

A Function of Thymine DNA Glycosylase- Initiated DNA Repair in Maintaining Epigenome Stability

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Angelika Jacobs
aus
Marburg, Deutschland

Basel, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Primo Schär (Fakultätsverantwortlicher und Dissertationsleiter)

Prof. Dr. Wolf Reik (Korreferent)

Basel, den 26.3.2013

Prof. Dr. Jörg Schibler

Dekan der Philosophisch-Naturwissenschaftlichen Fakultät

Acknowledgements

I would like to thank Primo Schär for supervising my thesis, and my PhD committee, Wolf Reik and Manuel Stucki for their critical evaluation of my work.

Moreover, I thank Primo for his contagious enthusiasm, but also for taking the time to steer his lab members through the ups and downs of their projects. My thanks also go to David Schürmann for his ongoing support, readiness to help and critical reading of parts of this thesis.

I would like to thank all past and present members of the group Schär for maintaining a really enjoyable work atmosphere. My heartfelt gratitude goes to Claudia Krawczyk, Annika Wirz, Anja Nusser and Katrin Martin for their friendship and support.

I thank my family, Monika and Matthias, who have encouraged and supported me during my whole studies. I especially thank my brother Matthias for triggering my scientific curiosity at young age.

Last but not least, I wish to thank Dimos for standing by me during the most challenging times, for late-night last-minute reading of this thesis, for his advice, his faith in me and for his love.

*“We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.”*

-- T.S. Eliot --

Table of Contents

1 Summary.....	5
2 Introduction.....	8
2.1. DNA Repair and the Concept of Genome Maintenance	8
2.1.1. Sources of DNA Damage and Modification	8
2.1.2. Molecular Mechanisms to Repair DNA Damage	10
2.1.3. Base Excision Repair and Single-strand Break Repair	13
2.2. The Thymine DNA Glycosylase (TDG).....	15
2.2.1. Classification and Characterization of TDG	15
2.2.2. Mechanism of Substrate Recognition and Processing	16
2.2.3. Biological Functions of TDG.....	17
2.2.4. Regulation of TDG Function by Post-Translational Modification	21
2.3. Epigenetic Regulation of Gene Expression	25
2.3.1. Chromatin Structure and Dynamics	25
2.3.2. Histone Modifications	26
2.3.3. DNA methylation	29
2.3.4. Epigenome dynamics in mammalian development	33
2.4. DNA demethylation.....	34
2.4.1. Concepts of DNA demethylation.....	34
2.4.2. Current models for BER-mediated active DNA demethylation.....	36
3 Aims of the Thesis	38
4 Results	39
4.1. Measuring the SUMO-Interaction Dynamics of the Thymine DNA Glycosylase by Fluorescence Resonance Energy Transfer (Appendix I)	39
4.2. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability (Appendix II)	42

4.3. TDG maintains a transitory equilibrium of CpG island methylation and oxidative demethylation during cell differentiation (Appendix III)	45
4.4. Supplementary Results.....	49
4.4.1. A cell-based assay to measure the epigenetic stability of gene promoters.....	49
4.4.2. Generation of 5-caC through inducing oxidative damage in live cells	56
4.4.3. Familial loss-of-function TET2 mutation does not correlate with the disease phenotype .	59
4.5. DNA Glycosylases: In DNA Repair and Beyond (Appendix IV).....	63
5 Concluding Discussion and Outlook	64
6 References.....	68

Appendix:

- I. Measuring the SUMO-Interaction Dynamics of the Thymine DNA Glycosylase by Fluorescence Resonance Energy Transfer
- II. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability
- III. TDG maintains a transitory equilibrium of CpG island methylation and oxidative demethylation during cell differentiation
- IV. DNA Glycosylases: In DNA Repair and Beyond

Curriculum vitae

Abbreviations

5-caC	5-carboxylcytosine
5-fC	5-formylcytosine
5-hmC	5-hydroxymethylcytosine
5-hmU	5-hydroxymethyluracil
5-mC	5-methylcytosine
A	adenine
AP	apurinic/apyrimidinic
APE1	AP endonuclease 1
BER	Base Excision Repair
BS	bisulphite
C	cytosine
caCDIP	5-caC DNA IP
Ceru	Cerulean
CGI	CpG island
ChIP	Chromatin IP
Cit	Citrine
DNA	Deoxyribonucleicacid
DNMT	DNA methyltransferase
ES cell	embryonic stem cell
FLPo	optimized flippase
FRET	Fluorescence Resonance Energy Transfer
FRT	FLP recognition target
G	guanine
IP	Immunoprecipitation
lacI	lac inhibitor
lacO	lac operator
LCMSMS	liquid chromatography – tandem MS
MeDIP	Methylated DNA IP
MEF	mouse embryonic fibroblast
MS	mass spectrometry
NER	Nucleotide Excision Repair
NP	neuronal progenitor

pIProX	plasmid lacO Promoter eXchange
PML-NB	promyelocytic leukemia protein nuclear bodies
Pol	polymerase
RA	retinoic acid
SBM	SUMO Binding Motif
seq	next generation sequencing
SUMO	Small Ubiquitin-like Modifier
T	thymine
TDG	Thymine DNA Glycosylase
TET	Ten Eleven Translocator
TSS	Transcription Start Site
wt	wildtype
XRCC1	X-ray repair complementing defective repair in chinese hamster cells 1
Δcat	catalytically dead mutant, TDG N151A
ΔS	SUMOylation deficient mutant, TDG K341R

1 Summary

The Thymine DNA Glycosylase (TDG) was initially discovered by its ability to excise the deamination products of cytosine and 5-methylcytosine (5-mC), and therefore thought to initiate base excision repair (BER) of the resulting G•U and G•T mismatches. Later, TDG was also found to act in concert with transcription factors in the regulation of gene expression. Whereas in some cases the function of TDG in gene regulation appeared to be a purely structural one (Chen et al. 2003), its role as a co-activator of the retinoic acid receptors (RAR/RXR), for instance, was shown to require its catalytic activity. In the attempt to connect these two seemingly distinct functions, TDG has been proposed to act as a DNA demethylase, removing 5-mC from the promoter regions of genes for transcriptional activation (Jost 1993; Jost et al. 1995; Zhu et al. 2000). However, as TDG appeared not to have direct 5-mC glycosylase activity its role in active DNA demethylation in mammals has remained controversial and the mechanism in general elusive.

With its apparently two-sided nature, TDG has riddled researchers for many years and the stimuli and interactions that control TDG function are still under investigation. The aim of my thesis was to dissect the role of TDG in DNA repair with a focus on its regulation by post-translational modification, and to investigate how TDG-initiated BER contributes to epigenetic stability at CpG islands (CGIs) during cell differentiation.

Both described functions of TDG, in DNA repair and in the regulation of gene expression, require its post-translational modification and non-covalent interaction with the small ubiquitin-like modifiers, SUMO1 and SUMO2/3 (Hardeland et al. 2002; Steinacher and Schar 2005; Mohan et al. 2007). Extensive biochemical studies by our laboratory have shown that SUMOylation of TDG may induce its dissociation from the abasic (AP-) site after base excision (Hardeland et al. 2002; Steinacher and Schar 2005). However, *in vivo* evidence corroborating an involvement of SUMOylation in TDG-dependent BER has been pending and the function of non-covalent SUMO-binding has remained elusive. I thus generated a Fluorescence Resonance Energy Transfer (FRET) system to monitor the interaction between TDG and SUMO1 or SUMO3 in cells. I was able to confirm a modulation of the SUMO1-TDG interaction dynamics in response to DNA damage, whereas the interaction with SUMO3 remained unaffected. This finding suggests that SUMO3 might regulate TDG function in a context other than DNA repair. In the light of recent findings that TDG is indeed involved in processes beyond canonical DNA repair, i.e. in maintaining the epigenetic stability of CGIs, our FRET system provides a powerful tool to investigate the role of SUMOylation and also specifically SUMO-binding in regulating TDG function in these pathways.

As the interaction partners of TDG range from DNA repair factors to transcription factors and even DNA methyltransferases (DNMTs) (Cortazar et al. 2007; Li et al. 2007; Boland and Christman 2008),

we set out to investigate the biological function of TDG genetically, by generating a *Tdg* knockout mouse. Much to our surprise, and in contrast to any other known DNA glycosylase, deletion of *Tdg* caused embryonic lethality. Further characterization of MEFs isolated from TDG-proficient and -deficient embryos revealed no evidence for a DNA repair defect, but a significant number of misregulated genes in differentiated *Tdg*^{-/-} cells. We found that *Tdg* knockout caused a loss of active histone marks, gain of repressive histone modifications and an accumulation of CpG methylation at CGI promoters. Interestingly, this phenotype became apparent only in differentiated but not in pluripotent cells. From these data, we proposed a dual function of TDG in maintaining active chromatin states at promoters during cell differentiation, first by structurally coordinating histone modifying enzymes and second by counteracting errors of the DNA methylation machinery by initiating repair of aberrantly methylated cytosines in CGIs. Consistent with a TDG-dependent engagement of DNA repair at such sites, we found BER factors to associate with these promoters and DNA repair intermediates to accumulate in differentiating cells in a TDG dependent manner.

These findings established a role of TDG in maintaining epigenome integrity in the context of cell differentiation. To investigate how TDG is involved in DNA methylation control, we mapped DNA methylation in the genomes of TDG-proficient and -deficient mouse embryonic stem cells (ESCs), neuronal progenitor cells (NPs) and MEFs and again found differential methylation to arise only with differentiation. Further characterization of the resulting differentially methylated regions (DMRs) revealed that those overlapping with a CGI were almost exclusively hypomethylated in TDG-deficient compared to -proficient cells, reflecting a failure to establish methylation at these CGIs during differentiation. In search of the reason for this failure in a 24 h differentiation timecourse, we found global 5-mC levels to rise with differentiation in cells lacking TDG activity, in parallel to the generation of the final products of TET-protein catalyzed 5-mC oxidation, 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), the latter two of which are proposed intermediates of active DNA demethylation and substrates for TDG. Differentiation thus appeared to induce methylation but also 5-mC oxidation to 5-fC and 5-caC for subsequent active demethylation by TDG. We therefore analyzed 5-mC and 5-caC levels at the CGI DMRs and found both to rise with differentiation in wildtype cells, suggesting that the loss of pluripotency induces a cycle of DNA methylation and demethylation at specific CGIs. In *Tdg* knockout cells, though, this induction appeared to fail whereas in cells expressing a catalytically dead mutant TDG (TDG_{Δcat}), the cycle of methylation and demethylation was induced but blocked by the inability of TDG_{Δcat} to excise 5-caC. In these cells, 5-caC will eventually be erased passively by DNA replication but this way of restoring an unmethylated C appears not to be sufficient to maintain the cycle and eventually establish methylation at these CGIs.

Taken together, in collaboration with colleagues from different laboratories I was able to show that differentiation triggers a state of high epigenetic plasticity at these CGIs and that catalytically active TDG is required to maintain an equilibrium of DNA methylation and demethylation. The imbalance of epigenetic marks resulting from knockout of TDG disrupts gene expression programs and the accumulation of aberrations eventually leads to loss of viability on the cellular and on the organismic level.

2 Introduction

2.1. DNA Repair and the Concept of Genome Maintenance

DNA is the carrier of genetic information, encoding the building blocks of all organisms. The DNA forms a double helical structure, consisting of two complementary single-strands that are made up of a deoxyribose-phosphate backbone supporting a sequential assembly of nucleosides with four different bases: the purines adenine (A) and guanine (G), and the pyrimidines thymine (T) and cytosine (C). The bases of opposite strands form specific hydrogen bonds with one another, establishing the complementarity of the two single-strands through the so-called Watson-and-Crick base pairing: G pairs with C via three hydrogen bonds, while A and T form two hydrogen bonds. All DNA-templated processes like replication, repair and transcription, rely on this basic principle.

2.1.1. Sources of DNA Damage and Modification

As an inherently instable molecule, the DNA is constantly at risk of being damaged by reacting with oxygen, water or other reactive agents of endogenous or exogenous origin.

Endogenous DNA damage

Endogenous DNA damage can arise, for instance, when the N-glycosidic bond between a DNA base and the deoxyribose hydrolyzes spontaneously, producing an apurinic or apyrimidinic (AP) site. Such AP-sites lack any instructive coding information and are therefore potentially mutagenic. Spontaneous damage can also derive from hydrolysis of the exocyclic amino groups of cytosine, 5-methylcytosine (5-mC), adenine and guanine, which converts them to uracil, thymine, hypoxanthine and xanthine, respectively. Thus, deamination of C and 5-mC generate C → T transition mutations, while the deamination product of A pairs preferentially with C, thus causing A → G transitions if unrepaired before the next round of DNA replication. As xanthine also pairs with C, deamination of G has no mutagenic effect. Another endogenous cause of DNA damage is the replication process itself, which can produce mismatches and double-strand breaks despite the high accuracy of the replicative DNA polymerases.

Oxidative damage can derive from reactions of the DNA with reactive oxygen species (ROS), which can arise as by-products of the normal cell metabolism. Oxidative damage to the bases can produce for example 8-oxoG which causes G → T mutations, or thymine glycol which interferes with replication fork progression.

Alkylation of bases can also be caused by endogenous agents; S-adenosylmethionine, for example, is an enzymatic cofactor which can accidentally generate 7-methylG, 3-methylA or O₆-methylG, the latter of which is pro-mutagenic as it pairs with T instead of C (Scharer 2003).

Endogenous DNA modification

It is striking that not all alkylation of bases is harmful to the integrity of the genetic code; certain base modifications have a function in cellular processes. Such modifications do not alter the base-pairing properties but change the DNA's overall charge and surface enough to affect DNA-protein interactions. The most prominent example is 5-mC, which occurs in mammals almost exclusively in the CpG context and serves as an epigenetic mark (see chapter 2.3.3.). The mechanisms that evolved to recognize small but harmful modifications to bases appear perfectly suited to read functional base modifications as well, a concept that is supported by the notion that members of the DNA glycosylase family of proteins, which initiate the Base Excision Repair (BER) pathway by recognizing and excising even bases with minor alterations, have been found to be involved in processes beyond canonical DNA repair (Jacobs and Schar 2012) (Appendix IV).

Deamination of bases not only occurs spontaneously but can also be enzymatically induced, which plays an important role in innate and adaptive immunity. Members of the apolipoprotein B mRNA editing catalytic polypeptide (APOBEC) family of cytosine deaminases confer innate immunity against retroviruses by deaminating cytosines in the viral cDNA and thus triggering its degradation by the concerted action of UNG2 and APE1 (Harris et al. 2003; Yang et al. 2007). Furthermore, the activation induced cytidine deaminase (AID) contributes to antibody maturation in adaptive immunity by deaminating cytosines in the course of somatic hypermutation to induce mutations in the light chain variable region of immunoglobulin loci (Pavri and Nussenzweig 2011). Also, deamination of cytosine followed by uracil processing that produces single-strand breaks required for the initiation of class switch recombination (Imai et al. 2003).

Exogenous DNA damage

Exogenous DNA damage can be caused for instance by UV light, which generates pyrimidine dimers by inducing the formation of cyclobutane rings between adjacent cytosines and/or thymines, thus

disturbing the double helical structure. These dimers are mostly paired with A during replication, which is why T-T dimers are less mutagenic than C-T and C-C dimers. The latter can cause C → T or CC → TT transition mutations if they remain unrepaired (Scharer 2003).

The integrity of the DNA structure can also be compromised by X-rays, α -, β - and γ -radiation, which are summarized as ionizing radiation. These can affect not only the bases but also the deoxyribose moieties of the DNA backbone, causing single- and double-strand breaks as well as DNA-DNA or DNA-protein-crosslinks. These effects arise either directly or through the generation of ROS, like hydrogen peroxide, superoxide radical anions or hydroxyl radicals (Scharer 2003).

Alkylating agents pose a threat to the DNA either by causing intra- or interstrand crosslinks which block replication fork progression or by aberrantly methylating DNA bases (Scharer 2003).

2.1.2. Molecular Mechanisms to Repair DNA Damage

To ensure the integrity of the genetic code, cells harbor an arsenal of repair mechanisms that address specifically different kinds of DNA damage. Figure 2-01 summarizes the most relevant causes of DNA damage, the type of DNA damage they produce and the repair pathway that fixes them.

Direct Repair

The most straightforward way a cell can address DNA damage is by direct chemical reversal of the lesion. Bacteria, plants and lower eukaryotes harbor so-called photolyases which can reverse UV light induced pyrimidine dimers in a light-dependent process called 'photoreactivation' (Essen and Klar 2006). Higher eukaryotes, on the other hand, depend on Nucleotide Excision Repair (NER) to address these photoadducts (see below).

O₆-alkylguanine transferases (AGTs, or in humans O₆-methylguanine methyltransferase, MGMT) are another example. These proteins convert O₆-alkylguanine back to guanine in an irreversible reaction that transfers the aberrant methyl-group from the guanine to an acceptor cysteine residue of the protein, after which they are inactive and targeted for proteolytic degradation (Pegg 2000). Finally, the bacterial iron and 2-ketoglutarate dependent oxygenase AlkB targets methyl-lesions (1-methylA and 3-methylC) in DNA and RNA, converting them to their original state through oxidative demethylation (Falnes et al. 2007).

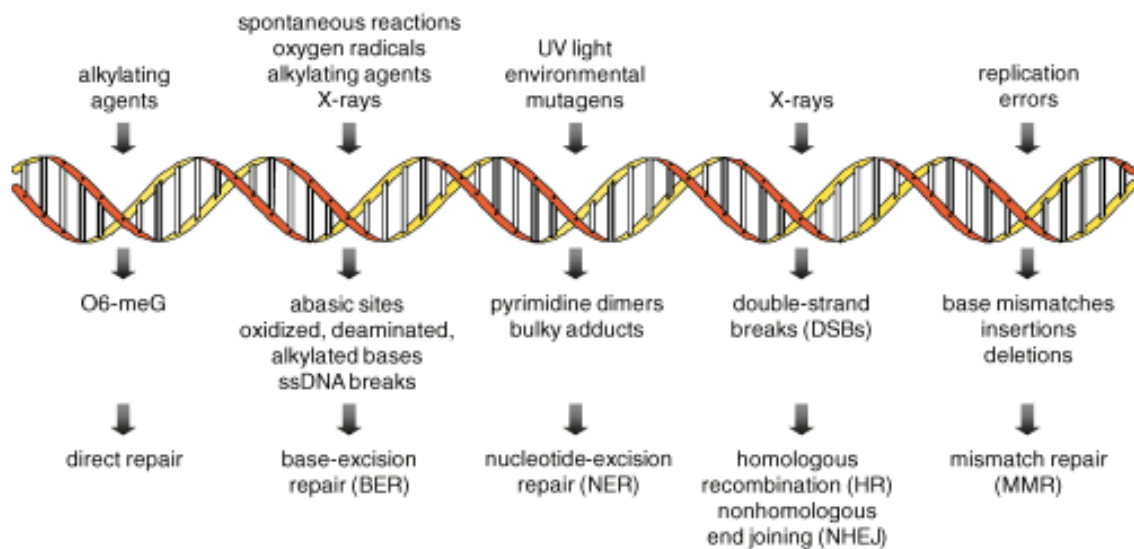


Figure 2-01: Overview of the most common DNA damaging agents, lesions and corresponding repair pathways. Summarized are the most important DNA damage causes (top), the lesions they induce (middle) and the pathways employed to repair these lesions (bottom) (Scharer 2003).

DNA Double-Strand Break Repair

DNA double-strand breaks (DSBs) can be the product of ionizing radiation or metabolic by-products like ROS but also arise during DNA replication or repair. DSBs are among the most harmful DNA lesions as they can lead to chromosomal aberrations and deletions, or if unrepaired cause cell death. In order to give the cellular repair machinery time to react, DSBs trigger cell cycle checkpoints that arrest cell cycle progression. Members of the phosphatidylinositol 3-kinase-like kinase (PIKK) family, namely ATM (ataxia-telangiectasia mutated) and ATR (ATM and Rad 3-related), are thought to activate the central checkpoint regulators p53, Chk1 and Chk2, which in turn activate Cdk2/CyclinE and Cdk2/CyclinB1 to mediate G1 and G2 arrest, respectively (Langerak and Russell 2011).

Depending on the cell cycle stage, DSBs are repaired preferentially either by homologous recombination (HR) or non-homologous end joining (NHEJ). HR is mostly employed in S and G2 phase when a sister chromatid is available as an identical copy of the damaged double-strand that can serve as a template for synthesis the original sequence across the DSB. This requires strand-invasion and the formation of so-called Holliday-junctions, the ultimate resolution of which may result in a reciprocal exchange of the engaged DNA double-strands, known as sister chromatid exchange. In G1 and early S phase, when a sister chromatid is not available, cells utilize the more error-prone NHEJ to repair DSBs and avoid cell death. In this pathway, free ends at the break are captured, tethered and

re-attached through minimal base-pairing, which often results in deletions and point mutations (Friedberg 2003).

Nucleotide Excision Repair

Bulky DNA lesions that perturb the overall structure of the DNA double helix and interfere with DNA replication and transcription are corrected by the Nucleotide Excision Repair (NER) machinery. NER can be subdivided by the mode of damage recognition into global genome repair (GGR) and transcription-coupled repair (TCR). As the name implies, TCR involves the transcription machinery sensing a lesion that blocks the progression of the RNA polymerase, whereas in the transcription-independent GGR pathway, a helix-distorting damage is recognized by a complex of XPC and HR23B. In both cases, the DNA is locally unwound and the damage-containing strand is cleaved 3' and 5' of the damage, generating a single-stranded DNA stretch of 24-32 nucleotides. The double-strand is restored through repair synthesis by polymerase δ and/or ϵ and ligation by DNA ligase I (Friedberg 2003).

Defects in NER are connected to a predisposition to cancer and premature ageing. Two syndromes are caused by loss-of-function of NER factors, namely Xeroderma pigmentosum caused by defects in GGR, and Cockayne syndrome resulting from defective TCR. The fact that certain types of Xeroderma pigmentosum are connected to neurological abnormalities and that patients with Cockayne syndrome display developmental and neurological defects, reflects the involvement of NER factors in processes beyond DNA repair (Kamileri et al. 2012).

Mismatch Repair

Although strictly taken no physical DNA damage, base mismatches and small insertions/deletions that derive from base misincorporation by and slippage of DNA polymerases during the DNA replication process have to be corrected to avoid mutations. Such replication errors are corrected by the Mismatch Repair (MMR) system that can distinguish between the template and the newly synthesized DNA strand. In eukaryotes, mismatched bases and small insertion/deletion loops are recognized by MutS α (heterodimer of hMSH2-hMSH6), larger loops by MutS β (heterodimer of hMSH2-hMSH3), which in subsequent steps recruit the MLH1/PMS2 heterodimer. This heterodimer, termed MutL α , recognizes the newly synthesized strand by replication-associated strand discontinuities, e.g. between unprocessed Okazaki fragments. Furthermore, the PMS2 subunit of

MutL α was found to possess endonuclease activity in concert with MutS α , PCNA and replication factor C (RFC) that could induce strand discontinuities (Kadyrov et al. 2006). Such a nick in close proximity to the mismatch serves as an entry point for excision of the newly synthesized strand by exonucleases (e.g. EXO1) through and beyond the mismatch, followed by DNA repair synthesis by a replicative DNA polymerase and sealing of the nick by DNA ligase I (Friedberg 2003).

2.1.3. Base Excision Repair and Single-strand Break Repair

Single-base damage caused by deamination, oxidation or alkylation is mainly addressed by Base Excision Repair (BER). As this pathway is central to my thesis, I summarize it here in more detail.

This pathway is initiated by DNA glycosylases, a family of enzymes that evolved to specifically recognize irregular DNA base modifications. DNA glycosylases recognize a substrate base by flipping it out of the double-helical context into an active site pocket. Subsequently, they cleave the N-glycosidic bond between the base and the deoxyribose, which produces an AP-site. In Appendix IV, I am presenting a detailed review on this ancient family of enzymes, their mode of action and their involvement in processes beyond canonical DNA repair, for instance in adaptive immunity and active DNA demethylation.

Two sub-pathways are distinguished in BER, depending on whether just a single nucleotide is replaced (short-patch) or a stretch of 2-13 nucleotides (long-patch, Fig. 2-02). Short-patch BER involves a mono- or bifunctional glycosylase, distinguished by their mode of base excision. Monofunctional glycosylases only catalyze base excision by utilizing an activated water molecule for nucleophilic attack on the N-glycosidic bond. Bifunctional glycosylases use the amino group of a conserved lysine for this purpose, resulting in a covalent Schiff's base intermediate. Their inherent 3' AP lyase activity enables bifunctional glycosylases to cleave the AP-site by β -elimination. The resulting 3' α,β -unsaturated aldehyde is processed by an AP endonuclease (e.g. APE1 in human and mouse) to produce the 3'OH required for repair synthesis.

Since monofunctional glycosylases have no AP lyase activity, nicking of the AP-site is performed by APE1 which generates a 3'OH and a 5'dRP end. DNA polymerase β (Pol β) harbors a 5'dRP lyase activity, thus hydrolyzing the 5'dRP prior to filling the single-nucleotide gap. The remaining nick is sealed either by DNA ligase I or by a complex of DNA ligase III with XRCC1 (Almeida and Sobol 2007) (Fig. 2-02).

The 5' end produced by APE1 upon incising an oxidized or reduced AP-site cannot be processed by the 5'dRP lyase activity of Pol β . In this case, a stretch of DNA 3' to the excised bases is displaced and resynthesized through long-patch BER. The current mechanistic model proposes that strand-

displacement is accomplished through DNA synthesis by polymerase δ , ϵ or β , aided by the Proliferating Cell Nuclear Antigen (PCNA), the replication factor C (RFC) and the poly (ADP-ribose) polymerase 1 (PARP1). The resulting DNA flap structure of 2-13 nucleotides is degraded by the flap endonuclease 1 (FEN1) and the ends are sealed by DNA ligase I (Fig. 2-02) (Almeida and Sobol 2007; Fortini and Dogliotti 2007).

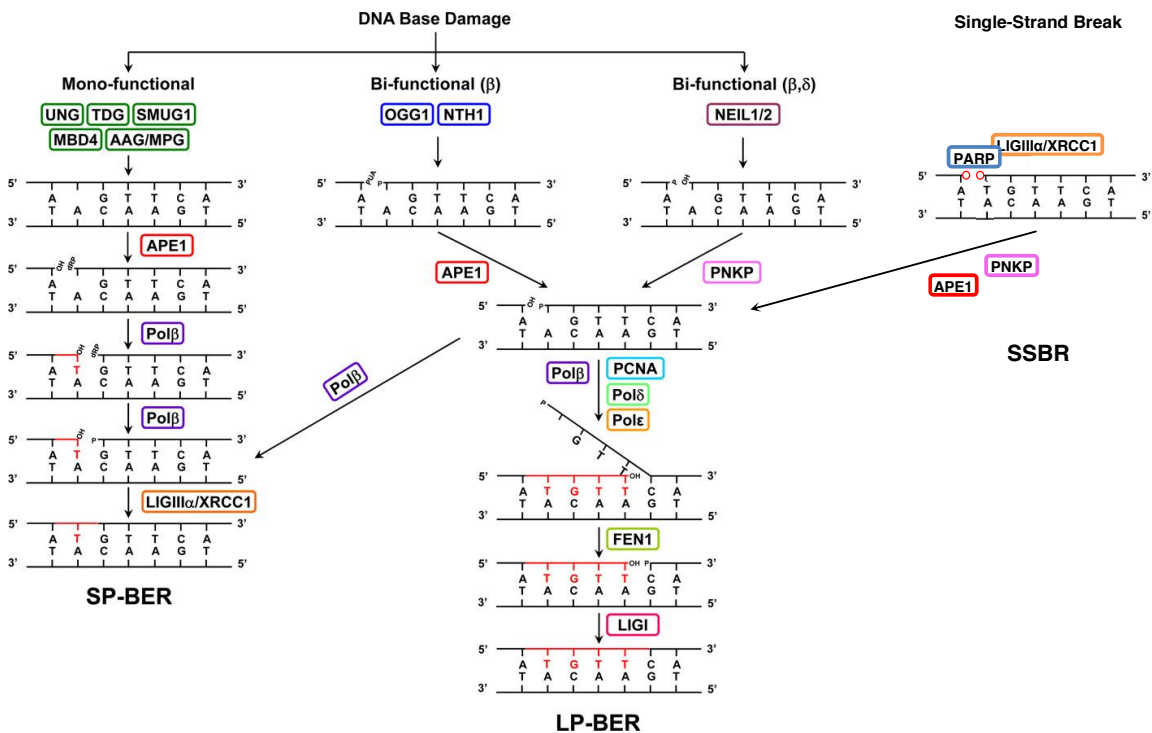


Fig. 2-02: Overview of short- and long-patch BER and SSBR in human cells. Short-patch (SP-) BER is initiated by a mono- or bifunctional glycosylase excising a substrate base. APE1 or the inherent AP lyase activity of a bifunctional glycosylase cleaves the AP-site, followed by processing of the free DNA ends to allow repair synthesis by Pol β . Finally, the remaining nick is ligated by LigaseIII (LIGIII α) in complex with XRCC1. Long-patch BER is employed when strand incision by APE1 produces ends that are refractory to end processing by Pol β . In this case, the damaged strand is displaced through DNA synthesis by Pol β , δ or ϵ aided by PCNA, the resulting DNA flap is cleaved by FEN1 and the ends are sealed by Ligase I (LIG1). An APE1-independent pathway is initiated by NEIL1 or 2, followed by end-processing by PNKP (PNKP). SSBR is initiated by PARP which recognize a single-strand break and recruits XRCC1/LIGIII α , which in turn provide a scaffold for the assembly of downstream factors. The damaged 5' and 3' termini (red circles) are further processed by APE1 or PNKP to yield 3'OH and 5' phosphate moieties. Subsequent steps are similar if not identical to SP- or LP-BER. Adapted from (Kim and Wilson 2012) with information from (Caldecott 2003).

Downstream of base excision and AP-site incision, BER converges with the single-strand break repair pathway (SSBR). Hence, BER and SSBR share many proteins and features except for the initiation step, which, in the latter case, appears to be coordinated by the poly (ADP-ribose) polymerase 1

(PARP1) instead of a DNA glycosylase. PARP1 recognizes the single-strand break and recruits XRCC1 which serves as a scaffold for the assembly of the SSBR complex (Masson et al. 1998). The DNA ends are processed to generate a 3'OH and a 5' phosphate moiety, which can involve several enzymes depending on the type of residue at the 3' and 5' end, e.g. APE1 or the polynucleotide kinase (PNK) (Fig. 2-02). The subsequent repair steps are similar, if not identical to those of BER with regard to the factors involved and the short- or long-patch repair sub-pathways (Fortini and Dogliotti 2007).

Oxidized bases like dihydrothymine or α -anomeric 2'-deoxynucleosides like α -2'-deoxyadenosine can also be repaired by nucleotide incision repair (NIR) a pathway equivalent to BER except for the initiation step. NIR is independent of DNA glycosylases but relies on APE1 which cleaves the DNA backbone 5' of these bases. NIR of deaminated purine bases is initiated by EndoV which incises one nucleotide 3' of the lesion (Dalhus et al. 2009).

Moreover, oxidized bases have been shown to be repaired independently of APE1, following the action of the bifunctional glycosylases NEIL1 and 2 that catalyze a simultaneous beta-delta elimination. The resulting 3' phosphate end is processed by PNK to generate the 3'OH required by Pol β for repair synthesis (Fig. 2-02) (Wiederhold et al. 2004; Das et al. 2006; Kim and Wilson 2012).

2.2. The Thymine DNA Glycosylase (TDG)

2.2.1. Classification and Characterization of TDG

DNA glycosylases can be categorized into three superfamilies: the Helix-hairpin-Helix (HhH) glycosylases, the Endonuclease VIII-Like (NEIL) glycosylases and the Uracil DNA glycosylases (UDG). The latter has been named after the *Escherichia coli* Uracil-N-Glycosylase (Ung), the first glycosylase found to excise uracil from DNA. Uracil in DNA can arise through deamination of cytosine or U misincorporation during DNA replication, producing G•U mismatches or non-mutagenic A•U pairs, respectively. While the recognition and repair of uracil as a foreign base in DNA appears relatively straightforward, the deamination of methylated cytosine (5-mC) gives rise to thymine, which as a normal DNA base cannot be easily recognized as damage. Thus, a subfamily of the UDGs evolved to recognize and excise thymine when it is mispaired with guanine. These monofunctional Mismatch-specific Uracil DNA Glycosylases (MUGs) form specific contacts with the opposing base to distinguish a T derived from deamination of 5-mC from a canonical A•T pair. The first member of this family isolated was the human TDG, identified by its ability to recognize and excise thymine opposite guanine (Wiebauer and Jiricny 1989; Neddermann and Jiricny 1993).

Members of the MUG protein family harbor an α/β fold structural domain that is characteristic of the UDG superfamily. All MUGs have a common and rather simple architecture, consisting of a conserved core domain containing the active site flanked by variable N- and C-terminal domains. Within the core domain, the MUG orthologs share 37-52% sequence identity on the amino acid level (Cortazar et al. 2007). Members of this protein family have been identified in organisms throughout the tree of life, from *E. coli* and *Schizosaccharomyces pombe* over *Drosophila melanogaster* to *Xenopus laevis* and mammals (Wiebauer and Jiricny 1989; Neddermann and Jiricny 1993; Gallinari and Jiricny 1996; Hardeland et al. 2003). One characteristic of the MUG proteins is their large catalytic cavity that can accommodate a broad spectrum of damaged bases for excision (Barrett et al. 1999).

Human TDG is made up of 410 amino acids. The mouse ortholog exists as two splice variants, TdgA and TdgB, the latter lacking the first 24 amino acids of the N-terminus (Neddermann and Jiricny 1993; Gallinari and Jiricny 1996). Whether these two isoforms serve distinct biological functions remains unclear. The N-terminal domain of TDG has been shown to be essential for efficient processing of G•T mismatches (see chapter 2.2.2.) and both terminal domains are involved in protein-protein interactions (see chapter 2.2.3).

2.2.2. Mechanism of Substrate Recognition and Processing

Structural studies on *E. coli* Mug have provided insight into the catalytic mechanism of damage recognition and base excision by the MUG proteins. While the damaged base is flipped into the catalytic pocket, conserved amino acid residues within that cavity form a wedge that takes the place of the damaged base in the double-stranded DNA, forming specific contacts and thus mimicking Watson-Crick base pairing with the widowed G. Due to this combined nucleotide flipping/intercalation mechanism, MUG proteins only process modified bases in double-stranded DNA and exhibit a strong preference for substrates opposite guanine (Barrett et al. 1998).

Extensive biochemical studies have shed light on the mechanism employed by TDG to search for, recognize and excise a damaged base. The flexible N-terminal domain mediates non-specific DNA binding, switching from an open to a clamp-like conformation upon binding to DNA. This DNA binding capacity of the N-terminal domain has also been found to be essential for effective processing of G•T, probably stabilizing the glycosylase-substrate complex (Hardeland et al. 2002; Hardeland et al. 2003; Steinacher and Schar 2005).

The clamp-like configuration presumably allows TDG to slide along the DNA in search of a substrate base. When encountering a lesion, TDG employs the same combined nucleotide flipping/intercalation mechanism as Mug. A highly conserved asparagine residue within the catalytic

pocket (in human N140, in mouse TdgA N151) positions an activated water molecule for hydrolytic attack on the N-glycosidic bond between the base and the deoxyribose (Barrett et al. 1998; Barrett et al. 1999; Hardeland et al. 2000). In contrast to Mug, TDG remains firmly bound to the AP-site after release of the damaged base. This product inhibition is mediated by the specific contacts formed with the widowed guanine and the non-specific DNA binding of the N-terminus (Hardeland et al. 2002; Steinacher and Schar 2005). As AP-sites represent a lesion that is even more hazardous than the original base damage as they can easily turn into single- and double-strand breaks, their hand-over to downstream acting repair factors has to be tightly controlled. The high affinity of TDG to its product AP-site probably serves the purpose of stabilizing this dangerous repair intermediate. It was shown that post-translational modification with Small Ubiquitin-like Modifiers (SUMOs) regulates this hand-over as well as the dissociation of TDG (see chapter 2.2.4.) (Hardeland et al. 2002; Steinacher and Schar 2005).

2.2.3. Biological Functions of TDG

Classical DNA repair

TDG recognizes and excises U or T mispaired with G, arising from spontaneous or enzymatic deamination of cytosine or 5-mC, respectively. Methylated cytosine (5-mC) is even more sensitive to spontaneous hydrolytic reactions at its exocyclic amino-group than cytosine (Ehrlich et al. 1990). 5-mC occurs predominantly in CpG dinucleotides which have been found to be a hotspot for mutation, possibly not only because of 5-mC deamination but also because alkylation damage to the neighboring G is repaired less efficiently by MGMT (Bentivegna and Bresnick 1994). Both, deamination of 5-mC and unrepaired O₆-methylguanine would result in C → T transition mutations. C → T transitions are among the most frequent mutations associated with human cancer (Sjoblom et al. 2006; Wood et al. 2007; Rubin and Green 2009), making up about 25% of all somatic mutations in the p53 tumor suppressor gene in human cancers, in certain tumors even ~50% (Petitjean et al. 2007).

Furthermore, the spontaneous deamination of 5-mC (but possibly also inefficient repair of O₆-methylguanine in a methylated CpG) has been proposed to have caused the underrepresentation of CpG dinucleotides observed in organism with DNA methylation. CpG islands (CGIs) are characteristically hypomethylated, which has been suggested to have preserved their high CpG content (Antequera 2003; Jones 2012). What mechanisms maintain CGIs in a hypomethylated state,

however, remains unclear although recent studies suggest an involvement of TDG-dependent BER in protecting CGIs from aberrant methylation ((Illingworth and Bird 2009) and Appendices II and III).

Four glycosylases have evolved capacities to counteract the deamination of cytosine, i.e. mutagenesis by G•U mismatches: the Uracil N-Glycosylase UNG, the Single-strand specific Monofunctional Uracil Glycosylase SMUG1, Methyl-binding domain glycosylase MBD4 and TDG. Of these, MBD4 and TDG also process G•T mismatches that arise from 5-mC deamination. In agreement with a function in mutation avoidance, knockout of *Mbd4* and knockdown of *Smug1* result in a mild increase in C → T transition mutations (Millar et al. 2002; Wong et al. 2002; An et al. 2005). Inhibition of UNG in human cells also leads to a mild increase in mutations. In agreement with a role of UNG in innate and adaptive immunity downstream of enzymatic cytosine deamination, knockout of *Ung* in mice results in the development of B-cell lymphomas, deficiencies in antibody diversification and defective innate immunity against retroviral infection (Radany et al. 2000; Rada et al. 2002; Nilsen et al. 2003). SMUG1 is thought to serve as a back-up for UNG in counteracting the accumulation of uracil in genomic DNA, which has recently been corroborated by the knockout of *Smug1* in mice (An et al. 2005; Kemmerich et al. 2012).

With its large catalytic cavity, TDG accommodates a broad spectrum of substrates comprising not only G•U and G•T mismatches but also uracil-derivates modified at the C5 position, e.g. 5-fluorouracil or 5-bromouracil, etheno-adducts like 3,N⁴-ethenocytosine and oxidized pyrimidines like thymine glycol (see Table 2-01) (Hardeland et al. 2003). Furthermore, TDG was reported to have 5-mC glycosylase activity, yet these findings could not be corroborated so far (Jost 1993; Zhu et al. 2000). Recently, TDG has been shown to process 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), the oxidation products of 5-hydroxymethylcytosine catalyzed by the Ten Eleven Translocator (TET) family of 5-mC hydroxylases (see chapter 2.4.) (He et al. 2011; Maiti and Drohat 2011).

In contrast to UNG, MBD4 and SMUG1, the knockout of *Tdg* does not increase mutation frequencies in standard mutation assays, suggesting that its function in classical DNA repair is neglectable, at most redundant. Interestingly though, in contrast to all other DNA glycosylases, we and others have found deletion of *Tdg* in mice to cause embryonic lethality, hinting at a non-redundant function of TDG in embryonic development (Cortazar et al. 2011; Cortellino et al. 2011).

Still, the DNA repair function of TDG is not entirely obsolete, as was demonstrated in the context of 5-FU processing. 5-FU is a base analog that is often used in chemotherapy as it is incorporated into RNA and DNA, disturbing RNA synthesis and DNA replication. TDG has been found to be one of the major contributors to 5-FU induced cytotoxicity. It excises this base analog from A•5-FU base pairs, lingering on the resulting AP-site and thus interfering with downstream repair processes, which leads to an accumulation of DNA strand breaks. Accordingly, depleting MEFs and HeLa cells of TDG rendered them resistant to 5-FU (Kunz et al. 2009).

Table 2-01: Substrate spectrum of TDG orthologs*

Substrate**	hsTDG	hsTDG Δ N	hs/mmTDG $_{\Delta$ cat	ecMug	spThp1p	dmThd1p
G•U	+++	+++	nd	+++	+++	+++
A•U	+	+	nd	+	+++	++
ss U	-	-	nd	-	+++	-
G•FU	+++	nd	++	nd	+++	+++
A•FU	++	nd	-	nd	+++	++
ss FU	++	nd	nd	nd	++	++
G•BrU	+++	nd	nd	nd	++	+++
A•BrU	+	nd	nd	nd	+	-
ss BrU	-	nd	nd	nd	+	-
G•hmU	+++	nd	nd	+	-	++
G•hU	+++	nd	nd	++	+++	nd
G•T	+++	-	-/+	-	-	++
G•Tg	++	nd	nd	nd	nd	nd
G•ϵC	+++	nd	nd	+++	+++	+++
A•ϵC	++	nd	nd	nd	+++	++
ss ϵC	-	nd	nd	+	+++	-
G•Hx	+	+	nd	nd	+++	+
T•Hx	-	-	nd	-	+++	-
ss Hx	-	nd	nd	nd	+++	-
G•ϵA	-	nd	nd	+	++	-
T•ϵA	-	nd	nd	-	++	-
ss ϵA	-	nd	nd	nd	+	-
G•mC	-/+	nd	-	nd	-	-
G•hmC	-	nd	-	nd	nd	nd
G•fC	+++	nd	nd	nd	nd	nd
G•caC	+++	nd	-/+	nd	nd	nd
ss caC	++	nd	nd	nd	nd	nd
G•heC	-	nd	nd	++	+++	nd
G•hpC	-	nd	nd	-	+++	nd
G•G	-	nd	nd	nd	+	-

* Relative processing efficiencies of recombinant human full size TDG (hsTDG), N-terminally truncated TDG (hsTDG Δ N) and human or murine catalytic amino acid residue mutated TDG (TDG $_{\Delta$ cat, N140A in human, N151A in mouse TdgA) and the orthologs of *E. coli* (ecMug), *S. pombe* (spThp1p) and *D. melanogaster* (dmThd1p). Base release efficiencies are indicated as: +++, high; ++, intermediate; +, low; -, insignificant; nd, not determined.

** The putative substrate base is 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-. Adapted from (Cortazar et al. 2007) with information from (Hardeland et al. 2001; He et al. 2011; Maiti and Drohat 2011) and Alain Weber, personal communication and Appendix III.

Transcriptional regulation

TDG has been shown to interact with several transcription factors, transcriptional co-activators and the DNA methyltransferases Dnmt3a and b. The first indication for an involvement of TDG in transcriptional regulation was provided by Chevray and colleagues in 1992, who reported a direct interaction between the glycosylase and c-Jun, a basic leucine zipper transcription factor that is part of the activator protein 1 (AP-1) complex (Chevray and Nathans 1992). The function of TDG as a co-activator of transcription was corroborated when it was found to interact with two nuclear receptors, the retinoic acid receptor (RAR) and the retinoid X receptor (RXR), to potentiate their binding to retinoic acid response elements (RAREs) and to enhance the transactivation of a reporter gene by RAR/RXR (Um et al. 1998). A structurally intact but catalytically inactive mutant TDG (N151A) failed to significantly stimulate RAR/RXR-mediated transcription, suggesting that its glycosylase activity is essential for its co-activator function in this context (Hardeland, U. and Schär, P., unpublished data). Furthermore, TDG was found to interact with estrogen receptor alpha (ER α) in a ligand-dependent manner. ER α and ER β are the nuclear receptors mediating the major responses to estradiol. TDG acts as a co-activator of ER α but, unlike with RAR/RXR, its glycosylase activity is dispensable for this function. Therefore, it was suggested that TDG serves as a structural scaffold in this case (Chen et al. 2003).

Similarly, the catalytic activity of TDG was found to be dispensable for a physical and functional interaction with the CREB binding protein (CBP) and its paralog p300. CBP/p300 are transcriptional co-activators with intrinsic histone acetyltransferase activity. They stimulate transcriptional activation in collaboration with a number of sequence-specific transcription factors, e.g. CREB and p53, through chromatin modeling and interactions with the basal transcription machinery (Goodman and Smolik 2000). TDG interacts with CBP/p300 through its N- and C-terminal domains, potentiates CBP-activated transcription and was found to be a substrate for acetylation by CBP/p300. Interestingly, TDG acetylation abolishes its interaction with CBP as well as with the downstream-acting BER factor APE1 (Tini et al. 2002).

Two reports describe a repressive effect of TDG on gene transcription. Rat TDG was found to interact with the thyroid transcription factor TTF1 and to repress TTF1-activated transcription in transient co-transfection experiments (Missero et al. 2001). Additionally, the interaction of TDG with Myocardin, a co-activator of the serum response factor (SRF) in the regulation of smooth muscle-specific gene expression, was found to interfere with the Myocardin-SRF interaction and binding of Myocardin to its target promoters (Zhou et al. 2008).

The function of TDG in the regulation of gene expression has been associated with the establishment and maintenance of epigenetic marks during cell differentiation. The role of TDG in this context is described in chapters 2.4.2. as well as in Results 4.2. (Appendix II) and 4.3. (Appendix III).

2.2.4. Regulation of TDG Function by Post-Translational Modification

Modification by Small Ubiquitin Like Modifiers (SUMOs)

The high affinity of TDG for its product AP-site benefits the cell by protecting these dangerously fragile sites and thereby preventing the spontaneous generation of DNA single- and double-strand breaks. On the other hand, this tight AP-site binding interferes with downstream steps of BER, which, for instance in the case of the base analog 5-FU that is frequently used in cancer therapy, results in the accumulation of unrepaired AP-sites that eventually trigger cell death (Kunz et al. 2009). Correct timing of the hand-over of this repair-intermediate to the downstream BER factors is therefore of utmost importance and requires a tight regulation. TDG was shown to be post-translationally modified by Small Ubiquitin-like Modifiers (SUMOs), which was implicated in regulating TDG's dissociation from the AP-site (Hardeland et al. 2002; Steinacher and Schar 2005).

SUMOs are structurally similar to ubiquitin but share less than 20% sequence homology. SUMOylation can modulate structural and functional features of its target protein, including the stability, subcellular localization, protein-protein interactions and activity. Four distinct SUMO genes have been identified in the human genome, encoding SUMO1-4. SUMO2 and 3 share 97% sequence homology and are therefore often referred to as SUMO2/3. While SUMO1 and 2/3 are ubiquitously expressed, SUMO4 is only found in kidney, lymph nodes and spleen (Geiss-Friedlander and Melchior 2007).

TDG has been shown to be modified by SUMO1 and 2/3 (Hardeland et al. 2002). The SUMO-conjugation process requires several enzymatic steps, starting with a SUMO-activating enzyme E1 (SAE1/SAE2), which catalyzes the formation of a thioester bond between SUMO and SAE2. Subsequently, SUMO is transferred to a cysteine in the SUMO-conjugating enzyme E2 (UBC9) and further to a lysine residue in a target protein. The SUMO acceptor lysine in TDG (K341 in mouse TdgA) lies within a C-terminal SUMOylation consensus motif (VKEE) (Hardeland et al. 2002). This last step of the conjugation process is often – but not always – catalyzed by a SUMO E3 ligase which is thought to mediate target specificity (Geiss-Friedlander and Melchior 2007). In the case of TDG-SUMOylation, such an E3 ligase has yet to be identified.

Extensive biochemical analyses have shed light onto the regulation of TDG function by SUMOylation (Fig. 2-03). The flexible N-terminus of TDG undergoes a conformational change upon binding to homoduplex DNA, mediating non-specific DNA interaction that has been found to be essential for effective G•T processing (Fig. 2-03 B) (Hardeland et al. 2002). This binding capacity keeps TDG firmly attached to the AP-site after base excision (Fig. 2-03 D). SUMOylation reverses the conformational switch in the N-terminus, abolishing the tight AP-site binding (Fig.2-03 E). Finally, deSUMOylation by a Sentrin-specific protease (SENP) restores the DNA binding ability of TDG (Steinacher and Schar 2005).

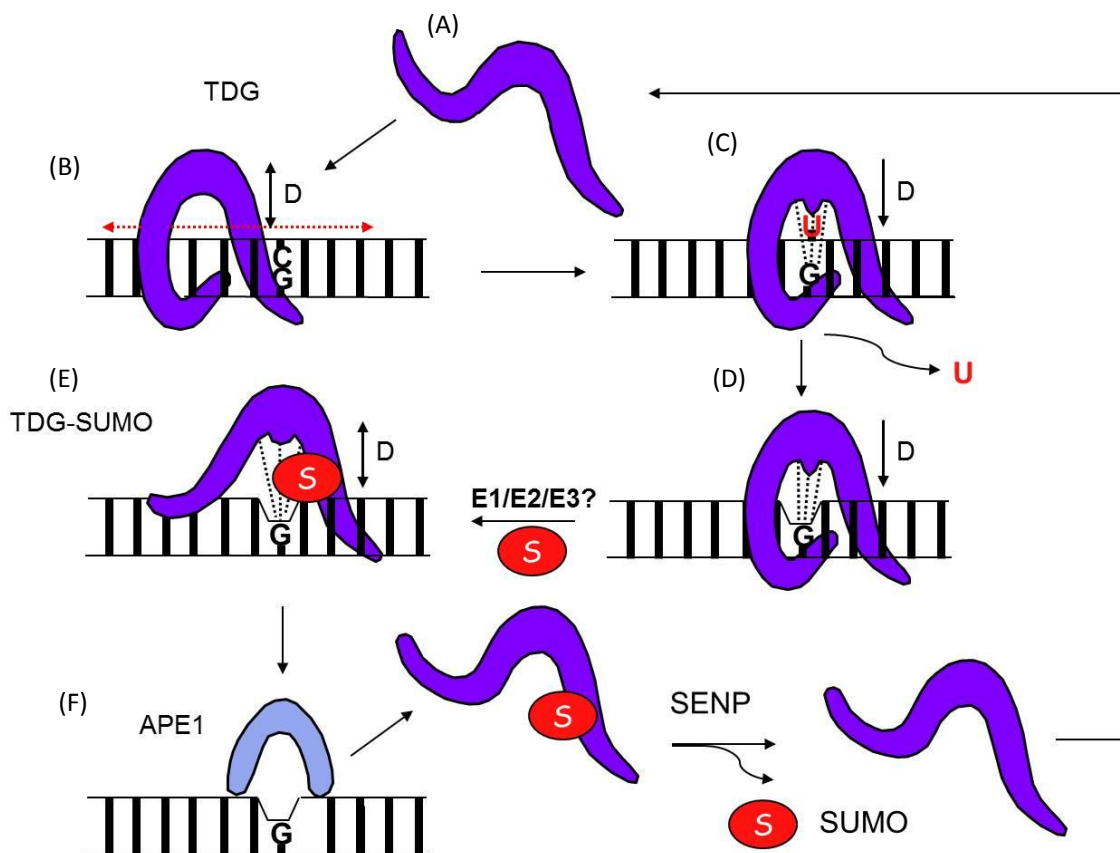


Fig.2-03: A model for SUMO modulated dynamic DNA interactions of TDG during BER. (A) DNA-free TDG is present in the cell nucleus in an open conformation. (B) Upon binding to DNA, the N-terminus forms a closed structure with the catalytic core domain. This clamp-like conformation may allow TDG to slide along the DNA in search of a potential substrate. (C and D) G•U- or G•AP-site-bound TDG reflects a third conformational state where the catalytic site forms specific contacts with the guanine opposite (dashed lines). The non-specific and the specific DNA contacts now cooperate to keep TDG firmly bound to the substrate. (E) SUMOylation of TDG then induces a fourth conformational state, neutralizing the non-specific DNA interactions of the N-terminus and facilitating the dissociation of the enzyme from the AP-site. (F) APE1 gains access to the AP-site and carries on the BER process. DeSUMOylation by SENP proteins allows recycling of TDG and SUMO. (Steinacher and Schar 2005)

A role of SUMOylation in regulating the release of TDG from the AP-site has been corroborated by structural analyses showing that SUMO-conjugation induces the formation of a protruded α -helix within the catalytic domain that was proposed to facilitate TDG dissociation (Baba et al. 2005, 2006). Furthermore, in agreement with a role of SUMOylation in timing the dissociation of TDG from the AP-site until downstream BER factors are in place to take over the repair intermediate, presence of APE1 stimulates the release of TDG (Fig. 2-03 F) (Waters et al. 1999).

In addition to being SUMOylated, TDG contains two SUMO Interaction/Binding Motif (SIM or SBM), one in the N- and one in the C-terminal domain (Mohan et al. 2007). SBMs mediate non-covalent SUMO-interactions and consist of a hydrophobic core with N- or C-terminally flanking acidic and/or serine residues (Minty et al. 2000; Song et al. 2004; Hecker et al. 2006).

The function of non-covalent SUMO binding of TDG is less clear than that of SUMOylation. While SUMOylation is thought to induce conformational changes in its target proteins that affect their affinity to other proteins, SUMO-binding is regarded as a kind of proteinaceous glue, stabilizing complexes through interactions of SBM-containing proteins with SUMO-conjugated factors. An example for this function is the formation of promyelocytic leukemia protein (PML) nuclear bodies (NBs). SUMOylation of PML has been suggested to induce the assembly of PML networks through non-covalent interactions with the SBM of other PML molecules (Shen et al. 2006). A similar scenario might also apply for TDG as it was shown to localize to PML-NBs in a SUMO-dependent manner. However, the role of SUMO-binding and SUMOylation in regulating the interaction between TDG and PML is not yet clearly defined. There have been contradictory reports, one claiming that SUMO-binding and SUMOylation of TDG are necessary for the TDG-localization to PML-NBs, the other proposing that SUMO-binding mediates and SUMOylation of TDG abolishes the interaction with PML (Takahashi et al. 2005; Mohan et al. 2007).

Furthermore, SUMO1-binding of TDG has been shown to be essential for the stimulating effect on CBP-dependent transcription. SUMOylation of TDG on the other hand proved to abolish the interaction with and acetylation by CBP (Mohan et al. 2010). As the interaction with CBP involves the N-terminal domain of TDG, the conformational change induced by SUMOylation might disrupt this interaction (Tini et al. 2002). Also, intramolecular interactions of conjugated SUMO with the two SBMs might interfere with interactions with other SUMOylated factors (Mohan et al. 2007). However, this point is not entirely clear as another study based on NMR spectroscopy revealed no competition between intermolecular and intramolecular SUMO-binding of the C-terminal SBM (Smet-Nocca et al. 2011). The same study also reported that SUMO1-binding but not SUMOylation induces a conformational change in TDG, thus contradicting previous work.

Taken together, SUMO-binding and SUMOylation play essential roles in regulating TDG function by mediating protein-protein interactions or abolishing protein-protein and protein-DNA interactions, respectively. However, the details of this regulation are not yet fully understood.

Modification of TDG by Ubiquitylation

The fact that four mammalian glycosylases with apparently redundant UDG activity have co-evolved suggests that they serve distinct functions. One way cells could diversify these enzymatically redundant functions is by cell cycle regulation. TDG was found to be degraded in S-phase through the ubiquitin-proteasome system, strikingly at the same time when expression of *UNG* is induced. While TDG levels are high in G2-M and G1 phase but is rapidly degraded at the onset of S-phase, the pattern of UNG2 levels is exactly inverse with a peak in early S-phase and ubiquitin-proteasome mediated degradation towards the end of the replication process (Fischer et al. 2004; Hardeland et al. 2007). While UNG2's main function appears to be the excision of misincorporated uracil (A•U) during DNA replication (Otterlei et al. 1999), TDG appears better suited for excision of uracil and thymine arising through deamination events in non-replicating DNA (G•U or G•T). The activity of TDG on A•U is rather low and its high affinity for AP-sites might interfere with the progression of replication forks. UNG2 on the other hand excises U opposite A with high efficiency and turnover. Exclusion of TDG from S-phase thus prevents its activity from interfering with the replication process (Hardeland et al. 2007).

Ubiquitylation, like SUMOylation, involves an ubiquitin-activating E1, a ubiquitin-conjugating E2 enzyme and an E3 ubiquitin ligase. The E3 responsible for TDG-ubiquitylation has yet to be identified. The ubiquitylation and SUMOylation systems have been reported to interact in antagonistic or non-antagonistic ways on certain target proteins, sometimes competing for the same acceptor site (Ulrich 2005). However, such a crosstalk in the case of TDG has not been described.

Phosphorylation and Acetylation

The N-terminal domain of TDG was found to be acetylated by CBP and phosphorylated by the protein Kinase C α (PKC α) in a mutually exclusive manner (Tini et al. 2002; Mohan et al. 2010). Both modifications occur on adjacent lysine (acetylation) and serine (phosphorylation) residues in a conserved sequence motif **SKKSGKS**. While phosphorylation can occur on free and DNA-bound TDG alike, DNA-binding reduces TDG susceptibility to acetylation, suggesting that this modification

requires release from the DNA. Additionally, acetylation was shown to decrease the homoduplex binding and G•T processing capacity of TDG but apparently does not affect AP-site binding (Mohan et al. 2010).

The role of phosphorylation in regulating the function of TDG is still unclear although it was proposed to preserve its G•T processing activity by preventing acetylation. Phosphorylation has been suggested to modulate the SUMO-modification of certain target proteins, e.g. in the case of I κ B α , the inhibitor of the transcription factor NF- κ B. Phosphorylation of I κ B α prevents its SUMOylation and primes it for degradation by the ubiquitin-proteasome system (Desterro et al. 1998). Whether phosphorylation of TDG has an effect on its modification by SUMO remains to be tested. However, phosphorylation may be involved in cell cycle regulation of TDG, i.e. by enhancing interactions with an E3 ubiquitin ligase (Gao and Karin 2005; Hardeland et al. 2007).

2.3. Epigenetic Regulation of Gene Expression

2.3.1. Chromatin Structure and Dynamics

In the nucleus of eukaryotic cells, genomic DNA is packaged into a protein-DNA complex termed chromatin. The basic building blocks of chromatin are the nucleosome core particles which consist of 147 base pairs (bp) of DNA wrapped in 1.7 superhelical turns around a histone octamer. These octamers consist of two H3-H4 and two H2A-H2B histone dimers. Nucleosome core particles are separated by linker DNA of 10-80 bp length associated with the linker histone H1, altogether forming a 10 nm fiber, resembling “beads on a string”, which is further compacted into a fibre of ~30 nm diameter (Felsenfeld and Groudine 2003; Hubner et al. 2012). Further compaction into higher-order structures involves long-range intra- and interchromosomal interactions mediated by chromatin associated structural organizer complexes such as the CCCTC-binding factor (CTCF), cohesins (Lee and Iyer 2012), and others. Chromosome Conformation Capture (3C) experiments have revealed that chromatin is packaged into fractal globules. All fractal globules of one chromosome appear to occupy a distinct area of the nucleus, a so-called chromosomal territory (Hubner et al. 2012).

The chromatin structure represents a second layer of genetic information “on top of” the DNA base sequence. This “epigenetic” information is shaped by developmental and environmental cues that instruct heritable patterns of chromatin structure to regulate gene expression in a cell type specific manner and thus determine cell fate. The sum of this “epigenetic” information is termed “epigenome”. Chromatin exists either in a condensed (heterochromatin) or in a more open, accessible form (euchromatin). The latter contains mostly transcriptionally active genes (but may

contain silent genes or heterochromatinized regions), whereas heterochromatin in its most condensed form is inaccessible for the large protein complexes involved in gene transcription. It thus contains mostly transcriptionally inactive genes and repetitive sequences (Hubner et al. 2012).

DNA replication, recombination, repair and activation or silencing of transcription, all involve alterations of the chromatin structure. This modulation can occur in three ways: First, by ATP-dependent chromatin remodeling complexes, second, through exchange of the core histones with other histone variants, third, through alterations of epigenetic modifications, namely histone modifications and DNA methylation (Felsenfeld and Groudine 2003).

For gene regulatory factors to be able to bind to their target sequences in the genome, their binding sites - which may lie deeply buried within a nucleosome - have to be made accessible without unraveling the overall chromatin architecture. This is accomplished by chromatin remodeling complexes that metabolize ATP to alter the DNA-histone interactions so that a nucleosome is relocated to a neighboring stretch of free DNA, transiently exposing the previously occupied sequence (Becker and Horz 2002).

The canonical histones can be replaced by histone variants to regulate transcriptional activation or silencing, chromatin structure, DNA repair and ES cell differentiation (Talbert and Henikoff 2010). For instance, a variant of H2A, H2AX, which is ubiquitously present throughout the genome, is phosphorylated in the course of DNA strand break repair, e.g. in the context of DNA damage or V(D)J recombination during antibody diversification (Fernandez-Capetillo et al. 2004). Furthermore, the histone variant H3.3 was found to be associated with active chromatin, whereas H3.2 correlates with a repressive chromatin state (Hake and Allis 2006).

2.3.2. Histone Modifications

The core histones consist of a globular domain and an amino-terminal tail, both of which were found to be subject to an ever-growing number of post-translational modifications, including phosphorylation, acetylation, methylation, mono- and poly-ADP-ribosylation, deimination, β -N-acetylglucosamine-modification, proline isomerization, SUMOylation and ubiquitylation. These modifications not only change the interactions between nucleosomes and between histones and DNA, but are also thought to form a "histone code" that is "read" by various protein complexes involved for instance in chromatin remodeling or transcription (Bernstein et al. 2007). Here, I will focus on histone acetylation and methylation, introduce the enzymes catalyzing these modifications (writers) and the factors that recognize them (readers and effectors). For an in-depth review on other histone modifications, refer to (Bannister and Kouzarides 2011).

Histone Acetylation

While the function of many of these modifications is still not fully understood, some basic principles have emerged. Acetylation of lysine residues almost always increases transcriptional activity and chromatin accessibility, presumably by neutralizing the charge interaction between the DNA backbone and the positively charged lysine residues of the histone tails (Bernstein et al. 2007; Bannister and Kouzarides 2011). Acetylation is highly dynamic and controlled by two families of enzymes with antagonistic functions: the histone acetyl-transferases (HATs) and the histone deacetylases (HDACs). HATs act as part of large multi-protein complexes (see also next sub-chapter) and can be subdivided into different groups with specific functions, type-A and type-B HATs. The latter is responsible for acetylating newly synthesized histones in the cytoplasm. Type-A HATs can modify histones that are already integrated into nucleosomes, and many members of this family are involved in transcriptional co-activation, e.g. CBP/p300. In contrast, HDACs are mostly transcriptional repressors as they restore the positive charge of a lysine through deacetylation, tightening the interaction between the histone tail and the DNA.

