


Article

Demethylation of Non-CpG Sites in DNA Is Initiated by TET2 5-Methylcytosine Dioxygenase

Aninda Sundar Dey ^{1,†} , Chayan Bhattacharya ^{1,†}, Yihong Guan ² , Babal Kant Jha ²  and Mridul Mukherji ^{1,*}

¹ Division of Pharmacology & Pharmaceutical Sciences, University of Missouri-Kansas City, Kansas City, MO 64112, USA; asdmr9@mail.umkc.edu (A.S.D.); cbg66@mail.umkc.edu (C.B.)

² Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH 44195, USA; guany2@ccf.org (Y.G.); jhab@ccf.org (B.K.J.)

* Correspondence: mukherjim@umkc.edu

† These authors contributed equally to this work.

Abstract: In the mammalian genome, cytosine methylation predominantly occurs at CpG sites. In addition, a number of recent studies have uncovered extensive C5 cytosine methylation (5mC) at non-CpG (5mCpH, where H = A/C/T) sites. Little is known about the enzyme responsible for active demethylation of 5mCpH sites. Using a very sensitive and quantitative LC-MS/MS method, we demonstrate that the human TET2, an iron (II)- and 2OG-dependent dioxygenase, which is a frequently mutated gene in several myeloid malignancies, as well as in a number of other types of cancers, can oxidize 5mCpH sites in double-stranded DNA in vitro. Similar to oxidation of 5mCpG, oxidation of 5mC at CpH sites produces 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) bases in DNA. After 5mCpG, which is the most preferred substrate, TET2 prefers 5mCpC as a substrate, followed by 5mCpA and then 5mCpT. Since the TDG/BER pathway and deformylation or decarboxylation of 5fC or 5caC, respectively, can convert 5fCpH and 5caCpH to an unmodified cytosine base in DNA, our results suggest a novel demethylation pathway of 5mCpH sites initiated by TET2 dioxygenase.

Keywords: epigenetics; DNA demethylation; TET dioxygenases; substrate specificity; 5-methylcytosine oxidation; non-CpG demethylation



Citation: Dey, A.S.; Bhattacharya, C.; Guan, Y.; Jha, B.K.; Mukherji, M. Demethylation of Non-CpG Sites in DNA Is Initiated by TET2 5-Methylcytosine Dioxygenase. *DNA* **2021**, *1*, 26–36. <https://doi.org/10.3390/dna1010004>

Academic Editor: Benoît Miotto

Received: 21 May 2021

Accepted: 6 September 2021

Published: 21 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Methylation of cytosine at carbon-5 (5mC) within the CpG dinucleotide (5mCpG) in DNA plays crucial roles in X-chromosome inactivation, gene imprinting, nuclear reprogramming, and tissue-specific gene expression in mammalian cells [1–3]. Dynamic epigenetic regulation of 5mCpG also plays critical roles in multiple stages of pluripotency, differentiation, and development [4]. There appear to be more than 20 million CpG dinucleotides in the human genome with 70–80% of CpG cytosines being methylated [5]. A delicate balance between cytosine methylation and demethylation within CpG dinucleotides shapes the final epigenetic pattern of a cell, and any imbalance in methylation patterns results in pathological conditions, including cancer [6,7].

In addition to 5mCpG marks, several recent genome-wide bisulfite sequencing studies at single-base resolution have uncovered a significant amount of 5mC at CpH sites in almost all human cells [8–12]. Particularly, in human embryonic stem cell lines, 67.85%, 6.68%, 1.48%, and 0.63% of all CpG, CpA, CpT, and CpC sites are methylated, respectively [11]. In brain cells, neurons, and different embryonic stem cells (ESCs), one quarter of all methylation marks identified were on CpH sites [8–11,13]. Generally, as in the case of 5mCpG marks, the presence of 5mCpH sites also inversely correlated with transcription [8–11,13]. In addition, the 5mCpH marks correlate with genomic imprinting and regulation of inter-chromosomal interactions between enhancer elements and receptor genes, and are predictive of genes that escape X-inactivation [1,14]. Further, several recent

studies have demonstrated a correlation between non-CpG methylation and gene expression in different types of cancers [15,16]. Finally, a number of studies have reported the presence of 5hmC in the context of CpH in human cells [17,18]. Collectively, these initial studies suggest critical roles for 5mCpH marks in the mammalian genome.

The 5mC modification within the CpG dinucleotide is carried out by DNA methyltransferases (DNMTs) [19], while its removal is initiated by the ten-eleven translocation family of dioxygenases (TET1–3) [20,21]. TET proteins, which belong to a class of iron (II)- and 2-oxoglutarate (2OG)-dependent dioxygenases, catalyze 5mC oxidation to form 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) sequentially, in a three-step oxidation process [22,23]. Finally, the 5fC and 5caC residues are replaced with unmodified cytosine bases by the thymine DNA glycosylase (TDG)/base excision repair (BER) pathway [23,24]. Although the role of TET2 (and its homologs, TET1 and TET3) in the oxidation of 5-methylcytosine (5mC) at CpG sites has been extensively studied, its role in 5mC oxidation at the non-5mCpG (or 5mCpH) sites has not been established.

Interestingly, in vitro substrate specificity studies with DNMT3A using oligonucleotides demonstrated that this enzyme could methylate CpH marks [25,26]. Further in vivo studies have established a strong correlation between the presence of 5mCpH marks and the expression of DNMT3A, DNMT3B, and DNMT3L genes [27–29]. Finally, studies in ESCs provided convincing evidence that DNMT3A and DNMT3B methylate CpH sites [30,31]. Although passive demethylation of 5mCpH sites during cell division may occur, in some instances dynamic active demethylation has been observed [13,17,32,33]. However, Hu et al. demonstrated that, although the human TET2 enzyme can rapidly oxidize 5mCpG sites, oxidation of both the 5mCpA and 5mCpC sites was negligible [34]. In contrast to human TET2 dioxygenase, a TET homolog from *N. gruberi*, NgTET1, which is structurally very similar to the human TET2, can oxidize DNA containing 5mCpG and 5mCpA sites with comparable efficiency [35]. These conflicting results raise questions with regards to the tolerance of bases at the +1 position in TET-mediated 5mC oxidation of DNA substrates.

In this study, using a very sensitive LC–MS/MS-based assay, we show that human TET2 can oxidize 5mCpH sites in double-stranded DNA. Similar to oxidation of 5mCpG, oxidation of 5mC in CpH produces 5hmC, 5fC, and 5caC oxidation marks. Since the human TDG enzyme can excise 5fC and 5caC from CpH sites, allowing their replacement with unmodified cytosine, along with the recently established deformylation and decarboxylation pathways, our results demonstrate a novel demethylation pathway of 5mCpH sites initiated by TET2 dioxygenase [36–40]. These results may also help to elucidate the emerging role of non-CpG methylation in gene expression and cancer.

