Supporting Online Material for

Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine

Shinsuke Ito, Li Shen, Qing Dai, Susan C. Wu, Leonard B. Collins,
James A. Swenberg, Chuan He, Yi Zhang*

*To whom correspondence should be addressed. E-mail: yi_zhang@med.unc.edu (Y.Z.)

Published xx July 2011 on *Science* Express

DOI: xxx/science. 1210597

This PDF file includes:

Figs. S1 to S8
Author contribution
Materials and Methods
References

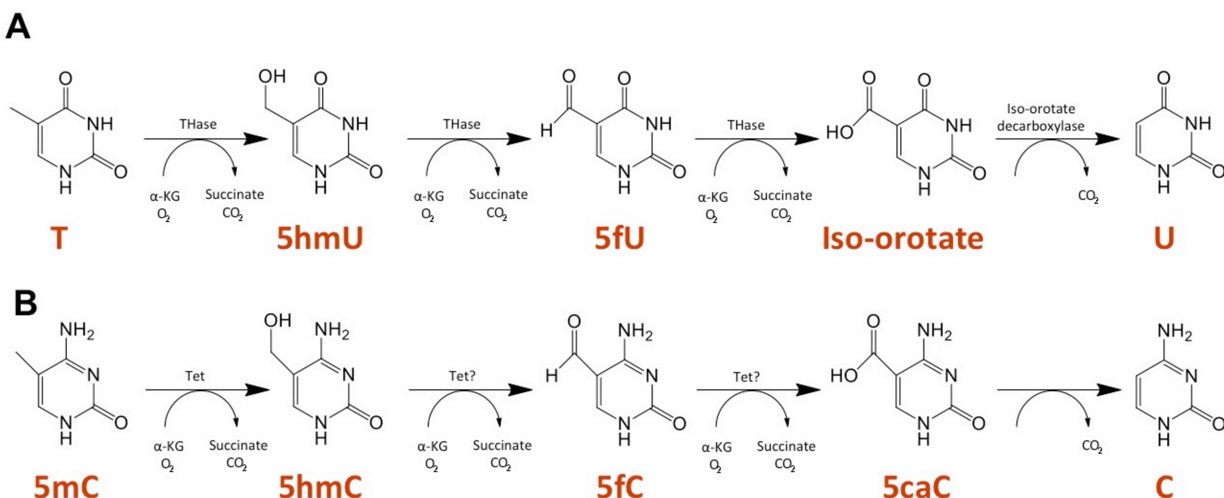

Figure S1. Chemistry involved in the thymine to uracil conversion in the thymidine salvage pathway and a proposed mechanism for 5mC demethylation

(**A**) Part of the thymidine salvage pathway. Thymine (T) is converted to 5-hydroxymethyl U (5hmU), 5-formyl U (5fU), and iso-orotate by thymine hydroxylase (THase) in three consecutive oxidation reactions, each requiring O_2, α-KG while releasing CO_2 and succinate. Iso-orotate is then converted to uracil by iso-orotate decarboxylase.

(**B**) Proposed mechanism of oxidative DNA demethylation initiated by Tet proteins. Similar to THase, Tet proteins can potentially oxidize 5mC to produce 5-hydroxymethyl C (5hmC), 5-formyl C (5fC), and 5-carboxyl C (5caC), which then may be converted to C by a decarboxylase.

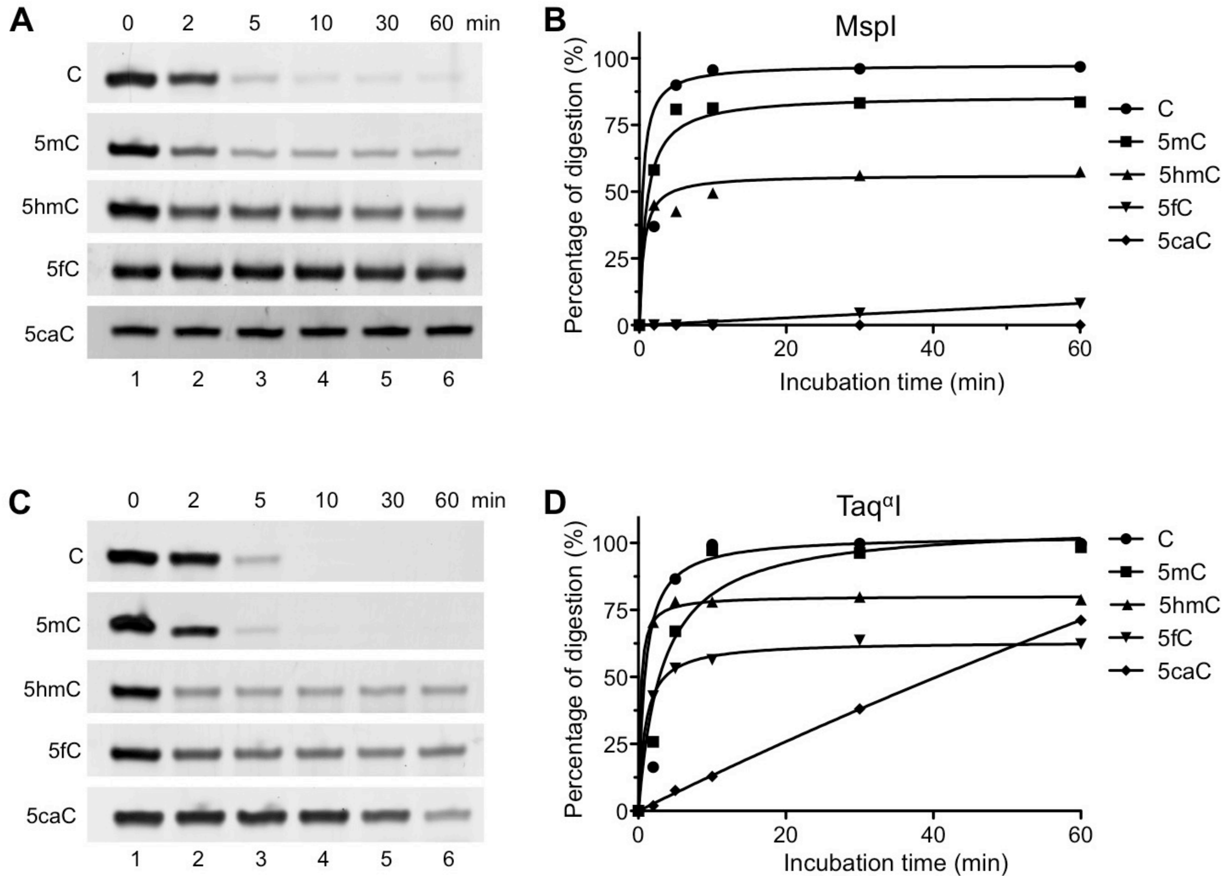

Figure S2. Modifications at the 5-position of cytosine affect efficiency of restriction enzyme digestion

(A) Effect of various modifications at the 5-position of the internal cytosine (C) of the MspI site on MspI digestion. MspI site-containing 20mer oligo DNA was digested by excessive amounts of MspI at different times. Shown are the SYTO 60 stained undigested oligo DNA after separation in 15% non-denaturing acrylamide gels. The only difference between the various oligos is at the 5-position of the internal C of the MspI site.

