



Citation for published version:

Bachman, M, Uribe-Lewis, S, Yang, X, Burgess, HE, Iurlaro, M, Reik, W, Murrell, A & Balasubramanian, S 2015, '5-Formylcytosine can be a stable DNA modification in mammals', *Nature Chemical Biology*, vol. 11, no. 8, pp. 555-557. <https://doi.org/10.1038/nchembio.1848>

DOI:

[10.1038/nchembio.1848](https://doi.org/10.1038/nchembio.1848)

Publication date:

2015

Document Version

Peer reviewed version

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

NCHEMBIO.1848_FINAL_ED

5-Formylcytosine can be a stable DNA modification in mammals

Martin Bachman^{1,2}, Santiago Uribe-Lewis², Xiaoping Yang², Heather E Burgess³, Mario Iurlaro³, Wolf Reik^{3,4}, Adele Murrell^{2,5} & Shankar Balasubramanian^{1,2*}

5-Formylcytosine (5fC) is a rare base found in mammalian DNA and thought to be involved in active DNA demethylation. Here, we show that developmental dynamics of 5fC levels in mouse DNA differ from those of 5-hydroxymethylcytosine (5hmC), and using stable isotope labeling *in vivo*, we show that 5fC can be a stable DNA modification. These results suggest that 5fC has functional roles in DNA that go beyond being a demethylation intermediate.

DNA of all mammalian cells and tissues is methylated at specific loci, mainly in the 5'-cytosine-phosphate-guanine-3' (CpG) context, to modulate the expression of genes¹. 5-Methylcytosine (5mC) is produced from cytosine (C) by dedicated DNA methyltransferases using S-adenosylmethionine (SAM) as a source of the methyl group². In 2009, two independent laboratories found that 5-hydroxymethylcytosine (5hmC) is present in mammalian DNA and is the product of ten-eleven translocation (TET) enzyme-mediated oxidation of 5mC^{3,4}. This oxidized base occurs in all mammalian cells and tissues with global levels ranging between 0.005% and 0.7% of all Cs^{5,6}. The iron(II)- and 2-oxoglutarate-dependent TET enzymes can also oxidize 5hmC further to 5fC and 5-carboxycytosine (5caC), which were found at levels below 0.002% (or 20 p.p.m.) of all Cs in the genomic DNA of mouse embryonic stem (mES) cells and several adult mouse tissues⁷⁻⁹.

One proposed role for these oxidized cytosine bases is to serve as intermediates of enzyme-mediated DNA demethylation initiated by oxidation of 5mC^{10,11}. Indeed, thymine-DNA glycosylase (TDG) can selectively recognize and excise 5fC and 5caC from the genome and trigger a repair process, which can lead to restoration of unmodified C^{7,12}. Moreover, mES cells lacking TDG show increased levels of 5fC and 5caC, suggesting that a proportion of these modifications is constantly being removed from the genome of mES cells^{7,12}. On the other hand, we recently demonstrated that 5hmC is a predominantly stable modification in mammalian DNA, especially in the adult mouse brain where 5hmC is most abundant⁶. Herein we investigate the temporal dynamics of 5fC in genomic DNA *in vivo* to consider whether this rare modification may be stable, rather than an active demethylation intermediate (Fig. 1a).

We first analyzed global levels of all cytosine modifications in the genomic DNA of C57BL/6 mouse tissues to see whether we could detect and quantify 5fC and identify a relationship between its levels and those of its precursors 5mC and 5hmC or its metabolite 5caC. We included a range of postnatal tissues from newborn (1 d old), adolescent (21 d old) and adult (15 week old) mice, as their genomic DNA is known to have different levels of 5hmC depending on the overall proliferation rate (and therefore the age) of the tissue⁶. We also included C57BL/6 embryos at 11.5 d post-fertilization as this is the lethal age for mice lacking TDG^{13,14}, and mES cells derived

