



Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

DNA methylation and the core pluripotency network

Siba Shanak^{a,b}, Volkhard Helms^{b,*}^a Faculty of Science, Arab-American University, Jenin, Palestine^b Center for Bioinformatics, Saarland University, Saarbruecken, Germany

ARTICLE INFO

Keywords:

DNA methylation
Stem cells
Transcription factor
Differentiation

ABSTRACT

From the onset of fertilization, the genome undergoes cell division and differentiation. All of these developmental transitions and differentiation processes include cell-specific signatures and gradual changes of the epigenome. Understanding what keeps stem cells in the pluripotent state and what leads to differentiation are fascinating and biomedically highly important issues. Numerous studies have identified genes, proteins, microRNAs and small molecules that exert essential effects. Notably, there exists a core pluripotency network that consists of several transcription factors and accessory proteins. Three eminent transcription factors, OCT4, SOX2 and NANOG, serve as hubs in this core pluripotency network. They bind to the enhancer regions of their target genes and modulate, among others, the expression levels of genes that are associated with Gene Ontology terms related to differentiation and self-renewal. Also, much has been learned about the epigenetic rewiring processes during these changes of cell fate. For example, DNA methylation dynamics is pivotal during embryonic development. The main goal of this review is to highlight an intricate interplay of (a) DNA methyltransferases controlling the expression levels of core pluripotency factors by modulation of the DNA methylation levels in their enhancer regions, and of (b) the core pluripotency factors controlling the transcriptional regulation of DNA methyltransferases. We discuss these processes both at the global level and in atomistic detail based on information from structural studies and from computer simulations.

1. Main text

Recent review articles provide excellent introductions into the fields of pluripotency (Li and Belmonte, 2017; Martello and Smith, 2014), stem cell differentiation (Alvarado and Yamanaka, 2014; Dixon et al., 2015; Jang et al., 2017; Keller, 2005), cellular programming (Prasad et al., 2016; Shi et al., 2017; Weinberger et al., 2016), and of the epigenetic changes taking place upon these transitions. Here, we will focus on the current mechanistic understanding of how epigenetic modifications and drugs targeting epigenetic enzymes exert their effects at the atomistic level. We will limit the discussion to the physiological processes related to pluripotency and not discuss epigenetic aberrations that are detected in cultivated human pluripotent stem cells as reviewed by Bar and Benvenisty (2019). Due to space limitations and also our own research interests, we will concentrate on the discussion of DNA methylation changes and will largely omit the equally important field of histone marks. We will start by briefly reviewing what pluripotent cells and DNA methylation are. Then, we will introduce the players of the core pluripotency network, followed by the role of cytosine methylation on DNA conformation and on DNA-protein interactions. The review is rounded up

by a quick look at available medications and drug candidates targeting DNA methyltransferases.

2. Core pluripotency network

2.1. The pluripotent state

Pluripotent stem cells (PSCs) are cells that have the ability to self-renew indefinitely and to differentiate into any cell type of an adult creature. The developmental potential declines from the totipotent cell state following the onset of fertilization. Before blastocyst implantation, the epigenome of the developing embryo is transformed to the ground state of pluripotency (Reik et al., 2001), which is also termed the 'naïve' state. Cells that exhibit this form of pluripotency can be isolated from the inner cell mass (ICM) (Guo et al., 2016). ICM cells that are not exposed to the fluid cavity will adopt the epiblast fate and eventually differentiate into embryonic tissue. The pre-implantation epiblast is in the ground state of pluripotency (Nichols and Smith, 2009). However, after implantation, the epiblast switches into the so-called 'primed' state of pluripotency (Nichols and Smith, 2009) that is adopted by epiblast stem

* Corresponding author.

E-mail address: volkhard.helms@bioinformatik.uni-saarland.de (V. Helms).<https://doi.org/10.1016/j.ydbio.2020.06.001>

Received 9 January 2020; Received in revised form 1 May 2020; Accepted 4 June 2020

Available online 18 June 2020

0012-1606/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

cells, or epiSCs (Brons et al., 2007; Tesar et al., 2007). Subsequently, in the early germ line most parental imprints are erased in primordial germ cells (PGCs) (Lee et al., 2002).

Murine pluripotent cells have been extensively studied *in vivo*. In comparison, we only have an incomplete understanding of how human pluripotent cells behave *in vivo* (Iurlaro et al., 2017), which is in part due to ethical considerations that limit the accessibility of human ESCs. Thus, in an attempt to understand the processes of reprogramming and differentiation, human and murine cells are grown *in vitro* to recapitulate the ICM state. Naïve pluripotent embryonic stem cells (ESCs) and embryonic germ cells (EGCs) are derived from the pre-implantation epiblast, respectively (Leitch et al., 2013; Nichols and Smith, 2009, 2012). Historically, murine ESCs (mESCs) have been grown in a medium containing fetal bovine serum and leukemia inhibitory factor (LIF). This resulted in a hypermethylated population of cells that resemble the early epiblast in terms of epigenetic and transcriptional aspects, but that are nonetheless in a primed state of pluripotency. Epigenetic reprogramming is currently investigated using a (serum-to-2i) model where naïve hypomethylated mESCs are derived from primed hypermethylated mESCs in a serum-free medium, to which two small molecule inhibitors (2i), PD0325901 (PD) and CHIR99021 (CHIR) (Ying et al., 2008), are added (van den Berg et al., 2010). The former substance inhibits mitogen-activated protein kinase kinase (MAPKK) that targets fibroblast growth factor/extracellular signal-regulated kinase (FGF/ERK), whereas the latter one inhibits glycogen synthase kinase 3 (GSK3) signaling cascades and also activates signaling via the canonical Wnt pathway (Galonska et al., 2015). mESCs grown under these conditions are in their naïve state and are globally hypomethylated, but sustain methylation at imprinted regions (Ficz et al., 2013; Leitch et al., 2013).

Throughout this manuscript we will adopt the following nomenclature: Human gene symbols are generally italicized with all letters in upper-case (e.g., *SHH*). Murine gene symbols generally are italicized too, however, starting with a capital letter, whereas the remaining letters are in lower-case (*Shh*). In contrast, murine and human protein names are typically not italicized and all letters are upper-case (SHH). To distinguish human and murine proteins, we will use a short superscript prefix ^hSHH and ^mSHH, respectively, whenever needed.

Cells in the naïve state of pluripotency can differentiate into all three germ layers *in vitro*, form teratomas and chimaeric embryos *in vivo*, and

are capable of tetraploid complementation (Evans and Kaufman, 1981; Martin, 1981; Nagy et al., 1993). PSCs derived from post-implantation epiblasts (epiSCs) are in the primed state of pluripotency and can only contribute to chimaera formation in the post-implantation stage (Brons et al., 2007; Huang et al., 2012; Kojima et al., 2014; Rossant, 2008; Tesar et al., 2007).

It is believed that the process of reprogramming in (serum-to-2i) medium can yield a general molecular understanding of how the state of pluripotency is established. Essentially, pluripotency is promoted by epigenetic features such as global DNA hypomethylation, DNA hypermethylation of imprinted gene loci (Okita et al., 2007), silencing of retroviral transgenes (Cherry et al., 2000), reactivation of the X chromosome in female iPSCs, and reorganization of chromatin fibers (Fussner et al., 2011; Meshorer and Misteli, 2006), e.g. in the promoters and enhancers of genes that are regulated during development (Mikkelsen et al., 2008; Spivakov and Fisher, 2007). One crucial change during reprogramming is the reactivation of endogenous pluripotency genes (Jaenisch and Young, 2008; Takahashi et al., 2007; Takahashi and Yamanaka, 2006). Beyond this, Li & Belmonte recently reviewed various types of post-transcriptional controls, particularly those induced by RNA-binding proteins and alternative splicing, as a further important regulatory layer of pluripotency (Li and Belmonte, 2018). Fig. 1 illustrates schematically how epigenetic marks differ between the naïve and primed/differentiated states.

2.2. Core pluripotency factors

Core pluripotency factors are those key transcription factors and co-factors, such as microRNAs and other proteins that are either able to reprogram cells into the pluripotent state or maintain the cells in this state (Orkin et al., 2008). Experimental and computational studies have characterized a tightly connected set of core transcription factors that maintain murine ESC self-renewal under defined conditions (Chen et al., 2008; Dunn et al., 2014; MacArthur et al., 2012; Niwa et al., 2009). Three eminent transcription factors, ^{h/m}OCT4, ^{h/m}SOX2 and ^{h/m}NANOG, serve as hubs in this pluripotency network and are being used as genetic markers for induced pluripotent stem cells (iPSCs) (Huangfu et al., 2008; Park et al., 2008; Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Yu et al., 2007). ^mOCT4 (also called *pou5f1*) is the octamer-binding

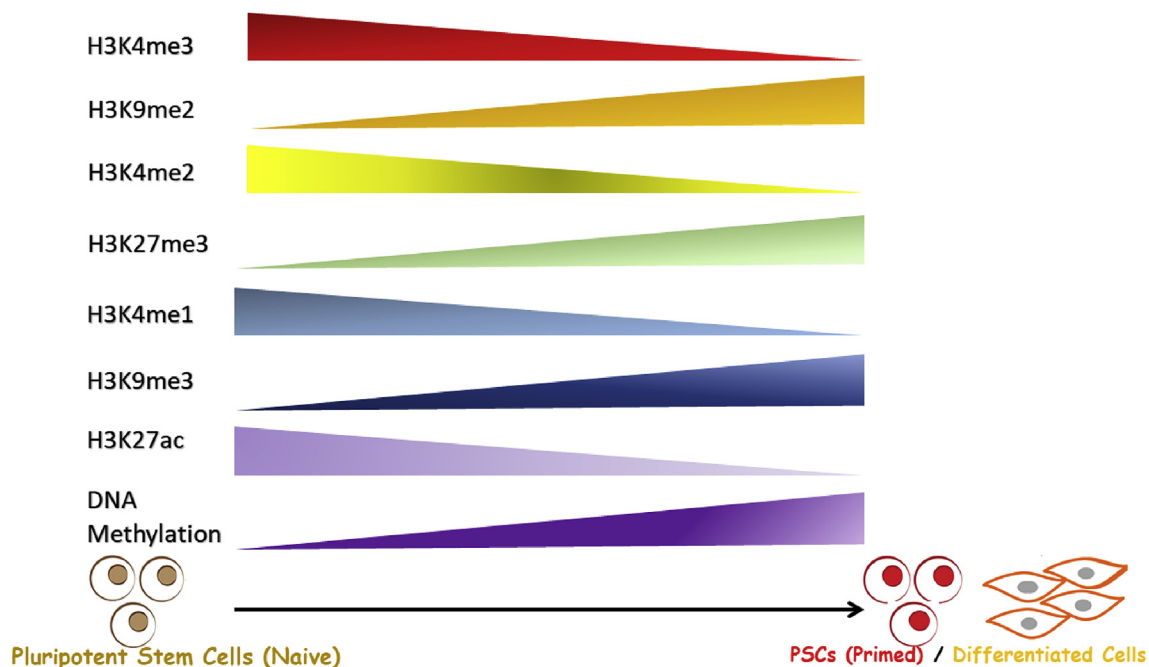


Fig. 1. Methylation/histone marks in the naïve, primed and differentiated states.

transcription factor 4 that is crucial for both *in vitro* and *in vivo* pluripotency (Nichols et al., 1998; Scholer et al., 1989). OCT4 is a member of the family of POU proteins, which consist of two well-conserved DNA-binding domains connected by a variable linker region (Tantin, 2013). ^mSOX2 (short for SRY-box 2) regulates OCT4 expression in ESCs and is involved in the process of epiblast formation (Avilion et al., 2003; Masui et al., 2007). The presence of ^mNANOG promotes the acquisition of pluripotency in the inner cell mass (Silva et al., 2009), and ^mNANOG is essentially involved in self-renewal of pluripotent cells (Mitsui et al., 2003). ^mNANOG also drives leukemia inhibitory factor (LIF)-independent self-renewal in pluripotent cells (Chambers et al., 2003, 2007; Mitsui et al., 2003; Silva et al., 2009; Suzuki et al., 2006). In addition, genome-wide binding of ^mNANOG alone can induce the state transition from naïve ESCs to primed epiblast-like cells independent of bone morphogenetic protein 4 (^mBMP4), and this is associated with epigenetic resetting of regulatory genes and activation of the enhancers of key germline transcription factors (Murakami et al., 2016). Fig. 2 illustrates the auto-regulatory and cross-regulatory effects among Oct4, Sox2, and Nanog.

