Cell

Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair

Salvatore Cortellino,[1] Jinfei Xu,[1] Mara Sannai,[1] Robert Moore,[1] Elena Caretti,[1] Antonio Cigliano,[1] Madeleine Le Coz,[1] Karthik Devarajan,[2] Andy Wessels,[4] Dianne Soprano,[5] Lara K. Abramowitz,[6] Marisa S. Bartolomei,[6] Florian Rambow,[7,8] Maria Rosaria Bassi,[1] Tiziana Bruno,[9] Maurizio Fanciulli,[9] Catherine Renner,[3] Andres J. Klein-Szanto,[3] Yoshihiro Matsumoto,[10,12] Dominique Kobi,[11] Irwin Davidson,[11] Christophe Alberti,[7,8] Lionel Larue,[7,8] and Alfonso Bellacosa[1,*]

[1]Cancer Biology Program and Epigenetics and Progenitor Cells Keystone Program
[2]Department of Biostatistics
[3]Department of Pathology
Fox Chase Cancer Center, Philadelphia, PA 19111, USA
[4]Department of Cell Biology and Anatomy, Medical University of South Carolina, Charleston, SC 29425, USA
[5]Department of Biochemistry, Temple University School of Medicine, Philadelphia, PA 19140, USA
[6]Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
[7]Institut Curie, Centre de Recherche, Developmental Genetics of Melanocytes, 91405 Orsay, France
[8]CNRS UMR3347 INSERM U1021, 91405 Orsay, France
[9]Regina Elena Cancer Center, 00158 Rome, Italy
[10]Department of Biological Sciences, East Stroudsburg University, East Stroudsburg, PA 18301, USA
[11]Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 67404 Illkirch, France
[12]Present address: University of New Mexico Cancer Research Facility, Albuquerque, NM 87131, USA
*Correspondence: alfonso.bellacosa@fccc.edu
DOI 10.1016/j.cell.2011.06.020

SUMMARY

DNA methylation is a major epigenetic mechanism for gene silencing. Whereas methyltransferases mediate cytosine methylation, it is less clear how unmethylated regions in mammalian genomes are protected from de novo methylation and whether an active demethylating activity is involved. Here, we show that either knockout or catalytic inactivation of the DNA repair enzyme thymine DNA glycosylase (TDG) leads to embryonic lethality in mice. TDG is necessary for recruiting p300 to retinoic acid (RA)-regulated promoters, protection of CpG islands from hypermethylation, and active demethylation of tissue-specific developmentally and hormonally regulated promoters and enhancers. TDG interacts with the deaminase AID and the damage response protein GADD45a. These findings highlight a dual role for TDG in promoting proper epigenetic states during development and suggest a two-step mechanism for DNA demethylation in mammals, whereby 5-methylcytosine and 5-hydroxymethylcytosine are first deaminated by AID to thymine and 5-hydroxymethyluracil, respectively, followed by TDG-mediated thymine and 5-hydroxymethyluracil excision repair.

INTRODUCTION

Cytosine methylation—the formation of 5-methylcytosine (5mC) at CpG sites—is an important epigenetic modification used by mammals to mediate transcriptional regulation, including transcriptional repression, X chromosome inactivation, imprinting, and suppression of parasitic sequences (Bird, 1992; Kass et al., 1997; Siegfried and Cedar, 1997). The establishment and maintenance of the correct DNA methylation patterns at CpG sites is essential in mammals during development, gametogenesis, and differentiation of somatic tissues. Indeed, alterations in DNA methylation patterns, with the associated chromatin changes, have profound consequences, as demonstrated by embryonic lethality in the absence of DNA methylation (Li et al., 1992; Okano et al., 1999), developmental defects and accelerated aging in cloned mammals (Rideout et al., 2001), and characteristic epigenetic changes in cancer, such as global genome hypomethylation and tumor suppressor gene hypermethylation (Feinberg and Tycko, 2004; Jones and Laird, 1999).

Whereas DNA methylation is mediated by de novo DNA methyltransferases (DNMT3a and DNMT3b) that act on unmethylated DNA and maintenance DNA methyltransferases (DNMT1) that act on newly replicated, transiently hemimethylated DNA, the demethylating activities or processes that remove methylation marks in mammals are largely unknown. Indeed, it has been controversial as to whether demethylation is an active process in mammals (Ooi and Bestor, 2008) and which mechanisms are involved (Wu and Zhang, 2010).

Demethylation can occur passively due to replication in the absence of remethylation, with consequent dilution of this modification. However, there is evidence supporting the occurrence of active demethylation in mammals, including demethylation of the paternal genome shortly after fertilization (Mayer et al., 2000; Oswald et al., 2000), demethylation to erase and reset imprinting in primordial germ cells (Reik et al., 2001; Surani et al., 2007), and demethylation during somatic differentiation of the developing embryo to establish tissue-specific gene expression patterns (Kress et al., 2006; Niehrs, 2009) and during gene activation in adult kidney (Kim et al., 2009) and brain (Ma et al., 2009). In addition, it is generally thought that active transcription contributes to the maintenance of the unmethylated state of promoter-associated CpG-rich sequences known as CpG islands, but the molecular details of protection from hypermethylation and the potential involvement of an active demethylation process are unknown (Illingworth and Bird, 2009).

Accumulating evidence in nonmammalian model organisms points to the involvement of DNA repair mechanisms in active demethylation (Gehring et al., 2009; Niehrs, 2009). In *Arabidopsis*, the base excision repair (BER) proteins Demeter and ROS1 affect demethylation by directly removing 5mC through their glycosylase activities (Gehring et al., 2006; Morales-Ruiz et al., 2006). In *Xenopus*, demethylation has been reported to be initiated by the genome stability protein Gadd45a (growth arrest and DNA damage-inducible protein 45 α) in a process dependent on the nucleotide excision repair protein XPG (Barreto et al., 2007); however, the role of mammalian GADD45 in demethylation (Barreto et al., 2007; Schmitz et al., 2009) has been challenged (Jin et al., 2008). In zebrafish embryos, rapid demethylation of exogenous and genomic DNA occurs in two coupled steps: enzymatic 5mC deamination to thymine by activation-induced deaminase (AID) or apolipoprotein B RNA-editing catalytic component 2b and 2a (Apobec2b, 2a), followed by removal of the mismatched thymine by the zebrafish thymine glycosylase MBD4, with Gadd45 promoting the reaction (Rai et al., 2008). Recently, 5-hydroxymethylcytosine (5hmC), an oxidative product of 5mC generated by the Tet hydroxylases (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009), has been proposed as a demethylation intermediate (Globisch et al., 2010; Wu and Zhang, 2010). During gene activation in the adult mouse brain, demethylation by TET1-mediated hydroxylation of 5meC to 5hmC was promoted by AID/Apobec deaminases in a process that generates 5-hydroxymethyluracil (5hmU) and ultimately requires BER, although the specific glycosylases involved were not identified (Guo et al., 2011).

Numerous in vitro studies have documented a potential role of the BER enzyme TDG (thymine DNA glycosylase) in transcriptional regulation and demethylation. Indeed, TDG interacts with several transcription factors, including retinoic acid receptor (RAR), retinoid X receptor (RXR) (Um et al., 1998), estrogen receptor α (ERα) (Chen et al., 2003), thyroid transcription factor 1 (TTF1) (Missero et al., 2001), and histone acetyl-transferases p300 and CBP (Tini et al., 2002). It has been proposed that TDG may be responsible for demethylation either through a direct 5mC glycosylase activity (Zhu et al., 2000) or indirectly by acting on G:T mismatches originated by a controlled deaminase activity of DNMT3a and DNMT3b (Métivier et al., 2008). Very recently, TDG was shown to be involved in maintaining active and bivalent chromatin marks in mouse embryo fibroblasts and ES cells undergoing neuronal differentiation, respectively, but the mechanism for such epigenetic effects and the requirement of its catalytic activity were not clarified (Cortázar et al., 2011). To investigate the functional role of TDG in epigenetic regulation, DNA demethylation, and mammalian development, we generated mice with targeted inactivation of the *Tdg* gene. *Tdg* null embryos die in midgestation and exhibit a complex developmental phenotype that appears to derive from the failure to establish and maintain proper DNA methylation patterns at promoters and enhancers. A knockin mutation that inactivates the glycosylase function of TDG is also embryonically lethal, and TDG is found in a complex with AID and GADD45a. These findings suggest a two-step catalytic mechanism for DNA demethylation that is essential for mammalian development.

RESULTS

Generation of *Tdg* Null Mice

The *Tdg* gene was deleted in mice by homologous recombination, using a targeting construct that removed exons 3–7, corresponding to most of the catalytic domain (Figure S1 available online). The resulting Tdg^{-} allele does not produce detectable protein by western blotting (Figure 1A), suggesting that it is null. TDG is dispensable for efficient uracil removal from G:U mismatches (Figure S2). However, removal of the mispaired thymine from G:T mismatch-containing oligonucleotides is virtually abrogated in $Tdg^{-/-}$ homozygous mouse embryo fibroblast (MEF, *vide infra*) nuclear extracts (Figure 1B), thus confirming that Tdg^{-} is a null allele and suggesting that TDG is the predominant G:T mismatch repair activity in MEFs.

Lethality and Complex Developmental Phenotype of *Tdg* Null Embryos

Whereas heterozygous $Tdg^{+/-}$ mice are viable and fertile and show no obvious phenotype, homozygosity for the null *Tdg* allele leads to embryonic lethality. In fact, when heterozygous $Tdg^{+/-}$ mice were interbred, no live birth $Tdg^{-/-}$ homozygotes were derived. Specifically, the numbers of wild-type, heterozygous, and homozygous mutant pups obtained were 11, 29, and 0, respectively, which is significantly different from the expected Mendelian ratios ($p < 0.0008$ by χ^2). To examine in detail the embryonic lethality, timed matings were set up between $Tdg^{+/-}$ heterozygotes, and pregnant mice were sacrificed at different gestation times, ranging from embryonic day (E) 10.5 to E14.5. This analysis revealed an arrest of development associated with *Tdg* nullizygosity at E11.5; at later gestation times, homozygous mutant embryos are beginning to be resorbed (E12.5–13.5) or are completely resorbed and never detected (E14.5) (Table S1). At E11.5, developmentally arrested *Tdg* null embryos manifest a complex phenotype characterized macroscopically by abdominal (liver) hemorrhage, pericardial edema/hemorrhage, hypoplastic branchial arches, delayed limb development, prominent telencephalic vesicles, and diffuse hemorrhagic lesions (Figures 1C, 1D, 1G, and 1H).

Microscopically, *Tdg* null embryos exhibit specific patterning defects of the developing heart, with the most significant

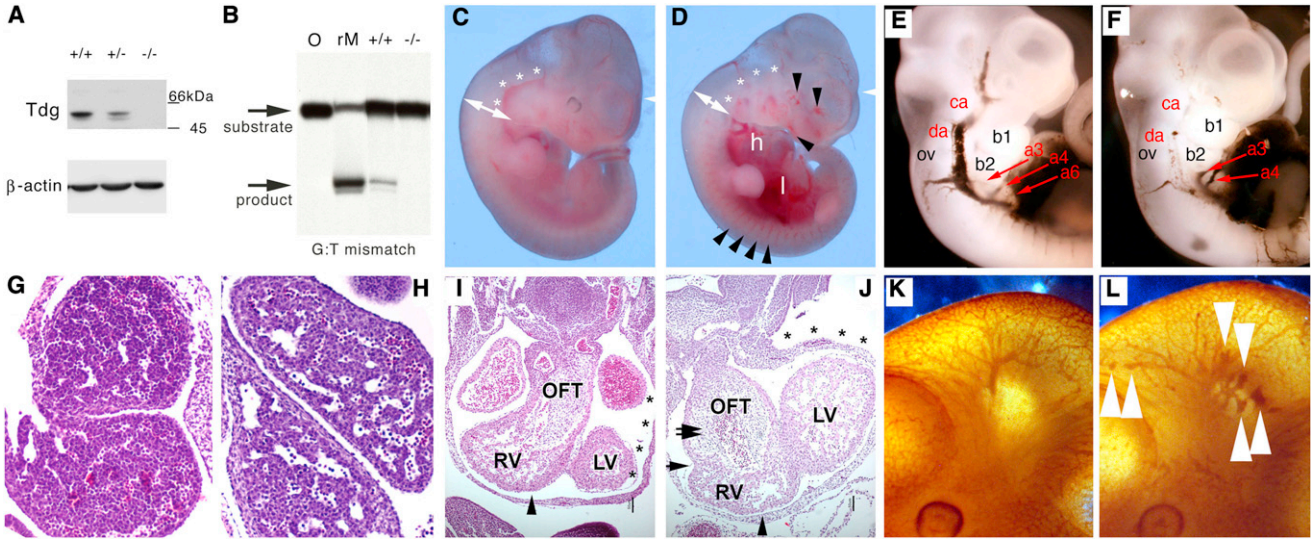

Figure 1. Developmental Defects in *Tdg* Null Embryos

(A) The expression of TDG was monitored by western blot. No expression was detected in MEFs homozygous for the Tdg^{-} allele, whereas heterozygous MEFs showed reduced expression compared to wild-type cells.

(B) TDG is the foremost G:T mismatch repair activity at CpG dinucleotides in MEFs. Repair of a double-stranded oligonucleotide containing a G:T mismatch by nuclear extracts of MEFs with different genotypes, in comparison to a substrate neither exposed to lysate nor enzyme (O). Reaction with recombinant MBD4/MED1 (rM) was used as a size marker for cleavage at the mismatched thymine.

(C and D) Gross phenotype of wild-type and $Tdg^{-/-}$ littermate embryos at embryonic day E11.5. Double arrows show constriction in the cervical region of *Tdg* null embryos (D) compared to wild-type embryos (C); white asterisks mark the carotid artery that is stenotic in *Tdg* null embryos; the enlarged heart with pericardial effusion (h) and hemorrhagic liver (l) are apparent; arrowheads point to hemorrhagic lesions in the cranium and enlarged and irregular segmental arteries.

(E and F) Cardiac perfusion with India ink in wild-type and *Tdg* mutant embryos at E11. In *Tdg* null embryo (F), circulatory insufficiency is demonstrated by reduced perfusion of the dorsal aorta (da) and carotid artery (ca), whereas the third (a3) and fourth (a4) branchial arch arteries are enlarged in comparison to wild-type embryos (E). The first (b1) and second (b2) branchial arches, as well as the otic vesicle (ov), are indicated.

(G and H) Transverse sections of the liver at E11. Compared to wild-type (G), the mutant liver has enlarged hepatic sinusoids, the likely proximal causes of abdominal hemorrhage.

