

Linking Active DNA Demethylation by Thymine DNA Glycosylase with Epigenetic Regulation of Gene Expression

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Annika Wirz

aus

Luzern, Schweiz

Basel, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Primo Schär (Fakultätsverantwortlicher und Dissertationsleiter)

Prof. Dr. Patrick Matthias (Korreferent)

Prof. Magnar Bjørås (Korreferent)

Basel, den 24.06.2014

Prof. Dr. Jörg Schibler
Dekan der Philosophisch-Naturwissenschaftlichen Fakultät

Acknowledgements

First of all, I want to express my gratitude to my PhD supervisor Primo Schär – for giving me the chance to learn so much; for trusting in me; for teaching, guiding and challenging me; for long discussions and his open door policy; for being a very nice in addition to a very smart person and for the much appreciated time in Norway.

I want to thank the co-examiners of my PhD thesis: Patrick Matthias for his input and time invested and Magnar Bjørås just as much for his input and time and, furthermore, for the traveling he undertook for my doctorate's sake. Additionally, I thank Antonius Rolink for being the chairman of my PhD defense.

I would like to thank all current and past members of the Schär lab.

Especially, I want to thank Christophe Kunz and Daniel Cortázar for introducing me to the lab and the bench, for motivating words and very patient discussions in the beginning of my PhD time. Additionally, I thank Christophe Kunz for critical reading of my thesis and for his extra support towards the end. Thanks also to David Schürmann for the generous sharing of his knowledge, his patience and his good sense of humour. I thank Joëlle Rüegg for her spirit and “Poison” and Faiza Noreen for her belief in me and motivating words. Thanks go also to Stefan Weis for his indestructible optimism and diverting neighborhood and to Claudia Krawczyk for her kind nature and loud laughs, her support, many hugs and many hours of running and kicking. I further want to thank Alain Weber for his friendship and fun (after 9 am) and Angelika Jacobs for professional team-working and motivating words.

I acknowledge Christian Beisel, Ina Nissen and Katja Eschbach from the D-BSSE (ETH Zürich) for the possibility to work in their lab from time to time. Further, I thank the people from the Genome Technology Access Center (GTAC) in St. Louis (Missouri, USA) for very efficient and thoughtful processing of our ChIP samples and an enjoyable cooperation. I also want to thank Robert Ivánek and Faiza Noreen for performing the bioinformatical analysis of my ChIP-seq data sets, I appreciate your effort.

I further want to thank Mirjam Zimmermann, Anja Nusser and Katrin Martin for lovely coffee and ice cream breaks and fun times outside the lab.

Last but not least, I want to thank my family, Sabine and Beat Wirz and Cédric Cattin. Thank you infinitely for everything you did and continue to do for me, for everything you are. Your love and support is most precious in my life.

Your assumptions are your windows on the world.

Scrub them off every once in a while,

or the light won't come in.

Isaac Asimov

Table of Contents

Abbreviations	i-iii
1 Summary	1
2 Introduction	5
2.1 The Genome and Its Organization	5
2.1.1 Genetic Maintenance	6
2.1.2 Epigenetic Maintenance	6
2.2 Types of DNA Damage and Appropriate DNA Repair	7
2.2.1 Base Excision Repair (BER)	11
2.2.2 DNA Glycosylases	13
2.2.3 The Thymine DNA Glycosylase (TDG)	15
2.3 Chromatin	18
2.3.1 Chromatin Remodeling and Histone Variants	20
2.3.2 Genomic Region- and Context-Specific Histone Modifications ..	23
2.3.3 Chromatin Modifying Complexes and Their Regulation	25
2.3.4 Epigenetic Modifications and Cell Plasticity	29
2.4 DNA Methylation	31
2.4.1 CpG Islands (CGIs)	33
2.4.2 Regulation of DNA Methylation During Development	34
2.4.3 The Role of DNA Methylation in Disease	36
2.5. DNA Demethylation	37
2.5.1 Mechanisms of DNA Demethylation	38
2.5.2 Ten-Eleven-Translocation (TET) Protein Family	42
2.6 Linking DNA Repair to Epigenetics	46
2.7 Transcription by RNA Polymerase II	49

2.7.1 Transcription Initiation	50
2.7.2 Transcription Elongation	52
2.7.3 Transcription Termination	53
3 Aims of the Thesis	55
4 Results	56
4.1 TDG Balances DNA Methylation and Oxidative Demethylation in Differentiating Cells (Appendix I)	56
4.2 TET1, TET2 and TDG Cooperate in a Locus-Specific Manner to Promote Chromatin Plasticity by Oxidative DNA Demethylation (Appendix II)...	59
4.3 Embryonic Lethal Phenotype Reveals a Function of TDG in Maintaining Epigenetic Stability (Appendix III)	63
4.4 Supplementary Results	67
Investigating the Effect of TDG and BRD4 on RNA Polymerase II Phosphorylation State and Productive Transcription	67
5 Concluding Discussion and Outlook	78
6 References	84

Appendix:

- I: TDG Balances DNA Methylation and Oxidative Demethylation in Differentiating Cells
- II: TET1, TET2 and TDG Cooperate Locus-Specifically to Promote Chromatin Plasticity by Oxidative DNA Demethylation
- III: Embryonic Lethal Phenotype Reveals a Function of TDG in Maintaining Epigenetic Stability

Curriculum Vitae

Abbreviations

2-HG	2-hydroxyglutarate
2-OG	2-oxoglutarate / α -ketoglutarate
3-meA	N3-methyladenine
5-caC	5-carboxylcytosine
5-fC	5-formylcytosine
5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
7-meG	N7-methylguanine
8-oxoG	7,8-dihydro-8-oxoguanine
A	adenine
AID	Activation-induced deaminase
APE1	AP endonuclease 1
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide
AP-site	apurinic/apyrimidinic site
bp	base pair
BER	Base Excision Repair
BRD4	Bromodomain containing 4
C	cytosine
cat	TDG catalytically dead mutant, TDG N151A
CBP/p300	CREB-binding protein and its homologue p300
ChIP	Chromatin immunoprecipitation
CIMP	CpG island methylator phenotype
CpG	C – G dinucleotide
CRC	colorectal cancer
CTD	C-terminal domain (of RNA Pol II)
DIP	DNA immunoprecipitation
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase

DSB	double-strand break
FXS	Fragile X Syndrome
G	guanine
GGR	Global Genome Repair
GTF	general transcription factor
H3K4me1	histone 3 monomethylated at lysine 4
H3K4me2	histone 3 dimethylated at lysine 4
H3K9me3	histone 3 trimethylated at lysine 9
H3K9ac	histone 3 acetylated at lysine 9
H3K27me3	histone 3 trimethylated at lysine 27
HAT	histone acetyl transferase
HD	Huntington Disease
HDAC	histone deacetylase
HMT	histone methyl transferase
HR	homologous recombination
kb	kilo base (1000 bps)
KDM	lysine demethylase
(l)ncRNA	(long)non-coding RNA
MBD4	Methyl-CpG-binding domain protein 4
mESC	mouse embryonic stem cell
MMR	Mismatch Repair
NER	Nucleotide-Excision Repair
NHEJ	nonhomologous end joining
NP	neuronal progenitor
OGG1	8-oxoguanine DNA glycosylase
(p)CGI	(promoter) CpG island
PGC	primordial germ cells
PIC	preinitiation complex
PRC1/2	Polycomb Repressive Complex 1 and 2
PTM	post-translational modification
(q)PCR	(quantitative) polymerase chain reaction

RA	retinoic acid
RAR/RXR	retinoic acid receptor / retinoid X receptor
RNA	ribonucleic acid
RNA Pol II	RNA polymerase II
ROS	reactive oxygen species
RT	reverse transcription
S2(p)	serine 2 on CTD (phosphorylated)
S5(p)	serine 5 on CTD (phosphorylated)
SAM	S-adenosyl-L-methionine
seq	next generation sequencing
SMUG1	Single-strand selective monofunctional uracil DNA glycosylase 1
SUMO	Small ubiquitin-like modifier
T	thymine
TCR	Transcription-Coupled Repair
TDG	Thymine DNA glycosylase
TET1-3	Ten-Eleven-Translocation family of proteins 1-3
TF	transcription factor
TSS	transcription start site
U	uracil
UNG	Uracil N-glycosidase
WB	Western Blot
wt	TDG wildtype

1. Summary

The correct regulation of epigenetic modifications is crucial for cell plasticity and the establishment of cell identity in vertebrates. However, the underlying molecular mechanisms are far from being understood. Recent development in this direction indicated a role for DNA repair proteins in this context, the investigation of which was the overall aim of my PhD thesis.

The Thymine DNA Glycosylase (TDG) was discovered as an enzyme excising uracil (U) and thymine (T) mismatched with G, deamination products of cytosine and 5-methylcytosine (5-mC), respectively, thereby initiating a base excision repair (BER) process that restores the canonical G•C base pairs (Neddermann and Jiricny, 1994; Wiebauer and Jiricny, 1990). Yet, TDG has also been implicated in transcriptional regulation, both in the role as a co-activator and co-repressor (Chen et al., 2003; Kim and Um, 2008; Missero et al., 2001; Tini et al., 2002). The interaction of TDG with the retinoic acid receptor (RAR) is a well studied example of a physical as well as a functional interaction, where TDG acts as a co-activator for RAR-targeted gene expression (Leger et al., 2014; Um et al., 1998). In this respect, TDG has also been postulated as a candidate enzyme that directly demethylates 5-methylcytosine (5-mC) in DNA (Jost, 1993; Jost et al., 1995; Zhu et al., 2000), in the context of regulating gene expression. The underlying evidence, however, was highly controversial and, unlike for specialized plant DNA glycosylases, a direct activity of TDG or any other vertebrate DNA glycosylase on 5-mC could never be substantiated. Only recently, it became clear that TDG is indeed critically involved in DNA demethylation. The mechanism, however, turned out to be more complex than direct excision of the methylated base and involves several enzymatic steps, including ten-eleven translocation (TET) proteins. A currently favored view of how active DNA demethylation is accomplished is that TET proteins iteratively oxidize 5-mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), and that TDG then excises 5-fC and 5-caC, thereby initiating a repair process that integrates a non-methylated C (He et al., 2011; Maiti and Drohat, 2011). Consistently, 5-caC appears to be the long sought canonical substrate for TDG (Maiti et al., 2013) and no other activity has yet been described to excise either 5-fC or 5-caC.

The exact mechanisms for the recruitment, targeting, regulation and – probably context-dependent – action of TET and TDG are still unclear. It emerges though that TDG, together with the TET proteins, is a key player in the epigenetic maintenance of regulatory genomic elements with developmental relevance. My work aimed at providing further mechanistic insight into the epigenetic function of this multifaceted DNA repair enzyme.

In a first part, I was involved in a team effort to characterize the phenotype of the first *Tdg* knockout mouse to be reported, which ultimately established the function of TDG-dependent DNA repair in epigenetic control. Unexpectedly and contrarily to all other knockouts of a DNA glycosylase, deletion of *Tdg* caused embryonic lethality. Further characterization of TDG-proficient and TDG-deficient cells revealed no notable defects in DNA repair, but instead a misregulation of developmental genes in differentiating mouse embryonic stem cells (ESCs) and mouse embryonic fibroblasts (MEFs). This was associated with a decrease in activating and an increase in repressive chromatin marks and a stochastic accumulation of DNA methylation in CGI promoters of developmental genes in TDG-deficient cells. These discoveries led to the proposal of a mechanistic model, implicating TDG-dependent DNA repair in the establishment and maintenance of an active chromatin state at gene promoters in cells undergoing cell lineage commitment (**Appendix III**).

We then wanted to understand the precise role of TDG-mediated active DNA demethylation in cell differentiation. We generated genome-wide DNA methylation profiles of TDG-proficient and TDG-deficient mESCs and *in vitro* differentiated neuronal progenitor cells (NPs) as well as terminally differentiated MEFs. Confirming earlier observations on candidate gene loci, we identified a large number of differentially methylated regions (DMRs) but only in differentiated cells. Surprisingly, the DMRs that overlapped with CGIs were almost exclusively hypomethylated in TDG-deficient NPs. As these same CGIs were also prone to acquire DNA methylation in TDG-proficient NPs, these results suggested a failure of TDG-deficient cells to establish *de novo* methylation at these sites upon differentiation. Measuring global levels of 5-mC and its oxidized derivatives 5-fC and 5-caC confirmed elevated levels of the latter in TDG-deficient mESCs, that further increased during retinoic acid (RA)-induced differentiation. We observed the same alterations at the very CGI DMRs, implicating an engagement of TDG activity in ongoing cycles of DNA

methylation and active demethylation at these sites. Our data thus showed that CGIs undergoing epigenetic transitions during cell differentiation are kept in a state of high epigenetic plasticity, whereby the catalytic function of TDG is required to maintain the balance between DNA methylation and demethylation (**Appendix I**).

In subsequent work, I wanted to characterize the genomic sites, where TET proteins and TDG engage to effect cyclic DNA methylation – demethylation. The goal was to provide insight into the genome-wide functional interaction and coordination between these proteins and relevant biological outputs. I therefore established procedures for chromatin-immunoprecipitation-coupled next generation sequencing (ChIP-seq) for TET1, TET2 and TDG and generated the respective ChIP-seq datasets from a 24 hour RA-induced mESC differentiation setup with mESCs expressing wildtype TDG, catalytically inactive TDG or no TDG. Bioinformatic analyses then revealed clear but differential correlation of TET and TDG association with gene regulatory elements, especially gene promoters and active enhancers. Notably, sites of 5-fC enrichment in TDG-deficient cells, CGIs as well as bivalent chromatin domains showed a preferential co-occupancy with all three proteins, indicating that highly dynamic active DNA demethylation may involve the TET1-TET2-TDG trio as a whole. Contrarily, sites where uniquely TDG shows chromatin association, were more often located in active enhancers. Moreover, a vast majority of TET1 and TET2 chromatin association events were independent on TDG protein or activity, consistent with TDG acting downstream of the TET proteins and the recruitment of these demethylation factors occurring in a sequential order. Remarkably, there was also a strong correlation particularly of TET1-TET2-TDG co-occupancy with the occurrence of the dynamic histone variants H3.3 and H2A.Z. This led us to hypothesize that the targeted oxidation and excision of 5-mC by the TET-TDG system at gene regulatory elements may in fact serve the purpose to generate single-stranded DNA breaks to trigger nucleosomal dynamics and, thus, the epigenetic plasticity required at certain enhancers or promoters in differentiating cells. The predominant occurrence of TET1-TET2-TDG at bivalent CGIs suggests that this process is specifically targeted to developmental genes, presumably to facilitate their transcription. These results further indicate a functional separation between distinct TDG protein fractions (**Appendix II**).

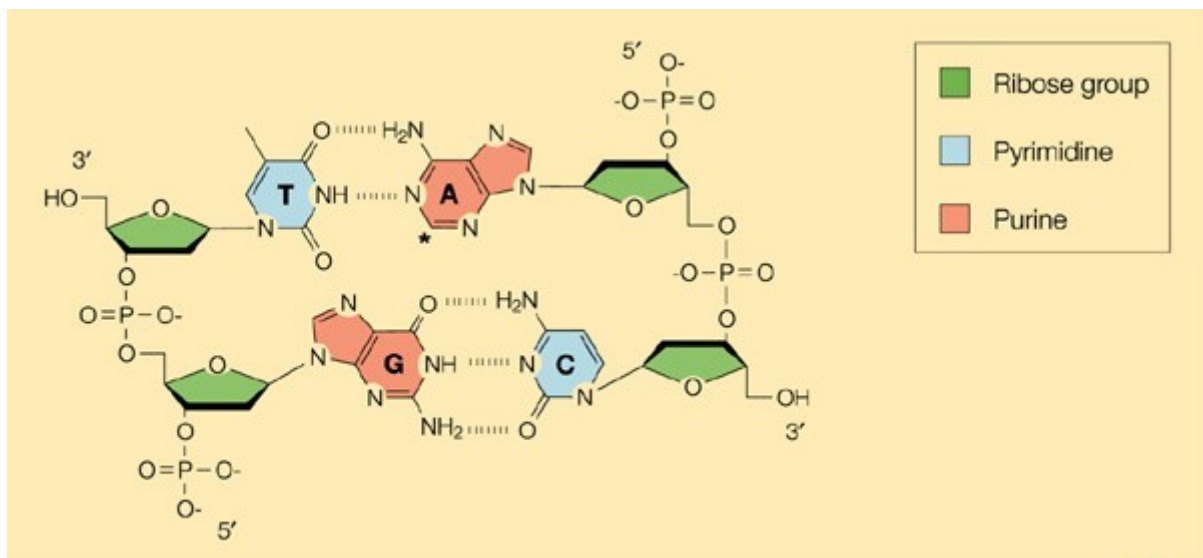
Preliminary evidence supports a role for TDG in the regulation of transcription, by directly affecting the assembly and progression of the transcription machinery (**Supplementary Results 4.4**). TDG's catalytic activity seems to be important for early elongation by RNA Polymerase II (RNA Pol II). At this early stage of transcriptional elongation, TDG might be in a complex with BRD4, which binds to acetylated histones and is a serine 2-kinase for the C-terminal domain (CTD) of RNA Pol II (Devaiah et al., 2012). This may directly link TDG's role in chromatin dynamics to the regulation of RNA Pol II-mediated transcription. Furthermore, the increased association of TDG to 3'UTRs of different splice variants of the *RARβ* locus indicated that TDG might be involved in co-transcriptional processes, like mRNA end processing or splicing.

Taken together, the work presented in my thesis contributes to the understanding of the epigenetic function of TDG-mediated active DNA demethylation with respect to time (during differentiation) and space (at gene regulatory elements). My **concluding model** depicts an attractive concept of a role for TDG in the maintenance of chromatin plasticity at critical genomic regions destined to undergo epigenetic regulation in response to developmental or environmental cues. In mESCs, these appear to be mainly CGI promoters and enhancers of developmental genes. The TET-TDG DNA demethylation machinery is recruited to these sites upon initiation of differentiation, presumably by developmental stage-specific transcription factors. There, TDG will excise TET-generated 5-fC and/or 5-caC and subsequent BER will produce single-strand breaks that will facilitate increased nucleosomal dynamics. This will enhance accessibility for the transcription machinery or for further chromatin modifying complexes. We propose that by the targeted demethylation of DNA, TDG assists in the maintenance of an open chromatin state and facilitates transcription.

2. Introduction

2.1 The Genome and Its Organization

Deoxyribonucleic acid (DNA) contains the genetic code, where every cell in an organism stores the information about its metabolism. This information can be copied and inherited to daughter cells and is mainly stored in “genes”, which are subject to tight regulation. The building blocks of DNA are the nucleotides: They consist of a base, the moiety containing the actual genetic information, and a sugar (deoxyribose) linked to a phosphate group. The bases are connected to the sugar-phosphate backbone by *N*-glycosidic bonds. In DNA there are four canonical bases: the purines Adenine (A) and Guanine (G) and the pyrimidines Cytosine (C) and Thymine (T); C methylated on the C5 position (5-methylcytosine; 5-mC) can be considered as a fifth base in the genome. In the cell nucleus, DNA exists as double-stranded α -helix, in which two anti-parallel complementary single-strands are paired. Complementarity is achieved through base pairing: A pairs with T via two hydrogen-bonds and C pairs with G via three hydrogen-bonds (**Figure 1**; (Ghosh and Bansal, 2003; Watson and Crick, 1953)).



Nature Reviews | Genetics

Figure 1: DNA base pairing according to Watson and Crick, with the hydrogen bonds indicated by the dashed lines. The bases are depicted in blue (pyrimidines) and red (purines) and sugar moieties are shown in green. The asterisk depicts the position, where A could have a third hydrogen bond (Szathmary, 2003).

The human genome consists of roughly 3 giga bases (Li, 2011) which translates to approximately 2 meters of DNA in every human cell. Thus, multiple rounds of

compaction are needed in order to fit the DNA into the nucleus of a cell which is only a few micrometers (μm) in size. For this reason, DNA is packed into chromatin, first by winding 147 base pairs (bp) of DNA around eight histone proteins (dimers of H2A, H2B, H3 and H4 each). This first compaction forms the nucleosome subunit and is called the primary chromatin structure. Further compaction with the help of linker histones (H1), which covers between 20 and 90 bp, induces zigzagging of the chromatin fibre. Continued looping of this roughly 30 nm thick structure can ultimately condense the DNA into a chromatin structure, which in its most compact form appears as a metaphase chromosome ((Sajan and Hawkins, 2012); **Chapter 2.3**).

2.1.1 Genetic Maintenance

The faithful maintenance and transmission of the DNA from one cell to another is of fundamental importance for living organisms. This, however, represents a major challenge, since DNA is under constant attack by physical and chemical agents of both exogenous and endogenous origin. Estimation of DNA damage events in a single human cell ranges from $10^4 - 10^6$ per day (Ciccia and Elledge, 2010; Hoeijmakers, 2009; Lindahl, 1993). Different kinds of DNA base damage can interfere with DNA-templated processes, such as replication and transcription, and thereby generate genetic mutations or induce cell death. These deleterious effects are counteracted by safe-guarding DNA repair mechanisms, which evolved manifold in order to assure the integrity of the genome; the most important DNA repair pathways are described in **Chapter 2.2** (reviewed in (Scharer, 2003)).

2.1.2 Epigenetic Maintenance

The different cell types of a multicellular organism contain the same genetic information – differences in cell morphology and function are established by epigenetic mechanisms. These effect cell-type-specific modifications on the DNA and histone proteins which program a specific reading of the genome. These modifications are thus referred to as epigenetic modifications; they include for instance methylation of the DNA, different chemical modifications on histone proteins as well as incorporation of non-canonical histone proteins. The greek prefix *epi*

means *on top*, hence suggesting that epigenetic modifications add a level of information onto the primary sequence information of the DNA. Genetic traits are thus determined first by the encoding DNA sequence including possible polymorphisms, and second, by their epigenetic state. In the revived field of epigenetics, many questions remained unanswered until now: How environmental cues are integrated into epigenetic instructions and whether and how these marks are inherited over generations is still not clear (Heijmans et al., 2008; Nilsson and Skinner, 2014; Wei et al., 2014). The mechanism of the so called “epigenetic memory”, which is responsible for the establishment and maintenance of stable patterns of epigenetic marks during ontogenesis is, although its existence is widely accepted, far from being understood (Cheedipudi et al., 2014). Accordingly, how and when histone modifications are deposited on newly assembled nucleosomes, how these modifications are maintained over time and how nucleosome reassembly throughout successive rounds of cell divisions is coordinated are questions that are heavily investigated at the moment (Alabert and Groth, 2012; Corpet and Almouzni, 2009). Although we know how DNA methylation is maintained in replicating cells, much less is known about whether and if yes, how DNA methylation patterns are inherited across generations.

2.2 Types of DNA Damage and Appropriate DNA Repair

Endogenous DNA damage concerns mostly DNA bases and can arise, for instance, from hydrolysis or reactive oxygen species (ROS), the latter resulting from endogenous metabolic processes. The reaction of ROS with DNA can give rise to more than 100 oxidative modifications in DNA (Beckman and Ames, 1997), most prominently the G or T oxidation products, 8-oxoguanine and thymine glycol, respectively. The former is a mutagenic, the latter a DNA and RNA polymerase blocking lesion. Contrarily, hydrolysis may lead to the cleavage of the *N*-glycosidic bonds between the base and the phosphate sugar backbone, leaving an apurinic/apyrimidic site or abasic site (AP-site). AP-sites are prone to generate mutations during replication (Choi et al., 2010). Moreover, AP-sites can lead to the formation of cytotoxic single-strand breaks and are thus highly deleterious for a cell. Furthermore, hydrolytic deaminations can occur at the exocyclic amino groups of C, 5-mC, A and G; generating U, T, hypoxanthine and xanthine bases, respectively

(Lindahl, 1993; Loeb and Preston, 1986). The most important DNA repair pathways for specific types of lesions will be briefly introduced below and are illustrated in **Figure 2**. Most relevant to this work is the DNA base excision repair pathway (BER), which is discussed in detail in a separate chapter (**2.2.1**).

Mismatch Repair (MMR)

Misincorporation of nucleotides by DNA polymerases during DNA replication or repair events is a major source of DNA mismatches. The average fidelity of the DNA polymerases Pol δ and Pol ϵ with an inherent proofreading activity is in the order of 1 error in 10^7 nucleotides synthesized. Contrarily, low fidelity DNA polymerases like Pol κ or Pol η , generate mismatches in the range of up to 1 in 10 synthesized nucleotides. The fact that the overall mutation rate in a human cell is only around 1 in 10^{10} nucleotides is thus largely owed to the ability of cells to recognize and correct DNA polymerase errors. The relevant mismatch correction activities are tightly associated with DNA replication and act downstream of the replication fork (Kunkel, 1992; Kunz et al., 2009b). Mismatched bases are usually not damaged and are *per se* not easy to distinguish from correctly paired bases. The mismatches are rather detected by unusual base-base pairing interactions. Importantly, repair has to be directed to the nascent DNA strand, since this is where the error was caused by DNA polymerases. It is not entirely clear, how strand discrimination and strand-directed repair is achieved in eukaryotes; presumably strand discontinuities serve this purpose. Upon the encounter of a mismatch, the mismatch recognition complexes MutS α or MutS β bind the mismatch. MutS α and MutS β are heterodimeric complexes consisting of the MSH2 and MSH6 or MSH3 proteins, respectively, and can be distinguished due to their complementary modes of mismatch recognition. Subsequently, the MutL complexes (MutL $\alpha/\beta/\gamma$) are recruited to MutS α/β . The function of the MutL complexes is not fully understood; however, they are also heterodimeric complexes, consisting of hMLH1 and hPMS2, hPMS1 or hMLH3, respectively. Upon the assembly of DNA replication proteins like the proliferative cell nuclear antigen (PCNA) and replication factor C (RFC) at the MutS/MutL complex, MutL α can act as an endonuclease, nicking the discontinuous strand 3' and 5' of the assembled complex. This generates an entry point for the exonuclease EXO1. EXO1 is activated by MutS α and degrades the newly synthesized strand in 5'-3' direction

towards and past the mismatch. DNA re-synthesis is achieved by Pol δ and the nick is sealed by DNA ligase I (reviewed in (Kunz et al., 2009b)). Defects in MMR genes, like MSH2 and MLH1, can predispose to cancer, such as in the hereditary non-polyposis colon cancer (HNPCC) familial form of colon cancers (Bronner et al., 1994; Leach et al., 1993).

Nucleotide Excision Repair (NER)

NER is the major pathway that removes bulky base lesions that thermodynamically destabilize the DNA, applying a “cut out and refill” mechanism. Accordingly, NER displays a remarkably broad substrate spectrum. Examples include UV-induced photoproducts (cyclopyrimidine dimers, 6-4 photoproducts), certain oxidative lesions (cyclopurines) as well as adducts formed by environmental mutagens like benzo [a] pyrene, which is found in cigarette smoke, or adducts formed by cancer chemotherapeutic drugs, such as cisplatin (reviewed in (Scharer, 2013)). Common to all of these lesions is their distorting effect on the DNA helical structure – there appears to be a positive correlation between the efficiency of repair and the degree of helical distortion caused by a lesion (Gunz et al., 1996; Sancar, 1996). There are two subpathways of NER: global genome NER (GG-NER) and transcription-coupled NER (TC-NER). GG-NER can occur anywhere in the genome and is initiated by the GG-NER-specific factor XPC-RAD23B, whereas TC-NER assures the rapid repair of lesions in the transcribed strand of active genes and is initiated by a stalled RNA polymerase and the TC-NER-specific factors CSA, CSB and XAB2. Both subpathways require the subsequent core NER system to restore the intact DNA, engaging roughly 30 proteins. After damage recognition, the general transcription and NER factor TFIIH is recruited to the site of the lesion. TFIIH consists of multiple proteins, whereof the helicases XPB and XPD are particularly important in NER, since they unwind the DNA around the lesion. This triggers the recruitment of XPA, the ssDNA binding protein RPA and the endonuclease XPG in order to form the preincision complex. Next, the endonuclease ERCC1-XPF is recruited to the NER complex by interaction with XPA. Once the two endonucleases (XPG and ERCC1-XPF) are positioned, dual incision is initiated. The first incision is made by ERCC1-XPF 5' to the lesion, the lesion-containing oligonucleotide is thereby released from the NER factors with TFIIH bound to it. DNA repair synthesis is then initiated by DNA

polymerases Pol δ , Pol ϵ or Pol κ together with the standard replication factors. Finally, XPG cleaves 3' of the lesion, allowing the final release of a 24-32 oligonucleotide as well as the sealing of the remaining nick by DNA Ligase I (reviewed in (Scharer, 2013)). Defects in NER factors are associated with several genetic disorders: (i) xeroderma pigmentosum is linked to defects in one of the seven xeroderma pigmentosum complementation groups (XPA through XPG) and is a GG-NER defect, (ii) Cockayne Syndrome is caused by defects in CSA, CSB, XPB, XPD and XPG and reflects a TC-NER defect, and (iii) trichothiodystrophy is caused by defective subunits of TFIIH (reviewed in (Scharer, 2013)).

Interestingly, several NER factors have been recently implicated in epigenetics as well as in regulation of transcription. For instance, the NER complex together with Gadd45 was shown to be recruited to promoters of rRNA genes by TAF12, which caused DNA demethylation at these sites (Schmitz et al., 2009). Also, the NER complex was suggested to facilitate a chromatin state that allows for transcription at active promoters, even in the absence of genotoxic attack (Le May et al., 2010). Furthermore, the endonucleases XPF and XPG were shown to promote chromatin looping together with CTCF, as well as DNA nicking, which finally caused DNA demethylation, indicating important additional functions of NER proteins in transcription control (Le May et al., 2012).

Double-strand break repair

DNA double-strand breaks (DSB) can be generated during endogenous processes, for instance during replication, either by the replication toward a single-strand break (SSB) in the template strand or by the collapse of a replication fork (Cox et al., 2000). Additionally, they can arise from sugar lesions that frequently disrupt the DNA backbone, thereby generating SSBs and DSBs, if these lesions occur in a clustered manner (Singh et al., 2011). Exogenous agents, such as ionizing radiation (e.g. X-rays), can also lead to the formation of DSBs (Mahaney et al., 2013). Taking the severity of the lesion into account, it is not astonishing that a single unrepaired DSB in a yeast cell can lead to cell death (Sandell and Zakian, 1993). The challenge of DSB repair lies in the acquisition of an appropriate homologous template, since no conventional template on the opposite strand is available for repair. There are two

pathways that repair DSBs that deal differently with this. Homologous recombination (HR) uses stretches of DNA homology on sister chromatids, present in S and G2 phases of the cell cycle, as repair templates and is thus an accurate repair pathway (reviewed in (van Gent et al., 2001)). HR involves the resolution of a complex repair intermediate connecting the broken and intact DNA duplexes, known as Holiday junction. Contrarily, in nonhomologous end joining (NHEJ), two broken DNA ends are simply rejoined during G1 phase. This involves DSB recognition, end processing and bridging as well as ligation steps and can be achieved with or without micro-homologies at the DNA ends. Therefore, the NHEJ pathway is not necessarily accurate and small deletions may result at the site of the resealed DSB (reviewed in (Scharer, 2003)).

Defective DSB repair can predispose to cancer. For instance, loss of function mutations of breast cancer 1 or 2 (BRCA1/2), which is involved in cell cycle checkpoint activation and DSB repair, was shown to increase the susceptibility for breast or ovarian cancer drastically (O'Donovan and Livingston, 2010).

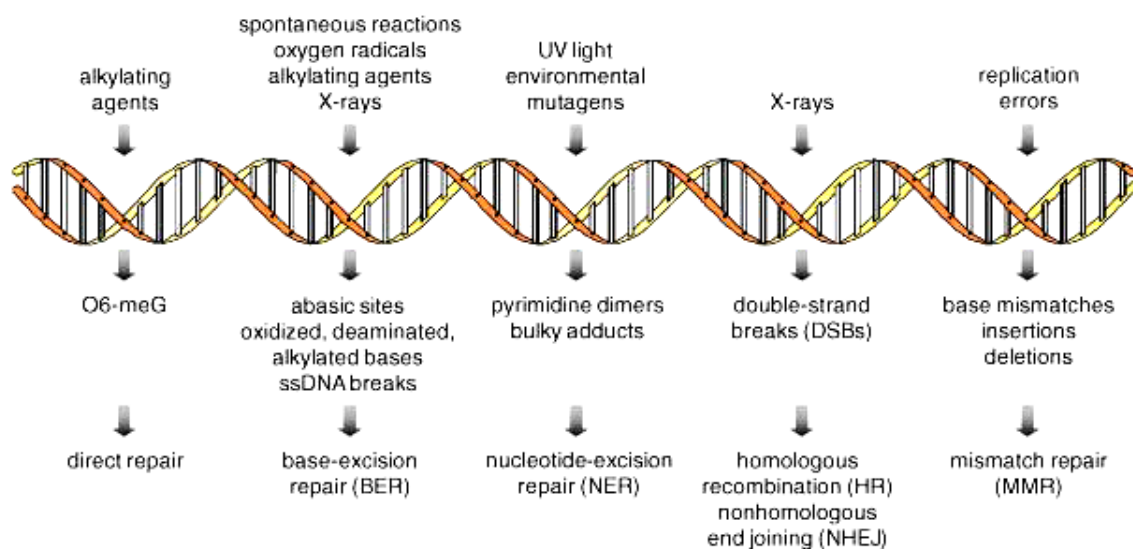
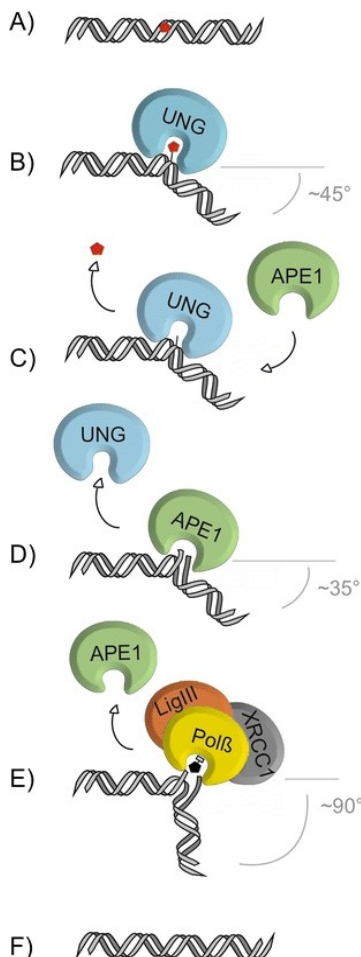


Figure 2: The most common DNA-damaging agents, the lesions they cause and the DNA repair pathways that restore the original state (Scharer, 2003).

2.2.1 DNA Base Excision Repair (BER)

DNA is constantly engaged in chemical reactions in its cellular environment, which can cause damage to the coding bases, the most vulnerable moieties in DNA. The types of damages include deamination, oxidation and alkylation and are usually non-distorting (Lindahl and Wood, 1999; Scharer, 2003). Base lesions can be pro-mutagenic when altering the Watson-Crick pairing properties of the base, meaning that they can give rise to genetic mutations if not repaired. Examples for this include the hydrolytic deamination of C or 5-mC which generates U or T, causing C-G → U-A or C-G → T-A transition mutations if replicated prior to repair, respectively. Another example is 7,8-dihydro-8-oxoguanine (8-oxoG) which arises through oxidation of G, and favors base-pairing with A, thereby giving rise to G-T → T-A transversion mutations, if unrepaired. Base lesions caused by alkylation, such as N7-



methylguanine (7-meG) or N3-methyladenine (3-meA), are cytotoxic because they can block replicative DNA polymerases (Lindahl and Wood, 1999). All these damages are mainly repaired by BER, which is initiated by damage-specific DNA glycosylases recognizing a cognate type of damaged base (**Chapter 2.2.2**; reviewed in (Jacobs and Schar, 2012)). The excision of the base occurs by flipping the base into the catalytic pocket of the DNA glycosylase where the *N*-glycosidic bond between the base and the sugar phosphate backbone of the DNA is hydrolyzed, leading to an AP-site in the DNA strand (**Figure 3**). Some DNA glycosylases display a very high affinity to AP sites; the dissociation from their product can be facilitated by the recruitment of downstream BER factors or also by posttranslational modifications (PTMs; (Hardeland et al., 2002; Parikh et al., 1998; Waters et al., 1999)). Hence, the release of the base and downstream repair events are well orchestrated. The AP endonuclease (APE1) subsequently

Figure 3: The damaged base [A] is recognized by a DNA glycosylase, which flips the base into its catalytic pocket and thereby kinks the DNA [B]. After base excision, APE1 is recruited to the glycosylase [C], which triggers the release of the glycosylase from the AP-site [D]. Pol β fills the gap, the nick is sealed by Lig III [E], thus restoring the initial state [F] (Jacobs and Schar, 2012).

hydrolyzes the phosphodiester bond 5' to the AP-site to generate a nick with a 3'-OH and a 5'- deoxyribose-5-phosphate (5'-dRP) end. In the major BER pathway (short-patch repair), Polymerase β (Pol β) will then incorporate a single nucleotide after it processed the 5' end of the nick in order to reconstitute the necessary 5'-phosphate (5'-P). This is achieved with Pol β 's additional AP lyase activity that excises the initial 5' dRP. Some DNA glycosylases are bifunctional and possess an additional AP lyase activity, so that the AP lyase activity of Pol β is not required. The remaining nick in the DNA backbone is then sealed by DNA ligase III (Lig III), which is bridged to Pol β by X-ray repair cross-complementing group 1 (XRCC1). Generally, the processing of the AP-site appears to be coordinated by protein-protein interactions: APE1 interacts with Pol β and XRCC1 directly; Pol β interacts with LigIII through XRCC1 (reviewed in (Scharer, 2003)).

There exists also a minor BER pathway (long-patch repair), where DNA synthesis is achieved by DNA polymerases δ/ϵ which introduce 2-6 nucleotides. The resulting oligonucleotide overhang is cleaved by flap endonuclease 1 (FEN1) and the nick is sealed by DNA Ligase I (Lig I; (Dianov and Lindahl, 1994; Pascucci et al., 1999; Scharer, 2003)).

2.2.2 DNA Glycosylases

The first DNA glycosylase was discovered in 1974. It was assumed that there must be a mechanism in cells that eliminates deoxyuridine monophosphate (dUMP) which is either misincorporated during DNA replication or arises by hydrolytic deamination of C (Lindahl, 1974). This led to the discovery of the uracil N-glycosidase (UDG) in extracts of *Escherichia coli* cells, which then ignited the identification of other DNA glycosylases in all kingdoms of life. Up to date, eleven DNA glycosylases have been identified in mammals and they can be classified into four structurally distinct super-families. These are (i) the uracil DNA glycosylases (UDGs), (ii) the helix-hairpin-helix (HhH) glycosylases, (iii) the 3-methyl-purine glycosylase (MPG) and (iv) the endonuclease VIII-like (NEIL) glycosylases (reviewed in (Jacobs and Schar, 2012)).

Different DNA glycosylases evolved to recognize specific types of base damage. However, they all use a common base flipping mechanism to accomplish catalysis, independent of the size of the catalytic pocket. Thereby, the base is pushed into the

catalytic pocket of the enzyme through intercalation of the latter in the minor groove of DNA. There are both very specific and rather unspecific catalytic pockets. For instance, the uracil-N glycosylase (UNG) and the single-strand specific monofunctional uracil DNA glycosylase 1 (SMUG1) establish specific contacts with the uracil (U) to be excised (Mol et al., 1995a; Mol et al., 1995b; Savva et al., 1995). UNG has a very tight-fitting catalytic pocket which contributes to its high substrate specificity (Krokan et al., 2001). In contrast, the MUG and TDG catalytic pockets can accommodate a variety of pyrimidine and purine derivatives without contacting the base to be hydrolyzed (Barrett et al., 1999), giving them a rather broad substrate spectrum. In case of the latter enzymes, damage recognition specificity is additionally ensured by interactions with the complementary DNA strand opposing the damaged base. This may explain the double-strand dependency of MUG proteins and the preference for a G opposite the damaged base (Barrett et al., 1998). Lately, this view has been challenged though, since TDG was found to form numerous contacts with the strand containing the AP-site, involving several phosphodiester groups, the abasic sugar as well as the Gs of both strands in the CpG dinucleotide containing the damaged base (Maiti et al., 2008). Furthermore, the recognition of 5-caC in TDG's catalytic pocket appears to be an exception, as it is highly specific compared to other substrates and its excision apparently functions in a pH-dependent manner (Maiti et al., 2013; Zhang et al., 2012).

Mechanistically, there are two classes of DNA glycosylases. Monofunctional DNA glycosylases perform only base excision and use a strategically positioned water molecule as a nucleophile to cleave the *N*-glycosidic bond (reviewed in (Scharer, 2003)). Contrarily, bifunctional glycosylases combine base excision with an additional AP-lyase step. They use a catalytic lysine side chain of the enzyme which is suitably positioned as a nucleophile to attack the *N*-glycosidic bond. This generates a 3'-OH and 5'-P end, which can directly be further processed (Bruner et al., 2000; Jacobs and Schar, 2012; Lau et al., 1998; Scharer, 2003).

Some DNA glycosylases, such as TDG and to a lesser extent also UNG, have a higher affinity for the product AP-site they generated compared to the actual substrate (Parikh et al., 1998; Waters et al., 1999). TDG thus additionally displays a strong product inhibition and remains bound to the product, until either APE1 is recruited and/or TDG gets posttranslationally modified by Small Ubiquitin-like

Modifier (SUMO), which induces a conformational change and causes TDG's dissociation (Hardeland et al., 2002; Steinacher and Schar, 2005; Waters et al., 1999). Thus, the release of DNA glycosylase from the AP-site can be considered the rate limiting step in BER initiated by TDG. In contrast, the turnover of UNG is extremely rapid and orders of magnitude higher than that of other uracil-DNA glycosylases (600-1000 nucleotides per minute; (Krokan et al., 2002)). Differences persist also in the timed action of glycosylases: whereas UNG is replication-associated and removes Uracil misincorporated by replicative polymerases, TDG is downregulated during S phase (Hardeland et al., 2007). TDG is highly expressed in G2 and G1 phases of the cell cycle, when UNG is downregulated, thus these two UDG family members underlie strictly anticyclic cell cycle regulation (Hardeland et al., 2007).

2.2.3 The Thymine DNA Glycosylase (TDG)

TDG is the most relevant DNA glycosylase for this work, thus I am specifically focusing on its biochemical and biological characteristics in this chapter.

Base Excision Mechanism

TDG belongs to the super-family of monofunctional uracil-DNA glycosylases (UDGs) and the family of MUG enzymes (reviewed in (Cortazar et al., 2007)). The current model is that upon contact with DNA, TDG undergoes a conformational change in its N-terminal domain, which results in a clamp-like structure that allows TDG to slide along the DNA (Steinacher and Schar, 2005). How exactly the damaged base is recognized is not entirely clear yet (reviewed in (Jacobs and Schar, 2012)). Once a G-mismatched substrate is encountered, the substrate base is flipped out of the DNA helix into the catalytic pocket of the enzyme. This is achieved by an insertion loop wedging into the DNA helix. This wedge stabilizes the base stack and forms specific hydrogen bonds with the widowed G (Barrett et al., 1998; Barrett et al., 1999; Maiti et al., 2008). The release of TDG from its product is facilitated by its PTM with SUMO. C-terminal SUMOylation of TDG significantly decreases the affinity for both substrate and product, increasing the turnover of the enzyme (Hardeland et al., 2002; Steinacher and Schar, 2005). Moreover, TDG can become acetylated by CBP/p300 as well as phosphorylated through the protein kinase C (PKC α). These modifications

influence enzyme turnover, interaction with other proteins or localization in the cell (Hardeland et al., 2002; Madabushi et al., 2013; Mohan et al., 2010), indicating that PTMs contribute critically to the coordination of TDG-initiated BER.

Substrates for TDG

Initially discovered as an enzyme removing T from G•T mismatches believed to arise from deamination of 5-mC, it later turned out, that TDG possesses a much larger substrate spectrum (**Table 1**; (Neddermann et al., 1996)). Besides processing T and U resulting from hydrolytic deamination of 5-mC and C, respectively, TDG was also shown to excise larger uracil-derivatives like 5-hydroxymethyluracil (5-hmU), 5-fluorouracil (5-FU) or 5-bromouracil (5-BrU) as well as bases with etheno-adducts (e.g. 3,N₄-ethenocytosine) or oxidized pyrimidines (e.g. thymine glycol; (Hardeland et al., 2003)). Interestingly, a 5-mC processing activity has been described for chicken TDG (Jost, 1993; Zhu et al., 2000). This activity, however, is questionable, since it could never be reproduced.

Table 1: Substrate* spectrum of human recombinant full length TDG

Base release efficiencies			
high	intermediate	low	insignificant
G• U	A• FU	A• U	ss U
G• FU	ss FU	A• BrU	ss BrU
G• BrU	G• Tg	G• Hx	ss εC
G• hmU	A• εC		T• Hx
G• hU	ss caC		ss Hx
G• T			G• εA
G• εC			ss εA
G• fC			G• hmC
G• caC			G• heC
			G• hpC
			G• G
			G• mC

* The putative substrate base is marked in bold letters.

ss, single strand; F, fluoro-; Br, bromo-; h, hydroxy-; hm, hydroxymethyl-; Tg, thymine glycol; ε, etheno-; Hx, hypoxanthine; f, formyl-; ca, carboxyl-; he, hydroxyethano-; hp, hydroxypropano-. Information collected from (Cortazar et al., 2007; Hardeland et al., 2003; Hardeland et al., 2001; He et al., 2011; Maiti and Drohat, 2011) and Alain Weber, unpublished data.

Notably, TDG was recently shown to excise 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), intermediates of active DNA demethylation (He et al., 2011; Maiti and Drohat, 2011). Importantly, TDG was shown to also process two single-stranded (ss) substrates, namely ss-5-caC and ss-5-FU ((Hardeland et al., 2000) and Alain Weber; unpublished data), establishing that it is not strictly mismatch-dependent. Interestingly, TDG shows redundancy with other DNA glycosylases for certain substrates, as for example U, which may also be excised by SMUG1, UNG and MBD4 (Visnes et al., 2009). However, there are at least two characteristics rendering TDG unique when comparing it to other DNA glycosylases: (i) TDG is the only glycosylase whose knockout causes embryonic lethality (Cortazar et al., 2011; Cortellino et al., 2011) and (ii) TDG is thus far the only glycosylase capable to excise 5-fC and 5-caC (He et al., 2011; Maiti and Drohat, 2011). Thus, albeit TDG was initially discovered as a DNA repair enzyme, it appears that its role in control of DNA methylation represents its primary task (Zheng et al., 2014).

Biological functions of TDG

The embryonic lethality of *Tdg* knockout mice was an unexpected and unique phenotype for a DNA glycosylase defect. Given that no other DNA glycosylase knockout shows such a severe phenotype and that there are other DNA glycosylases with redundant substrate spectra, pointed towards additional roles for TDG apart from classical DNA repair. Consequently, TDG has been implicated in several biological processes not immediately related to DNA repair, such as immunity, regulation of gene transcription, control of DNA methylation as well as epigenetic maintenance (reviewed in (Jacobs and Schar, 2012)).

TDG was shown to interact with the nuclear receptors retinoic acid receptor (RAR) and retinoid X receptor (RXR) and to enhance their binding to their respective response elements, thereby influencing the transcriptional outcome (Um et al., 1998). Another nuclear receptor, the estrogen receptor α (ER α), was also found to interact with TDG, where TDG is acting as a co-activator of transcription for ER α -regulated genes (Chen et al., 2003). Next, TDG was demonstrated to act as a transcriptional repressor at thyroid transcription factor 1 (TTF-1) targeted genes (Missero et al., 2001). Furthermore, TDG was shown to interact with the histone acetyltransferases

CREB-binding protein (CBP)/p300 and the resulting CBP-TDG complex is capable of binding DNA, repairing G•T/U mismatches and acetylating histones (Tini et al., 2002). TDG enhances the transcriptional co-activator function of CBP and serves itself as a substrate for acetylation by CBP. It was proposed that TDG acetylation abolishes the recruitment of APE1 (Tini et al., 2002). Furthermore, TDG was shown to act as a co-activator of transcription for SRC1 family members (Lucey et al., 2005) and p53 family proteins (Kim and Um, 2008).

DNA methylation control and epigenetic maintenance

We and others showed before, that loss of TDG leads to aberrant DNA methylation at promoters of developmental genes (Cortazar et al., 2011; Cortellino et al., 2011), as well as changes in histone modifications (**Appendix III**). The recent finding that TDG processes 5-fC and 5-caC clearly positions TDG as a key player in active DNA demethylation (**Chapter 2.5**; (He et al., 2011; Ito et al., 2011; Kohli and Zhang, 2013; Maiti and Drohat, 2011)). It emerges that TDG-mediated targeted DNA demethylation serves the maintenance of cellular plasticity during development through DNA methylation control at gene regulatory regions, further linking control of DNA methylation and regulation of transcription through TDG (**Appendix II**). I contributed to these discoveries during my PhD thesis; they are discussed in more detail in the results section.

2.3 Chromatin

In order to fit roughly 2 meters of DNA into a cell nucleus with an average diameter in the range of a few micrometers μm , organized compaction of the DNA is indispensable (**Figure 4**). The DNA primary structure is arranged into nucleosomes, with approximately 147 bp of DNA wrapped around a histone octamer. A nucleosome consists of two dimers of each of the canonical histones H2A, H2B, H3 and H4 (reviewed in (Horn and Peterson, 2002)). Additionally, canonical histone proteins can be exchanged for histone variants as e.g. H2A.Z or H3.3, and ATP-dependent nucleosome remodeling allows the control of nucleosomal dynamics (see **Chapter 2.3.1**). The highly basic N-terminal tails of these histone proteins protrude out of the

nucleosome and can be chemically modified by e.g. methylation or acetylation (reviewed in (Bannister and Kouzarides, 2011)). At the moment, at least eight distinct types of histone modifications are known and they can be placed on more than 60 different residues on the different histone tails. Histone modifications are capable to dictate the higher-order structure of chromatin (reviewed in (Kouzarides, 2007)).

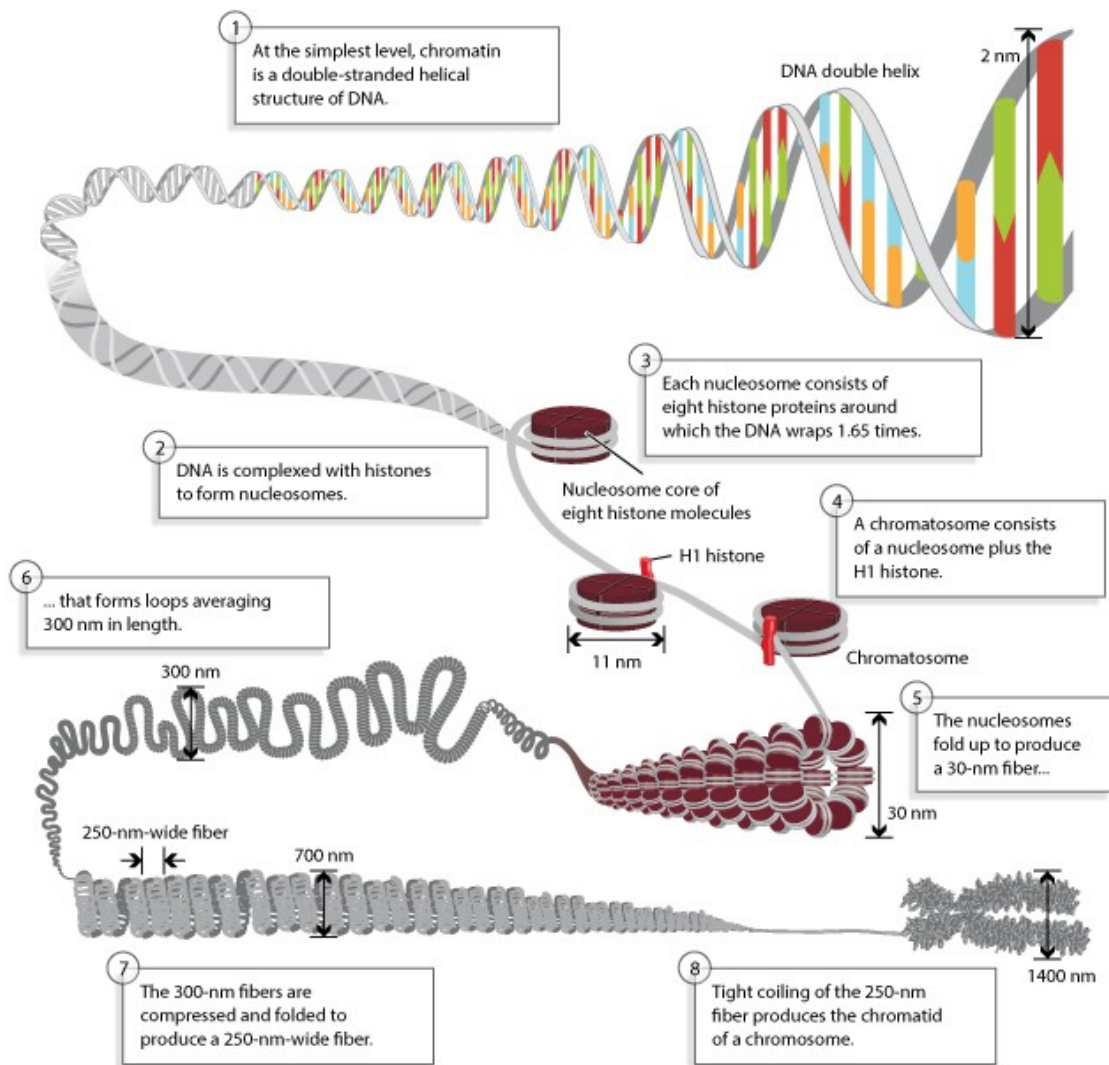


Fig 4: Condensation of the DNA into nucleosomes and higher-order chromatin (From: Annunziato, A. (2008), DNA packaging: Nucleosomes and chromatin. Nature Education 1(1):26).

Of all histone modifications identified so far, acetylation has the highest potential to impact the chromatin structure because it neutralizes the basic charge of the lysine, which leads to a less compacted chromatin state (Shogren-Knaak et al., 2006). The most important modifications and their preferential localization are summarized in **Chapter 2.3.2**. The primary chromatin structure is then further compacted with the help of linker histones (H1), which cover between 20 and 90 bp of DNA and induce zigzagging, resulting in a roughly 30 nm thick chromatin fiber. Additional looping and

tight coiling finally gives rise to one chromatid of a chromosome (Sajan and Hawkins, 2012). A certain consistency in the combination of these variable modifications provoked the creation of the term “histone code” (Jenuwein and Allis, 2001). Modifications are deposited by “writers” such as histone methyltransferases (HMTs), recognized by “readers” such as the CBX subunit of the polycomb repressive complex 1 (PRC1) and eliminated by “erasers” such as histone deacetylases (HDACs; (Jakovcevski and Akbarian, 2012)). The most important chromatin modifying complexes and their regulation are discussed in **Chapter 2.3.3**.

If chromatin is further compacted into a (constitutively) repressed heterochromatic state, it is often associated with the nuclear lamina and does not allow binding of DNA-associated proteins. Contrarily to that, transcription can occur at euchromatic regions where DNA-binding proteins still have access to the DNA due to the low level of its compaction (reviewed in (Zhou et al., 2011)). Histone modifications can thus either serve to establish global chromatin environments or to orchestrate DNA-based biological tasks, thereby influencing many key metabolic processes (reviewed in (Kouzarides, 2007)).

2.3.1 Chromatin Remodeling and Histone Variants

Chromatin remodeling complexes regulate the higher order chromatin structure and thereby the access to DNA and histone proteins. These complexes consist of numerous subunits and function in an ATP-dependent manner. Histone–DNA interactions must be loosened to facilitate access of protein complexes involved in DNA metabolic processes, such as transcription, replication, recombination and repair. This is achieved through the hydrolysis of ATP by chromatin remodeling complexes, which operate similar to DNA translocases to disrupt electrostatic interactions between DNA and histone proteins to facilitate nucleosome sliding, DNA exposure and nucleosome exchange (reviewed in (Hargreaves and Crabtree, 2011)). In mammals, there exist four different families of ATP-dependent chromatin remodeling complexes: SWI/SNF, INO80/SWR1, ISWI and CHD (reviewed in (Hargreaves and Crabtree, 2011)).

Contrarily to canonical histones which are deposited behind the replication fork during S phase, histone variants are incorporated independently of DNA replication

(reviewed in (Jin et al., 2005)). It emerges that most histone variants are deposited at preferential genomic locations and that they have relevant biological functions, for instance in DNA damage signaling or transcription. An example for the latter is H3.3, a variant of canonical H3: H3.3 is associated with actively transcribed regions and relatively enriched in PTMs characteristic for transcribed genes (H3K4me2, H3K4me3, H3K9ac, H3K14ac, H3K18ac, H3K23ac; (McKittrick et al., 2004; Wirbelauer et al., 2005)). The canonical histone H3, in contrast, is rather enriched in silenced chromatin domains that are positive for H3K9 methylation (McKittrick et al., 2004). Furthermore, H3.3 functions as an insulator against heterochromatin spreading in *Drosophila melanogaster* (Nakayama et al., 2007). H2A.X, a variant of H2A, is an ubiquitously incorporated histone variant and in its phosphorylated state (γ H2A.X), it is a marker for DNA lesions and initiates DNA damage response (Rogakou et al., 1998). Another variant of H2A, H2A.Z, is preferentially incorporated into promoter-proximal regions and plays a positive role in transcription. Whereas H3.3 is distinguished from H3 in only four amino acid residues, H2A.Z only shows 60% sequence identity with canonical H2A (reviewed in (Jin et al., 2005)). Still, the above mentioned histone variants are highly conserved in eukaryotes, further consolidating their fundamental biological roles (reviewed in (Jin et al., 2005)). Of course, also the regulation of the deposition of canonical histones plays a vital role, for instance in transcription: RNA polymerase II was shown to preferentially bind to nucleosomes that are depleted of one H2A-H2B dimer (Baer and Rhodes, 1983). Furthermore, elongation by RNA polymerase II contributes to histone dimer release (Levchenko et al., 2005; Levchenko and Jackson, 2004). This is in accordance with gene activation usually being accompanied by disassembly of nucleosomes in promoter regions (reviewed in (Henikoff and Ahmad, 2005; Jin et al., 2005)). The assembly of nucleosomes is assisted by specific histone chaperones, the histone regulatory homolog A (HIRA) e.g. deposits H2A variants and H3.3. Depending on the specific genomic location, H3.3 is deposited by different histone chaperones: HIRA deposits H3.3 at both active and repressed genes but not at telomeres and many TF binding sites; in contrast, ATRX (a member of SNF2 chromatin remodelers) is responsible for H3.3 deposition at telomeres and the death-domain associated protein (DAXX) incorporates H3.3 at pericentromeres (Drane et al., 2010; Goldberg et al., 2010; Lewis et al., 2010; Morozov et al., 2012). Contrarily, the canonical histone H3 is mainly deposited by the histone chaperone chromatin assembly factor

1 (CAF1; (Jin et al., 2005)). Interestingly, the chromatin remodeling complex INO80, which regulates the cell cycle checkpoint at G2/M for instance after a DSB, shows a preference for the incorporation of H2A.X and H2A.Z (Morrison and Shen, 2005; Papamichos-Chronakis et al., 2006; van Attikum and Gasser, 2005a; Watanabe and Peterson, 2010). Importantly, it was reported recently that conserved variation between amino acid residues on H3 may confer specificity to histone modifying enzymes. Specifically, it was shown in plants that the H3K27 methyltransferase ATXR5 methylates alanine 31 in H3.1 but not threonine 31 in H3.3. This can provide a way to assure accuracy in the mitotic inheritance of histone modifications (Jacob et al., 2014).

It was suggested that due to its independency of replication, H3.3 is incorporated to replace the nucleosomes that were evicted during the transcription of highly active genes (Wirbelauer et al., 2005). Yet, H3.3 may also serve to mark sites of continuous histone turnover to maintain accessibility to gene regulatory elements (reviewed in (Szenker et al., 2011)). H3.3 impairs higher-order chromatin folding, although it does not significantly alter the stability of mononucleosomes (Chen et al., 2013a). This is in line with the role of H3.3 in transcription: H3.3 is actively deposited into enhancers in mouse ESCs prior to gene induction by retinoic acid (RA; (Chen et al., 2013a)). Upon RA-dependent gene activation, H3.3 is then depleted from enhancers and deposited into the promoter region (Chen et al., 2013a). This suggests that H3.3 plays a critical role in the activation of inducible genes (Chen et al., 2013a). Furthermore, in mouse ESCs, H3.3 is highly enriched at the promoters of developmental genes characterized by a bivalent chromatin state (Goldberg et al., 2010). Here, the deposition of H3.3 by HIRA facilitates the binding and activity of PRC2 to establish bivalency in mouse ESCs (Banaszynski et al., 2013). A summary of H3.3 “deposition dynamics” and turnover is shown in **Figure 5**. In my thesis, I addressed a possible role of the regulation of DNA methylation in H3.3 deposition; these data are presented in more detail in **Appendix II**.

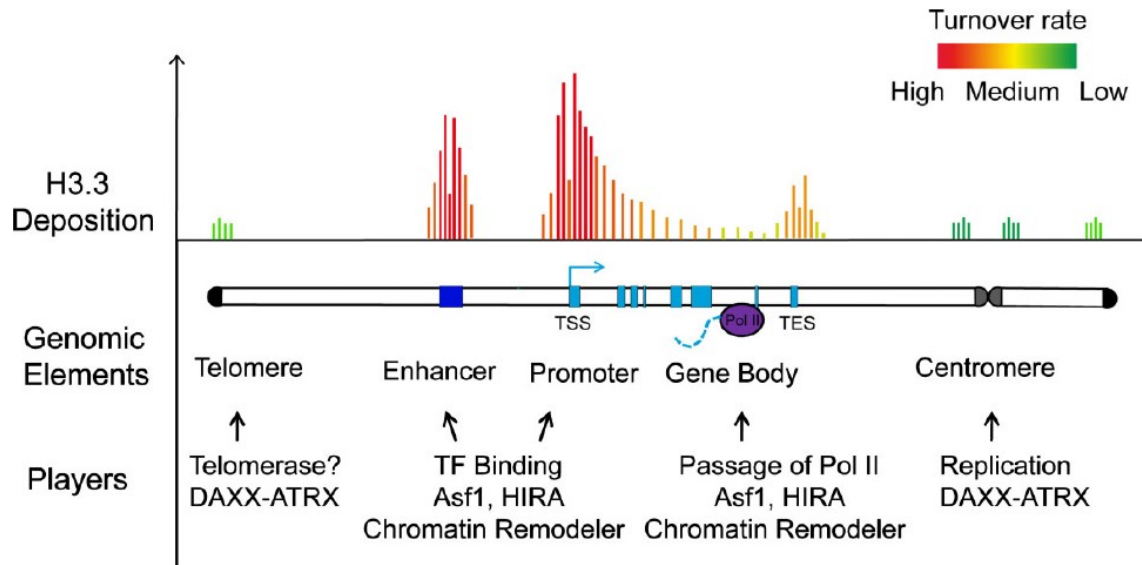


Figure 5: Preferential deposition of H3.3 at distinct genomic regions and the respective turnover profiles are depicted schematically (Huang and Zhu, 2014).

2.3.2 Genomic Region- and Context-Specific Histone Modifications

Active promoters

Gene promoter regions are characterized by an enrichment of histone 3 lysine 4 trimethylation (H3K4me3), independently of the expression status of the respective gene (Guenther et al., 2007). Methylated H3K4 is selectively recognized by the PHD finger domain of TAF3, which is part of the general transcription factor TFIID. This interaction can generate a docking site for further proteins and finally activate transcription (Vermeulen et al., 2007). H3K4me3-positive sites are often accompanied by histone acetylation: There exist positive correlations both for H3K4 methylation with H3 acetylation at lysines 9 and 18 as well as for the valency of the two modifications (Nightingale et al., 2007; Wang et al., 2001). Additionally, H3K4me3-positive sites are often occupied by the H3.3 histone variant (Goldberg et al., 2010; Guenther et al., 2007; Kim et al., 2005; Mikkelsen et al., 2007; Wang et al., 2008). Chromatin in an open, accessible conformation is devoid of DNA methylation, an explanation for this might be that methylation of H3K4 does not allow a physical interaction between the histone tail and Dnmt3L (Meissner et al., 2008; Mohn et al., 2008; Ooi et al., 2007). Many of these H3K4me3-positive promoter regions are also

bound by RNA Pol II and transcription is initiated there, although not all of these promoters produce detectable terminated transcripts (Guenther et al., 2007).

Repressed promoters

The typical chromatin mark for repressed promoters is the trimethylation of histone 3 at lysine 27 (H3K27me3). Often, repressed chromatin also contains histone 3 which is trimethylated at lysine 9 (H3K9me3) and the respective DNA is methylated (Zhou et al., 2011). Importantly, it appears that deacetylation of histones precedes the silencing of genes by H3K9 methylation, as the H3K9-specific histone methyltransferase SUV39h1 was found to interact with several histone deacetylases, explaining both the physical and functional interaction (Vaute et al., 2002).

Bivalent promoters, which are strongly associated with developmental genes and CpG islands (CGIs), are positive for H3K4me3 and H3K27me3 (Bernstein et al., 2006). These bivalent genes are poised for transcription; they can either be rapidly induced or shut down, depending on the developmental requirement (reviewed in (Aloia et al., 2013)).

Gene bodies

The gene bodies of transcribed genes are predominantly trimethylated on histone 3 at lysine 36 (H3K36me3) and dimethylated on histone 3 at lysine 79 (H3K79me2; (Mikkelsen et al., 2007; Schubeler et al., 2004)). Compared to introns, H3K36me3 is particularly strongly enriched in expressed exons (Kolasinska-Zwierz et al., 2009). This might be due to the interaction of RNA Pol II with SET2, the histone methyltransferase for H3K36 (Li et al., 2003).

Enhancers

The typical histone mark at enhancers is monomethylation of histone 3 at lysine 4 (H3K4me1) while H3K4me3 is absent (Heintzman et al., 2007). In contrast to poised enhancers, which are marked by H3K4me1 alone, active enhancers are characterized by additional acetylation of histone 3 at lysine 27 (H3K27ac; (Creyghton et al., 2010)). H3K27ac, together with the deposition of the histone variant H2A.Z, correlates with gene expression levels of the downstream target gene (Ernst and Kellis, 2010).

Heterochromatic Regions

Heterochromatin is very condensed and can be divided into two different types, constitutive and facultative heterochromatin. Constitutive heterochromatin is considered to remain compacted throughout an organism's life span, such as the pericentric satellite repeats and telomeres (reviewed in (Craig, 2005)), whereas facultative heterochromatin is established during development, for instance at the inactive X chromosome in females or at silenced genes (Brockdorff, 2002; Trojer and Reinberg, 2007). Constitutive heterochromatin is marked by hypoacetylation and trimethylation of both H3K9 (H3K9me3) and histone 4 at lysine 20 (H4K20me3). The situation is less clear for facultative heterochromatin: for the process of local gene silencing it appears that general histone hypoacetylation, dimethylation of H3K9 (H3K9me2), monomethylation of H4K20 (H4K20me1) as well as monoubiquitination of histone H2A on lysine 119 (H2AK119ub) are prevalent, whereas general histone hypoacetylation, H3K27me2/3, H2AK119ub and H4K20me3 are characteristic for long range gene silencing of e.g. Hox gene clusters (reviewed in (Trojer and Reinberg, 2007)). Frequently, the constitutive heterochromatin at centromeres is bound by the heterochromatin binding protein 1 (HP1; (Maison and Almouzni, 2004)).

2.3.3 Chromatin Modifying Complexes and Their Regulation

Histone Methyltransferases (HMTs)

The current understanding is that at least three lysine residues on histone tails are associated with the activation of transcription when methylated (H3K4, H3K36, H3K79) and three lysine residues in the repression of transcription when methylated (H3K9, H3K27, H4K20; reviewed in (Kouzarides, 2007)). In the following paragraphs, the regulation of the most relevant lysine methylation residues for my thesis are presented in more detail (H3K4, H3K27 and H3K9).

H3K4: The Trithorax complex (Trx) contains the mixed-lineage leukemia (MLL) HMT which catalyzes methylation of H3K4 through its SET domain (Dou et al., 2006). Trx proteins are found in different complexes with diverse functions, such as in cell cycle regulation, in DNA damage signaling and in stem cell maintenance. They are clearly associated with transcriptional activation and generally have opposing roles to the

polycomb repressive complexes (see below). Beside H3K4 methylation, Trx may also mediate H3K27 acetylation and H3K36 methylation (Schuettengruber et al., 2007). There are several H3K4 HMTs, among them five different MLL enzymes (MLL1/2/3/4/5) as well as SET1A, SET1B and ASH1 (reviewed in (Kouzarides, 2007)).

H3K27: H3K27me_{2/3} is deposited by EZH2 which is part of the polycomb repressive complex 2. In mammals, the polycomb group proteins exist in two major complexes: polycomb repressive complex 1 (PRC1) and 2 (PRC2). PRC2 consists of SUZ12, EED and EZH1/2. After the deposition of H3K27me_{2/3} by PRC2, this mark is then specifically recognized by the chromodomain of Polycomb (Pc), a subunit of PRC1, which is subsequently recruited to these sites thereby preventing the access of nucleosome remodeling factors and leading to the formation of a repressive chromatin state (Cao et al., 2002; Wang et al., 2004). Additionally, EZH2 interacts with all three DNMTs, although this interaction seems to be transient, leading to increased DNA methylation and thereby the connection of two epigenetic repression systems (Vire et al., 2006). Interestingly, genes that are bound by PRC2 and positive for H3K27me₃ in ESCs become significantly more often aberrantly methylated in cancer. Developmental genes that are normally repressed during differentiation become somehow additionally and erroneously pre-marked to permanently stay repressed. Presumably, this is triggered by the additional deposition of H3K9me₂ and H3K9me₃ at these sites in embryonic cells. This specific chromatin pattern and the silencing of these developmental genes in stem cells prime these genes for DNA hypermethylation and the heritability of the silencing can provoke tumor initiation and progression (Lee et al., 2006; Mikkelsen et al., 2007; Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). This may reflect a PRC2-DNMT-mediated interruption of the cyclic DNA methylation-demethylation observed at developmental genes during differentiation (**Appendices I, II and III**). Further, PRC2 recruits the H3K4me_{2/3} demethylase (JARID1a) to target genes, further generating a repressive chromatin state and highlighting the importance of the coordinated recruitment of histone modifying enzymes (Pasini et al., 2008).

H3K9: In mammals, there are several H3K9 methyltransferases; SETDB1, G9a, SUV39h1, SUV39h2 and EHMT1 (Peters et al., 2001). H3K9 trimethylation by the SUV39 proteins is required for the DNA methylation of major satellite repeats in

pericentromeric heterochromatin by DNMT3b (Lehnertz et al., 2003). G9a interacts with DNMT1 and PCNC in a ternary complex directly at the replication fork, in order to coordinate the correct duplication of both DNA methylation and H3K9 methylation during replication (Esteve et al., 2006). Furthermore, some H3K9 methyltransferases interact with HDACs and with the heterochromatin binding protein 1 (HP1). This led to a model where deacetylation of H3K9 facilitates H3K9 methylation, this stimulates HP1 binding and recruitment of further HDACs and HMTs in order to orchestrate gene silencing (Czermin et al., 2001; Vaute et al., 2002).

Lysine Demethylases (KDMs)

Depending on its interaction partners, the lysine-specific demethylase 1 (LSD1) acts on different lysines: in association with the CoREST repressor complex, it demethylates H3K4me2/me1, leading to the repression of target genes (Shi et al., 2005). Contrarily, in association with the androgen receptor (AR) it acts as a H3K9me2/me1 demethylase which induces the expression of AR-targeted genes (Metzger et al., 2005). The H3K27me3 demethylases JMJD3 and UTX, in conjunction with the H3K4me3 HMT MLL2, are able to resolve bivalent chromatin domains during ESC differentiation. This is accompanied by the displacement of PRC2 and transcriptional activation of certain gene clusters, such as the *HoxB* locus (Agger et al., 2007).

Histone Acetyltransferases (HATs)

Generally, the specificity of HATs is rather conferred to the histone type than to specific lysine residues. HAT1 for instance prefers both lysine 5 (K5) and lysine 12 (K12) on H4. The p300/CBP-associated factor (PCAF) or lysine-acetyl transferase 2 (KAT2) acetylate K9, K14 and K18 on H3. CBP/p300 targets all canonical core histones (H2A, H2B, H3 and H4) on several lysine residues (Kouzarides, 2007). CBP/p300 is an important transcriptional co-activator and associates with a number of chromatin-modifying proteins; like the histone chaperone DAXX, HATs, HMTs but interestingly also with the DNA repair enzyme TDG (Ernst et al., 2001; Kuo et al., 2005; Tini et al., 2002).