(B) Quantification of the data presented in panel A. Shown is the percentage of digested oligos as a function of digestion time.

(C) The same as panel A except that the MspI site in the oligo DNA is replaced by a TaqI site.

(D) Quantification of the data presented in panel C. Shown is the percentage of digested oligos as a function of digestion time.

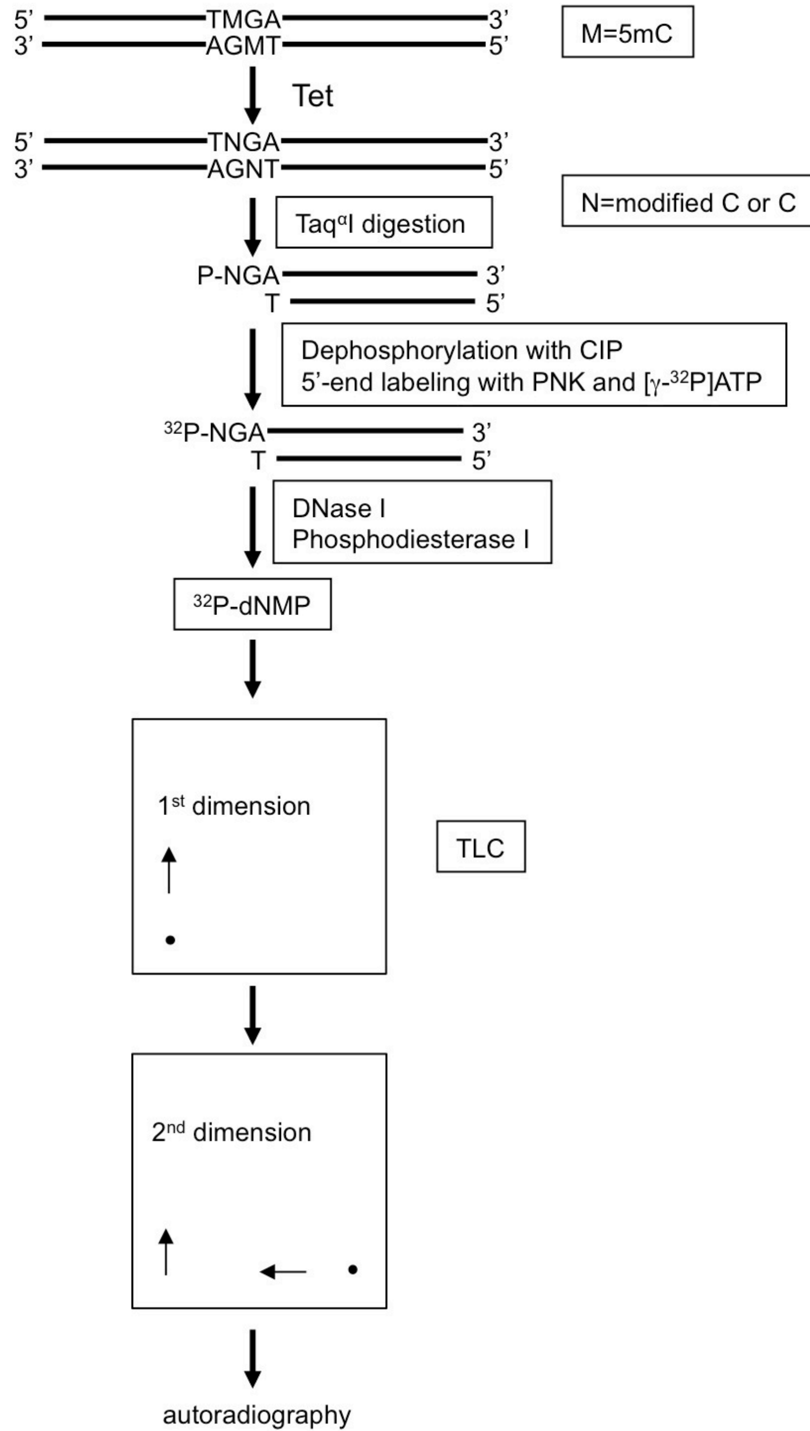

Figure S3. Diagram depicting the *in vitro* assay and 2D-TLC

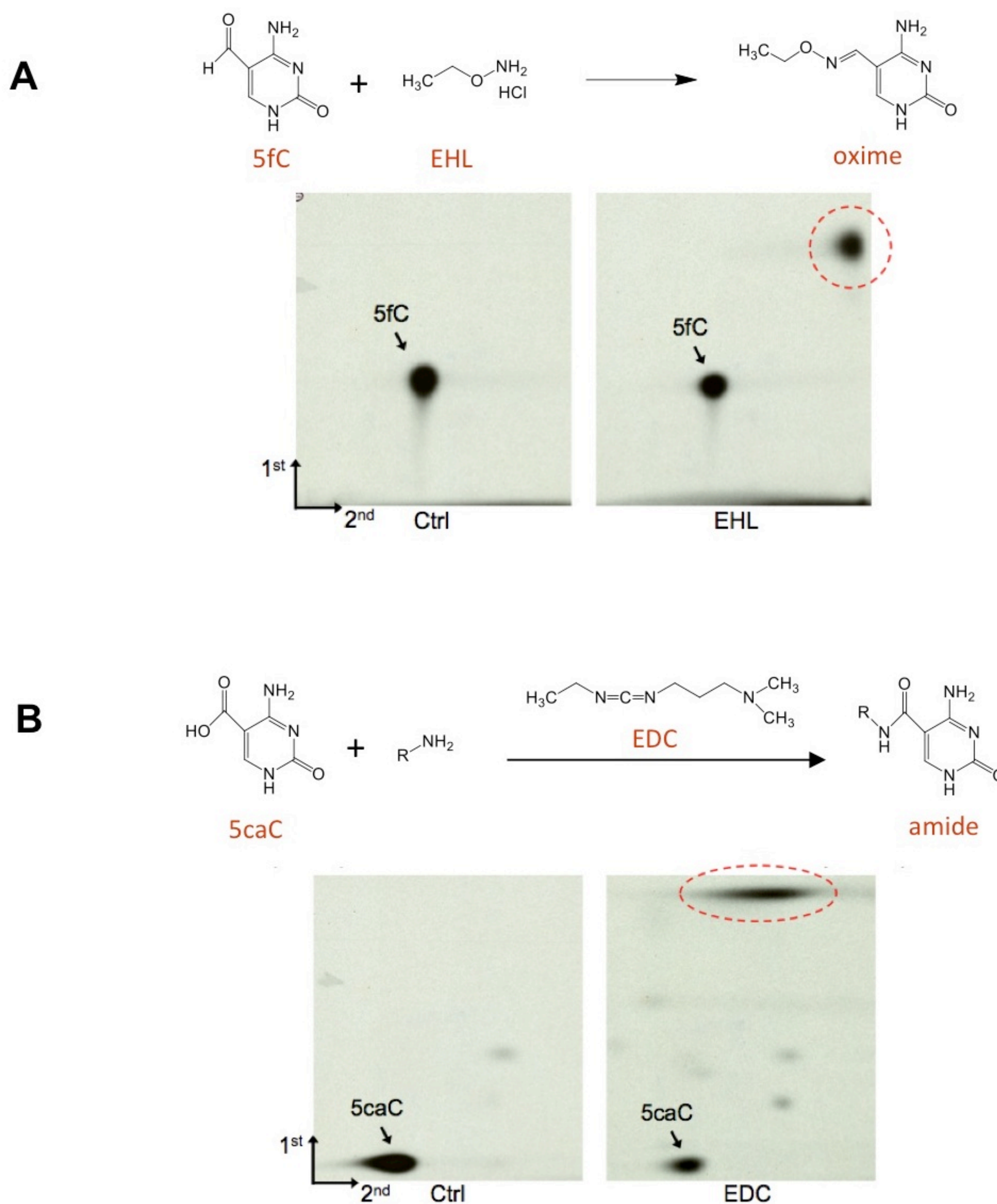

Figure S4. Chemical conversion of 5fC and 5caC nucleotides by treatment with O-ethylhydroxylamine hydrochloride (EHL) or 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC)

