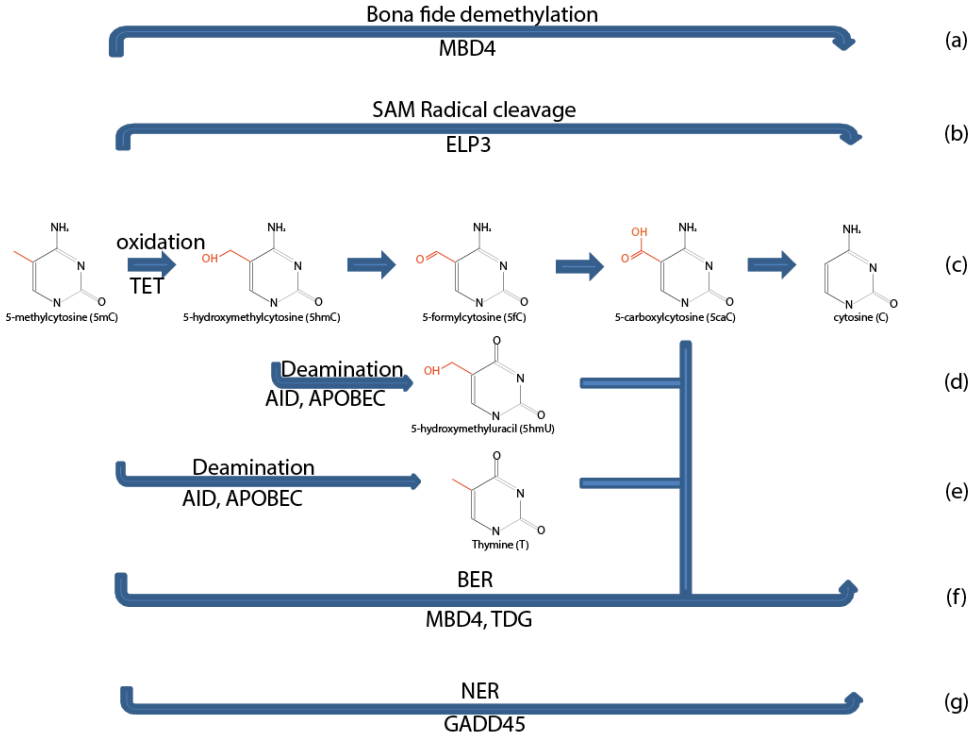


Figure 5: Summary for hypotheses of active DNA demethylation. Direct removal of the methyl group of 5-methylcytosine (5mC) involves breaking a carbon-carbon bond, which requires an enzyme with high catalytic energy. (a) Mbd4 and (b) Elp3 protein were reported to enzymatically remove the methyl moiety from 5mC. (c) Tet proteins were reported to oxidize the 5mC to 5hmC, followed by further oxidation to 5fC and 5caC, and the deformylation of 5fC or decarboxylation of 5caC is proposed to be required for completed demethylation. (e) The 5mC or (d) its oxidative intermediate 5hmC are also reported to undergo deamination, which are catalyzed by deaminase Aid or Apobec proteins, and produce T or 5hmU, respectively. The deamination products could be further processed by BER pathway. (f) BER mechanism is an important pathway that involved in DNA damage repair. After recognizing incorrect nucleotide, a key factor belonging to glycosylase family can catalyze hydrolysis reaction of the N-glycosidic bond. Tdg is found to execute the nucleotide excision step for a broad range of substrates, including thymine, uracil, 5hmU, 5fC and 5caC. (g) NER is also a mechanism of DNA repair, which is involved in large fragment removal of DNA strands. Gadd45 is reported to be a factor in NER pathway, and recently found to be recruited at active demethylation sites. Therefore it is also possible that nucleotide excision occurs during active DNA demethylation.

Direct enzymatic removal of methyl group

Demethylation at C5 of cytosine is not a trivial reaction, because cytosine is an electron-poor heterocyclic aromatic ring system and therefore high energy is required to break the strong carbon-carbon bond of 5mC. Therefore, a potential enzyme carrying out a direct methyl-removal process must overcome a high energy barrier. Methyl-CpG-binding domain protein 2 (Mbd2) has been claimed to directly erase a methyl group from DNA, releasing methanol as product (Bhattacharya et al., 1999). But this research did not substantiate; people found that paternal pronucleus fused with an Mbd2-lacking mouse oocyte still present normal methylation pattern (Santos et al., 2002). Additionally, although Mbd2 shows necessity for transcriptional repression, *mbd2*$^{-/-}$ mice show normal physiological phenotypes, such as viability, offspring production and normal DNA methylation patterns in different tissues (Hendrich et al., 2001), and these researches raise the doubt to the role of Mbd2 in DNA demethylation.

The group from Yi Zhang found that elongator acetyltransferase complex subunit 3 (Elp3), plays a role in zygotic demethylation (Okada et al., 2012). In addition to the previously assumed candidates, Mbd2 and Gadd45, many potential factors were selected according to their catalytic motifs in that study. Effects of those factors were examined by knocking-down in pronuclear stages and screened by a live cell imaging system in mouse zygotes. It was found that Elp3 knock-down impairs demethylation in the paternal genome. The functional entity of elongation complex is shown to be consisted with 6 subunits (Elp1-Elp6), termed as DNA polymerase II (Pol II) holo-elongator complex (Krogan and Greenblatt, 2001). Deficiency of other two components in elongation complex, Elp1 and Elp4, also results in similar DNA methylation abnormalities (Okada et al., 2012). Elp3 belongs to a radical S-adenosyl-L-methionine (SAM) family, which class contains a wide range of enzymes that are involved in DNA repair and biosynthesis of vitamins and coenzymes. They share a common CxxxCxxC motif, which can form a [4Fe-4S] $^{1+}$ iron-sulfur cluster. An iron-sulfur center serves as a strong reducing agent, and could generate a powerful oxidizing 5'-deoxyadenosyl radical together with the cofactor SAM. With a series of radical products, this mechanism can catalyze the cleavage of inactive hydrocarbon bond in alkyl group (Wang and Frey, 2007). The energetic similarity between C-H and C-C bonds makes this mechanism attractive to study active DNA demethylation. In addition to the SAM domain, Elp3 also contains a C-terminal histone acetylation (HAT) motif, which might indicate the linkage between histone and DNA modifications (Winkler et al., 2002).

Up until now, the best characterized pathway of DNA demethylation is an oxidative mechanism catalyzed by ten-eleven translocation (Tet) enzymes. Tet proteins belong to a 2-oxoglutarate (2OG) and iron (II) -dependent dioxygenase superfamily (Aravind and Koonin, 2001). This superfamily is widespread from prokaryotic to eukaryotic organisms, and modifies a range of substrates. Alpha-ketoglutarate-dependent dioxygenase (AlkB), kinetoplastid base J binding protein (Jbp) and Tet family perform hydroxylation of nucleic acid bases (Lyer et al., 2009). This model describes a stepwise oxidation of 5mC, first into 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009; Ito et al., 2010), which is further oxidized into 5-formylcytosine (5fC) and later into 5-carboxylcytosine (5caC) (Ito et al., 2011; Nable and Kohli, 2011; He et al., 2011). Thereupon, one possibility is that the deformylation of 5fC or decarboxylation of 5caC achieves the final erasure of cytosine modifications, and another possibility is that the oxidative intermediates associate with other pathways to achieve complete demethylation, for instance with base excision repair (BER) pathway. To support the importance of Tet proteins and 5hmC in zygotic demethylation, two reports have shown that the rapid loss of 5mC in mouse paternal haploid is accompanied by an accumulation of 5hmC (Iqbal et al. 2011; Wossidlo et al., 2011). Further details about Tet family are described in chapter 1.3.

Indirect demethylation

In plant, many genetic and biochemical studies indicate that active demethylation can be achieved by base excision repair (BER) pathways (Zhu, 2009; Gehring et al., 2006; Penterman et al., 2007). BER is one of the most important pathways involved in DNA

repair from mutagenic or cytotoxic damage. The BER pathway contains the following steps: recognition of an incorrect nucleotide, followed by an excision of the inappropriate base to leave a single nucleotide gap, then insertion of a new nucleotide (Sancar et al., 2004; Wilson and Bohr, 2007). The key factors that execute the nucleotide excision step belong to a glycosylase family, which catalyze hydrolysis reaction of the N-glycosidic bond (Dizdaroglu, 2005; Stivers and Jiang, 2003) and generate an abasic (AP) site (Sancar et al., 2004).

Arabidopsis is a very good model to display this active demethylation process (Zhu, 2009; Gehring et al., 2006; Penterman et al., 2007). Repressor of silencing 1 (ROS1, also known as DML1), Demeter (DME), DME-like 2 (DML2) and DML3 belong to Demeter (DME) DNA glycosylase family. This family was first identified when DME mutation was found to cause an imprinted gene *MEDEA* silenced (Choi et al., 2002). Researches demonstrated that DME members can target methylated DNA substrates, process glycosylation function, and are essential for gametogenesis and parental methylation pattern (Schoft et al., 2011; Gehring et al., 2006).

The model in plants raises the question whether similar mechanism exists also in mammalian cells. One possible candidate might be the thymine DNA glycosylase (Tdg), which belongs to the uracil DNA glycosylase family. In addition to biological functions in DNA repair and gene expression regulation (Cortaza et al., 2007), Tdg presents putative involvement in DNA demethylation (Cortaza et al., 2011; Metivier et al., 2008). Tdg knock-out leads to embryonic lethality and DNA hypermethylation in mice (Cortellino et al., 2011). Tdg is proposed to be involved in demethylation by either a direct glycosylation excision of 5mC (Zhu et al., 2000b) or by combinative pathways with other enzymes, which will be discussed later. The methyl-CpG binding domain 4 (Mbd4) protein is another candidate enzyme for glycosylating 5mC during DNA demethylation (Zhu et al., 2000a). Mbd4 is a nuclear protein and localizes at heterochromatin (Hendrich and Bird, 1998), consisted of an N-terminal MBD domain and a C-terminal glycosylase domain (Hendrich et al., 1999).

Nucleotide excision repair (NER) is another mechanism for DNA repair. UV light, chemotherapy drugs and toxic could cause large lesions on DNA, and NER pathway repairs such damages (Nouspikel, 2009; Dinant et al., 2012). NER includes following steps: lesion recognizing, opening of a denaturation bubble, incision of the damaged strand, displacement of the lesion-containing oligonucleotide, and gap sealing (Nouspikel, 2009). The deficiency of NER results in many inherited diseases: such as xeroderma pigmentosum, cockayne syndrome, trichothiodystrophy and UV-sensitive syndrome (Nouspikel, 2009; Mu et al., 1995).

The growth arrest and DNA-damage-inducible 45 alpha (Gadd45a) protein was firstly reported to be involved in NER pathway (Smith et al., 1996). In 2007, Gadd45a was found to be recruited to methylated reporter and activate silenced promoters by loci-specific and global demethylation (Barreto et al., 2007). Using zebrafish embryos as model, Gadd45 is shown to be involved in widespread demethylation, in cooperation with the activation-induced cytidine deaminase (Aid, also named Aicda) and a glycosylase

Mbd4 (Rai et al., 2008). It was found that active demethylation takes place at lineage-specific genes when adult stem cells differentiate into terminal cells. Gadd45a up-regulates this process, and knock-down of Gadd45a causes hypomethylation of those genes, leading to suppression of differentiation (Zhang et al., 2011). Mouse Gadd45a expression strongly increases at E7.5-E8.5 during embryonic stage (Kaufmann et al., 2011), which is consistent with the occurrence of the DNA demethylation in early development. Anyhow, conflicting evidence also exists; a normal DNA methylation pattern exhibits in *gadd45a*$^{-/-}$ cells (Engel et al., 2009).

Cooperated pathways for DNA demethylation

Some studies support that combinations of aforementioned enzymes may provide answer for how active DNA demethylation happens. One speculation is that a deaminase first converts 5mC into an intermediate nucleotide, such as thymine, followed by mismatch excision by BER pathway (Wu and Zhang, 2010; Figure 5). Deaminases, such as Aid or apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 (Apobec1), were reported to be involved in the demethylation (Metivier et al., 2008; Rai et al., 2008; Morgan et al., 2004).

Aid/Apobec protein family is well-known for participation in antigen-driven antibody diversification and maturation (Conticello, 2008). Aid and Apobec share structural and functional similarities, and both belong to a widespread superfamily of zinc-dependent deaminase (Conticello et al., 2007). Apobec was first described to catalyze deamination of cytidine to uridine in apolipoprotein B pre-mRNA (Scott, 1995), and Aid was reported to convert C to U in DNA of immunoglobulin loci (Neuberger et al., 2003). Later, Morgan et al. demonstrated that Aid and Apobec1 convert 5-methylcytosine to thymine, which leads to a T:G mismatch in DNA (Morgan et al., 2004). Aid and Apobec1 are detected to be expressed in oocytes and PGCs (Morgan et al., 2004), while Aid-deficient PGCs exhibit hypermethylation throughout genome (Popp et al., 2010). Knock-down and rescue experiments demonstrated that Aid is required for the onset of cellular reprogramming in iPS cell generation (Bhutani et al., 2013). These evidences suggest the involvement of deamination in DNA demethylation process.

A study in zebrafish embryos showed that Aid, together with a glycosylase Mbd4 and a promoting Gadd45, participate in widespread demethylation (Rai et al., 2008). *In vitro* studies showed that among the derivatives of 5mC, Mbd4 binds preferentially to deamination products rather than other types (Hendrich et al., 1999; Morera et al., 2012). In addition to an N-terminal region that recognizes 5mC, Mbd4 contains a C-terminal catalytic domain that shares sequence identity with many known DNA glycosylases. The catalytic domain of this enzyme excises thymine or uracil at T:G or U:G mismatch, regardless of the methylation status (Sjolund et al., 2012; Hendrich et al., 1999). Crystallization of the catalytic domain of human MBD4 together with mismatched DNA substrates shows the MBD4 can flip out and capture the target thymine or 5hmC base into binding pocket (Morera et al., 2012; Hashimoto et al., 2012b).

Participation of Tdg as glycosylase in demethylation process is also widely accepted. Biochemical methods indicated that Tdg interacts with Aid and Gadd45a (Cortellino et al., 2011). It is proposed that deamination of 5mC, thymine, is the substrate for Tdg (Cortaza et al., 2007). Besides thymine and uracil, systematic studies showed that Tdg has glycosylation activity on a broad range of oxidation-linked substrates, including 5-hydroxymethyluracil (5hmU), 5fC and 5caC, but not 5mC or 5hmC (Cortellino et al., 2011; He et al., 2011; Maiti and Drohat, 2011). Tdg-deficiency in mouse ES cells causes accumulation of 5fC and 5caC, and further mapping with modification-specific antibodies demonstrated that the 5fC and 5caC accumulate mostly in proximal and distal gene regulatory regions (He et al., 2011; Shen et al., 2013). Guo et al., proposed a stepwise processing of 5mC, including oxidation to 5hmC, deamination to 5hmU by Aid/Apobec in adult mouse brain (Guo et al., 2011; Figure 5). Although the concentration of 5hmU is relatively low in tissues (Globisch et al., 2010), co-crystallization of the catalytic domain of human TDG with G:T or G:5hmU mismatch DNA substrates demonstrated the binding between DNA and TDG, and also exhibits a flipping-out of the target nucleotide from the double-strand DNA, which supports the hypothesis (Hashimoto et al., 2012a).

1.3 Dioxygenase Tet family

As a well-known candidate for catalyzing active DNA demethylation, Tet proteins are found in several tissues of mouse and human, particularly at the developmental stages of zygote, blastocyst inner cell mass and PGCs formation, in which stages global demethylation and *de novo* methylation take place, indicating their functions in early embryonic development. The biological functions, expression profile and catalytic activity of Tet family bring great interest of research.

1.3.1 2-Oxoglutarate (2OG) and iron (II)-dependent dioxygenase superfamily

As described in chapter 1.2.2, Tet proteins belong to a 2-oxoglutarate (2OG) and iron (II) -dependent dioxygenase superfamily, which contains many enzymes with diverse biological functions. Two other members of this class, AlkB and Jbp, also have been reported to catalyze *in situ* hydroxylation of bases in nucleic acids (Lyer et al., 2009).

The *E. coli* enzyme AlkB catalyzes the oxidative demethylation of 1-methyladenine and N^3-methylcytosine (3meC) in the bacterial response to DNA alkylated base lesion, together with 2OG and Fe (II) as cofactors. Decarboxylation converts 2OG to succinate and carbon dioxide (Figure 6) (Falnes et al., 2002; Trewick et al., 2002; Mishina and He, 2006). The crystal structure of AlkB has been solved, which might be very helpful for understanding its homologues, such as the Tet proteins who also share the conserved 2OG-Fe^{2+} binding domain (Sedgwick et al., 2007; Yang et al., 2008). The 90 amino acids in the N-terminus of AlkB form a special structure, which consists of a β-strand and α-helix, and covers the surface of dioxygenase part. This N-terminal part is inferred to be the “nucleotide-recognition lid” of AlkB (Yu et al., 2006; Figure 6). A substrate-bound AlkB-DNA complex reveals conformational changes of amino acid residues within the active site, which is recognized to be important for binding damaged bases (Holland and Hollis, 2010).

The Jbp family is defined according to its involvement in synthesis of Beta-d-glucopyranosyloxymethyluracil (base J), a DNA modification that exists in kinetoplastid. The generation of base J begins with a hydroxylation of thymidine, followed by glycosylation (Borst and Sabatini, 2008). There are two proteins, Jbp1 and Jbp2, which catalyze the thymidine hydroxylation. Jbp1 and 2 have different C-terminus for specific protein interactions, but share a conserved N-terminal 2OG-Fe (II) dioxygenase domain (Cliffe et al., 2009; Yu et al., 2007).

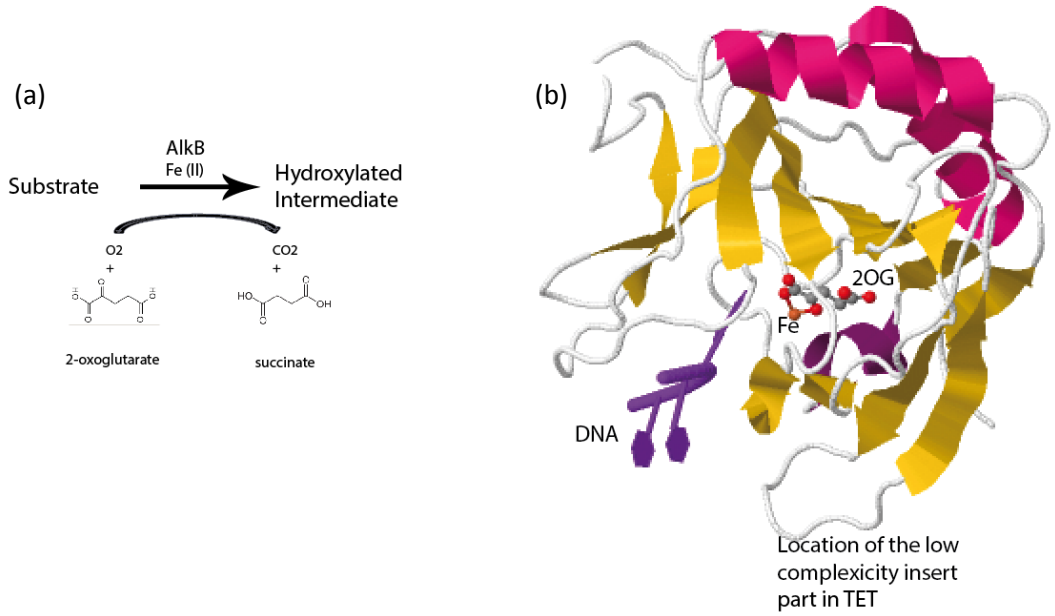


Figure 6: (a) Diagram of AlkB-catalyzing oxidative demethylation. AlkB belongs to the 2-oxoglutarate (2OG) and iron (II) -dependent dioxygenase superfamily. One atom of oxygen is transferred from 2OG into hydroxylation intermediate, releasing CO_2 and succinate. (b) Cartoon representation of AlkB protein (PDB: 2FD8; Yu et al., 2006). AlkB was shown under the presence of DNA substrate, Fe (II) ion and its cofactor 2-oxoglutarate (2OG). In the 2OG structure, carbon atoms are shown red and the rest of the structure carbons are shown in grey. Fe (II) is shown in orange. The α-helix of AlkB is present in pink and β-sheet in yellow. The figure was generated using the PyMol software (DeLano, 2002).

The Tet family contains three members: Tet1, 2 and 3, which are paralogues with Jbp proteins (Tahiliani et al., 2009). Tet family displays the typical double-strand β helix (DSBH) fold of 2OG-Fe (II) dioxygenase, including the conserved motif to chelate iron and 2OG (Lyer et al., 2009). Tet1 was the first described member of Tet family and initially discovered as a fusion partner with H3K4 methyltransferase MLL (mixed lineage leukemia protein), and therefore Tet family is named after the ten-eleven translocation (t (10; 11) (q22; q23)) (Lorsback et al., 2003; Ono et al., 2002) in acute myeloid leukemia (AML). Recently, it was shown that Tet proteins hydroxylate 5mC into 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009; Ito et al., 2010), and further oxidize 5hmC stepwise into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011; Nable and Kohli, 2011; He et al., 2011). One hypothesis proposes that the deformylation of 5fC or decarboxylation of 5caC achieves the active demethylation step (Figure 5).

1.3.2 Tet proteins and 5hmC

1.3.2.1 Cellular functions of Tet1

Tet1 was discovered in human MLL1-TET1 fusion protein from acute myeloid leukemia (AML) cells and was initially encoded as leukemia-associated protein with a CXXC domain (LCX) (Lorsback et al., 2003; Ono et al., 2002). MLL1, which is located on 11q23 in the human chromosome, belongs to a H3K4 methyltransferase family. *MLL1* is frequently found to be a target of recurrent specific chromosomal translocation in human hematological malignancies. These cytogenetic abnormalities implicate many different genes and over 30 different fusion partners have been described, such as RNA polymerase II elongation factor, transcriptional co-activator and histone acetyltransferase (Daser and Rabbitts, 2005). Those fusion proteins, including MLL-TET1, consistently retain the amino-terminal fragment of MLL1 and carboxyl-terminal domains from the fusion partners (Ayton and Cleary, 2001; Ayton et al., 2004).

Tet1 is of great importance in different developmental stages, particularly in blastocyst and PGCs formation. Tet1 is highly expressed in pluripotent mouse ES cells, in which an elevated level of 5hmC is also observed. Tet1 is drastically down-regulated when ES cells differentiate, indicating that Tet1 has a specific function for ES cells maintenance or for lineage specification at this stage (Szwagierczak et al., 2010). ES cells with either Tet1 knock-out or knock-down show reduction of 5hmC level and formation of large haemorrhagic teratomas, which is likely due to excessive number of trophoblast-derived cells. Therefore, deficiency of Tet1 leads to skew of extraembryonic lineages (Dawlaty et al., 2011; Koh et al., 2011). Tet1 depletion hyperactivates some key trophoblast regulators, like E74-like factor 5 (Elf5) and caudal-type homeobox 2 (Cdx2) (Ito et al., 2010; Koh et al., 2011; Ficz et al., 2011). At the same time, Tet1-depletion negatively regulates neurogenic differentiation factor 1 (Neurod1) and neural development gene paired box 6 (Pax6) (Dawlaty et al., 2011; Xu et al., 2011; Williams et al., 2011). These evidences suggest essential regulatory functions of Tet1 in the first lineage commitment stages.

In ES cells, a large amount of developmentally regulated factors and lineage-specific genes are under control of a “bivalent marker”, which describes a co-occupation of the marker for transcription active genes, H3K4me3, and the repressive marker H3K27me3. This dual-marker system keeps genes silenced, but is proposed to prepare genes for future initial transcription. This poised state is likely to be necessary for the development potential of ES cells (Pan et al., 2007; Bernstein et al., 2006). Comparing the list of Tet1 binding protein with identified genes in mouse ES cells, the enrichment of Tet1 shows a positive correlation with genes that either are transcriptional active or contain bivalent chromatin signature (Wu et al., 2011b; Wu and Zhang, 2011). 5hmC also shows a similarity of genomic occupation with Tet1 in ES cells. High-throughput sequencing revealed an enrichment of 5hmC within exons and near transcription start sites (TSSs), especially at the start sites of genes whose promoters bear the bivalent marks (Pastor et al., 2011), suggesting that Tet1 and 5hmC may play a key role in orchestrating the balance between pluripotent and lineage committed states. However, conflicting evidence

exists representing that Tet1 binds throughout the genome of ES cells, with the majority of binding sites located at TSSs of CpG-rich promoters and within genes (Williams et al. 2011).

Although Tet1 is important for lineage regulation in ES cells and inner cell mass, *tet1*$^{-/-}$ and *tet1*$^{-/-}$ *tet2*$^{-/-}$ mice form all three germ layers and are viable (Dawlaty et al., 2011; Dawlaty et al., 2013), indicating that Tet1 and Tet2 may have important but not crucial functions during mouse development.

As introduced in a previous chapter (chapter 1.2.2), PGCs undergo a rapid drop in DNA methylation levels. Simultaneously with the wave of demethylation, generation of 5hmC and transcriptional activity of Tet1/2 was detected (Vincent et al., 2013; Hackett et al., 2013). At embryonic day E9.5 to E10.5, a significant conversion of 5mC to 5hmC is observed *en masse* and at individual loci, consistent with the fact that numerous promoters and gene bodies are hypermethylated in Tet1 knock-down PGCs (Hackett et al., 2013). Investigation based on immunofluorescence measurements showed that the levels of 5fC and 5caC remained relatively stable during PGCs reprogramming, indicating that the oxidative excision might not occur during this stage (Yamaguchi et al., 2013).

Compared to normal germ cells, Tet1-deficient germ cells show developmental abnormality and dysregulation of a large set of genes. During E16.5-E18.5, nearly half of the *tet1*$^{-/-}$ female gamete samples suffer a defect in meiotic synapsis formation and show developmental arrest. Many meiosis-related genes seem to be activated by Tet1, including malate dehydrogenase (Mae1) and synaptonemal complex protein 1 (Sycp1) (Yamaguchi et al., 2012).

Conclusively, Tet1 plays an important role in determining lineage commitment at inner cell mass stage. It participates in the conversion of 5mC to 5hmC at the demethylation wave in PGCs, positively regulating many meiosis-related genes in germ cells. The functional perturbation of Tet1 significantly reduces female germ-cell numbers and fertility.

1.3.2.2 Cellular functions of Tet2

Compared to Tet1, Tet2 has been found to be highly expressed in many somatic tissues, as well as in ES cells and PGCs (Szwagierczak et al., 2010; Vincent et al., 2013). Apart from the collaborative functions together with Tet1 in embryogenesis and PGC formation, Tet2 is also linked to myeloid leukemia.

Deletions and mutations of TET2 were found in a wide range of human myeloid malignancies, including acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), myeloproliferative neoplasms (MPN) and secondary AML (sAML) (Langemeijer et al., 2009; Delhommeau et al., 2009; Konstandin et al., 2011). Tet2 mRNA is found in high levels in hematopoietic multipotent progenitors from bone marrow, maintains high levels in myeloid progenitors, and gets low in mature granulocytes (Ko et al., 2010), which suggests the tightly relationship between Tet2 and hematopoietic lineage.

1.3.2.3 Cellular functions of Tet3

Tet3 is proposed to be the answer for the mysterious rapid DNA demethylation wave in zygotic pronuclear stage. Recent researches proved that a massive DNA methylation oxidation exists in zygote, along with 5hmC accumulation in paternal pronucleus and reduction of 5mC (Wossidlo et al., 2011). Moreover, Tet3 is found to be enriched specifically in the male DNA. In Tet3 knock-out zygotes, conversion of 5mC into 5hmC fails to occur in paternal DNA, and the level of 5mC remains. Tet3-deficient oocytes also get reduced in their capability of reprogramming the injected somatic nuclei, and the Tet3-depleted germ cells show severely developmental failure at around E11.5 for unknown reason (Wossidlo et al., 2011; Gu et al., 2011).

In comparison to other tissues, the nervous system shows highest 5hmC level and prominent transcriptional activity of Tet3 (Szwagierczak et al., 2010). Xenopus laevis Tet3 participates in early eye and neural development by regulating some developmental related genes (Xu et al., 2012). Loss of Tet2 and Tet3 leads to a defect in neuronal differentiation from neural progenitor cells (Hahn et al., 2013), which indicates the functional role of Tet proteins in neural development.

1.3.2.4 5-Hydroxymethylcytosine (5hmC) - the sixth base in the genome

5-Hydroxymethylcytosine (5hmC) was discovered as a new composition of mammalian DNA (Penn et al., 1972). The discovery is of tremendous importance because it was the only other modified base in genome of higher organisms besides 5mC for a long time. Therefore, it is also called the sixth base in mammalian genome, namely after the four nucleotides and 5mC.

Evidence is increasing that DNA hydroxymethylation is strongly associated with actively transcribed genes, but the mechanism of this regulation function is still unknown (Ito et al., 2010; Ficz et al., 2011). The distribution of 5hmC varies in different tissues. Various methods were established to quantify the level of 5hmC, and 5hmC is found to be enriched in neuron system and ES cells (Szwagierczak et al., 2010; Kriaucionis and Heintz, 2009; Globisch et al., 2010). 5hmC constitutes around 0.6% of guanine in genome of Purkinje neurons and 0.2% in genome of granule cells (Kriaucionis and Heintz, 2009). In ES cells, 5hmC takes around 4% of all cytosine modification species (Tahiliani et al., 2009) and the 5hmC abundance in genome decreases very rapidly during differentiation of ES cells (Szwagierczak et al., 2010). Taken together, 5hmC is of a detectable portion in genome and is physiologically regulated by cell developmental stage. Mass spectrometry analysis also revealed that human induced pluripotent stem (iPS) cells show a high increase of 5hmC levels when compared to parental fibroblast cells, which indicates that the 5hmC is an important epigenetic change during cell reprogramming (Le et al., 2011). It was observed that 5hmC is enriched in gene bodies of active genes in mouse cerebellum and ES cells, in contrast to 5mC that is mainly enriched at CpG-rich transcription start sites (TSSs) (Song et al. 2011a; Williams et al. 2011). 5hmC can be further oxidized to 5fC and 5caC. 5fC in genome is quantified by

biochemical methods (Pfaffeneder et al., 2011; Song et al., 2013); around 0.02% (against guanine) of cytosine is 5fC in wide-type ES cells, which is a significant existence of such derivative.

1.3.3 Post-translational modification of Tet proteins

Post-translational modification (PTM) is an important step of protein biosynthesis. During translation step, amino acids are incorporated into polypeptide chains by ribosome. PTMs describe the chemical modifications of proteins after their translation, which largely increases the complexity and diversity of the proteome. Research showed that the PTMs are widely present in proteome (Khoury et al., 2011). Statistics of relative abundance for different post-translational modifications have been experimentally and putatively detected using proteome-wide information analysis. The importance of PTM on protein function and stability raises the question: whether Tet proteins are also under regulation of PTM systems.

PTMs widely extend the range of protein functions. For example, phosphorylation, a very widespread PTM, which describes the addition of a phosphate group to a protein, plays an important role in almost every aspect of eukaryotic cell regulation. Usually, the reversible phosphorylation reaction is involved in a transfer of a phosphoryl group from a high-energy organic compound, such as adenosine triphosphate or guanosine triphosphate, to the side chain of serine, threonine or tyrosine amino acid residues. Protein phosphorylation affects every basic cellular process, including metabolism, division, differentiation, motility, growth, signaling transduction and muscle contraction (Manning et al., 2002; Dephoure et al., 2008; Engholm-Keller and Larsen, 2013). Aberrant phosphorylation patterns are found to be implicated in many developmental abnormalities and human diseases, such as breast cancer, diabetes, cardiovascular and neuropsychiatric disorders (Hendriks and Pulido, 2013; Nunes-Xavier et al., 2013; Hendriks et al., 2013). Therefore, for many decades, phosphorylation is of central research interest for academic and therapeutic aspects.

Another abundant and essential post-translational modification event is O-linked N-acetylglucosamine (O-GlcNAc), which occurs both in cytoplasm and nucleus. In the 1980s, a PTM of serine or threonine residues by a monosaccharide, β-D-N-acetylglucosamine, was discovered (Torres and Hart, 1984), and later it became clear that this modification is intracellular (Holt and Hart, 1986). The O-GlcNAc modification is dynamic, with cycling addition and removal of the special monosaccharide β-D-N-acetylglucosamine to target proteins. The cycling of O-GlcNAc is a nutrient-responsive PTM that can influence the activity of a target protein, and thus effects cellular signaling transduction, protein turnover and mRNA transcription (Love et al., 2010; Love and Hanover, 2005).

Until recently, O-GlcNAc cycling was found to be a novel regulator that affects also the epigenome and high-order chromatin structure. A combination of techniques showed that acetylated histones are modified by O-GlcNAc. Over expression of O-GlcNAc transferase modestly increases chromatin condensation (Sakabe et al., 2010). A later research reported that histone H2B could be O-GlcNAcylated at a serine residue S112 *in vitro* and *in vivo*, which subsequently facilitates the mono-ubiquitination of a lysine residue K120 (Fujiki et al., 2011).

Crosstalk between different types of PTMs encodes numerous amount of information. Identification of key protein targets of O-GlcNAcylation demonstrated that O-GlcNAcylation may be as widespread as phosphorylation. Furthermore, there is a possibility of reciprocal regulation of O-GlcNAc and O-phosphate, because these two modifications always modify the same or adjacent sites on proteins (Kamemura and Hart, 2003). The mechanism and function of this extensive interplay between O-GlcNAc and O-phosphate are still elusive, and it is only known that they can either occupy different sites on same peptide, or competitively occupy a single site or proximal sites (Zeidan and Hart, 2010; Hart et al., 2011). Deficient O-GlcNAcylation could lead to neurodegenerative diseases and cancers. Abnormal crosstalk between O-GlcNAcylation and phosphorylation is involved in diabetes (Butkinaree et al., 2010).

Protein glycosylation is one of the most abundant PTMs in eukaryotic cells. Data showed that 1-3% of the human genome is dedicated to encoding glycoside hydrolase or glycosyltransferase enzymes (Davies et al., 2005). To date, two enzymes are known to regulate O-GlcNAc cycling in higher metazoans: O-GlcNAc transferase (Ogt), which catalyzes the addition of O-GlcNAc, and O-GlcNAcase (Oga), a single enzyme that catalyzes the selectively removal of O-GlcNAc from target substrates (Hanover et al., 2012; Butkinaree et al., 2010).

Ogt is a highly evolutionary conserved protein found in most organisms (Lubas et al., 1997; Love and Hanover, 2005). A conditional knock-out experiment of mouse Ogt demonstrated the essential role of Ogt in early development. Deficiency of Ogt leads to embryonic lethality at around E5.5 (Shafi et al., 2000). In mammals, the alternatively splicing produces three isoforms of Ogt, which differ in the length of their N-terminal tetratricopeptide repeats (TPR) and the cellular localization (Lubas et al., 1997; Kreppel and Hart, 1999; Hanover et al., 2012). The C-terminal domain of this protein represents the catalytic portion. The longest isoform, termed nucleocytoplasmic Ogt (ncOgt), contains 12 TPR motifs and is found to localize in both nucleus and cytoplasm. ncOgt protein binds to many transcriptional regulators through the TPR domain, and thus regulates the activity of these transcriptional regulators by O-GlcNAcylation (Hanover et al., 2012). It was found that Ogt binds with a switch-independent 3a (Sin3a), and thus interacts with a histone deacetylase complex, which normally represses gene transcription (Yang et al., 2002). The intermediate isoform, mitochondrial Ogt (mOgt), contains 9 TPRs, and the shortest Ogt isoform (sOgt) contains only 2 TPRs and is found to localize in both nucleus and cytoplasm (Hanover et al., 2012).

Oga, together with Ogt, regulates the O-GlcNAcylation cycling in higher eukaryotic cells (Braidman et al., 1974). Oga is found in both the nucleus and the cytoplasm, and catalyzes the removal of O-GlcNAc from target substrates. The structure of Oga bacterial homologues has been resolved, and thus provides the basis for understanding the mechanism of O-GlcNAc removal (Gloster and Vocadlo, 2010; Rao et al., 2006).

Until recently, Tet proteins were found to be a target of O-GlcNAcylation modification. Ogt was found to interact with Tet1, and the Tet1 expression level is regulated by its interaction with Ogt (Shi et al., 2013). In our research, we provide novel evidence for Ogt

interaction with the triplication Tet1/2/3, and demonstrate that the Tet proteins undergo O-GlcNAcylation modification *in vivo* (chapter 2.3).

1.3.4 CXXC-type zinc finger and Tet proteins

CXXC-type zinc finger domains widely exist in divergent chromatin-binding proteins, including Dnmt1, Mll proteins, Mbd1 and Tet1 (Lee et al., 2001; Ono et al., 2002; Thomson et al., 2010; Lorsbach et al., 2003; Birke et al., 2002). Because the CXXC domain requires zinc ions for efficient DNA binding activity, it is also named zinc finger (Lee et al., 2001). Data showed that CXXC domains of some these proteins specifically mediate binding to DNA templates containing unmethylated CpG sites (Thomson et al., 2010; Birke et al., 2002). However, there are also CXXC motifs being reported to have no DNA binding activity. For instance, Mbd1 protein contains three CXXC zinc fingers, but only one of them shows affinity with DNA substrate (Jorgensen et al., 2004).

CXXC of human DNMT1 is found to bind specifically to unmethylated CpG (Pradhan et al., 2008). Anyway, there is also conflicting data that fragment including AA 613-748 of mouse Dnmt1, which shows DNA affinity with a slight preference for hemi-methylated CpG sites (Fatemi et al., 2001). Deficiency of CXXC displays significant reduction in DNA methyltransferase enzymatic activity, indicating that Dnmt1 CXXC affects DNA methyltransferase activity (Pradhan et al., 2008). However, our independent research demonstrated conflicting results and this will be discussed in later chapters (Frauer et al., 2011, see chapter 2.1 and 3.1). Structural study of a DNA-Dnmt1 complex clearly reveals a contact between DNA and CXXC domain (Song et al., 2011b). It was shown that CXXC domain forms two short helical segments, with eight reserved cysteine residues in two clusters. A loop from one cluster penetrates into the major groove of DNA.

In addition to a CXXC motif in Dnmt1, members of dioxygenase Tet family are also find to be closely accompanied with CXXC domains. In human and mouse, Tet1 contains a CXXC motif in N-terminal part, which is referred to CXXC6 as well. And the *tet2* and *tet3* genes are adjacent to *cxxc4* and *cxxc10-1*, respectively. According to the sequence homology, all these CXXC motifs together identify a distinct subgroup within the whole CXXC domain family (Frauer et al., 2011, see chapter 2.1).

CXXC4, which is also termed as inhibitor of the Dvl and Axin complex (Idax), was reported as an inhibitor of Wnt signaling pathway by directly binding to the PDZ domain of Dishevelled (Dvl), a positive regulator of Wnt signaling (Hino et al., 2001; Michiue et al., 2004). Human CXXC4 gets increasingly expressed in colonic villous adenoma, which means CXXC4 might be involved in tumorigenesis (Nguyen et al., 2010). Further evidence showed that in renal cell carcinoma, expression of CXXC4 decreases and thus activates malignancy through WNT pathway (Kojima et al., 2009). On genomic level, both human and mouse *cxxc4* localizes near *tet2* in the chromosome, and a recent evolutionary study indicates that *cxxc4* was originally encoded within an ancestral *tet2* gene. A study hypothesized that chromosomal gene inversion occurs during evolution, thus separating the Tet2 CXXC domain from the catalytic domain (Ko et al., 2013).

Rat CXXC5 is found to be another modulator of Wnt signaling pathway in neural stem cells and positively regulated by Bmp4 (bone morphogenetic protein 4). CXXC5 shows partial sequence homology and functional similarity with CXXC4. CXXC5 is shown to bind

with Wnt mediator Dvl and antagonize Wnt signaling by competing with Axin (Andersson et al., 2009). Human CXXC5 is identified in normal and tumoral myelopoiesis, and is functionally required in the differentiation progress of normal blood progenitors and promyelocytic leukemia cells (Pendino et al., 2009).

In conclusion, the CXXC-type zinc finger motif is a widely existing domain found in many chromatin modifiers with diverse functions, and is usually found to be associated with DNA binding activity. CXXC motifs are accompanied with Tet family members. In this study, we provide experimental evidence for the claim that Tet3 harbors a CXXC domain, CXXC10-1. Additionally, interaction between Tet3 and CXXC4 were shown *in vivo*. Many properties, including DNA affinity, expression pattern, cellular localization and mobility of these zinc fingers, will also be discussed in this study (Frauer et al., 2011, see chapter 2.1; Liu et al., 2013, see chapter 2.2).

1.4 Aims of this work

The main objective of this work was the study of active DNA demethylation and the CXXC domain, cell cycle dependent localization, protein interaction and post-translational modifications of Tet proteins. The goal was, to get further insight into the complex regulatory mechanisms of Tet proteins and to identify the biological functions of different domains.

The first part of this study focused on elucidating the regulatory function and role of the zinc finger CXXC motifs of Tet proteins. CXXC domain is found in many chromatin-binding proteins. Tet1 protein has an ancestral intramolecular CXXC domain, and a *cxxc4* gene is encoded in the genome as a neighboring gene with *tet2*. This might suggest that the CXXC interacts or plays important role for Tet proteins. To address this hypothesis, tools and techniques to monitor Tet1/2/3 and 5hmC *in vitro* and *in vivo* had to be developed. GFP-trap was applied to study the DNA binding activity and preference of GFP-fused CXXC domains (chapter 2.1). Our research provided the experimental evidence that Tet3 co-transcribes with a CXXC domain (chapter 2.2). Photobleaching technology was used to detect the effect of mobility for CXXC. Isotope-labeling assay for detecting 5hmC abundance, which was established in our group, was applied for finding the influence on the hydroxylation activity of Tet proteins and their mutants.

It is also interesting to investigate how the Tet proteins get regulated during cell cycle (chapter 2.4). Specific monoclonal antibodies were generated as a tool for studying Tet1 and Tet2 in cellular context. Endogenous Tet1 and Tet2 were presented together with cell cycle marker. Furthermore, we performed direct DNA competition binding assay on isolated domains, and thus demonstrated the DNA binding platform of Tet proteins.

Moreover, our group found Tet proteins interact with some post-translational modifier proteins (chapter 2.3). The interaction between Tets and modifiers (Ogt, Oga) was studied by F3H and biochemical assays, including western blot, immunofluorescence and immunoprecipitation. The interaction was mapped by subcloning individual domains of Tet proteins.

2 RESULTS

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

OPEN ACCESS Freely available online

PLoS one

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

Carina Frauer[1☯], Andrea Rottach[1☯], Daniela Meilinger[1], Sebastian Bultmann[1], Karin Fellinger[1¤a], Stefan Hasenöder[1¤b], Mengxi Wang[1], Weihua Qin[1], Johannes Söding[2], Fabio Spada[1]*, Heinrich Leonhardt[1]*

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 ΔCXXC 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

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)

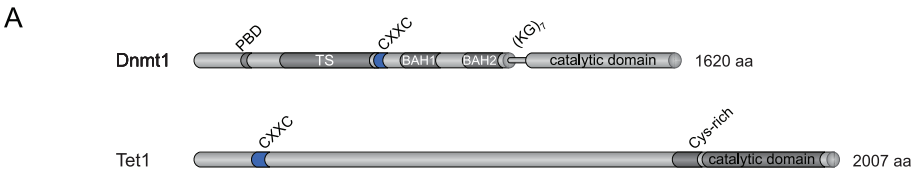
☯ These authors contributed equally to this work.

