Supporting Information for

2'-*O*-methyl-5-hydroxymethylcytidine – a second oxidative derivative of 5-methylcytidine in RNA

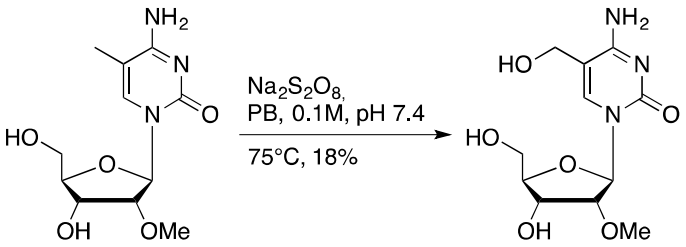
by

Sabrina M. Huber, Pieter van Delft, Arun Tanpure, Eric A. Miska, Shankar Balasubramanian

Materials:

2'-O-methyl-5-methylcytidine was obtained from *Carbosynth Limited*. 1,3-$^{15}N_2$-cytidine was obtained from commercial $^{15}N_2$-urea (*Sigma*) according to procedures previously reported by us.[1] All solvents and reagents were purchased from *Sigma Aldrich* or *Fisher Scientific* and used as received. TLC was performed on ALUGRAM SIL G/UV254 (*Macherey-Nagel*) pre-coated TLC sheets. Flash chromatography was carried out using CombiFlash *Rf* (*Teledyne Isco*) with puriFlash columns (*Interchim*). ^{1}H-NMR and ^{13}C-NMR spectra were recorded on a Bruker DRX-500 instrument and are referenced to the residual solvent peak. Chemical shifts are quoted in parts per million (ppm) using the following abbreviations: s, singlet; d, doublet; t, triplet; m, multiplet; br, broad. The coupling constants (*J*) are measured in Hertz. High resolution mass spectra (HRMS) were recorded on a Vion IMS QTof (*Waters*) mass spectrometer.

Synthesis 2'-O-methyl-5-hydroxymethylcytidine



2'-*O*-methyl-5-methylcytidine (90 mg, 0.33 mmol) was suspended in sodium phosphate buffer (9 ml, 0.1 M, pH 7.4). Sodium persulfate (79 mg, 0.33 mmol) was added and reaction mixture was heated at 75 °C for 4 h.[2] The reaction mixture was concentrated under reduced pressure. The crude product was purified by flash chromatography using a silica gel column (10–22 % methanol in dichloromethane) to afford the product 2'-*O*-methyl-5-hydroxymethylcytidine as white solid (17 mg, 0.06 mmol, 18 %). R_f = 0.4 (CH_2Cl_2/MeOH 8:2);

^{1}H NMR (500 MHz, DMSO-d_6): δ = 7.86 (*s*, 1H, H-C(6)), 7.39 (br, 1H, H(a)-N(4)), 6.60 (br, 1H, H(b)-N(4)), 5.86 (*d*, *J* = 4 Hz, 1H, H-C(1')), 5.10 (*t*, *J* = 5 Hz, 1H, OH-C(7)), 5.05 (*d*, *J* = 6 Hz, 1H, OH-C(3')), 4.98 (*t*, *J* = 5.5 Hz, 1H, OH-C(5')), 4.17 (*d*, *J* = 5 Hz, 2H, H_2C(7)), 4.06-4.05 (*m*, 1H, H-C(2')), 3.83-3.80 (*m*, 1H, H-C(3')), 3.69-3.64 (*m*, 2H, H_2C(5')), 3.58-3.54 (*m*, 1H, H-C(4')), 3.38 (*s*, CH_3-O (2')).

^{13}C NMR (125 MHz, DMSO-d_6): δ = 164.5 (C(4)), 155.0 (C(2)), 139.3 (C(6)), 105.9 (C(5)), 86.9 (C(1')), 84.3 (C(2')), 83.2 (C(3')), 68.1 (C(4')), 60.4 (C(5')), 57.5 (C(7) and C-O(2')).