(A) Conversion of 5fC to oxime by EHL treatment. Top panel depicts the chemical reaction. Bottom panels show the migration of 5fC and the reaction product oxime in 2D-TLC without and with the treatment of EHL.

(B) Conversion of 5caC to amide by EDC treatment. Top panel depicts the chemical reaction. Bottom panels show the migration of 5caC and the reaction product amide in 2D-TLC without and with the treatment of EDC.

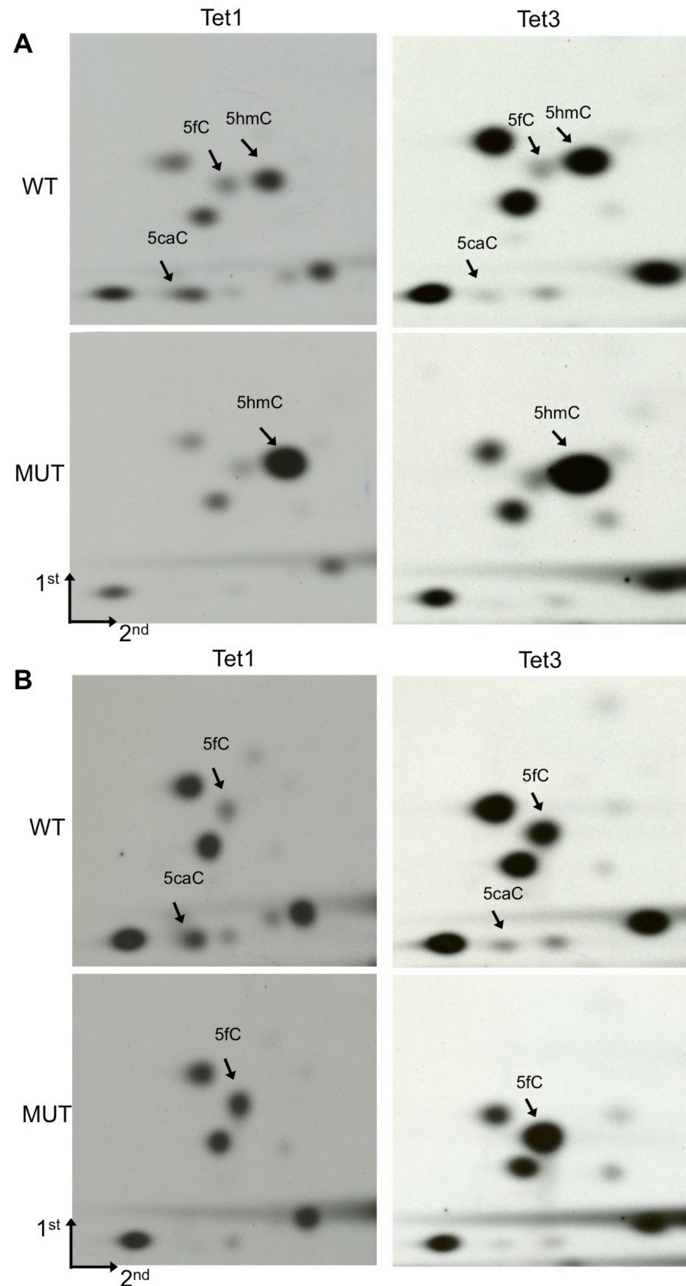

Figure S5. Enzymatic oxidation of 5hmC and 5fC-containing DNA by Tet1 and Tet3

(A) Autoradiographs of 2D-TLC analysis of samples derived from 5hmC-containing TaqI 20mer DNA oligos incubated with wild-type or catalytic-deficient mutant Tet1 or Tet3. Compared to their mutant controls, incubation of each Tet protein with substrate resulted in a decrease in the 5hmC levels concomitant with appearance of the 5fC and 5caC.

