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

Sabrina M. Huber[1,‡], Pieter van Delft[1,‡], Arun Tanpure[1], Eric A. Miska[2], Shankar Balasubramanian[1]*

1 Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

2 Wellcome Trust/Cancer Research UK Gurdon Institute, Tennis Court Road, Cambridge, CB2 1QN, UK

Supporting Information Placeholder

ABSTRACT: 5-hydroxymethylcytidine (hm^{5}C) was recently identified as a direct metabolite of m^{5}C in RNA. We investigated the stability of hm^{5}C in human cells using bio-isotopomers and high-resolution tandem LC-MS/MS. This has led to the discovery of a second oxidative metabolite of m^{5}C in RNA, namely 2'-*O*-methyl-5-hydroxymethylcytidine (hm^{5}Cm). Subsequent quantitative analysis of total RNA from higher organisms revealed varying levels and TET-independent formation of this new RNA modification.

There is a wide chemical diversity of ribonucleoside modifications in RNA.[1,2] While epigenetic events such as DNA methylation and histone modifications are understood to be dynamic and reversible processes, RNA modifications, have long been considered relatively static and stable marks. However, it was recently shown that the enzyme FTO mediates the oxidative demethylation of m^{6}A *via* N6-hydroxy- and N6-formylcytidine in messenger RNA.[3,4] This first example of reversible RNA methylation has opened up the possibility that RNA modifications may also be dynamic, with potential regulatory roles analogous to reversible epigenetic modifications. In support of this, we recently showed that m^{5}C undergoes similar oxidative metabolism in RNA to produce hm^{5}C and that the latter modification is conserved across Archaea, Bacteria and Eukarya.[5] Furthermore, Fu *et al.* reported the ability of TET enzymes to oxidize m^{5}C to hm^{5}C in synthetic RNA strands *in vitro* and showed the dependency of hm^{5}C on the TET3 enzyme in an *in vivo* knock-out mouse model.[6] Together, these studies established hm^{5}C as a new RNA modification that is introduced through active, enzyme-catalyzed oxidation, rather than passive, reactive oxygen species-mediated oxidation of m^{5}C.

In contrast with m^{6}A, which is predominantly an mRNA modification, we had determined by quantitative LC-MS/MS that hm^{5}C is enriched in tRNA fractions (Figure S1). The turnover of m^{5}C into hm^{5}C is of particular interest as the extent of m^{5}C modification at specific tRNA sites plays a key role in regulating the cellular stress response. For example, the absence of m^{5}C triggers increased stress-induced cleavage of tRNAs and sensitizes organisms to oxidative stress.[7,8] Furthermore, tRNA wobble modifications can change as a result of exposure to toxic agents and thereby trigger stress-specific enhancement of translation of proteins critical to the cell stress response.[9,10] These response mechanisms require a dynamic control of tRNA modifications which can either be achieved through their reversible introduction or specific tRNA turnover/degradation.