HRMS: *m/z* calcd. for $C_{11}H_{17}N_3O_6$ [M]+ = 287.1117, found = 287.1106

Sources of model organisms:

Mouse brain tissues were obtained from 62 days old C57BL/6J (JAX mice strain) male mice. *A. thaliana* total was obtained from Prof Baulcombe (University of Cambridge) and *D. melanogaster* was provided by Prof St Johnston (Cambridge University). Mouse embryonic stem (ES) cells and mouse ES cells with the depletion of all three *Tet* genes were described elsewhere.[3] HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (*Life Technologies*) supplemented with 10% fetal bovine serum, penicillin (100 U/ml), and streptomycin (100 g/ml) at 37 °C in 5% CO_2 atmosphere. The cells were isolated for RNA extraction at a confluence level of 75% by trypsinisation followed by centrifugation. The resulting cell pellet was washed twice with PBS before the addition of TRI reagent for RNA isolation. *C. elegans* (wild-type strain var. Bristol N2)[4] were grown under standard conditions at 20 °C.

Isotopic labelling of HEK293T cells:

HEK293T cells were cultured in methionine- and cystine-free Dulbecco's Modified Eagle Medium (DMEM) (*Life Technologies*) supplemented with 10% dialysed fetal bovine serum, penicillin (100 U/ml), streptomycin (100 g/ml), *L*-$^{13}CD_3$-methionine (30 mg/L), and L-cysteine-HCl at 37 °C in 5% CO_2 atmosphere. After 5 days the heavy methionine medium was removed and the cells were washed twice with ice-cold PBS. DMEM (*Life Technologies*) supplemented with 10% fetal bovine serum, penicillin (100 U/ml), streptomycin (100 g/ml) and $^{15}N_2$-cytidine (100 M final concentration) was subsequently added and cells were subsequently harvested by trypsination and centrifugation in hourly intervals over a period of 15 h.

Total RNA isolation:

0.2 ml of chloroform was added per ml of TRIsure used. The sample was shaken vigorously for 15 seconds, allowed to stand for 5 minutes at room temperature and centrifuged at 12,000 x g for 15 minutes at 4 °C. The aqueous phase was transferred to a fresh tube and 0.5 ml of 2-propanol was added per ml of TRIsure used in the sample preparation. The sample was allowed to stand 10 minutes at room temperature and centrifuged at 12,000 x g for 10 minutes at 4 °C to precipitate the RNA at the bottom and side of the tube. The supernatant was removed and the RNA pellet was washed with 75 % ethanol, air-dried and redissovled in nuclease-free water. Total RNA was then purified using RNA Clean & Concentrator (*Zymo Research*).

Small/large RNA fractionation:

Total RNA was fractionated into small (< 200 nt) and large (> 200 nt) RNAs either using the Quick-RNA MiniPrep kit (*Zymo Research*) according to the manufacturer's instructions or gel electrophoresis. In the ladder, total RNA and a low range ssRNA ladder (*NEB*) were run on a 15 % Novex TBE-urea gel (*Life Technologies*) for 60 min at 180 V. The ladder was excised from the gel, stained with CYBRGold and visualised under UV. Gel sections of the corresponding RNA bands of interest were excised from an unstained gel and RNA was eluted by overnight agitation at 4 °C in gel elution buffer (0.3 M NaCl, 0.25 % SDS, 1 mM EDTA (pH 8)). RNA was precipitated using isopropanol, washed with 75 % EtOH, air-dried and redissolved in water.