Like HATs, HDACs are present in several large protein complexes, often with other members of the HDAC family. HDAC1 and HDAC2, for instance, are subunits of the NuRD, Sin3a and Co-REST repressor complexes (Bannister and Kouzarides 2011).

Histone Methylation

Histone methylation does not alter the charge of the histone tail but rather functions as a recognition site for histone-code “readers”, proteins with a chromo-, bromo-, MBT, Tudor or PHD finger domain. Thus, methylation can have different effects, depending on which residues are modified and with how many methyl-groups. Methylation occurs mostly on lysine or arginine residues, in which lysine can be mono-, di- or tri-methylated and arginine mono- or di-methylated (symmetrically or asymmetrically). For instance, H3K4me3 marks promoters and transcriptional start sites of active and “poised” (to be activated) promoters, whereas H3K4me1 is associated with introns and distal regulatory elements termed enhancers (Black et al. 2012). H3K9 and H3K27 trimethylation on the other hand is generally associated with repression of transcription. The methyl-donor for both, lysine and arginine methylation is *S*-adenosylmethionine (SAM) (Bannister and Kouzarides 2011).

At the center of establishing and maintaining active or repressive chromatin states are trithorax (TrxG) and polycomb group (PcG) proteins, respectively. Both form large multi-protein complexes that contain both writers and readers of histone modifications, and that exhibits not only histone

methylating but also HAT or HDAC and other activities, reflecting the functional interlink between different histone-modifications.

TrxG proteins can be subdivided into three classes, one consisting of SET-domain containing factors that methylate histone tails (writers), a second including ATP-dependent chromatin remodeling factors that read the histone methylation established by the SET-domain proteins (readers and effectors), and a third class that binds directly to specific DNA sequences and serve multiple functions. Five PcG protein complexes have been identified to date. Polycomb repressive complex 1 (PRC1) and 2 (PRC2), for instance, establish repressive chromatin states, e.g. through their H3K9 methyltransferase, H3K4 and H3K36 demethylase and H3K27 methyltransferase activities (Lanzuolo and Orlando 2012).

All histone lysine methyltransferases (HKMTs) that modify histone tails have a common SET domain that harbors the enzymatic activity and catalyzes the transfer of a methyl-group from S-adenosylmethionine (SAM) to the ϵ -amino group of a target lysine. One of the first SET-domain TrxG proteins identified in mammals was MLL1, which catalyzes H3K4 trimethylation (following mono- and di-methylation) primarily at HOX genes. MLL1 was originally identified as a gene inducing human leukemia through aberrant fusion, e.g. with AF9 or TET1, caused by chromosomal rearrangements. (Schuettengruber et al. 2011). A prominent example for a SET-domain containing PcG protein is EZH2 (enhancer of zeste), the catalytic subunit of PRC2, which catalyzes di- and tri-methylation of H3K27. Interestingly, EZH2 also recognizes H3K27me₃, which makes it not only a writer but also a reader of this chromatin mark (Lanzuolo and Orlando 2012).

Lysine methylation is reversible through the action of a lysine-specific demethylase (LSD1), which can demethylate mono- and dimethylated lysine. Target specificity and biological role is conferred by the different protein complexes LSD1 is associated with. For instance, as a subunit of the Co-REST complex, LSD1 demethylates mono- or dimethylated H3K4, acting as a co-repressor, but in complex with the androgen receptor, it demethylates H3K9, functioning as a co-activator. Trimethylation of a histone lysine can be reversed by jumonji domain proteins, of which JMJD2 was the first trimethyl lysine demethylase identified. The histone demethylase JARID2 for example is a subunit of PRC2 (Marmorstein and Trievel 2009; Bannister and Kouzarides 2011).

Inheritance of Histone Modifications

How epigenetic information encoded by histone modifications is maintained during/after DNA replication remains poorly understood. Several models have been proposed over the years how histone modifications are inherited during mitotic divisions. The simplest, a semi-conservative inheritance mechanism, would rely on equal distribution of the two copies of each core histones to

the daughter strands upon DNA replication, and a symmetrical copying of the modification onto the newly deposited histones. However, it is unknown if histone modifications occur in a symmetric manner on both copies of each histone within a single nucleosome and whether nucleosomes are distributed semi-conservatively during DNA replication. Indeed, recent work questions such a mode of distribution of H3-H4, which carry the modifications most likely to be relevant for epigenetic phenomena (Zhu and Reinberg 2011).

The observation that many histone modifiers localize to the replication forks in S-phase has raised the hypothesis that factors within or associated with the DNA replication machinery recruit chromatin-modifying enzymes to restore the histone modifications on newly synthesized chromatin. Yet, this model cannot apply for certain histone modifications that appear to “mature” throughout the cell cycle (Zhu and Reinberg 2011).

It is clear that the inheritance of histone modifications requires some instructive pre-existing modifications to serve as templates for restoration of the chromatin states during or after DNA synthesis. Assuming a random distribution of nucleosomes to the nascent DNA strands, newly deposited histones might be modified analogous to a neighboring nucleosome, similar to a phenomenon termed “chromatin-modification spreading”. Such spreading of histone modifications has been found to occur with H4K16 acetylation, H3K9 and H3K27 methylation (Cockell et al. 1998; Grewal and Moazed 2003; Margueron et al. 2009).

Histone modifications are functionally interlinked with DNA cytosine methylation in the establishment of the epigenetic code, although the hierarchical order in which these modifications are written and read is not clear. While the mechanism of propagation of histone modifications is still under debate, the biochemistry of the DNA methylation system offers a logical and plausible concept for inheritance (see 2.3.3.).

2.3.3. DNA methylation

CpG Methylation System

CpG dinucleotides in mammalian cells are subject to methylation at the C5 position of cytosine. The occurrence of 5-mC is negatively correlated with the density of CpG dinucleotides, the CpG-richest sequences, also termed CpG islands (CGIs), showing the lowest methylation levels (Bird et al. 1985; Kafri et al. 1992; Illingworth and Bird 2009). Non-CpG methylation is common in plants but was found to be a characteristic exclusively of the pluripotent stem cell state in mammals and is lost with differentiation (Ramsahoye et al. 2000; Lister et al. 2009). In the human genome, over half of the

total amount of 5-mC is associated with repetitive sequences and contributes to genome stability by repressing recombination and the transcription of retrotransposons (Hu and Rosenfeld 2012).

Cytosine methylation is accomplished by three DNA methyltransferases, the maintenance DNA methyltransferase Dnmt1 and the *de novo* DNA methyltransferases Dnmt3a and b. Like histone methyltransferases, Dnmts use SAM as a methyl-group donor. Deletion of *Dnmt1* or *Dnmt3b* in mice is embryonic lethal and *Dnmt3a* knockout mice die about one month after birth (Li et al. 1992; Okano et al. 1999). Cultured triple knockout (*Dnmt1*^{-/-}*Dnmt3a*^{-/-}*Dnmt3b*^{-/-}) ES cells lose the ability to differentiate (Tsumura et al. 2006).

The *de novo* DNA methyltransferases Dnmt3a and Dnmt3b – in concert with the catalytically inactive cofactor Dnmt3L – establish methylation patterns in early development and contribute to the maintenance of these patterns through cell divisions (Jones and Liang 2009). How Dnmt3a and b are recruited to their target sites is not entirely clear although several reports suggest that they may be directed by histone modifications or associated factors (Fuks et al. 2003; Vire et al. 2006; Ooi et al. 2007; Dong et al. 2008). Faithful transmission of methylation patterns through mitotic divisions is conferred by Dnmt1 which is targeted to hemimethylated CpGs by Np95, e.g. following DNA replication, to restore the fully methylated state (Sharif et al. 2007). As DNA methylation and histone modifications are functionally interlinked, the Dnmt1-mediated inheritance of DNA methylation patterns might also support the faithful transmission of histone marks by providing cues for their restoration.

DNA Methylation in Transcriptional Regulation

DNA methylation plays a crucial role in the control of gene expression, e.g. in the allele-specific expression of imprinted genes and in establishing cell type-specific expression patterns during cell fate determination. Depending on its localization, DNA methylation can have different effects on transcriptional activity. Enrichment of 5-mC in close proximity to a transcription start site (TSS) inhibits transcription initiation but DNA methylation in the gene body has been found to even stimulate transcriptional elongation and may control splicing (Laurent et al. 2010; Jones 2012). The effect of DNA methylation at enhancer regions on gene expression is only beginning to become clear. Enhancers are mostly CpG poor regulatory sites located at variable distances from gene promoters and marked by transcription factor binding. Genome-wide studies of DNA methylation in mouse ES cells and neuronal precursors have associated transcription factor binding with the CpG methylation level of enhancers (Stadler et al. 2011). As a single CpG in a single cell can only be either methylated or unmethylated, the intermediate methylation levels observed at these regions possibly reflect a highly dynamic methylation state, shaped by competing DNA methylation and demethylation which

averages to intermediate levels in a cell-population. Indeed, these low methylated regions (LMRs) have been found to be associated with 5-hmC, a potential intermediate of DNA demethylation, and generation of 5-hmC appears to facilitate enhancer activation (Stadler et al. 2011; Serandour et al. 2012).

CpG Islands

CpG islands (CGIs) represent stretches of relatively high CpG density (observed/expected ratio >0.65) in the otherwise CpG-depleted genomes of all organisms with DNA methylation. CGIs overlap with the promoters of all ubiquitously expressed genes and also about 40% of those expressed in a tissue-specific pattern (Illingworth and Bird 2009). CGIs are maintained in a hypomethylated state which is thought to have protected these genomic regions from deamination-induced 5-mCpG→TpG mutation. However, the mechanisms conferring this immunity of CGIs to *de novo* methylation remain largely unknown. DNA sequence itself has been proposed to make CGIs refractory to the activity or binding of Dnmt3a and b (Fig. 2-04 A). Such a scenario is unlikely considering the high density of CpGs, which are the preferred substrate of Dnmts (Ramsahoye et al. 2000). Alternatively, the association of transcription factors might sterically block the access of Dnmts to CGIs (Fig. 2-04 C). Several lines of evidence support this hypothesis, e.g. the fact that the promoters of all constitutively expressed genes and 93% of those expressed during mouse embryogenesis contain CGIs (Ponger et al. 2001; Illingworth and Bird 2009). Moreover, the presence of the active histone mark H3K4me3 has been shown to inhibit binding of Dnmt3L (Fig. 2-04 C) (Ooi et al. 2007). H3K4 methyltransferase complexes are targeted to CGIs by the CXXC domain, conferring binding to unmethylated CpGs, of factors like Cfp1 or MLL1 (Clouaire et al. 2012).

Finally, transcription itself and the formation of R-loops (RNA-DNA hybrid structures) have been proposed to maintain hypomethylation at CGIs immediately downstream of a TSS (Ginno et al. 2012). A third possibility how CGIs might be protected from methylation is a targeted proof-reading complex that removes aberrant 5-mC (Fig.2-04 B). First evidence for such a proof-reading activity came with the finding that *in vitro* methylated CGIs become demethylated when introduced into ES cells (Frank et al. 1991). The discovery of the TET proteins as 5-mC hydroxylases has shed new light on possible mechanisms involved in CGI methylation proof-reading (Tahiliani et al. 2009), especially since TET1 localizes to CGIs (Williams et al. 2012). However, such a proof-reading complex might not be the first and foremost mechanism to keep CpG islands unmethylated but might rather function as a caretaker to correct aberrant methylation caused by the failure of the two other – or as yet unknown – mechanisms (see next chapter).

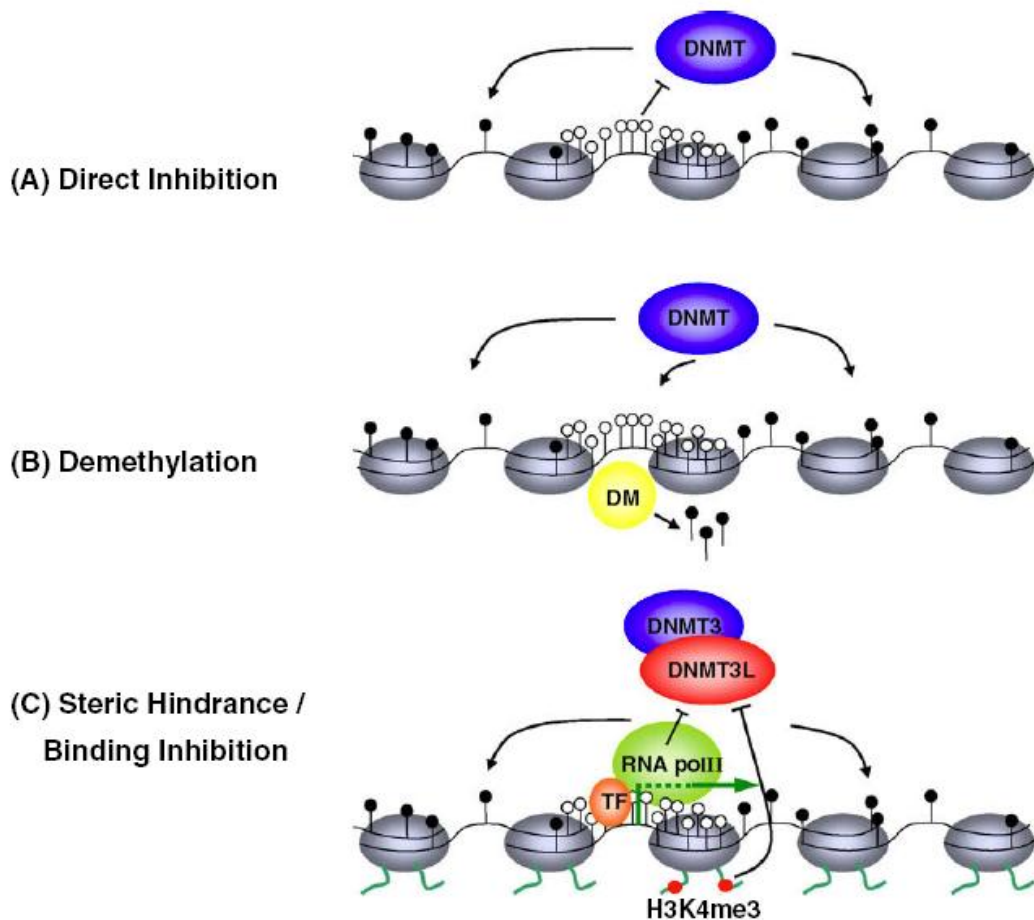


Fig. 2-04: Possible mechanisms maintaining hypomethylation at CGIs. (A) Intrinsic sequence properties might keep CpGs in CGIs unmethylated (white lollipops) by inhibiting activity or binding of DNA methyltransferases (DNMT). (B) CGIs might be subject to methylation but a targeted DNA demethylase (DM) complex might remove 5-mC (black lollipops) from these genomic loci. (C) Occupation by transcription factors (TF) and RNA polymerase II (RNA polII) might sterically block the access of DNMTs to CGIs. Also, the active histone mark H3K4me3 was shown to inhibit binding of the DNMT cofactor DNMT3L. (Illingworth and Bird 2009)

CpG Methylation Instability

Carcinogenesis and ageing are accompanied by a disruption of DNA methylation patterns. In general, cancer cells exhibit hypomethylation of the bulk genome punctuated by hypermethylation at CGIs. Hypomethylation occurs mostly in CpG poor regions and contributes to tumorigenesis through aberrant gene activation, loss of transposon and recombination repression and destabilization of the chromatin structure e.g. at telomeres (Schar and Fritsch 2011; Gokul and Khosla 2012). CGI hypermethylation at the promoters of tumor suppressor genes resulting in transcriptional downregulation is one of the hallmarks of carcinogenesis. (Hanahan and Weinberg 2000; Gronbaek

et al. 2007). Causes for aberrant DNA methylation can be mutations in or misregulation of the *Dnmt* genes, false targeting of the Dnmts or an imbalance of the methyl-group donor SAM (Schar and Fritsch 2011). Still, whether DNA methylation errors are the cause or the consequence of carcinogenesis remains unclear.

Yet, the fact that the DNA methylation machinery can erroneously methylate regions that should remain unmethylated makes it clear that proof-reading mechanisms must have evolved for this epigenetic mark (Schar and Fritsch 2011). Also, the erasure of DNA methylation, e.g. in order to reset the epigenome to a totipotent state (see chapter 2.3.4.), illustrates the need for reversibility of DNA methylation. In plants, 5-mC demethylation is accomplished by 5-mC glycosylases but mammals appear to employ more complex mechanisms (see chapter 2.4.) (Zhu 2009).

2.3.4. Epigenome dynamics in mammalian development

Multicellular organisms develop out of a single cell, the zygote. After fertilization, the epigenetic marks of the paternal and maternal genomes, present as distinct structures in the zygote, have to be erased to restore totipotency, the ability to generate any cell type. Newly established epigenetic programs determine cell identity and provide a heritable “cellular memory”. Although histone modifications play a crucial role in establishing the gene expression profiles that confer cell identity, this chapter focuses mostly on DNA methylation dynamics.

Upon entering the oocyte, the densely packed and transcriptionally silent paternal chromatin undergoes extensive remodeling to become transcriptionally inducible. This includes replacement of protamines, sperm-specific histone-analogs, with canonical histones provided by the oocyte, as well as DNA decondensation, which results in the formation of the paternal pronucleus (Jenkins and Carrell 2012). The paternal pronucleus undergoes rapid, global DNA demethylation, with the exception of imprinted loci, IAP retrotransposons and centromeres. As this global erasure of 5-mC precedes and coincides with the first round of DNA replication, it is thought to be an active process (see chapter 2.4.2.). The maternal pronucleus escapes this global DNA demethylation event through mechanisms that are as of yet not entirely clear, although an involvement of repressive histone modifications and the maternal factor Stella has been suggested (Nakamura et al. 2007; Wossidlo et al. 2011; Szabo and Pfeifer 2012). Demethylation of the maternal DNA is gradually achieved through several cell divisions upon reactivation of the oocyte’s cell cycle, consistent with a passive process by inhibition of maintenance methylation (Jenkins and Carrell 2012). The erasure of the majority of this epigenetic mark allows for the establishment of new, cell type-specific DNA methylation profiles.

A first divergence in the epigenetic profiles of the zygote-derived cells becomes evident at the morula stage, where cells in the periphery are primed to become extra-embryonic tissue, whereas cells at the center eventually give rise to the embryo proper (Johnson and Ziomek 1981). In the blastocyst, the peripheral trophectoderm cells reach the lowest levels of 5-mC, while in the inner cell mass DNA methylation has already begun to be restored by *de novo* methyltransferases (DNMT3). One of the first genes silenced through DNA methylation in the epiblast, which determines lineage trajectory, is *Elf5*, a trophectoderm-specific transcription factor (Seisenberger et al. 2013). During and following implantation of the mouse embryo, about 480 gene promoters become methylated and thus transcriptionally silenced, including a set of CGI promoters. Among them are many pluripotency factors, i.e. *Rex1*, but also germline-specific genes and tissue-specific factors that are re-expressed in later stages of development, e.g. hematopoietic genes (discussed below) (Borgel et al. 2010).

Primordial Germ Cells (PGCs), the precursors of the gametes, inherit a significant amount of DNA methylation from the epiblast they are derived from, including imprinted loci inherited from the parental generation. To reset these somatic methylation profiles to that of germ cells that are capable of forming a totipotent zygote, PGCs undergo global DNA demethylation that spares only the retrotransposons with the highest mutagenic potential. Male or female gamete-specific methylation profiles are established gradually until the prospermatogonia stage or in the growing oocyte, respectively. New imprints are established according to the embryo's gender (Seisenberger et al. 2013).

A subset of tissue-specific genes was found to be silenced in early development but reactivated in later stages. Among these are brain-, eye- or hematopoiesis-specific genes, i.e. *Mbp*, *Cryaa*, *Cplx4*, *Obf1*, *Tlr6* and *Cytip*. The promoters of these genes display intermediate CpG-density and were found to acquire DNA methylation in early embryogenesis. Yet, the promoters of the eye-specific genes *Cryaa* and *Cplx4* become demethylated during eye-development, as are those of *Tlr6* and *Cytip* in the hematopoietic lineage and that of *Obf1* during adult B-cell differentiation (Borgel et al. 2010). This targeted erasure of 5-mC is likely to employ different mechanisms than the global DNA demethylation observed in zygotes and PGCs (discussed in chapter 2.4.).

2.4. DNA demethylation

2.4.1. Concepts of DNA demethylation

The fact that plants have evolved enzymes with 5-mC glycosylase activity and appear to achieve DNA demethylation through BER has fueled the search for analogous enzymes in mammals. Passive

removal of 5-mC has long been established to occur through inhibition of DNA methylation maintenance by Dnmt1, e.g. to gradually achieve global demethylation of the maternal pronucleus. However, the rapid loss of 5-mC in the paternal pronucleus and in PGCs suggests the involvement of an enzymatic component.

Indeed, the discovery of the TET proteins as 5-mC hydroxylases has shed light on the mechanism of global DNA demethylation (Tahiliani et al. 2009; Ito et al. 2010). Recent studies have revealed that TET3 mediated conversion of 5-mC to 5-hmC and subsequent dilution by DNA replication causes a global loss of 5-mC in the paternal pronucleus (Gu et al. 2011; Inoue and Zhang 2011; Wossidlo et al. 2011). Furthermore, global demethylation of primordial germ cells has been demonstrated to be driven by TET-catalyzed conversion of 5-mC to 5-hmC, followed by a decline of 5-hmC levels consistent with replication-mediated dilution (Hackett et al. 2013).

Such a semi-active process of DNA demethylation, converting 5-mC to 5-hmC which is no longer maintained by Dnmt1 (Valinluck and Sowers 2007), appears perfectly suitable for global erasure of 5-mC. Alternative pathways that were proposed for global demethylation entail a DNA repair step (see chapter 2.4.2.), which would generate potentially hazardous AP-sites and DNA strand-breaks at an alarming number. Still, the BER factor APE1 and BER facilitator PARP1 were shown to be essential for global DNA demethylation in PGCs (Hajkova et al. 2010). How BER contributes to these global events remains to be elucidated. However, replication-independent pathways for enzymatic DNA demethylation may well be employed in controlling DNA methylation at specific sites.

Direct removal of the methyl-group has been proposed as a most straightforward way to actively convert 5-mC back to C. The 5-mC binding domain protein MBD2 has been reported to exhibit such activity but this finding could not be reproduced by other groups so far (Bhattacharya et al. 1999). Also, *Mbd2* knockout mice exhibit normal 5-mC levels and paternal pronucleus demethylation, shedding doubt over a biologically relevant MBD2-mediated DNA demethylation process (Hendrich et al. 2001; Santos et al. 2002).

An alternative to the removal of the methyl-group is excision of 5-mC through DNA repair. Two major pathways have been considered in such a scenario: nucleotide excision repair (NER), which would remove a stretch of DNA that contains the methylated CpG, and BER, which would remove 5-mC directly. The unmethylated CpG would be restored by repair synthesis in both cases.

The NER machinery has been implicated in transcriptional regulation in absence of DNA damage. NER factors are recruited to active gene promoters and contribute to chromatin remodeling for efficient transcription initiation (Schmitz et al. 2009; Le May et al. 2010). The NER endonucleases XPG and XPF were shown to be essential for controlling DNA demethylation at promoter and transcriptional termination sites although the mechanistic connection between strand incision by XPG or XPF and DNA demethylation remains unclear (Le May et al. 2012). As NER mostly recognizes and corrects

helix-distorting lesions, the cue triggering NER action in the context of DNA demethylation has yet to be identified. One possibility is the occurrence of AP-sites which have been shown to be repaired by NER under certain circumstances (Lin and Sancar 1989; Huang et al. 1994; Kim and Jinks-Robertson 2010).

The search for a mammalian 5-mC DNA glycosylase initiating DNA demethylation by BER has revealed an enzyme capable of excising 5-mC from hemimethylated DNA in chicken embryo extracts. This enzyme later turned out to be the chicken homolog of TDG, in complex with RNA and an RNA helicase (Jost 1993; Jost et al. 1995; Fremont et al. 1997; Jost et al. 1997; Jost et al. 1999; Zhu et al. 2000). Yet, the glycosylase activity of TDG on 5-mC could not be corroborated so far with recombinant enzymes. A more plausible pathway, which is supported by recent reports, is the conversion of 5-mC to a more suitable substrate for DNA glycosylases (discussed in chapter 2.4.2).

2.4.2. Current models for BER-mediated active DNA demethylation

As 5-mC is more susceptible to deamination than C and as several DNA glycosylases evolved to counteract such spontaneous deamination events, enzymatic deamination has long been considered as a potential first step of DNA demethylation, followed by BER. Candidate 5-mC deaminases in such a pathway include the activation induced deaminase (AID) and the apolipoprotein B mRNA-editing enzyme complex (APOBEC) family of proteins as well as Dnmt3a and b.

AID was shown to contribute to demethylation of the *Oct4* and *Nanog* gene promoters during somatic cell reprogramming (Bhutani et al. 2010). AID has also been implicated in global DNA demethylation in PGCs, as AID-deficient mouse PGCs exhibit threefold higher methylation levels at embryonic day 13.5 as their wildtype counterparts (Popp et al. 2010). Furthermore, a deamination-coupled DNA demethylation pathway was also observed in zebrafish embryos, involving deamination of 5-mC by AID and subsequent excision by MBD4 (Rai et al. 2008).

Dnmt3a and b have also been shown to deaminate 5-mC *in vitro* under conditions of SAM depletion. Whether such an environment is generated *in vivo* is unclear, but transient recruitment of the Dnmt3s and TDG has been shown to be associated with cyclical methylation and demethylation at estrogen-responsive promoters. Interestingly, reminiscent of the reports by J.P. Jost and colleagues, this putative DNA demethylation complex also includes the RNA helicase p68, suggesting the complex to be targeted and/or stabilized by an RNA component (Kangaspeska et al. 2008; Metivier et al. 2008).

The discovery of the TET proteins as 5-mC hydroxylases shed new light on a possible BER-coupled DNA demethylation pathway (Tahiliani et al. 2009; Ito et al. 2010). Interestingly, a 5-hmC glycosylase

activity had been identified in calf thymus long ago (Cannon et al. 1988), alas the responsible enzyme has never been identified and none of the mammalian DNA glycosylases characterized appears to be capable of excising 5-hmC on its own (He et al. 2011; Maiti and Drohat 2011). More recently, it was suggested that the concerted action of TET proteins and AID or APOBEC proteins facilitate 5-mC demethylation through subsequent generation of 5-hmC and deamination to 5-hydroxymethyluracil (5-hmU), followed by excision of 5-hmU by either TDG or SMUG1 (Boorstein et al. 2001; Hardeland et al. 2003). One study implicated TET1 and Apobec1-catalyzed deamination of 5-hmC in neuronal activity-induced site-specific active DNA demethylation in the adult mouse brain (Guo et al. 2011). Also, AID was reported to interact directly with TDG in a coupled deamination-BER pathway of DNA demethylation (Cortellino et al. 2011). However, a recent study of the enzymatic properties of AID and APOBEC proteins revealed that 5-hmC is no suitable substrate for deamination by these enzymes, whereas a residual activity on 5-mC was confirmed (Nabel et al. 2012).

Further insight into a potential mechanism of active DNA demethylation came with the finding that TET proteins can oxidize 5-hmC further to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), both of which are readily excised by TDG but – unlike the deamination products of 5-mC and 5-hmC – apparently by no other DNA glycosylase (He et al. 2011; Ito et al. 2011; Maiti and Drohat 2011). Given the fact that TDG is also the only DNA glycosylase known so far with a developmental phenotype (Cortazar et al. 2011; Cortellino et al. 2011; Jacobs and Schar 2012), its non-redundant 5-fC/5-caC glycosylase activity might hint at an essential role of TDG in DNA demethylation during development. However, the significance of such a TET-TDG mediated pathway during development has remained elusive.

3 Aims of the Thesis

TDG has been implicated in two seemingly disconnected biological processes, DNA repair and transcriptional regulation. Both functions have been shown to require SUMOylation or SUMO-binding (Hardeland et al. 2002; Steinacher and Schar 2005; Mohan et al. 2007). The regulation of TDG function by SUMOylation has been studied extensively in biochemical assays and we previously proposed that SUMOylation induces dissociation of TDG from its product AP-site (Steinacher and Schar 2005). However, an involvement of SUMOylation in TDG-dependent BER had not been corroborated by *in vivo* data.

The first goal of my PhD thesis was to generate a tool to monitor the SUMO-TDG interaction in living cells, unchallenged and in presence of DNA damaging agents. To this end, I established a Fluorescence Resonance Energy Transfer (FRET) system, which we used to characterize the covalent and non-covalent SUMO1- and SUMO3-binding of TDG.

In the attempt to reconcile its functions in DNA repair and in the regulation of gene expression, TDG was proposed to actively demethylate DNA by directly excising 5-mC (Zhu et al. 2000). As this model was not corroborated by independent experimental evidence, the involvement of TDG in active DNA demethylation has remained elusive. Still, we found deletion of *Tdg* in mouse to cause embryonic lethality and epigenetic aberrations at CGIs in *Tdg* knockout cells. These aberrations suggested a dual role of TDG in safeguarding CGIs, one in coordinating the maintenance of active histone marks and one in counteracting aberrant DNA methylation. The underlying mechanisms, however, have remained obscure.

The second goal of my PhD thesis was to investigate the mechanism by which TDG contributes to the epigenetic stability of CGIs. To this end, I – in collaboration with others – dissected the causes for aberrant methylation levels in *Tdg* knockout cells in the course of cell differentiation to address the involvement of BER, deamination and/or oxidation in controlling CGI methylation.

The study of factors involved in maintaining stable chromatin states raises the need for tools to investigate the rates of epigenetic change (epimutagenesis). However, unlike for genetic change (mutagenesis), no such tools have been designed nor developed thus far.

The third goal was thus to establish a system that allows quantitative monitoring of the epigenetic stability of promoters. To this end, I established an ES cell-based system that allows a timed release of an exchangeable promoter of interest from selective pressure and several read-out options to monitor promoter activity, applicable in high-throughput screening approaches as well as in studying stochastic effects on a clonal basis.

4 Results

The following section summarizes the results presented in the manuscripts provided in the appendix as well as supplementary results not included in the manuscripts.

4.1. Measuring the SUMO-Interaction Dynamics of the Thymine DNA Glycosylase by Fluorescence Resonance Energy Transfer (Appendix I)

Post-translational modification with Small Ubiquitin-like Modifiers (SUMOs) can affect the subcellular localization, enzymatic activity, three-dimensional structure or protein-interactions of a target protein and is thus involved in regulating and coordinating various cellular processes, e.g. in DNA replication and repair but also in pathways connected to development (Gill 2004; Lomeli and Vazquez 2011). SUMO is also bound non-covalently by proteins with a SUMO Interaction/Binding Motif (SIM/SBM). We have previously reported that SUMOylation of TDG facilitates the dissociation of the glycosylase from the AP-site by abolishing the non-specific DNA-interaction of the N-terminal domain (Hardeland et al. 2002; Steinacher and Schar 2005). TDG also interacts non-covalently with SUMO (Hardeland et al. 2002), via two SIM/SBM motifs located in the N-terminal and the C-terminal domains (Mohan et al. 2007)

In this part of my PhD thesis, I present the development of a Fluorescence Resonance Energy Transfer (FRET) system to monitor the SUMO1- and SUMO3-interactions of TDG in living cells. FRET is a physical process involving the transfer of an energy quantum between a donor and an acceptor fluorophore, in our case Cerulean and Citrine, respectively. This energy transfer requires an overlap of the emission spectrum of the donor with the absorption spectrum of the acceptor, so that an energy quantum emitted by the donor in an excited state can be absorbed by the acceptor fluorophore, which subsequently emits a photon in a wavelength beyond that of the donor. The transfer can only occur if the donor and acceptor are in a suitable orientation and in close proximity (10 to 80 Å) of one another. Thus, the interaction between two proteins of interest can be monitored by fusing one to the donor fluorophore and one to the acceptor (Siegel et al. 2000).

We generated and validated a series of constructs encoding wildtype TDG as well as a SUMOylation deficient mutant (K341R, TDG Δ S) and a catalytically dead mutant (TDG Δ cat) fused N- or C-terminally to the FRET acceptor Citrine. The donor Cerulean was N-terminally fused to SUMO1 and SUMO3. We could confirm full-length expression and catalytic activity of the TDG fusion proteins, although the N-terminal tag appears to interfere with efficient expression and the non-specific DNA interactions of

the N-terminal domain (Appendix I, Fig.2b, 2c and 3d). Additionally, we proved that the fluorophore tags on both, TDG and SUMO, do not interfere with the SUMO conjugation pathway or with their respective subcellular localization (Appendix I, Fig. 2d and Fig. 4).

We provide proof that the SUMO1- and SUMO3-interactions of TDG can be monitored in living cells and that these interactions occur at a surprisingly high steady state in the absence of induced DNA damage (Appendix I, Fig. 4 and 5). Including the SUMOylation-deficient mutant TDG Δ S in our system enabled us to distinguish between covalent and non-covalent SUMO-interactions. We found a significantly higher FRET signal for the SUMO3-interaction of wildtype TDG than for TDG Δ S, suggesting that a significant part of the cellular TDG pool is covalently attached to SUMO3 in the presence of endogenous levels of DNA damage (Appendix I, Fig.5b). On the other hand, it appears that SUMO1 is mostly associated with TDG through its SBMs rather than through a covalent bond as we do not find a significant difference between the FRET signal produced by the SUMO1-interaction of wildtype TDG and TDG Δ S (Appendix I, Fig. 5a).

To validate our system, we tested whether excess DNA damage alters the TDG-SUMO-interaction dynamics. By treating cos7 cells transiently expressing the FRET constructs with the base analog 5-FU which is incorporated into the DNA and represents a substrate for TDG-mediated repair, we were able to shift the balance between SUMO-binding and SUMO-modification of TDG. Interestingly, the treatment only affected the interaction with SUMO1 and not with SUMO3 (Fig. 5). TDG Δ S appears to lose the interaction with SUMO1 upon treatment with 5-FU, which is in agreement with previous reports of SUMO- and DNA-binding of recombinant TDG being mutually exclusive (Smet-Nocca et al. 2011). We observe a small increase of the SUMO1 interaction of wildtype TDG upon induction of DNA damage. As the FRET signal derived from the SUMO1-interactions of wildtype TDG is a mix of covalent and non-covalent SUMO-binding, and as the non-covalent interaction is reduced upon 5-FU treatment, we conclude that modification of TDG by SUMO1 is in fact stimulated by DNA damage (Appendix I, Fig. 5a). This observation supports a role of SUMOylation in regulating TDG function during DNA repair.

SUMO3, on the other hand, appears to be covalently attached to part of the cellular TDG pool even in absence of DNA damage. Upon 5-FU treatment, the association of SUMO3 with TDG through its SBMs appears to be stimulated, evident in a mildly increased FRET signal derived from the SUMO3-interactions of TDG Δ S (Appendix I, Fig.5b). As the overall interaction of wildtype TDG with SUMO3 is reduced, covalent SUMO3 binding appears to be dissolved upon induction of DNA damage.

Taken together, it appears that the interaction with SUMO1 and SUMO3 regulate TDG function in different contexts. As TDG is involved not only in canonical DNA repair but also in the regulation of gene expression, these two different processes might employ different SUMO family members. SUMOylation of TDG in response to exogenous DNA damage appears to involve mostly SUMO1,

whereas SUMO3 might control TDG function in the context of transcriptional regulation by mediating protein-protein interactions with transcription factors or histone modifying enzymes. SUMO1 was shown to mostly interact with the C-terminal SBM of TDG (Smet-Nocca et al. 2011), raising the question whether the N-terminal SBM serves a distinct function and mediates interactions within a different set of complexes, possibly involving SUMO3. Our FRET system provides a powerful tool for elucidating the different roles of SUMO1- and SUMO3 interaction of TDG and the conditions affecting these interaction dynamics.

We further designed and generated a FRET system to monitor the conformational switch of TDG between an open and a closed form by fusing the FRET donor Cerulean to the C-terminus and the acceptor Citrine to the N-terminus of TDG. We found that attaching fluorophore tags to both termini drastically increased the relative catalytic activity of TDG on a G•T substrate (Appendix 1 Fig. 3d). Therefore, double-tagged TDG might not be able to efficiently switch to or remain in a closed conformation. Still, pilot experiments revealed a robust FRET signal. As some biochemical and structural evidence suggests the formation of TDG homodimers ((Maiti et al. 2008; Morgan et al. 2011) and Roland Steinacher, personal communication), we included dimerization controls: single-tagged TDG fused N- or C-terminally to either Cerulean or Citrine. We found FRET to occur between single-tagged TDG molecules, although it remains unclear in which conformation, parallel or anti-parallel, this interaction occurs. N-terminally Citrine- and C-terminally Cerulean-tagged TDG produced a robust FRET signal which was significantly enhanced when both tags were fused to the C-terminus (Appendix 1, Fig.6). The FRET efficiency was potentiated further with double-tagged TDG, which is probably mostly due to doubling the number of fluorophores between which the energy transfer can occur, although we cannot exclude that part of the FRET signal derives from an intramolecular energy transfer. Furthermore, we found a significantly higher FRET signal produced by double-tagged TDG ΔS compared to wildtype TDG, suggesting that SUMO-modification interferes with dimerization. It remains to be tested whether this interaction involves only two TDG molecules or is in fact a polymerization, and whether it is mediated by DNA and/or SUMO-binding.

Dimerization of TDG on the DNA in the course of DNA repair might serve the purpose of inhibiting BER or NER of any damage occurring in close proximity of the produced AP-site on the opposite strand. The generation of two opposing AP-sites or the degradation of a whole stretch of DNA opposite an AP-site might result in double-strand breaks. It is likely that mechanisms evolved to avoid such a clash of two DNA repair processes. Such a function of TDG dimerization becomes especially appealing with regard to DNA demethylation in which TDG has been implicated. 5-mC occurs mostly symmetrically in a palindromic CpG dinucleotide and a BER-coupled DNA demethylation process would have to be restricted to one strand at a time to avoid double-strand breaks (see also Appendix III).

In conclusion, we have generated a FRET system that provides a powerful tool to elucidate the role of the SUMO-TDG-interaction in DNA repair and other contexts. Furthermore, we provide the first evidence of TDG dimerization in living cells, which may have a function in ensuring strand specificity during DNA demethylation processes.

Contribution:

I designed, generated, validated and characterized all FRET constructs used in this study. I cultured and transfected cos7 cells, performed protein extraction from cos7 cells, conducted Western blot analysis, base release assay, microscopy and image analysis, and wrote the manuscript in Appendix I.

4.2. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability (Appendix II)

The interaction partners of TDG are various, ranging from DNA repair factors to transcription factors and chromatin modifiers. Thus, TDG has been implicated not only in DNA repair but also in the regulation of gene expression. To elucidate the biological function of TDG, we generated *Tdg*^{-/-} mice and found TDG to be essential for embryonic development, unlike any other mammalian DNA glycosylase. TDG knockout embryos display internal haemorrhage by E10.5 and die around E11.5 (Appendix II, Fig. 1a). To investigate whether loss of the DNA repair function of TDG causes embryonic lethality, we tested the sensitivity of wildtype and *Tdg*^{-/-} immortalized MEFs to ionizing radiation and H₂O₂ but found that TDG deficiency had no effect on cell survival. Furthermore, a standard mutator assay revealed no increase in the frequency of spontaneous mutations (Appendix II, Supplementary Fig. 2). The function of TDG in canonical DNA repair, if any, is likely redundant and compensated for by other DNA glycosylases with an overlapping substrate spectrum.

We further investigated whether deletion of *Tdg* affects the regulation of gene expression in MEFs. Expression analysis revealed 461 misregulated genes in the *Tdg* knockout of which many encode transcription factors, thus suggesting direct as well as indirect gene expression effects of the TDG loss. The gene networks most affected were connected to embryogenesis and development (Appendix II, Supplementary Fig. 3a).

As TDG has been implicated in active DNA demethylation, we analyzed DNA methylation at the CGI containing promoters of genes downregulated in *Tdg* knockout MEFs. CGI are normally maintained in a hypomethylated state, but sequencing of Na-bisulphite converted DNA isolated from wildtype and

Tdg knockout MEFs revealed an accumulation of aberrant *de novo* DNA methylation in absence of TDG (Appendix II, Fig. 1c). We found that TDG is associated with these promoters by chromatin immunoprecipitation (ChIP) but found no enrichment of TDG on an intergenic control region or the transcriptionally silent promoters of *Oct4* and *Tuba3* (Appendix II, Fig. 1d). It thus appears that TDG is targeted to active gene promoters, possibly to protect them from acquiring aberrant *de novo* DNA methylation and transcriptional silencing.

Further analysis of the chromatin status of these gene promoters revealed a loss of di-methylation of lysine 4 on histone 3 (H3K4me2) which is associated with an active chromatin state. In parallel, we found these loci to gain repressive histone marks: active promoters which showed only H3K4me2 in a wildtype background (e.g. those of *Sfrp2* and *Twist2*) acquired H3K27 tri-methylation (H3K27me3), whereas promoters with a bivalent chromatin state (coinciding H3K4me2 and H3K27me3 e.g. at the promoters of *Hoxa10* and *Hoxd13*) acquired H3K9me3 (Appendix II, Fig. 1e). Stable complementation of *Tdg*^{-/-} MEFs with a vector encoding TDG rescued the chromatin state at the promoters and the expression of *Sfrp2* and *Twist2* (Appendix II, Fig. 2). However, we found that those promoters that lost H3K4me2 completely and gained H3K9me3 (e.g. those of *Hoxd13* and *Hoxa10*) could not be reversed to a bivalent chromatin state by re-introducing *Tdg* into knockout MEFs, suggesting that heterochromatinization of these promoters progressed beyond a point of reversibility in our experimental system.

To investigate the origin of the epigenetic instability we observe in absence of TDG, we profiled gene expression in *Tdg* positive and negative ES cells and after *in vitro* differentiation to neuronal progenitor cells (NPs). Interestingly, we found almost no differentially regulated genes in ES cells but a significant increase of transcriptional misregulation (297 genes) after differentiation to NPs (Fig. 3a). Analysis of gene networks most affected revealed a connection to developmental functions and many of the misregulated genes to have promoter CGIs and to be targets of the polycomb repressive system (Appendix II, Supplementary Fig. 7a). We could confirm enrichment of TDG at the promoter of misregulated genes in ES cells and NPs. Moreover, we found TDG associated with the promoters of *Oct4* and *Nanog* in ES cells but not in NPs (Appendix II, Fig. 3b) suggesting that TDG loses affinity to these promoters upon heterochromatinization, which may explain why TDG complementation failed to rescue the chromatin state of the heterochromatinized *Hoxd13* and *Hoxa10* promoters in MEFs.

Na-bisulphite sequencing revealed increased levels of DNA methylation at several promoters of genes downregulated in *Tdg*^{-/-} NPs but not in ES cells (Appendix II, Fig. 3c). Similarly, aberrant histone modifications only arose upon *in vitro* differentiation, evident in a gain of H3K27me3 at the promoters of *Hoxa10* and *Hoxd13* (Appendix II, Supplementary Fig. 8). Taken together, epigenetic aberrations in absence of TDG appear to arise only in differentiated cells, suggesting a function of TDG in establishing and maintaining epigenetic marks during cell differentiation.

To elucidate whether this function of TDG entails downstream BER, we tested for an association of key BER factors with these TDG-bound gene promoters. We found a TDG-dependent enrichment of XRCC1 and APE1 in MEFs at these genomic regions, whereas XRCC1 appears to bind independent of TDG in ES cells (Fig. 4a). Furthermore, we found an induction of chromatin associated XRCC1 foci in TDG positive but not in TDG-deficient ES cells that were differentiated for 8 h with retinoic acid (RA) (Appendix II, Supplementary Fig. 9). Additionally, *Tdg*^{-/-} cells exhibited reduced sensitivity to inhibition of poly(ADP-ribose) polymerase (PARP) upon differentiation (Appendix II, Supplementary Fig. 10), suggesting that differentiation in the presence of TDG induces the formation of single-strand DNA breaks the further processing of which requires XRCC1 and PARP.

Investigating the origin of the imbalance of histone modifications at these promoters, we found that the histone acetyltransferase CBP and the H3K4 specific methyltransferase MLL1 are enriched at the promoters of genes downregulated in TDG knockout MEFs, and that this association depends on the presence of TDG in MEFs but not in ES cells (Appendix II, Fig. 4b). Interestingly, CBP/p300 has been shown to interact with TDG and to protect promoters from H3K27 tri-methylation (Tini et al. 2002; Pasini et al. 2010).

In conclusion, our results suggest a dual role of TDG in maintaining an active chromatin state at CGI promoters during differentiation: first by structurally coordinating the association of chromatin modifiers responsible for establishing and maintaining active histone marks (e.g. MLL1 and CBP/p300), second by enzymatically contributing to protection of these CGIs against aberrant methylation (Appendix II, Fig. 4c). The induction of DNA single-strand breaks in response to RA differentiation in addition to the association of XRCC1 and APE1 with CGI promoters in a TDG-dependent manner suggests that this latter function entails BER, hinting at a repair mediated DNA demethylation process contributing to epigenome stability during cell differentiation.

Contribution: I established immunocytochemical staining of XRCC1 and PCNA (Supplementary Fig. 9a), performed imaging of ES cells at 0 and 8 h of RA differentiation and visual assessment of XRCC1 foci number in S-phase and non-S-phase cells in 5 independent, blinded experiments (Supplementary Fig. 9b). I further assessed XRCC1-foci number in H₂O₂ treated and mock treated ES cells (Supplementary Fig. 9c) and confirmed equal cell cycle profiles of TDG positive and negative ES cells after 8 h of RA differentiation (data not shown). Moreover, I established immunostaining of H3K4me₂, H3K27me₃ and H3K9me₃, which I performed on ES cells differentiated for 0, 48 and 96 h with RA in replicates. The global patterns of these chromatin modifications did not differ between TDG-proficient and -deficient cells within this timeframe of differentiation (data not shown).

For all experiments involving ES cells, I prepared Feeder cells, conditioned the ES cells for an undifferentiated state for 3 passages on Feeders and prepared frozen stocks. During the

establishment of Na-bisulphite sequencing, I prepared genomic DNA from ES cells at 0, 48 and 96 h of RA differentiation and conducted bisulphite conversion. PCR-amplified DNA fragments were cloned into bacterial vectors and transformed into *E.coli*. I prepared plasmid DNA from 117 bacterial clones for commercial sequencing. I further prepared total RNA from ES cells differentiated for 48 h with RA for gene expression analyses by qRT-PCR (Supplementary Fig. 6c). Finally, in the attempt to investigate the mechanism underlying TDG catalyzed DNA demethylation at CGIs, I generated DNA substrates for base release assays to test a potential 5-hmC glycosylase activity. We found no activity of TDG on 5-hmC (data not shown) which has by now been confirmed by other studies. The conversion of 5-hmC to 5-fC and 5-caC and TDG's glycosylase activity on these bases was reported later (He et al. 2011; Ito et al. 2011; Maiti and Drohat 2011).

4.3. TDG maintains a transitory equilibrium of CpG island methylation and oxidative demethylation during cell differentiation (Appendix III)

Our previous efforts to elucidate the biological function of TDG in embryo development has implicated TDG in safeguarding CGIs during cell differentiation by coordinating histone modifiers responsible for an active chromatin state as well as by counteracting aberrant DNA methylation. Recently, TDG has been implicated in a putative DNA demethylation pathway involving the Ten Eleven Translocator (TET) family of proteins. TET1-3 convert 5-mC to 5-hydroxymethylcytosine (Tahiliani et al. 2009) and further to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), the latter two representing excellent substrates for TDG (He et al. 2011; Maiti and Drohat 2011).

To investigate a function of TDG in DNA methylation control, we performed MeDIP in combination with next generation sequencing on DNA from TDG positive and negative ES cells, *in vitro* differentiated neuronal progenitors (NPs) and MEFs derived from litter-matched wildtype and TDG knockout embryos around E10.5. While we found no differentially methylated regions (DMRs) in ES cells, *in vitro* differentiation gave rise to more than 900 DMRs in NPs, a number that increased about 40 fold in MEFs, a more advanced stage of differentiation (Appendix III, Fig. 1b). Further characterization of the NP DMRs revealed that DMRs with a high CpG density were mostly hypomethylated in knockout NPs, whereas CpG poor DMRs were almost exclusively hypermethylated (Appendix III, Fig. 1c). Moreover, we wondered whether differential methylation affected regions involved in the regulation of gene expression. We found that hypomethylated DMRs were on

average localized closer to a TSS and more often associated with promoters (ENSEMBL TSS -1.5 kb and +0.5 kb) than hypermethylated DMRs (Appendix III, Fig. 1d and e).

As we previously proposed a function of TDG in safeguarding CGIs against aberrant DNA methylation, we tested whether the NP DMRs overlapped with CGIs. DMRs were indeed significantly enriched for CGIs and almost all these CGI DMRs were hypomethylated in the *Tdg* knockout background (Appendix III, Fig. 2a). This finding was counterintuitive given that CGIs are generally considered devoid of CpG methylation, demanding explanation. As, in principle, hypomethylation in TDG-deficient NPs can derive either from a loss of methylation already present at the ES cell stage or from a failure to *de novo* methylate sequences in absence of TDG during differentiation, we intersected the hypomethylated CGI DMRs with CGIs that are hypermethylated in wildtype NPs compared to wildtype ES cells. Almost all CGI DMRs overlapped with CGIs that are poised for methylation during differentiation (Appendix III, Fig. 2b). Thus, hypomethylation of CGIs in *Tdg* knockout NPs arises through a failure to *de novo* methylate these regions.

As only a minority of CGIs are differentially methylated across cell types and those that are do not necessarily overlap with gene promoters (Illingworth and Bird 2009; Jones 2012), we intersected the CGI DMRs with published datasets of histone modifications and DNA binding factors to characterize which type of CGI is affected by the absence of TDG. We found that the CGI DMRs were depleted for promoters (ENSEMBL TSS -1.5 kb and +0.5 kb), RNA polymerase II, p300 and H3K27ac. On the other hand, we found the CGI DMRs to be significantly enriched for TET1 binding sites, H3K4me1, H3K27me3 as well as distal regulatory regions that were recently characterized as low methylated regions (LMRs) (Stadler et al. 2011). Especially with NP-specific LMRs, i.e. binding sites of transcription factors only expressed or active in NPs, we found a striking overlap of 52%. The association of hypomethylation in the TDG-deficient NPs with LMRs expanded also to CpG poor DMRs, 20% of which were LMRs. Taken together, this characterization revealed that CGI DMRs overlap with polycomb targets (H3K27me3) (Kuzmichev et al. 2002) and enhancer regions (LMRs) (Stadler et al. 2011) which appear to be in a silent or poised state, enriched for H3K4me1 but not p300 or H3K27ac (Creyghton et al. 2010).

Reduced levels of 5-mC at the CGI DMRs can be explained either by a failure to recruit Dnmt3a and b, which were both shown to interact with TDG (Li et al. 2007; Boland and Christman 2008), or by a chemical conversion of 5-mC to a derivative that is no longer recognized by the 5-mC antibody used for MeDIP. Such a conversion could entail deamination, which would result in a pre-mutagenic G•T mismatch, or TET catalyzed oxidation to 5-hmC and further to 5-fC and 5-caC. Alternatively, 5-hmC could be deaminated to give rise to a G•5-hmU mismatch. Both deamination-coupled scenarios would result in C → T transition mutations.

To test these different hypotheses, we performed hairpin bisulphite sequencing (BS-seq), which allows simultaneous assessment of strand-specific methylation and mutation frequency at selected targets with a sequencing read depth of $\sim 10^4$. To be able to distinguish between a structural and a catalytic function of TDG, we performed *in vitro* differentiation in a complemented cell system (*Tdg*^{-/-} ES cells complemented with wildtype *Tdg*, catalytically inactive *Tdg*_{Δcat} or the empty complementation vector). Analysis of a set of representative CGI DMRs confirmed hypomethylation of 5 out of 7 targets in knockout NPs and cells expressing *Tdg*_{Δcat} (Appendix III, Fig. 3). As the catalytic inactive but structurally intact TDG should still be capable of recruiting other factors like Dnmt3a and b, it is unlikely that hypomethylation is caused by a failure to direct the *de novo* methyltransferases to these regions during differentiation. Importantly, we did not find any increase in the C → T mutation frequency in knockout and TDG_{Δcat}. We thus conclude that 5-mC is not lost due to deamination and ultimately mutation of CpGs at these CGIs.

As the product of TET mediated oxidation, namely 5-hmC, 5-fC and 5-caC are not maintained through DNA replication by Dnmt1 (Valinluck and Sowers 2007; Inoue et al. 2011), we performed a short (24 h) RA differentiation timecourse to avoid dilution of these 5-mC derivatives. Further, to reduce epigenetic heterogeneity, we preconditioned the ES cells in 2i medium for 4 passages prior to RA differentiation, which, *nota bene*, reduced the global 5-mC levels by approximately 50% (Appendix III, Fig. S2b). By monitoring RNA and protein levels of key factors like TET1 and TET2, but also of pluripotency genes and transcription factors silenced or induced by differentiation, we were able to confirm equal loss of pluripotency in all three complemented ES cell lines (Appendix III, Fig. S1). It is noteworthy that *AID* RNA levels remained extremely low throughout differentiation and protein levels were below the detection limit in Western blot analysis (Appendix III, Fig. S1c).

We investigated the global levels of 5-mC, 5-hmC, 5-fC, 5-caC and 5-hmU throughout the 24 hours timecourse of RA differentiation and found the levels of 5-mC, 5-fC and 5-caC to rise significantly with time in knockout and TDG_{Δcat} but not in wildtype cells (Appendix III, Fig. 4 and S2a). Differential 5-mC and 5-fC levels between TDG-proficient and -deficient cells thus arose with differentiation. 5-caC levels were already in the pluripotent state about 9 fold higher in cells lacking TDG activity, which is in agreement with previous observations in *Tdg* knockdown experiments (He et al. 2011). We also verified that TDG_{Δcat} retains no activity on 5-caC but found that it binds this substrate with higher affinity than C, 5-mC or 5-hmC (Appendix III, Fig. 5b and c). Taken together, our findings suggest that differentiation-induced methylation is accompanied by the generation of higher oxidized 5-mC derivatives which accumulate in TDG deficient cells.

To investigate whether the changes observed on the global scale reflect the situation at the CGI DMRs, we performed targeted analysis of 5-mC and 5-caC levels. We found the levels of 5-mC and 5-caC to increase in both, wildtype and TDG_{Δcat} cells but not in absence of TDG. TDG_{Δcat} cells displayed a

striking accumulation of 5-caC at the CGI DMRs (Appendix III, Fig. 5a). It thus appears that differentiation-induced methylation of the CGI DMRs is accompanied by stepwise oxidation of 5-mC. As the overall trajectory of the CGI DMRs points towards methylation in NPs (Appendix III, Fig. 3), the generation of not only 5-mC but also 5-caC suggests that 5-caC eventually has to be removed and replaced with C, the substrate for Dnmt3a and b. In knockout cells, on the other hand, RA differentiation induced neither methylation nor the generation of 5-caC at the CGI DMRs, suggesting a structural function of TDG in inducing or maintaining the cycle of cytosine methylation and demethylation.

To further investigate the structural and catalytic contribution of TDG to the cycle of CGI methylation and demethylation, we performed chromatin immunoprecipitation (ChIP) to monitor the association of TET1, the most highly expressed member of the TET family in ES cells (Appendix III, Fig. S1a and b), to the CGI DMRs during differentiation. We found that after incubation with RA, TET1 binds to the CGI DMRs with increasing affinity in wildtype cells but loses this affinity with differentiation in knockout and, surprisingly, also in TDG_{Δcat} cells (Appendix III, Fig. 6a and S4).

As TDG_{Δcat} is incapable of processing 5-caC but binds this substrate with high affinity, we reasoned that TDG_{Δcat} might bind the accumulating 5-caC at the CGI DMRs and thus destabilize TET1 association. We therefore performed ChIP to compare the affinity of wildtype and TDG_{Δcat} to the CGI DMRs. Indeed, we found that wildtype TDG is associated with the DMRs only transiently, whereas TDG_{Δcat} accumulated at the CGI DMRs (Appendix III, Fig. 6c and S5). The inability of TDG_{Δcat} to turn over might result in unusual DNA demethylation intermediates after the next round of DNA replication, i.e. hemi-caC sites which are no longer a substrate for TET1.

Taken together, we found that TDG controls the transition of CGI methylation states during cell lineage restriction by safeguarding the balance of 5-mC and its oxidized derivatives during a state of high epigenetic plasticity. We found that cell differentiation induces a cycle of DNA methylation and oxidative demethylation at CGIs associated with polycomb targets and poised enhancers and propose that TDG not only contributes catalytically to this cycle by excising the products of TET mediated 5-mC oxidation but also structurally coordinates initiation and/or maintenance of the cycle.

Contribution: I cultured the complemented ES cell lines before and during the 24h RA timecourse, checked the cell cycle profile of undifferentiated and differentiated cells by FACS (data not shown), prepared genomic DNA, RNA and protein extracts of a 2i control (-16h) and timepoints 0, 4, 8 and 24 h, and organized and coordinated the LCMSMS analysis of genomic levels of 5-mC derivatives in collaboration with researchers from the Hans Krokan laboratory in Trondheim, Norway. Moreover, I conducted qRT-PCR and Western Blot analyses for the 24 h timecourse, as well as statistical analysis of the LCMSMS, MeDIP, GLIB, caCDIP and ChIP results. Together with A. Wirz, I prepared chromatin

and performed TET1- and TDG-ChIP of timepoints 0, 8 and 24 h. I also performed and coordinated bioinformatic analyses to characterize NP DMRs with regard to the correlation between methylation change and CpG density, proximity to the nearest TSS and overlap with various (epi)genomic features. Finally, I wrote the manuscript in Appendix III.

4.4. Supplementary Results

4.4.1. A cell-based assay to measure the epigenetic stability of gene promoters

The finding that mechanisms exist to ensure the epigenetic stability of promoters during cell differentiation has raised the need for tools to assess the rate of stochastic epigenetic aberrations. To this end I set out to develop a cell-based assay to measure the epigenetic stability of promoters, which can be applied in high-throughput screens as well as to analyze stochastic events on a clonal basis.

In a pilot experiment, I explored a fluorescence microscopy-based approach (Meilinger et al. 2009) to compare the silencing kinetics of the cytomegalovirus (pCMV) promoter driving mCherry expression to the epigenetically more stable chicken actin promoter (pCAG) driving the expression of GFP. Both constructs were transiently co-transfected into TDG-proficient and -deficient MEFs and cells were imaged for 2, 4, 6 and 8 days after transfection (Fig. 4-01). Both constructs are non-replicating as they lack a mammalian origin of replication and, thus, we expect dilution by cell-replication to contribute to loss of fluorescence signal.

Although pCAG is regarded as an epigenetically more stable promoter, we found a rapid loss of GFP signal within the first 2-4 days. Furthermore, we observed a faster loss of signal of both constructs in TDG knockout compared to wildtype MEFs (Fig. 4-01b), which can be interpreted either as faster dilution of the constructs in knockout MEFs or as differential silencing of the constructs in absence of TDG.

The proliferation-rate and also the rate of random integration of the constructs into the genome might produce cell-line specific biases, raising the need for a more controlled approach to monitor promoter silencing, excluding such artifacts. Moreover, as this simple promoter silencing assay is limited in its applications, I designed and generated a construct for stable integration into ES cells that allows keeping the promoter of interest under selective pressure and timing its release to monitor gradual silencing. The construct, termed pProX, contains two modified Cre recombinase

target (lox) sites, lox511 and lox2272, facilitating site-directed, unidirectional exchange of the whole cassette. Upstream of lox511, we inserted 8 repeats of the lac operator (lacO) sequence to allow targeting of a factor of interest fused to the lac repressor (lacI). Such targeting of TDG has been shown to induce DNA demethylation of the targeted sequences (Gregory et al. 2012). Downstream of lox2272, we integrated a firefly luciferase reporter gene to allow for simplified expression monitoring (Fig. 4-02).

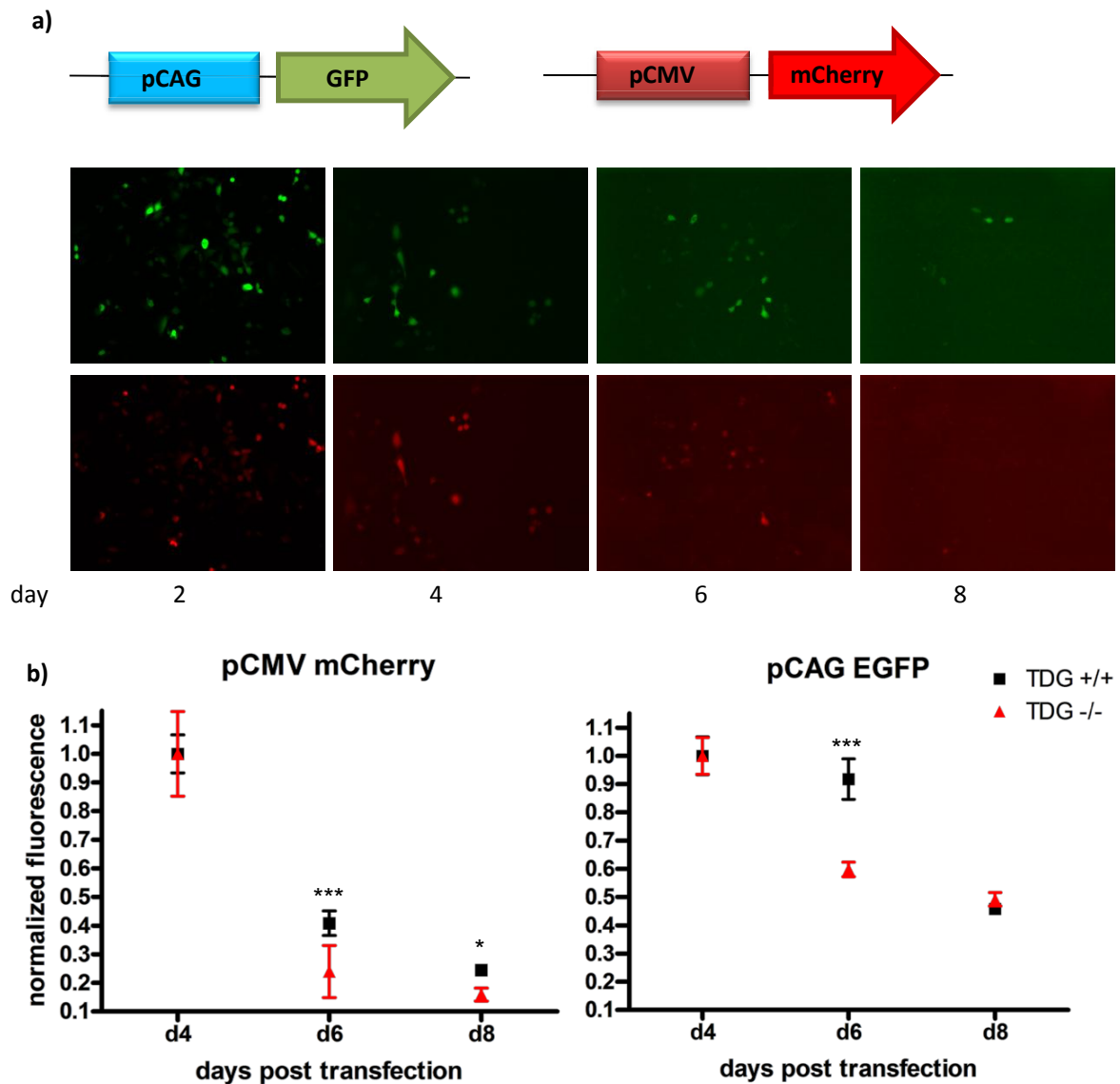


Figure 4-01: Silencing of transiently transfected fluorescence reporter constructs. a) pCMV mCherry and pCAG GFP were co-transfected into TDG wildtype and knockout MEFs. Images were acquired every second day with equal settings. b) Medians of 130 - 400 cells per time point and sample. Fluorescence intensities were normalized to the respective median at day 4; error bars, s.e.m. of normalized intensities, *, $p < 0.01$, ***, $p < 0.0001$, Mann-Whitney test.

