Organic & Biomolecular Chemistry

Base-Resolution Analysis of 5-hydroxymethylcytidine by Selective Oxidation and Reverse Transcription Arrest

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID	OB-ART-05-2021-000995.R1
Article Type:	Paper
Date Submitted by the Author:	29-Jun-2021
Complete List of Authors:	Koyama, Kenta; The University of Tokyo Hayashi, Gosuke; Nagoya University, Biomolecular Engineering Ueda, Hiroki; The University of Tokyo Ota, Satoshi; The University of Tokyo Nagae, Genta; The University of Tokyo Aburatani, Hiroyuki; The University of Tokyo Okamoto, Akimitsu; The University of Tokyo, Research Center for Advanced Science and Technology

ARTICLE

Base-Resolution Analysis of 5-hydroxymethylcytidine by Selective Oxidation and Reverse Transcription Arrest

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Kenta Koyama,[a] Gosuke Hayashi,[b] Hiroki Ueda,[c] Satoshi Ota,[c] Genta Nagae,[c] Hiroyuki Aburatani,[c] Akimitsu Okamoto*[,a, c]

While 5-hydroxymethylcytidine in RNA (^{hm5}C) is associated with cellular development and differentiation, its distribution and biological function remain largely unexplored because suitable detection methods are lacking. Here, we report a base-resolution sequencing method for ^{hm5}C in RNA by applying peroxotungstate-mediated chemical conversion of ^{hm5}C to trihydroxylated thymine (thT). Reverse transcription by SuperScript III terminated at the thT site, probably because of its unnatural nucleobase structure producing truncated cDNA. Consequently, base-resolution analysis of the ^{hm5}C sites in RNA was achieved with both Sanger sequencing and Illumina sequencing analysis by comparing sequencing data before and after peroxotungstate treatment.

Introduction

Epitranscriptome, which refers to the landscape of RNA chemical modifications, plays essential roles in post-transcriptional regulation, such as in translation, RNA degradation, and localization. 5-Methylcytidine (^{m5}C) is one of the major RNA modifications identified in noncoding RNA, rRNA, tRNA, and mRNA. It is known to stabilize tRNA secondary structure,[1] localize mRNA into the nucleus,[2] and regulate protein expression.[3] The ten-eleven translocation (TET) enzymes that oxidize 5-methylcytosine to 5-hydroxymethylcytosine in DNA[4] (^{5hm}C) can also oxidize ^{m5}C into 5-hydroxymethylcytidine in RNA (^{hm5}C).[5] Although the function of ^{5hm}C in DNA has been intensively investigated, the biological significance of ^{hm5}C is still largely unknown. ^{hm5}C has been shown to be enriched in *Drosophila melanogaster* mRNA, to increase mRNA translation, and to play a central role in *Drosophila* brain development,[6] even at the low levels found in cells (~100 ppm among modified C derivatives).[7-9] A recent study has suggested that ^{hm5}C contributes to the transcriptome flexibility required for mouse embryonic stem cell differentiation through mRNA destabilization.[10] To obtain precise information of ^{hm5}C distribution in a transcriptome-wide manner, sensitive and robust sequencing methods are demanded.

Base-resolution analysis of ^{5hm}C in genomic DNA has been achieved so far, in which a ^{5hm}C-selective chemical reaction[11] or enzymatic reaction[12] was applied and combined with bisulfite treatment. However, these bisulfite-based methods could not be applied to RNA ^{hm5}C because of the harsh conditions required, such as high temperature and wide pH fluctuations causing severe degradation of RNA. Several bisulfite-free methods have been developed for ^{5hm}C detection,[13-17] but some of these methods might be impossible to be applied to RNA ^{hm5}C possibly due to the backbone difference between DNA and RNA. The hydroxymethylcytosine RNA immunoprecipitation (hMeRIP) sequencing method using anti-^{5hm}C antibody has been developed to map ^{hm5}C in the *Drosophila* transcriptome. However, the resolution of hMeRIP-Seq is limited because of the nature of immunoprecipitation.[6]

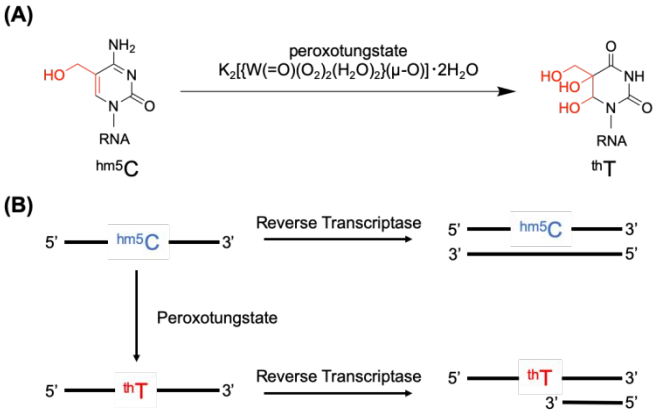


Fig. 1 ^{hm5}C detection strategy used in this study. (A) Peroxotungstate-mediated conversion of ^{hm5}C into thT. (B) Truncated cDNA production by reverse transcriptase arrest at the thT site.

In a previous study, we developed peroxotungstate-mediated selective conversion of ^{5hm}C in DNA to trihydroxylated thymine (thT) (Fig. 1A).[13, 18] To distinguish ^{5hm}C from canonical cytosine and 5-methylcytosine, we employed selective DNA cleavage at thT sites by hot piperidine treatment[18] or C-to-T transition at thT sites after the polymerase chain reaction (PCR) step because of the potential for thT to form hydrogen bonds with A.[13] The advantages of peroxotungstate-mediated oxidation are high selectivity to allylic alcohol structure in 5-hydroxymethylcytosine nucleobase and its mild reaction condition of neutral pH (pH 7.0). Our previous study showed that this reaction proceeded well in single-stranded DNA rather than double-stranded DNA, prompting us to apply the peroxotungstate-mediated oxidation to RNA, which usually exists as

[a] *Department of Chemistry and Biotechnology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[b] *Department of Biomolecular Engineering, Nagoya University, Furo-cho Chikusa-ku, Nagoya, Aichi 464-8603, Japan*
[c] *Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan*
†Electronic Supplementary Information (ESI) available. See
DOI: 10.1039/x0xx00000x

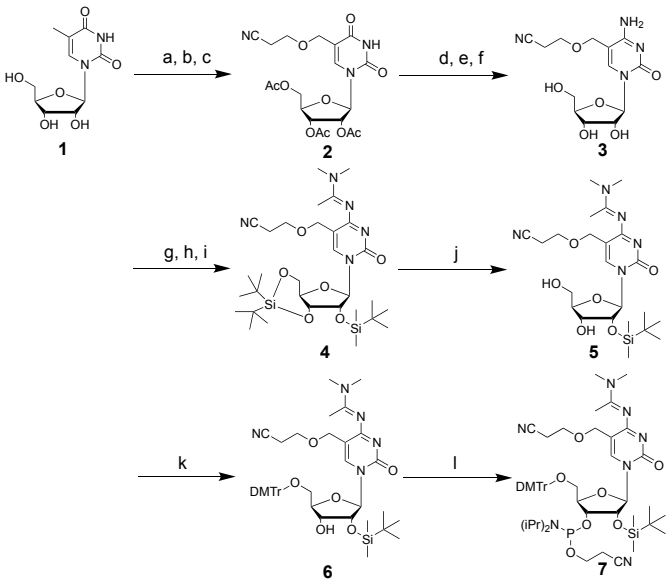
a single strand. In this study, we confirmed that ^{hm5}C in RNA sequences was efficiently converted to thT with negligeable damage to the RNA backbone, and that certain reverse transcriptases (RTases) stop at the thT site during cDNA synthesis to produce a truncated cDNA (Fig. 1B). The truncated sites could be identified by DNA sequencing as previously performed in miCLIP[19] and Pseudo-Seq.[20] As a result, we could provide a base-resolution analysis of ^{hm5}C, termed peroxotungstate-mediated truncation sequencing (WT-Seq), using the combination of the peroxotungstate-mediated oxidation and subsequent truncated cDNA production by RTase.

