

A Quantitative Sequencing Method for 5-Formylcytosine in RNA

Ruitu Lyu,⁺[a] Kinga Pajdzik,⁺[a] Hui-Lung Sun,^[a] Linda Zhang,^[a] Li-Sheng Zhang,^[a] Tong Wu,^[a] Lei Yang,^[b] Tao Pan,^{*,[a]} Chuan He,^{*,[a]} and Qing Dai^{*,[a]}

Abstract: 5-Formylcytosine (f⁵C) modification is present in human mitochondrial methionine tRNA (mt-tRNA^{Met}) and cytosolic leucine tRNA (ct-tRNA^{Leu}), with their formation mediated by NSUN3 and ALKBH1. f⁵C has also been detected in yeast mRNA and human tRNA, but its transcriptome-wide distribution in mammals has not been studied. Here we report f⁵C-seq, a quantitative sequencing method to map f⁵C transcriptome-wide in HeLa and mouse embryonic stem cells (mESCs). We show that f⁵C in RNA can be reduced to dihydrouracil (DHU) by pic-borane, and DHU

can be exclusively read as T during reverse transcription (RT) reaction, allowing the detection and quantification of f⁵C sites by a unique C-to-T mutation signature. We validated f⁵C-seq by identifying and quantifying the two known f⁵C sites in tRNA, in which the f⁵C modification fractions dropped significantly in ALKBH1-depleted cells. By applying f⁵C-seq to chromatin-associated RNA (caRNA), we identified several highly modified f⁵C sites in HeLa and mouse embryonic stem cells (mESC).

Keywords: 5-Formylcytosine · quantitative sequencing · transcriptome-wide · mutation rate · read-through rate

Over 100 naturally occurring RNA modifications have been identified so far, with some of them playing various roles in gene expression regulation.^[1–3] As the most abundant internal modification in eukaryotic mRNA, N⁶-methyladenosine (m⁶A) is dynamically regulated and involved in numerous aspects of mRNA metabolism, such as alternative splicing,^[4] nuclear export,^[5] stability,^[6] translation^[7,8] and decay.^[9] In recent years, studies on transcriptome-wide sequencing of other mRNA modifications have also been emerging. The reported sequencing methods can be grouped as: (1) Antibody-based MeRIP-seq for m⁶A,⁴ m¹A,^[10–13] ac⁴C^{14,15}, m⁵C¹⁶ and hm⁵C¹⁷. These methods rely on antibody-based enrichment but could neither achieve base precision nor reveal absolute modification fraction. (2) Reverse transcription (RT) stop-based methods such as CMC-based pseudouridine sequencing^[18] and low dNTP-based 2'-O-Me sequencing.^[19] While these methods can detect modification sites at base resolution, they usually have high false-positive rates since RT stop signatures could be generated non-specifically.^[20] (3) RT mutation-based approaches, such as methods to map m⁶A,^[21–24] m⁷G^[25–27] and m¹A^[28] that generate mutation signatures at modified sites in order to achieve single base resolution with low background. (4) RT deletion-based approaches, such as BS-Induced quantitative pseudouridine sequencing.^[29,30] Another consideration in RNA modification is the modification stoichiometry at each site. The modification fraction is a biological parameter that is directly related to the modification dynamics and their regulatory functions.

5-Methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), and 5-formylcytosine (5fC) are DNA modifications that are important intermediates in an active DNA 5mC

demethylation pathway. Sequencing methods for these modified bases in DNA have been well documented.^[31–37] However, these modified bases also occur naturally in RNA, and their biological roles remain to be elucidated. m⁵C has been reported to protect RNA from degradation,^[38] regulate mRNA export,^[39] and promote the pathogenesis of bladder cancer.^[40] Additionally, m⁵C on nuclear mRNA can serve as DNA damage codes to regulate DNA repair.^[41] hm⁵C has been

[a] R. Lyu,⁺ K. Pajdzik,⁺ H.-L. Sun, L. Zhang, L.-S. Zhang, T. Wu, T. Pan, C. He, Q. Dai

Department of Chemistry,
Department of Biochemistry and Molecular Biology,
Howard Hughes Medical Institute,
The University of Chicago,
Chicago, IL 60637, USA
E-mail: daiqing@uchicago.edu
chuanhe@uchicago.edu
taopan@uchicago.edu

[b] L. Yang

First Maternity & Infant Hospital,
School of Medicine,
Tongji University,
Shanghai, China

[⁺] These authors contributed equally.

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/ijch.202300111>

© 2023 The Authors. *Israel Journal of Chemistry* published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Research Article

detected in mRNA,^[42] and its presence was found to favor mRNA translation.^[17] f⁵C displays approximately 100% modification fraction at C34 in mt-tRNA^{Met}^[43] and a moderate modification fraction in ct-tRNA^{Leu} in human cells.^[43–45] In both cases, NSUN3 was reported to be the methyltransferase that converts the target C to m⁵C, while ALKBH1 further catalyzes the oxidation of m⁵C to f⁵C.^[43] f⁵C in tRNA is associated with several human diseases^[46] and f⁵C in tRNA-Leu-CAA promotes decoding under stress conditions.^[47] In addition, the presence of f⁵C in yeast and human mRNA has also been detected by LC–MS/MS.^[48,49] Here, we describe f⁵C-seq, a new method for quantitative sequencing of f⁵C in HeLa and mouse embryonic stem cells (mESCs). To detect 5fC in DNA, Zhu *et al* used malononitrile to specifically react with 5fC in DNA to generate a cyclized base which induces a C-to-T transition during DNA amplification.^[50] Recently, a new method that employs pic-borane to reduce 5fC in DNA to DHU which could be read as T during amplification was reported.^[51] We speculated that pic-borane reduction may also convert f⁵C base in RNA to DHU under optimized conditions, and RT enzyme may read through DHU efficiently and generate high C-to-T mutation rate to enable f⁵C detection and quantitation at base resolution in RNA.

