Identification and quantification of modified nucleosides in *Saccharomyces cerevisiae* mRNAs

Mehmet Tardu[1], Joshua D. Jones[1], Robert T. Kennedy[1], Qishan Lin[2], and Kristin S. Koutmou[1]*

[1]University of Michigan, Department of Chemistry, 930 N University, Ann Arbor, MI 48109, (734) 764-5650, kkoutmou@umich.edu, [2]Mass Spectrometry Consortium for Epitranscriptomics, University at Albany, 1400 Washington Ave, Albany, NY 12222.

This document contains: Methods, Supporting Figures (Figure S1, Figure S2, Figure S3, Figure S4, Figure S5, Figure S6, Figure S7, Figure S8 and Figure S9) and Supporting References.

Methods

Growth conditions and stress experiments

Wild-type and *rra1*Δ *Saccharomyces cerevisiae* cells were grown in YPD medium (non-stressed control, oxidative stress and heat-shock conditions) or in defined synthetic complete medium (SC) with 2% glucose (glucose starvation). For all studies, an individual colony was selected from a plate and inoculated into 10 ml of YPD or SC+glucose medium to grow overnight at 30 °C with agitation (200 rpm). Then, cells were diluted to OD_{600} of 0.05 with YPD or SC+glucose medium and grown to an OD_{600} of 0.6. This culture was used for stress experiments and sample collection. Before exposing cells to different stress conditions, 10 ml of cells grown in YPD medium (OD_{600} = 0.6) were collected and used as a control (un-stressed) to compare with stress-induced samples. To assess the effects of oxidative stress on the mRNA modification profiles of *S. cerevisiae*, cells (OD_{600} = 0.6) were incubated with 0.25 mM H_2O_2 for 30 minutes at 30 °C. For heat-shock experiments cultures of exponentially growing yeast (OD_{600} =0.6) in YPD medium at 30 °C were heat-shocked by adding an equal volume of fresh medium at 44°C, to immediately reach a final temperature of 37°C. Heat-shocked cells were incubated at 37°C for 45 minutes. Glucose-starvation experiments were carried out by growing cells to OD_{600}=0.6 in SC+glucose medium. Then, cells were harvested at 5000 x *g* for 2 minutes, washed three times with SC-glucose medium. After that cells were diluted into fresh SC-glucose medium to OD_{600}=0.6 and incubated at 30 °C for 60 minutes.

mRNA Purification, RNA-seq Library Preparation and qRT-PCR

Total RNA was extracted from 10 ml of yeast cells (OD_{600} = 0.6) using standard hot acidic phenol method[1]. Total RNA samples were treated with RNase-free DNase I (Thermo Scientific, USA) (1 U/µg). mRNA isolation was performed in two sequential steps. In the first step oligo-dT magnetic beads (Dynabeads, Invitrogen, USA) were used to selectively bind poly-adenylated RNAs; these beads hybridize to the poly(A) sequence terminating the 3' end of eukaryotic mRNAs. In the second step, we used a commercial rRNA depletion kit (RiboZero Gold, Illumina) to remove the residual 5S, 5.8S, 18S and 25S rRNAs from our samples. The purity of the isolated mRNA was evaluated using

Bioanalyzer RNA 6000 Pico Kit (Agilent, USA) prior to UHPLC-MS/MS analysis. For each sample, rRNA contamination percentage was calculated with the Bioanalyzer software (**Supplemental Figure S1)**. Additionally, we performed qRT-PCR to measure the levels of $tRNA^{Arg,UCU}$, $tRNA^{Glu,UUC}$, $tRNA^{Ser,UGA}$, 5S rRNA, 18S rRNA, and 25S rRNA to evaluate the purity of our isolated mRNAs (**Supplemental Figure S2)**.

We further evaluate the purity of the isolated mRNAs by RNA-seq (**Figure 1**). Briefly, 50 ng of total RNA and purified mRNA samples were fragmented using fragmentation buffer with the TruSeq RNA Sample Preparation Kit according to manufacturer's instructions (Illumina, USA). After that, fragmented RNAs were used for the first-strand cDNA synthesis using random hexamer-primer, and then the second strand was synthesized by using Second Strand Master Mix which includes DNA polymerase I and RNase H. The double strand cDNAs were purified with AMPure XP beads (Beckman Coulter, USA) and eluted with resuspension buffer followed by end-repair and 3'end adenine nucleotide addition. Finally, indexing adapters were ligated to the fragments and cDNA fragments were enriched by 14-cycles of PCR amplification. Single-end sequencing was performed for the two cDNA libraries (two biological replicates) for each sample (total RNA and purified mRNA) using the Illumina HiSeq 4000 sequencing platform (Illumina, USA). All sequence data are single-end 50 bp. Image processing, base calling, and quality value calculation were performed by the Illumina data processing pipeline. Raw reads were saved in FASTQ format.

qRT-PCR was performed to measure the mRNA levels of *CCT1*, *HSP30* and *HXT2* genes at different time points to verify that stress was induced under each condition (**Supplemental Figure S7**). qRT-PCRs were performed with Luminaris HiGreen qRT-PCR Master Mix using gene-specific primers (**Supplemental Table S6**), with *ACT1* as the internal reference gene. qRT-PCR data was analyzed using the Livak method ($2^{-(\Delta\Delta Ct)}$ method). Briefly, average Ct values for all the target genes and housekeeping gene (*ACT1*) in total RNA and mRNA samples were calculated. Then, ΔCt = Ct (gene of interest) – Ct (housekeeping gene) was calculated for each sample. After that, $\Delta\Delta$Ct = ΔCt (mRNA sample) – ΔCt (Total RNA sample) was calculated. Finally, relative gene level fold change was found by taking 2 to the power of negative $\Delta\Delta$Ct (Relative gene level fold change = $2^{-(\Delta\Delta Ct)}$).

