

# Small-molecule-catalysed deamination enables transcriptome-wide profiling of *N*<sup>6</sup>-methyladenosine in RNA

Received: 3 June 2024

Accepted: 11 March 2025

Published online: 17 April 2025

 Check for updates

Pingluan Wang <sup>1,2,5</sup>, Chang Ye <sup>1,2,5</sup>, Michelle Zhao<sup>1,5</sup>, Bochen Jiang <sup>1,2</sup> & Chuan He <sup>1,2,3,4</sup> ✉

The deamination reaction is important to both fundamental organic chemistry and biochemistry. Traditional chemical methods of deamination rely on the use of aryldiazonium salts under harsh acidic conditions, which limits the application scope for most biological substrates. Here we present an N-nitrosation strategy for deamination under mild conditions that DNA and RNA biological macromolecules can tolerate. Cooperative catalysis combining a carbonyl organocatalyst with a Lewis acid catalyst facilitates the formation of a carbon–nitro intermediate from a primary amine, which, on rearrangement into *N*-nitrosamine, leads to the selective deamination of unsubstituted canonical DNA/RNA bases under mild conditions. We used this approach to deaminate adenine into hypoxanthine, read as guanine by reverse transcriptases or DNA polymerases, while *N*<sup>6</sup>-methyladenosine sites resist deamination and remain identified as adenine. This reactivity enables a chemically mild, low-input detection method for sequencing of adenosine methylation at base resolution, named chemical cooperative catalysis-assisted *N*<sup>6</sup>-methyladenosine sequencing.

Chemical modifications in RNA and DNA play crucial roles in a wide range of biological processes, including transcription regulation, RNA degradation, protein translation and immune modulation<sup>1–6</sup>. These modifications have been quantitatively mapped at single-base resolution by new sequencing methods<sup>7–14</sup>, with nucleobase deamination being a highly effective strategy for the selective mapping of 5-methylcytosine in DNA<sup>9,10</sup> and *N*<sup>6</sup>-methyladenosine (m<sup>6</sup>A) in RNA<sup>12,13</sup>.

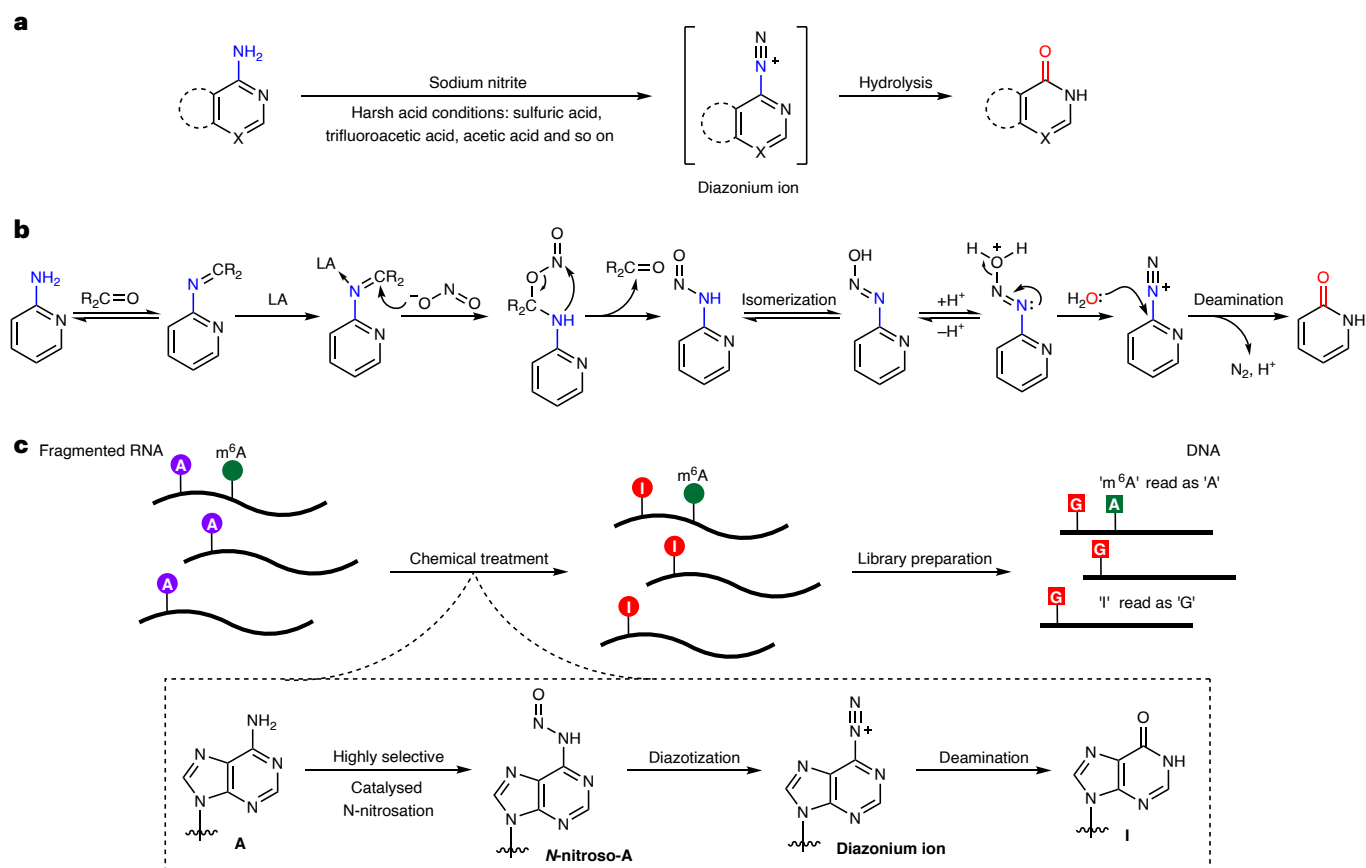
In nitrite ion-based RNA sequencing methods<sup>13,15–18</sup>, adenosine (A) or cytidine (C) is deaminated through stepwise nitrosation and diazotization under acidic conditions (Fig. 1a). NO-seq and NT-seq are simple, cost-effective chemical methods, but their low conversion efficiencies and harsh acid treatment can compromise detection accuracy and RNA integrity through degradation<sup>15,17</sup>. Recently, Wang and co-workers reported the chemical-based approach of GLORI for m<sup>6</sup>A sequencing, which uses a glyoxal protecting group on guanosine (G)

to substantially enhance deamination efficiency and selectivity for A in mammalian transcriptomes<sup>13,18</sup>. Despite being a major step forward, this method is still limited by RNA degradation under relatively harsh reaction conditions and reverse transcription stopping caused by incomplete glyoxal deprotection, necessitating a high input of RNA.

We sought to overcome the limitations of all previous nitrite ion-mediated nucleobase deamination approaches by designing a mild chemical deamination reaction that operates close to neutral pH, notably preserving RNA and improving the signal-to-noise ratio. While inspired by the evolved TadA-assisted *N*<sup>6</sup>-methyladenosine sequencing (eTAM-seq)<sup>12</sup>, GLORI<sup>13</sup> and the chemoselective deamination of unmethylated adenines in general<sup>15–17</sup>, we were also intrigued by a report from Keeper and Roller in 1973 on the carbonyl-catalysed N-nitrosation of secondary amines using nitrite ions under neutral and basic media<sup>19</sup>. Adopting their strategy, we additionally employed a Lewis acid to facilitate

<sup>1</sup>Department of Chemistry, The University of Chicago, Chicago, IL, USA. <sup>2</sup>Howard Hughes Medical Institute, The University of Chicago, Chicago, IL, USA.

<sup>3</sup>USA Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, USA. <sup>4</sup>Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA. <sup>5</sup>These authors contributed equally: Pingluan Wang, Chang Ye, Michelle Zhao. ✉e-mail: [chuanhe@uchicago.edu](mailto:chuanhe@uchicago.edu)



**Fig. 1 | Cooperative deamination of primary amines catalysed by a carbonyl organocatalyst and Lewis acid.** **a**, Nitrite ions mediate deamination through a sequential process of nitrosation and diazotization under harsh acidic conditions. **b**, This work and GLORI-seq<sup>13</sup>: cooperative deamination catalysed

by a carbonyl organocatalyst and Lewis acid (LA). **c**, Schematic of the m<sup>6</sup>A-CAM-seq method for whole transcriptome m<sup>6</sup>A sequencing. RNA samples were fragmented and treated under optimized conditions, followed by purification and sequencing library preparation.

N-nitrosation of the initial primary amine (Fig. 1b). We report here the development of a chemical sequencing method that mediates selective A-to-inosine (I) deamination in RNA to map m<sup>6</sup>A at base resolution under mild conditions (Fig. 1c). As the deamination is facilitated by cooperative catalysis using a carbonyl and Lewis acid as catalysts, we termed it chemical cooperative catalysis-assisted m<sup>6</sup>A sequencing (CAM-seq). CAM-seq promises a sensitive and robust mapping of m<sup>6</sup>A with extensive coverage by avoiding harsh acidic treatment of RNA, allowing for more comprehensive quantification of m<sup>6</sup>A sites using a low input of sample.

## Results

### General deamination of substrates under mild conditions

When applying the conditions of carbonyl-catalysed N-nitrosation to nucleobase deamination, we envisioned that a Lewis acid catalyst could enhance the reactivity between the carbonyl organocatalyst and the heteroaromatic amine of nucleobases. Through preliminary deamination screenings of various Lewis acids and carbonyl compounds using cytidine as the substrate, we identified the optimal reaction conditions for the deamination process (Supplementary Tables 1–3).