To ascertain whether pic-borane facilitates the efficient and quantitative conversion of f⁵C to DHU in RNA, we initiated our study with the treatment of a 5-mer RNA oligo containing an f⁵C modification with pic-borane under different conditions (Table S1). The reactions were monitored using MALDI-TOF MS. We found that the reduction products were temperature dependent. At 25 °C, f⁵C was primarily reduced to dihydro-f⁵C (DHf⁵C) *via* 3,4-reduction, where DHU was obtained as the sole product at 70 °C *via* further deformylation and subsequent deamination (Figure 1a–b). Moreover, we detected a small peak at 1,529 Daltons at 25 °C, which represents the intermediate of 3,4-reduction and deformylation, but without deamination (Figure 1a–b). This observation suggests that deformylation occurs after 3,4-reduction and prior to deamination, which differs from the proposed mechanism for the pic-borane reduction of 5fC in DNA, in which deamination was thought to occur before deformylation.^[51]

To determine whether RT enzymes can read through DHU and produce a C-to-T mutation, we conducted a primer extension reaction using the DHU counterpart of an f⁵C-containing 33mer RNA oligo (Table S1, Figure 1c). Notably, both the untreated f⁵C-containing probe and the pic-borane treated sample rendered full-length products using the SuperScript II RT enzyme. In contrast, the sample treated with malononitrile predominantly produced RT-stop byproducts. The resulting cDNA products were then amplified by RT-PCR followed by Sanger sequencing. Our analysis revealed that untreated f⁵C was read as C, while malononitrile treatment led to approximately 50% C-to-T mutations. Impressively, pic-borane treatment produced a significantly elevated C-to-T mutation rate of over 80% (Figure 1d). Collectively, these results suggest that pic-borane mediated conversion of f⁵C to DHU provides higher read-through and C-to-T mutation rates

compared with malononitrile treatment. Additionally, we observed no significant RNA degradation when a 45mer f⁵C-containing RNA oligo (Table S1) was treated with pico-borane at a temperature between 55 to 70 °C (Figure S1). The mild nature of pico-borane treatment paved the way for the development of f⁵C-seq, which performs reduction after integrating the RNA fragments into library construction, followed by high-throughput sequencing (Figure S2).

We next investigated whether the C-to-T mutation rate is dependent on f⁵C sequence context and whether there is a linear correlation between the mutation rate and f⁵C fractions.^[11,25,28] To do this, we treated fragmented small RNA isolated from HeLa cells treated with *E. coli* AlkB demethylase to remove the major tRNA methylations that block RT.^[52] We then added spike-in oligos with NNf⁵CNN motifs (N represents a mixture of A, C, G and U) and five pairs of RNA oligos with different f⁵C modification fractions (Table S1). After performing 3'- and 5'-ligations, we treated the ligated RNA with pic-borane followed by RT reaction, PCR amplification, and sequencing to determine the C-to-T mutation rates of the reduced f⁵C. Our results showed that the C-to-T mutation rates were consistently high in all 256 NNf⁵CNN oligos, suggesting that the C-to-T mutation rate is generally independent of the f⁵C sequence context (Figure 2a). To our delight, we observed a nearly linear calibration curve, which allows us to precisely deduce the f⁵C modification fraction from the observed C-to-T mutation rate in RNA (Figure 2b).

In order to construct libraries suitable for sequencing, an alkaline fragmentation step is necessary. Initially, we performed MALDI TOF MS analysis of f⁵C-containing oligo treated in 0.1 M NaHCO₃ pH 9.2 at 95 °C for 9 min to evaluate the potential impact of alkaline fragmentation on f⁵C in RNA. Our data shows that f⁵C remains unaffected under alkaline fragmentation condition (Figure S3). Previous studies have shown that ALKBH1 catalyzes f⁵C formation in both mt-tRNA^{Met} and ct-tRNA^{Leu} in human cells.^[43] Therefore, we used small RNA isolated from shControl and shALKBH1 HeLa cells to construct f⁵C-seq libraries (Figure S4a–b). We then examined the C-to-T mutation rates at the known f⁵C sites in tRNAs. In HeLa cells, we observed a high C-to-T mutation rate of approximately 80% at the mt-tRNA^{Met} f⁵C site and a low C-to-T mutation rate of around 15% at the ct-tRNA^{Leu} f⁵C site (Figure 2c, S5a), corresponding to f⁵C modification fractions of 87.2% and 16.4%, respectively, which is consistent with the previous reports based on mass spectrometry analysis.^[43] The bases surrounding the f⁵C sites had minimal background mutation (Figure 2c). Two known f⁵C sites at tRNAs also showed very low C-to-U mutation rates in the input libraries (Figure S5). Additionally, we observed a marked reduction of f⁵C modification fractions in ALKBH1-deficient HeLa cells, while the mutation frequencies at adjacent cytosine sites remained unchanged upon ALKBH1 knockdown (Figure 2c, S5a). Similarly, our findings also revealed that both C34 sites in mt-tRNA^{Met} and ct-tRNA^{Leu} in mESCs are f⁵C-modified (Figure 2d, S5b), with a similar f⁵C modification fraction to that in the corresponding HeLa