(B) Autoradiographs of 2D-TLC analysis of samples derived from 5fC-containing TaqI 20mer DNA oligos incubated with wild-type or catalytic-deficient mutant Tet1 or Tet3. Compared to their mutant controls, incubation of each Tet protein with substrate resulted in a decrease in the 5fC levels concomitant with appearance of the 5caC.

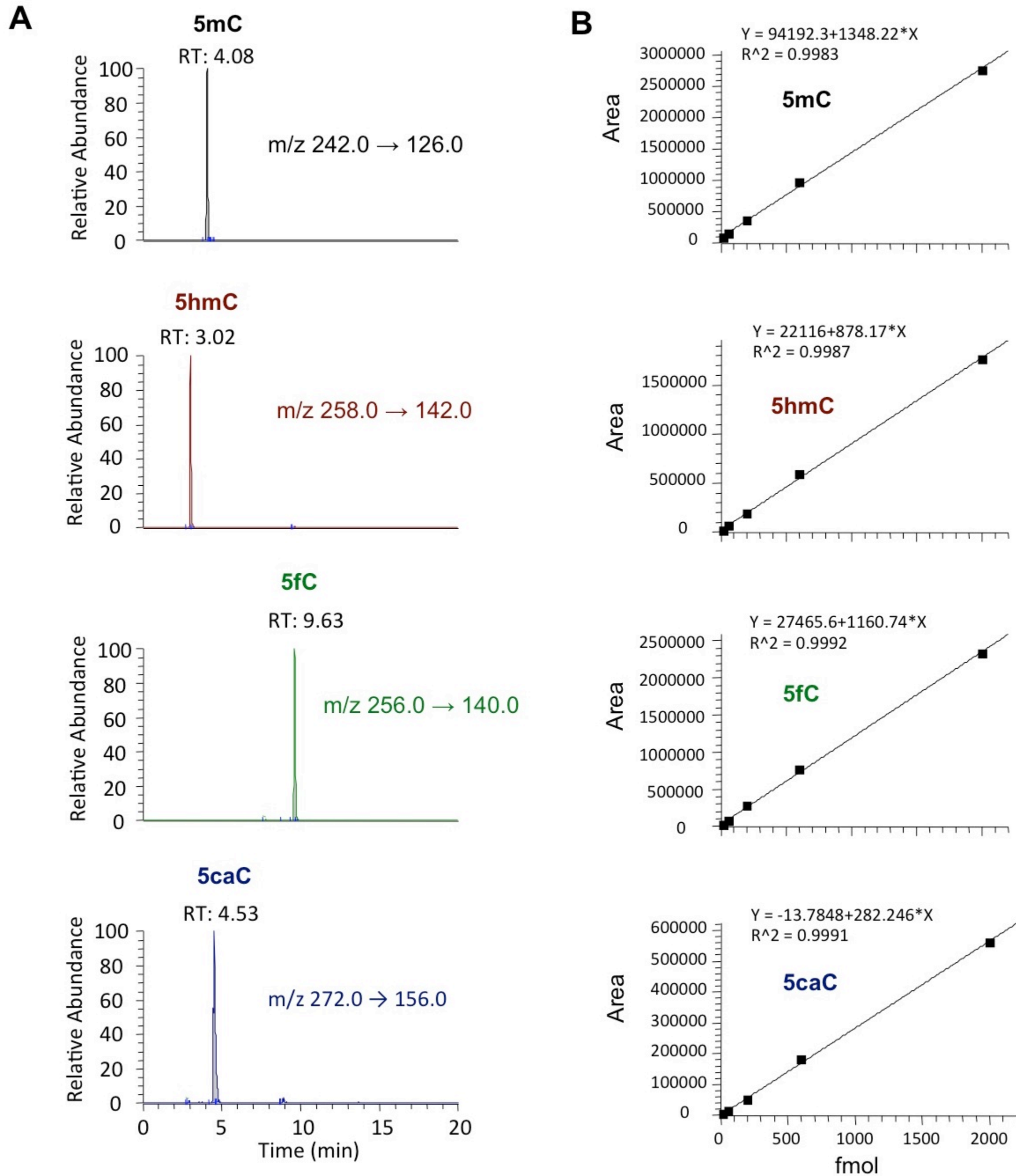

Figure S6. Separation and quantification of cytosine derivatives by LC-MS/MS

(A) Detection of 5mC, 5hmC, 5fC, and 5caC in multiple reaction monitoring (MRM) mode. Shown are typical LC-MS/MS chromatograms of 200 fmol external standards loaded on the column.

(B) Standard curves for 5mC, 5hmC, 5fC, and 5caC generated for quantification. Good linearity was achieved from 20-2000 fmol of each of the cytosine derivatives.

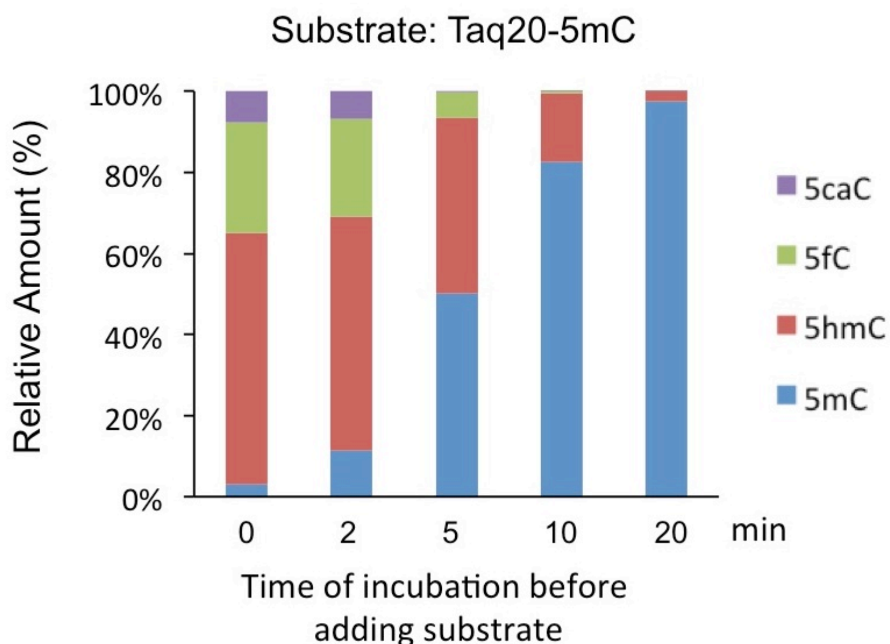

Figure S7. Tet2 is quickly inactivated under reaction conditions

Tet2 protein was incubated in the reaction conditions in the absence of substrate DNA and $Fe(NH_4)_2(SO_4)$ for the indicated time before the addition of the TaqI-5mC-containing 20mer substrate and $Fe(NH_4)_2SO_4$. After 10 min of reaction, the reaction mixture was analyzed by LC-MS/MS to quantify the various cytosine derivatives.