Histone Deacetylases (HDACs)

In mammals, one can distinguish between the classical HDACs, whose enzymatic activity requires Zn^{2+} , and sirtuins, which are NAD^+ -dependent enzymes (reviewed in (Reichert et al., 2012)). The classical HDAC family contains 11 members which are further assigned to four different classes according to phylogenetic analysis and sequence homologies. For instance, class I is comprised of HDAC1/2/3 and 8 (reviewed in (Walkinshaw et al., 2008)). HDACs are ubiquitously expressed and interact with numerous tissue-specific TFs, pointing to their important role in transcriptional regulation. Besides deacetylation of different histones (H3, H4), thereby generating a repressive chromatin environment; HDACs also deacetylate non-histone proteins such as TFs which is a further means to regulate transcription (reviewed in (Reichert et al., 2012)). Furthermore they constitute core components of several complexes: HDAC1 and 2 form the catalytic core of the (i) SIN3a co-repressor complex, they are part of the (ii) nucleosome remodeling and deacetylase (NuRD) complex and (iii) the transcriptional co-repressor complex CoREST (reviewed in (Reichert et al., 2012)). HDACs lack a DNA binding domain; they thus have to interact with e.g. TFs to establish contact with their substrates. Thereby, they can be recruited to cell cycle regulator genes (such as p21 and p53) by different TFs where they directly repress transcription of these genes (reviewed in (Reichert et al., 2012)).

Non-coding RNAs (ncRNAs)

Interestingly, long non-coding RNAs (lncRNAs) have been found to constitute scaffolds for histone modifying complexes, and act both *in trans* (e.g. HOTAIR) and *in cis* (e.g. Air, Kcnq1ot1). The human lncRNA HOTAIR (HOX Antisense Intergenic RNA), which is expressed from the *HOXC* locus, bridges the PRC2 and LSD1/REST/CoREST protein complexes, to perform H3K27 methylation and H3K4 demethylation at the *HOXD* locus and select genes on other chromosomes *in trans* (Rinn et al., 2007; Tsai et al., 2010). Another lncRNA called Air was shown to regulate genomic imprinting *in cis* in the embryo proper as well as in the placenta. Air itself is imprinted and monoallelically expressed from the paternal allele, transcribed from a locus overlapping the *Igf2r* gene (Nagano et al., 2008; Sleutels et al., 2002). It mediates silencing of the imprinted, maternally expressed, genes *Igf2r*, *Slc22a2* and *Slc22a3* (Sleutels et al., 2002). The underlying mechanism is not entirely clear, but

silencing of *Igf2r* does not include G9a (Nagano et al., 2008). In contrast, the imprinting of *Slc22a2* and *Slc22a3* in the placenta requires Air-mediated targeting of G9a to these loci, leading to H3K9 methylation and allelic silencing (Nagano et al., 2008). Similarly, *Kcnq1ot1*, another lncRNA, was shown to be involved in the epigenetic silencing of imprinted genes. Target genes include *Kcnq1*, *Cdkn1c*, *Slc22a18* and *Phlda1*, which are ubiquitously imprinted, as well as *Cd81*, *Osbpl5*, *Tssc4*, *Tspan32* and *Ascl2*, which are only imprinted in placenta ((Kanduri, 2011); see also **Chapter 2.4.2**). *Kcnq1ot1* was shown to target several chromatin modifying complexes involved in epigenetic repression to these loci, among them are the G9a, PRC1 and PRC2 complexes as well as DNMT1 (Mohammad et al., 2010; Pandey et al., 2008; Terranova et al., 2008; Umlauf et al., 2004). Again, this promotes the epigenetic silencing of the target genes and strikingly, the silencing is observed in a lineage-specific manner (Pandey et al., 2008). Furthermore, miRNA and RNAi pathway components may also contribute to the methylation of H3K27 by recruitment of the PRC2 complex (Kim et al., 2008).

2.3.4 Epigenetic Modifications and Cell Plasticity

The different tissues and cell types in an organism all contain the same genetic information. Cellular identity is thus not solely determined by DNA sequence, but also by tissue-specific transcription factors that cooperate with epigenetic modifiers to establish and maintain distinct chromatin states that instruct cell-type-specific gene expression patterns and hence cell identity. Initially, the term “epigenotype” was used to describe the development and inheritance of phenotypic traits (Waddington, 2012). Already in 1950, the occurrence of 5-mC in nucleic acids was described for the first time (Wyatt, 1950), although at that time, this was not seen as an epigenetic modification. The term “epigenetic modification” became custom only in the 1990s and today, it comprises modifications of DNA bases and histone protein tails. These modifications are tightly regulated and also inherited during DNA replication in order to propagate a stable cellular identity. The underlying mechanisms are not exactly clear; except that epigenetic modifications can be reset and that they do not impact the primary sequence of the DNA molecule (reviewed in (Alabert and Groth, 2012; Cedar and Bergman, 2009; Cheedipudi et al., 2014; Jenuwein and Allis, 2001; Kouzarides, 2007)).

The correct setting, reading and restoration of epigenetic marks is especially critical during periods of high cellular plasticity, such as embryonic development and meiosis. But also for cellular metabolism or proliferation, epigenetic marks need to be strictly regulated. During embryogenesis, the developmental potential of embryonic cells becomes restricted by differentiation programs, channeling the cell fate to finally establish a terminal cell identity. Accompanying the differentiation of embryonic cells to somatic lineages, there is thus a progressive decrease of cell plasticity, which results in a restricted cell fate, where cellular memory is retained and cellular specialization is conferred (reviewed in (Shah and Allegrucci, 2013)). A core network of transcription factors (TFs), including POU domain, class 5, transcription factor 1 (OCT4/POU5f1), SRY (sex determining region Y)-box 2 (SOX2) and Nanog homeobox (NANOG), define the ground state of pluripotent embryonic stem cells (ESCs). These factors are both transcription activators and repressors: cell proliferation and self-renewal genes become activated and the expression of lineage-specific genes becomes repressed (reviewed in (Young, 2011)). Interestingly, it is sufficient to bring the key pluripotency factors OCT4, SOX2, Kruppel-like factor4 (KLF4) and myelocytomatosis oncogene (MYC; OSKM) into a somatic cell to induce epigenetic reprogramming and give rise to an induced pluripotent stem cell (iPSC; (Takahashi and Yamanaka, 2006)). It was found later, that even two factors (OCT4 and SOX2) are sufficient to induce pluripotency, however, a higher number of TFs increases the reprogramming efficiency as well as the differentiation potential of the iPSCs (Lohle et al., 2012). Recently, it was shown that the methyl-CpG binding domain protein 3 (MBD3) together with the NuRD complex is responsible for blocking the efficient induction of pluripotency genes (Rais et al., 2013). The reduction of MBD3 protein levels in the presence of OSKM allows a nearly complete reversion to naïve pluripotency, as evidenced by characteristic changes in DNA methylation, thus clearly linking epigenetic reprogramming to DNA methylation. The absence of MBD3's inhibitory effects allows interactions of OSKM with pluripotency-promoting epigenetic regulators and channels the cell fate progressively towards pluripotency (Dos Santos et al., 2014; Rais et al., 2013). Hence, epigenetic landscapes are plastic and reversible and this allows the reset of the epigenetic make-up of one generation to another. The drawback is that this also allows aberrant reprogramming as observed for instance in cancer stem cells (reviewed in (Berdasco and Esteller, 2010; Munoz et al., 2012)).

2.4 DNA Methylation

DNA methylation is found in many organisms, both in prokaryotes and eukaryotes. In prokaryotes, DNA methylation can occur both on C and A and is involved in diverse biological processes, like replication initiation, strand discrimination during MMR or defense against invading foreign DNA, i.e. bacteriophages (reviewed in (Kumar and Rao, 2013)). In eukaryotes, DNA methylation is predominantly found at the C5 position of the Cytosine (C) and is generally associated with transcriptional repression. Examples include the inactivation of the second X chromosome in female mammals, silencing of retroviral DNA and transposons as well as the parent-of-origin-specific silencing of imprinted loci (reviewed in (Basu and Zhang, 2011; Plasschaert and Bartolomei, 2014; Reiss and Mager, 2007)). Non-CpG methylation is also found in eukaryotes, although to a much lesser extent, and its role is not clear yet (Baubec and Schubeler, 2014; Patil et al., 2014). Methylation is performed by DNA methyltransferases (DNMTs) which use S-adenosyl-L-methionine (SAM) as a methyl donor (Jurkowska et al., 2011). DNMTs can be classified into maintenance and *de novo* DNMTs; the disruption of which in mice is lethal (Klose and Bird, 2006).

De novo DNA Methylation

There are two active *de novo* DNMTs (DNMT3a and DNMT3b) in mammals. A third member (DNMT3L) is present but catalytically inactive and unable to bind the co-factor SAM. DNMT3L binds DNA very poorly but stimulates the activity of both DNMT3a and DNMT3b *in vivo* and *in vitro*, probably by induction of a conformational change in the active site of the enzyme, which facilitates SAM and DNA binding and thus catalysis (Chedin et al., 2002; Gowher et al., 2005; Kareta et al., 2006).

DNMT3L can hence be considered a positive regulator of *de novo* DNMTs. The preferential substrate for DNMTa and DNMT3b is the CpG dinucleotide, however, they can also methylate Cytosine in a non-CpG context (Gowher and Jeltsch, 2001; Jurkowska et al., 2011). Yet, it is unclear, how *de novo* DNMTs are recruited to their site of action. Several mechanisms have been proposed: (i) *de novo* DNMTs recognize chromatin directly via specific domains, such as the Proline-Tryptophane-Tryptophane-Proline motif (PWWP domain) which is essential for the chromatin targeting of enzymes (Ge et al., 2004)); or (ii) *de novo* DNMTs are recruited to chromatin through protein-protein interactions with transcriptional repressors, such as

MYC (Brenner et al., 2005). Interestingly, DNMTs were shown to preferentially methylate DNA which is in a chromatin environment characterized by unmethylated H3K4 (Hashimoto et al., 2010; Zhang et al., 2010).

Maintenance DNA Methylation

The current understanding is that the maintenance DNMT (DNMT1) uses hemimethylated DNA molecules as a substrate and copies the pre-existing methylation pattern onto the new DNA strand during DNA replication (Jurkowska et al., 2011), thereby ensuring the maintenance of cell type-specific methylation states. DNMT1 recruitment to replicating regions is dependent on ubiquitin-like containing PHD and RING finger domains 1 (UHRF1; or nuclear protein 95, NP95). UHRF1 possesses a SET and RING finger-associated (SRA) domain, by which it specifically binds to methylated or hemi-methylated CpG sites (Sharif et al., 2007). DNMT1 is then positioned right at the replication fork, where it catalyzes the transfer of the methylgroup with high processivity. It is able to maintain this velocity on long stretches of DNA without dissociating from the template and without swapping the target strand (Goyal et al., 2006; Hermann et al., 2004). Interestingly, the SRA-domain of UHRF1 was recently shown to also bind to 5-hmC (Frauer et al., 2011), which implies that the contribution of UHRF1 to maintenance (or change) of DNA methylation patterns may indeed be more complex than anticipated before.

DNA Methylation and Gene Activity

DNA methylation can repress gene transcription in two ways. It can directly inhibit TF binding if a binding site requires unmethylated C (Blattler and Farnham, 2013). Also, DNA methylation triggers chromatin changes that affect the accessibility of DNA more generally, leading for instance to the recruitment of co-repressors of transcription by methyl-CpG binding domain proteins ((MBDs; (Nan et al., 1998; Watt and Molloy, 1988)). Moreover, DNA methylation can influence co-transcriptional processes, such as mRNA splicing. Intragenic DNA methylation was shown to modulate alternative splicing genome-wide. Compared to the excluded exons, alternatively spliced exons which become included in the transcript, show significantly more DNA methylation. This is mediated by the methyl CpG binding protein 2 (MeCP2), which is recruited to the methylated exons and functions primarily in

promoting their recognition for inclusion. Further, MeCP2-mediated recruitment of histone deacetylases (HDACs) facilitates the inclusion of the exon probably through modulation of acetylation states of histones or other proteins (Maunakea et al., 2013).

2.4.1 CpG Islands (CGIs)

In mammals, DNA methylation occurs mainly at the C5 position of Cytosine and in the context of CpG dinucleotides. 5-mC has a higher deamination rate than C, thus generating at a steady rate TpG dinucleotides and hence, in mammals the CpG dinucleotide is strongly underrepresented (Lander et al., 2001). The majority of CpG dinucleotides is methylated at the C5 position of the Cytosine (reviewed in (Illingworth and Bird, 2009)). Nevertheless, mammalian genomes contain clusters of CpG dinucleotides, the so called CpG islands (CGIs). The original definition of CGIs considers length (>200 bp), GC content (>50%) and the observed/expected CpG density (>0.6; (Gardiner-Garden and Frommer, 1987)). However, these regions are variable in length, with an average of approximately 1 kilo base (kb), furthermore, they are characterized by low levels of Cytosine methylation (Gardiner-Garden and Frommer, 1987; Zhao and Han, 2009). Mammalian genomes contain nearly 30'000 CGIs which frequently span gene regulatory regions (e.g. promoters), and thus seem to play a role in gene regulation (Illingworth and Bird, 2009; Lander et al., 2001). To date it is unclear, how CGIs have evolved at all. There are two plausible scenarios: (i) actively transcribed CGIs are usually bound by TFs, transcriptional co-regulators and the transcription machinery; this could sterically hinder DNMT association and thus prevent DNA methylation (Brandeis et al., 1994); (ii) there is a targeted DNA demethylation and repair mechanism, which specifically maintains CGIs in an unmethylated state (Frank et al., 1991). In both scenarios, the absence of DNA methylation might keep the C→T mutation rate low. In support of the latter, we show that TDG and TET1/2 act together to regulate DNA methylation at CGIs, relevant for lineage commitment. This indicates that cell-type-specific CGIs are kept in an unmethylated state by a precisely targeted mechanism, involving TET and TDG, which at the same time can repair deamination of Cs or 5-mCs (**Appendices I and II**). Indeed, it becomes evident that Cytosine methylation and demethylation at CGIs (as well as of the associated histone proteins) is regulated in a cell-type- and developmental-stage-specific manner (Li et al., 2014). This goes together with the

transcriptional repression caused by CGI methylation, which can prevent expression of non-tissue-specific genes and also of tumorsupressor genes during carcinogenesis (Baylin et al., 2001; Futscher et al., 2002; You and Jones, 2012). About 50% of all CGIs contain transcriptional start sites (TSSs) as evident from their overlap with promoters of annotated genes (promoter CGI; **Figure 6**). But also of the remaining 50% inter- and intragenic CGIs, many represent not yet annotated promoters and indeed produce transcripts, often in a tissue-specific manner (orphan CGIs; (Illingworth et al., 2010)). Promoter CGIs are relatively free of nucleosomes (Ramirez-Carrozzi et al., 2009; Schwarzbauer et al., 2012). Nevertheless, characteristic histone modifications can be found at CGIs; among them elevated levels of H3 and H4 acetylation and H3K4me3 as well as loss of H1 and H3K36me2 (Blackledge et al., 2010; Fan et al., 2008; Mikkelsen et al., 2007; Tazi and Bird, 1990). Also, a unique composition of proteins, such as i.e. the CXXC finger protein 1 (CFP1) which specifically binds to non-methylated CpGs as well as the H3K36-specific lysine demethylase KDM2a, can be found at CGIs (Blackledge et al., 2010; Thomson et al., 2010).

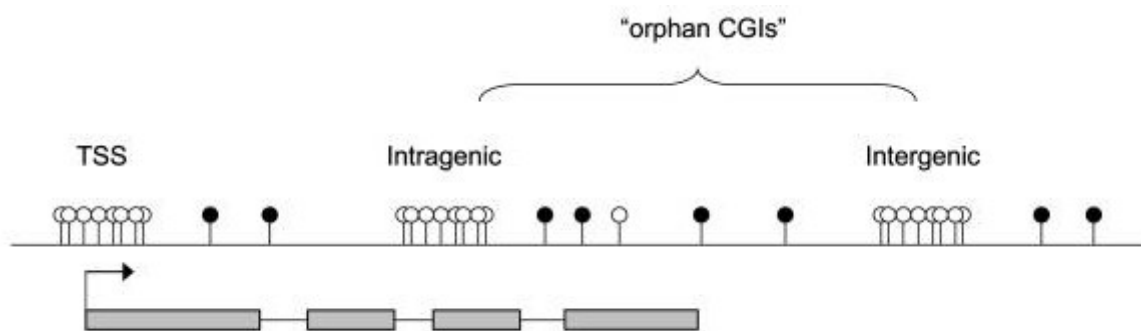


Figure 6: The promoter CpG island and both intra- and intergenic “orphan” CpG islands are characterized by a high density of CpG dinucleotides in an unmethylated state (Deaton and Bird, 2011).

2.4.2 Regulation of DNA Methylation During Development

The timing and placement of DNA methylation needs to be tightly regulated and is essential for cellular function and finally normal development. In the first place, targeting of DNMTs is achieved by the sequence context of the CpG dinucleotide. Since this is not a highly specific recognition sequence, it can be speculated that

there are additional means to target DNMTs. These could include specific histone modifications or nucleosome positioning (reviewed in (Rose and Klose, 2014)). Also, since DNMTs are SAM-dependent enzymes, SAM metabolism could play a role as well. Furthermore, DNMT3L was shown to affect overall DNA methylation levels, by selectively modulating DNMT3a activity at imprinted loci in germ cells (Aapola et al., 2000; Kaneda et al., 2004).

The successful and healthy development of organisms is highly dependent on the accurate timing of DNA methylation as well as DNA demethylation (see also **Chapter 2.5**). During embryonic development, when cells undergo lineage commitment, there are specific gene programs that are shut down by DNA methylation or activated by DNA demethylation. For instance, pluripotency genes become silenced very early upon differentiation, already within 24 hours (Rigbolt et al., 2011). Contrarily, developmental genes will rapidly become activated upon specific developmental clues (reviewed in (Kraushaar and Zhao, 2013)).

An additional aspect of DNA methylation in development is genomic imprinting. For the majority of genes, the maternal and the paternal allele are expressed or repressed likewise. A number of genes, however, is expressed monoallelically, in a parent-of-origin-specific manner (reviewed in (Plasschaert and Bartolomei, 2014)). These are designated imprinted genes and they include probably up to 1000 genes in humans (Kelsey and Bartolomei, 2012). Usually, imprinted genes are located in clusters of 3-12 genes and the clusters are characterized by allele-specific epigenetic marks, including DNA methylation and histone modifications. Imprinting is well conserved in mammals; however, it is not exactly known how the specific recognition of these genes is achieved. Nevertheless, it is clear that in both germlines, DNA methylation is established by DNMT3a and DNMT3L, albeit there may be several mechanisms involved. An example for paternal-specific methylation is the *H19/Igf2* locus. DNA methylation is deposited there already prior to the onset of meiosis, in the germline which generates the gametes. Contrarily, maternal-specific imprinting occurs postnatally in growing oocytes. Interestingly, imprinted genes include many with functions in placental growth, and is therefore assumed to have evolved from the fundamental maternal-fetal conflict on nutrient allocation (Fowden and Moore, 2012; Girardot et al., 2013; Plasschaert and Bartolomei, 2014).

2.4.3 The Role of DNA Methylation in Disease

DNA methylation is playing an important role in a variety of diseases. Of these, some types of cancer represent archetypes for misregulation of DNA methylation: They often exhibit hypomethylation at repetitive sequences, which may cause chromosomal instability (Cheung et al., 2009); whereas important genes are inactivated by aberrant hypermethylation of their CpG island promoters. This is in line with an aberrant silencing of tumor suppressor genes present in numerous cancers (Choi and Lee, 2013; Jones, 2002; Jones and Baylin, 2002). Since recently the TET proteins were identified as dioxygenases of 5-mC, aberrant DNA methylation in cancer was reconsidered. The question arose, whether malfunctioning *Tet* genes could be responsible for cancer-associated DNA methylation changes, including DNA hydroxymethylation. Indeed, in many different human cancer types, the levels of 5-hmC were found to be significantly reduced (Haffner et al., 2011; Jin et al., 2011; Kudo et al., 2012; Lian et al., 2012; Yang et al., 2013). Corroborating this finding, TET proteins were reported to be mutated or dysfunctional due to chromosomal translocations in many hematopoietic malignancies (Solary et al., 2014). In our lab, we recently found that the *Tet1* promoter is hypermethylated in a subtype of colorectal cancers (CRCs) characterized by the simultaneous hypermethylation of CpG islands (CpG Island Hypermethylator Phenotype, CIMP; Stefan Weis, unpublished data). The comparison between CIMP positive and CIMP negative CRC cell lines revealed elevated 5-hmC levels in CIMP positive cell lines (Annika Wirz, unpublished data). Similarly, mutations in *Idh1* and *Idh2* were reported to be sufficient to generate a CIMP phenotype in gliomas, possibly by depriving the TET proteins of its co-factor 2-oxoglutarate (2-OG; (Turcan et al., 2012)).

Trinucleotide repeat disorders are characterized by well-defined, unstable trinucleotide regions located within different genes, which undergo expansion for unknown reasons. The expanded state is then frequently linked to neurodegenerative disorders (reviewed in (McMurray, 2010)). Examples herefore are the myotonic dystrophy (DM1) or the fragile X syndrome (FXS) – both disorders, in which DNA methylation plays an important role (Evans-Galea et al., 2013). In FXS, the DNA methylation following and presumably triggered by repeat expansion, is responsible for the manifestation of the syndrome as evident from the rare cases showing an expanded but unmethylated repeat and an absence of disease (Loesch et al., 2012).

The silencing of the *Fmr1* gene, which was initiated by repeat expansion and followed by DNA methylation thus causes the loss of the gene product (FMRP1), which is responsible for the development of FXS.

The misregulation of imprinting during development can lead to a number of disorders, including the Beckwith-Wiedemann (BW) and the Silver-Russell (SR) growth syndromes. In SR patients, the imprinting control region (ICR) of the *Igf2/H19* locus, which is the element regulating the imprinting of the whole cluster, is characterized by hypomethylation of the paternal allele in SR individuals. This is associated with a loss of *Igf2* expression, thereby leading to the manifestation of the syndrome, which is characterized by intrauterine growth restriction, postnatal growth retardation, facial dysmorphism and body asymmetry (reviewed in (Girardot et al., 2013)). Contrarily, the BW syndrome is characterized by fetal and postnatal overgrowth as well as an increased risk of childhood cancers. A variety of underlying causes has been described to cause BW syndrome, among them is the defective imprinting observed at the *Kcnq1* imprinted domain, which leads to a strong reduction in *Cdkn1c* expression. CDKN1c is a negative regulator of the cell cycle and probably also a tumor suppressor gene (reviewed in (Girardot et al., 2013)).

2.5 DNA Demethylation

There are certain biological conditions (described in more detail below), where active DNA demethylation seems to occur. While the biochemistry of the methylation of Cytosine is relatively well-studied, knowledge about the reverse process, the DNA demethylation, has been lagging behind. At a first thought, direct removal of the methyl group from cytosine would represent an elegant and simple way to restore an unmethylated C. Plants like *Arabidopsis thaliana* possess the DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1) enzymes, two DNA glycosylases able to excise 5-mC, thus allowing restoration of an unmethylated C by BER (Ikeda and Kinoshita, 2009). Although there have been reports about the direct excision of 5-mC in mammals as well (Bhattacharya et al., 1999; Fremont et al., 1997; Jost et al., 1995; Zhu et al., 2000), the situation seems to be less clear here. Over the years, these activities could not be reproduced by others, thus potential DNA demethylation mechanisms remained under strong debate up to date. Recent evidence points to

mammalian DNA demethylation being a multi-step process, which might vary depending on, for instance, developmental stage and genomic location. The possible pathways that remove 5-mC are discussed below.

2.5.1 Mechanisms of DNA Demethylation

Passive DNA Demethylation

In the absence of maintenance methylation, i.e. DNMT1 / UHRF1, and upon successive rounds of DNA replication, 5-mC will be lost due to passive dilution. This is the scenario envisioned for the global erasure of 5-mC in the maternal pronucleus in mouse preimplantation embryos (Monk et al., 1987; Wu and Zhang, 2014). Along these lines, recruitment of DNMT1 to replication foci is strongly reduced upon the loss of UHRF1 in PGCs (Ohno et al., 2013). A further possibility of “passive” DNA demethylation includes the action of the TET proteins: Since DNMT1 is not able to recognize 5-hmC, oxidation of 5-mC by the TET enzymes will render these sites “invisible” for DNMT1. This will eventually lead to dilution of 5-hmC and thus demethylation of the DNA after several cycles of DNA replication (Inoue et al., 2011; Inoue and Zhang, 2011; Otani et al., 2013).

Suggested Pathways for Active DNA Demethylation

Since plants have ways to directly excise 5-mC by DNA glycosylases, an analogous mechanism was initially anticipated to exist in mammals. However, the reports about such a direct activity remained scarce: MBD2 was reported to directly remove the methyl-group from the 5-mC and restore the unmodified C (Bhattacharya et al., 1999) and a chicken homolog of TDG was reported to act on 5-mC together with an RNA (Fremont et al., 1997; Jost et al., 1995; Zhu et al., 2000). Furthermore, several notions implicated a deaminase activity, such as AID (activation-induced cytidine deaminase) or other members of the APOBEC family (apolipoprotein B mRNA editing enzyme, catalytic polypeptide), that would deaminate 5-mC or 5-hmC to T or 5-hmU, respectively (Guo et al., 2011; Morgan et al., 2004). The thereby generated mismatches would be suitable substrates for the TDG, MBD4 or SMUG1 DNA glycosylases (reviewed in (Jacobs and Schar, 2012) and subsequent BER would restore the original unmethylated C. However, there are some caveats to a prevalent deamination-based DNA demethylation pathway: First, the generation of

premutagenic lesions (G:T or G:hmU mismatches) would inevitably increase the risk of mutation. Yet, we could not find an elevated C→T mutation rate in a *Tdg*-knockout background, supporting the notion that a deamination-based pathway is rather unlikely to occur on a regular basis (**Appendix I**). Second, hmC was shown to be an unfavorable substrate for deamination by AID or APOBECs, due to steric constraints of the catalytic pocket (Nabel et al., 2012; Rangam et al., 2012).

Fascinatingly, it turned out, that the predominant DNA demethylation pathway works through the iterative oxidation of 5-mC by the Ten-Eleven Translocation family of proteins (TET1-3; (He et al., 2011; Hu et al., 2014; Maiti and Drohat, 2011; Raiber et al., 2012; Shen et al., 2013)). All three TET proteins were shown to have the ability to catalyze these oxidation reactions in a Fe²⁺- and 2-OG-dependent manner applying a base-flipping mechanisms (**Chapter 2.5.2**; (Hu et al., 2013; Ito et al., 2010; Ito et al., 2011; Tahiliani et al., 2009)). Another milestone result was the finding that TDG very efficiently excises 5-fC and 5-caC (He et al., 2011; Maiti et al., 2013). Apart from TDG, no other DNA glycosylase has been shown to process these substrates. Direct removal of the carboxyl-group has been discussed as well, but so far, no direct decarboxylase activity could be identified (**Figure 7**).

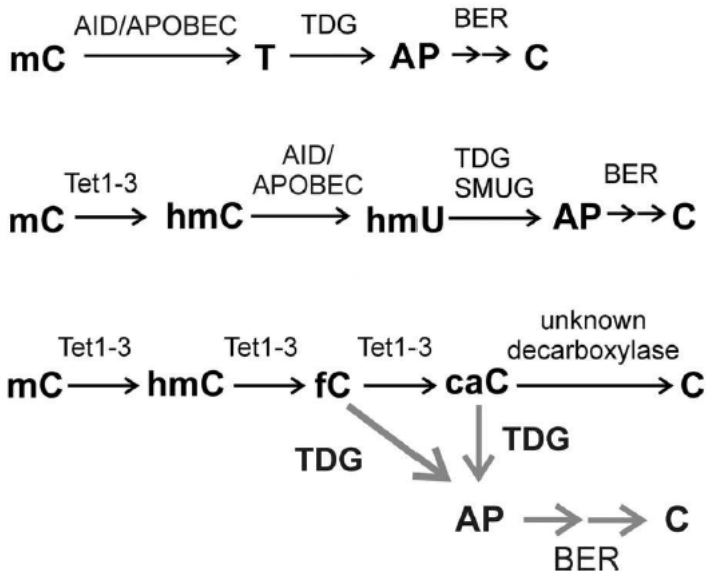


Figure 7: Implicated pathways for DNA demethylation include a deamination-repair based model [top], an oxidation-deamination-repair based model [middle] and an iterative oxidation-repair based model [bottom] (Maiti and Drohat, 2011).

There has been an intriguing report lately, that the *de novo* DNA methyltransferases DNMT3a and DNMT3b are able to dehydroxymethylate 5-hmC thereby restoring the unmethylated state. However, this is dependent on the redox-state of the protein: in an oxidized form, the dehydroxymethylase activity is enhanced and in a reduced form, the DNA methyltransferase activity is enhanced (Chen et al., 2012). Unfortunately, this has not been reproduced by others so far.

Global and Targeted DNA Demethylation

A global erasure of nearly all DNA methylation occurs twice in the development of mammals. The first event takes place in the paternal pronucleus in preimplantation embryos: Within 6-8 hours of fertilization and before the onset of DNA replication, the whole paternal genome is demethylated (Inoue and Zhang, 2011; Iqbal et al., 2011; Mayer et al., 2000; Oswald et al., 2000). This is achieved by TET3-mediated global oxidation of 5-mC to 5-hmC, which is subsequently diluted upon DNA replication (Inoue and Zhang, 2011). It is likely that different pathways cooperate in the erasure of both global and locus-specific DNA methylation, i.e. oxidation by the TET proteins followed by replication-dependent dilution or oxidation by the TET proteins followed by repair-mediated base excision (reviewed in (Wu and Zhang, 2014)). Interestingly, DNA in the maternal pronucleus as well as a few paternally imprinted loci are protected from TET3-mediated oxidation. These loci are characterized by the heterochromatic histone mark H3K9me2 to which the maternal factor DPPA3 (Stella / PGC7) exhibits strong affinity. This leads to the exclusion of TET3 binding and thereby prevention of DNA demethylation (Nakamura et al., 2012). The second global DNA demethylation event occurs during the development and migration of primordial germ cells (PGCs) and can be mechanistically divided into two distinct phases. The first phase occurs before embryonic day E9.5 and involves global depletion of 5-mC with the retention of locus-specific methylation at imprinted genes and repetitive elements (Hackett et al., 2013; Vincent et al., 2013). Although TET-mediated global DNA demethylation through the generation of 5-hmC can be excluded for the first phase (Vincent et al., 2013), the exact mechanism of methylation erasure is still not clear. Contrarily, the second phase of DNA demethylation in PGCs is initiated around E13.5 and involves DNA demethylation at maternally imprinted loci (Seisenberger et al., 2012). In contrast to the first phase, the second phase is initiated by targeted

TET1- and TET2-mediated oxidation (Vincent et al., 2013). DNA demethylation appears to be unidirectional in both phases, since *de novo* methylation is not observed (Seisenberger et al., 2012). However, to date, the mechanism that re-establishes DNA methylation patterns in late PGCs is not clear. It involves DNMT3a, DNMT3b and DNMT3L, but seems to vary between distinct genomic locations, such as imprinted regions or transposable elements. Interestingly, the result of global DNA methylation re-establishment differs between female and male germ cells: sperm cells are heavily methylated, whereas oocytes are moderately methylated (reviewed in (Seisenberger et al., 2013)).

The mechanism of targeted erasure of DNA methylation is more difficult to decipher. To this end, work from different laboratories including ours, suggests a TET-TDG-mediated oxidation-excision repair pathway at gene regulatory elements, such as CGIs and enhancers of developmentally important genes ((Raiber et al., 2012; Shen et al., 2013; Song et al., 2013) and **Appendices I and II**). It is neither established, how the TET proteins or TDG are recruited to these specific sites, nor whether they are recruited in a sequential order or together in a protein complex. If and how the recruitment to DNA and/or putative protein interactions affect or even regulate the catalytic activity of the TET proteins or TDG is not known. There are several possible scenarios which could explain the precise recruitment of the TETs and TDG. The first possibility includes a tissue-specific or developmental-stage-specific TF, such as EBF1 or NANOG for targeting the TET proteins or RAR for targeting TDG (Costa et al., 2013; Guilhamon et al., 2013; Um et al., 1998). A second possibility for the targeting of TET and/or TDG can be envisioned by the recruitment through RNA Polymerase II to sites of active or stalled transcription. It was shown, for instance, that 5-fC is mainly enriched in promoter CGIs and exons and that high levels of 5-fC coincided with active transcription and hence presence of RNA Pol II (Raiber et al., 2012). Additionally, both TETs and TDG have been shown to interact with several TFs, further strengthening the link to the transcription machinery (Chen et al., 2003; Tini et al., 2002; Williams et al., 2011).

2.5.2 Ten-Eleven-Translocation (TET) Protein Family

The recent discovery of the catalytic activity of the TET proteins evoked a lot of interest. TET proteins are named after the ten-eleven chromosomal translocation, fusing the *Mll1* gene on chromosome 10 to the *Tet1* gene on chromosome 11, which is found in certain forms of leukemia (Lorsbach et al., 2003). It turned out later that the TET proteins are key enzymes in DNA methylation control and have roles in embryonic development as well as regulation of gene transcription (reviewed in (Pastor et al., 2013)). TET proteins represent a subfamily of Base J-binding proteins (JBP), and belong to a large family of Fe²⁺- and 2-oxoglutarate-dependent (2-OG) dioxygenases (Iyer et al., 2008; Iyer et al., 2009). Based on the analysis of JBPs in *Trypanosoma brucei*, the TET proteins were predicted to have DNA modifying activity (Iyer et al., 2008). JBPs generate base J, which is a modified thymine. This is achieved by the oxidation of the 5-methyl group of T and the subsequent glucosylation of the resulting hydroxyl group by an unknown glucosyltransferase (Gommers-Ampt et al., 1993; Yu et al., 2007).

To date, three different TET proteins (TET1-3) have been described and all of them were shown to oxidize 5-mC to 5-hmC. Moreover, it was found that all three TET proteins are able to iteratively oxidize 5-hmC to 5-fC and 5-caC (He et al., 2011; Ito et al., 2011; Tahiliani et al., 2009). Metazoan TET proteins contain an N-terminal CXXC domain, enabling them to recognize and bind CpG dinucleotides in DNA; as well as a C-terminal catalytic domain, where the oxidation reaction occurs (**Figure 8**). The core of the catalytic domain is folded into a double-stranded β -helix (DSBH) which harbors a Cys-rich region that stabilizes the DNA above the catalytic core (Hu et al., 2013; Iyer et al., 2009; Tahiliani et al., 2009). In jawed vertebrates, the *Tet* gene underwent triplication giving rise to *Tet1*, *Tet2* and *Tet3*. A subsequent chromosomal inversion event at the *Tet2* locus resulted in the separation of its CXXC domain, which became a separate gene called *Idax* (Iyer et al., 2009; Ko et al., 2013). Interestingly, TET2 does physically and functionally interact with IDAX (Ko et al., 2013).

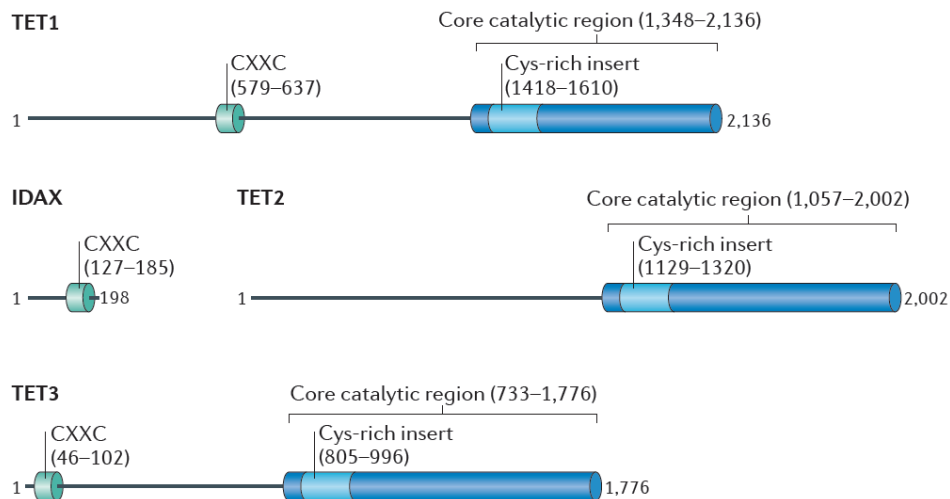


Figure 8: Gene maps of the *Tet* genes, depicting the CXXC domain (which is located in *Idax* for *Tet2*) and the core catalytic regions (Pastor et al., 2013).

The catalytic mechanism of the TET proteins was revealed recently and the applied base-flipping mechanism is reminiscent of both DNMTs as well as DNA glycosylases (Hu et al., 2013; Lloyd and Cheng, 1997). The crystal structure of TET2 bound to methylated DNA revealed that 5-mC is flipped into the catalytic cavity of the enzyme (Hu et al., 2013). This allows close contact to the catalytic Fe^{2+} iron, which is coordinated by a conserved Histidine residue in the C-terminal of the DSBH fold (Loenarz and Schofield, 2008, 2011). In addition to the aforementioned co-factors iron and 2-OG, ascorbic acid (vitamin C) was recently demonstrated to also act as a co-factor for TET proteins (Yin et al., 2013). Vitamin C enhances TET activity which leads to a global and fast increase in 5-hmC as well as 5-fC and 5-caC, resulting in a net loss of DNA methylation at promoters and thus a subsequent increase of gene expression (Blaschke et al., 2013; Minor et al., 2013; Yin et al., 2013).

Interestingly, the catalytic activity of the TET proteins can be inhibited by mutations in the *Idh1* or *Idh2* genes, which encode isocitrate dehydrogenases that produce 2-OG. These mutations lead to an accumulation of 2-hydroxyglutarate (2-HG) at the expense of 2-OG, thus depriving TET of its co-factor (Xu et al., 2011a). *Idh1* and *Idh2* mutations are often found in gliomas and acute myeloid leukemias, and *Idh1* mutations may account for a CpG island methylator phenotype in gliomas (G-CIMP), possibly by inhibition of TET activity (Turcan et al., 2012).

Expression of TET proteins

Although TET1 and TET2 are highly expressed in ESCs and play important roles in regulating mouse development, ESC survival and pluripotency, their expression is dispensable for development (reviewed in (Pastor et al., 2013)). With exception of distinct hematological abnormalities and enhanced self-renewal of hematopoietic stem cells in the case of *Tet2*, the single gene knockouts of *Tet1* and *Tet2* have surprisingly little effect on the health of mice (Dawlaty et al., 2011; Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011). Double *Tet1-Tet2* knockout animals, however, show a lethal phenotype with incomplete penetrance: half of the animals die perinatally, with visible defects in head development. The remaining animals survive and have fully developed organs but exhibit lower body weight and reduced fertility (Dawlaty et al., 2013; Pastor et al., 2013). In case of *Tet3*-knockout, about half of the embryos stop growing and die around E11.5 and the remaining ones die perinatally for unknown reasons (Gu et al., 2011).

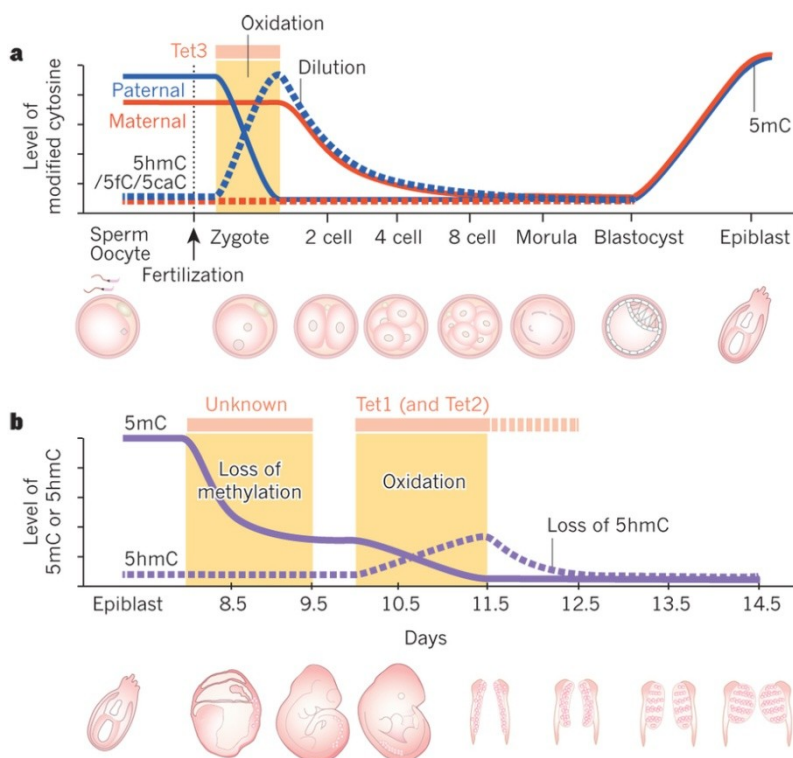


Figure 9: Expression profiles of TET3 during early events after fertilization (a) and expression profiles of TET1 and TET2 in primordial germ cells (PGCs; b), correlated in both panels with 5-mC and 5-hmC profiles (Kohli and Zhang, 2013).

TET3 is mainly expressed in the zygote where it oxidizes 5-mC in the male pronucleus. The concomitant increase in 5-hmC is not observed in TET3-deficient oocytes (Gu et al., 2011; Wossidlo et al., 2011). Contrarily, *Tet1* and *Tet2* genes are mainly expressed in the blastula stage of the embryo as well as in embryonic stem cells, but become quickly downregulated during differentiation ((Tahiliani et al., 2009) and reviewed in (Pastor et al., 2013); **Figure 9**).

Genome-wide binding patterns of the TET proteins and their products

TET1 binding sites are mainly found within gene bodies (exons and introns), with the highest density at the TSS (Williams et al., 2011). Similarly, TET2 binding sites show the highest density around the TSS (Chen et al., 2013b). These are genomic locations with high CpG density and interestingly, besides the TET proteins, also IDAX is strongly enriched at CpG-rich regions, often coinciding with promoter regions in close proximity to the TSS (Ko et al., 2013; Xu et al., 2011c). The oxidation products of TET1 and TET2 are also mostly found in regions of high CpG density, often located at gene regulatory elements (Booth et al., 2012; Shen et al., 2013). For instance, a strong enrichment of 5-hmC was found at enhancer regions and the levels of 5-hmC increased particularly at those enhancers, which become activated during differentiation of ESCs (Serandour et al., 2012; Stadler et al., 2011; Stroud et al., 2011; Szulwach et al., 2011; Yu et al., 2012).

Interaction partners of the TET proteins

All three TET proteins were shown to interact with the O-linked N-acetylglucosamine transferase (OGT) and to enhance and stabilize the association of OGT to chromatin (Chen et al., 2013b; Deplus et al., 2013; Ito et al., 2014; Shi et al., 2013; Vella et al., 2013). The binding of TET2/3 and OGT to chromatin was found to enhance the activity of the H3K4 methyltransferase SETD1a, thereby helping to activate gene expression (Deplus et al., 2013). Furthermore, TET1 and TET2 were shown to interact physically and functionally with NANOG to facilitate reprogramming of somatic cells to iPSCs. The reprogramming synergy between NANOG and the TET proteins was dependent on the catalytic activity of the latter (Costa et al., 2013). Additionally, the early B cell factor 1 (EBF1) was found to physically interact with TET2. In cell lines with impaired TET function due to increased levels of 2-HG, the

interaction with the DNA-binding protein EBF1 was suggested to serve the tissue-specific targeting of DNA demethylation (Guilhamon et al., 2013). Recently, TET1 was shown to form a protein complex with SUZ12, a member of the Polycomb Repressive Complex 2 (PRC2), in ESCs. It is believed that PRC2 tethers TET1 to bivalent promoters in order to generate 5-hmC there, which provides the epigenetic plasticity necessary for normal differentiation (Neri et al., 2013). This is further supported by a reported interaction of TET1 with Enhancer of zeste homolog 2 (EZH2), another component of PRC2 (Cartron et al., 2013). This report included the identification of additional interaction partners for TET1, belonging to the family of chromatin modifying enzymes, such as the histone deacetylases 1, 6 and 7 (HDAC1/6/7) and lysine-specific demethylase 1a (LSD1). In addition, they found an interaction of TET1 with PCNA, the sliding clamp crucial in DNA replication and other DNA metabolic processes; as well as with the methyl-CpG binding protein 2 (MeCP2; (Cartron et al., 2013)). Moreover, TET1 interacts with SIN3a, a transcriptional co-repressor complex (Cartron et al., 2013; Williams et al., 2011). The PR domain-containing transcriptional regulator (PRDM14) physically interacts with TET1 and TET2 and enhances the targeting of the TET proteins to chromatin, thereby promoting DNA demethylation of PRDM14 target genes (Okashita et al., 2014). The examination of TET interaction partners indicates that the TET proteins are involved in a variety of biological processes. These include the regulation of transcription, maintenance of epigenetic plasticity, as well as the coordination of histone modifications and DNA demethylation.

2.6 Linking DNA Repair to Epigenetics

It is evident that DNA repair processes play a crucial role in prevention from mutations and thereby the preservation of the primary genetic information, which is important to ensure cellular identity and viability. However, many DNA repair proteins have also been linked to epigenetic mechanisms, either by processing epigenetic modifications in DNA and/or by physically or functionally interacting with chromatin modifying enzymes. As an example for the former, the DNA glycosylase TDG has been shown to process epigenetic DNA modifications, namely oxidative DNA demethylation intermediates: TDG excises 5-fC and 5-caC and by the recruitment of downstream BER proteins is able to restore unmethylated C (He et al., 2011; Maiti

and Drohat, 2011). *Vice versa*, it is assumed that chromatin remodeling is necessary to allow access for DNA repair proteins, thus several chromatin modifying enzymes were shown to associate with and participate in various DNA repair events (reviewed in (Verger and Crossley, 2004)). Some of the interconnections between DNA repair enzymes and chromatin modifying enzymes are discussed below.

The histone chaperone CAF1 was shown to facilitate chromatin assembly after NER-mediated DNA repair (Gaillard et al., 1996). The recruitment of CAF1 is restricted to damaged sites and depends on NER and may involve PCNA (Green and Almouzni, 2003; Moggs et al., 2000), suggesting that CAF1 is involved in local chromatin rearrangements following DNA damage and repair (Verger and Crossley, 2004). Furthermore, the TC-NER-specific repair protein Cockayne Syndrome B protein (CSB) was the first DNA repair enzyme shown to possess nucleosome remodeling activity (Citterio et al., 2000; Scharer, 2013). CSB is a DNA-dependent ATPase and belongs to the SWI/SNF family of chromatin remodelers. Thus, CSB not only interacts with ds DNA but also with core histones and it is assumed that its nucleosome remodeling is important to generate space for the numerous other NER proteins (Citterio et al., 2000; Thoma, 1999; Verger and Crossley, 2004). Moreover, the HAT TIP60 was shown to be involved in DSB repair. The absence of TIP60 HAT activity appears to deprive the cell of its ability to signal the occurrence of DNA damage to the apoptotic machinery, resulting in the loss of apoptotic competence (Ikura et al., 2000). The yeast homologue of TIP60 was found to acetylate H4, a modification which was demonstrated to be essential for DSB repair (Bird et al., 2002). These results indicate that TIP60 participates in the signal transduction between DNA damage recognition and recruitment of DNA repair proteins.

Another example is the interaction between the HATs CBP/p300, which are transcriptional co-activators, with several DNA repair proteins, among them TDG and XPE which initiate BER and NER, respectively (Datta et al., 2001; Tini et al., 2002). TDG was shown to recruit CBP/p300 to DNA and, importantly, the resulting ternary complex consisting of CBP-TDG-DNA was competent for both the excision step of DNA repair as well as histone acetylation. TDG stimulated CBP's transcriptional activity and was itself a substrate for CBP/p300-mediated acetylation (Tini et al., 2002). The acetylation of TDG triggered the release of CBP from DNA and, furthermore, it prevented the recruitment of APE1, the downstream repair enzyme

(Tini et al., 2002). These results implicate CBP/p300 as a possible chromatin remodeling co-factor in TDG-mediated DNA repair.

As a last example, an effect of EZH2, the catalytic subunit of PRC2, on DNA damage signaling has been reported: EZH2-mediated gene silencing in cancer cells was found to determine the cellular response to DNA damage. The inhibition of EZH2 activity abrogated both G1 and G2/M cell cycle checkpoints and promoted apoptosis after treatment with DNA damaging agents in cancer, but not in healthy cells (Wu et al., 2011). In addition, it was shown that EZH2 controls the intracellular localization of Breast Cancer 1 (BRCA1), a protein involved in the repair of DNA DSBs. Upon upregulation of EZH2, BRCA1 is exported from the nucleus, causing aberrant mitosis and thus genomic instability; presumably due to the absence of BRCA1-dependent DSB repair (Gonzalez et al., 2011).

Due to the manifold connections between chromatin modifying enzymes and DNA repair factors, it is tempting to speculate about a certain epigenetic “code” for coordination of DNA repair. One example is the well-studied, rapid phosphorylation of the histone variant H2A.X (γ H2A.X), particularly at sites of DNA DSBs. γ H2A.X is itself a product of early DNA damage response and stabilizes DNA damage signaling, facilitating recruitment of DNA repair proteins (Svetlova et al., 2010; van Attikum and Gasser, 2005b). Furthermore, it can be envisioned that PTMs on both DNA repair as well as histone proteins could regulate repair activity. Examples include the acetylation of FEN1 by p300, which is significantly increased after DNA damage induction by UV, limiting its nuclease activity (Hasan et al., 2001) as well as the increased ADP ribosylation of histone proteins after DNA damage (Bouchard et al., 2003; Verger and Crossley, 2004).

A consistent functional endpoint of the co-operation between DNA repair proteins and chromatin modifiers appears to be the generation of a chromatin environment, facilitating DNA-templated processes like repair or transcription. In general, chromatin is still considered as a barrier for the repair of DNA damage (Lukas et al., 2011).

Genetics and Epigenetics in Cancer

The connection between genetics and epigenetics is exemplified in different cancers, where the interplay between the two is often misregulated. On one hand, changes in epigenetic mechanisms can lead to genetic mutations and on the other hand, genetic mutations in epigenetic regulators can have various effects on the epigenome (reviewed in (You and Jones, 2012)). For instance, many cancers were found to harbor distinct mutations in EZH2, the catalytic subunit of the PRC2 complex; if the mutation is inactivating, it is associated with poor prognosis (Chase and Cross, 2011). There were two main mechanisms described, which can lead to the loss of a functional gene product and are often observed during tumorigenesis: (i) mutations in genes controlling the cell cycle progression or DNA repair and (ii) promoter hypermethylation of e.g. tumor-suppressor or DNA repair genes (reviewed in (You and Jones, 2012)). An example for the latter is hypermethylation of the *Mlh1* promoter (Menigatti et al., 2001). MLH1 is a DNA repair protein involved in MMR, and its epigenetic silencing leads to defective MMR. This in turn negatively impacts genomic stability due to absent repair of base-base mismatches and insertion/deletion mispairs that are generated during DNA replication (Sameer et al., 2014). Such defects in MMR cause genetic hypermutability and instability mainly at short repeat sequences (micro-satellite instability; (MSI)), which is a hallmark of many tumors (Esteller et al., 1998). Bioinformatical clustering of the genome-wide occurrence of DNA methylation reveals certain recurrent patterns in tumors that allow the identification of tumor subgroups, like for instance the CpG Island Methylator Phenotype (CIMP) in CRCs or gliomas. This subtype is characterized by both specific epigenetic and genetic traits, namely simultaneous DNA hypermethylation of CGIs and genetic mutation of the BRAF gene, respectively (Hinoue et al., 2009; Noushmehr et al., 2010; Toyota et al., 1999; Xu et al., 2011b).

2.7 Transcription by RNA Polymerase II

A major interest during my PhD thesis was the investigation of how TDG affects transcription, both directly and indirectly (**Supplementary Results 4.4 and Appendix II**). In order to provide the essential background, the relevant processes during transcription are described here in detail.

DNA-dependent RNA polymerases (RNA Pol) transcribe the genetic information in the DNA into a corresponding RNA, the transcript. In mammalian nuclei, transcription is done by three distinct RNA Polymerases, RNA Pol I-III, each of them being responsible for the synthesis of a discrete subset of RNAs. While RNA Pol I and III are responsible for the synthesis of ribosomal RNA and tRNAs, respectively, it is RNA Pol II which transcribes protein encoding genes as well as small non-coding RNAs (e.g. small nuclear RNAs, small nucleolar RNAs, microRNAs) and lncRNAs (Jacobs et al., 2004). RNA Pol II consists of 12 subunits (Bushnell and Kornberg, 2003), whereof the largest subunit (Rbp1) developed a unique, highly conserved carboxy-terminal domain (CTD; (Barron-Casella and Corden, 1992)). In mammals, this CTD is composed of 52 tandem heptapeptides with the evolutionary conserved consensus motif YSPTSPS (Allison et al., 1985; Corden et al., 1985; Liu et al., 2010). The CTD can interact with many different proteins at different stages of the transcription cycle. This is achieved by the dynamic structural plasticity of the CTD together with the variety of specific binding surfaces generated by extensive PTMs on the heptapeptide repeats ((Komarnitsky et al., 2000) and reviewed in (Phatnani and Greenleaf, 2006)). For instance, the phosphorylation pattern on each heptapeptide repeat was shown to be involved in the recruitment of more than 100 different proteins with diverse roles (Phatnani et al., 2004). The manifold possibilities for PTMs on the heptapeptide repeats of the CTD – all seven amino acids having the potential of being modified – provoked the creation of the term “CTD code”. A great number of combinatorial variations can be achieved by exclusive or simultaneous phosphorylation of tyrosine, threonine and serine and/or glycosylation of threonine and serine and/or isomerization of proline (Kelly et al., 1993).

2.7.1 Transcription Initiation

Initiation of transcription by RNA Pol II is a highly regulated process. Regulation is accomplished by gene-specific activators and repressors as well as PTMs of the RNA Pol II itself (reviewed in (Grunberg and Hahn, 2013)). General transcription factors (GTFs) function in gene promoter as well as TSS recognition and RNA Pol II recruitment. Additionally, they interact with regulatory factors, like TFs or co-regulators of transcription (reviewed in (Grunberg and Hahn, 2013)).

The assembly of the preinitiation complex (PIC) is the earliest event during gene expression. The PIC contains the RNA Polymerase II, the TATA binding protein (TBP), the GTFs (TFIIA, TFIIB, TFIID, TFIIIE, TFIIF and TFIIH) and, in most cases, the Mediator complex. The latter can be considered as a general co-activator for RNA Pol II-mediated transcription, although it is not *per se* needed for transcription (reviewed in (Poss et al., 2013)). Interactions between DNA and the GTFs anchor the RNA Pol II to double-stranded promoter DNA (**Figure 10**). In a next step, TFIIH unwinds approximately 10 bp of promoter DNA in an ATP-dependent manner, generating the open complex state. The correct assembly of the PIC is thus crucial to position the RNA Pol II at the TSS and to prepare the DNA for subsequent transcription (reviewed in (Grunberg and Hahn, 2013)). TFIIH is the most complex GTF, comprised of ten subunits, whereof three contain ATP-dependent enzymatic activities, one of these being CDK7, a CTD-Serine 5 kinase (reviewed in (Grunberg and Hahn, 2013)). In addition to transcription initiation, TFIIH plays important roles in both GG- and TC-NER, where it unwinds DNA, thus allowing excision of the damage-containing single strand (Compe and Egly, 2012). The Mediator complex is even more complex and consists of 25-30 protein subunits. These are grouped into head, middle and tail domains as well as the CDK8 kinase module, which was shown to phosphorylate serine 5 of the CTD (Davis et al., 2002; Ramanathan et al., 2001). Different Mediator subunits make contact with various TFs, including activators, co-activators, GTFs and subunits of RNA Pol II (Ansari and Morse, 2013). For instance, Mediator was shown to functionally interact with CBP/p300, thereby enhancing ER α -dependent transcription (Acevedo and Kraus, 2003). H4 tails are also targeted by interactions with Mediator subunits (Liu and Myers, 2012). Interestingly, Mediator was recently shown to interact with Cohesin, which forms ring structures capable to embrace two double-stranded DNA segments, apparently to physically and functionally connect enhancers and core promoters of active genes in ESCs. This results in cell-type-specific DNA looping, ensuring the initiation of cell-type-specific gene expression (Kagey et al., 2010). Furthermore, lncRNAs can associate with Mediator and thereby modulate chromatin architecture and transcription (Lai et al., 2013).

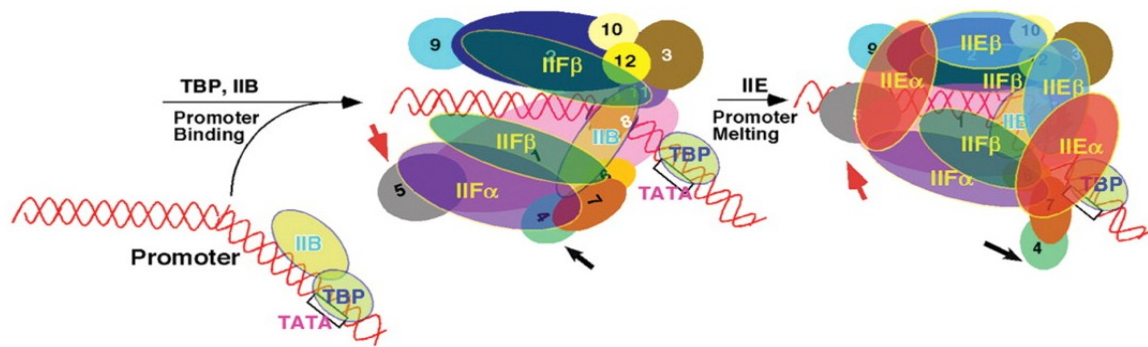


Figure 10: Recruitment to and anchoring of the RNA Polymerase II (the 12 subunits are displayed in numbers) at the promoter and subsequent conformation change upon stable binding to the TSS. The subunits (α , β) of several GTFs are also depicted (IIB: TFIIB; IIF: TFIIF; IIE: TFIIE; (Hirose and Ohkuma, 2007)).

2.7.2 Transcription Elongation

The establishment of the PIC is followed by transcription initiation and elongation. Two different modes of elongation by RNA Pol II can be distinguished: early and productive elongation. Early elongation is unproductive and the RNA Pol II is frequently stalling and pausing (reviewed in (Li and Gilmour, 2011; Nechaev and Adelman, 2011)). Upon transition to productive elongation, which is triggered by several RNA Pol II-interacting factors including the Positive Transcription Elongation Factor b (P-TEFb), the resulting mature RNA Pol II complex is remarkably stable (reviewed in (Nechaev and Adelman, 2011)). During productive elongation, RNA Pol II incorporates nucleotides into nascent RNA at a rate of $10\text{-}70\text{ s}^{-1}$ (Darzacq et al., 2007; Neuman et al., 2003) with a fidelity of about 1 error per 100'000 inserted bases (Ninio, 1991).

The productivity of the RNA Polymerase II is largely dependent on the temporal and spatial distribution of PTMs on its CTD (see also **Supplementary Results 4.4**). The levels of serine 5 phosphorylation (S5p) are highest at promoters and decrease successively towards the 3' end of a gene, whereas serine 2 phosphorylation (S2p) increases throughout the gene (**Figure 11**; reviewed in (Heidemann et al., 2013)). Kinases responsible for CTD phosphorylation are CDK7 (subunit of TFIIH) for S5 and S7, ERK2 for S5, CDK9 (subunit of P-TEFb) and BRD4 for S2 ((Devaiah et al., 2012; Tee et al., 2014) and reviewed in (Heidemann et al., 2013)). There is extensive cross-talk among the CDK7, CDK9 and BRD4 kinases. BRD4 and CDK9

phosphorylate each other at multiple sites. Phosphorylation by CDK9 enhances the CTD kinase activity of BRD4, while phosphorylation of BRD4 by CDK7 inhibits its CTD kinase activity. In addition to these direct interactions, the GTF TAF7 (subunit of TFIID) regulates the CTD kinase activities of CDK7, CDK9 and BRD4 indirectly. However, the sequential order of these interactions is still unclear. It can be speculated that once the CTD is phosphorylated at S5 (by CDK7 or ERK2), the BRD4 kinase is activated and thus the CTD can be phosphorylated at S2. This is in line with the finding that BRD4 preferentially phosphorylates S2 of CTD when it is pre-phosphorylated at S5 by CDK7 (Devaiah et al., 2012; Devaiah and Singer, 2012). *In vitro* assays with purified mammalian RNA Pol II revealed that 5-mC and 5-hmC do not deteriorate nucleotide incorporation efficiency much. Yet, both 5-fC and 5-caC reduced the incorporation rate significantly, indicating that the occurrence of 5-fC and/or 5-caC may perturb efficient elongation by RNA Pol II, possibly due to its stalling (Kellinger et al., 2012).

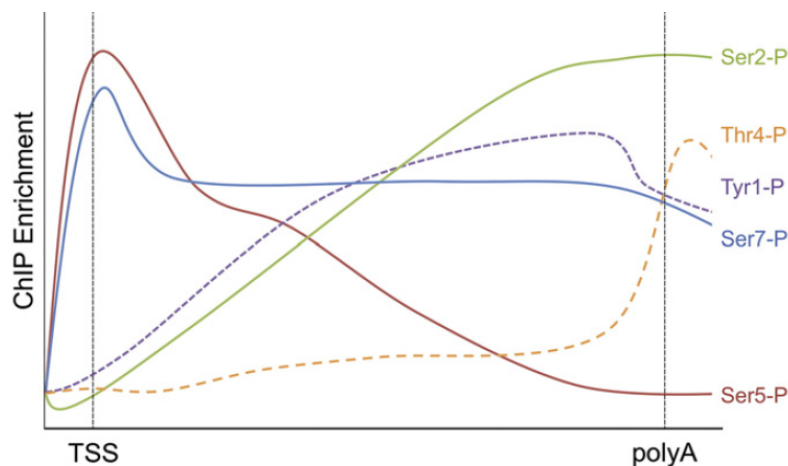


Figure 11: Typical phosphorylation pattern across a transcribed gene, from TSS to poly(A) tail (Heidemann et al., 2013).

2.7.3 Transcription Termination

After successful elongation, transcription is terminated, which implicates disengagement of both RNA Pol II and the transcript from the DNA template (reviewed in (Nechaev and Adelman, 2011)). After RNA Pol II transcribes past the poly(A) site located in the 3' end of the gene, the RNA is cleaved by the polyadenylation machinery and the RNA downstream of the cleavage site is

degraded (Kim et al., 2004). Polyadenylation of the transcript is tightly linked to the termination of transcription; polyadenylation factors are even required for transcription termination ((Kim et al., 2004) and reviewed in (Buratowski, 2009)). S2p of the CTD triggers the co-transcriptional recruitment of the cleavage and polyadenylation specificity factor (CPSF or PCF11). Subsequently, CPSF tethers further polyadenylation factors, such as the poly(A) polymerase (PAP), the cleavage stimulation factor (CstF) and cleavage factors I and II (CFI/II_m). The assembly of these proteins ensures mRNA maturation (reviewed in (Buratowski, 2009)). In yeast, the 5'-3' exonuclease Rat1 is recruited to the 3' ends of protein coding genes, indicating that poly(A) site cleavage by Rat1 triggers transcription termination ((Kim et al., 2004) and reviewed in (Buratowski, 2009)). However, termination of RNA Pol II transcription can also function through a non-canonical, poly(A)-independent pathway, especially for shorter transcripts, including those of the small nucleolar RNA genes (reviewed in (Nechaev and Adelman, 2011)). This appears to require the activity of the negative elongation factor (NELF), whose recruitment is promoted by TFIIF kinase activity (reviewed in (Buratowski, 2009)). Surprisingly, components of the RNA end processing as well as yeast transcriptional termination complexes have been detected at 5' ends of genes, interacting with RNA Pol II marked by S5p ((Vasiljeva et al., 2008) and reviewed in (Nechaev and Adelman, 2011)). At least half of the tandem heptadpeptide repeats are required for transcription termination (Ryan et al., 2002), providing evidence for the involvement of the versatile CTD domain of RNA Pol II in mRNA processing (McCracken et al., 1997; Proudfoot, 2004). It has been suggested that the released RNA Pol II may be "recycled" back to the promoter at genes, where multiple rounds of transcription take place in rapid succession, thus facilitating productive transcription (Yao et al., 2007).

3. Aims of My PhD Thesis

Recent research has shown that TDG, a *bona fide* DNA repair enzyme, has a role beyond classical repair of damaged DNA. It has emerged as a key player in active DNA demethylation initiated by the TET proteins, and thus as a master epigenetic regulator (Cortazar et al., 2011; He et al., 2011; Maiti et al., 2013; Shen et al., 2013; Song et al., 2013). Genome-wide studies have associated TET proteins and products of their 5-mC oxidation predominantly in gene bodies and at gene regulatory elements (Shen et al., 2013; Song et al., 2013; Song et al., 2011; Stadler et al., 2011) and both TET proteins and TDG were implicated in regulation of gene expression ((Williams et al., 2011) and reviewed in (Cortazar et al., 2007)). We showed before that TDG is essential for embryonic development and contributes to the maintenance of an active chromatin environment, i.e. counteracts aberrant DNA methylation and histone modifications and, thus, gene silencing in cells undergoing differentiation-associated epigenetic transitions ((Cortazar et al., 2011); **Appendix III**).

The molecular mechanism by which TDG prevents DNA from hypermethylation was not clear. Furthermore, the effect of TDG on both the regulation of gene expression and the productivity of the transcription machinery were not known. In my PhD thesis, I thus wanted to further investigate the exact role of TDG in epigenetic control of DNA methylation in the context of stem cell differentiation, the mechanism of its action, including its functional interactions and coordination with TET proteins and the resulting impact on gene regulation.

The **first goal** of my PhD thesis was to investigate the function and consequence of TDG-dependent DNA methylation control during early embryonic stem cell differentiation (**Appendix I**). The **second goal** was to explore genome-wide interactions and possible coordination of TDG and TET proteins in the control of epigenetic states at gene regulatory elements as well as in a putative direct association with the RNA polymerase II transcription machinery (**Appendix II, Supplementary Results 4.4**).

To these ends, I thoroughly characterized DNA methylation and demethylation dynamics during stem cell differentiation and I generated and functionally investigated genome-wide data sets for TDG, TET1 and TET2 chromatin association. Additionally, I explored the performance of transcription initiation and elongation events in dependence of the presence and functionality of TDG.

4. Results

In the following section, I will summarize the results presented in the manuscripts attached in the appendix, as well as a selection of additional results that are not contained in the manuscripts.

4.1 TDG balances DNA methylation and oxidative demethylation in differentiating cells (Appendix I)

Our work summarized in Cortázar et al. ((Cortazar et al., 2011); Appendix III) established a role for TDG in the protection of CpG island promoters of developmental genes from aberrant DNA methylation during cell differentiation. Meanwhile, work by others had clearly implicated TDG in active DNA demethylation by excising the 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) products generated by oxidation of 5-methylcytosine (5-mC) by the Ten Eleven Translocation (TET) family of proteins (He et al., 2011; Maiti and Drohat, 2011; Tahiliani et al., 2009). Thus, we wanted to investigate how the TET-TDG system of active DNA demethylation contributes to the epigenetic programming of cells during differentiation. We observed a failure of TDG-deficient mESCs to differentiate into late neuronal progenitor cells (NPs) and neurons, whereas no morphological difference could be observed in undifferentiated cells (Appendix I, Supplementary Figure 1a). To determine whether this observation is caused by defective DNA demethylation, we performed MeDIP-seq of mESCs, *in vitro*-differentiated NPs and MEFs in TDG-proficient and TDG-deficient cells, respectively. We could confirm that differentially methylated regions (DMRs) arose upon differentiation only: more than 900 DMRs were identified in NPs and more than 32'000 in MEFs, while none were detected in mESCs (Appendix I, Figure 1a). Interestingly, DMRs with a high CpG density, predominantly representing CGIs, were almost exclusively hypomethylated in TDG-deficient NPs, while CpG-poor DMRs were predominantly hypermethylated (Appendix I, Figure 1b). Most of the hypomethylated NP-specific DMRs turned out to be CGIs that become *de novo* methylated upon lineage commitment (Appendix I, Figure 1c). Importantly, these CGI DMRs are significantly more often bound by TET1 and marked by H3K27me3 and H3K4me1, and often they represent low-methylated regions (LMRs; Appendix I, Figure 1d). LMRs are CpG-poor distal regulatory regions

and they are occupied by cell-type-specific DNA-binding factors, which locally influence DNA methylation. This showed that the loss of TDG causes a misregulation of DNA methylation at regions important for gene regulation including CGIs and enhancers. Since the methylation analysis was done by MeDIP-seq with a 5-mC-specific antibody, the observed loss of DNA methylation can result from either of these different events; (i) the exchange of 5-mC with a C, (ii) the oxidation of 5-mC to 5-hydroxymethylcytosine (5-hmC), (iii) mutation of 5-mC to T. All these events would invariably lead to the loss of the 5-mC epitope for the antibody. 5-mC could be replaced by a C, either after replication when DNA maintenance methylation is deficient, or upon active DNA demethylation. The oxidation of 5-mC can be achieved catalytically by one of the TET proteins, or theoretically likewise through oxidative DNA damaging agents. C→T transition mutations can arise from deamination of 5-mC or 5-hmC to T or 5-hmU opposite G, respectively. This could occur if a deaminase, for instance AID, would be engaged on a regular basis. By hairpin Na-bisulphite sequencing (BS-seq), which allowed us to evaluate a putative elevation of the mutation frequency besides measuring the levels of DNA methylation, we could validate the presence of the above identified DMRs. However, we were not able to detect an elevated mutation frequency in TDG-deficient NPs: the frequency of C→T transition mutations did not rise above the error rate of the method and since the methylation changes we measured were in the higher percentage range, mutations cannot explain the methylation changes we observed. These results argued strongly against a deamination-based pathway for DNA demethylation at the DMRs tested (Appendix I, Figure 2).

Being interested in the levels of DNA demethylation intermediates in response to TDG (5-hmC, 5-fC and 5-caC), we established a 24h differentiation of mESCs with retinoic acid (RA), to minimize the effects of replication-dependent dilution of the above mentioned intermediates (Appendix I, Figure 3a, Supplementary Figure 2). The analysis of genomic DNA from time points 0h, 8h and 24h by LC-MSMS revealed a significant increase of 5-fC and 5-caC in cells without TDG or with catalytically inactive TDG (Appendix I, Figure 3b); the global levels of 5-mC, 5-fC and 5-caC increased with differentiation and this effect was more pronounced in cells lacking TDG activity (Appendix I, Supplementary Figure 3). In agreement with previous work (He et al., 2011), 5-caC levels were 9 fold higher, already in undifferentiated mESCs lacking TDG activity. Although catalytically inactive TDG is

unable to remove substrate bases, it binds to 5-caC with very high affinity (Appendix I, Figure 4b, Supplementary Figure 4b). Hence, the processing of DNA demethylation intermediates which accumulate upon cell differentiation is dependent on TDG's catalytic activity. To clarify whether these global changes also reflect changes at the CGI DMRs, we investigated 5-mC, 5-hmC and 5-caC levels at these loci. Compared to TDG-deficient mESCs, we found increasing 5-mC, 5-hmC and 5-caC levels at selected CGI DMRs in mESCs expressing wildtype or catalytically inactive TDG upon cell differentiation (Appendix I, Figure 4a, Supplementary Figure S4a). To investigate whether TET and TDG associate with the CGI DMRs, we performed ChIP-qPCR experiments, which revealed an enrichment of both at these loci. Interestingly, catalytically inactive TDG showed a particularly high enrichment, presumably due to its inability to process the substrates and continued re-association with 5-fC and 5-caC. TET1 enrichment was particularly high in a TDG wildtype background, indicating that TDG facilitates TET1 association, presumably by eliminating the oxidized 5-mC and, thus, giving way to a new round of methylation (Appendix I, Figures 5a and 5c, Supplementary Figures 5 and 6).

Differentiation towards NPs requires non-NP-specific genes to be repressed by *de novo* methylation. We propose that TET and TDG constitute an active DNA methylation – demethylation system that operates to maintain epigenetic plasticity at genomic loci that undergo programming in the context of cell lineage commitment. During differentiation, 5-mC is oxidized by TET, finally generating 5-caC which is excised by TDG. Re-methylation is catalyzed by DNMT3a or DNMT3b. This cycle will be exited eventually, once lineage commitment is accomplished and the point of exit determines the final CpG methylation state of the gene regulatory regions. Importantly, CGI DMRs in a TDG-deficient background did not show any changes in the levels of DNA demethylation intermediates, indicating a structural role of TDG in coordinating the DNA methylation – demethylation cycle, besides its crucial catalytic role in excising DNA demethylation intermediates (Appendix I, Figure 4a).

