Supporting Information

for

Direct decarboxylation of Ten-eleven translocation-produced 5-carboxylcytosine in mammalian genomes forms a new mechanism for active DNA demethylation

Yang Feng,[a,b] Juan-Juan Chen,[a] Neng-Bin Xie,[a] Jiang-Hui Ding,[a] Xue-Jiao You,[a] Wan-Bing Tao,[a] Xiaoxue Zhang,[c] Chengqi Yi,[c] Xiang Zhou,[a] Bi-Feng Yuan,*[,a,b] Yu-Qi Feng[a,b]

[a] Sauvage Center for Molecular Sciences, Department of Chemistry, Wuhan University, Wuhan 430072, China.

[b] School of Health Sciences, Wuhan University, Wuhan 430071, China.

[c] State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China.

* Corresponding author: Bi-Feng Yuan. Tel: +86-27-68755595; Fax: +86-27-68755595; E-mail: bfyuan@whu.edu.cn

The Supplementary Data includes following items:

HPLC purification of F-5caC

Purification of F-5caC was performed on a Shimadzu LC-20A HPLC system (Tokyo, Japan) equipped with UV/vis detector and binary HPLC pump. Chromatographic separation was fulfilled by a self-manufactured C18 column at 35°C using the mobile phases of 0.05% (v/v) formic acid in water (A) and methanol (B) at a flow rate of 2 mL/min. A gradient of 0-10 min 5% B, 10-40 min 5 to 80% B, 40-42 min 80 to 5% B, and 42-50 min 5% B, was used for the separation.

Preparation of F-5mC

F-5mC was obtained from feeding of F-dC to HEK293T cells. The genomic DNA of HEK293T cells was extracted after feeding of F-dC with different concentrations (100 μM, 200 μM, 300 μM, and 400 μM). Then isolated DNA was enzymatically digested to nucleosides followed by LC-ESI-MS/MS analysis. Upon conformation of the peak of F-5mC, the peak was purified by HPLC from 300 μM F-dC labeled genomic DNA. The concentration of purified F-5mC was determined by ultraviolet spectrophotometer.

Characterization of F-5caC, F-dC, and F-5mC by high-resolution LC/MS

High-resolution LC/MS spectra were obtained on an LTQ-Orbitrap Elite high-resolution mass spectrometer (Thermo-Fisher Scientific, Waltham, MA, USA) equipped with an ESI source and Dionex ultimate 3000 UPLC system (Thermo-Fisher Scientific, Waltham, MA, USA). The theoretical and measured *m/z* (ESI^+) of F-5caC were 290.0783 and 290.0784, respectively. The theoretical and measured *m/z* (ESI^+) of F-dC were 246.0885 and 246.0879,

respectively. The theoretical and measured *m/z* (ESI^+) of F-5mC were 260.1041 and 260.1038, respectively.

Characterization of F-5caC and F-dC by NMR

NMR spectra were recorded in deuterated DMSO on Bruker Avance III HD (Billerica, MA, USA).

F-5caC: ^{1}H NMR (400 MHz, DMSO-d_6): δ = 8.91 (s, 1 H), 7.93 (d, *J* = 16.0 Hz, 2 H), 5.91 (d, *J* = 17.2 Hz, 1 H), 5.55 (br. s., 1 H), 5.18 (br. s., 1 H), 5.09 - 4.80 (m, 1 H), 4.20 - 4.04 (m, 1 H), 3.91 (d, *J* = 8.6 Hz, 1 H), 3.81 (d, *J* = 12.1 Hz, 1 H), 3.67 - 3.51 (m, 1 H).

F-dC: ^{1}H NMR (400 MHz, DMSO-d_6): δ = 7.89 (d, *J* = 7.4 Hz, 1 H), 7.23 (d, *J* = 16.6 Hz, 2 H), 5.88 (d, *J* = 18.0 Hz, 1H), 5.71 (d, *J* = 7.4 Hz, 1 H), 5.54 (d, *J* = 6.4 Hz, 1 H), 5.15 (t, *J* = 5.0 Hz, 1 H), 4.87 (dd, *J* = 53.5, 3.6Hz, 1 H), 4.10 (dt, *J* = 18.6, 9.9 Hz, 1 H), 3.86 (d, *J* = 7.8 Hz, 1 H), 3.77 (d, *J* = 9.6 Hz, 1 H), 3.58 (dd, *J* = 8.2, 4.2 Hz, 1 H).

Nucleoside stability assay

For stability assay 1, F-5caC (100 μM) was incubated in DMEM medium at 37°C and 5% CO_2 for 3 d. Subsequently, Supel-Select SPE HLB tubes were used for the recovery of F-5caC. For stability assay 2, 100 μM of F-caC was added to enzymatic digestion buffer to a final volume of 30 μL. Then the solution was incubated at 95°C for 30 min. The nucleosides were then lyophilized to dryness, and then reconstituted in H_2O for LC-ESI-MS/MS analysis.

Enzymatic digestion of DNA

Nuclease digestion was carried out according to previous report with slight modification (1). Briefly, 10 μg of DNA substrate was suspended in 21 μL of H_2O, then supplied with 3 μL of 10× buffer (500 mM Tris–HCl, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM $ZnSO_4$, pH 7.0). The mixture was incubated at 95°C for 5 min and then kept on ice immediately. Subsequently, 2 μL of S1 nuclease (180 U/μL), 2 μL of phosphodiesterase I (0.001 U/μL), 1 μL of calf intestinal alkaline phosphatase (30 U/μL), and 1 μL of DNase I (5 U/μL) were added. Enzymatic digestion was carried out at 37°C for 3 h. After adding 470 μL of H_2O, the resulting solution was extracted with the same volume of chloroform twice. The aqueous layer was collected, lyophilized to dryness, and then reconstituted in H_2O for LC-ESI-MS/MS analysis.

Kinetics study

The F-5caC-DNA and 5caC-DNA were obtained from F-5caC metabolically labeled HEK293T cells and *TDG*$^{-/-}$ HEK293T cells, respectively. As for F-5caC-DNA, diverse amounts of F-5caC-DNA (1.25-20 μg) was incubated with 0.2 μM TDG protein at 37°C for 40 min. As for 5caC-DNA, diverse amounts of 5caC-DNA (1.25-20 μg) was incubated with 0.2 μM TDG protein at 37°C for 10 min. Then proteinase K (0.1 mg/mL) was added to inactive the TDG protein. After incubating at 95°C for 10 min, the solution was extracted with chloroform twice. The resulting aqueous layer was lyophilized to dryness for enzymatic digestion and LC-ESI-MS/MS analysis. The relative reaction velocity (v) was calculated from the ratio of the decrease of substrate amount over substrate amount as follows: $v \times t = (n - n_t)/n$, where t represents the reaction time. The apparent K_m and V_{max} values were obtained

from linear regression analysis of Lineweaver-Burk plots using the data points with different amounts of DNA substrate by three independent experiments.

Derivatization of AP

Girard's reagent P was used to derivatize F-AP. We used 2-deoxy-D-ribose to optimize the derivatization conditions. The reaction time, temperature and amount of Girard's reagent P were optimized to obtain the best derivatization efficiency. The reactions were performed in 100 μL H_2O with 10 pmol of 2-deoxy-D-ribose.

The derivatized product was examined by an LTQ-Orbitrap Elite high-resolution mass spectrometer. Full mass scans were acquired in positive ion mode with a mass range of *m/z* 50-300 at a resolution of 60000. A Waters Acquity UPLC® HSS T3 column (100 mm × 2.1 mm i.d., 1.8 μm) was used for the separation. The column temperature was set at 30°C. The 0.05% (v/v) formic acid in H_2O (A) and methanol (B) were employed as the mobile phases. A gradient of 0-5 min 3.5% B, 5-7 min 3.5 to 5% B, 7-13 min 5 to 80% B, 13-15 min 80% B, 15-15.5 min 80 to 3.5% B, and 15.5-20 min 3.5% B was employed. The flow rate was set at 0.3 mL/min.

Detection of F-AP in genomic DNA

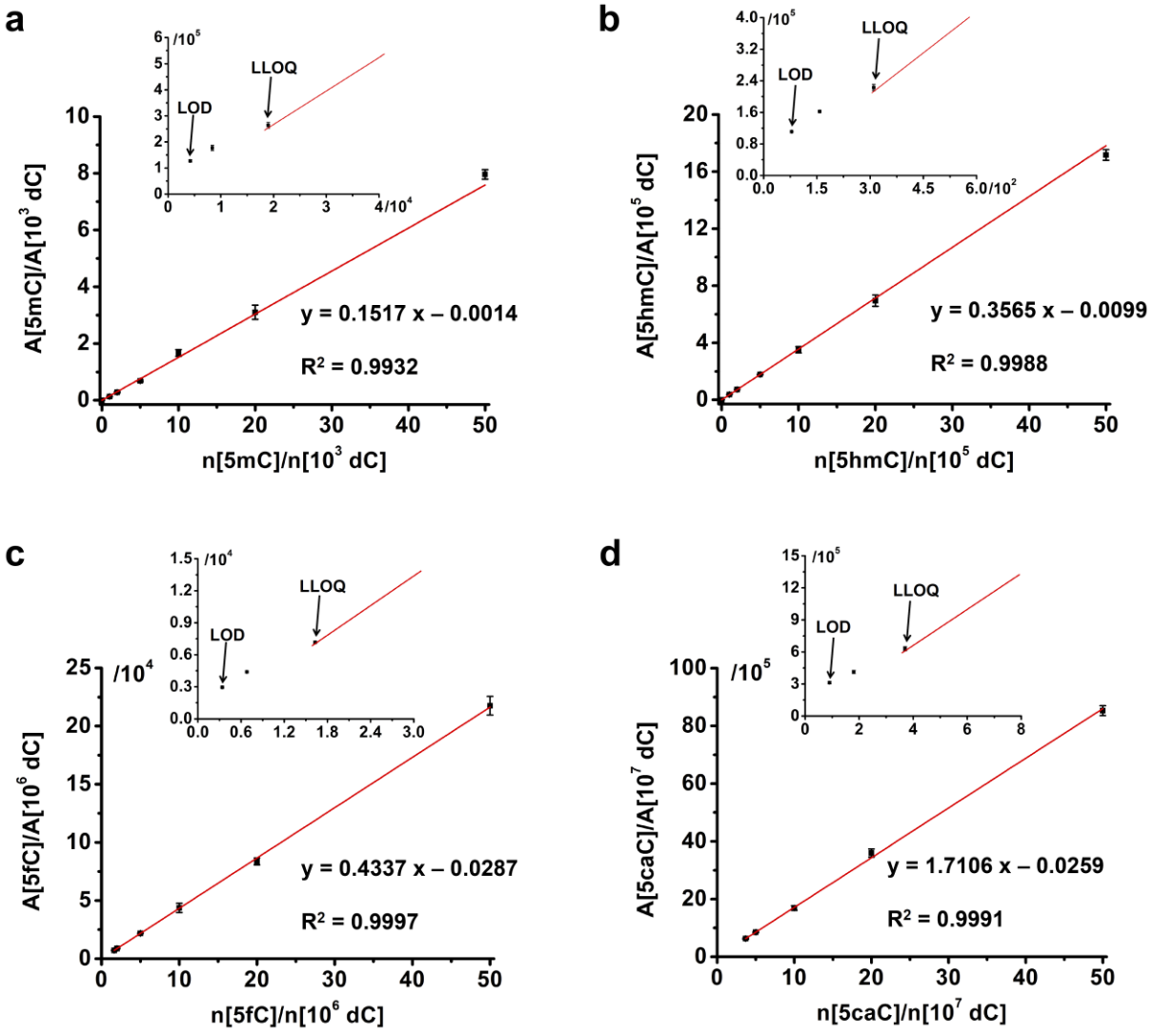
20 μg of F-5caC labeled genomic DNA from HEK293T cells and *TDG*$^{-/-}$ HEK293T cells were used to detect F-AP site. Derivatization of F-AP by 10 μM Girard's reagent P was performed in 100 μL H_2O. The mixture was incubated at 40°C for 10 min. Derivatized DNA was then precipitated with addition of glycogen (2 μL, 20 mg/mL), NH_4OAc (34 μL, 7.5 M),

and ethanol (544 μL). The mixture was incubated at -80°C for 2 h and centrifuged at 4°C at 13000 rpm for 30 min. Then the supernatant was removed, and the precipitation was washed twice by 500 μL of cold 75% ethanol. The resulting DNA pellet was dissolved in 23 μL H_2O and enzymatically digested to nucleosides for LC-ESI-MS/MS analysis. The detection of F-AP was performed by Shimadzu 8045 mass spectrometer using MRM under positive-ion mode.

