nature chemical biology

ARTICLE

PUBLISHED ONLINE: 8 JULY 2012 | DOI: 10.1038/NCHEMBIO.1042

AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation

Christopher S Nabel[1,2], Huijue Jia[3], Yu Ye[4], Li Shen[3], Hana L Goldschmidt[4], James T Stivers[4], Yi Zhang[3]* & Rahul M Kohli[1,2]*

Activation-induced deaminase (AID)/APOBEC-family cytosine deaminases, known to function in diverse cellular processes from antibody diversification to mRNA editing, have also been implicated in DNA demethylation, a major process for transcriptional activation. Although oxidation-dependent pathways for demethylation have been described, pathways involving deamination of either 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC) have emerged as alternatives. Here we address the biochemical plausibility of deamination-coupled demethylation. We found that purified AID/APOBECs have substantially reduced activity on 5mC relative to cytosine, their canonical substrate, and no detectable deamination of 5hmC. This finding was explained by the reactivity of a series of modified substrates, where steric bulk was increasingly detrimental to deamination. Further, upon AID/APOBEC overexpression, the deamination product of 5hmC was undetectable in genomic DNA, whereas oxidation intermediates remained detectable. Our results indicate that the steric requirements for cytosine deamination are one intrinsic barrier to the proposed function of deaminases in DNA demethylation.

© 2012 Nature America, Inc. All rights reserved.

AID/APOBEC enzymes are well characterized for their ability to deaminate cytosine to uracil. In various cellular settings, this modification alters and expands the genome's coding potential[1,2]. In the immunoglobulin locus of the maturing B cell, AID deaminates cytosine to trigger pathways that facilitate antibody affinity maturation or isotype switching. Similarly, APOBEC3 enzymes extensively hypermutate reverse-transcribed DNA as a powerful innate defense mechanism to antagonize retroviruses and retroelements. In addition, APOBEC1 deaminates cytosine to introduce a stop codon in the mRNA of an apolipoprotein, generating a protein variant that differentially influences lipid metabolism. In each of these canonical functions, the purposeful mutations achieved by AID/APOBEC enzymes are generated by the deamination of a canonical cytosine nucleobase in DNA or RNA (**Fig. 1**). However, the role of modified cytosine bases in epigenetic regulation of gene expression raises the question as to whether these enzymes can also act on noncanonical 5-substituted cytosines[1,3,4].

The best-understood cytosine modification is 5mC, which exerts transcriptionally repressive effects that are integral to such processes as genomic imprinting, X-chromosome inactivation, cellular differentiation and response to signaling stimuli[5]. This epigenetic modification is achieved through the action of DNA methyltransferase enzymes in mammalian cells, which introduce and maintain the methylation mark on the 5 position of cytosine bases within cytosine-guanine dinucleotide (CpG) motifs. An added layer of chemical complexity has emerged with the recent observation that 5mC is oxidized to 5hmC by the TET family of Fe(II)-dependent α-ketoglutarate–dependent oxygenase enzymes[6,7], though the consequences of 5hmC on gene expression are only starting to be elucidated[8–10].

To achieve DNA demethylation, an important process for reversing cytosine methylation and restoring transcription, cytosine deamination, oxidation and base excision repair (BER) have been invoked in a variety of possible combinations (**Fig. 1**)[1,11,12]. In this regard, two types of deamination-dependent mechanisms have been postulated. In one scenario, deamination of 5mC by an AID/APOBEC enzyme generates a T-G mismatch leading to subsequent repair by the BER enzyme thymidine DNA glycosylase (TDG)[13]. In the other scenario, deamination of 5hmC generates 5-hydroxymethyluracil (5hmU), which could also be reverted to cytosine by BER[14]. Recent studies have also demonstrated the feasibility of a deamination-independent pathway for DNA demethylation involving oxidation of 5mC by TET enzymes. The product of this oxidation, 5hmC, undergoes iterative oxidation to yield both 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[15,16]. These cytosine analogs with higher oxidation states are detectable in the genome of embryonic stem cells and are good substrates for excision by TDG, which could ultimately regenerate unmodified cytosine[15–19]. Notably, deficiency in TDG, a potential common mediator in the various proposed pathways for DNA demethylation, is associated with developmental methylation defects and embryonic lethality[20,21].

The plausibility of deamination-dependent demethylation has been difficult to establish because of the poorly characterized activities of AID/APOBEC enzymes on C5-modified cytosines and a lack of knowledge about the functional redundancy between AID/APOBEC family members[22]. Although prior studies suggest that AID's ability to deaminate 5mC is reduced relative to its ability to deaminate cytosine[13,23], other work proposes that the enzyme lacks any 5mC deaminase activity[24]. Additional ambiguity arises because the activities of other APOBEC enzymes on 5mC have not been directly investigated, and the biochemical activities of all AID/APOBECs against 5hmC remain entirely unknown.

In general, the presence of numerous AID/APOBEC family members presents a challenge to sorting out their potential roles in demethylation. A potential role for AID in demethylation arises from heterokaryon-based systems for the generation of stem cells[25] and the suggestion that AID deficiency perturbs the DNA methylome of primordial germ cells[26]. However, these observations are confounded by the finding that AID deficiency is viable[27], suggesting that other deaminases might serve functionally redundant roles

[1]Department of Medicine, Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [2]Department of Biochemistry and Biophysics, Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [3]Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina, USA. [4]Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. *email: rkohli@upenn.edu or yi_zhang@med.unc.edu

