Supporting Information

For

Novel dual methylation of cytidines in RNA of mammals

Ming-Yu Cheng,[1] Xue-Jiao You,[1] Jiang-Hui Ding,[1] Yi Dai,[1] Meng-Yuan Chen,[1] Bi-Feng Yuan,*[,1,2] Yu-Qi Feng[1,2]

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

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

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

The Supporting Information includes following items:

Evaluation of the purity of isolated 18S rRNA by real-time quantitative PCR

A series of different concentrations of 16S rRNA were reverse transcribed and then analyzed by real-time quantitative PCR to construct the calibration curves. The calibration curves were generated by plotting the Ct values against the concentrations of 16S rRNA template. The obtained linear equation ($Ct = -3.5lgc_{16S\ rRNA} - 23.0$) showed good linearity with the coefficient of determination (R^2) being 0.993 (Figure S4A in Supporting Information). The calculated amplification efficiency of the real-time quantitative PCR was 93%.

To evaluate the potential *E. coli* 16S rRNA level in the isolated HeLa 18S rRNA, the equal amounts of *E. coli* 16S rRNA and isolated HeLa 18S rRNA were used as the templates for real-time quantitative PCR. The potential *E. coli* 16S rRNA in HeLa 18S rRNA are calculated as follows:

$$\Delta Ct = Ct^{16S\ rRNA} - Ct^{18S\ rRNA}$$

$$\text{percentage of } \textit{E. coli} \text{ 16S rRNA in HeLa 18S rRNA } = 2^{\Delta Ct} \times 100\%$$

Using the above equation, the measured percentage of potential *E. coli* 16S rRNA in HeLa 18S rRNA was 0.001%. In other word, the purity of the isolated HeLa 18S rRNA was 99.999%

The measured content of m^4Cm in *E. coli* 16S rRNA was 0.28% (m^4Cm/C, Figure S4D in Supporting Information). With 0.001% of *E. coli* 16S rRNA in the isolated HeLa 18S rRNA, the theoretical content of m^4Cm should be 0.0000028% (m^4Cm/C). But the measured level of m^4Cm in the isolated HeLa 18S rRNA was 0.0009% (m^4Cm/C, Table S6 in Supporting Information), which was 321 folds higher than 0.0000028%. The results clearly showed that HeLa 18S rRNA contains m^4Cm and the potential *E. coli* 16S rRNA in the isolated HeLa 18S rRNA could be negligible.

Enzymatic digestion of RNA

Briefly, 2 uL of 10 × buffer (500 mM Tris–HCl, 10 mM $MgCl_2$, 100 mM NaCl, 10 mM $ZnSO_4$, pH 7.0), 1 uL of S1 nuclease (180 U/mL), 1 uL of venom phosphodiesterase I (0.001 U/mL), and 0.5 uL of CIAP (30 U/mL) were added to 3 ug of RNA dissolved in 15.5 uL H_2O. The mixture was incubated at 37℃ for 4 h. Then 180 μL of H_2O and 200 μL of chloroform were added followed by vortexing for 3 min and centrifugation at 12000 rpm for 5 min. The upper aqueous phase was collected. The chloroform extraction was repeated for three times. The aqueous phase containing nucleosides was collected and lyophilized to dry, and then subjected to LC-ESI-MS/MS analysis.

Table S1. The information of nucleoside standards.

No	Nucleosides	Abbreviation	CAS number	Molecular formula	Molecular weigh	Company
1	Adenosine	A	58-61-7	$C_{10}H_{13}N_5O_4$	267.24	Sigma-Aldrich
2	Cytidine	C	65-46-3	$C_9H_{13}N_3O_5$	243.22	Sigma-Aldrich
3	Uridine	U	58-96-8	$C_9H_{12}N_2O_6$	244.20	Sigma-Aldrich
4	Guanosine	G	118-00-3	$C_{10}H_{13}N_5O_5$	283.24	Sigma-Aldrich
5	N^4,2'-*O*-dimethylcytidine	m^4Cm	13048-95-8	$C_{11}H_{17}N_3O_5$	271.27	Carbosynth
6	5,2'-*O*-dimethylcytidine	m^5Cm	113886-70-7	$C_{11}H_{17}N_3O_5$	271.27	Carbosynth
7	2'-*O*-methylcytidine	Cm	2140-72-9	$C_{10}H_{15}N_3O_5$	257.25	Carbosynth
9	Cytidine -$^{13}C_5$	rC-$^{13}C_5$	-	$C^{13}{}_5C_4H_{13}N_3O_5$	248.18	CATO
8	3,2'-*O*-dimethylcytidine	m^3Cm	-	$C_{11}H_{17}N_3O_5$	271.27	Synthesized in the current study

Table S2. The MRM parameters for analysis of nucleosides by LC-ESI-MS/MS using a Shimadzu 8045 mass spectrometer.