Table S1. MRM transitions and collision energies for analysis of nucleosides by LC-ESI-MS/MS.

Analytes	Q1 Mass (Da)	Q3 Mass (Da)	Q1 Pre Bias (V)	Collision Energy (V)	Q3 Pre Bias (V)
dA	252.2	136.1	-20.0	-15.0	-20.0
T	243.2	127.0	-12.0	-10.0	-22.0
dC	228.2	112.1	-11.0	-40.0	-20.0
dG	268.2	152.1	-22.0	-10.0	-29.0
5mC	242.2	126.1	-20.0	-11.0	-24.0
5hmC	258.1	142.1	-12.0	-10.0	-27.0
5fC	256.0	140.3	-11.0	-11.0	-14.0
5caC	272.1	156.0	-11.0	-8.0	-29.0
rA	268.1	136.1	-20.0	-20.0	-20.0
U	245.1	113.1	-12.0	-11.0	-21.0
rC	244.1	112.1	-19.0	-15.0	-20.0
rG	284.1	152.1	-13.0	-13.0	-24.0
F-5caC	290.1	156.1	-22.0	-11.0	-16.0
F-dC	246.1	112.1	-18.0	-13.0	-20.0
F-5mC	260.1	126.1	-18.0	-13.0	-20.0
F-dA	270.2	136.1	-20.0	-15.0	-20.0
F-dG	286.2	152.1	-22.0	-10.0	-29.0
AP	268.1	178.1	-13.0	-28.0	-30.0
F-AP	286.1	196.1	-13.0	-28.0	-30.0

Figure S1. Calibration curves for quantification of 5mC, 5hmC, 5fC, and 5caC. (a-d) Calibration curves for 5mC, 5hmC, 5fC, and 5caC. (e) LODs and LLOQs of 5mC, 5hmC, 5fC, and 5caC.



e

	LOD [fmol]	LLOQ [fmol]
5mC	0.4	1.9
5hmC	0.8	3.1
5fC	3.4	16.3
5caC	0.9	3.7

Figure S2. Quantification of 5fC and 5caC levels in wt and *TDG*$^{-/-}$ HEK293T cells. (a) Extracted-ion chromatograms for monitoring 5fC. (b) Extracted-ion chromatograms for monitoring 5caC. (c) Quantification of 5fC level in wt and *TDG*$^{-/-}$ HEK293T cells. (d) Quantification of 5caC level in wt and *TDG*$^{-/-}$ HEK293T cells.

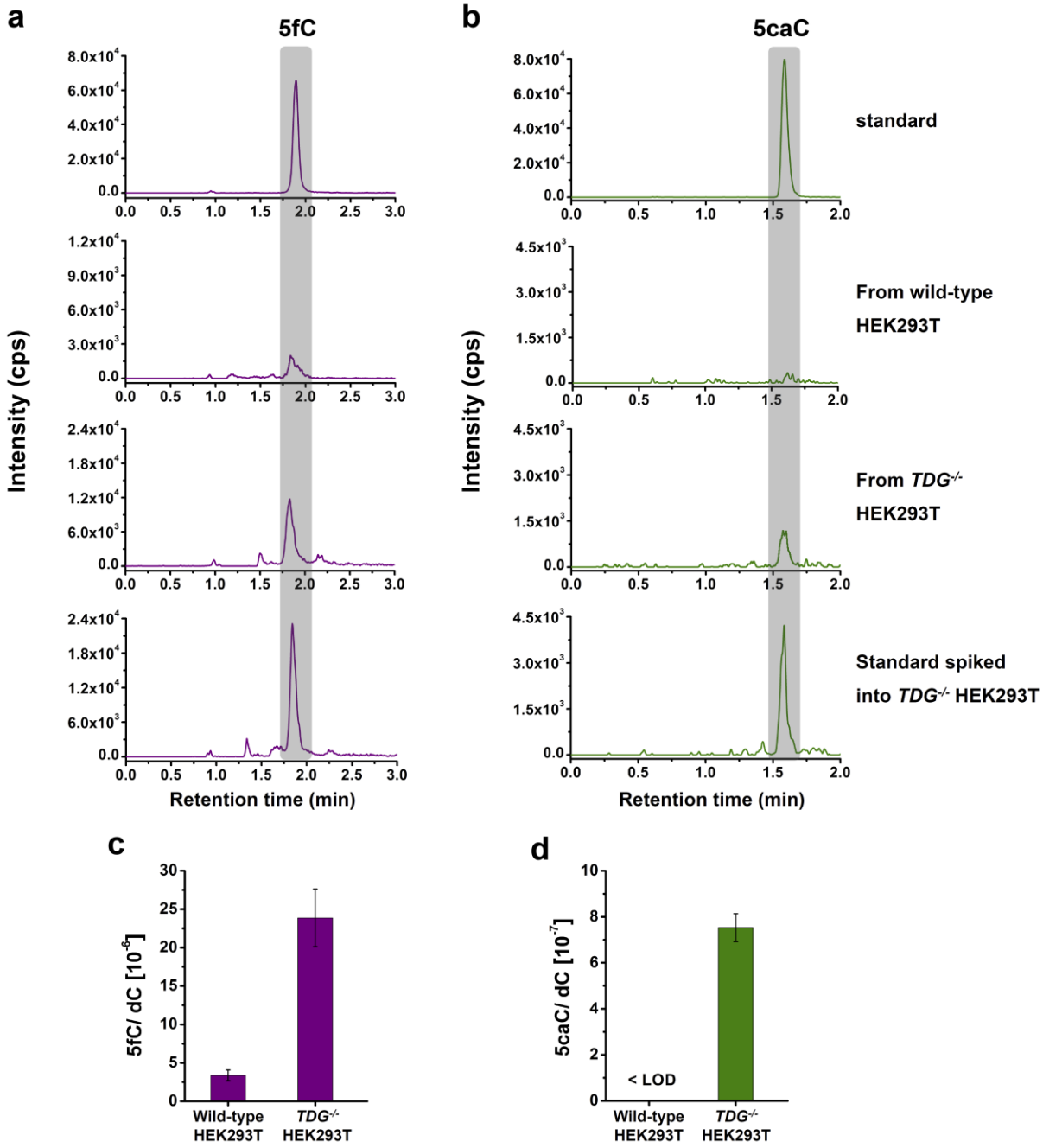


Figure S3. Quantification of 5caC level in the *in vitro* decarboxylation assay. (a) Extracted-ion chromatograms for monitoring 5caC. (b) Extracted-ion chromatograms for monitoring dC.

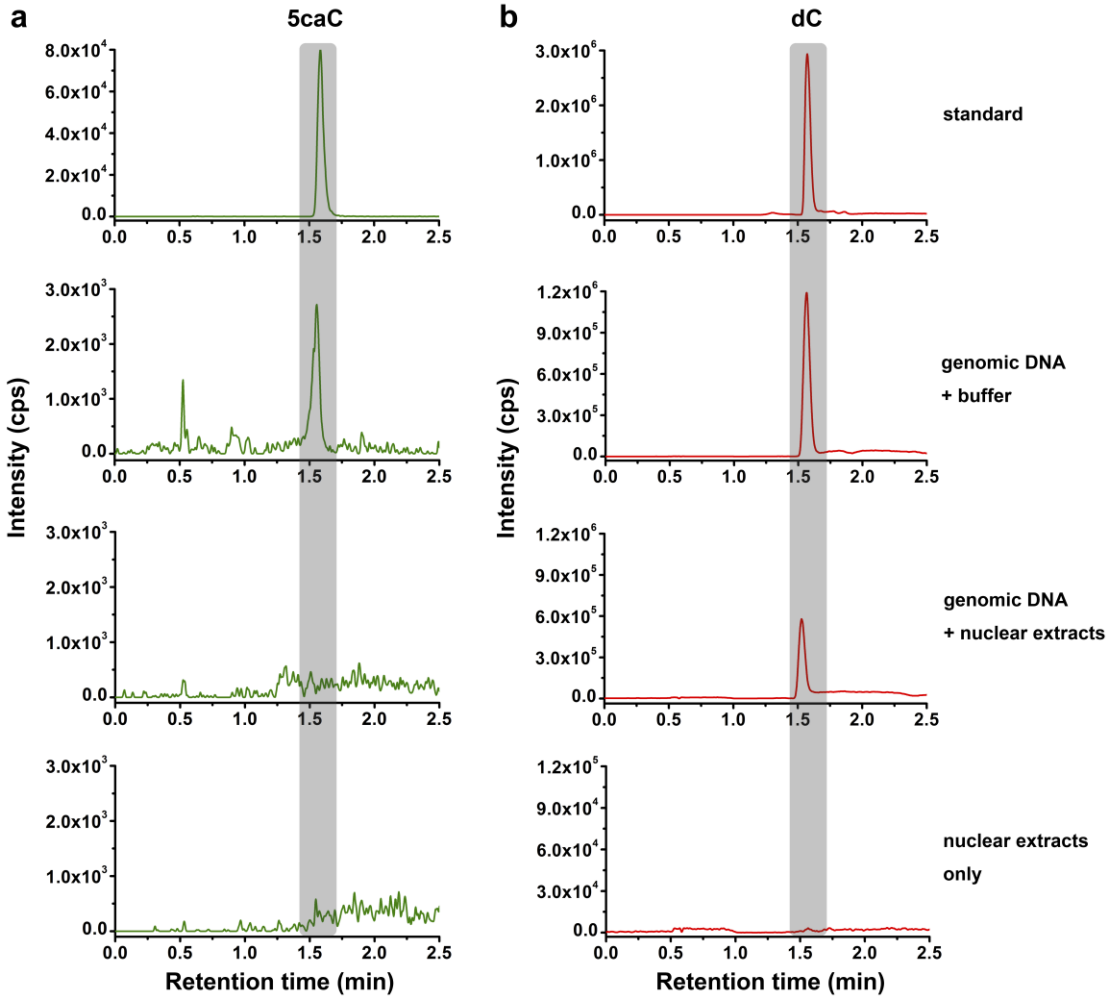


Figure S4. Quantification of 5hmC and 5fC levels in the *in vitro* decarboxylation assay. The levels of 5hmC (a) and 5fC (b) in genomic DNA upon incubation with nuclear extracts from *TDG*$^{-/-}$ HEK293T cells or extraction buffer.

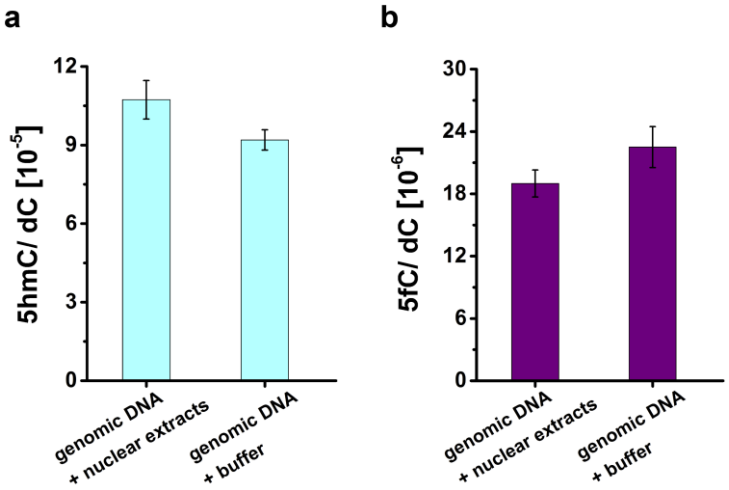


Figure S5. HPLC purification of F-5caC standard.