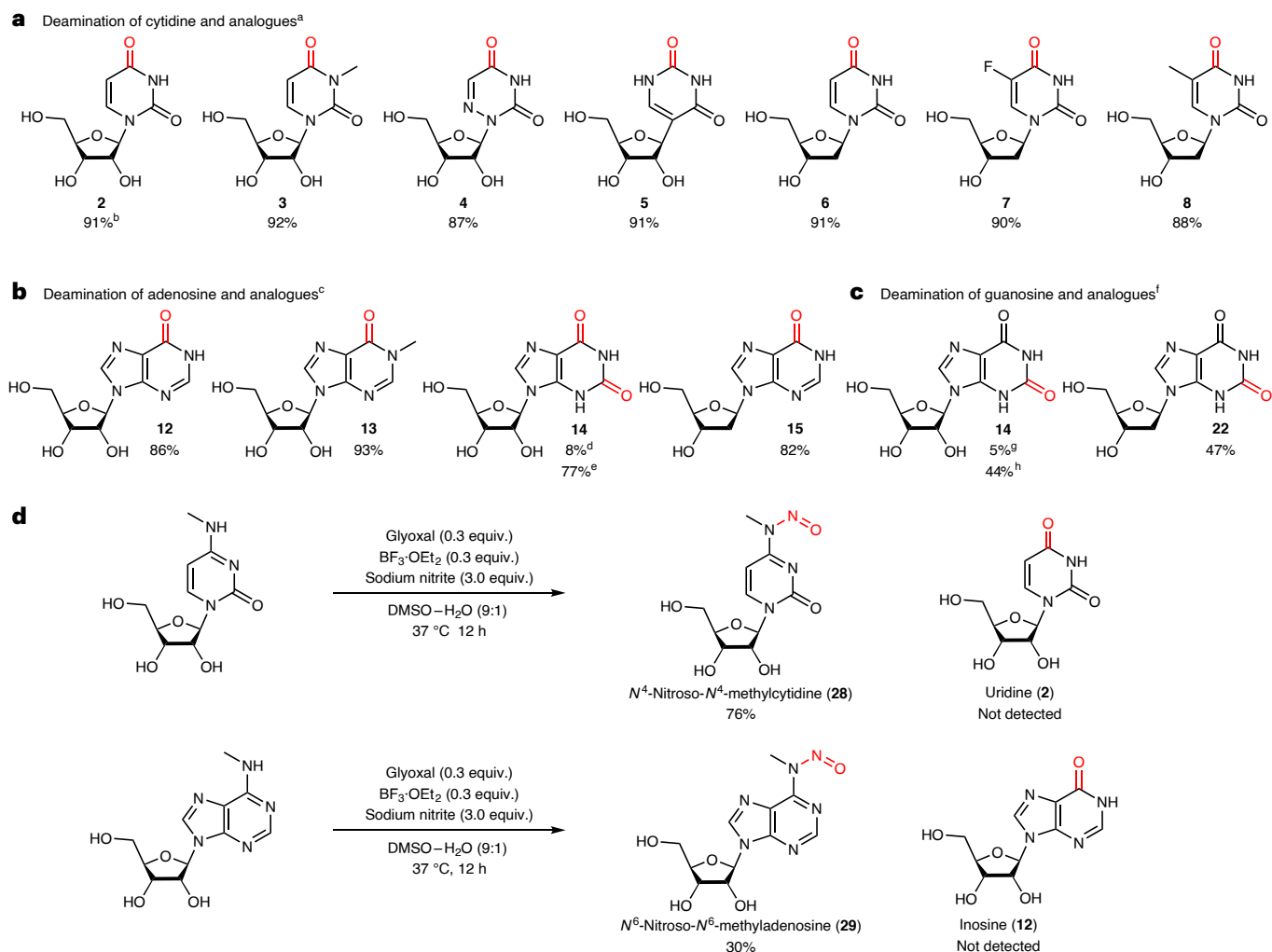
On assessing the deamination compatibility of A, C and G nucleobase analogues with our selected co-catalyst conditions, we found that glyoxal with boron trifluoride diethyl etherate (BF<sub>3</sub>·OEt<sub>2</sub>) facilitated the highly efficient deamination of a broad range of nucleobase substrates (Fig. 2a–c and Supplementary Fig. 2). Switching from a dicarbonyl organocatalyst to furfural, a monocarbonyl organocatalyst, substantially increased the yield of guanosine analogue deamination. We hypothesize that this occurs because guanosine-like products at position 6 can cyclize with glyoxal, forming a guanosine–glyoxal adduct (Supplementary Figs. 1a,b)<sup>20</sup>.

Reactions with m<sup>6</sup>A and N<sup>4</sup>-methylcytidine (m<sup>4</sup>C) were shown to stop at the nitrosylated intermediate stage, with no detection of deaminated inosine or uracil products (Fig. 2d), consistent with previous experiments that used nitrite to drive nucleobase deamination<sup>17</sup>; a highly efficient and selective conversion of A or C but not m<sup>6</sup>A or m<sup>4</sup>C under mild conditions would present a robust method to map these methylated bases. We noted that the rate of conversion of m<sup>6</sup>A to its N-nitroso intermediate was lower than that of A, which contradicts previous observations<sup>16</sup> and suggests a different mechanism behind the N-nitrosation step.

### Elucidation of the reaction mechanism

We first followed the reaction by <sup>15</sup>N nuclear magnetic resonance (NMR) spectroscopy, which revealed very poor condensation of [<sup>15</sup>N]adenosine with glyoxal in the absence of the Lewis acid even after 2 h of heating at 50 °C (Fig. 3a, middle). Addition of BF<sub>3</sub>·OEt<sub>2</sub> rapidly catalysed the conversion of <sup>15</sup>N-labelled adenosine to the intermediate **Int 3** within half an hour (Fig. 3a, bottom).

Two-dimensional <sup>1</sup>H–<sup>15</sup>N HMBC spectroscopy was used to monitor the Lewis acid-catalysed condensation with higher <sup>15</sup>N sensitivity (Supplementary Figs. 4a–c)<sup>21,22</sup>. This technique was also used to monitor the reaction steps from the imine intermediate (Supplementary Fig. 5a). In the presence of Na<sup>15</sup>NO<sub>2</sub>, the [<sup>15</sup>N]imine signal disappeared, indicating its transformation to an aliphatic nitro-substituted intermediate and subsequent rearrangement to the N-nitrosation adduct (Supplementary Fig. 5b). When unlabelled adenosine was used, a clear, single signal from <sup>15</sup>NO<sub>2</sub> was observed (Supplementary Figs. 1c and 5c). These results confirm the pathway through the imine intermediate.



**Fig. 2 | Deamination of nucleobases catalysed by an organocatalyst combined with a Lewis acid. a–c,** Substrate scope of the deamination reaction with cytidine (a), adenosine (b) and guanosine (c) and their analogues (products and yields are shown). <sup>a</sup>Deamination conditions A for cytidine and analogues: glyoxal (0.3 equiv.),  $\text{BF}_3 \cdot \text{OEt}_2$  (0.3 equiv.), nitrite ion (3.0 equiv.) in dimethylsulfoxide (DMSO)– $\text{H}_2\text{O}$  (9:1) at 37 °C for 12 h. <sup>b</sup>Isolated yield. <sup>c</sup>Deamination conditions B for adenosine and analogues: glyoxal (0.3 equiv.),  $\text{BF}_3 \cdot \text{OEt}_2$  (0.3 equiv.), nitrite

ion (3.0 equiv.) in DMSO– $\text{H}_2\text{O}$  (9:1) at 37 °C for 12 h. <sup>d</sup>2-Aminoadenosine was used as the starting material. <sup>e</sup>Isoguanosine was used as the starting material. <sup>f</sup>Deamination conditions C for guanosine and analogues: furfural (1.0 equiv.),  $\text{BF}_3 \cdot \text{OEt}_2$  (1.0 equiv.), nitrite ion (3.0 equiv.) in DMSO– $\text{H}_2\text{O}$  (9:1) at 50 °C for 12 h. <sup>g</sup>Guanosine deamination using conditions A. <sup>h</sup>Guanosine deamination using conditions C. **d,** Reactions of methylated nucleosides under the optimized catalytic conditions. See Supplementary Materials for details.

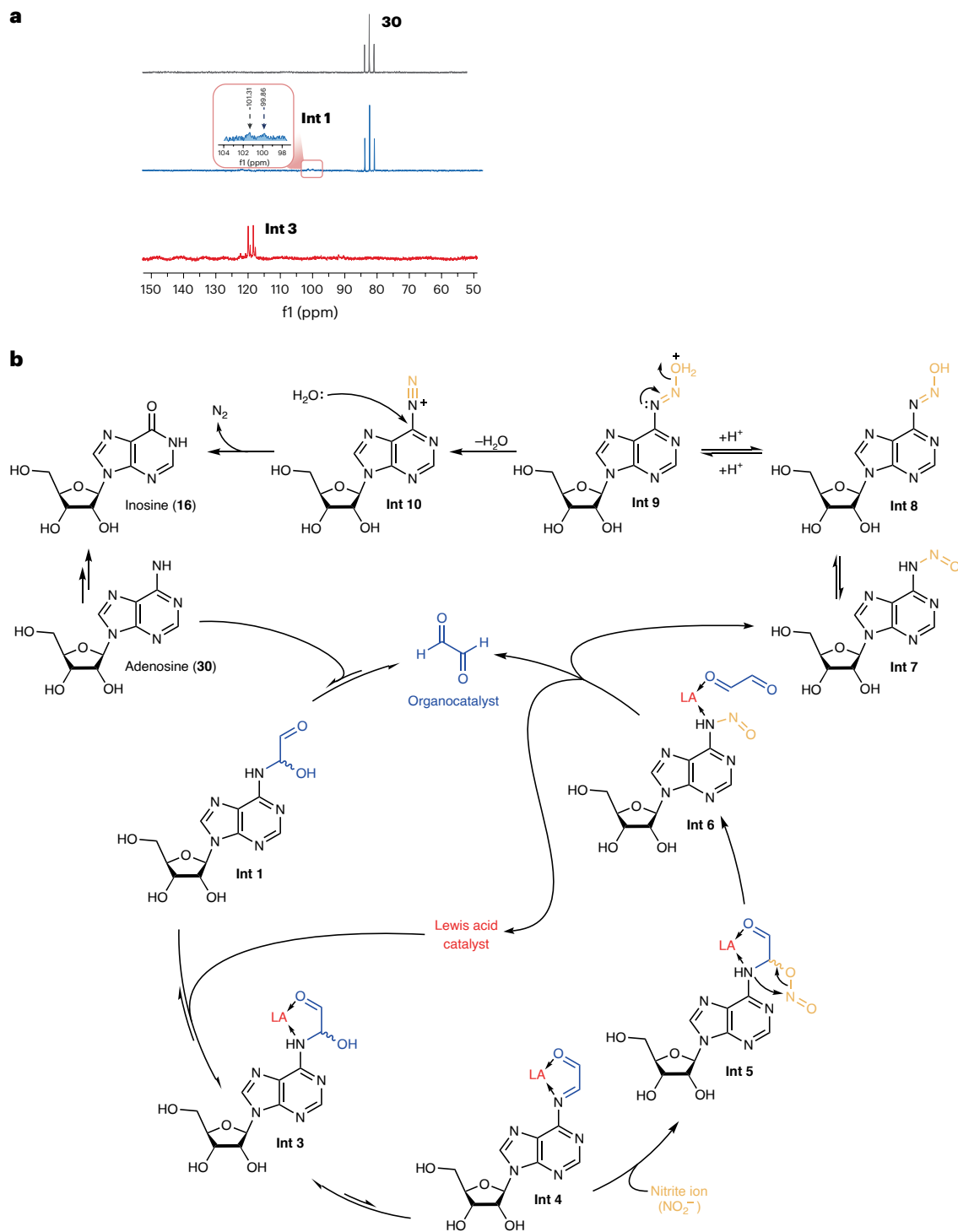
Building on these observations and previous studies of carbonyl-catalysed N-nitrosation, we proposed the reaction pathway depicted in Fig. 3b. Initially, the carbonyl catalyst and starting substrate condense to form a hemiaminal (**Int 1**), a process accelerated by the Lewis acid. Following dehydration to the imine intermediate **Int 4**, the nitrite ion interacts with the imine's lowest unoccupied molecular orbital to form the C–O–N–O intermediate **Int 5**. This intermediate rearranges to form an N–NO bond, culminating in the formation of nitrosamine **Int 7** and the release of the organocatalyst and Lewis acid. <sup>11</sup>B NMR experiments<sup>23</sup> suggested that the Lewis acid could interact with both the carbonyl co-catalyst and the imine intermediate (Supplementary Fig. 3), as shown in **Int 3–Int 5**. Inosine (**16**) emerges after diazotization and deamination under slightly acidic conditions.

