Identification of a Possible Role of Thymine DNA Glycosylase (TDG) in Epigenome Maintenance

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Daniel Cortázar

aus Deutschland

Basel, 2011

Genehmigt von der Philosodhisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Martin Spiess (Fakultätsverantwortlicher)

Prof. Dr. Primo Schär (Dissertationsleiter)

Prof. Dr. Antonius Rolink (Korreferent)

Basel, den 15. September 2009

Prof. Dr. Eberhard Parlow

List of Abbreviations

5-hmeC: 5-hydroxymethyl cytosine

5-meC: 5-methylcytosine

AP-site: Abasic site

BER: Base excision repair

ChIP: Chromatin IP

CSR: Class-switch recombination

DNMT: DNA methyltrasnferase

dRP: 5'-deoxyribose-phosphate

DSBR: Double strand break repair

ES cells: Embryonic stem cells

GGR: Global genome repair

H3K27: Histone 3 Lysine 27

H3K4: Histone 3 Lysine 4

H3K9 Histone 3 Lysine 9

HAT: Histone acetyltransferase

HDAC: Histone deacetylase

HMT: Histone methyltransferase

HR: Homologous recombination

ICM: Inner cell mass

IDL: Insertion deletion loop

IP: Immunoprecipitation

IR: Ionizing radiation

MBD: Methyl-binding domain

Me: Methylated

MeDIP: Methylated DNA immunoprecipitation

MEF: Mouse embryonic fibroblast

MMR: Mismatch repair

MS: Mass spectrometry

MUG: Mismatch-specific Uracil DNA glycosylase

NER: Nucleotide excision repair

NHEJ: Non-homologous end-joining

NP: Neuronal progenitor

PcG: Polycomb group

PRC: Polycomb repressive complex

RA: Retinoic acid

SHM: Somatic hypermutation

SUMO: Small ubiquitin modifier

TAP: Tandem Affinity Purification

TCR: Transcription coupled Repair

TDG: Thymine DNA Glycosylase

TEs: Transposable element

TE: Trophectoderm

TF: Transcription factor

UV: Ultrasviolett

1 Summary

Thymine DNA glycosylase (TDG) was discovered as an enzyme capable of removing uracil (U) and thymine (T) from G/U and G/T mispairs, respectively. Owing to this ability, TDG was proposed to initiate restoration of C/G pairs at sites of cytosine or 5-methycytosine (5-meC) deamination. In addition to products of base deamination, the substrate spectrum of TDG covers a wide range of DNA base damages resulting from base oxidation and alkylation [1]. TDG was also found to engage in physical and functional interactions with transcription factors [2-5], and more recent evidence supports additional interactions with the *de novo* DNA methyltransferases Dnmt3a and 3b in the context of gene transcription [6]. Together with its biochemical properties, these observations suggest that TDG might be targeted to gene regulatory sequences as part of a macromolecular assembly to control their functional integrity. TDG may counteract the mutagenic effects of C and 5-meC deamination in CG-rich regions and/or be involved in the maintenance of CpG promoter methylation patterns. A tight regulation of CpG methylation at gene regulatory regions is critical for accurate gene expression, proper cellular differentiation and embryonic development [7]. A somewhat surprising but in this context consistent finding was that, in contrast to other DNA glycosylases, TDG is essential for proper fetal development since a targeted knockout of the gene leads to embryonic lethality [1].

To gain insights into the biological functions of TDG, we aimed to establish and apply biochemical fractionation procedures for high affinity purification and structural and functional characterization of TDG containing proteins complexes.
The first part of the thesis was concerned with biochemical characterization of the protein interaction network of TDG in living mammalian cells. To this end, I applied different approaches allowing high affinity isolation of protein complexes from mammalian cells, such as the tandem affinity purification (TAP) method as well as immunoprecipitation of endogenous protein and of the TDGa isoform from *TdgA* overexpressing embryonic stem (ES) cells. These efforts, however, did not reveal any

TDG interacting partners in subsequent mass spectrometry (MS) analyses. These results were surprising, as TDG was previously reported to interact with transcription factors and DNA methyltransferases [2-6]. Remarkably, however, all previously identified protein interactors of TDG were discovered in screen with the respective partner proteins, and under conditions of simultaneous overexpression of both interacting proteins. The only proteins ever identified in screen with TDG were Sumo1 and Sumo3, which turned out to covalently modify the glycosylase [8, 9]. For this reason, we decided to pursue our search with classical cell fractionation experiments. We first did gel filtration experiments from total cell lysates and showed that TDG is indeed able to form distinct multiprotein complexes in undifferentiated mouse embryonic stem cells that may also contain the RNA helicase p68 [10]. Further subcellular fractionation experiments then revealed that TDG is present in all cell compartments, with a significant fraction of nuclear TDG being associated with chromatin, together with p68 and *de novo* DNA methyltransferases. Together with published findings, these results suggested that protein complexes containing TDG might act in a chromatin-associated context, at gene regulatory regions.

The developmental phenotype of *Tdg*$^{-/-}$ knockout mice and the interactions of TDG with factors involved in developmental gene regulation (e.g. retinoic acid receptors RAR/RXR) implicate a function of TDG during early development and cell differentiation, at times governed by dynamic changes in gene expression, DNA methylation and histone modifications. Such changes have been studied using a well-established during *in vitro* differentiation of ES cells to lineage committed neuronal progenitors (NPs) [11]. We thus aimed to address the function of TDG as part of chromatin associated protein complexes during the process of retinoic acid induced differentiation of ES cells to NPs.

In the second part of the thesis we made use of a this well-established *in vitro* differentiation system to examine the genome-wide localization of TDG to chromatin by TDG chromatin immunoprecipitation (ChIP) and to correlate TDG association to chromatin with gene expression and DNA methylation changes linked to cellular differentiation. TDG ChIP combined with high throughput sequencing showed that

TDG associates with high preference to CpG islands in promoters of actively transcribed genes or genes poised for transcriptional activation. Such CpG rich sequences are normally unmethylated in mammalian genomes. Interestingly, we found TDG to localize to promoters of many genes controlling pluripotency (e.g. *Oct4, Nanog*) and developmental processes (e.g. *Sfrp2, Tgfb2, Gata6*), thus, supporting a function of TDG in cell differentiation and/or embryonic development. As different lines of circumstantial evidence have associated TDG with changes in CpG methylation following activation of hormone responsive gene promoters, we went on to further test genome-wide promoter methylation in $Tdg^{+/-}$ and $Tdg^{-/-}$ NPs making use of a combination of methylated DNA immunoprecipitation (MeDIP) and microarray technology. This showed that the loss of TDG does not affect global promoter DNA methylation. Nevertheless, there were a number of significant differences, suggesting that TDG might affect the CpG methylation pattern at some promoters. Also, owing to the limited resolution of the MeDIP method, however, we could not exclude an involvement of TDG in the control of DNA methylation of specific promoter CpGs. Additional bisulfite sequencing of promoters of TDG bound developmental genes (e.g. *Sfrp2, Tgfb2*) in NPs and differentiated mouse embryonic fibroblasts (MEFs) have indeed proved that loss of TDG affects local changes in DNA methylation at particular CpGs.

Subsequent analysis of genome-wide gene expression profiles of ES cells and differentiated $Tdg^{+/-}$ and $Tdg^{-/-}$ NPs revealed that a limited number of genes (229) are differentially regulated in ES, whereas substantial differences in gene expression in were observed in NPs (1022 genes). This implicated a specific function of TDG in the regulation of cell differentiation triggered gene expression changes. Detailed analysis of the expression of the *Pax6* gene, accurate regulation of which is essential for proper neuron development [12], showed that its promoter is bound by TDG and that its transcription is inappropriately regulated upon further differentiation of $Tdg^{-/-}$ NPs into the neuronal lineage. Whereas $Tdg^{+/-}$ NPs efficiently downregulated *Pax6* (50x) and further differentiated into neuron-like cells, $Tdg^{-/-}$ NPs only partially downregulated *Pax6* gene expression (6x) and underwent apoptosis at day 2 after plating in neuronal medium. This phenotype was complemented by expression of

TDGa, clearly implicating TDG in the regulation of *Pax6* expression during differentiation of ES cells to terminal neurons.

We further observed misregulation of pluripotency genes (e.g. *Oct4*) regulated by TDG bound promoters during early differentiation of ES cells. In the absence of TDG, ES cells showed the tendency to enter spontaneous and/or RA induced differentiation, suggesting a role for TDG in the regulation of pluripotency. During RA induced differentiation we further observed the activation of the neuron specific gene *Lrrtm2* exclusively in TDG proficient cells. In addition, ChIP experiments showed that transcription factors involved in the activation of the Lrrtm2 gene (e.g. COUP-TFI, RAR) are not recruited to the respective promoter in $Tdg^{-/-}$ cells, suggesting that TDG might act passively as a scaffold factor important for the recruitment of transcription factors to promoter regions.

I set out to clarify the biological function of TDG by investigating its molecular interactions in mammalian cells. I found that TDG, as a DNA repair enzyme, associates tightly with chromatin, where it localizes with high preference to CpG island promoters of active genes and genes poised to be expressed. I also found that the loss of TDG causes misregulation of genes during cell differentiation and that this appears to be related to a function of TDG in establishing and/or maintaining CpG methylation pattern in gene regulatory sequences. These discoveries implicate a novel function of DNA repair, in the maintenance not only of the genome, but also the epigenome.

2 Introduction

2.1 DNA Damage and Repair

At the time when its structure was reported, DNA double helix was considered as a stable macromolecular entity [13]. This prevailing view initially hindered the consideration of the existence of DNA damage and of biochemical processes such as DNA mutation and DNA repair. Subsequent work on DNA metabolism, however, disproved the concept of an invulnerable DNA molecule.

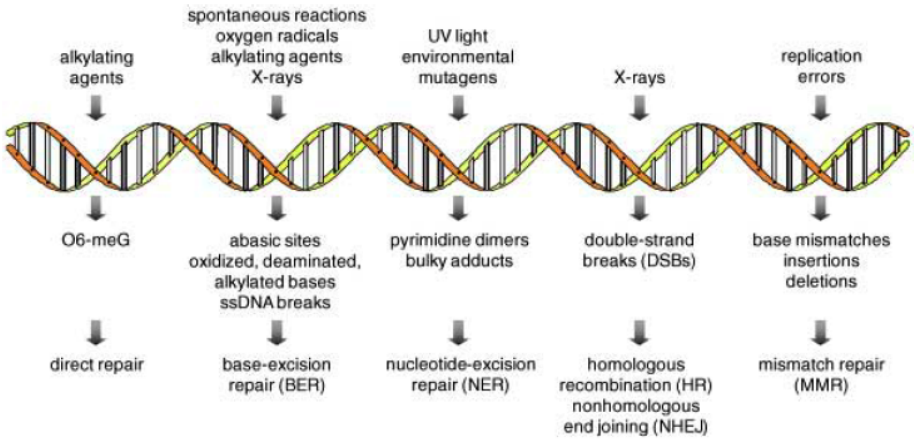


Figure 1. **DNA damage and repair**. Depicted are the most common sources of DNA damage , the lesions generated and the DNA Repair mechanisms concerned. See text for Details. *Adapted from Scharer, O.D. 2003.*

In all living organisms the highly dynamic DNA molecule is susceptible to damage induced by diverse cellular metabolites and exogenous DNA-damaging agents (Fig.1). Modifications of the primary DNA structure, if persisting, can lead to mutations, which alter genetic information or interfere with cellular transactions on the DNA template such as DNA replication and transcription. Ultimately, the distinct DNA lesions can trigger cell death and cause premature aging of multicellular organisms. In mammals, cancer and other age related diseases might be consequences of damaged DNA.

The survival of a species depends on the faithful transmission of genetic information from one generation to the next. This requires continuous surveillance of the

genome in order to minimize the number of heritable DNA lesions. Hence, nature has evolved highly sophisticated DNA damage response strategies, including DNA repair systems that recognize and remove lesions from the DNA, and DNA damage signaling pathways that induce transcriptional responses, cell cycle arrest or apoptosis (Fig.2) [14, 15].

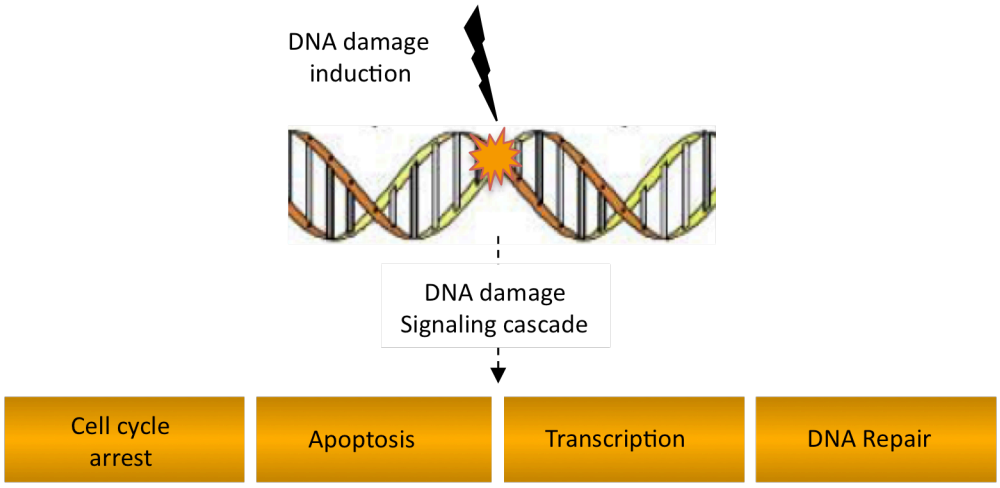


Figure 2. Cellular responses to DNA damage. DNA damage caused by a variety of sources induces a network of complex signaling pathways that can lead to arrest of the cell cycle, initiation of transcriptional programs, DNA repair and, if the damage is severe, to programmed cell death (apoptosis). *Inspired by Jan H. J. Hoeijmakers, 2001*

2.1.1 Types of DNA Damage

DNA can be chemically or physically damaged in a number of different ways. First, chemical bonds in DNA tend to spontaneously disintegrate under physiological conditions. Hydrolysis of the *N*-glycosyl bond of purine and pyrimidine bases leads to base loss and the formation of an abasic site (AP-site), which is cytotoxic, recombinogenic and mutagenic, if left unrepaired. Spontaneous or induced deamination of adenine, guanine, cytosine and 5-methylcytosine leads to the formation of mispairing hypoxanthine, xanthine, uracil and thymine, respectively [16]. Second, the primary structure of the DNA reacts spontaneously with (by-) products of normal cellular metabolism, including reactive oxygen species (superoxide anions, hydroxyl radicals and hydrogen peroxide) derived from oxidative respiration and products of lipid peroxidation [15, 17]. Altogether, these DNA base lesions represent the most frequent form of damage to DNA. Finally, environmental

agents, such as the ultraviolet (UV) radiation, ionizing radiation and numerous genotoxic chemicals cause damage to the DNA structure (e.g. DNA crosslinks, DNA strand breaks), which, if left unrepaired, may lead to mutations and eventually cause cancer [15, 18]. Figure 1 summarizes some of the most common types of DNA damage and their sources.

2.1.2 DNA Repair Mechanisms

Depending on the type of DNA damage, cells employ different strategies of repair. DNA double strand breaks (DSB) belong to the most toxic DNA lesions and are caused by ionizing radiation and radiomimetic chemicals. DSBs may also form spontaneously during DNA replication. Two conceptually distinct pathways for the repair of DSBs have evolved: homologous recombination repair (HR) and non-homologous end-joining (NHEJ). HR requires a homologuous donor sequence for DNA synthesis across the DSB and therefore repairs breaks with high accuracy. HR accounts for the repair of about 10% of DSB in mammalian cells and plays an important role during late S/G2 phase of the cell cycle and at stalled replication forks during DNAreplication [19]. NHEJ is a conceptually more simple, but error-prone repair mechanism for a DSB, directly joining the broken ends following removal of damaged nucleotides. NHEJ accounts for the repair of about 90% of DSB, at least in the G1/G0 phase [20].

Helix-distorting lesions (e.g. bulky DNA adducts) are generated by UV light and carcinogens. Such lesions stall DNA replication and transcription and are eliminated by the Nucleotide Excision Repair (NER) system. In a 'cut-and-patch'-type reaction, an oligonucleotide encompassing the lesion is cut from the DNA duplex and the resulting single-stranded gap is filled in using the intact complementary strand as template. The complexity of NER is illustrated by the fact that it requires more than 30 proteins that assemble into large enzyme complexes for the repair of the damage. NER can operate via two subpathways, Global Genome Repair (GGR) and Transcription-Coupled Repair (TCR). In contrast to GGR, which acts independently of transcription, TCR is specifically targeted to lesions in the transcribed strand of an

active gene and is thus dependent on RNA polymerase I and II driven transcription [18, 20, 21].

The integrity of genetic information depends also on the fidelity of DNA polymerases that replicate the DNA. Misincorporation of bases by DNA polymerases during DNA replication and insertion/deletion loops (IDL) that result from polymerase slippage during the synthesis of repetitive sequences are repaired by the mismatch repair (MMR) system. The MMR machinery also recognizes certain DNA lesions generated by normal intracellular metabolism and extracellular damaging agents [22].

Finally, certain DNA lesions are corrected by direct chemical reversal of the damage by simple enzymatic processes. In this way, DNA photolyases are responsible for the repair of photoproducts induced in DNA by UV light [15, 23, 24]. The O^6-alkylguanine-DNA alkyltransferase, AGT (O^6-methylaguanine methyltransferase (MGMT) in humans), converts O^6-alkylguanine to guanine [18], and the *E. coli* oxidative methyl transferase AlkB and presumably its human homologues demethylate N^1-methyladenine and N^3-methylcytosine to adenine and cytosine, respectively [25, 26].

2.1.3 DNA Base Excision Repair

Damage to DNA bases resulting from deamination, oxidation and alkylation is mainly repaired by the Base Excision Repair (BER) pathway, probably the most frequently engaged DNA repair mode in nature. The pathway is conserved through evolution and it has been possible to reconstitute the process with purified proteins from bacteria and mammals [27-29].

BER is initiated by DNA glycosylases that recognize damaged DNA bases and remove them from the DNA backbone. Based on their three-dimensional structure, DNA Glycosylases are classified into four structural superfamilies: *Helix-hairpin-Helix* (HhH), *Helix-Two-Turn-Helix* (H2TH), *Uracil DNA Glycosylases* (UDGs) and the *alkyladenine DNA Glycosylases* (AAG) [30, 31]. Whereas some DNA glycosylases recognize only one particular type of base damage, others possess rather broad substrate spectra. To date, 11 different mammalian glycosylases have been

characterized, altogether capable of recognizing a wide spectrum of base damages, though with redundancy for certain damages. A list of all known human and *E.coli* DNA glycosylases and their main enzymatic activities is provided in Table 1.

Table 1. Bacterial and human DNA glycosylases and principal DNA substrates [30, 32]

E.coli	Human	Preferred substrate(s)	Mono or bifunctional
Ung	UNG	U, U:G, U:A, 5-FU:G	Mono
Mug	TDG[a]	U:G, T:G, eC:G, eC:A, 5-FU:G, ^{Hm}U:G,…	Mono
	SMUG1	U, U:A, U:G	Mono
	MBD4	U:G, T:G, T:O^{6}meG,	Bi
Tag			Mono
AlkA	AAG/MPG	3-mA, 7-mA, 3-mG, 7-mG, Hx, eA, eG	Mono
Nth	NTH	T/C-glycol(g), Fapy, 5-OH-C/U	Bi
Nei	NEIL1	8-oxoG, Tg, FapyA, FapyG, fragmented Py	Bi
	NEIL2	Oxidized and fragmented Py	Bi
	NEIL3	Unknown	Bi
Fpg	OGG1	8-oxoG, FapyG, me-FapyG	Bi
MutY	MYH	A:G, A:8-oxoG, 2-OH-A:G	Mono

[a]See also [1] and Table 2.

Structural and biophysical studies of DNA glycosylases [31, 33, 34] bound to their cognate DNA substrates have revealed that DNA glycosylases remove damaged bases by a “base flipping” mechanism. This involves the insertion of specific amino acid residues of the enzymes catalytic core into the DNA helix at the position of the lesion, resulting in extrusion of the damaged nucleotide and fitting into a catalytic

pocket, kinking of the DNA and, ultimately, cleavage of the damaged base. DNA glycosylases can be either monofunctional or bifunctional enzymes. Although the damage recognition by both types of glycosylases is comparable, the removal of the damaged base is accomplished by different chemical reactions (Fig.3). Excision of the base by a monofunctional glycosylase involves hydrolysis of the *N*- glycosidic bond of the damaged deoxynucleotide, generating an apurinic/apymidinic site (AP-site). In human cells, further processing of the AP-site involves the AP-endonuclease APE1 (also called HAP-1 or Ref-1), which introduces a nick in the DNA backbone 5' to the AP-site, generating a 3'-OH and a 5'-deoxyribose-phosphate (dRP) [35]. The resulting dRP moiety is then removed by the dRP-lyase activity of DNA polymerase β. Bifunctional glycosylases possess an additional AP-lyase activity, that uses an amino group of an active site lysine residue to form a Schiff's base intermediate. This undergoes enzyme catalyzed β-elimination, which cleaves the phosphodiester bond 3' to the AP-site generating a 3' unsaturated aldehyde and a 5' phosphate end. Once the 3' aldehyde has been properly processed by APE1, a 1-nucleotide gap flanked by a 3'-OH and a 5'-phosphate is generated.

Irrespective of the type of DNA glycosylase involved, the 1-nucleotide gap intermediate is filled in by Polβ [17, 29, 36-38]. Resealing of the remaining nick is done by DNA Ligase III, which interacts with DNA Polymerase beta (Polβ) through the X-ray cross-complementation group 1 protein (XRCC1), a central scaffold factor and matchmaker protein in BER [39, 40].

In vivo, the dRP residues generated by APE1 from an oxidized or reduced AP-site cannot be removed by the β-lyase activity of Pol β. Such residues are removed through long patch BER. During this process, strand displacement can occur by DNA synthesis leading to longer repair patch sizes. In this long patch repair pathway, Polβ probably also incorporates the first nucleotide into the DNA [41], but the additional elongation relies on the activity of Polδ/ε together with PCNA and the clamp loader RFC, which are both necessary for polymerase anchoring. The repair synthesis introduces 2-10 nucleotides past the AP-site generating a short flap. This single strand overhang is excised by the flap endonuclease FEN-1 and the nick is then sealed by DNA Ligase I [28, 42].

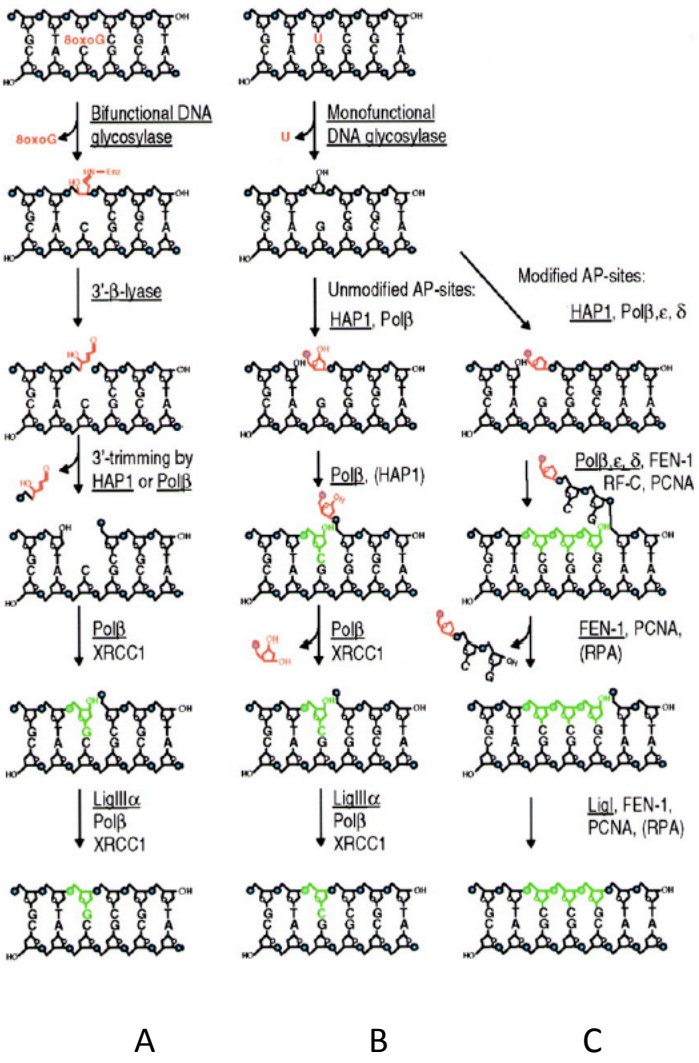


Figure 3. **General scheme of the base excision repair pathway**. The base excision repair (BER) pathway is a multistep process that involved removal of the damaged base by either monofunctional of bifunctional DNA glycosylases. Depending on the type of initiating glycosylase, BER follows two different directions, resulting in either short-patch BER (A, B) or long-patch repair (C) products. The catalytic proteins in each step are underlined. See text for Details. *Adapted from Nilsen, H. and Krokan, H.E. 2001*

Genetic studies in mouse have revealed an appreciable degree of functional redundancy at the level of base damage excision. Mice lacking a single DNA glycosylase often show no overt phenotype [43]. The exception here seems to be the thymine DNA glycoslyase, which as mentioned in the next chapter, represents a special case. The lack of a phenotype is somewhat surprising given the large amount of base damage generated in cells. On the other hand, the targeted deletion of genes that function downstream of the DNA gylcosylases usually leads to a very

severe and often lethal phenotype. This can be explained by the fact that BER, which can be initiated by many different DNA glycosylases, proceeds through identical downstream enzymes.

2.2 The Thymine DNA Glycosylase (TDG)

2.2.1 Classification of TDG and Primary Structure

Deamination of cytosine (C) and 5-methylcytosine (5meC) generates uracil (U) and thymine (T) opposite guanine (G), respectively. If not repaired, the resulting mispaires give rise to C:G to T:A transition mutations after replication. Because uracil itself represents a foreign base in the DNA, its removal appears a straightforward task. Any enzyme able to specifically recognize and excise uracil will suffice to initiate repair, independently of whether it is present in a matched or mismatched condition. In contrast, removal a mispaired thymine is more complicated, as thymine is a perfectly normal base in DNA, except that it is mispaired. For this reason it has to be recognized as “damage” in the context of 5-meC deamination. In the search for a specialized activity able to remove mismatched thymines, Josef Jiricny and colleagues where able to purify and subsequently clone the first G:T mismatch specific DNA glycosylase: the human Thymine DNA Glycosylase (TDG) [44-46]. Human TDG then turned out to be the founding member of the mismatch specific Uracil DNA Glycosylase (MUG) subfamily of DNA glycosylases, which comprises TDG orthologues in bacteria, insects, yeast and vertebrates [1]. Together with the UNG and SMUG1 family of glycosylases, the MUG proteins belong to the superfamily of the Uracil DNA Glycosylases (UDGs), able to remove uracil from DNA.

The eukaryotic TDG proteins are composed of a conserved catalytic core domain and non-conserved N- and C- terminal domains. Composition and conformation of these domains accounts for differences in substrate specificity, substrate interaction and kinetics of base removal [8, 47, 48]. Interestingly, the terminal domains of the human and murine TDG have been shown to be involved in physical and functional interactions with a number of transcriptional regulators including nuclear receptors

and chromatin remodeling enzymes [3-5, 49-55]. This suggests a species- and process-specific role of the terminal domains in targeting the DNA glycosylases to particular genomic regions under specific conditions (see "TDG and the regulation of gene expression").

Despite sharing only 10% amino-acid sequence homology, UDG family members exhibit extensive three-dimensional structure similarities. They all share a central four-stranded parallel twisted β–sheet flanked by α-helices (α/β -fold) [56, 57]. They also share the same intercalation/nucleotide flipping mechanism for base recognition and release, but still, substrate specificity varies between UDG family members. For instance, UNG and SMUG enzymes show high specificity for uracil due to a tight and highly selective catalytic pocket and strategically positioned residues that specifically interact with the substrate base [58-60]. In contrast, MUG proteins have a much larger catalytic cleft allowing the accommodation of various base derivatives [47, 48]. Nevertheless, substrate specificity is constrained due to specific hydrogen-bond interactions between enzyme active-site residues and the guanine in the complementary DNA strand, explaining the preference of these glycosylases for bases mispaired with guanine in a double stranded DNA context [61].

Consistent with 32% amino-acid sequence identity, crystallographic studies of the catalytic domain of human TDG and site-directed mutagenesis of critical residues therein showed that the human glycosylase and its bacterial orthologue share similar substrate binding and catalytic mechanisms [62-64]. However, the wider substrate spectrum and the kinetic properties of the TDG implicate an important role for the N-and C-terminal domains on the enzymatic activity, that cannot be explained by the structures of the glycosylase domains only.

2.2.2 TDG Substrates

MUG proteins of different origins were shown to act on a large number of substrates. These include derivatives of uracil (e.g. 5-fluorouracil), ethenoadducts (e.g 3,N^4-ethenocytosine), deaminated purines (e.g. hypoxanthine), oxidized pyrimidines (e.g. thymine glycol) and normal bases like thymine (Table 2) [47, 65, 66]. In addition, Jean Pierre Jost and his colleagues presented evidence for a 5-

methylcytosine (5-meC) activity associated with a mismatch specific G/T DNA glycosylase that turned out to be a chicken homologue of TDG [67-69]. Surprisingly, however no or only minor 5-meC processing activity was detected with the purified recombinant human and chicken TDG, respectively, suggesting a need for specific cofactors and/or posttranslational modifications to facilitate targeting of the Glycosylase and efficient processing of methylated cytosines.

Considering the broad spectrum of substrates, it seems that MUG proteins not only counteract mutations that arise from cytosine or 5-meC deamination but also protect DNA against base deamination and/or oxidation, as well as against base modifications by products of lipid peroxidation. Furthermore, in organisms that methylate cytosine in their DNA, MUGs may counteract the loss of methylation sites due to 5meC deamination and contribute to the establishment and/or maintenance of DNA methylation patterns, which essential for cell lineage determination and hence for proper development in mammals.

Table 2. TDG orthologs and their substrate preferences

Substrate[b]	hsTDG[c]	ΔNhsTDG[d]	ecMug[d]	spThp1p[c]	dmThd1p[c]
G·**U**	+++	+++	+++	+++	+++
A·**U**	+	+	+	+++	++
ss**U**	–	–	–	+++	–
G·^{F}**U**	+++	nd	nd	+++	+++
A·^{F}**U**	++	nd	nd	+++	++
ss^{F}**U**	++	nd	nd	++	++
G·^{Br}**U**	+++	nd	nd	++	+++
A·^{Br}**U**	+	nd	nd	+	–
ss^{Br}**U**	–	nd	nd	+	–
G·^{Hm}**U**	+++	nd	+	–	++
G·^{H}**U**	+++	nd	++	+++	nd
G·**T**	+++	–	–	–	++
G·**Tg**	++	nd	nd	nd	nd
G·**εC**	+++	nd	+++	+++	+++
A·**εC**	++	nd	nd	+++	++
ss**εC**	–	nd	+	+++	–
G·**Hx**	+	+	nd	+++	+
T·**Hx**	–	–	–	+++	–
ss**Hx**	–	nd	nd	+++	–
G·**εA**	–	nd	+	++	–
T·**εA**	–	nd	–	++	–
ss**εA**	–	nd	nd	+	–
G·^{m}**C**	–/+	nd	nd	–	–
G·^{He}**C**	–	nd	++	+++	nd
G·^{Hp}**C**	–	nd	–	+++	nd
G·**G**		nd	nd	+	

[a] Indicated are relative processing efficiencies of recombinant human full size (hsTDG) and N-terminally truncated TDG (ΔNhsTDG) and the orthologs of *E. coli* (ecMug), *S. pombe* (spThp1p) and *D. melanogaster* (dmThd1p). Base release efficiencies are indicated as: +++, high; ++, intermediate; +, low; –, insignificant.

[b] Putative substrate bases are indicated in bold letters. Abbreviations used are: ss, single-stranded DNA; ds, double-stranded DNA; ^{F}U, 5-fluorouracil; ^{Br}U, 5-bromouracil; ^{Hm}U, 5-hydroxymethyluracil; ^{H}U, 5-hydroxyuracil; Tg, thymine glycol; εC, 3,N^{4}-ethenocytosine; Hx, hypoxanthine; εA, 1,N^{6}-ethenoadenine; ^{m}C, 5-methylcytosine; ^{H}C, 5-hydroxycytosine; ^{He}C, 3,N^{4}-a-hydroxyethanocytosine; ^{Hp}C, 3,N^{4}-a-hydroxypropanocytosine; nd, not done.

[c] Fully AP-site inhibited; no enzymatic turnover.

[d] Partially AP-site inhibited; slow enzymatic turnover.

Cortazar *et al* 2007

2.2.3 Mechanism of Substrate Interaction and Processing

The removal of thymine from G/T mismatches arising from 5-meC deamination is complex and implies the involvement of a stringent and specific mechanism to avoid processing of the huge excess of thymines in canonical A/T base pairs. Indeed, it was shown that hTDG activity is 18'000 fold lower for A/T than for G/T base pairs [70]. Crystallographic analysis provided insight into the mechanistic basis of this mismatch specificity. A structure of the catalytic domain of hTDG bound to an AP-site revealed that the glycosylase forms a wedge invading the complementary strand at the opposite G [61, 71]. In addition, the wedge displaces the 5' located base and establishes a cleft around the remaining G at its 5' face, its Watson-Crick and minor groove boundaries. A structural basis for the specificity for lesions in a CpG context [72] is provided by specific contacts of $hTDG^{cat}$ to the G positioned 3' to the target nucleotide [71].

After removal of the mismatched base, hTDG remains bound to the AP-site because of its high affinity to such repair intermediates [64, 73]. Specific contacts to the Watson-Crick face of the complementary G contribute in part to the inability of full length TDG to dissociate from the AP-site DNA [61, 71]. Furthermore, the non-conserved N-terminal domain of TDG is suggested to stabilize the glycoslyase on the DNA facilitating efficient processing of the G/T substrate and contributing to the slow turnover of the enzyme [8]. Recent NMR data demonstrated that a small structural domain in the TDG N-terminal region encompassing 60 residues preceding the catalytic domain is not only implicated in dynamic molecular interactions with the catalytic domain (suggested to add supplementary substrate recognition surface), but also in nonselective interactions with dsDNA [74]. Thus, it seems that TDG protects the labile AP-site until downstream enzymes take over in the repair pathway at the cost of enzymatic turnover. In this regard, APE1, the enzyme supported to act downstream of TDG in BER was shown to induce TDG turnover on a G/T substrate [73], presumably by passive competition with the AP-site. This however, was efficient only in the presence of a huge excess of APE1 over TDG, implying a need for an active dissociation mechanism.

TDG was shown to be modified by the Small Ubiquitin Like Modifiers (SUMOs) SUMO-1 and SUMO-3 [9]. Covalent attachment of SUMO to a lysine residue (K330 in human TDG, K341 in mouse TDGA) in the VKEE consensus motif of the C-terminal domain of TDG was shown to reduce its affinity for the AP-site *in vitro,* thereby allowing turnover on G/U containing substrates. In addition, equimolar addition of APE1 was sufficient to enhance turnover of SUMOylated TDG in the repair process [9]. Interestingly, truncation of the N-terminal part affected TDG turnover to a similar extend as SUMOylation of the full-length protein. SUMOylation of the truncated protein did not further increase turnover, indicating that modification of the C-terminal domain with SUMO may directly influence the structure of the N-terminal region in the full-length TDG [8].

2.2.4 Biological Functions of TDG

DNA Repair

TDG was first proposed to counteract C to T transition mutation induced by deamination of cytosine and 5-methyl-cytosine. Following the observation that hTDG and its homologues in other organisms process a wider spectrum of DNA base damages MUG proteins were proposed to have a more general function in the repair of DNA base damage than initially predicted.

Nevertheless, processing of a G/T mismatch is most efficiently accomplished by the mammalian TDG, but not by the bacterial, yeast or fly orthologues [47]. The fact that mammals show a higher degree of cytosine methylation in CpG dinuleotides in their genomic DNA suggests that the mammalian TDG has acquired a specific role in the repair of damage resulting from deamination of 5-meC. A second DNA glycosylase, the methyl-CpG-binding domain protein *MBD4* processes a number of substrates in common with TDG, including G/T and G/U mismatches [75, 76]. Yet, it is not clear to what extend any of these glycosylases contributes to G/T processing *in vivo*. The frequency of C to T mutations in Mbd4 *knock-out* mice is slightly increased [77] but nuclear extracts from mouse Tdg *knock-out* cells show no

detectable G/T processing activity. This suggests that TDG is the prevalent glycosylase repairing products of 5-meC deamination. It seems likely though, that G/T repair in vertebrates is provided by both TDG and MBD4 in a partially redundant manner [1].

Although cells seem to possess a repair mechanism protecting against C to T mutation, methylated CpGs nevertheless are frequently mutated in mammalian genomes and C to T transitions are found in the DNA of human cancer cells. For instance, 25-50% of all cancer-associated mutations in the *p53* tumor suppressor gene are C to T transitions at CpG dinucleotides [78]. There are two explanations for such a discrepancy. Some G/T mispaires may escape repair and lead to mutations due to an imbalance between the frequency of 5-meC deaminations arising in the genome and the efficiency of repair of the deamination products. Or, unrepaired G/T mismatches may be targeted by the postreplicative mismatch repair (MMR) system. Not capable of discriminating between wildtype and mutated DNA strand in non-replicating DNA, MMR will fix the mutation by erroneously processing the correct strand (e.g. the G) in 50% of the cases.

Another attractive possibility is that G/T glycosylases act under specific physiological conditions within specific areas of the genome were cytosines are prone to become methylated. This is the case for CpG dinucleotides in promoter regions of vertebrate genes, where dynamic methylation/demethylation processes appear to be involved in the regulation of gene expression. In this context, TDG and/or MBD4 might contribute to a DNA-repair mediated alteration of CpG methylation and therefore in the regulation of gene activity, rather than in the repair of deaminated 5-meC elsewhere in the genome [79].

Although hydrolytic deamination of 5-meC occurs at a higher rate than deamination of the unmethylated cytosine, the generation of G/U mismatches by deamination of the latter is frequent on a genomic scale [1]. Unless repaired, G/U mismatches lead to C to T transitions after DNA replication. In addition, uracil can additionally arise in DNA through the misincorporation of dUMP during DNA replication, giving rise to non-mutagenic A/U mispaires.

In vitro, purified mammalian TDG processes G/U mispairs with high efficiency but shows relatively poor activity on a A/U substrate [80]. In cells, however,

additional uracil DNA glycosylases, such as UNG2, MBD4 and SMUG1, contribute to the removal or U in the DNA and the level of redundancy among these enzymes is not completely understood. Nevertheless, experimental evidence suggests that they act in a context-dependent manner. For instance, UNG2 appears to be the main glycosylase removing U misincorporated during DNA replication [81]. A contribution of TDG to the repair of uracil in this context is excluded by the fact that it is actively downregulated during S-phase [82]. In addition, *UNG* deficient mice accumulate significant amounts of dUMP despite the presence of TDG, SMUG1 and MBD4 [81]. Interestingly, the C to T mutation frequency is only slightly altered in UNG deficient mice, corroborating a redundancy of uracil repair activities at the level of mutation avoidance. Indeed, the mutation frequency was shown to increase synergistically upon downregulation of SMUG1 [83]. The fact that TDG interacts with various transcription factors suggests that the repair of G/U mismatches by this DNA glycosylase is constrained to specific genomic regions and/or physiological contexts.

One example of context specific generation of G/U mispaires is Somatic Hypermutation (SHM) and Class-Switch Recombination (CSR) of the immunoglobulin genes in the process of antibody diversification in activated B cells. Upon lymphocyte activation, the Activation Induced Cytosine Deaminase (AID) actively deaminates DNA [84], giving rise to G/U mismatches. These appear to be processed by UNG2 and the BER pathway and occasionally by the MMR system. According to the model, uracil removal would generate AP-sites that induce mutations during DNA replication or repair upon the action of error-prone polymerases [85]. Nevertheless, neither hypermutation at the variable Ig region nor CSR are completely defective in *UNG* deficient mice, indicating the contribution of other UDGs. In contrast to SMUG1, which is downregulated following B-cell activation [86], TDG is upregulated upon LPS treatment of naïve B-cells (C. Kunz, unpublished observations), making TDG a plausible candidate. Taking into account the slow turnover properties of the enzyme, a model in which TDG shields the AP-site from being processed until the next round of DNA replication involving error-prone DNA synthesis seems plausible.

TDG and the Regulation of Gene Expression

TDG was identified as a c-Jun-interacting protein by Yeast-Two-Hybrid screening almost 20 years ago [54]. This was the first indication of a function of TDG in the regulation of gene transcription. Later, TDG was found to interact with the nuclear receptors RAR and RXR through its catalytic domain in a ligand-independent manner. In addition, the interaction was shown to enhance binding of the receptors to their retinoic acid response elements (RAREs), potentiating nuclear receptor mediated transactivation of reporter genes in co-transfection experiments [3]. Interestingly, an active site mutant of TDG failed to stimulate RAR/RXR mediated transcription, suggesting an involvement of the glycosylase function.

Human TDG was also shown to interact with Estrogen Receptor alpha (ERα) *in vivo* and *in vitro*, another member of the nuclear receptor family, and to stimulate ERα activity in a ligand-dependent manner in reporter gene assays [5]. Interestingly, and in contrast to the stimulation of RXR activity, transcriptional activation did not require a functional TDG catalytic domain. However, results from other studies favor a role for TDG in DNA repair following site specific deamination of methylated cytosine in ERα responsive gene promoters [6, 87]. One explanation for these apparent discrepancies is that, depending on whether a promoter requires demethylation or not, TDG might act as a DNA glycosylase and/or a scaffold protein in gene regulatory protein complexes. TDG also acts with other transcriptional coactivators in a cooperative manner. It associates with SRC1, a co-activator of ERα, to form a complex that is recruited to estrogen-responsive gene promoters and to activate ERα-mediated gene expression in the absence of ligand [4].

Furthermore, Pierre Chambon's laboratory demonstrated that TDG associates with the transcriptional coactivators CREB binding protein (CBP) and its paralogue p300, which act as coactivators for a number of sequence-specific transcription factors and activate RNA polymerase II-mediated transcription through chromatin remodeling and interactions with the basal transcription machinery. *In vitro,* CBP-TDG complexes were demonstrated to recognize and process G/T and G/U mispaires and to acetylate histone tails. Moreover, CBP-dependent transcription was potentiated by the presence of TDG. In this case the glycosylase activity was dispensable for

transcriptional activation [50]. These results suggest that TDG might be recruited to promoter regions by transcription factors, where local chromatin remodeling involving CBP/p300 would facilitate G/T repair of deaminated 5-meC prior to transcriptional activation.

In two cases, TDG was reported to act as a repressor of transcription. Missero et al. showed an interaction between TDG and the thyroid transcription factor 1 (TTF1) [49]. The mechanism through which TDG represses TTF-1 mediated gene expression and whether its glycosylase activity is required for repression was not elucidated. In a recent report, TDG was shown to interact with the transcription factor Myocardin, a co-activator of SRF mediated transcriptional activation of smooth muscle-specific genes [2]. By disrupting the formation of myocardin/SRF complexes TDG was found to repress myocardin binding to target promoters of smooth muscle-specific genes. Furthermore, repression of mycardin activity was shown to be independent on TDG's glycosylase activity.

In addition to the interactions described above, several other transcription factors have been shown to associate with TDG [1]. Figure 4 summarizes the most important TDG interactions including those with DNA repair (XPC, XPG, XRCC1, APE1) and DNA damage signaling proteins (Rad9-Rad1-Hus1 complex), which will not further detailed here, but see [51, 55].

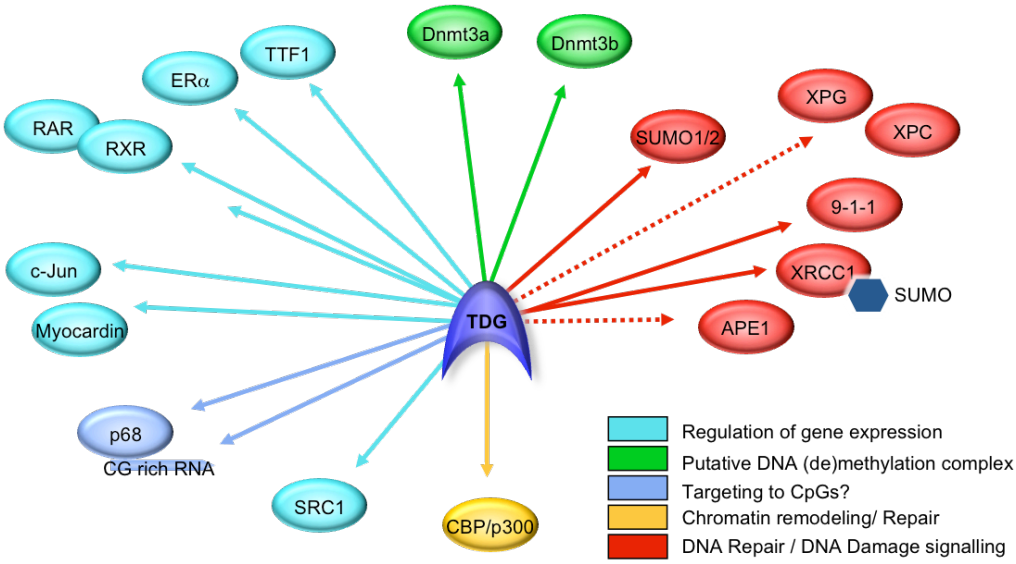


Figure 4. The TDG Interaction network. Shown are most important known TDG interaction partners based on literature search. TDG interacts with DNA Repair and DNA damage signaling molecules (*red*), transcription factors and nuclear receptors (*blue*), DNA methylatransferases (*green*), the RNA helicase p68 and CG-rich RNA (*light blue*) and chromatin modifying factors (*yellow*). The putative function of

TDG in the context of these interactions is indicated in the colour-legends. Details on the interactions are described in the text.

Under consideration of the enzymatic properties of TDG, these interactions indicate that TDG might be recruited to regulatory regions of actively expressed genes to interrogate and maintain the integrity of such sequences in order to prevent aberrant gene regulation.

TDG and CpG (De)Methylation

Genomic DNA methylation patterns change during biological processes, such as fertilization, differentiation and transformation. Hence, proper mammalian embryonic development and the maintenance of cell identity require a tight control of DNA methylation and possibly demethylation [7, 88, 89]. While the enzymology of DNA methylation is well established [90], the mechanism of DNA demethylation is still a matter or debate. First evidence for an active DNA demethylating process came from the laboratory of Jean-Pierre Jost. They isolated an activity from nuclear extracts of chicken embryos and mouse G8 myoblasts capable of demethylating 5-meC in an artificial hemimethylated DNA substrate [67, 91]. The purification of this activity led to the isolation of a 5-meC DNA glycosylase (5-MCDG) activity [92] that also processed T in G/T mismatches and was finally identified as the chicken orthologue of the human TDG. However, the recombinant 5-MCDG/TDG produced in *E.coli* showed only marginal activity on 5-meC [68], suggesting that additional cofactors are needed to constitute physiologically relevant cytosine demethylation activity. In further studies, the 5-meC and G/T mismatch specific DNA glycosylase activities purified from chicken embryos were shown to be associated with CpG rich RNA. RNA complementary to the methylated DNA strand of a synthetic hemimethylated DNA substrate was shown to be able to restore 5-meC DNA glycosylase activity in a previously RNAse treated preparation [93, 94]. Therefore, an RNA component of the glycosylase was suggested to facilitate targeting of the enzyme to sites were specific demethylation is needed (e.g. gene regulatory regions) and/or to stimulate its enzymatic activity. In addition, the purified DNA

demethylation complex also contained a DEAD box protein related to the human p68 RNA helicase [10]. The p68 helicase, which has established roles in transcription and RNA processing [95], might be responsible for the reorganization of the secondary structure of the CpG-rich RNA to make it available for the DNA glycosylase, in a way that the demethylation activity is targeted to a specific DNA sequence [10].

To corroborate the biological relevance of 5-MCDG/TDG for active DNA demethylation the group of Jean-Pierre Jost showed that overexpression of human TDG in human embryonic kidney cells stimulated the expression of a reporter gene controlled by an ecdysone-retinoic acid responsive enhancer-promoter element. Interestingly, specific demethylation was observed at the promote-enhancer region [69]. This results suggested that transcription factors might be involved in the targeting of the glycosylase to specific sites, i.e. CpG dinucleotides at promoters of active genes, that need to be demethylated or protected from *de novo* methylation.

Two recent reports demonstrate that cycles of DNA methylation and demethylation do occur upon transcriptional activation of gene promoters [6, 96]. Surprisingly, these studies propose the involvement of Dnmt3a and Dnmt3b in both DNA methylation and demethylation. By looking at specific CpG sites in the promoter of the estrogen-responsive gene *pS2 (trefoil factor 1; TFF1)* in human breast cancer cells, the authors found that the *pS2* promoter undergoes cyclic DNA de- and remethylation upon activation by ERα. Loss of methylation at the beginning of each transcriptionally productive cycle coincided with recruitment of Dnmt3a, Dnmt3b, TDG, APE1, DNA polymerase β and DNA ligase I to the promoter. Métivier et al. further showed that limiting levels of the methyl donor S-adenosyl methionine (SAM) induce Dnmt3a and Dnmt3b to deaminate C and 5-meC *in vitro,* and that TDG is able to process the resulting G/U and G/T mispaires, respectively [6]. Consistently, TDG was shown to interact with both Dnmt3a and Dnmt3b [52, 53]. This interaction, however, has initially been proposed to facilitate efficient remethylation of cytosines incorporated during repair of G/T mismatches resulting from spontaneous deamination of 5-meC, thereby allowing the restoration of the original methylation pattern. Finally, supporting an involvement in a putative “demethylation complex”, the p68 RNA helicase, a coactivator of ERα [97] and interaction partner of TDG[10], was also localized to the *pS2* promoter.

The observed but still putative function of TDG mediated BER in active cytosine demethylation implicates a contribution to the maintenance of CpG methylation states at specific genomic sites to which TDG is targeted through the interactions with transcription factors. Hence, in this way TDG might ensure accurate gene expression.

TDG in Embryonic Development

Our Laboratory recently showed that homozygous *Tdg* knockout embryos lose viability at midgestation (Primo Schär et al, Tetsuya Ono et al., manuscripts in preparation). Given that other glycosylases are dispensable for embryogenesis [1], this phenotype suggested that the non-redundant and essential function of TDG is distinct from classical DNA repair and may relate to its involvement in gene regulation. In this context, the cooperation of TDG with transcription factors that regulate developmental processes (e.g. RAR/RXR) might implicate a function in the establishment of cell-lineage specific gene expression patterns, possibly by controlling CpG methylation of gene regulatory sequences.

2.3 Epigenetic regulation of gene expression

Eukaryotic DNA is accommodated within the cellular nucleus where it is present as a nucleoprotein complex known as chromatin. The fundamental structural unit of the chromatin, the nucleosome core particle, is composed of 147 base pairs (bp) of DNA wrapped in approximately two superhelical turns around a histone octamer consisting of two copies of each of the histone proteins H2A, H2B, H3 and H4. Strings of thousands of nucleosome core particles joint by linker DNA and stabilized by histone H1 are then further packaged into higher order chromatin structures that remain poorly defined [98]

In higher eukaryotes, chromatin is subdivided into structurally distinct states known as hetero- and euchromatin. Heterochromatin has a compact and rigid conformation and mostly comprises transcriptionally inactive repetitive sequences and only a few active genes. In contrast, euchromatin assumes a more open

structure with less compact nucleosome architectures and is relatively gene rich and transcriptionally active. Today we know that the chromatin structure does more than simple DNA packaging. It constitutes a natural barrier limiting the accessibility of macromolecular protein complexes to particular DNA segments. As a result, mechanisms that dynamically modulate the chromatin structure, including chromatin remodeling, covalent histone modifications, utilization of histone variants and DNA methylation are therefore key components in the regulation of genetic processes such as DNA replication, DNA recombination, DNA repair and spatially and temporally coordinated gene expression [99].

2.3.1 Histone Modifications

Chromatin structure is determined and regulated by covalent modifications of histones, that altogether form a "histone code" [100]. The core histones consist of a globular domain and charged NH_2-terminal "tails", which span 19-25 amino acid residues. Both, globular domains and histone tails are subject to different posttranslational modifications such as acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP-ribosylation and proline isomerization [101]. Most of these modifications appear to be reversible. Modification and demodification of histones are specific enzymatic processes involving histone acetylases (e.g. CBP/p300) and deacetylases (e.g. SirT2), histone methyltransferases (e.g. G9a) and demethylases (e.g. LSD) in addition to protein kinases (e.g. Mst1) and ubiquitin ligases (e.g. Ring1A). Table 3 summarizes the most important histone modifications and their effect on gene activity. It is believed that the covalent histone modifications, alone or in combination, constitute a "histone code" for each nucleosome that acts as a scaffold for the recruitment of other histone modifiers, chromatin remodeling complexes, transcription factors and factors that participate in certain DNA-templated processes. Through specific recognition domains (e.g. Bromo-, chromo- and 14-3-3 domains), these specifically recognize particular histone modifications (Fig. 5A) [102-104].

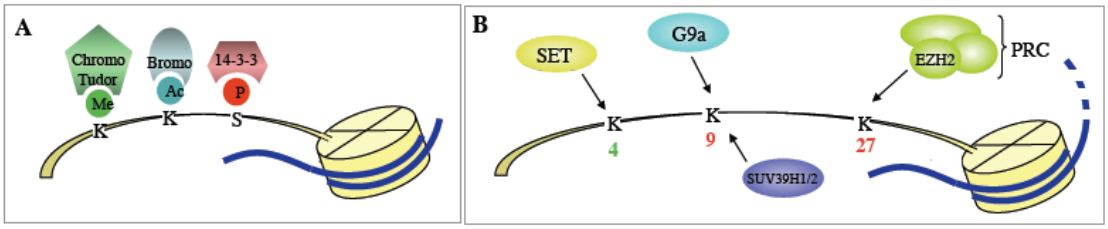


Figure 5. Modification of histone tails and protein recruitment to histones. A) Histone tail modifications such as acetylation, methylation or phosphorylation are recognized by chromatin remodeling complexes or transcription factors via specific recognition domains, such as Chromo, Tudor, Bromo and 14-3-3 domains. **B)** Important histone methylation marks on H3 lysine (K) residues involved in regulation of gene expression and maintenance of chromatin structure. SET domain proteins and G9a are involved in H3K4 and H3K9 methylation, respectively. Polycomb group repressing complexes (PRC) have been shown to be involved in methylation of H3 at K27, thereby reversibly silencing a number of developmental gens.
(K=Lysine, S=Serine; Me=Methylation, Ac=Acetylation, P= Phosphorylation. Green numbers="active marks", red numbers="repressive marks. See text for Details) Inspired by [102, 105]

For the purpose of this thesis, I will focus the description of histone modifications on those that are known to be functionally associated with DNA Repair and the regulation of gene expression. Details on additional histone modifications and their biological outcomes are provided in [102] and Table 3.