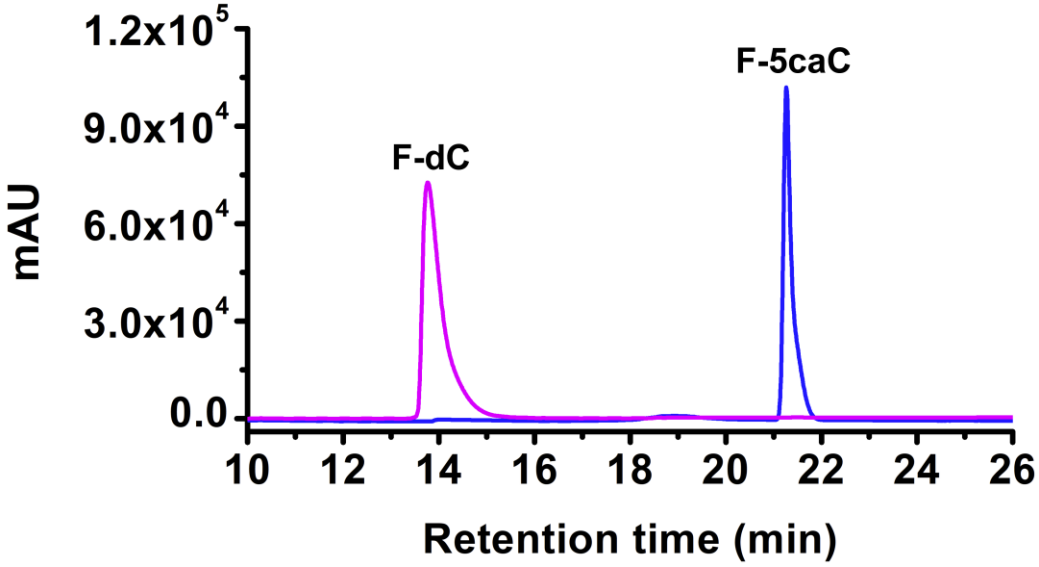


Figure S6. LC-HRMS spectra of purified F-5caC standard. (a) Extracted-ion chromatograms of F-5caC and F-dC from HPLC-purified F-5caC standard by direct injection. (b) High-resolution mass spectrum of HPLC-purified F-5caC standard.

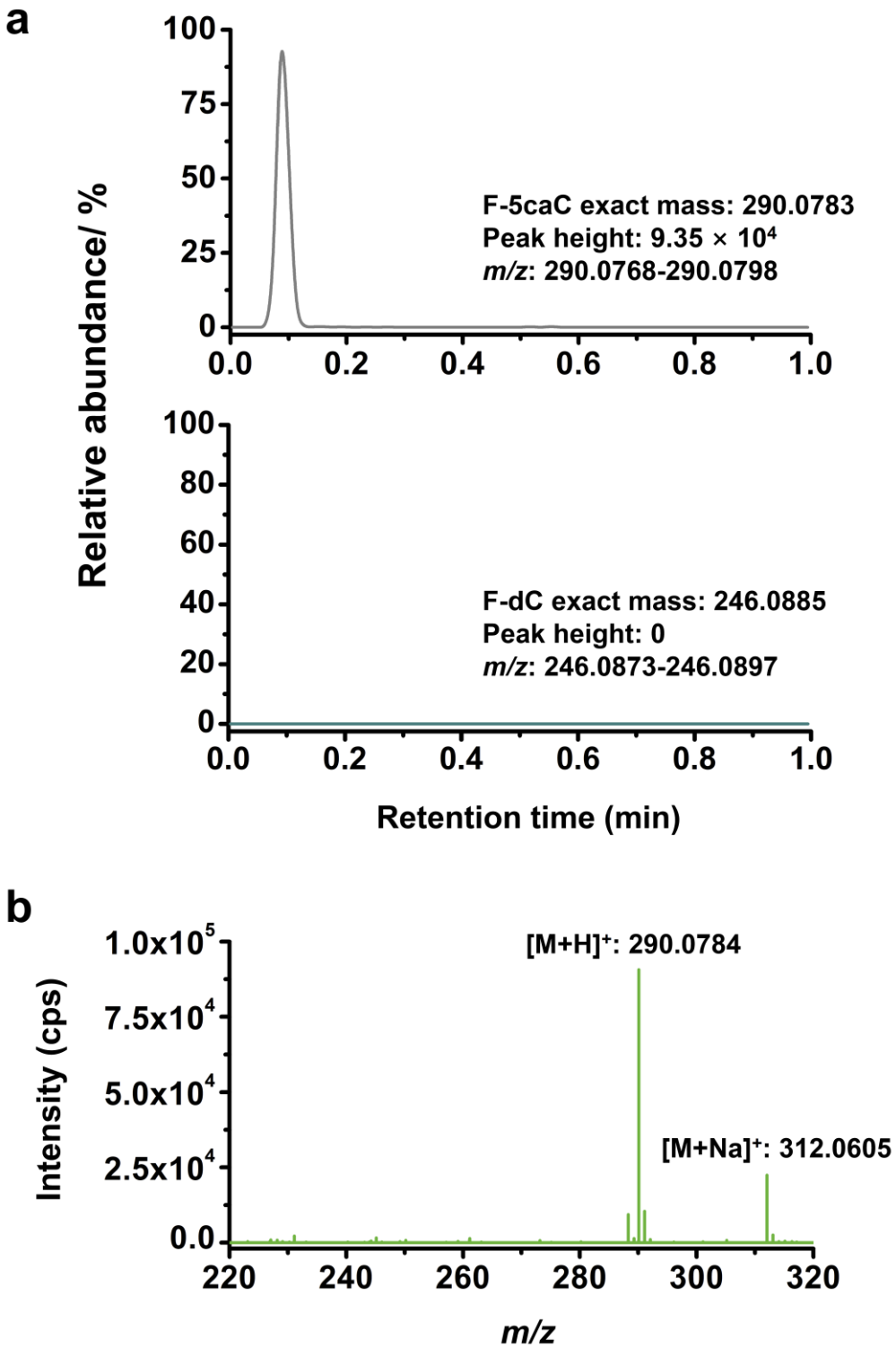


Figure S7. ^{1}H NMR spectrum of F-5caC in d_6-DMSO.

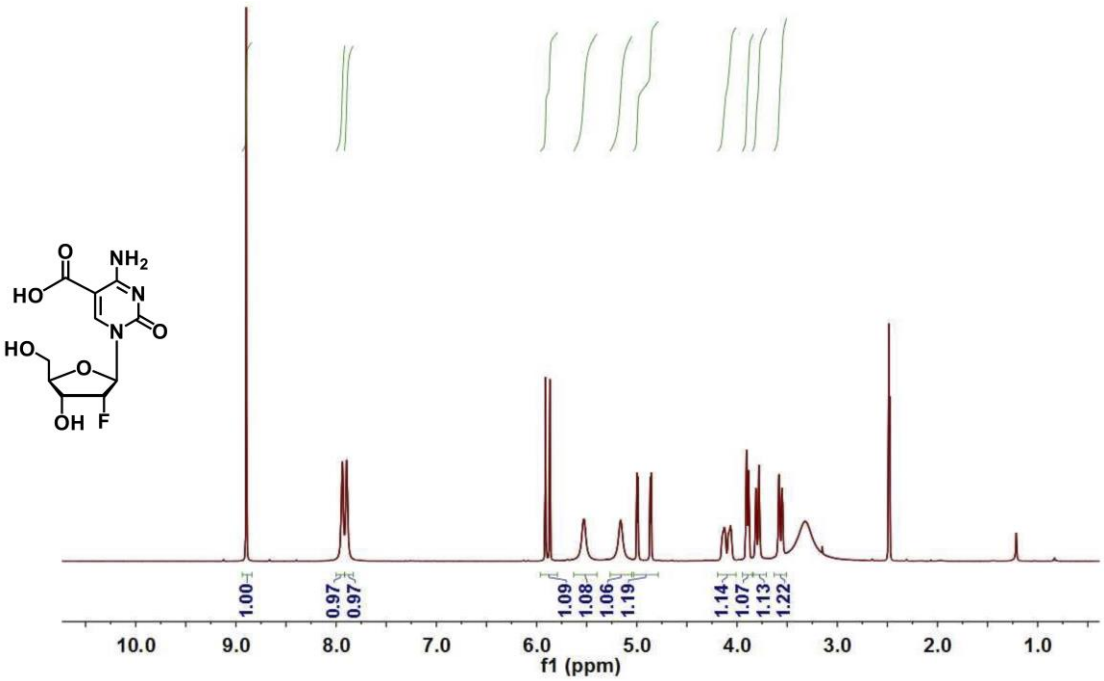


Figure S8. LC-HRMS spectra of F-dC. (a) Extracted-ion chromatogram of F-dC standard by direct injection. (b) High-resolution mass spectrum of F-dC standard.

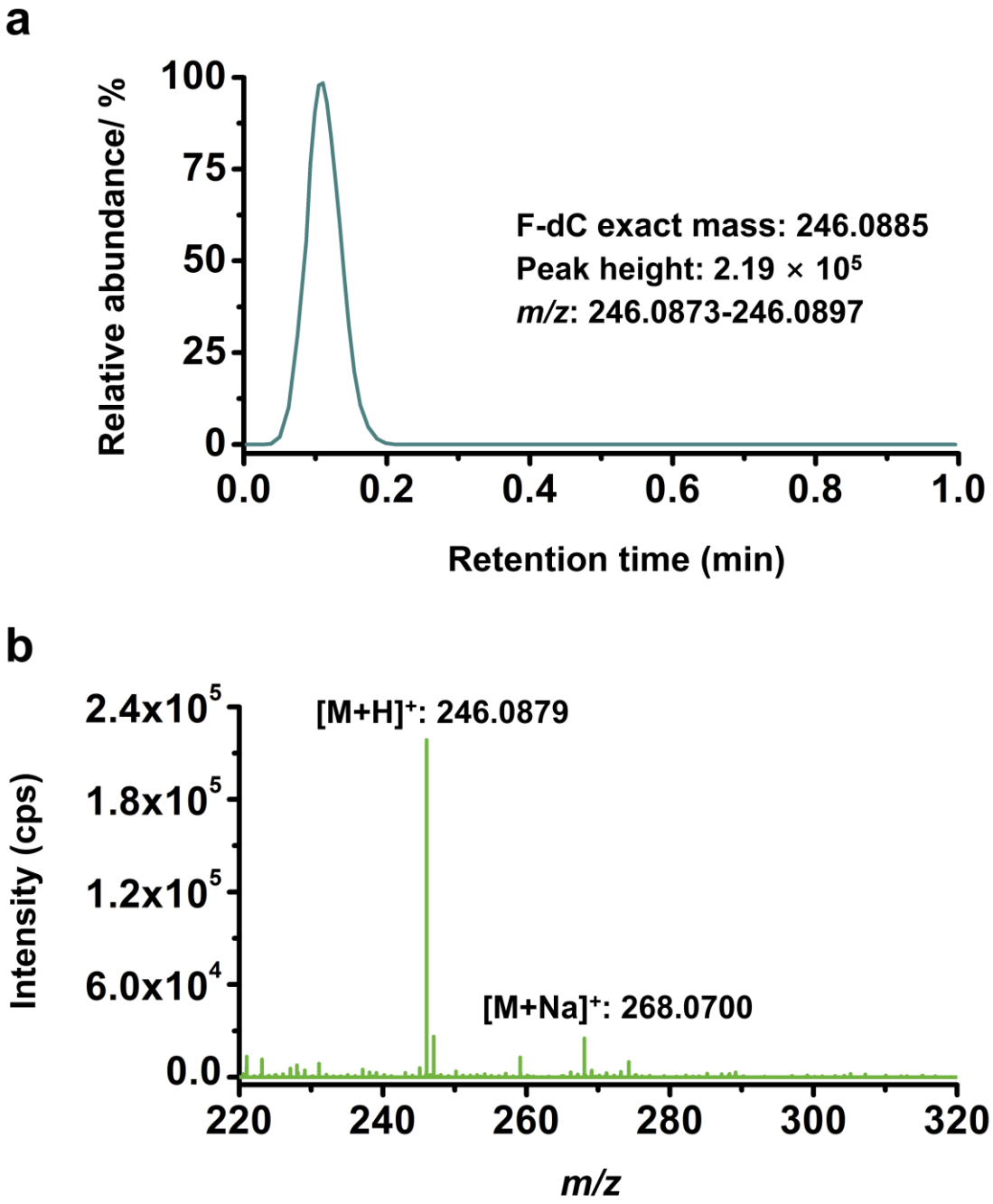


Figure S9. ^{1}H NMR spectrum of F-dC in d_6-DMSO.