RNA-seq Data Analysis

All sequence data were single-end 50 bp in length. FastQC (v0.11.8) was used to evaluate the initial quality of the raw reads. Then, cutadapt (v1.18) [2] was used to obtain high quality clean reads with the arguments -u 10 -m 15 -q 10. High-quality reads were saved in fastq files and deposited in the Gene Expression Omnibus (GEO) database at NCBI with the accession number GSE126405 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126405). After quality filtering, Bowtie2 (v2.3.4.3)[3] was used to align reads to *Saccharomyces cerevisiae* reference genome (R64-1-1) with the default parameters. Following alignment, mmquant tool (v0.1.0)[4] was used to count the number of mapped reads for each transcript (**Supplemental Table S5**). After that, gene_biotype feature in *S. cerevisiae* GTF file (R64-1-1.95) was used to classify RNA species as mRNA, rRNA, tRNA, etc. Finally, percentage of mapped reads to coding and non-coding RNA species was calculated simply by dividing number of mapped reads to each RNA species to the total number of mapped reads (i.e. [# of mapped reads to mRNAs]/[# of mapped reads to whole transcriptome]).

UHPLC-MS/MS Analysis

Briefly, RNAs were first hydrolyzed to the composite mononucleosides via a two-step enzymatic hydrolysis with nuclease P1 and bacterial alkaline phosphatase (BAP). We do not detect either ac^4C or f^5C in samples of control RNAs (MS2 bacteriophage) digested with the same enzymes. LC separations were performed using a Waters Acquity UPLC HSS T3 (50 x 2.1 mm, 1.8 µm, 100 Å). Mobile phase A was 0.01% (v/v) formic acid in water, and mobile phase B was 0.01% (v/v) formic acid in acetonitrile. The flowrate was 0.2 mL/min, and the gradient used was designed as previously published [5]. Tandem MS analysis of RNA nucleosides was performed on a Waters XEVO TQ-STM (Waters, USA) triple quadrupole mass spectrometer equipped with an electrospray ionization (ESI) source maintained at 500°C and the capillary voltage was set at 3.5 kV with extraction cone of 14 V. Nitrogen flow was maintained at 500 l/h and desolvation temperature at 500 °C. The cone gas flow was set to 150 l/h and nebulizer pressure to 7 bar. Each individual nucleoside modification was characterized by single infusion in positive mode ionization

over an m/z range of 100-500 amu. Further nucleoside characterization was produced by using Waters software part of Intellistart MS/MS method development where a ramp of collision and cone voltages is applied to find optimal collision energy parameters for all possible daughter ions (details in **Supplemental Table S7**). To quantify RNA modified nucleosides, calibration curves were prepared for 40 modified nucleosides including adenosine, cytidine, guanosine and uridine. [^{13}C][^{15}N]-G (1 pg/ul) was used as an internal standard.

A method to extract peak areas from raw data to allow quantification was developed using a combination of instrument manufactures suites, MassLynx V4.1 and TargetLynx (Waters, USA). These methods allowed extraction of information to produce calibration curves from each RNA modification standard. In addition, these programs were used to extract the peak areas to be extrapolated on the standard calibration curves (calibration curve examples given in **Supplemental Figure S8**) for quantification of RNA modifications (quantifications given in **Supplemental Table S1**). Python script / Production of calibration curves as well as quantification from samples was produced in Originlab software suite 2017.

LC-MS/MS analysis of *rra1*Δ nucleosides was performed using a Waters Acquity UPLC HSS T3 (100 x 2.1 mm, 1.8 µm, 100 A°) on a Vanquish ultrahigh-pressure liquid chromatograph (Thermo Fisher Scientific, Gemering, Germany) interfaced to a TSQ Quantum Ultra triple quadrupole mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Mobile phase A was 0.01% (v/v) formic acid in water, and mobile phase B was 0.01% (v/v) formic acid in acetonitrile. The flowrate was 0.4 mL/min, and the gradient used was designed as previously published [5]. The sample injection volume was 5 µL. The autosamples was kept at 4°C, and the column was held at 25°C in still air mode. Electrospray ionization was used in positive mode at 4 kV. The capillary temperature was 200°C, the vaporizer temperature was 350°C, the sheath gas was 10, and the auxiliary gas was 5. Ions were detected in tandem mass spectrometry (MS/MS) mode. To quantify RNA nucleosides, calibration curves were created for the four main bases and 13 nucleoside modifcations. [^{13}C][^{15}N]-G (10 nM) was used as an internal standard. Automated peak integration was performed using XCalibur 3.0 MS software. All peaks

were visually inspected to ensure proper integration. The sources of the nucleosides used as standards are given in **Supplemental Table S8**.

UHPLC-MS/MS Data Processing and Analysis

Each of our reported values **(Figures 1-4, Supplemental Table S1)** reflects data collected from experiments performed with two biological replicates and three technical replicates of each biological sample. The raw data were processed by discarding zero and negative read values. Further analyses were conducted using these data.