(I and J) Transverse sections of the heart at E11.5 show patterning defects in mutant embryos. The conal part of the OFT is severely hyperplastic in mutant embryos (twin arrows in J) when compared to the heterozygous specimen (I), generating an atypical indentation between the right ventricle (RV) and OFT (arrow head in J); the characteristic "dog leg bend" of the OFT, which is responsible for correctly positioning the OFT over the midline of left ventricle (LV) and RV, is not observed in mutant hearts. Instead, the OFT is situated right above the RV. As a result, the left part of the body wall is pushed out by the LV (asterisks).

(K and L) Immunostaining of the vascular labyrinth with a PECAM/CD31 antibody in wild-type (K) and *Tdg* null (L) embryos at E11 reveal a generalized disorganization of the vascular network in the latter. Arrowheads point to irregular branches of the internal carotid with varicosities, bulges, and ectasias.

See also Figure S1, Figure S2, and Figure S3.

abnormalities seen in the outflow tract (OFT) (Figure 1I and 1J and Figures S3A–S3D); vasculogenesis defects of dorsal aortae, carotid arteries, and branchial arteries (Figures 1E and 1F); and generalized defect of angiogenesis, particularly evident as altered branching of the internal carotid (Figure 1F) and the coronaries (Figures S3E and S3F).

TDG has two proposed roles in mutational avoidance (DNA repair) and transcriptional regulation (Cortázar et al., 2007). The embryo phenotype was fully penetrant and reproducible, which is inconsistent with an antimutagenic DNA repair defect that would be expected to yield a variable, heterogeneous phenotype caused by stochastic secondary mutations at different target genes. We therefore focused on the role of TDG in transcription as a possible mechanistic explanation of the phenotype.

Remarkably, many features of the phenotype of *Tdg* null embryos (altered vasculogenesis and angiogenesis, hemorrhagic lesions, heart abnormalities with thinning of the myocardium, and pericardial effusion) have been previously described for either $Cbp^{-/-}$ or $p300^{-/-}$ embryos (Tanaka et al., 2000; Yao et al., 1998). Similarly, some other specific phenotypic features in *Tdg* null embryos (OFT septation defects, hypoplastic myocardium, abnormal great arteries derived from branchial arches, delayed limbs) resemble those of embryos deficient in various *Rar* and *Rxr* genes (Mark et al., 2006) or hypomorphic for retinaldehyde dehydrogenase, the enzyme involved in RA biosynthesis (Vermot et al., 2003). We infer that the lethality phenotype is likely related to the inactivation of a developmentally relevant, transcription-related function of TDG. In order to investigate this issue directly, we established MEF lines from *Tdg* null embryos.

Attenuated RA-Dependent Transcription and Altered p300 Recruitment in *Tdg* Null MEFs

We hypothesized that, given the phenotypic features described above, it is possible that RAR/RXR and p300 activity might be reduced in the absence of TDG. Indeed, assay of p300 activity

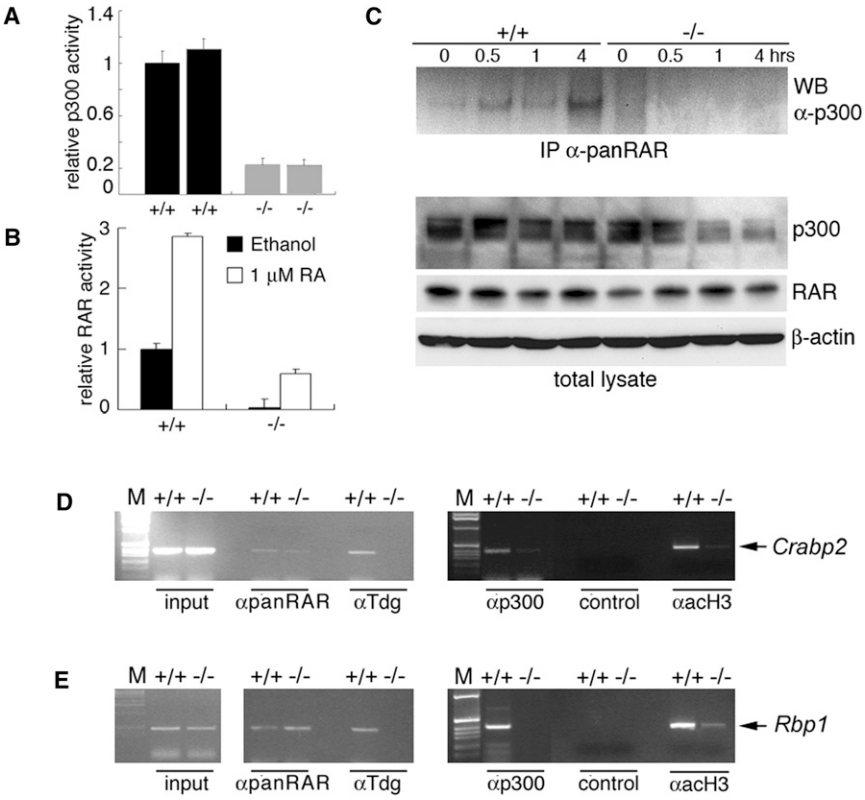

Figure 2. Involvement of TDG in Transcription and Composition of RAR-p300 Complexes

(A) Reduced p300-induced transcriptional activation in $Tdg^{-/-}$ MEFs. Luciferase activity of a *Gal4* operator luciferase reporter cotransfected with a Gal4 DNA-binding domain-p300 fusion construct and normalized to transfection efficiency using β-galactosidase expression. Because the transfected plasmids are unmethylated, this assay reflects only the coactivator role of TDG.

(B) Reduced retinoic-dependent RAR/RXR transcriptional activity in $Tdg^{-/-}$ MEFs. CAT activity of a *RARE*-containing reporter was normalized to transfection efficiency using β-galactosidase expression.

(C) Co-IP with an antibody capable of recognizing all of the RARs shows lack of association between RAR and p300 in $Tdg^{-/-}$ MEFs. Wild-type and mutant MEFs were stimulated with 1 μM RA for the indicated time. Approximately equal levels of p300 and RAR are present in wild-type and mutant cells. Detection of β-actin is shown as a loading control.

(D and E) Chromatin immunoprecipitation shows that TDG binds directly to the promoter of two differentially expressed RAR-RXR target genes, *Crabp2* (D) and *Rbp1* (E), and is required for p300 recruitment and histone H3 acetylation. Approximately equal amounts of input chromatin were used for immunoprecipitation. As negative control, immunoprecipitation with nonspecific immunoglobulins was performed.

Data are presented as mean ± standard error of the mean (SEM). See also Figure S4.

with reporter constructs indicated that transcriptional coactivation by this acetyltransferase is significantly reduced in two independent $Tdg^{-/-}$ MEF lines in comparison to wild-type MEF lines (Figure 2A). RA-dependent RAR/RXR transcriptional activity is also attenuated in *Tdg* null MEFs (Figure 2B).

To clarify the role of TDG in transcription, we compared the transcriptome of wild-type and *Tdg* null MEFs. In keeping with a role of TDG in transcriptional activation, of the 108 differentially expressed genes from 120 probe sets, approximately three-fourths of the genes were downregulated in its absence (Table S2). The differentially expressed genes were analyzed with gene ontology and pathway analysis applications. Remarkably, the pathway/network with the highest score was centered around RA (Figure S4) and comprised retinol biosynthesis and RA-dependent target genes downregulated in *Tdg* null MEFs, including those encoding cellular retinoic acid-binding protein 2 (*Crabp2*, 15.3-fold), retinol-binding protein 1 (*Rbp1*, 9.3-fold), *Igfbp6* (16.3-fold), embryonal fyn-associated substrate (*Efs*, 6-fold), and *Rai14* (1.9-fold). These observations indicate that TDG is a positive regulator of transcription, particularly p300- and RAR/RXR-dependent transcription.

To define the role of TDG in RA-dependent transcription, we examined the composition of RAR-containing complexes by coimmunoprecipitation (co-IP) and found that p300 is in a complex with RAR/RXR in wild-type MEFs, but not in *Tdg* null MEFs, despite the presence of approximately equal levels of p300 and RAR in total lysates (Figure 2C). In addition, RARs occupy retinoic acid response elements (*RARE*) on the *Crabp2* and *Rbp1* promoters in both wild-type and *Tdg* null cells, but in the absence of TDG, there is little recruitment of p300 and a reduction in the presence of its product, acetylated histone H3 (Figures 2D and 2E). These findings are consistent with the differential expression of *Crabp2* and *Rbp1* and indicate that TDG has an obligatory direct role in their proper transcriptional regulation.

Altered DNA Methylation Patterns in *Tdg* Null Cells and Tissues

Given the possible role of TDG in demethylation (Métivier et al., 2008; Zhu et al., 2000), we examined the DNA methylation patterns of promoters of select genes that were differentially expressed between wild-type and *Tdg* null MEFs using sodium bisulfite/DNA sequencing. In *Tdg* null MEFs, the downregulated genes contain a CpG island within 2 kb of sequence upstream of the transcriptional start site and are hypermethylated, including *Efs*, *Crabp2*, *Hoxa5*, and *H19* (Figures 3A–3D). Because these CpG islands and the maternal allele of the imprinted *H19* gene are unmethylated in zygotes, ES cells, and the soma (Mohn et al., 2008; Reese and Bartolomei, 2006; data not shown), this observation suggests that, in the absence of TDG, sequences that are normally kept unmethylated succumb to hypermethylation, likely as a consequence of unscheduled de novo methylation.

To rule out the possibility that hypermethylation was caused by the in vitro culture stress of MEFs (Pantoja et al., 2005), we

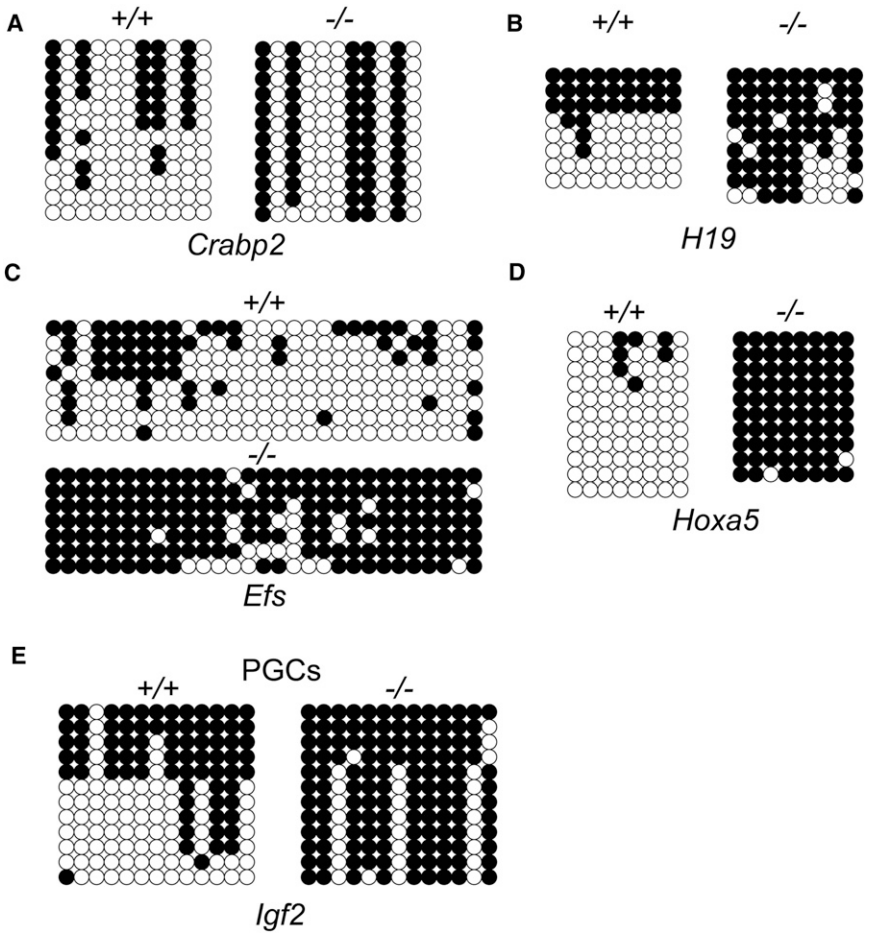

Figure 3. Hypermethylation of CpG Islands in the Absence of TDG

(A–D) DNA methylation analysis by sodium bisulfite modification and sequencing of cloned PCR products in wild-type and *Tdg* null cells. Open and closed circles represent unmethylated and methylated CpGs, respectively. *Crabp2* (A), *Efs* (C), and *Hoxa5* (D) promoters are unmethylated at various degrees in wild-type cells, whereas their CpG islands are hypermethylated in *Tdg* null MEFs. For the *H19* promoter (B), the first three clones in wild-type MEFs are likely derivatives of the inactive paternal allele, and the remaining ones are probably of maternal origin (*H19* is maternally expressed), whereas in *Tdg* null MEFs, both alleles are hypermethylated.

(E) Analysis of the 3′ half of the *Igf2* DMR2. The first five clones in wild-type PGCs are likely derived from the paternal (methylated) allele, and the remaining ones are probably of maternal origin, whereas in *Tdg* null PGCs, all of the alleles are hypermethylated.

See also Figure S4.

analyzed the methylation pattern of the imprinted gene *Igf2* in wild-type and *Tdg* mutant primordial germ cells (PGCs) isolated from E11 embryos (prior to the onset of lethality). Whereas wild-type E11 PGCs show the typical methylation profile of maternal unmethylated and paternal methylated alleles at the *Igf2* differentially methylated region 2 (DMR2), all of the alleles sequenced in *Tdg* mutant PGCs, presumably including those of maternal origin, are methylated (Figure 3E). Although it is currently unclear whether TDG has any specific role in the establishment or maintenance of imprinting, these data confirm a TDG-dependent protection from hypermethylation in early development.

The observed protective function begs the question of whether TDG might also have a role in DNA demethylation. During development, highly conserved noncoding elements and enhancers undergo demethylation in a process linked to tissue-specific gene expression and differentiation (Kress et al., 2006; Niehrs, 2009). One such example is the albumin gene (*Alb1*) enhancer, whose five CpG dinucleotides are progressively demethylated during liver development and are associated with *Alb1* mRNA transcription (Xu et al., 2007). Analysis of these sites revealed that they remain methylated in *Tdg* null liver at midgestation, in a configuration similar to that of a nonalbumin-producing organ, e.g., brain (Figures 4A and 4B). This correlated with inefficient *Alb1* mRNA transcription in the *Tdg* null liver (Figure 4C).

