Supporting Information

Pais et al. 10.1073/pnas.1417939112

SI Materials and Methods

Expression and Purification of NgTET1. A pTXB1 construct of the N-terminal 6× histidine-tagged full-length NgTET1 was expressed and purified as described previously, using a HiTrap Heparin HP column followed by a HisTrap HP column (GE-Healthcare) (1). The isolated protein with >95% homogeneity, as determined by Coomassie-stained PAGE, was stored at –20 °C in 20 mM Tris pH 7.5, 300 mM sodium chloride (NaCl), and 50% (vol/vol) glycerol. Site-directed mutagenesis was performed using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs), and all variant proteins were expressed and purified using the same protocol as the WT protein.

Preparation of DNA Substrates. Single-stranded (ss) oligonucleotides (oligos) were synthesized by IDT, and duplex oligos were prepared by annealing equimolar amounts of complementary ss-oligos at 95 °C for 3 min followed by slow cooling at room temperature for 1 h. Plasmid DNA substrates used in this study are either commercially available from New England Biolabs or isolated from *Escherichia coli* cells, which were either dcm^+ for C^{5m}CWGG-methylated (where W = A or T) plasmids, or dcm^- for plasmids harboring a methyltransferase gene to achieve in vivo methylation at the desired target sites on the plasmid. One plasmid substrate, pXba DNA, was in vitro methylated using M.HaeIII methyltransferase (New England Biolabs) following the recommended protocol. Mammalian gDNA and *Helicobacter pylori* 26695 gDNA (*H. pylori* gDNA) are available from New England Biolabs and ATCC, respectively. Other gDNA substrates used were isolated from *Escherichia coli* expressing different methyltransferase genes following standard procedures. All gDNA was sheared using a Covaris S2 Focused-ultrasonicator, according to the manufacturer's guidelines.

NgTET1 Activity Assay. For all NgTET1 reactions, substrate DNA was added to buffer containing 50 mM NaCl, 1 mM DTT, 2 mM ascorbic acid, 1 mM α-ketoglutarate (αKG), and 100 μM ferrous sulfate ($FeSO_4$). Unless otherwise noted, 50 mM Bis-Tris buffer at pH 6.0 was used; other buffers used were 50 mM Tris or 50 mM 3-(*N*-morpholino)propanesulfonic acid (Mops buffer) at the indicated pH values. Reactions had a total volume of 50 μL and were incubated for 1 h at 34 °C, unless otherwise noted, and subsequently incubated with 0.8 units (U) proteinase K (New England Biolabs) for 1 h at 50 °C. For time courses, reactions were incubated at 34 °C for the specified time, and then quenched by heating at 95 °C for 3 min. The samples were then placed on ice for 5 min, followed by digestion with proteinase K. DNA was purified using either a QIAquick Nucleotide Removal Kit (Qiagen) or a DNA Clean & Concentrator Kit (Zymo Research), and was analyzed for oxidized ^{5m}C modifications using either a restriction enzyme (RE)-based assay or a LC-MS–based assay (see below).

Activity assays with the C-terminal catalytic domain of mouse TET1 (mTET1CD) were performed according to the manufacturer's instructions (WiseGene). Reaction time points were quenched as described for NgTET1 time courses. All other DNA treatment steps were as described for NgTET1 reactions.

RE-Based NgTET1 Activity Assay. Purified DNA (300 ng) from each NgTET1 reaction, performed in Tris buffer (pH 6.6) for 30 min, was digested with 20 U of BamHI (New England Biolabs) (to linearize the plasmid) and 50 U of MspI (New England Biolabs) in New England Biolabs CutSmart buffer (pH 7.9) for 1 h at 37 °C in 20 μL total volume. The reaction products were resolved on a 1.8% agarose gel.

LC-MS–Based NgTET1 Activity Assay. Each DNA sample from the NgTET1 reaction (see above) was digested to nucleosides, as described previously (1). LC-MS analysis was performed as indicated either on an Agilent 1200 series [G1316A UV Detector, 6120 Mass Detector, Agilent equipped with a Waters Atlantis T3 column (4.6 × 150 mm, 3 μm; Waters) with in-line filter and guard] or on a 6490 Triple Quad LC-MS [1290 Infinity UV detector, 6490 Triple Quad Mass detector (Agilent), equipped with an XSelect HSS T3 column (2.1 × 100mm, 2.5 μm; Waters)]. The relative percentages of the reaction species were calculated according to the standard curves generated from intensity peak integration.