Results and discussion

Synthesis of ^{hm5}C phosphoramidite and ^{hm5}C-containing oligonucleotides

A synthetic route of ^{hm5}C phosphoramidite was initially developed to obtain ^{hm5}C-containing oligonucleotides (Scheme 1). 5-Methyluridine was used as a starting material and its ribose moiety was protected with a 2',3',5'-*O*-tri-acetyl protecting group. After a simple extraction process, the 5-methyl moiety was brominated via Wohl-Ziegler reaction with *N*-bromosuccinimide (NBS) in the presence of 2,2'-Azobis(isobutyronitrile) (AIBN) in benzene, followed by nucleophilic substitution reaction with 3-hydroxypropionitrile in *N*,*N*-dimethylformamide (DMF) to give compound **2**. Next, the C4 carbonyl group of **2** was activated through the reaction with 1,2,4-triazole in the presence of triethylamine and phosphorus oxychloride in acetonitrile, resulting in the triazole-activated derivative. The derivative was treated with aqueous ammonium hydroxide in dioxane for the aminolysis reaction at the activated C4 position to give a C4 amino group. Further treatment with aqueous ammonium in methanol deprotected the 2',3',5'-*O*-tri-acetyl protecting group, resulting in compound **3**. Compound **3** was then reacted with di-*tert*-butylsilyl (DTBS) ditriflate for protection of the 3'- and 5'-hydroxy groups, followed by protection of the 2'-hydroxy group with tert-butyldimethylsilyl (TBDMS) chloride. Subsequently, an amino group at the C4 position was protected with *N*,*N*-dimethylformamide dimethyl acetal to yield compound **4**. Cleavage of the 3'- and 5'-DTBS group with pyridine hydrofluoride in dichloromethane and pyridine gave compound **5**. Compound **5** was then transformed into the dimethoxytritylated compound **6** by dimethoxytrityl triflate. Finally, phosphitylation was executed with 2-cyanoethyl-*N*,*N*,*N'*,*N'*-tetraisopropylphosphoramidite in the presence of 1*H*-tetrazole in acetonitrile. Starting from 5-methyluridine, our route provides phosphoramidite **7** with 4% overall yield in 12 steps and with five chromatographic purifications; in total, 136 mg of phosphoramidite **7** was obtained in the course of this study.

Oligonucleotides containing ^{hm5}C (Table 1) were synthesized via the standard RNA solid-phase synthesis procedure. Coupling yields of the novel building block **7** were higher than 98%, estimated by monitoring the amount of released 4,4'-dimethoxytrityl cations from the coupled building block. Cleavage of the oligonucleotides from the solid support and their deprotection was performed using 28% ammonium hydroxide in water, followed by treatment with triethylamine trihydrofluoride in DMSO. The crude RNA oligomers, RNA1(^{hm5}C) and RNA2(^{hm5}C), were purified by reverse-phase HPLC under denaturing conditions (45 °C) (Fig. S1). The molecular weights of the purified RNAs were confirmed by MALDI-TOF MS.



Scheme 1. Synthesis of ^{hm5}C phosphoramidite (**7**). Reagents and conditions: (a) acetic anhydride, triethylamine, acetonitrile, r.t., 8 h; (b) NBS, AIBN, benzene, reflux, 4 h; (c) 3-hydroxypropionitrile, DMF, r.t., 26 h, 36% in three steps; (d) 1,2,4-triazole, phosphoryl chloride, triethylamine, acetonitrile, r.t., 5 h; (e) ammonium hydroxide, dioxane, r.t., 4 h; (f) ammonium hydroxide, methanol, r.t., overnight, 60% in three steps; (g) DTBS ditriflate, DMF, 4 °C, 3 h; (h) TBDMS chloride, imidazole, DMF, r.t., 1 day; (i) *N*,*N*-dimethylformamide dimethyl acetal, imidazole, DMF, r.t., 48 h, 73% in three steps; (j) pyridine hydrofluoride, pyridine, dichloromethane, 4 °C, 2 h, 48%; (k) 4,4'-dimethoxytrityl trifluotomethanesulfonate, pyridine, dichloromethane, r.t., 2.5 h, 84%; (l) 2-cyanoethyl-*N*,*N*,*N'*,*N'*-tetraisopropylphosphoramidite, 1*H*-tetrazole, acetonitrile, r.t., 2 h, 68%.

Table 1. RNA sequences used in this paper.[a]

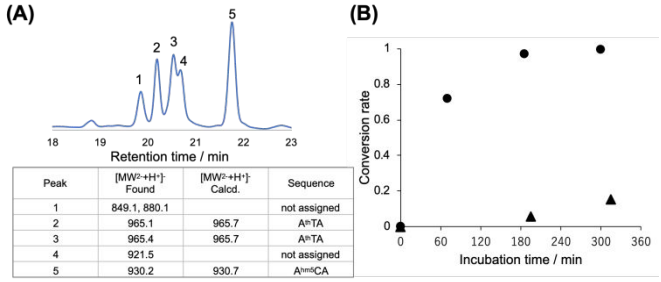
	Sequences (5' to 3')
RNA1(X)	A**X**A
RNA2(X)	CACU**X**GCUUCCUCCAGAUGA
RNA3(X)	GGGAGGUGAGAGUGAGAGUAUGUAUAGAAUUGAUAU**X**G AAAUGAGUAGGUGAUGGAAGUGGUAGGUAAGGGAA
RNA4(^{hm5}C)	GGGGUAGGAUGUGUGAUGAGAAA^{hm5}**C**UUAUAAGGAGUA AUGUGGGAUGUGUAGAGGAUUAGAUUAUGAGGUAGAU UGUGAUAUGAAGUGGAGAGUGAUGAGUGGUAAGGGAAU GAGUGGAAAUGA
RNA5(C)	GGGGUAUGAUGUAUGAGGAGAAA**C**UUAUGAGGAGUGAU GUGGGAUGUUGAGAAGAUUGAUUAAUGAGGAGUAUUG UGAUAGUAAGGUAGAGGUGGUAGUAGGUAAGGAGUAGA UGGGAAUGAA

[a] **X** in the sequences denotes ^{hm5}C, ^{m5}C, C, or thT. See Table S1 for the other nucleic acid sequences.