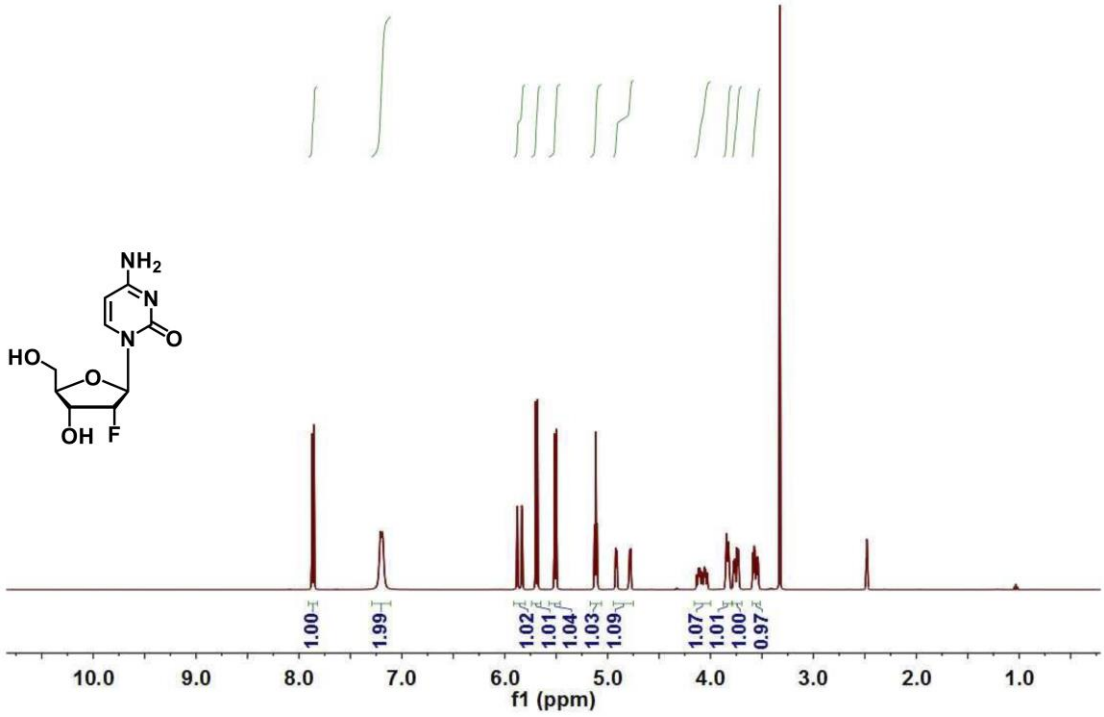


Figure S10. Purification of F-5mC standard. (a) Extracted-ion chromatograms for monitoring F-5mC released from genomic DNA of HEK293T cells upon feeding of F-dC with diverse concentrations. (b) Extracted-ion chromatogram of purified F-5mC by direct injection. (c) High-resolution mass spectrum of purified F-5mC standard.

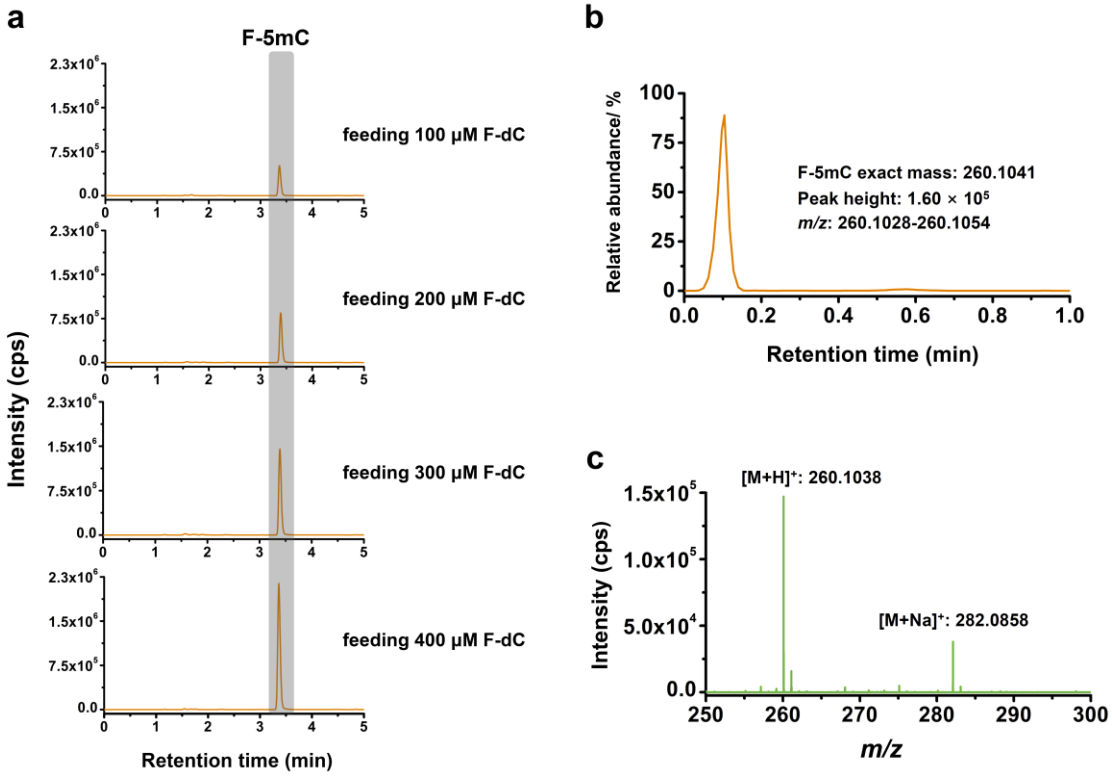
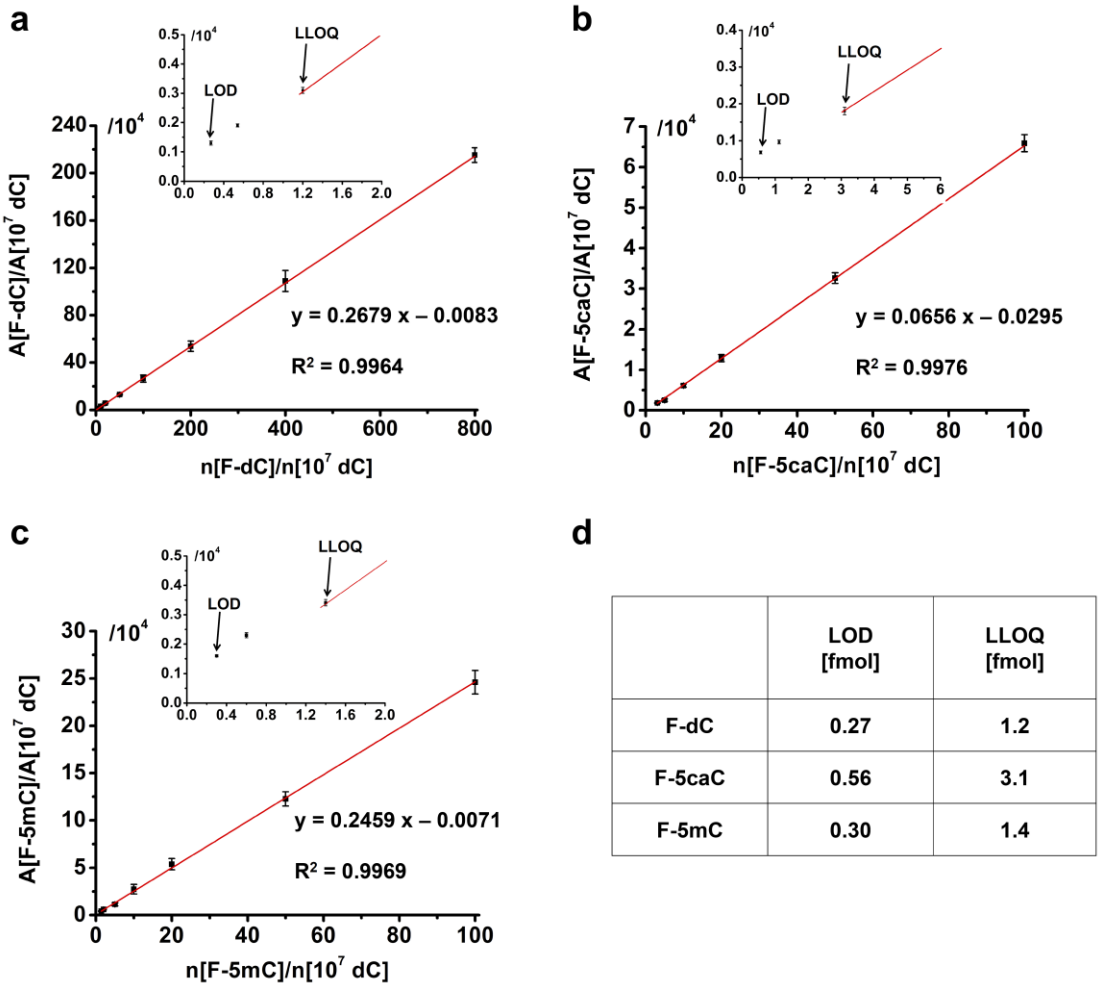


Figure S11. Calibration curves for quantification of F-dC, F-5caC, and F-5mC. (a-c) Calibration curves of F-dC, F-5caC, and F-5mC. (d) The LODs and LLOQs of F-dC, F-5caC, and F-5mC.



d

	LOD [fmol]	LLOQ [fmol]
F-dC	0.27	1.2
F-5caC	0.56	3.1
F-5mC	0.30	1.4

Figure S12. F-5caC metabolic labeling monitored by mass spectrometry. (a) Extracted-ion chromatograms of F-5caC enzymatically released from genomic DNA of HEK293T cells upon feeding of F-5caC with diverse concentrations. (b) Extracted-ion chromatograms of F-dC enzymatically released from genomic DNA of HEK293T cells upon feeding of F-5caC with diverse concentrations. The peak of rC+2 represents the natural isotope peaks of RNA cytidine.

