PLOS ONE

RESEARCH ARTICLE

Chemical and enzymatic modifications of 5-methylcytosine at the intersection of DNA damage, repair, and epigenetic reprogramming

Tuvshintugs Baljinnyam[1], Mark L. Sowers[1,2], Chia Wei Hsu[1,2], James W. Conrad[1], Jason L. Herring[1], Linda C. Hackfeld[1], Lawrence C. Sowers[1,3]*

1 Department of Pharmacology and Toxicology, University of Texas Medical Branch, Galveston, Texas, United States of America, 2 MD-PhD Combined Degree Program, University of Texas Medical Branch, Galveston, Texas, United States of America, 3 Department of Internal Medicine, University of Texas Medical Branch, Galveston, Texas, United States of America

* lasowers@UTMB.edu

Abstract

The DNA of all living organisms is persistently damaged by endogenous reactions including deamination and oxidation. Such damage, if not repaired correctly, can result in mutations that drive tumor development. In addition to chemical damage, recent studies have established that DNA bases can be enzymatically modified, generating many of the same modified bases. Irrespective of the mechanism of formation, modified bases can alter DNA-protein interactions and therefore modulate epigenetic control of gene transcription. The simultaneous presence of both chemically and enzymatically modified bases in DNA suggests a potential intersection, or collision, between DNA repair and epigenetic reprogramming. In this paper, we have prepared defined sequence oligonucleotides containing the complete set of oxidized and deaminated bases that could arise from 5-methylcytosine. We have probed these substrates with human glycosylases implicated in DNA repair and epigenetic reprogramming. New observations reported here include: SMUG1 excises 5-carboxyuracil (5caU) when paired with A or G. Both TDG and MBD4 cleave 5-formyluracil and 5caU when mispaired with G. Further, TDG not only removes 5-formylcytosine and 5-carboxycytosine when paired with G, but also when mispaired with A. Surprisingly, 5caU is one of the best substrates for human TDG, SMUG1 and MBD4, and a much better substrate than T. The data presented here introduces some unexpected findings that pose new questions on the interactions between endogenous DNA damage, repair, and epigenetic reprogramming pathways.

Introduction

Double-stranded DNA within cells is the repository for genetic information. DNA can be replicated and transcribed, but otherwise was initially considered to be metabolically inactive. In

the 1970s studies by Lindahl and coworkers revealed that DNA is chemically reactive and can undergo hydrolysis reactions, including the deamination of bases, and the loss of purines and pyrimidines from DNA by hydrolytic cleavage of the glycosidic bond [1–4]. Other studies revealed that DNA bases can also undergo oxidative damage [5–7]. Observations of the endogenous damage of DNA led to the discovery of DNA repair enzymes and pathways by which this damage could be repaired [8–11]. Today, it is well established that the DNA of living cells undergoes hundreds, or thousands, of endogenous DNA damage events every day [12–14] and that multiple, partially overlapping, DNA repair pathways exist to repair most of this damage.

In addition to the canonical bases A, T, G and C, a fifth base, 5-methylcytosine (5mC), was identified in 1948 [15]. Subsequent studies revealed an inverse association between 5mC in gene promoter regions and gene transcription [16]. It is now known that 5mC is an epigenetic marker that can significantly increase the binding affinity of DNA-binding proteins which contain a methyl-binding domain (MBD) [17–20]. The initial binding of these proteins to methylated sites leads to the recruitment of histone-modifying enzymes that create a compact chromatin structure that inhibits transcription [21, 22]. Usually, cytosines in both strands of a symmetrical CpG dinucleotide are methylated. Following DNA replication, a 5mC base in the parental strand can direct the maintenance methyltransferase, DNMT1, to methylate cytosine in the progeny strand re-establishing symmetrical methylation at that CpG dinucleotide [23, 24]. This mechanism for the heritability of methylation patterns suggested an important role for DNA methylation in cellular development and differentiation, and it was initially thought that methylation patterns, once established, would not be reversible.

Alterations in cytosine methylation patterns are tied to cancer etiology in two distinct ways. First, methylation patterns are perturbed in most cancer cells [25, 26]. Aberrant methylation can lead to the transcriptional silencing of tumor suppressor genes, while loss of methylation can lead to the inappropriate expression of transforming genes. Mechanisms by which methylation patterns can be altered are poorly understood. Second, C to T (C>T) transition mutations dominate the mutational landscape of human tumors and these transitions occur with hotspot frequency at methylated CpG dinucleotides [27–30]. The hydrolytic deamination of 5mC generates thymine resulting in the formation of a T:G mispair. T:G mispairs are poorly repaired in mammalian cells which may explain, in part, recurrent C>T mutations found in human tumors [31]. The thymine methyl group can also undergo oxidation, generating 5-hydroxymethyluracil (5hmU), 5-formyluracil (5foU) and 5-carboxyuracil (5caU) [32–35]. In addition to hydrolytic deamination of 5mC to T, the 5-methyl group of 5mC can undergo nonenzymatic oxidation to form 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5foC) and 5-carboxycytosine (5caC) [36–38].

In 2009, the Rao group identified an a-ketoglutarate dependent dioxygenase that could enzymatically convert 5mC to 5hmC in DNA [39]. The discovery of the TET enzymes led to a resurgence in studies on epigenetic reprogramming pathways [40–42]. The measurement of high levels of 5hmC in mammalian cells, particularly in neurons [43], established 5hmC as an important player in epigenetic reprogramming. Subsequent studies with the Tet enzymes revealed that 5mC could be converted not only to 5hmC, but to 5foC and 5caC as well (Fig 1) [44, 45]. The latter two products were then discovered to be substrates for Thymine DNA glycosylase [46]. This pathway has been proposed as the active DNA demethylation pathway.

Mechanistic studies of these DNA modification and repair pathways often rely upon synthetic oligonucleotides containing site-specific modified DNA bases. Our group developed methods to place both 5hmU [47] and 5hmC [48] into synthetic oligonucleotides. Subsequently, we showed that both modifications could interfere with DNA-protein interactions [49]. The replacement of 5mC by 5hmC not only inhibits the binding of MBD-containing proteins [20], but also blocks maintenance methylation by DNMT1 [24], providing mechanistic
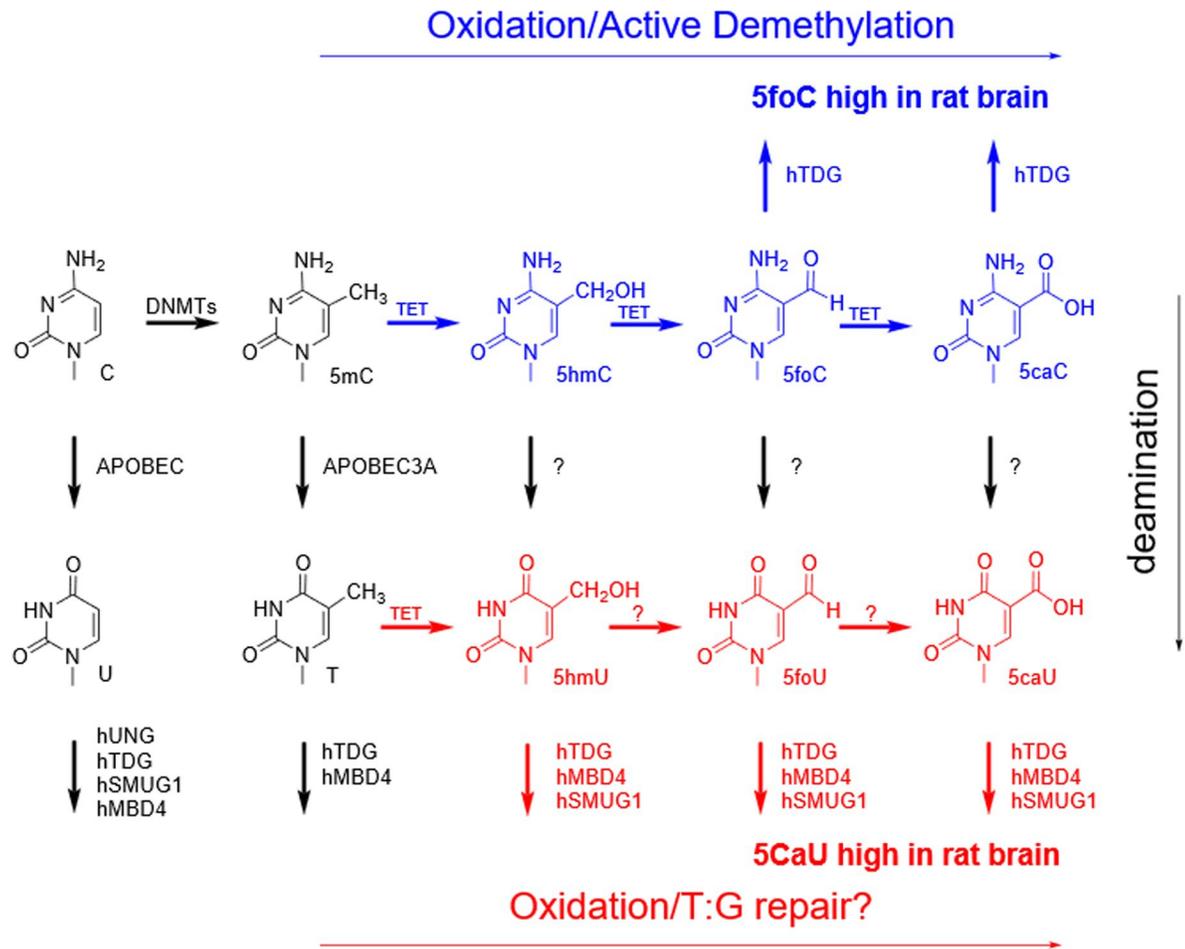
Fig 1. Possible active demethylation pathways resulting from the modification of 5mC. The oxidation of 5mC to 5foC and 5caC, and the removal of 5foC and 5caC by hTDG (blue) have been established. The glycosylase removal of 5hmU, 5foU and 5caU is established here (red). However, the enzymatic conversion of T to 5hmU, 5foU and 5caU (red) remains controversial.

https://doi.org/10.1371/journal.pone.0273509.g001

insight into how enzymatic modification of 5mC in DNA could alter epigenetics, even without further oxidation and or base excision repair.

Furthermore, additional DNA demethylation pathways utilizing the base excision repair (BER) pathway have also been suggested by invoking the deamination products of the cytosine analogs as shown in Fig 1 [50–53]. Following deamination, the corresponding uracil analogs could also be removed by members of the uracil DNA glycosylase superfamily, constituting a complete demethylation pathway. A family of DNA cytosine deaminases is known in human cells [50, 54, 55]. These deaminases account for hypermutation in variable regions of antibody genes in immune cells as well as the destruction of invading viral genomes. While some labs have presented evidence supporting a role for AID/APOBEC deamination of 5mC, and its oxidized analogs, other labs have presented contradictory evidence [50, 55].