2. Materials and Methods

2.1. Chemicals and Reagents

Analytical grade chemicals were purchased from Thermo Fisher Scientific or Sigma-Aldrich. All molecular biology kits were obtained from Superior Scientifics (Lenexa, KS, USA). Nucleoside standards and growth media were supplied by Carbosynth (Now Biosynth Carbosynth (Compton, UK)) and Difco Laboratories (Detroit, MI, USA), respectively.

2.2. Purification of TET2 Dioxygenase

2.2.1. Purification of hTET2 Catalytic Domain

In our earlier publication, we reported the cloning and expression of the human TET2 catalytic domain (TET2 1129–1936, Δ1481–1843) [41]. Purification of the TET2 catalytic domain was performed using cation exchange chromatography, which is described in our previous publication [42].

2.2.2. Purification of GST-Tagged hTET2 Catalytic Domain

Firstly, the GST-tagged catalytic domain of TET2 (TET2 1129–1936, Δ 1481–1843) was incorporated into pGEX4T-1 and, then, recombinant vectors were transformed into *E. coli* BL21(DE3) cells for protein expression. The BL21(DE3) cells were resuspended in lysis buffer (PBS: 10 mM Na_2HPO_4 , 1.8 mM KH_2PO_4 , 140 mM NaCl) at pH 7.5 and sonicated. Glutathione agarose resin was used for affinity tag TET2 purification, using a centrifugal force method. Elution of the protein was carried out with an elution buffer (50 mM Tris-HCl, 33 mM reduced glutathione, pH 7.5). Selected fractions were concentrated to 10 mL using ultrafiltration membranes (NMWL = 30 kDa).

2.2.3. Purification of Full-Length Myc-Tagged hTET2

The full-length *hTET2* was cloned with an N-terminal Myc-tag, as described earlier, and purified on a large scale with minor modifications [43]. Briefly, the Myc-TET2 was ectopically expressed in human embryonic kidney 293 cells, obtained from ATCC (Manassas, VA, USA), by plasmid transfection for 3 days. Protein was extracted in the RIPA buffer supplemented with 300 mM NaCl and 1X protease inhibitor. Protein concentration was measured using the BCA assay and diluted to 1 mg/mL, with 1X PBS for affinity purification using Myc-tag antibody-conjugated magnetic beads (Cell Signaling, Danvers, MA, USA). The Myc-TET2 protein was eluted using 0.5 mg/mL c-Myc peptide in Tris-buffered saline containing 0.05% Tween-20 for 5–10 min at 37 °C.

2.3. 5mC Oxidation by TET2 Dioxygenase

25-mer double-stranded DNA was used for the TET2 enzymatic assay. The substrates contain only one 5-methylcytosine residue on one strand only. The specific sequences of the substrates used are provided in Table 1A–C and Table 2. The reactions were performed in triplicate. Each reaction was carried out in 100 μL total reaction volume, containing 3 μg of substrate, 500 μg of TET2 enzyme, 75 μM FeSO_4 , 1 mM 2-oxoglutarate, 5 mM ascorbate, and 50 mM HEPES (pH 8.0). The enzymatic reactions of the GST-tagged enzyme (100 μg) and full-length TET2 enzyme were carried out under identical conditions. After incubating at 37 °C for 30 min, the reactions were stopped using 5 μL 500mM EDTA. Using an oligo purification technique, the DNA was separated from the reaction. Then, the DNA (20 μL) was digested with 2 units of DNase I and 60 units of S1 Nuclease in 40 μL reaction volume. The reactions were incubated for 12 h at 37 °C to generate individual nucleoside monophosphates. After this step, 2 units of calf-intestinal alkaline phosphatase (CIAP) were added to the reactions. After 4 h of incubation at 37 °C, individual nucleosides were produced. Thirty-five μL of each sample was analyzed by LC–MS/MS.

2.4. MS/MS Parameters for Nucleosides

A Sciex 3200 QTrap mass spectrometer (Foster City, CA, USA), coupled with a Shimadzu UFLC LC-20 system (Columbia, MD, USA), was used for LC–MS/MS. The system was run by Analyst v1.6.2 software. In positive mode, a C18 column (serial number: 0201449T)—dimensions: 150 \times 2 mm; particle size: 5 μM ; pore size: 100 Å—was used. In negative mode, a C18 column (serial number: 020439T)—dimensions: 100 \times 2 mm; particle size: 5 μM ; pore size: 60 Å—was used. These two columns were purchased from Keystone Scientific (Thermo Fischer Scientific, Lenexa, KS, USA). MS/MS parameters for nucleosides was determined by our previously described method [42]. Mass spectrometry parameters are provided under Supplementary Materials (Tables S1–S3).

Table 1. List of double-stranded DNA sequences containing 5mC sites used in this study.

1A	
DNA Substrates	Sequence
5mCpG Substrate 1	Forward Strand 5'-GCGCCGGTCGTA/Me-C/GGCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGCCGTACGACCGGCGC-3'
5mCpA Substrate 1	Forward Strand 5'-GCGCCGGTCGTA/Me-C/AGCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGCTGTACGACCGGCGC-3'
1B	
DNA Substrates	Sequence
5mCpG Substrate 2	Forward Strand 5'-GCGCCGGTCCCG/Me-C/GGCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGCCCGCGGGACCGGCGC-3'
5mCpG Substrate 3	Forward Strand 5'-GCGCCGGTCGTT/Me-C/GTCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGACGAACGACCGGCGC-3'
5mCpG Substrate 4	Forward Strand 5'-GCGCCGGTCAAC/Me-C/GACCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGTCCGTTGACCGGCGC-3'
5mCpA Substrate 2	Forward Strand 5'-GCGCCGGTCTAA/Me-C/AAACGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGTTTGTAGACCGGCGC-3'
5mCpA Substrate 3	Forward Strand 5'-GCGCCGGTCGTT/Me-C/ATCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGATGAACGACCGGCGC-3'
5mCpA Substrate 4	Forward Strand 5'-GCGCCGGTCGGA/Me-C/AACCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGTTGTCCGACCGGCGC-3'
1C	
DNA Substrates	Sequence
5mCpC Substrate	Forward Strand 5'-GCGCCGGTCCTG/Me-C/CCCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGGGCGAGGACCGGCGC-3'
5mCpT Substrate	Forward Strand 5'-GCGCCGGTCCTG/Me-C/TCCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGGGAGCAGGACCGGCGC-3'

2.5. Liquid Chromatographic Conditions for Nucleosides under Different MS/MS Modes

A neutral solvent system was used for chromatographic separation in negative mode. Solvent A was 10 mM ammonium acetate (pH 6.5), and solvent B was 80% acetonitrile/20% 10 mM ammonium acetate (pH 6.5). The gradient was 0% solvent B at 0–2 min, 0–20% at 2–5 min, 20–60% at 5–9 min, 60–0% solvent B at 9–10 min and a 5 min post-equilibration with solvent A. The overall flow rate was 0.3 mL/min. A water/methanol-based solvent system was used for chromatographic separation in positive mode. Solvent A was water (adjusted to pH 3.5 using formic acid) and solvent B was methanol (adjusted to pH 3.0 using formic acid). The gradient was 0% solvent B at 0–1 min, 0–2% at 1–12 min, 2–30% at 12–17 min, 30% at 17–18 min, 30–0% solvent B at 18–18.5 min and a 3.5 min post-equilibration with solvent A. The overall flow rate was 0.3 mL/min. Standard curves, the limit of detection (LOD), the lower limit of quantification (LLOQ), and the matrix effect for all eight nucleosides in the positive and negative modes were calculated using our developed method [42]. Mass spectrometry parameters and standard curves for modified cytosines are provided under Supplementary Materials (Tables S1–S3 and Figures S1 and S2). An *in vitro* TET2 enzymatic assay for each substrate was performed in triplicate. The amount of product formed during TET2 oxidation reactions across different DNA substrates was normalized by calculating the peak area of each product (i.e., 5hmC or 5fC or 5caC) and dividing it by the area represented by one deoxycytidine residue. For statistical analysis, standard deviation and standard error were calculated for each oxidative product generated in the assays performed in triplicate. Calculated standard error is represented in figures and tables.

3. Results and Discussion

3.1. Detection of Cytosine Derivatives Produced by Oxidative Demethylation of 5mCpG and 5mCpA Sites

Despite the critical roles played by cytosine derivatives (5hmC, 5fC, and 5caC) in human health and disease, in vitro quantification of these recently identified nucleic acid bases is still in an early stage. Recently, we described an improved positive/negative ion-switching-based LC–MS/MS method that can separate and quantify these modified cytosine bases [42]. This method improved the detection sensitivity for 5hmC and 5fC 6–8-fold, whereas 32-fold improvement was observed in the case of the 5caC base, compared to an earlier method [34]. Using this sensitive method, we determined the activity of the TET2 catalytic domain, using a 25-mer dsDNA containing one 5mC residue in a CpG mark (Table 1A). After TET2 enzymatic reactions, the DNA was purified and converted into nucleosides. The nucleoside mixtures were subjected to mass spectrometry analysis using our previously described method [42]. In the TET2 enzymatic reactions, in addition to A, T, G, C, and 5mC peaks, three new peaks corresponding to 5hmC and 5fC were observed in the positive mode, and 5caC was observed in the negative mode (Figure 1A).

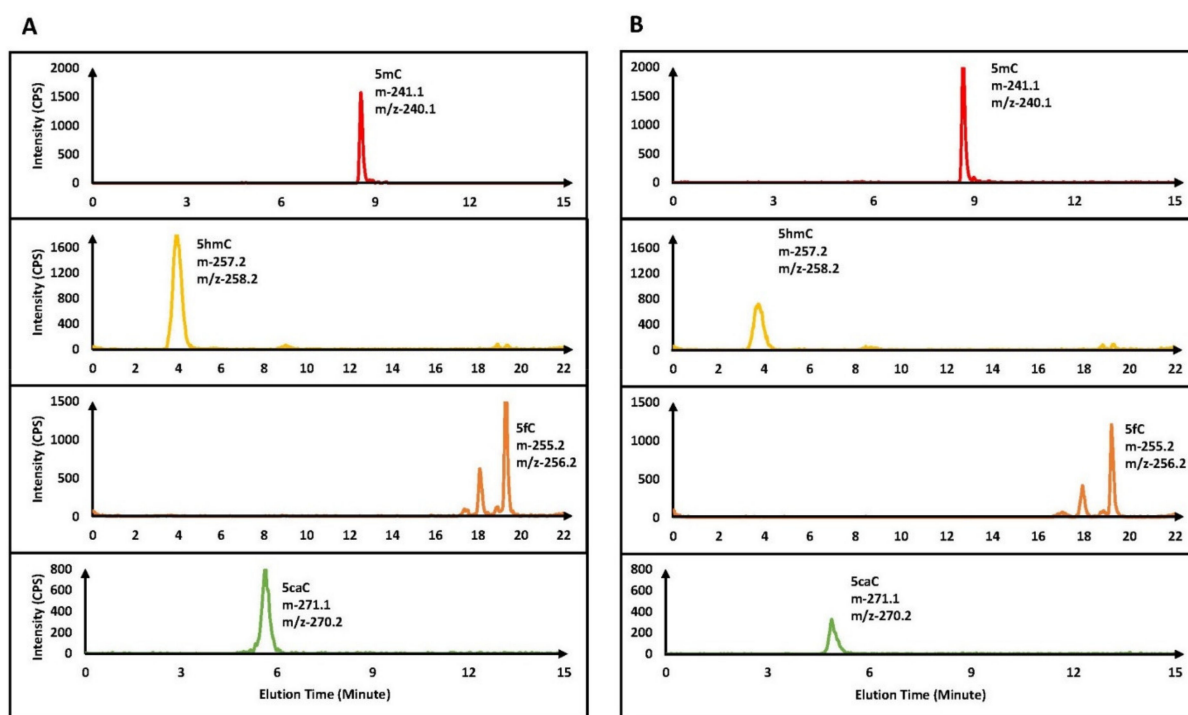


Figure 1. Top panels are for the negative control (without TET2 enzyme), which showed only 5mC but no 5hmC, 5fC, and 5caC. Lower panels are for the MRM chromatograms of 5mC, 5hmC, 5fC, and 5caC in TET2 catalyzed oxidation of DNA containing 5mCpG (A) and 5mCpA (B) sites. Please note that 5mC and 5caC were detected under the negative ion mode using a 15 min chromatographic gradient, while 5hmC and 5fC were detected under the positive ion mode using a 22 min gradient.