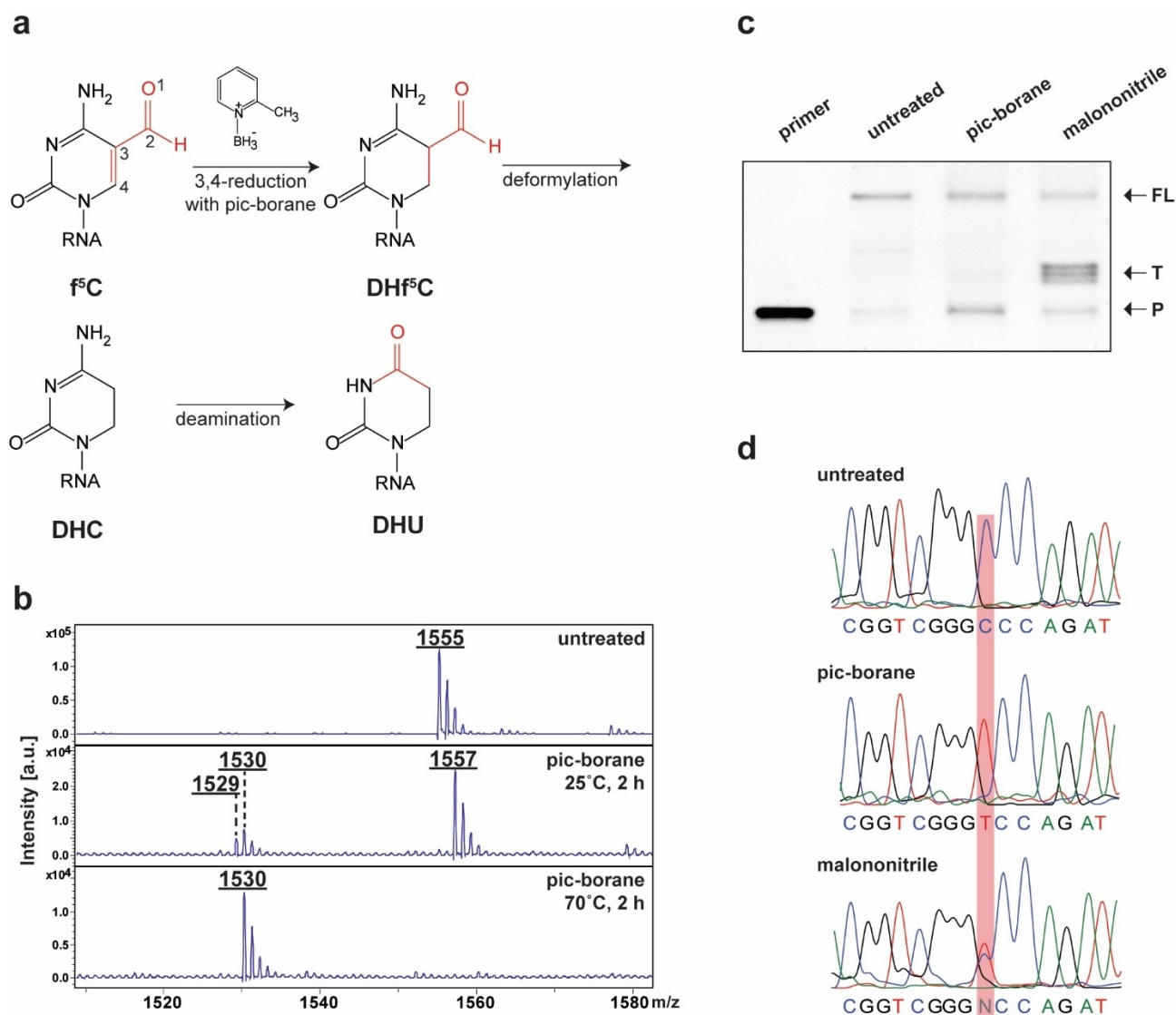


Figure 1. f^5C -seq and its chemical validation. a) Suggested pic-borane f^5C reduction mechanism based on the observed intermediates. b) MALDI-TOF MS analysis comparing an untreated f^5C -containing RNA probe with the same probe treated with pic-borane for 2 h at either 25 °C or 70 °C. The observed peaks at m/z values 1,555, 1,557, 1,529 and 1,530 represent oligos integrated with f^5C , dihydro-5-formylcytosine (DH f^5C), dihydrocytosine (DHC) and DHU, respectively. Notably, the peak at 1,557 represents the intermediate that f^5C is reduced via 3,4-reduction and undergoes subsequent deformylation process, yielding DHC followed by further deamination to produce DHU. c) Primer extension assay of 33-mer f^5C -containing RNA oligo treated with pic-borane and malonitrile. FL: full length; T: truncated product; P: primer. d) Sanger sequencing of untreated, pic-borane and malonitrile-treated f^5C -containing 33-mer RNA oligos followed by RT-PCR.

tRNAs. Notably, the f^5C fractions decreased to nearly undetectable levels in ALKBH1-deficient mES cells (Figure 2d, S5b). Taken together, these findings robustly confirm the accuracy and quantitative reliability of our f^5C -seq method in detecting f^5C modifications at base resolution in RNA.