Under specific conditions, some of these three core factors may also drive cell differentiation. For example, overexpression of Oct4 or Sox2 leads to differentiation of the germ layer. On the other hand, trophoderm differentiation is promoted by knocking out Oct4 or down-regulation of Sox2 (Ivanova et al., 2006; Nichols et al., 1998). The core pluripotency gene regulatory network involves protein-protein and protein-DNA interactions that may stabilize any of the two states, i.e., pluripotency or differentiation. One of the key interests in the stem cell field is to unravel what types of molecular changes (overexpression or downregulation of key genes and microRNAs) induce transitions between these two states (Boyer et al., 2005; Dunn et al., 2019).

Besides that, core pluripotency genes can also show differential enhancer binding in primed versus naïve pluripotency, where some factors switch from the Oct4 proximal (serum) to the distal (2i) element (Tesar et al., 2007). For example, upon being switched to 2i medium, the proximal enhancer of the Oct4 gene shows decreased binding to Nanog, whereas the distal enhancer shows an increased binding to Sox2 under the same conditions (Galonska et al., 2015). Whereas all three core transcription factors (^mOCT4, ^mSOX2 and ^mNANOG) are critically involved in maintaining the pluripotent state, there exist clear differences between NANOG on the one hand, and the other two transcription factors, ^mOCT4 and ^mSOX2, on the other hand. Oct4 and Sox2 are strongly expressed in both the naïve and primed-state pluripotency, whereas Nanog is only highly expressed in the naïve state of pluripotency (Chambers et al., 2007). Other murine pluripotency genes are also expressed in the absence of Nanog (Silva and Smith, 2008). When Nanog is knocked down in pluripotent stem cells, they can still self-renew but nonetheless have a high propensity for differentiation, unlike to

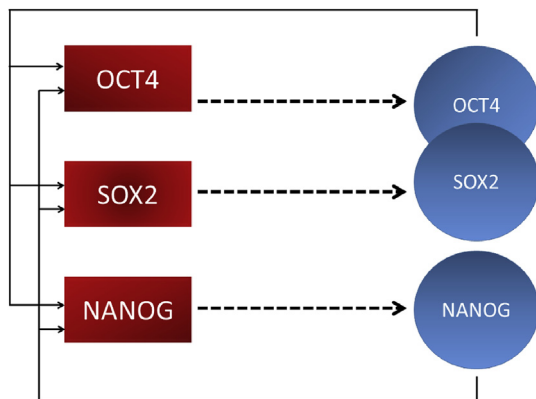


Fig. 2. Tight auto-regulatory and cross-regulatory feedback loops of the key pluripotency factors Oct 4, Sox2 and Nanog.

wild-type cells (Chambers et al., 2007; Festuccia et al., 2012). As such, Nanog seems to protect pluripotent cells from signals that induce differentiation.

Dunn and co-workers recently presented a computational model describing state transitions (on/off) in a core pluripotency network of 13 murine transcription factors (Dunn et al., 2019). They set up a so-called Boolean network that describes maintenance of naïve state ESCs. Interestingly, the model was also able to predict transcription factor behavior and potency during resetting from primed pluripotency. Computationally generated gene activation profiles were experimentally confirmed (with a predictive accuracy of 77%) at single-cell resolution by RT-qPCR.

2.3. Co-occupancy of core pluripotency transcription factors

The three core pluripotency transcription factors bind to the enhancer regions of their target genes and modulate, among others, the expression levels of genes that are associated with Gene Ontology terms related to differentiation and self-renewal (Young, 2011). Cooperative binding was found between ^{h/m}OCT4 and ^{h/m}SOX2, which were shown to form heterodimers (Chew et al., 2005), see below. The two transcription factors jointly bind and regulate the expression of many other genes (Chang et al., 2017), including the OSN triad itself in human and in mouse (OCT4, SOX2 and NANOG) (Okumura-Nakanishi et al., 2005; Rodda et al., 2005; Tomioka et al., 2002). Many murine genes regulated by ^mOCT4 and ^mSOX2 are also bound by ^mNANOG (Chen et al., 2008).

Molecular details of the formation of the ^mSOX2/^mOCT4 complex were explored via single-molecule imaging. First, ^mSOX2 dynamically engages with the target DNA sites and prepares them for ^mOCT4 binding. Then, binding of OCT4 stabilizes the heterodimeric ^mOCT4-^mSOX2 complex on the target binding sites (Chen et al., 2014). The highly conserved ^hOCT4:Lys156 residue provides stability to the ^hOCT4 protein, and a salt bridge between ^hOCT4:Lys151 and ^hSox2:Asp107 contributes to the stability of the ^hOCT4-^hSOX2 complex (Pan et al., 2016). In bladder cancer patients, post-translational modifications of Lys156 impaired the Lys151-Asp107 salt bridge and the ^hOCT4-^hSOX2 interaction. This resulted in the upregulation of mesendodermal genes and a subsequent epithelial-mesenchymal transition (Pan et al., 2016). The ^hOCT4-^hSOX2

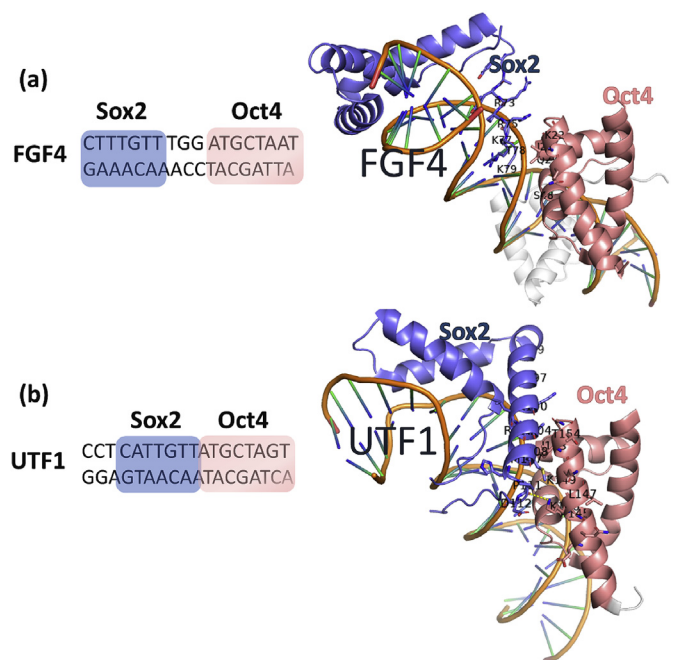


Fig. 3. Binding of the Oct4-Sox2 heterodimer to the target regulatory motifs; (a) the 3bp-separated motif (PDB ID: 1GT0 (Remenyi et al., 2003)), and (b) the no-gap motif (PDB ID: 6HT5).

heterodimer can bind to DNA in two alternative ways. In the first arrangement (Fig. 3(a)), the two TF binding motifs are separated by three base pairs. This conformation is seen, for example, in the *Fgf4/FGF4* promoter that is involved in embryonic development and morphogenesis (Jauch et al., 2011; Li and Belmonte, 2017; Tapia et al., 2015). In the other arrangement, the so-called no-gap motif, the binding motifs of ^hOCT4 and ^hSOX2 are arranged next to each other (Fig. 3(b)). This conformation is the canonical one found in the regulatory motifs of ^hOCT4, ^hNANOG, and ^hUTF1 (encoding undifferentiated embryonic cell transcription factor 1) (Pan et al., 2016), and it is crucial for somatic cell reprogramming and pluripotency in mouse and human (Li and Belmonte, 2017; Tapia et al., 2015).

Human and murine OSKM (OCT4, SOX2, KLF4, and MYC) are expressed both *in vitro* and *in vivo* in the inner cell mass of the blastocyst (Nichols and Smith, 2009). Several other transcription factors, cofactors, and co-repressors serve as additional layers of the pluripotency regulatory circuitry by occupying the regulatory sequences of hundreds of target genes, including their own promoters and enhancers. As a result, several cascades that comprise feedback and feedforward loops stabilize the pluripotent state (Adamo et al., 2011; Hackett and Surani, 2014; Young, 2011). Two prominent examples are the transcription factors ^mc-MYC and ^mKLF4. c-Myc is a proto-oncogenic target gene of ^mLIF-^mSTAT3 signaling that stimulates cell proliferation and self-renewal (Cartwright et al., 2005). Klf4 is a pluripotency factor whose expression may promote LIF-independent self-renewal (Niwa et al., 2009). Furthermore, ^mLIN28 regulates stem cell metabolism and conversion to the primed pluripotent state (Zhang et al., 2016a). β -catenin is a regulator of the WNT signaling cascade (Moon et al., 2002), and was shown to safeguard normal DNA methylation levels of murine ^mESCs and regulate genome stability (Theka et al., 2019). Esrrb is a ^mNANOG target that can replace ^mNANOG in murine NANOG-KO cells via inhibiting the Gsk3 signaling pathway (Festuccia et al., 2012). Krüppel-like factor 2 (^mKLF2) is a protein that is crucial in the naïve ground state of pluripotency. Ectopic expression of Klf2 can replace Mek/Erk inhibition in murine ^mESC (Yeo et al., 2014). ^hREX1 (short for reduced expression 1) is a protein whose addition to the reprogramming pool improves the reprogramming kinetics (Son et al., 2013). ^{h/m}STELLA is a pluripotency marker that is upregulated during the naïve state of pluripotency (Theunissen et al., 2014; Weinberger et al., 2016). ^mPRDM14 contributes to the regulation of pluripotency either by antagonizing the signaling of fibroblast growth factor receptor (^mFGFR), or by inhibiting synthesis of DNA methyltransferases (Yamaji et al., 2013).

2.4. Co-activators and co-repressors

During the processes of self-renewal of pluripotent cells and differentiation, a set of further transcriptional co-activators and co-repressors regulates the core transcription factors and aids them in their action. Some of these protein complexes do not bind to DNA directly but rather act via chromatin-mediated mechanisms. Examples for this are the protein complexes mediator and cohesin, which are crucial to the 3D genome organization, and facilitate physical as well as functional interactions between the enhancers and core promoters of the activated genes by forming chromatin loops in interphase nuclei of human and mouse (Gorkin et al., 2014; Kagey et al., 2010; Li et al., 2012). For example, binding of OCT4, SOX2, KLF4, mediator, and cohesin to the upstream enhancer of OCT4 facilitates the formation of contacts with the OCT4 promoter to induce expression of OCT4 (Kagey et al., 2010; Wei et al., 2013; Zhang et al., 2013). As shown by circular chromosome conformation capture with high-throughput sequencing (4C-seq), depletion of KLF4 causes breaking of the long-range interactions between enhancer and promoter, unloads cohesin off the enhancer, eliminates the OCT4 enhancer–promoter interaction, inhibits OCT4 expression, and

induces differentiation (Wei et al., 2013; Zhang et al., 2013). Furthermore, knockdown experiments of Med12 (component of mediator complex) or Smc1 (component of cohesin complex) in ESCs resulted in the repression of genes regulated through cohesin-mediator interactions (Phillips-Cremins et al., 2013).