To normalize each of our samples for comparison we first internally normalized the levels of each nucleoside by dividing its molar concentration to the total value of its corresponding unmodified nucleoside molar concentration (e.g. $m^7G_{normalized}$ = [m^7G]/[G]). We next determined the retention value for each nucleoside, as described below for m^7G as an example (**Figures 2, 3** and **Supplemental Table S2**).

$$\text{retention of } \mathrm{m^7G} = 100\% * \left(\frac{\mathrm{m^7G_{normalized,mRNA}}}{\mathrm{m^7G_{normalized,total\ RNA}}}\right)$$

To establish how modification levels vary in response to cellular stress we calculated the fold change of each nucleoside under different stress conditions. This was accomplished by dividing the level of each nucleoside in stress exposed mRNA-enriched sample by the level of the nucleoside in control mRNA-enriched sample (no-stress) as described below for m^7G as an example (**Figure 4, Supplemental Figure S6** and **Supplemental Table S4**).

$$\text{fold change of } \mathrm{m^7G} = \left(\frac{\mathrm{m^7G_{normalized,mRNA,stress\ condition}}}{\mathrm{m^7G_{normalized,mRNA,no-stress\ control}}}\right)$$

The number of each nucleoside per mRNA was calculated using an estimated average mRNA length of 1641 nucleotides (1641 nucleotides = average ORF length (1385 nt) + average total UTRs length (256 nt)). We used the internally normalized the levels of each nucleoside, and multiplied it by the average mRNA length and frequency

of each nucleotide in yeast genome to calculate frequency of nucleoside modification per mRNA molecule (e.g. frequency of m^7G nucleoside = ($m^7G_{normalized,mRNA}$)*(average mRNA length)*(frequency of guanine nucleotide in yeast transcriptome)). The number of mRNAs per nucleoside is expressed as 1/frequency of nucleoside (**Supplemental Table S2**).

5-formylcytosine (f^5C) Quantification

The f^5C level of total RNA and purified mRNA were measured using MethylFlash 5-formylcytosine (5-fC) Quantification Kit (Epigentek, USA) according to the manufacturer's instructions, including the negative and positive controls. 100 ng of RNA in 10 µl RNase-free dH2O was used to estimate f^5C concentration in the samples according to the standard curve constructed from the manufacturer provided controls (**Supplemental Figure S5**).