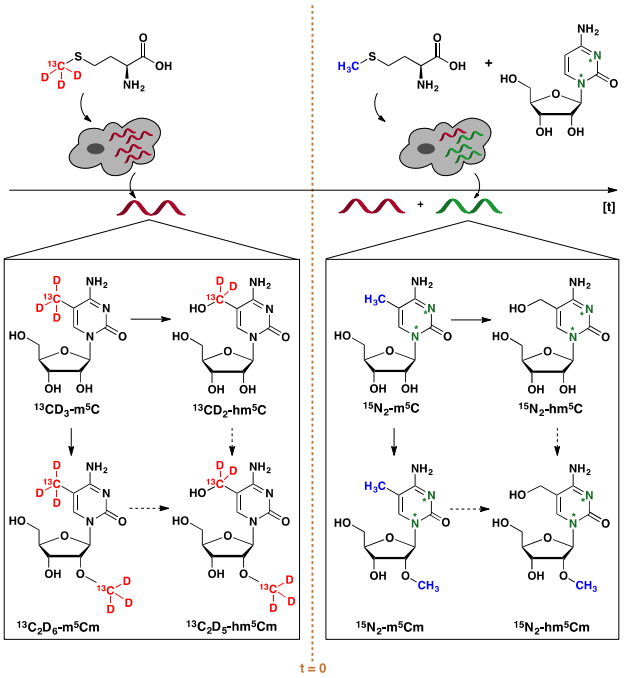


Figure 1. Overview of the stable isotope, dual-labeling approach. Cells were grown in the presence of [methyl-$^{13}CD_3$]-*L*-methionine until near quantitative isotope incorporation into m^{5}C was observed by LC-MS/HRMS. Heavy methionine (red) was then removed and a mixture of unlabeled *L*-methionine (blue) and 1,3-^{15}N-cytidine (green) was added.

We investigated whether hm^5C is indeed subject to dynamic turnover and looked for the existence of additional, novel oxidative metabolites of m^5C to examine the presence of an active cytidine-C5 demethylation pathway in RNA.

As a means to study the relative stabilities of m^5C and hm^5C in tRNA-enriched fractions, as compared to tRNA turnover, we selected stable isotope tracing monitored by mass spectrometry (Figure 1). We adapted methods previously reported by us in which we studied both RNA and DNA methylation and their oxidative pathways.[5,11,12] Briefly, human HEK293T cells were cultured in the presence of stable isotope labeled (SIL) methionine, $^{13}CD_3$-*L*-methionine, to metabolically $^{13}CD_3$ label the methyl group of m^5C in RNA (Figure 1). The medium was then replaced with medium containing unlabeled methionine and SIL labeled 1,3-$^{15}N_2$-cytidine (Figure 1, t = 0) and cells were collected at hourly intervals and subjected to total RNA isolation, over the course of 15 hours, approximately a complete cell cycle of a HEK293T cell. The total RNA fractions were each subsequently enriched for tRNAs by fractional precipitation, enzymatically digested into nucleosides and subjected to mass spectrometric analysis to quantify the SIL forms of both m^5C ($^{13}CD_3$-m^5C, $^{15}N_2$-m^5C) and hm^5C ($^{13}CD_2$-hm^5C, $^{15}N_2$-hm^5C). 1,3-$^{15}N_2$-cytidine was included to ensure the differential labeling of RNA synthesized during (Figure 1, before t = 0) and after (Figure 1, from t = 0 onwards) the $^{13}CD_3$-*L*-methionine labeling. This distinguishes the apparent $^{13}CD_n$ decay as a result of isotope dilution due to cell proliferation or tRNA turnover, rather than modification turnover.

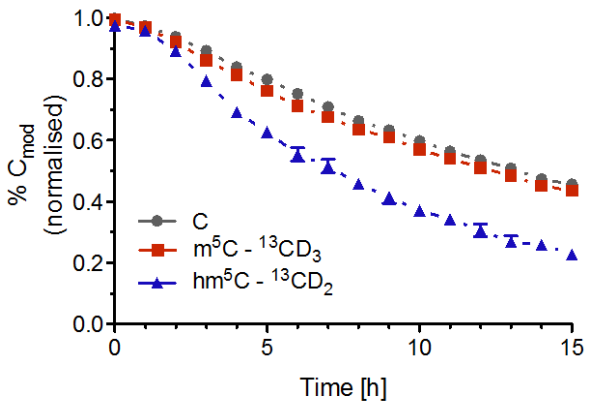


Figure 2. Amounts of the $^{13}CD_n$-labeled fractions of a modification relative to the sum of the total amounts of the same modification in small RNAs from $^{13}CD_n$-labeled HEK293T cells grown in the presence of 1,3-$^{15}N_2$-cytidine and absence of $^{13}CD_3$-*L*-methionine as a function of time (i.e. for m^5C: % $^{13}CD_3$-m^5C = {[$^{13}CD_3$-m^5C] / ([$^{13}CD_3$-m^5C] + [$^{15}N_2$-m^5C] + [m^5C])} * 100).

Thus, we measured the abundances of the different isotopologues of C, m^5C and hm^5C in the tRNA-enriched digests and calculated the amount of their $^{13}CD_n$ labeled (m^5C and hm^5C) and unlabeled (C) fractions relative to the sum of the total amounts of any given modification (Figure 2).