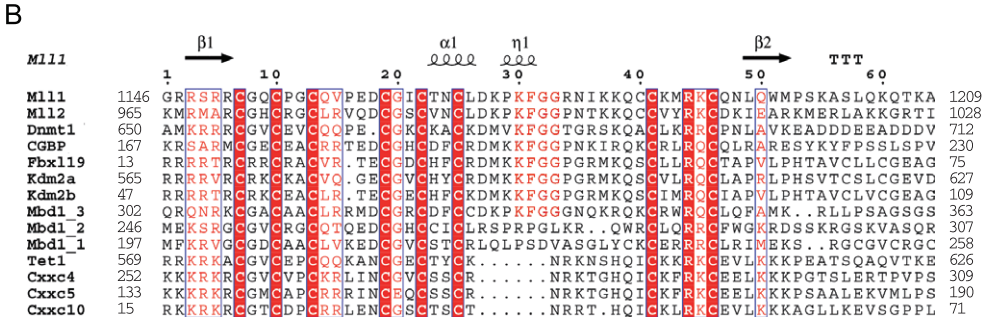
¤a Current address: Intervet International GmbH, Unterschleissheim, Germany
¤b Current address: Institute of Stem Cell Research, Helmholtz Zentrum München, Neuherberg, Germany

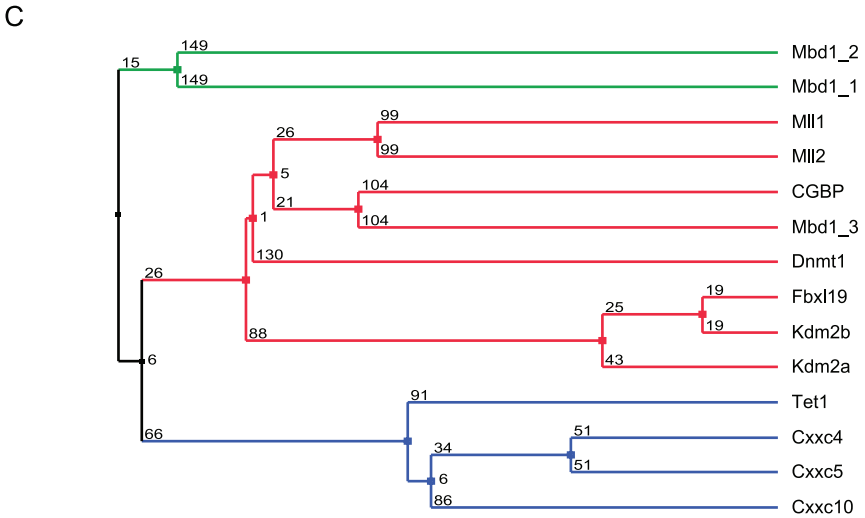
Introduction

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


A
Dnmt1
PBD
TS
CXXC
BAH1
BAH2
(KG)7
catalytic domain
1620 aa
Tet1
CXXC
Cys-rich
catalytic domain
2007 aa


B
Mll1
β1
α1
η1
β2
TTT
1 10 20 30 40 50 60
Mll1 1146 GRRSRRCGQCPGCQVPEDCGICTNCLDKPKFGGRNIKKQCKMRKCQNLQWMPSKASLQKQTKA 1209
Mll2 965 KMRMARCGHCRGCLRVQDCGSCVNCLDKPKFGGPNTKKQCVYRKCDKIEARKMERLAKKGRTI 1028
Dnmt1 650 AMKRRRCGVCEVCQQPE.CGKCKACKDMVKFGGTGRSKQACLKRRCPNLAVKEADDDEEADDDV 712
CGBP 167 KRSARMCGECEACRRTEDCGHCDFCRDMKKFGGPNKIRQKCRLRQCQLRARESYKYFPSSLSPV 230
Fbxl19 13 RRRRTRCRRCRACVR.TECGDCHFCRDMKKFGGPGRMKQSCLLRQCTAPVLPHTAVCLLCGEAG 75
Kdm2a 565 RRRRVRCRKCKACVQ.GECGVCHYCRDMKKFGGPGRMKQSCVLRQCLAPRLPHSVTCSLCGEVD 627
Kdm2b 47 RRRRTRCRKCEACLR.TECGECHFCKDMKKFGGPGRMKQSCIMRQCIAPVLPHTAVCLVCGEAG 109
Mbd1_3 302 QRQNRKCGACAACLRRMDCGRCDFCCDKPKFGGGNQKRQKCRWRQCLQFAMK..RLLPSAGSGS 363
Mbd1_2 246 MEKSRGCGVCRGCQTQEDCGHCCICLRSPRPGLKR..QWRCLQRRCFWGKRDSSKRGSKVASQR 307
Mbd1_1 197 MFKRVGCGDCAACLVKEDCGVCSTCRLQLPSDVASGLYCKCERRRCLRIMEKS..RGCGVCRGC 258
Tet1 569 RRKRKACGVCEPCQQKANCGECTYCK.....NRKNSHQICKKRKCEVLKKKPEATSQAQVTKE 626
Cxxc4 252 KKKRKRCGVCVPCKRLINCGVCSSCR.....NRKTGHQICKFRKCEELKKKPGTSLERTPVPS 309
Cxxc5 133 KKKRKRCGMCAPCRRRINCEQCSSCR.....NRKTGHQICKFRKCEELKKKPSAALEKVMLPS 190
Cxxc10 15 RKKRKRCGTCDPCRRLENCGSCTSCT.....NRRT.HQICKLRKCEVLKKKAGLLKEVSGPPL 71


C
Mbd1_2
Mbd1_1
Mll1
Mll2
CGBP
Mbd1_3
Dnmt1
Fbxl19
Kdm2b
Kdm2a
Tet1
Cxxc4
Cxxc5
Cxxc10

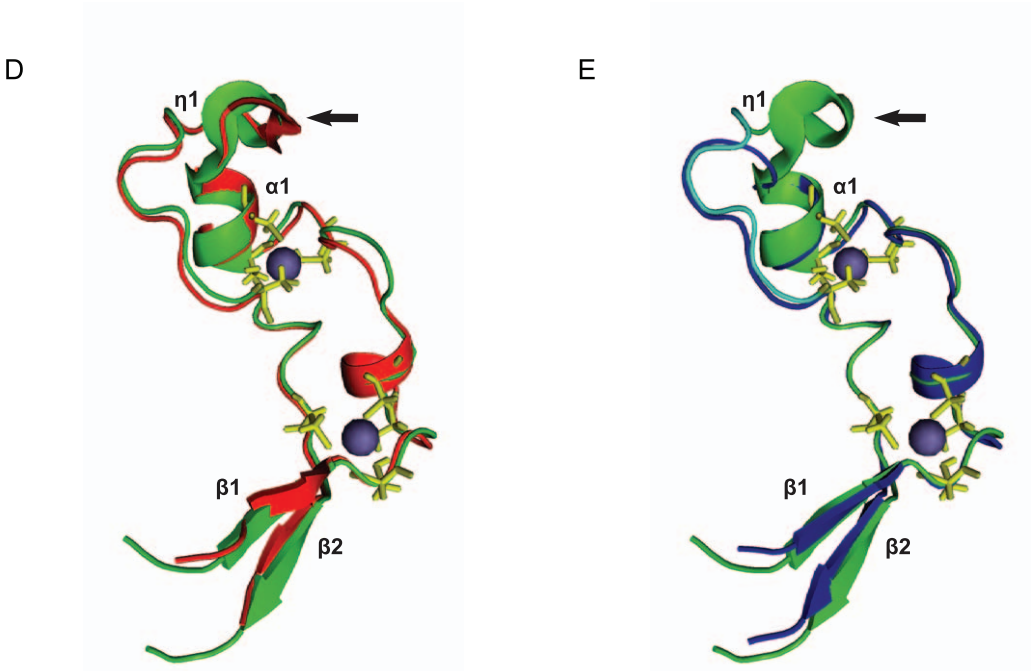

D
η1
α1
β1
β2
E
η1
α1
β1
β2

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 β1 and β2 (black arrows), α helices α1 and α2 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 Zn^{2+} 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 3_{10} helix (η1 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$^{Dnmt1KF/AA}$ 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

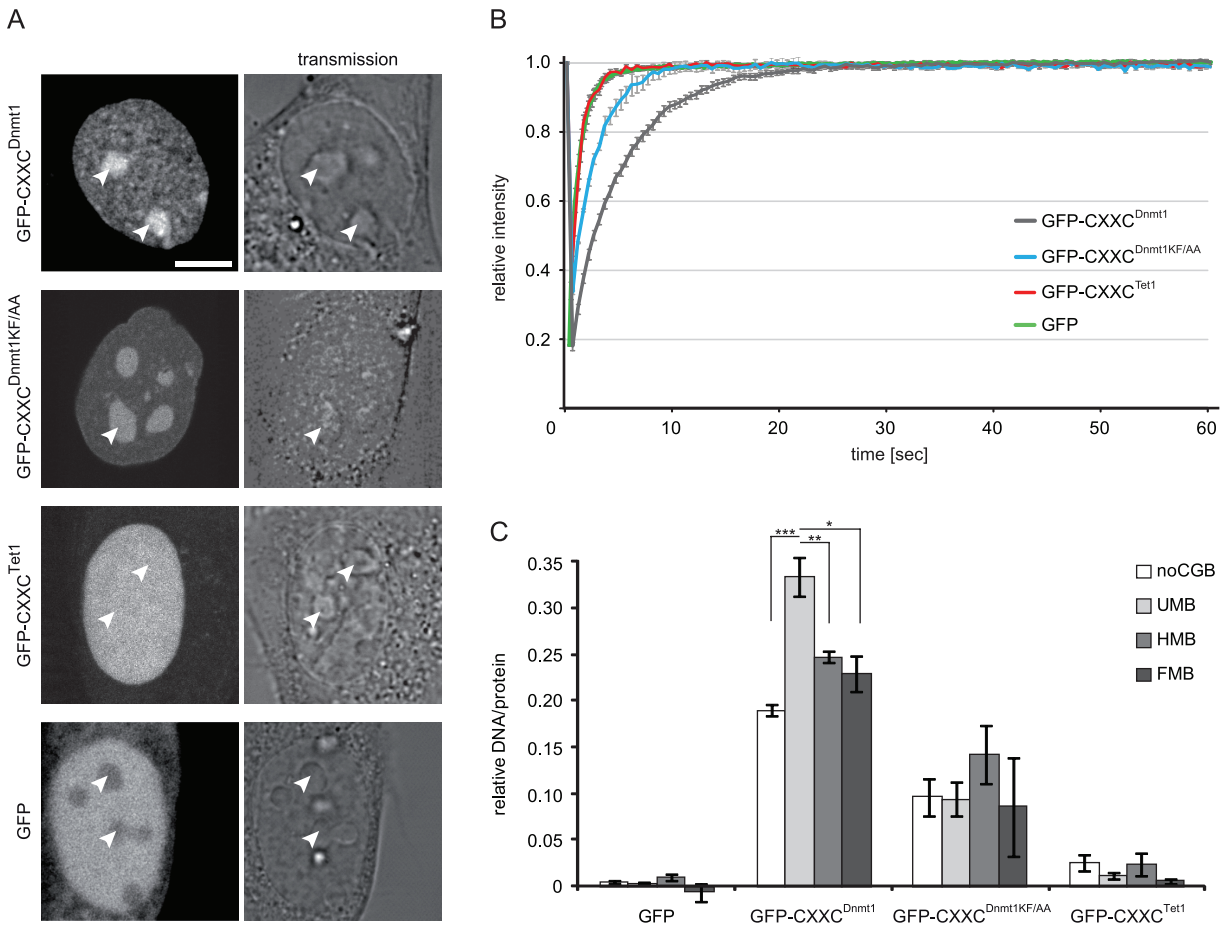


Figure 2. Properties of isolated Dnmt1 and Tet1 CXXC domains. (**A–B**) Subcellular localization (A) and binding kinetics (B) of GFP-CXXC^{Dnmt1}, GFP-CXXC$^{Dnmt1KF/AA}$, GFP-CXXC^{Tet1} and GFP in mouse C2C12 myoblasts. Localization and binding kinetics were independent from the cell cycle stage (Figures 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$^{\Delta CXXC}$). In C2C12 myoblasts GFP-Tet1 and GFP-Tet1$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ and GFP-NTR$^{\Delta CXXC}$, 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 ΔCXXC 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$, 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$^{\Delta CXXC}$. 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$^{\Delta CXXC}$ 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

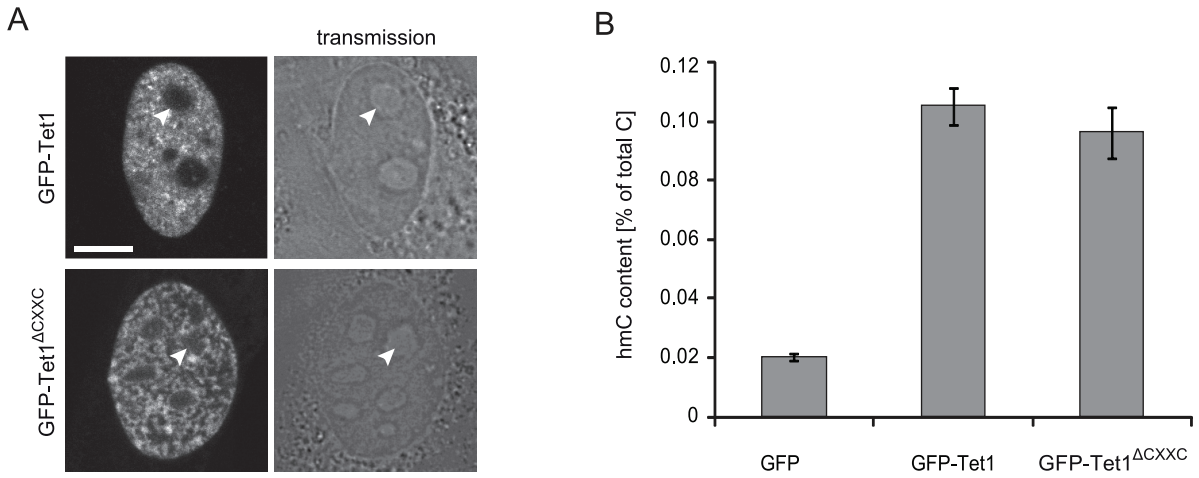


Figure 3. Cellular localization and *in vivo* catalytic activity of GFP-Tet1 and GFP-Tet1$^{\Delta CXXC}$. (A) Live images of C2C12 myoblasts expressing GFP-Tet1. Scale bar: 5 μm. (B) Genomic hmC content in HEK293T cells overexpressing GFP, GFP-Tet1 and GFP-Tet1$^{\Delta CXXC}$. Shown are mean values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

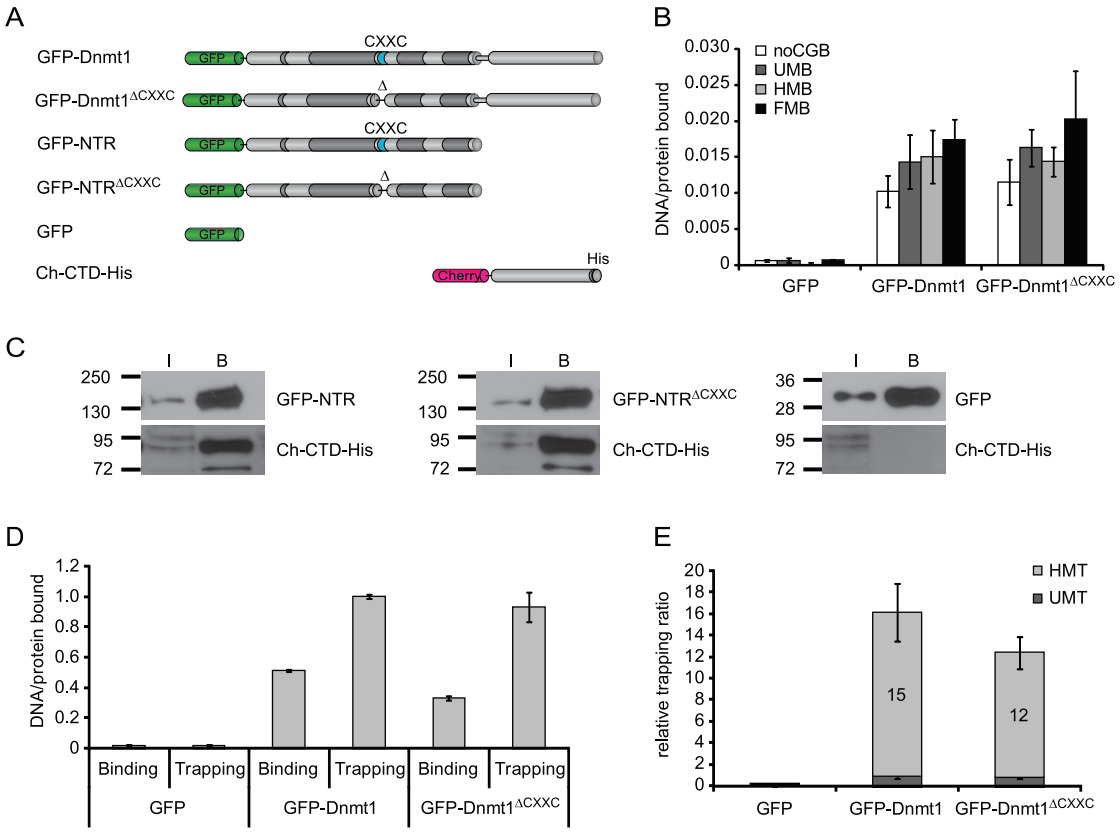


Figure 4. DNA binding specificity, intramolecular interaction and trapping of wild-type Dnmt1 and CXXC deletion constructs *in vitro*. (**A**) Schematic representation of Dnmt1 expression constructs. (**B**) DNA binding specificity of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ were assayed as described 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$^{\Delta CXXC}$, 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$^{\Delta CXXC}$ to monitor irreversible covalent complex formation with hemimethylated substrates. (**E**) Relative covalent complex formation rate of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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 [^{3}H]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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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.

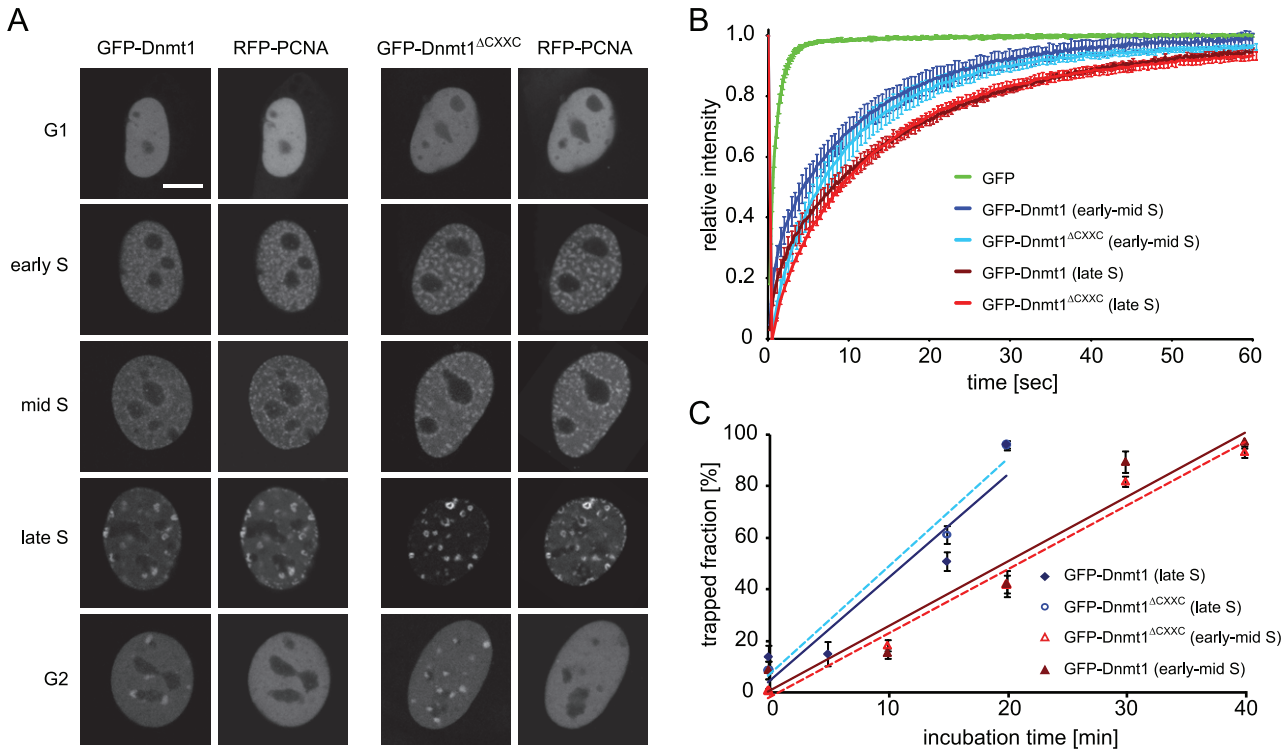


Figure 5. Cell cycle dependant cellular localization, protein mobility and trapping of wild-type Dnmt1 and CXXC deletion constructs in mouse C2C12 myoblasts. (**A**) Cell cycle dependent localization of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ constructs. Scale bar: 5 μm. (**B**) Analysis of binding kinetics of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ construct *in vivo*. We first compared localization and binding kinetics of GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ in mouse C2C12 myoblasts co-transfected with RFP-PCNA, which served as S-phase marker [41]. GFP-Dnmt1$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ to restore DNA methylation patterns in mouse *dnmt1*$^{-/-}$ ESCs. Cells transiently expressing either GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ 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.

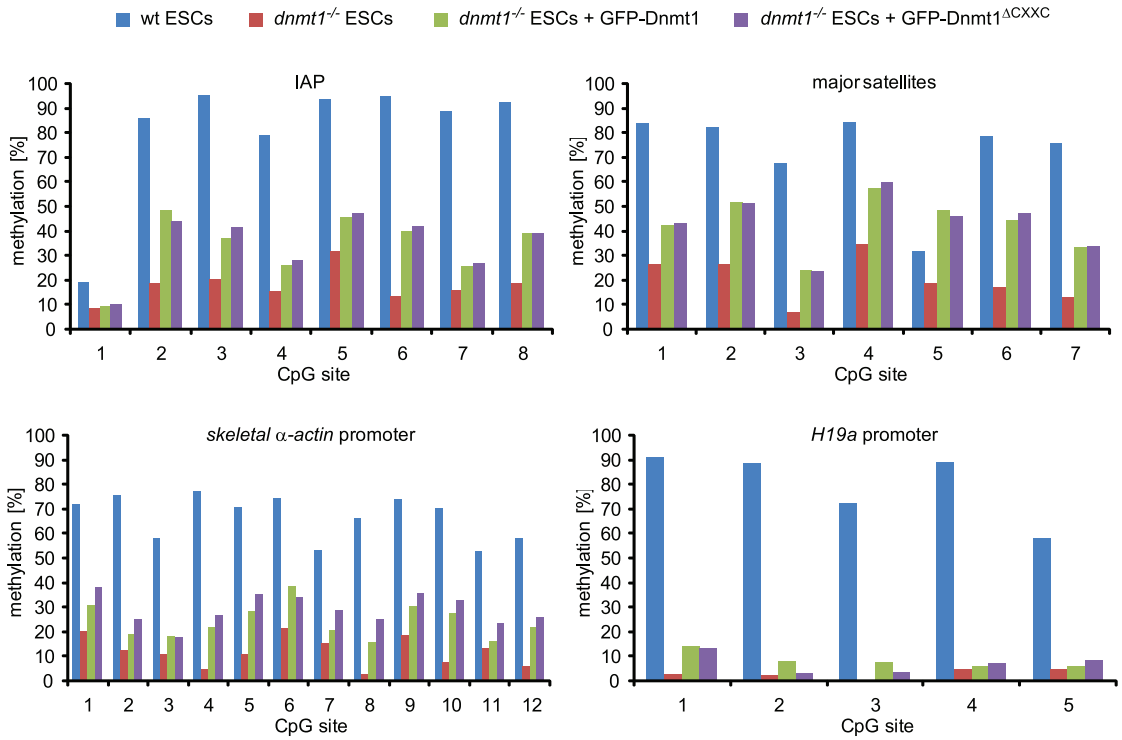


Figure 6. The CXXC deletion construct of Dnmt1 restores methylation in *dnmt1* null cells. Mouse *dnmt1*$^{-/-}$ ESCs transiently expressing GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ were isolated by FACS-sorting 48 h after transfection and CpG methylation levels within the indicated sequences were analyzed 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 ΔCXXC 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 Zn^{2+} 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$^{\Delta CXXC}$ 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$^{\Delta CXXC}$ construct was generated by subcloning the sequence coding for human DNMT1$^{\Delta CXXC}$ from the homonymous construct by Pradhan *et al.* [32] in the pEGFP-C2 vector (Clonetech). 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$^{\Delta CXXC}$ aa 569-621 of murine Tet1 were deleted from GFP-Tet1 using a type IIs 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$^{\Delta CXXC}$ was obtained by replacing the BglII-XhoI fragment of GFP-NTR with the same fragment of GFP-Dnmt1$^{\Delta CXXC}$. 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 ± detection bandwidth settings were used: 490/511±10 nm for GFP, 550/580±15 nm for ATTO550, 600/630±15 nm for ATTO590, 650/670±10 nm for ATTO647N and 700/720±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$^{\Delta CXXC}$ 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-[^{3}H]glucose donor (0.43 μM) 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 μM 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×/1.4 NA Plan-Apochromat oil immersion objective. Microscope settings were as described except that a smaller region of interest (3 μm×3 μm) 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 μg 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 AAAATGAGAAA-CATCCACTTG (forward primer) and CCATGATTTT-CAGTTTTCTT (reverse primer). For amplification we used Qiagen Hot Start Polymerase in 1x Qiagen Hot Start Polymerase buffer supplemented with 0.2 mM dNTPs, 0.2 μM forward primer, 0.2 μM 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 *dnmt1* null ESCs and Shriharsa Pradhan (New England BioLabs, Ipswich, MA) for the human DNMT1$^{\Delta CXXC}$ 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

1. 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.
2. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465–476.
3. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, et al. (2010) Dynamic changes in the human methylome during differentiation. Genome Res 320: 320–331.
4. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet 20: 116–117.
11. 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.
12. 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-Arteta AM, Di Giacco 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, Skalnik 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. Lorsbach 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, Samaranayke 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, Cascante 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. Fellinger 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. Thompson 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. Fellinger 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, Fellinger 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 nanotrap 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 Fellinger, 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'- CTCAACAACTAACTACCATCCGGACCAGAAGAGTCATCATGG -3'
MG-up	5'- CTCAACAACTAACTACCATCMGGACCAGAAGAGTCATCATGG -3'
noCG-up	5'- CTCAACAACTAACTACCATCTGGACCAGAAGAGTCATCATGG -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		550		Fill-In-550	5-aza-dCTP	Trapping
HMB 590	hemimethylated	590	MG-up	Fill-In-590	dCTP	Binding
HMB 647N		647N		Fill-In-647N		
HMT 550		550		Fill-In-550	5-aza-dCTP	Trapping
HMT 647N		647N		Fill-In-647N		
FMB 647N	fully methylated	647N	MG-up	Fill-In-647N	5methyl dCTP	Binding

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'- CCCTAATTAACTACAACCCA -3'
IAP LTR-2	5'- TGTAGTTAATTAGGGAGTGA -3'
Major Satellite-1	5'- AAAATGAGAAATATTTATTTG -3'
Major Satellite-2	5'- GAGAAATATATACTTTAGGA -3'

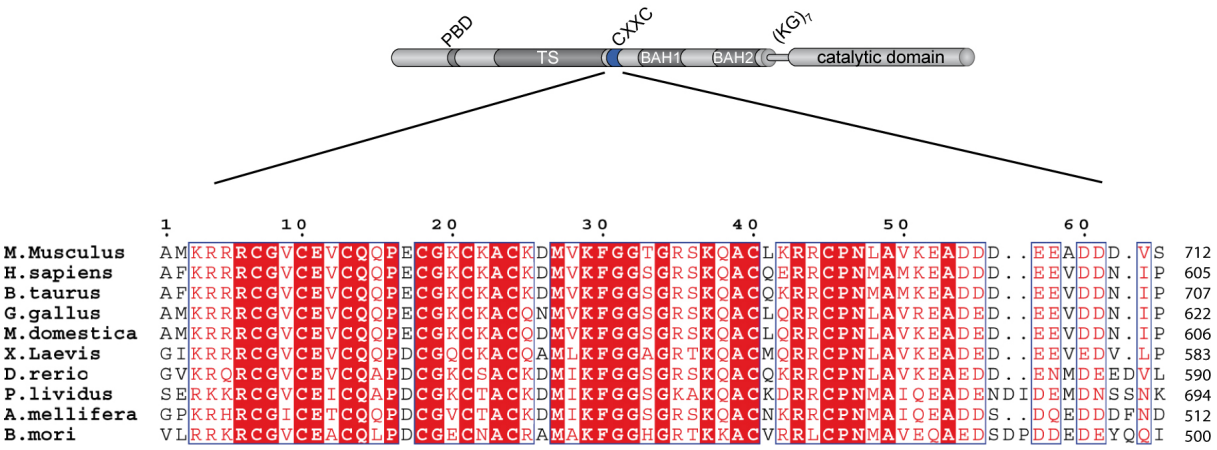


Figure S1. Dnmt1 domain structure and alignment of Dnmt1 CXXC domains from different species. Numbers on the right side indicate the position of the last amino acid in each sequence. PBD: PCNA binding domain; TS: targeting sequence; CXXC: CXXC-type zinc finger domain; BAH1 and 2: bromo-adjacent homology 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.

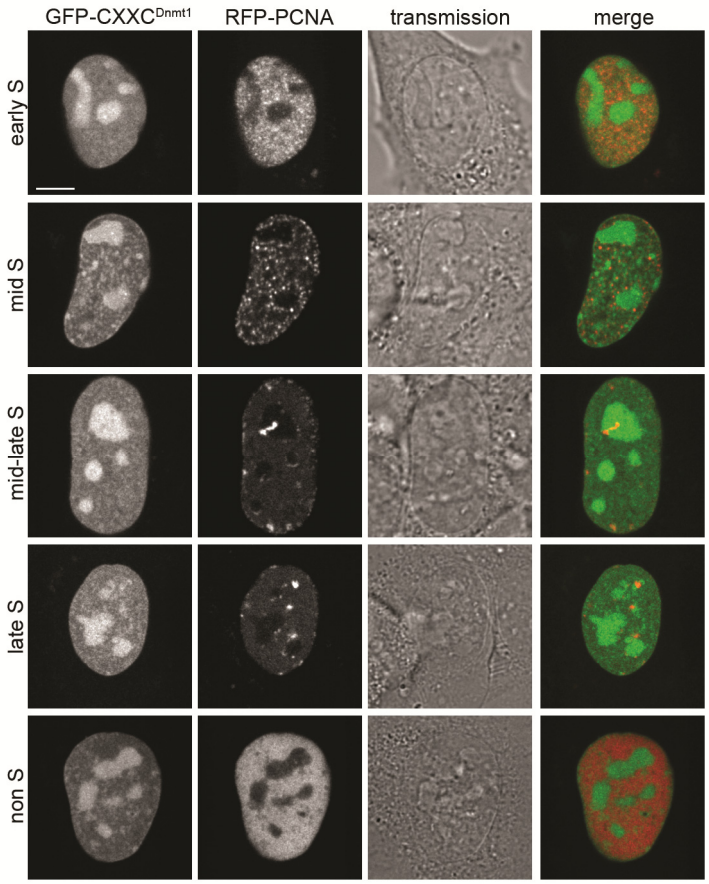


Figure S2. The cellular localization of GFP-CXXC^{Dnmt1} is independent of cell cycle stage. Live images of C2C12 mouse myoblasts cotransfected with expression constructs for GFP-CXXC^{Dnmt1} and RFP PCNA. The latter served for identification of the cell cycle stage. Scale bar: 5 µm.

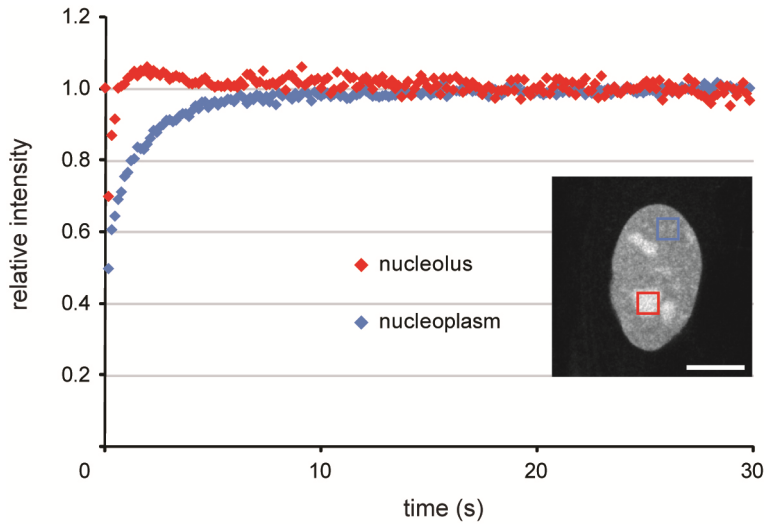


Figure S3. Differential mobility of GFP-CXXC^{Dnmt1} in nucleoli and nucleoplasm of mouse C2C12 myoblasts measured by FRAP analysis. Identical regions of interest over the nucleoplasm or nucleoli (as exemplified in the inset) were bleached and recovery curves were recorded over 30 seconds. GFP-CXXC^{Dnmt1} 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 µm.

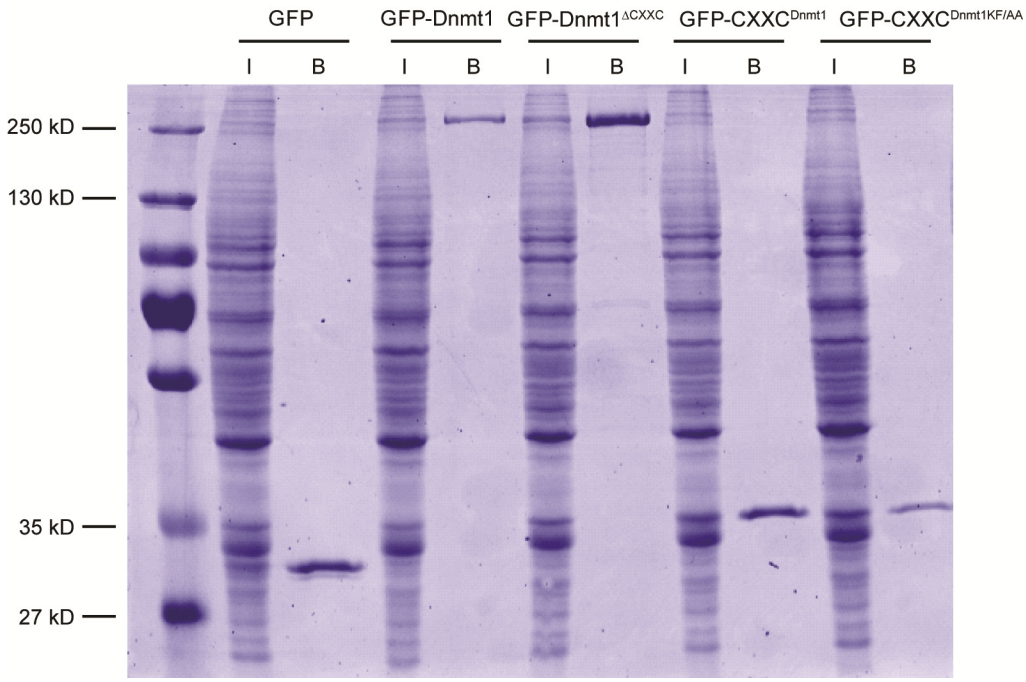


Figure S4. GFP fusion pulldowns from transiently transfected HEK293T cells using the GFP-trap. Shown is a SDS polyacrylamide gel stained with coomassie blue. I = input (1%); B = bound (10%).

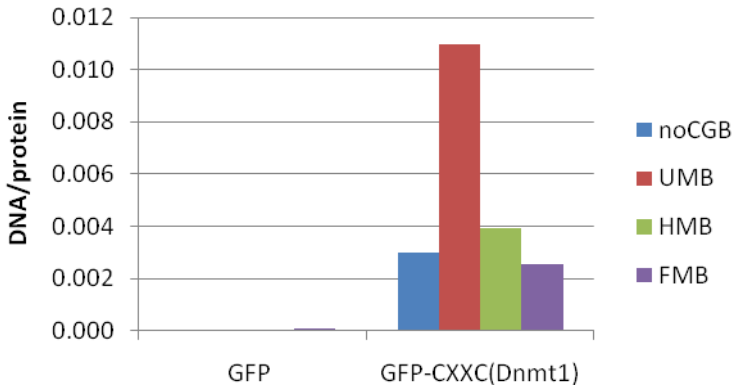


Figure S5. The CXXC domain of Dnmt1 preferentially binds unmethylated CpG sites. GFP and GFP-CXXC^{Dnmt1} purified from transiently transfected HEK293T cells with the GFP trap were challenged with fluorescent DNA substrates containing no CpG site or one central un-, hemi- or fully methylated CpG site 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.

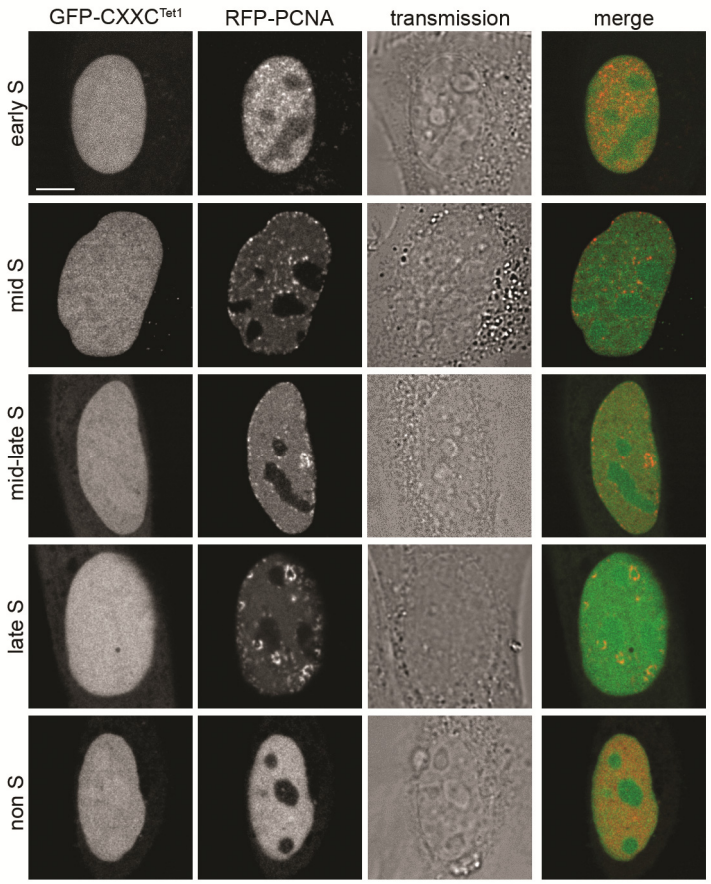


Figure S6. The cellular localization of GFP-CXXC^{Tet1} is independent of cell cycle stage. Live images of C2C12 mouse myoblasts cotransfected with expression constructs for GFP-CXXC^{Tet1} and RFP PCNA. The latter served for identification of the cell cycle stage. Scale bar: 5 µm.

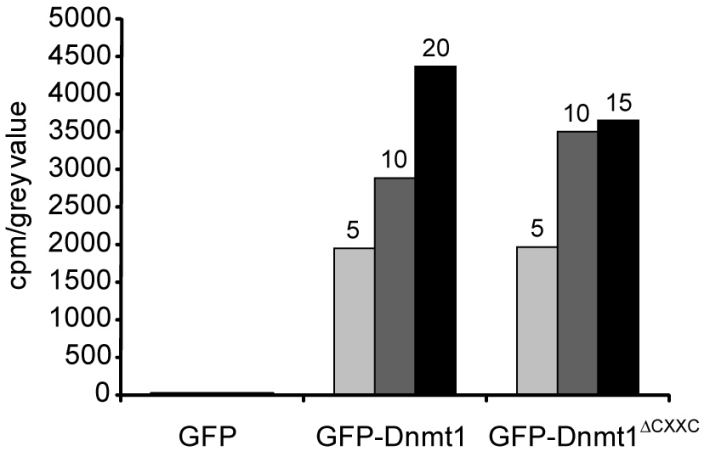


Figure S7. Radioactive methyltransferase activity assay for GFP Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$. The transfer of [3H]-methyl groups to poly(dI•dC)-poly(dI•dC) substrate was measured for increasing volumes of GFP fusion proteins immunopurified from transiently transfected HEK293T cells. Counts per 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.

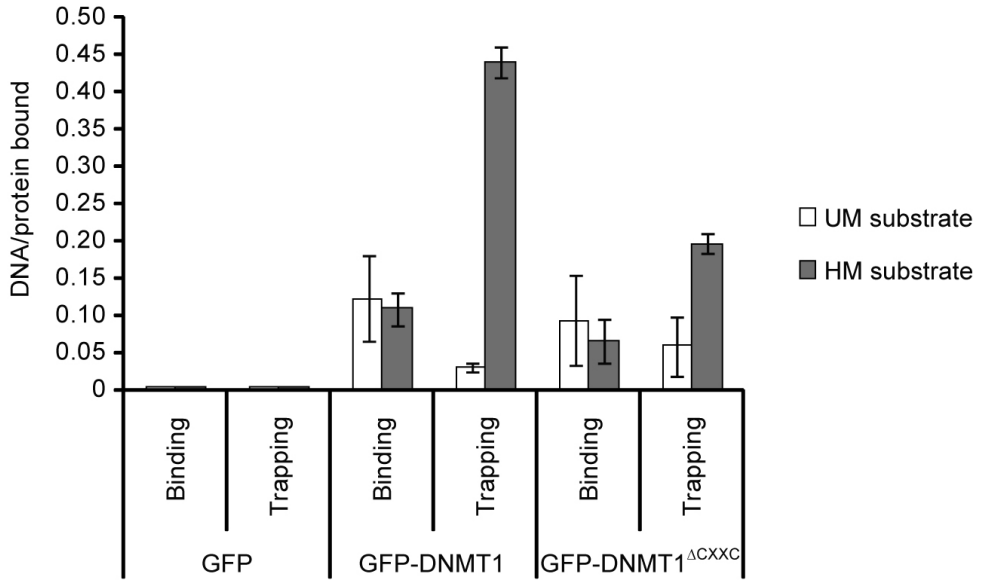


Figure S8. Competitive DNA binding and trapping assays for human GFP-DNMT1 and GFP-DNMT1$^{\Delta CXXC}$. GFP, GFP-DNMT1 and GFP-DNMT1$^{\Delta CXXC}$ were purified from transfected HEK293T cells using the GFP-trap and incubated with fluorescent DNA substrates containing one central unmethylated (UM) 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$^{\Delta CXXC}$ 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 NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.5 % Tween-20, 2 mM $MgCl_2$, 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 NaH_2PO_4 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 NaH_2PO_4 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-[^{3}H-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.

2.2 Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules

OPEN ACCESS Freely available online

PLOS ONE

Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules

Nan Liu[◔], Mengxi Wang[◔], Wen Deng, Christine S. Schmidt, Weihua Qin, Heinrich Leonhardt, Fabio Spada[*¤]

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

Abstract

Tet proteins are emerging as major epigenetic modulators of cell fate and plasticity. However, little is known about how Tet proteins are targeted to selected genomic loci in distinct biological contexts. Previously, a CXXC-type zinc finger domain in Tet1 was shown to bind CpG-rich DNA sequences. Interestingly, in human and mouse the *Tet2* and *Tet3* genes are adjacent to *Cxxc4* and *Cxxc10-1*, respectively. The CXXC domains encoded by these loci, together with those in *Tet1* and *Cxxc5*, identify a distinct homology group within the CXXC domain family. Here we provide evidence for alternative mouse Tet3 transcripts including the Cxxc10-1 sequence ($Tet3^{CXXC}$) and for an interaction between Tet3 and Cxxc4. *In vitro* Cxxc4 and the isolated CXXC domains of Tet1 and $Tet3^{CXXC}$ bind DNA substrates with similar preference towards the modification state of cytosine at a single CpG site. *In vivo* Tet1 and Tet3 isoforms with and without CXXC domain hydroxylate genomic 5-methylcytosine with similar activity. Relative transcript levels suggest that distinct ratios of $Tet3^{CXXC}$ isoforms and Tet3-Cxxc4 complex may be present in adult tissues. Our data suggest that variable association with CXXC modules may contribute to context specific functions of Tet proteins.

Citation: Liu N, Wang M, Deng W, Schmidt CS, Qin W, et al. (2013) Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules. PLoS ONE 8(5): e62755. doi:10.1371/journal.pone.0062755

Editor: Pierre-Antoine Defossez, Université Paris-Diderot, France

Received November 12, 2012; **Accepted** March 25, 2013; **Published** May 14, 2013

Funding: This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB 684, and SPP 1463, http://www.dfg.de/en/index.jsp) 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: Fabio.Spada@cup.uni-muenchen.de

◔ These authors contributed equally to this work.

¤ Current address: Faculty of Chemistry and Pharmacy, Ludwig Maximilians University Munich, Munich, Germany

Introduction

In higher eukaryotes methylation of genomic cytosine to 5-methylcytosine (mC) prominently contributes to epigenetic indexing of transcriptional activity. mC has long been regarded as a stable mark mediating permanent repression, but recent compelling evidence supports a highly dynamic modulation of transcriptional activity by both gain and loss of mC and several pathways for erasure of cytosine methylation have been proposed [1–3].

Recently, it has been shown that mC can be progressively oxidized to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxycytosine (caC) by a three member family of Tet α-chetoglutarate and Fe(II)-dependant dioxygenases [4–7]. The discovery of mC derivatives generated by enzymatic oxidation has kindled the idea that they represent intermediates in mC demethylation pathways. Although there is now support for hmC, fC and caC as demethylation intermediates, the relative abundance of hmC in tissues and the stability of its genomic patterns point to a role of this modification as an epigenetic mark with functional relevance distinct from mC [8–13]. Direct mutation of Tet2 or inhibition of its catalytic activity by 2-hydroxyglutarate generated through neomorphic IDH1/2 mutations lead to perturbed cytosine methylation patterns in hematopoietic progenitors and are associated with myeloid and lymphoid neoplasia [14–17]. Interestingly, Tet1 has been shown to mediate both transcriptional activation and repression and at least part of its repressive function has been proposed to be independent of its catalytic activity [18–20]. A role of Tet2 as transcriptional activator has been recently proposed [21], but it is not known whether Tet2 and Tet3 share the dual functional properties of Tet1. Maternally inherited Tet3 has been shown to oxidize paternal genomic mC in the zygote shortly after fertilization and is required for demethylation and subsequent efficient acitivation of the paternal *Oct4* and *Nanog* alleles [22].

Very few interactions involving Tet proteins have so far been reported [18,20,23] and even fewer known domains are identified in these proteins despite their relatively large size. As a consequence, little is known about how Tet proteins are targeted to specific genomic loci in distinct cell types and developmental stages. The only relatively well characterized modules in Tet proteins are the double-stranded β-helix fold typical of Fe(II)-dependent oxygenase domains and an N-terminal CXXC-type zinc finger in Tet1, thereby the latter has also been referred to as Cxxc6. The CXXC domains in these proteins, as well as that of Tet1, were shown to bind DNA sequences rich in CpG sites. Similar domains are also present in two factors, Cxxc4 and Cxxc5, shown to antagonize the canonical Wnt pathway and an additional CXXC domain is encoded in Cxxc10-1, a predicted ORF adjacent to the *Tet3* gene [24–27]. We have previously shown that the CXXC domains of Tet1, Cxxc4, Cxxc5 and Cxxc10-1 form a distinct homology group among CXXC domains [24]. Although human and mouse Tet3 have also been reported to

harbour a CXXC domain in recent reviews [28,29], experimental evidence for these claims was not available. CXXC domains are present in several other proteins with functions related to DNA and histone modification. Here we provide evidence for *cis* and *trans* association of mouse Tet3 isoforms with Cxxc10-1 and Cxxc4, respectively, and characterize the DNA binding properties of their CXXC domains with respect to the modification state of cytosine at CpG sites. Our data suggest that association with distinct CXXC domains may modulate Tet3 function.

Results

Identification and expression pattern of mouse Tet3 transcripts encoding a CXXC domain

The N-terminal region of Tet1 contains a CXXC-type zinc finger domain [4]. In contrast, none of the human and mouse annotated genomic or transcript sequences for Tet2 and Tet3 includes a sequence encoding such domain. However, in both the human and mouse genomes the *Tet2* and *Tet3* genes are adjacent to loci encoding CXXC domains, *Cxxc4* and *Cxxc10-1*, respectively (Fig. 1A) [24,30]. The *Cxxc4* and *Tet2* loci are 700 and 800 kb apart in the human and mouse genomes, respectively. These loci are transcribed in opposite orientations and encode distinct proteins, suggesting that they evolved through splitting of a *Tet1*-like ancestral gene and intergenic inversion. The Cxxc10-1 ORF was identified *in silico* about 13 kb upstream of the annotated transcriptional start site of *Tet3* and has the same orientation as the Tet3 ORF. Previously, we showed that the CXXC domains of Tet1, Cxxc10-1, Cxxc4 and Cxxc5 constitute a homology group distinct from CXXC domains present in several other factors with functions related to DNA or chromatin modification [24]. The proximity and co-orientation of the Cxxc10-1 and Tet3 ORFs in the human and mouse genomes suggest that alternative Tet3 transcripts may include the Cxxc10-1 ORF. This is also suggested by GenBank entries of Tet3 orthologues encompassing an N-terminal CXXC domain from other vertebrate species, including a *Xenopus* Tet3 transcript and a Tet3 protein homolog predicted from the genomic sequence of the naked mole rat (*Heterocephalus glaber*). Alignment of the CXXC domains from these Tet3 homologues with the CXXC domains of mouse Cxxc10-1, Tet1, Cxxc4 and Cxxc5 shows that they all belong to the same homology subgroup that we identified previously (Fig. 1B). In addition, the *Hydra* genome encodes a single Tet homolog and its predicted protein product contains an N-terminal CXXC domain with key features of this subgroup (Fig. 1B). These observations support the idea of a common ancestral *Tet* gene encoding a CXXC domain and that in addition to Tet1, this arrangement is preserved also in vertebrate Tet3.