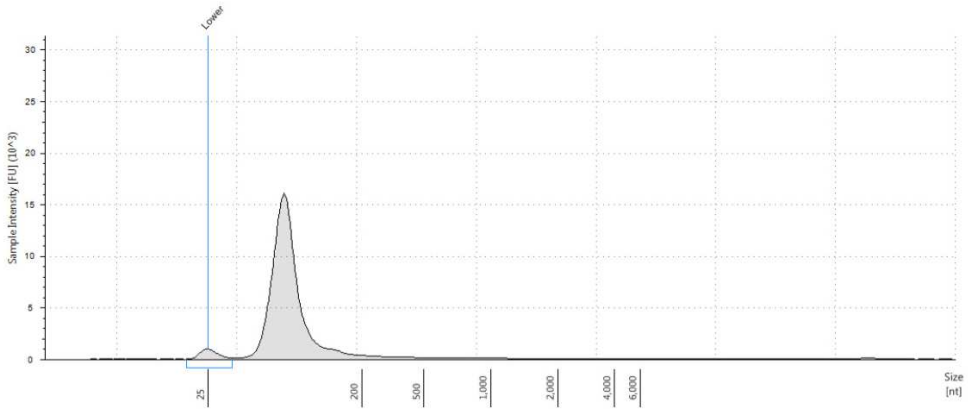


Figure S1: Representative Agilent TapeStation electropherogram of small (< 200 nt) RNA fraction.

RNA digestion:

Digestion enzyme master mix was prepared by combining benzonase (250 U/µl, 0.625 µl, *Sigma Aldrich*), phosphodiesterase I from Crotalus adamanteus venom (10 mU/µl, 10 µl, *Sigma Aldrich*) and Antarctic phosphatase (5 U/µl, 20 µl, *NEB*). Aqueous solutions of total RNA (1 µg in 13.25 µl final volume) were mixed with 5x digestion buffer (5 µl, Tris-HCl pH 8 (20 mM), $MgCl_2$ (20 mM), NaCl (100 mM)) and digestion enzyme stock solution (0.5 µl) and water (6.25 µL), followed by incubation at 37 °C for 14 h. The nucleoside mixture was subsequently cleaned-up by filtration over Amicon Ultra 0.5 ml (10 kDa MWCO, Merck-Milipore) spin columns.

LC-MS/MS analysis:

Quantitative LC-MS/MS analysis was carried out using an Ultimate3000 UPLC system (*Thermo Scientific*) coupled to a QExactive quadrupole orbitrap hybrid mass spectrometer (*Thermo Scientific*). LC was performed using a Waters Acquity UPLC HSS T3 column (100 x 2.1 mm, 1.8 µm particle size) kept at 50 °C, applying a gradient starting at 100% of 0.1% formic acid in water followed by increasing proportions of 0.1% formic acid in acetonitrile up to 15%, at a flow rate of 300 µl/min over 3 min. An additional 2 min were used to wash and re-equilibrate the column under the starting conditions. The MS was operated using positive electrospray ionisation in multiple reaction monitoring (MRM) mode to measure the analytes listed in Table 1 and the available internal standards, [2-^{13}C, 1,3-$^{15}N_2$]-cytidine, 5-[methyl-D_3]-[6-D]-cytidine and 5-hydroxymethyl-[2-^{13}C, 1,3-$^{15}N_2$]-cytidine.[1] The transitions and product ions used for these measurements are listed in Table 1. Calibration lines were prepared for all analytes in the range of 0.05 - 50,000 nM using 9 calibration points. Data was processed using AB Sciex Multiquant (Ver. 2.1.1) or Thermo Xcalibur Quanbrowser (Ver. 2.2.44) selecting for either external calibration (hm^{5}Cm and m^{5}Cm) or internal calibration for those analytes with available SIL standards (C, m^{5}C and hm^{5}C). Sample concentrations for each analyte were then back calculated from their respective calibration curves.