Acetylation within the N-terminal domains of histones H3 (e.g. K9, 14, 18), H4 (e.g. K5, 8, 12, 16), H2A (e.g. K5) and H2B (e.g. K12, 15) is the best-described type of modification linked to transcriptional regulation. It is believed that acetylation induces changes in nucleosome structure through charge neutralization of the histone tail. This modification has been primarily associated with transcriptional activation, while histone deacetylation is associated with gene repression. Therefore, the activity status of chromatin is determined by the opposing effects of histone acetyltransferases (HATs) and histone deacetylases (HDACs) [105].

Table 3. Mammalian histone modifications and their effect on transcription [102]

Histone modifications	Residues Modified	Effect on transcription
Acetylation		
HAT1	H4(K5/12)	Activation
CBP/p300	**H3(K14/18), H4(K5/8)**	**Activation**
	H2A (K5), H2B (K12,K15)	Activation
PCAF/GCN5	**H3 (K9/14/18)**	**Activation**
TIP60	H4 (K5/8/12/16), H3K14	Activation
HB01	H4 (K5/8/12/16)	Activation
Deacetylation		
SirT2	H4K16	Repression
Lysine methylation		
SUV39H1/H2	**H3K9**	**Repression**
G9a	**H3K9**	**Repression**
ESET/SETDB1	H3K9	Repression
GLP	H3K9	Repression
CLL8	H3K9	Repression
MLL1,2,3,4,5	H3K4	Activation
SET1B, SET1A	H3K4	Activation
ASH	H3K4	Activation
SET2	H3K36	Activation
NSD1	H3K4	Activation
SYMD2	H3K36	Activation
DOT1	H3K79	Activation
Pr-SET 7/8	H3K20	Repression
SUV4 20H1/H2	H3K20	Repression
EZH2	**H3K27**	**Repression**
RIZ1	H3K9	Repression
Lysine Demethylation		
LSD	H3K4	Repression
JHDM1a/1b	H3K36	Repression
JHDM2a/2b	H3K9	Activation
JMDJ2A/JMDJ3A	H3K9, H3K36	Repression
JMDJ2B	H3K9	Activation
JMDJ2C/GASC1	H3K9, H3K36	Repression
JMDJ2D	H3K9	Activation
Arginine Methylation		
CARM1	H3 (R2/17/26)	Activation
PRMT4	H4R3	Activation
PRMT5	H3R8, H4R3	Activation
Phosphorylation		
Haspin	H3T3	Activation
MSK1/2	H3S28	Activation
CKII	H4S1	Activation
Mst1	H2BS14	Activation
Ubiquitination		
Bmi/Ring1A	H2AK119	Repression
RNF20/49	H2BK120	Activation

In addition to target histone proteins, both HATs and HDACs have been shown to modify non-histone proteins including tumor suppressors (e.g. p53) (Ito et al. 2001), transcription factors (e.g. E2F, pRB) (Martinez-Balbas et al. 2000) and DNA repair proteins (e.g. TDG) [50]. Depending on which residue of the protein is acetylated the modification can modulate enzymatic properties of modified proteins (e.g. DNA

binding) influence protein stability, or regulate protein-protein interactions [106, 107].

Conjugation of methyl groups takes place at lysine (K) and arginine (R) residues of histone tails. Arginine methylation will not be discussed here and can be found elsewhere [108]. Methylation of lysine (K) residues is catalyzed by specific histone methyltransferases (HMTases) and occurs on histones H3 (K4, K9, K27, K36 and K79) and H4 (K20) ([109] and Table 3). Each of these lysine residues can be mono-, di-, or trimethylated, which appears to define distinct chromatin states, i.e silent or active chromatin, depending on the lysine modified (see below).

In vertebrates, methylation of H3K4 is associated with the 5′ region of actively transcribed genes and genes poised to be expressed in an open chromatin state [110]. In mammals, there are at least ten known or predicted HMTases for H3K4 and they are all characterized by the presence of a SET (Su(var)3-9, Enhancer of Zeste, Trithorax) domain, which is responsible for the transfer of the methyl-group to the target histone (Table 3 and Fig. 5B) [111]. During active transcription, H3K4 methylation correlates with histone acetylation (e.g. H3K9) and RNA polymerase II occupancy at gene promoters. In addition, H3K4 methylation was shown to prevent binding of negatively acting chromatin-remodeling complexes [112-114].

By contrast, methylation of H3 at lysine 9 (H3K9) has been associated with gene silencing. The main HMTases responsible for K9 methylation are the SET-domain containing HMTases SUV39H1 SUV39H2 and G9a (Fig. 5B) [115, 116]. H3K9 methylation was shown to be enriched at pericentrometic heterochromatin (H3K9me3) and at silent euchromatic regions (H3K9me, H3K9me2) [117]. The recognition of H3K9 methylation by the chromodomain containing HP1 proteins has been shown to be crucial for the establishment of a ‘silent’ chromatin state [118]. K9 methylation can also suppress transcription independently of HP1 through a mechanism involving the inhibition of histone acetylation by p300 (see Table 3) [119].

Another silencing histone modification is the methylation of H3K27. Methylated H3K27 was found to be associated with the binding of Polycomb group (PcG) proteins, initially discovered in the fruit fly *drosophila melanogaster*. PcG proteins are involved in maintaining pluripotency and cells identity via repression of

key developmental regulators in embryos and ES cells [120, 121]. Biochemical purification experiments identified two multimeric PcG complexes or polycomb repressor complexes (PRCs) with distinct enzymatic activities. Binding of the PRC2 initiation complex (formed by the proteins enhancer of zeste homologue 2 (EZH2), embryonic ectoderm development (EED) and suppressor of zeste 12 (SUZ12)) to the polycomb group target genes induces EZH2-mediated H3K27 di/tri-methylation. K27 trimethylation is followed by the recruitment of other repressive complexes, such as the PRC1 complex (comprising the core components polycomb (PC), posterior sex combs (PSC) and sex combs extra (SCE/RING)). PRC1 is thought to cooperate with PRC2 to mediate repression by inhibiting chromatin remodeling, blocking transcription and/or by mediating chromatin compaction [122]. Furthermore, recent reports suggest interplay between the polycomb pathway and DNA methylation. EZH2, as part of the PRC2 complex, was shown to recruit DNA methyltransferases (DNMTs) to selected target genes in cancer cell lines ([123] and chapter "DNA methylation"). In addition, several studies suggested preferential aberrant DNA methylation in human cancer cell lines and primary cancers at promoters of polycomb target genes in unrelated human ES cells in culture [124, 125].

2.3.2 Chromatin Remodeling

A major class of chromatin regulators are the chromatin remodeling complexes. These complexes are composed of diverse proteins that contain a catalytic ATPase subunit. In an ATP-dependent manner, these remodeling machines rearrange histone-DNA contacts resulting in nucleosome mobilization along the DNA, histone displacement and exchange. These changes modulate the accessibility of factors involved in DNA templated processes (e.g, DNA or RNA Polymerases, transcription factors, DNA repair proteins) to nucleosomal DNA [126].

Several remodeling multisubunit complexes have been identified. Based on the distinct domain structures of the ATPase subunits, these have been subdivided into four main classes (Fig. 6); the SWI/SNF (switching defective/sucrose non-fermenting), the ISWI (imitation SWI), the NuRD (nucleosome remodeling and deacetylation)/Mi-2/CHD (chromodomain, helicase, DNA binding) and the INO80

(inositol requiring 80) families of chromatin remodelers [126]. These complexes were shown to recognize and bind unmodified as well as acetylated, phosphorylated and methylated histone tails (See Histone Modifications) and to bind DNA directly through specific DNA recognition domains. [127-129].

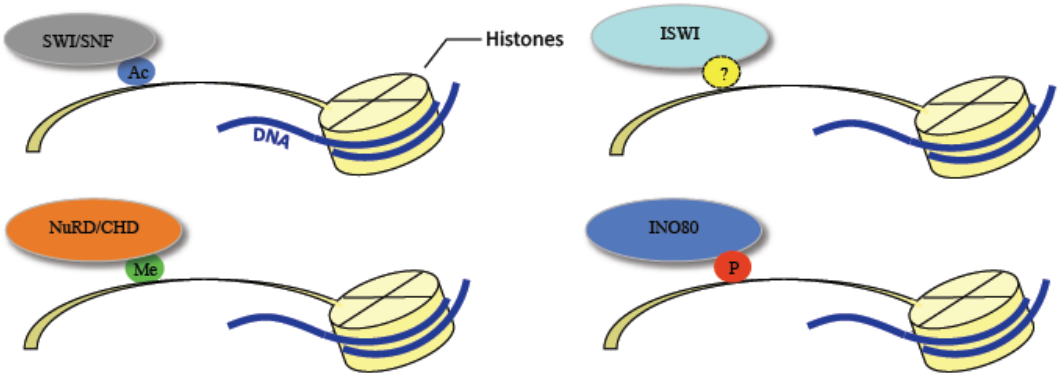


Figure 6. ATP-dependent chromatin remodelling complexes. The four best-characterized ATP-dependent chromatin remodeling complexes in yeast and mammalian systems interact with specific chromatin substrates (histone marks) to restructure and mobilize nucleosomes. *Ac, histone acetylation; Me, histone methylation; P, histone phosphorylation* [126].

Chromatin binding of these complexes was found to promote both gene activation and silencing and the precise outcome of their action seems to be dependent on a particular chromatin context, i.e. the presence of activation or silencing histone modifications, the level of DNA methylation at gene promoters and interactions with specific transcription factors [130-132]. In addition, ATP-dependent remodeling complexes carry out specialized roles in cellular processes other than gene regulation including DNA-damage repair [133, 134], DNA replication, chromosome condensation and segregation, and the maintenance of pluripotency in embryonic stem (ES) cells [126].

2.3.3 DNA Methylation

DNA methylation: the mark and its mediators

DNA methylation occurs at the 5' position of cytosines in a CpG sequence context. Depending on the species and tissues examined, this modification affects about 60-90% of CpG dinucleotides. In total, 5-methylcytosine represents about 1-5% of the

mammalian DNA bases [7].

DNA methylation in mammals is catalyzed by three DNA methyltransferases (DNMTs), DNMT1, DNMT3a and DNMT3b, which use S-adenosylmethionine (SAM) as a methyl donor [7]. DNMT1 is responsible for propagation and stable maintenace of DNA methylation patterns through mitosis [135]. On the other hand, DNMT3a and DNMT3b are highly expressed in early embryonic cells and are responsible for the *de novo* methylation of unmethylated CpG sites [136]. Genetic studies revealed that DNA methylation is essential for normal development. Mice devoid of DNMT1 die at 8.5 days post coitus whre the phenotype was mostly attributed to loss of imprinting and reactivation of transposons and other repetitive elements [137]. Also, whereas *Dnmt3a*$^{-/-}$ knockout mice die within a few weeks after birth, the *Dnmt3b*$^{-/-}$ knockout is more severe and emrbyos die around 9.5 days post coitus [138]. An additional protein, DNMT3L, which lacks intrinsic methyltransferase activity, was shown to physically associate with DNMT3a and DNMT3b and to modulate their catalytic activity [139, 140]. Furthermore, Dnmt3L is able to recruit Dnmt3a to unmethylated genomic regions via a direct interaction with unmethylated H3K4 [141]. Interestingly, mouse DNMT3L seems to be essential for establishing paternal and maternal methylation imprints and was reported to be involved in spermatogenesis [142, 143]. A fourth DNA methyltransferase, DNMT2, shows minimal DNA methyltransferase activity *in vitro* [144] and homozygous gene deletion has no obvious effects in mice or ES cells, suggesting either a different or more spezialized function for DNMT2 compared to the other DNMTs [145]. However, recent reports identified DNMT2 as RNA methyltransferase specific to aspartic acid t-RNA [146, 147].

DNA methylation behaves like a clonally inherited repressive epigenetic mark in mammalian cells and is therefore considered an epigenetic DNA modification [7]. Two basic mechanisms of CpG methylation dependent gene repression have been proposed: (I) The presence of a methyl group at the 5’ position of cytosines may directly inhibit binding of transcriptional activators to their cognate DNA sequences; (II) methylated DNA is recognized by methyl-CpG-binding proteins (MBPs), which in turn recruit histone modifying activities that act as co-repressors of transcription. Two major families of MBPs are known to date. The MBD family proteins, MeCP2,

MBD1, MBD2, MBD3 and MBD4, which share an N-terminal methyl-binding domain (MBD), and the Kaiso family proteins (Kaiso, ZBTB4, ZBTB38). MeCP2 and MBD1-3 were shown to mediate transcriptional repression and heterochromatin formation through the interaction with ATP dependent chromatin remodeling complexes, histone deacetylases (HDACs) and histone methyltransferases (HMTs) [148-150], thereby coupling DNA methylation to histone modifications and chromatin remodeling [79, 151]. Interestingly, MBD4 contains a C-terminal DNA glycosylase catalytic domain in addition to an N-terminal MBD and therefore appears to link DNA methylation with DNA repair. However, MBD4 was reported to contribute to gene silencing of CpG methylated promoters and to directly bind to the transcriptional repressor Sin3a and HDAC1. Similarly, the MBP Kaiso has been identified as part of the HDAC-containing transcriptional repressor complex [152].

In genomes of mammals and other species that undergo extensive DNA methylation, CpG dinucleotides are largely depleted, except at short genomic regions called *CpG islands*. Such CpG islands often locate in gene promoters and are mostly unmethylated (see below). It is thus generally accepted that the underrepresentation of CpG outside of CpG islands results from the inherent mutability of methylated cytosines. 5meC deaminates at a higher rate than unmethylated cytosine [153] generating G/T mispairs that, unless repaired, give rise to C to T transition mutations following DNA replication. Surprisingly, methylated CpGs are C to T mutation hotspots in mammalian genomes given the presence of two DNA repair enzymes, TDG and MBD4, that are able to recongnize and initiate repair of deamination induced G/T mismatches. One possible explanation for this discrepancy could be the fact that G/T glycosylases do not act globally in the genome but are targeted to specific sites [1]. Yet, there is growing evidence for methylated CpGs being vulnerable to other endogenous or exogenous mutagens that give rise to lesions that are not substrate of TDG and/or MBD4, leading equally to a decline of CpG dinucleotides in our genomes [154].

Biological Functions of Mammalian DNA Methylation

As an epigenetic mark, DNA methylation plays a substantial role in many vital cellular processes, such as DNA replication and repair, chromosomal organization and stability, genomic imprinting and X-chromosome inactivation [7, 89, 155, 156]. The importance of DNA methylation was established in mice and frogs, where targeted disruption of DNA methyltransferase genes resulted in developmental defects [138, 157-159]. Also, misregulation of DNA methylation has been linked to a number of different human diseases including cancer [7].

Mammalian development is associated with considerable changes in global DNA methylation. Specific developmental periods are characterized by genome wide resetting of DNA methylation. Once established, the patterns of DNA methylation have to be maintained to avoid loss of cell identity and tissue homeostasis. Thus, to assure proper fetal development and to maintain tissue homeostasis, DNA methylation and other epigenetic modifications have to be tightly regulated and controlled [89].

Genomic imprinting: Genomic imprinting is a developmental process, affecting a small subset of genes and resulting in paternal origin-specific mono-allelic expression. Imprinting marks are established during the late stages of germ cell differentiation and are then maintained after fertilization and throughout embryonic development.

Imprinted genes are often found clustered in the genome, allowing cluster-wide expression of paternal or maternal alleles, controlled by *cis*-regulatory elements called imprinting control regions (ICRs) or differentially methylated regions (DMRs) [160]. DMRs are usually located in CpG rich regions and often fulfill the criteria for CpG Islands [161]. CpGs of DMRs become methylated during germ cell differentiation by the *de-novo* Dnmt's, that appear to be targeted by Dnmt3L, which itself is recruited to these regions by histone arginine methylation [162, 163].

After oocyte fertilization, most of the genomes methylation is erased. The parental genome is actively demethylated (see 1.3.4) in the zygote, whereas the maternal genome undergoes a passive demethylation during cell divisions due to absence of Dnmt1 in the nucleus until the blastocyst stage, where methylation is re-established in a cell lineage dependent matter (Fig.6). General demethylation is important for

resetting the gamete genomes to a pluripotent state, a prerequisite for the establishment of the inner-cell-mass lineage in the blastocyst [155]. Strikingly, imprinted genes are protected from this epigenetic reprogramming process, and is has been suggested that the protein PGC7/*Stella* acts to protect imprinted loci from DNA demethylation at this stage [164]. Exactly how this is achieved is still obscure. Maintenance of DNA methylation imprints throughout pre-implantation development was shown to involve maternal and zygotic Dnmt1 isoforms [165].

In Primordial Germ Cells (PGCs), parental epigenetic imprints are widely erased. Current evidence indicates that both DNA methylation aswell as H3K9 methylation, accompanied by gain of H3K27 methylation are necessary for the establishment of new imprints at later stages (Fig.6) so that that in mature gametes they reflect the sex of the germline [166].

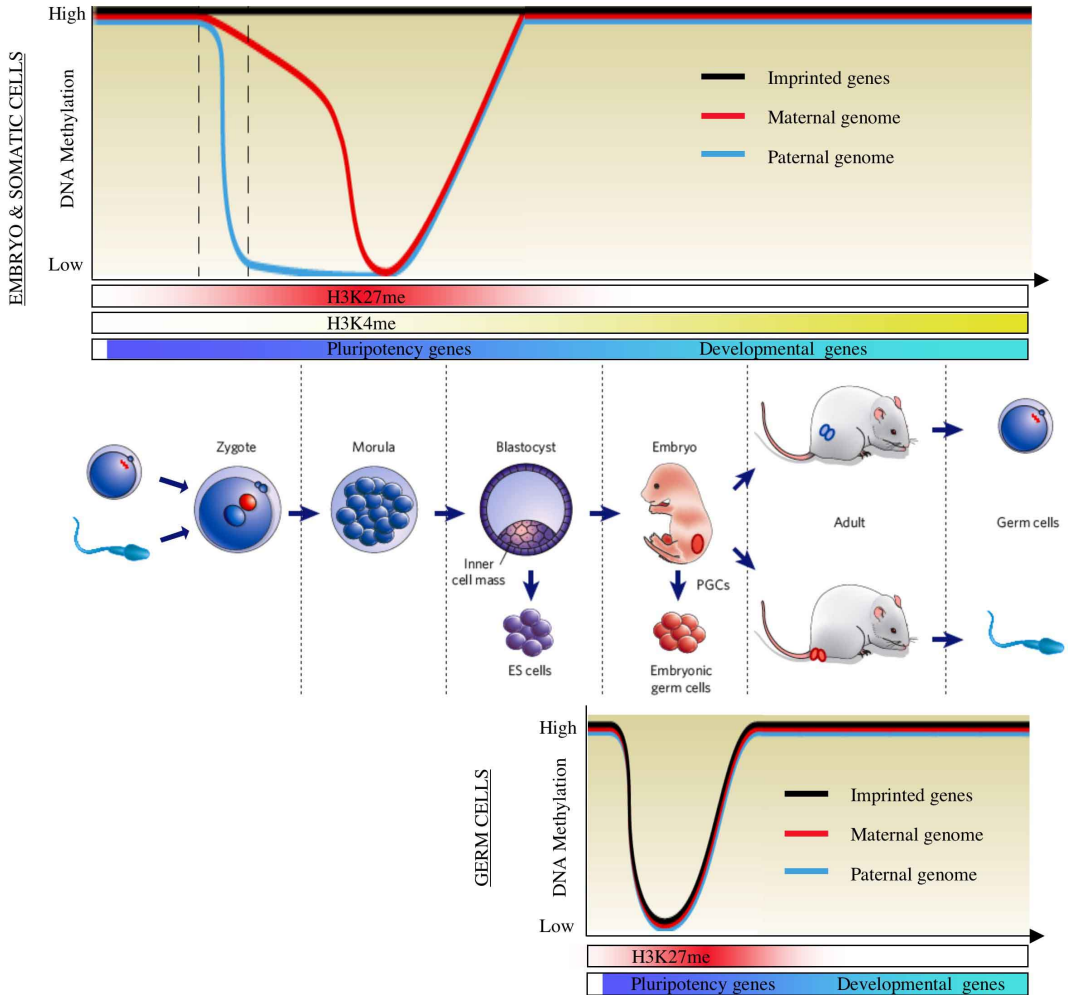


Figure 6. Epigenetic gene regulation during mammalian development. Important histone modifications, DNA methylation changes and gene expression patterns are shown along key developmental events in mammals. Early in development, DNA methylation is erased (except for imprinted genes), in accordance with the expression of pluripotency genes and repression of developmental genes. During differentiation of ES cells, pluripotency genes are silenced by DNA methylation and developmental genes become expressed, in concordance with H3K4 methylation and H3K27 demethylation. Similarly, during embryonic germ cell development, pluripotency genes become activated after DNA demethylation. Polycomb repressed genes become activated during germ cell development in the adult. At this stage, an increase in DNA methylation allows stable silencing of imprinted and pluripotency associated genes. *Adapted from Reik, 2007* and Reik 2001.

The mechanism of active DNA demethylation in PGCs as well as in the zygote is still unknown. However, as discussed below in the next section, the process might involve deamination of 5-methylcytosine to thymine followed by repair of the resulting G/T mismatch.

X-Chromosome inactivation: Compensation of X-linked gene dosage between XY male and XX female mammals occurs by random X chromosome inactivation in females. X-inactivation (XCI) comprises two processes; the induction of a silent state and the subsequent stable maintenance of the silent state. XCI is dependent on the X-inactivation center (Xic) and is initiated by the expression of *Xist* (X-inactive-specific transcript), which then acts to trigger events that lead to heterochromatinization [167]. The molecular mechanism is not clear; yet, it seems to involve coating of the chromosome with *Xist* that excludes RNA polymerase II from binding. Following *Xist* RNA accumulation on the inactive X, loss of active chromatin marks (H3K4 methylation, H3K9 acetylation), and appearance of repressive H3K9 methylation and H3K27 tri-methylation have been reported [168-170]. In addition, DNA methylation of CpG islands in inactive genes on the X-chromosome was shown to play an important role in the stabilization of the inactive state, at least in embryonic tissue. By contrast, Polycomb mediated silencing might dominate in extra-embryonic tissues [155, 169].

Heterochromatin formation: Genomes of most eukaryotes are structurally organized in two distinct chromatin compartments; euchromatin and heterochromatin. Heterochromatin comprises pericentromeric and telomeric regions, is enriched for repetitive DNA sequences including highly repeated tandem satellite sequences and

transposable elements (e.g. SINEs and LINEs). Heterochromatin is essential for chromosome segregation, centromere function and telomere protection [171].
Molecular characterization has led to the identification of several characteristic features of heterochromatin. Mouse heterochromatin shows hypoacetylated histones, extensive enrichment of H3K9 di- and trimethylation, which appears to be important for spreading and maintenance of the heterochromatic state [118], as well as H3K27 methylation [172, 173]. DNA cytosine methylation is a feature of constitutive heterochromatin, and appears to be required for both, the formation and maintenance of the heterochromatinized state. Mutations in the PWWP (Proline-tryptophan-tryptophan-proline) domain of *de novo* DNA methyltransferases, which is important for DNA binding *in vitro*, were shown to abolish enzyme targeting and subsequent DNA methylation at centromeric and pericentromeric satellite repeats in mouse ES cells and fibroblast [174].

Both histone modifications and DNA methylation contribute to genome integrity by ensuring a compact chromatin structure, preventing aberrant DNA replication and repair, processes that would lead to chromosomal instability [175].

Silencing of transposable elements: Transposable Elements (TEs) of various classes constitute a large fraction of eukaryotic genomes [176]. Active TEs are highly mutagenic due to their ability to integrate into protein coding regions or to induce chromosomal breakage, illegitimate recombination and genome rearrangement events. In addition, TEs can also influence the expression of flanking genes by affecting not only splicing and polyadenylation events or by acting as enhancers or promoters [177]. Small RNA mediated silencing and a number of epigenetic modifications including chromatin remodeling, histone modifications and DNA methylation were shown to suppress TE transcription [178]. Elevated transposon transcripts were reported in DNMT1-/- embryos, and DNMT3a and Dnmt3L were shown to be involved in TE silencing in male germ cells [179, 180]. DNA CpG methylation thus contributes to epigenetic silencing and stabilization of TEs in the mammalian genome during development and presumably also in the adult tissue [177, 178, 181].

2.3.4 DNA Demethylation

In mammals, DNA demethylation can occur as a genome-wide event during gametogenesis and early embryogenesis (Fig. 6) or in a regulated fashion to control the activity of individual genes [6, 155]. Demethylation of DNA can be achieved either by passive or active mechanisms, or a combination of both. Whereas passive DNA demethylation occurs upon inhibition or loss of maintenance methylation during DNA replication, active demethylation requires defined and targeted enzymatic reactions. Four mechanisms for active demethylation have been postulated on the basis of experimental evidence. A first involves the direct enzymatic removal of the methyl group from 5-meC. Human MBD2 was reported to posses such activity [182], but these finding could not be corroborated in subsequent replication studies in different laboratories. Also, *Mbd2* knockout mice show normal DNA demethylation during development, demonstrating that MBD2 is not essential for global demethylation *in vivo* [89, 183].

A second model proposed an involvement of growth arrest and DNA damage inducible protein alpha (Gadd45a) in active DNA demethylation by a nucleotide excision repair-based mechanism [184, 185]. Other studies then reported that Gadd45a does not contribute significantly to active demethylation [186, 187], which would be in agreement with a mouse Gadd45a phenotype that does not show developmental defects nor aberrant CpG methylation [186]. More recently though, Rai et al. provided evidence for a function of Gadd45a as a factor involved in DNA base excision repair mediated DNA demethylation in zebrafish embryos [188]. According to their model, active DNA demethylation may occur through a two-step process involving generation of a G/T mismatch by enzymatic deamination of 5-meC by AID (Activation induced Cytosine deaminase) coupled to base excision repair involving the DNA glycosylase MBD4. In this context, Gadd45a was shown to stimulate demethylation mediating an interaction between AID and MBD4. Whether or not this applies to mammals remains questionable, since neither Gadd45a nor MBD4 knockout mice show an epigenetic defect in development. Thus, if MBD4 is involved in active DNA cytosine demethylation, it must be restricted to specific genomic regions or be redundant with other activities that compensate for the lack

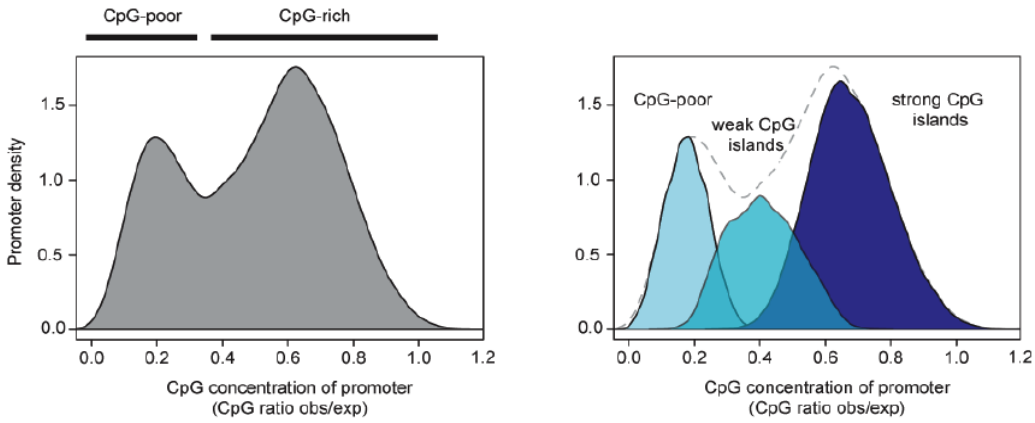
of MBD4 (e.g. TDG).

A recent article has provided evidence for DNA methylation regulated in mammalian cells by enzymatic conversion of 5-meC to 5-hydroxymethylcytosine (5-hmeC) both *in vitro* and *in vivo* [189]. Using computational and experimental analyses, Tahiliani and colleagues identified the 'ten-eleven translocation-1' (TET1) protein as an enzyme capable of catalyzing this hydroxylation reaction. Furthermore, 5-hmeC seems to be a physiological constituent of ES cell DNA with levels decreasing upon RNAi knockdown of TET1 [189]. These findings raise the possibility that hydroxylation of 5-meC is an epigenetic mechanism, either as an end product or as an intermediate in active DNA demethylation by DNA repair. Whether 5-hmeC is important for epigenetic gene regulation is not clear yet. However, it is striking that nuclei from Purkinje cells, which lack heterochromatin show significant levels of 5-methylcytosines conversion to 5-hmeC [190].

2.3.5 Epigenetic Landscapes in Embryonic Stem Cells and Differentiated Cells.

Based on the CpG ratio (obs/exp), the GC content and the length of the CpG-rich region, more than 15'000 gene promoter of the mouse genome have recently been classified into strong, weak and poor CpG Island promoters [191] (Fig. 7A). Further genome-wide surveys have looked at the DNA methylation status and the chromatin state of the different classes of promoters in embryonic stem (ES) cells and differentiated cells. These analyses studies revealed that strong, weak and poor CpG islands promoters are not only differentially methylated but also differentially regulated by active and repressive histone marks (Fig. 7B). This section will summarize the findings most relevant for this thesis only. [11, 192, 193].

A



B

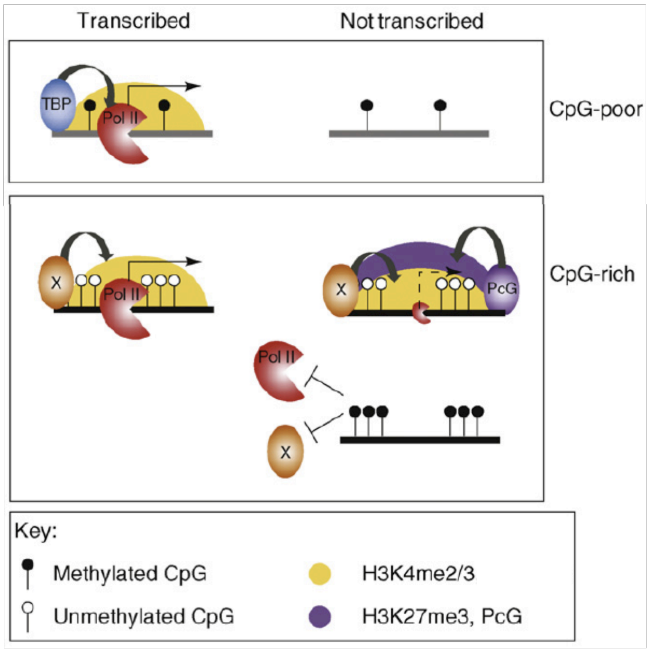


Figure 7. **A)** Promoter classification. Density plots showing the CpG ratio observed/expected for all 15.100 high-confidence promoters in the mouse genome. The total distribution (left) consisting of CpG-poor and CpG-rich promoters was split into 2 non-overlapping populations (right) of strong CpG islands (dark blue) and CpG-poor promoters (lightblue). The remaining intermediate group was classified as weak CpG island promoters (blue). *Taken from Mohn, 2008* [11]. **B)** Differential regulation of CpG-poor and CpG-rich promoters. CpG-rich promoters are mostly DNA unmethylated (white lollipops), decorated by H3K4me and H3/H4ac (*yellow*) and display low levels of Pol II (*red*) even when transcriptionally inactive. When repressed by Polycomb proteins (PcG) proteins (*purple*), the active environment persists, allowing transient repression. Upon DNA methylation (black lollipops), CpG-rich promoters lose active histone modifications, and become stably silenced. CpG-poor promoters only carry H3K4me and H3/H4ac marks when actively transcribed by Pol II. *Adapted from Mohn and Schubeler, 2009*

Strong and weak CpG island promoters (together referred to as CpG-rich), generally associated with pluripotency, housekeeping and key developmental genes, are predominantly free of CpG methylation and show H3K4 methylation in ES cells, irrespective of their transcriptional activity [11, 192]. Whereas promoters of active housekeeping and pluripotency genes are enriched for the active chromatin mark H3K4me2, promoters of developmental genes are found in bivalent chromatin domains (e.g. methylated at H3K4 and H3K27), meaning that they are generally silent but poised to be expressed. During cellular differentiation, when the pluripotency program must be stably repressed, promoters of pluripotency genes become *de novo* methylated and irreversibly silenced [11], whereas promoters regulating developmental genes become either activated after losing the repressive H3K27 methylation mark or else irreversibly silenced through DNA methylation, depending on the lineage choice.

CpG poor promoters, that control lineage- and tissue specific genes, appear to be regulated differently. They are found predominantly in a CpG methylated state in ES cells and differentiated cells, even if when the respective genes are not being actively transcribed [11]. Yet, when transcribed, CpG poor promoters present elevated levels of H3K4 methylation. Promoters of this class are most frequently devoid of the repressive H3K27 methylation mark [194]. At the same time, H3K9 methylation and heterochromatin protein 1 (HP1), which binds methylated H3K9, seem to repress some CpG poor promoters in ES cells and differentiated cells, and this may initially trigger DNA methylation [193, 195]. The CpG poor promoter of the pluripotency genes *Oct4* is a prominent example of differentiation dependent repression through H3K9 methylation and HP1 binding and subsequent CpG methylation [195]. Interestingly, a subset of CpG poor promoters were shown to become CpG demethylated and expressed during lineage commitment, suggesting that DNA demethylation of CpG poor promoters can occur in a tissue specific context [11]

2.3.6 Factors Regulating Pluripotency in ES cells

Embryonic development and cell fate determination require a strict coordination of epigenetic and genetic programs that together establish cell-type specific readouts of the genome. Genome reprogramming after fertilization is concomitant with embryonic genome activation that takes place during the early cell division stages of the fertilized egg and establishes the expression of distinct sets of genes during preimplantation development. The blastocyst, composed of an outer epithelial trophectoderm (TE) and the inner cell mass (ICM), represents the earliest differentiation stage in mammalian embryogenesis. The differentiation potential of TE cells is restricted to the trophoblast lineage. By contrast, cells of the ICM are able to differentiate to all other cell lineages of the organism and are therefore considered pluripotent. Embryonic stem (ES) cells are derived from the ICM and are excluded from the trophoblast lineage. Conversely, trophoblast stem (TS) cells are excluded from embryonic tissues.

Embryonic cell lineage specification is established in part by core pluripotency transcription factors (e.g. Oct4, Nanog, Sox2, Klf4, c-Myc) [196, 197]. Multiple physical and functional interactions between these proteins and gene regulatory regions indicate extensive co-regulatory mechanisms in the maintenance of ES cells (Fig. 8). For instance gene promoters that are bound by a small set of pluripotency transcription factors tend to be inactive or repressed, whereas promoters bound by more than four factors are largely active in the pluripotent state and become repressed upon differentiation [198-200]. Moreover, Oct4, Sox2 and Nanog may directly influence silencing of developmental gene programs through the recruitment of polycomb proteins such as Suz12 to regulatory gene regions [201] or indirectly through the regulation of polycomb group proteins [202].

In addition, the extended transcription factor network in ES cells has been shown to interact with two signaling pathways central for the maintenance of pluripotent ES cells [200]. It was found that STAT3 and Smad1, two key components of the LIF and BMP signaling pathways, respectively, share many targets with the Oct4, Sox2 and Nanog, indicating that external signals important for maintaining the pluripotent state are linked to downstream effectors through cell signaling mediators.

Stable maintenance and inheritance of lineage determination is then further supported by epigenetic modifications, including DNA methylation and posttranslational histone modifications. Thus, both transcription factor networks and epigenetic regulation are needed to maintain pluripotency in the earl embryo and in ES cells and to determine cell lineage identity upon differentiation [155]. Figure 8 summarizes the most important factors involved and their crosstalk in the maintenance of ES cell pluripotency and selfrenewal.

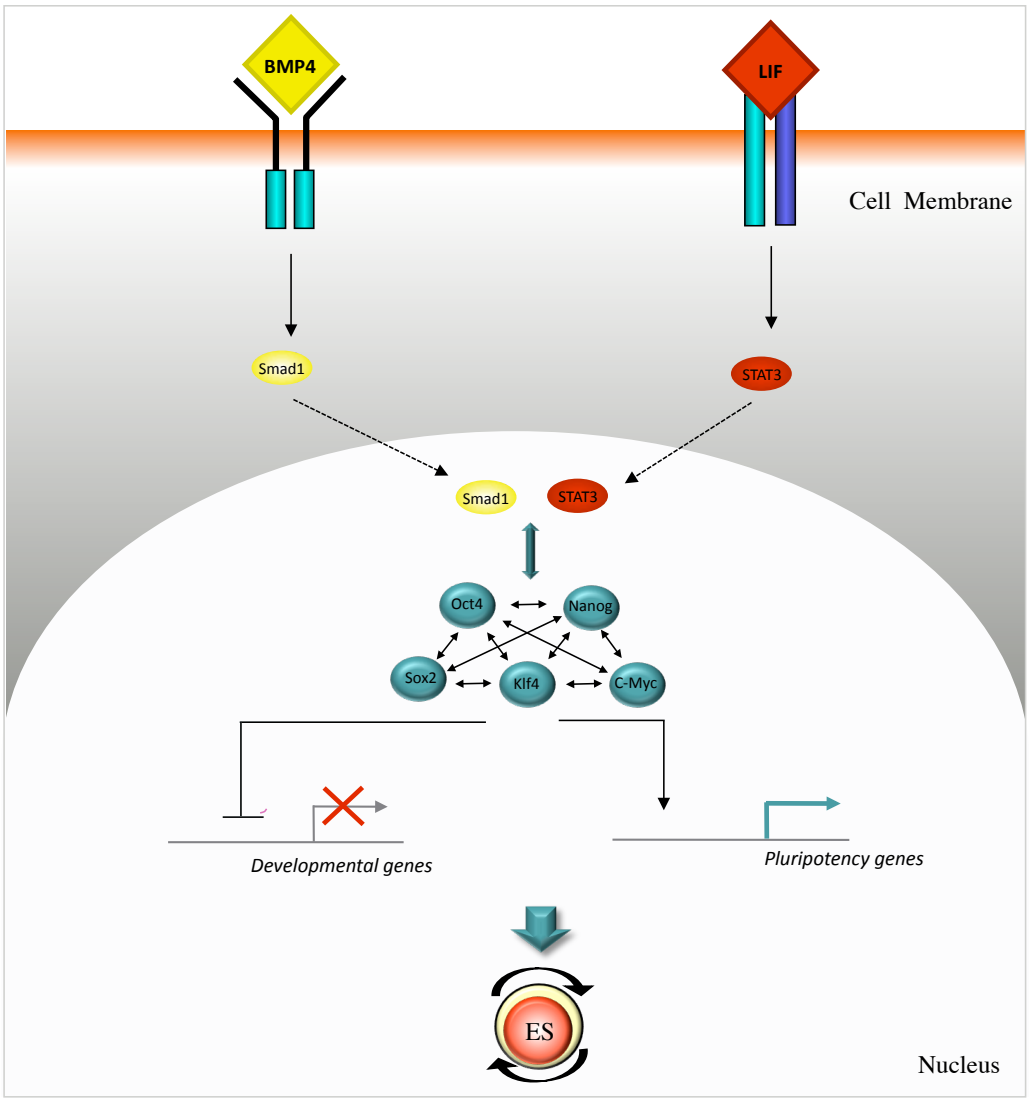


Figure 9. Simplified model for ES cell maintenance. Binding of extrinsic signaling molecules such as LIF and BMP4 activates downstream signaling pathways. As a result, downstream effector molecules such as STAT3 and Smad1 translocate to the nucleus, and interact with core pluripotency transcription factors. Complex protein complexes are then involved in the activation of targets genes that encode pluripotency genes. At the same time, they also repress the expression of developmental genes. These multifaceted regulation mechanisms lead to ES cell proliferation and self-renewal. *Inspired by* [155, 203].

3 Aim of the Thesis

TDG is a DNA glycosylase capable of recognizing and removing a wide range of mutagenic DNA base damages resulting from deamination, oxidation and alkylation. Its preferred substrates, however, are the deamination products of C and 5-meC, i.e. the G/U and G/T mismatches, respectively [1]. Besides this role in mutation avoidance, TDG was associated with the regulation of gene transcription by physical and functional interactions with transcription factors [2-5]. Additional interactions with the *de novo* DNA methyltransferases [6] in the context of transcriptional activation suggested that TDG might be targeted to gene promoters to assist cyclic DNA demethylation processes, thereby preventing mutations and/or erroneous methylation of CG-rich promoters. The severe developmental phenotype of *Tdg* knockout mice then showed than unlike other DNA glycosylases, TDG has an essential function in embryonic development.

These studies led us to the hypothesis that the function of TDG might be distinct from classical base excision repair and that it may involve the establishment of cell-lineage specific gene expression patterns, possibly by controlling CpG methylation at gene promoters.

The aim of this thesis was to gain deeper insights into the biological function(s) of TDG.

In a first part, the aim was to establish and apply biochemical high affinity purification and subcellular fractionation procedures allowing structural and functional characterization of TDG and TDG containing protein complexes.

The aim of the second part was to establish and validate a specific chromatin immunoprecipitation (ChIP) method, to perform genome-wide TDG ChIP analyses and to correlate TDG chromatin associations with gene expression and DNA methylation profiles.

4 Results

This results section shortly summarizes the data presented in the manuscripts in the appendix. Experiment and results not included in these manscripts are presented in this section.

4.1 Protein Interaction Studies Reveal a Tight Association of Thymine DNA Glycosylase With Epigentic Regulators in Chromatin (Appendix I)

Thymine DNA glycosylase (TDG) catalyzes the first step of DNA base excision repair (BER), the removal of the damaged base. TDG is able to excise a wide range of damaged DNA bases, resulting from deamination, oxidation and alkylation. Its preferred substrates are G/U and G/T mispairs resulting from the hydrolytic deamination of C and 5-meC, respectively. This substrate preference suggests that TDG plays a role in the avoidance of C to T transition mutations in DNA [1]. In this context, TDG dependent BER has also been proposed to participate in the regulation of gene expression by the maintenance of CpG methylation states through a coordinated action with sequence specific transcription factors [2-5] and *de novo* DNA methyltransferases [6].

Despite the detailed understanding of its biochemical and structural properties, the biological function has not been resolved.

To gain mechanistic insights into the biological functions of TDG, we followed different approaches in order to isolate TDG containing protein complexes from mammalian cells. In a first attempt, we purified TDG using the tandem affinity purification (TAP) method, a high affinity purification procedure that has been used for the identification of protein complexes under native conditions in yeast and mammalian cells [204-208]. This approach, even after extensive optimization of every single purification step and upscaling of the method, provided insufficient amounts of purified protein to identify proteins copurifying with TDG by mass spectrometry (MS). Further attempts to purify TDG interacting proteins by immunoprecipitation of endogenous TDG and TDGa from *TdgA* overexpressing ES cells also failed.

The negative results were surprising, taking into account that TDG protein interactors have been reported. However, it is remarkable, that they were all discovered in screens with the partner protein, with the exception of Sumo1 and Sumo3, under conditions of overexpression of both TDG and the respective interacting partner. For this reason, we then decided to apply more biochemical approaches. We first did fractionation of total cell extracts by gel filtration and showed that TDG is able to form distinct high molecular weight complexes in undifferentiated mouse embryonic stem cells. The RNA helicase p68 cofractionated with TDG in high molecular weight fractions, supporting the involvement of TDG and p68 in a common complex [10].

Previously described interactions of TDG with proteins acting on DNA [1, 3-5, 49] led us to the hypothesis that TDG interactions might preferentially take place in association with chromatin. Further biochemical fractionation in transcriptionally active and pluripotent mouse P19 teratocarcinoma cells and showed that TDG is present in all compartments of the cell. Interestingly, a significant fraction of nuclear TDG was associated with the insoluble chromatin fraction, a phenomenon also observed for *de novo* DNA methyltransferase Dnmt3a and Dnmt3b. We observed, that TDG can be partially eluted form the chromatin fraction by increasing salt concentrations or by DNA fragmentation using micrococcal nuclease (MNase). MNase treatment also partially released Dnmt3a and Dnmt3b. Interestingly, a significant fraction of TDG and Dnmt3a/b, and all p68, remained bound to chromatin even after efficient MNase treatment or high salt elution (1M NaCl) indicating that protein complexes containing p68 and Dnmt3a/b might form with high preference in association with chromatin.

Based on these results, future experiments will therefore focus on the biochemical identification and characterization of chromatin associated TDG-protein complexes in the context of gene regulation by classical biochemical methods combined with high affinity purification.

4.2 Thymine DNA Glycosylase Associates With CpG Islands to Maintain Their Epigenetic Function (Appendix II)

TDG has been proposed to be involved in the regulation of gene expression through the physical and functional interactions with sequence-specific transcription factors, histone modifying enzymes [2-5] and *de novo* DNA methyltransferases [6]. These interactions with factors essential in developmental gene regulation support the fact that TDG, unlike other glycosylases, has an essential function during embryonic development, as TDG *knockout* embryos die at midgestation.

We have shown that TDG associates tightly with chromatin in a DNA damage independent manner in P19 cells, together with the known interactors p68, an RNA helicase, and Dnmt3a and Dnmt3b (Appendix I). Altogether, these results implicated that TDG might localize and be targeted in a sequence- or site-specific manner to gene regulatory regions during fetal development and cell differentiation.

Here, we show that upon RA treatment of ES cells, TDG protein levels increase gradually, concomitant with the upregulation of Dnmt3a and downregulation of Dnmt3b. We further examined the subcellular localization of TDG in undifferentiated ES cells and retinoic acid (RA) treated cells. As previously observed in P19 cells, TDG tightly associates with DNA before and after RA induced differentiation. Also, the loss of TDG did not affect localization of Dnmt3b and Dnmt3a in undifferentiated and RA treated cells, respectively.

Based on these observations we started to work on the hypothesis that TDG dependent BER processes might have a more general role in the regulation of gene promoters undergoing dynamic DNA methylation changes during development and/or cell differentiation. Genome-wide dynamic changes of gene expression, along with epigenetic changes such as DNA methylation and histone modifications were previously studied during *in vitro* differentiation of mouse ES cells into homogenous populations of neuronal progenitor cells (NPs) [11, 214]. We thus addressed the role of TDG during the RA induced differentiation of $Tdg^{+/-}$ and $Tdg^{-/-}$ embryonic stem (ES) cells to neuronal progenitors (NPs).

We show that both $Tdg^{+/-}$ and $Tdg^{-/-}$ are able to differentiate to early NPs (4hNPs), a clear indicative that both cell line respond to RA. This was confirmed by downregulation of pluripotency markers such as *Oct4* and upregulation of the NP specific genes *Nes* and *Pax6.* However, early $Tdg^{-/-}$ NPs did not differentiate further

along the neuronal lineage and in contrast to *Tdg*$^{+/-}$ NPs, lost viability after 2 days of plating in neuronal medium.

We then examined genome-wide chromatin associations of TDG in ES cells and NPs and correlated these with changes in DNA methylation and gene expression during NP differentiation, to address the contribution of TDG to DNA (de)methylation and regulation of gene expression in cell fate determination.

First, we established and validated the TDG chromatin immunoprecipitation (ChIP) for ES cells and NPs by analyzing TDG occupancy at the *Rplp0 (P0*) proximal promoter, a previously described TDG binding region [213]. We then performed TDG-ChIP from *Tdg*$^{+/-}$and *Tdg*$^{-/-}$ ES cells and NPs coupled to high-throughput DNA sequencing to analyze genome-wide binding of TDG to chromatin.

Our studies led to the important discovery that TDG selectively binds to CpG island (CGI) promoters in ES cells and NPs. Also, CGI promoters bound by TDG were found exclusively in an active chromatin environment, as determined by the presence of dimethylated H3 at lysine 4 (H3K4me2). In most of these cases (72% and 74% in ES cells and NPs, respectively), these promoters belonged to actively transcribed genes (e.g. H3K4me2 and RNA polymerase II bound). The remaining was shown to regulate genes that are “poised” to be expressed during early and terminal cell differentiation (e.g. H3K4me2 and RNA polymerase II unbound). Among the genes regulated by TDG bound promoters in ES cells we found pluripotency genes such as *Oct4*, *Nanog* and *Gdf3*. Interestingly, differentiation-induced silencing of these genes, which is accompanied by DNA methylation of the respective promoters, correlated with loss of TDG association, indicating that TDG preferentially localizes to active and DNA demethylated gene promoters.

We also found TDG to localize promoters in bivalent chromatin domains (H3K4me2 and H3K27me3). Genes controlled by such promoters represent important polycomb target genes controlling developmental processes (e.g. *Sfrp2, Tgfb2, Gata6*) thus, supporting a function of TDG in cell differentiation and/or embryonic development. Binding of TDG to these regions was particularly interesting, as polycomb silenced genes in stem cells are prone to become aberrantly DNA methylated in cancer, suggesting that TDG might protect CGI promoters of polycomb target genes from aberrant methylation.

To address the function of TDG at the target regions identified, we first tested differential promoter DNA methylation in TDG proficient and deficient cells that underwent differentiation from ES the cell stage to NPs. Methylated DNA immunoprecipitation (MeDIP) revealed that global DNA methylation patterns in NPs are not affected in the absence of TDG. Nevertheless, we detected some significant changes at a number of promoters, indicating that TDG might regulate the DNA methylation pattern of some promoters. Also, owing to the limited resolution of the MeDIP method, we could not exclude an involvement of TDG in the control of DNA methylation of specific promoter CpGs. Additional bisulfite sequencing of promoters of TDG bound developmental genes (e.g. *Sfrp2, Tgfb2, HoxD13*) in NPs and differentiated mouse embryonic fibroblasts (MEFs; Christophe Kunz, personal communication) have then shown that loss of TDG affects local changes in DNA methylation at particular CpGs.

We also tested differential gene expression in TDG proficient and deficient ES cells and early NPs by using Illumina expression Arrays. First, this confirmed that differentiation was efficient in both cell types, as pluripotency genes (e.g. *Oct4, Nanog, Gdf3*) were downregulated upon differentiation. On the other hand, clear upregulation of NP specific genes (e.g. *Nes, Neurog2, Pax6*) was detected. When comparing $Tdg^{+/-}$ with $Tdg^{-/-}$ ES cells and $Tdg^{+/-}$ with $Tdg^{-/-}$ NPs, 229 and 1022 genes, respectively, showed significant differential expression. This indicates that the absence TDG affects the accuracy of reprogramming gene expression in cells undergoing cell differentiation. Pathway analysis of significantly differentially regulated genes revealed that a number of genes affecting important developmental and cellular differentiation processes are misregulated TDG deficient cells (e.g., *Sfrp2, Gata6*).

During further differentiation of NPs into the neuronal lineage, we observed that $Tdg^{-/-}$ 4hNPs failed to properly further differentiate after the early NP stage. This was accompanied by inefficient *Pax6* downregulation, which is essential for proper development of NPs to terminal neurons [12]. Interestingly, complementation of $Tdg^{-/-}$ ES cell with an external vector containing the TDGa isoform rescued *Pax6* misregulation during terminal neuron differentiation, indicating that TDG might be directly involved in the regulation of the *Pax6* promoter activity. Whether TDG acts a

coregulator of transcription, a scaffold protein and/or whether it influences CpG methylation pattern of the *Pax6* promoter will have to be carefully analyzed in future studies.

Our results indicate that TDG has a function in establishing and/or maintaining the correct CpG methylation pattern of CpG island promoters regulating key pluripotency and developmental genes. As a consequence, loss of TDG causes misregulation of genes during cell differentiation. Together, our results suggest that he base excision repair enzyme TDG has a role not only in maintaining genome but also epigenome stability in the course of early development.

4.3 Supplementary Results

4.3.1 TDG is Involved in The Regulation of Oct4 Gene Expression

Preliminary results from Yusuke Saito in our laboratory have indicated differential DNA methylation kinetics at the *Oct4* proximal promoter (PP) region following retinoic acid (RA) induced *in vitro* differentiation of $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells. In similar experiments with P19 cells, others have shown the involvement of Dnmt3a in the stable silencing of the *Oct4* gene through DNA methylation of specific CpGs in its proximal promoter region [195]. This observations indicated that TDG, in a complex with *de novo* methyltransferases and transcription factors [3, 6, 52], might be involved in the regulation of DNA methylation at the *Oct4* promoter. Analysis of the *Oct4* expression in ES cells before and after retinoic acid (RA) induced differentiation showed that although both $Tdg^{+/-}$ *and* $Tdg^{-/-}$ cells downregulate *Oct4* expression upon differentiation TDG deficient cells showed lower levels of *Oct4* expression than TDG proficient cells after 48 hours of RA treatment (Fig. 1B). Interestingly, differential *Oct4* expression was already observed at the undifferentiated ES cell stage. These data suggest that TDG proficient and deficient ES cells are both able to respond to RA, but cells lacking TDG seem to be more prone to enter spontaneous and/or induced differentiation.

We then addressed whether TDG is targeted to the *Oct4* proximal promoter region

in ES cells. As shown in Figure 1C, TDG ChIP from *Tdg*$^{+/-}$ cells clearly enriched Oct4 promoter sequence. Together, these results establish binding of TDG to the Oct4 proximal promoter in undifferentiated cells, suggesting a role for TDG in the regulation of pluripotency.

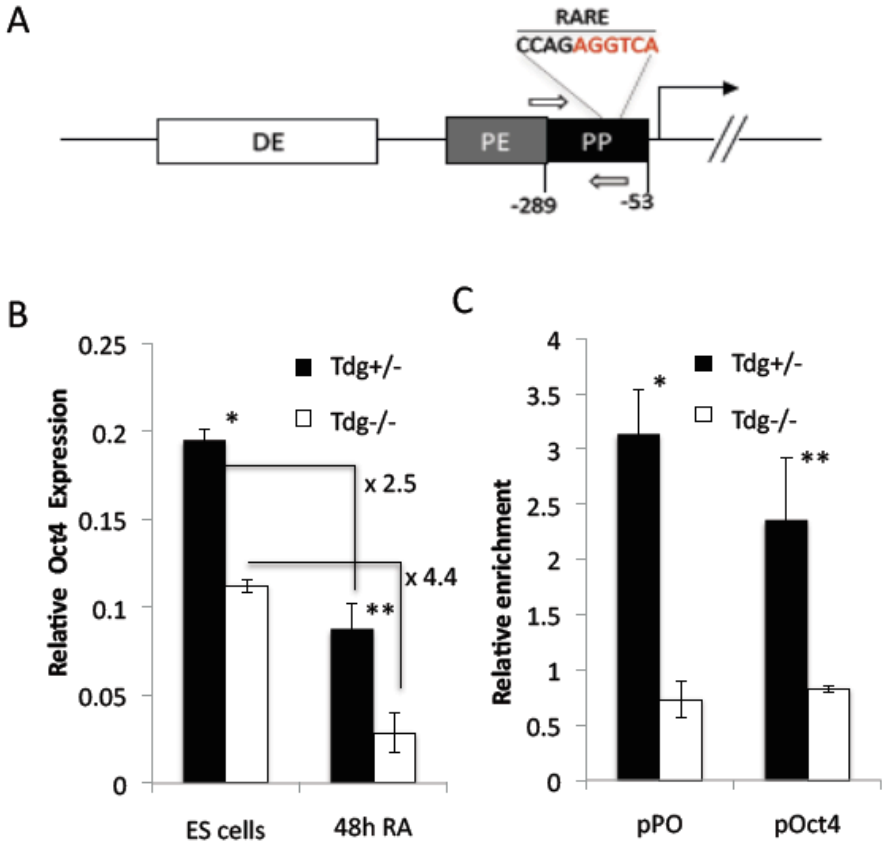


Figure 1. Involvement of TDG in the regulation of *Oct4* expression A) Schematic representation of the *Oct4* 5' flanking region indicating positions of the primers used for ChIP quantitative PCR (*open arrows*) and position of the retinoic acid response element (RARE) **B)** Tdg$^{+/-}$ and Tdg$^{-/-}$ ES cells were treated with 1µM retinoic acid (RA) for 48 hours. Total RNA was isolated from undifferentiated and RA treated cells and *Oct4* gene expression was determined by RT-PCR. *Gapdh* was used as internal normalization. *Tdg*$^{-/-}$ ES cells show decreased levels of *Oct4* expression at the undifferentiated stage as well as after treatment with 1µM RA for 48 hours. Shown is the mean of three independent experiments with error bars indicating ± SE. *p≤0.01; **p≤0.05 by students T-Test. **C)** TDG binding to the *Oct4* proximal promoter in undifferentiated ES cells was assessed by ChIP. The *Gapdh* gene-coding region served as control and was used for normalization. P0 promoter binding was used as positive ChIP control. Results are presented as relative enrichment (See Methods for calculation). Shown is the mean of three independent experiments with error bars indicating ± SE. * p≤0.01, **p≤0.05 by one-way ANOVA. See Appendix II for methods on the differentiation procedure.