The glucocorticoid-responsive unit (GRU) of the tyrosine aminotransferase (*Tat*) gene enhancer undergoes demethylation at midgestation in the developing rat liver in a process stimulated by the prenatal peak of glucocorticoids (Thomassin et al., 2001). Demethylation of this enhancer is associated with single-strand nicks 3′ to the 5mC, leading to the suggestion that a demethylating activity initiates base or nucleotide excision repair at these sites (Kress et al., 2006). We found that demethylation of five CpG sites at the GRU of the murine *Tat* enhancer begins at midgestation and is dependent on TDG (Figures 4D and 4E).

Taken together, these observations indicate that TDG is required for the establishment of proper DNA methylation patterns that are conducive to transcription of developmentally and hormonally regulated genes and of tissue-specific genes, both by guarding from CpG island hypermethylation and promoting selective demethylation events.

Involvement of TDG in Active DNA Demethylation

Whereas protection against CpG island hypermethylation might be, in principle, a reflection of the coactivator function of TDG, the involvement of this enzyme in DNA demethylation suggests an active catalytic role. In order to determine directly whether TDG is involved in active DNA demethylation, we studied the transcriptional reactivation of a heterologous in vitro-methylated *Oct4* pluripotency gene in embryonic carcinoma P19 cells or in the same cells expressing either an shRNA (C8) targeting TDG or a control shRNA (C7) (Figure 4F). We used an *Oct4* promoter::*EGFP* reporter assay in which reactivation of EGFP expression is

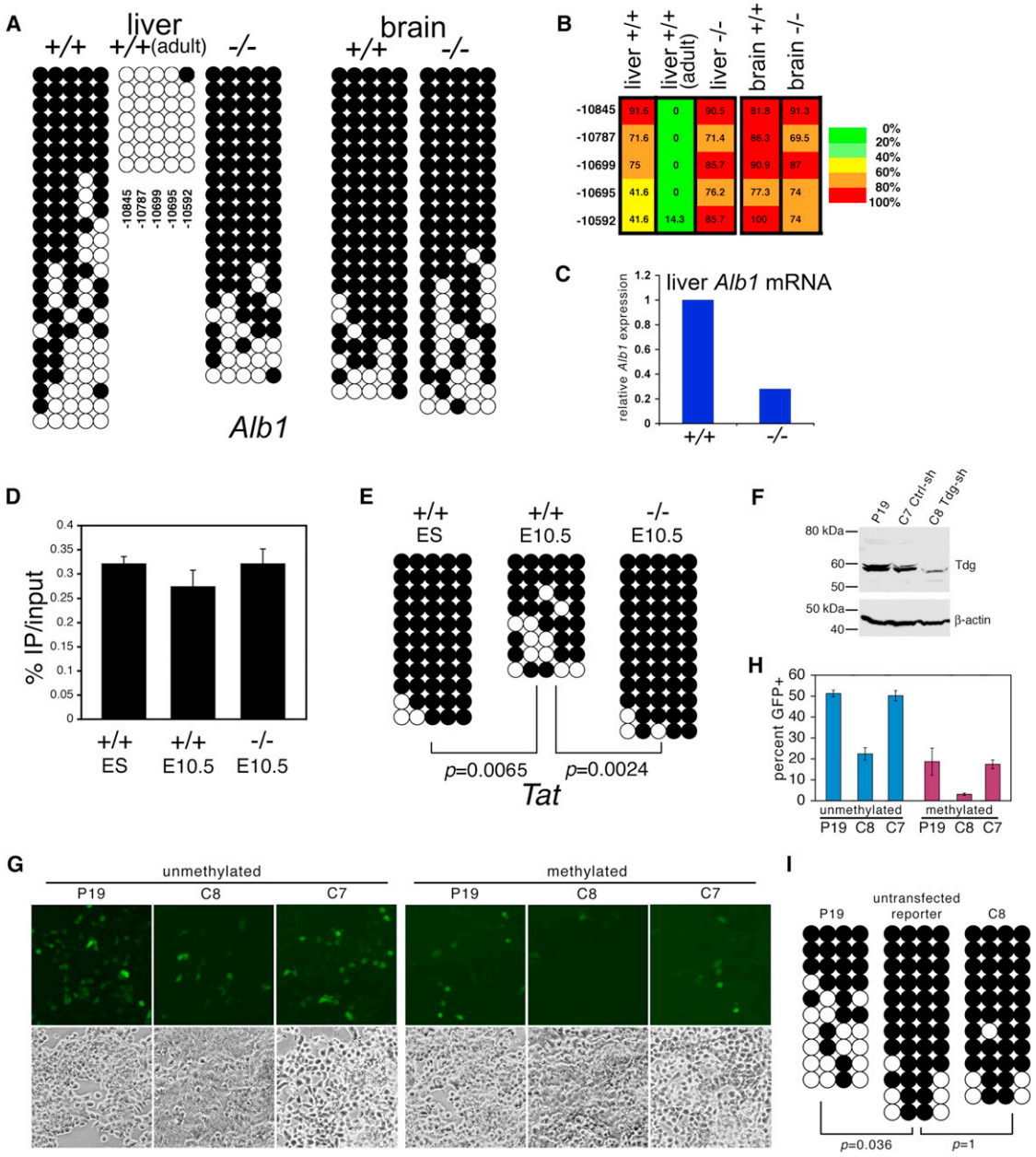

Figure 4. TDG Is Involved in DNA Demethylation

(A) DNA methylation analysis of five CpG dinucleotides of the *Alb1* enhancer in liver and brain of wild-type and *Tdg* null embryos at E11. The numbers refer to the position of CpGs relative to the transcription start site (TSS).

(B) Corresponding quantification and color-coded display of percent of DNA methylation at each CpG dinucleotide of the *Alb1* enhancer.

(C) Real-time RT-PCR quantification of *Alb1* mRNA expression, normalized to *Hprt* mRNA expression, in wild-type and *Tdg* null livers at E11.

(D) Methylated DNA immunoprecipitation quantitative PCR (MeDIP-qPCR) analysis of methylation levels at the *Tat* gene GRU, expressed as percent of immunoprecipitated DNA relative to input DNA, in ES cells, wild-type, and *Tdg* null embryos at E10.5 (headless embryo body dissected to enrich for liver).

(E) Methylation analysis by sodium bisulfite modification and sequencing of five CpG dinucleotides of the *Tat* gene GRU (at −2520, −2485, −2473, −2390, and −2386 bp relative to TSS) in ES cells, wild-type, and *Tdg* null embryos at E10.5 (dissected to enrich for liver). Two-sided Fisher's exact test at the 5% significance level.

(F) Western blot analysis showing effective downregulation of TDG in P19 C8 cells expressing a short hairpin RNA (shRNA) directed against murine *Tdg* mRNA in comparison to parental P19 cells and control shRNA P19 C7 cells.

(G) Detection by fluorescence of GFP+ cells (top) in cultures of parental P19, TDG-shRNA-containing P19 C8 cells and control shRNA C7 cells, transfected with unmethylated and SssI-methylated human *Oct4*::*EGFP* reporter. Cells were plated at approximately equal density, as evidenced by phase contrast microscopy (bottom).

(H) Quantitation of expression of unmethylated and SssI-methylated human *Oct4*::*EGFP* reporter in P19, TDG-shRNA C8, and control shRNA C7 cells.

due to demethylation of the heterologous *Oct4* promoter (Bar-reto et al., 2007). We found that the unmethylated *Oct4::EGFP* reporter was expressed within 12 hr of transfection in both parental P19 cells and its derivative line C8 bearing the TDG knockdown (the 2.3-fold expression differential between P19 and C8 likely reflects the coactivator function of TDG). In contrast, the methylated reporter is efficiently expressed only in parental P19 cells but not C8 cells (6-fold expression differential). The C7 cells expressing a control shRNA and P19 cells infected with a scrambled shRNA behaved similarly to the parental P19 line (Figures 4G and 4H and data not shown). A bisulfite sequencing analysis of the transgene recovered after 12 hr from transfected cell lines revealed that demethylation of the proximal region of the *Oct4* promoter is compromised in the *Tdg* knockdown cells, thus establishing a direct effect of TDG on demethylation (Figure 4I). The short time frame of the demethylation and the fact that the reporter plasmid used lacks an origin of replication rule out any potential effect of passive demethylation and confirm that TDG is involved in an active demethylation process.

The DNA Glycosylase Activity of TDG Is Required for Development and DNA Demethylation

If TDG has a catalytic role in DNA demethylation, the prediction is that an inactivating point mutation at the glycosylase active site would reproduce the embryonic lethality. On the other hand, lack of lethality of such mutation would suggest that TDG affects methylation patterns and development as a reflection of its coactivator function. We tested these two possibilities by generating a knockin mouse strain (Figure S1) expressing a point mutation (N151A) in the TDG glycosylase domain (Figures 5A and 5B) that eliminates the obligatory asparagine residue at the TDG active site (Hardeland et al., 2000) and abrogates glycosylase activity (Figure 5C). Live birth $Tdg^{N151A/N151A}$ homozygotes were never derived from the breeding of heterozygotes, indicating that, indeed, TDG glycosylase activity is required for embryonic development. Remarkably, analysis of timed matings revealed that lethality occurs even earlier than in knockout embryos, i.e., at ~E10.5. $Tdg^{N151A/N151A}$ embryos are much smaller than wild-type littermates and show general developmental delay, turning defect, and pericardial effusion (Figure 5D). Importantly, DNA demethylation of the *Tat* enhancer was abrogated in $Tdg^{N151A/N151A}$ embryos (Figure 5E). We conclude that the catalytic activity of TDG is essential for development and DNA demethylation.

TDG Is in a Complex with AID and GADD45a

A direct 5mC glycosylase activity of TDG has been reported (Zhu et al., 2000), but we were unable to detect such activity using a preparation of recombinant TDG that was extremely active on its cognate G:T mismatched substrates (Figure S5A). In zebrafish embryos, demethylation is initiated by enzymatic deamination of 5mC to T by AID, Apobec2a, or Apobec2b, followed by MBD4 glycosylase removal of the mismatched T in a reaction promoted by GADD45 (Rai et al., 2008). We therefore set out to determine whether TDG may mediate DNA demethylation in a similar two-step mechanism by interacting with AID/Apobec members and GADD45a in mammalian cells.

Co-IP experiments conducted in HEK293 cells transfected with cDNAs encoding tagged versions of TDG, AID, Apobec1, and GADD45a revealed that TDG forms a complex with AID (Figures 6A and 6B, lane 6) and GADD45a (Figures 6A and 6B, lane 8). In addition, transfected AID and GADD45a co-IP with endogenous TDG (Figure 6B, lanes 3 and 5). The interaction of TDG with AID is specific, as TDG does not co-IP with the AID-related family member Apobec1 (Figures 6A and 6B, lane 7). Furthermore, AID interacts with GADD45a (Figure 6C, lane 4) in a TDG-independent manner, as shown by co-IP in two cell lines having undetectable levels of TDG (Figures 6D and 6E).

We further tested whether these interactions are taking place at endogenous levels of expression in the developmentally relevant context of embryonic carcinoma P19 cells and teratocarcinoma F9 cells exhibiting high and low levels, respectively, of TDG and AID. Co-IP experiments conducted on endogenous lysates demonstrated that, in P19 cells, TDG interacts with AID (Figure 6F, top, lane 2) and that AID and GADD45a also interact (Figure 6F, second panel, lane 2). These interactions were not detected in F9 cells, likely due to the very low levels of TDG and AID expression. Interestingly, shRNA-mediated downregulation of TDG in the P19-derived C8 cells leads to reduction of AID expression (Figure 6G), suggesting that TDG may regulate the levels and/or the stability of AID. Recombinant AID and recombinant GADD45a-purified proteins directly interact, as do recombinant AID and recombinant TDG, albeit with lower affinity (Figure 6H). One possible difference to explain the stronger interaction detected in cells is that the latter may be mediated by posttranslational modifications not present in the recombinant proteins. We conclude that the TDG-AID-GADD45a interaction occurs in vivo and has functional consequences for AID levels.

Interestingly, the TDG N151A protein retains interaction with AID and GADD45a (Figure S6), suggesting that the more severe phenotype of the $Tdg^{N151A/N151A}$ compared to the *Tdg* knockout embryos might be due to a dominant-negative action of the catalytically dead mutant protein in sequestering AID and GADD45 or other interactors in nonfunctional, nonproductive complexes.

TDG Has Glycosylase Activity on 5hmU

Because 5hmC, the hydroxylation product of 5mC, has been proposed as an intermediate in demethylation, we assayed but failed to detect 5hmC glycosylase activity for TDG or any other related glycosylases (Figure S5B). However, during active DNA demethylation in mammalian cells, AID/Apobec deaminases convert 5hmC to 5hmU for subsequent processing by BER (Guo et al., 2011). We therefore assayed the 5hmU glycosylase activity of TDG and related glycosylases. As previously described

(l) DNA methylation analysis by sodium bisulfite modification and sequencing of the proximal region (region 8) (Deb-Rinker et al., 2005) of the human *Oct4* promoter from the untransfected methylated *Oct4::EGFP* reporter and the same reporter transfected and recovered from P19 and C8 cells. Two-sided Fisher's exact test at the 5% significance level.

Data are presented as mean ± SEM.