SMRT Sequencing. DNA was sheared to 2-kb fragments by a Covaris system, concentrated using the DNA Clean & Concentrator Kit (Zymo Research), and suspended in milliQ H_2O. Fragments were then treated with NgTET1 (or buffer only for the negative control), using 20-μL total volume containing 1 μg DNA and 10 μM NgTET1 (when present) in Mops buffer, pH 6.75. Reactions were incubated for 1 h at 34 °C, followed by the addition of 18 μL of 50 mM Tris, pH 8.0, and 0.8 U proteinase K for 1 h at 50 °C. Samples treated with T4-β-glucosyltransferase (T4-βGT; New England Biolabs) were first subjected to chemical reduction with 100 mM sodium borohydride ($NaBH_4$) following the proteinase K digestion step, and then 500 U/mL T4-βGT in NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) containing 50 μM uridine diphosphate glucose. These reactions were incubated for 1 h at 37 °C. All DNA fragments were then purified using the DNA Clean & Concentrator Kit and used to prepare SMRTbell template libraries, following the Procedure and Checklist for 2-kb Template Preparation and Sequencing from Pacific Biosciences, starting with an additional DNA repair step before the end repair step.

Analysis of SMRT Sequencing Data. Base modification detection and motif analysis was performed using the "RS_Modification_and_Motif_Analysis.1" protocol included in the SMRT Portal v2.2.0. The GenBank/RefSeq entry NC_000915.1 was used as a reference for *H. pylori* 26695, and the samples treated without NgTET1 were entered as control jobs in the SMRT portal to allow for comparative sequencing and detection of ^{5m}C modifications. A kinetic score threshold of quality value 30 was used for all samples. One or eight SMRT cells (each taking two 60-min movies) were used to sequence pRS(M.HpaII) plasmid DNA or *H. pylori* 26695 gDNA, respectively, yielding an average reference coverage of 5110 or 150. IPD ratio values were extracted from the raw modifications.csv files and the detected fractions were extracted from the modifications.gff files.

1. Hashimoto H, et al. (2014) Structure of a *Naegleria* Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* 506(7488):391–395.

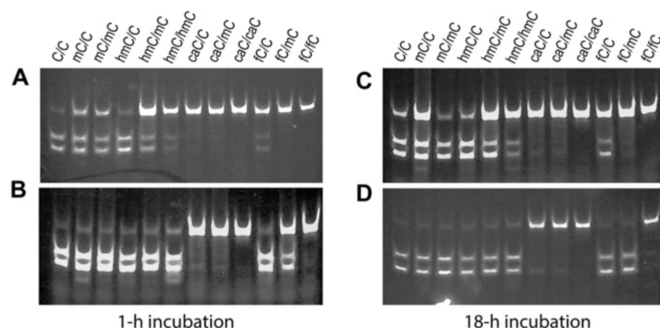

Fig. S1. MspI activity on oligo substrates containing ^{5m}C, ^{5hm}C, ^{5f}C, and ^{5ca}C. Fluorescently labeled 45-bp ss-oligos containing either unmodified (CCGG) or ^{5m}C-modified (C^{5m}CGG) sequences were made by oligo synthesis and were annealed with the respective counterparts to generate ds-oligos with the respective MspI recognition site CCGG/CCGG (referred to as C/C), C^{5m}CGG/CCGG (referred to as mC/C), and C^{5m}CGG/C^{5m}CGG (referred to as mC/mC) (Table S4, oligos 1–3). The fluorescently labeled ds-oligos containing a single ^{5hm}C, ^{5f}C or ^{5ca}C residue either on one or both strands were synthesized following a previously described protocol (1). The resulting ds-oligos (Table S4, oligos 4–12) contained modified MspI sites as follows: C^{5hm}CGG/CCGG (referred to as hmC/C); C^{5hm}CGG/C^{5hm}CGG (referred to as hmC/hmC); C^{5ca}CGG/CCGG (referred to as caC/C); C^{5ca}CGG/C^{5ca}CGG (referred to as caC/caC); C^{5f}CGG/CCGG (referred to as fC/C); C^{5f}CGG/C^{5f}CGG (referred to as fC/fC). Each ds-oligo substrate (10 pmol/10 μL NEBuffer 4) was cleaved with either 2 U (*A* and *C*) or 20 U (*B* and *D*) of MspI restriction endonuclease. The reactions were carried out at 37 °C for either 1 h (*A* and *B*) or 18 h (*C* and *D*). Reaction products were separated by electrophoresis on 10–20% polyacrylamide gels and visualized under UV light.