4.3.2 TDG is Involved in The Regulation of the Lineage Specific Gene Lrrtm2

In collaboration with Gilles Salbert and colleagues at the University of Rennes, (France) we then examined whether TDG also localizes to genes that are methylated in ES cell but become demethylated and activated during cell differentiation. One such gene is *Lrrtm2* (leucine rich repeat transmembrane neuronal 2 gene), the promoter of which was shown to be demethylated and activated upon neuronal differentiation [11]. Interestingly, after 48 hours of RA induced differentiation, *Lrrtm2* expression was induced in *Tdg*$^{+/-}$ cells, but not in *Tdg* $^{-/-}$ *knockout* cells (Fig. 2A). In addition, we detected by ChIP specific binding of TDG to the *Lrrtm2* promoter (Fig.2B). Activation of this promoter further correlated with the recruitment of RAR-α and the COUP transcription factor I (COUP-TFI) and Dnmt3a, which were previously shown cooperate in the transcriptional regulation of the *vitronectin* gene [213]. Strikingly, COUP-TFI, RAR and Dnmt3a engagement at the *Lrrtm2* promoter appeared to be dependent on TDG (Fig. 2B). Similar results were observed at the *vitronectin* promoter, which becomes partially demethylated during RA induced differentiation (Gilles Salbert, personal communication) (Fig. 2C).

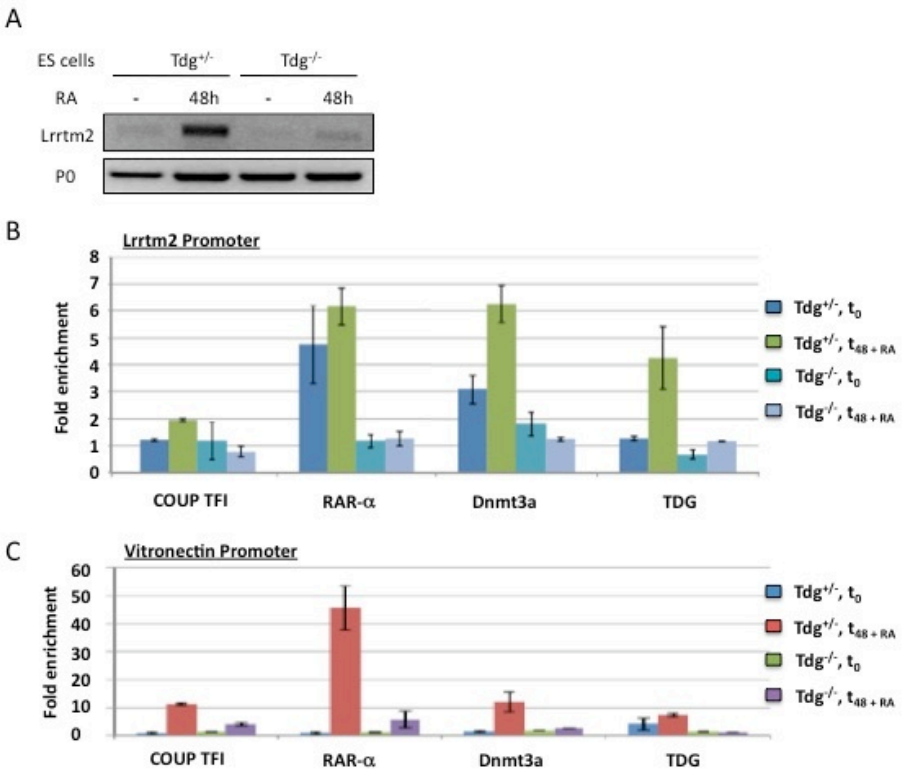


Figure 2. A) RT-PCR analysis of *Lrrtm2* and *P0* transcripts during retinoic acid induced differentiation of $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells. **B)** COUP-TFI, TDG, Dnmt3a and RAR-α binding to the *Lrrtm2* promoter before and after treatment with 1μM retinoic acid (RA) for 48 hours was assessed by ChIP in $Tdg^{+/-}$ and $Tdg^{-/-}$ cells (collaboration with Gilles Salbert, University of Rennes, France). Results are expressed as fold enrichment relative to Input DNA put DNA, with error bars representing SDs. **C)** COUP-TFI, TDG, Dnmt3a and RAR localization to the *vitronectin* promoter before and after treatment with 1μM RA for 48 hours. Results are expressed as fold enrichment over input DNA, with error bars representing SDs. See Appendix II for methods on the differentiation procedure.

These results appear to corroborate an involvement of TDG in the demethylation of the Lrrtm2 promoter by a mechanism involving Dnmt3a dependent deamination of 5-meC as proposed previously [6]. Whether or not TDG plays an active role as a DNA glycosylase, processing the resulting G/T mismatches, or a passive role as a scaffold factor important for the recruitment of transcription factors and Dnmt's to promoter regions remains to be clarified.

4.4 The Enigmatic Thymine DNA Glycosylase (Appendix III)

TDG has turned out to be a special DNA glycosylase, implicated in DNA repair and in the regulation of gene expression. Strikingly, either of these functions appears to be essential for proper embryonic development. The unexpected complexity of and the uncertainty about its biological function led us to review the last decade of research on this fascinating DNA repair protein. I contributed to an important chapter of this extended review concerning a proposed function of TDG in the active demethylation recognition and removal of 5meC. This review is found attached in Appendix III.

5 Summarizing Conclusions and Outlook

When it was first isolated from human cell extracts, TDG was found to be able to remove thymine from G/T mispairs. This implicated a function of this mismatch specific DNA glycosylase in counteracting the mutability of methylated cytosines in DNA [1]. These bases tend to spontaneously deaminate, leading to the formation of T mispaired with G in double stranded DNA. The experimental evidence that has emerged during the last decades, however, has indicated functions of TDG that are more complex than initiating the repair of damaged bases. Protein-protein interactions have implicated a role for TDG in the regulation of gene expression [1-5] and more recent studies have suggested that this might be linked with DNA glycosylase activity [6]. Together with the discovery that TDG deficient mice die during early embryogenesis, these observations have led to the current hypothesis, that TDG dependent BER might participate in the regulation of developmental gene expression by the maintenance of CpG methylation states through a coordinated action with sequence specific transcription factors and *de novo* DNA methyltransferases.

In a first part of my thesis, we aimed to provide novel insights into the biological function of TDG in mammalian cells by biochemical purification and structural as well as functional characterization of TDG containing protein complexes. Initial work concentrated on the purification of TDG from mouse embryonic fibroblasts (MEFs) and mouse embryonic stem (ES) cells. After extensive elaboration and optimization of various purification approaches, we were not successful in identifying any proteins interacting with TDG.

The biological function of TDG might indeed explain the difficulty to purify TDG-protein complexes from mammalian cells. So far all the interaction partners of TDG have been discovered in screens with the partner protein and under conditions where both TDG and the interacting protein were overexpressed. In addition, the reported partners are exclusively involved in dynamic biological processes taking place in chromatin [6]. For this reason, we decided to classical size fractionation methods to characterize the molecular environment of TDG. Gel filtration

experiment revealed that TDG indeed forms distinct protein complexes in cells and fractionation of the cell compartments further showed that a portion of nuclear TDG is tightly associated with chromatin. This suggested that TDG might interact with other proteins mainly in a chromatin context.

Future studies in the direction of identifying TDG interacting proteins will thus have to focus on the purification of chromatin associated TDG by cell fractionation, followed by classical size-exclusion chromatography to enrich for TDG complexes, which then can be further purified by immunoprecipitation and analyzed by MS.

In the second part of my thesis, I started to focus on chromatin associated TDG to investigate the hypothesis that TDG dependent repair has a general role in the regulation of gene promoters during cell differentiation, governed by dynamic gene expression and DNA methylation changes. We thus decided to investigate genome-wide TDG chromatin association during the differentiation of ES cells to NPs, and correlated our findings to global changes in DNA methylation in and gene expression.

We discovered that chromatin associated TDG selectively binds to CpG island promoters in an 'open' chromatin state, independently of the stage of cellular differentiation. In addition, promoters bound by TDG belonged either to actively transcribed genes or genes poised for activation during cell differentiation. Interestingly, pluripotency and developmental genes where represented amongst the TDG targets, a fraction of which is known to be silenced by polycomb repressive complexes (PRCs) in ES cells. Polycomb mediated silencing in stem cells is reversible and due to this reason, it does not involve DNA methylation. Nevertheless, *de novo* DNA methyltransferases have been shown to associate with PRCs and to aberrantly methylate promoters of developmental genes in cancer cells. As the mechanism involved in protection from methylation in untransformed cells is not known, the binding of TDG to polycomb targets is particularly interesting and opens the possibility of TDG being involved in the protection of promoter CpG from being aberrantly methylated. We are currently looking at the subcellular localization of TDG in colorectal cancer tissues. Preliminary data indicate that TDG is absent from a substantial fraction of tumors, while overexpressed in others, indicating that TDG

might control carcinogenesis. This will be carefully analyzed by methods including biochemical fractionation of and TDG ChIP from primary cancer cells lines, and by correlating cancer associated DNA methylation changes with the TDG expression and activity state.

To address a possible role of TDG in DNA methylation, we performed MeDIP analysis, assessing global DNA methylation changes in NPs. We identified DNA methylation changes at a limited number of promoters, however, the data suggested that global CpG methylation is not affected upon loss of TDG. Because of the limited sensitivity and resolution of the MeDIP technique, however, changes of local DNA methylation changes were probably not detected. Indeed, recent bisulfite sequencing of the promoters of developmental polycomb target genes regulated by TDG bound promoters revealed DNA methylation changes at specific CpG sites both in NPs and MEFs, indicating that TDG does have a role in regulating local DNA methylation in gene promoters.

TDG was previously shown to be involved in the regulation of gene expression. Genome-wide gene expression analysis in ES cells and NPs revealed that loss of TDG in ES cells leads to significant misregulation of genes in the course of neuronal differentiation. Moreover, we find that a significant number of differentially expressed genes between TDG proficient and deficient NPs are regulated by promoters bound by TDG in ES cells including important developmental genes (e.g. *Gata6, Pdgfra, Sfrp2*). We further showed that insufficient downregulation of the *Pax6* gene at the neuronal progenitor stage, a process that is indispensable for proper development of terminal neurons, is a direct consequence of TDG loss. In this context, it is not clear yet if TDG regulated the DNA methylation status of the Pax6 promoter or if it acts in a passive way as a scaffold factor. This will be subject of future studies.

Together our studies suggest that TDG is indeed involved in the regulation of gene expression during cell differentiation and presumably, embryonic development. Although TDG loss does not affect global DNA methylation in

differentiated cells, it appears to be involved in regulating local DNA methylation patterns of developmental genes. In addition, TDG appears to me involved in CpG demethylation of promoters of the neuronal lineage specific gene *Lrrtm2*, which become demethylated during cell differentiation. The underlying mechanism is unclear but there is evidence pointing out to an involvement of 5-meC deamination by Dnmt3a. Whether of not TDG plays an active role as a DNA glycosylase in the repair of the G/T mispairs generated by 5-meC deamination or rather a passive role as scaffold factor facilitating the recruitment of *de novo* Dnmt's and transcription factors, remains to be clarified. In this context, future studies involve the analysis of the local CpG methylation pattern of the Lrrtm2 promoter by bisulfite treatment and DNA sequencing of bisulfite converted DNA in TDG proficient and deficient cells differentiation. In addition, we will perform Re-ChIP experiment with Dnmt3a and to address TDG mediated recruitment of these proteins to CpG island promoters.

One important question to be resolved is how TDG is targeted specifically to CpG rich sites. Based on current evidence, this is probably achieved by physical interactions with sequence specific transcription factors. We will address this by doing ChIP and Re-ChIP experiment with TDG and known interacting partnern (e.g. RAR, RXR). In addition, TDG ChIP experiment in ES cells expressing a DNA glycosylase mutant of TDG will answer the question if the catalytic activity is needed for localization to chromatin and/or for the regulation of DNA methylation pattern at targets discovered during this work.

Still, TDG on its own might directly find to unmethylated CpG. In this context, it is interesting that TDG was found to be associated with CpG rich RNA [93]. The origin of the RNA is not known. However one could think of a mechanism involving i) association of TDG with non-protein-coding CpG rich transcripts followed by ii) subsequent targeting and binding to complementary CpG rich DNA prior of transcription.

Loss of TDG could also be expected to increase the C to T transition mutations rate at CpG rich regions. So far, we have not addressed this point. However, we observed that TDG is associated to CpG Islands, which are conserved among mammals. These sequence stretches have been protected from the genomic

loss of CpGs during evolution due to their preferentially unmethylated state, and the mechanism protecting these sites from methylation and/or mutation are not known. Our results open the possibility that TDG could be responsible for the protection of unmethylated CpG Islands against both aberrant DNA methylation and mutations. We will analyze prominent TDG bound CpG rich regions in *Tdg*$^{-/-}$ differentiated cells (e.g. MEFs, NPs) for the presence of mutations by sequencing.

The discoveries described in here represent a significant progress in understanding the biological function(s) of the enigmatic thymine DNA glycosylase. Although many questions remain to be resolved, we are now in position to refine our functional models and test novel, more precise predictions (Fig. 8, Appendix II). Currently we would propose that TDG associates to CpG island promoters to protect these from DNA methylation and/or mutation. Whereas in the latter case the involvement of the DNA glycosylase activity seems essential and straightforward, the protection from erroneous methylation could involve either an active role of the DNA glycosylase to process G/T mismatches after deamination of 5-meC and/or a passive role of TDG as scaffold factor important for the recruitment of the deaminating activity of *de novo Dnmts* or other deaminases. Still, an alternative mechanisms of protection might involve shielding of already unmethylated CpGs, where TDG could either inhibit binding of DNA methyltransferases or inhibit their methyltransferase activity, as it has been shown *in vitro* [52].Future work in the direction of resolving the newly emerging relationship between genome and epigenome maintenance promises to be highly interesting.

6 References

1. Cortazar, D., et al., *The enigmatic thymine DNA glycosylase.* DNA Repair (Amst), 2007. **6**(4): p. 489-504.
2. Zhou, J., et al., *TDG represses myocardin-induced smooth muscle cell differentiation by competing with SRF for myocardin binding.* J Biol Chem, 2008.
3. Um, S., et al., *Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase.* J Biol Chem, 1998. **273**(33): p. 20728-36.
4. Lucey, M.J., et al., *T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif.* Nucleic Acids Res, 2005. **33**(19): p. 6393-404.
5. 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, 2003. **278**(40): p. 38586-92.
6. Metivier, R., et al., *Cyclical DNA methylation of a transcriptionally active promoter.* Nature, 2008. **452**(7183): p. 45-50.
7. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
8. Steinacher, R. and P. Schar, *Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation.* Curr Biol, 2005. **15**(7): p. 616-23.
9. Hardeland, U., et al., *Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover.* Embo J, 2002. **21**(6): p. 1456-64.
10. Jost, J.P., et al., *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, 1999. **27**(16): p. 3245-52.
11. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-66.
12. Bel-Vialar, S., F. Medevielle, and F. Pituello, *The on/off of Pax6 controls the tempo of neuronal differentiation in the developing spinal cord.* Dev Biol, 2007. **305**(2): p. 659-73.
13. Friedberg, E.C., *The intersection between the birth of molecular biology and the discovery of DNA repair.* DNA Repair (Amst), 2002. **1**(10): p. 855-67.
14. Friedberg, E.C., *DNA damage and repair.* Nature, 2003. **421**(6921): p. 436-40.
15. Scharer, O.D., *Chemistry and biology of DNA repair.* Angew Chem Int Ed Engl, 2003. **42**(26): p. 2946-74.
16. Lindahl, T., *Instability and decay of the primary structure of DNA.* Nature, 1993. **362**(6422): p. 709-15.
17. Nilsen, H. and H.E. Krokan, *Base excision repair in a network of defence and tolerance.* Carcinogenesis, 2001. **22**(7): p. 987-98.

18. Sancar, A., et al., *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints.* Annu Rev Biochem, 2004. **73**: p. 39-85.
19. Helleday, T., et al., *DNA double-strand break repair: from mechanistic understanding to cancer treatment.* DNA Repair (Amst), 2007. **6**(7): p. 923-35.
20. Hoeijmakers, J.H., *Genome maintenance mechanisms for preventing cancer.* Nature, 2001. **411**(6835): p. 366-74.
21. Hanawalt, P.C., *Subpathways of nucleotide excision repair and their regulation.* Oncogene, 2002. **21**(58): p. 8949-56.
22. Kunz, C., Y. Saito, and P. Schar, *DNA Repair in mammalian cells: Mismatched repair: variations on a theme.* Cell Mol Life Sci, 2009. **66**(6): p. 1021-38.
23. Carell, T., et al., *The mechanism of action of DNA photolyases.* Curr Opin Chem Biol, 2001. **5**(5): p. 491-8.
24. Essen, L.O., *Photolyases and cryptochromes: common mechanisms of DNA repair and light-driven signaling?* Curr Opin Struct Biol, 2006. **16**(1): p. 51-9.
25. Sedgwick, B., *Repairing DNA-methylation damage.* Nat Rev Mol Cell Biol, 2004. **5**(2): p. 148-57.
26. Falnes, P.O., A. Klungland, and I. Alseth, *Repair of methyl lesions in DNA and RNA by oxidative demethylation.* Neuroscience, 2007. **145**(4): p. 1222-32.
27. Dianov, G. and T. Lindahl, *Reconstitution of the DNA base excision-repair pathway.* Curr Biol, 1994. **4**(12): p. 1069-76.
28. Pascucci, B., et al., *Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases delta and epsilon.* J Biol Chem, 1999. **274**(47): p. 33696-702.
29. Srivastava, D.K., et al., *Mammalian abasic site base excision repair. Identification of the reaction sequence and rate-determining steps.* J Biol Chem, 1998. **273**(33): p. 21203-9.
30. Baute, J. and A. Depicker, *Base excision repair and its role in maintaining genome stability.* Crit Rev Biochem Mol Biol, 2008. **43**(4): p. 239-76.
31. Huffman, J.L., O. Sundheim, and J.A. Tainer, *DNA base damage recognition and removal: new twists and grooves.* Mutat Res, 2005. **577**(1-2): p. 55-76.
32. Krokan, H.E., et al., *Base excision repair of DNA in mammalian cells.* FEBS Lett, 2000. **476**(1-2): p. 73-7.
33. Hitomi, K., S. Iwai, and J.A. Tainer, *The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair.* DNA Repair (Amst), 2007. **6**(4): p. 410-28.
34. Zharkov, D.O. and A.P. Grollman, *The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases.* Mutat Res, 2005. **577**(1-2): p. 24-54.
35. Lindahl, T. and R.D. Wood, *Quality control by DNA repair.* Science, 1999. **286**(5446): p. 1897-905.
36. Matsumoto, Y. and K. Kim, *Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair.* Science, 1995. **269**(5224): p. 699-702.
37. Hang, B. and B. Singer, *Protein-protein interactions involving DNA glycosylases.* Chem Res Toxicol, 2003. **16**(10): p. 1181-95.
38. Beard, W.A. and S.H. Wilson, *Structural design of a eukaryotic DNA repair polymerase: DNA polymerase beta.* Mutat Res, 2000. **460**(3-4): p. 231-44.

39. Kubota, Y., et al., *Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein.* Embo J, 1996. **15**(23): p. 6662-70.
40. Dianova, II, et al., *XRCC1-DNA polymerase beta interaction is required for efficient base excision repair.* Nucleic Acids Res, 2004. **32**(8): p. 2550-5.
41. Podlutsky, A.J., et al., *Human DNA polymerase beta initiates DNA synthesis during long-patch repair of reduced AP sites in DNA.* Embo J, 2001. **20**(6): p. 1477-82.
42. Matsumoto, Y., et al., *Reconstitution of proliferating cell nuclear antigen-dependent repair of apurinic/apyrimidinic sites with purified human proteins.* J Biol Chem, 1999. **274**(47): p. 33703-8.
43. Robertson, A.B., et al., *DNA repair in mammalian cells: Base excision repair: the long and short of it.* Cell Mol Life Sci, 2009. **66**(6): p. 981-93.
44. Neddermann, P. and J. Jiricny, *The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells.* J Biol Chem, 1993. **268**(28): p. 21218-24.
45. Neddermann, P. and J. Jiricny, *Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1642-6.
46. Neddermann, P., et al., *Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase.* J Biol Chem, 1996. **271**(22): p. 12767-74.
47. Hardeland, U., et al., *The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs.* Nucleic Acids Res, 2003. **31**(9): p. 2261-71.
48. O'Neill, R.J., et al., *Mismatch uracil glycosylase from Escherichia coli: a general mismatch or a specific DNA glycosylase?* J Biol Chem, 2003. **278**(23): p. 20526-32.
49. Missero, C., et al., *The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription.* J Biol Chem, 2001. **276**(36): p. 33569-75.
50. Tini, M., et al., *Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription.* Mol Cell, 2002. **9**(2): p. 265-77.
51. Shimizu, Y., et al., *Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase.* Embo J, 2003. **22**(1): p. 164-73.
52. Li, Y.Q., et al., *Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair.* Nucleic Acids Res, 2007. **35**(2): p. 390-400.
53. Boland, M.J. and J.K. Christman, *Characterization of Dnmt3b:Thymine-DNA Glycosylase Interaction and Stimulation of Thymine Glycosylase-Mediated Repair by DNA Methyltransferase(s) and RNA.* J Mol Biol, 2008.
54. Chevray, P.M. and D. Nathans, *Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 5789-93.
55. Guan, X., et al., *The human checkpoint sensor Rad9-Rad1-Hus1 interacts with and stimulates DNA repair enzyme TDG glycosylase.* Nucleic Acids Res, 2007. **35**(18): p. 6207-18.

56. Aravind, L. and E.V. Koonin, *The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates.* Genome Biol, 2000. **1**(4): p. RESEARCH0007.
57. Mol, C.D., et al., *Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis.* Cell, 1995. **80**(6): p. 869-78.
58. Pearl, L.H., *Structure and function in the uracil-DNA glycosylase superfamily.* Mutat Res, 2000. **460**(3-4): p. 165-81.
59. Wibley, J.E., et al., *Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1.* Mol Cell, 2003. **11**(6): p. 1647-59.
60. Mol, C.D., et al., *Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases.* J Mol Biol, 2002. **315**(3): p. 373-84.
61. Barrett, T.E., et al., *Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions.* Cell, 1998. **92**(1): p. 117-29.
62. Baba, D., et al., *Crystal structure of thymine DNA glycosylase conjugated to SUMO-1.* Nature, 2005. **435**(7044): p. 979-82.
63. Baba, D., et al., *Crystal structure of SUMO-3-modified thymine-DNA glycosylase.* J Mol Biol, 2006. **359**(1): p. 137-47.
64. Hardeland, U., et al., *Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis.* J Biol Chem, 2000. **275**(43): p. 33449-56.
65. Borys-Brzywczy, E., et al., *Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts.* Acta Biochim Pol, 2005. **52**(1): p. 149-65.
66. Yoon, J.H., et al., *Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair.* Nucleic Acids Res, 2003. **31**(18): p. 5399-404.
67. Jost, J.P., *Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine.* Proc Natl Acad Sci U S A, 1993. **90**(10): p. 4684-8.
68. 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 U S A, 2000. **97**(10): p. 5135-9.
69. 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 U S A, 2001. **98**(9): p. 5031-6.
70. Morgan, M.T., M.T. Bennett, and A.C. Drohat, *Excision of 5-Halogenated Uracils by Human Thymine DNA Glycosylase: ROBUST ACTIVITY FOR DNA CONTEXTS OTHER THAN CpG*
10.1074/jbc.M704253200. J. Biol. Chem., 2007. **282**(38): p. 27578-27586.
71. Maiti, A., et al., *Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition.* Proc Natl Acad Sci U S A, 2008. **105**(26): p. 8890-5.

72. Sibghat, U., et al., *Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA-glycosylase.* Biochemistry, 1996. **35**(39): p. 12926-32.
73. Waters, T.R., et al., *Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1.* J Biol Chem, 1999. **274**(1): p. 67-74.
74. Smet-Nocca, C., et al., *The Thymine-DNA Glycosylase Regulatory Domain: Residual Structure and DNA Binding.* Biochemistry, 2008.
75. Bellacosa, A., *Role of MED1 (MBD4) Gene in DNA repair and human cancer.* J Cell Physiol, 2001. **187**(2): p. 137-44.
76. Hendrich, B., et al., *The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites.* Nature, 1999. **401**(6750): p. 301-4.
77. Millar, C.B., et al., *Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice.* Science, 2002. **297**(5580): p. 403-5.
78. Greenblatt, M.S., et al., *Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis.* Cancer Res, 1994. **54**(18): p. 4855-78.
79. Kondo, E., et al., *The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes.* Mol Cell Biol, 2005. **25**(11): p. 4388-96.
80. Kunz, C., et al., *Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil.* PLoS Biol, 2009. **7**(4): p. e91.
81. Nilsen, H., et al., *Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication.* Mol Cell, 2000. **5**(6): p. 1059-65.
82. Hardeland, U., et al., *Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2.* Nucleic Acids Res, 2007. **35**(11): p. 3859-67.
83. An, Q., et al., *C --> T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases.* Embo J, 2005. **24**(12): p. 2205-13.
84. Petersen-Mahrt, S., *DNA deamination in immunity.* Immunol Rev, 2005. **203**: p. 80-97.
85. Fleck, O. and P. Schar, *Translesion DNA synthesis: little fingers teach tolerance.* Curr Biol, 2004. **14**(10): p. R389-91.
86. Di Noia, J.M., C. Rada, and M.S. Neuberger, *SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair.* Embo J, 2006. **25**(3): p. 585-95.
87. Reid, G., R. Gallais, and R. Metivier, *Marking time: The dynamic role of chromatin and covalent modification in transcription.* Int J Biochem Cell Biol, 2008.
88. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nat Rev Genet, 2008. **9**(6): p. 465-76.
89. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development.* Science, 2001. **293**(5532): p. 1089-93.
90. Jeltsch, A., *Molecular enzymology of mammalian DNA methyltransferases.* Curr Top Microbiol Immunol, 2006. **301**: p. 203-25.

91. Jost, J.P. and Y.C. Jost, *Transient DNA demethylation in differentiating mouse myoblasts correlates with higher activity of 5-methyldeoxycytidine excision repair.* J Biol Chem, 1994. **269**(13): p. 10040-3.
92. Jost, J.P. and Y.C. Jost, *Mechanism of active DNA demethylation during embryonic development and cellular differentiation in vertebrates.* Gene, 1995. **157**(1-2): p. 265-6.
93. Jost, J.P., et al., *The RNA moiety of chick embryo 5-methylcytosine- DNA glycosylase targets DNA demethylation.* Nucleic Acids Res, 1997. **25**(22): p. 4545-50.
94. Fremont, M., et al., *Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA.* Nucleic Acids Res, 1997. **25**(12): p. 2375-80.
95. Fuller-Pace, F.V., *DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation.* Nucleic Acids Res, 2006. **34**(15): p. 4206-15.
96. Kangaspeska, S., et al., *Transient cyclical methylation of promoter DNA.* Nature, 2008. **452**(7183): p. 112-5.
97. Endoh, H., et al., *Purification and identification of p68 RNA helicase acting as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor alpha.* Mol Cell Biol, 1999. **19**(8): p. 5363-72.
98. Felsenfeld, G. and M. Groudine, *Controlling the double helix.* Nature, 2003. **421**(6921): p. 448-53.
99. Horn, P.J. and C.L. Peterson, *Molecular biology. Chromatin higher order folding--wrapping up transcription.* Science, 2002. **297**(5588): p. 1824-7.
100. Peterson, C.L. and M.A. Laniel, *Histones and histone modifications.* Curr Biol, 2004. **14**(14): p. R546-51.
101. Zhang, Y. and D. Reinberg, *Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails.* Genes Dev, 2001. **15**(18): p. 2343-60.
102. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
103. Verger, A. and M. Crossley, *Chromatin modifiers in transcription and DNA repair.* Cell Mol Life Sci, 2004. **61**(17): p. 2154-62.
104. Seet, B.T., et al., *Reading protein modifications with interaction domains.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 473-83.
105. Sterner, D.E. and S.L. Berger, *Acetylation of histones and transcription-related factors.* Microbiol Mol Biol Rev, 2000. **64**(2): p. 435-59.
106. Kouzarides, T., *Acetylation: a regulatory modification to rival phosphorylation?* Embo J, 2000. **19**(6): p. 1176-9.
107. Martinez-Balbas, M.A., et al., *Regulation of E2F1 activity by acetylation.* Embo J, 2000. **19**(4): p. 662-71.
108. Stallcup, M.R., *Role of protein methylation in chromatin remodeling and transcriptional regulation.* Oncogene, 2001. **20**(24): p. 3014-20.
109. Lachner, M., R.J. O'Sullivan, and T. Jenuwein, *An epigenetic road map for histone lysine methylation.* J Cell Sci, 2003. **116**(Pt 11): p. 2117-24.

110. Ruthenburg, A.J., C.D. Allis, and J. Wysocka, *Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark.* Mol Cell, 2007. **25**(1): p. 15-30.
111. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation.* Nat Rev Mol Cell Biol, 2005. **6**(11): p. 838-49.
112. Schubeler, D., et al., *The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote.* Genes Dev, 2004. **18**(11): p. 1263-71.
113. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast.* Cell, 2005. **122**(4): p. 517-27.
114. Zegerman, P., et al., *Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex.* J Biol Chem, 2002. **277**(14): p. 11621-4.
115. Lachner, M. and T. Jenuwein, *The many faces of histone lysine methylation.* Curr Opin Cell Biol, 2002. **14**(3): p. 286-98.
116. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.
117. Rice, J.C., et al., *Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains.* Mol Cell, 2003. **12**(6): p. 1591-8.
118. Zhang, R., et al., *HP1 proteins are essential for a dynamic nuclear response that rescues the function of perturbed heterochromatin in primary human cells.* Mol Cell Biol, 2007. **27**(3): p. 949-62.
119. Stewart, M.D., J. Li, and J. Wong, *Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment.* Mol Cell Biol, 2005. **25**(7): p. 2525-38.
120. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
121. Boyer, L.A., et al., *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.
122. Margueron, R., et al., *Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms.* Mol Cell, 2008. **32**(4): p. 503-18.
123. Vire, E., et al., *The Polycomb group protein EZH2 directly controls DNA methylation.* Nature, 2006. **439**(7078): p. 871-4.
124. Schlesinger, Y., et al., *Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer.* Nat Genet, 2007. **39**(2): p. 232-6.
125. Widschwendter, M., et al., *Epigenetic stem cell signature in cancer.* Nat Genet, 2007. **39**(2): p. 157-8.
126. Wang, G.G., C.D. Allis, and P. Chi, *Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling.* Trends Mol Med, 2007. **13**(9): p. 373-80.
127. Hassan, A.H., et al., *Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes.* Cell, 2002. **111**(3): p. 369-79.
128. Boyer, L.A., R.R. Latek, and C.L. Peterson, *The SANT domain: a unique histone-tail-binding module?* Nat Rev Mol Cell Biol, 2004. **5**(2): p. 158-63.

129. Flanagan, J.F., et al., *Double chromodomains cooperate to recognize the methylated histone H3 tail.* Nature, 2005. **438**(7071): p. 1181-5.
130. Wysocka, J., et al., *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.
131. Liang, J., et al., *Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells.* Nat Cell Biol, 2008. **10**(6): p. 731-9.
132. Martens, J.A. and F. Winston, *Recent advances in understanding chromatin remodeling by Swi/Snf complexes.* Curr Opin Genet Dev, 2003. **13**(2): p. 136-42.
133. Gong, F., D. Fahy, and M.J. Smerdon, *Rad4-Rad23 interaction with SWI/SNF links ATP-dependent chromatin remodeling with nucleotide excision repair.* Nat Struct Mol Biol, 2006. **13**(10): p. 902-7.
134. van Attikum, H., et al., *Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair.* Cell, 2004. **119**(6): p. 777-88.
135. Hermann, A., R. Goyal, and A. Jeltsch, *The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites.* J Biol Chem, 2004. **279**(46): p. 48350-9.
136. Okano, M., S. Xie, and E. Li, *Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases.* Nat Genet, 1998. **19**(3): p. 219-20.
137. Walsh, C.P. and T.H. Bestor, *Cytosine methylation and mammalian development.* Genes Dev, 1999. **13**(1): p. 26-34.
138. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
139. Gowher, H., et al., *Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L.* J Biol Chem, 2005. **280**(14): p. 13341-8.
140. Suetake, I., et al., *DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction.* J Biol Chem, 2004. **279**(26): p. 27816-23.
141. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.
142. Webster, K.E., et al., *Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis.* Proc Natl Acad Sci U S A, 2005. **102**(11): p. 4068-73.
143. Arima, T., et al., *Loss of the maternal imprint in Dnmt3Lmat-/- mice leads to a differentiation defect in the extraembryonic tissue.* Dev Biol, 2006. **297**(2): p. 361-73.
144. Hermann, A., S. Schmitt, and A. Jeltsch, *The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity.* J Biol Chem, 2003. **278**(34): p. 31717-21.
145. Okano, M., S. Xie, and E. Li, *Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells.* Nucleic Acids Res, 1998. **26**(11): p. 2536-40.

146. Jurkowski, T.P., et al., *Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism.* Rna, 2008. **14**(8): p. 1663-70.
147. Rai, K., et al., *Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish.* Genes Dev, 2007. **21**(3): p. 261-6.
148. Jones, P.L., et al., *Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription.* Nat Genet, 1998. **19**(2): p. 187-91.
149. Le Guezennec, X., et al., *MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties.* Mol Cell Biol, 2006. **26**(3): p. 843-51.
150. Sarraf, S.A. and I. Stancheva, *Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly.* Mol Cell, 2004. **15**(4): p. 595-605.
151. Hendrich, B. and A. Bird, *Identification and characterization of a family of mammalian methyl-CpG binding proteins.* Mol Cell Biol, 1998. **18**(11): p. 6538-47.
152. Yoon, H.G., et al., *N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso.* Mol Cell, 2003. **12**(3): p. 723-34.
153. Shen, J.C., W.M. Rideout, 3rd, and P.A. Jones, *The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA.* Nucleic Acids Res, 1994. **22**(6): p. 972-6.
154. Pfeifer, G.P., *Mutagenesis at methylated CpG sequences.* Curr Top Microbiol Immunol, 2006. **301**: p. 259-81.
155. Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development.* Nature, 2007. **447**(7143): p. 425-32.
156. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.
157. Stancheva, I. and R.R. Meehan, *Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos.* Genes Dev, 2000. **14**(3): p. 313-27.
158. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
159. Lei, H., et al., *De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells.* Development, 1996. **122**(10): p. 3195-205.
160. Ferguson-Smith, A.C. and J.M. Greally, *Epigenetics: perceptive enzymes.* Nature, 2007. **449**(7159): p. 148-9.
161. Neumann, B., P. Kubicka, and D.P. Barlow, *Characteristics of imprinted genes.* Nat Genet, 1995. **9**(1): p. 12-3.
162. Kaneda, M., et al., *Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting.* Nature, 2004. **429**(6994): p. 900-3.
163. Jelinic, P., J.C. Stehle, and P. Shaw, *The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation.* PLoS Biol, 2006. **4**(11): p. e355.
164. Nakamura, T., et al., *PGC7/Stella protects against DNA demethylation in early embryogenesis.* Nat Cell Biol, 2007. **9**(1): p. 64-71.

165. Hirasawa, R., et al., *Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development.* Genes Dev, 2008. **22**(12): p. 1607-16.
166. Seki, Y., et al., *Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice.* Development, 2007. **134**(14): p. 2627-38.
167. Herzing, L.B., et al., *Xist has properties of the X-chromosome inactivation centre.* Nature, 1997. **386**(6622): p. 272-5.
168. Rougeulle, C., et al., *Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome.* Mol Cell Biol, 2004. **24**(12): p. 5475-84.
169. Heard, E. and C.M. Disteche, *Dosage compensation in mammals: fine-tuning the expression of the X chromosome.* Genes Dev, 2006. **20**(14): p. 1848-67.
170. Okamoto, I., et al., *Epigenetic dynamics of imprinted X inactivation during early mouse development.* Science, 2004. **303**(5658): p. 644-9.
171. Peng, J.C. and G.H. Karpen, *Epigenetic regulation of heterochromatic DNA stability.* Curr Opin Genet Dev, 2008. **18**(2): p. 204-11.
172. Peters, A.H., et al., *Partitioning and plasticity of repressive histone methylation states in mammalian chromatin.* Mol Cell, 2003. **12**(6): p. 1577-89.
173. Lehnertz, B., et al., *Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin.* Curr Biol, 2003. **13**(14): p. 1192-200.
174. Chen, T., N. Tsujimoto, and E. Li, *The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin.* Mol Cell Biol, 2004. **24**(20): p. 9048-58.
175. Pearson, C.E., K. Nichol Edamura, and J.D. Cleary, *Repeat instability: mechanisms of dynamic mutations.* Nat Rev Genet, 2005. **6**(10): p. 729-42.
176. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
177. Girard, L. and M. Freeling, *Regulatory changes as a consequence of transposon insertion.* Dev Genet, 1999. **25**(4): p. 291-6.
178. Slotkin, R.K. and R. Martienssen, *Transposable elements and the epigenetic regulation of the genome.* Nat Rev Genet, 2007. **8**(4): p. 272-85.
179. Bourc'his, D. and T.H. Bestor, *Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L.* Nature, 2004. **431**(7004): p. 96-9.
180. Liang, G., et al., *Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements.* Mol Cell Biol, 2002. **22**(2): p. 480-91.
181. Girard, A. and G.J. Hannon, *Conserved themes in small-RNA-mediated transposon control.* Trends Cell Biol, 2008. **18**(3): p. 136-48.
182. Bhattacharya, S.K., et al., *A mammalian protein with specific demethylase activity for mCpG DNA.* Nature, 1999. **397**(6720): p. 579-83.
183. Hendrich, B., et al., *Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development.* Genes Dev, 2001. **15**(6): p. 710-23.

184. Barreto, G., et al., *Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation.* Nature, 2007. **445**(7128): p. 671-5.
185. Schmitz, K.M., et al., *TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation.* Mol Cell, 2009. **33**(3): p. 344-53.
186. Engel, N., et al., *Conserved DNA methylation in Gadd45a(-/-) mice.* Epigenetics, 2009. **4**(2).
187. Jin, S.G., C. Guo, and G.P. Pfeifer, *GADD45A does not promote DNA demethylation.* PLoS Genet, 2008. **4**(3): p. e1000013.
188. Rai, K., et al., *DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45.* Cell, 2008. **135**(7): p. 1201-12.
189. Tahiliani, M., et al., *Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by the MLL Fusion Partner TET1.* Science, 2009.
190. Kriaucionis, S. and N. Heintz, *The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain.* Science, 2009.
191. Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.* Nat Genet, 2005. **37**(8): p. 853-62.
192. Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.
193. Mikkelsen, T.S., et al., *Dissecting direct reprogramming through integrative genomic analysis.* Nature, 2008. **454**(7200): p. 49-55.
194. Mohn, F. and D. Schubeler, *Genetics and epigenetics: stability and plasticity during cellular differentiation.* Trends Genet, 2009.
195. Feldman, N., et al., *G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis.* Nat Cell Biol, 2006. **8**(2): p. 188-94.
196. Yamanaka, Y., et al., *Cell and molecular regulation of the mouse blastocyst.* Dev Dyn, 2006. **235**(9): p. 2301-14.
197. Niwa, H., *How is pluripotency determined and maintained?* Development, 2007. **134**(4): p. 635-46.
198. Orkin, S.H., et al., *The Transcriptional Network Controlling Pluripotency in ES Cells.* Cold Spring Harb Symp Quant Biol, 2008.
199. Kim, J., et al., *An extended transcriptional network for pluripotency of embryonic stem cells.* Cell, 2008. **132**(6): p. 1049-61.
200. Chen, X., V.B. Vega, and H.H. Ng, *Transcriptional Regulatory Networks in Embryonic Stem Cells.* Cold Spring Harb Symp Quant Biol, 2008.
201. Lee, T.I., et al., *Control of developmental regulators by Polycomb in human embryonic stem cells.* Cell, 2006. **125**(2): p. 301-13.
202. Ura, H., et al., *STAT3 and Oct-3/4 control histone modification through induction of Eed in embryonic stem cells.* J Biol Chem, 2008. **283**(15): p. 9713-23.
203. Ng, R.K., et al., *Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5.* Nat Cell Biol, 2008. **10**(11): p. 1280-90.
204. Gingras, A.C., R. Aebersold, and B. Raught, *Advances in protein complex analysis using mass spectrometry.* J Physiol, 2005. **563**(Pt 1): p. 11-21.

205. Jin, J., et al., *A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1.* Mol Cell, 2006. **23**(5): p. 709-21.
206. Puig, O., et al., *The tandem affinity purification (TAP) method: a general procedure of protein complex purification.* Methods, 2001. **24**(3): p. 218-29.
207. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.
208. Yang, P., H.M. Sampson, and H.M. Krause, *A modified tandem affinity purification strategy identifies cofactors of the Drosophila nuclear receptor dHNF4.* Proteomics, 2006. **6**(3): p. 927-35.
209. Wang, H., et al., *Role of histone H2A ubiquitination in Polycomb silencing.* Nature, 2004. **431**(7010): p. 873-8.
210. Minsky, N., et al., *Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells.* Nat Cell Biol, 2008. **10**(4): p. 483-8.
211. Campalans, A., et al., *XRCC1 interactions with multiple DNA glycosylases: a model for its recruitment to base excision repair.* DNA Repair (Amst), 2005. **4**(7): p. 826-35.
212. Nimura, K., et al., *Dnmt3a2 targets endogenous Dnmt3L to ES cell chromatin and induces regional DNA methylation.* Genes Cells, 2006. **11**(10): p. 1225-37.
213. Gallais, R., et al., *Deoxyribonucleic acid methyl transferases 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation.* Mol Endocrinol, 2007. **21**(9): p. 2085-98.
214. Bibel, M., et al., *Differentiation of mouse embryonic stem cells into a defined neuronal lineage.* Nat Neurosci, 2004. **7**(9): p. 1003-9.
215. Zhu, B., et al., *The human PAF complex coordinates transcription with events downstream of RNA synthesis.* Genes Dev, 2005. **19**(14): p. 1668-73.
216. Kim, J., et al., *RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells.* Cell, 2009. **137**(3): p. 459-71.

7 Acknowledgements

First, I would like to thank Primo Schär, for giving me the opportunity to accomplish my PhD studies in his laboratory. I specially thank him for believing in me, for his immense support and for always having the door open for questions and critical and fruitful discussions.

Further I would like to thank Christophe Kunz, Yusuke Saito, David Schürmann and Olivier Fritsch for their scientific support during the past years and for encouraging me during hard times of the thesis.

I also want to express my thanks to all current and past members of the Schär laboratory, in particular to Frauke Focke, my lab neighbor for many years. We "survived" to hard times by encouraging each other, and ended up by having a good and successful time. I also want to thank our new lab member, Annika Wirz, for bringing a great atmosphere in and outside the lab during the writing part of my thesis.

I would like to acknowledge Prof. Martin Spiess and Prof. Antonius Rolink as members of my thesis committee, as well as Prof. Patrick Matthias as the examinations chairman.

Very special thanks go to Maja, for her friendship, her love, for her support and immeasurable encouragement during this thesis.

Finally, I want to thank my parents, for their great support during all this years. All this would not have been possible without their help.

Appendix

Appendix I: Manuscript: Protein Interaction Studies Reveal a Tight Association of Thymine DNA Glycosylase With Epigentic Regulators in Chromatin

Daniel Cortázar, Yusuke Saito, Daniel Hess, Primo Schär

Appendix II: Manuscript: Thymine DNA Glycosylase Associates with CpG Islands to maintain Their Epigenetic Function

Daniel Cortázar, Yusuke Saito, Christophe Kunz, Fabio Mohn, Fredy Siegrist, Dirk Schübeler, Primo Schär

Appendix III: Published Review: The Enigmatic Thymine DNA Glycosylase

Daniel Cortázar, Christophe Kunz, Yusuke Saito, Roland Steinacher, Primo Schär

DNA Repair (Amst), 2007. **6**(4): p. 489-504.

Protein Interaction Studies Reveal a Tight Association of Thymine DNA Glycosylase With Epigenetic Regulators in Chromatin

Daniel Cortázar[1], Yusuke Saito[1], Daniel Hess[2], Primo Schär[1]*

[1]Department of Biomedicine

University of Basel

Mattenstrasse 28

4058 Basel

Switzerland

[2]Friedrich Miescher Institute for Biomedical Research

Maulbeerstrasse 66

4058 Basel

Switzerland

* Corresponding author: primo.schaer@unibas.ch

Phone: +41 61 267 0767

Fax: +41 61 267 3566

Protein Interaction Studies Reveal a Tight Association of Thymine DNA Glycosylase With Epigenetic Regulators in Chromatin

Daniel Cortázar[1], Yusuke Saito[1], Daniel Hess[2], Primo Schär[1]*

[1]Department of Biomedicine

University of Basel

Mattenstrasse 28

4058 Basel

Switzerland

[2]Friedrich Miescher Institute for Biomedical Research

Maulbeerstrasse 66

4058 Basel

Switzerland

* Corresponding author: primo.schaer@unibas.ch

Phone: +41 61 267 0767

Fax: +41 61 267 3566

Word count:

Introduction, Results and Discussion: 3698

Materials and Methods: 3289

ABSTRACT

Thymine DNA glycosylase (TDG) initiates excision repair of a variety of DNA base lesions generated by deamination, oxidation or alkylation. Owing to its substrate preferences, however, TDG appears to play a major role in the maintenance of CpG dinucleotides by excising U and T from G/U and G/T mismatches that arise in DNA from deamination of C and 5meC, respectively. TDG has also been implicated in the regulation of gene expression via physical and functional interactions with transcription factors, histone modifying enzymes and *de novo* DNA methyltransferases. To address mechanistic aspects of the function of TDG, we made use of biochemical approaches to isolate and characterize TDG containing protein complexes from mammalian cells. We established protocols for fractionation and purification of active TDG protein from different cell types. Surprisingly, although TDG was frequently isolated as interaction partner of other proteins in yeast two hybrid assays, we found that affinity purification of TDG itself from mammalian cells yields no copurifying factors. Nevertheless, gel filtration experiments show that protein complexes comprising TDG may form in embryonic stem (ES) cells, and we further show that a significant fraction of nuclear TDG associates tightly with chromatin together with known interaction partners, such as *de novo* DNA methyltransferases and the RNA helicase p68.

INTRODUCTION

Thymine DNA glycosylase (TDG) excises a variety of DNA base lesions from double stranded DNA, including derivatives of uracil (e.g. 5-fluorouracil), ethenoadducts (e.g 3,N^4-ethenocytosine), deaminated purines (e.g. hypoxanthine) and oxidized pyrimidines (e.g. thymine glycol), thereby initiating DNA base excision repair (BER). Its preferred biological substrates, however, are uracil (U) and thymine (T) mispaired with guanine (G), suggesting a role of TDG in the avoidance of C to T transition mutations following deamination of C and 5-meC, respectively [1-7].

However, TDG was frequently found in protein interaction screens as partner of proteins implicated in gene and chromatin regulation [8-13]. Interactions with the *de novo* DNA methyltransferases Dnmt3a/3b associated TDG with processes of DNA methylation [14-16]. Recent evidence further indicates that demethylation of CpG in promoters of estrogen-responsive genes involves the coordinated recruitment of Dnmt3a/b, p68, TDG and downstream acting BER proteins during activation of gene transcription [17]

Available experimental evidence thus suggests that TDG, as part of a macromolecular assembly, might be targeted to CpG-rich gene regulatory regions to counteract the mutagenic effect of C and 5-meC deamination or to maintain specific cytosine methylation patterns in coordination with Dnmt3a/3b. In this context, TDG may act both as a DNA glycosylase to process eventually arising substrates or as a scaffold factor in protein complexes, thereby promoting complex assembly, stability and/or activity.

Thus, to gain novel insights into the biological function of TDG, we attempted to isolate and characterize TDG containing complexes present in cells under native conditions.

In a first attempt, we established and performed tandem affinity purification (TAP) of TDG followed by Mass Spectrometry (MS) analysis. We show that TAP-tagged TDG can be purified from mammalian cells using the original procedure. However, MS analysis showed that the system failed to isolate stoichiometrically co-purifying proteins. Similarly, immunoprecipitation of TDGa from lysates of TdgA

overexpressing *Tdg*$^{-/-}$ embryonic stem (ES) cells did not identify any TDG specific interacting partners. Nevertheless, classical gel filtration of ES cell extracts fractionated TDG in protein complexes of distinct molecular weights. The RNA helicase p68, a previously identified TDG interaction protein [17-19], cofractionated with TDG in some elution peaks.

Cell fractionation experiments revealed that TDG is present in all cellular compartments, with a significant fraction of nuclear TDG being tightly associated with chromatin. Immunoprecipitation and affinity purification of endogenous and TAP-tagged TDG, respectively, from salt-eluted chromatin-associated fractions, however did not reveal copurifying interacting partners, suggesting that TDG containing complexes might be disrupted at high salt concentrations. Fractionation of P19 cells revealed that insoluble chromatin bound TDG can be efficiently solubilized, together with *de novo* DNA methyltransferases, after micrococcal nuclease digestion of DNA. Strikingly, even after efficient DNA fragmentation, TDG was still found associated with chromatin, as were *de novo* DNA methyltransferases and p68. Co-fractionation with these proteins further corroborates the hypothesis of TDG, Dnmt3a/b and p68 being part of a chromatin bound protein complex.

MATERIALS AND METHODS

Antibodies

For Western Blot analyses, we used the following antibodies: rabbit anti-mTDG (raised against recombinant full length mouse TDG), mouse monoclonal anti-p68 (Upstate, USA), mouse monoclonal anti-PCNA (Leinco Technologies, USA), rabbit anti-Dnmt3a (Abcam, UK) and rabbit anti-Dnmt3b (Abcam, UK). For TDG immunoprecipitations we used rabbit anti-mouse and human TDG antibodies that were affinity purified with full-length recombinant mouse and human TDG proteins, respectively.

Cell culture

Mouse Embryonic Fibroblasts (MEFs) were cultured in the MEF Medium (DMEM, 10%FBS, 2mM L-Glutamine) at 37°C in a humidified atmosphere containing 5% carbon dioxide. Cells at 80% confluence were washed once with 1x PBS [pH 7.4] and treated with 1x Trypsin-EDTA solution at 37°C for 5min. Cell were then harvested after addition of 5 volumes of cell culture medium, counted in a Neubauer Hemacytometer and diluted 1:20 into a new culture dish.

Mouse Embryonic Stem (ES) cells were grown on primary mouse embryonic fibroblasts (feeder cells) in ES cell Medium (DMEM, 15% FBS, 2mM L-Glutamine, 1mM Na-Pyruvate, 1x MEM non-essential Amino Acids, 90 μM β-mercapthoethanol, 10^3 units/ml ESGRO® (LIF) (Millipore, USA)). Feeder cells were prepared as follows:

Primary MEFs from 13.5 ~ 14.5 dpc embryos are thawed and plated on 100 mm culture dishes. When confluent, cells were irradiated with γ-rays (50Gy). 24 hours after irradiation, cells were trypsinized and frozen with MEF medium + 10% DMSO at a density of 5×10^6 cells per ml.

Prior to culture of ES cells, feeder cells were thawed and plated on two 35mm-dishes (5×10^5 cells/dish). After 24 hours, frozen ES cells (2×10^6 cells) were quickly thawed at 37 °C and transferred to a 15 ml falcon tube containing 2 ml of ES medium. After centrifugation at 700 × *g* for 5 min at room temperature (RT), the supernatant was removed and cells were resuspended in 2 ml fresh ES medium and plated on feeder cells. In the case cell showed less than 50% confluence, ES medium was changed after 24 hours, otherwise they were passaged. Cells were washed once with PBS prior to the addition of 500 μl 0.25% Trypsin-EDTA (Invitrogen, USA) and incubated for 3 min at 37°C. Cell clumps were then gently detached by pipetting up and down 3-5 times. Trypsin was inactivated by the addition of 2 ml ES cell medium and cells were split at ratios ranging from 1:5 to 1:10 into fresh plates containing feeder cells. Low-density plating was avoided in order to keep homogenous undifferentiated cell cultures.

For biochemical experiments with ES cells, feeder cells were removed in order to avoid genotype/phenotype contaminations. For this purpose ES cells and feeder cells were plated simultaneously. After 30 min, most of feeder cells attached to the cell culture dishes, while ES cells were still in suspension. ES cells in suspension were

transferred to a new culture dish. To ensure pure ES cell cultures this procedure was repeated for one additional passage.

Stable transfection of MEFs

50% confluent cultures in 10 cm dishes were transfected with 5 µg linearized vector DNA using 6µl Transfectin Reagent (BioRad, USA). The efficiency of vector delivery was assessed by transfection of pEGFP under identical conditions and quantitation of GFP positive cells was done by fluorescence microscopy. Selection of stable clones was done by growing cells in the presence of 1.5 µg/µl puromycin for 1 week.

Preparation of total cell lysates

Cells grown in 10 cm dishes were harvested at 90% confluency by gentle trypsinization and centrifugation at 2000 rpm at 4°C for 5min. After three washes with ice cold 1x PBS [pH 7.4], cells were lysed in NP-40 lysis buffer (50mM Na-phosphate [pH 8.0], 125 mM NaCl, 1% NP-40, 0.5 mM EDTA, 1 mM PMSF, 1 mM DTT, 1x Complete™ protease inhibitors (Roche, Switzerland) for 30min on ice. For immunoprecipitation experiments from ES cells, whole cell extracts were further sonicated for 5 minutes (15sec on 30sec off, HIGH) using a BioRuptor (Diagenode, Belgium). Extracts were then clarified by centrifugation (15 min, 20'000g, 4°C).

Western Blotting

Soluble protein extracts were separated in SDS-polyacrylamide gels and transferred to a nitrocellulose membrane (Millipore, USA) according to standard prcedures. To assess the quality of transfer, membranes were stained with Ponceau staining solution (Sigma-Aldrich, USA). Membranes were washed once with water and once with TBS-T (100 mM Tris/HCl [pH 8.0], 150 mM NaCl, 0.2% Tween20). Membranes were then incubated with blocking buffer (TBS-T, 5% dry milk) for 1 h at room temperature. Membranes were incubated with the primary antibody for 1 h at 33 °C (anti-mTDG; 1:10'000) or RT (anti–PCNA, 1: 5'000; anti-p68, 1: 5'000; anti-Dnmt3a/b, 1:2'000) in blocking buffer. After hybridization, membranes were rinsed three times for 1 min and washed three times for 10 min at room temperature with TBS-T.

