


# m<sup>6</sup>Am-seq reveals the dynamic m<sup>6</sup>Am methylation in the human transcriptome

Hanxiao Sun<sup>1,5</sup>, Kai Li<sup>1,2,3,5</sup>, Xiaoting Zhang<sup>1</sup>, Jun'e Liu<sup>1</sup>, Meiling Zhang<sup>1</sup>, Haowei Meng<sup>1</sup>  & Chengqi Yi<sup>1,3,4</sup>  

N<sup>6</sup>,2'-O-dimethyladenosine (m<sup>6</sup>Am), a terminal modification adjacent to the mRNA cap, is a newly discovered reversible RNA modification. Yet, a specific and sensitive tool to directly map transcriptome-wide m<sup>6</sup>Am is lacking. Here, we report m<sup>6</sup>Am-seq, based on selective in vitro demethylation and RNA immunoprecipitation. m<sup>6</sup>Am-seq directly distinguishes m<sup>6</sup>Am and 5'-UTR N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) and enables the identification of m<sup>6</sup>Am at single-base resolution and 5'-UTR m<sup>6</sup>A in the human transcriptome. Using m<sup>6</sup>Am-seq, we also find that m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A respond dynamically to stimuli, and identify key functional methylation sites that may facilitate cellular stress response. Collectively, m<sup>6</sup>Am-seq reveals the high-confidence m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A methylome and provides a robust tool for functional studies of the two epitranscriptomic marks.

<sup>1</sup>State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. <sup>2</sup>Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. <sup>3</sup>Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. <sup>4</sup>Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing, China. <sup>5</sup>These authors contributed equally: Hanxiao Sun, Kai Li. ✉email: [chengqi.yi@pku.edu.cn](mailto:chengqi.yi@pku.edu.cn)

More than 160 different chemical modifications have been found so far<sup>1</sup>. N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) is the most abundant internal modification on mRNA and lncRNA in eukaryotes<sup>2–4</sup>. In addition to m<sup>6</sup>A, there exists another reversible modification in higher eukaryotes, called N<sup>6</sup>, 2'-O-dimethyladenosine (m<sup>6</sup>Am), which is precisely located at the first transcribed nucleotide and hence adjacent to the cap structure of mRNA<sup>5,6</sup>. A total of 50–80% of adenosine-starting mammalian mRNAs are believed to be m<sup>6</sup>Am modified<sup>6</sup>. It is catalyzed by PCIF1<sup>7–10</sup>, a protein that interacts with the phosphorylated CTD of RNAPII<sup>11</sup>, and could be removed by the m<sup>6</sup>A demethylase FTO<sup>12,13</sup>. Thus, m<sup>6</sup>Am is dynamically regulated by FTO and PCIF1, leading to the direction of the cap epitranscriptomics.

The development of transcriptome-wide sequencing technologies for various RNA modifications has greatly facilitated the field of epitranscriptomics<sup>14</sup>. The most widely used method for m<sup>6</sup>Am/m<sup>6</sup>A mapping relies on m<sup>6</sup>A antibodies, which do not discriminate between the two functionally distinct modifications<sup>15,16</sup>. While m<sup>6</sup>A is enriched around the stop codon, it is challenging to distinguish m<sup>6</sup>Am from 5'-UTR m<sup>6</sup>A, which constitutes a significant part of m<sup>6</sup>A methylome and has been shown to be functionally important in various biological processes<sup>17–19</sup>. There have been efforts aiming to distinguish the two modifications<sup>8–10,20,21</sup>; yet they suffer from the low efficiency of UV crosslinking, the inaccuracy of TSS annotation, and the limited activity of 5' exonuclease, thus compromising the sensitivity and precision of m<sup>6</sup>Am detection. While the use of PCIF1 knockout (KO) cell lines has improved the confidence of m<sup>6</sup>Am detection<sup>9,10</sup>, it is an indirect approach and is incompatible for epitranscriptome analysis in human tissues and biological samples where genetic manipulation is challenging. Thus, a sensitive and direct method for transcriptome-wide m<sup>6</sup>Am identification is highly desired.

Here, we develop m<sup>6</sup>Am-seq to investigate the prevalence, topology, and dynamics of m<sup>6</sup>Am in the human transcriptome. m<sup>6</sup>Am-seq relies on an *in vitro* demethylation reaction that selectively removes m<sup>6</sup>Am while keeps m<sup>6</sup>A intact, thereby discriminating genuine m<sup>6</sup>Am at the mRNA cap and 5'-UTR m<sup>6</sup>A. Using m<sup>6</sup>Am-seq, we identify 2166 m<sup>6</sup>Am sites at base resolution from 1652 peaks and 1307 5'-UTR m<sup>6</sup>A peaks throughout the human transcriptome. Moreover, we show that m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A respond dynamically to heat shock and hypoxia conditions, and identify hundreds of stress-induced m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A peaks that could mediate adaptive response to the stressors. Altogether, we provide a tool to directly and selectively distinguish and profile m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A; we anticipate that m<sup>6</sup>Am-seq will pave the way for future functional studies of m<sup>6</sup>Am in various biological systems.

## Results

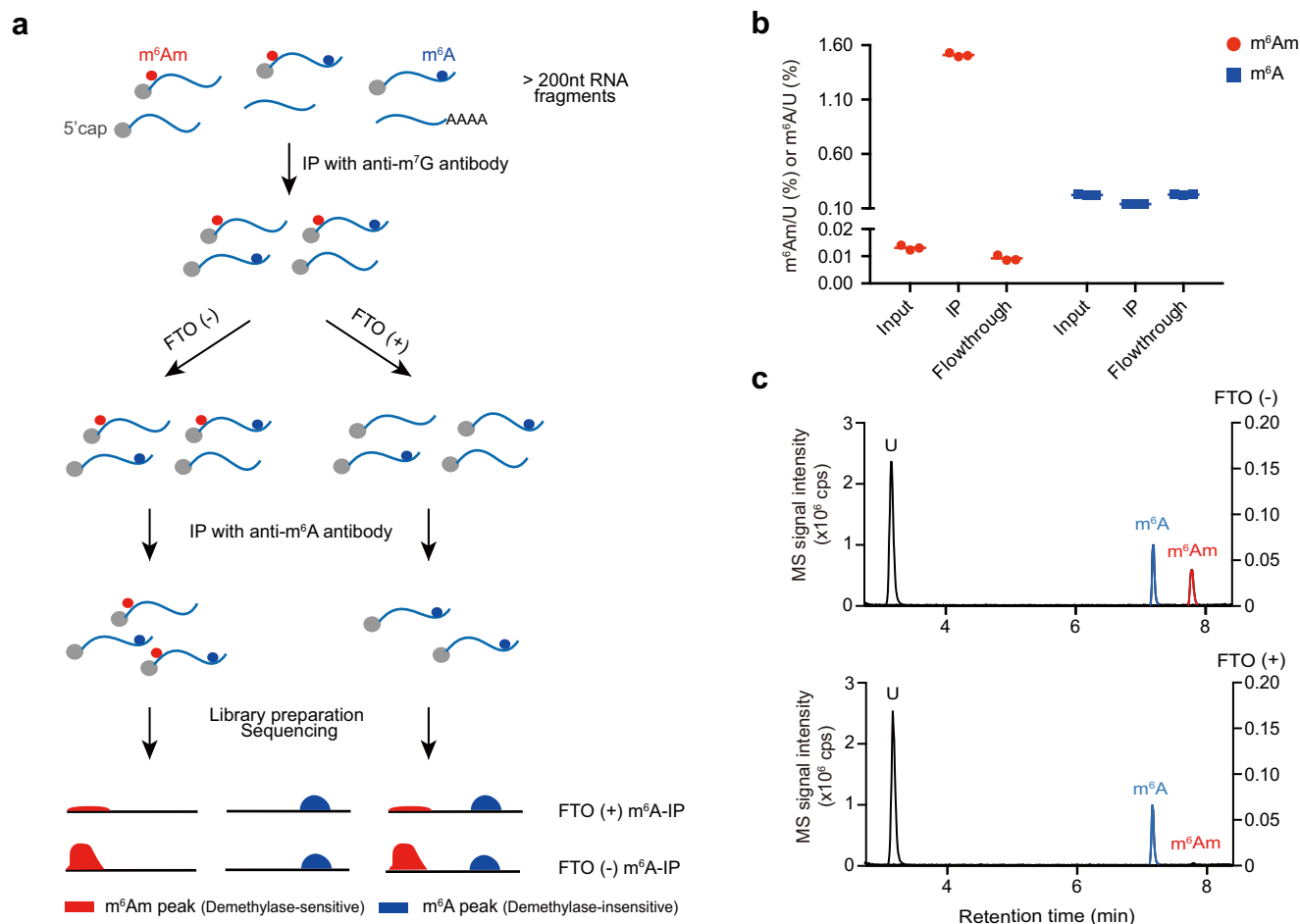
**A selective demethylation reaction to distinguish m<sup>6</sup>Am and m<sup>6</sup>A.** m<sup>6</sup>Am-seq utilizes a selective demethylation reaction to achieve specific and sensitive detection of m<sup>6</sup>Am. While both m<sup>6</sup>Am and m<sup>6</sup>A are recognized by antibodies, we envisioned that the unique cap-adjacent structural context of m<sup>6</sup>Am could be selectively read out *in vitro* by demethylase (Supplementary Fig. 1a). To this end, we tested ALKBH5 and FTO, two known RNA demethylases<sup>22,23</sup>. We found that ALKBH5 does not act on m<sup>6</sup>Am, consistent with previous biochemical data<sup>13</sup>. However, it does not demethylate m<sup>6</sup>A to completion (~80%, Supplementary Fig. 1b), complicating subsequent analysis. In contrast, FTO prefers m<sup>6</sup>Am but demethylates both modifications with high efficiency (>98% and ~60% for m<sup>6</sup>Am and m<sup>6</sup>A, respectively), under recommended conditions from literature<sup>13</sup> (Supplementary Fig. 1c).

Neither of the demethylation reactions would be good enough to distinguish m<sup>6</sup>Am and m<sup>6</sup>A in their present forms. The ideal condition would remove one modification to completion while leaving another intact. FTO and ALKBH5 are Fe(II)/2-KG-dependent dioxygenases, whose *in vitro* demethylation activity is promoted by L-ascorbic acid<sup>22,23</sup>. As a reducing agent, it is capable of reducing redox-active metals such as iron, thereby increasing the pro-oxidant chemistry of these metals<sup>24</sup>. We thus hypothesized that by omitting L-ascorbic acid from our conditions, we might be able to modulate the *in vitro* activity of FTO so that it only acts on the preferred, cap-adjacent m<sup>6</sup>Am but not the internal m<sup>6</sup>A. To our delight, after removing L-ascorbic acid and further optimizing the demethylation conditions, >95% of m<sup>6</sup>Am was removed without significantly affecting the m<sup>6</sup>A level (Fig. 1c). We further verified the selectivity of our demethylation condition using a dually modified probe containing both cap m<sup>6</sup>Am and internal m<sup>6</sup>A, which were 76 nt apart from each other (Supplementary Fig. 1d). Thus, we identified a demethylation reaction that is specific for m<sup>6</sup>Am.