Nucleosides	Parent ion (*m/z*)	Daughter ion (*m/z*)	Q1 (V)	CE (V)	Q3 (V)	RT (min)
m^3Cm	272.1	126.1	-21	-14	-22	3.92
m^4Cm	272.1	126.1	-21	-14	-22	3.36
m^5Cm	272.1	126.1	-11	-13	-23	4.19
D_3-m^3Cm	275.1	129.1	-21	-14	-22	3.92
m^3Cm-D_3	275.1	126.1	-21	-14	-22	3.92
D_3-m^3Cm-D_3	278.1	129.1	-21	-14	-22	3.92
D_3-m^4Cm	275.1	129.1	-21	-14	-22	3.36
m^4Cm-D_3	275.1	126.1	-21	-14	-22	3.36
D_3-m^4Cm-D_3	278.1	129.1	-21	-14	-22	3.36
D_3-m^5Cm	275.1	129.1	-21	-14	-22	4.19
m^5Cm-D_3	275.1	126.1	-21	-14	-22	4.19
D_3-m^5Cm-D_3	278.1	129.1	-21	-14	-22	4.19
A	268.1	136.0	-	-50	-	4.09
C	244.1	112.1	-	-50	-	1.37
G	284.1	152.0	-	-50	-	4.32
U	245.1	113.0	-	-50	-	2.38
rC-$^{13}C5$	249.2	112.1	-	-50	-	1.37

Table S3. Deuterated labeling percentages of ribose and nucleobase of m^3Cm, m^4Cm and m^5Cm.

Modification	Deuteration labeling percentage of nucleobase (%)	Deuteration labeling percentage of ribose (%)
m^3Cm	55.0±0.2	60.5±0.2
m^4Cm	64.3±1.1	68.6±0.9
m^5Cm	49.5±0.8	47.9±1.6

Table S4. Calibration curves, LODs and LOQs for the analysis of m^3Cm, m^4Cm and m^5Cm.

Analytes	Linear range	Calibration curve data			LOD	LOQ
		Slope	Intercept	R^2 value	(fmol)	(fmol)
C	0.05-5.0 (nmol)	9.213×10^{-1}	-1.170×10^{-3}	0.9997	11.15	37.19
m^3Cm	0.5-200.0 (fmol)	9.183×10^{-4}	3.378×10^{-6}	0.9897	0.0206	0.0686
m^4Cm	0.5-200.0 (fmol)	6.892×10^{-4}	6.547×10^{-5}	0.9990	0.0587	0.1960
m^5Cm	0.5-200.0 (fmol)	6.276×10^{-4}	2.515×10^{-5}	0.9950	0.0196	0.0653

Table S5. Accuracy and precision for the analysis of m^3Cm, m^4Cm and m^5Cm.

Analytes	QCs	Theoretical Value	Measured Value	Relative Error (%)	Intra-day RSD%, n=3	Inter-day RSD%, n=3
C	Low	0.10 (nmol)	0.10 (nmol)	1.8	0.9	2.9
	Medium	0. 50 (nmol)	0.50 (nmol)	0.5	0.9	3.3
	High	2.00 (nmol)	2.06 (nmol)	3.1	0.8	2.5
m^3Cm	Low	0. 50 (fmol)	0.51 (fmol)	2.8	2.2	8.3
	Medium	5.00 (fmol)	4.91 (fmol)	-1.9	3.0	6.9
	High	50.00 (fmol)	45.62 (fmol)	-8.8	1.5	6.2
m^4Cm	Low	0. 50 (fmol)	0.43 (fmol)	-13.7	4.8	8.3
	Medium	5.00 (fmol)	5.20 (fmol)	4.0	1.3	2.4
	High	50.00 (fmol)	48.82 (fmol)	-2.4	1.0	6.0
m^5Cm	Low	0. 50 (fmol)	0.51 (fmol)	-1.9	7.3	14.9
	Medium	5.00 (fmol)	5.31 (fmol)	6.9	3.0	12.6
	High	50.00 (fmol)	46.43 (fmol)	-7.2	2.2	4.4

Table S6. The measured contents of m^3Cm, m^4Cm and m^5Cm in total RNA, small RNA (< 200 nt) and 18S rRNA from different human cell lines.