**Selective conditions for adenine deamination in DNA and RNA**  
Our group has previously developed various enzyme-assisted methods for mapping m<sup>6</sup>A modifications at single-base resolution, including selective allyl chemical labelling and sequencing (SAC-seq)<sup>11,24</sup> and eTAM-seq<sup>12</sup>. However, we recognized the potential of a highly efficient, selective chemical deamination reaction as a cost-efficient alternative for m<sup>6</sup>A mapping,

such as CAM-seq, involving carbonyl and Lewis acid co-catalysts under mild conditions. To make CAM-seq specific to m<sup>6</sup>A, we further explored the conditions to optimize its selectivity for adenine over cytosine and guanine in oligonucleotides, as well as the rate of deamination.

We tested different Lewis acids by measuring the mutation ratio at adenine sites in an oligonucleotide. In contrast to the general deamination conditions, we found that boric acid exhibited the highest efficiency in enhancing reactivity (Fig. 4a and Supplementary Fig. 6a). With boric acid, dicarbonyl compounds achieved more efficient deamination than monocarbonyl compounds (Fig. 4b and Supplementary Fig. 6b). Using a liquid chromatography–tandem mass spectrometry (LC–MS/MS) assay, both trifluoropyruvic aldehyde and glyoxal demonstrated selectivity for adenine over cytosine (Fig. 4c and Supplementary Fig. 6c), but due to the notable DNA/RNA degradation caused by trifluoropyruvic aldehyde, glyoxal was selected for further study. This preference was similarly observed in RNA oligonucleotides (Supplementary Fig. 7a), confirming that dicarbonyl compounds are more reactive than monocarbonyl compounds.

We further evaluated these reactions using matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI–TOF



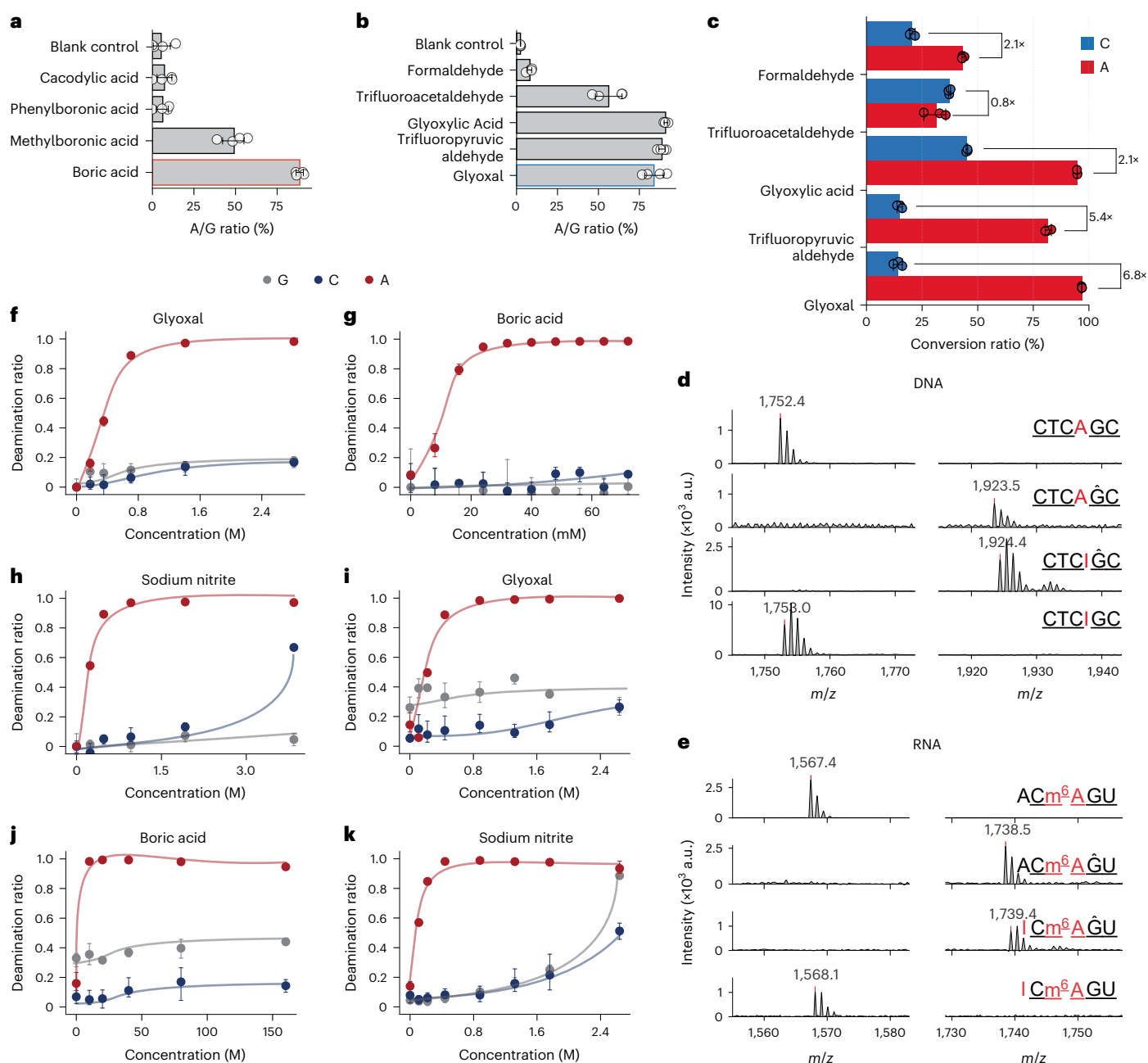
**Fig. 3 | Mechanistic investigations and catalytic cycle. a**, One-dimensional  $^{15}\text{N}$  NMR spectra of the deamination of  $^{15}\text{N}^6$ adenosine.  $^{15}\text{N}$  NMR spectra of  $^{15}\text{N}^6$  adenosine (top), the condensation product formed between  $^{15}\text{N}^6$ adenosine and the organocatalyst (middle) and the condensation product formed between

$^{15}\text{N}^6$ adenosine and the organocatalyst under Lewis acid catalysis (bottom). Inset: expansion of the  $^{15}\text{N}$  NMR spectrum around 100 ppm for **Int 1**. **b**, Proposed catalytic cycle for the deamination reaction.

MS)<sup>25</sup>. Complete deamination of A sites to inosines was observed in both DNA and RNA, whereas m<sup>6</sup>A sites remained unchanged (Fig. 4d,e). After optimizing the deamination conditions (Fig. 4f–k), glyoxal remained the most selective organocatalyst, as confirmed by the deep sequencing of a 35-base RNA oligonucleotide (Supplementary Fig. 7a).

Our optimized catalyst combination proved very robust across various temperatures and pH values, maintaining complete deamination

at A sites with slightly reduced deamination at C and G sites as the temperature or pH was decreased (Supplementary Fig. 8a,b). These optimized conditions offer a deamination reaction under nearly neutral conditions, minimizing the degradation of biomacromolecules. Ammonium nitrites showed less deamination at C sites than sodium nitrite (Supplementary Fig. 8c), indicating the potential for even more selective conditions for adenosine deamination in RNA, although further optimization is required.



**Fig. 4 | Optimization of deamination in oligonucleotides.** **a, b**, The mutation ratio at adenine sites in an oligonucleotide for various Lewis acid catalysts (**a**) and carbonyl organocatalysts (**b**). The boric acid (red outline) and glyoxal (blue outline) catalysts were selected for further optimization. The error bars represent the 95% confidence interval ( $n = 4$  technical replicates). **c**, Comparison of the catalytic efficiency of the deamination of A and C with various carbonyl organocatalysts. Dicarbonyl catalysts led to higher conversions on A sites than on C sites. The error bars represent the 95% confidence interval ( $n = 3$  technical replicates), and the relative fold change of A conversion to C conversion is labeled. **d**, A 6-base DNA oligonucleotide CTCAGC under the optimized conditions shows high reactivity in both G protection and A deamination. **e**, A 5-base m<sup>6</sup>A RNA oligonucleotide probe (ACm<sup>6</sup>AGU) was used to demonstrate the selectivity between A and m<sup>6</sup>A under conditions similar to those used in **d**. Protected G sites are indicated by ^ and deaminated A sites are underlined.

a.u., arbitrary units. **f–h**, Optimization of glyoxal (**f**), boric acid (**g**) and nitrite (**h**) concentrations for the deamination reaction of a DNA oligonucleotide, monitored via nucleotide composition analysis using LC–MS/MS ( $n = 3$  biological replicates). The data are presented as the mean with 95% confidence intervals. Increased concentrations of glyoxal and boric acid did not affect the selectivity between A and G/C; however, higher nitrite concentrations substantially increased the conversion of C. **i–k**, Optimization of glyoxal (**i**), boric acid (**j**) and nitrite (**k**) concentrations for the deamination reaction of an RNA oligomer, monitored via nucleotide composition analysis using LC–MS/MS ( $n = 3$  biological replicates). The data are presented as the mean with 95% confidence intervals. Increased concentrations of glyoxal and boric acid did not affect the selectivity between A and G/C; however, higher nitrite concentrations substantially increased the conversion of C.

### Optimized conditions for m<sup>6</sup>A mapping (CAM-seq)

We noticed that, despite the effectiveness of glyoxal as an organocatalyst, its tendency to form adducts with guanine, followed by incomplete deprotection, could cause reverse transcriptase to stop during reverse

transcription (known as RT stops). This is a limitation faced by GLORI in which glyoxal is used as a protecting group for G to minimize G-to-X conversions and ensure a selective deamination of adenosine. Therefore, we decided to use kethoxal as a protecting moiety due to its highly



selective yet reversible reaction with G<sup>25,26</sup>. Glyoxal-caged RNA oligonucleotides showed a higher rate of RT stops, with 75% RT stops at the first caged G compared with 40% for the first kethoxal-caged G in RNA oligonucleotides (Fig. 5a,b). Kinetic studies of the DNA oligonucleotide template revealed that kethoxal displayed a caging and decaging rate 32 times greater than that of glyoxal under neutral conditions (Fig. 5c–f). Stability of the kethoxal protecting group could be enhanced under slightly acidic pHs, while a slightly basic pH facilitated easier removal (Supplementary Figs. 9j–l). Most importantly, glyoxal showed no ability to displace kethoxal from the fully caged kethoxal–guanine complex in competitive assays (Fig. 5g), which reinforced our preference for kethoxal as the protecting group for guanine during CAM-seq.