1. Kinney SM, et al. (2011) Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *J Biol Chem* 286(28):24685–24693.

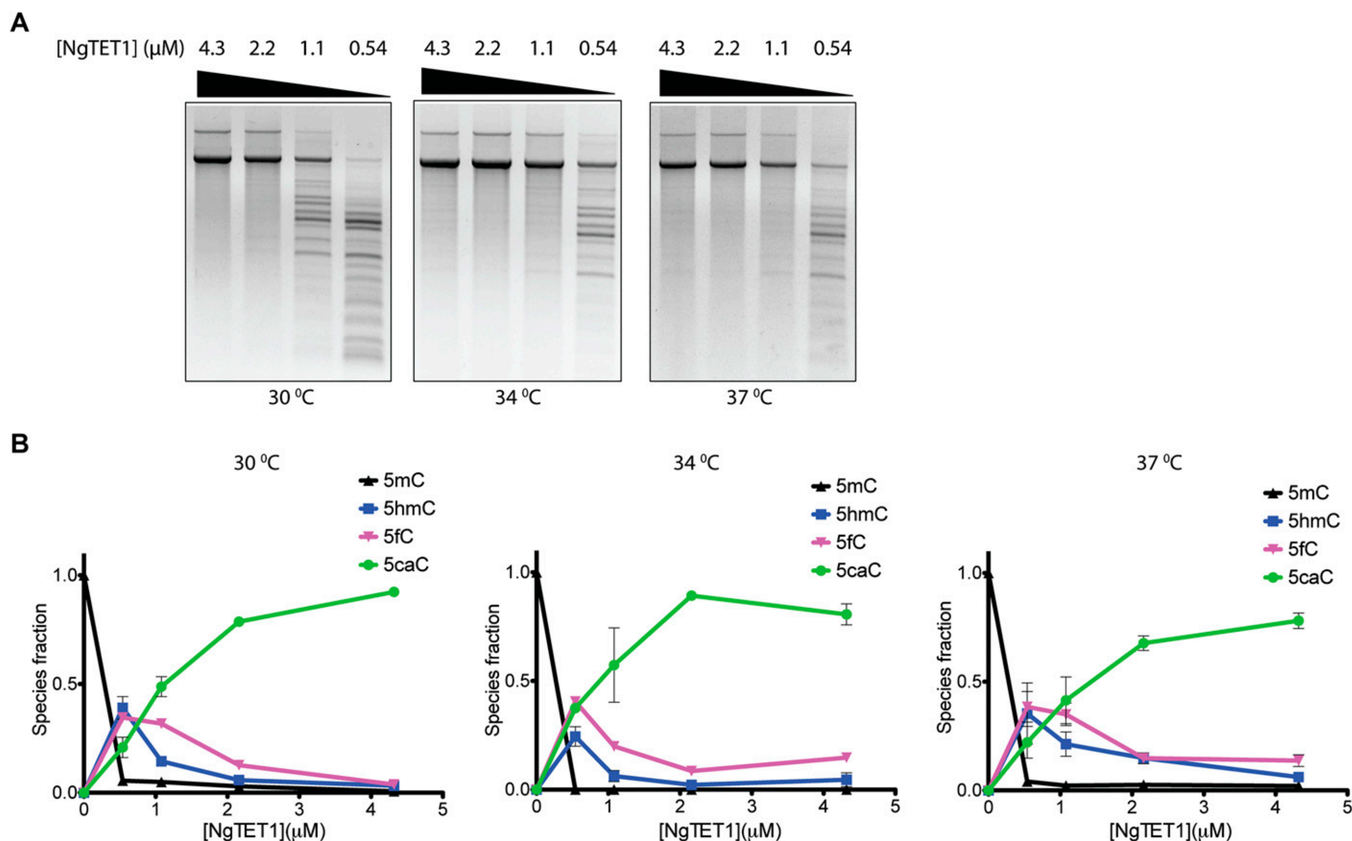

Fig. S2. Temperature dependence of NgTET1. (*A*) MspI-based assay to measure NgTET1 reaction efficiency at 30 °C, 34 °C, or 37 °C. NgTET1 reactions were carried out using 1 μg of pRS(M.HpaII) plasmid DNA at 34 °C for 30 min in 50 mM Tris, pH 6.6, 50 mM NaCl, 1 mM DTT, 2 mM ascorbic acid, 1 mM αKG, and 100 μM $FeSO_4$ in 50-μL total volume. DNA was purified and 300 ng were subjected to digestion with 20 U of BamHI (to linearize the plasmid) and 50 U of MspI in New England Biolabs CutSmart buffer (pH 7.9) for 1 h at 37 °C in 20-μL total volume. (*B*) Quantification of NgTET1 reaction products at different temperatures using the LC-MS–based assay, at varying concentrations of enzyme. NgTET1 reactions were conducted as described in *A* for the pRS(M.HpaII) plasmid. Purified DNA was digested to nucleosides and analyzed on an Agilent 1200 series (G1316A UV Detector, 6120 Mass Detector) with a Waters Atlantis T3 column (4.6 × 150 mm, 3 μM; Waters) with in-line filter and guard.