An array of pyrimidine analogs arising from both oxidation and deamination are now known to occur in DNA by both enzymatic modification and nonenzymatic spontaneous damage. Currently, it remains unclear which of these modifications of 5mC represent DNA damage, bonafide epigenetic markers, or intermediates in either pathway. The simultaneous presence of modified bases from potentially multiple pathways reveals an intersection of DNA damage, repair, and epigenetic reprogramming.

To elucidate details of potential demethylation and DNA repair pathways, we have prepared a series of synthetic oligonucleotides containing the five cytosine and five uracil analogs shown in Fig 1. We have probed oligonucleotides containing these analogs when paired with A or G with recombinantly expressed human uracil DNA glycosylases including Uracil-DNA Glycosylase 2 (hUNG2), Single-strand selective Monofunctional Uracil DNA Glycosylase 1 (hSMUG1), Thymine DNA Glycosylase (hTDG) and Methyl-CpG-Binding Domain protein 4 (hMBD4). This is the first systematic study of all possible analogs with all relevant members of the BER pathway. The results obtained should shed new light on epigenetic remodeling and repair pathways as well as their possible intersection.

Materials and methods

Oligonucleotide synthesis and characterization

Oligonucleotides were synthesized by standard phosphoramidite methods with an ABI Expedite 8909 nucleic acid synthesis system. Reagents for oligonucleotides synthesis were purchased from Glen Research, Sterling VA, including phosphoramidites of normal bases as well as for 5hmU, 5hmC, 5foC and 5caC. Oligonucleotides containing 5foU were made with unprotected 5-formyl-2'-deoxyuridine as described previously [56, 57]. Oligonucleotides containing 5caU were prepared according to the method previously reported [58, 59]. Sequences of oligonucleotides are shown in Fig 2.

Oligonucleotides were purified DMT-on with a PRP column and on a reverse phase column following detritylation. Oligonucleotides were characterized by gel electrophoresis and MALDI-TOF-MS. Oligonucleotide composition was verified by GC-MS analysis following formic acid hydrolysis and by HPLC following enzymatic hydrolysis as reported previously.

Protein expression and characterization

Human hUNG2 (UNG-1547H) was obtained from Creative BioMart (Shirley, NY).

Plasmid DNA (pET28c(+)) encoding full length hTDG (NM_003211.6) [60] with an N-terminal his-tag was obtained from Addgene (Watertown, MA, USA). Coding sequences for hSMUG1 (NM_001243788.2) [61] and hMBD4 (NM_003925.3) [62] also containing an N-terminal his-tag, were inserted into pET28a(+) vectors by Genscript (Piscataway, NJ).

Plasmid DNAs were transformed into *E. coli* strain BL21 (DE3) and transformants were selected on LB + 1.4% agar plates containing kanamycin (50 mg/mL). After confirmation of target protein expression, cells were grown in 100 mL of LB broth supplemented with kanamycin (50 mg/mL) at 37°C until the OD_{600} increased to 0.6. Cells were then induced with IPTG

```
5'-6FAM-CGT-GGC-YGG-CCA-CGA-CGG-3'

      3'-GCA-CCG-PCC-GGT-GCT-GCC-5'

Y = C, 5mC,5hmC,5foC,5caC,U,T,5hmU,5foU,5caU

P = A or G
```

Fig 2. Sequence of the oligonucleotide duplex used in this study. See Fig 1 for structures of the pyrimidine analogs, 'Y'.

(0.5 mM) and grown at 15˚C for 22 h (hTDG) or 30˚C for 6 h and then 15˚C overnight (hSMUG1), or 30˚C for 6 h (hMBD4). Cells were harvested by centrifugation at 4,100 rpm for 5 min and stored at -20˚C until proteins were isolated.

The cell pellets were thawed and suspended in 4 mL of lysis buffer and sonicated for 8 cycles for 30 s with 30 s breaks on ice. Lysis buffer consisted of 50 mM potassium phosphate pH 8, 300 mM NaCl, 10 mM b-mercaptoethanol (bME), 20 mM imidazole, 1% Triton-X100, and 1 mM PMSF. After centrifugation at 12,000 rpm for 10 min, supernatants were loaded on previously equilibrated HisPur Ni-NTA resin and incubated for 1.5 hours at 4˚C on a see-saw shaker. The suspension of HisPur Ni-NTA Resin beads and cell lysate was centrifuged at 1,000 g, 4˚C for 5 min. Beads were washed sequentially with 3 mL of wash buffers A (x2), wash buffer B (x2), and wash buffer C (x6). Wash buffers contained 50 mM potassium phosphate pH 8, 300 mM NaCl, 10 mM bME, and progressively higher concentrations of imidazole: A, 5 mM; B, 10 mM; and C, 20 mM. His-tagged proteins were eluted from beads in 1 mL of elution buffer containing 300 mM NaCl, 10 mM bME and 100 mM imidazole. Protein concentrations were quantified with a Bradford protein assay using bovine serum albumin as a standard. Proteins were electrophoresed on a 12% tris-glycine SDS gel which was stained with Coomassie brilliant blue. Stained gels were scanned with ImageJ software and protein purity determined by densitometry.

Glycosylase assays measured by gel electrophoresis

In the studies presented here, glycosylase excision of a target base generates an abasic site. Subsequent cleavage of the phosphodiester backbone at the abasic site generates a shorter, labelled oligonucleotide that can be resolved by polyacrylamide gel electrophoresis. Gels bands can be visualized and quantified to determine quantitatively the magnitude of glycosylase excision. Among the modified bases examined here, additional challenges exist with 5foU and 5foC. Our previous studies [57] have revealed that oligonucleotides containing 5foU can spontaneously cleave, especially when heated. Therefore, in the studies reported here, cleavage bands observed in the absence of a glycosylase were used as a control and subtracted from the glycosylase generated band.

To examine the substrate specificity of each glycosylase, 2.5 pmol of a FAM-labelled duplex containing a target pyrimidine pair opposite A or G was incubated with hUNG2 (0.07 mg, MW 40.0 kDa, 1.75 pmol), hTDG (1.5 mg, MW 48.4 kDa, 31 pmol), hSMUG1 (0.5 mg, 30.7 kDa, 16.3 pmol) in buffer containing 10 mM K_2HPO_4, 30mM NaCl, 40mM KCl, pH 7.9 or hMBD4 (0.5 mg, 66.9 kDa, 7.5 pmol) in buffer containing 20 mM tris-HCl, 1 mM DTT and 1mM EDTA at pH 8.0 and incubated for 1 h at 37˚C. Recombinant hyTDG(Y163K) (0.5 mg, MW 29.7 kDa, 16.8 pmol) with no glycosylase activity, but efficient AP endonuclease activity [63, 64], was added and incubated at 65˚C for 1 h to cleave the phosphodiester backbone at the abasic site. An equal volume of formamide was added and samples were loaded onto a 20% polyacrylamide gel containing 6M urea and 1xTBE buffer and electrophoresed at 180 V for 40 min using a Bio-Rad Mini-PROTEAN Tetra Cell (Hercules, CA). Gels were visualized with a Storm 860 gel imager.

Following target bases excision, oligonucleotide cleavage can be accomplished with an AP lyase or NaOH. In this study, when analogs were paired with G, the phosphodiester backbone were cleaved with a Y163K mutant of the hyTDG we recently described [63], which has lost glycosylase activity but gained AP lyase activity [64]. The advantage of the AP lyase is that neutral conditions employed do not result in chemical damage to modified bases. The disadvantage is that the AP lyase (Y163K) retains its selectivity for cleavage opposite G, underestimating glycosylase cleavage when the target pyrimidine is paired with A.

Alternatively, oligonucleotides containing abasic sites can be cleaved in alkali, 0.1 M NaOH and heated to 96˚C for 10 min. However, 0.1 M NaOH causes chemical cleavage of oligonucleotides containing 5foC. Therefore, we measured the cleavage of oligonucleotides containing target pyrimidines paired with A in the presence and absence of glycosylase and treated with NaOH. Cleavage in the absence of a glycosylase was then subtracted from any observed glycosylase cleavage.

Glycosylase activity measured by GC-MS

Isotope standards for 5hmU, 5foU, 5caU, 5hmC, 5foC and 5caC were prepared as previously described [65]. Free bases released by glycosylases were measured using GC-MS. Briefly, 110 pmol of each oligonucleotide containing a target pyrimidine was combined with 2 equivalents of a complementary oligonucleotide. To that, a mixture of isotope-enriched pyrimidine standards containing 55 pmol of each standard described above, was mixed in a 10 mM potassium phosphate buffer, pH 7.7, 40 mM NaCl and 50 mM KCl. 100 pmol of each glycosylase was then added and samples were incubated at 37˚C for 2 h in a total volume of 100 μL.

300 mL of de-ionized water was then added to each sample and vortexed. Released free bases were separated from oligonucleotides and proteins by spin filtration using a 3 kDa Amicon filter at 14,000 g for 40 min. The flow through was dried in a GC vial under reduced pressure. Pyrimidines were derivatized in 20 mL acetonitrile and 20 mL of N-(t-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA) with 1% tert-butyldimethylchlorosilane (TBDMCS) at 140˚C for 40 min. Retention times and expected ions of free bases released from oligonucleotides as well as isotope-labeled standards are listed in S1 Table in S1 File.

0.5 mL of each sample was injected onto an Agilent 7890A GC equipped with an Agilent J&W DB-5MS + DG column (30 m × 0.25 mm id, film thickness 0.25 μm) using helium carrier gas at 1 mL/min constant flow. The GC oven temperature was held at 100˚C for 2 min, ramped to 260˚C at 30˚C /min then held for 10 min. The GC was directly coupled to an Agilent 5975C Mass Selective Detector. Data was collected in the selected ion mode using ions appropriate for each analyte. Pyrimidines were quantified by comparing the peak area of the released pyrimidine to the peak area of the corresponding isotope standard.

Results

Preparation and validation of oligonucleotides and DNA uracil glycosylases

Oligonucleotides containing modified bases require special protection and deprotection strategies. Oligonucleotides containing the bases of interest (Fig 2) were synthesized using previously published methods, including methods for the synthesis of oligonucleotides containing 5hmU [47], 5foU [56, 57], 5caU [58, 59], 5hmC [48], 5foC [66–68] and 5caC [66, 67].

Oligonucleotides containing modified bases must be carefully characterized because deprotection may not occur as anticipated, unintended side reactions can occur, and the modified bases themselves can undergo further reactions. In this study, oligonucleotides were characterized by gel electrophoresis and MALDI-TOF mass spectrometry. The composition of each oligonucleotide was confirmed by HPLC analysis following enzymatic hydrolysis and by GC-MS analysis following formic acid hydrolysis as described previously.