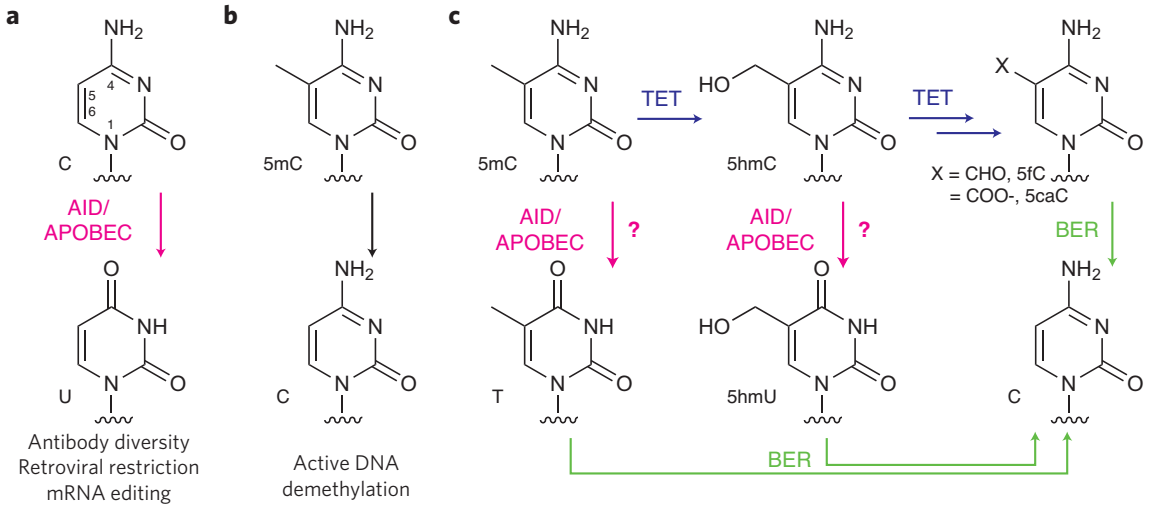


Figure 1 | Proposed noncanonical role for AID/APOBEC enzymes acting on modified cytosine substrates in DNA. (**a**) Deamination of cytosine has known physiological roles in adaptive immunity (AID), innate immunity against retroviruses (APOBEC3 enzymes) and mRNA editing (APOBEC1). These canonical roles involve deamination of cytosine to generate uracil. (**b**) By contrast, the proposed function of AID/APOBEC family members on modified cytosine residues remains poorly understood despite their implication in potential pathways for active DNA demethylation. (**c**) Proposed pathways for DNA demethylation. Deamination of 5mC or 5hmC, the product of TET-mediated oxidation, could generate thymidine or 5hmU, respectively. BER could subsequently excise the deaminated bases and replace them with unmodified cytosine. An alternative deamination-independent pathway involves iterative oxidation, generating 5fC or 5caC. BER-mediated excision of the oxidized cytosine would result in reversion to unmodified cytosine.

© 2012 Nature America, Inc. All rights reserved.

in demethylation. In support of this proposal, APOBEC2 enzymes are postulated to have a role in zebrafish DNA demethylation[28], and APOBEC1 is implicated in neuronal DNA demethylation[14]. Thus, biochemical characterization of the similarities and differences between these deaminases would address the functional redundancy of these enzymes in DNA demethylation.

The previous implications that deaminases might be involved in DNA demethylation[1,3,4] make it important to examine their activity on 5-substituted cytosine bases in DNA. Here we elucidate the substrate preferences of AID/APOBEC family members and examine the plausibility of deamination-dependent versus oxidation-dependent pathways for cytosine demethylation. We found that all AID/APOBEC family members preferentially deaminate unmodified cytosine and strongly discriminate against 5-substituted cytosine substrates with increasing steric bulk. We also observed these preferences in cells, where the deamination products of the bulky base 5hmC are not detected when deaminases are overexpressed. Further, intermediates in the iterative oxidation pathway were readily found in genomic DNA, and their formation was not diminished by the overexpression of deaminases. We argue that, relative to the oxidation-mediated pathway, the potential for deaminases to participate in DNA demethylation is limited.

RESULTS

AID/APOBECs deaminate unmodified cytosine

We wished to profile the reactivity of representative AID/APOBEC family members with modified cytosine nucleobases. We chose to investigate the mouse enzyme family, which has only a single gene for each family member (**Supplementary Results**, **Supplementary Fig. 1a**), rather than the human family, where extensive gene duplication and specialization at the A3 locus have generated at least seven variants named A3A–A3H. Mouse APOBEC1 (mA1), APOBEC2 (mA2), APOBEC3 (mA3) and AID (mAID) were generated as N-terminal maltose-binding protein fusion constructs. Although mAID was inactive under these conditions (data not shown), we had previously expressed and characterized active human AID (hAID) by expression of the enzyme together with the chaperone trigger factor in *Escherichia coli*[29]. Using this expression system, the expanded cohort of AID/APOBEC enzymes were all soluble and were partially purified over amylose resin (**Supplementary Methods**, **Supplementary Fig. 1b**).

We designed DNA oligonucleotides containing a single cytosine residue with a few criteria in mind. First, because each AID/APOBEC family member prefers to deaminate cytosine in a different trinucleotide sequence context[29–31], we selected a universal sequence that would be acted upon by multiple family members (TGC). Next, a guanine was introduced directly downstream of the cytosine to create a CpG motif, an important consideration given that epigenetic modifications via methylation, hydroxymethylation and demethylation are highly linked to CpG sites and islands in the mammalian genome. All oligonucleotides were synthesized using standard phosphoramidite chemistry (**Supplementary Methods**).

AID/APOBEC family members were assayed against the cytosine-containing substrate S30-TGC using a discontinuous, uracil DNA glycosylase (UDG)-coupled assay (**Fig. 2a**). At the end of the deamination period, the reaction product was hybridized to a complementary strand, yielding duplexed DNA containing a U-G mismatch in the deaminated product. Treatment with UDG generated abasic sites in the deaminated oligonucleotides and left unreacted substrates intact. Cleavage of the abasic sites under alkaline conditions allowed for specific detection of product after separation on a denaturing gel. Under these conditions, the mA1, mA3 and hAID variants all showed deaminase activity against S30-TGC (**Fig. 2b** and **Supplementary Fig. 2**). As anticipated from prior studies, mA2 showed no detectable cytosine deaminase activity[32].

To next examine the deamination activities on physiologically relevant 5-modified cytosines, we synthesized S30-TG5mC and S30-TG5hmC. Upon deamination, these target bases would convert to T and 5hmU, respectively. Alternate DNA glycosylases were used to detect these species: TDG, which excises thymidine from T-G mismatches in CpG contexts, and human SMUG1 (SMUG), which excises 5hmU mispaired to G (**Fig. 2a**). We established that these glycosylase-coupled assays accurately detect small amounts of deamination products (as little as 0.5% product under the condition of our deamination reaction; **Supplementary Fig. 3**). Using