Peroxotungstate-mediated oxidation to ^{hm5}C-containing oligonucleotides

Peroxotungstate-mediated oxidation to the synthetic trinucleotide RNA1(^{hm5}C) (Table 1) was investigated using MALDI-TOF MS with HPLC monitoring. After 1 h incubation with 5 mM peroxotungstate at 50 °C, the HPLC peak of the starting RNA1(^{hm5}C) decreased by approximately 72%, and four new peaks appeared (Fig. 2A). MALDI-TOF MS revealed that the molecular weights of the two major peaks were identical to RNA1(thT) (5'-AthTA-3'), suggesting that the diastereoisomers of the thT were produced in similar yields (Fig. S2). The other products were presumably derived from hydrolyzed thT structures. Time-course HPLC analysis showed that more than 99%

of RNA1(^{hm5}C) reacted to generate RNA1(thT) after 5 h (Fig. 2B). On the other hand, tri-nucleotide RNA1(C) (5'-ACA-3') showed little change in HPLC spectra even after 5 h oxidation. 15% of RNA1(C) seemed to be degraded in 5 hours (Fig. 2B). These results suggest that ^{hm5}C nucleobase existing in RNA can be selectively converted to thT by the peroxotungstate treatment without severe degradation of RNA.



Peak	[MW²⁻+H⁺]⁻ Found	[MW²⁻+H⁺]⁻ Calcd.	Sequence
1	849.1, 880.1		not assigned
2	965.1	965.7	AthTA
3	965.4	965.7	AthTA
4	921.5		not assigned
5	930.2	930.7	A^{hm5}CA

Fig. 2 Reactions of RNA1(^{hm5}C) with peroxotungstate. (A) HPLC trace of RNA1(^{hm5}C) after 1-h incubation with peroxotungstate, showing several product and byproduct peaks (peaks 1–5). The product peaks (peaks 1–5) were analyzed using MALDI-TOF-MS. (B) Time-course study of peroxotungstate-mediated oxidation of trinucleotides RNA1(^{hm5}C) (circle) and RNA1(C) (triangle) to demonstrate conversion rate and specificity of the reaction.

Reverse transcription of oxidized ^{hm5}C-containing RNA.

We then investigated the effect of thT during cDNA synthesis by using oxidized and unoxidized oligonucleotide RNA2(^{hm5}C) (Table 1). After peroxotungstate treatment of RNA2(^{hm5}C), production of thT-containing RNA2(thT) was identified by MALDI-TOF MS (Fig. S3). Single-base primer extension assays were conducted with fluorescein-labeled primer and SuperScript III RTase. The template RNA-primer DNA2 duplex (Fig. S4A) was incubated with dATP, dGTP, dCTP, or dTTP in the presence of the RTase at 50 °C for 10 min. Denaturing PAGE analysis revealed that dGTP incorporation was clearly decreased after oxidation (Fig. S4B). On the other hand, the other dNTPs did not show any elongated bands. Next, we tested four commercially available RTases with primer extension in the presence of all dNTPs. Interestingly, all of the RTases showed truncations near the thT site, although two of them gave fully-extended bands (Fig. S5). Notably, SuperScript III RTase strongly induced truncation, while efficient full-length synthesis occurred on intact RNA2(^{hm5}C). To clarify whether the truncated cDNA production was caused by RTase arrest at the thT site but not by cutoff of template RNA, we analyzed RNA2(^{hm5}C) by denaturing PAGE before and after peroxotungstate treatment. As a result, the RNA2(^{hm5}C) was not cleaved after the oxidation reaction (Fig. S6). Then, the reverse transcription conditions of SuperScript III were further optimized under a range of temperatures, reagent concentrations, and RTase amounts to achieve the best truncation efficiency using a 14-nt primer to evaluate truncation itself or a 12-nt primer to observe overall reverse-transcription reaction (Fig. S7). The amount of enzyme largely contributed to the truncation efficiency, whereas neither the incubation temperature nor dNTPs or DTT concentration were significant for the truncation. The optimized conditions were applied to synthesized template RNA2(thT), demonstrating efficient production of a truncated cDNA (Fig. 3).

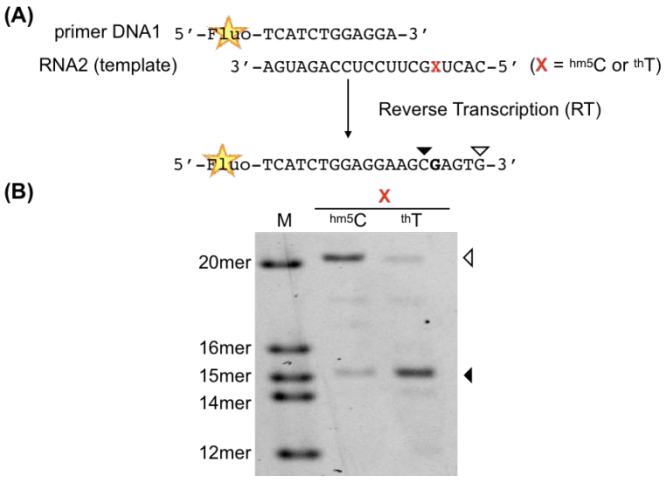


Fig. 3 Reverse transcription of non-oxidized or oxidized ^{hm5}C-containing RNA using SuperScript III RTase. (A) Sequences of synthesized RNA2(X) (X = ^{hm5}C or thT) and fluorescein-labeled primer DNA1. (B) Denaturing PAGE analysis of cDNA templated non-oxidized RNA2 (lane 2) and cDNA templated oxidized RNA2 (lane 3).

Sanger sequencing analysis of ^{hm5}C-containing RNA

Identification of truncation sites by sequencing analysis led to the detection of ^{hm5}C at a single-base resolution. We conducted Sanger sequencing to investigate whether a ^{hm5}C site in a 20-mer RNA2(^{hm5}C) could be detected by comparing data before and after peroxotungstate treatment. To identify the truncation site in cDNA, we followed the iCLIP procedure,[19, 21] in which cDNA was circularized and amplified by PCR (Fig. 4A). Although the sequencing data contained some noise signals probably due to imperfect RTase stalling (Fig. 3B), the sequencing result after 4 h peroxotungstate treatment showed 5-base truncation compared with that without treatment (Fig. 4B and C), indicating production of truncated cDNA. Therefore, we reasoned that WT-seq can provide a base-resolution detection of ^{hm5}C location by the combinatorial use of peroxotungstate treatment, SuperScript III RTase, and the iCLIP procedure.