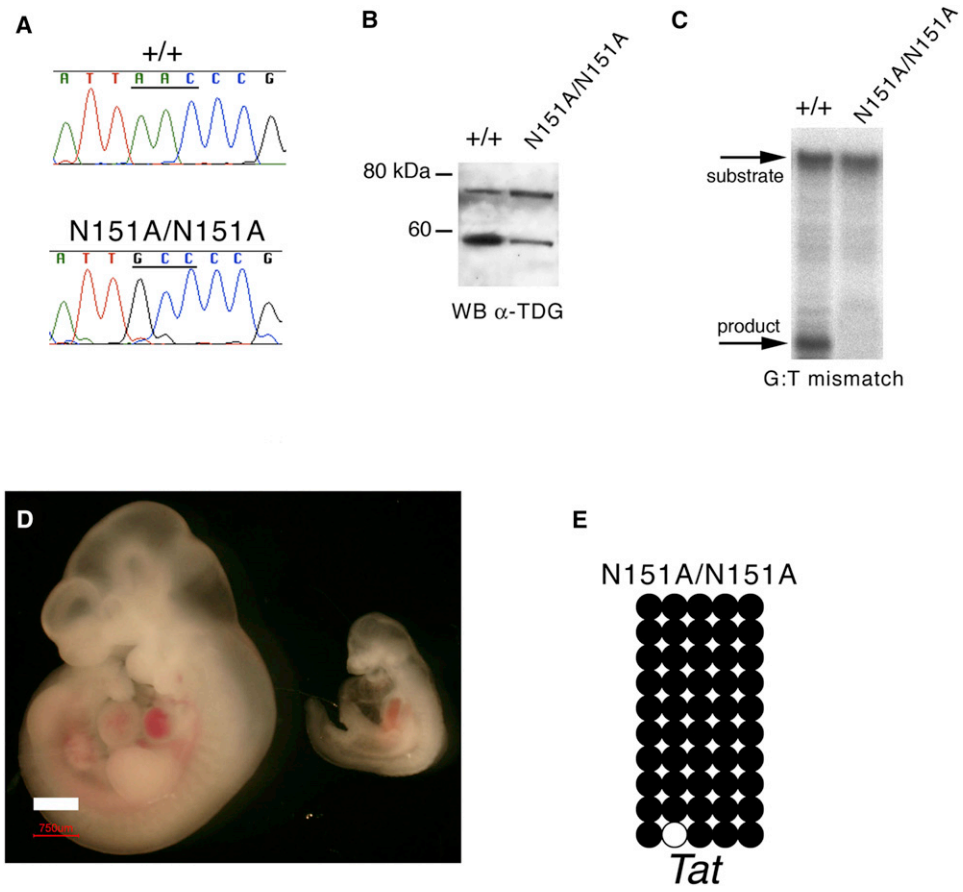

Figure 5. The DNA Glycosylase Activity of TDG Is Required for Development and DNA Demethylation

(A) Sequence analysis of a cDNA fragment encompassing the relevant *Tdg* exon 4 region from wild-type and $Tdg^{N151A/N151A}$ E10.5 embryo total RNA confirms expression of the knockin allele.

(B) Western blot analysis with an anti-TDG antibody reveals expression of the wild-type and TDG^{N151A} protein in E10.5 embryo lysates of corresponding genotypes.

(C) Repair of a double-stranded oligonucleotide containing a G:T mismatch by nuclear extracts of whole E10.5 embryo extracts of the indicated genotypes.

(D) Gross phenotype of wild-type (left) and $Tdg^{N151A/N151A}$ (right) littermate embryos at embryonic day E10.5. Scale bar, 750 μm.

(E) Methylation analysis by sodium bisulfite modification and sequencing of five CpG dinucleotides of the *Tat* gene GRU in a $Tdg^{N151A/N151A}$ E10.5 embryo (headless embryo body preparations dissected to enrich for liver). Comparison of methylation levels with the wild-type embryo in Figure 4E was made using the two-sided Fisher's exact test at the 5% significance level and revealed a p value equal to 0.0004.

See also Figure S1.

(Haushalter et al., 1999), SMUG1 has activity on 5hmU in single-strand DNA or when paired with adenine in double-strand DNA to resemble an expected product of thymine oxidation (Figure 7A). Remarkably, similar to a sequence-unrelated thermophilic Tdg (Baker et al., 2002), mammalian TDG exhibits robust activity, comparable to SMUG1, on a 5hmU:G mismatch in double-strand DNA, the expected product of deamination following hydroxylation of 5mC (Figure 7A). This result expands the suggested role of TDG downstream of deamination to include initiation of BER following hydroxylation of 5mC (Figure 7B).

Lack of Promoter Mutations in *Tdg* Mutant Cells

Deamination of 5mC or 5hmC in the absence of TDG-mediated repair of the resulting G:T or G:5hmU mismatch is expected to increase G:C-to-A:T transition mutations. For this reason, we conducted a sequence analysis of nonbisulfite-modified DNA of promoters undergoing TDG-dependent protection from hypermethylation or DNA demethylation. No mutation was found in *H19*, *Efs*, *HoxA5*, and *Crabp2* promoters in Tdg mutant MEFs or at the *Oct4* promoter of the *Oct4::EGFP* transgene recovered from the C8 Tdg knockdown cells (data not shown). Although a possible compensation by MBD4, SMUG1, or mismatch repair cannot be ruled out, this result suggests the possibility that the deamination and glycosylase steps are coordinated, such that deamination does not occur in the absence of TDG.

DISCUSSION

Our results demonstrate that TDG is required for normal mammalian development and the establishment of proper promoter/enhancer DNA methylation patterns that are conducive to transcription during embryogenesis. Failure to establish and

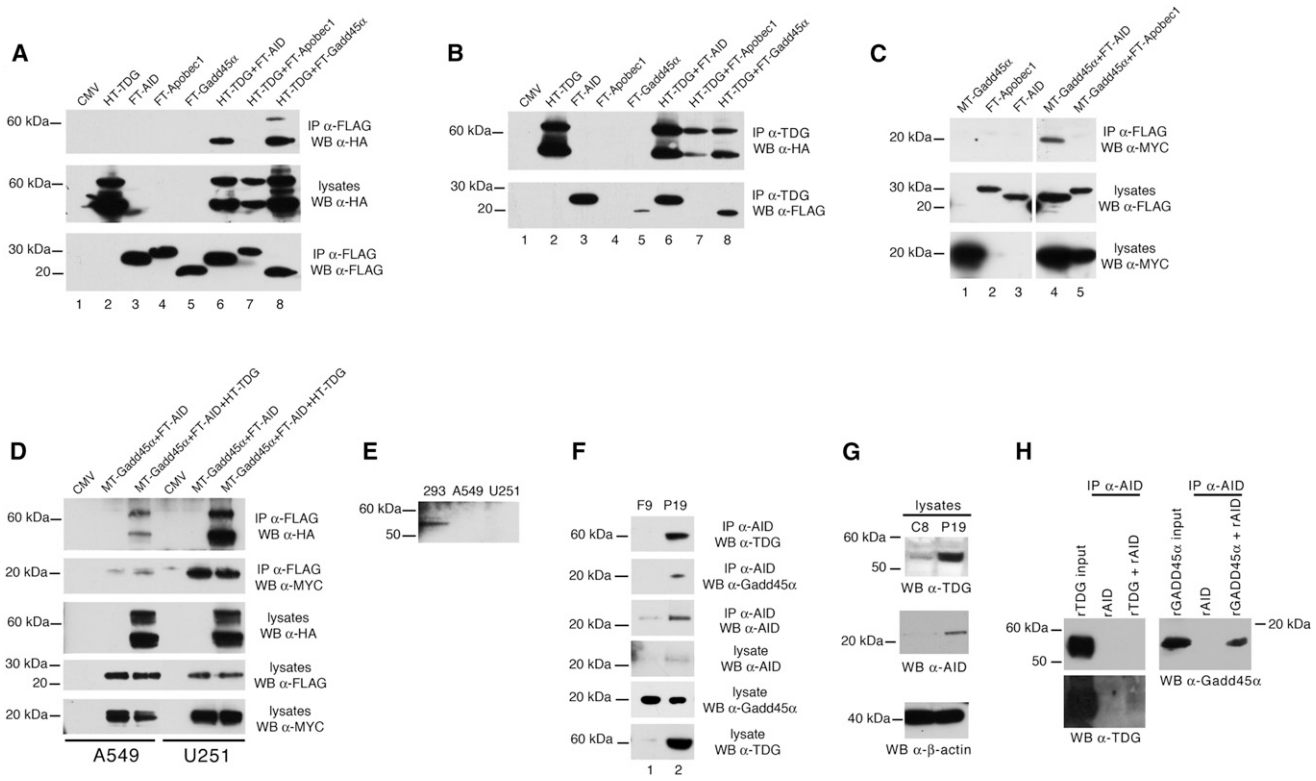

Figure 6. TDG Is in a Complex with AID and GADD45a

(A–D) Immunoprecipitates of lysates of HEK293 (A–C) or A549 and U251 (D) cells transfected with hemagglutinin (HA)-, FLAG-, or MYC-tagged expression constructs were resolved by PAGE and detected by western blotting with the indicated antibodies. Western blotting of lysates shows that identical tagged constructs were expressed at approximately equal levels. In these and other western blots, the presence of additional TDG bands is the result of SUMOylation or other posttranslational modifications (Cortázar et al., 2007). Note that, in (B), FLAG-tagged AID and FLAG-tagged GADD45a immunoprecipitate not only with exogenous HA-tagged TDG (lanes 6 and 8), but also with endogenous TDG (lanes 3 and 5).

(E) Western blotting with an anti-TDG antibody reveals TDG expression in HEK293 cells, but not in A549 or U251 cells.

(F) Coimmunoprecipitation experiments with the indicated antibodies show that AID forms a complex with TDG and GADD45a at endogenous levels of expression in P19 embryonic carcinoma cells. Western blotting with anti-AID antibody confirms that AID was actually immunoprecipitated.

(G) AID levels are reduced by shRNA in the P19 derivative, TDG knockdown cell line C8, as evidenced by western blotting of lysates with the indicated anti-TDG and anti-AID antibodies. Western blotting with an anti-β-actin antibody acts as a loading control.

(H) The indicated recombinant proteins were premixed, and the mixtures were immunoprecipitated with an anti-AID antibody. Immunoprecipitates along with the input recombinant TDG or GADD45a were detected by western blotting with an anti-TDG or anti-GADD45a antibody, as indicated. GADD45a is readily detected, whereas only a small amount of TDG (visible in the longer exposure, bottom-left) precipitates with AID, suggesting a low-affinity interaction.

See also Figure S5 and Figure S6.

maintain correct DNA methylation patterns is likely the cause of lethality and the observed complex developmental phenotype, as it is known that even small changes in DNA methylation cause abnormalities or lethality (Gaudet et al., 2003).

In maintaining the proper epigenetic states, TDG apparently has a dual role: protection from aberrant hypermethylation (examples include the CpG islands of *Efs*, *HoxA5*, and *Crabp2* in MEFs and the maternal alleles of *H19* and *Igf2* DMR2 in MEFs and PGCs) and promotion of demethylation (exemplified by the *Alb1* and *Tat* enhancers in hepatoblasts and the *Oct4::EGFP* reporter in P19 cells).

Our combined biochemical and developmental data suggest that demethylation is an active process requiring TDG catalytic activity immediately downstream of the deaminase-catalyzed conversion of 5mC into thymine and/or of 5hmC into 5hmU. Thus, both the 5mC deamination and hydroxylation-deamination pathways may converge on TDG (Figure 7B), and perhaps this convergence may explain the absolute requirement of TDG during embryogenesis, although the relative prevalence of each pathway in different developmental contexts is currently unknown. The extent of possible partial compensation by activities of MBD4 and SMUG1 is also unknown, although the observed lethality of the *Tdg* knockout suggests that it is not sufficient to maintain embryogenesis.

Similarity of the lethal phenotype of the *Tdg* knockin and knockout embryos and, in turn, their resemblance to the *p300/Cbp* and *Rar/Rxr* mutant embryos strongly suggest the possibility that protection from aberrant de novo methylation may also involve an active process, requiring TDG glycosylase activity to constantly antagonize methylation. However, a noncatalytic role of TDG cannot be ruled out completely, and it is possible that protection from hypermethylation may occur at

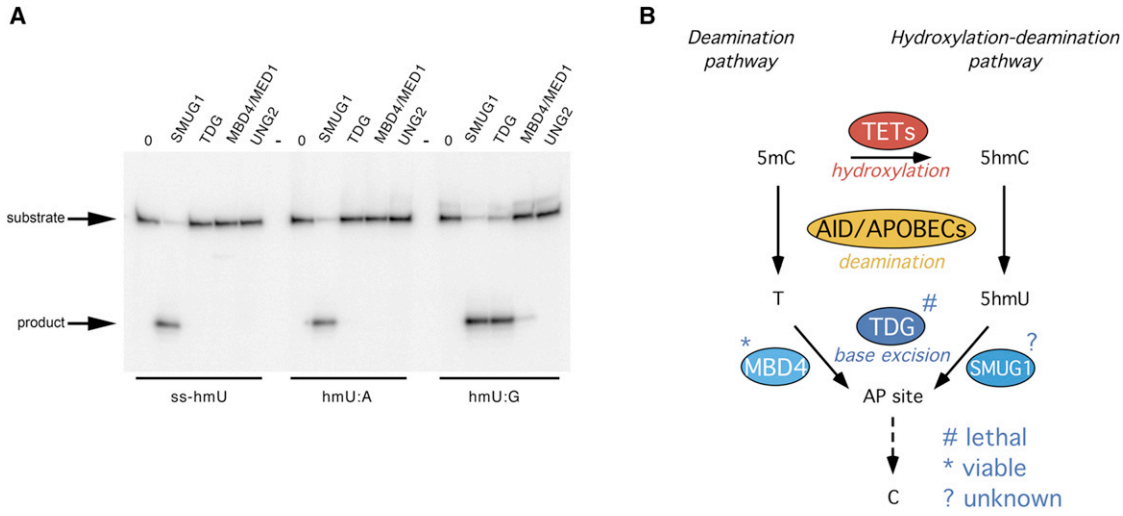

Figure 7. TDG Glycosylase Activity on 5hmU and Model of the Role of TDG in DNA Demethylation Pathways

(A) Recombinant TDG and related glycosylases were incubated with 5hmU-containing single-strand oligonucleotide or double-strand oligonucleotides bearing 5hmU:A pairing or 5hmU:G mismatch, all ^{32}P labeled on the 5hmU strand. The resulting AP site was cleaved with alkali at high temperature.

(B) Schematic of the involvement of TDG in both the deamination and hydroxylation-deamination pathways of DNA demethylation. Lethality/viability of the glycosylase knockout mice is indicated.

See also Figure S5 and Figure S7.

least in part via TDG inhibition of de novo DNA methyltransferases (Li et al., 2007).

Our observation that TDG is required for the interaction of RAR/RXR with p300 both on and off the DNA suggests a model in which transcription factor binding (which is TDG independent; Figures 2D and 2E) may be responsible for tethering TDG to the promoter/enhancer (Figure S7). Thus, it is possible that the TDG-provoked demethylation of differentiation-associated promoters/enhancers depends on TDG tethering by tissue-specific transcription factors. Of note, GADD45 is known to bind to nuclear hormone receptors (Yi et al., 2000) and preferentially to hyperacetylated nucleosomes (Carrier et al., 1999), suggesting that GADD45 may also be involved in targeting promoters for demethylation. In this model, proper targeting of AID may, in turn, depend on its interaction with GADD45 and TDG. Future binding studies conducted on a genomic scale in different tissues and at various developmental stages will further define the promoters/enhancers that are direct targets of the TDG-AID-GADD45a complex.