Further, using this sensitive method, we determined the activity of the TET2 catalytic domain using a 25-mer dsDNA containing one 5mC residue in a CpA mark (Table 1A). To our surprise, similar to the oxidation of 5mC in CpG sites, three new peaks were observed in the LC–MS/MS analysis. These novel peaks in the TET2 enzymatic reactions using 5mCpA as a substrate had identical mass characteristics and elution profiles to the 5hmC, 5fC, and 5caC bases (i.e., 5hmC and 5fC observed in the positive mode, while 5caC was observed in the negative mode) [42] (Figure 1B).

3.2. Quantification of Cytosine Derivatives Produced by TET2-Mediated Oxidation of 5mCpH Sites; Effect of +1 Base

The oxidation of 5mC in the CpA site (~50% oxidation product compared to 5mCpG) was surprising, because a previous study reported that the human TET2 predominantly oxidized 5mCpG sites (>85% 5mC oxidized), while the oxidation of 5mCpA was negligible (<2% 5mC oxidized) [34]. The previous observation was reasoned by a specific hydrogen bond between the phosphate groups of +1G of the 5mCpG mark with the S1290-TET2 residue [34]. The crystal structure further demonstrated a base-stacking interaction between the +1G:C' base pair (where the C' is present in the reverse strand, opposite the +1G base) of DNA and the Y1294-TET2 residue, resulting in specific recognition of the +1G:C' base pair in the 5mCpG dinucleotide by TET2.

To validate our initial observation that TET2 can significantly oxidize 5mCpA sites, we tested the activity of the TET2 catalytic domain with three additional sequences with 5mCpA sites (Table 1B). For comparison, we studied TET2 activity with similar substrates (i.e., length and sequences) which included 5mCpG sites instead of 5mCpA sites in our study (Table 1). Repeatedly, we observed that TET2 can significantly oxidize 5mC in CpA (~50%) compared to 5mCpG (Figure 2). Specifically, little difference was observed in the formation of 5hmC and 5fC, while the amount of 5caC, in the case of 5mCpA, was significantly less than in the case of 5mCpG. Intrigued by these results, we further tested the activity of the TET2 catalytic domain with dsDNA sequences containing one 5mC residue in CpC and CpT sites (Table 1C). Indeed, our experiments demonstrated that the TET2 catalytic domain can oxidize 5mC in CpH sites (Figure 3). Our results demonstrated that, after 5mCpG, TET2 prefers 5mCpC over 5mCpA, with substrates with 5mCpT being the least preferred. Preference for the 5mCpC substrate over 5mCpA and 5mCpT could be due to a base-stacking interaction between the +1C:G' base pair of the substrate and the Y1294-TET2 residue, as suggested by the crystal structure of the TET2 enzyme [34].

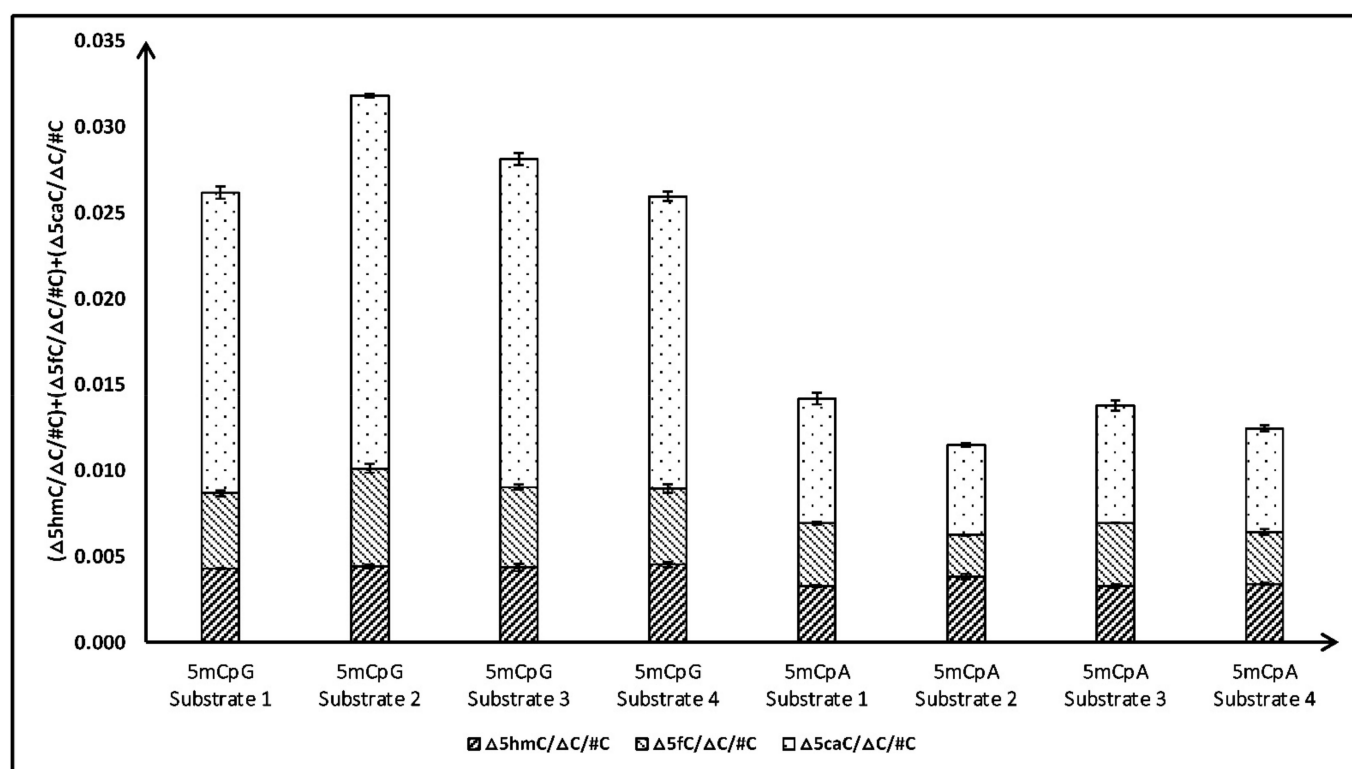


Figure 2. Comparison of total products (5hmC + 5fC + 5caC) formed after TET2 catalyzed oxidation of DNA containing 5mCpG and 5mCpA sites.