Four members of the uracil DNA glycosylase superfamily are thought to be involved in the repair of deamination and methyl group oxidation and shown in Fig 1. Of these, hUNG2 was obtained commercially and three were prepared in our laboratory. Each enzyme was examined by gel electrophoresis and their sequence confirmed by LC-MS/MS following trypsin digestion.

Glycosylase assay substrate selectivity

Oligonucleotide duplexes were prepared by mixing fluorescently tagged oligonucleotides containing target pyrimidines with complementary strands placing either A or G opposite the target pyrimidine (Fig 2). Oligonucleotide duplexes were then incubated with a molar excess of each glycosylase, followed by incubation with an AP-endonuclease to cleave the phosphodiester bond at abasic sites generated by glycosylase excision. Oligonucleotides were then separated by gel electrophoresis and visualized. Bands corresponding to intact and cleaved oligonucleotides were quantified. This was repeated a total of three times for each substrate.

Results of glycosylase excision for the oxidized uracil analogs paired with G or A are shown in Fig 3. For the oxidized cytosine analogs, the results are shown in Fig 4. As expected, hUNG2 had the narrowest substrate range. Among the substrates examined, hUNG2 cleaved only U and when present in a U:G mispair or U:A base pair. On the other hand, hSMUG1 had an increased substrate range that included U, 5hmU, 5foU and 5caU, but not T. As with hUNG2, target pyrimidines were excised when paired with G or with A. The overall substrate preference under these conditions for hSMUG1 was U ~ 5caU> 5hmU > 5foU when mispaired with G. When paired with A the order was 5caU > U > 5hmU ~ 5foU.

hTDG also had the broadest substrate range and cleaved all the oxidized uracil analogs, as well as T. However, target pyrimidines were excised much more efficiently when paired opposite G. The rank order for cleavage opposite G was 5caU > U > 5hmU ~ 5foU >> T. There was modest cleavage of U:A but other uracil analogs when paired with A were not detectable. Among the glycosylases examined here, hTDG was the only to cleave any of the cytosine analogs. The cytosine analogs 5foC and 5caC were excised when paired opposite either G or A.

The last glycosylase examined, hMBD4, had a narrower substrate range than hSMUG1 or hTDG. hMBD4 excised U, > 5foU > 5caU > hmU ~ T. Similar to hTDG, hMBD4 only removed U when paired with A, but not the other uracil analogs.

In this study as well as some previous studies, substantial cleavage of oligonucleotides containing 5foU could occur in the absence of a glycosylase, due to thermal hydrolysis of the glycosidic bond (65). Data presented in Table 1 was adjusted to account for observed spontaneous cleavage of 5foU in the negative control lane where no enzyme was added. Alkaline hydrolysis of 5foC paired with A similarly caused artifactual cleavage (Fig 4). Data presented in Table 1 was also adjusted to account for alkaline hydrolysis of 5foC-containing oligonucleotides in the absence of a glycosylase.

Stable-isotope dilution gas-chromatography glycosylase validation

Glycosylase reactions were repeated using a GC-MS based method (Fig 5). This was done to verify that cleavage of target pyrimidines indicated by the fluorescent gel-based assay was due to excision of the modified free base by breaking the N-glycosidic bond, consistent with monofunctional glycosylase activity. Oligonucleotide duplexes containing target pyrimidines were incubated with a glycosylase, and released free bases were separated with a spin filter and analyzed by GC-MS using stable-isotope-enriched standards.

A total of 24 confirmatory GC-MS analyses were conducted, and all confirmed the findings from the gel-based assay (S5 to S28 Figs in S1 File). As an example, the release of 5caU when paired with G was verified for hTDG, hSMUG1 and hMBD4 (Fig 6). The expected retention time was confirmed by coelution with the authentic isotope standard, and the GC peak was collected when monitoring for the appropriate ions. In this case 441 amu was monitored for 5CaU released from the oligonucleotide and 443 amu for the isotope standard, 5CaU^{+2}.
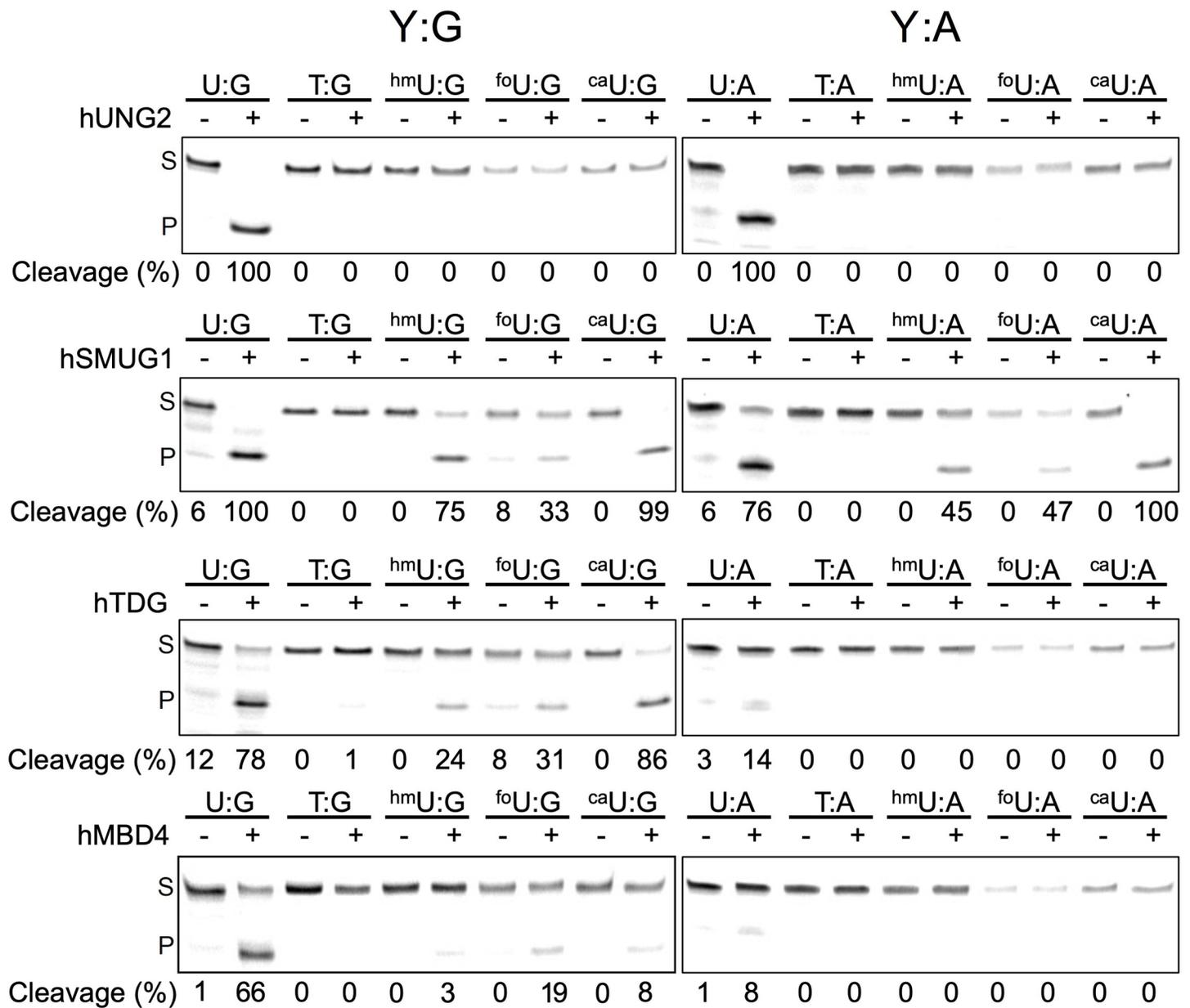
Fig 3. The excision specificity of human uracil family glycosylases for a series of 5-substituted uracil analogs (Y) paired with G (left) or A (right). FAM-labelled oligonucleotide duplexes (2.5 pmol) in buffer appropriate for each enzyme were incubated with a glycosylase at 37°C for 1 h. Duplexes paired with G were then incubated with hyTDG-lyase (Y163K) (65°C, 1 h) and those paired with A incubated with NaOH solution (96°C, 10 min) to cleave the abasic site. Oligonucleotides were resolved on a 20% polyacrylamide/urea gel, visualized on a STORM imager, and quantified with "Image quant" software. The amount of cleavage is shown below the lanes in each gel (S = substrate, P = product). Some artifact cleavage of 5foU-containing oligonucleotides is observed due to its thermal lability. Replicate gel images are shown in S1 and S2 Figs in S1 File.