from the same strain were added for comparison. To achieve quantification of the rare modifications (5fC and 5caC) with the highest possible sensitivity and accuracy, we employed a nano high-performance liquid chromatography–tandem high-resolution mass spectrometry (nanoHPLC-MS/HRMS) method, which is able to resolve genuine rare modified bases (5fC and 5caC) from potential impurities of the same nominal mass and retention time, and can detect down to 0.1 p.p.m. of total Cs in as little as 100 ng of digested genomic DNA. In addition, the use of isotopically labeled internal standards (IS) of C, 5mC and 5hmC substantially improved the quality of the measurements, ensured excellent reproducibility between technical replicates and excluded spontaneous oxidation of 5hmC as the source of 5fC or 5caC. Example mass spectra, extracted ion chromatograms and calibration curves are shown in **Supplementary Results, Supplementary Figures 1–5**. 5mC and 5hmC were present in all tissues (**Supplementary Fig. 6**), and the levels that we detected were in good agreement with available published data^{5,6,8,9}. While 5mC levels show a relatively uniform distribution between tissues, global 5hmC content is highly correlated to the proportion of proliferating cells in the tissue, as we have shown previously⁶. We found 5fC to be also present in all studied tissues at levels ranging between 0.2 p.p.m. and 15 p.p.m. of all Cs (**Fig. 1b** and **Supplementary Fig. 6**). Notably, 5caC was not detected in any postnatal tissues from C57BL/6 mice, even in those with high 5fC content, but several tissues from C57BL/6 embryos (**Fig. 1b**) and adult (12 week old) CD1 mice (**Supplementary Fig. 7**) contained up to 2 p.p.m. of this rare DNA base modification. Overall, we found no correlation between the levels of 5fC and the levels of its precursors 5mC or 5hmC (**Supplementary Fig. 8**), nor did we find any clear pattern of DNA modification changes as the tissues age. They can retain the levels of 5fC while gaining 5hmC (e.g., brain), lose 5fC while retaining the levels of 5hmC (e.g., heart) or even lose 5fC while gaining 5hmC (e.g., liver) (**Fig. 1c–e**). We found that DNA from mES cells lacking all three TET enzymes (TET triple-knockout (TET-TKO))¹⁵ contained no detectable 5hmC, 5fC or 5caC (**Fig. 1a** and **Supplementary Fig. 6**), confirming that 5hmC is the only source of 5fC and 5caC in mES cell DNA. Although we have no measure of tissue-specific susceptibility to oxidation (such as the quantity of the oxidative lesion 8-oxoguanine), the lack of correlation between global levels of 5hmC and 5fC (**Supplementary Fig. 8**) together with the lack of positional overlap between 5hmC and 5fC in mES cells^{16,17} strongly suggest that 5fC and 5caC are not generated by spontaneous oxidation of 5hmC and 5fC.

To elucidate the stability of 5fC in genomic DNA toward turnover *in vivo*, we applied a stable isotope tracing method consisting of feeding cultured cells and mice with [*methyl*-¹³CD₃]_L-methionine,

¹Department of Chemistry, University of Cambridge, Cambridge, UK. ²Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK.

³Babraham Institute, Babraham, UK. ⁴Wellcome Trust Sanger Institute, Hinxton, UK. ⁵Present address: Centre for Regenerative Medicine, Department of Biology and Biochemistry, University of Bath, Bath, UK. *e-mail: sb10031@cam.ac.uk

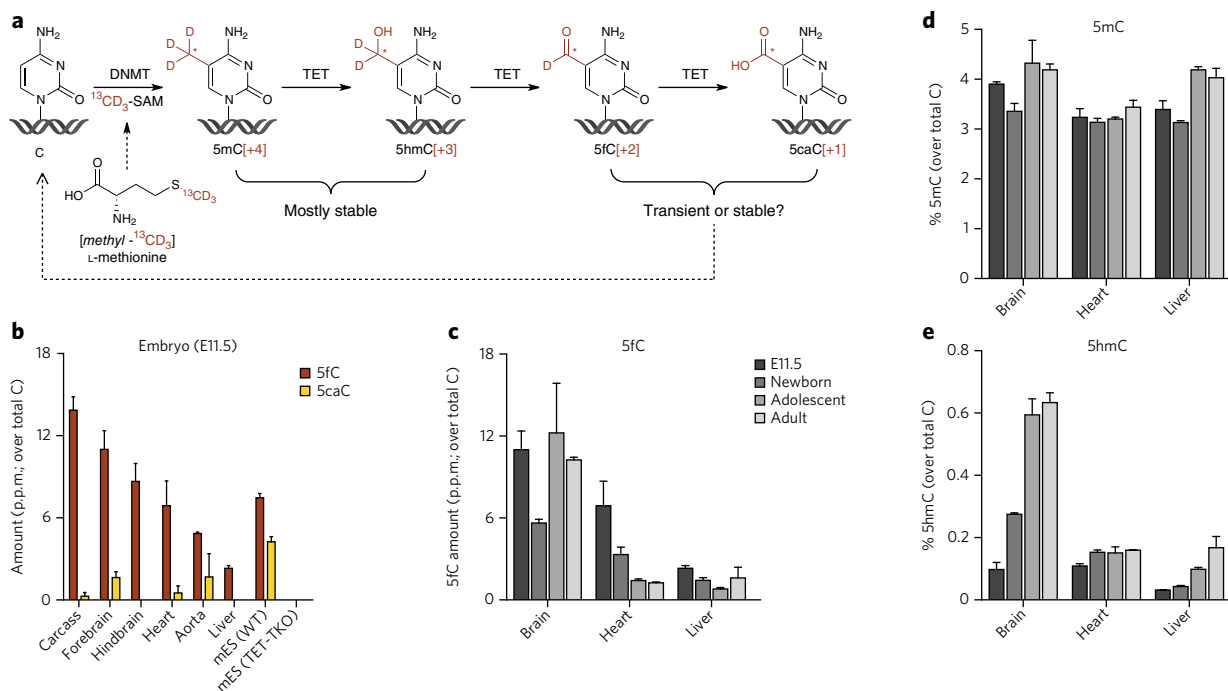


Figure 1 | Dynamics of global levels of 5fC during mouse development are distinct from those of 5hmC. (a) Metabolism of cytosine modifications in DNA. Although the majority of 5mC and 5hmC was known to persist in the genomic DNA, the stability of 5fC and 5caC *in vivo* was unknown. Feeding [*methyl*- $^{13}\text{CD}_3$]-L-methionine can be used as a means to measure the lifetime of cytosine modifications in cells and *in vivo*. Labeling pattern is indicated in red. See also **Figure 2**. (b) Global levels of 5fC and 5caC in genomic DNA from mouse embryos (embryonic day 11.5, E11.5). Shown are mean \pm s.e.m. of three animals. Each sample was analyzed in technical duplicate and the mean value was used. WT, wild type. (c–e) Changes of global 5fC (c), 5mC (d) and 5hmC (e) levels during development in selected C57BL/6 mouse tissues (further data in **Supplementary Fig. 6**). Shown are mean \pm s.e.m. of three embryos (E11.5) (data from **Fig. 1b**), three newborns (1 d old), and two adolescent (21 d old) and two adult (15 week old) mice. See also **Supplementary Figures 7 and 8**.

as we have done previously to study the lifetime of 5hmC⁶. The *methyl*- $^{13}\text{CD}_3$ group enters the intracellular pool of SAM and is transferred into newly methylated Cs by the action of DNA methyltransferase (DNMT) enzymes. The TET enzymes can then convert labeled [*methyl*- $^{13}\text{CD}_3$]-5mC (5mC[+4]) into [*hydroxymethyl*- $^{13}\text{CD}_2$]-5hmC (5hmC[+3]) and [*formyl*- ^{13}CD]-5fC (5fC[+2]) (**Fig. 1a**). The labeling ratios (e.g., % 5fC[+2] over total 5fC) change according to the dynamics and half-life of the given modification in the genomic DNA. For example, a modification that is quickly turning over in DNA would show a high labeling ratio, whereas a very stable modification

would show no labeling in nonproliferating cells or tissues. The maximum obtainable ratio also depends on the activity of other biosynthetic pathways feeding into the one-carbon metabolism. The labeling ratios can be determined very accurately for each modified C using LC-MS/HRMS due to unique masses of the labeled base fragments (**Supplementary Fig. 9** and **Supplementary Table 1**).

We first cultured mES cells in the labeled ([*methyl*- $^{13}\text{CD}_3$]-L-methionine) medium for 8 d, and found that the labeling ratio of 5fC increased much more slowly than that of 5mC and 5hmC. This indicates either a substantial time lag in making 5fC from

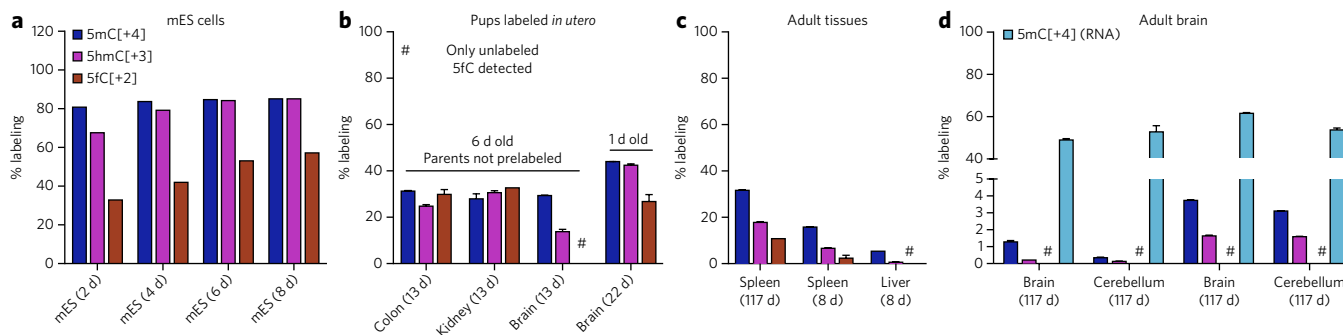


Figure 2 | 5fC can be a stable DNA modification *in vivo*. (a) Labeling ratios of 5mC, 5hmC and 5fC in the genomic DNA of mES cells cultured in the presence of [*methyl*- $^{13}\text{CD}_3$]-L-methionine. Shown are single measurements; total labeling time is given in brackets. (b–d) Labeling ratios of 5mC, 5hmC and 5fC in the genomic DNA of C57BL/6 mice fed with the [*methyl*- $^{13}\text{CD}_3$]-L-methionine diet. (b) 6-d-old pups labeled from 1 week before birth (total labeling time of 13 d) or 1-d-old newborns (total labeling time of 22 d, parents on labeled diet for 52 d before conception). Shown are mean \pm s.e.m. of two animals (6 d old pups) or two technical replicates (1 d old pup). (c,d) Mice labeled in adulthood. Shown are mean \pm s.e.m. of at least two technical replicates from individual mice, and total labeling time is shown in brackets. The absence of 5fC[+2] in the brain (d), where 5fC is most abundant (see **Fig. 1** and **Supplementary Fig. 6**), indicates minimal or no further generation of 5fC once placed in postmitotic tissues. Moreover, if 5fC was involved in cycles of methylation and demethylation, its labeling ratio would be similar to that of 5mC in RNA (d). See also **Supplementary Figure 9** and **Supplementary Table 1**.

newly formed 5hmC or the presence of a population of more slowly dividing or nondividing (unlabeled) cells with higher global levels of 5fC than the quickly dividing (labeled) population of mES cells (Fig. 2a).

We then analyzed genomic DNA from C57BL/6 mice fed with a diet in which all L-methionine was replaced with [*methyl*-¹³CD₃]L-methionine. To gain information about 5fC in developing tissues, we fed a pregnant female the labeled diet starting from 7 d before birth, and kept the family on the diet for 6 more days (so that the 6-d-old pups had been labeled for 13 d when their tissues were harvested). The genomic DNA in tissues such as kidney or colon showed uniform labeling of around 30% for all detectable modifications (5mC, 5hmC and 5fC) (Fig. 2b). However, brain tissue from the same pups showed much less 5hmC[+3] and no detectable 5fC[+2]. This indicates that 5fC was formed in these tissues before the start of labeling and remained there for 13 d until the DNA was harvested. 1-d-old newborns labeled from conception and with pre-labeled parents (starting 52 d before conception, so that total labeling time of pup was 22 d) already showed higher 5fC labeling in the brain, but the ratio was still lower than those of 5mC and 5hmC (27% vs. 44% and 42.5%, respectively) (Fig. 2b). During the gestation period, the labeling ratio of the methionine pool in the pregnant female was still increasing, and therefore this observation is consistent with 5fC being more abundant in the older or more slowly proliferating DNA, as we concluded above. Proliferating tissues (e.g., spleen) from adult mice showed a similar trend in which the 5fC labeling ratio was always smaller than that of 5mC or 5hmC, even in animals labeled for as long as 4 months (117 d) (Fig. 2c). This effect is best explained by 5fC being mostly stable, and again by the presence of nondividing (unlabeled) cells alongside a population of proliferating (labeled) cells that have lower global 5fC levels than the nondividing cells.

Finally, in the mostly nondividing adult brain, where 5fC is most abundant (see Fig. 1 and Supplementary Fig. 6) and only 1.3% and 3.7% of 5mC becomes labeled during the 117-d feeding period, there was no detectable labeled 5fC. Again, this is not due to the lack of an intracellular pool of [*methyl*-¹³CD₃]SAM, as we could measure more than 50% [*methyl*-¹³CD₃]5mC in RNA in adult brain and cerebellum (Fig. 2d). If 5fC was a short-lived DNA modification and was constantly being turned over, its labeling ratio would be close to the labeling ratio of intracellular SAM and 5mC in RNA. If 5fC was short-lived and produced only from pre-existing unlabeled 5hmC in the adult brain, the levels of 5hmC would be depleted over time, which is inconsistent with the high levels of 5hmC in this tissue and is the opposite of what has been described for aging brain¹⁸. Therefore, the lack of 5fC labeling in the adult brain means that this modified base must be stable in the genome as opposed to generally acting as a dynamic intermediate of active DNA demethylation.

In summary, we present the first direct evidence that 5fC (derived from mC by TET-mediated oxidation) can be a stable DNA modification *in vivo*, and provide quantitative measurements of the levels of all modified Cs in mouse tissues across several developmental stages. 5fC levels do not correlate with those of its precursors 5mC and 5hmC or its metabolite 5caC or with the age of the individual. Although there is precedent for removal of 5fC and 5caC from the genome (for example, in mES cells), probably in the process of active DNA demethylation^{7,12}, our findings suggest that the bulk of 5fC can be stable. 5fC has been identified as having

more protein binders than 5mC or 5hmC^{19,20} and having a genomic profile distinct from those of 5mC, 5hmC or 5caC at single-base resolution^{12,16,17,21–23}. Moreover, 5fC has recently been shown to alter the structure of the DNA double helix²⁴. We therefore conclude that such stably 5fC-modified DNA could have profound consequences for the regulation of gene expression that may be distinct to those caused by 5mC and 5hmC. However, direct evidence regarding the biological function of 5fC remains to be demonstrated.

Received 23 February 2015; accepted 15 May 2015; published online 22 June 2015

Methods

Methods and any associated references are available in the online version of the paper.

References

- Ooi, S.K., O'Donnell, A.H. & Bestor, T.H. *J. Cell Sci.* **122**, 2787–2791 (2009).
- Goll, M.G. & Bestor, T.H. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
- Kriaucionis, S. & Heintz, N. *Science* **324**, 929–930 (2009).
- Tahiliani, M. *et al. Science* **324**, 930–935 (2009).
- Globisch, D. *et al. PLoS ONE* **5**, e15367 (2010).
- Bachman, M. *et al. Nat. Chem.* **6**, 1049–1055 (2014).
- He, Y.F. *et al. Science* **333**, 1303–1307 (2011).
- Ito, S. *et al. Science* **333**, 1300–1303 (2011).
- Pfaffeneder, T. *et al. Nat. Chem. Biol.* **10**, 574–581 (2014).
- Song, C.X. & He, C. *Trends Biochem. Sci.* **38**, 480–484 (2013).
- Schiesser, S. *et al. J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
- Neri, F. *et al. Cell Reports* **10**, 674–683 (2015).
- Cortázar, D. *et al. Nature* **470**, 419–423 (2011).
- Cortellino, S. *et al. Cell* **146**, 67–79 (2011).
- Hu, X. *et al. Cell Stem Cell* **14**, 512–522 (2014).
- Booth, M.J., Marsico, G., Bachman, M., Beraldi, D. & Balasubramanian, S. *Nat. Chem.* **6**, 435–440 (2014).
- Song, C.X. *et al. Cell* **153**, 678–691 (2013).
- Kraus, T.F., Guibourt, V. & Kretzschmar, H.A. *J. Neural Transm.* doi:10.1077/s00702-014-1346-4 (2014).
- Iurlaro, M. *et al. Genome Biol.* **14**, R119 (2013).
- Spruijt, C.G. *et al. Cell* **152**, 1146–1159 (2013).
- Booth, M.J. *et al. Science* **336**, 934–937 (2012).
- Yu, M. *et al. Cell* **149**, 1368–1380 (2012).
- Lu, X. *et al. Cell Res.* **25**, 386–389 (2015).
- Raiber, E.A. *et al. Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).

Acknowledgments

We thank C. d'Santos and D. Oxley for their support with mass spectrometry and G. Xu for kindly providing TET-TKO mES cells. This work was supported by Cancer Research UK (C14303/A17197, S.B.), The Wellcome Trust (WT099232, S.B.; WT095645/Z/11/Z, W.R.) and the Biotechnology and Biological Sciences Research Council UK (BB/K010867/1, W.R.).

Author contributions

M.B., S.U.-L. and S.B. conceived the study; S.U.-L., M.B., H.E.B. and M.I. performed experiments; M.B. and X.Y. carried out mass spectrometry and data analysis; S.B., A.M. and W.R. supervised the project; M.B. and S.B. wrote the manuscript with contributions from all authors.

Competing financial interests

The authors declare competing financial interests: details accompany the online version of the paper.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to S.B.

ED: Please check link.

ONLINE METHODS

Animals. All *in vivo* experiments were performed under the terms of a UK Home Office license. C57BL/6 and CD1 mice were bred and housed according to UK Home Office guidelines. Custom L-methionine-free mouse diet supplemented with [*methyl*-¹³CD₃]L-methionine (Sigma) was manufactured by TestDiet.

Cell culture. mES cells were derived by X. Zou at the CRUK Cambridge Institute from a C57BL/6 mouse (Charles River) and cultured on a gelatin-coated plate in a DMEM-KO medium (Invitrogen) supplemented with 10% FCS, MEM non-essential amino acids, glutamine, sodium pyruvate, penicillin, streptomycin, mouse leukemia inhibitory factor (mLIF) and 2i (mitogen-activated protein kinase kinase (MEK) and glycogen synthase kinase (GSK-3 β) inhibitors) as described by Ying *et al.*²⁵. TET-TKO mES cells were obtained from G. Xu¹⁵ and cultured in the 2i conditions as above. All cells were regularly tested for mycoplasma contamination. For isotopic labeling experiments, cells were maintained in a custom L-methionine-free DMEM-KO medium (Invitrogen) supplemented with 30 mg/L of [*methyl*-¹³CD₃]L-methionine (Cambridge Isotope) and the respective components above.

Genomic DNA extraction. Tissues and cells were resuspended in lysis buffer (100 mM Tris, pH 5.5, 5 mM EDTA, 200 mM NaCl, 0.2% SDS) supplemented with 400 μ g/ml proteinase K (Invitrogen), and were incubated at 55 °C overnight. DNA was purified using phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) and Phase Lock Gel (5 Prime), precipitated from 70% ethanol and resuspended in ultrapure HPLC-grade water.