Cell lines	RNA	m^3Cm/C (%)	m^4Cm/C (%)	m^5Cm/C (%)
HeLa	Total RNA	0.0003±0.00002	0.0004±0.00005	0.0003±0.00003
	Small RNA	0.002±0.00004	-	0.002±0.0003
	18S rRNA	-	0.0009±0.0002	-
HepG2	Total RNA	0.0003±0.00001	0.0004±0.0002	0.0002±0.00002
	Small RNA	0.002±0.0002	-	0.0008±0.00006
	18S rRNA	-	0.001±0.0004	-
HL-7702	Total RNA	0.0004±0.00001	0.001±0.0003	0.0002±0.00001
	Small RNA	0.002±0.0003	-	0.0007±0.00006
	18S rRNA	-	0.003±0.0006	-
MCF-7	Total RNA	0.0008±0.00001	0.0006±0.0001	0.00009±0.00001
	Small RNA	0.007±0.0006	-	0.0006±0.00003
	18S rRNA	-	0.002±0.0004	-

Figure S1. Characterization of the synthesized m^3Cm by NMR.

^1H NMR (400 MHz, D_2O) δ 8.46 (s, 1H), 8.15 (d, *J* = 8.0 Hz, 1H), 6.30 (d, *J* = 8.0 Hz, 1H), 6.00 (d, *J* = 2.7 Hz, 1H), 4.29 (dd, *J* = 7.2, 5.2 Hz, 1H), 4.17 – 4.12 (m, 1H), 4.03 (dd, *J* = 5.1, 2.7 Hz, 1H), 3.90 (ddd, *J* = 16.9, 13.0, 3.3 Hz, 2H), 3.57 (s, 3H), 3.51 (s, 3H).