Given that f^5C has previously been identified within human mRNA, we then tried to map transcriptome-wide f^5C sites in polyA⁺ RNA isolated from both HeLa and mES cells using f^5C -seq. Although we identified several hundred f^5C sites in both cell lines, the f^5C fraction at each site did not exceed 10%. Interestingly, when we employed f^5C -seq on chromatin-

associated RNA (caRNA) from HeLa and mES cells, we identified multiple f^5C sites with high fraction levels (Table S2). This includes a site on the MER68 ERVL endogenous retrovirus-related Long Terminal Repeats (LTR) in HeLa cells (Figure 3a), and another on the U3 snRNA repeats in mES cells (Figure 3b). While the modification fraction at the MER68 ERVL LTR site remained relatively stable following ALKBH1 knockdown (Figure 3c), we observed a notable reduction at the U3 snRNA repeat site. Specifically, upon ALKBH1 depletion, the modification fraction declined markedly from 39.04% to 26.62%, which corresponds to a

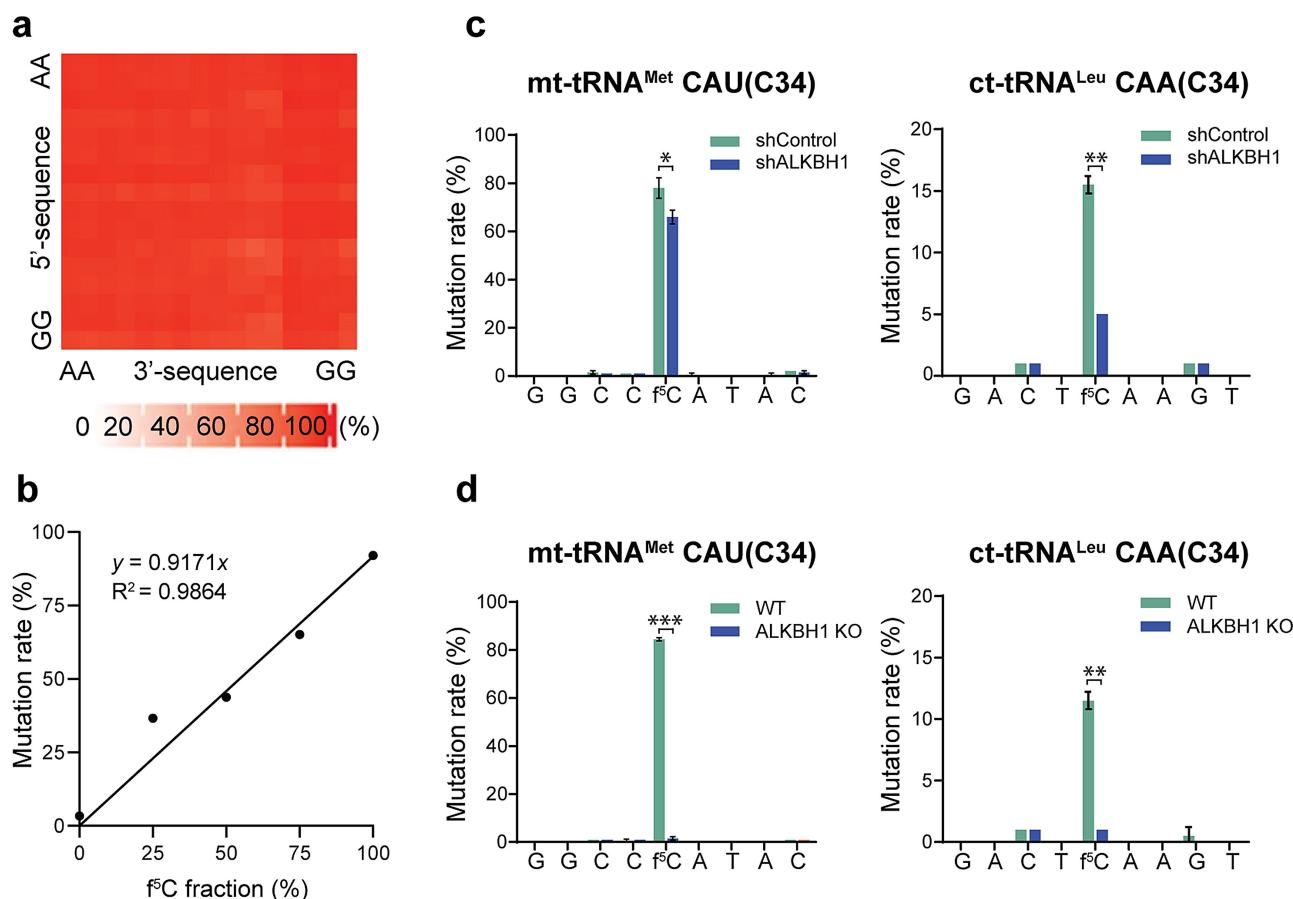


Figure 2. Validation of f^C-seq by identifying two known f^C sites in human tRNA with next generation sequencing. a) Mutation rate is independent of sequence context around the f^C site. b) Calibration curve of spike-in oligos containing f^C with varying f^C fractions and C-to-T mutation rates. c) C-to-T mutation rates of mt-tRNA^{Met} CAU(C34) and ct-tRNA^{Leu} CAA(C34) sites, as well as their neighboring sites, in shControl and shALKBH1 HeLa cells. d) C-to-T mutation rates of mt-tRNA^{Met} CAU(C34) and ct-tRNA^{Leu} CAA(C34) sites, as well as their neighboring sites, in WT and ALKBH1-KO mESC. Bars represent mean of two technical replicates ± SD. Statistical significance was determined by t-test using the Holm-Sidak method (*p ≤ 0.01, **p ≤ 0.01, ***p ≤ 0.001).

decrement in the f^C fraction from 42.57% to 29.27% (Figure 3d). This data presents a compelling avenue for further exploration into the dynamic roles and regulation of f^C modifications in RNA biology.