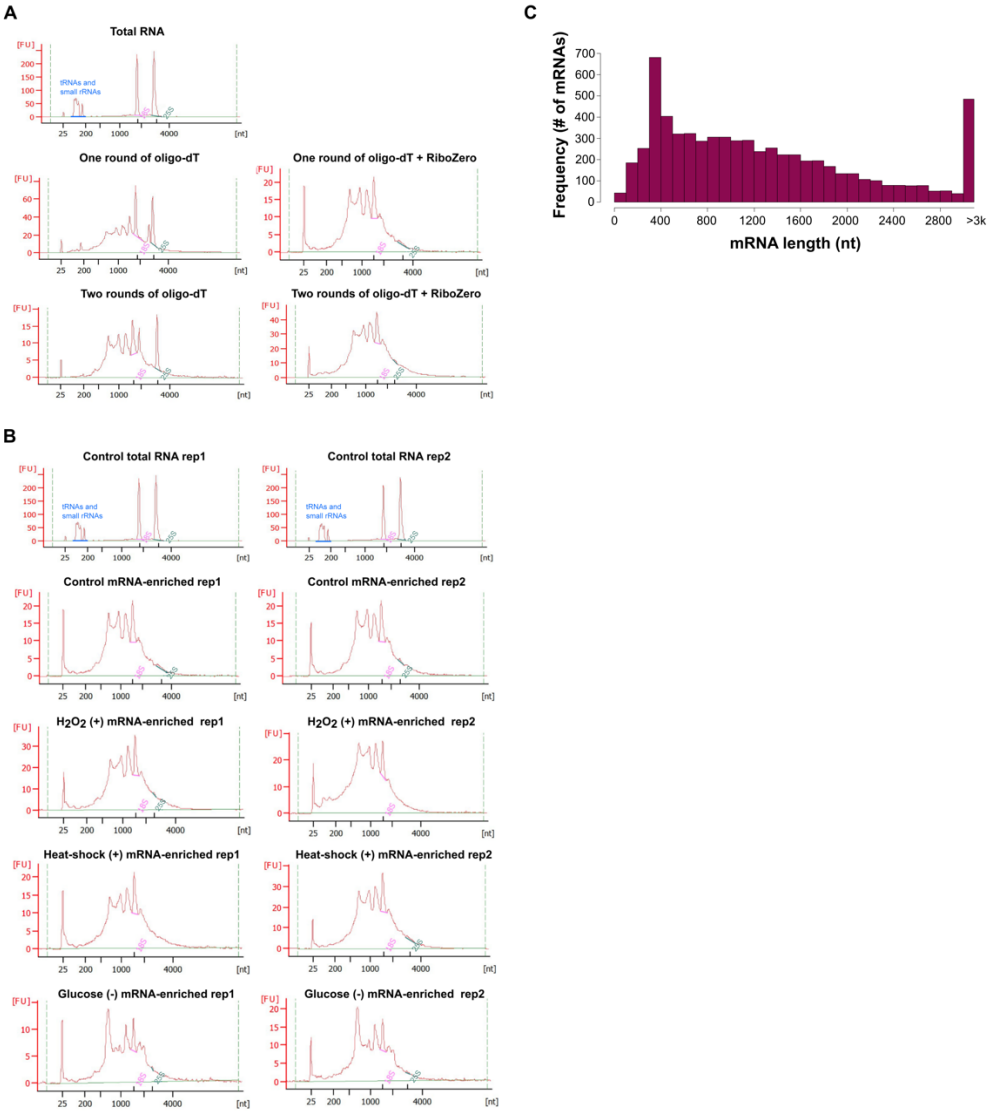

Figure S1: Electropherogram images of total RNA and mRNA-purified samples. Each panel indicates **A)** Control total RNA, and different mRNA-purification schemes. **B)** Electropherogram for mRNAs collected by one oligo-dT pull-downs and RiboZero kit from cells grown under unstressed, H_2O_2, heat-shock, and glucose-starvation conditions with two biological replicates. The x-axis represents size distribution of each samples in nucleotides [nt] while the y-axis depicts signal intensity measured in fluorescence units [FU]. Note that fluorescence levels of 18S and 25S rRNAs decreased from ~200 FU to ~20 FU after mRNA enrichment protocol. Bioanalyzer is a chip-based capillary electrophoresis technique that separates RNAs by size, but does not provide any sequence information. Therefore, the peak on the Bioanalyzer electropherogram marked as 18S rRNA actually reports on all of the RNA species roughly the length of 18S rRNA (1700-1900 nucleotides). There are > 350 yeast mRNAs of this length (shown in **(C)**). When the area under this aggregate peak is quantified, we find that ~3-4% of the total RNA in the Bioanalyzer chromatogram is between 1700-1900 nucleotides in length. Taking into account the fact that there are many mRNAs 1700-1900 nucleotides in length, our Bioanalyzer assays agree with the RNA-seq results shown in Figure 1 indicating that our mRNA samples have < 1% 18S rRNA contamination.

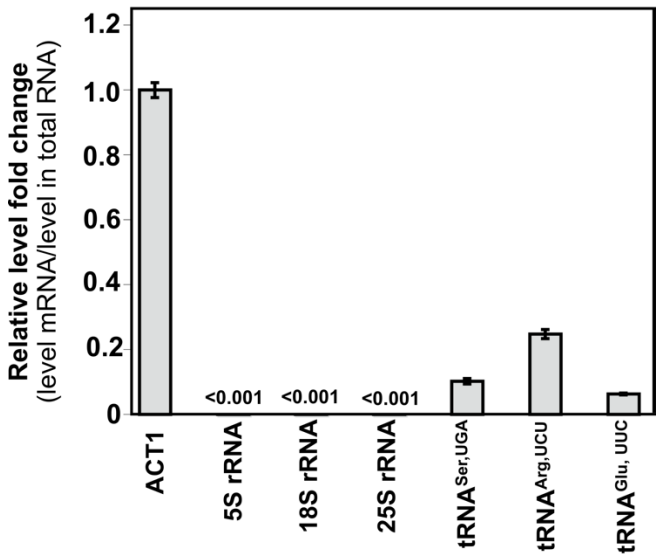

Figure S2: Non-coding RNAs are depleted in mRNA samples. qRT-PCR demonstrates that the levels of rRNAs and tRNAs in our mRNA samples are very low. Notably, because using qRT-PCR can be problematic as an approach to quantify tRNAs and standard RNA-seq does not reliably report on the presence of tRNAs, we heavily relied on the results of our extremely sensitive UHPLC-MS/MS assay to assess how much of our signal could potentially arise from contaminating tRNAs. We were unable detect signals for 13 naturally occurring yeast tRNA modifications in our purified mRNA samples (i^6A, m^1A, m^1I, $cmnm^5U$, m^3U, mnm^5s^2U, mnm^5U, mo^5U, s^4U, m_2^2G, Um, mcm^5U, mcm^5s^2U). Chromatograms demonstrating the presence of several tRNA modifications in total RNA, and the absence of the same tRNA modifications in mRNA, are displayed in Figure 1C and Supplementary Figure S4. Our inability to detect any signal in our mRNA samples for the majority of common tRNA modifications that we assayed provides strong evidence that our samples are deplete of tRNAs below our limit of detection.

We critically assessed our ability to detect modifications on the tRNA that gave the highest signal in our qRT-PCR experiment, $tRNA^{Arg}$. $tRNA^{Arg}$ is highly modified and we were concerned that potential $tRNA^{Arg}$ contamination could impact the conclusions of our study. We find that the mcm^5U modification only found in $tRNA^{Arg}$ in yeast is measurable in our total RNA samples, but NOT detectable in our mRNA samples. Additionally, we see no evidence for a second modification on $tRNA^{Arg}$, m^1A, in our mRNA samples. This suggests that signal originating from any potential contamination $tRNA^{Arg}$ is below our limit of detection (as highlighted in Figure 1B). Furthermore, because we were concerned about ac^4C in particular, we analyzed the retention levels and concentrations of modifications on the two tRNA species where ac^4C is found – $tRNA^{Ser}$ and $tRNA^{Leu}$ (see Figure 3). We find that all of the modifications on tRNAs where ac^4C are present are either not detected (e.g. m^1A, i^6A), or retained at levels ~10-fold less than ac^4C (see Figure 3B).

