A Photo-triggered Wittig-Cyclization for Detecting and Sequencing 5-Formylcytidine in RNA

Xiao-Yang Jin,[1,4] Zu-Rui Huang,[2,4] Li-Jun Xie,[1] Li Liu,[1,4] Da-Li Han,[2,4,*] Liang Cheng[1,3,4,*]

[1] Beijing National Laboratory for Molecular Sciences (BNLMS), CAS Key Laboratory of Molecular Recognition and Function, CAS Research/Education Center for Excellence in Molecular Sciences, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China

[2] China National Center for Bioinformation, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

[3] Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

[4] University of Chinese Academy of Sciences, Beijing 100049, China

* chengl@iccas.ac.cn (L.C); handl@big.ac.cn (D.L.H.)

ABSTRACT

The simultaneous detection of a specific post-transcriptional modification that decorates the natural RNA can provide comprehensive information to better understand cellular and molecular functions of this variation. 5-Formylcytidine f5C is one of the natural nucleotides that was discovered in the tRNA of many species during the oxidative demethylation of 5-methylcytidine m5C. Despite the evident importance of f5C in playing an important role in gene expression regulation, little is known about the exact amount and positions of this modified cytidine in RNAs, as well as its exact functions. To capture this type of information, we performed a site-specific dual functionalization of f5C with semi-stabilized ylide cyanomethylene triphenylphosphorane that can simultaneously seizure the C5-formyl and N4-amino groups in the cytidine residue under light irradiation. The photochemical labelling with f5C-containing oligonucleotides imparts a high selectivity towards f5C and allows distinction from structurally similar 5-formyluridine f5U. We implemented a detection strategy of f5C modification in RNA based on the fluorescence signal of the cyclization product 4,5-pyridin-2-amine-cytidine paC, which exhibited a relatively high fluorescence quantum yield. The results could clearly identify f5C with a limit of detection LOD at 0.58 nM. The chemical labelling was capable of altering the hydrogen bonding activity of the parent cytidine and enabled the modulation of the reverse transcription signature of f5C in sequencing profile. We showed that using this Pan-Seq strategy, f5C can be detected from tRNA segments with a single-base resolution. Taken together, this approach represents a sensitive, robust, antibody-free and applicable detection ad sequencing to f5C RNA.

INTRODUCTION

Chemical modifications to ribonucleic acids (RNA) play an essential role for the control of gene expression on different levels.[1,2] The simplest methylation in mRNA, N^6-methyladenine m6A, for example, affects the cell viability and development, and is associated with mRNA localization, structure, stability, splicing, and translation.[3] Similar to that adenosine modification, 5-methylcytosine m^5C is another renowned epigenetic mark that also commonly occurs in various RNA molecules. It has recently gained increasing attention with accumulating evidence suggesting the involvement of this cytidine modification in a variety of cellular processes.[4] In eukaryotes, the installation of C5-methyl at cytosine is catalyzed by enzymes of the NOL1/NOP2/SUN domain (NSUN) family and the DNA methyltransferase homologue DNMT2.[5] Its removal,

on the other hand, is induced by ten-eleven translocation (Tet) family of Fe(II)- and 2-oxoglutarate-dependent dioxygenases (in mammals), the same enzymes also being responsible for the active oxidation/demethylation of 5-methyldeoxycytidine 5mC in DNA.[6-9] In this process, 5-formylcytidine f^5C, one of the unique post-modifications in RNA involving a formyl motif,[10] was identified as one of the key intermediates in the active demethylation pathway of m^5C (Figure 1). It was discovered in both coding and noncoding RNAs,[11,12] along with its precursor 5-hydroxymethylcytidine hm^5C[7,13] and hyper-oxidized product 5-carboxylcytidine ca^5C.[14,15] f^5C was mainly detected at position 34 (wobble base) of mammalian ct-pre-$tRNA^{Leu(CAA)}$[16,17] and mt-$tRNA^{Met}$ in mitochondrial RNA.[18-23] Results from NMR and biophysical characterizations have confirmed that the electron-withdrawing formyl substituent significantly circumscribed the conformational dynamics to the anticodon base region, resulting a specific geometry for a canonical Watson-Crick base pair with both G of AUG and A of AUA, respectively.[22,23] However, its physiological roles are still largely unknown. A major challenge for studying f^5C RNA and its impact on gene regulation is the development of robust, simple-to-use, and cost-efficient analysis tools that allow absolute quantification and determining correct stoichiometry of this epigenetic modification with both analytical chemistry accuracy and high-throughput sequencing.[24]

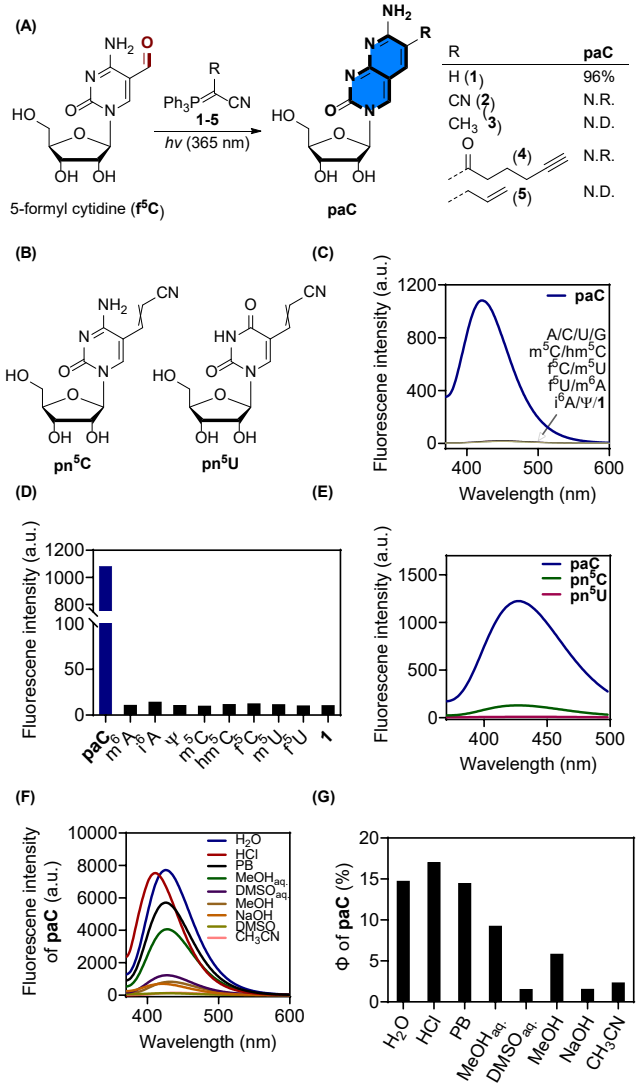

Figure 1. Design and optimization of photochemical labelling of f^{5}C. (A) Design of chemical labelling of f^{5}C with variable substituted Wittig reagents. Representative Wittig reagents used in the optimization process were presented at the right. (B) Labelling intermediates of f^{5}C and f^{5}U during the procedure or before light irradiation. (C) Labelling discrimination of f^{5}C from other canonical nucleosides and common modifications by fluorescence measurement of the process. (D) Fluorescence intensity measurement of paC with other completing nucleosides. (E) Fluorescence intensity measurement of paC with labelling intermediates of f^{5}C and f^{5}U, pn^{5}C and pn^{5}U, respectively. (F) Fluorescence intensity measurement of paC in different solvents. (G) The fluorescence efficiency of paC in different solvents.