It is worth mentioning that during the preparation of this manuscript, several other RNA f^C sequencing methods have been published.^[53–55] One of these methods uses pyridine borane as a reductant,^[53] while the other was based on the selective and efficient malonitrile-mediated labeling of f^C residues to generate adducts that are read as C-to-T mutations upon reverse transcription.^[54] However, our f^C-seq method distinguishes itself by utilizing pic-borane as a reductant, akin to the pyridine borane used in published method. Through extensive analysis, we demonstrated that pic-borane can proficiently reduce f^C to DHU, similarly inducing C-to-T transitions at f^C sites during RT-PCR, which facilitates f^C single-base resolution detection. When contrasted with other methods that employ malonitrile or photo-mediated labeling, our technique stands out for its simplicity and efficiency

in mapping transcriptome-wide f^C sites. These newly developed methods represent exciting developments in the field and offer alternative approaches to sequencing f^C modifications. The emergence of multiple methods for detecting f^C modifications highlights the growing interest in this area of research and suggests that there is still much to be learned about the function and regulation of these modifications in various cellular contexts. As the field continues to evolve, it will be important to compare the strengths and limitations of different approaches and to identify the best methods for studying f^C modifications in different biological systems.

In summary, we have developed f^C-seq, a quantitative sequencing method for mapping f^C modification in RNA. Our method is based on the chemical principle that f^C in RNA can be specifically and quantitatively reduced to DHU by pic-borane at a higher temperature, and DHU is read as T instead of C in RNA sequencing. It is worth noting that although in principle ca⁵C can also be converted to DHU by

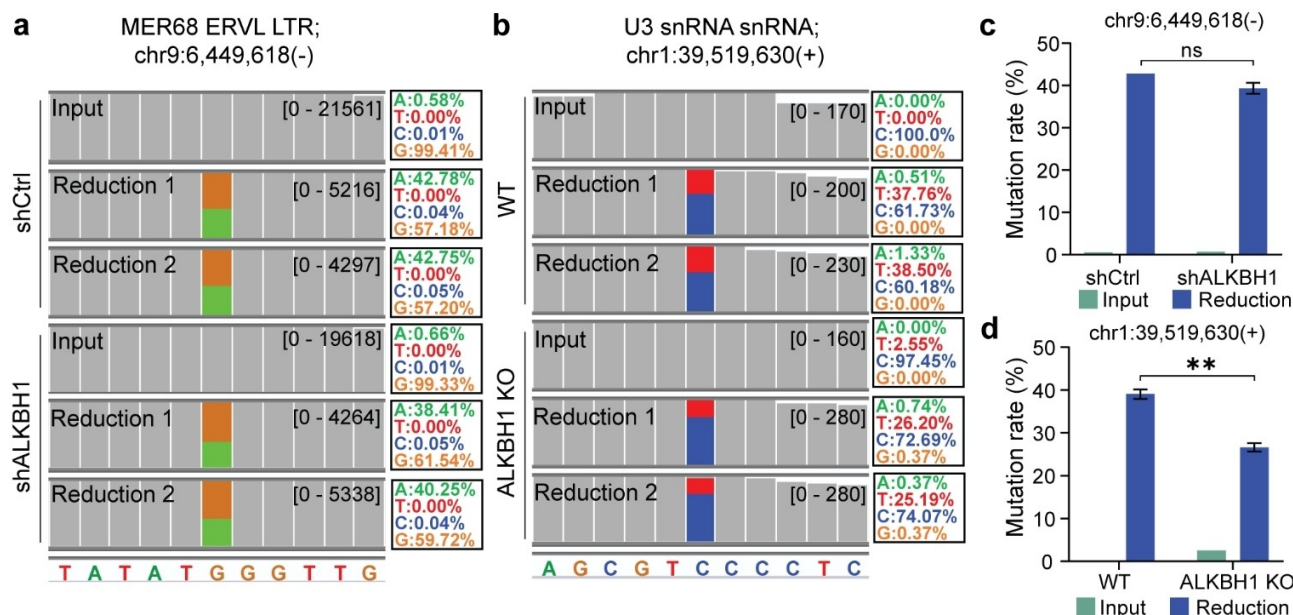


Figure 3. Overview of f^5C sites detected in caRNA in HeLa and mES cells. a) IGV tracks showing the mutation signature of the identified f^5C site on caRNA MER68 ERVL LTR in shCtrl and shALKBH1 HeLa cells. b) IGV tracks showing the mutation signature of identified f^5C site on caRNA U3 snRNA repeats in wild-type (WT) and ALKBH1 knockout (KO) mES cells. c) C-to-T mutation rates of detected tRNA f^5C site on caRNA MER68 ERVL LTR in shCtrl and shALKBH1 HeLa cells. Statistical significance was determined by t-test using the Holm-Sidak method (ns: not significant). d) C-to-T mutation rate of detected f^5C site on caRNA U3 snRNA repeats in WT and ALKBH1 KO mES cells. Bars represent mean of two technical replicates \pm SD. Statistical significance was determined by t-test using the Holm-Sidak method (** $p \leq 0.01$).

pic-borane to generate C to T mutation, so far, no ca^5C has been detected in RNA. Using f^5C -seq, we verified the two known f^5C sites of mt-tRNA^{Met} (C34) and ct-tRNA^{Leu} (C34) in human tRNA and confirmed that their f^5C modification fractions are sensitive to ALKBH1 knockdown. Further sequencing confirmed that ALKBH1 is also responsible for the formation of these two f^5C sites in mES cells. We then sequenced f^5C in HeLa and mESCs polyA⁺ RNA. The f^5C levels in identified hundreds of sites were low and did not exceed 10%. This result is consistent with the low f^5C levels (~1.7 ppm) measured in HEK293 C polyA⁺ RNA by LC-MS/MS by Arguello et al.^[56] We also applied f^5C -seq to caRNA from HeLa and mES cells and detected several highly modified f^5C sites that were not reported previously. We found that f^5C located on mouse U3 snRNA repeats was sensitive to ALKBH1 depletion, suggesting that ALKBH1 is also responsible for f^5C formation at this position. Interestingly, the f^5C fraction identified on human MER68 ERVL LTR did not change upon ALKBH1 knockdown (KD). Further studies are needed to unravel the enzyme responsible for the formation of f^5C on human MER68 ERVL LTR. Other RNA modifications, notably m⁶A, have been co-transcriptionally integrated into various caRNAs in mammalian cells. These modifications play a pivotal role in controlling RNA abundance, which in turn influences gene transcription through alterations in chromatin accessibility.^[57] Intriguingly, in our studies, we have identified multiple sites on caRNA with pronounced f^5C modifications in both HeLa and mouse ES cells. Noteworthy