Taken together, we found that TDG safeguards the equilibrium between 5-mC and DNA demethylation intermediates during lineage restriction, which is a state of high epigenetic plasticity. We describe a “DNA methylation – oxidative demethylation cycle” which occurs during cell differentiation at CGIs that are preferentially associated with the polycomb-mediated mark H3K27me3 and poised enhancers, and

are prone to be methylated in the process of lineage commitment. Hence, TDG might be selectively targeted to cell-fate-specific regulatory sequences where it actively contributes to DNA demethylation by excising TET-mediated 5-mC oxidation products and coordinates the initiation and/or maintenance of the cycle in a structural manner (Appendix I, Figure 6).

Contribution: I conducted MeDIP, GLIB and caC-DIP experiments, performed quantitative PCRs for the targets shown and further positive and negative controls (data not shown) as well as extensive quantitative analysis and quality controls of the DIP and GLIB data sets. Further, I prepared chromatin extracts of the timecourse samples prepared by Angelika Jacobs and conducted TET1 and TDG ChIP experiments as well as subsequent qPCRs. I quantitatively analysed and conducted quality controls of these ChIP data sets. I further did fC-DIP including qPCR analysis, expression analysis of the DMRs and surrounding genes (RT-qPCR), TET1 and TDG ChIPs including qPCR from TET1-knockout mESCs, and I established and performed TET2 ChIP (data not shown). Additionally, I prepared samples for oxBis and MAB-seq analysis by Pascal Giehr and coordinated the collaboration with him.

4.2 TET1, TET2 and TDG cooperate locus-specifically to promote chromatin plasticity by oxidative DNA demethylation (Appendix II)

TET1 and TET2 have been shown to associate genome-wide with gene bodies, TSSs and enhancers, fundamental features for productive transcription (Chen et al., 2013b; Stadler et al., 2011; Williams et al., 2011). Furthermore, the products of TET-mediated 5-mC oxidation (5-hmC, 5-fC and 5-caC) have been shown to arise predominantly at gene regulatory elements of developmental genes (Raiber et al., 2012; Serandour et al., 2012; Shen et al., 2013; Song et al., 2013; Stadler et al., 2011). This hints at a mechanism where DNA methylation of gene regulatory elements is controlled by the TET proteins in the context of gene regulation. To elucidate a putative involvement of and a functional interaction with TDG, we wanted to determine the genome-wide association of TDG and further clarify a possible dependency of TET chromatin-interactions on the presence or activity of TDG. We thus performed ChIP followed by deep sequencing (ChIP-seq) for TDG, TET1 and TET2. ChIP-seq was performed in isogenic mouse embryonic stem cells (mESCs) deficient for TDG and complemented with minigenes encoding either a wildtype TDG (*Tdg*[wt]), a catalytically inactive but structurally intact TDG (*Tdg*[cat]) or the empty

vector (*Tdg*[null]). Since TDG function becomes apparent in differentiating rather than in pluripotent mESCs (Appendix I), the cells were differentiated for 24 hours in the presence of all-*trans* retinoic acid (RA). Datasets were obtained for TDG-ChIP-seq in *Tdg*[wt] and *Tdg*[cat] backgrounds at 24h RA and TET1- and TET2-ChIP-seq in *Tdg*[wt], *Tdg*[cat] and *Tdg*[null] backgrounds at 0h RA and 24h RA. After aligning the reads and peak calling, we were able to identify 14'144 TDG peaks in *Tdg*[wt] and 28'164 in *Tdg*[cat]. Notably, we found only a minor overlap (3'058), which can be explained with the different enzymatic properties of wildtype and catalytically inactive TDG (Appendix II, Figure 1a). We explain this divergence by the fact that the catalytically inactive TDG exhibits reduced turnover and is thus potentially more efficiently detected by ChIP than the wildtype TDG. Catalytically active TDG may well be associated with the same sites as the inactive TDG, but due to its rapid turnover not as well detectable by ChIP. When we mapped the average distance of the TDG peaks to the nearest TSS, we found a narrow peak around the TSS and a broader peak ranging from 1 to 1000 kb away from the TSS (Appendix II, Figure 1b). Interestingly, there was no difference between *Tdg*[wt] and *Tdg*[cat]. Next, we characterized the preferential genomic features that TDG associates with: we found an increased association of both active and inactive TDG to regions of intermediate to high CpG density, including a significant enrichment of CGIs. Functional characterization then revealed a preference for introns and intergenic regions (Appendix II, Figures 1c and 1d). Furthermore, a significant portion of TDG peaks in both *Tdg*[wt] and *Tdg*[cat] coincides with histone marks associated with transcription (H3K36me3, H3K4me3, H3K9ac), the latter two characteristic for active promoters (Appendix II, Figure 2c). Additionally, there was a striking enrichment of the polycomb-mediated histone mark H3K27me3 in TDG peaks, which occurs predominantly at CGI promoters of developmental genes. Contrarily, the heterochromatin mark H3K9me3 did not show a comparable enrichment (Appendix II, Figure 2b), indicating that TDG is not targeted randomly to chromatin. Excitingly, TDG peaks (in *Tdg*[wt] and *Tdg*[cat]) are strongly enriched for sites of DNase hypersensitivity, dynamic histone variants (H3.3 and H2A.Z) as well as histone marks characteristic for active enhancers (H3K4me1, H3K27ac), implicating a possible function for TDG at gene regulatory elements which are characterized by high nucleosome dynamics and depletion (Appendix II, Figures 2a, 2d and 2e). Generally,

TDG peaks in *Tdg[wt]* and *Tdg[cat]* showed very similar properties, suggesting that they represent subfractions of the same functional entity.

Next, we were interested in the proportion of TDG peaks that are also bound by TET1 or TET2, respectively. We found 38% and 31% of TDG peaks (in *Tdg[wt]* and *Tdg[cat]*, respectively) co-bound by TET1 in *Tdg[wt]* mESCs and only 10% and 4% of TDG peaks (in *Tdg[wt]* and *Tdg[cat]*, respectively) co-bound by TET2 (Appendix II, Figure 3a). Compared to the TET1/2 occupancy in the whole genome, the increase of TET1/2 peaks within the TDG peaks was highly significant. Interestingly, the co-localization of both TET1 and TET2 with TDG was higher in *Tdg[wt]* compared to *Tdg[cat]*, indicating that blocking of the removal of 5-fC and 5-caC will interfere with *de novo* TET binding and the subsequent cyclic DNA methylation and oxidative demethylation. This suggests that TET and TDG are not engaged as a complex but individually. The triple overlap between active TDG and inactive TDG as well as TET1 and TET2 revealed 633 and 825 targeted loci, respectively (Appendix II, Figure 3b). Interestingly, only a minor fraction of TDG-TET1 peaks is additionally associated with TET2, whereas the major fraction of TDG-TET2 peaks is additionally bound by TET1, indicating a close functional relationship between TDG and TET2. Gene ontology analysis of the triple-positive peaks identified mainly pathways involved in gene expression and cellular development (Appendix II, Figure 3e). Especially CGIs are significantly enriched within these triple-positive peaks, compared to the peaks that are only bound by TDG (Appendix II, Figures 3c), indicating that the triple-positive peaks could indeed account for the peaks located in CpG-dense regions (Appendix II, Figure 1d). We conclude from these data, that the triple-positive TET1-TET2-TDG sites are predominantly targeted to promoter CGIs, whereas the unique TDG peaks show an enrichment for active enhancers. This is consistent with the histone modifications that were found to be enriched in TDG peaks, however, it indicates that there is a functional separation between TDG only and triple-positive fractions.

We could not determine a dependency of TET1 or TET2 association to chromatin on TDG presence at 24 hours of differentiation; the majority of TET1 peaks in *Tdg[wt]* were also bound in *Tdg[cat]* (64.6%) and in *Tdg[null]* (64.1%), indicating that presence and functionality of TDG have no influence on the major portion of TET1 binding events. This was not unexpected, since TET1 is implicated upstream of TDG

(Appendix II, Figure 4a). Interestingly, the overlap between TET2 binding sites in the different TDG backgrounds at 24h RA exposure was lower than that observed for TET1; only 42% of total TET2 peaks in *Tdg[wt]* were also bound by TET2 in *Tdg[cat]* and 48% were in the *Tdg[null]* background (Appendix II, Figure 4b). Fascinatingly, whereas the number of TET2 peaks in a background without or with inactive TDG stayed roughly the same throughout the 24 hours of differentiation, it decreased dramatically in a *Tdg[wt]* background, which is unlike the situation for TET1. This dependency seems to arise only upon differentiation, indicating that the turnover of TET2 in particular might be regulated by wildtype TDG during differentiation (Appendix II, Supplementary Figure 6).

To determine whether the combined targeting of the TET proteins and TDG to CGIs might involve active DNA demethylation, we next investigated the correlation between the triple-positive binding sites and the occurrence of DNA demethylation intermediates (5-hmC and 5-fC). Indeed, we found a significant overlap between the triple-positive peaks and 5-hmC as well as 5-fC. The correlation was higher than that obtained for unique TDG peaks, indicating that it is the joint action of both TET proteins and TDG that is necessary for DNA demethylation (Appendix II, Figure 5a). The same observation was made for the bivalent histone marks (H3K4me3 and H3K27me3); they were strongly increased in triple-positive peaks, both in *Tdg[wt]* and *Tdg[cat]* (Appendix II, Figure 5b). This may indicate, that the joint action of both TET proteins and TDG induces DNA demethylation-mediated gene activation at developmental genes during cell differentiation. Since the deposition of the dynamic histone variants H3.3 and H2A.Z was reported to mark gene regulatory elements and regulate the transcriptional outcome of inducible genes (Chen et al., 2013a), we were interested in clarifying whether occurrence of dynamic histone variants is correlated with TDG occupancy. Indeed, we found a significant fraction of both triple-positive as well as unique TDG peaks to be co-occupied by H3.3 (Appendix II, Figure 5c). Moreover, H2A.Z was also highly enriched, particularly in the triple-positive peaks (Appendix II, Figure 5d). Notably, in the unique TDG peaks, the occurrence of H2A.Z was lower in a *Tdg[wt]* compared to *Tdg[cat]* background, indicating that TDG's catalytic activity plays a role in nucleosome turnover. This led us to propose a model, where TET-TDG-mediated DNA demethylation facilitates histone exchange and nucleosome turnover in order to promote a state of chromatin plasticity at regulatory elements of developmental genes in differentiating mESCs.

In summary, we propose a model where TET proteins together with TDG are targeted to gene regulatory regions, particularly CpG islands, where they coordinately engage in active DNA demethylation (Appendix II, Figure 6). The tight spatio-temporal control ensures high plasticity of chromatin and gene expression, a necessity during differentiation. Additionally, the strong correlation with the highly dynamic histone variants H3.3 and H2A.Z suggests a putative involvement of TET and TDG in the mediation of nucleosomal dynamics.

Contribution: I extracted chromatin from short-differentiation timecourses (prepared by Angelika Jacobs and Zeinab Barekati), established ChIP procedures and performed all ChIPs for TDG, TET1 and TET2. I analyzed the samples by qPCR and conducted thorough quantitative and qualitative analyses of the data. I prepared the samples for next generation sequencing and sent them to the Genome Technology Access Center (GTAC) at the Washington University in St. Louis, USA. I coordinated the processing and the analysis of our samples there, and I participated in the coordination of the subsequent bioinformatical analysis. Finally, I wrote the manuscript in Appendix II.

4.3 Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability (Appendix III)

Besides being a DNA repair protein, TDG was also shown to be a co-regulator of transcription. To further clarify the biological role of TDG, we generated *Tdg* knockout mice and, quite unexpectedly, found TDG to be essential for embryonic development; TDG knockout embryos died at E11.5 (Appendix III, Figure 1a, Supplementary Figure 1). TDG-deficient embryos revealed signs of internal hemorrhage and hemorrhagic necrosis, but apart from that no informative pathology. The severity of the phenotype was surprising, since other DNA glycosylases have a substrate spectrum partially redundant with TDG, and additionally, no other knockout of a DNA glycosylase displays embryonic lethality. We thus aimed to investigate the reason for developmental failure of TDG knockout mice. To address a potential DNA repair defect, we subjected TDG-proficient and –deficient MEFs to ionizing radiation and H₂O₂ treatment, but could not find an effect on cell survival dependent on TDG. Also, using different mutation assays, we were not able to measure increased mutation upon loss of TDG (Appendix III, Supplementary Figure 2). This argued strongly against a function of TDG in canonical DNA repair to account for its developmentally

essential role. Next, we assessed the effect of TDG on gene expression in MEFs and this revealed 461 genes differentially regulated in a TDG knockout background. Interestingly, the catalytic activity of TDG was necessary to restore gene expression upon re-introduction of TDG into TDG-deficient MEFs (Appendix III, Figure 2a). These 461 genes comprised many transcription factors (TF) and therefore, the loss of TDG probably has direct and indirect, TF-mediated effects. Gene ontology analysis of these differentially expressed genes identified mostly pathways in embryogenesis and development (Appendix III, Figure 1b, Supplementary Figure 3) and since TDG had been implicated in one way or another in active DNA demethylation and because many of the differentially expressed genes are known to be epigenetically regulated during development, we next analyzed the DNA methylation state of promoter CpG islands (pCGIs) of genes down-regulated upon TDG knockout. Bisulfite sequencing data then revealed aberrant *de novo* DNA methylation in TDG-deficient MEFs (Appendix III, Figure 1c, Supplementary Figure 4). By TDG-ChIP we could also confirm an association of TDG to these developmental gene promoters (Appendix III, Figure 1d). Together with DNA methylation changes, we found a loss of the activating histone mark H3K4me2 and a gain of the repressive histone marks H3K27me3 and H3K9me3 in TDG-deficient MEFs at affected loci. More specifically, active promoters (positive for H3K4me2 in TDG-proficient cells: *Sfrp2* and *Twist2*) acquired H3K27me3; whereas bivalent promoters (positive for H3K4me2 and H3K27me3 in TDG-proficient cells: *HoxD13* and *HoxA10*) acquired H3K9me3 (Appendix III, Figure 1e). This indicated that TDG is recruited to specific gene promoters to protect them from aberrant epigenetic silencing. In case of *Sfrp2* and *Twist2*, we found that the chromatin state as well as gene expression was rescued by the stable complementation of TDG-deficient MEFs with wildtype TDG. In case of *HoxD13* and *HoxA10*, which completely lost H3K4me2 and additionally gained H3K9me3, however, re-introduction of TDG could not restore the bivalent chromatin state. This indicated that, in the latter case, the epigenetic silencing has progressed extensively and thus became irreversible (Appendix III, Figure 2). To investigate the origin of the epigenetic aberrations observed in TDG-deficient MEFs, we performed gene expression analysis in TDG-proficient and TDG-deficient mESCs and NPs derived from these mESCs (Appendix III, Supplementary Figure 6). Interestingly, gene expression differences were minor in the undifferentiated mESCs (16 genes), but increased upon differentiation (297 genes in NPs; Appendix III, Figure 3a). Again,

gene ontology analysis identified a predominant misregulation of developmental genes in TDG-deficient cells (Appendix III, Supplementary Figure 7) and TDG was associated with these gene promoters both in mESCs and in NPs. In line with this, we found aberrant DNA methylation to arise only in the differentiated state and not in mESCs. Interestingly, ectopic expression of TDG during differentiation prevented from the gain of *de novo* DNA methylation at these specific gene promoters (Appendix III, Figures 3b and 3c). Similarly, altered histone methylation patterns were only observed in TDG-deficient differentiated NPs but not in TDG-deficient undifferentiated mESCs (Appendix III, Supplementary Figure 8). To elucidate whether the epigenetic function of TDG involves active DNA repair, as implicated by the necessity of TDG's catalytic activity to restore gene expression, we assessed the association of downstream BER proteins with the above indicated developmental genes by ChIP. Indeed, we found XRCC1 and APE1 to be enriched at these sites in TDG-proficient MEFs, but not in TDG-proficient mESCs (Appendix III, Figure 4a). Thus, TDG induces DNA BER at loci where chromatin state and DNA methylation patterns are altered upon the loss of TDG. Similarly, we found an increase of chromatin-associated XRCC1 foci after treatment of mESCs with all-trans retinoic acid (RA) for 8 hours in the presence but not in the absence of TDG. This was further supported by a higher sensitivity of TDG-proficient cells to PARP inhibition following RA treatment compared to TDG-deficient cells (Appendix III, Supplementary Figures 9 and 10). This strongly indicated that cell-differentiation-induced TDG activity initiates PARP and XRCC1-dependent DNA single strand break repair. We further found association of histone modifying proteins to the promoter CGIs of developmental genes; specifically, we found a significant enrichment of the H3K4-specific methyltransferase MLL1 and the histone acetyltransferases CBP/p300 only in TDG-proficient MEFs. In undifferentiated mESCs, however, we did not detect MLL1 or CBP/p300 enrichment, independent of the TDG background (Appendix III, Figure 4b). The thereby induced changes in histone modifications were only observed in a TDG-proficient background.

In summary, this study established a novel role of TDG and DNA repair in the epigenetic maintenance of CpG island promoters during cell differentiation (Appendix III, Figure 4c). In that context, TDG apparently has structural as well as catalytic functions; structurally, TDG complexes with activating histone modifiers to maintain the chromatin state in an active or bivalent conformation during cell differentiation.

Upon loss of TDG, chromatin modifications are imbalanced towards a more repressive state. Catalytically, TDG activity contributes to erasure of aberrant DNA methylation and recruitment of downstream BER proteins (XRCC1 and APE1). This suggests that TDG is part of a repair-mediated control system which contributes to epigenome stability at critical DNA sequences during cell differentiation.

Contribution: I contributed to the preparation of chromatin extracts from MEFs and mESCs and to ChIP experiments for H3K4me2, H3K9me3 and H3K27me3 in MEFs (Appendix III, Figure 1e; together with Daniel Cortázar). I performed ChIP experiments in wildtype-complemented MEFs for all histone marks shown (Appendix III, Figure 2b); as well as for XRCC1 and CBP in MEFs (Appendix III, Figures 4a and 4b). Additionally, I performed ChIP experiments for H3K4me2 and H3K27me3 in mESCs (Appendix III, Supplementary Figure 8). Moreover, I performed ChIP experiments for H3K4me2, H3K9me3 and H3K27me3 in MEFs carrying a catalytic-dead TDG (data not shown). Finally, I performed qPCR analysis of TDG ChIPs in MEFs and XRCC1 ChIPs in NPs (together with Daniel Cortázar).

4.4 Supplementary Results

Investigating the Effect of TDG and BRD4 on RNA Polymerase II Phosphorylation State and Productive Transcription

Wirz, A. & Schär, P.

Department of Biomedicine, University of Basel, Switzerland

Correspondence:

primo.schaer@unibas.ch; Tel. +41 61 267 3561; Fax: +41 61 267 3566

An attractive hypothesis is that the DNA repair function of TDG and its role in cyclic DNA methylation and demethylation at gene regulatory elements reflects different aspects of one and the same function in the context of transcriptional activity. Thymine DNA glycosylase (TDG) has been linked to transcription through its manifold interactions with transcription factors (TF). TDG was shown to have both co-activator and co-repressor functions (Chen et al., 2003; Kim and Um, 2008; Missero et al., 2001; Tini et al., 2002). For example, TDG was reported to stimulate transcription by interaction with the nuclear TFs retinoic acid receptor (RAR) and retinoid X receptor (RXR), resulting in enhanced binding of these factors to their cognate response elements and increased transcriptional activity (Leger et al., 2014; Um et al., 1998). Our laboratory recently discovered an interaction of TDG with BRD4 in an yeast two-hybrid screen (Roland Steinacher and Primo Schär, unpublished data); BRD4 is a bromodomain protein, interacting with acetylated histones (e.g. H3K9ac) and a number of chromatin modifiers and it is also an atypical kinase for serine 2 on the C-terminal domain (CTD) of the RNA Polymerase II (RNA Pol II; (Devaiah et al., 2012; Liu et al., 2008; Rahman et al., 2011)). This latter discovery is very interesting as it may link TDG's role in chromatin dynamics directly to the regulation of RNA Polymerase II.

Therefore, I performed pilot experiments to test this possibility. I analyzed the effect of catalytically inactive and wildtype TDG on transcriptional outcome of the *RARβ* locus. *RARβ* was chosen because its expression is induced by RA and because our ChIP data confirmed that TDG is strongly associated with this locus. Furthermore, I characterized the phosphorylation pattern of RNA Pol II's C-terminal domain (CTD) – which is highly indicative of transcriptional progression (Bataille et al., 2012; Gebara et al., 1997) – as a function of TDG, to be able to distinguish between effects on transcriptional initiation and productive elongation. Inhibition of BRD4 with the small-molecule inhibitor JQ1 was shown to abrogate its binding to acetylated histones, which coincides with the inhibition of BRD4-dependent RNA Pol II phosphorylation (Devaiah et al., 2012). I thus wanted to additionally investigate the effect of BRD4 inhibition on the phosphorylation pattern of the CTD in the presence or absence of TDG or its catalytic activity.

First, I confirmed the inducibility of *RARβ* expression upon the treatment of mouse embryonic fibroblasts (MEFs) with all-*trans* retinoic acid (RA). Induction was

observed independent of the *Tdg* background (*Tdg*^{-/-} MEFs complemented with minigenes encoding either wildtype TDG (*Tdg*[wt]), catalytically inactive TDG (*Tdg*[cat]) or the empty vector (*Tdg*[null])). Interestingly, compared to mock-treated MEFs (DMSO), the induction of *RARβ* expression was lowest in *Tdg*[wt] (20-fold, *p* = 0.013) compared to *Tdg*[cat] (53-fold, *p* = 0.001) and *Tdg*[null] (65-fold, *p* = 0.0015); **Figure 1a**). Treatment of these MEFs with JQ1 showed an inverse effect on *RARβ* gene expression in *Tdg*[wt] and *Tdg*[null]: expression was decreased in *Tdg*[wt] but was significantly increased in *Tdg*[null] upon JQ1 treatment compared to control treatment (DMSO; *p* = 0.015; **Figure 1b**), although expression levels were generally low. These results indicate a synergistic effect of the combined absence of TDG and BRD4 on induction of *RARβ* expression.

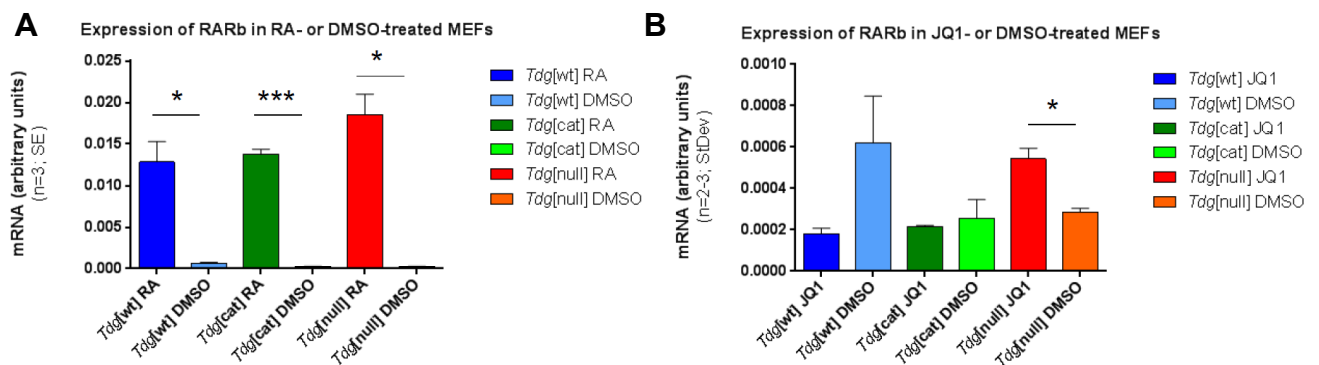


Figure 1: Expression analysis of the *RARβ* gene in MEFs in different *Tdg* backgrounds. A: RT-qPCR of mRNA extracted from MEFs, treated with all-*trans* retinoic acid (RA; 1 μ M for 24h) or with DMSO as a mock control. **B:** RT-qPCR of mRNA extracted from MEFs, treated with the BRD4 inhibitor JQ(+1) (250 nM for 24h) or with DMSO as a mock control. Normalization was conducted against β -*Actin* and *B2m* (in **1a** and **1b**). *n*=2 or 3; error bars are standard deviations or standard errors, respectively. two-tailed t-tests; *p* < 0.05: *; *p* < 0.01: ***.

To investigate a possible role of TDG in transcription initiation, I then checked the influence of RA-treatment on the formation of the preinitiation complex (PIC), which is characterized by an unphosphorylated state of the RNA Pol II CTD (Lolli, 2009), in MEFs expressing wildtype, catalytically inactive or no TDG. This was achieved by chromatin immunoprecipitation (ChIP) with an antibody directed against the CTD of RNA Pol II followed by quantitative PCR (qPCR). Indeed, I found RA-treatment promoting an elevated association of the RNA Pol II to the promoter region of *RARβ* (**Figures 2a and 2b**). This was most pronounced when TDG was catalytically inactive (*Tdg*[cat]) or absent (*Tdg*[null]), suggesting that loss of TDG activity is beneficial for PIC formation. To see how the *Tdg* background further affects

downstream steps of transcription, I performed ChIP-qPCR experiments for CTD phosphorylated at serine 5 (S5p; the mark for promoter escape and abortive transcription) and serine 2 (S2p; the mark for productive transcription elongation) (Devaiah and Singer, 2012; Heidemann et al., 2013; Phatnani and Greenleaf, 2006). To facilitate investigation of the phosphorylation pattern and the association of RNA Pol II along the entire *RAR β* gene, I defined regions of interest along the gene (primer pairs 1 through 7; **Table 2 and Supplementary Figure 1**). Upon treatment with RA, I detected a significantly increased S5 phosphorylation in *Tdg[cat]*, however, the distribution of S5p RNA Pol II did not differ significantly between *Tdg[wt]* and *Tdg[null]* throughout the *RAR β* locus (**Figures 2c and 2d**). S5 phosphorylation of the CTD in *Tdg[cat]* was most pronounced in the promoter region (Map 1), but was additionally also enriched towards the 3' end of the gene (Map 6 and Map 7). The genuinely higher levels of S5p in the *Tdg[cat]* background for RA-treated MEFs (**Figure 2c**) could be explained by an effect on co-transcriptional alternative splicing. It is along these lines, that the CTD is rather phosphorylated at S5 than S2 upon slowing down of the RNA Pol II due to an alternative splicing event at an exon variant (Batsche et al., 2006). The fact that this is more pronounced in *Tdg[cat]* compared to *Tdg[wt]* and *Tdg[null]* might be explained by the delay of RNA Pol II progression due to a prolonged association to or constant association and dissociation from chromatin by catalytically inactive TDG (compare to (Hardeland et al., 2000; Hardeland et al., 2002)). This could implicate that TDG activity contributes to alternative splicing or spliceosome assembly at particular intron-exon configurations. Interestingly, DNA methylation has been reported to regulate alternative splicing events (Maunakea et al., 2013); this constitutes yet another possibility for TDG to participate in the regulation of transcription, namely by actively demethylating alternatively spliced exons, resulting in aberrant splicing.

As expected, levels of S2 phosphorylation increased towards the 3' end of the *RAR β* locus (**Figures 2e and 2f**), which is in accordance with a previous report (Mahony et al., 2011). Interestingly however, the increase was strongest, though not significant, in the absence of TDG (*Tdg[null]*) at Map 6 and Map 7. This could indicate two things; on one hand the structure of TDG may be inhibitory for productive transcriptional elongation, on the other hand, absence of TDG (*Tdg[null]*) could cause the transcription machinery to slow down, thus facilitating detection by ChIP-qPCR due to more efficient cross-linking.

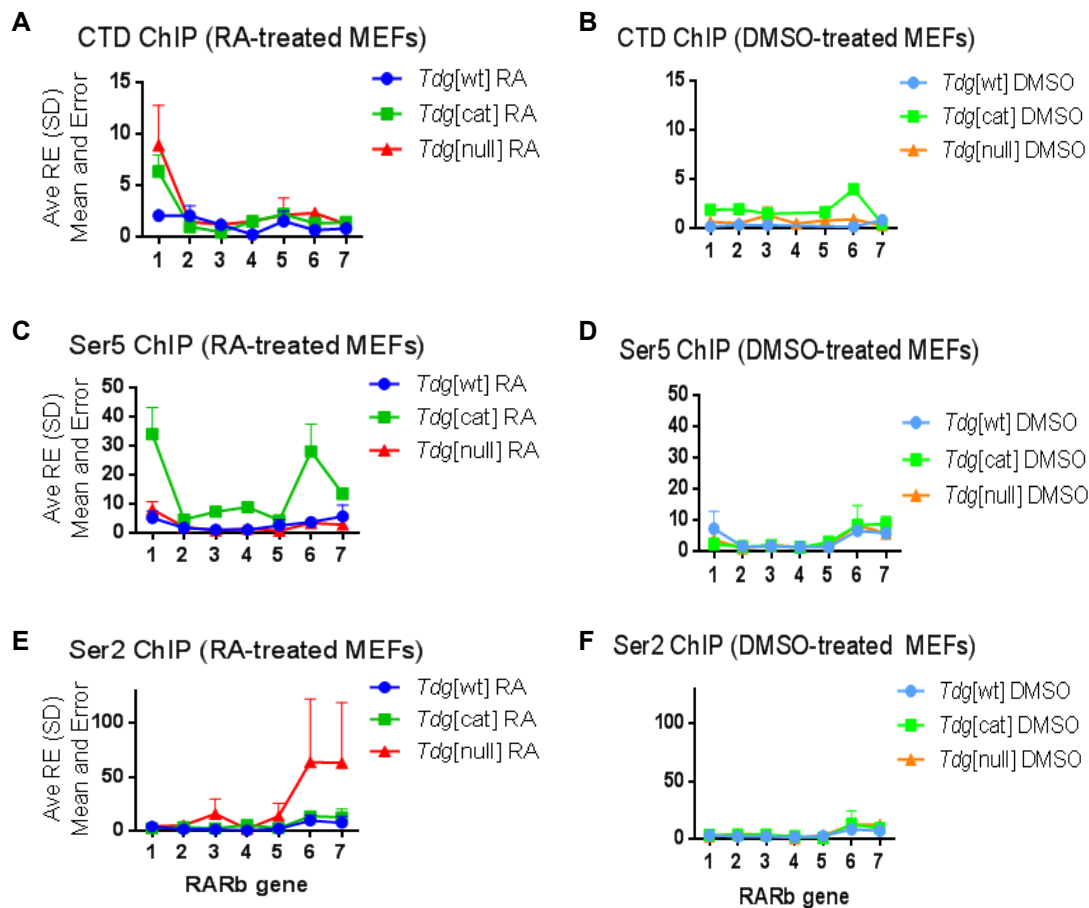


Figure 2: Association of the RNA Polymerase II to the *RARβ* gene upon RA-treatment. **A:** RNA Pol II CTD ChIP and mapping along the *RARβ* gene by qPCR in RA-treated MEFs. **B:** RNA Pol II CTD ChIP and mapping along the *RARβ* gene by qPCR in DMSO-treated MEFs. **C:** RNA Pol II S5p ChIP and mapping along the *RARβ* gene by qPCR in RA-treated MEFs. **D:** RNA Pol II S5p ChIP and mapping along the *RARβ* gene by qPCR in DMSO-treated MEFs. **E:** RNA Pol II S2p ChIP and mapping along the *RARβ* gene by qPCR in RA-treated MEFs. **F:** RNA Pol II S2p ChIP and mapping along the *RARβ* gene by qPCR in DMSO-treated MEFs. % input is normalized against an external control (chr2:75'408'932-75'409'164); n=2, error bars are standard deviations. Mapping positions: 1: promoter CGI in exon 1; 2: intron 1, 15 kb downstream of TSS; 3: exon 2, 25 kb downstream of TSS, 3'UTR of one isoform; 4: intron 3, high CpG content (45%), 70 kb downstream of TSS; 5: exon 4, 120 kb downstream of TSS; 6: exon 6, 5 kb upstream of 3'UTR of main isoform; 7: just after exon 8, 3'UTR.

BRD4 was shown to bind to acetylated histone proteins, function as a regulator of cell cycle progression and serve as a scaffold for the recruitment of the positive transcription elongation factor b (P-TEFb). Further, BRD4 was shown to be an atypical kinase that phosphorylates S2 on the CTD of RNA Pol II (Devaiah et al., 2012). S2 phosphorylation of RNA Pol II's CTD can be inhibited by JQ1, which functions by competitively binding to BRD4's BET-bromo-domain, thereby preventing its binding to chromatin (Delmore et al., 2011). I was thus interested if and how the loss of BRD4 binding affects the progression of the RNA Pol II depending on the TDG status. For this purpose, I treated MEFs with JQ1 and investigated the

association of RNA Pol II to the *RARβ* locus as well as the phosphorylation pattern RNA Pol II's CTD. In TDG-deficient cells (*Tdg[cat]* and *Tdg[null]*) I measured a trend to more CTD association upon BRD4 inhibition (**Figures 3a and 3b**). However, the levels of CTD association were generally low, consistent with the low levels of *RARβ* expression observed (**Figure 1**).

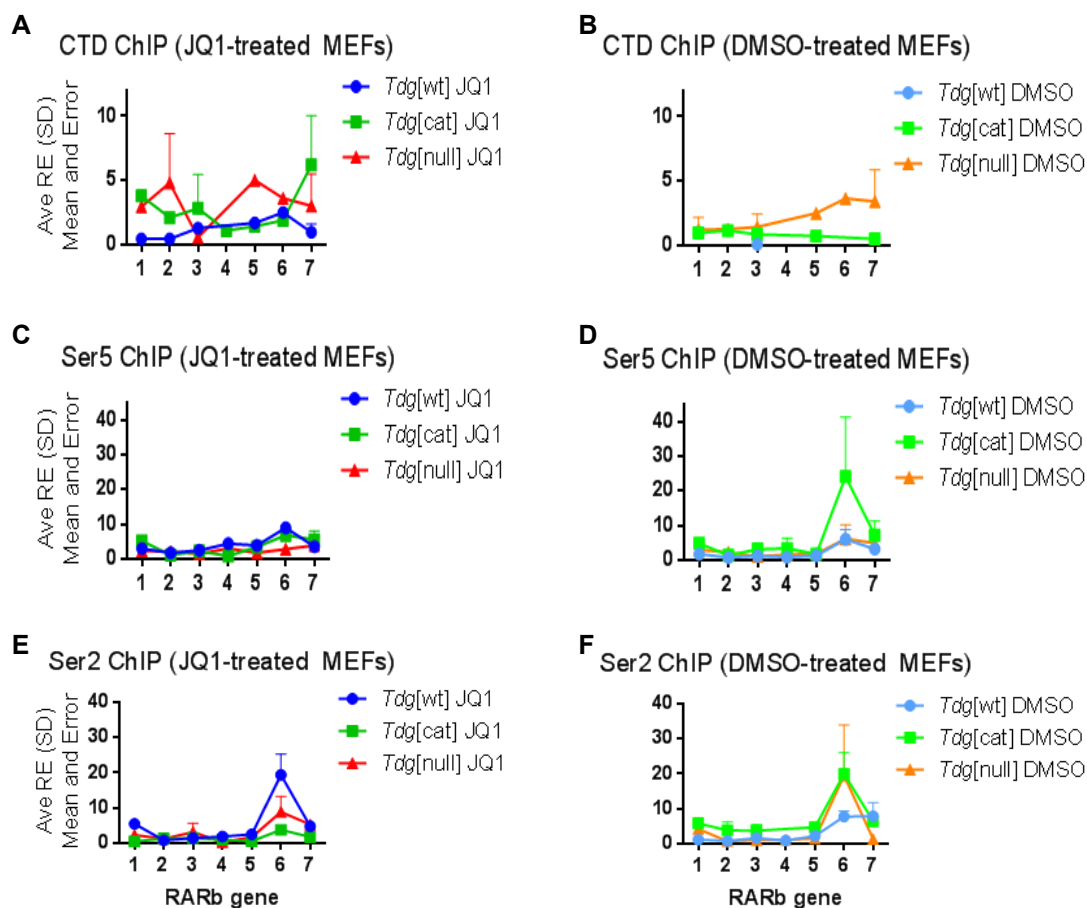


Figure 3: Association of the RNA Polymerase II to the *RARβ* gene upon JQ1-treatment. A: RNA Pol II CTD ChIP and mapping along the *RARβ* gene by qPCR in JQ1-treated MEFs. **B:** RNA Pol II CTD ChIP and mapping along the *RARβ* gene by qPCR in DMSO-treated MEFs. **C:** RNA Pol II S5p ChIP and mapping along the *RARβ* gene by qPCR in JQ1-treated MEFs. **D:** RNA Pol II S5p ChIP and mapping along the *RARβ* gene by qPCR in DMSO-treated MEFs. **E:** RNA Pol II S2p ChIP and mapping along the *RARβ* gene by qPCR in JQ1-treated MEFs. **F:** RNA Pol II S2p ChIP and mapping along the *RARβ* gene by qPCR in DMSO-treated MEFs. % input is normalized against an external control (chr2:75'408'932-75'409'164); n=2; t-tests corrected for multiple comparisons with the Holm-Sidak method; error bars are standard deviations. Mapping positions: 1: promoter CGI in exon 1; 2: intron 1, 15 kb downstream of TSS; 3: exon 2, 25 kb downstream of TSS, 3'UTR of one isoform; 4: intron 3, high CpG content (45%), 70 kb downstream of TSS; 5: exon 4, 120 kb downstream of TSS; 6: exon 6, 5 kb upstream of 3'UTR of main isoform; 7: just after exon 8, 3'UTR.

Similarly, the level of S5p was also quite low overall and the treatment with BRD4-inhibitor (JQ1) did not lead to different phosphorylation levels in *Tdg[cat]* and

Tdg[null] backgrounds (**Figures 3c and 3d**). Surprisingly, the enrichment was most pronounced towards the 3' end of the gene (Map 6) and not in the promoter region as expected. Analysis of S2 phosphorylation levels of CTD along the *RARβ* locus showed a clear effect of BRD4 inhibition (**Figures 3e and 3f**): the enrichment of S2p was relatively small along the gene body (**Supplementary Figure 1**), but again especially increased at positions 6 and 7. The effect of JQ1 treatment was beneficial for S2 phosphorylation of the RNA Pol II CTD in *Tdg*[wt] but inhibitory in *Tdg*[cat] and *Tdg*[null]. This was significant for *Tdg*[wt] ($p = 0.004$) and *Tdg*[cat] ($p = 0.002$) and interestingly, the trends observed for S2p are perfectly contrary to the trends observed for CTD enrichment and gene expression (**Figures 3a and 3b, Figure 1**). It is interesting that I observed more S2p, which is the mark for productive transcriptional elongation, but less CTD and gene expression in *Tdg*[wt] compared to *Tdg*[cat]. This might indicate that the RNA Pol II requires both functional TDG as well as RA to facilitate gene expression.

Next, I wanted to assess the effect of BRD4 inhibition by JQ1 on the association of TDG to the *RARβ* locus by TDG-ChIP-qPCR. Little enrichment was found in *Tdg*[wt] and *Tdg*[null] independent of the treatment (JQ1 or DMSO). However, BRD4 inhibition increased TDG association in *Tdg*[cat] markedly (**Figure 4a**). This may be explained by the lack of turn-over of catalytically inactive TDG and hence more efficient cross-linking and detection by ChIP (Wirz et al, manuscript in preparation, 2014). I further checked for CDK9 enrichment at the *RARβ* locus, because CDK9 is another S2 kinase and part of P-TEFb, which is recruited by BRD4. In untreated MEFs, I observed little enrichment, possibly because CDK9 levels are generally low in MEFs (Devaiah et al., 2012). There may be a trend for more CDK9 enrichment in a TDG-deficient background (*Tdg*[null]), especially towards the 3' end of the *RARβ* locus (**Figure 4b**). An explanation for this could be that in the absence of TDG, BRD4 is not recruited to the chromatin and thus cannot act as a S2 kinase. Thus, CDK9 may take over, especially at the 3' end of the gene, where S5 of the CTD is markedly decreased, which is the favorite substrate for CDK9 (Devaiah et al., 2012; Tietjen et al., 2010).

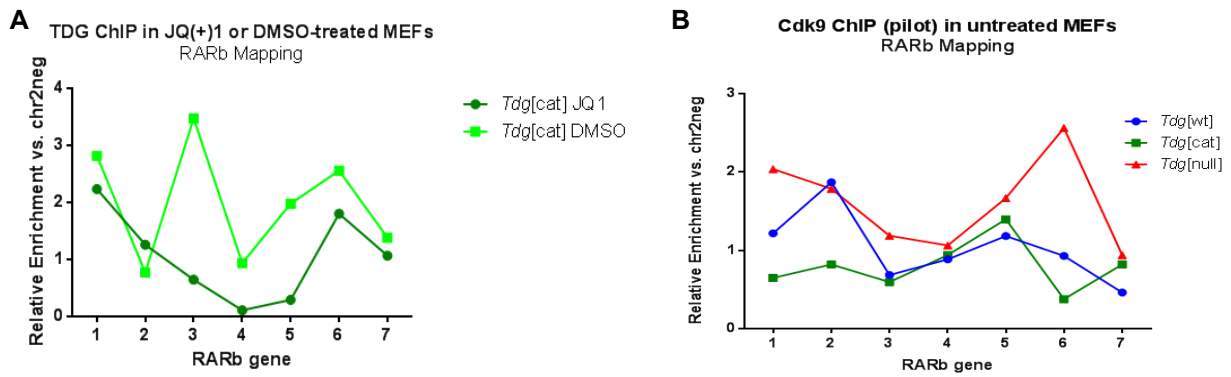


Figure 4: Association of TDG to the *RARβ* gene upon JQ1-treatment and association of the CDK9 serine 2 kinase to the *RARβ* gene in untreated MEFs. **A:** TDG ChIP-qPCR in JQ1- or DMSO-treated MEFs, mapping along the *RARβ* locus. **B:** CDK9 ChIP-qPCR in untreated MEFs, mapping along the *RARβ* locus. The % input is normalized against an external control (chr2:75'408'932-75'409'164) in **4a** and **4b**. Single experiments are shown. Mapping positions: 1: promoter CGI in exon 1; 2: intron 1, 15 kb downstream of TSS; 3: exon 2, 25 kb downstream of TSS, 3'UTR of one isoform; 4: intron 3, high CpG content (45%), 70 kb downstream of TSS; 5: exon 4, 120 kb downstream of TSS; 6: exon 6, 5 kb upstream of 3'UTR of main isoform; 7: just after exon 8, 3'UTR.

Taken together, I could show that RA-treatment increases the association of unphosphorylated CTD to the promoter region of *RARβ*. In this context, functional TDG seems to play an inhibitory role. Additionally, I found significantly increased levels of S5p of the RNA Pol II CTD in MEFs expressing catalytically inactive TDG, not only in the promoter region (position 1), but towards the 3' end of the gene (positions 6 and 7). This could indicate that alternative splicing takes place and is misregulated in the absence of TDG catalytic activity. TDG catalytic activity seems to be important for early elongation, where TDG might be in a complex with BRD4, which was suggested to be the initial S2 kinase (Devaiah et al., 2012). Upon productive transcriptional elongation, however, the absence of TDG is beneficial, which is also mirrored in the most significantly elevated gene expression induction upon RA-treatment. The loss of BRD4 chromatin association seems to effect the PIC formation mostly. Additionally, TDG association to the *RARβ* locus is increased, mainly at positions of particular interest; located in the promoter CGI, in the 3'UTR of a splice variant and near the 3'UTR of the main splice variant, again indicating that TDG might be involved in co-transcriptional processes like mRNA end processing or splicing.

I conclude that in the absence of TDG protein or activity, normal transcription at the *RARβ* locus is disturbed, depending to some extent on the functions of BRD4. This may provide an entry point towards investigation of the mechanistic links between DNA demethylation and the control of RNA Pol II-mediated transcription.

Material and Methods

Cell culture and treatment of cells

SV40-immortalized MEF cell lines were previously described (Kunz et al., 2009a) and cultivated in growth medium (Dulbecco's Modified Eagle's Medium (DMEM), containing 10% Fetal Bovine Serum (FBS), 2 mM L-glutamine and 1x Penicillin / Streptomycin) at 37 °C in a humidified atmosphere containing 5% CO₂. Before starting retinoic-acid or JQ(+)-1 inhibitor (JQ1) treatment, MEFs were grown for two passages. Treatment with all-trans retinoic acid (RA) was conducted at 1 μM for 24h. Treatment with JQ(+)-1 inhibitor (BioVision) was conducted at 250nM for 24h. DMSO mock controls were included for both treatments. After 24h, cells were harvested to perform downstream experiments. All cell culture components were from Sigma if not denoted differently.

Chromatin immuno-precipitation (ChIP) and qPCR analysis

Chromatin was prepared and the immunoprecipitation was performed as described in **Appendix I**. The only adaptation was the addition of phosphatase inhibitors (5mM NaF and 2mM Na₃VO₄) to all three lysis buffers and the dilution buffer for S5p and S2p ChIPs. Quantitative PCR (qPCR) analysis was performed on a Rotor-Gene 3000 thermocycler (Qiagen) using 1x QuantiTect SYBR Green (Qiagen). Antibodies and primers used are shown in **Table 1** and **Table 2**, respectively. Statistical analysis was performed with GraphPad Prism 6, by *t*-tests corrected for multiple comparisons by Holm-Sidak.

RNA extraction, reverse transcription and expression analysis

RNA was extracted with the TRI Reagent (Sigma) according to the manufacturer's protocol and as described in **Appendix I**. RNA was then subjected to DNase I digestion (Fermentas) and reverse transcribed with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems), as indicated by the manufacturers. qPCR

analysis was performed using 1x QuantiTect SYBR Green (Qiagen) with a Rotor-Gene 3000 thermocycler (Qiagen), primers are shown in **Table 2**. Statistical analysis was performed on independent duplicate or triplicate experiments by non-paired, two-tailed *t*-tests.

Table 1: Antibodies used in this study

Antibody	Product number	Manufacturer
Anti-RNA polymerase II CTD repeat YSPTSPS [8WG16] antibody	ab817	Abcam
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody	ab5131	Abcam
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) antibody	ab5095	Abcam
Anti-Cdk9 antibody (C-20)	sc-484	Santa Cruz Biotechnology

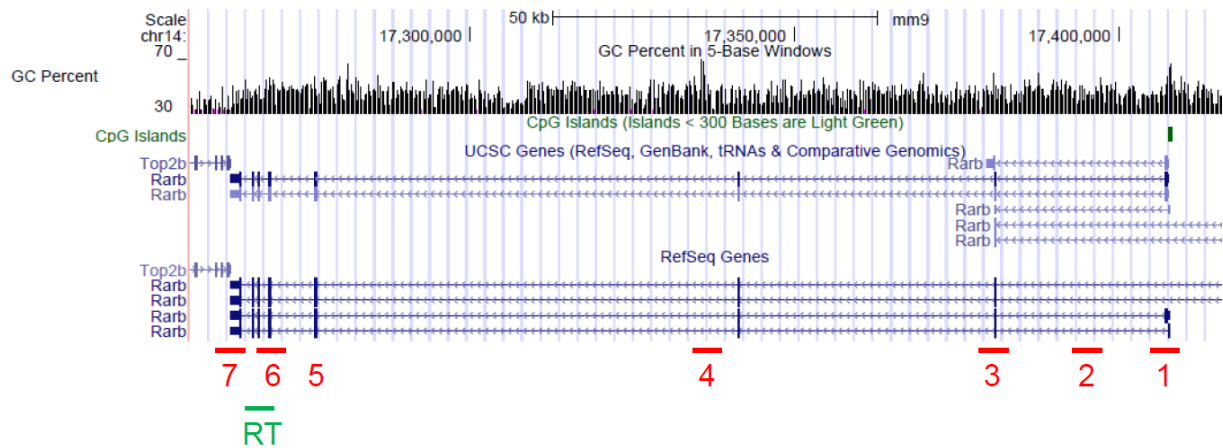
The anti-TDG antibody used for CHIP was produced and affinity purified in our lab, for further information see (Cortazar et al., 2011; Hardeland et al., 2002; Neddermann et al., 1996).

Table 2: Primers used in this study

Primer	Forward (5'-3' sequence)	Reverse (5'-3' sequence)
RAR β -1	GAT CCC AAG TTC TCC CTT CC	GGC AGG AGG GTC TAT TCT TTG
RAR β -2	TTG CCA GGT AGT CAG GAA GG	CCC ACT GCA ACA GCC TAG AG
RAR β -3	CTG ACG CCA TAG TGG TAG CC	TTG TAG CCA TCG AGA CAC AGA
RAR β -4	GCC ACA TAA AGA GGA GGA GGA	TTG ATG AAG GCT ACC CCT TG
RAR β -5	CAG TGG CTC TTA CCG TGG TG	GCT GTT AGG AAT GAC AGG AAC A
RAR β -6	GGG TCT GGT GAC GGT TTC TA	TGA TGG CCT TAC ACT AAA TCG
RAR β -7	CTG GGC AAC ATA AGG GAA AG	AAG ACA CCG AGG TTG TGG AG
chr2neg	AGCACAGCCTGAAGCCTCTA	AGAGGGCATTTCGGTCTTTT
RAR β (RT)	TTAATCTGTGGAGACCGCCAG	TTACACGTTCCGGCACCTTTTCG
Actin β (RT)	CGT CGA CAA CGG CTC CGG CAT	CCA CCA TCA CAC CCT GGT GCC TAG G
B2m (RT)	TCA CGC CAC CCA CCG GAG AA	TCT CGA TCC CAG TAG ACG GTC TTG G

All primers were from Microsynth, Switzerland. RT: reverse transcription.

Supplementary Information



Supplementary Figure 1: Schematic summary of the *RARβ* gene from the Genome Browser (<http://genome-euro.ucsc.edu/index.html>). UCSC (top) and RefSeq (bottom) Genes are depicted, as well as CGIs (in green). Primer positions are indicated in red and numbered as shown in Figures 2, 3 and 4 (cf. **Table 2**). RT primer positions are indicated below (in green, RT).

Contributions: I designed and performed all the experiments described in this section.

5. Concluding Discussion and Outlook

Although the initial discovery of TDG as a mismatch-specific DNA glycosylase implicated a function in DNA repair, it emerged quickly that TDG contributes to more biological processes beyond canonical DNA repair. TDG has been linked physically and functionally to transcriptional activation and, although less frequently, to transcriptional repression, findings that are underlined by the manifold interactions of TDG with different transcription factors. In further support of a link to transcription is the fact that TDG was found to associate with RNA and an RNA helicase (p68), which was proposed to enhance its catalytic activity (Boland and Christman, 2008; Fremont et al., 1997; Gallais et al., 2007).

Other lines of investigation established that TDG participates in the control of DNA methylation. TDG was demonstrated to interact physically and functionally with both *de novo* DNA methyltransferases DNMT3a and DNMT3b (Boland and Christman, 2008; Li et al., 2007). Early evidence indicated that TDG plays a role in direct DNA demethylation, which was by virtue of its ability to excise 5-mC from DNA, a finding that was highly controversial (Jost, 1993; Jost et al., 1995). Instead, it turns out, that TDG accomplishes the last step in an enzymatic cascade for active DNA demethylation, which requires precedent catalytic activity of the TET proteins (He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011; Tahiliani et al., 2009). These findings, together with the fact that the loss of TDG leads to aberrations in DNA and histone modifications, clearly place TDG in the context of epigenome regulation and maintenance.

The focus of my work lay on the characterization of the epigenetic functions of TDG. During my PhD thesis, the mechanistic involvement of TDG in DNA demethylation, downstream of the action of the TET or AID proteins, was clearly demonstrated. The main part of my work focused on the investigation of TDG's role in DNA methylation control in the context of epigenetic programming in stem cell differentiation, its functional interactions and coordination with TET proteins and the resulting impact on gene regulation (**Appendices I and II**).

In a mouse model, we were able to characterize the *Tdg*-knockout phenotype. Intriguingly, we found that loss of *Tdg* causes embryonic lethality which is unlike the

phenotype of any other DNA glycosylase knockout. We could rule out TDG's DNA repair function to be causative for this severe phenotype and, instead, established aberrant epigenetic modulation as a reason: at promoter CGIs of developmental genes, we found a shift towards a more repressed state upon loss of TDG – both in terms of histone marks as well as DNA methylation. Since this was only found in differentiated but not in pluripotent cells, we concluded that TDG safeguards the epigenetic state of developmental genes upon cell lineage commitment. Importantly, we found a TDG-dependent association of both downstream BER proteins as well as histone modifying proteins with these developmental genes, indicating that TDG triggers the regulation of the epigenetic state of gene promoters. On the basis of these data, we proposed that TDG-dependent DNA repair has evolved to provide epigenetic stability during cell lineage commitment (**Appendix III**).

We continued to investigate TDG's role in DNA methylation control by genome-wide approaches and discovered that TDG-deficient mouse embryonic stem cells accumulated differentially methylated regions (DMRs) in the course of differentiation to neuronal progenitor cells. Again, we observed a peculiarity with respect to CGIs, namely, the DMRs located in CGIs were almost exclusively hypomethylated in TDG-deficient cells. This implicated that CGIs are subject to DNA methylation and demethylation, challenging the prevailing view that CGIs are *bona fide* unmethylated. Upon further characterization of the methylation state of these CGI DMRs, we indeed found evidence for ongoing DNA methylation – oxidative demethylation cycles at these loci. *Nota bene*, this cycle was disturbed in the case of catalytically inactive TDG, as demonstrated by the aberrant accumulation of DNA demethylation intermediates (5-fC and 5-caC). As the CGIs affected were sequences that become *de novo* methylated during cell differentiation, this pointed at the necessity for TDG-mediated DNA repair and/or demethylation activity for developmental programming of gene expression. Consistently, these CGI DMRs showed a significant enrichment of genomic features of distal gene regulatory regions. These results postulate a novel mechanism for the establishment and maintenance of gene expression patterns based on the epigenetic regulation of enhancer elements by a DNMT-TET-TDG-mediated DNA methylation and oxidative demethylation. As there is an ongoing debate about a possible deamination-based DNA demethylation pathway, we analysed the occurrence of mutations at CGI DMRs in TDG-deficient and –proficient cells. Since the implicated AID- or APOBEC-mediated deamination of C or 5-mC

would give rise to premutagenic lesions (T•G or U•G, respectively), which are substrates for TDG, we predicted to find an elevated C→T mutation frequency in a background without or with inactive TDG. Since this was not the case, as measured with hairpin Na-sulfite deep-sequencing, we concluded that a deamination-based DNA demethylation pathway plays no or only a minor role in the demethylation events observed at the CGI DMRs investigated. A highly dynamic alteration of DNA methylation at transcriptionally stimulated promoters as well as other gene regulatory elements has been described before and linked to the orchestrated action of DNMTs, TDG and BER (Kangaspeska et al., 2008; Metivier et al., 2008) as well as the TET proteins and TDG (Shen et al., 2013; Song et al., 2013). Our findings additionally link the aberrant DNA methylation patterns caused by the loss of TDG to a biological phenotype, namely the failure to undergo lineage commitment, as evidenced by increased cell death in *Tdg* knockout cells subjected to retinoic acid-induced neuronal differentiation (**Appendix I**).

To further disentangle the mechanisms underlying the *Tdg* knockout phenotype, I analyzed the genome-wide association of TDG to chromatin in mESCs subjected to 24 hours of RA-treatment. Catalytically inactive TDG, which binds but does not remove substrate bases from DNA and, hence, shows a severely reduced DNA dissociation, localized to the genome in roughly 28'000 peaks that showed a higher than 2 fold enrichment ($\log_2 > 1$). TDG peaks showed a bimodal distribution; a narrow peak was encompassing the TSS and a broader peak was located roughly 1 to 1000 kb away from TSS. Further, TDG peaks were characterized by an intermediate to high CpG density, showing a clear enrichment of CGIs. Moreover, TDG peaks were mainly positive for activating histone marks (H3K36me3, H3K4me3, H3K9ac) as well as histone marks characteristic for active enhancers (H3K4me1, H3K27ac), supporting the notion that TDG is targeted to gene regulatory sites to regulate DNA methylation and thereby the epigenetic state. It will be interesting to clarify whether TDG facilitates transcription by the maintenance of epigenetic plasticity at gene regulatory elements only, or whether TDG is actively involved in facilitating and establishing enhancer-promoter loops and thereby the assembly of the transcription machinery. In support of the latter, we found a clear correlation of TDG association to genomic loci with occurrence of the dynamic histone variants H3.3 and H2A.Z. These histone variants have been linked to actively transcribed genes, promoters and enhancers and, particularly at the latter, were found to rapidly turn over (Kraushaar et

al., 2013). We further found a striking enrichment of both TET1 and TET2 occupancy in TDG peaks and, interestingly, the triple-positive TET1-TET2-TDG peaks showed a significant preference for CGIs, whereas the unique TDG peaks did not. Contrarily, unique TDG peaks were preferentially located in active enhancers, indicating a functional separation between the distinct TDG protein fractions. In line with this notion is the finding that triple-positive peaks showed a drastic increase in DNA demethylation intermediates as well as bivalent chromatin marks, in contrast to the unique TDG peaks. This suggests that the orchestrated, TET1-TET2-TDG-mediated DNA demethylation occurs mostly at developmental gene promoters, characterized by a high CpG density (CGIs). The increased co-occupancy with dynamic histone variants at triple-positive sites is in further support of this; enhanced nucleosome turnover might be promoted by TDG-mediated single-strand breaks generated in the context of excision repair-mediated active DNA demethylation. This is an attractive hypothesis, supported by an increased sensitivity to PARP inhibitors of differentiated TDG-proficient cells when compared to TDG-deficient cells (Cortazar et al., 2011). The hypersensitivity to PARP1 inhibitor suggests that single-strand breaks accumulate in a TDG-dependent manner whose further processing would require PARP-coordinated single-strand break repair. The above mentioned results are consistent with a mechanistic model where TDG is recruited to gene regulatory regions together with the TET proteins to perform DNA demethylation by 5-mC oxidation and excision. This will generate a DNA single-strand break, which facilitates nucleosome eviction or turnover and thus gene expression. The preferential co-occupancy of TET1-TET2-TDG with bivalent CGIs indicates that this may occur predominantly in the context of developmental gene activation (**Appendix II**).

TDG was shown to physically and functionally interact with the retinoic acid receptor (RAR/RXR; (Um et al., 1998)), in addition to several other nuclear receptors and transcription factors. I investigated the assembly and the progression of the transcription machinery as a function of TDG and BRD4, the latter having been identified as a TDG interacting protein in our laboratory and, thus, possibly providing a direct link between TDG and the RNA Polymerase II (data not shown). In a series of pilot experiments, I analyzed the phosphorylation state of RNA Pol II's C-terminal domain (CTD) along the *RARβ* locus upon retinoic acid (RA)-induced transcription as well as upon BRD4 inhibition in cells expressing wildtype, catalytically inactive or no TDG. Interestingly, I observed elevated levels of pre-initiation complex (PIC)

formation upon RA-induced transcription in the absence of functional TDG. Additionally, phosphorylation of CTD's Serine 5 (early elongation) and Serine 2 (productive elongation) were increased in cells expressing catalytically inactive or no TDG, respectively. It is not clear at this point whether this somewhat unexpected finding reflects indeed an overall increase of productive transcriptional elongation in the absence of TDG, or whether it represents an increase of stalled or paused RNA Polymerase II, which is more readily cross-linked and detected by ChIP-qPCR. The inhibition of the putative bridging factor BRD4 generally reduced transcriptional activity, reflected by low mRNA expression levels, low amounts of PIC formation and low presence of early elongation states. However, inhibition of BRD4 increased productive transcriptional elongation in a TDG wildtype background, raising again the question whether this really reflects increased transcriptional levels or rather an increased stalling of the RNA Polymerase II. When investigating the association of TDG to the *RARβ* locus, I found an increase in TDG association upon BRD4 inhibition; the highest enrichment of TDG was observed in the 3' UTRs of different splice variants of *RARβ* (besides the promoter CGI), indicating a possible involvement of TDG in co-transcriptional processes, like mRNA end processing or splicing. These data strengthen a direct role of TDG in transcriptional regulation, possibly by means of modulating PIC assembly, productive transcriptional elongation and/or co-transcriptional processes (**Supplementary Results 4.4**). Further studies will be needed to dissect the effective role of TDG herein and to specify the engagement of TET-TDG in transcription initiation.

Taken together, I was able to further specify and provide novel insight into TDG's role in the epigenetic programming and maintenance of gene regulatory elements during cell lineage commitment. Together with the TET proteins, TDG participates in targeted DNA methylation and demethylation at gene promoters and enhancers, which is tightly associated with the maintenance of a dynamic chromatin structure, required for the setup of developmental stage-specific gene expression programs. Hopefully, this work constitutes a basis for answering many questions that still remain open. For instance, what is the mechanistic relationship between TET-TDG-mediated cyclic DNA demethylation, and how are the deposition of H3.3 and H2A.Z and the depletion of nucleosomes orchestrated at such sites? And how is TET-TDG-mediated 5-mC oxidation and excision regulated such that an accumulation of DNA strand breaks and genomic instability can be avoided? Since the TET proteins and TDG co-

localize at many genomic loci, further questions to be addressed are: what are the underlying targeting mechanisms and is TET-TDG co-occupancy simultaneous or successive? Moreover, TET proteins and TDG cooperate in cycles of DNA methylation and demethylation at sites determined by cell differentiation stage and cell type. How is the exit from cycling regulated such that one region can end up terminally methylated and another terminally unmethylated once a cell fate is determined? Finally, TET proteins have been assigned to different developmental stages but have catalytically redundant roles – so are there differences in the recruitment efficiency and activity of TDG, depending on the different TET proteins?

6. References

- Aapola, U., Kawasaki, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Minoshima, S., Krohn, K., Antonarakis, S.E., *et al.* (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* *65*, 293-298.
- Acevedo, M.L., and Kraus, W.L. (2003). Mediator and p300/CBP-steroid receptor coactivator complexes have distinct roles, but function synergistically, during estrogen receptor alpha-dependent transcription with chromatin templates. *Mol Cell Biol* *23*, 335-348.
- Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* *449*, 731-734.
- Alabert, C., and Groth, A. (2012). Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol* *13*, 153-167.
- Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* *42*, 599-610.
- Aloia, L., Di Stefano, B., and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. *Development* *140*, 2525-2534.
- Ansari, S.A., and Morse, R.H. (2013). Mechanisms of Mediator complex action in transcriptional activation. *Cell Mol Life Sci* *70*, 2743-2756.
- Baer, B.W., and Rhodes, D. (1983). Eukaryotic RNA polymerase II binds to nucleosome cores from transcribed genes. *Nature* *301*, 482-488.
- Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsasser, S.J., Chapgier, A., Goldberg, A.D., Canaani, E., *et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. *Cell* *155*, 107-120.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* *21*, 381-395.
- Barrett, T.E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L.H. (1998). Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* *92*, 117-129.
- Barrett, T.E., Scharer, O.D., Savva, R., Brown, T., Jiricny, J., Verdine, G.L., and Pearl, L.H. (1999). Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J* *18*, 6599-6609.
- Barron-Casella, E., and Corden, J.L. (1992). Conservation of the mammalian RNA polymerase II largest-subunit C-terminal domain. *J Mol Evol* *35*, 405-410.
- Basu, R., and Zhang, L.F. (2011). X chromosome inactivation: a silence that needs to be broken. *Genesis* *49*, 821-834.
- Bataille, A.R., Jeronimo, C., Jacques, P.E., Laramée, L., Fortin, M.E., Forest, A., Bergeron, M., Hanes, S.D., and Robert, F. (2012). A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* *45*, 158-170.
- Batsche, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* *13*, 22-29.
- Baubec, T., and Schubeler, D. (2014). Genomic patterns and context specific interpretation of DNA methylation. *Curr Opin Genet Dev* *25C*, 85-92.
- Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., and Herman, J.G. (2001). Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet* *10*, 687-692.
- Beckman, K.B., and Ames, B.N. (1997). Oxidative decay of DNA. *J Biol Chem* *272*, 19633-19636.
- Berdasco, M., and Esteller, M. (2010). Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell* *19*, 698-711.

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

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

Bird, A.W., Yu, D.Y., Pray-Grant, M.G., Qiu, Q., Harmon, K.E., Megee, P.C., Grant, P.A., Smith, M.M., and Christman, M.F. (2002). Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature* *419*, 411-415.

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

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., *et al.* (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* *500*, 222-226.

Blattler, A., and Farnham, P.J. (2013). Cross-talk between site-specific transcription factors and DNA methylation states. *J Biol Chem* *288*, 34287-34294.

Boland, M.J., and Christman, J.K. (2008). Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *J Mol Biol* *379*, 492-504.

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

Bouchard, V.J., Rouleau, M., and Poirier, G.G. (2003). PARP-1, a determinant of cell survival in response to DNA damage. *Exp Hematol* *31*, 446-454.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature* *371*, 435-438.

Brenner, C., Deplus, R., Didelot, C., Lorient, A., Vire, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., *et al.* (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J* *24*, 336-346.

Brockdorff, N. (2002). X-chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends Genet* *18*, 352-358.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., *et al.* (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* *368*, 258-261.

Bruner, S.D., Norman, D.P., and Verdine, G.L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* *403*, 859-866.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Mol Cell* *36*, 541-546.

Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* *100*, 6969-6973.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* *298*, 1039-1043.

Cartron, P.F., Nadaradjane, A., Lepape, F., Lalier, L., Gardie, B., and Vallette, F.M. (2013). Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer* *4*, 235-241.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* *10*, 295-304.

Chase, A., and Cross, N.C. (2011). Aberrations of EZH2 in cancer. *Clin Cancer Res* *17*, 2613-2618.

Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A* *99*, 16916-16921.

Cheedipudi, S., Genolet, O., and Dobrova, G. (2014). Epigenetic inheritance of cell fates during embryonic development. *Front Genet* *5*, 19.

Chen, C.C., Wang, K.Y., and Shen, C.K. (2012). The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem* **287**, 33116-33121.

Chen, D., Lucey, M.J., Phoenix, F., Lopez-Garcia, J., Hart, S.M., Losson, R., Buluwela, L., Coombes, R.C., Chambon, P., Schar, P., *et al.* (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *J Biol Chem* **278**, 38586-38592.

Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X., *et al.* (2013a). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes Dev* **27**, 2109-2124.

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

Cheung, H.H., Lee, T.L., Rennert, O.M., and Chan, W.Y. (2009). DNA methylation of cancer genome. *Birth Defects Res C Embryo Today* **87**, 335-350.

Choi, J.D., and Lee, J.S. (2013). Interplay between Epigenetics and Genetics in Cancer. *Genomics Inform* **11**, 164-173.

Choi, J.Y., Lim, S., Kim, E.J., Jo, A., and Guengerich, F.P. (2010). Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases alpha, delta, eta, iota, kappa, and REV1. *J Mol Biol* **404**, 34-44.

Ciccio, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. *Mol Cell* **40**, 179-204.

Citterio, E., Van Den Boom, V., Schnitzler, G., Kanaar, R., Bonte, E., Kingston, R.E., Hoeijmakers, J.H., and Vermeulen, W. (2000). ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol Cell Biol* **20**, 7643-7653.

Compe, E., and Egly, J.M. (2012). TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol* **13**, 343-354.

Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A* **82**, 7934-7938.

Corpet, A., and Almouzni, G. (2009). Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends Cell Biol* **19**, 29-41.

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

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

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

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

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., and Marians, K.J. (2000). The importance of repairing stalled replication forks. *Nature* **404**, 37-41.

Craig, J.M. (2005). Heterochromatin--many flavours, common themes. *Bioessays* **27**, 17-28.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936.

Czermin, B., Schotta, G., Hulsman, B.B., Brehm, A., Becker, P.B., Reuter, G., and Imhof, A. (2001). Physical and functional association of SU(VAR)3-9 and HDAC1 in *Drosophila*. *EMBO Rep* **2**, 915-919.

Darzacq, X., Shav-Tal, Y., de Turriz, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14**, 796-806.

Datta, A., Bagchi, S., Nag, A., Shiyonov, P., Adami, G.R., Yoon, T., and Raychaudhuri, P. (2001). The p48 subunit of the damaged-DNA binding protein DDB associates with the CBP/p300 family of histone acetyltransferase. *Mutat Res* 486, 89-97.

Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell* 10, 409-415.

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

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

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev* 25, 1010-1022.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J., *et al.* (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 146, 904-917.

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

Devaiah, B.N., Lewis, B.A., Cherman, N., Hewitt, M.C., Albrecht, B.K., Robey, P.G., Ozato, K., Sims, R.J., 3rd, and Singer, D.S. (2012). BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A* 109, 6927-6932.

Devaiah, B.N., and Singer, D.S. (2012). Cross-talk among RNA polymerase II kinases modulates C-terminal domain phosphorylation. *J Biol Chem* 287, 38755-38766.

Dianov, G., and Lindahl, T. (1994). Reconstitution of the DNA base excision-repair pathway. *Curr Biol* 4, 1069-1076.

Dos Santos, R.L., Tosti, L., Radzsheuskaya, A., Caballero, I.M., Kaji, K., Hendrich, B., and Silva, J.C. (2014). MBD3/NuRD Facilitates Induction of Pluripotency in a Context-Dependent Manner. *Cell Stem Cell*.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* 13, 713-719.

Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* 24, 1253-1265.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 28, 817-825.

Ernst, P., Wang, J., Huang, M., Goodman, R.H., and Korsmeyer, S.J. (2001). MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. *Mol Cell Biol* 21, 2249-2258.

Esteller, M., Levine, R., Baylin, S.B., Ellenson, L.H., and Herman, J.G. (1998). MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene* 17, 2413-2417.

Esteve, P.O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev* 20, 3089-3103.

Evans-Galea, M.V., Hannan, A.J., Carroddus, N., Delatycki, M.B., and Saffery, R. (2013). Epigenetic modifications in trinucleotide repeat diseases. *Trends Mol Med* 19, 655-663.

Fan, S., Zhang, M.Q., and Zhang, X. (2008). Histone methylation marks play important roles in predicting the methylation status of CpG islands. *Biochem Biophys Res Commun* 374, 559-564.

Fowden, A.L., and Moore, T. (2012). Maternal-fetal resource allocation: co-operation and conflict. *Placenta* 33 Suppl 2, e11-15.

Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A., and Cedar, H. (1991). Demethylation of CpG islands in embryonic cells. *Nature* *351*, 239-241.

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

Fremont, M., Siegmann, M., Gaulis, S., Matthies, R., Hess, D., and Jost, J.P. (1997). Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA. *Nucleic Acids Res* *25*, 2375-2380.

Futscher, B.W., Oshiro, M.M., Wozniak, R.J., Holtan, N., Hanigan, C.L., Duan, H., and Domann, F.E. (2002). Role for DNA methylation in the control of cell type specific maspin expression. *Nat Genet* *31*, 175-179.

Gaillard, P.H., Martini, E.M., Kaufman, P.D., Stillman, B., Moustacchi, E., and Almouzni, G. (1996). Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. *Cell* *86*, 887-896.

Gallais, R., Demay, F., Barath, P., Finot, L., Jurkowska, R., Le Guevel, R., Gay, F., Jeltsch, A., Metivier, R., and Salbert, G. (2007). Deoxyribonucleic acid methyl transferases 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation. *Mol Endocrinol* *21*, 2085-2098.

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

Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A., and Xu, G.L. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J Biol Chem* *279*, 25447-25454.

Gebara, M.M., Sayre, M.H., and Corden, J.L. (1997). Phosphorylation of the carboxy-terminal repeat domain in RNA polymerase II by cyclin-dependent kinases is sufficient to inhibit transcription. *J Cell Biochem* *64*, 390-402.

Ghosh, A., and Bansal, M. (2003). A glossary of DNA structures from A to Z. *Acta Crystallogr D Biol Crystallogr* *59*, 620-626.

Girardot, M., Feil, R., and Lleres, D. (2013). Epigenetic deregulation of genomic imprinting in humans: causal mechanisms and clinical implications. *Epigenomics* *5*, 715-728.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., *et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* *140*, 678-691.

Gommers-Ampt, J.H., Van Leeuwen, F., de Beer, A.L., Vliegenthart, J.F., Dizdaroglu, M., Kowalak, J.A., Crain, P.F., and Borst, P. (1993). beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* *75*, 1129-1136.

Gonzalez, M.E., DuPrie, M.L., Krueger, H., Merajver, S.D., Ventura, A.C., Toy, K.A., and Kleer, C.G. (2011). Histone methyltransferase EZH2 induces Akt-dependent genomic instability and BRCA1 inhibition in breast cancer. *Cancer Res* *71*, 2360-2370.

Gowher, H., and Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J Mol Biol* *309*, 1201-1208.

Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. (2005). Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J Biol Chem* *280*, 13341-13348.

Goyal, R., Reinhardt, R., and Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic Acids Res* *34*, 1182-1188.

Green, C.M., and Almouzni, G. (2003). Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. *EMBO J* *22*, 5163-5174.

Grunberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* *38*, 603-611.

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

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* *130*, 77-88.

Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G.A., Feber, A., Teschendorff, A.E., Gomez, V., Hergovich, A., Tirabosco, R., Fernanda Amary, M., *et al.* (2013). Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nat Commun* 4, 2166.

Gunz, D., Hess, M.T., and Naegeli, H. (1996). Recognition of DNA adducts by human nucleotide excision repair. Evidence for a thermodynamic probing mechanism. *J Biol Chem* 271, 25089-25098.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell Cycle* 10, 2662-2668.

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

Haffner, M.C., Chaux, A., Meeker, A.K., Esopi, D.M., Gerber, J., Pellakuru, L.G., Toubaji, A., Argani, P., Iacobuzio-Donahue, C., Nelson, W.G., *et al.* (2011). Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* 2, 627-637.

Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem* 275, 33449-33456.

Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2003). The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic Acids Res* 31, 2261-2271.

Hardeland, U., Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J., and Schar, P. (2001). Thymine DNA glycosylase. *Prog Nucleic Acid Res Mol Biol* 68, 235-253.

Hardeland, U., Kunz, C., Focke, F., Szadkowski, M., and Schar, P. (2007). Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic Acids Res* 35, 3859-3867.

Hardeland, U., Steinacher, R., Jiricny, J., and Schar, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J* 21, 1456-1464.

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* 21, 396-420.

Hasan, S., Stucki, M., Hassa, P.O., Imhof, R., Gehrig, P., Hunziker, P., Hubscher, U., and Hottiger, M.O. (2001). Regulation of human flap endonuclease-1 activity by acetylation through the transcriptional coactivator p300. *Mol Cell* 7, 1221-1231.

Hashimoto, H., Vertino, P.M., and Cheng, X. (2010). Molecular coupling of DNA methylation and histone methylation. *Epigenomics* 2, 657-669.

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

Heidemann, M., Hintermair, C., Voss, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* 1829, 55-62.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* 105, 17046-17049.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311-318.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. *Annu Rev Cell Dev Biol* 21, 133-153.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 279, 48350-48359.

Hinoue, T., Weisenberger, D.J., Pan, F., Campan, M., Kim, M., Young, J., Whitehall, V.L., Leggett, B.A., and Laird, P.W. (2009). Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling. *PLoS One* 4, e8357.

Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J Biochem* 141, 601-608.

Hoeijmakers, J.H. (2009). DNA damage, aging, and cancer. *N Engl J Med* 361, 1475-1485.

Horn, P.J., and Peterson, C.L. (2002). Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science* 297, 1824-1827.

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P., and Xu, Y. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* 155, 1545-1555.

Hu, X., Zhang, L., Mao, S.Q., Li, Z., Chen, J., Zhang, R.R., Wu, H.P., Gao, J., Guo, F., Liu, W., *et al.* (2014). Tet and TDG Mediate DNA Demethylation Essential for Mesenchymal-to-Epithelial Transition in Somatic Cell Reprogramming. *Cell Stem Cell*.

Huang, C., and Zhu, B. (2014). H3.3 turnover: A mechanism to poise chromatin for transcription, or a response to open chromatin? *Bioessays*.

Ikeda, Y., and Kinoshita, T. (2009). DNA demethylation: a lesson from the garden. *Chromosoma* 118, 37-41.

Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* 102, 463-473.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. *FEBS Lett* 583, 1713-1720.

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

Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res* 21, 1670-1676.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334, 194.

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

Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., and Yokoyama, A. (2014). TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* 19, 52-65.

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

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

Iyer, L.M., Anantharaman, V., Wolf, M.Y., and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 38, 1-31.

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

Jacob, Y., Bergamin, E., Donoghue, M.T., Mongeon, V., LeBlanc, C., Voigt, P., Underwood, C.J., Brunzelle, J.S., Michaels, S.D., Reinberg, D., *et al.* (2014). Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. *Science* 343, 1249-1253.

Jacobs, A.L., and Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma* 121, 1-20.

Jacobs, E.Y., Ogiwara, I., and Weiner, A.M. (2004). Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. *Mol Cell Biol* 24, 846-855.

Jakovcevski, M., and Akbarian, S. (2012). Epigenetic mechanisms in neurological disease. *Nat Med* 18, 1194-1204.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074-1080.

Jin, J., Cai, Y., Li, B., Conaway, R.C., Workman, J.L., Conaway, J.W., and Kusch, T. (2005). In and out: histone variant exchange in chromatin. *Trends Biochem Sci* 30, 680-687.

Jin, S.G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q., and Pfeifer, G.P. (2011). 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 71, 7360-7365.

Jones, P.A. (2002). DNA methylation and cancer. *Oncogene* 21, 5358-5360.

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3, 415-428.

Jost, J.P. (1993). Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. *Proc Natl Acad Sci U S A* 90, 4684-4688.

Jost, J.P., Siegmund, M., Sun, L., and Leung, R. (1995). Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. *J Biol Chem* 270, 9734-9739.

Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. *Chembiochem* 12, 206-222.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430-435.

Kanduri, C. (2011). Kcnq1ot1: a chromatin regulatory RNA. *Semin Cell Dev Biol* 22, 343-350.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900-903.

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

Kareta, M.S., Botello, Z.M., Ennis, J.J., Chou, C., and Chedin, F. (2006). Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. *J Biol Chem* 281, 25893-25902.

Kellinger, M.W., Song, C.X., Chong, J., Lu, X.Y., He, C., and Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol* 19, 831-833.

Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* 268, 10416-10424.

Kelsey, G., and Bartolomei, M.S. (2012). Imprinted genes ... and the number is? *PLoS Genet* 8, e1002601.

Kim, D.H., Saetrom, P., Snove, O., Jr., and Rossi, J.J. (2008). MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* 105, 16230-16235.

Kim, E.J., and Um, S.J. (2008). Thymine-DNA glycosylase interacts with and functions as a coactivator of p53 family proteins. *Biochem Biophys Res Commun* 377, 838-842.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeá, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517-522.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature* 436, 876-880.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31, 89-97.

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

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci U S A* *108*, 14566-14571.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* *502*, 472-479.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* *41*, 376-381.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* *14*, 2452-2460.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* *128*, 693-705.

Kraushaar, D.C., Jin, W., Maunakea, A., Abraham, B., Ha, M., and Zhao, K. (2013). Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol* *14*, R121.

Kraushaar, D.C., and Zhao, K. (2013). The epigenomics of embryonic stem cell differentiation. *Int J Biol Sci* *9*, 1134-1144.

Krokan, H.E., Drablos, F., and Slupphaug, G. (2002). Uracil in DNA--occurrence, consequences and repair. *Oncogene* *21*, 8935-8948.

Krokan, H.E., Otterlei, M., Nilsen, H., Kavli, B., Skorpen, F., Andersen, S., Skjelbred, C., Akbari, M., Aas, P.A., and Slupphaug, G. (2001). Properties and functions of human uracil-DNA glycosylase from the UNG gene. *Prog Nucleic Acid Res Mol Biol* *68*, 365-386.

Kudo, Y., Tateishi, K., Yamamoto, K., Yamamoto, S., Asaoka, Y., Ijichi, H., Nagae, G., Yoshida, H., Aburatani, H., and Koike, K. (2012). Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci* *103*, 670-676.

Kumar, R., and Rao, D.N. (2013). Role of DNA methyltransferases in epigenetic regulation in bacteria. *Subcell Biochem* *61*, 81-102.

Kunkel, T.A. (1992). DNA replication fidelity. *J Biol Chem* *267*, 18251-18254.

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

Kunz, C., Saito, Y., and Schar, P. (2009b). DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cell Mol Life Sci* *66*, 1021-1038.

Kuo, H.Y., Chang, C.C., Jeng, J.C., Hu, H.M., Lin, D.Y., Maul, G.G., Kwok, R.P., and Shih, H.M. (2005). SUMO modification negatively modulates the transcriptional activity of CREB-binding protein via the recruitment of Daxx. *Proc Natl Acad Sci U S A* *102*, 16973-16978.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* *494*, 497-501.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* *409*, 860-921.

Lau, A.Y., Scharer, O.D., Samson, L., Verdine, G.L., and Ellenberger, T. (1998). Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* *95*, 249-258.

Le May, N., Fradin, D., Iltis, I., Bougneres, P., and Egly, J.M. (2012). XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol Cell* *47*, 622-632.

Le May, N., Mota-Fernandes, D., Velez-Cruz, R., Iltis, I., Biard, D., and Egly, J.M. (2010). NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Mol Cell* *38*, 54-66.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., Nystrom-Lahti, M., *et al.* (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215-1225.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313.

Leger, H., Smet-Nocca, C., Attmane-Elakeb, A., Morley-Fletcher, S., Benecke, A.G., and Eilebrecht, S. (2014). A TDG/CBP/RARalpha ternary complex mediates the retinoic acid-dependent expression of DNA methylation-sensitive genes. *Genomics Proteomics Bioinformatics* **12**, 8-18.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* **13**, 1192-1200.

Levchenko, V., Jackson, B., and Jackson, V. (2005). Histone release during transcription: displacement of the two H2A-H2B dimers in the nucleosome is dependent on different levels of transcription-induced positive stress. *Biochemistry* **44**, 5357-5372.

Levchenko, V., and Jackson, V. (2004). Histone release during transcription: NAP1 forms a complex with H2A and H2B and facilitates a topologically dependent release of H3 and H4 from the nucleosome. *Biochemistry* **43**, 2359-2372.

Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A* **107**, 14075-14080.

Li, B., Howe, L., Anderson, S., Yates, J.R., 3rd, and Workman, J.L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **278**, 8897-8903.

Li, J., and Gilmour, D.S. (2011). Promoter proximal pausing and the control of gene expression. *Curr Opin Genet Dev* **21**, 231-235.

Li, R., Mav, D., Grimm, S.A., Jothi, R., Shah, R., and Wade, P.A. (2014). Fine-tuning of epigenetic regulation with respect to promoter CpG content in a cell type-specific manner. *Epigenetics* **9**.

Li, W. (2011). On parameters of the human genome. *J Theor Biol* **288**, 92-104.

Li, Y.Q., Zhou, P.Z., Zheng, X.D., Walsh, C.P., and Xu, G.L. (2007). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res* **35**, 390-400.

Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118**, 4509-4518.

Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., *et al.* (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* **150**, 1135-1146.

Lindahl, T. (1974). An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc Natl Acad Sci U S A* **71**, 3649-3653.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.

Lindahl, T., and Wood, R.D. (1999). Quality control by DNA repair. *Science* **286**, 1897-1905.

Liu, P., Kenney, J.M., Stiller, J.W., and Greenleaf, A.L. (2010). Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol* **27**, 2628-2641.

Liu, Y., Wang, X., Zhang, J., Huang, H., Ding, B., Wu, J., and Shi, Y. (2008). Structural basis and binding properties of the second bromodomain of Brd4 with acetylated histone tails. *Biochemistry* **47**, 6403-6417.

Liu, Z., and Myers, L.C. (2012). Med5(Nut1) and Med17(Srb4) are direct targets of mediator histone H4 tail interactions. *PLoS One* **7**, e38416.

Lloyd, R.S., and Cheng, X. (1997). Mechanistic link between DNA methyltransferases and DNA repair enzymes by base flipping. *Biopolymers* **44**, 139-151.

Loeb, L.A., and Preston, B.D. (1986). Mutagenesis by apurinic/apyrimidinic sites. *Annu Rev Genet* **20**, 201-230.

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

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

Loesch, D.Z., Sherwell, S., Kinsella, G., Tassone, F., Taylor, A., Amor, D., Sung, S., and Evans, A. (2012). Fragile X-associated tremor/ataxia phenotype in a male carrier of unmethylated full mutation in the FMR1 gene. *Clin Genet* 82, 88-92.

Lohle, M., Hermann, A., Glass, H., Kempe, A., Schwarz, S.C., Kim, J.B., Poulet, C., Ravens, U., Schwarz, J., Scholer, H.R., *et al.* (2012). Differentiation efficiency of induced pluripotent stem cells depends on the number of reprogramming factors. *Stem Cells* 30, 570-579.

Lolli, G. (2009). Binding to DNA of the RNA-polymerase II C-terminal domain allows discrimination between Cdk7 and Cdk9 phosphorylation. *Nucleic Acids Res* 37, 1260-1268.

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

Lucey, M.J., Chen, D., Lopez-Garcia, J., Hart, S.M., Phoenix, F., Al-Jehani, R., Alao, J.P., White, R., Kindle, K.B., Losson, R., *et al.* (2005). T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif. *Nucleic Acids Res* 33, 6393-6404.

Lukas, J., Lukas, C., and Bartek, J. (2011). More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nat Cell Biol* 13, 1161-1169.

Madabushi, A., Hwang, B.J., Jin, J., and Lu, A.L. (2013). Histone deacetylase SIRT1 modulates and deacetylates DNA base excision repair enzyme thymine DNA glycosylase. *Biochem J* 456, 89-98.

Mahaney, B.L., Hammel, M., Meek, K., Tainer, J.A., and Lees-Miller, S.P. (2013). XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair. *Biochem Cell Biol* 91, 31-41.

Mahony, S., Mazzone, E.O., McCuine, S., Young, R.A., Wichterle, H., and Gifford, D.K. (2011). Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol* 12, R2.

Maison, C., and Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* 5, 296-304.

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

Maiti, A., Michelson, A.Z., Armwood, C.J., Lee, J.K., and Drohat, A.C. (2013). Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J Am Chem Soc* 135, 15813-15822.

Maiti, A., Morgan, M.T., Pozharski, E., and Drohat, A.C. (2008). Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc Natl Acad Sci U S A* 105, 8890-8895.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* 23, 1256-1269.

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

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357-361.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A* 101, 1525-1530.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet* 11, 786-799.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* *454*, 766-770.

Menigatti, M., Di Gregorio, C., Borghi, F., Sala, E., Scarselli, A., Pedroni, M., Foroni, M., Benatti, P., Roncucci, L., Ponz de Leon, M., *et al.* (2001). Methylation pattern of different regions of the MLH1 promoter and silencing of gene expression in hereditary and sporadic colorectal cancer. *Genes Chromosomes Cancer* *31*, 357-361.

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

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* *437*, 436-439.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* *448*, 553-560.

Minor, E.A., Court, B.L., Young, J.I., and Wang, G. (2013). Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J Biol Chem* *288*, 13669-13674.

Missero, C., Pirro, M.T., Simeone, S., Pischetola, M., and Di Lauro, R. (2001). The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription. *J Biol Chem* *276*, 33569-33575.

Moggs, J.G., Grandi, P., Quivy, J.P., Jonsson, Z.O., Hubscher, U., Becker, P.B., and Almouzni, G. (2000). A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. *Mol Cell Biol* *20*, 1206-1218.

Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* *137*, 2493-2499.

Mohan, R.D., Litchfield, D.W., Torchia, J., and Tini, M. (2010). Opposing regulatory roles of phosphorylation and acetylation in DNA mismatch processing by thymine DNA glycosylase. *Nucleic Acids Res* *38*, 1135-1148.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* *30*, 755-766.

Mol, C.D., Arvai, A.S., Sanderson, R.J., Slupphaug, G., Kavli, B., Krokan, H.E., Mosbaugh, D.W., and Tainer, J.A. (1995a). Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* *82*, 701-708.

Mol, C.D., Arvai, A.S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H.E., and Tainer, J.A. (1995b). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* *80*, 869-878.

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

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., *et al.* (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* *20*, 11-24.

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

Morozov, V.M., Gavrilova, E.V., Ogryzko, V.V., and Ishov, A.M. (2012). Dualistic function of Daxx at centromeric and pericentromeric heterochromatin in normal and stress conditions. *Nucleus* *3*, 276-285.

Morrison, A.J., and Shen, X. (2005). DNA repair in the context of chromatin. *Cell Cycle* 4, 568-571.

Munoz, P., Iliou, M.S., and Esteller, M. (2012). Epigenetic alterations involved in cancer stem cell reprogramming. *Mol Oncol* 6, 620-636.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* 8, 751-758.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717-1720.

Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 486, 415-419.

Nakayama, T., Nishioka, K., Dong, Y.X., Shimojima, T., and Hirose, S. (2007). Drosophila GAGA factor directs histone H3.3 replacement that prevents the heterochromatin spreading. *Genes Dev* 21, 552-561.

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

Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta* 1809, 34-45.

Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J.J., Wiebauer, K., and Jiricny, J. (1996). Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J Biol Chem* 271, 12767-12774.

Neddermann, P., and Jiricny, J. (1994). Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc Natl Acad Sci U S A* 91, 1642-1646.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. (2013). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol* 14, R91.

Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., and Block, S.M. (2003). Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* 115, 437-447.

Nightingale, K.P., Gendreizig, S., White, D.A., Bradbury, C., Hollfelder, F., and Turner, B.M. (2007). Cross-talk between histone modifications in response to histone deacetylase inhibitors: MLL4 links histone H3 acetylation and histone H3K4 methylation. *J Biol Chem* 282, 4408-4416.

Nilsson, E.E., and Skinner, M.K. (2014). Environmentally induced epigenetic transgenerational inheritance of disease susceptibility. *Transl Res*.

Ninio, J. (1991). Connections between translation, transcription and replication error-rates. *Biochimie* 73, 1517-1523.

Noushmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F., Pelloski, C.E., Sulman, E.P., Bhat, K.P., *et al.* (2010). Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* 17, 510-522.

O'Donovan, P.J., and Livingston, D.M. (2010). BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis* 31, 961-967.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., *et al.* (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* 39, 237-242.

Ohno, R., Nakayama, M., Naruse, C., Okashita, N., Takano, O., Tachibana, M., Asano, M., Saitou, M., and Seki, Y. (2013). A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development* 140, 2892-2903.

Okashita, N., Kumaki, Y., Ebi, K., Nishi, M., Okamoto, Y., Nakayama, M., Hashimoto, S., Nakamura, T., Sugawara, K., Kojima, N., *et al.* (2014). PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development* 141, 269-280.

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

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

Otani, J., Kimura, H., Sharif, J., Endo, T.A., Mishima, Y., Kawakami, T., Koseki, H., Shirakawa, M., Suetake, I., and Tajima, S. (2013). Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PLoS One* *8*, e82961.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* *32*, 232-246.

Papamichos-Chronakis, M., Krebs, J.E., and Peterson, C.L. (2006). Interplay between Ino80 and Swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptation in response to DNA damage. *Genes Dev* *20*, 2437-2449.

Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E., and Tainer, J.A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J* *17*, 5214-5226.

Pascucci, B., Stucki, M., Jonsson, Z.O., Dogliotti, E., and Hubscher, U. (1999). Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases delta and epsilon. *J Biol Chem* *274*, 33696-33702.

Pasini, D., Hansen, K.H., Christensen, J., Agger, K., Cloos, P.A., and Helin, K. (2008). Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* *22*, 1345-1355.

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

Patil, V., Ward, R.L., and Hesson, L.B. (2014). The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* *9*.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., *et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* *107*, 323-337.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* *20*, 2922-2936.

Phatnani, H.P., Jones, J.C., and Greenleaf, A.L. (2004). Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry* *43*, 15702-15719.

Plasschaert, R.N., and Bartolomei, M.S. (2014). Genomic imprinting in development, growth, behavior and stem cells. *Development* *141*, 1805-1813.

Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* *48*, 575-608.

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* *16*, 272-278.

Quivoron, C., Couronne, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H., *et al.* (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* *20*, 25-38.

Rahman, S., Sowa, M.E., Ottinger, M., Smith, J.A., Shi, Y., Harper, J.W., and Howley, P.M. (2011). The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol* *31*, 2641-2652.

Raiber, E.A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol* *13*, R69.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., *et al.* (2013). Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* *502*, 65-70.

Ramanathan, Y., Rajpara, S.M., Reza, S.M., Lees, E., Shuman, S., Mathews, M.B., and Pe'ery, T. (2001). Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *J Biol Chem* *276*, 10913-10920.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* *138*, 114-128.

Rangam, G., Schmitz, K.M., Cobb, A.J., and Petersen-Mahrt, S.K. (2012). AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS One* *7*, e43279.

Reichert, N., Choukrallah, M.A., and Matthias, P. (2012). Multiple roles of class I HDACs in proliferation, differentiation, and development. *Cell Mol Life Sci* *69*, 2173-2187.

Reiss, D., and Mager, D.L. (2007). Stochastic epigenetic silencing of retrotransposons: does stability come with age? *Gene* *390*, 130-135.

Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal* *4*, rs3.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* *129*, 1311-1323.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* *273*, 5858-5868.

Rose, N.R., and Klose, R.J. (2014). Understanding the relationship between DNA and histone lysine methylation. *Biochim Biophys Acta*.

Ryan, K., Murthy, K.G., Kaneko, S., and Manley, J.L. (2002). Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. *Mol Cell Biol* *22*, 1684-1692.

Sajan, S.A., and Hawkins, R.D. (2012). Methods for identifying higher-order chromatin structure. *Annu Rev Genomics Hum Genet* *13*, 59-82.

Sameer, A.S., Nissar, S., and Fatima, K. (2014). Mismatch repair pathway: molecules, functions, and role in colorectal carcinogenesis. *Eur J Cancer Prev*.

Sancar, A. (1996). DNA excision repair. *Annu Rev Biochem* *65*, 43-81.

Sandell, L.L., and Zakian, V.A. (1993). Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell* *75*, 729-739.

Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* *373*, 487-493.

Scharer, O.D. (2003). Chemistry and biology of DNA repair. *Angew Chem Int Ed Engl* *42*, 2946-2974.

Scharer, O.D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol* *5*, a012609.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., *et al.* (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* *39*, 232-236.

Schmitz, K.M., Schmitt, N., Hoffmann-Rohrer, U., Schafer, A., Grummt, I., and Mayer, C. (2009). TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol Cell* *33*, 344-353.

Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J., *et al.* (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* *18*, 1263-1271.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* *128*, 735-745.

Schwarzbauer, K., Bodenhofer, U., and Hochreiter, S. (2012). Genome-wide chromatin remodeling identified at GC-rich long nucleosome-free regions. *PLoS One* 7, e47924.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* 48, 849-862.

Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci* 368, 20110330.

Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L., *et al.* (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* 40, 8255-8265.

Shah, M., and Allegrucci, C. (2013). Stem cell plasticity in development and cancer: epigenetic origin of cancer stem cells. *Subcell Biochem* 61, 545-565.

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

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153, 692-706.

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

Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. *Mol Cell* 19, 857-864.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844-847.

Singh, S.K., Wang, M., Staudt, C., and Iliakis, G. (2011). Post-irradiation chemical processing of DNA damage generates double-strand breaks in cells already engaged in repair. *Nucleic Acids Res* 39, 8416-8429.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810-813.

Solary, E., Bernard, O.A., Tefferi, A., Fuks, F., and Vainchenker, W. (2014). The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia* 28, 485-496.

Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., *et al.* (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153, 678-691.

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

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., *et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490-495.

Steinacher, R., and Schar, P. (2005). Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr Biol* 15, 616-623.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* 12, R54.

Svetlova, M.P., Solovjeva, L.V., and Tomilin, N.V. (2010). Mechanism of elimination of phosphorylated histone H2AX from chromatin after repair of DNA double-strand breaks. *Mutat Res* 685, 54-60.

Szathmary, E. (2003). Why are there four letters in the genetic alphabet? *Nat Rev Genet* 4, 995-1001.

Szenker, E., Ray-Gallet, D., and Almouzni, G. (2011). The double face of the histone variant H3.3. *Cell Res* 21, 421-434.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., *et al.* (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* 7, e1002154.

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

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

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell* 60, 909-920.

Tee, W.W., Shen, S.S., Oksuz, O., Narendra, V., and Reinberg, D. (2014). Erk1/2 Activity Promotes Chromatin Features and RNAPII Phosphorylation at Developmental Promoters in Mouse ESCs. *Cell* 156, 678-690.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 15, 668-679.

Thoma, F. (1999). Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair. *EMBO J* 18, 6585-6598.

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

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S., *et al.* (2010). Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* 17, 1154-1161.

Tini, M., Benecke, A., Um, S.J., Torchia, J., Evans, R.M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol Cell* 9, 265-277.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* 96, 8681-8686.

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* 28, 1-13.

Tsai, M.C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689-693.

Turcan, S., Rohle, D., Goenka, A., Walsh, L.A., Fang, F., Yilmaz, E., Campos, C., Fabius, A.W., Lu, C., Ward, P.S., *et al.* (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483, 479-483.

Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., and Chambon, P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J Biol Chem* 273, 20728-20736.

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat Genet* 36, 1296-1300.

van Attikum, H., and Gasser, S.M. (2005a). ATP-dependent chromatin remodeling and DNA double-strand break repair. *Cell Cycle* 4, 1011-1014.

van Attikum, H., and Gasser, S.M. (2005b). The histone code at DNA breaks: a guide to repair? *Nat Rev Mol Cell Biol* 6, 757-765.

van Gent, D.C., Hoeijmakers, J.H., and Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2, 196-206.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 15, 795-804.

Vaute, O., Nicolas, E., Vandel, L., and Trouche, D. (2002). Functional and physical interaction between the histone methyl transferase Suv39H1 and histone deacetylases. *Nucleic Acids Res* 30, 475-481.

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

Verger, A., and Crossley, M. (2004). Chromatin modifiers in transcription and DNA repair. *Cell Mol Life Sci* 61, 2154-2162.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 131, 58-69.

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

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M., *et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871-874.

Visnes, T., Doseeth, B., Pettersen, H.S., Hagen, L., Sousa, M.M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G., and Krokan, H.E. (2009). Uracil in DNA and its processing by different DNA glycosylases. *Philos Trans R Soc Lond B Biol Sci* 364, 563-568.

Waddington, C.H. (2012). The epigenotype. 1942. *Int J Epidemiol* 41, 10-13.

Walkinshaw, D.R., Tahmasebi, S., Bertos, N.R., and Yang, X.J. (2008). Histone deacetylases as transducers and targets of nuclear signaling. *J Cell Biochem* 104, 1541-1552.

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., and Zhang, Y. (2001). Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol Cell* 8, 1207-1217.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14, 637-646.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., *et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40, 897-903.

Watanabe, S., and Peterson, C.L. (2010). The INO80 family of chromatin-remodeling enzymes: regulators of histone variant dynamics. *Cold Spring Harb Symp Quant Biol* 75, 35-42.

Waters, T.R., Gallinari, P., Jiricny, J., and Swann, P.F. (1999). Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J Biol Chem* 274, 67-74.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737-738.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* 2, 1136-1143.

Wei, Y., Yang, C.R., Wei, Y.P., Zhao, Z.A., Hou, Y., Schatten, H., and Sun, Q.Y. (2014). Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proc Natl Acad Sci U S A* 111, 1873-1878.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., *et al.* (2007). Epigenetic stem cell signature in cancer. *Nat Genet* 39, 157-158.

Wiebauer, K., and Jiricny, J. (1990). Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. *Proc Natl Acad Sci U S A* 87, 5842-5845.

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

Wirbelauer, C., Bell, O., and Schubeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes Dev* *19*, 1761-1766.

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

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* *156*, 45-68.

Wu, Z., Lee, S.T., Qiao, Y., Li, Z., Lee, P.L., Lee, Y.J., Jiang, X., Tan, J., Aau, M., Lim, C.Z., *et al.* (2011). Polycomb protein EZH2 regulates cancer cell fate decision in response to DNA damage. *Cell Death Differ* *18*, 1771-1779.

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

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Xiao, M.T., Liu, L.X., *et al.* (2011a). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* *19*, 17-30.

Xu, Y., Hu, B., Choi, A.J., Gopalan, B., Lee, B.H., Kalady, M.F., Church, J.M., and Ting, A.H. (2011b). Unique DNA methylome profiles in CpG island methylator phenotype colon cancers. *Genome Res*.

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

Yang, H., Liu, Y., Bai, F., Zhang, J.Y., Ma, S.H., Liu, J., Xu, Z.D., Zhu, H.G., Ling, Z.Q., Ye, D., *et al.* (2013). Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* *32*, 663-669.

Yao, J., Ardehali, M.B., Fecko, C.J., Webb, W.W., and Lis, J.T. (2007). Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Mol Cell* *28*, 978-990.

Yin, R., Mao, S.Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., *et al.* (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J Am Chem Soc* *135*, 10396-10403.

You, J.S., and Jones, P.A. (2012). Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* *22*, 9-20.

Young, R.A. (2011). Control of the embryonic stem cell state. *Cell* *144*, 940-954.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., *et al.* (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* *149*, 1368-1380.

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

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H., and He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol* *8*, 328-330.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., *et al.* (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res* *38*, 4246-4253.

Zhao, Z., and Han, L. (2009). CpG islands: algorithms and applications in methylation studies. *Biochem Biophys Res Commun* *382*, 643-645.

Zheng, G., Fu, Y., and He, C. (2014). Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics. *Chem Rev*.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* *12*, 7-18.

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

Appendix:

- I: TDG Balances DNA Methylation and Oxidative Demethylation in Differentiating Cells

- II: TET1, TET2 and TDG Cooperate in a Locus-Specific Manner to Promote Chromatin Plasticity by Oxidative DNA Demethylation

- III: Embryonic Lethal Phenotype Reveals a Function of TDG in Maintaining Epigenetic Stability

Appendix I

TDG Balances DNA Methylation and Oxidative Demethylation at CpG Islands in Differentiating Cells

Jacobs, A.L.^{1,*}, Cortázar, D.^{1,*}, **Wirz, A.**^{1,*}, Arand, J.², Steinacher, R.¹, Broberg Vågbø, C.³, Giehr, P.², Weber, A.¹, Wilson, G.⁴, Galashevskaya, A.³, Kunz, C.¹, Reik, W.⁵, Beck, S.⁴, Walter, J.², Krokan, H.³, Schär, P.¹

¹ Department of Biomedicine, University of Basel, Switzerland;

² Department of Biological Sciences, Institute of Genetics/Epigenetics, University of Saarland, Saarbrücken, Germany;

³ Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway;

⁴ Medical Genomics, UCL Cancer Institute, University College London, London, United Kingdom;

⁵ Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, United Kingdom

* These authors contributed equally to this study

Correspondence:

primo.schaer@unibas.ch; Tel. +41 61 267 3561; Fax: +41 61 267 3566

Contribution: Extraction of genomic DNA from timecourse samples and performing MeDIP, GLIB and caC-DIP experiments thereof. qPCRs for the targets shown and further positive and negative controls (data not shown) as well as extensive quantitative analysis and quality control of the DIP and GLIB data sets. Preparation of chromatin extracts of timecourse samples and conducting TET1 and TDG ChIP experiments as well as subsequent qPCRs. Extensive quantitative analysis of these data and quality control of these ChIP data sets. Further, fC-DIP including qPCR analysis, expression analysis of the DMRs and surrounding genes (RT-qPCR), TET1 and TDG ChIPs including qPCR of TET1-knockout mESCs and establishing and performing TET2 ChIP (data not shown). Additionally, sample preparation for oxBis and MAB-seq analysis (by PG); coordination of this collaboration.

Abstract

The Thymine DNA Glycosylase (TDG) initiates Base Excision Repair of G•T mismatches arising from deamination of 5-methylcytosine (5-mC). Due to this substrate specificity, TDG has been suggested to act in a deamination-coupled 5-mC demethylation process. More recently, TDG has been found to process 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), the final products of TET mediated 5-mC oxidation, implicating TDG in an oxidative DNA demethylation process. However, the significance of either of these proposed pathways in the context of epigenetic programming during cell differentiation is yet unclear. Here, we report that TDG is required to establish DNA methylation at CpG islands during differentiation by controlling a transitory cycle of DNA methylation and demethylation. We provide evidence that this cycle does not entail a deaminase but stepwise oxidation of 5-mC and that TDG structure and catalytic activity both contribute to controlling the epigenetic transitions from a pluripotent to a differentiated state.

Introduction

Cell type specific patterns of gene expression are shaped by chemical modifications of histone proteins and the DNA, termed “epigenetic”. The C5-position of cytosine is subject to methylation by DNA methyltransferases (DNMTs) (Goll and Bestor, 2005). 5-methylcytosine (5-mC) occurs predominantly in CpG dinucleotides, the vast majority of which is methylated throughout the genome, with the exception of CpG islands (CGIs) (Bird et al., 1985). These regions of high CpG density are maintained unmethylated and colocalize with the promoters of all ubiquitously expressed genes but also with about 40% of those with tissue-specific expression patterns. However, a small but significant proportion of CGIs, many of which are distal to promoters, is differentially methylated between cell types (Illingworth and Bird, 2009).

In contrast to histone modifications that are highly dynamic (Bannister and Kouzarides, 2011), cytosine methylation has long been regarded as a stable epigenetic mark that is established and maintained by DNMT3a or b and DNMT1, respectively (Goll and Bestor, 2005). However, two global DNA demethylation events have been described to occur in the mammalian life cycle (Seisenberger et al., 2013) and both have recently been shown to involve the activity of the Ten Eleven Translocator (TET) family of proteins. These 5-mC hydroxylases convert 5-mC to 5-hydroxymethylcytosine (5-hmC) (Tahiliani et al., 2009), which is not maintained by DNMT1 and thus diluted through DNA replication (Hackett et al., 2013; Hashimoto et al., 2012; Iqbal et al., 2011; Valinluck and Sowers, 2007; Wossidlo et al., 2011). Whereas such passive removal presents a plausible pathway for global demethylation, it appears that specific loci are actively demethylated independent of DNA replication (Bruniquel and Schwartz, 2003; Kangaspeska et al., 2008; Metivier et al., 2008).

Efforts to identify an enzyme that actively demethylates 5-mC have implicated the thymine DNA glycosylase (TDG) as a prime candidate. Its protein interactions have placed TDG in the context of DNA methylation control and regulation of gene expression (Cortazar et al., 2007). In further support of such a function, we and others have found deletion of *Tdg* in mice to be embryonic lethal and cause epigenetic aberrations at CpG island promoters (Cortazar et al., 2011; Cortellino et al., 2011).

As TDG recognizes and processes the deamination product of 5-mC, a G•T mismatch, it has been proposed to act downstream of a cytosine deaminase in a putative DNA demethylation pathway, e.g. the activation induced deaminase (AID) or the apolipoprotein B RNA-editing catalytic component (APOBEC) enzymes. Several lines of evidence support an involvement of these deaminases in both, global and targeted DNA demethylation (Bhutani et al., 2010; Popp et al., 2010; Rai et al., 2008). Another hypothetical DNA demethylation pathway entails oxidation, converting 5-mC to 5-hmC, and subsequent deamination of 5-hmC to 5-hydroxymethyluracil (5-hmU), which can be excised by TDG (Cortellino et al., 2011; Hardeland et al., 2003). However, a recent study has cast doubt on such a pathway since

AID and the APOBEC deaminases appear to be mostly inactive on 5-hmC (Nabel et al., 2012).

G•T and G•5-hmU are processed also by the Methyl-CpG Binding Domain protein 4 (MBD4) and the Single-strand specific Monofunctional Uracil Glycosylase 1 (SMUG1), respectively. As neither MBD4 nor SMUG1 are essential for embryonic development (Kemmerich et al., 2012; Wong et al., 2002) and neither can compensate for the loss of TDG, it appears that TDG acts in a non-redundant pathway essential for embryo development. Such a pathway has taken shape with the finding that the TET proteins can oxidize 5-hmC further to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), both of which are excellent substrates for TDG-mediated Base Excision Repair (BER) and appear not to be processed by any other DNA glycosylase (He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011). However, the significance of a putative DNA demethylation pathway coupling the activities of TET and TDG through processes of cell fate determination has not been studied. Here, we report that TDG is essential for establishing differentiation-induced methylation at CpG islands by structurally and enzymatically supporting an equilibrium of DNA methylation and oxidative demethylation during a transitory state of high epigenetic plasticity.

Results

Aberrant DNA methylation in *Tdg*^{-/-} cells

We reported previously that TDG is essential for embryonic development and that TDG deficient cells accumulate aberrant DNA methylation at CpG island (CGI) promoters (Cortazar et al., 2011). To investigate the role of TDG in the regulation of DNA methylation, we performed MeDIP combined with next generation sequencing on DNA from TDG proficient (*Tdg*^{+/-}) and deficient (*Tdg*^{-/-}) embryonic stem cells (ESCs), early (4h) neuronal progenitors (NPs) derived by *in vitro* differentiation (Fig. S1a) and MEFs isolated from *Tdg*^{+/+} and *Tdg*^{-/-} embryos (Wilson, 2012). Whereas *Tdg*^{+/-} and *Tdg*^{-/-} ESCs showed no significant

differences in their DNA methylation patterns, *in vitro* neuronal differentiation gave rise to 942 differentially methylated regions (DMRs), and comparison of the MEFs revealed 32976 DMRs (Fig. 1a). This phenotype indicated a failing DNA methylation control in the TDG deficient cells that deteriorates with differentiation. This notion is supported by the observation that *Tdg*^{-/-} ESCs fail to form terminal neurons *in vitro* and rapidly lose cell viability in neuronal differentiation medium (Fig. S1a).

Of the 942 DMRs found in NPs, 609 are hypermethylated and 333 hypomethylated in *Tdg*^{-/-} cells compared to *Tdg*^{+/-}. As DNA methylation is not equally distributed throughout the genome (Meissner et al., 2008), we characterized the relationship between DMRs and CpG density. In the absence of TDG, CpG poor DMRs were preferentially hypermethylated while DMRs with a higher CpG density were associated with a loss of DNA methylation (Fig. S1b). We also analysed the average distance of the DMRs to the nearest transcriptional start site (TSS). Hypomethylated DMRs were on average located closer to a TSS (24.3 kb ± 45.0 kb) than the hypermethylated (47.7 kb ± 77.4 kb) (Fig. S1c). Accordingly, 57% of the hypomethylated DMRs but only 34% of the hypermethylated lie within 10 kb of a TSS. Intersection of the DMRs with promoter regions confirmed that only a minority of the DMRs overlap with promoters (1kb upstream and 0.5kb downstream of a TSS), but that hypomethylated DMRs are more often promoter-associated than hypermethylated ones (Fig. S1d). We thus conclude that the hypomethylated DMRs are more likely to affect gene expression, but aberrant methylation appears to affect mostly regions distal to promoters.

De novo methylation of CGIs requires TDG

We found 123 DMRs to overlap with CGIs (UCSC) (Gardiner-Garden and Frommer, 1987). Unexpectedly, 122 of these 123 differentially methylated CGIs, henceforth called CGI DMRs, were hypomethylated in *Tdg*^{-/-} NPs, whereas DMRs not classified as CGIs were predominantly hypermethylated (hyper: hypomethylated = 3:1) (Fig. 1b). Although a vast majority of CGIs is maintained in a hypomethylated state in ESCs, a subset was shown to

acquire *de novo* methylation during neuronal differentiation (Mohn et al., 2008). We therefore asked whether hypomethylation at CGIs in *Tdg^{-/-}* NPs represents a loss of DNA methylation present in ESCs or a failure to establish methylation during NP differentiation. We thus intersected the CGI DMRs with all CGIs acquiring *de novo* methylation with differentiation (Wilson, 2012). 117 of the 122 CGI DMRs overlap with CGIs that acquire methylated in NP differentiation (Fig. 1c), revealing diminished differentiation-associated *de novo* methylation to be the cause of hypomethylation in absence of TDG.

To further explore the genomic features of the CGI DMRs, we intersected them with published datasets of genome-wide protein-binding sites and histone modifications in ESCs. This revealed that the CGI DMRs were significantly depleted for gene promoters (Ensembl TSS -1kb and +0.5kb), sites of RNA-polymerase II (RNA-Pol II), histone acetyltransferase p300 and H3K27ac enrichment (Fig. S1e). On the other hand, we found the CGI DMRs to be enriched in sites of TET1 binding and H3K4me1 and H3K27me3 modification (Fig. 1d), suggesting that a large proportion of these CGIs represent enhancer elements and targets of the polycomb repressive complex 2 (PRC2) which trimethylates H3K27 (Kuzmichev et al., 2002). Enhancer elements were shown to be marked by H3K4 monomethylation and bound by TET1 (Heintzman et al., 2007; Serandour et al., 2012) but the fact that the CGI DMRs are depleted for H3K27ac and p300 suggests that these enhancer elements are inactive or poised in ES cells (Creyghton et al., 2010). Interestingly, we found a highly significant overlap of CGI DMRs with low methylated regions (LMRs) that represent transcription factor binding sites at distal regulatory regions (Stadler et al., 2011); 52% of the CGI DMRs coincided with NP-specific LMRs and 7% with ESC-specific LMRs (Fig. 1d), whereas constitutive LMRs showed no significant overlap (Fig. S1e). The CGI DMRs thus appear to be enriched for polycomb targets and poised enhancer elements.

Loss of 5-mC is not caused by mutation

The diminished differentiation-triggered methylation of CGIs and, thus, the apparent hypomethylation of such regions in *Tdg* knockout NPs can be explained in two ways: 1) by failure to target the DNA methylation machinery to these regions or 2) by conversion of 5-mC to another base that would no longer be recognized by the 5-mC antibody used in MeDIP. Conversion of 5-mC could occur by deamination, i.e. by AID, which would generate a G•T mismatch that – unless repaired by TDG or MBD4 – will give rise to C→T mutations, or by oxidation of 5-mC to 5-hmC and further to 5-fC and 5-caC by the TET proteins.

To test these hypotheses and to be able to distinguish between a structural and enzymatic role of TDG in this context, we performed *in vitro* differentiation in a complemented cell system, in which either wildtype TDG (wt), a catalytically dead mutant TDG N151A (TDG_{Δcat}) or vector control (ko) were stably expressed in *Tdg*^{-/-} ESCs. Genomic DNA from NPs derived from these ESCs was subjected to hairpin Na-bisulfite sequencing (BS-seq) to allow simultaneous analysis of strand-specific methylation status and mutation frequency (Arand et al., 2012). The analysis of 7 representative hypomethylated CGIs (Supplementary Table 1) sequenced with a coverage of ~10'000 reads confirmed the hypomethylation in 5 targets (Fig. 2). Furthermore, the frequency of C→T mutations we observed did not rise above the error rate of the method and cannot explain the loss of 5-mC, which is in the higher percentage range. We thus conclude that the hypomethylation appearing in NPs is not caused by deamination of 5-mC or 5-hmC, as both deamination products (T and 5-hmU) are pre-mutagenic and would give rise to appreciable amounts of C→T mutations in *Tdg* knockout cells. Compensation by other DNA glycosylases like MBD4 and SMUG1 is unlikely, as neither is capable of compensating the loss of 5-mC nor the developmental phenotype of the *Tdg* knockout.

Notably, 2 of the 7 targets chosen for hairpin Na-bisulfite sequencing (DMR36 and 8) exhibited hypermethylation in TDG_{Δcat} and (only in DMR8) knockout NPs but hypomethylation in MeDIP-seq in TDG deficient cells. 5-mC and 5-hmC are not distinguishable by BS-seq,

whereas MeDIP-seq relies on an antibody specific for 5-mC. The discrepancy between the results from BS-seq and MeDIP-seq thus suggests accumulation of 5-hmC at the respective targets. This notion is supported by the increased appearance of hemimethylated CpGs at the same targets; 5-hmC is not maintained by DNMT1 (Hashimoto et al., 2012; Valinluck and Sowers, 2007) and therefore is expected to occur more often opposite an unmethylated CpG.

5-fC and 5-caC rise with differentiation

TDG was proposed to be the only DNA glycosylase capable of excising 5-fC and 5-caC and, consistently, the levels of these C-modifications were shown to increase following a knockdown of TDG in ESCs (He et al., 2011). We wanted to investigate the generation of these derivatives in the context of ESC differentiation, i.e. when differential methylation in TDG proficient and deficient cells becomes apparent. Yet, as neither 5-hmC nor the higher oxidized C-modifications are maintained by DNMT1 (Inoue et al., 2011; Valinluck and Sowers, 2007), the quantitative analysis of these modifications is perturbed by dilution through DNA replication. To minimize this dilution effect, we performed a 24 h retinoic acid (RA) differentiation time course, allowing a maximum of two rounds of DNA replication to occur. To reduce epigenetic heterogeneity often observed in ESC culture, we conditioned our complemented ESC lines in 2i medium prior to RA-induced differentiation (Ying et al., 2008). We observed that culturing in 2i medium decreased the global 5-mC levels in comparison to cells cultured exclusively in ESC medium with LIF (Fig. S3b) by about 50%, irrespective of *Tdg* genotype. This suggested that active inhibition of differentiation in the 2i medium changes the epigenetic ground state of our ESCs in a TDG independent manner and consistent with previous observations ((Leitch et al., 2013) and Reik, W., personal communication).

We harvested genomic DNA and chromatin after 0, 8 and 24 h of incubation with RA (Fig. 3a). By testing the mRNA levels of *Nanog*, *Oct4*, *Rex1* and *Gata6* throughout the timecourse,

we verified the downregulation of pluripotency genes and induction of developmental genes within these 24 h of differentiation (Fig. S2a), and we confirmed at mRNA and protein level that TET1 and TET2 expression was equal in all three cell lines (Fig. S2a and b and Fig. 5b). The levels of AID mRNA were extremely low and protein levels were below the detection limit in Western blot analysis (Fig. S2c).

We then measured levels of 5-mC, 5-hmC, 5-fC, 5-caC and 5-hmU in the genomic DNA of undifferentiated and differentiated cells by liquid chromatography-tandem mass spectrometry (LCMSMS). In agreement with previous findings in TDG knockdown experiments (He et al., 2011), we found a significant enrichment of 5-fC and 5-caC (~2- and 9-fold, respectively) in *Tdg* knockout as well as catalytically inactive ($TDG_{\Delta cat}$) cells (Fig. 3b). We also found the global levels of 5-mC, 5-fC and 5-caC to rise with differentiation. This effect was more pronounced in cells lacking TDG activity and specifically induced by RA (Fig. S3a and c). Accordingly, global 5-mC levels became significantly different between TDG proficient and deficient (ko, $TDG_{\Delta cat}$) cells at 24 h of differentiation (Fig. 3b). Interestingly, global 5-fC levels in mutant cells increased significantly above wildtype levels already after 8 h of differentiation. By contrast, 5-caC levels were ~9 fold higher in undifferentiated mutant cells compared to wildtype and only the mutant cell lines showed a further significant increase in 5-caC with differentiation (Fig. S3a). Global 5-hmC and 5-hmU levels were not significantly different neither between any of the genotypes nor within the differentiation timecourse.

From these results, we conclude that the loss of pluripotency triggers a turnover of global 5-mC, 5-hmC, 5-fC and 5-caC, a process which is disturbed in the absence of active TDG. Overall, cytosine modification levels are equally affected in *Tdg* knockout or $Tdg_{\Delta cat}$ cells, implicating a catalytic active role of TDG in controlling transitions in CpG methylation.

TDG activity balances 5-mC and 5-caC

The absence of increased C→T mutations at CGI DMRs indicated that the hypomethylation observed in TDG deficient NPs is not a result of deamination-mediated loss of CpG sites but, instead, may originate from targeted 5-mC oxidation by the TET proteins. To address this hypothesis, we analyzed 5-mC (MeDIP), 5-hmC (GLIB) and 5-caC (caCDIP) levels at 4 CGI DMRs and compared their change in the 24 h interval of RA-induced ESC differentiation in the presence or absence of TDG protein and/or activity.

While the effects we observed varied to some extent with the genomic context of the target, certain trends became apparent across targets. We observed a slight but consistent differentiation-induced increase of 5-mC in TDG wildtype and TDG_{Δcat} set against no change or even a decrease in knockout cells (Fig. 4a). While the levels of 5-hmC showed a high variability between replicate experiments and, thus, no consistent difference between *Tdg* genotypes, 5-caC levels were clearly increased in TDG_{Δcat} across the targets when compared to wildtype and knockout cells (Fig. 4a, aggregated p-value < 0.0001, Anova). Comparing the proportions of 5-mC and 5-caC modifications at 0 and 24 h revealed a shift in the equilibrium between these modifications in a time- and genotype-dependent manner (Fig. S4a). While in wildtype and in *Tdg* ko cells, both modifications remained equilibrated at the CGI DMRs during the 24 h interval of differentiation, this balance tipped towards an increase in 5-caC in differentiated TDG_{Δcat} cells (Fig. S4a).

The increase of 5-mC and 5-caC at CGI DMRs in RA-stimulated wildtype and TDG_{Δcat} ESCs indicates that the transition to a higher methylated state of these CGIs in the course of cell differentiation is accompanied by the generation of higher oxidized 5-mC-modifications. Subsequent re-methylation by DNMT3a or DNMT3b requires 5-caC to be excised and replaced with an unmodified C. The fact that *Tdg*^{-/-} cells did not show an increase in 5-mC but rather a reduction of 5-caC suggests that TDG is structurally involved in the initiation and/or maintenance of cyclic methylation and oxidative demethylation.

This cycle of methylation and demethylation appears to be disrupted in TDG_{Δcat}, resulting in an increase of 5-caC over time that surpasses that observed in wildtype. To elucidate how this disruption occurs, we characterized TDG_{Δcat} biochemically regarding its activity on and association with 5-caC. We tested the activity of recombinant wildtype TDG and TDG_{Δcat} in a standard base release assay on double-stranded oligonucleotide substrates with one fluorescence-labeled strand containing a single thymine or modified cytosine opposite guanine. We found TDG_{Δcat} to be virtually inactive on 5-caC (Fig. 4b), which is in agreement with the accumulation of 5-caC in TDG_{Δcat} cells (Fig. 3b). We then tested the ability of the catalytic-dead TDG to bind the 5-caC substrate in electrophoretic mobility-shift assays with fluorescence-labelled substrate oligonucleotides (G•T, G•5-mC, G•5-hmC, G•5-caC) in the presence of a 10 or 20 fold excess of unlabelled competitor DNA containing either unmodified C or a 5-caC. These competition assays identified 5-caC as the substrate most efficiently bound by TDG (Fig. 4c); the binding specificity and efficiency of the catalytic-dead protein appeared to be comparable to that of the wildtype TDG (Fig. S4b) with the caveat that the assay with the latter most likely reflects binding of to the product abasic-sites (Hardeland et al., 2000). We thus conclude that TDG_{Δcat} binds 5-caC with higher affinity than 5-mC and 5-hmC.

TDG facilitates TET1 occupancy at CGIs

The differential effects of the *Tdg* disruption and the catalytic-dead mutant on the methylation-demethylation equilibrium at the CGI DMRs suggest a structural role of TDG in this context. To elucidate this role, we tested whether the presence or absence of TDG influences the association of TET1 – the most highly expressed of the TET proteins in ES cells (Fig. S2a) – with these regions. Chromatin immunoprecipitation (ChIP) revealed that TET1 enrichment increases at all selected CGI DMRs over time of RA stimulation in wildtype cells. Whereas the association of TET1 to the CGI DMRs appears to be independent of TDG in pluripotent cells, differentiation induces a gradual loss of TET1 occupancy at the CGI

DMRs in both knockout and TDG $_{\Delta cat}$ cells (Fig. 5a and S4). These findings corroborate that initiation of ESC differentiation activates a cycle of DNA methylation and demethylation involving 5-mC oxidation at specific CGIs.

Notably, while TDG $_{\Delta cat}$ is sufficient to support the stepwise oxidation of 5-mC to 5-caC structurally (Fig. 4a), the absence of the catalytic activity in this mutant significantly destabilizes – or suppresses – TET1 association to the CGI DMRs (Fig. 5a). The high affinity of TDG $_{\Delta cat}$ to 5-caC combined with its inability to turn over (Fig. 4b and c) is likely to result in an accumulation of TDG $_{\Delta cat}$ at these CGIs, thus blocking the progression of the cycle. Indeed, by ChIP, we found TDG $_{\Delta cat}$ to be clearly enriched at the CGI DMRs, while association of wildtype TDG hardly rose above the background measured in *Tdg* knockout (Fig. 5c, for relative enrichment, controls and statistics see Fig. S6).

It thus appears that both TDG structure and catalytic activity are essential for stabilizing TET1-occupancy at the CGI DMRs but that TDG structure can become an obstacle without the ability to turnover on 5-caC.

Discussion

We have found that *in vitro* differentiation of TDG deficient ESCs is accompanied by an increasing disturbance of DNA methylation patterns. Intersection of the DMRs with CGIs revealed that 99% of all differentially methylated CGIs are hypomethylated in cells lacking TDG. The vast majority of these regions are CGIs that undergo *de novo* methylation during *in vitro* differentiation of ESCs to NPs, suggesting that TDG is essential for establishing cell-type specific methylation of CGIs in the course of cell differentiation. Interestingly, many of these CGI DMRs appear to be inactive or poised enhancers and overlap with NP-specific LMRs (Stadler et al., 2011).

We found that a catalytically-dead but structurally intact TDG variant (TDG_{Δcat}) fails to rescue differentiation-triggered methylation at these CGI DMRs, showing that the establishment and maintenance of methylation patterns at these regions depends on the active excision of DNA bases. Furthermore, we exclude that loss of 5-mC at these CGIs arises through deamination of either 5-mC or 5-hmC, which was proposed to be catalysed by AID in direct interaction with TDG (Cortellino et al., 2011). By hairpin bisulfite-sequencing with high coverage of a representative set of hypomethylated CGIs, we did not detect increased mutation levels, which would be the consequence of unrepaired deaminated 5-mC, neither in *Tdg* knockout cells nor in cells expressing TDG_{Δcat}. The results with the catalytic inactive but structurally intact TDG allow us to exclude that the lack of mutations may reflect a failure to recruit AID to these genomic loci. Our data strongly argue against a deamination-dependent process accounting for the loss of 5-mC at these CGIs.

In contrast, our data support a model connecting the loss of 5-mC at CGIs in TDG deficient NPs with the conversion of 5-mC to 5-hmC and 5-fC/caC by TET proteins. In a 24 h timecourse of RA-induced ESC differentiation, we found the genomic levels of all 5-mC-derivatives to increase in *Tdg* wildtype, knockout and TDG_{Δcat} cells, but more extensively so in the mutant cell lines. This suggests that differentiation triggers the generation of 5-mC and of oxidative demethylation intermediates, the latter of which accumulate globally in absence of TDG.

The analysis of local C-modification levels at representative CGI DMRs revealed an imbalance of 5-mC and 5-caC in TDG_{Δcat} but no significant shift of C-modification levels in knockout cells, indicating a dual function of TDG in controlling these levels during differentiation: one as an enzymatic component and one as a structural scaffold. Indeed, we found that TET1 associates with the CGI DMRs independent of TDG in ESCs but is rapidly lost upon differentiation in *Tdg* knockout cells.

Additionally, we found that the failure of turnover of TDG_{Δcat} combined with its high affinity for 5-caC disrupts the methylation-demethylation complex, i.e. TET1-occupancy at the CGI

DMRs. This may occur either through steric hindrance or by generation of unusual demethylation intermediates, i.e. hemi-5-caC sites that cannot be remethylated by DNMT1 after DNA replication. Also, DNMT3a and DNMT3b cannot target 5-caC for methylation and may not be able to methylate a CpG opposite a 5-caCpG bound by TDG_{Δcat}. Without BER resetting the methylation state, methylation cannot be established correctly at these CGIs, which eventually results in the hypomethylation observed in NPs. The passive removal of 5-hmC/fC/caC by DNA replication does not appear to be sufficient to establish methylation at these loci as the process clearly depends on functional TDG.

We propose that at these CGIs, RA-induced differentiation triggers a cycle of DNA methylation and demethylation involving the iterative oxidation of 5-mC and enzymatic removal of 5-fC/caC by TDG and BER (Fig. 6). This establishes a transient equilibrium of methylation and demethylation intermediates that, at later stages of differentiation (early NPs), is tipped towards methylation (Fig. 6), suggesting that this cycle represents a transitory state that accompanies the loss of pluripotency. Such a cycle induced by differentiation is in agreement with our previous observation that RA induces DNA repair processes involving TDG that increase the number of XRCC1 foci and the sensitivity to PARP inhibitors in differentiating wildtype compared to *Tdg*^{-/-} cells (Cortazar et al., 2011).

Previous reports of cyclical methylation and demethylation of the *pS2* promoter in response to estrogen-induced transcriptional activation have implicated DNMT3a and DNMT3b as well as TDG in coordinating these epigenetic transitions between transcriptionally active and silent states (Kangaspeska et al., 2008; Metivier et al., 2008). Others have proposed such a cycle involving oxidation of 5-mC by the TET proteins. However, the biological function of this cycle has not been described. We provide the first phenotypic evidence for a biological function of this cycle in establishing methylation patterns during cell-fate determination as its disruption by deletion of *Tdg* causes epigenetic aberrations and compromises cell viability.

Methods

Cell culture and ES cell differentiation

For NP differentiation, ESCs were grown on Feeders at 37°C in ES cell medium (ECM: DMEM, 15% heat-inactivated FCS, 1x non-essential amino acids, 1 mM Na-pyruvate, 2 mM L-glutamine and 90 µM β-mercaptoethanol) with LIF (1'000Uml⁻¹) in a humidified atmosphere containing 5% CO₂.

Prior to differentiation, ESCs were grown without Feeders for 2 passages. For embryoid body formation, 4x10⁶ *Tdg*^{+/-}, *Tdg*^{-/-} or *Tdg*^{-/-} pWt, pΔcat and pVec ESCs were plated onto non-adherent bacterial dishes (Greiner Bio-one) in differentiation medium (ECM without LIF and with 10% FCS) and grown at 37°C with a medium exchange after 2 days. After 4 days, 5 µM all-trans retinoic acid (RA) was added and cells were further incubated for 4 days with a medium exchange after 2 days. Embryoid bodies were washed twice with 1x PBS and dissociated with freshly prepared trypsin solution (0.05% TPCK-treated trypsin in 0.05% EDTA/PBS) at 37°C for 3 min. Dissociated embryoid bodies were resuspended in 10 ml differentiation medium and collected by centrifugation at 700xg for 5 min at room temperature. The pellet was resuspended in N2 medium (DMEM-F12 nutrient mixture 1:1, 1xN2 supplement) and the cell suspension filtered through a 40 µm nylon cell strainer (BD). Filtered cells were immediately plated onto poly-L-lysine and laminin-coated dishes at a density of 5x10⁶ cells per 60 mm dish or 1.5x10⁷ cells per 100 mm dish. The N2 medium was exchanged 2 and 24 h after plating. For MeDIP-sequencing and hairpin BS-sequencing, cells were collected after 4 h in N2 medium.

Complemented ES cell lines were derived by transfection of *Tdg*^{-/-} ES cells with the complementation vectors pTCO2 TDG wt, pTCO2 TDG_{Δcat} and empty pTCO2 (Cortazar et al., 2011) using jetPEI® (Polyplus Transfections) according to the manufacturer's

recommendations. Cells were cultivated in ECM supplemented with $1.5 \mu\text{g ml}^{-1}$ puromycin to select stable clones.

For the 24 h RA differentiation, complemented ES cells were cultured on Feeders for 2 passages, then conditioned for 4 passages without Feeders in 2i medium (Neurobasal medium and DMEM/F-12 1:1, 1x N2 supplement, 1x B27 supplement, LIF ($1'000 \text{ Uml}^{-1}$), 2 mM L-glutamine, 90 μM β -mercaptoethanol, 3 μM CHIR99021 and 1 μM PD0325901 (University of Dundee) and 1x penicillin/streptomycin). Prior to RA differentiation, ESCs were seeded at suitable cell numbers for each time point onto two 140 mm dishes (for Chromatin and genomic DNA extraction) or two 30 mm dishes (for Protein and RNA extraction) in ECM. For differentiation, the medium was exchanged for ECM without LIF but supplemented with 5 μM RA (5 mM stock in DMSO). Chromatin, genomic DNA were harvested at 0, 8 and 24 h, Protein and RNA at -16 h (2i control), 0, 4, 8 and 24 h. For the DMSO control for LCMSMS, ES cells were treated accordingly but incubated with DMSO 1:1'000 in ECM.

If not indicated otherwise, cell culture components were obtained from Gibco® Life Technologies, chemicals from Sigma and LIF from Merck Millipore.

MeDIP-Sequencing

5 μg of DNA was sonicated giving fragment sizes $<500\text{bp}$. Fragments were end repaired, phosphorylated, 3' adenylated and ligated to Illumina adapters in accordance with the Illumina Multiplex Sample Preparation protocol (PE-930-1001). These samples were then subjected to MeDIP as described previously (Weber et al., 2005), with 3 μg 5-mC antibody (Euogentec) per 1 μg DNA. The immunoprecipitated (IP) sample was purified using the DNA Clean & Concentrator™-5 kit (Zymo Research) according to the manufacturer's instructions. The sample isolated by MeDIP then underwent gel electrophoresis and library size selection (150-200 bp), prior to PCR amplification using Illumina paired-end primers for 18 cycles. During this step, the libraries were tagged with a unique identifier, or index, as per Multiplex

Sample Preparation Oligonucleotide protocol (PE-400-1001). Libraries were quantified using an Agilent Bioanalyzer 2100.

MeDIP-seq Data Analysis

The generated MeDIP-seq data were analyzed using the computational pipeline MeDUSA (v1.0.0)(Wilson, 2012) and the MEDIPS (v1.0.0) R bioconductor package (Chavez et al., 2010). MeDUSA comprises several analysis steps. Firstly, BWA (v0.5.8) (Li and Durbin, 2009) was used to align the paired end sequence data to the reference mouse genome (Build mm9) using default settings. Filtering was performed to remove reads that were unable to be aligned as a viable pair and also those pairs in which neither read scored an alignment score of ≥ 10 . In cases of non-unique reads, possibly caused by PCR artifacts, all but one pair was removed. Quality control was performed using the tool FastQC (v0.9.4) (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>) and MEDIPS. The USeq (v6.8) suite of tools (Nix et al., 2008), specifically MultipleReplicaScanSeqs (MRSS) and EnrichedRegionMaker, were used to identify DMRs between cohorts. MRSS processes Point data for use in the BioConductor package DESeq (Anders and Huber, 2010). Window size was set at 500. Only regions containing a minimum of 10 reads summed from the cohorts being compared were included for DMR analysis. The dataset was initially described in (Wilson, 2012), and is available in the GEO repository (GSE27468).

To determine the overlap between DMRs and other genomic features, the “operate on genomic intervals” tool of the Galaxy project was used (<http://usegalaxy.org/>; (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010). DMRs were intersected with promoters, defined as Ensembl TSS plus 1kb upstream and 0.5kb downstream, RNA-Pol II (GSM918749), p300 (GSM918750), H3K4me1 (GSM1000121), H3K27ac (GSM1000126), H3K4me3 (GSM769008) and H3K27me3 (GSM1000089) peaks for ES cells, generated by ENCODE/LICR (Dunham et al., 2012), as well as with CGI coordinates (Gardiner-Garden and Frommer, 1987) and LMRs (Stadler et al., 2011). TET1 binding sites (GSM706672) were

converted from mm8 to mm9 using liftOver (Kent et al., 2002) prior to intersection with the DMRs.

Hairpin bisulfite deep sequencing for selected genomic regions

The analysis was performed according to (Arand et al., 2012). Briefly, genomic DNA was digested with a restriction enzyme cutting in the selected DMRs, specified in Supplementary Table 1, followed by a ligation of a hairpin linker to link the upper to the lower strand. After bisulfite treatment the selected regions were amplified. Restriction enzymes and primers used in this analysis are given in Supplementary Table 2. The amplified products were sequenced by 454 sequencing. The sequencing data was then analyzed by BiQAnalyzerHT (Lutsik et al., 2011) for accurate alignment and methylation evaluation, followed by merging of the methylation information of the upper and lower strand using python scripts. Average methylation, hemimethylation and mutation rates were calculated in Microsoft Excel.

LCMSMS analysis of global C-modification levels

Genomic DNA was enzymatically hydrolyzed to nucleosides essentially as described (Crain, 1990), followed by addition of 3 volumes of methanol and centrifugation (16'000xg, 30 min, 4°C). The supernatants were dried and dissolved in 50 µl 5% methanol in water (v/v) for LCMSMS analysis of the deoxynucleosides 5-hm(dC), 5-f(dC), 5-ca(dC), and 5-hm(dU). A portion of each sample was diluted for the quantification of 5-m(dC) and unmodified deoxynucleosides (dA, dC, dG, and dT). Chromatographic separation was performed on a Shimadzu Prominence HPLC system with a Zorbax SB-C18 2.1x150 mm i.d. (3.5 µm) column equipped with an Eclipse XDB-C8 2.1x12.5 mm i.d. (5 µm) guard column (Agilent Technologies). The mobile phase consisted of water and methanol (both supplemented with 0.1% formic acid), for 5-m(dC), 5-hm(dC), 5-f(dC), and 5-ca(dC) starting with a 5 min gradient of 5-60% methanol, followed by 6 min re-equilibration with 5% methanol, and for

unmodified nucleosides maintained isocratically with 85% methanol. hm(dU) was gradient chromatographed with a mobile phase of only water and methanol. Mass spectrometry detection was performed using an MDS Sciex API5000 triple quadrupole (Applied Biosystems) operating in positive electrospray ionization mode for the mass transitions 258.1/ 142.1 (5-hm(dC)), 256.1/ 140.1 (5-f(dC)), 272.1/ 156.1 (5-ca(dC)), 242.1/ 126.1 (5-m(dC)), 252.1/136.1 (dA), 228.1/112.1 (dC), 268.1/152.1 (dG), and 243.1/127.1 (T), or negative electrospray ionization mode for the mass transitions 257.1/ 124.1, 257.1/ 141.1, and 257.1/ 214.1 (5-hm(dU), quantifier and qualifier ions).

Purification of recombinant TDG

See (Kunz et al., 2009), briefly, TDG wt and TDG_{Δcat} were expressed from vectors pET28c-mTDGa.0 and pET28c-mTDGa.1 as described. Cell lysis was carried out in NiNTA lysis buffer (50 mM Na-phosphate [pH 7.5], 500 mM NaCl, 20% glycerol, 0.1% Tween-20, 20 mM imidazole, 20 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride) by sonication followed by extract clarification. The clear supernatant was loaded onto a 5 ml HisTrap FF crude column (GE Healthcare), bound protein was eluted with 400 mM imidazole and dialyzed against Heparin buffer (25 mM Na-phosphate [pH 7.0], 250 mM NaCl, 20% glycerol, 20 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride). The dialyzed fractions were loaded onto a 5 ml HiTrap Heparin HP column (GE Healthcare) and bound protein was eluted with a linear gradient of 250 mM – 1.5 M NaCl. For ion exchange, relevant fractions were pooled, dialyzed against AIEX buffer (50mM Bicine [pH 8.8], 25 mM NaCl, 20% glycerol, 20 mM β-mercaptoethanol, 0.11 mM phenylmethylsulfonyl fluoride) and loaded onto a 1 ml Resource Q column (GE Healthcare). Bound protein was eluted with a linear salt gradient of 25 mM – 1 M NaCl and purest fractions finally dialyzed against storage buffer (50 mM Tris-HCl [pH 8.0], 50 mM NaCl, 10% glycerol, 1 mM dithiothreitol), frozen on dry-ice and stored at -80°C.

Base release assay

60-mer double-stranded oligonucleotide substrates containing different modifications were prepared by annealing of an unlabeled upper strand oligonucleotide (5'-TAGACATTGCCCTC GAGGTACCATGGATCCGATGTCGACCTCAAACCTAGACGAATTCCG-3') to a (5'-fluorescein-labeled lower oligonucleotide strand 5'-F-CGGAATTCGTCTAGGTTTGAGGTXGACATCGGATCCATGGTACCTCGAGGG CAATGTCTA-3', where X = T, 5mC, 5hmC or 5caC.

Base release assays were carried out in a total volume of 20 μ l containing 0.5 pmol of recombinant protein and 0.5 pmol of the labeled DNA substrate in 1x reaction buffer (50 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1 mM DTT, 1 mg/ml BSA) for 15 min at 37°C. Generated AP-sites were cleaved by the addition of NaOH to a final concentration of 100 mM and heating to 99°C for 10 min. Subsequently, DNA was ethanol precipitated overnight at -20°C in 0.3 M Na-acetate (pH 5.2) and in the presence of 0.4 mg/ml carrier tRNA. The DNA was collected by centrifugation (20 min, 20'000g, 4°C) and washed in 80% ethanol. Air-dried pellets were resuspended in loading buffer (1x TBE, 90% formamide), heated at 99°C for 5 min, and then immediately chilled on ice. Reaction products were separated on 15% denaturing polyacrylamide gels in 1x TBE. The fluorescein-labeled DNA was visualized with a Typhoon 9400 (GE Healthcare) and quantified using the ImageQuant TL software (GE Healthcare).

Electrophoretic mobility shift assay

EMSA were performed to measure the DNA-binding ability of wild-type and mutant TDG protein, using the double-stranded oligonucleotide substrates described above. Standard EMSA were carried out in a total reaction volume of 10 μ l containing 2 pmol of recombinant

protein and 1 pmol of labeled DNA substrate with varying amounts of unlabeled competitor DNA in 1x reaction buffer (50 mM Tris-HCl [pH 8.0], 1 mM DTT, 5% glycerol, and 1 mM EDTA). After 15 min incubation at 37°C the reactions were loaded immediately onto 6% native polyacrylamide gels and separated in 0.5x TBE for 50 min at 100 V at room temperature. The fluorescein-labeled DNA was also visualized with a Typhoon 9400 (GE Healthcare) and quantified using the ImageQuant TL software (GE Healthcare).

DNA Immunoprecipitation and GLIB

Genomic DNA was prepared from cells by incubation in lysis buffer (20 mM Tris-HCl pH 8.0, 4 mM EDTA, 20 mM NaCl, 1% SDS and 1 mg ml⁻¹ proteinase K) at 55°C for 8-12 h and subsequent phenol/chloroform extraction and Na-acetate/ethanol precipitation. DNA pellets were resuspended in 10 mM Tris-HCl pH 8 and concentration was measured by absorbance at 260 nm. RNA was removed by incubation with 2.5 µg RNaseA per µg DNA for 30 min at 37°C, followed by Na-acetate/ethanol precipitation. Quality of the DNA tested by standard agarose gel electrophoresis.

5-mC and 5-caC were detected by MeDIP and caCDIP, performed essentially as described in (Weber et al., 2005). DNA was sonicated to yield fragments of 100-500bp followed by NaCl (400 mM)/ethanol precipitation in the presence of glycogen-carrier. 1 µg fragmented DNA in TE was denatured and incubated with 0,3 µg of monoclonal anti-5-methylcytidine or 2 µg polyclonal anti-5-carboxylcytosine antibody (Supplementary Table 3) at 4°C for 2 h in 1x immunoprecipitation (IP) buffer (10 mM sodium phosphate pH 7.0, 140 mM NaCl, 0.05% Triton X-100). Immuno-complexes were precipitated by the addition of 20 µl M-280 sheep anti-mouse IgG antibody coupled Dynabeads (Invitrogen) and incubation at 4°C for 2 h followed by three washes in IP buffer. Bound material was treated with 250 µl proteinase K digestion buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5% SDS and 0.25 mg ml⁻¹ proteinase K) at 50°C for 3 h. Immunoprecipitated methylated DNA was purified by

phenol/chloroform extraction followed by NaCl/ethanol precipitation and re-suspended in 10 mM Tris-HCl pH 8.0.

5-hmC containing DNA fragments were captured with the Hydroxymethyl Collector kit from Active Motif as described in the manufacturer's instructions.

qPCR analysis of sonicated genomic input DNA and Me/cacDIP/GLIB DNA with target specific primers (Supplementary Table 4) was performed using Quantitect SYBR Green (Qiagen) with a Rotor-Gene 3000 thermocycler (Qiagen). Statistical analysis was performed on Graphpad Prism Software.

Chromatin Immunoprecipitation

To crosslink protein-bound DNA, ES cells were incubated in 1% formaldehyde/PBS at room temperature. The reaction was quenched after 10 min by addition of glycine to a final concentration of 125 mM. After washing three times with ice cold PBS, cells were collected using a cell scraper and subsequent centrifugation at 600xg and 4 °C. Supernatant was discarded and the cells snap-frozen until further processing. After thawing on ice, nuclei were isolated by incubation in 400 µl of cold ChIP Lysis Buffer I (10 mM HEPES pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, 1 mM PMSF) for 5 min on ice followed by two incubations of 5 min on ice in 400 µl cold ChIP Lysis buffer II (10 mM HEPES pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 200 mM NaCl, 1mM PMSF). All centrifugation steps were conducted at 600xg and 4°C for 5 min. Pelleted nuclei were lysed in 400 µl ChIP Lysis buffer III (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5% Triton X-100, 1% SDS, 1 mM PMSF) for 10 min on ice followed by sonication for 15 min (15 sec on, 30 sec off, power high) using a Bioruptor sonicator (Diagenode) to yield fragments of ~200-500 bp. The solution was cleared of remaining cell debris by centrifugation at 14'000xg and 4°C for 10 min. For ChIP of TDG and TET1, 150 µg of chromatin were diluted 1:10 in ChIP dilution buffer (50 mM Tris-HCl pH 8.0,

1 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 1x protease inhibitor cocktail, 1 mM PMSF). After removing 1% (volume) for input analysis, diluted chromatin was pre-cleared at 4°C for 1 h with 30 µl of a 50% slurry of magnetic Protein G beads (Invitrogen) pre-blocked with 1 mg ml⁻¹ BSA and 1 mg ml⁻¹ tRNA. Pre-cleared chromatin was incubated with 1-2 µg of the respective antibody (Supplementary Table 3) overnight at 4°C under slow rotation. Immuno-complexes were precipitated with 40 µl of a 50% slurry of blocked Protein G beads and further incubated at 4°C for 2 h. Beads were then serially washed with 500 µl ChIP wash buffer I (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100), twice with 500 µl ChIP wash buffer II (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100). After two additional washes with 500 µl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), bound complexes were eluted by two sequential incubations with 250 µl elution buffer (1% SDS, 0.1 M NaHCO₃) at 65°C for 10 min shaking. Crosslink reversal of eluates and respective input samples was done in the presence of 200 mM NaCl at 65°C for 4 h followed by proteinase K digestion (50 µg ml⁻¹) in the presence of 10 mM EDTA and 40 mM Tris-HCl pH 6.5 at 45°C for 1 h. DNA was purified by phenol/chloroform extraction and NaCl/ethanol precipitation, and resuspended in 10 mM Tris-HCl pH 8.0. qPCR analysis with target specific primers (Supplementary Table 4) was performed using Quantitect SYBR Green (Qiagen) with a Rotor-Gene 3000 thermocycler (Qiagen). Statistical analysis was performed on Graphpad Prism Software.