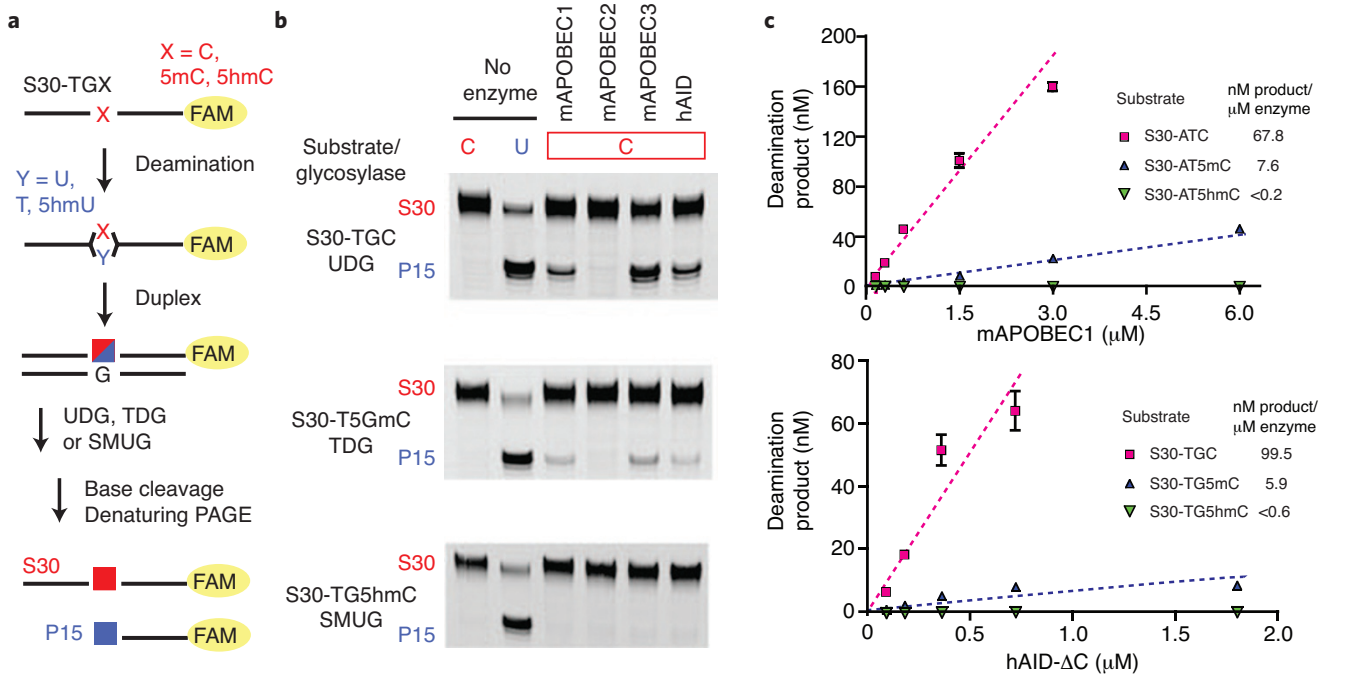


Figure 2 | AID/APOBEC enzymes preferentially deaminate unmodified cytosine. (**a**) Fluorophore (FAM)-labeled 30-mer oligonucleotides (S30) synthesized with a single internal modified cytosine (red) embedded in a CpG motif were incubated with AID/APOBEC family members. After incubation, oligonucleotides were duplexed with a complementary strand generating U-G, T-G or 5hmU-G mismatches with the deaminated substrates (blue). Treatment with UDG (reactive with U-G), TDG (reactive with T-G) or SMUG (reactive with 5hmU-G), followed by base-mediated cleavage, fragments the deaminated products to a 15-mer (P15). (**b**) The reaction products resulting from incubation of S30-TGX substrates with mA1, mA2, mA3 or hAID are shown separated on a denaturing gel. The substrate and product controls, without incubation with AID/APOBEC enzymes, are shown on the left. Uncropped gels are shown in **Supplementary Figure 2**. (**c**) The fraction of deaminated substrates are plotted as a function of increasing concentrations of enzyme for mA1 and hAID-ΔC assayed against substrates containing a cytosine (magenta), 5mC (blue) or 5hmC (green). hAID-ΔC was incubated with the S30-TGX series of substrates, whereas mA1 was incubated with the S30-ATX substrates. Error bars represent s.d. from the mean of at least three independent replicates. For relative comparison of substrates, the slope of each plot in the region where product formation is linear with enzyme (dashed line) is listed. The standard error for the measurement of enzyme dependent product formation is ±20% of the reported value for all measurements. The values given for 5hmC substrates are based on the lower limits of detection with these substrates.

© 2012 Nature America, Inc. All rights reserved.

the TDG-coupled assay, deamination of S30-TG5mC was detectable with all active AID/APOBEC enzymes, though product formation was decreased relative to S30-TGC (**Fig. 2b**). By contrast, no detectable deamination activity was evident against S30-TG5hmC, despite robust deamination of S30-TGC under identical conditions with mA1, mA3 and hAID. mA2 was also inactive against both S30-TG5mC and S30-TG5hmC *in vitro*.

To assure that the discrimination against 5-substituted cytosine bases was not simply limited to a single sequence context, or perhaps the single enzyme concentration used in the above studies, we performed additional investigations. First, we tested a series of substrates containing C, 5mC and 5hmC in an ATX trinucleotide context (where X represents the cytosine base variant), which is a preferred sequence for mA1. As expected, reaction of S30-ATC with mA1 led to robust deamination, whereas S30-AT5mC was compromised, and deamination of S30-AT5hmC was undetectable (**Fig. 2c**). With varying concentrations of mA1, deamination was linearly dependent on the amount of enzyme used. On the basis of our detection limits, mA1 was estimated to have ~10-fold discrimination against 5mC relative to unmodified C and >300-fold discrimination against 5hmC. We conducted a similar enzyme-dependent analysis of deamination of S30-TGX substrates with a C-terminal truncation variant of hAID (hAID-ΔC), which is associated with hyperactive deamination (approximately three-fold) without affecting sequence-dependent targeting[29]. We reasoned that low deamination might be easier to detect with this hyperactive variant. As with mA1, we found a similar degree of discrimination against 5mC (~16-fold) and substantial discrimination against 5hmC (>150-fold; **Fig. 2c**). A quantitatively similar pattern can be observed with full-length hAID (**Supplementary Fig. 4**). Finally, mA3 also demonstrated clear discrimination against the naturally modified cytosine nucleobases (**Supplementary Fig. 4**). We conclude that discrimination against 5-substituted cytosines is an intrinsic property evident in the entire AID/APOBEC enzyme family regardless of the source organism or canonical function.