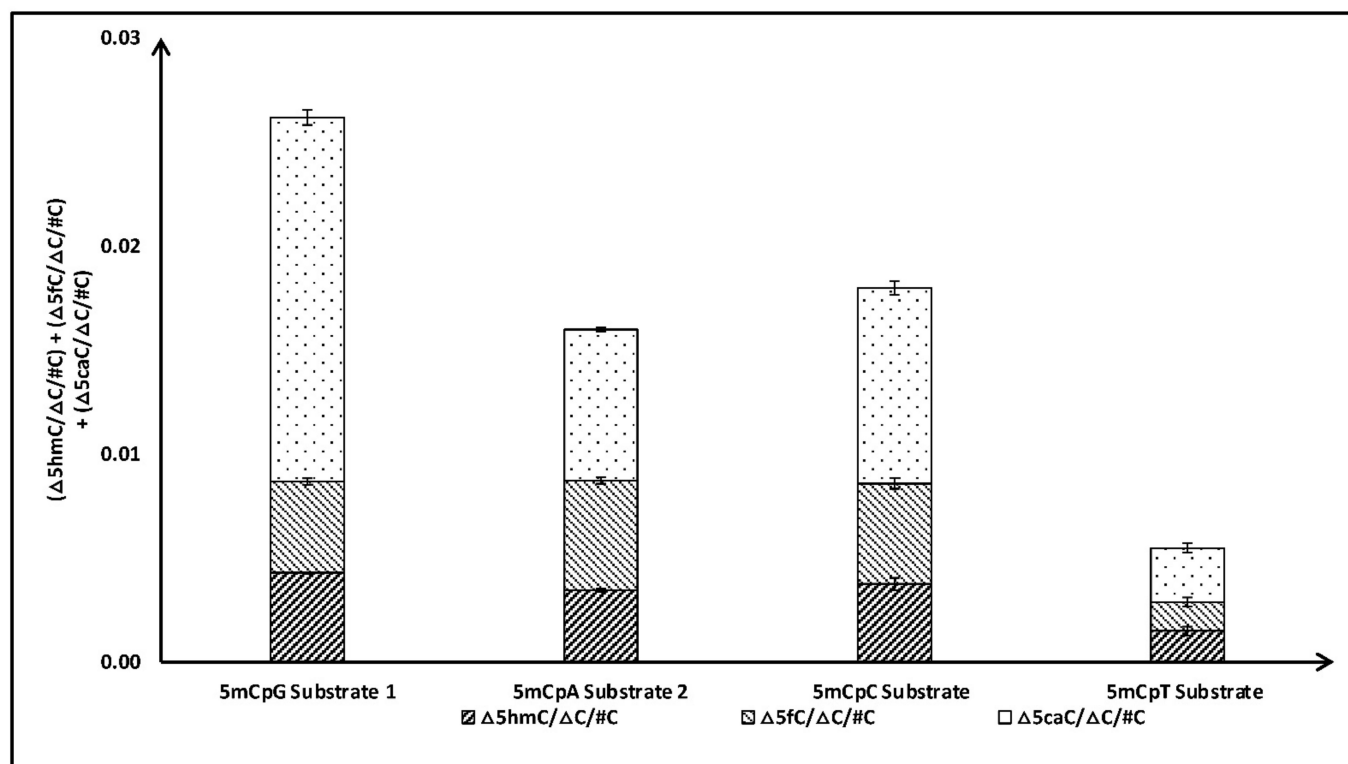


Figure 3. Comparison of total products (5hmC + 5fC + 5caC) formed after TET2 catalyzed oxidation of DNA containing 5mCpG, 5mCpA, 5mCpC, and 5mCpT sites.

3.3. TET2 Can Initiate Demethylation of 5mCpH Sites

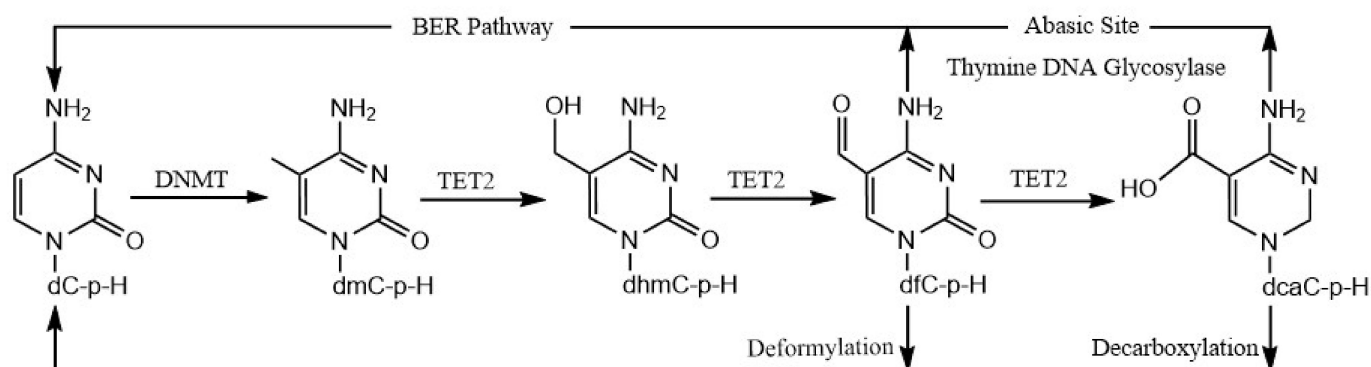
The TET2 catalytic domain (TET2 1129–1936, Δ1481–1843) was identified as the minimal catalytically active domain present on the C-terminus of full-length human TET2 enzyme [34]. Hu et al. produced this minimal catalytically active TET2 domain as a GST-tagged protein in bacteria to solve its structure. Since this clone is an engineered TET2 protein and does not present the native enzyme, we produced full-length human TET2 enzyme in human embryonic kidney 293 cells. To compare the activity of the untagged TET2 catalytic domain, which was predominantly used in this study, we performed oxidation assays with the GST-TET2 catalytic domain and full-length human TET2, using 5mCpG and 5mCpA (which is the most common CpH methylation found in the human genome) as substrates. Indeed, both the GST-TET2 catalytic domain and full-length human TET2 oxidized 5mC on CpA sites in DNA and converted it into 5hmC, 5fC, and 5caC (Table 2). Thus, our results confirm that human TET2 dioxygenase can carry out iterative oxidation of 5mC in CpA dinucleotides, similar to the oxidation of 5mCpG sites in DNA.

During the course of this submission, DeNizio et al. demonstrated, consistently with our results, that TET2 can oxidize 5mCpH sites, albeit less efficiently than 5mCpG [40]. Similar to our results, they demonstrated that oxidation of 5mCpH produces 5hmCpH, 5fCpH, and 5caCpH. Interestingly, DeNizio et al. demonstrated that the rate of formation of 5hmC in CpG substrates is higher than in CpH substrates; however, there was no difference in the rate of formation of 5fC and 5caC in substrates with CpG vs. CpH. DeNizio et al. further demonstrated that the order of TET2 substrate preference, in the case of 5mCpH oxidation, is 5mCpH ≥ 5hmCpH > 5fCpH, which is slightly different from the oxidation of 5mCpG (i.e., 5mCpG > 5hmCpG > 5fCpG).

