N4-acetylcytidine and 5-formylcytidine are present in *Saccharomyces cerevisiae* mRNAs

Mehmet Tardu[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.

Abstract

Chemical modifications of RNAs have long been appreciated as key modulators of non-coding RNA structure and function. There is an emerging realization that chemical modification of protein-coding mRNAs also plays critical roles in the cell. Nonetheless, of the over 100 known RNA modifications found in biology only a handful have been identified in mRNAs. Here we use an ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) method to identify and quantify modifications present in mRNAs of yeast cells. We detect the presence of four modified nucleosides in mRNAs at relatively high abundances: N7-methylguanosine, N6-methyladenosine, N4-acetylcytidine and 5-formylcytidine. Additionally, we investigate how the levels of mRNA modifications change in response to cellular stress. We find that the concentrations of nine mRNA modifications including N6-methyladenosine and N4-acetylcytidine change in response to heat stress, glucose starvation and/or oxidative stress. Our findings suggest that mRNA modification may provide a potential mechanism for cells to rapidly respond to environmental stressors.

Main text

The presence of modified nucleosides was first reported in rRNAs and tRNAs over six decades ago[1-2]. To date over 100 RNA modifications have been identified across phylogeny[3]. These modifications are integral to biology and serve as crucial determinants of non-coding RNA structure and function[4]. There is a growing appreciation that RNA modifications are also found in coding messenger RNAs (mRNAs), and thus may potentially contribute to the stability and translation of mRNAs[5-6]. However, the specific

biological role, frequency and distribution of modified ribonucleosides in mRNAs is just being recognized. Although the list of chemical modifications present in mRNAs is growing, there are still fewer than 15 varieties[7-8] of RNA chemical modifications known in mRNAs. Given the diversity of RNA chemical modifications found in non-coding RNA molecules, it is likely that the full catalog of RNA modifications present in mRNAs has not yet been uncovered. Recent advances in deep-sequencing technology unveiled the incorporation of a limited set of modifications in mRNAs including N6-methyladenosine (m^6A), 5-methylcytosine (m^5C), N1-methyladenosine (m^1A)[9], and pseudouridine (Ψ)[6, 10-15]. However, there is still some debate about the precise location and prevalence of several modifications; with m^5C and m^1A, in particular, being actively investigated. It is widely believed that mRNA modification is highly dynamic and likely serves as a gene regulation mechanism since the enzymatic incorporation of mRNA modifications has the potential to modulate mRNA stability and folding, protein-recruitment, and translation in a programmed manner[3, 10, 16]. However, much still remains to be understood about how the levels of modified mRNA nucleosides, resulting from either RNA damage or enzymatic incorporation, alter the stability, structure, function, and translatability of mRNAs.

Here we seek to expand the variety of known nucleoside modifications present in mRNAs and to determine how the levels of specific mRNA modifications can be modulated. To accomplish this we examined the nucleoside modification profile of *Sacchromyces cerevisiae* mRNAs using high throughput ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS)[17] and quantified how the levels of modified mRNA nucleosides change in response to environmental stress. Our work reveals the presence of two chemical modifications previously not annotated in mRNAs, and evaluates if the levels of 12 mRNA modifications, identified by ourselves and others, are impacted by cellular stress. We find that the levels of nine modified nucleosides in mRNAs are altered under different environmental stress conditions, consistent with the idea that the mRNA epitranscriptome is not static.

To determine the best conditions for purifying yeast mRNAs we tested four different purification schemes: single oligo-dT pull down, two oligo-dT bead pull-downs, single oligo-dT bead pull-down followed by a RiboZero rRNA depletion kit, and two oligo-dT pull downs followed by RiboZero depletion. Ultimately, we selected to use a two-step

protocol (oligo-dT pull-down followed by a RiboZero kit) because it yielded mRNAs of equal or greater purity than all other tested methods, and at a sufficient concentration for analysis. We first evaluated the quality and quantity of our mRNA samples by BioAnalyzer (**Supplementary Figure S1**). We then verified the purity of our isolated mRNA samples by qRT-PCR. Our qRT-PCR assays directly measured the levels of rRNAs (5S rRNA, 18S rRNA, 25S rRNA) and a diverse set of tRNAs mRNAs ($tRNA^{Arg,UCU}$, $tRNA^{Glu,UUC}$, $tRNA^{Ser,UGA}$). qRT-PCR data indicate that our mRNA samples lack rRNAs and are depleted of tRNAs **(Figure 1 and Supplementary Table S1, Supporting Information)**.