DNA degradation to 2'-deoxynucleosides and LC-MS analysis. 1–2 μ g of DNA was incubated with 5 U of DNA Degradase Plus (Zymo Research) in a total volume of 30 μ l for 4 h at 37 °C. Samples were filtered through a pre-washed Amicon 10 kDa centrifugal filter unit (Millipore) before LC-MS analysis.

LC-MS analysis of global 5mC, 5hmC, 5fC and 5caC levels. Analysis of global levels of 5mC, 5hmC, 5fC and 5caC was performed on a Q Exactive

mass spectrometer (Thermo) fitted with an UltiMate 3000 RSLCnano HPLC (Dionex) and a self-packed hypercarb column (20 mm \times 75 μ m, 3 μ m particle size) at a flow rate of 0.75 μ l/min and a gradient of 0.1% formic acid in water and acetonitrile. Calibration curves were generated using a mixture of synthetic standards 2'-deoxycytidine (Sigma), 5-methyl-, 5-hydroxymethyl-, 5-formyl- and 5-carboxy-2'-deoxycytidine (Berry & Associates), in the ranges of 0.5 nM–5 μ M for C, 0.025–50 nM for 5mC and 0.005–50 nM for 5hmC, 5fC and 5caC. Samples and synthetic standards were spiked with an isotopically labeled mix containing 100 nM of 2'-deoxycytidine-(¹⁵N,_d₂) (synthesis and characterization in Bachman *et al.*⁶), 5-methyl-2'-deoxycytidine-(d₃) and 5-hydroxymethyl-2'-deoxycytidine-(d₃) (both from Toronto Research Chemicals). Target ions were fragmented in a positive ion mode at 10% normalized collision energy, and full scans (50–300 Da) were acquired. The inclusion list contained the following masses: C (228.1), C_{IS} (231.1), 5mC (242.1), 5mC_{IS} (245.1), 5hmC (258.1), 5hmC_{IS} (261.1), 5fC (256.1), 5caC (272.1). Extracted ion chromatograms of base fragments (see **Supplementary Fig. 1**) were used for quantification. Results are expressed as a % or p.p.m. of total cytosines.

LC-MS analysis of isotope incorporation into genomic DNA. Analysis of isotope incorporation into DNA was performed using the same instrumental set up as above, targeting ions of masses 242.1 (mC), 246.1 (mC[+4]), 258.1 (5hmC and 5fC[+2]), 261.1 (5hmC[+3]) and 256.1 (5fC). Extracted ion chromatograms of base fragments were used for quantification of labeling ratios (see also **Supplementary Fig. 8**). Results are expressed as % labeling (for example, % 5fC[+2] represents the percentage of labeled 5fC[+2] in total 5fCs).

LC-MS analysis of isotope incorporation into 5mC in RNA. Total RNA was carried through during genomic DNA extraction (no RNase treatment), during hydrolysis to nucleosides and LC-MS/HRMS analysis. An additional mass of 262.1 was targeted for RNA 5mC[+4] (unlabeled 5mC was present in the 258.1 channel), and base fragments 126.0662 (5mC) and 130.0884 (5mC[+4]) were used for quantification of % labeling as above.

25. Ying, Q.L. *et al. Nature* **453**, 519–523 (2008).

5-Formylcytosine can be a stable DNA modification in mammals

Martin Bachman, Santiago Uribe-Lewis, Xiaoping Yang, Heather E Burgess, Mario Iurlaro, Wolf Reik, Adele Murrell & Shankar Balasubramanian

EDITORIAL SUMMARY

AOP: 5-Formylcytosine (5fC), produced by TET-mediated oxidation of 5-methylcytosine, is considered an intermediate in active DNA demethylation. Labeling studies and LC/MS analysis across mouse developmental stages reveals that 5fC modifications are more persistent in the genome and may have other functional roles.