Western Blot analyses

Denaturing protein extracts were prepared by washing the ES cells twice in cold PBS before addition of lysis buffer (50 mM Tris-HCl pH 7.5, 1% SDS, 5 mM DTT). The lysate was collected using a cell scraper and processed by two cycles of heating to 65°C and sonication for 5 min (15 sec on, 30 sec off, power high), followed by 10 min centrifugation at 20'000xg and 4°C. The concentration of the supernatant was estimated by a standard Bradford assay by diluting the extract 1:800 in ddH₂O before adding Bradford reagent (final volume 1 ml).

40 µg of protein extract was separated on a 10% PAA gel (for AID) or a Mini-Protean pre-cast gradient gel (BioRad) and transferred to a nitrocellulose membrane (Millipore). For TET1, TET2 and TDG, 10% methanol and 0.002% SDS were added to the transfer buffer (25 mM Tris, 192 mM Glycine), for AID no SDS but 20% methanol. Membranes were washed once with TBS-T (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.2% Tween-20), blocked with blocking buffer (TBS-T, 10% low fat dry milk) at room temperature for 1 h and incubated with the primary antibody at 33°C (anti-mTDG) or room temperature (anti-TET1, anti-TET2, anti-AID, anti-DNMT3b, anti-β-actin) for 1 h in 7.5% dry milk/TBS-T. Dilutions were 1:10'000 for the rabbit anti-mTDG and the mouse anti-β-actin antibodies; 1:2'000 for the rabbit anti-TET1 (Millipore) antibody; 1:500 for the monoclonal mouse anti-AID (gift by S.K. Petersen-Mahrt) and 1:100 for the monoclonal rat anti-TET2 antibody (gift by H. Leonhard). Washing steps after hybridization were once at 33°C and twice at room temperature for 15 min for anti-mTDG, or three times at room temperature for 10 min for all other antibodies. Membranes were incubated with secondary HRP-conjugated antibodies diluted 1:5'000 (goat anti-rabbit and goat anti-mouse) or 1:20'000 (anti-rat) in 5% dry milk/TBS-T at room temperature for 1 h. After three washing steps of 10 min at room temperature, detection of the signals was performed using the WesternBright Quantum Chemiluminescent HRP Substrate (Advansta).

Quantitative RT-PCR

1 µg total RNA extracted with TRI Reagent (Sigma) was reverse transcribed with the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer's protocol. qPCR with target specific primers (Supplementary Table 5) was performed using Rotor-Gene SYBR Green PCR mix with a Rotor-Gene 3000 thermocycler (Qiagen). Conditions for each target were validated by standard and melting curve analyses. Target-specific amplifications were normalized to the average of TBP, B2m and β-actin. Data of three independent experiments were analyzed by Anova to test for differences between genotypes.

Figure Legends

Figure 1: Differentiated *Tdg*^{-/-} cells accumulate aberrant methylation patterns. a) MeDIP-sequencing revealed an increasing number of differentially methylated regions (DMRs) arising with *in vitro* differentiation of *Tdg*^{+/-} and *Tdg*^{-/-} ES cells to neuronal progenitors. b) Volcano-plot of methylation fold change versus FDR-adjusted p-value (narrow peaks), CpG islands (CGIs) in red. 99% of the DMRs that overlap with a CGI are hypomethylated in knockout NPs, non-overlapping ones (other) are mostly hypermethylated. c) Overlap of CGIs hypomethylated in knockout versus wildtype with CGIs that become *de novo* methylated with differentiation; $p < 0.0001$, Chi-square with Yates correction. Hypomethylation at these CGIs is caused by diminished differentiation-driven methylation. d) Overlap of broad peaks with published datasets. To test for enrichment or depletion of a feature in the hypomethylated CGIs, we compared the proportion of CGI DMRs overlapping (positive) or not overlapping (negative) with a feature to the analogous proportion within CGIs that were not differentially methylated. Percentages of CGIs overlapping (positive, grey) or not overlapping (negative, white). ** $p < 0.005$; *** $p < 0.0001$, Chi-square with Yates correction.

Figure 2: Hairpin bisulfite-sequencing of representative CGI DMRs in NPs. Strand-specific methylation (5-mC and 5-hmC) and mutation analysis of CGI DMRs, specified with characteristics in Supplementary Table 1, in *Tdg* knockout NPs complemented with wildtype TDG (wt), a catalytically dead mutant (TDG_{Δcat}) or the empty complementation vector (ko). Insulin growth factor 2 (*Igf2*) served as a control (Arand et al., 2012). The bars indicate average percentage of fully methylated, hemi-methylated and mutated CpGs, the according numbers are presented in the tables. The heat maps display neighboring CpG dinucleotides, and each line represents one sequencing read. Catalytic activity of TDG is essential to establish methylation. Hypomethylation at these CGIs appears to not be caused by deamination of 5-mC, as the rate of C→T mutation is within the error rate of the method.

Figure 3: Global levels of 5-mC and its derivatives in a 24 h differentiation timecourse.

a) Scheme of experimental setup. ES cells preconditioned in 2i medium were seeded in ESC medium (ECM) with LIF 14-16 h prior to differentiation, then transferred to ECM without LIF and with 5μM retinoic acid (RA). Samples were harvested at the indicated timepoints. b) LCMSMS analysis of global C-modification and 5-hmU levels in genomic DNA prepared at 0, 8 and 24 h of differentiation. We observe a significant rise of the global levels of 5-mC, 5-fC and 5-caC with differentiation in knockout and TDG_{Δcat} (for statistics see Fig. S3). Shown are Log2 fold changes compared to wildtype (mean with s.e.m.); statistical analysis was performed on absolute numbers (Fig. S3), asterisks indicate significant difference to the respective timepoint in wt, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0001$, one-way Anova.

Figure 4: Targeted MeDIP, GLIB and caCDIP analysis of CGI DMRs. a) Relative enrichment (RE) of 5-mC (n=2) and 5-caC (n=3) normalized to a randomly chosen CpG-poor

region (neg.contr.). TDG and TDG_{Δcat} accumulate 5-mC and 5-caC over time, 5-caC especially accumulates in TDG_{Δcat}. Mean with s.e.m. Asterisk, significant difference of RE between 0 h and 24 h within a genotype, * p<0.05, one-way Anova. Statistical comparison of genotypes with regard to the log2-transformed 24 h versus 0 h fold change across targets: aggregated p-value < 0.0001. b) Biochemical analysis of TDG and TDG_{Δcat} glycosylase activity on 5-mC derivatives. Double-stranded 60mer oligonucleotide substrates containing a single T, 5-mC, 5-hmC or 5-caC on a fluorescence labeled strand were incubated with recombinant wildtype TDG or TDG_{Δcat}. Glycosylase activity on the indicated substrates produces an abasic site which is converted to a single-strand break by heating under alkaline conditions, giving rise to a shorter fragment of 23 nt. Shown are reaction products separated on a denaturing polyacrylamide gel and quantification of 3 independent experiments (mean with standard deviation). c) Electrophoretic mobility shift assay with labeled 60mer oligonucleotides with the indicated modification (Substrate*), incubated with wildtype TDG or TDG_{Δcat} and varying amounts of unlabeled competitor substrate containing an unmodified C or 5-caC. See also text.

Figure 5: TET1 fails to stably associate with CGI DMRs during differentiation in absence of TDG activity. TET1-ChIP on chromatin prepared at 0, 8 or 24h of differentiation. a) Log2 fold changes of relative enrichment (RE) at 8 or 24 h versus the respective RE at 0 h (median with range). TET1 loses affinity to the CGI DMRs with ongoing differentiation in absence of TDG activity. For RE values and controls, refer to Fig. S5. Asterisks, significant difference between the genotypes with regard to their fold change versus 0h, *p<0.05, ** p<0.005, one-way Anova with Bonferroni post-test. b) TET1 and β-actin protein levels in SDS protein extracts detected by Western blot. c) TDG-ChIP, analogous to a. Log2 fold changes of relative enrichment (RE) versus the respective relative enrichment in TDG ko cells serving as a background control (median with range). TDG_{Δcat} is enriched at CGI DMRs due to its high affinity to 5-caC present at these loci. Wildtype TDG is capable of rapid turnover on 5-caC and appears to be associated very transiently with the CGI DMRs. For RE values and statistical analysis, refer to Fig. S6. d) TDG and β-actin protein levels in SDS protein extracts detected by Western blot.

Figure 6: Model for a dual function of TDG in a differentiation-driven DNA methylation and demethylation cycle. At the onset of differentiation, a cycle of DNA methylation and demethylation is triggered. CpG methylation is catalyzed by Dnmts. 5-mC is oxidized by the TET proteins in a stepwise manner and the final products 5-fC/caC are excised by TDG (wt, top). With ongoing differentiation, the equilibrium of the different C-modifications is shifted towards 5-mC, evident as CGI *de novo* methylation in early NPs (“normal methylation”). We propose that apart from its active function in catalyzing the final step of demethylation TDG additionally provides a structural scaffold to allow the recruitment of the key factors involved. In absence of TDG, initiation of the cycle fails or stops immediately after initiation (*Tdg* ko, middle), resulting in lower methylation levels in ko NPs than in wt NPs (“hypomethylation”). TDG_{Δcat} on the other hand provides the scaffold for assembly and coordination of the different steps of the cycle but upon binding to 5-caC fails to turn over (bottom). The lacking catalytic activity of TDG leads to the accumulation of 5-caC which is bound with high affinity by TDG_{Δcat}. As the cycle fails to proceed, TET1 association is destabilized. Removal of 5-caC by dilution through DNA replication eventually leads to the hypomethylation observed in NPs.

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11, R106.
- Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., *et al.* (2012). In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet* 8, e1002750.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* 21, 381-395.
- Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463, 1042-1047.
- Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91-99.
- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol Chapter 19*, Unit 19 10 11-21.
- Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat Immunol* 4, 235-240.
- Chavez, L., Jozefczuk, J., Grimm, C., Dietrich, J., Timmermann, B., Lehrach, H., Herwig, R., and Adjaye, J. (2010). Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. *Genome Res* 20, 1441-1450.
- Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007). The enigmatic thymine DNA glycosylase. *DNA Repair (Amst)* 6, 489-504.
- Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., *et al.* (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* 470, 419-423.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., *et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146, 67-79.
- Crain, P.F. (1990). Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol* 193, 782-790.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 107, 21931-21936.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fritze, S., Harrow, J., Kaul, R., *et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.
- Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J Mol Biol* 196, 261-282.
- Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., *et al.* (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15, 1451-1455.
- Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11, R86.
- Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74, 481-514.
- Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339, 448-452.

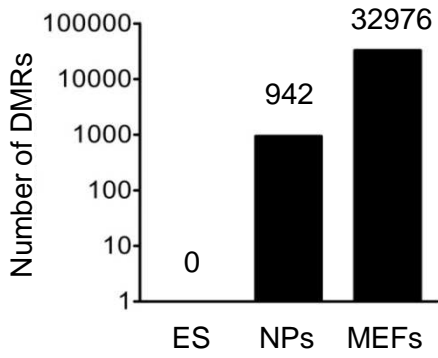
- Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem* 275, 33449-33456.
- Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2003). The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic Acids Res* 31, 2261-2271.
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* 40, 4841-4849.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303-1307.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311-318.
- Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. *FEBS Lett* 583, 1713-1720.
- Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res* 21, 1670-1676.
- Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* 108, 3642-3647.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300-1303.
- Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. *Nature* 452, 112-115.
- Kemmerich, K., Dingler, F.A., Rada, C., and Neuberger, M.S. (2012). Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung^{-/-}Msh2^{-/-} mice. *Nucleic Acids Res* 40, 6016-6025.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* 12, 996-1006.
- Kunz, C., Focke, F., Saito, Y., Schuermann, D., Lettieri, T., Selfridge, J., and Schar, P. (2009). Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol* 7, e91.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 16, 2893-2905.
- Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., *et al.* (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol* 20, 311-316.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
- Lutsik, P., Feuerbach, L., Arand, J., Lengauer, T., Walter, J., and Bock, C. (2011). BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing. *Nucleic Acids Res* 39, W551-556.
- Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286, 35334-35338.
- Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766-770.

- Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., *et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452, 45-50.
- Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30, 755-766.
- Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* 8, 751-758.
- Nix, D.A., Courdy, S.J., and Boucher, K.M. (2008). Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics* 9, 523.
- Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101-1105.
- Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135, 1201-1212.
- Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci* 368, 20110330.
- Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Paliarne, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L., *et al.* (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* 40, 8255-8265.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., *et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490-495.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935.
- Valinluck, V., and Sowers, L.C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 67, 946-950.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37, 853-862.
- Wilson, G.A. (2012). Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers. *Gigascience* 1.
- Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M., Lipkin, M., *et al.* (2002). Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* 99, 14937-14942.
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2, 241.
- Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519-523.

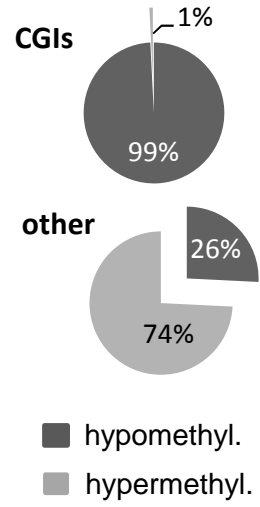
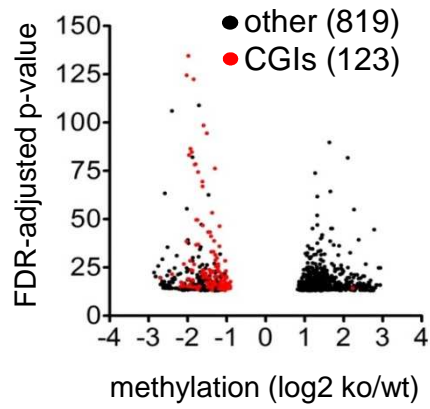
Figure 1

Appendix I

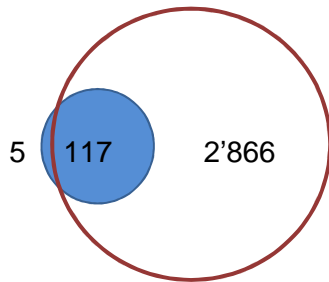
a)



b)



c)



Hypomethylated CGI
ko vs wt

CGIs *de novo*
methylated in NPs

d)

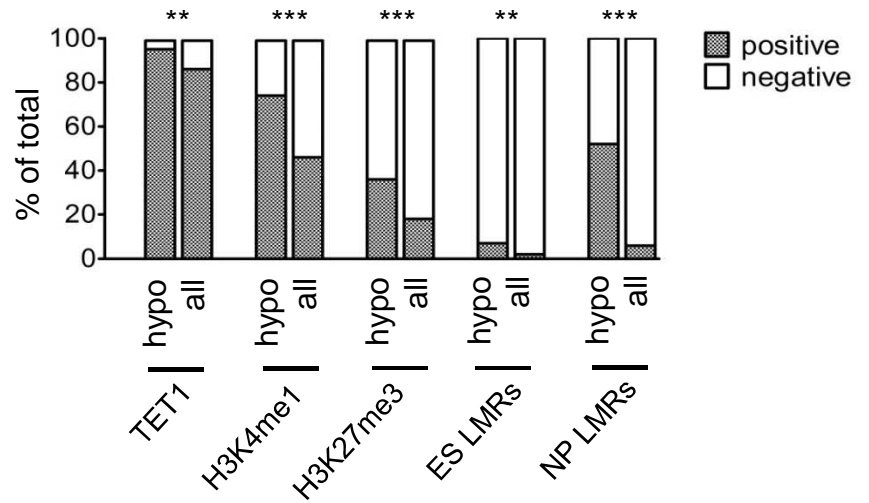


Figure 2

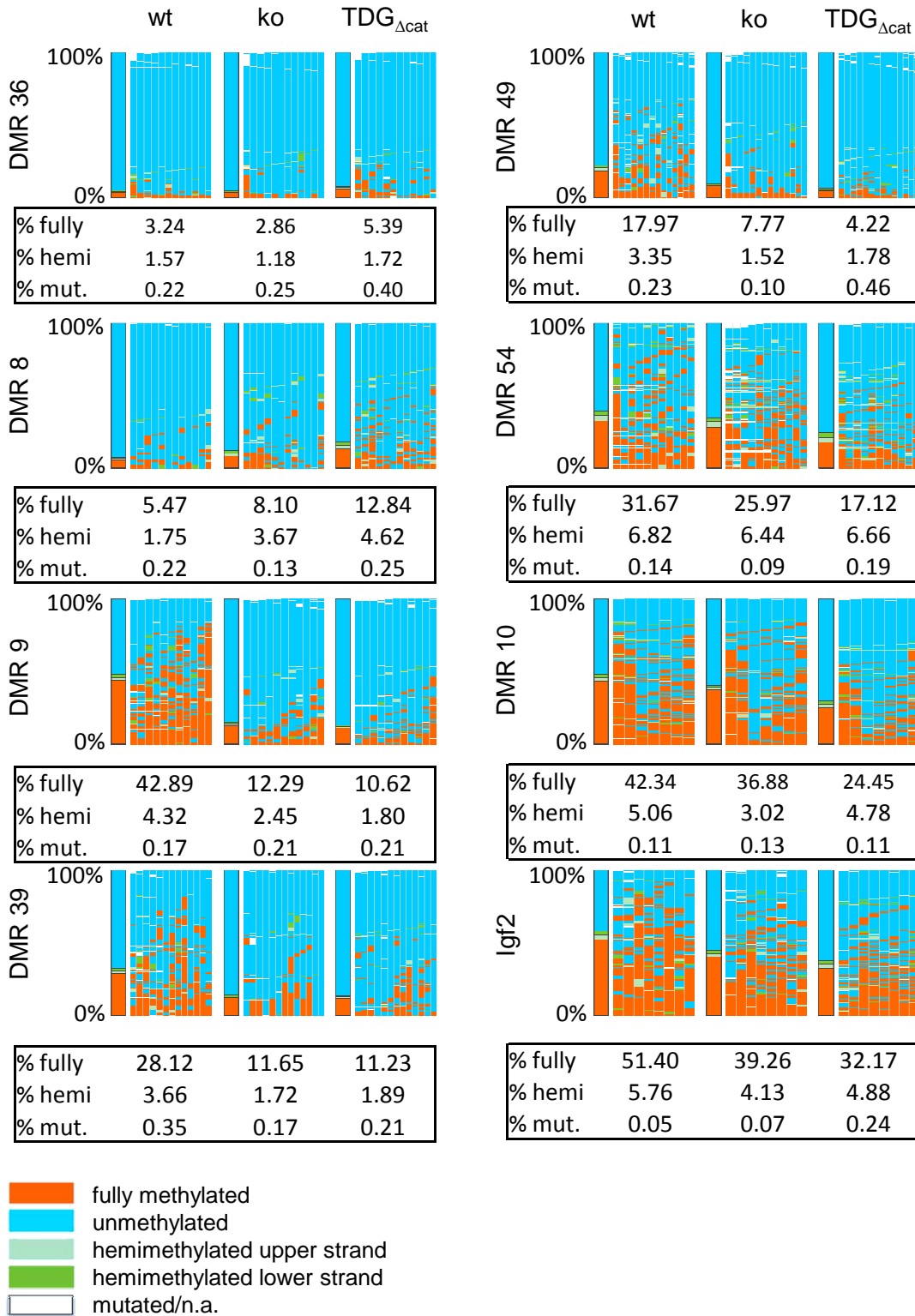


Figure 3

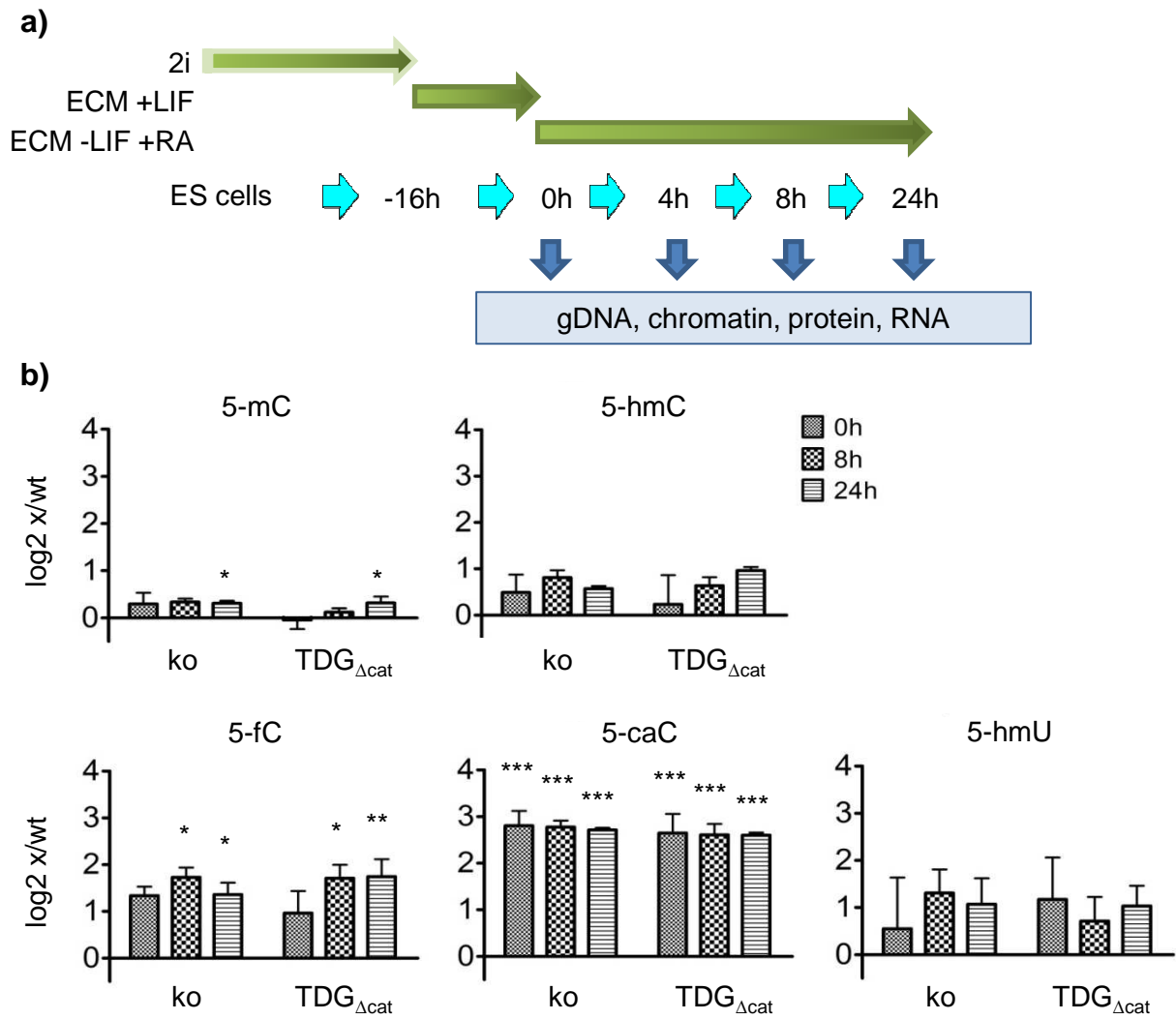


Figure 4

Appendix I

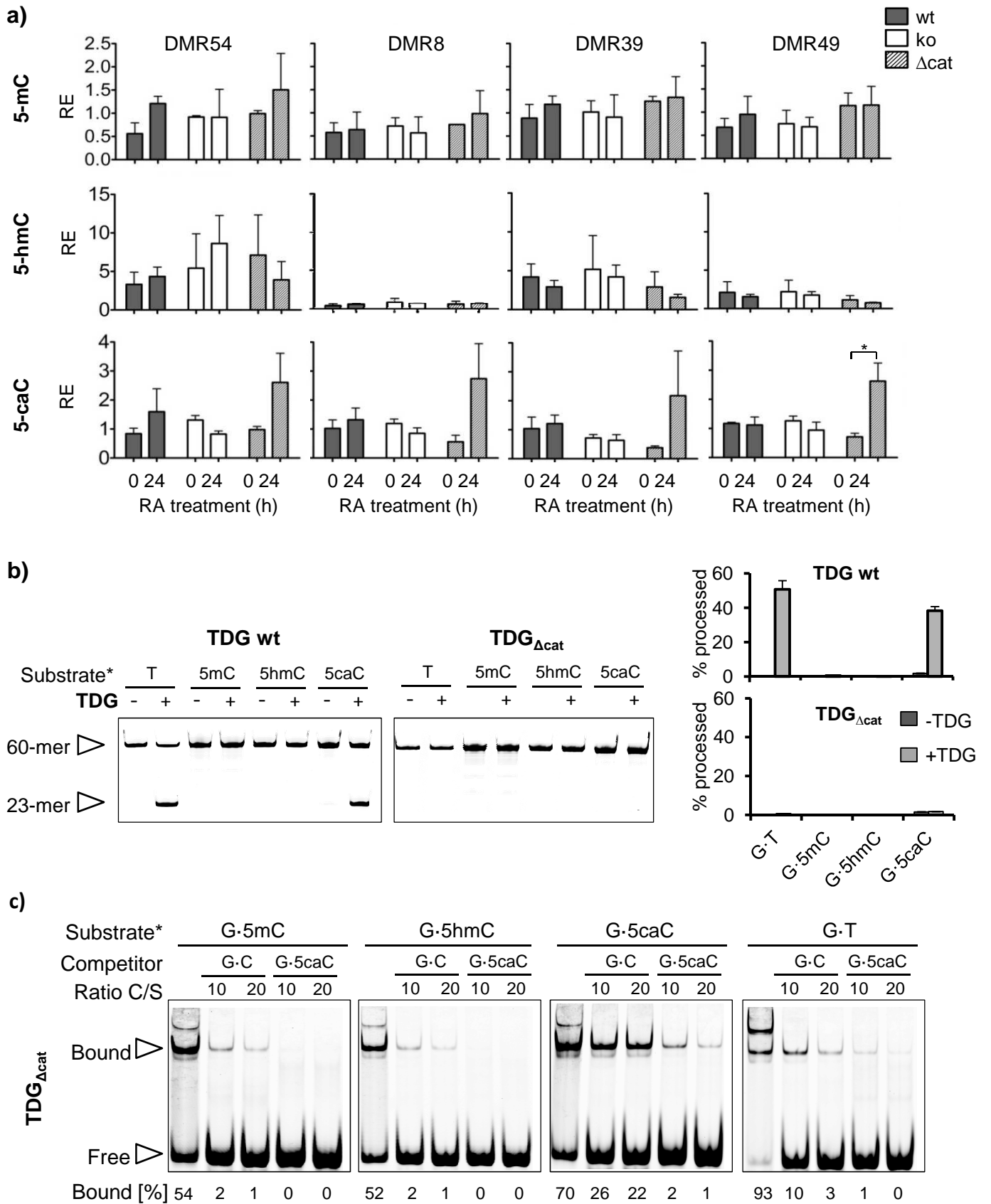


Figure 5

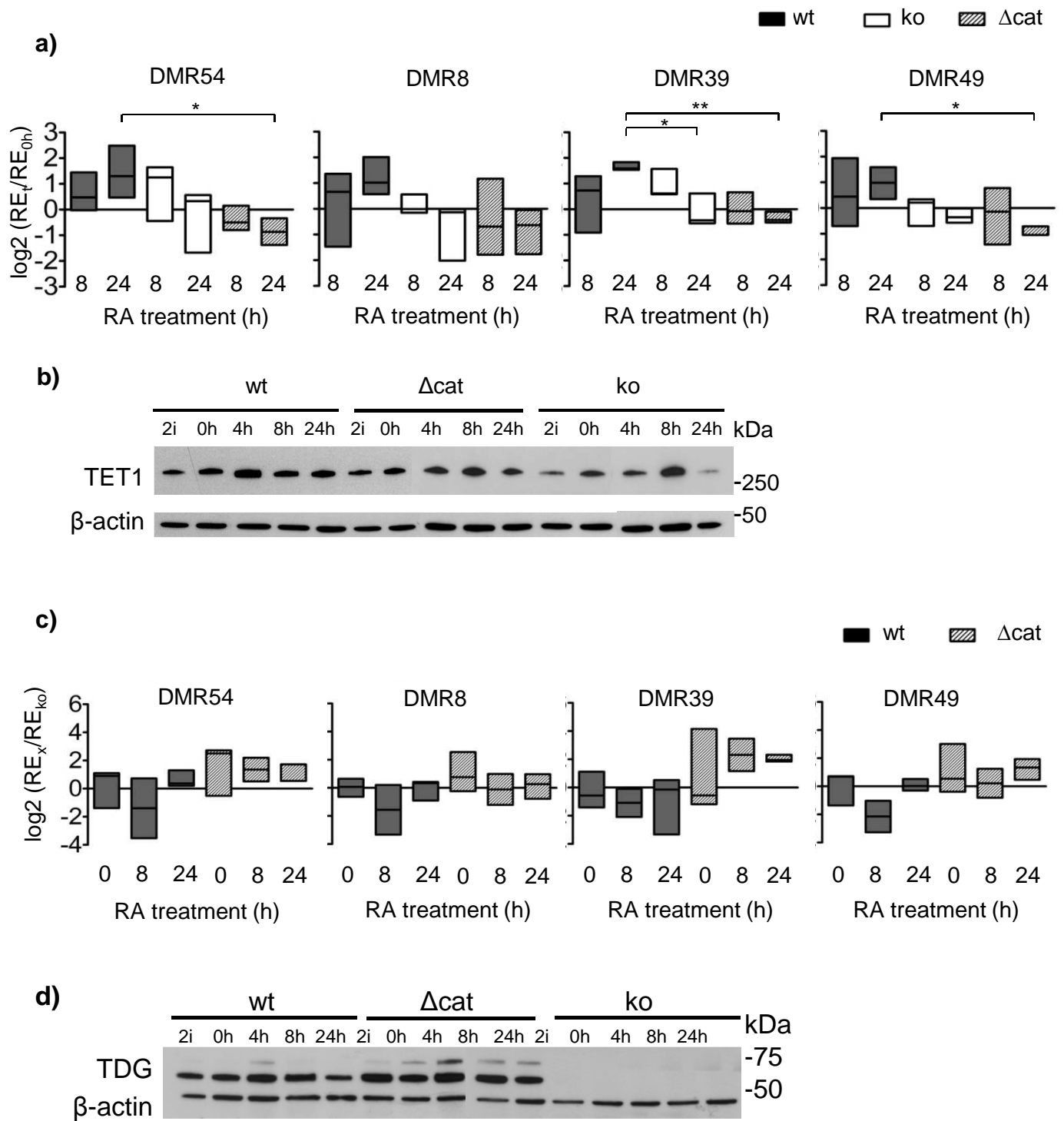
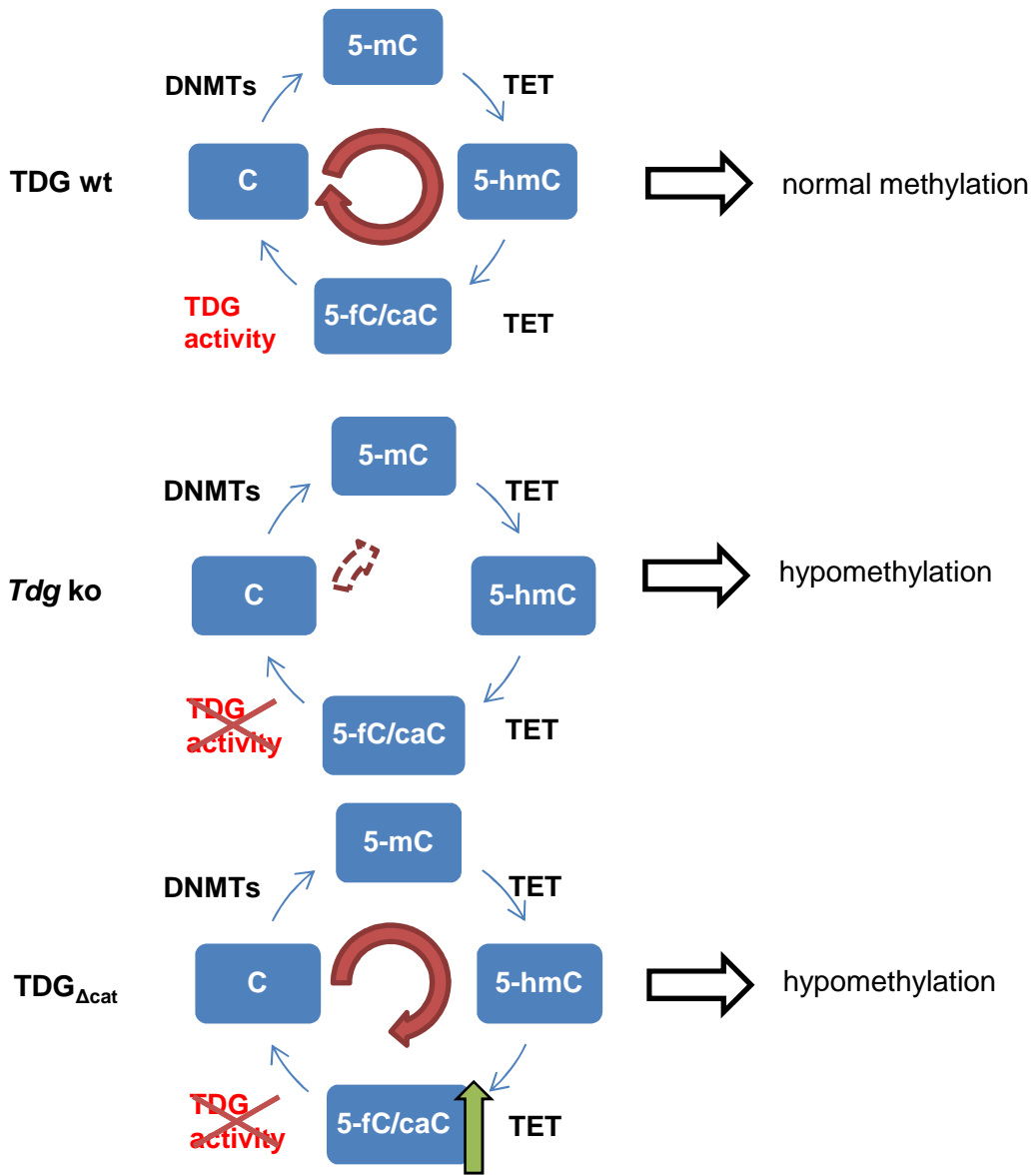


Figure 6



Supplementary Material

TDG balances DNA methylation and oxidative demethylation at CpG islands in differentiating cells

Jacobs, A.L.^{1,*}, Cortázar, D.^{1,*}, Wirz, A.^{1,*}, Arand, J.², Steinacher, R.¹, Broberg Vågbø, C.³, Giehr, P.², Weber, A.¹, Wilson, G.⁴, Galashevskaya, A.³, Kunz, C.¹, Reik, W.⁵, Beck, S.⁴, Walter, J.², Krokan, H.³, Schär, P.¹

¹ Institute of Biochemistry and Genetics, Department of Biomedicine, University of Basel, Switzerland;

² Department of Biological Sciences, Institute of Genetics/Epigenetics, University of Saarland, Saarbrücken, Germany;

³ Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway;

⁴ Medical Genomics, UCL Cancer Institute, University College London, London, United Kingdom;

⁵ Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, United Kingdom

* These authors contributed equally to this study

Correspondence:

primo.schaer@unibas.ch; Tel. +41 61 267 35 61; Fax: +41 61 267 35 66

Supplementary Figure Legends

Figure S1: Characterization of NP DMRs

a) *In vitro* differentiation of *Tdg^{+/+}* and *Tdg^{-/-}* ES cells to neuronal progenitors; *Tdg^{-/-}* cells rapidly lose viability at later stages of NP differentiation (24-48h). b) DMRs that are hypomethylated in knockout NPs are associated with higher CpG density, whereas CpG poor DMRs are almost exclusively hypermethylated; CpG density and methylation fold change refer to narrow peaks. c) Density plot of the log₁₀ distance of hypo- or hypomethylated NP DMRs (center of broad peaks) to the nearest TSS (Ensembl), ***p*<0.005, Mann-Whitney T-test on average of distances (linear). d) Promoter-association of hypo- versus hypermethylated NP DMRs (broad peaks). e) CGI DMRs (hypo) or not differentially methylated CGIs (all) overlapping (positive) or not overlapping (negative) with sites of binding or enrichment of the indicated factor. ** *p*<0.005; *** *p*<0.0001, Chi-square with Yates correction.

Figure S2: mRNA and protein levels in a 24h differentiation timecourse. a) mRNA levels of the pluripotency markers Oct4, Nanog and Rex1, and the differentiation-induced transcription factor Gata6, detected by qRT-PCR, confirm loss of pluripotency within 24h of differentiation. All three cell lines appear to differentiate with similar efficiency. TET1 mRNA levels decrease towards 24h but protein levels remain stable throughout this time window and are equal between cell lines (Fig.6b). mRNA levels were normalized to the average of TBP, B2m and β-actin. b) TET2 mRNA and protein levels, detected by qRT-PCR and Western blot, respectively, show no differences between the cell lines. Note that mRNA levels of TET2 are 5-10 times lower than those of TET1. c) mRNA and protein levels of AID. AID mRNA levels are extremely low and protein levels are below the detection limit of Western blot. Activated B-cells served as a positive control for detection of AID in Western blot. Shown are means with s.e.m.; statistical analysis by one-way Anova revealed no significant differences across genotypes.

Figure S3: Global C-modification levels measured by LCMSMS at 0, 8 and 24h of differentiation. a) Absolute numbers per 10⁶ unmodified nucleotides. b) 5-mC levels decrease with culturing in 2i medium. Without 2i, n=1, with 2i n=3. c) Control for unspecific effects, 0h or 24h incubation with DMSO; n=2. DMSO does not induce any significant change of C-modification levels. Error bars, s.e.m.; statistical test between time points, * *p*<0.05, ** *p*<0.005, *** *p*<0.0001, Anova.

Figure S4: Targeted analysis of 5-mC, 5-caC and 5-hmC at CGI DMRs. a) 5-mC and 5-caC proportions of the sum of both average RE. In *Tdg_{Δcat}* the equilibrium of 5-mC and 5-caC is tipped towards 5-caC upon differentiation.

Figure S5: TET1-association at CGI DMRs. Relative enrichment normalized to a randomly chosen CpG-poor region (neg.contr.). Bars indicate the mean, error bars the s.e.m. The promoter regions of Oct4, Nanog and HoxA10 served as control regions. See also Fig.7.

Figure S6: TDG-association at CGI DMRs. Relative enrichment normalized to a random CpG-poor region (neg.contr.). Bars indicate the mean, error bars the s.e.m. The promoter regions of Oct4, Nanog and HoxA10 served as control regions. Statistical analysis, two-way Anova: genotypes significantly different, 3 $p < 0.05$, ** $p < 0.005$. See also Fig.8.

Supplementary Tables

Supplementary Table 1: Targets of hairpin BS-seq with characteristics

DMR	genic/intergenic	gene	CGI	genomic features	NP vs ES
8	5' exon	Tbx3	weak	TET1	hyper
9	intron	Kdm2b	strong	TET1, H3K4me1, H3K4me3, NP-LMR	hyper
10	3' exon	Zfp282	strong	TET1, H3K4me1, H3K4me3	hyper
36	intron	Gm5089	strong	TET1, H3K4me1, H3K4me3, NP-LMR	hyper
39	5' exon	Ldoc1l	strong	TET1, H3K4me1, H3K4me3, NP-LMR	hyper
49	intergenic	-----	strong	TET1, H3K4me1, H3K4me3, NP-LMR	hyper
54	exon	Mgat4b	strong	CTCF, H3K4me1, H3K4me3, NP-LMR	hyper

Supplementary Table 2: Enzymes and primers used in hairpin BS-sequencing

DMR	restriction enzyme	Primer sequences
8	MspI	F GATAAGGATATTGAGTTAGAGGA
		R AAAAACACTAAACCAAAAAAC
9	TaqI	F TTTTAGGAGATATAAAGAATAGTTT
		R AAAAACACAAAAACAACCTC
10	PstI	F AGAAGAGTTTTAATTGTTATTTTGG
		R AAACCTCAACTACCACTCTAACC
36	BamHI	F TTTGATATTTTTTTTAGTTTT
		R CCTAACACTTTCTCTTAATTT
39	TaqI	F GGATGTAGGTATTGATTAT
		R ACCTACCAAACCTTTACAA
49	TaqI	F GTGTATAGTTGGGTTTGTAGTG
		R TAAAAAACTAAAATATCCCCTC
54	MspI	F TAGGATTGTGTTGTTTTAGATTT
		R CACCTATACCTTTCTCAACCA

Igf2 according to (Arand et al. 2012).

Supplementary Table 3: Antibodies used in this study

Antibody	Product Nr.	Manufacturer
Anti-5mC monoclonal antibody 33D3	Mab-081-100	Diagenode
Anti-5-Carboxylcytosine antibody	61225	Active Motif
Anti-Methylcytosine dioxygenase TET1 antibody	09-872	Millipore
Anti- β -actin monoclonal	ab8226	Abcam
Anti-TET2 monoclonal		H. Leonhard
Anti-AID monoclonal 4.26.1		S.K. Petersen-Mahrt
Anti-Dnmt3b	ab2851	Abcam
Anti-TDG L58 polyclonal		our laboratory
Anti-mouse Ig (horse radish peroxidase linked)	NXA931	GE Healthcare
ECL TM Anti-rabbit IgG (HRP linked)	NA934V	GE Healthcare
Anti-rat Ig (HRP linked)	A9037-1ML	Sigma

The anti-TDG antibody used for ChIP was produced and affinity purified in our lab, for further information see (Neddermann et al. 1996; Hardeland et al. 2002).

Supplementary Table 4: Primers used in ChIP, MeDIP, GLIB and caCDIP qPCR

Primer	5'-3' sequence
neg. contr. F	AGC ACA GCC TGA AGC CTC TA
neg. contr. R	ACA CAG CAT GGC ATC TTG AA
DMR 54 F	ACCCAGCAAATCTCACCTG
DMR 54 R	GACACTGGACAGGGCTCCA
DMR 39 F	GAGCTGGATAGCCCTTGTAGAATG
DMR 39 R	TTGGCAGCGGAGGGAGCAG
DMR 8 F	CTGGCCACAGCTTTACCATC
DMR 8 R	AAGGACACTGAGCCAGAGGA
DMR 49 F	GCTGGGTTTGTAGTGGGAAC
DMR 49 R	GCAGGACCACACCTCACATC
Nanog P_2 F	GAGGATGCCCCCTAAGCTTTCCCTCCC
Nanog P_2 R	CCTCCTACCCTACCCACCCCTATTCTCCC
Oct4_PP F	GTGAGGTGTCGGTGACCCAAGGCAG
Oct4_PP R	GGCGAGCGCTATCTGCCTGTGTC
pHoxA10_T1 F	CACTCCCAGTTTGGTTTCGT
pHoxA10_T1 R	GGGGGTACAGGTTCAAGAGC

F forward primer; R reverse primer.

Supplementary Table 5: Primers used in quantitative RT-PCR

Primer	5'-3' sequence
AID RT fw	TTC GGC GCA TCC TTT TGC CCT
AID RT rev	GGC GGT CCT GTG CAG CTC AA
β -actin RT fw	CGT CGA CAA CGG CTC CGG CAT
β -actin RT rev	CCA CCA TCA CAC CCT GGT GCC TAG G
B2m RT fw	TCA CGC CAC CCA CCG GAG AA
B2m RT rev	TCT CGA TCC CAG TAG ACG GTC TTG G
Gata6 RT fw	TCG AAA CGC CGG TGC TCC AC
Gata6 RT rev	CCG TGA TGA AGG CAC GCG CT
Nanog RT fw	CCT TCC CTC GCC ATC ACA CTG ACA
Nanog RT rev	GAG GAA GGG CGA GGA GAG GCA GC
Oct4 RT fw	GTC CCC CAA GTT GGC GTG GAG
Oct4 RT rev	CAT GTC CTG GGA CTC CTC GGG AG
Rex1 RT fw	GGA CTA AGA GCT GGG ACA CG
Rex1 RT rev	TCC TGC TTT TTG GTC AGT GGT
TBP RT fw	CCT AAA GAC CAT TGC ACT TCG TG
TBP RT rev	ACT GAA AAT CAA CGC AGT TGT CC
TET1 RT fw	ACA CAC CTT GGG GCA GGA CCA
TET1 RT rev	TCT GAT CAC CCA CTT GGC GAC C
TET2 RT fw	GGA AGC AAG ATG GCT GCC CTG TA
TET2 RT rev	GAA TGA ATC CAG CAG CAC CGT CCC

Arand, J., et al. (2012). "In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases." *PLoS Genet* **8**(6): e1002750.

Hardeland, U., et al. (2002). "Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover." *EMBO J* **21**(6): 1456-1464.

Neddermann, P., et al. (1996). "Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase." *J Biol Chem* **271**(22): 12767-12774.

Figure S2: mRNA and protein levels, 24h RA timecourse

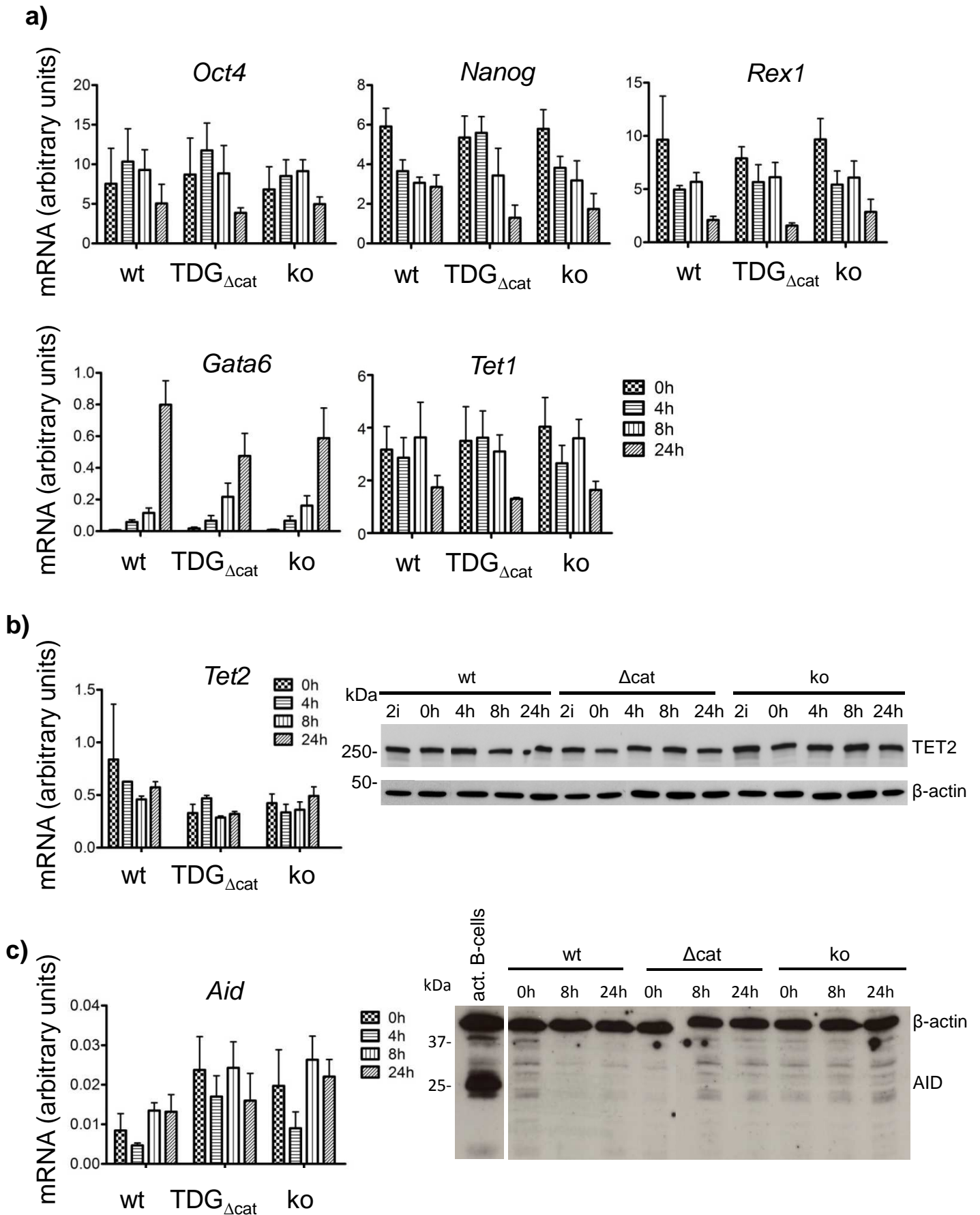


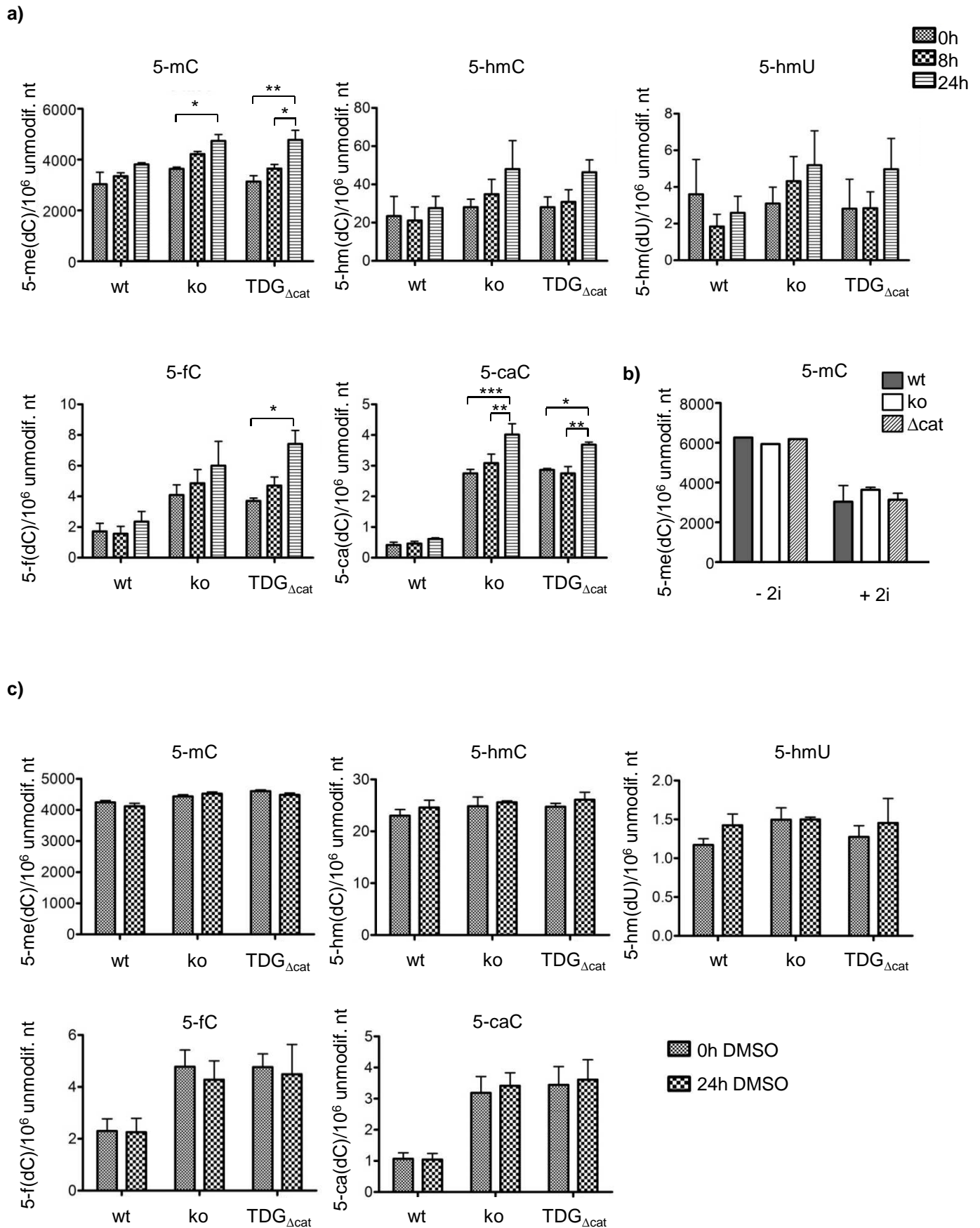
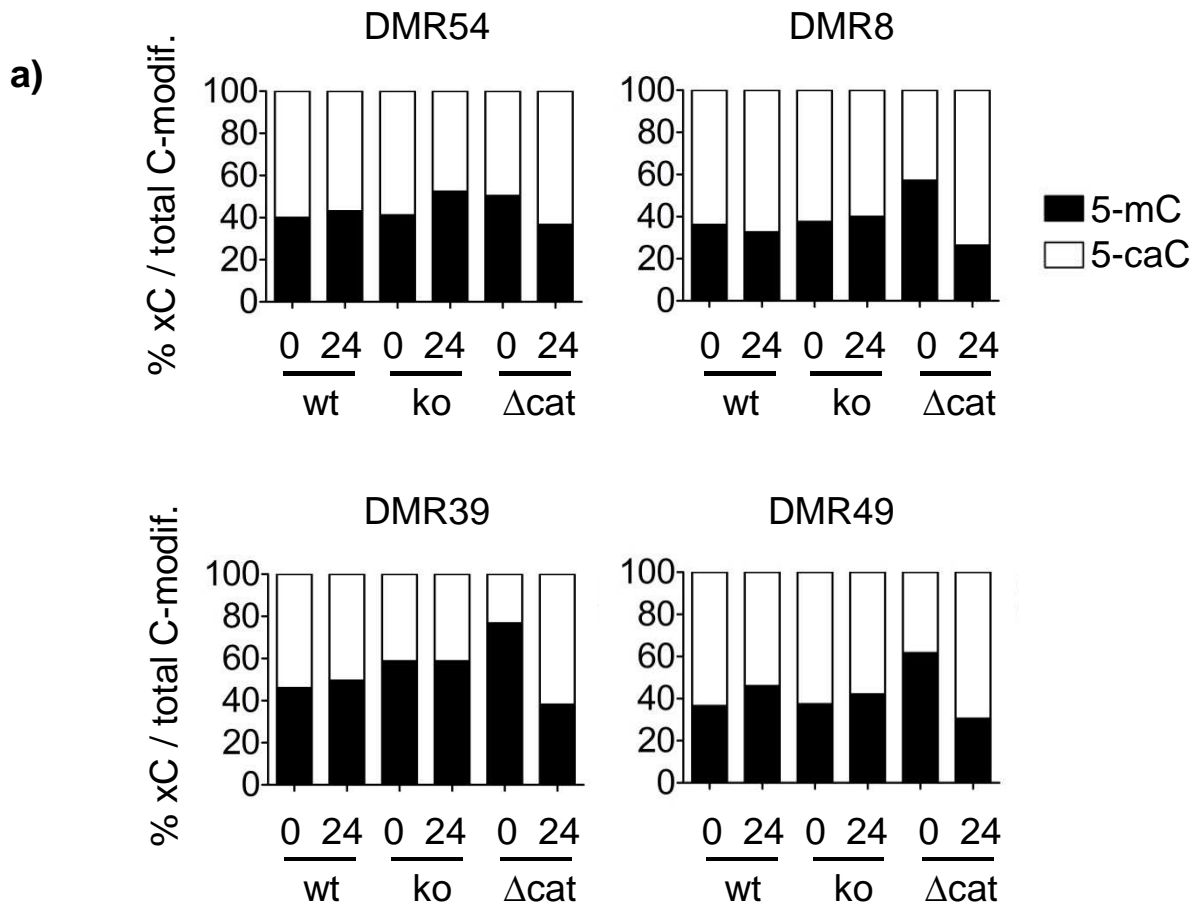
Figure S3: global C-modification levels, absolute numbers per 10⁶ unmodified nt

Figure S4: Targeted analysis of 5-mC, 5-hmC and 5-caC levels at CGI DMRs



b)

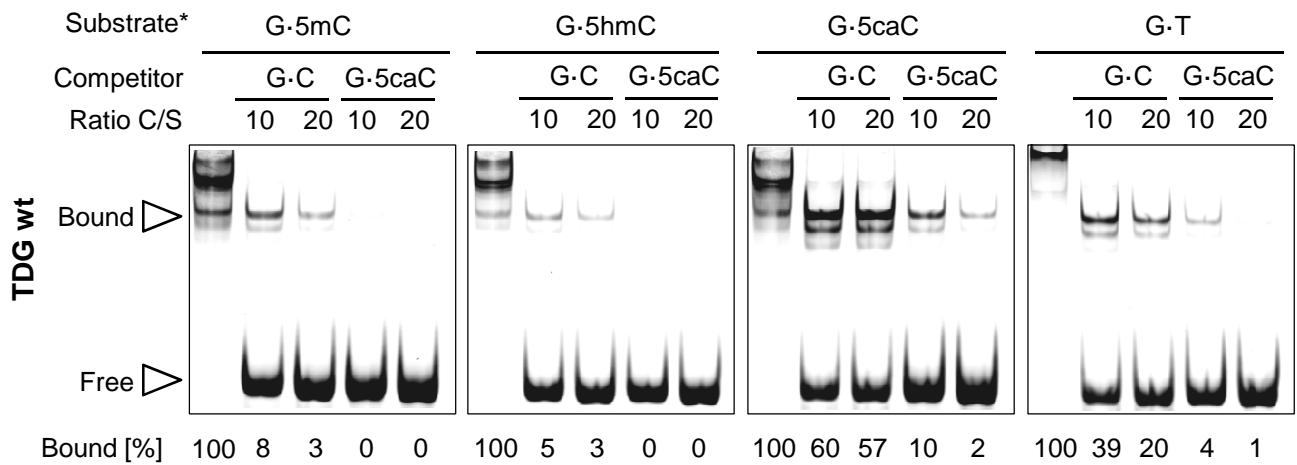


Figure S5: TET1-ChIP, relative enrichment versus chr2neg

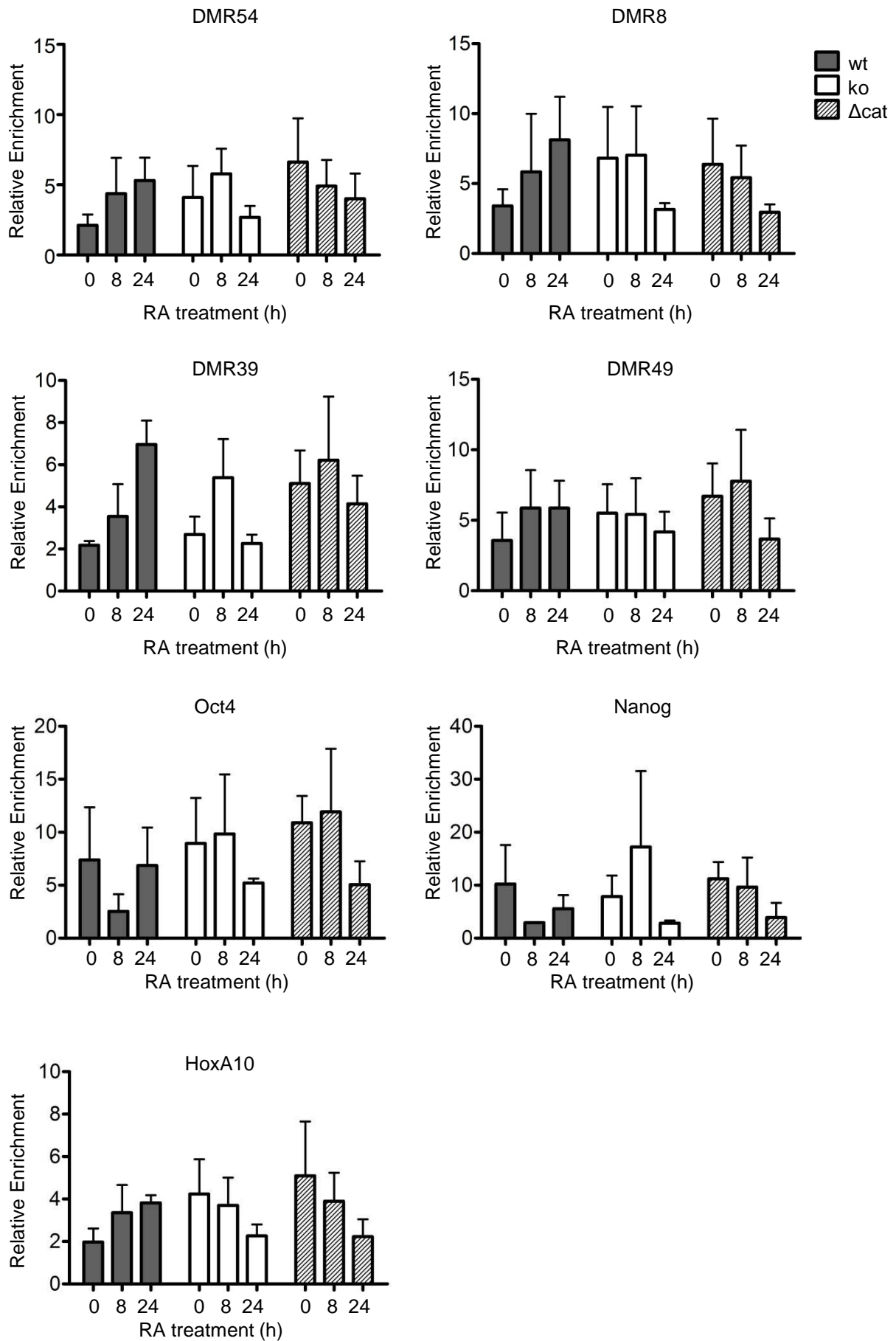
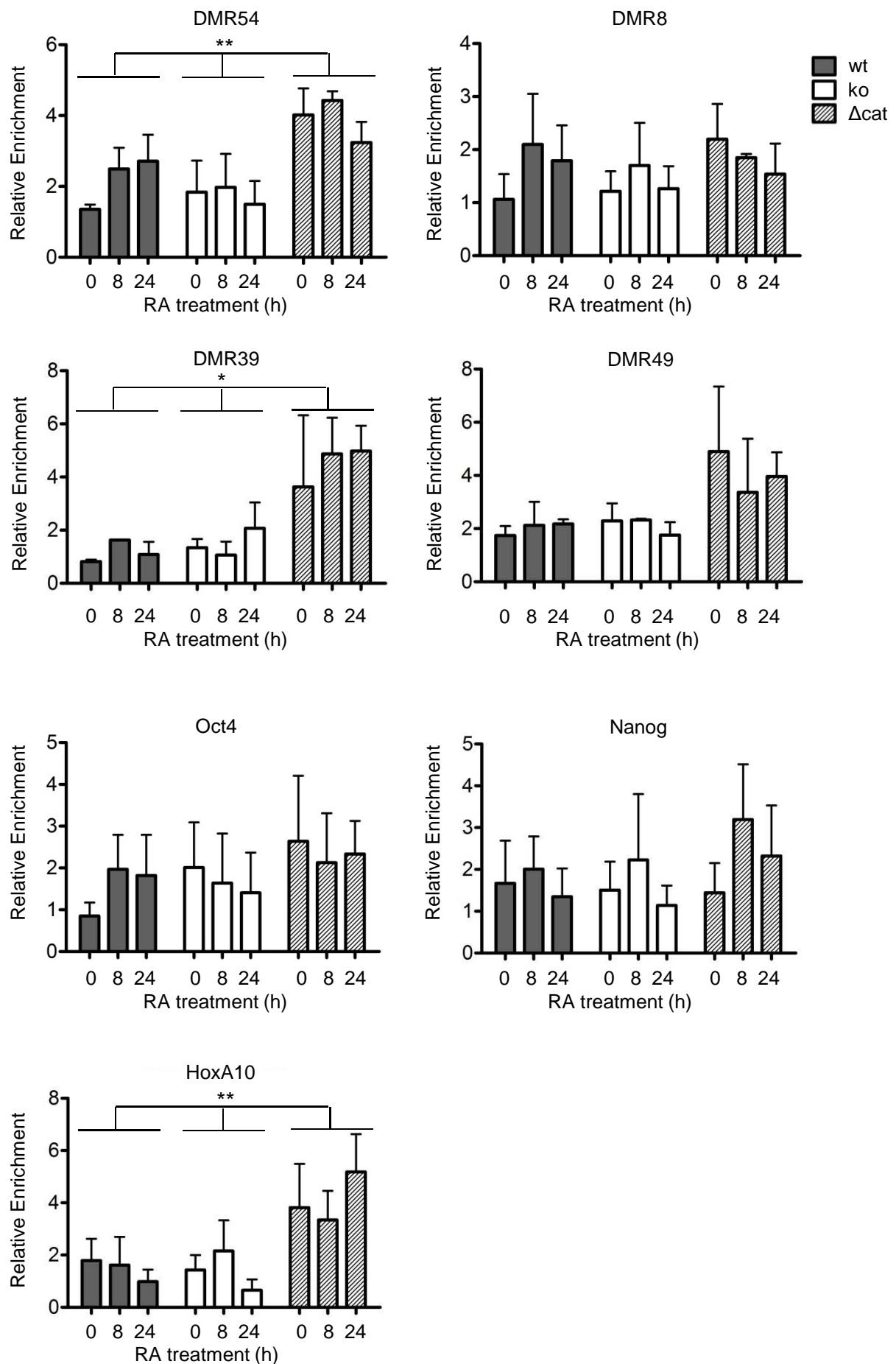


Figure S6: TDG-ChIP, relative enrichment versus chr2neg



TET1, TET2 and TDG Cooperate in a Locus-Specific Manner to Promote Chromatin Plasticity by Oxidative DNA Demethylation

Wirz, A., Noreen, F., Ivánek, R., Barekati, Z., Jacobs, A. L., Kunz, C., Schär, P.*

Department of Biomedicine, University of Basel, Switzerland

* Correspondence:

primo.schaer@unibas.ch; Tel. +41 61 267 3561; Fax: +41 61 267 3566

Abstract

The establishment of developmental stage-specific gene expression programs is crucial for successful mammalian development. DNA methylation and demethylation processes play an important role in this context, and involve, besides DNA methyltransferases, TET proteins and the DNA glycosylase TDG that oxidize and excise 5-methylcytosine from the DNA. We performed ChIP-seq experiments to investigate the engagement and co-operation of these factors in mouse embryonic stem cells undergoing differentiation. We found TDG to localize with high preference to gene regulatory elements, including transcriptional start sites and enhancers, as well as to CpG islands (CGIs). TET1, TET2 and TDG co-operate in a locus-specific manner. They preferentially co-localize at genomic regions showing increased levels of DNA demethylation intermediates, a bivalent chromatin environment and an enrichment of the dynamic histone variants H3.3 and H2A.Z. These sites predominantly represent gene regulatory elements associated with developmental genes, including CGIs. A large fraction of TDG peaks is not co-occupied by TET proteins. These regions are depleted of CGIs as well as low-methylated regions (LMRs), contain less activating and bivalent chromatin marks, and show lower co-occupancy with DNA demethylation intermediates and with the dynamic histone variants H3.3 and H2A.Z, indicating reduced nucleosomal dynamics. Based on our results, we propose a mechanistic model where TDG associates with regulatory elements of developmental genes to promote epigenetic plasticity required in programming of cell identity. It does so together with TET1 and TET2 by oxidation and excision of 5mC, thereby generating DNA single-strand breaks that trigger changes in nucleosomal dynamics.

Introduction

The spatio-temporally accurate establishment of DNA methylation patterns plays a crucial role in transcriptional regulation, embryonic development, cell lineage commitment and other vital biological processes (Kraushaar and Zhao, 2013). DNA methylation is catalyzed by DNA methyltransferases (DNMTs). In mammals this occurs predominantly in CpG dinucleotide sequences (Klose and Bird, 2006). Genome-wide analysis of the distribution of DNA methylation in differentiated cells

revealed that the bulk genome is highly methylated, whereas certain islands of CpG-rich regions remain unmethylated. These unmethylated regions often overlap with transcription start sites (TSS) and are known as CpG-islands (CGIs; (Illingworth et al., 2010)). There are two events of epigenetic reprogramming which involve global DNA demethylation; the first takes place in pre-implantation embryos to set up pluripotency and the second in migrating primordial germ cells to erase parental-specific imprints (Hajkova et al., 2002; Mayer et al., 2000). Mechanisms and pathways underlying DNA demethylation have been subject of intense research, and there seem to occur both passive and active mechanisms. One emerging concept for active DNA demethylation entails a multi-step process involving the enzymatic oxidation and excision repair of 5-methylcytosine (5mC). This is achieved by a concerted action of Tet dioxygenases (TET1/2/3) that iteratively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC; (Ito et al., 2011; Pastor et al., 2013; Tahiliani et al., 2009)), and the Thymine DNA Glycosylase (TDG), which excises 5fC and 5caC from the DNA and thereby initiates the replacement of the modified with an unmodified C by repair synthesis (He et al., 2011; Maiti et al., 2013). Consistent with a role of non-canonical, TDG-dependent DNA repair in DNA methylation control, we and others showed previously that homozygous *Tdg* knockout mice display developmental failure associated with aberrant epigenetic programming of developmental gene promoters (Cortazar et al., 2011; Cortellino et al., 2011; Jacobs and Schar, 2012; Saito et al., 2012). Intermediates of oxidative DNA demethylation may have epigenetic functions, particularly 5hmC, which was proposed to play a role in ES cell fate, regulation of gene expression and neurodevelopment (Ito et al., 2010; Kriaucionis and Heintz, 2009; Szulwach et al., 2011). How the transition from 5mC to 5hmC and finally 5caC is controlled, is currently not known; one possibility constitutes the differential engagement of TET and TDG proteins.

Genome-wide analyses showed that TET1 and TET2 associate with chromatin in a non-random but targeted manner. Gene bodies, transcription start sites (TSS) and enhancers are preferred sites of interaction for TET1; however, TET1 binding does not predict transcriptional activity of a TSS (active, poised, inactive), but it might be an early event of enhancer activation (Serandour et al., 2012; Williams et al., 2011). There is less evidence available for genome-wide TET2 association, however, it appears that TET2 also preferentially interacts with TSS and gene bodies (Chen et

al., 2013b). Loss of TET2 causes a decrease of 5hmC predominantly in gene bodies and at the boundaries of highly expressed exons (Huang et al., 2014). There appear to be differential patterns of TET protein association and their oxidation products, relating to chromatin state and transcriptional activity. For instance, 5hmC is generally more enriched in exons and active enhancers compared to introns and poised enhancers, respectively (Neri et al., 2013). In mouse embryonic stem cells (ESCs) a fraction of TET1 is centered around the TSS of target genes enriched only for H3K4 trimethylation (H3K4me3) marks while another fraction is centered downstream of the TSS together with the bivalent histone marks H3K4me3 and H3K27me3. Also, TET proteins have been reported to associate with histone modifying enzymes; TET1 was shown to form a complex with the suppressor of zeste 12 (SUZ12), a component of the Polycomb repressive complex 2 (PRC2; (Neri et al., 2013; Wu et al., 2011)). Furthermore, TET1 interacts with SIN3a, a transcriptional co-repressor (Williams et al., 2011). Consistently, at H3K27me3-negative gene promoters, TET1 enrichment coincides with SIN3a enrichment whereas at H3K27me3-marked genes, TET1 appears bimodally distributed showing overlaps with SIN3a at the TSS and with SUZ12 downstream of the TSS (Neri et al., 2013). TET2 was shown to interact with the O-linked β -N-acetylglucosamine transferase (OGT), which glycosylates the canonical histones H2B, H3 and H4, thereby regulating transcription (Chen et al., 2013b; Deplus et al., 2013; Vella et al., 2013; Zentner and Henikoff, 2013). Importantly, the target genes for TET proteins in ESCs are prevalently associated with cell morphogenesis, (neural) development and differentiation (Raiber et al., 2012; Stadler et al., 2011; Williams et al., 2011). At the same time, gene ontology analysis revealed exactly these processes as predominantly impaired upon the knock-out of TDG (Cortazar et al., 2011). Similar to the TET proteins, TDG was shown to physically and functionally interact with chromatin associating and modulating proteins, including transcription factors such as retinoid- and estrogen-receptors (Chen et al., 2003; Kim and Um, 2008; Missero et al., 2001; Um et al., 1998), the histone acetyl transferase CBP/p300, and the *de novo* DNA methyltransferases DNMT3A and DNMT3B (Boland and Christman, 2008; Li et al., 2007; Tini et al., 2002). However, genome-wide analysis of chromatin association has not been reported for TDG.