Secondary horseradish peroxidase–conjugated antibodies (anti-rabbit IgG, GE Helathcare, UK) were diluted 1:5'000 in blocking buffer and hybridized to the membranes for 1 h at room temperature. After washing at room temperature (3x 10 min in TBS-T), detection of the signals was carried out using the Immobilon Western Chemiluminescent HRP Substrate (Millipore).

Generation of TdgA overexpressing mouse embryonic stem cells

For expression of the transactivator (tTA), a 1.7 kb *SalI–EcoRI* fragment containing the CAG promoter from the pPyCAG-IP vector was inserted into the pBluescript vector. The 1.6 kb *EcoRI-PvuII* fragment containing tTA and polyA from pTet-Off vector (Clontech, USA) was inserted downstream (*EcoRI-SmaI*) to the CAG promoter. Finally, the CAG-tTA-polyA construct (*KpnI-NotI*) was inserted into the *Kpn* I-*Not* I sites of the expression vector pCAG-Tet off Hyg (Carrying a Hygromycin resistance cassette). To generate the Tet-respon vector fro Tdg expression, the mouse TdgA cDNA (*EcoRV-NotI*) fragment was inserted into *PvuII-NotI* sites of pTRE2 vector (Clontech, USA) carrying the $P_{hCMV*-1}$ promoter. The pPyCAG-IP vector was then digested with *SalI* and *NotI* to remove the CAG promoter and the $P_{hCMV*-1}$-Tdg fragment (*XhoI-NotI*) was ligated into this site to generate the vector p$P_{hCMV*-1}$-TdgA-IP.

Tdg$^{-/-}$ ES cells were grown without feeder cells following standard procedures. Cells were trypsinized and kept in suspension in ES-medium. For the first transfection, 1 μg of pCAG-Tet-off Hyg vector in 50 μl DMEM was mixed with 2.5 μl of lipofectamine^{TM} 2000 (Invitrogen, USA) in 50 μl DMEM at room temperature. After 20 minutes incubation, 5×10^5 ES cells were gently mixed with the complexes and plated into 60 mm gelatin-coated culture dishes. After 48 hours incubation, cells were cultured with 150 μg/ml hygromycin in ES medium or 2 weeks. Hygromycin-resistance colonies were picked and further cultured with 100 μg/ml hygromycin. To confirm the expression of tTA, several independent ES clones were transiently transfected with p$P_{hCMV*-1}$-EGFP-IP vector and EGFP expression was monitored in the absence of tetracycline. One EGFP-positive clone was then used for the second transfection with the p$P_{hCMV*-1}$-TdgA-IP vector. Cells were cultured with 1.5 μg/ml

puromycin for 2 weeks and TDGa protein levels of puromycin-resistance colonies were then monitored.

Cloning of TAP tagged TdgA

Construction of the vectors pCor1 and pCor2 encoding C- and N-terminal TAP-tagged mouse TdgA, respectively was done in two steps. To obtain pCor2, the open reading frame of the mouse TdgA gene was PCR amplified from the pHH25-mTdgA plasmid (Ch.Kunz) with primers designed to give a product flanked by *EcoRI* restriction sites. Following *EcoRI* digestion, the PCR fragment was then inserted into the unique *EcoRI* site of the vector pZome-1-N (Cellzome, USA). To obtain pCor1, the TdgA gene was amplified with primers designed to give a product flanked by *BamHI* restriction sites. Digested fragments were then inserted into the unique BamHI site of the pZome-1-C vector (Cellzome, USA). Accuracy of all plasmid constructs was confirmed by DNA sequencing. Primer sequences for N-terminal tagging: cTdgA TN (F): 5'-CTT GAA TCC ATG GAC GCA GAG GCC GC-3', cTdgA TN (R): 5'-CTT GAA TT CTA AGC GTG GCT CTC TTC-3'; and for C-terminal tagging: cTdgA TC (F): 5'-CTT GGA TCC ATG GAC GCA GAG GCC GC-3', cTdgA TC (R): 5'-CTT GGA TCC AGC GTG GCT CTC TTC TTC-3'. Primers for sequencing were: mTDG885s: 5'-CCA GAC ACA GAA ACT CTG-3', mmTDGa436: 5'-GTG ACT ATT GGC ATT AAC C-3', exon1/1043: 5'-CAC CTC ACT CAG CCC CGA C-3'

Tandem affinity purification of C-and N-TAP tagged TDG

The tandem affinity purification procedure was adapted from protocols of the Seraphin Lab and the protocol of Anne-Claude Gingras [20]. The following purification steps were conducted at 4°C. Mouse embryonic fibroblasts expressing either the N-terminally or C-terminally TAP tagged TdgA (vectors pCor2 and pCor1, respectively) were grown on 150 mm plates to a confluency of 80-90%. Cells were directly lysed on the plates by addition of 750 µl of TAP-lysis buffer (20 mM HEPES-KOH [pH 8.0], 150 mM NaCl, 10% glycerol, 0.1% NP-40, 2 mM EDTA [pH 8.0], Complete™Protease Inhibitor Cocktail (Roche, Switzerland), 2 mM DTT) and 30 minutes incubation at 4°C. Cells were subsequently scraped off and the resulting homogenate was cleared by centrifugation at 14'000 rpm for 15 minutes. Protein

concentration was then determined by the Bradford Protein Assay (BioRad, Switzerland).

IgG sepharose 6 Fast Flow beads (GE healthcare, UK) (10µl packed beads per 10 mg lysate) were washed three times with 10 packed volumes of TAP-lysis buffer. Total cell lysates were then incubated with the equilibrated beads during 4 hours on a rotor shaker. After binding to IgG-Sepharose beads, the unbound fraction was recovered after centrifugation (5 min, 500 x g) and the beads were washed three times with 10 bead volumes TAP-lysis buffer and three times with TEV buffer (10 mM HEPES-KOH [pH 8.0], 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA [pH 8.0], 1 mM DTT). Washed beads were then resuspended in TEV buffer supplemented with *ac*TEV protease (Invitrogen, USA) (100U in 150µl/100µl beads), and cleavage was allowed to proceed for 4 hours at 4°C. After TEV cleavage, beads were centrifuged 5 min at 1'500 x g and 120 µl TEV eluate were transferred to a new tube. 20 µl TEV eluate was kept for Western Blot analysis. IgG beads were then washed 3 times with 120 µl Calmodulin binding buffer (10 mM HEPES-KOH [pH 8.0], 150 mM NaCl, 0.1% NP-40, 1 mM Mg-OAc, 1 mM Imidazole, 10 mM β-mercaptoethanol, 2 mM $CaCl_2$) and the washes were added to the previously collected TEV eluate, to give a total volume of 480µl.

After addition of 1/250 volumes of 1M CaCl2 the sample was incubated with 50 µl equilibrated Calmodulin Sepharose 4B (GE healthcare, UK) beads. Equilibration of beads was done by washing three times with 10 volumes of Calmodulin binding buffer. Binding of TEV eluates was allowed to proceed over night at 4°C on a rotor shaker. The beads were then washed three times with Calmodulin Binding Buffer and three times with Calmodulin Rinsing Buffer (50 mM NH_4HCO_3, 75 mM NaCl, 1 mM MgOAc, 1 mM Imidazole, 2 mM $CaCl_2$). For native elutions, calmodulin beads were incubated with Calmodulin elution buffer (50 mM NH_4HCO_3, 25 mM EGTA). For elution under denaturing conditions, beads were resuspended in standard SDS loading buffer and incubation at 95°C for 5 min. Following elution, purified fractions were analysed by SDS-PAGE and Western blotting.

Isolation of chromatin bound proteins

1 x 10^7 – 2 x 10^7 cells were washed three times with ice cold PBS [pH 7.4] and

resuspended in 200 µl Buffer A (10 mM HEPES [pH 7.9], 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M Sucrose, 10% Glycerol, 1 mM DTT, 0.1% Triton X-100, Protease inhibitor cocktail (Roche, Switzerland)). Cells were then incubated on ice for 6-8 minutes and centrifuged for 5 minutes at 1,300 x g and 4°C. The supernatant (cytoplasmic fraction) was clarified by high-speed centrifugation at 20'000 x g for 5 minutes at 4 °C. The nuclear pellet was washed twice with Buffer A and lysed for 30 minutes in 200 µl Buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, Protease inhibitor cocktail (Roche)). After centrifugation at 1,700 x g for 5 minutes at 4°C, the pellet fraction containing the insoluble chromatin fraction was washed twice with 200 µl Buffer B. For salt extraction of chromatin-associated proteins, the insoluble chromatin was resuspended in 1 v/v of Buffer C (50 mM Na-phosphate [pH 8.0], 3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, 0.5% NP-40, 10% glycerol, Complete™Protease inhibitors (Roche, Switzerland) supplemented with NaCl at concentrations from 0.1 M to 1 M and incubated on ice for 1 hour. Solubilized proteins were recovered by centrifugation at 20'000 x g for 10 min at 4°C and analyzed by SDS-PAGE and Western blotting. For micrococcal nuclease digestion, the insoluble chromatin fraction was washed three times with 200 µl MNase buffer (10 mM Tris [pH 7.5], 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 3 mM $CaCl_2$, PMSF). Chromatin associated proteins were solubilized by digestion of DNA with 60 U micrococcal nuclease (Fermentas) in 25 µl MNase buffer for 5 min at 37°C. The reaction was stopped by addition of 5 mM EDTA/2.5 mM EGTA for 10 min at 4°C. Solubilized proteins were recovered by centrifugation at 20'000 x g for 5 min at 4°C and analyzed by SDS-PAGE and Western blotting.

Immunoprecipitations

Immunoprecipitation of TDGa from ES cells: 5mg total cell extracts (see above) at a concentration of 5mg/ml in IP buffer (50 mM Na-phosphate [pH 8.0], 125 mM NaCl, 1% NP-40, 0.5 mM EDTA, 1 mM PMSF, 1 mM DTT, 1x Complete™ protease inhibitor Cocktail (Roche, Switzerland)) were first precleared for one hour with 50 µl of a 50% Dynabeads® Protein G magnetic beads (Invitrogen, USA) slurry, previously blocked with 1mg/ml BSA. Precleared lysates were then subjected to immunoprecipitation for 4 hours at 4 °C with 10 µg affinity purified anti-mouse TDG antibody. Antibody

protein complexes were recovered by 4h incubation at 4°C with 50 µl of a 50% slurry of preblocked 50% Protein G beads slurry. Beads were then washed three times with 10 bead-volumes of IP buffer and boiled at 95°C in SDS-loading buffer for SDS-PAGE and Western blotting. For MS analysis, the antibody was directly crosslinked to the beads according to the manufacturers instructions. In this case, Antibody-beads were incubated 4 hours at 4°C with total cell lysates. Immunoprecipitated material was eluted twice with 50 µl 0.1 M Glycine [pH 2.5]. Each eluate was immediately neutralized with 10 µl 1 M Tris-HCl [pH 8.0]. Pooled elutions were then precipitated through the addition of 80 µl of a 50% TCA solution and 5 minutes incubation at -20°C followed by over-night incubation at 4°C. Precipitated material was washed once with 500µl cold acetone. After centrifugation, the pellet was air-dried and frozen at -20°C until further processing.

Immunoprecipitation of chromatin bound endogenous TDG from MEFs: 4mg total cell extract at a concentration of 10 mg/ml in Buffer C containing 0.25 M or 0.5 M NaCl (see Isolation of chromatin bound proteins) were subjected to immunoprecipitation for 4 hours at 4°C with 10µl of a 50% slurry of Tosylactivated magnetic beads (Invitrogen, USA) coated with affinity purified anti-TDG antibody. Coupling of antibody to the beads was done following the manufacturers instruction using 80 µg purified antibody per 200µl of a 50% slurry of Tosylactivated beads. After immunoprecipitation, beads were washed three times with 10 bead-volumes of Buffer C with the appropriate salt concentration and boiled at 95°C in SDS-loading buffer for SDS-PAGE and Western blot analysis.

Gel filtration

Gel filtrations were performed with Superdex 200 10/300GL colums (GE Healthcare, USA)) using an ÄKTA™ Purifier 10 according to the manufacturers instructions. Total cell lysates were prepared as described above. 10 mg extract at a concentration of 10 mg/ml were dialysed three times for 2 hours at 4°C in 500 ml Dialysis Buffer (50mM Na-phosphate [pH 8.0], 0.5mM EDTA, 1 mM PMSF, 1mM DTT, 10% Glycerol) containing the desired NaCl concentration. After dialysis, extracts were centrifuged for 15 min at 14'000 x g and filtered through a 0.4 µm filter (Sarsted, Germany) to remove residual debris and cleared lysates were then loaded onto the gel filtration

column. Column washing, loading and sampling of the fractions was done according to the manufacturers instructions. 0.5 ml fractions were collected and 0.1 ml of each fraction was Methanol/Chloroform precipitated for SDS-PAGE and Western blot analysis.

Base Release Assay

For base release assays, 50 µg of whole cell extract or 1/10th of either the TEV elution, IgG bound or calmodulin bound fractions were incubated with 1 pmol of a fluorescein-labeled G/T mismatched DNA substrate in reaction buffer (50 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1 mM DTT, 1 mg/ml BSA) for 40 min at 37 °C. Resulting AP-sites were chemically 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 0.3 M Na-acetate [pH 5.2] and in the presence of 0.4 mg/ml carrier t-RNA. The DNA was collected by centrifugation (20 min, 20'000 x g, 4 °C) and washed in 80% ethanol. Air-dried pellets were resuspended in loading buffer (1× TBE, 90% formamide), heated at 95 °C for 5 min, and then immediately chilled on ice. Reaction products were separated on 15% denaturing polyacrylamide gels in 1×TBE. The fluorescein-labeled DNA was visualized with a Typhoon 9400 (GE Healthcare, UK) scanner.

Mass spectrometry analysis

The LC-MS analysis was essentially done as described in Hess et al. [21]. Shortly, precipitated proteins were solubilized in 10 µl reduction buffer (RB) (500 mM Tris-HCl [pH 8.6], 6M GnHCl) and 2µl 100 mM Tris (2-Carboxyethyl) phosphine (TCEP) in RB. Samples were flushed for 15 seconds with Argon and reduced for 30 min at 37^{o}C. 2 µl 250 mM Iodoacetamide (46 µg/µl) in RB was added and the samples were alkylated for 1 hour in the dark. 79 µl digestion buffer (50 mM Tris-HCl [pH 7.4], 5 mM $CaCl_2$, 2 µl Acetonitrile and 5 µl (0.1 µg/µl) Trypsin (Promega, USA) was added and the proteins were cleaved over night at 37°C. The generated peptides were analyzed by capillary liquid chromatography tandem mass spectrometry (LC–MSMS) using a Magic C18 100 µm x 10 cm HPLC column (Swiss BioAnalytics, Switzerland) connected on-line to a 4000Q Trap (MDS Sciex). The peptides were loaded manually

onto a peptide cap-trap (Michrom BioResources, Inc.). The trap was washed at 10 µl/min with 0.05% TFA and 2% acetonitrile in H_2O. The HPLC buffers were: A (0.1% formic acid, 2% acetonitrile in H_2O) and B (0.1% formic acid, 80% acetonitrile in H_2O). A linear gradient from 0% to 10% B in A (3 min) followed by 10% to 40% B (80 min) and 40% to 55% B (15 min) was delivered with an 1100 Nano-HPLC system (Agilent) at 500 nl/min. The eluting peptides were ionized by electrospray ionization and the peptide ion masses were measured in the linear ion trap. The detected ions were automatically selected in Quadrupol 1, fragmented in Quadrupol 2 and the generated ions were measured in the linear ion trap. Individual MSMS spectra, containing sequence information for a single peptide, were compared against the protein sequence database UNIPROT 14.0 using Mascot software.

For the analysis of TAP-tagged purified fractions, gel slices were washed once with 50 µl 25 mM NH_4HCO_3 for 30 min (under agitation) and twice with 50 µl of 50% 25 mM NH_4HCO_3/ 50% acetonitrile for 30 min (under agitation). Gel slices were then resuspended in 100µl 1mM TECP and incubated for 30 minutes at room temperature. Liquid was then removed and the gel slices were resuspended in 100µl 20mM Iodoacetamide and incubated for 45 minutes at room temperature in the dark. Gel slices were then washed once with 50 µl 25 mM NH_4HCO_3 for 30 min and twice with 50 µl of 50% 25 mM NH_4HCO_3/ 50% acetonitrile for 30 min as described above and subjected to in-gel trypsin digestion. The peptides generated were analyzed as described before.

RESULTS

Tandem Affinity Purification of TDG protein complexes

Tandem Affinity Purification (TAP) has been successfully used for purifications of protein complexes on a large scale in yeast and mammals [20, 22-25]. The method involves fusion of a TAP-tag to a protein of interest, which is then expressed as a fusion protein in a suitable cell type. The TAP-tag consists of two distal Immunoglobulin G (IgG) binding units of *Staphylococcus aureus* protein A (Prot A), a cleavage site for tobacco etch virus (TEV) protease and a proximal calmodulin-

binding peptide (CBP) domain. Thus, two sequential affinity purification steps allow specific recovery of the fusion protein and interacting components (Fig. 1A). Due to the high specificity, the TAP procedure allows protein purification under physiological conditions, thereby preserving fragile multiprotein complexes.

The pZome-1N and pZome-1C (Euroscarf, Germany) plasmids were used as backbones for the construction of the vectors encoding N- and C-terminal TAP tagged TDGa. To obtain the N-terminally tagged variant, the open reading frame of mouse TdgA was PCR amplified from the pHH25-mTdgA plasmid with primers designed to give a product flanked by *EcoRI* restriction sites. The PCR fragment was then inserted into the unique *EcoRI* site of the vector pZome-1N to give rise to the vector p-N-TAP-TdgA. To obtain the C-terminally tagged TDG, the TdgA gene was amplified with primers designed to give a product flanked by *BamHI* restriction sites and fragments were then inserted into the unique *BamHI* site of the pZome-1C vector to give rise to the vector p-C-TAP-TdgA. We then stably transfected $Tdg^{-/-}$ knockout mouse embryonic fibroblasts (MEFs) with the two vectors and examined the expression of the fusion proteins of 75kDa for the N-TAP-TDG and 85kDa for the C-TAP-TDG by Western blot analysis of total cell lysates with a polyclonal anti-mouse TDG antiserum (Fig.1C). Two bands corresponding to 85 and 75kDa were detectable for the C-terminally TAP-tagged TDG that might indicate a truncation or processing of the native N-terminus of TDG. Note that the expression levels of the tagged proteins are comparable to endogenous levels in $Tdg^{+/+}$ MEFs, which is preferable to minimize the co-purification of false positive interactions. Total cell lysates of stable cell clones expressing endogenous levels of N- or C-TAP-TDG were made and then subjected to the Tandem Affinity Purification. 50 to 100 mg of lysate containing about 12.5 to 25µg of TDG protein (Fig. S1) were incubated with IgG beads as described in the methods section. About 10% of C- and N-TAP-tagged TDG protein was usually recovered after binding to IgG beads (Fig. 2A, lanes 4). Following TEV mediated proteolytic cleavage most of specifically IgG-bound material could be released from the IgG beads (Fig. 2A, lanes 7). Yet, 50-60% of released protein bound unspecifically to the IgG sepharose beads (Fig 2A, lanes 6, arrows). Note that only one isoform of the C-terminally TAP tagged TDG is being eluted. TEV eluted material

was found to bind efficiently to calmodulin beads in the second purification step allowing recovery of 30-50% of TEV eluate. About 125ng of TDG were eluted in the experiment shown in figure 2A, which gives an overall TDG recovery of about 0.5% after TAP purification. Elution from calmodulin beads was not successful under mild conditions using EGTA (Fig. 2A, lane 10). By contrast, applying denaturing conditions efficiently released 100% of the total calmodulin bound TDG (Fig. 2A, lane 9), suggesting that a large fraction of TDG binds unspecifically to the calmodulin sepharose matrix.

To test whether the TAP tagging affects the enzymatic activity of TDG, we assayed G/T Glycosylase activity in the TAP fractions of N- and C-terminally tagged TDG. Making use of a well-established base release assay with a synthetic G/T DNA substrate, we were able to detect thymine excision activity in the TEV eluates and in the calmodulin bound fractions (Fig. 2B). To address the enrichment of specific thymine DNA glycosylase activity during purification, we compared the total cell lysates (Fig.2B, lanes 3 and 6) with 1/10 of TEV the eluates and 1/5 of calmodulin bound proteins (Fig.2B, lanes 4+7 and 5+8, respectively) by quatifying the intensity of the product band. Clearly, the purified fractions after TEV cleavage and calmodulin binding showed an enrichment of specific activity when compared to input lysates, indicating successful TDG purification. Also, unspecific cleavage products generated by unspecific nucleases in total cell extracts are not detected after TAP purification, indicating the removal of nuclease contaminants upon TAP. By comparing the base release activity of the TAP purified fractions with that of recombinant TDG protein (Fig.2B, lane 2), we saw that the lower activity of the purified fractions correlates with a lower amount of TDG protein in these fractions. Together, these results show that TAP tagging of TDG does not affect its glycosylase activity on a G/T DNA substrate.

To identify putative TAP-TDG copurifying proteins, we performed by capillary liquid chromatography tandem mass spectrometry (LC-MSMS) analysis of eluates after TEV cleavage (Fig.2A, lane 7) and after denaturing elution from calmodulin sepharose beads (Fig.2A lane 9). As a control, we used equivalent elutions from a purification performed from extracts of cells expressing the TAP vector without TdgA insert. LC-MSMS analysis failed to detect specific TDG associated proteins or protein

complexes, presumably due to a low yield of purified material. However, attempts to upscale the procedures and optimizing all purifications steps were successful in terms of TDG recovery, but still did not identify any copurifying proteins. These results show that TDG can be purified from cell extracts using the TAP method. Nevertheless, bait purification seems to yield levels of TDG and copurifying proteins below the detection limit of MS identification.

Identification of TDG protein complexes in ES cells

Because of the developmental defect of *Tdg* deficient mouse embryos and the implicated role of TDG in gene regulatory processes associated with cell differentiation, we decided to investigate protein interactions in mouse ES cells.

For this purpose, we generated *Tdg*$^{-/-}$ ES cells expressing *TdgA* under the control of an inducible promoter, making use of the Tet-Off system [26]. For this purpose we generate *Tdg*$^{-/-}$ knockout ES cells stably expressing both the Tet-responsive transcriptional activator (tTA) from the CMV/actin/globin (CAG) promoter, and *TdgA* under the control of the Tet-responsive $P_{\text{hCMV*-1}}$ promoter. In this system, tTA induces transcription of the transgene in the absence of Tet by binding to the Tet-response element upstream of the $P_{\text{hCMV*-1}}$ promoter (Fig. 3A). Western blot analysis confirmed that *Tdg*$^{-/-}$ knockout ES cells expressed the TdgA transgene at levels up to 100-fold higher than *Tdg* wild-type ES cells (Fig. 3B). In addition, base release assays further showed that total cell extracts from TdgA overexpressing ES cell are eight to ten times more active on a synthetic G/T DNA substrate than total cell lysates isolated from *Tdg*$^{+/-}$ ES cells (Fig. 3C).

We next performed immunoprecipitation (IP) of TDGa from total cell lysates of *TdgA* overexpressing *Tdg*$^{-/-}$ knockout ES cells. Extracts from *Tdg*$^{-/-}$ knockout ES cells expressing EGFP were used as control. At physiological salt concentration, TDGa was efficiently and specifically immunoprecipitated (Fig. 4A). Staining of IP fractions after SDS-polyacrylamide electrophoresis using a highly sensitive Coomassie Blue staining solution revealed minor differences in protein band patterns between precipitates from *TdgA* overexpressing cells and EGFP control cells, and a control IP without specific antibody (Fig. 4B, asterisks). However, because TDG-specific and unique protein bands were difficult to detect, we did LS-MSMS analysis of the total eluate of

the IPs, containing about 50 to 100ng of TDGa, to obtain as much information as possible. To this end, total eluates were digested in-solution by trypsin and the peptides generated were analyzed by LC–MSMS. Individual MSMS spectra, containing sequence information for a single peptide, were compared against the protein sequence database UNIPROT 14.0 using Mascot software. Although MS analysis confirmed the presence of TDGa only in immunoprecipitated samples from TdgA expressing cells, no TDG specific interacting proteins could be identified.

These data and the negative results from the TAP tagging approach indicate that the fraction of TDG that interacts with other proteins in a complex is inaccessibly enclosed in protein complexes. Indeed, although TDG has many interaction partners, all of them were identified in screens with the partner protein under conditions overexpressing both proteins. For this reason, we examined whether TDG forms part of macromolecular assemblies in ES cells. For this purpose, we analyzed TDG complexes in total cell lysates of *TdgA* overexpressing ES cells using size exclusion chromatography. At physiological salt concentrations (e.g. 125 mM NaCl), TDG eluted in four main fractions ranging from 17 kDa to 670 kDa (Fig.5A, I-IV). Further gel filtrations at 150mM and 300mM NaCl showed that the broad peak fraction at 75-158 kDa (Fig.5A, I) can be further fractionated into two distinct peaks at 140-158 kDa and 60-75 kDa (Fig.5B, III).

We also monitored elution of the TDG interacting protein p68 (Fig. 5B). At 150 mM NaCl, p68 coeluted with TDG at 140-180 kDa but showed a main peak at 40-60 kDa. At 200 and 300 mM NaCl however, p68 eluted only at 40-60 kDa. Co-elution of TDG and p68 at low salt concentrations might reflect the presence of both proteins as part of a common protein complex. These findings suggest that TDG is part of high molecular weight protein complexes in ES cells. These appear to be differentially resistant to high salt concentrations.

Protein complexes including TDG are bound to chromatin

Next, we investigated the subcellular distribution of TDG by biochemical cell fractionation of mouse ES cells (Fig. 6A). We found TDG to be present in the nuclear compartments of the cell. Interestingly, a significant fraction of nuclear TDG was associated with the insoluble chromatin fraction, a phenomenon previously

observed for *de novo* Dnmts [27] but not expected for a DNA glycosylase. We also observed that TDG can be partially eluted from the chromatin by increasing salt concentrations (Fig. 6A, lanes 3 to 8). Chromatin bound TDG is eluted upon treatment of chromatin with 200 and 300 mM NaCl (Fig. 6A, lanes 4 and 5). A fraction of the remaining chromatin bound TDG is further eluted at 500mM NaCl (Fig. 6A, lane 7). Interestingly, a significant amount of TDG remained bound to chromatin even after high salt elution up to 1M NaCl (Fig. 6A, lane 9). These experiments suggest that TDG is tightly associated with chromatin, to which it appears to be constitutively targeted by interactions with transcription factors.

To identify proteins interacting with chromatin bound TDG we first eluted chromatin bound TDG from $Tdg^{+/+}$, $Tdg^{-/-}$ MEFs and $Tdg^{-/-}$ MEFs expressing the C-TAP TDG by incubation of the insoluble chromatin fraction with either 250 mM or 500 mM NaCl (Fig.6B). Eluted endogenous TDG from $Tdg^{+/+}$ and $Tdg^{-/-}$ MEFs was then subjected to immunoprecipitation with a polyclonal anti-TDG antibody. In parallel, eluted chromatin bound C-TAP-TDG was allowed to bind to IgG sepharose beads. Endogenous TDG can be efficiently immunoprecipitated from chromatin fractions eluted at both 250 and 500 mM NaCl (Fig.6B lanes 3 and 9). Similarly, we see efficient binding of chromatin eluted C-TAP-TDG to IgG sepharose beads (Fig.6C, lanes 3 and 9). These results show that native and TAP-tagged TDG can be purified from chromatin after salt elution. However, immunoprecipitation of native TDG did not show a specific enrichment of copurifying proteins in $Tdg^{+/+}$ cells. Purification of chromatin bound C-TAP-TDG revealed differences in protein band patterns after SDS-PAGE and Coomassie Blue staining, when compared to purifications performed from extracts of cells stably expressing the C-TAP vector without TdgA insert. The differences however, turned out to be difficult to reproduce.. Although the bands between 75 and 250kDa represent TDG interacting proteins, their irreproducibility suggests that interactions with TDG might be randomly rearranged at high salt concentrations (Fig.6C, lanes 3 and 6, asterisks). The prominent band at 45kDa appears to a degradation product of the Ig heavy chain (Fig.6C, lanes 3, 6 and 13 crosses). Together, these data suggest that TDG interacts with other proteins or protein complexes in chromatin. These interaction, however, appear to be destroyed at high salt concentrations.

TDG has been shown to interact with Dnmt3a/b during activation of gene transcription in mouse P19 teratocarcinoma cells [16, 17]. We thus decided to address the subcellular localization of TDG in these cells. As observed in ES cells, we find TDG in the cytoplasma (25% of total) and in the nucleus (Fig.7A), where and a significant fraction of TDG (about 30%) is associated with chromatin. Because high salt extraction seems to disrupt TDG containing chromatin complexes (Fig. 6), we extracted TDG from the chromatin by digestion of the DNA with micrococcal nuclease (MNase) (Fig.7A, lane 5). MNase treatment liberated 60% of chromatin bound TDG, together with parts of chromatin bound Dnmt3a/3b. Yet, even after efficient MNase treatment (Fig. 7B) some TDG remained attached to insoluble chromatin, again together with Dnmt3b, a lower amount of Dnmt3a (Fig. 7A, lane 7) and almost all of p68. Together, these experiments suggested that a significant fraction of TDG is tightly associated with chromatin to which it might be constitutively targeted by interactions with transcription factors and/or *de novo* DNA methyltransferases. Co-fractionation with p68 and Dnmt3a/b corroborates our gel filtration results and further supports the hypothesis of TDG being part of a macromolecular assembly including Dnmt3a/b and p68 as proposed by others [17].

DISCUSSION

To gain novel insights into the biological function of TDG, we tried to identify and characterize TDG containing protein complexes in different cell models. Making use of various biochemical fractionation methods, we realized that while TDG can be enriched, co-purifying proteins are difficult to detect. This came as a surprise, since TDG was previously reported to interact with a number of proteins; including mainly transcription factors but also DNA repair proteins, *de novo* DNA methyltransferases and chromatin remodeling factors [7, 14, 15, 17, 28].

We first performed purification of TDG from mouse embryonic fibroblast by the tandem affinity purification (TAP) method, originally developed in yeast for the analysis of protein interactions under physiological conditions. We made use of the originally decribed TAP tag consisting of two IgG binding domains of Protein A, a TEV protease cleavage site and a calmodulin binding peptide (CBP) domain [20, 22-25]. Despite several attempts to optimize the TAP procedure, an isolation of co-purifying

TDG interacting proteins was not successfull. Although we started with large amounts of total cell extract, the overall TDG yield after purification lied around 0.5%. Most of TDG is lost during purification on the IgG sepahrose matrix. First, only 1/10 of total input TDG was recovered after binding to IgG beads. Second, unspecific binding to IgG sepharose beads upon cleavage of the Protein A tag significantly lead to the loss of purified protein needed for further purification.

The efficiency of the original TAP method to purify protein complexes might also be limited by the size of the tag (21kDa), which could interfere with interactions of TDG with other proteins. Nevertheless, the glycosylase activity of tagged TDG is not affected, indicating that the catalytic domain is largely intact, despite the fusion of the TAP tag. Future studies in these direction should make use of smaller tag combinations, such as Strep2-Flag, Flag-HA or Flag-His [24, 29-31]. Strikingly, however, immunoprecipitations of TDG from total cell lysates of *TdgA* overexpressing $Tdg^{-/-}$ ES cells also failed to identify specific TDG interacting proteins. This indicated that apart from technical constraints associated with TAP-tagging, the molecular environment of TDG might directly explain the difficulty to purify TDG-protein complexes from total cell lysates.

Nevertheless, fractionation of total cell lysates of TdgA overexpressing $Tdg^{-/-}$ ES cells by classical gel filtrations showed that TDG associates with different high molecular weight protein complexes. This suggested that although TDG interacts with proteins in macromolecular assemblies it could be enclosed in the latter, making it inaccessible for binding by TDG specific antibodies under native conditions. In this context, it should be emphasized that TDG interacting proteins identified so far have been identified in yeast two hybrid screens using the partner protein as bait, and under conditions where both interaction partners were overexpressed. However, the fact that a majority of the proteins previously reported to interact with TDG have chromatin associated functions [17] also suggested that interactions involving TDG might occur primarily on chromatin. Indeed, ES cell fractionations show that a significant fraction of nuclear TDG is tightly associated with chromatin. However, immunoprecipitation of TDG after high salt elution from chromatin of $Tdg^{+/+}$ MEFs and $Tdg^{-/-}$ knockout MEFs complemented with TAP-tagged TDG did not reveal any specific copurifying proteins or protein complexes, suggesting that TDG-

containing protein complexes are susceptible to high salt concentrations. This would also be consistent with our gel filtration results, showing high molecular weight complexes containing TDG are dissolved under high salt conditions.

Finally, we found that TDG is tightly associated with chromatin in mouse P19 cells, together with Dnmt3a/3b and p68, from which it could only be partially extracted by extensive MNase digestion of DNA, again together with Ddnmt3a/3b. Based on these results, we expected TDG to interact with other proteins in MNase digested fractions. However, immunoprecipitation experiments in these fractions would copurify large amounts of contaminating chromatin at low salt concentrations, making the identification of specific TDG interacting proteins difficult if nor impossible.

Together, our findings show that a significant fraction of TDG is included in distinct chromatin associated macromolecular assemblies in mammalian cells, some of which might contain the *de novo* DNA methyltransferases Dnmt3a/3b and the RNA helicase p68. A tight association with chromatin would be consistent with an involvement of TDG in a putative DNA demethylation complex proposed to act during transcriptional activation of the hormone responsive pS2 promoter [17]. Together with published interactions with transcription factors, the enzymatic properties of TDG and the developmental phenotype of $Tdg^{-/-}$ mice, TDG-containing protein complexes associated with chromatin could be predicted to act mainly during DNA (de)methylation processes at transcriptionally active gene promoters during embryonic development and/or cellular differentiation. Future studies will have to focus on size exclusion chromatography of MNase digested samples followed by MS analysis in order to identify and characterize chromatin bound TDG containing high molecular weight protein complexes. In addition, experiments to address the role of chromatin associated TDG-protein complexes will include Chromatin Immunoprecipitation (ChIP) experiments with TDG and known interacting proteins to address their colocalization at particular genomic regions, such as gene promoters. In addition, characterization and purification of TDG-complexes will include size exclusion experiments and TDG Immunoprecipitation from chromatin bound protein fractions followed by MS analysis.

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation.

REFERENCES

1. Neddermann, P. and J. Jiricny, *The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells.* J Biol Chem, 1993. **268**(28): p. 21218-24.
2. Neddermann, P. and J. Jiricny, *Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1642-6.
3. Borys-Brzywczy, E., et al., *Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts.* Acta Biochim Pol, 2005. **52**(1): p. 149-65.
4. Hardeland, U., et al., *The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs.* Nucleic Acids Res, 2003. **31**(9): p. 2261-71.
5. Yoon, J.H., et al., *Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair.* Nucleic Acids Res, 2003. **31**(18): p. 5399-404.
6. Kunz, C., et al., *Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil.* PLoS Biol, 2009. **7**(4): p. e91.
7. Cortazar, D., et al., *The enigmatic thymine DNA glycosylase.* DNA Repair (Amst), 2007. **6**(4): p. 489-504.
8. Chevray, P.M. and D. Nathans, *Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 5789-93.
9. Um, S., et al., *Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase.* J Biol Chem, 1998. **273**(33): p. 20728-36.
10. Missero, C., et al., *The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription.* J Biol Chem, 2001. **276**(36): p. 33569-75.
11. 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, 2003. **278**(40): p. 38586-92.
12. Shimizu, Y., et al., *Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase.* Embo J, 2003. **22**(1): p. 164-73.
13. Lucey, M.J., et al., *T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif.* Nucleic Acids Res, 2005. **33**(19): p. 6393-404.
14. Li, Y.Q., et al., *Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair.* Nucleic Acids Res, 2007. **35**(2): p. 390-400.

15. Boland, M.J. and J.K. Christman, *Characterization of Dnmt3b:Thymine-DNA Glycosylase Interaction and Stimulation of Thymine Glycosylase-Mediated Repair by DNA Methyltransferase(s) and RNA.* J Mol Biol, 2008.
16. Gallais, R., et al., *Deoxyribonucleic acid methyl transferases 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation.* Mol Endocrinol, 2007. **21**(9): p. 2085-98.
17. Metivier, R., et al., *Cyclical DNA methylation of a transcriptionally active promoter.* Nature, 2008. **452**(7183): p. 45-50.
18. Endoh, H., et al., *Purification and identification of p68 RNA helicase acting as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor alpha.* Mol Cell Biol, 1999. **19**(8): p. 5363-72.
19. Jost, J.P., et al., *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, 1999. **27**(16): p. 3245-52.
20. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.
21. Hess, D., et al., *Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective Omicron-glycosylation of thrombospondin type 1 repeats.* J Biol Chem, 2008. **283**(12): p. 7354-60.
22. Bauer, A. and B. Kuster, *Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes.* Eur J Biochem, 2003. **270**(4): p. 570-8.
23. Gingras, A.C., et al., *A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity.* Mol Cell Proteomics, 2005. **4**(11): p. 1725-40.
24. Jin, J., et al., *A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1.* Mol Cell, 2006. **23**(5): p. 709-21.
25. Knuesel, M., et al., *Identification of novel protein-protein interactions using a versatile Mammalian tandem affinity purification expression system.* Mol Cell Proteomics, 2003. **2**(11): p. 1225-33.
26. Mizuguchi, H. and T. Hayakawa, *The tet-off system is more effective than the tet-on system for regulating transgene expression in a single adenovirus vector.* J Gene Med, 2002. **4**(3): p. 240-7.
27. Nimura, K., et al., *Dnmt3a2 targets endogenous Dnmt3L to ES cell chromatin and induces regional DNA methylation.* Genes Cells, 2006. **11**(10): p. 1225-37.
28. Guan, X., et al., *The human checkpoint sensor Rad9-Rad1-Hus1 interacts with and stimulates DNA repair enzyme TDG glycosylase.* Nucleic Acids Res, 2007. **35**(18): p. 6207-18.
29. Burckstummer, T., et al., *An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells.* Nat Methods, 2006. **3**(12): p. 1013-9.
30. Gloeckner, C.J., et al., *A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes.* Proteomics, 2007. **7**(23): p. 4228-34.

31. Yang, P., H.M. Sampson, and H.M. Krause, *A modified tandem affinity purification strategy identifies cofactors of the Drosophila nuclear receptor dHNF4.* Proteomics, 2006. **6**(3): p. 927-35.

FIGURE LEGENDS

Figure 1. **TDG Tandem Affinity Purification (TAP) of TDG. A)** TAP protocol. TAP-tagged proteins (X) are bound to IgG sepharose (1) and released by TEV-protease cleavage (2). TEV protease–cleaved proteins are then bound to calmodulin sepharose (3) and eluted for subsequent MS analysis. **B)** TAP-tag constructs used to purify TDG from mouse embryonic fibroblasts (MEFs). Shown are the N- and C-terminal TAP tag TDG variants. **C)** Immunoblot with anti-mouse TDG antibody showing endogenous TDG levels in $Tdg^{+/+}$ MEFs and corresponding expression levels of N-and C-terminally tagged proteins in total cell extracts from stably transfected $Tdg^{-/-}$ MEFs. *Calm, Calmodulin; CBD, Calmodulin binding domain; TEV, Tobacco Etch Virus. Asterisks = Sumo-TDG.*

Figure 2. Purification of TAP-tagged TDG. A) TDG purification from MEFs. Recovery of TAP-TDG was monitored by immunoblotting (upper two panels) using an anti-mouse TDG antibody and Coomassie Blue Stain (lower panel). 50 µg $Tdg^{+/+}$ WCL, 50 µg TAP-TDG WCL or indicated amounts of purification fractions were analyzed. **B)** Base release activity of recombinant TDG and purified TAP-TDG on a G/T mispair containing 60mer DNA duplex. Shown are intact substrates (S) and cleaved products (P) resolved on denaturing polyacrylamide gels after 40 minutes incubation with the respective fractions. Shown is the product intensity in % relative to the positive control using recombinant protein set as 100 (lane 2). On the right, schematic representation of the base release assay. *WCL, whole cell lysate; TEV, TEV protease; FT, Flow-through = unbound; CALM, Calmodulin; CBP, Calmodulin binding peptide. Asterisks: Sumo-native TDG*

Figure 3. Expression of *TdgA* in mouse embryonic stem cells A) *TdgA* expression strategy used. Mouse *TdgA* is expressed in the absence of tetracycline (Tet) because the transactivator (tTA) factor binds to the promoter of the pP_{CMV*-1} -*TdgA*

expression vector and activates transcription. Addition of Tet to the culture medium causes a conformational change in tTA, which leads to its dissociation from the promoter and thus, inactivation of mouse *TdgA* transcription. **B)** TDG immunoblot of total cell lysates of $Tdg^{+/+}$ wildtype ES cells and $Tdg^{-/-}$ knockout ES cells overexpressing TDGa. **C)** Base release activity in total cell extracts from $Tdg^{-/-}$, $Tdg^{+/-}$ and EGFP or *TdgA* overexpressing ES cells on a G/T mispair containing 60-mer DNA duplex. *TdgA* overexpressing cells (Lane A) show a ~10 fold higher base release activity than $Tdg^{+/-}$ ES cell extracts. Shown is a representative denaturing polyacrylamide gel showing intact substrate (S) and cleaved products (P) at the top and bottom respectively. Ctr = no extract. Band quantification was done with ImageJ (version 1.42). *P_{CMV*-1}: Tet-responsive cytomegalovirus promoter, P_{CAG}: CMV early enhancer/chicken β-actin promoter, pA: poly A, Hyg: hygromycin resistance gene, , IRES: Internal Ribosome Entry Site, puro: puromycin resistance gene*.

Figure 4. TDG Immunoprecipitation (IP) from undifferentiated mouse ES cells. A) TDG immunoblot of input and TDG precipitates from total cell lysates of $Tdg^{-/-}$ knockout ES cells overexpressing TDGa (lanes 1 and 3) or EGFP as control (lanes 2 and 4). **B)** Representative 12% SDS-PAGE of TDG-immunoprecipitates stained with Coomassie Blue. 50 µg total cell lysate (Input) (lanes 1 and 2) and unbound fraction (FT) (lanes 3 and 4) were used for analysis. 50% of the total IP material, either with (lanes 5 and 6) or without anti-mouse TDG antibody (lanes 7 and 8) were loaded.

Figure 5. TDG associates with distinct high molecular weight protein complexes. A) Total extracts from mouse ES cells prepared at physiological salt concentrations were fractionated through a gel filtration column. Selected fractions were then resolved by SDS-PAGE (12%) and analyzed by Western blotting to identify TDG. Four apparent elution peaks (I, II, III, IV) were identified between 670kDa and 17kDa (dashed boxes). Molecular weights of gel filtration standards are indicated. The lower panel shows the corresponding SDS-PAGE gel stained with Coomassie Blue. Band quantification was done with ImageJ. **B)** Cell extracts prepared with 150, 200 and 300 mM NaCl were fractionated through a gel filtration column as in A. The elution peak in shown in A (fractions 9 to 12) was resolved into two distinct elution

peaks after increasing salt concentration (ii and iii). **C)** The presence of p68 was monitored in TDGa containing fractions 8, 10, 13 and 14 between 180 and 40kDa.

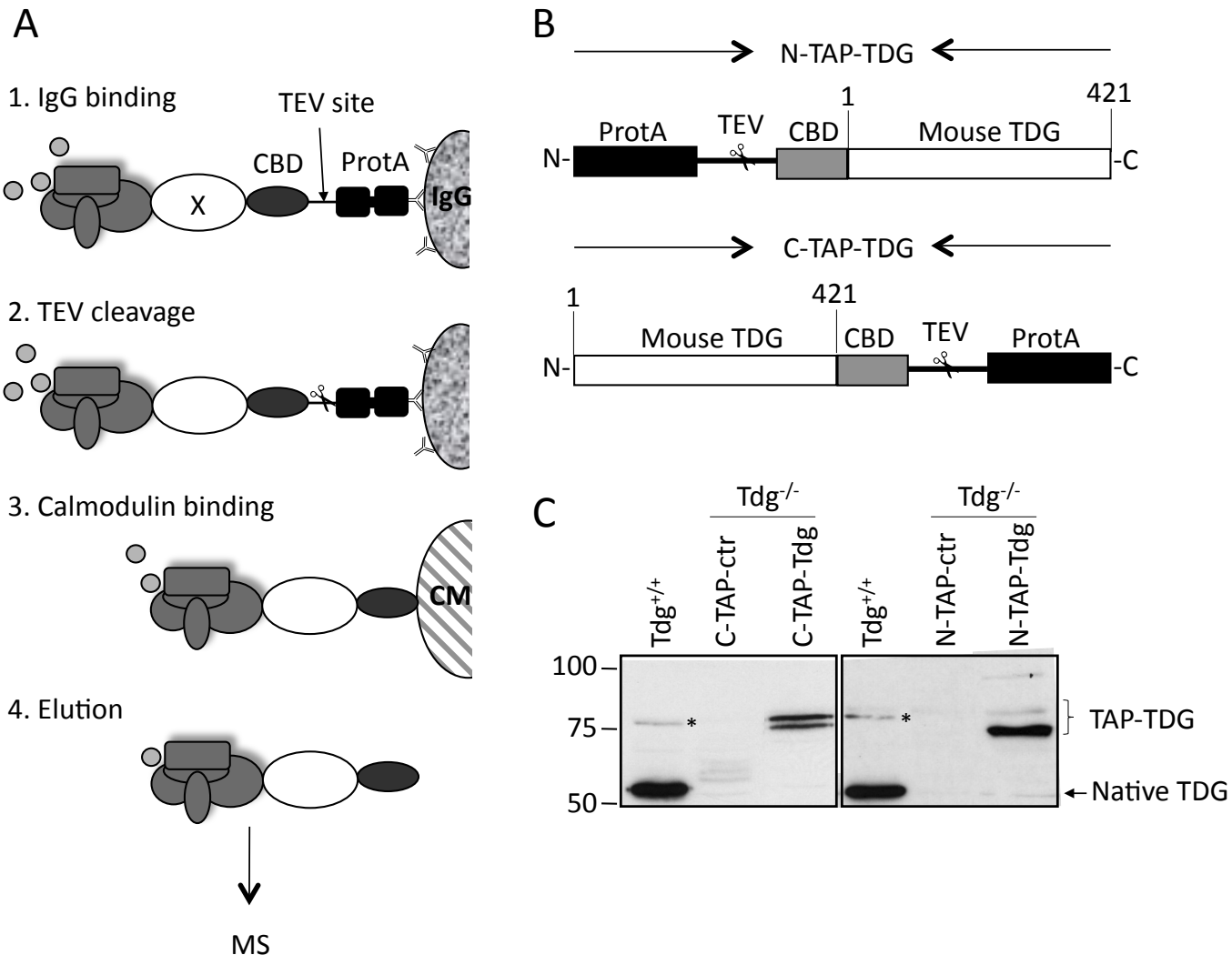
Figure 6. Immunoprecipitation of chromatin bound TDG. A) Solubilization of chromatin bound proteins in ES cell extracts. Selected fractions were analyzed by SDS-PAGE and Western blotting with a polyclonal anti-mouseTDG antibody. Salt concentrations indicated were used to extract chromatin bound proteins. PCNA was used as comparison. Chromatin bound TDG eluted at 0.2-0.3M NaCl (lanes 4 and 5), and at 0.5 M NaCl (lane 7). A fraction of TDG remains chromatin associated (Lane 9) even after elution with 1M NaCl (lane 8). *C: cytoplasmic, N: nuclear, P: pellet* **B)** Whole cell lysates of of $Tdg^{+/+}$ and $Tdg^{-/-}$ MEFs were prepared and chromatin bound fractions eluted in lysis buffer containing 250 mM or 500 mM NaCl prior immunoprecipitation with anti-mouseTDG antiserum. Immunoprecipitated material was analysed by SDS-PAGE (upper panels) and Western blotting (lower panels). 50 µg Input (I) and flow-through (FT) and 1/10 of the total immunoprecipitated (IP) fraction were loaded. **C)** Chromatin bound proteins were isolated from $Tdg^{-/-}$ MEFs expressing C-terminally TAP tagged TDGa (C-TAP-TDG) or the empty TAP vector (C-TAP). Eluted proteins were allowed to bind to IgG Sepharose beads and recovery was analyzed by SDS-PAGE (upper panels) and Western blotting (lower panels). 50 µg Input (I) and flow-through (FT) and 1/10 of the total IgG bound fraction were loaded.

Figure 7. Subcellular localization of TDG in P19 cells A) Presence of TDG in the soluble cytoplasmic (C), nuclear (N) and in the insoluble chromatin fraction of P19 cells. SDS extracts (SDS) were used as control for total protein amount. Insoluble, chromatin bound proteins were further solubilized by treatment with micrococcal nuclease (MNase) (lane 5). TDG is found in all cellular fractions. Insoluble nuclear TDG is eluted together with Dnmt3a/b from chromatin upon DNA digestion with MNase (lane 5). A substantial amount of TDG remained bound to the insoluble chromatin, as did Dnmt3b, lower amount of Dnmt3a and total p68 (lane 7). **B)** Representative agarose gel electrophoresis showing the efficiency of MNase digestion used in A.

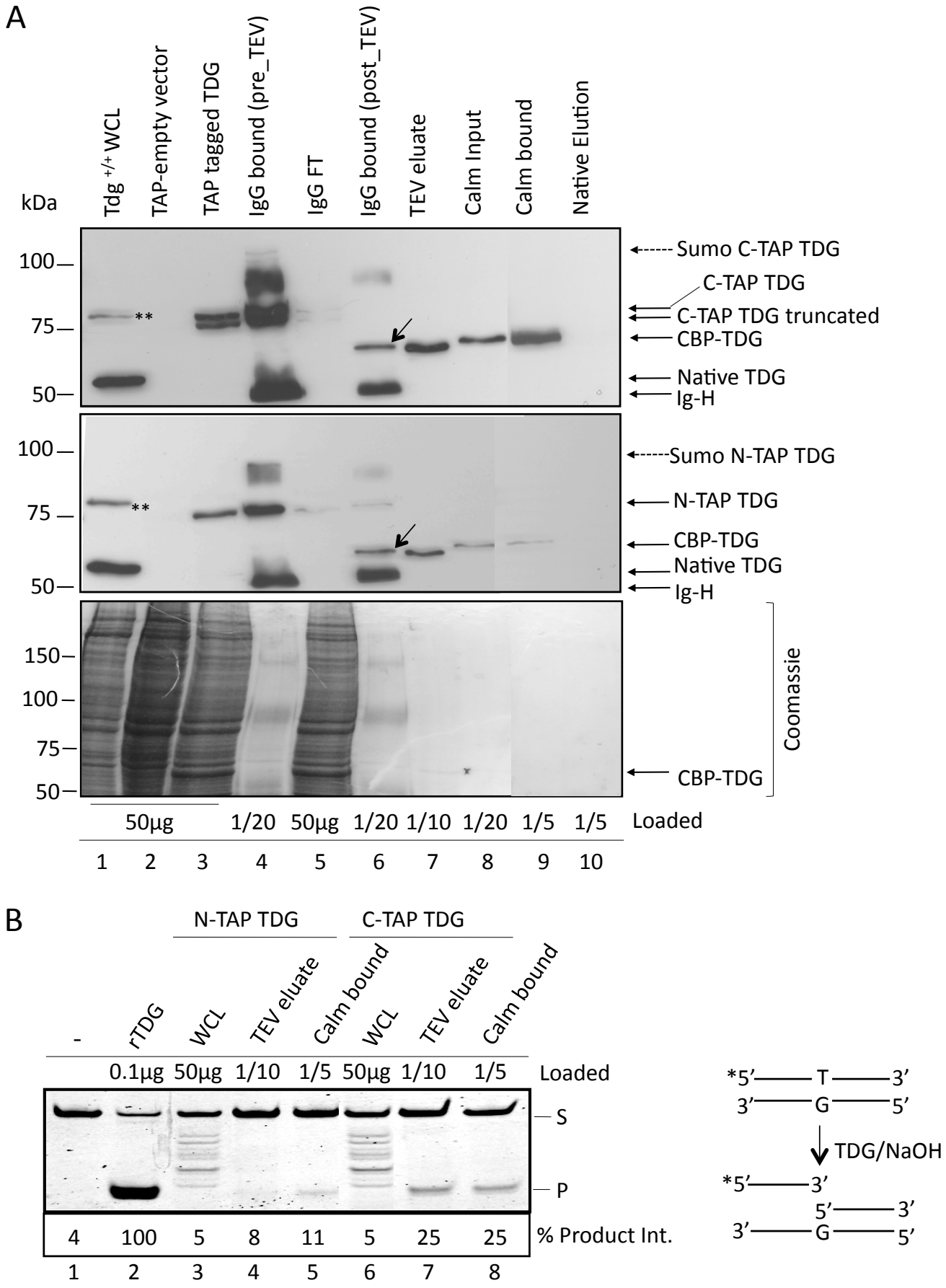
Figure S1. Quantitation of TDG protein amount in total cell extracts. 50 µg of total MEF lysate and different amounts of recombinant TDG were separated by SDS-PAGE and analysed by Western blotting with an anti-mouse TDG antibody. Band intensities were calculated with ImageJ. 50µg of total cell lysates (last lane) are estimated to contain 0.02% of TDG protein (10ng).