among these are the f^5C sites present on the caRNA MER68 ERVL LTR and U3 snRNA repeats. These f^5C sites exhibit diverse responses to ALKBH1 KD, which hints at the potential diverse roles of f^5C in orchestrating chromatin states, influencing transcription, and governing alternative splicing. This diverges from its established regulatory function in translation. Furthermore, given the prevalence of m⁵C sites on both caRNA^[58] and mRNA,^[59] we speculate that f^5C could serve as an intermediate in a potential RNA demethylation process. Taken together, f^5C -seq provides a quantitative tool for future studies on the biological function of f^5C in RNA.

Associated Content

Experimental protocols, supporting figures and oligonucleotides sequences are to be found in **Supporting Information**. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Acknowledgements

This study was supported by a grant from National Institutes of Health (NIH) Grant RM1 HG008935 (C.H.)

Author Contributions

All authors have given approval to the final version of the manuscript.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

- [1] D. Dominissini, *Science* **2014**, *346*, 1192.
- [2] B. S. Zhao, I. A. Roundtree, C. He, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 31–42.
- [3] Y. Saletore, K. Meyer, J. Korlach, I. D. Vilfan, S. R. Jaffrey, C. E. Mason, *Genome Biol.* **2012**, *13*.
- [4] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, *Nature* **2012**, *485*, 201–206.
- [5] I. A. Roundtree, G. Z. Luo, Z. Zhang, X. Wang, T. Zhou, Y. Cui, J. Sha, X. Huang, L. Guerrero, P. Xie, E. He, B. Shen, C. He, *eLife* **2017**, *6*, 1–28.
- [6] E. Kierzek, R. Kierzek, *Nucleic Acids Res.* **2003**, *31*, 4472–4480.
- [7] K. L. Heilman, R. A. Leach, M. T. Tuck, *Int. J. Biochem. Cell Biol.* **1996**, *28*, 823–829.
- [8] K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S. B. Qian, S. R. Jaffrey, *Cell* **2015**, *163*, 999–1010.
- [9] K. Chen, Z. Lu, X. Wang, Y. Fu, G. Z. Luo, N. Liu, D. Han, D. Dominissini, Q. Dai, T. Pan, C. He, *Angew. Chem. Int. Ed.* **2015**, *54*, 1587–1590.
- [10] D. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N. Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, G. Zheng, T. Pan, O. Solomon, E. Eyal, V. Herschkovitz, D. Han, L. C. Doré, N. Amariglio, G. Rechavi, C. He, *Nature* **2016**, *530*, 441–446.
- [11] X. Li, X. Xiong, M. Zhang, K. Wang, Y. Chen, J. Zhou, Y. Mao, J. Lv, D. Yi, X. W. Chen, C. Wang, S. B. Qian, C. Yi, *Mol. Cell* **2017**, *68*, 993–1005.
- [12] X. Li, X. Xiong, K. Wang, L. Wang, X. Shu, S. Ma, C. Yi, *Nat. Chem. Biol.* **2016**, *12*, 311–316.
- [13] M. Safra, A. Sas-Chen, R. Nir, R. Winkler, A. Nachshon, D. Bar-Yaacov, M. Erlacher, W. Rossmann, N. Stern-Ginossar, S. Schwartz, *Nature* **2017**, *551*, 251–255.
- [14] W. R. Sinclair, D. Arango, J. H. Shrimp, T. T. Zenggeya, J. M. Thomas, D. C. Montgomery, S. D. Fox, T. Andresson, S. Oberdoerffer, J. L. Meier, *ACS Chem. Biol.* **2017**, *12*, 2922–2926.
- [15] D. Arango, D. Sturgill, N. Alhusaini, A. A. Dillman, T. J. Sweet, G. Hanson, M. Hosogane, W. R. Sinclair, K. K. Nanan, M. D. Mandler, S. D. Fox, T. T. Zenggeya, T. Andresson, J. L. Meier, J. Collier, S. Oberdoerffer, *Cell* **2018**, *175*, 1872–1886.
- [16] T. Amort, D. Rieder, A. Wille, D. Khokhlova-Cubberley, C. Riml, L. Trixl, X. Y. Jia, R. Micura, A. Lusser, *Genome Biol.* **2017**, *18*, 1–16.