Several other cofactors engage in this process by establishing protein-protein interactions with the core transcription factors. For example, the RNA polymerase-associated factor 1 (^mPAF1) complex physically interacts with the ^mOCT4 protein and thereby contributes to maintaining of self-renewal (Ponnusamy et al., 2009). The binding of the 60 kDa Tat-interactive protein (^mTIP60)–^mp400 chromatin-remodeling complex to its target promoters appears to be driven by the binding of ^mNANOG in ESCs and histone H3 lysine 4 tri-methylation (H3K4me3) signals separately (Fazio et al., 2008). Additionally, the co-repressors CCR4-NOT transcription complex subunit 3 (^mCNOT3) and tripartite motif-containing protein 28 (^mTRIM28) co-occupy many gene promoters together with ^mc-MYC and ^mZFX and thus aid in promoting self-renewal of embryonic stem cells. Nonetheless, ^mCNOT3 and ^mTRIM28 show no interaction with any component of the core pluripotency (OSN) triad, what suggests that a different module is active in the self-renewal network than in the core pluripotency module (Hu et al., 2009).

Estrogen-related receptor b (Esrrb) is one of the few prominent Δ Nanog-responsive genes, i.e., genes whose expression patterns are affected when cells lack Nanog. Normally, ^mNANOG binds to the Esrrb locus and promotes the binding of RNA Polymerase II and downstream Esrrb transcription. When Esrrb is overexpressed, pluripotency and self-renewal are preserved. However, it is noteworthy that Nanog(-/-) ESCs possess the same activity. Moreover, Esrrb can reprogram cells lacking Nanog (Festuccia et al., 2012). Orthodenticle homeobox 2 (Otx2) was recently shown to function during the transition from the naïve to primed states of pluripotency, whereas Otx2-null ESCs maintain the naïve pluripotent state. Nanog, Oct4, and Sox2 are direct targets of ^mOTX2. The strongest OTX2-binding site was found in the Nanog promoter, where ^mOTX2 enhances the expression of Nanog. When the amount of Nanog is low or is totally absent, this is causing cells to be redirected to the naïve state (Acampora et al., 2016).

The orphan nuclear hormone receptor ^mNR0B1 (Nuclear receptor subfamily 0, group B, member 1), also termed ^mDAX1, was identified as an OCT4-interacting protein in a yeast two-hybrid screen. ^mNR0B1 binds to the POU-specific domain of ^mOCT4. In ESCs, ^mNR0B1 acts as a repressor of Oct4 and abolishes its DNA binding activity. ^mNR0B1 also decreases the activities of Nanog and Rex1 promoters (Sun et al., 2009). Overexpression of Nr0b1 maintains self-renewal of pluripotent stem cells (Zhang et al., 2014). ^mNR0B1, alone or in cooperativity with ^mOCT4, inhibits trophectoderm differentiation. Both ESCs and induced pluripotent states are kept in the pluripotent state by the synergistic activities of ^mNANOG and ^mNR0B1 (Zhang et al., 2014). Testis-expressed sequence 10 protein (^mTEX10) has been revealed via immunoprecipitation as a novel transcription cofactor in the core pluripotency network that forms complexes with the ^mSOX2 protein in ESCs, and functions in ESC maintenance and efficient reprogramming. ^mTEX10 is enriched at ESC-specific distal enhancers and functions as a co-activator, where it modulates histone acetylation and DNA demethylation by recruiting mp300 and ^mTET1 (Ding et al., 2015). CBFA2/RUNX1 translocation partner 2 (^hCBFA2T2) is another co-repressor that is important for the regulation of pluripotency. ^hCBFA2T2 was found to interact and colocalize throughout the genome with the PR domain-containing 14 (^hPRDM14) (Tu et al., 2016), a pluripotency factor that regulates DNA methylation in mouse (Leitch et al., 2013; Yamaji et al., 2013). Overexpression of CBFA2T2 also enhances iPSC reprogramming efficiency in a similar manner as PRDM14. ^hCBFA2T2 functions synergistically with core pluripotency factors, where it oligomerizes to form a scaffold that stabilizes the binding of ^hOCT4 and ^hPRDM14 (Tu et al., 2016).

3. How does cytosine methylation affect DNA conformation and DNA-protein interactions?

3.1. DNA methylation of cytosine bases

In mammals, DNA methylation at the 5-position of cytosine (5 mC) plays a key role in various processes, including maintenance of genomic integrity, regulation of transcription, and genomic imprinting, X chromosome inactivation, as well as during reprogramming and differentiation (Bird, 2007; Bogdanovic and Lister, 2017; Goldberg et al., 2007; Sasaki and Matsui, 2008). DNA methylation of promoter regions is believed to generally inhibit gene expression, possibly through altering chromatin structure (Razin, 1998). Also, DNA methylation of enhancer regions is considered as a sign of transcriptional inactivity (Sharifi-Zarchi et al., 2017). On the other hand, so-called gene body methylation, i.e. high levels of DNA methylation in exons, is frequently encountered even for actively transcribed genes but so far not fully understood. Further below, we will summarize our current understanding of how DNA methylation affects the three-dimensional conformation of DNA and its dynamics.

With respect to the focus of this review article, DNA methylation dynamics is also pivotal during embryonic development and in the process of reprogramming murine stem cells to the naïve state of pluripotency, where global demethylation is a cornerstone (Leitch et al., 2013). Initially, oocyte and sperm show intermediate to high levels of DNA methylation. After fertilization and before the two nuclei merge, the genomes inherited from both parent mice undergo global demethylation (Messerschmidt et al., 2014). In murine and human PGCs, global DNA demethylation is coupled to the erasure of histone H3K9 di-methylation (H3K9me2) (Eguizabal et al., 2016) and strong gains of histone H3K27 tri-methylation (H3K27me3) (Seki et al., 2005, 2007). Yet, some loci and regulatory elements, such as imprinted genes and transposable elements, maintain their DNA methylation levels (Hirasawa et al., 2008; Sasaki and Matsui, 2008). Indeed, DNA methylation has a pivotal role in reshaping the 3D genome structure and in the involvement of polycomb complexes in 3D genome re-organization in naïve pluripotency (McLaughlin et al., 2019).

Global genome DNA demethylation during reprogramming and its subsequent re-methylation are orchestrated processes resulting from the dynamic interplay between three main routes, (1) *de novo* methylation by the DNA methyltransferases ^mDNMT3a/^mDNMT3b, whereby new methyl marks are acquired, (2) maintenance methylation by ^mDNMT1, whereby methyl marks are maintained during replications, and (3) active demethylation. Active demethylation is a replication-independent mechanism that involves the action of Ten-eleven translocation (^mTET) protein-mediated iterative oxidation (He et al., 2011; Ito et al., 2010, 2011; Tahiliani et al., 2009), followed by the excision of oxidized bases by thymine DNA glycosylase (^{h/m}TDG) (He et al., 2011; Kohli and Zhang, 2013; Maiti and Drohat, 2011). We will now summarize our current structural understanding of the protein-DNA complexes associated with the processes just mentioned.

3.2. Protein-DNA complexes

Proteins recognize specific DNA sequences by two general mechanisms: The ‘direct readout’ mechanism involves the formation of hydrogen-bonds between the side chains of amino acids and hydrogen-bond donor and acceptor atoms of the target nucleotide bases. For example, bases in the major groove were found to have distinctive hydrogen-bond signatures (Garvie and Wolberger, 2001). On the other hand, the ‘indirect readout’ mechanism encompasses deviations from the canonical B-DNA conformation and subsequent conformational changes that optimize the protein-DNA interface (Otwinowski et al., 1988; Travers, 1989). In this readout, no direct contacts of DNA bases are needed with the target protein (Rohs et al., 2009).

3.3. Structural insights on the DNA maintenance methyltransferase DNMT1

^{h/m}DNMT1 is active during the S phase of the interphase/cell cycle when it copies methylation patterns from the parental strand to the newly synthesized daughter strand (Esteve et al., 2011; Leonhardt et al., 1992). ^hDNMT1 is recruited to replication foci by the proliferating cell nuclear antigen (PCNA) and other factors including ^hUHRF1 (Esteve et al., 2006). Von Meyenn et al. found that the major drivers of global demethylation in naïve pluripotent stem cells are of passive rather than active nature (von Meyenn et al., 2016). In this respect, impaired maintenance of pre-existing methyl groups takes place when ^mUHRF1 and ^mDNMT1 are repressed at the protein level. Additionally, ^mUHRF1 mutants that are unable to bind H3K9me2/3 cannot recruit ^mDNMT1 to the replication foci (Liu et al., 2013). ^hDNMT1 interacts directly with ^hSUV39H1, ^hSUV39H2 or ^hG9a, and this interaction may function in recruiting both histone methyltransferases to suitable binding sites during replication (Esteve et al., 2006).

As suggested by the available crystal structures of DNMT1:DNA complexes, ^{h/m}DNMT1 appears to establish a balance between auto-inhibitory and active mechanisms in DNA recognition, thus ensuring that ^{h/m}DNMT1 catalyzes methylation of the hemi-methylated form of DNA, but not of the unmethylated form (Song et al., 2011). When viewed from N-terminus to C-terminus, ^hDNMT1 is composed of an N-terminal regulatory domain, a conserved (Gly-Lys)_n repeat, and a C-terminal methyltransferase domain. The regulatory N-terminal domain consists of a nuclear localization sequence, sequences responsible for the interaction of ^hDNMT1 with other proteins (Chuang et al., 1997), a domain responsible for allocating DNMT1 to the DNA replication fork (Leonhardt et al., 1992), two bromo-adjacent homology (BAH) domains that play an important role by linking DNA methylation, replication as well as transcriptional regulation (Callebaut et al., 1999), and a zinc finger CXXC (Cys-X-X-Cys) domain that recognizes unmethylated DNA sequences with high specificity (Pradhan et al., 2008). The C-terminal methyltransferase domain contains two subdomains, the catalytic domain and the target recognition domain (TRD).

Upon formation of a complex between DNMT1 and hemimethylated DNA, the TRD domain ‘leans’ toward the DNA major groove by about 2–3 Å (Song et al., 2012) in comparison to the unbound conformation of DNMT1. This conformational transition induces an opening in the central dinucleotide step of DNA, guanine (G7) is translated by one step along the DNA helix toward the 3’ end, and the downstream residue C8 is flipped out of the DNA helical conformation. The catalytic loop of DNMT1 penetrates into the DNA from the minor groove via the side chain of Met1235, and occupies the space vacated by the extruded flipped out base fC7’ carbon atom on the target strand (Song et al., 2012). Song et al. also determined structures of partial mouse and human DNMT1 bound to unmethylated DNA. Specific binding of the CXXC to an unmethylated CpG dinucleotide was found to induce structural changes such that the CXXC–BAH1 linker is positioned between the active site of DNMT1 and the DNA sequence. As a result, DNA methylation cannot take place. In addition, the target recognition domain (TRD) is inhibited when a loop of BAH2 interacts with it, thus preventing it from penetrating into the major groove of DNA (Song et al., 2011).