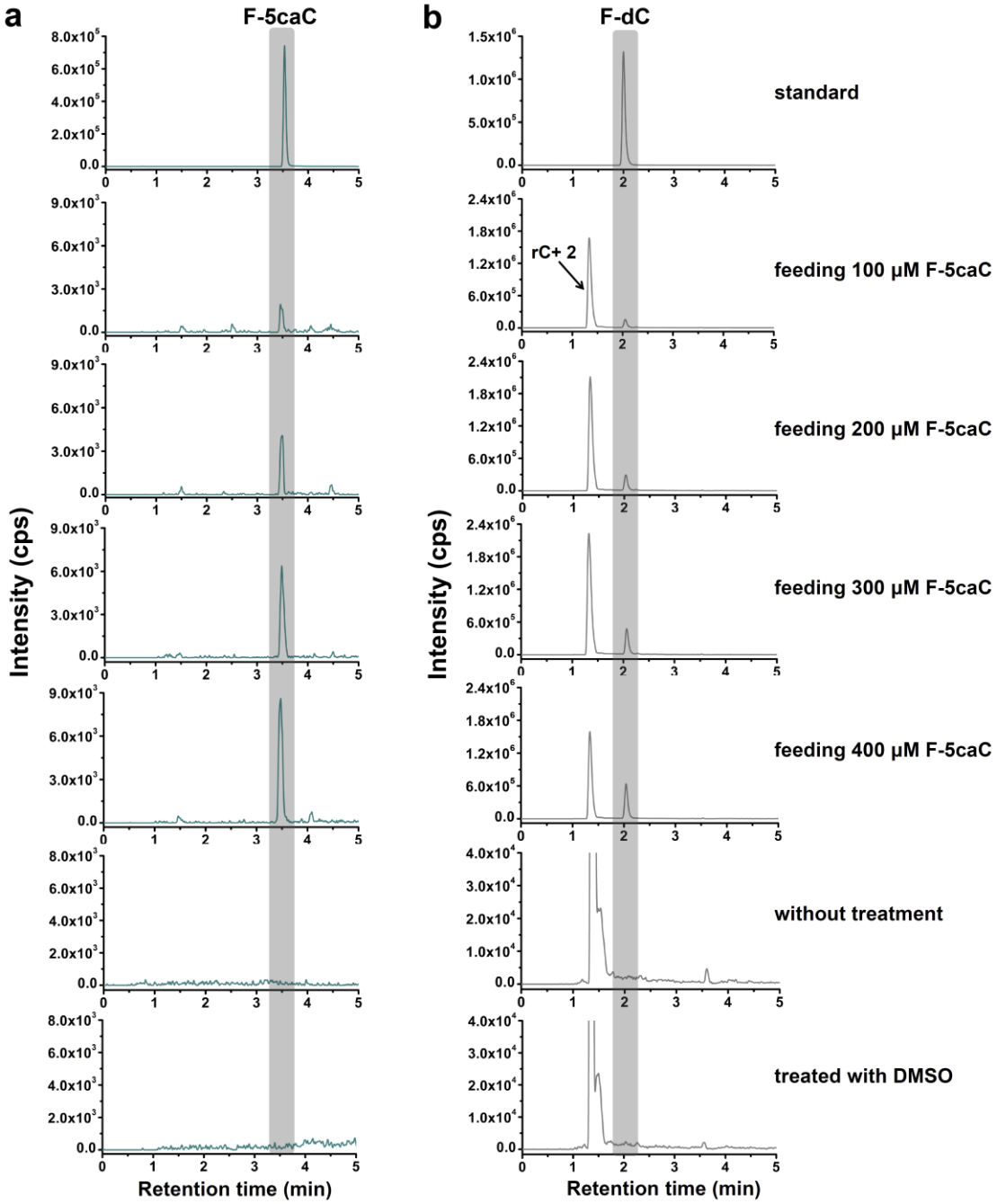


Figure S13. F-5caC metabolic labeling in wt and *TDG*$^{-/-}$ HEK293T cells. (a) Extracted-ion chromatograms for monitoring F-5caC, F-dC, and F-5mC released from genomic DNA of wt HEK293T cells upon feeding 300 μM of F-5caC. (b) Extracted-ion chromatograms for monitoring F-5caC, F-dC, and F-5mC released from genomic DNA of *TDG*$^{-/-}$ HEK293T cells upon feeding 300 μM of F-5caC. The peaks of rC+2 and m5rC+2 represented the natural isotope peaks of RNA cytidine and 5-methylcytidine, respectively.

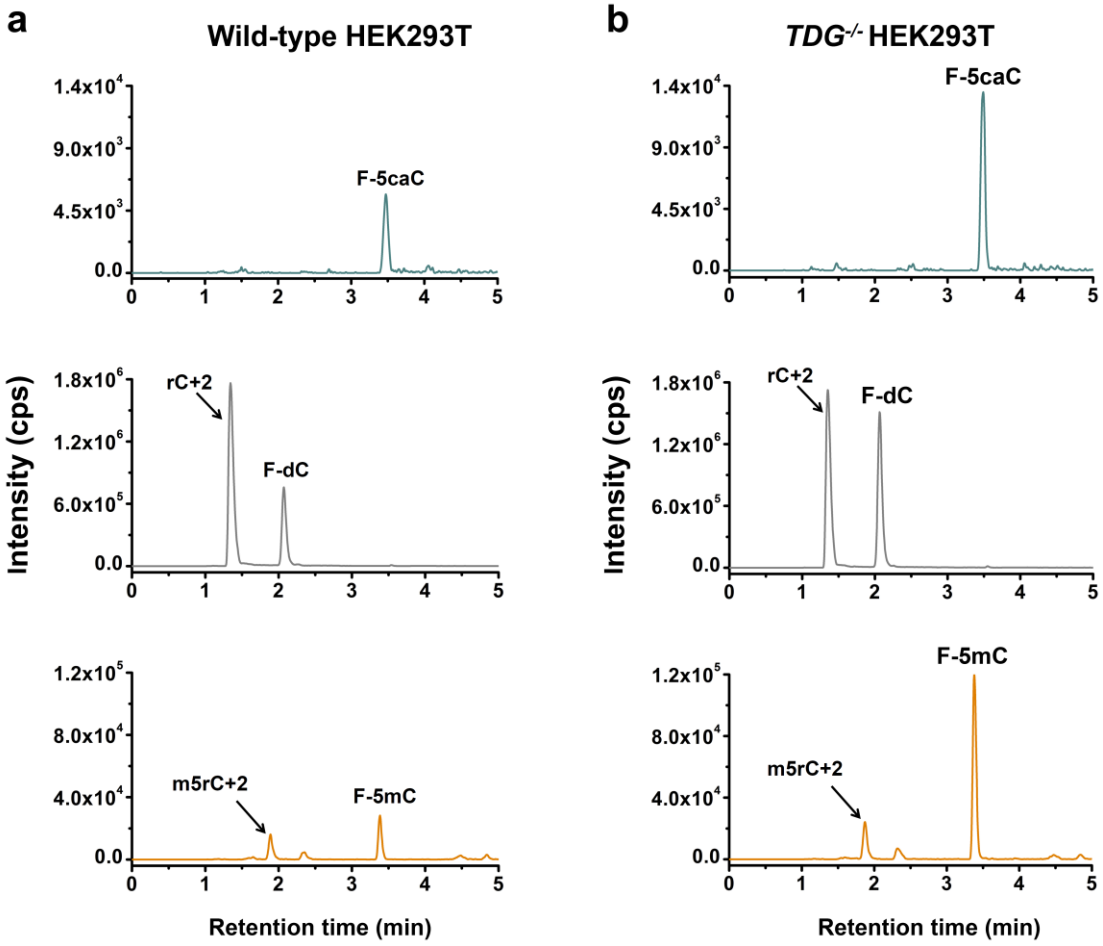


Figure S14. SDS-PAGE analysis of the recombinant human TDG protein.

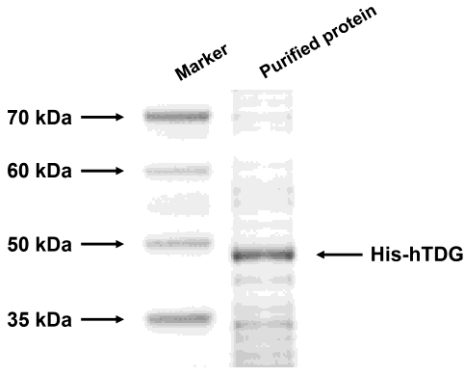


Figure S15. Assay for the evaluation of recombinant human TDG activity. (a) Schematic overview of the experimental procedure for evaluating TDG activity. Double-stranded fluorescein labeled DNA containing T/G mismatch was incubated with recombinant TDG protein. TDG protein can cleave the *N*-glycosidic bond of T, leading to the generation of AP site. The DNA would break at the AP site upon treatment with NaOH. (b) Gel electrophoresis analysis of the product of fluorescein labeled DNA containing T/G mismatch upon treatment with recombinant TDG protein.

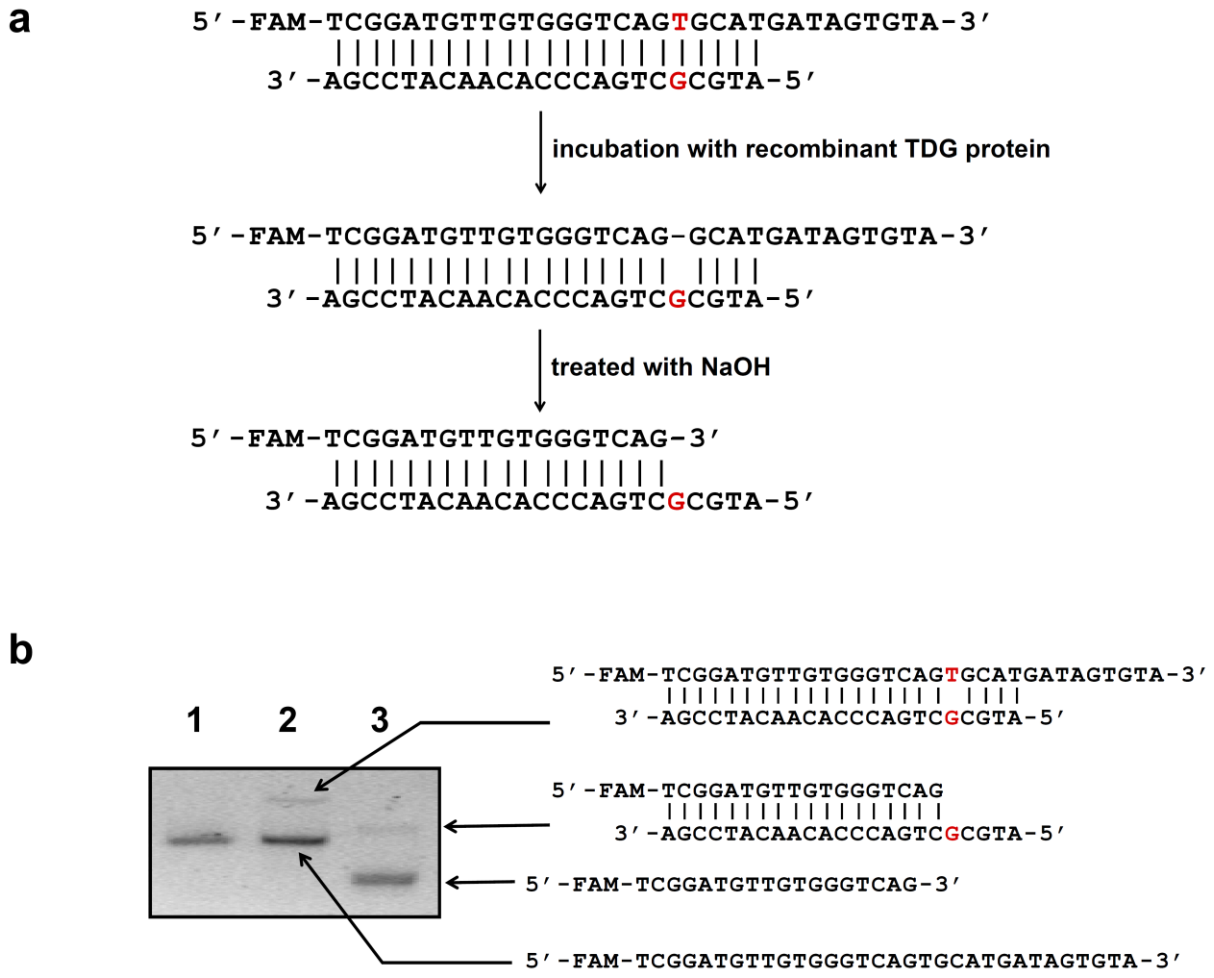


Figure S16. Evaluation of the cleavage of *N*-glycosidic bond of F-5caC by recombinant human TDG protein. (a, b) Extracted-ion chromatograms for monitoring F-5caC and F-dC released from F-5caC (300 μM) labeled genomic DNA of HEK293T cells upon incubation with recombinant TDG protein or buffer. (c) Peak areas of F-5caC and F-dC released from F-5caC (300 μM) labeled genomic DNA of HEK293T cells upon incubation with recombinant TDG protein or buffer.