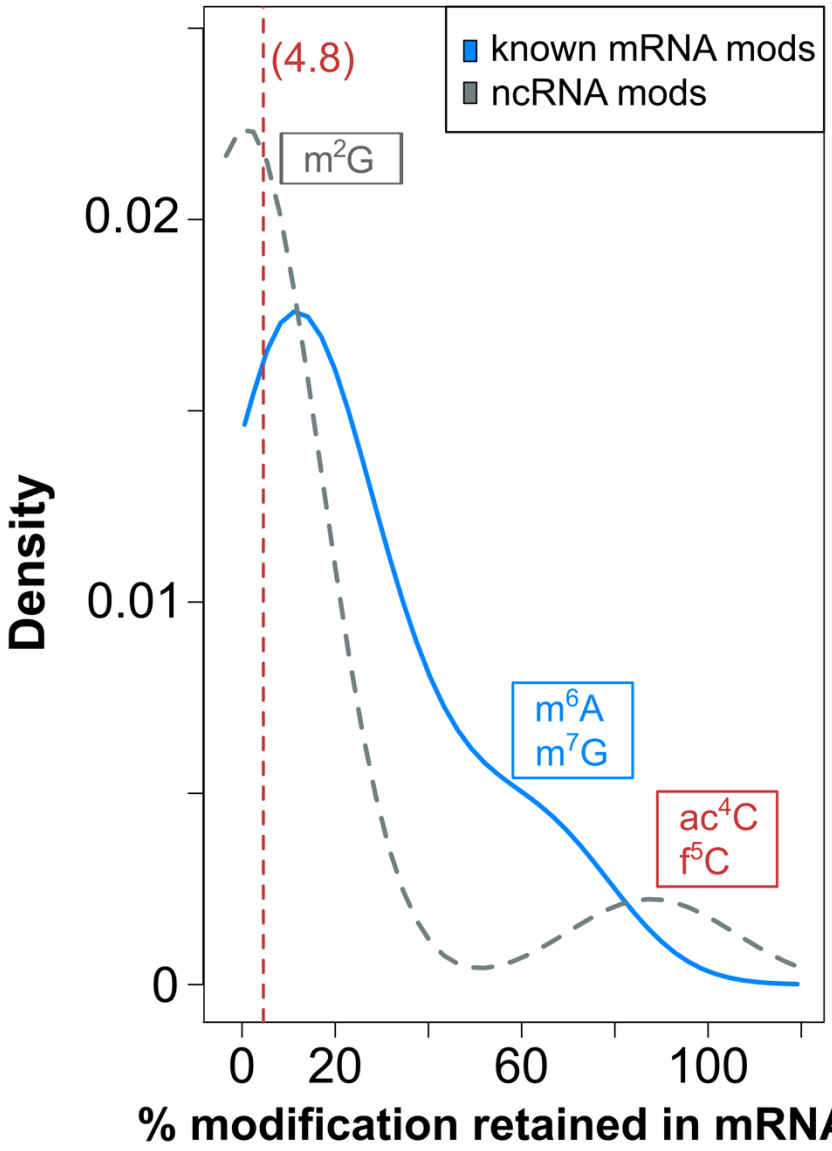

Figure S3: Distribution of nucleoside modification retention levels in our mRNA samples. Blue line indicates known mRNA modifications while dotted grey line indicates modifications previously identified only in non-coding RNA species. As expected, most known mRNA modifications are well retained, and 90% of the annotated non-coding RNA modifications were either not detectable or poorly retained with average retention of 4.8 ± 2.4% (red vertical dashed line). The most highly retained modifications only previously annotated as non-coding modification were m^2G (6.8 %), ac^4C (80%) and f^5C (90%); these are denoted on the plot. We empirically set a cut-off for denoting positive 'hits' for new mRNA modifications at two standard deviations greater than the average retention value of most detectable non-coding modifications, 9.48 %.

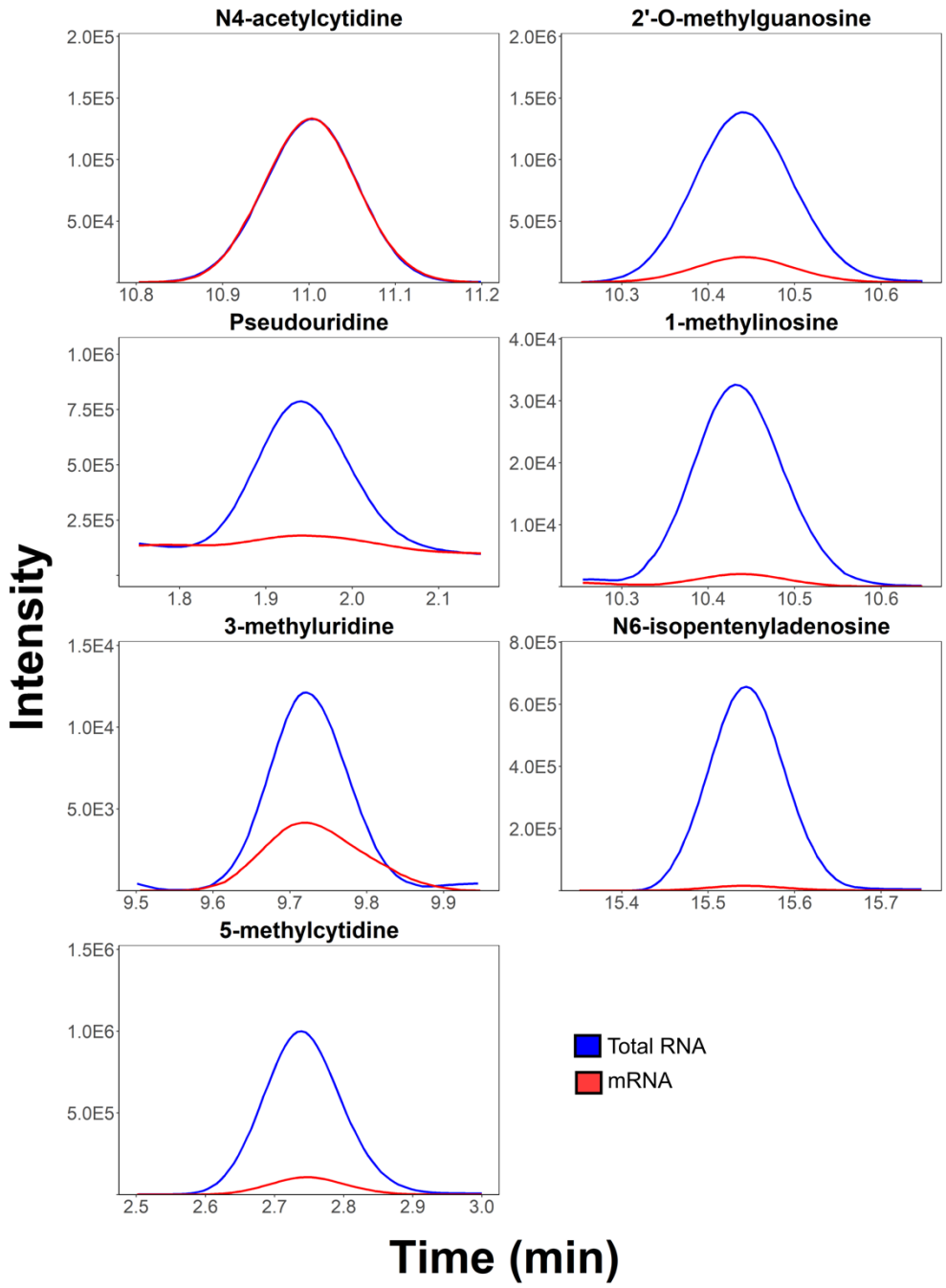

Figure S4: Chromatograms of nucleoside modifications analyzed by UHPLC-MS/MS. Overlaid extracted ion chromatograms of seven modification in total RNA (blue line) and in purified mRNA (red line). The N4-acetylcytidine chromatograms are from total RNA and purified mRNA from oxidative stress growth conditions. The rest of the chromatograms are from no-stress growth conditions.

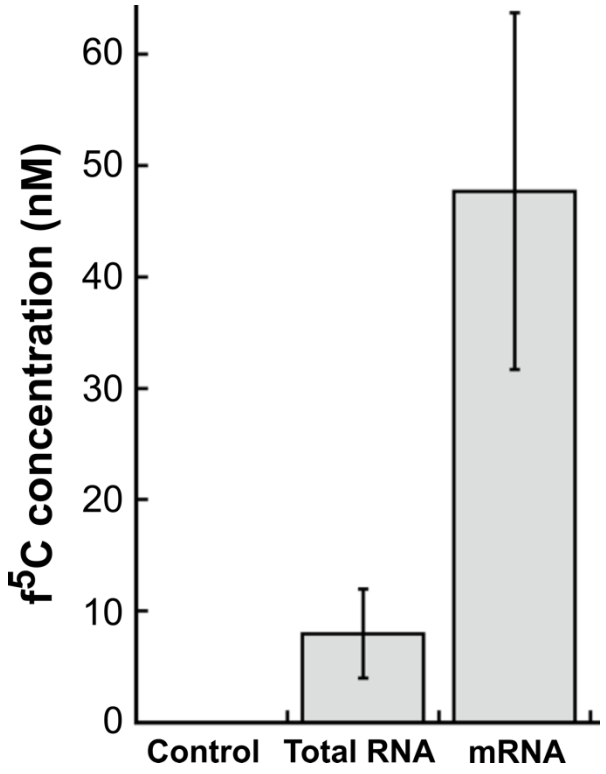

Figure S5: f^5C-levels in total RNA and mRNAs. The concentrations of f^5C were measured with an orthogonal antibody based detection kit on an unmodified oligo (negative control), total RNA and mRNA samples from H_2O_2 stressed cells. The concentrations of f^5C were calculated using the standard curve constructed from the manufacturer provided controls. The values displayed reflect the average concentraitons in two different biological samples, and the error bars represent the standard error on the mean.

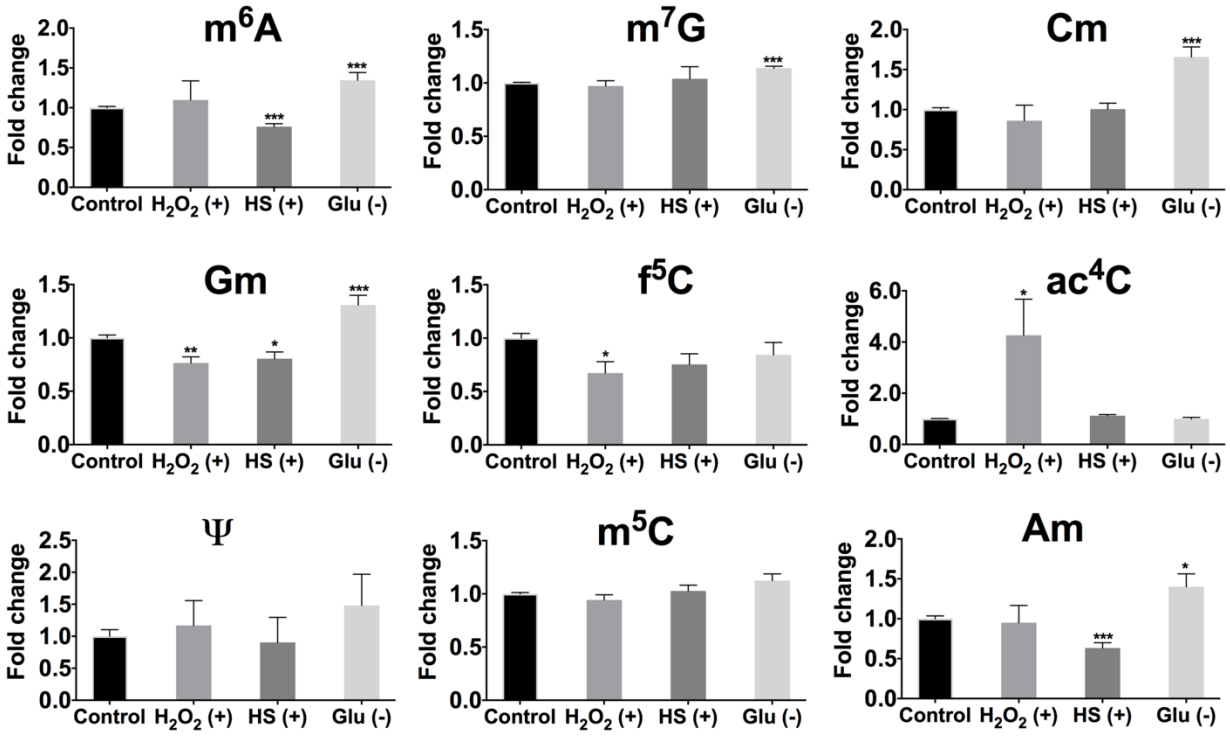

Figure S6: **Post-transcriptional modifications in mRNA of *S. cerevisiae* exhibit differential responses to environmental stressors.** The fold change for all modifications identified either here or previously upon stress induction was calculated as the ratio of nucleoside level in the mRNA sample of stress-treated condition (H_2O_2 (+), heat-shock (HS) (+) and glucose (Glu) (-)) and nucleoside level in the mRNA sample of no-stress condition (control). Error bars represent the standard error of mean. *p<0.05, **p<0.01, ***p<0.005, Student's t test.

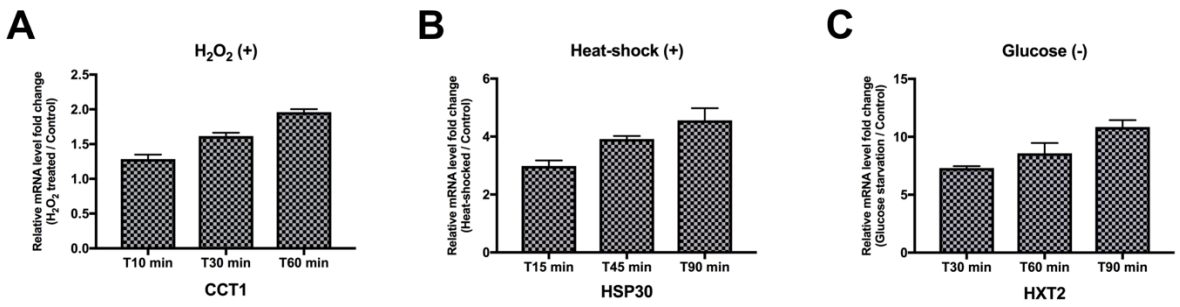

Figure S7: Validation of stress induction protocols by qRT-PCR on known stress response genes. Stress induction for **A)** oxidative-stress **B)** heat-shock stress and **C)** glucose-starvation conditions was confirmed by measuring the mRNA levels of stress-specific *CCT1*, *HSP30* and *HXT2* genes, respectively. Each bar represents relative mRNA level fold changes of genes at indicated stress exposure time with respect to unstressed control samples. Results were obtained from three independent biological replicates (n = 3), and error bars represent the standard deviation between these values.