https://doi.org/10.1371/journal.pone.0273509.g003

Discussion

Role of 5-methylcytosine in cancer and epigenetic regulation

The modified base, 5mC, is the primary epigenetic mark in the DNA of mammals. It has long been implicated as a contributor to cancer development. Transition mutations at methylated CpG sites are one of the most abundant single base changes observed in human tumors
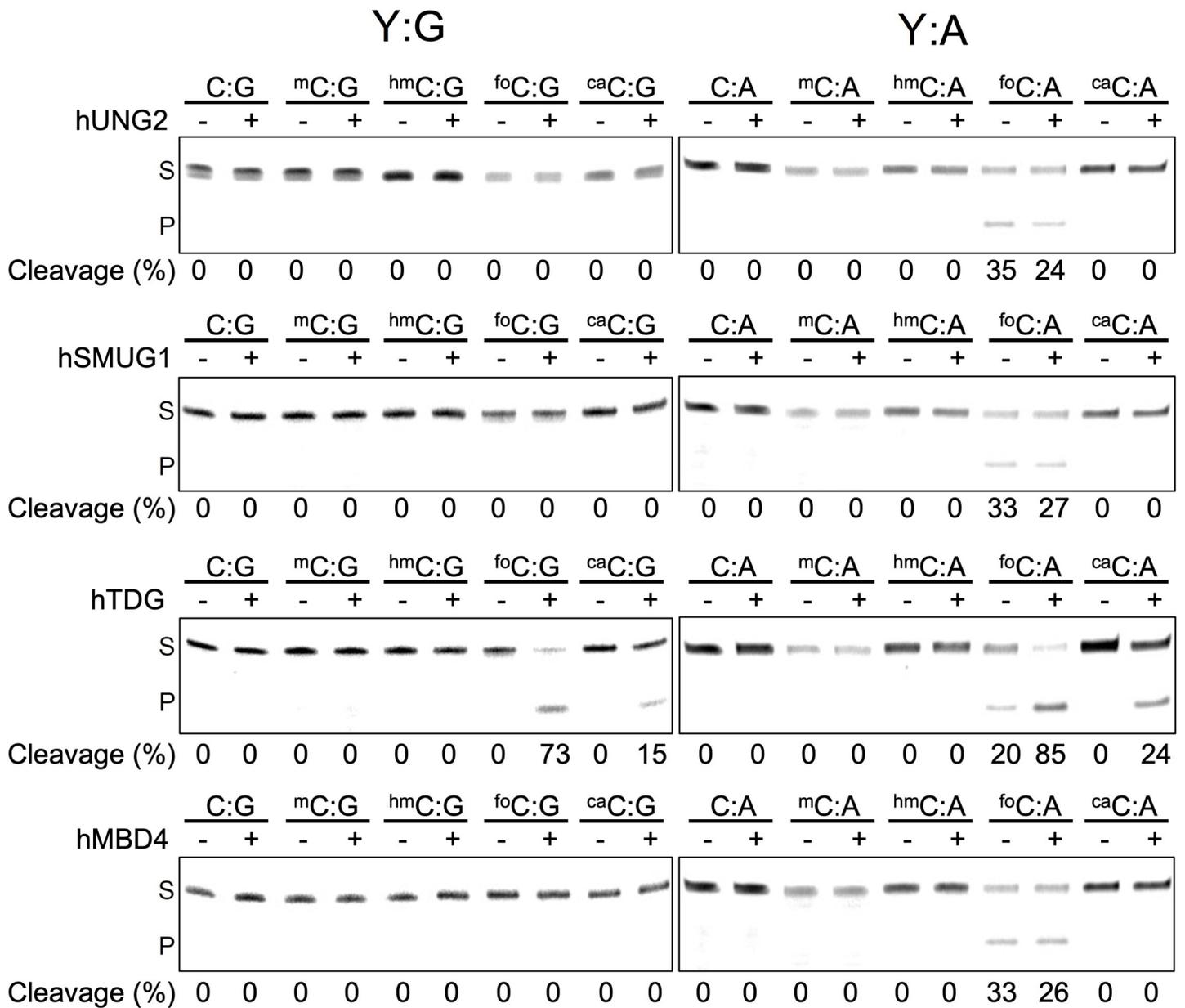
Fig 4. The excision specificity of human uracil family glycosylases for a series of 5-substituted cytosine analogs (Y) paired with G (left) or A (right). FAM-labelled oligonucleotide duplexes (2.5 pmol) in buffer appropriate for each enzyme were incubated with a glycosylase at 37°C for 1 h. Duplexes paired with G were then incubated with hyTDG-lyase (Y163K) (65°C, 1 h) and those paired with A incubated with NaOH solution (96°C, 10 min) to cleave the abasic site. Oligonucleotides were resolved on a 20% polyacrylamide/urea gel, visualized on a STORM imager, and quantified with "Image quant" software. The amount of cleavage is shown below the lanes in each gel (S = substrate, P = product). Some artifact cleavage of 5foC-containing oligonucleotides is observed due to its alkaline lability. Replicate gel images are shown in S3 and S4 Fig in S1 File.

https://doi.org/10.1371/journal.pone.0273509.g004

[27–30]. Additionally, cytosine methylation patterns are perturbed in most, if not all cancer cells. This results in inappropriate expression of transforming genes or silencing of tumor suppressor genes [25, 26].

Understanding mechanisms by which methylation patterns are established, and reprogrammed, may reveal potential defects in these pathways that could result in human disease. Our prior studies suggested that 5hmC might be an important intermediate in epigenetic

Table 1. Extent of cleavage for indicated bases in duplex oligonucleotides by members of the human uracil glycosylase family. Excision studies for analogs paired with G used a hyTDG-lyase (Y163K) to cleave the phosphodiester backbone whereas those paired with A were cleaved with NaOH. Values shown in the table represent the averages and standard deviations from three independent experiments. Cleavage of 5foU-containing oligonucleotides were adjusted for thermal hydrolysis by subtraction of cleavage observed in no enzyme controls. Cleavage of 5foC:A base pairs was adjusted for alkaline hydrolysis by subtraction of cleavage in no enzyme controls. Glycosylase excision of 5caC paired with A by hTDG was also measured using hyTDG-lyase (Y163K) and a value of 50.3±3.9% cleavage was observed. Phosphodiester backbone cleavage using the hyTDG-lyase (Y163K) underestimates glycosylase cleavage when the target pyrimidine is paired with A rather than G.

	Substrate	hUNG2 (n = 3)	hSMUG1 (n = 3)	hTDG (n = 3)	hMBD4 (n = 3)
1	U:G	100±0	94±1.5	61±6.2	60±6.4
2	T:G	0	0	1±1	2±2.0
3	5hmU:G	0	76±3.2	19±8.9	2±0.6
4	5foU:G	0	28±5.5	29±11	23±13
5	5caU:G	0	98±3.2	81±6.4	9±4.0
6	U:A	100±0	67±4.9	12±4.9	9±4.0
7	T:A	0	0	0	0
8	5hmU:A	0	44±1.0	0	0
9	5foU:A	0	45±2.0	0	0
10	5caU:A	0	100±0	0	0
11	C:G	0	0	0	0
12	mC:G	0	0	0	0
13	5hmC:G	0	0	0	0
14	5foC:G	0	0	79±6.5	0
15	5caC:G	0	0	28±11	0
16	C:A	0	0	0	0
17	mC:A	0	0	0	0
18	5hmC:A	0	0	0	0
19	5foC:A	0	0	65±3.5	0
20	5caC:A	0	0	23±1.7	0

https://doi.org/10.1371/journal.pone.0273509.t001

reprogramming [20, 24, 51]; however, this pathway was not firmly established until the discovery of the TET family of oxidases. Subsequent studies have revealed additional 5mC analogs
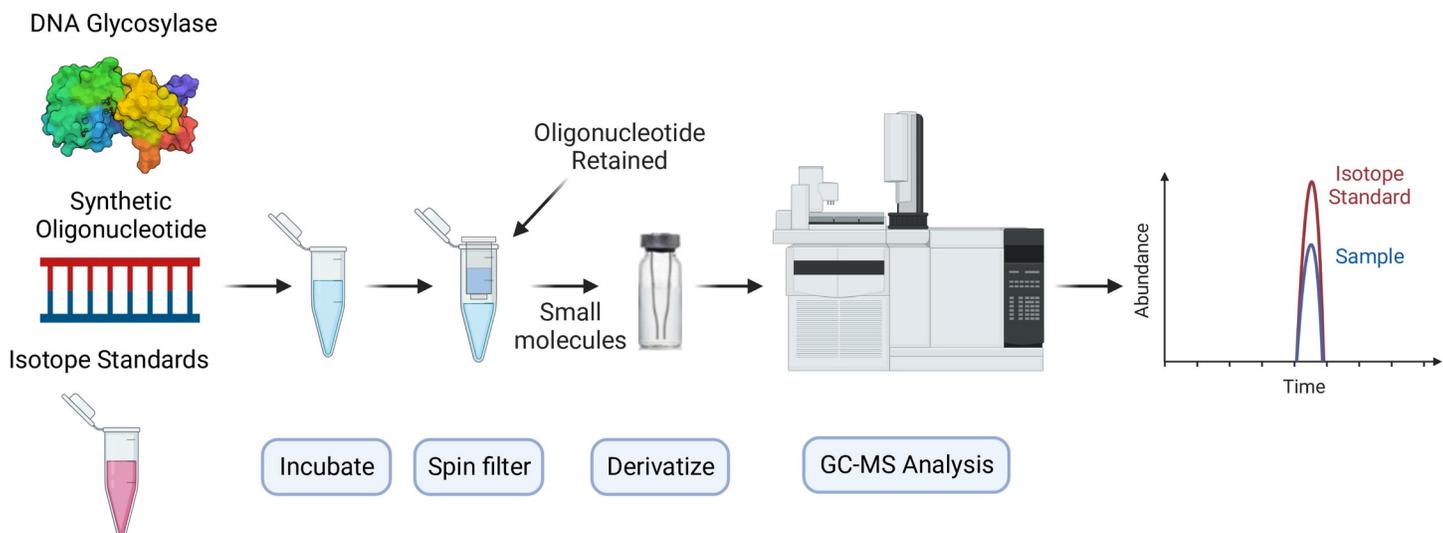
Fig 5. GC-MS scheme. This workflow illustrates the GC-MS based glycosylase assay to determine excision and base release from a synthetic oligonucleotide using stable-isotope standards.

https://doi.org/10.1371/journal.pone.0273509.g005
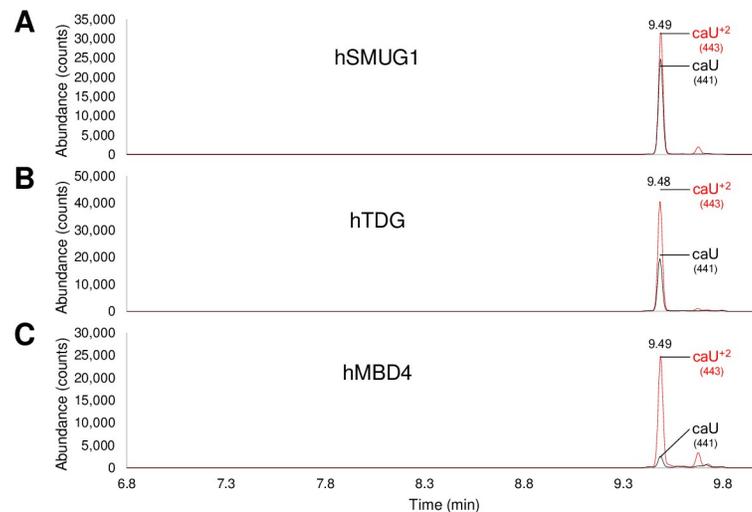
Fig 6. Confirmation of 5caU removal measured by GC-MS. A) 5caU:G treated with hSMUG1, B) 5caU:G treated with hTDG, and C) 5caU:G treated with hMBD4. A mixture of duplex oligo, containing a 5caU:G mispair (110 pmol) with a 2-fold excess of the complementary strand, and a 5caU+2 isotope labeled standard (55 pmol) was prepared in buffer: 10 mM potassium phosphate pH 7.7, 30 mM sodium chloride and 40 mM potassium chloride. The oligonucleotides and free base standard were incubated with approximately 100 pmol of either hSMUG1, hTDG or hMBD4 at 37°C for 2 h in a 110 μL reaction volume. Free bases were separated from enzymes and oligonucleotides by spin filtration, 90% of the sample was dried and analyzed by GC-MS. Derivatized free bases were separated by GC and identified by MS with selected ion monitoring. The ions for 5caU released by glycosylases (441 m/z, black) and 5caU+2 isotope enriched standard (443 m/z, red) were monitored, and the relative sizes of the integrated peak areas were used to calculate the amount of 5caU released. There was approximately 39, 24 and, 5 pmoles 5caU released by hSMUG1, hTDG and hMBD4, respectively. The overall removal of 5caU when paired opposite G by human glycosylases was hSMUG1 > hTDG > hMBD4, consistent with the trend seen by gel-based analysis of excision.

that may serve as epigenetic marks. The discovery of the enzymatic removal of 5foC and 5caC by TDG [46, 52, 69–73] provided a complete enzymatic pathway for active DNA demethylation. The participation of the BER pathway in epigenetic reprogramming, by excising analogs that could also be endogenous DNA damage products, introduces many intriguing questions.