- [17] B. Delatte, E. Collignon, E. Bonvin, R. Deplus, E. Calonne, B. Hassabi, P. Putmans, C. Creppe, M. Defrance, F. Fuks, F. Wang, S. Minakhina, R. Steward, L. V. Ngoc, R. Soin, C. Gueydan, V. Kruijs, S. Awe, J. Kreher, A. Brehm, C. Wetzel, P. A. Limbach, *Science* **2016**, *351*, 282–285.
- [18] T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli, W. V. Gilbert, *Nature* **2014**, *515*, 143–146.
- [19] D. Incarnato, F. Anselmi, E. Morandi, F. Neri, M. Maldotti, S. Rapelli, C. Parlato, G. Basile, S. Oliviero, *Nucleic Acids Res.* **2017**, *45*, 1433–1441.
- [20] M. Zaringhalam, F. N. Papavasiliou, *Methods* **2016**, *107*, 63–72.
- [21] L. Hu, S. Liu, Y. Peng, R. Ge, R. Su, C. Senevirathne, B. T. Harada, Q. Dai, J. Wei, L. Zhang, Z. Hao, L. Luo, H. Wang, Y. Wang, M. Luo, M. Chen, J. Chen, C. He, *Nat. Biotechnol.* **2022**, *40*, 1210–1219.
- [22] R. Ge, C. Ye, Y. Peng, Q. Dai, Y. Zhao, S. Liu, P. Wang, L. Hu, C. He, *Nat. Protoc.* **2022**, *18*, 626–657.
- [23] C. Liu, H. Sun, Y. Yi, W. Shen, K. Li, Y. Xiao, F. Li, Y. Li, Y. Hou, B. Lu, W. Liu, H. Meng, J. Peng, C. Yi, J. Wang, *Nat. Biotechnol.* **2023**, *41*, 355–366.
- [24] Y. L. Xiao, S. Liu, R. Ge, Y. Wu, C. He, M. Chen, W. Tang, *Nat. Biotechnol.* **2023**, *41*, 993–1003.
- [25] L. S. Zhang, C. Liu, H. Ma, Q. Dai, H. L. Sun, G. Luo, Z. Zhang, L. Zhang, L. Hu, X. Dong, C. He, *Mol. Cell* **2019**, *74*, 1304–1316.
- [26] C. Enroth, L. D. Poulsen, S. Iversen, F. Kirpekar, A. Albrechtsen, J. Vinther, *Nucleic Acids Res.* **2019**, *47*, e126.
- [27] L. Pandolfini, I. Barbieri, A. J. Bannister, A. Hendrick, B. Andrews, N. Webster, P. Murat, P. Mach, R. Brandi, S. C. Robson, V. Migliori, A. Alendar, M. d’Onofrio, S. Balasubramanian, T. Kouzarides, *Mol. Cell* **2019**, *74*, 1278–1290.
- [28] H. Zhou, S. Rauch, Q. Dai, X. Cui, Z. Zhang, S. Nachtergaele, C. Sepich, C. He, B. C. Dickinson, *Nat. Methods* **2019**, *16*, 1281–1288.
- [29] M. Zhang, Z. Jiang, Y. Ma, W. Liu, Y. Zhuang, B. Lu, K. Li, J. Peng, C. Yi, *Nat. Chem. Biol.* **2023**, *19*, 1185–1195.
- [30] Q. Dai, L. S. Zhang, H. L. Sun, K. Pajdzik, L. Yang, C. Ye, C. W. Ju, S. Liu, Y. Wang, Z. Zheng, L. Zhang, B. T. Harada, X. Dou, I. Irklyenko, X. Feng, W. Zhang, T. Pan, C. He, *Nat. Biotechnol.* **2023**, *41*, 344–354.
- [31] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [32] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, N. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
- [33] C. X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68–75.
- [34] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368–1380.
- [35] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678–691.
- [36] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435–440.
- [37] M. J. Booth, M. R. Branco, G. Ficiz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934–937.
- [38] F. Tuorto, R. Liebers, T. Musch, M. Schaefer, S. Hofmann, S. Kellner, M. Frye, M. Helm, G. Stoecklin, F. Lyko, *Nat. Struct. Mol. Biol.* **2012**, *19*, 900–905.
- [39] X. Yang, Y. Yang, B. F. Sun, Y. S. Chen, J. W. Xu, W. Y. Lai, A. Li, X. Wang, D. P. Bhattarai, W. Xiao, H. Y. Sun, Q. Zhu, H. L. Ma, S. Adhikari, M. Sun, Y. J. Hao, B. Zhang, C. M. Huang, N. Huang, G. Bin Jiang, Y. L. Zhao, H. L. Wang, Y. P. Sun, Y. G. Yang, *Cell Res.* **2017**, *27*, 606–625.