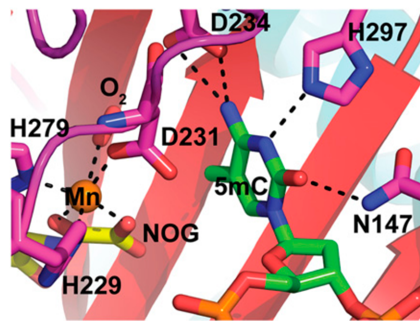

Fig. S3. View of active site and interactions with ^{5m}C in the crystal structure of NgTET1 in complex with DNA (PDB ID code 4LT5) (1).

1. Hashimoto H, et al. (2014) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* 506(7488):391–395.

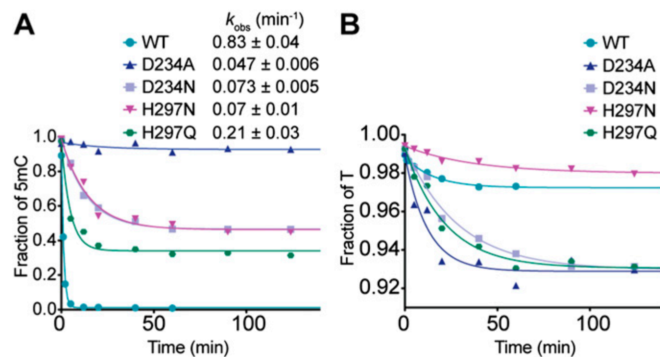

Fig. S4. Kinetic traces plotting the decay of (*A*) ^{5m}C or (*B*) T for WT compared with variants, using 4 μM NgTET1 WT or variant proteins with 2 μM oligo C^{5m}CGG. Reaction species were detected and quantified by LC-MS (Agilent 1200). The data are fit to a single exponential and the observed rate constants with SEM are provided.

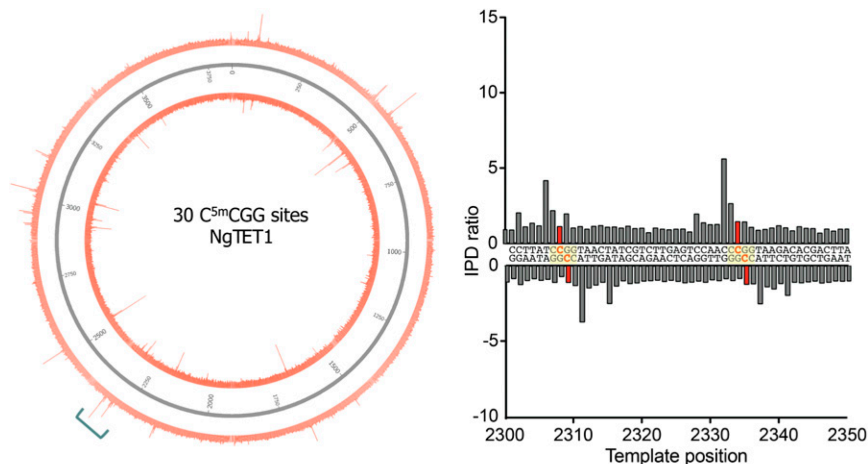

Fig. S5. Circos plot showing the plasmid-wide view of IPD ratios for a C^{5m}CGG methylated plasmid, pRS(M.HpaII), treated with NgTET1 and subjected to SMRT sequencing. The outer and inner circles denote the forward and reverse strands, respectively, and the brackets indicate the 50-bp segment shown at base resolution in the graph on the right, where the methylation motifs are highlighted in yellow and the methylated position is shown in red.