Table 2. List of double stranded DNA sequences containing 5mC in CpG and CpA context, TET2 enzymes used in the enzymatic assay and the oxidation products.

2A				
DNA Substrate		DNA Sequence		
5mCpG		Forward Strand 5'-GCGCCGGTCCCG/Me-C/GGGCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGTTTGTTAGACCGGCGC-3		
Enzyme		Full Length human TET2		
dC Area (Counts) ± Standard Error	5hmC Area (Counts) ± Standard Error	5fC Area (Counts) ± Standard Error	5caC Area (Counts) ± Standard Error	
(2.00 × 10 ⁷) ± (6.11 × 10 ⁵)	(3.11 × 10 ⁴) ± (2.17 × 10 ³)	(2.25 × 10 ⁴) ± (1.29 × 10 ³)	(6.12 × 10 ³) ± (3.71 × 10 ²)	
Enzyme		GST Tagged human TET2		
dC Area (Counts)	5hmC Area (Counts)	5fC Area (Counts)	5caC Area (Counts)	
(6.63 × 10 ⁷) ± (3.33 × 10 ⁵)	(3.17 × 10 ⁵) ± (2.11 × 10 ⁴)	(2.50 × 10 ⁵) ± (2.67 × 10 ³)	(7.37 × 10 ⁴) ± (1.89 × 10 ³)	
2B				
DNA Substrate		DNA Sequence		
5mCpA		Forward Strand 5'-GCGCCGGTCTCTG/Me-C/ACCCGCTCCCGC-3' Reverse Strand 5'-GCGGGAGCGTTTGTAGACCGGCGC-3		
Enzyme		Full Length human TET2		
dC Area (Counts) ± Standard Error	5hmC Area (Counts) ± Standard Error	5fC Area (Counts) ± Standard Error	5caC Area (Counts) ± Standard Error	
(1.56 × 10 ⁷) ± (6.43 × 10 ⁵)	(1.91 × 10 ⁴) ± (1.17 × 10 ³)	(1.67 × 10 ⁴) ± (1.00 × 10 ³)	(2.89 × 10 ³) ± (4.88 × 10 ²)	
Enzyme		GST Tagged human TET2		
dC Area (Counts)	5hmC Area (Counts)	5fC Area (Counts)	5caC Area (Counts)	
(7.73 × 10 ⁷) ± (3.33 × 10 ⁵)	(2.17 × 10 ⁵) ± (1.18 × 10 ⁴)	(1.41 × 10 ⁵) ± (3.67 × 10 ³)	(2.37 × 10 ⁴) ± (1.65 × 10 ³)	

A previous study reported that, in addition to excising 5fC and 5caC residues in CpG sites and replacing them with unmodified cytosine bases by the TDG/BER pathway [39], it can act on 5fC and 5caC residues in CpH sites [36]. DeNizio et al. further reestablished, similarly to a previously published study by Papin et al. [36], that TDG can excise fC and caC in CpH contexts [36,40], which may allow the replacement of these modified cytosine bases with cytosine. Further, in a recent, elegant study, an alternate demethylation process demonstrated that 5fC and 5caC deformylate and decarboxylate, respectively, to directly reinstall unmodified cytosine bases at previously methylated sites [37,38]. As such, this pathway will not distinguish the +1 base between the CpG and CpH sites. Taken together, our results suggest a novel demethylation of 5mCpH sites through the oxidation of 5mCpH sites by TET2 into 5hmCpH, 5fCpH, and 5caCpH, followed by the replacement of 5fCpH and 5caCpH by either the TDG/BER pathway or direct C–C bond cleavage (Figure 4).

**Figure 4.** Schematic representation of 5mCpH demethylation mediated via TET2 dioxygenase.

4. Conclusions

TET2 is one of the most frequently mutated genes in different types of leukemia, including myelodysplastic syndromes (MDS), MDS-myeloproliferative neoplasms (MDS-MPN), and acute myeloid leukemia derived from MDS and MDS-MPN (sAML). Mutations in the *TET2* gene have been reported in many diverse solid tumors as well. The *TET2* gene codes for 5-methylcytosine dioxygenase, an enzyme that epigenetically regulates 5mC levels in the genome. Although the role of TET2 in the oxidation of 5mC at CpG sites was well documented in the literature, its role in the demethylation of 5mC at CpH sites remained questionable. Using a sensitive LC–MS/MS-based method in this study, we provided multiple lines of evidence that the human TET2 can oxidize 5mCpH sites in DNA. Like oxidation of 5mCpG, the oxidation of 5mC at CpH sites produces 5hmC, 5fC, and 5caC marks in DNA. The preferred order of TET2 substrates is 5mCpG > 5mCpC > 5mCpA >> 5mCpT. Since the TDG/BER pathway and deamination or decarboxylation of 5fC or 5caC, respectively, can replace 5fCpH and 5caCpH with an unmodified cytosine base in DNA, our results suggest a novel demethylation of 5mCpH sites by TET2 dioxygenase. Since our study, along with the study by DeNizio et al., was performed in vitro, in vivo confirmation of our results will provide more insights regarding the demethylation of 5mCpH sites by TET2. TET2 uses a base flipping mechanism for its interaction with DNA, like many other DNA modifying enzymes [34,44]. The activity of some of these base flipping enzymes depends on flanking bases [45,46]. Since Hu et al. studied non-CpG demethylation using only one DNA sequence, further follow-up investigations based on our study reported here, along with crystallography, will help to improve our understanding of non-CpG demethylation [34].

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/dna1010004/s1>, Figure S1: Standard curves of different nucleosides in negative mode, Figure S2: Standard curves of different nucleosides in positive mode, Table S1: The chemical structure of four different modified deoxyribonucleosides, indicating the charge fragmentation position and the mass transition used for MRM detection, Table S2: Mass spectrometric parameters for the most intense MS/MS transitions of the eight nucleosides in negative mode detection, Table S3: Mass spectrometric parameters for the most intense MS/MS transitions of the eight nucleosides in positive mode detection.

Author Contributions: A.S.D., C.B., and Y.G. performed most of the experiments. A.S.D., C.B., and M.M. designed the experiments. A.S.D., C.B., B.K.J., and M.M. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US Department of Defense in the form of an Idea Award (W81XWH-13-1-0174), an Aplastic Anemia & MDS Foundation Grant, and a UMRB grant to M.M. and Leukemia Lymphoma Society Translational Research Program grant 6582-20-LLS to B.K.J.