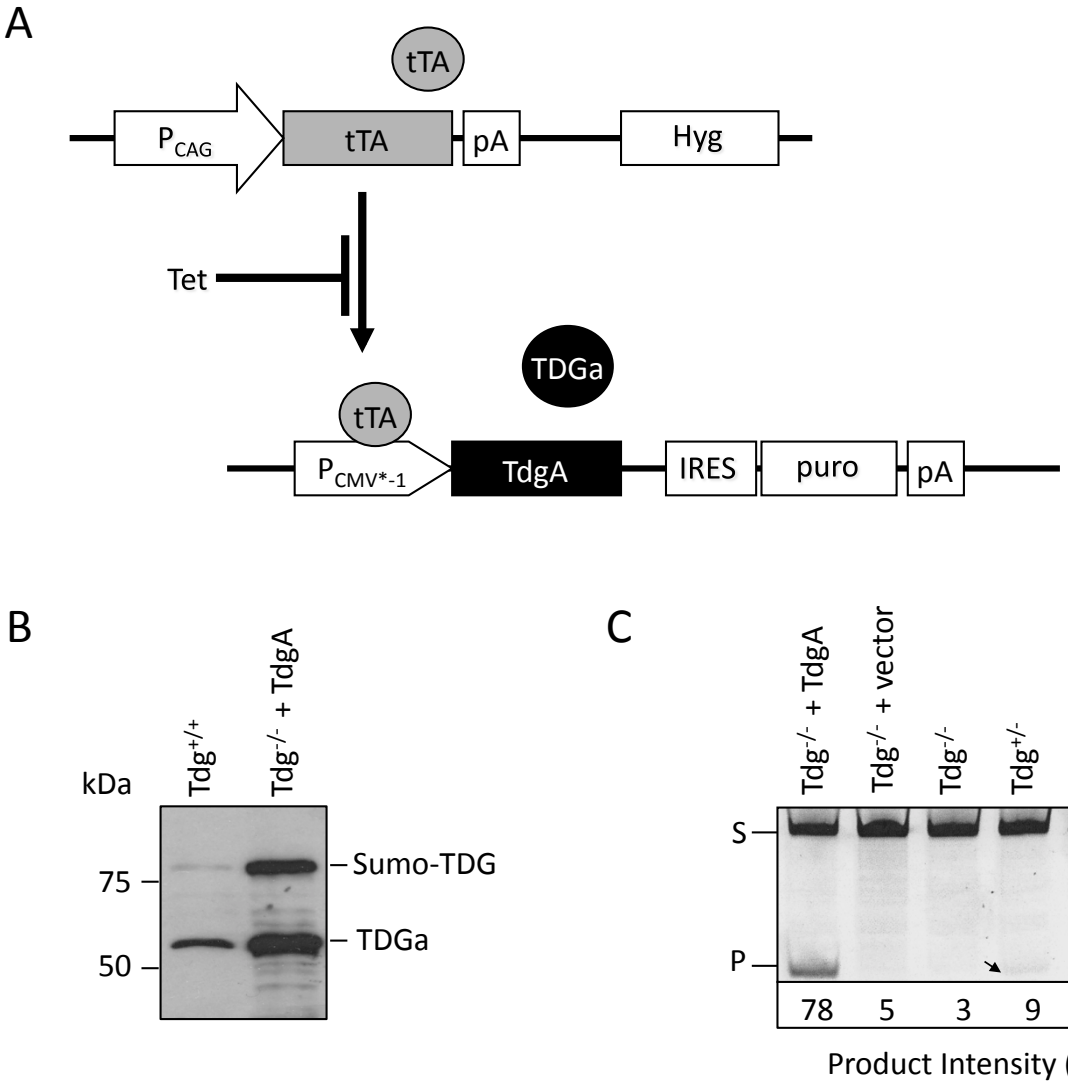
FIGURES

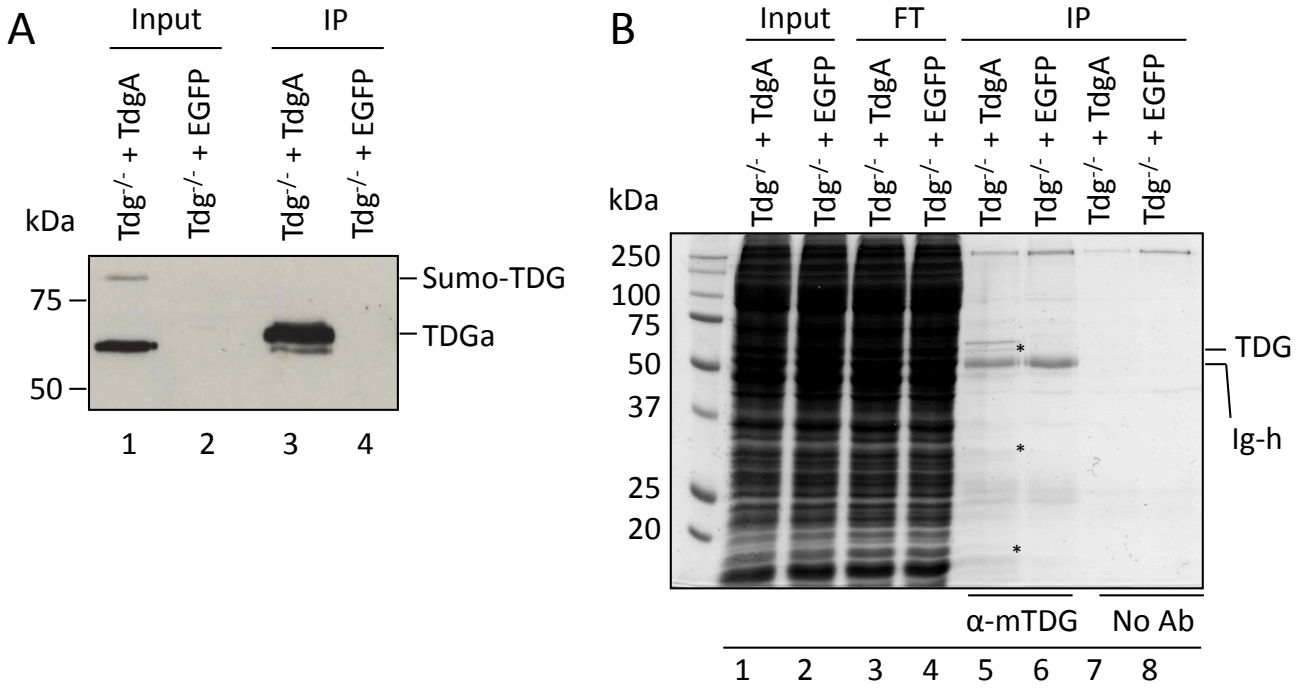
Cortázar et al., Figure 1



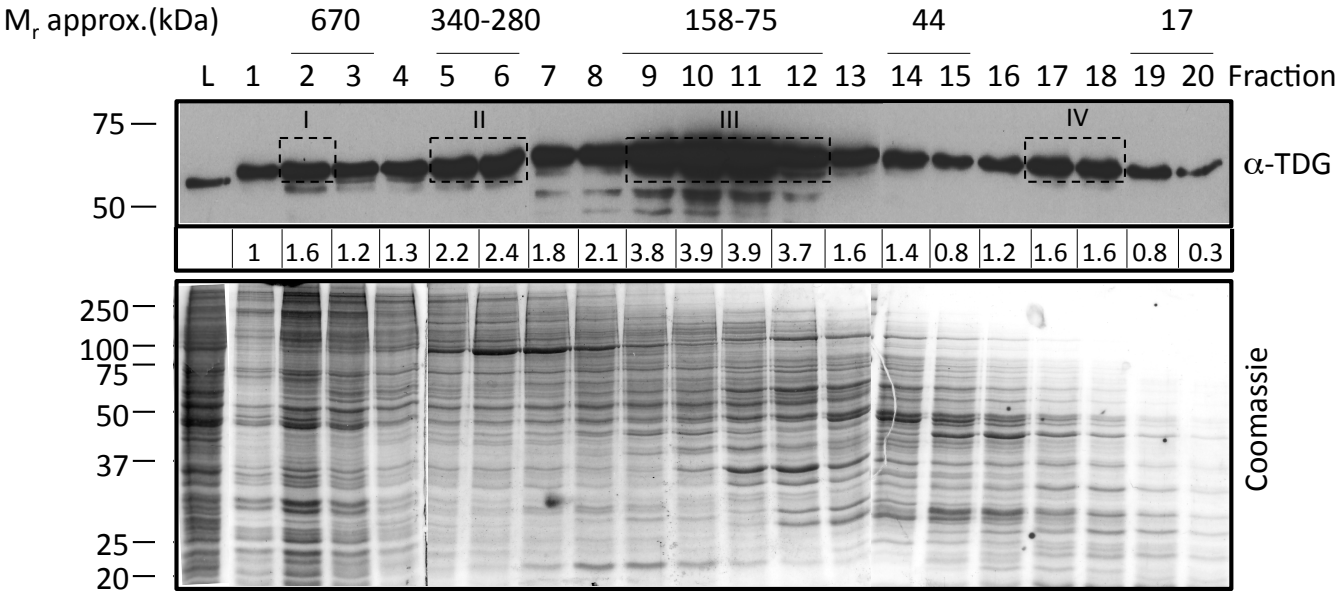
Cortázar et al., Figure 2



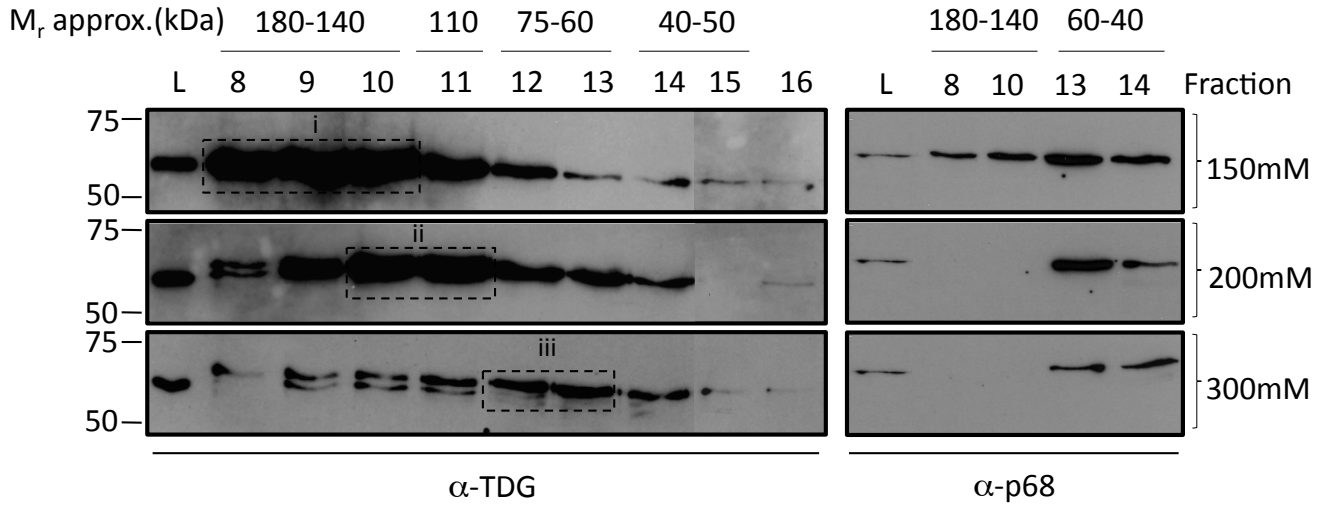

A
tTA
P CAG
tTA
pA
Hyg
Tet
TDGa
tTA
P CMV*-1
TdgA
IRES
puro
pA
B
kDa
Tdg+/+
Tdg-/- + TdgA
75
50
Sumo-TDG
TDGa
C
Tdg-/- + TdgA
Tdg-/- + vector
Tdg-/-
Tdg+/-
-
S
P
78
5
3
9
1
Product Intensity (%)


A
Input
IP
Tdg-/- + TdgA
Tdg-/- + EGFP
Tdg-/- + TdgA
Tdg-/- + EGFP
kDa
75
50
Sumo-TDG
TDGa
1
2
3
4
B
Input
FT
IP
Tdg-/- + TdgA
Tdg-/- + EGFP
Tdg-/- + TdgA
Tdg-/- + EGFP
Tdg-/- + TdgA
Tdg-/- + EGFP
Tdg-/- + TdgA
Tdg-/- + EGFP
kDa
250
100
75
50
37
25
20
TDG
Ig-h
α-mTDG
No Ab
1
2
3
4
5
6
7
8

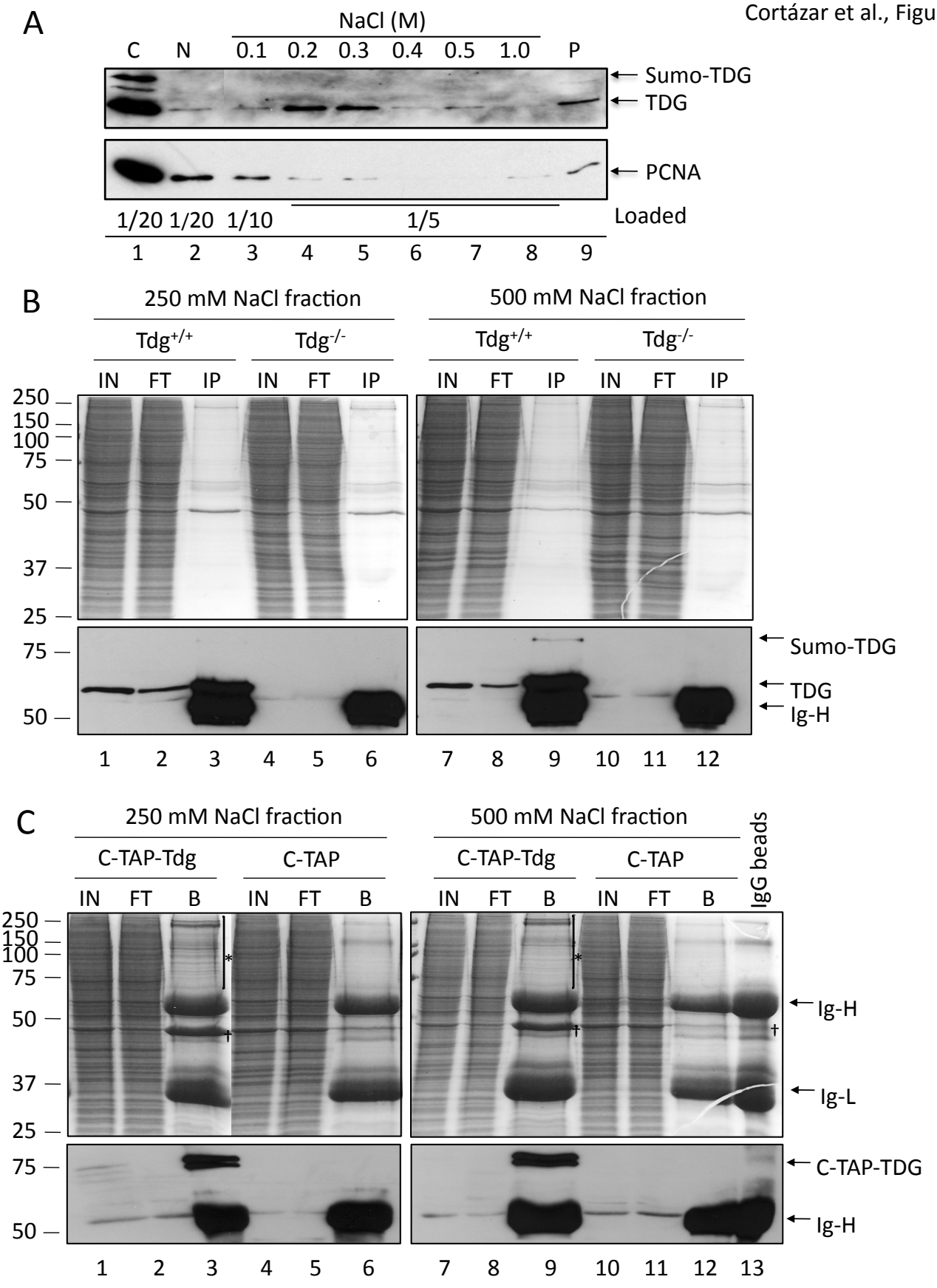
A


M$_r$ approx.(kDa)
670
340-280
158-75
44
17
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Fraction
75
50
I
II
III
IV
α-TDG
1 1.6 1.2 1.3 2.2 2.4 1.8 2.1 3.8 3.9 3.9 3.7 1.6 1.4 0.8 1.2 1.6 1.6 0.8 0.3
250
100
75
50
37
25
20
Coomassie

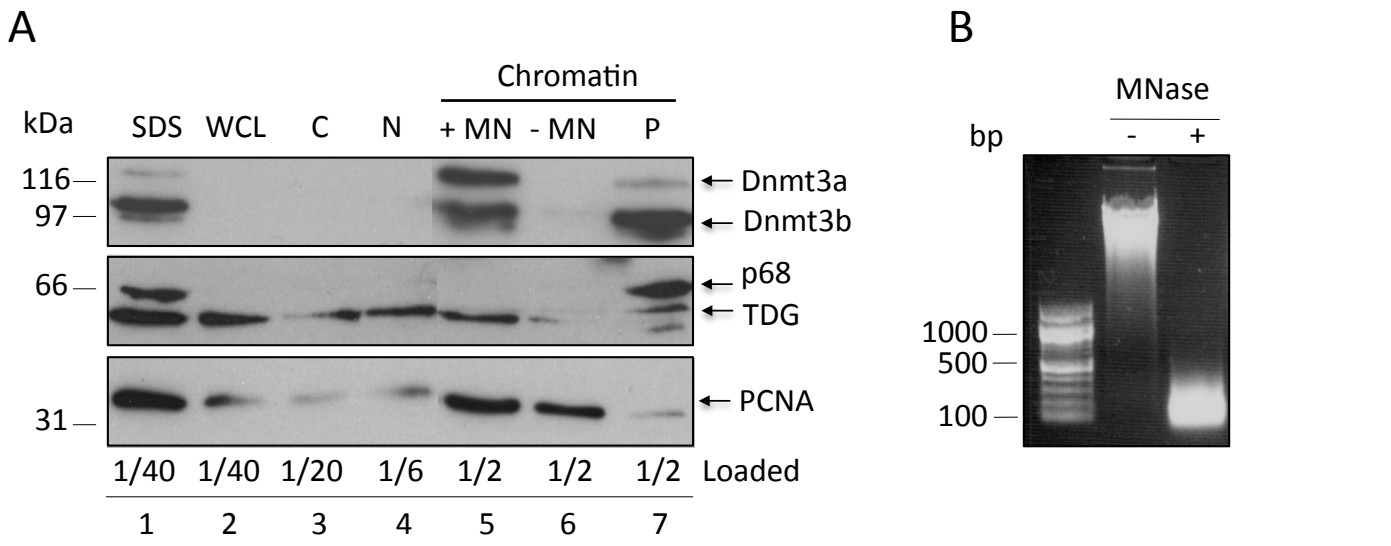
B


M$_r$ approx.(kDa)
180-140
110
75-60
40-50
180-140
60-40
L 8 9 10 11 12 13 14 15 16 L 8 10 13 14 Fraction
75
50
i
ii
iii
150mM
200mM
300mM
α-TDG
α-p68

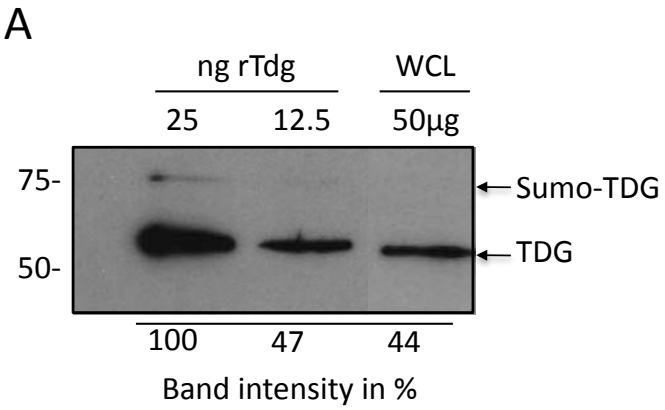
Cortázar et al., Figure 6



Cortázar et al., Figure 7



Cortázar et al., Figure S1



Thymine DNA Glycosylase Associates With CpG Islands to Maintain Their Epigenetic Function

Daniel Cortázar[1], Yusuke Saito[1], Christophe Kunz[1], Fabio Mohn[2], Fredy Siegrist[3], Dirk Schübeler[2], Primo Schär[1]*

[1]Department of Biomedicine
University of Basel
Mattenstrasse 28
4058 Basel
Switzerland

[2]Friedrich Miescher Institute for Biomedical Research
Maulbeerstrasse 66
4058 Basel
Switzerland

[3]Non-Clinical Drug Safety (NCS), Hoffmann-La Roche Ltd
Grenzacherstrasse 2,
4002 Basel,
Switzerland

* Corresponding author: primo.schaer@unibas.ch
Phone: +41 61 267 0767
Fax: +41 61 267 3566

Thymine DNA Glycosylase Associates With CpG Islands to Maintain Their Epigenetic Function

Daniel Cortázar[1], Yusuke Saito[1], Christophe Kunz[1], Fabio Mohn[2], Fredy Siegrist[3], Dirk Schübeler[2], Primo Schär[1]*

[1]Department of Biomedicine
University of Basel
Mattenstrasse 28
4058 Basel
Switzerland

[2]Friedrich Miescher Institute for Biomedical Research
Maulbeerstrasse 66
4058 Basel
Switzerland

[3]Non-Clinical Drug Safety (NCS), Hoffmann-La Roche Ltd
Grenzacherstrasse 2,
4002 Basel,
Switzerland

* Corresponding author: primo.schaer@unibas.ch

Phone: +41 61 267 0767

Fax: +41 61 267 3566

Word count:

Introduction, Results and Discussion: 5235

Materials and Methods: 2280

ABSTRACT

Thymine DNA glycosylase (TDG) was discovered as an enzyme capable of removing uracil and thymine from G/U and G/T mispairs, respectively. Owing to this ability, TDG was proposed to initiate restoration of C/G pairs at sites of cytosine and 5-methylcytosine deamination.

TDG has been shown to interact with transcription factors [1-4], histone modifying enzymes [5] and *de novo* DNA methyltransferases [6-8]. Recent studies have further demonstrated that TDG participates in promoter de- and remethylation of CpG sites during transcriptional activation of estrogen responsive genes [7]. Consistently, we have recently shown that TDG associates tightly with chromatin and cofractionates with Dnmt3a/b. Together, these studies suggest that TDG is targeted to CpG dinucleotides at gene promoters to either interrogate the integrity of such sequences in order to ensure faithful transcription and/or to regulate promoter activity through the modulation of DNA methylation.

We now show that TDG localizes to CpG islands promoters of transcribed genes and of genes 'poised' to be expressed during early and terminal cell differentiation. Among the identified targets, we find a significant number of pluripotency and key developmental genes, some of them belonging to the class of polycomb target genes. We further show that failure of TDG to localize to the promoters of these genes affects the establishment of correct gene expression patterns during cell differentiation. In addition, investigation of the DNA methylation pattern of gene promoters in NPs revealed that, although TDG has no effect on the genome-wide establishment of DNA methylation patterns, it affects DNA methylation at specific CpG sites of promoters regulating developmental genes in differentiated cells.

Our results suggest that TDG loss leads to local DNA methylation changes leading to the misregulation of pluripotency and developmental genes, adding a role for this base excision repair protein in epigenome maintenance.

INTRODUCTION

DNA cytosine methylation in mammals plays a critical role in many cellular processes, such as the epigenetic regulation of gene expression, the maintenance of genome stability, genetic imprinting and X-chromosome inactivation [9-12]. Cytosine methylation is an essential component of epigenetic regulation of embryonic development [13-16], and in humans, aberrant methylation has been linked to a number of diseases, including cancer [10, 17].

Cytosine methylation in mammals is restricted to CpG dinucleotide sequences. Species that undergo widespread genomic DNA methylation have lost CpG dinucleotides during evolution. One exception are the so-called CpG Islands, CpG-rich sequences which are commonly associated with gene promoter regions and other gene regulatory elements. CpG Islands show a high degree of mammalian conservation and have been shown to be predominantly unmethylated [18-20].

Genomic loss of CpGs is thought to be a consequence of the inherent mutability of methylated cytosines (C), due to a higher rate of spontaneous hydrolytic deaminations of 5-methylcytosine (5-meC) compared to cytosine. Deamination of 5-meC generates thymines (T) and thus leads to the formation of G/T mispairs in double stranded DNA. These give rise to C to T transition mutations DNA following DNA replication. Two enzymes, notably thymine DNA glycosylase (TDG) and methyl binding domain 4 (MBD4), have been suggested to initiate the restoration of methylated CpGs following 5-meC deamination. Whereas *Mbd4* knock out mice show a discrete increase in C to T mutations [21], extracts from *Tdg* knock-out mouse embryonic fibroblasts (MEF) and embryonic stem (ES) cells show undetectable levels of G/T processing activity, placing TDG as the prevalent DNA glycosylase repairing products of 5-meC deamination [22, 23]. Nonetheless, CpG dinucleotides remain hotspots for mutations in mammalian genomes, a phenomenon also linked to the development of cancers. For instance, up to 40% of all cancer-associated mutations in the *p53* tumor

suppressor gene are C to T transitions at methylated CpG dinucleotides [24]. This suggests that the action of G/T glycosylases might be confined to specific genomic sites where CpG sequences and or cytosine methylation have a genetic function, such as in gene regulatory regions. Indeed, TDG was shown to interact with transcriptional regulators [1-5, 25] and *de novo* DNA methylatransferases [6, 8]. Recent studies have further demonstrated that cyclic promoter de- and remethylation of estrogen-responsive genes involves the coordinated recruitment of Dnmt3a/b, p68, TDG and other BER proteins to promoter CpGs during active transcription [7].

A tight regulation of CpG methylation at gene regulatory regions, involving mutation avoidance and/or protection from aberrant methylation, is critical for accurate gene expression, proper cellular differentiation and embryonic development.

In addition to its function in DNA repair, TDG seems to be implicated in the regulation of developmental gene regulation through the interaction with transcription factors, such retinoic acid receptors RAR and RXR, histone modifying enzymes and *de novo* DNA methyltransferases (Dnmt3a and Dnmt3b) [3, 5, 6, 8, 25] all of which are essential for cell differentiation. Consistently TDG, unlike other DNA glycosylases [21, 26-29], has a non-redundant and essential function during embryonic development; homozygous *Tdg* null-embryos lose viability at midgestation (Primo Schär et al, Tetsuya Ono et al., manuscripts in preparation).

Recent subcellular fractionation studies in P19 cells have shown that a fraction of nuclear TDG is tightly associated with chromatin in a DNA damage independent manner, together with the TDG interacting proteins Dnmt3a/b and the RNA helices p68 (Cortázar et al., Manuscript in preparation). These data suggested that TDG might be constantly targeted to chromatin at regions undergoing dynamic changes in DNA methylation. Whether and how this is related to TDGs function in DNA repair and/or gene regulation remains to be clarified.

Dynamic changes of gene expression, along with epigenetic changes such as DNA methylation and histone modifications have been previously studied during *in vitro* differentiation of ES cells into pure populations of neuronal progenitor cells (NPs) [20, 30]. We decided to make use of this well-established differentiation system in order to address the function of TDG in developmental processes. Under the hypothesis that TDG might have a role at gene regulatory regions, we initially examined the association of TDG to DNA and investigated the consequences of TDG loss during *in vitro* differentiation of ES cell to NPs in respect to changes in gene expression and genome-wide DNA methylation patterns. We initially performed TDG chromatin immunoprecipitation (ChIP) coupled to high-throughput DNA sequencing to assess genome-wide localization of TDG in both ES cells and NPs. We show that TDG binds to CpG island promoters with high preference in both ES cells and NPs. Most of these promoters belong to actively transcribed genes or genes poised to be expressed during early and terminal cell differentiation. Among the genes regulated by TDG bound promoters we found a significant number of developmental genes. To elucidate a possible function of TDG in the regulation and/or maintenance of CpG methylation at the respective promoters we compared promoter methylation levels in NPs derived from $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells. We show that although some changes in DNA methylation are observed at some promoters, the steady state of global DNA methylation patterns as detectable by methylated DNA immunoprecipitation (MeDIP) are not significantly different between TDG proficient and deficient NPs. However, bisulfite sequencing revealed local changes at particular promoter CpGs of developmental genes in $Tdg^{-/-}$ NPs, indicating that TDG loss affects DNA methylation patterns at these sites.

In addition, we analyzed global gene expression changes during neuronal differentiation of $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells. First, we observed differential gene expression between *Tdg heterozygous* and *knockout* ES and NPs. Importantly, most of the changes in gene expression between $Tdg^{+/-}$ and $Tdg^{-/-}$ cells occurred at the NP stage, indicating that the

loss of TDG affects the establishment of gene expression patterns early during differentiation. Among the misregulated genes, we found a significant number of developmental genes and genes relevant for neuron development and regulated by TDG target promoters.

Our results show that TDG plays, initially described as a DNA repair protein involved in genome maintenance, plays an important role in maintaining the integrity of the genome.

MATERIALS AND METHODS

Antibodies

For Western blot analysis, we used the following antibodies: The polyclonal rabbit anti-mTDG antiserum was newly generated by immunization with recombinant full-length mouse TDGa (Primm Labs). Polyclonal anti-Dnmt3a (ab2850, Abcam, UK), Dnmt3b (ab2851 Abcam, UK), anti-p68 (IgG1 PAb204, Upstate, USA) For TDG ChIP and MeDIP, we used affinity-purified anti-mTDG (kindly provided by Gilles Salbert, University of Rennes, France), affinity purified anti-mTDG (made in our laboratory) and anti-5meC (Eurogentec 29, clone 33D3, USA), respectively. For HeK4me2 ChIP, we used polyclonal anti-H3K4me2 (# 07-030; Millipore, USA)

Cell culture

Tdg$^{+/-}$ and *Tdg*$^{-/-}$ ES cells used for neuronal differentiation experiments were cultured without feeder cells in ES cell medium (DMEM containing 15 % heat inactivated FCS, LIF (1,000 U/ml), nonessential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine and 90 μM β-mercaptoethanol) for two passages before differentiation was started. Differentiation into the neuronal lineage based on the protocol from Bibel *et al.* [30]. For embryonic body formation, 4 x 10^6 ES cells were plated onto nonadherent Greiner bacterial dishes in differentiation medium (ES medium without LIF and only 10 % FCS) and incubated for 4 days

at 37 °C in a humidified atmosphere containing 5 % CO_2. Medium was changed after two days. At day four, 5µM *all-trans* retinoic acid (RA) was added and cells were further incubated for another four days with a medium exchange at day 6. Embryonic bodies were dissociated at day eight and single cells were plated on poly-L-lysine (PLL) and laminin-coated dishes. For this purpose, cell culture dishes were coated with a solution of 10 µg/ml PLL in 1x PBS and placed overnight in the incubator. After washing the plates three times with PBS, laminin (0.5 µg/cm^2) was added directly to the PBS solution and the plates were returned to the incubators for at least 2 h. Embryonic bodies were washed twice with PBS and trypsinized by incubation for 3 minutes in a water bath at 37 °C in a 0.05 % trypsin solution in 0.05 % EDTA/PBS (freshly prepared with trypsin powder, TPCK-treated, Sigma-Aldrich, USA). Embryonic bodies were then gently but thoroughly resuspended in 10 ml differentiation medium, then centrifuged for 5 minutes at 1'000 rpm in a SORVAL®RT6000D centrifuge at room temperature (RT). The pellet was resuspended in N2 medium (DMEM, F12 nutrient mixture (1:1), N2 supplement) and the cell suspension filtered through a 40 µm nylon cell strainer (BD, USA). After removal of the laminin solution from the plates, the cell suspension was immediately added at a density of 5×10^6 cells/60 mm dish or 1.5×10^7 cells/100 mm dish. The N2 medium was changed after 2 hours and again after one day. After two days, the N2 medium was replaced by complete B27 Medium (Neurobasal Medium, B27 supplement, 2 mM L-glutamine).

RA induced differentiation of ES cells for 2 and 4 days was done as follows. ES cell were cultured without feeder cells in ES cell medium for two passages before differentiation was started. ES cells were then washed twice in ES cell medium without LIF and further cultured in ES cell medium without LIF and complemented with 1µM RA. RA containing medium was changed every 24 hours until cell harvesting at the desired timepoints. For late timepoints, ES cell were spit when reaching 80% confluence and plated in fresh RA containing medium.

Western Blot Analysis

Soluble proteins were separated in 10 % SDS-polyacrylamide gels and transferred to a nitrocellulose membrane (Millipore, USA). Membranes were washed once with TBS-T (100 mM Tris/HCl [pH 8.0], 150 mM NaCl, 0.2% Tween20) and incubated with blocking buffer (TBS-T, 5 % dry milk) for 1 h at RT. Membranes were incubated with the primary antibody for 1 h at 33 °C (anti-mTDG; dilution 1:10'000) or RT (anti–PCNA, dilution 1: 5'000; anti-p68, dilution 1: 5'000; dilution anti-Dnmt3a/b, 1:2'000) in blocking buffer. After hybridization, membranes were washed three times for 1min followed by three washes for 15 min at RT. Secondary horseradish peroxidase–conjugated antibodies were diluted 1:5'000 in blocking buffer and hybridized for 1 h at RT. After three washing steps of 10 min at RT, detection of the signals was carried out using the Immobilon Western Chemiluminescent HRP Substrate (Millipore, USA).

Cellular fractionations

$1x10^7$ – $2x10^7$ ES cells were washed 3x with ice cold PBS [pH 7.4] and resuspended in 200 µl buffer A (10 mM HEPES [pH 7.9], 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 0.1% Triton X-100, 1 x protease inhibitor cocktail (Roche)). Cells were then incubated on ice for 6 - 8 minutes and centrifuged at 1'300 g for 5 min at 4°C. The supernatant corresponding to the cytoplasmic fraction was clarified by centrifugation at 20'000 g for 5 minutes at 4 °C. The nuclear pellet was washed twice with buffer A and lysed for 30 minutes at 4°C in 200 µl buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, 1 x protease inhibitor cocktail (Roche, Switzerland). After centrifugation at 1'700 g for 5 min at 4°C the pellet fraction containing the insoluble chromatin fraction was washed two times with 200 µl buffer B and three times with 200 µl MNase buffer (10 mM Tris [pH 7.5], 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 3 mM $CaCl_2$, PMSF). Chromatin associated proteins were solubilized by treatment with 60 U micrococcal nuclease (Fermentas, USA) in 25 µl MNase buffer for 5

min at 37°C. The reaction was then stopped by addition of 5 mM EDTA/2.5 mM EGTA for 10 min at 4°C. Solubilized proteins were recovered by centrifugation at 20'000 g for 5 min at 4°C and analyzed by SDS-PAGE and Western blot analysis.

Chromatin Immunoprecipitation (ChIP) Assay

For ChIP, mouse ES cells and NPs were chemically crosslinked by the addition of fresh PBS [pH 7.4] containing 1% formaldehyde for 10 min at RT. Glycine was added to a final concentration of 125 mM to stop the crosslinking reaction. Cells were then rinsed twice with cold PBS and were collected using a silicon cell scraper. After centrifugation at 500 g cell pellets were washed once with 200 μl cold ChIP lysis buffer I (10 mM HEPES [pH 6.5], 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) and twice with 200 μl cold ChIP lysis buffer II (10 mM HEPES [pH 6.5], 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl) with centrifugation steps of 5 minutes at 500 g and 4°C in between the washing steps. Cell lysis was done for 10 minutes in 500 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) followed by 15 minutes sonication (15 seconds ON, 30 seconds OFF, Power at HIGH) in a BIORUPTOR Sonication device (Diagenode, UK) to produce random DNA fragments ranging from 300 to 1'000 bp in size. After centrifugation at 14'000 g for 10 minutes, the supernatant was transferred to a fresh tube and DNA concentration was measured and adjusted to 0.75 mg/ml. 15 μg of sonicated chromatin were used as input control for further analysis. For immunoprecipitation reaction, 150 μg of sonicated chromatin were diluted 10 times in ChIP dilution buffer (50 mM Tris-HCl [pH 8.0], 1 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 1 mM PMSF) and were precleared for 1 hour at 4°C after the addition of 40 μl of a 50% magnetic Protein G bead (Invitrogen, USA) slurry preblocked with 1mg/ml BSA and 1mg/ml tRNA. Protein G beads were removed with an MPCS® magnet (DYNAL) and the supernatant subjected to overnight immunoprecipitation with 4 μg of affinity purified antibody against mTDG (Gilles Salbert) or our own affinity

purified antibody against mTDG at 4 °C. For H3K4me2 ChIP, 5µl of polyclonal anti-H3K4me2 antibody were used. Antibody bound protein-DNA complexes were then recovered by the addition of 40 µl of a 50% Protein G bead slurry preblocked with BSA/tRNA and further incubation at 4°C for 3 hours. Bead bound precipitates 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) and twice with 500 µl TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). Beads-bound precipitates from H3K4me2 ChIP were washed one with ChIP wash buffer I, once with ChIP wash buffer II and once with ChIP wash buffer III (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 250 mM LiCl, 1% sodium deoxycholate, 1% NP-40) and twice with TE. Bound complexes were eluted from the beads by two sequential incubations with 150 µl elution buffer (1% SDS, 0.1 M $NaHCO_3$) for 10 minutes at 65°C. Crosslink reversal of immunoprecipitated DNA and input DNA involved an incubation step at 37°C for 30 minutes in the presence of 0.5 mg/ml RNAseA, followed by the addition 0.5 mg/ml Proteinase K and further incubation at 65°C for 5 hours. DNA was finally purified by phenol/chloroform extraction and Na-acetate/ethanol precipitation. Inputs were resuspended in 10mM Tris-HCl [pH 8.0] at a concentration of 25ng/µl. ChIP elutions were resuspended in 20 µL 10mM Tris-HCl [pH 8.0]. 25 ng of input and 2 µl of immunoprecipitated DNA were used for subsequent qPCR analysis. For library preparation and Illumina Sequencing, the total eluted material from two ChIP experiments was used.

Sequencing, Library Preparation and Illumina/Solexa Sequencing

Library preparation, cluster generation and Illumina sequencing were performed by Dr. Christian Beisel and Ina Nissen at the Department of Biosystems Science and Engineering (D-BSSE) according to the manufacturer's instructions, ETH Zürich, Switzerland.

Methylated DNA Immunoprecipitation (MeDIP) assay

MeDIP assays were performed as described in [31]. Shortly, genomic DNA from 5 x 10^6 cells was prepared by resuspension of the cell pellets in lysis buffer (20 mM Tris-HCl pH 8.0, 4 mM EDTA, 20 mM NaCl, 1% SDS and 1mg/ml proteinase K) and incubation for 5 hours at 55°C. DNA was purified by phenol-chloroform extraction and ethanol/Na-acetate precipitation. DNA pellets were resuspended in TE containing 20 µg/ml RNase. Before carrying out MeDIP, genomic DNA was sonicated for 15 minutes (15 seconds ON, 30 seconds OFF, Power at HIGH) in a BIORUPTOR Sonication device (Diagenode, UK) to produce random fragments ranging from 300 to 1,000 bp. Fragmented DNA was then precipitated with 400 mM NaCl, 100% ethanol and glycogen-carrier. 4 µg of fragmented DNA were diluted in 450 µl TE and boiled for 10 min at 95 °C, cooled on ice for 10 min and immunoprecipitated for 2 hours at 4°C with 10 µl of monoclonal antibody against 5-methylcytidine in a final volume of 500µl after the addition of 10x IP buffer (100 mM sodium phosphate [pH 7.0], 1.4 M NaCl, 0.5% Triton X-100). The mixture was then incubated with 40 µl of Dynabeads coupled with M-280 sheep anti mouse IgG antibody (Dynal Biotech) for 2 h at 4°C and washed three times with 700 µl IP buffer. Beads were then 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 proteinase K) for 3 hours at 50°C. Immunoprecipitated methylated DNA was recovered by phenol-chloroform extraction followed by ethanol precipitation and resuspended in 60 µl TE and kept at -20°C. 20 ng sonicated genomic input DNA and 1 µl of the MeDIP reaction were used for subsequent qPCR. For hybridization to microarrays, ¾ of a MeDIP reaction and 50 ng of sonicated input DNA were amplified in parallel using the Whole Genome Amplification Kit (WGA) (Sigma-Aldrich, USA).

Microarray design and Promoter Microarray Hybridization and Analysis

385K CpG Island Plus Promoter Arrays (NimbleGen Systems Inc., UK) are single array designs that include all UCSC-annotated CpG islands and 1.8 kb tiled regions around currently annotated RefSeq gene promoters in the mouse MM8 genome build. Oligonuclotide probes are isothermal (50-75bp) and all regions are on average tiled at 100bp spacing between the probes. Sample labeling, hybridization, and microarray scanning were performed by NimbleGen Systems Inc. according to standard procedures. For analysis, raw fluorescent intensity values were used to calculate the log2 of the bound/input ratios for each individual oligonucleotide. The arrays were further normalized in order to account for potential labeling dye artifacts in the low signal range. The log2 ratios of all oligos from one promoter, which are located within a 900 bp window (+200 to -700) around the transcription start site (TSS) were averaged to determine a single log2 value per promoter. Subsequently, for comparison all arrays were normalized to a median log2 = 0 and scaled to have the same median absolute deviation using the LIMMA package in R (Smyth, 2004; Smyth and Speed, 2003).

Illumina gene expression analysis

Total RNA was extracted from *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ embryonic stem cells and neuronal progenitors using the TRIZOL® (Sigma-Aldrich, USA) according to the manufacturers instructions. For each cell line, RNA was prepared from three independent experiments. Starting with 500 ng / sample of total RNA, total RNA was transcribed and labeled with Illumina(r) TotalPrep RNA Amplification Kit (Ambion, USA). cRNA was quantified with QuantIT (Invitrogen, USA) against ribosomal RNA standard (Invitrogen, USA). 1500 ng cRNA were hybridized on two Illumina MouseWG-6 v2 slides according to the manufacturers instructions. Biological replicates were distributed between both slides. Fluorescent signals were imaged by laser scanning using the Illumina iScan system. The data was processed with Illumina genome studio software (gene expression module 1.0.6) with no background

correction or normalization. VST (variance stabilization) transformation 'lumiT()', quantile normalization 'lumiN(, method = "quantile")' and statistical analysis were performed with R (2.9.0) using the package lumi (1.10.1) and libraries contained therein, in particular limma (2.16.4).

RESULTS

TDG is tightly bound to chromatin in the absence of DNA damage

TDG is implicated in the regulation of gene expression through its interaction with sequence-specific transcription factors, nuclear receptors, chromatin remodeling complexes and *de novo* DNA methyltransferases Dnmt3a and Dnmt3b [6-8, 25]. The described interactions of TDG with factors involved in early development and cell differentiation led us to hypothesize that TDG might be targeted to and associate with chromatin in a site-specific and DNA damage independent manner during differentiation [32].

We first examined TDG and Dnmt3a and Dnmt3b protein expression levels during retinoic acid (RA) induced differentiation. We observed a gradual increase in TDG protein levels after treatment of ES cells with RA for 4 and 8 days (Fig. 1A). In addition, we observed an increase in Dnmt3a protein levels upon induction of differentiation, whereas the opposite was seen for Dnmt3b (Fig. 1B). Importantly, Dnmt3a and 3b protein levels were not affected in *Tdg*$^{-/-}$ ES cells, when compared to *Tdg*$^{+/-}$ cells. Based on these results and to address our initial question, we then studied the subcellular distribution of TDGa and Dnmt3b in undifferentiated ES cells overexpressing the *TdgA* isoform (Cortazar et al, Manuscript in preparation), which express high levels of TDGa protein (Fig. 1C) by biochemical fractionation. Similarly, the subcellular localization of TDGa and Dnmt3a was analyzed in RA treated ES cells. In undifferentiated ES cells, we found TDGa to be distributed in the

cytoplasmic and nuclear compartments of the cell (Fig. 1D, lane 1 and 2). Interestingly, a significant fraction of nuclear TDGa was associated with the insoluble chromatin fraction and it can be partially eluted from this fraction upon treatment of the insoluble chromatin with 100 mM and 500 mM NaCl (Fig. 1D, lanes 4 and 5). Interestingly, a significant amount of TDGa remained bound to chromatin (Fig. 1D, lane 6), together with Dnmt3b, which cannot be solubilized under these conditions. Although a little increase of chromatin-bound Dnmt3b was observed in *Tdg*$^{-/-}$ ES cells expressing *TdgA* (Fig. 1E, lanes 3 and 6) when compared to control cell expressing vector only, TDGa seems to be dispensable for the recruitment of Dnmt3b to chromatin. Next, we treated ES cells with RA to induce Dnmt3a protein expression and targeting to the DNA upon cellular differentiation. As in undifferentiated ES cells, TDGa was found in the cytoplasm, in soluble nuclear extracts as well as in the insoluble chromatin fraction, in which we found all Dnmt3a (Fig. 1F). Insoluble chromatin fractions from nuclear extracts were then solubilized by high salt treatment (1M NaCl) and DNA fragmentation following sonication (Fig. 1F, lanes 3 and 7). Although chromatin bound TDGa and Dnmt3a were released into the supernatant under these conditions, a significant fraction of both proteins remained associated with the residual insoluble chromatin (Fig. 1F, lanes 4 and 8). Interestingly, in the absence of TDG, Dnmt3a showed the tendency to dissociate more easily from the insoluble chromatin (Fig. 1F, lanes 3 and 7), indicating that TDG might partially influence Dnmt3a targeting to the DNA upon cell differentiation.

Together, our results show that TDG expression increases upon RA induced differentiation, concomitant with the upregulation and downregulation of Dnmt3a and Dnmt3b, respectively. Second, TDG deficient cells also respond to RA in in vitro differentiation experiments. Futhermore, Dnmt3a/3b protein expression levels are not affected in the absence of TDG. We show that TDG is tightly associated with chromatin, together with Dnmt3a/3b, however, it's not essential for recruitment of the latter to DNA. Thus, TDG

might be constitutively targeted to the chromatin through the interaction with transcription factors and/or Dnmt3b/a in undifferentiated and differentiating cells, respectively.

TDG is essential for proper neuronal differentiation

The strong chromatin association of TDG and previously published evidence for its involvement in regulation of gene transcription [1-4, 7, 33] prompted us to identify possible TDG targets at a genomic level. Judged from its interactions with proteins involved in developmental gene regulation and on the developmental phenotype of TDG deficiency in the mouse, TDG's function appears to be required early in development, a period of embryonic development which is characterized by highly dynamic changes of gene expression and concomitant alterations of DNA methylation and histone modifications. Such genetic and epigenetic changes have been extensively studied during *in vitro* differentiation of ES cells into pure populations of neuronal progenitor cells (NPs) and terminally differentiated neurons [20, 30]. We first established this differentiation system for $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells in our laboratory to address the role of TDG during early RA-induced neuronal differentiation (Fig. 2A). We performed three independent differentiations starting from both $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells. Analyzing the downregulation of expression of the pluripotency gene *Oct3/4* and upregulation of the NP-specific gene *Pax6* we monitored successful differentiation of $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells into NPs [20]. For all differentiations, we observed efficient downregulation of *Oct3/4* and a gain in *Pax6* expression at the NP stage (Fig. 2B), 4 hours after embryonic body (EB) dissociation and plating into neuronal medium (4h NPs). This indicated that both $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells are able to differentiate into early NPs (Fig. 2C, a and b). Strikingly, however, 48h later, upon culturing NPs in complete neuronal medium, TDG deficient cells went through massive cell death (Fig. 2C, c and d), showing that the absence TDG affects further differentiation of NPs into the neuronal lineage.

TDG binds to CpG-rich promoters

We hypothesized, that the absence of TDG might affect global DNA (de)methylation of gene promoters and/or expression of the respective genes during *in vitro* differentiation of NPs. We thus decided to address genome-wide TDG chromatin associations on a genomic scale by chromatin immunoprecipitation (ChIP) to examine a with global changes in DNA methylation after differentiation, together with changes in gene expression at both differentiation stages.

We first established a TDG-ChIP protocol in our laboratory. We made use of an affinity purified antibody against mouse TDG and validated the ChIP method with $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells and NPs by analyzing TDG occupancy at the Rplp0 (P0) proximal promoter, a previously described TDG binding region [34]. As shown in figure 3A, a reproducible and significant 4- and 3-fold enrichment of the P0 promoter region (Fig. 3A, black arrows) was detected in TDG precipitated DNA from $Tdg^{+/-}$ ES cells and NPs, respectively, when compared to $Tdg^{-/-}$ cells and normalized against an external control region (Fig. 3A, white arrows). This established both efficiency and specificity of our TDG ChIP.

Enriched DNA fragments after TDG ChIP were then subjected to DNA sequencing by using the Illumina high-throughput DNA sequencing technology. Sequencing of enriched DNA following specific TDG ChIP generated 6 million sequence reads, of which 3 million mapped to unique regions in the mouse genome. Non-uniquely mapping reads, derived from faulty reads, reads with SNPs or repetitive sequences that could not be mapped to the mouse genome or were mapped to multiple genomic sites, were ignored for further analysis. The unique reads were subsequently used for peak finding employing MACS [35], whereby TDG ChIP sequencing data from $Tdg^{-/-}$ ES cells was used as a background dataset to exclude non TDG-specific enrichments. This defined 6109 putative TDG binding sites in ES cells and 2748 in NPs (Fig 3B). Further analysis of underlying sequences of significant peaks revealed that

59% and 31% of TDG binding sites mapped to gene promoters, defined to localize at +/-2kb of the transcriptional start site (TSS +/-2kb), in ES cells and NPs, respectively (Fig. 3C, black). In addition, TDG binding sites were located in intergenic regions (23% in ES cells and 42% in NPs) and or within genes (18% in ES cells and 27% in NPs) (Fig. 3C, grey and white).

Sequences of the total number peaks corresponding to sequences significantly enriched over the *Tdg*$^{-/-}$ background, in addition to TDG binding sites that mapped to gene promoters (TSS +/-2kb) and to intergenic regions were then classified into strong and weak CpG Islands (CGIs) and CpG poor sequences, according to *Weber et al* [36]. This revealed that 79% and 53% of the total significant TDG bound sequences in ES cells and NPs, respectively, contained CGIs (strong and weak CGIs) (Fig. 3D, left columns). 96% in ES cells and 97% in NPs of sequences located in promoters (TSS) contained CGIs (Fig. 3D, middle columns). In contrast, 60% and 36% of TDG bound intergenic regions in ES cells and NPs, respectively, contained CGIs (Fig. 3D, right columns). Together, these data show that TDG is preferentially associated with CpG-rich promoters, both in ES cells and lineage-committed neuronal precursors. At intergenic and genic regions, however, we observe a decrease of TDG-bound weak CGIs and a concomitant increase of CpG poor sequences at the NP stage, indicating that the localization of TDG to sites is dynamic and may change with cell differentiation. The difference observed might also be explained by the fact that at regions containing CpG poor sequences TDG binding might be less specific making its repair function visible.

Because TDG associated with CGI promoters, a great majority of which are embedded in an open chromatin environment and are free of CpG methylation [19, 20], we correlated TDG binding with the presence of H3 dimethylated at lysine 4 (H3K4me2), an established marker for transcriptionally permissive chromatin [37]. Using highly significant H3K4me2 reads from previous H3K4me2 ChIP experiments [20], we observed the presence of H3K4me2 at 96% and 89% of TDG-bound promoter-associated CGIs in ES cells and NPs, respectively (Fig 4A

and B, left pie). As H3K4 methylation and CpG methylation in CG-rich sequences are mutually exclusive [20], this indicated that TDG binding is restricted to unmethylated promoter sequences in an open chromatin conformation.

We then examined the transcriptional activity of genes regulated by TDG-bound promoters by means of RNA polymerase II binding as an indication of ongoing transcription. Using highly significant unique RNA polymerase II (Pol II) reads from previous Pol II ChIP experiments [20] we observed that 72% and 74% of H3K4 dimethylated CpG island promoters bound by TDG are being actively transcribed in ES cells and NPs, respectively (Fig. 4A and B, middle pies). Thus, TDG preferentially associated with CpG island promoters of actively transcribed genes and to promoters in a "transcriptionally permissive" chromatin environment (H3K4me2) [38-40]. To analyse the effect of TDG absence on H3K4 dimethylation at TDG bound promoters, we performed H3K4me2 ChIP in $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells. As shown in figure 4D, H3K4 dimethylation is present in both $Tdg^{+/-}$ ES cells and cells devoid of TDG, proving that TDG is not involved in the regulation of the active this histone mark.

Interestingly, a significant number of the untranscribed but H3K4 dimethylated TDG bound promoters in ES (106) and NPs (14) were found to coincide with sites of H3K27 trimethylation (H3K27me3) (Fig. 4A and B, right pies, and Fig. 4C), a mark of silent chromatin, characteristic for polycomb target genes [20]. CpG-rich promoters with bivalent H3K4 and H3K27 methylation were previously shown to reversibly silence key developmental genes and lineage-specific transcription factors [20, 41]. An association of TDG with such promoters is particularly interesting, as polycomb regulation during development appears to predispose to aberrant promoter methylation later found in cancer [42, 43] (Table 1).

To examine if TDG binding is excluded once a promoter gets methylated, we analyzed genes that become silenced by DNA methylation upon differentiation. Using the USCS genome browser tracks, we compared unique sequencing reads derived from *Tdg*$^{+/-}$ ES cell ChIP with those from the *Tdg*$^{+/-}$ NPs ChIP. We observed TDG binding to promoters regulating important pluripotency genes including *Oct4* and *Nanog* (Fig S2, A-C and Table 2). During cellular differentiation, *de novo* DNA methylation is the prevalent mechanism for the repression of these genes [20, 44, 45]. TDG ChIP results show that the differentiation induced silencing of these genes correlates with the loss of TDG binding to the respective promoters (Fig S2, A-C and Table 2). For instance, TDG binding to the Oct4 promoters is lost in differentiated mouse embryonic fibroblasts (Fig. S2, D). By contrast, TDG binding to promoters of representative housekeeping genes, which are actively transcribed and protected from promoter DNA methylation throughout cellular differentiation, was detected in both ES cells and NPs (Table 2). These observations indicated that promoter CpG methylation is incompatible with TDG binding and further support a function of TDG in the protection and/or shielding of promoters from DNA methylation.

TDG loss does not affect genome-wide DNA methylation changes

The correlation of TDG binding with an unmethylated state of CpG island promoter might suggest a function of TDG in protecting promoter CpG islands from aberrant methylation. Therefore, we examined the global methylation status for a validated set of mouse promoters and see possible TDG-dependent methylation changes. We enriched methylated DNA from *in vitro* differentiated *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ NPs using methylated DNA immunoprecipitation (MeDIP) technology [31] and analyzed the precipitated sequences by hybridization to a promoter microarray representing 17'381 high confidence mouse promoters. Our analysis revealed that the methylation patterns of 98% of CGI promoters represented in the array was not globally different in *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ NPs (Fig. 5A, grey

dots). Nevertheless, significant differences in methylation were detectable. 237 and 151 promoters were found to be hypermethylated and hypomethylated, respectively, in *Tdg*$^{-/-}$ NPs. In addition, we observed that a fraction of TDG bound promoters in ES presented DNA methylation changes in NPs (12.5%) (Fig. 5B). Thus, these data suggest that TDG is not involved in the regulation of genome-wide DNA methylation patterns at CGI promoters, but do not exclude the possibility that the absence of TDG might account for local changes in DNA methylation at CGI promoters. For instance, sequencing of bisulfite converted genomic DNA from *Tdg*$^{-/-}$ NPs revealed changes in DNA methylation at specific CpG sites of the *Tgfb2* promoter region bound by TDG in ES cells (Fig. 5C). In addition similar analysis in *Tdg*$^{-/-}$ mouse embryonic fibroblasts showed more drastic changes in DNA methylation at the promoters of other developmental genes (e.g. *Sfrp2, HoxD13*), indicating that in the course of differentiation, TDG loss leads to the accumulation of aberrantly methylated CpG sites at CGI promoters

TDG loss affects the establishment of gene expression during NP differentiation

TDG has been implicated as a coregulator of transcription through the interaction with transcription factors, including nuclear receptors that are critical for cell differentiation [1-4, 25]. We therefore examined whether TDG affects gene regulation during differentiation of ES cells to NPs. Using Illumina BeadArrays we analyzed gene expression patterns in *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ ES cells and NPs from three independent differentiation experiments. In ES cells, 229 genes were differentially expressed by more than 1.2 fold ($p<0.05$) (Fig. 6A, green and red dots). 128 and 101 genes were down- and upregulated in *Tdg*$^{-/-}$ ES cells, respectively, when compared to *Tdg*$^{+/-}$ ES cells (Fig. 6C, left). In NPs, the number of differentially regulated genes under the same criteria increased to 1022 (Fig. 6B, green and red dots), of which 572 and 450 genes were down- and upregulated in *Tdg* deficient NPs, respectively, when compared to *Tdg* heterozygous cells (Fig. 6C, right). The

fact that the majority of differences in gene expression between *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ cells arise upon cellular differentiation indicates that the loss of TDG affects the establishment of gene expression patterns early during differentiation, when cells encounter dynamic changes in gene expression and epigenetic modifications.