To date, approaches for quantifying f^{5}C RNA are basically classified into two types: 1) Extracted RNA was digested, chemically labelled with small organic molecules such as 2-bromoketones and followed by mass spectrometry analysis[11,12] and 2) Fluorogenic labeling of the formyl modification with organic probes like 2,3,3-trimethylindole derivatives[25] or 2-aminobenzenethiol[26] followed by fluorescence intensity measurement. While the former digestion-labeling-analysis procedure loses sequence and origin information and therefore cannot locate the modified nucleoside within the specific RNA, the latter approach could also label another formyl-containing modification 5-formyluridine f^{5}U as well.[10,25,26] Therefore, the results obtained from fluorescence quantitative detection are actually higher than the actual content of f^{5}C. On the other hand, sensitive and accurate methods for sequencing f^{5}C have just emerged until recently. Kleiner, Meier, Minczuk, and Zhou simultaneous disclosed two distinct chemical reduction and alkylation-based sequencing of f^{5}C in RNA.[18,27-29] The mostly used borohydride sequencing relied on the high polarity and electron deficiency of C5=C6 double bond conjugated with the electron-withdrawing formyl group, which was reduced under acidic conditions (pH 1 or 5.2) to afford the dihydrouridine DHU.[18,27,29] A subsequent PCR reaction enabled the conversion of DHU to thymine and thus uncovered the f5C modification in the following sequencing process. While this method has produced signals directly at the f^{5}C modification sites, it could not intrinsically differentiate it with other electron-deficient cytidines. *N*^{4}-acetylcytidine ac^{4}C[30,31] and ca^{5}C,[32] for example, would also be reduced under the same conditions and underwent deamination to generate DHU, causing false positive signals and decreasing the accuracy and sensitivity of the analysis. It is therefore of great interest to develop an affordable and accurate assay to detect f^{5}C RNA with the goal of both detecting and sequencing capabilities.

In this *Article*, we describe a photochemical labeling of f^{5}C RNA with the triphenyl phosphonium ylide **1** (the Wittig reagent) for its accurate detection and single-base sequencing (Figure 1A). The 4,5-cyclization with phosphonium ylide was exceptionally specific towards f^{5}C, which took advantage of the electrophilicity of the C5-carbonyl group and another control reaction within *N*^{4}-amino to discriminate it from ac^{4}C, ca^{5}C and any other modifications. The generated cyclized cytidine analogue 4,5-pyridin-2-amine-cytidine, named paC, exhibited an excellent fluorescence quantum yield, and thus enabled an accurate determination of f^{5}C with a limit of detection (LOD) at 0.58 nM. This straightforward protocol for quantitation of f^{5}C level in cellular RNA eliminated tedious sample preparation but afforded a degree of accuracy comparable to that from the standard measurement.[11,12] Besides, the 4,5-cyclization with f^{5}C impeded base-pairing with guanine and therefore the reverse transcription (RT) signatures can be affected, yielding a sequencing profile distinct from canonical cytidine. We showed that using this strategy, named Pan-Seq, f^{5}C can be detected from the segments within tRNA^{Met} with high accuracy. Taken together, as our approach is sensitive, robust, antibody-free and applicable to

this special RNA species, we expect it will help shed new light on previously intractable biology aspects of f^5C RNA.

RESULTS AND DISSCUSIONS

Photo-triggered Wittig-cyclization for selective chemical labeling and capture of f^5C. We previously have reported a visible-light facilitated and catalyst-free detection of f^5C with 2-aminobenzenethiol to generate the fluorogenic 5-(benzothiazol-2-yl)cytidine.[26] However, this method, along with other labelling reactions, can not distinguish f^5C from another formyl containing f^5U. In our continuous effort to explore opto-bioorthogonal chemistry of RNA modifications,[33-38] we envisioned that a 4,5-type cyclization with both N^4-amino and C5-formyl functionalities would be an ideal solution to this. We firstly explored a bunch of nucleophiles that could selectively react with the carbonyl substituent (Table S1). While most carbon-carbon bond formation reaction (Aldol reaction, Henry addition, etc.) showed inert reactivity under physiological neutral conditions, Wittig olefination, a classic chemical reaction of an aldehyde or ketone with triaryl phosphonium ylides or their precursors,[39] attracted our attention due its exclusive preference towards carbonyls, orthogonality with other biological residues, and compatibility with aqueous medium.[40-42] We therefore investigated a series of phosphonium ylides and their corresponding triphenylphosphonium salts (Figure 1A, Table S1). Interestingly, stable ylides like those with formyl, ester or ketone substituents at the methylene linkage were inert with f^5C (Figure 1A, Table S1). Unstabilized ylides were also tested but they rapidly decomposed during the process (Table S1).[43,44] However, the semi-stabilized ylide cyanomethylene triphenylphosphorane **1**, which was utilized in the condensation with 5fC, was identified as an efficient olefination reagent.[45-48] A pair of *E*/*Z*-5-(3-propenenitrileyl)cytidine (pn^5C) (59%) was obtained by using freshly prepared **1** (Figure 1B, Figure S12), along with a highly fluorescent nucleoside that was later identified as 4,5-pyridin-2-amine-cytidine (paC), however, in a low yield (22%, Table S2). We anticipated that the new paC was produced through a spontaneous cyclization of the less stable *Z*-stereoisomeric pn^5C, in which the spatially close N^4-amino facilitated an amino-nitrile cyclization.[49,50] We therefore adopted the ultraviolet light irradiated $E \rightarrow Z$ isomerization protocol[46,51] in this process and the thermodynamically more stable *E*-pn^5C was completely converted to *Z*-pn^5C upon light irradiation within 5 hours, providing paC in an isolated yield of 78%. However, di-substituted methylenetriphenylphosphoranes **2-5** were unsuccessful for the olefination or cyclization (Figure 1A), probably due to steric hindrances.