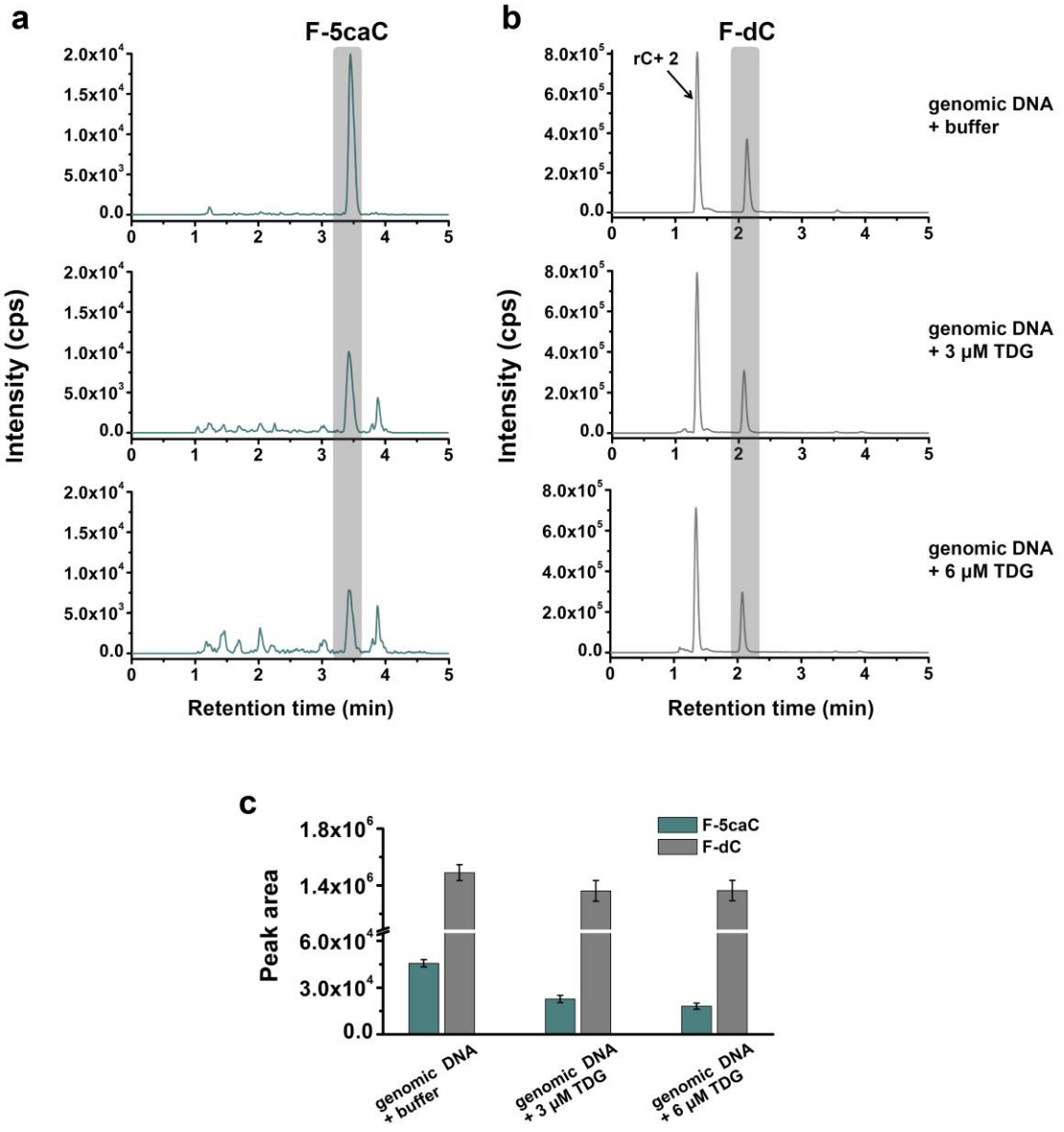
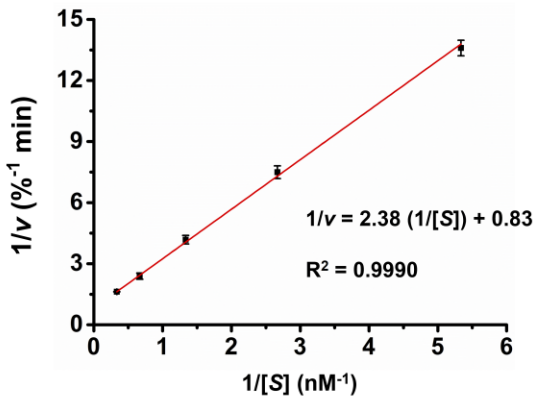
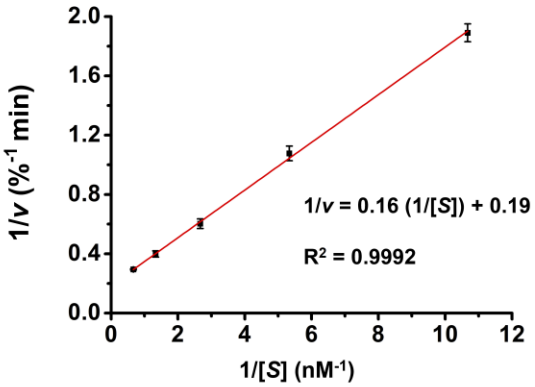


Figure S17. The kinetics of TDG in catalyzing F-5caC-DNA and 5caC-DNA substrates. (a-b) Lineweaver-Burk plots of TDG in catalyzing F-5caC-DNA and 5caC-DNA substrates. (c) The kinetics parameters for TDG in catalyzing F-5caC-DNA and 5caC-DNA substrates.

a



b



c

Enzyme	Substrate	V_{max} (% min^{-1})	K_m (nM)	V_{max}/K_m (min^{-1} nM^{-1})
TDG	F-5caC-DNA	1.2 ± 0.4	2.9 ± 0.6	4.1×10^{-3}
	5caC-DNA	5.3 ± 2.1	0.8 ± 0.2	66.3×10^{-3}

Figure S18. Derivatization of 2-deoxy-D-ribose by Girard's reagent P. (a) Derivatization reaction between 2-deoxy-D-ribose and Girard's reagent P. (b) Extracted-ion chromatograms of 2-deoxy-D-ribose and Girard's reagent P labeled 2-deoxy-D-ribose by high-resolution mass spectrometry analysis. (c) Product ion spectrum of Girard's reagent P labeled 2-deoxy-D-ribose under 28 V of collision energy.

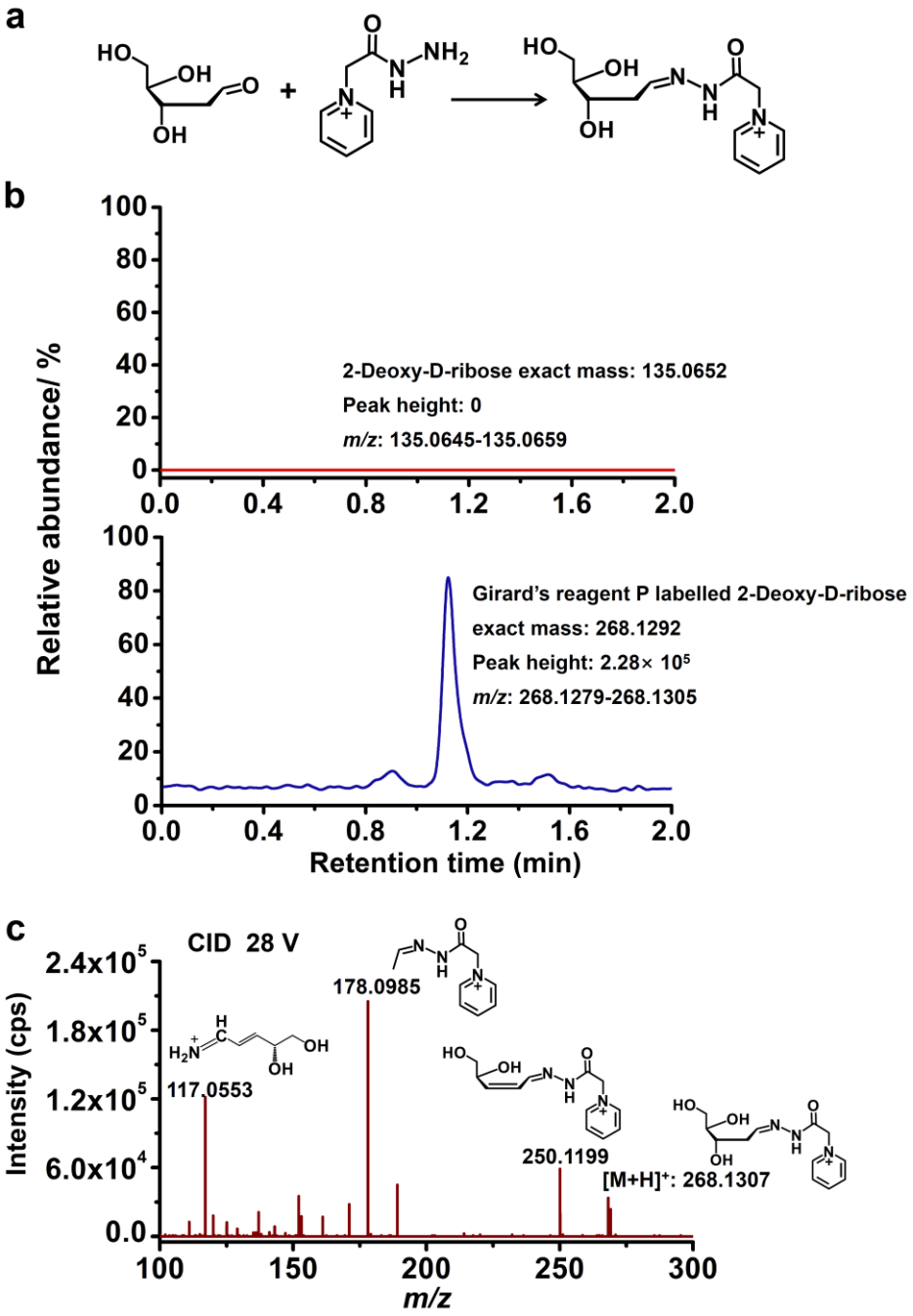


Figure S19. Optimization of derivatization reaction conditions between 2-deoxy-D-ribose and Girard's reagent P. (a) Optimization of reaction time. (b) Optimization of reaction temperature. (c) Optimization of molar ratios of Girard's reagent P over 2-deoxy-D-ribose. The optimized derivatization conditions for 2-deoxy-D-ribose by Girard's reagent P were under 40°C for 10 min with the molar ratios of Girard's reagent P/2-deoxy-D-ribose being over 100.

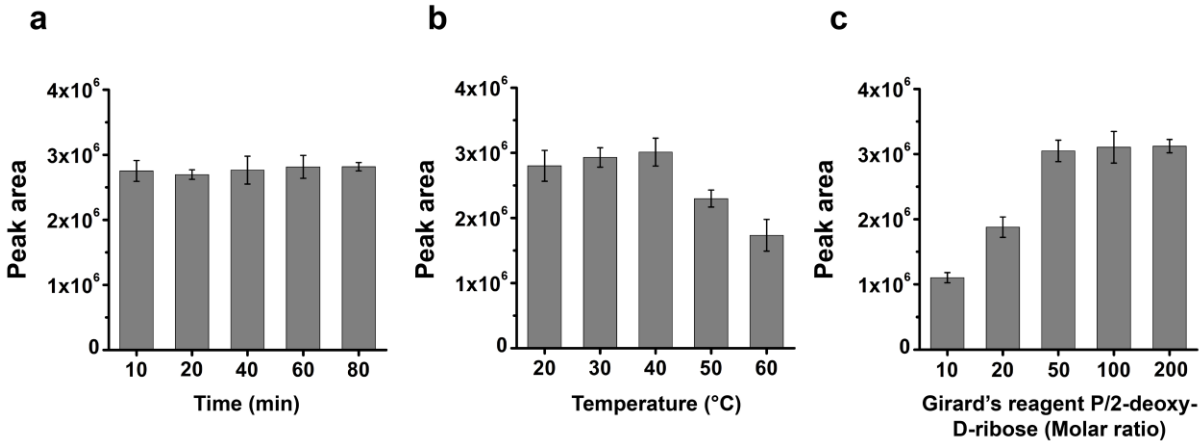


Figure S20. Detection of F-AP in genomic DNA. (a) The strategy for detecting F-AP in genomic DNA of wt and *TDG*$^{-/-}$ HEK293T cells upon feeding 300 μM of F-5caC. (b) The precursor ion and product ion of Girard's reagent P labeled F-AP site under 28 V of collision energy. (c) Extracted-ion chromatograms of Girard's reagent P labeled F-AP enzymatically released from genomic DNA of wt and *TDG*$^{-/-}$ HEK293T cells upon feeding 300 μM of F-5caC. Control, genomic DNA was omitted. (d) Quantitative data of F-AP in genomic DNA of wt and *TDG*$^{-/-}$ HEK293T cells.