**m<sup>6</sup>Am-seq directly detects m<sup>6</sup>Am.** To map the m<sup>6</sup>Am in the human transcriptome, we developed m<sup>6</sup>Am-seq, which combines the selective demethylation reaction with RNA immunoprecipitation (Fig. 1). Two key RNA immunoprecipitation procedures are utilized, which are a cap-m<sup>7</sup>G IP and an m<sup>6</sup>A IP step. Because the global m<sup>6</sup>A level is around 20-fold higher than that of m<sup>6</sup>Am in total RNA (Supplementary Fig. 1e), and the vast majority of m<sup>6</sup>Am is localized in mRNA cap (Supplementary Fig. 1f), we used the cap-m<sup>7</sup>G IP step to enrich for m<sup>6</sup>Am and deplete m<sup>6</sup>A that is mostly found in the 3'-UTR. As expected, immunoprecipitation with a cap-m<sup>7</sup>G antibody for fragmented RNA resulted in more than 100-fold enrichment of m<sup>6</sup>Am and a simultaneous 2-fold reduction of m<sup>6</sup>A (Fig. 1b). Moreover, ~90% of peaks are located at and near 5'-UTR, suggesting 5'-end of mRNAs were successfully and specifically enriched using size selected total RNA (>200 nt) as input, without the prerequisite to purify mRNA (Supplementary Fig. 2a).

The immunoprecipitated RNA was then subjected to the selective demethylation reaction or mock treatment, which was followed by the second IP step utilizing an m<sup>6</sup>A antibody (Fig. 1a). The FTO (–) sample would contain both m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A signals, while the FTO (+) sample would contain only m<sup>6</sup>A. By comparing the methylome profiles of the two samples, we were able to detect m<sup>6</sup>Am, which is sensitive to demethylase treatment, from 5'-UTR m<sup>6</sup>A, which is insensitive and conserved in the two samples. Indeed, we observed a bimodal-like distribution of fold change of peak intensity: peaks on the right side experienced reduced intensity upon demethylase treatment, suggesting they are demethylase-sensitive m<sup>6</sup>Am modification; while peaks on the left were not significantly altered or even slightly increased, suggesting they are retained 5'-UTR m<sup>6</sup>A peaks (Supplementary Fig. 2b, c). Thus, m<sup>6</sup>Am-seq selectively distinguishes the two modifications.

**m<sup>6</sup>Am-seq detects transcriptome-wide m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A.** We next applied m<sup>6</sup>Am-seq to characterize m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A methylome in HEK293T cells. We found that m<sup>6</sup>Am-seq results are highly correlated between biological replicates (correlation coefficient = 0.9821, Supplementary Fig. 2d), demonstrating its robustness. We identified 1652 and 1307 high-confidence m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A peaks from 1635 and 1297 genes, exhibiting the characteristic BCA (B = C/U/G) and GGACH (H = A/C/U) motif, respectively (Fig. 2a, b and Supplementary Data 1, 2). m<sup>6</sup>Am shows the highest enrichment at annotated TSSs, while the summit of m<sup>6</sup>A is clearly within 5'-UTR; the peak density of m<sup>6</sup>Am is also 3-fold higher than that of



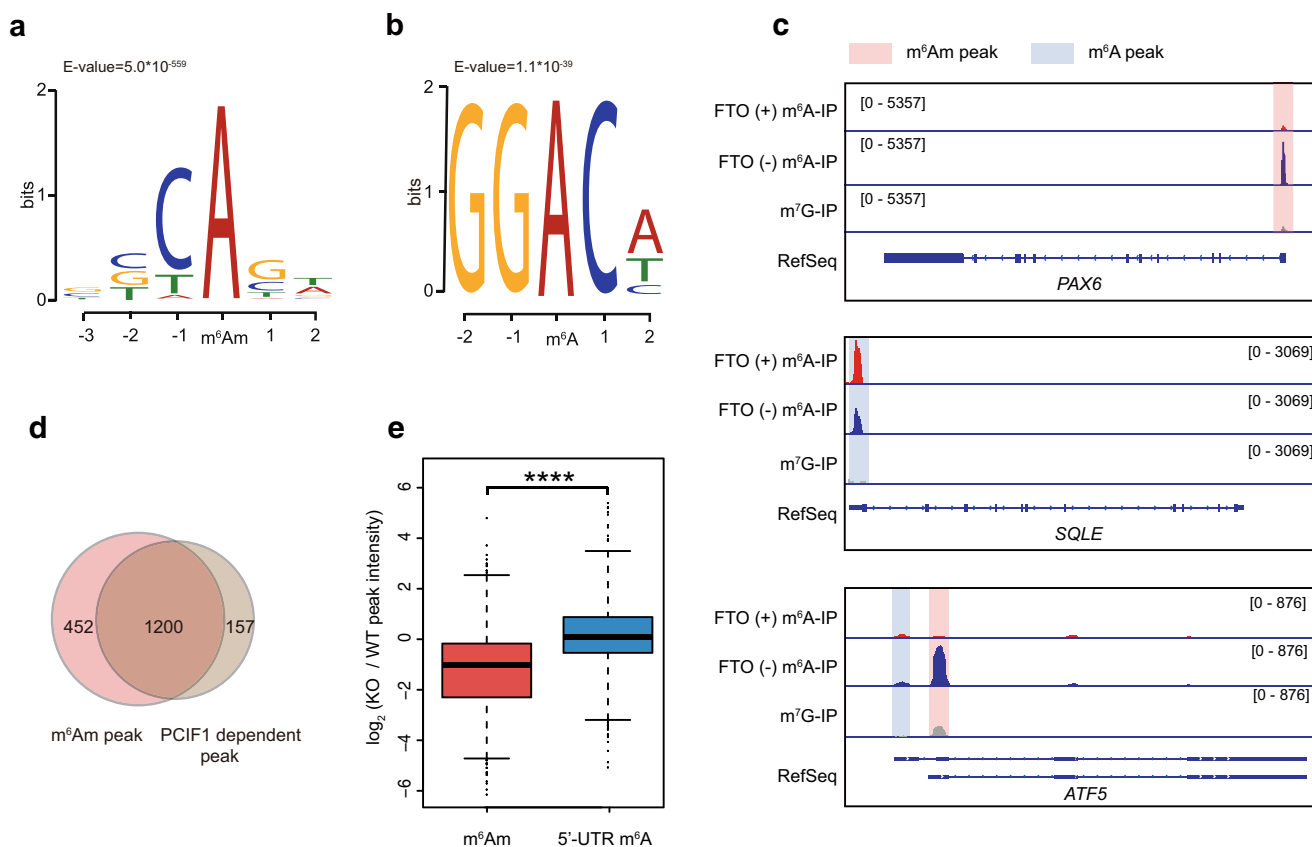
**Fig. 1**  **$m^6\text{Am}$ -seq combines RNA immunoprecipitation with an optimized demethylation reaction to enrich and distinguish  $m^6\text{Am}$  and 5'-UTR  $m^6\text{A}$ .** **a** Scheme of  $m^6\text{Am}$ -seq. Total RNA (>200 nt) was fragmented (input) and immunoprecipitated with cap- $m^7\text{G}$  antibody ( $m^7\text{G}$ -IP). The cap structure-containing RNA was subjected to [FTO (+)] or [FTO (-)] treatment. The [FTO (+)] or [FTO (-)] samples were immunoprecipitated with an  $m^6\text{A}$  antibody and sequenced for  $m^6\text{Am}$  detection. The sequencing profiles of the [FTO (+)  $m^6\text{A}$ -IP] and [FTO (-)  $m^6\text{A}$ -IP] samples were normalized and compared. The “Demethylase-sensitive peaks” and “Demethylase-insensitive peaks” were identified. The demethylase-sensitive peaks bear the  $m^6\text{Am}$ , while the demethylase-insensitive peaks bear the  $m^6\text{A}$  (Online Methods). **b** The  $m^6\text{Am}$  and  $m^6\text{A}$  levels of input,  $m^7\text{G}$ -IP, and flowthrough samples were quantified by LC-MS/MS, respectively.  $m^6\text{Am}$  was significantly enriched (>100 fold) by cap- $m^7\text{G}$  antibody while  $m^6\text{A}$  level decreased mildly. Values represent mean  $\pm$  SD ( $n = 3$  independent samples in “Input”, “IP” and “Flowthrough” respectively). **c** The  $m^6\text{Am}$  and  $m^6\text{A}$  levels of [FTO (+)] and [FTO (-)] samples were quantified by LC-MS/MS, respectively. >95%  $m^6\text{Am}$  was demethylated and  $m^6\text{A}$  was negligibly altered under optimized demethylation conditions. The chromatograms of U (black) and  $m^6\text{Am}$  (red) are scaled to the left Y axis, and the chromatograms of  $m^6\text{A}$  (blue) are scaled to the right Y axis. Source data are provided as a Source Data file.

5'-UTR  $m^6\text{A}$ , demonstrating that  $m^6\text{Am}$  is more concentrated than  $m^6\text{A}$  around the TSSs (Supplementary Fig. 3a, b). Representative methylation profiles are shown in Fig. 2c. In one example, the *ATF5* gene has multiple isoforms with distinct TSS and alternative first exons; we observed  $m^6\text{Am}$  modification for one transcript and 5'-UTR  $m^6\text{A}$  for another transcript (Fig. 2c), suggesting differential cis-regulatory elements in regulating the epitranscriptomic status of different transcripts from the same gene. Gene ontology (GO) terms analysis showed that transcripts containing  $m^6\text{Am}$  and 5'-UTR  $m^6\text{A}$  are most strongly enriched in cell cycle and RNA metabolism, respectively, suggesting their differential functions (Supplementary Fig. 3c, d).

To further demonstrate the accuracy of  $m^6\text{Am}$ -seq, we applied standard  $m^6\text{A}$  IP experiments to PCIF1-KO HEK293T cells. Based on the reduction of signals compared to WT cells, we detected 1357 PCIF1-dependent peaks; notably, 1200 out of the 1357 peaks were identified by  $m^6\text{Am}$ -seq (Fig. 2d and Supplementary Fig. 3e), suggesting the high confidence of  $m^6\text{Am}$ -seq. The remaining peaks all exhibited  $m^6\text{Am}$ -like

property (i.e., sensitive to demethylation) in  $m^6\text{Am}$ -seq but were not called due to the stringent cutoff we used. On the other hand, 5'-UTR  $m^6\text{A}$  peaks identified by  $m^6\text{Am}$ -seq were not affected upon PCIF1 depletion (Fig. 2e and Supplementary Fig. 3f), further demonstrating the specificity of  $m^6\text{Am}$ -seq. Taken together, these results prove that  $m^6\text{Am}$ -seq identifies  $m^6\text{Am}$  with high confidence. It is also worth mentioning that  $m^6\text{Am}$ -seq is applicable for  $m^6\text{Am}$  detection in tissues or cell populations where PCIF1 KD or KO is not feasible.