3.4. Structural insights on DNA *de novo* methyltransferases DNMT3a/3b/3L

^mDNMT3a and ^mDNMT3b have major roles in *de novo* methylation in mammalian germ cells (La Salle and Trasler, 2006) and during development. ^hDNMT3L is a paralogue of these two proteins. It is enzymatically inactive because it cannot bind the cofactor S-adenosyl-L-methionine at its methyltransferase catalytic domain (CD). A crystal structure determined by Jia and colleagues shows ^hDNMT3a and ^hDNMT3L in a tetrameric configuration (DNMT3L–DNMT3a–DNMT3a–DNMT3L) whereby two ^hDNMT3a–^hDNMT3L interfaces and one

^hDNMT3a-^hDNMT3a interface are formed. ^hDNMT3L plays a role in stabilizing the active loop of ^hDNMT3a. The two active sites of the ^hDNMT3a-^hDNMT3a homodimer are separated by a single helical turn of the DNA (Jia et al., 2007). ^hDNMT3L was found to be capable of activating DNA methyltransferase (Jia et al., 2007) and binding to the unmethylated lysine 4 on the histone 3 tail (Ooi et al., 2007) via the ATRX-DNMT3-DNMT3L (ADD) domain of ^hDNMT3a, thus recruiting it to the chromatin (Otani et al., 2009). Guo et al. determined crystal structures of DNMT3a-DNMT3L in the active form when being bound to H3, and in the inactive form. When H3 is not bound to the DNMT3a-DNMT3L dimer, the ADD domain interacts with the CD domain of DNMT3a and inhibits it. However, the presence of H3 interferes with this interaction and blocks the autoinhibitory process (Guo et al., 2015).

DNMT3b seems to be more important for embryonic development than DNMT3a. This is suggested by a study conducted by Okano and colleagues, where DNMT3a-null mice could survive until delivery, whereas no DNMT3b-null viable mice were retrieved at birth (Okano et al., 1999). Additionally, DNMT3b was found to assist the differentiation process of human ES cells. The *de novo* methyltransferases ^mDNMT3a/b are downregulated in naïve pluripotent cells, and there is a minor role for ^mTET1-3 mediated demethylation affecting a short portion of the genome (van den Berg et al., 2010). Reprogramming into the induced state of pluripotency was activated in knockdown experiments of ^hDNMT3b via an ectopic expression of the four pluripotency genes, *OCT4*, *SOX2*, *c-MYC*, and *KLF4* (Wongtrakongate et al., 2014). The C-terminal CD domain and the N-terminal ADD domain are not exclusive to ^hDNMT3a, but are also part of the ^hDNMT3b protein. In addition, the two proteins possess a PWWP domain, which is a member of the Royal superfamily of domains that bind simultaneously to histone ^hH3K36me3 and DNA via a conserved “aromatic cage”. Additional layers of DNA methylation as well as a crosstalk with other epigenetic marks can take place, as the PWWP domain can cooperate with other DNA reader or modifier proteins (Qin and Min, 2014). Rondelet et al. determined the first crystal structure of the PWWP domain of ^hDNMT3b in complex with histone ^hH3K36me3. They revealed a crucial conserved water molecule that mediates the interaction between trimethylated Lys36 and the Ser270 residue of ^hDNMT3b. The trimethyl ammonium group is stabilized in the cage by four conserved residues, a phenylalanine, two tryptophan residues and an aspartate (Rondelet et al., 2016) that are conserved in the PWWP domains of ^hDNMT3a and ^hDNMT3b.

3.5. Structural insights into Tet-mediated demethylation

Three TET (ten-eleven translocation) proteins, ^mTET1-TET3, have been identified so far and were shown to possess 5 mC to 5hmC oxidizing activity (Ito et al., 2010; Tahiliani et al., 2009). Although replication-dependent passive dilution is the major driving factor for zygotic DNA demethylation, ^mTET3 was found to actively contribute to the demethylation process as long as DNA replication takes place, especially in maternal pro-nuclei (Shen et al., 2014), and to some degree in paternal pro-nuclei as well (Wossidlo et al., 2011). This finding supports the crucial role of active demethylation (Okamoto et al., 2016) during the erasure of CpG methylation (5 mC) in PGCs, which occurs via the conversion to 5-hydroxymethylcytosine (5hmC) by TET1 and TET2. Increased expression of TET1, TET2, and transiently elevated levels of 5hmC have been reported in both human and mouse PGCs (Tang et al., 2015). Mice knockouts for either Tet1 or Tet2 are viable. Additionally, Tet1 and Tet2 double-knockout ESCs remain pluripotent (Dawlaty et al., 2013).

Global DNA demethylation in PGCs occurs passively in one stage. In a next stage, ^mTET-mediated locus-specific DNA demethylation affects imprinting control regions (ICRs) and meiotic genes. The two-stage mechanism maintains the ability to transmit DNA from parent to offspring (Hargan-Calvopina et al., 2016). The Tet-mediated oxidation of 5 mC is followed by the excision of the oxidized bases by thymine DNA glycosylase (^mTDG). In a study by Guo and colleagues, the demethylation

process was unaffected by the deletion of ^mTDG from the zygote, which hints at the involvement of other demethylation mechanisms that are TDG-independent (Guo et al., 2014).

Full-length TET1 and TET3 of human and mouse have a CXXC-type zinc finger domain at the amino terminus, whereas the CXXC domain originally belonging to TET2 became a separate gene encoding the IDAX (or CXXC4) protein (Hu et al., 2013; Pastor et al., 2013). In mouse, expression of full-length *Tet1* is limited to early developmental stages whereas somatic cells express a truncated isoform of *Tet1* lacking the N-terminal CXXC domain (Zhang et al., 2016c). The truncated form of ^mTET1 has weaker demethylation capacity and a reduced level of chromatin binding compared to the full-length form (Zhang et al., 2016c). In human, the truncated form of *TET1* without the CXXC domain is repressed in embryonic stem cells, but is active in embryonic and adult tissues. Two truncated forms of *Tet3* exist in mouse. One of them, the *Tet3o* isoform, is exclusively expressed in oocytes. The full-length isoform containing the CXXC domain (*Tet3FL*) may play a crucial protective role against neurodegeneration (Jin et al., 2016).

At their C-terminal ends, ^{h/m}TET proteins have a catalytic domain, which is composed of a double-strand β -helix (DSBH) and a domain rich in cysteines. The catalytic domain contains a nuclear localization sequence and can oxidize 5 mC (Hu et al., 2013; Ito et al., 2010; Tahiliani et al., 2009). ^hTET proteins are Fe²⁺/ α -ketoglutarate (Fe(II)/ α -KG)-dependent dioxygenases. In a crystal structure of ^hTET2 bound to methylated DNA, two of the three zinc fingers bring the cysteine-rich domain and the DSBH domain into close proximity. The catalytic cavity contains Fe(II) and an analog of α -KG. The 5 mC inserts into the cavity and is positioned close to the Fe(II) to be ready for oxidation (Hu et al., 2013).

^hTET2 was shown to specifically recognize CpG dinucleotides in different oxidation forms although the 5 mC methyl group apparently does not form direct TET2-DNA contacts (Hu et al., 2013). DNA bound to the binding interface of ^hTET2 is pushed into a narrow groove formed by loops L1 and L2 surrounding the DSBH core. The TET2-DNA binding interface is rich in hydrophobic interactions and hydrogen-bonds. A water-mediated hydrogen-bond is formed between the guanidinium group of Arg1262 of ^hTET2 and methylated cytosine (mC6) (mediated by the oxygen in the pentose ring of the C6 nucleotide). Further hydrogen-bonds are established between the phosphate groups of mC6 and neighboring nucleotides and lysine, arginine, and serine residues at the TET2 interface. In addition, hydrophobic contacts exist in the minor groove between C5:G5' and G8:C8'. The side chains of Met1293 and Tyr1294 push G6' out of its normal base stacking position between G5' and mC7' and take its place. As a result, flipping of mC6 out of the duplex renders the methylated cytosine “vulnerable” in the catalytic pocket (Hu et al., 2013).

3.6. Methyl-CpG-binding proteins

Methyl-CpG-binding proteins that recognize methylated base pairs are called ‘reader’ or ‘effector’ proteins. Reader proteins can be classified into three families based on the domain type with which they interact with DNA (Liu et al., 2014): The first family are the methyl-CpG-binding domain (^{h/m}MBD) proteins MBD1, MBD2, MBD4, and MeCP2 that recognize methylation in fully methylated CpG dinucleotides (Zou et al., 2012). The second family are the SET and RING-finger associated (SRA) domain proteins that recognize hemi-methylated DNA sequences. This family includes ^mURHF1 (its human ortholog is called ^hICBP90) and ^hURHF2 (Hashimoto et al., 2009). The third family includes Kaiso and Kaiso-like C2H2 zinc-finger proteins, e.g., ^hZBTB4, ^hZBTB38, ^hZFP57, and ^hKLF4. Proteins in this family preferentially bind to methylated CpG within a longer specific DNA sequence (Filion et al., 2006; Lopes et al., 2008).

The protein methyl-binding CpG domain protein 2 (^hMBD2) was found to play an indispensable role for the two opposing processes of pluripotency and differentiation. Therefore, one may wonder how this is achieved. In fact, via alternative splicing, ^hMBD2 actually codes for two

protein isoforms, ^hMBD2a and ^hMBD2c, that each perform a discordant function (Lu et al., 2014). Both ^hMBD2a and ^hMBD2c are enriched at the promoter regions of the core pluripotency genes, ^hOCT4, and ^hSOX2 (Lu et al., 2014). The OCT4 protein, a splicing regulator termed serine-arginine-rich splicing factor 2 (SRSF2) and splice products of MBD2 (MBD2a and MBD2c) participate in a positive feedback loop that stabilizes a self-renewing pluripotent ground state of the cell. Remarkably, decreased levels of SRSF2 or of OCT4 in hESCs led to an increase in the MBD2a isoform, but also to a decrease in MBD2c. MBD2a promoted hPSC differentiation via the interaction with repressive NuRD chromatin remodeling factors, whereas the expression of MBD2c boosted reprogramming of fibroblasts to pluripotency (Lu et al., 2014).

Both direct and indirect readouts appear to be involved in the recognition at the protein-DNA binding interface when ^{h/m}MBD binds to the major groove of methylated DNA. X-ray analysis of a methylated DNA-MBD complex of the MBD domain of MeCP2 with methylated DNA revealed (Ho et al., 2008) tightly bound crystal water molecules, or ‘structural waters’, at the binding interface, which contributed to a total of 23 hydrophilic contacts with the two methylated cytosines. It is worth pointing out that two of the structural waters formed CH...O interactions (Gu et al., 1999) with the two methyl groups of the mCpG dinucleotide pair (Ho et al., 2008) meaning that the mC5 methyl groups form water-bridged hydrophilic contacts with ^hMBD. This is unlike an MBD1:DNA complex, where the methylated CpG site makes contacts with five conserved protein residues forming a hydrophobic patch on the MBD1 surface (Ohki et al., 2001).