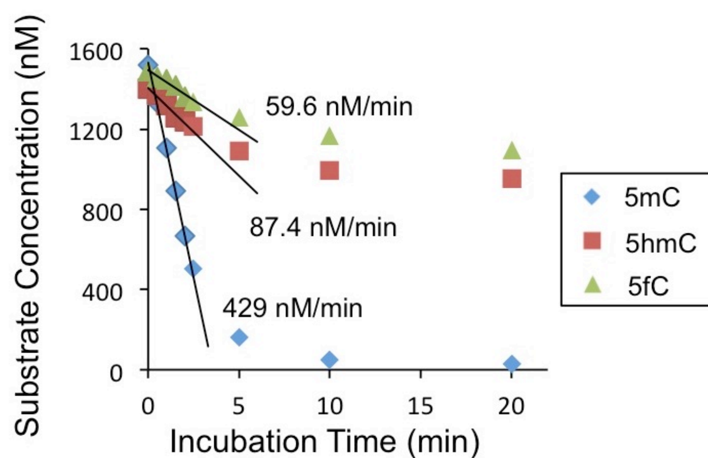

Figure S8. Determination of the initial reaction rate for Tet2 with 5mC, 5hmC, and 5fC-containing substrate

Substrate concentration at different time points of the reaction for each of the 5mC, 5hmC, and 5fC-containing substrate. The initial reaction rate for each of the substrate is derived using the data acquired in the first 2.5 min of the reaction.

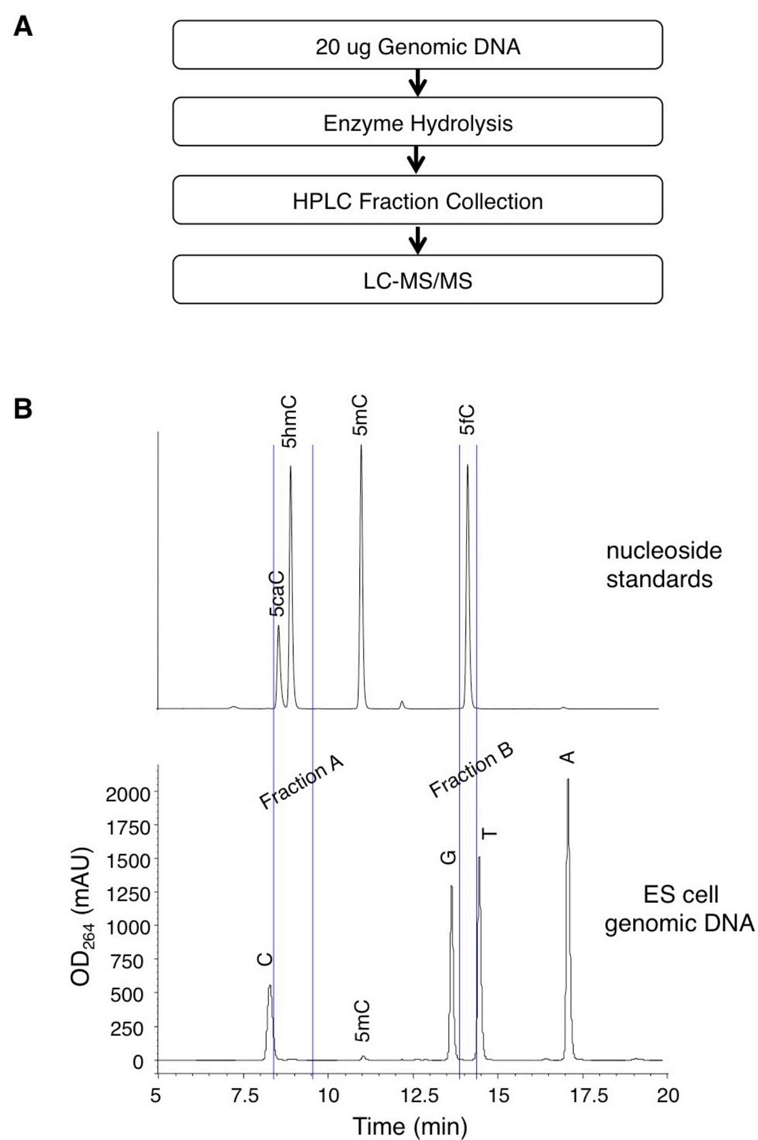

Figure S9. Procedures for quantification of cytosine derivatives in genomic DNA

(A) Schematic representation of the procedures used to quantify cytosine derivatives in genomic DNA.

(B) HPLC analysis of standard nucleosides (top panel) and enzyme hydrolyzed mouse ES genomic DNA. The fractions corresponding to the retention time of 5hmC/5caC (fraction A) and 5fC (fraction B) were collected for further LC-MS/MS analysis and quantified using the standard curves shown in Fig. S11A.