Systematic study of potential 5-methylcytosine analogs and DNA glycosylase activity

We have for the first time performed a systematic study of the activities of the four human glycosylases in uracil DNA glycosylase superfamily on the complete set of pyrimidines that can arise from oxidation and deamination of 5mC. Data are presented graphically in Fig 7. Studies with synthetic oligonucleotides containing modified bases allow direct observation of the activities of glycosylases. While the synthesis of oligonucleotides containing normal DNA bases is routine, the inclusion of modified bases that require additional selective protection, or are chemically labile, requires more challenging chemistry.

Our laboratory has previously developed a synthetic method to incorporate 5hmU into DNA in 1993 [47], and subsequently 5hmC in 1997 [48]. The availability of these analogs allowed studies of DNA-protein interactions and glycosylase activities. The discovery of the TET enzymes published in 2009 [39] drove a resurgence in synthetic activity and currently, phosphoramidites for the synthesis of 5hmC, 5foC, 5caC and 5hmU are commercially available. Phosphoramidites for 5foU and 5caU [56–59] have been developed but are not currently available commercially and were synthesized herein.
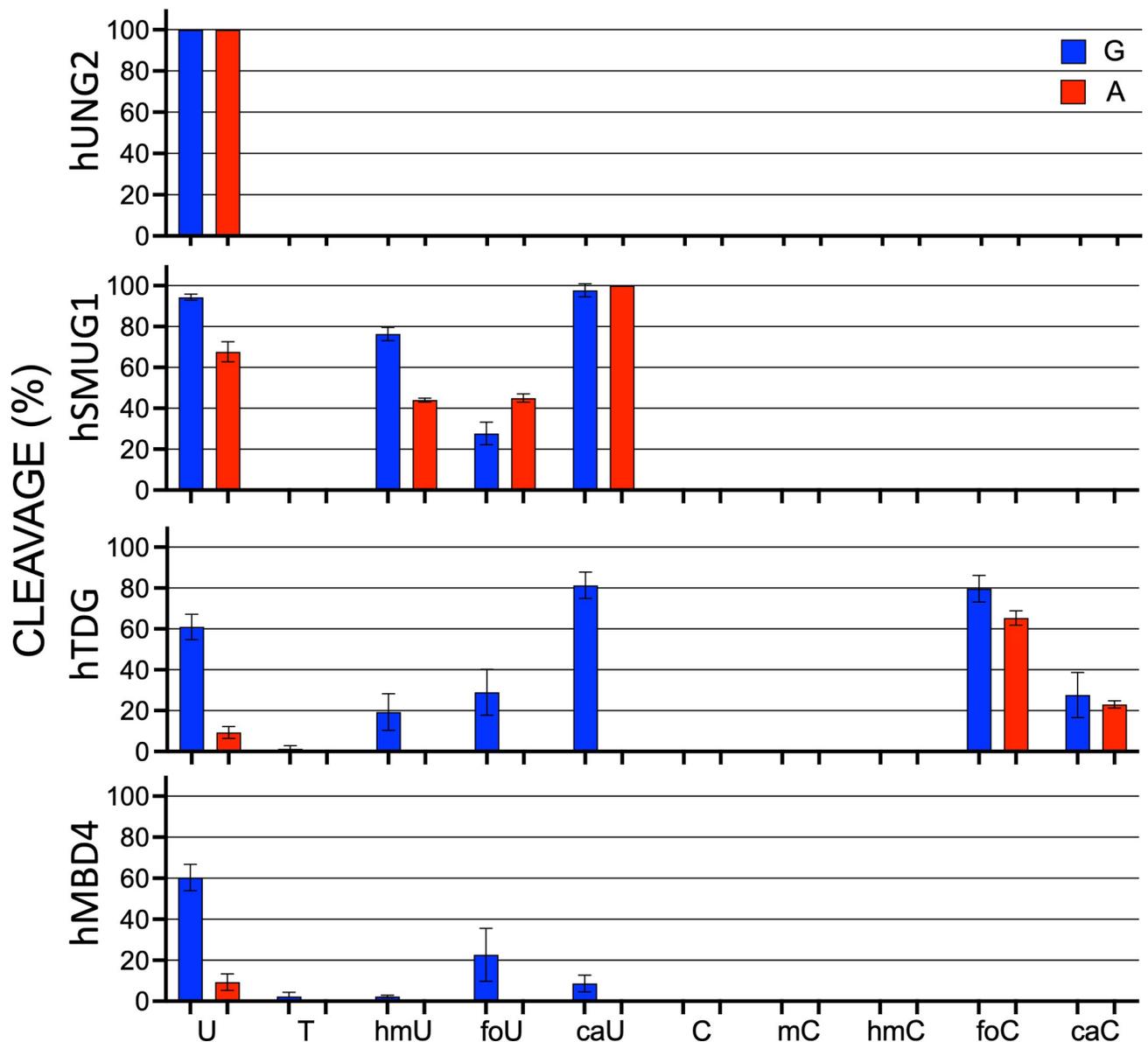
Fig 7. Summary for human glycosylase specificity for substrate. U, T, 5hmU, 5foU, 5caU, C, 5mC, 5hmC, 5foC and 5caC (mispaired with G (blue) or A (red)) oligonucleotides were treated with indicated enzymes for 1 hour at 37°C, and subsequently treated with hyTDG-lyase (Y163K) or NaOH to break the DNA backbone at the abasic site, mixed with an equal volume of formamide, and separated in 20% polyacrylamide denaturing gel. Gels were visualized using a STORM gel imager. Data presented as average value ± SD (n = 3).

https://doi.org/10.1371/journal.pone.0273509.g007

Among the glycosylases examined here, hUNG2 has the most limited substrate range. Uracil is efficiently removed from DNA by hUNG2, when paired with either A or G, and even in single-stranded DNA [74–76]. hUNG2 functions to remove uracil resulting from deamination of cytosine (U:G). It also removes uracil in a U:A base pair resulting from the misincorporation of dUMP during DNA replication or repair synthesis. While U:A is not mutagenic, uracil in DNA paired with A can interfere with sequence-specific DNA protein interactions [77], possibly explaining the role for repair of U:A. Prior structural studies have shown that a tyrosine residue in the pyrimidine binding pocket prevents hUNG2 from excising pyrimidines with 5-substituents larger than hydrogen or fluorine atoms [76]. Additionally, a strategically placed

asparagine residue, locked in place by a network of water molecules, distinguishes U from C analogs [75]. We confirm here that hUNG2 does not remove any of the U or C analogs with oxidized 5-methyl groups. While hUNG2 has an important role in protecting DNA from endogenous DNA damage, it is unlikely involved in epigenetic reprogramming.

hSMUG1 [61, 78–81] was initially identified as a single-strand selective uracil glycosylase [61]. However, subsequent studies examining buffer conditions revealed it to be active on duplex DNA [78]. hSMUG1 does not remove T from a T:G mispair. Previously, hSMUG1 was shown to remove 5hmU, 5foU and 5caU when mispaired with G and 5hmU and 5foU when paired with A [79–81]. We show here that hSMUG1 also removes 5caU when paired with A. Surprisingly, 5caU is removed as efficiently as U when paired with A or G. Previous structural studies have suggested the presence of a water molecule within the catalytic pocket of hSMUG1 that would block entry of the T-methyl group [78–81]. Alternatively, the water molecule could be displaced by polar substituents such as 5hmU, 5foU and 5caU which could hydrogen bond with amides previously holding the water molecule. Collectively, these results suggest that an important endogenous function of hSMUG1 is the removal of oxidation damage products derived from T. As hSMUG1 also repairs these damaged bases when paired with G, it may also play a role in epigenetic reprogramming.

hTDG has been shown to excise both 5foC and 5caC, but not 5hmC, when paired with G as previously described [46, 69, 72, 82–84]. It is surprising that hTDG shows activity against all the thymine oxidation analogs, including 5hmU, the cytosine oxidation analogs 5foC and 5caC, but not 5hmC. High resolution crystal structures reveal that the active site of hTDG with a substrate bound demonstrate various specific DNA-protein interactions (Fig 8) [69]. Importantly an asparagine (N191) can hydrogen bond with the pyrimidine ring and the amide backbone of tyrosine 152 (Y152) can also potentially hydrogen bond with the 5-substituent. However, 5hmC due to an internal hydrogen bond could potentially clash with the hydrogen on Y152. This may explain the preference for 5hmU rather than 5hmC.

Prior studies have also shown that hTDG preferentially cleaves analogs opposite G. It was therefore surprising to see that hTDG cleaves 5foC and 5caC opposite A with similar efficiency. The excision of 5foC and 5caC by hTDG is in accord with its participation in epigenetic reprogramming.

In accord with our own results, hTDG has been shown to remove mispaired uracil efficiently; however, its activity on mispaired thymine is remarkably weak. hTDG has also been reported to remove mispaired 5hmU as well as mispaired 5foU, but not as efficiently as mispaired U. Unlike the other U analogs, 5foU is more chemically labile. A distinct cleaved band is frequently observed in the control lane, especially when alkali and heat are used to cleave the DNA phosphate backbone. The chemical lability of 5foU would make it a poor epigenetic signal.

The methyl binding domain-4 protein (hMBD4) has two domains. A methyl binding domain on the N-terminus and a glycosylase activity on the C-terminus [85–90]. hMBD4 is the second of two glycosylases with T:G activity; however, like TDG, the T:G activity is relatively weak. Also like hTDG, hMBD4 can excise the oxidized T analogs when paired with G. The presence of the methyl binding domain has been suggested to localize hMBD4 to methylated CpG dinucleotides to promote repair of 5mC adducts.