Furthermore, in the atomistic structures of MBD2–mDNA and MeCP2–mDNA complexes from chicken, arginine residues at the binding interfaces stacked with the methyl-cytosines and formed hydrogen-bonds with the adjacent guanine in the mCpG step (Scarsdale et al., 2011). This arrangement has been termed ‘stair motif’ and involves two conserved arginines (Zou et al., 2012). A molecular dynamics simulation study aimed at revealing further mechanistic principles through which ^hMBD recognizes methylated DNA (Zou et al., 2012). In contrast to experiments, such computer simulations are able to compare the physiological scenario where ^hMBD binds to methylated DNA to the hypothetical case when ^hMBD is put into contact with non-methylated DNA. In simulations of the complex with methylated DNA, hydrogen-bonds between Arg22 and Arg44 and the guanine bases of the mCpG step showed very small fluctuations and shifted the guanines to the minor groove. Thus, the stacking between guanine and 5-methyl-cytosine was reduced, whereas that of the arginines with either base was increased. Also, in the methylated form, tighter hydrogen-bonds having smaller length fluctuations were established between the arginines and cytosines (root mean square deviation (RMSD) 1.0 Å and 0.8 Å, respectively) than in the hypothetical complex with non-methylated DNA (RMSD 1.8 Å and 1.5 Å) (Zou et al., 2012).

Schenkelberger et al. studied the role of the methyl-CpG binding domain of MeCP2 as a potential transcriptional modulator of the *BDNF* (Brain-Derived neurotrophic factor) promoter (Schenkelberger et al., 2017). In a cell-free expression translation extract from *E. coli*, ^hMBD2 functioned as a specific methylation- and sequence-dependent inhibitor. According to molecular dynamics simulations that were part of this study, we noticed changes in the fractional occupancy of the B-DNA conformer, as well as in the handedness and twisting of DNA in the methylated form upon MBD2 binding. These are characteristic for cooperative conformational transitions in the promoter region. Also, lower handedness values of DNA were adopted upon binding (untwisting took place), and the B-DNA conformation was partially disrupted altogether. The major groove width also reduced upon binding (Schenkelberger et al., 2017). In another purely computational study, we showed that methylation entropically favors the binding of the MBD domain of the human MeCP2 protein to C5-cytosine methylated DNA. The contribution of the binding enthalpy (bonded interactions, electrostatic interactions, van der Waal interactions, Poisson Boltzmann solvation, and surface area contribution) was found to be very small (Shanak et al.,

2017). Fig. 4 displays a representative snapshot from the molecular dynamics simulations illustrating the contact surface formed between MeCP2 and the DNA.

3.7. Mechanistic insights into UHRF1 binding

^{h/m}UHRF1 is a multi-domain protein that binds to methylated DNA with its SRA domain via a base-flipping mechanism (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). With its tandem Tudor domain, ^hUHRF1 also binds to methylated Lys9 on histone ^hH3 (H3K9me2) (Rottach et al., 2010). In mouse, mutations that targeted the two domains resulted in decreased levels of maintenance methylation (Liu et al., 2013). ^mUHRF1 is an essential cofactor that supports DNMT1 in maintenance methylation during DNA replication, which is a crucial event for epigenome inheritance (Kurimoto et al., 2008). ^mDNMT1 is recruited to the replication fork by the proliferative cell nuclear antigen (^mPCNA) and to hemi-methylated sites by ^mUHRF1 (Bostick et al., 2007; Sharif et al., 2007). Global demethylation in naïve embryonic stem cells appears to be driven by reduced protein levels of ^mUHRF1 and ^mDNMT1 (Iurlaro et al., 2017). Upon transition to the 2i stage, there is a global loss of ^mUHRF1 in concurrence with the loss of ^mH3K9me2 methylation required for chromatin binding of ^mUHRF1 as mentioned before (van den Berg et al., 2010). In murine imprinting control regions (ICR), H3K9 methylation functions as an anchor of local maintenance DNA methylation against global DNA demethylation and is assisted by several DNA binding proteins such as KAP1 and ZFP57 (Quenneville et al., 2011). Moreover, UHRF1 has an E3 ubiquitin-protein ligase activity (Nishiyama et al., 2013).

Recently, two studies explored the molecular details of the interaction between UHRF1, methylated CpGs and histone methylation. Fang et al. proposed that the binding of ^hUHRF1 to hemi-methylated DNA induces conformational changes that enhance its ability to recognize ^hH3K9me3. In addition, this resulted in a downstream interaction of the SRA domain with the RFTS sequence on ^hDNMT1 (Fang et al., 2016). In parallel, Harrison et al. showed that binding of ^hUHRF1 to hemi-methylated DNA and ^hH3K9me2/me3 via reciprocal cooperativity of DNA and histone binding domains, activates ubiquitylation of multiple lysines on the ^hH3 tail in the vicinity of the ^hUHRF1 histone-binding site (Harrison et al., 2016). This binding process was found to be required for DNA methylation but nonetheless is dispensable for chromatin interactions (Harrison et al., 2016).

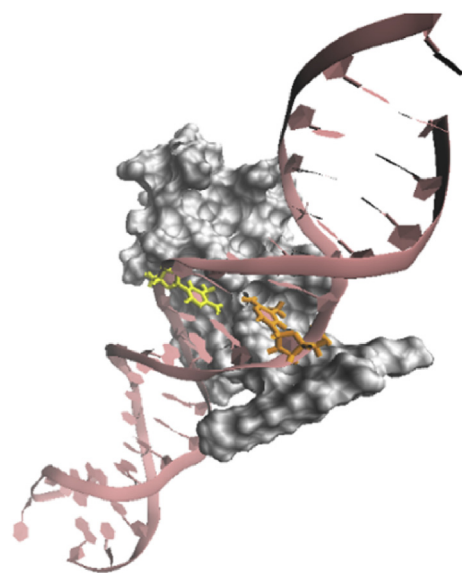


Fig. 4. MeCP2 bound to the methylated *BDNF* promoter sequence. Shown is a snapshot after 100 ns of molecular dynamics simulation. MeCP2 is shown in grey surface representation. The two methylated cytosine bases are shown as stick models.

3.8. Structural insights into C2H2 zinc finger proteins

The transcription factor Kaiso induces transcriptional repression of its target genes. In particular, Kaiso from *Xenopus laevis* binds specifically to the promoters of several genes of the Wnt signaling pathway (Kim et al., 2004; Park et al., 2005) that contributes to the maintenance of pluripotency in mouse and human ES cells, and promotes reprogramming of somatic cells to pluripotency (Marson et al., 2008). Interestingly, the Kaiso zinc finger DNA-binding domain uses similar mechanisms to bind either to the non-methylated, sequence-specific DNA target KBS, or to the promoter region of E-cadherin, which is symmetrically methylated. Superposition of the lowest energy structures of the NMR ensemble of both DNA sequences bound to Kaiso showed a high degree of structural alignment (Buck-Koehntop et al., 2012).

The bound Kaiso protein is composed of three domains, namely ZF1, ZF2, and ZF3. ZF3 is highly disordered in the unbound form and becomes ordered upon binding to DNA, when the third β -strand of the domain wraps around the backbone phosphate of DNA forming a unique $\beta\beta\alpha\beta$ motif. Additionally, the bound protein-DNA interface includes hydrophobic packing and hydrogen-bonding interactions in the ZF2 domain. Base-specific readouts are enabled by classical and methyl $\text{CH}\cdots\text{O}$ hydrogen-bonding interactions of DNA with the side chains of the N-terminal helices of the ZF1 and ZF2 domains. Gln563 in the ZF3 domain penetrates into the major groove of DNA to form hydrogen-bonds with G32 and C7 in the non-methylated form of DNA. In contrast, the methylated form of DNA has a slightly different geometry that makes it impossible for Gln563 to induce hydrogen-bonds with the aforementioned bases in the major groove. Instead, Gln563 forms interactions with the phosphate backbone of C5. The zinc fingers are anchored to methylated as well as unmethylated DNA by a set of direct hydrogen-bonds, as well as water-mediated contacts to the phosphate backbone. Additionally, van der Waals interactions with the sugar rings are established (Buck-Koehntop et al., 2012).

^mZFP57 is another C2H2 zinc finger protein that binds to specific stretches of DNA sequences. Together with its cofactor ^mKAP1, ^mZFP57 establishes and reinforces the activity of H3K9me, which functions as an anchor for local maintenance of DNA methylation (in imprinting control regions) in the face of simultaneous global demethylation. As a result, mH3K9 methylases, such as mSETDB1 are recruited. ZFP57, its cofactor KAP1, and other effectors bind in a selective manner to imprinting control regions (ICRs) in ES cells that are methylated and modified by the aforementioned enzymes. ZFP57 is also involved in imprint establishment and the maintenance of paternal and maternal imprinted loci (Li et al., 2008). Deletions in mZFP57 or mDNMTs lead to ICR DNA demethylation (Quenneville et al., 2011).

Liu and colleagues determined the crystal structure of a mouse Zfp57 fragment containing the two zinc fingers ZF2 and ZF3 bound to a 10bp oligonucleotide stretch. The two zinc fingers contain two β -strands and one helix that coordinate a zinc ion tetrahedrally via two cysteines and two histidines from the β -strands and the α -helix, respectively. Both zinc fingers bind in the major groove of DNA and have ‘direct readouts’. Therein, six base-pairs can form either direct or water-mediated hydrogen-bonds with the target amino acids. The carbonyl O4 atom of thymine 4 (T4) forms hydrogen-bonds with the hydroxyl group of Ser153, while the exocyclic amine group of adenine (A4) forms water-mediated bonds with Asp151. The three guanines in the GC stretch each form hydrogen-bonds with one arginine (namely Arg157, Arg178, Arg185). These bonds can either be direct contacts between protein and DNA residues or mediated via water molecules. The two methylated cytosines bind to the protein via completely different mechanisms: whereas 5mC at position 7 forms water contacts and is surrounded by an ordered layer of water molecules, 5mC at position 8 forms van der Waals contacts with the guanidium group of arginine 178. Additionally, one of the carboxylate oxygen atoms of glutamate 182 interacts with the N4 atom of cytosine 8. The residues involved in base recognition are the same in mouse and human (Liu et al., 2012).

^mKLF4 (one of the Yamanaka factors) stands for Krüppel-like factor 4 protein and functions to maintain the naïve pluripotency state (Guo et al., 2009). It is involved in regulating the expression of core pluripotency genes by binding to their upstream enhancer regions. By recruiting the protein cohesin, ^mKLF4 is crucial for maintaining the enhancer-promoter contacts of its target genes (Wei et al., 2013). A decline in *Klf4* levels induces an unloading of cohesin and a consequent transcriptional repression of *Oct4*, which causes differentiation (Wei et al., 2013). Being a target of LIF/STAT3 signaling cascades, ^mKLF4 induces self-renewal of the pluripotency state (Hall et al., 2009). Upon overexpression, *Klf4* promotes self-renewal of pluripotent cells even without LIF signaling (Niwa et al., 2009).

Mechanistically, ^mKLF4 recognizes the CpG dinucleotide in a G/C rich stretch of nucleotides (Liu et al., 2014) both in the methylated and the non-methylated forms. The consensus binding element for both forms shares a central GG(C/T)G, where the main constituent is either a ‘usually’ methylated CpG, or a TpG that is methylated on the complementary strand. ^mKLF4 contains three tandem C2H2 zinc fingers that bind in the major groove of DNA. The N-terminal ZF1 interacts with the 3’ side of the DNA sequence, whereas ZF2 and ZF3 bind in the central region and on the 5’ side, respectively. The interactions formed by the two methylated cytosines are very similar to the case of ZFP57-DNA interaction. One 5-methylcytosine displays water mediated contacts with the protein, while the guanidium group of Arg443 bridges the contact between the other 5-methylcytosine (via van der Waals interactions) and the adjacent guanine G6 (by forming hydrogen-bonds). Arg449 seems not to be involved in the interactions between protein and DNA in its non-methylated form. The 5 mC methyl group forms a weak (3.6 Å long) C–H \cdots O hydrogen-bond with the carboxylate group of Glu446 (Gu et al., 1999). KLF4 exhibits similar affinities to DNA in the fully methylated, hemi-methylated, and non-methylated forms, with a slight preference for the fully methylated form (Liu et al., 2014).