**Comparisons of existing  $m^6\text{Am}$  detection methods.** We then compared  $m^6\text{Am}$ -seq results with  $m^6\text{Am}$ -Exo-seq and miCLIP<sup>9,10</sup>. Under normalized sequencing depth, the averaged peak coverage of  $m^6\text{Am}$ -seq (~387) was ~4 $\times$  and 21 $\times$  higher than that of  $m^6\text{Am}$ -Exo-seq (~95) and miCLIP (~18) (Fig. 3a). In addition,  $m^6\text{Am}$  peaks by  $m^6\text{Am}$ -seq exhibit a strong tendency towards the annotated TSS, in comparison to a modest enrichment within 5'-UTR by miCLIP and  $m^6\text{Am}$ -Exo-Seq (Fig. 3b). Thus,  $m^6\text{Am}$ -seq outperforms existing methods in  $m^6\text{Am}$  detection.



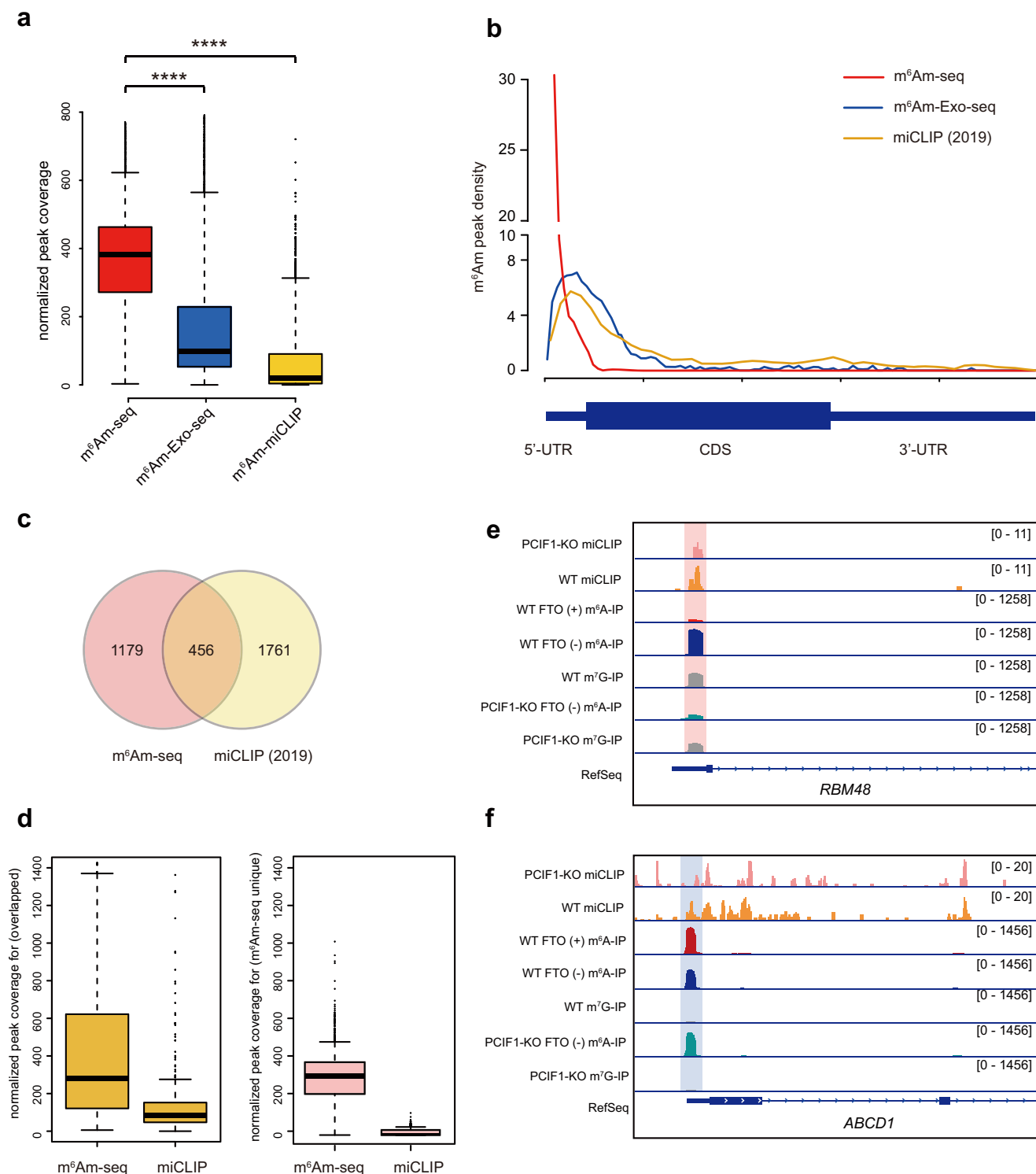
**Fig. 2**  $m^6Am$ -seq uncovers the high-confidence  $m^6Am$  and 5'-UTR  $m^6A$  methylation in the human transcriptome. Motif analysis of  $m^6Am$  peaks (a) and 5'-UTR  $m^6A$  peaks (b) identified by  $m^6Am$ -seq. c Representative views of typical  $m^6Am$  peaks and 5'-UTR  $m^6A$  peaks on mRNA. The methylation peak in the 5'-UTR of *PAX6* was lost in “FTO (+)  $m^6A$ -IP” samples, suggesting a demethylase-sensitive  $m^6Am$  modification (top); the methylation peak in the 5'-UTR of *SQLE* retained well in “FTO (+)  $m^6A$ -IP” samples, suggesting a demethylase-insensitive  $m^6A$  modification (middle); while multiple isoforms of *ATF5* were differentially marked by  $m^6Am$  and 5'-UTR  $m^6A$  (bottom). Pink and blue background colors denote  $m^6Am$  and  $m^6A$  signals, respectively. d Venn diagram showing the overlap between  $m^6Am$  peaks identified directly by  $m^6Am$ -seq and  $m^6Am$  peaks identified indirectly by  $m^6A$  IP using PCIF1-KO samples. e Boxplot showing that the peak intensity of  $m^6Am$  (red,  $n = 1652$ ), but not that of  $m^6A$  (blue,  $n = 1307$ ), was significantly reduced upon PCIF1 depletion. Statistical significance of the difference was determined by unpaired two-sided Mann-Whitney  $U$ -test. \*\*\*\* $P < 2.2e^{-16}$ . Boxes represent 25th–75th percentile (line at median) with whiskers at  $1.5 \times IQR$ . Source data are provided as a Source Data file.

Because  $m^6Am$ -seq and miCLIP can both detect  $m^6Am$  and  $m^6A$ , we performed comparisons of the two methods<sup>10,20</sup> (Fig. 3c and Supplementary Fig. 4a, b). We used the recent miCLIP data relying on PCIF1-KO for detailed comparison<sup>10</sup>. Approximately 21% (456/2217) of  $m^6Am$  marked genes by miCLIP overlap with that of  $m^6Am$ -seq (Fig. 3c). We found that miCLIP data sets showed very low coverage, providing an explanation for the limited sensitivity of miCLIP (Fig. 3d, e). While it is possible that the stringent cutoff of  $m^6Am$ -seq may favor specificity over sensitivity, a significant portion of miCLIP-identified  $m^6Am$  peaks could be mis-annotated and are instead 5'-UTR  $m^6A$  peaks (Fig. 3f).

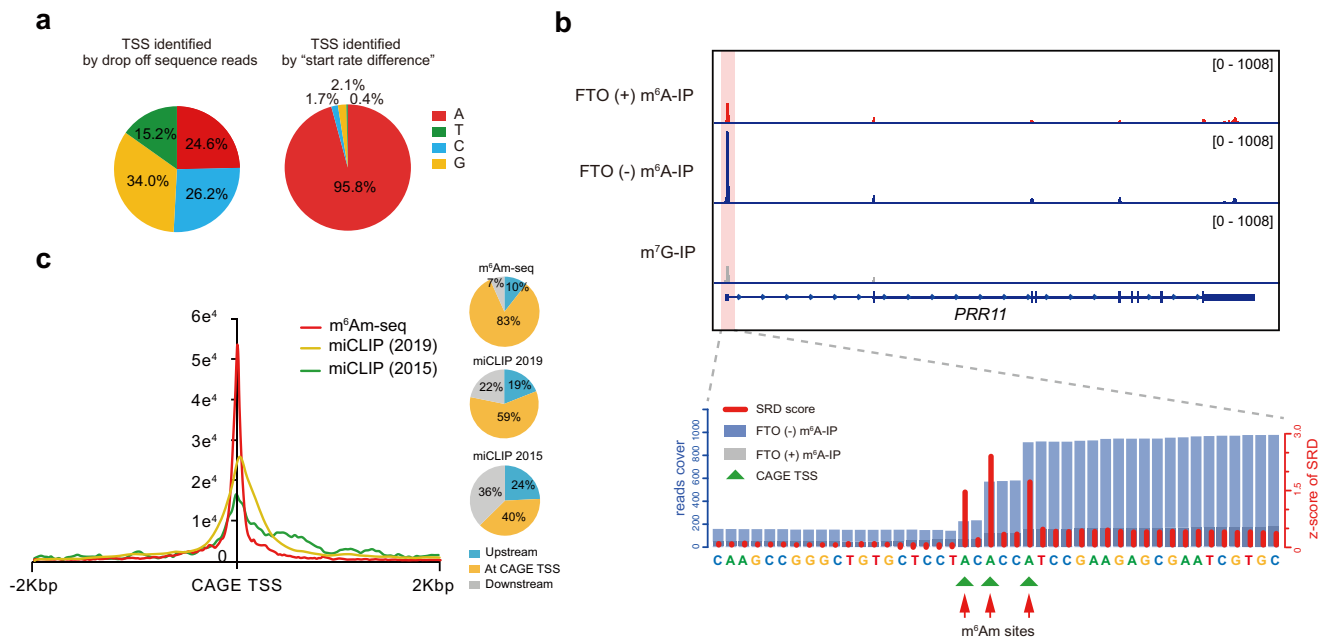
It was proposed that  $m^6Am$  controls mRNA stability<sup>12</sup>, based on the  $m^6Am$  methylome identified by miCLIP. Given the observation that miCLIP may mis-annotate the two modifications, we compared the expression of  $m^6Am$ -containing genes, which are identified by  $m^6Am$ -seq, in wild-type and PCIF1-KO cells. Different from the previous conclusion<sup>12</sup>, we found that the loss of  $m^6Am$  did not alter the expression of their host mRNA (Supplementary Fig. 4c).