Fig. S6. LC-MS traces depicting the conversion of ^{5m}C to ^{5ca}C and ^{5gm}C using *E. coli* MG1655 genomic DNA before and after treatment with NgTET1 followed by addition of $NaBH_4$ and T4-βGT. The estimated final product distribution achieved using this protocol was 68% ^{5gm}C, 3% ^{5m}C, and 39% ^{5ca}C. Reactions were carried out in 50 mM Bis-Tris, pH 6, 50 mM NaCl, 1 mM DTT, 2 mM ascorbic acid, 1 mM αKG, and 100 μM $FeSO_4$ using 2 μg of 1.5-kb sheared DNA and 20 μM NgTET1 in 50-μL total volume. Following the proteinase K digestion step, the samples were reduced with 100 mM $NaBH_4$ and incubated for 1 h at 37 °C in NEBuffer 4 containing 30U T4-βGT and 50 μM UDP-glucose. LC-MS analysis was done on an Agilent 1200 series (G1316A UV Detector, 6120 Mass Detector) with a Waters Atlantis T3 column (4.6 × 150 mm, 3 μM; Waters) with in-line filter and guard.

Fig. S7. IPD ratio plots for the sequences detected (*Left*) versus undetected (*Right*) as belonging to the ^{5m}CCTC methylation motif for *H. pylori* gDNA samples treated with NgTET1 followed by $NaBH_4$/T4-βGT treatment. NgTET1 reactions and SMRT sequencing were performed as described in *SI Materials and Methods*.

Table S1. List of oligo substrate names and sequences used in this study

Name	Figure	Oligonucleotide sequence
24X	Fig. 1*C*	5′-TACTCTATACTCTACTCATCATTACA^{5m}CG^{5m}CG^{5m}CGATAT^{5m}CGTTAA^{5m}CGATAATT^{5m}CG-^{5m}CG^{5m}CGATTA^{5m}CGAT^{5m}CGATAA^{5m}CG^{5m}CGTTAATATGAGATATGAGATGTGTATG-3′ 3′-ATGAGATATGAGATGAGTAGTAATGTG^{5m}CG^{5m}CG^{5m}CTATAG^{5m}CAATTG^{5m}CTATTAAG^{5m}CG-^{5m}CG^{5m}CTAATG^{5m}CTAG^{5m}CTATTG^{5m}CG^{5m}CAATTATACTCTATACTCTACACATAC-5′
C^{5m}CGG	Fig. 1 *D* and *E*, Fig. 2 *A* and *D*, Fig. S4	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGG^{5m}CCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CGG(hemi_1)	Figs. 1*D* and 3*E*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGCCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CGG(hemi_2)	Fig. 1*D*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCCCGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGG^{5m}CCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CGG (ss_1)	Fig. 1*D*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CGGACTCTGATGACCAGGGCATCACA-3′
C^{5m}CGG (ss_2)	Fig. 1 *D* and *E*	3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGG^{5m}CCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CGGG	Fig. 3*C*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CGGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGCCCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CAGG	Fig. 3*C*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CAGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGTCCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CTGG	Fig. 3*C*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CTGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGACCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CCGG	Fig. 3*C*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CCGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGGCCTGAGACTACTGGTCCCGTAGTGT-5
A^{5m}CGGG	Fig. 3*C*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCA^{5m}CGGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGTGCCCTGAGACTACTGGTCCCGTAGTGT-5′
T^{5m}CGGG	Fig. 3*C*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCT^{5m}CGGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGAGCCCTGAGACTACTGGTCCCGTAGTGT-5′
G^{5m}CGGG	Fig. 3*C*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCG^{5m}CGGGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGCGCCCTGAGACTACTGGTCCCGTAGTGT-5′
ss_TA	Fig. 3*D*	5′-TATAATACCCTAGTAATAAGAGTACGGGGCGGAGCCTATAGAACCCGGAAACGCCG-3′
ss_TG	Fig. 3*D*	5′-TGTGATGCCCTGGTGATGAGAGTGCGGGGCGGAGCCTGTGGAACCCGGAAACGCCG-3′
ss_TC	Fig. 3*D*	5′-TCTCATCCCCTCGTCATCAGAGTCCGGGGCGGAGCCTCTCGAACCCGGAAACGCCG-3′
C^{5m}CAG(hemi)	Fig. 3*E*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CAGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGTCTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CTG(hemi)	Fig. 3*E*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CTGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGACTGAGACTACTGGTCCCGTAGTGT-5′
C^{5m}CCG(hemi)	Fig. 3*E*	5′-CGGCGTTTCCGGGTTCCATAGGCTCCGCCC^{5m}CCGACTCTGATGACCAGGGCATCACA-3′ 3′-GCCGCAAAGGCCCAAGGTATCCGAGGCGGGGGCTGAGACTACTGGTCCCGTAGTGT-5′