Deamination decreases with increasing steric bulk at C5

The molecular basis for recognition of cytosine by AID/APOBEC enzymes is a matter of speculation given the lack of structural information on these enzymes complexed with nucleic acid substrates. To probe the molecular impact of substitution at the 5 position of cytosine, we synthesized additional substrates with unnatural C5 substituents of varied steric and electronic character (**Fig. 3a**) using a sequence context appropriate for hAID (S30-TGX) or mA1 (S30-ATX). One potential determinant of reactivity in this series is the electron-withdrawing ability of the C5 substituents. Electronegative C5 groups could potentially enhance deamination by making the C4 of cytosine more electrophilic or by altering the pK_a of N3. Alternatively, the hydrophobicity of the 5-position substituent could influence selectivity[33]. Finally, the size of the substituent at the 5 position could dominate the rate effect.

To study deamination of these unnatural cytosines, we first determined the specificity of UDG, TDG and SMUG against DNA substrates containing either the modified cytosine or the corresponding 5-substituted uracil. On the basis of the determined substrate preferences, we selected UDG to assay the 5-fluoro (5F)

substrates and SMUG for the 5-hydroxy (5OH), 5-bromo (5Br) and 5-iodo (5I) substrates (**Supplementary Fig. 5**). We also constructed standard curves to allow for accurate quantification of deamination of each unnatural 5-modified substrate using our glycosylase-coupled assay (**Supplementary Fig. 6**).

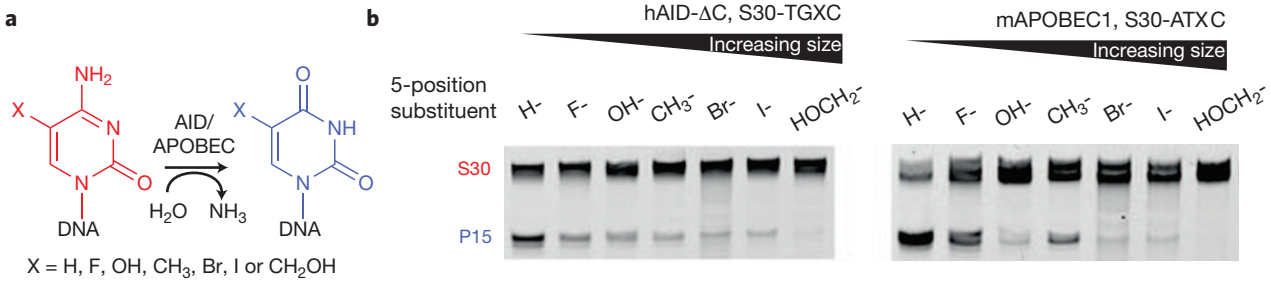
Each 5-substituted S30-TGX substrate was incubated with hAID-ΔC and then duplexed and treated with the appropriate glycosylase to assay for deamination. For hAID-ΔC, any substitution resulted in decreased efficiency of deamination relative to unmodified cytosine (**Fig. 3b** and **Supplementary Fig. 7**). To allow for relative comparison of substrates, we next calculated the product formation under conditions where deamination was linearly proportional to enzyme concentration. Across this series of substrates, representing a >150-fold difference in reactivity, the size of the substituent at the 5 position seemed to be an important determinant of deamination (**Fig. 3c** and **Supplementary Fig. 8**). The smallest unnatural substituent, 5-fluorocytosine, was deaminated most readily, although it remained half as reactive as unmodified cytosine. Bulkier halogen substituents were relatively poor substrates compared to the smaller 5mC. In addition to the influence of sterics, the poor hydrophobic character of 5hmC may have an additional role in the reactivity decrease observed between (5I)C, which has detectable deamination, and 5hmC, which has no detectable deamination.

To assess the generality of reactivity determinants in the deaminase family, we additionally profiled mA1 against a series of unnatural substrates in its preferred S30-ATX context (**Fig. 3**). Again, 5-fluorocytosine was a good substrate for deamination, though it was approximately only one-quarter as reactive as cytosine, whereas bulkier substituents were increasingly poor substrates (**Fig. 3c**). Though smaller size remained a prerequisite for efficient deamination with mA1, hydrophobicity also seems to have a key role. This role is most notable with (5OH)C, which undergoes negligible deamination, whereas the larger 5mC is readily deaminated. Our extensive data show that for two distinct, active AID/APOBEC family enzymes, efficient deamination of cytosine has steric requirements at the 5 position that contribute to the lack of any detectable enzymatic activity on 5hmC.

Deaminases do not alter amounts of epigenetic bases

Among the multiple potential pathways for DNA demethylation is the possibility of collaboration between oxidation and deamination to generate 5hmU from 5mC[14]. To complement our *in vitro* findings, we addressed whether 5hmC deamination could be detected in genomic DNA. Accordingly, we overexpressed one isoform of the TET oxidase family, TET2, in HEK 293T cells, resulting in a high prevalence of genomic 5hmC in a genome otherwise devoid of detectable oxidized cytosine bases[15]. TDG and the individual AID/

© 2012 Nature America, Inc. All rights reserved.



Nucleobase	C	(5F)C	(5OH)C	5mC	(5Br)C	(5I)C	5hmC
Size* (Å^3)	10.5	16.8	19.4	30.5	33.0	37.5	40.8
Hydrophobicity**	0.00	0.14	–0.67	0.56	0.86	1.12	–1.03
hAID†	35.8 (1.0)	–	–	4.4 (0.12)	–	–	<1 (<0.04)
hAID-ΔC†	99.5 (1.0)	50.6 (0.51)	15.1 (0.15)	5.9 (0.06)	9.7 (0.10)	7.6 (0.08)	<0.6 (<0.006)
mAPOBEC1†	67.8 (1.0)	17.8 (0.26)	0.3 (0.005)	7.6 (0.11)	1.0 (0.015)	1.0 (0.015)	<0.2 (<0.003)
mAPOBEC3†	37.6 (1.0)	–	–	6.8 (0.18)	–	–	<0.3 (<0.008)

Figure 3 | DNA deamination decreases as a function of increasing steric bulk at the 5 position of cytosine. (**a**) Schematic of the deamination assay with DNA oligonucleotides containing unnatural modifications at C5. (**b**) At left, the denaturing gel for hAID-ΔC with S30-TGX substrate series is shown. At right, the denaturing gel with mAPOBEC1 assayed against S30-ATX substrates. Samples were run in order of increasing steric bulk of the 5-position substituent. Gels are shown without cropping in **Supplementary Figure 7.** (**c**) Shown are the electrostatic potential maps of each of the modified cytosine bases as determined using the SPARTAN program (6-31G* basis set). The electrostatic potential is colored from maximal negative (red) to positive (blue). The volume (*) is determined by linking the 5-position substituent to a single hydrogen atom and calculating the total volume. The hydrophobic substituent constant (**) is derived from partitioning studies of substituted benzenes between octanol and water, and negative values for hydroxy and hydroxymethyl substituents represent less hydrophobicity[33]. Reported are the values for enzyme-dependent product formation (nM product/μM enzyme) for each substrate examined with each active deaminases (†). The data for C, 5mC and 5hmC deamination are consolidated from **Figure 2c** and **Supplementary Figure 4**; data for unnaturally modified substrates are based on plots in **Supplementary Figure 8**. The relative activity for each modified substrate when compared to unmodified cytosine for each deaminase is reported parenthetically, allowing for comparisons across a row with each enzyme.