The acidic conditions used in previous nitrite deamination methods can cause notable RNA degradation<sup>13,15–17</sup>. We found that operating CAM-seq at a neutral to slightly acidic pH substantially reduced RNA degradation. While the optimized conditions caused a small amount of degradation when using long, full-length messenger RNA (Fig. 5h), negligible degradation was observed using 200-nucleotide mRNA fragments as input RNA (Fig. 5i).

Despite CAM-seq reagents achieving nearly 100% conversion efficiency in LC–MS/MS assays, ~1% of A sites remained unchanged in sequencing data. We reasoned that RT reactions may also yield false positives. Partial pairing of inosine with 2'-deoxythymidine (dT) could lead to the misinterpretation of m<sup>6</sup>A as unmodified A, and RT stops at inosine sites could create bias in quantification<sup>27,28</sup>. To address potential RT-induced artefacts, we examined 11 different reverse transcriptases and 3 different dNTP ratios, that is, 2'-deoxythymidine triphosphate (dTTP) to 2'-deoxycytidine triphosphate (dCTP) ratios of 1:1, 1:40 and 1:400. CAM-seq processing of 10 ng polyA-tailed RNA from HEK293T cells spiked in with m<sup>6</sup>A-modified RNA oligonucleotide (Supplementary Note I) revealed varying propensities of reverse transcriptases to incorporate dT during RT reactions (Fig. 5j and Supplementary Figs. 10–12). Considering that the conversion ratio and RT efficiency could be affected by local sequence context, representing the overall performance with a single oligonucleotide sequence may be overly simplistic. We therefore calculated the unconverted ratio of all A sites in the human transcriptome for each condition to better estimate the signal-to-noise ratio (Supplementary Fig. 12). GAC motifs, which are highly enriched for m<sup>6</sup>A modifications, served as a positive control, while UAG motifs, less enriched, were used as a negative control. Comparing signal-to-noise ratios led us to select condition RevertAid RT with a 1:40 dTTP/dCTP ratio for its ~fourfold decrease in background noise, with the signal from GAC motifs reduced by less than one-fifth. Under these optimized RT conditions, the background noise from unmodified A sites was less than 0.5% (Supplementary Fig. 12).

### Comprehensive m<sup>6</sup>A mapping in human transcriptome

The progression of m<sup>6</sup>A sequencing technologies has evolved over the past 10 years with notable advances in resolution<sup>29–31</sup> and quantification<sup>30,31</sup>. The latest techniques<sup>11–13</sup> combine single-base accuracy with quantification but still face some challenges in bias. The lessened effectiveness of SAC-seq for AAC motifs<sup>11</sup> can potentially

produce false negatives, while eTAM-seq<sup>12</sup> and GLORI-seq<sup>13</sup> encounter a background noise of 1–2% that escalates to >10% in highly structured regions such as ribosomal RNA (Supplementary Fig. 13). To attain a thorough grasp of the m<sup>6</sup>A landscape across the transcriptome, a method with much lower background noise devoid of false negatives is essential. Our optimized CAM-seq protocol selectively converts adenosine sites into inosine; after RT, inosine is paired with dCTP and read as an A-to-G mutation, while m<sup>6</sup>A reads as A after PCR (Supplementary Fig. 14). The optimized protocol also dramatically reduces RNA degradation and background noise, enabling its application to low input samples. Thus, CAM-seq was used to map m<sup>6</sup>A throughout the transcriptome in both human and plant samples using 10 ng polyA-enriched RNA.

Using the optimized RT conditions (RevertAid RT, 1:40 dTTP/dCTP ratio) for the human HEK293T cell line, we identified 282,281 highly confident m<sup>6</sup>A sites ( $P < 0.001$ ). Of these, 235,173 and 180,056 sites exhibited modification ratios greater than 5% and 10%, respectively (Supplementary Fig. 15a,b). The GAC motif emerged as the most enriched, with 152,292 sites (54% of confident sites) within the HEK293T transcriptome (Fig. 6a), followed by the AAC motif (underlined A indicates modified site), which accounted for 76,475 sites (27%). Notably, YAC (CAC/UAC) and RAU (GAU/AAU) motifs, which differ by only one nucleotide from the RAC motif, also showed a degree of m<sup>6</sup>A modification. Other motifs, such as UAA/UAG/UAU, showed relatively rare m<sup>6</sup>A modifications. For example, of all the detected m<sup>6</sup>A sites, only 0.054% were in UAG motifs. m<sup>6</sup>A-enriched motifs such as GAC, but not m<sup>6</sup>A-rare motifs such as UAG, exhibited a distinct enrichment signal at the 3'-UTR region near the stop codon in the human transcriptome (Fig. 6a and Supplementary Fig. 15c,d), further confirming that CAM-seq provides high-quality m<sup>6</sup>A mapping across the entire transcriptome. We also calculated the ratio of UAG to GAC as a measure of the signal-to-noise level and for comparison with other methods.

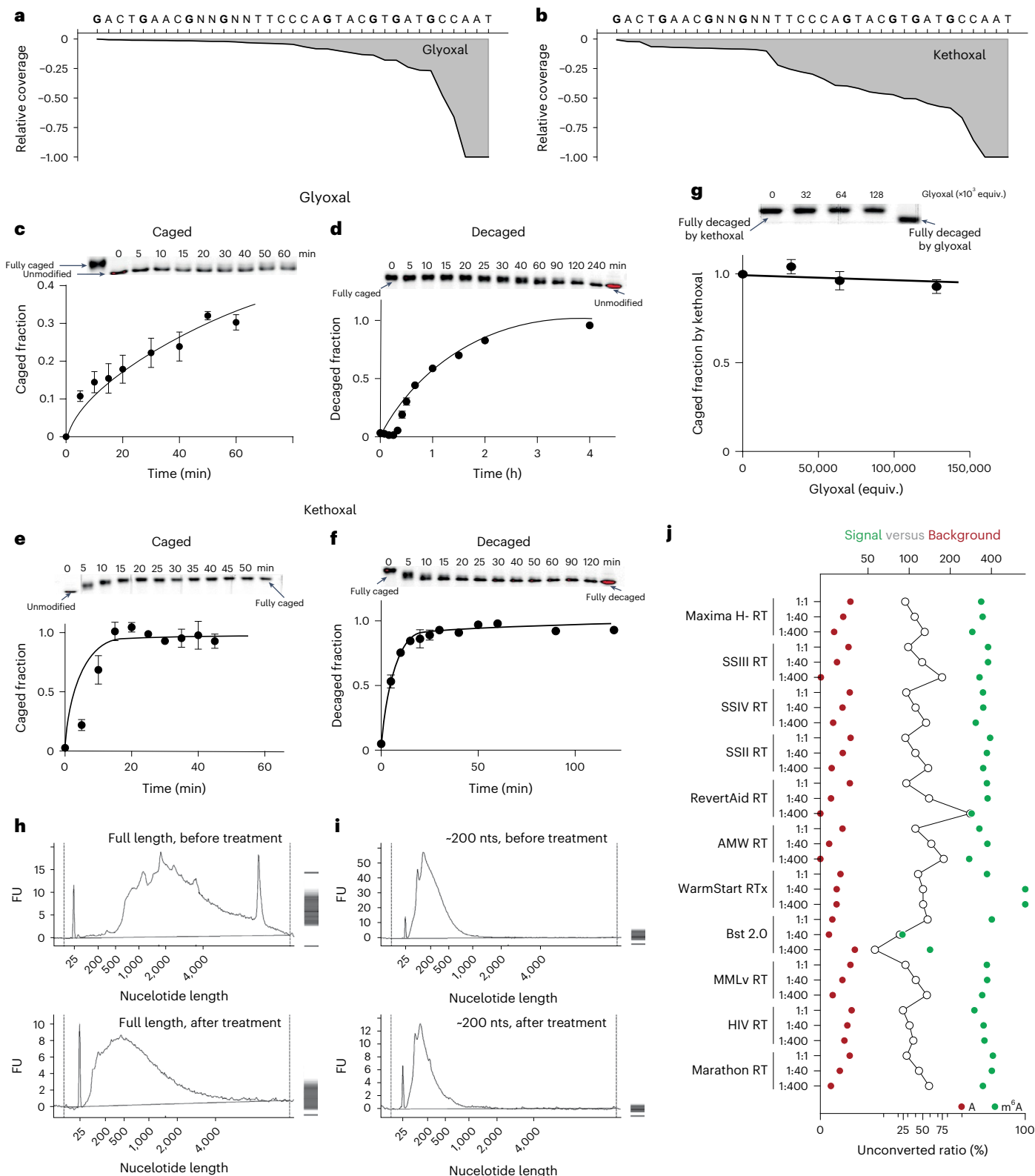
This study presents the most comprehensive dataset for m<sup>6</sup>A detection in the human transcriptome so far (Fig. 6b). The signal-to-noise level of CAM-seq, calculated on the basis of the ratio of m<sup>6</sup>A on the UAG motif to m<sup>6</sup>A on the GAC motif, is approximately 0.05%, which is ten times lower than published results obtained in the same HEK293T cell line using the other two deamination-based methods, eTAM-seq and GLORI-seq. Enzyme-based methods, such as MAZTER-seq, m<sup>6</sup>A-miCLIP-seq, DART-seq and SAC-seq, exhibit motif bias, failing to detect all m<sup>6</sup>A sites. These results align with the observation in previous studies that a notable proportion of sites show an overestimation of m<sup>6</sup>A levels (Fig. 6c), probably due to incomplete deamination. Before the development of m<sup>6</sup>A sequencing methods, liquid chromatography–mass spectrometry (LC–MS) determined that approximately 30% of m<sup>6</sup>A sites were located in AAC motifs<sup>32</sup>; however, it has been challenging to validate this observation using high-throughput sequencing data due to the potential for false positives and false negatives. The ability of CAM-seq to uncover low-ratio motifs and eliminate false positives provides a more accurate alignment with LC–MS observations. By leveraging the CAM-seq data, we calculated the overall motif composition by sum of the (m<sup>6</sup>A ratio × site