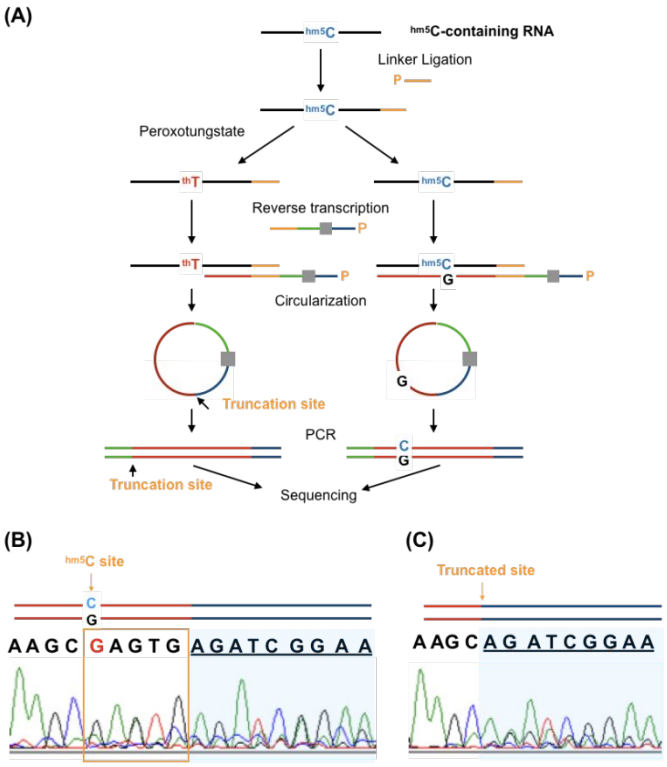


Fig. 4 Sanger-sequencing analysis of RNA2(^{hm5}C). (A) Scheme of WT-Seq. (B, C) Sanger sequencing results using (B) native RNA2(^{hm5}C) and (C) oxidized-RNA2(^{hm5}C). The orange region in RNA2(^{hm5}C) was removed in oxidized-RNA2(^{hm5}C), showing that the red-colored G site, a base-pair of ^{hm5}C, was the truncation site. The blue-colored region shows the sequence ligated to the 3'-end or truncation site in cDNA during circularization.

Illumina sequencing analysis of ^{hm5}C-containing RNA

We attempted to demonstrate the applicability of WT-seq to Illumina sequencing with low copies of ^{hm5}C-containing RNA in the presence of excess amounts of contaminating RNA. To increase the detection sensitivity, hMeRIP was tested for selective enrichment of ^{hm5}C-containing RNA. We first prepared three different in vitro transcribed RNAs, RNA3(X) (X = ^{hm5}C, ^{m5}C, or C) (Table 1). Denaturing PAGE analysis of immunoprecipitated C, ^{m5}C, or ^{hm5}C-containing RNA3 suggested that rabbit polyclonal anti-^{5hm}C antibody can selectively recognize ^{hm5}C-containing RNA3(^{hm5}C) (Fig. S8). We confirmed oxidation of RNA3(^{hm5}C) did not show clear degradation (Fig. S9).

We then conducted WT-seq using the Illumina sequencing platform to detect ^{hm5}C in in vitro transcribed RNA4(^{hm5}C) (Table 1), which was mixed with nonmodified RNA5(C) for evaluation of hMeRIP efficiency. First, 20 ng of a mixture of RNAs containing RNA4(^{hm5}C) and RNA5(C) was mixed with 50 μg of fragmented Poly(A)$^+$ RNA, isolated from *Drosophila* S2 cells. The mixed RNA was immunoprecipitated by hMeRIP, oxidized with peroxotungstate, and converted to a cDNA library according to the iCLIP procedure.[19, 21] MiSeq analysis of the cDNA library confirmed that after immunoprecipitation, the ratio of read number mapped to RNA4(^{hm5}C) increased compared with RNA5(C), showing ^{hm5}C-selective enrichment of RNA (Fig. 5A). After oxidation, the ^{hm5}C position in oxidized-RNA4(^{hm5}C) was detected by a peak value calculation using Pseudo-Seq methods[20] (Fig. 5B and S10). Although the truncation in cDNA of RNA3 was not complete, only the ^{hm5}C site showed a high peak value, demonstrating that WT-seq has potential to selectively detect ^{hm5}C. These results suggest that ^{hm5}C sites in low copy numbers of ^{hm5}C-containing RNAs derived from cells might be detected at a base-resolution level when WT-seq is combined with hMeRIP.

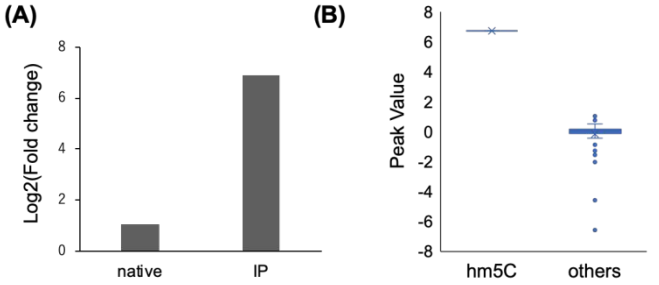


Fig. 5 MiSeq analysis of RNA4(^{hm5}C). (A) Enrichment analysis of the hMeRIP procedure. The ratio of ^{hm5}C-containing RNA4(^{hm5}C) (black boxes) and C-containing RNA5(C) (light gray stripes) was evaluated before (native) and after immunoprecipitation (IP). (B) Truncation efficiency at ^{hm5}C site or other sites in oxidized-RNA4(^{hm5}C). The data were visualized as peak values: 6.564 at the ^{hm5}C site, 0.034 (median) at the others with variance 0.904. The peak values of 87 bases (each base from 12th to 99th except hm5C) were plotted in the "others" column.

Conclusions

In this study, we demonstrated that one-step chemical conversion of ^{hm5}C to thT leads to a new base-resolution analysis of ^{hm5}C. The truncation in RTase-mediated cDNA synthesis at the thT site was confirmed by primer extension analysis and Sanger sequencing. WT-seq using Illumina sequencing analysis clearly detected the ^{hm5}C site in in vitro transcribed RNA, even in the presence of excess amounts of cell-derived RNA, even though a previous study using peroxotungstate was not applied to the Illumina sequencing platform.[22] When combined with hMeRIP, our method could detect the ^{hm5}C location in in vitro transcribed RNA at a base-resolution in the presence of cell-extracted RNA for the first time. Although efficiency of WT-seq needs to be improved for transcriptome-wide ^{m5}C detection, WT-seq may become more useful for the identification of ^{hm5}C distribution in combination with more efficient ^{hm5}C selective enrichment strategies and more rigorously optimized RTases.

Experimental section

General methods and materials

All reactions involving moisture- or air-sensitive reagents were carried out under an argon atmosphere. Reagents were purchased from commercial suppliers. Thin layer chromatography was performed on TLC Silica gel 60 F254 plates (Merck Millipore, 1.05554.0001). Column chromatography was carried out on a spherical silica gel 60 (Kanto Chemical, 37564-84). All NMR experiments were recorded on a 600 MHz instrument. ^{1}H NMR chemical shifts (δ) were referenced to TMS (0 ppm) as an internal standard. ^{13}C NMR chemical shifts (δ) were referenced to $CDCl_3$ (77.0 ppm) or DMSO-d_6 (39.5 ppm). Assignments were based on ^{1}H, and ^{1}H-COSY experiments. Mass spectrometric analysis (HRMS) of nucleoside derivatives were performed on an ESI TOF device. MA analysis of oligonucleotides were performed on a MALDI-TOF MS device.

Synthesis of ^{hm5}C phosphoramidite.