We observed a 50% decrease of unlabeled cytidines (Figure 2, grey trace), consistent with the cell population and total RNA doubling in this time, which dilutes the $^{15}N_2$ label by 50% due to the addition of 1,3-$^{15}N_2$-cytidine at t = 0. We observed 5-methylcytidine (Figure 2, red trace) had comparable stability to cytidine, indicating the bulk of m^5C residues was not subject to active turnover. Since only a small fraction of m^5C residues is converted to hm^5C (~ 0.1 %), this turnover was not sufficient to detect by our approach. However, when we considered the turnover of hm^5C (Figure 2, blue trace), we observed a strikingly steeper initial slope for its decay and a much lower relative abundance of the $^{13}CD_2$ isotopologue after 15 h as compared to m^5C. Consequently, hm^5C-containing RNA transcripts could either be unstable and subject to accelerated degradation, *or* hm^5C itself could be actively metabolized within its RNA transcript. To explore the latter hypothesis, we screened the dual SIL RNA samples for other oxidative derivatives of m^5C. Recently, it was shown that the Fe(II)-dependent oxygenase ALKBH1/ABH1 oxidizes m^5C at position 34 in human mitochondrial $tRNA^{Met}$ to f^5C and hm^5C was not observed as an intermediate in this study.[13] We therefore considered the hitherto unknown 2'-OH methylated derivative of hm^5C, 2'-*O*-methyl-5-hydroxymethylcytidine (hm^5Cm), as a potential downstream product of hm^5C metabolism in subsequent LC-MS/HRMS analyses. 2'-*O*-methylation has been observed in several RNA classes[14] and close examination of previous extracted ion counts and fragmentation patterns of hm^5C led us to hypothesize the presence of hm^5Cm in RNA. Thus, using the tRNA-enriched digests from the eight hour time point, we targeted the mass spectrometry for 2'-*O*-[$^{13}CD_3$-methyl]-5-[$^{13}CD_2$-hydroxymethyl]-cytosine, 2'-*O*-methyl-1,3-[$^{15}N_2$]-5-hydroxymethylcytosine and the minor, completely unlabeled hm^5Cm isotopologues. As depicted in Figure 3 (left), we could extract all the corresponding product ions. This, together with the observed co-elution of all the isotopologues during liquid chromatography, provided the first evidence for the presence of hm^5Cm in RNA. The slightly earlier elution of deuteriated compounds, as observed for $^{13}C_2D_5$ labeled hm^5Cm (Figure 3, left, red trace), is commonly observed in liquid chromatography of deuterium labeled compounds due to their different polarity, polarizability and molecular volume compared to their lighter isotopologues.[15]