The Cre-lox exchangeable cassette consists of the promoter of interest, driving expression of the puromycin N-acetyltransferase (PuroR) that confers resistance to puromycin. PuroR allows not only selecting for stable integration of the construct in the ES cell genome but also for keeping the promoter of interest under selective pressure and thus active. PuroR is flanked by flippase (FLP) recognition target (FRT) sites which allow removal of PuroR by FLP-FRT recombination. Removing the selective pressure mediated by PuroR allows gradual silencing of the promoter of interest which can be monitored as transcription of a reporter gene encoding a fusion of GFP with the HSV-1 thymidine kinase (TK) downstream of the second FRT site. Expression of GFP allows live cell sorting to attain a pure population of cells having gone through FLP-FRT recombination to start the silencing time course. Silencing can be monitored as gradual loss of GFP fluorescence by microscopy, as a reduction in GFP-positive cells by FACS, by measuring GFP-TK mRNA levels and by selection for silencing with ganciclovir for clonal analyses (St Clair et al. 1987).

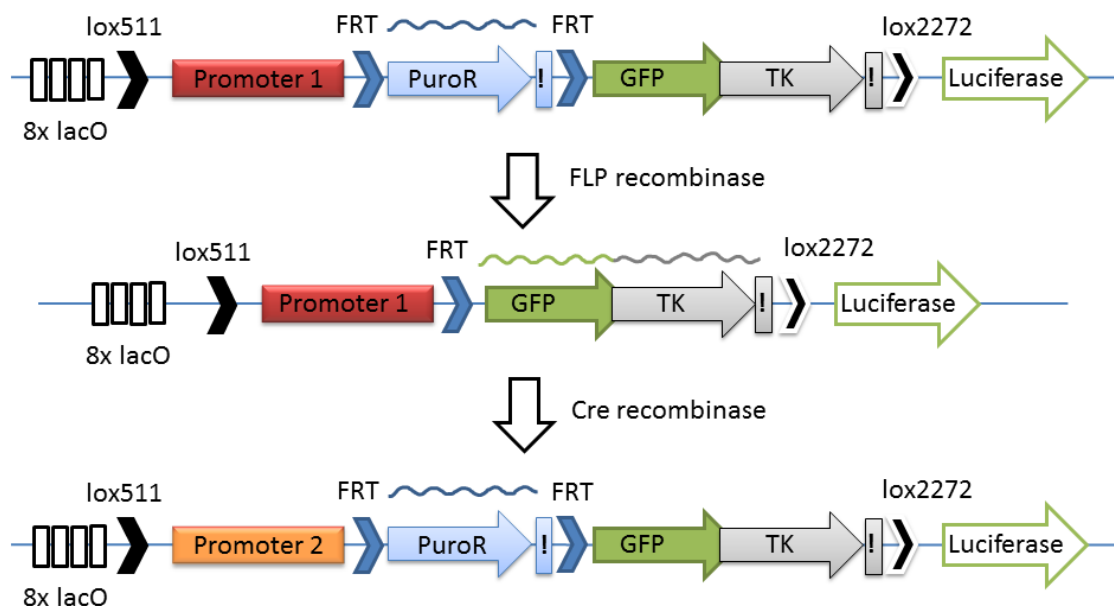


Figure 4-02: Scheme of the pProX construct. FLP-FRT recombination allows removal of PuroR and thus release of the selective pressure on promoter 1. Additionally, transcription of a GFP-TK reporter gene is switched on, which allows cell sorting for a pure population of GFP-TK expressing cells and subsequent monitoring of silencing by fluorescence microscopy, FACS, qRT-PCR and ganciclovir selection. The whole cassette can be exchanged by Cre-lox recombination to analyze other promoters of interest. A sequence encoding firefly luciferase downstream of lox2272 allows a simplified analysis of gene expression. The line above PuroR or GFP-TK indicates transcription, ! marks the position of transcription terminator sequences.

I generated a first construct with the CMV promoter, which provides sufficient expression levels of PuroR for initial selection, and selected for stable integration of the construct into the genome of *Tdg* knockout ES cells in the presence of puromycin. An initial screen of the stable pProX clones was performed by PCR on genomic DNA, testing for the presence of lox511 and lox2272 (Fig. 4-03a).

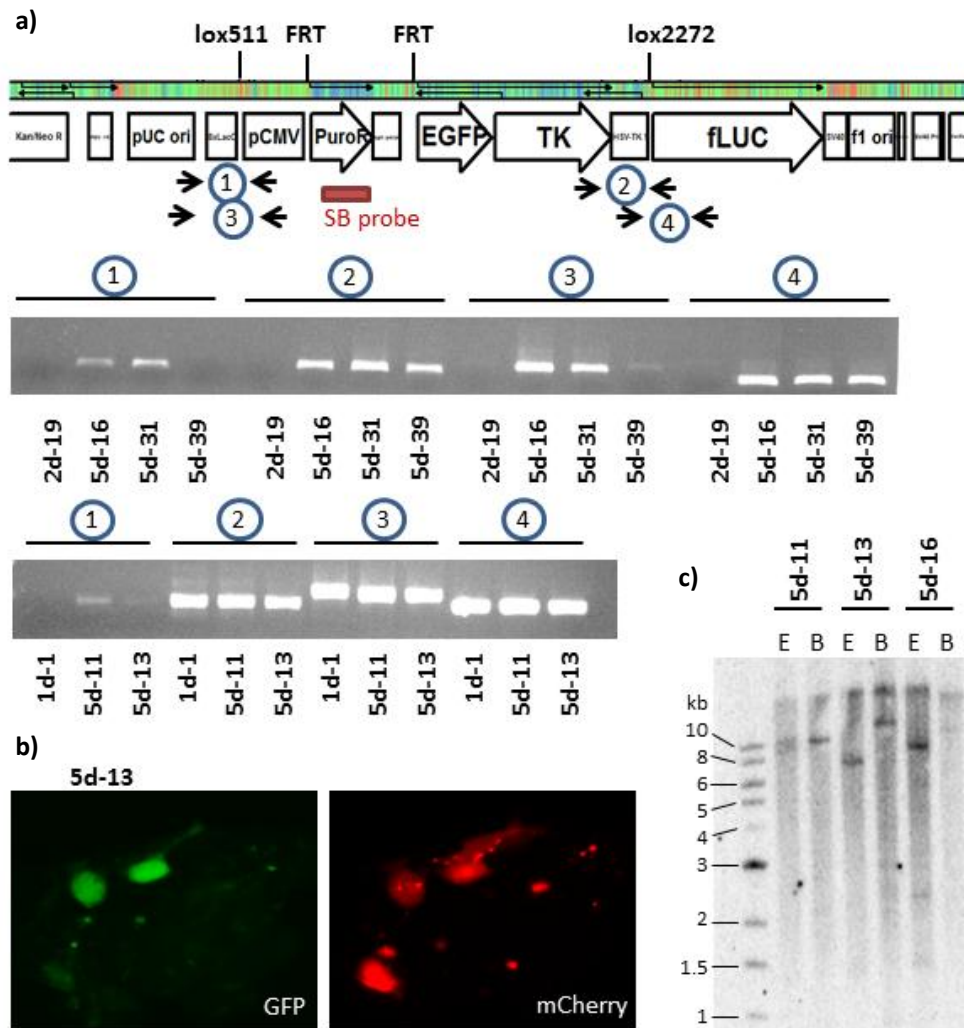


Fig. 4-03: Characterization of ES cell clones with stable integration of pProX. a) Initially, presence of the lox sites was verified by PCR on genomic DNA with 2 primer pairs each (circled numbers). b) Clones with verified lox sites were transiently co-transfected with vectors encoding mCherry and an optimized flippase (FLPo), the latter catalyzing the removal of FRT-site flanked PuroR, resulting in the expression of GFP-TK. c) Genomic DNA from clones with FLP-activatable GFP-expression was subjected to Southern blot (SB) analysis to test for single-copy integration of pProX. The hybridization site of the probe is indicated in a. E, *EcoRI*, B, *BglIII*-digested DNA.

Additionally, clones that were verified to contain the construct were co-transfected with a vector encoding an optimized FLP recombinase (FLPo) and one encoding mCherry as a transfection control. Images were acquired two days after transfection (Fig. 4-03b). Clones that expressed GFP after transfection with FLPo were further characterized with regard to copy number of the construct integrated into the genome. Genomic DNA was digested either with *EcoRI* or with *BglII*, both of which cut once within the pProX construct. The DNA fragments containing the construct were subsequently detected by Southern blot with a radioactive probe hybridizing to the sequence of PuroR (Fig. 4-03a and c). In collaboration with S. Diggelmann, I identified the clones 1d-4 (characterization not shown), 5d-13, 5d-16 and 5d-31 as containing a single copy, the latter two of which yield a robust GFP signal in FACS analysis (data not shown).

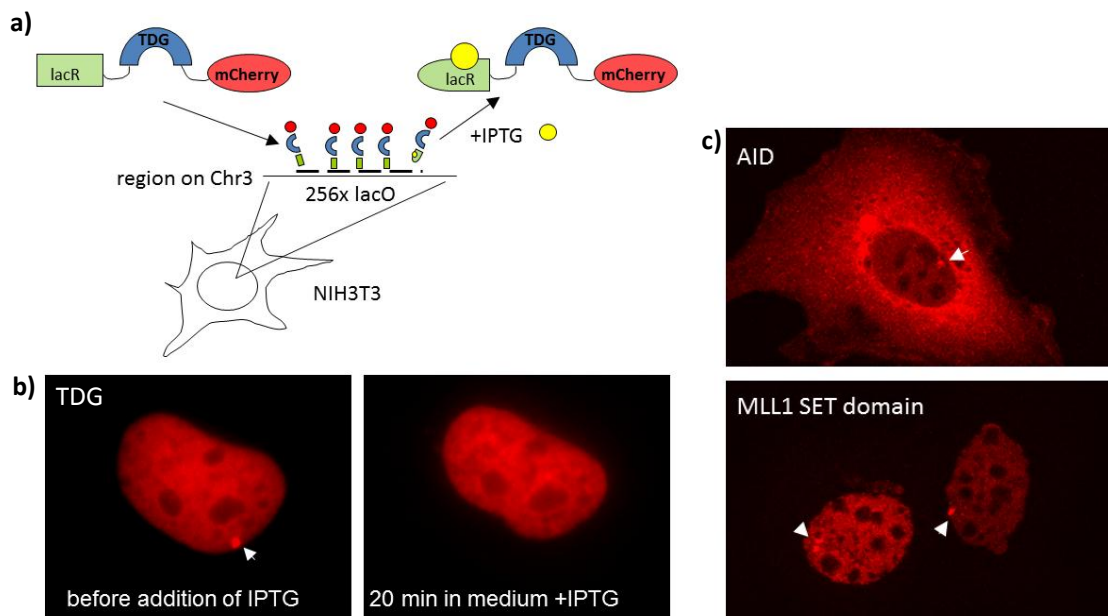


Fig. 4-04: LacO-lacI mediated targeting of mCherry-tagged factors. a) Scheme of targeting to 256 lacO repeats in NIH3T3 cells and IPTG-induced dissociation of lacI-tagged TDG. b) lacI-TDG-mCherry transiently expressed in NIH3T3. The arrow indicates the position of the lacO array within the cell nucleus, evident as a focus of mCherry fluorescence (left panel). After addition of 5 mM Isopropyl- β -D-thiogalactopyranosid (IPTG) to the medium, lacI-TDG-mCherry dissociates from the lacO array (right panel). c) Targeting of lacI-AID-mCherry and lacI-MLL1 SET-mCherry in NIH3T3. The arrows indicate the focal accumulation of the lacI-fusion protein on the lacO array.

Further characterization of these clones will entail ganciclovir selection, test for efficient Cre-lox recombination and a preliminary characterization of silencing rates. Finally, two independent clones will be complemented with a vector containing a *Tdg* minigene controlled by the authentic *Tdg* promoter and allowing alternative splicing to produce isoforms A and B. For separation of function

studies, we also generated a variant of this minigene expressing a catalytic inactive but structurally intact isoform of TDG.

As our system includes the option to target factors of interest to the promoter through lacO lacI interaction, we generated constructs encoding AID, the MLL1 SET domain and TDG fused to mCherry and lacI. In parallel to establishing the pProX system, we validated targeting and IPTG-induced dissociation of the fusion proteins in an available lacO cell system (Soutoglou et al. 2007) (Fig. 4-04). The lacI-fusion constructs can thus be used to address mechanistic questions with the pProX system by measuring the effect of the targeted activities on silencing rates or even reactivation of the promoter.

Taken together, the pProX system provides a novel and versatile method to monitor the epigenetic stability of gene regulatory sequences and to facilitate high throughput screens as well as the isolation and analysis of stochastic silencing events in a cell population, equivalent to the fluctuation test available for the assessment of genetic mutation rates.

Methods:

Construction of pProX. Essentially as described in Appendix I. All components were PCR amplified with primers containing additional restriction sites, FRT or lox sites where required and inserted into EGFP-1 (Clontech). The gene encoding firefly luciferase replaced EGFP and introduced lox2272. The CMV promoter introduced lox511, PuroR the 5' FRT site. The GFP-TK fusion with a 5' FRT site was assembled in EGFP-N1 and subsequently inserted between PuroR and lox2272. All PCR-amplified fragments were verified by sequencing (Microsynth, Switzerland).

Cell culture. All cell lines were incubated at 37°C, 5% CO₂ and 100% moisture. ES cells were cultured essentially as described in Appendix III. *Tdg*^{-/-} cells were cultured in ECM with LIF and transfected with 1, 2 or 5 µg of endotoxin free pProX plasmid (Qiagen, Switzerland) using jetPEI™ (Polyplus Transfections, France). Clones were selected with 1.5 µg/mL puromycin for 5-7 days. Single clones were constantly cultured with 1 µg/mL puromycin. Transfection of mCherry-N1 and FlpO was performed with jetPEI.

NIH3T3 cells with lacO repeats (NIH2/4) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FCS and 2 mM L-glutamine and transfected with lacI-fusion constructs using jetPEI™. To reverse targeting of lacI-fusion proteins, culture medium was supplemented with 5 mM IPTG.

Extraction of genomic DNA. As described in Appendix III.

Validation of lox sites by PCR. Fragments surrounding the lox sites were amplified with Paq5000 polymerase in standard polymerase chain reaction according to the manufacturer's recommendations. Primer pairs used are listed in Table 4-01.

Southern Blot. 15 µg genomic DNA was digested with 5U *EcoRI* or *BglII* per µg DNA for 5 hours at 37°C, then with additional 5U per µg DNA overnight. DNA was purified by ethanol precipitation and resuspended for ~5 h at 55°C shaking at 14000 rpm, followed by separation on a 0.7% agarose gel in TBE with 50 V for 8 h. DNA was stained with 0.3 µg/mL ethidium bromide and visualized under UV. DNA was depurinated by incubating the gel in 0.1% HCl for 10 min. The gel was rinsed in ddH₂O, equilibrated for 10min in 4N NaOH and the DNA transferred overnight to a nylon membrane (Zeta-probe, BioRad, Switzerland) in 4N NaOH. The membrane was briefly rinsed 3 times in 2x SSC, DNA was crosslinked to the membrane by incubation for 2 h at 80°C. The probe detecting the sequence of PuroR was PCR amplified and 100ng of the purified PCR product was radiolabelled with 50µCi [α -³²P]-dCTPs (Perkin Elmer, Switzerland) using the Fermentas Hexalabel DNA labeling kit (Fermentas, Switzerland). Excess nucleotides were removed with the QIAquick Nucleotide Removal Kit (Qiagen, Switzerland). The membrane was pre-hybridized in hybridization buffer (0.25M Na₂HPO₄, pH 7.2, 7% SDS) for 15 min - 2 h at 65°C before addition of the probe and incubation overnight at 65°C. Subsequently, the membrane was washed at 65°C, three times in Wash Solution I (40 mM Na₂HPO₄, pH 7.2, 5% SDS), twice in Wash Solution II (40 mM Na₂HPO₄, pH 7.2, 1% SDS), followed by incubation for 48 h on a phospho-imager screen in a light-tight cassette and scan with a Typhoon fluorescence scanner (GE Healthcare, Switzerland).

Table 4-01: Primers used in clone characterization.

	Name	5'-3' sequence
Primer	pEGFP F	CCGTATTACCGCCATGCATTAG
pair 1	CMV 5' R	GTTATGTAACGCGGAACTCC
Primer	TK seq 3	CTG CTG CAA CTT ACC TCC GG
pair 2	TK BglII NotI rev	CGA GCG GCC GCA GAT CTG GGT CGT CCA CCA GAC CCC ACG
Primer	Across lox511 fw	GGC CTT TTG CTG GCC TTT TGC TC
pair 3	Across lox511 rev	GTC AAT GGG CGG GGG TCG TTG
Primer	Across lox2272 fw	CAG GGC TCG CAG CCA ACG TC
pair 4	Across lox2272 re	CGC GCC CAA CAC CGG CAT AA
Southern	PuroR probe fw	ACA TCG AGC GGG TCA CCG AG
Blot probe	PuroR probe rev	CCA TAG AGC CCA CCG CAT CCC

All primers were obtained from Microsynth, Switzerland.

Contribution: I designed and generated the pIProX construct and the TDG complementation vectors. Furthermore, I selected ES cell clones stably transfected with pIProX and performed initial characterization by PCR and fluorescence microscopy. Moreover, I supervised the Master project of S. Diggelmann who continues characterization of the clones, e.g. by Southern blot.

4.4.2. Generation of 5-caC through inducing oxidative damage in live cells

The TET proteins were found to oxidize 5-mC to 5-hmC and further to 5-fC and 5-caC (Tahiliani et al. 2009; He et al. 2011; Ito et al. 2011). LCMSMS analysis of global 5-fC and 5-caC levels in *Tdg* knockout and *Tdg*_{Δcat} cells revealed 2-9 fold enrichments compared to wildtype cells, which is in agreement with TDG being the primary DNA glycosylase responsible for the excision of 5-fC and 5-caC. However, as these higher oxidized 5-mC derivatives are not maintained by Dnmt1 (Valinluck and Sowers 2007; Inoue et al. 2011) and, thus, diluted by DNA replication, it is striking that such a high level of 5-fC and 5-caC accumulate in TDG-deficient cells. We therefore asked whether 5-fC and 5-caC might arise spontaneously through non-enzymatic oxidation of 5-hmC and, thus, if oxidative stress might have an impact on epigenetic stability.

We treated *Tdg*^{-/-} ES cells complemented with wildtype *Tdg* or *Tdg*_{Δcat} with H₂O₂ to induce oxidative stress and subsequently performed immunofluorescence staining with a specific anti 5-caC antibody. Surprisingly, we found 5-caC to localize in distinct foci throughout the nucleus in absence of oxidative damage and independent of cell cycle stage, irrespective of the TDG status. These foci did not co-localizing with focal PCNA during S-phase (Fig. 4-04a), indicating that they do not associate with progressing replication forks. As 5-caC localization in ES cells has not been described so far, neither by immunofluorescence staining nor by genome wide mapping, it remains to be verified that 5-caC indeed clusters in foci. However, 15min of incubation in H₂O₂ appeared to increase the number of foci in a dose dependent manner in wildtype cells but not in TDG_{Δcat} (Fig. 4-04b), suggesting that the latter is either refractory to H₂O₂-induced generation of 5-caC or that the already elevated levels of 5-caC in TDG_{Δcat} represent a ground state that allows no further induction of 5-caC without erratic effects. The wider range of foci numbers observed in TDG_{Δcat} might indicate the latter.

We next measured the global levels of 5-mC, 5-hmC, 5-fC, 5-caC and - as a control - 8-oxoG by LCMSMS. 15min of H₂O₂ exposure plus 15min of recovery increased the global 8-oxoG levels approximately 7 fold in wildtype genomic DNA compared to mock treated cells (Fig. 4-04c).

Interestingly, 5-mC levels were decreased and 5-hmC, 5-fC and 5-caC levels were increased by H₂O₂ in wildtype cells, suggesting that the application of oxidative stress indeed oxidized 5-mC to 5-hmC and further. However, the H₂O₂ treatment appeared ineffective in *Tdg*_{Δcat} and knockout cells, as evident from the unchanged levels of 8-oxoG. The variable efficacy of the treatment might be caused by the instability of H₂O₂. Surprisingly though, the effect observed in wildtype cells appeared to be reversed in *Tdg*_{Δcat} and knockout cells suggesting that the treatment converted 5-hmC, 5-fC and 5-caC in a way not detectable in our approach.

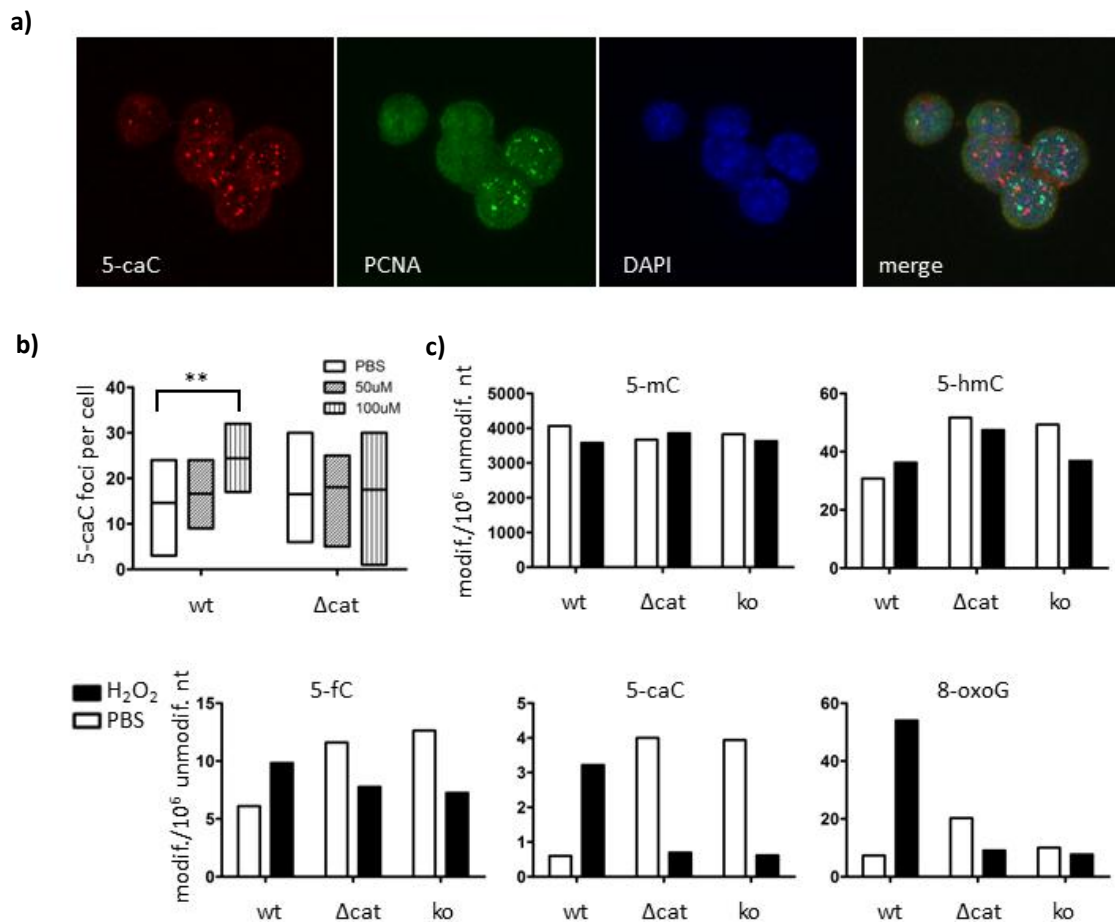


Fig. 4-04: Non-enzymatic generation of 5-fC and 5-caC. a) Immunofluorescence staining of 5-caC and PCNA in wildtype ES cells mock treated in PBS for 15min. DAPI was used to stain DNA. b) Number of 5-caC foci per cell (n=7-12), visually assessed in z-stacks acquired on a confocal microscope (Leica SP5). Floating bars indicate median plus range. Statistical analysis by two-way Anova with Bonferroni post-test. ** p<0.001. c) LCMSMS quantification of the indicated modified bases in the genome of wildtype (wt), *TDG*_{Δcat} (Δcat) and knockout (ko) cells. Bars indicate number per 10⁶ unmodified bases (A, C, T and G).

These results were obtained in a single experiment the reproduction of which failed due to the erratic effects of H₂O₂ on the induction of oxidative damage. Still, my preliminary results indicate that higher oxidized 5-mC derivatives could indeed arise non-enzymatically, suggesting that oxidative stress, besides inducing genetic mutations, might also destabilize cell programming. The question how oxidative damage affects 5-mC and its derivatives thus remains an important one and will have to be readdressed in a more robust set-up. Other reagents which are more stable, i.e. menadione, might prove more suitable to study the generation of 5-fC and 5-caC under oxidative stress.

Methods:

ES cell culture. Essentially as described in Appendix III. Cells were cultured in 2i medium prior to treatment with 100 µM (if not indicated otherwise) H₂O₂ in PBS or just PBS (mock control) for 15 min. Cells were left for 15 min in culture medium to recover followed by further processing.

Immunocytochemistry. Essentially as described in Appendix II. Cells were fixed in -20°C cold methanol for 5 min on ice and then rehydrated for 2x 15 min in PBS. After permeabilization with 0.2% TritonX-100/PBS for 5 min on ice, cells were rinsed with PBS and DNA was depurinated with 2 N HCl for 20 min. Cells were washed with PBS and neutralized by two 10 min washes in 100mM Tris-HCl pH 8. After blocking with blocking solution (BS; 2% BSA, 0.05% Tween20, PBS) for 1 h at room temperature, cells were incubated with the primary antibodies (anti 5-caC 1:500, Active Motif, Belgium; anti-PCNA FITC 1:400, Leinco Technologies, Missouri USA) in BS over night at 4°C. Three 10 min washes with BS were followed by incubation with secondary antibody (goat anti rabbit Alexa 594 1:500, Invitrogen – Life Technologies Switzerland) in BS for 1 h. After two 10 min washes in BS and one in PBS, DNA was stained with 5 µg/mL DAPI. Excess DAPI was removed by two 5 min washes in PBS followed by mounting in Vectashield (Vector Labs, UK). Z-stacks were acquired on a Leica SP5 with 405 nm, 488nm and 594 nm laser lines (Leica, Switzerland).

LCMSMS. As described in Appendix III. Conditions for 8-oxo(dG) were as for C-modifications. Mass spectrometry detection of 8-oxo(dG) was performed using an MDS Sciex API5000 triple quadrupole (Applied Biosystems – Life Technologies, Norway) operating in positive electrospray ionization mode for the mass transition 284.1/168.2.

Contribution: I cultured the ES cells and performed the treatment with H₂O₂. Moreover, I established the 5-caC immunofluorescence staining, conducted microscopy-based analysis and visual assessment

of 5-caC foci number. I further prepared genomic DNA from treated and mock treated cells for LCMSMS analysis.

4.4.3. Familial loss-of-function TET2 mutation does not correlate with the disease phenotype

Alexey Veligodskiy¹, Franz X. Schaub¹, Angelika Jacobs², Primo Schär², and Radek C. Skoda¹

(Manuscript in preparation)

¹ Experimental Hematology, ² Institute of Biochemistry and Genetics, Department of Biomedicine, University Hospital Basel, Basel, Switzerland

Members of the TET family of proteins have been implicated in myeloid malignancies. Only recently, TET proteins have raised considerable attention in the epigenetics field, owing to the discovery of their biochemical activity as iron- and alpha-ketoglutarate- dependent dioxygenases, converting 5-methylcytosine to 5-hydroxymethylcytosine, and ultimately to 5-formylcytosine and 5-carboxylcytosine (chapter 2.4.2. and (Tan and Shi 2012)). Whereas TET1 was originally identified as a fusion partner of MLL, generated by aberrant chromosomal rearrangements in acute myeloid leukemia (AML) (Lorsbach et al. 2003), *TET2* mutations have been identified in patients with hematological diseases like myeloproliferative neoplasms (MPN), myelodysplastic syndromes and AML (Tefferi et al. 2009). Among these mutations are frameshifts, nonsense and missense mutations that abolish its catalytic activity. Hence, patients with *TET2* mutations display lower overall 5-hmC levels and hypermethylation at promoter regions (Ko 2010, Figueroa 2011, Liu X 2013). The impact of *TET2* inactivation on myeloid malignancies is not yet fully understood although it was proposed to promote expansion of mutant progenitor or stem cells in the course of hematopoiesis (Delhommeau et al. 2009; Figueroa et al. 2010; Schaub et al. 2010; Swierczek et al. 2011). However, it is unclear whether *TET2* inactivation is sufficient to cause disease.

The frameshift mutation *TET2*-D1858fs was identified as a heterozygous germline mutation in a familial case of MPN (Schaub et al. 2010). The frameshift is caused by a deletion of 4 nt that results in a stretch of 27 amino acids unrelated to the original sequence before a premature stop-codon. The

mutation thus abolishes two out of three amino acid moieties required for binding of Fe(II) ions and 2-oxoglutarate, namely H1881 and R1896 (Tailiani 2009, Ito 2010). Despite the fact that knockdown of TET2 resulted in expansion of progenitor and stem cells in a mouse model (Figuroa et al. 2010), two individuals carrying the heterozygous mutation were found to be asymptomatic regarding blood counts or reported hematological diseases but displayed decreased levels of 5-hmC in DNA isolated from peripheral blood mononuclear cells (Veligodskiy et al, manuscript in preparation). Therefore, we tested the catalytic activity of TET2-D1858fs. In addition to the analysis of global 5-hmC levels in genomic DNA of HEK293T cells expressing wildtype or mutant *TET2* by dot blot (Veligodskiy et al, manuscript in preparation), we tested the ability of TET2-D1858fs to induce the generation of 5-hmC in HEK293T cells by immunofluorescence staining and confocal microscopy. We found that wildtype TET2 readily generated 5-hmC, whereas the naturally low levels of 5-hmC in HEK293T cells remained unchanged in cells expressing TET2-D1858fs or a previously described truncated and inactive TET2-S1848x (Langemeijer et al. 2009) (Fig. 4-05). We conclude that the frameshift mutation D1858fs abolished TET2 activity and that the absence of a phenotype in these patients may be explained with compensation by the functional copy of *TET2*.

In conclusion, we report two cases of a heterozygous germline loss-of-function mutation in *TET2* in individuals without any symptoms of hematological disease. Whereas the functional copy of *TET2* might be sufficient to suppress a phenotype, 5-hmC levels are decreased in peripheral blood mononuclear cells from these individuals (Veligodskiy et al., manuscript in preparation). Also, the development of symptoms in members of the same family with haploinsufficiency of *TET2* suggests an involvement of other, as yet unknown factors. Notably, knockdown of *TET2* by 40-80% in murine hematopoietic stem cells (which might be comparable with haploinsufficiency) was shown to impair differentiation (Figuroa et al. 2010) and mice with targeted deletion of *TET2* develop myeloid malignancies (Li et al. 2011).

Taken together, the decrease in 5-hmC levels alone appears to not be sufficient to explain the development of symptoms, raising the question whether TET2 is structurally rather than enzymatically required for controlling haematopoiesis. A recent study revealed that TET2 targets the O-linked β -N-acetylglucosamine transferase (OGT) to chromatin to modify a serine in H2B (Chen et al. 2013; Deplus et al. 2013; Vella et al. 2013). As this TET2-OGT complex is potentially involved in the regulation of gene expression, it remains to be elucidated whether truncated TET2, i.e. TET2-D1858fs, is still capable of interacting with and targeting OGT.

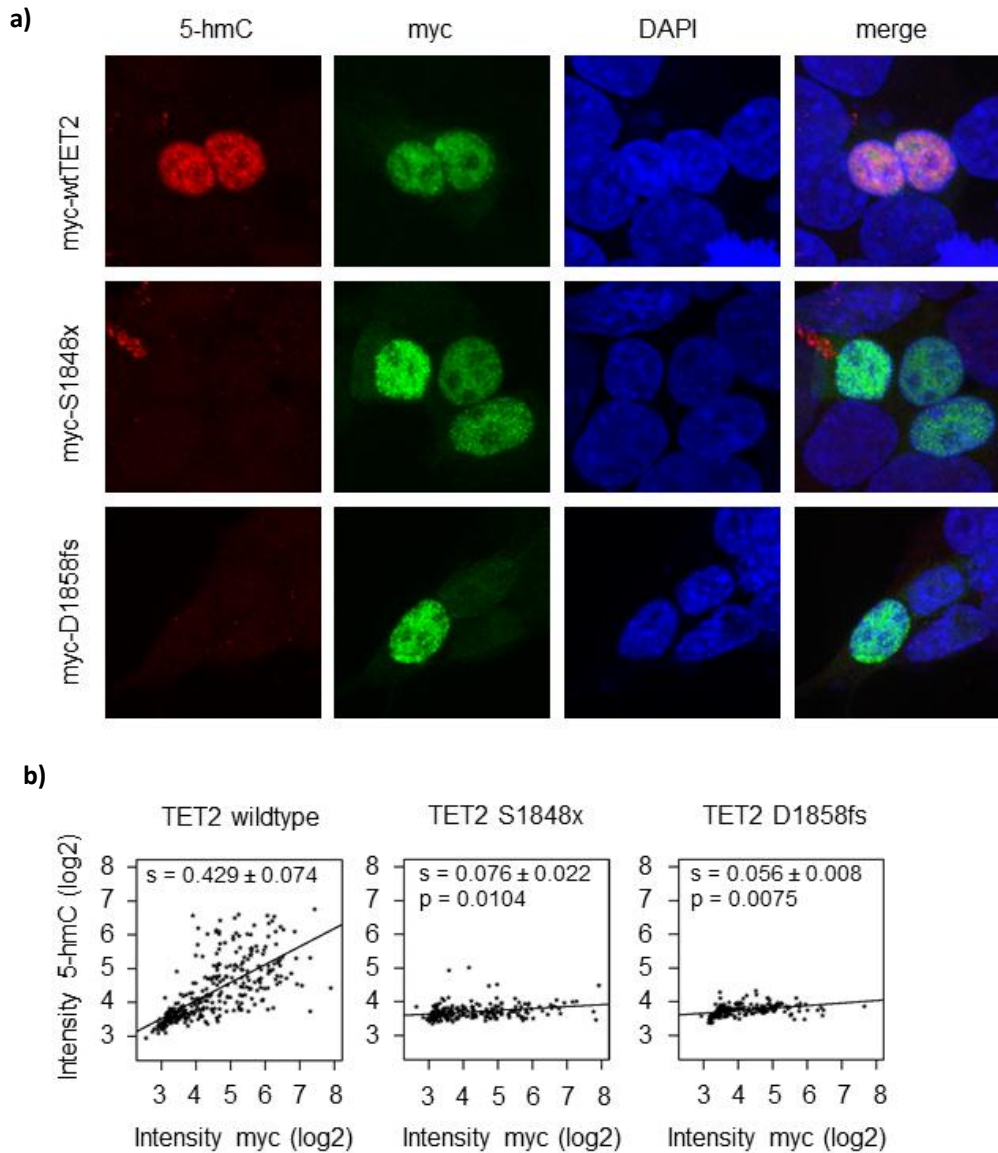


Fig. 4-05: 5-hmC levels in HEK293T cells expressing wildtype and mutant myc-tagged TET2. a) Representative confocal images (z-stack projection) acquired 44-48 h after transfection. TET2-S1848x and TET2-D1858fs do not generate 5-hmC. b) Scatter plot of log₂-transformed average intensities of myc and 5-hmC signals (0 to 255) per cell nucleus with linear regression. *s*, average slope resulting from 3 replicate experiments ± s.e.m.; *p*-values of t-test versus wildtype.

Methods:

Immunocytochemistry. HEK293T cells were seeded onto sterile coverslips in 12-well plates at 3×10^4 cells/well one day prior to transfection with expression constructs encoding myc-tagged wildtype TET2, D1858fs, S1848x and empty vector (myc only) using JetPEI™ (Polyplus Transfections, France)

according to the manufacturer's instructions. 44-48 h post transfection, cells were fixed for 10 min in 3.7% paraformaldehyde/PBS, washed briefly in PBS and incubated for 5 min in pre-chilled methanol at -20°C. After 30 min of rehydration in PBS, cells were permeabilized for 30 min in 0.2% TritonX100/PBS, incubated for 15 min in 2 N HCl and neutralized for 10 min in 100mM Tris-HCl pH 8.4. Cells were blocked for 1 h at room temperature in 3%BSA, 0.1% Tween 20 in PBS (BS), incubated over night at 4°C with primary antibodies (mouse anti-myc 9E11, Santa Cruz, rabbit anti-hydroxymethylcytosine, Active Motif, Belgium) 1:200 in BS, followed by detection with secondary antibodies (goat anti-rabbit Alexa594, goat anti-mouse Alexa488, Invitrogen - Life Technologies, Switzerland) 1:200 in BS for 30-60 min at room temperature. Nuclei were stained with 5 µg/mL of 4',6-diamidino-2-phenylindole (DAPI) and cover slips were mounted in VectaShield (Vector Labs, UK). Images were collected as z-stacks on a Leica SP5 confocal microscope (Leica, Switzerland), using the 20x objective for quantification and 63x objective for visualization of cellular TET2 and HMC distribution. Four z-stacks per sample were acquired with the 405nm, 488nm and 594nm laser lines, and subsequently converted to 2D tiff images by maximum intensity projection in ImageJ (<http://imagej.nih.gov/ij/>) for further image analysis.

Image analysis. Image quantification was performed with the CellProfiler™ software (www.cellprofiler.org). Cell nuclei were identified in the DAPI images with the "IdentifyPrimaryObjects" module (between 20 and 80 pixel units in diameter), with a lower boundary of the "Otsu PerObject" derived threshold of 0.05. Misinterpreted objects were manually removed with the "EditObjectsManually" module. HMC and mycTET2 intensities were quantified in the remaining regions using the "MeasureObjectIntensities" module. Average intensities per cell (ranging from 0 to 255) were log2 transformed with a pseudocount of 1 ($y=\log_2(x+1)$). The corresponding figures were generated in R (www.r-project.org).

Contribution: I transfected HEK293T cells with wildtype and mutant myc-tagged TET2. Subsequently, I established immunocytochemical co-staining of 5-hmC and myc-tagged TET2, conducted image acquisition, automated image analysis with the CellProfiler software (www.cellprofiler.org) and statistical analysis with R.

4.5. DNA Glycosylases: In DNA Repair and Beyond (Appendix IV)

The integrity of the genome is safeguarded by a number of cellular DNA repair mechanisms that have evolved to specifically address different kinds of damage. Single-base lesions caused by oxidation, deamination or alkylation are repaired by the Base Excision Repair (BER) pathway which is initiated by DNA Glycosylases. These enzymes are capable of detecting base damage that does not distort the overall structure of the DNA double helix but is pro-mutagenic or cytotoxic by disturbing a base's Watson-Crick base-pairing properties or interferes with DNA replication. Detection of such small irregularities is possible through a mode of action common to all DNA glycosylases: they flip a damaged base out of the double helix and into a catalytic site cavity, in which specific interactions between the base and key amino acid residues ensure recognition of the damage and alignment of the base for hydrolysis of the N-glycosidic bond.

While all DNA glycosylases share this base-flipping mechanism, they differ in how they scan the DNA for base lesions, how such lesions are initially recognized and how excision of a damaged base is catalyzed. I performed a comprehensive study of the literature and explored the principles of design relating to the function of DNA glycosylases in the recognition of DNA base modifications in the context of canonical DNA repair and other processes, e.g. adaptive immunity and DNA demethylation (Appendix IV).

5 Concluding Discussion and Outlook

Following its discovery as mismatch directed Thymine DNA Glycosylase, studies on TDG function have revolved mainly around classical aspects of DNA repair and genome stability. However, recent discoveries pointed into a quite different direction, implicating TDG dependent repair processes in the maintenance of epigenetic rather than genetic stability. These important discoveries have shaped the progress of my PhD thesis, the aims of which was to investigate TDG function both mechanistically and biologically.

TDG has been isolated as interaction partner of many different kinds of proteins. Most often these were transcription factors, implicating a contributing of TDG to the regulation of gene expression. Besides transcription factors, Dnmt3a and Dnmt3b were also found to interact with TDG, connecting the glycosylase with DNA methylation control (Cortazar et al. 2007; Li et al. 2007; Boland and Christman 2008). The search for interaction partners with TDG as a bait, however, was much less yielding, and produced only a few hits. The strongest interaction partners identified in this way were Small Ubiquitin like Modifiers (SUMOs), and it turned out that TDG is indeed covalently modified but also non-covalently interacts with SUMOs. Biochemical studies have implicated SUMOylation in controlling the dissociation of TDG from its product AP-site (Hardeland et al. 2002; Steinacher and Schar 2005; Mohan et al. 2007). To investigate the function of TDG SUMOylation *in vivo*, I have established a Fluorescence Resonance Energy Transfer (FRET) system and was able to show first evidence for an involvement of SUMOylation in TDG dependent DNA repair in live cells. Preliminary studies of the SUMO-TDG interaction indicated that SUMO1 and SUMO3 might control TDG function in distinct contexts, as I was able to shift the balance of covalent and non-covalent SUMO1-binding by inducing DNA damage, whereas the SUMO3-interaction remained mostly unaffected. Moreover, I found a surprisingly high level of SUMO-TDG interaction already in absence of exogenous DNA damaging agents. These studies were based on the classical paradigm that TDG is a bona fide DNA repair protein counteracting DNA damage. However, this view has changed with the finding that TDG contributes to epigenetic stability.

In the light of recent findings that TDG is involved in chromatin maintenance and more specifically in maintaining a cycle of DNA methylation and demethylation (Appendix II and III), the steady state of SUMO-TDG interaction we observed might reflect the contribution of SUMO to regulating TDG-dependent processes beyond canonical DNA repair. Be it in counteracting DNA damage or in safeguarding chromatin states, TDG operates in an intricate framework of enzymatic cascades that need to be tightly controlled to ensure efficient substrate processing, handover and dissociation in a complex of structural and catalytic factors. SUMOylation and SUMO-binding are very likely involved in regulating these processes and transactions, creating a need for tools to monitor the SUMO-

interactions of the key players of genome and epigenome maintenance in living cells. Our FRET system allows analyzing the SUMO-TDG interaction dynamics not only in response to DNA damage but also for instance in the course of ES cell differentiation. By including a SUMOylation deficient mutant TDG, we can distinguish between the contribution of SUMO-binding and SUMOylation to controlling TDG function in these processes.

Remarkably, our attempt to generate a FRET system to monitor the conformational changes of TDG revealed that part of the cellular TDG pool exists as homodimers. In the context of active DNA demethylation, it is tempting to speculate that homodimerization of TDG might inhibit simultaneous repair processes on opposing DNA strands. As 5-mC mostly occurs symmetrically on a palindromic CpG dinucleotide, mechanisms must have evolved to avoid the simultaneous conversion of both 5-mCs to a BER-substrate and subsequent generation of two adjacent AP-sites. On the other hand, the high affinity of TDG for its product AP-site might already block the access of TET proteins to the opposite 5-mC. These hypotheses may be tested in the future by making use of biochemical assays, for instance by testing whether pre-incubation of TDG_{Δcat} with a substrate containing 5-caCpG/5-hmCpG inhibits oxidation of 5-hmC to 5-fC/5-caC by TET. Combination with the FRET system using purified FRET fusion proteins (single- or double-tagged TDG) would allow testing a) whether dimerization depends on DNA and b) whether a potential exclusion of TET from the opposite 5-hmC involves TDG dimerization.

Establishing the function of TDG in safeguarding epigenetic transitions, we found that knockout of *Tdg* in mice is embryonic lethal and that CGI promoters acquire aberrant methylation during *in vitro* differentiation in TDG-deficient ES cells. Furthermore, we observed that RA-induced *in vitro* differentiation increases the number of XRCC1 foci and the sensitivity to PARP inhibition in wildtype but not in *Tdg* knockout cells (Appendix II), indicating that single-strand breaks are induced upon differentiation in a TDG dependent manner. We found further support for the hypothesis that cell differentiation is accompanied by TDG-dependent DNA repair processes when we observed that the loss of pluripotency triggers a state of high epigenetic plasticity that is characterized by the generation of 5-mC and its higher oxidized derivatives 5-fC and 5-caC at CGIs. TDG appears to contribute to this cycle of DNA methylation and oxidative demethylation not only enzymatically by initiating BER of 5-fC and 5-caC to restore an unmethylated C, but also structurally as we found initiation and/or maintenance of the cycle to fail in *Tdg* knockout cells. Interestingly, *Tdg*_{Δcat} appears to structurally support initiation but due to its inability to excise 5-caC fails to maintain the cycle and - by binding with high affinity to the accumulating 5-caC at these genomic regions - destabilizes the association of TET1.

In the absence of TDG, differentiation stimuli might trigger undirected and uncoordinated methylation. A failure to assemble the protein complex required to maintain the balance between methylation and demethylation might cause the stochastic patterns of hypermethylation we observed at CGI promoters in *Tdg* knockout NPs by pyrosequencing of bisulphite converted genomic DNA. In apparent contrast to this finding, genome-wide methylation analysis by MeDIP-seq revealed that most differentially methylated CGIs in TDG knockout NPs are hypomethylated. However, keeping in mind that MeDIP relies on an antibody for the specific enrichment of 5-mC containing DNA fragments and bisulphite sequencing does not distinguish between 5-mC and 5-hmC, it may very well be that the aberrant methylation we observed by bisulphite sequencing was in fact an accumulation of 5-hmC. Moreover, the fact that this hypermethylation appeared rather mild might reflect that 5-fC and 5-caC, which are indistinguishable from C in bisulphite sequencing, accumulate rather than 5-hmC at these CGI promoters. As methods for mapping 5-fC and 5-caC are being developed (Raiber et al. 2012), it will be possible to analyze the balance between the different C-modifications on a genome-wide scale. The combination of, for instance, 5-caCDIP with next generation sequencing holds the promise of shedding light onto which sites are affected by this differentiation induced cycle of DNA methylation and demethylation. Intersecting such data with published data sets might reveal 1) what signals are necessary to target the key factors involved to these specific sites and 2) what determines the overall trajectory of this cycle towards a methylated or an unmethylated state.

Another question to be addressed is what role TET2 and TET3 play in this cycle. As TET1 is gradually downregulated upon differentiation, it is likely that the other members of the TET family take over at later stages of differentiation and in terminally differentiated cells, providing altered target specificities that might be required in these cells.

To facilitate mechanistic studies into the maintenance of epigenetic stability, I developed a model system that allows assessment of the rates of spontaneous epigenetic aberrations by monitoring the silencing of gene promoters or other sequences of interest. While the core reporter system is established and functional and first experiments in stable TDG proficient and deficient backgrounds are pending, pilot experiments with a transient transfection setup studies already indicated that differential silencing dynamics between TDG-proficient and -deficient cells can be measured. Besides measuring silencing rates, the goal was also to make the tool applicable for analyzing stochastic epigenetic aberrations at a clonal level, and for high-throughput screening for factors or even chemical compounds affecting the epigenetic stability of gene promoters. Moreover, by facilitating the targeting of factors like MLL1, AID, TET1 or TDG to the promoter of interest by lacO-lacI

interaction, we hope to provide a tool for mechanistic investigations into the signals required for assembling the complex mediating epigenetic changes.

Epigenetic instability and thus the failure of pathways proof-reading the epigenome is also a hallmark of many human cancers. The mechanisms behind this epigenetic instability, especially with regard to DNA methylation, are only beginning to be uncovered. For instance, fusions of TET1 with MLL have been identified in several patients with acute myeloid leukemia (Lorsbach et al. 2003) and *TET2* mutations have been associated with myeloid malignancies (Tefferi et al. 2009). Future studies will have to address if mutation, misregulation or mistargeting of the TET proteins - and TDG for that matter - might be involved in carcinogenesis and cancer progression. In a collaboration, we addressed the correlation of a heterozygous loss-of-function *TET2* mutation found in a familial case of MPN with disease phenotype. Strikingly, this mutation was found in family members both with and without myeloid disease symptoms, illustrating the complexity of the interlink between disease and the system that maintains a “healthy” epigenome.

In conclusion, our studies have revealed an essential function of TDG in development and more particularly as part of the cellular machinery that safeguards and possibly directs epigenetic transitions during cell differentiation. In the course of my thesis, I produced data that may guide future studies and generated tools that can be exploited in the analyses of the intricate framework of factors maintaining epigenome stability, to investigate the dynamic protein-protein interactions involved and their effect on promoter stability.

6 References

- Almeida, K. H. and R. W. Sobol (2007). "A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification." DNA Repair (Amst) **6**(6): 695-711.
- An, Q., et al. (2005). "C → T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases." EMBO J **24**(12): 2205-2213.
- Antequera, F. (2003). "Structure, function and evolution of CpG island promoters." Cell Mol Life Sci **60**(8): 1647-1658.
- Baba, D., et al. (2005). "Crystal structure of thymine DNA glycosylase conjugated to SUMO-1." Nature **435**(7044): 979-982.
- Baba, D., et al. (2006). "Crystal structure of SUMO-3-modified thymine-DNA glycosylase." J Mol Biol **359**(1): 137-147.
- Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Res **21**(3): 381-395.
- Barrett, T. E., et al. (1998). "Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions." Cell **92**(1): 117-129.
- Barrett, T. E., et al. (1999). "Crystal structure of a thwarted mismatch glycosylase DNA repair complex." EMBO J **18**(23): 6599-6609.
- Becker, P. B. and W. Horz (2002). "ATP-dependent nucleosome remodeling." Annu Rev Biochem **71**: 247-273.
- Bentivegna, S. S. and E. Bresnick (1994). "Inhibition of human O6-methylguanine-DNA methyltransferase by 5-methylcytosine." Cancer Res **54**(2): 327-329.
- Bernstein, B. E., et al. (2007). "The mammalian epigenome." Cell **128**(4): 669-681.
- Bhattacharya, S. K., et al. (1999). "A mammalian protein with specific demethylase activity for mCpG DNA." Nature **397**(6720): 579-583.
- Bhutani, N., et al. (2010). "Reprogramming towards pluripotency requires AID-dependent DNA demethylation." Nature **463**(7284): 1042-1047.
- Bird, A., et al. (1985). "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA." Cell **40**(1): 91-99.
- Black, J. C., et al. (2012). "Histone lysine methylation dynamics: establishment, regulation, and biological impact." Mol Cell **48**(4): 491-507.
- Boland, M. J. and J. K. Christman (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**(3): 492-504.
- Boorstein, R. J., et al. (2001). "Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1." J Biol Chem **276**(45): 41991-41997.
- Borgel, J., et al. (2010). "Targets and dynamics of promoter DNA methylation during early mouse development." Nat Genet **42**(12): 1093-1100.
- Caldecott, K. W. (2003). "Protein-protein interactions during mammalian DNA single-strand break repair." Biochem Soc Trans **31**(Pt 1): 247-251.
- Cannon, S. V., et al. (1988). "5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue." Biochem Biophys Res Commun **151**(3): 1173-1179.

- Chen, D., 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**(40): 38586-38592.
- Chen, Q., et al. (2013). "TET2 promotes histone O-GlcNAcylation during gene transcription." Nature **493**(7433): 561-564.
- Chevray, P. M. and D. Nathans (1992). "Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun." Proc Natl Acad Sci U S A **89**(13): 5789-5793.
- Clouaire, T., et al. (2012). "Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells." Genes Dev **26**(15): 1714-1728.
- Cockell, M., et al. (1998). "Targeting Sir proteins to sites of action: a general mechanism for regulated repression." Cold Spring Harb Symp Quant Biol **63**: 401-412.
- Cortazar, D., et al. (2007). "The enigmatic thymine DNA glycosylase." DNA Repair (Amst) **6**(4): 489-504.
- Cortazar, D., et al. (2011). "Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability." Nature **470**(7334): 419-423.
- Cortellino, S., et al. (2011). "Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair." Cell **146**(1): 67-79.
- Creyghton, M. P., et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.
- Dalhus, B., et al. (2009). "DNA base repair--recognition and initiation of catalysis." FEMS Microbiol Rev **33**(6): 1044-1078.
- Das, A., et al. (2006). "NEIL2-initiated, APE-independent repair of oxidized bases in DNA: Evidence for a repair complex in human cells." DNA Repair (Amst) **5**(12): 1439-1448.
- Delhommeau, F., et al. (2009). "Mutation in TET2 in myeloid cancers." N Engl J Med **360**(22): 2289-2301.
- Deplus, R., et al. (2013). "TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS." EMBO J **32**(5): 645-655.
- Desterro, J. M., et al. (1998). "SUMO-1 modification of I κ B α inhibits NF- κ B activation." Mol Cell **2**(2): 233-239.
- Dong, K. B., et al. (2008). "DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity." EMBO J **27**(20): 2691-2701.
- Ehrlich, M., et al. (1990). "Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues." Mutat Res **238**(3): 277-286.
- Essen, L. O. and T. Klar (2006). "Light-driven DNA repair by photolyases." Cell Mol Life Sci **63**(11): 1266-1277.
- Falnes, P. O., et al. (2007). "Repair of methyl lesions in DNA and RNA by oxidative demethylation." Neuroscience **145**(4): 1222-1232.
- Felsenfeld, G. and M. Groudine (2003). "Controlling the double helix." Nature **421**(6921): 448-453.
- Fernandez-Capetillo, O., et al. (2004). "H2AX: the histone guardian of the genome." DNA Repair (Amst) **3**(8-9): 959-967.
- Figueroa, M. E., et al. (2010). "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation." Cancer Cell **18**(6): 553-567.

- Fischer, J. A., et al. (2004). "Proteolytic degradation of the nuclear isoform of uracil-DNA glycosylase occurs during the S phase of the cell cycle." DNA Repair (Amst) **3**(5): 505-513.
- Fortini, P. and E. Dogliotti (2007). "Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways." DNA Repair (Amst) **6**(4): 398-409.
- Frank, D., et al. (1991). "Demethylation of CpG islands in embryonic cells." Nature **351**(6323): 239-241.
- Fremont, M., et al. (1997). "Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA." Nucleic Acids Res **25**(12): 2375-2380.
- Friedberg, E. C. (2003). "DNA damage and repair." Nature **421**(6921): 436-440.
- Fuks, F., et al. (2003). "The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase." Nucleic Acids Res **31**(9): 2305-2312.
- Gallinari, P. and J. Jiricny (1996). "A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase." Nature **383**(6602): 735-738.
- Gao, M. and M. Karin (2005). "Regulating the regulators: control of protein ubiquitination and ubiquitin-like modifications by extracellular stimuli." Mol Cell **19**(5): 581-593.
- Geiss-Friedlander, R. and F. Melchior (2007). "Concepts in sumoylation: a decade on." Nat Rev Mol Cell Biol **8**(12): 947-956.
- Gill, G. (2004). "SUMO and ubiquitin in the nucleus: different functions, similar mechanisms?" Genes Dev **18**(17): 2046-2059.
- Ginno, P. A., et al. (2012). "R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters." Mol Cell **45**(6): 814-825.
- Gokul, G. and S. Khosla (2012). "DNA methylation and cancer." Subcell Biochem **61**: 597-625.
- Goodman, R. H. and S. Smolik (2000). "CBP/p300 in cell growth, transformation, and development." Genes Dev **14**(13): 1553-1577.
- Gregory, D. J., et al. (2012). "Selective DNA demethylation by fusion of TDG with a sequence-specific DNA-binding domain." Epigenetics **7**(4): 344-349.
- Grewal, S. I. and D. Moazed (2003). "Heterochromatin and epigenetic control of gene expression." Science **301**(5634): 798-802.
- Gronbaek, K., et al. (2007). "Epigenetic changes in cancer." APMIS **115**(10): 1039-1059.
- Gu, T. P., et al. (2011). "The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes." Nature **477**(7366): 606-610.
- Guo, J. U., et al. (2011). "Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain." Cell **145**(3): 423-434.
- Hackett, J. A., et al. (2013). "Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine." Science **339**(6118): 448-452.
- Hajkova, P., et al. (2010). "Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway." Science **329**(5987): 78-82.
- Hake, S. B. and C. D. Allis (2006). "Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis"." Proc Natl Acad Sci U S A **103**(17): 6428-6435.
- Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.
- Hardeland, U., et al. (2000). "Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis." J Biol Chem **275**(43): 33449-33456.

- Hardeland, U., et al. (2003). "The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs." Nucleic Acids Res **31**(9): 2261-2271.
- Hardeland, U., et al. (2001). "Thymine DNA glycosylase." Prog Nucleic Acid Res Mol Biol **68**: 235-253.
- Hardeland, U., et al. (2007). "Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2." Nucleic Acids Res **35**(11): 3859-3867.
- 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.
- Harris, R. S., et al. (2003). "DNA deamination mediates innate immunity to retroviral infection." Cell **113**(6): 803-809.
- He, Y. F., et al. (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." Science **333**(6047): 1303-1307.
- Hecker, C. M., et al. (2006). "Specification of SUMO1- and SUMO2-interacting motifs." J Biol Chem **281**(23): 16117-16127.
- Hendrich, B., et al. (2001). "Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development." Genes Dev **15**(6): 710-723.
- Hu, Q. and M. G. Rosenfeld (2012). "Epigenetic regulation of human embryonic stem cells." Front Genet **3**: 238.
- Huang, J. C., et al. (1994). "Substrate spectrum of human excinuclease: repair of abasic sites, methylated bases, mismatches, and bulky adducts." Proc Natl Acad Sci U S A **91**(25): 12213-12217.
- Hubner, M. R., et al. (2012). "Chromatin organization and transcriptional regulation." Curr Opin Genet Dev.
- Illingworth, R. S. and A. P. Bird (2009). "CpG islands--'a rough guide'." FEBS Lett **583**(11): 1713-1720.
- Imai, K., et al. (2003). "Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination." Nat Immunol **4**(10): 1023-1028.
- Inoue, A., et al. (2011). "Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development." Cell Res **21**(12): 1670-1676.
- Inoue, A. and Y. Zhang (2011). "Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos." Science **334**(6053): 194.
- Ito, S., et al. (2010). "Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification." Nature **466**(7310): 1129-1133.
- Ito, S., et al. (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." Science **333**(6047): 1300-1303.
- Jacobs, A. L. and P. Schar (2012). "DNA glycosylases: in DNA repair and beyond." Chromosoma **121**(1): 1-20.
- Jenkins, T. G. and D. T. Carrell (2012). "Dynamic alterations in the paternal epigenetic landscape following fertilization." Front Genet **3**: 143.
- Johnson, M. H. and C. A. Ziemek (1981). "The foundation of two distinct cell lineages within the mouse morula." Cell **24**(1): 71-80.
- Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nat Rev Genet **13**(7): 484-492.
- Jones, P. A. and G. Liang (2009). "Rethinking how DNA methylation patterns are maintained." Nat Rev Genet **10**(11): 805-811.

- 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**(10): 4684-4688.
- Jost, J. P., et al. (1997). "The RNA moiety of chick embryo 5-methylcytosine- DNA glycosylase targets DNA demethylation." Nucleic Acids Res **25**(22): 4545-4550.
- Jost, J. P., et al. (1999). "A chicken embryo protein related to the mammalian DEAD box protein p68 is tightly associated with the highly purified protein-RNA complex of 5-MeC-DNA glycosylase." Nucleic Acids Res **27**(16): 3245-3252.
- Jost, J. P., et al. (1995). "Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase." J Biol Chem **270**(17): 9734-9739.
- Kadyrov, F. A., et al. (2006). "Endonucleolytic function of MutLalpha in human mismatch repair." Cell **126**(2): 297-308.
- Kafri, T., et al. (1992). "Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line." Genes Dev **6**(5): 705-714.
- Kamileri, I., et al. (2012). "Nucleotide excision repair: new tricks with old bricks." Trends Genet **28**(11): 566-573.
- Kangaspeska, S., et al. (2008). "Transient cyclical methylation of promoter DNA." Nature **452**(7183): 112-115.
- Kemmerich, K., et al. (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**(13): 6016-6025.
- Kim, N. and S. Jinks-Robertson (2010). "A basic sites in the transcribed strand of yeast DNA are removed by transcription-coupled nucleotide excision repair." Mol Cell Biol **30**(13): 3206-3215.
- Kim, Y. J. and D. M. Wilson, 3rd (2012). "Overview of base excision repair biochemistry." Curr Mol Pharmacol **5**(1): 3-13.
- Kunz, C., et al. (2009). "Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil." PLoS Biol **7**(4): e91.
- Kuzmichev, A., et al. (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein." Genes Dev **16**(22): 2893-2905.
- Langemeijer, S. M., et al. (2009). "Acquired mutations in TET2 are common in myelodysplastic syndromes." Nat Genet **41**(7): 838-842.
- Langerak, P. and P. Russell (2011). "Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair." Philos Trans R Soc Lond B Biol Sci **366**(1584): 3562-3571.
- Lanzuolo, C. and V. Orlando (2012). "Memories from the polycomb group proteins." Annu Rev Genet **46**: 561-589.
- Laurent, L., et al. (2010). "Dynamic changes in the human methylome during differentiation." Genome Res **20**(3): 320-331.
- Le May, N., et al. (2012). "XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes." Mol Cell **47**(4): 622-632.
- Le May, N., et al. (2010). "NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack." Mol Cell **38**(1): 54-66.