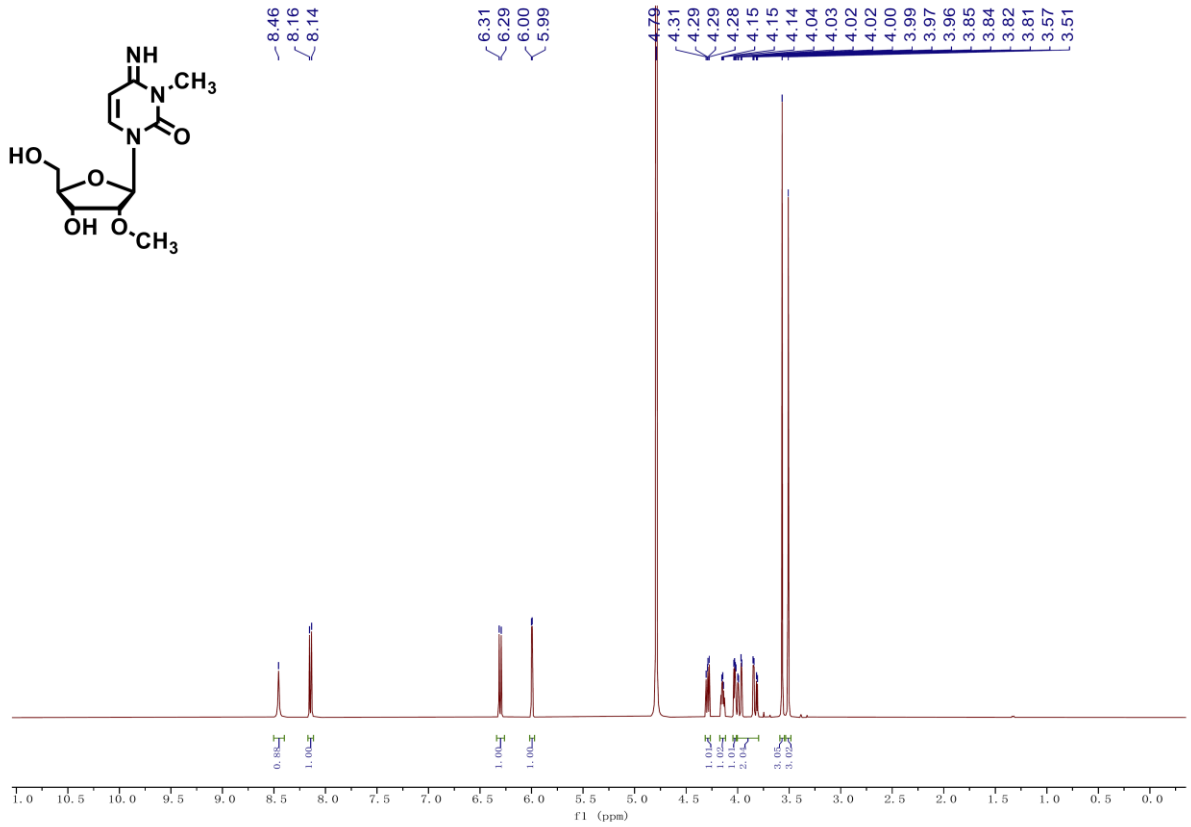

Figure S2. Characterization of Cm by NMR.

^{1}H NMR (400 MHz, D_2O) δ 7.86 (d, *J* = 7.6 Hz, 1H), 6.04 (d, *J* = 7.6 Hz, 1H), 5.97 (d, *J* = 3.5 Hz, 1H), 4.29 (dd, *J* = 6.5, 5.4 Hz, 1H), 4.09 (ddd, *J* = 6.8, 4.3, 2.8 Hz, 1H), 4.00 (dd, *J* = 5.3, 3.6 Hz, 1H), 3.87 (ddd, *J* = 17.2, 12.9, 3.6 Hz, 2H), 3.52 (s, 3H).