APOBEC family members were transfected together into HEK 293T cells along with TET2 to determine their impact on the amount of modified nucleoside. Similar degrees of expression for each enzyme were confirmed by western blot (**Supplementary Fig. 9**). Using methodology we previously used to detect the products of iterative oxidation by TET[15], the genomic DNA was isolated and digested to generate a genomic nucleoside pool. For highly sensitive detection of the modified bases, the obtained nucleosides were subjected to LC-MS/MS in multiple reaction monitoring mode and quantified by comparison to known standards, including that of genomic cytosine, to allow for approximation of the genomic prevalence of modified bases. This methodology enables detection of 10 fmol of 5hmU from 0.75 μg of genomic DNA, which converts to a detection limit of approximately 2–3 5hmU bases in 10^5 dC bases (**Supplementary Figs. 10** and **11**).

As previously observed, TET2 overexpression in isolation leads to 5hmC amounts that are about one-sixth that of genomic 5mC (**Fig. 4a**). The products of iterative oxidation, 5fC and 5caC, can also be detected at ~1% the amount of 5mC[15,16]. We posited that if AID/APOBEC enzymes could deaminate 5hmC, three changes should be observed: 5hmU should be detected in genomic DNA, the amount of 5hmC should decrease and the amount of 5fC and 5caC should also decrease owing to the deamination of their precursor 5hmC. However, none of these three predicted changes were observed when TET2 was overexpressed along with either mA1 or hAID (**Fig. 4a**). We conclude that in cells overexpressing TET2 and hAID or mA1, the genome contains less than 2–3 5hmU bases per 10^5 dC bases. To further examine whether other family members might influence the genomic levels of modified bases, we screened all mouse AID/APOBEC enzymes with similar results (**Supplementary Fig. 11**). To determine whether our inability to detect 5hmU could be due to its rapid enzymatic excision from the genome, we measured the 5hmU glycosylase activity from HEK 293T nuclear extracts. We observed negligible 5hmU glycosylase activity under conditions where robust nuclear uracil glycosylase activity was observed (**Supplementary Fig. 12**). Together these data suggest that if deamination of 5hmC occurs, it falls below the detection limits of this sensitive methodology.

To confirm that the overexpressed deaminases were active despite the lack of detectable genomic 5hmU, we first examined nuclear lysates for deaminase activity using oligonucleotides containing unmodified cytosine. In line with our biochemical studies, deamination of cytosine was only detectable in nuclear lysates from cells overexpressing either hAID or mA1, but not a catalytic mutant of hAID (hAID^{E58A}) (**Fig. 4b** and **Supplementary Fig. 13**). As a second validation of deaminase activity in the cellular setting, we examined whether overexpressed deaminases have an appreciable impact on the amount of genomic deoxyuridine (dU). To overcome the rapid and efficient processing of genomic dU by multiple DNA repair pathways, we inhibited the major pathway involving UDG by overexpressing the small-protein uracil DNA glycosylase inhibitor (UGI) along with hAID, hAID^{E58A}, mA1 or an empty plasmid control (**Supplementary Fig. 14**). The nuclear lysates associated with UGI overexpression were unable to excise uracil from duplexed oligonucleotides, confirming that UGI effectively inhibited the major pathway for uracil excision. Notably, this observation also independently suggested low TDG and SMUG activity in 293T cells, as low uracil excision was observed despite the fact that these enzymes are not inhibited by UGI[34]. In these cells, we next quantified the amount of genomic dU using our highly sensitive LC-MS/MS methodology and demonstrated a consistent increase in genomic dU in hAID samples relative to that in hAID^{E58A} or the empty plasmid control. Unlike for hAID, it is unknown whether mA1 can act upon genomic cytosine. In our analysis, mA1 overexpression also yielded increases in genomic dU compared to overexpression of an empty vector control. Thus, by two measures, in lysates and through analysis of genomic DNA, the deaminases were active on unmodified cytosine under conditions where no deamination of 5hmC was detectable.