- Lee, B. K. and V. R. Iyer (2012). "Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation." J Biol Chem **287**(37): 30906-30913.
- Li, E., et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-926.
- Li, Y. Q., et al. (2007). "Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair." Nucleic Acids Res **35**(2): 390-400.
- Li, Z., et al. (2011). "Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies." Blood **118**(17): 4509-4518.
- Lin, J. J. and A. Sancar (1989). "A new mechanism for repairing oxidative damage to DNA: (A)BC excinuclease removes AP sites and thymine glycols from DNA." Biochemistry **28**(20): 7979-7984.
- Lister, R., et al. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-322.
- Lomeli, H. and M. Vazquez (2011). "Emerging roles of the SUMO pathway in development." Cell Mol Life Sci **68**(24): 4045-4064.
- Lorsbach, R. B., et al. (2003). "TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23)." Leukemia **17**(3): 637-641.
- Maiti, A. and A. C. Drohat (2011). "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites." J Biol Chem **286**(41): 35334-35338.
- Maiti, A., et al. (2008). "Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition." Proc Natl Acad Sci U S A **105**(26): 8890-8895.
- Margueron, R., et al. (2009). "Role of the polycomb protein EED in the propagation of repressive histone marks." Nature **461**(7265): 762-767.
- Marmorstein, R. and R. C. Trievel (2009). "Histone modifying enzymes: structures, mechanisms, and specificities." Biochim Biophys Acta **1789**(1): 58-68.
- Masson, M., et al. (1998). "XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage." Mol Cell Biol **18**(6): 3563-3571.
- Meilinger, D., et al. (2009). "Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells." EMBO Rep **10**(11): 1259-1264.
- Metivier, R., et al. (2008). "Cyclical DNA methylation of a transcriptionally active promoter." Nature **452**(7183): 45-50.
- Millar, C. B., et al. (2002). "Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice." Science **297**(5580): 403-405.
- Minty, A., et al. (2000). "Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif." J Biol Chem **275**(46): 36316-36323.
- Missero, C., et al. (2001). "The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription." J Biol Chem **276**(36): 33569-33575.
- Mohan, R. D., et al. (2010). "Opposing regulatory roles of phosphorylation and acetylation in DNA mispair processing by thymine DNA glycosylase." Nucleic Acids Res **38**(4): 1135-1148.

- Mohan, R. D., et al. (2007). "SUMO-1-dependent allosteric regulation of thymine DNA glycosylase alters subnuclear localization and CBP/p300 recruitment." Mol Cell Biol **27**(1): 229-243.
- Morgan, M. T., et al. (2011). "Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA." Nucleic Acids Res **39**(6): 2319-2329.
- Nabel, C. S., et al. (2012). "AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation." Nat Chem Biol **8**(9): 751-758.
- Nakamura, T., et al. (2007). "PGC7/Stella protects against DNA demethylation in early embryogenesis." Nat Cell Biol **9**(1): 64-71.
- Neddermann, P. and J. Jiricny (1993). "The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells." J Biol Chem **268**(28): 21218-21224.
- Nilsen, H., et al. (2003). "Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas." Oncogene **22**(35): 5381-5386.
- Okano, M., et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.
- Ooi, S. K., et al. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-717.
- Otterlei, M., et al. (1999). "Post-replicative base excision repair in replication foci." EMBO J **18**(13): 3834-3844.
- Pasini, D., et al. (2010). "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**(15): 4958-4969.
- Pavri, R. and M. C. Nussenzweig (2011). "AID targeting in antibody diversity." Adv Immunol **110**: 1-26.
- Pegg, A. E. (2000). "Repair of O(6)-alkylguanine by alkyltransferases." Mutat Res **462**(2-3): 83-100.
- Petitjean, A., et al. (2007). "Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database." Hum Mutat **28**(6): 622-629.
- Ponger, L., et al. (2001). "Determinants of CpG islands: expression in early embryo and isochore structure." Genome Res **11**(11): 1854-1860.
- Popp, C., et al. (2010). "Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency." Nature **463**(7284): 1101-1105.
- Rada, C., et al. (2002). "Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice." Curr Biol **12**(20): 1748-1755.
- Radany, E. H., et al. (2000). "Increased spontaneous mutation frequency in human cells expressing the phage PBS2-encoded inhibitor of uracil-DNA glycosylase." Mutat Res **461**(1): 41-58.
- Rai, K., et al. (2008). "DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45." Cell **135**(7): 1201-1212.
- Raiber, E. A., et al. (2012). "Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase." Genome Biol **13**(8): R69.
- Ramsahoye, B. H., et al. (2000). "Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a." Proc Natl Acad Sci U S A **97**(10): 5237-5242.
- Rubin, A. F. and P. Green (2009). "Mutation patterns in cancer genomes." Proc Natl Acad Sci U S A **106**(51): 21766-21770.

- Santos, F., et al. (2002). "Dynamic reprogramming of DNA methylation in the early mouse embryo." *Dev Biol* **241**(1): 172-182.
- Schar, P. and O. Fritsch (2011). "DNA repair and the control of DNA methylation." *Prog Drug Res* **67**: 51-68.
- Scharer, O. D. (2003). "Chemistry and biology of DNA repair." *Angew Chem Int Ed Engl* **42**(26): 2946-2974.
- Schaub, F. X., et al. (2010). "Clonal analysis of TET2 and JAK2 mutations suggests that TET2 can be a late event in the progression of myeloproliferative neoplasms." *Blood* **115**(10): 2003-2007.
- Schmitz, K. M., et al. (2009). "TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation." *Mol Cell* **33**(3): 344-353.
- Schuettengruber, B., et al. (2011). "Trithorax group proteins: switching genes on and keeping them active." *Nat Rev Mol Cell Biol* **12**(12): 799-814.
- Seisenberger, S., et al. (2013). "Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers." *Philos Trans R Soc Lond B Biol Sci* **368**(1609): 20110330.
- Serandour, A. A., et al. (2012). "Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers." *Nucleic Acids Res* **40**(17): 8255-8265.
- Sharif, J., et al. (2007). "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." *Nature* **450**(7171): 908-912.
- Shen, T. H., et al. (2006). "The mechanisms of PML-nuclear body formation." *Mol Cell* **24**(3): 331-339.
- Siegel, R. M., et al. (2000). "Measurement of molecular interactions in living cells by fluorescence resonance energy transfer between variants of the green fluorescent protein." *Sci STKE* **2000**(38): pl1.
- Sjoblom, T., et al. (2006). "The consensus coding sequences of human breast and colorectal cancers." *Science* **314**(5797): 268-274.
- Smet-Nocca, C., et al. (2011). "SUMO-1 regulates the conformational dynamics of thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity." *BMC Biochem* **12**: 4.
- Song, J., et al. (2004). "Identification of a SUMO-binding motif that recognizes SUMO-modified proteins." *Proc Natl Acad Sci U S A* **101**(40): 14373-14378.
- Soutoglou, E., et al. (2007). "Positional stability of single double-strand breaks in mammalian cells." *Nat Cell Biol* **9**(6): 675-682.
- St Clair, M. H., et al. (1987). "Inhibition by ganciclovir of cell growth and DNA synthesis of cells biochemically transformed with herpesvirus genetic information." *Antimicrob Agents Chemother* **31**(6): 844-849.
- Stadler, M. B., et al. (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." *Nature* **480**(7378): 490-495.
- Steinacher, R. and P. Schar (2005). "Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation." *Curr Biol* **15**(7): 616-623.
- Swierczek, S. I., et al. (2011). "Extent of hematopoietic involvement by TET2 mutations in JAK2V(6)(1)(7)F polycythemia vera." *Haematologica* **96**(5): 775-778.
- Szabo, P. E. and G. P. Pfeifer (2012). "H3K9me2 attracts PGC7 in the zygote to prevent Tet3-mediated oxidation of 5-methylcytosine." *J Mol Cell Biol* **4**(6): 427-429.
- Tahiliani, M., et al. (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." *Science* **324**(5929): 930-935.

- Takahashi, H., et al. (2005). "Noncovalent SUMO-1 binding activity of thymine DNA glycosylase (TDG) is required for its SUMO-1 modification and colocalization with the promyelocytic leukemia protein." *J Biol Chem* **280**(7): 5611-5621.
- Talbert, P. B. and S. Henikoff (2010). "Histone variants--ancient wrap artists of the epigenome." *Nat Rev Mol Cell Biol* **11**(4): 264-275.
- Tan, L. and Y. G. Shi (2012). "Tet family proteins and 5-hydroxymethylcytosine in development and disease." *Development* **139**(11): 1895-1902.
- Tefferi, A., et al. (2009). "Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML." *Leukemia* **23**(7): 1343-1345.
- Tefferi, A., et al. (2009). "Mutation in TET2 in myeloid cancers." *N Engl J Med* **361**(11): 1117; author reply 1117-1118.
- Tini, M., et al. (2002). "Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription." *Mol Cell* **9**(2): 265-277.
- Tsumura, A., et al. (2006). "Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b." *Genes Cells* **11**(7): 805-814.
- Ulrich, H. D. (2005). "Mutual interactions between the SUMO and ubiquitin systems: a plea of no contest." *Trends Cell Biol* **15**(10): 525-532.
- Um, S., et al. (1998). "Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase." *J Biol Chem* **273**(33): 20728-20736.
- Valinluck, V. and L. C. Sowers (2007). "Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1." *Cancer Res* **67**(3): 946-950.
- Vella, P., et al. (2013). "Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells." *Mol Cell* **49**(4): 645-656.
- Vire, E., et al. (2006). "The Polycomb group protein EZH2 directly controls DNA methylation." *Nature* **439**(7078): 871-874.
- Waters, T. R., et al. (1999). "Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1." *J Biol Chem* **274**(1): 67-74.
- Wiebauer, K. and J. Jiricny (1989). "In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells." *Nature* **339**(6221): 234-236.
- Wiederhold, L., et al. (2004). "AP endonuclease-independent DNA base excision repair in human cells." *Mol Cell* **15**(2): 209-220.
- Williams, K., et al. (2012). "DNA methylation: TET proteins-guardians of CpG islands?" *EMBO Rep* **13**(1): 28-35.
- Wong, E., et al. (2002). "Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation." *Proc Natl Acad Sci U S A* **99**(23): 14937-14942.
- Wood, L. D., et al. (2007). "The genomic landscapes of human breast and colorectal cancers." *Science* **318**(5853): 1108-1113.
- Wossidlo, M., et al. (2011). "5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming." *Nat Commun* **2**: 241.
- Yang, B., et al. (2007). "Virion-associated uracil DNA glycosylase-2 and apurinic/aprimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA." *J Biol Chem* **282**(16): 11667-11675.

- Zhou, J., et al. (2008). "Thymine DNA glycosylase represses myocardin-induced smooth muscle cell differentiation by competing with serum response factor for myocardin binding." J Biol Chem **283**(51): 35383-35392.
- Zhu, B. and D. Reinberg (2011). "Epigenetic inheritance: uncontested?" Cell Res **21**(3): 435-441.
- Zhu, B., et al. (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**(10): 5135-5139.
- Zhu, J. K. (2009). "Active DNA demethylation mediated by DNA glycosylases." Annu Rev Genet **43**: 143-166.

Appendix:

- I. Measuring the SUMO-Interaction Dynamics of the Thymine DNA Glycosylase by Fluorescence Resonance Energy Transfer
- II. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic
- III. TDG maintains a transitory equilibrium of CpG island methylation and oxidative demethylation during cell differentiation
- IV. DNA Glycosylases: In DNA Repair and Beyond

Measuring the SUMO-Interaction Dynamics of the Thymine DNA Glycosylase by Fluorescence Resonance Energy Transfer

Angelika L. Jacobs, David Schürmann, Yusuke Saito, Primo Schär*

Department of Biomedicine, University of Basel, Mattenstrasse 28, CH-4058 Basel, Switzerland

*To whom correspondence should be addressed: E-mail primo.schaer@unibas.ch; Tel. +41 61 267 3561; Fax: +41 61 267 3566

Author contributions:

A.L.J. generated the FRET constructs, performed Western Blot analysis, base release assay, fibroblast cell culture, microscopy, image processing, FRET analysis and statistics; A.L.J. and D.S. designed the FRET constructs; Y.S. cultured and transfected ES cells and prepared native protein extracts. P.S. designed, coordinated and supervised the study.

Abstract

The Thymine DNA Glycosylase (TDG) initiates Base Excision Repair (BER) of single-base damage, preferentially the deamination products of cytosine and 5-methylcytosine, uracil and thymine, respectively. By hydrolyzing the N-glycosidic bond between the damaged base and the deoxyribose, TDG produces an abasic (AP) site to which TDG remains firmly attached. We have previously shown that modification of TDG with Small Ubiquitin-like Modifiers (SUMOs) facilitates the dissociation of TDG from its product AP-site by inducing a conformational change from a closed to an open conformation. Since this hypothesis was based solely on biochemical studies, the significance of the SUMO-TDG interaction for BER *in vivo* remains unclear. Here, we present a Fluorescence Resonance Energy Transfer (FRET) system that allows monitoring the SUMO-interactions of TDG in living cells, and we provide evidence that the interaction dynamics measured by FRET can be altered by inducing DNA damage. Furthermore, we established a FRET system for the dimerization of TDG and report that part of the cellular TDG pool exists as homodimers.

Introduction

The DNA encoding all genetic information is under constant risk of damage by deamination, oxidation or alkylation. Most frequently, such chemical modifications concern the DNA bases, which are then corrected through the Base Excision repair (BER) system. BER is initiated by DNA glycosylases which recognize the damage base in the DNA and flip it out of the DNA double helix into a catalytic pocket for hydrolysis of the N-glycosidic bond (Lindahl and Wood 1999). The resulting abasic (AP) site is then further processed by an AP endonuclease (APE1) producing a single-strand break which is subsequently repaired through short-patch or long-patch BER (Fortini and Dogliotti 2010).

The Thymine DNA Glycosylase (TDG) was first discovered by its ability to excise the deamination products of cytosine (C) and 5-methylcytosine (5-mC), uracil and thymine mispaired with guanine, respectively, but its substrate spectrum encompasses a broad range of pyrimidine derivatives, e.g. 5-fluorouracil (5-FU) or 5-hydroxymethyluracil (5-hmU) (Cortazar et al. 2007; Jacobs and Schar 2012). Recently, 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), the oxidation products of the Ten Eleven Translocator (TET) family of 5-mC hydroxylases, have been shown to be efficiently excised from DNA by TDG, suggesting an involvement of TDG in the epigenetic regulation of gene expression through a putative DNA demethylation pathway (He et al. 2011; Maiti and Drohat 2011). In this context, it is noteworthy that among the interaction partners of TDG are not only DNA repair factors but also transcription factors like the retinoic acid receptor (RAR) or the estrogen receptor α (ER α) and the *de novo* DNA methyltransferases Dnmt3a and b (Cortazar et al. 2007).

The search for interaction partners of TDG has revealed an involvement of post-translational modification with Small Ubiquitin-like Modifiers (SUMOs) in TDG-initiated BER (Hardeland et al. 2002). SUMOylation represents a fast and (in most cases) reversible way to regulate protein stability, subcellular localization, enzymatic activity or interaction dynamics. The SUMO protein family includes three genes in mammals, SUMO1, 2 and 3; the latter two share 97% sequence identity and are often referred to as SUMO2/3. While most of the cellular SUMO1 pool appears to be constitutively attached to target proteins, SUMO2/3 appear to become conjugated to target proteins primarily as a reaction to environmental stress (Saitoh and Hinchey 2000). A fourth gene, encoding SUMO4, is present only in the human genome and is expressed mostly in lymph nodes, kidney and spleen (Geiss-Friedlander and Melchior 2007).

SUMO is synthesized as an inactive precursor and processed by a SUMO-specific isopeptidase (SENtrin-specific Protease, SENP) to yield a glycine-glycine motif at the C-terminus, which can form an isopeptide bond with an acceptor lysine of a target protein. The SUMO proteins are covalently attached to their respective target protein in a multi-step process, initiated by a SUMO activating enzyme E1 (heterodimer of SAE1/SAE2) which adenylates SUMO, facilitating the formation of a

thioester bond with SAE2. SUMO is then transferred to the SUMO-conjugating enzyme (E2) UBC9, and further to a lysine residue in the target protein, this last step often but not always being mediated by SUMO E3 ligases. The E3 ligase responsible for SUMOylating TDG has so far remained elusive. The modification is reversible through the action of SENPs (Geiss-Friedlander and Melchior 2007).

A majority of the SUMO-targets appear to be nuclear proteins, many of which are involved in DNA replication and transcription or in the regulation of chromosome structure and dynamics, suggesting SUMO to be a key player in genome maintenance and stability (Gill 2004). Moreover, knockout of any of the non-redundant components of the SUMO conjugation pathway has been reported to be embryonic lethal, suggesting a crucial role of SUMOylation in development (Lomeli and Vazquez 2011).

Apart from forming covalent bonds, SUMO has been shown to interact non-covalently with proteins containing a so-called SUMO Interaction/Binding Motif (SIM or SBM), which consist of a hydrophobic core, flanked N- or C-terminally by acidic and/or serine residues (Minty et al. 2000; Song et al. 2004). TDG interacts covalently as well as non-covalently with SUMO1 and SUMO2/3 (Hardeland et al. 2002). The SUMO acceptor lysine K341 (in murine splice-variant TdgA) lies within a C-terminal SUMOylation consensus motif (VKEE). Additionally, two SBMs have been identified in TDG, one in the N- and one in the C-terminal domain (Mohan et al. 2007).

Extensive biochemical studies have revealed that SUMOylation induces a conformational change in TDG that regulates the dynamics of the initiation of BER (Steinacher and Schar 2005). Upon binding to homoduplex DNA, the N-terminal domain of TDG forms a clamp-like structure that allows the glycosylase to slide along the DNA in search of a substrate base. When encountering a damaged base and flipping it into the catalytic pocket, specific interactions with the opposite guanine in addition to the homoduplex DNA-binding activity of the N-terminal domain cooperate to bind TDG firmly to the substrate for excision, after which TDG stays firmly bound to the AP-site (Waters et al. 1999; Hardeland et al. 2000). SUMOylation of the C-terminus facilitates the dissociation of TDG by opening

the clamp-like conformation, thus neutralizing the homoduplex binding activity of the N-terminus (Hardeland et al. 2002; Steinacher and Schar 2005). Downstream-acting repair factors like the AP-endonuclease1 (APE1) and XRCC1 stimulate SUMOylation of TDG, suggesting that AP-site binding may serve the purpose of stabilization of these hazardously fragile sites until the BER complex is in place for hand-over of the repair intermediate ((Steinacher and Schar 2005) and unpublished data).

To confirm this function of TDG SUMOylation as well as the conformational switch of TDG between an open and a closed form *in vivo*, we developed a tool for monitoring the SUMO-interactions and the conformational changes of murine TDG in live cells. We established a Fluorescence Resonance Energy Transfer (FRET) system, which allows measuring the interaction between two proteins by fusing one to a donor fluorophore and one to an acceptor fluorophore with a suitable absorption spectrum. If the two proteins of interest bring the fluorophores in close proximity (10 to 80 Å) and a favorable orientation, an energy quantum is transferred from the donor in its excited state to the acceptor fluorophore. The thus excited acceptor emits light in a wavelength beyond the emission spectrum of the donor. Here, we report that the inter- and intramolecular interaction dynamics of TDG can be measured *in vivo* by FRET and that the SUMO-interaction dynamics can be altered by inducing DNA damage.

Materials and Methods

Vector construction

The FRET constructs were assembled by standard cloning techniques in IRES constructs holding either the chicken actin promoter and a Puromycin resistance cassette downstream of the IRES (pCAIp used for TDG constructs, the positive control Citrine-Cerulean and Citrine) or the CMV promoter and a Hygromycin resistance (pCMIh, analogous to pIRESHyg, Clontech, used for SUMO constructs and Cerulean). Murine TDG, SUMO and the fluorophores were PCR amplified with adaptor primers

holding the required restriction sites. All oligonucleotides used for cloning can be found in Supplementary Table 2. A schematic overview of the cloning strategy is shown in Supplementary Fig. 1.

The sequence encoding for the wildtype Tdg (murine splice variant A, mTdgA) or the SUMOylation-deficient mutant mTdgA K341R (TDG Δ S) was inserted into the *EcoRI* and *BamHI* sites of pCAIp, introducing additional restriction sites for *NheI*, *BsrGI*, *XhoI* and *Clal*. For N-terminal fusion with EGFP or Citrine, the respective coding sequence was inserted into the *EcoRI* and *BsrGI* sites upstream of TDG. C-terminal fusion was accomplished by introducing the fluorophore sequence into the downstream *XhoI* and *BamHI* sites. The DNA fragment encoding for Cerulean was digested with *XbaI* and *NheI* and inserted into the *NheI* site at the 5' end of *TdgA*. For the C-terminal fusion, the coding sequence was inserted into the *XhoI* and *Clal* sites 3' of *TdgA*.

The sequences encoding human SUMO1 and 3 (identical with murine proteins at the amino acid level) with a glycine-glycine motif at the C-terminus were PCR-amplified from pCDNA3-HA-SUMO1GG or pCDNA3-HA-SUMO3GG (kindly provided by Ron Hay), respectively, and inserted into the *EcoRV* and *NotI* sites of pCMIh, additionally introducing *XhoI*, *NheI* and *BsrGI* sites. The Cerulean encoding sequence was introduced into the *XhoI* and *NheI* sites upstream of the SUMO1/3 encoding sequence. The FRET positive control vector pCAIp Citrine-Cerulean was derived from a vector encoding double tagged TDG (pCAIp Cit-TDG-Ceru), by replacing TDG with a linker sequence encoding VQSGGDASGGSSST. The negative controls pCAIp Citrine and pCMIh Cerulean were assembled through introduction of the fluorophore sequences into the *EcoRI* and *BamHI* sites of the respective expression vector. All PCR-amplified fragments were validated by sequencing.

Cell culture and transfection

All cell lines were incubated at 37°C, 5% CO₂ and 100% humidity. Cos7 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FCS and 2 mM L-glutamine. Murine ES cells were cultivated on feeder cells in DMEM with 15% heat-inactivated FCS, 2mM L-

glutamine, 1x sodium pyruvate, 1x MEM non-essential amino acids, 0.1mM β -mercaptoethanol and 1000U/mL LIF. Transfections were performed using TransFectin™ (BioRad, Switzerland) or JetPEI™ (Polyplus Transfection, France) according to the manufacturer's manuals. ES cells were transfected with TransFectin in suspension after feeder-removal and subsequently plated for further incubation. DNA damage was induced by adding 30 μ M 5-FU or 10 μ M BrdU to the culture medium followed by 24-28h of incubation.

Protein extraction

Denaturing SDS-extracts were prepared 40-48 h after transfection by adding 100 μ L of SDS sample buffer (45 mM Tris-HCl pH8, 10% glycerol, 1% SDS, 0.01% bromophenol blue, 50 mM DTT) to one 35 mm plate, scraping the cells off the surface and transferring the lysate to an Eppendorf tube. The samples were heated to 95°C for 10 min and centrifuged at 16'000xg and 4°C for 10 min.

For native protein extracts, mammalian cells were harvested, washed with PBS and lysed in an appropriate volume of NP40 lysis buffer (50 mM sodium phosphate buffer pH8, 125 mM NaCl, 0.5 mM EDTA pH8, 1% NP40, 1 mM PMSF, 1 mM DTT, 1x Complete EDTA-free Protease Inhibitor Cocktail) by vortexing and incubating them on ice for 30 min. To increase cell breakage, samples were sonicated 5 times for 20 sec with intervals of 40 sec. After centrifuging at 16'000xg and 4°C for 20 min, the supernatants were shock-frozen and stored at -80°C. The Bradford Protein Assay (BioRad, Switzerland) was used to determine the approximate protein concentration in the cell lysates according to the manufacturer's instructions.

Denaturing Gel electrophoresis and Western Blotting

Gels were cast using the Mini-PROTEAN® 3 system (BioRad, Switzerland). For one 0.75 mm minigel, 5 mL of separating gel were prepared (375 mM Tris-HCl pH8.8, 0.1% SDS, x% Acryl/Bis™ 37.5:1, 0.1% APS, 0.05% TEMED). To facilitate homogeneous polymerization, the gel solution was overlaid with 2-propanol. After complete polymerization of the separating gel, the 2-propanol was removed and the

stacking gel (125 mM Tris-HCl pH6.8, 0.1% SDS, 0.05% APS, 0.1% TEMED) was cast. 10-15 μ L of protein samples were loaded in SDS sample buffer (45mM Tris-HCL pH 8, 10% SDS, 0.01% bromphenol blue, 50mM DTT) and separated at 30 mA for 1 hour.

SDS-PAGE separated proteins were transferred onto a nitrocellulose membrane (Schleicher & Schuell Bioscience) by the wet-blot technique using the Mini-Transblot system (BioRad, Switzerland). Blotting was done at 1 mA per cm^2 gel in pre-cooled transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol) at 4°C over night. The membrane was blocked with 10% blocking milk in TBS (100 mM Tris-HCl pH8, 150 mM NaCl) followed by incubation with the two primary antibodies in 7.5% blocking milk in TBS + 0.2% Tween20 (TBST) at 33°C for 1 h (anti-TDG 1:10'000; anti-GFP 1:1'000, Roche #11814460001; anti- β -actin 1:10'000, Abcam ab8226). After two rinses in TBST, the membrane was washed three times for 10 min in TBST, once at 33°C, twice at room temperature. The fluorescence-labeled secondary antibodies (goat anti-rabbit IRDye 800CW, LICOR® P/N 926-32211; goat anti-mouse IRDye 680; LICOR® P/N 926-32220) were diluted 1:5'000 in 5% blocking milk in TBST and incubated on the membrane for 30 min, protected from light. After three washes in TBST, the membrane was rinsed three times with TBS to remove residual Tween20 and scanned with a LICOR® Odyssey scanner (LI-COR, Germany).

Base release assay

The heteroduplex G•T-substrate was prepared by annealing the oligonucleotides "Subs60uG" and the FITC-labeled "Subs60IT-F" (Supplementary Table 2). The reaction mixture for the base release assay contained 1x nicking buffer (50 mM Tris-HCl pH8, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA), 1 pmol heteroduplex G•T-substrate and 100 μ g NP40 protein extract in a total volume of 50 μ L. After an overnight incubation at 37°C, the reaction was stopped by adding 10 μ g Proteinase K and incubating at 37°C for another 30 min. To induce a single-strand break at the product AP-site, 5 μ L of 1 N NaOH was added, the sample was mixed well and heated to 99°C for 10 min. Afterwards, the

DNA was precipitated by adding 5 μ L 3 M sodium acetate pH5.2, 150 μ L of 99% ethanol and incubating at -20°C for 2 h.

After centrifugation at 4°C and 16'000xg for 20 minutes, the supernatant was discarded and the pellet washed with 180 μ L of 80% ethanol. After another 5 min of centrifugation, the supernatant was discarded and the pellet air-dried at room temperature for 5 min. The pellet was finally resuspended in 10 μ L formamide buffer (90% formamide, 1x TBE), heated to 99°C for 5 min to denature the DNA double-strands, and chilled on ice. The cold samples were loaded onto a pre-warmed and pre-run 15% denaturing DNA gel (2M urea, 500 μ L H₂O, 1 mL 5x TBE, 1.88 mL 40% PAA 19:1, 0.05% APS, 7.5 μ L TEMED) in pre-warmed buffer. After running the gel at 200 V for 30 min, labeled DNA was detected with the blue fluorescence mode of the Typhoon 9400 (GE Healthcare, Switzerland) and analyzed by ImageQuant TL software (v7.0 GE Healthcare).

Fluorescent Microscopy and FRET analysis

Fluorescent images were acquired on a Leica DMI 6000B equipped for live cell imaging with a temperature-controlled incubation chamber, using the MetaMorph® software. Cos7 cells transiently expressing the FRET constructs were imaged in 3 channels, the CFP- (Cerulean), the YFP- (Citrine) and the FRET-channel. The specifications of the filter cubes used for acquisition are listed in Supplementary Table 3. Settings were adjusted first roughly on TDG-Citrine and Cerulean-SUMO expressing cells and then fine-tuned for each channel to render equal brightness for equal amounts of fluorophore using the positive control Citrine-Cerulean. Spectral bleed-through was measured on images of cos7 cells expressing either only the donor or only the acceptor fluorophore and calculated as $BT_X = I_{FRET} / I_X$ (I=intensity, X= CFP or YFP) (Xia and Liu 2001). Images were processed in ImageJ using the PixFRET plugin (Feige et al. 2005) which calculates FRET values pixel by pixel as $FRET = I_{FRET} - BT_{CFP} * I_{CFP} - BT_{YFP} * I_{YFP}$, normalizes for different intensity levels of the donor and acceptor using the formula $NFRET = FRET / \sqrt{I_{CFP} * I_{YFP}}$ and visualizes the NFRET value of each pixel in grey scale. As this formula normalizes only within certain limits of different intensities, cells exceeding a ratio of 5:1

were excluded. NFRET values were measured within a random region in the cell nucleus and normalized to the average NFRET of the positive control. All graphs and statistical analyses (Mann-Whitney T-test) were generated in GraphPad Prism.

Results

Setup of a FRET system to monitor SUMO1/3 interactions and the conformational switch of TDG

Biochemical data showed that modification of TDG with SUMO1 or 2/3 induces the dissociation of the glycosylase from its product AP-site. Based on these data, we proposed SUMO modification as a means to control the hand-over of the fragile AP-site repair intermediate to downstream acting BER factors (Hardeland et al. 2002; Steinacher and Schar 2005). To test whether SUMOylation is indeed associated with TDG function in the context of DNA repair, we set up a Fluorescence Resonance Energy Transfer system to measure the TDG-SUMO-interaction dynamics in live cells (Fig. 1). Although crystal structures of SUMOylated TDG were available, they described the core-domain only (Baba D et al, 2005 Baba D et al JBC 2006) and therefore, were not suitable to predict an optimal placement of the GFP-tag on full-length TDG to achieve the proximity and orientation of the TDG-fused fluorophore relative to the SUMO-fused fluorophore required for FRET. For this reason, we generated constructs for N- and C-terminally Citrine-tagged mouse TdgA (Suppl. Figure 1 A).

We also wanted to test whether the predicted conformational change TDG undergoes upon binding to DNA can be monitored by FRET. For this purpose, we tagged TDG N-terminally with Citrine and C-terminally with Cerulean. Since in this context a FRET signal could also derive from the formation of TDG homodimers, we generated control constructs for Cerulean-tagged TDG in addition to the Citrine-TDG fusions described above. All fusion constructs were generated in parallel with a

SUMOylation-deficient mutant TdgA K341R (TDG Δ S) to allow covalent and non-covalent SUMO-interactions to be distinguished, and with a catalytic-dead mutant TdgA N151A (TDG Δ cat) to allow substrate and product AP-site binding to be examined separately.

The FRET donor, Cerulean, was fused to the flexible N-terminus of SUMO1 and SUMO3 with an HA-tag as a spacer and additional epitope for antibody detection (Suppl. Figure 1 B). As the C-termini of SUMO are engaged in the covalent linkage to target proteins, we refrained from generating C-terminally tagged SUMO constructs.

All available FRET constructs are listed in Supplementary Table 1, which also included additional control constructs not further described in this study.

Characterization of expression and functionality of the FRET fusion proteins

To characterize the FRET constructs functionally, we tested whether their transfection into mammalian cells resulted in the synthesis of full-length fusion proteins, whether tagged TDG could still be modified by endogenous SUMO on one hand and whether tagged SUMO could be covalently attached to tagged TDG on the other hand. Protein extracts from transiently transfected cos7 cells were subjected to SDS-PAGE and multiplex Western blotting with an anti-GFP (red channel) and an anti-TDG antibody (Fig. 2). This revealed that Citrine or Cerulean-tagged TDG (wildtype and mutants) were expressed to produce full-length proteins, migrating at approximately 85 kDa (N-terminally tagged) or 90 kDa (C-terminally tagged). Notably, the N-terminally tagged TDG fusions were consistently expressed at lower levels than the C-terminally tagged ones (Fig. 2B and C). The intramolecular FRET constructs (Citrine-TDG-Cerulean) were also expressed in full length, producing proteins of approximately 120 kDa (Fig. 2D).

Western blotting also revealed that tagged TDG wildtype and TDG Δ cat can still be SUMOylated while TDG Δ S – as expected – is not; N-terminally tagged SUMOylated wildtype and catalytically dead TDG migrated at molecular weight of approximately 120 kDa in SDS-PAGE, the C-terminally tagged TDGs migrated at approximately 200 kDa, and both these bands were missing in the extracts containing

the SUMOylation-deficient TDG (Fig. 2B and C). This also confirms that the fluorophore-tags are not SUMOylated and fusion to TDG did not produce a new SUMO acceptor site. The intramolecular FRET fusion proteins were also still proficient for SUMOylation, evident from additional bands at 180 kDa for TDG wildtype and TDG Δ cat but not TDG Δ S (Fig. 2D).

The Cerulean tag on SUMO1 does not interfere with the SUMO-conjugation process as is evident from the appearance of an additional band running above 250 kDa (Fig. 2E). The fact that in the TDG-channel both bands, the one corresponding to endogenous SUMO attached to TDG-Citrine and the one representing Cerulean-SUMO1 linked to TDG-Citrine, show a similarly strong signal suggests that SUMOylation efficiency is not reduced by either of the tags (Suppl. Fig. 2).

We tested further whether a fluorophore tag on TDG has an effect on its glycosylase activity. To this end we utilized a standard base-release assay (Fig. 3A), testing the ability of differently tagged TDG to process a 60 bp substrate oligomer containing a single G•T mismatch. This oligomer was incubated with native protein extracts from *Tdg* knockout embryonic stem (ES) cells transiently expressing untagged, N-terminally, C-terminally or double-tagged TDG. *Tdg*^{-/-} ES cells were chosen as they tolerate overexpression of TDG (Y. Saito, personal communication) and to start with a clean TDG deficient background. Following incubation of the 60mer substrate with the protein extracts, quantification of glycosylase activity utilizes a Fluorescein-tag at the 5' end of the strand containing the mismatched T: heating under alkaline conditions cleaves the AP-site produced by TDG and generates a shorter 23-mer, which can be separated by denaturing PAGE from the uncleaved substrate (Fig. 3A and B).

We calculated the relative efficiency of base excision as the ratio between the percentage of processed substrate (Fig. 3B) and the relative TDG protein levels in the same extracts as determined by Western blotting (Fig. 3C), both normalized to the levels observed for untagged TDG. For testing the glycosylase activity of single-tagged TDG, EGFP-fusion constructs analogous to the Citrine-fusion ones were used. The N-terminal tag appears to increase the relative G•T processing efficiency of TDG approximately 3 fold (Fig. 3D), possibly by interfering with the non-specific DNA-binding capacity of

the N-terminal domain that keeps TDG firmly attached to the AP-site after base excision and, hence, prevents enzymatic turnover. This product-inhibition is further decreased by fusing fluorophores to both termini of TDG (Citrine-TDG-Cerulean), thus increasing the relative catalytic activity of TDG 5 times above the level of untagged TDG. It is likely that the double-tagged TDG cannot efficiently switch to or remain in the closed clamp-like conformation and so does not stay firmly attached to the AP-site.

From these results we conclude that the C-terminally tagged TDG FRET construct is similar to untagged TDG concerning expression and catalytic activity and would thus be the favorable construct for FRET experiments monitoring the SUMO-TDG interaction.

Visualization of the SUMO-interaction of TDG

To test the subcellular localization of the FRET constructs as well as visualizing the interaction of TDG with SUMO, we co-transfected cos7 cells with TDG-Citrine and Cerulean-SUMO1 or -SUMO3, as well as with controls: TDG-Citrine + Cerulean, Cerulean-SUMO1 or -SUMO3 + Citrine and TDG Δ S-Citrine + Cerulean-SUMO1 or -SUMO3. The nuclear localization of wildtype TDG and TDG Δ S was not perturbed by the fluorescence tag (Fig. 4, YFP-channel, line 3, 4, 6 and 7), while the free fluorophores were distributed equally between the nucleus and the cytoplasm (Fig. 4, CFP-channel, line 1, YFP-channel, line 2 and 5). The distribution of Cerulean-SUMO1 and 3, with a bias towards nuclear localization (Fig. 4, CFP-channel, line 2-7), is in agreement with previously reported immunofluorescence stainings of the endogenous SUMO proteins (Saitoh and Hinchev 2000; Evdokimov et al. 2008).

As similar levels of FRET donor and acceptor are essential for accurate FRET measurement and an excess of either one of them results in a false positive FRET signal after normalization, microscope settings were adjusted using cells expressing a direct fusion of Citrine and Cerulean, which also served as a positive control. Images of cells with similar donor and acceptor levels were acquired in 3 channels: the CFP (Cerulean), the YFP (Citrine) and the FRET channel. The images were processed

using the PixFRET ImageJ plugin (Feige et al. 2005) which calculates FRET values pixel per pixel, normalized for different intensity levels of donor and acceptor, and visualizes them in grey scale. As this formula is only accurate within certain limits of differential intensity levels, cells with a ratio of intensities above 5:1 were excluded. The false positive FRET signal deriving from different intensities beyond these limits is evident for instance as a cytoplasmic halo in cells expressing the nuclear TDG-Citrine and the ubiquitously distributed Cerulean (Fig. 4, NFRET, line 1).

The FRET signal within the nucleus of cells expressing TDG-Citrine and Cerulean-SUMO3 is increased 1.5 - 2 fold ($p < 0.0001$, Mann-Whitney test) compared to the highest “background” FRET signal observed in control cells, namely those expressing TDG-Citrine and free Cerulean (Fig. 4 and 5B). The same applies for the interaction between TDG and SUMO1 (Fig. 4 and 5A). The fact that cells expressing TDG ΔS instead of the wildtype produces a FRET signal significantly above the background (approximately 1.5 fold, $p < 0.0001$, Mann-Whitney test) and in the same range as wildtype TDG indicates that our FRET system does not solely measure the covalent but also the non-covalent SUMO-interaction of TDG. TDG ΔS therefore functions as a control to estimate which changes in the FRET signal derive from covalent and which derive from non-covalent SUMO-interactions.

Induction of DNA damage alters the TDG-SUMO interaction dynamics measured by FRET

It is noteworthy that a significant part of the cellular TDG pool appears to interact with SUMO1 and 3 without the presence of DNA damaging agents. To test the hypothesis that SUMOylation of TDG is involved in regulating processes during DNA repair, in particular the dissociation of TDG from the abasic site, we treated the cells with the base analog 5-FU which, incorporated into the DNA, represents a lesion addressed by TDG (Hardeland et al. 2003; Morgan et al. 2007). We have previously reported that TDG contributes to the cytotoxic effect of 5-FU by a delayed repair of the AP-site after excision of 5-FU from 5-FU•A mismatches, possibly through a saturation of the SUMO system (Kunz et al. 2009).

We treated cells with 30 μ M 5-FU and measured the SUMO1- and SUMO3-interactions of TDG. The FRET signal observed with wildtype TDG, reflecting the compound signal produced by covalent and non-covalent SUMO1-interactions, remained unchanged after treatment with 5-FU. Yet, the FRET signal of SUMO1 and TDG Δ S was significantly decreased in 5-FU-treated cells by approximately one third ($p < 0.0001$, Mann-Whitney test) (Fig. 5A), possibly reflecting a previously reported effect that DNA associated TDG fails to engage in non-covalent SUMO-interactions (Mohan et al. 2007; Smet-Nocca et al. 2011). As 5-FU treatment appears to reduce SUMO1-binding but the FRET signal derived from the sum of SUMO1-binding plus SUMO1-modification of wildtype TDG remains unchanged, we conclude that SUMO1-modification of TDG is stimulated by the treatment.

In contrast to SUMO1, SUMO3 shows a significantly higher FRET signal (by approximately 20%, $p < 0.001$, Mann-Whitney test) produced by SUMO3 with wildtype TDG compared to SUMO3 with TDG Δ S in untreated cells. This suggests that at least 1/5 of the signal observed with wildtype TDG represents SUMO3-modified protein. Thus it appears that TDG is preferentially modified by SUMO3 rather than by SUMO1 in unchallenged cells (Fig. 5B). Treatment with 5-FU, however, does not alter the FRET efficiency significantly although we observed certain small trends: the signal derived from wildtype TDG and SUMO3 seems to be slightly reduced whereas the one derived with TDG Δ S is mildly increased.

Taken together, it appears that the interaction of TDG with SUMO1 and SUMO3 occur in the context of different processes as SUMO1-binding is influenced by the induction of DNA damage while SUMO3-binding remains mostly unaffected at least under the conditions used in this study.

Part of the cellular TDG pool exists as a homodimer

We fused TDG N- and C-terminally to Citrine and Cerulean with the aim to monitor the conformational switch TDG undergoes upon binding to homoduplex DNA. Since there is biochemical evidence that suggests the formation of TDG homodimers (Roland Steinacher, personal communication), we included single-tagged TDG-donor- and -acceptor fusions as dimerization

controls to distinguish between FRET deriving from intramolecular or from intermolecular interactions. We found that TDG-Citrine and TDG-Cerulean generated a significant FRET signal, similar to but surpassing Citrine-TDG and TDG-Cerulean. Double-tagged TDG produced an even higher FRET signal, which was significantly increased in TDG ΔS (Fig. 6). Assuming that TDG ΔS remains longer in the closed conformation, this increase suggests that at least part of the FRET signal derives from the conformational switch bringing the donor and acceptor fluorophores closer together, although it remains unclear whether this involves one or two TDG proteins.

Still, the majority of the FRET signal appears to be generated through dimerization of TDG, evident from the FRET signal produced by two single-tagged TDG fusion proteins, which is enhanced approximately 2-fold by addition of a second fluorophore tag. Moreover, as the two tags appear to weaken non-specific DNA interactions of TDG (homoduplex DNA binding), accelerating enzymatic turnover of TDG (Fig. 3D), it is likely that the conformational switch is not induced efficiently upon DNA binding. We therefore conclude, that this system does not allow a precise quantitative monitoring of the conformational switch of TDG as the fluorophore tags may interfere with the efficient formation of a clamp-like closed conformation, but we provide the first *in vivo* evidence of TDG homodimerization. Whether this is mediated by DNA remains to be tested.

Discussion

TDG operates in the BER pathway which excises and replaces damaged bases that are continuously generated in the DNA. For BER to be beneficial for cells, it requires a tight regulation of protein complex assembly, of hand-over of potentially hazardous repair intermediates and of the dissociation of the factors involved. Post-translational modification with SUMO, in addition to non-covalent SUMO-binding of key proteins such as TDG, is likely to function as a platform for protein-protein

interactions and as a regulatory switch to control the interplay of proteins and DNA in these processes (Geiss-Friedlander and Melchior 2007).

We present a FRET system to monitor the SUMO-interaction dynamics of TDG in living cells, which produces a robust FRET signal in cos7 cells (Fig. 4). By including a SUMOylation-deficient mutant TDG, our system allows to distinguish between covalent and non-covalent SUMO-binding and, thus, to draw conclusions about the dynamics of SUMO modification and interaction of TDG.

Interestingly, a considerable proportion of the cellular TDG pool appears to interact with SUMO1 and SUMO3 non-covalently (TDG ΔS in Fig. 4 and 5). Little is known about the function of the non-covalent interaction between TDG and SUMO, although it has been found to be essential for its role as a co-activator of CBP/p300 and for its translocation to and/or from PML bodies (Takahashi et al. 2005; Mohan et al. 2007). Examples of other factors harboring an SBM suggest that TDG might interact with SUMOylated proteins through its SBM. The transcriptional repressor Daxx, for example, was shown to interact with SUMO covalently and non-covalently. The SBM of Daxx is essential for its repressor function as it mediates the interaction with other SUMOylated factors (Lin et al. 2004; Kuo et al. 2005; Lin et al. 2006). If a similar scenario applies for TDG, its interaction with SUMO-conjugated proteins might mediate the assembly of downstream acting BER factors, which is then followed by SUMOylation of TDG and its release from the AP-site. This hypothesis is supported by proteomics approaches that revealed several BER factors to be target of SUMO modification, e.g. XRCC1, DNA Ligase III and PARP1 (Gocke et al. 2005; Bruderer et al. 2011). Moreover, non-denaturing pull-down of protein complexes with SUMOylated subunits also revealed Pol β , which fills the nucleotide gap in the final steps of BER, to be associated with SUMOylated proteins without being a target of modification itself (Bruderer et al. 2011).

Although we can measure solely non-covalent SUMO-binding with our FRET system by using TDG ΔS , it is unclear whether the interaction occurs with free or with conjugated SUMO. To provide a means to distinguish these two possibilities in the future, we generated a Citrine-tagged mutant SUMO1, in which we replaced the C-terminal Gly-Gly motif with Ala-Ala, thus abolishing conjugation capacity.

By inducing DNA damage, we were able to alter the interaction dynamics between TDG and SUMO1 but not with SUMO3. Interestingly, SUMO1 was reported to exist almost exclusively in the conjugated form, whereas SUMO2/3 conjugation appears to occur mainly in response to stress (Geiss-Friedlander and Melchior 2007). We found 5-FU to boost SUMO1-conjugation but decreases non-covalent SUMO1-binding (Fig. 5a). This shift towards SUMOylation can be explained by the highly efficient processing of 5-FU by TDG. TDG would repeatedly bind a substrate 5-FU•G or 5-FU•A, immediately excise 5-FU and get SUMOylated to dissociate from the abasic site (Hardeland et al. 2003). The decrease in non-covalent SUMO-binding hints at the disassembly of complexes (involving either free or conjugated SUMO) and/or an increased engagement of TDG in DNA repair which shifts the balance of covalent and non-covalent interaction towards the covalent modification. Furthermore, dissolving any non-covalent SUMO-interactions is necessary to allow DNA-binding, as these two were shown to be mutually exclusive (Smet-Nocca et al. 2011).

The fact that the SUMO3-interaction of TDG remains mostly unaffected by the incorporation of 5-FU into the DNA (Fig. 5b) suggests that the interaction with this SUMO variant may be involved in another pathway. Given the recent findings that TDG may be involved in processes other than classical DNA repair, e.g. the maintenance of epigenetic stability of CpG islands (reviewed in (Jacobs and Schar 2012)), it is likely that SUMO in general and the SUMOylation of TDG in particular play an essential role in regulating the interactions and conformational changes needed in these epigenetic pathways. SUMO1 and SUMO3 may play distinct roles in assembling different TDG-associated complexes, and our FRET system provides a tool to elucidate their functions as it allows monitoring these interactions under various conditions in live cells.

We also designed the TDG-FRET system to monitor the predicted conformational switch of TDG from an open to a closed form upon DNA interaction. These attempts, however, were complicated by the fact that the N- and C-terminal fluorophore tags needed for FRET appeared to interfere with the formation of the predicted clamp-like configuration, as inferred from an increased relative efficiency of the double-tagged TDG in G•T mismatch processing (Fig. 3d).

However, data we obtained from pilot FRET analyses with these double-tagged fusion proteins provide evidence that TDG may form homodimeric complexes in living cells (Fig. 6): The dimerization controls included to distinguish between intra- and intermolecular FRET showed clearly that TDG interacts with itself *in vivo* (Fig. 6). The biological function of this interaction may be further stabilization of the fragile AP-site and its direct vicinity (discussed below).

Whether this interaction really involves only two TDG proteins or more and if the orientation of TDG within these dimers is parallel or anti-parallel is not clear. If DNA-binding mediates this interaction, it is possible that the termini of two TDG molecules in a closed conformation bring the donor and acceptor fluorophores in a suitable proximity and orientation for FRET. A published crystal structure of DNA-bound TDG indeed suggests that TDG homodimers might assemble on DNA (Maiti et al. 2008; Morgan et al. 2011).

Further studies are necessary to test whether DNA- and SUMO-binding are involved in TDG dimerization and what the biological function of this interaction is. Assuming for example the binding of a second TDG opposite an AP-site-bound TDG, such an interaction might support stabilization of this fragile site, possibly by interfering with repair processes addressing damage in the opposing strand. Such a function of TDG dimerization appears plausible in the light of recent findings implicating TDG in active DNA demethylation (He et al. 2011; Maiti and Drohat 2011). Methylated cytosine 5-mC occurs (in most cases) symmetrically in palindromic CpG dinucleotides. During BER-coupled DNA demethylation, mechanisms must be in place to inhibit simultaneous generation of AP-sites as demethylation intermediates on both strands. The generation of a second AP-site in the opposing strand would inevitably result in the formation of a double-strand break which would have to be avoided at all costs. Other repair pathways, i.e. long-patch BER, nucleotide excision and mismatch repair, involve the degradation of a whole stretch of single-strand containing a mismatch or lesion, which would also have to be avoided vis-à-vis an ongoing BER process. It is thus likely that mechanisms evolved to avoid the clash of two repair events on opposing strands, and these might include the SUMO coordinated binding and release of repair proteins to and from the DNA.

In conclusion, our FRET system provides a powerful tool to investigate the dynamic SUMO- and auto-interactions of TDG in DNA repair and other processes, i.e. DNA demethylation and maintenance of epigenetic stability.

References

- Bruderer, R., et al. (2011). "Purification and identification of endogenous polySUMO conjugates." EMBO Rep 12(2): 142-148.
- Cortazar, D., et al. (2007). "The enigmatic thymine DNA glycosylase." DNA Repair (Amst) 6(4): 489-504.
- Evdokimov, E., et al. (2008). "Loss of SUMO1 in mice affects RanGAP1 localization and formation of PML nuclear bodies, but is not lethal as it can be compensated by SUMO2 or SUMO3." J Cell Sci 121(Pt 24): 4106-4113.
- Feige, J. N., et al. (2005). "PixFRET, an ImageJ plug-in for FRET calculation that can accommodate variations in spectral bleed-throughs." Microsc Res Tech 68(1): 51-58.
- Fortini, P. and E. Dogliotti (2010). "Mechanisms of dealing with DNA damage in terminally differentiated cells." Mutat Res 685(1-2): 38-44.
- Geiss-Friedlander, R. and F. Melchior (2007). "Concepts in sumoylation: a decade on." Nat Rev Mol Cell Biol 8(12): 947-956.
- Gill, G. (2004). "SUMO and ubiquitin in the nucleus: different functions, similar mechanisms?" Genes Dev 18(17): 2046-2059.
- Gocke, C. B., et al. (2005). "Systematic identification and analysis of mammalian small ubiquitin-like modifier substrates." J Biol Chem 280(6): 5004-5012.
- Hardeland, U., et al. (2000). "Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis." J Biol Chem 275(43): 33449-33456.
- Hardeland, U., et al. (2003). "The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs." Nucleic Acids Res 31(9): 2261-2271.
- 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.
- He, Y. F., et al. (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." Science 333(6047): 1303-1307.
- Jacobs, A. L. and P. Schar (2012). "DNA glycosylases: in DNA repair and beyond." Chromosoma 121(1): 1-20.
- Kunz, C., et al. (2009). "Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil." PLoS Biol 7(4): e91.
- Kuo, H. Y., et al. (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(47): 16973-16978.
- Lin, D. Y., et al. (2004). "Negative modulation of androgen receptor transcriptional activity by Daxx." Mol Cell Biol 24(24): 10529-10541.
- Lin, D. Y., et al. (2006). "Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors." Mol Cell 24(3): 341-354.
- Lindahl, T. and R. D. Wood (1999). "Quality control by DNA repair." Science 286(5446): 1897-1905.
- Lomeli, H. and M. Vazquez (2011). "Emerging roles of the SUMO pathway in development." Cell Mol Life Sci 68(24): 4045-4064.

- Maiti, A. and A. C. Drohat (2011). "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites." J Biol Chem 286(41): 35334-35338.
- Maiti, A., et al. (2008). "Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition." Proc Natl Acad Sci U S A 105(26): 8890-8895.
- Minty, A., et al. (2000). "Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif." J Biol Chem 275(46): 36316-36323.
- Mohan, R. D., et al. (2007). "SUMO-1-dependent allosteric regulation of thymine DNA glycosylase alters subnuclear localization and CBP/p300 recruitment." Mol Cell Biol 27(1): 229-243.
- Morgan, M. T., et al. (2007). "Excision of 5-halogenated uracils by human thymine DNA glycosylase. Robust activity for DNA contexts other than CpG." J Biol Chem 282(38): 27578-27586.
- Morgan, M. T., et al. (2011). "Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA." Nucleic Acids Res 39(6): 2319-2329.
- Saitoh, H. and J. Hinchey (2000). "Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3." J Biol Chem 275(9): 6252-6258.
- Smet-NoCCA, C., et al. (2011). "SUMO-1 regulates the conformational dynamics of thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity." BMC Biochem 12: 4.
- Song, J., et al. (2004). "Identification of a SUMO-binding motif that recognizes SUMO-modified proteins." Proc Natl Acad Sci U S A 101(40): 14373-14378.
- Steinacher, R. and P. Schar (2005). "Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation." Curr Biol 15(7): 616-623.
- Takahashi, H., et al. (2005). "Noncovalent SUMO-1 binding activity of thymine DNA glycosylase (TDG) is required for its SUMO-1 modification and colocalization with the promyelocytic leukemia protein." J Biol Chem 280(7): 5611-5621.
- Waters, T. R., et al. (1999). "Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1." J Biol Chem 274(1): 67-74.
- Xia, Z. and Y. Liu (2001). "Reliable and global measurement of fluorescence resonance energy transfer using fluorescence microscopes." Biophys J 81(4): 2395-2402.

Figure Legends

Fig. 1: Measuring SUMO interactions and conformational changes of TDG in the process of base excision by FRET. A: Structure-function model of TDG action. In the open and unmodified form, TDG has a high affinity for DNA, facilitating the search for and binding of a substrate base. Upon binding to double-stranded DNA, TDG undergoes a conformational change involving the N-terminal domain, switching to a closed form. Upon excision of the substrate base, involving the catalytically essential N141 (N), TDG stays firmly bound to the abasic site. SUMOylation of K341 (K) facilitates dissociation of TDG by switching it back to an open form, thus reducing its affinity to DNA. Fusion of TDG with Citrine and SUMO with Cerulean allows for measuring their interaction following base-release, since SUMOylation of TDG brings the two fluorophores in the required proximity for FRET. B: Rationale of the intramolecular FRET approach; fusing TDG N- and C-terminally to Citrine and Cerulean respectively will allow measuring intramolecular FRET if the N- and the C-terminus of the closed form bring the donor and acceptor fluorophore in a suitable proximity for the energy transfer.

Fig. 2: Analysis of expression of the FRET constructs and SUMOylation efficiency. SDS-extracts of *cos7* cells transiently expressing the indicated constructs were separated by SDS-PAGE, blotted and probed with anti-GFP (red channel) and anti-TDG (green channel) antibodies. Shown are overlays of both channels. A: Schematic overview of the constructs. B and C: Test for full-length expression and SUMOylation of Citrine- or Cerulean-tagged TDG, comparison of N- and C-terminally tagged TDG, C: test for full-length expression of double-tagged TDG (Citrine-TDG-Cerulean), D: test for full-length expression and ability of HA- and Cerulean-tagged SUMO1 to modify Citrine-tagged TDG; numbers represent molecular weight in kDa; the asterisk indicates a truncated peptide recognized by the anti-GFP antibody.

Fig. 3: Glycosylase activity of the wildtype TDG-fluorophore fusion proteins. Shown are the results of a base-release assay (Hardeland *et al.* 2000) with native whole cell extracts of murine ES cells transiently expressing the fusion constructs. A: Schematic overview of the base-release assay. A 60-mer G•T-heteroduplex DNA substrate (green star, Fluorescein-label) was incubated with cell extract. Functional TDG excises the T from the G•T mismatch. The resulting AP-site was cleaved by heating under alkaline conditions, inducing β - δ -elimination. The nicked substrate DNA strand was separated from the unprocessed oligomers by denaturing PAGE. The Fluorescein-labelled DNA was detected by fluorescence scanning. B: Results of the base-release assay; TDG activity is indicated by a shortened DNA substrate strand (23 nt). Depending on the 3' moiety (OH or P), the shortened substrate DNA strands migrate slightly differently, giving rise to a double band. Overexpressed TDGB was used as positive control; negative control, no extract; Cit, Citrine; Ceru, Cerulean; mock transfected, without DNA; C: Western blot of native ES cell extracts to estimate relative expression level; green channel anti-TDG (upper), red channel anti- β -actin (lower). The asterisk indicates SUMOylated TDG. D: quantification of the relative catalytic activity of the wildtype TDG-fluorophore fusion proteins on a G•T-substrate, calculated as the ratio between relative amount of processed substrate and relative expression level, both normalized to the values of untagged TDG.

Fig. 4: Normalized FRET images of cos7 cells transiently expressing the indicated FRET constructs. Images of cells with similar expression levels were analysed with the PixFRET plugin for ImageJ, calculating normalized FRET values pixel by pixel, visualizing them in grey scale. An imbalance of donor and acceptor expression causes a false positive FRET signal; the halo in the cell expressing TDG-Citrine and Cerulean (upper row) is caused by TDG being restricted to the nucleus while Cerulean is ubiquitously distributed.

Fig. 5: FRET analysis of the SUMO1 and 3-interactions of TDG upon induction of DNA damage. Cos7 cells were transiently transfected with the indicated constructs. A: Cells were incubated with 30 μ M 5-FU for 24 h prior to FRET analysis, n=26-37. B: as A, n=33-99. Relative NFRET, percent of the average of the positive control, each dot representing one cell. **, p < 0.005, ***, p < 0.0001, Mann-Whitney test.

Fig. 6: The majority of the FRET signal of double-tagged TDG derives from homodimer formation. cos7 cells were transiently transfected with the indicated constructs. n=34-39. Relative NFRET, percent of the average of the positive control, each dot representing one cell. ***, p < 0.0001, Mann-Whitney test.

FIGURES

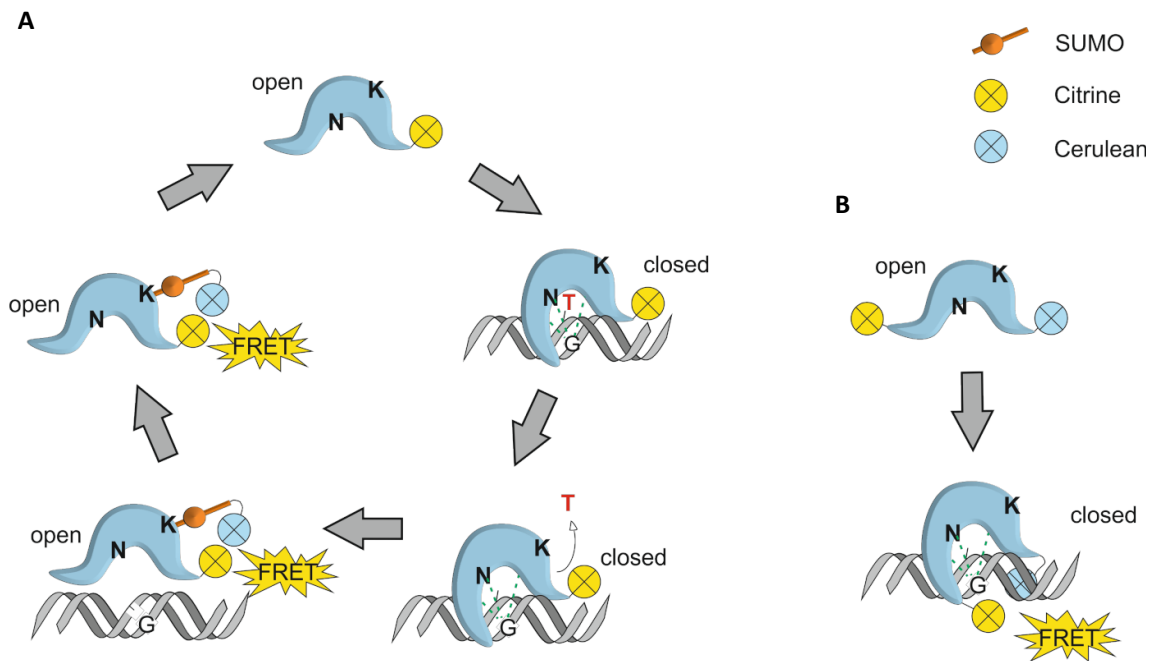


Figure 1: Measuring SUMO interactions and conformational changes of TDG in the process of base excision by FRET.

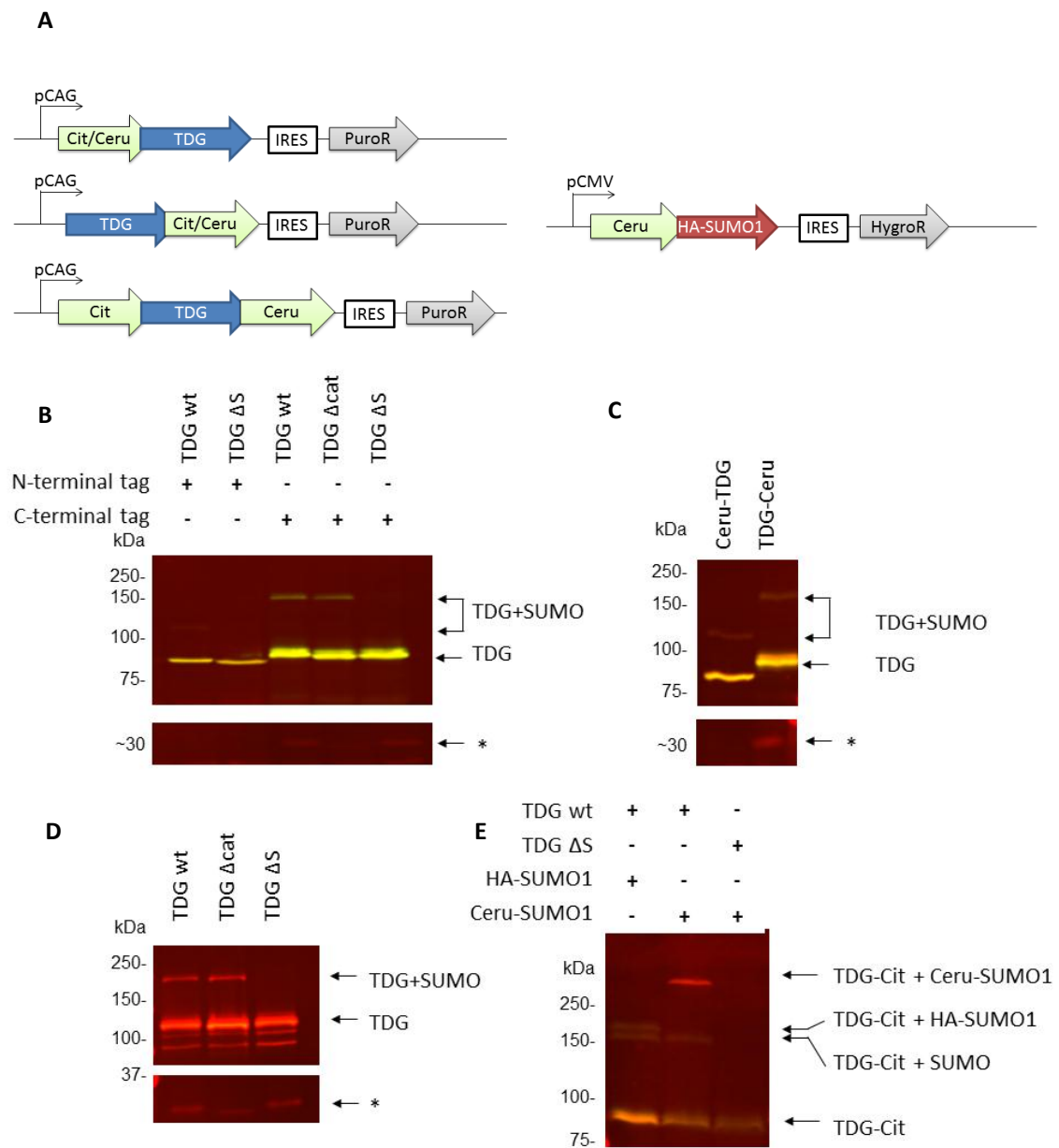


Fig. 2: Analysis of expression of the FRET constructs and SUMOylation efficiency.