Thus, we set out to verify whether Tet3 transcripts including the Cxxc10-1 ORF are expressed in the mouse. To this aim we performed conventional PCR on total cDNA template from a neural stem cell (NSC) line derived by *in vitro* differentiation of E14 embryonic stem cells (ESCs; Fig. S1). We used primer pairs spanning from the Cxxc10-1 ORF to the Tet3 ORF in exon 3 according to the annotated *Tet3* sequence. Cloning and sequencing of products identified two alternative transcripts where the exon containing the Cxxc10-1 ORF is spliced to the first position of either exon 2 or exon 3 of the annotated *Tet3* gene (Fig. 2A,B). These splicing events set the Cxxc10-1 ORF in frame with the annotated Tet3 coding sequence through its exon 2 and/or exon 3 sequences representing part of the 5′UTR in the annotated Tet3 transcript. Rapid amplification of cDNA 5′ ends (RACE) identified a 5′UTR sequence upstream of the Cxxc10-1 ORF including an additional exon upstream of the one encoding the Cxxc10-1 ORF (Fig. 2A). To verify the expression and size of alternative Tet3 transcripts we first performed northern blotting of RNA from the same NSC line and parental ESCs (Fig. 2D). In NSCs a cDNA probe comprising exons 3–6 of the annotated Tet3 transcript detected two bands with estimated sizes of 10.9 and 11.6 kb, roughly corresponding to the sizes of the annotated Tet3 transcript and those encoding the Cxxc10-1 ORF, respectively, assuming the same splicing events downstream of the annotated exon 3 (Fig. 2A). A probe spanning the Cxxc10-1 ORF detected only the 11.6 kb band. Each of these probes detected the same respective bands in RNA from ESCs, but their intensity was much weaker than for NSCs (not visible in Fig. 2C) despite the same amount of RNA was loaded. We found no evidence for independent expression of the Cxxc10-1 sequence in these samples, as no other distinct band was detected in the blots (Fig. S2). As final evidence for the expression of the Tet3 transcript including the Cxxc10-1 ORF and the annotated exon 2 (hereafter referred to as $Tet3^{CXXC}L$) we amplified its entire coding sequence as a single fragment (5412 bp encoding a polypeptide of 1803 aa) using cDNA from NSCs as template and confirmed its primary structure by sequencing (NCBI accession number JX946278). These results show that the use of an alternative promoter and alternative splicing lead to the expression of Tet3 transcripts containing the Cxxc10-1 ORF (altogether referred to as $Tet3^{CXXC}$) and that these transcripts share the same splicing organization with the previously annotated Tet3 transcript (hereafter referred to as Tet3) downstream of its exons 2 ($Tet3^{CXXC}L$) or 3 ($Tet3^{CXXC}S$; Fig. 2A).

To characterize the expression patterns of Tet3 and $Tet3^{CXXC}$ transcripts we performed real time PCR (qPCR) on cDNAs from stem cell lines and various adult mouse tissues (Fig. 3A). We set primer pairs for selective amplification of the $Tet3^{CXXC}$ transcript including exon 2 of the Tet3 transcript, the Cxxc10-1 ORF and exons 1–3 of Tet3. The levels of Tet3 and $Tet3^{CXXC}$ transcripts varied widely across the samples and were very low in ESCs, confirming our northern blot data. Notably, the ratio of Tet3 to $Tet3^{CXXC}$ transcripts was higher in brain regions relative to other tissues.

Cxxc4 interacts with Tet3 *in vivo* and is expressed in the adult brain

The evolutionary association of Tet proteins with a distinct group of CXXC domains *in cis* raises the question as to whether they associate with this type of CXXC module also *in trans*. Therefore we probed the interaction of each of the three Tet proteins with Cxxc4 and Cxxc5 using a mammalian fluorescent three hybrid assay (F3H). In this assay baits fused to GFP are anchored to a *lac* operator array integrated in the genome of BHK cells and challenged with preys fused to a red fluorescent protein [31–33]. The colocalization of prey and bait at the *lac* operator array reflects their interaction (Fig. 4 and Fig. S3). The pair Tet3-Cxxc4 tested positive in both prey-bait combinations, while all other Tet-Cxxc4/5 pairs showed no interaction. However, we could not detect coimmunoprecipitation of Tet3 and Cxxc4 fluorescent fusion constructs overexpressed in HEK293T cells (not shown), which may be due to the lack or limiting endogenous levels of bridging factors in these cells. Cxxc4 and 5 have been shown to antagonize canonical Wnt signaling by binding to cytoplasmic Disheveled [25–27]. However, expression of fluorescent fusions revealed a prevalently nuclear localization of Cxxc4 in BHK cells, C2C12 myoblasts and ESCs (Fig. 4 and Fig. S4). In this regard we note that the KKKRK sequence (Fig. 1B) at the N-terminus of the CXXC domain in both Cxxc4 and 5 is a perfect match to the minimal prototypic nuclear localization sequence of

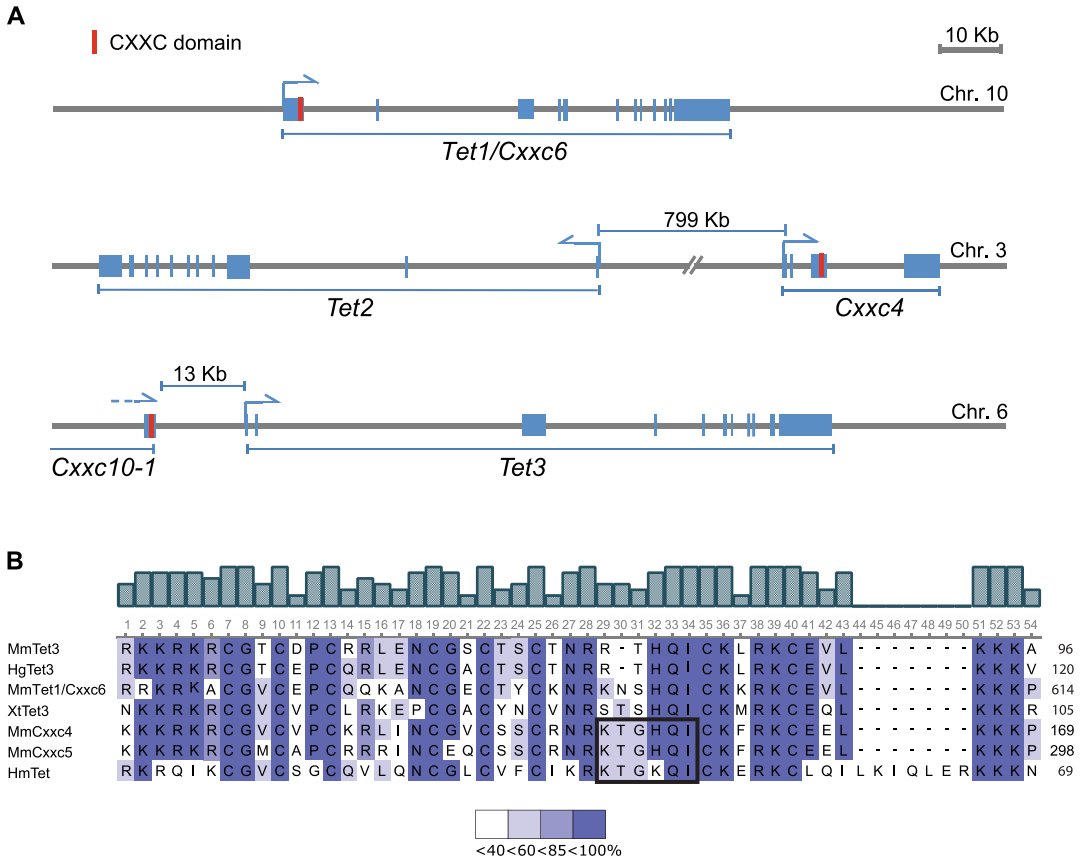


Figure 1. Genomic arrangement of mouse *Tet* genes and adjacent *Cxxc* loci (A) and homology of CXXC domains from mouse Cxxc4, Cxxc5 and Tet homologues in various animal species (B). (A) Schematic representation of mouse *Tet1*, *Tet2/Cxxc4* and *Tet3/Cxxc10* loci. Exons are depicted as blue rectangles. Annotated transcriptional start sites and transcription orientation are indicated with half arrows. (B) Alignment of CXXC domains from mouse Cxxc4, Cxxc5 and Tet homologues in various animal species (Mm, *Mus Musculus*; Hg, *Heterocephalus glaber*; Xt, *Xenopus tropicalis*; Hm, *Hydra mangipallata*). The alignment was generated with Unipro UGENE [64]. Numbers on the right side indicate the position of the last amino acid in the corresponding protein. The KTXXXI motif, previously identified as determinant for the interaction of Cxxc4 with Dvl [54], is boxed (see Discussion). The scale at the bottom indicates the upper limit of percent identity represented by each color. GenBank accession numbers: MmCxxc10, JX946278; XtTet3, NP_001090656.1; HgTet3, EHB01729.1; MmTet1, NP_081660.1; MmCxxc4, NP_001004367; MmCxxc5, NP_598448; HmTet, XP_002161163.1.
doi:10.1371/journal.pone.0062755.g001

the SV40 large T antigen [34,35], and that Cxxc5 was also found to be predominantly nuclear in various cell types [27,36].

Next we determined the levels of Cxxc4 and Cxxc5 transcripts in adult mouse tissues and stem cell lines (Fig. 3B). Interestingly, among adult tissues Cxxc4 was expressed mainly in the brain, where Tet3 transcripts that do not encode the CXXC domain were more abundant relative to Tet3^{CXXC} transcripts. In contrast, Cxxc5 mRNA was detected ubiquitously and apart from ESCs its levels were substantially higher than those of Cxxc4. No obvious correlation could be found between the levels of Cxxc5 transcripts and those of any of the Tet transcripts analyzed (Fig. S5).

The CXXC domains of Tet1, Tet3^{CXXC}, CXXC4 and CXXC5 bind CpG containing DNA substrates

Previously, we showed that a construct corresponding to the isolated CXXC domain of mouse Tet1 (aa 561–614) with an N-terminal GFP tag (GFP-CXXC^{Tet1}) has very low DNA binding activity *in vitro* [24]. In contrast, Xu *et al.* showed that a larger fragment of mouse Tet1 including the CXXC domain (aa 512–671) binds CpG rich DNA sequences [37]. To resolve this discrepancy we directly compared the DNA binding activity of the isolated CXXC domain of Tet1 with GFP fused either to its N-terminus (the GFP-CXXC^{Tet1} construct we used previously) or to its C-terminus (CXXC^{Tet1}-GFP), as well as the same Tet1 fragment used by Xu *et al.* with an N-terminal GFP tag (GFP-Tet1$^{512–671}$; Fig. S6A). These constructs were overexpressed in HEK293T cells, immunopurified and challenged with fluorescent DNA substrates bearing a single CpG site that was either unmodified, symmetrically methylated or symmetrically hydroxymethylated in direct competition [24,38–41]. GFP-Tet1$^{512–671}$ and CXXC^{Tet1}-GFP showed similar and substantial binding activity toward substrates containing unmodified and symmetrically methylated CpG sites and were preferred to the substrate with the hydroxymethylated CpG, consistent with previous data [37]. Instead, a much lower DNA binding activity was confirmed for GFP-CXXC^{Tet1} (Fig. S6B). We conclude that the DNA binding properties observed for the Tet1$^{512–671}$ fragment are attributable to the CXXC domain and that direct fusion of GFP at the N-terminus of the isolated CXXC domain interferes with DNA binding.

These results and the high similarity shared by the CXXC domains of Tet1, Tet3^{CXXC} and the Tet3 interactor Cxxc4 prompted us to compare their DNA binding properties. Cxxc4-GFP, Cxxc5-GFP, GFP-Tet1, CXXC^{Tet1}-GFP as well as full

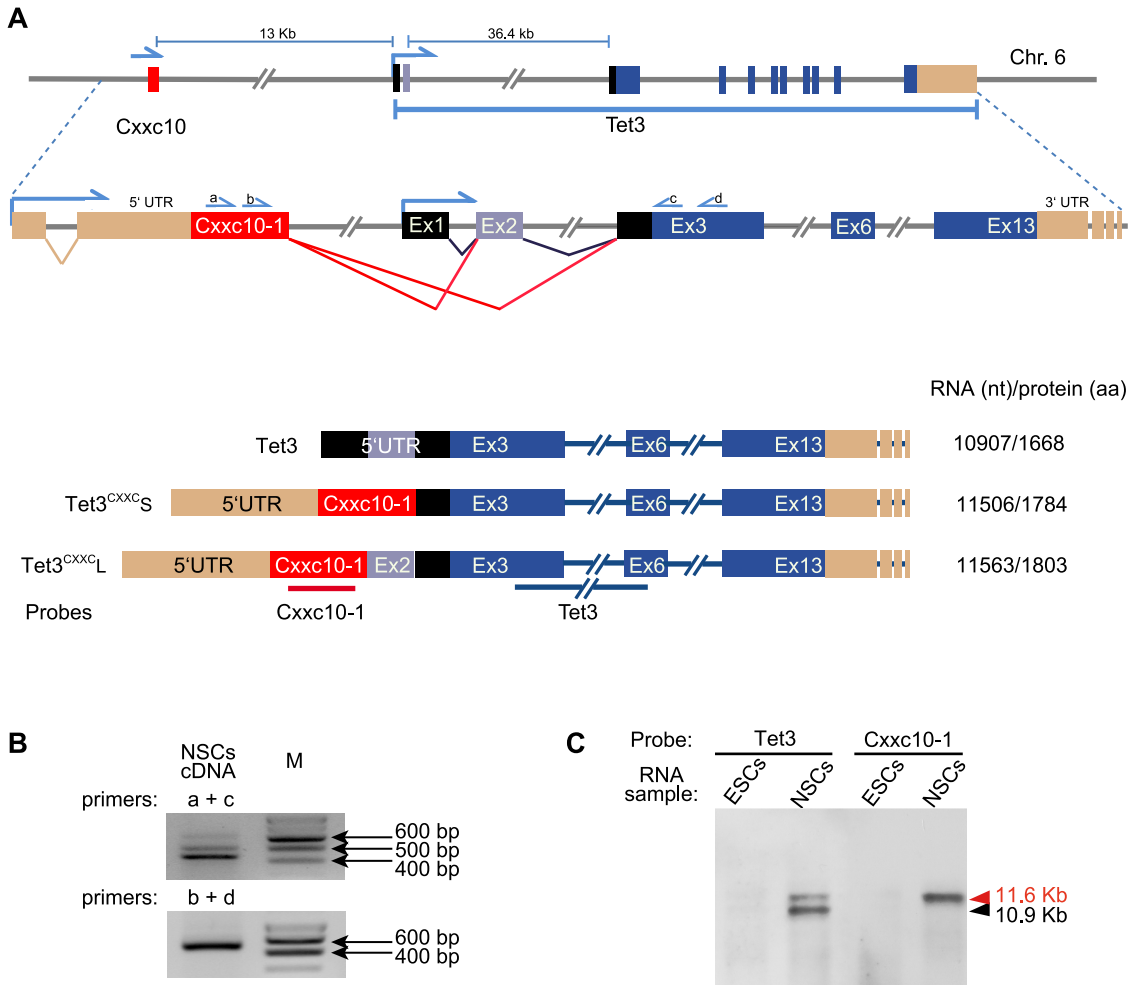


Figure 2. Identification of mouse Tet3 transcript variants encoding a CXXC domain. (A) Drawing illustrating the generation of alternative transcripts from the *Tet3/Cxxc10-1* locus. The positions of primers used in B are reported. The lower part reports a schematic representation of alternative Tet3 transcripts. The positions of the probes used for northern blotting in C are reported. (B) Amplification of fragments from NSCs cDNA identifying Tet3 transcripts that include the Cxxc10-1 ORF. (C) Northern blot detection of alternative Tet3 transcripts in ESCs and NSCs (see Fig. S1 for full and additional blots).
doi:10.1371/journal.pone.0062755.g002

length Tet1, Tet3 and Tet3^{CXXC}L constructs with an N-terminal GFP tag were subjected to similar DNA binding assays as above (Fig. 5 and Fig. S7). CXXC^{Tet3}-GFP corresponds to the isolated CXXC domain of the Cxxc10-1 ORF with GFP fused to its C-terminus and is therefore analogous to CXXC^{Tet1}-GFP. Although we could not detect interactions between Tet proteins and Cxxc5, we investigated the DNA binding potential of the latter as its CXXC domain is also highly homologous to that of Tet1. CXXC domains belonging to a distinct homology class, including the CXXC domain of Dnmt1 (CXXC^{Dnmt1}), were shown to preferentially bind CpG-containing sequences [24,42–46]. Therefore, we first determined the binding preference of our constructs with respect to DNA substrates differing only for the presence or absence of a single central CpG site and compared it to that of the CXXC domain of Dnmt1 (GFP-CXXC^{Dnmt1}; Fig. S7). Cxxc4, Cxxc5 and all Tet constructs showed higher DNA binding activity as well as similar and substantial preference for the substrate containing a CpG site as compared to GFP-CXXC^{Dnmt1}.We then determined the binding preference with respect to substrates containing a single central CpG site with distinct cytosine modifications as shown above for CXXC^{Tet1} constructs. Cxxc4-GFP, Cxxc5-GFP and CXXC^{Tet3}-GFP displayed similar binding properties, with decreasing preference for substrates with the unmodified, symmetrically methylated and symmetrically hydroxymethylated CpG site. In contrast and as shown above, CXXC^{Tet1}-GFP did not discriminate between substrates with unmodified and symmetrically methylated CpG. In the case of full length Tet1, Tet3 and Tet3^{CXXC}L constructs, incubation with a 4-fold molar excess of DNA substrates is expected to minimize potential competition among multiple DNA binding sites. GFP-Tet1 displayed the same substrate preference as the isolated CXXC domain of Tet1 (CXXC^{Tet1}-GFP), albeit with an 8-fold increase in binding activity, indicating that sequences outside the CXXC domain (very likely the catalytic domain) contribute to the affinity for DNA without altering the substrate preference. In contrast, both GFP-Tet3 and GFP-Tet3^{CXXC}L showed a relative increase in binding activity toward the substrate with methylated CpG site as compared to CXXC^{Tet3}-GFP. Thus, in Tet3^{CXXC}L features outside the CXXC domain override the binding preference of the latter.

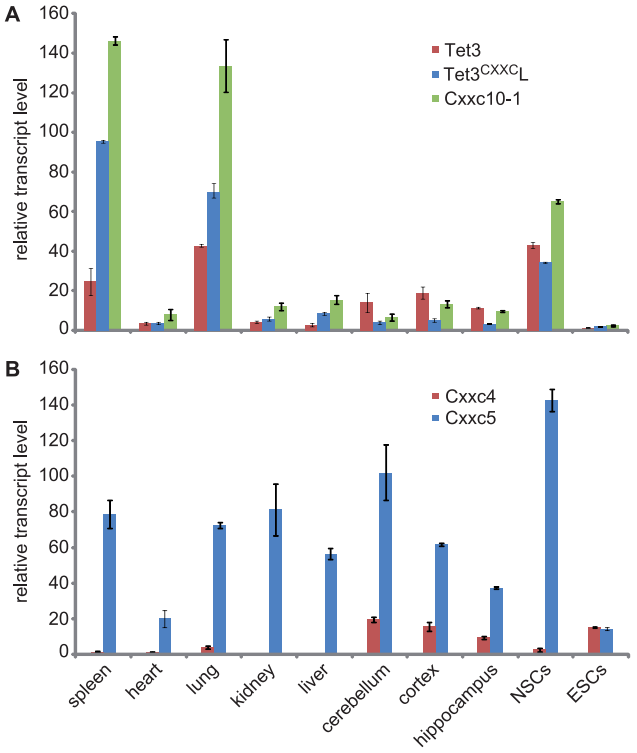


Figure 3. Levels of Tet3, Cxxc4 and Cxx5 transcripts in mouse adult tissues, NSCs and ESCs. Transcript levels were determined by qPCR analysis of total cDNA. (A) Amplfied fragments identify the Tet3 mRNA refseq NM_183138 (Tet3), the alternative Tet3 transcript containing the Cxxc10-1 ORF and exon 2 of NM_183138 (Tet3^{CXXC}L) and all transcripts including the Cxxc10-1 ORF. (B) Cxxc4 and Cxx5 transcript levels. Data relative to kidney, liver, cerebellum and cortex samples are from three biological replicates (two 6 week old 129Sv mice and a 30 week old C57BL/6 mouse). Data relative to spleen, heart, lung and hippocampus are from two biological replicates (a 6 week old 129/Sv mouse and a 30 week old C57BL/6 mouse). Data relative to NSCs and ESCs are from three independent cultures each. Shown are mean values and standard errors of the mean (SEM).
doi:10.1371/journal.pone.0062755.g003

Tet3^{CXXC} oxidizes genomic mC *in vivo* and shows slightly lower mobility than the Tet3 isoform lacking the CXXC domain

We then compared the activity of Tet1 and Tet3 isoforms with or without CXXC domain by determining global levels of genomic hmC in HEK293T cells transiently transfected with GFP-tagged constructs (Fig. 6). A similar increase of hmC levels was observed in cells transfected with GFP-Tet1, GFP-Tet3 and GFP-Tet3^{CXXC}L, the latter possibly showing higher conversion of mC to hmC. As further characterization of Tet3 isoforms we compared nuclear localization and mobility of GFP-Tet3 and GFP-Tet3^{CXXC}L in C2C12 myoblasts. Both constructs were diffusely distributed throughout the nucleus with exclusion of nucleoli and large clusters of pericentric heterochromatin (chromocenters; Fig. S8A). After photobleaching half of the nucleus the fluorescence of GFP-Tet3^{CXXC}L recovered more slowly and reached a plateau at a lower level than that of GFP-Tet3 (Fig. S8B). These differences were small, but reproducible.

Thus, the presence of the CXXC domain in Tet3 does not affect and perhaps promotes conversion of mC to hmC, while it reduces its mobility and slightly increases the immobile fraction, suggesting that the CXXC domain contributes to additional nuclear interactions.

Discussion

Very limited information is available as to how Tet family dioxygenases target selected genomic loci in distinct developmental and cellular contexts. CXXC-type zinc finger modules have been shown to direct chromatin modifying activities, including Tet1, to CpG rich sequences where they contribute to the establishment of a transcriptionally competent environment [37,46–48]. We now provide evidence that alternative mouse Tet3 isoforms associate with distinct CXXC modules also endowed with DNA binding activity. Alternative presence of an intrinsic CXXC domain or interaction with Cxxc4 may constitute the basis for differential targeting of Tet3 isoforms. In this regard we note that the ratio of Tet3 to Tet^{CXXC} transcripts was higher in brain tissues where Cxxc4 transcripts were more abundant. However, we found that *in vitro* Cxxc4 and the CXXC domain of Tet3^{CXXC} isoforms have similar binding preference with respect to the modification state of cytosine at CpG sites and that DNA binding elements other than the CXXC domain dominate the global DNA substrate preference of Tet3^{CXXC}. Further investigation is required to assess how DNA binding by Cxxc4 and the CXXC domain of Tet3^{CXXC} contribute to Tet3 function *in vivo*.

While the current manuscript was under review a report was published showing a role for CXXC domain-containing Tet3 orthologues in early neural and eye development of Xenopus [49]. In the same publication the cloning of human and mouse Tet3 isoforms containing a CXXC domain was reported, the latter being identical to our mouse Tet3^{CXXC}L, but no expression or functional data were provided for these mammalian isoforms. Importantly, their isothermal titration calorimetry data on the DNA binding properties of the CXXC domain from *Xenopus* and human TET3 isoforms are fully consistent with the results of our DNA binding assays with the CXXC domain of mouse Tet3^{CXXC}.

Association with distinct CXXC domains may also modulate Tet protein function by additional mechanisms. Interestingly, Cxxc4 and Cxxc5 were shown to antagonize Wnt signaling by competing with Axin for binding to Dishevelled (Dvl), thus leading to destabilization of β-catenin [25–27]. Although β-catenin stabilization by Dvl occurs in the cytoplasm, nuclear Dvl has been shown to interact with a two megadalton TCF/β-catenin transcriptional complex and to be required for activation of Wnt pathway target genes [34,50,51]. Importantly, we found that Cxxc4, like Cxxc5, is predominantly nuclear. Interestingly, other factors interacting with Dvl such as DP1 and NFAT are known to shuttle between cytoplasm and nucleus [52,53]. DP1 was shown to play dual regulatory roles in Wnt signaling depending on its nucleocytoplasmic localization, while dephosphorylated NFAT was proposed to inhibit canonical Wnt signaling by sequestering Dvl from transcriptional complexes in the nucleus. Therefore, it will be interesting to investigate whether Cxxc4 and Tet3 are involved in nuclear TCF/β-catenin complexes and affect transcription of their target genes. A KTXXXI motif within the CXXC domain of Cxxc4 was previously shown to be minimally required for the interaction with Dvl [54], but is poorly conserved in the CXXC domain of vertebrate Tet3^{CXXC} isoforms (Fig. 1B). Differential expression of Tet3 isoforms and interaction with Cxxc4 may therefore modulate the recruitment of Tet3 to TCF/β-catenin complexes. Thus, our results warrant further investigation on the functional relevance of the association between Tet proteins and CXXC modules.

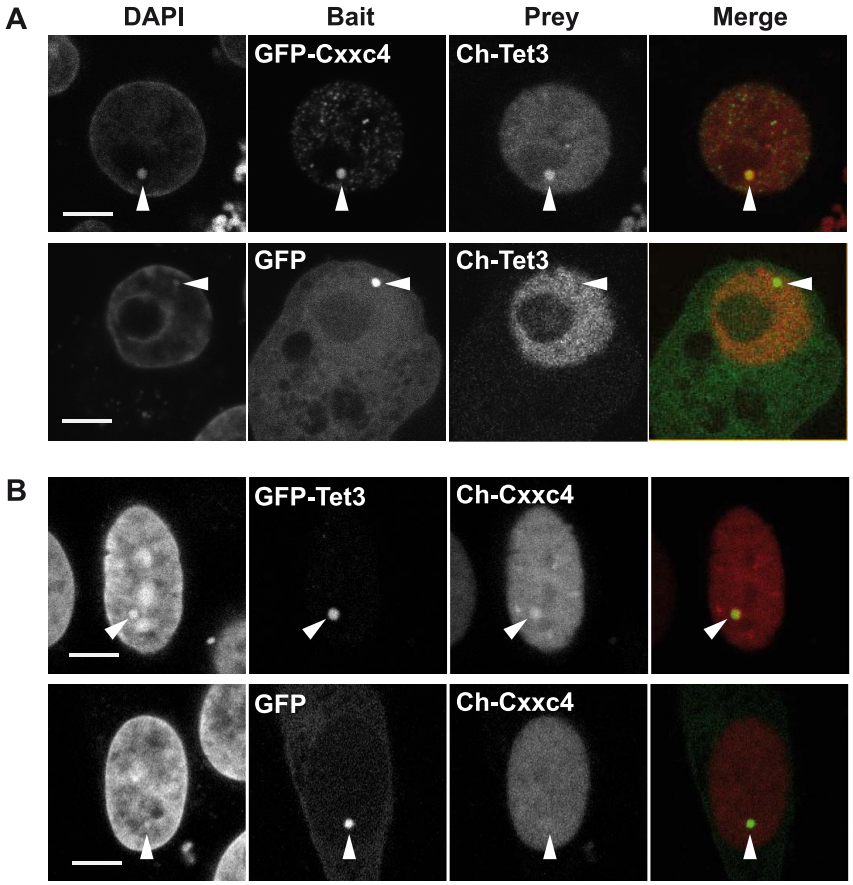


Figure 4. Tet3 and Cxxc4 interact *in vivo*. The interaction was detected by the F3H assay in BHK cells harboring a *lac* operator array (see text and Fig. S2 for explanations). (A) An N-terminal fusion of Tet 3 with Cherry (Ch) was used as prey and GFP-Cxxc4 (upper row) or GFP (as control; lower row) as baits. Localization patterns are representative of 8 (upper row) and 9 (lower row) out of 10 imaged cells. (B) Ch-Cxxc4 was used as prey and GFP-Tet3 (upper row) or GFP (as control; lower row) as baits. Localization patterns are representative of 4 out of 5 (upper row) and 6 out of 7 (lower row) imaged cells. Arrowheads indicate the position of the *lac* operator array as identified by bait signals (GFP channel). Scale bars: 5 μm.
doi:10.1371/journal.pone.0062755.g004

Materials and Methods

Ethics statement

Collection of animal tissues was performed in accordance with the German Animal Protection Law. No experiment was performed on live animals. Mice were painlessly killed under anesthesia with Isofuran before harvesting organs and tissues. According to the German Animal Welfare Act (Part III: "Killing of animals", Section 4, May 18, 2006) postmortem collection of tissues and organs does only require summary notification to the animal protection institution, but does not require any special permission. Therefore, this study was not registered as an animal experiment and the animal tissues used are registered only in the annual report of animals sacrificed for research and study to the relevant authority.

Cell culture

E14 [55] and CGR8 [56] ESCs were maintained in gelatin coated flasks with DMEM high glucose containing 16% FBS, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin (all from PAA Laboratories GmbH), 1× MEM Non-essential Amino Acid Solution and 0.1 mM β-mercaptoethanol (both from Invitrogen) and supplemented with 3 μM CHIR 99021 and 1 μM PD0325901 ("2i"; both from Axon Medchem). The NSC line ENC1 used throughout this study was derived from E14 ESCs as described [57] and was maintained in Knockout-DMEM/F12 containing 2 mM GlutaMAX-I (both from Invitrogen) 100 U/ml penicillin, 100 μg/ml streptomycin, and supplemented with 1% N2 (custom made according to [58]) and 20 ng/ml each FGF-2 and EGF (PeproTech). ENC1 cells homogeneously expressed NSC markers Nestin, Pax6 and Olig2 (Fig. S1). C2C12 myoblasts [59] HEK293T [60] cells and BHK cells with a stably integrated *lac* operator array [61] were cultured as described [24,32,33].

Expression constructs

Throughout this study enhanced GFP and monomeric Cherry fusion constructs were used and are referred to as GFP and Cherry fusions, respectively, for brevity. GFP-Tet1 and GFP-CXXC^{Tet1} were described previously [24]. For other GFP and Cherry fusions cDNA was generated from either ENC1 NSCs (Tet3, Tet3^{CXXC}L, CXXC^{Tet3}, Cxxc5) or parental E14 ESCs (Cxxc4) with the RevertAid Premium First Strand cDNA Synthesis kit (Thermo Scientific). Coding sequences were amplified using Phusion High-Fidelity DNA polymerase (New England Biolabs) and primers listed in Table S1. Sequences coding for Tet3, Tet3^{CXXC}L and Tet1$^{512–671}$ were inserted into the pCAG-GFP-IB vector [62] or the derived pCAG-Cherry-IB vector to generate N-terminal GFP and Cherry fusions, respectively. Sequences coding for CXXC^{Tet1} CXXC^{Tet3}, Cxxc4 and Cxxc5 were inserted into pCAG-Tev-GFP (derived from pCAG-GFP-IB) to generate C-terminal GFP

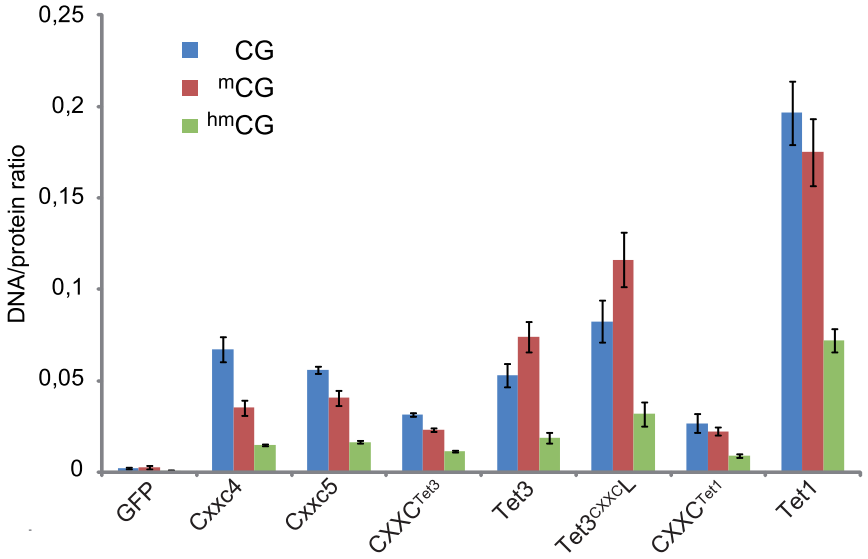


Figure 5. *In vitro* DNA binding properties of Cxxc4 and 5, isolated CXXC domains and full length constructs of Tet1 and Tet3^{CXXC}. All proteins were expressed as GFP fusion constructs in HEK293T cells and affinity purified using a GFP-trap. Fluorescently labeled DNA substrates with the same sequence and a single CpG site either unmethylated, symmetrically methylated or symmetrically hydroxymethylated were incubated in direct competition. Shown are mean values of bound substrate/protein ratios and SEM from n independent replicate experiments: Tet1, n = 10; Tet3, CXXC^{Tet3}, n = 6; Tet3^{CXXC}L, n = 7; CXXC^{Tet1}, Cxxc4 and GFP, n = 3; Cxxc5, n = 2.
doi:10.1371/journal.pone.0062755.g005

fusions. Cxxc4 and Cxxc5 coding sequences were also inserted into pCAG-Cherry-IB to generate N-terminal Cherry fusions. All constructs were verified by DNA sequencing and their expression by western blotting (Fig. S9).

Northern blotting, cDNA synthesis and qPCR

Total RNA was extracted using the NucleoSpin Triprep Kit and the poly(A)$^+$ fraction was enriched with the Nucleotrap mRNA Mini kit (both from Macherey-Nagel). Northern blotting was performed according to the DIG Application Manual for Filter Hybridization (Roche). Probes were generated and labeled by PCR using DIG-dUTP and primers listed in Table S2. Ten micrograms each of total RNA from ESCs and NSCs were separated on formaldehyde-agarose gels, transferred to Hybond-N+ nylon membranes (GE healthcare) and immobilized by UV crosslinking. Blots were prehybridized with DIG Easy hyb (Roche) at 50°C for 30 min followed by overnight hybridization at 50°C. Probes were applied at a final concentration of 100 ng/ml in DIG Easy hyb. After washing, the blots were incubated with blocking solution (Roche) for 30 min, followed by incubation with alkaline phosphatase conjugated anti-digoxygenin antibody (Roche) for 30 min at room temperature. The membrane was washed twice, equilibrated with detection buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5) and chemiluminescence with CDP-Star substrate (Roche) was used to detect the bound antibody.

Tissue samples were prepared from 6 week old 129Sv and 30 week old C57BL/6 mice (see legend to Fig. 3 for details). Total RNA (500 ng) was reverse transcribed with High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instruction. Primers for conventional PCR indicated in Fig. 2A,B are listed in Table S2. Real-time PCR was

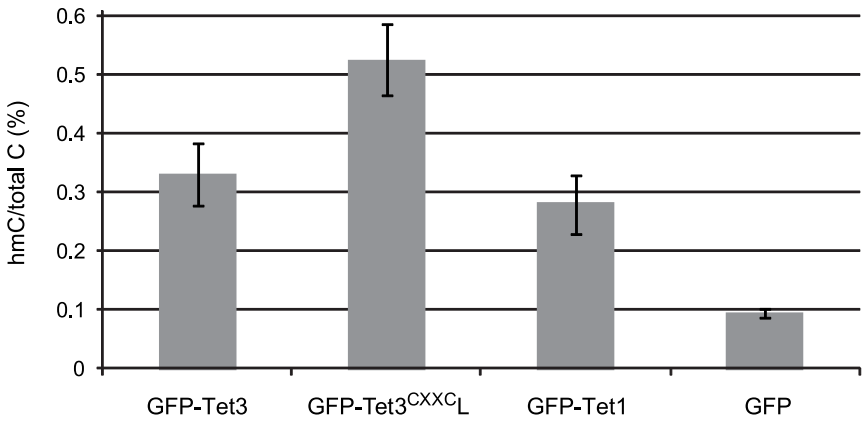


Figure 6. Tet3^{CXXC} oxidizes genomic mC *in vivo*. GFP or GFP-Tet fusions were transiently overexpressed in HEK293T cells and genomic hmC levels were determined using an *in vitro* glucosylation assay with T4 β-glucosyltransferase and UDP-[^{3}H]glucose. Shown are mean percentages and SEM of hmC over total C from 2 (GFP-Tet1) or 3 (all others) independent transfections.
doi:10.1371/journal.pone.0062755.g006

performed using Power SYBR Green PCR Master Mix (Applied Biosystems) on a 7500 Fast Real-Time PCR System (Applied Biosystems) with primers listed in Table S3. Glyceraldehyde phosphate dehydrogenase (GAPDH) was used for normalization and the comparative CT method was used to analyze expression data.

5′ RACE

5′ RACE was performed as described [63] and primers are listed in Table S2. Briefly, 100 ng of total RNA from ENC1 NSCs were reverse transcribed as described above, but using the gene-specific primer1 (GSP1). To remove excess primer, the reaction was purified with a silica mini-column (Nucleospin Gel and PCR Clean-up; Macherey-Nagel). After tailing with terminal deoxynucleotide transferase and dATP the tailed cDNA was subjected to nested PCR reactions with Phusion High-Fidelity DNA Polymerase (New England Biolabs). In the first reaction the upstream primers were $(dT)_{17}$-adaptor primer and adaptor primer, while the downstream primer was gene-specific primer2 (GSP2). Cycling parameters were as follows: one cycle of 98°C for 30 s, 94°C for 5 min, 50°C for 5 min, and 72°C for 40 min, followed by 30 cycles of 94°C for 40 s, 54°C for 1 min, and 72°C for 3 min, with a final cycle of 94°C for 40 s, 54°C for 1 min, and 72°C for 15 min. In the second reaction the upstream primer was adaptor primer and the downstream primer was gene specific primer 3 (GSP3). Cycling parameters were as follows: 98°C for 30 s, (98°C for 15 s, 55°C for 20 s, and 72°C for 30 s) 30 cycles, 72°C for 10 min. PCR products were purified by gel electrophoresis followed by silica column purification, cloned into pCR-Blunt with Zero Blunt PCR Cloning Kit (Invitrogen) and analyzed by sequencing.

F3H assay

F3H assay (Fig. S3) was performed as described [33]. Briefly, BHK cells with a stably integrated *lac* operator array [61] were seeded on coverslips, cotransfected with GFP binding protein (GBP)-lacI, GFP-bait and Ch-prey constructs, fixed and imaged 16 h after transfection.

In vitro DNA binding assay

In vitro DNA binding assays were performed as described previously [24,38,39]. Briefly, two or three double stranded DNA oligonucleotides labeled with different ATTO fluorophores were used as substrates in direct competition. DNA oligonucleotide substrates with identical sequence contained an unmodified, symmetrically methylated or symmetrically hydroxymethylated cytosine at a single, central CpG site (CG, mCG and hmCG substartes), while the noCG substrate contained a TpG site at the same position and had otherwise the same sequence (Tables S4, S5, and S6). GFP fusion constructs were expressed in HEK293T cells by transient transfection and immunopurified from cell lysates using the GFP-trap (ChromoTek). GFP-trap beads were washed three times before incubating with DNA substrates at a final concentration of 160 nM each. After removal of unbound substrates, protein amounts (GFP fluorescence) and bound DNA were measured with an Infinite M1000 plate reader (Tecan).

Determination of global genomic hmC levels

Global hmC levels in genomic DNA from transiently transfected HEK293T cells were determined by the *in vitro* glucosylation assay as described previously [11,24] with minor modifications. Briefly, 50 µl reactions containing 150 mM NaCl, 20 mM Tris, pH 8.0, 25 mM CaCl2, 1 mM DTT, 3.5 µM UDP-[^{3}H]glucose (20 Ci/mmol; Hartmann Analytic GmbH), 500 ng of sheared genomic DNA and 40 nM recombinant T4 β-glucosyltransferase were incubated for 20 min at room temperature and terminated by heating at 65°C for 10 min. DNA fragments were purified by silica column chromatography (Nucleospin, Macherey-Nagel) and radioactivity was determined by liquid scintillation. Radioactive counts were converted to percentages of hmC over total C using curves from PCR generated standards containing variable hmC/C ratios as previously described [11]. The values for all GFP-Tet constructs were corrected for differences in expression levels using GFP-fluorescence measurements. This correction was not applied to control samples transfected with GFP as the latter is expressed at least at ten times higher levels than GFP-Tet1 constructs, which would lead to artificially enhanced differences between basal hmC levels and those resulting by overexpression of Tet constructs.

Supporting Information

Figure S1 Expression of NSCs markers in ENC1 cells. Epifluorescence images of immunofluorescent stainings with antibodies to the indicated markers. Antibody sources: Nestin, mouse monoclonal antibody Rat-401 (Developmental Studies Hybridoma Bank, University of Iowa); Pax6, rabbit polyclonal antibody (PRB-278P, Covance). Olig2, rabbit polyclonal antibody (AB9610, Millipore). Scale bars: 10 µm.
(EPS)

Figure S2 Northern blot analysis of Tet3 and Tet3^{CXXC}L transcripts in NSCs and ESCs (related to Fig. 2). On the right the same blot as in Fig. 2D is shown uncropped. In this blot total RNA was loaded [without poly(A)$^+$ enrichment], resulting in stronger crosshybridization with 28S and 18S ribosomal RNAs.
(EPS)

Figure S3 Schematic representation of the mammalian F3H assay (related to Fig. 4).
(EPS)

Figure S4 Nuclear localization of GFP-Cxxc4 in C2C12 myoblasts and CGR8 ESCs (related to Fig. 4). Epifluorescence images of transiently transfected cells. Scale bars: 5 µm.
(EPS)

Figure S5 Transcript levels of Cxxc4, Cxx5 and Tet1–3 in adult mouse tissues ESCs and NSCs (related to Fig. 3). In (A) the same plot as in Fig. 3B is reported for ease of comparison between transcript levels of Cxxc4/5 (A) and Tet1–3 (B). In (B) cumulative levels of all Tet3 transcripts were determined using a primer set spanning common sequences downstream exon 3 of the annotated *Tet3* gene. Shown are mean values and SEM. Sample sources and replicates are as for Fig. 3.
(EPS)

Figure S6 *In vitro* DNA binding properties of GFP-Tet1$^{512–671}$, GFP-CXXC^{Tet1} and CXXC^{Tet1}-GFP. (A) Schematic representation of assayed Tet1 constructs. Start and end positions relative to full length Tet1 protein are reported. (B) DNA binding assay as in Fig. 5. Shown are mean values and SEM from 4 independent experiments.
(EPS)

Figure S7 *In vitro* binding of various full length Cxxc domain-containing proteins and isolated CXXC domains to DNA substrates containing one or no CG site), but otherwise identical sequence (related to Fig. 5). All constructs are GFP fusions. Shown are mean values of bound substrate/protein ratios and SEM from n independent replicate experiments: GFP and CXXC^{Tet3}-GFP, n = 5; GFP-Tet1, Cxxc4-

GFP, Cxxc5-GFP and GFP-CXXC^{Dnmt1}, n = 4; GFP-Tet3, GFP-Tet3^{CXXC}L and CXXC^{Tet1}-GFP, n = 3.
(EPS)

Figure S8 Localization and mobility of Tet3 and Tet3^{CXXC}L isoforms in C2C12 nuclei. (A) Optical sections of fixed C2C12 cotransfected with GFP-Tet3^{CXXC}L and Ch-Tet3 constructs as indicated. Arrowheads indicate the position of large chromocenters from which GFP-Tet3^{CXXC}L and Ch-Tet3 signals are excluded. (B) FRAP curves of GFP-Tet3 and GFP-Tet3^{CXXC}L in transiently transfected C2C12 myoblasts. Images were taken every 150 ms in the first 60 s, and then at intervals of 1 s for the next 120 s. Shown are mean values and SEM from 12 (GFP-Tet3) and 10 cells (GFP-Tet3^{CXXC}L). Live cell imaging and FRAP analysis was performed as described (Schermelleh et al., 2007, Nucl Acids Res 35: 4301) with the following minor modifications. The images were Gauss-filtered (2 pixel radius) and data sets showing lateral movement were corrected by image registration using the StackReg plug-in of ImageJ, starting with a time frame where approximately half recovery was reached.
(EPS)

Figure S9 Western blot analysis of fluorescent fusion proteins. (A) GFP-CXXC^{Dnmt1}, CXXC^{Tet3}-GFP, CXXC^{Tet1}-GFP, Cxxc4-GFP, Cxxc5-GFP. (B) GFP-Cxx4 and GFP-Cxxc5. (C) GFP-Tet1, GFP-Tet3 and GFP-Tet3^{CXXC}L. (D) Cherry-Tet3. Blots were probed with an anti-GFP antibody (A–C) or with an anti-RFP antibody recognizing an epitope present in both RFP and Cherry (D). In all cases the major reacting band migrated as a peptide with the expected mass of the specific, full length fluorescence fusion and in no case peptides with mass corresponding to the fluorescent protein moiety (GFP or Cherry) were detected.
(EPS)

Table S1 Primer sequences for cloning of coding sequences in expression constructs.
(DOCX)

Table S2 Primer sequences for 5′ RACE, conventional RT-PCR, northern blotting probes.
(DOCX)

Table S3 Primer sequences for qPCR.
(DOCX)

Table S4 Sequences of oligonucleotides used for preparation of double stranded DNA substrates.
(DOCX)

Table S5 CG, mCG and hmCG containing DNA substrates used for *in vitro* binding assay (related to Fig. 5).
(DOCX)

Table S6 CG and noCG containing DNA substrates used for *in vitro* binding assay (related to Fig. S7).
(DOCX)

Combined Supporting Information File S1
(PDF)

Acknowledgments

We are grateful to Adrian Bird (Wellcome Trust Centre for Cell Biology, Edinburgh) and Haruiko Koseki (RIKEN Research Center for Allergy and Immunology, Yokohama) for CGR8 and E14 ESC lines, respectively, and to Bijàn Montazeri (LMU Munich) for discussions.

Author Contributions

Conceived and designed the experiments: FS HL. Performed the experiments: NL MW FS WD CSS. Analyzed the data: NL MW FS WD CSS WQ. Wrote the paper: FS. Acquisition of funding: HL. Final approval: FS HL.

References

1. Branco MR, Ficz G, Reik W (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet 13: 7–13.
2. Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol 11: 607–620.
3. Niehrs C, Schäfer A (2012) Active DNA demethylation by Gadd45 and DNA repair. Trends in Cell Biology 22: 220–227.
4. 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.
5. 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.
6. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. (2011) Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science 333: 1300–1303.
7. He Y-F, Li B-Z, Li Z, Liu P, Wang Y, et al. (2011) Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science 333: 1303–1307.
8. Szulwach KE, Li X, Li Y, Song C-X, Wu H, et al. (2011) 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat Neurosci 14: 1607–1616.
9. Sérandour AA, Avner S, Oger F, Bizot M, Percevault F, et al. (2012) Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucl Acids Res 40: 8255–8265.
10. Doege CA, Inoue K, Yamashita T, Rhee DB, Travis S, et al. (2012) Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature 488: 652–655.
11. 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.
12. Münzel M, Globisch D, Bruckl T, Wagner M, Welzmiller V, et al. (2010) Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. Angew Chem Int Ed Engl 49: 5375–5377.
13. Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, et al. (2010) Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. PLoS ONE 5: e15367.
14. Langemeijer SMC, Kuiper RP, Berends M, Knops R, Aslanyan MG, et al. (2009) Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet 41: 838–842.
15. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, et al. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468: 839–843.
16. Konstandin N, Bultmann S, Szwagierczak A, Dufour A, Ksienzyk B, et al. (2011) Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. Leukemia 25: 1649–1652.
17. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, et al. (2010) Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. Cancer Cell 18: 553–567.
18. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PAC, et al. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473: 343–348.
19. Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, et al. (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 473: 389–393.
20. Yildirim O, Li R, Hung J-H, Chen PB, Dong X, et al. (2011) Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. Cell 147: 1498–1510.
21. Kallin EM, Rodríguez-Ubreva J, Christensen J, Cimmino L, Aifantis I, et al. (2012) Tet2 Facilitates the Derepression of Myeloid Target Genes during CEBPα-Induced Transdifferentiation of Pre-B Cells. Molecular Cell 48: 266–276.
22. Gu T-P, Guo F, Yang H, Wu H-P, Xu G-F, et al. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477: 606–610.
23. Mancini M, Veljkovic N, Leo E, Aluigi M, Borsi E, et al. (2012) Cytoplasmatic compartmentalization by BCR-ABL promotes TET2 loss-of-function in chronic myeloid leukemia. Journal of Cellular Biochemistry 113: 2765–2774.
24. 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: e16627.