2',3',5'-Tri-*O*-acetyl-5-(2-cyanoethoxy)methyluridine (2). 5-Methyluridine (25.0 g, 96.8 mmol) was suspended in acetonitrile (192.4 mL). Triethylamine (81.0 mL, 580.8 mmol) and acetic anhydride (36.6 mL, 387.2 mmol) were added to the suspension. The mixture was stirred at room temperature for 8 h. After concentrating, ethyl acetate (300 mL) and brine (300 mL) were added to the residue. The product was extracted to the organic layer. The organic layer was washed with saturated sodium chloride (300 mL) twice, dried over magnesium sulfate, and filtered. The solvent was removed by evaporation, and the residue was dried under reduced pressure. *N*-Bromosuccinimide (18.9 g, 106.5 mmol) and benzene (192.4 mL) were added to the residue. The mixture was stirred and repeatedly degassed by injection of argon gas. After addition of AIBN (1.59 g, 9.68 mmol), the mixture was refluxed for 4 h. The reaction mixture was cooled and the precipitate was removed by filtration. The solvent was removed by evaporation and the residue was dried under reduced pressure. *N*,*N*-Dimethylformamide (DMF, 96.8 mL) and 3-hydroxypropionitrile (39.7 mL,580.8 mmol) were added to the residue. The mixture was stirred at room temperature for 26 h. Ethyl acetate (300 mL) and brine (300 mL) were added to the reaction mixture. The product was extracted to the organic layer. The organic layer was washed with brine (300 mL) twice, dried over magnesium sulfate, and filtered. The solvent in the organic phase was removed by evaporation. The product was purified by silica gel column chromatography (6:4 - 3:7 hexane/ethyl acetate). The product **2** was obtained as a white foam (15.7 g, 36%): R_f 0.3 (hexane/ethyl acetate 3:7); ^{1}H NMR ($CDCl_3$): δ 9.71 (1H, s), 7.53 (1H, s), 6.02 (1H, d), 5.42 (1H, t), 5.36 (1H, d), 4.35–4.40 (5H, m), 3.77 (2H, t), 2.66 (2H, t), 2.05–2.17 (9H, m). ^{13}C NMR ($CDCl_3$): δ 170.42, 169.81, 169.73, 162,73, 162.55, 150.19, 137.89, 117.94, 112.01, 88.14, 80.13, 72.79, 70.43, 65.50, 65.13, 63.33, 20.86, 20.59, 20.54, 20.51, 18.90. HRMS (ESI) calcd for $C_{19}H_{23}N_3O_{10}Na^+$ ([M + Na]$^+$) 476.1276, found 476.1289.

5-(2-cyanoethoxy)methylcytidine (3). Nucleoside **2** (14.96 g, 33 mmol) and 1,2,4-triazole (20.50 g, 297 mmol) were dissolved in acetonitrile (583 mL). Triethylamine (46.06 mL, 330 mmol) and phosphorous oxychloride (5.26 mL, 55.8 mmol) were added to the solution with stirring. The reaction mixture was stirred at room temperature for 5 h. The solvent was evaporated, and ethyl acetate (400 mL) and brine (400 mL) were added to the residue. The product was extracted to the organic layer. The organic layer was washed with brine twice and dried over magnesium sulfate. The solvent was removed by evaporation, and the residue was dried under reduced pressure. 1,4-Dioxane (735 mL) and 28% ammonia (110 mL) were added to the residue. The mixture was stirred at room temperature for 4 h. The solvent was removed by evaporation. Methanol (110 mL) and 28% ammonia (110 mL) were added to the residue. The mixture was stirred at room temperature overnight. The solvent was removed by evaporation. The product was dissolved in small amount of methanol, precipitated by addition of ethanol and washed by ethanol. The product **3** was obtained as a white solid (6.5 g, 60%): ^{1}H NMR (DMSO-d_6): δ 7.98 (1H, s), 7.48 (1H, s), 6.72 (1H, s), 5.75 (1H, d), 5.14 (1H, t), 5.02 (1H, d), 4.22 (2H, q), 3.94–3.98 (2H, m), 3.82–3.84 (1H, m), 3.68–3.71 (1H, m), 3.54–3.59 (3H, m), 2.78 (2H, t). ^{13}C NMR (DMSO-d_6): δ 164.50, 155.23, 141.90, 119.40, 101.79, 89.30, 84.15, 74.05, 69.28, 66.21, 63.97, 60.53, 18.03. HRMS (ESI) calcd for $C_{13}H_{18}N_4O_6Na^+$ ([M + Na]$^+$) 349.1124, found 349.1165.

2'-*O*-tert-butyldimethylsilyl-4-*N*,*N*-di-n-butylformamidine-3',5'-*O*-di-*tert*-butylsilyl-5-(2-cyanoethoxy)methylcytidine (4). Nucleoside **3** (3.0 g, 9.2 mmol) was dissolved in DMF (18.4 mL). The solution was kept on ice with stirring. Di-*tert*-butylsilyl bis(trifluoromethanesulfonate) (3.3 mL, 10.1 mmol) was added to the solution dropwise over 15 min. The reaction mixture was stirred for 3 h on ice. Imidazole (3.13 g, 46.0 mmol) and *tert*-butyldimethyl chlorosilane (1.66 g, 11.0 mmol) were added. In 5 min, the solution was kept stirred at room temperature for 1 day. Ethyl acetate (200 mL) and brine (200 mL) were added to the reaction mixture. The product was extracted to the organic layer. The organic layer was washed with brine (200 mL) twice, dried over magnesium sulfate, and filtered. The solvent in the organic phase was removed by evaporation. To the pellet, DMF (42 mL) and *N*,*N*-dimethylformamide dimethylacetal (10.8 mL were added. The solution was stirred at room temperature overnight. Ethyl acetate (200 mL) and brine (200 mL) were added to the reaction mixture. The product was extracted to the organic layer. The organic layer was washed with brine (200 mL) twice, dried over magnesium sulfate, and filtered. The solvent in the organic phase was removed by evaporation. The product was purified by silica gel column chromatography (100:0 – 100:1 chloroform/methanol). The product **4** was obtained as a white foam (4.3 g, 73%): R_f 0.6 (chloroform/methanol 10:1); ^{1}H NMR ($CDCl_3$): δ 8.73 (1H, s), 7.48 (1H, s), 5.72 (1H, s), 4.34–4.47 (3H, m), 4.24 (1H, d), 4.14–4.18 (1H, m), 4.00 (1H, t), 3.78–3.83 (1H, m), 3.65–3.69 (2H, m), 3.09 (3H, s), 3.06 (3H, s), 2.56 (2H, t), 0.81–0.97 (27H, m), 0.18 (3H, s), 0.09 (3H, s). ^{13}C NMR ($CDCl_3$): δ 169.77, 158.61, 155.97, 139.62, 118.07, 110.53, 94.62, 76.07, 75.78, 74.90, 68.16, 67.09, 65.50, 41.71, 35.49, 27.82, 27.35, 26.26, 23.10, 20.64, 19.25, 18.54, 18.35, –4.61. HRMS (ESI) calcd for $C_{28}H_{48}N_5O_6Si2^+$ ([M + H]$^+$) 636.3613, found 636.3635.