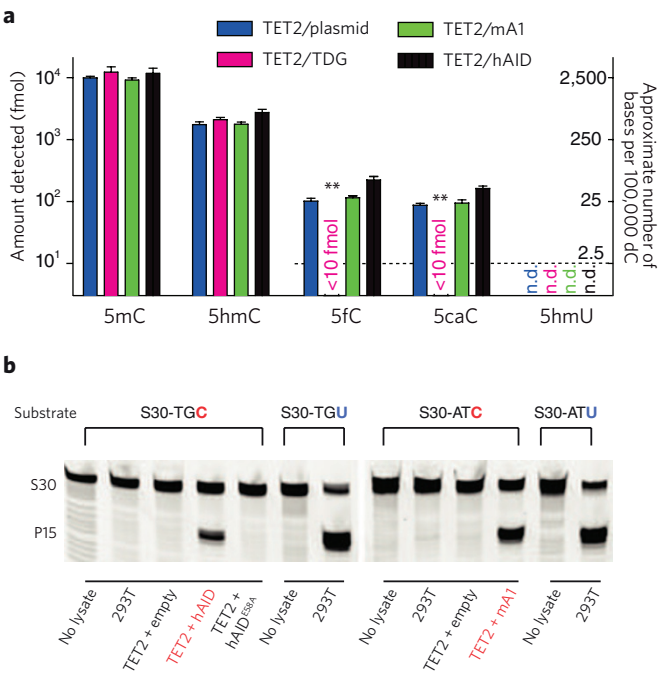


Figure 4 | AID/APOBEC enzymes do not perturb levels of 5mC oxidation intermediates in genomic DNA. (**a**) Genomic DNA was extracted from HEK 293T cells expressing TET2 along with an empty vector control, TDG, mA1 or hAID. The genomic DNA was then digested to single nucleosides and analyzed via MS for the presence of C, 5mC, 5hmC, 5fC, 5caC and 5hmU. The left axis depicts the absolute fmol of nucleoside on a logarithmic scale, and the right axis represents the approximate conversion from absolute fmol to the genomic prevalence of modified bases for every 10^5 dC bases. The dashed line demonstrates the lowest examined amount of 5fC, 5caC and 5hmU standards. Error bars indicate the s.d. from the mean for two to three biological replicates. Asterisks: $P \leq 10^{-3}$ for 5fC and 5caC in samples with TDG in comparison to plasmid-only control. (**b**) To confirm that overexpressed hAID and mA1 are catalytically active, nuclear extracts were tested for deaminase activity against the unmodified cytosine. At left, TET2-hAID nuclear extracts and negative controls (no extract, nuclear extract from untransfected 293T cells and TET2-hAID^{E58A}) were incubated with S30-TGC substrate; at right, TET2-mA1 nuclear extracts and negative controls (no extract and nuclear extract from untransfected 293T cells) were incubated with S30-ATC substrate. The lanes demonstrating deamination of S30-TGC by TET2-hAID and S30-ATC substrate by TET2-mA1 extracts are highlighted (red). Nuclear extracts were also incubated with S30-TGU or S30-ATU to verify robust uracil excision activity. Uncropped gels are shown in **Supplementary Figure 13**.

By contrast to our results with overexpression of AID/APOBEC family members, we reasoned that if iterative oxidation coupled to BER is a feasible pathway for demethylation, the amounts of modified cytosine bases should be readily perturbed by coexpression of TET and TDG. Indeed, we found that TDG overexpression led to substantial reductions in the highly oxidized 5fC and 5caC species (**Fig. 4a**). Notably, no overall change was observed in the amounts of 5mC and 5hmC. Although this observation could be explained by the sparsity of highly oxidized species relative to 5mC and 5hmC, it also suggests that TDG-mediated depletion of 5fC and 5caC does not promote further oxidation and consumption of genomic 5hmC. Potential explanations include the possibility that some genomic 5hmC is sheltered from further oxidation or has roles independent of demethylation[10]. Though the dynamics of iterative oxidation will require further intensive study, from our data we conclude

© 2012 Nature America, Inc. All rights reserved.

that the global prevalence of modified cytosine nucleobases in the genome can be altered by overexpression of the members of the deamination-independent pathway but not by those of the deamination-dependent pathway.

DISCUSSION

In all of the well-established roles whereby DNA deaminase enzymes modulate the genome, the enzymes act upon unmodified cytosine: in antibody diversity (AID), retroviral restriction (A3) and RNA editing (A1). Here, we have considered the possibility that deaminases could also have a role in DNA demethylation via deamination of 5mC or 5hmC by assessing the ability of AID/APOBEC enzymes to deaminate modified forms of cytosine. Our observation that the steric bulk of the 5-position substituent decreased the efficiency of deamination presents a mechanistic rationale for discrimination against 5mC and 5hmC. Although these enzymes have diverged to assume distinctive cellular functions, the consistent pattern of substrate selectivity across the AID/APOBEC family indicates a conserved active site architecture that is primarily tuned for the recognition and deamination of unmodified cytosine.

Both steric and electronic effects of pyrimidine C5 substitutions are known to affect various enzymes that act on DNA. For the BER enzyme TDG, electron-withdrawing effects from the C5 substituent promote nucleobase hydrolysis by weakening the N-glycosidic bond[35]. This finding fits well with TDG's epigenetic role in removing 5fC and 5caC from DNA because these bases have substituents that activate cleavage of the glycosidic bond[17]. By contrast, steric exclusion of C5 substituents is important for the BER glycosylase UDG, which uses a tyrosine side chain to selectively exclude thymidine and excise uracil with high proficiency[36]. It will be informative to see whether structural studies on the AID/APOBEC enzymes in complex with DNA will demonstrate a similar molecular mechanism for discrimination against bulky 5-position substituents on cytosine.

We can now reconcile the known characteristics of AID/APOBEC enzymes with their proposed function in DNA demethylation by examining the multiple levels of targeting involved in deamination. At the level of the nucleobase, our results suggest that all deaminase enzymes have evolved an active site that is designed to deaminate unmodified cytosine preferentially. However, deamination of 5mC to T can occur, albeit at ~10% reduced rate relative to cytosine deamination. Thus, strictly from the perspective of biochemical feasibility, deamination of 5mC may constitute a viable pathway for demethylation in some situations, though other constraints are important to consider (described below). By contrast, we demonstrated that deamination of 5hmC was detectable neither *in vitro* nor in cells when relevant enzymes were overexpressed. Our results contrast with those of a prior study that used immunoblotting of DNA[14], a method of uncertain specificity, to report detection of genomic hmU[37]. However, our findings are in good agreement with studies on embryonic stem cells, where 5hmU is not detectable in genomic DNA when probed with a sensitive and specific MS methodology[18,38] and with results from other groups who have not detected deamination of 5hmC by AID *in vitro* (S. Petersen-Mahrt, IFOM Fondazione Istituto FIRC di Oncologia Molecolare, personal communication). Together, our biochemical and cellular data provide a strong argument against the proposed collaboration between oxidation, deamination and BER as a pathway for DNA demethylation in mammalian cells.

Beyond the target cytosine, the local sequence context provides an additional potential barrier to efficient deamination. Each AID/APOBEC family member acts at preferred trinucleotide hotspots[29–31,39] (**Supplementary Fig. 1a**), yet methylated CpG motifs can be found in all common sequence contexts, even those that may be disfavored by the individual deaminases. Further, as all known AID/APOBEC enzymes are specific for single-stranded DNA[23,40], it remains unclear how methylated CpG motifs in genomic DNA might be sufficiently targeted by AID/APOBECs. Although transcription or replication could generate single-stranded DNA, active demethylation of CpG islands does not always require replication or transcription[41,42]. Further, only AID and human APOBEC3A have been shown to deaminate mammalian host genomic DNA[43,44], and these mutations seem to be localized to expressed genes. Finally, expression of the various deaminase family members is restricted to particular cell lineages, suggesting that a single deaminase is unlikely to have a universal role in demethylation.