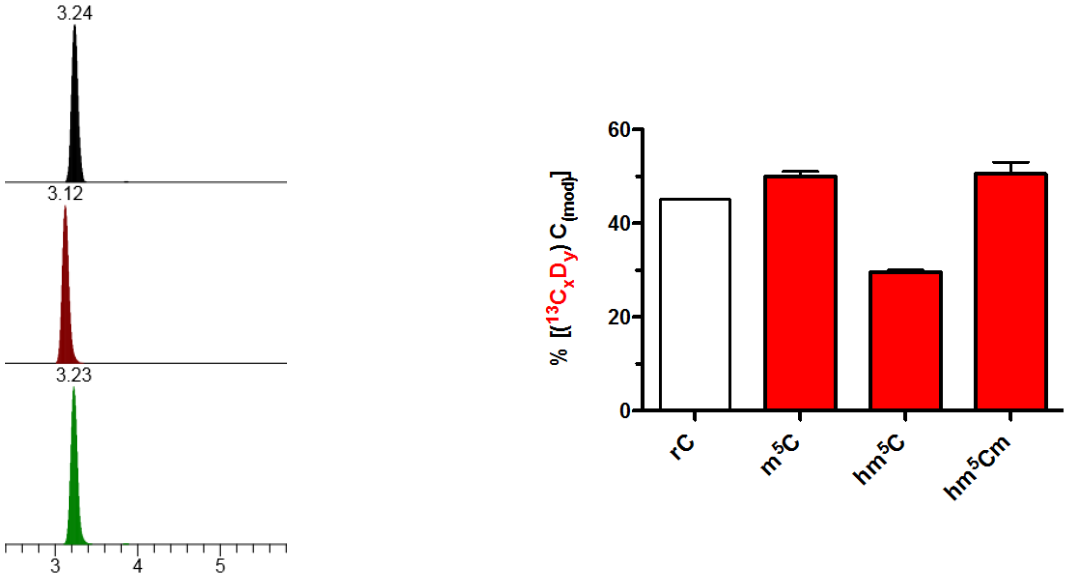


Figure 3. Left) Differential labeling of hm^{5}Cm. LC-MS/MS analysis of RNA obtained from HEK293T cells grown in regular (top), [methyl-$^{13}CD_3$]-*L*-methionine- (middle) or 1,3-$^{15}N_2$-cytidine-supplemented (bottom) medium. Extracted ion counts are shown for hm^{5}C, $^{13}CD_2$-hm^{5}C and $^{15}N_2$-hm^{5}C. Right) Analysis of the levels of $^{13}C_xD_y$ labeled modifications after 15 hours.

To assess the stability of this novel RNA modification in relation to C, m^{5}C and hm^{5}C we determined the levels of $^{13}C_xD_y$-labeled modifications after 15 h, from our previous time decay study. As shown in Figure 3 (right), around 50% of all hm^{5}Cm are still $^{13}C_2D_5$-labeled after a complete HEK293T cell cycle. This is comparable to that observed for rC (45%) and m^{5}C (50%), two residues that we identified as highly stable. In contrast, $^{13}CD_2$-hm^{5}C accounts for only 29% of all hm^{5}C residues after 15 hours. These data demonstrate that hm^{5}Cm is a stable modification. To unequivocally establish hm^{5}Cm as a novel RNA modification, we synthesized a reference standard for hm^{5}Cm by sodium persulfate-mediated oxidation of commercially obtained m^{5}Cm[16] and performed quantitative LC-MS/HRMS analysis of total RNA samples form a variety of organisms. Thereby, we measured the abundance of m^{5}C, hm^{5}C, m^{5}Cm, and hm^{5}Cm (Figure 4). We selected HEK293T cells and murine brain tissue as human and mammalian examples, respectively. Furthermore, we chose models with previously reported low (*C. elegans*), high (*A. thaliana*) and undetermined (*D. melanogaster*) absolute levels of hm^{5}C.[5,17] As shown in Figure 5, we measured hm^{5}C levels that agreed with those previously described by us and others.[5,6] *A. thaliana* RNA exhibited the highest hm^{5}C levels (130 ppm), while *C. elegans* RNA showed the lowest (< 10 ppm). While the presence of hm^{5}C in *D. melanogaster* total RNA was previously demonstrated by dot blot experiments, we could not verify these results by LC-MS/MS.[17] On the other hand, we could readily observe the 2'-OH methylated form of hm^{5}C, hm^{5}Cm, in the latter organism. This may indicate that currently used antibodies cannot discriminate between hm^{5}C and hm^{5}Cm.[17] In general, organisms exhibiting a very low or undetectable level of hm^{5}C, actually showed a relatively high, detectable level of the 2'-OH methylated form, hm^{5}Cm. In addition to *D. melanogaster*, this is exemplified by *C. elegans*, which contains 30 ppm hm^{5}Cm in total RNA. For the human cells and mouse brain the abundances of methylated and unmethylated hm^{5}C are comparable. These results suggest that oxidation of C5-methylated cytidines (m^{5}C and/or m^{5}Cm) is a widespread process and eukaryotes seem to largely select for either 2'-OH methylated or unmetylated derivative with only mammalian RNA containing both forms of C5-hydroxymethylation (hm^{5}C and hm^{5}Cm).