The identification of a ternary complex containing TDG, AID, and GADD45a is consistent with the recent recognition of AID as a factor required for DNA demethylation during reprogramming of somatic cells (Bhutani et al., 2010) and erasure of DNA methylation at imprinted and other loci in PGCs (Popp et al., 2010), as well as with the role of GADD45a (Barreto et al., 2007; Schmitz et al., 2009) and the related GADD45b (Ma et al., 2009) in demethylation of specific promoters. It is also consistent with the role of BER in genome-wide active DNA demethylation in PGCs (Hajkova et al., 2010). However, *Aid*-deficient mice are viable and fertile (Muramatsu et al., 2000; Revy et al., 2000), which suggests that TDG may also function downstream of Apobec deaminases and possibly engage different GADD45 proteins. In fact, various TDG-deaminase-GADD45 complexes and possibly different TET proteins may be utilized for demethylation associated with distinct developmental processes and reprogramming events.

Our observation of the lack of deamination-induced transition mutations in *Tdg* mutant MEFs and *Tdg* knockdown cells at promoters that undergo TDG-dependent protection from hypermethylation and demethylation suggests that TDG has a role not only in repairing deamination products, but also in initiating the DNA demethylation process, thus controlling the potentially mutagenic deaminase activity of AID. It is possible that initiation is regulated by targeting of the AID-TDG-GADD45a complex to relevant promoter/enhancers or by the optimal reciprocal amounts of these proteins in the complex, as TDG affects AID levels and/or stability (Figure 6G). Given the frequent CpG transition mutations and hypermethylation of tumor suppressor gene promoters in human cancer, inactivation of TDG and its demethylating complex or altered relative amounts of TDG, AID, and GADD45a may play a role in tumor formation.

EXPERIMENTAL PROCEDURES

Derivation of *Tdg* Null and *Tdg* Knockin Mice

Tdg null and *Tdg* knockin mice were generated by homologous recombination of positive-negative selection targeting constructs in R1 ES cells. ES clones carrying the targeted *Tdg* locus were injected into C57/BL6 blastocysts to generate chimeric mice. The chimeric mice were backcrossed into the C57/BL6 line for at least eight generations.

Isolation of MEFs

Mouse embryo fibroblasts (MEFs) were prepared as previously described (Cortellino et al., 2003) from embryos harvested at E10.5 and were grown in DMEM supplemented with 15% fetal bovine serum.

Isolation of Primordial Germ Cells

PGCs were isolated from E11 germinal ridges dissected by pipetting up-down and trypsin digestion. PGCs stained with anti-SSEA1 antibody were sorted by FACS. The morphology of the collected PGCs was evaluated after phosphatase alkaline staining to confirm cellular specificity.

Analysis of DNA Methylation by Bisulfite Modification Sequencing

Genomic DNA was subjected to the sodium bisulfite modification reaction, as previously described (Howard et al., 2009). Products from the bisulfite reactions were amplified by PCR using primers designed with the MethPrimer software at http://www.urogene.org/methprimer/. Purified PCR products were subcloned into pGEM T-Easy vector (Invitrogen), and individual inserts from 10–15 clones were sequenced. Comparisons of methylation levels were made using the two-sided Fisher's exact test at the 5% significance level.

Analysis of Active DNA Demethylation

P19, P19 C8, P19 C7, and P19 scrambled shRNA cells were transfected with unmethylated or SssI in vitro-methylated *Oct4*::*EGFP* reporter plasmid. Cells were visualized for GFP expression 12 hr after transfection. Transfected plasmid was recovered with QIAGEN Midikit and then processed for bisulfite treatment.

ACCESSION NUMBERS

Microarray data were submitted to ArrayExpress with accession number E-MEXP-2610.

SUPPLEMENTAL INFORMATION

Supplemental information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at doi:10.1016/j.cell.2011.06.020.

ACKNOWLEDGMENTS

We thank Drs. F. De Angelis, H.-Y. Fan, R. Katz, O. Segatto, K. Zaret, and R. Zhang for critical reading of the manuscript; Drs. P.K. Cooper, F. Roegiers, and K. Soprano for comments and advice; G. Albergo, K. Brewer, C. Garnier, B. Lurie, and M. Oliver for mouse genotyping; Dr. Z.P. Zhang for help with CAT assays; Dr. J. Kulkosky, A. Kowalski, T. Stulkivska, and W. Schroeder for *Crabp2* and *Oct4* methylation analysis; Dr. M. Xu for the targeting plasmid (with permission from Dr. G. Martin); Dr. A.J. Furnace for recombinant GADD45a; Dr. K. Sugasawa for anti-TDG antibody; Drs. L. Bagella and P.L. Puri for p300 reporters; Drs. S. Peri and Y. Zhou for ingenuity pathway analysis; Dr. F. Alt for AID cDNA; Dr. N. Davidson for Apobec-1 cDNA; Dr. W. Cui for human *Oct4*::*EGFP* reporter; Drs. C. Niehrs and A. Schafer for advice on transfected plasmid recovery; Dr. J. Thorvaldsen for advice on bisulfite modification of PGC DNA; and R. Sonlin for secretarial assistance. We thank the following core facilities at the Fox Chase Cancer Center: Genotyping, Cell Culture, Transgenic and Knock-out, Laboratory Animal, and Fannie E. Rippel Biotechnology Facility. Y.M. would like to thank Dr. Alan Tomkinson for providing hospitality to conduct some of the experiments in his laboratory. This study was supported by NIH grants CA78412, CA06927, and DK067558; an appropriation from the Commonwealth of Pennsylvania to the Fox Chase Cancer Center; and the Italian Association for Cancer Research (AIRC). S.C. was supported, in part, by an American-Italian Cancer Foundation Post-Doctoral Research Fellowship. M.L.C. was supported by an Epigenetics and Progenitor Cells Keystone Program Fellowship. L.K.A. was supported by NIH training grant T32GM008216. A.B. would like to dedicate this article to the memory of his father Pancrazio who taught him the importance of hard work, perseverance and dedication.

Received: February 18, 2011
Revised: May 17, 2011
Accepted: June 12, 2011
Published online: June 30, 2011

REFERENCES

Baker, D., Liu, P., Burdzy, A., and Sowers, L.C. (2002). Characterization of the substrate specificity of a human 5-hydroxymethyluracil glycosylase activity. Chem. Res. Toxicol. *15*, 33–39.

Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F., and Niehrs, C. (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671–675.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042–1047.

Bird, A. (1992). The essentials of DNA methylation. Cell *70*, 5–8.

Carrier, F., Georgel, P.T., Pourquier, P., Blake, M., Kontny, H.U., Antinore, M.J., Gariboldi, M., Myers, T.G., Weinstein, J.N., Pommier, Y., and Fornace, A.J., Jr. (1999). Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin. Mol. Cell. Biol. *19*, 1673–1685.

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

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

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

Cortellino, S., Turner, D., Masciullo, V., Schepis, F., Albino, D., Daniel, R., Skalka, A.M., Meropol, N.J., Alberti, C., Larue, L., and Bellacosa, A. (2003). The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. Proc. Natl. Acad. Sci. USA *100*, 15071–15076.

Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M., and Walker, P.R. (2005). Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. J. Biol. Chem. *280*, 6257–6260.

Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. Nat. Rev. Cancer *4*, 143–153.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489–492.

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

Gehring, M., Reik, W., and Henikoff, S. (2009). DNA demethylation by DNA repair. Trends Genet. *25*, 82–90.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE *5*, e15367.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. Cell *145*, 423–434.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science *329*, 78–82.

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

Haushalter, K.A., Todd Stukenberg, M.W., Kirschner, M.W., and Verdine, G.L. (1999). Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. Curr. Biol. *9*, 174–185.

Howard, J.H., Frolov, A., Tzeng, C.W., Stewart, A., Midzak, A., Majmundar, A., Godwin, A.K., Heslin, M.J., Bellacosa, A., and Arnoletti, J.P. (2009). Epigenetic downregulation of the DNA repair gene MED1/MBD4 in colorectal and ovarian cancer. Cancer Biol. Ther. *8*, 94–100.

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

Jin, S.G., Guo, C., and Pfeifer, G.P. (2008). GADD45A does not promote DNA demethylation. PLoS Genet. *4*, e1000013.

Jones, P.A., and Laird, P.W. (1999). Cancer epigenetics comes of age. Nat. Genet. *21*, 163–167.

Kass, S.U., Landsberger, N., and Wolffe, A.P. (1997). DNA methylation directs a time-dependent repression of transcription initiation. Curr. Biol. *7*, 157–165.

Kim, M.S., Kondo, T., Takada, I., Youn, M.Y., Yamamoto, Y., Takahashi, S., Matsumoto, T., Fujiyama, S., Shirode, Y., Yamaoka, I., et al. (2009). DNA demethylation in hormone-induced transcriptional derepression. Nature *461*, 1007–1012.

Kress, C., Thomassin, H., and Grange, T. (2006). Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks. Proc. Natl. Acad. Sci. USA *103*, 11112–11117.

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

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

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

Ma, D.K., Jang, M.H., Guo, J.U., Kitabatake, Y., Chang, M.L., Pow-Anpongkul, N., Flavell, R.A., Lu, B., Ming, G.L., and Song, H. (2009). Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. Science *323*, 1074–1077.

Mark, M., Ghyselinck, N.B., and Chambon, P. (2006). Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. Annu. Rev. Pharmacol. Toxicol. *46*, 451–480.

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

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

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

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

Morales-Ruiz, T., Ortega-Galisteo, A.P., Ponferrada-Marín, M.I., Martínez-Macías, M.I., Ariza, R.R., and Roldán-Arjona, T. (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. Proc. Natl. Acad. Sci. USA *103*, 6853–6858.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553–563.

Niehrs, C. (2009). Active DNA demethylation and DNA repair. Differentiation *77*, 1–11.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

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

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

Pantoja, C., de Los Ríos, L., Matheu, A., Antequera, F., and Serrano, M. (2005). Inactivation of imprinted genes induced by cellular stress and tumorigenesis. Cancer Res. *65*, 26–33.

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

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

Reese, K.J., and Bartolomei, M.S. (2006). Establishment and maintenance of H19 imprinting in the germline and preimplantation embryo. Cytogenet. Genome Res. *113*, 153–158.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science *293*, 1089–1093.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., et al. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565–575.

Rideout, W.M., III, Eggan, K., and Jaenisch, R. (2001). Nuclear cloning and epigenetic reprogramming of the genome. Science *293*, 1093–1098.

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

Siegfried, Z., and Cedar, H. (1997). DNA methylation: a molecular lock. Curr. Biol. *7*, R305–R307.

Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. Cell *128*, 747–762.

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

Tanaka, Y., Naruse, I., Hongo, T., Xu, M., Nakahata, T., Maekawa, T., and Ishii, S. (2000). Extensive brain hemorrhage and embryonic lethality in a mouse null mutant of CREB-binding protein. Mech. Dev. *95*, 133–145.

Thomassin, H., Flavin, M., Espinás, M.L., and Grange, T. (2001). Glucocorticoid-induced DNA demethylation and gene memory during development. EMBO J. *20*, 1974–1983.

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

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

Vermot, J., Niederreither, K., Garnier, J.M., Chambon, P., and Dollé, P. (2003). Decreased embryonic retinoic acid synthesis results in a DiGeorge syndrome phenotype in newborn mice. Proc. Natl. Acad. Sci. USA *100*, 1763–1768.

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

Xu, J., Pope, S.D., Jazirehi, A.R., Attema, J.L., Papathanasiou, P., Watts, J.A., Zaret, K.S., Weissman, I.L., and Smale, S.T. (2007). Pioneer factor interactions

and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. Proc. Natl. Acad. Sci. USA *104*, 12377–12382.

Yao, T.P., Oh, S.P., Fuchs, M., Zhou, N.D., Ch'ng, L.E., Newsome, D., Bronson, R.T., Li, E., Livingston, D.M., and Eckner, R. (1998). Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell *93*, 361–372.

Yi, Y.W., Kim, D., Jung, N., Hong, S.S., Lee, H.S., and Bae, I. (2000). Gadd45 family proteins are coactivators of nuclear hormone receptors. Biochem. Biophys. Res. Commun. *272*, 193–198.

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

EXTENDED EXPERIMENTAL PROCEDURES

Design of the *Tdg* Targeting Construct and Derivation of *Tdg* Null and *Tdg* Knockin Mice

For construction of the *Tdg* targeting vector, genomic clones from the mouse *Tdg* locus were obtained by screening a 129/Sv-J lambda library (Stratagene) with a probe derived from the murine *Tdg* cDNA (Image clone 4195230). Since the lambda clones obtained did not cover the full gene, murine genomic probes from the lambda clones, corresponding to portions of introns 1 and 9, were used to isolate BAC 526 A17 from a 129/Sv-J library (Research Genetics). This BAC was used as a template for long-range PCR reactions that amplified the following: a 3.5 kb fragment containing exon 2 and approximately 2/3 of intron 2 along with 3′ engineered BglII and *lox*P site (left arm), a 3.3 kb fragment containing exons 3 to 7 along with a *lox*P site (central arm), and a 4.4 kb fragment containing intron 7, exons 8, 9 and the first half of exon 10 (right arm). These fragments were fully sequenced and then inserted in pKO Scrambler v.903 (Lexicon Genetics), along with the neo^R gene flanked by *FRT* sites from pK11/pM-30 for positive selection, and the diphteria toxin gene driven by the *pol2* promoter for negative selection. Targeting vectors were linearized with NotI and electroporated into R1 ES cells. ES cell colonies bearing gene replacement were selected with G418 (300 μg/ml), and further screened by Southern Blot analysis using 5′- and 3′-flanking probes. Five positive ES cell clones were isolated out of 85. Two positive ES clones carrying the targeted *Tdg* locus were injected into C57/BL6 blastocysts to generate chimeric mice. Male chimeras were mated to C57/BL6 females for germline transmission of the mutation. The *Tdg* allele produced after germ-line transmission was referred to as $Tdg^{neoflox}$.

The constitutive null allele of *Tdg*, termed Tdg^-, was created by mating heterozygous $Tdg^{neoflox}$ mice with homozygous *EIIa::Cre* mice that express the Cre recombinase ubiquitously (Lakso et al., 1996). F1 progeny that demonstrated a high degree of mosaicism counter to $Tdg^{neoflox}$ were bred with C57/BL6 mice. The F2 progeny was genotyped for *EIIa::Cre*, Tdg^- and $Tdg^{neoflox}$; mice inheriting the Tdg^- allele but lacking *EIIa::Cre* were bred further by backcrossing into the C57/BL6 line.