25. Hino S, 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.
26. Michiue T, Fukui A, Yukita A, Sakurai H, Danno H, et al. (2004) XIdax, an inhibitor of the canonical Wnt pathway, is required for anterior neural structure formation in Xenopus. Developmental Dynamics 230: 79–90.
27. Andersson T, Södersten E, Duckworth JK, Cascante A, Fritz N, et al. (2009) CXXC5 Is a Novel BMP4-regulated Modulator of Wnt Signaling in Neural Stem Cells. Journal of Biological Chemistry 284: 3672–3681.
28. Williams K, Christensen J, Helin K (2011) DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep 13: 28–35.
29. Tan L, Shi YG (2012) Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development 139: 1895–1902.
30. Katoh M, Katoh M (2004) Identification and characterization of human CXXC10 gene in silico. Int J Oncol 25: 1193–1199.
31. Zolghadr K, Mortusewicz O, Rothbauer U, Kleinhans R, Goehler H, et al. (2008) A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells. Mol Cell Proteomics 7: 2279–2287.
32. Meilinger D, Fellinger 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 reports 10: 1259–1264.
33. Dambacher S, Deng W, Hahn M, Sadic D, Fröhlich J, et al. (2012) CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3: 101–110.
34. Itoh K, Brott BK, Bae G-U, Ratcliffe MJ, Sokol SY (2005) Nuclear localization is required for Dishevelled function in Wnt/β-catenin signaling. Journal of Biology 4: 3.
35. Jans DA, Xiao C-Y, Lam MHC (2000) Nuclear targeting signal recognition: a key control point in nuclear transport? BioEssays 22: 532–544.
36. Pendino F, Nguyen E, Jonassen I, Dysvik B, Azouz A, et al. (2009) Functional involvement of RINF, retinoid-inducible nuclear factor (CXXC5), in normal and tumoral human myelopoiesis. Blood 113: 3172–3181.
37. Xu Y, Wu F, Tan L, Kong L, Xiong L, et al. (2011) Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Molecular Cell 42: 451–464.
38. Frauer C, Leonhardt H (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. Nucl Acids Res 37: e22–e22.
39. 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: e21306.
40. Pichler G, Wolf P, Schmidt CS, Meilinger D, Schneider K, et al. (2011) Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. Journal of Cellular Biochemistry 112: 2585–2593.
41. Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, et al. (2009) The Multi-Domain Protein Np95 Connects DNA Methylation and Histone Modification. Nucl Acids Res 38: 1796–1804.
42. Pradhan M, Estève P-O, Chin HG, Samaranayke M, Kim G-D, et al. (2008) CXXC Domain of Human DNMT1 Is Essential for Enzymatic Activity. Biochemistry 47: 10000–10009.
43. Lee J-H, Voo KS, Skalnik DG (2001) Identification and Characterization of the DNA Binding Domain of CpG-binding Protein. J Biol Chem 276: 44669–44676.
44. Birke M, Schreiner S, García-Cuéllar M-P, 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.
45. Jørgensen 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.
46. Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, et al. (2010) CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. Molecular Cell 38: 179–190.
47. Erfurth FE, Popovic R, Grembecka J, Cierpicki T, Theisler C, et al. (2008) MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. PNAS 105: 7517–7522.
48. 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.
49. Xu Y, Xu C, Kato A, Tempel W, Abreu JG, et al. (2012) Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for Xenopus Eye and Neural Development. Cell 151: 1200–1213.
50. Torres MA, Nelson WJ (2000) Colocalization and Redistribution of Dishevelled and Actin during WNT-Induced Mesenchymal Morphogenesis. J Cell Biol 149: 1433–1442.
51. Gan X, Wang J, Xi Y, Wu Z, Li Y, et al. (2008) Nuclear Dvl, c-Jun, β-catenin, and TCF form a complex leading to stabiLization of β-catenin–TCF interaction. J Cell Biol 180: 1087–1100.
52. Kim W, Kim H, Katanaev VL, Lee SJ, Ishitani T, et al. (2012) Dual functions of DP1 promote biphasic Wnt-on and Wnt-off states during anteroposterior neural patterning. The EMBO Journal 31: 3384–3397.
53. Huang T, Xie Z, Wang J, Li M, Jing N, et al. (2011) Nuclear Factor of Activated T Cells (NFAT) Proteins Repress Canonical Wnt Signaling via Its Interaction with Dishevelled (Dvl) Protein and Participate in Regulating Neural Progenitor Cell Proliferation and Differentiation. J Biol Chem 286: 37399–37405.
54. London TBC, Lee H-J, Shao Y, Zheng J (2004) Interaction between the internal motif KTXXXI of Idax and mDvl PDZ domain. Biochemical and Biophysical Research Communications 322: 326–332.
55. Handyside AH, O'Neill GT, Jones M, Hooper ML (1989) Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem cell line. Development Genes and Evolution 198: 48–56.
56. Mountford P, Zevnik B, Düwel A, Nichols J, Li M, et al. (1994) Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression. PNAS 91: 4303–4307.
57. Conti L, Pollard SM, Gorba T, Reitano E, Toselli M, et al. (2005) Niche-Independent Symmetrical Self-Renewal of a Mammalian Tissue Stem Cell. PLoS Biology 3: e283.
58. Nichols J, Ying Q-L (2006) Derivation and propagation of embryonic stem cells in serum- and feeder-free culture. In: Turksen K, editor. Embryonic Stem Cell Protocols. Methods in Molecular Biology. Humana Press, Vol. 329. pp. 91–98.
59. Blau HM, Pavlath GK, Hardeman EC, Chiu CP, Silberstein L, et al. (1985) Plasticity of the differentiated state. Science 230: 758–766.
60. 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.
61. Tsukamoto T, Hashiguchi N, Janicki SM, Tumbar T, Belmont AS, et al. (2000) Visualization of gene activity in living cells. Nature Cell Biology 2: 871–878.
62. Schermelleh L, Haemmer A, Spada F, Rösing 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.
63. Nature Publishing Group (2005) Rapid amplification of 5′ complementary DNA ends (5′ RACE). Nature Methods 2: 629–630.
64. Okonechnikov K, Golosova O, Fursov M (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28: 1166–1167.

Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules

Nan Liu, Mengxi Wang, Wen Deng, Christine S. Schmidt, Weihua Qin, Heinrich Leonhardt and Fabio Spada

Department of Biology II, Ludwig Maximilians University Munich, Planegg-Martinsried, Germany.

SUPPORTING INFORMATION FILE

Figures S1-9.

Tabels S1-6.

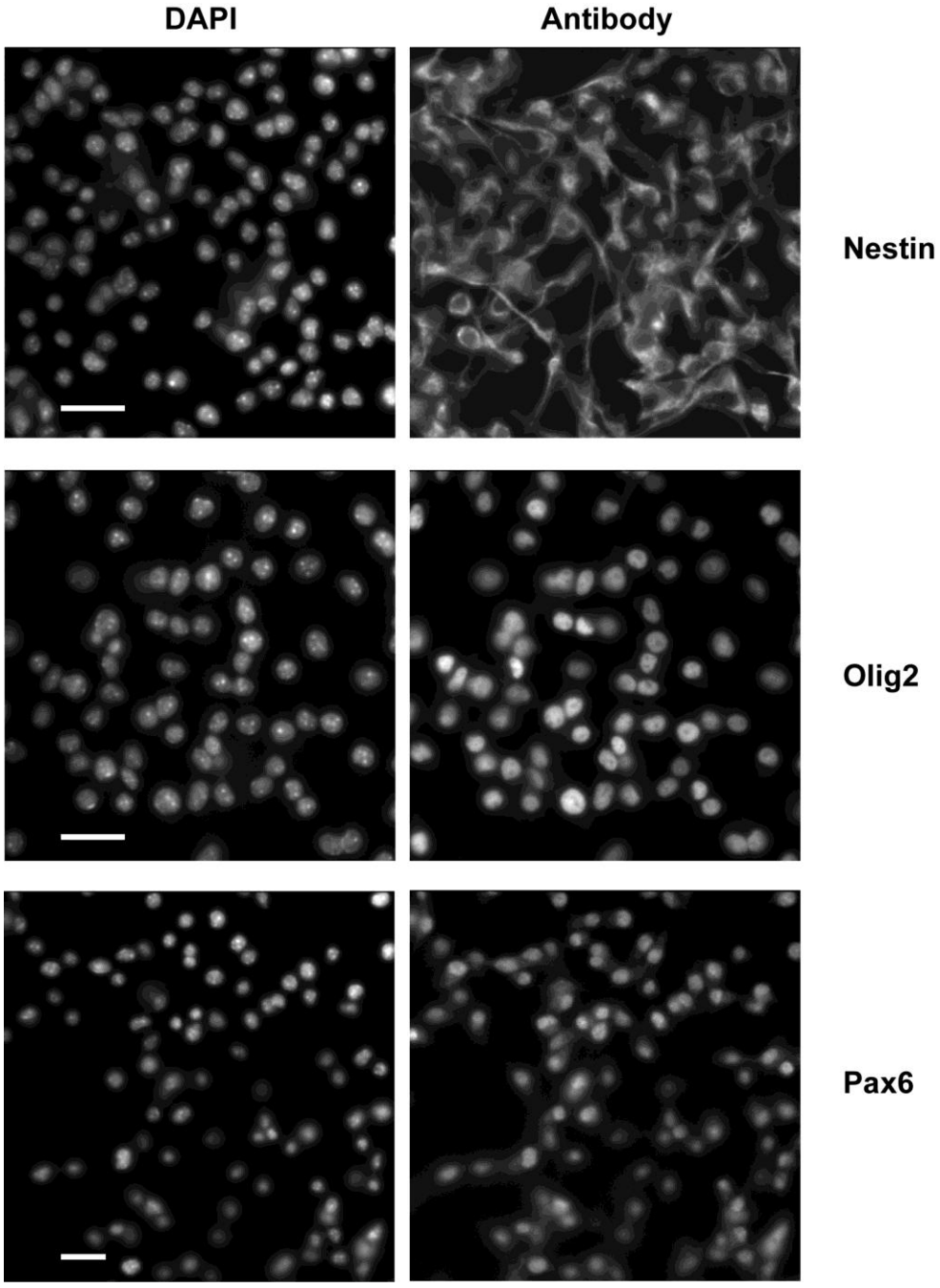


Figure S1. Expression of NSCs markers in ENC1 cells. Epifluorescence images of immunofluorescent stainings with antibodies to the indicated markers. Antibody sources: Nestin, mouse monoclonal antibody Rat-401 (Developmental Studies Hybridoma Bank, University of Iowa); Pax6, rabbit polyclonal antibody (PRB-278P, Covance). Olig2, rabbit polyclonal antibody (AB9610, Millipore). Scale bars: 10 µm.

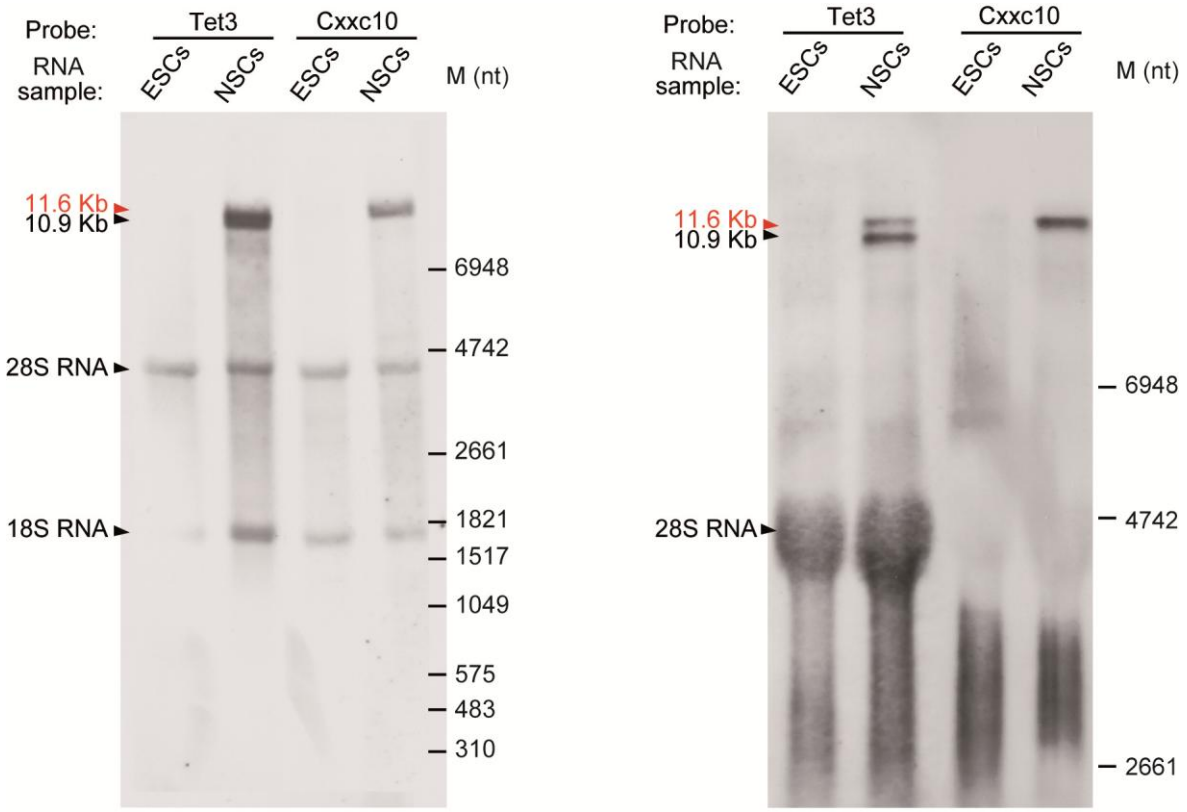


Figure S2. Northern blot analysis of Tet3 and Tet3^{CXXC}L transcripts in NSCs and ESCs (related to Fig. 2). On the right the same blot as in Fig. 2D is shown uncropped. In this blot total RNA was loaded [without poly(A)$^+$ enrichment], resulting in stronger crosshybridization with 28S and 18S ribosomal RNAs.

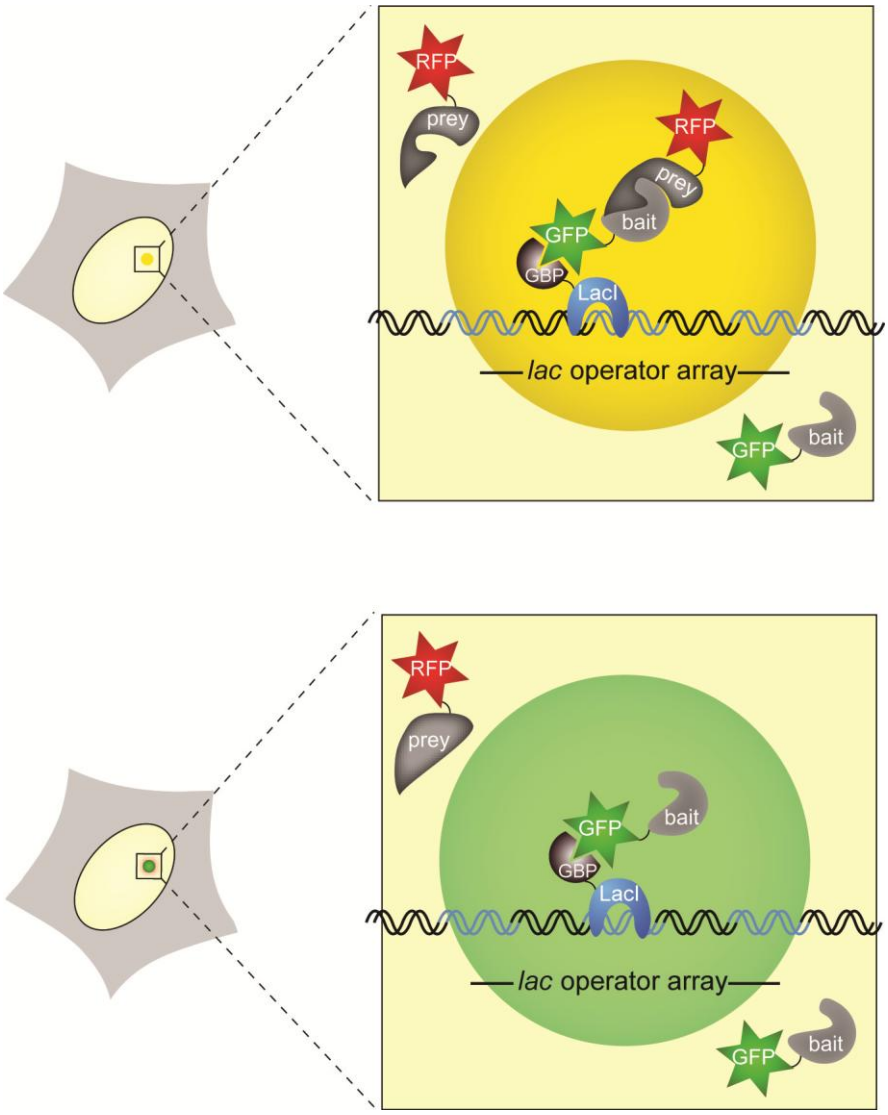


Figure S3. Schematic representation of the mammalian F3H assay (related to Fig. 4).

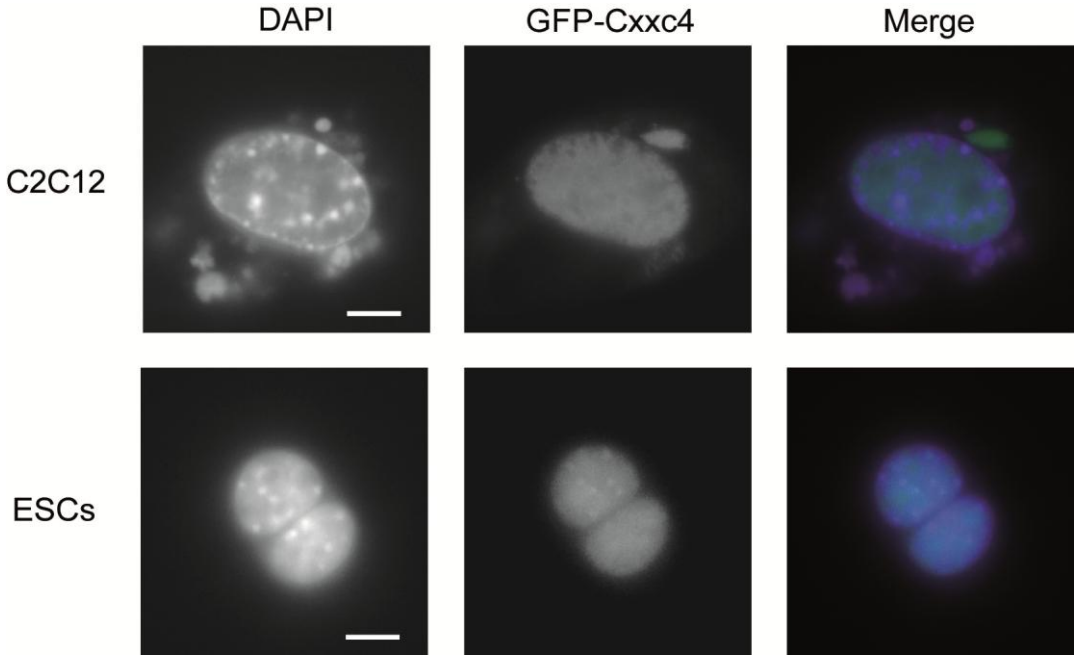


Figure S4. Nuclear localization of GFP-Cxxc4 in C2C12 myoblasts and CGR8 ESCs (related to Fig. 4). Epifluorescence images of transiently transfected cells. Scale bars: 5 µm.

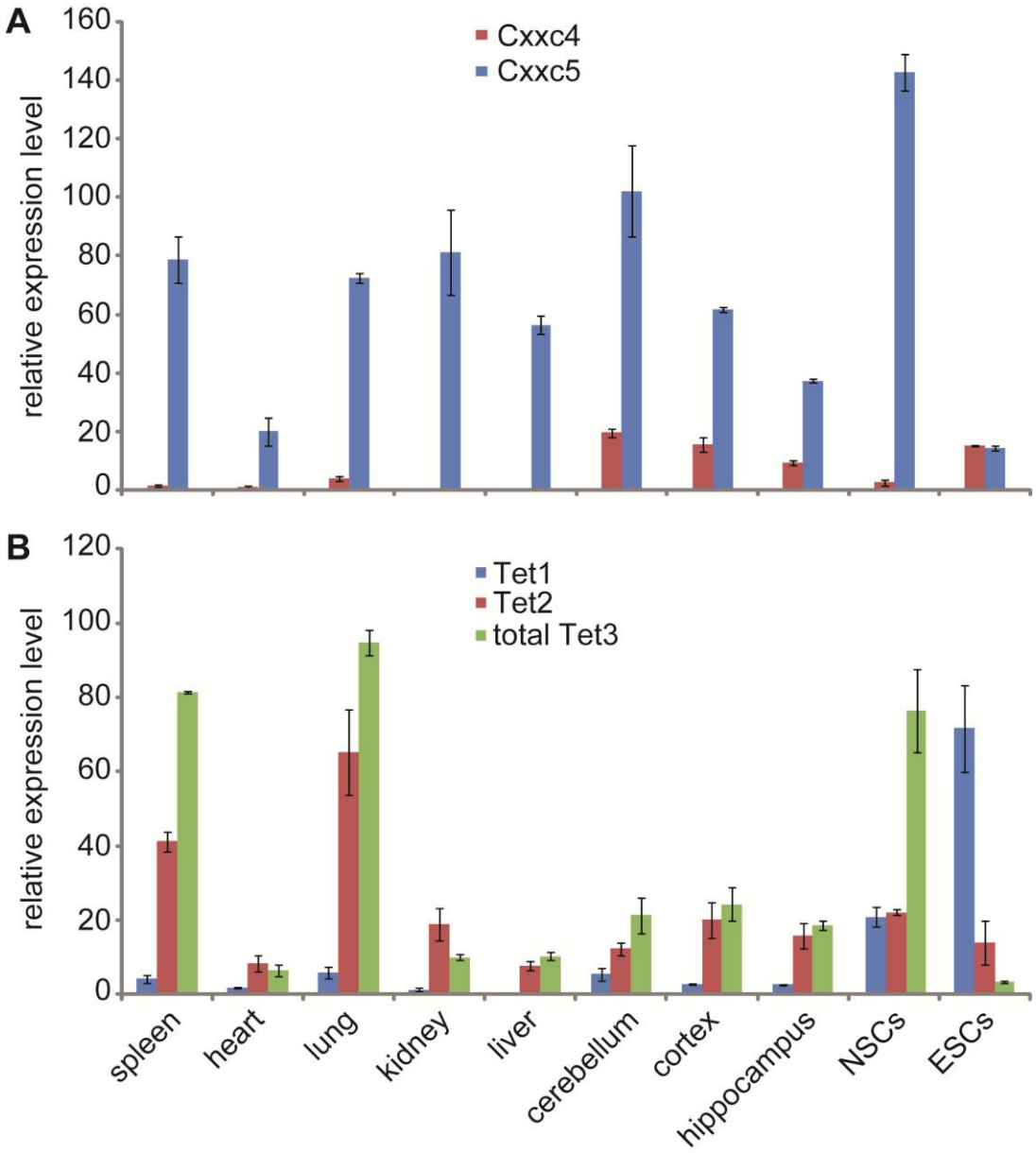


Figure S5. Transcript levels of Cxxc4, Cxx5 (A) and Tet1-3 (B) in adult mouse tissues ESCs and NSCs (related to Fig. 3). In (A) the same plot as in Fig. 3B is reported for ease of comparison between transcript levels of Cxxc4/5 (A) and Tet1-3 (B). In (B) cumulative levels of all Tet3 transcripts were determined using a primer set spanning common sequences downstream exon 3 of the annotated *Tet3* gene. Shown are mean values and SEM. Sample sources and replicates are as for Fig. 3.

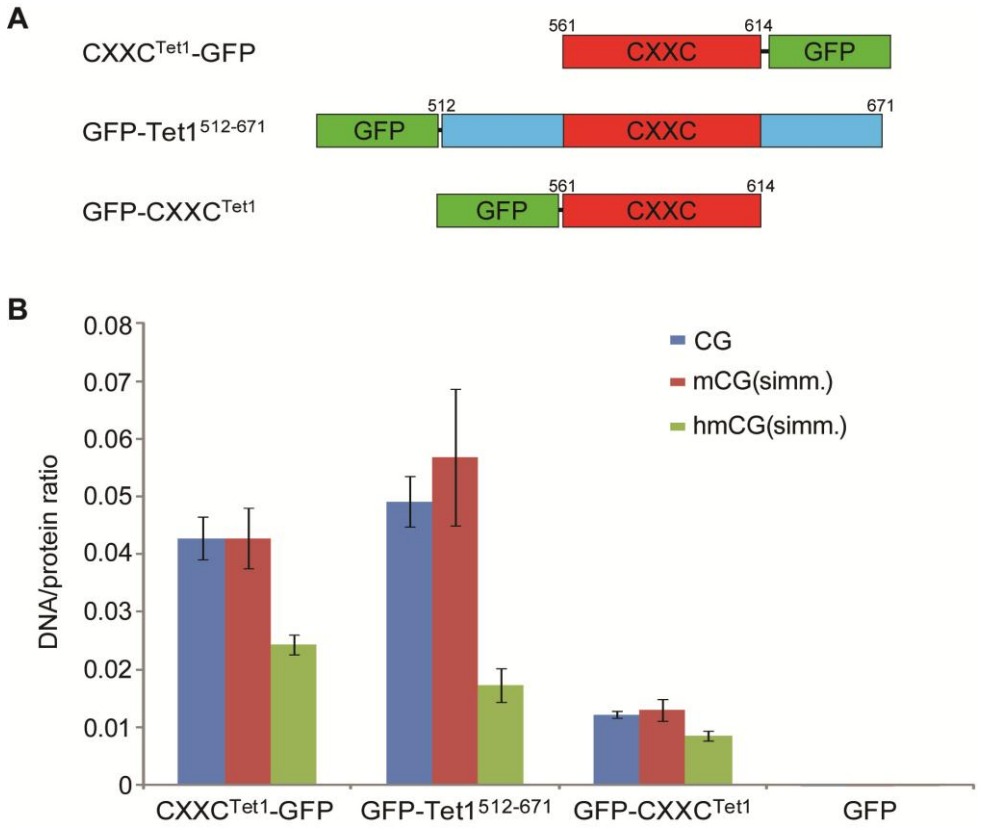


Figure S6. *In vitro* DNA binding properties of GFP-Tet1$^{512-671}$, GFP-CXXC^{Tet1} and CXXC^{Tet1}-GFP. (A) Schematic representation of assayed Tet1 constructs. Start and end positions relative to full length Tet1 protein are reported. (B) DNA binding assay as in Fig. 5. Shown are mean values and SEM from 4 independent experiments.

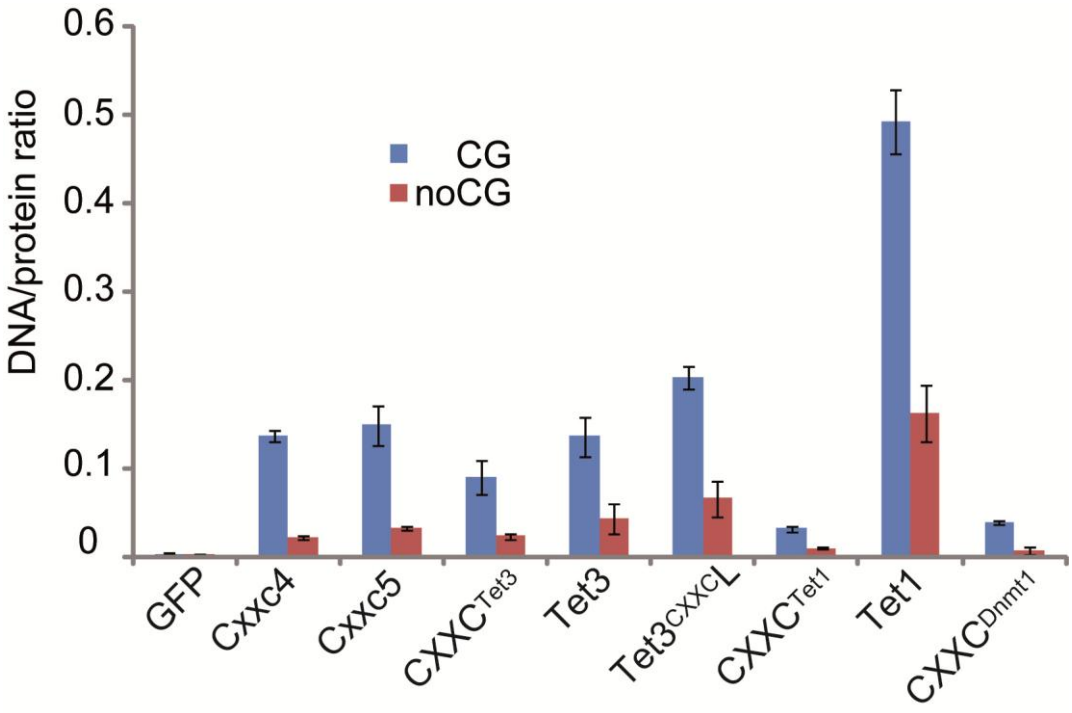


Figure S7. *In vitro* binding of various full length CXXC domain-containing proteins and isolated CXXC domains to DNA substrates containing one or no CpG site (noCG; TpG instead of CpG), but otherwise identical sequence (related to Fig. 5). All constructs are GFP fusions. Shown are mean values of bound substrate/protein ratios and SEM from n independent replicate experiments: GFP and $CXXC^{Tet3}$-GFP, n=5; GFP-Tet1, Cxxc4-GFP, Cxxc5-GFP and GFP-$CXXC^{Dnmt1}$, n=4; GFP-Tet3, GFP-$Tet3^{CXXC}$L and $CXXC^{Tet1}$-GFP, n=3.

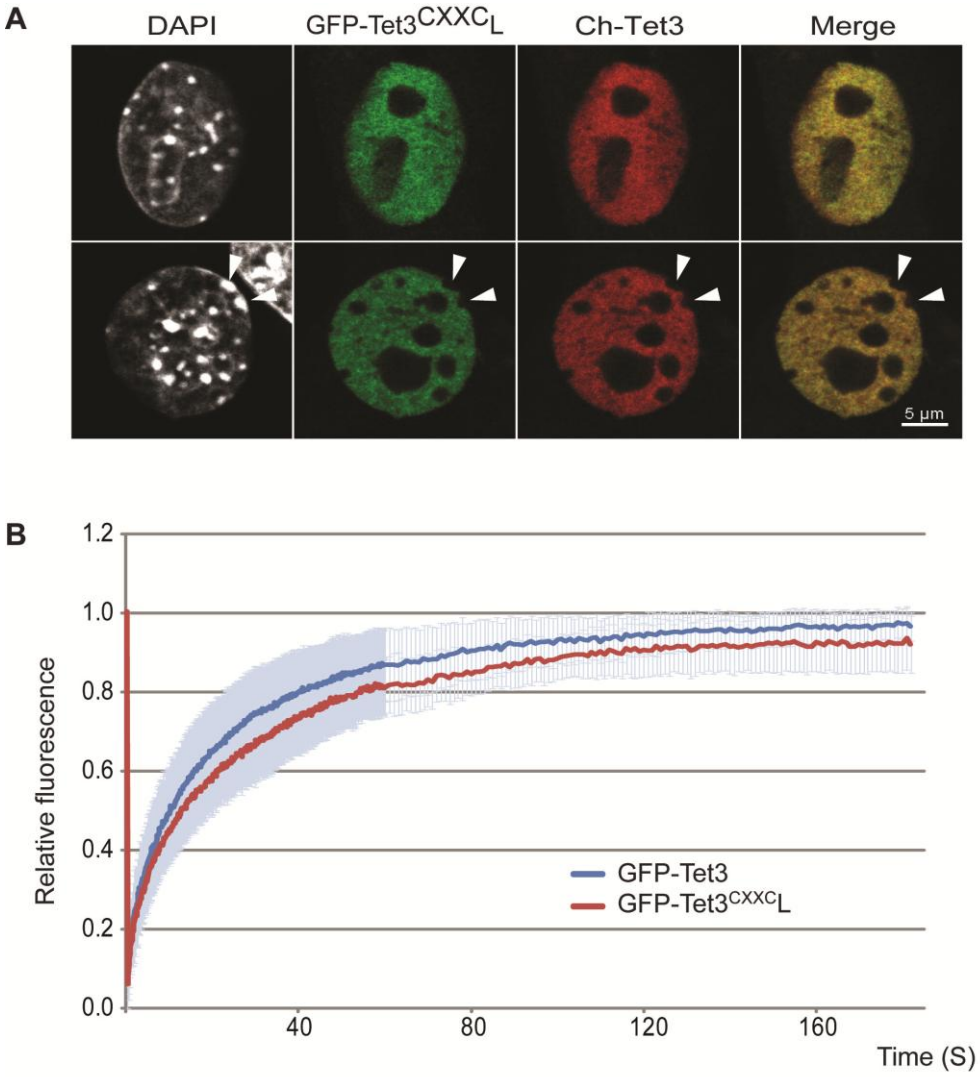


Figure S8. Localization (A) and mobility (B) of Tet3 and Tet3^{CXXC}L isoforms in C2C12 nuclei. (A) Optical sections of fixed C2C12 cotransfected with GFP-Tet3^{CXXC}L and Ch-Tet3 constructs as indicated. Arrowheads indicate the position of large chromocenters from which GFP-Tet3^{CXXC}L and Ch-Tet3 signals are excluded. (B) FRAP curves of GFP-Tet3 and GFP-Tet3^{CXXC}L in transiently transfected C2C12 myoblasts. Images were taken every 150 ms in the first 60 s, and then at intervals of 1 s for the next 120 s. Shown are mean values and SEM from 12 (GFP-Tet3) and 10 cells (GFP-Tet3^{CXXC}L). Live cell imaging and FRAP analysis was performed as described (Schermelleh et al., 2007, Nucl Acids Res 35: 4301) with the following minor modifications. The images were Gauss-filtered (2 pixel radius) and data sets showing lateral movement were corrected by image registration using the StackReg plug-in of ImageJ, starting with a time frame where approximately half recovery was reached.

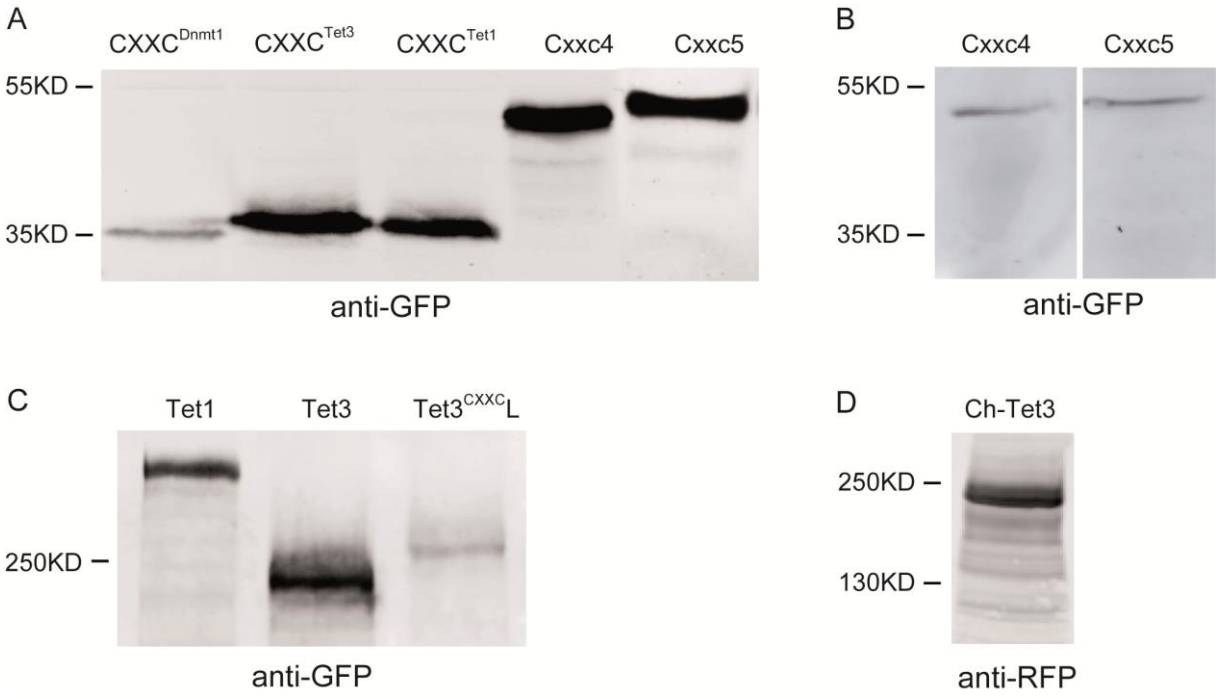


Figure S9. Western blot analysis of fluorescent fusion proteins. (A) GFP-CXXC^{Dnmt1}, CXXC^{Tet3}-GFP, CXXC^{Tet1}-GFP, Cxxc4-GFP, Cxxc5-GFP. (B) GFP-Cxx4 and GFP-Cxxc5. (C) GFP-Tet1, GFP-Tet3 and GFP-Tet3^{CXXC}L. (D) Cherry-Tet3. Blots were probed with an anti-GFP antibody (A-C) or with an anti-RFP antibody recognizing an epitope present in both RFP and Cherry (D). In all cases the major reacting band migrated as a peptide with the expected mass of the specific, full length fluorescence fusion and in no case peptides with mass corresponding to the fluorescent protein moiety (GFP or Cherry) were detected.

Table S1. Primers for amplification and insertion of coding sequences in expression constructs.

Construct	Primer
$Tet1^{512-671}$	5'-AAG CGA TCG CTT AGA TCT TAC CCA GGG-3'
	5'-TTG CGG CCG CCA AAT CCA ACC TTT GC-3'
$CXXC^{Tet1}$	5'-GGC GAT CGC ATG TCT ACG CCG CCA ATG-3'
	5'-CGC GGC CGC CTG GCT TCT TTT TGA GCA-3'
Cxxc4	5′-ATG CAC CAC CGG AAC GAC TCC CAG CG-3'
	5'-TTA AAA GAA CCA TCG GAA CGC TTC AGC-3'
Cxxc5	5′-AAG CGA TCG CAT GTC GAG CCT CGG CGG TGG-3′
	5′-GCG CGG CCG CTC ACT GAA ACC ACC GGA AGG-3′
$CXXC^{Tet3}$	5′-ATG CGA TCG CAT GCT GCG AGG GGG TGG AGA T-3′
	5′-ATG CGG CCG CCC GCT TTT TTC TTC AGC ACC TC-3′
$Tet3^{CXXC}L$	5′-GGG CGA TCG CAT GAG CCA GTT TCA GGT GCC CTT GG-3′
	5′-GCG GCC GCC TAG ATC CAG CGG CTG TAG GGG CC-3′

Table S2. Primer sequences for 5′ RACE, conventional RT-PCR (primers a-d indicated in Fig. 2A,B) and generation of probes for northern blotting.

Name	Sequence
GSP1	5′ -AGG TCC ATC AAC TGG GCT-3′
$(dT)_{17}$-adaptor	5′-GAC TCG AGT CGA CAT CGA $(T)_{17}$-3′
adaptor primer	5′-GAC TCG AGT CGA CAT CG-3′
GSP2	5′-AGC ACC TCA CAC TTG CG-3′
GSP3	5′-GCA GCT GGT ACA AGA CC-3′.
Primer a	5′- GCG ATC GCA TGA GCC AGT TTC AGG -3′
Primer c	5′- AAG CGG CCG CCA GTC GGG CTT CTG GTC TAC -3′
Primer b	5′- ATG GCT GGG AGT GAG AC -3′
Primer d	5′- ATC GCA GGT GCA GTT GGG TG -3′
CXXC10 probe for	5′-CAC ACC CAT TGG CTC ACC T-3′
CXXC10 probe rev	5′-GGG TCT CAC TCC CAG CCA-3′
Tet3 probe for	5′-GCT CTC AAC TAC CTG CTT CC-3′
Tet3 probe rev	5′-CAT TGA GGC CAC ATC TCC G-3′

Table S3. Primer sequences for Real-time PCR.

Name	Sequence
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′
Total Tet3 forward*	5′-GAG CAC GCC AGA GAA GAT CAA-3′
Total Tet3 reverse*	5′-CAG GCT TTG CTG GGA CAA TC-3′
Cxxc4 forward	5′-ACC TGG CAC TTC GCT AGA GAG A-3′
Cxxc4 reverse	5′-TTG CCC TTC ATT CCC AAA TG-3′
Cxxc5 forward	5′-CAG CAG TTG TAG GAA CCG AAA GA-3′
Cxxc5 reverse	5′-TCC CGA CGG AAG CAT CAC-3′
Cxxc10 forward	5′-GTG GAG ATG GGC GGA AGA A-3′
Cxxc10 reverse	5′-GAT CTG GTG TGT GCG ACG AT-3′
$Tet3^{CXXC}L$ forward	5′-ATC GTC GCA CAC ACC AGA TC-3′
$Tet3^{CXXC}L$reverse	5′-TCC TTC ACG AGC ATT TAT TTC CA-3′
Tet3 forward	5′-GCG GCC GAT GCA GTA GTG-3′
Tet3 reverse	5′-ATC AAC TGG GCT GAG CTC TGA-3′

* 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

Table S4. Sequences of oligonucleotides used for preparation of double stranded DNA substrates.

M: 5-methylcytosine X: 5-hydroxymethylcytosine

Name	Sequence
CGup	5'- CTCAACAACTAACTACCATCCGGACCAGAAGAGTCATCATGG -3'
um647N	5'- ATTO647N-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
MGup	5'- CTCAACAACTAACTACCATCMGGACCAGAAGAGTCATCATGG -3'
mC700	5'- ATTO700-CCATGATGACTCTTCTGGTCMGGATGGTAGTTAGTTGTTGAG -3'
hmCGup	5'- CTCAACAACTAACTACCATCXGGACCAGAAGAGTCATCATGG -3'
hmC550	5'- ATTO550-CCATGATGACTCTTCTGGTCXGGATGGTAGTTAGTTGTTGAG -3'
um550	5'- ATTO550-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
um700	5'- ATTO700-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
um590	5'- ATTO590-CCATGATGACTCTTCTGGTCCGGATGGTAGTTAGTTGTTGAG -3'
noCGup	5'- CTCAACAACTAACTACCATCTGGACCAGAAGAGTCATCATGG -3'
noCG647N	5'- ATTO647N-CCATGATGACTCTTCTGGTCTGGATGGTAGTTAGTTGTTGAG -3'

Table S5. CG, mCG and hmCG containing DNA substrates used for *in vitro* binding assay (referes to Fig. 5).

	Name	CpG site	Label	Oligo I	Oligo II
sample set	647N-CG	unmethylated	647N	CGup	um647N
	700-mC	fully methylated	700	MGup	mC701
	550-hmC	fully hydroxymethylated	550	hmCGup	hmC550
control set	647N-CG	unmethylated	647N	CGup	um647N
	550-CG		550		um550
	700-CG		700		um700

Table S6. CG and noCG containing DNA substrates used for *in vitro* binding assay (referes to Fig. S7).

	Name	CpG site	Label	Oligo I	Oligo II
sample set	590-CG	ummethylated	590	CGup	um590
	647N-noCG	no CpG site	647N	noCGup	noCG647N
control set	590-CG	unmethylated	590	CGup	um590
	647N-CG		647N		um647N

2.3 Ogt interacts with all three Tet proteins, mediates their O-GlcNAcylation and links metabolism to epigenetics

Ogt interacts with all three Tet proteins, mediates their O-GlcNAcylation and links metabolism to epigenetics

Andrea Rottach[1]*, Christina Bauer[1], Klaus Göbel[1], Mengxi Wang[1], Udo Müller[1], Elisabeth Kremmer[2], and Heinrich Leonhardt[1]*

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

[2] Helmholtz Center Munich, German Research Center for Environmental Health, Institute of Molecular Immunology, Munich, Germany.

* Corresponding author

Running title: Glycosylation of Tet proteins

Key words: Ogt; Tet1; Tet2; Tet3; O-GlcNAcylation; DNA methylation; DNA hydroxylation

This manuscript contains 5 main figures and no table

Abstract **(203 words and 1479 characters)**

DNA methylation plays a central role in the epigenetic regulation of gene expression during development and disease and was considered as a relatively stable, repressive DNA modification. Recently, Ten-eleven translocation (Tet) proteins have been discovered to convert 5-methylcytosine to 5-hydroxymethylcytosine (5hmC) and by further oxidation to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), likely contributing to an active DNA demethylation process. Using a set of newly generated monoclonal Tet antibodies we performed co-immunoprecipitation assays followed by mass spectrometry analyses and identified several interaction partners. Most prominently, we found O-linked N-acetylglucosamine (GlcNAc) transferase (Ogt), an enzyme that catalyzes the addition of a regulatory GlcNAc to serine or threonine residues. In turn, O-GlcNAcase (Oga) removes O-GlcNAc modifications. Using a fluorescent-3-hybrid (F3H) and biochemical assays, we could confirm the interaction of all three Tet proteins with Ogt and Oga. Fine-mapping experiments revealed the catalytic domain of Tets as major interaction surface with Ogt and Oga. Additionally, we could detect Ogt-mediated glycosylation of all Tet proteins mainly at the N-terminus that can be removed by Oga, indicating the involvement of GlcNAc-cycling in Tet regulation. In summary, our results suggest a link between the new DNA modifications and glycosylation that could possibly alter gene regulation in response to changing environmental conditions such as nutrient availability.

Introduction

DNA methylation plays a central role in the epigenetic regulation of gene expression during development and disease (Bird et al., 2002; Rottach et al., 2009; Reik et al., 2011). Established and maintained by three DNA methyltransferases (Dnmts), cell type and differentiation-specific DNA methylation was considered as a relatively stable, repressive DNA modification. Then, Ten-eleven translocation 1–3 (Tet1–3) proteins have been discovered to convert 5-methylcytosine to 5-hydroxymethylcytosine (5hmC) and by further oxidation steps modify 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), leading to an active DNA demethylation process (Tahiliani et al., 2009; Mohr et al. 2011; Guo et al., 2011; Ito et al., 2011; Tan et al 2012; reviewed in Pastor et al., 2013). Several methods based on either purified enzymes, monoclonal antibodies (mAbs) or mass spectrometry analyses enabled a fast and accurate quantification and mapping of 5hmC in genomic DNA samples of various mouse tissues and differentiating embryonic stem cells (Sun et al., 2013; Szwagierczak et al., 2011; Szwagierczak et al., 2010; Jin et al., 2010). Here, 5hmC levels showed a correlation with differential expression of *tet* genes. Furthermore, sAML patients showed aberrant low hmC levels directly correlated with TET2 and IDH2 mutations (Konstandin et al., 2011).