4. The impact of cytosine methylation on the interaction of core pluripotency factors with DNA in the pluripotent state

Global DNA hypomethylation is a hallmark of the naïve epigenome in pluripotent cells, whereby low-methylated regions are active distal regulatory regions. Numerous studies explored the global hypomethylation as a major scheme in the induction of the ‘ground state’ induced by 2i together with the accompanying transcriptional profile (Hackett et al., 2013; Leitch et al., 2013; Sim et al., 2017). So far, the causal effectors for this global demethylation are little understood. Notably, when changing the ESC medium from serum to ground-state conditions (2i), the occupancies of core transcription factors to the genome reorganize despite marginal changes in genome-wide DNA methylation in mouse (Galonska et al., 2015). Global DNA demethylation is observed in conventional murine ESCs 1–3 days after inducing the transition to the ground state (Ficz et al., 2013). This, together with evidence on chromatin structure and transcription factor activity supports the hypothesis that binding of the transcription factors of pluripotency is generally not affected by the methylation status of their binding sites. There is indeed evidence from mouse experiments that the binding landscape of the core pluripotency factors reshapes local demethylation; i.e. this binding is a cause for and not a consequence of low methylation levels (Stadler et al., 2011). The forward transition from naïve to primed pluripotency in murine ESCs involves global genomic reorganization of the binding states of core pluripotency and a subsequent remodeling of the enhancer landscape (Buecker et al., 2014; Factor et al., 2014). However, these dynamic changes seem mostly not to result directly from alterations in the expression of core transcription factors, but to be rather caused by a change in the global binding landscape and the redirection of core transcription factors by their binding partners (for example ^mOTX2) (Galonska et al., 2015).

Chromatin immunoprecipitation (ChIP) time course experiments showed that ^mOCT4 initiates the demethylation of H3K9me2 and

depletion of H3 via recruiting the histone lysine demethylase ^mJMJD1C. Consequently, the DNA is prone to DNMT3a-mediated methylation (Shakya et al., 2015). ChIP-seq analysis confirmed the binding of ^mOCT4, ^mJMJD1C, and ^mFACT (facilitates chromatin transcription) to the enhancer region of *Oct4*, the *Nanog* promoter, as well as other gene targets of ^mOCT4. These proteins are functional cofactors that are crucial for reprogramming fibroblasts to pluripotency (Shakya et al., 2015).

^mOCT4 possesses a linker region (AA76-AA92) that differs from the linker sequence in other members of the Oct family but is nonetheless conserved across species (Esch et al., 2013). Linker residues exposed to the ^mOCT4 surface were suggested to contribute to the biological activity of OCT4. Mutations in asparagine residues 76–79 led to the formation of significantly fewer iPS cell colonies. Additionally, Leu80 achieved the strongest effect in abolishing any colony formation by the pluripotent stem cells, as the capacity to reprogramming was fully abolished. In murine embryonic stem cells, no difference between the wild-type and the mutants was detected with respect to the binding process of the OCT4-SOX2 heterodimer. This indeed signifies the presence of a different interaction partner for OCT4 whose interaction interface is misshaped (Esch et al., 2013). Other studies then analyzed the full protein interactome involving ^mOCT4 (Ding et al., 2012; Pardo et al., 2010; van den Berg et al., 2010). ^mSMACRCA4, a member of the SWI/SNF chromatin remodeling complex, was found to be one of the two proteins in the interactome network that exhibited dramatically reduced levels in the mutant interactome network of the L80A ^mOCT4 protein (Esch et al., 2013). Analysis of the pluripotency proteome showed that *Smarca4/Brg1* is significantly overexpressed in 2i cells compared to serum (Taleahmad et al., 2015). The other protein which had lower levels in the mutant interactome of the L80A OCT4 protein is Chd4, a helicase belonging to the NuRD complex. As previously mentioned, ^{h/m}MBD2 interacts with the NuRD complex and directs it to methylated DNA (Zhang et al., 1999).

5. DNA methylation levels in regulatory elements of core pluripotency factors

The upstream region of the mouse *Oct4* gene contains three regulatory elements, namely distal enhancer (DE), proximal enhancer (PE) and proximal promoter (PP). The two enhancers are differentially active depending on the developmental stage of the mouse embryo (Kellner and Kikyo, 2010). DE activates *Oct4* expression in ICM, ES cells and primordial germ cells, whereas PE drives *Oct4* expression in epiblast cells. DNA methylation of these three elements mirrors the expression of the *Oct4* gene. In ES cells, they are unmethylated. In somatic cells which do not express *Oct4*, these elements are methylated. DNA methylation at the Sox2 super enhancer was found to have distinct effects on the cellular differentiation state (Song et al., 2019). Elevated methylation levels in a subset of naïve pluripotency gene promoters were found during the transition from pluripotency in mouse ESCs, including the *Nanog* promoter (Kalkan et al., 2017). However, the promoters of other naïve and general pluripotency factors did not gain methylation, showing that pluripotency-associated genes acquire methylation with different kinetics. Since expression was only poorly correlated to methylation levels, the authors suggested that promoter methylation may not be a major driver of transcriptional changes during exit from the naïve state, although they acknowledged that methylation might contribute to repressing certain genes that are relevant for the transition (Kalkan et al., 2017).

During differentiation of ESCs as well as in the mouse post-implantation embryo, DNMT3a and DNMT3b were found to mutually stimulate each other and interact synergistically to methylate the promoters of the *Oct4* and *Nanog* genes (Li et al., 2007). Double knockdown cells of *Dnmt3a* and *Dnmt3b* are associated with the downstream reactivation and transcription of *Oct4* and *Nanog* (Li et al., 2007). Nonetheless, the two methyltransferases seem to be dispensable for reprogramming to the pluripotent state in mouse (Pawlak and Jaenisch, 2011). Conversely, ChIP analysis was conducted by Wu and colleagues.

The results showed that in mouse NIH/3 T3 cells and CCE cells, ^mOCT4 binds directly to the –554 to –294 fragment of the upstream regulatory sequence of *Dnmt1*. Transfecting ^mOCT4 siRNA into mouse CCE cells resulted in the downregulation of *Dnmt1* expression (Wu et al., 2018).

^mTET1-^mTET3 were found to be involved in active demethylation of the regulatory regions of the core pluripotency genes. ^mTET2 mediates the conversion to 5-hydroxymethylcytosine (5hmC) in the promoters of *Nanog*, some of its target genes, as well as of the estrogen-related receptor-β (ESRRB). *Esrrb* is ubiquitously expressed in the naïve state of pluripotency and in the reprogramming system of OCT4-SOX2-KLF4-cMYC. It was noted in OSKM-transduced mouse embryonic fibroblast cells that hydroxymethylation can be a distinct epigenetic mark from demethylation (Doege et al., 2012). The ^mTET1 protein is highly expressed in the ground state of pluripotency, but it is downregulated during differentiation (Tahiliani et al., 2009). ^mTET1 was shown to interact with the ^mSIN3A co-repressor complex (Williams et al., 2011). ^{h/m}OCT4 is considered as the hub of the core pluripotency network that regulates its own expression and the expression of ^{h/m}NANOG, ^{h/m}SOX2, as well as of ^{h/m}TET1 and ^{h/m}TET2 (Boyer et al., 2005; Koh et al., 2011). It was shown that Tet1 is capable of replacing Oct4 during somatic cell reprogramming, together with Sox2, Klf4, and c-Myc in a secondary reprogramming circuitry. Additionally, ^mTET1 was demonstrated to physically and functionally interact with the ^mNANOG protein (Costa et al., 2013) and to co-occupy many genomic loci, e.g., *Esrrb*. Neither ^mTET1 nor ^mTET2 was sufficient for the induction of pluripotency, but either enzyme can partner with ^mNANOG to enhance reprogramming of somatic cells to naïve pluripotency (Costa et al., 2013). ^mTET1 and ^mNANOG co-occupy the promoter of *Oct4* and induce a rise in its 5hmC levels prior to reprogramming (Costa et al., 2013). Thus, demethylation and hydroxy-methylation both contribute to the process of reactivating regulatory genes that are vital for pluripotency in mouse (Gao et al., 2013).

Olariu and colleagues proposed a stochastic model for methylation of the promoter of *Oct4* (Olariu et al., 2016). This model suggests a positive feedback loop between ^mOCT4, ^mTET1, and ^mNANOG. According to this computational model, not only is *Oct4* a main target for regulation by NANOG-TET1, but TET1 also regulates its own expression (Olariu et al., 2016). Such regulatory feedback control loops were studied intensively in the literature. Commonly, *Oct4*, *Nanog* and *Tet1* are the cornerstones in these circuits, supporting the direct involvement of DNA methylation in the regulation of the regulatory regions for core pluripotency factors. This circuit is additionally affected by the changing environment, what affects cell fate (Papatsenko et al., 2018). Recent studies added several protein layers to this network topology, with positive and negative interactions (Dunn et al., 2014; Xu et al., 2014). Ravichandran and colleagues found that the retinoid inducible nuclear factor in mouse, ^mRinf, forms complexes with ^mTet1-^mNanog-^mOCT4 and facilitates the proper recruitment of these factors to the regulatory regions of their own genes (Ravichandran et al., 2019).

6. General roles of cytosine methylation in cell differentiation

Excellent reviews have documented the changes of global DNA methylation patterns upon differentiation of ESCs (Kim and Costello, 2017). Here, in a continuation of the previous section, we will briefly review the existing knowledge whether and how differentiation affects the methylation levels at the enhancers and promoters of key pluripotency genes and what the relevant writer and eraser proteins for this are.

In preparation for differentiation, DNA hypermethylation is induced at the promoters of many pluripotency and germline-specific genes, e.g., *Oct4* and *Nanog*. During differentiation, DNA methylation co-occurs with nucleosome assembly so that the binding of transcription factors is inhibited. It was shown that both ^mDNMT3a and ^m3b are required for *de novo* methylation of the promoters of *Oct4* and *Nanog* (Li et al., 2007). On the other hand, nucleosome-depleted regions (NDRs) are enriched in the proximal promoters or the distal enhancers of pluripotency factors, e.g.,

^hNANOG and ^hOCT4, respectively. Such NDRs are normally unmethylated and hence prone to binding by core pluripotency transcription factors, which again promotes the expression of ^hOCT4 and ^hNANOG. Differentiation leads to increased nucleosome occupancy, which precedes *de novo* DNA methylation. For example, forced expression of ^hOCT4 during differentiation was found to restore the NDRs but only in cells with unmethylated enhancers (You et al., 2011).