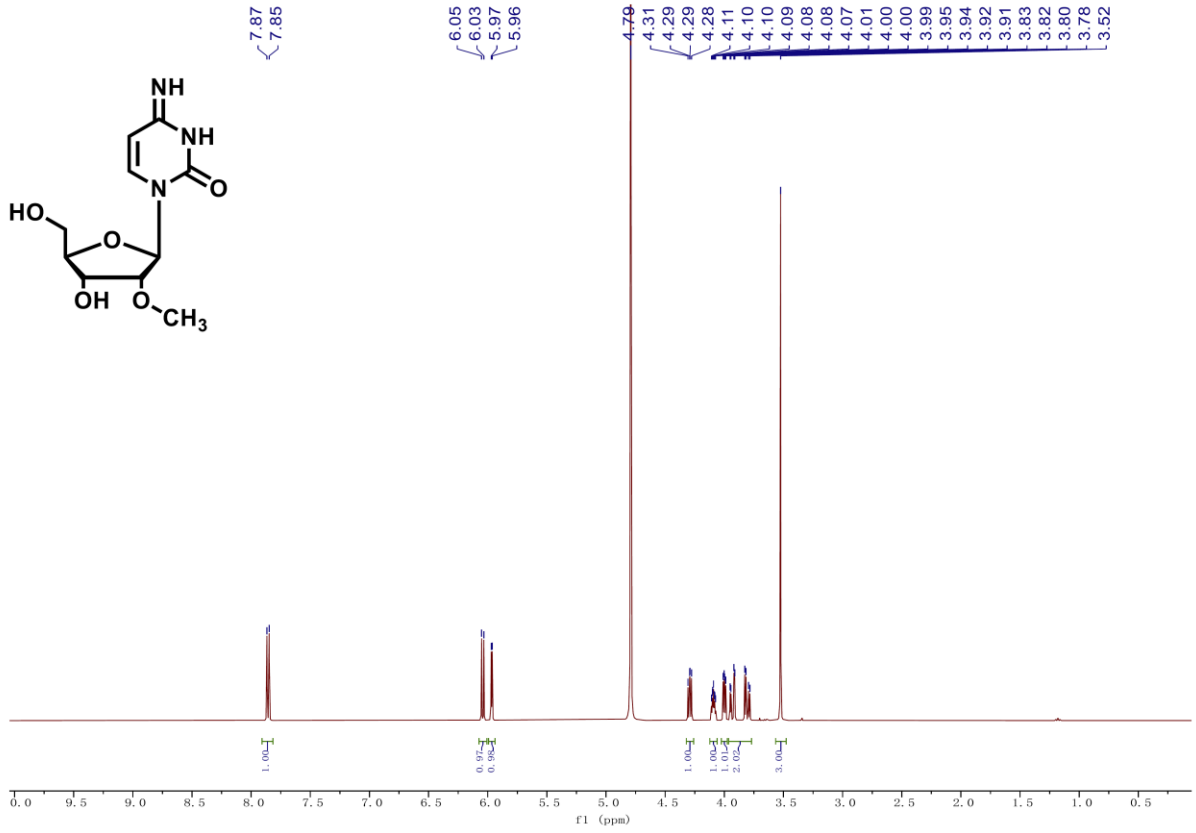

Figure S3. Separation of the isomers of m^3Cm, m^4Cm and m^5Cm. (A-D) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm using 0.05% FA/H_2O (pH 2.8, 3.0, 4.0 and 5.0) as solvent A with gradient 1 at a flow rate of 0.3 mL/min. (E) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm using 2 mM NH_4HCO_3 as solvent A with gradient 1 at a flow rate of 0.3 mL/min. (F) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm using 0.05% FA/H_2O (pH 2.8) as solvent A with gradient 2 at a flow rate of 0.3 mL/min. (G) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm using 0.05% FA/H_2O (pH 2.8) as solvent A with gradient 2 at a flow rate of 0.2 mL/min. Gradient 1: 0–2 min, 3% B; 2–10 min, 3–80% B; 10–12 min, 80% B; 12–13 min, 80–3% B; and 13–20 min, 3% B. Gradient 2: 0–3 min, 3% B; 3–10 min, 3–80% B; 10–12 min, 80% B; 12–13 min, 80–3% B; and 13–20 min, 3% B.