The most surprising finding from this study is that 5caU is a very good substrate for three of the glycosylases examined here. It is the best substrate for hSMUG1 and hTDG, and third best for hMBD4. Mispaired T is excised by hTDG and hMBD4, but it is the worst substrate of the uracil analogs examined. Upon the basis of these findings, the primary function of hTDG and hMBD4 appears to be the removal of the oxidized T analogs, and that T removal seems almost to be an incidental activity.
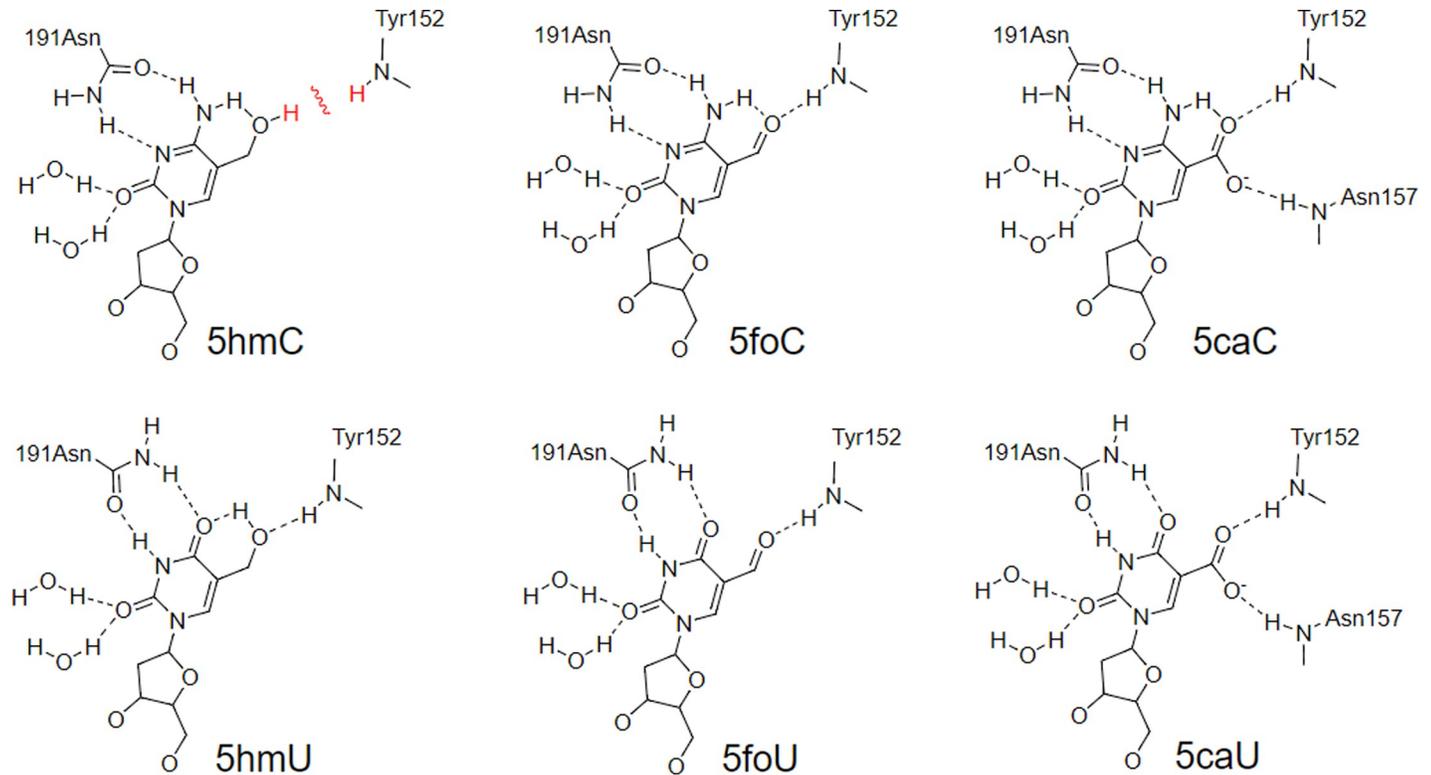
Fig 8. Potential hydrogen bonding interactions in the catalytic pocket of hTDG that allows recognition of 5foC, 5caC, 5hmU, 5foU, and 5caU, but not 5hmC.

5-Carboxyuracil is found *in vivo*

Recently we examined the free pyrimidines in extracts from rat brain. We found high levels of 5caU, which is not an intermediate in pyrimidine synthetic pathways, that were comparable to levels of 5foC measured [65]. Furthermore, Guerniou et al. have shown repair activities for 5hmU:G, 5foU:G and 5caU:G in Hela cell extracts [91]. Coupled with the glycosylase results presented here, good evidence exists for the formation of 5caU:G in DNA and its subsequent removal by the BER pathway. However, this poses additional questions of what role it plays in the genome.

5mC deamination: The Achilles heel of the genome

The T:G mispair constitutes an Achilles heel for the human genome. The deamination of 5mC: G to T:G is responsible for a disproportionate number of transition mutations found in human tumors. Why then is T:G so poorly repaired? The discovery of uracil glycosylase by Lindahl and coworkers was rationalized by the need to remove U:G, which would result in a mutation if not repaired. The deamination of 5mC in DNA is only slightly faster than C, but generates a slowly repaired T:G mispair [92]. Although the T:G mispair could also be repaired by the mismatch repair pathway [93], mutations at methylated CpG dinucleotides appear as mutational hotspots in human tumors, which has been explained by the poor repair of T:G [31]. Our results demonstrate that the capacity of enzymes attributed to T:G repair *in vivo*, have instead been better optimized to repair analogs with polar 5-substituents, which would diminish the repair of T:G.

The formation of 5hmU:G, 5foU:G and 5caU:G mispairs would require the oxidation of 5mC and deamination, or deamination followed by oxidation (Fig 1). The formation of these

analogs by endogenous chemical pathways is unlikely as we have discussed previously [51]. Alternatively, enzymatic pathways have been proposed. The enzymatic deamination of cytosine and cytosine analogs in DNA has been demonstrated. However, the known deaminases do not function efficiently on the oxidized 5mC analogs [55]. The TET enzymes can oxidize 5mC to 5caC in DNA, but they appear to act predominantly on 5mC and not T. Although enzyme activities that oxidize T are known [94], only weak T oxidation has been demonstrated for TET family enzymes or in mammalian cells [95, 96].

Pathways for the formation of 5caU:G in DNA remain an enigma although the glycosylases proposed to be involved in epigenetic reprogramming have robust 5caU:G activity, Hela cells extracts have abundant 5caU:G excision activity, and the resulting glycosylase product, 5caU, has been measured in rat brain at significant levels. The oxidation of 5mC to 5foC and 5caC, followed by hTDG removal represents an active demethylation pathway supported by results from enzymatic assays on TET oxidases and hTDG. Perhaps the capacity of hTDG to remove 5foC and 5caC when paired with G, as well as 5foU and 5caU when paired with G, provides some backup to inadvertent deamination of intermediates in epigenetic reprogramming.

Intersection of DNA damage, repair and epigenetic reprogramming

The results of our studies generate many additional questions on mechanisms of T:G repair and the intersection of DNA damage, repair and epigenetic reprogramming. For example, it is intriguing to hypothesize that the role of enzymatic repair of deaminated and oxidized pyrimidines *in vivo* could be to reduce the incidence of transition mutations by converting a poorly repaired substrate, a T:G mispair, into several significantly better substrates. However, it is unclear if the formation of the oxidized uracil analogs occurs by either spontaneous and or enzymatic pathways. In addition, while in replicating cells a T:G mispair could lead to a C>T transition mutation, this would not be the case in post-mitotic cells, such as neurons. Interestingly, MBD proteins do not appear to discriminate between 5mC:G or T:G mispair [20]. Therefore, it begs the question if repair of the oxidation products of 5mC deamination could also serve to preserve DNA-protein interactions that could be critical to proper cellular function or serve as epigenetic markers in their own right.

Supporting information

S1 File.
(DOCX)

S1 Raw images.
(PDF)

Author Contributions

Conceptualization: Mark L. Sowers, Chia Wei Hsu, Lawrence C. Sowers.

Data curation: Tuvshintugs Baljinnyam, James W. Conrad, Lawrence C. Sowers.

Formal analysis: Tuvshintugs Baljinnyam, Mark L. Sowers, Jason L. Herring, Lawrence C. Sowers.

Investigation: Tuvshintugs Baljinnyam, Jason L. Herring.

Methodology: Tuvshintugs Baljinnyam, Mark L. Sowers, Jason L. Herring, Linda C. Hackfeld, Lawrence C. Sowers.

Project administration: Lawrence C. Sowers.

Resources: Tuvshintugs Baljinnyam, Mark L. Sowers, Jason L. Herring, Linda C. Hackfeld, Lawrence C. Sowers.

Supervision: Lawrence C. Sowers.

Validation: Tuvshintugs Baljinnyam, Mark L. Sowers, Jason L. Herring.

Visualization: Tuvshintugs Baljinnyam, Mark L. Sowers, Jason L. Herring, Lawrence C. Sowers.

Writing – original draft: Tuvshintugs Baljinnyam, Mark L. Sowers, Lawrence C. Sowers.

Writing – review & editing: Tuvshintugs Baljinnyam, Mark L. Sowers, Chia Wei Hsu, James W. Conrad, Jason L. Herring, Lawrence C. Sowers.

References

1. Lindahl T, Nyberg B. Rate of Depurination of Native Deoxyribonucleic Acid. Biochemistry. 1972; 11: 3610–3618. https://doi.org/10.1021/bi00769a018 PMID: 4626532

2. Lindahl T, Andersson A. Rate of Chain Breakage at Apurinic Sites Double-Stranded Deoxyribonucleic Acid. Biochemistry. 1972; 11: 3618–3623. https://doi.org/10.1021/bi00769a019 PMID: 4559796

3. Lindahl T, Nyberg B. Heat-induced deamination of cytosine residues in DNA. Biochemistry. 1974; 13: 3405–3410.

4. Lindahl T. My Journey to DNA Repair. Genomics, Proteomics Bioinforma. 2013; 11: 2–7. https://doi.org/10.1016/j.gpb.2012.12.001 PMID: 23453014

5. ImLay J, Linn S. DNA Damage and Oxygen Radical Toxicity. Science (80-). 1988; 240: 1302–1309. https://doi.org/10.1126/science.3287616 PMID: 3287616

6. Yu Y, Cui Y, Niedernhofer LJ, Wang Y. Occurrence, Biological Consequences, and Human Health Relevance of Oxidative Stress-Induced DNA Damage. Chem Res Toxicol. 2016; 29: 2008–2039. https://doi.org/10.1021/acs.chemrestox.6b00265 PMID: 27989142

7. Kumar N, Raja S, Van Houten B. The involvement of nucleotide excision repair proteins in the removal of oxidative DNA damage. Nucleic Acids Res. 2020; 48: 11227–11243. https://doi.org/10.1093/nar/gkaa777 PMID: 33010169

8. Barnes DE, Lindahl T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu Rev Genet. 2004; 38: 445–476. https://doi.org/10.1146/annurev.genet.38.072902.092448 PMID: 15568983

9. Bauer NC, Corbett AH, Doetsch PW. The current state of eukaryotic DNA base damage and repair. Nucleic Acids Res. 2015; 43: 10083–10101. https://doi.org/10.1093/nar/gkv1136 PMID: 26519467