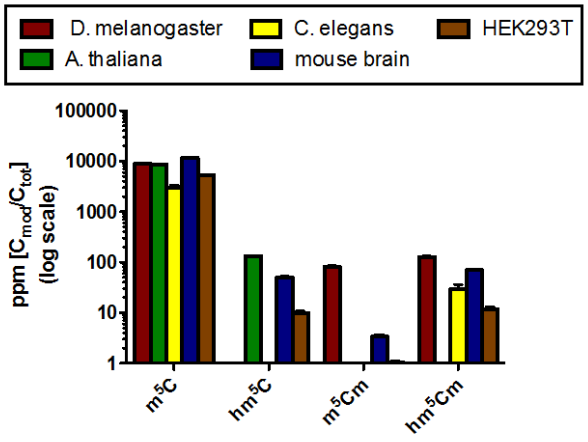


Figure 4. The abundance of m^{5}C and its derivatives as determined by quantitative LC-MS/HRMS.

Whilst further studies are required to fully discern the function of hm^{5}Cm, it should be noted that methylation of 2'-hydroxyl groups in tRNA molecules has been previously observed to occur at the first position of the anticodon to promote codon-anticodon interaction.[18] Furthermore 2'-O-methylation can block the ability of the 2'-position of the nucleoside to serve as a proton donor and therefore prevents RNA hydrolysis, increasing the lifetime of the RNA.[19] We therefore propose that hm^{5}Cm may promote the stability of tRNAs themselves and the stability of duplex formation with complementary RNA molecules. It is noteworthy that the oxidative derivative of hm^{5}Cm, 2'-O-methyl-5-formylcytidine is already known to be present at the wobble position of cytoplasmic tRNAs.[20]

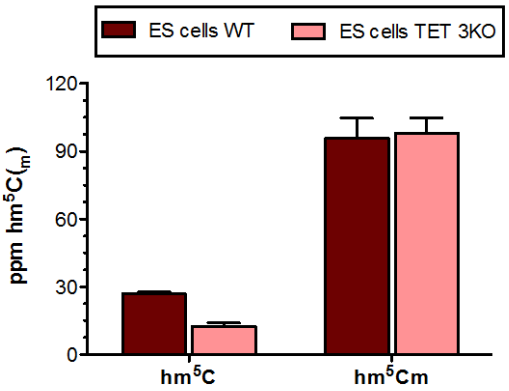


Figure 5. The abundance of hm^{5}C and hm^{5}Cm as determined by quantitative LC-MS/HRMS in mouse embryonic WT and TET triple KO stem cells.

The TET family of enzymes was previously reported to be capable of oxidizing m^{5}C to hm^{5}C in RNA both *in vitro* and *in vivo*.[6,17] To shed light on whether hm^{5}Cm is also TET-dependent, we measured its levels in TET triple knockout (TKO) mouse embryonic stem cells that have been mutated in the catalytic domain of all three TET enzymes and therefore have no residual TET activity. (Figure 5).[21] Interestingly, as we show here, hm^{5}Cm *is not* TET-dependent. RNA obtained from TET wild type and TET TKO cells show equal amounts of the 2'-*O*-methylated version of hm^{5}C. This shows that hm^{5}Cm is generated by an enzyme other than TET, which is in accordance with the findings that hm^{5}Cm is highest in organisms that do not express TET (*C. elegans*) or express TETs at a reduced level (*D. melanogaster*).

In conclusion, we have identified a novel derivative of C5-methylated ribonucleosides in RNA from mammalian cells, tissue and several organisms. The exact functional roles of hm^{5}C and hm^{5}Cm and the relationship between them will be the subject of future studies.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website and contains detailed experimental procedures, supporting figures and tables.

AUTHOR INFORMATION

Corresponding Author

* sb10031@cam.ac.uk

Author Contributions

‡ These authors contributed equally.

ACKNOWLEDGMENT

This work was supported by a Wellcome Trust Senior Investigator award to SB (grant no. 099232/z/12/z). The Balasubramanian group is supported by core funding from Cancer Research UK (C14303/A17197). Prof. Daniel St. Johnston is thanked for the provision of *D. melanogaster* for total RNA isolation.