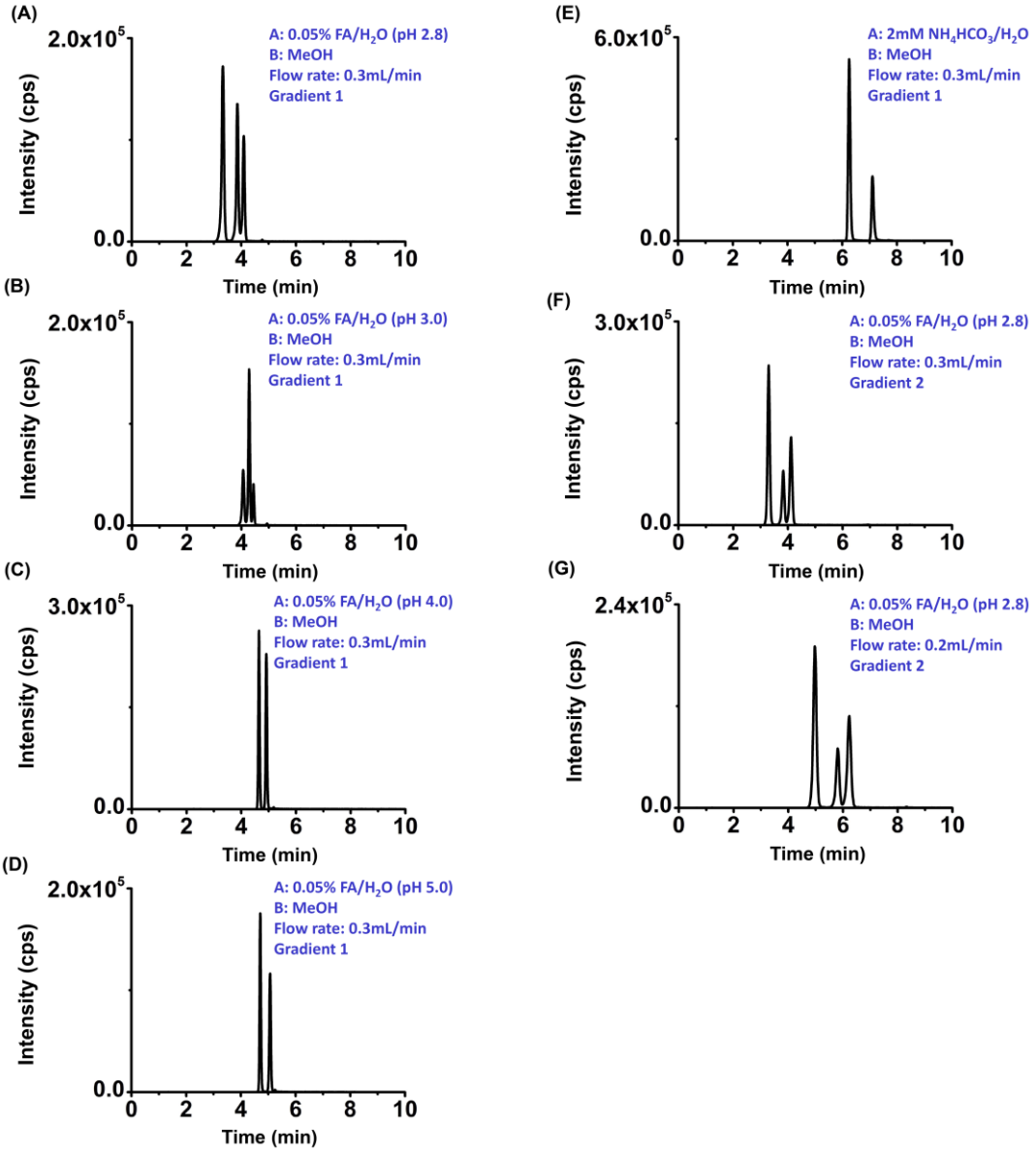

Figure S4. Evaluation of the purity of HeLa 18S rRNA. (A) Calibration curve using different concentrations of *E. coli* 16S rRNA as the template. (B) Evaluation of the purity of the isolated HeLa 18S rRNA by real-time quantitative PCR. (C) Detection of m^4Cm in 16S rRNA of *E. coli*. (D) Quantification of m^4Cm in 16S rRNA of *E. coli*.