Gene Ontology analysis for differentially regulated genes ($p<0.05$) revealed enrichment of transcription factor activity, gene expression and cell differentiation and proliferation, cell development and maintenance, cell death, in addition to cancer (Fig. 6D), highlighting the point that TDG loss affect the expression pattern of important developmental genes.

A comparison of TDG ChIP sequencing data in ES cells with the expression data revealed that number of developmental genes that are differentially expressed at the neuronal progenitor stage, such as the polycomb target genes *Gata6*, *Tgfb2*, *Sfrp2* and *Pdgfra* (Fig. 4C) are regulated by promoters to which TDG associates in ES cells. This indicates that TDG might be targeted to CGI promoters to directly regulate the expression of the respective genes in the process of cellular differentiation.

Among the differentially regulated genes between *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ NPs we also found a number of neuronal lineage specific genes, whose proper regulation is critical for terminal neuronal differentiation, such as *Nes*, *Neurog2*, and *Pax6*. *Pax6* is an important developmental transcription factor involved at different levels of neuron differentiation [46]. For instance, whereas expression of *Pax6* is essential for the development of neuronal progenitors, its downregulation is necessary to provoke further differentiation of neuronal progenitors to terminal neurons. Based on the fact that TDG binds to the *Pax6* promoter in ES cells (Fig. 4C), and since *Tdg*$^{-/-}$ ES cells started to die after 48 hours of plating of progenitors (Fig. 2C, right), we decided to investigate the levels of *Pax6* mRNA at 4 hours and 48 hours (48hNPs) after plating the progenitors. Whereas there is a 50-fold downregulation of *Pax6* gene expression during differentiation of *Tdg*$^{+/-}$ 4hNP to 48hNPs,

Tdg$^{-/-}$ 4h NPs only showed a 8-fold decrease in *Pax6* levels during the same period (Fig. 7A). To confirm that misregulation of *Pax6* is dependent on TDG, we performed differentiation experiments with *Tdg*$^{-/-}$ ES cells, overexpressing the *TdgA* isoform. *TdgA* expression in *Tdg*$^{-/-}$ 48hNPs clearly complemented misregulation of Pax6 (Fig. 7B). These results show that insufficient downregulation of the *Pax6* gene during further differentiation of NPs along the neuronal lineage, a process that in indispensable for proper development of terminal neurons, is a direct consequence of TDG loss. Although TDG binds to the *Pax6* promoter in ES cells and might act as a repressor of *Pax6* transcription directly, we cannot exclude the possibility that TDG might indirectly involved in the regulation of *Pax6* through the regulation of upstream molecules.

DISCUSSION

TDG is the prevalent DNA glycosylase involved in the restoration of C/G pairs at sites of C and 5-meC deamination. In addition to this DNA repair related function, TDG has been implicated in the regulation of gene expression in cooperation with transcription factors and histone modifying enzymes [1-4, 33], and, together with other base excision repair factors, in the de- and remethylation process of estrogen responsive gene promoters during gene activation in complex with *de novo* DNA methyltransferases [7, 34].

We have shown that, in contrast to other DNA glycosylases [32], TDG is strongly associated with chromatin. In the absence of DNA damage, these are rather surprising results that suggest that TDG is constantly associated with and/or is being constantly being targeted to chromatin. To examine the hypothesis that TDG dependent repair has a role in the regulation of gene promoters, we established TDG ChIP in undifferentiated ES cells and lineage committed neuronal precursors (NPs) and show that TDG associates with high preference to CpG Island promoters in an active chromatin environment, enriched for the active H3K4me2 mark, independently of the stage of cellular differentiation. Among these,

TDG localizes to CpG promoters of transcribed genes (e.g. enriched for RNA pol II), TDG associates with silent promoters, which are however “poised” for activation during cell differentiation (e.g. H3K4me2+; RNA pol II-). Since the presence of H3K4me and DNA methylation at CpG rich promoters are mutually exclusive, our data indicate that TDG associates preferentially with unmethylated CpG sequences.

In ES cells, we found that promoters of pluripotency genes, such as *Oct4* and *Nanog* are targeted by TDG. We observed that silencing of pluripotency genes, which is accompanied by DNA methylation of the respective promoters [47], correlated with concomitant dissociation of TDG from the latter. This suggests that DNA methylation might be the consequence of TDG dissociation, and implicates a function of TDG in protecting such promoters from premature DNA methylation, thereby contributing to the stability of the pluripotent state. Consistently, we have observed that *Tdg*$^{-/-}$ ES cells have the tendency to spontaneously differentiate and to lose pluripotency during normal ES cell culture conditions (Daniel Cortázar, Yusuke Saito, Data not shown).

In addition, TDG associates with promoters of important developmental genes (e.g. *Sfrp2, Tgfb2, Gata6, Pdgfra*). Developmental genes are often found in bivalent chromatin domains, methylated at H3K4 and trimethylated at H3K27, the silencing mark catalyzed by polycomb-group (PcG) proteins. We show indeed, that a significant number of polycomb targets are regulated by TDG bound promoters. Silencing of polycomb targets in stem cells has to be reversible and thus, does not involved DNA methylation. However, promoters of such genes are prone to become aberrantly methylated in human cancers [31]. Since polycomb repression targets CpG islands for DNA methylation in cancer and the process seems to involve the action of *de novo* DNA methyltransferases [48, 49], binding of TDG to these promoters is particularly interesting and suggests a role of TDG in the protection of CpG island promoters of developmental genes from being aberrantly methylated.

TDG has been previously shown to be involved in the regulation of gene expression

of a limited number of genes, both as a co-activator and co-repressor of transcription. We observed a significantly higher number of genes misregulated in NPs than in undifferentiated cells, indicating that TDG is involved in establishing gene expression patterns early in the course of cell differentiation. Pathway analysis of differentially regulated genes in NPs ($p<0.05$) revealed enrichment of mainly developmental genes, and ChIP data support the idea that the misregulation of developmental genes might be a direct consequence of TDG absence at the respective CpG island promoters.

Another developmental TDG target in ES cells is the *Pax6* gene, whose proper regulation is essential for proper neuron development. We show that *Pax6* is upregulated in both $Tdg^{+/-}$ and $Tdg^{-/-}$ 4hNPs. However, $Tdg^{-/-}$ NPs show only a partial downregulation of Pax6 during differentiation of 4hNP to terminal neurons, a fact that partially explains their lethal phenotype after 2 days of plating in neuronal medium. We further show that complementation of $Tdg^{-/-}$ ES cells with *TdgA* leads to efficient downregulation of Pax6 in 4hNPs when compared to $Tdg^{-/-}$ cells and leads to proper differentiation into terminal neurons. This result indicates that TDG might be directly involved in the regulation of *Pax6* expression.

We addressed a possible role of TDG in establishing genome-wide DNA methylation patterns. We examined the global DNA methylation status of gene promoters in lineage-committed 4hNPs by MeDIP analysis. Although we identified a significant number of promoters with differential DNA methylation patterns between $Tdg^{+/-}$ and $Tdg^{-/-}$ NPs, our MeDIP results rather indicate that TDG is not involved in establishing global DNA methylation changes in the course of cell differentiation. However, due to the limited of the MeDIP technique, local DNA methylation changes at specific promoter CpG sites cannot be excluded. For instance, a specific enrichment for hypermethylated CpG sites in an overall hypomethylated promoter environment is difficult, if not impossible to detect. We thus performed bisulfite sequencing of differentially regulated genes in NPs whose promoters

are bound by TDG in undifferentiated ES cells (e.g. *Tgfb2*) and show that loss of TDG affects the DNA methylation status of specific promoter CpG sites at the NP stage. In addition, recent bisulfite sequencing of additional TDG targets (e.g. *Sfrp2, HoxD13*) in differentiated mouse embryonic fibroblasts (MEFs) shows that at late differentiation timepoints, DNA methylation changes are even more drastic than in NPs (Christophe Kunz, personal communication), indicating that during development and/or cellular differentiation, there is a gradual accumulation of aberrantly methylated promoter CpGs in TDG deficient cells.

Our results suggest that promoter CpG sites are protected from DNA methylation by a mechanism involving the action of TDG. At actively transcribed genes, loss of TDG would lead to promoter methylation and consequent gene silencing. In the case of pluripotency genes, this would lead to the loss of the pluripotent state in normally undifferentiated cells (Fig. 8A). At polycomb target genes, the absence of TDGs protective function would leads to the stable silencing of otherwise reversibly silenced genes through DNA methylation of their promoters, a phenomenon that has been observed in human cancer (Fig. 8B).

Mechanistically, TDG mediated protection from DNA methylation could involve either the inhibition of binding of *de novo* Dnmts to DNA by shielding CpG sites (Fig. 8C, 1) or the modulation of their enzymatic activity (Fig. 8C, 2) [6]. Alternatively, as part of a putative DNA demethylation complex, TDG and Dnmt3a/3b might be involved in the demethylation of methylated CpG by a mechanism involving Dnmt3a/3b mediated 5-meC deamination and repair of thereby generated G/T mispairs by the base excision repair system [7] (Fig. 8C, 3).

One important question to be resolved is how TDG is targeted specifically to unmethylated CpG rich sequences. TDG targeting to promoter CpG islands of genes could be mediated by the interaction with transcription factors recognizing specific response elements in gene promoter regions. In addition, the fact that TDG associates almost exclusively with H3K4 dimethylated regions might also suggest that the histone modification itself or H3K4 methyltransferases could be involved in the recruitment process.

It is noteworthy, that TDG is associated to CpG Islands, which are generally conserved among mammals. These sequence stretches have been protected from the genomic loss of CpGs during evolution due to their preferentially unmethylated state, and the mechanism protecting these sites from methylation and/or mutation are not known. Our results make it plausible to think about TDG having a role in the protection of unmethylated CpG Islands against both aberrant DNA methylation and mutations.

It appears then, that TDG, as a base excision repair enzyme, is not only involved in maintaining the integrity of the genome, but also in the maintenance of the epigenome.

ACKNOWLEDGEMENTS

We thank Michael Stadler and Dimos Gaidatzis from the Friedrich Miescher Institute (FMI) (Basel, Switzerland) for assistance in Data generation and bioinformatic analysis of ChIP sequencing data. We acknowledge Christian Beisel and Ina Nissen from the Deep Sequencing Unit of The Department of Biosystems Science and Engineering (D-BSSE), ETH Zürich, Basel Switzerland) for library generation and Illumina sequencing.

This work was supported by the Swiss National Science Foundation

REFERENCES

1. 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, 2003. **278**(40): p. 38586-92.
2. Lucey, M.J., et al., *T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif.* Nucleic Acids Res, 2005. **33**(19): p. 6393-404.
3. Um, S., et al., *Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase.* J Biol Chem, 1998. **273**(33): p. 20728-36.

4. Zhou, J., et al., *TDG represses myocardin-induced smooth muscle cell differentiation by competing with SRF for myocardin binding.* J Biol Chem, 2008.
5. Tini, M., et al., *Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription.* Mol Cell, 2002. **9**(2): p. 265-77.
6. Li, Y.Q., et al., *Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair.* Nucleic Acids Res, 2007. **35**(2): p. 390-400.
7. Metivier, R., et al., *Cyclical DNA methylation of a transcriptionally active promoter.* Nature, 2008. **452**(7183): p. 45-50.
8. Boland, M.J. and J.K. Christman, *Characterization of Dnmt3b:Thymine-DNA Glycosylase Interaction and Stimulation of Thymine Glycosylase-Mediated Repair by DNA Methyltransferase(s) and RNA.* J Mol Biol, 2008.
9. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
10. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.
11. Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development.* Nature, 2007. **447**(7143): p. 425-32.
12. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development.* Science, 2001. **293**(5532): p. 1089-93.
13. Stancheva, I. and R.R. Meehan, *Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos.* Genes Dev, 2000. **14**(3): p. 313-27.
14. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
15. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
16. Lei, H., et al., *De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells.* Development, 1996. **122**(10): p. 3195-205.
17. Esteller, M., *Epigenetics in cancer.* N Engl J Med, 2008. **358**(11): p. 1148-59.
18. Fouse, S.D., et al., *Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation.* Cell Stem Cell, 2008. **2**(2): p. 160-9.
19. Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.
20. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-66.
21. Millar, C.B., et al., *Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice.* Science, 2002. **297**(5580): p. 403-5.

22. Neddermann, P. and J. Jiricny, *The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells.* J Biol Chem, 1993. **268**(28): p. 21218-24.
23. Kunz, C., et al., *Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil.* PLoS Biol, 2009. **7**(4): p. e91.
24. Greenblatt, M.S., et al., *Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis.* Cancer Res, 1994. **54**(18): p. 4855-78.
25. Cortazar, D., et al., *The enigmatic thymine DNA glycosylase.* DNA Repair (Amst), 2007. **6**(4): p. 489-504.
26. Nilsen, H., et al., *Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication.* Mol Cell, 2000. **5**(6): p. 1059-65.
27. Minowa, O., et al., *Mmh/Ogg1 gene inactivation results in accumulation of 8-hydroxyguanine in mice.* Proc Natl Acad Sci U S A, 2000. **97**(8): p. 4156-61.
28. Takao, M., et al., *Novel nuclear and mitochondrial glycosylases revealed by disruption of the mouse Nth1 gene encoding an endonuclease III homolog for repair of thymine glycols.* Embo J, 2002. **21**(13): p. 3486-93.
29. Engelward, B.P., et al., *Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13087-92.
30. Bibel, M., et al., *Differentiation of mouse embryonic stem cells into a defined neuronal lineage.* Nat Neurosci, 2004. **7**(9): p. 1003-9.
31. Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.* Nat Genet, 2005. **37**(8): p. 853-62.
32. Campalans, A., et al., *XRCC1 interactions with multiple DNA glycosylases: a model for its recruitment to base excision repair.* DNA Repair (Amst), 2005. **4**(7): p. 826-35.
33. Missero, C., et al., *The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription.* J Biol Chem, 2001. **276**(36): p. 33569-75.
34. Gallais, R., et al., *Deoxyribonucleic acid methyl transferases 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation.* Mol Endocrinol, 2007. **21**(9): p. 2085-98.
35. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
36. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nat Genet, 2007. **39**(4): p. 457-66.
37. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome.* Cell, 2007. **128**(4): p. 669-81.
38. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells.* Cell, 2007. **130**(1): p. 77-88.

39. Pan, G., et al., *Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells.* Cell Stem Cell, 2007. **1**(3): p. 299-312.
40. Zhao, X.D., et al., *Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells.* Cell Stem Cell, 2007. **1**(3): p. 286-98.
41. Boyer, L.A., et al., *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.
42. Ohm, J.E. and S.B. Baylin, *Stem cell chromatin patterns: an instructive mechanism for DNA hypermethylation?* Cell Cycle, 2007. **6**(9): p. 1040-3.
43. Ohm, J.E., et al., *A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing.* Nat Genet, 2007. **39**(2): p. 237-42.
44. Gidekel, S. and Y. Bergman, *A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element.* J Biol Chem, 2002. **277**(37): p. 34521-30.
45. Deb-Rinker, P., et al., *Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation.* J Biol Chem, 2005. **280**(8): p. 6257-60.
46. Bel-Vialar, S., F. Medevielle, and F. Pituello, *The on/off of Pax6 controls the tempo of neuronal differentiation in the developing spinal cord.* Dev Biol, 2007. **305**(2): p. 659-73.
47. Feldman, N., et al., *G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis.* Nat Cell Biol, 2006. **8**(2): p. 188-94.
48. Schlesinger, Y., et al., *Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer.* Nat Genet, 2007. **39**(2): p. 232-6.
49. Vire, E., et al., *The Polycomb group protein EZH2 directly controls DNA methylation.* Nature, 2006. **439**(7078): p. 871-4.

FIGURE LEGENDS

Figure 1. Analysis of TDG protein level and subcellular localization in ES cells.

A) TDG immunoblot of total cell lysates of Tdg+/- ES cells before in undifferentiated and retinoic acid (RA) treated cells for 4 and 8 days. TDG proteins levels increase upon RA induced ES cell differentiation. **B)** Dnmt3a and Dnmt3b immunoblots of SDS extracts during

RA induced differentiation. In both *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ cells Dnmt3a and Dnmt3b protein levels increase, respectively, after 48h of RA treatment. **C)** TDG immunoblot of total cell lysates of *Tdg*$^{+/+}$ wildtype ES cells and *Tdg*$^{-/-}$ knockout ES cells overexpressing TDGa. Complemented cells express about 100 times more TDG when compared to wildtype cells. **D)** TDG and Dnmt3b immunoblot showing the subcellular localization both proteins in undifferentiated *Tdg*$^{-/-}$ ES cells complemented with *TdgA*. TDG is found in the cytoplasm (C) soluble nuclear (N) fractions. Insoluble nuclear TDG can be solubilized upon treatment of insoluble chomatin with 0.1 and 0.5M NaCl. A fraction of TDGa stays associated with chromatins, together with Dnmt3b. **E)** Dnmt3b immunoblot of SDS extracts from *Tdg*$^{-/-}$ ES cells complemented with a *TdgA* expression construct or the vector alone. TDG might be involved but is not essential for Ddnmt3b binding to chromatin. **F)** RA differentiated *Tdg*$^{-/-}$ ES cells complemented with a *TdgA* expression construct or the vector alone. TDG is found in the cytoplasmic (C) and nuclear fraction. Chromatin associated proteins were further solubilized by high salt (1 M NaCl) and sonication (S). SDS extracts (SDS) were used as control. Treatment of the insoluble fraction allowed partial release of TDGa. However, a substantial amount remained bound to the insoluble chromatin. A similar behavior was detected for Dnmt3a.

Figure 2. Cellular Differentiation System and Experimental Setup

A) Overview of the *in vitro* differentiation of embyonic stem (ES) cells to neuronal progenitors (NPs). Experiments were done in biological triplicates and both ES cells and NPs at 4h after plating were harvested for the analysis of TDG localization using ChIP, global DNA methylation by MeDIP, and gene expression analysis. **B)** Differentiation of *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ ES cells to 4hNPs was monitored by analyzing the expression of the pluripotency gene *Oct4* and the NP-specific marker *Pax6.* Cell efficiently downregulated Oct4 during *in vitro* differentiation and started to express Pax6 at the NP stage. **C)** Cell morphology of *Tdg*$^{+/-}$ and

Tdg$^{-/-}$ NPs after 4 (a and b) and after 48 hours (c and d) of culture in neuronal medium. Whereas *Tdg*$^{+/-}$ further develop along the neuronal lineage, *Tdg*$^{-/-}$ start to die after 48h of plating dissociated EBs in neuronal medium. RA, retinoic acid; EB, embryonic bodies

Figure 3. TDG localizes to promoter CpG Islands

A) Establishment of the TDG chromatin immunoprecipitation method for ES cells and NPs. Relative enrichment at the P0 promoter region was normalized againts a protein coding region of the gapdh gene. Black and white arrows indicate the position of primers used for PCR analysis of ChIPed sequences.

TDG specific enrichment of P0 promoter sequence was observed at both differentiation stages. **B)** Sequencing of ChIP enriched fragments yielded 6 million reads, from which 3 million were mapped to unique sites in the mouse genome. Unique reads were subsequently used for peak finding employing MACS [35], defining 6109 and 2748 putative TDG binding sites in ES cells and NPs, respectively. Unique reads from the control ChIP of Tdg$^{-/-}$ cells used to normalize the dataset for peakfinding in order to exclude non TDG-specific enrichments. Analysis of all TDG binding sites revealed 3622 and 844 promoter-associated sequences (TSS +/-2kb) in undifferentiated *Tdg*$^{+/-}$ ES cells and NPs, respectively. **C)** Percentage of total TDG bound sites associated to promoters (TSS +/-2kb), intergenic and genic regions for ES cells and NPs. In ES cells TDG localizes preferentially to promoter regions. In NPs, TDG shows preference for both promoters and intergenic sequences. **D)** The CpG content of total TDG bound sequences, promoter (TSS) associated, integenic and genic sequences was analyzed. Based on CpG content, TDG bound sequences were classified into strong and weak CpG Island and poor CpG regions according to [36]. TDG binds preferentially to strong CpG island promoters in both ES and NPs, whereas intergenic and genic sites are enriched for weak CpG islands and CpG poor sequences.

Figure 4. TDG binding to gene promoters is associated with an active chromatin environment.

A)/B) TDG bound CpG promoters (TSS) in ES cells/NPs are located in an active chromatin environment, as shown by the presence of the active histone mark H3K4me2. H3K4 methylation positive TDG-bound promoters are mostly actively transcribed. RNA polymerase II (Pol II) binding was used as a measure of transcriptional activity. A significant number of Pol II negative genes in ES and NPs are found in 'bivalent' chromatin domains, methylated at H3K4 and H3K27. **C)** Validation of TDG ChIPseq data by quantitative PCR at selected promoters in ES cells. Shown is the TDG specific enrichment of promoters regulating pluripotency (*Oct4, Nanog*) and developmental (*Pax6, Sox1, Sfrp2, HoxD13, Gata6, Tgfb2, Pdgfra)* genes in addition to a known TDG target (P0), used as a positive control for all TDG ChIP experiments. Developmental genes depicted are known polycomb targets and are prone to become aberrantly DNA methylated in cancer cells. **D)** H3K4me2 ChIP in $Tdg^{+/-}$ and $Tdg^{-/-}$ ES cells. Loss of TDG does not affect the H3K4 methylation pattern at bound promoters in ES cells. ChIP qPCR data is expressed as relative enrichment [% Input $_{(target)}$ / % Input $_{(control)}$]. TDG ChIPs were normalized against a chromosomal region devoid of any genes and CpG islands. H3K4me2 ChIPs were normalized against the methylated IAP LTR. Shown is the mean of three independent experiments with error bars indicating ±SE.

* $p<0.05$ by students T-test.

Figure 5. TDG loss does not affect global DNA methylation patterns

A) Scatter plot comparing DNA methylation values from replicate microarrays for 17'381 validated promoters in $Tdg^{+/-}$ (x-axis) and $Tdg^{-/-}$ (y-axis) NPs. Final promoter DNA methylation $\log_2$ ratios of immunoprecipitated (MeDIP) over input signal represent the average of two independent experiments. **B)** Scatter plot as in A) showing averaged DNA methylation values of all tested promoters (grey dots) and averaged DNA methylation

values of promoters shown to be bound TDG in ES cells by ChIP. 12.5% of TDG bound promoters show differential methylation between *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ NPs. **C)** Bisulfite sequencing of the region around the transcription start (bended arrow) of *Tgfb2* shows an increase in DNA methylation in *Tdg*$^{-/-}$ NPs, when compared to *Tdg*$^{+/-}$ cells.

Figure 6. TDG loss affects the establishment of gene expression patterns during differentiation

A) Effect of tdg knock out was more pronounced in neuronal precursor (NP) cells than in embryonal stem (ES) cells. Illumina gene expression array analysis results in more significantly changed gene transcripts in the neuronal differentiated stage at day 4 between heterozygotes and knockout mouse cells. A) Scatter plot showing averaged normalized fluorescence levels of three independent experiments, combining an arcsinh transformation for variance stabilization with quantile normalization from *Tdg*$^{+/-}$ (x-axis) and *Tdg*$^{-/-}$ ES cells (y-axis). Gray: 44573 probes are not significantly changed after correction of the *p* value to control the false discovery rate (FDR). Green: 695 significant probes (0.001 < *p* value < 0.05) comparing the two genotypes by moderated t. Red: 13 highly significant probes (*p* value < 0.001) **B)** Scatter plot showing averaged normalized fluorescence levels of three independent experiments from *Tdg*$^{+/-}$ (x-axis) and *Tdg*$^{-/-}$ ES cells (y-axis). Gray: 42968 probes are not significantly changed. Green: 2228 significant probes (0.001 < *p* value < 0.05) comparing the two genotypes by moderated t. Red: 85 highly significant probes (*p* value < 0.001) **C)** Significant (p<0.05) differentially regulated genes between *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ ES (229) and NPs (1022) were grouped into downregulated and upregulated genes.

Differential gene expression between genotypes was more pronounced in NPs than in ES cells, indicating that TDG loss affects the establishment of gene expression during cell differentiation. **D)** Gene ontology classifications of differentially regulated genes between *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ NPs. The GO term is on the x-axis, and the –log of the *P* value indicating

significance of enrichment ($P < 0.05$; yellow line). Data were analyzed through the use of Ingenuity Pathways Analysis (Ingenuity® Systems, www.ingenuity.com).

Figure 7. A) Total RNA was isolated from *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ 4hNPs and 48hNPs and *Pax6* expression was examined by RT-PCR. *Gapdh* was used as internal normalization. *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ ES cells differentiate into 4hNPs, confirmed by the increase in *Pax6* expression. After two days of plating in neuronal medium, *Tdg*$^{+/-}$ NPs efficiently downregulate Pax6 (50-fold, downregulation) whereas *Tdg*$^{-/-}$ NPs downregulate *Pax6* to a much lower extend (8-fold) Shown is the mean of three independent experiments with error bars indicating ±SE. *p<0.05 by students T-test. **B)** *Tdg*$^{+/-}$, *Tdg*$^{-/-}$ and TdgA overexpressing *Tdg*$^{-/-}$ ES cells were allowed to differentiate into 48hNPs. Total RNA was isolated at the neuronal stage and *Pax6* expression was determined by RT-PCR. *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ cells complemented with *TdgA* efficiently downregulate *Pax6* expression, whereas *Tdg*$^{-/-}$ still express high level of *Pax6*, indicating that misregulation of *Pax6* might be directly dependent on TDG. *Gapdh* was used as internal normalization. Shown is the mean of three independent experiments with error bars indicating ±SE. *p<0.05 by students T-test.

Figure 8. Working Model

A) TDG binds to the promoters of pluripotency and housekeeping genes when these are being actively transcribed, thus in an active chromatin state (H3K4me2+) and bound by RNA polymerase II (Pol II). Loss of TDG binding leads to premature methylation of CpG sequences and ultimately to gene silencing. As a consequence, the pluripotent state is lost in ES cells devoid of TDG. **B)** TDG binds to promoters of polycomb target genes, which are kept in an open chromatin environment but reversibly silenced by trimethylation of H3K27. Silencing by polycomb repressive complexes (PRC) in stem cells seem to target CpG promoters for

aberrant methylation of promoter CpGs in cancer cells. TDG might be involved in the protection of such sites from *de novo* DNA methylation. **C)** Possible mechanisms involving TDG mediated protection of CpG island promoters from premature and/or aberrant methylation. TDG might be involved in shielding unmethylated promoter CpG from methylation, in a process that involves inhibition of binding of Dnmt3a/3b to target CpGs (1). Alternatively, TDG could inhibit the methyltransferase activity DNMT3a/3b (2). Finally, TDG might be part of a DNA demethylation complex in which Dnmt3a/b deaminate methylated cytosines allowing TDG and other base excision repair proteins to repair the generated G/T mispairs, restoring the original C/G pair (3).

Figure S1. Localization of TDG to gene promoters in an active chromatin environment

TDG ChIP sequencing reads from ES cell were plotted in the UCSC genome browser (red) and overlapped with significant RNA polymerase II (RNA Pol II, blue) and H3K4me2 (green) ChIP sequencing reads. Shown are examples of TDG binding to the promoter CpG island of **A)** the actively transcribed *Dnmt3b* gene and the neuron specific gene *Neurog2*, which is found in an active chromatin state without being transcribed, and therefore "poised" to be expressed. reads from T(red) detected in the pluripotency genes **A)** Fes, **B)** Amn, **C)** Smc1b and **D)** Pou5f1 for ES cells (top) and differentiated NPs (bottom). CpG Islands and associated genes are shown in *green* and *blue*, respectively. TDG associates to the promoters of pluripotency genes only in undifferentiated ES cells, when these are not DNA methylated.

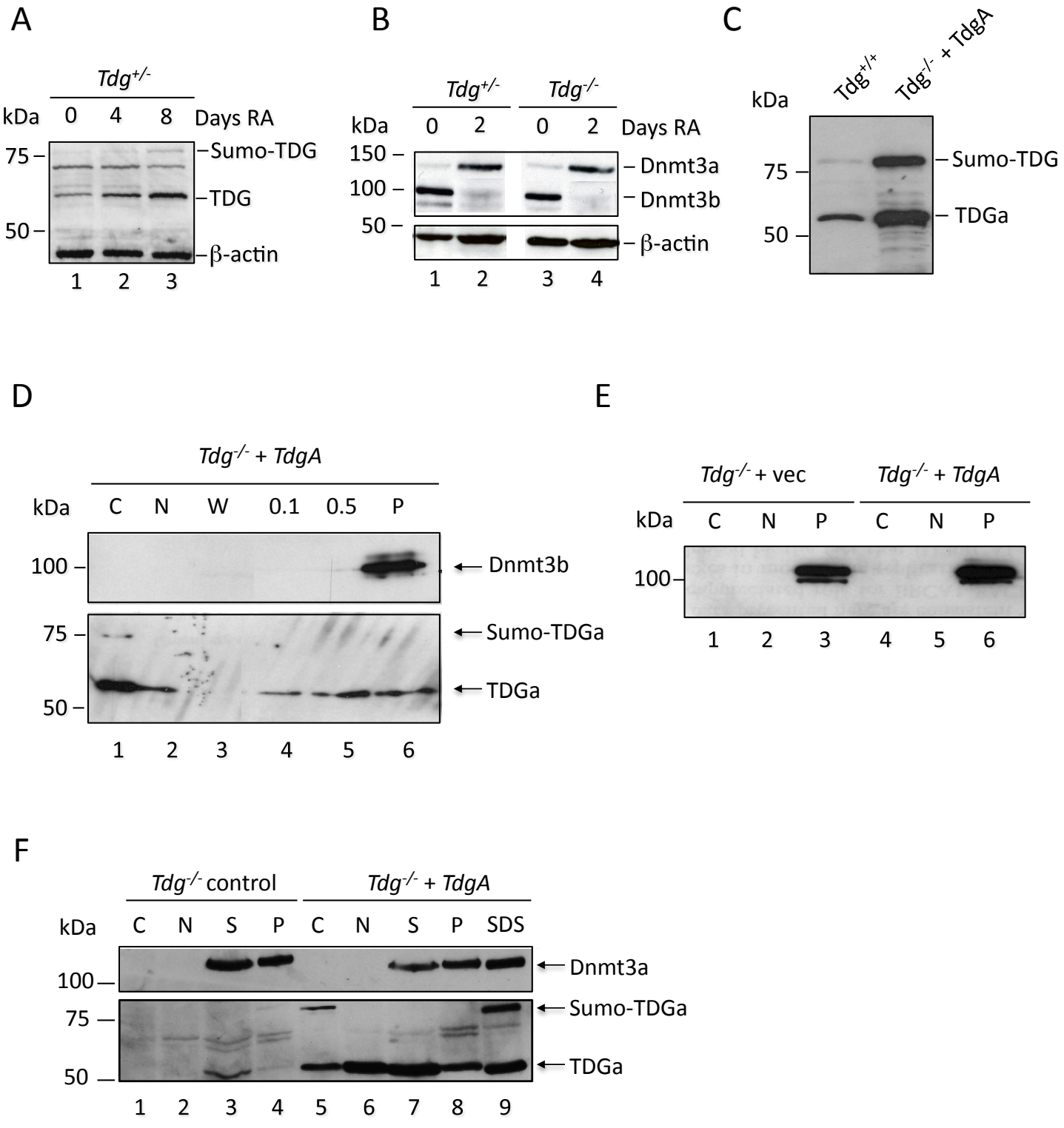
Figure S2. TDG only bind to pluripotency genes when the respective promoters are DNA unmethylated

A-C) TDG ChIP sequencing reads from ES cells and NPs were plotted in the UCSC genome browser (red). Enrichment of TDG binding was observed at the promoters of pluripotency genes in ES cells (shown in blue), whereas loss of binding is observed after *in vitro*

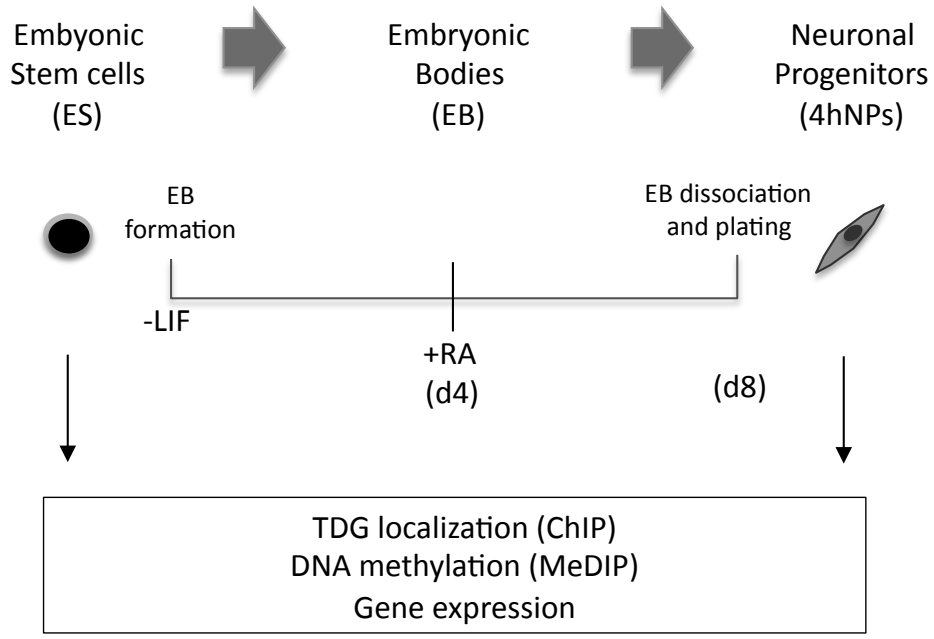
differentiation to NPs. **D)** TDG ChIP experiment in ES cells and mouse embryonic fibroblast (MEFs). *Oct4* promoter sequence was significantly enriched in *Tdg*$^{+/-}$ ES cells, whereas no enrichment could be detected in MEFs. TDG ChIPs were normalized against a chromosomal region devoid of any genes and CpG islands. Shown is the mean of three independent experiments with error bars indicating ±SE. *$p<0.05$ by students T-test.

FIGURES

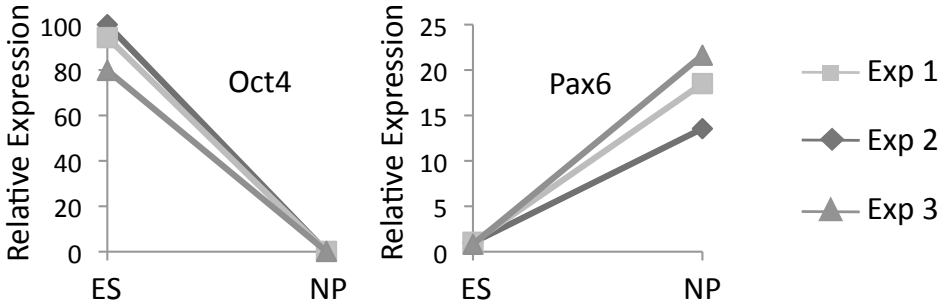
Cortázar et al., Figure 1



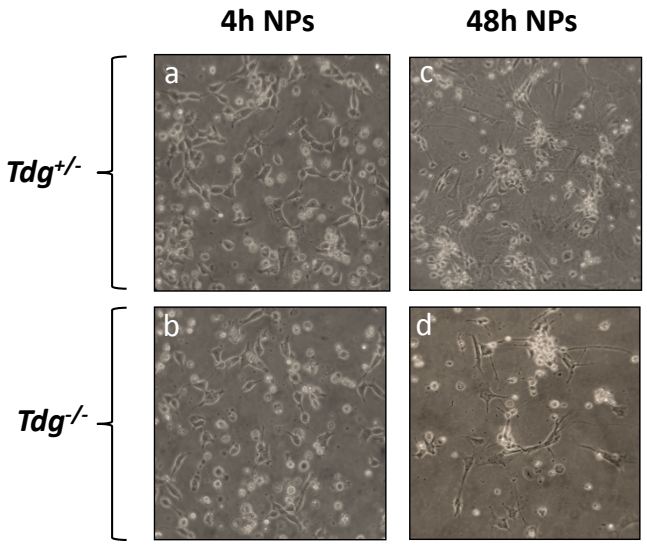
A


Embyonic
Stem cells
(ES)
Embryonic
Bodies
(EB)
Neuronal
Progenitors
(4hNPs)
EB
formation
EB dissociation
and plating
-LIF
+RA
(d4)
(d8)
TDG localization (ChIP)
DNA methylation (MeDIP)
Gene expression

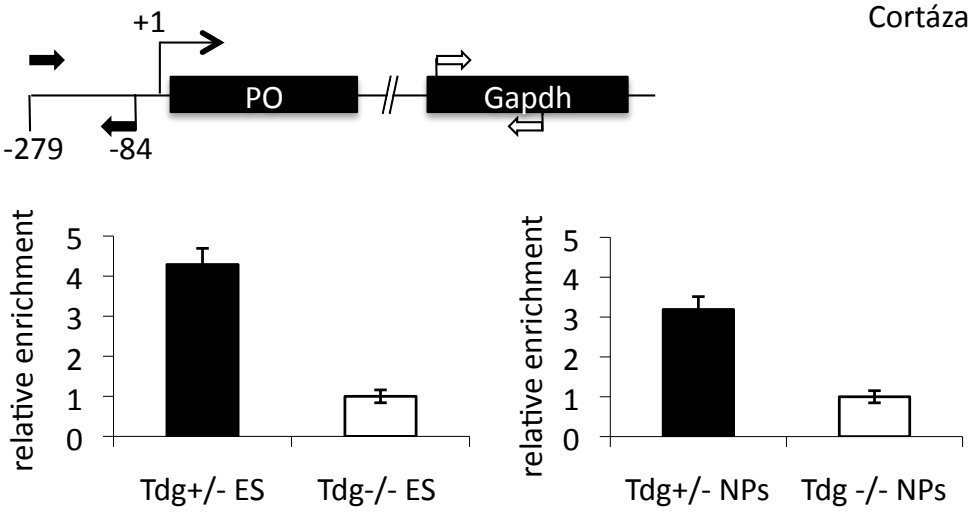
B


Relative Expression
100
80
60
40
20
0
Oct4
ES
NP
Relative Expression
25
20
15
10
5
0
Pax6
ES
NP
Exp 1
Exp 2
Exp 3

C


4h NPs
48h NPs
Tdg+/-
Tdg-/-
a
b
c
d

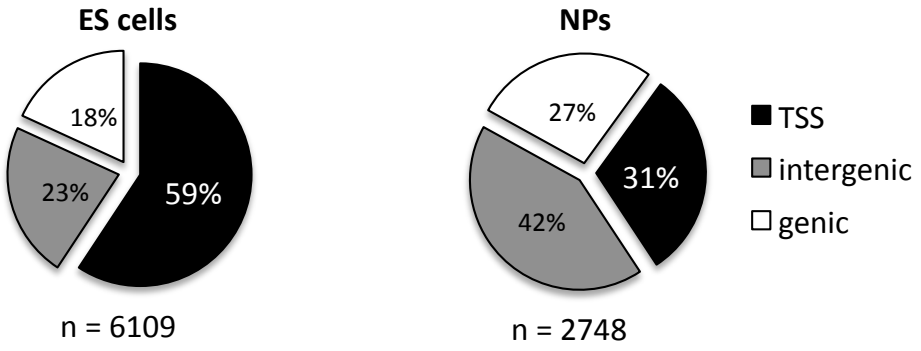
A



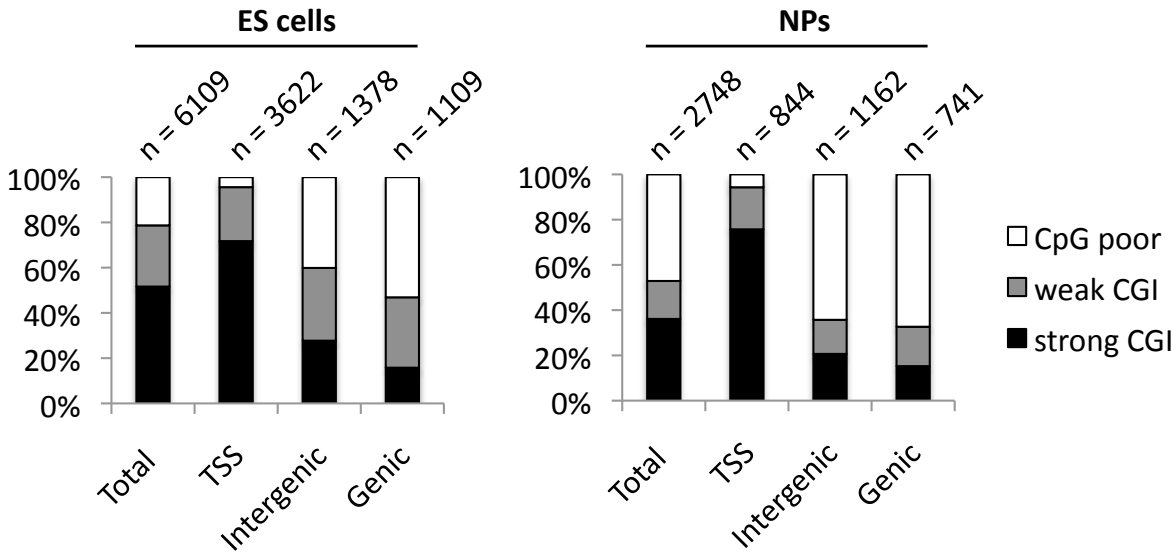
B

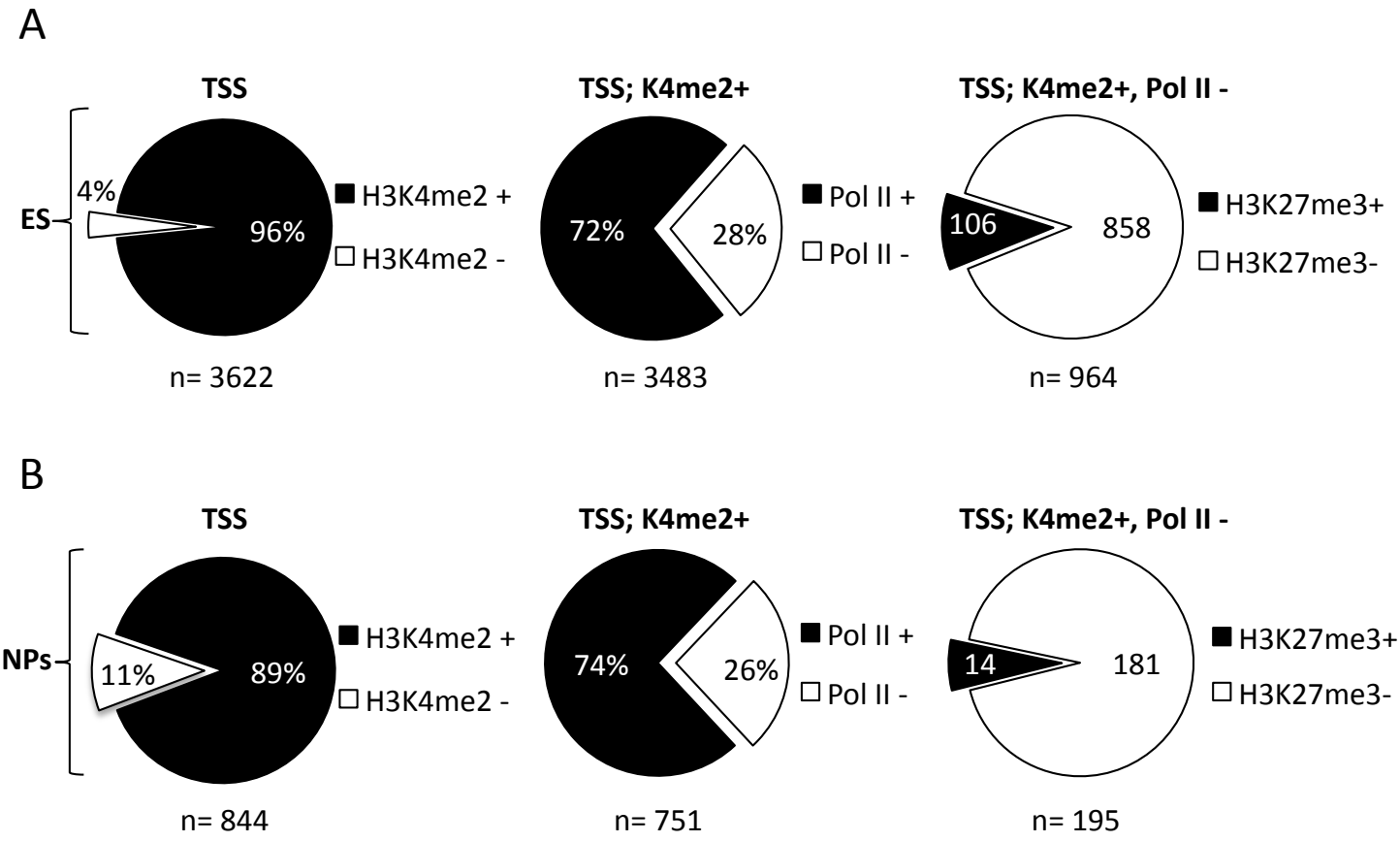
TDG ChIP Seq Data processing	ES cells	NPs
Total reads	6 x 10^6	6 x 10^6
Total mapped unique reads	3 x 10^6	3 x 10^6
MACS peak finder	6109	2748
TSS associated peaks	3622	844

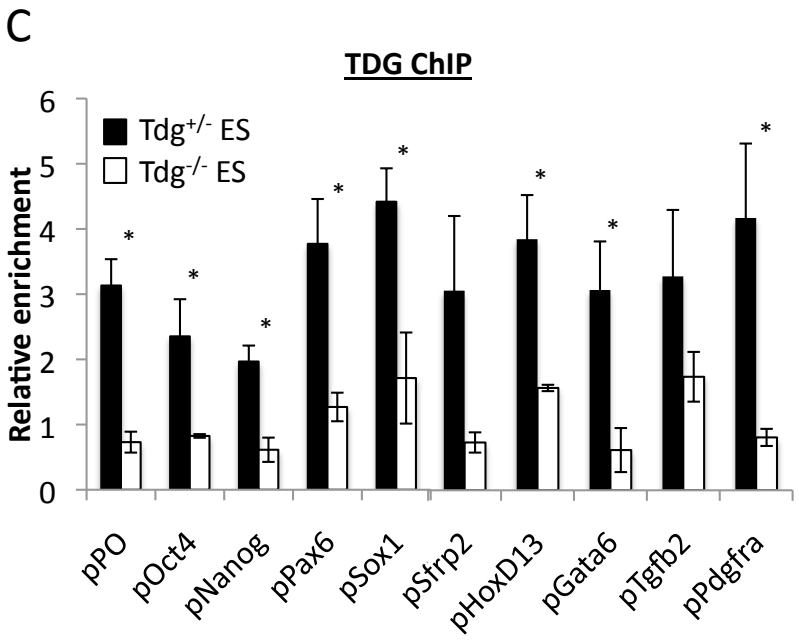
C



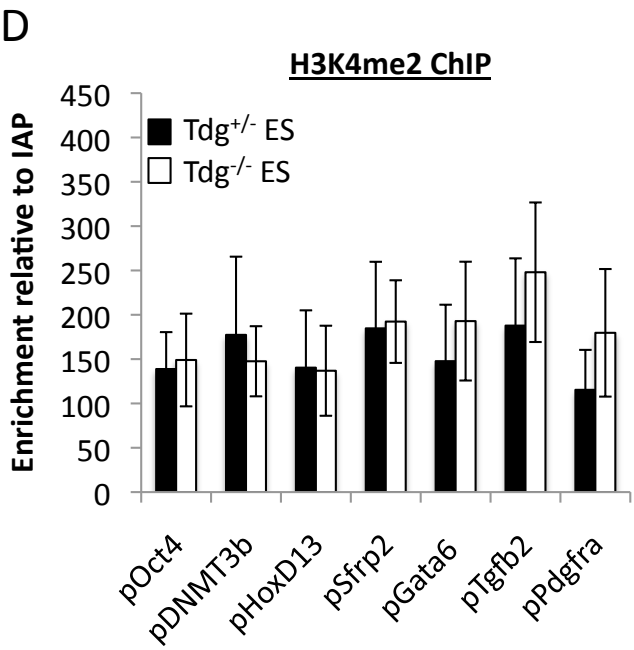
D



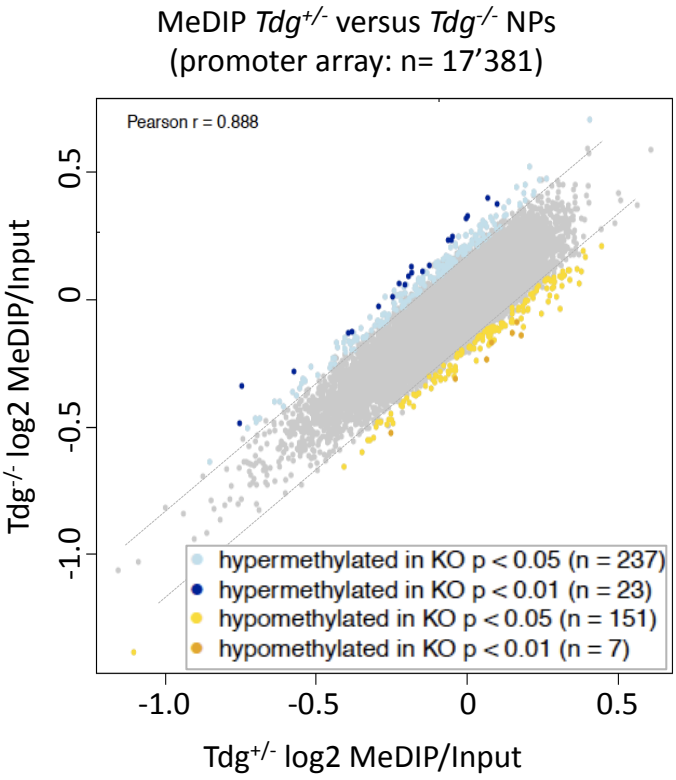




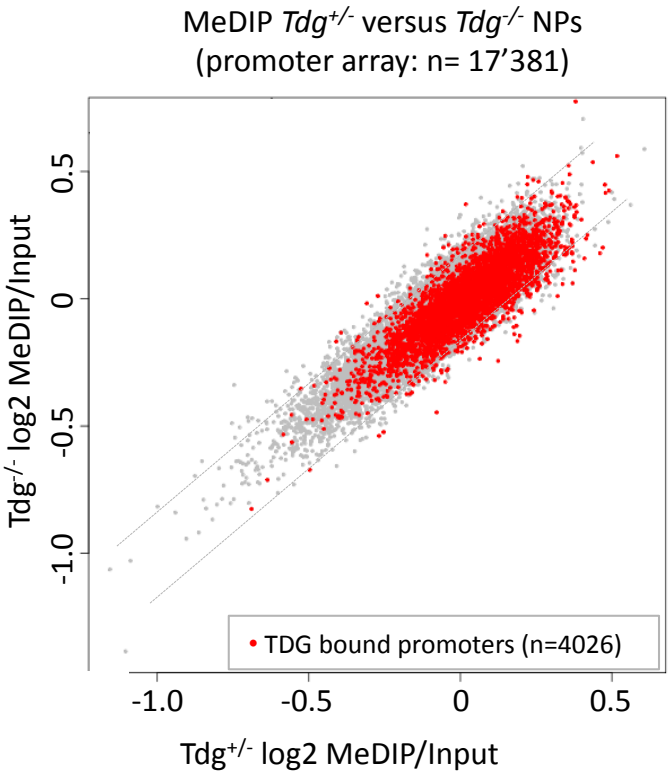
Polycomb target	+	+	+	+	+	+	+
Aberrant DNA methylation in cancer	+	+	+	+	+	+	+