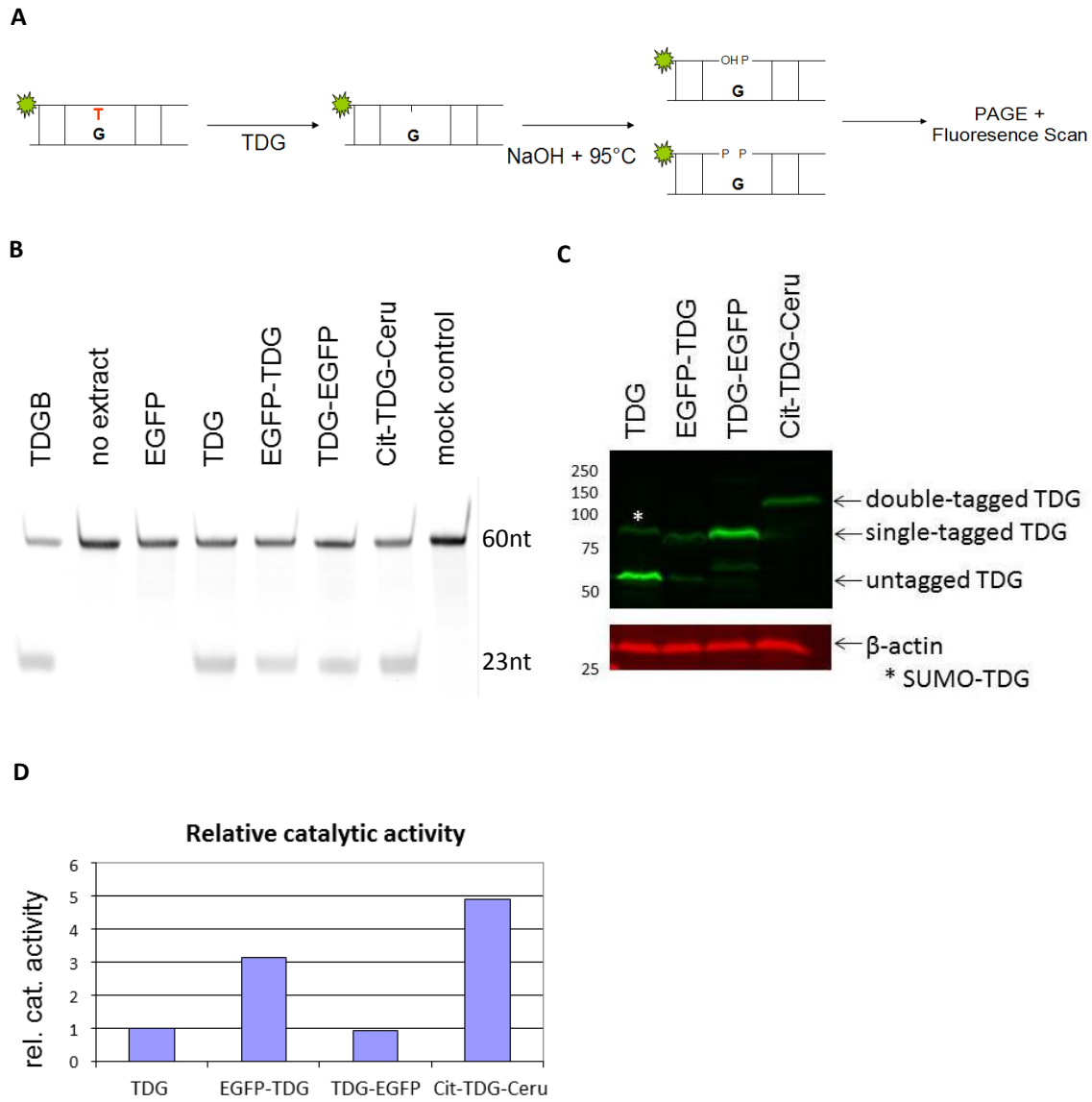


Fig. 3: Glycosylase activity of the wildtype TDG-fluorophore fusion proteins.

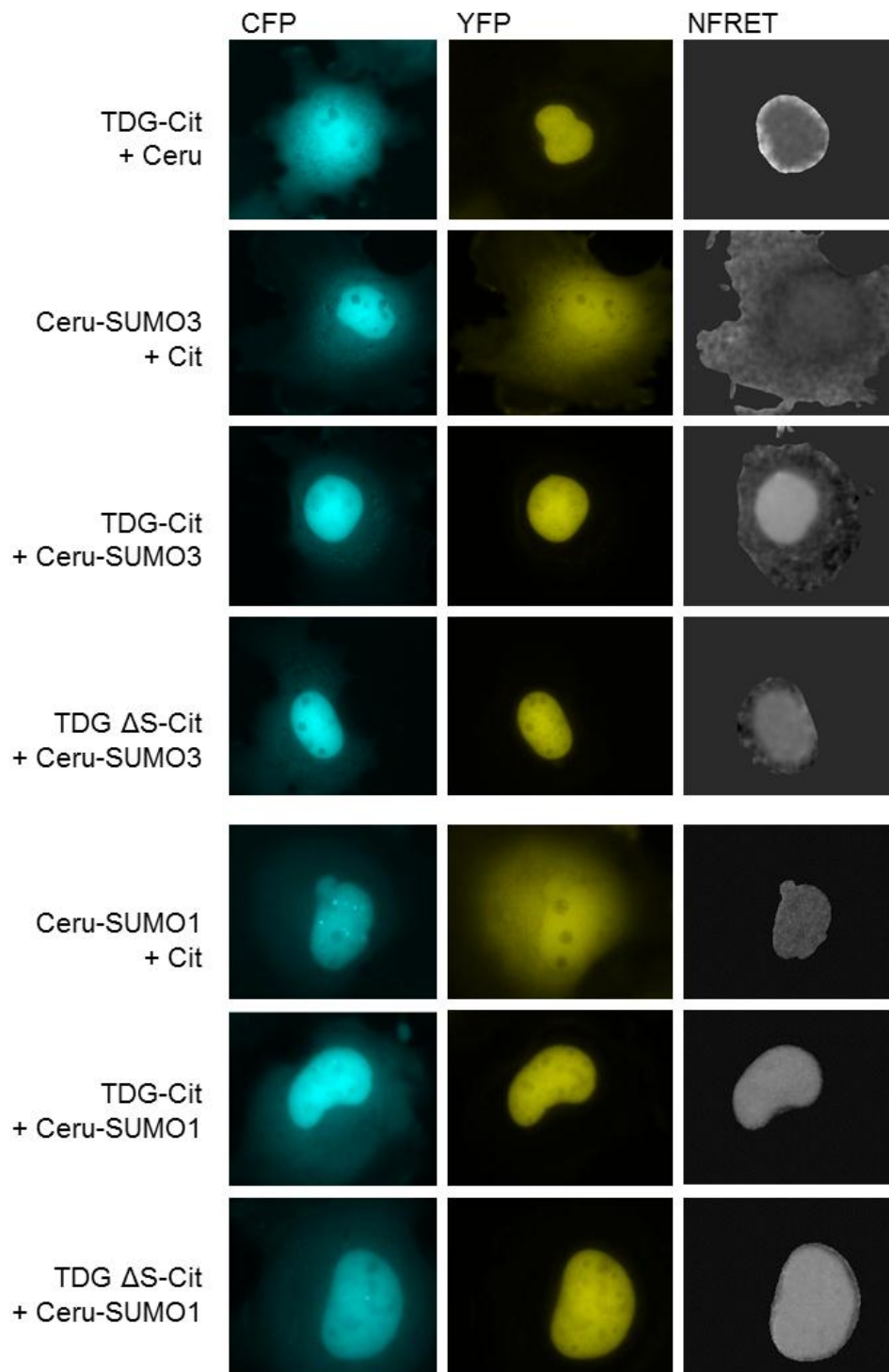


Fig. 4: Normalized FRET images of cos7 cells transiently expressing the indicated FRET constructs.

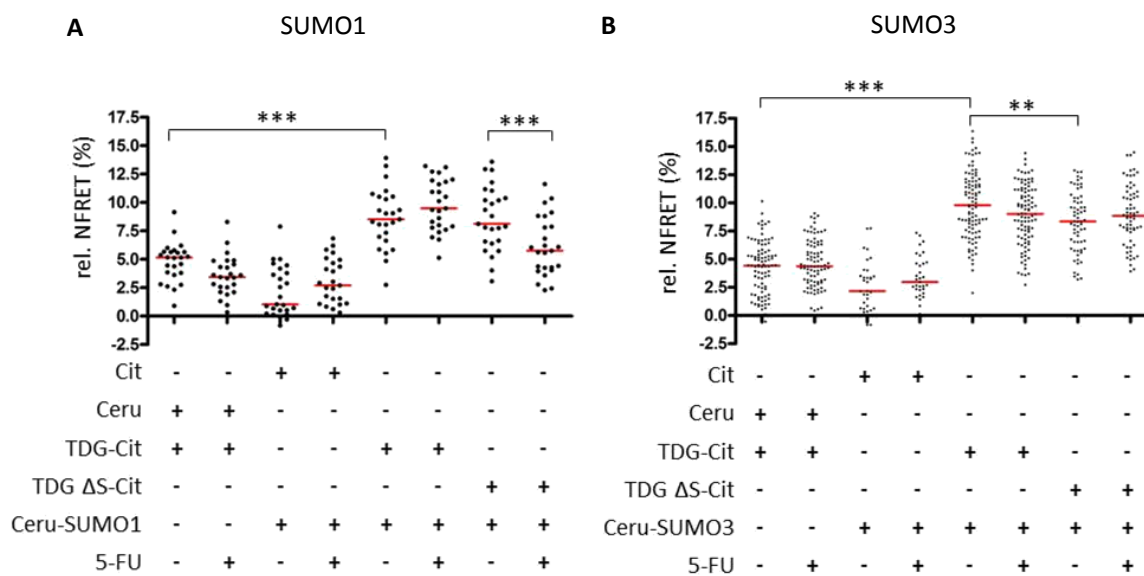


Fig. 5: FRET analysis of the SUMO1 and 3-interactions of TDG upon induction of DNA damage.

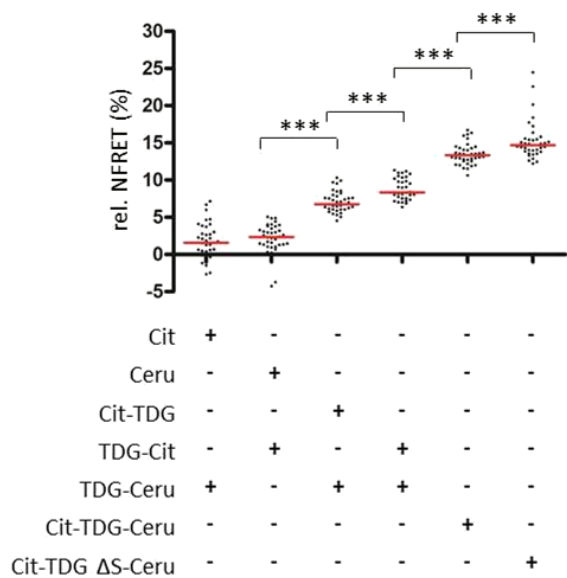


Fig. 6: The majority of the FRET signal of double-tagged TDG derives from homodimer formation.

Supplementary Material

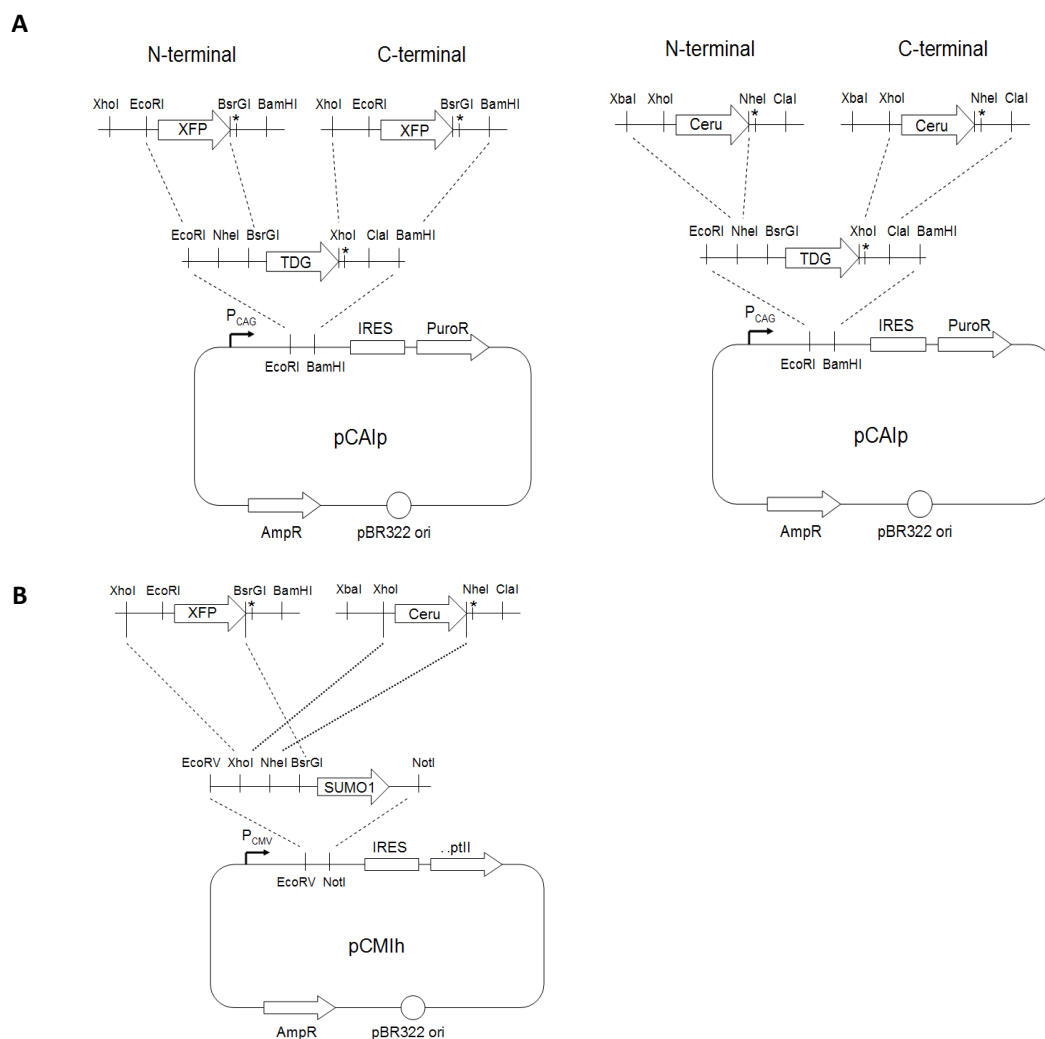
Measuring the SUMO-Interaction Dynamics of the Thymine DNA Glycosylase by Fluorescence Resonance Energy Transfer

Angelika L. Jacobs, David Schürmann, Yusuke Saito, Primo Schär*

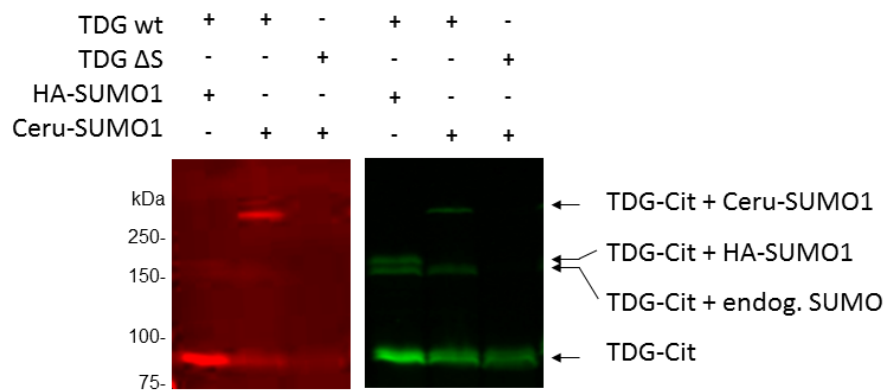
Department of Biomedicine, University of Basel, Mattenstrasse 28, CH-4058 Basel, Switzerland

*To whom correspondence should be addressed: E-mail primo.schaer@unibas.ch; Tel. +41 61 267
3561; Fax: +41 61 267 3566

Supplementary Figures



Suppl. Figure 1: Cloning strategy for the TDG- and SUMO-fluorophore fusion constructs. A: *TdgA* (wildtype, Δcat , ΔS) was digested with the indicated restriction enzymes and inserted into the mammalian expression vector pCAIp, containing an internal ribosome entry site (IRES), a eukaryotic Puromycin and a prokaryotic Ampicillin resistance cassette (PuroR and AmpR), as well as the chicken actin promoter (P_{CAG}). In a second step, the DNA fragments coding for the fluorophores (left panel: XFP = EGFP and Citrine; right panel: Ceru = Cerulean) were digested and inserted into the vector via the indicated restriction sites. The linker between TDG and the N-terminal tag codes for the amino acids NVQST for XFP, for ASSVQST for Ceru. The linker between TDG and the C-terminal tag codes for SSSRLEFA for XFP, for SSSST for Cerulean. The original amino acid sequence of TDG was extended by four serine residues, the first three of which are part of the spacer between TDG and the C-terminal tag. **B:** The PCR-amplified SUMO1 and SUMO3 genes were inserted into the expression vector pCMIh, containing an internal ribosome entry site (IRES), a eukaryotic Hygromycin and a prokaryotic Ampicillin resistance cassette (HptII and AmpR), as well as the cytomegalovirus promoter (P_{CMV}). In a second step, the DNA fragments coding for the fluorophores (XFP = EGFP and Citrine, Ceru = Cerulean) were inserted into the vector via the indicated restriction sites. The linker between XFP and SUMO codes for the amino acids NVQT, between Cerulean and SUMO1 for ASSVQT. Asterisks indicate the position of the stop codons.



Suppl. Figure 2: Modification of Citrine-tagged TDG with Cerulean-SUMO1 is not hindered by either of the fluorophore tags. SDS-extracts of cos7 cells transiently expressing the indicated constructs were separated by SDS-PAGE, blotted and probed with anti-GFP (red channel) and anti-TDG (green channel) antibodies. HA- and Cerulean-tagged SUMO1 can modify Citrine-tagged TDG with an efficiency equal to that of endogenous SUMO; numbers represent molecular weight in kDa.

Supplementary Tables

Supplementary Table 1: Constructs generated in this study

		N-terminal	C-terminal	Promoter	Selection	Name
controls	Citrine-Cerulean			pCAG	Puromycin	pCAIp Cit-Ceru
	Citrine			pCAG	Puromycin	pCAIp Cit
	Cerulean			CMV	Hygromycin	pCMIh Ceru
TDG	TDG wt	Citrine	-	pCAG	Puromycin	pCAIp Cit-TDG wt
	TDG wt	-	Citrine	pCAG	Puromycin	pCAIp TDG wt-Cit
	TDG wt	Citrine	Cerulean	pCAG	Puromycin	pCAIp Cit-TDG wt-Ceru
	TDG wt	Cerulean	-	pCAG	Puromycin	pCAIp Ceru-TDG wt
	TDG wt	-	Cerulean	pCAG	Puromycin	pCAIp TDG wt-Ceru
	TDG wt	EGFP	-	pCAG	Puromycin	pCAIp GFP-TDG wt
	TDG wt	-	EGFP	pCAG	Puromycin	pCAIp TDG wt-GFP
	TDG K341R	Citrine	-	pCAG	Puromycin	pCAIp Cit-TDG ΔS
	TDG K341R	-	Citrine	pCAG	Puromycin	pCAIp TDG ΔS-Cit
	TDG K341R	Citrine	Cerulean	pCAG	Puromycin	pCAIp Cit-TDG ΔS-Ceru
	TDG K341R	Cerulean	-	pCAG	Puromycin	pCAIp Ceru-TDG ΔS
	TDG K341R	-	Cerulean	pCAG	Puromycin	pCAIp TDG ΔS-Ceru
	TDG K341R	EGFP	-	pCAG	Puromycin	pCAIp GFP-TDG ΔS
	TDG K341R	-	EGFP	pCAG	Puromycin	pCAIp TDG ΔS-GFP
	TDG N141A	Citrine	-	pCAG	Puromycin	pCAIp Cit-TDG Δcat
	TDG N141A	-	Citrine	pCAG	Puromycin	pCAIp TDG Δcat-Cit
	TDG N141A	Citrine	Cerulean	pCAG	Puromycin	pCAIp Cit-TDG Δcat-Ceru
	TDG N141A	Cerulean	-	pCAG	Puromycin	pCAIp Ceru-TDG Δcat
	TDG N141A	-	Cerulean	pCAG	Puromycin	pCAIp TDG Δcat-Ceru
	TDG N141A	EGFP	-	pCAG	Puromycin	pCAIp GFP-TDG Δcat
TDG N141A	-	EGFP	pCAG	Puromycin	pCAIp TDG Δcat-GFP	
SUMO	SUMO1	Cerulean-HA	-	pCMV	Hygromycin	pCMIh Ceru-SUMO1-GG
	SUMO1	Citrine-HA	-	pCMV	Hygromycin	pCMIh Cit-SUMO1-GG
	SUMO1	EGFP-HA	-	pCMV	Hygromycin	pCMIh EGFP-SUMO1-GG
	SUMO1-AA	Citrine-HA	-	pCMV	Hygromycin	pCMIh Cit-SUMO1-AA
	SUMO3	Cerulean-HA	-	pCMV	Hygromycin	pCMIh Ceru-SUMO3-GG
SUMO3	Cit-HA	-	pCMV	Hygromycin	pCMIh Cit-SUMO3-GG	

Supplementary Table 2: Oligonucleotides*

Name	Use	Sequence 5'--> 3'
Ceru fw X X +K	PCR of Cerulean	GCGCTCTAGACTCGAGCACCATGGTGAGCAAGGGCGA GGAGC
Ceru rev N*C	PCR of Cerulean	CGCATCGATGTTAGCTAGCCTTATACAGCTCGTCCATGC CGAGAGTGA
EGFP fw X E +K	PCR of EGFP and Citrine	GCGCGCTCGAGGCTCGAATTCGCCATGGTGAGCAAGGG CGAGGAGC
EGFP rev B*B	PCR of EGFP and Citrine	GCGCGGATCCATGCATCTATTGTACATTGTATAGCTCGT CCATGCCGAGAGTGATC
HA fw XhoNheBsrGI	PCR of HA-SUMO1 and 3	ATCCTCGAGTAGCTAGCAGTGTACAGACCATGGCTTCAT ATCCTTACG
Linker B N X fw	Spacer for Cit-Ceru	GTACAGAGCGGTGGCAATGCATCAGGAGGTAGC
Linker B N X rev	Spacer for Cit-Ceru	TCGAGCTACCTCCTGATGCATTGCCACCGCTCT
mTDGa fw E N B +K	PCR of mTdgA	CGCGAATTCTAGCTAGCAGTGTACAGAGCACCATGGAC GCAGAGGCCGC
mTDGa rev X*C B	PCR of mTdgA	CGCGGATCCATCGATTAGCTCGAGCTAGAAGCGTGGCT CTCTTCTCCTG
SP6	PCR of HA-SUMO1 and 3	TGAATTTAGGTGACACTATAG
SUMO1AA rev*N	PCR of HA-SUMO1-AA	GCGTATCAGCGCCGCTAAGCCGCCGTTTGTTCCTGATAAAC
Subs60IT-F	Base release assay	CGGAATTCGTCTAGGTTTGAGGTTGACATCGGATCCATG GTACCTCGAGGGCAATGTCTA
Subs60uG	Base release assay	TAGACATTGCCCTCGAGGTACCATGGATCCGATGTCGAC CTCAAACCTAGACGAATTCCG

*all oligonucleotides were provided by Microsynth AG, Switzerland

Supplementary Table 3: Filter cube* specifications for fluorescent microscopy

Name	Excitation band	Excitation filter	Dichroic mirror	Emission band	Emission filter
Dapi	325-375 nm	350/50	400	435-485 nm	460/50
GFP	450-490 nm	470/40	495	500-550 nm	525/50
CFP	418-442 nm	430/24	455	458-482 nm	470/24
YFP	510-530 nm	500/20	515	520-550 nm	535/30
FRET	418-442 nm	430/24	455	520-550 nm	535/30

*source: Chroma Technology Corp®

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.

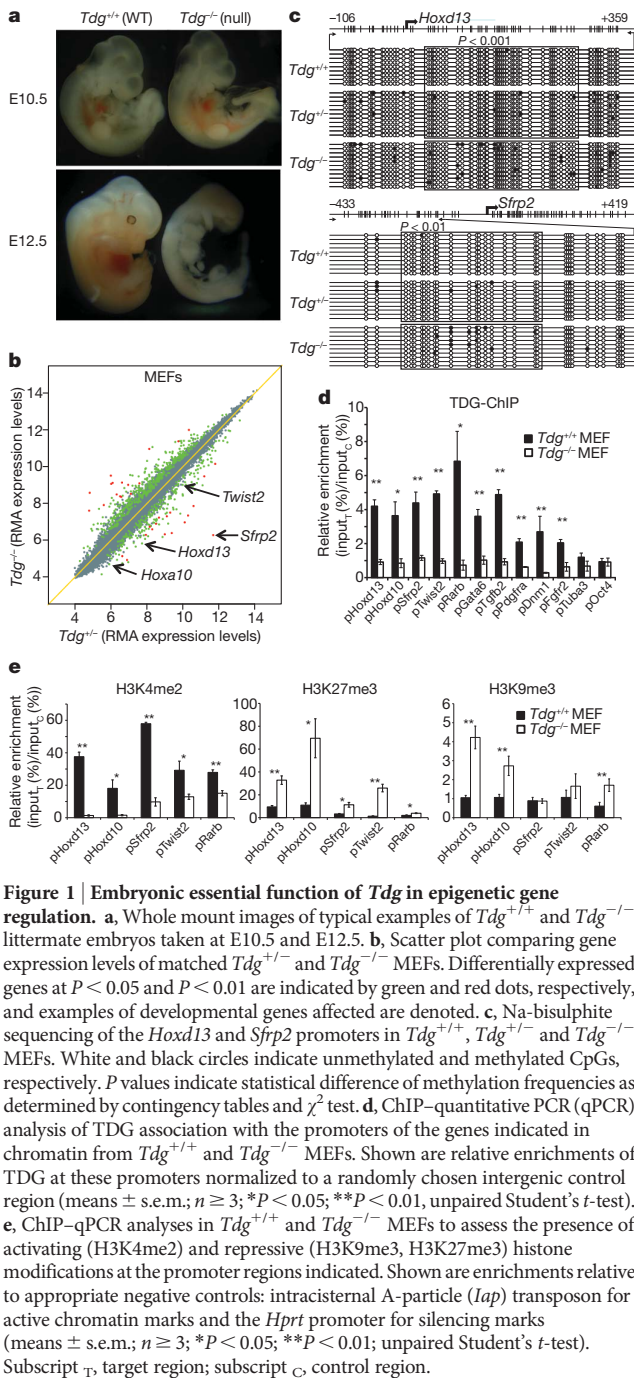
We next investigated a possible involvement of TDG in gene regulation by expression profiling of TDG-proficient and -deficient MEFs. To limit potential clonal biases, we compared the transcriptomes of early passages of litter-matched populations of SV40 immortalized MEFs. This identified 461 differentially 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*, *Sfp2*, *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.

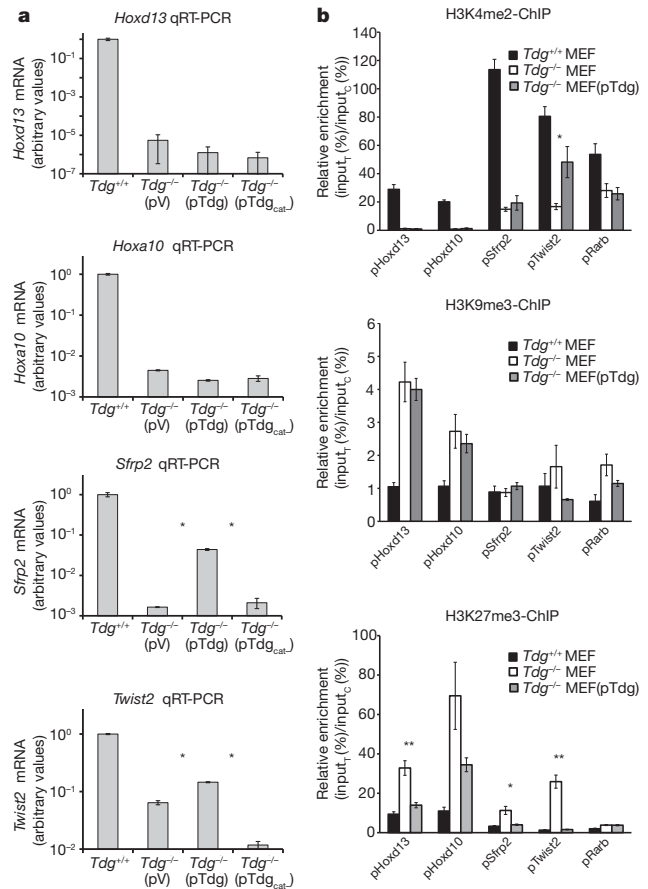
†Present addresses: European Commission, Joint Research Centre, Institute for Environment and Sustainability, 21027 Ispra, Italy (T.L.); Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK (R.S.).

*These authors contributed equally to this work.



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



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 ≥ 1.5) but increased significantly upon differentiation to neuronal progenitor cells (297 genes, $P \leq 0.05$, FC ≥ 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

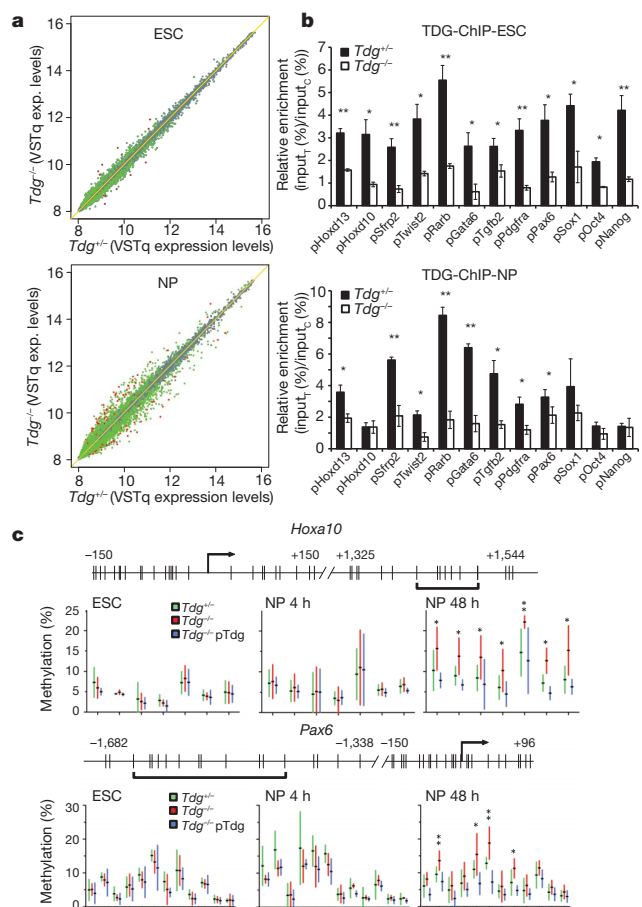


Figure 3 | TDG-dependent differences in gene expression and chromatin status arise during cell differentiation. **a**, Scatter plots comparing gene expression profiles of *Tdg*^{+/-} and *Tdg*^{-/-} ESCs or *in vitro* differentiated neuronal progenitors. Green and red dots indicate differentially expressed 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

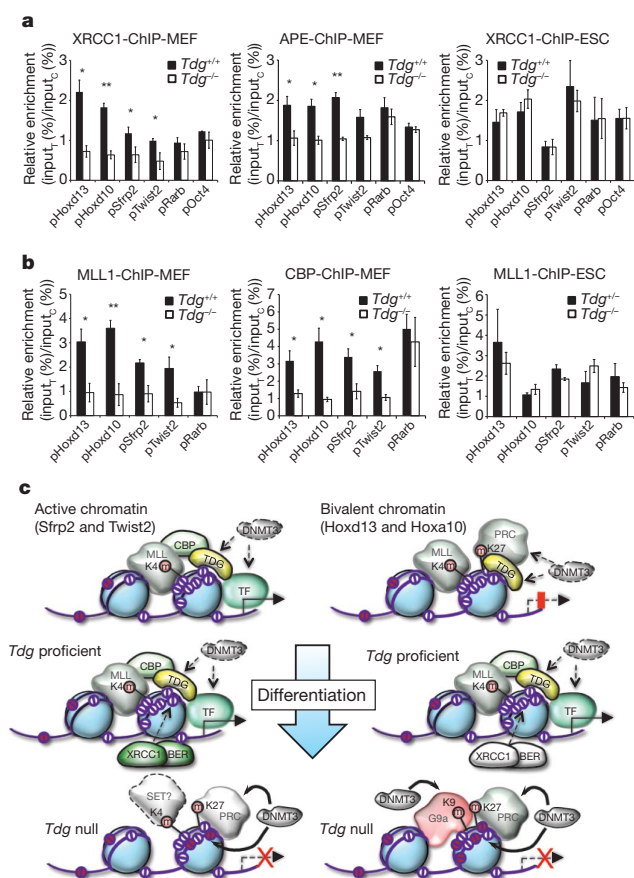


Figure 4 | Structural and catalytic functions of TDG in epigenetic maintenance. **a**, ChIP-qPCR analysis of XRCC1 and APE1 association with the gene promoters indicated in chromatin of TDG-proficient and -deficient MEFs and ESCs. Shown are relative enrichments of XRCC1 and APE1 at these promoters normalized 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).
- Kangaspekka, 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 Peroxiredoxin 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 (3100A0-108436; 3100A0-122574/) 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 serum, 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) pre-blocked 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.

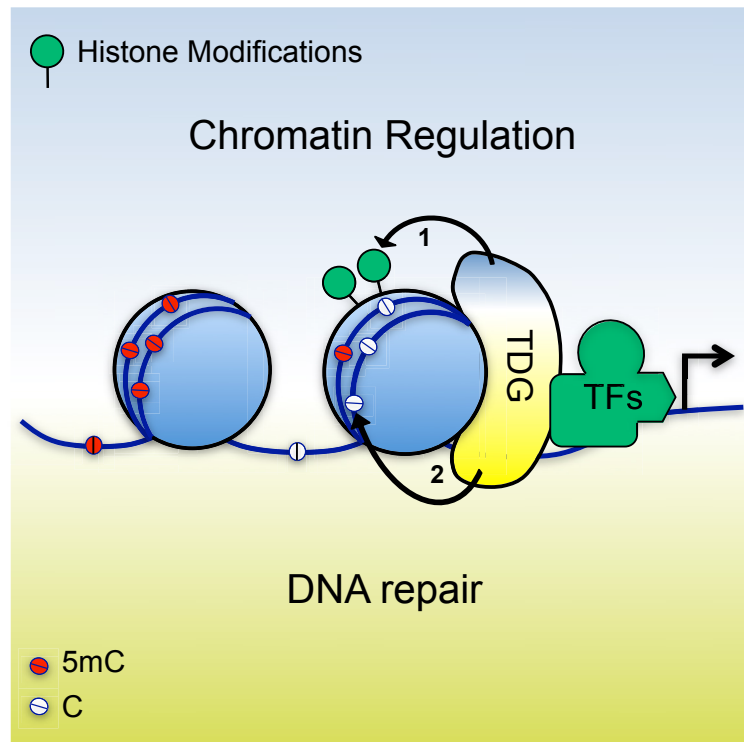
Base release assay. For base release assays, 25–50 µg of ESC whole-cell extracts were incubated with 0.5 pmol of a fluorescein-labelled GC/TG, GCm/CG or GCm/mCG DNA substrate in reaction buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT, 1 mg ml⁻¹ BSA) at 37 °C for 1 h (GC/TG) or overnight (methylated 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).

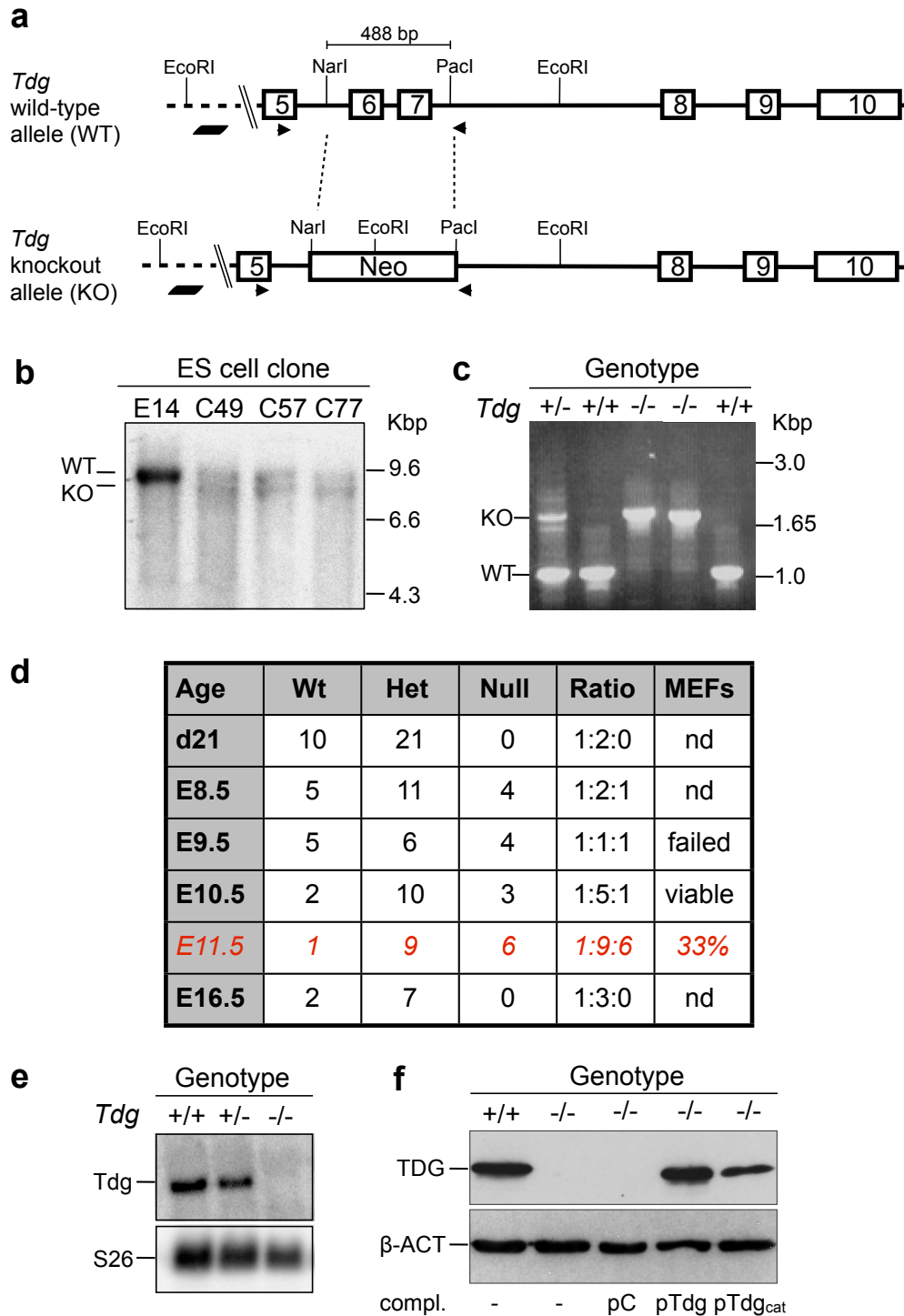
30. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37**, 853–862 (2005).

Summarizing Figure



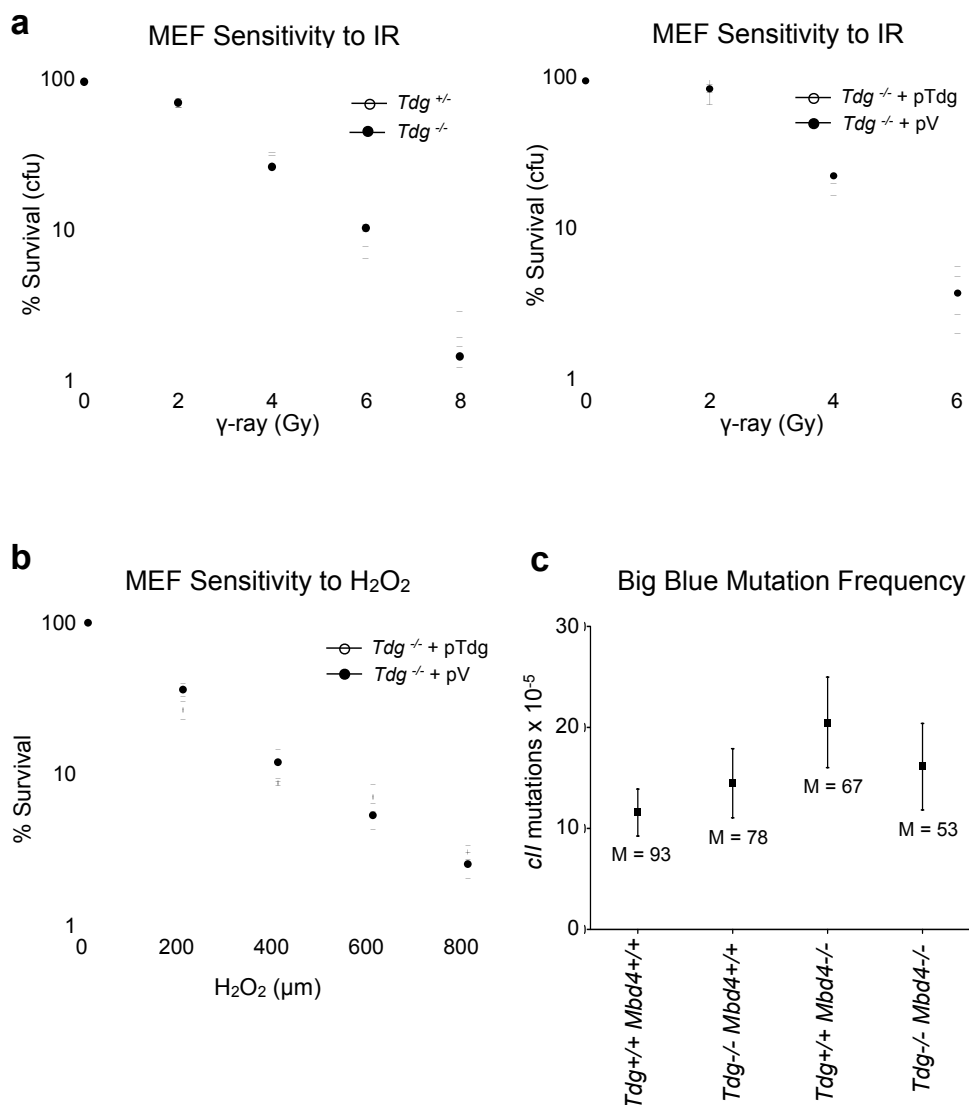
Summarizing Figure. The role of TDG in epigenetic control. TDG sustains proper (permissive) epigenetic states at gene promoters. As a structural component of transcription regulatory complexes, it contributes to the establishment and/or maintenance of accurate histone modification patterns (1), as a 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.

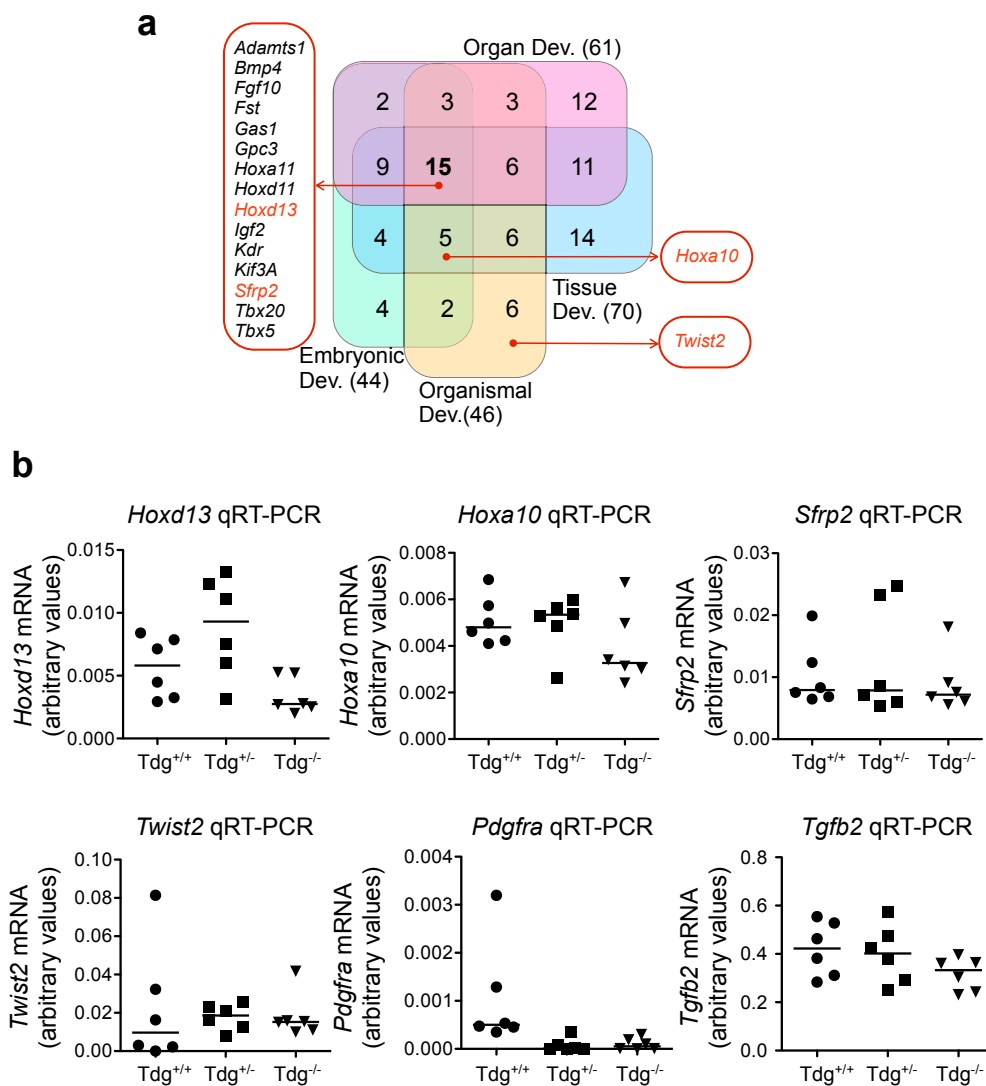
Supplementary Fig. 2



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

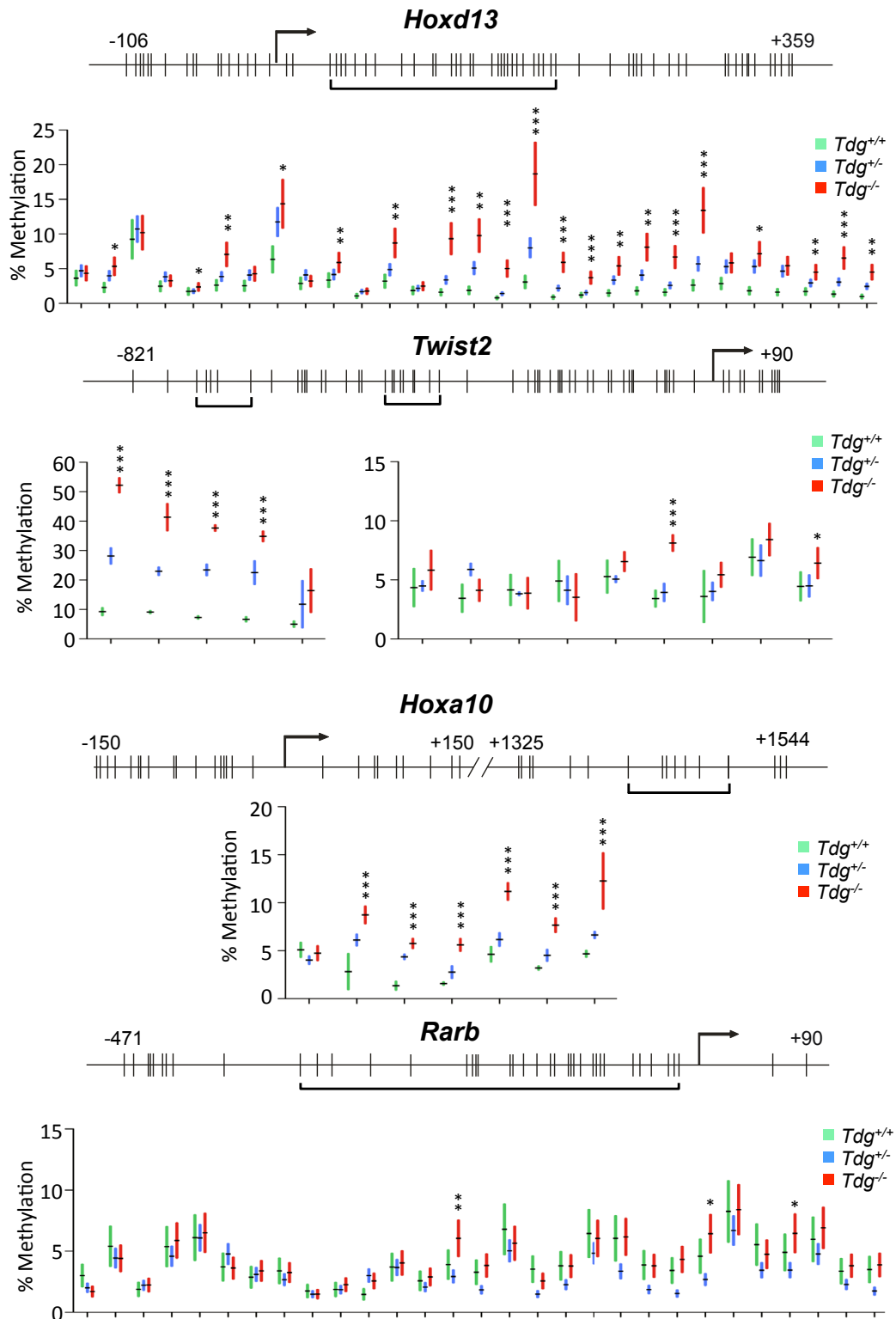
phage screened. Shown are mutation frequencies with 95% confidence intervals as calculated from the binominal proportions, with M indicating the actual number of mutant plaques scored for each genotype.

Supplementary Fig. 3



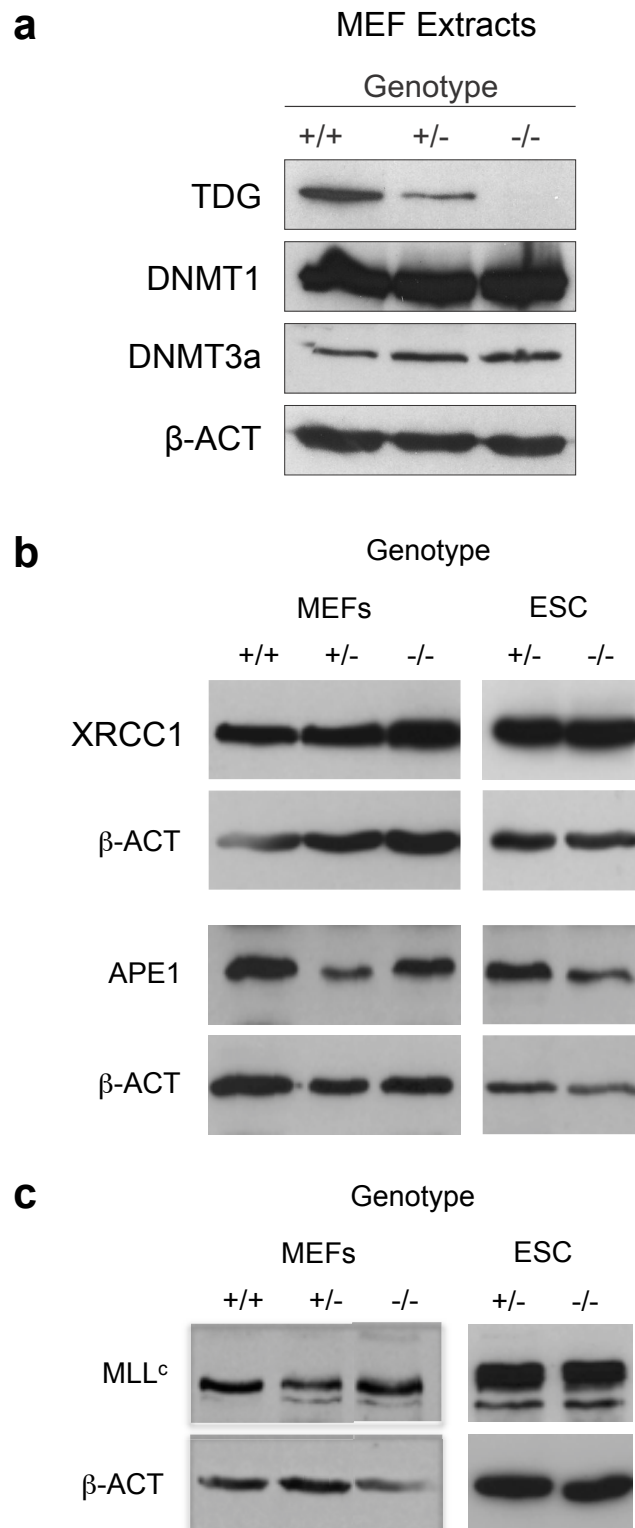
Supplementary Figure 3 | Gene ontology analysis and expression of selected targets in primary MEF isolates. **a**, Gene ontology (GO) annotations of the 200 most differentially regulated genes ($p < 0.05$) reveal a significant enrichment of developmental pathways (Ingenuity Pathway Analysis). **b**, 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

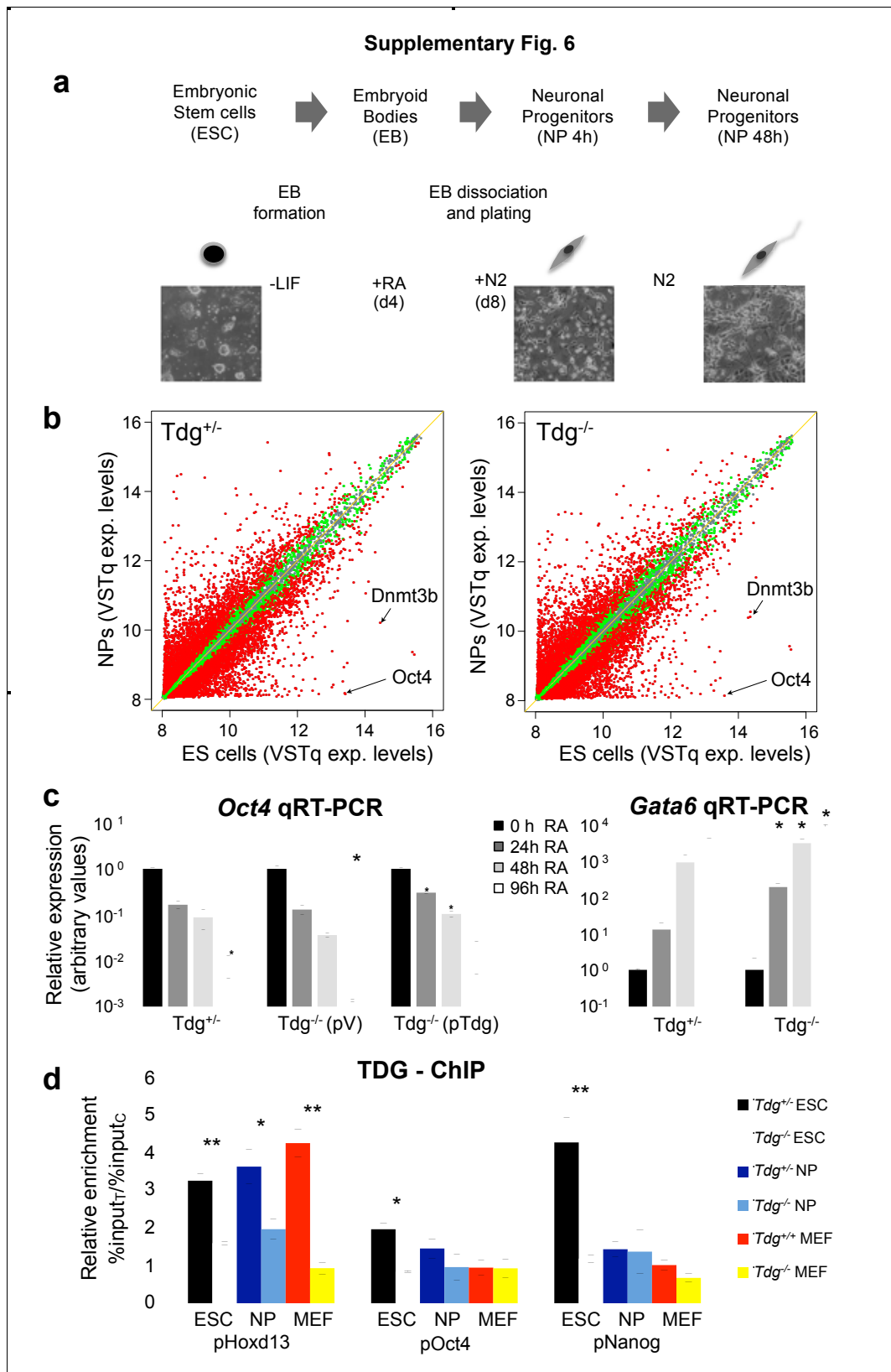


Supplementary Figure 4 | CpG methylation states of selected target promoters in MEFs. DNA methylation analysis by bisulfite pyrosequencing of *Hoxd13*, *Twist2*, *Hoxa10* and *Rarb* promoter regions in *Tdg^{+/+}*, *Tdg^{+/-}* and *Tdg^{-/-}* MEFs. Promoter regions are depicted schematically with vertical 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



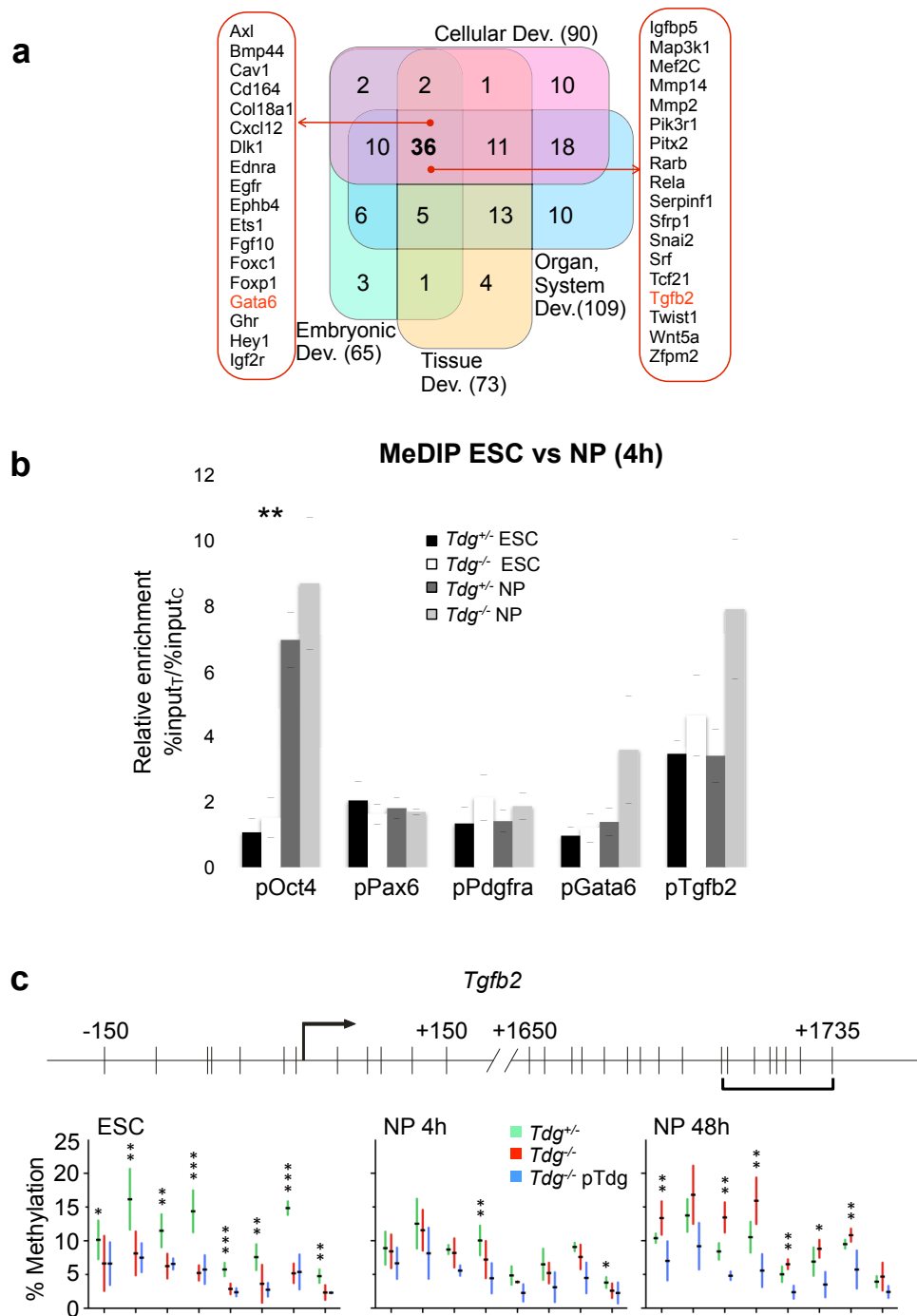
Supplementary Figure 5 | Validation of proteins levels and ChIP analysis of DNMT3a in TDG proficient and deficient MEFs. **a**, Western blots showing protein levels of TDG, DNMT1 and DNMT3a in whole cell extracts (WCE) of *Tdg*^{+/+}, *Tdg*^{+/-} and *Tdg*^{-/-} MEFs with β -ACT as loading control. 50 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).