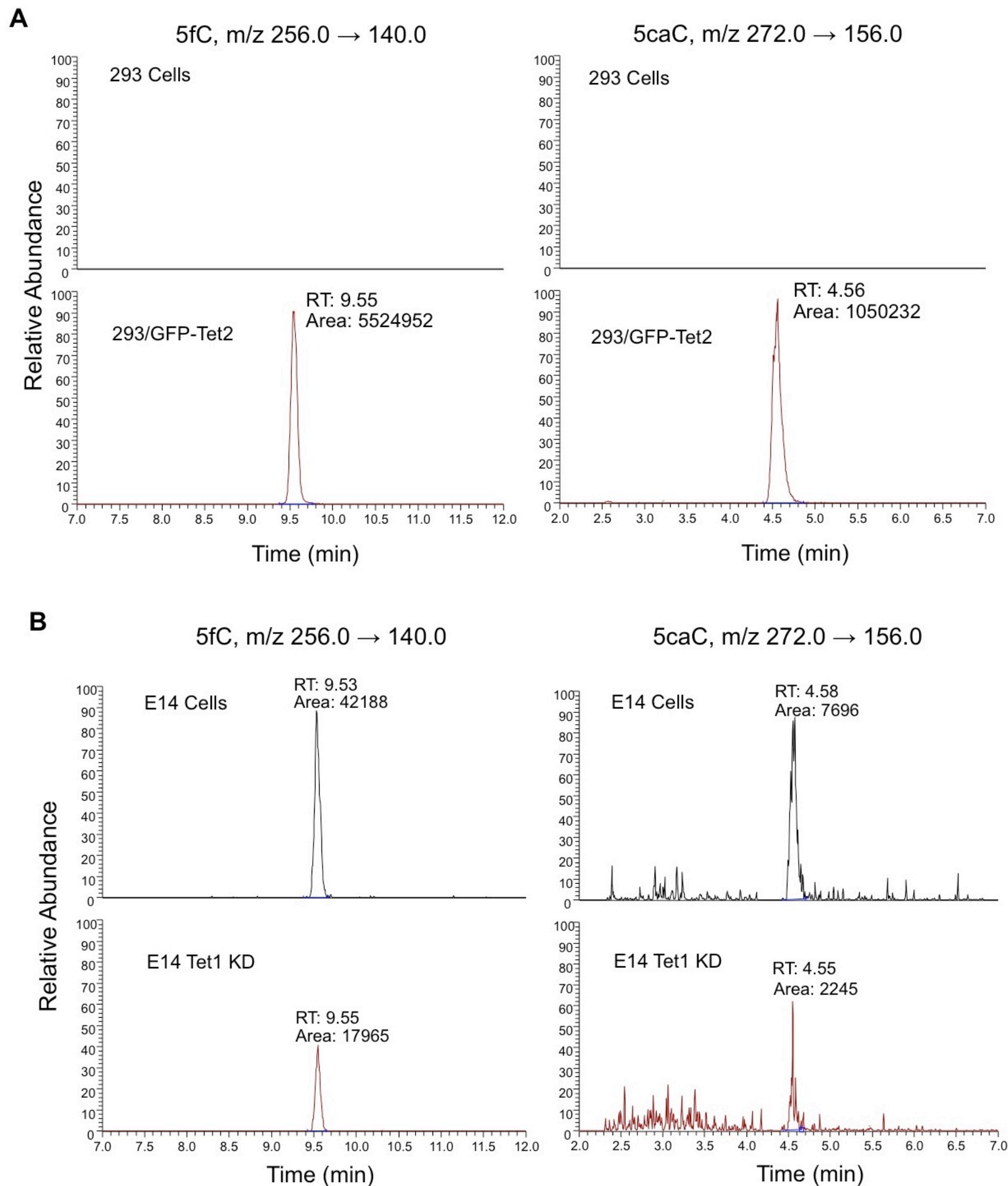

Figure S10. Detection of 5fC and 5caC in the genomic DNA of 293 cells overexpresing Tet2 or the genomic DNA of mouse ES cells with or without Tet1 knockdown

(A) Mass spectrometry detection of 5fC (left panels) and 5caC (right panels) in genomic DNA of 293 cells without (top panels) or with Tet2 (bottom panels) overexpression.

(B) Mass spectrometry detection of 5fC (left panels) and 5caC (right panels) in genomic DNA of E14 mouse ES cells without (top panels) or with Tet1 (bottom panels) knockdown (KD).

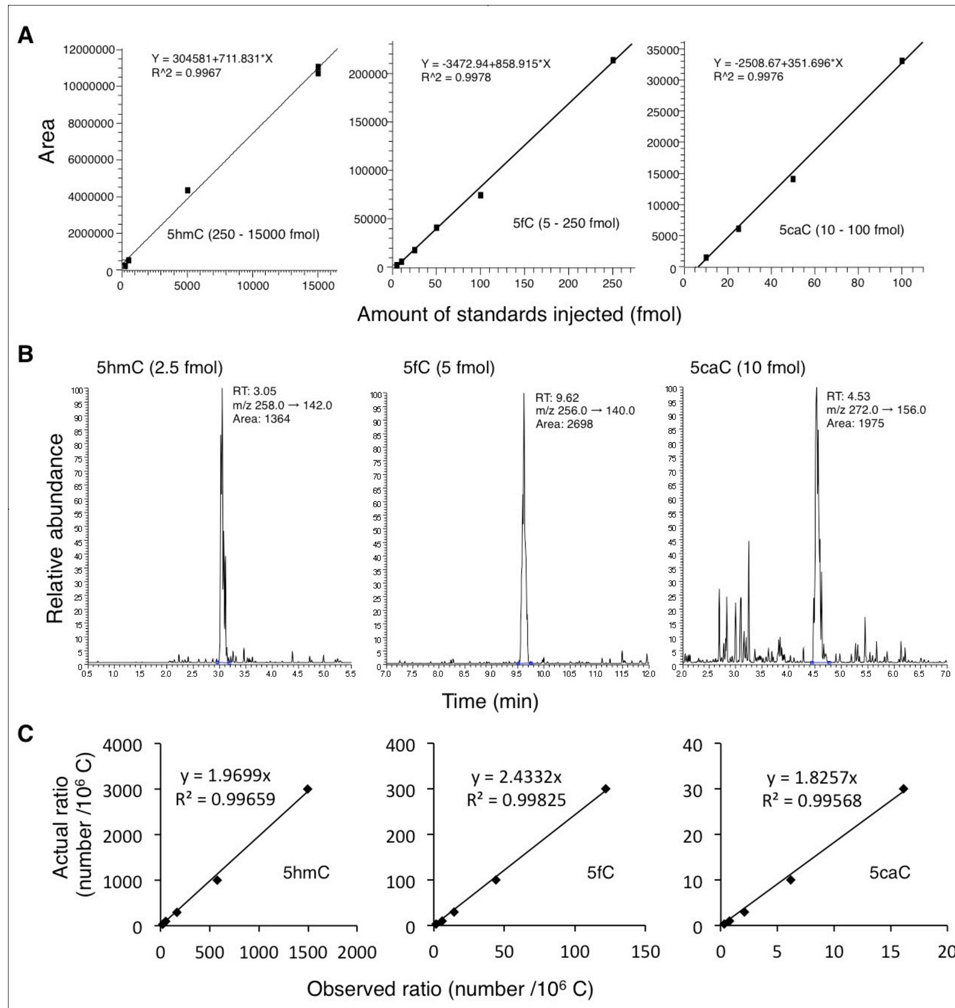

Figure S11. Standard curves and detection limits of mass spectrometry for 5hmC, 5fC, and 5caC

(A) Standard curves for 5hmC, 5fC, and 5caC generated for quantification in genomic DNA. Good linearity was achieved from 250-15000 fmol for 5hmC, 5-250 fmol for 5fC, and 10-100 fmol for 5caC.