Although a large set of readers for the Tet oxidation products 5hmC, 5fC and 5caC has been identified (Valinluck et al., 2004; He et al., 2011; Spruijt et al., 2013), only little is known about the regulatory network controlling Tet proteins. That Tet proteins are highly regulated was demonstrated recently in Wang et al, showing a direct effect of calcium-dependent proteases, the calpains, on Tet protein stability (Wang et al., 2014). In addition, Zhang et al. (Zhang et al., 2014) indicated a correlation between human Tet3 restriction to the cytoplasm and reduced 5hmC levels. This effect was glucose dependent. Together with other publications, these data suggest a correlation between metabolism and epigenetic regulation of gene expression (Zhang et al., 2014; Hayakawa et al., 2013).

O-linked N-GlcNAc transferase (Ogt), an enzyme that catalyzes the addition of a regulatory GlcNAc to serine or threonine residues was discovered recently as one potential interactor of Tet proteins (Chen et al 2013; Deplus et al 2013; Vella et al 2013 and Zhang et al 2014). The formation of this highly abundant post-translational modification is linked to nutrient availability and crosstalks with cell cycle progression, gene expression, splicing, chromatin remodeling, pluripotency and reprogramming (Lubas et al. 1997; Kreppel et al 1999; Hanover et al., 2005; Butkinaree et al., 2010; Jang et al., 2012; Krause et al., 2012). In turn, O-GlcNAcase (Oga)

removes O-GlcNAc modifications, making O-linked glycosylation a dynamic process known as O-GlcNAc cycling (Hart et al., 2011; reviewed in Hanover et al. 2012). In mammals, Ogt is present in three different isoforms: nuclear ncOgt, mitochondrial mOgt and a short sOgt (reviewed by Hanover et al., 1987; Lazarus et al., 2011; Holt et al. 1986). This compartmental restriction might also influence GlcNAcylation of different proteins in a spatial and or temporal context (Zhang et al., 2013). In line, ncOgt has been shown to directly glycosylate important nuclear core components like histones or transcriptional regulators including transcription factors (TFs) (e.g. PRCs, Sin3a, HDACs, HCF) and PolIICTDs, whereas the cytoplasmic isoform sOgt modifies mainly proteasomal proteins, pore components and e.g. MAP-kinases (Comer et al., 2000; Zeidan et al., 2010; Ranuncolo et al., 2012; Wysocka et al., 2012; Myers et al 2011; McDonel et al., 2011; Cole et al; reviewed in Hanover et al.,2012; Vella et al., 2013). Besides the three different isoforms of Ogt, the sheer number of potential binding partners reflects the involvement of Ogt in a multitude of cellular pathways.

Functional studies of mOgt displayed its potential in inducing apoptosis and to be involved in the lipid/carbohydrate metabolism. Also, Ogt and O-GlcNAcylation seem to play an important role in onset of neurodegenerative disorders, cardiovascular diseases and insulin resistance (Cole et al., 2001; Lazarus et al., 2009; Fulop et al., 2007; Ngoh et al., 2010; Yang et al., 2008). In addition, the involvement of O-GlcNAc cycling in disease development might be causally linked with the Tet enzymes (Krzeslak et al.; 2012; Freudenberg et al., 2012; Li et al., 2011; Hart et al., 2011). Even though there are several indications to a tight regulatory interplay between Tets and GlcNAc-cycling, the functional relevance of the interaction between Tets and Ogt/Oga and its contribution to epigenetic regulation of gene expression, however, remains unclear.

In this study, we have addressed this question by generating a set of monoclonal antibodies against all three Tet proteins. Using these new tools, we performed co-immunoprecipitation assays of endogenous as well as overexpressed Tet proteins, followed by mass spectrometry analyses and identified several Tet interaction partners, including members of the Ogt/Sin3a complex. Using a previously developed cell biological F3H assay (Eskat et al., 2012; Zolghadr et al., 2008) and biochemical analyses, we could show the interaction of all three Tet proteins with Ogt and Oga, respectively. We further characterized the interaction and observed Ogt-mediated direct glycosylation of all three Tet proteins. In summary, our results suggest a regulation of Tet proteins through GlcNAc-cycling, which might represent a mechanism affecting DNA demethylation, and possibly allows differential gene expression as an adaption to changing environmental situations.

Results and Discussion

Interactome analysis reveals Ogt as major partner of all three Tet proteins.

Tet proteins are involved in a step-wise conversion of 5mC to 5hmC and by further oxidation to 5fC and 5caC (Fig. 1A). Even though the abundance of Tet proteins and their catalytic products are well characterized (Tahiliani et al., 2009; Xu et al., 2011; Wu et al., 2011; Ito et al., 2011; reviewed in Pastor et al., 2013), only little is known about Tet regulation, targeting or their interaction network. To investigate the functional role of Tet proteins together with their interactome *in vitro* and *in vivo*, we initially generated a set of monoclonal antibodies (mAbs) against all three Tet proteins via the hybridoma technology (Koehler and Milstein, 1975) (Fig. 1B). These mAbs are suited for immunofluorescence stainings, western blot (WB) applications, or immunoprecipitation (IP) except for Tet3.

First, we fluorescently stained endogenous Tet1 in wt J1 embryonic stem cells (ESCs) and found a distinct nuclear pattern. Furthermore, we could show a direct correlation between endogenous Tet1 localization, 5hmC and histone marks like H3K4me3 representing chromatin with actively transcribing genes (Fig. 1B and Suppl. Fig. 1). Similar results were obtained for Tet2 (Suppl. Fig. 2 and data not shown).

Second, we performed co-IP assays with either endogenous Tet1 and Tet2 in ESC lines (Fig. 1C, left panel) or transiently expressed GFP-Tet fusion proteins in HEK 293T cells (Fig. 1C right panel), followed by LC-MS/MS and identified a huge set of Tet interaction partners (Fig. 1D). Most prominently, we detected O-linked N-GlcNAc transferase (Ogt), an enzyme that catalyzes the addition of a regulatory GlcNAc to serine or threonine residues (Fig. 1D). The formation of this highly abundant post-translational modification is linked to nutrient availability and crosstalks with cell cycle progression, gene expression, splicing and chromatin remodeling (as reviewed by Hanover, Krause et al. 2012). In line with previous publications, we found a large number of known Ogt-complex members like Hdacs, Sin3a, Sap30, Hsp90 or Kpnb1 co-purifying with Tets (Fig. 1D) (McDonel et al. 2011; Wysocka et al 2012; Vella et al 2013). Most of the isolated interaction partners were found in both, endogenous and GFP-fusion pull down experiments (Fig. 1D, right panel). Taken together, our data suggest that Tet proteins and Ogt are tethered together in a large complex.

Notably, most of our identified Tet interactors are representatives of a repressive chromatin environment (Fig. 1D). However, Tet localization was associated mainly with euchromatin

marks, such as H3K4me3 (Fig. 1B and Suppl. Fig. 1 / 2). This is in accordance with previous findings (Wu et al., 2011; Shen et al., 2013) and suggests a dual role of Tet proteins in both, gene activation and silencing. Spatial and temporal complex composition as well as downstream regulatory pathways involving Tet proteins are still elusive.
It is important to note, that all three Tet proteins showed an association with Ogt which was independently confirmed by co-IP analyses of GFP-Tets and mCh-Ogt as a short and long transcript variant using GFP-trap precipitation and an anti-Ogt/cherry antibody for detection (Dambacher et al., 2012) (Fig. 1D; 2A; 2B; and Suppl. Fig. 3). This is in contrast to previously published work, showing Ogt interaction with either only one or maximal two Tet proteins, but not with all three Tet family members (Chen et al., 2013; Deplus et al., 2013 and Vella et al., 2013). No interaction of mCh-Ogt was observed with GFP-Dnmt1 and the negative control GFP alone (Fig. 2B). Moreover, even washing with high salt conditions (500 mM NaCl) was not sufficient to disrupt the interaction of Tet1 catalytic domain (CD) and mCh-Ogt, demonstrating a relatively strong interaction (data not shown).

Ogt and Oga interact with all three Tet proteins *in vivo*.

Since O-GlcNAcase (Oga) that in turn removes O-GlcNAc modifications is known to coexist with Ogt in higher metazoans (reviewed in Hanover et al., 2012), we focused on a potential involvement of Oga in Tet interaction and regulation. Using a previously developed cell biological F3H assay (Zolghadr et al., 2008; Eskat et al., 2012), we validated the interaction of all three Tet proteins with Ogt. In short, GFP-fusion proteins were recruited to the lac-operator-array by the GFP-binder and used as bait for potential mCh-fused interaction partners (Fig. 2C). For the first time, we could show an interaction of Tet proteins with Oga, suggesting the involvement of both opposing partners in Tet regulation (Fig. 2B and 2D). Reciprocal F3H (bait/prey and color swop) analyses showed similar results (data not shown).
Interestingly, both Ogt and Oga were mainly localized in the cytoplasm in control cells (lower panel). However, upon co-transfection, Ogt was translocated into the nucleus, even showing co-localization at distinct Tet foci, indeed arguing for a strong interaction or even targeting of Ogt by Tet. In contrast, the cytoplasmic fraction of Oga remains unaffected, when Tet proteins are coexpressed. However, a nuclear depletion of Oga towards the F3H spot can be observed. Thus, also Tet proteins themselves are able to recruit Ogt/Oga and not only vice versa as suggested previously (Deplus et al., 2013; Vella et al., 2013; Chen et al., 2012). Interestingly,

this interaction is independent from Ogt enzymatic activity, since the inactive Ogt mutant (Ogt^{H508A}) shows unaffected association (Fig. 4A and 4D).

Mapping the interaction interface between Ogt or Oga and Tets.

To fine map the interaction interface between Tet and Ogt or Oga, we generated a series of single domain constructs and deletion mutants (Fig 3A), and applied an F3H interaction screen (Fig 3B and 3C). Here, we identified the catalytic domain (CD) of Tets as the major interaction platform for Ogt and Oga (Fig. 3 and Suppl. Fig. 3). This is in line with a previous publication showing the DSBH of Tet2 as main interaction domain (Chen et al., 2012). However, subtle differences were observed for the subdomains forming the CD. The large and unstructured insert domain of Tet2 shows a similar localization pattern as full length (FL) Tet2 and is sufficient to mediate the interaction with Ogt. In contrast, Oga cannot be targeted by the Tet2 insert domain. The same results were obtained with the N-terminal region of Tet2. Taken together, our results point to a multi-domain interaction interface of Tet2, where the individual subdomains likely act in a cooperative manner.

Interaction with Ogt / Oga regulates the glycosylation status of Tet proteins.

Since we found a strong association of both counteracting proteins Ogt and Oga with Tet proteins, we investigated potential direct GlcNAcylation of Tet proteins *in vitro* and *in vivo*. In co-IP experiments we found a strong GlcNAcylation of Tet1 upon co-expression of wt Ogt. In contrast, only basal O-GlcNAc levels on Tets were observed with the catalytic inactive variant of Ogt (Ogt^{H508A}) and Oga (Fig. 4A). Similar results were obtained for Tet2 and Tet3 (Fig. 4B and data not shown). Fine-mapping experiments revealed that O-GlcNAc sites are mainly present in the N-terminal region of Tets. As an exception, Tet2 harbors O-GlcNAcylation also in the CD, arguing for an additional regulation mechanism (Fig. 4B and 4C). Furthermore, we could show a direct glycosylation of all Tets *in vivo* (Fig. 4D).

In general, post-translational modifications such as O-GlcNAcylation are known to act in a complex crosstalk, highly coordinated and to be involved in the regulation of e.g. protein localization, activity and stability (Hanover et al., 2005; Wang et al. 2008; Hu et al., 2010; Hart et al., 2011). In summary, we can show a tight association of Tets with Ogt and Oga, mainly

mediated by the CDs of Tets (Fig. 5A, 5B and 5C). In addition, attachment of GlcNAc moieties to serine/threonine residues could be narrowed down to the N-terminus of Tets. This covalent addition is also reflected by a band shift of the bound Tet protein fraction. Together, our data suggest a potential O-GlcNAc-cycling mechanism on Tet proteins mediated by Ogt and Oga (Fig. 5C).

So far it has been shown, that Tet proteins display differential expression patterns in different tissues and during development (Sun et al., 2013; Szwagierczak et al., 2011; Szwagierczak et al., 2010; Jin et al., 2010). Thus Tet proteins are highly regulated on transcriptional levels. Here, we provide evidence for an additional layer of regulation on posttranslational/protein levels, namely glycosylation of Tets by Ogt. In accordance with Chen et al., 2012, we could not detect a direct effect of Ogt complex formation on Tet catalytic activity (data not shown), hence arguing for another regulatory mechanism. In line, Zhang et al indicated a glucose-dependent restriction of Tet3 in the cytoplasm mediated by Ogt interaction. As major interacting partner, Ogt might thus be a key regulator of Tet proteins, and therefore have indirect effects on DNA methylation patterns. Furthermore, this complex formation might specifically regulate Tet protein activity, localization and/or stability, resulting in an adaption of gene expression to the environmental situation, such as nutrient levels, and allowing cells to flexibly respond to environmental changes.

Materials and Methods

Expression constructs.

Expression constructs for GFP-Tet1 (wt, full length), GFP-Tet2, GFP-Tet3, GFP-Dnmt1$^{1-1111}$, GFP and mCherry (mCh) were described previously (Frauer et al., 2011; Liu et al., 2013; Spruijt et al., 2013). To generate the mCh-ncOgt^{short}, mCh-ncOgt^{ong} and mCh-Oga constructs, the coding sequences were amplified using cDNA from mouse E14 ESCs as template and subcloned into the pCAG-Cherry-IB vector (Frauer et al., 2011). Expression constructs for mCh-Ogt^{H508A} were derived from mCh-ncOgt^{long} by overlap extension PCR. Expression constructs for mCherry or GFP fusions of the various Tet fragments were cloned into pCAG-Cherry-IB vector or pCAG-GFP-IB by PCR amplification. All constructs were verified by DNA sequencing.

Antigen purification, immunization, generation of hybridomas and ELISA screening.

The His-tagged insert region in the CD of each Tet protein was expressed in *E. coli* BL21 (DE3) cells (Novagen, Darmstadt, Germany) and purified with the TALONTM Superflow Metal Affinity Resin system (Clontech, Saint Germain, France) under native conditions as described in Rottach et al 2007a, b, Jost et al 2012. Approximately 100 µg of each antigen was injected both intraperitoneally (i.p.) and subcutaneously (s.c.) into Lou/C rats using CPG2006 (TIB MOLBIOL, Berlin, Germany) as adjuvant. After eight weeks a boost was given i.p. and s.c. three days before fusion. Fusion of the myeloma cell line P3X63-Ag8.653 with the rat immune spleen cells was performed using polyethylene glycol 1500 (PEG 1500, Roche, Mannheim, Germany). After fusion, the cells were plated in 96 well plates using RPMI1640 with 20% fetal calf serum, Penicillin/streptomycin, pyruvate, nonessential amino acids (PAA, Linz, Austria) supplemented by aminopterin (Sigma, St. Louis, MO). Hybridoma supernatants were tested in a solid-phase immunoassay. Microtiter plates were coated over night with His-tagged Tet proteins at a concentration of 3-5 µg/ml in 0.1 M sodium carbonate buffer, [pH 9.6]. After blocking with non fat milk (Frema, Neuform, Zarrentin, Germany), hybridoma supernatants were added. Bound rat mAbs were detected with a cocktail of biotinylated mouse mAbs against the rat IgG heavy chains, avoiding IgM mAbs (α-IgG1, α-IgG2a, α-IgG2b (ATCC, Manassas, VA), α-IgG2c (Ascenion, Munich, Germany)). The biotinylated mAbs were visualized with peroxidase-labelled avidin (Alexis, San Diego, CA) and o-phenylenediamine as chromogen in the peroxidase reaction. Tet1 5D6 (rat IgG2a), Tet2 9F7 (rat IgG2a), Tet3 23B9 (rat IgG2a) were stably subcloned and further characterized.

Cell culture, transfection and F3H assay.

HEK293T, BHK and ESCs were cultured and transfected as described (Meilinger et al., 2009; Szwagierczak et al., 2010). F3H assay was performed as described in Dambacher et al., 2012.

Co-immunoprecipitation (Co-IP).

For detection of glycosylated Tet-GFP fusion proteins and interaction partners, cells were lysed in buffer containing 150 mM NaCl, 50 mM Tris-HCl (pH7.4), 5mM MgCl2, 0,1% Np40, 1xPI. Cell lysates were cleared by centrifugation at 4°C for 10 min and supernatants were incubated with GFP-trap® beads for 2 h at 4°C with gentle rotation as described in Dambacher et al., 2012. The beads were then washed three times with lysis buffer and resuspended in SDS-PAGE sample buffer. Endogenous Tet1 and Tet2 proteins were pulled out via monoclonal antibodies (5D6 and 9F7, respectively) coupled to protein G sepharose beads as described in Rottach et al., 2007.

Mass spectrometry.

After Co-IP, protein samples were digested on beads with trypsin according to standard protocols. Peptide mixtures were analyzed using electrospray tandem mass spectrometry in collaboration with the Core Facility of the Max-Planck-Institute for Biochemistry, Martinsried. Experiments were performed with an LTQ Orbitrap mass spectrometer (Thermo Scientific). Spectra were analyzed with the Mascot™ Software (Matrix Science) using the NCBInr Protein Database.

Western blot (WB) and immunofluorescence (IF) analysis.

For WB analyses, proteins samples were separated by SDS-PAGE and blotted onto a nitrocellulose membrane. Antigens were detected with a mouse monoclonal anti-GFP antibody (Roche) and a rat monoclonal anti-red antibody (Rottach et al., 2008). GlcNAc modification was detected using a mouse antibody (RL2), Ogt was detected via a rabbit polyclonal antibody (Abcam; ab96718). Alexa488-, 594-, or 647N-conjugated secondary antibodies (Sigma) were used for fluorescence detection via the typhoon (GE Healthcare). For IF stainings, cells were grown on cover slips, fixed with 3.7 % formaldehyde in PBS for 10 minutes and permeabilized with 0.5 % Triton X-100 for 5 minutes. After blocking with 3% BSA in PBS for 1 h same primary and secondary antibodies were used as for WB. Cells were counterstained with DAPI and mounted in Vectashield (Vector Laboratories). Images were obtained using a TCS SP5 AOBS confocal laser scanning microscope (Leica) using a 63x/1.4 NA Plan-Apochromat oil immersion objective. Fluorophores were excited with 405, 488, 561 or 633 nm lasers.

Acknowledgments

We thank Aleksandra Szwagierczak and Christoph Lehnert for their help in testing the monoclonal Tet antibodies. C.B. is a member of the IMPRS-LS. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB646, SFB684, SFB1064), the Nanosystem Initiative Munich (NIM), the CiPSm (center of integrated protein science munich) and the Center for NanoScience (CeNS) to H.L.

Figure Legends

Figure 1: Generation of Tet specific antibodies applied to co-precipitate Tet interaction partners.

(**A**) Step-wise conversion of 5mC to 5hmC, 5fC and 5caC. Methylation of cytosine is set by Dnmts. Further oxidation steps are catalyzed by the Tet protein family. (**B**) Generation of monoclonal antibodies (mAbs) by the hybridoma technology. Shown is a typical workflow starting from immunization of Tet antigen until the final application of the mAbs in either ELISA, immunofluorescence, immunoprecipitation (IP) or western blot (WB). Lower panel shows immunostaining of endogenous Tet1 using a 5D6 mAb in wt J1 embryonic stem cells. Cells are co-stained with an anti-5hmC antibody. DNA was counterstained with DAPI. Scale bar represents 5 µm. (**C**) Schematic outline of two different co-immunoprecipitation approaches. The left panel shows a classical pulldown of endogenous Tet proteins using protein G sepharose beads coupled with a specific Tet-antibody. On the right, a GFP-trap pulldown of transiently expressed GFP-Tets from HEK239T cells is shown. (**D**) Ogt and other Ogt complex members are identified as major Tet interaction partners by mass spectrometry. Unique peptide counts are indicated.

Figure 2: Ogt and Oga interact with all three Tet proteins *in vitro* and *in vivo*.

(**A**) Domain structure of the Tet protein family, Dnmt1, Ogt and Oga shown as either GFP or mCherry fusions. CxxC: zinc finger domain; Cys-R: cystein rich region; D: Dioxygenase domain; I: insert domain; CD: catalytic domain; PBD: PCNA-binding domain; TS: targeting sequence; BAH: bromo-adjacent homology domain; TPR: tetratricopeptide repeat; NLS: nuclear localization signal; NB: nucleotide binding domain; PB: phosphatidylinositol-3,4,5-triphosphate binding; HAT: histone acetyl transferase. (**B**) Co-immunoprecipitation analyses shows interaction of Ogt with all three Tet proteins but not with Dnmt1 and GFP alone. Similar results were obtained for Oga. The blots were probed with an anti-GFP antibody and an anti-red monoclonal antibody that recognizes several red fluorescent proteins including mCherry. I=input; FT=flow-through and B=bound fractions. (**C+D**) Co-IP results were confirmed by a fluorescent -3-hybrid assay (F3H) in BHK cells harboring a stably integrated lac-operator-array. Positive or negative interaction is marked by a solid or empty arrowhead, respectively.

Figure 3: Mapping the interaction interface between Ogt and Tets.

(**A**) Schematic depiction of the Tet2 domain structure, single domains and deletion constructs used for fine-mapping the interaction with Ogt and Oga. (**B+C**) The interaction was detected by the F3H assay in BHK cells harboring a lac-operator-array. GFP-fused Tet2 domains or GFP alone was used as bait. A fusion of Ogt/Oga with mCherry (mCh) was used as prey. Scale bar represents 5 µm. Positive or negative interaction is marked by a solid or empty arrowhead, respectively. 8-10 BHK cells were imaged for each sample. Percentages indicate the number of cells that show positive interaction.

Figure 4: GlcNAcylation of Tet proteins is regulated by Ogt and Oga.

(**A**) Direct glycosylation of Tet1 is shown after co-immunoprecipitation with Ogt. Basal endogenous glycosylation levels were similar in samples coexpressing either the catalytic inactive Ogt mutant (Ogt^{H508A}), Oga or in the Tet1 sample without coexpression. The blots were probed with an anti-GlcNAc antibody, an anti-GFP antibody, an anti-Ogt antibody and an anti-red monoclonal antibody that recognizes both mCherry-Oga and mCh-Ogt/Ogt^{H508A} fusion proteins. I=input; FT=flow-through and B=bound fractions. Molecular weight of the fusion proteins is indicated on the right. (**B**) Glycosylation was shown after co-IP experiments for all three Tet proteins and could be fine-mapped mostly to the N-terminal domains. Endogenous GlcNAc-levels are depicted on the right. (**C**) Quantitative evaluation of GlcNAc-levels of the Co-IP experiment shown in (B). Mean intensities were detected as mean grey values by the Image J software. Values were substracted by background intensities, normalized to GFP input signal. The strongest intensity signal was set to 100 %. (**D**) *In vivo* glycosylation assay using the F3H as targeting system. GlcNAc staining was detected as focal enrichment at the Lac-operator array in the mCh-Ogt coexpression sample (solid arrowhead). mCh-Ogt^{H508A} shows co-localization with Tet2 (asterisk), but no catalytic activity (empty arrowhead). Scale bar represents 5 µm.

Figure 5: Reversible and dynamic O-GlcNAcylation on Tet proteins.

(**A**) Schematic depiction of the Tet proteins, domain structure and single domains used for fine-mapping the interaction with Ogt. (**B**) Glycosylation and interaction was shown after co-IP experiments for all three Tet proteins. GlcNAcylation occurs mainly on the N-terminus, whereas Ogt interaction is mediated by the catalytic domains of Tets. Relative intensities are given on the right: ++ reflects very strong signal; + reflects strong signal; +/- reflects medium signal and - reflects no signal. Green color code indicates major GlcNAc signal or interaction signal comparing N- or C-terminus of the respective Tet protein. (**C**) Levels of O-GlcNAcylation on Tet proteins are regulated by the opposing functions of Ogt and Oga. Ogt interacts with the C-terminal part of Tet proteins and attaches GlcNAc on serine/threonine residues predominantly at the N-terminal domain. The removal of O-GlcNAc is catalyzed by Oga resulting in a dynamic change of the GlcNAcylation patterns.

References

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev 16, 6-21.

Butkinaree, C., Park, K., and Hart, G.W. (2010). O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. Biochim Biophys Acta 1800, 96-106.

Chen Q, Chen Y, Bian C, Fujiki R, & Yu X (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493(7433): 561-564.

Cole RN, Hart GW. 2001. *Cytosolic O-glycosylation is abundant in nerve terminals. J. Neurochem. 79(5): 1080-89*

Comer FI, Hart GW. 2000. O-glycosylation of nuclear and cytosolic proteins - Dynamic interplay between O-GlcNAc and O-phosphate. J Biol Chem 275: 29179-82

Dambacher S, Deng W, Hahn M, Sadic D, Fröhlich J, et al. (2012) CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3: 101–110.

Deplus R*, et al.* (2013) TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO journal* 32(5):645-655.

Eskat A, Deng W, Hofmeister A, Rudolphi S, Emmerth S, et al. (2012) Step-Wise Assembly,

Maturation and Dynamic Behavior of the Human CENP-P/O/R/Q/U Kinetochore Sub-Complex. PLoS ONE 7(9): e44717.

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: e16627

Freudenberg JM, Ghosh S, Lackford BL, Yellaboina S, Zheng X, et al. 2012. Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. Nucleic acids research 40: 3364-77

Fulop N, Marchase RB, Chatham JC. 2007. Role of protein O-linked N-acetyl-glucosamine in mediating cell function and survival in the cardiovascular system. Cardiovascular research 73: 288-97

Gu TP, Guo F, Yang H, Wu HP, Xu GF, et al. 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477: 606-10

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: 423-34

Hanover JA, Cohen CK, Willingham MC, Park MK. 1987. O-Linked N-Acetylglucosamine Is Attached to Proteins of the Nuclear-Pore - Evidence for Cytoplasmic and Nucleoplasmic Glycoproteins. J Biol Chem 262: 9887-94

Hanover JA, Krause MW, Love DC. 2012. Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. Nature reviews 13: 312-323

Hanover JA, Love DC. 2005. The Hexosamine Signaling Pathway: Deciphering the "O-GlcNAc Code" *Sci. STKE*, 312: p. re13

Hanover JA, Roos MD. 2000. Structure of O-Linked GlcNAc Transferase: Mediator of Glycan-Dependent Signaling. Biochem Biophys Res Commun 271: 275-80.

Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. 2011. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. Annu Rev Biochem 80: 825-58

He YF, Li BZ, Li Z, Liu P, Wang Y, et al. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333: 1303-7

Holt GD, Hart GW. 1986. The Subcellular-Distribution of Terminal N-Acetylglucosamine Moieties - Localization of a Novel Protein-Saccharide Linkage, O-Linked Glcnac. J Biol Chem 261: 8049-57

Hu P, Shimoji S, Hart GW. 2010. Site-specific interplay between O-GlcNAcylation and phosphorylation in cellular regulation. FEBS Lett. 584(12): 2526-38

Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. 2011. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science 333: 1300-03

Köhler G, and C. Milstein; Continuous cultures of fused cells secreting antibody of predefined specificity. Nature, 1975, 256 (5517): 495−497.

Konstandin N*, et al.* (2011) Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* 25(10):1649-1652.

Kreppel, L.K., Blomberg, M.A., and Hart, G.W. (1997). Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. J Biol Chem 272, 9308-9315.

Kreppel RK, Hart GW. 1999. Regulation of a Cytosolic and Nuclear *O*-GlcNAc Transferase: Role of the tetratricopeptide repeats. J Biol Chem 274: 32015-32022

Krzeslak A, Forma E, Bernaciak M, Romanowicz H, Brys M. 2012. Gene expression of *O*-GlcNAc cycling enzymes in human breast cancers. Clin Exp Med 12(1): 61-65

Lazarus BD, Love DC, Hanover JA. 2009. O-GlcNAc cycling: implications for neurodegenerative disorders. The international journal of biochemistry & cell biology 41: 2134-46

Lazarus MB, Nam YS, Jiang JY, Sliz P, Walker S. 2011. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature 469: 564-U168

Li Z, Cai X, Cai CL, Wang J, Zhang W, et al. 2011. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood 118: 4509-18

Liu N, Wang, M, Deng W, Schmidt CS, Qin W, Leonhardt H, and Spada, Fabio (2013): Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules. PloS one 8(5)

Lubas WA, Frank DW, Krause M, Hanover JA. 1997. O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. J Biol Chem 272: 9316-24

McDonel, P., Demmers, J., Tan, D.W., Watt, F., and Hendrich, B.D. (2012). Sin3a is essential for the genome integrity and viability of pluripotent cells. Dev Biol 363, 62-73.

Meilinger D, Fellinger 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 reports 10: 1259–1264. doi:10.1038/embor.2009.201.

Mohr F, Dohner K, Buske C, Rawat VP. 2011. TET genes: new players in DNA demethylation and important determinants for stemness. Experimental hematology 39: 272-81

Ngoh GA, Facundo HT, Zafir A, Jones SP. 2010. O-GlcNAc signaling in the cardiovascular system. Circulation research 107: 171-85

Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14, 341-356.

Pastor WA*, et al.* (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473(7347):394-397.

Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., and Lewis, B.A. (2012). Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. J Biol Chem 287, 23549-23561.

Reik W, Dean W, Walter J. 2001. Epigenetic reprogramming in mammalian development. Science 293: 1089-93

Rottach A, Kremmer E, Nowak D, Leonhardt H, Cardoso MC (2008) Generation and Characterization of a Rat Monoclonal Antibody Specific for Multiple Red Fluorescent Proteins. Hybridoma 27: 337-343.

Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, et al. (2009) The Multi-Domain Protein Np95 Connects DNA Methylation and Histone Modification. Nucl Acids Res. Available:http://nar.oxfordjournals.org/content/early/2009/12/21/nar.gkp1152. Accessed 18 April 2012.

Shen L*, et al.* (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153(3):692-706.

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.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152, 1146-1159.

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: 930-5

Tan L, Shi YG. 2012. Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development 139: 1895-902

Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC. 2004. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic acids research 32: 4100-8

Vella P*, et al.* (2013) Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Molecular cell* 49(4):645-656.

Wang Z, Gucek M, Hart GW. 2008. Cross-talk between GlcNAcylation and phosphorylation: Site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. Proc Natl Acad USA 105(37): 13793-98

Wang, Y., and Zhang, Y. (2014). Regulation of TET Protein Stability by Calpains. Cell Rep 6, 278-284.

Williams K, Christensen J, Helin K. 2012. DNA methylation: TET proteins-guardians of CpG islands? EMBO reports 13: 28-35

Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, et al. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473: 389-93

Wu H, Zhang Y. 2011. Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle 10: 2428-36

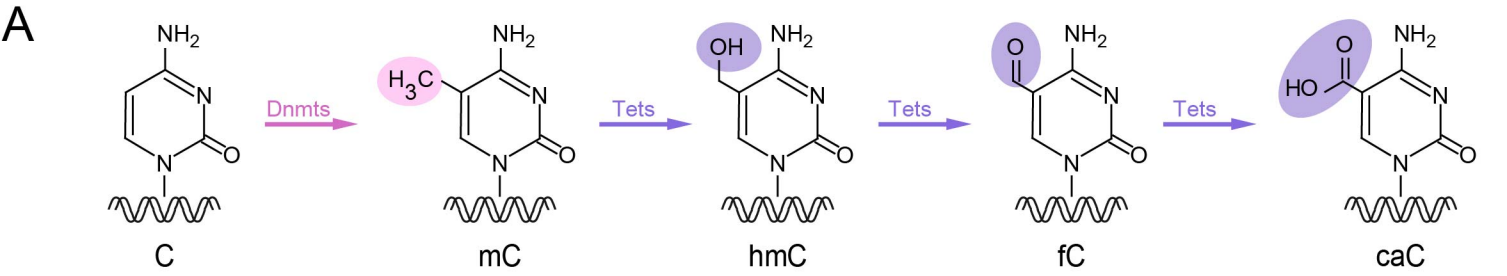
Wysocka J, Myers MP, Laherty CD, Eisenman RN, & Herr W (2003) Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes & development* 17(7):896-911.

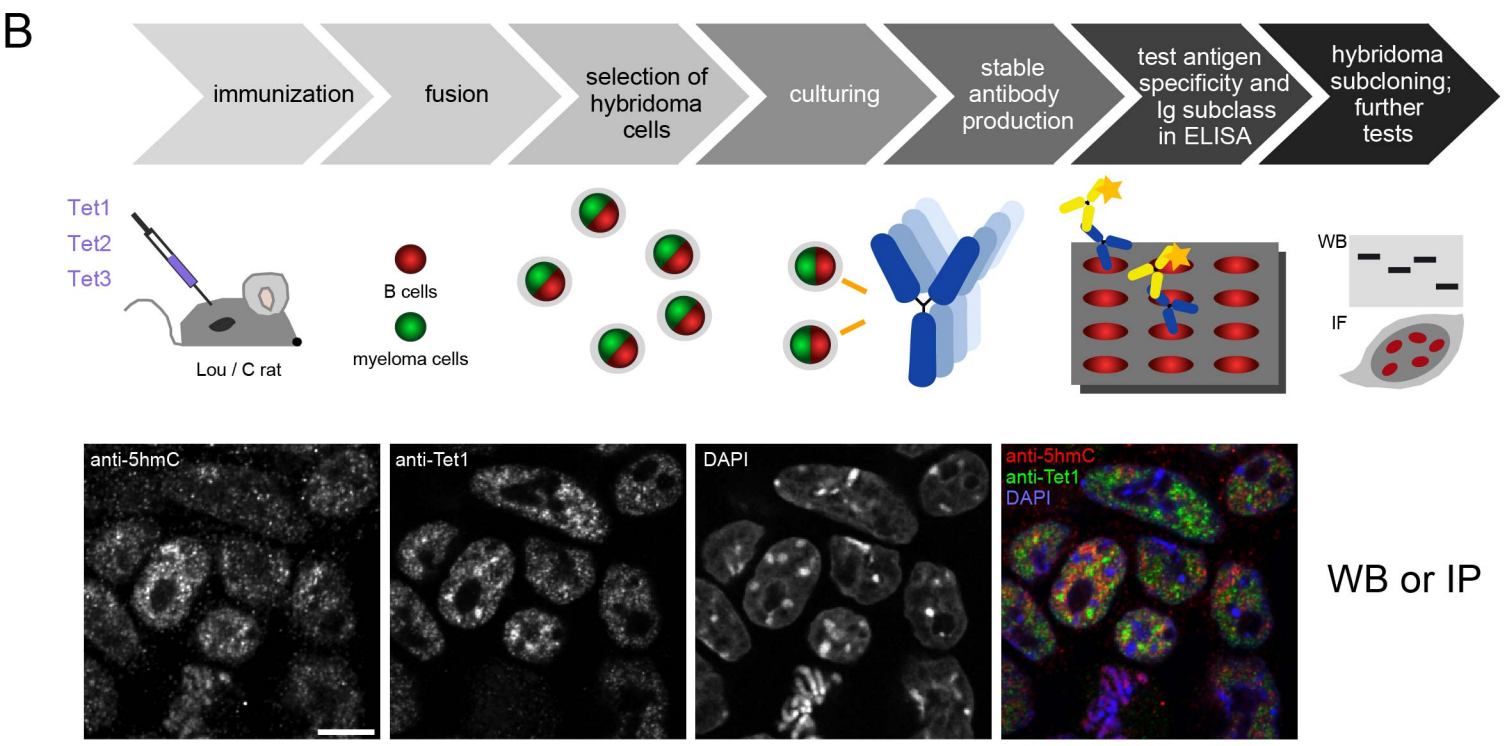
Xu Y, Wu F, Tan L, Kong L, Xiong L, et al. 2011. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Molecular cell 42: 451-64

Yang X, Ongusaha PP, Miles PD, Havstad JC, Zhang F, So VW, Kudlow JE, Michel RH, Olefsky JM, Field SJ, Evans RM. 2008. Phosphoinositide signalling links *O*-GlcNAc transferase to insulin resistance. *Nature* 451: 964-969

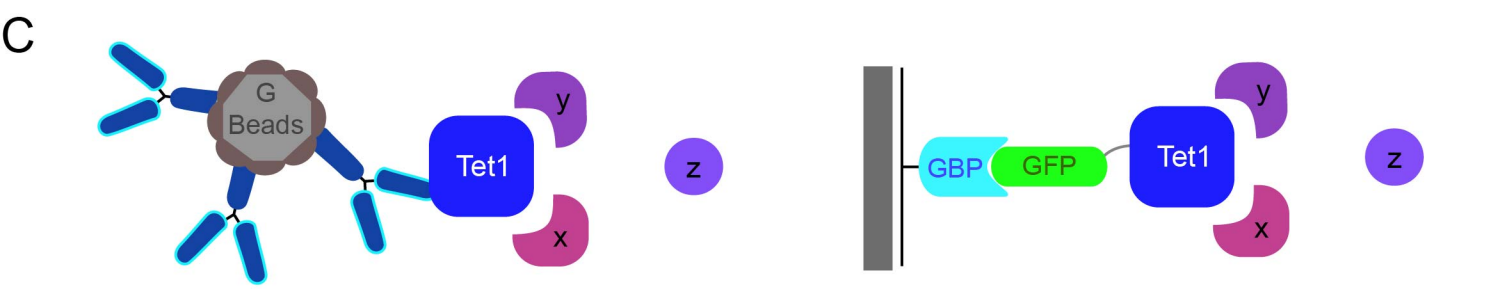
Zeidan Q, Wang ZH, De Maio A, Hart GW. 2010. O-GlcNAc Cycling Enzymes Associate with the Translational Machinery and Modify Core Ribosomal Proteins. Mol Biol Cell 21: 1922-36

Zolghadr K, Mortusewicz O, Rothbauer U, Kleinhans R, Goehler H, et al. 2008. A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells. Molecular & cellular proteomics : MCP 7: 2279-87