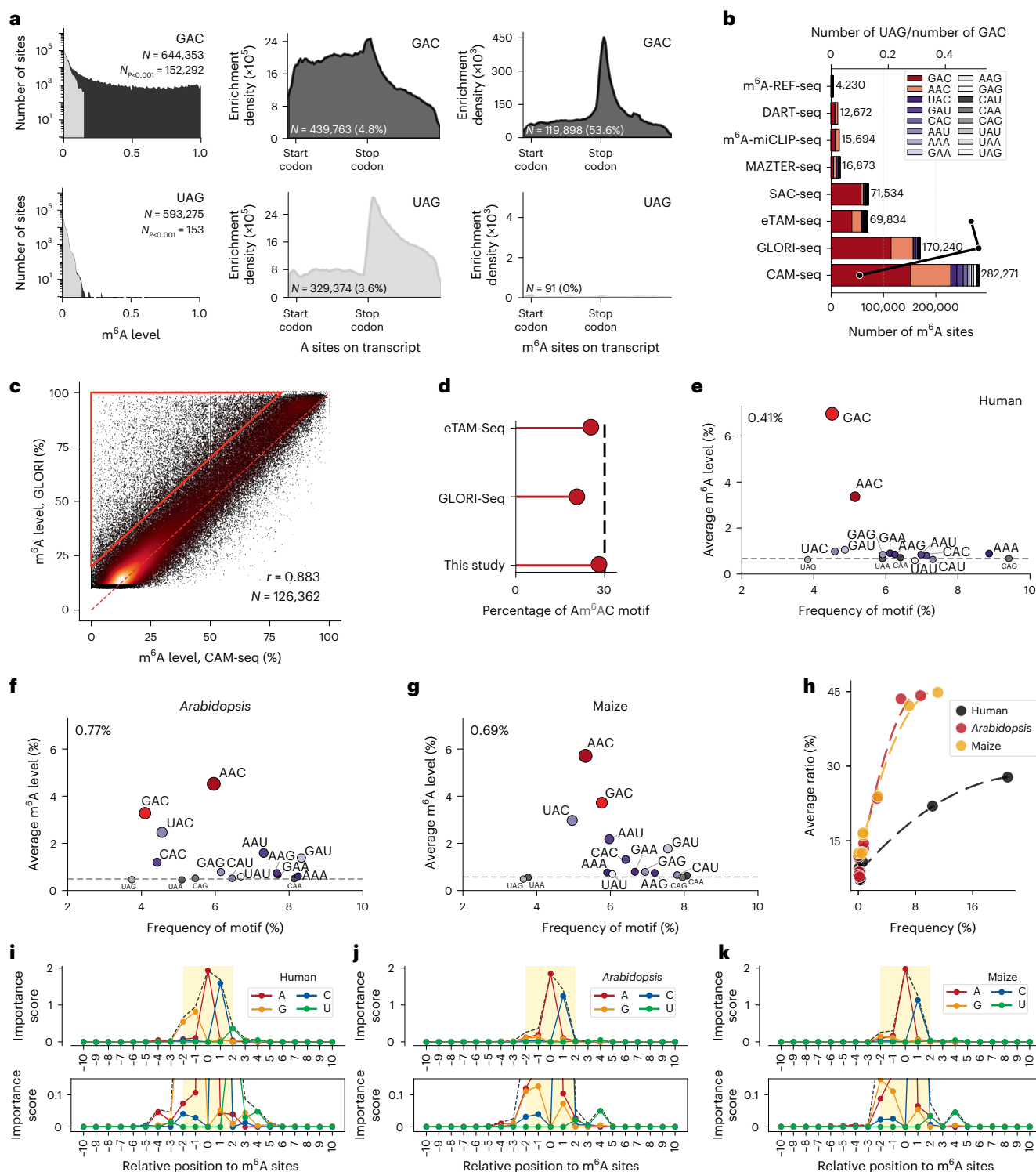
**Fig. 5 | Optimization of the CAM-seq protocol.** **a**, RT stopping was tested on an RNA oligonucleotide probe with G sites protected by glyoxal. The RT stop ratio reached approximately 75% at the first glyoxal-protected G site. **b**, A kethoxal-protected RNA oligonucleotide probe was used to demonstrate the RT stop ratio under conditions similar to those in **a**. The RT stop ratio reached only 40% at the first kethoxal-protected G site. **c–f**, Comparison of the caging (**c,e**) and decaging (**d,f**) kinetics of a DNA oligonucleotide using glyoxal (**c,d**) and kethoxal (**e,f**) under neutral conditions. The oligonucleotide was separated using PAGE gel, and the caged/decaged fraction was estimated from intensity. Data are presented as the mean with 95% confidence interval. **g**, Competition for the kethoxal fully caged DNA oligonucleotide by glyoxal under neutral conditions. The oligonucleotide was separated using PAGE gel, and the caged/decaged

fraction was estimated from intensity. Data are presented as the mean with 95% confidence interval. **h,i**, RNA degradation assay using both full-length RNA (**h**) and 200-nucleotide fragmented RNA (**i**). The nucleotide length (nt) is shown on x-axis, and the RNA concentration measured by fluorescence units (FU) is shown on the y-axis. The gel image from the fragment analyzer is shown on the right. Following chemical treatment, full-length mRNA shows partial degradation to a 500-nt fragment, while fragmented RNA does not show notable degradation. **j**, Optimization of the reverse transcriptase condition was achieved through spike-in oligonucleotide sequencing data. Eleven reverse transcriptases and three dTTP/dCTP ratios (1:1, 1:40 and 1:400) were assayed. A stand-out condition (RevertAid RT at a 1:40 ratio) exhibited a background noise level below 0.36%. Open circles represent the signal-to-background ratio.

coverage of m<sup>6</sup>A sites on the AAC motif) over the sum of the (m<sup>6</sup>A ratio × site coverage of all m<sup>6</sup>A sites), deriving values of 25.4%, 20.5% and 28.1% for eTAM-seq, GLORI-seq and CAM-seq, respectively (Fig. 6d). These results suggest that CAM-seq can accurately uncover the absolute modification profile of a given sample, more accurately aligning with LC-MS observations and providing a reliable representation of the m<sup>6</sup>A landscape.

**Cross-species study uncovers motif-dependent m<sup>6</sup>A variation**  
Motivated by this observation, we investigated whether this m<sup>6</sup>A modification profile at the motif level is consistent across animal and plant species. *Arabidopsis* and maize RNA samples were extracted and sequencing libraries were prepared using the optimized RT condition (RevertAid RT, 1:40 dTTP/dCTP ratio). With a stringent cut-off *P* value of <0.001, we detected 111,121 and 141,990 m<sup>6</sup>A sites in *Arabidopsis*





**Fig. 6 | Accurate quantification of m<sup>6</sup>A modification reveals m<sup>6</sup>A deposition at the motif level.** **a**, Distribution of m<sup>6</sup>A modification levels for all A sites (coverage > 20), focusing on GAC and UAG motifs as examples. Left: sites with  $P < 0.001$  are marked in black and non-significant sites are marked in grey. The total number of A sites ( $N$ ) and filtered m<sup>6</sup>A sites ( $N_{P < 0.001}$ ) are indicated. **b**, Comparison of detected m<sup>6</sup>A sites and estimated background noise for multiple methods. The bars show the total number of m<sup>6</sup>A sites, with GAC and AAC motifs in shades of red and other motifs in shades of purple. The black line indicates the UAG/GAC ratio, used as an estimate of the upper bound for false positives. **c**, Correlation between CAM-seq and GLORI-seq m<sup>6</sup>A levels for 126,362 overlapping sites ( $r = 0.883$ ). Red dashed line mark the equal levels. The area marked by the red triangle highlights sites where GLORI-seq overestimates m<sup>6</sup>A levels. **d**, Fraction of AAC motifs among all m<sup>6</sup>A sites observed in three

deamination-based methods, calculated as the sum of (m<sup>6</sup>A ratio × coverage) for AAC sites over the sum of (m<sup>6</sup>A ratio × coverage) for all sites. **e**, Average methylation level versus relative motif frequency across all A sites (no filter) for human samples. GAC and AAC motifs (red) stand out above the background. The overall m<sup>6</sup>A level is indicated as 0.41%. **f, g**, Average methylation level versus relative motif frequency across all A sites for *Arabidopsis* (f) and maize (g). The overall m<sup>6</sup>A levels are indicated as 0.77% and 0.69%, respectively. **e–g**, the dotted-dashed lines indicate the background noise, defined by the rarely methylated motifs (UAG/UAA/CAG/CAA). **h**, Relationship between motif frequency and average methylation level for the 16 three-letter motifs passing filters in human, *Arabidopsis* and maize. **i–k**, Nucleotide importance scores at each position relative to m<sup>6</sup>A sites in human (i), *Arabidopsis* (j) and maize (k). The dashed black lines show the total importance at each position.



(Supplementary Fig. 16a,b) and maize (Supplementary Fig. 17a,b) samples, respectively. Similar to the human transcriptome, RAC motifs were the motifs most prevalently enriched with m<sup>6</sup>A in these plants. However, RAC motifs constitute only 40% of all m<sup>6</sup>A sites in plants, compared with about 85% in human samples. Interestingly, the AAC motif is the most abundant in *Arabidopsis* and maize, in contrast to the GAC motif in humans. Motifs that differ by one nucleotide, such as YAC and RAU, account for ~50% of the sites in plants, whereas they represent only ~10% in humans. This variation in m<sup>6</sup>A motif enrichment between plants and humans raises questions about whether m<sup>6</sup>A regulation is less stringent in the plant kingdom. On further investigation, the frequency of one m<sup>6</sup>A-rare motif (UAG) was found to be 0.043% and 0.051% in *Arabidopsis* and maize, respectively. These observations suggest that m<sup>6</sup>A writers in plants may exhibit a wider motif spectrum, but are still tightly regulated, similarly to humans (Fig. 6e–g).

The frequencies of m<sup>6</sup>A sites within all three-letter motifs were analysed against their average modification level, revealing an overall positive correlation. This means that motifs with higher abundances tend to have higher m<sup>6</sup>A modification levels (Fig. 6h). Interestingly, the plant samples exhibited a nearly identical trend, while the data for humans showed a smaller slope. This indicates that the m<sup>6</sup>A modification in human samples is more evenly distributed than in plants, where the m<sup>6</sup>A modification is more concentrated at specific sites. The analysis of five-letter motifs showed similar trends (Supplementary Figs. 15b–17b). For example, GGACT and CGACT both belong to the GAC motif, but 34,261 versus 30 modified sites were detected, with modification levels of 38% versus 7%, respectively. This suggests that the flanking bases at positions –2 and +2, in addition to those at positions –1 and +1, play important roles in regulating m<sup>6</sup>A deposition in humans. To unravel the rules governing m<sup>6</sup>A deposition at the motif level, we developed an importance score for each nucleotide at positions around the m<sup>6</sup>A site. The importance score represents the relative contribution of each position within the motifs, ranging from nucleotides –10 to +10 flanking the m<sup>6</sup>A site, based on the level of m<sup>6</sup>A methylation of each motif through the calculation of entropy (see Methods). The modification level is considered in this calculation rather than just the relative motif frequency. Consistent with previous studies, the modification level in the human transcriptome is significantly influenced by the cytosine (C) base at position +1 (Fig. 6i). In addition, the guanine (G) base at positions –2 and –1 and uracil (U) at position +2 notably affect the modification level of m<sup>6</sup>A motifs. Surprisingly, positions –4, –3, +3 and +4 also contribute to the m<sup>6</sup>A level: adenine (A) at position –4 and uracil (U) at position +4 are associated with a higher m<sup>6</sup>A modification level, whereas positions –3 and +3 show no clear nucleotide preference. Distal sites more than four nucleotides away do not significantly impact the m<sup>6</sup>A level.