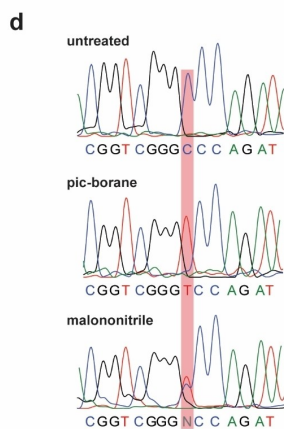
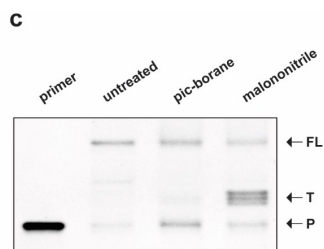
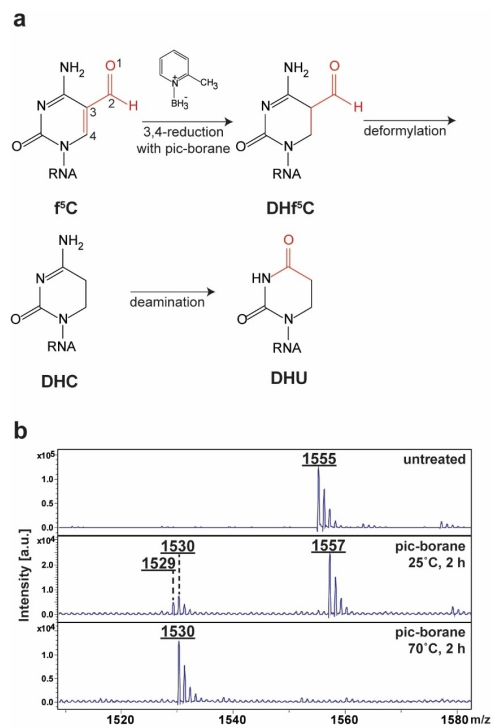
- [40] X. Chen, A. Li, B. Sun, Y. Yang, Y. Han, X. Yuan, R. Chen, *Nucleic Acids Res.* **2018**, *46*, 1565–1583.
- [41] H. Chen, H. Yang, X. Zhu, T. Yadav, J. Ouyang, S. S. Truesdell, J. Tan, Y. Wang, M. Duan, L. Wei, L. Zou, A. S. Levine, S. Vasudevan, L. Lan, *Nat. Commun.* **2020**, *11*, 3–14.
- [42] S. M. Huber, P. Van Delft, L. Mendil, M. Bachman, K. Smollett, F. Werner, E. A. Miska, S. Balasubramanian, *ChemBioChem* **2015**, *16*, 752–755.
- [43] L. Kawarada, T. Suzuki, T. Ohira, S. Hirata, K. Miyauchi, T. Suzuki, *Nucleic Acids Res.* **2017**, *45*, 7401–7415.
- [44] S. Haag, K. E. Sloan, N. Ranjan, A. S. Warda, J. Kretschmer, C. Blessing, B. Hübner, J. Seikowski, S. Dennerlein, P. Rehling, M. V. Rodnina, C. Höbartner, M. T. Bohnsack, *EMBO J.* **2016**, *35*, 2104–2119.
- [45] K. Asano, T. Suzuki, A. Saito, F. Y. Wei, Y. Ikeuchi, T. Numata, R. Tanaka, Y. Yamane, T. Yamamoto, T. Goto, Y. Kishita, K. Murayama, A. Ohtake, Y. Okazaki, K. Tomizawa, Y. Sakaguchi, T. Suzuki, *Nucleic Acids Res.* **2018**, *46*, 1565–1583.
- [46] N. Jonkhout, J. Tran, M. A. Smith, N. Schonrock, J. S. Mattick, E. M. Novoa, *RNA* **2017**, *23*, 1754–1769.
- [47] A. E. Arguello, A. Li, X. Sun, T. W. Eggert, E. Mairhofer, R. E. Kleiner, *Nat. Commun.* **2022**, *13*, 1–17.
- [48] M. Tardu, J. D. Jones, R. T. Kennedy, Q. Lin, K. S. Koutmou, *ACS Chem. Biol.* **2019**, *14*, 1403–1409.
- [49] W. Huang, M. D. Lan, C. B. Qi, S. J. Zheng, S. Z. Wei, B. F. Yuan, Y. Q. Feng, *Chem. Sci.* **2016**, *7*, 5495–5502.
- [50] C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, *Cell Stem Cell* **2017**, *20*, 720–731.e5.
- [51] Y. Liu, P. Siejka-Zielińska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Böckler, C. X. Song, *Nat. Biotechnol.* **2019**, *37*, 424–429.
- [52] G. Zheng, Y. Qin, W. C. Clark, Q. Dai, C. Yi, C. He, A. M. Lambowitz, T. Pan, *Nat. Methods* **2015**, *12*, 835–837.
- [53] Y. Wang, Z. Chen, X. Zhang, X. Weng, J. Deng, W. Yang, F. Wu, S. Han, C. Xia, Y. Zhou, Y. Chen, X. Zhou, *ACS Chem. Biol.* **2022**, *17*, 77–84.
- [54] A. Li, X. Sun, A. E. Arguello, R. E. Kleiner, *ACS Chem. Biol.* **2022**, *17*, 503–508.
- [55] X. Y. Jin, Z. R. Huang, L. J. Xie, L. Liu, D. L. Han, L. Cheng, *Angew. Chem. Int. Ed.* **2022**, *61*, e202210652.
- [56] A. E. Arguello, A. Li, X. Sun, T. W. Eggert, E. Mairhofer, R. E. Kleiner, *Nat. Commun.* **2022**, *13*.
- [57] J. Liu, X. Dou, C. Chen, C. Chen, C. Liu, M. Michelle Xu, S. Zhao, B. Shen, Y. Gao, D. Han, C. He, *Science.* **2020**, *367*, 580–586.
- [58] F. Aguilo, S. De Li, N. Balasubramanian, A. Sancho, S. Benko, F. Zhang, A. Vashisht, M. Rengasamy, B. Andino, C. hung Chen, F. Zhou, C. Qian, M. M. Zhou, J. A. Wohlschlegel, W. Zhang, F. J. Suchy, M. J. Walsh, *Cell Rep.* **2016**, *14*, 479–492.
- [59] X. Chen, A. Li, B. F. Sun, Y. Yang, Y. N. Han, X. Yuan, R. X. Chen, W. S. Wei, Y. Liu, C. C. Gao, Y. S. Chen, M. Zhang, X. D. Ma, Z. W. Liu, J. H. Luo, C. Lyu, H. L. Wang, J. Ma, Y. L. Zhao, F. J. Zhou, Y. Huang, D. Xie, Y. G. Yang, *Nat. Cell Biol.* **2019**, *21*, 978–990.

Manuscript received: July 26, 2023

Revised manuscript received: September 27, 2023

Version of record online: ■■, ■■

RESEARCH ARTICLE



R. Lyu, K. Pajdzik, H.-L. Sun, L. Zhang, L.-S. Zhang, T. Wu, L. Yang, T. Pan*, C. He*, Q. Dai*

1 – 8

A Quantitative Sequencing Method for 5-Formylcytosine in RNA