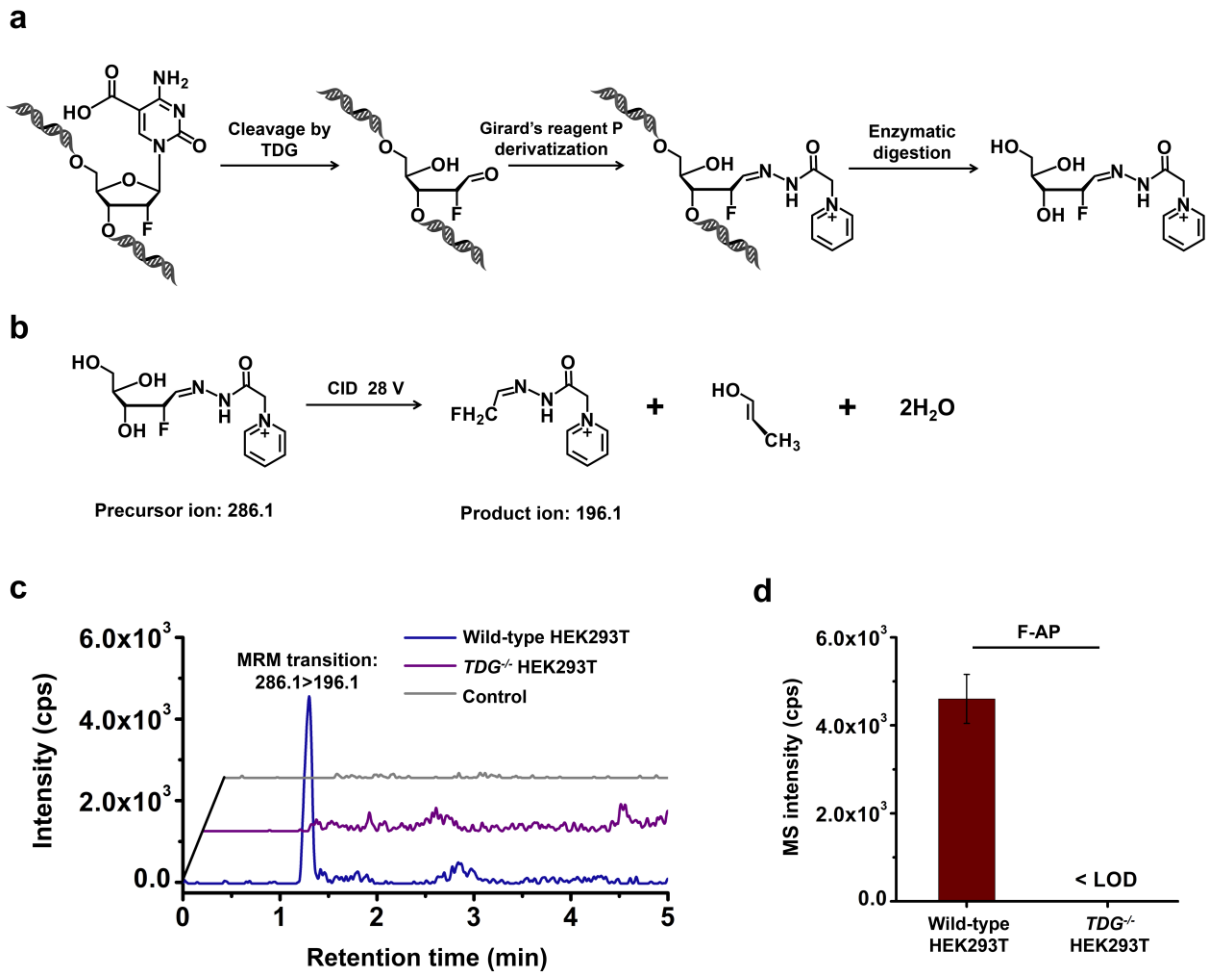


Figure S21. Schematic overview of experimental procedure for the extraction of nucleoside/nucleotide soluble pool.

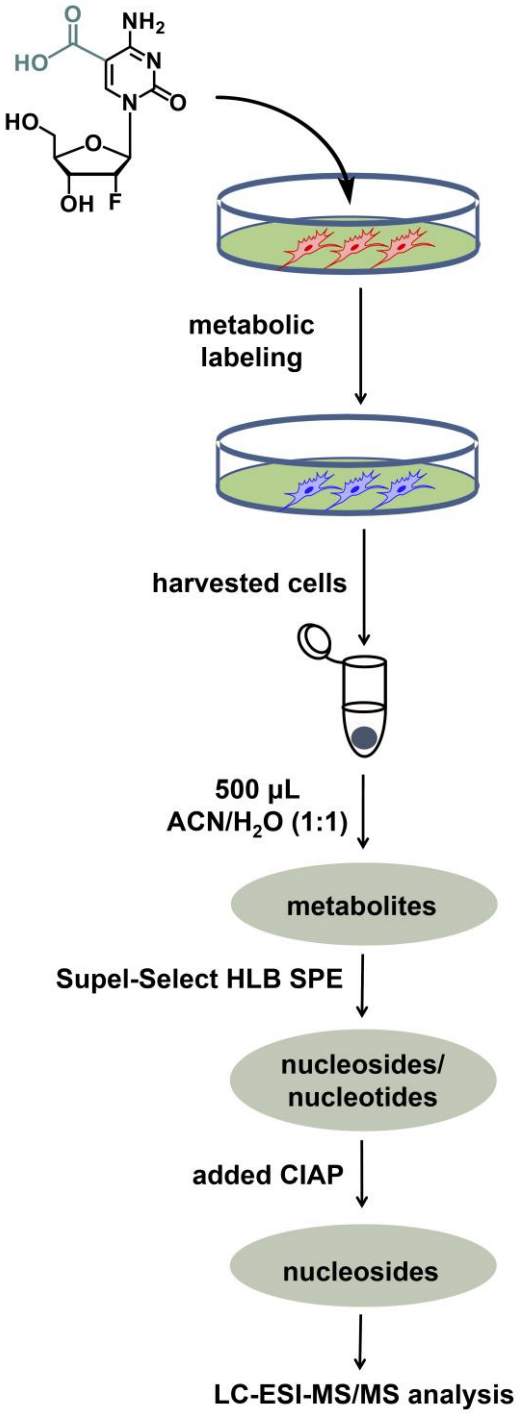


Figure S22. Evaluation of the decarboxylation of F-5caC in the soluble nucleoside/nucleotide pool. (a, b) Extracted-ion chromatograms for monitoring F-5caC and F-dC in the soluble pool after feeding 300 μM of F-5caC for different time. (c, d) Quantitative data of F-5caC and F-dC in the soluble pool after feeding 300 μM of F-5caC for different time.

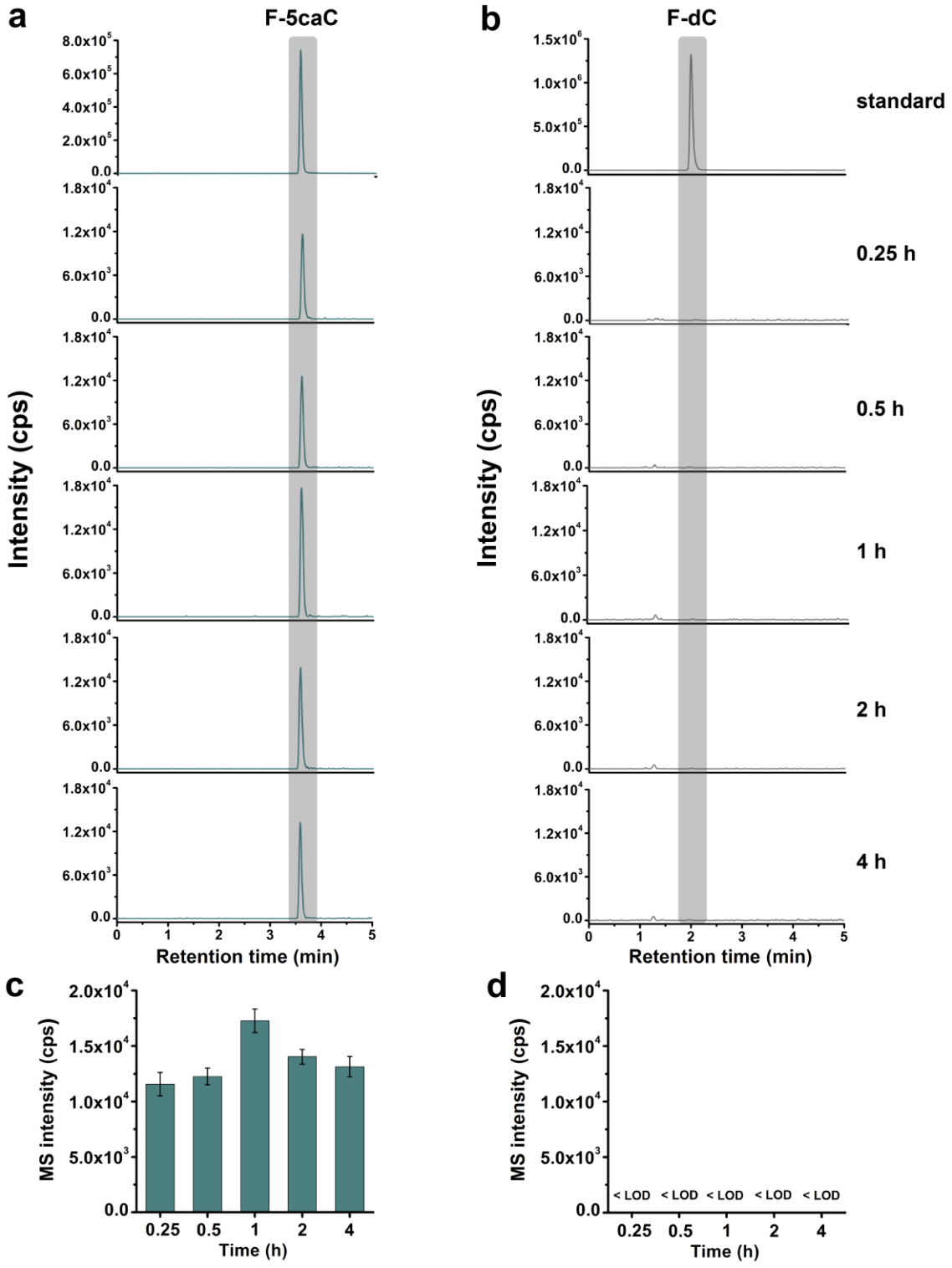


Figure S23. Extracted-ion chromatograms for monitoring F-dC in genomic DNA of HEK293T cells upon feeding F-dC with diverse concentrations (2 nM – 500 nM).

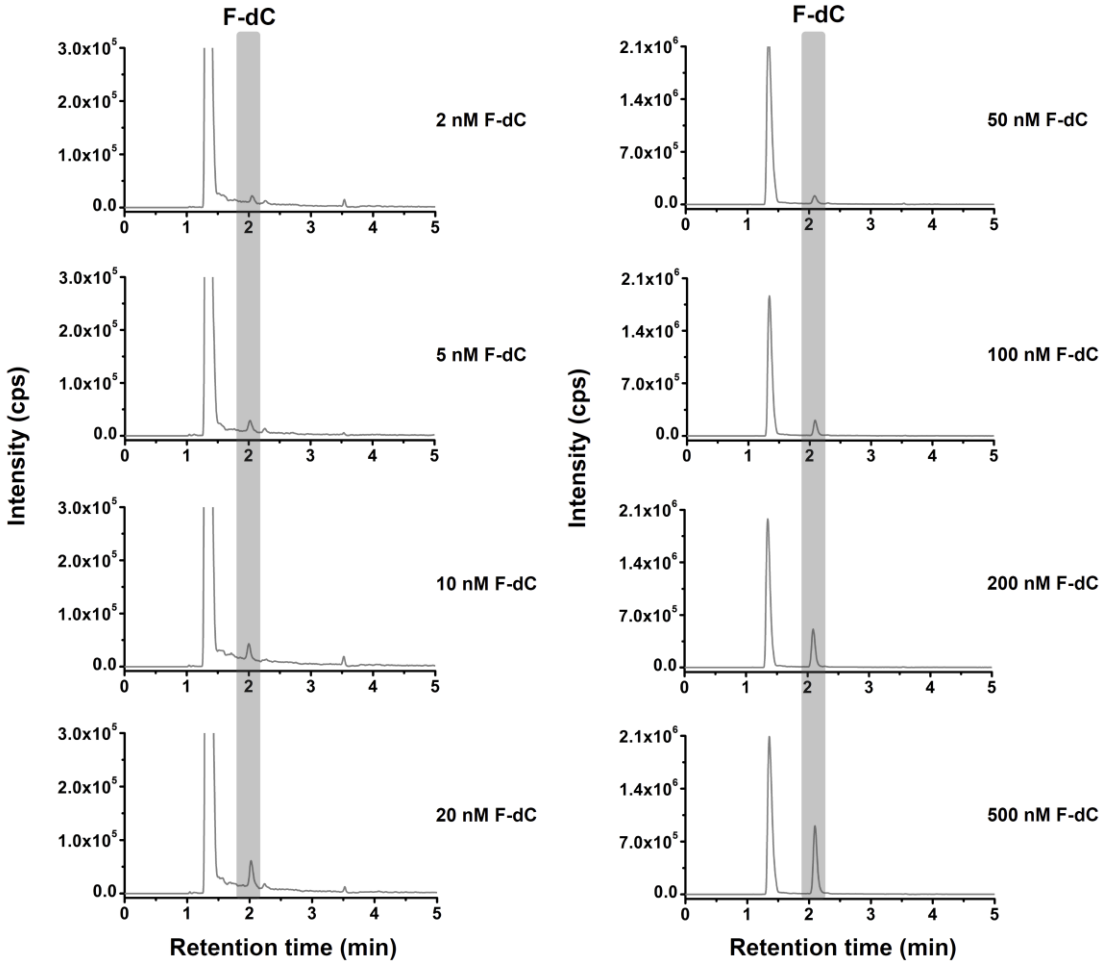


Figure S24. Evaluation of the stability of F-5caC under different conditions. (a, b) Extracted-ion chromatograms for monitoring F-5caC and F-dC. (c, d) Quantification of the levels of F-5caC and F-dC. Assay 1, F-5caC was incubated in DMEM medium at 37°C for 3 d. Assay 2, F-5caC was incubated in the enzymatic digestion buffer at 95°C for 30 min.