The accumulation of the DNA demethylation intermediates 5fC and 5caC at proximal and distal gene regulatory elements indicates that the TET proteins may engage at

these loci to induce DNA demethylation and hence facilitate gene expression (Raiber et al., 2012; Shen et al., 2013; Song et al., 2013). However, since not all TET-bound sites become DNA demethylated, it can be envisioned that the TET proteins and TDG operate at these regulatory elements as well as in gene bodies in different ways, either to fully demethylate or to generate 5hmC, which might serve as an epigenetic mark itself. The quick and dynamic adaptation of gene expression programs to developmental clues is of particular importance for pluripotent cells. ESCs therefore maintain gene regulatory regions, such as promoters and enhancers, in a state of epigenetic plasticity. At promoters, bivalent chromatin domains (marked by H3K4me3 and H3K27me3) can be instructed to become constitutively active or silent in the context of lineage restriction (Aloia et al., 2013); similarly, enhancers become additionally marked by H3K27ac besides H3K4me1 upon activation (Creyghton et al., 2010; Zhu et al., 2013). Interestingly, gene promoters in a bivalent chromatin state usually have RNA polymerase II bound in a stalled configuration, thus, indicating that they are poised for transcription (Brookes et al., 2012). The histone variants H3.3 and H2A.Z were also reported to be highly enriched at gene regulatory regions like promoters and enhancers, where they display high turnover (Barski et al., 2007; Chen et al., 2013a; Kraushaar et al., 2013). Interestingly, H3.3 and H2A.Z preferentially associate in conjunction with each other in nucleosomes (Yukawa et al., 2014). Notably, H3.3 was suggested to serve as a chromatin bookmark to promote transcription recovery after genotoxic stress (Adam et al., 2013), consistent with a reported interaction of the H3.3-specific chaperone HIRA with the RNA Polymerase II (Ray-Gallet et al., 2011). The transient occupancy of H3.3 and H2A.Z at gene promoters and enhancers implicates a general role in gene regulation (Henikoff, 2009). This, together with the finding of highly dynamic DNA methylation and demethylation at actively transcribed promoters (Kangaspeska et al., 2008; Metivier et al., 2008) suggests that DNA oxidation and excision repair mediated DNA demethylation promotes chromatin dynamics required to maintain transcriptional plasticity in differentiating cells. The evolution of a targeted mechanism to regulate DNA methylation and demethylation at gene regulatory elements thus appears logical. Such a system would ideally target further chromatin modifying proteins and transcriptional co-factors to facilitate the silencing of certain genes and the initiation of transcription at others. How the targeting of e.g. TET proteins and

TDG is orchestrated is not known to date, further, it is unclear how the accompanying changes in chromatin are induced.

To address whether and how TET1, TET2 and TDG chromatin association is coordinated in differentiating mouse ESCs and whether differential occupancy can account for differential patterns of 5mC oxidation and nucleosomal composition, we set out to perform comprehensive TET1-, TET2-, and TDG-ChIP-seq analyses in mESC, generating the first chromatin association profiles for wildtype and catalytic-dead TDG in mESCs. We investigated genomic and chromatin features determining the enrichment of TET and TDG to refine the understanding of the respective functional interactions. We found that TDG preferentially associates with a chromatin environment, which is either bivalent or marked by the characteristic histone marks of enhancers. Furthermore, we found that the common peaks for TET1, TET2 and TDG are characterized by elevated levels of DNA demethylation intermediates as well as dynamic histone variants.

Results

To elucidate genome-wide chromatin association patterns for TDG, and correlate these with TET1 and TET2 occupancy, we performed chromatin immunoprecipitation followed by deep-sequencing (ChIP-seq) analyses for all proteins. As a cell model, we used TDG-deficient mouse embryonic stem cells (mESCs), complemented with TDG minigenes expressing either a wildtype TDG or a catalytically inactive variant (N151A) or the empty vector (*Tdg*[wt], [cat] and [null], respectively; (Cortazar et al., 2011)). Since TDG function becomes apparent in differentiating mESCs rather than in the pluripotent state (Jacobs et al., manuscript in preparation, 2014), and because TDG ChIP from pluripotent mESCs was inefficient and highly variable (data not shown), we performed our experiments with mESCs preconditioned in 2i medium and then subjected to a short differentiation for 24 hours induced by all-*trans* retinoic acid (RA). Importantly, TDG, TET1 and TET2 are present both in mESCs before and after RA exposure (Supplementary Figure 1). With this experimental setup, we generated ChIP-seq datasets for TDG from *Tdg*[wt] and *Tdg*[cat] mESCs, and for TET1 and TET2 from *Tdg*[wt], *Tdg*[cat] and *Tdg*[null] mESCs. For the TET proteins, we also performed ChIP-seq before RA exposure of the cells.

ChIP-seq analysis reveals a predominant association of TDG with transcription start sites as well as with intergenic regions

We then focused our analyses on the datasets obtained from mESCs at 24h RA. Peak calling for the TDG ChIP-seq data (MACS) first confirmed a good reproducibility of TDG enrichments in two independent experiments with *Tdg[cat]* mESCs ($R = 0.85$; Supplementary Figure 2). Further analyses identified a total of 14'144 sites in *Tdg[wt]* and 28'164 sites in *Tdg[cat]* mESC showing a higher than 2 fold enrichment ($\log_2 > 1$). Notably, only a small fraction of peaks in the two datasets overlapped; 21.6% of *Tdg[wt]* peaks are also present in *Tdg[cat]* and the correlation between the two was negative ($R = -0.3$; Figure 1a). This discrepancy may be explained by the different enzymatic properties of the wildtype and catalytic-dead TDG. While the wildtype TDG will excise a substrate base, i.e. 5fC or 5caC, and dissociate to give way for excision repair, the catalytic-dead enzyme will bind but fail to dissociate and turn over due to its inability to excise the base ((Hardeland et al., 2000; Waters and Swann, 1998) and Jacobs et al., manuscript in preparation, 2014). Hence, the catalytic-dead TDG will be caught at sites accumulating unprocessed substrate and, thus, more efficiently cross-link in ChIP. Given this, the higher number of peaks in *Tdg[cat]* compared to *Tdg[wt]* is expected and it seems likely that peaks showing up with the catalytic-dead TDG only are bound by the wildtype protein as well but, due to its higher turnover, are not detectable under the ChIP conditions applied. On the other hand, peaks appearing only with the wildtype TDG may reflect the limited dynamic capacity of the catalytic-dead TDG, which would affect the re-localization of TDG to newly activated sites of 5fC and 5caC generation in the course of ESC differentiation.

To proceed in an unbiased way, we kept *Tdg[wt]* and *Tdg[cat]* peaks separate in all subsequent analyses. Both, wildtype and catalytic-dead TDG showed a similar bimodal distribution with regard to their distance to the nearest transcriptional start site (TSS; RefSeq; Figure 1b), one distinct fraction of peaks showing in a narrow window encompassing the TSS and a second located roughly between 1 kb and 1000 kb away from the TSS. Consistent with this distribution pattern, most TDG peaks in both *Tdg[wt]* and *Tdg[cat]* were located in introns (41.4 % and 42.4%, respectively) and intergenic region (45.4% and 44.2%, respectively; Figure 1c). Less than 10% of TDG peaks were located in coding sequences (7.5% and 8.0%, respectively) and less than 5% were located in 5' (4.1% and 3.1%, respectively) and

3' untranslated regions (1.8% and 2.4%, respectively). Addressing the DNA sequence context, we found a preferential association of TDG with an intermediate CpG density, both in *Tdg*[wt] and *Tdg*[cat]. Nevertheless, we found a significant enrichment of CGIs within the TDG peaks (in *Tdg*[wt] and *Tdg*[cat]) compared to the overall genome (Figure 1d), supporting the notion that TDG associates to key sites for gene regulation. Figure 1e depicts the *HoxD* cluster as an example (Genome Browser). Altogether, these chromatin association patterns of both wildtype and catalytic-dead TDG could reflect a role in the regulation of transcription, both at the TSS and at distal regulatory elements, or in co-transcriptional processes.

TDG preferentially associates with DNase I hypersensitive sites and chromatin marked by active histone marks, RNA Polymerase II and dynamic nucleosomes

We showed before that differentiating TDG-deficient mESCs accumulate aberrant DNA methylation and histone modifications at CG-rich promoters of developmental genes (Cortazar et al., 2011). To decipher the chromatin environment disposing for TDG association, we investigated the coincidence of TDG peaks with distinct chromatin features, including histone modifications and nucleosome composition, and functional genomic elements implicated by DNase I hypersensitive sites or CTCF, p300 and RNA Polymerase II occupancy (Figure 2 and Supplementary Figure 5). This revealed very high enrichments for specific features in both *Tdg*[wt] and *Tdg*[cat] peaks when compared to the overall genome. Regarding genomic elements, the enrichment was most significant for DNase I hypersensitive sites (7.5 and 6.9 fold in *Tdg*[wt] and *Tdg*[cat], respectively) and RNA Pol II occupancy (9.8 and 5.6 fold, respectively), whereas p300 and CTCF binding sites were less prominently enriched. This indicated that TDG primarily occupies genomic regions characterized by low nucleosome occupancy and associated with transcription. Regarding repressive chromatin features, the H3K27me3 polycomb mark showed the highest enrichment in *Tdg*[wt] and *Tdg*[cat] peaks (14.3 and 9.0 fold, respectively), whereas the H3K9me3 heterochromatin mark showed little co-occupancy (1.2 and 1.7 fold, respectively; Figure 2b, Supplementary Figure 5b). This clearly associated TDG with the polycomb repressive system, consistent with its role in maintaining developmental genes in a poised state. Histone marks associated with active transcription, i.e. H3K4me3, H3K36me3 and H3K9ac, were all enriched in *Tdg*[wt] and *Tdg*[cat] peaks, although

this was more prominent for H3K4me3 and H3K9ac (7.3 to 10.2 fold enrichments) than for the transcriptional elongation mark H3K36me3 (1.8 to 2.3 fold enrichments; Figure 2c, Supplementary Figure 5c). Notably, strong enrichments were also apparent for the characteristic enhancer marks H3K4me1 (3.4 to 5.6 fold enrichments) and H3K27ac (3.4 to 4.4 fold enrichments). When we combined the latter with DNase I hypersensitivity to interrogate the chromatin environment of active enhancers (Thurman et al., 2012), we again found a significantly enriched occurrence of these in *Tdg[wt]* and *Tdg[cat]* peaks. The Low-methylated regions (LMRs), which represent a further distinct class of distal gene regulatory elements, were only moderately enriched within the TDG peaks, both in *Tdg[wt]* and *Tdg[cat]* (1.2 and 2.7 fold enrichment, respectively; Figure 2d, Supplementary Figure 5d). LMRs are characterized by a lower CpG content and reduced length compared to CGIs and they are occupied by DNA-binding factors in a cell-type-specific manner (Stadler et al., 2011). Furthermore, we found a significant enrichment of the dynamic histone variants H3.3 and H2A.Z in the *Tdg[wt]* and *Tdg[cat]* peaks. Enrichment was particularly high for H2A.Z (10.6 to 12.7 fold enrichment, for *Tdg[wt]* and *Tdg[cat]*, respectively), further strengthening the link to gene regulatory elements (Figure 2e, Supplementary Figure 5e). We conclude from these results that TDG preferentially localizes to DNase I hypersensitive sites which are further characterized by presence of RNA Polymerase II and dynamic histone variants. These sites are marked by histone modifications which are indicative of gene regulatory elements, as for instance found at active and bivalent gene promoters as well as enhancers. Contrarily, we do not observe an enrichment of TDG at genomic sites which are characterized by heterochromatic histone marks.

A substantial fraction of TDG peaks is co-occupied by the TET proteins and these sites are significantly enriched for CGIs

We next explored the coincidence of chromatin associations of TDG, TET1 and TET2. Thus, after peak calling from the TET1 and TET2 ChIP-seq datasets (MACS; Supplementary Figures 3 and 4) we asked how many TDG peaks (*Tdg[wt]* and *Tdg[cat]*) are co-occupied by TET1 or TET2 in *Tdg[wt]* ESCs (showing a higher than 2 fold enrichment; $\log_2 > 1$). Overall, we found TET1 and TET2 peaks to be significantly enriched in regions of *Tdg[wt]* and *Tdg[cat]* chromatin association in

mESC (Figure 3a), the co-occupancy being more pronounced for TET1 than for TET2 (37.7% versus 10.2% in *Tdg[wt]*, respectively). Notably, the co-occupancy of TDG with TET1 and TET2 is higher in *Tdg[wt]* compared to *Tdg[cat]* mESC, indicating that the catalytic activity of TDG facilitates TET binding. Next, we examined the mutual co-occupancies of all three proteins, intersecting the TDG peaks obtained for *Tdg[wt]* and *Tdg[cat]* with TET1 and TET2 peaks obtained from the *Tdg[wt]* background (Figure 3b). Of all TDG peaks in *Tdg[wt]*, 67% were uniquely bound by TDG, 24% were additionally bound by TET1, 4% by TET2, and 5% were triple-positive. Similarly, of all TDG peaks in *Tdg[cat]*, 66% were uniquely bound by TDG, 30% were additionally bound by TET1, 1% by TET2 and 3% were triple-positive. It is fair to assume that the fractions of unique, double- or triple-positive peaks reflect different functional engagements of TDG alone or together with TET1 and/or TET2. Notably, only a minor fraction of TDG-TET1 peaks is additionally associated with TET2, but a major fraction of TDG-TET2 peaks is additionally bound by TET1. Hence, if TET2 associates with TDG it does so mainly together with TET1, whereas TET1 and TDG appear to form an independent functional unit. There are also considerable differences in the genomic locations of these two distinct pools: While more than 32% of triple-positive peaks both in *Tdg[wt]* and *Tdg[cat]* mESC coincide with CGIs, less than 3.1 % of unique TDG peaks do so. By contrast, triple-positive and unique TDG peaks associate equally with active enhancers, both in *Tdg[wt]* (1.1% and 1.5%, respectively) and in *Tdg[cat]* mESCs (1.8% and 7.1%, respectively; Figure 3c). Furthermore, we found a significant fraction of triple-positive peaks in *Tdg[wt]* and *Tdg[cat]* (14.1% and 18.2%, respectively) to overlap with LMRs, as identified by Bis-seq ((Stadler et al., 2011); Figure 3d).

We then addressed potential differences in the molecular and cellular functions associated with the TET1-TET2-TDG positive sites and the TDG only peaks. Gene ontology analysis of the genes (Ingenuity Pathway Analysis) nearest to triple-positive peaks in both *Tdg[wt]* (633) and *Tdg[cat]* (825) revealed gene expression and cellular development as the most significant functions. Nearest genes to the peaks uniquely bound by TDG in *Tdg[wt]* (9431) and *Tdg[cat]* (18'468) yielded predominant functions in cellular assembly and organization as well as cellular function and maintenance (Figure 3e). These associations are consistent with previously reported findings that the loss of TDG in differentiating mESCs affects expression of genes involved mainly in embryonic, tissue, organ and organismal development, many of which were

encoding for transcription factors (Cortazar et al., 2011). From these data we conclude that TET1, TET2 and TDG show different patterns of genome-wide co-occupancy, depending on the DNA sequence context and functional sequence elements. While triple-positive peaks locate predominantly in CpG-rich regions (e.g. CGIs), TDG only enrichments are present at sites of intermediate CpG density (Figures 1d and 3c).

TET1 and TET2 associate with chromatin independent of TDG – but active TDG may regulate TET2 association with chromatin upon differentiation

Next, we asked if the chromatin association of TET1 and TET2 in mESC following 24h of RA-induced differentiation is affected by the presence or absence of TDG protein and/or its catalytic activity. We identified a total of 22'318 TET1 peaks in *Tdg[wt]* mESCs, which is in the range of previously reported results (Williams et al., 2011). Of these, 14'436 were also apparent in *Tdg[cat]* (64.7% of TET1 peaks in *Tdg[wt]*) and 16'218 in *Tdg[null]* mESC (72.6%, Figure 4a). This high overlap is consistent with TET1 associating with chromatin independently of TDG or its catalytic activity, and in agreement with TDG acting downstream of TET in the DNA demethylation process. Yet, there is a substantial fraction of TET1 peaks uniquely appearing in *Tdg[wt]*, *Tdg[cat]* or *Tdg[null]* mESCs (5178, 2553 and 4384, respectively), possibly representing distinct pools of chromatin-associated TET1 engaged in processes with different dissociation and association kinetics of both TET1 and TDG. Examining TET2, we identified 5129 enrichment peaks in *Tdg[wt]*, 9458 in *Tdg[cat]* and 14'623 in *Tdg[null]* (Figure 4b). The overlap between TET2 chromatin associations in the different TDG backgrounds was lower than that observed for TET1; 41.7% and 47.7% of TET2 peaks in *Tdg[cat]* and *Tdg[null]*, respectively, coincided with the total TET2 peaks in the *Tdg[wt]* cells (Figure 4b). Also, it appears that the absence or catalytic inactivity of TDG stimulates TET2 chromatin association in mESCs at 24h RA, which is unlike the situation with TET1. Notably, this dependency seems to arise only upon differentiation since pluripotent mESCs (0h RA) show slightly more TET2 in the presence of active TDG (Supplementary Figure 6b), and this goes in line with another difference between TET1 and TET2: while the total number of TET1 chromatin associations increased during the 24h of RA-induced differentiation irrespective of the TDG status, it was

reduced for TET2, most dramatically in the *Tdg*[wt] background (Figure 4, Supplementary Figure 6). From these data, we conclude that TET1 and TET2 association with chromatin is largely independent of TDG, but that the turnover of TET2 in particular may be regulated by TDG. In the context of differentiation-induced DNA demethylation, active TDG appears to reduce TET2 association with chromatin.

A significant fraction of sites co-occupied by TET1, TET2 and TDG is marked by DNA demethylation intermediates and dynamic histone variants

To clarify, whether and how differential TET1, TET2 and TDG co-occupancy is related to DNA demethylation activity, we correlated our ChIP-seq profiles with publicly available datasets on genome-wide DNA demethylation intermediates, including 5hmC and 5fC (hMe-Seal-seq and fC-Seal-seq by (Song et al., 2013)). For 5fC, the analysis was performed in a TDG-deficient background to be able to measure enrichments. Comparing unique TDG peaks with a higher than 2 fold enrichment ($\log_2 > 1$) with the triple-positive peaks in *Tdg*[wt] mESCs, we observed a significantly higher overlap of 5fC accumulating sites with triple-positive peaks than with TDG only peaks, while the overlap with 5hmC was more pronounced in TDG only peaks (Figure 5a). Hence, higher oxidation of 5mC occurs preferentially at sites where TET1, TET2 and TDG are tightly associated. The overlap of 5hmC and/or 5fC was similar for triple-positive peaks in *Tdg*[wt] and *Tdg*[cat]. However, it was significantly different for TDG only peaks; TDG only peaks in *Tdg*[cat] mESC showed a striking enrichment of sites double-positive for 5hmC and 5fC at the expense of unmodified C, 5mC or 5caC (“None”; Figure 5a). We conclude from these results first, that there is a functional separation between uniquely TDG-bound and triple-positive sites and second, that catalytic-dead TDG blocks the progression of the DNA methylation cycle.

We next addressed the chromatin states associated with TDG only and TET1, TET2 and TDG occupied sites, focusing on the activating and silencing histone modifications H3K4me3 and H3K27me3 that are prominently enriched at sites of TDG chromatin association. This revealed a very strong enrichment of bivalent chromatin marks in triple-positive peaks compared to TDG only peaks (23.8-fold and 12.0-fold in *Tdg*[wt] and *Tdg*[cat], respectively), in fact, bivalent chromatin marks

were nearly absent in TDG only peaks. Concomitantly, we noticed a remarkable increase of H3K4me3-positive sites in the triple-positive peaks (8.8-fold and 2.8-fold, respectively, Figure 5b). Similarly to the distribution of DNA demethylation intermediates, triple-positive peaks in *Tdg[wt]* and *Tdg[cat]* were nearly identical with respect to the representation of H3K4me3 and H3K27me3 marks. Differences were greater in the unique TDG peaks, where *Tdg[cat]* shows a higher fraction of H3K4me3-positive peaks compared to *Tdg[wt]*. This may again be a reflection of the failure of the catalytic-dead TDG to turn over at sites where 5fC and 5caC substrate is generated, which would in this case be at sites activating chromatin marks. Importantly, sites in a bivalent chromatin state are most significantly enriched in TET1, TET2, and TDG triple-positive peaks. In differentiating mESCs, bivalent chromatin marks are found predominantly at developmental genes (Bernstein et al., 2006). Developmentally regulated loci are also characterized by deposition of the dynamic histone variant H3.3 (Goldberg et al., 2010), the turnover of which is particularly high at enhancer and promoter regions (Kraushaar et al., 2013). Moreover, active enhancers are marked by H3K4me1 and H3K27ac and occupied by 5hmC, 5fC and 5caC (Serandour et al., 2012; Shen et al., 2013; Song et al., 2013). Notably, all these characteristics are enriched in the TET1, TET2, and TDG triple-positive peaks. We therefore tested whether deposition of the histone variants H3.3 and H2A.Z is associated with TET and TDG occupancy (Figures 5c and 5d). This revealed a very high coincidence of H3.3 with both triple-positive and unique TDG peaks, both in *Tdg[wt]* and *Tdg[cat]* mESCs. In the *Tdg[wt]* background, the H3.3 co-occupancy was increased in the triple-positive peaks when compared to the TDG only peaks. This suggests that H3.3 deposition is higher at sites of triple-positive TDG peaks. Again, the H3.3 co-occupancy in the triple-positive peaks in both *Tdg[wt]* and *Tdg[cat]* were nearly identical (91.5% and 90.7%, respectively), whereas the most significant difference was in the comparison of the TDG only peaks in *Tdg[wt]* and *Tdg[cat]* (84.6% and 95.9% co-occupancy, respectively; Figure 5c). A similar pattern was found for H2A.Z co-occupancy: 62.9% and 58.5% of the TET1, TET2 and TDG triple-positive peaks were co-occupied by H2A.Z in *Tdg[wt]* and *Tdg[cat]*, respectively (Figure 5d). H2A.Z co-occupancy was reduced in TDG only peaks in both *Tdg[wt]* and *Tdg[cat]* (24.1% and 35.9%, respectively), however, the reduction was more pronounced in a *Tdg[wt]* background, as is the case for H3.3 co-occupancy. This seems comprehensible by the argument that the catalytic-dead TDG

accumulates at sites where 5fC and 5caC is generated and thereby inhibits active DNA demethylation, and thus repair-mediated DNA strand-break formation, which in turn may affect H3.3 and H2A.Z deposition and turnover. We thus propose a mechanistic model, where in differentiating cells, TDG together with the TET proteins associate at CpG-rich gene regulatory regions to induce DNA demethylation and thereby facilitate nucleosome turnover and accompanying changes in chromatin modifications (Figure 6).

Discussion

Dynamic regulation of DNA methylation and gene expression is a hallmark of cell differentiation and development. The recent discovery of DNMT-TET-TDG-mediated cyclic methylation, oxidation and excision of cytosine constitutes a novel possibility for fine-tuning of gene expression. To address the co-operation of TET and TDG proteins at a genomic scale, we set out to establish genome-wide association patterns of these proteins in differentiating mESCs. We found a preference for intermediate to high CpG density and a bimodal distribution of TDG chromatin association sites; one peak was encompassing the TSS and one peak was located 1 to 1000 kb away from the TSS. TDG preferentially associated to a chromatin environment marked by DNase I hypersensitive sites, RNA Pol II occupancy, and histone marks characteristic for bivalent or active chromatin. Further, TDG peaks showed a significant co-occupancy with TET1 and TET2 chromatin association. Interestingly, the TET1-TET2-TDG triple-positive peaks were clearly enriched for CGIs, bivalent chromatin, DNA demethylation intermediates and dynamic nucleosomes, such as H3.3 and H2A.Z. Contrarily, unique TDG peaks were associated with active enhancers, not marked by bivalent chromatin and showed a lower fraction of DNA demethylation intermediates and dynamic nucleosomes, suggesting a functional separation between distinct TDG protein fractions.

ChIP-seq data for wildtype and catalytic-dead TDG variants revealed different genomic distributions. These can largely be explained by different biochemical properties of the proteins, i.e. a reduced turnover of the catalytic inactive TDG and an increased accumulation of unprocessed substrates in *Tdg[cat]* mESCs. Consistently, the genomic characteristics of the interaction sites identified with active and inactive

TDG were largely identical, although the actual overlap between enrichment peaks was rather small. It is thus likely that active TDG does associate with many sites observed in *Tdg[cat]*, but because of its turnover, it more frequently escapes crosslinking and, hence, detection by ChIP. We found most TDG peaks located in regions of intermediate to high CpG density; functional characterization revealed a preference for gene bodies (mostly introns) and intergenic regions (such as enhancers), indicating that TDG is associated with the transcription machinery upon induction of differentiation. The high percentage of intergenic peaks, together with the prominent occurrence of TDG peaks in the vicinity of the TSS and between 1kb and 1000kb away from it, indicates a preference for association with gene regulatory elements. This is supported by other features coinciding at TDG peaks such as the enrichment of the histone modifications H3K4me3, H3K9ac, H3K4me1 and H3K27ac, the former two characterizing active promoters, the latter two active enhancers. Moreover, TDG peaks also correlate with DNase I hypersensitive sites, as well as p300 and RNA Polymerase II enrichment, pointing towards an interaction with nucleosome depleted, transcriptionally active chromatin, another feature of gene promoters. Interestingly, we also found a clear enrichment of bivalent chromatin marks (H3K4me3 and H3K27me3), indicating that many developmental genes are targeted by TDG, which is in line with a previous report from our lab (Bernstein et al., 2006; Cortazar et al., 2011). By contrast, TDG does not interact with constitutively silenced genomic regions, as H3K9me3 was not comparably enriched in TDG peaks, showing that TDG is not randomly distributed in chromatin.

TDG peaks were also significantly enriched in TET1 and TET2 binding sites, implicating a functional interaction between these proteins. Notably, TET1 and TET2 association with chromatin is largely independent of TDG. However, the turnover of TET2 in particular appears to be regulated by TDG: upon differentiation-induced DNA demethylation, active TDG seems to reduce TET2 association with chromatin, which was not observed for TET1. This is in line with the finding that most TET2 peaks are co-occupied with TDG and TET1, whereas TET1 peaks occur largely independently of TET2. Thus, TET2-TDG appear to be more intimately linked compared to TET1-TDG. Nevertheless, there is a substantial number of both TET1 and TET2 peaks which is observed exclusively in *Tdg[wt]*, *Tdg[cat]* or *Tdg[null]*, respectively. Presumably, TET1 peaks in a *Tdg[wt]* background represent sites which become activated upon differentiation by DNA demethylation, in contrast to *Tdg[null]*-specific

TET1 peaks, which may rather reflect targets of TET1-mediated gene regulation. We conclude from these findings that the targeting of the distinct protein pools (e.g. TDG only, TET1-TDG, TET2-TDG, TET1-TET2-TDG) depends at least partly on its composition and may thus reflect functional differences of the proteins.

Triple-positive peaks, showing enrichments for TET1, TET2 and TDG, accounted for roughly 5% of the total TDG peaks. It was reported before that TET1 associates with chromatin in two distinct patterns, the first centered at the TSS together with the transcriptional co-repressor Sin3a, the second located downstream and co-localizing with 5-hmC and PRC2 (Neri et al., 2013). Similarly, our data show a differential association of TET proteins with TDG throughout the genome, implicating the existence of locus-specific physical and functional interactions. Triple-positive peaks are significantly more often located in CGIs compared to the unique TDG peaks. Thus, it is possible that the triple-positive peaks reflect ongoing active DNA demethylation at sites of transcriptional initiation. Contrarily, unique TDG peaks are rather associated with enhancers, possibly reflecting sites of lower turnover. We conclude that the triple-positive peaks clearly associate with promoter CGIs, where TET activity is high and unique TDG peaks associate with enhancers, where TET activity might be lower. The positive correlation between TET1, TET2, TDG and the DNA demethylation intermediates is a strong argument for the constant, highly orchestrated turnover of DNA methylation at CGIs. The enhanced binding of catalytically inactive TDG at sites of 5fC and 5caC accumulation might reflect an inhibition of *de novo* methylation and TET association.

So far, we conclude that during differentiation, the cycling DNA methylation and oxidative demethylation is required to maintain epigenetic plasticity at gene regulatory elements (e.g. CGI promoters) that ultimately need to be programmed for cell-type specific expression. This may contradict the conventional view that CGIs are unmethylated regions in the genome, but is supported by an increasing body of evidence, at least for CGIs associated with the regulation of developmental genes. Several studies demonstrate a modulation of the DNA methylation status of intra- and intergenic enhancer regions during differentiation and specifically show massive DNA demethylation at enhancers during cell differentiation (Aran and Hellman, 2013; Schmidl et al., 2009; Xie et al., 2013). The high dynamics of DNA methylation at the triple-positive loci coincided with a bivalent chromatin environment. Bivalent domains

mark gene promoters characterized by a high degree of regulatory plasticity; they can rapidly become activated or silenced in response to appropriate transcriptional instructions (Aloia et al., 2013). Unlike the unique TDG peaks, the triple-positive peaks show a clear enrichment for both bivalent and H3K4me3-positive chromatin. Both H3K4me3-marked sites and bivalent chromatin are positively correlated with CGIs as well as TSS (Bernstein et al., 2006). The lack of enrichment of both bivalent chromatin as well as CGIs in the unique TDG peaks compared to the triple-positive peaks thus further supports a functional separation of the distinct TDG fractions.

This goes in line with the high co-occupancy of both the triple-positive but also unique TDG peaks with the dynamic histone variants H3.3 and H2A.Z. Roughly 90% of the triple-positive peaks were positive for H3.3, irrespective of the TDG background. Contrarily, the unique TDG peaks showed a difference between TDG backgrounds: in *Tdg[wt]*, H3.3 occurred less compared to *Tdg[cat]*, hinting at a role for TDG's catalytic activity in nucleosome turnover. Notably, H3.3 occupancy was shown to change at cell-type specific genes and regulatory elements during cell differentiation, as evidenced by enhanced nucleosome exchange (Goldberg et al., 2010). Thus, deposition of the rapidly turning-over H3.3 has been implicated in a mechanism that poises chromatin for transcriptional activation (Huang and Zhu, 2014). H3.3 was further demonstrated to facilitate the association of PRC2 with developmental genes, thus establishing the correct setting for H3K27 methylation in bivalent chromatin (Banaszynski et al., 2013). Interestingly, a solid body of circumstantial evidence indicates that DNA demethylation at gene regulatory regions in the context of transcriptional activation includes the formation of DNA single-strand breaks (Fry et al., 2006; Kress et al., 2006; Wossidlo et al., 2010). It is thus appealing to think of a mechanism for gene activation during development, where the TET-TDG-mediated DNA demethylation is followed by the generation of a single-strand break, presumably by APE1, which in turn facilitates nucleosome exchange and the opening up of chromatin. Nucleosome exchange could either be achieved by the sheer relaxation of DNA around the histone octamers induced by the single-strand break. Alternatively, single-strand breaks were shown to attract PARP1 which would PARylate the surrounding histones, enhancing their turnover (Meyer-Ficca et al., 2011a; O'Donnell et al., 2013). How common this is and how the strand break is repaired and prevented from generating a double-strand break is currently unknown.

We describe two functionally different fractions of chromatin associated TDG: the first is strongly co-occupied by the TET proteins and the second consists of TDG only. The TET associated fraction correlates with (presumably promoter) CGIs at genes involved in gene expression and cellular development, the TDG only fraction correlates rather with regions characterized by an intermediate CpG-density and is enriched for active enhancers. Based on our findings, we postulate the existence of a mechanism which uses highly coordinated DNA demethylation and specifically targets gene regulatory elements of developmental genes to establish epigenetic plasticity during cell lineage commitment. This process is initiated at lineage-specific proximal and distal gene regulatory elements, where the presence of TET1-TET2-TDG promotes transient cycling of DNA methylation and demethylation. This facilitates the deposition and turnover of the histone variants H3.3 and H2A.Z at enhancers and gene promoters, presumably through PARP1-mediated histone PARylation upon the generation of a BER-mediated DNA single-strand break (Kress et al., 2006; Meyer-Ficca et al., 2011b; O'Donnell et al., 2013). TET1-TET2-TDG peaks are further significantly enriched for active histone marks (H3K4me1, H3K4me3, H3K9ac, H3K27ac), all of which were shown to positively correlate with DNA looping events between TSS and distal regulatory elements (Sanyal et al., 2012). We envision TET1-TET2-TDG to operate in the context of promoter-enhancer-interaction loops to facilitate nucleosomal dynamics at both sites. The deposition of the dynamic histone variants H3.3 and H2A.Z is central as this was shown to regulate the transcriptional outcome of inducible genes (Chen et al., 2013a).

For the first time, we could show the genome-wide distribution of TDG chromatin association and demonstrate a differential co-occupancy with TET1 and TET2. It will be important to further investigate the relationship between cyclic DNA methylation and active, TET-TDG mediated demethylation as well as nucleosomal composition and turnover, and how this mechanism is targeted to specific sites in the genome.

Material & Methods

Cell culture

Complemented ES cell lines were derived by transfection of TDG^{-/-} ES cells with the complementation vectors pTCO2 TDG wt, pTCO2 TDG Δ cat and empty pTCO2 (Cortazar et al., 2011) using jetPEI® (Polyplus Transfections) according to the manufacturer's recommendations. Cells were cultivated in ES cell medium (ECM; DMEM, 15% heat-inactivated FCS, 1x non-essential amino acids, 1 mM Na-pyruvate, 2 mM L-glutamine and 90 μ M β -mercaptoethanol) and LIF (1'000U/ml-1) supplemented with 1.5 μ g/ml-1 puromycin to select stable clones.

For the 24 h RA differentiation, complemented ES cells were cultured on Feeders for 2 passages and then conditioned for 4 passages without Feeders in 2i medium (Neurobasal medium and DMEM/F-12 1:1, 1x N2 supplement, 1x B27 supplement, LIF (1'000 U/ml-1), 2 mM L-glutamine, 90 μ M β -mercaptoethanol, 3 μ M CHIR99021 and 1 μ M PD0325901 (University of Dundee) and 1x penicillin/streptomycin). Prior to RA differentiation, ESCs were seeded at suitable cell numbers onto two 140 mm dishes (for Chromatin) or two 60 mm dishes (for Protein and RNA extraction) in ECM. For differentiation, the medium was exchanged for ECM supplemented with 5 μ M RA (5 mM stock in DMSO). Chromatin, protein and RNA were harvested at 0 and 24 h. If not indicated otherwise, cell culture components were obtained from Gibco® Life Technologies, chemicals from Sigma and LIF from Merck Millipore.

Chromatin immuno-precipitation (ChIP)

To crosslink protein-bound DNA, ES cells were incubated in 1% formaldehyde/PBS at room temperature. The reaction was quenched after 10 min by addition of glycine to a final concentration of 125 mM. After washing three times with ice cold PBS, cells were collected using a cell scraper and subsequent centrifugation at 600xg and 4 °C. Supernatant was discarded and the cells snap-frozen until further processing. After thawing on ice, nuclei were isolated by incubation in 400 μ l of cold ChIP Lysis Buffer I (10 mM HEPES pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, 1 mM PMSF) for 5 min on ice followed by two incubations of 5 min on ice in 400 μ l cold ChIP Lysis buffer II (10 mM HEPES pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 200 mM NaCl, 1mM PMSF). All centrifugation steps were conducted at 600xg and 4°C for 5 min. Pelleted nuclei were lysed in 400 μ l ChIP Lysis buffer III (50 mM Tris-HCl pH 8.0,

1 mM EDTA, 0.5% Triton X-100, 1% SDS, 1 mM PMSF) for 10 min on ice followed by sonication for 5x10 min (30 sec on, 30 sec off, power high) using a Bioruptor sonicator (Diagenode) to yield fragments of ~200-500 bp. The solution was cleared of remaining cell debris by centrifugation at 14'000xg and 4°C for 10 min. For ChIP of TDG, TET1 and TET2, 150 µg of chromatin were diluted 1:10 in ChIP dilution buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 1x protease inhibitor cocktail, 1 mM PMSF). After removing 1% (volume) for input analysis, diluted chromatin was pre-cleared at 4°C for 1 h with 30 µl of a 50% slurry of magnetic Protein G beads (Invitrogen) pre-blocked with 1 mg ml⁻¹ BSA and 1 mg ml⁻¹ tRNA. Pre-cleared chromatin was incubated with 1-2 µg of the respective antibody (Supplementary Table 1) overnight at 4°C under slow rotation. Immuno-complexes were precipitated with 40 µl of a 50% slurry of blocked Protein G beads and further incubated at 4°C for 2 h. Beads were then serially washed with 500 µl ChIP wash buffer I (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100) and twice with 500 µl ChIP wash buffer II (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100). After two additional washes with 500 µl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), bound complexes were eluted by two sequential incubations with 250 µl elution buffer (1% SDS, 0.1 M NaHCO₃) at 65°C shaking for 10 min. Crosslink reversal of eluates and respective input samples was done in the presence of 200 mM NaCl at 65°C for 4 h followed by proteinase K digestion (50 µg ml⁻¹) in the presence of 10 mM EDTA and 40 mM Tris-HCl pH 6.5 at 45°C for 1 h. DNA was purified by phenol/chloroform extraction and NaCl/ethanol precipitation, and resuspended in 10 mM Tris-HCl pH 8.0. qPCR analysis with target specific primers (Supplementary Table 2) was performed using Quantitect SYBR Green (Qiagen) with a Rotor-Gene 3000 thermocycler (Qiagen). Statistical analysis was performed on Graphpad Prism Software.

Deep-sequencing and bioinformatical analysis

Several ChIP experiments were pooled and 10 ng of ChIP elution was sent to the Genome Technology Access Center (GTAC; <https://gtac.wustl.edu/index.php>) in the Department of Genetics at Washington University School of Medicine in St. Louis, Missouri, USA. After quality control (Bioanalyzer, Agilent Technologies Inc.), ChIP

DNA was blunt ended, Adenosine was added to the 3' end and sequencing adapters were ligated to the ends. The fragments were size selected to 200-600 bp and underwent amplification for 15 cycles. The resulting libraries were sequenced using the Illumina HiSeq-2500 as single reads extending 50 bases (Illumina Inc.). The raw data was demultiplexed and the reads were mapped to the mouse genome (NCBI37/mm9) using the Bowtie (version 1.0.0; (Langmead et al., 2009)) implemented in QuasR package (www.bioconductor.org) allowing up to 10 best alignment positions in the genome. The sample specific fragment sizes were estimated from cross correlation profiles of read density on both chromosomal strands using the Chipcor software (ref: <http://ccg.vital-it.ch/chipseq>). Reads were shifted by half of the fragment size (90bp) towards the middle of the fragment. Peaks were detected using MACS (version 1.3.7.1., with parameters `--tsize=50 --pvalue=1e-5 --mfold=8 --lambdaset='1000,5000,10000' --nomodel --shiftsize=90`; (Zhang et al., 2008)) using a pool of read alignments from all biological replicates and time points in combination with pooled input controls. Peaks were filtered based on the enrichment over the input control samples. ChIP enrichment of each peak was calculated as

$e = \log_2\left(\frac{n_{fg}/N_{fg} \cdot \min(N_{fg}, N_{bg}) + p}{n_{bg}/N_{bg} \cdot \min(N_{fg}, N_{bg}) + p}\right)$, where n_{fg} and n_{bg} are the number of overlapping foreground and background (input chromatin) read alignments, respectively. N_{fg} and N_{bg} are the total number of aligned reads in foreground and background samples, and p is a pseudocount constant ($p=8$) used to minimize the sampling noise for peaks with very low counts. Only peaks with \log_2 fold enrichment higher than 1 were used for further analysis.

For comparative analysis, the following GEO Datasets were downloaded: GSE37074 (DNase), GSE49847 (H3K4me3, H3K4me1, CTCF, RNA Pol II, p300, H3K27me3, H3K27ac, H3K36me3, H3K9ac, H3K9me3), GSE41545 (5-hmC, 5-fC), GSE16893 (H3.3), GSE40065 (H2A.Z) and GSE30206 (LMRs).

Supplementary Table 1: Antibodies used in this study

Antibody	Product Number	Manufacturer
Anti-Methylcytosine dioxygenase TET1 antibody	09-872	Millipore
Anti-TET2 (N2) antibody	R1086-4	Abiocode

The anti-TDG antibody used for ChIP was produced and affinity purified in our lab, for further information see (Cortazar et al., 2011; Hardeland et al., 2002; Neddermann et al., 1996).

Supplementary Table 2: Primers used in this study

Primer	Forward (5' – 3' sequence)	Reverse (5' – 3' sequence)
chr2neg	AGCACAGCCTGAAGCCTCTA	AGAGGGCATTTCGGTCTTTT
HoxA10	CACTCCCAGTTTGGTTTCGT	GGGGGTACAGGTTCAAGAGC
Oct4	GTGAGGTGTCGGTGACCCAAGGCAG	GGCGAGCGCTATCTGCCTGTGTC
Nanog	GAGGATGCCCCCTAAGCTTTCCTCCC	CCTCCTACCCTACCCACCCCCTATTCTCCC
DMR 8	CTGGCCACAGCTTTACCATC	AAGGACACTGAGCCAGAGGA
DMR 39	GAGCTGGATAGCCCTTGTAGAATG	TTGGCAGCGGAGGGAGCAG
DMR 49	GCTGGGTTTGTAGTGGGAAC	GCAGGACCACACCTCACATC
DMR 54	ACCCAGCAAATCTCACCTG	GACACTGGACAGGGCTCCA

All primers were from Microsynth, Switzerland.

Acknowledgements

We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1 TR000448 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH.

References

- Adam, S., Polo, S.E., and Almouzni, G. (2013). Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. *Cell* *155*, 94-106.
- Aloia, L., Di Stefano, B., and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. *Development* *140*, 2525-2534.
- Aran, D., and Hellman, A. (2013). DNA methylation of transcriptional enhancers and cancer predisposition. *Cell* *154*, 11-13.
- Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsasser, S.J., Chappier, A., Goldberg, A.D., Canaani, E., *et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. *Cell* *155*, 107-120.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* *129*, 823-837.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* *125*, 315-326.
- Boland, M.J., and Christman, J.K. (2008). Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *J Mol Biol* *379*, 492-504.
- Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., *et al.* (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* *10*, 157-170.
- Chen, D., Lucey, M.J., Phoenix, F., Lopez-Garcia, J., Hart, S.M., Losson, R., Buluwela, L., Coombes, R.C., Chambon, P., Schar, P., *et al.* (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *J Biol Chem* *278*, 38586-38592.
- Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X., *et al.* (2013a). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes Dev* *27*, 2109-2124.
- Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013b). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* *493*, 561-564.
- Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., *et al.* (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* *470*, 419-423.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., *et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* *146*, 67-79.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* *107*, 21931-21936.
- Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Mendez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., *et al.* (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* *32*, 645-655.
- Fry, R.C., DeMott, M.S., Cosgrove, J.P., Begley, T.J., Samson, L.D., and Dedon, P.C. (2006). The DNA-damage signature in *Saccharomyces cerevisiae* is associated with single-strand breaks in DNA. *BMC Genomics* *7*, 313.
- Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., *et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* *140*, 678-691.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* *117*, 15-23.

- Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem* 275, 33449-33456.
- Hardeland, U., Steinacher, R., Jiricny, J., and Schar, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J* 21, 1456-1464.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303-1307.
- Henikoff, S. (2009). Labile H3.3+H2A.Z nucleosomes mark 'nucleosome-free regions'. *Nat Genet* 41, 865-866.
- Huang, C., and Zhu, B. (2014). H3.3 turnover: A mechanism to poise chromatin for transcription, or a response to open chromatin? *Bioessays*.
- Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W.A., Kang, J., Zepeda-Martinez, J.A., Pape, U.J., Jacobsen, S.E., Peters, B., *et al.* (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 111, 1361-1366.
- Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* 6, e1001134.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129-1133.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300-1303.
- Jacobs, A.L., and Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma* 121, 1-20.
- Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. *Nature* 452, 112-115.
- Kim, E.J., and Um, S.J. (2008). Thymine-DNA glycosylase interacts with and functions as a coactivator of p53 family proteins. *Biochem Biophys Res Commun* 377, 838-842.
- Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31, 89-97.
- Kraushaar, D.C., Jin, W., Maunakea, A., Abraham, B., Ha, M., and Zhao, K. (2013). Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol* 14, R121.
- Kraushaar, D.C., and Zhao, K. (2013). The epigenomics of embryonic stem cell differentiation. *Int J Biol Sci* 9, 1134-1144.
- Kress, C., Thomassin, H., and Grange, T. (2006). Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks. *Proc Natl Acad Sci U S A* 103, 11112-11117.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929-930.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Li, Y.Q., Zhou, P.Z., Zheng, X.D., Walsh, C.P., and Xu, G.L. (2007). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res* 35, 390-400.
- Maiti, A., Michelson, A.Z., Armwood, C.J., Lee, J.K., and Drohat, A.C. (2013). Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J Am Chem Soc* 135, 15813-15822.
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* 403, 501-502.

- Metivier, R., Gallais, R., Tiffocche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., *et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452, 45-50.
- Meyer-Ficca, M.L., Ihara, M., Lonchar, J.D., Meistrich, M.L., Austin, C.A., Min, W., Wang, Z.Q., and Meyer, R.G. (2011a). Poly(ADP-ribose) metabolism is essential for proper nucleoprotein exchange during mouse spermiogenesis. *Biol Reprod* 84, 218-228.
- Meyer-Ficca, M.L., Lonchar, J.D., Ihara, M., Meistrich, M.L., Austin, C.A., and Meyer, R.G. (2011b). Poly(ADP-ribose) polymerases PARP1 and PARP2 modulate topoisomerase II beta (TOP2B) function during chromatin condensation in mouse spermiogenesis. *Biol Reprod* 84, 900-909.
- Missero, C., Pirro, M.T., Simeone, S., Pischetola, M., and Di Lauro, R. (2001). The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription. *J Biol Chem* 276, 33569-33575.
- Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J.J., Wiebauer, K., and Jiricny, J. (1996). Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J Biol Chem* 271, 12767-12774.
- Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. (2013). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol* 14, R91.
- O'Donnell, A., Yang, S.H., and Sharrocks, A.D. (2013). PARP1 orchestrates variant histone exchange in signal-mediated transcriptional activation. *EMBO Rep* 14, 1084-1091.
- Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* 14, 341-356.
- Raiber, E.A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol* 13, R69.
- Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D.C., Pchelintsev, N.A., Adams, P.D., Jansen, L.E., *et al.* (2011). Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol Cell* 44, 928-941.
- Saito, Y., Ono, T., Takeda, N., Nohmi, T., Seki, M., Enomoto, T., Noda, T., and Uehara, Y. (2012). Embryonic lethality in mice lacking mismatch-specific thymine DNA glycosylase is partially prevented by DOPS, a precursor of noradrenaline. *Tohoku J Exp Med* 226, 75-83.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109-113.
- Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res* 19, 1165-1174.
- Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L., *et al.* (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* 40, 8255-8265.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153, 692-706.
- Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., *et al.* (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153, 678-691.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., *et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490-495.
- Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., *et al.* (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* 7, e1002154.

- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935.
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., *et al.* (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75-82.
- Tini, M., Benecke, A., Um, S.J., Torchia, J., Evans, R.M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol Cell* 9, 265-277.
- Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., and Chambon, P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J Biol Chem* 273, 20728-20736.
- Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., *et al.* (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* 49, 645-656.
- Waters, T.R., and Swann, P.F. (1998). Kinetics of the action of thymine DNA glycosylase. *J Biol Chem* 273, 20007-20014.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343-348.
- Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., Scholer, H., and Walter, J. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J* 29, 1877-1888.
- Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473, 389-393.
- Xie, M., Hong, C., Zhang, B., Lowdon, R.F., Xing, X., Li, D., Zhou, X., Lee, H.J., Maire, C.L., Ligon, K.L., *et al.* (2013). DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* 45, 836-841.
- Yukawa, M., Akiyama, T., Franke, V., Mise, N., Isagawa, T., Suzuki, Y., Suzuki, M.G., Vlahovicek, K., Abe, K., Aburatani, H., *et al.* (2014). Genome-wide analysis of the chromatin composition of histone H2A and H3 variants in mouse embryonic stem cells. *PLoS One* 9, e92689.
- Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* 20, 259-266.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.
- Zhu, Y., Sun, L., Chen, Z., Whitaker, J.W., Wang, T., and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res* 41, 10032-10043.

Figure Legends

Figure 1: TDG ChIP-seq reveals a preferential association of TDG with intergenic regions as well as introns

A: Density plot depicting the negative correlation between total TDG peaks (MACS) in *Tdg[wt]* and *Tdg[cat]* (left; density: red > blue). The VENN diagram (right) reveals little overlap between TDG peaks ($\log_2 > 1$) in *Tdg[wt]* and *Tdg[cat]*. **B:** Density diagram showing TDG peaks in relation to TSSs (RefSeq). **C:** Pie charts showing the percentage of TDG peaks (*Tdg[wt]* top and *Tdg[cat]* bottom) uniquely in the indicated genomic feature (cds: coding sequence; UTR: untranslated region). **D:** Density plot showing the correlation between TDG enrichment and CpG density (left). Bar charts indicating the percentage of TDG peaks located in a CGI (top, right) and the fold enrichment of CGIs within the TDG peaks compared to the occurrence of CGIs in the total genome (bottom, right). **E:** Genome Browser snapshot as an illustrating example.

Figure 2: The enrichment of active as well as bivalent chromatin marks is most pronounced in TDG-targeted sites

Bar charts are depicted, showing the fold enrichment of the indicated feature compared to the occurrence of the respective feature in the total genome (left): in *Tdg[wt]* (middle) and in *Tdg[cat]* (right). **A:** DNase hypersensitive sites, RNA Polymerase II, p300 and CTCF. **B:** Repressive histone marks: H3K27me3 and H3K9me3. **C:** Activating histone marks associated with transcription: H3K4me3, H3K36me3, H3K9ac. **D:** Features associated with active enhancers: H3K4me1, H3K27ac as well as the combination of these two marks together with DNase hypersensitivity in “active enhancers”, low-methylated regions (LMRs). **E:** Histone variants: H3.3, H2A.Z.

Figure 3: Co-localization of TDG with TET1 and TET2 indicates a coordinated function in gene regulation

A: Barplot on the left shows the comparison between the percentage of TET1 occupancy (in *Tdg[wt]*) in the total genome (top) as a reference, compared to the percentage of TET1 occupancy within TDG peaks (*Tdg[wt]*: middle, *Tdg[cat]*: bottom). The barplot on the right shows the same for TET2. **B:** Barplot illustrating the different overlaps between TDG, TET1 and TET2 in both *Tdg[wt]* (left) and *Tdg[cat]* (right). **C:** The percentage of the triple-positive peaks from **3B** located in a CGI or active enhancer (aE) is shown in a pie chart (left; *Tdg[wt]*: top and *Tdg[cat]*: bottom). As a comparison, the percentage of TDG only peaks in the respective genomic features is also shown (right). **D:** The fraction of TDG only and triple-positive peaks overlapping with an LMR is shown for *Tdg[wt]* (left) and *Tdg[cat]* (right). **E:** Gene ontology analysis (Ingenuity Pathway Analysis) of the fractions generated in **3B**. The *p* values of the five most significant molecular and cellular functions are depicted.

Figure 4: Distinct subgroups of TET1 and TET2 associate to chromatin – one subgroup depends on functional TDG, others do not

A: VENN diagram showing the overlap between all TET1 peaks in different TDG backgrounds (*Tdg[wt]* green, *Tdg[cat]* red, *Tdg[vec]* blue; left) at 24h RA. Bar diagram indicating the total number of peaks for TET1 in the different TDG backgrounds (right). **B:** VENN diagram showing the overlap between all TET2 peaks in different TDG backgrounds (*Tdg[wt]* green, *Tdg[cat]* red, *Tdg[vec]* blue; left) at 24h RA and Bar diagram indicating the total number of peaks for TET2 in the different TDG backgrounds (right).

Figure 5: TDG-TET1-TET2 co-localize at sites of active DNA demethylation and co-localize with dynamic histone variants

A: Barplot showing the fraction of unique TDG peaks and triple-positive peaks in *Tdg[wt]* (left) and *Tdg[cat]* (right) overlapping with DNA demethylation intermediates. Fractions are indicated for neither 5hmC nor 5fC containing DNA (None); 5fC only; 5hmC only; as well as the double-positive for 5fC and 5hmC. **B:** Barplot showing the fraction of unique TDG peaks and triple-positive peaks in *Tdg[wt]* (left) and *Tdg[cat]* (right) backgrounds overlapping with bivalent chromatin modifications. Fractions are indicated for neither H3K4me3 nor H3K27me3 containing histones (None); H3K4me3 only; H3K27me3 only; as well as the bivalent chromatin marks H3K4me3 together with H3K27me3. **C:** Barplot showing the fraction of unique TDG peaks and triple-positive peaks in *Tdg[wt]* (left) and *Tdg[cat]* (right) backgrounds overlapping with the histone variant H3.3. **D:** Barplot showing the fraction of unique TDG peaks and triple-positive peaks in *Tdg[wt]* (left) and *Tdg[cat]* (right) backgrounds overlapping with the histone variant H2A.Z.

Figure 6: Model

Model describing the association of the TET proteins and TDG at CpG-rich sites (promoter CGIs) and enhancers, where they induce active DNA demethylation and presumably nucleosome exchange. We propose that the TET proteins and TDG thereby facilitate enhancer-promoter contacts, and hence gene expression, of developmental genes during differentiation.

Supplementary Figure 1: Protein levels of TDG, TET1 and TET2 in mESCs

Western blots for TDG (**A**), TET1 (**B**) and TET2 (**C**) in mESCs with different TDG backgrounds (*Tdg[wt]*, *Tdg[cat]* and *Tdg[null]*; denaturing protein extracts). Different conditions are shown: mESCs from 2i preconditioning medium (2i) and 0h, 4h, 8h and 24h after RA-induced differentiation. β -Actin is shown as a loading control. Cross-reference: these Western blots are also shown in Jacobs et al., manuscript in preparation, 2014.

Supplementary Figure 2 (to Figure 1): Correlation between TDG ChIP samples

The correlation of all called peaks is shown for all TDG ChIP-seq samples in *Tdg[cat]* and *Tdg[wt]* at 24h RA-induced differentiation in a density plot. R values are indicated.

Supplementary Figure 3 (to Figure 3): Correlation between TET1 ChIP samples

The correlation of all called peaks is shown for all TET1 ChIP-seq samples in *Tdg[cat]*, *Tdg[wt]* and *Tdg[null]* at 0h and 24h RA-induced differentiation in a density plot. R values are indicated.

Supplementary Figure 4 (to Figure 3): Correlation between TET2 ChIP samples

The correlation of all called peaks is shown for all TET2 ChIP-seq samples in *Tdg[cat]*, *Tdg[wt]* and *Tdg[null]* at 0h and 24h RA-induced differentiation in a density plot. R values are indicated.

Supplementary Figure 5 (to Figure 2): TDG preferentially localizes in an active and nucleosome-poor chromatin environment

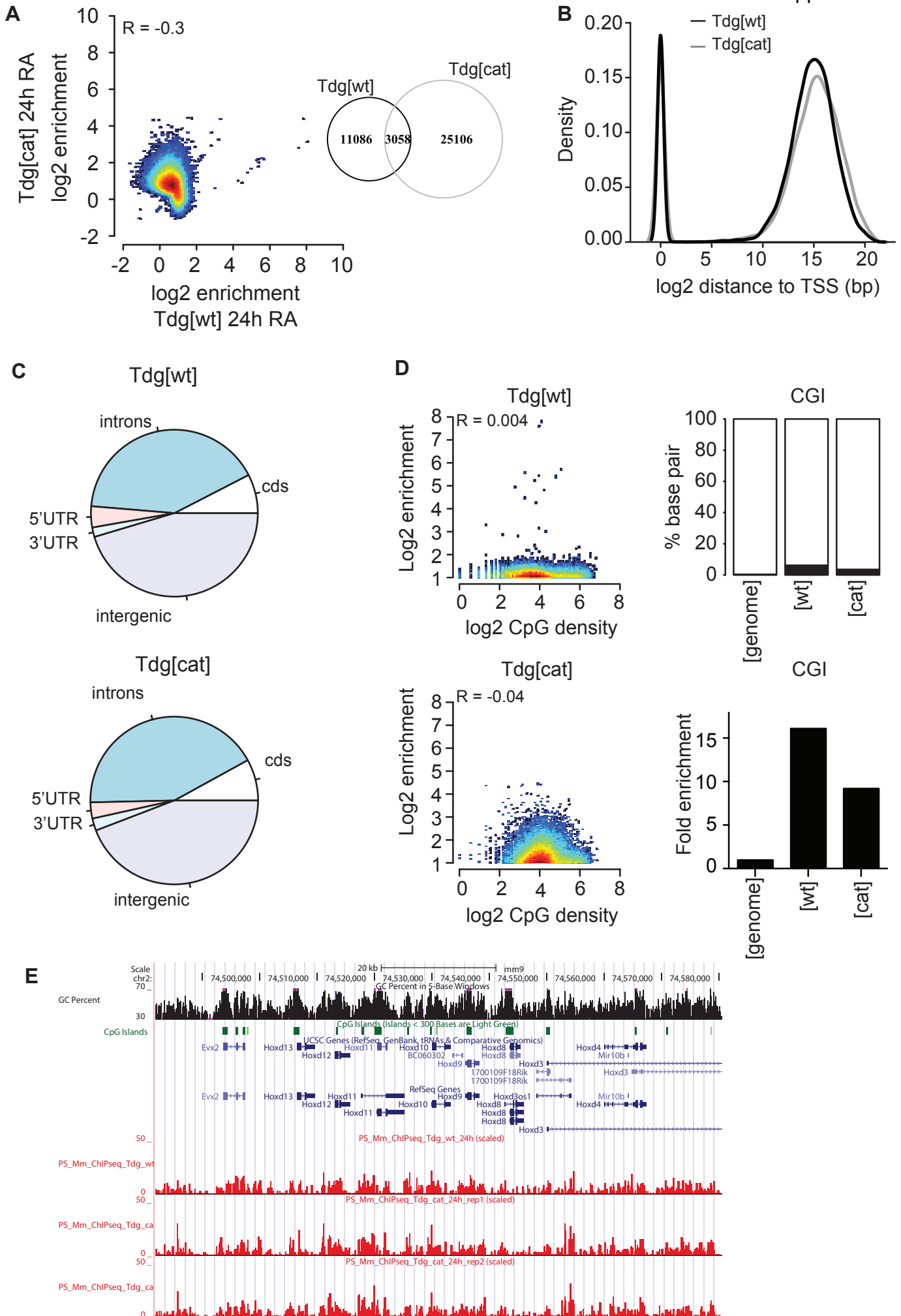
Bar charts are depicted, showing the percentage of coverage in the genome of the indicated feature as a reference (left), the percentage of the indicated feature occurring in *Tdg[wt]* peaks (middle) and *Tdg[cat]* peaks. **A:** DNase hypersensitive sites, RNA Polymerase II, p300 and CTCF. **B:** Repressive histone marks: H3K27me3 and H3K9me3. **C:** Active histone marks associated with transcription: H3K4me3, H3K36me3, H3K9ac. **D:** Features associated with active enhancers: H3K4me1, H3K27ac as well as the combination of these two marks together with DNase hypersensitivity in “active enhancers”, low-methylated regions (LMRs). **E:** Histone variants: H3.3 and H2A.Z.

Supplementary Figure 6 (to Figure 4): TET2 association to chromatin is regulated by active TDG during differentiation

A: VENN diagram showing the overlap between all TET1 peaks in different TDG backgrounds (*Tdg*[wt] green, *Tdg*[cat] red, *Tdg*[null] blue; left) at 0h RA. Bar diagram indicating the total number of TET1 peaks in the different TDG backgrounds (right).

B: VENN diagram showing the overlap between all TET2 peaks in different TDG backgrounds (*Tdg*[wt] green, *Tdg*[cat] red, *Tdg*[null] blue; left) at 0h RA. Bar diagram indicating the total number of TET2 peaks in the different TDG backgrounds (right).

Figure 1



Appendix II

Figure 2

Appendix II

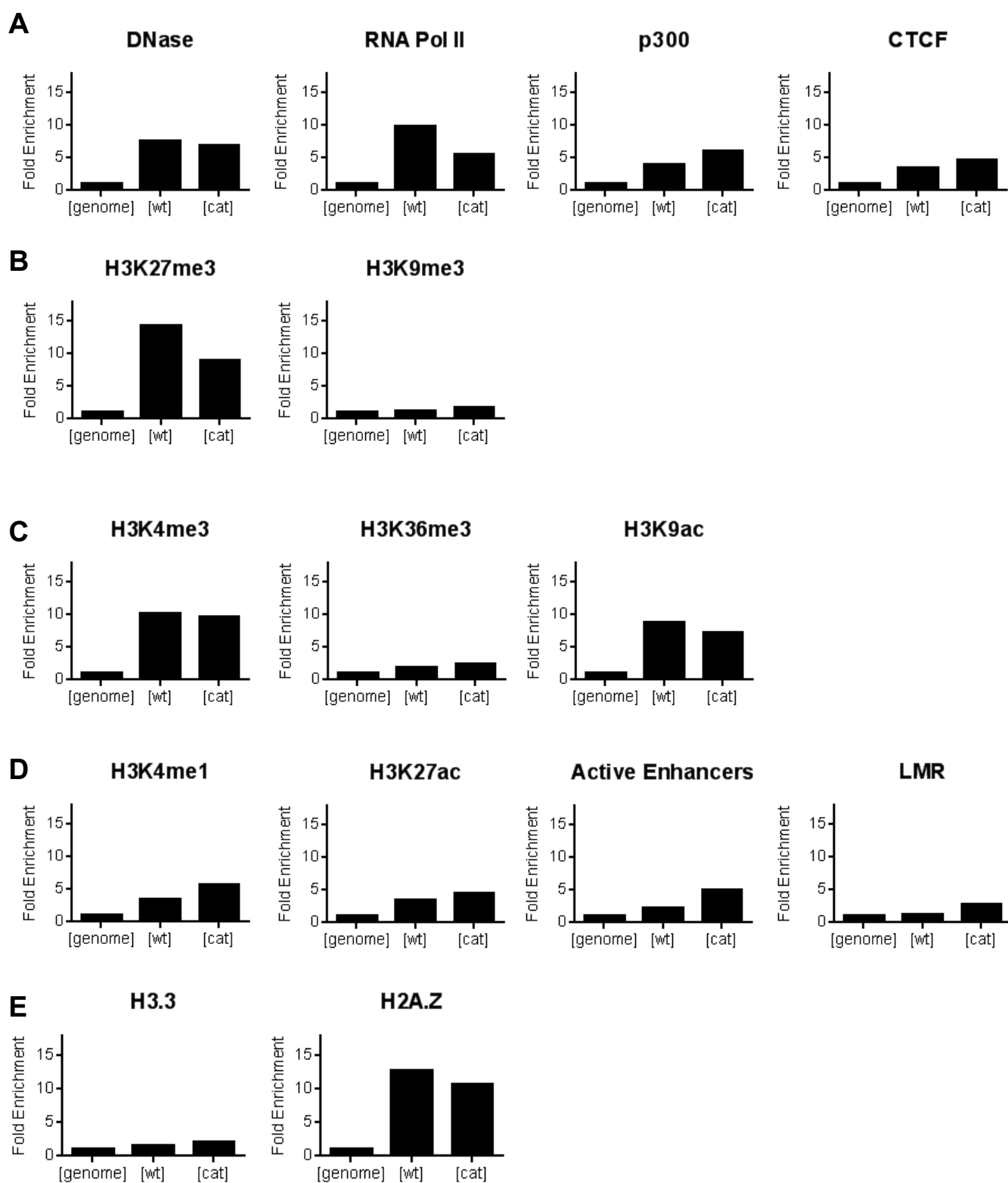
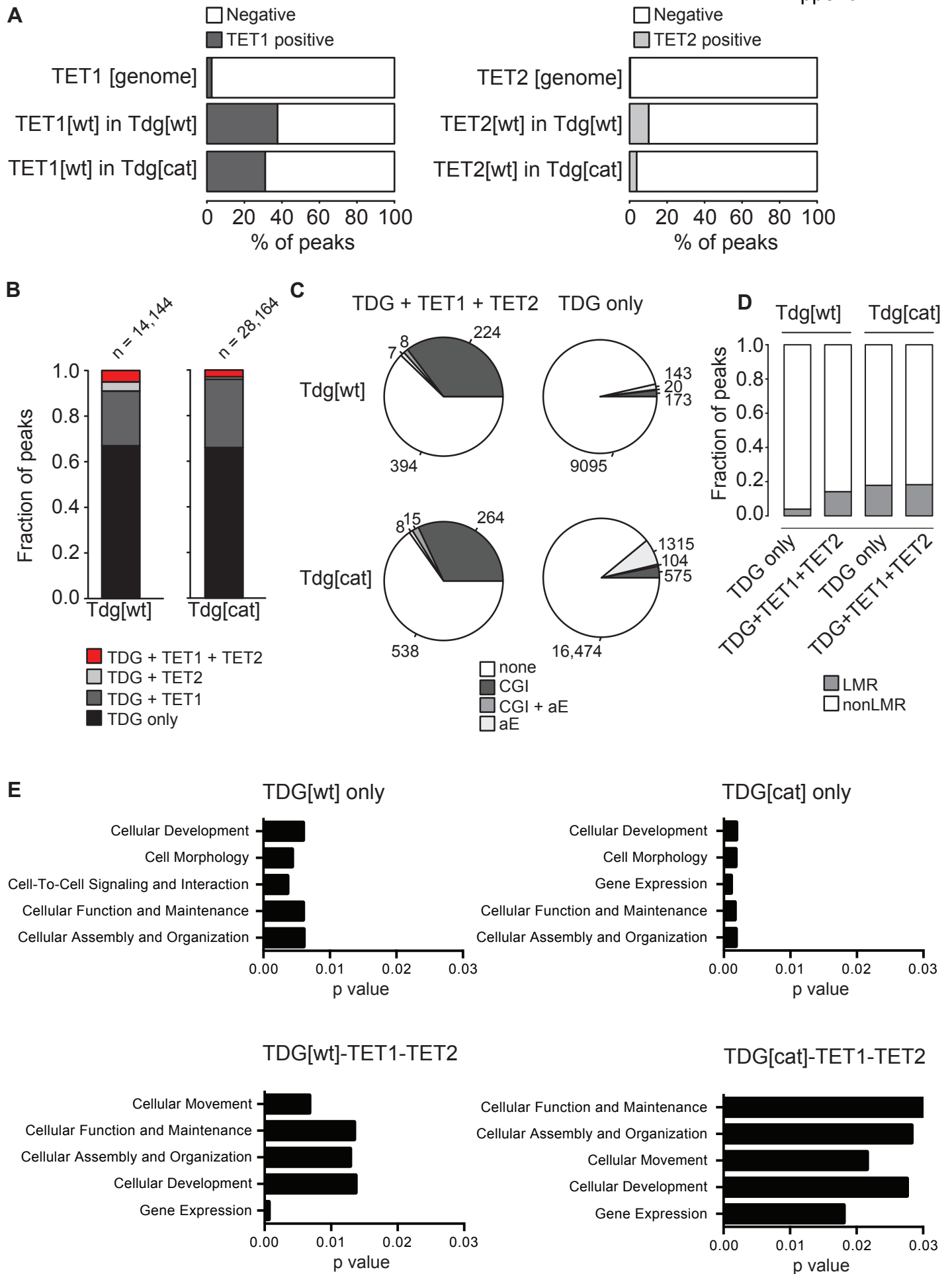


Figure 3



Appendix II

Figure 4

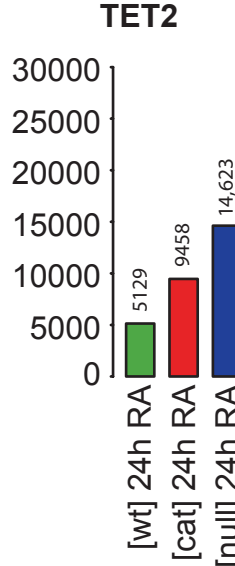
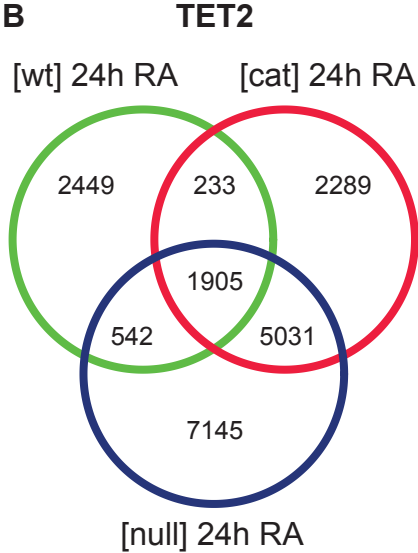
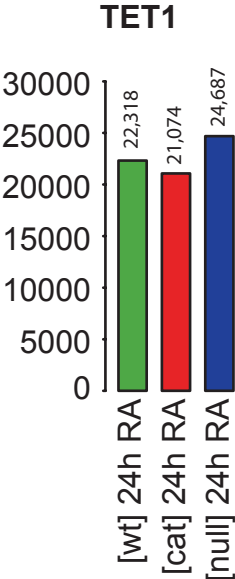
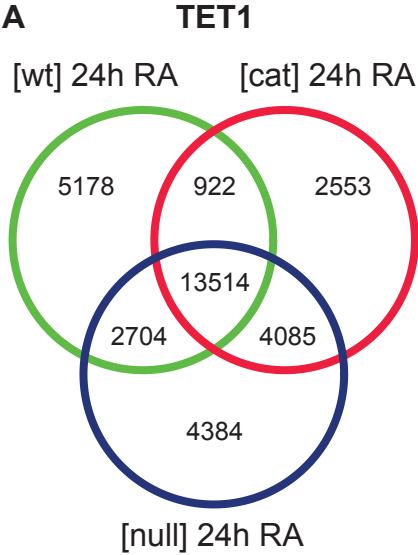