analyte	parent ion $[M+H]^+$	fragment ion $[M+H]^+$
C	244	112.05054
m^{5}C	258	126.06619
hm^{5}C	274	142.06110
f^{5}C	272	140.04545
hm^{5}Cm	288	142.06110
m^{5}Cm	272	126.06619
$^{13}C^{15}N_2$-C	247	115.04796
D_4-m^{5}C	262	130.09130
$^{13}C^{15}N_2$-hm^{5}C	277	145.05853
$^{13}CD_3$-m^{5}C	262	130.08837
$^{13}CD_2$-hm^{5}C	277	145.07701
$^{15}N_2$-m^{5}C	260	128.06026
$^{15}N_2$-hm^{5}C	276	144.05517
$^{15}N_2$-hm^{5}Cm	290	144.05517
$^{13}C_2D_5$-hm^{5}Cm	295	145.07701

Table S1: Analytes and their transitions used for LC-MS/MS analysis.

rC [%]		m^5C [%]		hm^5C [%]		hm^5Cm [%]	
BR1	*BR2*	*BR1*	*BR2*	*BR1*	*BR2*	*BR1*	*BR2*
45	45	51	49	29	30	53	48

Table S2: The percentage levels of $^{13}C_xD_y$ labeled rC, m^5C, hm^5C and hm^5Cm in RNA from HEK293T cells 15 hours after removal of L-$^{13}CD_3$-methionine and addition of unlabelled L-methionine and 1,3-$^{15}N_2$-cytidine.

Origin	m^5C [ppm]			hm^5C [ppm]		
	BR1	*BR2*	*BR3*	*BR1*	*BR2*	*BR3*
D. melanogaster	8930	8910	8890	BLD	BLD	BLD
A. thaliana	8490	8730	8550	131	135	130
C. elegans	3370	3250	2290	BLQ	BLQ	BLQ
HEK293T	5190	5560	5300	9.50	8.50	11.60
mouse brain	11,700	11,580		53.10	47.60	

Origin	m^5Cm [ppm]			hm^5Cm [ppm]		
	BR1	*BR2*	*BR3*	*BR1*	*BR2*	*BR3*
D. melanogaster	87.30	83.60	76.00	120.30	131.50	132.40
A. thaliana	BLD	BLD	BLD	0.50	0.60	0.70
C. elegans	0.60	0.50	0.20	33.80	38.60	17.30
HEK293T	1.00	1.20	1.00	9.60	14.30	12.00
mouse brain	3.70	3.00	na	70.70	69.50	na

Table S3: The levels (ppm of total C) of m^5C, hm^5C, m^5C and hm^5Cm in total RNA samples isolated from different eukaryotic model organisms and HEK293T cells. 3 biological replicates (BR) were measured. BLD = below limit of detection, BLQ = below limit of quantification, na = not available.

	WT mouse ES cells			TET 3 KO mouse ES cells		
	BR1	*BR2*	*BR3*	*BR1*	*BR2*	*BR3*
hm^5C	26.8	25.1	28.30	12.5	9.60	150
hm^5Cm	103.3	77.40	106.2	100.00	85.80	108.6

Table S4: The levels (ppm of total C) of hm^5C and hm^5Cm in total RNA samples isolated from TET wildtype or TET triple knockout embryonic stem cells. 3 biological replicates (BR) were measured.

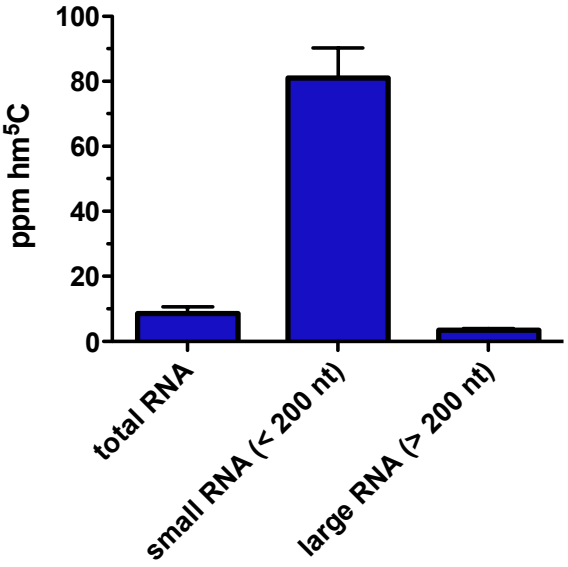


Figure S1: The levels (ppm of total C) of hm^5C and hm^5Cm in total RNA samples isolated from different HEK293T RNA fractions. 3 biological replicates (BR) were measured.