REFERENCES

(1) Cantara, W. A.; Crain, P. F.; Rozenski, J.; Mc Closkey, J. A.; Harris, K. A.; Zhang, X.; Vendeix, F. A.; Fabris, D.; Agris P. F. *Nucleic Acids Res.* **2011**, *39*, D195-D201.

(2) Machnicka, M. A.; Milanowska, K.; Osman Oglou, O.; Purta, E.; Kurkowska, M.; Olchowik, A.; Januszewski, W.; Kalinowski, S.; Dunin-Horkawicz, S.; Rother, K. M.; Helm, M.; Bujnicki, J. M.; Grosjean H. *Nucleic Acids Res.* **2013**, *41*, D262-D267.

(3) Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y. G.; He, C. *Nat. Chem. Biol.* **2011**, *7*, 885-887.

(4) Fu, Y.; Jia, G.; Pang, X.; Wang, R. N.; Wang, X.; Li, C. J.; Smemo, S.; Dai, Q.; Bailey, K. A.; Nobrega, M. A.; Han, K. L.; Cui, Q.; He, C. *Nat. Commun.* **2013**, *4*, 1798.

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

(6) Fu, L.; Guerrero, C. R.; Zhong, N.; Amato, N. J.; Liu, Y.; Liu, S.; Cai, Q.; Ji, D.; Jin, S. G.; Niedernhofer, L. J.; Pfeifer, G. P., Xu, G. L.; Wang, Y. *J. Am. Chem. Soc.* **2014**, *136*, 11582-11585.

(7) Schaefer, M.; Pollex, T.; Hanna, K.; Tuorto, F.; Meusburger, M.; Helm, M.; Lyko, F. *Genes Dev.* **2010**, *24*, 1590-1595.

(8) Blanco, S.; Dietmann, S.; Flores, J. V.; Hussain, S.; Kutter, C.; Humphreys, P.; Lukk, M.; Lombard, P.; Treps, L.; Popis, M.; Kellner, S.; Hölter, S. M.; Garrett, L.; Wurst, W.; Becker, L.; Klopstock, T.; Fuchs, H.; Gailus-Durner, V.; Hrabë de Angelis, M.; Káradóttir, R. T.; Helm, M.; Ule, J.; Gleeson, J. G.; Odom, D. T.; Frye, M. *EMBO J.*, **2014**, *33*, 2020-2039.

(9) Chan, C. T.; Dyavaiah, M.; De Mott, M. S.; Taghizadeh, K.; Dedon, P. C.; Begley, T. J. *PLoS Genet.* **2010**, *6*, e1001247.

(10) Chan, C. T., Pang, Y. L.; Deng, W.; Babu, I. R.; Dyavaiah, M.; Begley, T. J., Dedon, P. C. *Nat. Commun.* **2012**, *3*, 937.

(11) Bachman, M.; Uribe-Lewis, S.; Yang, X.; Williams, M.; Murrell, A.; Balasubramanian, S. *Nat. Chem.* **2014**, *6*, 1049-1055.

(12) Bachman, M.; Uribe-Lewis, S.; Yang, X.; Burgess, H. E.; Iurlaro, M.; Reik, W.; Murrell, A.; Balasubramanian, S. *Nat. Chem. Biol.* **2015**, *11*, 555-557.

(13) Haag, S.; Sloan, K. E.; Ranjan, N.; Warda, A. S.; Kretschmer, J.; Blessing, C.; Hübner, B.; Seikowski, J.; Dennerlein, S.; Rehling, P.; Rodnina, M. V., Höbartner, C.; Bohnsack, M. T. *EMBO J.* **2016**, *35*, 2104-2119.
(14) Kiss, T. *Cell* **2002**, *109*, 145-148.
(15) Wade, D. *Chem. Biol. Interact.* **1999**, *117*, 191-217.
(16) Itahara T. *Chem. Lett.* **1991**, *20*, 1591-1594.
(17) Delatte, B.; Wang, F.; Ngoc, L. V.; Collignon, E.; Bonvin, E.; Deplus, R.; Calonne, E.; Hassabi, B.; Putmans, P.; Awe, S.; Wetzel, C.; Kreher, J.; Soin, R.; Creppe, C.; Limbach, P. A.; Gueydan, C.; Kruys, V.; Brehm, A.; Minakhina, S.; Defrance, M.; Steward, R.; Fuks, F. *Science* **2016**, *351*, 282-285.
(18) Satoh, A.; Takai, K.; Ouchi, R.; Yokomaya, S.; Takaku, H. *RNA* **2000**, *6*, 680-686.
(19) Motorin, Y.; Helm, M. *Biochemistry* **2010**, *49*, 4934-4944.
(20) Païs de Barros, J. P; Keith, G.; El Adlouni, C.; Glasser, A. L.; Mack, G; Dirheimer, G; Desgrès, J. *Nucleic Acids Res.* **1996**, *24*, 1489-1496.
(21) 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.