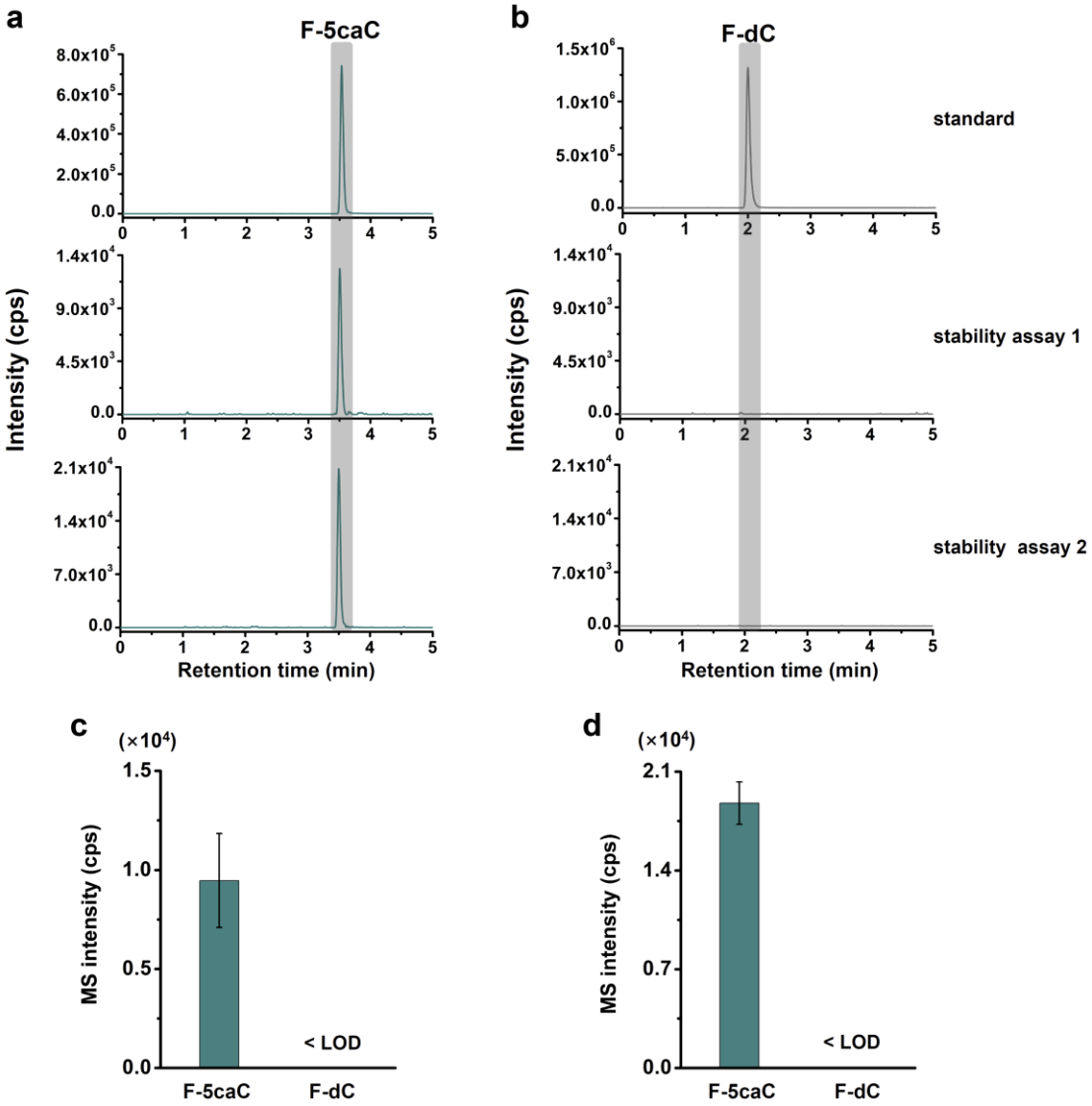


Figure S25. The time-course of decarboxylation experiment. Extracted-ion chromatograms for monitoring F-5caC (a), F-dC (b), and F-5mC (c) released from genomic DNA of HEK293T cells upon feeding 300 μM of F-5caC for different time.

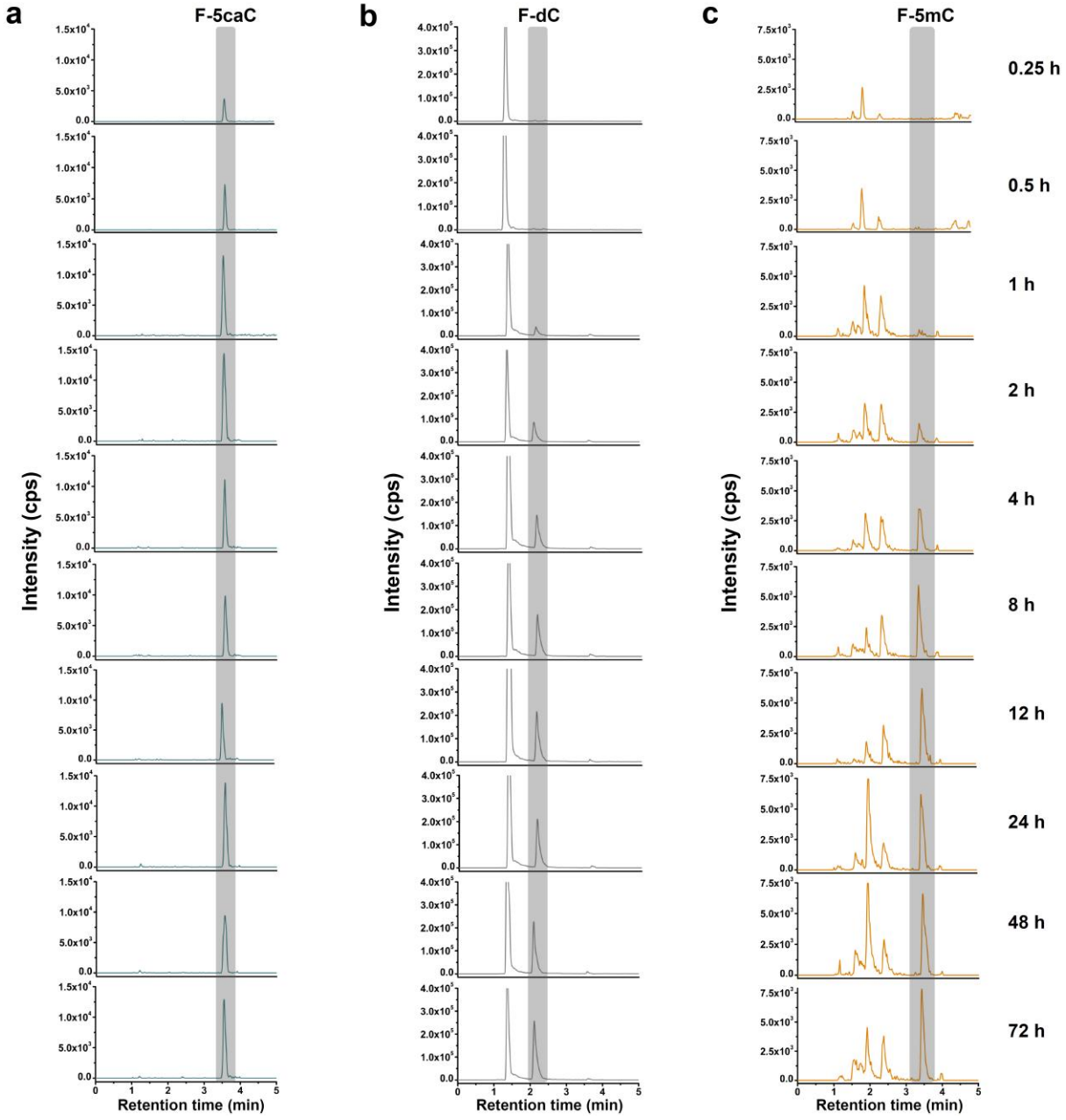
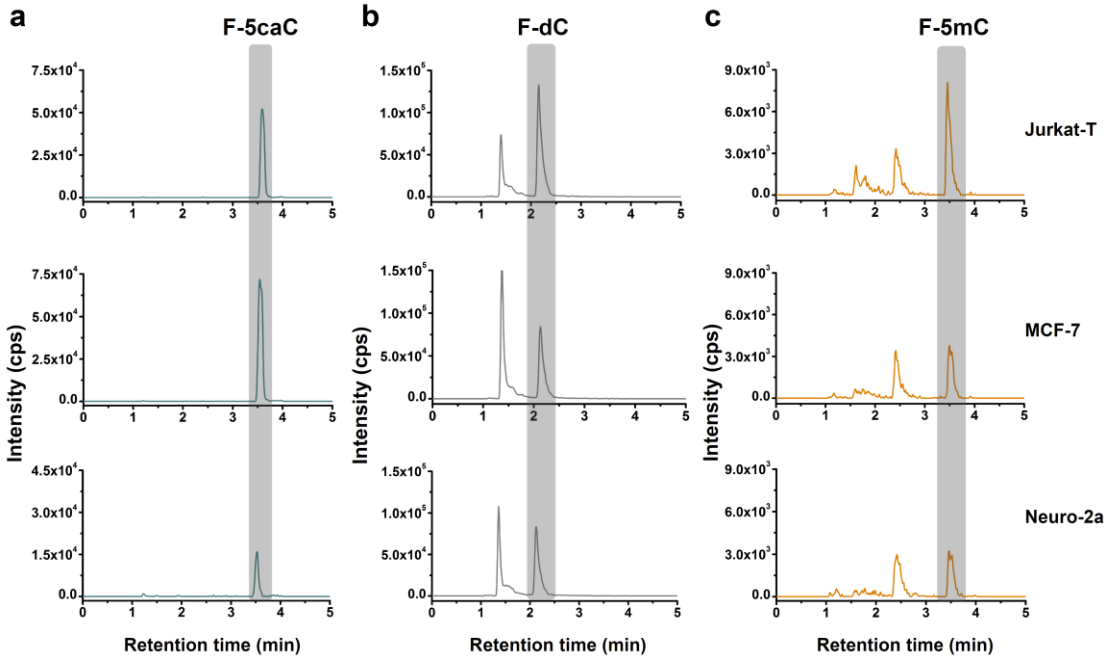


Figure S26. Decarboxylation of F-5caC occurs in different mammalian cells. Extracted-ion chromatograms for monitoring F-5caC (a), F-dC (b), and F-5mC (c) released from genomic DNA of various cell lines upon feeding 300 μM of F-5caC for 3 d.



Reference

1. Xiong, J., Ye, T.T., Ma, C.J., Cheng, Q.Y., Yuan, B.F. and Feng, Y.Q. (2019) N6-Hydroxymethyladenine: a hydroxylation derivative of N6-methyladenine in genomic DNA of mammals. *Nucleic Acids Res*, **47**, 1268-1277.