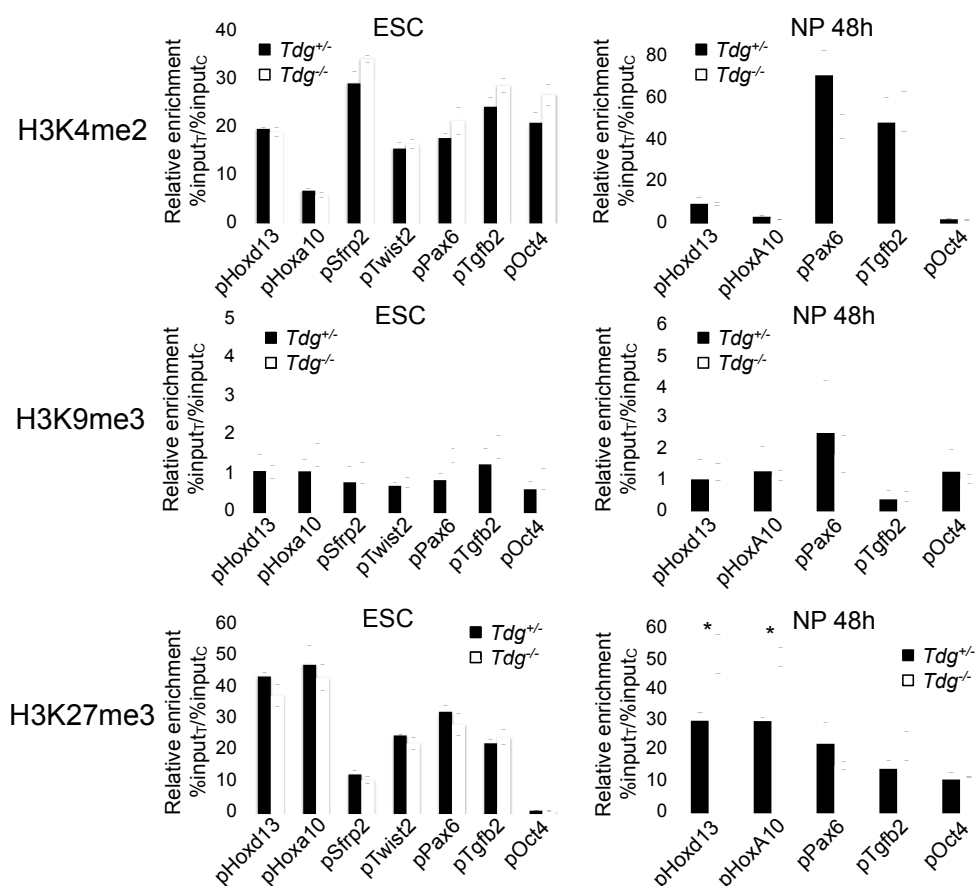
Supplementary Fig. 7



Supplementary Figure 7 | Gene ontology and DNA methylation analyses of TDG controlled genes during ESC - NP differentiation. **a**, Gene ontology (GO)

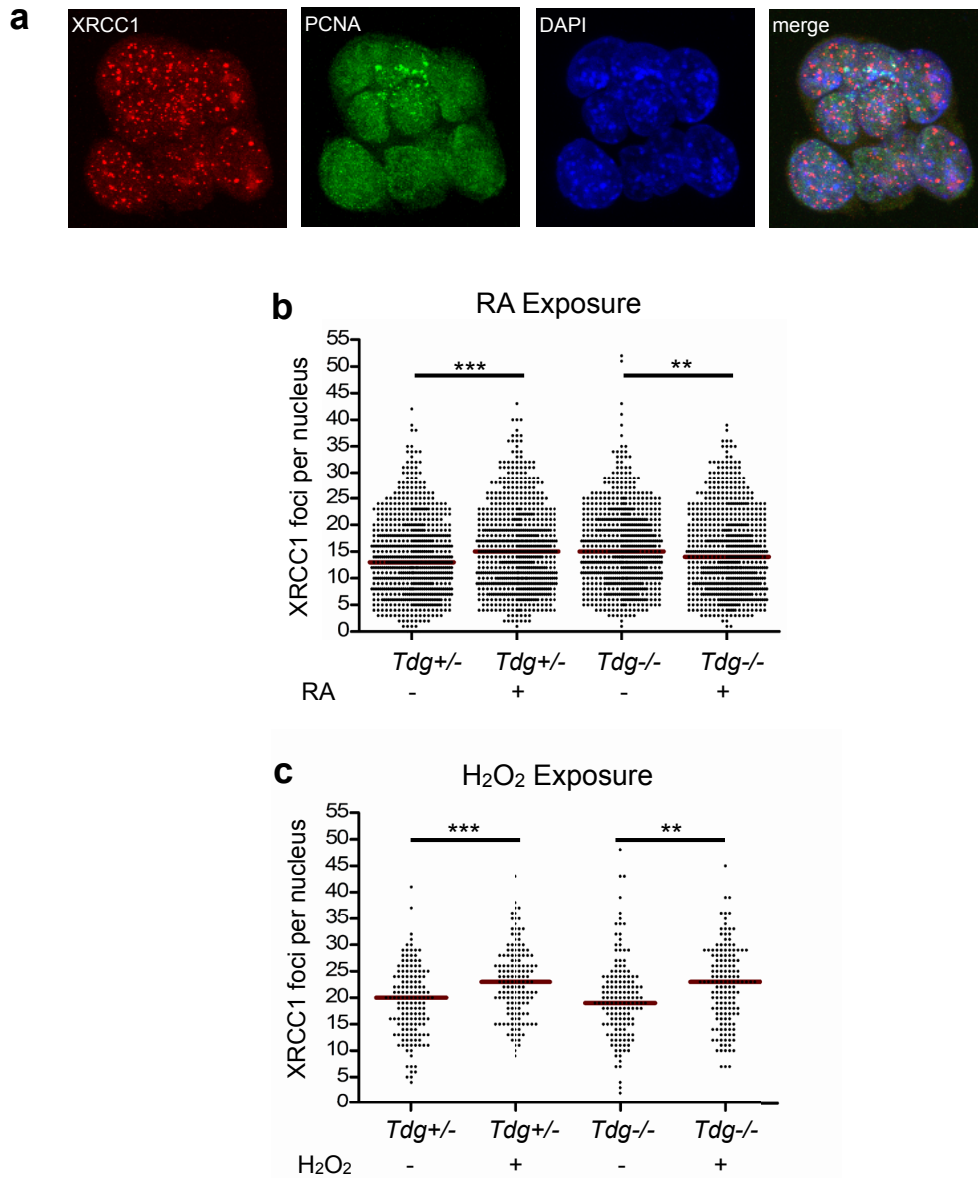
annotations of the 200 most differentially regulated genes (all $p < 0.05$) reveal a significant enrichment of developmental pathways (Ingenuity Pathway Analyses). **b**, The DNA methylation status at the *Oct4*, *Pax6*, *Pdgfra*, *Gata6* and *Tgfb2* promoters was analysed by MeDIP-qPCR in *Tdg*^{+/+} 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.

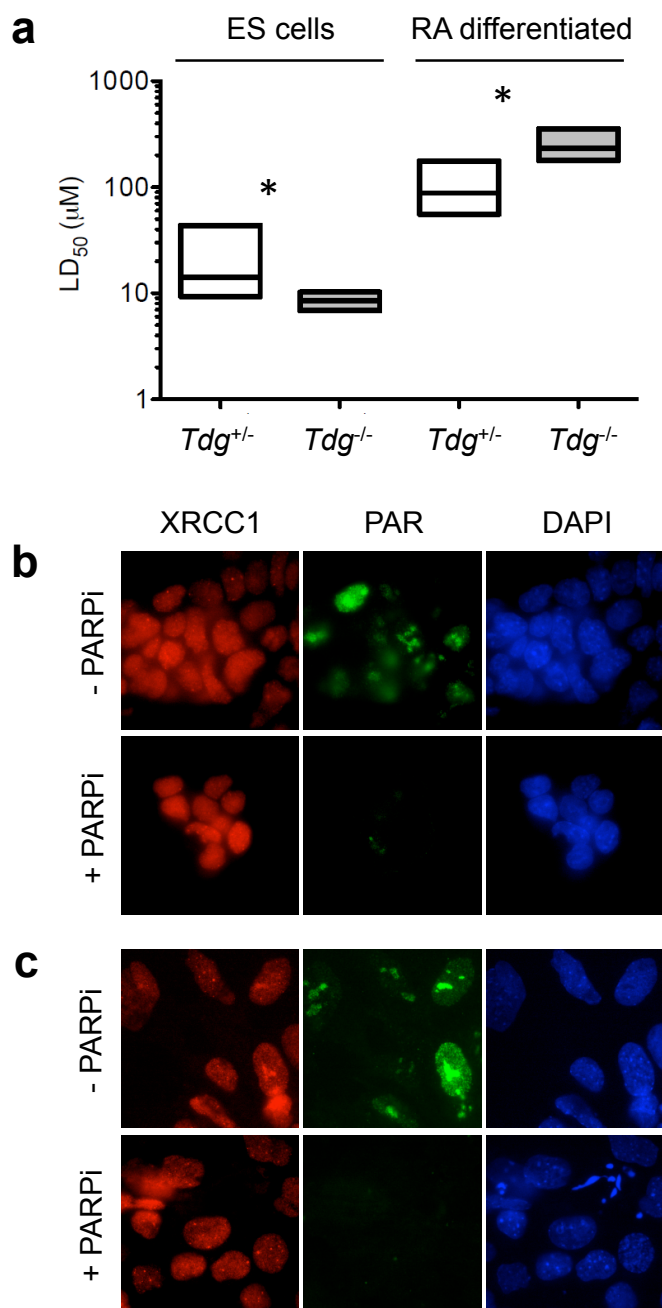
Supplementary Fig. 9



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 μM H_2O_2 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 H_2O_2 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 (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

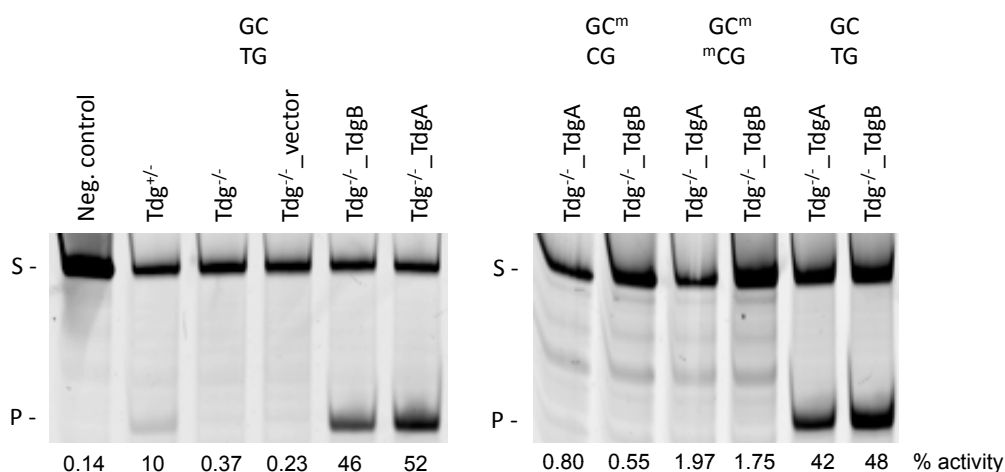
Supplementary Fig. 10



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	TTGTAAGGTATTTAAATAAGTAG
HoxD13 F	GGGTTATGAGTAGTTAGGGGATTTGGGATATGGATGG
HoxD13 R	<u>GTCAGTCCAGTCCAGGTCAGGGTGAAGTATAGTATAGAGGTTGAG</u> GTTGAATTTTAAAT
HoxD13 S1	GGGGATTTGGGATATG
HoxD13 S2	GTAGTAGAGTTTGGTTAG
Pax6 F	GAGTGGGGTGGGGGGAAAAT
Pax6 R	b-TTCACCCTAACTCCCACCCTTATCC
Pax6 S1	GGGAAAATGGGTAGG
Pax6 S2	GGTTTAGGTATAGTTGTGTTA
Rarβ F	GTTAGATTGGTTGGGTTATTTGAAGGTTAG
Rarβ R	<u>GTCAGTCCAGTCCAGGTCAGGATCTTTTTCCCAACCCCAATCATA</u> AATTATAACAA
Rarβ S1	GGGTTATTTGAAGGTTAGTA
Rarβ S2	GTTTGAAGGGAGAAT
Rarβ S3	GATTGGGATGTAGAGG
Rarβ S4	GGGGGATTAGAGTTT
Tgfβ2 F	TAATAGTATTAGGGATTTATTGTAGGAGAAGGTAAG
Tgfβ2 R	b-AATTTACAAACCTATAAATCCCTCTCCATC
Tgfβ2 S	GGGATTTATTGTAGGAGAAG
Twist2 F	<u>GTCAGTCCAGTCCAGGTCAGGGTTGTGATGTTTAAGTTATAAAGTAT</u> TTAGGGGGTAG
Twist2 R	TCTCCTAAAACAAATTTAACCCCTACCAAATTC
Twist2 S1	TTTCTAACTACTTCAACCTA
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
pRarβF	GGGAGTTTTTAAGCGCTGTG
pRarβR	CGGAGCAGCTCACTTCCTAC
pTgfβ2F	AAGGGACGAGACGAGAAGGT
pTgfβ2R	ACATCCACACGCACACTCAT
pPax6F	CGGTGAAAGAAGCCACTAGG
pPax6R	TAGGGCGTTTGTTCCTAAAT
pOct4F	GTGAGGTGTCGGTGACCCAAGGCAG
pOct4R	GGCGAGCGCTATCTGCCTGTGTC
pGata6F	AGTTTTCCGGCAGAGCAGTA
pGata6R	AGGAGGAAACAACCGAACCT
pDnm1F	ATTCGCGGACTGGTCACTAT
pDnm1R	TTAGCACCCCTAGCCATCAC
pPdgfraF	GGACGAGCGATCTGGAATAA
pPdgfraR	CCGTGCAGAAAAGACTCCAC
pFgfr2F	CTCCAGAATCCAAGGACCA
pFgfr2R	CATCCCAATGCTGACATCTG
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)	GGATGCAGGGATGTTC
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

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

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

²Department 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

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 implicated in active DNA demethylation propagated by the TET proteins as it processes 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), the final products of TET mediated 5-mC oxidation. 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 state of high epigenetic plasticity mediated by a 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. As these modifications encode heritable information about cell identity that is not laid down in the sequence of the DNA bases, they are 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 in being placed, removed and replaced by a cohort of histone modifying enzymes (Bannister and Kouzarides 2011), cytosine methylation has long been regarded as a stable epigenetic mark that, once established by the *de novo* DNA methyltransferases DNMT3a and DNMT3b, is preserved through cell proliferation by the maintenance DNA methyltransferase DNMT1 (Goll and Bestor 2005). However, two global DNA demethylation events have been described to occur in the mammalian life cycle, one in the paternal pronucleus in the zygote and one in primordial germ cells (PGCs). Both events have recently been shown to be mediated by 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 in the case of global DNA demethylation appears to be subsequently diluted by DNA replication as 5-hmC is not maintained by DNMT1 (Valinluck and Sowers 2007; Iqbal et al. 2011; Wossidlo et al. 2011; Hashimoto et al. 2012; Hackett et al. 2013). Whereas the conversion of 5-mC to 5-hmC followed by passive removal presents a plausible and safe pathway for global demethylation, several lines of evidence suggest that at specific loci, DNA demethylation can occur by an active process not requiring DNA replication (Bruniquel and Schwartz 2003; Kangaspeska et al. 2008; Metivier et al. 2008). Both, passive and active mechanisms are likely involved in global and targeted reprogramming during development.

Interestingly, plants have been shown to utilize DNA glycosylases to excise 5-mC, followed by restoration of an unmethylated C by the base excision repair (BER) pathway (Zhu 2009). Efforts to identify analogous enzymes in mammals have implicated the thymine DNA glycosylase (TDG) as a prime candidate for a 5-mC demethylase. Apart from its ability to excise the deamination product of 5-mC, thymine, from G•T mismatches, interactions with DNMT3a and DNMT3b (Li et al. 2007; Boland and Christman 2008) and various transcription factors 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, which suggests a non-redundant role of TDG in development. Furthermore, TDG-deficient cells accumulate epigenetic aberrations with differentiation (Cortazar et al. 2011; Cortellino et al. 2011). While a direct 5-mC glycosylase activity of TDG has been proposed (Zhu et al. 2000), these findings still pend corroboration and it appears that mammalian cells apply a more complex mechanism than plants in erasing 5-mC.

A plausible alternative to direct excision of 5-mC by a DNA glycosylase is the conversion of 5-mC to a more favorable substrate for DNA glycosylases. Three major pathways have been proposed, the most straightforward suggesting the deamination of 5-mC by a cytosine deaminase, e.g. the activation induced deaminase (AID) or the apolipoprotein B RNA-editing catalytic component (APOBEC) enzymes, resulting in a G•T mismatch that could be processed by TDG but also the Methyl-CpG Binding Domain protein 4 (MBD4). Support for such a pathway came with the finding that AID contributes to demethylation of the *Oct4* and *Nanog* gene promoters during somatic cell reprogramming and to global DNA demethylation in primordial germ cells (PGCs) (Bhutani et al. 2010; Popp et al. 2010). Furthermore, a deamination-coupled DNA demethylation pathway involving the coupled action of AID and MBD4 was described in zebrafish embryos (Rai et al. 2008). A second putative DNA demethylation pathway entails the coupled action of the TET proteins, converting 5-mC to 5-hmC, and subsequent deamination of 5-hmC to 5-hydroxymethyluracil (5-hmU), which can be excised by either TDG or the single-strand specific monofunctional glycosylase 1 (SMUG1) (Hardeland et al. 2003; Cortellino et al. 2011). However, a recent study has cast doubt on such a pathway since AID and the APOBEC family of deaminases appear to be mostly inactive on 5-hmC (Nabel et al. 2012).

As neither MBD4 nor SMUG1 are essential for embryonic development (Wong et al. 2002; Kemmerich et al. 2012) and neither can compensate for the loss of TDG, it appears that TDG acts in a distinct pathway essential for embryo development that is not coupled to deamination. Such a pathway has taken shape with the finding that the TET proteins catalyze not only the generation of 5-hmC but also its further oxidation to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), both of

which are excellent substrates for TDG-mediated BER and appear not to be processed by any other DNA glycosylase (He et al. 2011; Ito et al. 2011; Maiti and Drohat 2011).

We have previously proposed a dual function of TDG in maintaining a permissive chromatin state at CpG island promoters during cell differentiation, first by structurally supporting the maintenance of active histone marks and second by counteracting errors of the DNA methylation machinery (Cortazar et al. 2011). However, how exactly TDG supports epigenetic stability through processes of cell fate determination has remained elusive. 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

Neuronal differentiation in *Tdg*^{-/-} cells is accompanied by an increasing disturbance of DNA methylation patterns

We reported previously that TDG is essential for embryonic development and that TDG deficient cells accumulate aberrant DNA methylation at CpG island (CGI) promoters, accompanied by a loss of active and gain of repressive histone marks in differentiated cells (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 and deficient embryonic stem cells (ESCs), early (4h) neuronal progenitors (NPs) derived by *in vitro* differentiation (Fig.1a) and MEFs isolated from *Tdg*^{+/+} and *Tdg*^{-/-} embryos (Wilson 2012). Whereas *Tdg*^{-/-} ESCs showed no significant differences in their DNA methylation patterns compared to TDG proficient cells, *in vitro* neuronal

differentiation gave rise to 942 differentially methylated regions (DMRs), and the comparison of the MEFs revealed 32976 DMRs (Fig.1b). 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.1a).

Of the 942 DMRs found in NPs, 609 are hypermethylated and 333 hypomethylated in *Tdg* knockout cells compared to wildtype. As DNA methylation is not equally distributed throughout the genome and was found to negatively correlate with the density of CpG dinucleotides (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.1c). We also analyzed 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.1d). 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, but that a greater proportion of the hypomethylated DMRs are promoter-associated than of the hypermethylated (Fig.1e). We thus conclude that the hypomethylated DMRs are more likely to affect gene expression than the hypermethylated, but aberrant methylation appears to affect mostly regions distal to promoters.

Differentiation-associated *de novo* methylation of CGIs is diminished in absence of TDG

We found 123 DMRs to overlap with CGIs as defined in the UCSC genome analysis tools (Gardiner-Garden and Frommer 1987). Unexpectedly, 122 of these 123 differentially methylated CGIs, henceforth called CGI DMRs, were hypomethylated in NPs derived from TDG deficient ESCs, whereas DMRs not classified as CGIs were predominantly hypermethylated (hyper:hypomethylated = 3:1)

(Fig.2a). 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 the hypomethylation at CGIs in *Tdg* knockout 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 showing increased methylation in wildtype NPs compared to wildtype ESCs in our setup (Wilson 2012). This revealed that 117 of the 122 CGI DMRs overlap with CGIs that become methylated with differentiation (Fig.2b). Thus, hypomethylation of CGIs in *Tdg* knockout NPs is associated with diminished differentiation-associated *de novo* methylation.

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 and tested for the enrichment or depletion of specific elements (Fig.2c). 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. On the other hand, we found the CGI DMRs to be enriched in sites of TET1 binding and H3K4me1 and H3K27me3 modification, 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 enriched for H3K4me1 and TET1 but 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.2c), whereas constitutive LMRs showed no significant overlap. The CGI DMRs thus appear to be enriched for polycomb targets and poised enhancer elements that acquire *de novo* methylation during differentiation.

Cells lacking TDG activity exhibit hypomethylation of CGIs but no rise in C→T mutations

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 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 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.3). 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 5-mC loss nor the developmental knockout 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 both protected from deamination by Na-bisulfite and can

thus not be distinguished by this method. MeDIP-seq, on the other hand, discriminates between the two C modifications as it 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 the maintenance DNA methyltransferase DNMT1 (Valinluck and Sowers 2007; Hashimoto et al. 2012) and therefore is expected to occur more often opposite an unmethylated CpG.

Aberrant levels of higher oxidized C-modifications accumulate in the absence of TDG

TDG was proposed to be the only DNA glycosylase capable of excising the higher oxidized 5-hmC-derivatives 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, since neither 5-hmC nor the higher oxidized C-modifications are maintained by DNMT1 (Valinluck and Sowers 2007; Inoue et al. 2011), the quantitative analysis of the generation of these modifications is likely 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 for a homogeneously undifferentiated state in 2i medium prior to RA-induced differentiation (Ying et al. 2008). Remarkably, 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.S2b) by about 50%, irrespective of the status of TDG activity. 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 (Reik, W., personal communication).

We harvested genomic DNA and chromatin after 0, 8 and 24 hours of incubation with RA, RNA and protein at two additional timepoints (2i medium control and 4h) (Fig.4a). By testing the mRNA levels of *Nanog*, *Oct4*, *Rex1* and *Gata6* we verified the loss of expression of pluripotency genes and induction of developmental genes within these 24 h of RA differentiation (Fig.S1a), and we confirmed at mRNA and protein level that TET1 and TET2 expression was equal in all three cell lines (Fig.S1a and b and Fig.6b). The levels of AID mRNA were extremely low and protein levels were below the detection limit in Western blot analysis (Fig.S1c).

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.4b). We also found the global levels of 5-mC, 5-fC and 5-caC to rise with differentiation, and this effect was more pronounced in cells lacking TDG activity and specifically induced by RA (Fig.S2a 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.4b). Similarly, global 5-fC levels in knockout and $TDG_{\Delta cat}$ cells increased significantly above wildtype levels only after 8 and 24 h of differentiation. By contrast, 5-caC levels were ~9 fold higher in undifferentiated knockout and $TDG_{\Delta cat}$ cells than in wildtype ESCs and only the mutant cell lines showed a further significant increase in 5-caC with differentiation (Fig.S2a). 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 catalytic-dead cells, implicating a catalytic active role of TDG in controlling transitions in CpG methylation.

TDG activity is required to balance 5-mC and 5-caC at CGIs

The absence of increased C→T mutations at CGI DMRs in NPs derived from TDG deficient ESCs indicated that the hypomethylation observed 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} ESCs set against no change or even a small decrease in knockout cells (Fig.5a). 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.5a, aggregated p-value < 0.0001, Anova). Comparing the proportions of 5-mC and 5-caC modifications at 0 and 24 h (Fig.S3a) revealed a shift in the equilibrium between these modifications in a time- and genotype-dependent manner. 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.S3a).

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, thus requiring 5-caC to be excised and replaced with an unmodified C for subsequent re-methylation by DNMT3a or DNMT3b. The fact that *Tdg* knockout 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 with respect to its activity on and association with 5-caC. As TDG_{Δcat} shows residual glycosylase activity on substrates with a weak N-glycosidic bond, e.g. 5-FU (Kunz et al. 2009), 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.5b), which is in agreement with the accumulation of 5-caC in TDG_{Δcat} cells (Fig.4b). We then tested the ability of the catalytic-dead TDG to bind the 5-caC substrate in electrophoretic mobility-shift assays with fluorescence-labeled substrate oligonucleotides (G•T, G•5-mC, G•5-hmC, G•5-caC) in the presence of a 10 or 20 fold excess of unlabeled 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.5c); the binding specificity and efficiency of the catalytic-dead protein appeared to be comparable to that of the wildtype TDG (Fig.S3b) 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-activity facilitates stable association of TET1 to CGIs during differentiation

Given the differential effects of the *Tdg* disruption and the catalytic-dead mutant on the methylation-demethylation equilibrium at the CGI DMRs, 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.S1a) – 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.6a 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.5a), the absence of the catalytic activity in this mutant significantly destabilizes – or suppresses – TET1 association to the CGI DMRs (Fig.6a). The high affinity of TDG $_{\Delta cat}$ to 5-caC combined with its inability to turn over (Fig.5b 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.6c, for relative enrichment, controls and statistics see Fig.S5).

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. The majority of differentially methylated regions arising with lineage restriction of *Tdg* knockout cells show aberrantly increased levels of CpG methylation,

which is in agreement with a previously proposed function of TDG in preventing *de novo* methylation errors (Cortazar et al. 2011). Remarkably though, one third of the DMRs showed reduced levels of methylation in the *Tdg* knockout cells. These hypomethylated regions display a higher CpG density, closer proximity to TSS and a greater overlap with gene promoters than the hypermethylated DMRs, suggesting that the hypomethylation phenotype is connected with the regulation of gene expression. Hypomethylated DMRs arising in NPs are highly enriched for CGIs, and this translates into 99% of all differentially methylated CGIs being 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, evident from a 75% overlap of CGI DMRs with H3K4me1 enriched regions but only a minor overlap with sites of H3K27ac and p300 enrichment (Creyghton et al. 2010). Furthermore, 52% of the CGI DMRs overlap with NP-specific LMRs (Stadler et al. 2011). At first glance, the overlap with LMRs, which are distal regulatory regions that display variable methylation levels due to transcription factor binding, and the gain of methylation at these enhancer CGIs, which correlates with a transcriptionally inactive state of the corresponding gene, might appear counter-intuitive. However, it is important to bear in mind that also the binding of transcriptional repressors can produce an LMR (Stadler et al. 2011).

We found that a catalytically-dead but structurally intact TDG variant ($TDG_{\Delta cat}$) 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 catalyzed by AID in direct interaction with TDG (Cortellino et al. 2011). The deamination products of 5-mC (T) and 5-hmC (5-hmU) are both pre-mutagenic and would result in C→T transition mutations if unrepaired prior to DNA replication. Yet, by hairpin bisulfite-sequencing with high coverage of a representative set of hypomethylated CGIs, we did not detect

increased mutation levels, 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. Consistently, we did not observe a significant increase of 5-hmU in the genomic DNA of cells lacking TDG activity, neither with regard to differentiation nor to genotype. Together with recent reports of AID and other APOBEC proteins showing little to no reactivity on 5-mC and 5-hmC (Nabel et al. 2012), 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, with 5-fC and 5-caC specifically accumulating in TDG deficient cells, which is agreement with previous reports of TDG being the only DNA glycosylase excising these bases (He et al. 2011; Maiti and Drohat 2011). From the similarly elevated levels of 5-fC and 5-caC in TDG_{Δcat} and from a glycosylase activity assay with recombinant TDG_{Δcat}, we conclude that the N151A mutation abolishes TDG's 5-caC processing capacity (Fig.5b). However, the mutation does not disrupt binding of 5-caC, which we found TDG to associate more tightly with than with C, 5-mC or 5-hmC (Fig.5c). It is somewhat striking that relatively high levels of 5-fC and 5-caC accumulate in TDG deficient ESCs, considering that these undergo DNA replication every 13-16h and 5-mC oxidation products fail to support maintenance methylation by DNMT1 (Valinluck and Sowers 2007; Inoue et al. 2011; Hashimoto et al. 2012). This finding thus suggests that the higher oxidized 5-mC-modifications are continuously generated at the same rate as replication-dependent dilution occurs. To what extent they arise as consequence of epigenetic modeling or simply by chemical oxidation of the considerable genomic 5-mC and 5-hmC pool remains to be tested. In any case, it is clear that TDG is required to keep control of the global 5-fC/caC levels.

The analysis of local C-modification levels at representative CGI DMRs revealed a differentiation-induced increase of 5-mC and 5-caC levels in wildtype, a slight rise in 5-mC coupled with pronounced

accumulation of 5-caC in TDG $_{\Delta cat}$ but no significant shift of C-modification levels in knockout cells (Fig.5a and Fig.S3a). 5-hmC levels appear to change only slightly during differentiation, which may reflect that this oxidation intermediate is rapidly converted to 5-fC and 5-caC.

Although the differences that arise in the 24 h interval of *in vitro* differentiation are not dramatic, they indicate clearly that the loss of pluripotency coincides with a transition of DNA methylation states at certain CGIs that involves the generation of higher oxidized derivatives of 5-mC. As the overall trajectory at these CGIs is towards 5-mC enrichment, it is evident that these higher oxidized 5-mC-modifications have to be erased and replaced again by 5-mC. 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 thus 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.7). This establishes a transient equilibrium of methylation and demethylation intermediates that, at later stages of differentiation (early NPs), is tipped towards methylation (Fig.7), suggesting that this cycle represents a transitory state that accompanies the loss of pluripotency.

As only wildtype and even more the TDG $_{\Delta cat}$ cells exhibit an accumulation of 5-caC at these CGIs, we propose a dual function of TDG, one as an enzymatic component and one as a structural scaffold in the assembly of key factors involved in the cycle of DNA methylation and demethylation. Indeed, we found that TET1 associates with these loci independent of TDG in ESCs but is rapidly lost upon differentiation in *Tdg* knockout cells. TDG $_{\Delta cat}$, on the other hand, can still function as a scaffold, as evident by the accumulation of 5-caC in a differentiation-dependent manner. Notably, this occurs despite the concomitant reduction of TET1 association with these regions in these cells, suggesting that the specific 5-caC binding capacity of TDG actively blocks the turnover of the 5-caC generated. This implies that with the accumulation of 5-caC, an increasing number of TDG $_{\Delta cat}$ molecules would bind at these loci but fail to catalyze the final step of demethylation, as indeed indicated by the

increased enrichment of TDG_{Δcat} at these DMRs (Fig.6c and Fig.S5). Wildtype TDG on the other hand is capable of rapid turnover on 5-caC and accordingly associates with the CGI DMRs only transiently.

While the failure of turnover of TDG_{Δcat} might disrupt the methylation-demethylation complex by steric hindrance, excision and subsequent repair of 5-caC itself might produce signals necessary for associating TET1. Alternatively, the failure to excise 5-caC might result in 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, the TET proteins lack a substrate and would thus lose association with the CGI DMRs. Passive erasure of 5-fC/caC through DNA replication would eventually result in the hypomethylation observed in NPs.

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). Here, we describe a cycle driven by the activity of DNMTs, TET1 (possibly also TET2) and TDG that reflects a transitory state at CpG islands that undergo methylation during differentiation. This cycle of CpG methylation and demethylation appears to be induced by differentiation which 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). In this cycle, TDG fulfills two functions: first, structural stabilization of a complex that drives the cycle and second, catalyzing the final step of 5-caC removal. However, what initiates the cycle at these specific CGIs and which signals determine the overall trajectory towards a methylated or an unmethylated state remains to be elucidated.

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 mm 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 µg 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/>; (Giardine et al. 2005; Blankenberg et al. 2010; 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 ALEX 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

EMSAs 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.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." Genome Biol **11**(10): R106.

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

Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Res **21**(3): 381-395.

- Bhutani, N., et al. (2010). "Reprogramming towards pluripotency requires AID-dependent DNA demethylation." Nature **463**(7284): 1042-1047.
- Bird, A., et al. (1985). "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA." Cell **40**(1): 91-99.
- Blankenberg, D., et al. (2010). "Galaxy: a web-based genome analysis tool for experimentalists." Curr Protoc Mol Biol **Chapter 19**: Unit 19 10 11-21.
- Boland, M. J. and J. K. Christman (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**(3): 492-504.
- Bruniquel, D. and R. H. Schwartz (2003). "Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process." Nat Immunol **4**(3): 235-240.
- Chavez, L., et al. (2010). "Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage." Genome Res **20**(10): 1441-1450.
- Cortazar, D., et al. (2007). "The enigmatic thymine DNA glycosylase." DNA Repair (Amst) **6**(4): 489-504.
- Cortazar, D., et al. (2011). "Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability." Nature **470**(7334): 419-423.
- Cortellino, S., et al. (2011). "Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair." Cell **146**(1): 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., et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.
- Dunham, I., et al. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.
- Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-282.
- Giardine, B., et al. (2005). "Galaxy: a platform for interactive large-scale genome analysis." Genome Res **15**(10): 1451-1455.
- Goecks, J., et al. (2010). "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." Genome Biol **11**(8): R86.
- Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.

- Hackett, J. A., et al. (2013). "Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine." *Science* **339**(6118): 448-452.
- Hardeland, U., et al. (2000). "Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis." *J Biol Chem* **275**(43): 33449-33456.
- Hardeland, U., et al. (2003). "The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs." *Nucleic Acids Res* **31**(9): 2261-2271.
- Hashimoto, H., et al. (2012). "Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation." *Nucleic Acids Res* **40**(11): 4841-4849.
- He, Y. F., et al. (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." *Science* **333**(6047): 1303-1307.
- Heintzman, N. D., et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." *Nat Genet* **39**(3): 311-318.
- Illingworth, R. S. and A. P. Bird (2009). "CpG islands--'a rough guide'." *FEBS Lett* **583**(11): 1713-1720.
- Inoue, A., et al. (2011). "Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development." *Cell Res* **21**(12): 1670-1676.
- Iqbal, K., et al. (2011). "Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine." *Proc Natl Acad Sci U S A* **108**(9): 3642-3647.
- Ito, S., et al. (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." *Science* **333**(6047): 1300-1303.
- Kangaspeska, S., et al. (2008). "Transient cyclical methylation of promoter DNA." *Nature* **452**(7183): 112-115.
- Kemmerich, K., et al. (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**(13): 6016-6025.
- Kent, W. J., et al. (2002). "The human genome browser at UCSC." *Genome Res* **12**(6): 996-1006.
- Kunz, C., et al. (2009). "Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil." *PLoS Biol* **7**(4): e91.
- Kuzmichev, A., et al. (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein." *Genes Dev* **16**(22): 2893-2905.
- Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." *Bioinformatics* **25**(14): 1754-1760.
- Li, Y. Q., et al. (2007). "Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair." *Nucleic Acids Res* **35**(2): 390-400.

- Lutsik, P., et al. (2011). "BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing." *Nucleic Acids Res* **39**(Web Server issue): W551-556.
- Maiti, A. and A. C. Drohat (2011). "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites." *J Biol Chem* **286**(41): 35334-35338.
- Meissner, A., et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." *Nature* **454**(7205): 766-770.
- Metivier, R., et al. (2008). "Cyclical DNA methylation of a transcriptionally active promoter." *Nature* **452**(7183): 45-50.
- Mohn, F., et al. (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." *Mol Cell* **30**(6): 755-766.
- Nabel, C. S., et al. (2012). "AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation." *Nat Chem Biol* **8**(9): 751-758.
- Nix, D. A., et al. (2008). "Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks." *BMC Bioinformatics* **9**: 523.
- Popp, C., et al. (2010). "Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency." *Nature* **463**(7284): 1101-1105.
- Rai, K., et al. (2008). "DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45." *Cell* **135**(7): 1201-1212.
- Serandour, A. A., et al. (2012). "Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers." *Nucleic Acids Res* **40**(17): 8255-8265.
- Stadler, M. B., et al. (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." *Nature* **480**(7378): 490-495.
- Tahiliani, M., et al. (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." *Science* **324**(5929): 930-935.
- Valinluck, V. and L. C. Sowers (2007). "Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1." *Cancer Res* **67**(3): 946-950.
- Weber, M., et al. (2005). "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." *Nat Genet* **37**(8): 853-862.
- Wilson, G. A. (2012). "Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers." *Gigascience* **1**(3).
- Wong, E., et al. (2002). "Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation." *Proc Natl Acad Sci U S A* **99**(23): 14937-14942.

- Wossidlo, M., et al. (2011). "5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming." Nat Commun **2**: 241.
- Ying, Q. L., et al. (2008). "The ground state of embryonic stem cell self-renewal." Nature **453**(7194): 519-523.
- Zhu, B., et al. (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**(10): 5135-5139.
- Zhu, J. K. (2009). "Active DNA demethylation mediated by DNA glycosylases." Annu Rev Genet **43**: 143-166.

Figure Legends

Figure 1: Differentiated *Tdg*^{-/-} cells accumulate aberrant methylation patterns. 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) MeDIP-seq reveals an increasing number of differentially methylated regions (DMRs) with advancing differentiation state. c) 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. d) 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). e) Promoter-association of hypo- versus hypermethylated NP DMRs (broad peaks).

Figure 2: Differentiation-dependent methylation at CpG islands fails in absence of TDG. a) 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. b) 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. c) 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 3: 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 4: 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.S2). Shown are Log₂ fold changes compared to wildtype (mean with s.e.m.); statistical analysis was performed on absolute numbers (Fig.S2), asterisks indicate significant difference to the respective timepoint in wt, * p<0.05, **p<0.005, *** p<0.0001, one-way Anova.

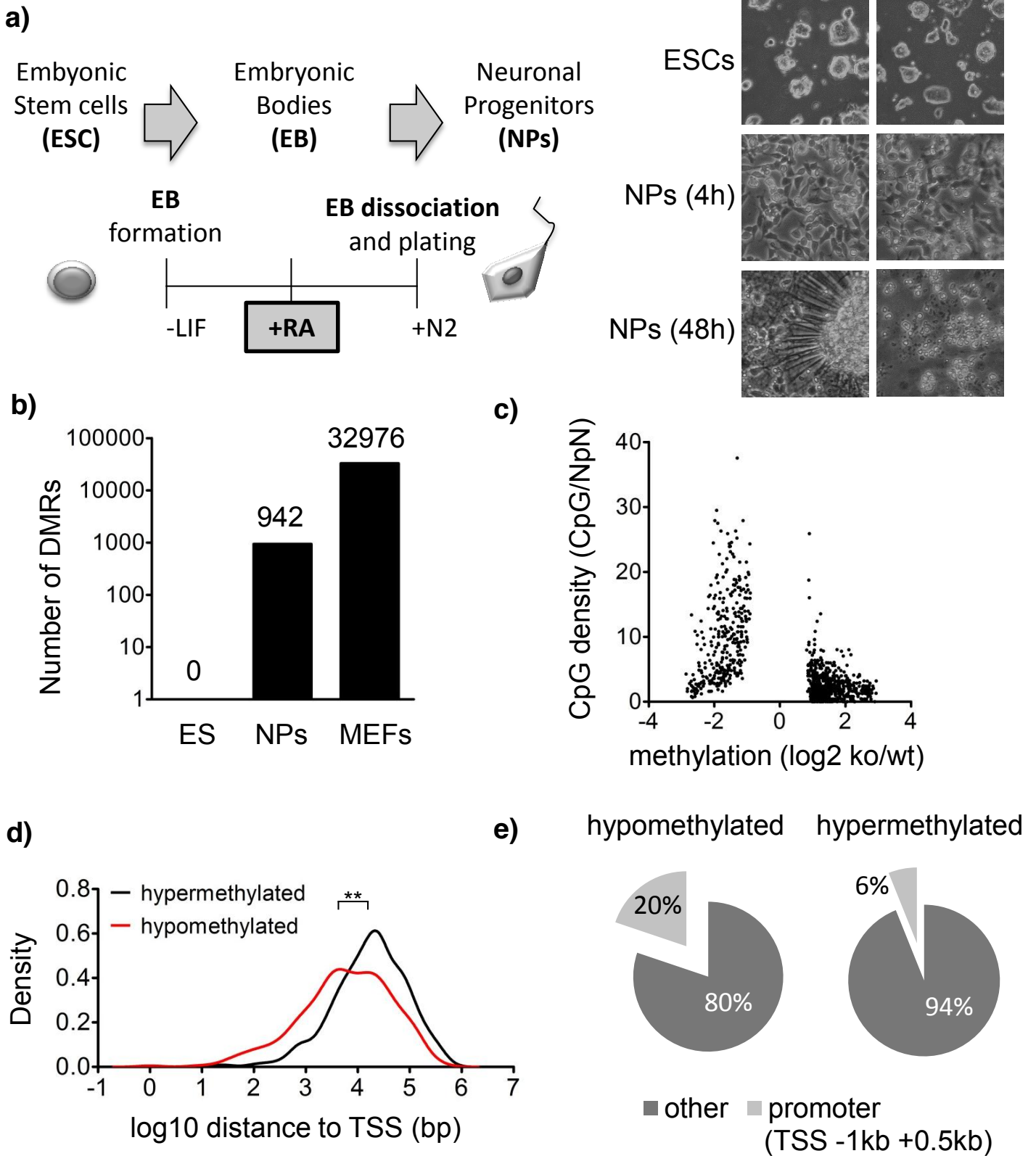
Figure 5: Targeted MeDIP, GLIB and caCDIP analysis of CGI DMRs. a) Relative enrichment (RE) of 5-mC (n=2) and 5-caC (n=3) normalized to a randomly chosen CpG-poor region (neg.contr.). TDG and TDG_{Δcat} accumulate 5-mC and 5-caC over time, 5-caC especially accumulates in TDG_{Δcat}. Mean with s.e.m. 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 6: 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. S4. 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. S5. d) TDG and β-actin protein levels in SDS protein extracts detected by Western blot.

Figure 7: 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.

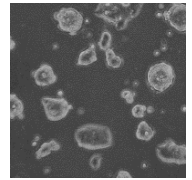
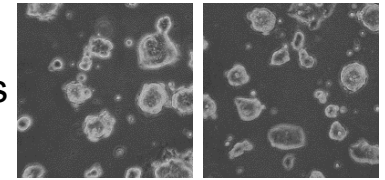
Figure 1



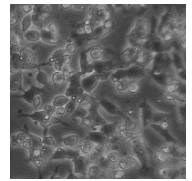
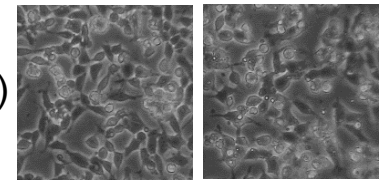
Appendix III

Tdg^{+/-} *Tdg*^{-/-}

ESCs



NPs (4h)



NPs (48h)

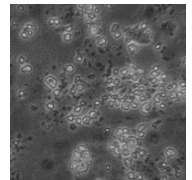
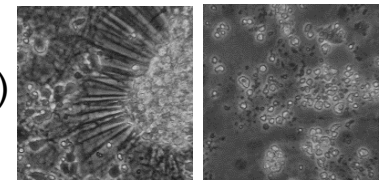


Figure 2

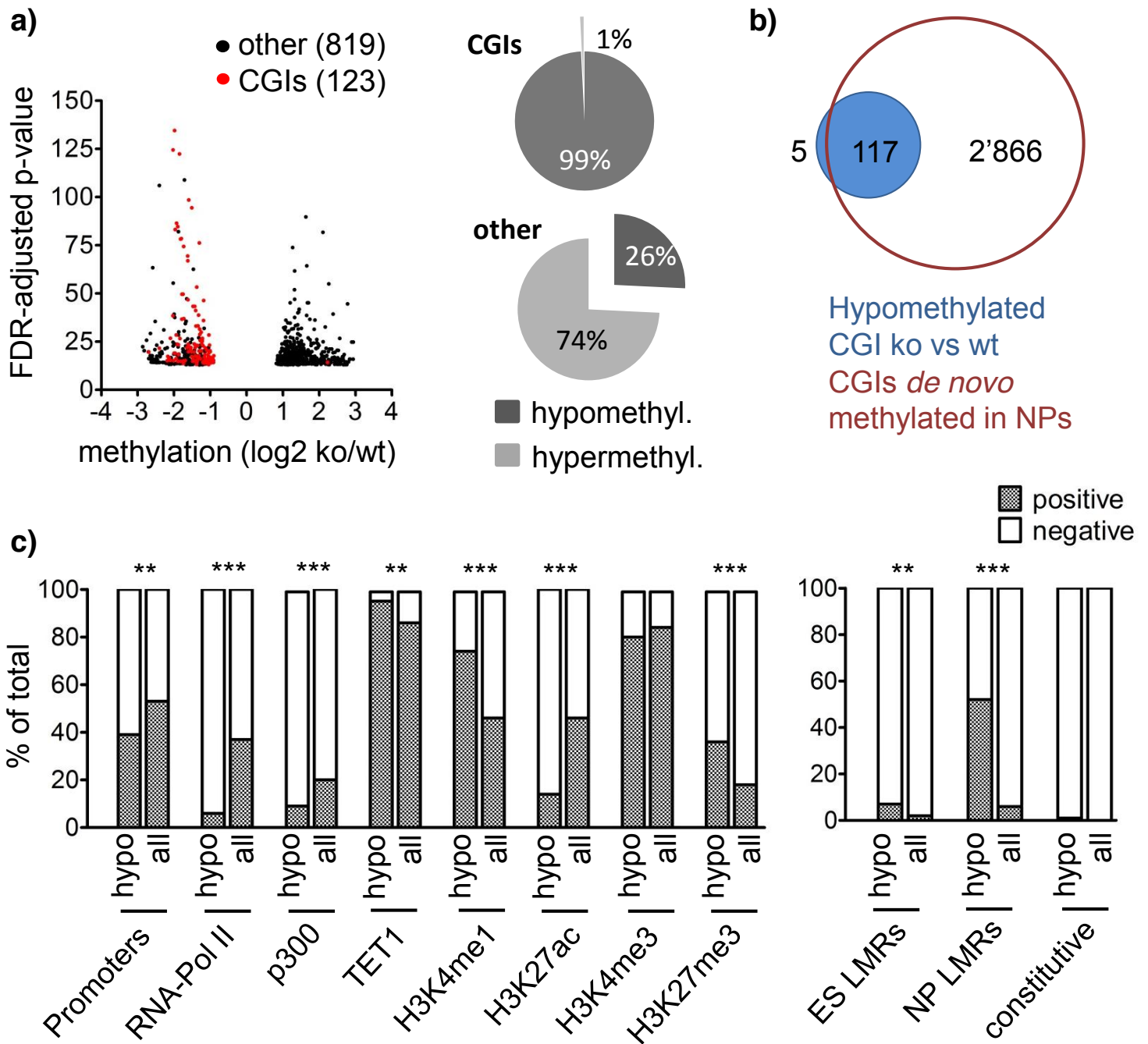


Figure 3

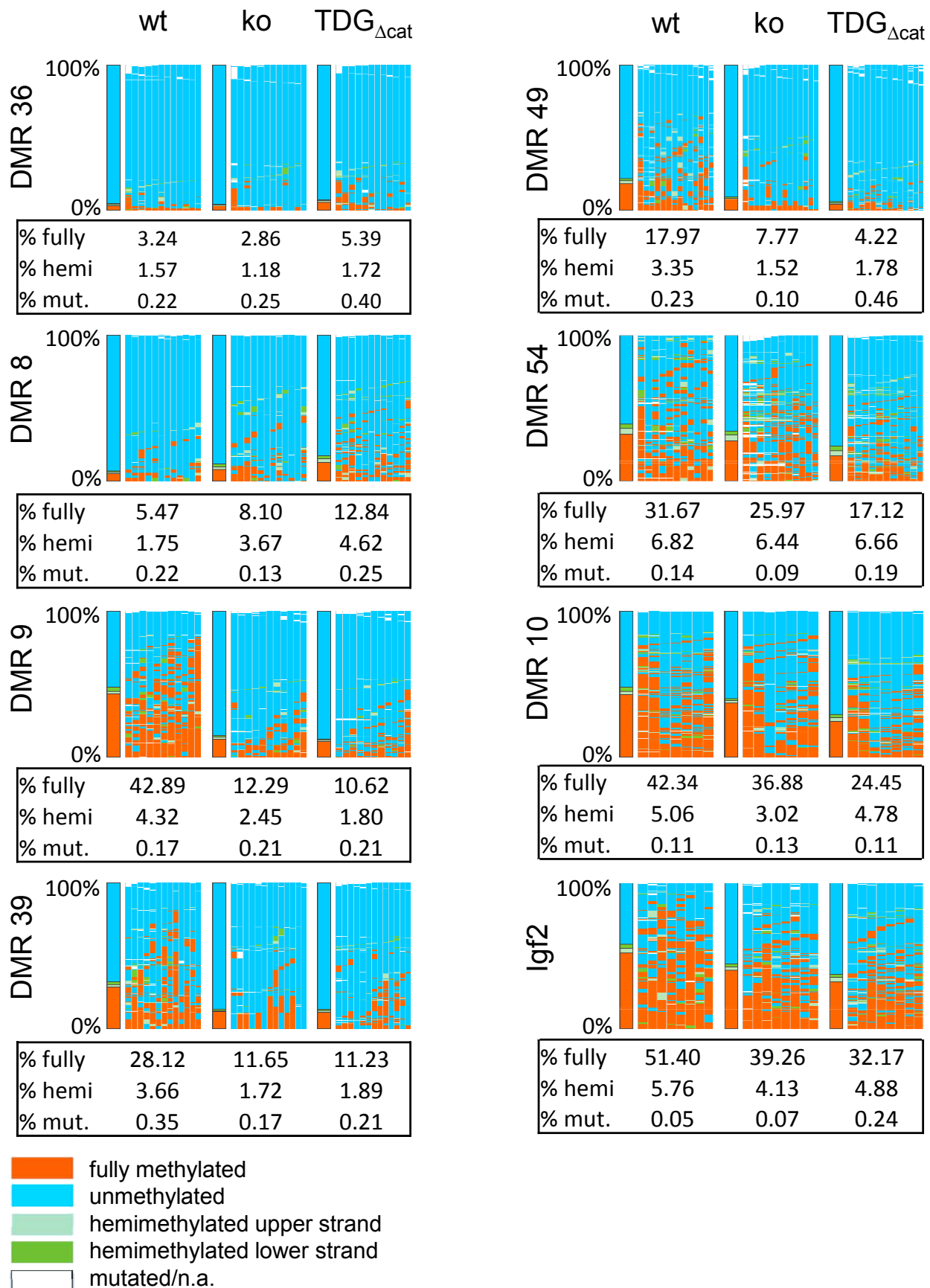


Figure 4

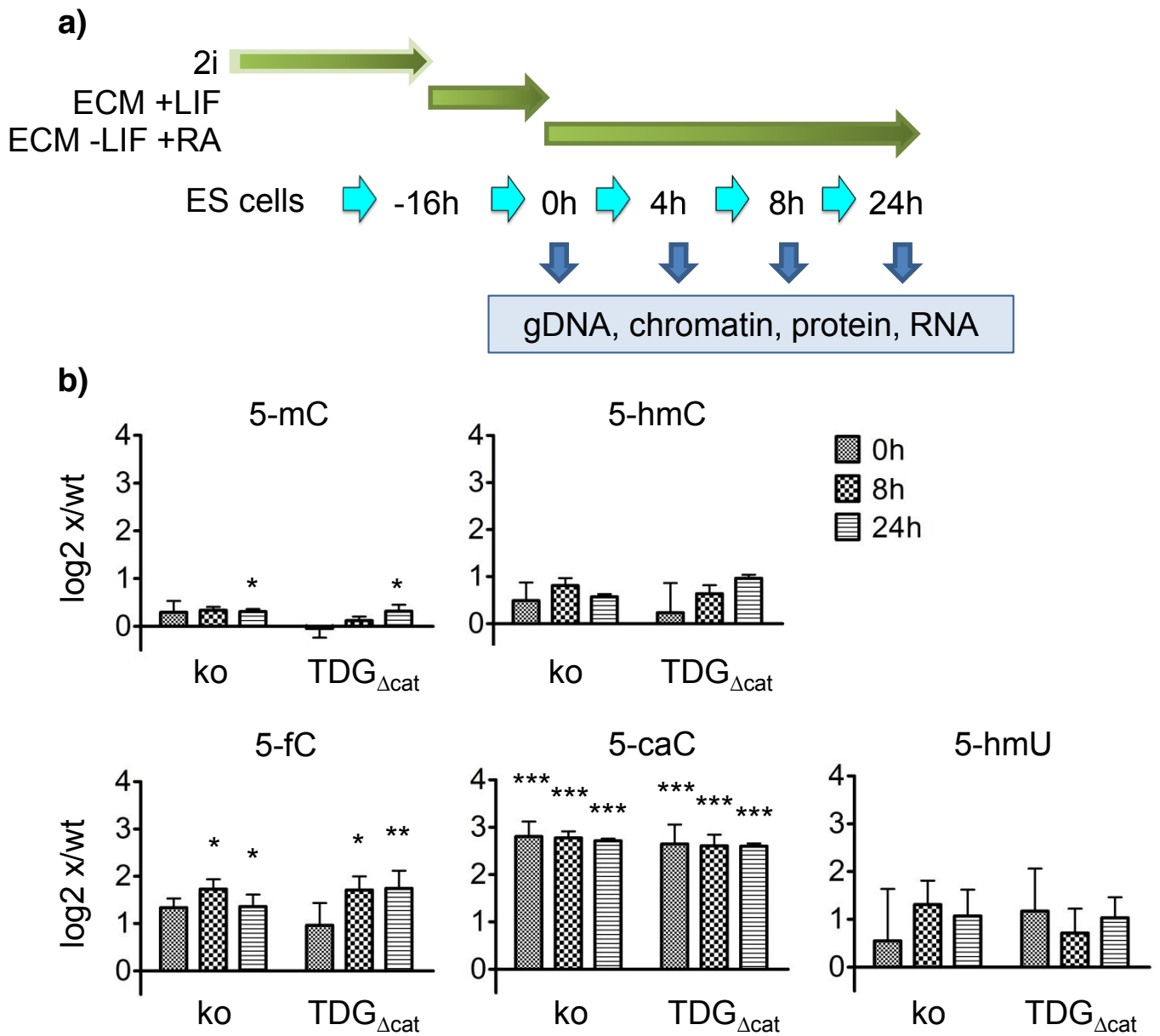


Figure 5

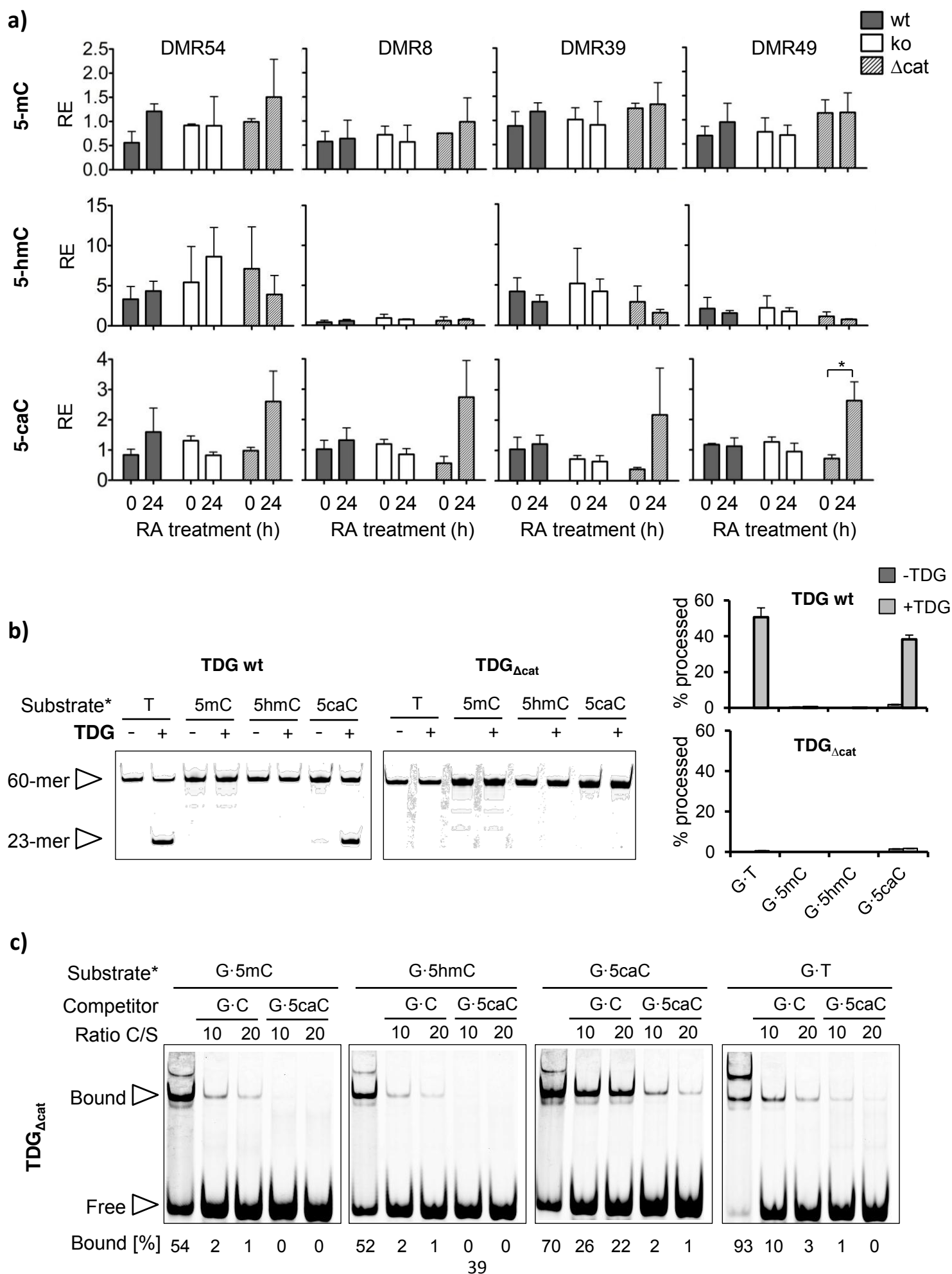


Figure 6

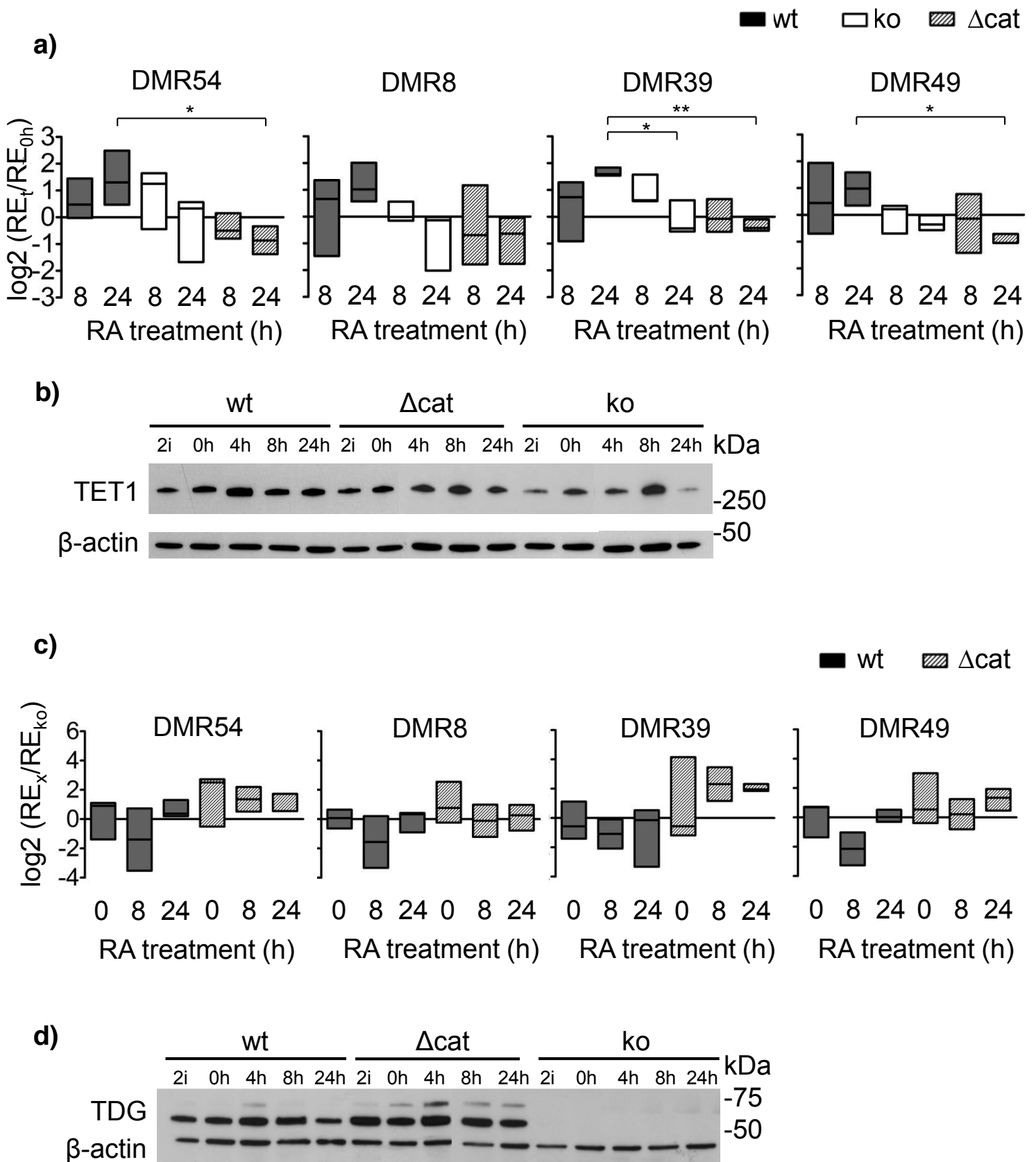
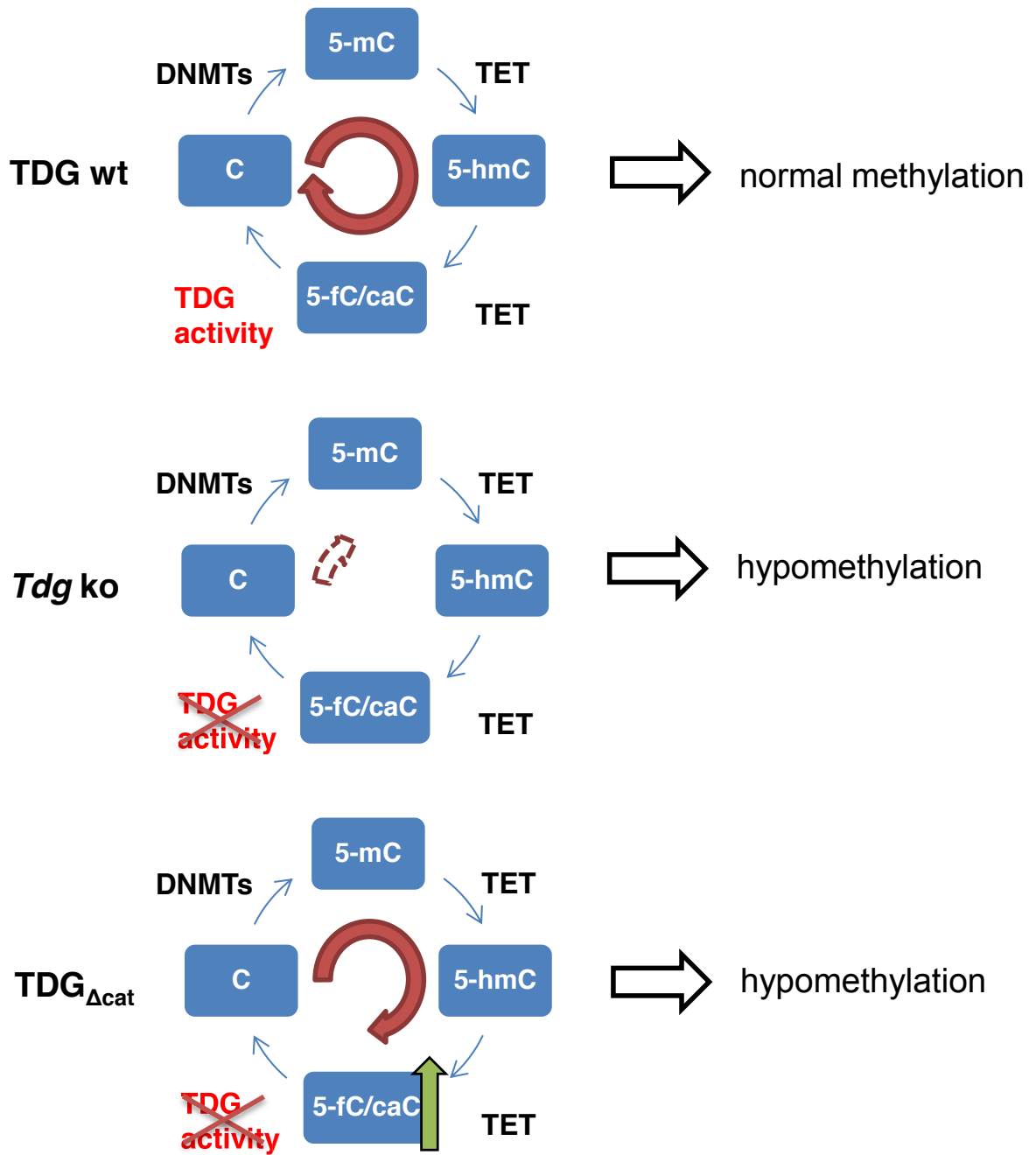


Figure 7



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 3561; Fax: +41 61 267 3566

Supplementary Figure Legends

Figure S1: 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 S2: Global C-modification levels measured by LCMSMS at 0, 8 and 24h of differentiation. a) Absolute numbers per 10^6 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 S3: 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 Td $g_{\Delta cat}$ the equilibrium of 5-mC and 5-caC is tipped towards 5-caC upon differentiation.