WB or IP



D

Tet1 overexpression

Protein	Peptides	Coverage %
Ogt	20	24.8
Sin3a	13	12.7

Tet2 overexpression

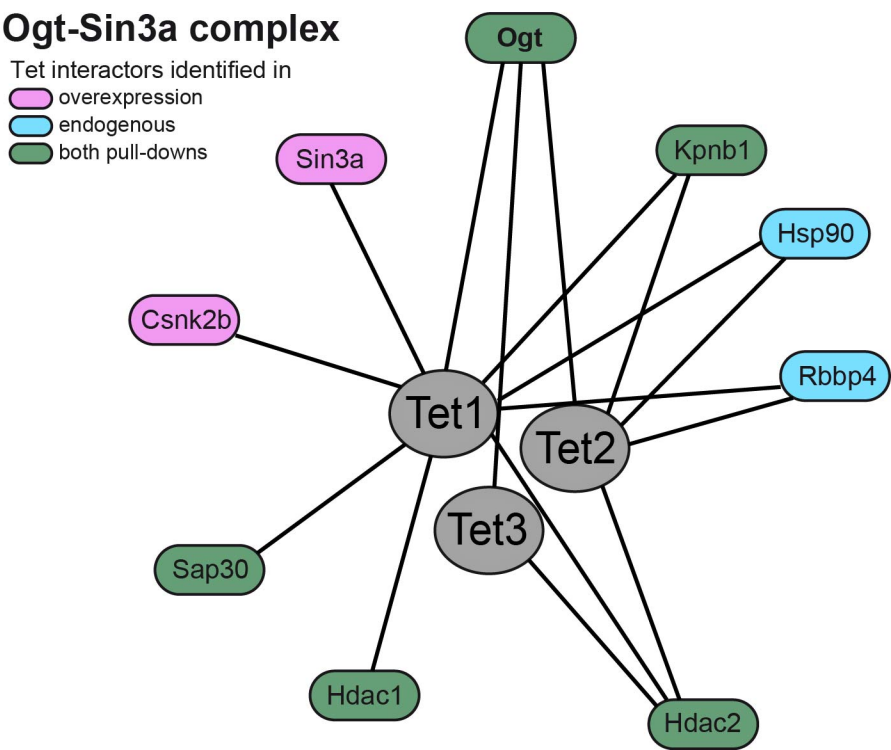
Protein	Peptides	Coverage %
Ogt	16	22.3

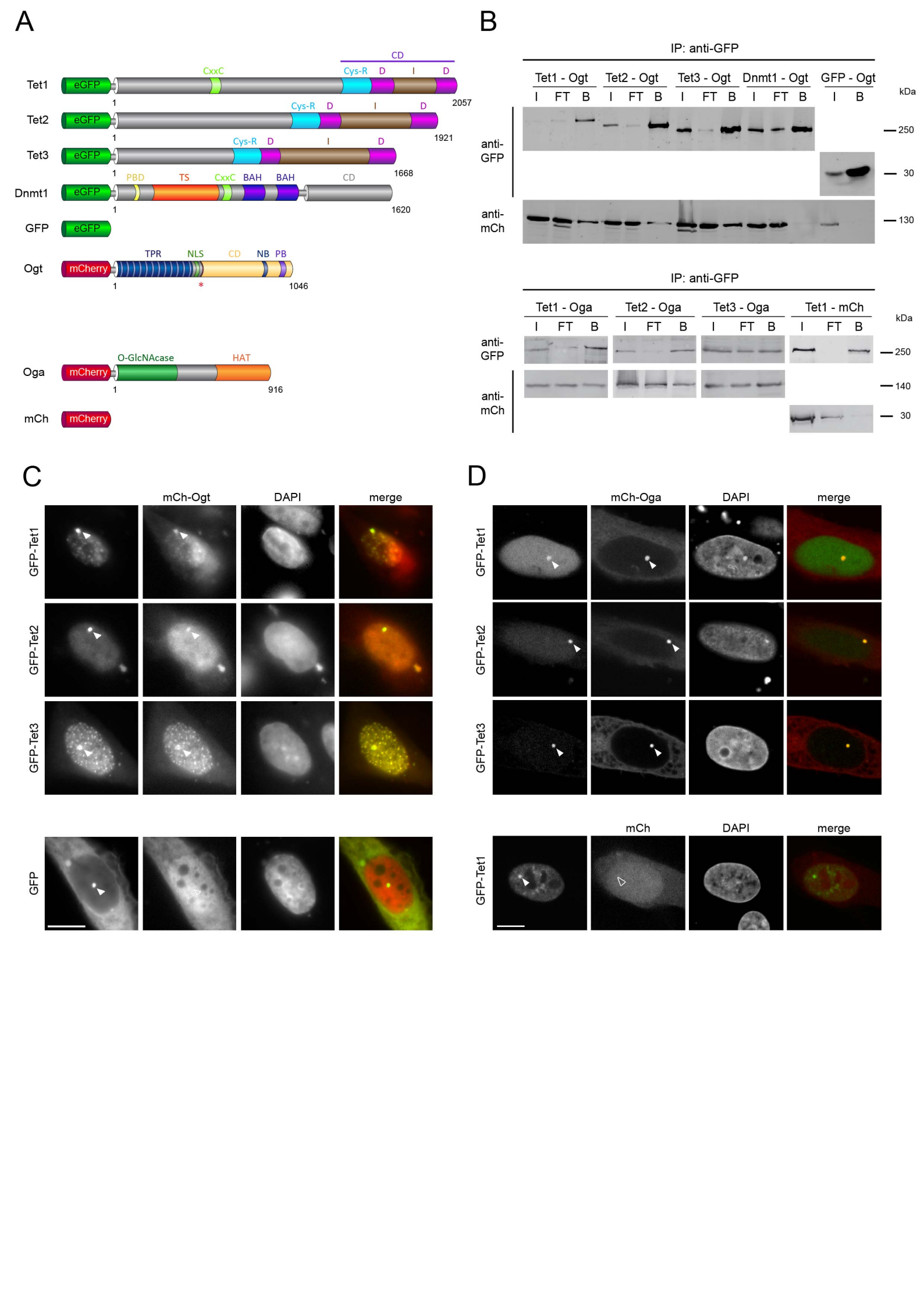
Tet3 overexpression

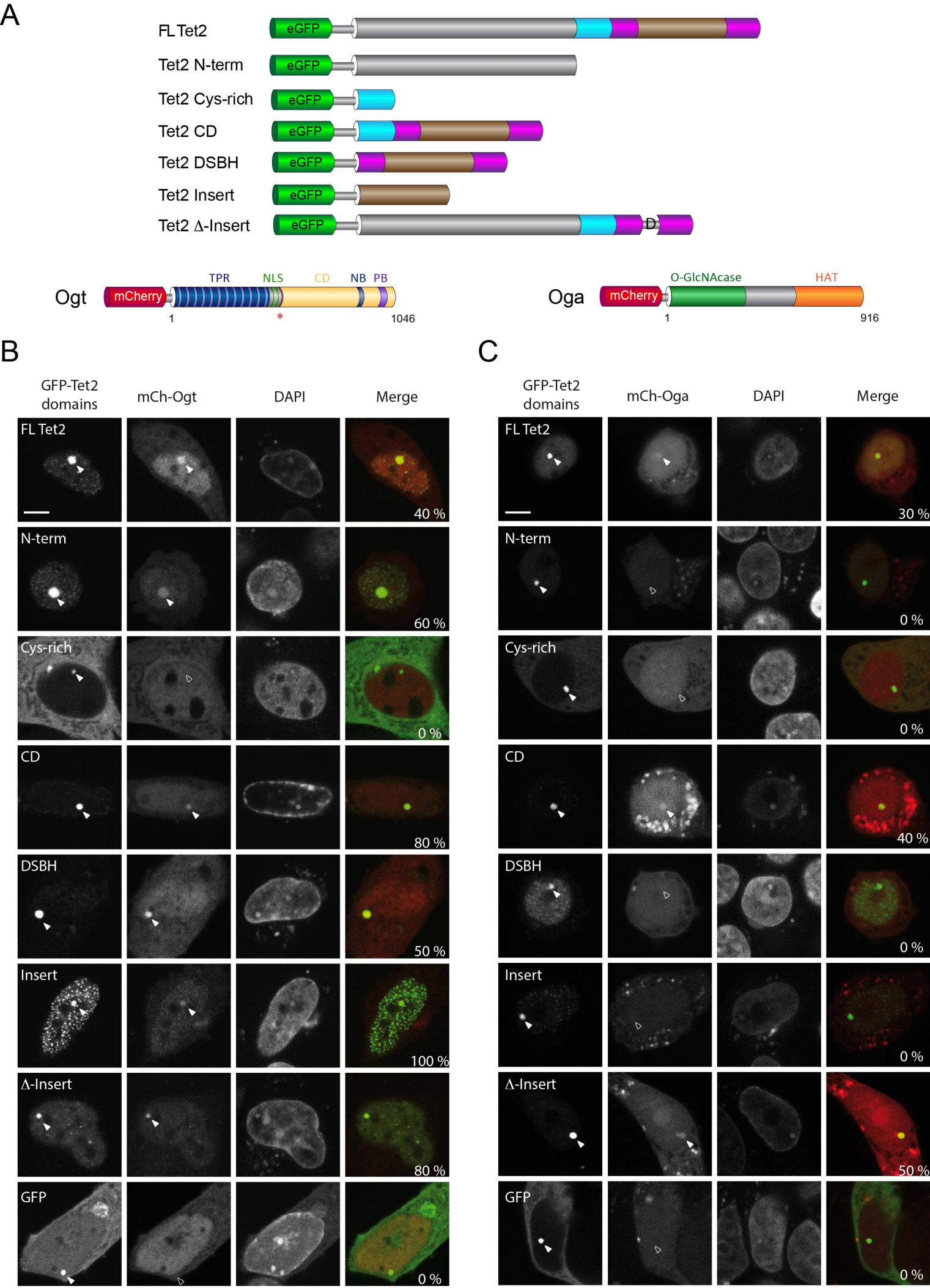
Protein	Peptides	Coverage %
Ogt	46	55.7

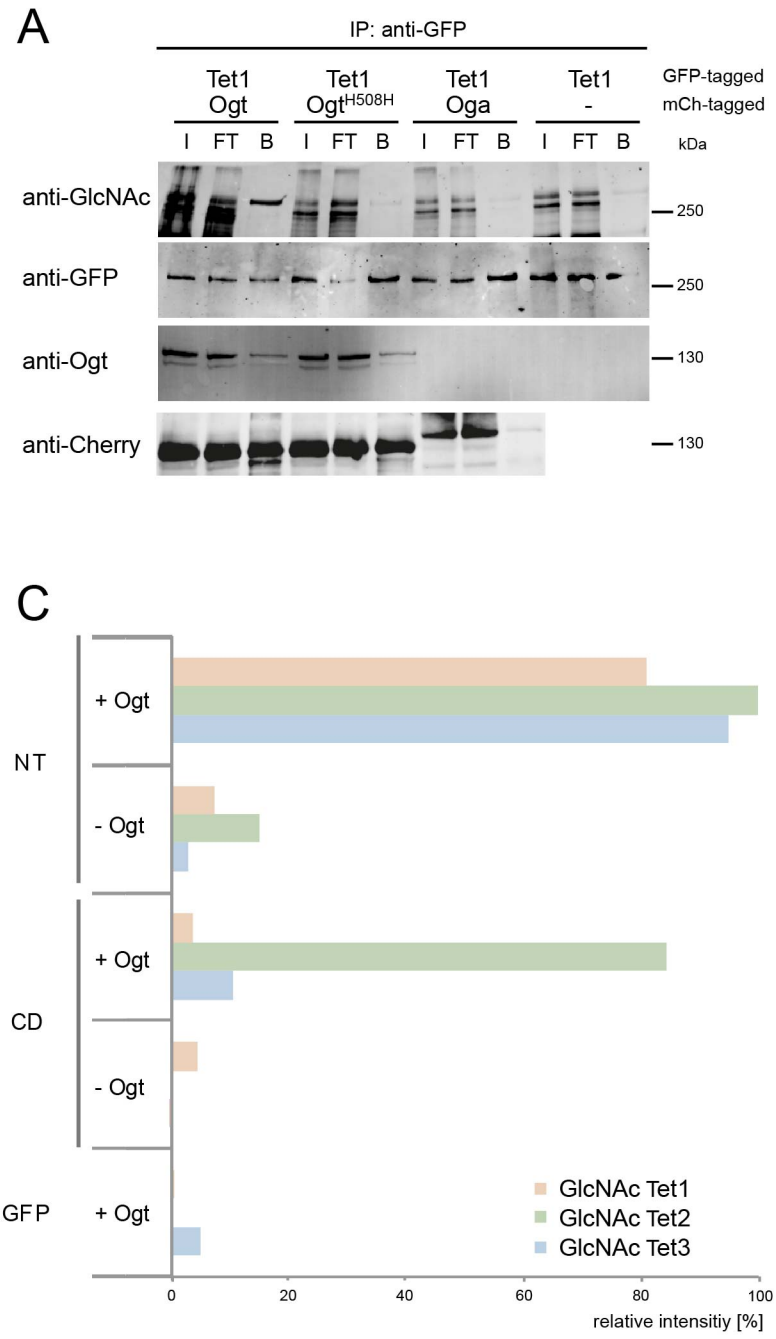
endogenous pull-downs

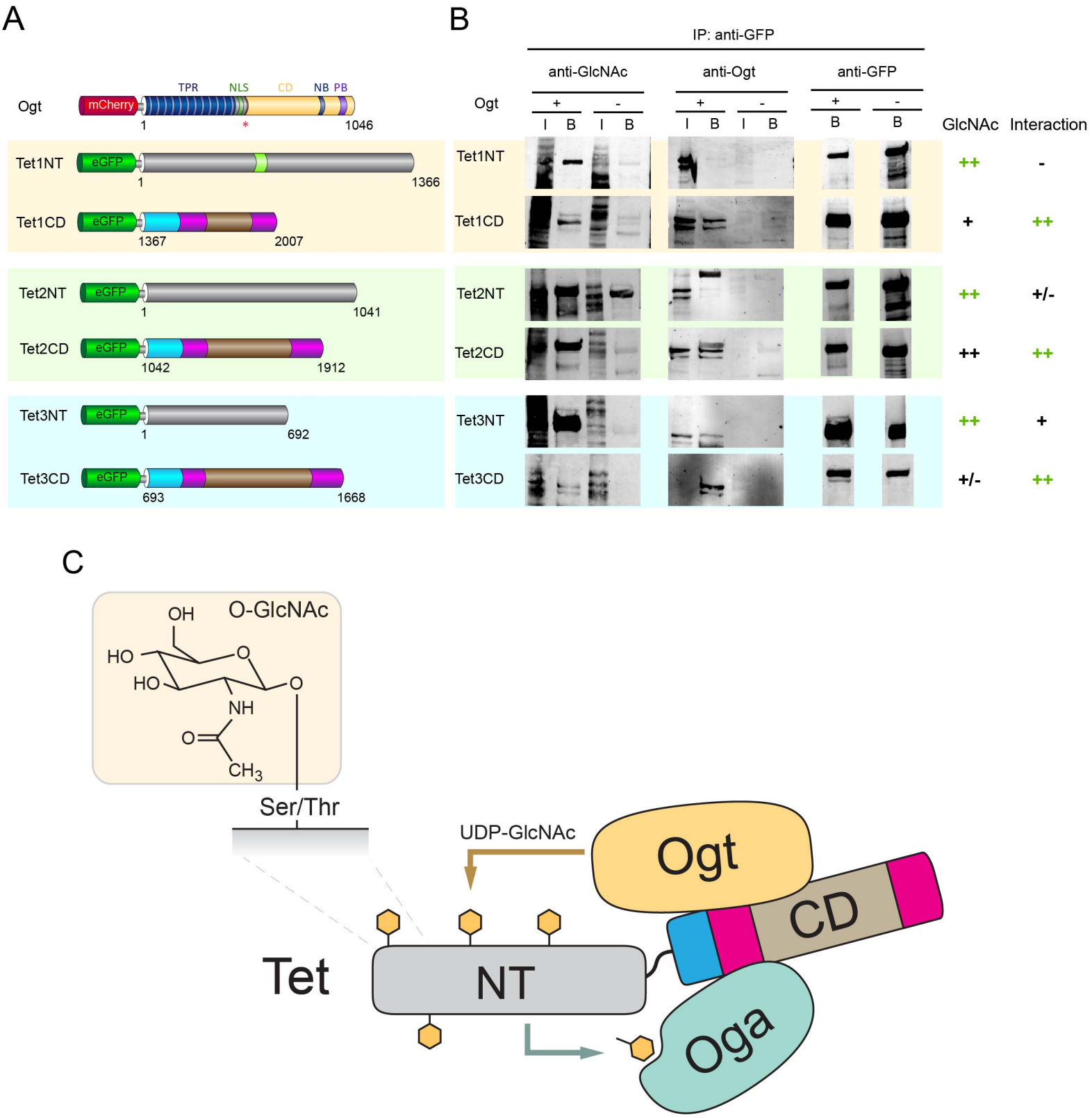
Tet1		Tet2	
Protein	Peptides	Protein	Peptides
Kpnb1	35	Kpnb1	5
Hsp90	7	Hsp90	7
Ogt	7	Ogt	2




A
CD
CxxC
Cys-R D I D
Tet1
eGFP
1
2057
Tet2
1921
Tet3
1668
PBD TS CxxC BAH BAH CD
Dnmt1
1620
GFP
TPR NLS CD NB PB
Ogt
mCherry
1046
O-GlcNAcase HAT
Oga
916
mCh
B
IP: anti-GFP
Tet1 - Ogt
Tet2 - Ogt
Tet3 - Ogt
Dnmt1 - Ogt
GFP - Ogt
I FT B
kDa
anti-GFP
250
30
anti-mCh
130
IP: anti-GFP
Tet1 - Oga
Tet2 - Oga
Tet3 - Oga
Tet1 - mCh
140
C
mCh-Ogt
DAPI
merge
GFP-Tet1
GFP-Tet2
GFP-Tet3
GFP
D
mCh-Oga
mCh

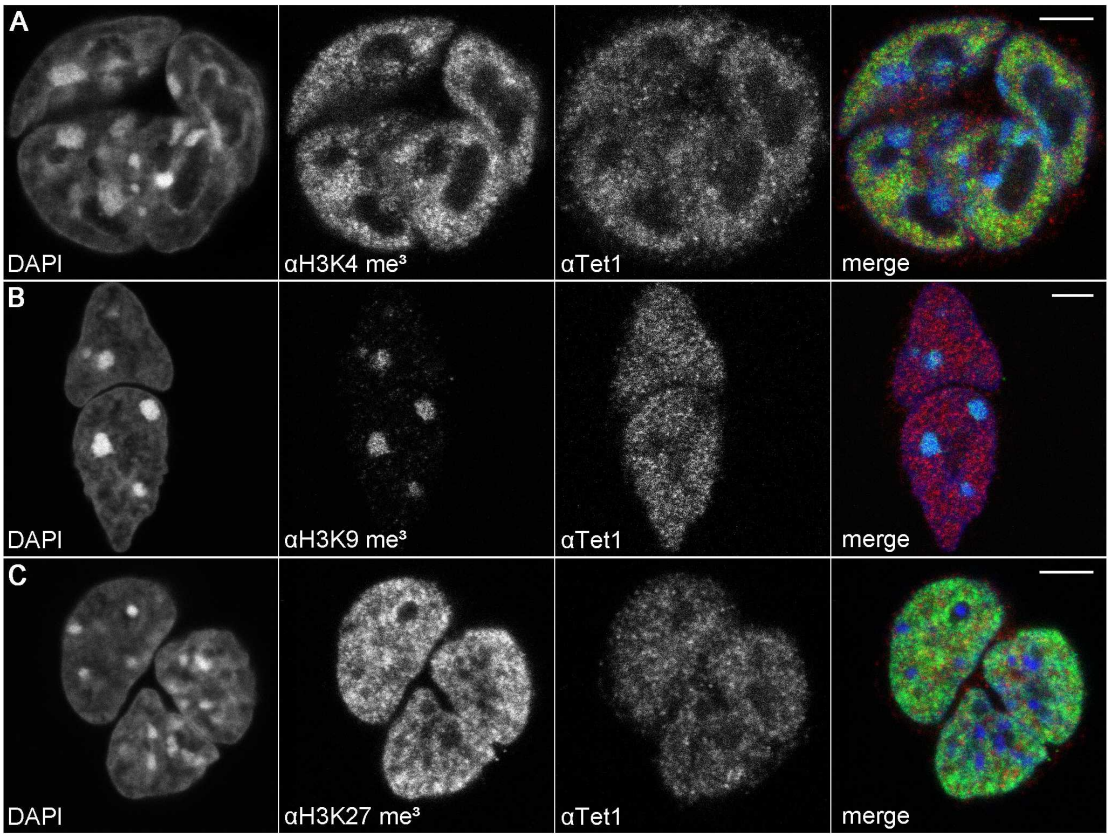

A
FL Tet2
eGFP
Tet2 N-term
Tet2 Cys-rich
Tet2 CD
Tet2 DSBH
Tet2 Insert
Tet2 Δ-Insert
D
Ogt
mCherry
TPR
NLS
CD
NB
PB
1
*
1046
Oga
O-GlcNAcase
HAT
916
B
C
GFP-Tet2 domains
mCh-Ogt
mCh-Oga
DAPI
Merge
FL Tet2
N-term
Cys-rich
CD
DSBH
Insert
Δ-Insert
GFP
40 %
60 %
0 %
80 %
50 %
100 %
30 %
40 %


A
IP: anti-GFP
Tet1 Ogt
Tet1 Ogt^H508H
Tet1 Oga
Tet1 -
GFP-tagged
mCh-tagged
I FT B I FT B I FT B I FT B
kDa
anti-GlcNAc
anti-GFP
anti-Ogt
anti-Cherry
250
250
130
130
B
IP: anti-GFP
anti-GlcNAc
anti-GFP
+Ogt
-
I B B I B B
GFP-Tet1NT
GFP-Tet2NT
GFP-Tet3NT
GFP-Tet1CD
GFP-Tet2CD
GFP-Tet3CD
GFP
C
NT
CD
GFP
+ Ogt
- Ogt
+ Ogt
- Ogt
+ Ogt
GlcNAc Tet1
GlcNAc Tet2
GlcNAc Tet3
0 20 40 60 80 100
relative intensitiy [%]
D
GFP
mCh-Ogt
anti-GlcNAc
merge
GFP-Tet2
mCh-Ogt
anti-GlcNAc
merge
GFP-Tet2
mCh-Ogt^H508H
anti-GlcNAc
merge

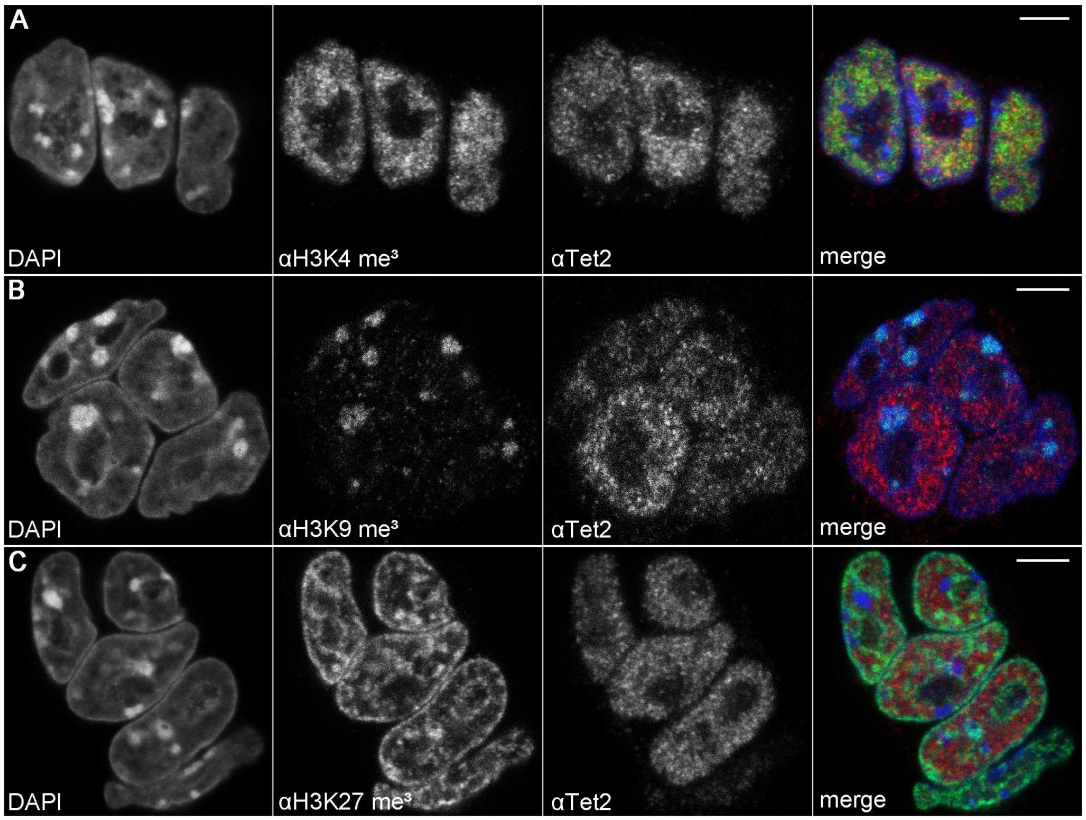

A
Ogt
mCherry
TPR
NLS
CD
NB PB
1
*
1046
Tet1NT
eGFP
1
1366
Tet1CD
eGFP
1367
2007
Tet2NT
eGFP
1
1041
Tet2CD
eGFP
1042
1912
Tet3NT
eGFP
1
692
Tet3CD
eGFP
693
1668
B
IP: anti-GFP
anti-GlcNAc
anti-Ogt
anti-GFP
Ogt
+
-
I B I B
I B I B
B
B
GlcNAc
Interaction
Tet1NT
++
-
Tet1CD
+
++
Tet2NT
++
+/-
Tet2CD
++
++
Tet3NT
++
+
Tet3CD
+/-
++

C
OH
O-GlcNAc
HO
O
HO
O
NH
O
CH3
Ser/Thr
UDP-GlcNAc
Ogt
CD
Tet
NT
Oga

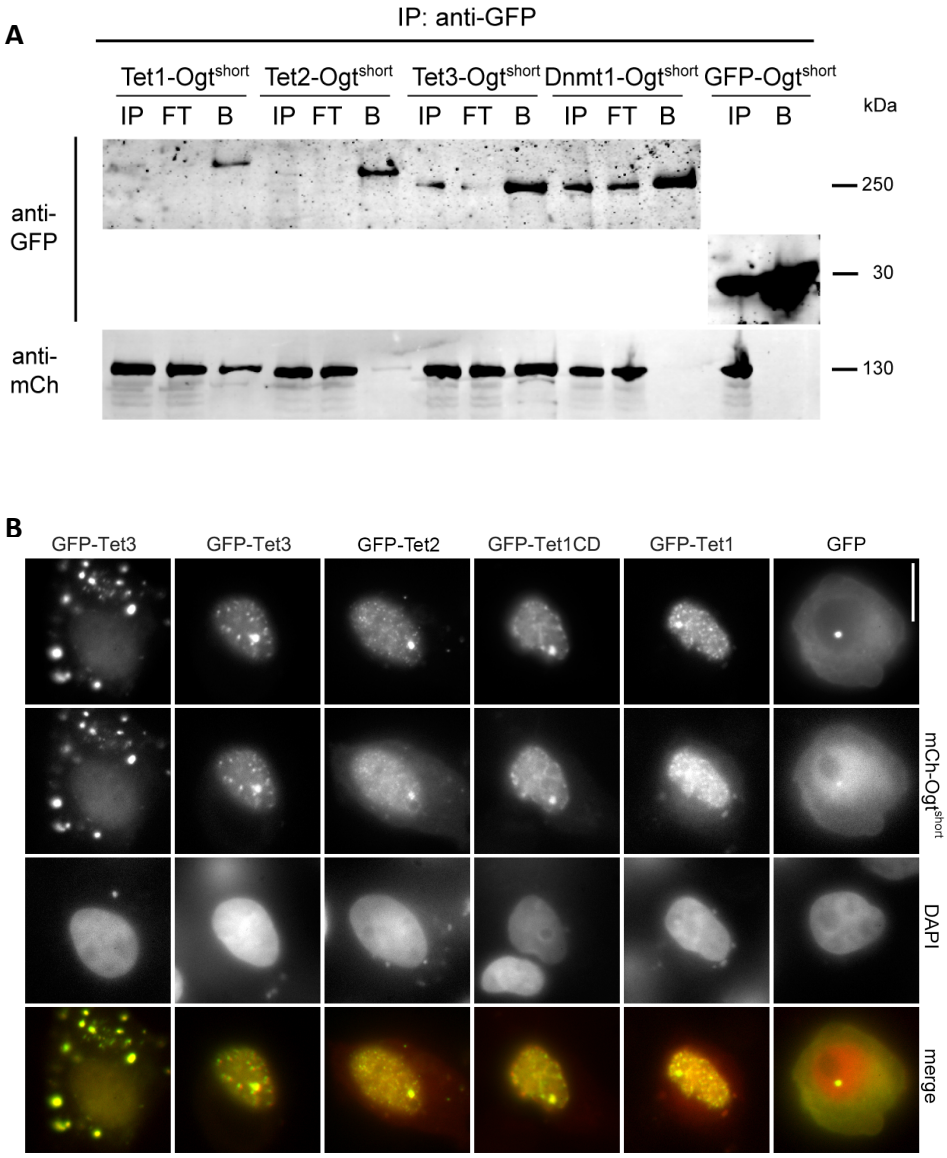
Supplementary Information:



Supplementary Figure 1: Sub-cellular distribution of endogenous Tet1 in respect to H3K4 me³, H3K9 me³, H3K27 me³ in wild type mouse embryonic stem cells.
(**A+B+C**) show immunostainings of endogenous Tet1 using a 5D6 mAb in wt J1 embryonic stem cells. Cells are co-stained with antibodies against the respective chromatin modification. DNA was counterstained with DAPI (blue). Secondary antibodies are either coupled to Alexa Fluor® 488 or Alexa Fluor® 594. Scale bar represents 5 µm. (**A**) shows wt J1 ESCs stained for H3K4 me³ (green) and Tet1 (red). (**B**) shows stainings against H3K9 me³ (green) and Tet1 (red). (**C**) H3K27 me³ with respective Tet1 staining.



Supplementary Figure 2: Sub-cellular distribution of endogenous Tet2 in respect to H3K4 me³, H3K9 me³, H3K27 me³ in wild type mouse embryonic stem cells.
(**A+B+C**) show immunostainings of endogenous Tet2 using a 9F7 mAb in wt J1 embryonic stem cells. Cells are co-stained with antibodies against the respective chromatin modification. DNA was counterstained with DAPI (blue). Secondary antibodies are either coupled to Alexa Fluor® 488 or Alexa Fluor® 594. Scale bar represents 5 µm. (**A**) shows wt J1 ESCs stained for H3K4 me³ (green) and Tet2 (red). (**B**) shows stainings against H3K9 me³ (green) and Tet2 (red). (**C**) H3K27 me³ with respective Tet2 staining.



Supplementary Figure 3: The short isoform of Ogt interacts with all three Tet proteins *in vitro* and *in vivo*.

(**A**) Co-immunoprecipitation analyses shows interaction of the short isoform of Ogt (Ogt^{short}) with all three Tet proteins but not with Dnmt1 or GFP alone. The blots were probed with an anti-GFP antibody and an anti-red monoclonal antibody that recognizes several red fluorescent proteins including mCherry (mCh). I=input; F=flow-through and B=bound fractions. (**B**) Co-IP results were confirmed by a fluorescent -3-hybrid assay (F3H) in BHK cells harboring a stably integrated lac-operator-array. Also, the catalytic domain (CD) of Tet1 shows an interaction with Ogtsi, thus mapping a potential interaction surface. Upon co-expression, Tet3 shows either a nuclear focal enrichment or was translocated to the cytoplasm. Scale bar represents 5 µm.

2.4 The domain structure and cell cycle of Tet proteins

The domain structure and cell cycle of Tet proteins

Introduction:

During early embryonic development, dynamic changes of the epigenetic profile take place. The epigenome confers stability of gene expression and is very crucial for toti- and pluripotency, correct initiation of gene expression, and is involved in early lineage differentiation in the embryo (Feil, 2009; Feng et al., 2010; Shi and Wu, 2009). Cells undergo chromatin reprogramming during the life cycle in two phases: during gametogenesis and preimplantation development, which involves the erasure and remodeling of epigenetic marks (Morgan et al., 2005). The landmark events of these two reprogramming processes are the waves of large-scale active DNA demethylation that occur in the zygotic paternal pronucleus and primordial germ cells (Santos et al., 2002; Oswald et al., 2000; Surani, 2001). Among several hypotheses and enzyme candidates, oxidative demethylation of 5mC catalyzed by Tet proteins is a well-studied and creditable mechanism (Tahiliani et al., 2009; Ito et al., 2010). Although the enzymatic function of Tet proteins has been recently discovered, the exact mechanism how Tet proteins are regulated in the cellular context is still elusive.

To address this question, we investigated Tet proteins by several methods. For this aim I developed monoclonal antibodies against mouse Tet proteins, together with Dr. Elizabeth Kremmer (Helmholtz). After identifying the best antibodies for immunofluorescence, we stained for Tet1 and Tet 2 in pluripotent ES cell. This staining analysis demonstrated that Tet1 and Tet2 express in pluripotent ES cells and are enriched at scattered regions, independently from the replication foci. Furthermore, a direct DNA competition binding assay revealed that hydroxymethylation cytosine has a specific inhibitory effect on DNA affinity of Tet proteins. This phenomenon is hypothesized to enable Tet proteins to discriminate between target genes. Using an *in vivo* mammalian fluorescent three hybrid assay (F3H assay) for protein-protein interaction, we provided experimental evidence that Tet1 is not a target of multimerization. These findings will provide new insights in understanding of the regulatory mechanisms of localization and activity of Tet proteins.

Results:

Generation of rat monoclonal antibodies against mouse Tet1 and Tet2

The understanding of the complicated Tet proteins is highly dependent on the ability to detect and visualize them in cellular context. Antibodies are an essential tool in many biochemical and cellular researches. A loop fragment of the Tet protein, which is located in the C-terminal catalytic core and of low evolutionary conservation (Figure 7), was cloned into N-terminal His_6-tag construct, expressed in Escherichia coli, and purified for antigen production of Tet1/2/3. The recombinant proteins were precipitated into inclusion body. Therefore the proteins were denatured and later refolded in native conditions. After concentrating, refolded proteins were finally purified via a gel-filtration chromatography. Elution fractions with correct molecular weights were collected, and then tested by SDS-PAGE. The yield of purified protein was approximately 0.8-1mg/ml (Figure 7).

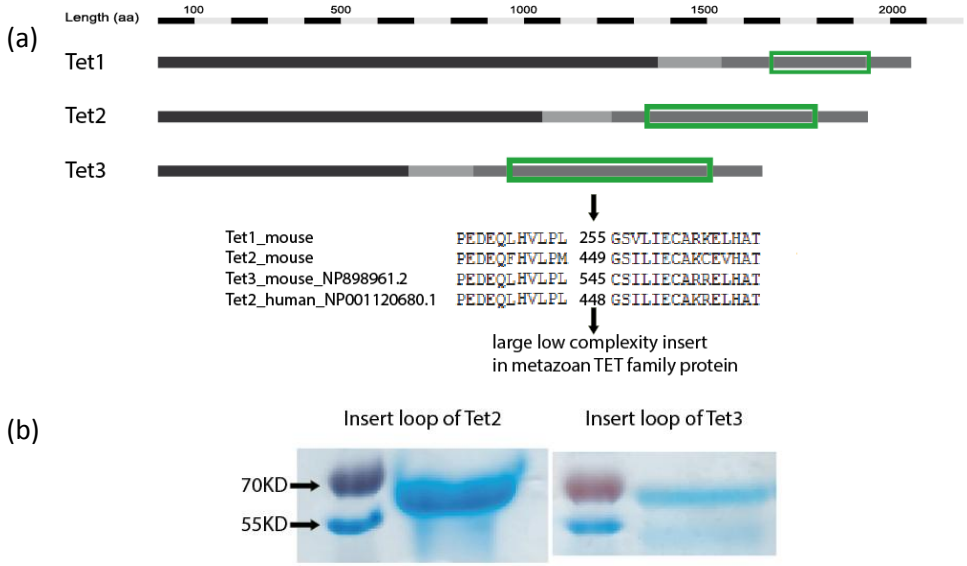


Figure 7: Purification of antigens against mouse Tet proteins (a) scheme of corresponding fragments for antigens. Mammalian Tet family members share similar domain structure, including an N-terminal domain, an extension cysteine-rich domain and a core double stranded β-helix (DSBH) of catalytic domain at C-terminus. The loose-loop insert part, which locates in the C-terminal DSBH and varies of size and sequence among Tet family members, was used to immunize rats. The corresponding domains are boxed. (b) Purified loop fragment of Tet proteins were subjected to gradient SDS-PAGE and stained with coomassie. The molecular weight markers are indicated on the left.

Dr. Kremmer generated the antibodies as described in early reference (Rottach et al., 2008). For this purpose, His-tagged fragments of Tet1/2/3 were injected into rats, hybridomas generated and were cultured, and different supernatants produced by the hybridoma technology were initially screened by the enzyme-linked immunosorbent assay (ELISA) (Rottach et al., 2008). The positive hybridoma supernatants were tested by immunoblotting assay using HEK293T cell extraction with exogenous expression of GFP-fused Tet1/2/3 proteins, as well as whole cell extraction from mouse ES cells. After comparison among these supernatants, single clones of Tet1 (code: Tet1-5D6) and Tet2 (code: Tet2-9F7) were selected and expanded.

Immunofluorescence staining of endogenous Tet1 and 2 in embryonic stem cells

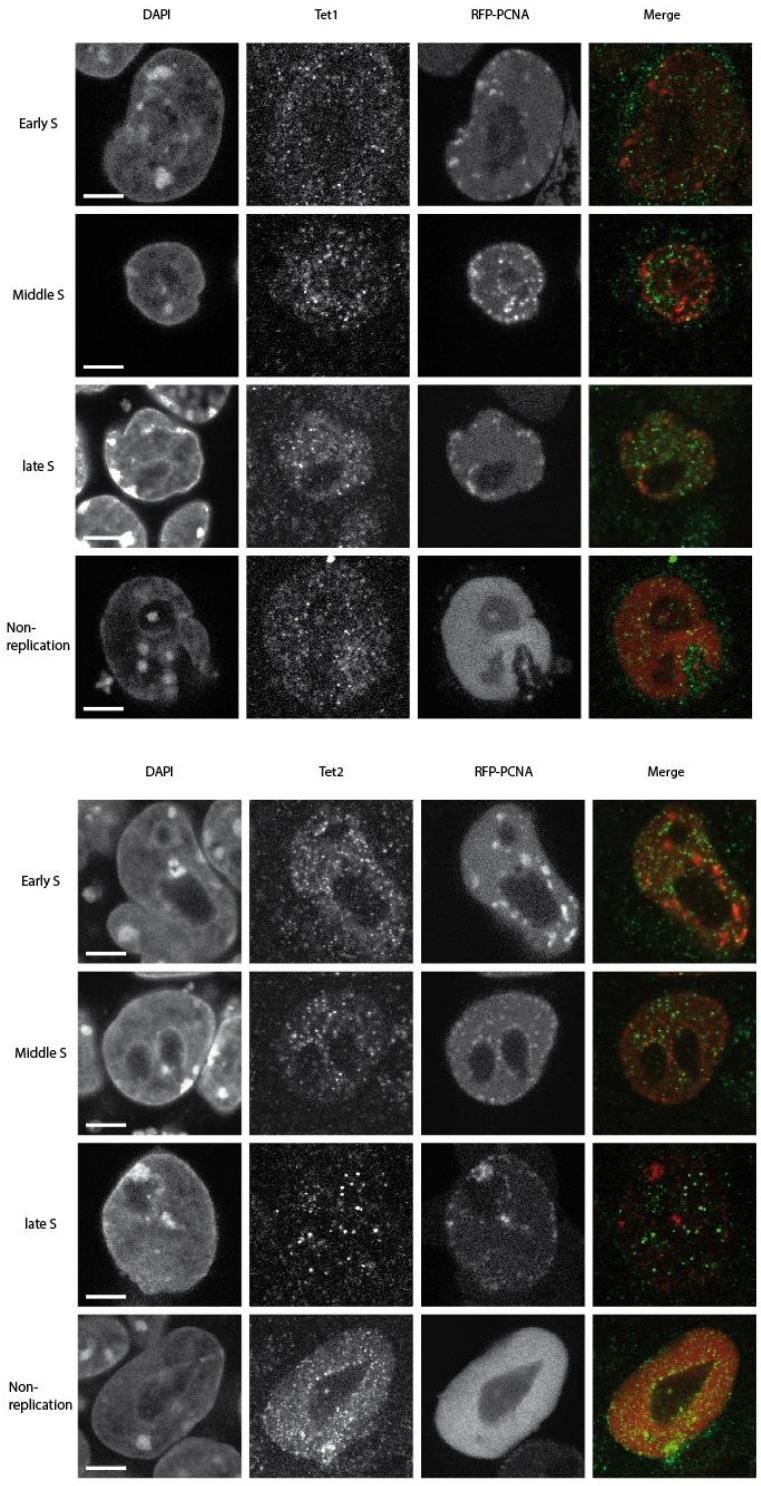


Figure 8: The localization pattern of endogenous Tet1 and 2. Immunofluorescence staining was performed in the J1 ES cell line. Rat monoclonal antibodies Tet1-5D6 and Tet2-9F7 were applied to detect Tet1 and Tet2 proteins, respectively. Anti-rat second antibody was labeled with ATTO488 fluorescence group. RFP-PCNA was excessively expressed in cells as a marker to distinguish cell cycle stages. Scale bars: 5 µm.

As described in chapter 1.3, Tet1 and 2 are expressed in pluripotent embryonic stem (ES) cells, which are derived from the inner cell mass of blastocysts, and play a very

important role in pluripotency maintenance and lineage determination. To investigate how the Tet1 and Tet2 get regulated according to the cell cycle and DNA replication, I performed *in vivo* immunofluorescence staining of endogenous Tet1 and 2 in ES cells, using the previously described antibodies Tet1-5D6 and Tet2-9F7.

RFP-PCNA was used as marker to distinguish cell cycle stages. PCNA is a key factor of the nuclear replication machinery and accumulates at DNA replication foci. Replication foci are stably localized in the nucleus and the pattern changes throughout S phase (Leonhardt et al., 2000).

From the immunofluorescence staining profile, I conclude that Tet1 and Tet2 are expressed in ES cells and are located in nucleus, which is consistent with published mRNA expression data (Szwagierczak et al., 2011). Tet1 and 2 show recruitment pattern throughout S phase and also in non-replicating cell cycle stages. The scattered enrichment spots do not co-localize with DNA replication foci (Figure 8).

Tet proteins bind to DNA substrates *in vitro*

I examined the *in vitro* DNA binding capability of mouse Tet1/2/3 and their domains by a direct competition binding assay. Differentially fluorescently labeled DNA substrates were incubated with GFP labeled Tet constructs that purified from HEK293T cells by GFP-trap as previously described (Liu et al., 2013, see chapter 2.2).

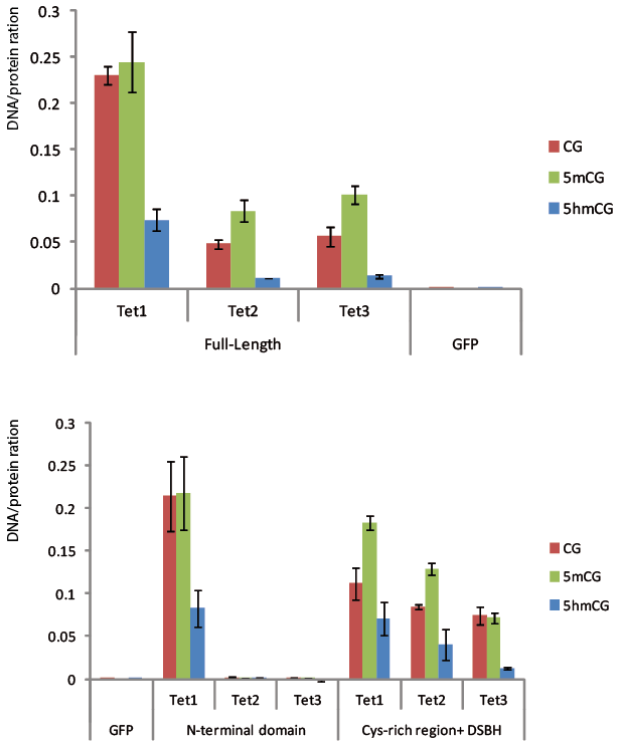


Figure 9: *In vitro* DNA binding properties of Tet family proteins and isolated domains. All constructs were expressed as GFP fusion proteins in HEK293T cells and affinity-purified with a GFP-trap. Direct competition binding assay was performed by incubating proteins with fluorescently labeled DNA substrates, which harbor the same sequence with a single CpG site either unmethylated, symmetrically methylated or symmetrically hydroxymethylated. Shown are mean values of bound substrate/protein ratios and SEM from 3 independent replicate experiments. Note that mouse Tet3 is the isoform NP_898961.2.

The above data shows that GFP-Tet1/2/3 have similar and dominant binding affinity towards substrates containing unmodified and methylated CpG sites, but the binding was inhibited by the substrate with the hydroxymethylated CpG (Figure 9). This preference shared by the isolated domains of N- and C-terminus of Tet proteins. $Tet1^{1-1366}$, $Tet2^{1-1041}$, $Tet3^{1-692}$, $Tet1^{1367-2057}$, $Tet2^{1042-1921}$ and $Tet3^{693-1668}$ were subjected to similar DNA binding assay as described above. GFP- $Tet1^{1-1366}$, $Tet2^{1-1041}$ and $Tet3^{1-692}$ constructs respectively correspond to the N-terminal domains of Tet1, 2 and 3 with GFP fused to its N-terminus, and $Tet1^{1367-2057}$, $Tet2^{1042-1921}$ and $Tet3^{693-1668}$ correspond to the catalytic part constructs (Cys-rich region and DSBH) of Tet1/2/3, respectively.

Therefore, I firstly determined the binding preference of our constructs towards substrates containing unmodified and methylated CpG sites. Secondly, I determined that the catalytic fragments of Tet proteins contribute to the affinity for DNA without altering the substrate preference. Thirdly, Tet1 N-terminus shows significant affinity with DNA substrates, which might be due to the presence of CXXC domain in Tet1 or also with some other unidentified structures in the N-terminal domain of Tet1.

Multimerization was not detected in Tet1 by *in vivo* F3H assay

Previous studies demonstrated that Tet proteins are multi-domain structural (Iyer et al., 2009; Iyer et al., 2011). For better understanding of the function and interaction between domains (Figure 12, see chapter 3.2), I checked whether Tet1 forms multimerization or whether interaction exists within the domains of Tet1.

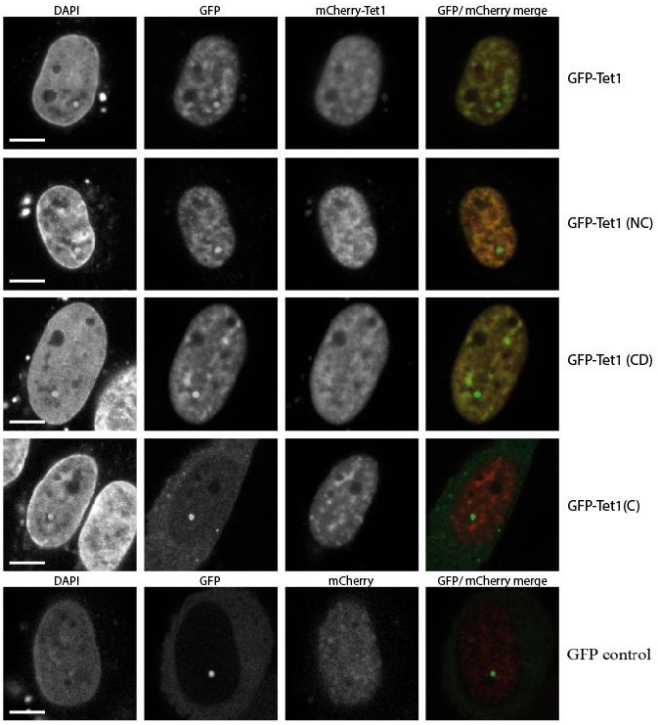


Figure 10: The interaction was detected by the F3H assay and was performed as described, in BHK cells harboring a *lac* operator array (Dambacher et al., 2012). A full-length Tet1 with mCherry-labeling was used as prey, and GFP-fused Tet1 domain constructs (upper rows) or GFP (as control; lower row) as baits. N: N-terminal domain; C: Cys-rich region; D: DSBH. Scale bars: 5 µm.

I tested the interaction of mouse Tet1 protein with each subcloned domain using a mammalian fluorescent three hybrid assay (F3H) (Dambacher et al., 2012). In this assay, GFP fused baits are anchored to a *lac* operator array, which is integrated in the genome of BHK cells, and challenged with preys fused to a red fluorescent protein. All the tested pairs did not show interaction (Figure 10). Meanwhile expression of fluorescence-fused proteins revealed a nuclear localization of full-length, N-terminal and DSBH domains of Tet1 in BHK cells, while isolated Cys-rich domain is localized in cytoplasm.

3 DISCUSSION

Although the sequence of the mammalian Tet protein family has been reported and its function in the oxidative DNA demethylation has been explosively studied, the exact mechanism of Tet proteins' regulation and the domain structure remains to be elucidated.

In my Ph.D work, the regulatory mechanisms of Tet proteins were investigated. The impacts of CXXC-type zinc finger domains in Tet1 and Tet3 were studied, including the cellular localization, gene expression, transcription, interaction partners and molecular mobility. In this research, a new transcription isoform of Tet3 was discovered and influence of CXXC domain on dioxygenation activity of Tet1 and Tet3 was reported. I optimized our powerful 5hmC detection assay, and thus provide insight into the complex regulation on dioxygenase activity of Tet proteins. A similar CXXC domain in the DNA methyltransferase Dnmt1 was also investigated for structure and functions. Furthermore, the interaction partners of Tet1/2/3, Ogt and Oga, were characterized. For these purposes, we purified peptides from Tet proteins to produce specific antigens, in order to elucidate the biochemical and cellular properties of Tet proteins.

3.1 Zinc finger domain: a widespread motif in chromatin-binding proteins

The main objective of this study is to investigate the role of CXXC-type zinc finger modules in the regulation of DNA methyltransferase Dnmt1 and Tet methylcytosine dioxygenases. To address this question, we cloned the isolated CXXC domains and their corresponding elimination mutants for our objective proteins. Various technologies were applied with these constructs to demonstrate their *in vivo* and *in vitro* properties. In addition, conventional PCR and northern hybridization provided evidence for alternative mouse Tet3 transcripts, which contains a CXXC sequence in open reading frame (ORF).

3.1.1 CXXC motif is important for Tet family

CXXC modules have been shown to mediate chromatin binding. Even though several CXXC-type domains show a high sequence similarity, they differ drastically in their binding specificity (Lee et al., 2001; Ono et al., 2002; Thomson et al., 2010; Lorsbach et al., 2003; Birke et al., 2002). In our study, a homology tree was generated from the sequence alignment and structural models were based on the reported MLL1 CXXC domain structure. MLL1 contains a zinc finger with the KFGG motif, which is shown to be essential for DNA binding activity (Ayton et al., 2004). Additionally, there are two highly evolutionary conserved CGXCXXC motifs through the CXXC family, which was also demonstrated by our study (Frauer et al., 2011, see chapter 2.1). Structure of MLL1 CXXC was resolved by multidimensional NMR spectroscopy in 2006, and the DNA-binding interface was determined by combination of many techniques (Allen et al., 2006). The structure reveals that each of the two CGXCXXC motifs chelates a zinc ion within a small helix. In addition, the so-called extended residues, which are situated between the KFGG motif and the two distal cysteines, form a surface loop that is in close contact with the DNA (Allen et al., 2006). This structure study provides basic information of CXXC modules, and therefore we used MLL1 CXXC domain as a template to generate the structural model of mouse Dnmt1 and Tet1.

The chromatin-binding proteins containing CXXC domains are divided into three subgroups (Frauer et al., 2011, see chapter 2.1). The first subgroup includes the first and second ZnF domains of Mbd1, namely Mbd1_1 and Mbd1_2, which were reported to have no DNA binding activity (Jorgensen et al., 2004). The second set includes CXXC domains of many chromatin-related proteins such as Mbd1_3, CXXC domains of Dnmt1, Mll1, CXXC finger protein 1 (CXXC1, also named as Cfp1) and Lys-specific demethylase (Kdm2) family proteins. These CXXC domains are mostly found to specifically recognize unmodified CpG sequences and target the proteins to CpG-rich locus in DNA (Lee et al., 2001; Birk et al., 2002; Jorgensen et al., 2004). The third subgroup is consisted with the Tet-related CXXC, such as CXXC domain in Tet1 (CXXC6), CXXC4 protein and CXXC10-1 motif. This data clearly demonstrates the sequence similarity of intrinsic and extrinsic CXXC domains of Tet1/2/3 (Frauer et al., 2011, see chapter 2.1).

The CXXC domains are characterized by two clusters of 6 and 2 cysteine residues. These two clusters are separated by an extending sequence, which varies among

different types of CXXC modules (Lee et al., 2001). According to the reported human MLL1 structure (Allen et al., 2006), the CXXC domains are delimited by an antiparallel β-sheet, and therefore we reasoned that the motif forming this discrete structure element could present the properties of CXXC domains. Based on the homology and structural models, we designed isolated CXXC motifs of Dnmt1 (AA 645-696) and Tet1 (AA 561-614) proteins, cloned them into mammalian expression constructs and characterized the binding activity and biological functions *in vivo* and *in vitro* by different experimental approaches. The paralogues of CXXC domains in Tet proteins lack a KFGG motif and are expected to have distinct properties in comparison to other subgroups. In my Ph.D study, I focused on DNA binding, protein interaction, cellular localization and mobility of the CXXC motifs in Tet family.

Our research in 2011 showed that a construct encoding the isolated CXXC domain of mouse Tet1 (AA 561-614) with an N-terminal GFP-tag has very low DNA affinity in an *in vitro* DNA binding assay (Frauer et al., 2011, see chapter 2.1). Discrepant results showed DNA binding activity of a larger fragment containing CXXC6 domain in Tet1 (Xu et al., 2011; Zhang et al., 2010). To resolve this contradictive result, we cloned the expanded fragment (AA 512-671) and our defined CXXC6 domain (AA 561-614) into different GFP-tag vectors. The results confirmed that both the expanded fragment with N-terminal GFP-tag and CXXC6 domain (AA 561-614) with C-terminal GFP-tag showed a similar and substantial DNA binding activity (Liu et al., 2013, see chapter 2.2). We conclude that direct fusion of GFP fluorescent group at the N-terminus of the CXXC domain might interfere with its DNA binding property, and the extra separation sequence between GFP and CXXC6 might prevent this obstruction.

Our studies also demonstrated that the CXXC4, CXXC5, CXXC6 and CXXC-10 bind CpG-rich DNA sequences. Among the unmodified, symmetrically methylated or symmetrically hydroxymethylated DNA substrates, isolated CXXC domains share a similar preference for a cytosine modification in the *in vitro* direct competition assay. Hydroxylated 5mC shows a significant inhibition effect on DNA binding. Considering the fact that hydroxymethylated cytosine is the main product of the oxidation catalyzed by Tet proteins (Tahiliani et al., 2009; Ito et al., 2010), the DNA binding preference of CXXC modules is consistent with the enzymatic function of Tet proteins. Therefore, this data might provide evidence on the mechanism for how Tet dioxygenases are targeted to specific genomic loci in different cellular contexts. The specific DNA affinity of CXXC modules may contribute to the cellular and subcellular distribution of Tet proteins and 5hmC, which is observed in different tissues (Globisch et al., 2010; Ficz et al., 2011). In addition, it was also assumed that the alternative association between Tet3 and an intrinsic CXXC domain or interaction with CXXC4 might contribute to the differential DNA binding of Tet3 isoforms. However, *in vitro* DNA binding assays suggested that CXXC4 and the CXXC-containing Tet3 isoform have similar DNA targeting (Liu et al., 2013, see chapter 2.2).

We not only reported the evidence for *in cis* association of mouse Tet3 with CXXC10-1, but also demonstrated the *in trans* interaction between Tet3 and CXXC4. Hence, other

possibilities in explaining the mechanism for how zinc fingers regulate Tet proteins are conceivable. Neural system cells present a high abundance of 5hmC in comparison to other somatic tissues (Szwagierczak et al., 2010; Kriaucionis and Heintz, 2009; Globisch et al., 2010). We note that the ratio of CXXC-10 to Tet3 transcripts is lower in brain tissues, and in contrast CXXC4 transcripts are more expressed. A recent publication described the down-regulation effect of CXXC4 and CXXC5 to Tet2 stability on protein level and 5hmC level *in vivo* (Ko et al., 2013). Therefore, it could be hypothesized that the different CXXC motifs play regulatory functions in expression and stability of the three Tet proteins, or even in an antagonistic manner. In addition, CXXC4 and CXXC5 were previously found to be modulators of Wnt signaling pathway, by inhibiting the Dvl and Axin complex (Hino et al., 2001; Michiue et al., 2004; Andersson et al., 2009). We found that subcellular localization of CXXC4 and CXXC5 is predominantly nuclear. Although many Wnt signaling factors are found in the cytoplasm (Itoh et al., 2005; Sokol 2011), it still needs further elucidation of the functional link between the Tet proteins, CXXC modules and Wnt/β-catenin signaling transduction system.