A similar strategy was used to generate the Tdg^{N151A} knock-in mice. In this case, the targeted recombinant $Tdg^{neoflox\ N151A}$ allele (Figure S6) was produced in ES cells after homologous recombination. Mice bearing this targeted allele were bred with mice expressing the Flp recombinase at the Rosa26 locus (*Rosa26::FlpeR*), and the offspring were genotyped to score for Flp-mediated recombination of the *FRT* sites, with consequent removal of the *neo* cassette and generation of the knock-in Tdg^{N151A} allele. Mice bearing the knock-in Tdg^{N151A} allele were then backcrossed into the C57/BL6 line.

All mouse experiments were performed in accordance with relevant guidelines and regulations, and were approved by the Fox Chase Cancer Center Institutional Animal Care and Use Committee.

Isolation of MEFs

Mouse embryo fibroblasts (MEFs) were prepared as previously described (Cortellino et al., 2003) from embryos harvested at E10.5, and grown in DMEM supplemented with 15% fetal bovine serum.

Preparation of Nuclear Extracts and Repair Assay

Nuclear extracts were prepared according to previous reports (Challberg and Kelly, 1979; Holmes et al., 1990). The following oligonucleotides were purified by 8.3 M urea/15%–20% PAGE (Turner et al., 2006), and annealed: CAATCCTAGCTGACACGATGTGGC CAATGGCATGACT (top) and GAGTCATGCCATTGGCCACATYGTGTCAGCTAGGATT (bottom), where Y is either T or U. Prior to annealing, the bottom strand oligonucleotide was radiolabeled at the 5′ end with T4 polynucleotide kinase (New England Biolabs) and γ-^{32}P-ATP (NEN Dupont). Repair (excision) assays were carried out with 15 μg of nuclear extract and 10 nM of double-stranded oligonucleotide substrate at 30°C for 16 hr in reaction buffer (15 mM HEPES pH 7.9, 0.75 mM DTT, 0.75 mM EDTA) in a final volume of 20 ml. Reactions were terminated with 50 μl stop buffer (1 M Tris-HCl, 0.5 M EDTA, 10% SDS, 10 mg/ml proteinase K). Oligonucleotides were phenol-chloroform extracted and ethanol precipitated. For the reactions with recombinant DNA glycosylases, substrate DNA was incubated with purified recombinant MED1 catalytic domain or purified recombinant TDG in the same reaction buffer supplemented with 0.8 μg/μl of bovine serum albumin at 37°C for 30 min (Petronzelli et al., 2000b), and the reaction was treated with NaOH at 90°C for 30 min to cleave the abasic site. Substrate and product bands were separated by 8.3 M urea/25% PAGE and exposed to autoradiography.

For the glycosylase assays on 5-hydroxymethyldeoxyuracil, a 41-mer oligonucleotide (CAGGTACGTCAACGGAACGAHmUAC CACTAGAGCTGGGCCGTC, where HmU = 5-hydroxymethyldeoxyuridine) was labeled at the 5′-terminus with ^{32}P. The labeled oligonucleotide was directly used for the single-strand assay or used after annealing with a 43-mer oligonucleotide (CGTCCATG CAGTTGCCTTGCTXTGGTGATCTCGACCCGGCAGT, where X = either A or G) for HmU:A or HmU:G assays. Alternatively, the oligonucleotides CAGGTACGTCAACGGAAHmUGATACCACTAGAGCTGGGCCGTC and GAGTCATGCCATTGGCCACATHmUGTGT CAGCTAGGATT were labeled at the 5′-terminus with ^{32}P and annealed to TGACGGCCCAGCTCTAGTGGTATCGTTCCGTTGACG TACCTGC and CAATCCTAGCTGACACGATGTGGCCAATGGCATGACT, respectively. For the glycosylase assays on 5-hydroxymethylycytosine, a 37-mer oligonucleotide (GAGTCATGCCATTGGCCACATHmCGTGTCAGCTAGGATT where HmC = 5-hydroxymethyldeoxycytidine) was labeled at the 5′-terminus with ^{32}P and annealed with the 37-mer oligonucleotide CAATCCTAGCTGACAC GATGTGGCCAATGGCATGACT. Assays were conducted in a 10 μl reaction mixture containing 20 mM HEPES, pH7.5, 100 mM NaCl, 1 mM EDTA, 0.1 μg/μl BSA, 0.1 pmole of the indicated oligonucleotides and 4 pmole of the indicated DNA glycosylase at

37°C for 30 min. Reactions were terminated by the addition of 1 μl of 1 M NaOH followed by incubation at 97°C for 3 min, and subsequently neutralized with 1 μl of 1 M acetic acid. After addition of 20 μl formamide/dyes, 5 μl of each sample was subjected to electrophoresis in a 20% Sequagel (National Diagnostics), and visualized with a Fuji BAS phosphorimager.

Embryo Harvest, Histology, and Immunohistochemistry

Timed matings were set up between heterozygous mice and the embryos were dissected at different time points. Embryos were fixed in 4% paraformaldehyde overnight at 4°C, dehydrated in increasing ethanol series (70, 80, 90% and absolute) for 1 hr each step. Before embedding in paraffin, embryos were incubated with ethanol:xylene (50:50) and xylene for 1 hr each. Sections (5-10 μm) were prepared for hematoxylin and eosin staining. After deparaffinization and rehydration using xylene and decreasing ethanol series (absolute, 95, 80%), the sections were stained with hematoxylin for 3 min, rinsed in deionized water, destained in acid ethanol, and washed again with deionized water. Eosin staining was performed by incubating the sections in eosin for 30 s, ethanol 95% and absolute for 15 min each, and xylene for 45 min. Slides were mounted using Permount media.

For immunohistochemistry, embryos were fixed in 4% (v/v) paraformaldehyde at 4°C overnight, and, after several washes with 0.1% Tween-phosphate-buffered saline (PBST), stored in methanol at –20°C. Embryos were rehydrated, bleached with 5% H_2O_2 in methanol for 1 hr at room temperature, washed briefly in PBST and incubated in blocking solution, consisting of 4% bovine serum albumin (BSA) in PBST. Whole-mount embryos were then incubated overnight at 4°C with rat anti-mouse platelet endothelial cell adhesion molecule (PECAM)-1 antibody (PharMingen) diluted 1/10 in blocking solution, washed in blocking solution for 2 hr at 4°C, incubated overnight at 4°C with peroxidase-conjugated anti-rat immunoglobulin diluted 1/100 in blocking solution, and washed as above. Staining was revealed by incubation with diaminobenzidine-hydrogen peroxide.

Western Blot Analysis and Coimmunoprecipitation

Cells were lysed on ice in RIPA buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 1% sodium deoxycholate, 1% Triton X-100, 0.1% SDS, 10 mM NaF, 1 mM each of sodium pyrophosphate, sodium orthovanadate, dithiothreitol, and EDTA), plus protease inhibitors. Lysates were fractionated by SDS-PAGE and transferred to PVDF membranes (Millipore). Membranes were blocked in 4% nonfat dry milk in PBS and incubated with anti-Tdg antibody, 1/400 dilution in 2% nonfat dry milk in PBS (anti-β-actin, anti-RAR and anti-p300 from Santa Cruz). Detection was performed using enhanced chemiluminescence (Amersham).

For co-immunoprecipitation, cells were lysed on ice in buffers containing 0.2% Nonidet P-40, 40 mM Tris, pH8, 150 mM NaCl, 10% glycerol, 1 mM $ZnCl_2$, 10 mM NaF, and supplemented with complete proteinase inhibitor (Roche), as previously described (Bellacosa et al., 1999). Lysates were sonicated 4 times for 30 s on ice, centrifuged at 13,000 g for 10 min and the supernatants were incubated with anti-pan retinoic acid receptor (pan-RAR) antibody (Santa Cruz) for 16 hr at 4°C, or with anti-AID (Cell Signaling), anti-FLAG (Sigma), and anti-TDG antibody for 2 hr at 4°C. Immune complexes were collected on protein A-protein G beads for 1 hr, washed 3-4 times with lysis buffer, incubated at 42°C for 7 min, resolved by 4%–12% SDS-PAGE and processed for western as above, using an anti-p300 (Calbiochem), anti-TDG, anti-Gadd45α (Cell Signaling), anti-HA (Covance), anti-MYC (Cell Signaling), and anti-FLAG (Sigma) antibodies. For the immunoprecipitations with recombinant proteins, AID (Enzymax), GADD45a and TDG (0.5 μg each) were mixed for 30 min at 4°C in 150 mM NaCl, 40 mM Tris-HCl pH8, 1 mM $MgCl_2$, and 1 mM $ZnCl_2$, 10% glycerol, and then processed for immunoprecipitation as above.

Transcriptional Activation Assays

For p300 transcriptional activation assays, MEFs were transfected with the following CMV promoter-based plasmids: β-galactosidase (0.5 μg), *Gal4* operator-luciferase reporter (1 μg), and either Gal4 DNA binding domain alone or Gal4 DNA binding domain fused to p300 (1.5 μg). Transfections were performed with a Nucleofector device, using MEF2 Nucleofector solution and the T-20 program (Amaxa). Cells were lysed 48 hr after transfection with 1x Report Lysis Buffer (Promega), and luciferase and β-galactosidase assays were conducted following the manufacturer's recommendations. For each experiment, luciferase activity was normalized to β-galactosidase. The change in normalized luciferase activity was calculated relative to that of wild-type cells transfected with reporter plasmids and Gal4 DNA binding domain alone, which was set as 1. Values are the means ± standard deviations of a representative experiment performed in triplicate.

For retinoic acid-dependent transcription assay (Scafonas et al., 1997; Tairis et al., 1994), MEFs were transfected as above with 1 μg of CMV-β-galactosidase and 2 μg of pSG5-DR5 retinoic acid response element-chloramphenicol acetyltransferase (*RARE*-CAT), with or without 3 μg of either empty pSG5 or pSG5-RARα or pSG5-RXRα. Four hours after transfection, the medium was changed to DMEM supplemented with 10% charcoal-stripped fetal bovine serum; after additional twenty hours, cells were treated with 1 μM retinoic acid (RA) or with vehicle (ethanol) alone. After 24 hr, cells were harvested and processed for CAT activity (Scafonas et al., 1997; Tairis et al., 1994). Normalization of CAT activity to β-galactosidase activity was done as above. For quantitation, the change in normalized CAT activity was calculated relative to that of wild-type cells that were transfected with empty vector DNA and treated with ethanol, which was set as 1. Values are the means ± standard deviations of a representative experiment performed in triplicate.

Analysis of DNA Methylation by Bisulfite Modification Sequencing

Genomic DNA was prepared from MEFs or mouse tissues by phenol-chloroform extraction, following standard procedures (Sambrook et al., 1989). For bisulfite sequencing analysis, genomic DNA was subjected to the sodium bisulfite modification reaction, as previously described (Howard et al., 2009). Products from the bisulfite reactions were amplified by PCR using primers, designed with the MethPrimer software at http://www.urogene.org/methprimer/, and mapping within 2 kb upstream of the transcriptional start site of the genes of interest. Typical PCR conditions consisted of an initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 50-54°C for 45 s and 72°C for 1 min. PCR products were purified using a PCR Purification Kit (QIAGEN). Purified products were either directly sequenced or subcloned into pGEM T-Easy vector (Invitrogen), in which case, individual inserts from 10-15 clones were sequenced with M13 (−20) reverse primer. Oligonucleotide sequences and details of the PCR conditions are available upon request. Comparisons of methylation levels were made using the two-sided Fisher's Exact Test at the 5% significance level.

Expression and Purification of Recombinant TDG, SMUG1, UNG2, and MBD4/MED1

Expression constructs of human TDG protein were prepared in the vector pET-FPHn, which provides a Flag-PKA-6-histidine tag, and were propagated in *E. coli* strain XL-1 Blue and expressed in Rosetta(DE3) (LacI) cells upon induction with 1 mM isopropyl-1-thio-®-D-galactopyranoside at 16°C for 16 hr. Cells collected by centrifugation were resuspended in 10 ml P200 Buffer (40 mM HEPES, pH 7.5, 200 mM NaCl, 10% glycerol, and "Complete" protease inhibitors (Roche Molecular Biochemicals) and sonicated in ice water for 4 min. After clarification by centrifugation at 10,000 rpm, the supernatant was diluted with 10 ml P0 Buffer (40 mM HEPES, pH 7.5, 10% glycerol) and imidazole was added to the sample to a final concentration of 30 mM. The sample was applied to a 1-ml HiTrap Chelate column charged with Ni^{2+} (Amersham Pharmacia Biotech). The column was washed in the following order with PI30 buffer (40 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol, 30 mM imidazole), P1000 buffer (40 mM HEPES, pH 7.5, 1 M NaCl, 10% glycerol), PI70 buffer (40 mM HEPES, pH 7.5, 70 mM NaCl, 10% glycerol, 70 mM imidazole), PI100 buffer (40 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol, 100 mM imidazole) and then eluted with PI300 buffer (40 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol, 300 mM imidazole). The PI300 eluate was diluted with 4 volumes of P0 buffer (40 mM HEPES, pH 7.5, 20 mM NaCl, 10% glycerol) and loaded onto 1-ml HiTrap Q column preequilibrated with P20 buffer. After wash with P50 (40 mM HEPES, pH 7.5, 50 mM NaCl, 10% glycerol) and P100 buffer (40 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol), the TDG protein was eluted with buffer P200 (40 mM HEPES, pH 7.5, 200 mM NaCl, 10% glycerol) and P400 (40 mM HEPES, pH 7.5, 400 mM NaCl, 10% glycerol). Throughout the purification, fractions were followed by SDS-polyacrylamide gel electrophoresis (PAGE) and assayed for glycosylase activity. The final TDG preparation was estimated to be > 95% pure by SDS-PAGE.

A similar approach was used for the purification of the human enzymes SMUG1 and UNG2. They were expressed from the pET28 (Novagen)-derived vectors by which a hexahistidine tag was attached at their N-terminus. Expressed proteins were purified by column chromatography with Ni Sepharose and Q Sepharose (GE Healthcare) for SMUG1, or Ni Sepharose and SP Sepharose for UNG2.

The purification of human MBD4/MED1 was conducted as previously described (Petronzelli et al., 2000a; Petronzelli et al., 2000b).

Microarray Analysis of Gene Expression

Total RNA was extracted from three independent wild-type and three independent *Tdg* null MEF cultures using guanidinium isothiocyanate. Biotinylated cRNA (15 μg), generated by using the One-Cycle Target Labeling and Control Reagent kit (Affymetrix) was fragmented, prehybridized for 10 min, and hybridized to MOE4302.0 Affymetrix chips for 16 hr, following the manufacturer's instructions, as previously described (Caretti et al., 2008). Arrays were washed, stained with anti-biotin antibody and streptavidin phycoerythrin, and placed inside a GeneChip Scanner 3000 (Affymetrix) for data acquisition.