2'-*O*-tert-butyldimethylsilyl-4-*N*,*N*-di-*n*-butylformamidine-5-(2-cyanoethoxy)methylcytidine (5). Nucleoside **4** (172.4 mg, 0.27 mmol) was dissolved in dichloromethane (1.08 mL) and kept on ice. Hydrogen fluoride-pyridine complex (hydrogen fluoride ~70%, pyridine ~30%) (29 μL) in pyridine (173.5 μL) was added to the solution dropwise. The reaction mixture was stirred on ice for 2 h. Dichloromethane and saturated sodium bicarbonate were added and the product was extracted to the organic layer. The organic layer was dried over potassium sulfate, filtered and evaporated. The product was purified by silica gel column chromatography (100:0 - 50:1 chloroform/methanol). The product **5** was obtained as a white form (62.2 mg, 48%): R_f 0.4 (chloroform/methanol 10:1); ^{1}H NMR ($CDCl_3$): δ 8.68 (1H, s), 7.85 (1H, s), 5.52 (1H, d), 4.64 (1H, t), 4.37 (2H, dd), 4.15 (1H, s), 4.06 (1H, s), 3.79 (2H, dd), 3.65 (2H, t), 3.08 (3H, s), 3.06 (3H, s), 2.58 (2H, t), 0.81 (9H, s), 0.00 (3H, s), –0.01 (3H, s). ^{13}C NMR ($CDCl_3$): δ 170.06, 158.59, 156.56, 143.29, 118.62, 111.05, 94.81, 85.94, 74.64, 70.80, 66.84, 65.32, 62.09, 41.77, 35.57, 26.08, 19.20, 18.34, –4.61. HRMS (ESI) calcd for $C_{22}H_{38}N_5O_6Si^+$ ([M + H]$^+$) 496.2591, found 496.2584.

2'-*O*-tert-butyldimethylsilyl-4-*N*,*N*-di-*n*-butylformamidine-5-(2-cyanoethoxy)methyl-5'-*O*-(4,4'-dimethoxytrityl)cytidine (6). Silver triflate (250.5 mg, 0.975 mmol) was added to 4,4'-dimethoxytritylchloride (330.4 g, 0.975 mmol) in dichloromethane (3.25 mL) at room temperature to prepare 4,4'-dimethoxytrityl triflate salt. The mixture was stirred for 1.5 h. The supernatant 4,4'-dimethoxytrityl triflate salt solution (650 μL,0.195 mmol) was added to a solution of nucleoside **5** (64.5 mg,0.13 mmol) in dichloromethane-pyridine (2.6 mL, 1:1, v/v) at room temperature and stirred for 2.5 h. The reaction was diluted with dichloromethane (50 mL) and washed with saturated aqueous sodium bicarbonate solution (50 mL) twice. The aqueous layer was back-extracted with dichloromethane (50 mL). The organic layer was dried over sodium

sulfate , filtered and evaporated. The residue was purified by silica gel column chromatography (200:1 - 100:1 chloroform/methanol with 4 % triethylamine). The product **6** was obtained as a white foam (87.3 mg, 84 %): R_f 0.5 (chloroform/methanol 10:1); ^{1}H NMR ($CDCl_3$): δ 8.84 (1H, s), 8.05 (1H, s), 6.85–7.48 (13H, m), 6.07 (1H, d), 4.40 (1H, dd), 4.28 (1H, d), 4.12–4.15 (2H, m), 3.81 (6H, s), 3.72 (1H, d), 3.60 (1H, dd), 3.36 (1H, dd), 3.29 (2H, ddd), 3.16 (3H, s), 3.14 (3H, s), 2.12 (2H, ddd), 0.93 (9H, s), 0.27 (3H, s), 0.17 (3H, s). ^{13}C NMR ($CDCl_3$): δ 170.10, 158.74, 158.71, 158.43, 156.18, 144.73, 141.03, 135.69, 135.60, 130.33, 130.24, 128.32, 128.07, 127.09, 117.80, 113.35, 113.34, 110.41, 90.01, 86.62, 83.24, 70.14, 66.44, 64.94, 62.57, 55.35, 46.32, 41.42, 35.23, 25.95, 25.91, 18.19, 18.07, 11.63, 11.59, -4.61. HRMS (ESI) calcd for $C_{43}H_{56}N_5O_8Si^+$ ([M + H]$^+$) 798.3898, found 798.3880.

2'-*O*-tert-butyldimethylsilyl-4-*N*,*N*-di-*n*-butylformamidine-5-(2-cyanoethoxy)methyl-5'-*O*-(4,4'-dimethoxytrityl) cytidine-3'-*O*-2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite (7). Nucleoside **6** (161.6 mg, 0.20 mmol) and 1*H*-tetrazole (28.0 mg, 0.40 mmol) were dissolved in degassed acetonitrile (2.86 mL). To the mixture, 2-cyanoethyl *N*,*N*,*N'*,*N'*-tetraisopropylphosphoramidite (191 μL, 0.60 mmol) was added. The reaction was stirred at room temperature for 2 h. After concentration, ethyl acetate (20 mL) was added to the residue, washed with saturated aqueous sodium bicarbonate solution (20 mL) and with brine twice. The organic layer was dried over magnesium sulfate, filtered and evaporated. The residue was purified by silica gel column chromatography (1:1 - 0:1 hexane/ethyl acetate with 4 % triethylamine). The product **7** was obtained as a white foam (136.3 mg, 68 %): R_f 0.2 (ethyl acetate); ^{1}H NMR ($CDCl_3$): δ 8.82 (1H, d), 8.06 (1H, d), 8.82 (1H, d), 6.82–7.49 (13H, m), 6.42 (1H, s), 4.19–4.22 (2H, m), 4.12 (2H, d), 3.79 (6H, d), 3.47–3.56 (7H, m), 3.12–3.14 (6H, d), 2.74–2.78 (4H, m), 1.27 (18H, d), 0.89 (9H, d). ^{31}P NMR ($CDCl_3$): δ 150.65. HRMS (ESI) calcd for $C_{52}H_{73}N_7O_9PSi^+$ ([M + H]$^+$) 998.4977, found 998.4997.

Synthesis of RNA Oligonucleotides.

Phosphoramidite chemistry was applied for automated RNA strand elongation using 2'-*O*-TBS nucleoside building blocks (Glen Research) and a polysterene support (Glen Research) in combination with synthesized ^{hm5}C phosphoramidte. All oligonucleotides were synthesized on an NTS H-6 DNA/RNA synthesizer using standard procedures. The solid support after synthesis was removed and treated with ammonium hydroxide (1 mL, 28%) for 6 h at 75 °C. The supernatant was evaporated to dryness and heated with triethylamine trihydrofluoride (125 μL) in dimethyl sulfoxide (100 μL) at 65 °C for 2.5 h. The reaction was quenched by the addition of triethylammonium acetate buffer (1.5 mL, 2M, pH 7.4). The crude RNA in the solution was purified by HPLC on a 5C$_{18}$-MS-II column at 45 °C; 0 %–25 % acetonitrile/100 mM triethylammonium acetate buffer over 30 min. The purified RNA was analyzed by MALDI-TOF MS. RNA sequences synthesized in this study can be found in Table 1 and S1.

Time-course study of peroxotungstate-mediated oxidation.

RNA1(^{hm5}C) was mixed with phosphate buffer (50 mM phosphate, 100 mM sodium chloride, pH 7.0) and peroxotungstate in water (5 mM). The mixture was incubated at 50 °C for 0–7 h, followed by HPLC analysis using a 5C$_{18}$-MS-II column at 45 °C; 0 %–25 % acetonitrile/100 mM triethylammonium acetate buffer over 30 minutes. The purified RNA was analyzed by MALDI-TOF MS.