The epigenomic machinery responsible for differentiation is an orchestrated cellular “program” involving different players. First, a repressor binds to the proximal enhancer regions of the target genes, then the protein ^mG9a, also known as euchromatic histone-lysine N-methyltransferase 2 (^mEHMT2), mediates ^mH3K9 methylation, followed by recruitment of heterochromatin protein 1 (^mHP1) and then *de novo* DNA methylation (Feldman et al., 2006). Artificial targeting of HP1 to the promoter region of *OCT4* induces *OCT4* silencing. In cells other than pluripotent cells, this silencing remains heritable after the removal of HP1. Nonetheless, in mouse embryonic fibroblasts derived from ESCs, this removal leads to concurrent *OCT4* demethylation and expression (Hathaway et al., 2012). ^mDNMT3a and ^mDNMT3b are recruited and can initiate methylation at the proximal enhancer (Athanasiadou et al., 2010). The histone H3 Lys9 methylases G9a and GLP were shown to be essential to maintenance of DNA methylation at the imprinted loci in murine ESCs as they recruit the *de novo* DNMTs (Zhang et al., 2016b). Other studies showed that G9a may not be required for DNA methylation at the *Oct4* enhancer (Athanasiadou et al., 2010).

Establishing DNA methylation at specific gene promoters requires cooperative action of ^mLSH, a member of the SNF2 ATPase family (encoded by the *Hells* gene), and the ^mG9a/GLP complex of histone methylases (Myant et al., 2011). Differentiating ESCs lacking ^mLSH showed discrepancies in methylation between neighboring CpG regions. As a result, many genes were deregulated in fibroblasts lacking the LSH protein (Athanasiadou et al., 2010).

^mLSD1 (also termed Kdm1A) is a histone demethylase that removes methylation at ^mH3K4me and ^mH3K9me. The ^mLsd1-Mi2/NuRD complex was found to be involved in a regulatory switch that stimulates the interaction of H3K4-unmethylated histone tails with the ^mDNMT3 ATRX-DNMT3-DNMT3L (ADD) domain and subsequent DNA methylation at the enhancers of some pluripotency genes (Petell et al., 2016). The interaction of the ADD domain of ^{h/m}DNMT3a with H3K4 is blocked by H3K4 methylation (Guo et al., 2015). Petell and colleagues suggested that for ^mDNMT3a to be active, ^mLSD1-Mi2/NuRD-mediated histone deacetylation and demethylation should precede ^mDNMT3a-induced methylation of the enhancer region of pluripotency genes (Petell et al., 2016). This histone-mediated effect is triggered by the dissociation of the OSN (OCT4-SOX2-NANOG) co-activator complex. Suppression of some genes in DNMT3a-KO cells was only partly affected when compared to wild-type cells. As such, DNMT3a-mediated DNA methylation is one of several processes that together cause the inactivation of pluripotency genes. On the other hand, inhibition of ^mLSD1 was not correlated with the level of ^mDNMT1 protein during induction of differentiation of ESCs, nor was ^mLSD1 associated with *global* DNA methylation. In addition, ^mDNMT3-KO cells almost lacked any methylation in the differentiated states. This indeed suggests that the presence of ^mDNMT1 cannot compensate for the absence of ^mDNMT3a in initiating methylation at the enhancers of pluripotency genes (Petell et al., 2016).

In the process of hematopoietic maturation, ^mLSD1 ensures that the regulatory regions of hematopoietic stem and progenitor cell (HSPC) genes are inhibited (Kerenyi et al., 2013). After loss of ^mDNMT3a, multipotency genes are upregulated, whereas genes responsible for differentiation are downregulated (Challen et al., 2012). This indeed puts forward a possible similar mechanism for methylation followed during hematopoiesis as in differentiation.

ESCs lacking DNMT1 are not viable. When differentiation is induced, they enter into apoptosis. On the other hand, ESCs deficient for DNMT3a/DNMT3b partially or completely lack their ability to differentiate. These cells remain viable in the pluripotent state, and they have

high histone acetylation levels. If DNA methylation is reestablished in these cells, they regain their ability to differentiate (Jackson et al., 2004). Lastly, recruitment of DNMTs to murine imprinting control regions was found to antagonize TET-dependent DNA demethylation (Zhang et al., 2016b).

7. Effects of epigenetic drugs

Epigenetic drugs, or ‘epidrugs’, are a new class of drug molecules that induce or inhibit histone-modifying enzymes, DNA methyltransferases, or the readers of the resulting chromatin modifications. As the activity of the drug targets directly affects chromatin state, epidrugs crucially affect a wide spectrum of cellular processes. Naturally, the efficacy of epidrugs is also dependent on the 3D structure of chromatin, which in turn is context-dependent as well. Hence, one can clearly expect that epidrugs will have different effects in different cell types. It has been argued that epidrugs have the potential to target a wide variety of ailments, including neurodegenerative diseases (Delgado-Morales et al., 2017), tumors, regenerative medicine and metabolic disorders (Altucci and Rots, 2016).

One obvious starting point where epigenetic drugs may have a potential use as therapeutic agents is the observation that anomalies in DNA methylation often occur during tumorigenesis. In particular, tumor genomes show global hypomethylation coupled with hypermethylation of specific loci. Many of the sequence segments showing hypomethylation in tumors are originally transposable and parasitic elements, which are hypermethylated under normal conditions (Bestor and Tycko, 1996; Lee et al., 2012; Xie et al., 2013). Hypermethylation, on the other hand, was shown to impair the beneficial activity of tumor-suppressor genes and induce inactivation of DNA-repair genes, especially in the context of CpG islands (Baylin and Herman, 2000). In the light of the observed global hypomethylation, it may come as a surprise that the expression levels of ^hDNMT1 were shown to be higher in tumor than in normal tissues, e.g. in colon cancer (El-Deiry et al., 1991) as well as in acute and chronic myelogenous leukemia (Mizuno et al., 2001) (here also ^hDNMT3a and ^hDNMT3b), which can be associated with the hypermethylation of CpG islands (Vertino et al., 1996). The role of DNMT1 in tumorigenesis is not uniform. On the one hand, DNMT1 was shown to possess tumor suppressor activity in some tumor forms; e.g., in early-stage murine prostate cancer (Kinney et al., 2010). On the other hand, it had an oncogenic activity in later stages of the same cancer, thus playing opposing roles (Kinney et al., 2010). Other mechanisms were suggested to attribute tumors to aberrant expression of ^{h/m}DNMTs during different stages of the cell cycle (Jones, 1996; Zhang and Xu, 2017).

One of the proposed mechanisms involves aberrant protein-protein interactions including ^{h/m}DNMTs. DNMT1 interacts, for example, with proteins that affect its nuclear localization (e.g., ^{h/m}PCNA) or chromatin targeting (e.g., ^{h/m}HDAC classes I and II) (Robertson, 2001). Two DNMT inhibitors, azacytidine and decitabine, have been approved by the FDA as treatments for acute myeloid leukemia and chronic myelomonocytic leukemia, respectively (Mazzone et al., 2017). Azacytidine and decitabine are structurally related cytidine nucleoside analogs that become intracellularly activated by triphosphorylation. Decitabine-TP is exclusively incorporated into DNA, while azacytidine-TP is incorporated primarily into RNA (Oellerich et al., 2019). It was recently shown that decitabine-TP, but not azacytidine-TP, is an activator and substrate of the triphosphohydrolase SAMHD1 (Oellerich et al., 2019). Fig. 5 shows the active site region in the X-ray structure of decitabine-TP bound to SAMHD1. Ongoing clinical trials of next-generation hypomethylating agents (HMAs), such as guadecitabine and cedazuridine are described in (Pan et al., 2020).

DNMTi drugs also modulate several layers of the immune system. In this respect, substances inducing DNA hypomethylation can be utilized as a feasible approach in tumor therapy. Cancer testis antigens (CTA) are a family of tumor-associated antigens (TAA), which elicit immune response (by cytotoxic T cells) against tumors. These proteins are normally not expressed in human somatic tissues, except for testis and

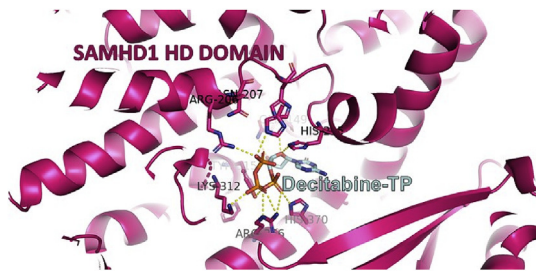


Fig. 5. X-ray structure of the epigenetic inhibitor decitabine-TP bound to SAMHD1 (PDB code: 6CM2) (Oellerich et al., 2019).

placenta, as well as in several tumors. In normal tissues, CTA genes have methylated promoters so that they are transcriptionally inactive (Fratta et al., 2011), and they are activated by demethylation (De Smet et al., 1999). Antigen presentation is another immunological mechanism that the body uses in its defense mechanism. In mouse or human tumors, DNMTi were found to induce this machinery and propagate the downstream lysis of neoplastic cells by cytotoxic T cells (Khan et al., 2008; Manning et al., 2008; Nie et al., 2001; Sigalotti et al., 2004).

8. Conclusions

It is now well accepted that there exists a compact group of core pluripotency factors that are responsible for maintenance of pluripotency and/or induction of differentiation via regulating the expression of hundreds to thousands of target genes. Furthermore, the importance of epigenetic alterations (DNA methylation and histone marks) in these processes have also been well documented. As presented here, there exists an intricate interplay between the core pluripotency factors and the enzymes of the DNMT family. For example, as discussed in section 4, the transcriptional activities of ^mDnmt1 were regulated by direct binding of Oct4 to the upstream regulation element of Dnmt1 (Wu et al., 2018). On the other hand, DNMT3a and DNMT3b were found to mutually stimulate each other and interact synergistically to methylate the promoters of the *Oct4* and *Nanog* genes (Li et al., 2007). Similar ideas are contained in the computational model of Olariu et al. (2016), who formulated a regulatory network of Oct4, Nanog and Tet1 including positive feedback loops involving DNA-demethylation around the promoters of Oct4 and Tet1. We suggest that following up on these three studies would be a worthwhile goal that may reveal further surprising insights into the interplay of transcriptomic and epigenomic regulation of differentiation processes. A challenge hereby is to place or remove epigenetic marks at specified genomic positions under *in vivo* conditions and studying their phenotypic effects. We hope that this review will contribute to stimulate further work along these lines.

Acknowledgements

We thank Thorsten Will for critical comments on the text, Afnan Sultan for background research of the literature, Michael Hutter for comments on the text, and Raghad Albzoor for help with producing graphics. This work was financially supported by a Zamalah fellowship to SS and by Deutsche Forschungsgemeinschaft, Germany through SFB 1027 (VH and SS).

References

Acampora, D., Omodei, D., Petrosino, G., Garofalo, A., Savarese, M., Nigro, V., Di Giovannantonio, L.G., Mercadante, V., Simeone, A., 2016. Loss of the Oct2-binding site in the Nanog promoter affects the integrity of embryonic stem cell subtypes and specification of inner cell mass-derived epiblast. *Cell Rep.* 15, 2651–2664. <https://doi.org/10.1016/j.celrep.2016.05.041>.

Adamo, A., Sese, B., Boue, S., Castano, J., Paramonov, I., Barrero, M.J., Belmonte, J.C.I., 2011. LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat. Cell Biol.* 13, 652–659. <https://doi.org/10.1038/ncb2246>.

Altucci, L., Rots, M.G., 2016a. Epigenetic drugs: from chemistry via biology to medicine and back. *Clin. Epigenet.* 8, 56. <https://doi.org/10.1186/s13148-016-0222-5>.

Alvarado, A.S., Yamanaka, S., 2014. Rethinking differentiation: stem cells, regeneration, and plasticity. *Cell* 157, 110–119. <https://doi.org/10.1016/j.cell.2014.02.041>.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., Shirakawa, M., 2008. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455, 818–821. <https://doi.org/10.1038/nature07249>.