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: 7

Synthesis 2'-O-methyl-5-hydroxymethylcytidine 7

Sources of model organisms: 9

Isotopic labelling of HEK293T cells: 9

Total RNA isolation: 9

Small/large RNA fractionation: 10

Figure S1: Agilent TapeStation electropherogram of small RNA fraction. 10

RNA digestion: 10

LC-MS/MS analysis: 11

Table S1: Analytes and their transitions used for LC-MS/MS analysis. 12

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. 13

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 14

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 14

Figure S1: The levels (ppm of total C) of hm^5C and hm^5Cm in total RNA samples isolated from different HEK293T RNA fractions 14

References: 15

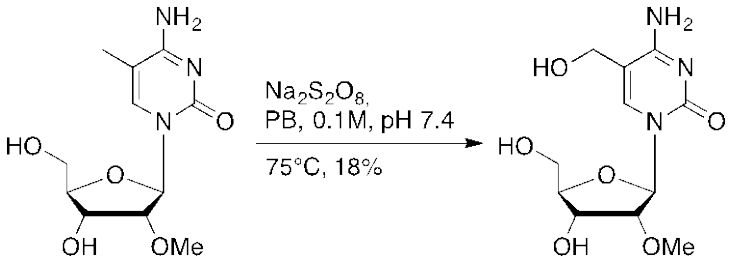
^{1}H-NMR 2'-O-methyl-5-hydroxymethylcytidine: S6 16

^{13}C-NMR 2'-O-methyl-5-hydroxymethylcytidine: 16

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,S7H-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 redisso- vled 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.

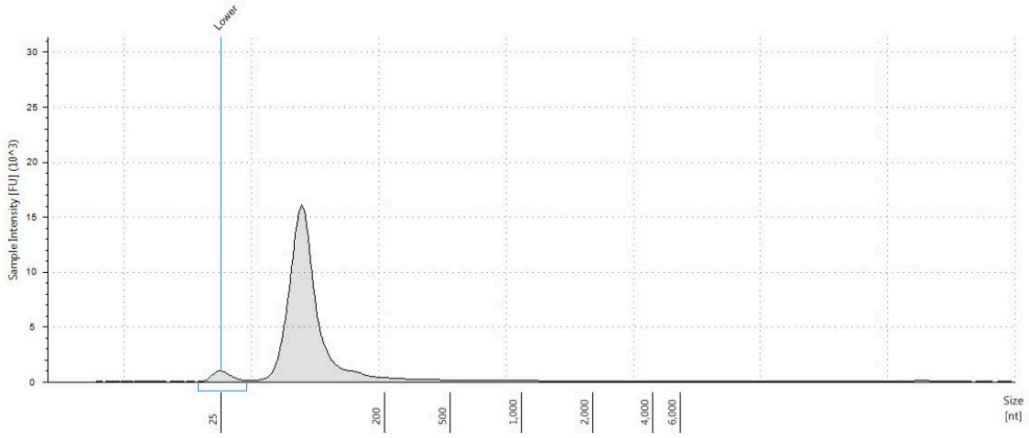


Figure S1: Representative Agilent TapeStation electropherogram of small (< 200 nt) RNA fraction.

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 col-

umns.

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.