**$m^6Am$ -seq detects  $m^6Am$  at single-base resolution.** Because  $m^6Am$  is the first transcribed nucleotide, one would expect to identify  $m^6Am$  at base resolution by searching for A-starting

transcripts (i.e., those showing sudden drop-off of sequencing reads at the 5' end). However, even when we used a library preparation procedure that preserves the first transcribed nucleotide, we did not observe an enrichment of A nucleotide in the  $m^6Am/m^6A$  enriched FTO (–) data set (Fig. 4a), suggesting that its TSS location alone is not sufficient to detect  $m^6Am$  sites. Thus, we defined a “start rate difference” score (SRD score) for each nucleotide to quantify sequencing reads beginning from it (see “Methods” section). Application of these metrics immediately resulted in accurate identification of the  $m^6Am$  site in synthetic spike-in RNAs (Supplementary Fig. 5a) and ~95.8% of A nucleotides in the  $m^6Am/m^6A$  enriched FTO (–) data set (Fig. 4a). In total, we identified 2166  $m^6Am$  sites from 1459 genes (Supplementary Data 3). For instance, mRNA-coding gene *PRR11* and a lncRNA *LOC101929147* both have alternative TSSs that start with an A nucleotide, and we showed that all these transcripts are marked by  $m^6Am$  (Fig. 4b and Supplementary Fig. 5b). In addition, while  $m^6Am$  is associated with a BCA motif, our single-base  $m^6Am$  methylome revealed ~34% non-BCA  $m^6Am$  sites (Supplementary Fig. 6a). We found also a weaker signal of GRO-seq and H3K27ac ChIP-seq for the non-BCA  $m^6Am$  sites, but no difference in H3K4me3, H3K4me1, and H3K9me3 signals (Supplementary Fig. 6b–f). Altogether, these data indicate that  $m^6Am$ -seq enables  $m^6Am$  detection at single-base resolution.



**Fig. 3 Comparison of m<sup>6</sup>Am methylation detected by m<sup>6</sup>Am-seq and existing m<sup>6</sup>Am detection methods.** **a** Boxplot showing that the normalized peak coverage by m<sup>6</sup>Am-seq (red,  $n = 1652$ ), m<sup>6</sup>Am-Exo-seq (blue,  $n = 3169$ ), and miCLIP (yellow,  $n = 2217$ ). Statistical significance of the difference was determined by unpaired two-sided Mann-Whitney  $U$ -test. \*\*\*\* $P < 2.2e^{-16}$  for m<sup>6</sup>Am-seq vs m<sup>6</sup>Am-Exo-seq, m<sup>6</sup>Am-seq vs miCLIP. Boxes represent 25th–75th percentile (line at the median) with whiskers at 1.5 $\times$ IQR. **b** Metagene profiles of m<sup>6</sup>Am distribution identified by the three sequencing methods. Each segment was normalized according to its average length in RefSeq annotation. **c** Venn diagram showing the overlap of m<sup>6</sup>Am-marked genes by m<sup>6</sup>Am-seq and by miCLIP. **d** Boxplot showing the normalized peak coverage of shared m<sup>6</sup>Am peaks (yellow,  $n = 456$ , identified in both m<sup>6</sup>Am-seq and miCLIP) and m<sup>6</sup>Am-seq-unique m<sup>6</sup>Am peaks (pink,  $n = 1179$ ) in the m<sup>6</sup>Am-seq and miCLIP data sets, respectively. Boxes represent 25th–75th percentile (line at the median) with whiskers at 1.5 $\times$ IQR. **e** An example of m<sup>6</sup>Am (in the *RBM48* gene), which was missed by miCLIP due to its limited sensitivity but is clearly identified by m<sup>6</sup>Am-seq. The m<sup>6</sup>Am peak is lost in PCIF1-KO data sets, further proving the detection confidence of m<sup>6</sup>Am-seq. Pink background color denotes m<sup>6</sup>Am signal. **f** An example of mis-annotated m<sup>6</sup>Am (in the *ABCD1* gene) by miCLIP. This peak is not affected by FTO treatment nor PCIF1-KO, showing that it is a 5'-UTR m<sup>6</sup>A modification. Blue background color denotes m<sup>6</sup>A signal. Source data are provided as a Source Data file.



**Fig. 4** m<sup>6</sup>Am-seq detects m<sup>6</sup>Am methylome at single-base resolution. **a** Nucleotide proportion of TSS identified by the sudden drop-off of sequencing reads (left) and by "start rate difference" (SRD) metrics. **b** A representative view of three single m<sup>6</sup>Am sites on the transcripts of *PRR11*, which were all supported by CAGE data. Three adenosine residues with a high SRD score (red bar) were defined as m<sup>6</sup>Am sites, which overlapped exactly with CAGE sites (green triangles). Each segment was normalized according to its average length in RefSeq annotation. Pink background color denotes m<sup>6</sup>Am signal. **c** A metaplot showing that the vast majority of m<sup>6</sup>Am sites by m<sup>6</sup>Am-seq overlapped with CAGE sites, suggesting that these m<sup>6</sup>Am sites are indeed TSSs. Compared with miCLIP, m<sup>6</sup>Am sites identified by m<sup>6</sup>Am-seq were much closer to the CAGE TSSs. Two miCLIP data sets were used for comparison: an m<sup>6</sup>Am list reported in 2015<sup>20</sup> and an updated m<sup>6</sup>Am list in 2019<sup>10</sup>, based on PCIF1-KO. Source data are provided as a Source Data file.

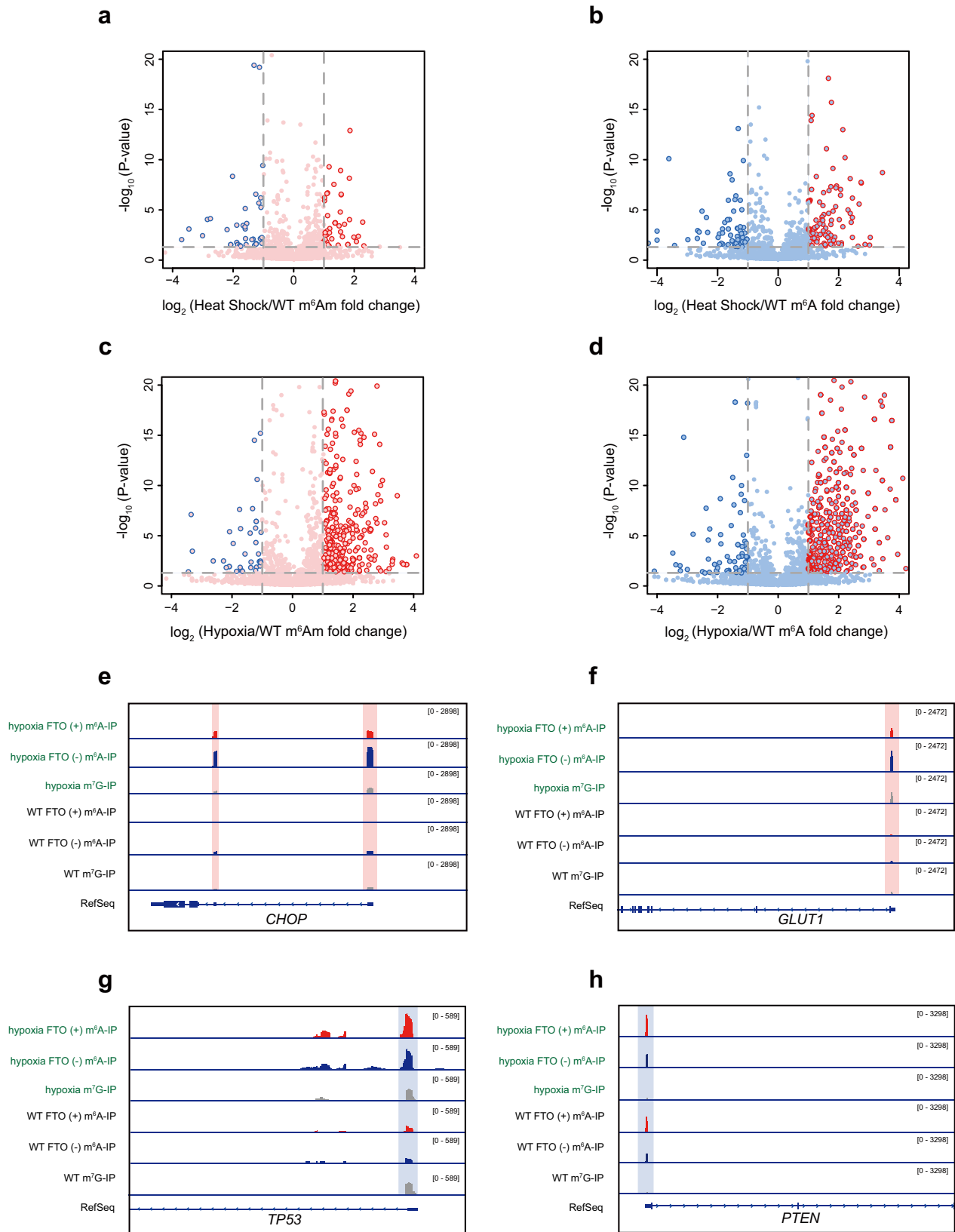
To further characterize the accuracy of the identified m<sup>6</sup>Am sites, we performed the cap-analysis of gene expression (CAGE) experiments and compared the m<sup>6</sup>Am sites with annotated and CAGE-identified TSSs. We found that most of the m<sup>6</sup>Am sites are present downstream of the RefSeq-annotated TSSs, with only 14% of m<sup>6</sup>Am sites mapped at the annotated TSSs (Supplementary Fig. 7a). This observation suggested that existing methods based on annotated TSSs information can be inaccurate for m<sup>6</sup>Am mapping. For instance, the m<sup>6</sup>Am site of *TMET241* would have been located internally within 5'-UTR if the annotated TSSs were used (Supplementary Fig. 7b); and two m<sup>6</sup>Am peaks are clearly identified for *RAB5A*, which has only one annotated TSS (Supplementary Fig. 7c). We next plotted the distance from m<sup>6</sup>Am sites to their nearest CAGE TSS sites, and observed that the vast majority of m<sup>6</sup>Am sites overlapped with CAGE TSSs (Fig. 4c), suggesting that these m<sup>6</sup>Am sites are indeed TSSs. Compared with miCLIP<sup>10,20</sup>, m<sup>6</sup>Am sites by m<sup>6</sup>Am-seq were much closer to the CAGE TSSs (Fig. 4c). These data highlight the reliability of m<sup>6</sup>Am-seq in obtaining high-confidence m<sup>6</sup>Am sites.