Similar analyses of *Arabidopsis* (Fig. 6j) and maize (Fig. 6k) samples revealed that, despite the evolutionary diversity of these two plants, the rules for m<sup>6</sup>A deposition are relatively conserved. In contrast to human samples, only cytosine (C) at position +1 showed a significant importance score in these plants. Adenine (A) at position –1 plays a more crucial role than guanine (G) at the same position, while adenine (A) and guanine (G) at position –2 demonstrate comparable levels of importance. Interestingly, although position +2 contributes to m<sup>6</sup>A deposition, its total importance score is as low as that of position +4. Moreover, positions +2 and +3 also do not exhibit a preference for any specific nucleotide, although uracil (U) at position +4 contributes more to the m<sup>6</sup>A level, in alignment with the observations in human samples.

### Saturated profiling links transcription rate to m<sup>6</sup>A level

Despite extensive research into the role of motif patterns in regulating m<sup>6</sup>A levels, the variability in the degree of methylation among identical motifs remains unexplained. Over the past decade, several studies have identified factors that regulate the level of m<sup>6</sup>A site modification, such as the proximity to the stop codon, internal exon length, last exon

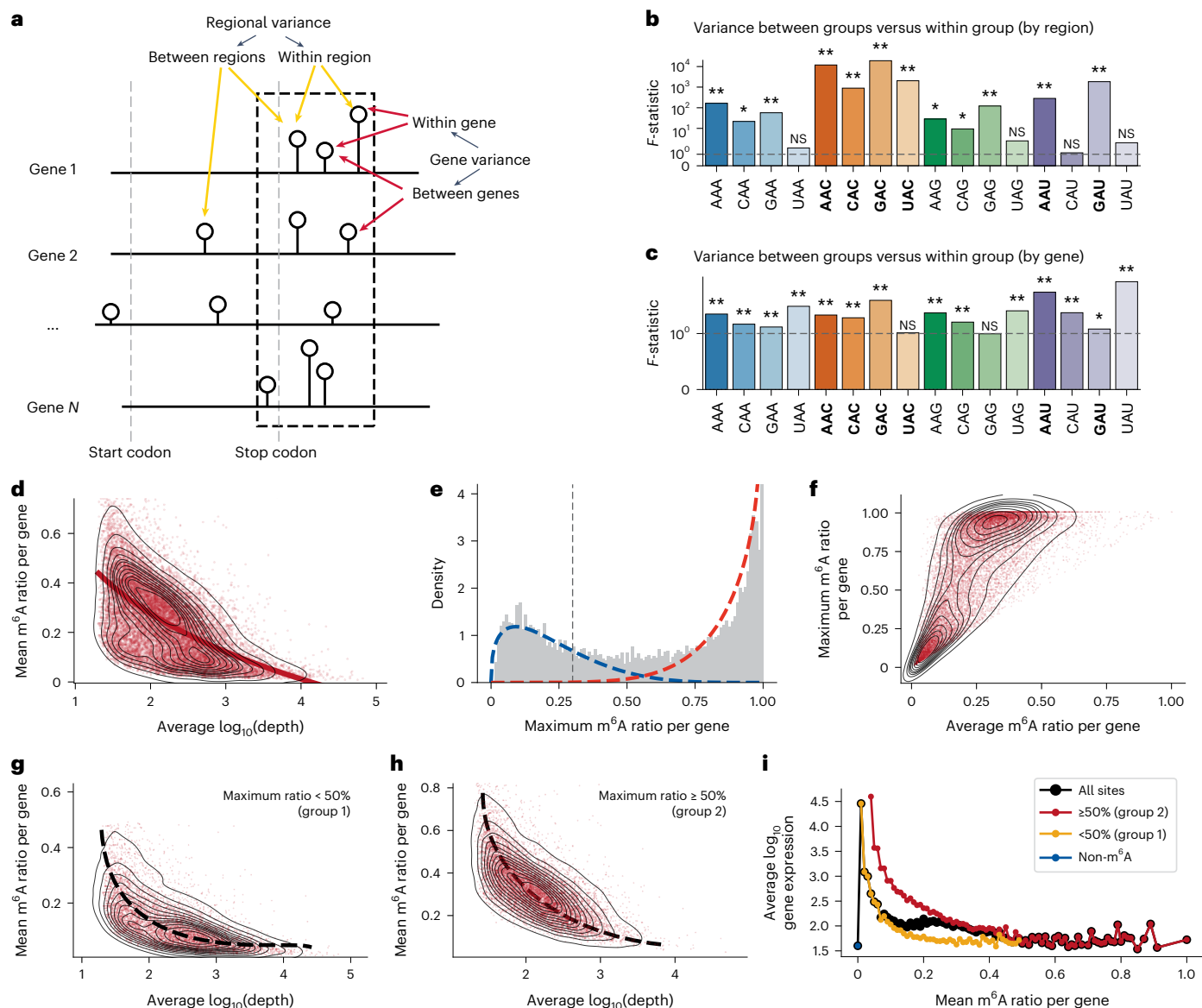
localization and the presence of the exon junction complex. We refer to these influencing factors as regional variance, which can be demonstrated through global-scale statistical analyses, even in noisy and sparse datasets (Fig. 7a). For instance, by comparing the modification levels of sites within the –20 to +180 region near the stop codon with those outside this region, we observed that the regional distribution significantly impacts most A motifs (Fig. 7b).

Another variance exists at the gene level. Even when identical regions are considered, with all other features such as the distance to the stop codon being constant, notable differences in levels of m<sup>6</sup>A methylation are observed between genes (Fig. 7a). Previous challenges in studying this variance arose due to the absence of methods capable of generating a saturated map of m<sup>6</sup>A modifications, making the confirmation of m<sup>6</sup>A sites across genes uncertain. Using the saturated m<sup>6</sup>A dataset in this Article (Supplementary Figs. 18a,b), we can now accurately quantify the modification of all A sites with sufficient sequencing coverage. Our analysis revealed that gene-level factors significantly affect m<sup>6</sup>A variance (Fig. 7c), prompting an investigation into how genes influence m<sup>6</sup>A levels.

We observed a generally negative association between gene expression level and average m<sup>6</sup>A level (Fig. 7d). However, analysis of the density of the genes showed that the data points for the genes could be roughly separated into two clusters. This observation led us to speculate that the average m<sup>6</sup>A level of different genes might follow a bimodal distribution, but that was not the case (Supplementary Fig. 19c). We further reasoned that the level of modification of m<sup>6</sup>A could be influenced by motif pattern, regional distribution and the length of exons, and that variance of these factors might obscure comparison of m<sup>6</sup>A differences at the gene level. Consequently, the maximum modification level of a gene, indicative of its maximum capacity for m<sup>6</sup>A modification, might be a better parameter for distinguishing genes. When distinguishing genes by their maximum modification level, we found that all m<sup>6</sup>A-modified genes in the human transcriptome could be divided into two distinct groups (Fig. 7e,f): one with generally low modification across all m<sup>6</sup>A sites and the other with the potential for highly modified sites. Thus, we attempted to classify these genes into two groups based on whether the maximum modification level is below or above 50% (group 1 and group 2, respectively). Both groups showed a notable negative correlation between gene expression level and m<sup>6</sup>A level (Fig. 7g,h). We propose two possibilities to explain this observation: (1) genes with high m<sup>6</sup>A modification levels tend to be degraded, leading to a systematic drop in the expression level, and (2) genes with high expression levels show high transcription rates, and within cells, a higher transcription rate may contribute to a lower probability of methylation. Comparison of non-m<sup>6</sup>A-modified genes, group 1 genes and group 2 genes revealed that group 2 genes, with a higher capacity for m<sup>6</sup>A modification, showed higher expression levels than group 1, and group 1 genes showed higher expression levels than non-m<sup>6</sup>A genes (Fig. 7i). These observations suggest that m<sup>6</sup>A-mediated degradation may not be the primary factor in this trend, while the transcription rate might play an important role.

## Discussion

We have reported here a nucleoside deamination reaction involving cooperative catalysis with a carbonyl organocatalyst and Lewis acid. We have shown that this catalytic reaction proceeds through N-nitrosation, diazotization and denitrogenation. Our mechanistic studies revealed that boric acid acts as a Lewis acid during deamination, implying a role beyond that of a protective additive for guanine, as previously reported<sup>15</sup>. The harsh acidic reaction conditions of previous deamination procedures have limited their application in RNA modification sequencing<sup>14,16</sup>. We have shown here that the use of a boric acid buffer to adjust the pH to nearly neutral is crucial to maximize the conversion of adenine and minimize RNA degradation under the mild deamination conditions.



**Fig. 7 | Factors affecting  $m^6A$  variation.** **a**, Illustration showing how the regional and gene distribution of A sites may affect the variance of the  $m^6A$  modification level in genes. **b**, The region from  $-20$  to  $+180$  nucleotides near the stop codon is identified as an  $m^6A$  deposition hotspot based on metagenesis analysis. All A sites are divided into two categories: those within this window and those outside. The variance of  $m^6A$  sites between and within these regions was compared, followed by a one-way analysis of variance test. The  $F$ -statistic represents the ratio of the variance between groups to the variance within groups, with the  $P$  value indicating the significance of this comparison (\*\* $P < 10^{-10}$ , \* $P < 10^{-2}$ , NS, not significant). The dashed lines indicate an  $F$ -statistic of 1. Major  $m^6A$  motifs in the human transcriptome are highlighted in bold. **c**, Corresponding data for the variance between genes and within a single gene. Only A sites in the  $-20$  to

$+180$  hotspot region were analysed as other regions were minimally modified (\*\* $P < 10^{-10}$ , \* $P < 10^{-2}$ , NS, not significant). **d**, Relationship between  $\log_{10}$  gene expression and average  $m^6A$  level, shown as a contour plot with a dotted trend line. **e**, Distribution of the maximum  $m^6A$  modification level of all the genes. The fitting curve for low methylation and high methylation groups are marked by a blue and red dashed line, respectively. **f**, Correlation between the average and maximum  $m^6A$  modification levels within genes. **g, h**,  $\log_{10}$  gene expression level versus the average  $m^6A$  level for lowly modified genes only (**g**) and highly modified genes only (**h**). **i**, Overall relationship between  $m^6A$  level and gene expression for all genes (black), genes without  $m^6A$  modification (blue), low- $m^6A$ -modified genes (yellow) and high- $m^6A$ -modified genes (red).