With the labeling under very mild reaction conditions in hand, we first investigated the selectivity of **1** with potential intervening functionalities (Figure 1C). A collection of canonical nucleosides A / U / C / G and other epigenetic modifications including m^5C / hm^5C / 5-methyluridine (m^5U) / m^6A / N^6-isopentenyladenosine (i^6A) / pseudouridine (ψ) were all fluorescent inactive and exhibited inert reactivity when incubating with **1**. However, the photo-triggered Wittig-cyclization with f5C induced a remarkable fluorescent enrichment of up to 106-fold at 420 nm (λ_{ex} = 350 nm) (Figures 1C, D, Table S7). We ascribed this change to the creation of fluorescent 4,5-pyridin-2-amine-cytidine. Interestingly, the olefination intermediate pn^5C emitted a very weak fluorescence (Figure 1E). A 10-fold enhancement of signal upon light irradiation was observed and hence enabled a visible confirmation of the cyclization process during RNA labeling. Most importantly, this method exhibited an ultra-high selectivity for f^5C over the structurally similar f^5U, which was not distinguishable in aldehyde condensation protocols.[25,26] A 161-fold difference in fluorescence intensity was detected due to the

covalent cyclization with N^4-amino group in *Z*-pn^{5}C while in the condensation product 5-(3-propenenitrileyl)uridine (pn^{5}U) no further cyclization was possible no matter with under light irradiation or not (Figures 1B, E, Figures S13, S14). We next tested the fluorescent behavior of paC in a series of solvents (Figure 1F, Table S6). Emission spectra of paC in neutral or basic solutions were quite similar, exhibiting a maximum around 426~432 nm. The most intense fluorescence was observed in pure water or phosphate-buffered solution, which was of great benefit to the fluorescence measurement as no organic solvent would be needed in the cell lysate labeling (Figure 1G, Table S8). A high quantum yield around 0.15 is crucial for a wide range of applications enables many exciting possibilities for f^{5}C displays and bioimaging. For paC in acidic hydrochloride solution, the emission maximum was shifted to shorter wavelength (411 nm), probably due to the protonation of the remote amino group.

To explore the potential of the photochemical labeling of f^{5}C modification in RNA, two oligoes f^{5}C-RNA-1 (5'-GGG AGA f^{5}CGU A-3') and f^{5}C-RNA-2 (5'-AGC UAU CGG GCC f^{5}CAU ACC CCG AAA A-3') (Table S3), whose sequence was adapted from wobble regions of human mt-tRNA^{Met},[52] were synthesized following the standard phosphoramidite chemistry (Figure 2A, Figure S15).[53] To identify an optimal labeling condition for the selective modification of f^{5}C nucleotides, we began with screening of variable Wittig reagent amounts by employing f^{5}C-RNA-1 as a model system (Figure 2B, Table S4). Gratifyingly, a rapid conversion of f^{5}C RNA (93%) was observed with an exclusive formation of product paC RNA (83%) when sufficient 1 was employed (Figures 2B, 2C). The skeleton stability of the ribonucleotides during the photochemical cyclization was unambiguously proven based on the high-performance liquid chromatography (HPLC) analysis (Figure 2C). Two new peaks were observed after the incubation with 1 before radiation, one of which turned out to be the anticipated paC-RNA-1 based on the matrix-assisted laser desorption/ionization coupled to time-of-flight (MALDI-TOF) mass spectrometry analysis (Figure S17), while the other peak was tentatively assigned as the pn^{5}C-RNA, which was later isomerized to its cis-form and underwent a spontaneous cyclization to achieve a full conversion to paC-RNA-1 under light irradiation. Enzymatic digestion of the RNA oligo supported the conclusion that the f^{5}C modification has been selectively converted to paC, exhibiting a new peak at 260 nm and 350 nm, respectively (Figure 2D). Another mt-tRNA^{Met} mimic, f^{5}C-RNA-2 with a completely different sequence, also underwent a complete transformation with **1** under the photochemical cyclization (77%) (Figure 2E, Table S5). MALDI-TOF analysis (Figure S18) and enzymatic cleavages (Figure 2F) also confirmed that the chemical recognition and functionalization with Wittig reagent **1** showed no sequence or conformation preference but a high selectivity for f^{5}C modification, which is consistent with our previously developed chemical modulation of RNA epigenetics with small organic molecules.[37,38,54]