**Heat shock induces dynamic m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A methylations.** Stress-induced RNA modification is an important feature of dynamic epitranscriptomic regulation under physiological conditions<sup>15,25–27</sup>. For instance, it has been shown that in response to heat shock stress, m<sup>6</sup>A deposited to the 5'-UTR of newly transcribed mRNAs promotes cap-independent translation initiation, providing a mechanism of selective mRNA translation during general translation inhibition<sup>17,19</sup>. In light of this, we applied m<sup>6</sup>Am-seq to characterize the dynamic m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A methylome under heat shock conditions. We identified 45 upregulated m<sup>6</sup>Am peaks and 90 5'-UTR m<sup>6</sup>A peaks, as well as 36 and 67 downregulated m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A peaks (Fig. 5a, b and Supplementary Fig. 8a, b and Supplementary Fig. 9a–d and Supplementary Data 4). Among them, we found key heat shock-

associated transcripts with elevated 5'-UTR m<sup>6</sup>A: notable examples include *HSPA1A*, *HSPA1B*, and *DNAJB1* (Supplementary Fig. 10a, b), whose transcripts undergo m<sup>6</sup>A-dependent but cap-independent selective translation<sup>17</sup>. These results by m<sup>6</sup>Am-seq accurately recapitulated heat shock-induced, functional 5'-UTR m<sup>6</sup>A methylation.

Unexpectedly, a subset of transcripts previously proposed to rely on m<sup>6</sup>A-dependent translation under heat shock<sup>17</sup> did not contain m<sup>6</sup>A modification in the 5'-UTR; instead, m<sup>6</sup>Am-seq reveals that they are dynamically modified by m<sup>6</sup>Am (Supplementary Fig. 10c). For instance, using a method that does not distinguish between m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A, upregulation of 5'-UTR m<sup>6</sup>A level in *EGR1* was believed to mediate the stress-induced selective translation. However, using m<sup>6</sup>Am-seq, we found that *EGR1* contained m<sup>6</sup>Am rather than 5'-UTR m<sup>6</sup>A (Supplementary Fig. 10d). Therefore, cap-independent translation of such transcripts including *EGR1* is unlikely to be mediated by m<sup>6</sup>A. Whether or not m<sup>6</sup>Am, which is located at TSS and immediately after the m<sup>7</sup>G cap, can also promote stress-induced selective translation would be of interest for future studies. Overall, m<sup>6</sup>Am-seq provides accurate methylome maps to enable the study of the physiological roles of the two RNA modifications.

**Dynamic m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A methylome under hypoxia conditions.** We next determined whether m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A methylation can be dynamically regulated by hypoxia. We were able to identify 362 and 638 elevated m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A peaks respectively, with 31 and 63 decreased m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A peaks (Fig. 5c, d and Supplementary Fig. 8c, d and Supplementary Data 5). While some of the hypoxia-specific genes show inducible m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A modifications, many transcripts with unaltered expression levels are also dynamically modified (Supplementary Fig. 9e–h). Overall, we observed a more



**Fig. 5 m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A are dynamically regulated by different stress conditions.** **a-d** Volcano plot showing stress-induced m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A: heat shock-inducible m<sup>6</sup>Am peaks (**a**) and 5'-UTR m<sup>6</sup>A peaks (**b**), and hypoxia-inducible m<sup>6</sup>Am peaks (**c**), and 5'-UTR m<sup>6</sup>A peaks (**d**). Red circles denote stress-upregulated peaks; blue circles denote stress-downregulated peaks. Representative views of elevated m<sup>6</sup>Am peaks in *CHOP* (**e**) and *GLUT1* (**f**) under hypoxia stress. Representative views of elevated 5'-UTR m<sup>6</sup>A peaks in *TP53* (**g**) and *PTEN* (**h**) under hypoxia stress. Pink and blue background colors denote m<sup>6</sup>Am and m<sup>6</sup>A signals, respectively.

evident increase of m<sup>6</sup>Am and m<sup>6</sup>A methylome under hypoxia conditions than heat shock conditions.

To cope with hypoxia, mammals have evolved key adaptive mechanisms including cellular adaptations of protein synthesis, energy metabolism, mitochondrial respiration, lipid, and carbon metabolism as well as nutrient acquisition<sup>28</sup>. Given the dramatic change of m<sup>6</sup>Am level, we examined whether dynamic m<sup>6</sup>Am may participate in the adaptive mechanisms during the hypoxia condition. Intriguingly, our GO analysis showed that genes with elevated m<sup>6</sup>Am levels are linked to the endoplasmic reticulum (ER) stress-response regulation (Supplementary Fig. 11a). For instance, it has been documented that transcriptional and translational upregulation of both *CHOP* and *ATF3* are important to activate the expression of downstream genes in response to stress<sup>29,30</sup>; in line with this, we observed significantly elevated m<sup>6</sup>Am levels in their mRNAs (Fig. 5e and Supplementary Data 5). In addition to ER stress, oxygen deprivation also reprograms intracellular metabolism<sup>31</sup>. For instance, hypoxia can enhance the mRNA and protein levels of HIF1 target gene *GLUT1* to achieve glucose transport<sup>31</sup>; in agreement with this, we also observed increased m<sup>6</sup>Am level in the *GLUT1* mRNA (Fig. 5f). Thus, we identified a dynamic m<sup>6</sup>Am program in key transcripts that are utilized by cells to adapt to hypoxia.

Distinct from the m<sup>6</sup>Am-marked genes, GO analysis of genes with elevated 5'-UTR m<sup>6</sup>A indicated an enrichment of the p53 signaling pathway (Supplementary Fig. 11b). In fact, hypoxia has frequently been described to be a p53 inducer<sup>32</sup>. Along this line, p53 and many of its downstream genes, such as *MDM2*, *PTEN*, *PIDD1*, etc., were found to contain elevated 5'-UTR m<sup>6</sup>A under hypoxia stress (Fig. 5g, h and Supplementary Data 5). Since induced 5'-UTR m<sup>6</sup>A by heat shock has been shown to enable cap-independent translation, it is possible that protein synthesis of the p53 signaling pathway could be modulated by the stress-induced m<sup>6</sup>A. Altogether, our results suggest that dynamic m<sup>6</sup>Am and m<sup>6</sup>A as key components of the adaptive response to hypoxia.

## Discussion

More than 40 years have passed since the initial documentation of m<sup>6</sup>Am was made<sup>5,6</sup>. The recent discovery of its “eraser” and “writer” has led to a renewed interest in this reversible modification<sup>33</sup>. Yet its biological functions remain enigmatic, due to a lack of sensitive methods that detect the modification at the transcriptome-wide level. Here, we report m<sup>6</sup>Am-seq, which enables specific and robust profiling of m<sup>6</sup>Am in the human transcriptome.

The development of m<sup>6</sup>Am-seq is inspired by existing epigenomic and epitranscriptomic methods utilizing biologically important demethylases. For instance, TAB-Seq<sup>34</sup> and TAPS<sup>35</sup> rely on the TET proteins, which oxidize 5-methylcytosine (5mC), to detect 5mC and its oxidative derivatives in the genome. In DM-tRNA-seq<sup>36</sup>, ARM-seq<sup>37</sup>, and m<sup>1</sup>A-MAP<sup>38</sup>, the RNA/DNA demethylase AlkB has been used to map N<sup>1</sup>-methyladenosine (m<sup>1</sup>A) in tRNA and mRNA. In this study, we expand the available toolbox by repurposing FTO to detect m<sup>6</sup>Am and m<sup>6</sup>A, two dynamic and reversible epitranscriptomic marks. Different from the existing methods that use demethylases to indiscriminately remove all substrates, we manipulated the *in vitro* activity of FTO to selectively demethylate just one of the two cognate substrates. This is achieved by the discovery of a controlled *in vitro* demethylation reaction, which omits the key cofactor L-ascorbic acid from the biochemical condition. Owing to its capacity to reduce Fe<sup>3+</sup> to the catalytically active Fe<sup>2+</sup>, L-ascorbic acid can enhance the *in vitro* demethylation activity<sup>24,39</sup>. By leaving out L-ascorbic acid from the *in vitro* reactions, FTO still possesses a high m<sup>6</sup>Am demethylation activity while minimally affects the level of m<sup>6</sup>A, hence allowing selective detection of m<sup>6</sup>Am.

Interestingly, the *in vivo* bioavailability of ascorbate, the dominant form of vitamin C under physiological pH conditions, has been demonstrated to impact the epigenome via TETs and JmjC domain-containing demethylases, ultimately leading to phenotypic changes in development, aging, cancer and other diseases<sup>24</sup>. It remains to be seen whether or not the deficiency of vitamin C might alter m<sup>6</sup>Am or m<sup>6</sup>A methylome and hence epitranscriptome reprogramming inside of cells.

Except for the selective demethylation condition omitting L-ascorbic acid, major steps of m<sup>6</sup>Am-seq are well-known meRIP protocols. Hence, m<sup>6</sup>Am-seq is straightforward and highly reproducible. Its high sensitivity and accuracy are also enabled by the combination of cap-m<sup>7</sup>G IP and m<sup>6</sup>A IP, which builds on the knowledge that m<sup>6</sup>Am is adjacent to the m<sup>7</sup>G cap. It should be noted that m<sup>6</sup>Am-seq does not detect internal m<sup>6</sup>Am, although the amount of the internal m<sup>6</sup>Am in mRNAs seems negligible. Only one internal m<sup>6</sup>Am site in U2 snRNA has been documented so far. In m<sup>6</sup>Am-seq, we used a demethylation reaction to distinguish m<sup>6</sup>Am from m<sup>6</sup>A, other selective chemistry could be exploited for the determination of additional modifications. For instance, dNTP-sensitivity during reverse transcription<sup>40</sup> and peroxidate oxidation<sup>41</sup> could be combined with cap-m<sup>7</sup>G IP to detect ribose methylation at the first and second transcribed nucleotides of mRNA. In addition, given the diversity of RNA cap structures<sup>42,43</sup>, including NAD<sup>+</sup>, NADH, desphospho-CoA, and 2,2,7-trimethylguanosine, tailored combination of IP enrichment and a discrimination procedure, as demonstrated in m<sup>6</sup>Am-seq, could have broad applications in analyzing the epitranscriptomic state of the transcriptome.

m<sup>6</sup>Am-seq distinguishes m<sup>6</sup>Am from m<sup>6</sup>A and provides a high-confidence m<sup>6</sup>Am methylome. While a previous study used miCLIP for m<sup>6</sup>Am identification and has indicated a role of m<sup>6</sup>Am in controlling mRNA stability<sup>12</sup>, this conclusion was disputed by more recent studies led by independent groups<sup>7,9,13</sup>. Using m<sup>6</sup>Am-seq, we corroborated the latter studies and found that m<sup>6</sup>Am does not affect mRNA stability. Because miCLIP does not separate well m<sup>6</sup>Am from m<sup>6</sup>A, it is likely that the observed effect in mRNA stability is caused by m<sup>6</sup>A<sup>13</sup>, which has been shown to increase mRNA stability when recognized by certain m<sup>6</sup>A readers<sup>44</sup>. Nevertheless, such disagreement underscores the importance of a specific method to accurately map m<sup>6</sup>Am for future functional studies.