We selected kethoxal as the reactive and reversible protecting group for guanosine, avoiding harsh deprotection and reverse transcription stalling due to incomplete deprotection. Combined with optimal RT conditions, we have developed CAM-seq, a transcriptome-wide  $m^6A$  sequencing method that has lower input RNA requirements, a shorter reaction time, streamlined procedure and reduced costs compared with other sequencing methods. CAM-seq can identify approximately 200,000  $m^6A$  sites using as little as 10 ng of input mRNA with background noise as low as 0.5%.

Saturated  $m^6A$  profiling throughout the transcriptome using CAM-seq advances analyses from semi-quantitative to nearly

quantitative, moving from limited profiling to comprehensive profiling. While previous methods such as meRIP-seq,  $m^6A$ -REF-seq,  $m^6A$ -SAC-seq, eTAM-seq and GLORI-seq provided valuable insights, they were limited to semi-quantitative analysis or limited profiling, although  $m^6A$ -SAC-seq<sup>11,24</sup> and related chemical approaches<sup>14</sup> could directly read out  $m^6A$  and retain sequence complexity. Our CAM-seq method, with its remarkably low background noise, facilitates systematic quantification of  $m^6A$  sites in most regions within most transcripts, enabling unbiased comparisons at the gene level. Using this sequencing method, we systematically explored  $m^6A$  variance at different layers: motif dependency, region variance and gene level. Consistent with

previous studies, we found that m<sup>6</sup>A modifications are enriched at RAC motifs and near stop codons. Beyond this, we quantitatively compared the m<sup>6</sup>A distribution among different motifs in human and plant samples and found that YAC (CAC/UAC) and RAU (GAU/AAU) motifs, which differ by only one nucleotide from the RAC motif, also show enriched m<sup>6</sup>A modification, especially in plant samples. More importantly, our results suggest that the transcription rate is a potential determinant of m<sup>6</sup>A levels. Specifically, genes with high transcription rates tend to have lower m<sup>6</sup>A levels, possibly due to a limited availability of the m<sup>6</sup>A deposition complex or dynamic association of the methyltransferase complex with transcription machineries within the cell.

Overall, CAM-seq represents an important advance in RNA modification mapping, providing a robust, efficient and truly quantitative method for comprehensive m<sup>6</sup>A sequencing. Furthermore, it will enhance our understanding of m<sup>6</sup>A dynamics in various biological contexts with quantitative and saturated data.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41557-025-01801-3>.

## References

- Chen, X., Xu, H., Shu, X. & Song, C.-X. Mapping epigenetic modifications by sequencing technologies. *Cell Death Differ.* <https://doi.org/10.1038/s41418-023-01213-1> (2023).
- Kim, S. C. et al. Modifications of mRNA vaccine structural elements for improving mRNA stability and translation efficiency. *Mol. Cell. Toxicol.* **18**, 1–8 (2022).
- Liu, A. & Wang, X. The pivotal role of chemical modifications in mRNA therapeutics. *Front. Cell Dev. Biol.* **10**, 901510 (2022).
- Harcourt, E. M., Kietrys, A. M. & Kool, E. T. Chemical and structural effects of base modifications in messenger RNA. *Nature* **541**, 339–346 (2017).
- Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **8**, 24 (2015).
- Delaunay, S., Helm, M. & Frye, M. RNA modifications in physiology and disease: towards clinical applications. *Nat. Rev. Genet.* **25**, 104–122 (2024).
- Khoddami, V. et al. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl Acad. Sci. USA* **116**, 6784–6789 (2019).
- Dai, Q. et al. Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-022-01505-w> (2022).
- Frommer, M. et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89**, 1827–1831 (1992).
- Dai, Q. et al. Ultrafast bisulfite sequencing detection of 5-methylcytosine in DNA and RNA. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-02034-w> (2024).
- Hu, L. et al. m<sup>6</sup>A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-022-01243-z> (2022).
- Xiao, Y.-L. et al. Transcriptome-wide profiling and quantification of N<sup>6</sup>-methyladenosine by enzyme-assisted adenosine deamination. *Nat. Biotechnol.* **41**, 993–1003 (2023).
- Liu, C. et al. Absolute quantification of single-base m<sup>6</sup>A methylation in the mammalian transcriptome using GLORI. *Nat. Biotechnol.* **41**, 355–366 (2023).
- Nappi, M., Hofer, A., Balasubramanian, S. & Gaunt, M. J. Selective chemical functionalization at N<sup>6</sup>-methyladenosine residues in DNA enabled by visible-light-mediated photoredox catalysis. *J. Am. Chem. Soc.* **142**, 21484–21492 (2020).
- Werner, S. et al. NOseq: amplicon sequencing evaluation method for RNA m<sup>6</sup>A sites after chemical deamination. *Nucleic Acids Res.* **49**, e23 (2021).
- Mahdavi-Amiri, Y., Chung, K. C. K. & Hili, R. Single-nucleotide resolution of N<sup>6</sup>-adenine methylation sites in DNA and RNA by nitrite sequencing. *Chem. Sci.* **12**, 606–612 (2021).
- Li, X. et al. NT-seq: a chemical-based sequencing method for genomic methylome profiling. *Genome Biol.* **23**, 122 (2022).
- Shen, W. et al. GLORI for absolute quantification of transcriptome-wide m<sup>6</sup>A at single-base resolution. *Nat. Protoc.* **19**, 1252–1287 (2024).
- Keeper, L. K. & Roller, P. P. N-Nitrosation by nitrite ion in neutral and basic medium. *Science* **181**, 1245–1247 (1973).
- Lai, C., Lin, G., Wang, W. & Luo, H. Absolute configurations and stability of cyclic guanosine mono-adducts with glyoxal and methylglyoxal. *Chirality* **23**, 487–494 (2011).
- Marek, R. & Lycka, A. <sup>15</sup>N NMR spectroscopy in structural analysis. *Curr. Org. Chem.* **6**, 35–66 (2002).
- Marek, R., Lycka, A., Kolehmainen, E., Sievanen, E. & Tousek, J. <sup>15</sup>N NMR spectroscopy in structural analysis: an update (2001–2005). *Curr. Org. Chem.* **11**, 1154–1205 (2007).
- Arkhipenko, S. et al. Mechanistic insights into boron-catalysed direct amidation reactions. *Chem. Sci.* **9**, 1058–1072 (2018).
- Ge, R. et al. m<sup>6</sup>A-SAC-seq for quantitative whole transcriptome m<sup>6</sup>A profiling. *Nat. Protoc.* **18**, 626–657 (2023).
- Weng, X. et al. Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.* **16**, 489–492 (2020).
- Wu, T., Lyu, R., You, Q. & He, C. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. *Nat. Methods* **17**, 515–523 (2020).
- Martin, F. H., Castro, M. M., Aboul-ela, F. & Tinoco, I. Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Res.* **13**, 8927–8938 (1985).
- Zheng, Y. Y., Reddy, K., Vangaveti, S. & Sheng, J. Inosine-induced base pairing diversity during reverse transcription. *ACS Chem. Biol.* **19**, 348–356 (2024).
- Meyer, K. D. et al. Comprehensive Analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
- Linder, B. et al. Single-nucleotide-resolution mapping of m<sup>6</sup>A and m<sup>6</sup>Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
- Garcia-Campos, M. A. et al. Deciphering the “m<sup>6</sup>A code” via antibody-independent quantitative profiling. *Cell* **178**, 731–747.e16 (2019).
- Wei, C.-M. & Moss, B. Nucleotide sequences at the N<sup>6</sup>-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16**, 1672–1676 (1977).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025



## Methods

### General procedure for deamination

A 4-ml screw-cap vial was charged with the substrate nucleobase (0.1 mmol, 1.0 equiv.) and solvent mixture. The organocatalyst (0.03 mmol, 0.3 equiv.) and Lewis acid catalyst (0.03 mmol, 0.3 equiv.) were then added in one portion at room temperature, followed by sodium nitrite (3.0 equiv.). The reaction vial was sealed and incubated at 37 °C for 12 h. The crude reaction mixture was then concentrated under reduced pressure (in vacuo) and purified by flash column chromatography on silica gel to obtain the product.

### General procedure for MALDI-TOF MS analysis

A 1.0 µl aliquot of a 100 µM synthetic DNA oligonucleotide (5'-CTCAGC-3') was mixed with 6.0 µl nuclease-free water. Next, 2.0 µl of 5× reaction buffer (0.5 M sodium cacodylate, 50 mM MgCl<sub>2</sub>, pH 7.0) and 1.0 µl of 1.0 M N<sub>3</sub>-kethoxal (dissolved in DMSO) were added, and the mixture was incubated at 37 °C for 10 min. A solution containing 20 µl of 8.8 M glyoxal and 5 µl of 10× PBS buffer was then added, followed by 15 µl H<sub>3</sub>BO<sub>3</sub> and 10 µl nuclease-free water at ambient temperature. This reaction mixture was then incubated at 50 °C for 30 min. Subsequently, 40 µl of a deamination buffer (comprising 10 µl saturated sodium nitrite solution, 10 µl of 5× HEPES buffer (0.5 M, pH 6.0) and 20 µl nuclease-free water) was added and mixed thoroughly cycles, after which the reaction was held at 4 °C until the next step. The DNA was precipitated with ethanol, reconstituted in 5.0 µl nuclease-free water and used directly for MALDI-TOF MS analysis. For the MALDI-TOF MS analysis, a matrix solution was prepared by mixing 2',4',6'-trihydroxyacetophenone (10 mg ml<sup>-1</sup> in 50% CH<sub>3</sub>CN-H<sub>2</sub>O) and ammonium citrate (50 mg ml<sup>-1</sup> in H<sub>2</sub>O) in a ratio of 8:1 (v/v). A 1.0 µl aliquot of the purified reaction product was combined with 1.0 µl of the matrix, spotted onto the MALDI-TOF sample plate, allowed to dry and analysed using a Bruker Ultraflex extreme MALDI-TOF/TOF mass spectrometer.