Figure S4: 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 S5: 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.

Figure S1: mRNA and protein levels, 24h RA timecourse

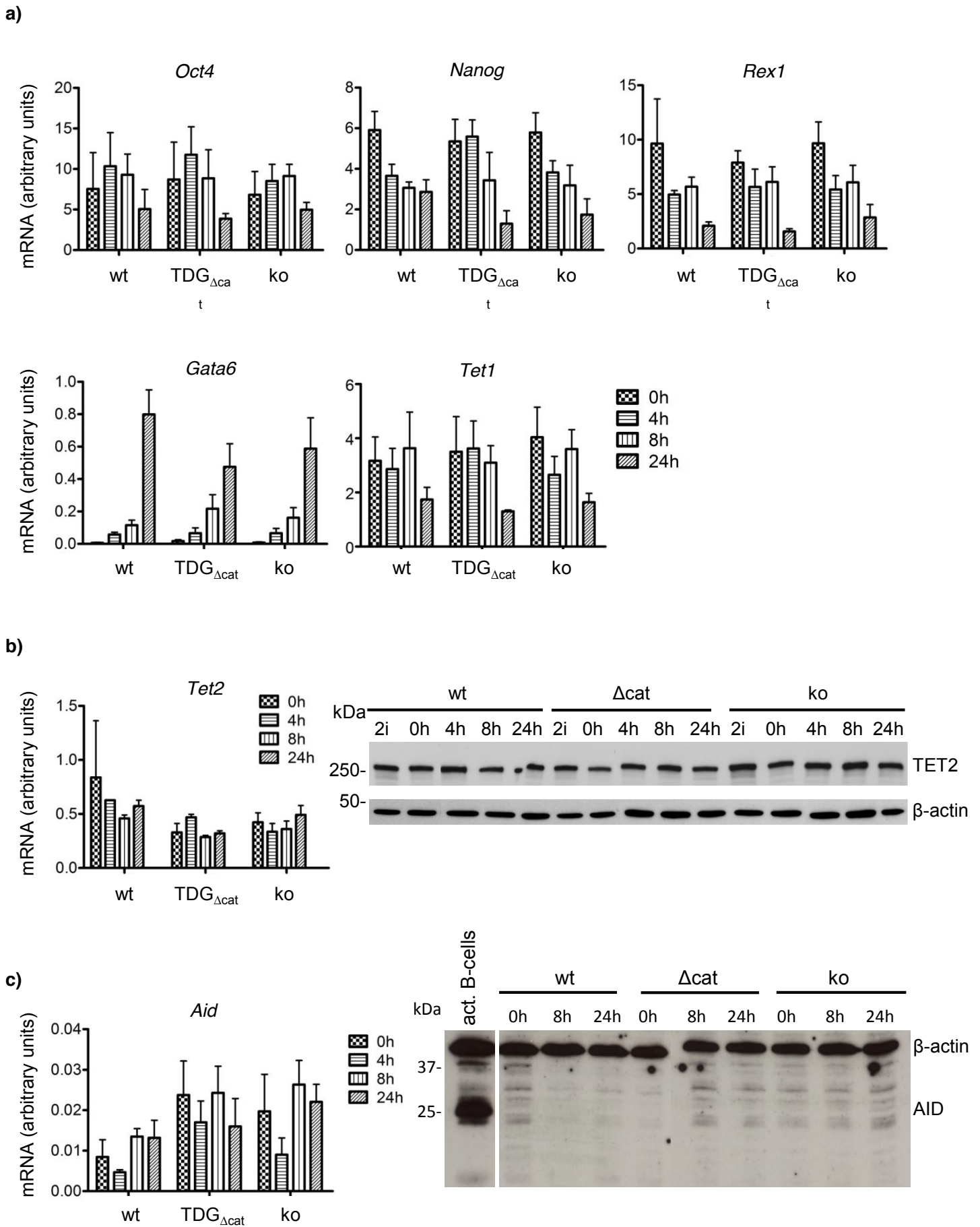


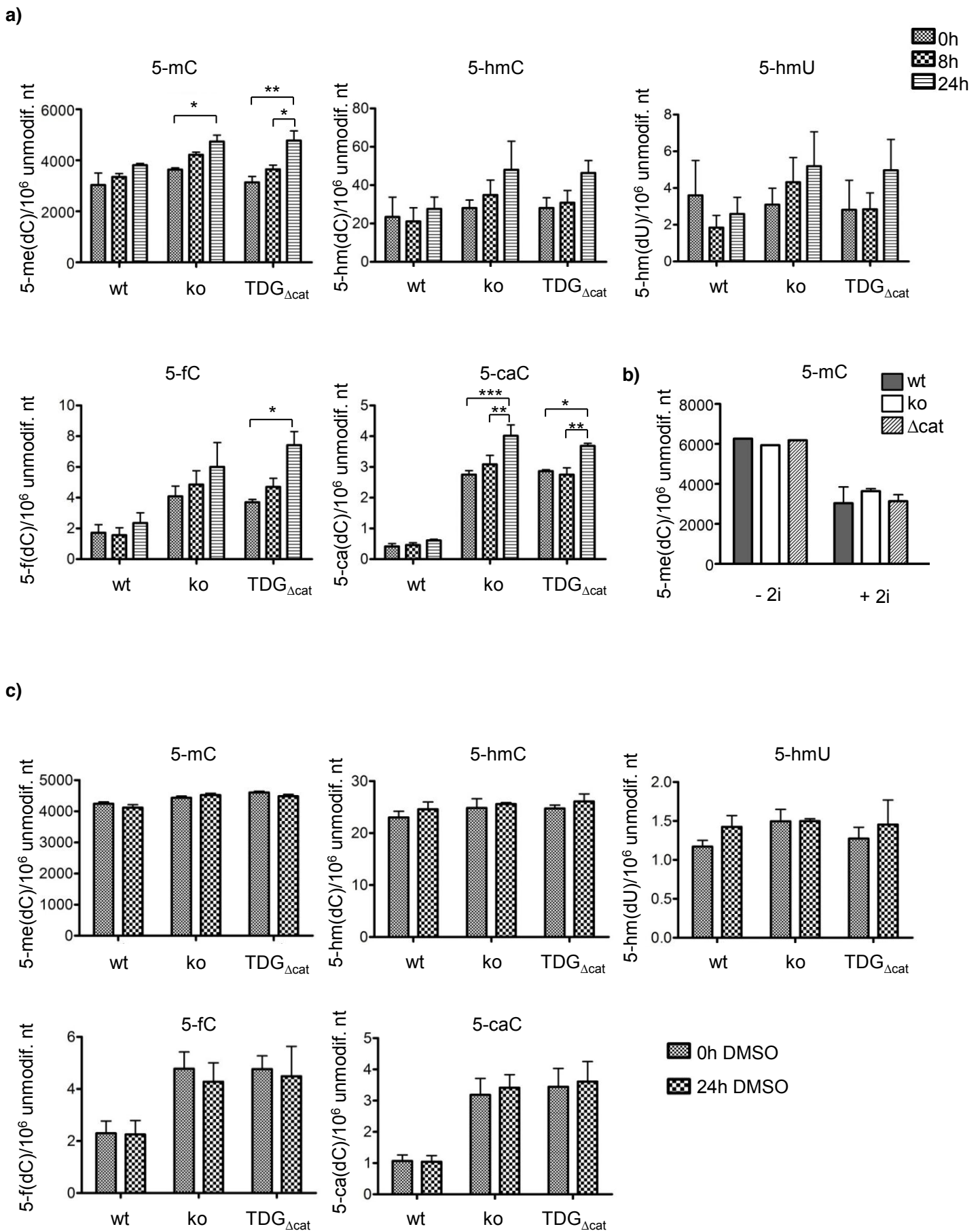
Figure S2: global C-modification levels, absolute numbers per 10⁶ unmodified nt

Figure S3: Targeted analysis of 5-mC, 5-hmC and 5-caC levels at CGI DMRs

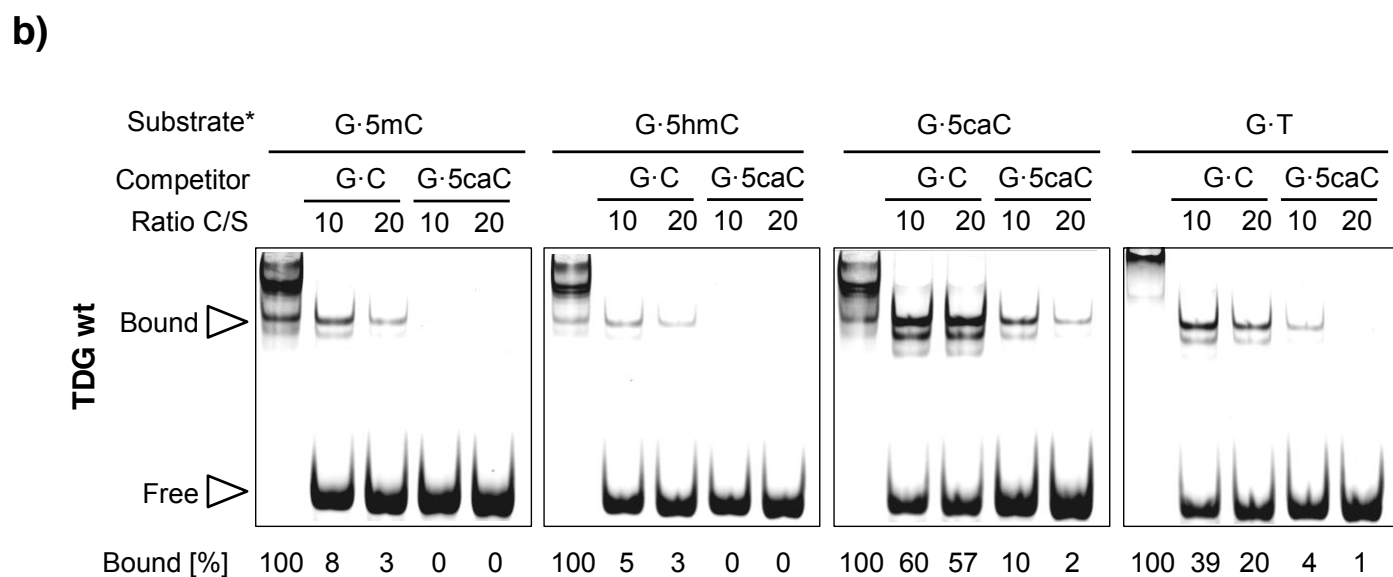
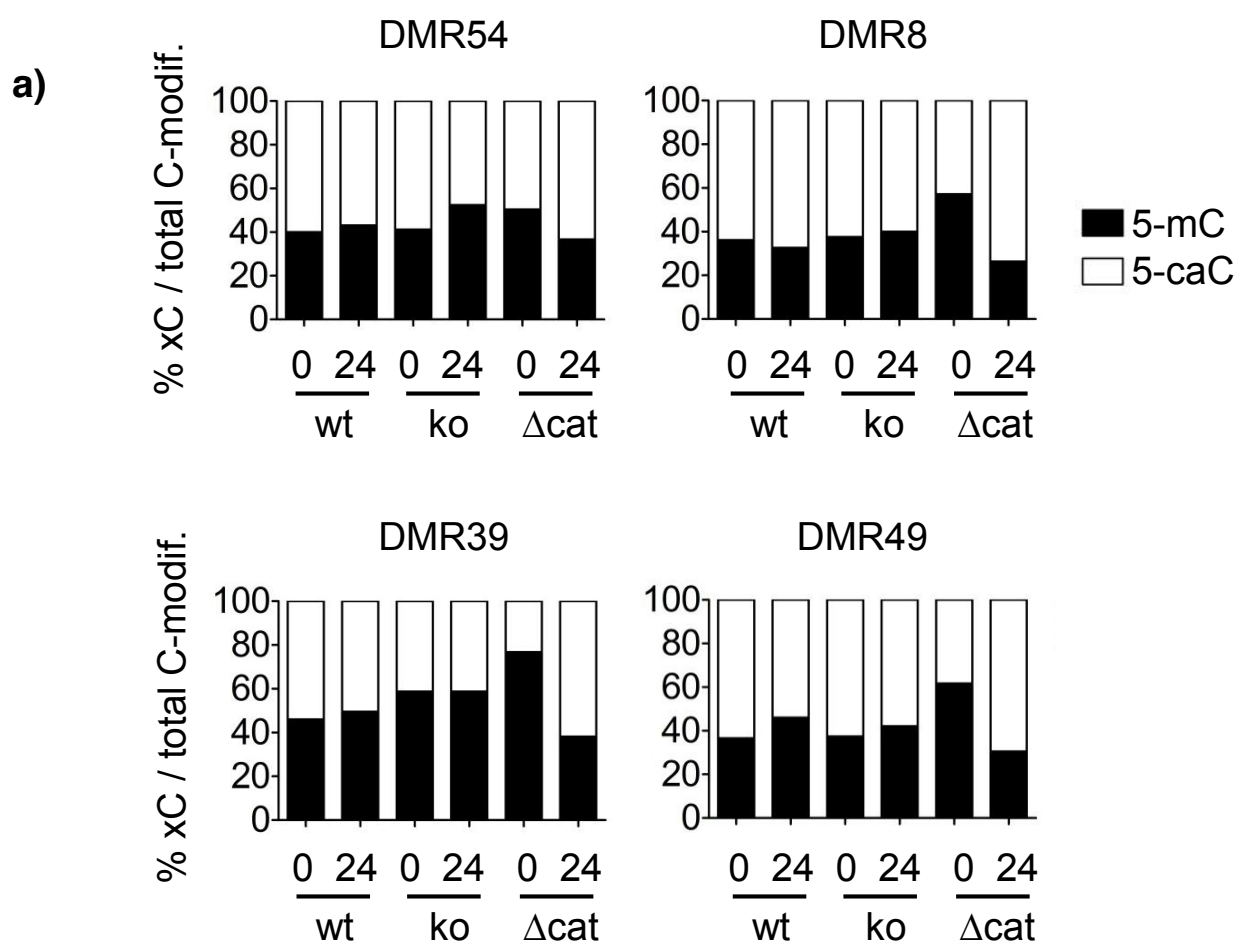


Figure S4: TET1-ChIP, relative enrichment versus chr2neg

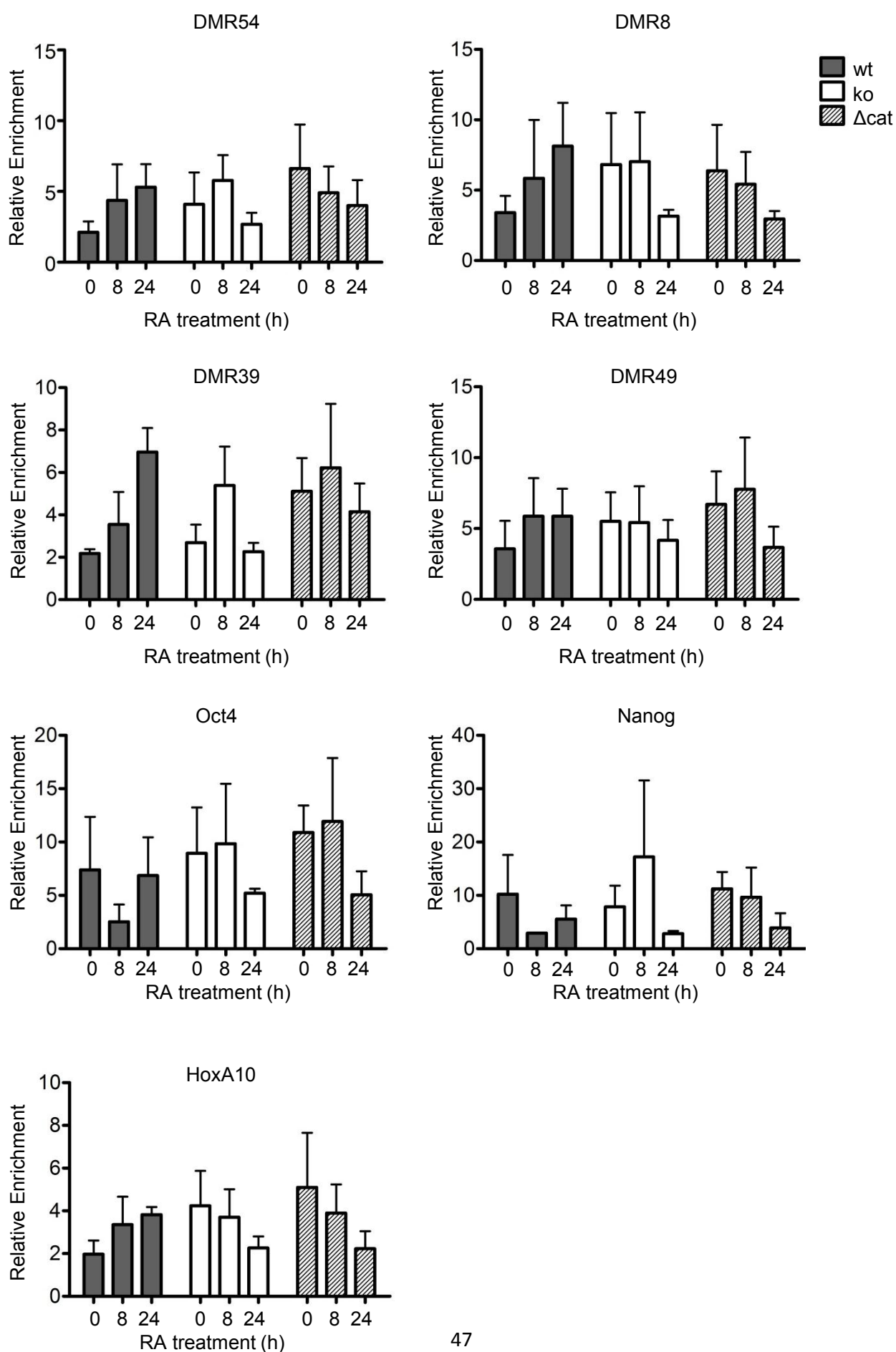
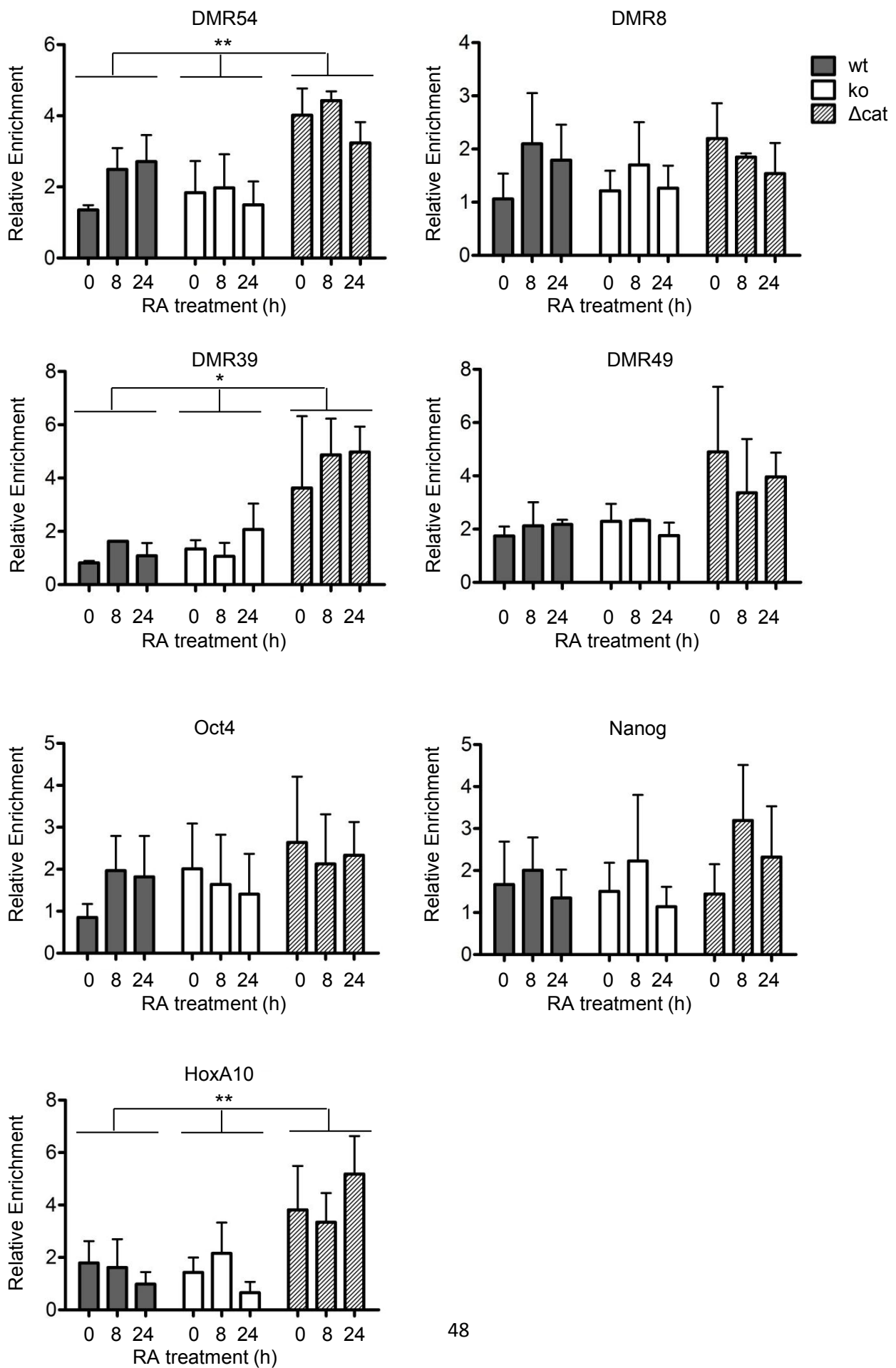


Figure S5: TDG-ChIP, relative enrichment versus chr2neg



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 AAAAACTAAACCAAAAAAC
9	TaqI	F TTTTAGGAGATATAAAGAATAGTTT
		R AAAAAACAAAAAACAATC
10	PstI	F AGAAGAGTTTAATTGTTATTTTGG
		R AAACCTCAACTACCACTCAACC
36	BamHI	F TTTGATATTTTTTTTAGTTTT
		R CCTAACACTTTCTCTTAATTT
39	TaqI	F GGATGTAGGTATTGATTAT
		R ACCTACCAAACCTTACAA
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™ 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	ACCCAGCAAAATCTCACCTG
DMR 54 R	GACTACTGGACAGGGCTCCA
DMR 39 F	GAGCTGGATAGCCCTGTAGAATG
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	GGGGGTACAGTTCAAGAGC

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.

DNA glycosylases: in DNA repair and beyond

Angelika L. Jacobs · Primo Schär

Received: 6 September 2011 / Revised: 10 October 2011 / Accepted: 11 October 2011 / Published online: 3 November 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The base excision repair machinery protects DNA in cells from the damaging effects of oxidation, alkylation, and deamination; it is specialized to fix single-base damage in the form of small chemical modifications. Base modifications can be mutagenic and/or cytotoxic, depending on how they interfere with the template function of the DNA during replication and transcription. DNA glycosylases play a key role in the elimination of such DNA lesions; they recognize and excise damaged bases, thereby initiating a repair process that restores the regular DNA structure with high accuracy. All glycosylases share a common mode of action for damage recognition; they flip bases out of the DNA helix into a selective active site pocket, the architecture of which permits a sensitive detection of even minor base irregularities. Within the past few years, it has become clear that nature has exploited this ability to read the chemical structure of DNA bases for purposes other than canonical DNA repair. DNA glycosylases have been brought into context with molecular processes relating to innate and adaptive immunity as well as to the control of DNA methylation and epigenetic stability. Here, we summarize the key structural and mechanistic features of DNA glycosylases with a special focus on the mammalian enzymes, and then review the evidence for the newly emerging biological functions beyond the protection of genome integrity.

Introduction

The integrity of genetic information is under constant threat by the tendency of DNA to engage in chemical reactions in its cellular environment. These can damage the DNA in various ways, most frequently by oxidation, alkylation, or deamination of the coding bases (Lindahl and Wood 1999). Damage to DNA bases may affect their base-pairing properties and, therefore, needs to be fixed to maintain the template function of the DNA (Kunz et al. 2009a). Many base lesions are pro-mutagenic, i.e., they give rise to genetic mutations if not repaired. One such example is 7,8-dihydro-8-oxoguanine (8-oxoG), a frequent product of DNA oxidation. 8-oxoG tends to base-pair with adenine, thus giving rise to G•C to T•A transversion mutations. Likewise, hydrolytic deamination of cytosine and 5-methylcytosine (5-meC) gives rise to uracil and thymine mispaired with guanine, respectively, causing C•G→T•A transition mutations if not repaired. Alkylation can generate a variety of DNA base lesions comprising O6-methylguanine (6-meG), N7-methylguanine (7-meG), or N3-methyladenine (3-meA). While 6-meG is pro-mutagenic by its property to pair with thymine, 7-meG and 3-meA block replicative DNA polymerases and are therefore cytotoxic (Lindahl and Wood 1999).

These and many other forms of DNA base damage arise in cells at least 10,000 times every day and only the continuous action of specialized DNA repair systems can prevent a rapid decay of genetic information. Single-base lesions are eliminated by base excision repair (BER), a pathway initiated by DNA glycosylases that recognize and excise damaged bases. Base removal by a DNA glycosylase generates a so-called apurinic/apyrimidinic site (AP-site) in DNA, which is then further processed by specific AP-endonuclease, DNA polymerase, and DNA ligase activities

Communicated by: E. Nigg

A. L. Jacobs · P. Schär (✉)
Department of Biomedicine,
Institute of Biochemistry and Genetics,
University of Basel,
Mattenstrasse 28,
4058 Basel, Switzerland
e-mail: primo.schaer@unibas.ch

to restore the original DNA sequence (Fig. 1) (Almeida and Sobol 2007). Accordingly, cells lacking DNA glycosylase functions generally show increased levels of base damage in their DNA, elevated mutation rates, and hypersensitivity to specific DNA damaging agents. Surprisingly, however, the phenotype of DNA glycosylase disruptions in mice is usually rather moderate (reviewed in Robertson et al. 2009), the only known exception being the thymine DNA glycosylase (TDG), which was recently reported to be essential for embryonic development in mouse (Cortazar et al. 2011; Cortellino et al. 2011).

In this review, we will focus on the mammalian DNA glycosylases, for which we will briefly summarize the key structure–function concepts and discuss their role in the repair of DNA base lesions. We will further elaborate on their newly emerging functions beyond canonical DNA repair, e.g., in innate and adaptive immunity and in DNA methylation control.

DNA glycosylases—an ancient family of DNA repair proteins

The consideration that cells must possess an ability to remove uracil from DNA, which arises either by misincorporation of deoxyuridine monophosphate (dUMP) during DNA replication or by hydrolytic deamination of cytosine, led to the discovery of an enzyme capable of cleaving uracil–deoxyribose bonds, the uracil N-glycosidase (Ung) in *Escherichia coli* (Lindahl 1974). This finding was followed by the isolation of many other DNA glycosylases in species from all kingdoms of life. Eleven DNA glycosylases have been identified in mammals and these can be subdivided into four structurally distinct superfamilies; the uracil DNA glycosylases (UDGs), the helix-hairpin-helix (HhH) glycosylases, the 3-methyl-purine glycosylase (MPG), and the endonuclease VIII-like (NEIL) glycosylases (Table 1).

The uracil DNA glycosylases

E. coli Ung turned out to be the founding member of a large superfamily of glycosylases, which now includes six subfamilies, three of which are present in the eukarya. Besides the UNG subfamily, these include the mismatch-specific uracil DNA glycosylases (MUGs) (Gallinari and Jiricny 1996) and the single-strand-specific monofunctional uracil DNA glycosylases (SMUGs) (Haushalter et al. 1999). Despite a considerable amino-acid sequence divergence, all UDGs share a common alpha–beta fold structured catalytic domain (Aravind and Koonin 2000).

Members of the UNG subfamily have been characterized in organisms from bacteria and yeasts to humans and large

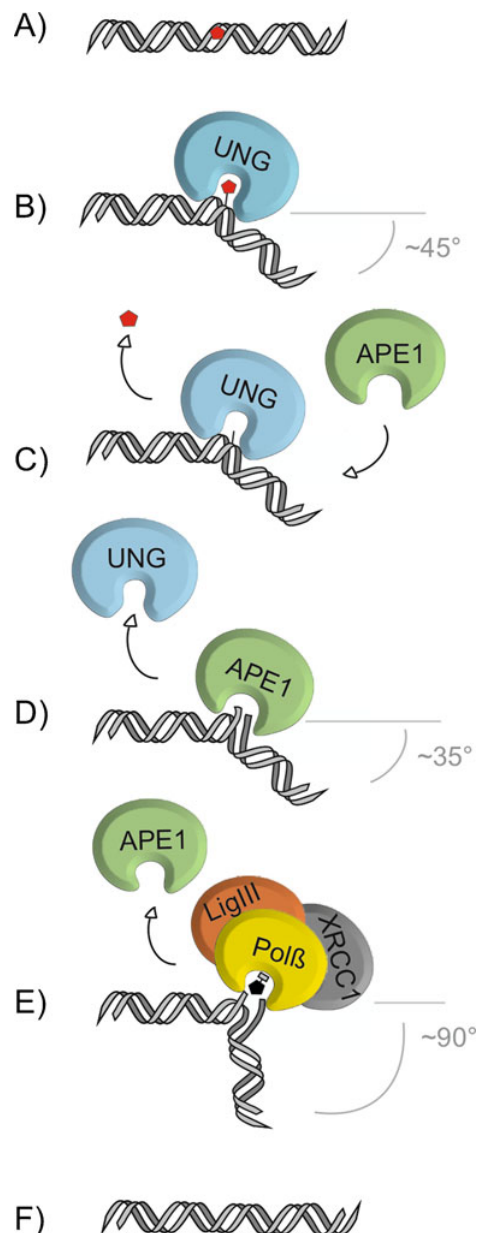


Fig. 1 The core pathway short-patch BER. The base-excision repair pathway addresses single-base lesions (a). BER is initiated by a DNA glycosylase, e.g., UNG, specifically recognizing and binding a base lesion. Upon encountering a substrate base, e.g., uracil for UNG, the glycosylase flips the base out of the base-stack into its catalytic site pocket where specific contacts examine the substrate base and position it for nucleophilic attack to the N-glycosidic bond (b). Release of the substrate base results in an abasic site (c), which is further processed by the AP-endonuclease, APE1, that cleaves the phosphate backbone 5' to the abasic site, producing a 3'OH and a 5' deoxyribose-phosphate moiety (5'dRP) (d). Polymerase β (*Pol* β) hydrolyzes the 5'dRP and fills in the single nucleotide gap, which is subsequently sealed by the DNA ligase III (*LigIII*), supported by the scaffold protein XRCC1 (e), thus restoring the original base sequence (f). The increase of DNA bending from UNG to *Pol* β might support the directionality of the handover from one BER factor to the next

Table 1 Mammalian DNA glycosylases, their main substrates, modes of action, and mutant phenotypes

Type of base lesion	Name	Name	Physiological substrates	Mono (M)/ bi(B) functional	Mouse knockout (ko)/ knockdown (kd) phenotype
Uracil in ssDNA dsDNA	UNG	Uracil-N glycosylase	U, 5-FU, ss and dsDNA	M	ko: viable, B-cell lymphomas, disturbed antibody diversification
	SMUG1	Single-strand-specific monofunctional uracil DNA glycosylase 1	U, 5-hmU, 5-FU, ss and dsDNA	M	kd: moderate increase in mutation frequency (C→T)
Pyrimidine derivates in mismatches	MBD4	Methyl-binding domain glycosylase 4	T, U, 5-FU, εC, opposite G, dsDNA	M	ko: viable, elevated mutation frequency (C→T)
	TDG	Thymine DNA glycosylase	T, U, 5-FU, εC, 5-hmU, 5-fC, 5-caC; opposite G, dsDNA	M	ko: embryonic lethal, aberrant DNA methylation and imbalanced chromatin marks in CpG-rich promoters
Oxidative base damage	OGG1	8-OxoG DNA glycosylase 1	8-oxoG, FaPy, opposite C, dsDNA	B	ko: viable, accumulation of 8-oxoG, elevated mutation frequency (G→T)
	MYH	MutY homolog DNA glycosylase	A opposite 8-oxoG, C or G, 2-hA opposite G, dsDNA	M	ko: viable, see OGG1
Alkylated purines	MPG	Methylpurine glycosylase	3-meA, 7-meG, 3-meG, hypoxanthine, εA, ss and dsDNA	M	ko: viable, elevated levels of ethenoA and hypoxanthine
Oxidized, ring-fragmented or –saturated pyrimidines	NTHL1	Endonuclease III-like 1	Tg, FaPyG, 5-hC, 5-hU, dsDNA	B	ko: viable
	NEIL1	Endonuclease VIII-like glycosylase 1	Tg, FaPyG, FaPyA, 8-oxoG, 5-hU, 5-hC, ss and dsDNA	B	ko: metabolic syndrome, increased damage levels in mitochondrial DNA; kd: hypersensitive to γ radiation
	NEIL2	Endonuclease VIII-like glycosylase 2	As NTHL1 and NEIL1	B	Unknown
	NEIL3	Endonuclease VIII-like glycosylase 3	FaPyG, FaPyA, prefers ssDNA	B	ko: normal

U, uracil; A, adenine; T, thymine; C, cytosine, G, guanine; ss single stranded; ds, double stranded; 5-hm, 5-hydroxymethyl; 5-FU, 5-fluorouracil; ε, etheno; 5-fC, 5-formylcytosine; 5-caC, 5-carboxylcytosine; 8-oxoG, 8-oxo-7,8-dihydroguanine; Tg, thymine glycol; FaPy, 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine; me, methyl; h, hydroxyl

"eukaryotic" viruses. These are highly conserved both at the amino-acid sequence and gene structure levels; the yeast and human proteins share 40.3% amino-acid sequence similarity and the human, mouse, and fish genes have identical exon–intron boundaries, indicating that the exon–intron organization has not changed for more than 450 million years (Krokan et al. 1997). Alternative splicing as well as transcription from two distinct start sites gives rise to the specific mitochondrial and nuclear isoforms UNG1 and UNG2 in mouse and human cells (Nilsen et al. 1997). UNG is highly specific for processing of uracil in DNA but also excises DNA-incorporated 5-fluorouracil (5-FU), a uracil analog used in cancer therapy (Pettersen et al. 2011). Interactions with PCNA and RPA target the nuclear UNG2 to sites of DNA synthesis, where its main function is to rapidly excise uracil that gets incorporated opposite from adenine (Otterlei et al. 1999). Accordingly, mouse cells deficient in UNG accumulate ~100-fold increased levels of uracil in their DNA but, notably, do not show a significant mutator phenotype (Nilsen et al. 2000). This is unlike

human cells where inhibition of UNG appears to elevate the mutation frequency mildly (Radany et al. 2000). However, mice lacking *Ung* do develop B-cell lymphomas and show disturbances of antibody diversification, implicating a specific function of UNG in processing deamination-induced U•G mismatches at immunoglobulin loci to facilitate somatic hypermutation and class switch recombination (Rada et al. 2002; Nilsen et al. 2003). Consistently, mutations in the human *UNG* gene have been associated with a subgroup of hyper-IgM syndrome patients, showing impaired class switch recombination (Imai et al. 2003).

The MUG subfamily of UDGs emerged with the identification of TDG, an enzyme capable of excising thymine from G•T mismatches. Nonetheless, the family was named after the *E.coli* Mug protein (Gallinari and Jiricny 1996), giving credit to the fact that the G•U rather than the G•T mismatch represents the common most efficiently processed substrate for the members of this subfamily. MUG orthologs have been described in *Schizosaccharomyces pombe*, *Drosophila melanogaster* (Hardeland et al. 2003), and mammals

(Neddermann et al. 1996). Crystal structural analyses revealed a striking resemblance in the overall fold of the glycosylase domains of *E. coli* Mug and Ung, despite the absence of a notable sequence similarity (Barrett et al. 1998). Unlike UNG, however, the MUG glycosylases have a spacious and rather non-discriminating active site pocket, accommodating a broad range of substrates including pyrimidine derivatives like 5-FU, 5-hydroxymethyluracil (5-hmU), and 3,N4-ethenocytosine (Table 1) (reviewed in Cortazar et al. 2007), and they process these substrates with an extremely low turnover rate (Waters and Swann 1998; Hardeland et al. 2000). Compared to *E. coli* Mug, which consists of the catalytic core only, TDG contains additional N- and C-terminal domains, providing non-specific DNA interaction and regulatory functions (Hardeland et al. 2002; Steinacher and Schär 2005; Baba et al. 2005). Knockout of *Tdg* in mouse is embryonic lethal, suggesting that, unlike other UDGs, it has a non-redundant essential function in embryonic development (Cortazar et al. 2011; Cortellino et al. 2011). TDG-deficient cells do not show increased sensitivity towards agents that would cause TDG-relevant DNA base lesions, nor do they show increased levels of spontaneous mutations (Cortazar et al. 2011), implicating functions beyond canonical DNA repair, which will be discussed below. An involvement of TDG in DNA repair becomes obvious, however, in the processing of the anti-cancer drug 5-FU when incorporated into the DNA. In this special case, the repair activity of TDG does not provide drug resistance as might be expected; excision of the base analog by TDG results in an accumulation of toxic AP-site intermediates and DNA strand breaks and, thus, mediates the DNA-directed cytotoxic effect of 5-FU (Kunz et al. 2009b).

SMUG-family glycosylases were initially identified as a uracil-excising activity in *Xenopus*, insect, and human cells (Haushalter et al. 1999). As such, it appears to serve as back-up for UNG in limiting uracil accumulation (U•A) in genomic DNA and in preventing C→T mutation following cytosine deamination (U•G) (Haushalter et al. 1999; An et al. 2005). While *Xenopus* SMUG has a preference for uracil in single-stranded DNA, hence the name “single-strand-specific monofunctional uracil DNA glycosylase”, the human homolog processes uracil also in double-stranded DNA (Table 1) (Haushalter et al. 1999; Kavli et al. 2002). SMUGs show only limited amino-acid sequence similarity with members of other UDG subfamilies and the conservation seems restricted to catalytic site residues, showing mosaic features of the UNG and MUG enzymes. Crystallographic analysis of SMUG1 identified a pyrimidine binding pocket topologically similar to other UDGs and implicated a water displacement/replacement mechanism to account for the enzyme's preference for uracil over thymine (Wibley et al. 2003). Like TDG, SMUG1 is active on 5-FU but, unlike TDG,

appears to protect cells from the cytotoxic effects of the drug as shown in siRNA knockdown experiments (An et al. 2007). Notably, a 5-hmU DNA glycosylase activity originally discovered in calf thymus was later identified as SMUG1 (Cannon-Carlson et al. 1989; Boorstein et al. 2001). So, like TDG, SMUG1 is capable of processing the deamination product of 5-hydroxymethylcytosine (5-hmC), a substrate that has recently gained attention in the context of active DNA demethylation.

Additional subfamilies of UDGs appear to have evolved in archaeal and bacterial organisms thriving under extreme environmental conditions such as high temperature, favoring hydrolytic deamination of cytosine and 5-mC. As these will not be further discussed here, the reader is referred to the excellent classification of the UDG superfamily originally published by Aravind and Koonin (2000).

The helix-hairpin-helix glycosylases

The second superfamily of DNA glycosylases, characterized by a shared helix-hairpin-helix (HhH) domain, comprises a diverse group of enzymes present in organisms throughout all kingdoms of life. Phylogenetic analysis in 94 genomes from bacteria, archaea, and eukaryotes identified six distinct families of HhH DNA glycosylases: Nth (homologs of the *E. coli* EndoIII protein), OggI (8-oxoG DNA glycosylase I), MutY/Mig (A/G-mismatch-specific adenine glycosylase), AlkA (alkyladenine-DNA glycosylase), MpgII (N-methylpurine-DNA glycosylase II), and OggII (8-oxoG DNA glycosylase II) (Denver et al. 2003). The Nth and MutY/Mig family glycosylases as well as some of the MpgII type proteins contain iron–sulfur [4Fe4S] clusters that are thought to play a structural role in DNA binding and substrate recognition (Cunningham et al. 1989; Kuo et al. 1992; Guan et al. 1998; Porello et al. 1998; Begley et al. 1999).

The founding member of the Nth family was originally discovered as an endonuclease activity (EndoIII) in *E. coli* (Radman 1976) but then turned out to be a DNA glycosylase with an associated AP-lyase activity. Nth proteins appear to be the most highly conserved subfamily within the HhH glycosylases (Denver et al. 2003). The mammalian homolog, NTHL1 (endonuclease III-like 1), acts on ring fragmented purines or oxidized pyrimidine residues like thymine glycol (Tg), formamidopyrimidine (FaPy), 5-hydroxycytosine (5-hC), and 5-hydroxyuracil (5-hU), preferentially when placed opposite guanine (Table 1) (Dizdaroglu et al. 1999; Eide et al. 2001). *Nth1* knockout mice show no overt abnormalities presumably because the loss of its repair function can be compensated for by NEIL glycosylases (see below) (Ocampo et al. 2002; Takao et al. 2002).

MutY was first identified in *E. coli* as an enzyme excising adenine from A•G mispairs (Au et al. 1988), while Mig.Mth was discovered in hyperthermophilic archaea by its ability to excise U or T mispaired with G (Horst and Fritz 1996). MutY/Mig homologs are widespread in bacterial genomes, but only about half of the eukaryotes and less than a third of the archaeal species analyzed have glycosylases of this family (Denver et al. 2003). The mammalian homolog of MutY, termed MYH, excises adenine opposite 8-oxoG, guanine, or cytosine (Table 1) (McGoldrick et al. 1995), contributing to a multimodal defense against the mutability of guanine oxidation (van Loon et al. 2010). Facilitating the replacement of A opposite 8-oxoG with a C, MYH produces the preferred substrate for the 8-oxoG directed DNA glycosylase OGG1. Hence, disruption of *Myh* in mice does not produce a mutator phenotype *per se* because it is masked by OGG1, which corrects the bulk of oxidized guanines before replicative DNA polymerases get a chance to misinsert adenine opposite the damaged base. Knocking out both *Myh* and *Ogg1*, however, results in a synergistic increase in G→T mutations (Russo et al. 2004). In humans, germline mutations in the *MYH* gene have been associated with a predisposition to colorectal cancer (Al-Tassan et al. 2002; Jones et al. 2002).

The OGG1 protein family is less well represented across the phylogeny. While present in most eukaryotic genomes, OGG1 encoding genes seem to be missing in bacteria and archaea (Denver et al. 2003). OGG1 was originally discovered in yeast and later also identified in mammals (Nash et al. 1996; Lu et al. 1997; van der Kemp et al. 1996; Radicella et al. 1997), where it provides the major activity for the removal of 8-oxoG opposite cytosine (Friedberg et al. 2006). It does, however, also excise other oxidized pyrimidines or ring-fragmented purines like formamidopyrimidine (FaPy) (Table 1) (Dherin et al. 1999; Karahalil et al. 1998). In *Saccharomyces cerevisiae*, inactivation of *OGG1* results in an accumulation of G→T transversion mutations (Thomas et al. 1997). *Ogg1* null mice are viable but exhibit a 2-fold increase in chromosomal 8-oxoG and moderately elevated spontaneous mutation frequencies (Klungland et al. 1999). Polymorphisms in the human *OGG1* gene impairing the 8-oxoG incision activity were found to be associated with non-small cell lung cancer (Janik et al. 2011) and an increased risk of childhood acute lymphoblastic leukemia (Stanczyk et al. 2011).

Although closely related to the Nth and MutY proteins, MBD4, also known as MED1, is special in two ways; it has a methyl-CpG binding domain (MBD) and is therefore also a member of the MBD protein family, and it functionally interacts with MLH1, a protein of the postreplicative mismatch repair system (Hendrich and Bird 1998; Bellacosa et al. 1999). Like the structurally unrelated TDG, MBD4 is a

mismatch-directed DNA glycosylase processing a wide range of G-mispaired base lesions, including thymine, uracil, 5-FU, and 3,N4-ethenocytosine (Table 1) (Petronzelli et al. 2000; Cortellino et al. 2003). Its methyl-CpG binding domain in addition to its activity on the deamination product of 5-meC has made MBD4 a prime candidate for an active DNA demethylase, an epigenetic function likely to be important in embryogenesis. However, *Mbd4* knockout mice show no developmental defects, but a mild increase in C→T mutation frequency and a predisposition to gastrointestinal cancer in APC-deficient tumor models (Millar et al. 2002; Wong et al. 2002), consistent with a role of MBD4 in repair of cytosine or 5-meC deamination damage.

E. coli AlkA, the founding member of the AlkA family of HhH glycosylases, acts on alkylated bases, e.g., 3-meA. While homologs are present in many bacterial and eukaryotic genomes (Denver et al. 2003), mammals appear to be devoid of this particular class of enzyme. Instead, they use a structurally unrelated enzyme, MPG, to eliminate specific forms of base alkylation damage.

The 3-methyl-purine glycosylase (MPG)

MPG, also known as AAG or MDG, originally identified in rat (O'Connor and Laval 1990) and later in human, is a DNA glycosylase excising a range of alkylated bases from DNA, including 3-meA, 7-meG, 3-methylguanine (3-meG) as well as ethylated bases in single- and double-stranded DNA (Table 1) (O'Connor 1993; Lee et al. 2009). MPGs form a structurally distinct class of glycosylases; they lack helix-hairpin-helix motifs nor do they have an alpha-beta fold structure characteristic of UDGs. Mice lacking MPG are viable and show a mild increase in the frequency of spontaneous mutation (Engelward et al. 1997; Hang et al. 1997), and they are more prone to develop azoxymethane-induced colon cancer than their wild-type counterparts (Wirtz et al. 2010). Reminiscent of TDG's role in the DNA-directed cytotoxic effect of 5-FU, MPG drives alkylation-induced retinal degeneration in mice by generating cytotoxic BER intermediates (Meira et al. 2009).

The endonuclease VIII-like glycosylases

Although their substrate spectrum overlaps with that of endonuclease III (Nth), the homologs of *E. coli* endonuclease VIII, encoded by the *nei* gene, are structurally related to the formamidopyrimidine-DNA glycosylase Fpg and form a separate family of DNA glycosylases. Nei was discovered in *E. coli* as a second activity next to Nth acting on thymine glycol (Tg) and urea (Melamede et al. 1994).

The mammalian counterparts are termed Nei-like (NEIL)1, 2, and 3, and share a conserved helix-two-turn-helix motif with the *E. coli* Fpg and EndoVIII proteins. The

preferred substrates of NEIL1 and NEIL2 are oxidized pyrimidines such as Tg, 5-hC, FaPyA, and FaPyG (Hazra et al. 2002; Morland et al. 2002; Rosenquist et al. 2003), but also 5-hydroxyuracil (5-hU) and 8-oxoG in DNA bubble structures (Table 1) (Dou et al. 2003). NEIL3, on the other hand, excises FaPy but is inactive on 8-oxoG (Liu et al. 2010). Mice with a targeted inactivation of the *Neil1* gene exhibit a phenotype reminiscent of the metabolic syndrome, as well as increased levels of DNA damage in mitochondrial DNA (Vartanian et al. 2006). *Neil3* knockout mice are viable and fertile, and the protein appears to be preferentially expressed in hematopoietic tissues (Torisu et al. 2005), implicating a possible function in hematopoiesis or the immune system.

Structure function aspects

DNA glycosylases evolved to counter the many different forms of chemical damage occurring to DNA bases. They are highly specialized enzymes with distinct structures and substrate specificities, but they all share a common principle of action. DNA glycosylases recognize their cognate substrates by rotating bases out of the DNA helix into a specifically fitting pocket that harbors the active site. Substrate selectivity is mostly achieved by steric exclusion from the binding pocket of bases that are not to be processed, and to some extent also by the catalytic efficiency of the active site configuration. If base fitting is successful, cleavage of the N-glycosidic bond will occur, resulting in the release of a free base and the generation of a base-less sugar, an AP-site in the DNA.

Detection and verification of base damage

Detecting a single damaged base in a vast excess of regular bases in the genome amounts to the proverbial task of searching a needle in a haystack, and this is not helped by the fact that the lesions addressed by BER do not usually cause notable distortions to the DNA helix. A human cell suffers about 10^4 base lesions per day, translating into roughly one lesion every 10 s within a genome of about 14 billion nucleotides that must be spotted and repaired. How DNA glycosylases manage to efficiently search for and recognize these lesions is not clear but biochemical and structural work has provided some insight into possible mechanisms.

Scanning the genome for damaged bases

Little is known about how DNA glycosylases find damaged bases in the genome. One idea, proposed on the basis of biochemical evidence and theoretical considerations, postu-

lates the association of the glycosylase with undamaged DNA by non-specific interactions, facilitating sliding along the DNA duplex for a certain distance and scanning the sequence for irregular bases (Berg et al. 1981). Considering the structural and functional diversity of DNA glycosylases, however, there are likely to be different translocation mechanisms, variably involving features of tracking, diffusion, and hopping on the DNA (Blainey et al. 2009; Steinacher and Schär 2005; Friedman and Stivers 2010). Recently, an appealing DNA scanning concept was proposed for DNA glycosylases harboring an [4Fe4S] cluster. The underlying observation was that *E. coli* MutY and Nth change the oxidation state of their iron-sulfur cluster from [4Fe4S]²⁺ to [4Fe4S]³⁺ upon contact with DNA, which stabilizes the interaction. Thus, if such DNA glycosylases bind in the vicinity of each other, they might act as electron donors and acceptors for each other, making use of the DNA for charge transfer. This may facilitate the dissociation of one glycosylase upon binding of another by reduction of its [4Fe4S] cluster. If the electron transfer involved is perturbed by a base lesion between the two glycosylases, however, both will stay bound to the DNA, increasing the likelihood of damage detection (Boal et al. 2009). This way, [4Fe4S] clusters may support the search for base damage without a need of scanning the entire DNA sequence.

Detection of DNA base damage by DNA glycosylases ultimately requires a full examination of the chemical surface of single bases. To minimize the effort, DNA glycosylases employ strategies of damage pre-selection. OGG1 and UNG, for instance, were proposed to pre-select substrates by establishing superficial base contacts through conserved residues in close proximity to the mouth of their catalytic pocket. This allows potential substrates to be identified without fully inserting every base into the active site pocket (Fig. 2) (Banerjee et al. 2005; Parker et al. 2007). Consistently, NMR studies on human UNG showed the glycosylase to undergo a conformational change upon DNA binding, allowing for oscillation between an open form, loosely interacting with the DNA in an unspecific manner, and a closed form engaged in base examination without fully rotating the base out of the helix (Fig. 2) (Friedman et al. 2009). Base pre-scanning is likely to be facilitated by DNA breathing, considering that many base lesions affect base pairing dynamics to some extent and are therefore likely to enhance local DNA melting.

Formation of a mature enzyme substrate complex

For final damage verification, the base needs to be flipped out of the DNA helix and accommodated in the active site cavity of the glycosylase. This increases the surface for molecular interactions, providing for a sensitive discrimi-

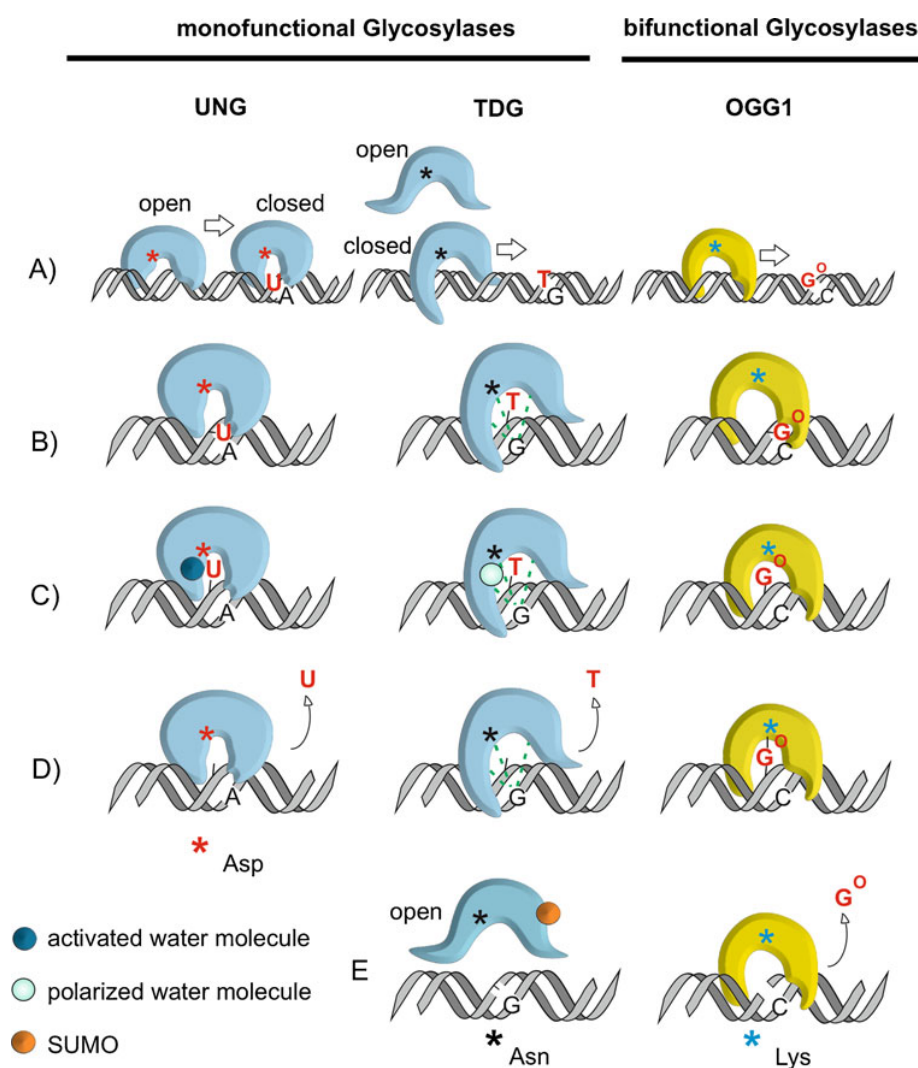


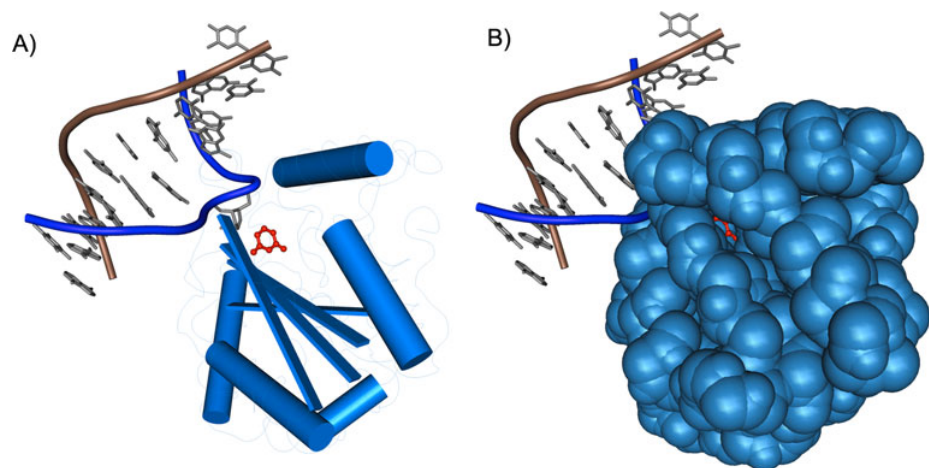
Fig. 2 Mechanism of base removal by monofunctional and bifunctional glycosylases. While all DNA glycosylases share a common principle of action, they differ in details of damage recognition, and excision. Among the monofunctional DNA glycosylases, UNG searches for uracil in DNA by oscillating between an open conformation, making unspecific contacts with the DNA, and a closed conformation for base-interrogation (a), involving conserved structures at the mouth of the catalytic site pocket (b). Pre-selected bases are flipped out of the DNA duplex and inserted into the catalytic pocket, where specific hydrogen bonds align the base for nucleophilic attack by an activated water molecule strategically positioned by a conserved aspartic acid residue (*red asterisk*, c). TDG differs from UNG by its non-specific interaction with DNA through a clamp formed by the N-terminal domain (a), its larger catalytic pocket that accommodates a broader range of substrates, and its ability to involve

the opposing base in lesion recognition (b). The catalytic residue in TDG is an asparagine (*black asterisk*) that positions, but does not activate, a water molecule which can then act as a weak nucleophile (c). TDG stays firmly bound to the abasic site upon base release (d) until SUMOylation induces dissociation by neutralizing the N-terminal DNA binding activity (e). Similar to UNG, the bifunctional OGG1 pre-selects bases for flipping by interactions with conserved amino-acid residues at the mouth of the catalytic cavity (a, b), and similar to TDG, it also takes the opposing base into account for substrate selection (c). Bifunctional glycosylases use a conserved amino-acid residue (*blue asterisk*) for nucleophilic attack, which results in a covalent intermediate between the glycosylase and the DNA substrate (d). Resolution of this intermediate produces a DNA nick that is further processed by APE1 (e)

nation of even minor base alterations and, ultimately, catalysis of the base release. One of the best-studied glycosylases with regard to damage recognition is UNG. On its surface, UNG forms a positively charged groove that accommodates the minor groove of the DNA and contains

the active site cavity. Upon encountering a uracil in DNA, UNG uses a hydrophobic side chain of a conserved leucine to push the uracil out of the base stack and insert it into the catalytic pocket (Fig. 3). The same residue stabilizes the double helix by occupying the vacated space (Mol et al.

Fig. 3 The base-flipping intermediate captured in a crystal structure of substrate bound human UNG. The UNG double mutant L272R/D145N, stabilizing the glycosylase–substrate complex, was co-crystallized with an oligonucleotide bearing a U•G mismatch. Uracil (*red*) is flipped into an extrahelical position (**a**) and inserted deeply into the tight fitting active site pocket of UNG (**b**), where it is positioned by specific molecular interactions for the nucleophilic attack on the N-glycosidic bond (Slupphaug et al. 1996)



1995; Savva et al. 1995). At the same time, UNG pinches and compresses the double-stranded DNA backbone, thereby inducing DNA bending by about 45° at the site of the damage (Fig. 1) (Parikh et al. 1998). Selectivity for uracil is mediated by several structural features: purine bases are sterically excluded from the active site by the overall narrow geometry of the binding pocket, while the entry of thymine and other 5-methylated pyrimidines is blocked by a side chain of a conserved tyrosine residue. Accordingly, UNG can be converted into a thymine-processing enzyme by an amino acid substitution that eliminates this sterical block (Kavli et al. 1996). Cytosine can enter the active site but, due to unfavorable hydrogen bonding at the bottom of the cavity, cannot be positioned correctly for catalysis (Kavli et al. 1996; Slupphaug et al. 1996).