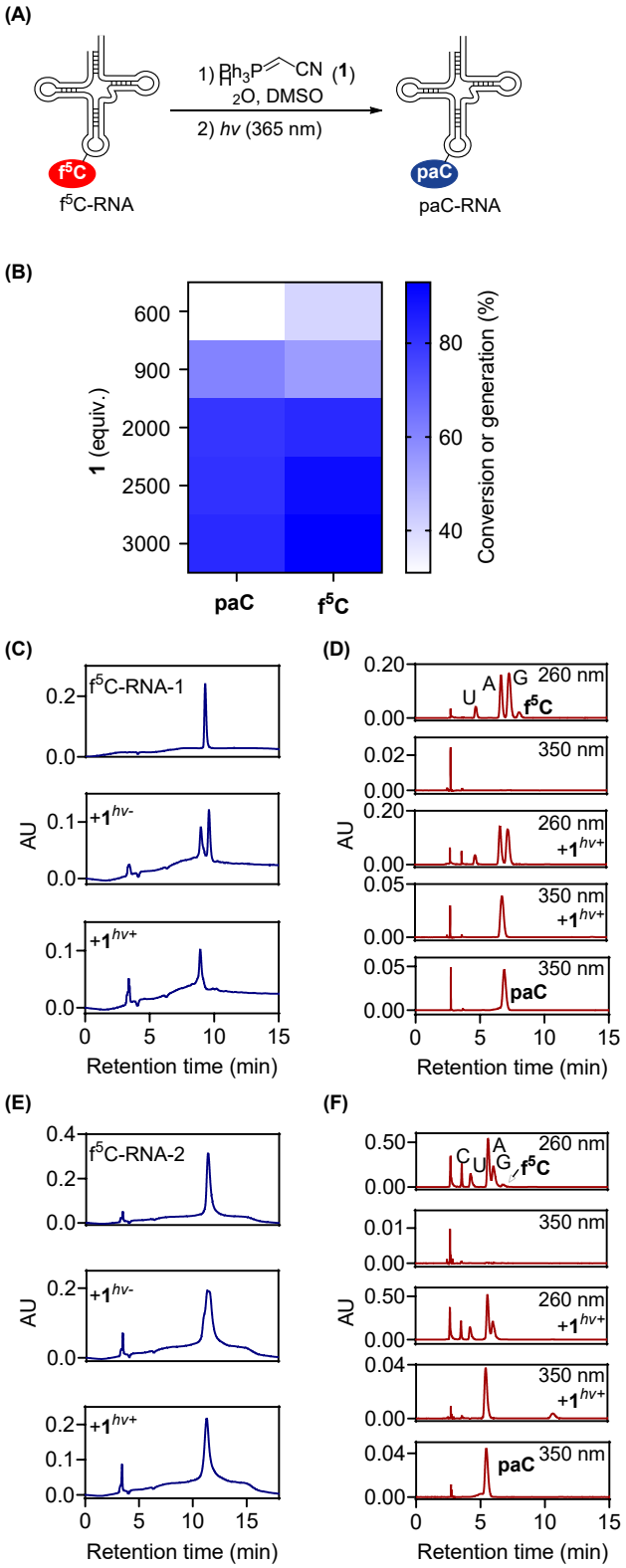

Figure 2. Photo-triggered Wittig-cyclization for selective chemical labeling and capture of f^{5}C. (A) A schematic description of the chemical labelling of f^{5}C within a fragment of tRNA at the wobble position. (B) The conversion and generation of f^{5}C and paC upon incubation with variable amounts of the Wittig reagent **1**. (C) HPLC analysis of f^{5}C-RNA-1 and its chemical labelling with Wittig reagent **1** before ($h\nu^-$) and after light irradiation ($h\nu^+$). (D) HPLC analysis of the digestion mixture of f^{5}C-RNA-1 and its chemical labelling mixture

with Wittig reagent **1** before ($h\nu^-$) and after light irradiation ($h\nu^+$). HPLC chromatograms with a wavelength of 260 nm and 350 nm were recorded respectively. The lowest HPLC was recorded with the pure paC sample. (E) HPLC analysis of f^{5}C-RNA-2 and its chemical labelling with Wittig reagent 1 before ($h\nu^-$) and after light irradiation ($h\nu^+$). (F) HPLC analysis of the digestion mixture of f^{5}C-RNA-2 and its chemical labelling mixture with Wittig reagent 1 before ($h\nu^-$) and after light irradiation ($h\nu^+$). HPLC chromatograms with a wavelength of 260 nm and 350 nm were recorded respectively.

Fluorescence measurement of f^{5}C modification in RNA. Having achieved quantitative labeling and HPLC resolution of f^{5}C under the bio-compatible conditions, efforts were directed to use this fluorescence analysis as an accurate approach to quantitative f^{5}C modification in RNA (Figure 3). First we demonstrated that this fluorogenic labeling could be utilized to assay f^{5}C nucleoside in a digested RNA sample (Figures 3A, B). Current methods for measuring f^{5}C nucleoside in RNA employed base-promoted alkylation with arylacyl bromides derivatives to improve the LC-ESI-MS/MS detection sensitivity,[11,12] or acid-facilitated aldol type condensation with trimethylindoleninium to produce hemicyanine-like chromophores.[25] Considering the facts that paC was constructed under physiological pH and exhibited maximum quantum yield and largest emission intensity under the same conditions (Figures 1F, G), we anticipated that our method may serve as a rapid and sensitive fluorescence-based assay to quantify cellular f^{5}C nucleosides in the routine RNA treatment. This assay relied on the principle that only the fluorescence coming from the fluorophore of interest 4,5-pyridin-2-amine-cytidine can be detected. The concentration of limiting f^{5}C was thus directly proportional to the fluorescence generated by avoiding any other unspecific signals (Figure 3A). The assay exhibited excellent linearity ($R^2 > 0.99$) and could be modified to detect between 1 and 800 nM of f^{5}C. The limits of detection (LOD) for f^{5}C modification was defined as 0.58 nM (Figure 3B). One additional advantage of the current assay over the conventional mass-based assays is that this method could be applied in RNA samples without any interference from the endogenous cellular nucleotides (Figure 1C) or intermediates generated during the labelling process (Figure 1E). No fluorescence signal above background was detected with f^{5}C RNA before the labelling while the photochemical treatment with **1** resulted a strong blue light emission under an UV lamp (Figure 3C). A significant increase in fluorescence enhancement (106-fold) was observed (Figure 3D, Figure S7). To evaluate the feasible application of the highly fluorescent of paC for the detection of f^{5}C in RNA samples, a quantitative measurement of paC RNA using the spectrofluorometric approach was explored (Figures 3E, F). The fluorescence signal changes of paC RNA with cumulative concentrations were given in Figure 3E and clearly indicated that this assay could speedily quantify even trace amount of f^{5}C within RNA. The linear relationship and lowest detection limit of the current approach was shown in Figure 3F, in which the emission intensities of the paC RNA were plotted against variable concentrations from which a calibration curve was obtained. The calibration diagram indicated that the proposed sensing strategy was in good fit with concentration ranging from 0.1 to 10 μM, while the LOD was calculated as 19.2 nM. By this approach, determination of absolute f^{5}C amounts in total RNAs of the extracts from HeLa cells (328 nM) is possible which is of sufficient accuracy for the majority of cell-based applications without tedious chemical treatment and specialized instrumentation.

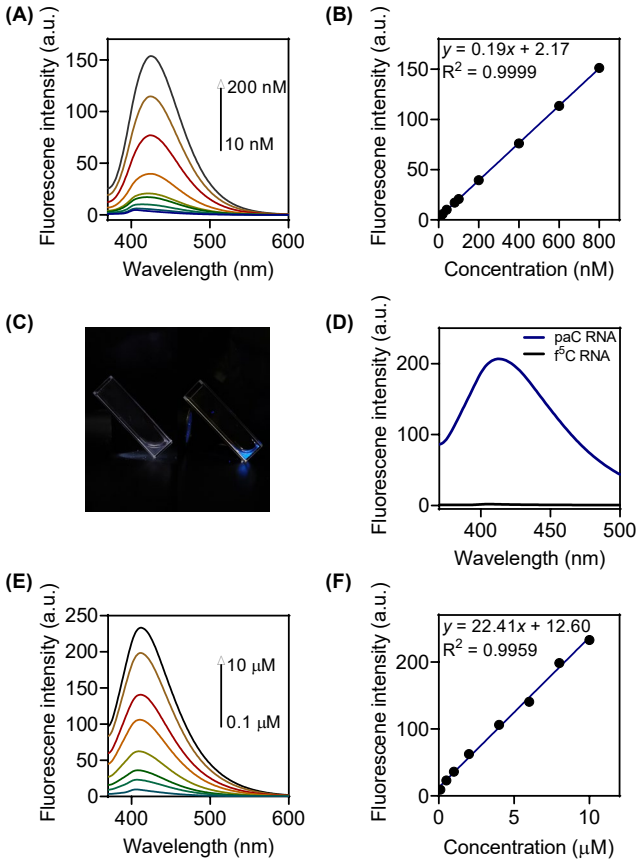

Figure 3. The fluorescence measurement of f^{5}C modification in RNA. (A) Fluorescence intensity measured by variation of paC nucleoside from 10 nM to 200 nM. (B) The linear relationship between the concentrations of paC with the maximum absorptions at λ of 422 nm. (C) The fluorescence snapshots of an f^{5}C RNA solution before (left) and after photochemical labeling (right). (D) The fluorescence determination of f^{5}C RNA before and after the photochemical labelling. paC RNA indicated the purified oligo after purification. (E) Fluorescence intensity measured by variation of paC RNA from 0.1 μM to 10 μM. (F) The linear relationship between the concentrations of paC RNA with the maximum absorptions at λ of 410 nm.