Figure 5

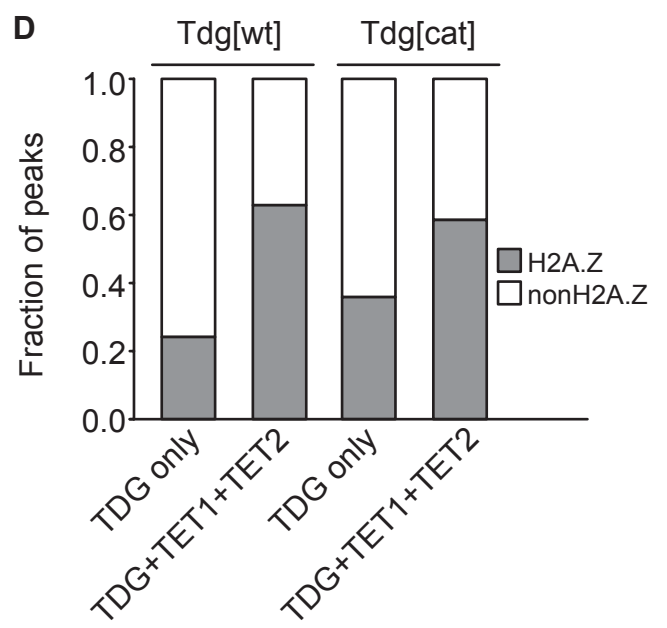
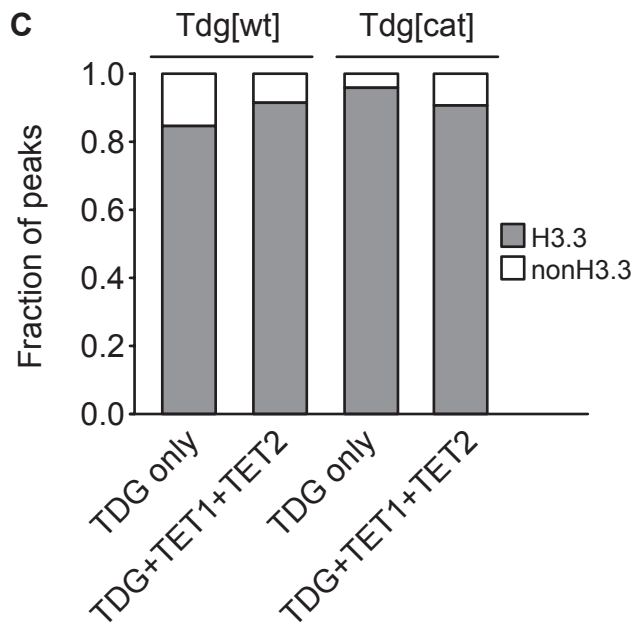
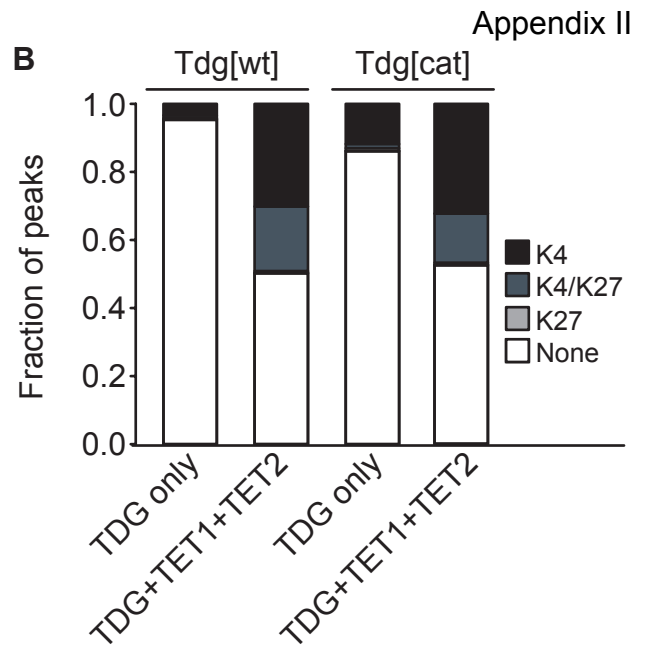
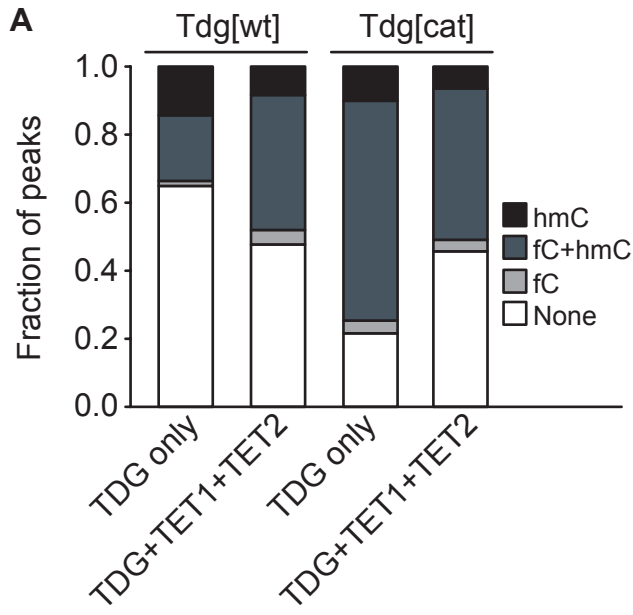
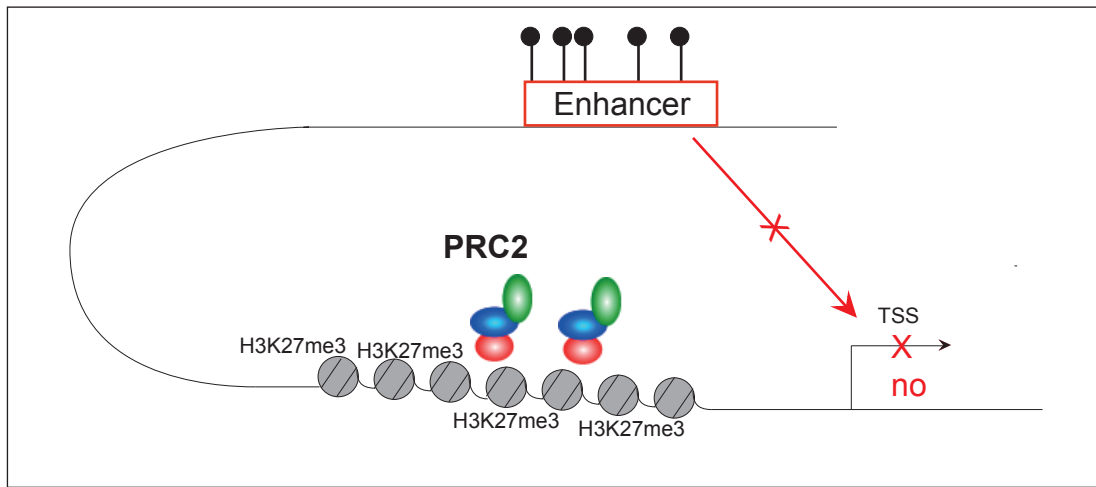
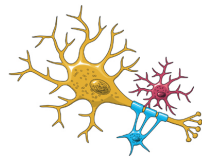


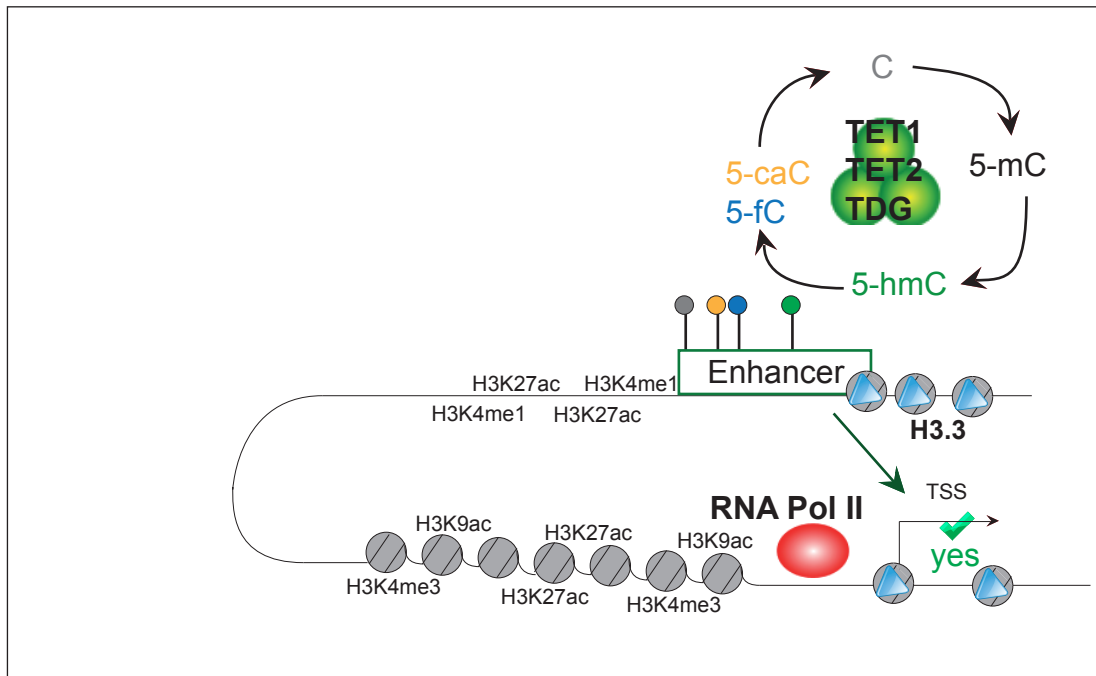
Figure 6



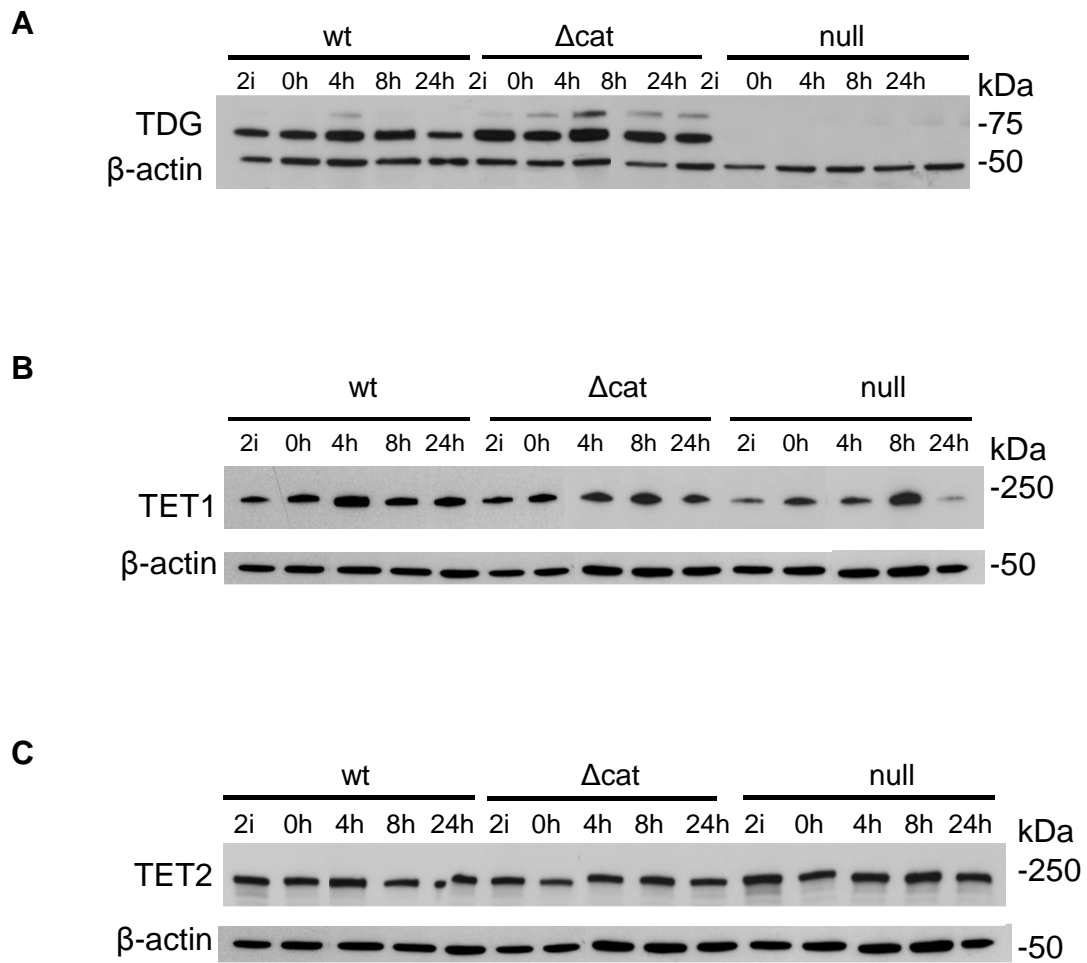
Differentiation

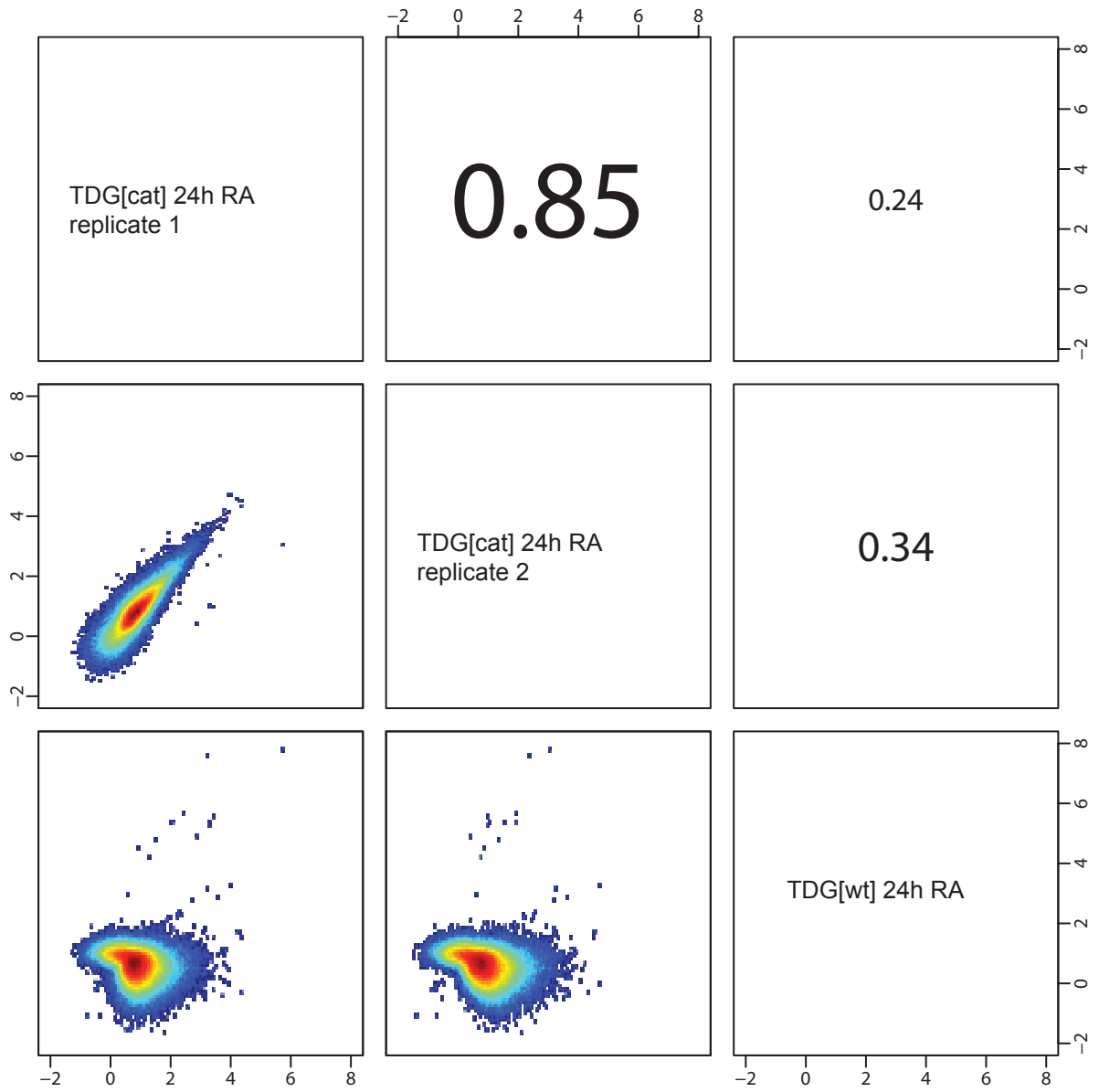


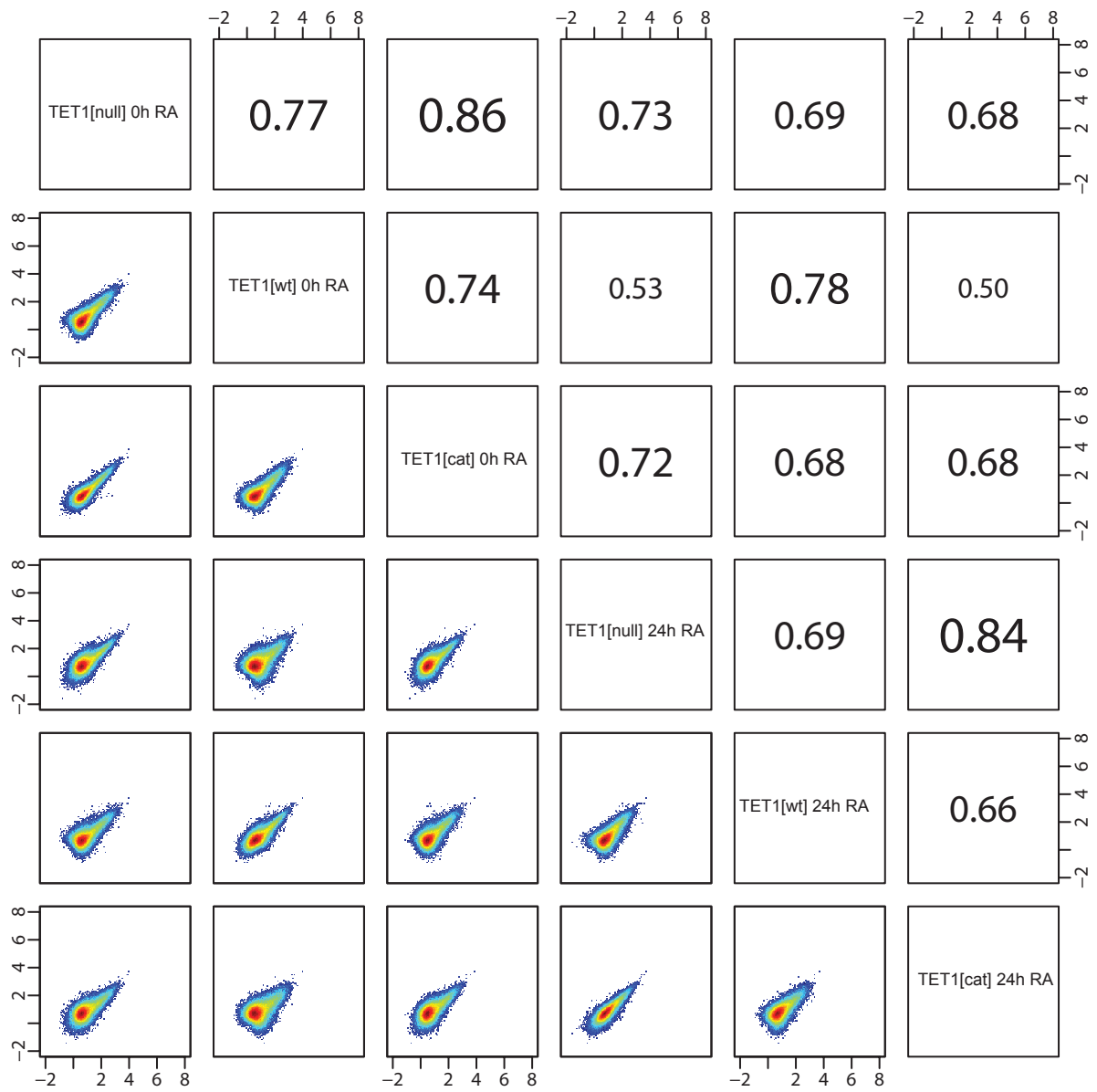
e.g. differentiated neuron
(at specific loci)

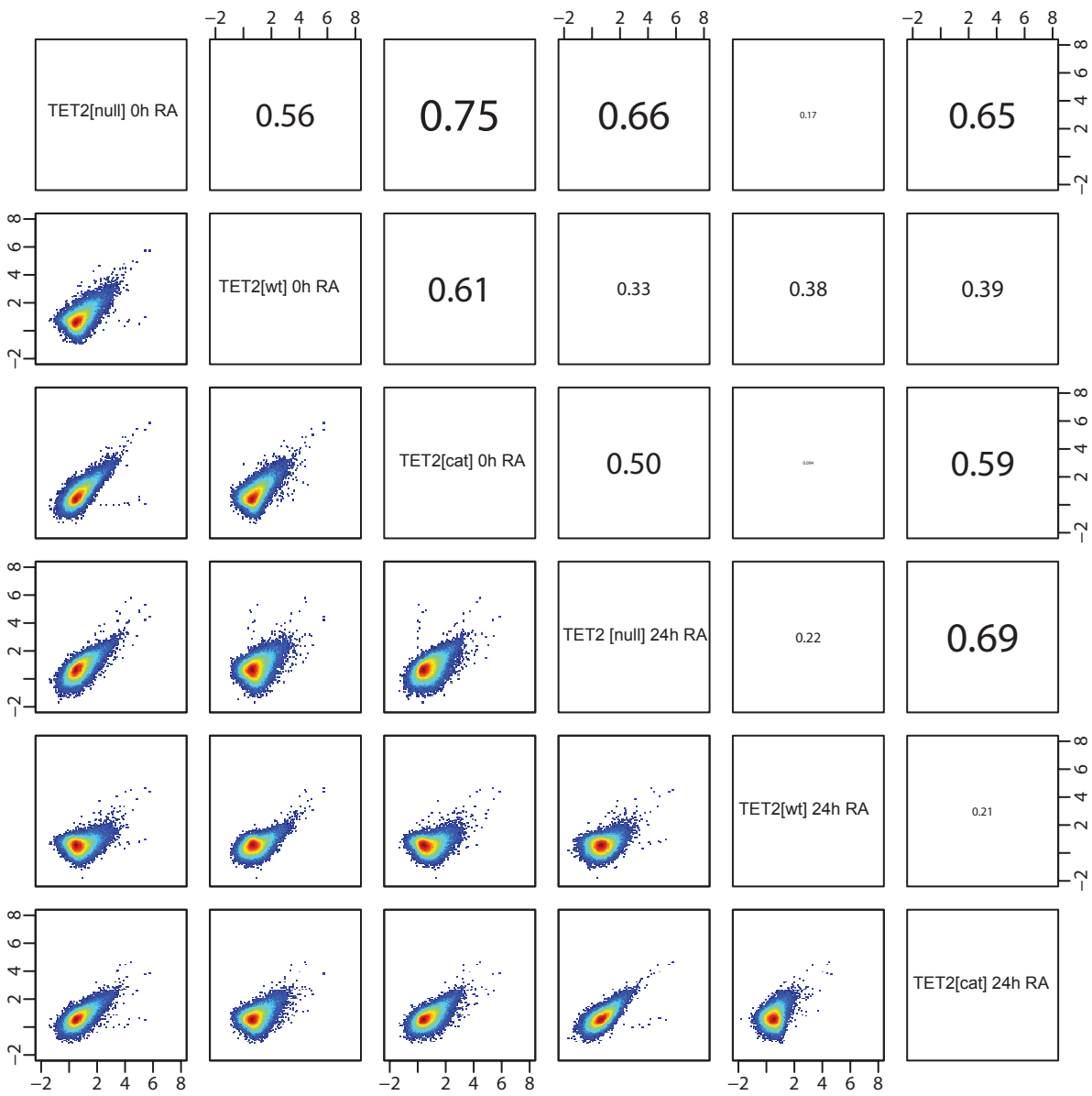


Supplementary Figure 1



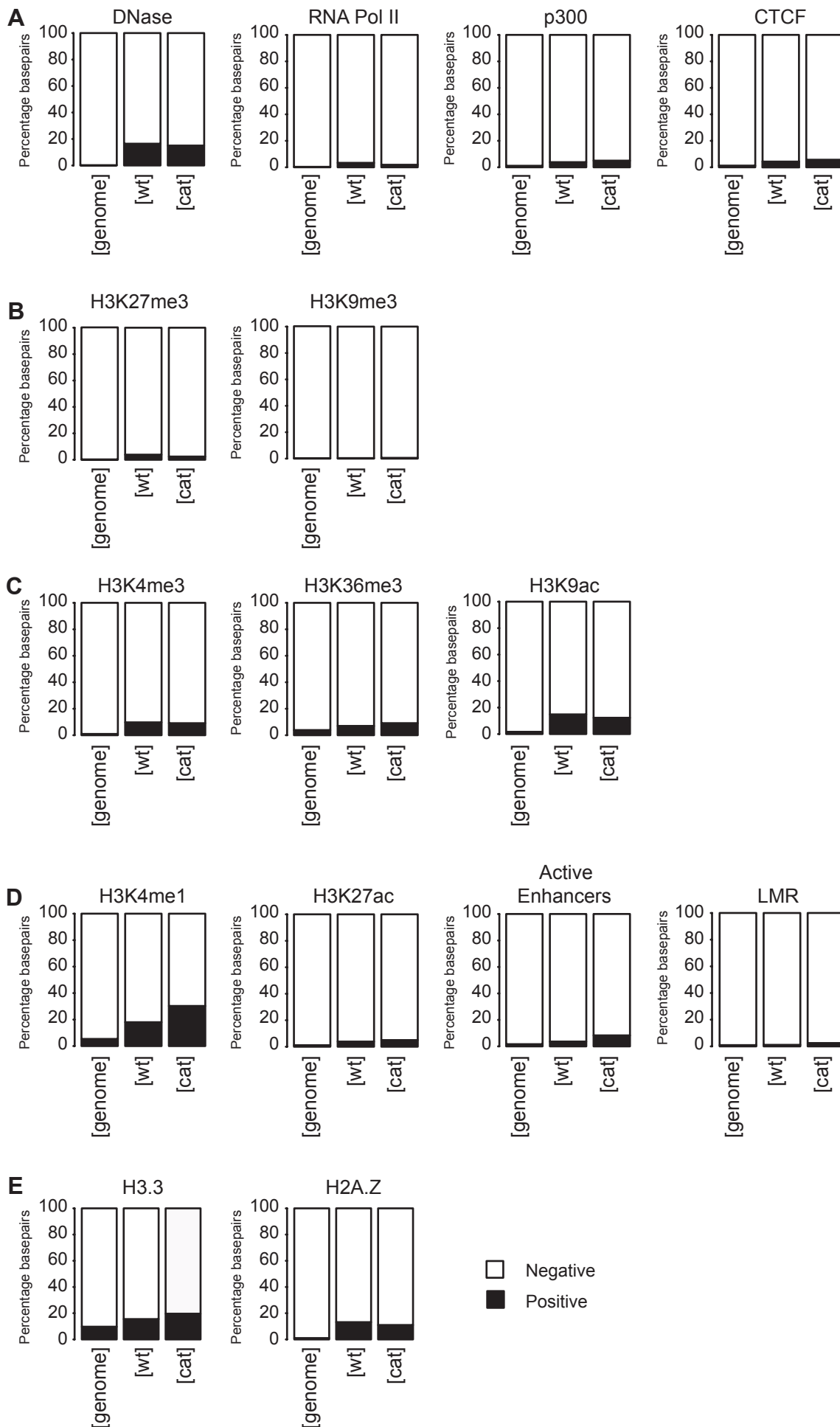


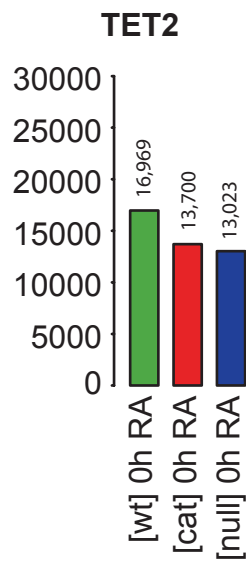
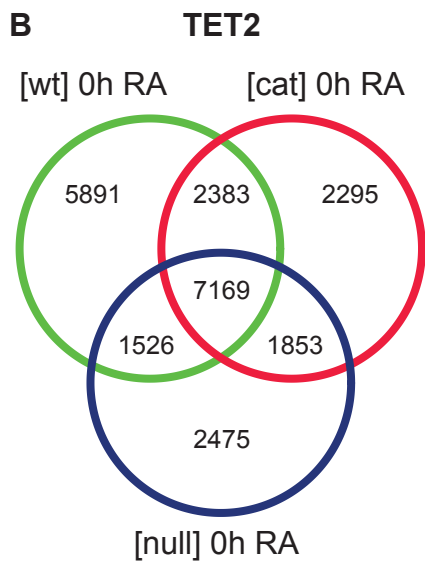
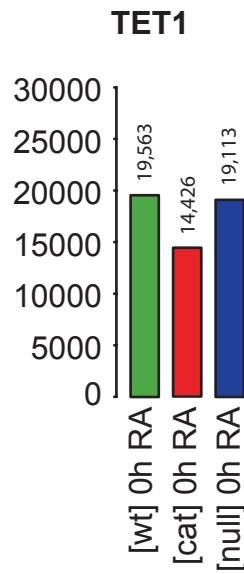
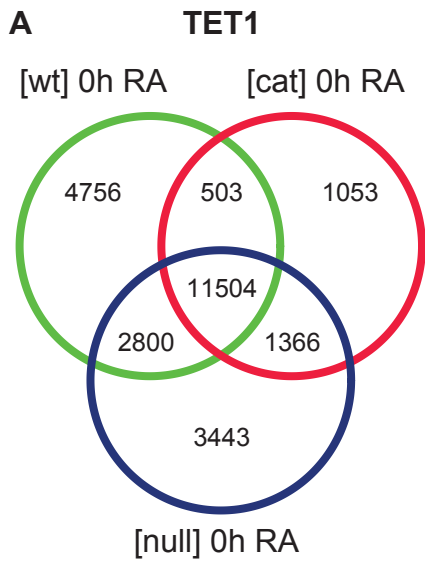




Supplementary Figure 5

Appendix II





Embryonic Lethal Phenotype Reveals a Function of TDG in Maintaining Epigenetic Stability

Cortázar, D.^{1,*}, Kunz, C.^{1,*}, Selfridge, J.², Lettieri, T.³, Saito, Y.¹, MacDougall, E.², **Wirz, A.**¹, Schuermann, D.¹, Jacobs, A. L.¹, Siegrist, F.⁴, Steinacher, R.¹, Jiricny, J.³, Bird, A.², Schär, P.¹

¹ Department of Biomedicine, University of Basel, Switzerland;

² The Wellcome Trust Centre for Cell Biology, University of Edinburgh, UK;

³ Institute of Molecular Cancer Research, University of Zürich, Switzerland;

⁴ Pharmaceutical Research, Global Preclinical Safety, F. Hoffmann-La Roche Ltd., Basel, Switzerland

* These authors contributed equally to this study

Correspondence:

primo.schaer@unibas.ch; Tel. +41 61 267 3561; Fax: +41 61 267 3566

Nature 470, 419–423 (17 February 2011); doi:10.1038/nature09672

Contribution: Contribution to the preparation of chromatin extracts from MEFs and mESCs and to ChIP experiments (H3K4me2, H3K9me3 and H3K27me3 in MEFs; Appendix III, Figure 1e; with DC). ChIP experiments in wildtype-complemented MEFs for all histone marks shown (Appendix III, Figure 2b) as well as for XRCC1 and CBP in MEFs (Appendix III, Figures 4a and 4b). ChIP experiments for H3K4me2 and H3K27me3 in mESCs (Appendix III, Supplementary Figure 8). ChIP experiments for H3K4me2, H3K9me3 and H3K27me3 in TDG catalytically dead MEFs (data not shown). Assistance with qPCR analysis of TDG ChIPs in MEFs and XRCC1 ChIPs in NPs (with DC).

Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability

Daniel Cortázar^{1*}, Christophe Kunz^{1*}, Jim Selfridge², Teresa Lettieri^{3†}, Yusuke Saito¹, Eilidh MacDougall², Annika Wirz¹, David Schuermann¹, Angelika L. Jacobs¹, Fredy Siegrist⁴, Roland Steinacher^{1†}, Josef Jiricny³, Adrian Bird² & Primo Schär¹

Thymine DNA glycosylase (TDG) is a member of the uracil DNA glycosylase (UDG) superfamily of DNA repair enzymes. Owing to its ability to excise thymine when mispaired with guanine, it was proposed to act against the mutability of 5-methylcytosine (5-mC) deamination in mammalian DNA¹. However, TDG was also found to interact with transcription factors^{2,3}, histone acetyltransferases⁴ and *de novo* DNA methyltransferases^{5,6}, and it has been associated with DNA demethylation in gene promoters following activation of transcription^{7–9}, altogether implicating an engagement in gene regulation rather than DNA repair. Here we use a mouse genetic approach to determine the biological function of this multifaceted DNA repair enzyme. We find that, unlike other DNA glycosylases, TDG is essential for embryonic development, and that this phenotype is associated with epigenetic aberrations affecting the expression of developmental genes. Fibroblasts derived from *Tdg* null embryos (mouse embryonic fibroblasts, MEFs) show impaired gene regulation, coincident with imbalanced histone modification and CpG methylation at promoters of affected genes. TDG associates with the promoters of such genes both in fibroblasts and in embryonic stem cells (ESCs), but epigenetic aberrations only appear upon cell lineage commitment. We show that TDG contributes to the maintenance of active and bivalent chromatin throughout cell differentiation, facilitating a proper assembly of chromatin-modifying complexes and initiating base excision repair to counter aberrant *de novo* methylation. We thus conclude that TDG-dependent DNA repair has evolved to provide epigenetic stability in lineage committed cells.

TDG is one of four enzymes with UDG activity in mammalian cells, but its biological function has remained enigmatic¹⁰. We thus set out to generate and phenotypically investigate a *Tdg* knockout mouse (Supplementary Fig. 1a–c). ESC clones carrying the targeted allele gave rise to healthy heterozygous *Tdg* knockout mice but attempts to generate homozygous null mutants failed, indicating that TDG-deficiency may cause embryonic lethality. This was unexpected, given the generally mild phenotype of other DNA glycosylase knockouts¹¹. In timed matings, *Tdg* null embryos isolated up to embryonic day (E) 10.5 appeared alive and normal, whereas those isolated at E12.5 were dead, and none were detectable at E16.5 (Fig. 1a and Supplementary Fig. 1d). *Tdg* null embryos at E10.5 produced viable fibroblasts (MEFs) but only a third of E11.5 embryos did so, suggesting that by this stage most of them were dead. We thus concluded that lethality in *Tdg* null embryos occurs around E11.5. For the actual cause of lethality, closer examination of the *Tdg* null embryos at E10.5 indicated internal haemorrhage, and evidence for haemorrhagic necrosis (data not shown), but otherwise did not reveal an informative pathology.

We then explored the essential function of TDG in MEFs and ESCs, first addressing a potential DNA repair defect by classical genotoxicity and mutator analyses. The TDG status did not affect cell survival

following ionizing radiation or H₂O₂ exposure, both of which induce DNA base lesions processed by TDG *in vitro*¹⁰, nor did it affect mutation frequencies in a Big Blue transgenic mutation assay (Supplementary Fig. 2). We therefore concluded that the role of TDG in the repair of canonical base damage is minor and therefore unlikely to account for its essential function in mouse embryogenesis.

We next investigated a possible involvement of TDG in gene regulation by expression profiling of TDG-proficient and -deficient MEFs. To limit potential clonal biases, we compared the transcriptomes of early passages of litter-matched populations of SV40 immortalized MEFs. This identified 461 differentially transcribed genes ($P \leq 0.05$, fold change (FC) ≥ 1.5 , Fig. 1b), comprising many transcription factors and, thus, likely reflecting both direct and indirect consequences of TDG loss. Global pathway analyses revealed gene networks associated with embryogenesis and development as being most significantly misregulated in the absence of TDG (Supplementary Fig. 3a). Four out of six target genes analysed showed TDG-dependent differential expression also in independently isolated primary MEFs (Supplementary Fig. 3b).

Considering its putative involvement in DNA demethylation^{7–9}, we next investigated a possible occurrence of aberrant promoter methylation in TDG-deficient cells. We examined the CpG islands in the promoters of *Hoxa10*, *Hoxd13*, *Sfrp2*, *Twist2* and *Rarb*, all of which were downregulated in TDG-deficient MEFs (Fig. 1b and Supplementary Fig. 3a). These genes are developmentally regulated by the polycomb repressive system¹² and their promoter CpG islands are unmethylated in most normal tissues but subject to aberrant *de novo* methylation in human cancers^{13,14}. Na-bisulphite sequencing of the respective CpG islands revealed an increased occurrence of *de novo* methylation in the TDG-deficient MEFs (Fig. 1c and Supplementary Figs 4 and 5a). The patterns and frequency of these methylation events indicated that the loss of TDG generates hotspots of *de novo* methylation in certain gene promoters. We then used chromatin immunoprecipitation (ChIP) to examine a possible association of TDG with the promoters of these and additional differentially expressed genes. Compared with a randomly chosen intergenic sequence or the silent promoters of *Oct4* and *Tuba3*, DNA fragments surrounding the promoters of all genes examined were significantly enriched in the TDG precipitates (Fig. 1d). This indicated that TDG is targeted to specific gene promoters, possibly to protect them from acquiring aberrant CpG methylation and eventual epigenetic silencing. Consistently, further examination of the chromatin status revealed a general loss of activating (H3K4me2) and a concomitant increase of repressive histone marks (H3K27me3, H3K9me3) in TDG-deficient cells with promoter-specific patterns (Fig. 1e): a complete loss of H3K4 dimethylation was accompanied by a strong increase of H3K27 and/or H3K9 trimethylation at the *Hoxd13* and *Hoxa10* promoters; a partial reduction of H3K4me2 coincided with an enrichment of H3K27me3 but not H3K9me3 at

¹Institute of Biochemistry and Genetics, Department of Biomedicine, University of Basel, 4048 Basel, Switzerland. ²The Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK. ³Institute of Molecular Cancer Research, University of Zürich, 8057 Zürich, Switzerland. ⁴Pharmaceutical Research, Global Preclinical Safety, F. Hoffmann-La Roche Ltd., 4058 Basel, Switzerland. [†]Present addresses: European Commission, Joint Research Centre, Institute for Environment and Sustainability, 21027 Ispra, Italy (T.L.); Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK (R.S.).

*These authors contributed equally to this work.

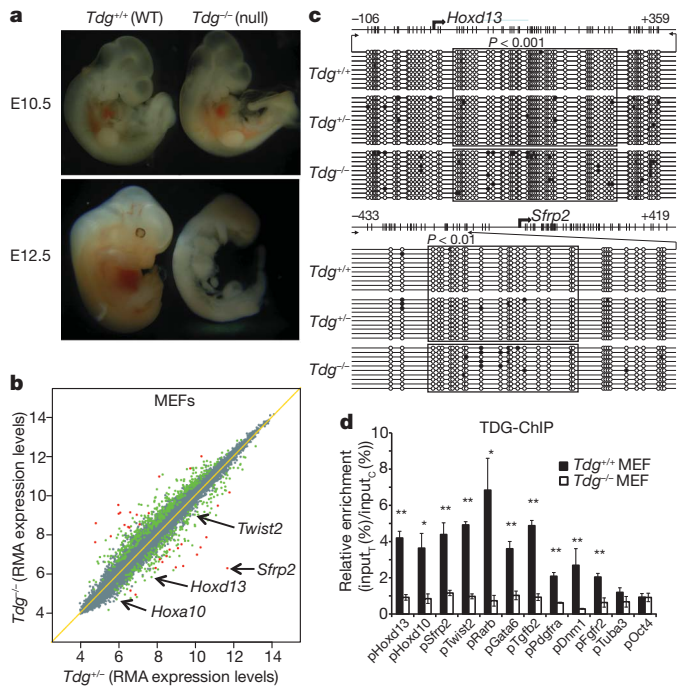


Figure 1 | Embryonic essential function of *Tdg* in epigenetic gene regulation. **a**, Whole mount images of typical examples of *Tdg*^{+/+} and *Tdg*^{-/-} littermate embryos taken at E10.5 and E12.5. **b**, Scatter plot comparing gene expression levels of matched *Tdg*^{+/+} and *Tdg*^{-/-} MEFs. Differentially expressed genes at $P < 0.05$ and $P < 0.01$ are indicated by green and red dots, respectively, and examples of developmental genes affected are denoted. **c**, Na-bisulphite sequencing of the *Hoxd13* and *Sfrp2* promoters in *Tdg*^{+/+}, *Tdg*^{+/-} and *Tdg*^{-/-} MEFs. White and black circles indicate unmethylated and methylated CpGs, respectively. P values indicate statistical difference of methylation frequencies as determined by contingency tables and χ^2 test. **d**, ChIP-quantitative PCR (qPCR) analysis of TDG association with the promoters of the genes indicated in chromatin from *Tdg*^{+/+} and *Tdg*^{-/-} MEFs. Shown are relative enrichments of TDG at these promoters normalized to a randomly chosen intergenic control region (means \pm s.e.m.; $n \geq 3$; $*P < 0.05$; $**P < 0.01$, unpaired Student's t -test). **e**, ChIP-qPCR analyses in *Tdg*^{+/+} and *Tdg*^{-/-} MEFs to assess the presence of activating (H3K4me2) and repressive (H3K9me3, H3K27me3) histone modifications at the promoter regions indicated. Shown are enrichments relative to appropriate negative controls: intracisternal A-particle (*Iap*) transposon for active chromatin marks and the *Hprt* promoter for silencing marks (means \pm s.e.m.; $n \geq 3$; $*P < 0.05$; $**P < 0.01$; unpaired Student's t -test). Subscript _T, target region; subscript _C, control region.

the *Sfrp2* and *Twist2* promoters; and reduction of H3K4me2 was coupled with an increase in H3K9me3 but not H3K27me3 at the *Rarb* promoter. Thus, promoter *de novo* methylation in TDG-deficient cells is associated with a loss of H3K4 dimethylation and a concomitant increase in trimethylation of H3K27 more than H3K9.

Stable expression of a TDG encoding complementary DNA (cDNA) in *Tdg*^{-/-} MEFs (Supplementary Fig. 1f) restored activity to the *Sfrp2* and *Twist2* genes (Fig. 2a). This correlated with a loss of H3K27 trimethylation in their promoters and an increase in H3K4 dimethylation in the case of *Twist2* (Fig. 2b). Expression of *Hoxd13* and *Hoxa10*, however, was not restored although a partial reduction of H3K27 trimethylation also occurred. This indicated that, once H3K4 methylation is lost (*Hoxd13*, *Hoxa10*), the repressive chromatin maintained by

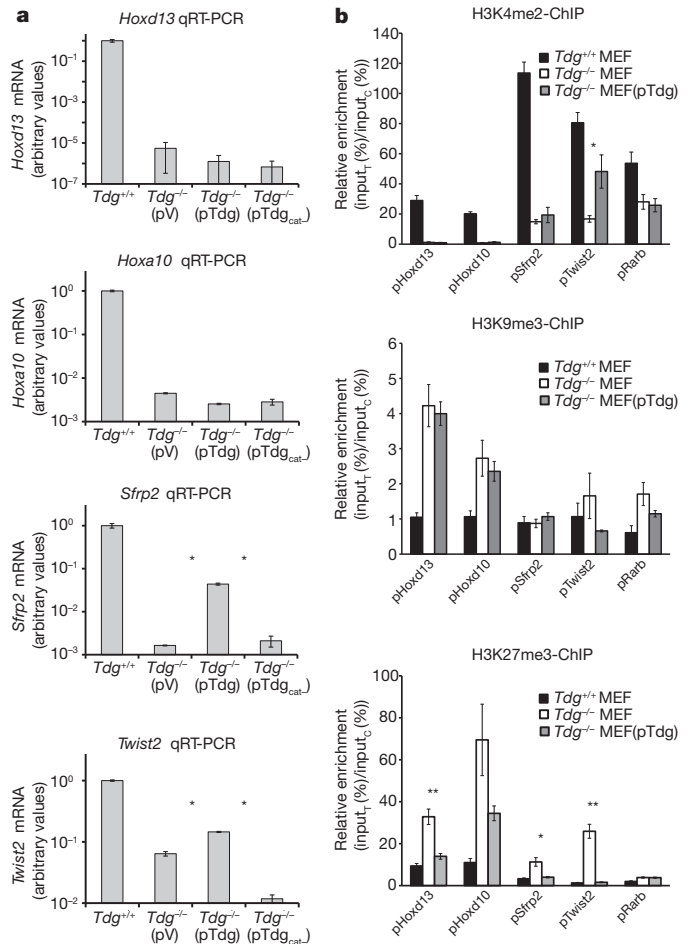


Figure 2 | Complementation of the loss of gene expression depends on the chromatin state of the promoter. **a**, *Hoxd13*, *Hoxa10*, *Sfrp2* and *Twist2* expression in *Tdg*^{+/+} and *Tdg*^{-/-} MEFs complemented with vectors expressing either a wild-type (pTdg) or a catalytically deficient *Tdg* (pTdg_{cat}, N151A), or a vector control (pV). Target-specific messenger RNA (mRNA) levels were assessed by qRT-PCR and normalized to *Gapdh* mRNA; values represent arbitrary units (means \pm s.d.; $n \geq 3$; $*P < 0.05$; unpaired Student's t -test). **b**, ChIP-qPCR analyses to detect H3K27me3 and H3K4me2 marks at the gene promoters indicated in chromatin of *Tdg*^{+/+}, *Tdg*^{-/-} and *Tdg*^{-/-} MEFs complemented with a wild-type *Tdg* cDNA. *IAP* and the *Hprt* promoter were used as normalizers for active and repressive chromatin marks, respectively (means \pm s.e.m.; $n = 3$; $*P < 0.05$; $**P < 0.01$; unpaired Student's t -test).

H3K9 and H3K27 methylation and aberrant CpG methylation cannot be reversed to an active state by re-expression of *Tdg*. If residual H3K4 methylation is present, however, promoter reactivation is possible, and this requires the catalytic function of TDG¹⁵ as shown for *Sfrp2* and *Twist2* (Fig. 2a).

To address the origin of the epigenetic aberrations in *Tdg* null MEFs, we investigated gene expression and chromatin states in TDG-proficient and -deficient ESCs before and after retinoic-acid-induced *in vitro* differentiation to neuronal progenitor cells¹⁶ (Supplementary Fig. 6a). Strikingly, gene expression differences were minor in ESCs (16 genes, $P \leq 0.05$, $FC \geq 1.5$) but increased significantly upon differentiation to neuronal progenitor cells (297 genes, $P \leq 0.05$, $FC \geq 1.5$) (Fig. 3a). This was not due to an inability of TDG-deficient ESCs to respond transcriptionally to retinoic acid (Supplementary Fig. 6b), although they showed somewhat faster kinetics of silencing pluripotency genes (*Oct4*, *Nanog*) and activating developmental genes (for example, *Gata6*, *Pax6*) (Supplementary Fig. 6c). Similar to the situation in MEFs, the genes most significantly misregulated in TDG-deficient neuronal progenitor cells control developmental functions, most of

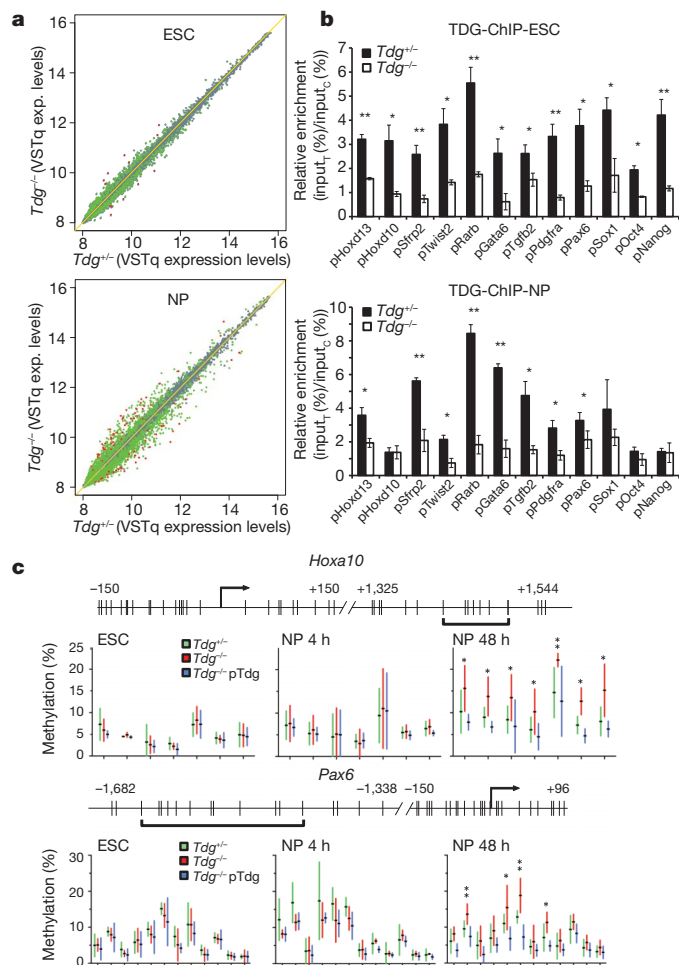


Figure 3 | TDG-dependent differences in gene expression and chromatin status arise during cell differentiation. **a**, Scatter plots comparing gene expression profiles of *Tdg*^{+/-} and *Tdg*^{-/-} ESCs or *in vitro* differentiated neuronal progenitors. Green and red dots indicate differentially expressed genes at $P < 0.05$ and $P < 0.01$, respectively. **b**, ChIP-qPCR analysis of TDG association with the gene promoters indicated in chromatin from *Tdg*^{+/-} and *Tdg*^{-/-} ESCs and neuronal progenitor (NP) cells. Shown is the relative enrichment of TDG at these promoters normalized to a randomly chosen intergenic control region (means \pm s.e.m.; ESCs, $n = 3$; neuronal progenitor cells, $n = 3$; * $P < 0.05$, ** $P < 0.01$; unpaired Student's *t*-test). **c**, DNA methylation states at the *Hoxa10* and *Pax6* promoters in TDG-deficient ESCs and neuronal progenitor cells analysed by bisulphite pyrosequencing. Promoter regions are depicted schematically at the top. Vertical tick marks indicate CpG sites, bent arrows transcription start sites, and horizontal brackets the CpGs for which methylation data are presented in the graphs. Methylation levels are given as the percentage of methylated cytosines at each CpG analysed. Shown are means with the 95% confidence intervals (bars) obtained from three differentiation experiments. * $P < 0.05$; ** $P < 0.01$; unpaired Student's *t*-test.

them having CpG islands in their promoters and being targets of the polycomb repressive system (Supplementary Fig. 7a). Using ChIP, we confirmed an enrichment of TDG at the promoters of differentially expressed genes both in ESCs and in neuronal progenitor cells (Fig. 3b). This also revealed that TDG associates with the promoters of *Oct4* and *Nanog* in ESCs but not in neuronal progenitor cells and MEFs (Fig. 3b and Supplementary Fig. 6d), suggesting that its interaction is lost upon heterochromatinization of these promoters. Notably, the inability to associate with heterochromatinized promoters may explain why re-expression of TDG in *Tdg* null MEFs failed to restore *Hoxd13* and *Hoxa10* transcription (Fig. 2).

Next, we examined the status of CpG methylation in gene promoters downregulated in TDG-deficient neuronal progenitor cells, making use of Na-bisulphite (pyro)sequencing and methylated

DNA-immunoprecipitation (MeDIP). Although MeDIP only detected trends for methylation differences at specific promoters (Supplementary Fig. 7b and unpublished observations), pyrosequencing revealed significantly increased DNA methylation in *Tdg* null neuronal progenitor cells at three out of five gene promoters tested (*Hoxa10*, *Pax6*, *Tgfb2*). Notably, these methylation differences were not present in ESCs nor in freshly dissociated embryonic bodies, they arose within 48 h of cultivation of the neuronal progenitor cells in progenitor medium (Fig. 3c and Supplementary Fig. 7c), and the phenotype was complemented by ectopic expression of *Tdg* during cell differentiation. Similarly, histone methylation marks were not different between TDG-proficient and -deficient ESCs but arose in neuronal progenitor cells with an enrichment of H3K27me3 at the promoters of *Hoxd13*, *Hoxa10* (Supplementary Fig. 8) and *Pdgfra* (unpublished observations). Thus, differences in DNA methylation and histone modifications became apparent at the neuronal progenitor cell stage but were not as pronounced as in MEFs, indicating an epigenetic phenotype that may progress upon further differentiation and/or cultivation. Consistently, attempts to differentiate TDG-deficient neuronal progenitor cells to terminal neurons failed because of a rapid loss of cell viability in neuronal-rich medium.

We then wondered whether this epigenetic function of TDG involves active DNA repair, as implicated by the inability of a catalytic-dead TDG (N151A) to complement the loss of *Sfrp2* and *Twist2* expression in *Tdg* null MEFs (Fig. 2). To monitor a possible engagement of downstream base excision repair, we first performed ChIP for XRCC1¹⁷. This revealed a specific, TDG-dependent enrichment of this critical base excision repair protein at the *Hoxd13*, *Hoxa10*, *Sfrp2* and *Twist2* promoters in MEFs but not in ESCs (Fig. 4a and Supplementary Fig. 5b). Hence, in MEFs, where TDG helps maintain these promoters in an active state, its presence correlates with an enrichment of XRCC1. In ESCs, however, where TDG also associates with these promoters but does not affect their chromatin status, XRCC1 enrichment is not observed. Besides XRCC1, we also found APE1, another component of base excision repair, to associate with these promoters in a TDG dependent manner in MEFs (Fig. 4a). Moreover, retinoic acid treatment of ESCs for 8 h increased the number of chromatin-associated XRCC1 foci in the presence but not in the absence of TDG (Supplementary Fig. 9), and TDG-proficient cells were significantly more sensitive to poly(ADP-ribose) polymerase (PARP) inhibition than TDG-deficient cells upon retinoic-acid-induced differentiation (Supplementary Fig. 10). These observations strongly suggest that cell differentiation-induced TDG activity feeds into PARP and XRCC1-dependent DNA single-strand break repair¹⁸.

Addressing the phenotype on histone modifications, we then found by ChIP that the absence of TDG also compromises the association of the H3K4-specific methyltransferase MLL1 with the promoters of *Hoxd13*, *Hoxa10*, *Sfrp2* and *Twist2* (Fig. 4b). This was apparent in TDG-deficient MEFs but not in ESCs, with the former indeed showing a loss of H3K4 methylation and an occurrence of aberrant CpG methylation at gene promoters reminiscent of the phenotype of MLL defects^{19–21}. Similar to MLL, the binding of CBP/p300 to these promoters was significantly reduced in the *Tdg* null MEFs (Fig. 4b). CBP/p300 is a transcription-activating histone acetyltransferase known to interact with TDG⁴ and, notably, its association with gene promoters has been reported to protect from polycomb-mediated H3K27 trimethylation²².

Taken together, our data suggest structural and catalytic functions of TDG in epigenetic maintenance (Fig. 4c). As a structural component, TDG complexes with activating histone modifiers (for example, MLL, CBP/p300) to maintain states of active (H3K4me2) and bivalent (H3K4me2, H3K27me3) chromatin during cell differentiation. In the absence of TDG, the assembly and function of such complexes is distorted and, consequently, chromatin modifications imbalanced towards repressive states. TDG also provides DNA repair capacity to erase CpG methylation locally. Aberrant methylation arises at GC-rich promoters in TDG-deficient cells following lineage commitment, and the frequencies and patterns of these events indicate an underlying

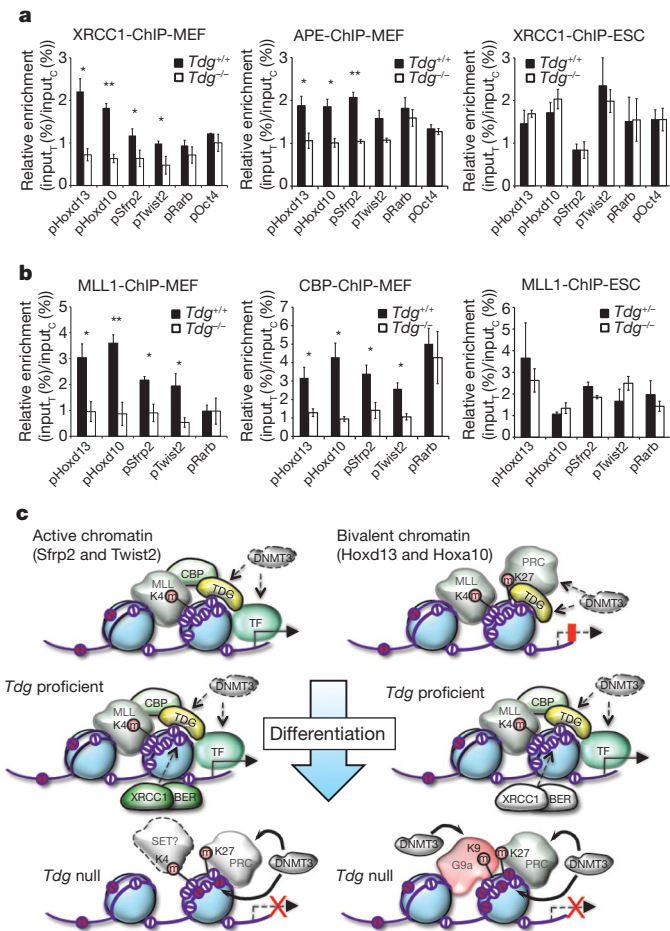


Figure 4 | Structural and catalytic functions of TDG in epigenetic maintenance. **a**, ChIP-qPCR analysis of XRCC1 and APE1 association with the gene promoters indicated in chromatin of TDG-proficient and -deficient MEFs and ESCs. Shown are relative enrichments of XRCC1 and APE1 at these promoters normalized to a randomly chosen intergenic control region (means \pm s.e.m.; $n \geq 3$; * $P < 0.05$; ** $P < 0.01$; unpaired Student's *t*-test). **b**, ChIP-qPCR analysis of MLL1 and CBP/p300 association with the gene promoters indicated in chromatin of TDG-proficient and -deficient MEFs and ESCs. Shown are relative enrichments of MLL1 and CBP/p300 at these promoters normalized to a randomly chosen intergenic control region (means \pm s.e.m.; $n \geq 3$; * $P < 0.05$; ** $P < 0.01$; unpaired Student's *t*-test). **c**, Model summarizing epigenetic aberrations and implicated functions observed in the absence of TDG. In ESCs TDG associates with gene promoters in an active 'open' (H3K4me₂, for example; *Sfrp2* and *Twist2*, left side) or transiently silent 'bivalent' chromatin conformation (H3K4me₂ and H3K27me₃, for example; *Hoxd13* and *Hoxa10*, right side). In active chromatin, the lack of TDG results in a partial loss of H3K4 dimethylation and a gain of H3K27 trimethylation as well as in sporadic DNA hypermethylation (red balls) upon cell differentiation. Differentiation-associated activation of promoters in 'bivalent' chromatin involves the demethylation of H3K27me₃ and transcription factor binding. The absence of TDG results in an aberrant loss of H3K4 dimethylation accompanied by a gain in repressive H3K9 and H3K27 trimethylation and in DNA methylation, eventually directing irreversible transcriptional silencing. In both cases, the loss of active and the gain in repressive histone marks can be accounted for by a failure of TDG-deficient cells to target MLL and CBP to these promoters. We propose that TDG, as part of transcription regulatory complexes, assures the establishment and the maintenance of proper epigenetic states at developmentally regulated gene promoters. As a DNA glycosylase, it protects these regions from aberrant CpG methylation in a process that engages XRCC1 and APE1, factors essential for downstream base excision repair.

stochastic process of *de novo* methylation. Hence, TDG keeps *de novo* DNMT activities in check to avoid erroneous methylation, and the engagement of XRCC1 and APE1 suggests that it operates through

base excision repair. Several previous studies have implicated TDG in active DNA demethylation^{8,9,23}. Mechanistically, it may do so on its own, acting as a 5-mC DNA glycosylase²³, or it may cooperate with a 5-mC deaminase (for example, AID/Apobec^{24,25} or DNMTs⁸), or a 5-mC hydroxylase (for example, TET1^{26,27}) that would convert 5-mC into a favourable substrate for TDG. Numerous efforts to reproduce 5-mC glycosylase activity for mouse and human TDG have failed (Supplementary Fig. 11 and unpublished observations). We therefore consider a deamination or hydroxylation-mediated, TDG-dependent repair process a preferable model for active cytosine demethylation. The mouse *Tdg* knockout phenotype shows that such an epigenetic control system has evolved to protect critical DNA sequences from *de novo* methylation and heterochromatinization during development.

METHODS SUMMARY

***Tdg* knockout mouse and cell lines.** The *Tdg*-targeting construct (Supplementary Fig. 1) was generated by replacement of a *NarI*-*PacI* fragment enclosing exons 6 and 7 by a neomycin resistance cassette in a cloned fragment spanning exons 5–10 of the *Tdg* locus. This construct was used to target the *Tdg* allele in 129 mouse ESCs, which were then used to generate chimaeras and, ultimately, *Tdg*^{+/-} heterozygotes by backcrossing to C57BL/6. The generation and establishment of MEFs and *Tdg*^{-/-} ESCs was previously described²⁸.

***In vitro* differentiation.** *In vitro* differentiation of ESCs was performed essentially according to the protocol published in ref. 16. RNA isolation for transcriptome analysis of MEFs or ESCs and neuronal progenitor cells was performed using the RNeasy Mini Kit (Qiagen) or the Trizol reagent (Invitrogen), respectively. Antibodies and sequences of oligonucleotides used for PCR with reverse transcription (RT-PCR), bisulphite sequencing and ChIP are listed in Supplementary Tables 1–4.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 February; accepted 17 November 2010.

Published online 30 January 2011.

- Gallinari, P. & Jiricny, J. A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature* **383**, 735–738 (1996).
- Um, S. *et al.* Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* **273**, 20728–20736 (1998).
- Chen, D. *et al.* T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *J. Biol. Chem.* **278**, 38586–38592 (2003).
- Tini, M. *et al.* Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol. Cell* **9**, 265–277 (2002).
- Li, Y. Q., Zhou, P. Z., Zheng, X. D., Walsh, C. P. & Xu, G. L. Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res.* **35**, 390–400 (2007).
- Gallais, R. *et al.* Deoxyribonucleic acid methyl transferases 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation. *Mol. Endocrinol.* **21**, 2085–2098 (2007).
- Zhu, B. *et al.* Overexpression of 5-methylcytosine DNA glycosylase in human embryonic kidney cells EcR293 demethylates the promoter of a hormone-regulated reporter gene. *Proc. Natl Acad. Sci. USA* **98**, 5031–5036 (2001).
- Metivier, R. *et al.* Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45–50 (2008).
- Kangaspeska, S. *et al.* Transient cyclical methylation of promoter DNA. *Nature* **452**, 112–115 (2008).
- Cortazar, D., Kunz, C., Saito, Y., Steinacher, R. & Schär, P. The enigmatic thymine DNA glycosylase. *DNA Repair (Amst.)* **6**, 489–504 (2007).
- Robertson, A. B., Klungland, A., Rognes, T. & Leiros, I. DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cell. Mol. Life Sci.* **66**, 981–993 (2009).
- Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
- Furuta, J. *et al.* Silencing of Peroxisome oxidin 2 and aberrant methylation of 33 CpG islands in putative promoter regions in human malignant melanomas. *Cancer Res.* **66**, 6080–6086 (2006).
- Cheng, Y. Y. *et al.* Frequent epigenetic inactivation of secreted frizzled-related protein 2 (SFRP2) by promoter methylation in human gastric cancer. *Br. J. Cancer* **97**, 895–901 (2007).
- Hardeland, U., Bentele, M., Jiricny, J. & Schär, P. Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.* **275**, 33449–33456 (2000).
- Bibel, M., Richter, J., Lacroix, E. & Barde, Y. A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols* **2**, 1034–1043 (2007).

17. Caldecott, K. W. XRCC1 and DNA strand break repair. *DNA Repair (Amst.)* **2**, 955–969 (2003).
18. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
19. Yu, B. D., Hanson, R. D., Hess, J. L., Horning, S. E. & Korsmeyer, S. J. MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proc. Natl Acad. Sci. USA* **95**, 10632–10636 (1998).
20. Erfurth, F. E. *et al.* MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. *Proc. Natl Acad. Sci. USA* **105**, 7517–7522 (2008).
21. Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.* **29**, 6074–6085 (2009).
22. Pasini, D. *et al.* Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* **38**, 4958–4969 (2010).
23. Zhu, B. *et al.* 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc. Natl Acad. Sci. USA* **97**, 5135–5139 (2000).
24. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytosine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279**, 52353–52360 (2004).
25. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201–1212 (2008).
26. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
27. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
28. Kunz, C. *et al.* Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol.* **7**, e91 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Klewe-Nebenius for preparations of mouse primary fibroblast, and F. Mohn and D. Schübeler for discussions and assistance in the setup and evaluation of the ChIP experiments. The work was supported by project grants from the Swiss National Science Foundation (3100AO-108436; 3100AO-122574/) and the Association For International Cancer Research (01-330).

Author Contributions D.C. established and performed the ChIP and MeDIP analyses and the *in vitro* differentiation experiments, and contributed to writing the paper; C.K. established and characterized MEF lines, designed and performed gene expression and DNA methylation analyses, and contributed to writing the paper; J.S. did blastocyst injections, established the first heterozygous *Tdg* knockout mice, characterized the lethal phenotype of the *Tdg* null embryos and provided SV40-immortalized MEFs; T.L. and Y.S. generated *Tdg*-targeting constructs and established heterozygous and homozygous *Tdg* knockout ES cell lines; Y.S. established *in vitro* differentiation protocols; E.MacD. performed the Big Blue mutation assays; A.W. performed ChIP experiments; D.S. isolated primary MEFs and performed RT-PCR validations of gene expression differences and the PARP inhibitor experiments; A.L.J. established and performed immunofluorescence experiments including XRCC1 foci analyses; F.S. performed bioinformatic analyses of gene expression array data; R.S. affinity-purified anti-TDG antibodies for ChIP; J.J. contributed genomic *Tdg* clones and supervised initial work of T.L.; A.B. was involved in study design (mutation analyses) and supervised the work of J.S. and E.MacD.; P.S. designed, coordinated and supervised the study, analysed the data and wrote the paper. All authors discussed the results and commented on the paper.

Author Information Gene expression array data have been deposited in the NCBI Gene Expression Omnibus under accession number GSE20693. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S. (primo.schaer@unibas.ch).

METHODS

Tdg knockout strategy. The *Tdg*-targeting construct (Supplementary Fig. 1) was generated by replacement of a *NarI*-*PacI* fragment enclosing exons 6 and 7 by a neomycin resistance cassette in a cloned fragment spanning exons 5–10 of the *Tdg* locus. This construct was used to target the *Tdg* allele in 129 mouse ESCs, which were then used to generate chimaeras and, ultimately, *Tdg*^{+/-} heterozygotes by backcrossing to C57BL/6. The generation and establishment of MEFs and *Tdg*^{-/-} ESCs was previously described²⁸.

Cell culture and ESC differentiation. SV40-immortalized MEF cell lines were previously described²⁹ and cultivated in growth medium (DMEM, 10% FCS, 2 mM L-glutamine) at 37 °C in a humidified atmosphere containing 5% CO₂. For growth of cell lines complemented with *Tdg*-expressing vectors, the growth medium was additionally supplemented with 1 μg ml⁻¹ puromycin.

For isolation of primary MEFs, 10.5 days post-coitum embryos were dissected, homogenized and cells dissociated in 0.05% trypsin-EDTA for 5 min before plating in modified ES cell medium without LIF (DMEM, 10% FCS seraplus, 1× non-essential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine and 50 μM β-mercaptoethanol, 1× penicillin/streptomycin) and cultivation for 10 days.

ESCs were grown in the presence of feeder cells at 37 °C in ES cell medium (ECM: DMEM, 15% heat-inactivated FCS, LIF (1,000 U ml⁻¹), 1× non-essential amino acids, 1 mM Na-pyruvate, 2 mM L-glutamine and 90 μM β-mercaptoethanol) in a humidified atmosphere containing 5% CO₂.

Before starting retinoic-acid-induced differentiation, ESCs were grown in the absence of feeder cells for two passages. For embryoid body formation during neuronal differentiation, 4 × 10⁶ *Tdg*^{+/-} or *Tdg*^{-/-} ESCs were plated into non-adherent bacterial dishes (Greiner Bio-one) in differentiation medium (ECM without LIF and with 10% FCS) and grown at 37 °C with a medium exchange after 2 days. After 4 days, 5 μM all-*trans* retinoic acid was added and cells were further incubated for 4 days with a medium exchange after 2 days. Embryoid bodies were washed twice with 1× PBS and dissociated with freshly prepared trypsin solution (0.05% TPCK-treated trypsin in 0.05% EDTA/1× PBS) at 37 °C for 3 min. Dissociated embryoid bodies were re-suspended in 10 ml differentiation medium and collected by centrifugation at 700g for 5 min at room temperature. The pellet was re-suspended in N2 medium (DMEM-F12 nutrient mixture 1:1, 1× N2 supplement) and the cell suspension filtered through a 40-μm nylon cell strainer (BD). Filtered cells were immediately plated onto poly-L-lysine and laminin-coated dishes at a density of 5 × 10⁶ cells per 60-mm dish or 1.5 × 10⁷ cells per 100-mm dish. The N2 medium was exchanged 2 and 24 h after plating, and cells were collected after 4 and 48 h for further analysis.

Retinoic-acid-induced differentiation of ESCs for time course, PARP inhibitor and immunofluorescence experiments was induced in ECM without LIF in the presence of 1 or 5 μM retinoic acid. The retinoic-acid-containing medium was exchanged every 24 h, and cells were collected at the indicated time points. For immunofluorescence experiments, 10⁵ ESCs were seeded onto gelatin-coated cover slips 1 day before differentiation. For the analysis of PARP inhibition on cell survival during differentiation, 10⁵ ESCs were seeded into gelatin-coated 12-well dishes, 1 day before the addition of 5 μM retinoic acid or further cultivation in ECM. After 24 h, increasing concentrations of the PARP inhibitor AG-014699 (a gift of SelleckChem) were added and cell numbers determined 24 h later with the CASYcell counter. The 50% lethal dose of the inhibitor and statistical differences between *Tdg*-proficient and -deficient cells were calculated on triplicate experiments by linear regression with 95% confidence intervals using GraphPad Prism software.

Microarray gene expression analysis. For the analysis of differential gene expression between *Tdg*^{+/-} and *Tdg*^{-/-} MEFs, total RNA was isolated from three independent cultures of each cell line using the RNeasy Mini Kit (Qiagen), cDNA synthesized from 13 μg RNA with the SuperScript double-Stranded cDNA Synthesis Kit (Invitrogen) followed by *in vitro* transcription reactions using the MEGA Script T7 Kit (Ambion) supplemented with 1.5 mM Bio-11-CTP and Bio-16-UTP (Enzo Life Sciences). cDNAs and cRNAs were purified using the GeneChip Sample Cleanup Module (Qiagen). cRNA (15 μg) was fragmented and hybridized to GeneChip Mouse Expression Arrays 430A (Affymetrix). Hybridized arrays were stained and washed according to the manufacturer's protocol and scanned with an Affymetrix Scanner 3000 7G. Scanned images were processed with Microarray Suite software and obtained 'cel'-files used for further data analysis.

For gene expression analysis of ESCs and *in vitro* differentiated neuronal progenitor cells, total RNA was extracted from independent triplicates using the Trizol reagent (Invitrogen). RNA was quantified using the Quant-iT RiboGreen RNA Assay (Invitrogen) and 500 ng of total RNA subjected to cDNA synthesis and subsequent *in vitro* transcription to biotinylated cRNA using the Illumina TotalPrep RNA Amplification Kit (Ambion, USA). cRNA (1.5 μg) was hybridized to MouseWG-6v2 slides (Illumina) according to the manufacturer's protocol. Bead arrays were washed and stained using FluoroLink Cy3 Streptavidin (GE

Healthcare). Fluorescent signals were imaged using the iScan system (Illumina). Scanner images files were processed to probe intensity files by the manufacturer's software and further processed with the genome studio software (Illumina) without normalization and background correction.

Affymetrix data and Illumina probe intensity data were either processed by robust multi-array average or variance stabilization transformation, respectively, using R/Bioconductor software and 'affy' or 'lumi' libraries, followed by quantile normalization. Significance of effects for probes (Illumina) or probe-sets (Affymetrix) was tested in R/Bioconductor ('limma' library) using a moderated *t*-test and the false discovery rate (=5%) method of Benjamini and Hochberg for multiple testing correction. No unspecific filter was applied and multiple probe-sets per gene or probe-sets with ambiguous genomic targets were retained.

Methylation analyses. Genomic DNA from MEFs, ESCs and neuronal progenitor cells was isolated with the QIAamp DNA Mini Kit (Qiagen). DNA (2 μg) was subjected to bisulphite conversion using the EZ DNA Methylation Kit (Zymo Research). Respective target regions were amplified from bisulphite-treated DNA with TrueStart Taq polymerase (New England Biolabs). For conventional bisulphite sequencing, *Hoxd13* or *Sfrp2* promoter regions were amplified from converted DNA and cloned into the *XhoI* and *BamHI* restriction sites of pBluescript SK- before sequencing of individual clones. For pyrosequencing, potential regions of hypermethylation were first identified by COBRA. Pyrosequencing primers (Supplementary Table 1) were designed using the PyroMark Assays Design software (version 2.0.1.15, Qiagen). Primer pairs included either one biotinylated primer or one primer containing a universal region. In the latter case, products were subjected to a second amplification using a biotinylated universal primer and Phusion Hot Start High-Fidelity DNA Polymerase (Finnzymes). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), quantified and 300–500 ng were used for pyrosequencing in a PyroMark Q24 (Qiagen). Reactions were analysed using PyroMark Q24 software (version 2.0.6, Qiagen). Significance of methylation differences between different *Tdg*-proficient and -deficient cell lines at individual CpG sites was evaluated by unpaired, two-tailed *t*-tests.