While UNG is highly specialized to the excision of DNA uracil, TDG is a good example of a broad spectrum UDG showing mismatch dependency. Unlike UNG, the bacterial and human TDG orthologs have rather spacious pyrimidine-binding pockets, accommodating a large variety of base derivatives (Barrett et al. 1998, 1999; Baba et al. 2005; Maiti et al. 2008), although the preferred substrates are G-mismatched deaminated pyrimidines (Waters and Swann 1998; Hardeland et al. 2003). Because TDG acts on thymine, it needs to be able to discriminate between a regular thymine in DNA (A•T) and one that resulted from deamination of 5-meC (G•T). Substrate selection thus has to take into account not only the base structure itself but also the opposite base. Structural and biochemical studies of *E. coli* Mug and human TDG (Barrett et al. 1998; Maiti et al. 2008, 2009) have shed light on how this might be achieved. In contact with DNA, TDG undergoes a conformational change in its N-terminal domain, forming a clamp-like structure that permits TDG to track along the DNA (Fig. 2) (Steinacher and Schär 2005). Upon encountering a G-mismatched substrate (G•T), the substrate base is pushed out of the DNA helix by an insertion loop wedging into the DNA helix. This same wedge

stabilizes the base stack and forms specific hydrogen bonds with the widowed guanine, mimicking Watson–Crick base pairing (Barrett et al. 1998, 1999; Maiti et al. 2008). These interactions then cooperate with the non-specific DNA binding activity of the N terminus to stabilize the TDG–substrate complex for efficient base excision (Hardeland et al. 2000; Steinacher and Schär 2005).

Another DNA glycosylase well studied with regard to damage recognition is OGG1. Following successful pre-selection of a potential 8-oxoG substrate (Fig. 2) (Banerjee et al. 2005), the oxidized base is flipped into the active site cavity of OGG1, inducing DNA bending of about $\sim 70^\circ$ due to a tyrosine residue that wedges between the opposing C and its 5' neighbor. Unstacking of the widowed C facilitates enzyme–DNA contacts that maintain opposite base selectivity. The void generated by 8-oxoG rotation is occupied by a conserved asparagine residue, which contributes to hydrogen bonding with the opposing cytosine. While A, C, and T are excluded from the active site pocket by several strategically positioned amino-acids, a conserved glycine is the only residue discriminating between G and 8-oxoG (Bruner et al. 2000).

Catalysis of base removal

The catalytic mechanism subdivides DNA glycosylases into monofunctional and bifunctional enzymes (Table 1). Monofunctional glycosylases perform base excision only, using an activated water molecule for nucleophilic attack on the N-glycosidic bond, while bifunctional glycosylases use an amino group of a lysine side chain for the same purpose, forming a Schiff-base intermediate, and subsequently cleave the DNA backbone 3' to the lesion.

Monofunctional DNA glycosylases

To illustrate the monofunctional mode of action, we will focus on UNG and TDG as two well-studied UDGs

showing contrasting catalytic features. Upon recognition, uracil is tightly fitted into the active site pocket of UNG (Fig. 3). Interactions between the uracil and amino-acid residues at the bottom of the pocket position the N-glycosidic bond for hydrolysis (Mol et al. 1995; Savva et al. 1995; Slupphaug et al. 1996). Structure and mutational analyses have established a catalytic mechanism involving the polarization of the N-glycosidic bond by a conserved histidine to make it susceptible for nucleophilic attack, and the positioning and deprotonation of a water molecule by a conserved catalytic aspartate, which then attacks the C1 of the deoxyribose (Fig. 2) (Mol et al. 1995; Slupphaug et al. 1996). N-glycosidic bond cleavage is completed by addition of the water nucleophile to uracil, resulting in a free base and an AP-site.

The mismatch-specific uracil glycosylases, e.g., Mug and TDG, interact less specifically with the substrate base within the catalytic pocket and use a less potent mechanism of catalysis, as first revealed in the crystal structure of *E. coli* Mug (Barrett et al. 1998, 1999). In place of the catalytic aspartate in UNG, Mug/TDG enzymes have an asparagine. This asparagine positions a water molecule but, unlike the aspartate in UNG, is unable to protonate it for an efficient nucleophilic attack on the N-glycosidic bond (Fig. 2). Also, an appropriate residue for polarization of the N-glycosidic bond appears to be missing in Mug/TDG, altogether explaining the comparably low catalytic efficiency of the MUG enzymes (Hardeland et al. 2000; Maiti et al. 2009). These differences illustrate that the mode of catalysis can vary considerably even within one DNA glycosylase superfamily, most likely reflecting the requirement to fine tune substrate spectrum and catalytic efficiency in the evolution of subfamilies with distinct biological functions.

Bifunctional DNA glycosylases

DNA glycosylases that use an amino group of a lysine side chain as a nucleophile for base cleavage form a covalent Schiff-base intermediate with the substrate. The resolution of this reaction intermediate incises the DNA 3' to the product AP-site, generating a strand break with 3'phosphate and 5'OH ends. Thus, these enzymes couple base excision with an AP-lyase step, as best illustrated by the bifunctional mechanism proposed for OGG1 (Sun et al. 1995; Nash et al. 1997). Once an 8-oxoG is stably fitted in the active site cavity of OGG1, the side chain amino group of a suitably positioned catalytic lysine is activated as a nucleophile to attack the C1 of the deoxyribose. The resulting rearrangement to a covalent Schiff-base intermediate releases the damaged base from the DNA, which is then held in position to further participate in the catalysis of the β -lyase reaction, cleaving the DNA strand at the 3'phosphate. Notably,

all steps of the OGG1 catalyzed base release involve the transfer of protons and this is promoted by the excised base itself (Bruner et al. 2000; Fromme et al. 2003). An interesting variation on the bifunctional mode of action is displayed by the mammalian NEIL proteins and their *E. coli* counterpart Nei. These enzymes couple base excision to beta and delta elimination, incising the DNA strand both 3' and 5' to the AP-site (Takao et al. 2002). Moreover, unlike OGG1, *E. coli* Nei ejects the excised base from the catalytic site immediately, i.e., is capable to perform the AP-lyase reaction without contribution of the damaged base (Zharkov et al. 2002).

AP-site dissociation and turnover of glycosylases

Upon base release, DNA glycosylases tend to stay bound to the product of their action, the AP-site. In fact, many glycosylases display a higher affinity to AP-sites than to their actual substrate base (Parikh et al. 1998; Waters et al. 1999; Hardeland et al. 2000; Hill et al. 2001; Pope et al. 2002). The release of the glycosylase from the AP-site is thus rate limiting in the BER process (Fig. 1). Since AP-sites are chemically unstable and lack base coding potential, the binding of the glycosylase might serve to protect cells against their cytotoxic and mutagenic effects. This considered, it appears reasonable that the release of the AP-site is coordinated with the recruitment and assembly of the downstream acting BER factors. Consistently, the AP-endonuclease APE1 was reported to stimulate the turnover of several DNA glycosylases including UNG2, TDG, and OGG1 (Parikh et al. 1998; Waters et al. 1999; Hill et al. 2001), and a similar effect was observed for the XPC protein on TDG and SMUG1 (Shimizu et al. 2003, 2010). It remains to be clarified, though, to what extent these stimulatory effects reflect active processes or simply a passive competition for a common DNA substrate.

Another, perhaps more sophisticated, way to regulate AP-site interaction of the glycosylase is by posttranslational modification. In this direction, UNG2 was reported to be cell cycle specifically phosphorylated at serine 23, which markedly increases its association with replicating chromatin but also its turnover rate, apparently to facilitate efficient correction of misincorporated uracil during ongoing DNA replication (Hagen et al. 2008). In the case of TDG, which binds AP-sites very rigidly, dissociation is regulated by modification of a C-terminal lysine residue with small ubiquitin-like modifiers (SUMOs). This induces a conformational change that neutralizes the non-specific DNA-binding capacity of its N-terminal domain, thereby facilitating AP-site dissociation (Fig. 2) (Hardeland et al. 2002; Steinacher and Schär 2005). SUMO modification might be triggered by the presence of downstream acting BER factors, providing for a controlled handover of the AP-site intermediate in the

repair process (R. Steinacher and P. Schär, unpublished data).

It appears to be a general feature of DNA glycosylases to bend DNA upon establishing a mature enzyme–substrate complex; UNG for example induces a bend of $\sim 45^\circ$ (Parikh et al. 1998), OGG1 of even $\sim 70^\circ$ (Bruner et al. 2000). On the basis of these and similar observations with other BER factors, it was proposed that the DNA bending might serve as a structural determinant to orchestrate the handover from one step in the repair process to the next (Fig. 1) (Parikh et al. 1999; Wilson and Kunkel 2000).

Functions of DNA glycosylases beyond DNA repair

Although DNA glycosylases are optimally suited for the repair of damaged DNA bases, their structural and biochemical properties would support a much wider spectrum of genetic functions. The ability to recognize and excise chemically modified bases can be used to edit the DNA at specifically marked sites. The UDG superfamily in particular appears to comprise enzymes with specialized functions, e.g., in innate immunity and antibody diversification, as well as in the regulation of gene expression and epigenetic maintenance.

DNA glycosylases in immunity

UDGs have been functionally associated with mechanisms providing innate immunity against viral infection as well as antibody diversity in the adaptive immune system. All these processes are triggered by enzymatic deamination of cytosine by members of the apolipoprotein B mRNA editing catalytic polypeptide (APOBEC) family of proteins, including the activation-induced cytidine deaminase (AID) (Conticello et al. 2005).

DNA glycosylases providing innate immunity

Proteins of the human APOBEC3 subfamily were shown to inhibit replication of a variety of retroviruses, including human immunodeficiency viruses (HIV) (reviewed in Holmes et al. 2007). APOBEC3G was discovered by its ability to restrict replication of a mutant HIV-1 lacking the viral infectivity factor (Vif) (Sheehy et al. 2003). Cells infected with such a virus package APOBEC3G into HIV-1 virions. When these infect new cells, APOBEC3G will deaminate multiple cytosines in the viral cDNA during reverse transcription, which inactivates the provirus (reviewed in Holmes et al. 2007). While APOBEC3G appears to intervene with the viral life cycle at several steps, the antiviral activity mediated by uracilation of the viral genome is coupled to the action of UNG2 and APE1. UNG2 is thought to introduce AP-sites into the deaminated

viral cDNA, triggering the cleavage and thus degradation by APE1 (Harris et al. 2003; Yang et al. 2007). Consistently, inhibition and/or downregulation of either UNG2 or APE1 was shown to decrease the antiviral potency of APOBEC3G (Yang et al. 2007). Given that HIV-1 evolved Vpr, a small protein specifically targeting UNG2 and SMUG1 for degradation by the ubiquitin–proteasome system, supports a critical role of uracil base excision in antiviral defense (Schrofelbauer et al. 2005, 2007).

DNA glycosylases providing adaptive immunity

The cooperation of cytidine deaminase and UDG activities is also central to the genetic transactions associated with antibody diversification in the adaptive immune system, i.e., somatic hypermutation (SHM) and class switch recombination (CSR) (Fig. 4) (Muramatsu et al. 2000; Rada et al. 2002).

SHM introduces point mutations in the light chain variable (V) region of immunoglobulin loci. Mutagenesis is initiated by AID, which is targeted to these loci to deaminate multiple cytosines in single-stranded DNA occurring during transcription (reviewed in Pavri and Nussenzweig 2011). Mutations can arise from these U•G mismatches in several ways (Fig. 4): (1) in the absence of uracil excision, replication across the U•G mismatch would generate a C→T mutation in one of the daughter strands; (2) uracil excision by a UDG and subsequent replication across the resulting AP-site can potentially give rise to any type of base substitution at the site of deamination, although replicative DNA polymerases preferentially insert A opposite an AP-site; (3) long-patch BER, initiated by a UDG, or MMR coupled to error-prone DNA synthesis would cause mutations in proximity to the deaminated cytosine, allowing for mutations to occur also at A•T base pairs.

An involvement of UNG2 in antibody diversification was first implicated by a general perturbation of both SHM and CSR in an UNG2 inhibited chicken B-cell line as well as in UNG-deficient mice (Di Noia and Neuberger 2002; Rada et al. 2002). While the loss of UNG mainly affected SHM at G•C base pairs, inactivation of the mismatch repair system (MMR) was found to diminish hypermutation at A•T pairs (Rada et al. 1998; Wiesendanger et al. 2000). This suggested that error prone MMR contributes to mutagenesis at sites away from the deaminated cytosine (Wilson et al. 2005). Later work then implicated that MMR can indeed operate at an AID induced G•U mismatch if assisted by UNG2, providing a nick at a nearby G•U for initiation of strand excision (Frieder et al. 2009; Schanz et al. 2009).

AID, UNG2, and APE1 were also shown to play a crucial role in the initiation of CSR, a specialized recombination process switching the Ig isotype of an

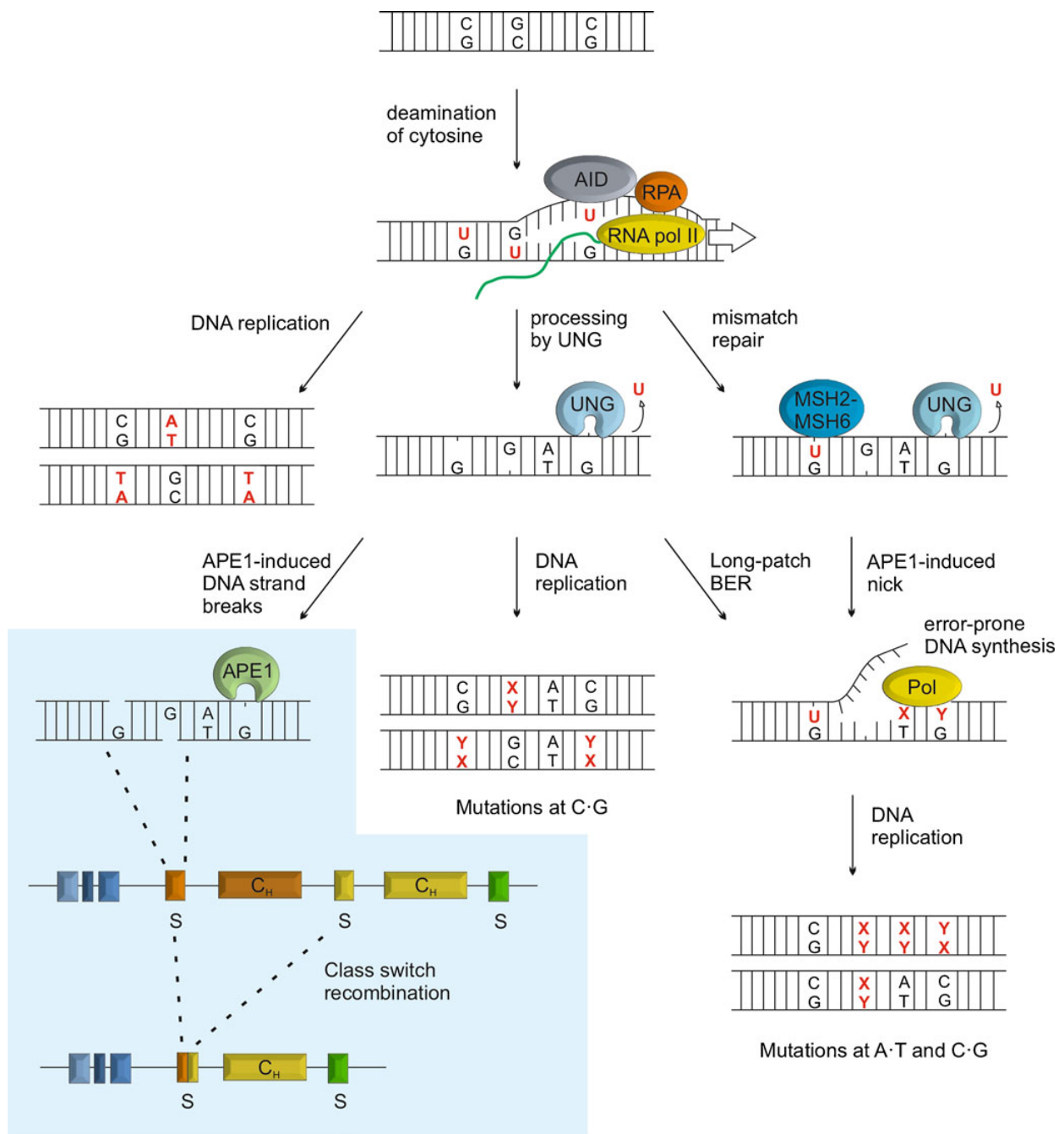


Fig. 4 Uracil DNA glycosylase in antibody diversification. UNG plays a central role in somatic hypermutation (SHM) and class switch recombination (CSR). UNG2 is targeted to immunoglobulin loci in activated B-cells by AID converting cytosine to uracil in single-stranded DNA during transcription. While replication across unprocessed uracil itself will generate C→T mutations, uracil excision by UNG provides for a wider range of mutations both at the C•G and nearby A•T base pairs; (1) by generating non-instructive AP-sites

which, upon DNA replication give rise to transition or transversion mutations, (2) by initiating long-patch BER or, (3) following cleavage by APE1, by providing a DNA strand nick for activation of MMR. Error-prone synthesis associated with long-patch BER and MMR would then produce mutations at A•T base pairs. Similarly, DNA strand breaks occurring through BER of deaminated cytosines in switch regions of immunoglobulin heavy chain loci may induce CSR (*blue panel*)

antibody without affecting its antigen specificity. CSR rearranges the antibody constant region of heavy chain genes (C_H) by strand breakage in and joining of two selected switch (S) regions located upstream of every C_H -coding segment. The DNA strand breaks required for initiation of CSR were shown to arise through cytosine deamination, subsequent uracil excision, and AP-site cleavage by AID, UNG2, and an AP-endonuclease, respectively (Fig. 4) (Imai et al. 2003). Consistently, murine B-cells lacking UNG2 show a severe reduction of CSR and in humans, recessive mutations in the *UNG* gene have been associated with the hyper-IgM syndrome caused by a deficiency in CSR (Rada et al. 2002; Imai et al. 2003). It is also reported, however, that the catalytic activity of UNG2 is dispensable for efficient CSR, while an N-terminal sequence motif appears to be important, suggesting that UNG2 is not responsible for the DNA cleavage step of CSR (Begum et al. 2007, 2009). It is possible that UNG2 simply marks AID-induced G•U mismatches for further processing by other factors, whereby the N terminus is required to mediate specific protein–protein interactions.

Residual CSR and SHM in UNG2-deficient mice hint at a possible contribution of other UDGs. Indeed, overexpression of SMUG1 in MSH2 UNG2 double-deficient cells could partially restore SHM and CSR. A biological role of SMUG1 in antibody diversification, however, is questionable as it is downregulated upon B-cell activation (Di Noia et al. 2006). Likewise, the U•G mismatch-directed glycosylase MBD4 is unlikely to play a major role in either SHM or CSR as a knockout of the gene in mouse showed no effect on either processes (Bardwell et al. 2003). The situation is less clear for TDG. While the impact of a TDG-deficiency on antibody diversification remains to be investigated, its upregulation in activated mouse B-cells hints at a specific function in B-cell maturation (Christophe Kunz and Primo Schär, unpublished data). In this context, TDG might be simply required to prevent mutations at non-Ig genes arising from mistargeted AID. It might, however, also directly contribute to SHM and CSR. Considering its tight interaction with AP-sites, TDG would be optimally suited to delay processing of these repair intermediates, thereby favoring error-prone translesion synthesis and recombination repair.

DNA glycosylases in DNA methylation control

Distinct patterns of DNA methylation and histone modifications are established during cell lineage restriction to determine and maintain cell-type-specific gene expression programs. In mammals, DNA methylation occurs mainly in the form of 5-meC in CpG dinucleotide sequences and is controlled by a methylation machinery consisting of both methylating and demethylating components. While the *de*

novo establishment and the maintenance of CpG methylation can be rationalized by the biochemical features of the DNA methyltransferase (DNMT) activities involved (Hermann et al. 2004), the reverse process of demethylation has remained elusive. In principle, however, DNA demethylation can be achieved through active and/or passive mechanisms. Whereas passive demethylation occurs upon DNA replication with downregulation or inhibition of the maintenance DNA methyltransferase, active demethylation describes the enzymatic removal of 5-meC and replacement with C. Active mechanisms have been implicated in several important biological processes, including the demethylation of the paternal pronucleus in the murine zygote (Mayer et al. 2000; Oswald et al. 2000), the genome-wide methylation erasure and reset of parental imprinting during gametogenesis (Monk et al. 1987; Kafri et al. 1992), and the reprogramming of methylation patterns observed after transfer of somatic cell nuclei to enucleated oocytes (Dean et al. 2001; Simonsson and Gurdon 2004). Active demethylation has also been reported to be targeted to select gene regulatory regions during neurogenesis, memory formation, and immune response (Bruniquel and Schwartz 2003; Miller and Sweatt 2007; Ma et al. 2009).

Mechanistically, the conversion of 5-meC to C in DNA can occur in different ways: (1) by a direct removal of the methyl group, (2) the replacement of 5-meC with C by excision repair, or (3) the deamination or oxidation of 5-meC followed by replacement of the nucleotide by BER. Consistent with an excision repair scenario, an increasing number of observations point at an engagement of DNA glycosylases in active demethylation. In plants, the removal of 5-meC by at least four bifunctional DNA glycosylases is well established; ROS1, DEMETER (DME), and the DME-like (DML) 2 and 3 all process 5-meC in CpG and non-CpG sequence contexts (reviewed in Zhu 2009). Mutations in these glycosylases affect cytosine methylation at specific loci but not in the overall genome, suggesting that they act in a targeted rather than a global manner (Penterman et al. 2007; Lister et al. 2008). In vertebrates, the situation is less clear; MBD4 and TDG have been implicated in DNA demethylation but they do not seem to be potent enough to excise 5-meC directly.

Concepts of DNA glycosylase-mediated demethylation in vertebrates

First evidence for an involvement of DNA glycosylases in active demethylation in vertebrates came with the discovery of a 5-meC DNA glycosylase activity in extracts of chicken embryos (Jost et al. 1995). The responsible enzyme later revealed itself as a homolog of the human TDG (Zhu et al. 2000b). 5-meC DNA glycosylase activity was then also reported for the human TDG and MBD4 proteins by the

same group (Zhu et al. 2000a, 2001). Yet, as several attempts to reproduce these findings were unsuccessful, the idea of TDG and MBD4 acting as demethylating glycosylases did not achieve broad acceptance. Recent observations, however, indicate that the difficulty in reproduction might lie in the necessity of unknown co-factors that either boost the catalytic efficiency of these glycosylases or convert 5-meC into a more favorable substrate.

MBD4 was recently reported to control CpG methylation in the context of parathyroid (PTH) hormone-induced gene activation. This was shown for the *CYP27B1* promoter, which undergoes active demethylation upon hormone stimulation (Kim et al. 2009). Both promoter activation and cytosine demethylation coincided with and depended on the physical association of MBD4 and downstream BER factors, but not of the functionally related TDG. Remarkably, this study further showed that phosphorylation of MBD4 by

PKC may potentiate its activity to process 5-meC, suggesting that posttranslational modification might be required to unleash a potentially harmful but dormant 5-meC glycosylase activity in certain DNA glycosylases for targeted demethylation under specific conditions (Fig. 5a). This might apply as well to TDG, given its propensity to posttranslational modification by SUMOylation (Hardeland et al. 2002), ubiquitylation (Hardeland et al. 2007), phosphorylation (Um et al. 1998), and acetylation (Tini et al. 2002). It will thus be necessary to revisit TDG's activity as a 5-meC glycosylase under conditions that support the formation of these posttranslational modifications.

Other lines of investigation support demethylation scenarios that involve the conversion of 5-meC to more favorable substrates for DNA glycosylases. Deamination of 5-meC by a cytidine deaminase is one possibility. This would generate a G•T mismatch that can be acted on by

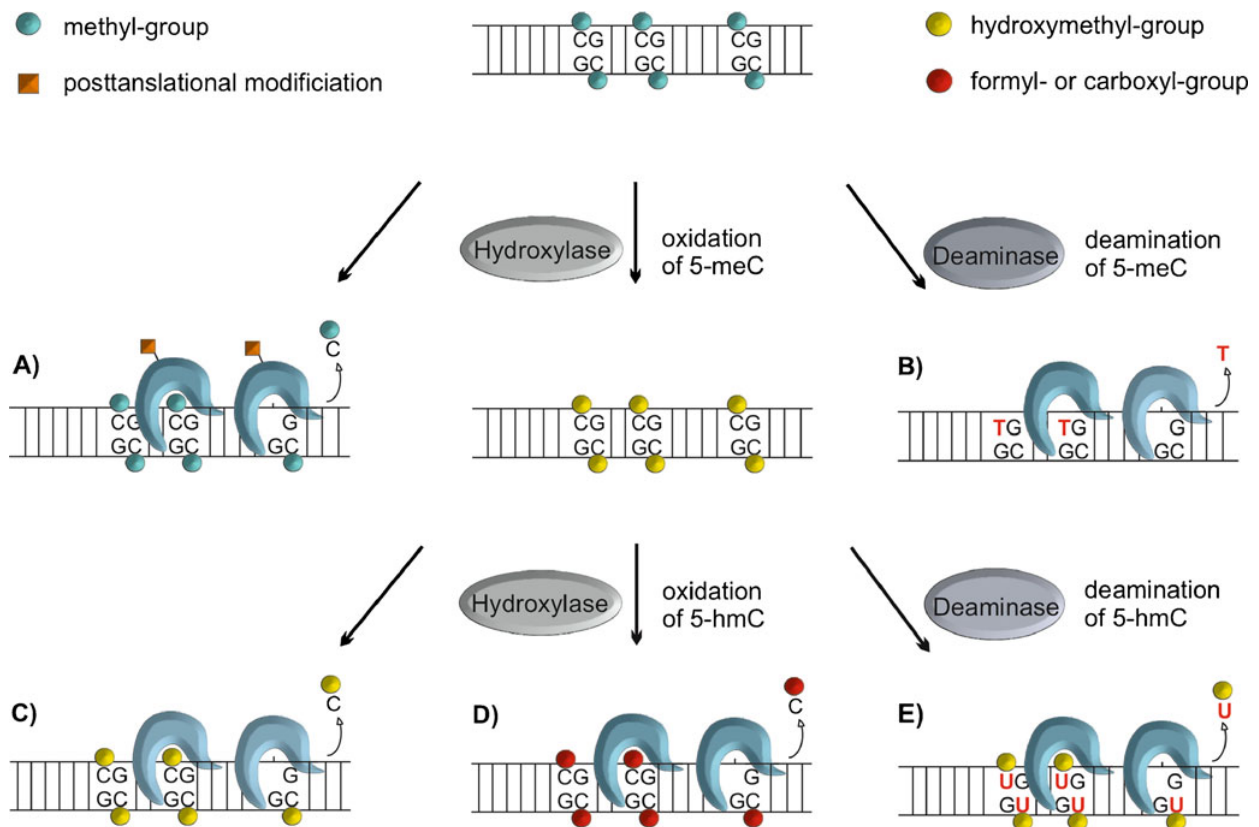


Fig. 5 Possible pathways of active DNA demethylation involving BER. Enzymatic removal of 5-meC has been suggested to be accomplished by different DNA glycosylase mediated mechanisms. Direct excision of 5-meC by mammalian DNA glycosylases has been tested with contradicting results, suggesting the possibility that posttranslational modification might induce a shift in the substrate spectrum, allowing for the removal of 5-meC. (a) Another possibility is the enzymatic conversion of 5-meC to substrates more favorable to

DNA glycosylases, such as deamination to thymine (b), hydroxylation to 5-hydroxymethylcytosine (c) and/or further oxidation or deamination of 5-hmC to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (d) or 5-hydroxymethyluracil (e), respectively. All these processes could be accomplished by a cooperation of 5-meC deaminases like AID, hydroxylases like the TET proteins, and DNA glycosylases like TDG (G•T, G•5hmU, G•fC, G•caC), MBD4 (G•T, ?) and possibly SMUG1 (G•T, G•5hmU, ?)

MBD4 or TDG (Fig. 5b). Along these lines, it was shown that the estrogen responsive pS2/TFF1 promoter undergoes cyclic DNA methylation and demethylation during transcriptional activation, involving 5-meC deamination by the methyltransferase Dnmt3a/b itself and subsequent TDG-dependent BER (Metivier et al. 2008; Kangaspeska et al. 2008). The demethylation complex also includes the RNA helicase p68, implicating an RNA component in either targeting and/or stabilization of the complex, consistent with previously reported observations on the chicken 5-methylcytosine DNA glycosylase (Jost et al. 1997; Schwarz et al. 2000). Additional support for a deamination-coupled demethylation pathway came from a study in zebrafish embryos, implicating AID in the deamination of 5-meC and MBD4 in the excision of the resulting G•T mismatch, as well as Gadd45 as an auxiliary factor (Rai et al. 2008). The same concept was adopted more recently to explain TDG-mediated 5-meC demethylation as it may occur during somatic differentiation of the developing mouse embryo (Cortellino et al. 2011). This, however, was mainly inferred from the co-immunoprecipitation of overexpressed AID and Gadd45 with TDG and therefore needs to be corroborated by more direct functional evidence. Finally, an AID-dependent and, thus, deamination-mediated mechanism has been associated with global DNA demethylation occurring in primordial germ cells as well as during the reprogramming of somatic cell nuclei towards pluripotency (Bhutani et al. 2010; Popp et al. 2010). Obviously, genome-wide 5-meC deamination would generate massive amounts of G•T mismatches, and all of these would have to be repaired by the G•T-directed DNA glycosylases MBD4 or TDG. A strong prediction of a deamination-based demethylation model is therefore that a failure of G•T repair would result in genome-wide C→T mutagenesis. This has not been observed so far in MBD4 and/or TDG-deficient cells (Cortazar et al. 2011) and will have to be tested more carefully.

The recent discovery of DNA dioxygenases acting on 5-meC in DNA introduced another conceptual framework for active demethylation. The principle of oxidative demethylation of DNA bases was first described for the bacterial DNA repair protein AlkB. AlkB belongs to a large superfamily of Fe(II)/2-oxoglutarate (2-OG)-dependent hydroxylases and catalyzes the hydroxylation of N-methylated bases like 1-methyladenine (1-meA) and 3-methylcytosine (3-meC), ultimately resulting in demethylation upon release of the hydroxymethyl moiety as formaldehyde (Falnes et al. 2002; Trewick et al. 2002). The superfamily of Fe(II)/2-OG hydroxylases also contains the kinetoplastid base J binding proteins (JBP). Base J stands for β -D-glucosyl(hydroxymethyl)uracil, an abundant base in the genome of kinetoplastida, synthesized through a 5-hmU intermediate generated by enzymatic hydroxylation of thymine by JBP1/2 (reviewed in Borst and Sabatini 2008).

Computational analyses identified the mammalian oncogenic TET proteins as close relatives of the JBPs (Iyer et al. 2009). The ultimate finding that TETs comprise a catalytic domain capable of catalyzing the oxidation of 5-meC to 5-hmC uncovered a functional link between these proteins and DNA methylation, possibly demethylation (Tahiliani et al. 2009; Ito et al. 2010). Indeed, 5-hmC was suggested to trigger passive demethylation through inhibition of the maintenance methyltransferase DNMT1 (Valinluck and Sowers 2007). Besides that, 5-hmC might represent an intermediate of active DNA demethylation through stepwise oxidation of 5-meC coupled to either excision repair or decarboxylation (Fig. 5d).

Direct excision of 5-hmC by a DNA glycosylase would seem a plausible scenario (Fig. 5c). A 5-hmC DNA glycosylase was reported to be active in calf thymus extracts (Cannon et al. 1988). The responsible protein, however, has never been purified, nor has 5-hmC glycosylases activity been associated with any known mammalian DNA glycosylase. Given their activity on 5-hmU, TDG and SMUG1 would seem good candidates for 5-hmC processing, but recent evidence shows that at least TDG fails to do so at an appreciable rate (He et al. 2011; Maiti and Drohat 2011). Thus, while direct 5-hmC excision by DNA glycosylases may occur and contribute to DNA demethylation, the underlying enzymatic pathway remains to be clarified.

Another possible route of demethylation would be the further conversion of 5-hmC to an intermediate for DNA glycosylase mediated excision. The deamination of 5-hmC by a specific deaminase (e.g., AID), for example, would give rise to 5-hmU mismatched with G (Rusmintrati and Sowers 2000), which is a substrate for SMUG1 and TDG (Fig. 5e) (Boorstein et al. 2001; Hardeland et al. 2003; Cortellino et al. 2011). As discussed above for 5-meC deamination, however, the concept of demethylation through a pro-mutagenic 5-hmU intermediate is debatable as this would require an immensely efficient and accurate coupling of the deamination and repair processes, particularly in densely methylated DNA sequences, if genome integrity is to be maintained.

Less problematic in this regard seems a more recently implicated mode of TET-mediated demethylation. The key discovery was that TET proteins do not only generate 5-hmC but can oxidize this intermediate further to 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC) (Ito et al. 2011), which are good substrates for excision by TDG (He et al. 2011; Maiti and Drohat 2011) (Fig. 5d). Given this, a DNA glycosylase-mediated conversion of 5-meC to C may thus occur without a need to deaminate, i.e., to generate a mutagenic intermediate. However, such a pathway would still trigger massive DNA incision activity in the context of genome-wide active demethylation, which seems a genetically risky and energetically wasteful way to

erase an epigenetic mark. A more straightforward way to eliminate 5-caC under these conditions would be to couple the 5-meC oxidation cascade with a decarboxylation step to generate C and CO₂ as final products. A 5-caC decarboxylase is thus an activity to watch out for.

G•T-directed DNA glycosylases and epigenetic stability

Active demethylation of 5-meC in mammalian cells occurs under different circumstances for different purposes, such as genome-wide epigenetic reprogramming, activation of tissue-specific genes during embryogenesis, and the maintenance of active and bivalent chromatin states during and after cell-lineage commitment. These circumstances generate specific demands for a demethylation mechanism (e.g., targeting, catalysis, processivity, efficiency) which are most likely addressed by distinct pathways.

Any of the above considered routes of DNA repair-mediated active demethylation of 5-meC requires an enzyme capable of recognizing and excising a cytosine derivative (5-meC, 5-hmC, 5-caC, T, 5-hmU) in a base (mis)pairing configuration with guanine. On the basis of their substrate spectra, the two DNA glycosylases MBD4 and TDG appear to be most suitable for this purpose, and both have been implicated in one way or another in DNA demethylation as discussed. The biological functions associated with these activities, however, still need to be clarified. Considering, for instance, the reported involvement of MBD4 in hormone-induced promoter demethylation, and the requirement of such demethylation for derepression of select genes during embryogenesis, it is surprising that *Mbd4* knockout mice develop normally and display no apparent epigenetic abnormality (Millar et al. 2002; Wong et al. 2002). So, either promoter demethylation during embryogenesis is not developmentally important, or MBD4 is not or only redundantly involved in these processes. The phenotype of *Tdg* knockout mice, however, has provided more direct evidence for an epigenetic function.

Disruption of TDG in mouse causes embryonic lethality (Cortazar et al. 2011; Cortellino et al. 2011), most likely because TDG-deficient cells fail to properly establish and/or maintain cell-type-specific gene expression programs during cell lineage commitment (Cortazar et al. 2011). This phenotype coincides with the occurrence of aberrant chromatin modifications at promoters of misregulated genes: a loss of active histone marks (H3K4me²), a gain of repressive histone marks (H3K9me³, H3K27me³) and, *nota bene*, a gradual accumulation of CpG methylation (Cortazar et al. 2011). Together with evidence for a TDG-dependent engagement of BER at affected gene promoters (Cortazar et al. 2011; Cortellino et al. 2011), this suggests that TDG protects CpG-rich promoters from aberrant

hypermethylation by active demethylation of erroneously methylated cytosines.

TDG may also contribute to active demethylation of originally methylated sequences. Tissues of TDG deficient embryos showed hypermethylation at the *Alb1* enhancer and the *Tat* glucocorticoid-responsive unit, both undergoing demethylation in the process of tissue specific gene activation (Cortellino et al. 2011). These results can, however, be interpreted in two ways; while the hypermethylation measured in the absence of TDG can indeed be explained by inefficient active CpG demethylation during tissue differentiation, it can equally well be accounted for by a lack of TDG-dependent maintenance of the unmethylated state following successful active demethylation through a different pathway. Hence, while the concept of TDG-mediated active demethylation in the maintenance of hypomethylated states at CpG-rich gene promoters is well supported, its potential contribution to demethylation of methylated sequences requires further validation.

The functions of MBD4 and TDG in mediating DNA demethylation are clearly distinct but there might be a partial overlap; TDG might compensate for the loss of MBD4 in knockout mice but obviously not *vice versa*. A plausible scenario would be that MBD4 is primarily involved in the demethylation of methylated sequences in the context of developmental gene activation, whereas TDG's function is to protect unmethylated promoter sequences across the genome from aberrant *de novo* methylation. This would be supported by the ability of MBD4 to bind to methylated CpGs, and the preferential association of TDG with unmethylated gene promoters.

Conclusion

Given their ability to interrogate the surface of DNA bases by flipping them out of the helix into a selective active site pocket, DNA glycosylases represent efficient tools to specifically recognize unduly modified bases and eliminate them from the DNA, thereby enforcing genetic integrity. Yet, while being perfectly equipped to function in DNA repair, their structure and mode of action could provide for more. This is supported by accumulating evidence for non-canonical functions of these DNA-probing enzymes, the most recent and most intensely discussed being an involvement in DNA demethylation and epigenetic control. Further studies will shed light on the mechanism surrounding the action of DNA glycosylases in such a context, from the signals required for the temporal and spatial regulation of their action to the co-factors necessary for efficient base-flipping and excision.

Acknowledgments We wish to thank Christophe Kunz and David Schuermann for helpful discussion and critical reading of the manuscript. We also like apologize to all colleagues whose work we could not reference due to limited space. This work was supported by a grant from the Swiss National Science Foundation (31003A-122574).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Almeida KH, Sobol RW (2007) A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair (Amst)* 6:695–711
- Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP (2002) Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 30:227–232
- An Q, Robins P, Lindahl T, Barnes DE (2005) C→T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *EMBO J* 24:2205–2213
- An Q, Robins P, Lindahl T, Barnes DE (2007) 5-Fluorouracil incorporated into DNA is excised by the Smug1 DNA glycosylase to reduce drug cytotoxicity. *Cancer Res* 67:940–945
- Aravind L, Koonin EV (2000) The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol* 1, RESEARCH0007
- Au KG, Cabrera M, Miller JH, Modrich P (1988) Escherichia coli mutY gene product is required for specific A-G—C.G mismatch correction. *Proc Natl Acad Sci USA* 85:9163–9166
- Baba D, Maita N, Jee JG, Uchimura Y, Saitoh H, Sugawara K, Hanaoka F, Tochio H, Hiroaki H, Shirakawa M (2005) Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* 435:979–982
- Banerjee A, Yang W, Karplus M, Verdine GL (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* 434:612–618
- Bardwell PD, Martin A, Wong E, Li Z, Edelman W, Scharff MD (2003) Cutting edge: the G-U mismatch glycosylase methyl-CpG binding domain 4 is dispensable for somatic hypermutation and class switch recombination. *J Immunol* 170:1620–1624
- Barrett TE, Savva R, Panayotou G, Barlow T, Brown T, Jiricny J, Pearl LH (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* 92:117–129
- Barrett TE, Schärer OD, Savva R, Brown T, Jiricny J, Verdine GL, Pearl LH (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J* 18:6599–6609
- Begley TJ, Haas BJ, Noel J, Shekhtman A, Williams WA, Cunningham RP (1999) A new member of the endonuclease III family of DNA repair enzymes that removes methylated purines from DNA. *Curr Biol* 9:653–656
- Begum NA, Izumi N, Nishikori M, Nagaoka H, Shinkura R, Honjo T (2007) Requirement of non-canonical activity of uracil DNA glycosylase for class switch recombination. *J Biol Chem* 282:731–742
- Begum NA, Stanlie A, Doi T, Sasaki Y, Jin HW, Kim YS, Nagaoka H, Honjo T (2009) Further evidence for involvement of a noncanonical function of uracil DNA glycosylase in class switch recombination. *Proc Natl Acad Sci USA* 106:2752–2757
- Bellacosa A, Cicchillitti L, Schepis F, Riccio A, Yeung AT, Matsumoto Y, Golemis EA, Genuardi M, Neri G (1999) MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc Natl Acad Sci USA* 96:3969–3974
- Berg OG, Winter RB, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929–6948
- Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463:1042–1047
- Blainey PC, Luo G, Kou SC, Mangel WF, Verdine GL, Bagchi B, Xie XS (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* 16:1224–1229
- Boal AK, Genereux JC, Sontz PA, Gralnick JA, Newman DK, Barton JK (2009) Redox signaling between DNA repair proteins for efficient lesion detection. *Proc Natl Acad Sci USA* 106:15237–15242
- Boorstein RJ, Cummings A Jr, Marenstein DR, Chan MK, Ma Y, Neubert TA, Brown SM, Teebor GW (2001) Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J Biol Chem* 276:41991–41997
- Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* 62:235–251
- Bruner SD, Norman DP, Verdine GL (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* 403:859–866
- Bruniquel D, Schwartz RH (2003) Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat Immunol* 4:235–240
- Cannon SV, Cummings A, Teebor GW (1988) 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. *Biochem Biophys Res Commun* 151:1173–1179
- Cannon-Carlson SV, Gokhale H, Teebor GW (1989) Purification and characterization of 5-hydroxymethyluracil-DNA glycosylase from calf thymus. Its possible role in the maintenance of methylated cytosine residues. *J Biol Chem* 264:13306–13312
- Coticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. *Mol Biol Evol* 22:367–377
- Cortazar D, Kunz C, Saito Y, Steinacher R, Schär 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 AL, Siegrist F, Steinacher R, Jiricny J, Bird A, Schär P (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* 470:419–423
- Cortellino S, Turner D, Masciullo V, Schepis F, Albino D, Daniel R, Skalka AM, Meropol NJ, Alberti C, Larue L, Bellacosa A (2003) The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc Natl Acad Sci USA* 100:15071–15076
- Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination–base excision repair. *Cell* 146:67–79
- Cunningham RP, Asahara H, Bank JF, Scholes CP, Salerno JC, Surerus K, Munck E, McCracken J, Peisach J, Emptage MH (1989) Endonuclease III is an iron–sulfur protein. *Biochemistry* 28:4450–4455
- Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E, Reik W (2001) Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc Natl Acad Sci USA* 98:13734–13738

- Denver DR, Swenson SL, Lynch M (2003) An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Mol Biol Evol* 20:1603–1611
- Dherin C, Radicella JP, Dizdaroglu M, Boiteux S (1999) Excision of oxidatively damaged DNA bases by the human alpha-hOgg1 protein and the polymorphic alpha-hOgg1(Ser326Cys) protein which is frequently found in human populations. *Nucleic Acids Res* 27:4001–4007
- Di Noia J, Neuberger MS (2002) Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 419:43–48
- Di Noia JM, Rada C, Neuberger MS (2006) SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J* 25:585–595
- Dizdaroglu M, Karahalil B, Senturker S, Buckley TJ, Roldan-Arjona T (1999) Excision of products of oxidative DNA base damage by human NTH1 protein. *Biochemistry* 38:243–246
- Dou H, Mitra S, Hazra TK (2003) Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J Biol Chem* 278:49679–49684
- Eide L, Luna L, Gustad EC, Henderson PT, Essigmann JM, Demple B, Seeberg E (2001) Human endonuclease III acts preferentially on DNA damage opposite guanine residues in DNA. *Biochemistry* 40:6653–6659
- Engelward BP, Weeda G, Wyatt MD, Broekhof JL, de Wit J, Donker I, Allan JM, Gold B, Hoeijmakers JH, Samson LD (1997) Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase. *Proc Natl Acad Sci USA* 94:13087–13092
- Falnes PO, Johansen RF, Seeberg E (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* 419:178–182
- Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T (2006) DNA repair and mutagenesis, 2nd edn. ASM, Washington
- Frieder D, Larijani M, Collins C, Shulman M, Martin A (2009) The concerted action of Msh2 and UNG stimulates somatic hypermutation at A. T base pairs. *Mol Cell Biol* 29:5148–5157
- Friedman JI, Stivers JT (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* 49:4957–4967
- Friedman JI, Majumdar A, Stivers JT (2009) Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res* 37:3493–3500
- Fromme JC, Bruner SD, Yang W, Karplus M, Verdine GL (2003) Product-assisted catalysis in base-excision DNA repair. *Nat Struct Biol* 10:204–211
- Gallinari P, Jiricny J (1996) A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature* 383:735–738
- Guan Y, Manuel RC, Arvai AS, Parikh SS, Mol CD, Miller JH, Lloyd S, Tainer JA (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat Struct Biol* 5:1058–1064
- Hagen L, Kavli B, Sousa MM, Torseth K, Liabakk NB, Sundheim O, Pena-Diaz J, Otterlei M, Horning O, Jensen ON, Krokan HE, Slupphaug G (2008) Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. *EMBO J* 27:51–61
- Hang B, Singer B, Margison GP, Elder RH (1997) Targeted deletion of alkylpurine-DNA-N-glycosylase in mice eliminates repair of 1, N6-ethenoadenine and hypoxanthine but not of 3, N4-ethenocytosine or 8-oxoguanine. *Proc Natl Acad Sci U S A* 94:12869–12874
- Hardeland U, Bentele M, Jiricny J, Schär 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, Schär P (2002) Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J* 21:1456–1464
- Hardeland U, Bentele M, Jiricny J, Schär 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, Kunz C, Focke F, Szadkowski M, Schär 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
- Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell* 113:803–809
- Haushalter KA, Todd Stukenberg MW, Kirschner MW, Verdine GL (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr Biol* 9:174–185
- Hazra TK, Izumi T, Boldogh I, Imhoff B, Kow YW, Jaruga P, Dizdaroglu M, Mitra S (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc Natl Acad Sci USA* 99:3523–3528
- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* [Epub ahead of print]
- Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18:6538–6547
- Hermann A, Gowher H, Jeltsch A (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol Life Sci* 61:2571–2587
- Hill JW, Hazra TK, Izumi T, Mitra S (2001) Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. *Nucleic Acids Res* 29:430–438
- Holmes RK, Malim MH, Bishop KN (2007) APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci* 32:118–128
- Horst JP, Fritz HJ (1996) Counteracting the mutagenic effect of hydrolytic deamination of DNA 5-methylcytosine residues at high temperature: DNA mismatch N-glycosylase Mig. *Mth of the thermophilic archaeon Methanobacterium thermoautotrophicum* THF. *EMBO J* 15:5459–5469
- Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, Forveille M, Kavli B, Krokan HE, Ochs HD, Fischer A, Durandy A (2003) Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 4:1023–1028
- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, 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 SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* [Epub ahead of print]
- Iyer LM, Tahiliani M, Rao A, 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
- Janik J, Swoboda M, Janowska B, Ciesla JM, Gackowski D, Kowalewski J, Olinski R, Tudek B, Speina E (2011) 8-Oxoguanine incision activity is impaired in lung tissues of NSCLC patients with the polymorphism of OGG1 and XRCC1 genes. *Mutat Res* 709–710:21–31

- Jones S, Emmerson P, Maynard J, Best JM, Jordan S, Williams GT, Sampson JR, Cheadle JP (2002) Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C→T:A mutations. *Hum Mol Genet* 11:2961–2967
- Jost JP, Siegmund M, Sun L, 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
- Jost JP, Fremont M, Siegmund M, Hofsteenge J (1997) The RNA moiety of chick embryo 5-methylcytosine-DNA glycosylase targets DNA demethylation. *Nucleic Acids Res* 25:4545–4550
- Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J, Cedar H, Razin A (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev* 6:705–714
- Kangaspekka S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G (2008) Transient cyclical methylation of promoter DNA. *Nature* 452:112–115
- Karahalil B, Girard PM, Boiteux S, Dizdaroglu M (1998) Substrate specificity of the OggI protein of *Saccharomyces cerevisiae*: excision of guanine lesions produced in DNA by ionizing radiation- or hydrogen peroxide/metal ion-generated free radicals. *Nucleic Acids Res* 26:1228–1233
- Kavli B, Slupphaug G, Mol CD, Arvai AS, Peterson SB, Tainer JA, Krokan HE (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J* 15:3442–3447
- Kavli B, Sundheim O, Akbari M, Otterlei M, Nilsen H, Skorpén F, Aas PA, Hagen L, Krokan HE, Slupphaug G (2002) hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J Biol Chem* 277:39926–39936
- Kim MS, Kondo T, Takada I, Youn MY, Yamamoto Y, Takahashi S, Matsumoto T, Fujiyama S, Shirode Y, Yamaoka I, Kitagawa H, Takeyama K, Shibuya H, Ohtake F, Kato S (2009) DNA demethylation in hormone-induced transcriptional derepression. *Nature* 461:1007–1012
- Klungland A, Rosewell I, Hollenbach S, Larsen E, Daly G, Epe B, Seeberg E, Lindahl T, Barnes DE (1999) Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proc Natl Acad Sci USA* 96:13300–13305
- Krokan HE, Standal R, Slupphaug G (1997) DNA glycosylases in the base excision repair of DNA. *Biochem J* 325:1–16
- Kunz C, Saito Y, Schär P (2009a) DNA repair in mammalian cells: mismatched repair: variations on a theme. *Cell Mol Life Sci* 66:1021–1038
- Kunz C, Focke F, Saito Y, Schuermann D, Lettieri T, Selfridge J, Schär P (2009b) Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol* 7:e91
- Kuo CF, McRee DE, Fisher CL, O'Handley SF, Cunningham RP, Tainer JA (1992) Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. *Science* 258:434–440
- Lee CY, Delaney JC, Kartalou M, Lingaraju GM, Maor-Shoshani A, Essigmann JM, Samson LD (2009) Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG). *Biochemistry* 48(9):1850–1861
- Lindahl T (1974) An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc Natl Acad Sci USA* 71:3649–3653
- Lindahl T, Wood RD (1999) Quality control by DNA repair. *Science* 286:1897–1905
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133:523–536
- Liu M, Bandaru V, Bond JP, Jaruga P, Zhao X, Christov PP, Burrows CJ, Rizzo CJ, Dizdaroglu M, Wallace SS (2010) The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *Proc Natl Acad Sci USA* 107:4925–4930
- Lu R, Nash HM, Verdine GL (1997) A mammalian DNA repair enzyme that excises oxidatively damaged guanines maps to a locus frequently lost in lung cancer. *Curr Biol* 7:397–407
- Ma DK, Jang MH, Guo JU, Kitabatake Y, Chang ML, Pow-Anpongkul N, Flavell RA, Lu B, Ming GL, Song H (2009) Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* 323:1074–1077
- Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* [Epub ahead of print]
- Maiti A, Morgan MT, Pozharski E, Drohat AC (2008) Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc Natl Acad Sci USA* 105:8890–8895
- Maiti A, Morgan MT, Drohat AC (2009) Role of two strictly conserved residues in nucleotide flipping and N-glycosyl bond cleavage by human thymine DNA glycosylase. *J Biol Chem* 284:36680–36688
- Mayer W, Niveleau A, Walter J, Fundele R, Haaf T (2000) Demethylation of the zygotic paternal genome. *Nature* 403:501–502
- McGoldrick JP, Yeh YC, Solomon M, Essigmann JM, Lu AL (1995) Characterization of a mammalian homolog of the *Escherichia coli* MutY mismatch repair protein. *Mol Cell Biol* 15:989–996
- Meira LB, Moroski-Erkul CA, Green SL, Calvo JA, Bronson RT, Shah D, Samson LD (2009) Aag-initiated base excision repair drives alkylation-induced retinal degeneration in mice. *Proc Natl Acad Sci USA* 106:888–893
- Melamede RJ, Hatahet Z, Kow YW, Ide H, Wallace SS (1994) Isolation and characterization of endonuclease VIII from *Escherichia coli*. *Biochemistry* 33:1255–1264
- Metivier R, Gallais R, Tiffocche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452:45–50
- Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297:403–405
- Miller CA, Sweatt JD (2007) Covalent modification of DNA regulates memory formation. *Neuron* 53:857–869
- Mol CD, Arvai AS, Slupphaug G, Kavli B, Alseth I, Krokan HE, Tainer JA (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* 80:869–878
- Monk M, Boubelik M, 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
- Morland I, Rolseth V, Luna L, Rognes T, Bjoras M, Seeberg E (2002) Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res* 30:4926–4936
- Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T (2000) Class switch recombination and hypermutation

- require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102:553–563
- Nash HM, Bruner SD, Schärer OD, Kawate T, Addona TA, Spooner E, Lane WS, Verdine GL (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr Biol* 6:968–980
- Nash HM, Lu R, Lane WS, Verdine GL (1997) The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOgg1: direct identification, ablation and chemical reconstitution. *Chem Biol* 4:693–702
- Neddermann P, Gallinari P, Lettieri T, Schmid D, Truong O, Hsuan JJ, Wiebauer K, Jiricny J (1996) Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J Biol Chem* 271:12767–12774
- Nilsen H, Otterlei M, Haug T, Solum K, Nagelhus TA, Skorpen F, Krokan HE (1997) Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic Acids Res* 25:750–755
- Nilsen H, Rosewell I, Robins P, Skjelbred CF, Andersen S, Slupphaug G, Daly G, Krokan HE, Lindahl T, Barnes DE (2000) Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol Cell* 5:1059–1065
- Nilsen H, Stamp G, Andersen S, Hrivnak G, Krokan HE, Lindahl T, Barnes DE (2003) Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene* 22:5381–5386
- Ocampo MT, Chaung W, Marenstein DR, Chan MK, Altamirano A, Basu AK, Boorstein RJ, Cunningham RP, Teebor GW (2002) Targeted deletion of mNth1 reveals a novel DNA repair enzyme activity. *Mol Cell Biol* 22:6111–6121
- O'Connor TR (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic Acids Res* 21:5561–5569
- O'Connor TR, Laval F (1990) Isolation and structure of a cDNA expressing a mammalian 3-methyladenine-DNA glycosylase. *EMBO J* 9:3337–3342
- Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J (2000) Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* 10:475–478
- Otterlei M, Warbrick E, Nagelhus TA, Haug T, Slupphaug G, Akbari M, Aas PA, Steinsbekk K, Bakke O, Krokan HE (1999) Post-replicative base excision repair in replication foci. *EMBO J* 18:3834–3844
- Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE, Tainer JA (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J* 17:5214–5226
- Parikh SS, Mol CD, Hosfield DJ, Tainer JA (1999) Envisioning the molecular choreography of DNA base excision repair. *Curr Opin Struct Biol* 9:37–47
- Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* 449:433–437
- Pavri R, Nussenzweig MC (2011) AID targeting in antibody diversity. *Adv Immunol* 110:1–26
- Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL (2007) DNA demethylation in the *Arabidopsis* genome. *Proc Natl Acad Sci USA* 104:6752–6757
- Petronzelli F, Riccio A, Markham GD, Seeholzer SH, Genuardi M, Karbowski M, Yeung AT, Matsumoto Y, Bellacosa A (2000) Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *J Cell Physiol* 185:473–480
- Pettersen HS, Visnes T, Vagbo CB, Svaasand EK, Doseth B, Slupphaug G, Kavli B, Krokan HE (2011) UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation. *Nucleic Acids Res*. doi:gkr563 [pii]
- Pope MA, Porello SL, David SS (2002) *Escherichia coli* apurinic-apyrimidinic endonucleases enhance the turnover of the adenine glycosylase MutY with G:A substrates. *J Biol Chem* 277:22605–22615
- Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463:1101–1105
- Porello SL, Cannon MJ, David SS (1998) A substrate recognition role for the [4Fe-4S]₂₊ cluster of the DNA repair glycosylase MutY. *Biochemistry* 37:6465–6475
- Rada C, Ehrenstein MR, Neuberger MS, Milstein C (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 9:135–141
- Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS (2002) Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 12:1748–1755
- Radany EH, Dornfeld KJ, Sanderson RJ, Savage MK, Majumdar A, Seidman MM, Mosbaugh DW (2000) Increased spontaneous mutation frequency in human cells expressing the phage PBS2-encoded inhibitor of uracil-DNA glycosylase. *Mutat Res* 461(1):41–58
- Radicella JP, Dherin C, Desmaze C, Fox MS, Boiteux S (1997) Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 94:8010–8015
- Radman M (1976) An endonuclease from *Escherichia coli* that introduces single polynucleotide chain scissions in ultraviolet-irradiated DNA. *J Biol Chem* 251:1438–1445
- Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135:1201–1212
- Robertson AB, Klungland A, Rognes T, Leiros I (2009) DNA repair in mammalian cells: base excision repair: the long and short of it. *Cell Mol Life Sci* 66:981–993
- Rosenquist TA, Zaika E, Fernandes AS, Zharkov DO, Miller H, Grollman AP (2003) The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. *DNA Repair (Amst)* 2:581–591
- Rusmintrati V, Sowers LC (2000) An unexpectedly high excision capacity for mispaired 5-hydroxymethyluracil in human cell extracts. *Proc Natl Acad Sci USA* 97:14183–14187
- Russo MT, De Luca G, Degan P, Parlanti E, Dogliotti E, Barnes DE, Lindahl T, Yang H, Miller JH, Bignami M (2004) Accumulation of the oxidative base lesion 8-hydroxyguanine in DNA of tumor-prone mice defective in both the Myh and Ogg1 DNA glycosylases. *Cancer Res* 64:4411–4414
- Savva R, McAuley-Hecht K, Brown T, Pearl L (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* 373:487–493
- Schanz S, Castor D, Fischer F, Jiricny J (2009) Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc Natl Acad Sci USA* 106:5593–5598
- Schrofelbauer B, Yu Q, Zeitlin SG, Landau NR (2005) Human immunodeficiency virus type 1 Vpr induces the degradation of the UNG and SMUG uracil-DNA glycosylases. *J Virol* 79:10978–10987
- Schrofelbauer B, Hakata Y, Landau NR (2007) HIV-1 Vpr function is mediated by interaction with the damage-specific DNA-binding protein DDB1. *Proc Natl Acad Sci USA* 104:4130–4135
- Schwarz S, Bourgeois C, Soussaline F, Homsy C, Podesta A, Jost JP (2000) A CpG-rich RNA and an RNA helicase tightly associated

- with the DNA demethylation complex are present mainly in dividing chick embryo cells. *Eur J Cell Biol* 79:488–494
- Sheehy AM, Gaddis NC, Malim MH (2003) The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* 9:1404–1407
- Shimizu Y, Iwai S, Hanaoka F, Sugawara K (2003) Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase. *EMBO J* 22:164–173
- Shimizu Y, Uchimura Y, Dohmae N, Saitoh H, Hanaoka F, Sugawara K (2010) Stimulation of DNA glycosylase activities by XPC protein complex: roles of protein–protein interactions. *J Nucleic Acids* 2010
- Simonsson S, Gurdon J (2004) DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. *Nat Cell Biol* 6:984–990
- Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* 384:87–92
- Stanczyk M, Sliwinski T, Cuchra M, Zubowska M, Bielecka-Kowalska A, Kowalski M, Szemraj J, Mlynarski W, Majsterek I (2011) The association of polymorphisms in DNA base excision repair genes XRCC1, OGG1 and MUTYH with the risk of childhood acute lymphoblastic leukemia. *Mol Biol Rep* 38:445–451
- Steinacher R, Schär P (2005) Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr Biol* 15:616–623
- Sun B, Latham KA, Dodson ML, Lloyd RS (1995) Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. *J Biol Chem* 270:19501–19508
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930–935
- Takao M, Kanno S, Kobayashi K, Zhang QM, Yonei S, van der Horst GT, Yasui A (2002) A back-up glycosylase in Nth1 knock-out mice is a functional Nei (endonuclease VIII) homologue. *J Biol Chem* 277:42205–42213
- Thomas D, Scot AD, Barbey R, Padula M, Boiteux S (1997) Inactivation of OGG1 increases the incidence of G. C→T. A transversions in *Saccharomyces cerevisiae*: evidence for endogenous oxidative damage to DNA in eukaryotic cells. *Mol Gen Genet* 254:171–178
- Tini M, Benecke A, Um SJ, Torchia J, Evans RM, Chambon P (2002) Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol Cell* 9:265–277
- Torisu K, Tsuchimoto D, Ohnishi Y, Nakabeppu Y (2005) Hematopoietic tissue-specific expression of mouse Neil3 for endonuclease VIII-like protein. *J Biochem* 138:763–772
- Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419:174–178
- Um S, Harbers M, Benecke A, Pierrat B, Losson R, 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
- Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 67:946–950
- van der Kemp PA, Thomas D, Barbey R, de Oliveira R, Boiteux S (1996) Cloning and expression in *Escherichia coli* of the OGG1 gene of *Saccharomyces cerevisiae*, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. *Proc Natl Acad Sci USA* 93:5197–5202
- van Loon B, Markkanen E, Hubscher U (2010) Oxygen as a friend and enemy: how to combat the mutational potential of 8-oxoguanine. *DNA Repair (Amst)* 9:604–616
- Vartanian V, Lowell B, Minko IG, Wood TG, Ceci JD, George S, Ballinger SW, Corless CL, McCullough AK, Lloyd RS (2006) The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. *Proc Natl Acad Sci USA* 103:1864–1869
- Waters TR, Swann PF (1998) Kinetics of the action of thymine DNA glycosylase. *J Biol Chem* 273:20007–20014
- Waters TR, Gallinari P, Jiricny J, Swann PF (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
- Wibley JE, Waters TR, Haushalter K, Verdine GL, Pearl LH (2003) Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol Cell* 11:1647–1659
- Wiesendanger M, Kneitz B, Edelmann W, Scharff MD (2000) Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J Exp Med* 191:579–584
- Wilson SH, Kunkel TA (2000) Passing the baton in base excision repair. *Nat Struct Biol* 7:176–178
- Wilson TM, Vaisman A, Martomo SA, Sullivan P, Lan L, Hanaoka F, Yasui A, Woodgate R, Gearhart PJ (2005) MSH2-MSH6 stimulates DNA polymerase ϵ , suggesting a role for A:T mutations in antibody genes. *J Exp Med* 201:637–645
- Wirtz S, Nagel G, Eshkind L, Neurath MF, Samson LD, Kaina B (2010) Both base excision repair and O6-methylguanine-DNA methyltransferase protect against methylation-induced colon carcinogenesis. *Carcinogenesis* 31:2111–2117
- Wong E, Yang K, Kuraguchi M, Werling U, Avdievich E, Fan K, Fazzari M, Jin B, Brown AM, Lipkin M, Edelmann W (2002) Mbd4 inactivation increases C→T transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci USA* 99:14937–14942
- Yang B, Chen K, Zhang C, Huang S, Zhang H (2007) Virion-associated uracil DNA glycosylase-2 and apurinic/aprimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. *J Biol Chem* 282:11667–11675
- Zharkov DO, Golan G, Gilboa R, Fernandes AS, Gerchman SE, Kycia JH, Rieger RA, Grollman AP, Shoham G (2002) Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J* 21:789–800
- Zhu JK (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet* 43:143–166
- Zhu B, Zheng Y, Angliker H, Schwarz S, Thiry S, Siegmann M, Jost JP (2000a) 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res* 28:4157–4165
- Zhu B, Zheng Y, Hess D, Angliker H, Schwarz S, Siegmann M, Thiry S, Jost JP (2000b) 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc Natl Acad Sci USA* 97:5135–5139
- Zhu B, Benjamin D, Zheng Y, Angliker H, Thiry S, Siegmann M, Jost JP (2001) 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