Pan-Seq for single base-resolution sequencing of f^{5}C RNA. To further understand how the dynamic oxidation of m^{5}C shapes RNA, a detailed analysis of the distribution of its oxidative intermediates including hm^{5}C,[55] f^{5}C and ca^{5}C is highly desirable as these modifications are actively involved in the demethylation pathway. Unfortunately, f^{5}C behaved equally to cytosine in bisulfite sequencing[56] and similarly to ca^{5}C and ac^{4}C in protonation-dependent hydride reduction based sequencings.[18,27,29] Besides, its low abundance in transcript RNA (10^{-6} f^{5}C/G) makes it challenging to efficiently employ antibody-based immunoprecipitation that normally worked well with other commonly existing modifications. To overcome such technical bottleneck, we consider that the hydrogen bond donor and acceptor sites of paC have changed significantly compared with cytosine (Figure S26), and it may no longer pair with guanine during the reverse transcription process, resulting in distinct changes in the reading signal and hence an exact indication of the formyl modification sites. Based on our previous work on the discovery of base resolution analysis of DNA modifications pertaining to cytidine and adenosine,[57-60] we developed a new sequencing technology to detect f^{5}C in single base-resolution named paC-seq and used two f^{5}C RNA oligos and their corresponding paC RNAs (Table S13) to test the sensitivity and specificity

(Figure 4). To our delight, 70.2~70.9% of paC was mutated after reverse transcription and PCR amplification (Figure 4A, B), of which 98% were C-to-U mutations (Figure 4C), which was at the leading level in the reported chemical reduction and alkylation-based sequencing methods (29~60%).[27,28] These results indicated that our methods can sensitively detect f^{5}C sites by recognizing the C-to-U mutation and exhibit an extremely high selectivity toward paC over unmodified C with no cross-reactivity to other bases. The consistent performance in oligonucleotides with distinct sequences from f^{5}C-RNA-1/2 also indicated the stability of our method. Taken together, the excellent mutation from paC to U has enable the detection of f5C from unmodified C in a single base-resolution.

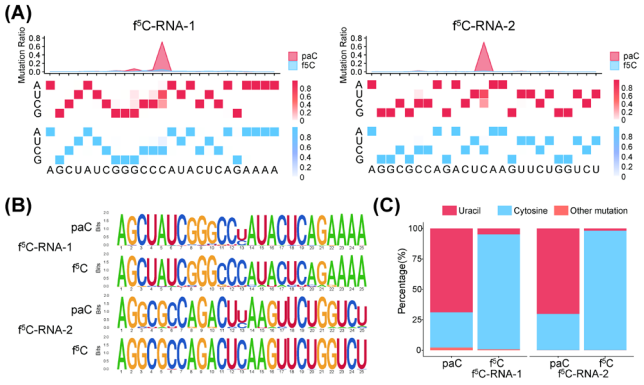

Figure 4. f^{5}C RNA modification detected with high sensitivity and specificity. (A) Mutation Ratio (line chart) and base frequency distribution (heat map) in single base-resolution in f^{5}C-RNA-1 (left) and f^{5}C-RNA-2 (right) shows paC has higher mutation ratio and the mutation of other sites was almost unaffected. In heat map, the identities of single base substitutions are showed on the left; reference sequence represented below. The right bar shows base frequency (from light to dark). Red and blue represent paC and f^{5}C, respectively. (B) Sequence logo shows the base composition and conservation of each site. Height of each site shows the site conservation, and the height of each base represents the frequency of each base. Bits = $\log_2(N)$-H(X), N = 4 and H(X) represents comentropy. (C) A stacking histogram shows the base composition of the paC and f^{5}C sites in f^{5}C-RNA-1 (left) and f^{5}C-RNA-2 (right). paC shows higher and stable conversion from cytosine to uracil than f^{5}C in each RNA oligonucleotide.

CONCLUSIONS

The f^{5}C modification has received cumulative attention in recent years. It was believed to play an important role in RNA epigenetic functions, however it remains challenging to quantify this modification in an unbiased manner. Here, we presented a conceptually new method for the distinction of f^{5}C in RNA from other related modifications. We first introduced a site-specific dual functionalization of the C5-formyl and N^4-amino in cytosine with semi-stabilized ylide cyanomethylene triphenylphosphorane. This fluorogenic switch-on labeling of f^{5}C produced an intramolecularly cyclized nucleobase paC under a cascade Wittig and photo-promoted isomerization reaction, leaving other canonical nucleosides and structurally similar f^{5}U untouched. The excellent fluorescence properties of paC have ensured a qualitative and quantitative detection of f^{5}C moieties in tRNA derivatives with a high accuracy that was comparable to quantitative mass spectrometry. Chemical labelling for fluorescent imaging apparatuses such as fluorescent microscopy is one of the most effective and widely used modern technologies in life sciences for the analysis of the quantitative behavior of

biomacromolecules and metabolites in living cells, tissues, and organisms.[61] While numerous imaging tools like electron microscopy, autoradiography, and immunochemistry have been developed, an accurate fluorescent labeling of RNA modifications without background noise and their subsequent observation and detection with optical instruments are still of great attention. A high temporal resolution of designated modification is needed to be analyzed for understanding the epigenetics dynamics. The current assay circumvented this issue by avoiding the need for a fluorescent labelling reagent and employing a fluorescence-inert reagent for the signal OFF to ON determination.

Second, we developed a photochemically assisted sequencing (Pan-Seq) method for the base-resolution detection of f^5C. The conversion of the exocyclic 4-amino group for the construction of the connected pyridine ring, which acted as a hydrogen binding donor in f^5C, has enabled an altered base pairing with A and thus led to C-to-T conversion during the following polymerase extension. With that technique in hand, we revealed the possibility of a single-base resolution analysis of f^5C in two mt-$tRNA^{Met}$ mimic oligonucleotides. The Illumina sequencing results disclosed a great potential of this method in the single-base resolution analysis of f^5C. A similar design could be further applied to both qualitative and quantitative detection and single-base resolution analysis of natural f^5C modification in more cell and tissue studies in the near future. We expect it will help shed new light on previously intractable aspects of f^5C RNA and its related biology events.

REFERENCES

[1] Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **2017**, *169* (7), 1187–1200.

[2] Traube, F. R.; Carell, T. The Chemistries and Consequences of DNA and RNA Methylation and Demethylation. *RNA Biol.* **2017**, *14* (9), 1099–1107.