We showed that m<sup>6</sup>Am is dynamically regulated by heat shock and hypoxia, hinting that m<sup>6</sup>Am is important for cellular stress response. While a specific role of m<sup>6</sup>Am remains to be elucidated, we found that many well-known stress-induced genes exhibiting elevated protein levels showed increased m<sup>6</sup>Am methylation levels upon stimuli. Such a positive correlation appears to be the opposite in normal human tissues<sup>21</sup>; nevertheless, this does not rule out the possibility of a unique role of m<sup>6</sup>Am under stress. Consistent with the positive association, ribosome profiling experiments also revealed a positive role of m<sup>6</sup>Am in translation<sup>7</sup>. Given that 5'-UTR m<sup>6</sup>A has been demonstrated to mediate cap-independent translation<sup>17,19</sup>, it is tempting to speculate that m<sup>6</sup>Am plays a role in stress-induced mRNA translation.

Whether or not m<sup>6</sup>Am may participate in cap-independent translation is an open question. It is known that the mRNA cap structure plays an important role in translation initiation. Aki-chika et al. showed that m<sup>6</sup>A does not modulate the binding affinity of eIF4E to the cap structure, suggesting that other cap-binding factors could be involved<sup>7</sup>. For instance, it has been shown that eIF3 and ABCF1 can selectively recognize 5'-UTR m<sup>6</sup>A and mediate the cap-independent translation<sup>18,19</sup>. Thus, future identification of potential m<sup>6</sup>Am readers, particularly those in the integrated stress response, could help elucidate the detailed function and mechanism of m<sup>6</sup>Am.



In summary, we report a tool for transcriptome-wide mapping of m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A, and reveal the dynamic features of the two epitranscriptomic marks under stress conditions. We anticipate that m<sup>6</sup>Am-seq will lead to new territories of RNA biology and open opportunities in the field of “cap epitranscriptomics”.

## Methods

**Antibodies and cell culture.** Rabbit anti-m<sup>6</sup>A antibody was purchased from EMD Millipore (ABE572). Anti-m<sup>7</sup>G-cap monoclonal antibody was obtained from MBL (RN016M). HEK293T cells were cultured at 37 °C in DMEM medium supplemented with 10% FBS, 1% GlutaMAX and 0.5% penicillin/streptomycin with 5% CO<sub>2</sub>. For heat shock, cells were grown to a confluence of 80% in 10-cm plastic dishes, placed in a 43 °C water bath for 2 h, and then recover for 6 h in normal growing conditions. For hypoxia treatment, cells were grown to 80%, carried out in Hypoxia Hood which was maintained at 37 °C, 5% CO<sub>2</sub>, 1% oxygen for 24 h.

**RNA isolation.** TRIzol was used for total RNA isolation according to the manufacturer's protocol. <200 nt RNA was deleted using the MEGAclear Transcription Clean-Up kit (Ambion, AM1908).

**Quantification of m<sup>6</sup>A and m<sup>6</sup>Am by LC-MS/MS.** For m<sup>6</sup>Am, 150 ng RNA was deacetylated by 10 units of RppH (NEB, M0356S) in Thermopol buffer for 5 h at 37 °C. RNA was purified by RNA clean & concentrator kit (Zymo Research, R1015). Purified RNA was digested by nuclease P1 (Sigma, N8630) with 10 mM ammonium acetate at 42 °C for 6 h, then mixed with 50 mM MES buffer (pH 6.5) and 0.5 U Shrimp Alkaline Phosphatase (NEB, M0371S), and incubated at 37 °C for another 6 h. For m<sup>6</sup>A, 150 ng RNA was digested into single nucleosides by nuclease P1 with 10 mM ammonium acetate at 42 °C for 6 h, then incubated with 50 mM MES buffer (pH 6.5) and 0.5 U Shrimp Alkaline Phosphatase for another 6 h at 37 °C.

The digested nucleosides were separated by the ultra-performance liquid chromatography on C18 column, and detected by the triple-quadrupole mass spectrometer (AB SCIEX QTRAP 6500). The multiple reaction-monitoring (MRM) mode was monitored: *m/z* 268.0–136.0 (A), *m/z* 245.0–113.1 (U), *m/z* 282.0–150.1 (m<sup>6</sup>A), *m/z* 296.0–150.0 (m<sup>6</sup>Am). The concentrations of m<sup>6</sup>A and m<sup>6</sup>Am levels in RNA samples were calculated from the standard curves.

**Synthesis of m<sup>6</sup>Am/Am and m<sup>6</sup>A probes.** cap-m<sup>6</sup>Am/Am RNA: template was obtained by PCR, which contains the T7 promoter sequence and ~150 bp DNA sequence. Primers used to amplify the spike-in template was provided (Supplementary Data 6). In vitro transcription reaction contained: reaction buffer (ThermoFisher Scientific, AM1354), 800 ng template DNA, T7 Enzyme Mix (ThermoFisher Scientific, AM1354), 15 mM ATP/UTP/CTP/GTP, and 15 mM cap analogs (Trilink, N-7113 or N-7102, N-7102 was ordered via email addressed to orders@trilinkbiotech.com). The reaction was incubated overnight at 37 °C. After incubation, 1 µl of TURBO DNase (ThermoFisher Scientific, AM1354) was added to the reaction, and the reaction was incubated at 37 °C for 1 h. The RNA was purified by phenol–chloroform isolation and ethanol precipitation. The precise RNA was purified from 7 M urea/15% acrylamide gels using elution buffer (0.3 M sodium acetate, 1 mM EDTA, 0.05% SDS) at 37 °C for 12 h. Next, the RNA was precipitated with ethanol. The probes were used as spike-ins and for probe demethylation assay. The sequence of cap-m<sup>6</sup>Am/Am RNA probes was list as follows:

m<sup>6</sup>Am(Am)GGAGAAAAUCACUCAGGGUCAUUGCCAGCGCUUCGUU  
AAUACAGAUGUAGGUUCCACAGGGUAGCCAGCAGCAUCCGCGCAU  
GAGAUCGGGAACAUAUAGGUGCAGGGCGCUGACUUCGGCGUUUCA  
GACUUUACGAAACACGGA.

The sequence of m<sup>6</sup>A probe was list as follows:  
AUCUACCGUCCAGUAGCCUUCAGGAUCAUGCUGUCUGACUUGCU  
GGm<sup>6</sup>ACAUCAUUCUAGUGCCAUAACUUCAGC.

**Synthesize the dually modified m<sup>6</sup>Am&m<sup>6</sup>A probe.** The dually modified m<sup>6</sup>Am & m<sup>6</sup>A probe was obtained by utilizing T4 RNA ligase 2 to ligate the 3' end of RNA probe to the 5' end of another RNA probe. The 30nt for m<sup>6</sup>Am probe was obtained by in vitro transcription reaction. For phosphorylation of 5' end of m<sup>6</sup>A probe, RNA was treated by PNK (NEB, M0201S) with incubation for 1 h at 37 °C. 1 µl of 20 µM m<sup>6</sup>Am probe, 2 µl of 20 µM m<sup>6</sup>A probe, 1.5 µl of 20 µM splint DNA (CTACTGGACAGGTAGATAGTCTGAAGGATTC), 1 µl 10× T4 DNA ligase buffer and 3 µl nuclease-free water were heated at 65 °C for 3 min, followed by 5 min at 25 °C in a thermocycler. Add 1 µl of 10 units/µl T4 RNA ligase 2 (NEB, M0239S) and 0.5 µl RiboLock RNase inhibitor (ThermoFisher Scientific, E00381) into the reaction mixture and incubate at 37 °C for 2 h. Synthetic m<sup>6</sup>Am & m<sup>6</sup>A probe was gel purified and sliced in the desired size. The sequence of the m<sup>6</sup>Am & m<sup>6</sup>A probe was listed as follows:

Capm<sup>6</sup>AmGGAAUAGCAGGCAUGGAAUCCUUAAGACUAUCUACCCUGU  
CCAGUAGCCUUCAGGAUCAUGCUGUCUGACUUGCUGm<sup>6</sup>ACAUCAU  
UCUAGUGCCAUAACUUCAGC.

**In vitro probe assay.** To investigate the FTO and ALKBH5 demethylation activity for m<sup>6</sup>Am and m<sup>6</sup>A, we performed in vitro demethylation assay using purified demethylases and synthesized probes contained m<sup>6</sup>A and m<sup>6</sup>Am. The demethylation assay was performed in 20 µl of mixture containing 200 ng probe, 50 mM MES (pH 6.5), 100 mM KCl, 2 mM MgCl<sub>2</sub>, 100 µM (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>·6H<sub>2</sub>O, 300 µM 2-ketoglutarate, 0 or 2 mM L-ascorbic acid, 0.4 U/ml SUPERase-In RNase inhibitor (Invitrogen, AM2694) and 2 µM purified FTO. The reaction mixture was incubated at 37 °C for 20 min. As the control, inactive demethylases were using for incubation with RNA.

**CAGE-seq library preparation.** CAGE allows mapping of the transcriptional start sites of capped RNAs. 5 µg of total RNA purified from HEK293T cells was used for the CAGE-seq<sup>45</sup>. Reverse transcription was carried out using PrimeScript reverse transcriptase (Takara, 2680A) with the RT-N15-EcoP primer. The RNA-cDNA hybrid was diol oxidized with NaIO<sub>4</sub> and then the 5' Cap and 3' end of RNA were biotinylated. The noncapped biotinylated RNA-cDNA hybrid was eliminated by RNase digestion and the biotin-trapping step. The complete cDNA was released from RNA and then the Bar-coded 5' linker was ligated to the single-stranded cDNA. The second-strand cDNA was synthesized using the biotin-modified primer. The cDNA was digested with EcoP15I (NEB, R0646S) and the 3' linker was ligated to the 3' end of cDNA. The cDNA was performed PCR amplification. The CAGE library was sequenced on Illumina HiSeq 2000. The sequences of primers and linkers were provided (Supplementary Data 6).