***In vitro* transcription of ^{hm5}C-containing RNA.**

In vitro transcription was performed using MEGAscript T7 Transcription Kit (Thermo Fisher Scientific; AM1333), according to the manufacturer's instructions. For m5C and ^{hm5}C-containing transcripts, ribo-CTP nucleotides were replaced with ribo-mCTP or ribo-hmCTP (TriLink Biotechnologies). The transcription was conducted at 37 °C for 17 h. After denaturing PAGE analysis of products, the transcripts were purified by phenol-chloroform-isoamyl alcohol extraction and lithium chloride precipitation. The precipitates were washed with 70 % ethanol and dried up. The RNA was dissolved in RNase-free water and analyzed by Nanodrop. RNA sequences transcribed in this study can be found in Table 1.

Primer extension assay.

Synthesized or transcribed RNA was denatured with 3 M urea in phosphate buffer (50 mM phosphate, 100 mM sodium chloride, pH 7.0) at 65 °C for 5 min, followed by incubation on ice for 5 min. To the solution, 5 mM peroxotungstate was added and heated at 50 °C for 5 h. For synthesized RNA, oxidized product was purified by HPLC. For transcribed RNA, the RNA in the reaction was purified with micro-Biospin column 6. The oxidized RNA was mixed with fluorescein-labelled RT primer (Table S1) and denatured at 65 °C for 5 min, followed by incubation on ice for 5 min. To the solution, 1× reverse transcription buffer, dNTP (30 μM), and SuperScript III reverse transcriptase were added and reacted at 37 °C for 20 min. The reaction was quenched by the addition of denaturing PAGE loading buffer and heated at 90 °C for 5 min. The solution was analyzed by denaturing PAGE.

Oxidation, Reverse Transcription and Sanger Sequencing of truncated cDNA.

Synthesized RNA2 (100 ng) and adenylated linker DNA (Bioo Scientific, #510201, 1 μM) were ligated by T4 RNA ligase 2 (truncated) (200 U) in the 1× T4 RNA ligase 2 buffer with 15% PEG8000 and SUPERase•In (20 U) at 16 °C overnight, followed by ethanol precipitation.

The precipitated RNA was mixed with phosphate buffer (50 mM phosphate, 100 mM sodium chloride, 3.5 M urea, pH 7.0) and denatured at 80 °C for 90 s. To the solution, 5 mM peroxotungstate was added and incubated at 50 ° for 5 h. The oxidized RNA was analyzed by 15 % denaturing PAGE and purified by gel extraction and subsequent ethanol precipitation.

The RNA was dissolved in RNase-free water and reverse transcription primer (Table S1) was added. The mixture was heated at 65 °C for 5 min and cooled on ice for 5 min. The solution was incubated with SuperScript III reverse transcriptase (100 U) at 50 °C for 20 min after the addition of 1× provided buffer, dNTP mix (30 μM), and DTT (2 mM). After reverse transcription, the reaction was quenched by the addition of denaturing PAGE loading buffer and analyzed by 8 % denaturing PAGE, followed by purification via gel extraction and ethanol precipitation.

The reverse transcribed cDNA was circularized by the incubation at 60 °C for 2 h with 1× circ ligase buffer, 50 μM ATP, 2.5 mM manganese (II) chloride, and 50 U CircLigase ssDNA Ligase. The reaction was stopped by heating at 80 °C for 10 min, and purified by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation.

The circularized cDNA was used as a template for PCR. One fourth amount of cDNA was mixed with 1× Phusion HF buffer, 0.2 mM dNTP mix, 0.5 μM forward and reverse primer (Table S1) and 0.4 U Phusion

HF DNA polymerase, and amplified by PCR. The amplicons were analyzed by 10% native PAGE, followed by purification via gel extraction and ethanol precipitation. The purified DNA was sequenced at Genewiz.

Immunoprecipitation of ^{hm5}C-containing oligonucleotides.

Transcribed RNA (1 µg) was denatured by heating at 65 °C for 5 min, followed by incubation on ice for 5 min. The denatured RNA was incubated with 2 µg of anti-5hmC antibody (Active Motif; 39679) in 1× IP buffer (10 mM Tris-HCl pH 7.5, 150 mM sodium chloride, 0.5 % Igepal CA-630, RNasin 400 U/ml, and RVC 2 mM) supplemented with protease inhibitors cOmplete Mini (Roche) at 4 °C overnight. The samples were incubated with 10 µL equilibrated Dynabeads Protein G (Life Technologies) at 4 °C for 2.5 h, washed three-times with 50 µL 1× IP buffer for 5 minutes, and eluted by the addition of 50 µL TriPure Isolation Reagent followed by RNA extraction according to the manufacturer's instructions. The precipitated RNA was analyzed by denaturing PAGE.

Cell culture.

Drosophila S2 cells were cultured in Schneider's Drosophila medium supplemented with 10% fetal bovine serum (FBS) and 1× Antibiotic-Antimycotic (Gibco; 15240062). Cells were grown at 27 °C under standard laboratory conditions.

Extraction of total and poly(A)$^+$ RNA.

Drosophila total RNA was isolated by TriPure Isolation Reagent (Sigma-Aldrich; 11667157001). Poly(A)+ RNA was isolated from 5 mg total RNA using oligo dT cellulose beads (NEB; S1408S).

WT-Seq of ^{hm5}C-containing oligonucleotides mixed with extracted RNA.

Extracted Poly(A)+ RNA (50 µg) was fragmented in 1x Fragmentation buffer (10 mM zinc chloride and 30 mM Tris•HCl pH 7.0) at 95 °C for 100 s and quenched by the addition of 50 mM EDTA (pH 8.0). The fragmented RNA was purified by RNA Subcellular Isolation Kit.

The fragmented RNA was mixed with 20 ng of spike-in RNA (Table S1) and denatured at 65 °C for 5 min, followed by incubation on ice for 5 min. To the solution, 1× IP buffer, RNase•In, and 25 µg of anti-^{5hm}C antibody (Active Motif, 39679) were added and incubated at 4 °C overnight. The antibody-RNA conjugates were bound to 200 µL of equilibrated Dynabeads Protein G at 4 °C for 2.5 h and immunoprecipitated RNA was purified as mentioned above.

The immunoprecipitated RNA was denatured at 80 °C for 90 s and dephosphorylated using T4 Polynucleotide Kinase (NEB, M0201A, 10U) in 1× T4 polynucleotide kinase buffer supplemented with SUPERase•In (20 U) at 37 °C for 1 h. The reaction was quenched at 70 °C for 10 min, and the dephosphorylated RNA was purified by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation.

The dephosphorylated RNA and adenylated linker DNA (Bioo Scientific, #510201, 1 µM) were ligated by T4 RNA ligase 2 (truncated, K227Q) (200 U) in the 1× T4 RNA ligase 2 buffer with 15% PEG8000 and SUPERase•In (20 U) at 16 °C overnight, followed by ethanol precipitation.