Our data suggested that interactions with distinct CXXC domains may modulate Tet function (Liu et al., 2013, see chapter 2.2; Frauer et al., 2011, see chapter 2.1). In line with our own publications, a recent manuscript reported cloning of human and mouse Tet3 isoforms containing a CXXC domain (Xu et al., 2012) and showed that Tet3 is essential in Xenopus laevis early eye and neural development by hydroxymethylating the promoters of many neuronal developmental genes. Not only the catalytic domain, but also the Tet3 zinc finger (CXXC) domain is found to be critical for Tet3 targeting and regulating genes. Tet3 protein comprising CXXC mutations or deletions loses the specific association with target gene promoters as well as functional rescue ability *in vivo* (Xu et al., 2012). These evidences indicate the importance of CXXC domain in biological functions of Tet3 protein, which is consistent with our research (Liu et al., 2013, see chapter 2.2).

In addition to the study on isolated CXXC motifs, we reported the co-transcription of CXXC domain with Tet3. As discussed above (chapter 1.3), Tet family members either comprise an N-terminal CXXC domain or are genetically situated in close proximity to a separated CXXC motifs. Tet1 includes an intrinsic CXXC-type zinc finger in N-terminal part (Iyer et al., 2009); *tet2* is 700 and 800 kb apart from an isolated *cxxc4* in human and mouse chromosome location, and *cxxc4* is proposed to be originally encoded within an ancestral *tet2* gene and separated later (Ko et al., 2013). Although human and mouse *tet3* has also been predicted to harbor ORF adjacent to the *cxxc-10* gene (Williams et al., 2011; Tan et al., 2012), our data were among the first to report the experimental claims for intrinsic connection of Tet3 with CXXC-10 motifs, and thus proved that Tet3 also contains an ancestral CXXC domain, which is confirmed by both conventional PCR and northern hybridization (Liu et al., 2013, see chapter 2.2).

The specific inhibitory effect of 5hmC on DNA binding of CXXC domains represents a possible mechanism of substrate discrimination, in which CXXC domain could act as regulatory factor by blocking Tet3 from interacting with hydroxymethylated cytosine

associated genomic elements. To examine this hypothesis, we cloned the two isoforms of Tet3, with or without the CXXC-10 motif, and compared their catalytic activity, nuclear localization, mobility and DNA binding activity by a series of *in vivo* and *in vitro* methods. Surprisingly, data showed that the CXXC domain is dispensable from Tet3's DNA binding specificity *in vitro* and the nuclear localization *in vivo*. Moreover the dioxygenase activity on genomic DNA shows similarity between the two isoforms when ectopically expressed in cells. However, FRAP analysis showed a slightly lower mobility and weakly increasing in the immobile fraction of the CXXC-10 encoding Tet3 isoform. This difference in the *in vivo* binding kinetics between these two isoforms might indicate the involvement of the CXXC domain in other nuclear interactions. Further investigation is required to assess how the CXXC domain contributes to Tet3 function *in vivo.* For example, we are generating the antibody against CXXC-10, for further investigation on the properties of CXXC-10 in different cell lineages and throughout cell cycle.

3.1.2 Dnmt1 contains a zinc finger in the regulatory domain

Dnmt1 contains a zinc finger domain of the CXXC type, which is highly conserved among Dnmt1 sequences from various animal species (Pradhan et al., 2008). As described in the homology model (Frauer et al., 2011, see chapter 2.1), some of CXXC modules, which are shown to mediate specific binding to double stranded DNA templates containing unmethylated CpG site, are clustered in a distinct homology subset and share a KFGG motif, which is found to be essential for DNA binding of CXXC domain in human MLL1 (Allen et al., 2006).

As previously reported, the deletion mutant lacking up to the first 580 amino acids in N-terminal part of human DNMT1 can still function similar to the full-length protein regarding catalytic activity and binding preference, but the mutant lacking the 672 amino acids in N-terminus could no longer form covalent complex with the DNA methylation cofactor SAM (Pradhan and Esteve, 2003). The core CXXC motif that contains a cluster of eight cysteine residues (AA 652-697) is partially disrupted in the elimination of 672 amino acids, indicating the function of CXXC domain in the enzymatic activity of Dnmt1 (Pradhan et al., 2008). However, the binding preference of CXXC was conflictingly reported in earlier studies (Pradhan et al., 2008; Fatemi et al., 2001). In our own research, we showed that this isolated mouse Dnmt1 CXXC domain preferentially binds to unmethylated DNA, which was also demonstrated for many other CXXC-containing mammalian proteins, including the methyl-CpG binding protein Mbd1 (Jorgensen et al., 2004), the histone H3K4 methylase Mll1 (Allen et al., 2006; Ayton et al., 2004; Cierpicki et al., 2010) and Cfp1 (Butler et al., 2008; Lee et al., 2001). In addition, using quantitative FRAP analysis, we showed that the Dnmt1 CXXC module interacts with chromatin structure *in vivo*.

This DNA binding preference of the Dnmt1 CXXC domain to unmethylated DNA substrates provided a possible mechanism of nucleotide discrimination and targeting, which properties are necessary for Dnmt1's biological function. To test this hypothesis, we defined and constructed a deletion mutant of Dnmt1, which lacks the corresponding CXXC motif, trying not to disrupt the folding of remaining Dnmt1 domains. When comparing the functions of wide-type Dnmt1 and its deletion mutant by *in vitro* and *in vivo* approaches, we found it is surprising that the CXXC domain is dispensable for Dnmt1's DNA binding activity and many functional representations, including intramolecular interaction, covalently binding to substrate, methyltransferase activity and DNA methylation rescue ability for *dnmt1*$^{-/-}$ ES cells (Frauer et al., 2011, see chapter 2.1).

This result is discrepant with previous researches, which demonstrate even point mutations in CXXC domain or the complete deletion could abolish the DNA methyltransferase activity of Dnmt1 (Fatemi et al., 2001; Pradhan et al., 2008). This inconsistence of CXXC domain functions could be resulted from the experimental setups. The difference between our and Pradhan's results might originate from the extent of CXXC motif. The larger size of human DNMT1 (AA 647-690) might disrupt protein folding or surrounding protein structure (Pradhan et al., 2008).

3.1.3 MLL-TET1: another combination of CXXC and TET1 catalytic activity

As described in previous chapters (chapter 2.1 and 3.1.1), the CXXC domain structure in histone H3K4 methylase MLL1 has been resolved and provides a fundamental understanding for other CXXC modules. MLL1 CXXC domain shows specific binding to unmethylated DNA substrate and contains a KFGG motif. In contrast, the Tet1 CXXC domain belongs to another homology group, which lacks the KFGG motif and demonstrates distinct DNA binding preference in respect to the cytosine modification state. Therefore, a chimeric protein between mixed lineage leukemia (MLL) and TET1, which was discovered in human acute myeloid leukemia (AML), evokes research interest (Lorsback et al., 2003; Ono et al., 2002).

This human MLL1-TET1 translocation is generated by an in-frame fusion of the amino-terminus of MLL with the catalytic part of TET1. MLL1 has been reported to be fused with over 30 different partners and frequently been discovered to be involved in translocated leukemia (Daser and Rabbitts 2005). Those fusion proteins, including MLL-TET1, consistently retain the AT hook and CXXC motif of MLL1. Intriguingly, the CXXC appears to be essential for myeloid transformation (Ayton and Cleary, 2001; Ayton et al., 2004). The MLL1-TET1 fusion protein lacks the CXXC domain of TET1, but keeps the CXXC from MLL1 (Figure 11a). At the C-terminus, the fusion protein retains the entire double strand β-helix domain, particularly the characteristic His-Xxx-Asp/Glu...His triad dioxygenase catalytic motif.

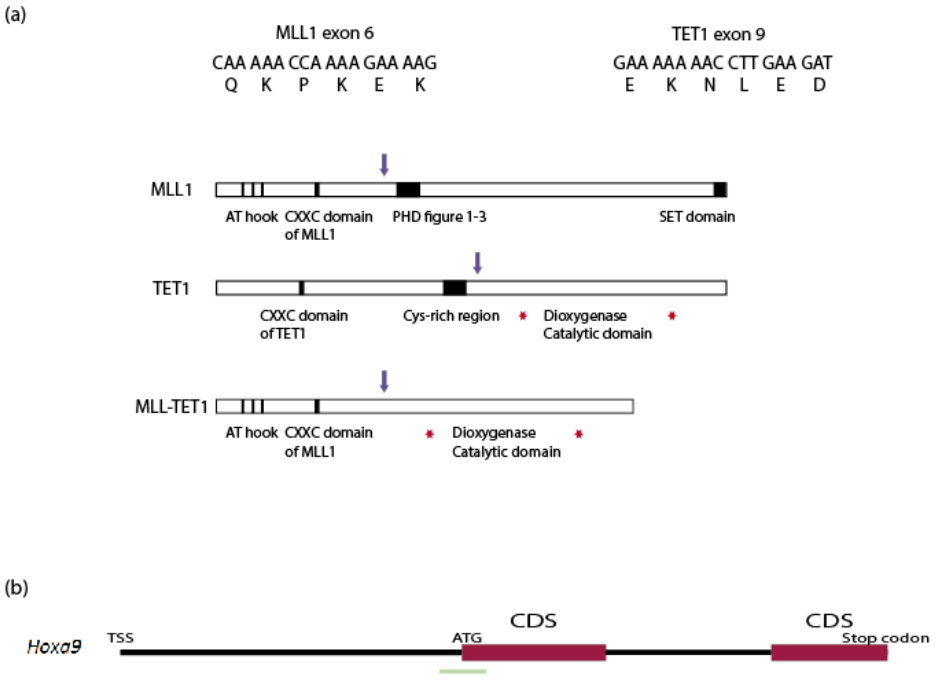


Figure 11: (a) Schematic presentation of human MLL, TET1, and the rearranged MLL-TET1 fusion protein. The arrowhead indicates the translocation breakpoint. This translocation is generated by an in-frame fusion of N-terminus of MLL with catalytic part of TET1. SET: Suvar3-9, Enhancer-of-zeste, Trithorax; PHD: plant homeodomain. Protein sequences are aligned by online server ClustalW version 2 (Larkin et al., 2007). (b) Scheme of DNA methylation analysis of *hoxa9* gene promoter. CDS indicates coding DNA sequence, TSS indicates transcription start site. The target CpG cluster is marked by green bar.

Considering the predominant DNA binding function of the CXXC superfamily, the substitution of CXXC type in TET1 N-terminus represents a possible mechanism of leukemia transformation, where the replaced N-terminus, including CXXC motif, might alter target genes of TET1, for example gene homeobox A9 (*HOXA9*). The *HOXA9* gene has been reported as the most important specific regulation targets of MLL1 (Milne et al., 2002). HOXA9 belongs to the HOX gene cluster and is essential for normal embryonic development. MLL1-mediated leukemia is frequently accompanied with persistent expression of HOXA9 (Krivtsov and Armstrong, 2007), and in leukemia the *HOXA9* gene is directly bound and activated by MLL1 fused proteins (Milne et al., 2005). Importantly, HOXA9 indicates an essential role in human MLL-rearranged leukemias survival (Faber et al., 2009).

Therefore assumingly, the transformation of MLL-TET1 could facilitate incorrectly targeting of TET dioxygenase, and consequentially leads to human tumorigenesis. My hypothesis is: if the MLL1 N-terminus navigates MLL-TET1 chimeric protein to *HOXA9* gene instead of normal TET1 target genes through either a direct CXXC-binding or an indirect protein interaction via N-terminus of MLL1, the retaining catalytic domain of TET1 could hydroxylate the methylcytosines in the promoter of *HOXA9* gene, followed by a loci-specific demethylation and abnormal activation of *HOXA9*. A supporting report was that a Tet1 fragment, lacking the N-terminal domain, can still catalyze the hydroxylation of 5mC (Tahiliani et al., 2009). Considering the importance of HOXA9 in development and hematopoietic tumorigenesis (Faber et al., 2009; Krivtsov and Armstrong, 2007), it might provide us a very good model in understanding the pathological function of CXXC domain and MLL-TET1 translocation.

To test this hypothesis, the methylation status of *hoxa9* promoter from mouse genomic DNA samples (obtained from Prof. Bohlander's group) was examined using bisulfite sequencing assay. The genomic DNA was isolated from a BaF3 mouse cell line harboring transgenic human *MLL-TET1*. The analyzed CpG-rich region includes both promoter and coding DNA sequence (CDS) (Figure 11b). Unexpectedly, I could not detect significant demethylation in this region, indicated by unconverted CpG sites after bisulfite treatment, which is similar to the negative control (data not shown). Hence, this data indicated that with human *MLL-TET1* translocation, no obvious complete erasure of methylation marks in *hoxa9* gene. However, the outlook of this investigation is to test the *in situ* hydroxylation of 5mC on the *hoxa9* gene promoter.

In contrast to my results, a recent research partially supported my earlier hypothesis about TET1's local activation on *HOXA9* in MLL-rearranged leukemia. There, Huang et al. published evidence that TET1 is a direct target of MLL-fusion proteins, and is up-regulated in MLL-rearranged leukemia (Huang et al., 2013). Importantly, although direct loci methylation analysis of *HOXA9* promoter was not shown, TET1 was confirmed to play a pivotal role through interaction with MLL1-fusion proteins in transcriptional activating a set of important oncogenic co-targets including *HOXA9*, which usually indicates the unmethylated status of the gene promoters (Busslinger et al., 1983; Jaenisch and Bird,

2003; Naveh-Many and Cedar, 1981). However, the mechanism of this regulation was not elucidated.

The inconsistence between my preliminary data and the Huang's paper could be due to many reasons. Firstly, it was human *MLL-TET1* translocated in a mouse BaF3 cell line in our research. The BaF3 cell line is a bone marrow-derived pro-B-cell line. Both, the species difference and the differentiated cell lineage could alter the methylation status of *hoxa9*. Another possible explanation is that 5mC and 5hmC are indistinguishable in the bisulfite conversion (as discussed in chapter 3.3). Thus, if TET1 catalytic domain only catalyzes partial demethylation process in this cell line, it cannot be reflected from this traditional assay. Although our data can not represent evidence for complete demethylation of *hoxa9* gene promoter, the re-targeting of TET dioxygenase activity is still very likely to be an important mechanism for regulating oncogenic genes during tumorigenesis. It warrants further investigation on the functional and pathological relevance between Tet proteins and CXXC modules.

3.2 Domain architecture and biological functions of Tet proteins

3.2.1 General review of Tet domain structure

As introduced in chapter 1.3, Tet methylcytosine dioxygenase family-mediated oxidation of 5mC is a well-characterized DNA modification, likely involved in active DNA demethylation pathways. In most animals, the Tet family undergoes a triplication to spawn the Tet1, Tet2 and Tet3 genes, which are characterized by an amino-terminal domain and a carboxyl-terminal catalytic dioxygenase double stranded β-helix (DSBH) domain, harboring an inserted cysteine-rich domain that contains 9 conserved cysteines and 1 histidine (Cys-rich domain) between N-terminus and DSBH (Iyer et al., 2009; Iyer et al., 2011; Figure 12).

The N-terminus varies among the Tet family members and is largely uncharacterized. The N-terminal domains occupy large fragments, which might indicate a functional relevance of protein interaction, activity regulation or post-translational modification (PTM). But to date, only little is known about N-terminal domains of Tet proteins. In this study, we analyzed the function, structure and DNA binding preference of the known Tet1 CXXC motif (also referred to CXXC6) that is situated in N-terminus (Frauer et al., 2011, see chapter 2.1). We also described for the first time an additional transcript variant in Tet3 including another motif called CXXC-10 in N-terminus, proving that Tet3 contains an ancestral CXXC domain (Liu et al., 2013, see chapter 2.2). In addition, our research demonstrated that glycosylation of Tet proteins mainly presents in the N-terminal regions (chapter 2.3). These results provide a further understanding on the regulatory mechanism of Tet proteins by their domain structure.

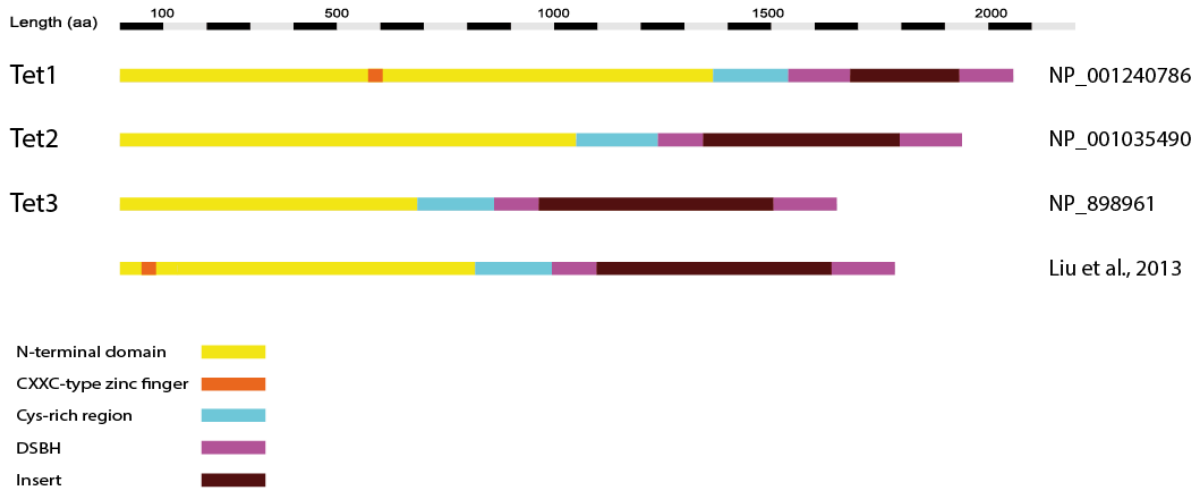


Figure 12: Domain structure of the mouse Tet1/2/3. Mammalian Tet family members share similar domain structure, including a low-conserved N-terminal domain, an extension with 9 conserved cysteines and 1 histidine (Cys-rich domain) and a core double stranded β-helix (DSBH) of catalytic domain at C-terminus. Length of domain is scaled to its real size (Tet1: NP_001240786; Tet2: NP_001035490; Tet3: NP_898961; Tet3 transcript variant: Liu et al., 2013 (see chapter 2.2).

Metazoan Tet proteins are distinguishable from other members in the Tet/Jbp family by a Cys-rich domain that extends from the N-terminus to the DSBH core (Iyer et al., 2009). The multiple alignment and bioinformatic analyses described the Cys-rich domain as a

Zn-chelating unit, containing 9 conserved cysteines and 1 histidine (Figure 13). Interestingly, the Cys-rich domain of Tet proteins is located at a similar position with the N-terminal part of AlkB (Yu et al., 2006). However, no evidence existed so far, how the presence of this domain influences Tet proteins. Further discussion about the Cys-rich domain is continued in chapter 3.2.2.

The catalytic domain of Tet family members share a double strand β-helix (DSBH) structure that binds to Fe (II) and 2OG, in which case the structure couples the two-electron oxidation of substrate to the oxidative decarboxylation of 2OG and gives succinate and CO_2 (Figure 6, see chapter 1.3). The DSBH contains a characteristic His-Xxx-Asp/Glu...His triad motif (Loenarz and Schofield, 2008; Loenarz and Schofield, 2011). Based on the resolved crystal structure of orthologue AlkB (Yu et al., 2006), secondary structure prediction of Tet proteins is summarized (Figure 13), and an extra insert loop in DSBH domain is thus identified (Figure 6, see chapter 1.3). This low complexity insert loop varies greatly in size and sequence among Tet1/2/3. It was speculated to be located on the exterior surface on one side of the catalytic domain, but to date, no experimental evidence for this hypothesis is available (Iyer et al., 2009). Thus it raises the question about the biological contribution of this insert loop in Tet proteins. Potentially, this domain might be involved in protein-protein interactions or might be target of post-translational modifications. In my Ph.D study, we found that the insert part of Tet2 is involved in interacting with Ogt, but the functional relevance of this interaction is unclear (chapter 2.3).

To better understand the biological functions and peculiarities of the Tet proteins was one of the initial aims of this thesis. Therefore we defined the fragments of isolated domains according to the previously reported bioinformatic analysis (Iyer et al., 2009) and secondary structure prediction by online server Jpred (Cole et al., 2008; Figure 13). We then cloned the isolated domains of Tet1/2/3 and their corresponding deletion mutants into mammalian expression vectors, in order to investigate them with different *in vitro* and *in vivo* methodologies.

In an earlier study, we could show that Tet1 comprising a CXXC domain, which has a DNA binding with a preference to unmethylated and methylated cytosine (Frauer et al., 2011, see chapter 2.1; Liu et al., 2013, see chapter 2.2). Thus interesting questions arise: whether the DNA binding properties observed for the Tet1 CXXC fragment are attributable to the whole protein? And how the Cys-rich region and DSBH dioxygenase domains of Tet proteins contribute to specific DNA targeting?

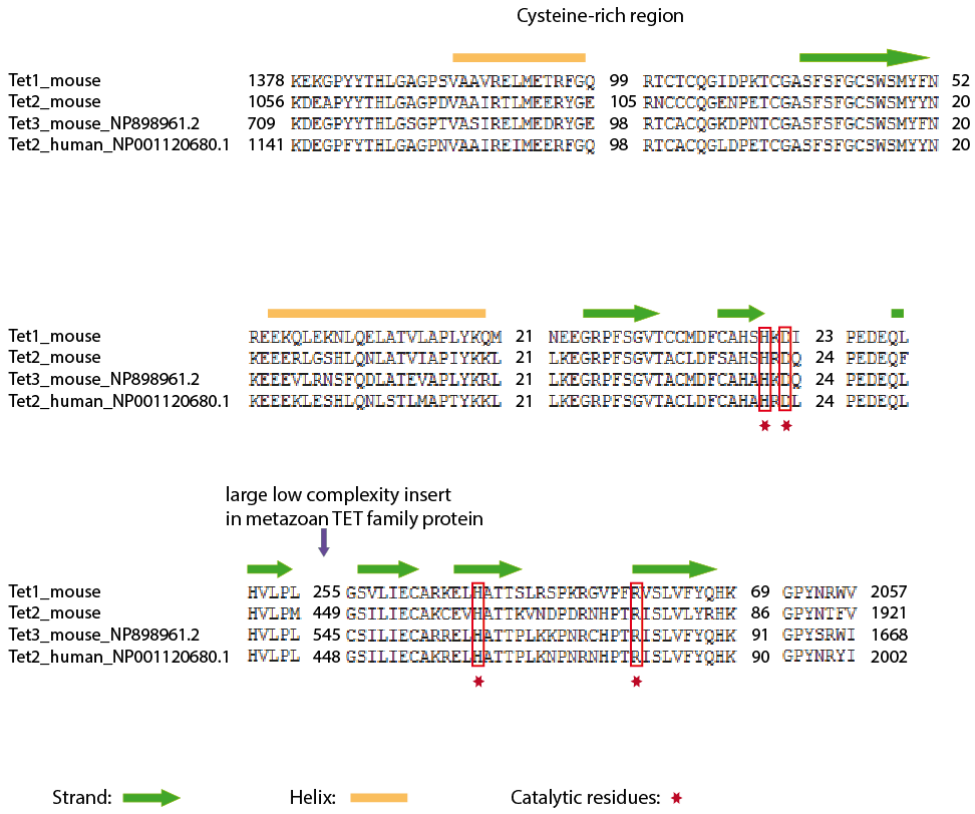


Figure 13: Multiple sequence alignment and secondary structure prediction of catalytic domain in human and mouse Tet proteins. Tet proteins belong to the nucleic-acid-modifying 2-oxoglutarate and Fe (II)-dependent dioxygenase superfamily. Conserved cysteine-rich regions extend from N-terminal domains to dioxygenase cores. Characteristic double stranded β-helix (DSBH) structures, containing unstructured insert loops, occupy the carboxyl-terminus of Tet proteins. Multiple protein sequences are aligned by online server ClustalW version 2 (Larkin et al., 2007). Second structure prediction is made by online server Jpred (Cole et al., 2008).

A direct DNA competition binding assay demonstrated *In vitro* that wide-type mouse Tet1/2/3 shows obvious and similar binding preference in respect of the modification state of DNA substrates (Figure 9, see chapter 2.4). For individual domains, only the N-terminus of Tet1 encoding CXXC6 binds to DNA; in contrast, all three catalytic domains of Tet1/2/3 show DNA affinity. The preference of these constructs is analogous: nucleotides containing unmodified and symmetrically methylated CpG sites represent higher affinity with Tet proteins. And interestingly, substrates with the centered hydroxymethylated CpG site drastically reduce DNA binding activity of the constructs. This result is consistent with our earlier study (Liu et al., 2013, see chapter 2.2). It indicates that Tet1 applies dual DNA affinity regions and Tet1/2/3 chiefly use the catalytic box to bind DNA. In line, it was reported that a Tet1 fragment, lacking the complete N-terminal domain, can still catalyze the oxidative reaction (Tahiliani et al., 2009), consistent with the independent DNA association of C-terminus. Hence, the N-terminus seems to be negligible for catalytic activity or to chiefly discriminate DNA binding and gene targeting.

3.2.2 Cysteine-rich region: an evolutionary conserved domain in Tet family

Dnmt proteins, which transfer methyl moieties onto cytosine, play a contradictory biological role to Tet proteins, which removes methyl moieties by oxidation. Fellinger et al. demonstrated that the N-terminal domain of Dnmt1 can form a stable dimer by hydrophobic interactions, and this aggregation might play an important role in the process of Dnmt1 targeting discrimination of hemi-methylated DNA substrate from other modification states (Fellinger et al., 2009). In addition, Tahiliani et al. presented very interesting data supporting the hypothesis that multimerization also exists in Tet proteins. Using recombinant proteins, immunoblotting showed that the entire catalytic domain of human TET1 (TET1-CD) forms multimers when exposed to oxidizing native condition, but not in denaturing SDS condition (Tahiliani et al., 2009). The tendency of TET1-CD to multimerize appears to at least partly involve the disulfide bond formation by the Cys-rich region, because the DSBH domain alone cannot form the dimerization bond. Whether a potential multimerization also exists *in vivo* and whether this might influence the function of Tet1 are still elusive. Therefore it attracts my interest to investigate whether multimerization of Tet1 occurs in cells.

Therefore, I decided to apply *in vivo* F3H assay to investigate this question (Dambacher et al., 2012). A combination between fluorescently labeled full-length Tet1 and truncated Tet proteins was used to fine map potential dimerization interphases. I expected to see the co-localization of green and red fluorescence if a multimerization among Tet molecules exists.

In contrast to the publication from Tahiliani, the F3H assay does not support the hypothesis of Tet1 multimerization in a cellular environment, at least not in the engineered BHK fibroblast cell line. With isolated Cys-rich region, extended to N-terminus or C-terminus, or full-length Tet1, co-localization was not detected (Figure 10, see chapter 2.4). In conclusion, protein multimerization was not observed for Tet1 in the F3H assay. This discrepancy might be due to experimental differences between the *in vivo* F3H and *in vitro* protein interaction detections. Under *in vitro* conditions, intermolecular disulfide bonds can be formed during extraction, which are resistant to reducing agent (Tahiliani et al., 2009). In contrast, F3H enables the study of *in vivo* molecular interaction.

As described in chapter 3.2.1, the Cys-rich domain is specifically conserved through metazoan Tet proteins, and shows little sequence homology with other members of the 2OG-Fe (II) dioxygenase family. Hence, the functional relevance of the Cys-rich region is still an interesting question. Tet2 mRNA expression is relatively high in hematopoietic progenitors, and deletions or mutations of Tet2 were found in a wide range of myeloid malignancies, indicating the diverse biological functions of Tet2 in hematopoietic cell lineage (Ko et al., 2010; Langemeijer et al., 2009; Delhommeau et al., 2009; Konstandin et al., 2011). Among the numerous mutations of human TET2 reported in acute myeloid leukemia (AML), myeloproliferative neoplasms (MPN) and chronic myelomonocytic leukemia (CMML), two point mutations were described to locate in the Cys-rich region (Abdel-Wahab et al., 2009). They are C1135Y in AML and C1194F in MPN patients, both of which belong to the nine characteristic and conserved cysteines (Figure 13). This

inspired me to find out the implication of Cys-rich domain using the mouse Tet2 constructs with the corresponding residues mutated, namely C1050Y and C1108F. In the outlook, it might be interesting to correlate functional properties of these leukemia-related mutations, such as cellular localization, DNA-binding activity and the influence of these two mutations on generating genomic 5hmC, with the onset and progression of the diseases. In conclusion, the domain architecture might encode many clues for Tet1/2/3 evolution, distinct biological functions and diseases.

3.3 Cellular functions of Tet proteins: methodology development and comparison

3.3.1 5-Hydroxymethylcytosine detection: a key step forward to a detailed understanding of DNA demethylation process

DNA methylation was considered as a relatively stable, repressive epigenetic mark. With the discovery of Tet proteins, it was shown that 5mC can be further modified to 5hmC, 5fC and 5caC, reflecting a so far undescribed DNA demethylation pathway (chapter 1.2). Speculation that Tet proteins and the hydroxylation product 5hmC play a central role in the DNA demethylation process is supporting by more and more recent investigations (Tahiliani et al., 2009; Ito et al., 2010; Nable and Kohli, 2011; Iqbal et al. 2011; Wossidlo et al., 2011). To dissect the individual steps of the demethylation process in quality, quantity and time, the detection and quantification of 5hmC abundance becomes an important task. For better understanding the mechanistic basis of regulation and functions of Tet proteins, we developed our own β-glucosyltransferase (Bgt)-dependent 5hmC detection assay. Hereon I discussed the advantage and disadvantage of these approaches, in comparison with the method established in our lab. Using our isotope-labeled method, we examined the genomic 5hmC levels in different tissues and investigated the regulatory mechanism of Tet proteins, including the dioxygenase activity of Tet1 and its CXXC-deletion, as well as the activity of different Tet3 isoforms, in order to elucidate the dioxygenation activity of Tet proteins under different regulatory mechanisms.

Before the discovery of 5hmC and other oxidative products of 5mC in genome, only two major states of the cytosine base were described in mammalian genome: either methylated or unmodified (Jaenisch and Bird, 2003). At that time, methods and assays were designed based on the previous knowledge. Most techniques, including bisulfite conversion and methylation-sensitive DNA restriction endonuclease assay, were set up to distinguish only these two states (Fraga and Esteller, 2002).

To date, a simple and fast method in detecting 5mC has been widely accepted and applied, based on methylation-sensitive DNA restriction endonucleases, for example HpaII and its isoschizomer MspI (Walder et al., 1983). HpaII cleaves only unmodified cytosine in a 5'-CCGG-3' context. In contrast, MspI cleaves the fragments regardless of the modification. But now, it is found the occurrence of either methylation or hydroxymethylation blocks the cleavage. Even though several methylation-sensitive methods exist, the set of endonuclease enzymatic methods cannot discriminate 5hmC from 5mC (Nestor et al., 2010; Ichiyanagi 2012). Hence, this method becomes invalid in representing the cytosine modification status if 5hmC is present.

Before 5hmC, 5fC and 5caC were discovered, bisulfite sequencing (BS) conversion was another traditional and often-used method for detecting DNA methylation states. Bisulfite treatment converts unmethylated cytosine to uracil, and subsequent PCR amplification of the converted DNA results in thymine. In contrast, this reaction leaves 5mC unaffected

(Fraga and Esteller, 2002). The comparative analysis between original and converted sequences will represent the methylation status of DNA. Recently, it was found that bisulfite conversion on 5hmC will yield cytosine 5-methylenesulfonate (CMS), and therefore 5hmC is indistinguishable from 5mC since the C-to-T transmission does not occur, meaning that BS-based methods are not applicable for detecting the intermediates of a potential demethylation process either (Huang et al., 2010; Nestor et al., 2010).

Consequently, the discovery of 5hmC in mammalian genome requires review of DNA modification. Many publications showed that 5mC and 5hmC are not only structural similar, but are also experimentally indistinguishable from each other by using traditional 5mC mapping methods (Nestor et al., 2010; Huang et al., 2010). Thus, previous DNA methylation data might require re-evaluation in the context of 5hmC. Meanwhile, it is also necessary to develop methods to specifically detect 5hmC. An overlook of recently reported methods for 5hmC is summarized in table 1, with comparison with the Bgt-dependent 5hmC measurement method in our group (Szwagierczak et al., 2010).

Mass spectrometry based methods (Le et al., 2011; Song et al., 2013) and high-pressure liquid chromatography (HPLC) with UV detection (Kriaucionis and Heintz, 2009; Liutkeviciute et al., 2009) were used to quantify oxidative intermediates of 5mC. These assays are well-established and sensitive. Digested genomic DNA containing around 0.1% 5hmC can be precisely measured (Kriaucionis and Heintz, 2009; Le et al., 2011). However, none of these procedures is easily applicable in a mid- or high-throughput manner.

A comparative method is based on selective oxidation by potassium perruthenate ($KRuO_4$) of 5hmC to 5fC. Bisulfite treatment can deaminate 5fC to uracil. The comparison of sequence data with or without chemical treatment can facilitate a single-base resolution mapping of 5hmC. However, considering the low abundance of 5hmC in genome, a very high read coverage is required for understanding genome-wide 5mC hydroxylation. The high cost of this method restricts its application to most researches (Booth et al., 2012).

The most straight-forward method to analyze global or local 5hmC level is based on 5hmC immunoprecipitation. There, a specific antibody is used to enrich and analyze 5hmC-containing sequences. A method named hydroxymethylated-DNA immunoprecipitation-sequencing (hMeDIP-seq) was established to analyze the global distribution of 5hmC in genome (Williams et al., 2011; Ficz et al., 2011; Wu et al., 2011a). Nevertheless, there are also disadvantages being reported, that hMeDIP-seq does not work for scattered 5hmC sites, and the precipitation efficiency is dependent of 5hmC density (Ko et al., 2010). As shown in reference, a comparative analysis even demonstrated that 5hmC-specific antibody basing methods show predominant enrichment of precipitating with poly-CA repeats, rising doubts in existing hMeDIP-seq data (Matarese et al., 2011).

Instead of directly immunizing against 5hmC, the bisulfite-treatment intermediate CMS is found to be highly immunogenic, resulting in highly specific poly-or monoclonal

antibodies. With these tools, it is possible to immunoprecipitate DNA with CMS, and thus this method could be applied to indirectly map genome-wide 5hmC level by the detection of CMS (Pastor et al., 2011; Huang et al., 2012). However, this method is also designed for bulky and high-intensity CMS, but does not work well in genomic regions with sparse 5hmC either.

A distinctive property of 5hmC is the glucose affinity and a series of methods are developed using β-glucosyltransferase (Bgt) for identification of 5hmC. Glucosyltransferases of T-even bacteriophages can transfer glucose from UDP-glucose to hydroxyl group of 5hmC in DNA (Szwagierczak et al., 2010; Kornberg et al., 1961; Georgopoulos and Revel, 1971), and thus the abundance of 5hmC could be determined by the covalently attached glucose. The engineered UDP-glucose alternates and is read out by different techniques.

Song et al. reported a single-base resolution method in 2012. An azide-substituted glucose is added to 5hmC by Bgt, then a biotin tag is attached by click chemistry reaction, and finally the read-out is done using real-time sequencing (Song et al., 2012a). Another laboratory combined many enzymatic and chemical steps to biotinylate 5hmC. This method also begins with a glucose transferring and is named as glucosylation, periodate oxidation, biotinylation (GLIB), describing the series of reactions. GLIB enables precipitation of DNA fragments containing very low density of 5hmC, but produces obviously higher unspecific background precipitation of DNA (Pastor et al., 2011; Pastor et al., 2012).

Our research group was among the first, which applied Bgt-catalyzing UDP-glucose transferring reaction on 5hmC detection (Szwagierczak et al., 2010; Song et al., 2011a). This method uses tritium-labeled UDP-glucose as adduct to 5hmC, and has been shown to be highly sensitive and accurate for many reasons. The Bgt is shown to modify all tested 5hmC-containing DNA strands and does not exhibit sequential or structural bias. The one-step covalent affinity is of high efficiency, strong and highly specific, so we reasoned that the incorporated UDP-[^{3}H] glucose could reflect the actual abundance of 5hmC. Thus this technology promises high robustness as compared to potential immune-based methods. Tritium-labeling is a widely-used and advantageous tag in many biological applications, due to the ease that compounds can be synthesized with a high specific activity. With a similar size of ordinary hydrogen atom, a tritium atom is not likely to structural interfere with the molecular surrounding of the target 5hmC, preventing the physical interruption between labels and DNA, and thus increases the sensitivity of detection. Experimental data confirmed that the attachment of [^{3}H] glucose to 5hmC dramatically enhances the sensitivity and simplicity of the 5hmC detection and quantification in bulky biological samples (Szwagierczak et al., 2010; Liu et al., 2013, see chapter 2.2).

Technologically, there were some technological improvements after the initial publishing of the assay (Szwagierczak et al., 2010): we later used abundant UDP-[^{3}H] glucose, instead of a mixture of “hot” and “cold” UDP-glucose in order to enhance the signal and

improve sensitivity. For better DNA retrieval after purification, we have tested DNA precipitation by trichloroacetic acid (TCA) and silica membrane binding, and the optimized protocol applies the silica column chromatography and vacuum. For quantification, we assume that all DNA samples, including controls, get clearly free from non-reactive UDP-[^{3}H] glucose and get retrieved with the same ratio. A few essential steps need to be considered and controlled to ensure accurate measurements, such as the homogenization of genomic DNA into fragment of a size ranging from 200-1000bp by sonification. Calibration curves should be made using a mixture of 5hmC-containing and unmodified reference fragments, with percentage gradient of 5hmC. The high linear relationship between [^{3}H]-glucose incorporation and percentage of 5hmC indicates the robustness of the method (Figure 14).

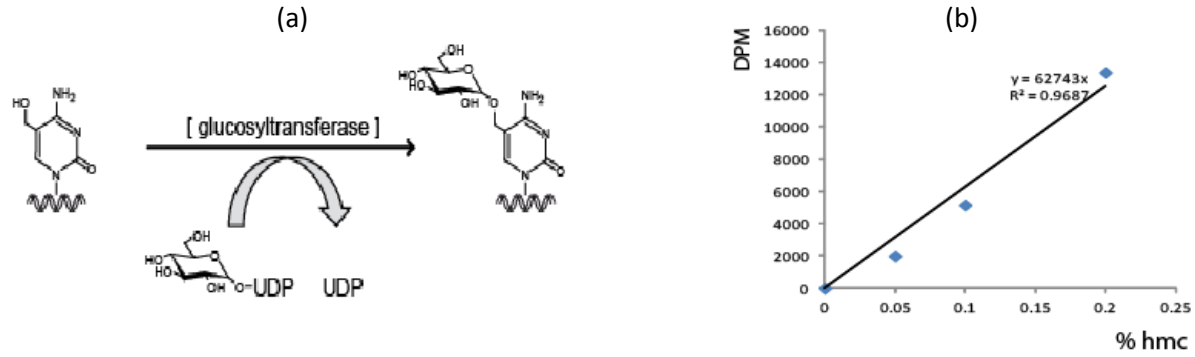


Figure 14: (a) Scheme of the 5hmC glucosylation reaction catalyzed by β-glucosyltransferase (Bgt), UDP-glucose with tritium-labeled as cofactor. Glucose is transferred to hydroxyl moiety of 5hmC in DNA. The 5hmC is radioactive and can be detected by liquid scintillation (Szwagierczak et al., 2010). (b) Calibration curve using mixtures of 5hmC-containing and unmodified reference fragment, with percentage gradient of 5hmC. Note the high linearity between [^{3}H]-glucose incorporation and percentage of 5hmC.

In conclusion, along with the development in the research field of DNA demethylation, it is found that the Tet protein-catalyzing oxidative modifications on the cytosine base leads to a demethylation pathway. New challenges arise, re-evaluation of established DNA methylation data is necessary, and most importantly new methodology of efficient 5hmC-detection is required. Up to date, a few of methods for 5hmC quantification, enrichment and mapping exist (Table 1). The advantages of our tritium-labeling glucose transferring method are that it is specific, accurate, high-throughput, available for genomic-wide detection, allowing parallel processing of large sample number, and is easy to operate (Szwagierczak et al., 2010; Song et al., 2011a; Liu et al., 2013, see chapter 2.2). And we could demonstrate this method is a useful tool in cell biology and biochemistry researches.

Table 1: The 5hmC toolbox, summary of currently established methodologies for detection and quantization of 5hmC

Methods	References	Descriptions	Advantages	Disadvantages
Isotope labeled assay	Szwagierczak et al., 2010; Liu et al., 2013	Tritium-labeling glucose is transferred to the hydroxyl group of 5hmC by Bgt. Radioactivity is detected by liquid scintillation	Specific, accurate, high-throughput, available for genomic-wide detection, allowing parallel processing of large number of samples, and is easy to operate	
Mass spectrometry	Le et al., 2011; Song et al., 2013	Mass spectrometry based assays	Sensitive, fast and accurate	Not applicable to large number of samples
HPLC	Kriaucionis and Heintz, 2009	High-pressure liquid chromatography (HPLC) with UV detection	Sensitive and accurate	Not applicable to large number of samples; Relies heavily on the chromatographic separation to avoid contamination from other nucleotides
hMeDIP-seq (hydroxymethylated-DNA immunoprecipitation-sequencing	Ficz et al., 2011; Williams et al., 2011;	Antibody against 5hmC allows detection	Allowing processing of large number of DNA samples	Higher background pull-down; high density dependence of CpG; artifact of precipitating CA and CT repeats
GLIB (glucosylation, periodate oxidation, biotinylation)	Pastor et al., 2011; Pastor et al., 2012	Glucose is transferred to 5hmC by Bgt, oxidized with sodium periodate to yield aldehydes, and reacted with the aldehyde reactive probe (ARP), yielding two biotins at the site of every 5hmC. Pull down 5hmC-containing DNA fragments by streptavidin beads, followed by sequencing.	high specificity and sensitivity	higher background precipitation of DNA
CMS-based DNA sequencing	Pastor et al., 2011; Huang et al., 2012; Ko et al., 2010	5hmC is converted to 5-methylenesulfonate (CMS) with sodium bisulfite. Special antisera against cytosine CMS pull down the DNA, and sequencing.	More sensitive than anti-5hmC in DNA dot blot assays; low background	High density dependence of CpG
Click-chemistry	Song et al., 2011a	An azide-containing glucose is transferred onto 5hmC by Bgt. The azide group is chemically modified with biotin for enrichment and sequencing	Single-base resolution; high specificity	

Table 1: The 5hmC toolbox, summary of currently established methodologies for detection and quantization of 5hmC

3.3.2 Monoclonal antibodies against Tet proteins: a useful tool to understand *in vivo* properties

As described in previous chapters 1.2.3, Tet1/2/3 proteins are expressed differentially during development and play different roles in cellular processes, despite their similar domain structure among the Tet proteins. Important questions arise, including what are the cellular features of Tets and how are they differentially regulated. To better understand Tet proteins *in vivo*, specific tools to analyze Tet proteins in different biological assays should be applied. To date, no specific antibody against Tet proteins exists. Hence, we generated antigens of mouse Tet1/2/3 by expressing them in *E. coli* and followed by purification process. The insert loops in C-terminus of low evolutionary conservation were selected as targeting sequence for the antibodies, in order to minimize the cross-talk among Tet1/2/3 proteins (Figure 7, see chapter 2.4). For antibody production, these antigens were individually used to immunize rats (Rottach et al., 2008), which were further applied in many biochemical methods (chapter 2.4). Using these antibodies, we pull-down the endogenous Tet proteins, identified their interaction partners and characterized their glycosylation modification (chapter 2.3). In addition, I analyzed the localization pattern of endogenous Tet1 and Tet2 in ES cells in the early, middle and late S phase (chapter 2.4).

Study of the human ES cell cycle showed that the pluripotent ES cells maintain the four canonical cell cycle stages G1, S, G2, and M. However, stem cells represent unique cell cycle characteristics, such as an absent G0 phase, short doubling time due to abbreviated G1 phase, and differentially regulated cell cycle checkpoint (Becker et al., 2006; Barta et al., 2013). Therefore we adopted PCNA as the marker to distinguish non-replication and DNA synthesis stage. Our immunofluorescence result demonstrated that Tet1 and Tet2 are expressed in ES cells (Figure 8, see chapter 2.4), which is consistent with published mRNA expression data (Szwagierczak et al., 2010) and the high abundance of 5hmC in ES cells (Szwagierczak et al., 2010, Ficz et al., 2011). The subcellular localization of Tet1 and 2 shows enrichment through the whole synthesis and also in non-replication stage. Dnmt1 is known to be recruited to replication forks by interaction with PCNA (Chuang et al., 1997; Leonhardt et al., 1992). In contrast, the enriched spots of Tet proteins do not co-localize with DNA replication foci, indicating the onset of hydroxylation of target genes is separated from the replication and methylation apparatus. The mechanism of how Tet proteins specifically discriminate and target genes remains largely unknown.

In addition, another question is interesting: whether expression levels of Tet1 and Tet2 fluctuate according to the cell cycle stages? Live-cell imaging system is a very useful tool for tracing fluorescent-labeled protein expression profiles *in vivo*. Biochemical analysis *in vitro* usually applies synchronization cells, meaning arresting cells at certain cell cycle stage and releasing them synchronically from the same phases. Therefore, my future plan is to analyze the expression fluctuation of endogenous Tet proteins during the cell cycle by our antibodies.

3.4 Interaction of Tet proteins with Ogt and Oga

As introduced in chapter 1.2 and 1.3, Tet proteins play a pivotal role in DNA demethylation and developmental reprogramming. Although the expression of Tet proteins and the abundance of their oxidative products are explosively studied (Szwagierczak et al., 2010; Tahiliani et al., 2009; Ito et al., 2011; Pastor et al., 2011; Williams et al., 2011), little is known about the regulation, interaction network and post-translational modification of Tet proteins. In our research, we investigated the interaction partner and modification of Tet proteins via series of *in vivo* and *in vitro* technologies (chapter 2.3).

We initially generated monoclonal antibodies against Tet proteins for cellular and biochemical applications. Using these tools, we immunoprecipitated (IP) Tet1/2 and analyzed by LC-MS/MS. Thus, a set of interaction proteins of Tet1 and 2 were identified. A dominant interaction partner, O-GlcNAc transferase (Ogt), was demonstrated. Through co-IP and F3H assay, we provided novel evidence for Ogt interaction with all three mouse Tet proteins (chapter 2.3). This is in contrast with previous researches finding that Ogt interacts with only one or two members of Tet family (Shi et al., 2013; Chen et al., 2012; Vella et al., 2013).

Functionally, Ogt is responsible for catalyzing the addition of O-linked N-acetylglucosamine (O-GlcNAc) to serine or threonine residues of proteins, and another enzyme O-GlcNAcase (Oga) catalyzes the selectively removal of O-GlcNAc from target substrates (Hanover et al., 2012; Butkinaree et al., 2010). In addition to Ogt, we also confirmed the interaction between Oga and Tets, which indicates that the O-GlcNAc cycling of Tet proteins is possible to occur.