Table S2. List of plasmid and genomic DNA substrates used in this study

Name	Figures	Type of DNA	Methylation motif	Description
pRS(M.HpaII)	Figs. 1 *A* and *C* and 3*B*	Plasmid	5′-C^{5m}CGG-3′ 3′-GG^{5m}CC-5′	Plasmid pUC19 encoding the M.HpaII methyltransferase gene, isolated from *E. coli* cells (*dcm*$^{-}$). Contains 15 CCGG sites on each strand.
IMR90	Fig. 1 *B* and *C*	Genomic	Complex	gDNA isolated from human diploid cell strain IMR90.
HeLa	Figs. 1*E* and 3*A*	Genomic	Complex	Human female HeLa gDNA (New England Biolabs).
M.Fnu4HI	Figs. 1*E* and 2*B*	Genomic	5′-G^{5m}CNGC-3′	gDNA isolated from *E. coli* cells (*dcm*$^{-}$) expressing the *M.Fnu4HI* methyltransferase gene.
pUC19	Figs. 2*C* and 3*B*	Plasmid	5′-C^{5m}CWGG-3′ 3′-GGW^{5m}CC-5′	Plasmid pUC19, isolated from *E. coli* cells (*dcm*$^{+}$).
5-Aza-dC Jurkat	Fig. 3*A*	Genomic	Complex	gDNA isolated from human male Jurkat cells that are treated with 5-aza-2-deoxycytidine (5-Aza-dC).
Jurkat	Figs. 2*B* and 3*A*	Genomic	Complex	Human male Jurkat gDNA (New England Biolabs).
CpG Jurkat	Fig. 3*A*	Genomic	Complex	Human male Jurkat gDNA enzymatically methylated with CpG methylase (M.SssI) (New England Biolabs).
CpG HeLa	Fig. 3*A*	Genomic	Complex	Human female HeLa gDNA enzymatically methylated with CpG methylase (M.SssI) (New England Biolabs).
NIH 3T3 mouse	Figs. 2*B* and 3*A*	Genomic	Complex	NIH 3T3 (mouse embryonic fibroblast cell line) gDNA (New England Biolabs).
CpG NIH 3T3 mouse	Fig. 3*A*	Genomic	Complex	NIH 3T3 (mouse embryonic fibroblast cell line) gDNA enzymatically methylated with CpG methylase (M.SssI) (New England Biolabs).
MG1655	Fig. 3*A*	Genomic	5′-C^{5m}CWGG-3′ 3′-GGW^{5m}CC-5′	gDNA isolated from *E. coli* strain K-12 MG1655 (*dcm*$^{+}$).
M.AluI	Fig. 3*A*	Genomic	5′-AG^{5m}CT-3′ 3′-T^{5m}CGA-5′	gDNA isolated from *E. coli* cells (*dcm*$^{-}$) expressing the *M.AluI* methyltransferase gene.
M2.HgaI	Fig. 3*A*	Genomic	5′-GA^{5m}CGC-3′ 3′-CTGCG-5′	gDNA isolated from *E. coli* cells (*dcm*$^{-}$) expressing the *M2.HgaI* methyltransferase gene.
H.pylori	Figs. 3*A* and 4	Genomic	5′-^{5m}CCTC-3′; 5′-C^{5m}CTTC-3′; 5′-G^{5m}CGC-3′	gDNA isolated from *H. pylori* strain 26695 (ATCC).
M.HaeIII	Fig. 3*B*	Plasmid	5′-GG^{5m}CC-3′ 3′-C^{5m}CGG-5′	Plasmid pXba was methylated in vitro using M.HaeIII methyltransferase (New England Biolabs) following the recommended protocol.
M. MspI	Fig. 3*B*	Plasmid	5′-^{5m}CCGG-3′ 3′-GGC^{5m}C-5′	Plasmid pUC19 encoding the *M.MspI* methyltransferase gene, isolated from *E. coli* cells (*dcm*$^{-}$).