10. Baiken Y, Kanayeva D, Taipakova S, Groisman R, Ishchenko AA, Begimbetova D, et al. Role of Base Excision Repair Pathway in the Processing of Complex DNA Damage Generated by Oxidative Stress and Anticancer Drugs. Front Cell Dev Biol. 2021; 8: 1–15. https://doi.org/10.3389/fcell.2020.617884 PMID: 33553154

11. Shafirovich V, Geacintov NE. Excision of oxidatively generated guanine lesions by competitive dna repair pathways. Int J Mol Sci. 2021; 22: 1–11. https://doi.org/10.3390/ijms22052698 PMID: 33800059

12. Ames BN. Mutagenesis and carcinogenesis: Endogenous and exogenous factors. Environ Mol Mutagen. 1989; 14: 66–77. https://doi.org/10.1002/em.2850140614 PMID: 2659334

13. Mullaart E, Lohman PHM, Berends F, Vijg J. DNA damage metabolism and aging. Mutat Res. 1990; 237: 189–210. https://doi.org/10.1016/0921-8734(90)90001-8 PMID: 2079959

14. Marnett LJ, Plastaras JP. Endogenous DNA damage and mutation. Trends Genet. 2001; 17: 214–221. https://doi.org/10.1016/s0168-9525(01)02239-9 PMID: 11275327

15. Hotchkiss RD. The Quantitative Seperation of Purines, Pyrimidines, and Nucleosides by Paper Chromatography. J Biol Chem. 1948; 175: 315:332. https://doi.org/10.1016/S0021-9258(18)57261-6

16. Razin A, Cedar H. DNA methylation and gene expression. Microbiol Rev. 1991; 55: 451–458. https://doi.org/10.1128/mr.55.3.451-458.1991 PMID: 1943996

17. Zingg JM, Jones PA. Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation and carcinogenesis. Carcinogenesis. 1997; 18: 869–882. https://doi.org/10.1093/carcin/18.5.869 PMID: 9163670

18. Meehan R, Lewis JD, Bird AP. Characterization of MECP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. 1992; 20: 5085–5092. https://doi.org/10.1093/nar/20.19.5085 PMID: 1408825

19. Bird AP, Wolffe AP. Methylation-induced repression-belts, braces, and chromatin. Cell. 1999; 99: 451–454. https://doi.org/10.1016/s0092-8674(00)81532-9 PMID: 10589672

20. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res. 2004; 32: 4100–4108. https://doi.org/10.1093/nar/gkh739 PMID: 15302911

21. Nguyen CT, Gonzales FA, Jones PA. Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: Correlation of accessibility, methylation, MeCP2 binding and acetylation. Nucleic Acids Res. 2001; 29: 4598–4606. https://doi.org/10.1093/nar/29.22.4598 PMID: 11713309

22. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem. 2003; 278: 4035–4040. https://doi.org/10.1074/jbc.M210256200 PMID: 12427740

23. Pradhan S, Bacolla A, Wells RD, Roberts RJ. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of novo and maintenance methylation. J Biol Chem. 1999; 274: 33002–33010. https://doi.org/10.1074/jbc.274.46.33002 PMID: 10551868

24. Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. 2007; 67: 946–950. https://doi.org/10.1158/0008-5472.CAN-06-3123 PMID: 17283125

25. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol. 2005; 45: 629–656. https://doi.org/10.1146/annurev.pharmtox.45.120403.095832 PMID: 15822191

26. Herman JG, Baylin SB. Gene Silencing in Cancer in Association with Promoter Hypermethylation. N Engl J Med. 2003; 349: 2042–2054. https://doi.org/10.1056/NEJMra023075 PMID: 14627790

27. Greenblatt MS, Bennett WP, Harris CC, Hollstein M. Mutations in the p53 Tumor Suppressor Gene: Clues to Cancer Etiology and Molecular Pathogenesis. Cancer Res. 1994; 54: 4855–4878. PMID: 8069852

28. Magewu AN, Jones PA. Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. Mol Cell Biol. 1994; 14: 4225–4232. https://doi.org/10.1128/mcb.14.6.4225-4232.1994 PMID: 8196660

29. Pfeifer GP, Besaratinia A. Mutational spectra of human cancer. Hum Genet. 2009; 125: 493–506. https://doi.org/10.1007/s00439-009-0657-2 PMID: 19308457

30. Iengar P. An analysis of substitution, deletion and insertion mutations in cancer genes. Nucleic Acids Res. 2012; 40: 6401–6413. https://doi.org/10.1093/nar/gks290 PMID: 22492711

31. Schmutte C, Yang AS, Beart RW, Jones PA. Base Excision Repair of U:G Mismatches at a Mutational Hotspot in the p53 Gene Is More Efficient Than Base Excision Repair of T:G Mismatches in Extracts of Human Colon Tumors. Cancer Res. 1995; 55: 3742–3746. PMID: 7641186

32. Alcántara R, Wang SY. Photochemistry of Thymine in Aqueous Solution. Photochem Photobiol. 1965; 4: 473–476. https://doi.org/10.1111/j.1751-1097.1965.tb09762.x

33. Teebor GW, Frenkel K, Goldstein MS. Ionizing radiation and tritium transmutation both cause formation of 5-hydroxymethyl-2'-deoxyuridine in cellular DNA. Proc Natl Acad Sci U S A. 1984; 81: 318–321. https://doi.org/10.1073/pnas.81.2.318 PMID: 6582490

34. Frenkel K, Cummings A, Solomon J, Cadet J, Steinberg JJ, Teebor GW. Quantitative Determination of the 5-(Hydroxymethyl)uracil Moiety in the DNA of γ-Irradiated Cells†. Biochemistry. 1985; 24: 4527–4533. https://doi.org/10.1021/bi00338a007 PMID: 3904829

35. Téoule R. Radiation-induced DNA damage and its repair. Int J Radiat Biol. 1987; 51: 573–589. https://doi.org/10.1080/09553008414552111 PMID: 3034813

36. Khattak MN, Green JH. Gamma-irradiation of nucleic-acid constituents in de-aerated aqueous solutions: II. 5-methyl cytosine. Int J Radiat Biol. 1966; 11: 137–143. https://doi.org/10.1080/09553006614550901 PMID: 5296548

37. Privat E, Sowers LC. Photochemical deamination and demethylation of 5-methylcytosine. Chem Res Toxicol. 1996; 9: 745–750. https://doi.org/10.1021/tx950182o PMID: 8831819

38. Castro GD, Díaz Gómez MI, Castro JA. 5-Methylcytosine attack by hydroxyl free radicals and during carbon tetrachloride promoted liver microsomal lipid peroxidation: Structure of reaction products. Chem Biol Interact. 1996; 99: 289–299. https://doi.org/10.1016/0009-2797(95)03680-6 PMID: 8620576

39. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (80-). 2009; 324: 930–935. https://doi.org/10.1126/science.1170116 PMID: 19372391

40. Ito S, Dalessio AC, Taranova O V., Hong K, Sowers LC, Zhang Y, et al. Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010; 466: 1129–1133. https://doi.org/10.1038/nature09303 PMID: 20639862

41. Wu SC, Zhang Y. Active DNA demethylation: Many roads lead to Rome. Nat Rev Mol Cell Biol. 2010; 11: 607–620. https://doi.org/10.1038/nrm2950 PMID: 20683471

42. Guo JU, Su Y, Zhong C, Ming GL, Song H. Emerging roles of TET proteins and 5-hydroxymethylcyto-sines in active DNA demethylation and beyond. Cell Cycle. 2011; 10: 2662–2668. https://doi.org/10.4161/cc.10.16.17093 PMID: 21811096

43. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. Science (80-). 2009; 324: 929–930. https://doi.org/10.1126/science.1169786 PMID: 19372393

44. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (80-). 2011; 333: 1300–1303. https://doi.org/10.1126/science.1210597 PMID: 21778364

45. Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydro-xymethylcytosine and search for active demethylation intermediates. PLoS One. 2010; 5: 1–9. https://doi.org/10.1371/journal.pone.0015367 PMID: 21203455

46. Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcyto-sine: Potential implications for active demethylation of CpG sites. J Biol Chem. 2011; 286: 35334–35338. https://doi.org/10.1074/jbc.C111.284620 PMID: 21862836

47. Sowers LC, Beardsley GP. Synthesis of oligonucleotides containing 5-(hydroxymethyl)-2'-deoxyuridine at defined sites. J Org Chem. 1993; 58: 1664–1665. https://doi.org/10.1021/jo00059a011 PMID: 35859906

48. Tardy-Planechaud S, Fujimoto J, Lin SS, Sowers LC. Solid phase synthesis and restriction endonucle-ase cleavage of oligodeoxynucleotides containing 5-(hydroxymethyl)-cytosine. Nucleic Acids Res. 1997; 25: 553–558. https://doi.org/10.1093/nar/25.3.553 PMID: 9016595

49. Rogstad DK, Liu P, Burdzy A, Lin SS, Sowers LC. Endogenous DNA lesions can inhibit the binding of the AP-1 (c-Jun) transcription factor. Biochemistry. 2002; 41: 8093–8102. https://doi.org/10.1021/bi012180a PMID: 12069602

50. Schutsky EK, Nabel CS, Davis AKF, DeNizio JE, Kohli RM. APOBEC3A efficiently deaminates methyl-ated, but not TET-oxidized, cytosine bases in DNA. Nucleic Acids Res. 2017; 45: 7655–7665. https://doi.org/10.1093/nar/gkx345 PMID: 28472485

51. Rusmintratip V, Sowers LC. An unexpectedly high excision capacity for mispaired 5-hydroxymethylura-cil in human cell extracts. Proc Natl Acad Sci U S A. 2000; 97: 14183–14187. https://doi.org/10.1073/pnas.97.26.14183 PMID: 11121024

52. Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell. 2011; 145: 423–434. https://doi.org/10.1016/j.cell.2011.03.022 PMID: 21496894

53. Bordin DL, Lirussi L, Nilsen H. Cellular response to endogenous DNA damage: DNA base modifications in gene expression regulation. DNA Repair (Amst). 2021; 99: 103051. https://doi.org/10.1016/j.dnarep.2021.103051 PMID: 33540225

54. Rangam G, Schmitz KM, Cobb AJA, Petersen-Mahrt SK. Aid enzymatic activity is inversely proportional to the size of cytosine c5 orbital cloud. PLoS One. 2012; 7: 3–8. https://doi.org/10.1371/journal.pone.0043279 PMID: 22916236

55. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, et al. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol. 2012; 8: 751–758. https://doi.org/10.1038/nchembio.1042 PMID: 22772155