[3] Wang, X.; Zhao, B. S.; Roundtree, I. A.; Lu, Z.; Han, D.; Ma, H.; Weng, X.; Chen, K.; Shi, H.; He, C. N(6)-Methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **2015**, *161* (6), 1388–1399.

[4] Xue, C.; Zhao, Y.; Li, L. Advances in RNA Cytosine-5 Methylation: Detection, Regulatory Mechanisms, Biological Functions and Links to Cancer. *Biomark. Res.* **2020**, *8* (1), 43.

[5] Bohnsack, K. E.; Höbartner, C.; Bohnsack, M. T. Eukaryotic 5-Methylcytosine (m5C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease. *Genes* **2019**, *10* (2), 102.

[6] Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **2011**, *333* (6047), 1300–1303.

[7] Fu, L.; Guerrero, C. R.; Zhong, N.; Amato, N. J.; Liu, Y.; Liu, S.; Cai, Q.; Ji, D.; Jin, S.-G.; Niedernhofer, L. J.; Pfeifer, G. P.; Xu, G.-L.; Wang, Y. Tet-Mediated Formation of 5-Hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **2014**, *136* (33), 11582–11585.

[8] Schiesser, S.; Pfaffeneder, T.; Sadeghian, K.; Hackner, B.; Steigenberger, B.; Schröder, A. S.; Steinbacher, J.; Kashiwazaki, G.; Höfner, G.; Wanner, K. T.; Ochsenfeld, C.; Carell, T. Deamination, Oxidation, and C-C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine. *J. Am. Chem. Soc.* **2013**, *135* (39), 14593–14599.

[9] Kohli, R. M.; Zhang, Y. TET Enzymes, TDG and the Dynamics of DNA Demethylation. *Nature* **2013**, *502* (7472), 472–479.

[10] Runtsch, L. S.; Stadlmeier, M.; Schön, A.; Müller, M.; Carell, T. Comparative Nucleosomal Reactivity of 5-Formyl-Uridine and 5-Formyl-Cytidine. *Chem.-Eur. J.* **2021**, *27* (50), 12747–12752.

[11] Huang, W.; Lan, M.-D.; Qi, C.-B.; Zheng, S.-J.; Wei, S.-Z.; Yuan, B.-F.; Feng, Y.-Q. Formation and Determination of the Oxidation Products of 5-Methylcytosine in RNA. *Chem. Sci.* **2016**, *7* (8), 5495–5502.

[12] Feng, Y.; Ma, C.-J.; Ding, J.-H.; Qi, C.-B.; Xu, X.-J.; Yuan, B.-F.; Feng, Y.-Q. Chemical Labeling - Assisted Mass Spectrometry Analysis for Sensitive Detection of Cytidine Dual Modifications in RNA of Mammals. *Anal. Chim. Acta* **2020**, *1098*, 56–65.

[13] Huber, S. M.; van Delft, P.; Mendil, L.; Bachman, M.; Smollett, K.; Werner, F.; Miska, E. A.; Balasubramanian, S. Formation and Abundance of 5-Hydroxymethylcytosine in RNA. *ChemBioChem* **2015**, *16* (5), 752–755.

[14] Basanta-Sanchez, M.; Wang, R.; Liu, Z.; Ye, X.; Li, M.; Shi, X.; Agris, P. F.; Zhou, Y.; Huang, Y.; Sheng, J. TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. *ChemBioChem* **2017**, *18* (1), 72–76.

[15] Trixl, L.; Lusser, A. The Dynamic RNA Modification 5-Methylcytosine and Its Emerging Role as an Epitranscriptomic Mark. *Wiley Interdiscip. Rev. RNA* **2019**, *10* (1), e1510.

[16] Païs de Barros, J. P.; Keith, G.; El Adlouni, C.; Glasser, A. L.; Mack, G.; Dirheimer, G.; Desgrès, J. 2'-O-Methyl-5-Formylcytidine (f5Cm), a New Modified Nucleotide at the "wobble" of Two Cytoplasmic TRNAs Leu (NAA) from Bovine Liver. *Nucleic Acids Res.* **1996**, *24* (8), 1489–1496.

[17] Brzezicha, B.; Schmidt, M.; Makalowska, I.; Jarmolowski, A.; Pienkowska, J.; Szweykowska-Kulinska, Z. Identification of Human TRNA:M5C Methyltransferase Catalysing Intron-Dependent M5C Formation in the First Position of the Anticodon of the Pre-TRNA Leu (CAA). *Nucleic Acids Res.* **2006**, *34* (20), 6034–6043.

[18] Van Haute, L.; Minczuk, M. Detection of 5-Formylcytosine in Mitochondrial Transcriptome. *Methods Mol. Biol.* **2021**, *2192*, 59–68.

[19] Moriya, J.; Yokogawa, T.; Wakita, K.; Ueda, T.; Nishikawa, K.; Crain, P. F.; Hashizume, T.; Pomerantz, S. C.; McCloskey, J. A.; Kawai, G. A Novel Modified Nucleoside Found at the First Position of the Anticodon of Methionine TRNA from Bovine Liver Mitochondria. *Biochemistry* **1994**, *33* (8), 2234–2239.

[20] Watanabe, Y.; Tsurui, H.; Ueda, T.; Furushima, R.; Takamiya, S.; Kita, K.; Nishikawa, K.; Watanabe, K. Primary and Higher Order Structures of Nematode (Ascaris Suum) Mitochondrial TRNAs Lacking Either the T or D Stem. *J. Biol. Chem.* **1994**, *269* (36), 22902–22906.

[21] Takemoto, C.; Spremulli, L. L.; Benkowski, L. A.; Ueda, T.; Yokogawa, T.; Watanabe, K. Unconventional Decoding of the AUA Codon as Methionine by Mitochondrial tRNAMet with the Anticodon f5CAU as Revealed with a Mitochondrial in Vitro Translation System. *Nucleic Acids Res.* **2009**, *37* (5), 1616–1627.

[22] Bilbille, Y.; Gustilo, E. M.; Harris, K. A.; Jones, C. N.; Lusic, H.; Kaiser, R. J.; Delaney, M. O.; Spremulli, L. L.; Deiters, A.; Agris, P. F. The Human Mitochondrial tRNAMet: Structure/Function Relationship of a Unique Modification in the Decoding of Unconventional Codons. *J. Mol. Biol.* **2011**, *406* (2), 257–274.

[23] Cantara, W. A.; Murphy, F. V., 4th; Demirci, H.; Agris, P. F. Expanded Use of Sense Codons Is Regulated by Modified Cytidines in tRNA. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (27), 10964–10969.

[24] Schaefer, M.; Kapoor, U.; Jantsch, M. F. Understanding RNA Modifications: The Promises and Technological Bottlenecks of the "Epitranscriptome." *Open Biol.* **2017**, *7* (5), 170077.