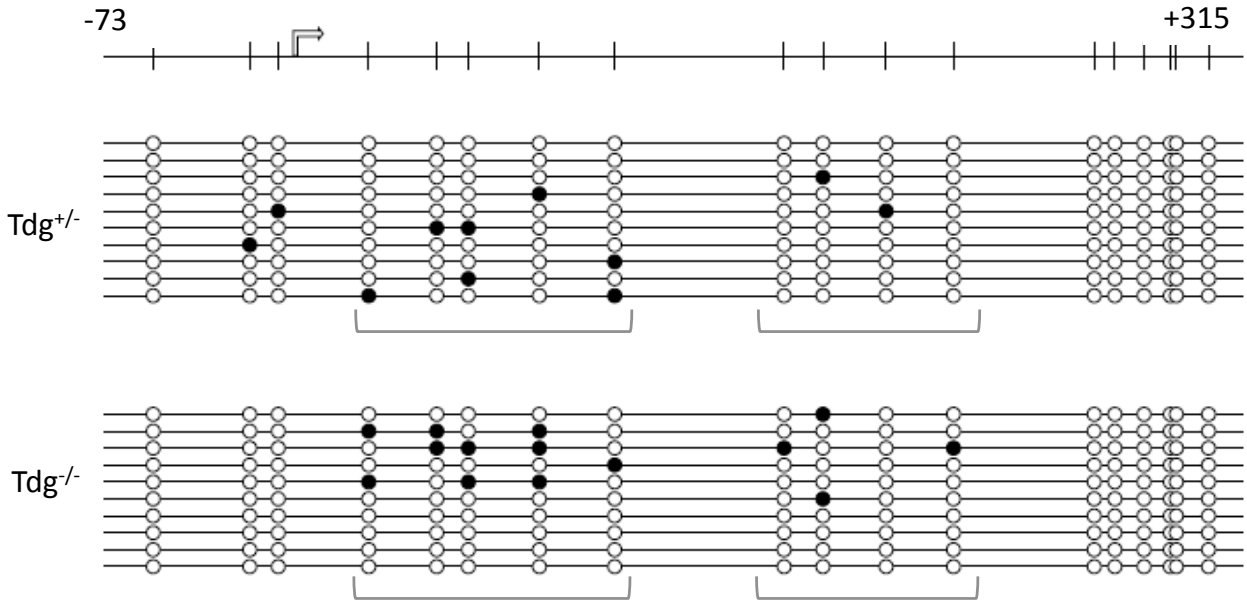
A

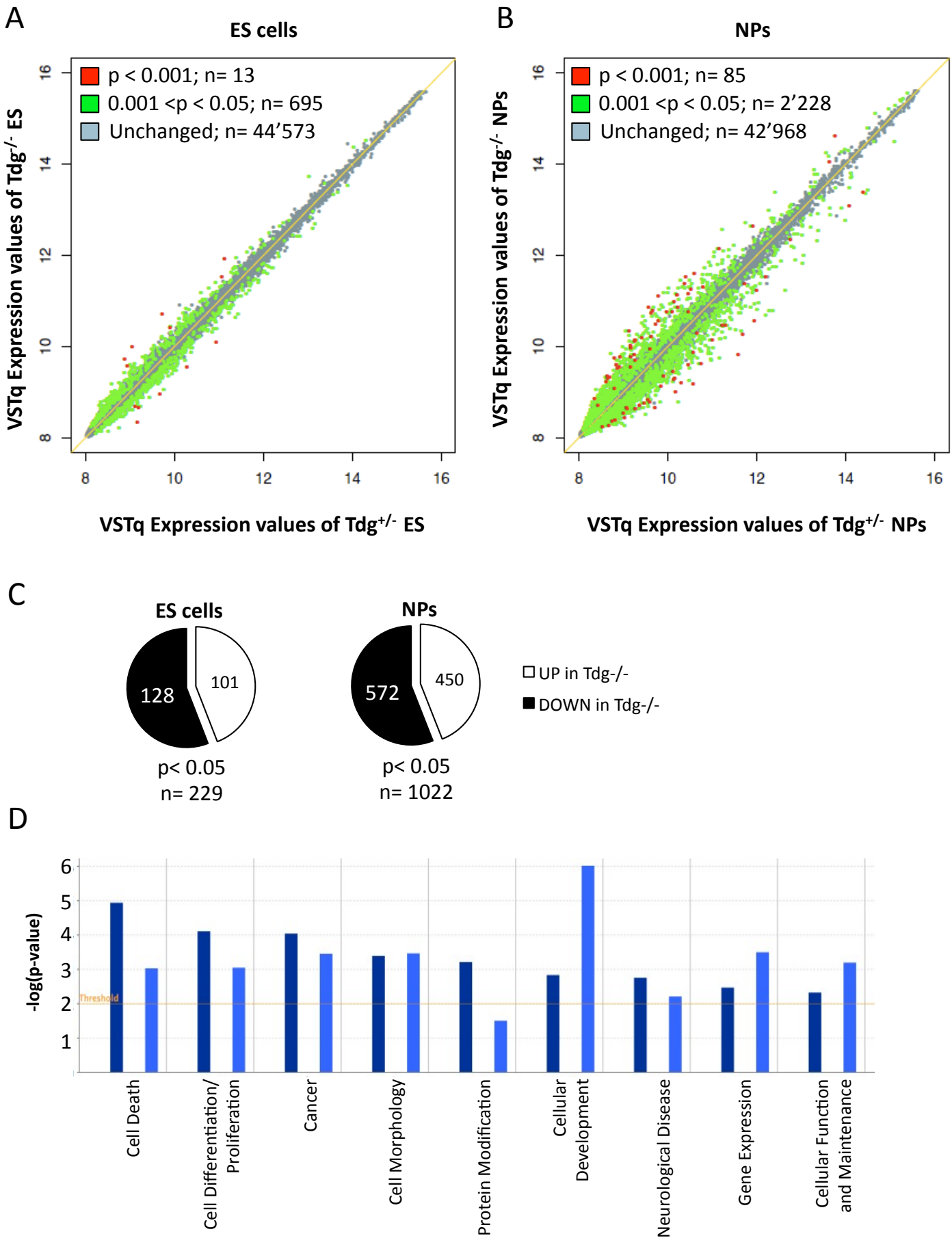


B

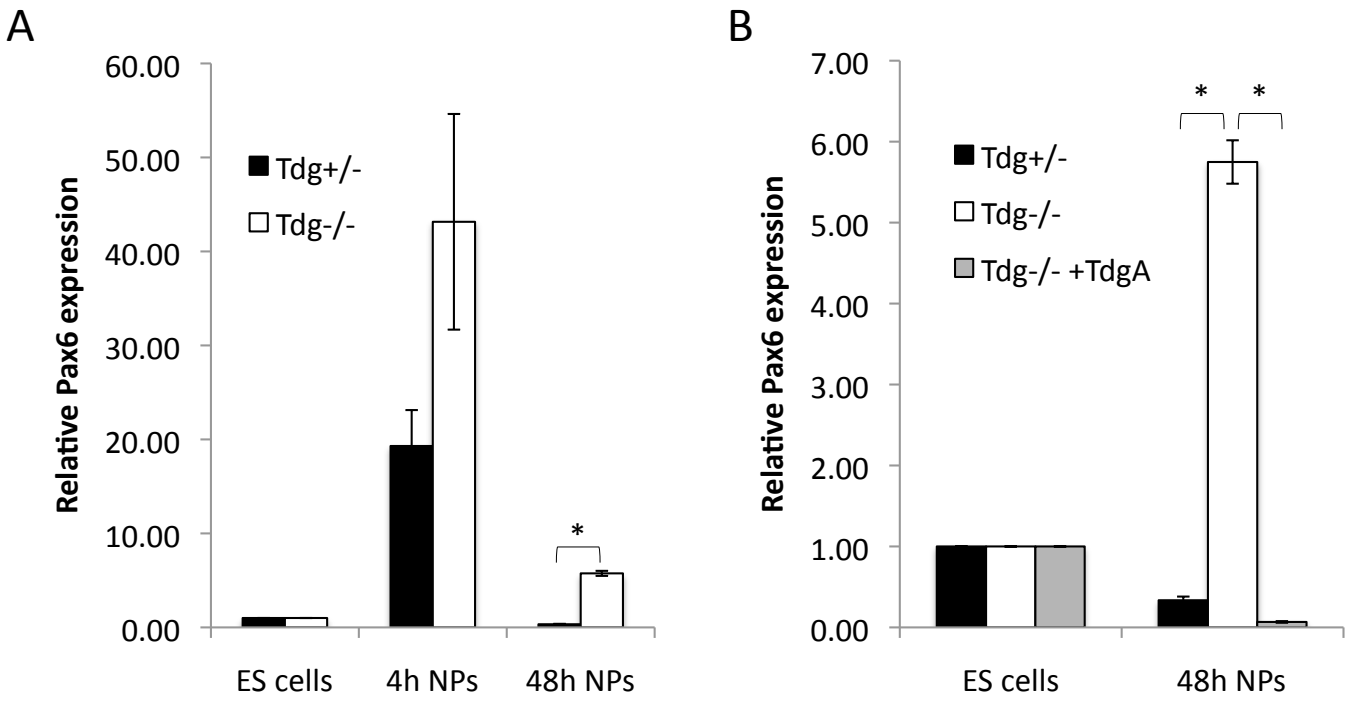


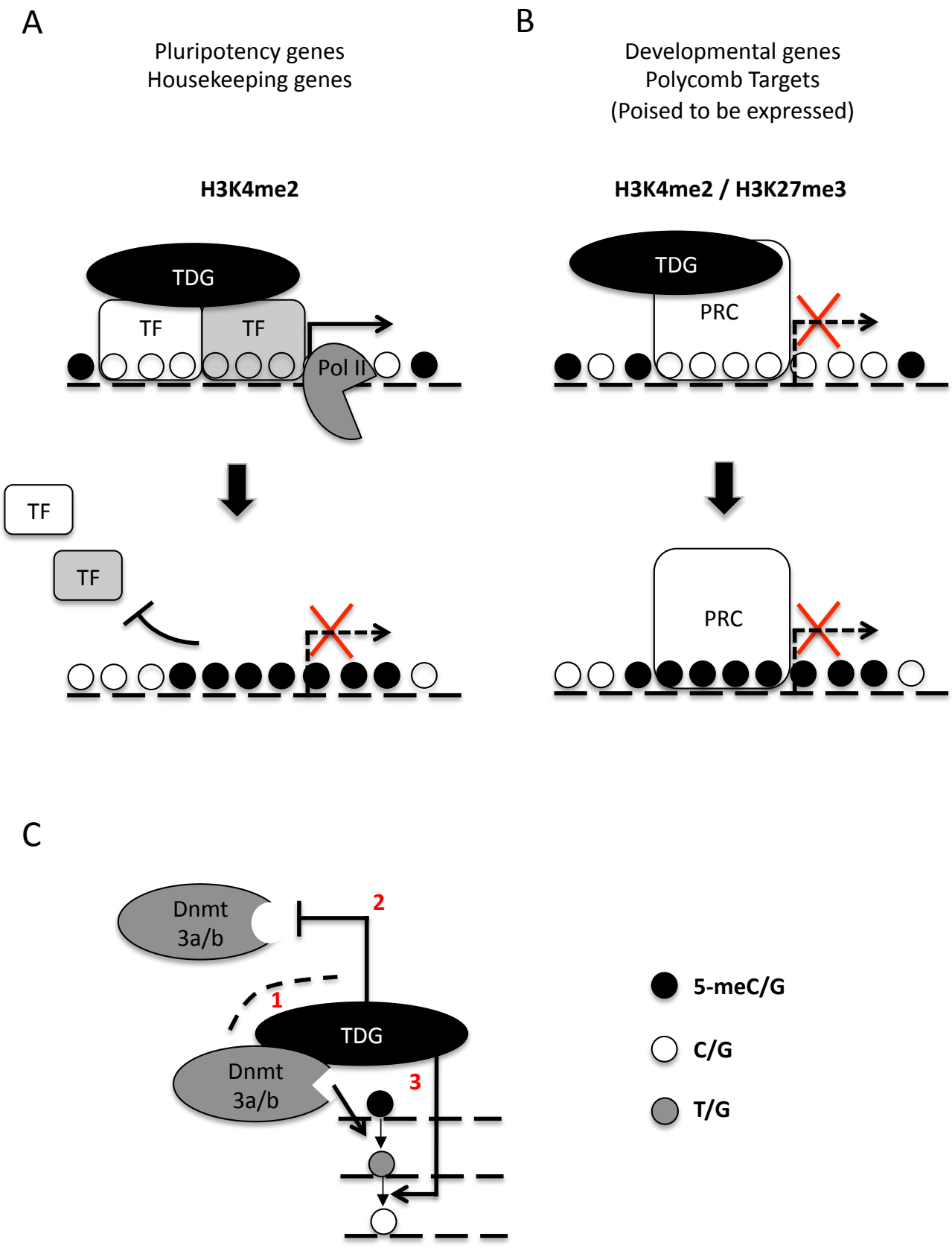
C Tgfb2 promoter region (-73-; +315)



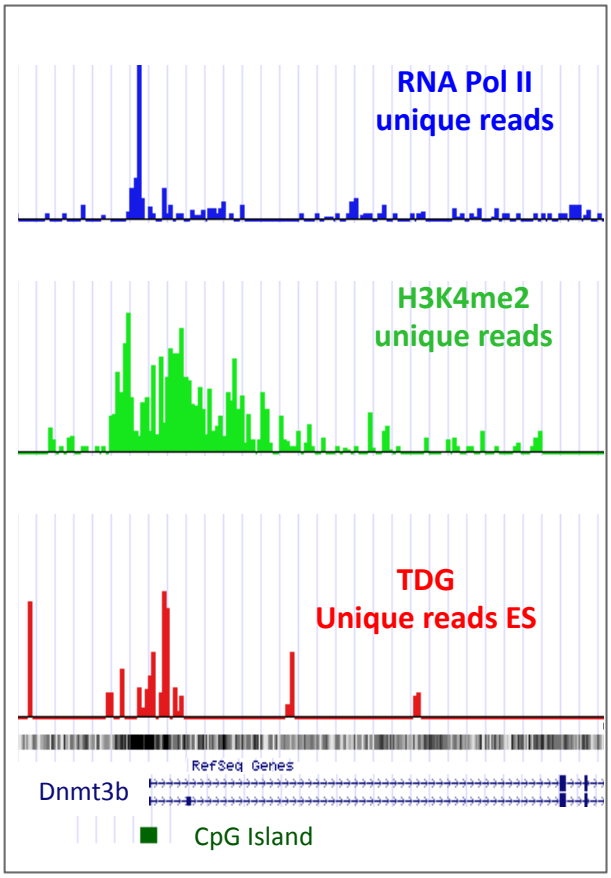

A
ES cells
p < 0.001; n= 13
0.001 <p < 0.05; n= 695
Unchanged; n= 44'573
VSTq Expression values of Tdg-/- ES
VSTq Expression values of Tdg+/- ES
B
NPs
p < 0.001; n= 85
0.001 <p < 0.05; n= 2'228
Unchanged; n= 42'968
VSTq Expression values of Tdg-/- NPs
VSTq Expression values of Tdg+/- NPs
C
ES cells
128
101
p< 0.05
n= 229
NPs
572
450
p< 0.05
n= 1022
UP in Tdg-/-
DOWN in Tdg-/-
D
-log(p-value)
Threshold
Cell Death
Cell Differentiation/ Proliferation
Cancer
Cell Morphology
Protein Modification
Cellular Development
Neurological Disease
Gene Expression
Cellular Function and Maintenance

Cortázar et al., Figure 7

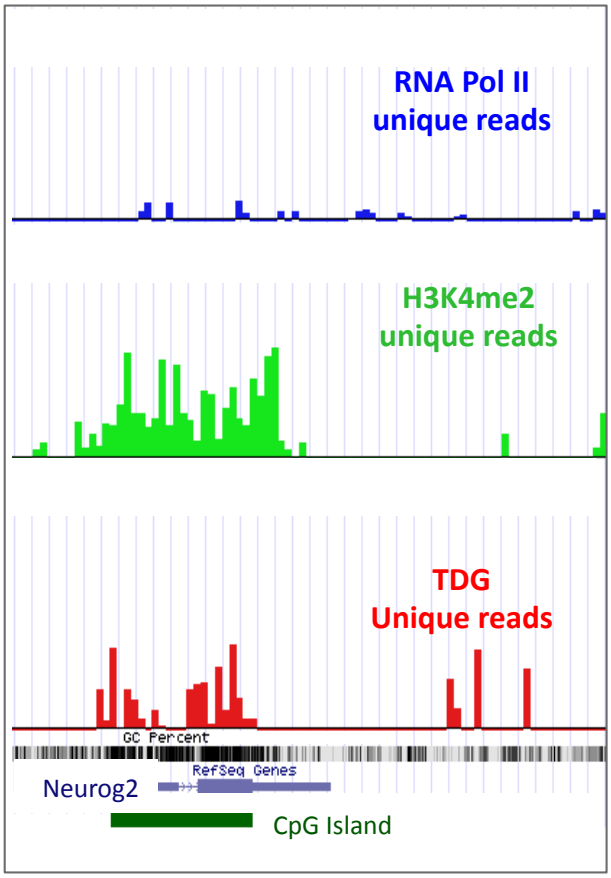



A
Pluripotency genes
Housekeeping genes
H3K4me2
TDG
TF
TF
Pol II
TF
TF
B
Developmental genes
Polycomb Targets
(Poised to be expressed)
H3K4me2 / H3K27me3
TDG
PRC
PRC
C
Dnmt
3a/b
2
1
TDG
Dnmt
3a/b
3
5-meC/G
C/G
T/G

A

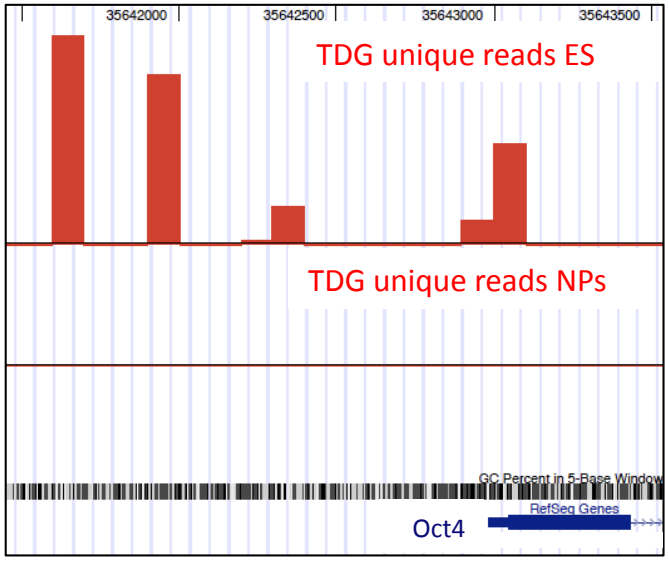

RNA Pol II
unique reads
H3K4me2
unique reads
TDG
Unique reads ES
RefSeq Genes
Dnmt3b
CpG Island

B


RNA Pol II
unique reads
H3K4me2
unique reads
TDG
Unique reads
GC Percent
RefSeq Genes
Neurog2
CpG Island

A

Oct4 Promoter

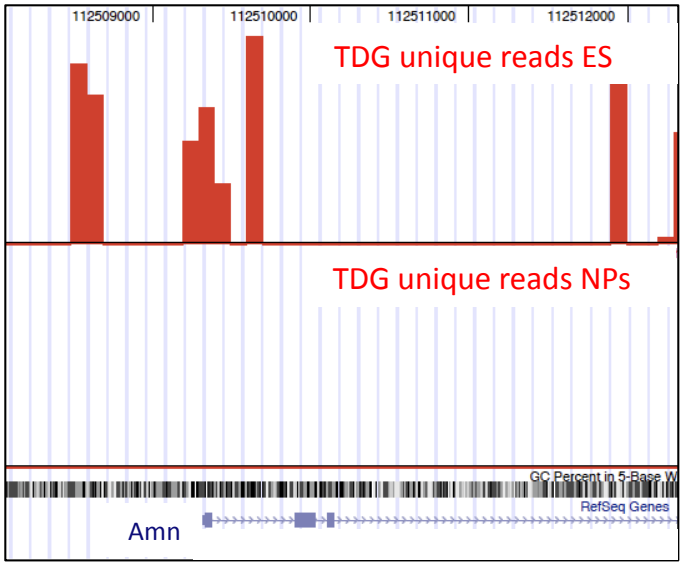


B

Fes Promoter

C

Amn Promoter



D

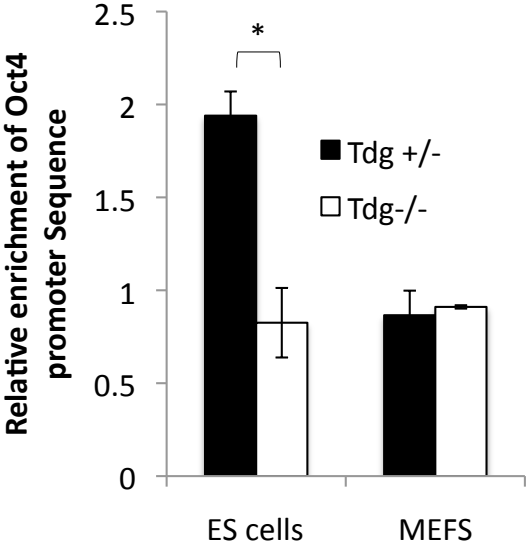


Table 1. Polycomb target pathways and respective genes regulated by TDG bound promoters

Polycomb Target Pathway	Targets bound by TDG in ES cells
HOX	HoxA, B, C, D clusters
Retinoic Acid	RARa, RORa
PAX	Pax1, 2, 3, 5, 6, 7, 9
FGF	Fgf1, 11, 12, 14, 20, 23
SOX	Sox1, 6, 9, 12, 13, 17
NOTCH	Notch1, 4, Hes5, Hes6
TGF	Smad1, 3, 4, 5, 6, Bmp2, 4, 7
WNT signaling	Wnt2, 5a, 10b, Sfrp2, Bcl, Lef1
HEDGEHOG	Shh
POLYCOMB	Cbx3, 4, 7

***** Observations are based on the analysis of TDG ChIP unique peaks for ES cells on the UCSC genome browser

Table 2. Correlation of TDG bound promoters regulating pluripotency and housekeeping genes in ES cells and NPs with *de novo* DNA methylation of promoter CpGs during differentiation from ES cells to NPs*.

Gene	Bound in ES	Bound in NPs	De novo methylated
	Y/N	Y/N	Y/N
Sall4	Y	Y	N
Sox2	Y	Y	N
Gdf3	Y	N	Y
Nanog	Y	N	Y
Oct4	Y	N	Y
Tdgf1	Y	N	Y
Amn	Y	N	Y
Sox30	Y	N	Y
Gapdh	Y	Y	N
Mlh1	Y	Y	N

***** Observations are based on the comparison of TDG ChIP unique sequence reads in ES cells and NPs using the UCSC genome browser. DNA methylation data was taken from *Mohn et. al, 2008.*

Supplementary Table 1. Primers used for ChIP and MeDIP qPCR

Oligo name	Sequence 5'->3'	Target
gapdh NAD (F)	AAC GAC CCC TTC ATT GAC	gapdh NAD domain
gapdh NAD (R)	TCC ACG ACA TAC TCA GCA C	gapdh NAD domain
gapdh (F)	CTC TGC TCC TCC CTG TTC C	Gapdh TSS
gapdh (R)	TCC CTA GAC CCG TAC AGT GC	Gapdh TSS
Oct4 PP(F)	GTG AGG TGT CGG TGA CCC AAG GCA G	Oct4 Promoter
Oct4 PP(R)	GGC GAG CGC TAT CTG CCT GTG TC	Oct4 Promoter
pDnmt3b(F)	GGT TAA GCG GCC CAA GTA A	Dnmt3b promoter
pDnmt3b(R)	ACC AGG GGA CCT GTC GTC	Dnmt3b promoter
pNanog(F)	GAG GAT GCC CCC TAA GCT TTC CCT CCC	Nanog promoter
pNanog(R)	CCT CCT ACC CTA CCC ACC CCC TAT TCT CCC	Nanog promoter
pOct4(F) fm1	ACC TCC GTC TGG AAG ACA CA	Oct4 promoter
pOct4(R) fm2	TCA CCT AGG GAC GGT TTC AC	Oct4 promoter
pPax6(F)	CGG TGA AAG AAG CCA CTA GG	Pax6 promoter
pPax6(R)	TAG GGC GTT TGT TTC CAA AT	Pax6 promoter
pPO (R)	CTG GTT CCA TCG ACT GTC CT	PO promoter
pPO(F)	CGC AGA AAG TTG TTT TGC TG	PO promoter
pSfrp2 (F)	GAC TTT CGT TGC CTC CTC CT	Sfrp2 promoter
pSfrp2 (R)	AGG CCG GTC ACT ACT TTC TG	Sfrp2 promoter
pSox1(F)	GCA CAG CTC TCT GGC TCT CT	Sox1 promoter
pSox1(R)	CTC TTG TCG GCT CGA AGT CT	Sox1 promoter

pGata6 (F)	AGT TTT CCG GCA GAG CAG TA	Gata6 promoter
pGata6 (R)	AGG AGG AAA CAA CCG AAC CT	Gata6 promoter
pHoxD13 (F)	TGG GCT ATG GCT ACC ACT TC	HoxD13 promoter
pHoxD13 (R)	GAC ACT TCC TTG GCT CTT GC	HoxD13 promoter
pTgfb2(F)	AAG GGA CGA GAC GAG AAG GT	Tgfb2 promoter
pTgfb2(R)	ACA TCC ACA CGC ACA CTC AT	Tgfb2 promoter
pPdgfra(F)	GGA CGA GCG ATC TGG AAT AA	Pdgfra promoter
pPdgfr2(R)	CCG TGC AGA AAA GAC TCC AC	Pdgfra promoter
pChr2_neg(F)	AGC ACA GCC TGA AGC CTC TA	ChIP control intergenic
pChr2-neg(R)	AGA GGG CAT TTC CGT CTT TT	ChIP control intergenic
IAP (F)	CTC CAT GTG CTC TGC CTT CC	IAP LTR
IAP (R)	CCC CGT CCC TTT TTT AGG AGA	IAP LTR

The Enigmatic Thymine DNA Glycosylase

Daniel Cortázar, Christophe Kunz, Yusuke Saito, Roland Steinacher, Primo Schär*

Department of Biomedicine

University of Basel

Mattenstrasse 28

4058 Basel

Switzerland

* Corresponding author: primo.schaer@unibas.ch

Phone: +41 61 267 0767

Fax: +41 61 267 3566

DNA REPAIR 6 (2007) 489–504

available at www.sciencedirect.com

ScienceDirect

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

The enigmatic thymine DNA glycosylase

*Daniel Cortázar, Christophe Kunz, Yusuke Saito, Roland Steinacher, Primo Schär**

Centre for Biomedicine, Department of Clinical Biological Research, University of Basel, Basel, Switzerland

ARTICLE INFO

Article history:
Published on line 20 November 2006

Keywords:
BER
TDG
SUMO
Deamination

ABSTRACT

When it was first isolated from extracts of HeLa cells in Josef Jiricny's laboratory, the thymine DNA glycosylase (TDG) attracted attention because of its ability to remove thymine, i.e. a normal DNA base, from G·T mispairs. This implicated a function of DNA base excision repair in the restoration of G·C base pairs following the deamination of a 5-methylcytosine. TDG turned out to be the founding member of a newly emerging family of mismatch-directed uracil-DNA glycosylases, the MUG proteins, that act on a comparably broad spectrum of base lesion including G·U as the common, most efficiently processed substrate. However, because of its apparent catalytic inefficiency, some have considered TDG a poor DNA repair enzyme without an important biological function. Others have reported 5-meC DNA glycosylase activity to be associated with TDG, thrusting the enzyme into limelight as a possible DNA demethylase. Yet others have found the glycosylase to interact with transcription factors, implicating a function in gene regulation, which appears to be critically important in developmental processes. This article reviews all these developments in view of possible biological functions of this multifaceted DNA glycosylase.

1. Introduction

Within cells, the chemically unstable DNA is under permanent hydrolytic and chemical attack. Hydrolytic reactions occur at a significant rate and include the deamination of DNA bases with exocylic amino groups, i.e. cytosine (C) and 5-methylcytosine (5-meC), adenine (A) and guanine (G) [1]. Deamination of C and 5-meC generates uracil (U) and thymine (T) mispaired with guanine, respectively, both giving rise to C·G to T·A transition, unless repaired. While U is a foreign base in DNA and is easily recognized and repaired as such, the correction of a deaminated 5-meC, i.e. a T, requires a higher level of sophistication at damage recognition, since the "damage" in this case is a perfectly normal DNA base, except that it is mispaired. Such thoughts led Josef Jiricny and colleagues to search for a DNA repair function that processes T when mispaired with G to restore a canonical G·C base pair. In transfection experiments with G·T mismatched SV40 DNA they indeed identified a G·T directed repair activity in African green monkey kidney cells that efficiently replaced the T with a C [2]. The subsequent purification of a G·T binding and processing enzyme from nuclear extracts of HeLa cells and the molecular cloning of the respective cDNA eventually led to the discovery of the human thymine DNA glycosylase (TDG) [3–5], the first mismatch-specific DNA glycosylase to be described. Its ability to hydrolyze thymine and uracil from G·T and G·U mispairs *in vitro* [6] implicated a specific biological role in base excision repair (BER) of deaminated 5-meC and C, i.e. in countering deamination-induced C → T mutation.

During the last decade, research on TDG has seen an impressive expansion into different disciplines. Enzymatic and structural studies provided insight into different aspects of its functionality. The identification and characterization of homologs and orthologs of species across the phylogeny shed

* *Corresponding author*. Tel.: +41 61 267 0767; fax: +41 61 367 3566.
E-mail address: primo.schaer@unibas.ch (P. Schär).
1568-7864/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.dnarep.2006.10.013

light on the evolution of this family of DNA glycosylases and facilitated first genetic approaches towards unraveling biological functions. While so far all efforts have failed to assign the human TDG to a well-defined cellular process, they have established lines of evidence that support three main working hypotheses. The biochemical and structural properties of TDG support a function in DNA BER of damaged or modified pyrimidine bases; biochemical and cell biological evidence has suggested a role in the active removal of 5-meC from methylated CpG dinucleotides in DNA; and protein–protein interactions have implicated TDG in the regulation of gene expression. What seems to be clear from all these studies is that, as a DNA glycosylase, TDG has some rather unusual features and that these may hint towards a link between DNA repair, the control of epigenetic DNA modification and the regulation of gene expression. Whether and how these seemingly divergent aspects of TDG function can be reconciled in a unifying mechanistic model remains to be addressed in the future. The objective of this article is to review the results of the last decade of research on TDG and to evaluate the emerging concepts for a biological function.

2. TDG—protein structure and enzymology

2.1. *Primary structure and the evolutionary aspects*

The cloning and sequencing of the cDNA encoding the human TDG [5] facilitated the search for related proteins in other organisms. This disclosed a broad phylogeny with orthologs in bacteria (e.g. *Escherichia coli*) [7], yeasts (e.g. *Schizosaccharomyces pombe*), insects (e.g. *Drosophila melanogaster*) [8] and frogs (e.g. *Xenopus laevis*). All these proteins belong to the MUG branch of the superfamily of monofunctional uracil-DNA glycosylases (UDGs) that share a common α/β-fold structure [9] (Fig. 1). Although the human TDG is the founding member, the family was named after the *E. coli* Mug protein, a mismatch-specific uracil-DNA glycosylase, to account for the fact that U rather than T processing is a common trait of these proteins.

The MUG proteins have a simple domain architecture, they are composed of a conserved core that constitutes the active site, and non-conserved N- and C-terminal extensions of variable lengths (Fig. 2). Within their catalytic domains, all orthologs share between 37% and 52% amino-acid sequence identity but no significant similarities with members of other UDG families, e.g. UNG and SMUG proteins. The molecular masses of the MUGs range between 18 kDa and 46 kDa with one notable exception. Owing to its uniquely long N- and C-terminal sequences, the Drosophila ortholog (Thp1) is a remarkably sized DNA glycosylase of more than 191 kDa. Similarly large proteins with active DNA glycosylase domains have thus far been described only in plants (i.e. DEMETER, ROS1), where they appear to control cytosine methylation mediated gene silencing [10,11].

Recent studies shed some light on possible functions of the divergent N- and C-terminal domains of the eukaryotic MUGs. Variations in the composition and configuration of these termini appear to correlate with changes in substrate specificity, substrate interaction and the kinetics of base release; they may thus be there to modulate the enzymatic activity of

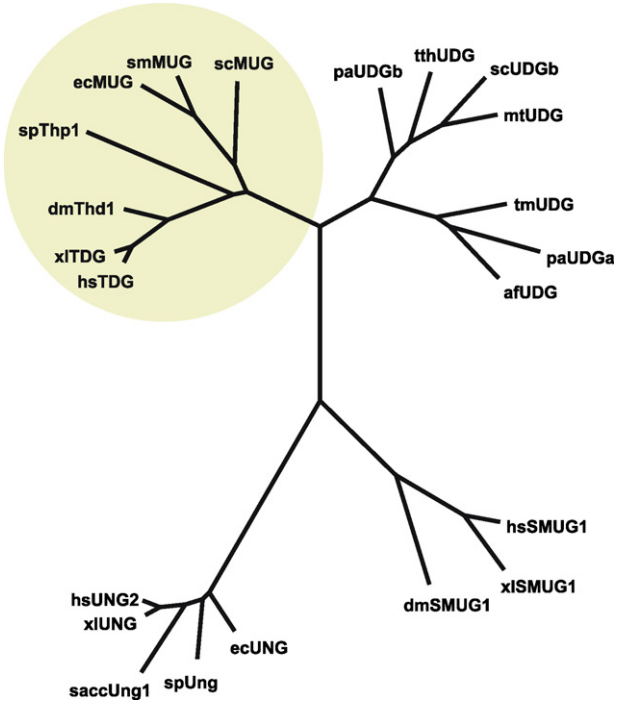


Fig. 1 – Evolutionary conservation of MUG proteins. (A) Shown are the clustered relationships between representative members of UDG superfamily. The MUG family is highlighted in green. Included are *Homo sapiens* TDG (hsTDG, accession no. Q13569), UNG2 (hsUNG2, P22674) and SMUG1 (hsSMUG1, Q53HV7); *Xenopus leavis* UNG (xlUNG, AAH72313), TDG (xlTDG, AAH77465.1) and SMUG1 (xlSMUG1, Q9YGN6); *Drosophila melanogaster* Thd1 (dmThd1, Q9V4D8) and Smug1 (dmSMUG1, Q9VEM1); *Schizosaccharomyces pombe* Ung1 (spUng1, O74834) and Thp1 (spThp1, O59825); *Saccharomyces cerevisiae* Ung1 (saccUng1, P12887); *Escherichia coli* Ung (ecUNG, P12295) and Mug (ecMUG, P0A9H1); *Serratia marcescens* Mug (smMUG, P43343); *Streptomyces coelicolor* MUG (scMUG, NP_625542) and UDGb (scUDGb, NP_626251); *Archaeoglobus fulgidus* UDG (afUDG; NP_071102); *Thermotoga maritima* UDG (tmUDG, NP_228321); *Pyrobaculum aerophilum* UDGa (paUDGa, NP_558739) and UDGb (paUDGb, NP_559226); *Mycobacterium tuberculosis* UDG (mtUDG, NP_335742); and *Thermus thermophilus* UDG (tthUDG, CAD29337). The tree was generated with the neighbor-joining algorithm of the MEGA 3.1 software applied to a multiple alignment produced with the ClustalW routine (Blosum matrix; pairwise alignments: Gap opening penalty 10, Gap extension penalty 0.1; multiple alignment: Gap opening penalty 10, Gap extension penalty 0.1).

TDG in a process-dependent manner [7,8,12,13]. The terminal domains of the human and murine TDGs were also found to mediate diverse physical and functional interactions with other proteins, including nuclear receptors and other transcriptional regulators. This emphasizes a possible species-specific role of these domains in targeting the glycosylase to specific DNA sequences in the genome where its activity is needed [14–20].

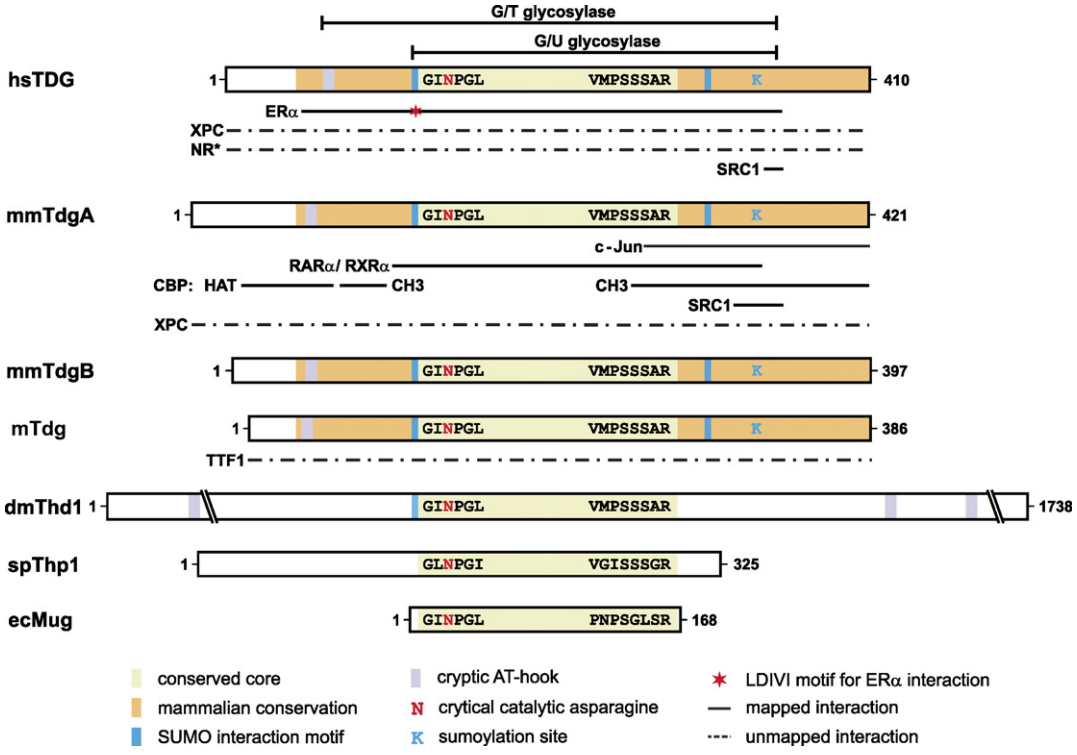


Fig. 2 – Schematic alignment of MUG proteins. Conserved sequences and protein motifs as well as known interactions with other proteins are shown as indicated in the legend. The highly conserved central domain harbors the sequence motifs G(I/L)NPG(L/I) and VMPSSSAR (hsTDG), representing critical residues of the active site. Similarities in the N-and C-terminal parts of the mammalian TDGs are confined to the SUMO-interaction motifs and the SUMOylation consensus motif VKEE. Bars on top of the human TDG designate the minimal sequence requirement for G·U or G·T processing. The predicted AT-hook motifs present in the mammalian and insect N- or C-termini may provide non-specific DNA binding capacity. Bars underneath the respective MUG orthologs indicate identified protein interactions. NR, nuclear receptors; androgen receptor, glucocorticoid receptor, progesterone receptor, peroxisome proliferator-activated receptor a, thyroid hormone receptor a, Vitamin D3 receptor.

2.2. *Three-dimensional structure and implicated mechanisms*

A deeper understanding of structure–function aspects came with the resolution of the three-dimensional structure of the *E. coli* Mug protein. Laurence Pearl and collaborators compiled an α/β-fold for Mug that closely resembles that of the functionally related UNG and SMUG proteins. In the light of the very limited amino-acid sequence homology between the members of these UDG families, this was a rather surprising finding. Crystals capturing Mug bound to its DNA substrate then confirmed that, like other UDGs, it utilizes a combined intercalation/nucleotide flipping mechanism for base recognition and binding [21,22]. A closer look into the active site, however, revealed some unique features of Mug. Whereas UNG and SMUG enzymes achieve a high substrate selectivity through strategic configurations of active site residues that establish specific contacts with the uracil [23–25], Mug forms a large hydrophobic catalytic cavity that accommodates a variety of pyrimidine and purine derivatives without contacting the base to be hydrolyzed [22]. Hence, MUG proteins have a comparably broad substrate spectrum that includes lesions as bulky as ethenoadducts of C and A (Table 1) [8,12]. Another unique feature of Mug is its interaction with the complementary DNA strand opposite from the damaged base. Residues within the catalytic pocket (Gly143, Leu144 and Arg146) form a 'wedge' that intercalates into the DNA base stack from the minor groove and occupies the space of the substrate base. This wedge establishes specific hydrogen-bonding interactions with the widowed G in a configuration that mimics Watson–Crick base pairing. These may account for the strict double-strand dependency of the MUG proteins and, since they are absolutely specific for G, also explain their opposite G preference [21]. Thus, Mug uses the complementary base for substrate discrimination, while other UDGs establish specific contacts with the substrate base.

Mutational analyses and a recently resolved crystal structure of the catalytic core of the human TDG are fully consistent with the structure–function model postulated on the basis of the *E. coli* Mug structure. It therefore appears that the basic catalytic mechanism can be extrapolated from Mug to human TDG and probably to the entire MUG family [26,27]. However, the Mug model fails when it comes to explain the effects of the N- and C-terminal domains on the substrate spectrum and the kinetic properties of the eukaryotic MUGs.

Table 1 – Substrate preferences of TDG orthologs[a]

Substrate[b]	hsTDG[c]	ΔNhsTDG[d]	ecMug[d]	spThp1p[c]	dmThd1p[c]
G·**U**	+++	+++	+++	+++	+++
A·**U**	+	+	+	+++	++
ss**U**	–	–	–	+++	–
G·^{F}**U**	+++	nd	nd	+++	+++
A·^{F}**U**	++	nd	nd	+++	++
ss^{F}**U**	++	nd	nd	++	++
G·^{Br}**U**	+++	nd	nd	++	+++
A·^{Br}**U**	+	nd	nd	+	–
ss^{Br}**U**	–	nd	nd	+	–
G·^{Hm}**U**	+++	nd	+	–	++
G·^{H}**U**	+++	nd	++	+++	nd
G·**T**	+++	–	–	–	++
G·**Tg**	++	nd	nd	nd	nd
G·**εC**	+++	nd	+++	+++	+++
A·**εC**	++	nd	nd	+++	++
ss**εC**	–	nd	+	+++	–
G·**Hx**	+	+	nd	+++	+
T·**Hx**	–	–	–	+++	–
ss**Hx**	–	nd	nd	+++	–
G·**εA**	–	nd	+	++	–
T·**εA**	–	nd	–	++	–
ss**εA**	–	nd	nd	+	–
G·^{m}**C**	–/+	nd	nd	–	–
G·^{He}**C**	–	nd	++	+++	nd
G·^{Hp}**C**	–	nd	–	+++	nd
G·**G**		nd	nd	+	

[a] Indicated are relative processing efficiencies of recombinant human full size (hsTDG) and N-terminally truncated TDG (ΔNhsTDG) and the orthologs of *E. coli* (ecMug), *S. pombe* (spThp1p) and *D. melanogaster* (dmThd1p). Base release efficiencies are indicated as: +++, high; ++, intermediate; +, low; –, insignificant.

[b] Putative substrate bases are indicated in bold letters. Abbreviations used are: ss, single-stranded DNA; ds, double-stranded DNA; ^{F}U, 5-fluorouracil; ^{Br}U, 5-bromouracil; ^{Hm}U, 5-hydroxymethyluracil; ^{H}U, 5-hydroxyuracil; Tg, thymine glycol; εC, 3,N^4-ethenocytosine; Hx, hypoxanthine; εA, 1,N^6-ethenoadenine; ^{m}C, 5-methylcytosine; ^{H}C, 5-hydroxycytosine; ^{He}C, 3,N^4-a-hydroxyethanocytosine; ^{Hp}C, 3,N^4-a-hydroxypropanocytosine; nd, not done.

[c] Fully AP-site inhibited; no enzymatic turnover.

[d] Partially AP-site inhibited; slow enzymatic turnover.

2.3. *Enzymatic properties of TDG*

2.3.1. *Substrates of TDG*

Several laboratories have investigated substrate processing features of MUG proteins. Some of these studies made use of specific chemical modifications at or near the target base (U or T) or its mispaired vis-à-vis (G) to explore mechanisms of substrate interaction and base hydrolysis by TDG. Others examined synthetic candidate base lesions with the aim to identify biologically relevant substrates [28]. We will focus our account here on a few representative substrates that illustrate the functional versatility and typical mechanistic properties of MUG proteins and implicate main lines of possible biological functions.

Although the human TDG is best known for its ability to remove T from a T·G mismatch, MUG proteins of different origin were shown to have rather broad substrate spectra with U mispaired to G being the common, most efficiently processed physiological DNA lesion (Table 1) [7,8,12,29]. Derivatives of U with modifications or substituents at the 5-carbon position such as 5-hydroxy-U, 5-hydroxymethyl-U, 5-fluoro-U (5-FU) and 5-bromo-U (5-BrU) turned out to be very efficiently and universally processed substrates as well [8]. By itself, this would suggest that the driving force for the evolution of the MUG protein family was the potential to counter mutagenesis by deamination and/or oxidation of C. It seems, however, that these glycosylases have learned to do more than that (Table 1). They act on DNA lesions as divergent as ethenoadducts (e.g. 3,N^4-ethenocytosine) [8,30,31], deaminated purines (e.g. hypoxanthine) [8] and thymine glycol [32], and even on normal DNA bases including T and 5-meC [33]. The biochemistry of the latter, however, has remained somewhat unclear. While the chicken MUG, normally referred to as 5-meC DNA glycosylase (5-MCDG), was shown to co-purify with an appreciable 5-meC processing activity from extracts of chicken embryos, this activity is comparably low, not to say marginal, when assayed with bacterially expressed chicken 5-MCDG or human TDG [8,33,34]. This may indicate a requirement for auxiliary factors that help target the glycosylase to the methylated C and, thus, facilitate the processing of this rather suboptimal substrate.

Interestingly though, the substrate spectra vary considerably between MUGs of different phylogenetic origin. Thymine and derivatives, for instance, are processed with a significant rate by the mammalian, chicken and drosophila enzymes but not by their bacterial and yeast counterparts [8,33], whereas 5-meC appears to be substrate for the vertebrate enzymes only (Table 1) [8,33,34]. Variable is also the degree of DNA

double-strand and mismatch-dependency. While the human TDG processes most of its substrate bases only in a mismatch with G, the fission yeast ortholog removes U from DNA also when it arises opposite A or in a single-stranded DNA context. It therefore appears that the MUG proteins have evolved with little selective pressure so that, in accordance with the specific needs of individual species, enzymes or enzyme complexes with rather distinct functionalities could develop.

If we try to infer a biological function for MUG proteins from their substrate preferences *in vitro*, a picture as follows emerges. Besides the elimination of mutagenic bases arising by hydrolytic deamination of C (e.g. G·U), MUG proteins seem to protect more generally against DNA base deamination (e.g. hypoxanthine) and/or oxidation (5-hydroxyuracil), as well as against base modifications by products of lipid peroxidation (e.g. 3,N^4-ethenocytosine). In organisms that methylate C in their DNA, MUG orthologs have acquired an ability to additionally deal with the corresponding deamination/oxidation products (e.g. G·T, 5-hydroxymethyluracil), and possibly, to contribute to the establishment and the stabilization of genomic DNA methylation patterns (e.g. 5-meC).

2.3.2. *Mechanism of substrate interaction*

Studies with non-cleavable substrate analogs provided first insight into how human TDG interacts with its substrate. In footprinting experiments, TDG protected an approximately 20 base pair stretch of DNA surrounding the mispaired U from DNaseI cleavage and made specific contacts to the N7 position of the G flanking the U at the 3′ side. The latter may explain a slight preference of TDG for G·T and G·U mispairs in a CpG sequence context [35]. No contacts with the "complementary" G opposite the lesion were detected in these studies although, by inference from the crystal structure of substrate bound Mug [21,22] and the strict requirement of this base for substrate recognition, TDG is expected to establish such an interaction. This would, however, involve the Watson-Crick surface of the widowed G, which cannot be seen easily by methylation interference [35]. The ability to recognize the substrate base through specific interactions with the nucleotide in complementary position seems important for a DNA glycosylase that has the potential to attack normal DNA bases. In the case of TDG, this assures that T is recognized as a substrate only when it is mispaired with G, i.e. originates from a deaminated 5-meC, but not when it is correctly base paired with A.

A mechanistic feature of the MUG proteins that, at least in part, relates to their ability to establish specific complementary base contacts, is their tight binding to the product of their reaction, the abasic site (AP-site). Human TDG, for instance, was shown to bind to an AP-site opposite G with an affinity that is higher than that to any of its preferred substrates (i.e. G·U or G·T mismatches) [26,36]. Indeed, purified full-length TDG is virtually unable to dissociate from the AP-site and is therefore fully product inhibited in base excision assays with G mispaired substrate. When the opposite base is not a G, however, slow dissociation of TDG is possible. Although the processing of such substrates is generally less efficient because of rate limitations at the level of substrate recognition, this corroborates the contribution of specific opposite base interactions to product inhibition.

Some degree of AP-site inhibition is common to most DNA glycosylases (e.g. [37–41]), but the extent observed with TDG is truly exceptional, even by comparison with other members of the MUG family. The *E. coli* Mug protein, for instance, turns over on a G·U substrate, albeit slowly [12], implicating that the opposite G interactions cannot fully account for the inability of TDG to dissociate from the AP-site. Indeed, deletion mutagenesis showed that truncation of the N-terminus converts TDG into Mug-like enzyme that processes G·U with a slow turnover, but fails to act on a G·T substrate with an appreciable efficiency [7,13]. Concurrently, deletion of the N-terminus abolishes a non-specific homoduplex DNA binding activity of TDG, which, compared to Mug and other DNA glycosylases, is quite appreciable [13,18,26].

This could all be explained by the presence of HMGA (HMG-I/Y)-box-like sequences in the N-terminal domain of TDG (Fig. 2). HMGA-boxes are frequently associated with nuclear proteins and have been shown to act as auxiliary structural motifs to provide non-specific DNA binding functionality [42,43]. Consistently, the Drosophila Thd1, which also processes G·T mispairs, contains related AT-hook motifs in its N- and C-terminal domains, whereas the fission yeast and the bacterial orthologs that fail on G·T substrates are devoid of such sequence motifs (Fig. 2). These correlations suggest that the non-conserved N-terminus of TDG has evolved to allow non-specific DNA binding, facilitating the processing of energetically suboptimal substrates such as G·T or G·5meC at the cost of free enzymatic turnover. Unfortunately, despite some considerable efforts, full-length TDG has eluded three-dimensional-structural analyses so that we do not know exactly how the N-terminal domain cooperates with the catalytic domain in DNA interaction. Analyses of partial tryptic digests of free and DNA bound TDG, however, have shed some light onto the problem [13]. Such experiments implicated that, upon encountering DNA, TDG undergoes a dramatic conformational change that depends on the presence of and involves the N-terminal domain (Primo Schär, Roland Steinacher, unpublished results). Consistently, crystallographic data revealed that *E. coli* Mug, which lacks a comparable N-terminus, undergoes only a minor structural rearrangement when it binds to DNA [22].

Thus, the experimental evidence available supports a mechanistic model in which the N-terminal domain of TDG forms a flexible "clamp" that holds the glycosylase onto the DNA. In this state, TDG may slide along the DNA in search of a G mismatched substrate. At substrate recognition and target base flipping, residues of the catalytic pocket establish the specific hydrogen-bonding interactions with the widowed G vis-à-vis. Following base release, these G contacts and the non-specific DNA contacts mediated by the N-terminus then cooperate to prevent free dissociation of TDG from the AP-site. Such a model would predict the need for a release factor that stimulates the displacement of TDG so that BER can proceed. A possible candidate would be the BER enzyme acting downstream of TDG, i.e. the (AP)-endonuclease (APE1). Indeed, experiments with purified human proteins showed that APE1 is able to stimulate the turnover of TDG on a G·T substrate [29]. However, the requirement of a high molar excess of APE1 and the fact that any other AP-site interacting protein tested has a similar stimulatory impact on TDG turnover implicates that

these are passive rather than active and specific effects (Primo Schär, Ulrike Hardeland, unpublished data), as proposed also for the APE1 mediated stimulation of other DNA glycosylases (e.g. [44]).

2.4. TDG and SUMO

A search for proteins interacting with human TDG led to the isolation of Small Ubiquitin like Modifiers (SUMOs) [18]. SUMOs are small polypeptides structurally related to ubiquitin that interact with and/or are attached to other proteins. Mechanistically, covalent SUMO modification (SUMOylation) is similar to ubiquitylation but requires its own set of E1, E2 and E3 conjugating enzymes [45]. Most of the targets of SUMOylation appear to be nuclear proteins with diverse functions [46]. Among them are proteins involved in DNA replication and repair as well as mediators of chromosome structure and dynamics, implicating a prominent role of SUMO modification in genome maintenance and stability [47]. The functional consequences of SUMO conjugation are still poorly understood, although an emerging theme is that it induces conformational rearrangements in target proteins in a way that alters their molecular interactions properties. Hence, depending on the target, SUMO-dependent changes in intra- or intermolecular interactions may affect protein localization, stability and enzymatic activity.

Not only was TDG found to interact with, but also to be modified by SUMO-1 and SUMO-3. SUMO conjugation involves lysine 330 (K330) located in a C-terminal SUMOylation consensus motif (VKEE) (Fig. 2), it is ATP-dependent and, when performed in cell extracts, stimulated by the presence of DNA. SUMO attachment to K330 affects structural and enzymatic properties of TDG. The modified glycosylase is not longer able to interact with free SUMO or SUMO-conjugated proteins (Primo Schär, Roland Steinacher, unpublished data), or to bind stably to AP-sites or any other DNA. Yet, it processes a G·U substrate with enhanced efficiency due to an induced enzymatic turnover but, at the same time, loses its ability to hydrolyze T from a G·T substrate. Obviously, SUMO modification alters the way TDG interacts with DNA so that AP-site dissociation and, hence, a slow turnover on an energetically favourable substrate (G·U) becomes possible, but processing of a suboptimal substrate that requires tight DNA interactions (G·T) is less efficient [13,18,26].

Thus, SUMO modification in the C-terminus converts TDG to an enzyme with Mug-like properties, as does the deletion of the N-terminus. As it turned out, this is not pure coincidence. A systematic assessment of the enzyme kinetic effects of SUMOylation on different domain truncation variants of TDG revealed that SUMO conjugation to full-length TDG and deletion of the N-terminus affect the same underlying mechanism of AP-site product inhibition. The results suggested that SUMOylation neutralizes the non-specific DNA binding activity of the N-terminal domain of TDG. The DNA-dependent conformational change of the N-terminus, i.e. the DNA clamp formation, was not seen when TDG was SUMOylated [13]. Hence, the same as SUMO modification prevented TDG from assuming a DNA binding conformation in these experiments, it may induce the opening of the "clamp" when it is conjugated to DNA bound TDG.

Other work addressed the non-covalent interaction of TDG with SUMO [18]. Mutational analyses identified a C-terminal SUMO-interaction motif (residues 304–316) that is distinct from the site of covalent attachment but appears to be essential for SUMO conjugation *in vivo* [48]. A putative second SUMO-interaction site is discernible in the N-terminal part of TDG, where the sequence "$I_{134}V_{135}I_{136}I_{137}$" matches an experimentally deciphered SUMO-interaction consensus [49]. An interaction of SUMO with these or nearby amino-acids might have considerable functional consequences as they are part of the conserved glycosylase active site and also overlap with a motif that mediates interaction with the estrogen receptor α [19]. Whether or not TDG uses these interfaces to interact with free SUMO or with SUMO-modified proteins in the context of its biological function is currently unknown.

The interactions of TDG with SUMO were ultimately visualized in a crystal structure. Although these analyses were done with a truncated TDG (residues 112–339), lacking the N-terminus and a large part of the C-terminal domain, the structure showed that upon attachment to K330, the TDG–SUMO conjugate assumes a highly organized configuration that builds on specific intermolecular contacts between the two protein moieties. These contacts engage residues of the C-terminal, non-covalent SUMO-interaction motif of TDG, which forms a β-strand that wraps around a β-strand of the SUMO to form an intermolecular antiparallel B-sheet structure. This conformation is further stabilized by extensive polar and hydrophobic contacts between residues of the intertwined B-strands [27]. Thus, consistent with our finding that SUMOylated TDG loses its ability to interact with non-conjugated SUMO, the structure shows that the C-terminal SUMO-interaction site of TDG is fully occupied when SUMO is attached and therefore no longer free for further non-covalent binding of SUMO.

As a consequence of SUMO conjugation, an α-helical peptide of TDG, containing the SUMO attachment site, forms a protrusion on the surface of the TDG–SUMO complex. From structural modeling, it was concluded that the protruding α-helix would interfere with DNA binding. Consistently, SUMO conjugation to the truncated TDG used in this study slightly reduced AP-site binding [27]. Yet, because these data were obtained with a N-terminal truncation of TDG, this cannot be the mechanism to account for the SUMOylation-induced AP-site dissociation observed with full-length TDG, where the tight interaction with the AP-site results from cooperative DNA binding by the N-terminal domain (unspecific DNA clamp) and the glycosylase active site (specific opposite G contacts) (see Section 2.3.2); a TDG lacking the non-specific DNA binding function dissociates from AP-sites regardless whether SUMO is attached or not [13]. Hence, while the structure provided valuable insight into the architecture of the TDG–SUMO conjugate, further studies with full-length TDG will be necessary for an ultimate resolution of the mechanism of SUMO-induced AP-site dissociation.

Taken together, biochemical and structural evidence supports the concept that SUMOs interact covalently and non-covalently with TDG and thereby induce changes in protein conformation that are required for its functionality, in particular, for its release from the AP-site. SUMOylation can thus be considered an integral regulatory component of TDG medi-

ated BER. If so, the apparent inconsistency that SUMOylation enhances G·U processing while abolishing G·T processing must be reconciled in a mechanistic model for TDG function in cells. This is possible if we appreciate first, that SUMOylation is a highly dynamic, i.e. reversible, protein modification, and second, that the majority of TDG protein in cells is SUMO-free, i.e. competent to recognize and process the full range of its substrates. Given this, a model can be postulated that invokes SUMOylation as a temporary TDG modification, affecting only DNA bound TDG and allowing the glycosylase to dissociate from the product AP-site following base release, so that BER can proceed. Dissociated TDG will then be readily de-modified by a SUMO-specific isopeptidase [50] to make it available for *de novo* recognition of G·U and G·T mispairs that might be generated [18].