ChIP. To crosslink protein-bound DNA, MEFs, ESCs and neuronal progenitor cells were incubated in freshly prepared crosslinking solution (PBS pH 7.4, 1% formaldehyde) at room temperature. The reaction was quenched after 10 min by addition of glycine to a final concentration of 125 mM. After washing twice with cold PBS, cells were collected using a cell scraper and subsequent centrifugation at 600g and 4 °C. Nuclei were isolated by incubation in 200 μl of cold CHIP Buffer I (10 mM HEPES pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) for 5 min on ice followed by two incubations of 5 min on ice in 200 μl cold CHIP buffer II (10 mM HEPES pH 6.5, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl). Pelleted nuclei were lysed in 400 μl ChIP buffer III (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5% Triton X-100, 1% SDS, 1 mM PMSF) for 10 min on ice followed by sonication for 15 min (15 s on, 30 s off, power high) using a Bioruptor sonicator (Diagenode) to produce random chromatin fragments ranging from 300 to 1,000 base pairs. The solution was cleared by centrifugation at 14,000g and 4 °C for 10 min and the concentration of chromatin was estimated by absorbance at 260 nm. For ChIP of TDG, MLL and APE1 100–150 μg of chromatin were diluted ten times in ChIP dilution buffer I (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 1 mM PMSF). For histone ChIPs, 25–75 μg of chromatin were diluted in ChIP dilution Buffer II (16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 1.1% Triton X-100, 0.01% SDS, 1 mM PMSF). Diluted chromatin was pre-cleared at 4 °C for 1 h with 40 μl of a 50% slurry of magnetic Protein G beads (Invitrogen) preblocked with 1 mg ml⁻¹ BSA and 1 mg ml⁻¹ tRNA (TDG, XRCC1, APE1 and MLL-ChIPs) or salmon sperm single-stranded DNA (histone ChIPs). Pre-cleared chromatin was incubated with 2–5 μg of the respective antibody (Supplementary Table 2) overnight at 4 °C under slow rotation. Immuno-complexes were precipitated with 40 μl of a 50% slurry of blocked Protein G beads and further incubated at 4 °C for 2 h. Beads were then serially washed with 500 μl ChIP wash buffer I (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100), 500 μl ChIP wash buffer II (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100) and 500 μl ChIP wash buffer III (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% sodium deoxycholate, 1% NP-40). For TDG, APE1 and MLL ChIPs, beads were washed once with 500 μl ChIP wash buffer I and twice with 500 μl ChIP wash buffer II. After two additional washes with 500 μl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), bound complexes were eluted by two sequential incubations with 150 μl elution buffer (1% SDS, 0.1 M NaHCO₃) at 65 °C for 10 min. Crosslink reversal of eluates and respective input samples (1% of chromatin used for ChIP) was done in the presence of 200 mM NaCl at 65 °C for 4 h followed by proteinase K digestion (50 μg ml⁻¹) in the presence of 10 mM EDTA at 45 °C for 1 h. DNA was purified by phenol/chloroform extraction and Na-acetate/ethanol precipitation, and re-suspended in 10 mM Tris-HCl pH 8.0. qPCR analysis with target specific primers (Supplementary Table 3) was performed using Quantitect SYBR Green (Qiagen) with

a Rotor-Gene 3000 thermocycler (Qiagen). The significance of different ChIP efficiencies among Tdg-proficient and -deficient cell lines was evaluated from triplicate experiments by non-paired, two-tailed *t*-tests.

MeDIP. MeDIP assays were performed as described in ref. 30. Briefly, genomic DNA was prepared from 5×10^6 cells by incubation in lysis buffer (20 mM Tris-HCl pH 8.0, 4 mM EDTA, 20 mM NaCl, 1% SDS and 1 mg ml⁻¹ proteinase K) at 55 °C for 5 h and subsequent phenol/chloroform extraction and Na-acetate/ethanol precipitation. DNA pellets were re-suspended in TE containing 20 µg ml⁻¹ RNase. DNA was sonicated as described for ChIP followed by NaCl (400 mM)/EtOH precipitation in the presence of glycogen-carrier. Fragmented DNA (4 µg) in 450 µl TE was denatured at 95 °C for 10 min and immediately chilled on ice. After addition of 10× immunoprecipitation buffer (100 mM sodium phosphate pH 7.0, 1.4 M NaCl, 0.5% Triton X-100), the DNA was incubated with 10 µg of a monoclonal anti 5-methylcytidine antibody (clone 33D2, Eurogentec) at 4 °C for 2 h. Immuno-complexes were precipitated by the addition of 40 µl M-280 sheep anti mouse IgG antibody coupled Dynabeads (Invitrogen) and incubation at 4 °C for 2 h followed by three washes in 700 µl IP buffer. Bound material was treated with 250 µl proteinase K digestion buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5% SDS and 0.25 mg ml⁻¹ proteinase K) at 50 °C for 3 h. Immunoprecipitated methylated DNA was purified by phenol/chloroform extraction followed by Na-acetate/ethanol precipitation and re-suspended in TE. qPCR analysis of sonicated genomic input DNA and MeDIP DNA with target specific primers (Supplementary Table 3) was performed as described for ChIP, and significance of MeDIP efficiencies tested by non-paired, two-tailed *t*-tests.

Quantitative RT-PCR analyses. Total RNA (2–4 µg) extracted by RNeasy Mini Kit or by Trizol methods was reverse transcribed with the RevertAid H Minus M-MuLV Kit (Fermentas) according to the manufacturer's protocol. qPCR with target specific primers (Supplementary Table 4) was performed using Power SYBR Green Master Mix (Applied Biosystems) with a Rotor-Gene 3000 thermocycler. Conditions for each target were validated by standard and melting curve analyses. Target-specific amplifications were normalized to a GAPDH control and data of at least three independent experiments were analysed by unpaired, two-tailed *t*-tests. *Tdg* genotype-specific target gene expression in primary MEFs was analysed by the non-parametric Kruskal–Wallis test and *post hoc* Dunn's multiple comparison.

Western blot analyses. Whole-cell extracts were prepared by cell lysis in lysis buffer (50 mM Na-phosphate pH 8.0, 125 mM NaCl, 1% NP-40, 0.5 mM EDTA, 1 mM PMSF, 1 mM DTT, 1× complete protease inhibitor, 2× phosphatase inhibitor cocktail 1 and 2) on ice for 30 min and clarification by centrifugation (15 min, 20,000g, 4 °C). Chromatin extracts were isolated as described for ChIP assays. Soluble proteins (50 µg) were separated on 7% or 10% SDS-polyacrylamide gels and transferred to a nitrocellulose membrane (Millipore). Membranes were washed once with TBS-T (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Tween-20), blocked with blocking buffer (TBS-T, 5% dry milk) at room temperature for 1 h and incubated with the primary antibody at 33 °C (anti-mTDG) or room temperature (anti-DNMT1, anti-DNMT3a, anti-XRCC1, anti-APE1, anti-MLL, anti-β-actin) for 1 h in blocking buffer. Dilutions were 1:10,000 for the rabbit anti-mTDG, the mouse anti-β-actin and the anti-DNMT1 antibodies; 1:1,000 for the anti-DNMT3a and anti-XRCC1 antibodies; 1:500 for the anti-APE1 and anti-MLL antibodies. Washing steps after hybridization were once at 33 °C and twice at room temperature for 15 min for anti-mTDG, or three times at room temperature for 10 min for all other antibodies. Membranes were incubated with secondary HRP-conjugated antibodies diluted 1:5,000 in blocking buffer and at room temperature for 1 h. After three washing steps of 10 min at room temperature, detection of the signals was performed using the Immobilon Western Chemiluminescent HRP Substrate (Millipore).

Cytotoxicity assays. For measurement of γ-ray sensitivity, MEF single-cell suspensions at a cell density of 2×10^5 cells ml⁻¹ in PBS were irradiated with the

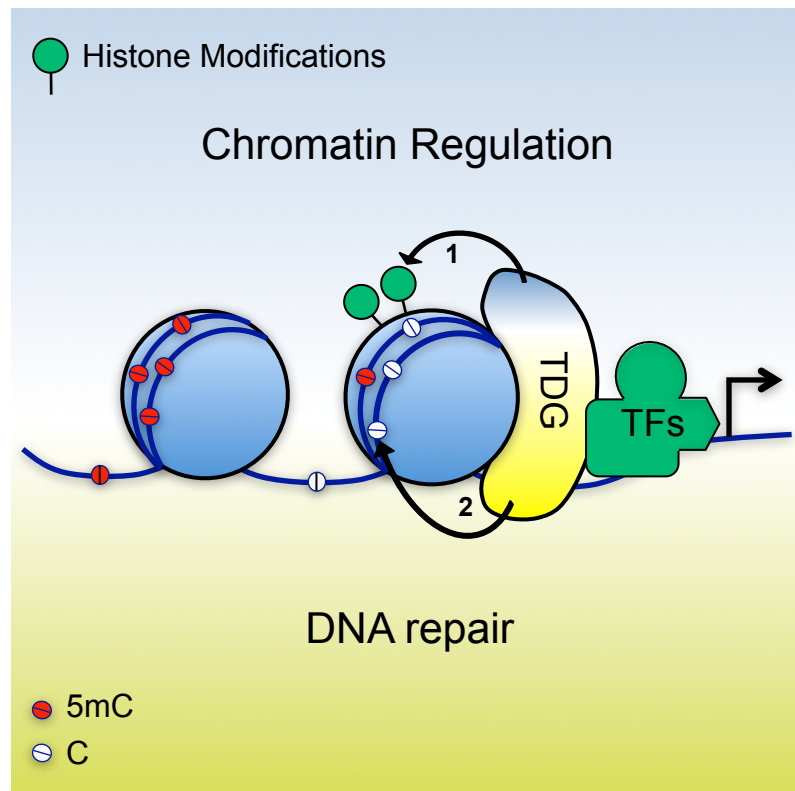
indicated doses in a Gammacell 40 irradiator using ¹³⁷Cs as a radioactive source. Irradiated cells were plated in 96-well microtitre plates at a density of 1000 cells per well in growth medium, and survival was measured after 3 days using the Cell Counting Kit-8 (Dojindo). Alternatively, survival was determined by clonogenic growth by plating 500–2000 cells in triplicate in 10-cm dishes containing growth medium and counting of Giemsa-stained colonies after 10 days. To measure sensitivity to H₂O₂, cells were plated at 2,500 cells per well in 96-well plates. After 24 h cells were treated for 15 min with the indicated concentrations of H₂O₂, washed with PBS and incubated in fresh growth medium for a further 24 h before measurement of survival with the Cell Counting Kit-8. Survival was determined as the percentage of mock-treated cells.

Base release assay. For base release assays, 25–50 µg of ESC whole-cell extracts were incubated with 0.5 pmol of a fluorescein-labelled GC/TG, GCm/CG or GCm/mCG DNA substrate in reaction buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT, 1 mg ml⁻¹ BSA) at 37 °C for 1 h (GC/TG) or overnight (methylated substrates). Resulting AP-sites were cleaved by the addition of NaOH to a final concentration of 100 mM and heating to 95 °C for 10 min. Subsequently, DNA was ethanol-precipitated overnight at –20 °C in the presence of 0.3 M Na-acetate pH 5.2 and 0.4 mg ml⁻¹ carrier t-RNA. DNA was collected by centrifugation (20 min, 20,000g, 4 °C) and washed with 80% ethanol. Air-dried pellets were re-suspended in loading buffer (1× TBE, 90% formamide), heated at 95 °C for 5 min and immediately chilled on ice. Reaction products were separated on denaturing 8 M urea/15% polyacrylamide gels in 1× TBE. The fluorescein-labelled DNA was visualized with a Typhoon 9400 and quantified using ImageQuant TL software (GE Healthcare).

Immunofluorescence. For detection of XRCC1 foci during retinoic acid stimulation, cells were fixed in ice-cold methanol for 5 min, then permeated in 0.2% Triton X-100/PBS pH 7.4 and 0.2% Triton X-100/0.2% NaBH₄/PBS pH 7.4 on ice for 5 min each. The inducibility of XRCC1 foci formation in ESCs was tested by incubation with H₂O₂ (50 µM in PBS) or PBS for 15 min at 37 °C and an additional 5 min in ECM with LIF before further processing. Coverslips were blocked in blocking buffer (1% BSA/0.05% Tween20/PBS pH 7.4), stained with rabbit anti-XRCC1 antibody (1:100 in blocking buffer) at room temperature for 1 h and washed three times for 10 min with blocking buffer before labelling with goat anti-rabbit Alexa Fluor 594 (1:200 in blocking buffer) for 30–60 min. After two washes of 10 min with blocking buffer, cells were again fixed in –20 °C cold methanol, incubated in blocking buffer for 1 h and stained with a mouse monoclonal anti-PCNA antibody (1:100 dilution) in blocking buffer overnight at 4 °C. Slides were counterstained for DNA with 50 ng ml⁻¹ DAPI and mounted in VectaShield mounting medium (Vector Lab). Slides were randomized and blinded before z-stacks were acquired on a Leica SP5 with the 405-nm diode, argon 488 nm and He–Ne 594-nm laser lines. XRCC1 foci numbers for individual cells were determined by visual inspection of the three-dimensional stacks. One hundred and fifty (retinoic acid stimulation) or 50 (H₂O₂) cells per sample were analysed. For co-staining of PAR and XRCC1 during retinoic acid differentiation, cells were fixed with 2% formaldehyde/PBS at room temperature for 30 min and permeabilized with PBS/0.2% Triton-X100 for 30 min. Antigen detection was done with a 1:250 diluted monoclonal α-PAR antibody 10H (Enzo Life Sciences) and a polyclonal α-XRCC1 as described above, but using 1:250 diluted anti-rabbit Alexa Fluor 594 and anti-mouse Alexa Fluor 488 secondary antibodies (Invitrogen). Pictures were acquired with a Nikon Diaphot 300 epifluorescence microscope using identical settings for all slides.

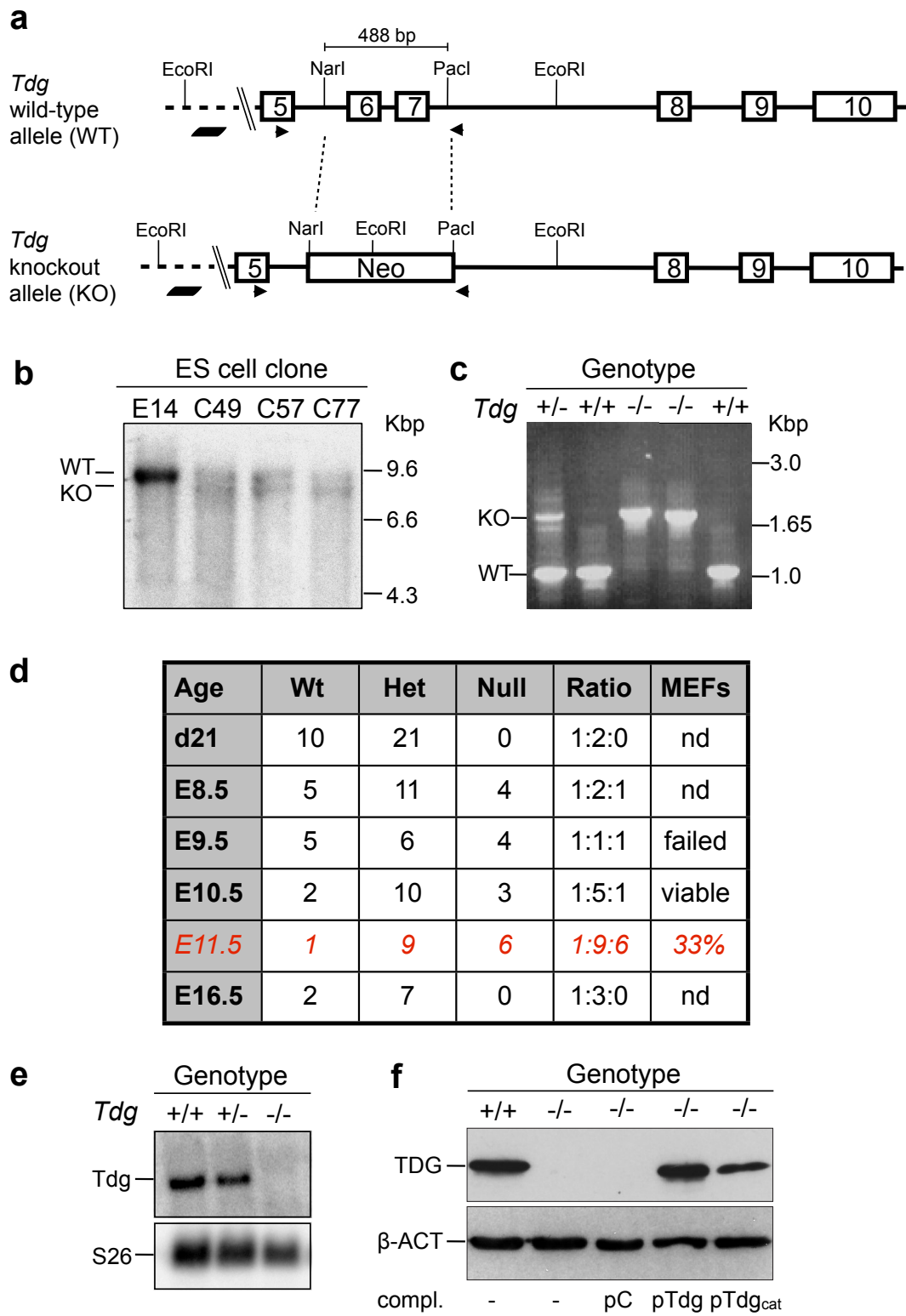
29. Kunz, C. *et al.* Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol.* **7**, e91 (2009).
30. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37**, 853–862 (2005).

Summarizing Figure



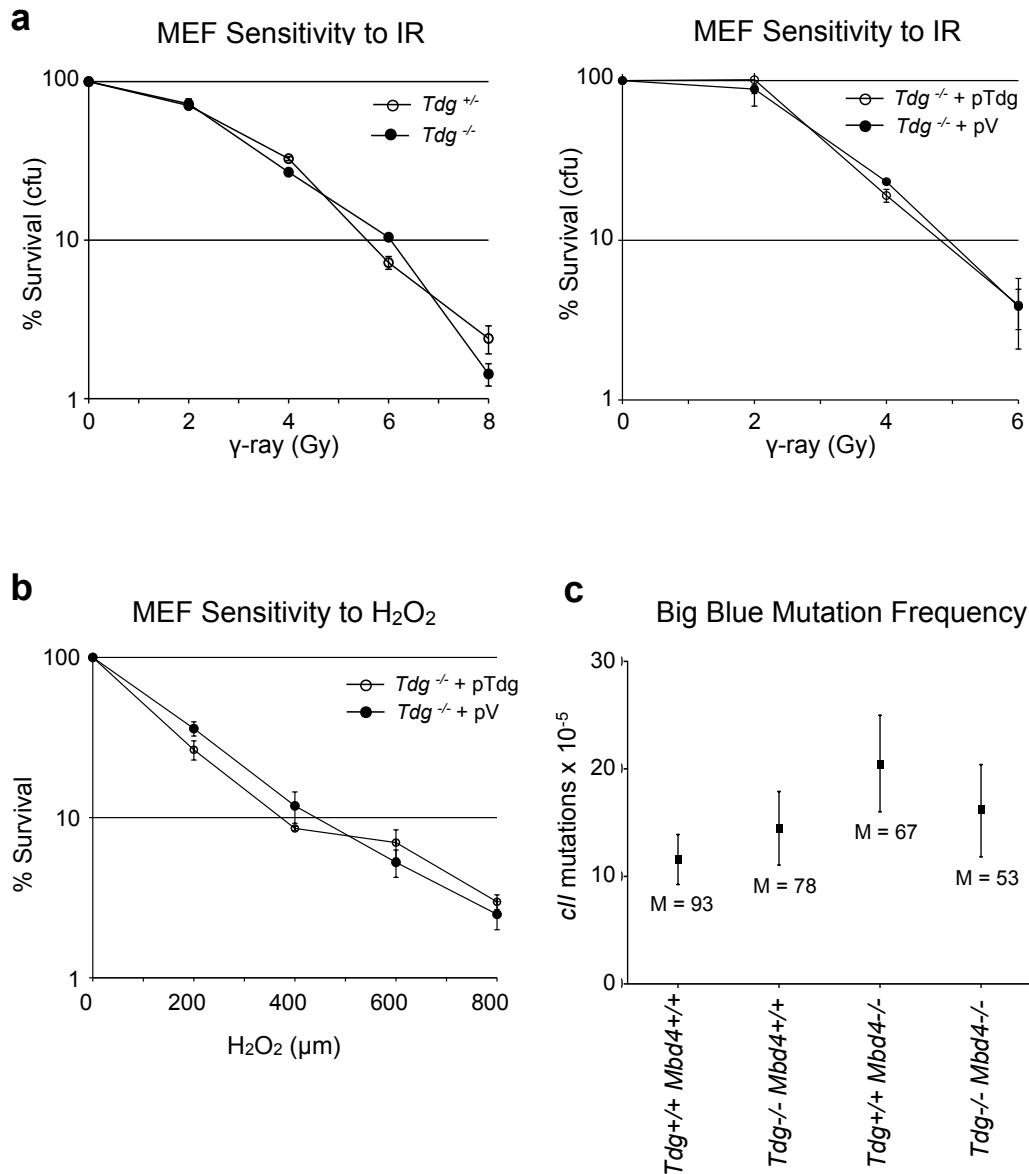
Summarizing Figure. The role of TDG in epigenetic control. TDG sustains proper (permissive) epigenetic states at gene promoters. As a structural component of transcription regulatory complexes, it contributes to the establishment and/or maintenance of accurate histone modification patterns (1), as a DNA repair enzyme, it corrects occasional aberrant de novo methylation of cytosine bases (2).

Supplementary Fig. 1



Supplementary Figure 1 | *Tdg* knockout strategy and validation. **a**, Schematic of the mouse *Tdg* locus representing exons 5-10. The insertion of the neomycin-resistance cassette to replace exons 6-7 is indicated, as well as the positions of probes used for Southern blotting (b) and primers for genotyping (c). **b**, Southern blot of *Eco*RI digested genomic DNA extracted from three E14 ESC clones (C49, C57, C77) with targeted *Tdg* locus. *Eco*RI digestion generated 9 kbp and 7.8 kbp DNA fragments for the wild-type and targeted *Tdg* alleles, respectively, here detected with a flanking probe external to the targeting construct as indicated in (a). **c**, PCR genotyping of *Tdg* knockout embryos. DNA was isolated from portions of embryos and analyzed by PCR using a primer pair amplifying both the targeted (1.7 kbp) and wild-type *Tdg* alleles (1.1 kbp). Shown are the PCR results of consecutive samples representing two *Tdg*^{+/+}, one *Tdg*^{+/-} and two *Tdg*^{-/-} genotypes. **d**, Pre-natal recovery of *Tdg*^{+/+}, *Tdg*^{+/-} and *Tdg*^{-/-} embryos after timed matings. Note that the *Tdg* null embryos isolated at E12.5 were all dead. **e**, Northern blot analysis of *Tdg* expression in MEFs isolated from *Tdg*^{+/+}, *Tdg*^{+/-}, and *Tdg*^{-/-} embryos. Blots were probed using a cDNA fragment spanning *Tdg* exons 8 to 10, amplified by RT-PCR. **e**, Western blot analysis of whole-cell protein extracts derived from SV40 immortalized *Tdg*^{+/+} and *Tdg*^{-/-} MEFs and *Tdg*^{-/-} complemented with wild-type (pTdg) and catalytically deficient (pTdg_{cat}) TDG or a vector control (pC). TDG was stained with a highly specific polyclonal anti-mouse TDG antibody (TDG) and staining for β-ACT served as loading control.

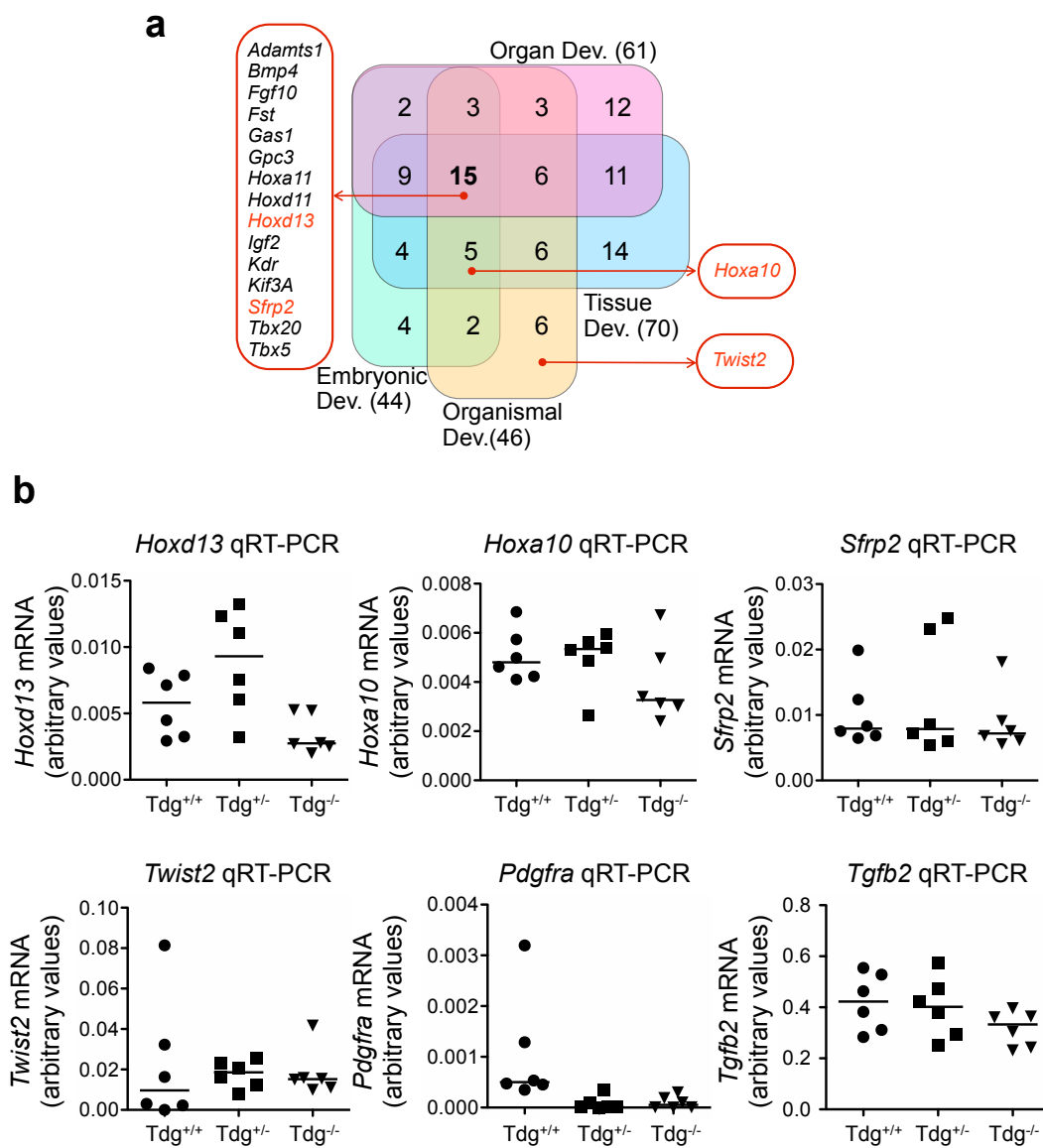
Supplementary Fig. 2



Supplementary Figure 2 | Lack of DNA repair associated phenotypes in TDG deficient cells. **a**, Sensitivities towards ionizing radiation (γ -ray) or hydrogen peroxide (H_2O_2) of $Tdg^{+/-}$, $Tdg^{-/-}$ or complemented $Tdg^{-/-}$ MEFs. Shown are survival curves as percentages of mock-treated cells (means \pm s.e.m., $n=3$). pV, vector control; pTdg, Tdg-expressing vector. **c**, *cII* mutation frequencies in *Tdg* and *Mbd4* single or double mutant MEFs. The *cII* mutant frequency is the ratio of *cII*⁻ plaques to the total number of λ

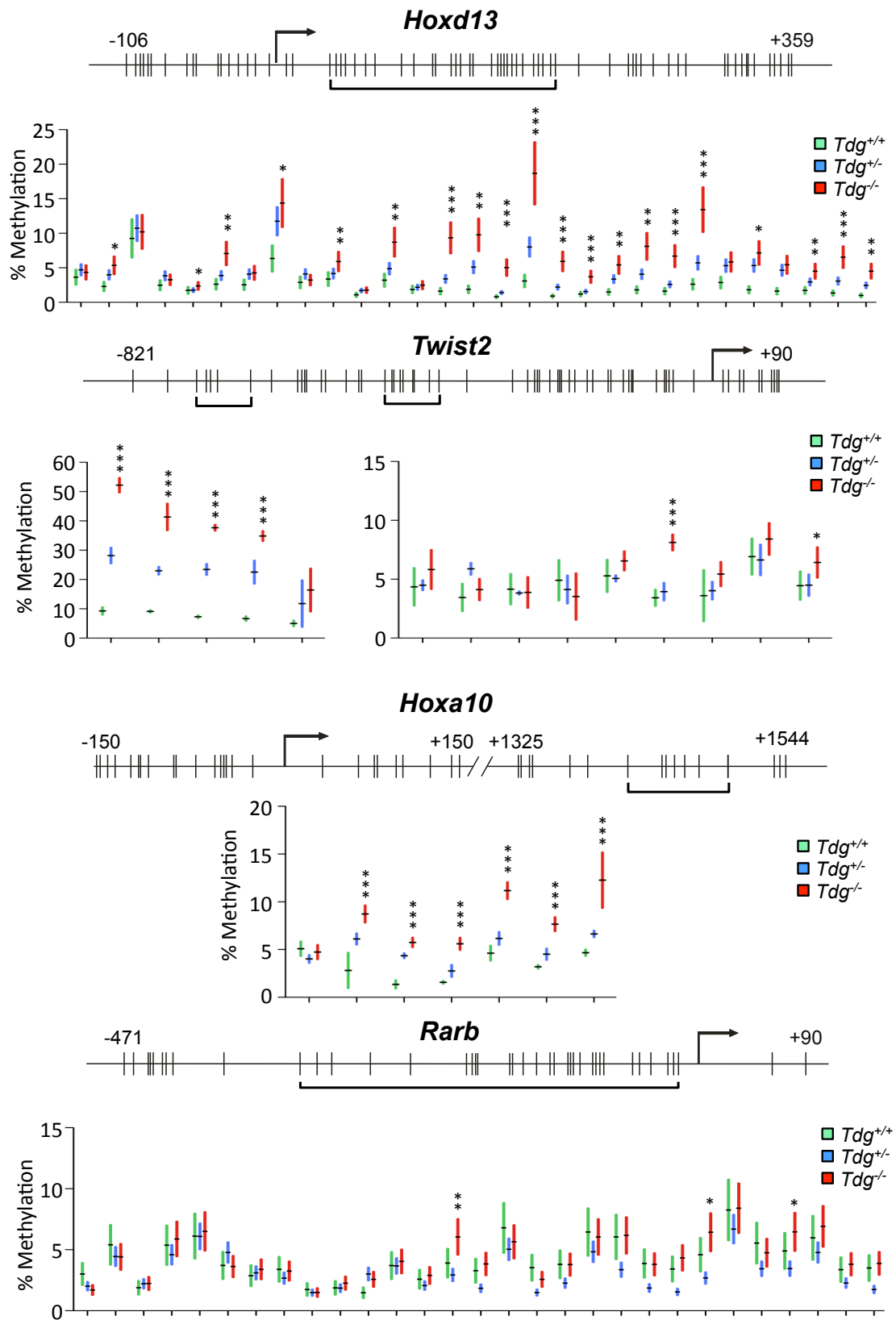
phage screened. Shown are mutation frequencies with 95% confidence intervals as calculated from the binominal proportions, with M indicating the actual number of mutant plaques scored for each genotype.

Supplementary Fig. 3



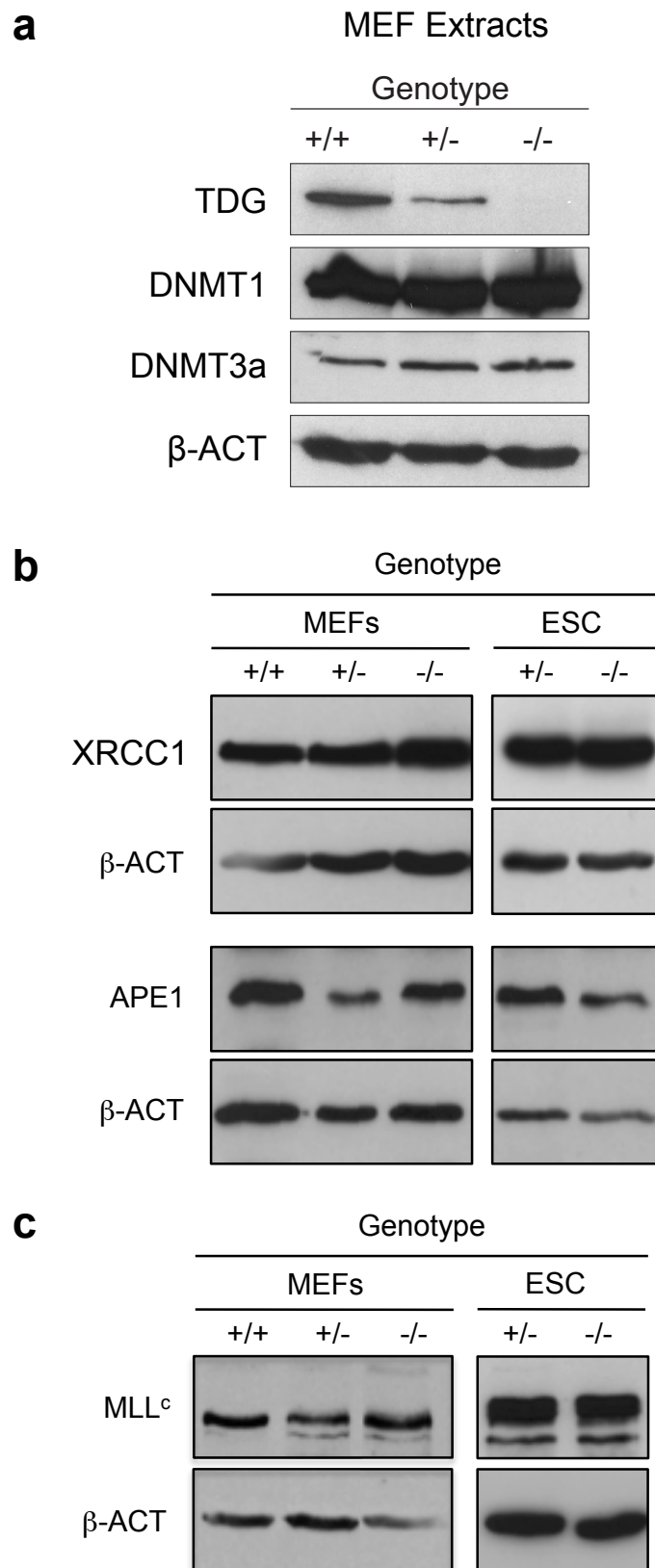
Supplementary Figure 3 | Gene ontology analysis and expression of selected targets in primary MEF isolates. **a**, Gene ontology (GO) annotations of the 200 most differentially regulated genes ($p < 0.05$) reveal a significant enrichment of developmental pathways (Ingenuity Pathway Analysis). **b**, Expression levels of selected genes in primary MEFs isolated from $Tdg^{+/+}$, $Tdg^{+/-}$, and $Tdg^{-/-}$ embryos at 10.5 dpc and cultured for 10 days. Gene expression was assessed by qRT-PCR, mRNA levels were normalized to *Gapdh* mRNA. Values represent arbitrary units with medians of six independent MEF isolates indicated by horizontal bars.

Supplementary Fig. 4

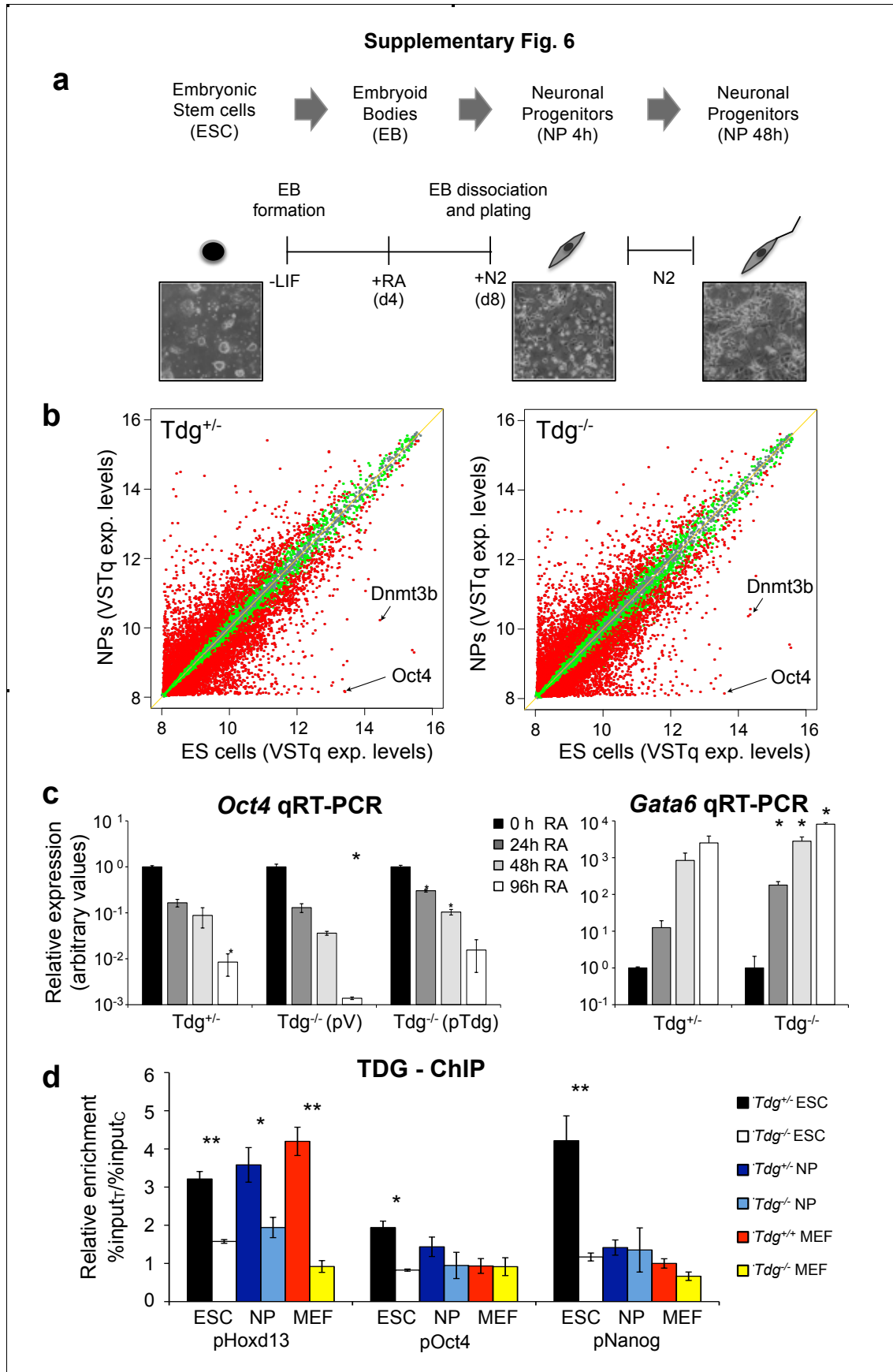


Supplementary Figure 4 | CpG methylation states of selected target promoters in MEFs. DNA methylation analysis by bisulfite pyrosequencing of *Hoxd13*, *Twist2*, *Hoxa10* and *Rarb* promoter regions in *Tdg*^{+/+}, *Tdg*^{+/-} and *Tdg*^{-/-} MEFs. Promoter regions are depicted schematically with vertical tick marks indicating CpG sites, bent arrows denoting transcription start sites, and horizontal brackets highlighting the CpGs for which methylation data is presented in the graphs below. Methylation levels are given as percentage of methylated cytosines at each CpG analyzed. Shown are means with 95% confidence intervals (bars) as obtained from at least 3 independent DNA isolations and bisulfite conversions for each genotype. *, p<0.05; **, p<0.01; ***, p<0.001; unpaired Student's t-test.

Supplementary Fig. 5



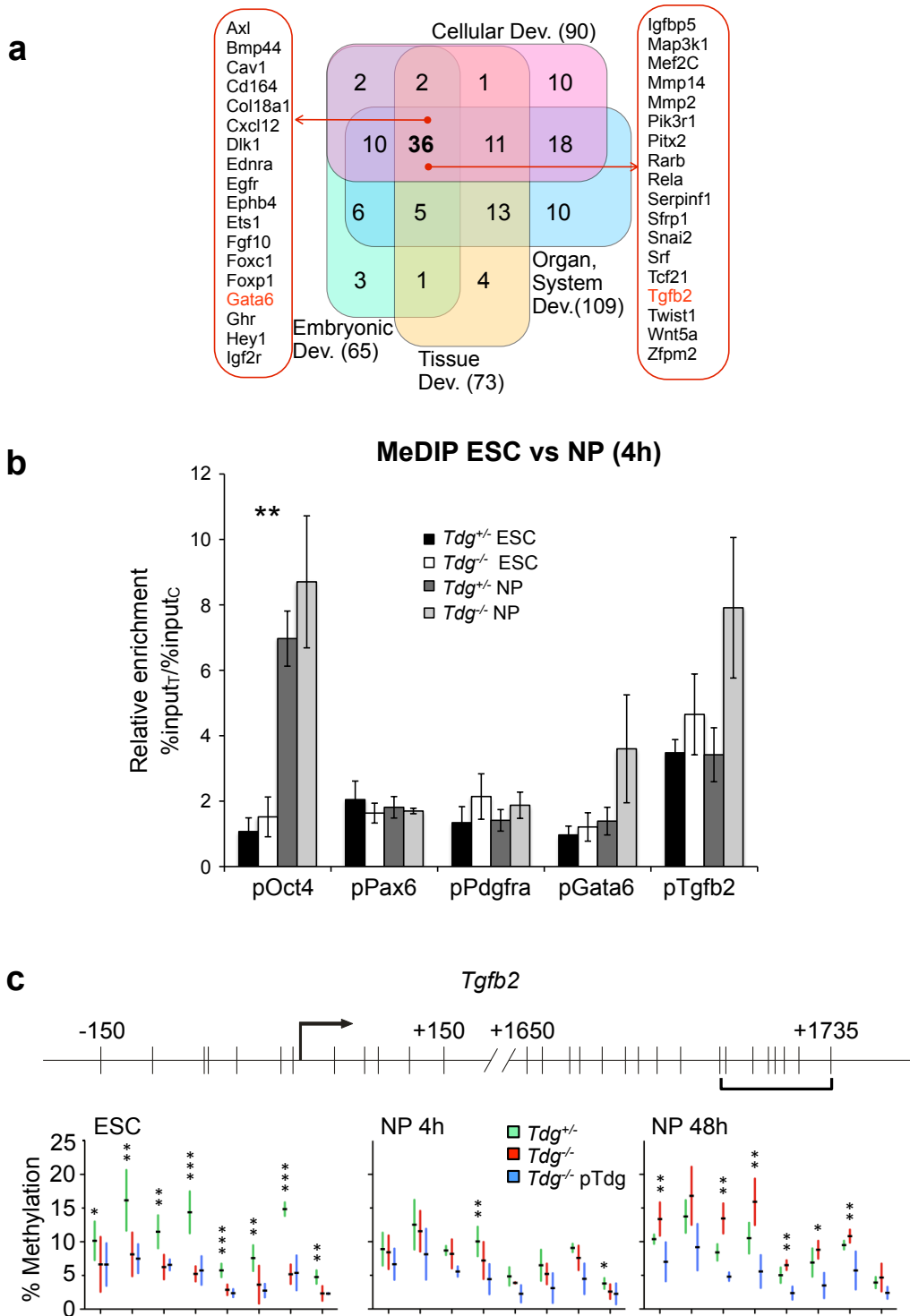
Supplementary Figure 5 | Validation of proteins levels and ChIP analysis of DNMT3a in TDG proficient and deficient MEFs. a, Western blots showing protein levels of TDG, DNMT1 and DNMT3a in whole cell extracts (WCE) of *Tdg*^{+/+}, *Tdg*^{+/-} and *Tdg*^{-/-} MEFs with β -ACT as loading control. 50 μ g of WCE were loaded in parallel on 10% (TDG, β -ACT) or 7% (TDG, DNMT1, DNMT3a) polyacrylamide gels and proteins detected with the respective antibodies after protein transfer. **b,** Western blots showing XRCC1, APE1 protein levels in 50 μ g chromatin extract of *Tdg*^{+/+}, *Tdg*^{+/-} and *Tdg*^{-/-} MEFs and *Tdg*^{+/-} and *Tdg*^{-/-} ESCs. β -ACT was used as loading control. **c,** Western blot showing MLL^c protein levels in 50 μ g of chromatin extracts of *Tdg*^{+/+}, *Tdg*^{+/-} and *Tdg*^{-/-} MEFs and *Tdg*^{+/-} and *Tdg*^{-/-} ESCs with β -ACT as loading control.



Supplementary Figure 6 | *In vitro* differentiation of ESCs to the neuronal lineage.

a, Schematic of the protocol used for *in vitro* differentiation of ESCs to NPs. ESCs were differentiated into embryoid bodies (EB) in the absence of LIF. EBs were treated with RA prior to dissociation and plating in N2 medium. ESCs and NPs at 4 or 48 hours after EB dissociation and plating were harvested for ChIP, DNA methylation and gene expression analyses. All differentiation experiments were done in biological triplicates. LIF, leukemia inhibitory factor; RA, all-trans retinoic acid. **b**, Scatter plots comparing gene expression before and after differentiation of *Tdg*^{+/-} or *Tdg*^{-/-} ESCs to NPs. Green (p<0.05) and red (p<0.01) dots represent differentially expressed genes. **c**, Validation of regulation of *Oct4* and *Gata6* expression following a time course of RA-induced cell differentiation. Shown are expression levels (qRT-PCR) relative to undifferentiated ESCs of the same genotype (mean±s.e.m., n=3, * p<0.05, unpaired Student's t-test). **d**, ChIP analysis of TDG association with the promoters of *Hoxd13*, *Oct4* and *Nanog* in chromatin of *Tdg*^{+/-} and *Tdg*^{-/-} ESCs, 48h NPs and MEFs. Shown are relative enrichments normalized to a random intergenic control region as determined by qPCR (mean±s.e.m., n=3; *, p<0.05; **, p<0.01; unpaired Student's t-test).

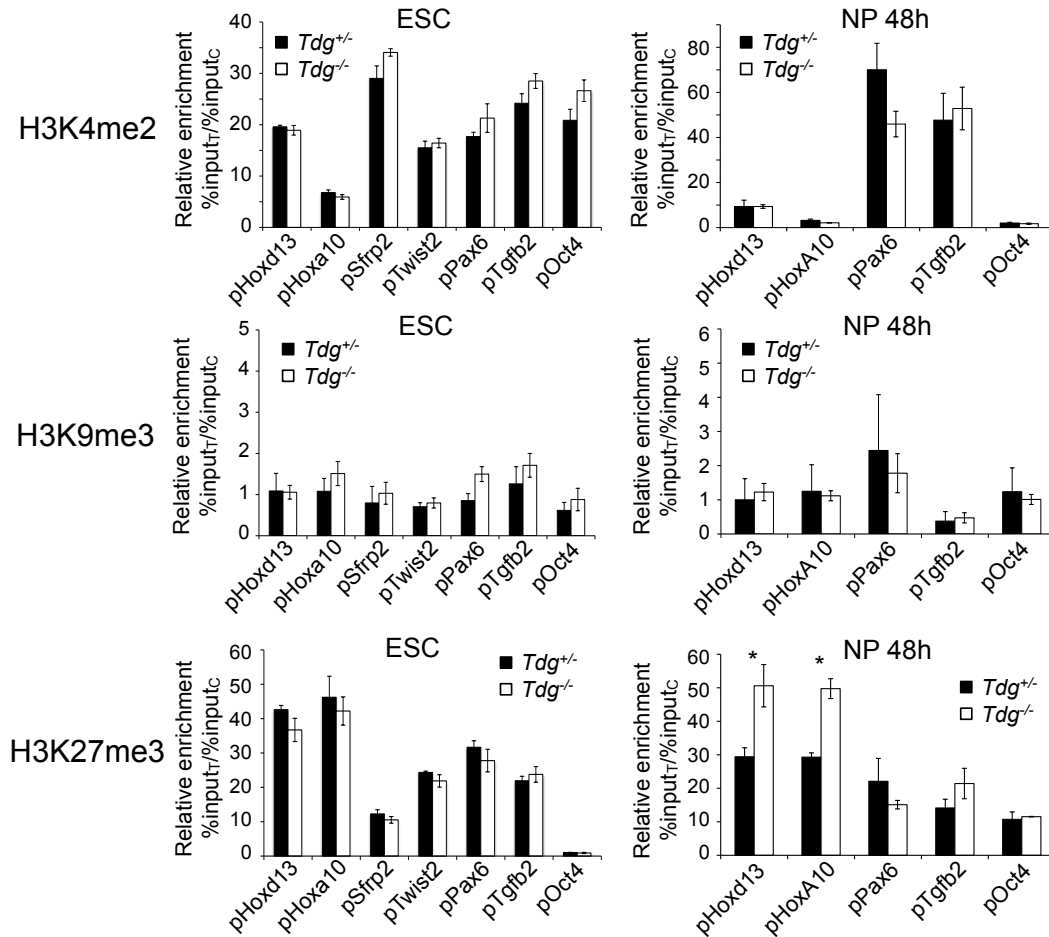
Supplementary Fig. 7



Supplementary Figure 7 | Gene ontology and DNA methylation analyses of TDG controlled genes during ESC - NP differentiation. **a**, Gene ontology (GO)

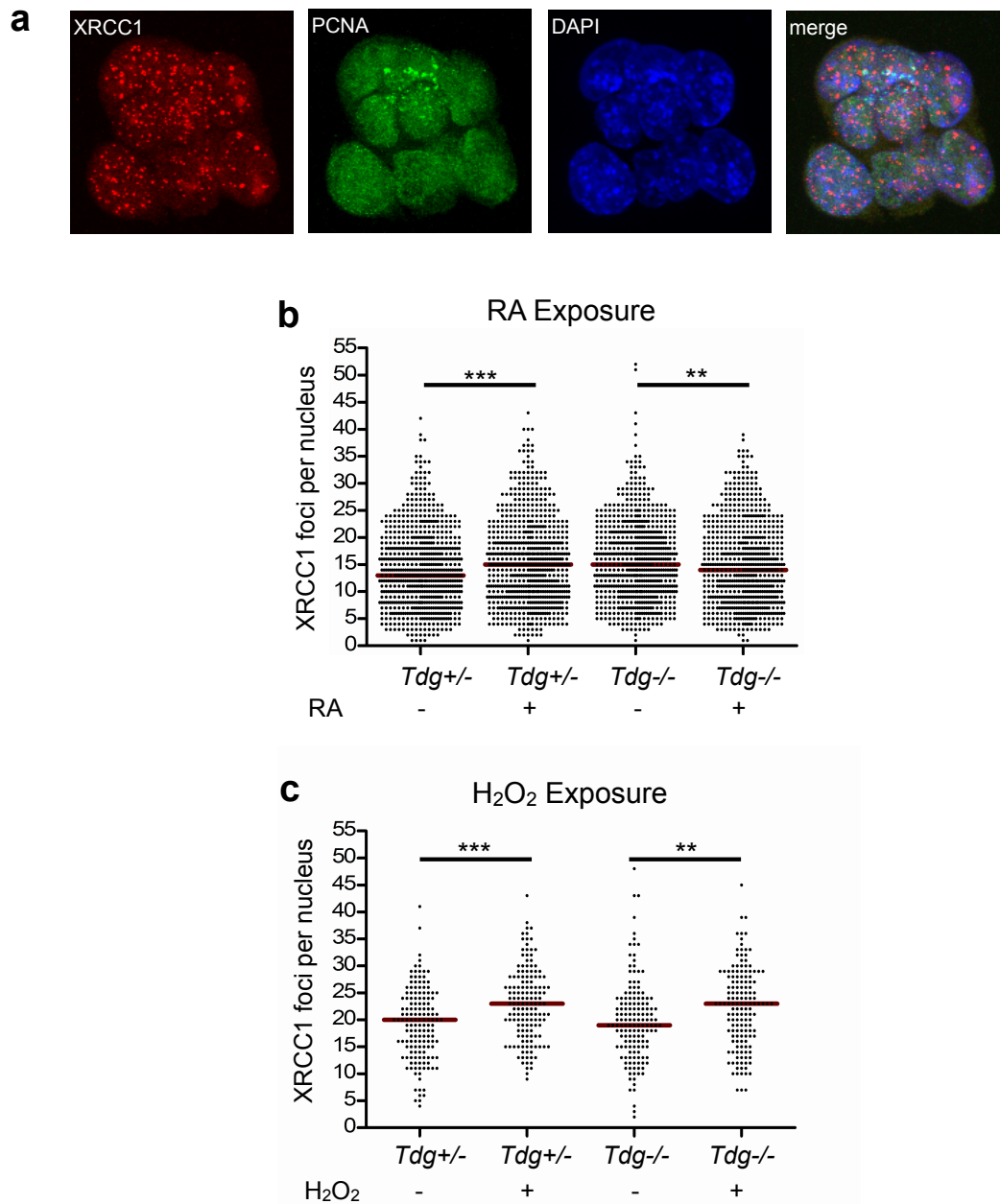
annotations of the 200 most differentially regulated genes (all $p < 0.05$) reveal a significant enrichment of developmental pathways (Ingenuity Pathway Analyses). **b**, The DNA methylation status at the *Oct4*, *Pax6*, *Pdgfra*, *Gata6* and *Tgfb2* promoters was analysed by MeDIP-qPCR in *Tdg*^{+/-} and *Tdg*^{-/-} ESCs and 4h NPs. The promoter region of *Gapdh* was used as internal normalizer (means \pm s.e.m., n=3, ** $p < 0.01$, unpaired Student's t-test), T, target region; C, control region. **c**, Bisulfite pyrosequencing analysis of CpG methylation in the *Tgfb2* promoter region in ESCs and NPs at 4 and 48h after plating of embryoid bodies in N2 medium. Promoter regions are depicted schematically with vertical tick marks indicating CpG sites, bent arrows denoting transcription start sites, and horizontal brackets highlighting the CpGs for which methylation data is presented in the graphs below. Methylation levels are given as percentage of methylated cytosines at each CpG analyzed. Shown are means with 95% confidence intervals (bars) as obtained from three differentiation experiments. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ (unpaired Student's t-test).

Supplementary Fig.8



Supplementary Figure 8 | Histone modification states in TDG deficient ESCs and NPs. ChIP-qPCR analyses performed on chromatin derived from *Tdg*^{+/-} and *Tdg*^{-/-} ESCs and NPs to assess the chromatin status at the TDG target promoters indicated. Data is expressed as relative enrichment normalized to *Iap* and the *Hprt* promoter for active and repressive chromatin marks, respectively (means±s.e.m., n=3; *, p<0.05; unpaired Student's t-test). T, target region; C, control region.

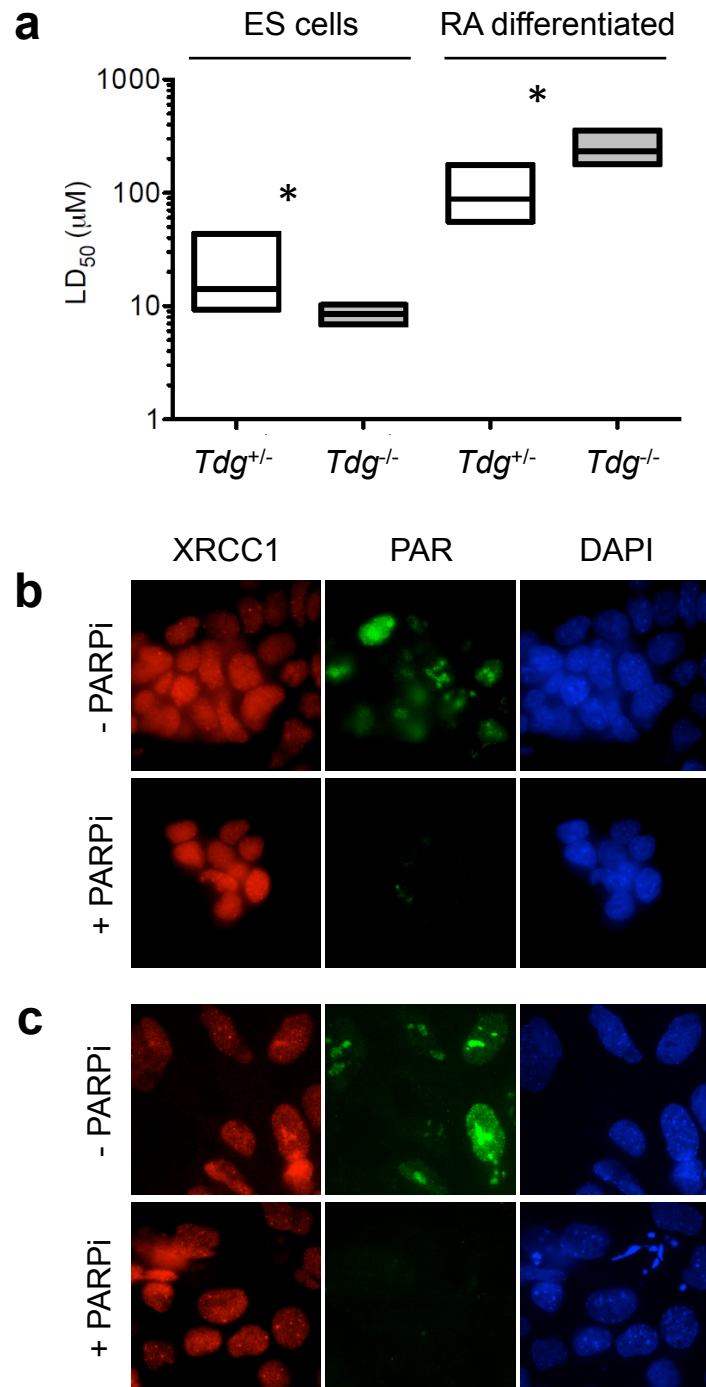
Supplementary Fig. 9



Supplementary Figure 9 | TDG dependent DNA repair activity upon RA induced ESC differentiation. Immunofluorescence staining of XRCC1 and PCNA in *Tdg*^{+/-} and *Tdg*^{-/-} ESCs before (RA-, LIF+) and after 8 hours induction of differentiation by 5 μ M retinoic acid (RA+, LIF-). **a**, Maximum intensity projections of confocal z-stacks for XRCC1 and PCNA immunofluorescence and for DNA counterstaining with DAPI.

PCNA staining was used as an indicator of S-phase cells to monitor and control for potential proliferation difference. **b**, Induction of XRCC1 foci following RA exposure. Shown are numbers of XRCC1 foci per cell as determined in 5 independent experiments. 150 cells per sample and experiment were analyzed for the number of XRCC1 foci. **c**, Positive control of damage dependent induction of XRCC1 foci. Shown are numbers of XRCC1 foci per cell after treatment with 50 μ M H₂O₂ in PBS (+) or PBS alone (-), as determined in 3 independent experiments. 50 cells per sample and experiment were analyzed. Note that the higher background of XRCC1 foci in the H₂O₂ experiments results from the prolonged incubation of the cells in PBS. Dots indicate individual cells, red lines the medians, and asterisks statistical significance determined by the Mann-Whitney-U-test (*, p<0.05; **, p<0.01; ***, p< 0.001).

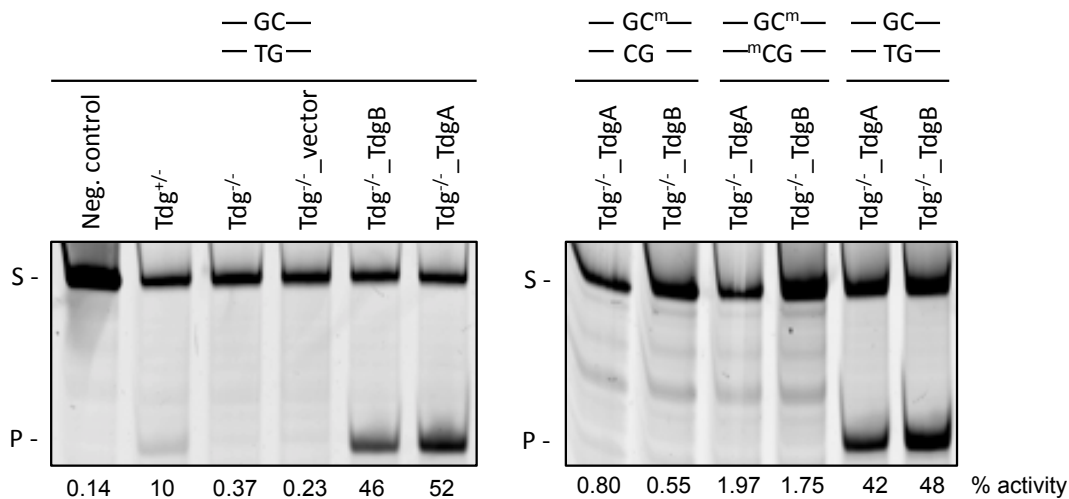
Supplementary Fig. 10



Supplementary Figure 10 | TDG sensitizes differentiating cells to the inhibition of PARP activity. **a**, ES cells were kept undifferentiated (+LIF, -RA) or differentiated (-LIF, +5 μ M RA) for 48 hours in the presence of increasing concentrations of the PARP

inhibitor (PARPi) AG-014699. Survival of *Tdg* proficient and deficient cells was measured and the LD50 determined by regression analysis (box, 95% confidence interval; line, LD₅₀; *, $p < 0.05$). Shown are representative epifluorescence images (100x magnification) of immunostainings for XRCC1 and poly(ADP)-ribose (PAR) in TDG deficient ES (b) and differentiating cells (c) treated without or with 10 μ M PARP inhibitor.

Supplementary Fig. 11



Supplementary Figure 11 | TDG has no 5-mC DNA glycosylase activity on its own.

Base release assays with whole cell extracts from *Tdg*^{+/-}, *Tdg*^{-/-} and *Tdg*^{-/-} ESC expressing either TdgA, TdgB or harbouring the vector only. Synthetic 60-mer DNA duplexes containing either a GC/TG mispair (left panel), or hemi- (GC^m/CG) or fully methylated (GC^m/C^mG) CpGs (right panel) were incubated with 25 μ g and 50 μ g of cell extracts at 37°C for 1 hour or overnight, respectively. Shown is a representative denaturing polyacrylamide gel showing the intact substrate (S) and cleaved product (P) at the top and bottom respectively with numbers at the bottom of the lanes representing the amounts of cleaved substrate (%). Neg. control = no extract.

Supplementary Table 1: Pyrosequencing primers

Primer	5'-3' Sequence
HoxA10 F	GAGGGGTAGGGAGGAAAAGTGGT
HoxA10 R	b-AACCATTCTAAATTTCAACTCTAAACCCA
HoxA10 S	TTTGTAAGGTATTTAAAATAAGTAG
HoxD13 F	GGGTTATGAGTAGTTAGGGGATTTGGGATATGGATGG
HoxD13 R	<u>GTCAGTCCAGTCCAGGTCAGGGTGAAGTATAGTATAGAGGTTGAG</u> GTTGAATTTTAAAT
HoxD13 S1	GGGGATTTGGGATATG
HoxD13 S2	GTAGTAGAGTTTGGTTAG
Pax6 F	GAGTGGGGTGGGGGGAAAAT
Pax6 R	b-TTCACCCTAACTTCCCACCCCTTATCC
Pax6 S1	GGGAAAATGGGTAGG
Pax6 S2	GGTTTAGGTATAGTTGTGTTA
Rar β F	GTTAGATTGGTTGGGTTATTTGAAGGTTAG
Rar β R	<u>GTCAGTCCAGTCCAGGTCAGGATCTTTTTCCCAACCCCAATCATA</u> AATTATAACAA
Rar β S1	GGGTTATTTGAAGGTTAGTA
Rar β S2	GTTTGGAAGGGAGAAT
Rar β S3	GATTGGGATGTAGAGG
Rar β S4	GGGGGGATTAGAGTTT
Tgf β 2 F	TAATAGTATTAGGGATTTATTGTAGGAGAAGGTAAG
Tgf β 2 R	b-AATTTACAAACCTATAAATCCCTCTCCATC
Tgf β 2 S	GGGATTTATTGTAGGAGAAG
Twist2 F	<u>GTCAGTCCAGTCCAGGTCAGGGTTGTGATGTTTAAGTTATAAAGTAT</u> TTAGGGGGTAG
Twist2 R	TCTCCTAAAACAAATTTAACCCCTACCAAATTC
Twist2 S1	TTTCTAAACTACTTCAACCTA
Twist2 S2	CCAAACCCAAATATACTC
Unique	b- <u>GTCAGTCCAGTCCAGGTCAGG</u>

b-, biotinylated primer; F, forward primer; R, reverse primer; S, sequencing primer; underlined sequence, universal primer

Supplementary Table 2: Antibodies

Antibody	Product Nr.	Manufacturer
Anti-H3K4me2	07-030	Millipore, USA
Anti-H3K9me3	pAb-056-050	Diagenode, UK
Anti-H3K27me3	07-449	Millipore, USA
Anti-MLL ^c	05-765	Millipore, USA
Anti-Ref-1 (APE1; C-20)	sc-334	Santa Cruz Biotechnology, Inc., USA
Anti-Dnmt3a (H-295)	sc-20703	Santa Cruz Biotechnology, Inc., USA
Anti-CBP (A-22)	Sc-369	Santa Cruz Biotechnology, Inc., USA
Anti-Dnmt3a	ab2850	Abcam, UK
Anti-Dnmt1	ab5208	Abcam, UK
Anti-beta Actin	ab8226	Abcam, UK
Anti-5-MeCyd (33D2)	BI-MECY-0100	Eurogentec, Belgium
Anti-XRCC1	X0629	Sigma-Aldrich, USA
Anti-PAR (10H)	ALX-804-220	Enzo Life Sciences
Anti-PCNA-Fluorescein	P105	Leinco Technologies, USA
Anti-rabbit-HRP	NA934	GE Healthcare, USA
Anti-mouse-HRP	NXA931	GE Healthcare, USA
Anti-rabbit-Alexa594	A-11012	Invitrogen, USA
Anti-mouse-Alexa488	A-11017	Invitrogen, USA

Supplementary Table 3: ChIP and MeDIP primers

Primer	5'-3' Sequence
pHoxD13F	TGGGCTATGGCTACCACTTC
pHoxD13R	GACACTTCCTTGGCTCTTGC
pHoxA10F	CACTCCCAGTTTGGTTTCGT
pHoxA10R	GGGGGTACAGGTTCAAGAGC
pSfrp2F	GACTTTCGTTGCCTCCTCCT

pSfrp2R	AGGCCGGTCACTACTTTCTG
pTwist2F	TCGCTGTGATGCCTAAG
pTwist2R	CACGATCTCGCCTCTAGGAT
pRar β F	GGGAGTTTTTAAGCGCTGTG
pRar β R	CGGAGCAGCTCACTTCCTAC
pTgf β 2F	AAGGGACGAGACGAGAAGGT
pTgf β 2R	ACATCCACACGCACACTCAT
pPax6F	CGGTGAAAGAAGCCACTAGG
pPax6R	TAGGGCGTTTGTTCCTCAAT
pOct4F	GTGAGGTGTCGGTGACCCAAGGCAG
pOct4R	GGCGAGCGCTATCTGCCTGTGTC
pGata6F	AGTTTTCCGGCAGAGCAGTA
pGata6R	AGGAGGAAACAACCGAACCT
pDnm1F	ATTCGCGGACTGGTCACTAT
pDnm1R	TTAGCACCCCTAGCCATCAC
pPdgfraF	GGACGAGCGATCTGGAATAA
pPdgfraR	CCGTGCAGAAAAGACTCCAC
pFgfr2F	CTCCAGAATCCAAGGACCA
pFgfr2R	CATCCAATGCTGACATCTG
lapF	CTCCATGTGCTCTGCCTTCC
lapR	CCCCGTCCCTTTTTTAGGAGA
pHprtF	CCAAGACGACCGCATGAGAG
pHprtR	CAACGGAGTGATTGCGCATT
Chr2negF	AGCACAGCCTGAAGCCTCTA
Chr2negR	AGAGGGCATTTCGGTCTTTT

Supplementary Table 4: qRT-PCR primers

Primer Name	5'-3' Sequence
GapDH(U)	TGCACCACCAACTGCTTA
GapDH(R)	GGATGCAGGGATGTTTC
HoxA10a RT F	CTCCCTGGGCAGTTCCAAAG
HoxA10a RT R	CGCTACGGCTGATCTCTAGG
HoxD13 RT1 F	CGACATGGTGTCCACTTTTG
HoxD13 RT1 R	TGGTGTAAGGCACCCTTTC
RT Sfrp2 fw3	GCCGGCCACAGAGGAAGCTC
RT Sfrp2 rev3	GGTCCCTTTCGGACACGCCG
Twist2 RT F	CGTCTCAGCTACGCCTTCTC
Twist2 RT R	CTGAGATGTGCAGGTGGGTC
Rar-b RT F	TTAATCTGTGGAGACCGCCAG
Rar-b RT R	TTACACGTTCCGGCACCTTTCG
Pdgfra RT F	CGAGGTCGTTGACCTGCAGTGG
Pdgfra RT R	CGACGAAGCCTTCTCGTGGACC
Tgfb2 RT F	AGAATCGTCCGCTTTGATGT
Tgfb2 RT R	GCTGGGTGGGAGATGTTACG
Oct3/4(U)	GGCGTTCTCTTTGGAAAGGTGTTTC
Oct3/4(R)	CTCGAACCACATCCTTCTCT
Gata6 RT F	TCCATGGGGTGCCTCGACCA
Gata6 RT R	ACCCCTGAGGTGGTCGCTTGT

Born on 4th November 1984 in Basel

Education

- 10/2009 – 07/2014 **Doctoral Thesis in Molecular Genetics at the University of Basel in Prof. Primo Schär's laboratory**
Focus: Epigenetics, carcinogenesis, cell biology
- *Managing and guiding different international research projects*
 - *Performing state-of-the-art laboratory work, e.g. ChIP-sequencing*
 - *Performing data analysis and basic bioinformatical analysis*
 - *Presenting my work in front of different audiences (in English)*
 - *Instructing medical students in practical courses and tutorials*
- Grade: "summa cum laude"*
- 04/2009 – 09/2009 **Internship (100%) in biomedical laboratory**
Prof. Primo Schär (Molecular Genetics), University of Basel
- 02/2008 – 02/2009 **Master Thesis** at the Department of Environmental Sciences, Section of Conservation Biology, University of Basel
Grade: 5.0 (of 6.0; good)
- 08/2007 – 01/2008 **Exchange semester in Sweden**
Studies with the Erasmus program at the University of Lund
- 10/2004 – 07/2007 **Bachelor of Science in Biology**, University of Basel
Grade: 4.9 (of 6.0; good)
- 08/2000 – 12/2003 **Matura** with focus on biology and chemistry (Schwerpunkt B) at the Gymnasium Münchenstein
Grade: 4.6 (of 6.0; good)

Languages

German	native speaker
English	proficient, oral and written (official working language)
French	good knowledge, oral and written
Swedish	good knowledge, oral and written
Spanish	good knowledge, oral and written

IT skills

Office applications	Proficient in Word, Excel, PowerPoint
Statistics	Proficient in Prism GraphPad and basic knowledge of R
Bioinformatics	Genome Browser, Ingenuity Pathway Analysis (IPA)

Publications in peer-reviewed journals

Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability

D. Cortázar*, C. Kunz*, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, **A. Wirz**,
D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird & P. Schär.
Nature 470, 419-423, 2011. * contributed equally

TDG balances DNA methylation and oxidative demethylation at CpG islands in differentiating cells

A. L. Jacobs*, D. Cortázar*, **A. Wirz***, J. Arand, R. Steinacher, C. Broberg Vågbø, P. Giehr,
A. Weber, G. Wilson, A. Galashevskaya, C. Kunz, W. Reik, S. Beck, J. Walter, H. Krokan &
P. Schär.
(Manuscript in preparation, 2014) * contributed equally

TET1, TET2 and TDG cooperate locus-specifically to promote chromatin plasticity by oxidative DNA demethylation

A. Wirz, F. Noreen, R. Ivánek, Z. Barekati, A. L. Jacobs, C. Kunz & P. Schär.
(Manuscript in preparation, 2014)

Attended Conferences and Meetings

- Talks**
- 1st PhD Retreat of the Department of Biomedicine in Saas Grund (CH), 2012
 - DNA Replication, Repair, Recombination Meeting in Zürich (CH), 2013
- Chair**
- 2nd PhD Retreat of the Department of Biomedicine in Hasliberg (CH), 2013
- Posters**
- EMBO conference on Chromatin and Epigenetics in Heidelberg (D) 2011
 - Swiss Meeting on Genome Stability in Weggis (CH) 2012
 - 3rd Erling Seeberg Conference on DNA Repair in Trondheim (N) 2012

References are available on demand.