[25] Samanta, B.; Seikowski, J.; Höbartner, C. Fluorogenic Labeling of 5-Formylpyrimidine Nucleotides in DNA and RNA. *Angew. Chem. Int. Ed Engl.* **2016**, *55* (5), 1912–1916.

[26] Wang, R.-L.; Jin, X.-Y.; Kong, D.-L.; Chen, Z.-G.; Liu, J.; Liu, L.; Cheng, L. Visible-light Facilitated Fluorescence "Switch-on" Labelling of 5-formylpyrimidine RNA. *Adv. Synth. Catal.* **2019**, *361* (23), 5406–5411.

[27] Link, C. N.; Thalalla Gamage, S.; Gallimore, D.; Kopajtich, R.; Evans, C.; Nance, S.; Fox, S. D.; Andresson, T.; Chari, R.; Ivanic, J.; Prokisch, H.; Meier, J. L. Protonation-Dependent Sequencing of 5-Formylcytidine in RNA. *Biochemistry* **2022**, *61* (7), 535–544.

[28] Li, A.; Sun, X.; Arguello, A. E.; Kleiner, R. E. Chemical Method to Sequence 5-Formylcytosine on RNA. *ACS Chem. Biol.* **2022**, *17* (3), 503–508.

[29] Wang, Y.; Chen, Z.; Zhang, X.; Weng, X.; Deng, J.; Yang, W.; Wu, F.; Han, S.; Xia, C.; Zhou, Y.; Chen, Y.; Zhou, X. Single-Base Resolution Mapping Reveals Distinct 5-Formylcytidine in Saccharomyces Cerevisiae mRNAs. *ACS Chem. Biol.* **2022**, *17* (1), 77–84.

[30] Sas-Chen, A.; Thomas, J. M.; Matzov, D.; Taoka, M.; Nance, K. D.; Nir, R.; Bryson, K. M.; Shachar, R.; Liman, G. L. S.; Burkhart, B. W.; Gamage, S. T.; Nobe, Y.; Briney, C. A.; Levy, M. J.; Fuchs, R. T.; Robb, G. B.; Hartmann, J.; Sharma, S.; Lin, Q.; Florens, L.; Washburn, M. P.; Isobe, T.; Santangelo, T. J.; Shalev-Benami, M.; Meier, J. L.; Schwartz, S. Dynamic RNA Acetylation Revealed by Quantitative Cross-Evolutionary Mapping. *Nature* **2020**, *583* (7817), 638–643.

[31] Thomas, J. M.; Briney, C. A.; Nance, K. D.; Lopez, J. E.; Thorpe, A. L.; Fox, S. D.; Bortolin-Cavaille, M.-L.; Sas-Chen, A.; Arango, D.; Oberdoerffer, S.; Cavaille, J.; Andresson, T.; Meier, J. L. A Chemical Signature for Cytidine Acetylation in RNA. *J. Am. Chem. Soc.* **2018**, *140* (40), 12667–12670.

[32] Liu, Y.; Siejka-Zielińska, P.; Velikova, G.; Bi, Y.; Yuan, F.; Tomkova, M.; Bai, C.; Chen, L.; Schuster-Böckler, B.; Song, C.-X. Bisulfite-Free Direct Detection of 5-Methylcytosine and 5-Hydroxymethylcytosine at Base Resolution. *Nat. Biotechnol.* **2019**, *37* (4), 424–429.

[33] Xie, L.-J.; Lin, C.-L.; Liu, L.; Cheng, L. A Chemical Labeling of N6-Formyl Adenosine (f6A) RNA. *Chin. Chem. Lett.* **2022**, *33* (3), 1563–1566.

[34] Lan, L.; Sun, Y.-J.; Liu, L.; Cheng, L. A Photo-Responsive Chemical Modulation of m6A RNA Demethylase FTO. *Chem. Commun.* **2021**, *57* (81), 10548–10551.

[35] Lan, L.; Sun, Y.-J.; Jin, X.-Y.; Xie, L.-J.; Liu, L.; Cheng, L. A Light-Controllable Chemical Modulation of m6A RNA Methylation. *Angew. Chem. Int. Ed Engl.* **2021**, *60* (33), 18116–18121.

[36] Jin, X.-Y.; Wang, R.-L.; Xie, L.-J.; Kong, D.-L.; Liu, L.; Cheng, L. A Chemical Photo-oxidation of 5-methyl Cytidines. *Adv. Synth. Catal.* **2019**, *361* (20), 4685–4690.

[37] Xie, L.-J.; Yang, X.-T.; Wang, R.-L.; Cheng, H.-P.; Li, Z.-Y.; Liu, L.; Mao, L.; Wang, M.; Cheng, L. Identification of Flavin Mononucleotide as a Cell-Active Artificial N6-Methyladenosine RNA Demethylase. *Angew. Chem. Int. Ed Engl.* **2019**, *58* (15), 5028–5032.

[38] Xie, L.-J.; Wang, R.-L.; Wang, D.; Liu, L.; Cheng, L. Visible-Light-Mediated Oxidative Demethylation of N6-Methyl Adenines. *Chem. Commun.* **2017**, *53* (77), 10734–10737.

[39] Maryanoff, B. E.; Reitz, A. B. The Wittig Olefination Reaction and Modifications Involving Phosphoryl-Stabilized Carbanions. Stereochemistry, Mechanism, and Selected Synthetic Aspects. *Chem. Rev.* **1989**, *89* (4), 863–927.

[40] Shi, Y.; Fu, L.; Yang, J.; Carroll, K. S. Wittig Reagents for Chemoselective Sulfenic Acid Ligation Enables Global Site Stoichiometry Analysis and Redox-Controlled Mitochondrial Targeting. *Nat. Chem.* **2021**, *13* (11), 1140–1150.

[41] Dambacher, J.; Zhao, W.; El-Batta, A.; Anness, R.; Jiang, C.; Bergdahl, M. Water Is an Efficient Medium for Wittig Reactions Employing Stabilized Ylides and Aldehydes. *Tetrahedron Lett.* **2005**, *46* (26), 4473–4477.

[42] El-Batta, A.; Jiang, C.; Zhao, W.; Anness, R.; Cooksy, A. L.; Bergdahl, M. Wittig Reactions in Water Media Employing Stabilized Ylides with Aldehydes. Synthesis of Alpha,Beta-Unsaturated Esters from Mixing Aldehydes, Alpha-Bromoesters, and Ph3P in Aqueous NaHCO3. *J. Org. Chem.* **2007**, *72* (14), 5244–5259.

[43] Robiette, R.; Richardson, J.; Aggarwal, V. K.; Harvey, J. N. Reactivity and Selectivity in the Wittig Reaction: A Computational Study. *J. Am. Chem. Soc.* **2006**, *128* (7), 2394–2409.