(B) Representative LC-MS/MS chromatograms at the limits of detection (LOD) for 5hmC (2.5 fmol), 5fC (5 fmol), and 5caC (10 fmol), showing excellent signal to noise ratio.

(C) Calibration curves for estimating the recovery rate of each nucleoside. Standard 5hmC, 5fC and 5caC nucleosides were mixed with 15 nmol each of C, G, T, A at 5 different levels, then samples were analyzed by the same procedures as for genomic DNA.

Samples	C	5mC		5hmC		5fC		5caC	
	Amount (nmol)	Amount (nmol)	% of C	Amount (fmol)	% of C	Amount (fmol)	num. /10^6 C	Amount (fmol)	num. /10^6 C
293 cells	11.1	0.29	2.6	514	0.009	20.0	4.4	ND	ND
293/GFP-CD2	10.9	0.14	1.3	41669	0.754	7289.2	1628.7	3082.5	516.8
293/GFP-CD2 MUT	11.5	0.31	2.6	847	0.014	8.2	1.7	ND	ND
E14 ctrl 1	17.1	0.53	3.0	9715	0.112	125.7	17.9	33.3	3.6
E14 ctrl 2	8.7	0.26	2.9	6066	0.138	70.6	19.8	14.8	3.1
E14 Tet1 KD 1	16.0	0.52	3.1	4343	0.053	38.6	5.9	8.3	1.0
E14 Tet1 KD 2	7.2	0.22	3.0	2279	0.063	28.2	9.6	ND	ND
Brain cortex 1	17.1	0.54	3.1	53692	0.619	146.3	20.8	3.3	0.4
Brain cortex 2	10.5	0.34	3.1	38490	0.719	47.3	10.9	ND	ND
Kidney 1	19.7	0.56	2.7	14248	0.142	24.0	3.0	ND	ND
Kidney 2	14.8	0.41	2.7	21026	0.280	18.4	3.0	ND	ND
Lung 1	18.8	0.53	2.7	14595	0.153	46.1	6.0	4.7	0.5
Lung 2	18.9	0.54	2.8	16638	0.173	11.0	1.4	ND	ND
Heart 1	14.4	0.40	2.7	10159	0.139	33.4	5.6	ND	ND
Heart 2	17.9	0.50	2.7	13521	0.149	26.1	3.6	ND	ND
Pancreas 1	17.3	0.46	2.6	8321	0.095	26.2	3.7	ND	ND
Pancreas 2	6.5	0.16	2.4	3023	0.092	21.4	8.0	ND	ND
Liver 1	18.7	0.53	2.7	11526	0.121	51.8	6.7	ND	ND
Liver 2	14.3	0.41	2.8	12001	0.166	38.5	6.6	ND	ND
Spleen 1	18.2	0.53	2.8	6099	0.066	57.6	7.7	0.9	0.1
Spleen 2	15.4	0.47	3.0	3824	0.049	98.8	15.7	ND	ND
Thymus 1	21.2	0.68	3.1	2133	0.020	10.9	1.3	ND	ND
Thymus 2	22.3	0.68	3.0	3076	0.027	65.6	7.2	ND	ND

Supplementary Table 1. Quantification of the 5mC, 5hmC, 5fC, and 5caC in genomic DNA of 293 cells, mouse ES cells and mouse organs. The percentages of 5hmC, 5fC and 5caC in total C were corrected with the calibration curve shown in Fig. S11C.

Author Contributions

Y.Z. conceived the project and wrote the manuscript. S.I., L.S, S.C.W., and L.B.C. performed the experiments. Q.D. and C.H. synthesized the 5fC and 5caC-nucleoside as well as the 5fC and 5caC-containing oligos. C.H. suggested the EHL and EDC experiments. J.A.S provided guidance in mass spectrometric analysis. The authors declare no competing interests.

Materials and Methods

MspI and TaqI digestion tests. The sequences of the 20mer MspI or TaqI restriction site-containing DNA oligos are 5'-TTTCAGCTCmCGGTCACGCTC-3' and 5'-GTTCAGCTTmCGATCACGCTC-3' (m represents the various modifications), respectively. Synthesis of the 5fC and 5caC containing oligos was described (*1*). Various double-stranded oligos (50 ng) were digested with MspI (10 U; NEB) or TaqI (5 U; NEB), and separated in 15% polyacrylamide gel. After staining with SYTO 60 fluorescent nucleic acid stain (Invitrogen), the gels were scanned with Odyssey infrared imager (LI-COR). The percentage of digestion was calculated using the amount undigested at a given time.

Tet activity and TLC assays. Expression and purification of recombinant Tet proteins encoding Flag-tagged catalytic domains of Tet1 (aa1367-2039), Tet2 (aa916-1921), and Tet3 (aa697-1668), as well as their catalytic mutants were performed as previously described (*2*). For *in vitro* enzymatic activity assays, 0.5 µg of various double stranded DNA oligos that contain unmodified or fully modified 5mC, 5hmC, or 5fC in the TaqI site (TCGA) were incubated with 1.2 µg (1:6 enzyme/substrate ratio) wild-type or catalytic mutant Tet proteins in the presence of 50 mM HEPES (pH 7.9), 100 mM NaCl, 75 µM $Fe(NH_4)_2(SO_4)_2$, 2 mM ascorbate, and 1mM α-ketoglutate at 37 ºC for 1-40 min. Then, the oligonucleotide substrates were purified using Qiaquick Nucleotide Removal Kit (Qiagen) and digested with TaqI. After treatment with calf intestinal alkaline phosphatase (CIAP) and labeling with [γ-^{32}P]ATP and T4 polynucleotide kinase, the DNA fragments were ethanol-precipitated and digested with 10 µg of DNase I and 10 µg Phosphodiesterase I in the presence of 15 mM $MgCl_2$ and 2 mM $CaCl_2$. The samples were then separated by 2D-TLC on PEIcellulose TLC plates (Merck) using the following solvents:

first dimension, isobutyric acid: NH_4OH: H_2O=66:2:20, second dimension: isopropanol: HCl: H_2O=70:15:15. After drying, the TLC plate was exposed to X-ray film. For labeling of 5fC and 5caC nucleosides, deoxycytidine kinase (Proteinkinase.de) was used in the presence of [γ-^{32}P]ATP.