### General procedure for triple quad analysis

First, 1.0 µl of a 100 µM solution of synthetic 60-base DNA oligonucleotide was mixed with 6.0 µl nuclease-free water, 2.0 µl 5× reaction buffer (0.5 M sodium cacodylate, 50 mM MgCl<sub>2</sub>, pH 7.0) and 1.0 µl of 1.0 M N<sub>3</sub>-kethoxal (DMSO solution). The mixture was then incubated at 37 °C for 10 min. Next, 20 µl of 8.8 M glyoxal solution and 5 µl of 10× PBS were added to the reaction mixture, followed by 15 µl H<sub>3</sub>BO<sub>3</sub> and 10 µl nuclease-free water under ambient conditions. The reaction mixture was then incubated at 50 °C for 30 min. Finally, 40 µl of a deamination buffer (comprising 10 µl saturated sodium nitrite solution, 10 µl of 5× HEPES buffer (0.5 M, pH 6.0) and 20 µl nuclease-free water) was added and the solution mixed well. The reaction mixture was heated at 37 °C for 5 min and then at 18 °C for 30 min for six cycles, and then held indefinitely at 4 °C until further purification. DNA was precipitated from the solution using ethanol and reconstituted in 50 µl nuclease-free water. Untreated deaminated oligonucleotides (50–150 ng) were diluted in 17 µl nuclease-free H<sub>2</sub>O and denatured at 95 °C for 5 min, then immediately chilled on ice for 2 min. Next, 1 µl nuclease P1 (1 U µl<sup>-1</sup>, Wako, 145-08221) and 2 µl of 100 mM NH<sub>4</sub>OAc solution were added to the reaction mixture and incubated at 42 °C for 2–4 h (or overnight). Then, 1 U of FastAP thermosensitive alkaline phosphatase and 2.7 µl of 10× FastAP buffer (Thermo Scientific, EF0651) were added, followed by incubation at 37 °C for 2–4 h (or overnight). Digested oligonucleotides were diluted to 60 µl with nuclease-free H<sub>2</sub>O and filtered through a 0.22 µm filter (Millipore, SLGVR04NL). Samples were analysed directly in a SCIEX Triple Quad 6500+ LC-MS/MS system. For each analysis, 10 µl of sample was injected and the nucleosides were separated by reversed-phase ultra-high pressure liquid chromatography on a C18 column (Agilent, 927,700-092), followed by detection by MS. Nucleosides were quantified using nucleoside precursor ion to base ion mass transitions of *m/z* 253.1 to 135.0 for dI and *m/z* 252.1 to 136.0 for dA. The

nucleoside concentration was quantified using the calibration curves obtained from nucleoside standards measured under the same conditions. The final ratios of dI/dA and I/A were calculated by subtracting the background (mock control) generated by digestion enzymes.

### Cell culture

HEK293T cells were obtained from the American Type Culture Collection. The cells were maintained at 37 °C under 5% CO<sub>2</sub> in a Heracell VIOS 160i incubator (Thermo Scientific). All cell lines were cultured in DMEM (GIBCO, 11995), supplemented with 10% (v/v) fetal bovine serum and 1% penicillin–streptomycin (GIBCO).

### Plant material collection

The Col-0 accession of *Arabidopsis thaliana* and maize inbred line B73 (*Zea mays*) were used in this Article. *Arabidopsis* plants were grown at 22 °C with 16 h of light per 24 h. The seedlings were collected after growing on 1/2 Murashige and Skoog medium plates for 7 days. Maize inbred line B73 seeds were germinated at 28 °C with 14 h of light per 24 h. The shoots of 7-day-old maize plants were collected. The collected tissues were flash-frozen in liquid nitrogen, ground using a mortar and pestle, and stored at –80 °C until further use.

### PolyA-tailed RNA isolation

Total RNA was extracted from cells or plant tissue using TRIzol reagent (Ambion by Life Technologies) and Direct-zol RNA miniprep kit (Zymo Research) following the manufacturer's protocol. After RNA extraction, two rounds of poly(A) enrichment were conducted using DynaBeads mRNA direct purification kit (Thermo Fisher Scientific) following the manufacturer's protocol with some modifications. Accordingly, 200 µl Dynabeads were washed with 200 µl lysis/binding buffer and then mixed with 100 µg total RNA in 300 µl lysis/binding buffer. The samples were then incubated on the roller mixer at room temperature for 20 min. The beads were washed twice with washing buffer A and once with washing buffer B in the kit. The beads were resuspended in 50 µl of 10 mM Tris-HCl (pH 7), incubated at 70 °C for 3 min for washing and then eluted with 10 µl buffer. The eluate was subjected to a second round of poly(A) enrichment following the same procedure.

### General procedure for library preparation

Wild-type HEK293T cells and plant tissues were collected in three biological replicates. PolyA+ RNA was extracted from total RNA and 10–20 ng of the RNA was fragmented using RNA fragmentation reagents (NEB, E6150S) at 94 °C for 2 min, followed by purification with the RNA Clean and Concentrator kit (Zymo Research). After chemical treatment, 3'-end repair was performed on both the 'input' and 'treated' samples using T4 polynucleotide kinase (PNK, NEB, M0201S). RNA was combined with 2 µl of 10× T4 PNK reaction buffer (NEB, B0201S) and 3 µl T4 PNK, diluted to a final volume of 20 µl and incubated at 37 °C for 60 min. The mixture was then purified using the RNA Clean and Concentrator kit (Zymo Research) and eluted with 10 µl RNase-free water. For 3'-adapter ligation, 10 µl of 3'-repaired RNA fragments were incubated with 1 µl of 20 µM RNA 3'-adapter (/5rApp/AGATCGGAAGAGCGTCGTG/3Bio/) at 70 °C for 2 min and immediately placed on ice. The reaction was supplemented with 2.5 µl of 10× T4 RNA ligase reaction buffer (NEB, M0373L), 7.5 µl poly(ethylene glycol) 8000 (PEG8000; 50%), 1 µl SUPERase-In RNase Inhibitor and 1 µl T4 RNA ligase 2 truncated KQ (NEB, M0373L). The mixture was incubated at 25 °C for 2 h, followed by incubation overnight at 16 °C for 12 h. The reaction mixture was diluted to a final volume of 47 µl with nuclease-free water and excess adapters were removed by adding 2 µl 5'-deadenylase (NEB, M0331S) and incubating at 30 °C for 30 min, followed by the addition of 1 µl RecJf (NEB, M0264L) for single-stranded DNA digestion at 37 °C for 30 min. The 3'-end-ligated RNA was purified using the RNA Clean and Concentrator kit (Zymo Research) and eluted with 10 µl RNase-free water. The 'input' and 'treated' samples were both

mixed with 1  $\mu$ l of 2.0  $\mu$ M RT primer (5'-ACACGACGCTCTCCGATCT-3'), heated at 65 °C for 2 min and immediately placed on ice. To this mixture, 1  $\mu$ l of 5 $\times$  RT buffer, 3.05  $\mu$ l dNTP solution mix, 1  $\mu$ l RNaseOUT recombinant ribonuclease inhibitor (Thermo Scientific, 10777019) and 1  $\mu$ l RevertAid H Minus reverse transcriptase (Thermo Scientific, EP0452) were added. The mixture was thoroughly mixed and incubated at 42 °C for 1 h, followed by the addition of 1  $\mu$ l RNase H (NEB, M0297L) and incubation at 37 °C for 20 min. The reaction mixture was then heated at 70 °C for 5 min and the resulting complementary DNA was purified using the DNA Clean and Concentrator kit (Zymo Research). The purified cDNA was mixed with 2.0  $\mu$ l of 10  $\mu$ M cDNA adapter (5'-/5Phos/NNNNNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/-3'), heated at 70 °C for 2 min and immediately placed on ice. To this mixture, 5  $\mu$ l of 10 $\times$  T4 RNA ligase reaction buffer, 0.5  $\mu$ l of 25 mM ATP, 25  $\mu$ l PEG8000 (50%), 1.25  $\mu$ l of 40 mM Co(NH<sub>3</sub>)<sub>6</sub>Cl<sub>3</sub> solution, 3.75  $\mu$ l DMSO and 1  $\mu$ l T4 RNA ligase 1 (high concentration, NEB, M0437M) were added. The reaction mixture was then thoroughly mixed and incubated at 25 °C for 12 h, followed by purification using the DNA Clean and Concentrator kit (Zymo Research), eluting with 20  $\mu$ l RNase-free water. The eluted cDNA (20  $\mu$ l) was stored at -80 °C, and 8  $\mu$ l was used for each 15-cycle PCR amplification. PCR was performed using NEBNext Multiplex Oligos for Illumina (96 unique dual index primer pairs, New England BioLabs, E6440S). The resulting libraries were purified on a 3.5% low-melting-point agarose gel and sequenced on an Illumina NovaSeq 6000 device with a single-end 100-base pair read length. For a more detailed description of the process, see Supplementary Note II.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Sequencing data have been deposited in the Gene Expression Omnibus (GEO) with the following accession numbers: [GSE268871](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268871) for the human samples, [GSE268872](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268872) for the maize samples and [GSE268873](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268873) for the *Arabidopsis* samples. All other data are available in the paper or the Supplementary Information. Source data are provided with this paper.

### Code availability

Reads mapping and m<sup>6</sup>A sites detection scripts are available at GitHub via <https://github.com/y9c/m6A-CAMseq> (ref. 33).

## References

33. Chang, Y. m6A-CAM-seq. <https://github.com/y9c/m6A-CAMseq> (2024).

## Acknowledgements

We thank Y. Ge, Y. Gao and Y. Xiao for discussions. We thank the Genomics Facility of the University of Chicago and the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility for assistance with sequencing. This study was supported by the National Institutes of Health (grant HG008935, C.H.). C.H. is an investigator of the Howard Hughes Medical Institute.

## Author contributions

C.H., P.W. and C.Y. conceived the original idea and designed the original studies; P.W. performed most of the experiments with help from M.Z. and B.J.; C.Y. performed most of the bioinformatics analyses with input from P.W.; B.J. provided the plant materials. All authors approved the final paper.

## Competing interests

C.H. is a scientific founder, a member of the scientific advisory board and equity holder of Aferna Bio and Ellis Bio, a scientific cofounder and equity holder of Accent Therapeutics, and a member of the scientific advisory board of Rona Therapeutics and Element Biosciences. The other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41557-025-01801-3>.

**Correspondence and requests for materials** should be addressed to Chuan He.

**Peer review information** *Nature Chemistry* thanks Thomas Carell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection MALDI-TOF data was collected by analysis software from Bruker Ultraflexextreme; QQQ data was collected from SCIEX Triple Quad 6500+ platform using MultiQuant 3.0; High throughput sequencing data were generated by Illumina sequencing platform.

Data analysis All the code for data analysis is available via <https://github.com/y9c/m6A-CAM-seq>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data have been deposited in NCBI's Gene Expression Omnibus (GEO) under the accession number GSE226161.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Data exclusions

Replication

Randomization

Blinding

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a  Involved in the study

Antibodies

Eukaryotic cell lines

Palaeontology and archaeology

Animals and other organisms

Clinical data

Dual use research of concern

### Methods

n/a  Involved in the study

ChIP-seq

Flow cytometry

MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Authentication

Cell lines were not authenticated after purchase from ATCC.

Mycoplasma contamination

All cell lines used in this study were confirmed to be negative of mycoplasma contamination based on sequencing data.

Commonly misidentified lines  
(See [ICLAC](#) register)

No commonly misidentified line was used.