In addition, for further understanding the details of the interaction between Tet proteins and Ogt/Oga, we fine mapped the interaction domains of Tet2 by *in vivo* F3H assay. Here, we identified that the catalytic domain of Tet2 is the main platform for interaction, which might indicates a correlation between the dioxygenase activity and O-GlcNAcylation for Tet proteins, in accordance with a previous research indicating that the DSBH domain of Tet2 interacts with Ogt (Chen et al., 2012).

O-GlcNAcylation is a widespread post-translational modification (PTM), which was reported to have a multitude of biological functions in cells, including responding to nutrient/stress availability (Love et al., 2010; Love and Hanover 2005), regulating the epigenome and high-order chromatin structure (Sakabe et al., 2010) and interplaying with other PTM for metabolic controlling of signaling transduction, gene transcription and cytoskeletal functions (Kamemura and Hart, 2003; Zeidan and Hart, 2010; Hart et al., 2011). PTM system, describing the chemical modification of a protein after its translation, largely increases the diversity of the proteome and is also involved in protein functions. Therefore the understanding of PTM regulation on proteins is a topic of interest and attracts a great deal of research effort.

In our study, O-GlcNAcylation modification of Tet proteins is confirmed by immunofluorescence *in vivo* and biochemical approaches *in vitro* (chapter 2.3). The interactors Ogt/Oga are shown to regulate the glycosylation status of Tet proteins. Consistent with Chen's result (Chen et al., 2012), we did not detect the impact of Ogt/Oga over-expression on genomic 5hmC abundance (data not shown). In a previous manuscript, O-GlcNAcylation was reported to be essential for Tet1 localization and 5hmC enrichment on Tet1-target genes (Shi et al., 2013). However, the biological function and mechanism of O-GlcNAcylation regulation on Tet proteins are still largely unclear. Hence, we will expand our experimental repertoire and test other hypothesis, such as the cross-talk among different PTMs.

Interplays between different PTMs drastically encode large amount of information. Cross-talk exists among acetylation, ubiquitination and O-GlcNAcylation, mostly controlling degradation of misfolding, damaged and unwanted proteins, and that is essential for cellular homeostasis (Ruan et al., 2013). For example, research in human demonstrated that DNMT1 can be acetylated by a protein lysine acetyltransferase 5 (KAT5) and this reaction triggers ubiquitination of DNMT1, which promotes proteasomal degradation of DNMT1. In contrast, DNMT1 is deacetylated by histone deacetylase 1 (HDAC1) and deubiquitinated by USP7, by which it protects and stabilizes DNMT1. These antagonistic regulation pathways link acetylation and ubiquitination on the same protein (Du et al., 2011). Conclusively, different PTMs might interplay each other in promoting or inhibiting manners, and carry out cooperative functions.

Our future plan includes the investigation of regulation link between ubiquitination, acetylation and O-GlcNAcylation for Tet proteins. Furthermore, both a research about the Tet1 and our own data from mass spectrometry, F3H assay and co-immunoprecipitation represented a significant protein interaction between histone deacetylase Hdac1/2 with Tet1/2/3 (William et al., 2011; unpublished data), which indicates involvement of acetylation on Tet proteins. Although the multifaceted post-translational modification is implied from our data, whether and how these PTMs serve for the Tet functions, stability and epigenetic system of cells still needs to be further elucidated.

4 ANNEX

4.1 Reference

Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., et al. (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood 114, 144-147.

Achour, M., Jacq, X., Ronde, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A.D., 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.

Agoston, A.T., Argani, P., Yegnasubramanian, S., De Marzo, A.M., Ansari-Lari, M.A., Hicks, J.L., Davidson, N.E., and Nelson, W.G. (2005). Increased protein stability causes DNA methyltransferase 1 dysregulation in breast cancer. J Biol Chem 280, 18302-18310.

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. EMBO J 25, 4503-4512.

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

Aoki, F., Worrad, D.M., and Schultz, R.M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. Dev Biol 181, 296-307.

Aravind, L., and Koonin, E.V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol 2, RESEARCH0007.

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

Ayton, P.M., Chen, E.H., and Cleary, M.L. (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.

Ayton, P.M., and Cleary, M.L. (2001). Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. Oncogene 20, 5695-5707.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27, 361-368.

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

Barta, T., Dolezalova, D., Holubcova, Z., and Hampl, A. (2013). Cell cycle regulation in human embryonic stem cells: links to adaptation to cell culture. Exp Biol Med (Maywood) 238, 271-275.

Barton, S.C., Arney, K.L., Shi, W., Niveleau, A., Fundele, R., Surani, M.A., and Haaf, T. (2001). Genome-wide methylation patterns in normal and uniparental early mouse embryos. Hum Mol Genet 10, 2983-2987.

Baylin, S.B., Hoppener, J.W., de Bustros, A., Steenbergh, P.H., Lips, C.J., and Nelkin, B.D. (1986). DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer Res 46, 2917-2922.

Beaujean, N., Hartshorne, G., Cavilla, J., Taylor, J., Gardner, J., Wilmut, I., Meehan, R., and Young, L. (2004). Non-conservation of mammalian preimplantation methylation dynamics. Curr Biol 14, R266-267.

Becker, K.A., Ghule, P.N., Therrien, J.A., Lian, J.B., Stein, J.L., van Wijnen, A.J., and Stein, G.S. (2006). Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. J Cell Physiol 209, 883-893.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-485.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev 23, 781-783.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315-326.

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

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum Mol Genet 9, 2395-2402.

Bestor, T.H. (2005). Transposons reanimated in mice. Cell 122, 322-325.

Bestor, T.H., and Verdine, G.L. (1994). DNA methyltransferases. Curr Opin Cell Biol 6, 380-389.

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

Bhutani, N., Decker, M.N., Brady, J.J., Bussat, R.T., Burns, D.M., Corbel, S.Y., and Blau, H.M. (2013). A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. FASEB J 27, 1107-1113.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev 16, 6-21.

Bird, A. (2007). Perceptions of epigenetics. Nature 447, 396-398.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res 30, 958-965.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science 336, 934-937.

Borst, P., and Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol 62, 235-251.

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

Braidman, I., Carroll, M., Dance, N., Robinson, D., Poenaru, L., Weber, A., Dreyfus, J.C., Overdijk, B., and Hooghwinkel, G.J. (1974). Characterisation of human N-acetyl-beta-hexosaminidase C. FEBS Lett 41, 181-184.

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., et al. (2004). Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat Biotechnol 22, 707-716.

Bronner, C. (2011). Control of DNMT1 abundance in epigenetic inheritance by acetylation, ubiquitylation, and the histone code. Sci Signal 4, pe3.

Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat Immunol 4, 235-240.

Busslinger, M., Hurst, J., and Flavell, R.A. (1983). DNA methylation and the regulation of globin gene expression. Cell 34, 197-206.

Butkinaree, C., Park, K., and Hart, G.W. (2010). O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. Biochim Biophys Acta 1800, 96-106.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643-655.

Chang, C.W., Lai, Y.S., Pawlik, K.M., Liu, K., Sun, C.W., Li, C., Schoeb, T.R., and Townes, T.M. (2009). Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells. Stem Cells 27, 1042-1049.

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

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

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R.J. (1993). Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74, 299-307.

Cheng, X., and Roberts, R.J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res 29, 3784-3795.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell 110, 33-42.

Choy, M.K., Movassagh, M., Goh, H.G., Bennett, M.R., Down, T.A., and Foo, R.S. (2010). Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. BMC Genomics 11, 519.

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

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

Cierpicki, T., Risner, L.E., Grembecka, J., Lukasik, S.M., Popovic, R., Omonkowska, M., Shultis, D.D., Zeleznik-Le, N.J., and Bushweller, J.H. (2010). Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat Struct Mol Biol 17, 62-68.

Cliffe, L.J., Kieft, R., Southern, T., Birkeland, S.R., Marshall, M., Sweeney, K., and Sabatini, R. (2009). JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. Nucleic Acids Res 37, 1452-1462.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36, W197-201.

Conticello, S.G. (2008). The AID/APOBEC family of nucleic acid mutators. Genome Biol 9, 229.

Conticello, S.G., Langlois, M.A., Yang, Z., and Neuberger, M.S. (2007). DNA deamination in immunity: AID in the context of its APOBEC relatives. Adv Immunol 94, 37-73.

Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007). The enigmatic thymine DNA glycosylase. DNA Repair (Amst) 6, 489-504.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature 470, 419-423.

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

Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R., and Feinberg, A.P. (2003). Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. Science 299, 1753-1755.

Dambacher, S., Deng, W., Hahn, M., Sadic, D., Frohlich, J., Nuber, A., Hoischen, C., Diekmann, S., Leonhardt, H., and Schotta, G. (2012). CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3, 101-110.

Daser, A., and Rabbitts, T.H. (2005). The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. Semin Cancer Biol 15, 175-188.

Davies, G.J., Gloster, T.M., and Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr Opin Struct Biol 15, 637-645.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell 24, 310-323.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell 9, 166-175.

Dean, W., Santos, F., and Reik, W. (2003). Epigenetic reprogramming in early mammalian development and following somatic nuclear transfer. Semin Cell Dev Biol 14, 93-100.

Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E., and Reik, W. (2001). Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc Natl Acad Sci U S A 98, 13734-13738.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J.P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in myeloid cancers. N Engl J Med 360, 2289-2301.

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A 105, 10762-10767.

Dinant, C., Bartek, J., and Bekker-Jensen, S. (2012). Histone Displacement during Nucleotide Excision Repair. Int J Mol Sci 13, 13322-13337.

Dizdaroglu, M. (2005). Base-excision repair of oxidative DNA damage by DNA glycosylases. Mutat Res 591, 45-59.

Dodge, J.E., Ramsahoye, B.H., Wo, Z.G., Okano, M., and Li, E. (2002). De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. Gene 289, 41-48.

Du, Z., Song, J., Wang, Y., Zhao, Y., Guda, K., Yang, S., Kao, H.Y., Xu, Y., Willis, J., Markowitz, S.D., et al. (2010). DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. Sci Signal 3, ra80.

Easwaran, H.P., Schermelleh, L., Leonhardt, H., and Cardoso, M.C. (2004). Replication-independent chromatin loading of Dnmt1 during G2 and M phases. EMBO Rep 5, 1181-1186.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. Science 300, 455.

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D.A. (2009). Conserved DNA methylation in Gadd45a(-/-) mice. Epigenetics 4, 98-99.

Engholm-Keller, K., and Larsen, M.R. (2013). Technologies and challenges in large-scale phosphoproteomics. Proteomics 13, 910-931.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., et al. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat Struct Mol Biol 15, 1176-1183.

Esteller, M. (2008). Epigenetics in cancer. N Engl J Med 358, 1148-1159.

Esteve, P.O., Chin, H.G., Benner, J., Feehery, G.R., Samaranayake, M., Horwitz, G.A., Jacobsen, S.E., and Pradhan, S. (2009). Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. Proc Natl Acad Sci U S A 106, 5076-5081.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-156.

Faber, J., Krivtsov, A.V., Stubbs, M.C., Wright, R., Davis, T.N., van den Heuvel-Eibrink, M., Zwaan, C.M., Kung, A.L., and Armstrong, S.A. (2009). HOXA9 is required for survival in human MLL-rearranged acute leukemias. Blood 113, 2375-2385.

Falnes, P.O., Johansen, R.F., and Seeberg, E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature 419, 178-182.

Fatemi, M., Hermann, A., Pradhan, S., and Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol 309, 1189-1199.

Feil, R. (2009). Epigenetic asymmetry in the zygote and mammalian development. Int J Dev Biol 53, 191-201.

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.

Fellinger, K., Rothbauer, U., Felle, M., Langst, G., and Leonhardt, H. (2009). Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. J Cell Biochem 106, 521-528.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science 330, 622-627.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 473, 398-402.

Fraga, M.F., and Esteller, M. (2002). DNA methylation: a profile of methods and applications. Biotechniques 33, 632, 634, 636-649.

Frauer, C., and Leonhardt, H. (2009). A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. Nucleic Acids Res 37, e22.

Frauer, C., and Leonhardt, H. (2011). Twists and turns of DNA methylation. Proc Natl Acad Sci U S A 108, 8919-8920.

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F., et al. (2011). Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. PLoS One 6, e16627.

Freitag, M., and Selker, E.U. (2005). Controlling DNA methylation: many roads to one modification. Curr Opin Genet Dev 15, 191-199.

Fujiki, R., Hashiba, W., Sekine, H., Yokoyama, A., Chikanishi, T., Ito, S., Imai, Y., Kim, J., He, H.H., Igarashi, K., et al. (2011). GlcNAcylation of histone H2B facilitates its monoubiquitination. Nature 480, 557-560.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem 278, 4035-4040.

Fulka, H., Mrazek, M., Tepla, O., and Fulka, J., Jr. (2004). DNA methylation pattern in human zygotes and developing embryos. Reproduction 128, 703-708.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol 196, 261-282.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124, 495-506.

Georgopoulos, C.P., and Revel, H.R. (1971). Studies with glucosyl transferase mutants of the T-even bacteriophages. Virology 44, 271-285.

Gjerset, R.A., and Martin, D.W., Jr. (1982). Presence of a DNA demethylating activity in the nucleus of murine erythroleukemic cells. J Biol Chem 257, 8581-8583.

Globisch, D., Munzel, M., Muller, M., Michalakis, S., Wagner, M., Koch, S., Bruckl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 5, e15367.

Gloster, T.M., and Vocadlo, D.J. (2010). Mechanism, Structure, and Inhibition of O-GlcNAc Processing Enzymes. Curr Signal Transduct Ther 5, 74-91.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74, 481-514.

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 tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311, 395-398.

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

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

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 339, 448-452.

Hackett, J.A., Zylicz, J.J., and Surani, M.A. (2012). Parallel mechanisms of epigenetic reprogramming in the germline. Trends Genet 28, 164-174.

Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P., et al. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep 3, 291-300.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech Dev 117, 15-23.

Hanover, J.A., Krause, M.W., and Love, D.C. (2012). Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. Nat Rev Mol Cell Biol 13, 312-321.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. Nat Genet 43, 768-775.

Hart, G.W., Slawson, C., Ramirez-Correa, G., and Lagerlof, O. (2011). Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. Annu Rev Biochem 80, 825-858.

Hashimoto, H., Hong, S., Bhagwat, A.S., Zhang, X., and Cheng, X. (2012a). Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. Nucleic Acids Res 40, 10203-10214.

Hashimoto, H., Zhang, X., and Cheng, X. (2012b). Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. Nucleic Acids Res 40, 8276-8284.

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., and Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. Nat Genet 34, 187-192.

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

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18, 6538-6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev 15, 710-723.

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401, 301-304.

Hendriks, W.J., Elson, A., Harroch, S., Pulido, R., Stoker, A., and den Hertog, J. (2013). Protein tyrosine phosphatases in health and disease. FEBS J 280, 708-730.

Hendriks, W.J., and Pulido, R. (2013). Protein tyrosine phosphatase variants in human hereditary disorders and disease susceptibilities. Biochim Biophys Acta 1832, 1673-1696.

Hermann, A., Gowher, H., and Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. Cell Mol Life Sci 61, 2571-2587.

Hino, S., Kishida, S., Michiue, T., Fukui, A., Sakamoto, I., Takada, S., Asashima, M., and Kikuchi, A. (2001). Inhibition of the Wnt signaling pathway by Idax, a novel Dvl-binding protein. Mol Cell Biol 21, 330-342.

Holland, P.J., and Hollis, T. (2010). Structural and mutational analysis of Escherichia coli AlkB provides insight into substrate specificity and DNA damage searching. PLoS One 5, e8680.

Holm, T.M., Jackson-Grusby, L., Brambrink, T., Yamada, Y., Rideout, W.M., 3rd, and Jaenisch, R. (2005). Global loss of imprinting leads to widespread tumorigenesis in adult mice. Cancer Cell 8, 275-285.

Holt, G.D., and Hart, G.W. (1986). The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc. J Biol Chem 261, 8049-8057.

Horton, J.R., Zhang, X., Maunus, R., Yang, Z., Wilson, G.G., Roberts, R.J., and Cheng, X. (2006). DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. Nucleic Acids Res 34, 939-948.

Hsieh, C.L. (1999). In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. Mol Cell Biol 19, 8211-8218.

Hu, J.L., Zhou, B.O., Zhang, R.R., Zhang, K.L., Zhou, J.Q., and Xu, G.L. (2009). The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. Proc Natl Acad Sci U S A 106, 22187-22192.

Huang, H., Jiang, X., Li, Z., Li, Y., Song, C.X., He, C., Sun, M., Chen, P., Gurbuxani, S., Wang, J., et al. (2013). TET1 plays an essential oncogenic role in MLL-rearranged leukemia. Proc Natl Acad Sci U S A 110, 11994-11999.

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.

Huang, Y., Pastor, W.A., Zepeda-Martinez, J.A., and Rao, A. (2012). The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. Nat Protoc 7, 1897-1908.

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., and Melton, D.A. (2008). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol 26, 1269-1275.

Huisinga, K.L., Brower-Toland, B., and Elgin, S.C. (2006). The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma 115, 110-122.

Ichiyanagi, K. (2012). Inhibition of MspI cleavage activity by hydroxymethylation of the CpG site: a concern for DNA modification studies using restriction endonucleases. Epigenetics 7, 131-136.

Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A 108, 3642-3647.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466, 1129-1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333, 1300-1303.

Itoh, K., Brott, B.K., Bae, G.U., Ratcliffe, M.J., and Sokol, S.Y. (2005). Nuclear localization is required for Dishevelled function in Wnt/beta-catenin signaling. J Biol 4, 3.

Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. Prog Mol Biol Transl Sci 101, 25-104.

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.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl, 245-254.

Jair, K.W., Bachman, K.E., Suzuki, H., Ting, A.H., Rhee, I., Yen, R.W., Baylin, S.B., and Schuebel, K.E. (2006). De novo CpG island methylation in human cancer cells. Cancer Res 66, 682-692.

Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem 3, 274-293.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell 128, 683-692.

Jorgensen, H.F., Ben-Porath, I., and Bird, A.P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol Cell Biol 24, 3387-3395.

Kamemura, K., and Hart, G.W. (2003). Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription. Prog Nucleic Acid Res Mol Biol 73, 107-136.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature 452, 112-115.

Karpf, A.R., and Matsui, S. (2005). Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. Cancer Res 65, 8635-8639.

Kaufmann, L.T., Gierl, M.S., and Niehrs, C. (2011). Gadd45a, Gadd45b and Gadd45g expression during mouse embryonic development. Gene Expr Patterns 11, 465-470.

Kersh, E.N., Fitzpatrick, D.R., Murali-Krishna, K., Shires, J., Speck, S.H., Boss, J.M., and Ahmed, R. (2006). Rapid demethylation of the IFN-gamma gene occurs in memory but not naive CD8 T cells. J Immunol 176, 4083-4093.

Khoury, G.A., Baliban, R.C., and Floudas, C.A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 1.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell 143, 313-324.

Kim, J.K., Esteve, P.O., Jacobsen, S.E., and Pradhan, S. (2009). UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. Nucleic Acids Res 37, 493-505.

Kimmins, S., and Sassone-Corsi, P. (2005). Chromatin remodelling and epigenetic features of germ cells. Nature 434, 583-589.

Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell 76, 357-369.

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijo, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lahdesmaki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature 497, 122-126.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468, 839-843.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 8, 200-213.

Kojima, T., Shimazui, T., Hinotsu, S., Joraku, A., Oikawa, T., Kawai, K., Horie, R., Suzuki, H., Nagashima, R., Yoshikawa, K., et al. (2009). Decreased expression of CXXC4 promotes a malignant phenotype in renal cell carcinoma by activating Wnt signaling. Oncogene 28, 297-305.

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

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.

Kreppel, L.K., and Hart, G.W. (1999). Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. J Biol Chem 274, 32015-32022.

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

Krivtsov, A.V., and Armstrong, S.A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer 7, 823-833.

Krogan, N.J., and Greenblatt, J.F. (2001). Characterization of a six-subunit holo-elongator complex required for the regulated expression of a group of genes in Saccharomyces cerevisiae. Mol Cell Biol 21, 8203-8212.

Kuehn, M.R., Bradley, A., Robertson, E.J., and Evans, M.J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature 326, 295-298.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., and Wilson, G.G. (1994). The DNA (cytosine-5) methyltransferases. Nucleic Acids Res 22, 1-10.

Kunz, C., Focke, F., Saito, Y., Schuermann, D., Lettieri, T., Selfridge, J., and Schar, P. (2009). Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. PLoS Biol 7, e91.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev 22, 1617-1635.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S.Y., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. Mol Cell Biol 25, 2475-2485.

Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A.G., Rechavi, G., and Cedar, H. (2007). Role of DNA methylation in stable gene repression. J Biol Chem 282, 12194-12200.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 35, 88-93.

Langemeijer, S.M., Kuiper, R.P., Berends, M., Knops, R., Aslanyan, M.G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A.G., Raymakers, R.A., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet 41, 838-842.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Le, T., Kim, K.P., Fan, G., and Faull, K.F. (2011). A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. Anal Biochem 412, 203-209.

Lee, B., and Muller, M.T. (2009). SUMOylation enhances DNA methyltransferase 1 activity. Biochem J 421, 449-461.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development 129, 1807-1817.

Lee, J.H., Voo, K.S., and Skalnik, D.G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. J Biol Chem 276, 44669-44676.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71, 865-873.

Leonhardt, H., Rahn, H.P., Weinzierl, P., Sporbert, A., Cremer, T., Zink, D., and Cardoso, M.C. (2000). Dynamics of DNA replication factories in living cells. J Cell Biol 149, 271-280.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915-926.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315-322.

Liu, N., Wang, M., Deng, W., Schmidt, C.S., Qin, W., Leonhardt, H., and Spada, F. (2013). Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules. PLoS One 8, e62755.

Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., and Klimasauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. Nat Chem Biol 5, 400-402.

Loenarz, C., and Schofield, C.J. (2008). Expanding chemical biology of 2-oxoglutarate oxygenases. Nat Chem Biol 4, 152-156.

Loenarz, C., and Schofield, C.J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. Trends Biochem Sci 36, 7-18.

Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia 17, 637-641.

Love, D.C., and Hanover, J.A. (2005). The hexosamine signaling pathway: deciphering the "O-GlcNAc code". Sci STKE 2005, re13.

Love, D.C., Krause, M.W., and Hanover, J.A. (2010). O-GlcNAc cycling: emerging roles in development and epigenetics. Semin Cell Dev Biol 21, 646-654.

Lubas, W.A., Frank, D.W., Krause, M., and Hanover, J.A. (1997). O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. J Biol Chem 272, 9316-9324.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol 13, 436-447.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251-260.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1, 55-70.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem 286, 35334-35338.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science 298, 1912-1934.

Margot, J.B., Aguirre-Arteta, A.M., Di Giacco, B.V., Pradhan, S., Roberts, R.J., Cardoso, M.C., and Leonhardt, H. (2000). Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. J Mol Biol 297, 293-300.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78, 7634-7638.

Matarese, F., Carrillo-de Santa Pau, E., and Stunnenberg, H.G. (2011). 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol Syst Biol 7, 562.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466, 253-257.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature 403, 501-502.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature 452, 45-50.

Michiue, T., Fukui, A., Yukita, A., Sakurai, K., Danno, H., Kikuchi, A., and Asashima, M. (2004). XIdax, an inhibitor of the canonical Wnt pathway, is required for anterior neural structure formation in Xenopus. Dev Dyn 230, 79-90.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell 10, 1107-1117.

Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G., and Hess, J.L. (2005). MLL associates specifically with a subset of transcriptionally active target genes. Proc Natl Acad Sci U S A 102, 14765-14770.

Mishina, Y., and He, C. (2006). Oxidative dealkylation DNA repair mediated by the mononuclear non-heme iron AlkB proteins. J Inorg Biochem 100, 670-678.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113, 631-642.

Morera, S., Grin, I., Vigouroux, A., Couve, S., Henriot, V., Saparbaev, M., and Ishchenko, A.A. (2012). Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA. Nucleic Acids Res 40, 9917-9926.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem 279, 52353-52360.

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). Epigenetic reprogramming in mammals. Hum Mol Genet 14 Spec No 1, R47-58.

Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M.C., and Leonhardt, H. (2005). Recruitment of DNA methyltransferase I to DNA repair sites. Proc Natl Acad Sci U S A 102, 8905-8909.

Mu, D., Park, C.H., Matsunaga, T., Hsu, D.S., Reardon, J.T., and Sancar, A. (1995). Reconstitution of human DNA repair excision nuclease in a highly defined system. J Biol Chem 270, 2415-2418.

Nabel, C.S., and Kohli, R.M. (2011). Molecular biology. Demystifying DNA demethylation. Science 333, 1229-1230.

Nady, N., Lemak, A., Walker, J.R., Avvakumov, G.V., Kareta, M.S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., et al. (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. J Biol Chem 286, 24300-24311.

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J., Matoba, R., Ko, M.S., et al. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. Mol Cell Biol 26, 7772-7782.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386-389.

Naveh-Many, T., and Cedar, H. (1981). Active gene sequences are undermethylated. Proc Natl Acad Sci U S A 78, 4246-4250.

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.

Neuberger, M.S., Harris, R.S., Di Noia, J., and Petersen-Mahrt, S.K. (2003). Immunity through DNA deamination. Trends Biochem Sci 28, 305-312.

Nguyen, A.V., Albers, C.G., and Holcombe, R.F. (2010). Differentiation of tubular and villous adenomas based on Wnt pathway-related gene expression profiles. Int J Mol Med 26, 121-125.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379-391.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24, 372-376.

Noma, K., Allis, C.D., and Grewal, S.I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 293, 1150-1155.

Nouspikel, T. (2009). DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. Cell Mol Life Sci 66, 994-1009.

Nunes-Xavier, C.E., Martin-Perez, J., Elson, A., and Pulido, R. (2013). Protein tyrosine phosphatases as novel targets in breast cancer therapy. Biochim Biophys Acta 1836, 211-226.

Okada, Y., Yamagata, K., Hong, K., Wakayama, T., and Zhang, Y. (2010). A role for the elongator complex in zygotic paternal genome demethylation. Nature 463, 554-558.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19, 219-220.

Olek, A., and Walter, J. (1997). The pre-implantation ontogeny of the H19 methylation imprint. Nat Genet 17, 275-276.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Res 62, 4075-4080.

Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. Cell 133, 1145-1148.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714-717.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol 10, 475-478.

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.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell 1, 299-312.

Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14, 341-356.

Pastor, W.A., Huang, Y., Henderson, H.R., Agarwal, S., and Rao, A. (2012). The GLIB technique for genome-wide mapping of 5-hydroxymethylcytosine. Nat Protoc 7, 1909-1917.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 473, 394-397.

Pei, D. (2009). Regulation of pluripotency and reprogramming by transcription factors. J Biol Chem 284, 3365-3369.

Pendino, F., Nguyen, E., Jonassen, I., Dysvik, B., Azouz, A., Lanotte, M., Segal-Bendirdjian, E., and Lillehaug, J.R. (2009). Functional involvement of RINF, retinoid-inducible nuclear factor (CXXC5), in normal and tumoral human myelopoiesis. Blood 113, 3172-3181.

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.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A 104, 6752-6757.

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

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463, 1101-1105.

Pradhan, M., Esteve, P.O., Chin, H.G., Samaranayke, M., Kim, G.D., and Pradhan, S. (2008). CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry 47, 10000-10009.

Pradhan, S., and Esteve, P.O. (2003). Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. Biochemistry 42, 5321-5332.

Qin, W., Leonhardt, H., and Pichler, G. (2011a). Regulation of DNA methyltransferase 1 by interactions and modifications. Nucleus 2, 392-402.

Qin, W., Leonhardt, H., and Spada, F. (2011b). Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. J Cell Biochem 112, 439-444.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell 135, 1201-1212.

Rao, F.V., Dorfmueller, H.C., Villa, F., Allwood, M., Eggleston, I.M., and van Aalten, D.M. (2006). Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. EMBO J 25, 1569-1578.

Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. Science 210, 604-610.

Reid, G., Metivier, R., Lin, C.Y., Denger, S., Ibberson, D., Ivacevic, T., Brand, H., Benes, V., Liu, E.T., and Gannon, F. (2005). Multiple mechanisms induce transcriptional silencing of a subset of genes, including oestrogen receptor alpha, in response to deacetylase inhibition by valproic acid and trichostatin A. Oncogene 24, 4894-4907.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., et al. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature 416, 552-556.

Roll, J.D., Rivenbark, A.G., Jones, W.D., and Coleman, W.B. (2008). DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines. Mol Cancer 7, 15.

Rossant, J., and Mummery, C. (2012). NOBEL 2012 Physiology or medicine: Mature cells can be rejuvenated. Nature 492, 56.

Roth, M., Helm-Kruse, S., Friedrich, T., and Jeltsch, A. (1998). Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the EcoRV adenine-N6-methyltransferase. J Biol Chem 273, 17333-17342.

Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F., and Leonhardt, H. (2010). The multi-domain protein Np95 connects DNA methylation and histone modification. Nucleic Acids Res 38, 1796-1804.

Rottach, A., Kremmer, E., Nowak, D., Leonhardt, H., and Cardoso, M.C. (2008). Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. Hybridoma (Larchmt) 27, 337-343.

Rottach, A., Leonhardt, H., and Spada, F. (2009). DNA methylation-mediated epigenetic control. J Cell Biochem 108, 43-51.

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

Ruan, H.B., Nie, Y., and Yang, X. (2013). Regulation of protein degradation by O-GlcNAcylation: crosstalk with ubiquitination. Mol Cell Proteomics.

Ruzov, A., Shorning, B., Mortusewicz, O., Dunican, D.S., Leonhardt, H., and Meehan, R.R. (2009). MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. Development 136, 2277-2286.

Saitou, M., and Yamaji, M. (2012). Primordial germ cells in mice. Cold Spring Harb Perspect Biol 4.

Sakabe, K., Wang, Z., and Hart, G.W. (2010). Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc Natl Acad Sci U S A 107, 19915-19920.

Sakai, Y., Suetake, I., Itoh, K., Mizugaki, M., Tajima, S., and Yamashina, S. (2001). Expression of DNA methyltransferase (Dnmt1) in testicular germ cells during development of mouse embryo. Cell Struct Funct 26, 685-691.

Sancar, A., Lindsey-Boltz, L.A., Unsal-Kacmaz, K., and Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem 73, 39-85.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol 241, 172-182.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A 103, 1412-1417.

Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. Nucleic Acids Res 35, 4301-4312.

Schoft, V.K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.S., Park, G.T., et al. (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci U S A 108, 8042-8047.

Scholer, H.R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. Nature 344, 435-439.

Scott, J. (1995). A place in the world for RNA editing. Cell 81, 833-836.

Sedgwick, B., Bates, P.A., Paik, J., Jacobs, S.C., and Lindahl, T. (2007). Repair of alkylated DNA: recent advances. DNA Repair (Amst) 6, 429-442.

Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature 463, 563-567.

Shafi, R., Iyer, S.P., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., and Marth, J.D. (2000). The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. Proc Natl Acad Sci U S A 97, 5735-5739.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450, 908-912.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. Cell 153, 692-706.

Shi, F.T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M.A., Wan, M., and Songyang, Z. (2013). Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. J Biol Chem 288, 20776-20784.

Shi, L., and Wu, J. (2009). Epigenetic regulation in mammalian preimplantation embryo development. Reprod Biol Endocrinol 7, 59.

Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Scholer, H.R., and Ding, S. (2008). A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell 2, 525-528.

Shimozaki, K., Nakashima, K., Niwa, H., and Taga, T. (2003). Involvement of Oct3/4 in the enhancement of neuronal differentiation of ES cells in neurogenesis-inducing cultures. Development 130, 2505-2512.

Sjolund, A.B., Senejani, A.G., and Sweasy, J.B. (2012). MBD4 and TDG: Multifaceted DNA glycosylases with ever expanding biological roles. Mutat Res.

Smith, M.L., Kontny, H.U., Zhan, Q., Sreenath, A., O'Connor, P.M., and Fornace, A.J., Jr. (1996). Antisense GADD45 expression results in decreased DNA repair and sensitizes cells to u.v.-irradiation or cisplatin. Oncogene 13, 2255-2263.

Sokol, S.Y. (2011). Maintaining embryonic stem cell pluripotency with Wnt signaling. Development 138, 4341-4350.

Song, C.X., Clark, T.A., Lu, X.Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C., and Korlach, J. (2012a). Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat Methods 9, 75-77.

Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013). Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. Cell 153, 678-691.

Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., et al. (2011a). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 29, 68-72.

Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011b). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331, 1036-1040.

Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012b). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335, 709-712.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Langst, G., and 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.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152, 1146-1159.

Stivers, J.T., and Jiang, Y.L. (2003). A mechanistic perspective on the chemistry of DNA repair glycosylases. Chem Rev 103, 2729-2759.

Sugiyama, Y., Hatano, N., Sueyoshi, N., Suetake, I., Tajima, S., Kinoshita, E., Kinoshita-Kikuta, E., Koike, T., and Kameshita, I. (2010). The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1delta/epsilon. Biochem J 427, 489-497.

Sun, L., Zhao, H., Xu, Z., Liu, Q., Liang, Y., Wang, L., Cai, X., Zhang, L., Hu, L., Wang, G., et al. (2007). Phosphatidylinositol 3-kinase/protein kinase B pathway stabilizes DNA methyltransferase I protein and maintains DNA methylation. Cell Signal 19, 2255-2263.

Surani, M.A. (2001). Reprogramming of genome function through epigenetic inheritance. Nature 414, 122-128.

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.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 16, 1779-1791.

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

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861-872.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-676.

Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A., and Tajima, S. (2011). Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A 108, 9055-9059.

Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development 139, 1895-1902.

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

Ting, A.H., Jair, K.W., Schuebel, K.E., and Baylin, S.B. (2006). Differential requirement for DNA methyltransferase 1 in maintaining human cancer cell gene promoter hypermethylation. Cancer Res 66, 729-735.

Torres, C.R., and Hart, G.W. (1984). Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. J Biol Chem 259, 3308-3317.

Torres-Padilla, M.E., Bannister, A.J., Hurd, P.J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int J Dev Biol 50, 455-461.

Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell 128, 651-654.

Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T., and Sedgwick, B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature 419, 174-178.

Turner, B.M. (2002). Cellular memory and the histone code. Cell 111, 285-291.

van der Heijden, G.W., Dieker, J.W., Derijck, A.A., Muller, S., Berden, J.H., Braat, D.D., van der Vlag, J., and de Boer, P. (2005). Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. Mech Dev 122, 1008-1022.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., et al. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell 49, 645-656.

Vincent, J.J., Huang, Y., Chen, P.Y., Feng, S., Calvopina, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., et al. (2013). Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. Cell Stem Cell 12, 470-478.

Voso, M.T., Santini, V., Finelli, C., Musto, P., Pogliani, E., Angelucci, E., Fioritoni, G., Alimena, G., Maurillo, L., Cortelezzi, A., et al. (2009). Valproic acid at therapeutic plasma levels may increase 5-azacytidine efficacy in higher risk myelodysplastic syndromes. Clin Cancer Res 15, 5002-5007.

Walder, R.Y., Langtimm, C.J., Chatterjee, R., and Walder, J.A. (1983). Cloning of the MspI modification enzyme. The site of modification and its effects on cleavage by MspI and HpaII. J Biol Chem 258, 1235-1241.

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., et al. (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet 41, 125-129.

Wang, S.C., and Frey, P.A. (2007). S-adenosylmethionine as an oxidant: the radical SAM superfamily. Trends Biochem Sci 32, 101-110.

Webster, K.E., O'Bryan, M.K., Fletcher, S., Crewther, P.E., Aapola, U., Craig, J., Harrison, D.K., Aung, H., Phutikanit, N., Lyle, R., et al. (2005). Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proc Natl Acad Sci U S A 102, 4068-4073.

Wernig, M., Meissner, A., Cassady, J.P., and Jaenisch, R. (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell Stem Cell 2, 10-12.

Williams, K., Christensen, J., and Helin, K. (2012). DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep 13, 28-35.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473, 343-348.

Wilson, D.M., 3rd, and Bohr, V.A. (2007). The mechanics of base excision repair, and its relationship to aging and disease. DNA Repair (Amst) 6, 544-559.

Winkler, G.S., Kristjuhan, A., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2002). Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. Proc Natl Acad Sci U S A 99, 3517-3522.

Woodcock, C.L., and Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. Curr Opin Genet Dev 11, 130-135.

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

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011a). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev 25, 679-684.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 473, 389-393.

Wu, H., and Zhang, Y. (2011). Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle 10, 2428-2436.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol 11, 607-620.

Wyatt, G.R. (1950). Occurrence of 5-methylcytosine in nucleic acids. Nature 166, 237-238.

Wyatt, G.R. (1951). Recognition and estimation of 5-methylcytosine in nucleic acids. Biochem J 48, 581-584.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol Cell 42, 451-464.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell 151, 1200-1213.

Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K., and Zhang, Y. (2013). Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res 23, 329-339.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature 492, 443-447.

Yang, C.G., Yi, C., Duguid, E.M., Sullivan, C.T., Jian, X., Rice, P.A., and He, C. (2008). Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. Nature 452, 961-965.

Yang, X., Zhang, F., and Kudlow, J.E. (2002). Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. Cell 110, 69-80.

Yen, R.W., Vertino, P.M., Nelkin, B.D., Yu, J.J., el-Deiry, W., Cumaraswamy, A., Lennon, G.G., Trask, B.J., Celano, P., and Baylin, S.B. (1992). Isolation and characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Res 20, 2287-2291.

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.

Yoder, J.A., Soman, N.S., Verdine, G.L., and Bestor, T.H. (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol 270, 385-395.

Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R., and Hunt, J.F. (2006). Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. Nature 439, 879-884.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, II, and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science 324, 797-801.

Yu, Z., Genest, P.A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J.M., Hausinger, R.P., et al. (2007). The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res 35, 2107-2115.

Zeidan, Q., and Hart, G.W. (2010). The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. J Cell Sci 123, 13-22.

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. (2010). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell Res 20, 1390-1393.

Zhang, R.P., Shao, J.Z., and Xiang, L.X. (2011). GADD45A protein plays an essential role in active DNA demethylation during terminal osteogenic differentiation of adipose-derived mesenchymal stem cells. J Biol Chem 286, 41083-41094.

Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., Thiry, S., Siegmann, M., and Jost, J.P. (2000a). 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. Nucleic Acids Res 28, 4157-4165.

Zhu, B., Zheng, Y., Hess, D., Angliker, H., Schwarz, S., Siegmann, M., Thiry, S., and Jost, J.P. (2000b). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. Proc Natl Acad Sci U S A 97, 5135-5139.

Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet 43, 143-166.

4.2 Abbreviation

2OGFeDO: 2-oxoglutarate and iron (II) dependent dioxygenase superfamily

5caC: 5-carboxylcytosine

5fC: 5-formylcytosine

5hmC: 5-hydroxymethylcytosine

5mC: 5-methylcytosine

AA: amino acid

AID: activation-induced cytidine deaminase

AlkB: Alpha-ketoglutarate-dependent dioxygenase

AML: acute myeloid leukemia

APOBEC1: apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1

BAH: bromo adjacent homology domain

BER: base excision repair

BGT: β-glucosyltransferase

CpG: cytosine-phosphatidyl-guanine

CMML: chronic myelomonocytic leukemia

C-MYC: Myc proto-oncogene protein

DNA: deoxyribonucleic acid

DNMT: DNA methyltransferase

E3-ligase: ubiquitin ligase

ELP: elongator complex protein

ES cells: embryonic stem cells

GADD45: growth arrest and DNA damage inducible protein 45

GFP: green fluorescent protein

H3K9me: H3 lysine 9 methyltransferase

HDAC: histone deacetylase

HMTs: histone methyltransferases

HP1: heterochromatin protein 1

KAT5: lysine acetyltransferase 5

MBD: methyl-CpG binding domain protein

MLL: mixed lineage leukemia

MPN: myeloproliferative neoplasms

MTase: 5mC methyltransferase

NER: Nucleotide excision repair

NP95: nuclear protein with 95 kilodalton

OCT4: octamer binding transcription factor 4

OGA: O-GlcNAcase

O-GlcNAc: O-linked N-acetylglucosamine

OGT: O-GlcNAc transferase

PBD: PCNA binding domain

PC center: prolylcysteinyl dipeptide center

PCNA: proliferating cell nuclear antigen

PDB: protein data bank

PHD: plant homeodomain

PGCs: primordial germ cells

Ring: really interesting new gene

PRC: polycomb repressive complex

PTM: post-translational modification

PWWP: proline-tryptophan-proline motif

SAH: S-adenosyl-L-homocysteine

SAM: S-adenosyl-L-methionine

S phase: synthesis phase

SRA: SET and Ring-associated

TDG: thymidine glycosylase

TET: ten-eleven translocation

TS: targeting sequence

UDP: Uridine diphosphate

UHRF1: ubiquitin-like, containing PHD and RING finger domains, 1

USP7: ubiquitin specific peptidase 7

ZnF: zinc finger

4.3 Contributions

Declaration of contributions to "Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules"

This study was initiated by Dr. Fabio Spada and me. In this project, I optimized and performed the measurement of genomic 5-hydroxymethylcytosine, prepared and contributed figure 6. The interaction between CXXC4 and Tet3 was detected via F3H assay in BHK cells, and I contributed and prepared figure 4. In addition, I performed *in vivo* FRAP analysis and live cell imaging and prepared figure S8. I performed the DNA binding assay of CXXC motifs and Tet proteins, preparing figure S6 and contributing figure 5. Nuclear localization of CXXC4 was determined in both stem cells and somatic cells, and I prepared and contributed figure S4. I performed western blot analysis for detecting CXXC4, Tet1 and Tet3, preparing and contributing to figure S9. I also contributed to the preparation of mouse tissues for RNA isolation, which contributed figure 3 and S5.

Declaration of contributions to "Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1"

I designed and constructed deletion mutants of the CXXC motif in Tet1, which contributed to figure 3. Furthermore, I investigated the *in vitro* DNA binding activity on both N-terminal and C-terminal GFP fusion versions of the CXXC domain, in order for the further clarify the DNA affinity of CXXC modules.

Declaration of contributions to "Ogt interacts with all three Tet proteins, mediates O-GlcNAcylation and links nutrients/metabolism to epigenetics"

I designed and constructed the truncation versions of Tet proteins, contributing figure 4 and 5. I performed the F3H assay for the fine mapping of Tet2 with Ogt and Oga proteins, identified the active interaction domain of Tet2 protein, and prepared figure 3. Furthermore, I expressed and purified the antigens of Tet2 and Tet3, contributing to antibodies testing.

Declaration of contributions to "The domain structure and cell cycle of Tet proteins"

I designed and performed all experiments and prepared the figures. I wrote the manuscript.

4.4 Declaration

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den ....

(unterschift)

Erklärung

Hiermit erkläre ich, *

☐ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist

☐ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

☐ dass ich mich mit Erfolg der Doktorprüfung im Hauptfach .....

und in den Nebenfächern .....

bei der Fakultät für ..... der .....

(Hochschule/ Universität)

unterzogen habe

☐ dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den ....

(unterschrift)

*) Nichtzutreffendes streichen

4.5 Acknowledgments

First of all, I would like to thank my doctor father and supervisor Prof. Dr. Heinrich Leonhardt for giving me the great opportunity to carry on my PhD research in your lab and for your supporting. It is always a great inspiration to following your supervision, kindly motivating and smart ideas for new experiments and projects. All these years of studying will guide me for my future academic and personal development.

Many thanks to Dr. Andrea Rottach for all the supervision, ideas and enthusiasm, especially during my thesis writing. Also I deeply appreciate Dr. Daniela Meilinger for the advices on my thesis. Thank you for your introduction about culturing embryonic stem cells, DNA methylation analysis and so on during my first two months in this lab. This guide helped me to get familiar with the new lab and our staff in a short time.

Special thanks go to the members of my office, Patricia Wolf, Katrin Schneider and Katharina Thanisch for the nice ambience. Thank Katrin deeply for always rescuing me from computer, printer and microscope problems. Thank Patricia and Katharina for many good advices and guidance for getting used to living in Munich.

I would like to thank colleagues in Heinrich's group for the friendship and support, particularly in the final stage of my study. I would like to thank Anja Gahl and Susanne Breitsameter for organizing the items in lab and thus the lab running smoothly. I am very grateful to Dr. Fabio Spada for his supervision, ideas and discussions; thank you for always spending time in answering my questions and providing helpful information. Many thanks to Dr. Weihua Qin, for your encouragement during my study. The same is true for Dr. Wen Deng, additionally thank you for creative discussions in science and guiding me through performing and analyzing F3H assay.

Furthermore, I am sincere grateful for all team members. Andreas, Boris, Carina, Christina, Congdi, Hsin-Yi, Irina, Jonas, Jürgen, Kamila, Martha, Nan, Sebastian, Sylvia, Tobias, Udo, Yolanda. Only teamwork makes work successful.

I would also like to thank Ola, Josef and Silvia. Thank you for your friendship and making me happy during the time I spend in Munich. It was really beautiful and joyful in the past four years. I will always miss this period of time in the future.

Last but not least, I thank my family for their support during all these years. Thank you for believing me, encouraging me and being always together with me. Mama and Papa, I am really proud of being your daughter and your love is the fundamental of my life.

5 Curriculum Vitae

Personal information:

Name: Mengxi Wang

[redacted]

Education:

10/2009 - present: PhD candidate in biology

Munich Ludwig Maximilians University, Munich
Department of biology II,
Under the supervision of Prof. Heinrich Leonhardt

09/2006 - 03/2009: Master student in Biomedical engineering

Aachen Rheinisch-Westfaelische Technische Hochschule (RWTH) Aachen
Aachen, Germany

10/2001 - 09/2005: Bachelor student in Biology

Beijing Peking University (PKU)
Beijing, China

Scientific experience and accomplishment

02/2003 - 08/2006: National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences (CAS)

Project: Investigation on G-protein for the aggregation, GTPase activity and function, as well as the interaction between regulator protein (RGS2) and G-protein, involving in Biochemistry, Molecular Biology, Molecular Imagine and Proteome techniques.

08/2007 - 12/2007: Internship

Helmholtz Institute of Biomedical Engineering, RWTH Aachen University, Germany

Project: Investigation on influence of cytokines, chemokines and MEF niche to the proliferation and multipotency of hematopoietic stem cell (HSC) *in vitro*, involving in Molecular Biology and Cell Biology techniques
02/2008 - 07/2008: HIWI student job
08/2008 – 01/2009: Master Thesis
Helmholtz Institute of Biomedical Engineering, RWTH Aachen University, Germany
Project: Reprogramming cell to pluripotent status by cell-cell fusion method and analysis of DNA methylation status for specific genes for the reprogrammed cells, involving in Biochemistry and Stem Cell Biology techniques.

List of publications:

1. Ni, J., Qu, L., Yang, H., **Wang, M.**, and Huang, Y. (2006). Palmitoylation and its effect on the GTPase-activating activity and conformation of RGS2. Int J Biochem Cell Biol 38, 2209-2218.
2. Yang, H., Qu, L., Ni, J., **Wang, M.**, and Huang, Y. (2008). Palmitoylation participates in G protein coupled signal transduction by affecting its oligomerization. Mol Membr Biol 25, 58-71.
3. Yang, H., Wan, L., Song, F., **Wang, M.**, and Huang, Y. (2009). Palmitoylation modification of Galpha(o) depresses its susceptibility to GAP-43 activation. Int J Biochem Cell Biol 41, 1495-1501.
4. Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., **Wang, M.**, Qin, W., Soding, J., Spada, F., et al. (2011). Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. PLoS One 6, e16627.
5. Liu, N.*, **Wang, M.***, Deng, W., Schmidt, C.S., Qin, W., Leonhardt, H., and Spada, F. (2013). Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. PLoS One 8, e62755.