Chemical treatment of *in vitro* Tet-reaction products. *In vitro* reaction mixtures digested with DNase I and phosphodiesterase I were incubated with either 1) 100 mM sodium borohydride ($NaBH_4$) (Sigma) at room temperature for 1 hour, 2) 25 mM O-ethylhydroxylamine hydrochloride (EHL) (Sigma) at 37 ºC for 3 hours, or 3) 200 mM l-Ethyl-3- (3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) (Pierce Chemical) at room temperature for 16 hours. After each treatment, samples were separated by 2D-TLC.

Confirming the identities of X and Y spots with mass spectrometry. After 2D-TLC, the spots corresponding to X and Y were scraped, extracted with the 2D-TLC second developing buffer, lyophilized, redissolved with water before treated with CIAP and filtered through Nanosep3K (PALL) for LC-MS/MS analysis using an UPLC (Waters) coupled to a TSQ-Quantum Ultra triple-quadrupole mass analyzer (ThermoFinnigan) using heat assisted electrospray ionization (HESI) in positive mode (spray voltage of 3000 V, API temperature of 250 °C, sheath gas flow rate 35 arb, AUX gas flow rate 25 arb, capillary temperature of 285 °C). Liquid chromatography (LC) was performed with a 2.1 x 100 mm HSS T3 1.8 μm column (Waters) with gradient elution at flow rate of 200 μl/min using 0.02% acetic acid in water as mobile phase A and methanol as mobile phase B. The gradient was 0→3.5 min, 3% B, 3.5→12.5 min, 3%→16.2%B, 12.5 →13min, 16.2%B→30%B, 13→15 min, 30%B, 15→16min, 30%→3%B, 16→20min, 3%B. The eluant was directed to the mass spectrometer that was running in the product scan mode (60-280 m/z) selecting 256.0 m/z or 272.0 m/z precursor ion (5fC and 5caC, respectively). The collision energy used was 30 V.

Quantification of *in vitro* reaction products with mass spectrometry. The *in vitro* reactions were stopped by adding 10 volumes of ice-cold Buffer PN (Qiagen), and oligonucleotides were purified with Qiaquick nucleotide Removal Kit (Qiagen) and digested with 5 U DNaseI, 0.5 μg Phosphodiesterase I and 2 U CIAP. The samples were filtered through Nanosep3K (PALL)

before LC-MS/MS analysis. Mass spectrometer was running in multiple reaction monitoring (MRM) mode, monitoring the transition of m/z 242.0 to 126.0 (5mC), m/z 258.0 to 142.0 (5hmC), m/z 256.0 to 140.0 (5fC) and m/z 272.0 to 156.0 (5caC), and the collision energy was 15 V. External standard calibration was used for the quantification of the various cytosine derivatives.

Detection and quantification of 5hmC, 5fC, and 5caC in genomic DNA from HEK293, ES, and mouse organs. Mass spectrometric quantification standard curve for each of the cytosine derivatives were generated using standard 5mC, 5hmC, 5fC, and 5caC. 5mC and 5hmC were purchased from USBiological (cat# M3400) and Berry & Associates (cat# PY7588), respectively. 5fC and 5caC were synthesized as described (*1*). HEK293 cells were transfected with GFP-Tet2 expression vectors using Fugene 6 transfection reagent (Roche) according to manufacturer's instructions. Three days after transfection, GFP positive cells were FACS sorted and genomic DNA was prepared by resuspending the cells in 0.5 ml DNA extraction buffer (10 mM Tris HCl pH8.0, 150 mM NaCl, 10 mM EDTA, and 0.1% SDS) containing 0.1 mg/ml RNase A. After 10 min incubation at room temperature, 0.5 mg/ml Proteinase K was added for overnight incubation at 55 ºC. DNA was extracted with equal volumes of phenol: chloroform: isoamyl alcohol (25:24:1) twice. After ethanol precipitation, DNA was resuspended in water. Similar procedures were also used for the isolation of genomic DNA from mouse ES cells and organs. For 2D-TLC assay, Ten μg of genomic DNA was digested with Taq$^{\alpha}$I in the presence of RNaseA overnight. CIAP was then added. After 1 hour incubation at 37 ºC, DNA was purified using Qiaquick Nucleotide Removal Kit (Qiagen), and end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP for 1 hour at 37 ºC. Radiolabeled DNA was digested with DNase I and phosphodiesterase I at 37 ºC overnight before being separated by 2D-TLC.

For mass spectrometric quantification, a method previously used for quantification of rare DNA adducts in genomic DNA was used (*3*). As depicted in Fig. S6, genomic DNA (20 μg) was hydrolyzed with DNaseI (50U) in the digestion buffer (20 mM Tris-HCl, 5 mM $MgCl_2$, pH 7.6) at 37 °C for 30 min, followed by the addition of phosphodiesterase I (5 μg) and CIAP (10 U) for an additional 1 hour. The digested samples were then filtered with Nanosep3K (PALL) and subjected to HPLC (Agilent 1200, 4.6 x 150 mm, 3 μm, Atlantis T3 column). The gradient used

is 0→15 min, 3%→30%B, 15→17 min, 30%B, 17→18 min, 30%→3%B, 18→25 min, 3%B (mobile phase A: 5 mM ammonium formate in water, pH 4.0; mobile phase B: methanol). During the HPLC separation, fractions corresponding to the retention time of 5hmC, 5caC and 5fC were collected, and C and 5mC were quantified with external standard curve. The collected fractions were then lyophilized, and redissolved with water for LC-MS/MS analysis. The recovery rates of 5hmC, 5fC, and 5caC were estimated as shown in Fig. S11C to correct their percentages in total cytosines.

References

1. Q. Dai, C. He, *Org Lett* **13**, 3446 (Jul 1, 2011).
2. S. Ito *et al.*, *Nature* **466**, 1129 (Jul 18, 2010).
3. G. Boysen *et al.*, *J Chromatogr B Analyt Technol Biomed Life Sci* **878**, 375 (Feb 1, 2010).