For each sample, raw signal intensities from Affymetrix. CEL files were preprocessed using the Robust Multi-chip Average (RMA) method (Irizarry et al., 2003). Linear Models for Microarray Data (LIMMA) from the Bioconductor package (http://www.r-project.org, http://www.bioconductor.org) were applied to identify differentially expressed genes between wild-type and *Tdg* null MEFs. The Benjamini-Hochberg step-up method (Benjamini and Hochberg, 1995) was then applied to control the false discovery rate. Genetic networks were generated through the use of Ingenuity Pathways Analysis (Ingenuity Systems, http://www.ingenuity.com).

Microarray data were submitted to ArrayExpress with accession number E-MEXP-2610.

Chromatin Immunoprecipitation Assay

Chromatin immunoprecipitation (ChIP) was performed with antibodies anti-TDG (gift of Dr. K. Sugasawa), anti-pan-RAR (Santa Cruz), anti-p300 (Upstate) and anti-acetylated histone H3 (Upstate). Immunoprecipitations with non-specific immunoglobulins (Santa Cruz) were performed as negative controls. In each experiment, signal linearity was tested by amplifying increasing amounts of the DNA template. Generally, DNA representing 0.005%–0.01% of the total chromatin sample (input) or 1%–10% of the immunoprecipitate was amplified using promoter-specific primers. The following primer pairs were used: for *Crabp2*, GGCGATCCAGAGGCCCTTCT GAGAGGCA and CAGGAGCGGCCTCGGAGGCACGCGGGTA; for *Rbp1*, AGGTGAGAGTCCCTGGTGTTCAGC and GGGCCCTTT CCATACCTGTGCACGG. PCR conditions consisted of 1 cycle at 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 67°C (61°C for Rbp1) for 45 s and 72°C for 1 min.

Knock-down of *Tdg* and Analysis of Active DNA Demethylation

The *Tdg* shRNA lentivirus was prepared by co-transfecting 293T cells with *Tdg* shRNA pLKO1 (Open Biosystem), pLP1, pLP2 and pVSVG plasmids (Invitrogen) using lipofectamine 2000. Parallel transfections were conducted with scrambled shRNA and inactive *Tdg* shRNA pLKO1. Twelve hr after transfection, cells were incubated for 48 hr with Optimem media. The virus supernatant was filtered through 40 μm filter and applied onto 60% confluent P19 cells for 16 hr with 6 μg/ml of polybrene to facilitate infection. Both P19 and the infected P19 C8, C7 and scrambled shRNA cells were maintained in Alpha medium supplemented with 37.5% calf serum and 12.5% fetal bovine serum. After 48 hr, puromycin selection of infected cells was carried out for 2 weeks and then TDG expression was tested by western. P19 and P19 C8, C7 and scrambled shRNA cells were transfected with unmethylated *Oct4*::*EGFP* reporter plasmid or with the same plasmid methylated by incubation with SssI CpG methylase (New England Biolabs) in the presence of the methyl donor, S-adenosylmethionine. Cells were visualized for GFP expression 12 hr after transfection. For recovery of transfected plasmid, P19 and C8 cells were collected in PBS. Cell pellets were processed with QIAGEN Midikit according to manufacturer's procedure. The recovered DNA was dissolved in TE and then processed for bisulfite treatment.

Isolation of Primordial Germ Cells

Tdg heterozygous mice were intercrossed to generate embryos with different genotype. The pregnant female mice were sacrificed at E11 and the embryos were collected in PBS. After the removal of head and internal organs, the germinal ridges were dissected from the dorsal wall and harvested in PBS. The germinal ridges were minced mechanically by pipetting up-down through pipette tips and then digested with 0.4% trypsin at 37°C for 5 min. The trypsin was neutralized by adding 10% fetal bovine serum (FBS). The samples were pelleted, washed and resuspended in PBS containing 10% FBS. The samples were then incubated for 30 min in wet ice with anti-SSEA1 antibody (Millipore) diluted in PBS containing 10% FBS. After three washes with PBS, the samples were incubated for 30 min in ice with anti-Alexafluor 594 antibody (Molecular Probe). PGCs labeled with Alexafluor 594 were sorted by FACS. The morphology of the collected PGCs was evaluated after phosphatase alkaline staining (Millipore) to confirm cellular specificity.

Mutational Analysis of Promoters

The promoter region of *H19*, *Crabp2*, *HoxA5*, *Efs* and *Oct4* was amplified with the proofreading Expand High fidelity PCR system (Roche) on DNA extracted from wild-type and *Tdg* null MEFs and P19 and C8 cell lines, using the following Thermocycler conditions: 5 min at 94°C, 3 cycles at 94°C for 1 min, 58°C for 1 min, 72°C for 1 min, following by 32 cycles at 94°C for 45 s, 58°C for 30 s and 72°C for 30 s and ending with the final extension at 72°C for 10 min. The amplicons were subcloned into pGEM-T-easy vector, transformed in DH5α competent cells and the DNA extracted from the bacteria clones was used for sequencing. Oligonucleotide sequences are available upon request.

Methylated DNA Immunoprecipitation-Quantitative PCR

MeDIP was performed according to Mohn et al. (Mohn et al., 2008). Briefly, 3 μg of sonicated (300–1000 bp) genomic DNA were immunoprecipitated with 10 μg antibody against 5-methylcytosine (Eurogentec). For qPCR of the *Tat* enhancer, 20 ng sonicated genomic input DNA and 1/40 of MeDIP reaction were used with the primers AAATTTCCCCATGTCCAACA and GCTGGGATTTCAGGGACATA.

SUPPLEMENTAL REFERENCES

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proc. Natl. Acad. Sci. USA *96*, 3969–3974.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc., B *57*, 289–300.

Caretti, E., Devarajan, K., Coudry, R., Ross, E., Clapper, M.L., Cooper, H.S., and Bellacosa, A. (2008). Comparison of RNA amplification methods and chip platforms for microarray analysis of samples processed by laser capture microdissection. J. Cell. Biochem. *103*, 556–563.

Challberg, M.D., and Kelly, T.J., Jr. (1979). Adenovirus DNA replication in vitro. Proc. Natl. Acad. Sci. USA *76*, 655–659.

Cortellino, S., Turner, D., Masciullo, V., Schepis, F., Albino, D., Daniel, R., Skalka, A.M., Meropol, N.J., Alberti, C., Larue, L., and Bellacosa, A. (2003). The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. Proc. Natl. Acad. Sci. USA *100*, 15071–15076.

Holmes, J., Jr., Clark, S., and Modrich, P. (1990). Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines. Proc. Natl. Acad. Sci. USA *87*, 5837–5841.

Howard, J.H., Frolov, A., Tzeng, C.W., Stewart, A., Midzak, A., Majmundar, A., Godwin, A.K., Heslin, M.J., Bellacosa, A., and Arnoletti, J.P. (2009). Epigenetic downregulation of the DNA repair gene MED1/MBD4 in colorectal and ovarian cancer. Cancer Biol. Ther. *8*, 94–100.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. *31*, e15.

Lakso, M., Pichel, J.G., Gorman, J.R., Sauer, B., Okamoto, Y., Lee, E., Alt, F.W., and Westphal, H. (1996). Efficient in vivo manipulation of mouse genomic sequences at the zygote stage. Proc. Natl. Acad. Sci. USA *93*, 5860–5865.

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

Petronzelli, F., Riccio, A., Markham, G.D., Seeholzer, S.H., Genuardi, M., Karbowski, M., Yeung, A.T., Matsumoto, Y., and Bellacosa, A. (2000a). Investigation of the substrate spectrum of the human mismatch-specific DNA *N*-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. J. Cell. Physiol. *185*, 473–480.

Petronzelli, F., Riccio, A., Markham, G.D., Seeholzer, S.H., Stoerker, J., Genuardi, M., Yeung, A.T., Matsumoto, Y., and Bellacosa, A. (2000b). Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA *N*-glycosylase. J. Biol. Chem. *275*, 32422–32429.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: a Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Scafonas, A., Wolfgang, C.L., Gabriel, J.L., Soprano, K.J., and Soprano, D.R. (1997). Differential role of homologous positively charged amino acid residues for ligand binding in retinoic acid receptor alpha compared with retinoic acid receptor beta. J. Biol. Chem. *272*, 11244–11249.

Tairis, N., Gabriel, J.L., Gyda, M., III, Soprano, K.J., and Soprano, D.R. (1994). Arg269 and Lys220 of retinoic acid receptor-beta are important for the binding of retinoic acid. J. Biol. Chem. *269*, 19516–19522.

Turner, D.P., Cortellino, S., Schupp, J.E., Caretti, E., Loh, T., Kinsella, T.J., and Bellacosa, A. (2006). The DNA N-glycosylase MED1 exhibits preference for halogenated pyrimidines and is involved in the cytotoxicity of 5-iododeoxyuridine. Cancer Res. *66*, 7686–7693.

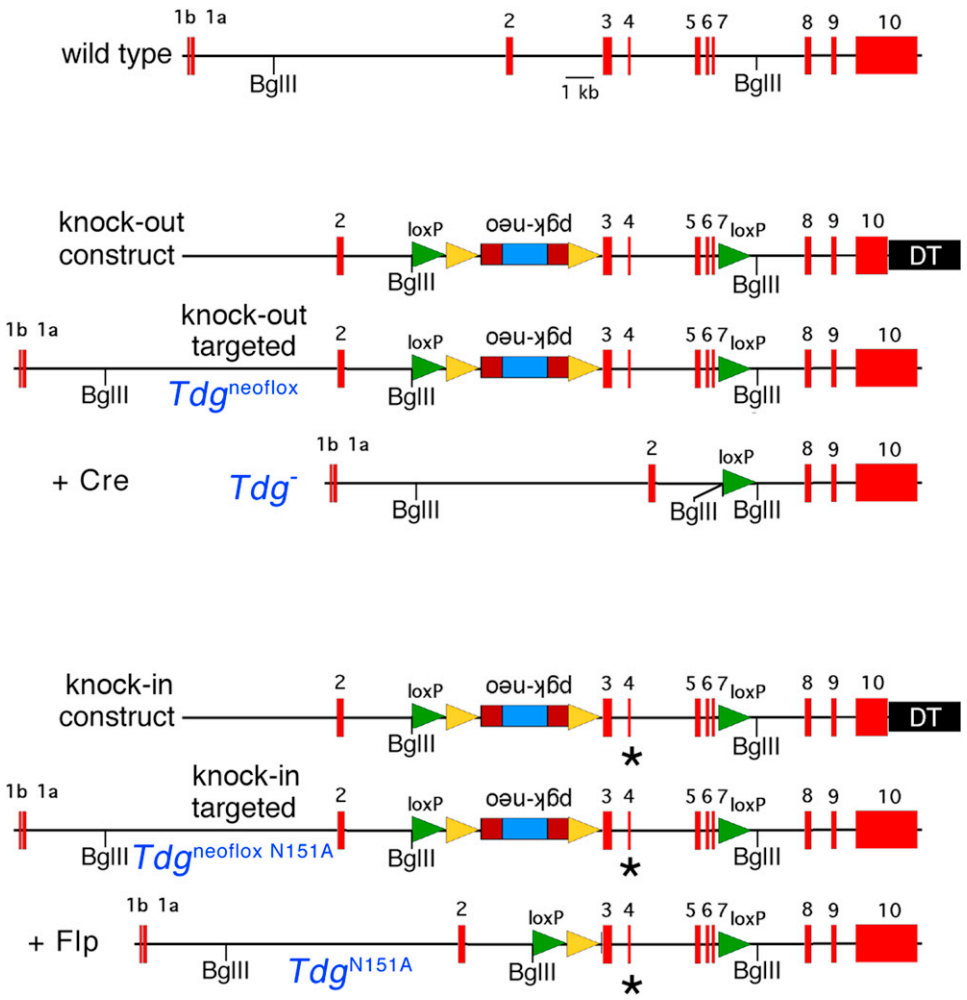

Figure S1. Strategy for Knockout and Knockin Inactivation of *Tdg* in the Murine Germline, Related to Figure 1 and Figure 5

In the *Tdg* knock-out targeting construct, a *pgk-neo* cassette flanked by *FRT* sites (yellow triangles) along with a *loxP* site (green triangle) is located between exons 2 and 3, while a *loxP* site is placed immediately downstream of exon 7; the *diphteria toxin* gene is marked as DT. The knock-out targeted recombinant $Tdg^{neoflox}$ allele was produced in ES cells after homologous recombination. In the wild-type allele, the BglII restriction sites are located approximately 3 and 2 kb downstream of exons 1a and 7, respectively. In the knock-out targeted $Tdg^{neoflox}$ allele, a new BglII site is engineered 2.5 kb downstream of exon 2. The null Tdg^{-} allele, in which Cre-mediated recombination between the *loxP* sites generates a deletion of exons 3 through 7, was produced by crossing $Tdg^{neoflox/+}$ mice with *EIIa-Cre* transgenic mice.

The *Tdg* knock-in targeting construct was generated by inserting the active site AAC > GGC mutation (asterisk) in *Tdg* exon 4, in a 2-kb fragment of the *Tdg* knock-out targeting construct. This fragment was sequenced and reinserted in the general targeting construct. The knock-in targeted recombinant $Tdg^{neoflox\ N151A}$ allele was produced in ES cells after homologous recombination. Mice bearing this targeted allele were bred with mice expressing the Flp recombinase at the Rosa26 locus (*Rosa26::FlpeR*), and the offspring were genotyped to score for Flp-mediated recombination of the *FRT* sites, with consequent removal of the *neo* cassette and generation of the knock-in Tdg^{N151A} allele.

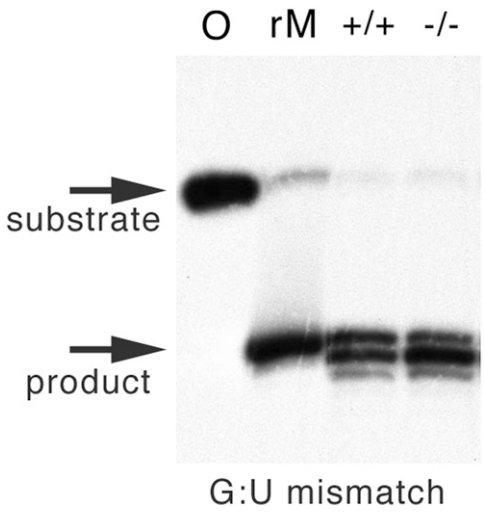

Figure S2. TDG Is Not Required for Efficient G:U Mismatch Repair at CpG Dinucleotides in MEFs, Related to Figure 1

Repair of a double-stranded oligonucleotide containing a G:U mismatch in a CpG context by nuclear extracts of MEFs with different genotypes. Reaction with recombinant MED1 (rM) was used as a size marker for cleavage at the mismatched uracil. One reaction received neither lysate not recombinant enzyme (0). Migration of the substrate and product bands is indicated.