[44] Byrne, P. A.; Gilheany, D. G. The Mechanism of Phosphonium Ylide Alcoholysis and Hydrolysis: Concerted Addition of the O-H Bond across the P=C Bond. *Chem.-Eur. J.* **2016**, *22* (27), 9140–9154.

[45] Ono, A.; Okamoto, T.; Inada, M.; Nara, H.; Matsuda, A. Nucleosides and Nucleotides. 131. Synthesis and Properties of Oligonucleotides Containing 5-Formyl-2'-Deoxyuridine. *Chem. Pharm. Bull.* **1994**, *42* (11), 2231–2237.

[46] Zhou, Q.; Li, K.; Li, L.-L.; Yu, K.-K.; Zhang, H.; Shi, L.; Chen, H.; Yu, X.-Q. Combining Wittig Olefination with Photoassisted Domino Reaction to Distinguish 5-Formylcytosine from 5-Formyluracil. *Anal. Chem.* **2019**, *91* (15), 9366–9370.

[47] Zhou, Q.; Li, K.; Yu, K.-K.; Li, N.; Shi, L.; Chen, H.; Chen, S.-Y.; Yu, X.-Q. Aqueous Wittig Reaction-Mediated Fast Fluorogenic Identification and Single-Base Resolution Analysis of 5-Formylcytosine in DNA. *Chem. Commun.* **2020**, *56* (81), 12158–12161.

[48] Zhou, Q.; Li, K.; Liu, Y.-H.; Li, L.-L.; Yu, K.-K.; Zhang, H.; Yu, X.-Q. Fluorescent Wittig Reagent as a Novel Ratiometric Probe for the Quantification of 5-Formyluracil and Its Application in Cell Imaging. *Chem. Commun.* **2018**, *54* (97), 13722–13725.

[49] Hirota, K.; Kubo, K.; Sajiki, H. Reactivities of 6-Amino-1,3-Dimethyl-5-Thioformyluracil toward Nucleophiles and Its Application to Synthesis of Pyrido(2,3-d)Pyrimidines. *Chem. Pharm. Bull.* **1997**, *45* (3), 542–544.

[50] Zinchenko, A. N.; Muzychka, L. V.; Biletskii, I. I.; Smolii, O. B. Synthesis of New 4-Amino-Substituted 7-Iminopyrido[2,3-d]Pyrimidines. *Chem. Heterocycl. Compd.* **2017**, *53* (5), 589–596.

[51] Gilmour, R.; Metternich, J. Photocatalytic E → Z Isomerization of Alkenes. *Synlett* **2016**, *27* (18), 2541–2552.

[52] Kawarada, L.; Suzuki, T.; Ohira, T.; Hirata, S.; Miyauchi, K.; Suzuki, T. ALKBH1 Is an RNA Dioxygenase Responsible for Cytoplasmic and Mitochondrial tRNA Modifications. *Nucleic Acids Res.* **2017**, *45* (12), 7401–7415.

[53] Tanpure, A. A.; Balasubramanian, S. Synthesis and Multiple Incorporations of 2'-O-Methyl-5-Hydroxymethylcytidine, 5-Hydroxymethylcytidine and 5-Formylcytidine Monomers into RNA Oligonucleotides. *ChemBioChem* **2017**, *18* (22), 2236–2241.

[54] Cheng, H.-P.; Yang, X.-H.; Lan, L.; Xie, L.-J.; Chen, C.; Liu, C.; Chu, J.; Li, Z.-Y.; Liu, L.; Zhang, T.-Q.; Luo, D.-Q.; Cheng, L. Chemical Deprenylation of N6 -Isopentenyladenosine (I6 A) RNA. *Angew. Chem. Int. Ed Engl.* **2020**, *59* (26), 10645–10650.

[55] Delatte, B.; Wang, F.; Ngoc, L. V.; Collignon, E.; Bonvin, E.; Deplus, R.; Calonne, E.; Hassabi, B.; Putmans, P.; Awe, S.; Wetzel, C.; Kreher, J.; Soin, R.; Creppe, C.; Limbach, P. A.; Gueydan, C.; Kruys, V.; Brehm, A.; Minakhina, S.; Defrance, M.; Steward, R.; Fuks, F. Transcriptome-Wide Distribution and Function of RNA Hydroxymethylcytosine. *Science* **2016**, *351* (6270), 282–285.

[56] Van Haute, L.; Dietmann, S.; Kremer, L.; Hussain, S.; Pearce, S. F.; Powell, C. A.; Rorbach, J.; Lantaff, R.; Blanco, S.; Sauer, S.; Kotzaeridou, U.; Hoffmann, G. F.; Memari, Y.; Kolb-Kokocinski, A.; Durbin, R.; Mayr, J. A.;

Frye, M.; Prokisch, H.; Minczuk, M. Deficient Methylation and Formylation of Mt-tRNAMet Wobble Cytosine in a Patient Carrying Mutations in NSUN3. *Nat. Commun.* **2016**, *7* (1), 12039.

[57] Liu, J.; Yue, Y.; Han, D.; Wang, X.; Fu, Y.; Zhang, L.; Jia, G.; Yu, M.; Lu, Z.; Deng, X.; Dai, Q.; Chen, W.; He, C. A METTL3-METTL14 Complex Mediates Mammalian Nuclear RNA N6-Adenosine Methylation. *Nat. Chem. Biol.* **2014**, *10* (2), 93–95.

[58] Xia, B.; Han, D.; Lu, X.; Sun, Z.; Zhou, A.; Yin, Q.; Zeng, H.; Liu, M.; Jiang, X.; Xie, W.; He, C.; Yi, C. Bisulfite-Free, Base-Resolution Analysis of 5-Formylcytosine at the Genome Scale. *Nat. Methods* **2015**, *12* (11), 1047–1050.

[59] Han, D.; Lu, X.; Shih, A. H.; Nie, J.; You, Q.; Xu, M. M.; Melnick, A. M.; Levine, R. L.; He, C. A Highly Sensitive and Robust Method for Genome-Wide 5hmC Profiling of Rare Cell Populations. *Mol. Cell* **2016**, *63* (4), 711–719.

[60] Dominissini, D.; Nachtergaele, S.; Moshitch-Moshkovitz, S.; Peer, E.; Kol, N.; Ben-Haim, M. S.; Dai, Q.; Di Segni, A.; Salmon-Divon, M.; Clark, W. C.; Zheng, G.; Pan, T.; Solomon, O.; Eyal, E.; Hershkovitz, V.; Han, D.; Doré, L. C.; Amariglio, N.; Rechavi, G.; He, C. The Dynamic N(1)-Methyladenosine Methylome in Eukaryotic Messenger RNA. *Nature* **2016**, *530* (7591), 441–446.

[61] Stephens, D. J.; Allan, V. J. Light Microscopy Techniques for Live Cell Imaging. *Science* **2003**, *300* (5616), 82–86