References:

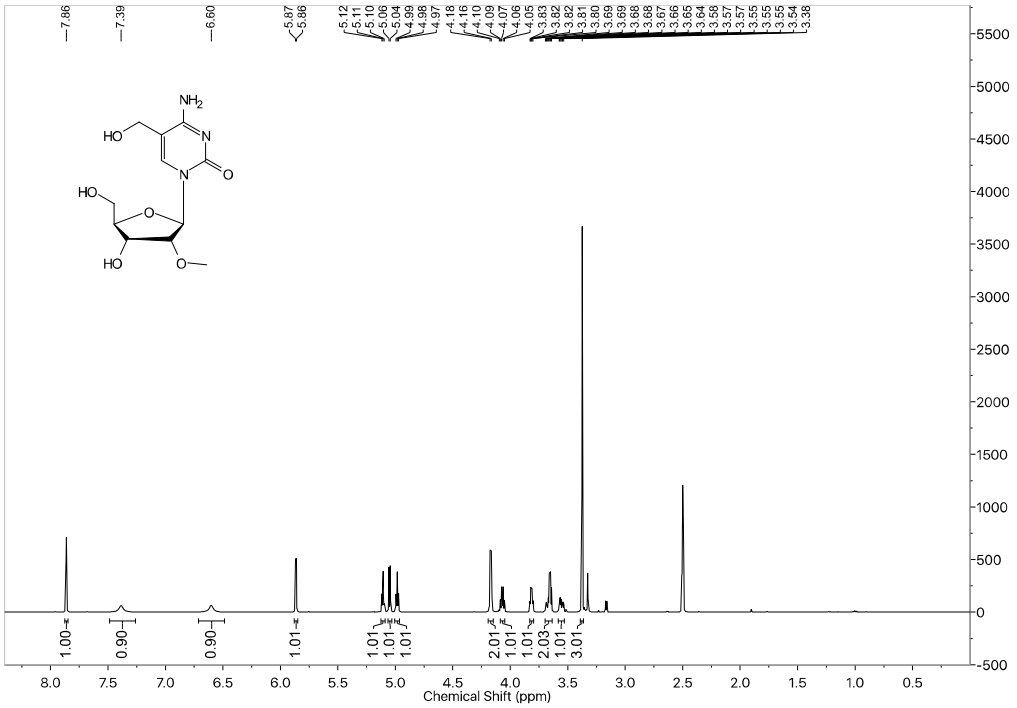
[1] Huber, S. M.; van Delft, P.; Mendil, L.; Bachman, M.; Smollett, K.; Werner, F.; Miska, E. A.; Balasubramanian, S. *ChemBioChem* **2015**, *16*, 752-755.

[2] Ithara, H. *Chem. Lett.* **1991**, *20*, 1591-1594.

[3] Hu, X.; Zhang, L.; Mao, S.-Q.; Li, Z.; Chen, J.; Zhang, R.-R., Wu, H.-P.; Gao, J.; Guo, F.; Liu, W.; Xu, G.-F.; Dai, H.-Q.; Shi, Y.-G.; Li, X.; Hu, B.; Tang, F.; Pei, D.; Xu, G.-L. *Cell Stem Cell* **2014**, *14*, 512-522

[4] Brenner, S. *Genetics* **1974**, *77*, 71-94.

[1]H-NMR 2'-O-methyl-5-hydroxymethylcytidine:



[13]C-NMR 2'-O-methyl-5-hydroxymethylcytidine: