nature communications

Article

https://doi.org/10.1038/s41467-025-66090-3

5-Formylcytosine is not a prevalent RNA modification in mammalian cells

Received: 6 February 2025

Accepted: 28 October 2025

Published online: 11 November 2025

Check for updates

Jasmin A. Dehnen[1,3], Alexander V. Gopanenko[1,3], Carola Scholz[1], Michael U. Musheev[1] & Christof Niehrs[1,2]

The RNA modification 5-formylcytidine (f5C) is poorly explored in mammals. Low f5C levels reported in mRNA may reflect spurious 5-methylcytidine (m5C) oxidation or targeted demethylation by TET or ALKBH1 dioxygenases. We analyzed f5C in RNA of mouse embryonic stem cells (mESCs) using LC-MS/MS and chemical-assisted sequencing. We reveal that the previously reported pyridine-borane-sequencing misidentifies N4-acetylcytidine (ac4C) and unmodified, hyper-reactive cytidines in a **C**UMC context as f5C. To overcome these limitations, we developed FIBo-seq with enhanced specificity and sensitivity for f5C-sequencing. We find no evidence for a role of TET enzymes in generating f5C, unlike for ALKBH1. Moreover, no f5C sites are detectable in mRNA. Instead, the bulk of mammalian f5C resides in the well-established mitochondrial tRNA Methionine (*mt-tRNAMet*) and is mediated by ALKBH1. The results argue against an instructive function for f5C outside tRNA in mammals.

RNA modifications play a crucial role in regulating gene expression, RNA stability, and translation[1–3]. Among them, 5-formylcytidine (f5C) has garnered interest as a potential epitranscriptomic mark[4]. However, the presence and function of f5C in RNA remain largely unexplored. In mammals, f5C is found in mitochondrial tRNA Methionine (*mt-tRNAMet*) at position 34 (f5C34), where it is essential for codon expansion during translation. f5C enables the recognition of both AUA and AUG codons, facilitating efficient protein synthesis in mitochondria[5–8]. f5C in *mt-tRNAMet* is produced stepwise, beginning with the methylation of cytidine to 5-methylcytidine (m5C) by the methyltransferase NSUN3, followed by oxidation to 5-hydroxymethylcytidine (hm5C) and finally to f5C by the dioxygenase ALKBH1[7,9].

Putative f5C sites were detected in mRNA of mouse liver[10], HEK293T cells[4], and yeast[11,12], raising the question if f5C has a broader distribution and epitranscriptomic function in RNA beyond *mt-tRNAMet*. The very low f5C levels in mRNA observed (in the ppm range, i.e. 1000x lower than m5C[4]) could also reflect spurious, substoichiometric m5C oxidation rather than specific enzymatic deposition at discrete sites.

TET enzymes (TET1/2/3) were proposed as candidates for mediating cytidine formylation in cellular RNA[10,13], since they catalyze 5-formyl-2′-deoxycytidine (5fdC) formation in DNA through a similar oxidative pathway as ALKBH1 in RNA[14,15]. Supporting this idea, TETs bind to RNA in cells and use m5C RNA as a substrate to catalyze formation of f5C in vitro[16–18]. Moreover, TET-mediated oxidation of m5C in RNA occurs in different cellular models and is linked to post-transcriptionally gene expression regulation[19–21] and leukemogenesis[22]. TET2 produces hm5C in tRNA and mRNA of mESCs and regulates RNA stability and translation[17,23,24]. In *Drosophila*, Tet-induced hm5C facilitates mRNA translation[25,26]. Nevertheless, the aforementioned studies generally did not investigate the potential function of TET-mediated f5C production in RNA, whose importance remains unknown.

Monitoring f5C in RNA is challenging due to its low quantity. Mass spectrometry, LC-MS/MS, is commonly used to quantify total f5C in RNA but lacks the spatial resolution to pinpoint f5C within transcripts. Additionally, contaminating non-coding RNA can mislead quantification of f5C[27,28]. Hence, chemical-assisted sequencing methods have been developed to map f5C at base-resolution (Table 1). Among them, pyridine-borane- (PyBo), 2-picoline-borane- (PiBo), and malononitrile- (Mal) assisted sequencing have been applied to RNA of different species to monitor f5C site-specifically. These reagents convert f5C into

[1]Institute of Molecular Biology (IMB), Mainz, Germany. [2]Division of Molecular Embryology, DKFZ-ZMBH Alliance, Heidelberg, Germany. [3]These authors contributed equally: Jasmin A. Dehnen, Alexander V. Gopanenko. e-mail: c.niehrs@imb-mainz.de

Table 1 | Chemical-assisted sequencing methods for single-base resolution mapping of f5C

Method	Chemistry & sensitivity (S)	f5C specificity	target RNA	Number of f5C sites reported	Reference
fCAB RNA-seq	O-ethylhydroxylamine & bisulfite treatment S: low	not tested	mitochondrial-enriched RNA (human)	1 (f5C34 of *mt-tRNAMet*)	52,53
RedBS RNA-seq	Na-borohydride & bisulfite treatment S:low	not tested	mitochondrial-enriched RNA (human)	1 (f5C34 of *mt-tRNAMet*)	52,53
f5C-seq	Pyridine borane (PyBo) S:high	not tested	poly(A)+ RNA (yeast)	1892	12
Protonation-dependent sequencing	Cyanoborohydride S: low	ac4C and ca5C cross-reactivity; No cross-reactivity with m5C, hm5C and C	*mt-tRNAMet* (human)	1 (f5C34 of *mt-tRNAMet*)	54
Mal-seq	Malononitrile (Mal) S: low	no cross-reactivity with m5C and hm5C	*mt-tRNAMet* (human, mouse, yeast, *Drosophila*, *C. elegans*)	human, mouse, *C. elegans*: 1 (f5C34 of *mt-tRNAMet*), yeast and Drosophila: 0	29
f5C-seq	2-picoline borane (PiBo) S: high	not tested	small RNA, poly(A)+ RNA, chromatin-associated RNA (caRNA) of human and mouse cells	human and mouse small RNA: 2 (f5C34 of *mt-tRNAMet* and f5C/f5Cm34 of *ct-tRNALeu*); poly(A)+ RNA: >100, caRNA: 3 sites in human and mouse, respectively	30
FIBo-seq	Immunoprecipitation & PyBo S: high	high	rRNA-depleted RNA from mESCs	1 (f5C34 of *mt-tRNAMet*)	This study

derivatives that are read as uridine during reverse transcription, leading to C-to-T conversions upon sequencing[12,29,30].

Here, we investigate the presence and distribution of f5C in RNA from mouse embryonic stem cells (mESCs) using a combination of LC-MS/MS and chemical-assisted sequencing techniques. Our analysis highlights significant limitations in pyridine-borane-assisted sequencing (PyBo-seq), because it misidentifies unmodified hyper-reactive cytidines in a **C**UMC context as f5C. We also demonstrate that PyBo produces a C-to-T signature at ac4C sites in rRNA and in in vitro-transcribed (IVT) RNA oligonucleotides, making ac4C indistinguishable from f5C. To address this problem, we developed an improved f5C sequencing method, FIBo-seq (f5C-immunoprecipitation-pyridine-borane sequencing), with enhanced specificity and sensitivity for f5C detection. We confirm the established role of ALKBH1 in producing f5C in RNA but find no role for TET enzymes in this process. Moreover, we detect no f5C sites in mRNA and validate this result by LC-MS/MS, suggesting that f5C does not play an epitranscriptomic role in mRNA regulation. Instead, we confirm that f5C in mammalian RNA is concentrated in *mt-tRNAMet*, where it is formed by ALKBH1. These findings suggest that in mammals, f5C is primarily restricted to tRNA and lacks a broader regulatory function in mRNA.

Results

PyBo-seq reveals ALKBH1- and TET-independent putative f5C sites in mESCs

To profile f5C in the transcriptome of mESCs at base resolution, we employed PyBo[12]. By LC-MS/MS, we confirmed that PyBo treatment of total RNA from mESCs effectively converts f5C to dihydrouridine (DHU) (Fig. 1a, b) without affecting total levels of C, m5C, and hm5C (Supplementary Fig. 1a). To evaluate the efficiency of PyBo in causing C-to-T conversions at specific f5C sites, we applied PyBo followed by reverse transcription (RT) and sequencing of an f5C-modified RNA oligonucleotide containing a single f5C site. We observed nearly complete C-to-T conversion at the f5C site by Sanger sequencing (Fig. 1c), confirming a high misincorporation during RT.

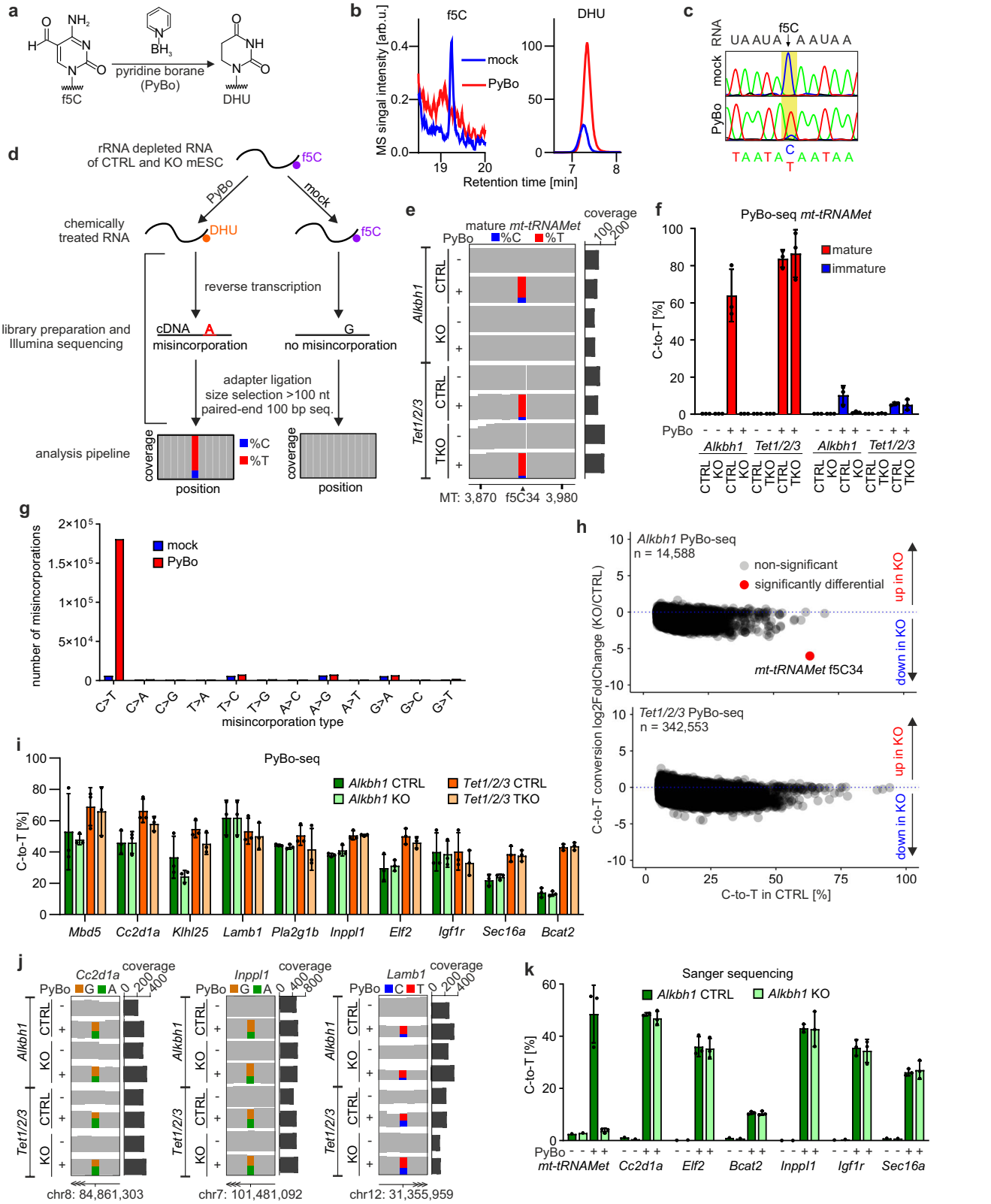
To dissect the role of ALKBH1 and TET enzymes on f5C in RNA, we generated *Alkbh1* knockout mESCs (*Alkbh1* KOs) (Supplementary Fig. 1b–e) and employed established *Tet1/2/3* triple knockout mESCs (*Tet1/2/3* TKOs)[31]. To control for clonal variation, each mutant line was compared to its own corresponding parental cell line (*Alkbh1* CTRL, *Tet1/2/3* CTRL). Following ribosomal RNA (rRNA) depletion, remaining RNA was subjected to PyBo-assisted ultra-deep high-throughput sequencing (PyBo-seq) with 200 million reads/sample on average (Fig. 1d; Supplementary Data 1). We obtained ~6-fold higher coverage compared to standard RNA-seq with 40 million reads/sample (Supplementary Fig. 1f), resulting in a high sequencing depth (avg. >100 reads/position) for ~8000 transcripts (Supplementary Fig. 1g; Supplementary Data 1). A high sequencing depth is crucial for modification detection, as this improves reproducibility of modification stoichiometry quantification (Supplementary Fig. 1h, i). We investigated >19,000 unique transcripts, including coding RNA, lncRNA, snRNA, pseudogenes and tRNA. Among them more than 12,734 protein-coding transcripts (matching 58% of all annotated murine genes) had an average coverage of >5 reads/position (Supplementary Fig. 1g) and therefore could be analyzed for the presence of C-to-T conversions. Thus, our ultra-deep sequencing approach was appropriate for the detection of f5C with moderate to low stoichiometry and the identification of C-to-T conversions even in lowly expressed transcripts.

As a positive control, we first analyzed f5C34 reads in *mt-tRNAMet*. To monitor mature *mt-tRNAMet* transcripts, we excluded reads mapping beyond *mt-tRNAMet* gene boundaries, which correspond to unprocessed pre-mitochondrial RNA[32]. Examination of sequencing reads at f5C34 revealed that PyBo treatment caused C-to-T sites at cytidine formylation, which were absent in *Alkbh1* KOs but remained unaffected in *Tet1/2/3* TKOs (Fig. 1e; Supplementary Data 2), confirming that ALKBH1 catalyzes f5C34. In control mESCs, we observed 60–80% C-to-T conversion at f5C34 in PyBo-seq, showing excellent agreement with Sanger sequencing results from the same samples (Fig. 1f; Supplementary Fig. 1j). Similar f5C34 modification levels were previously reported in mouse *mt-tRNAMet*[29,33].

To monitor f5C in immature, unspliced *mt-tRNAMet*, we identified polycistronic, unprocessed transcripts by selecting the reads that map to the locus and its flanking regions. We observed a C-to-T conversion below 5% at f5C34 (Fig. 1f; Supplementary Data 2), indicating that formation of f5C occurs on mature rather than immature *mt-tRNAMet*.

Analyzing the whole transcriptome, we expectedly found increased C-to-T conversions upon PyBo treatment (Fig. 1g). We identified 7516 high-confidence C-to-T sites shared between all CTRL mESC lines ($n = 6$) (Supplementary Data 3). Most sites showed only

https://doi.org/10.1038/s41467-025-66090-3



5–10% conversion but 146 sites had a conversion greater than 30% (Supplementary Fig. 1k). C-to-T sites with >30% conversion typically appeared only once per transcript (Supplementary Fig. 1l). To identify C-to-T sites that are ALKBH1- or TET-dependent, we performed differential analyses of C-to-T levels by comparing *Alkbh1* KOs or *Tet* TKOs with their corresponding CTRL mESCs. Surprisingly, we did not detect any ALKBH1-dependent C-to-T sites beyond f5C34 *mt-tRNAMet* nor did we detect TET1/2/3-dependent C-to-T sites (Fig. 1h). C-to-T sites in selected RNAs with high conversion were all unaffected in KO cells (Fig. 1i, j; Supplementary Data 2). Furthermore, we validated the reproducibility of high C-to-T conversion sites with an independent PyBo treatment and Sanger sequencing (Fig. 1k), thus excluding the possibility of artifacts in NGS-based modification mapping. We conclude that PyBo-seq reliably detects the ALKBH1-dependent f5C site in

Fig. 1 | Pyridine-borane-seq reveals ALKBH1- and TET-independent C-to-T sites in mESCs. a The reaction of pyridine borane (PyBo) with f5C to DHU. **b** LC-MS/MS chromatograms of f5C and DHU in mock-treated or PyBo-treated total RNA of mESCs. **c** PyBo-induced C-to-T signature of an in vitro-transcribed f5C-RNA oligonucleotide after reverse transcription and Sanger sequencing. **d** Scheme of pyridine-borane-assisted RNA sequencing (PyBo-seq). **e** Integrative genomics viewer (IGV) browser views of f5C34 in *mt-tRNAMet* from PyBo-seq. Sequencing reads were extracted for mature *mt-tRNAMet*. Blue/red bars indicate the proportion of C or T reads at f5C34. Coverage: bars on the right show the number of reads mapping to the C-to-T site at position f5C34. **f** Quantification of C-to-T conversion at f5C34 in CTRLs, *Alkbh1* KO and *Tet1/2/3* TKO mESC from PyBo-seq in mature and immature *mt-tRNAMet*. Mean ± SD values are shown ($n = 3$ biological replicates). **g** Analysis of nucleobase-conversion pattern upon PyBo treatment in *Alkbh1* CTRL samples relative to the reference genome. Shown is the mean of $n = 3$ biological replicates. **h** C-to-T site MA plots for *Alkbh1* KO vs. *Alkbh1* CTRL (top) and *Tet1/2/3* TKO vs. *Tet1/2/3* CTRL (bottom). Only sites that were covered in CTRL and KO are shown. For *mt-tRNAMet*, C-to-T levels from mature transcripts were analyzed. **i** Quantification of high C-to-T conversions in mRNAs from PyBo-seq in CTRLs, *Alkbh1* KO and *Tet1/2/3* TKO mESC. Mean ± SD values are shown ($n = 3$ biological replicates). C-to-T positions are detailed in Supplementary Data 2. **j** IGV browser views of C-to-T sites in three representative mRNAs. Coverage: bars on the right show the number of reads mapping to the C-to-T site. For RNA transcribed from the minus strand, C-to-T events appear as G-to-A conversions. **k** Quantification of C-to-T conversions from Sanger-seq chromatograms in selected mRNAs identified by PyBo-seq. Mean ± SD values are shown (n = 3 biological replicates for PyBo-treated samples; $n = 1$ for mock-treated samples). Source data are provided as a Source Data file for panels **b**, **f**, **g**, **h**, **i** and **k**.

mt-tRNAMet and reports numerous other ALKBH1- and TET-independent C-to-T sites.

Putative f5C sites lack sequence features of known m5C modifications

The f5C modification requires prior methylation of cytidine to m5C, which is catalyzed by enzymes of the NSUN family. NSUN3 specifically targets position C34 in *mt-tRNAMet*, but the mechanism by which it recognizes this substrate remains unclear. No additional substrates for NSUN3 have been identified[7] and its loss does not affect global mRNA methylation levels[18]. In contrast, NSUN2, NSUN5, and NSUN6 catalyze m5C formation in mRNA within defined sequence contexts. Given that most C-to-T sites detected by PyBo-seq occur in mRNA, we assessed their potential to be substrates of NSUN2, NSUN5, and NSUN6 by examining whether the sequence context of the C-to-T sites aligns with the known recognition motifs of these enzymes (NSUN2: **C**AGG and **C**KGGG (K = G or U); NSUN5: **C**ARAU (R = G or A); NSUN6: **C**UCCA)[18].

We focused our analysis on PyBo-induced C-to-T sites within the CDS, 3'UTR or 5'UTR of mRNAs that exhibited at least 5% conversion across all CTRL mESCs. The frequencies of NSUN2 and NSUN5 motifs at PyBo-induced C-to-T sites matched random expectation (Supplementary Fig. 2a). The NSUN6 motif was ~ twice as frequent as expected (69 of 4780 sites) but none of these C-to-T sites overlaps with experimentally confirmed NSUN6 cross-link sites[18] (Supplementary Fig. 2b; Supplementary Data 4), indicating that C-to-T- and NSUN6 sites are functionally unrelated.

To investigate how the C-to-T sites relate to known m5C sites, we compared our data to previously mapped m5C sites in mESCs but observed no overlap (Supplementary Fig. 2c). Collectively, these findings do not support that the putative f5C sites originate from m5C, aligning with their ALKBH1- and TET independence.

Mal-seq fails to confirm putative f5C sites in mESC RNA

Given the unexpected finding of m5C- and ALKBH1- and TET-independent C-to-T sites, we re-evaluated the specificity of PyBo-sequencing. We tested whether PyBo treatment produces C-to-T conversions in IVT RNA oligonucleotides harboring other modified cytidine nucleotides, including ca5C, ac4C, m5C, hm5C and Cm (Fig. 2a). Both ca5C and ac4C RNA oligonucleotides also showed PyBo-induced C-to-T conversions at the modified site. ca5C was not pursued further because its levels are exceedingly low in cellular RNA as measured by LC-MS/MS[4] and hence ca5C is unlikely to account for the numerous C-to-T sites with high conversion observed in PyBo-seq. ac4C, on the other hand, is prevalent in various RNA, including rRNA, tRNA and mRNA[34]. However, PyBo-induced conversion at ac4C was only 13% and hence rather inefficient (Fig. 2a). We confirmed by LC-MS/MS that PyBo treatment converts ac4C to DHU (Supplementary Fig. 3a). Two ac4C modified sites in *18S rRNA*, ac4C1337 and ac4C1842, are highly conserved with modification levels of 79% and 99%, respectively[35]. Indeed, PyBo-seq retrieved these two ac4C sites in samples without rRNA-depletion (Fig. 2b; Supplementary Fig. 3b), confirming that ac4C is detectable by PyBo-seq. The C-to-T conversions at these ac4C sites ranged from 8 to 28%, suggesting that the conversions underestimate ac4C stoichiometry (Fig. 2b; Supplementary Fig. 3b). Other *18S rRNA* modifications (Cm, m6A, pseudouridine) were not detected by PyBo-induced conversions, confirming the results of IVT RNA oligonucleotides and expanding it beyond cytidine modifications (Supplementary Fig. 3c). The low C-to-T conversion efficiency of ac4C in *18S rRNA* suggests that high conversion sites in PyBo-seq are not caused by ac4C. Consistently, siRNA knockdown of *Nat10*, the only known ac4C writer protein in mESCs[36], reduced global ac4C levels by 50% but left C-to-T conversions in *Cc2d1a*, *Elf2* and *Bcat2* transcripts unaffected (Supplementary Fig. 3d).

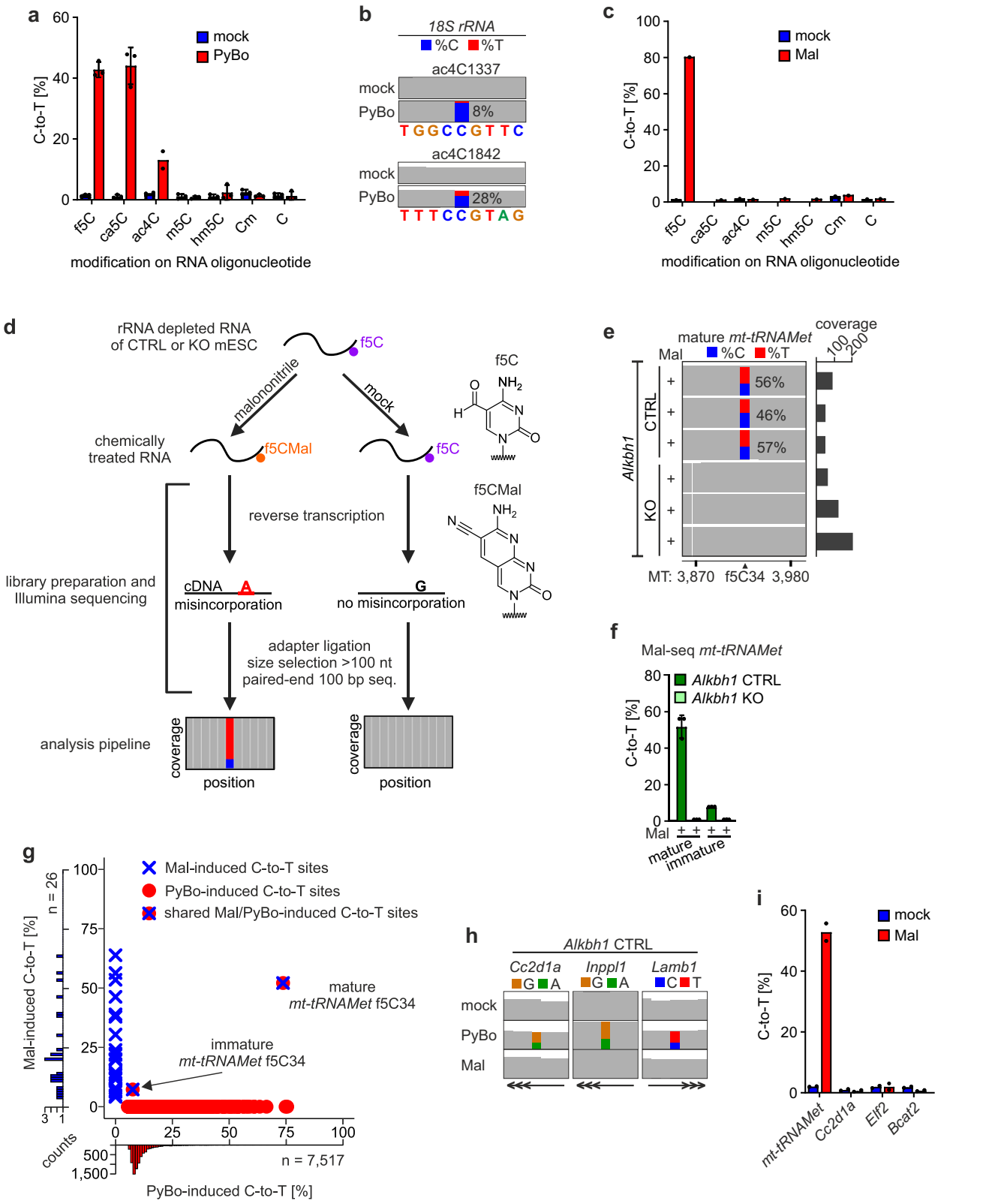
Given the non-specificity of PyBo-seq, we explored alternative reagents that generate C-to-T signatures at f5C sites. We reanalyzed high throughput RNA-seq data from samples treated with 2-picoline borane (PiBo), a structural analog of PyBo, and found that PiBo also cross-reacts with ac4C in high-throughput sequencing (Supplementary Fig. 3e, f). In contrast, malononitrile (Mal) showed greater specificity to f5C in comparison to PyBo when tested on IVT RNA oligonucleotides with modified cytidine nucleotides (Fig. 2c). We noticed that both PyBo and Mal also react with 5-formyl-2'-O-methylcytidine (f5Cm) (Supplementary Fig. 3g), which is expected because the chemical reaction engages the base and is independent of the sugar moiety. We performed transcriptome-wide ultra-deep sequencing with Mal (Mal-seq) in *Alkbh1* CTRLs and KOs (Fig. 2d; Supplementary Data 1; Supplementary Data 5). Mal-seq reliably detected f5C sites in *mt-tRNAMet* but was less efficient than PyBo in inducing C-to-T conversions (Fig. 2e, f; Supplementary Data 2), a known limitation of Mal treatment[4,12]. We compared the positions of Mal- and PyBo-induced C-to-T conversions. Surprisingly, f5C34 in *mt-tRNAMet* was the only shared C-to-T site between Mal-seq and PyBo-seq datasets (Fig. 2g, h). Sanger sequencing of independently Mal-treated RNA also failed to support putative f5C sites identified by PyBo-seq (Fig. 2i).

FIBo-seq shows superior specificity and sensitivity for f5C profiling in RNA

Given the low f5C conversion efficiency in Mal-seq, we thought to overcome the specificity-issues of PyBo-seq by developing a two-dimensional approach, FIBo-seq (f5C-immunoprecipitation-pyridine-borane-seq). The procedure consists of pre-enriching f5C-containing RNA with an f5C antibody followed by PyBo-seq (Fig. 3a). We scored C-to-T events in peaks only as f5C if the C-to-T signal was absent in mock-treated RNA and if the conversion efficiency after enrichment was similar to that observed at the enriched f5C34 position in *mt-tRNAMet*.

Using *mt-tRNAMet* as a control, we confirmed by RT-qPCR >500-fold enrichment of f5C-containing transcripts with an f5C antibody. RNA that was f5C-depleted (Δf5C) by PyBo treatment showed no

https://doi.org/10.1038/s41467-025-66090-3



enrichment for *mt-tRNAMet* (Fig. 3b). We conducted FIBo-seq with RNA from *Tet1/2/3* CTRL and TKO mESCs (Supplementary Data 1) and first analyzed the *mt-tRNAMet* locus for read coverage and C-to-T conversions (Fig. 3c). Control and mutant cell lines yielded comparable results for *mt-tRNAMet*, demonstrating a high reproducibility across samples. Antibody-based enrichment of f5C resulted in over 30-fold increase of *mt-tRNAMet* transcripts compared to input (Fig. 3d). Consistently, for f5C34 in *mt-tRNAMet*, FIBo-seq returned 98% of reads with a C-to-T conversion without the need to filter for reads belonging to mature *mt-tRNAMet* (Fig. 3e). While this increase in read coverage should have greatly improved the detection of cytidine formylation, no additional f5C sites were identified using 90% C-to-T conversion as threshold for the most prominent C-to-T event in each peak (Fig. 3f 'max'; Supplementary Data 6). In fact, only 24 peaks had a C-to-T event,

Fig. 2 | PyBo- and Mal-seq fail to detect novel f5C sites in mESC RNA. **a** Quantification of PyBo-induced C-to-T conversions from Sanger-seq chromatograms for in vitro-transcribed RNA oligonucleotides single-site modified as indicated. Mean ± SD values are shown ($n = 3$ experimental replicates; $n = 2$ for ac4C samples). **b** IGV browser views of the two known ac4C sites in *18S rRNA* from samples without rRNA depletion in PyBo-seq. Blue/red bars indicate the proportion of C or T reads at the position of ac4C. **c** Quantification of Mal-induced C-to-T conversions from Sanger-seq chromatograms for the RNA oligonucleotides as in (**a**) ($n = 1$ experimental replicate). **d** Scheme of malononitrile-assisted RNA sequencing (Mal-seq). **e** IGV browser views of f5C34 in *mt-tRNAMet* from Mal-seq. Sequencing reads were extracted for mature *mt-tRNAMet*. Blue/red bars indicate the proportion of C or T reads at f5C34. Coverage: bars on the right show the number of reads mapping to the C-to-T site at position f5C34. **f** Quantification of C-to-T conversion at f5C34 in *Alkbh1* CTRL and KO mESC from Mal-seq in mature and immature *mt-tRNAMet*. Mean ± SD values are shown ($n = 3$ biological replicates). **g** Comparison of C-to-T sites and their conversion from Mal-seq and PyBo-seq. f5C34 *mt-tRNAMet* is shown twice, corresponding to the mature and immature transcript. **h** IGV browser views of C-to-T sites in three representative mRNAs from PyBo-seq and Mal-seq samples. For RNA transcribed from the minus strand, C-to-T events appear as G-to-A conversions. **i** Quantification of Mal-induced C-to-T conversions from Sanger-seq chromatograms in *Alkbh1* CTRL for selected sites identified by PyBo-seq ($n = 2$ biological replicates). Source data are provided as a Source Data file for panels **a**, **c**, **f**, **g**, and **i**.

with *Set* showing the second highest conversion of 9.3% C-to-T, after *mt-tRNAMet* (Fig. 3g). To find peaks with near 100% C-to-T signal, we summed up the conversion levels of all sites in each peak, expecting that the cumulative conversion for f5C-containing peaks be at least 98%, as was the case for f5C34 *mt-tRNAMet*. This analysis addressed the possibility that cytidine formylation can occur at variable positions within a designated region rather than at a single cytidine and that peaks arise from multiple f5C sites in close proximity. Cumulative analysis returned ~110% C-to-T conversion in *mt-tRNAMet* peaks, although its only known f5C site is f5C34. This suggests that the cumulative C-to-T analysis reported some background from PyBo treatment. Regardless, even this sensitized analysis returned no new f5C sites (Fig. 3f 'sum'; Supplementary Data 6). In summary, FIBo-seq has superior specificity and sensitivity for f5C profiling in RNA. Nevertheless, in mESCs, we found no apparent role for TET enzymes in cytidine formylation as the only f5C site retrieved was f5C34 *mt-tRNAMet*.

Depletion of *mt-tRNAMet* from the transcriptome of mammalian cells abolishes f5C signal in LC-MS/MS

To independently verify that f5C in mESC RNA is confined to *mt-tRNAMet*, we quantified f5C levels in RNA using ultrasensitive LC-MS/MS. The absolute levels of f5C in total RNA normalized to unmodified C were between ~2-7 ppm in different mESC clonal control lines (Fig. 4a), as reported for mESCs (~ 3 ppm[37]), and human cell lines (~ 10 ppm in HEK293T[4], ~9 ppm in HeLa and HEK293T cells[13]). In *Alkbh1* KO cells, f5C became undetectable, indicating that essentially all f5C is ALKBH1-dependent. f5Cm, a derivative of f5C found in cytosolic *tRNALeu(CAA)*[9], also became undetectable in *Alkbh1* KOs (Fig. 4a), indicating that this enzyme accounts for f5Cm generation as reported[9]. In contrast, f5C and f5Cm levels were not significantly affected in *Tet1/2/3* TKO and *Tet1/2* double knockout (DKO) mESCs (Fig. 4a). Unlike f5C, hm5C is generated by ALKBH1 and TET enzymes[4]. Hence, we quantified hm5C and m5C levels in total RNA of mESCs and found for hm5C between 7–16 ppm and for m5C 500–1000 ppm. In *Alkbh1* and *Tet1/2/3* KO cells, hm5C levels were expectedly reduced while m5C remained unaffected (Fig. 4a). LC-MS/MS analysis from *Tet1/2/3* CTRL cells confirmed the absence of f5C and f5Cm in mRNA (Fig. 4a), consistent with FIBo-seq results. Low levels of hm5C and m5C were detectable, confirming that these modifications are present in mRNA[18,23].

To analyze f5C in cellular RNA by LC-MS/MS excluding *mt-tRNAMet*, we removed the transcript from total RNA of mESCs by hybrid capture using biotinylated antisense-DNA oligonucleotides (Fig. 4b). RT-qPCR analysis confirmed ~31,000-fold enrichment of *mt-tRNAMet* in the bound fraction and near complete depletion in the unbound fraction, while non-target transcripts (*tRNALys*, *Tbp*) were unaffected (Fig. 4c). LC-MS/MS analysis of the fractions showed 241-fold f5C enrichment in the bound fraction and f5C disappearance in the *mt-tRNAMet*-depleted fraction (Fig. 4d). f5Cm, hm5C and m6A decreased in the bound fraction, indicating that these modifications are absent in *mt-tRNAMet*. m5C was readily detected in the bound fraction supporting that *mt-tRNAMet* is m5C modified[33,38].

Since depletion of *mt-tRNAMet* abolished f5C signals in cellular RNA, this analysis supports the conclusion that f5C is restricted to the *mt-tRNAMet* transcript. To test if this conclusion extends beyond mESCs, we conducted *mt-tRNAMet* hybrid capture in two additional mouse cell lines (NIH/3T3 and C2C12) and two human cell lines (HeLa and HEK293T). *mt-tRNAMet* was enriched up to 10,700-fold in the bound fraction and was largely depleted from the unbound fraction in NIH/3T3, HeLa and C2C12 cells (Fig. 4e, g; Supplementary Fig. 4a). Notably, f5C was undetectable in cellular RNA following *mt-tRNAMet* depletion (Fig. 4f, h; Supplementary Fig. 4b), indicating that the absence of f5C outside *mt-tRNAMet* is consistent across multiple cell types. In HEK293T cells, we obtained ~9,200-fold RNA enrichment and there was substantial *mt-tRNAMet* RNA left in the unbound fraction (Supplementary Fig. 4c). LC-MS/MS signal of f5C in the bound fraction was also 2.5x lower compared to mESCs, altogether suggesting less efficient hybrid capture than in mESCs. Nevertheless, the unbound fractions showed >50% reduction of f5C (Supplementary Fig. 4d). Given that residual f5C in the unbound fraction is likely due to incomplete *mt-tRNAMet* removal, we conclude that *mt-tRNAMet* accounts for most- if not all- f5C signal also in HEK293T cells.

Pyridine borane induces C-to-T sites at unmodified cytidines in exposed CUMC motifs

We sought to understand the origin of PyBo-induced C-to-T sites that cannot be attributed to f5C, ac4C, or ca5C. Sequence motif analysis in CDS of mRNAs from PyBo-seq revealed a four-base **C**UMC motif (M = A or C) at positions 0 to +3 relative to the C-to-T site (Fig. 5a). The prominence for the **C**UMC motif increased with higher C-to-T conversion level. We also observed the **C**UMC motif in 5′UTR and 3′UTR sequences (Supplementary Fig. 5a). To investigate whether this motif is associated with PyBo-induced C-to-T sites across species, we analyzed public PyBo-seq datasets from *S.cerevisiae* and four human cell lines (HEK293T, HeLa, HepG2 and MCF-7)[12]. In yeast, 70% of C-to-T sites with a conversion >20% were located within a **C**UMC motif (Supplementary Fig. 5b). Similarly, in human mRNAs, seven of the ten sites with highest C-to-T conversion shared across all four cell lines occurred within this motif (Supplementary Fig. 5c; Supplementary Data 7). The most prominent site, located in the *KIF11* transcript, exhibited a **C**UAC sequence context and 85% C-to-T conversion (Supplementary Data 7). Independent validation of this site by Sanger sequencing confirmed a strong reactivity with PyBo (90% C-to-T) but no signal with malononitrile (Supplementary Fig. 5d), indicating that this site resembles murine PyBo-induced C-to-T sites and does not represent f5C. Together, we demonstrated that the **C**UMC motif is a consistent feature of highly PyBo-reactive cytidines across species.

Comparison of the number of C-to-T sites with the number of **C**UMC motifs in the respective UTRs and CDS revealed that there are more **C**UMC motifs in the transcriptome than reported PyBo-induced C-to-T conversions (Supplementary Fig. 5e). We also noted a higher ratio of C-to-T events to **C**UMC motifs in the 5′UTR compared to the CDS and 3′UTR. The fact that there are more **C**UMC motifs in the transcriptome than PyBo-induced C-to-T conversions implies that the

https://doi.org/10.1038/s41467-025-66090-3

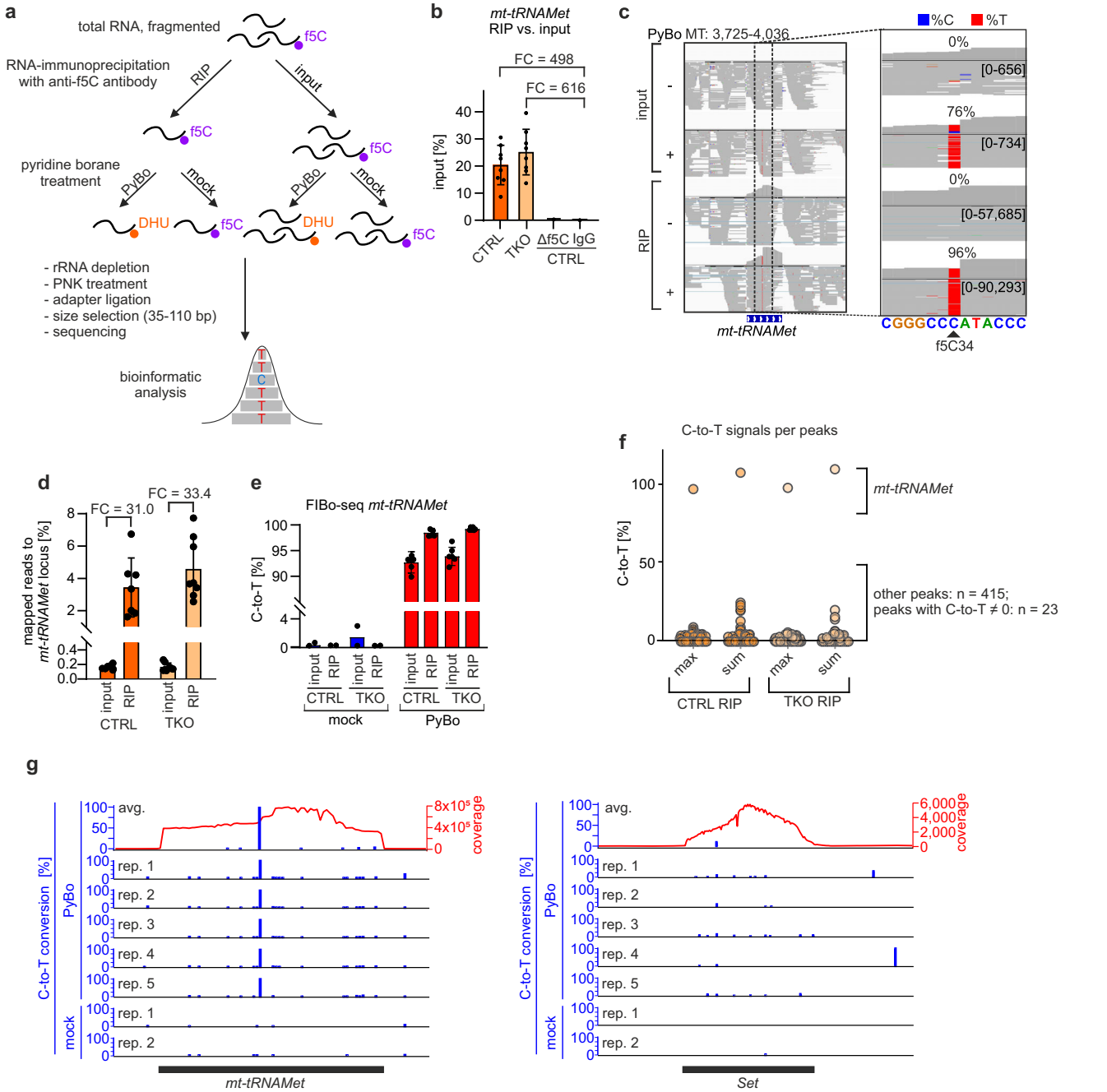


Fig. 3 | FIBo-seq has high sensitivity and specificity towards f5C. a Scheme of FIBo-seq (f5C-immunoprecipitation-pyridine-borane-seq). **b** RT-qPCR quantification of *mt-tRNAMet* pulldown efficiency with an f5C-antibody in *Tet1/2/3* CTRL and TKO total RNA samples. In the Δf5C sample, f5C was depleted from the total RNA through the reaction with PyBo prior to antibody-pulldown. In the IgG control, total, untreated RNA was incubated with rabbit IgG instead of f5C-antibody. Data are normalized to input samples. Mean ± SD values are shown ($n = 8$ experimental replicates; $n = 1$ for Δf5C and IgG). FC = fold change. **c** IGV browser views of f5C34 in *mt-tRNAMet* from FIBo-seq. Blue/red bars indicate the proportion of C or T reads at f5C34. The maximal coverage is shown in brackets. **d** Quantification of reads in FIBo-seq samples mapping to the *mt-tRNAMet* locus, shown as percentage of total mapped reads. Mean ± SD values are shown ($n = 8$ experimental replicates; $n = 7$ for CTRL RIP). FC = fold change. **e** Quantification of C-to-T conversion at f5C34 of *mt-tRNAMet* in FIBo-seq data from *Tet1/2/3* CTRL and TKO mESC. Mean ± SD values are shown ($n = 2$ experimental replicates for mock samples; $n = 6$ for PyBo samples except $n = 5$ for PyBo CTRL RIP). **f** Quantification of C-to-T conversions within FIBo-seq peaks from RIP samples. Max = single C-to-T site with the highest conversion level per peak. Sum = cumulative C-to-T conversion level of all C-to-T sites within one peak. **g** Examples of peaks with C-to-T conversions in FIBo-seq. Black bars indicate peak boundaries. Source data are provided as a Source Data file for panels **b**, **d**–**f**.

CUMC motif alone is insufficient to explain cytidine reactivity and suggests that RNA structure is also important for C-to-T conversions. Concordantly, RNAfold-based structure prediction for RNA regions around C-to-T sites showed stem-loop-like structures, with PyBo-susceptible cytidines locating to the predicted loops (Fig. 5b, c; Supplementary Fig. 5f). These results suggest that unmodified, unpaired cytidines in **C**UMC motifs are hyper-reactive with PyBo. To test this possibility, we selected three transcripts (*Cc2d1a, Inppl1, Elf2*) from PyBo-seq with high C-to-T conversion at **C**UMC motifs (Supplementary Data 2). We synthesized unmodified RNA fragments by in vitro transcription (Cc2d1a 232mer$^{86C\text{-}to\text{-}T}$, Inppl1 390mer$^{247C\text{-}to\text{-}T}$, Elf2 339mer$^{157C\text{-}to\text{-}T}$) encompassing the sequences required to form the

https://doi.org/10.1038/s41467-025-66090-3

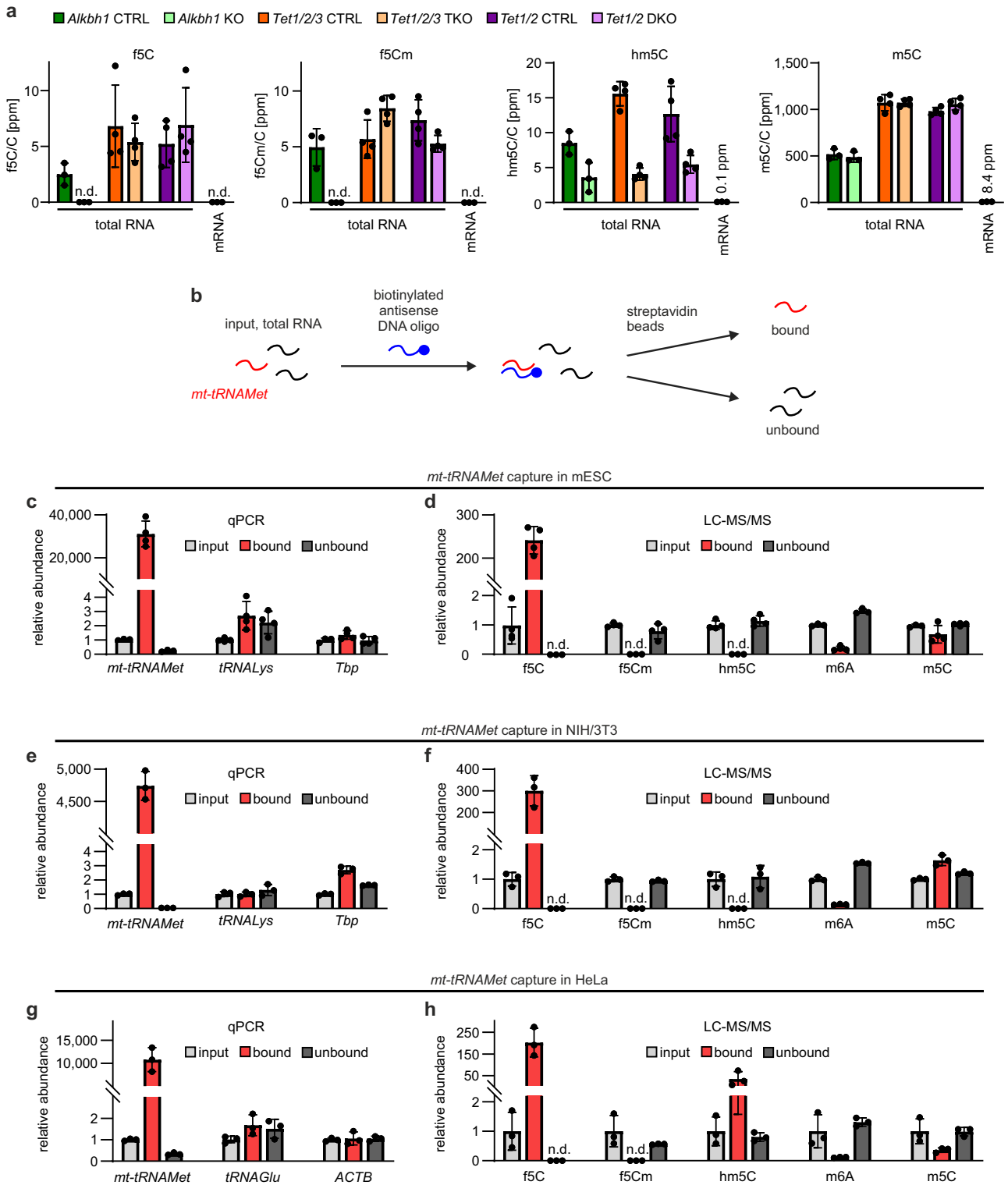
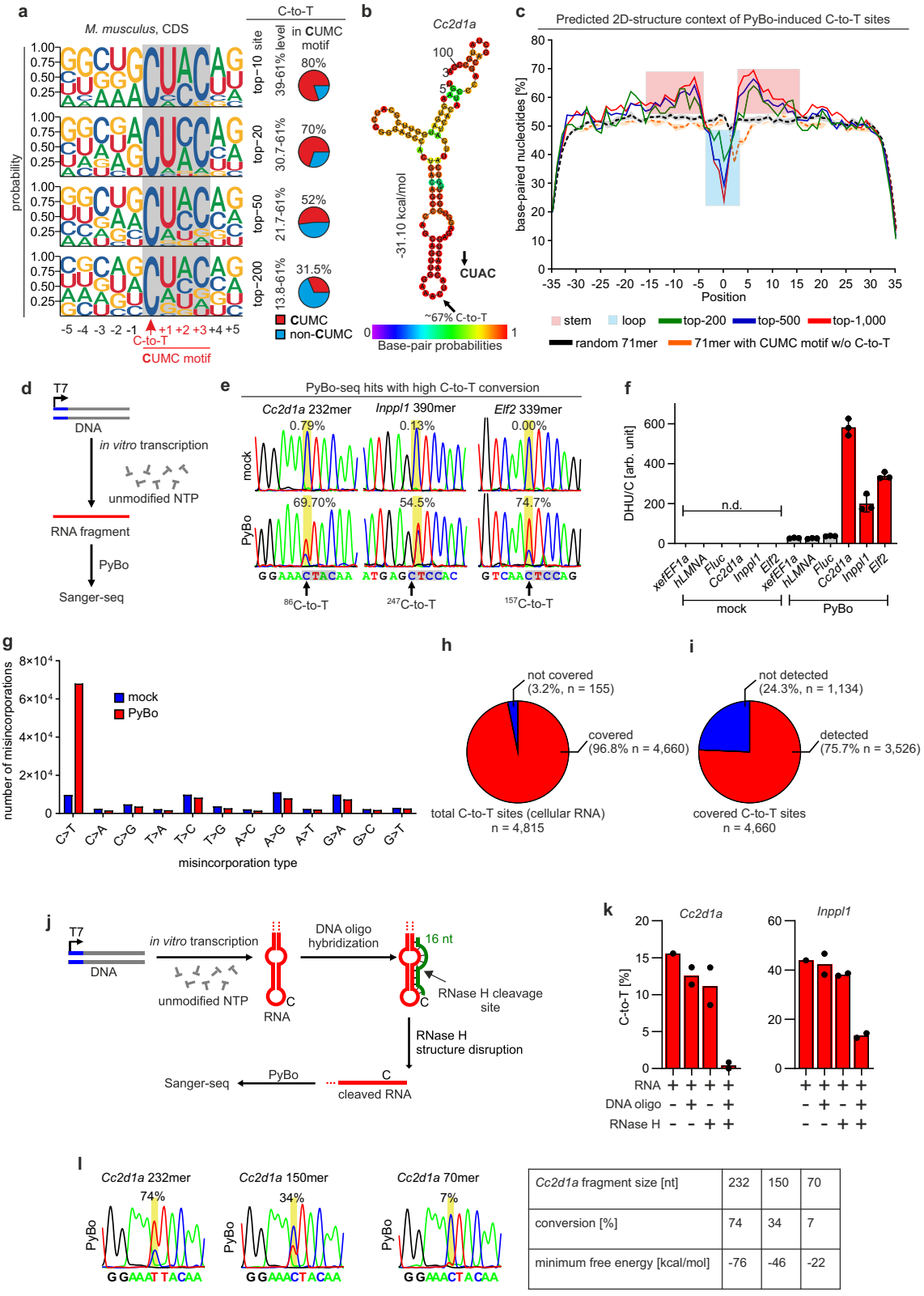


Fig. 4 | Most f5C is restricted to *mt-tRNAMet*. **a** LC-MS/MS quantification of indicated base modifications in total RNA from CTRL, *Alkbh1* KO, *Tet1/2/3* TKO and *Tet1/2* DKO mESCs and mRNA from *Tet1/2/3* CTRL mESCs. Mean ± SD values are shown ($n = 3$ biological replicates of total RNA from *Alkbh1* CTRL, KO and mRNA from *Tet1/2/3* CTRL; $n = 4$ biological replicates of total RNA from *Tet1/2/3* CTRL and TKO), n.d. = not detected. **b** Hybrid capture of *mt-tRNAMet*. **c** RT-qPCR quantification of *mt-tRNAMet* and control transcripts in the bound and unbound fraction after *mt-tRNAMet* capture relative to input total RNA of mESCs. RNA levels were normalized to *18S rRNA*. Mean ± SD values are shown ($n = 4$ experimental replicates). **d** LC-MS/MS quantification of modified nucleosides in the bound and unbound fraction after *mt-tRNAMet* capture relative to input total RNA of mESCs. Data for modified nucleosides were normalized to unmodified nucleoside. Mean ± SD values are shown ($n = 4$ experimental replicates), n.d. = not detected. As in (**c**) and (**d**) but with RNA of NIH/3T3 cells (**e**, **f**) and HeLa cells (**g**, **h**). Mean ± SD values are shown ($n = 3$ experimental replicates), n.d. = not detected. Source data are provided as a Source Data file for panels **a**, **c**–**h**.

https://doi.org/10.1038/s41467-025-66090-3



Cc2d1a fragment size [nt]	232	150	70
conversion [%]	74	34	7
minimum free energy [kcal/mol]	-76	-46	-22

predicted hairpins for each of the three RNAs (Fig. 5d). Upon PyBo treatment, Sanger sequencing of all three RNAs indeed showed robust C-to-T conversion at the expected **C**UMC sites (Fig. 5e). Moreover, using LC-MS/MS, we found that DHU, the reaction product of cytidine with PyBo, was greatly increased in the three RNAs but not in control RNAs (*xefEF1α*, *hLMNA, Fluc*) (Fig. 5f). We next applied PyBo-seq to a modification-free transcriptome generated by in vitro transcription of poly(A)+ RNA from mESCs. C-to-T conversions increased 7-fold following PyBo treatment (Fig. 5g), indicating that a substantial number of unmodified cytidines throughout the transcriptome is susceptible to PyBo. Of the 4815 C-to-T sites identified in cellular mRNA, 4660 sites (96.8%) were evaluable (Fig. 5h; Supplementary Data 8). Remarkably, 75.7% (3,526 sites) of these overlapped with C-to-T sites detected in IVT RNA (Fig. 5i; Supplementary Data 8), and most representative high-

Fig. 5 | PyBo reacts with hyper-reactive cytidines in exposed CUMC motifs. **a** Sequence motifs surrounding C-to-T sites identified by PyBo-seq in the CDS of mRNAs from CTRL mESCs. Transcripts were grouped according to the C-to-T conversion levels. Logos represent nucleotide probability at each position. Pie charts indicate the percentage of transcripts with **C**UMC motif at positions 0 to +3 relative to the C-to-T sites. **b** RNAFold-prediction of the hairpin loop structure in *Cc2d1a* mRNA (CDS) containing a C-to-T site within a **C**UMC motif. Arrows indicate location of the C-to-T site. **c** Metaprofiles of secondary structures of PyBo-induced C-to-T sites and flanking regions (± 35 nt) with C-to-T site at position 0 compared to random 71mers and 71mers with **C**UMC motif without (w/o) C-to-T site. **d** Outline of in vitro transcription and PyBo treatment of selected C-to-T sites from PyBo-seq. **e** Sanger-seq chromatograms with quantification of C-to-T conversion in unmodified in vitro-transcribed RNA after mock- or PyBo treatment. **f** LC-MS/MS quantification of DHU in in vitro-transcribed control RNAs (*xefEF1α*, *hLMNA*, *Fluc*) and RNAs from (**e**). Mean ± SD values are shown ($n = 3$ experimental replicates). **g** Analysis of nucleobase-conversion pattern after PyBo treatment in IVT RNA relative to the reference genome. Shown is the mean of n = 2 experimental replicates. **h** Fraction of C-to-T sites (cellular RNA) for which the corresponding cytidine position has sufficient read depth (> 5 reads) in the IVT transcriptome PyBo-seq dataset. **i** Fraction of PyBo-induced C-to-T sites from (**h**) present in IVT RNA, i.e. representing concordance with mESC RNA. **j** Outline of the RNase H assay to disrupt secondary structure of in vitro-transcribed RNAs. **k** Quantification of C-to-T conversions from Sanger-seq results as outlined in (**j**). $n = 2$ experimental replicates; $n = 1$ for RNA only sample. **l** Sanger-seq chromatograms of PyBo-treated in vitro-transcribed RNA containing the **C**UMC site of *Cc2d1a* and flanking regions of different length. The minimum free energy of structures predicted by RNAFold is indicated. Source data are provided as a Source Data file for panels **c**, **f**, **g**, and **k**.

conversion sites (> 30%) exhibited similar C-to-T levels in cellular and IVT RNA (Supplementary Fig. 5g). These findings corroborate that C-to-T sites in mESC RNA originate mostly from unmodified but hyper-reactive cytidines.

To confirm that RNA folding is essential for the C-to-T conversion, we disturbed the RNA structure of *Cc2d1a* and *Inppl1* RNA fragments prior to PyBo treatment. We induced site-directed RNA cleavage by targeting RNase H eight nucleotides downstream (*Cc2d1a*) or upstream (*Inppl1*) from the **C**UMC sites by single-stranded DNA oligonucleotides (Fig. 5j). RNA cleavage greatly reduced PyBo-induced C-to-T signals (Fig. 5k; Supplementary Fig. 5h). Additionally, we synthesized shorter *Cc2d1a* RNA fragments of 150- and 70-nucleotide length with the **C**UMC site at the center. In comparison to the 232mer RNA oligonucleotide, the truncated RNA oligonucleotides show greatly reduced C-to-T conversions upon PyBo treatment, supporting the importance of structural elements that stabilize the RNA stem-loop structure (Fig. 5l). Overall, the results indicate that most high C-to-T conversion sites detected by standard PyBo-seq in this and previous studies correspond to unmodified, hyper-reactive cytidines.

Discussion

While the role of TET2 and hm5C RNA modification has taken center stage lately, the occurrence and distribution of f5C in the transcriptome as well as an involvement of TET2 therein remains poorly understood. Our study i) highlights that in multiple mammalian cell lines, f5C is essentially confined to *mt-tRNAMet*, ii) indicates the non-involvement of TET enzymes in f5C formation in vivo, iii) reveals PyBo-seq artifacts, and to overcome them iv) introduces FIBo-seq as a robust tool for base resolution analysis of f5C.

As a two-dimensional method, FIBo-seq utilizes antibody-based enrichment of f5C-containing RNA and PyBo-induced C-to-T conversions at f5C modifications for more accurate f5C detection. Such enhancement is crucial for studying modifications that occur at low stoichiometry or in lowly expressed transcripts and reduces false positives stemming from PyBo reactivity with unmodified C and ac4C. We conclude that previous PyBo-seq and PiBo-seq studies overestimated the presence of f5C in various RNAs. Non-specificity of reagents for detection of RNA modification was previously reported[39] and hence thorough method validation is essential to avoid misinterpretation of epitranscriptomic data[2,40,41]. As shown here, combining antibody enrichment with chemical sequencing can reduce false positives and may be extended to other RNA modifications where similar issues with specificity and cross-reactivity exist.

Using FIBo-seq, we provide a comprehensive analysis of f5C distribution in the transcriptome of mESCs. The results indicate that in mESCs, the occurrence of f5C is essentially confined to *mt-tRNAMet*. Hybrid-capture LC-MS/MS across mouse and human cell lines likewise identified *mt-tRNAMet* as the dominant f5C source; any additional sites, if present, are vanishingly rare and likely not physiologically relevant.

We confirm that the f5C writer for *mt-tRNAMet* is ALKBH1[7,9], with f5C levels falling below detectability in *Alkbh1* KO cells but showing no reduction in *Tet* mutants. In DNA, TET enzymes readily oxidize 5mdC to 5hmdC, 5fdC, and 5cadC. In RNA, the major product of TET-mediated in vitro oxidation is hm5C in most experimental setups. While a recent study reports considerable f5C formation under specific conditions[18], others detect little to no f5C[16,19,24,42], suggesting that TETs oxidize hm5C to f5C inefficiently in vitro. Therefore, a kinetic barrier for TETs to oxidize hm5C to f5C may account for their lack of a physiological role as f5C writers in RNA compared to DNA.

Another unexpected finding was the pronounced reactivity of PyBo with certain unmodified Cs, which compromises the interpretation of previous PyBo-seq data. The concordance of C-to-T sites between modification-free and cellular RNA corroborates the presence of hyper-reactive unmodified cytidines, not enzyme-driven modifications. Specifically, cytidines that reside in a **C**UMC motif undergo C-to-T conversion when they are in predicted RNA loop regions. Consistently, reanalysis of previous PyBo-seq data suggests that most of the previously called f5C from yeast are false-positive **C**UMC sites. On a positive note, their PyBo hyper-reactivity indicates that exposed **C**UMC sites may harbor cytidines with unusual properties, such as protonation at near neutral pH (i.e. pKa-shifted cytidines), which play a critical role in riboswitches and catalytic RNAs such as ribozymes and self-splicing RNAs[43–46]. Hence, PyBo-seq may be a technique to identify RNAs with hyper-reactive cytidines and characterize their potential function, while FIBo-seq may be used to sequence f5C in other organisms.

Methods

Cell culture

mESCs (line V6.5) CTRL clones[31,47], *Tet1/2* DKO clone[47], *Tet1/2/3* TKO clones[31] and *Alkbh1* CTRL and KO clones (this study) were cultured on tissue culture plates coated with 0.1% EmbryoMax Gelatin Solution (Sigma-Aldrich) in 2i medium (50% Neurobasal and 50% DMEM/F-12 medium (Gibco), supplemented with 1x N2 (Gibco), 1x B27 (Gibco), 2 mM L-Glutamine (Gibco), 1000 U/ml Leukemia inhibitory factor (Millipore), 100 U/ml Pen-Strep (Gibco), 1 μM PD0325901 (Sigma-Aldrich), 3 μM CHIR99021 (Sigma-Aldrich) and 50 μg/ml BSA (Sigma-Aldrich)) at 37 °C in 5% CO_2 and 20% O_2. NIH/3T3 cells (CRL-1658, ATCC), HeLa cells (CCL-2, ATCC), C2C12 (CRL-1772, ATCC) and HEK293T cells (CRL-11268, ATCC) were cultured in DMEM (Gibco) supplemented with 10% FBSGold (PAA), 2 mM L-Glutamine, and 100 U/ml Pen-Strep at 37 °C in 5% CO_2 and 20% O_2. All cell lines were tested negative for mycoplasma contamination.

Generation of *Alkbh1* knockout cells

Homozygous *Alkbh1* knockout cell lines from mESC CTRL cells[31] were generated by CRISPR-Cas9 gene editing following a previously published protocol[48]. Two pSpCas9(BB)−2A-GFP (PX458) plasmids (48138,

Addgene) with single-guide RNAs (sgRNA) flanking the catalytic site of *Alkbh1* (sequence information Supplementary Data 9) were transfected in equimolar ratio using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. For generation of *Alkbh1* CTRL cells, pSpCas9(BB)−2A-GFP with non-targeting sgRNA was transfected. Two days post-transfection, GFP-positive cells were sorted as single cells into 96-well plates using the Bigfoot spectral cell sorter (Invitrogen) and clones were expanded. Knockout was confirmed by genomic PCR and western blotting. Three independent *Alkbh1* KO clones and three independent CTRL clones were used for further experiments.

Genotyping

Genomic DNA from mESCs was purified using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacture's protocol for cultured cells. The CRISPR/Cas9 targeted genetic locus was amplified using gene-specific primers (Supplementary Data 9) and Q5 High-Fidelity DNA polymerase (NEB) following the manufacturer's protocol. Amplicons were visualized on a 1% agarose gel, purified using the QIAquick PCR purification kit (Qiagen) and sequenced by the Sanger sequencing service at StarSEQ GmbH, Germany.

Western blotting

mESCs were lysed with RIPA buffer supplemented with protease inhibitors (Roche) for 10 min on ice. Protein lysates were cleared by centrifugation and quantified using a BCA assay (Sigma-Aldrich). 25 μg protein lysate was denatured in 4x NuPAGE LDS sample buffer (Invitrogen). Samples were run on 4–15% Criterion TGX Precast Midi Protein Gel (Bio-Rad) and blotted on a Trans-Blot Turbo Midi 0.2 μm PVDF membrane (Bio-Rad) using the Trans-Blot Turbo Transfer System (Bio-Rad). Membrane was blocked with 5% milk in TBS-T buffer and incubated with anti-ALKBH1 antibody (1:1,000; ab195376, clone EPR6176, Abcam) or anti-Vinculin antibody (1:10,000; MCA465GA, clone V284, Bio-Rad) in 5% milk in TBS-T buffer at 4 °C overnight. Membranes were washed and incubated with goat anti-rabbit IgG-HRP (1:5,000, 111-035-144, Dianova) or goat anti-mouse IgG-HRP (1:5,000, 115-035-146, Dianova) at room temperature for 2 h. The membrane was incubated with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Scientific) and imaged with the ChemiDoc system (Bio-Rad).

Nat10 knockdown

CTRL mESCs were transiently transfected with mouse *Nat10* siRNAs (siGENOME Mouse Nat10 siRNA SMARTPool, M-054588-01-0005, Dharmacon) or control siRNA (siGENOME non-targeting siRNA pool #2, D-001206-14-05, Dharmacon) using Lipofectamine RNAiMAX (Invitrogen) and Opti-MEM (Gibco) according to the manufacturer's protocol. The final concentration of siRNA used was 60 nM and the incubation time on cells was 24 h. Transfection were done in triplicates. RNA from cells was purified and used for RT-qPCR, ac4C quantification by LC-MS/MS and PyBo treatment followed by Sanger sequencing as described below.

RNA isolation and poly(A) selection

Total RNA was isolated from cultured cells using QIAzol Lysis Reagent (Qiagen) according to the manufacturer's protocol followed by TURBO DNase (Invitrogen) treatment and another round of QIAzol purification. RNA was precipitated with ammonium acetate and ethanol and suspended in nuclease-free H2O. RNA integrity was assessed in an agarose gel and/or in an RNA 6000 Nano chip using a 2100 Bioanalyzer (Agilent Technologies). Poly(A)+ RNA was purified from total RNA using Oligo d(T)25 Magnetic Beads (Thermo Fisher Scientific), following the manufacturer's protocol, with three consecutive rounds of selection. The purity of poly(A)+ RNA was verified using a 2100 Bioanalyzer (Agilent Technologies).

In vitro transcription of RNA oligonucleotides

RNA oligonucleotides were in vitro transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB) with canonical CTP or modified CTP as specified in Supplementary Data 9. A PCR amplified DNA oligonucleotide with T7 promoter served as DNA template for RNA oligonucleotides with one (modified) cytidine in the sequence (thereafter called: 1x cytidine RNA oligonucleotides) and the Cc2d1a 70mer. To generate DNA templates for in vitro transcription of *Cc2d1a*-, *Inppl1*- and *Elf2*-fragments, PCR amplicons containing the designated C-to-T site were ligated into the pJET1.2 vector using the CloneJet PCR cloning kit (Thermo Fisher Scientific) following the manufacturer's instructions and plasmids were linearized with XbaI (NEB). To obtain a DNA template for the Cc2d1a 150mer RNA synthesis, cDNA was amplified with a primer containing the T7 promoter sequence. Primers for template generation are listed in Supplementary Data 9. Control IVT RNA oligonucleotides for LC-MS/MS analysis were transcribed from FLuc control template DNA (NEB), pTRI-Xef (Invitrogen) and pJET1.2-hLMNA (gift from Deepa Jayaprakashappa). Synthesized RNA oligonucleotides were purified using DNase I (NEB) and the RNA Clean & Concentrator kit (Zymo Research) following the manufacturer's protocol. RNA integrity was assessed in an agarose gel and/or in an RNA 6000 Nano chip using a 2100 Bioanalyzer (Agilent Technologies).

Pyridine borane treatment

50 ng–1 μg RNA was incubated in 900 mM pyridine borane in 600 mM sodium acetate (pH 5.2) at 37 °C for 24 h as previously described[12]. For mock-treated samples, water was added instead of pyridine borane. Thereafter, RNA was purified by ethanol precipitation or with the RNA Clean & Concentrator kit (Zymo research) and eluted in RNase-free water.

Malononitrile treatment

50 ng–1 μg RNA was incubated in 150 mM malononitrile in 1x TE buffer (pH 7.4) at 37 °C for 24 h as previously described[29]. For mock-treated samples, water was added instead of malononitrile. Thereafter, RNA was purified by ethanol precipitation or with the RNA Clean & Concentrator kit (Zymo Research) and eluted in RNase-free water.

C-to-T quantification by Sanger sequencing

Purified, chemically-treated or mock-treated RNA was reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen) and random primers following the manufacturer's protocol. For 1x cytidine RNA oligonucleotides a gene-specific RT-primer (Supplementary Data 9) was used. cDNA was PCR amplified with gene specific primers (Supplementary Data 9) and Q5 High-Fidelity DNA polymerase (NEB) following the manufacturer's protocol. For *mt-tRNAMet* and 1x cytidine oligonucleotide cDNA amplifications, the forward primer contained a 5′ overhang sequence (Supplementary Data 9) for Sanger sequencing. Amplicons were purified using the QIAquick PCR purification kit (Qiagen) and sequenced by the Sanger sequencing service at StarSEQ GmbH, Germany (sequencing primer information Supplementary Data 9). Raw signal values for C and T at the designated position were extracted from.ab1 files and the percentage of C-to-T conversion was calculated with the formula: T/(C + T)*100.

RT-qPCR

cDNA from RNA was synthesized using SuperScript II Reverse Transcriptase (Invitrogen) and random primers, following the manufacturer's protocol. qPCR was performed in technical duplicates using SYBR Green I Master (Roche) or LightCycler 480 Probes Master (Roche) with hydrolysis probes (Universal ProbeLibrary technology, Roche) on a LightCycler 480 II instrument (Roche). Primer sequences and hydrolysis probe numbers are listed in Supplementary Data 9.

PyBo-seq and Mal-seq of rRNA-depleted RNA

Ribosomal RNA was depleted from triplicates of DNase-treated, total RNA using the NEBNext rRNA Depletion Kit v2 (Human/Mouse/Rat) (E7400, NEB) following the manufacturer's protocol and purified using the RNA Clean & Concentrator kit (Zymo Research). Thereafter, 200 ng of RNA was treated with PyBo, Mal or mock as described above. RNA was precipitated in ammonium acetate and ethanol and suspended in nuclease-free H2O. RNA library preparation for next-generation sequencing was performed using the TruSeq Stranded mRNA Library Prep kit (20020594, Illumina), entering the protocol at the RNA fragmentation step and excluding the poly(A) enrichment. 10 ng RNA was used as starting amount. Libraries were size-selected for >100 bp fragments, pooled in equimolar ratios and sequenced on an Illumina NovaSeq 6000 S4 flow cell in 100 bp paired-end mode. The same RNA samples used for NGS sequencing were also analyzed by Sanger sequencing to determine the % C-to-T conversion at f5C34 *mt-tRNAMet* as described above.

PyBo-seq of ribosomal RNA

DNase-treated, total RNA from one *Tet1/2/3* CTRL and one TKO clone was chemically fragmented to 120-180 nt-long fragments using the RNA fragmentation reagents (Invitrogen) and PyBo treated as described above. Thereafter, RNA was precipitated with ammonium acetate and ethanol and suspended in nuclease-free H2O. RNA was T4 PNK (NEB) treated before library construction with the NEXTFLEX Small RNA-Seq Kit V3 (Bioo Scientific) following Step A to Step G of Bioo Scientific's standard protocol (V19.01) using the NEXTFLEX 3' SR Adaptor and 5' SR Adaptor. Step A (NEXTFLEX 3' 4 N Adenylated Adapter Ligation) was performed overnight at 20 °C and for step F (Bead Cleanup) the modified protocol was used to retain small RNAs. Libraries were prepared with a starting amount of 1.4 ng RNA and amplified in 21 PCR cycles. Libraries were size-selected for 18 – 362 nt insert size in an 8% TBE gel, profiled in a High Sensitivity DNA Chip on a 2100 Bioanalyzer (Agilent Technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). Samples were pooled in equimolar ratio and sequenced on one NextSeq 500 midoutput flow cell, paired-end for 2×80 cycles plus 6 cycles for the index read.

PyBo- and Mal-seq data processing and analysis

Quality assessment and alignment. Raw read data were quality-assessed by FastQC (v.0.11.9). Alignment to the reference genome was performed using STAR (v. 2.7.10a) allowing soft-clipping with non-default parameters: --outFilterScoreMinOverLread 0.33, --outFilterMatchNminOverLread 0.33 and --outFilterMismatchNoverReadLmax 0.02, that ensure mapping of overlapping reads and allowed for up to 2% of mismatches. The reference (GRCm39.dna_sm.primary_assembly.fa) was obtained from Ensembl release 107 and the genome index was generated using M.mus.GRCm39.107.gtf for annotation and --sjdbOverhang 99 (read length − 1). The % reads mapped to rRNA was estimated by mapping reads to the 45S pre-rRNA reference (NR_046233.2) using BWA-mem (v 0.7.17) with default parameters. To check RNA-seq quality, read counts were estimated and assigned to genomic features by R/Bioconductor package Rsubread (v. 2.12.3) using paired-end mode, taking into account rev-strand specificity of the library and counting only fragments longer than 25 bp (isPairedEnd = TRUE, strand_specificity = 2, minFragLength = 25). R/Bioconductor package biomaRt (v. 2.56.1) and mmusculus_gene_ensembl database (v.107) were used for annotation. Multi-mapping reads were filtered out using samtools (v.1.10) with −q 10 parameter (MAPQ > 10).

C-to-T calling and annotation. C-to-T events were identified using JACUSA2 (v.2.0.4) applying 2 conditions mode (call-2, PyBo/Mal vs mock), scanning only positions with coverage >5 reads (-c 5), minimum base quality >20 (-q 20) and mapping quality >20 (-m 20), taking into account rev-strand specificity of the libraries (P RF-FIRSTSTRAND) and using bam files with only uniquely mapped reads as an input. To identify PyBo/Mal-induced C-to-T events for *Tet1/2/3* or *Alkbh1* CTRL samples, we applied the following filtering criteria for raw JACUSA2 output files: (1) JACUSA2 score >2, (2) C-to-T conversion level >5% in each of 3 replicates of PyBo/Mal-treated sample, (3) ≥ 5 reads (average of 3 replicates) supporting mismatched T in PyBo/Mal-treated samples, (4) C-to-T conversion level <2% in each of 3 replicates of mock-treated samples, (5) and <5 reads (average of 3 replicates) supporting mismatched T in mock-treated samples. The C-to-T events that passed filtering were annotated with hiAnnotator (v. 1.34.0) using Mus_musculus.GRCm39.107.gtf as a gene model. In cases when the C-to-T site overlapped with multiple transcripts, we employed the following strategy to clarify the transcript annotation. Firstly, the gene with the correct orientation for the C-to-T site and the highest expression level was selected. Secondly, the transcript variant was selected with the following prioritization: (1) highest expression (2) marked as canonical and (3) the longest transcript variant. The transcript expression level was determined using kallisto (v. 0.44.0).

Retrieving reads belonging to mature/immature mt-tRNAMet. To retrieve fragments corresponding to mature *mt-tRNAMet*, we extracted reads with start and end within the *mt-tRNAMet* locus borders. To do that, reads that were mapped within *mt-tRNAMet* locus and satisfy the following criteria: insert width <70, abs(isize) >20, abs(isize) <70, start > MT:3845, end <MT:3913 were retrieved using R/Bioconductor package GenomicAlignments (v. 1.36.0), and FilterSamReads function from Picard tool (v. 2.20.3). Reads that mapped to the *mt-tRNAMet* locus and overlapped with flanking genomic regions were considered to represent immature *mt-tRNAMet*.

Misincorporation analysis. To investigate all types of substitution deviating from the reference genome (GRCm39), JACUSA2 (v.2.0.4) was run in single-condition mode (only PyBo-, or mock-treated samples) with the following parameters: call-1 -c 50 -q 35 -m 30 -P RF-FIRSTSTRAND. Only positions having at least 5% of reads supporting substitution were considered.

Differential C-to-T sites analysis and MA-plot generation. DESeq2 (v. 1.40.2) was utilized to test transcriptome-wide differences in C-to-T conversions between *Tet1/2/3* CTRL and *Tet1/2/3* TKO mESC and between *Alkbh1* CTRL and *Alkbh1* KO mESC. The following parameters were applied for statistical testing: test = "Wald", alpha = 0.01. The MA plots were generated using the R package ggplot2 (v. 3.5.1) and the log2(x + 0.01) transformed values.

Motif analysis and structure prediction. Only sequences flanking C-to-T sites shared across six PyBo-treated samples from *Alkbh1* and *Tet1/2/3* CTRLs were analyzed. The sequences of CDS, 5'UTR, and 3'UTR were extracted using biomaRt (v. 2.56.1) and mmusculus_gene_ensembl database (v.107). The motifs logos were created using ggseqlogo (v. 0.1) and ggplot2 (v. 3.5.1). The secondary structures were predicted using RNAFold from ViennaRNA Package (v. 2.6.2) via LncFinder (v. 1.1.5) in R/RStudio environment.

FIBo-seq

f5C-RNA-Immunoprecipitation (f5C-RIP). DNase-treated, total RNA was chemically fragmented to 120-180 nt-long fragments using the RNA Fragmentation Reagents (Invitrogen). f5C-RIP was performed in technical duplicates with 2 × 80 μg RNA and 2% input was removed from each sample beforehand. RIP binding buffer (final concentration: 50 mM Tris HCl pH 7.5, 100 mM NaCl, 0.05% (v/v) NP-40), RNase inhibitor, and anti-f5C antibody (1 μg anti f5C antibody/5 μg RNA; 61227, Lot no. 34419003, Active Motif) were added to f5C-RIP samples and incubated at 4 °C for 16 h rotating. Thereafter, Dynabeads Protein

A (Invitrogen) were washed three times with 0.1% BSA in PBS, once in RIP binding buffer, and incubated with f5C-RIP samples at 4 °C for 2 h with rotation. Samples were washed three times with wash buffer (final concentrations: 50 mM Tris HCl pH 7.5, 250 mM NaCl, 0.1% (v/v) NP-40) and twice with RIP binding buffer. Any remaining liquid was removed from the beads, QIAzol Lysis Reagent was added and samples were vortexed well before frozen at −80 °C. RNA was eluted using the Qiazol/chloroform extraction method and suspended in nuclease-free H2O. To assess the enrichment efficiency of f5C-containing transcripts, the enrichment of *mt-tRNAMet* was determined by RT-qPCR and thereafter the duplicates were pooled into one sample.

PyBo treatment, rRNA depletion and T4 PNK treatment. f5C-RIP and input samples were PyBo treated as described above. Thereafter RNA was precipitated with ammonium acetate and ethanol and suspended in nuclease-free H2O. Ribosomal RNA was depleted from samples using the NEBNext rRNA Depletion Kit v2 (Human/Mouse/Rat) (E7400, NEB) following the manufacturer's protocol and purified using the Qiazol/chloroform extraction method. RNA was T4 PNK (NEB) treated before library construction.

Library preparation and sequencing. NGS library construction from biological triplicates of each CTRL and TKO clone was performed with NEXTFLEX Small RNA-Seq Kit V4 with UDIs (PerkinElmer) using the NEXTFLEX small RNAseq PreArrayed UDI Barcode primers. Libraries were prepared with a starting volume of 4 μl for the f5C-RIP samples (~ 0.823 ng) and a starting amount of 0.823 ng RNA for the input samples following Step A to Step F of Perkin Elmer's standard protocol (V23.04). Step A (NEXTFLEX 3′ 4 N Adenylated Adapter Ligation) was performed overnight at 16 °C and Step F (PCR amplification) with 24 PCR cycles. Thereafter, the no size selection option in the protocol was followed. Libraries were profiled in a High Sensitivity DNA Chip on a 2100 Bioanalyzer (Agilent Technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life Technologies) and pooled together in equimolar ratio. The pool was purified by running a 3% agarose gel cassette on a PippinHT (Sage Science) and size-selected for library fragments in 175–250 bp size range (adapter size + insert size). The pool was sequenced on a NextSeq 2000 P3 (100 cycles) flow cell, paired-end for 2 × 60 cycles plus 2 × 8 cycles for the dual index read.

FIBo-seq quality assessment and alignment. Raw read data were quality-assessed by FastQC (v.0.11.9) and aligned to the reference genome using STAR (v. 2.7.10a) with the following relaxed parameters to map short overlapping reads containing C-to-T mismatches: outFilterScoreMinOverLread 0.1, --outFilterMatchNminOverLread 0.1 and --outFilterMismatchNoverReadLmax 0.05. The genome index was generated with --sjdbOverhang 50 parameter. The reads were also mapped to *18S rRNA* (NR_003278.3) using BWA with default parameters. deepTools (v. 3.5.1) was used to generate bigWig tracks.

Peak calling and C-to-T sites matching. exomePeak2 (v. 1.10.0) was utilized for peak calling from input and RIP samples. The following parameters were applied to scan the gene coding part of the genome (mode = "exon") using DESeq2 for statistical testing (test_method = "DESeq2", p_cutoff = 0.001, bin_size = 25, step_size = 25). Unstranded library type was taken into account (strandness = "unstrand"). Fragment length parameter was estimated as the average insert size for all samples. The peaks were filtered using the following criteria: FDR < 0.01, RPM.input >10, RPM.IP > 50, log2FC > 3.32. Peak coordinates were used to count the number of reads belonging to each peak using bamCount function from the R/Bioconductor bamsignals package (v. 1.30.0). The peaks were scanned for the presence of PyBo-induced C-to-T conversions using the pileup function from R/Bioconductor package Rsamtools (v. 2.16.0) with non-default parameters to filter low-quality bases as well as low-quality mapped reads (min_base_quality = 30, min_mapq = 10). The C-to-T site with maximum conversion level within peak boundaries was defined as "max", and the sum of all conversion levels of C-to-T sites within peak boundaries was defined as "cumulative". The average C-to-T levels were reported as non-zero values only in cases when the mean(5 replicates of PyBo) > 2 x Standard Deviation(5 replicates of PyBo) and mean(2 replicates of mock) <1% C-to-T, and mean(5 replicates of PyBo) > 10 x mean(2 replicates of mock).

f5C-RIP-RT-qPCR

f5C-RIP from DNase-treated, total RNA was performed as described above. To obtain a f5C-depleted RNA sample, total RNA was treated with PyBo before immunoprecipitation. In the IgG sample, the f5C antibody was replaced by rabbit IgG (ab171870, Abcam). cDNA synthesize and RT-qPCR with SYBR Green I Master (Roche) were performed as described above. RT-qPCR data from f5C-RIP samples were normalized to input samples and presented as percentage input recovery.

LC-MS/MS analysis

RNA sample preparation. Prior to LC-MS/MS analysis, RNA samples were digested to nucleosides using a previously described procedure[49]. In brief, 0.003 U nuclease P1 (Sigma-Aldrich) and 0.01 U snake venom phosphodiesterase (Worthington Biochemical) were incubated with RNA at 37 °C for 2 h. Thereafter, 0.1 U alkaline phosphatase (Thermo Scientifc) was added to the samples and the reaction was incubated at 37 °C for 2 h. Isotopic labeled nucleosides were spiked into the digested RNA as standards before injection into the LC-MS/MS system.

Preparation of isotopic labeled nucleoside standard. Isotopic labeled nucleosides were obtained commercially ($^{15}N_3$-C (Silantes), D_3-m6A (Toronto Research Chemicals), $^{13}C_{10}$-A (Silantes) and $^{13}CD_2$-hm5C (Toronto Research Chemicals)). ^{13}CD-f5C and $^{13}C_2D_4$-f5Cm were made in-house from total RNA of HEK293T cells grown in L-Methionine-free DMEM (Gibco) supplemented with 0.2 M of $^{13}CD_3$-L-Methionine (Sigma) for 14 days. Cells were split every second day with full media exchange. RNA was extracted and digested to nucleosides as described above. The LC-MS/MS elution window of ^{13}CD-f5C and $^{13}C_2D_4$-f5Cm was identified by analytical LC-MS/MS using a Reprosil 100 C18 column (3 μm, 250 × 4.6 mm, Dr. Maisch) with 5 mM ammonium acetate pH 6.9 (solvent A) and acetonitrile (solvent B). The gradients were as follows: 100% A for first 12 min, gradual increase of B from 0 to 15% in next 24 min, constant 15% B for 1.5 min, increase of B from 15 to 60% within next 6 min and finally 100% A for last 9.5 min. Flow rates were 0.5 ml/min for first 36 min, then increased to 1 ml/min within 1.5 min, kept at 1 ml/min for next 6 min and switched back to 0.5 ml/min within 7.5 min and kept at 0.5 ml/min for last 2 min. ^{13}CD-f5C and $^{13}C_2D_4$-f5Cm where purified by preparative HPLC and their concentration was experimentally determined by LC-MS/MS using commercial f5C and f5Cm (both Carbosynth) of known concentrations. Finally, all isotopic labeled nucleotides were combined and this mixture was added to the samples before running the LC-MS/MS analysis.

Measurement of nucleosides. Detection of nucleosides from up to 2 μg RNA was performed using a LC-MS/MS system consisting of an Agilent 1290 UHPLC connected to an Agilent 6490 triple-quadrupole mass spectrometer. The chromatographic separation was performed with a ReproSil 100 C18 column (3 μm 150×4.6 mm, Dr. Maisch) maintained at 30 °C. The running solutions were 5 mM ammonium acetate pH 6.9 (solvent A) and acetonitrile (solvent B). The following gradients were used: 0% of solvent B from 0 min to 8 min, linear increase to 15% solvent B for next 16 min, hold at 15% for one minute, and then gradually increased to 60% solvent B over 4 min, and finally kept constant at 0% solvent B for 5 min. The flow rate was at

0.5 ml/min, except between 24 to 34 min, when it first increased to 1 ml/min within one minute, stayed constant at 1 ml/min for 4 min and then returned to 0.5 ml/min over 5 minutes. The MS settings as well as the monitored precursor ion to product ion *m/z* transitions are listed in Supplementary Data 10.

LC-MS/MS data analysis and quantification. The data were analyzed using Agilent MassHunter Qualitative- (version B.08.00) and Quantitative- (version B.09.00) Analysis Software (Agilent Technologies). Peak identity was confirmed with isotopic standards where applicable. For DHUm and f5Cm-mal no standards were available. They were identified by their expected molecular masses and the absence of signals in the mock-treated samples. Measurements of C, $^{15}N_3$-C, A and $^{13}C_{10}$-A were performed in 100-fold diluted samples. Absolute amounts of the modifications were calculated using the stable isotope dilution technique, which is described in detail elsewhere[50].

mt-tRNAMet hybrid capture

Hybrid capture of *mt-tRNAMet* was performed with 55 μg DNase-treated, total RNA from mESCs, HeLa or HEK293T cells and 10 μg DNase-treated, total RNA from NIH/3T3 and C2C12 cells. 10 μg of total RNA was saved as input. The hybridization step was done with 200 pmol biotinylated antisense DNA probes (Supplementary Data 9) in 1x SSC buffer by incubation at 95 °C for five minutes and cooling to 22 °C with −0.1 °C/sec temperature ramp. Dynabeads MyOne Streptavidin T1 (Invitrogen) were washed three times in wash buffer (final concentration: 5 mM Tris HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl) once in 5x SSC buffer, and incubated with RNA:DNA hybrid samples in the presence of RNase inhibitor at room temperature for 30 min with rotation. Thereafter, the supernatant (unbound fraction) was removed and saved for later use. The beads were washed three times each with 3x SSC, 1x SSC, and 0.1x SSC for 2 min, rotating. Any remaining liquid was removed and beads were suspended in RNase-free water. RNA was eluted with TURBO DNase (Invitrogen) at room temperature for 45 min. For RNA purification, TRIzol LS Reagent (Invitrogen) and chloroform were added to all samples and the upper aqueous phase was further purified with RNA Clean & Concentrator kit (Zymo research) following the manufacturer's protocol. One aliquot of RNA was analyzed by LC-MS/MS with isotopic labeled standards as described above. The other aliquot was used for cDNA synthesis and RT-qPCR with SYBR Green I Master (Roche) as described above. RT-qPCR data were normalized to *18S rRNA* and presented relative to input samples. In HeLa cells, the signal for *mt-tRNAMet* in the unbound fraction was not reduced after a first round of hybrid capture. For that reason, a second round of hybrid capture was performed.

PyBo-seq of modification-free transcriptome

IVT RNA generation and library preparation. 400 ng poly(A)+ RNA isolated from mESCs was incubated with 1.67 μM oligo(dT)-VN RT primer and 1.67 mM dNTPs for 5 min at 75 °C. For RT reaction, template switching RT enzyme mix and buffer (NEB) and 3.75 μM template switching oligo (Supplementary Data 9) was added and incubated for 90 min at 42 °C and 5 min at 85 °C. 2nd strand cDNA synthesis was performed with Q5 Hot Start High-Fidelity DNA Polymerase (NEB), RNase H and 0.2 μM T7 extension primer (Supplementary Data 9) for 15 min at 37 °C, 1 min at 95 °C and 15 min at 65 °C. Double-stranded DNA was purified with HighPrep PCR beads (0.8x beads, MagBio) and used for in vitro transcription with canonical NTPs as described above. RNA was size selected with HighPrep RNA Elite beads (0.6x beads, MagBio), PyBo treated and purified as described above. RNA library preparation for next-generation sequencing was performed using the TruSeq Stranded mRNA Library Prep kit (20020594, Illumina) and 20 ng RNA as described above and sequenced on an Illumina NovaSeq X Plus flow cell in 100 bp paired-end mode.

Data analysis of PyBo-seq on the IVT transcriptome. Quality assessment, alignment to the reference genome and misincorporation analysis were performed as described for PyBo-seq on the cellular transcriptome. Rsamtools (v2.16.0) was used to quantify how many C-to-T sites identified in cellular RNA were detected in the modification-free RNA, using the parameters min_base_quality = 30 and min_mapq = 10. C-to-T sites in cellular RNA were identified in CTRL mESC samples as described above, except that only sites within the CDS, 3′UTR, or 5′ UTR of annotated protein-coding transcripts and shared across all six PyBo-treated CTRL samples were considered. For IVT RNA, C-to-T sites were retained if they showed (1) a C-to-T signal greater than 0% in PyBo-treated samples, (2) a C-to-T signal below 2% in mock-treated samples and (3) at least a 5-fold higher C-to-T signal in PyBo-treated samples compared to mock. The average C-to-T signal was calculated as mean of two PyBo-treated replicates. If one replicate had insufficient coverage (< 5 reads), the value from the replicate with adequate coverage was reported.

RNase H treatment

Hybridization of in vitro-transcribed RNA oligonucleotides was done by incubating 3 μg RNA with 20-fold molar excess of DNA probes (Supplementary Data 9) in hybridization buffer (final concentration: 10 mM Tris HCl pH 7.5, 50 mM NaCl, 1 mM EDTA) at 95 °C for two minutes and cooling to 22 °C with −0.1 °C/s temperature ramp. The DNA probe for *Cc2d1a* is complementary to the 16-nt region directly downstream of the C-to-T site. The DNA probe for *Inppl1* is complementary to the 16-nt region directly upstream of the C-to-T site. RNA in the RNA:DNA duplex was cleaved with RNase H (NEB) following the manufacturer's protocol. Samples were treated with TURBO DNase (Invitrogen), purified with the RNA Clean & Concentrator kit (Zymo Research) and eluted in RNase-free water. Products were visualized in a RNA 6000 Nano chip using a 2100 Bioanalyzer (Agilent Technologies). RNA was PyBo treated and the C-to-T conversion was quantified by Sanger sequencing as described above. Primers contained a 5′ overhang (Supplementary Data 9) for adequate amplification and sequencing.

Processing of publicly available data

f5C-seq data for poly(A)+ RNA of *S. cerevisiae* was retrieved from GEO:GSE133138 (S.cer, NCBI SRA accessions SRR12879649-SRR1287952) and reanalyzed similarly as described above for PyBo-seq. Saccharomyces_cerevisiae.R64-1-1.dna.toplevel.fa genome was used as a reference. The detected C-to-T sites were used to identify sequence motifs as described above.

PyBo-assisted RNA-seq datasets for HEK293, HeLa, MCF7, and Hep2G cell lines were retrieved from NCBI BioProject PRJNA550080. Sample details and treatment conditions are provided in Supplementary Data 11. Data processing and C-to-T detection followed the same workflow as described for PyBo-seq. To extract motif information from C-to-T sites in human samples, we first pre-selected sites based on the following criteria: (1) an average C-to-T conversion greater than 0% from both PyBo-treated replicates of each cell line and (2) an average C-to-T conversion below 2% in mock-treated samples in at least three out of four cell lines. Sites with an average conversion >25% were manually inspected in IGV and subjected to further filtering: (1) presence in all PyBo-treated samples across all four cell lines, (2) localization within CDS, 5′UTR or 3′UTR regions, and (3) exclusion of sites mapping to pseudogene loci, as these sequences occur at multiple genomic locations.

Pico-borane-assisted RNA-seq (PiBo-seq) data for mESC was obtained from GEO:GSE156933 (NCBI SRA accessions SRR12524852-SRR12524857) and reanalyzed by mapping reads to tRNAs (mm39-tRNAs.fa) and *18S rRNA* (NR_003278.3) references using BWA with default parameters. To profile *18S rRNA* for C-to-T events after PiBo-conversion, the pileup function from R/Bioconductor package Rsamtools (v. 2.16.0) with non-default parameters to filter low-quality bases

as well as low-quality mapped reads (min_base_quality = 30, min_mapq = 10) was used.

NSUN2/5/6 motif enrichment analysis and NSUN6 CLIP comparison

Previously reported sequence motifs of NSUN2, NSUN5 and NSUN6 in mESCs were used for the enrichment analysis: **C**AGG and **C**KGGG (K = G or U) for NSUN2; **C**ARAU (R = G or A) for NSUN5 and **C**UCCA for NSUN6[18]. Expected motif frequencies were calculated based on their occurrence in mRNAs (CDS and UTR regions), considering only mRNAs that contained at least one PyBo-induced C-to-T site. To assess enrichment significance, a two-sided binom.test (x = number of C-to-T sites in the motif, n = total number of C-to-T sites in the analysis, p = expected probability) was applied.

To compare NSUN6 binding sites with PyBo-induced C-to-T sites, the NSUN6 CLIP dataset reported in Lu et al., 2024 (GSE242724) was used. Specifically, the signal data provided in the bigWig tracks was compared with the positions of PyBo-induced C-to-T sites.

Comparison of PyBo-induced C-to-T sites with reported m5C sites

To compare the positions of C-to-T signal from PyBo-seq with previously reported m5C modifications in the mESC transcriptome, we used single-nucleotide resolution datasets from Lu et al, 2024[18] (Table S5 of that study) and Amort et al, 2017[51] (Table S3 of that study). The genomic coordinates of m5C sites were converted from mm10 to mm39, using the leftOver tool with mm10ToMm39.over.chain.gz as chain file.

General

IGV genomic browser (v. 2.17.4) was used to visualize NGS data; GraphPad Prism (v. 10.2.3) and ggplot2 (v. 3.5.1) were utilized for visualization; R (v. 4.3.1) and RStudio (v. 2024.04.01) were used for running custom scripts; tidyverse (v. 2.0.0), reshape2 (v. 1.4.4) were used for data processing. ChatGPT was used for language editing purposes.

Reporting summary

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

Data availability

The data supporting the findings of this study are available from the corresponding authors upon request. The NGS data generated in this study have been deposited in GEO under accession number GSE288507. Accession numbers of publicly available datasets used in this study are provided in Supplementary Data 11. Source data for the figures and Supplementary Figs. are provided as a Source Data file.

References

1. Linder, B. & Jaffrey, S. R. Discovering and Mapping the Modified Nucleotides That Comprise the Epitranscriptome of mRNA. *Cold Spring Harb Perspect Biol.* **11**, https://doi.org/10.1101/cshperspect.a032201 (2019).
2. Wiener, D. & Schwartz, S. The epitranscriptome beyond m(6)A. *Nat. Rev. Genet* **22**, 119–131 (2021).
3. Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Biol.* **22**, 375–392 (2021).
4. Arguello, A. E. et al. Reactivity-dependent profiling of RNA 5-methylcytidine dioxygenases. *Nat. Commun.* **13**, 4176 (2022).
5. Moriya, J. et al. A novel modified nucleoside found at the first position of the anticodon of methionine tRNA from bovine liver mitochondria. *Biochemistry* **33**, 2234–2239 (1994).
6. Takemoto, C. et al. Unconventional decoding of the AUA codon as methionine by mitochondrial tRNAMet with the anticodon f5CAU as revealed with a mitochondrial in vitro translation system. *Nucleic Acids Res.* **37**, 1616–1627 (2009).
7. Haag, S. et al. NSUN3 and ABH1 modify the wobble position of mt-tRNAMet to expand codon recognition in mitochondrial translation. *EMBO J.* **35**, 2104–2119 (2016).
8. Bohnsack, M. T. & Sloan, K. E. The mitochondrial epitranscriptome: the roles of RNA modifications in mitochondrial translation and human disease. *Cell Mol. Life Sci.* **75**, 241–260 (2018).
9. Kawarada, L. et al. ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Res.* **45**, 7401–7415 (2017).
10. Huang, W. et al. Formation and determination of the oxidation products of 5-methylcytosine in RNA. *Chem. Sci.* **7**, 5495–5502 (2016).
11. Tardu, M., Jones, J. D., Kennedy, R. T., Lin, Q. & Koutmou, K. S. Identification and Quantification of Modified Nucleosides in Saccharomyces cerevisiae mRNAs. *ACS Chem. Biol.* **14**, 1403–1409 (2019).
12. Wang, Y. et al. Single-Base Resolution Mapping Reveals Distinct 5-Formylcytidine in Saccharomyces cerevisiae mRNAs. *ACS Chem. Biol.* **17**, 77–84 (2022).
13. Zhang, H. Y., Xiong, J., Qi, B. L., Feng, Y. Q. & Yuan, B. F. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. *Chem. Commun. (Camb.)* **52**, 737–740 (2016).
14. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
15. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
16. DeNizio, J. E., Liu, M. Y., Leddin, E. M., Cisneros, G. A. & Kohli, R. M. Selectivity and Promiscuity in TET-Mediated Oxidation of 5-Methylcytosine in DNA and RNA. *Biochemistry* **58**, 411–421 (2019).
17. He, C. et al. TET2 chemically modifies tRNAs and regulates tRNA fragment levels. *Nat. Struct. Mol. Biol.* **28**, 62–70 (2021).
18. Lu, L. et al. Base-resolution m(5)C profiling across the mammalian transcriptome by bisulfite-free enzyme-assisted chemical labeling approach. *Mol. Cell* **84**, 2984–3000.e2988 (2024).
19. Shen, Q. et al. Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature* **554**, 123–127 (2018).
20. Guallar, D. et al. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat. Genet.* **50**, 443–451 (2018).
21. Li, Y. et al. TET2-mediated mRNA demethylation regulates leukemia stem cell homing and self-renewal. *Cell Stem Cell* **30**, 1072–1090 e1010 (2023).
22. Zou, Z. et al. RNA m(5)C oxidation by TET2 regulates chromatin state and leukaemogenesis. *Nature* **634**, 986–994 (2024).
23. Lan, J. et al. Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. *Nat. Commun.* **11**, 4956 (2020).
24. Shen, H. et al. TET-mediated 5-methylcytosine oxidation in tRNA promotes translation. *J. Biol. Chem.* **296**, 100087 (2021).
25. Delatte, B. et al. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**, 282–285 (2016).
26. Singh, B. N. et al. Tet-dependent 5-hydroxymethyl-Cytosine modification of mRNA regulates axon guidance genes in Drosophila. *PLoS One* **19**, e0293894 (2024).
27. Legrand, C. et al. Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* **27**, 1589–1596 (2017).
28. Jones, J. D. et al. Methylated guanosine and uridine modifications in S. cerevisiae mRNAs modulate translation elongation. *RSC Chem. Biol.* **4**, 363–378 (2023).

29. Li, A., Sun, X., Arguello, A. E. & Kleiner, R. E. Chemical Method to Sequence 5-Formylcytosine on RNA. *ACS Chem. Biol.* **17**, 503–508 (2022).
30. Lyu, R. et al. A Quantitative Sequencing Method for 5-Formylcytosine in RNA. *Isr. J. Chem.* **64**, e202300111 (2023).
31. Dawlaty, M. M. et al. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* **29**, 102–111 (2014).
32. Rackham, O. & Filipovska, A. Organization and expression of the mammalian mitochondrial genome. *Nat. Rev. Genet.* **23**, 606–623 (2022).
33. Trixl, L. et al. RNA cytosine methyltransferase Nsun3 regulates embryonic stem cell differentiation by promoting mitochondrial activity. *Cell Mol. Life Sci.* **75**, 1483–1497 (2018).
34. Schiffers, S. & Oberdoerffer, S. ac4C: a fragile modification with stabilizing functions in RNA metabolism. *RNA* **30**, 583–594 (2024).
35. Taoka, M. et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **46**, 9289–9298 (2018).
36. Chen, L. et al. NAT10-mediated N4-acetylcytidine modification is required for meiosis entry and progression in male germ cells. *Nucleic Acids Res.* **50**, 10896–10913 (2022).
37. Xiong, J. et al. Heavy Metals Induce Decline of Derivatives of 5-Methycytosine in Both DNA and RNA of Stem Cells. *ACS Chem. Biol.* **12**, 1636–1643 (2017).
38. Van Haute, L. et al. NSUN2 introduces 5-methylcytosines in mammalian mitochondrial tRNAs. *Nucleic Acids Res.* **47**, 8720–8733 (2019).
39. Grozhik, A. V. et al. Antibody cross-reactivity accounts for widespread appearance of m(1)A in 5'UTRs. *Nat. Commun.* **10**, 5126 (2019).
40. Kong, Y., Mead, E. A. & Fang, G. Navigating the pitfalls of mapping DNA and RNA modifications. *Nat. Rev. Genet* **24**, 363–381 (2023).
41. Baquero-Perez, B. et al. N(6)-methyladenosine modification is not a general trait of viral RNA genomes. *Nat. Commun.* **15**, 1964 (2024).
42. Fu, L. et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136**, 11582–11585 (2014).
43. Moody, E. M., Lecomte, J. T. & Bevilacqua, P. C. Linkage between proton binding and folding in RNA: a thermodynamic framework and its experimental application for investigating pKa shifting. *Rna* **11**, 157–172 (2005).
44. Gottstein-Schmidtke, S. R. et al. Building a stable RNA U-turn with a protonated cytidine. *RNA* **20**, 1163–1172 (2014).
45. Nixon, P. L. & Giedroc, D. P. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J. Mol. Biol.* **296**, 659–671 (2000).
46. Ruckriegel, S., Hohmann, K. F. & Furtig, B. A Protonated Cytidine Stabilizes the Ligand-Binding Pocket in the PreQ(1) Riboswitch in Thermophilic Bacteria. *Chembiochem* **24**, e202300228 (2023).
47. Dawlaty, M. M. et al. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24**, 310–323 (2013).
48. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
49. Kellner, S. et al. Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, e142 (2014).
50. Kienhofer, S. et al. GADD45a physically and functionally interacts with TET1. *Differentiation* **90**, 59–68 (2015).
51. Amort, T. et al. Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* **18**, 1 (2017).
52. Van Haute, L. et al. Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. *Nat. Commun.* **7**, 12039 (2016).
53. Van Haute, L. & Minczuk, M. Detection of 5-formylcytosine in Mitochondrial Transcriptome. *Methods Mol. Biol.* **2192**, 59–68 (2021).
54. Link, C. N. et al. Protonation-Dependent Sequencing of 5-Formylcytidine in RNA. *Biochemistry* **61**, 535–544 (2022).

Acknowledgements

We thank Meelad M. Dawlaty for mESC lines, Deepa Jayaprakashappa for the pJET1.2-hLMNA plasmid and Sanja Nikolić for technical support in generating *Alkbh1* KOs. We gratefully acknowledge technical support and advice by the IMB Genomics Core Facility, DKFZ NGS Core Facility and the IMB Cytometry Core Facility. Funding of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) supported the acquisition of the Agilent 6490 triple-quadrupole mass spectrometer (#240891705, Christof Niehrs), the Illumina NextSeq 500 (#329045328, IMB Genomics Core Facility) and the Invitrogen Bigfoot Cell Sorter (#511658729, IMB Cytometry Core Facility).

Author contributions

Conceptualization: J.D. and C.N.; Experiments performed by J.D. and C.S.; Bioinformatics analysis: A.G.; LC-MS/MS analysis: J.D. and M.M.; Writing-Original Draft: J.D.; Writing-Review and Editing: A.G., C.N., J.D. and M.M.; Supervision: C.N. and J.D. Funding acquisition: C.N.

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/s41467-025-66090-3.

Correspondence and requests for materials should be addressed to Christof Niehrs.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.