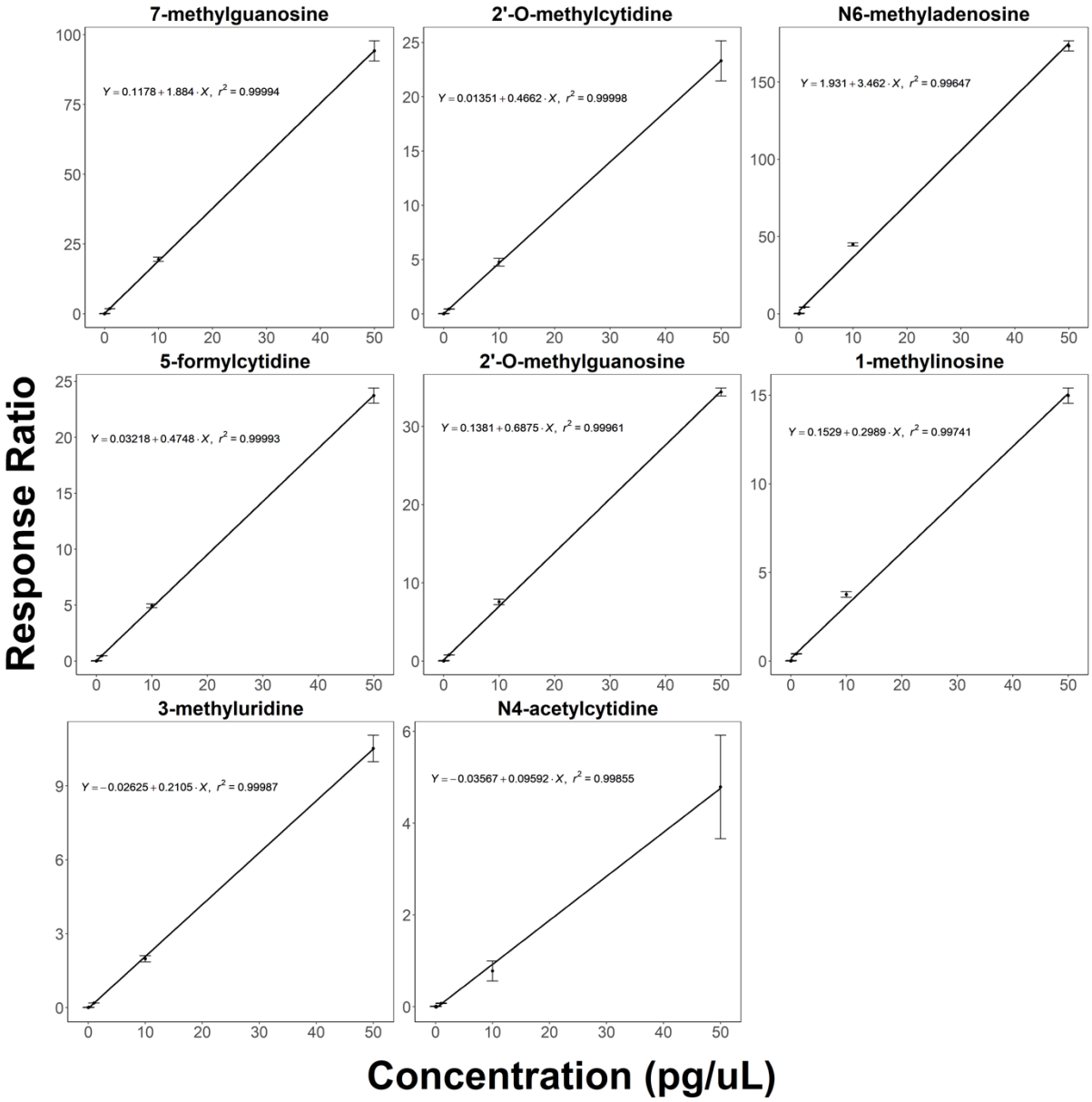

Figure S8: Example calibration curves used to quantitate modified nucleosides during UHPLC-MS/MS analysis. Response ratio (ratio between peak area of modification and internal standard) versus concentration of standard (0.001 to 50 pg/μL). Three technical replicates were performed for each standard. The error bars represent the standard deviation of the response ratios of the technical replicates.

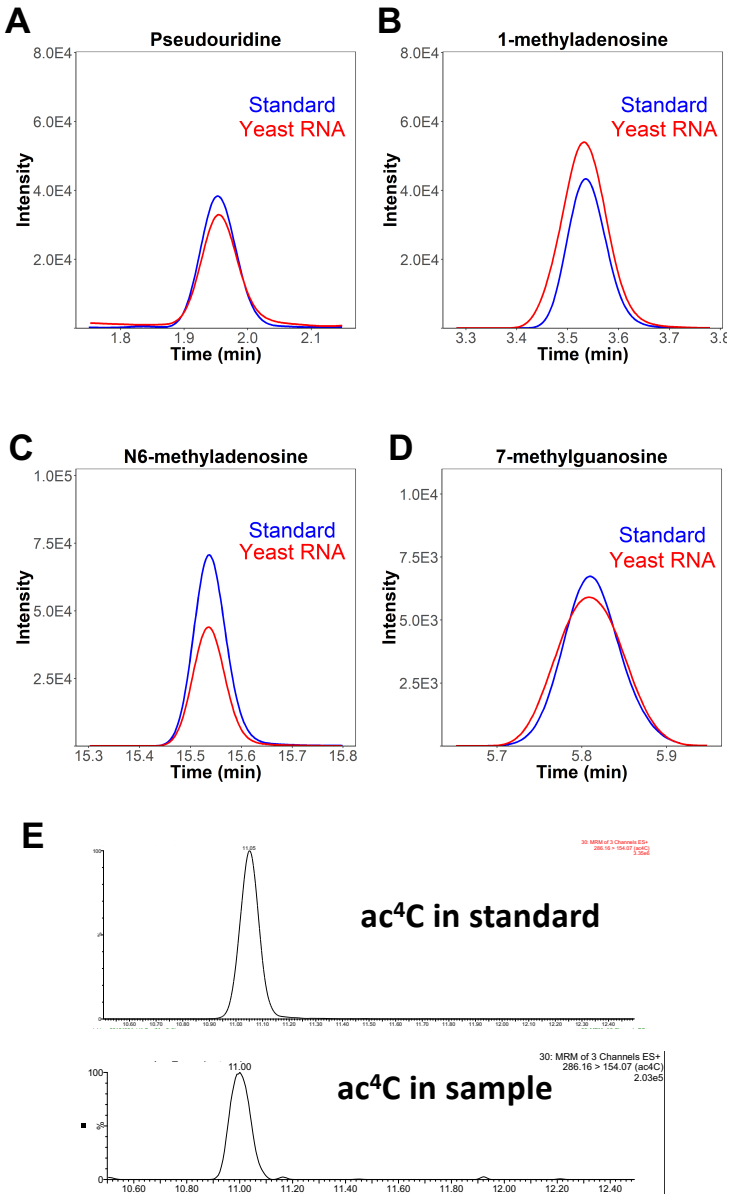

Figure S9. Example overlaid chromatograms of nucleoside standards and yeast total RNA samples. Overlaid chromatograms of standard (blue line) and yeast total RNA (red line) samples of **(A)** pseudouridine, **(B)** 1-methyladenosine, **(C)** N6-methyladenosine, and **(D)** 7 methylguanosine during the LC-MS/MS nucleoside analysis. The analysis was performed using a Waters Acquity UPLC HSS T3 (100 x 2.1 mm, 1.8 µm, 100 Å) as explained in supporting information. **(E)** Chromatograms of ac^4C in standard and yeast total RNA. Not displayed as overlay demonstrate the dearth of other nucleosides retained on a similar time scale.

References

1. Schmitt, M. E., Brown, T. A., and Trumpower, B. L. (1990) A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae, *Nucleic Acids Res 18*, 3091-3092.
2. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet.journal 17(1)*, 10--12.
3. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol 10*, R25.
4. Zytnicki, M. (2017) mmquant: how to count multi-mapping reads?, *BMC Bioinformatics 18*, 411.
5. Basanta-Sanchez, M., Temple, S., Ansari, S. A., D'Amico, A., and Agris, P. F. (2016) Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells, *Nucleic Acids Res 44*, e26.