56. Berthod T, Pétillot Y, Guy A, Cadet J, Forest E, Molko D. Synthesis and mass spectrometry analysis of oligonucleotides bearing 5- formyl-2'-deoxyuridine in their structure. Nucleosides and Nucleotides. 1996; 15: 1287–1305. https://doi.org/10.1080/07328319608002430

57. Rogstad DK, Heo J, Vaidehi N, Goddard WA, Burdzy A, Sowers LC. 5-Formyluracil-induced perturba-tions of DNA function. Biochemistry. 2004; 43: 5688–5697. https://doi.org/10.1021/bi030247j PMID: 15134443

58. Guerniou V, Gasparutto D, Sauvaigo S, Favier A, Cadet J. New synthesis of 5-carboxy-2′-deoxyuridine and its incorporation into synthetic oligonucleotides. Nucleosides, Nucleotides and Nucleic Acids. 2003; 22: 1073–1075. https://doi.org/10.1081/NCN-120022739 PMID: 14565347

59. Darwanto A, Theruvathu JA, Sowers JL, Rogstad DK, Pascal T, Goddard W, et al. Mechanisms of base selection by human single-stranded selective monofunctional uracil-DNA glycosylase. J Biol Chem. 2009; 284: 15835–15846. https://doi.org/10.1074/jbc.M807846200 PMID: 19324873

60. Schuermann D, Scheidegger SP, Weber AR, Bjørås M, Leumann CJ, Schär P. 3CAPS -A structural AP-site analogue as a tool to investigate DNA base excision repair. Nucleic Acids Res. 2016; 44: 2187–2198. https://doi.org/10.1093/nar/gkv1520 PMID: 26733580

61. Haushalter KA, Stukenberg PT, Kirschner MW, Verdine GL. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. Curr Biol. 1999; 9: 174–185. https://doi.org/10.1016/s0960-9822(99)80087-6 PMID: 10074426

62. Hendrich B, Bird A. Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. Mol Cell Biol. 1998; 18: 6538–6547. https://doi.org/10.1128/MCB.18.11.6538 PMID: 9774669

63. Hsu CW, Sowers ML, Baljinnyam T, Herring JL, Hackfeld LC, Tang H, et al. Measurement of deaminated cytosine adducts in DNA using a novel hybrid thymine DNA glycosylase. J Biol Chem. 2022; 298: 101638. https://doi.org/10.1016/j.jbc.2022.101638 PMID: 35085553

64. Begley TJ, Cunningham RP. Methanobacterium thermoformicicum thymine DNA mismatch glycosylase: conversion of an N-glycosylase to an AP lyase. Protein Eng Des Sel. 1999; 12: 333–340. https://doi.org/10.1093/protein/12.4.333 PMID: 10325404

65. Patel JP, Sowers ML, Herring JL, Theruvathu JA, Emmett MR, Hawkins BE, et al. Measurement of Postreplicative DNA Metabolism and Damage in the Rodent Brain. Chem Res Toxicol. 2015; 28: 2352–2363. https://doi.org/10.1021/acs.chemrestox.5b00359 PMID: 26447562

66. Münzel M, Lischke U, Stathis D, Pfaffeneder T, Gnerlich FA, DeimL CA, et al. Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. Chem—A Eur J. 2011; 17: 13782–13788. https://doi.org/10.1002/chem.201102782 PMID: 22069110

67. Dai Q, He C. Syntheses of 5-formyl- and 5-carboxyl-dc containing DNA oligos as potential oxidation products of 5-hydroxymethylcytosine in DNA. Org Lett. 2011; 13: 3446–3449. https://doi.org/10.1021/ol201189n PMID: 21648398

68. Xuan S, Wu Q, Cui L, Zhang D, Shao F. 5-Hydroxymethylcytosine and 5-formylcytosine containing deoxyoligonucleotides: Facile syntheses and melting temperature studies. Bioorganic Med Chem Lett. 2015; 25: 1186–1191. https://doi.org/10.1016/j.bmcl.2015.01.070 PMID: 25704892

69. Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, et al. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol. 2012; 8: 328–330. https://doi.org/10.1038/nchembio.914 PMID: 22327402

70. Cadet J, Wagner JR. TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. Mutat Res—Genet Toxicol Environ Mutagen. 2014; 764–765: 18–35. https://doi.org/10.1016/j.mrgentox.2013.09.001 PMID: 24045206

71. He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science (80-). 2011; 333: 1303–1307. https://doi.org/10.1126/science.1210944 PMID: 21817016

72. Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: Its structural basis and implications for active DNA demethylation. Nucleic Acids Res. 2012; 40: 10203–10214. https://doi.org/10.1093/nar/gks845 PMID: 22962365

73. Pidugu LS, Dai Q, Malik SS, Pozharski E, Drohat AC. Excision of 5-Carboxylcytosine by Thymine DNA Glycosylase. J Am Chem Soc. 2019; 141: 18851–18861. https://doi.org/10.1021/jacs.9b10376 PMID: 31693361

74. Schormann N, Ricciardi R, Chattopadhyay D. Uracil-DNA glycosylases—Structural and functional perspectives on an essential family of DNA repair enzymes. Protein Sci. 2014; 23: 1667–1685. https://doi.org/10.1002/pro.2554 PMID: 25252105

75. Pearl LH. Structure and function in the uracil-DNA glycosylase superfamily. Mutat Res—DNA Repair. 2000; 460: 165–181. https://doi.org/10.1016/s0921-8777(00)00025-2 PMID: 10946227

76. Kavli B, Slupphaug G, Mol CD, Arvai AS, Petersen SB, Tainer JA, et al. Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. EMBO J. 1996; 15: 3442–3447. https://doi.org/10.1002/j.1460-2075.1996.tb00710.x PMID: 8670846

77. Jancso A, Botfield MC, Sowers LC, Weiss MA. An altered-specificity mutation in a human POU domain demonstrates functional analogy between the POU-specific subdomain and phage lambda repressor. Proc Natl Acad Sci. 1994; 91: 3887–3891. https://doi.org/10.1073/pnas.91.9.3887 PMID: 8171007

78. Masaoka A, Matsubara M, Hasegawa R, Tanaka T, Kurisu S, Terato H, et al. Mammalian 5-Formyluracil—DNA Glycosylase. 2. Role of SMUG1 Uracil—DNA Glycosylase in Repair of 5-Formyluracil and Other Oxidized and. Biochemistry. 2003; 42: 5003–5012.

79. Wibley JEA, Waters TR, Haushalter K, Verdine GL, Pearl LH. Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. Mol Cell. 2003; 11: 1647–1659. https://doi.org/10.1016/s1097-2765(03)00235-1 PMID: 12820976

80. Matsubara M, Tanaka T, Terato H, Ohmae E, Izumi S, Katayanagi K, et al. Mutationl analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase. Nucleic Acids Res. 2004; 32: 5291–5302. https://doi.org/10.1093/nar/gkh859 PMID: 15466595

81. Knvælsrud I, Slupphaug G, Leiros I, Matsuda A, Ruoff P, Bjelland S. Opposite-base dependent excision of 5-formyluracil from DNA by hSMUG1. Int J Radiat Biol. 2009; 85: 413–420. https://doi.org/10.1080/09553000902818915 PMID: 19365746

82. Bennett MT, Rodgers MT, Hebert AS, Ruslander LE, Eisele L, Drohat AC. Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. J Am Chem Soc. 2006; 128: 12510–12519. https://doi.org/10.1021/ja0634829 PMID: 16984202

83. Morgan MT, Bennett MT, Drohat AC. Excision of 5-halogenated uracils by human thymine DNA glyco-sylase: Robust activity for DNA contexts other than CpG. J Biol Chem. 2007; 282: 27578–27586. https://doi.org/10.1074/jbc.M704253200 PMID: 17602166

84. Morgan MT, Maiti A, Fitzgerald ME, Drohat AC. Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. Nucleic Acids Res. 2011; 39: 2319–2329. https://doi.org/10.1093/nar/gkq1164 PMID: 21097883

85. Hardeland U, Bentele M, Jiricny J, Schar P. Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. J Biol Chem. 2000; 275: 33449–33456. https://doi.org/10.1074/jbc.M005095200 PMID: 10938281

86. Petronzelli F, Riccio A, Markham GD, Seeholzer SH, Genuardi M, Karbowski M, et al. Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): Funda-mental role of the catalytic domain. J Cell Physiol. 2000; 185: 473–480. https://doi.org/10.1002/1097-4652(200012)185:3<473::AID-JCP19>3.0.CO;2-# PMID: 11056019

87. Zhang W, Liu Z, Crombet L, Amaya MF, Liu Y, Zhang X, et al. Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4. Biochem Biophys Res Com-mun. 2011; 412: 425–428. https://doi.org/10.1016/j.bbrc.2011.07.091 PMID: 21820404

88. Moréra S, Grin I, Vigouroux A, Couvé S, Henriot V, Saparbaev M, et al. Biochemical and structural char-acterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA. Nucleic Acids Res. 2012; 40: 9917–9926. https://doi.org/10.1093/nar/gks714 PMID: 22848106

89. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. 2012; 40: 4841–4849. https://doi.org/10.1093/nar/gks155 PMID: 22362737

90. Hashimoto H, Zhang X, Cheng X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA gly-cosylase domain: Structural basis and implications for active DNA demethylation. Nucleic Acids Res. 2012; 40: 8276–8284. https://doi.org/10.1093/nar/gks628 PMID: 22740654

91. Guerniou V, Rapin D, Millau JF, Bufflier E, Favier A, Cadet J, et al. Repair of oxidative damage of thy-mine by HeLa whole-cell extracts: Simultaneous analysis using a microsupport and comparison with traditional PAGE analysis. Biochimie. 2005; 87: 151–159. https://doi.org/10.1016/j.biochi.2004.12.001 PMID: 15760707

92. Shen J-C, Rideout WM, Jones PA. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res. 1994; 22: 972–976. https://doi.org/10.1093/nar/22.6.972 PMID: 8152929

93. Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LSS. Structure of the Human MutSα DNA Lesion Recognition Complex. Mol Cell. 2007; 26: 579–592. https://doi.org/10.1016/j.molcel.2007.04.018 PMID: 17531815

94. Neidigh JW, Darwanto A, Williams AA, Wall NR, Sowers LC. Cloning and characterization of Rhodotor-ula glutinis thymine hydroxylase. Chem Res Toxicol. 2009; 22: 885–893. https://doi.org/10.1021/tx8004482 PMID: 19341313

95. Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat Chem Biol. 2014; 10: 574–581. https://doi.org/10.1038/nchembio.1532 PMID: 24838012

96. Pais JE, Dai N, Tamanaha E, Vaisvila R, Fomenkov AI, Bitinaite J, et al. Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. Proc Natl Acad Sci U S A. 2015; 112: 4316–4321. https://doi.org/10.1073/pnas.1417939112 PMID: 25831492