Athanasidou, R., de Sousa, D., Myant, K., Merusi, C., Stancheva, I., Bird, A., 2010. Targeting of de novo DNA methylation throughout the Oct-4 gene regulatory region in differentiating embryonic stem cells. *PLoS One* 5, e9937. <https://doi.org/10.1371/journal.pone.0009937>.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., Lovell-Badge, R., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Gene Dev.* 17, 126–140. <https://doi.org/10.1101/gad.224503>.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y.J., Duan, S.L., Bronner, C., Arrowsmith, C.H., Dhe-Paganon, S., 2008. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455, 822–825. <https://doi.org/10.1038/nature07273>.

Bar, S., Benvenisty, N., 2019. Epigenetic aberrations in human pluripotent stem cells. *EMBO J.* 38, e101033. <https://doi.org/10.15252/emj.2018101033>.

Baylin, S.B., Herman, J.G., 2000. DNA hypermethylation in tumorigenesis - epigenetics joins genetics. *Trends Genet.* 16, 168–174. [https://doi.org/10.1016/s0168-9525\(99\)01971-x](https://doi.org/10.1016/s0168-9525(99)01971-x).

Bestor, T.H., Tycko, B., 1996. Creation of genomic methylation patterns. *Nat. Genet.* 12, 363–367. <https://doi.org/10.1038/ng0496-363>.

Bird, A., 2007. Perceptions of epigenetics. *Nature* 447, 396–398. <https://doi.org/10.1038/nature05913>.

Bogdanovic, O., Lister, R., 2017. DNA methylation and the preservation of cell identity. *Curr. Opin. Genet. Dev.* 46, 9–14. <https://doi.org/10.1016/j.gde.2017.06.007>.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., Jacobsen, S.E., 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764. <https://doi.org/10.1126/science.1147939>.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.R., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., Young, R.A., 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956. <https://doi.org/10.1016/j.cell.2005.08.020>.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B.W., Lopes, S., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., Vallier, L., 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448, 191–195. <https://doi.org/10.1038/nature05950>.

Buck-Koehntop, B.A., Stanfield, R.L., Ekiert, D.C., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., Wright, P.E., 2012. Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proc. Natl. Acad. Sci. U. S. A.* 109, 15229–15234. <https://doi.org/10.1073/pnas.1213726109>.

Buecker, C., Srinivasan, R., Wu, Z.X., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M.J., Swigut, T., Wysocka, J., 2014. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14, 838–853. <https://doi.org/10.1016/j.stem.2014.04.003>.

Callebaut, I., Courvalin, J.C., Mornon, J.P., 1999. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett.* 446, 189–193. [https://doi.org/10.1016/s0014-5793\(99\)00132-5](https://doi.org/10.1016/s0014-5793(99)00132-5).

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., Dalton, S., 2005. LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* 132, 885–896. <https://doi.org/10.1242/dev.01670>.

Challen, G.A., Sun, D.Q., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H.C., Xi, Y.X., Liang, S.D., Lu, Y., Darlington, G.J., Meissner, A., Issa, J.P.J., Godley, L.A., Li, W., Goodell, M.A., 2012. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* 44, 23–31. <https://doi.org/10.1038/ng.1009>.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., Smith, A., 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113, 643–655. [https://doi.org/10.1016/s0092-8674\(03\)00392-1](https://doi.org/10.1016/s0092-8674(03)00392-1).

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., Smith, A., 2007. Nanog safeguards pluripotency and mediates germline development. *Nature* 450, 1230–1234. <https://doi.org/10.1038/nature06403>.

Chang, Y.M.K., Srivastava, Y., Hu, C.Z., Joyce, A., Yang, X.X., Zuo, Z., Havranek, J.J., Stormo, G.D., Jauch, R., 2017. Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. *Nucleic Acids Res.* 45, 832–845. <https://doi.org/10.1093/nar/gkw1198>.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W.W., Jiang, J.M., Loh, Y.H., Yeo, H.C., Yeo, Z.X., Narang, V., Govindarajan, K.R., Leong, B., Shahab, A., Ruan, Y.J., Bourque, G., Sung, W.K., Clarke, N.D., Wei, C.L., Ng, H.H., 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117. <https://doi.org/10.1016/j.cell.2008.04.043>.

Chen, J.J., Zhang, Z.J., Li, L., Chen, B.C., Revyakin, A., Hajji, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., Tjian, R., Liu, Z., 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* 156, 1274–1285. <https://doi.org/10.1016/j.cell.2014.01.062>.

Cherry, S.R., Biniszkiewicz, D., van Parijs, L., Baltimore, D., Jaenisch, R., 2000. Retroviral expression in embryonic stem cells and hematopoietic stem cells. *Mol. Cell Biol.* 20, 7419–7426. <https://doi.org/10.1128/mcb.20.20.7419-7426.2000>.

- maintenance if the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* 62, 848–861. <https://doi.org/10.1016/j.molcel.2016.04.025>.
- Wei, Z., Gao, F., Kim, S., Yang, H.Z., Lyu, J., An, W.J., Wang, K., Lu, W.G., 2013. Klf4 organizes long-range chromosomal interactions with the Oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* 13, 36–47. <https://doi.org/10.1016/j.stem.2013.05.010>.
- Weinberger, L., Ayyash, M., Novershtern, N., Hanna, J.H., 2016. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* 17, 155–169. <https://doi.org/10.1038/nrm.2015.28>.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., Helin, K., 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348. <https://doi.org/10.1038/nature10066>.
- Wongtrakoon, P., Li, J., Andrews, P.W., 2014. DNMT3B inhibits the re-expression of genes associated with induced pluripotency. *Exp. Cell Res.* 321, 231–239. <https://doi.org/10.1016/j.yexcr.2013.11.024>.
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., Walter, J., 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* 2, 241 <https://doi.org/10.1038/ncomms1240>.
- Wu, F.R., Wu, Q.Q., Li, D.K., Zhang, Y., Wang, R., Liu, Y., Li, W.Y., 2018. Oct4 regulates DNA methyltransferase 1 transcription by direct binding of the regulatory element. *Cell. Mol. Biol. Lett.* 23, 39. <https://doi.org/10.1186/s11658-018-0104-2>.
- Xie, M.C., Hong, C.B., Zhang, B., Lowdon, R.F., Xing, X.Y., Li, D.F., Zhou, X., Lee, H.J., Maire, C.L., Ligon, K.L., Gascard, P., Sigaroudinia, M., Tlsty, T.D., Kadlecik, T., Weiss, A., O'Geen, H., Farnham, P.J., Madden, P.A.F., Mungall, A.J., Tam, A., Kamoh, B., Cho, S., Moore, R., Hirst, M., Marra, M.A., Costello, J.F., Wang, T., 2013. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.* 45, 836–841. <https://doi.org/10.1038/ng.2649>.
- Xu, H.L., Ang, Y.S., Sevilla, A., Lemischka, I.R., Ma'ayan, A., 2014. Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Comput. Biol.* 10, e1003777 <https://doi.org/10.1371/journal.pcbi.1003777>.
- Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., Saitou, M., 2013. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* 12, 368–382. <https://doi.org/10.1016/j.stem.2012.12.012>.
- Yeo, J.C., Jiang, J.M., Tan, Z.Y., Yim, G.R., Ng, J.H., Goke, J., Kraus, P., Liang, H.Q., Gonzales, K.A.U., Chong, H.C., Tan, C.P., Lim, Y.S., Tan, N.S., Lufkin, T., Ng, H.H., 2014. Klf2 is an essential factor that sustains ground state pluripotency. *Cell Stem Cell* 14, 864–872. <https://doi.org/10.1016/j.stem.2014.04.015>.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., Smith, A., 2008. The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523. <https://doi.org/10.1038/nature06968>.
- You, J.S., Kelly, T.K., De Carvalho, D.D., Taberlay, P.C., Liang, G.N., Jones, P.A., 2011. OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 14497–14502. <https://doi.org/10.1073/pnas.1111309108>.
- Young, R.A., 2011. Control of the embryonic stem cell state. *Cell* 144, 940–954. <https://doi.org/10.1016/j.cell.2011.01.032>.
- Yu, J.Y., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, Thomson, J.A., 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917–1920. <https://doi.org/10.1126/science.1151526>.
- Zhang, W.H., Xia, W.K., Wang, Q.J., Towers, A.J., Chen, J.Y., Gao, R., Zhang, Y., Yen, C.A., Lee, A.Y., Li, Y.Y., Zhou, C., Liu, K.L., Zhang, J., Gu, T.P., Chen, X.Q., Chang, Z., Leung, D., Gao, S.R., Jiang, Y.H., Xie, W., 2016c. Isoform switch of TET1 regulates DNA demethylation and mouse development. *Mol. Cell* 64, 1062–1073. <https://doi.org/10.1016/j.molcel.2016.10.030>.
- Zhang, W., Xu, J., 2017. DNA methyltransferases and their roles in tumorigenesis. *Biomarker Research* 5, 1. <https://doi.org/10.1186/s40364-017-0081-z>.
- Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., Reinberg, D., 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Gene Dev.* 13, 1924–1935. <https://doi.org/10.1101/gad.13.15.1924>.
- Zhang, H., Jiao, W.W., Sun, L., Fan, J.Y., Chen, M.F., Wang, H., Xu, X.Y., Shen, A.D., Li, T., Niu, B.B., Ge, S.F., Li, W., Cui, J.W., Wang, G.J., Sun, J.N., Fan, X.Q., Hu, X., Mrsny, R.J., Hoffman, A.R., Hu, J.F., 2013. Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. *Cell Stem Cell* 13, 30–35. <https://doi.org/10.1016/j.stem.2013.05.012>.
- Zhang, J.L., Liu, G.K., Ruan, Y., Wang, J.L., Zhao, K., Wan, Y., Liu, B., Zheng, H.T., Peng, T., Wu, W., He, P., Hu, F.Q., Jian, R., 2014. Dax1 and Nanog act in parallel to stabilize mouse embryonic stem cells and induced pluripotency. *Nat. Commun.* 5, 5042 <https://doi.org/10.1038/ncomms6042>.
- Zhang, J., Ratanasirintrao, S., Chandrasekaran, S., Wu, Z.T., Ficarro, S.B., Yu, C.X., Ross, C.A., Cacchiarelli, D., Xia, Q., Seligson, M., Shinoda, G., Xie, W., Cahan, P., Wang, L.F., Ng, S.C., Tintara, S., Trapnell, C., Onder, T., Loh, Y.H., Mikkelsen, T., Sliz, P., Teitell, M.A., Asara, J.M., Marto, J.A., Li, H., Collins, J.J., Daley, G.Q., 2016a. LIN28 regulates stem cell metabolism and conversion to primed pluripotency. *Cell Stem Cell* 19, 66–80. <https://doi.org/10.1016/j.stem.2016.05.009>.
- Zhang, T., Termanis, A., Ozkan, B., Bao, X.X., Culley, J., Alves, F.D., Rappsilber, J., Ramsahoye, B., Stancheva, I., 2016b. G9a/GLP complex maintains imprinted DNA methylation in embryonic stem cells. *Cell Rep.* 15, 77–85. <https://doi.org/10.1016/j.celrep.2016.03.007>.
- Zou, X.Q., Ma, W., Solov'yov, I.A., Chipot, C., Schulten, K., 2012b. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res.* 40, 2747–2758. <https://doi.org/10.1093/nar/gkr1057>.