Integration across these layers of targeting, from the nucleobase to the cellular level, enables us to assess the plausibility of the early steps in the various proposed DNA demethylation pathways. First, our findings show that deamination of 5mC remains a plausible demethylation pathway based on enzymatic function, although other known limitations in targeting seem to present several barriers to its efficient function. It is possible that deamination of 5mC operates in nonphysiological systems, such as heterokaryon-based reprogramming[25]. In other physiological settings, it will be important to examine whether these enzymes may have roles in particular niches[21]. Although our findings exclude 5hmC deamination as a detectable enzymatic activity, a requirement for AID/APOBEC enzymes in the conversion of 5hmC to C in neurons has been suggested in ref. 14. Notably, in that study, AID/APOBEC enzymes were not shown to be directly capable of 5hmC deamination, and a requirement for catalytically active AID in conversion of 5hmC to C was also not established. If catalysis were to be required, it is feasible that AID/APOBEC-mediated deamination of C or 5mC indirectly stimulates TET or DNA damage response pathways to promote excision of genomic 5hmC. Our biochemical and cellular data make it unlikely that deamination of 5hmC by AID/APOBEC enzymes is involved in DNA demethylation. Finally, our data suggest that under the very same conditions where the product of 5hmC deamination (5hmU) is below detectable limits, the proposed intermediates in the iterative-oxidation pathway, 5fC and 5caC, are readily detected. As we have addressed for the AID/APOBEC enzymes, this work raises the issue of targeting by the TET enzymes. From their relative preferences for oxidation of 5mC, 5hmC and 5fC to their regulated expression in the appropriate cellular settings, much remains to be understood. These topics form an important area of focus for further explorations on the mechanism of DNA demethylation.

METHODS

AID/APOBEC protein expression and purification. Expression vectors contained the genes encoding mA1, mA2 and mA3 downstream of an N-terminal maltose-binding protein in pET41 (Novagen), as previously described for hAID[14,29]. AID/APOBEC constructs were expressed and purified as described in **Supplementary Methods**.

Deamination assays. Synthesis of oligonucleotides is detailed in **Supplementary Methods**. Deamination reactions were carried out in buffer DA (20 mM Tris-Cl, pH 8.0, 1 mM dithiothreitol, 5 mM EDTA) with 0.1 μg μl^{-1} RNaseA, using 200 nM oligonucleotide substrate. Unless otherwise detailed, incubations were performed for 12 h with 2 μM final enzyme concentration. Reactions with mA1 against AT-hotspot oligonucleotides were incubated for 15 min given higher overall catalytic activity. Reactions were incubated at 30 °C and terminated by 20 min at 95 °C. Subsequently, excess complementary oligonucleotide was added (150 nM reaction oligonucleotide to 250 nM complement) and annealed by slow cooling from 95 °C. Duplexed DNA (final 100 nM) was incubated with the appropriate glycosylase using either 0.25 U μl^{-1} UDG (New England Biolabs) or 140 nM hSMUG in 1× buffer DA with 0.1 mg ml^{-1} BSA and incubated at 37 °C for 45 min. For TDG reactions, 1.6 μM TDG was used in 1× buffer DA supplemented with 0.1 mg ml^{-1} BSA, and reactions were incubated at 16 °C for 12 h. The selection of the glycosylase used was based on a comprehensive survey of glycosylases (**Supplementary Fig. 5**) in which we selected a glycosylase with robust activity against a given uracil analog with no detectable activity on the associated cytosine analog. Cytosine and (5F)C substrates were assayed for deamination with UDG; 5mC substrates were assayed with TDG; and (5OH)C, (5Br)C, (5I)C and 5hmC were assayed with SMUG. The duration of each incubation was selected to maximize the excision of the deaminated bases. DNA glycosylase reactions were quenched in 50% (v/v) formamide and 150 mM NaOH and heated to 95 °C for 20 min to cleave the abasic sites generated by glycosylases.

© 2012 Nature America, Inc. All rights reserved.

Analysis of deamination assays. Reaction products were separated by denaturing PAGE (20% (v/v) acrylamide, 1× TBE (100 mM Tris, 90 mM boric acid, 1 mM EDTA), 7 M urea), run at 50 °C and imaged on a Typhoon 9410 (Amersham Biosciences). The intensities of the product and substrate bands were determined in QuantityOne (Bio-Rad) after subtracting background intensities. To account for incomplete excision of deaminated bases, standard curves for quantifying the fraction of deaminated product were generated using S30-ATX substrates (X = C, 5mC or 5hmC) and S30-ATY products (Y = U, T or 5hmU), mixed in various ratios and treated with UDG (for C or U), SMUG (for 5hmC or 5hmU) or TDG (for 5mC or T). Identical conditions were used to generate standard curves with the unnaturally modified S30-TGX substrates and S30-TGY products, with UDG (for (5F)C and (5F)U) and SMUG (for (5OH)C and 5(OH)U, (5Br)C and (5Br)U, and (5I)C and (5I)U). The calculated fraction of cleaved product was determined by the intensity of the fluorescent product over the total product plus substrate. For all deamination assays, the actual fraction of deaminated product was determined with reference to the standard curve generated for the corresponding glycosylase (**Supplementary Figs. 3** and **6**).

Transient expression of candidate enzymes. TDG, mA1 and mA3 from pCMV-SPORT6 vectors (Open Biosystems) were each cloned into a Flag-tagged pcDNA3 vector (pcDNA3β-FLAG). mA2 amplified from mouse embryonic stem cell cDNA was cloned into pcDNA3β-FLAG. The cloned TET2 construct, untagged human AID and the catalytic mutant hAID^{E58A} were previously described[15,39]. A synthetic gene encoding UGI was cloned into pIRESneo3 (Clontech). HEK 293T cells were transfected 18 h after plating using Fugene HD transfection reagents (Roche) and harvested 48 h after transfection. Additional details can be found in **Supplementary Methods**.

Mass spectrometric experiments. We used the DNeasy Kit (Qiagen) to isolate 2.5 μg genomic DNA, which was then heat denatured and hydrolyzed with 90 U of Nuclease S1 (Sigma) in buffer (0.5 mM $ZnSO_4$, 14 mM sodium acetate, pH 5.2) at 37 °C for 1 h, followed by the addition of 5 μl 10× buffer 2 (560 mM Tris-Cl, 30 mM NaCl, 10 mM $MgCl_2$, pH 8.3), 0.5 μg of phosphodiesterase I (Worthington) and 2 U of calf intestinal alkaline phosphatase (New England Biolabs) for an additional 1 h (final volume 50 μl). Digested DNA was then filtered with Nanosep3K (Pall Corporation), and filtered samples were subjected to LC-MS/MS analysis as described previously[15], with an additional transition for 5hmU (*m/z* 259.0 to 125.0) and, after enrichment, for dU (*m/z* 229.0 to 113.0). The absolute amount of nucleoside was quantified by comparison to known standards.