**CAGE data analysis.** CAGE tags were mapped to the human genome (hg19, UCSC Genome Browser) using TopHat2 (version 2.0.13) with default parameters that allow 0 mismatches per seed (22 nt), and CTSS pipeline was used to identify CAGE TSS<sup>45</sup>. Only uniquely mapped reads were used in downstream analysis within R and custom scripts. The mapped reads were sorted and the TSS tag was calculated. All unique 5' ends of CAGE tag-supported TSS and the number of tags in each CTSS represents expression levels. Tag counts in each CTSS were normalized using tag pre million reads (TPMs). We also used Deeptools (version 3.5.0) to analyze the enrichment of m<sup>6</sup>Am in the –2 Kb to +2 Kb region around the CAGE TSS. The miCLIP-2015 m<sup>6</sup>Am list and miCLIP-2019 m<sup>6</sup>Am list were obtained from previous studies<sup>10,13</sup>.

**Cloning, expression, and purification of FTO and ALKBH5.** Wild-type human FTO gene and ALKBH5 gene with deletion of the 66 amino acids were cloned into pET-28a. The recombinant plasmids were then transduced into *E. coli* BL21 (DE3). FTO and ALKBH5 proteins were purified following the procedures as previously described<sup>46</sup>. Bacteria were grown at 37 °C until the OD<sub>600</sub> reached 0.8–1, and then induced with 0.5 mM IPTG overnight at 15 °C. Cells were harvested by centrifugation and then resuspended in ice-cold lysis buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM imidazole), disrupted on ice, and centrifuged at 11,000×g for 1 h at 4 °C. The supernatant was purified on HiTrap His column (GE Healthcare) and eluted with buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 500 mM imidazole. Next, the fractions were purified by HiTrap Q column (GE Healthcare). Finally, the fractions were purified by Gel-filtration chromatography (HiLoad 16/600 Superdex 200 pg column, GE Healthcare) equilibrated in storage buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 3 mM dithiothreitol (DTT). Purified protein was stored at –80 °C with the addition of 20% glycerol.

## m<sup>6</sup>Am-seq

**cap-m<sup>7</sup>G RNA immunoprecipitation.** The total RNAs were extracted from control, heat shock-treated, and hypoxia-treated cells. 100 µg >200 nt RNA was fragmented into ~150 nt-long fragments using RNA fragmentation buffer (NEB, E6150S). Adding fragment stop solution buffer stopped the reaction, and fragments were purified and concentrated by ethanol precipitation. 10 ng of fragments were used as “input” and the remaining fragments were used for followed treatments. The fragmented RNA was denatured at 65 °C for 5 min, followed by chilling in ice. The denatured RNA was incubated overnight at 4 °C with 2 µg cap-m<sup>7</sup>G antibody (1:100 dilution) and 5 µl RiboLock RNase inhibitor in IPP buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% NP-40). 20 µl of protein A/G UltraLink Resin (ThermoFisher Scientific, 53132) were washed twice with IPP buffer, and then resuspended in a 200 µl IPP buffer contained 5 µl RiboLock RNase inhibitor. The antibody-RNA mixture was then incubated with protein A/G UltraLink Resin at 4 °C for 3 h. The Resin was washed twice with 1 ml of IPP buffer, once with 1 ml of low-salt IP buffer (10 mM Tris-HCl, pH 7.4, 75 mM NaCl, 0.05% NP-40), once with 1 ml of high-salt IP buffer (10 mM Tris-HCl, pH 7.4, 200 mM NaCl, 0.05% NP-40) and twice with 1 ml TET buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.05% NP-40) at 4 °C. The beads-antibody-RNA was resuspended in 1 ml Trizol and rotated for 10 min at room temperature. cap-m<sup>7</sup>G-containing RNA fragments

were extracted and precipitated. Around 150 ng cap-m<sup>7</sup>G-immunoprecipitated RNA fragments could be obtained.

**In vitro demethylation treatment.** 1 ng of the cap-m<sup>7</sup>G-immunoprecipitated RNA was used as “m<sup>7</sup>G-IP”. RNA was subjected to [FTO (+)] or [FTO (-)] treatment. About 100 ng of RNA was used for [FTO (+)] treatment. RNA was denatured at 65 °C for 5 min, followed by chilling in ice. And then, the demethylation assay was performed in a 20 µl mixture which contained 50 mM MES (pH 6.5), 100 mM KCl, 2 mM MgCl<sub>2</sub>, 100 µM (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>·6H<sub>2</sub>O, 300 µM 2-ketoglutarate, 0.4 U/ml SUPERase-In RNase inhibitor and 1 µM purified FTO. After incubation at 37 °C for 20 min, the demethylated RNA was purified by phenol–chloroform extraction. The remaining RNA (around 50 ng) directly performed m<sup>6</sup>A RNA immunoprecipitation.

**m<sup>6</sup>A RNA immunoprecipitation.** The [FTO (-)] and [FTO (+)] samples were immunoprecipitated with m<sup>6</sup>A antibody. The m<sup>6</sup>A RNA immunoprecipitation was performed following the procedures as previously described<sup>21</sup>. 10 µl of protein A magnetic beads (ThermoFisher Scientific, 10002D) and 10 µl of protein G magnetic beads (ThermoFisher Scientific, 10004D) were mixed and washed twice with 500 µl of IPP buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% NP-40). The beads were resuspended in 500 µl of IPP buffer containing 2 µg of anti-m<sup>6</sup>A antibody (Millipore, ABE572). The mixture was incubated at 4 °C for at least 6 h. The beads-antibody mixture was washed twice in 500 µl of IPP buffer and resuspended with 200 µl mixture containing 40 µl of 5× IPP buffer, denatured RNA (the [FTO (-)] control and [FTO (+)] treated samples) and 5 µl of RNasin Plus RNase Inhibitor (Promega, N2615) at 4 °C for 2 h (1:100 dilution of anti-m<sup>6</sup>A antibody). The beads-antibody-RNA mixture was washed twice with 500 µl of IPP buffer, twice with 500 µl of low-salt IP buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.1% NP-40), and twice with 500 µl of high-salt IP buffer (10 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.1% NP-40). And then, bound fragments were eluted from beads with 200 µl of IPP buffer containing 6.7 mM N<sup>6</sup>-methyladenosine (Sigma-Aldrich, M2780) and 5 µl of RiboLock RNase inhibitor. An additional phenol–chloroform isolation and ethanol precipitation were used to purify the RNA. The bound RNA was called [FTO (+) m<sup>6</sup>A-IP] and [FTO (-) m<sup>6</sup>A-IP].

**Library preparation and sequencing.** The “input”, “m<sup>7</sup>G-IP”, [FTO (+) m<sup>6</sup>A-IP] and [FTO (-) m<sup>6</sup>A-IP] were subjected to library construction using SMARTer<sup>®</sup> Stranded Total RNA-Seq Kit v2—Pico Input Mammalian (634413, Takara – Clontech, Japan) according to the manufacturer’s protocol. The libraries were sequenced on Illumina HiSeq X10 with PE150.

**Reads pre-processing and alignment.** Strand orientation of the original RNA was preserved on the process of library construction and reads R2 yields sequences sense to the original RNA. Thus, only reads R2 was used in our study. Raw sequencing reads were firstly subjected to Trim\_galore (version 0.6.6, [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) and cutadapt software (version 1.18) for quality control and trimming adaptor. The quality threshold was set to 30, and the minimum length required for reads after trimming was 20 nt. Processed reads were mapped to genome (hg19, UCSC Genome Browser) using HISAT2 (version 2.1.0)<sup>47</sup> with default parameters, separated by strand with in-house scripts, and reads coverage were showed by IGV. The mpileup files are generated from mapped BAM files using samtools mpileup command (version 1.9).

**Transcriptome-wide identification of 5'-UTR peaks in m<sup>6</sup>A-IP sample.** Mapped reads were subjected to the RNA methylation peak caller exomePeak<sup>48</sup>, the FOLD\_ENRICHMENT was set to 3, the GENE\_ANNOTATION were set RefSeq\_hg19 for peak calling. For genome-base peak caller MACS2 (version 2.1.1)<sup>49</sup>, the effective genome size was set to 2.7×10<sup>9</sup> for human, under the option of *-nomodel* and *-q-value* cutoff 0.01. In order to identify high-confidence peak in m<sup>6</sup>A-IP sample, peak must meet the following conditions: (1) peak was overlapped between m<sup>6</sup>A-IP and m<sup>7</sup>G-IP samples, and (2) fold enrichment in m<sup>6</sup>A-IP higher than m<sup>7</sup>G-IP samples. Only peaks that can meet both requirements were used for m<sup>6</sup>A/m<sup>6</sup>Am signal identification in our study.

**Definition and calculation of the demethylase-sensitivity peaks.** In order to more accurately extract the m<sup>6</sup>Am peak and 5'-UTR m<sup>6</sup>A peak, we identify the demethylase-sensitivity peak between the FTO (+) and FTO (-) sample. Firstly, 5'-UTR peak defined in the previous section of “Transcriptome-wide identification of 5'-UTR peaks in m<sup>6</sup>A-IP sample” were used for analysis. Secondly, differential methylation peak between FTO (+) and FTO (-) samples were identified using exomePeak (*P* < 0.01) and DESeq2 (version 3.11)<sup>50</sup> in the R environment (version 3.6). Thirdly, when the ratio of FTO (-) to FTO (+) is >2, is defined as demethylase-sensitivity peak.

**Identification of m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A peaks.** An m<sup>6</sup>Am peak was identified when: (1) peak was high-confidence in m<sup>6</sup>A-IP sample; (2) peak was demethylase sensitivity between the FTO (+) and FTO (-) sample. The 5'-UTR m<sup>6</sup>A peak was identified when: (1) peak was high-confidence in m<sup>6</sup>A-IP sample; (2) peak was

demethylase insensitivity between the FTO (+) and FTO (-) sample, and was not decrease after FTO treatment.

**Identification of m<sup>6</sup>Am site.** To identify the potential m<sup>6</sup>Am site, we defined a “start rate difference” score (SRD score) for each nucleotide within an m<sup>6</sup>Am peak. SRD score takes into account normalized sequencing start rate in untreated samples  $m1 = \frac{[FTO (-) \text{ start reads}]/[FTO (-) \text{ depth}]}{[FTO (+) \text{ start reads}]/[FTO (+) \text{ depth}]}$  and the reads coverage difference within the FTO (+) and FTO (-) samples  $m2 = \frac{[FTO (-) \text{ depth} - FTO (+) \text{ depth}]/[FTO (-) \text{ depth}]}{[FTO (+) \text{ depth} - FTO (-) \text{ depth}]}$ . *m1* reflects the fact that m<sup>6</sup>Am site is the first transcribed nucleotide, while *m2* is a measure for demethylase sensitivity. Because the start rate in FTO (+) samples could be very low and hence prone to the high background, we used FTO (+) coverage in *m2* instead of FTO (+) start rate.

$$SRD \text{ score} = m1 + m2 \quad (1)$$

The SRD score was used to identify the m<sup>6</sup>Am site. The defined a position *i* was defined to be m<sup>6</sup>Am when the following criteria were met: (1) position *i* must be located within m<sup>6</sup>Am peak and carry adenosine residue; (2) the start reads of position *i* are not <20 in FTO (-) sample; (3) the start rate of position *i* in FTO (-) sample is greater than that in m<sup>7</sup>G-IP sample; (4) the start reads coverage in m<sup>7</sup>G-IP sample is greater than that in input; (5) the position *i* possesses the top 3 of SRD scores (SRD score > 1).