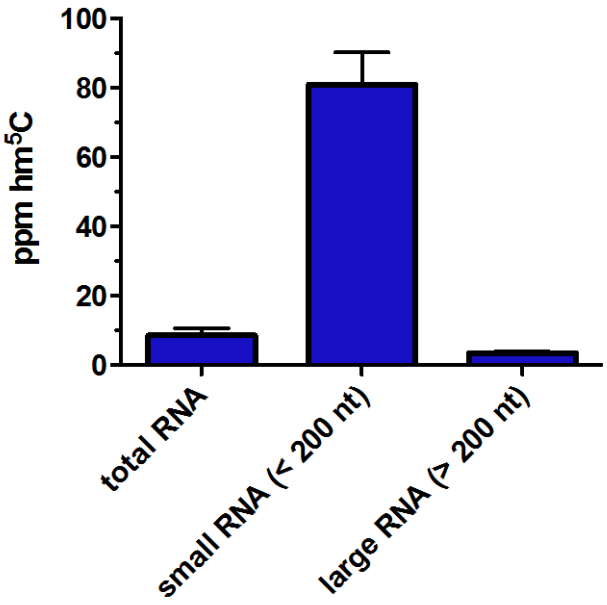


Figure S1: The levels (ppm of total C) of hm^5C and hm^5Cm in total RNA samples isolated from different HEK293T RNA fractions. 3 biological replicates (BR) were

measured.

References:

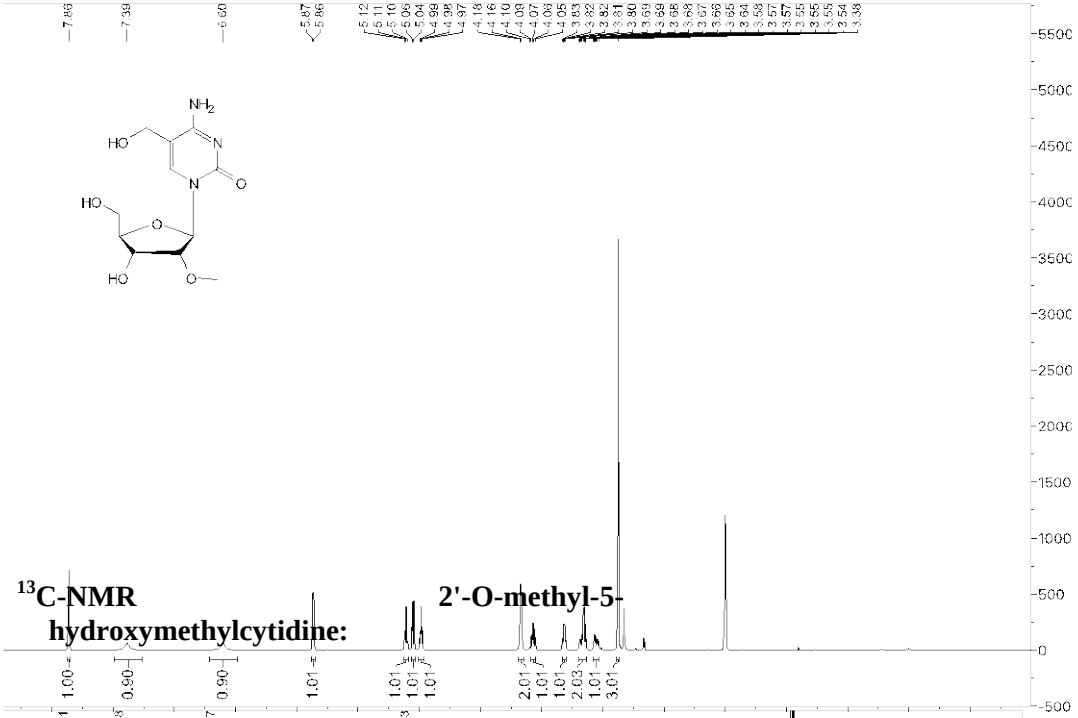
[1] Huber, S. M.; van Delft, P.; Mendil, L.; Bachman, M.; Smollett, K.; Werner, F.; Miska, E. A.; Balasubramanian, S. *ChemBioChem* **2015**, *16*, 752-755.

[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: