Epigenetics

International Edition: DOI: 10.1002/anie.201700424
German Edition: DOI: 10.1002/ange.201700424

Quantitative LC–MS Provides No Evidence for m^6dA or m^4dC in the Genome of Mouse Embryonic Stem Cells and Tissues

*Sarah Schiffers, Charlotte Ebert, René Rahimoff, Olesea Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, Fabio Spada, Stylianos Michalakis, Jörg Nickelsen, Markus Müller, and Thomas Carell**

Abstract: *Until recently, it was believed that the genomes of higher organisms contain, in addition to the four canonical DNA bases, only 5-methyl-dC (m^5dC) as a modified base to control epigenetic processes. In recent years, this view has changed dramatically with the discovery of 5-hydroxymethyl-dC (hmdC), 5-formyl-dC (fdC), and 5-carboxy-dC (cadC) in DNA from stem cells and brain tissue. N^6-methyldeoxyadenosine (m^6dA) is the most recent base reported to be present in the genome of various eukaryotic organisms. This base, together with N^4-methyldeoxycytidine (m^4dC), was first reported to be a component of bacterial genomes. In this work, we investigated the levels and distribution of these potentially epigenetically relevant DNA bases by using a novel ultrasensitive UHPLC–MS method. We further report quantitative data for m^5dC, hmdC, fdC, and cadC, but we were unable to detect either m^4dC or m^6dA in DNA isolated from mouse embryonic stem cells or brain and liver tissue, which calls into question their epigenetic relevance.*

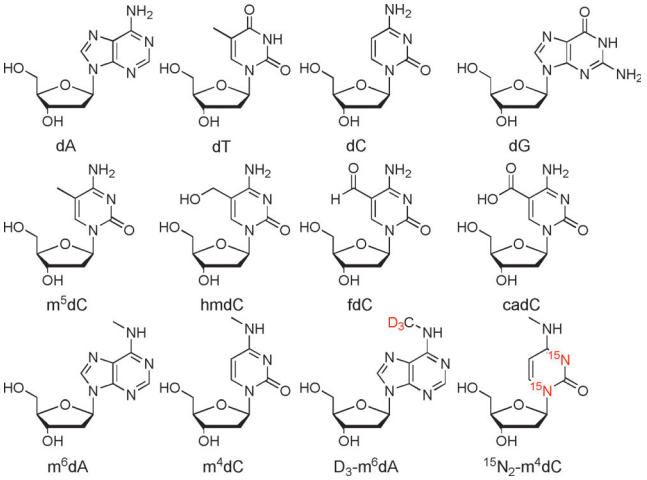


Figure 1. Depiction of the four canonical DNA bases and the epigenetic DNA marks m^5dC, hmdC, fdC, and cadC, as well as the bases m^6dA and m^4dC together with the synthesized isotopologues.

The genetic material of living organisms is constructed from the four canonical nucleobases dA, dC, dG, and dT, which establish the sequence information that, in multicellular organisms, is stored in the nucleus of every cell (Figure 1). In addition to the canonical bases, the methylated dC base 5-methyldeoxycytidine (m^5dC) is frequently found.[1] The presence or absence of this base in specific promoter segments determines whether the gene is actively transcribed or silenced.[1] The cell-type-specific distribution of m^5dC thus determines the identity of a given cell. Recently, 5-hydroxymethyldeoxycytidine (hmdC) was found as a sixth base of the genetic system[2,3] and in 2011, 5-formyldeoxycytidine (fdC)[4,5] and 5-carboxydeoxycytidine (cadC)[5,6] were also discovered, particularly in DNA isolated from stem cells, but also in brain DNA. It is currently believed that fdC and cadC are intermediates in an active demethylation process that allows cells to change the methylation pattern and hence the activity state of specific genes.[7,8] For fdC, separate epigenetic functions are also envisaged.[9]

While the genomes of bacteria are known to also contain N^4-methyldeoxycytidine (m^4dC)[10] and N^6-methyldeoxyadenosine (m^6dA),[11] attempts to detect these bases in the DNA of higher organisms have failed until recently.[12–15] m^6dA has now been found in algae (0.4 mol% m^6dA/A),[12] fruit flies (0.001%-0.07% m^6dA/A),[14] and *C. elegans* (0.01%-0.4% m^6dA/A),[13] and its presence has even been reported in the DNA of vertebrates (0.00009% in *X. laevis*[16] and 0.00019–0.003% of dA in murine cells and tissue[17]). These discoveries, especially concerning the DNA of vertebrates, have spurred a worldwide research interest in unraveling the function of these new bases in human genomic DNA.[18–20]

In this study, we developed an ultrasensitive triple quadrupole mass spectrometry (QQQ-MS) method, which in combination with ultra-high-pressure chromatography (UHPLC) enables m^4dC and m^6dA to be searched for and quantified in parallel to the more established new epigenetic DNA marks m^5dC, hmdC, fdC and cadC.

[*] S. Schiffers, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher, Dr. F. Spada, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPSM) at the Department of Chemistry, LMU München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: http://www.carellgroup.de
Dr. A.-V. Bohne, Prof. Dr. J. Nickelsen
Biocenter of the LMU
Dept. Biologie 1 – Botanik, Molekulare Pflanzenwissenschaften
Grosshaderner Strasse 2–4, 82152 Planegg-Martinsried (Germany)
Dr. S. Michalakis
CiPS[M], Department of Pharmacy, LMU München
Pharmacology for Life Sciences
Butenandtstr. 7, 81377 München (Germany)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: http://dx.doi.org/10.1002/anie.201700424.

For the quantitative measurements, we first chemically synthesized the two isotopologues of m^6dA and m^4dC shown in Figure 1 as internal standards for the analytical method. The prepared compounds D_3-m^6dA and $^{15}N_2$-m^4dC are three and two mass units heavier, respectively, than the natural bases. Despite these molecular-weight differences, they have identical properties during the UHPL chromatography step so that they strictly coelute with their natural counterparts, thus allowing them to enter the mass spectrometer at exactly the same time as the internal standards. The availability of these isotopologues makes the method highly reliable and strictly quantitative. The syntheses of the two compounds, together with all analytical data, are given in the Supporting Information.

We first benchmarked our study with an investigation of genomic DNA isolated from the unicellular green algae *Chlamydomonas reinhardtii* and the cyanobacterium *Synechocystis*. In both cases, DNA was isolated after cell lysis using a standard method (see the Supporting Information). The isolated DNA was subsequently digested with a mixture of three commercially available digestion enzymes (Nuclease S1, Antarctic Phosphatase, and Snake Venom Phosphodiesterase; see the Supporting Information). We next added the isotope-labelled standards D_3-m^6dA and $^{15}N_2$-m^4dC to the obtained digestion mixture and performed UHPLC-QQQ analysis. For the mass spectrometry detection, we selected fragmentation of the glycosidic bond as the indicative and hence recorded mass transition. This is $m/z = 266.12 \rightarrow 150.08$ for m^6dA and $m/z = 269.14 \rightarrow 153.10$ for its isotopologue D_3-m^6dA. For m^4dC, we also used fragmentation of the glycosidic bond, which gives a mass transition of $m/z = 242.11 \rightarrow 126.07$ for the natural compound m^4dC and $m/z = 244.11 \rightarrow 128.07$ for its isotopologue $^{15}N_2$-m^4dC (Figure 2 A).

We next modified the reported UHPLC-QQQ method[10] for the simultaneous quantification of m^4dC and m^6dA, together with the other epigenetically relevant bases m^5dC, hmdC, fdC, and cadC. To this end, the UHPLC gradient was fine-tuned to enable full separation of all six compounds. Finally, we measured precise calibration curves for all of the compounds (see Figures S1 and S2 in the Supporting Information). This subsequently allowed exact quantification of all of the discussed epigenetic DNA marks in a given sample (Figure 2 B–D).

Since m^4dC and m^6dA are well known in bacteria, we first analysed the cyanobacterium *Synechocystis* (PCC6803), and we indeed found both bases (Figure 2 B). The base m^6dA was detected at a level of 8.4×10^{-3} per dN and for m^4dC we measured a value of 5.9×10^{-3} per dN. The constitutional isomer m^5dC and all other dC-derived epigenetic DNA marks were detectable, but were not quantified in this experiment.

Next, we analyzed two different strains of *Chlamydomonas reinhardtii* (CC-3491 and wt 7d+), in which m^6dA has just recently been discovered,[12] and the levels of m^6dA were determined to be 8.4×10^{-4} per dN for CC-3491 and 6.9×10^{-4} per dN for wt 7d+ (Figure 2 C). This corresponds to about 3000 m^6dA bases per *Chlamydomonas* genome (genome size 1.2×10^8), which at 0.7% of the dA is a relatively high number. In both strains, m^4dC was not detected, thus showing that this base is unlikely to be a component of the genetic material of *Chlamydomonas*. This is interesting because *Synechocystis* is considered a relative of the chloroplasts present in *Chlamydomonas*.

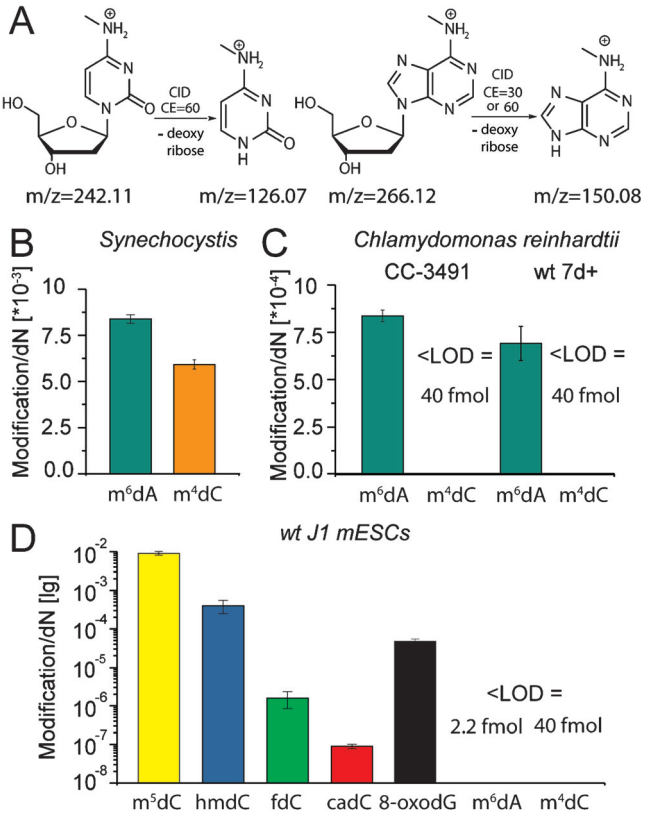


Figure 2. A) Fragmentation patterns of m^4dC and m^6dA. B–D) Quantitative data of the bases m^4dC and m^6dA in *Synechocystis* (B), *Chlamydomonas* (C) and of these bases and the other epigenetic DNA marks hmdC, fdC, cadC, and m^5dC in mouse embryonic stem cells (D).

With these positive results in hand, we extended our study to mouse embryonic stem cells (ESCs; wt J1, Figure 2 D). m^6dA in particular was recently reported to occur as an epigenetically relevant DNA mark in mouse ESCs (mESC cell line wt TT2).[17] When performing the measurements, we turned the mass spectrometer to maximum sensitivity. But even in this mode, we were unable to detect a signal for m^6dA within the detection limits of our system Table S2. In contrast, the other epigenetically relevant bases hmdC, fdC, cadC, and even the oxidative lesion 8-oxodG, which we also quantified in parallel, were clearly detectable. The 8-oxodG level was 4.8×10^{-5} per dN. The rare and difficult to detect cadC was clearly seen even at levels of only 9.0×10^{-8} per dN. For m^6dA, in contrast, a signal did not appear. We also re-measured wt TT2 cells as described and still did not detect m^6dA over background levels (see Figure S4).[17] To obtain unequivocal proof that m^6dA is not present in stem cells, we added $^{13}CD_3$-methionine to the mESC culture. Methionine provides the methyl group for the biosynthesis of m^6dA. With $^{13}CD_3$-methionine, this would lead to an m/z-shift of +4. We tuned the mass spectrometer to the new m/z-transition and again were unable to see any signal for $^{13}CD_3$-m^6dA (Figure S8).

We subsequently turned our attention to adult mouse tissue and analysed DNA isolated from liver and whole brain

(Figure S3) using our UHPLC-QQQ method. Figure 3 shows the data obtained from mouse liver. The middle column shows the data we obtained for hmdC. The already reported D_2-$^{15}N_2$-hmdC standard elutes at a retention time of 2.25 min