Institutional Review Board Statement: This research was reviewed and approved by the relevant institutional committees.

Acknowledgments: The authors would like to thank William Gutheil and Navid Ayon for their help with mass spectrometry.

Conflicts of Interest: The authors declare that they have no known conflicts of interest.

Abbreviations

TET, ten-eleven translocation; 5mdC/5mC, 5-methyldeoxycytidine/5-methylcytosine; 5hmdC/5hmC, 5-hydroxymethyldeoxycytidine/5-hydroxymethylcytosine; 5fdC/5fC, 5-formyldeoxycytidine/5-formylcytosine; 5cadC/5caC, 5-carboxydeoxycytidine/5-carboxycytosine; 2OG, 2-oxoglutarate; DN-MTs, DNA methyl transferases; TDG, thymine DNA glycosylase; BER, base excision repair; LC–MS/MS, liquid chromatography–tandem mass spectrometry.

References

- Schultz, M.D.; He, Y.; Whitaker, J.W.; Hariharan, M.; Mukamel, E.; Leung, D.; Rajagopal, N.; Nery, J.R.; Urich, M.A.; Chen, H.; et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **2015**, *523*, 212–216. [[CrossRef](#)] [[PubMed](#)]
- Bonasio, R.; Tu, S.; Reinberg, D. Molecular Signals of Epigenetic States. *Science* **2010**, *330*, 612–616. [[CrossRef](#)] [[PubMed](#)]
- Feng, S.; Jacobsen, S.E.; Reik, W. Epigenetic Reprogramming in Plant and Animal Development. *Science* **2010**, *330*, 622–627. [[CrossRef](#)] [[PubMed](#)]
- Smith, Z.D.; Meissner, A. DNA methylation: Roles in mammalian development. *Nat. Rev. Genet.* **2013**, *14*, 204–220. [[CrossRef](#)]
- Jabbari, K.; Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **2004**, *333*, 143–149. [[CrossRef](#)]
- Portela, A.; Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **2010**, *28*, 1057–1068. [[CrossRef](#)]
- Song, C.-X.; He, C. Balance of DNA methylation and demethylation in cancer development. *Genome Biol.* **2012**, *13*, 173. [[CrossRef](#)]
- Lister, R.; Pelizzola, M.; Downen, R.H.; Hawkins, R.D.; Hon, G.; Tonti-Filippini, J.; Nery, J.R.; Lee, L.; Ye, Z.; Ngo, Q.-M.; et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **2009**, *462*, 315–322. [[CrossRef](#)]
- Laurent, L.; Wong, E.; Li, G.; Huynh, T.; Tsigos, A.; Ong, C.T.; Low, H.M.; Sung, W.-K.; Rigoutsos, I.; Loring, J.; et al. Dynamic changes in the human methylome during differentiation. *Genome Res.* **2010**, *20*, 320–331. [[CrossRef](#)]
- Lister, R.; Pelizzola, M.; Kida, Y.; Hawkins, R.D.; Nery, J.R.; Hon, G.; Antosiewicz-Bourget, J.; O'Malley, R.; Castanon, R.; Klugman, S.; et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **2011**, *471*, 68–73. [[CrossRef](#)] [[PubMed](#)]
- Ziller, M.J.; Müller, F.; Liao, J.; Zhang, Y.; Gu, H.; Bock, C.; Boyle, P.; Epstein, C.B.; Bernstein, B.E.; Lengauer, T.; et al. Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genet.* **2011**, *7*, e1002389. [[CrossRef](#)] [[PubMed](#)]
- Ravichandran, M.; Jurkowska, R.Z.; Jurkowski, T.P. Target specificity of mammalian DNA methylation and demethylation machinery. *Org. Biomol. Chem.* **2017**, *16*, 1419–1435. [[CrossRef](#)] [[PubMed](#)]
- Guo, J.U.; Su, Y.; Shin, J.H.; Shin, J.; Li, H.; Xie, B.; Zhong, C.; Hu, S.; Le, T.; Fan, G.; et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **2013**, *17*, 215–222. [[CrossRef](#)] [[PubMed](#)]
- Patil, V.; Ward, R.; Hesson, L.B. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* **2014**, *9*, 823–828. [[CrossRef](#)]
- Xiao, W.; Liu, X.; Niu, X.; Li, C.; Guo, Y.; Tan, J.; Xiong, W.; Fan, L.; Li, Y. The frequency of CpG and non-CpG methylation of Notch3 gene promoter determines its expression levels in breast cancer cells. *Exp. Cell Res.* **2019**, *386*, 111743. [[CrossRef](#)]
- Li, C.; Xiong, W.; Liu, X.; Xiao, W.; Guo, Y.; Tan, J.; Li, Y. Hypomethylation at non-CpG/CpG sites in the promoter of HIF-1 α gene combined with enhanced H3K9Ac modification contribute to maintain higher HIF-1 α expression in breast cancer. *Oncogenesis* **2019**, *8*, 26. [[CrossRef](#)] [[PubMed](#)]
- Kinde, B.; Gabel, H.W.; Gilbert, C.S.; Griffith, E.; Greenberg, M.E. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc. Natl. Acad. Sci.* **2015**, *112*, 6800–6806. [[CrossRef](#)]
- Sun, Z.; Dai, N.; Borgaro, J.G.; Quimby, A.; Sun, D.; Correa, I.; Zheng, Y.; Zhu, Z.; Guan, S. A Sensitive Approach to Map Genome-wide 5-Hydroxymethylcytosine and 5-Formylcytosine at Single-Base Resolution. *Mol. Cell* **2015**, *57*, 750–761. [[CrossRef](#)]
- Bestor, T.H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **2000**, *9*, 2395–2402. [[CrossRef](#)] [[PubMed](#)]
- Tahiliani, M.; Koh, K.P.; Shen, Y.; Pastor, W.A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L.M.; Liu, D.R.; Aravind, L.; et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **2009**, *324*, 930–935. [[CrossRef](#)] [[PubMed](#)]
- Schübeler, D. Function and information content of DNA methylation. *Nature* **2015**, *517*, 321–326. [[CrossRef](#)] [[PubMed](#)]
- Ito, S.; Shen, L.; Dai, Q.; Wu, S.C.; Collins, L.B.; Swenberg, J.A.; He, C.; Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **2011**, *333*, 1300–1303. [[CrossRef](#)]
- He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; et al. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **2011**, *333*, 1303–1307. [[CrossRef](#)] [[PubMed](#)]
- Ponnaluri, V.C.; Maciejewski, J.P.; Mukherji, M. A mechanistic overview of TET-mediated 5-methylcytosine oxidation. *Biochem. Biophys. Res. Commun.* **2013**, *436*, 115–120. [[CrossRef](#)]
- Ramsahoye, B.H.; Biniszkiwicz, D.; Lyko, F.; Clark, V.; Bird, A.; Jaenisch, R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5237–5242. [[CrossRef](#)]
- Gowher, H.; Jeltsch, A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: The enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. *J. Mol. Biol.* **2001**, *309*, 1201–1208. [[CrossRef](#)]
- Shirane, K.; Toh, H.; Kobayashi, H.; Miura, F.; Chiba, H.; Ito, T.; Kono, T.; Sasaki, H. Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. *PLoS Genet.* **2013**, *9*, e1003439. [[CrossRef](#)]
- Barres, R.; Osler, M.E.; Yan, J.; Rune, A.; Fritz, T.; Caidahl, K.; Krook, A.; Zierath, J.R. Non-CpG Methylation of the PGC-1 α Promoter through DNMT3B Controls Mitochondrial Density. *Cell Metab.* **2009**, *10*, 189–198. [[CrossRef](#)] [[PubMed](#)]
- Ichiyanagi, T.; Ichiyanagi, K.; Miyake, M.; Sasaki, H. Accumulation and loss of asymmetric non-CpG methylation during male germ-cell development. *Nucleic Acids Res.* **2012**, *41*, 738–745. [[CrossRef](#)] [[PubMed](#)]