**Identification of stress-induced peak.** In order to identify the stress-induced peak, differential methylation peaks between the untreated and stress sample were calculate using exomePeak (version 2.13, *P* < 0.01) and DESeq2. The stress-induced peak was defined, when the difference between the stress and untreated sample is >2.

**Identification of PCIF1-dependent peak.** The m<sup>6</sup>Am peak in WT and PCIF1-KO sample was calculated in section “Identification of m<sup>6</sup>Am and 5'-UTR m<sup>6</sup>A peak”. To identification, the PCIF1-dependent peak, differential methylation peaks between WT and PCIF1-KO samples were calculate using exomePeak and DESeq2. The PCIF1-dependent peak was extracted, when the differential methylation ratio of WT to PCIF1-KO sample is >2 and *P*-value < 0.05.

**Analysis of RNA-seq data.** Paired-end, adapter-clean reads were aligned to the human genome (hg19, UCSC Genome Browser) using TopHat2 (version 2.0.13) with default parameters. The gene expression level was quantified as FPKM by Cufflinks (version 2.2.1)<sup>51</sup>.

**Correlations analysis of m<sup>6</sup>Am site with histone modification and TSS.** H3K27ac, H3K4me3, H3K4me1 and H3K9me3 ChIP-seq data was downloaded from ENCODE portal<sup>52</sup> (<https://www.encodeproject.org/>) with the following identifiers: ENCSR000DTU, ENCSR000FCJ, ENCSR000FCH, ENCSR000FCG. GRO-seq data were downloaded from the GEO database (GSE92375)<sup>53</sup>. We divided the m<sup>6</sup>Am sites into two parts with and without BCA motif, m<sup>6</sup>Am sites was extended 500 bp for each site, the overlapped site was removed. DeepTools were used to analyze ChIP-seq enrichment in -5 Kb to +5 Kb region around the m<sup>6</sup>Am sites.

**Motif discovery and GO enrichment analysis.** Peaks were annotated by homer software (version 4.10) with hg19 reference genome to annotate the marked genes. For the analysis of sequence consensus, the top 1000 peaks were chosen for de novo motif analysis with MEME (version 4.12.0)<sup>54</sup>, with 100 nt-long peak summit centered sequences as input. Gene Ontology (GO) enrichment analyses were performed using DAVID web-based tool (version 6.8)<sup>55</sup>.

**Statistical analysis and quantification.** For the statistical analysis of results, statistical evaluation was performed by unpaired two-sided Mann–Whitney *U*-test.

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

## Data availability

The sequence data generated in this study have been deposited in the NCBI GEO, under accession code [GSE180253](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE180253) that is publicly accessible at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE180253>. For H3K27ac, H3K4me3, H3K4me1, and H3K9me3 ChIP-seq public data were downloaded from ENCODE portal<sup>52</sup> (<https://www.encodeproject.org/>) with the following identifiers: ENCSR000DTU, ENCSR000FCJ, ENCSR000FCH, and ENCSR000FCG. GRO-seq data were downloaded from the GEO database (GSE92375)<sup>53</sup>. Source data are provided with this paper.

## Code availability

Custom code used for m<sup>6</sup>Am-seq analysis is available on GitHub ([https://github.com/KaiLi-bioinfo/m6Am\\_seq/](https://github.com/KaiLi-bioinfo/m6Am_seq/)).

Received: 17 March 2021; Accepted: 20 July 2021;

Published online: 06 August 2021

## References

- Boccaletto, P. et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
- Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
- Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G. & Suzuki, T. RNA modifications: what have we learned and where are we headed? *Nat. Rev. Genet.* **17**, 365–372 (2016).
- Gilbert, W. V., Bell, T. A. & Schaening, C. Messenger RNA modifications: form, distribution, and function. *Science* **352**, 1408–1412 (2016).
- Wei, C., Gershowitz, A. & Moss, B. N<sup>6</sup>, O<sup>2′</sup>-dimethyladenosine a novel methylated ribonucleoside next to the 5′ terminal of animal cell and virus mRNAs. *Nature* **257**, 251–253 (1975).
- Keith, J. M., Ensinger, M. J. & Moss, B. HeLa cell RNA (2′-O-methyladenosine-N<sup>6</sup>-)-methyltransferase specific for the capped 5′-end of messenger RNA. *J. Biol. Chem.* **253**, 5033–5039 (1978).
- Akichika, S. et al. Cap-specific terminal N(6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* **363**, eaav0080 (2019).
- Sun, H., Zhang, M., Li, K., Bai, D. & Yi, C. Cap-specific, terminal N(6)-methylation by a mammalian m(6)Am methyltransferase. *Cell Res.* **29**, 80–82 (2019).
- Sendinc, E. et al. PCIF1 catalyzes m(6)Am mRNA methylation to regulate gene expression. *Mol. Cell* **75**, 620–630.e629 (2019).
- Boulias, K. et al. Identification of the m(6)Am methyltransferase PCIF1 reveals the location and functions of m(6)Am in the transcriptome. *Mol. Cell* **75**, 631–643.e638 (2019).
- Fan, H. et al. PCIF1, a novel human WW domain-containing protein, interacts with the phosphorylated RNA polymerase II. *Biochem. Biophys. Res. Commun.* **301**, 378–385 (2003).
- Mauer, J. et al. Reversible methylation of m(6)Am in the 5′ cap controls mRNA stability. *Nature* **541**, 371–375 (2017).
- Wei, J. et al. Differential m(6)A, m(6)Am, and m(1)A demethylation mediated by FTO in the cell nucleus and cytoplasm. *Mol. Cell* **71**, 973–985.e975 (2018).
- Li, X., Xiong, X. & Yi, C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **14**, 23–31 (2016).
- Dominissini, D. et al. Topology of the human and mouse m(6)A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature* **485**, 201–206 (2012).
- Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
- Zhou, J. et al. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
- Meyer, K. D. et al. 5′ UTR m(6)A promotes cap-independent translation. *Cell* **163**, 999–1010 (2015).
- Coots, R. A. et al. m(6)A facilitates eIF4F-independent mRNA translation. *Mol. Cell* **68**, 504–514.e507 (2017).
- Linder, B. et al. Single-nucleotide-resolution mapping of m(6)A and m(6)Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
- Liu, J. et al. Landscape and regulation of m(6)A and m(6)Am methylome across human and mouse tissues. *Mol. Cell* **77**, 426–440.e426 (2020).
- Jia, G. et al. N<sup>6</sup>-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
- Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
- Young, J. L., Zuchner, S. & Wang, G. Regulation of the epigenome by vitamin C. *Annu. Rev. Nutr.* **35**, 545–564 (2015).
- Li, X. et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
- Li, X. et al. Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat. Chem. Biol.* **12**, 311–316 (2016).
- Chan, C. T. et al. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* **6**, e1001247 (2010).
- Lee, P., Chandel, N. S. & Simon, M. C. Cellular adaptation to hypoxia through hypoxia inducible factors and beyond. *Nat. Rev. Mol. Cell Biol.* **21**, 268–283 (2020).
- Blais, J. D. et al. Activating transcription factor 4 is translationally regulated by hypoxic stress. *Mol. Cell Biol.* **24**, 7469–7482 (2004).
- Bi, M. et al. ER stress-regulated translation increases tolerance to extreme hypoxia and promotes tumor growth. *EMBO J.* **24**, 3470–3481 (2005).
- Semenza, G. L. Hypoxia-inducible factors in physiology and medicine. *Cell* **148**, 399–408 (2012).
- An, W. G. et al. Stabilization of wild-type p53 by hypoxia-inducible factor 1α. *Nature* **392**, 405–408 (1998).
- Zhao, L. Y., Song, J., Liu, Y., Song, C. X. & Yi, C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell* **11**, 792–808 (2020).
- Yu, M. et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
- Liu, Y. et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **37**, 424–429 (2019).
- Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12**, 835–837 (2015).
- Cozen, A. E. et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* **12**, 879–884 (2015).
- Li, X. et al. Base-resolution mapping reveals distinct m(1)A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* **68**, 993–1005.e1009 (2017).
- Blaschke, K. et al. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).
- Incarnato, D. et al. High-throughput single-base resolution mapping of RNA 2′-O-methylated residues. *Nucleic Acids Res.* **45**, 1433–1441 (2017).
- Dai, Q. et al. Nm-seq maps 2′-O-methylation sites in human mRNA with base precision. *Nat. Methods* **14**, 695–698 (2017).
- Bird, J. G. et al. The mechanism of RNA 5′ capping with NAD<sup>+</sup>, NADH and desphospho-CoA. *Nature* **535**, 444–447 (2016).
- Karjilovich, J. & Yu, Y. T. Spliceosomal snRNA modifications and their function. *RNA Biol.* **7**, 192–204 (2010).
- Huang, H. et al. Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.* **20**, 285–295 (2018).
- Takahashi, H., Lassmann, T., Murata, M. & Carninci, P. 5′ end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat. Protoc.* **7**, 542–561 (2012).
- Jia, G. et al. Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.* **582**, 3313–3319 (2008).
- Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
- Meng, J. et al. A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods* **69**, 274–281 (2014).
- Feng, J. X., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying CHIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
- Sloan, C. A. et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).
- Bouvry-Livrand, M. et al. Analysis of primary microRNA loci from nascent transcriptomes reveals regulatory domains governed by chromatin architecture. *Nucleic Acids Res.* **45**, 12054 (2017).
- Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

## Acknowledgements

We would like to thank Prof. Guifang Jia and Mr. Wei Zhang for their assistance with recombinant FTO and ALKBH5 plasmids, and Xiaoyu Li and Xushen Xiong for discussions. We thank the National Center for Protein Sciences at Peking University in Beijing, China for assistance with RNA quality control and High Performance Computing Platform of the Center for Life Science for assistance with the analysis. This work was supported by the National Key R&D Program of China (Nos. 2019YFA0110900 and 2019YFA0802200), the National Natural Science Foundation of China (Nos. 91940304, 21825701, 31861143026, and 91740112 to C.Y.).

## Author contributions

H.S. and C.Y. conceived the project, designed the experiments; H.S., K.L., and C.Y. wrote the manuscript; H.S. performed the experiments with the help of M.Z. and J.L.; K.L. designed and performed the bioinformatics analysis with the help of X.Z. and H.M.; J.L. performed the CAGE experiment; M.Z. performed the PCIF1-KO cell line and participated in the discussion.

## Competing interests

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-25105-5>.

**Correspondence** and requests for materials should be addressed to C.Y.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021