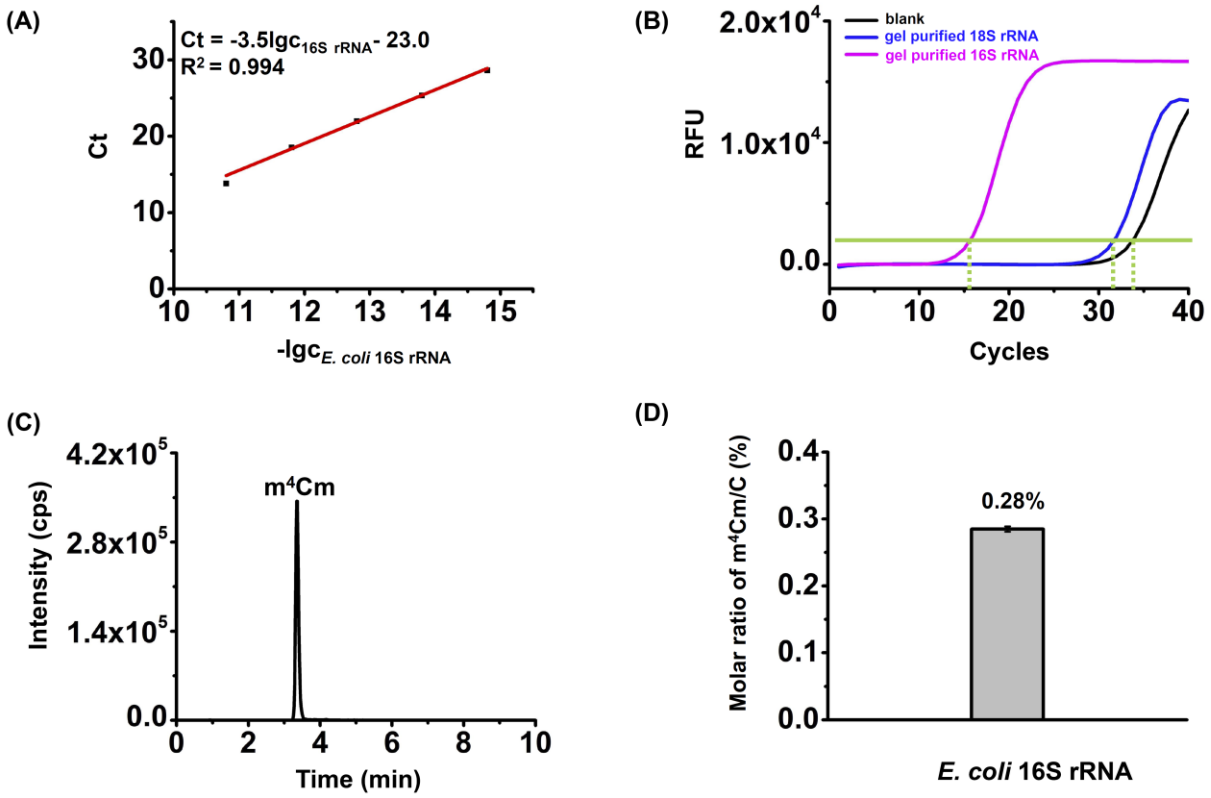

Figure S5. Enrichment of cytidine modifications by offline HPLC-UV. (A) Chromatograms of A, G, C and U standards. (B) Chromatograms of m^3Cm, m^4Cm and m^5Cm standards. (C) Chromatograms of digested nucleosides from small RNA (< 200 nt) of HeLa cells. (D) Chromatograms of digested nucleosides from 18S rRNA of HeLa cells. The digested cytidine modifications from small RNA (< 200 nt) or 18S rRNA of HeLa cells were collected from the indicated red dashed line.

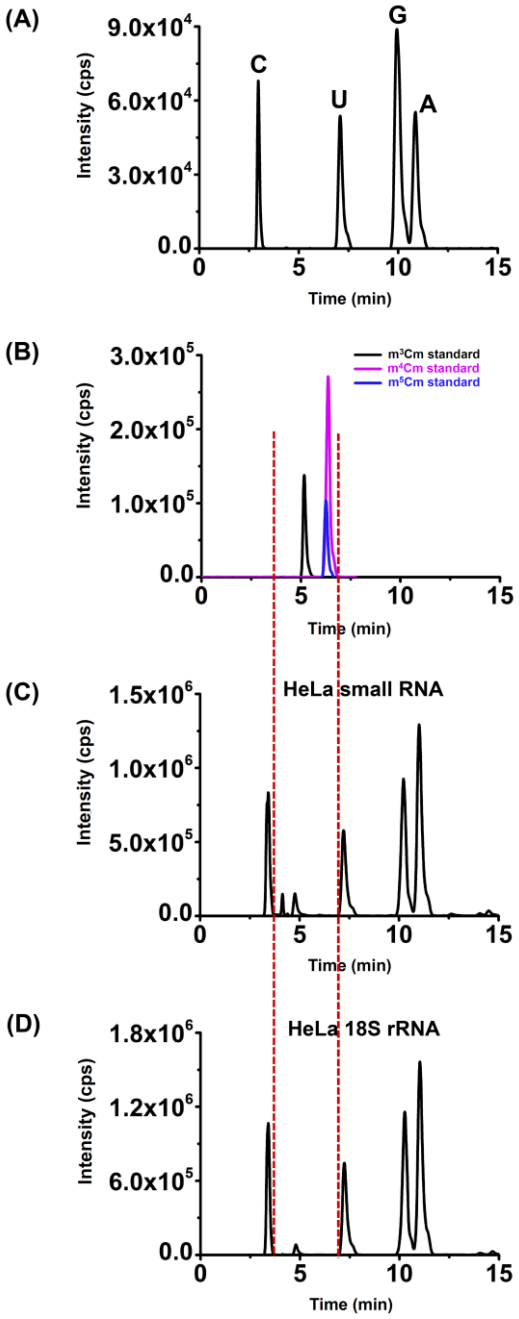

Figure S6. Structures of the deuterated labeled m^3Cm, m^4Cm and m^5Cm by stable isotope tracing monitored by mass spectrometry. HeLa cells were cultured in DMEM medium supplemented with 0.3 mM of D_3-Met to label RNA with the CD_3 group. Theoretically, single and dual CD_3 could be added to m^3Cm, m^4Cm and m^5Cm.