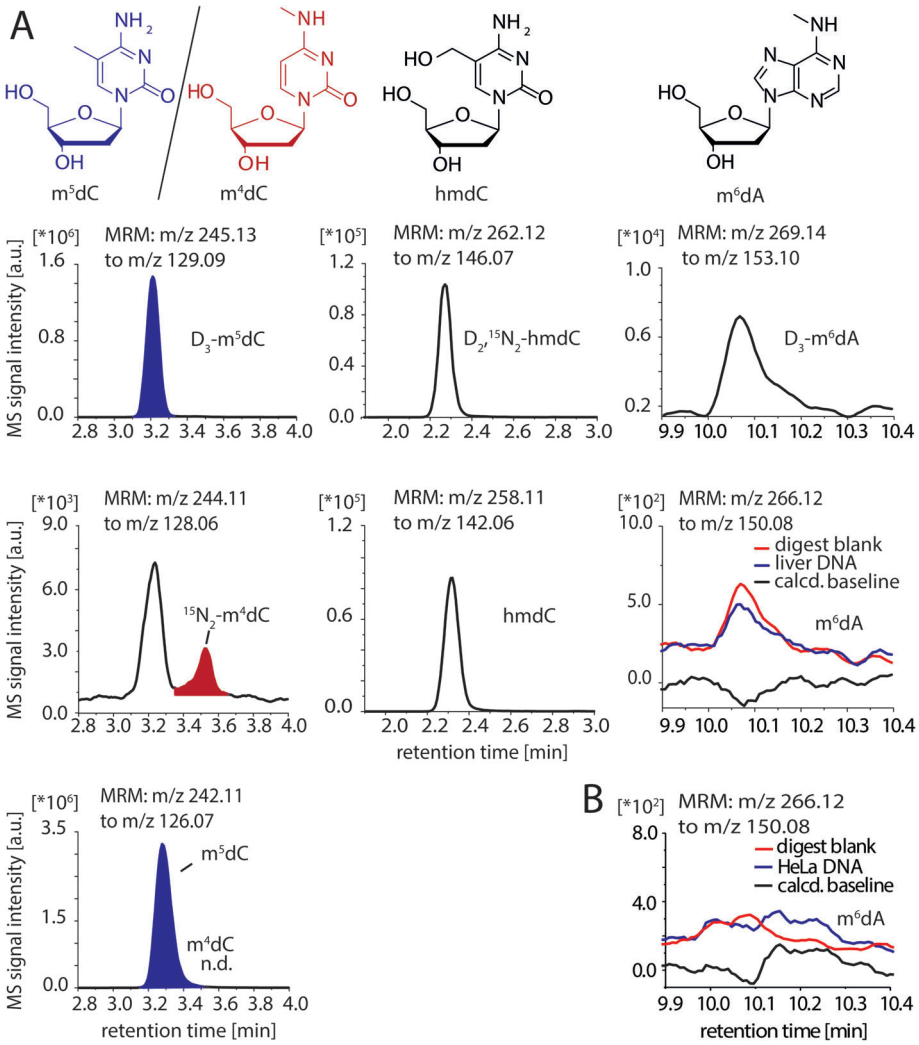


Figure 3. A) Chromatograms of the mass signal of mouse liver DNA. UHPLC-QQQ data obtained for m^5dC and m^4dC (left), hmdC (middle), and m^6dA (right, blue line) and their corresponding isotopic standards are shown. Additionally, for m^6dA, the chromatogram for the digest blank is shown (red line) and a computed baseline (black line), which was determined by subtracting the digest blank from the sample. B) A chromatogram of the mass signal from UHPLC-QQQ data obtained for m^6dA in DNA from HeLa cells.

and shows the expected fragmentation of the glycosidic bond, providing the fragmentation signature $m/z = 262.12 \rightarrow 146.07$, which allowed assignment of the signal. The naturally occurring hmdC is detected at exactly the same retention time with a mass transition of $m/z = 258.11 \rightarrow 142.06$, thus unequivocally demonstrating the presence of hmdC in murine liver DNA. Regarding the different monomethylated dC compounds m^5dC and m^4dC (left column), the epigenetic DNA mark m^5dC is clearly detected at a retention time of 3.2 min, but for m^4dC with a retention time of 3.5 min, there is obviously no signal present.

The m^6dA data are highly interesting (Figure 3 right column). While the D_3-m^6dA standard was clearly detectable at a retention time of 10.1 min, the unlabelled m^6dA provided a very weak signal. We then performed a control experiment to determine the limit of detection and investigated the digestion solution alone, which contains all of the commercial enzymes but no isolated DNA (red chromatogram). A weak signal for m^6dA was again detectable at a retention time of 10.1 min. After subtracting this background signal (red) from the measured chromatogram (blue), we obtained the black line showing that the original signal at 10.1 min is purely caused by background derived from the enzyme mixture. Here it is important to note that most of these proteins are recombinant proteins obtained from bacterial expression systems and bacterial DNA contains plenty of m^6dA. To support the evidence that m^6dA is not present in vertebrate DNA, we performed the same experiment with HeLa cells and also observed no signal for m^6dA (Figure 3B).

We then determined our limit of detection for m^6dA to be 3.5×10^{-7} per dN, which corresponds to 170 m^6dA bases per murine genome. This is a very small number and demonstrates the excellent sensitivity of our method. It shows that the maximum number of m^6dA that could be theoretically present and would not be detectable by our method is less than 170 m^6dA bases per genome, which led us to conclude that m^6dA is likely not epigenetically relevant but rather formed as a DNA lesion, perhaps by misguided methyltransferases. Spiking tests with synthetic nucleoside and DNA from *Chlamydomonas* nevertheless confirmed the sensitivity of the method, since the input amount equalled the found amount (see Figure S6).

To find a potential source for m^6dA in mESC DNA that could explain previous sequencing data,[17] we thought that m^6dA-containing bacterial DNA that gets degraded could provide the m^6dA nucleoside, which then might get incorporated into mESC DNA. This is indeed a possibility. When we added the m^6dA nucleoside to a mESCs culture, we indeed saw incorporation of some m^6dA into the genome (Figure S7).

We also cannot fully exclude the possibility that the presence of few m^6dA bases, at levels below our detection limit, could have a biological function. In addition, it is possible that at certain stages of organismal development, certain methyltransferases are activated that may induce high m^6dA levels at specific time points that may have escaped our detection.[18] Our data, however, show clearly that the maximum possible levels of m^6dA in the analysed organisms and mESCs under normal conditions are far lower than so far believed.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1032 (TP-A5), SFB749 (TP-A4), SPP1784, CA275 and the Excellence Cluster CiPS^M.

Conflict of interest

The authors declare no conflict of interest.

Keywords: DNA · epigenetics · methyldeoxyadenosine · methyldeoxycytidine · mass spectrometry

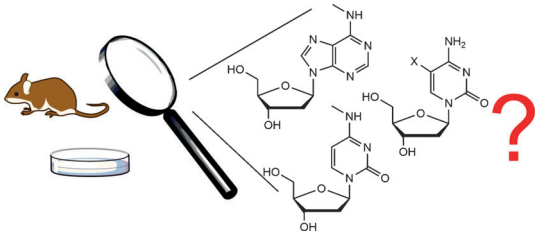
[1] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204 – 220.
[2] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930 – 935.
[3] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929 – 930.
[4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008 – 7012; *Angew. Chem.* **2011**, *123*, 7146 – 7150.
[5] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300 – 1303.
[6] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303 – 1307.
[7] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472 – 479.
[8] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593 – 14599.
[9] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797 – 11800; *Angew. Chem.* **2016**, *128*, 11974 – 11978.
[10] M. Yu, L. Ji, D. A. Neumann, D. H. Chung, J. Groom, J. Westpheling, C. He, R. J. Schmitz, *Nucleic Acids Res.* **2015**, *43*, e148.
[11] M. Ehrlich, M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, C. W. Gehrke, *Nucleic Acids Res.* **1985**, *13*, 1399 – 1412.
[12] Y. Fu, G. Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L. C. Dore, X. Weng, Q. Ji, L. Mets, C. He, *Cell* **2015**, *161*, 879 – 892.
[13] E. L. Greer, M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizabal-Corrales, C. H. Hsu, L. Aravind, C. He, Y. Shi, *Cell* **2015**, *161*, 868 – 878.
[14] G. Zhang, H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang, D. Chen, *Cell* **2015**, *161*, 893 – 906.
[15] S. Hattman, C. Kenny, L. Berger, K. Pratt, *J. Bacteriol.* **1978**, *135*, 1156 – 1157.
[16] M. J. Koziol, C. R. Bradshaw, G. E. Allen, A. S. Costa, C. Frezza, J. B. Gurdon, *Nat. Struct. Mol. Biol.* **2016**, *23*, 24 – 30.
[17] T. P. Wu, T. Wang, M. G. Seetin, Y. Lai, S. Zhu, K. Lin, Y. Liu, S. D. Byrum, S. G. Mackintosh, M. Zhong, A. Tackett, G. Wang, L. S. Hon, G. Fang, J. A. Swenberg, A. Z. Xiao, *Nature* **2016**, *532*, 329 – 333.
[18] J. Liu, Y. Zhu, G. Z. Luo, X. Wang, Y. Yue, X. Wang, X. Zong, K. Chen, H. Yin, Y. Fu, D. Han, Y. Wang, D. Chen, C. He, *Nat. Commun.* **2016**, *7*, 13052.
[19] W. Huang, J. Xiong, Y. Yang, S.-M. Liu, B.-F. Yuan, Y.-Q. Feng, *RSC Adv.* **2015**, *5*, 64046 – 64054.
[20] G. Z. Luo, F. Wang, X. Weng, K. Chen, Z. Hao, M. Yu, X. Deng, J. Liu, C. He, *Nat. Commun.* **2016**, *7*, 11301.

Manuscript received: January 13, 2017
Revised: February 15, 2017
Final Article published: ■■ ■■, ■■■■

Communications

Epigenetics

S. Schiffers, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher, A.-V. Bohne, F. Spada, S. Michalakis, J. Nickelsen, M. Müller, T. Carell* ——— ▮▮▮▮–▮▮▮▮



Quantitative LC–MS Provides No Evidence for m^6dA or m^4dC in the Genome of Mouse Embryonic Stem Cells and Tissues

Previous results challenged: Highly sensitive mass spectrometry reveals that m^6dA is most likely not an epigenetic base in the mouse genome. It appears that in vertebrates, the levels of this modified base, and that of N^4-methyldeoxycytidine, are much lower than previously thought.