Table S3. Quantification of oxidation reaction species, as determined by LC-MS–based NgTET1 activity assay

Figure	DNA substrate	Fraction of ^{5m}C	Fraction of ^{5hm}C	Fraction of ^{5f}C	Fraction of ^{5ca}C
Fig. 1*C*	Oligo (24X)	0.014 ± 0.003	0.039 ± 0.013	0.037 ± 0.008	0.910 ± 0.013
	Plasmid (M.HpaII)	0	0.050 ± 0.015	0.080 ± 0.016	0.870 ± 0.008
	Genomic (IMR90)	0.003 ± 0.001	0.035 ± 0.012	0.088 ± 0.013	0.874 ± 0.024
Fig. 1*D*	symmDNA (C^{5m}CGG)	0.015 ± 0.001	0.070 ± 0.007	0.166 ± 0.018	0.748 ± 0.020
	hemiDNA (C^{5m}CGG hemi_1)	0.015 ± 0.006	0.060 ± 0.001	0.059 ± 0.013	0.866 ± 0.011
	hemiDNA (C^{5m}CGG hemi_2)	0.012 ± 0.002	0.069 ± 0.011	0.039 ± 0.008	0.881 ± 0.019
	ssDNA (C^{5m}CGG ss_1)	0.004 ± 0.003	0.026 ± 0.005	0.024 ± 0.005	0.945 ± 0.011
	ssDNA (C^{5m}CGG ss_2)	0.021 ± 0.006	0.045 ± 0.014	0.031 ± 0.007	0.903 ± 0.020
Fig. 1*E*	HeLa + TET1	0.003 ± 0.003	0.020 ± 0.013	0.052 ± 0.044	0.926 ± 0.054
	HeLa + NgTET1	0.023 ± 0.008	0.014 ± 0.001	0.052 ± 0.002	0.910 ± 0.008
	C^{5m}CGG + TET1	0.015 ± 0.009	0.229 ± 0.016	0.126 ± 0.030	0.630 ± 0.030
	C^{5m}CGG + NgTET1	0.008 ± 0.001	0.043 ± 0.005	0.133 ± 0.007	0.815 ± 0.011
	C^{5m}CGG (ss_2) + TET1	0.022	0.036	0.049	0.893
	C^{5m}CGG (ss_2) + NgTET1	0.004 ± 0.001	0.007 ± 0.001	0.009 ± 0.002	0.980 ± 0.002
	M.Fnu4HI + TET1	0.049 ± 0.018	0.255 ± 0.047	0.232 ± 0.015	0.463 ± 0.053
	M.Fnu4HI + NgTET1	0.083 ± 0.004	0.207 ± 0.006	0.215 ± 0.002	0.495 ± 0.009
Fig. 3*A*	5-Aza-dC Jurkat	0.007 ± 0.004	0.035 ± 0.019	0.060 ± 0.014	0.897 ± 0.021
	Jurkat	0.002 ± 0.002	0.026 ± 0.004	0.066 ± 0.010	0.904 ± 0.010
	CpG Jurkat	0.001 ± 0.001	0.015 ± 0.004	0.040 ± 0.004	0.943 ± 0.003
	HeLa	0.003 ± 0.002	0.027 ± 0.003	0.066 ± 0.010	0.904 ± 0.010
	CpG HeLa	0.004 ± 0.002	0.014 ± 0.001	0.072 ± 0.018	0.910 ± 0.015
	NIH 3T3 mouse	0.003 ± 0.002	0.018 ± 0.004	0.062 ± 0.014	0.917 ± 0.012
	CpG NIH 3T3 mouse	0.004 ± 0.002	0.015 ± 0.003	0.086 ± 0.020	0.895 ± 0.017
	MG1655	0.147 ± 0.021	0.487 ± 0.026	0.073 ± 0.008	0.294 ± 0.009
	M.AluI	0.136 ± 0.004	0.615 ± 0.057	0.208 ± 0.088	0.042 ± 0.042
	M2.HgaI	0.014 ± 0.008	0.070 ± 0.041	0.062 ± 0.016	0.854 ± 0.057
	H.pylori	0.053 ± 0.017	0.151 ± 0.036	0.281 ± 0.004	0.515 ± 0.048
Fig. 3*B*	M.HpaII	0.001 ± 0.001	0.044 ± 0.014	0.084 ± 0.013	0.872 ± 0.007
	pUC19	0.071 ± 0.010	0.285 ± 0.025	0.151 ± 0.028	0.493 ± 0.019
	M.HaeIII	0.043 ± 0.004	0.390 ± 0.005	0.320 ± 0.087	0.247 ± 0.093
	M.MspI	0.061 ± 0.007	0.408 ± 0.058	0.230 ± 0.060	0.302 ± 0.086
Fig. 3*C*	C^{5m}CGGG	0.015 ± 0.001	0.055 ± 0.004	0.214 ± 0.006	0.715 ± 0.008
	A^{5m}CGGG	0.011 ± 0.002	0.014 ± 0.001	0.009 ± 0.001	0.967 ± 0.002
	G^{5m}CGGG	0.008 ± 0.001	0.061 ± 0.021	0.215 ± 0.027	0.716 ± 0.048
	T^{5m}CGGG	0.011 ± 0.001	0.020 ± 0.001	0.015 ± 0.002	0.954 ± 0.002
	C^{5m}CAGG	0.009 ± 0.001	0.396 ± 0.014	0.419 ± 0.013	0.176 ± 0.012
	C^{5m}CTGG	0.080 ± 0.056	0.463 ± 0.017	0.305 ± 0.038	0.151 ± 0.019
	C^{5m}CCGG	0.285 ± 0.128	0.522 ± 0.102	0.144 ± 0.022	0.048 ± 0.005