Deamination and base excision activity of nuclear lysates. Nuclear lysates were prepared from HEK 293T cells as previously described[45]. For analysis of deaminase activity, lysates (final protein concentration 0.3 μg μl^{-1}) were incubated with single-stranded oligonucleotide substrates (2 μM) in 1× buffer DA and 0.1 mg μl^{-1} BSA, for 30 min. Reactions were quenched by addition of phenol:chloroform:isoamyl alcohol (25:24:1). Formamide and NaOH were added to DNA extracted from the aqueous phase (50% v/v, 150 mM final concentrations). Samples were processed and analyzed as described for deamination assays. For analysis of base excision activity against uracil or 5hmU in nuclear lysates, lysates (final concentration 0.2–0.3 μg μl^{-1} as detailed) were preincubated at 37 °C for 20 min before addition of 1 μM duplex DNA. Reactions were quenched and analyzed as described above.

Statistical analysis. All data shown represent the mean value for a given number of replicates, as specified in each figure legend. Error bars represent s.d. from the mean. Student's *t*-test determined statistical significance between groups.

Received 7 May 2012; accepted 28 June 2012; published online 8 July 2012

References

1. Nabel, C.S., Manning, S.A. & Kohli, R.M. The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential. **7**, 20–30 *ACS Chem. Biol.* (2012).
2. Rosenberg, B.R. & Papavasiliou, F.N. Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. *Adv. Immunol.* **94**, 215–244 (2007).
3. Teperek-Tkacz, M., Pasque, V., Gentsch, G. & Ferguson-Smith, A.C. Epigenetic reprogramming: is deamination key to active DNA demethylation? *Reproduction* **142**, 621–632 (2011).
4. Fritz, E.L. & Papavasiliou, F.N. Cytidine deaminases: AIDing DNA demethylation? *Genes Dev.* **24**, 2107–2114 (2010).
5. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
6. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
7. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
8. Münzel, M., Globisch, D. & Carell, T. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angew. Chem. Int. Edn Engl.* **50**, 6460–6468 (2011).
9. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
10. Pastor, W.A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
11. Hajkova, P. Epigenetic reprogramming in the germline: towards the ground state of the epigenome. *Phil. Trans. R. Soc. Lond. B* **366**, 2266–2273 (2011).
12. Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
13. Morgan, H.D., Dean, W., Coker, H.A., Reik, W. & Petersen-Mahrt, S.K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279**, 52353–52360 (2004).
14. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
15. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
16. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
17. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
18. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Edn Engl.* **50**, 7008–7012 (2011).
19. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **8**, 328–330 (2012).
20. Cortázar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
21. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
22. Conticello, S.G., Langlois, M.A., Yang, Z. & Neuberger, M.S. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv. Immunol.* **94**, 37–73 (2007).
23. Bransteitter, R., Pham, P., Scharff, M.D. & Goodman, M.F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA* **100**, 4102–4107 (2003).
24. Larijani, M. *et al.* Methylation protects cytidines from AID-mediated deamination. *Mol. Immunol.* **42**, 599–604 (2005).
25. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).
26. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
27. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
28. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201–1212 (2008).
29. Kohli, R.M. *et al.* A portable hotspot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *J. Biol. Chem.* **284**, 22898–22904 (2009).
30. Larijani, M., Frieder, D., Basit, W. & Martin, A. The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in *ung*$^{(-/-)}$ *msh2*$^{(-/-)}$ mice. *Immunogenetics* **56**, 840–845 (2005).
31. Beale, R.C. *et al.* Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J. Mol. Biol.* **337**, 585–596 (2004).
32. Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F. & Chen, X.S. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **445**, 447–451 (2007).
33. Hansch, C. *et al.* "Aromatic" substituent constants for structure-activity correlations. *J. Med. Chem.* **16**, 1207–1216 (1973).
34. Pearl, L.H. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* **460**, 165–181 (2000).
35. Bennett, M.T. *et al.* Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **128**, 12510–12519 (2006).
36. Kavli, B. *et al.* Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J.* **15**, 3442–3447 (1996).
37. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell Cycle* **10**, 2662–2668 (2011).
38. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
39. Kohli, R.M. *et al.* Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J. Biol. Chem.* **285**, 40956–40964 (2010).
40. Pham, P., Bransteitter, R., Petruska, J. & Goodman, M.F. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* **424**, 103–107 (2003).
41. Bruniquel, D. & Schwartz, R.H. Selective, stable demethylation of the

© 2012 Nature America, Inc. All rights reserved.

interleukin-2 gene enhances transcription by an active process. *Nat. Immunol.* **4**, 235–240 (2003).
42. Martinowich, K. *et al.* DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* **302**, 890–893 (2003).
43. Liu, M. *et al.* Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**, 841–845 (2008).
44. Landry, S., Narvaiza, I., Linfesty, D.C. & Weitzman, M.D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.* **12**, 444–450 (2011).
45. Grogan, B.C., Parker, J.B., Guminski, A.F. & Stivers, J.T. Effect of the thymidylate synthase inhibitors on dUTP and TTP pool levels and the activities of DNA repair glycosylases on uracil and 5-fluorouracil in DNA. *Biochemistry* **50**, 618–627 (2011).

Acknowledgments

We are grateful to M. Bartolomei, M. Weitzman and M. Lazar for helpful discussions, K. Gajula and S. Manning for technical assistance and the University of North Carolina Biomarker Mass Spectrometry Facility for guidance. We are also grateful for A. Drohat (University of Maryland), A. Guminski (Johns Hopkins University) and J. Guo and H. Song (Johns Hopkins University) for providing reagents. This work was supported in part by the Rita Allen Foundation (to R.M.K.), the W. W. Smith Charitable Trust (to R.M.K.) and US National Institutes of Health grants K08-AI089242 (to R.M.K.), GM056834 (to J.T.S.) and U01DK089565 (to Y.Z.). Y.Z. is an investigator of the Howard Hughes Medical Institute.

Author contributions

R.M.K., J.T.S. and Y.Z. conceived the project. R.M.K., J.T.S., Y.Y., C.S.N. and Y.Z. designed the experiments. C.S.N., H.J., Y.Y., L.S. and H.L.G. performed the experiments. C.S.N., H.J. and L.S. analyzed the data. C.S.N., H.J., L.S., H.L.G., J.T.S., Y.Z. and R.M.K. interpreted the data. C.S.N., J.T.S. and R.M.K. wrote the manuscript, and all authors edited the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to R.M.K.

© 2012 Nature America, Inc. All rights reserved.