3. Biological functions of TDG

3.1. *DNA repair*

Given the presence of MUG proteins in species across a broad phylogeny, one might conclude that their function is of fundamental biological importance and has therefore been conserved during evolution. Initially discovered in HeLa cells as an activity that catalyzes the excision of U and T mispaired with a G, the human TDG was proposed to act against mutation when Cs and 5-meCs deaminate. However, later studies, comparing the enzymatic properties of MUGs of different species, identified a number of well-processed DNA substrates that are not generated by deamination of C or 5-meC. These include a range of oxidized pyrimidines as well as some rather bulky ethenoadducts and damaged purine bases (see Section 2.3.1). Hence, it appears that MUG proteins have a more general function in the repair of DNA base damage than originally postulated. It is interesting though that different MUGs have different substrate spectra, suggesting that their function has not been strictly conserved in evolution (Table 1). With G·U being the common most efficiently processed substrate, the MUGs most likely originated from an ancestral uracil processing activity but then diverged to specialized enzymes to suit specific needs of the respective hosts—hence the evolution of non-conserved N- and C-terminal domains around a conserved catalytic core (see Section 2.3.1, Fig. 2). This is perhaps best exemplified by the G·T substrate. The potential to process this substrate appears to correlate with the degree of cytosine methylation in the genomes of the host organisms; it is highest for the mammalian TDGs where 5% of bases are methylated, poor for the Drosophila Thd1 where less than 1% of bases are methylated, and absent from fission yeast Thp1 where cytosine methylation is undetectable [51–53]. Thus, G·T processing might represent an extra feature of the mammalian enzymes that helps avoid genetic instability at sites of cytosine methylation. An assessment of the biological function of TDG must therefore be done from a perspective of individual substrates.

3.1.1. *Repair of G·T mismatches*

Considering a role of TDG in the restoration of methylated G·C pairs following 5-meC deamination, a second DNA glycosylase with G·T processing ability present in vertebrate cells must be taken into account. This enzyme, called MBD4/MED1, belongs to the family of methyl-CpG-binding domain (MBD) proteins and consists of an N-terminal MBD domain that is linked to a C-terminal DNA glycosylase domain. Although structurally unrelated, MBD4/MED1 has enzymatic properties very similar to those of TDG; it releases T and U from G·T and G·U mismatches, respectively, and processes a number of other substrates in common with TDG [54,55]. To what degree any of these glycosylases contributes to G·T processing in living cells is uncertain. Disruption of *MBD4/MED1* in mouse causes a small increase in C → T mutations at CpG sites, which would be consistent with a defect in the repair of deaminated 5-meC that cannot be fully compensated for by the presence of TDG [56]. On the other hand, we found that inactivation of TDG in mouse embryonic stem cells and fibroblasts reduces G·T processing in cell extracts below detection, indicating that TDG provides the predominant activity against the products of 5-mC deamination in these cells (Christophe Kunz, Yusuke Saito, Primo Schär, manuscript in preparation). In the light of this apparent discrepancy, the only firm conclusion that can be drawn at this point is that the G·T repair capacity in vertebrate cells is provided by at least two distinct DNA glycosylases that may act in a partially redundant manner.

In the light of such powerful defense, however, it seems surprising that methylated CpGs are mutation hotspots in the mammalian genome, and that C → T transitions at such sites are often seen in the DNA of human cancer cells [57]. One could argue that, given the rather inefficient processing of G·T substrates by TDG and MBD4/MED1, the number of substrates generated by deamination may exceed the repair capacity of the cell. G·T mispairs escaping repair would then give rise to C → T transition mutations upon DNA replication. Alternatively, the postreplicative mismatch repair system (MMR) might gain access to such G·T mispairs occasionally and, not being able to discriminate mutant from wild-type sequence in the context of non-replicating DNA, erroneously process the G strand and, hence, fix the mutation.

Another hypothesis worth considering is that the G·T glycosylases do not act globally in the genome but are targeted to specific sites. Indeed, TDG was reported to interact with transcription factors and to co-regulate gene expression at specific promoters (see Section 3.2). Similarly, MBD4/MED1 was shown to repress transcription of a reporter gene controlled by hypermethylated *p16INK4a* and *hMLH1* promoters [58]. These findings are consistent with G·T repair by TDG and MBD4/MED1 being restricted to specific areas of the genome and associated with defined physiological processes that involve the activation or inactivation of genes. Consequently, some areas of the genome would be more susceptible to mutagenesis through 5-meC deamination, whereas other would be safeguarded by TDG and/or MBD4/MED1.

Finally, while all these explanations seem logical from a mechanistic point of view, there is one more possibility to be considered, one that challenges the generally accepted dogma that deamination of 5-meC is the predominant cause of the decline of CpG dinucleotides in our genomes. There is growing evidence for methylated CpGs not only being hyper-susceptible to deamination but also to various forms of endogenous and environmental genotoxic stress that can

give rise to mutagenic lesions that are not substrate for the G·T glycosylases [59]. Preferential mutagenesis at methylated CpGs through pathways that do not or not exclusively involve 5-meC deamination could thus resolve the dilema why mutations occur at these sites despite the presence of the G·T repair enzymes TDG and MBD4/MED1.

3.1.2. *Repair of mismatched uracil*

Cytosines suffer hydrolytic deamination at a three- to four-fold lower rate but, on a genomic scale, still more frequently than 5-meC. In double-stranded DNA, this generates a U mispaired with a G. Unless repaired, the U will pair with A during DNA synthesis and, thus, give rise to a C → T transitions. Alternatively, U can also arise in DNA through dUMP incorporation during replication, in which case it is base paired with A and, hence, non-mutagenic. The biochemical evidence would predict TDG to act efficiently on G·U mispairs but hardly on an A·U base pair (Table 1). Considering TDG function in the cellular context, however, we are again confronted with a complex situation of redundancy. In a mammalian cell nucleus, TDG finds itself in good company with at least three additional UDGs; i.e. the very potent uracil-DNA glycosylase (UNG2), the MBD4/MED1 protein, and the "single-strand selective monofunctional uracil-DNA glycosylase (SMUG1)". Certainly, these enzymes have not evolved side by side just to back-up each other in U excision; there must be specific functional niches for all of them, which remain to be identified. Experimental evidence showing UNG2 interacting with replication proteins and localizing to replication foci suggests that this UDG is specialized for the rapid removal of dUMPs that happen to be misincorporated during DNA replication [60]. This is consistent with the phenotype of UNG deficient cells. Despite the presence of TDG, MBD4/MED1 and SMUG1, these accumulate significant amounts of dUMP in their DNA [61]. This, in keeping with the poor activity on the A·U substrate, would argue against a significant contribution of TDG to the elimination of U that gets misincorporated opposite A. A direct replication associated function of TDG is further excluded by the fact that the protein is actively degraded by the proteasome pathway at the G1/S boundary of the cell cycle and then remains undetectable during the entire S-phase (Ulrike Hardeland et al., manuscript submitted). Unlike mammalian TDG, however, Thp1 of *S. pombe* releases U from A·U base pairs [8] and is not cell cycle regulated. Consistently, a fission yeast *ung1* mutant does not accumulate significant amounts of dUMP in its genome, unless *Thp1* is genetically inactivated as well (Marc Bentele et al., manuscript submitted). Thus, in fission yeast, unlike in mammalian cells, the replicative uracil-DNA glycosylase Ung1 acts synergistically with the TDG ortholog to eliminate U that gets misincorporated during DNA replication.

Considering the repair of mutagenic uracil that arises from cytosine deamination, the fact that inactivation of *UNG* in mouse did not significantly alter the mutation frequency in the *Big Blue* assay argues for G·U correction being achieved by redundant activities. UNG2 and SMUG1 are good candidates as the C → T transition frequency at the *hprt* locus of Ung deficient mouse cells increases synergistically when *Smug1* is silenced by siRNA [62]. However, TDG is likely to act on deaminated cytosine as well since it is highly active on a G·U substrate and is expressed in most mammalian cell types. A role of TDG in G·U processing is also evident from genetic data obtained from the *S. pombe* model, where the concurrent inactivation of *thp1* and *ung1* increases the C → T transition rate synergistically (Marc Bentele at al., manuscript submitted). Surely, with four enzymes competing for uracil excision in mammalian cells, the situation is more complex and implies some form of functional separation. Separation could be temporal and/or spatial, as indicated by the cell cycle regulation of TDG and/or the interaction of TDG and MBD4/MED1 with transcription factors. Hence, whereas UNG and SMUG1 may have more global genome repair activity, the G·U processing of TDG (and MBD4) may be confined to certain areas of the genome and/or to specific physiological states of the cells.

One example of localized generation of G·U mispairs is the AID (activation-induced cytidine deaminase) catalyzed cytosine deamination that occurs when B-cells of the immune system induce somatic hypermutation (SHM) and class switch recombination (CSR) in immunoglobulin genes [63]. Mice with a defect in Ung have a reduced frequency of AID-induced transversion mutations, implicating a defect in the removal of U, as well as low IgG serum levels, implicating inefficient CSR [64]. Therefore, Ung2 was proposed to act on AID generated G·U mispairs in the process of SHM and CSR [63]. According to this model, AP-sites generated by Ung2 either give rise to point-mutations (SHM) when translesion polymerases synthesize across, or induce recombination when converted to single-strand breaks by the action of an AP-endonuclease. However, neither transversion mutagenesis nor CSR is fully defective in an Ung deficient background, indicating a contribution of other UDGs. In principle, these could be attributed to Smug1, Tdg and/or Mbd4. *Smug1* overexpression was indeed found to partially complement the SHM and CSR defect of $msh2^{-/-}ung2^{-/-}$ cells, but an involvement of this protein in antibody diversification is questionable as it is downregulated following B-cell activation [65]. By contrast, TDG is relatively abundant in B-cells and appears to be upregulated upon B-cell activation *in vitro* (Christophe Kunz and Primo Schär, unpublished results). It may therefore take part in SHM and CSR. From a mechanistic point of view, one might argue that TDG is the enzyme optimally suited to induce transversion mutations following AID catalyzed cytosine deamination. Unlike Ung2, which is a high turnover enzyme designed for rapid and complete repair of U, TDG is slow because it remains bound to the AP-site until it is actively induced to dissociate (see Sections 2.3 and 2.4). This possibility of delaying the processing of the AP-site seems desirable in a condition where a non-instructive lesion is generated for the purpose of "instructing" a mutation. So, despite the undisputable effect of an Ung defect on SMH and CSR, TDG must be considered a candidate glycosylase for the generation of AID-induced mutation and recombination.

3.1.3. *Repair of other forms of base damage*

TDG has a broad range of substrates that includes oxidation, alkylation and deamination products of C, 5-meC, T and A (see Section 2.3.1 [8,32]), implicating a rather general function in the repair of DNA base damage. Corroborating genetic evidence, however, is still missing with one notable exception, the repair of 3,N^4-ethenocytosine (εC) in *E. coli*. DNA ethenoadducts including εC arise by reaction of DNA with metabolic products of carcinogens such as vinyl chloride or

through membrane lipid peroxidation [66,67]. Levels up to 28 adducts per 10^7 bases have been detected in the DNA of various mammalian tissues [68,69]. εC has a mutagenic potential and produces most frequently εC → A transversions and εC → T transitions, and it was shown to be a reasonably good substrate for MUG proteins, including TDG [8,70]. In *E. coli*, Mug appears to be the only enzyme capable of excising εC from the DNA. Consistently, Mug deficient strains display hypersensitivity to εC [71], a phenotype that can be complemented by human TDG [72]. In mammalian cells, SMUG1 and MBD4 also process εC [73,74], but compared to TDG, their activity seems relatively weak. Hence, TDG might constitute the main repair activity protecting cells from εC-induced mutagenesis.

3.2. *Regulation of gene expression*

Already in 1992, Chevray and Nathans published a physical interaction of mouse TDG with the transcription factor c-Jun [14]. This was the first of a number of reports by several laboratories of physical and functional interactions of TDG with various transcription factors that, altogether argued for a role of the glycosylase in the regulation of gene expression. Pierre Chambon' s laboratory found TDG to interact with the nuclear receptors RAR (retinoic acid receptor) and RXR (retinoid X receptor) [15]. RAR and RXR form dimeric complexes that bind retinoic acid response elements (RAREs) to regulate gene activity in a ligand-dependent manner [75]. TDG interacts with RAR and RXR through its central catalytic domain (residues 122–346) in a ligand-independent manner (Fig. 2). This enhances the binding of the receptor complexes to RARE containing DNA substrate *in vitro* and potentiates transactivation of reporter genes in co-transfection experiments [15]. An active site mutant of TDG (N140A), however, failed to stimulate RAR/RXR-mediated transcription significantly above background, suggesting an involvement of the glycosylase function (Ulrike Hardeland, Primo Schär, unpublished data).

Estrogen receptor α (ERα), mediating estrogen responses, is another member of the nuclear receptor family that was shown to associate physically and functionally with TDG [19]. The interaction is established through the ligand-binding domain of the receptor (LBD/AF2) and involves residues 116–146 of human TDG (Fig. 2). This sequence contains a putative α-helical motif related to the LXXLL signature that is known to mediate interactions with nuclear receptors. Chromatin immunoprecipitation (ChIP) confirmed that TDG is indeed recruited to estrogen-responsive promoters in MCF7 cells exposed to estradiol (E2), presumably through its interaction with ERα. Notably, in this case, the glycosylase function of TDG is dispensable for transcriptional co-activation; in transient co-transfection experiments the catalytic mutant (N140A) co-activated an ERα responsive reporter gene as efficiently as the wild-type protein. This stands in contrast to RAR/RXR-dependent transcription, where TDG seems to have an active role. One possible explanation for this apparent discrepancy is that TDG has both, DNA glycosylase and scaffold functions in gene regulatory protein complexes, which may be differentially required in different biological contexts, e.g. at methylated versus non-methylated gene promoters (see Section 3.3).

Moreover, TDG was found to associate with SRC1, a p160 co-activator of ERα [20]. Here, the interaction involves regions containing tyrosine-repeat motifs located between residues 334–346 of human TDG and residues 989–1240 of human SRC1 (Fig. 2). ChIP experiments indicated that TDG and SRC1 are both recruited to estrogen-responsive gene promoters in E2 stimulated cells, presumably by ERα. The TDG-SRC1 complex appears to activate ERα-mediated gene expression in the absence of estrogen, suggesting cooperative functional interactions between the two co-activators.

Tini et al. demonstrated an interaction of TDG with yet another type of transcriptional co-activator, the CREB binding protein (CPB) and its paralog p300 [17]. Both play an important role in RNA polymerase II-mediated gene transcription, have histone acetyltansferase (HAT) activity and interact with various other transcription factors. Through acetylation of histone tails, CBP/p300 is thought to induce changes in chromatin structure that make promoter regions accessible for transcription factor binding [76]. Interactions with TDG occur through the HAT and CH3 domains of CBP/p300 and involve both, its N- and the C-terminal domains (Fig. 2). The resulting CBP–TDG complex binds DNA, processes G·T and G·U mismatches, is competent for histone acetylation *in vitro*, and enhances CBP-activated transcription of a reporter gene in transient co-transfection experiments. All this implicates a functional interaction. As in the case of the TDG–ERα complex, however, the glycosylase activity is dispensable for the stimulation of CBP-activated transcription. The interaction with CBP/p300 also leads to the acetylation of lysine residues in the hydrophobic N-terminal region of TDG. The function of this modification is not entirely clear; it does not seem to affect the enzymatic activity of the glycosylase (Ulrike Hardeland, Primo Schär, unpublished data) but reduces the stability of a ternary TDG–CBP–DNA complex and prevents a DNA mediated interaction with APE1. It was therefore proposed that TDG acetylation may have a regulatory role in the context of chromatin remodeling, gene regulation and DNA repair.

In one case, TDG was reported to act as a repressor of transcription. This is when it interacts with the thyroid transcription factor 1 (TTF1), a member of Nkx2 family of homeodomain proteins that is essential for the embryonal differentiation of the thyroid, lung and brain, as well as for thyroid and lung-specific gene expression in adult tissue [16]. Rat TDG was found to repress TTF1-activated transcription in thyroid and non-thyroid cells in transient co-transfection experiments. Whether or not this repressor function requires the glycosylase activity has not been resolved.

In addition to the interactions described above, TDG was seen to associate with several other transcription factors of the nuclear receptor family, including androgen receptor (AR), glucocorticoid receptor (GR), progesterone receptor (PR), peroxisome proliferators-activated receptor α (PPARα), thyroid hormone receptor a (TRa) and Vitamin D3 receptor (VDR) [19]. Although the biological significance of these interactions is unclear, these findings suggest that a cooperation of TDG with transcriptional activators is the rule rather than the exception.

Given all these interactions with transcription factors, the question arises whether and how the DNA glycosylase activity of TDG can be reconciled with a role in gene regulation in a plausible functional concept. This is possible if we con-

sider that, in vertebrate genomes, cytosine methylation in CpG dinucleotides is an important feature of gene regulation that can be corrupted by base deamination. TDG might be responsible for initiating correction of G·T mismatches that arise at methylated CpGs or, equally, of G·U mismatches at non-methylated CpGs. Thus, the specific recruitment to gene regulatory elements through the interaction with transcription factors would allow TDG to interrogate the integrity of such sequences, including the CpG islands found in promoters of many vertebrate genes. Transcription factors, on the other hand, could play a role in region-specific DNA repair by 'sensing' DNA damage in actively expressed areas of the genome through their ability to recruit DNA repair enzymes like TDG.

Methylation of CpG dinucleotides at gene promoters is often associated with gene silencing and is a key epigenetic regulator of gene expression. It may be that, in some circumstances, the reversal of such methylation is necessary for the re-activation of genes. The mechanism of such demethylation is unclear, but it has been suggested that the process could involve BER, starting with the excision of 5-meC by a 5-methylcytosine DNA glycosylase (5-MCDG) that may turn out to be TDG (see Section 3.3).

3.3. *TDG and CpG (de)methylation*

The possible association of TDG with the regulation of DNA cytosine methylation has raised a vivid (scientific) debate. Because this is an important issue, this chapter is dedicated to a critical assessment of the relevant experimental evidence.

DNA methylation in vertebrates occurs at the 5′ position of Cs immediately followed by G, affecting about 60–90% of all CpG dinucleotides depending on the species and/or the tissue examined. Being an epigenetic modification, cytosine methylation is essential for genomic imprinting and X-chromosome inactivation, but also affects gene expression and genomic stability. CpG methylation in vertebrate genomes is largely laid down during embryogenesis. There, dynamic changes in DNA methylation and histone modifications contribute critically to the establishment of cell-type- and tissue-specific gene expression patterns, which are a prerequisite for a coordinated development of the fetus [77,78]. Hence, a tight control of DNA methylation and demethylation processes during embryogenesis is of vital importance. Aberrant DNA methylation in adult tissue has been associated with aging and various human diseases including imprinting disorders and cancer [79].

While the enzymology of cytosine methylation by DNA methyltransferases is reasonably well established [80], mechanisms of active demethylation are largely obscure. Such a function might be required to enforce fidelity on the methylation process or to reactivate silenced genes. Human MBD2b was associated with an enzymatic activity that has the power to remove the methyl-group from 5-meC [81], but the reproducibility of this finding has been questioned since. Jean-Pierre Jost' s laboratory, on the other hand, found a quite different activity in nuclear extracts from developing chicken embryos and from differentiating mouse G8 myoblasts. This activity promoted "demethylation" of 5-meC in a hemimethylated DNA substrate by a process implicating excision repair [82,83]. The purification of this activity from extracts of chicken embryos led to the isolation of a 5-meC DNA glycosylase (5-MCDG) [84], which also processed T in G·T mismatches and eventually turned out to be the chicken ortholog of TDG [33]. However, when produced as a recombinant protein in *E. coli*, the 5-MCDG/TDG processed 5-meC with an extremely poor efficiency [33], suggesting that the glycosylase on its own cannot constitute a physiologically relevant demethylation activity.

Other studies by the same group showed that the 5-meC and G·T mismatch specific DNA glycosylase activities isolated from chicken embryos were sensitive to RNAse digestion, suggesting an involvement of an RNA component [85,86]. Indeed, heterogeneous and CpG rich RNA could be recovered from purified chicken 5-MCDG. Strikingly, when made complementary to the methylated strand of a hemimethylated DNA substrate, synthetic RNAs were able to restore 5-meC DNA glycosylase activity to a previously RNAse treated preparation [86]. It therefore appears that the RNA is an integral part of the glycosylase that may facilitate the targeting of the enzyme to sites where demethylation is needed. Still, some discrepancies in the published experimental evidence remain to be resolved before a firm conclusion about the function of this RNA in 5-meC demethylation can be drawn. Besides the RNA, the purified native 5-meC DNA glycosylase activity also contained a DEAD box protein related to the mammalian p68 RNA helicase [87]. Exactly how this RNA helicase contributes to demethylation is unclear. Yet, according to a model [87], the p68 RNA helicase might be responsible for the rearrangement of the secondary structure of the CpG-rich RNA components to make them suitable for targeting the DNA glycosylase to the site that is to be kept free of methylation.

To assess the relevance of 5-MCDG/TDG for active demethylation in living cells, the Jost group expressed human TDG in human embryonic kidney cells and examined the inducibility of a stably integrated reporter gene controlled by an ecdysone-retinoic acid responsive enhancer-promoter element. Overexpression of 5-MCDG/TDG resulted in the specific demethylation of CpG sites downstream of the hormone response elements in the promoter–enhancer region [34]. Since no genome-wide demethylation was observed, this would fit with the idea that transcription factors target the glycosylase to sites that need be demethylated or protected from *de novo* methylation. Another study, however, addressing 5-meCpG demethylation during mouse myoblast differentiation led the same authors to conclude that 5-MCDG/TDG contributes to global demethylation [88].

The work by the Jost laboratory showed that an excision repair process is associated with the removal of 5-meC from DNA. A concept whereby a DNA glycosylase initiates this process seems plausible, and the enzyme in question could indeed be 5-MCDG/TDG. Yet, can DNA excision repair indeed be regarded an appropriate strategy for global demethylation as it is seen during embryonic development? The answer is: "We don't know". Certain is, tough, that if global demethylation occurred by BER, a high degree of coordination would have to be involved to assure that mutation, rearrangement and fragmentation of the genome through excessive repair is avoided. By its ability to bind and protect AP-sites until it gets SUMOylated, TDG would provide for such coordination at a critical step of the excision repair process. However, considering its poor affinity for 5-meC·G base pairs, its cat-

alytic inefficiency on this substrate and its slow turnover, TDG mediated BER seems a highly unproductive and, thus, risky way of global demethylation, and is therefore unlikely a realistic scenario. More likely is that TDG contributes to site-specific stability of CpG methylation. Targeted by auxiliary proteins, i.e. transcription factors, to such sites, it might demethylate 5-meCpGs in gene promoters upon gene activation and/or protect unmethylated CpGs in promoters of active genes from accidental *de novo* methylation. This would be consistent with the co-purification of 5-meC glycosylase activity in 5-MCDG/TDG containing fractions of nuclear extract of developing embryos [33,89]. Whether or not 5-MCDG/TDG is a genuine 5-meC DNA glycosylase is still uncertain. The activity measured in partially purified preparations of 5-MCDG/TDG may result form the concerted action of the glycosylase with auxiliary factors [86,87]. The possible functions of such factors are equally unclear. They may help targeting the glycosylase to the methylated cytosine; they may facilitate the disruption of the hydrogen bonds of the 5-meC·G base pair so that the methylated C can be accommodated in the active site pocket of the glycosylase; or they may promote enzymatic deamination of 5-meC to T [90], which would then be a better substrate for TDG. All this would be consistent with purified recombinant 5-MCDG/TDG being very inefficient, if not impotent, in processing of 5-meC [8,33]. Thus, given the right conditions and physiological environment, TDG may act as a 5-MCDG and contribute to demethylation of 5-meC at specific sites in the genome, but this remains to be confirmed.

3.4. *TDG in embryonic development*

TDG cooperates with transcription factors that are essential for developmental processes (e.g. [91–93]). Embryonic development, on the other hand, is known to be associated with dynamic changes in CpG methylation [78], which, at least in certain areas of the genome, correlates with gene activation or inactivation [94]. Does this implicate TDG in developmental gene regulation? Tdg is readily detectable and highly active in mouse embryonic stem (ES) cells, and the levels even increase when these cells are induced to differentiate *in vitro* (Yusuke Saito, Primo Schär, in preparation). In the mouse embryo itself, *Tdg*-specific mRNA is seen to distribute ubiquitously and uniformly across the fetus from days 7.5 to 13.5 post-coitum. Later (at day 14.5), the mRNA is enriched in certain tissues including the developing nervous system, thymus, lung, liver, kidney and the intestine [95]. Whatever the role of TDG in embryogenesis may be, it is essential for proper development of the fetus, as homozygous *Tdg* null-embryos lose viability at midgestation (Primo Schär et al.; Tetsuya Ono et al., manuscripts in preparation). Given that Ogg1, Nth1, Mpg, Ung, and Mbd4 are all dispensable for embryogenesis [56,61,96–98], this is a rather unusual phenotype for a DNA glycosylase defect. It must therefore be concluded that TDG has a non-redundant essential developmental function that may relate mechanistically to BER but is distinct from the simple elimination of damaged DNA bases.

It is tempting to speculate that the TDG defect affects gene expression controlled by its interaction partners RAR/RXR, CBP/p300, c-Jun, and others that have an essential role in embryonic development. The resulting imbalance in gene expression may then disrupt the developmental program and cause the embryo to die. This function in co-regulation of gene transcription may relate to Tdg's ability to process 5-meC [33], as an imbalance in CpG methyltransferase activities in embryos [99] also causes dysregulation of gene expression and lethality [100–102]. It may be that Tdg, in conjunction with specific targeting factors, contributes to the establishment and/or the maintenance of proper CpG methylation patterns in certain regions of the genome and thereby assures accurate control of gene expression. At this point, however, this is little more than an interesting hypothesis that is worth being tested.

3.5. *TDG and cancer*

3.5.1. *TDG and carcinogenesis*

All considered, TDG could contribute to tumor suppression in a number of different ways. It may help maintain genomic stability through the repair of mutagenic DNA base damage (e.g. deaminated C or 5-meC); it may provide epigenetic stability through the excision of erroneously methylated Cs in gene regulatory sequences; and/or it may assure proper cell differentiation and, thus, control the number of stem cells and/or tumor progenitor cells in certain tissues by its ability to cooperate with nuclear receptors and other transcription factors that integrate differentiation signals.

CpG dinucleotides, most of which are methylated in vertebrate genomes, are indeed hotspots for mutations, and correlations between CpG mutagenesis and cancer development have long been established. Approximately 25% of all cancer associated mutations in the *p53* tumor suppressor gene are C → T transitions located at CpG sites; in colon and gastric cancer, this proportion rises to about 50% [57]. Although human TDG is expressed in most, if not all, tumor relevant tissues, its contribution to the avoidance of such mutations is speculative. The gene was mapped to a chromosomal region (12q24.1) that is frequently affected by loss of heterozygosity in gastric cancers, but inactivating mutations in *TDG* have not yet been identified in such tissue [103,104]. However, the number of tumors or cancer cell lines screened so far are too small to allow firm conclusions regarding the role of TDG defects in carcinogenesis.

Considering its role in controlling gene expression, one might also expect dominant, i.e. oncogenic, effects of TDG defects. In this regards, it is interesting that *Tdg* expression levels were found increased in mammary gland tumors that developed in HA-*ras* or *c-myc* transgenic mice, or in osteosarcomas that arose in *p53* heterozygous mice [95].

Be it as a tumor suppressor or as an oncogene, an involvement of TDG in carcinogenesis remains to be established.

3.5.2. *TDG and cancer therapy*

5-FU is an antimetabolite used in chemotherapy against a wide range of human cancers. Within cells 5-FU is converted to three main active metabolites; fluorodeoxyuridine monophosphate (FdUMP), fluorodeoxyuridine triphosphate (FdUTP), and fluorouridine triphosphate (FUTP). These interfere with DNA and RNA metabolism. FdUMP inhibits thymidylate synthase (TS), a consequence of which is that dUTP levels increase at the expense of dTTP. This imbalance in the nucleotide pool

gives rises to increased misincorporation of dUMP into the DNA (U·A base pairs). This, together with the direct incorporation of FdUTP (5-FU·A base pairs), is thought to account for the DNA directed cytotoxicity of 5-FU, although the underlying mechanism has remained obscure [105]. 5-FU is an excellent substrate for the MUGs, irrespective of whether it is paired with G or A (Table 1) [8]. These, however, do not seem to provide 5-FU resistance to cells as one might expect, they rather kill; a fission yeast *thp1*$^{-}$ mutant is significantly hyperresistant to 5-FU treatment (Marc Bentele et al., manuscript submitted). Similarly, inactivation of TDG in mouse embryonic fibroblasts causes hyperresistance to moderate doses of 5-FU. This hyperresistance correlates with a decrease in 5-FU-induced DNA single- and double-strand breaks and the loss of activation of an intra S-phase DNA damage checkpoint (Christophe Kunz, Primo Schär, manuscript in preparation). Thus, TDG contributes significantly to the DNA directed cytotoxicity of 5-FU and, considering that A·U is a very poor substrate for TDG, this is best explained by its ability to excise 5-FU opposite A. This would generate AP-site intermediates, which, through further processing, would give rise to the DNA strand-breaks that become visible upon 5-FU treatment of cells. Although this is a straightforward explanation for the TDG mediated cytotoxicity of 5-FU, other scenarios, such as an effect of its transcription associate function cannot be excluded. Whatever the mechanism, these findings suggest a non-redundant function of TDG in mediating cytotoxicity towards 5-FU. It will therefore be important to examine if the activity of TDG in human tumors correlates with their response to 5-FU-based chemotherapy.

The G·T processing function of TDG may also bear chemotherapeutic relevance. T is a substrate for human TDG also when it is mispaired with an O6-methylated guanine (O6-meG·T) [106,107]. O6-meG is a prominent mutagenic and cytotoxic DNA lesion that arises either spontaneously or when cells are exposed to Sn1-type methylating agents, such as *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG) that are widely used in cancer chemotherapy. During DNA replication O6-meG pairs with C or T, thus giving rise to O6-meG·T (mis)matches. If generated in excess, the processing of such mispairs by the postreplicative mismatch repair system (MMR) was shown to generate DNA strand-breaks, chromosomal instability and eventually trigger cell death [108]. Given the ability of TDG to act on the same substrate, it has the potential to compete with MMR and, thus, to affect the cytotoxicity of O6-menthylguanine-inducing drugs; an interesting hypothesis that merits investigation.

4. Concluding remarks

This has become a rather extended review, mainly because of the uncertainty about the biological function of TDG (Fig. 3). There are currently quite a few possibilities suggested by experimental evidence that merit careful consideration. Judged from its structure and biochemical properties, TDG is a DNA glycosylase involved in the repair of damaged DNA bases; judged from its interactions with other proteins, it is a co-regulator of gene expression; and judged from the phenotypes of fission yeast and mouse mutants, it may indeed be doing both, repairing DNA base damage and regulating gene expression. Any one of these activities appears to be required for normal mouse embryonic development, setting TDG functionally apart from all other DNA glycosylases that have been genetically studied in mouse. Future research will have to address whether TDG is indeed a multifunctional protein that repairs DNA base damage at one place and controls gene activity at another, or whether the regulation of gene expression involves an as yet unknown function of TDG mediated BER.

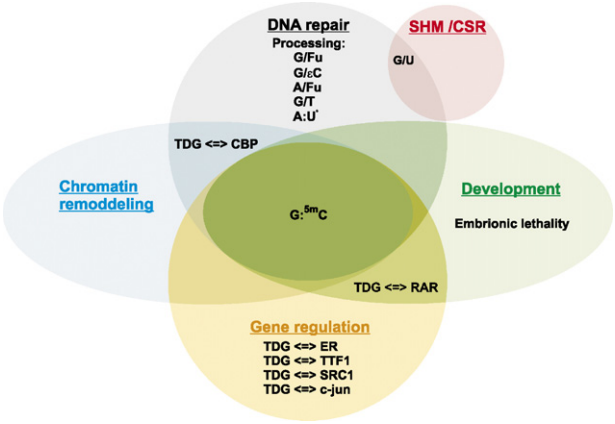


Fig. 3 – Biological processes with a possible involvement of TDG. The diagram illustrates implicated biological functions of TDG along with the relevant experimental observations. Biochemical properties suggest a role of TDG in the repair of DNA base damage (substrate are listed in descending order of cleavage efficiency). Interactions with transcription factors and the lethality of Tdg deficient mouse embryos suggest a function of TDG in gene regulation, cell differentiation and development. The putative 5-meC glycosylase activity of TDG may indicate that the gene regulatory function involves changes in DNA methylation and, thus, chromatin structure. TDG might also contribute to somatic hypermutation (SHM) and/or class switch recombination (CSR); it processes G·U with high efficiency and protects the AP-site until it is actively induced to dissociate by SUMOylation. <=>, interaction; *, A:U is substrate only for *S. pombe* Thp1p.

Acknowledgements

First of all, thank you Joe for a very interesting and still puzzling discovery. Many thanks also to Drs Tetsuya Ono and Adrian Bird for providing unpublished information on the mouse knock out phenotype of TDG. The generous support of the Swiss National Science Foundation and the Association for International Cancer Research (AICR), and the Krebsliga Beider Basel is gratefully acknowledged.

REFERENCES

[1] T. Lindahl, Instability and decay of the primary structure of DNA, Nature 362 (1993) 709–715.

[2] T.C. Brown, J. Jiricny, Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells, Cell 54 (1988) 705–711.
[3] K. Wiebauer, J. Jiricny, In vitro correction of G/T mispairs to G/C pairs in nuclear extracts from human cells, Nature 339 (1989) 234–236.
[4] P. Neddermann, J. Jiricny, The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells, J. Biol. Chem. 268 (1993) 21218–21224.
[5] P. Neddermann, P. Gallinari, T. Lettieri, D. Schmid, O. Truong, J.J. Hsuan, K. Wiebauer, J. Jiricny, Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase, J. Biol. Chem. 271 (1996) 12767–12774.
[6] P. Neddermann, J. Jiricny, Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 1642–1646.
[7] P. Gallinari, J. Jiricny, A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase, Nature 383 (1996) 735–738.
[8] U. Hardeland, M. Bentele, J. Jiricny, P. Schär, The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs, Nucl. Acids Res. 31 (2003) 2261–2271.
[9] L. Aravind, E.V. Koonin, The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates, Genome Biol. 1 (2000) 0007.
[10] Z. Gong, T. Morales-Ruiz, R.R. Ariza, T. Roldan-Arjona, L. David, J.K. Zhu, ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase, Cell 111 (2002) 803–814.
[11] Y. Choi, M. Gehring, L. Johnson, M. Hannon, J.J. Harada, R.B. Goldberg, S.E. Jacobsen, R.L. Fischer, DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis, Cell 110 (2002) 33–42.
[12] R.J. O'Neill, O.V. Vorob'eva, H. Shahbakhti, E. Zmuda, A.S. Bhagwat, G.S. Baldwin, Mismatch uracil glycosylase from *Escherichia coli*: a general mismatch or a specific DNA glycosylase? J. Biol. Chem. 278 (2003) 20526–20532.
[13] R. Steinacher, P. Schär, Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation, Curr. Biol. 15 (2005) 616–623.
[14] P.M. Chevray, D. Nathans, 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 (1992) 5789–5793.
[15] S. Um, M. Harbers, A. Benecke, B. Pierrat, R. Losson, P. Chambon, Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase, J. Biol. Chem. 273 (1998) 20728–20736.
[16] C. Missero, M.T. Pirro, S. Simeone, M. Pischetola, R. Di Lauro, The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription, J. Biol. Chem. 276 (2001) 33569–33575.
[17] M. Tini, A. Benecke, S.J. Um, J. Torchia, R.M. Evans, P. Chambon, Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription, Mol. Cell 9 (2002) 265–277.
[18] U. Hardeland, R. Steinacher, J. Jiricny, P. Schär, Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover, EMBO J. 21 (2002) 1456–1464.
[19] D. Chen, M.J. Lucey, F. Phoenix, J. Lopez-Garcia, S.M. Hart, R. Losson, L. Buluwela, R.C. Coombes, P. Chambon, P. Schär, S. Ali, T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha, J. Biol. Chem. (2003).
[20] M.J. Lucey, D. Chen, J. Lopez-Garcia, S.M. Hart, F. Phoenix, R. Al-Jehani, J.P. Alao, R. White, K.B. Kindle, R. Losson, P. Chambon, M.G. Parker, P. Schär, D.M. Heery, L. Buluwela, S. Ali, T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif, Nucl. Acids Res. 33 (2005) 6393–6404.
[21] T.E. Barrett, R. Savva, G. Panayotou, T. Barlow, T. Brown, J. Jiricny, L.H. Pearl, Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions, Cell 92 (1998) 117–129.
[22] T.E. Barrett, O.D. Scharer, R. Savva, T. Brown, J. Jiricny, G.L. Verdine, L.H. Pearl, Crystal structure of a thwarted mismatch glycosylase DNA repair complex, EMBO J. 18 (1999) 6599–6609.
[23] C.D. Mol, A.S. Arvai, R.J. Sanderson, G. Slupphaug, B. Kavli, H.E. Krokan, D.W. Mosbaugh, J.A. Tainer, Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA, Cell 82 (1995) 701–708.
[24] L.H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily, Mutat. Res. 460 (2000) 165–181.
[25] J.E. Wibley, T.R. Waters, K. Haushalter, G.L. Verdine, L.H. Pearl, Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1, Mol. Cell 11 (2003) 1647–1659.
[26] U. Hardeland, M. Bentele, J. Jiricny, P. Schär, Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis, J. Biol. Chem. 275 (2000) 33449–33456.
[27] D. Baba, N. Maita, J.G. Jee, Y. Uchimura, H. Saitoh, K. Sugasawa, F. Hanaoka, H. Tochio, H. Hiroaki, M. Shirakawa, Crystal structure of thymine DNA glycosylase conjugated to SUMO-1, Nature 435 (2005) 979–982.
[28] U. Hardeland, M. Bentele, T. Lettieri, R. Steinacher, J. Jiricny, P. Schär, Thymine DNA glycosylase, Prog. Nucl. Acid Res. Mol. Biol. 68 (2001) 235–253.
[29] T.R. Waters, P.F. Swann, Kinetics of the action of thymine DNA glycosylase, J. Biol. Chem. 273 (1998) 20007–20014.
[30] M. Saparbaev, S. Langouet, C.V. Privezentzev, F.P. Guengerich, H. Cai, R.H. Elder, J. Laval, 1N(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of *Escherichia coli* mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-*N*-glycosylase, J. Biol. Chem. 277 (2002) 26987–26993.
[31] E. Borys-Brzywczy, K.D. Arczewska, M. Saparbaev, U. Hardeland, P. Schär, J.T. Kusmierek, Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts, Acta Biochim. Pol. 52 (2005) 149–165.
[32] J.H. Yoon, S. Iwai, T.R. O'Connor, G.P. Pfeifer, Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair, Nucl. Acids Res. 31 (2003) 5399–5404.
[33] B. Zhu, Y. Zheng, D. Hess, H. Angliker, S. Schwarz, M. Siegmann, S. Thiry, J.P. Jost, 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 (2000) 5135–5139.
[34] B. Zhu, D. Benjamin, Y. Zheng, H. Angliker, S. Thiry, M. Siegmann, J.P. Jost, 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. U.S.A. 98 (2001) 5031–5036.

[35] O.D. Scharer, T. Kawate, P. Gallinari, J. Jiricny, G.L. Verdine, Investigation of the mechanisms of DNA binding of the human G/T glycosylase using designed inhibitors, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 4878–4883.
[36] T.R. Waters, P. Gallinari, J. Jiricny, P.F. Swann, Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1, J. Biol. Chem. 274 (1999) 67–74.
[37] J.W. Hill, T.K. Hazra, T. Izumi, S. Mitra, Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair, Nucl. Acids Res. 29 (2001) 430–438.
[38] H. Nilsen, K.A. Haushalter, P. Robins, D.E. Barnes, G.L. Verdine, T. Lindahl, Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase, EMBO J. 20 (2001) 4278–4286.
[39] F. Miao, M. Bouziane, T.R. O'Connor, Interaction of the recombinant human methylpurine-DNA glycosylase (MPG protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites, Nucl. Acids Res. 26 (1998) 4034–4041.
[40] F. Petronzelli, A. Riccio, G.D. Markham, S.H. Seeholzer, J. Stoerker, M. Genuardi, A.T. Yeung, Y. Matsumoto, A. Bellacosa, Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase, J. Biol. Chem. 275 (2000) 32422–32429.
[41] K. Krusong, E.P. Carpenter, S.R. Bellamy, R. Savva, G.S. Baldwin, A comparative study of uracil DNA glycosylases from human and herpes simplex virus type 1, J. Biol. Chem. (2005).
[42] L. Aravind, D. Landsman, AT-hook motifs identified in a wide variety of DNA-binding proteins, Nucl. Acids Res. 26 (1998) 4413–4421.
[43] R. Reeves, Molecular biology of HMGA proteins: hubs of nuclear function, Gene 277 (2001) 63–81.
[44] A.E. Vidal, I.D. Hickson, S. Boiteux, J.P. Radicella, Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: bypass of the AP lyase activity step, Nucl. Acids Res. 29 (2001) 1285–1292.
[45] R.T. Hay, SUMO: a history of modification, Mol. Cell 18 (2005) 1–12.
[46] R.J. Dohmen, SUMO protein modification, Biochim. Biophys. Acta 1695 (2004) 113–131.
[47] G. Gill, SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? Genes Dev. 18 (2004) 2046–2059.
[48] H. Takahashi, S. Hatakeyama, H. Saitoh, K.I. Nakayama, 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 (2005) 5611–5621.
[49] J. Song, L.K. Durrin, T.A. Wilkinson, T.G. Krontiris, Y. Chen, Identification of a SUMO-binding motif that recognizes SUMO-modified proteins, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 14373–14378.
[50] F. Melchior, M. Schergaut, A. Pichler, SUMO: ligases, isopeptidases and nuclear pores, Trends Biochem. Sci. 28 (2003) 612–618.
[51] M. Ehrlich, M.A. Gama-Sosa, L.H. Huang, R.M. Midgett, K.C. Kuo, R.A. McCune, C. Gehrke, Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells, Nucl. Acids Res. 10 (1982) 2709–2721.
[52] L.M. Field, F. Lyko, M. Mandrioli, G. Prantera, DNA methylation in insects, Insect Mol. Biol. 13 (2004) 109–115.
[53] F. Antequera, M. Tamame, J.R. Villanueva, T. Santos, DNA methylation in the fungi, J. Biol. Chem. 259 (1984) 8033–8036.
[54] B. Hendrich, U. Hardeland, H.H. Ng, J. Jiricny, A. Bird, The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites, Nature 401 (1999) 301–304.
[55] A. Bellacosa, Role of MED1 (MBD4) gene in DNA repair and human cancer, J. Cell Physiol. 187 (2001) 137–144.
[56] C.B. Millar, J. Guy, O.J. Sansom, J. Selfridge, E. MacDougall, B. Hendrich, P.D. Keightley, S.M. Bishop, A.R. Clarke, A. Bird, Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice, Science 297 (2002) 403–405.
[57] M.S. Greenblatt, W.P. Bennett, M. Hollstein, C.C. Harris, Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis, Cancer Res. 54 (1994) 4855–4878.
[58] E. Kondo, Z. Gu, A. Horii, S. Fukushige, The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes, Mol. Cell. Biol. 25 (2005) 4388–4396.
[59] G.P. Pfeifer, Mutagenesis at methylated CpG sequences, Curr. Top. Microbiol. Immunol. 301 (2006) 259–281.
[60] M. Otterlei, E. Warbrick, T.A. Nagelhus, T. Haug, G. Slupphaug, M. Akbari, P.A. Aas, K. Steinsbekk, O. Bakke, H.E. Krokan, Post-replicative base excision repair in replication foci, EMBO J. 18 (1999) 3834–3844.
[61] H. Nilsen, I. Rosewell, P. Robins, C. Skjelbred, S. Andersen, G. Slupphaug, G. Daly, H.E. Krokan, T. Lindahl, D.E. Barnes, Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication, Mol. Cell 5 (2000) 1059–1065.
[62] Q. An, P. Robins, T. Lindahl, D.E. Barnes, C → T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases, EMBO J. 24 (2005) 2205–2213.
[63] G.S. Lee, V.L. Brandt, D.B. Roth, B cell development leads off with a base hit: dU:dG mismatches in class switching and hypermutation, Mol. Cell 16 (2004) 505–508.
[64] C. Rada, G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, M.S. Neuberger, Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice, Curr. Biol. 12 (2002) 1748–1755.
[65] J.M. Di Noia, C. Rada, M.S. Neuberger, SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair, EMBO J. 25 (2006) 585–595.
[66] H.M. Bolt, Roles of etheno-DNA adducts in tumorigenicity of olefins, Crit. Rev. Toxicol. 18 (1988) 299–309.
[67] M. Saparbaev, J. Laval, Enzymology of the repair of etheno adducts in mammalian cells and in *Escherichia coli*, IARC Sci. Publ. 150 (1999) 249–261.
[68] L.J. Marnett, P.C. Burcham, Endogenous DNA adducts: potential and paradox, Chem. Res. Toxicol. 6 (1993) 771–785.
[69] J. Nair, A. Barbin, I. Velic, H. Bartsch, Etheno DNA-base adducts from endogenous reactive species, Mutat. Res. 424 (1999) 59–69.
[70] M. Saparbaev, J. Laval, 3,N^4-Ethenocytosine, a highly mutagenic adduct, is a primary substrate for *Escherichia coli* double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 8508–8513.
[71] E. Lutsenko, A.S. Bhagwat, The role of the *Escherichia coli* mug protein in the removal of uracil and 3,*N*(4)-ethenocytosine from DNA, J. Biol. Chem. 274 (1999) 31034–31038.
[72] J. Jurado, A. Maciejewska, J. Krwawicz, J. Laval, M.K. Saparbaev, Role of mismatch-specific uracil-DNA glycosylase in repair of 3,N^4-ethenocytosine in vivo, DNA Rep. 3 (2004) 1579–1590.
[73] B. Kavli, O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P.A. Aas, L. Hagen, H.E. Krokan, G. Slupphaug,

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 (2002) 39926–39936.
[74] F. Petronzelli, A. Riccio, G.D. Markham, S.H. Seeholzer, M. Genuardi, M. Karbowski, A.T. Yeung, Y. Matsumoto, A. Bellacosa, 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 (2000) 473–480.
[75] J. Bastien, C. Rochette-Egly, Nuclear retinoid receptors and the transcription of retinoid-target genes, Gene 328 (2004) 1–16.
[76] E. Kalkhoven, CBP and p300: HATs for different occasions, Biochem. Pharmacol. 68 (2004) 1145–1155.
[77] A. Bird, DNA methylation patterns and epigenetic memory, Genes Dev. 16 (2002) 6–21.
[78] W. Reik, W. Dean, J. Walter, Epigenetic reprogramming in mammalian development, Science 293 (2001) 1089–1093.
[79] K.D. Robertson, DNA methylation and human disease, Nat. Rev. Genet. 6 (2005) 597–610.
[80] A. Jeltsch, Molecular enzymology of mammalian DNA methyltransferases, Curr. Top. Microbiol. Immunol. 301 (2006) 203–225.
[81] S.K. Bhattacharya, S. Ramchandani, N. Cervoni, M. Szyf, A mammalian protein with specific demethylase activity for mCpG DNA, Nature 397 (1999) 579–583.
[82] J.P. Jost, 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 (1993) 4684–4688.
[83] J.P. Jost, Y.C. Jost, Transient DNA demethylation in differentiating mouse myoblasts correlates with higher activity of 5-methyldeoxycytidine excision repair, J. Biol. Chem. 269 (1994) 10040–10043.
[84] J.P. Jost, Y.C. Jost, Mechanism of active DNA demethylation during embryonic development and cellular differentiation in vertebrates, Gene 157 (1995) 265–266.
[85] M. Fremont, M. Siegmann, S. Gaulis, R. Matthies, D. Hess, J.P. Jost, Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA, Nucl. Acids Res. 25 (1997) 2375–2380.
[86] J.P. Jost, M. Fremont, M. Siegmann, J. Hofsteenge, The RNA moiety of chick embryo 5-methylcytosine-DNA glycosylase targets DNA demethylation, Nucl. Acids Res. 25 (1997) 4545–4550.
[87] J.P. Jost, S. Schwarz, D. Hess, H. Angliker, F.V. Fuller-Pace, H. Stahl, S. Thiry, M. Siegmann, 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, Nucl. Acids Res. 27 (1999) 3245–3252.
[88] J.P. Jost, E.J. Oakeley, B. Zhu, D. Benjamin, S. Thiry, M. Siegmann, Y.C. Jost, 5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation, Nucl. Acids Res. 29 (2001) 4452–4461.
[89] J.P. Jost, M. Siegmann, L. Sun, R. Leung, Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase, J. Biol. Chem. 270 (1995) 9734–9739.
[90] H.D. Morgan, W. Dean, H.A. Coker, W. Reik, S.K. Petersen-Mahrt, Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming, J. Biol. Chem. 279 (2004) 52353–52360.
[91] R.S. Johnson, B. van Lingen, V.E. Papaioannou, B.M. Spiegelman, A null mutation at the c-jun locus causes embryonic lethality and retarded cell growth in culture, Genes Dev. 7 (1993) 1309–1317.
[92] P. Kastner, J.M. Grondona, M. Mark, A. Gansmuller, M. LeMeur, D. Decimo, J.L. Vonesch, P. Dolle, P. Chambon, Genetic analysis of RXR alpha developmental function: convergence of RXR and RAR signaling pathways in heart and eye morphogenesis, Cell 78 (1994) 987–1003.
[93] Y. Oike, N. Takakura, A. Hata, T. Kaname, M. Akizuki, Y. Yamaguchi, H. Yasue, K. Araki, K. Yamamura, T. Suda, Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis, Blood 93 (1999) 2771–2779.
[94] A.F. Wilks, P.J. Cozens, I.W. Mattaj, J.P. Jost, Estrogen induces a demethylation at the 5′ end region of the chicken vitellogenin gene, Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 4252–4555.
[95] K. Niederreither, M. Harbers, P. Chambon, P. Dollé, Expression of T:G mismatch-specific thymidine-DNA glycosylase and DNA methyl transferase genes during development and tumorigenesis, Oncogene 17 (1998) 1577–1585.
[96] O. Minowa, T. Arai, M. Hirano, Y. Monden, S. Nakai, M. Fukuda, M. Itoh, H. Takano, Y. Hippou, H. Aburatani, K. Masumura, T. Nohmi, S. Nishimura, T. Noda, Mmh/Ogg1 gene inactivation results in accumulation of 8-hydroxyguanine in mice, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 4156–4161.
[97] M. Takao, S. Kanno, T. Shiromoto, R. Hasegawa, H. Ide, S. Ikeda, A.H. Sarker, S. Seki, J.Z. Xing, X.C. Le, M. Weinfeld, K. Kobayashi, J. Miyazaki, M. Muijtjens, J.H. Hoeijmakers, G. van der Horst, A. Yasui, Novel nuclear and mitochondrial glycosylases revealed by disruption of the mouse Nth1 gene encoding an endonuclease III homolog for repair of thymine glycols, EMBO J. 21 (2002) 3486–3493.
[98] B.P. Engelward, G. Weeda, M.D. Wyatt, J.L. Broekhof, J. de Wit, I. Donker, J.M. Allan, B. Gold, J.H. Hoeijmakers, L.D. Samson, Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 13087–13092.
[99] R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals, Nat. Genet. 33 (2003) 245–254.
[100] E. Li, T.H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality, Cell 69 (1992) 915–926.
[101] M. Okano, D.W. Bell, D.A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, Cell 99 (1999) 247–257.
[102] D. Biniszkiewicz, J. Gribnau, B. Ramsahoye, F. Gaudet, K. Eggan, D. Humpherys, M.A. Mastrangelo, Z. Jun, J. Walter, R. Jaenisch, Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality, Mol. Cell. Biol. 22 (2002) 2124–2135.
[103] L. Sard, S. Tornielli, P. Gallinari, F. Minoletti, J. Jiricny, T. Lettieri, M.A. Pierotti, G. Sozzi, P. Radice, Chromosomal localizations and molecular analysis of TDG gene-related sequences, Genomics 44 (1997) 222–226.
[104] C. Schmuttle, P.A. Jones, Involvement of DNA methylation in human carcinogenesis, Biol. Chem. 379 (1998) 377–388.
[105] D.B. Longley, D.P. Harkin, P.G. Johnston, 5-Fluorouracil: mechanisms of action and clinical strategies, Nat. Rev. Cancer 3 (2003) 330–338.
[106] U. Sibghat, P. Gallinari, Y.Z. Xu, M.F. Goodman, L.B. Bloom, J. Jiricny, R. Day III, Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA-glycosylase, Biochemistry 35 (1996) 12926–12932.

[107] S.U. Lari, F. Al-Khodairy, M.C. Paterson, Substrate specificity and sequence preference of G:T mismatch repair: incision at G:T, O6-methylguanine:T, and G:U mispairs in DNA by human cell extracts, Biochemistry 41 (2002) 9248–9255.

[108] B. Kaina, A. Ziouta, K. Ochs, T. Coquerelle, Chromosomal instability, reproductive cell death and apoptosis induced by O6-methylguanine in Mex−, Mex+ and methylation-tolerant mismatch repair compromised cells: facts and models, Mutat. Res. 381 (1997) 227–241.

Curriculum Vitae

Personals

Daniel Cortázar
Mittlere Strasse 110
CH-4056 Basel

Private: +41(0)61 321 43 29
Mobile: +41(0)78 600 84 45
E-mail: daniel.cortazar@unibas.ch

Date of Birth: 29. March 1980
Marital Status: Single
Nationality: German. Foreigner ID: B
Born in: Santafé de Bogotá/ Colombia

Working experience

06/2004- 05/2009	PhD Thesis in Biochemistry with Prof. Primo Schär **Department of Biomedicine**, **University of Basel.** Thesis Title: „Identification of a Possible Role of Thymine DNA Glycosylase in Epigenome Maintenance"
10/2003-05/2004	Assistant for Microscopy and Genetics **Institute of Anatomy, University of Basel**

Publications

Cortazar, D., et al., *The enigmatic thymine DNA glycosylase.* DNA Repair (Amst), 2007. **6**(4): p. 489-504.

Studies

03/2004	Diploma in Biology, Grade: 5.1/6.0 **University of Basel**
10/2002-10/2003	Diploma Thesis with Prof. Antonius Rolink **Institute of Molecular Immunology,** **Pharmacenter, University of Basel** Subject of research: „Effects of *In Vitro* Generated Dendritic Cells on the Immune System of the Mouse"
10/2001-10/2002	Blockcourses in Biology II (Molecular Biology, Biochemistry, Genetics, Cell Biology, Biophysics) **Biocenter, University of Basel**

10/1999-07/2001	Basic studies in biochemistry. Grade: 5.0/6.0 **University of Freiburg, Switzerland**
10/1998-07/1999	**Vorbereitungskurse für das Hochschulstudium in der Schweiz (VKHS)** Swiss Matura, Grade: 5.4/6.0 Freiburg, Switzerland
09/1992-06/1998	Graduation Diploma at the Swiss School „Colegio Helvetia“ Santafé de Bogotá, Colombia: „Bachiller Académico“

Languages

Spanish:	Mothers language
German:	2nd Mothers language Grosses Deutsches Sprachdiplom (GDS), Level $C2^+$ in the European Language Standard
English:	Very good knowledge
French:	Certificate of Practical French Studies, 2nd degree (CEFP2)
Italian:	Working knowledge

Informatics

Windows, MacOS X, MS-Office (Word, Excel, Power Point)

Hobbies

Tennis, Squash, Badminton, Mountainbike, Soccer

Daniel Cortázar
Basel
September 2009