Table S4. ^{5m}C methylation motifs detected for *H. pylori* 26695 using NgTET1/$NaBH_4$/T4-βGT and SMRT sequencing

Methyltransferase*	^{5m}C motif (5′-3′)	Total number in genome	No. methylated	% Detected methylated
M2.HpyAVI	^{5m}CCTC	4,885	4,349	89.0
M.HpyAV	C^{5m}CTTC	1,304	1,278	98.0
M.HpyAVIII	G^{5m}CGC	12,538	7,188	57.3

*Methyltransferase assignments are as previously described (1).

1. Krebes J, et al. (2014) The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res* 42(4):2415–2432.

Table S5. List of ds-oligo substrates containing MspI site, used in Fig. S1

No.	Name	Sequence
1	C/C	5′-FAM-ACACCCATCACATTTACACCGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGGCCTATTCTCAACTTACATCTCAACC-FAM-5′
2	mC/C	5′-FAM-ACACCCATCACATTTACAC^{5m}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGGCCTATTCTCAACTTACATCTCAACC-FAM-5′
3	mC/mC	5′-FAM-ACACCCATCACATTTACAC^{5m}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGG^{5m}CCTATTCTCAACTTACATCTCAACC-FAM-5′
4	hmC/C	5′-FAM-ACACCCATCACATTTACAC^{5hm}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGGCCTATTCTCAACTTACATCTCAACC-FAM-5′
5	hmC/mC	5′-FAM-ACACCCATCACATTTACAC^{5hm}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGG^{5m}CCTATTCTCAACTTACATCTCAACC-FAM-5′
6	hmC/hmC	5′-FAM-ACACCCATCACATTTACAC^{5hm}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGG^{5hm}CCTATTCTCAACTTACATCTCAACC-FAM-5′
7	caC/C	5′-FAM-ACACCCATCACATTTACAC^{5ca}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGGCCTATTCTCAACTTACATCTCAACC-FAM-5′
8	caC/mC	5′-FAM-ACACCCATCACATTTACAC^{5ca}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGG^{5m}CCTATTCTCAACTTACATCTCAACC-FAM-5′
9	caC/caC	5′-FAM-ACACCCATCACATTTACAC^{5ca}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGG^{5ca}CCTATTCTCAACTTACATCTCAACC-FAM-5′
10	fC/C	5′-FAM-ACACCCATCACATTTACAC^{5f}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGGCCTATTCTCAACTTACATCTCAACC-FAM-5′
11	fC/mC	5′-FAM-ACACCCATCACATTTACAC^{5f}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGG^{5m}CCTATTCTCAACTTACATCTCAACC-FAM-5′
12	fC/fC	5′-FAM-ACACCCATCACATTTACAC^{5f}CGGATAAGAGTTGAATGTAGAGTTGG-3′ 3′-TGTGGGTAGTGTAAATGTGG^{5f}CCTATTCTCAACTTACATCTCAACC-FAM-5′