The linker-ligated RNA was mixed with phosphate buffer (50 mM phosphate, 100 mM sodium chloride, 3.5 M urea, pH 7.0) and denatured at 80 °C for 90 s. To the solution, 5 mM peroxotungstate was added and incubated at 50 °C for 5 h. The oxidized RNA was analyzed by 5 % denaturing PAGE and purified by gel extraction and subsequent ethanol precipitation.

The RNA was dissolved in RNase-free water and reverse transcription primer (Table S1, 0.4 µM) was added. The mixture was heated at 65 °C for 5 min and cooled on ice for 5 min. The solution was incubated with SuperScript III reverse transcriptase (100 U) at 50 °C for 1 h after the addition of 1× provided buffer, dNTP mix (0.5 mM), and DTT (2 mM). After reverse transcription, the reaction was quenched by the addition of denaturing PAGE loading buffer and analyzed by 5 % denaturing PAGE, followed by purification via gel extraction and ethanol precipitation.

The reverse transcribed cDNA was circularized by the incubation at 60 °C for 16 h with 1× circ ligase II buffer, 1 M Betaine, 2.5 mM manganese (II) chloride, and 50 U CircLigase II ssDNA Ligase. The reaction was stopped by heating at 80 °C for 10 min, and purified by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation.

The circularized cDNA was used as a template for PCR. One fourth amount of cDNA was mixed with 1× Phusion HF buffer, 0.2 mM dNTP mix, 0.5 µM forward and reverse primer (Table S1) and 0.4 U Phusion HF DNA polymerase, and amplified by PCR. The amplicons were analyzed by 10% native PAGE, followed by purification via gel extraction and ethanol precipitation. The purified DNA was sequenced on an Illumina MiSeq (Paired-end, 75bpx2).

Read Trimming and Mapping.

Adapter sequence (5'-ATTGATGGTGCCTACAG-3') was removed from the second paired-end sequencing read (read 2) using Cutadapt, and both of the first paired-end sequencing read (read 1) and read 2 were processed by Fastq with the following parameters (-q 15 -n 10 -t 1 -T 1 -l 20). The reads were aligned to the drosophila reference mRNA (refMrna, dm6) and spiked oligonucleotide sequences using Bowtie 2 with the options (-a -L 15 -N 1 --local). PCR duplicates were removed with the Genome Analysis Toolkit program MarkDuplicates (Picard).

Identification of ^{hm5}C sites.

For a library pair with or without treatment of peroxotungstate-mediated oxidation, one library was scaled to the size of the other library. Peak values were calculated for each base position following Pseudo-Seq procedure[20].

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI 18H03931, 18H05504, and 19K22245 (A.O.), 18K05313, 19H05287, and 20H04704 and JST PRESTO JPMJPR19K6 (G.H.).

Notes and references

1. Y. Motorinand M. Helm, *Biochemistry,* 2010, **49**, 4934-4944.
2. 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. B. Jiang, Y. L. Zhao, H. L. Wang, Y. P. Sunand Y. G. Yang, *Cell. Res.,* 2017, **27**, 606-625.
3. W. Wang, *Wiley Interdiscip. Rev. RNA.,* 2016, **7**, 838-842.
4. X. Wuand Y. Zhang, *Nat. Rev. Genet.,* 2017, **18**, 517-534.
5. L. Fu, C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S. G. Jin, L. J. Niedernhofer, G. P. Pfeifer, G. L. Xuand Y. Wang, *J. Am. Chem. Soc.,* 2014, **136**, 11582-11585.
6. B. Delatte, F. Wang, L. V. Ngoc, E. Collignon, E. Bonvin, R. Deplus, E. Calonne, B. Hassabi, P. Putmans, S. Awe, C. Wetzel, J. Kreher, R. Soin, C. Creppe, P. A. Limbach, C. Gueydan, V. Kruys, A. Brehm, S. Minakhina, M. Defrance, R. Stewardand F. Fuks, *Science* 2016, **351**, 282-285.
7. W. Huang, M. D. Lan, C. B. Qi, S. J. Zheng, S. Z. Wei, B. F. Yuanand Y. Q. Feng, *Chem. Sci.,* 2016, **7**, 5495-5502.
8. S. M. Huber, P. van Delft, L. Mendil, M. Bachman, K. Smollett, F. Werner, E. A. Miskaand S. Balasubramanian, *ChemBioChem,* 2015, **16**, 752-755.
9. H. Y. Zhang, J. Xiong, B. L. Qi, Y. Q. Fengand B. F. Yuan, *Chem. Commun.,* 2016, **52**, 737-740.
10. J. Lan, N. Rajan, M. Bizet, A. Penning, N. K. Singh, D. Guallar, E. Calonne, A. Li Greci, E. Bonvin, R. Deplus, P. J. Hsu, S. Nachtergaele, C. Ma, R. Song, A. Fuentes-Iglesias, B. Hassabi, P. Putmans, F. Mies, G. Menschaert, J. J. L. Wong, J. Wang, M. Fidalgo, B. Yuanand F. Fuks, *Nat. Commun.,* 2020, **11**, 4956.
11. M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reikand S. Balasubramanian, *Science* 2012, **336**, 934-937.
12. 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. Renand C. He, *Cell* 2012, **149**, 1368-1380.
13. G. Hayashi, K. Koyama, H. Shiota, A. Kamio, T. Umeda, G. Nagae, H. Aburataniand A. Okamoto, *J. Am. Chem. Soc.,* 2016, **138**, 14178-14181.
14. Y. Liu, Z. Hu, J. Cheng, P. Siejka-Zielinska, J. Chen, M. Inoue, A. A. Ahmedand C. X. Song, *Nat. Commun.,* 2021, **12**, 618.
15. E. K. Schutsky, J. E. DeNizio, P. Hu, M. Y. Liu, C. S. Nabel, E. B. Fabyanic, Y. Hwang, F. D. Bushman, H. Wuand R. M. Kohli, *Nat. Biotechnol.,* 2018. **36**, 1083-1090.
16. Z. Sun, J. Terragni, J. G. Borgaro, Y. Liu, L. Yu, S. Guan, H. Wang, D. Sun, X. Cheng, Z. Zhu, S. Pradhanand Y. Zheng, *Cell. Rep.,* 2013, **3**, 567-576.
17. H. Zeng, B. He, B. Xia, D. Bai, X. Lu, J. Cai, L. Chen, A. Zhou, C. Zhu, H. Meng, Y. Gao, H. Guo, C. He, Q. Daiand C. Yi, *J. Am. Chem. Soc.,* 2018, **140**, 13190-13194.
18. A. Okamoto, K. Sugizaki, A. Nakamura, H. Yanagisawaand S. Ikeda, *Chem. Commun.,* 2011, **47**, 11231-11233.
19. B. Linder, A. V. Grozhik, A. O. Olarerin-George, C. Meydan, C. E. Masonand S. R. Jaffrey, *Nature Methods* 2015, **12**, 767-772.
20. T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoliand W. V. Gilbert, *Nature* 2014, **515**, 143-146.
21. J. Konig, K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombeand J. Ule, *Nat. Struct. Mol. Biol.,* 2010, **17**, 909-915.
22. F. Yuan, Y. Bi, P. Siejka-Zielinska, Y. L. Zhou, X. X. Zhangand C. X. Song, *Chem. Commun.,* 2019, **55**, 2328-2331.