30. Arand, J.; Spieler, D.; Karius, T.; Branco, M.; Meilinger, D.; Meissner, A.; Jenuwein, T.; Xu, G.; Leonhardt, H.; Wolf, V.; et al. In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genet.* **2012**, *8*, e1002750. [[CrossRef](#)] [[PubMed](#)]
31. Lee, J.-H.; Park, S.-J.; Nakai, K. Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. *Sci. Rep.* **2017**, *7*, 1–11. [[CrossRef](#)] [[PubMed](#)]
32. Fusco, A.; Ferraguti, G.; Grandoni, F.; Ruggeri, R.; Scarpa, S.; Strom, R.; Lucarelli, M. Early demethylation of non-CpG, CpC-rich, elements in the myogenin 5'-flanking region. *Cell Cycle* **2010**, *9*, 3965–3976. [[CrossRef](#)] [[PubMed](#)]
33. Tognini, P.; Napoli, D.; Pizzorusso, T. Dynamic DNA methylation in the brain: A new epigenetic mark for experience-dependent plasticity. *Front. Cell. Neurosci.* **2015**, *9*, 331. [[CrossRef](#)] [[PubMed](#)]
34. Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y.G.; Zhu, J.; Wang, P.; Xu, Y. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **2013**, *155*, 1545–1555. [[CrossRef](#)] [[PubMed](#)]
35. Pais, J.E.; Dai, N.; Tamanaha, E.; Vaisvila, R.; Fomenkov, A.I.; Bitinaite, J.; Sun, Z.; Guan, S.; Correa, I.; Noren, C.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. USA* **2015**, *112*, 4316–4321. [[CrossRef](#)] [[PubMed](#)]
36. Papin, C.; Ibrahim, A.; Le Gras, S.; Velt, A.; Stoll, I.; Jost, B.; Menoni, H.; Bronner, C.; Dimitrov, S.; Hamiche, A. Combinatorial DNA methylation codes at repetitive elements. *Genome Res.* **2017**, *27*, 934–946. [[CrossRef](#)]
37. Feng, Y.; Xie, N.; Tao, W.; Ding, J.; You, X.; Ma, C.; Zhang, X.; Yi, C.; Zhou, X.; Yuan, B.; et al. Transformation of 5-Carboxylcytosine to Cytosine Through C–C Bond Cleavage in Human Cells Constitutes a Novel Pathway for DNA De-methylation. *CCS Chem.* **2020**, *2*, 994–1008. [[CrossRef](#)]
38. Iwan, K.; Rahimoff, R.; Kirchner, A.; Spada, F.; Schröder, A.S.; Kosmatchev, O.; Ferizaj, S.; Steinbacher, J.; Parsa, E.; Müller, M.; et al. 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **2017**, *14*, 72–78. [[CrossRef](#)]
39. 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.* **2011**, *286*, 35334–35338. [[CrossRef](#)]
40. DeNizio, J.E.; Dow, B.J.; Serrano, J.C.; Ghanty, U.; Drohat, A.C.; Kohli, R.M. TET-TDG Active DNA Demethylation at CpG and Non-CpG Sites. *J. Mol. Biol.* **2021**, *433*, 166877. [[CrossRef](#)]
41. Bhattacharya, C.; Dey, A.; Ayon, N.J.; Gutheil, W.G.; Mukherji, I. Efficient Purification and LC-MS/MS-based Assay Development for Ten-Eleven Translocation-2 5-Methylcytosine Dioxygenase. *J. Vis. Exp.* **2018**, *140*, e57798. [[CrossRef](#)] [[PubMed](#)]
42. Dey, A.S.; Ayon, N.J.; Bhattacharya, C.; Gutheil, W.G.; Mukherji, M. Positive/negative ion-switching-based LC-MS/MS method for quantification of cytosine derivatives produced by the TET-family 5-methylcytosine dioxygenases. *Biol. Methods Protoc.* **2020**, *5*, bpaa019. [[CrossRef](#)]
43. Guan, Y.; Greenberg, E.F.; Hasipek, M.; Chen, S.; Liu, X.; Kerr, C.M.; Gackowski, D.; Zarakowska, E.; Radivoyevitch, T.; Gu, X.; et al. Context dependent effects of ascorbic acid treatment in TET2 mutant myeloid neoplasia. *Commun. Biol.* **2020**, *3*, 1–13. [[CrossRef](#)]
44. Hong, S.; Cheng, X. DNA Base Flipping: A General Mechanism for Writing, Reading, and Erasing DNA Modifications. *Adv. Exp. Med. Biol.* **2016**, *945*, 321–341. [[CrossRef](#)]
45. Handa, V.; Jeltsch, A. Profound Flanking Sequence Preference of Dnmt3a and Dnmt3b Mammalian DNA Methyltransferases Shape the Human Epigenome. *J. Mol. Biol.* **2005**, *348*, 1103–1112. [[CrossRef](#)] [[PubMed](#)]
46. Adam, S.; Anteneh, H.; Hornisch, M.; Wagner, V.; Lu, J.; Radde, N.E.; Bashtrykov, P.; Song, J.; Jeltsch, A. DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. *Nat. Commun.* **2020**, *11*, 1–15. [[CrossRef](#)] [[PubMed](#)]