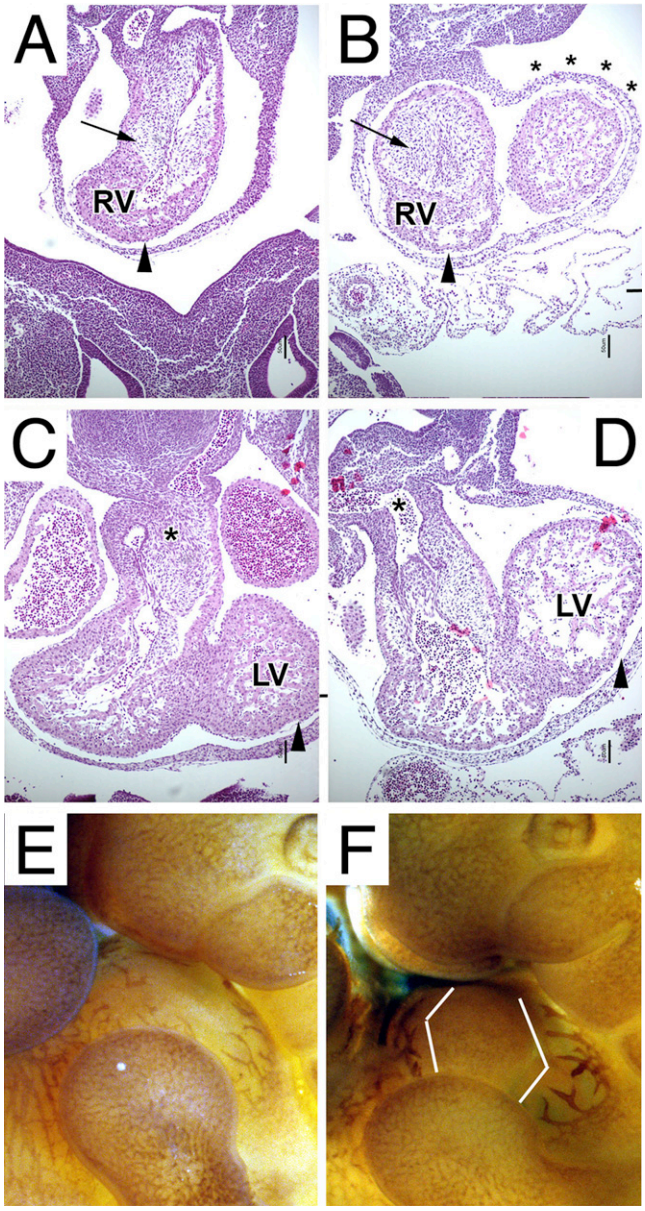

Figure S3. Additional Features of the Cardiovascular Phenotype in *Tdg* Null Embryos, Related to Figure 1

(A–D) The conal part (arrows in A and B) of the outflow tract (OFT) is severely hyperplastic and bulky in the mutant embryo (B) when compared to the wild-type specimen (A). The location of the OFT right above the right ventricle (RV) results in a relative shift of the left ventricle (LV) and a pronounced expansion of the left body wall (asterisks in B; see also Figure 1J). The septational complex in the distal OFT of the mutant does not form normally/completely (asterisks in D, compare with C). In addition to the OFT abnormalities, the RV is underdeveloped, with the myocardial wall thinner compared to that of wild-type littermates (arrowheads in B, compare to A); the overall morphology of this part of the hearts suggests that the RV chamber has not completely developed and that only the lower part of the RV has formed. The myocardial wall of the LV in mutant embryos also appears thinner than normal (arrowhead in D, compare to C).

(E and F) Immunostaining with a PECAM/CD31 antibody in wild-type (E) and *Tdg* null (F) embryos at E11 demonstrates, in mutant hearts, defective branching of the coronary arteries that are coarse and short, failing to reach the central portion of the epicardium (delimited by white brackets).

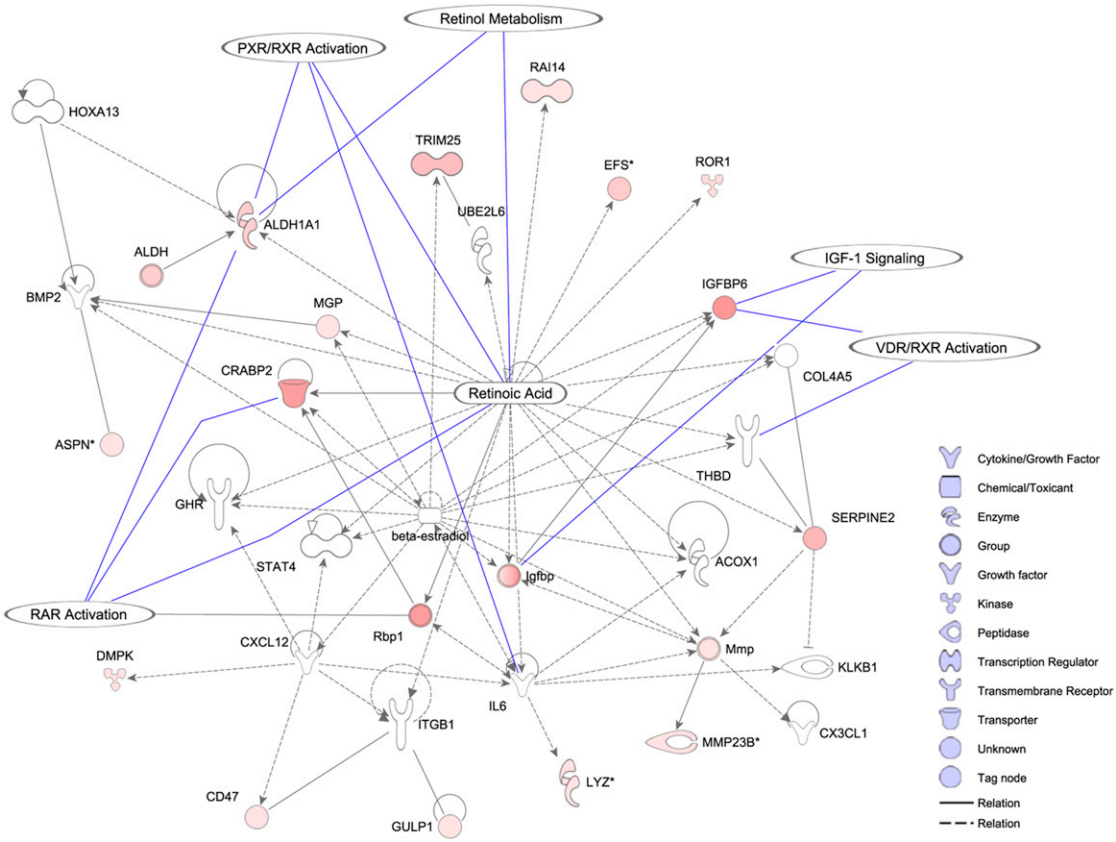

Figure S4. Reduced Retinol Metabolism and Retinoic Acid Receptor/Retinoid X Receptor-Dependent Transcription in *Tdg*$^{-/-}$ MEFs, Related to Figure 2 and Figure 3

Pathway analysis of differentially expressed genes between wild-type and *Tdg* mutant MEFs highlights reduced levels of retinol metabolism genes (*e.g. Aldh1a1*) and retinoic acid target genes (*e.g. Crabp2, Rbp1, Efs, Rai14*). Direct and indirect relationships are indicated by continuous or broken lines, respectively. Shades of red are proportional to the downregulation of the relevant mRNA (Table S2). Symbols refer to the function of the indicated proteins. General processes are circled with blue lines pointing to the relevant proteins.

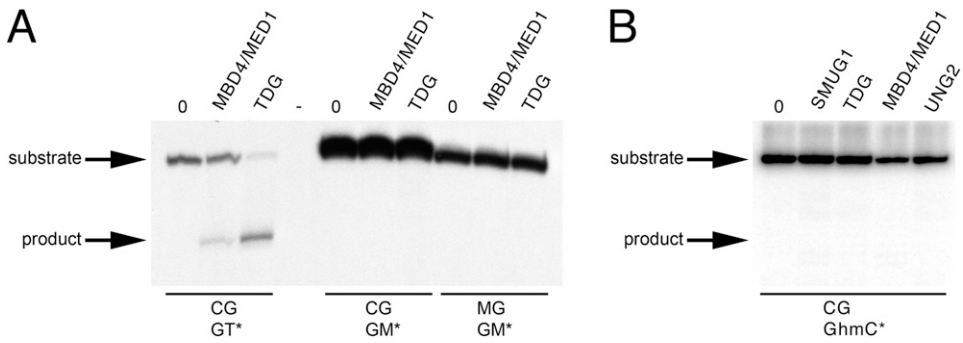

Figure S5. Recombinant TDG Lacks 5-Methylcytosine and 5-Hydroxymethylcytosine Glycosylase Activity, Related to Figure 6 and Figure 7

(A) Bacterially produced recombinant TDG was incubated at 37°C for 30 min with 37-mer double-stranded oligonucleotides ^{32}P-labeled on the bottom strand (marked by the asterisk) and bearing a G:T mismatch, or a hemimethylated or fully methylated CpG site, as indicated. The reactions were then treated with NaOH at 90°C for 20 min, in order to cleave the sugar-phosphate backbone at the AP site. A band representing a cleavage product was detected for the G:T-containing oligonucleotide substrate, but not for the hemimethylated or fully methylated CpG site. Arrows mark the expected migration of the substrate and product bands. Reactions with recombinant MED1 were used as additional size marker for cleavage at the mismatched thymine. Reactions that did not receive recombinant enzyme are indicated as (0).

(B) The indicated bacterially produced glycosylases were assayed as above for activity on a double-stranded oligonucleotide ^{32}P-labeled on the bottom strand (marked by the asterisk) and bearing a G:5-hmC pairing.

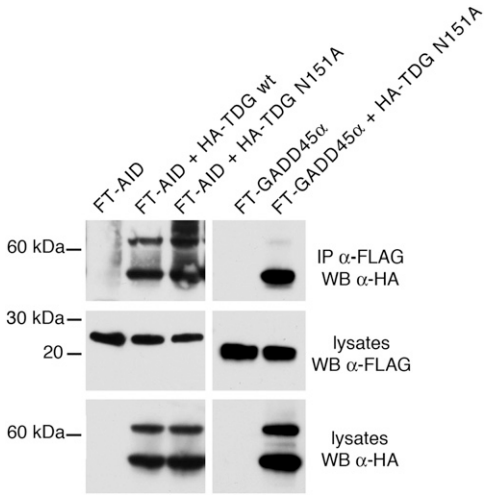

Figure S6. Catalytically Inactive TDG Forms a Complex with AID and GADD45a, Related to Figure 6

Co-immunoprecipitation analysis of lysates of HEK293 cells transfected with the indicated hemagglutinin (HA)-, and FLAG-tagged expression constructs reveals that the TDG^{N151A} protein interacts with AID and GADD45a.

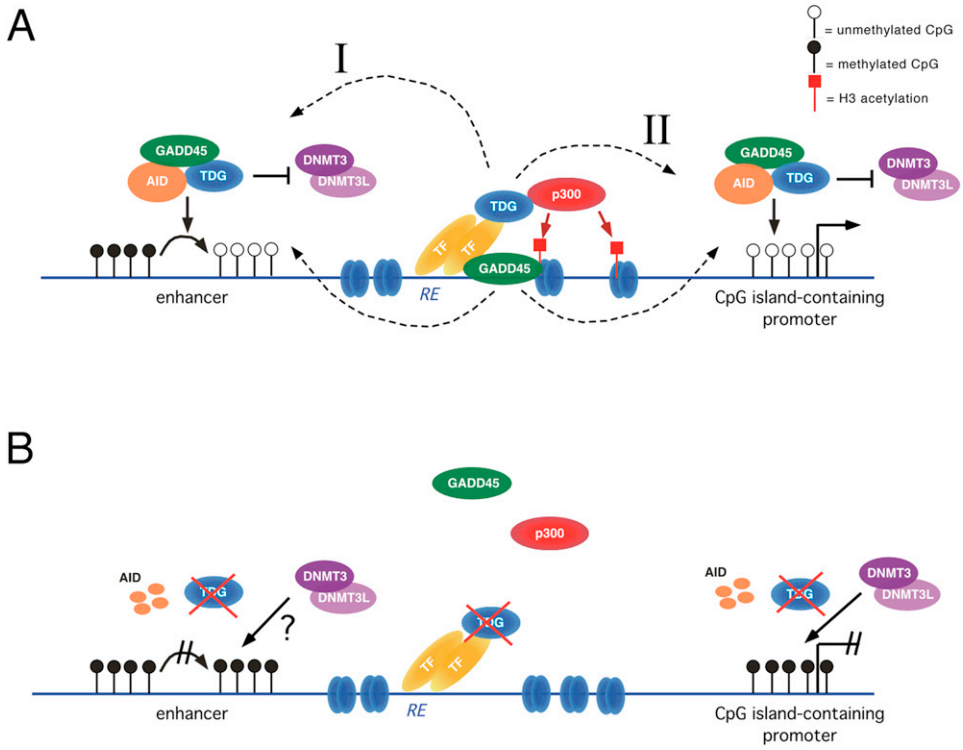

Figure S7. Model of the Dual Role of TDG in DNA Demethylation and Protection from Hypermethylation, Related to Figure 7

(A) In wild-type cells, binding to transcription factors (TF) at responsive elements (*RE*) tethers TDG and GADD45 to promoters; in turn, TDG promotes recruitment of p300 and the p300-acetylated nucleosomes (marked by red-squared lollipops) enhance GADD45 binding. The recruited TDG and GADD45 bind to AID and mediate DNA demethylation, by base excision repair, of CpG sites at enhancers, indicated by the white lollipops (function I, demethylation). The recruited TDG also helps maintaining CpG islands at promoters in their unmethylated state (function II, protection from hypermethylation). Both functions may be aided by TDG inhibition of de novo DNMTs.

(B) In the absence of TDG, p300 is not recruited, and the active demethylation of enhancers and protection of unmethylated promoters do not take place, leading to absent demethylation and hypermethylation, respectively. In addition, AID is destabilized and its targeting does not occur, preventing the accumulation of transition mutations at CpG sites.