To further investigate the purity of our mRNA samples we used the results of our UPHLC-MS/MS assay, which contains many internal negative and positive controls. UHPLC-MS/MS is a powerful tool that allows us to simultaneously quantify the levels of multiple nucleosides with high accuracy, sensitivity and selectivity; quantifying modification levels down to attomolar concentrations (10^{-18} moles/L)[17]. We used this technique to measure the levels of 42 known RNA modifications simultaneously and assessed the levels of all modifications in mRNA samples relative to their levels in total RNA (level in mRNA/level in total RNA; **Figure 1, Supplementary Table S2, and Supporting information**). As positive controls we evaluated the levels of the widely accepted mRNA modifications N7-methylguanosine (m^7G) and N6-methyladenosine (m^6A). For comparison, we examined the levels of multiple modifications previously found in non-coding RNAs (e.g. N6-isopentyladenosine (i6A), 3-methyluridine (m^3U), 5-methyluridine (m^5U), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine ($m^1acp^3\Psi$), N2,N2-dimethylguanosine ($m^2{}_2G$), 1-methylguanosine (m^1G), dihydrouridine (D), andN2-dimethylguanosine (m^2G)) **(Supplementary Table S3)**. We find that m^7G and m^6A were retained at expected levels consistent with previous reports (40-90%), while the levels of the comparison non-coding RNA modifications are depleted below background **(Figure 1 and Supporting Information).** Since the distribution of modification levels is clearly bimodal, we can use a modification's level to determine the likelihood that it is present in mRNA. The cut-off for considering modifications as in mRNA was based on modification retention in mRNAs, modification concentration, BioAnalyzer and qRT-PCR, as described in detail in the **Supporting Information**. Only three modifications previously identified as non-coding were retained in mRNAs at levels (80-

90%) and concentrations comparable to m^7G and m^6A: formylcytidine (f^5C), N4-acetylcytidine (ac^4C) and 5-methylaminomethyluridine (mnm^5U); these modifications are further scrutinized below **(Figure 1 and Supplementary Table S2)**.

We estimated the occurrence of each of the 42 modified nucleotides measured in our assay per mRNA **(Supporting Information)** and find that the most common modification in mRNAs is m^7G **(Supplementary Table S4)**. Our results indicate that m^7G is incorporated roughly once per mRNA (~1.2-times per mRNA). This was not entirely unexpected because m^7G recently been reported to be incorporated internally in mRNAs[18]. Similarly, the m^6A prevalence we estimate, approximately once per every 1-2 mRNAs, is consistent with previous estimates based on transcriptome wide m^6A mapping studies conducted in human cells[9] **(Supplementary Table S4)**. As mentioned above, we also found f^5C, mnm^5U and ac^4C present every once every 1-10 mRNAs **(Supplementary Table S4)**. Interestingly, both f^5C and ac^4C are conserved in all kingdoms of life, similar to 90% of the other mRNA modifications identified to date **(Figure 2A)**. It is also clear how both modifications are made in cells; f^5C is the oxidative product of m^5C through 5-hydroxymethylcytidine (hm^5C) [19] and ac^4C is incorporated by the enzyme Rra1[20]. In contrast, mnm^5U has only been observed in bacteria and eukaryotes and little is known about where this modification is found, beyond its location in yeast mitochondria[8]. We therefore are hesitant to draw any firm conclusions about mnm^5U in yeast mRNAs until the modification itself is better understood in eukaryotes, and we possess the knowledge necessary to design controls to evaluate the levels the non-coding RNAs where it can be found. There were four additional modifications retained in our mRNA samples that we find present only at very low concentrations; less than once per every 100 mRNAs (5-methyoxyuridine (mo^5U), 5-methylaminomethyl-2-thiouridine (mnm^5s^2U), 5-methoxycarbonylmethyluridine (mcm^5U), 2-thiouridine (s^2U)) **(Supplementary Table S4)**. The infrequency of these modifications suggests they could result from off-target enzyme activity or low levels of mRNA damage, and we posit that the biological role of these modifications in mRNAs – if any – is restricted to very specific circumstances. We do not classify these nucleosides as mRNA modifications without substantial follow-up studies.

Ac^4C is found in both tRNAs and rRNAs and we were concerned about the possibility that our mRNA samples could be enriched for non-coding RNAs containing the modification. However, our qRT PCR studies measuring the levels of rRNA and tRNA species where ac^4C is found (25S rRNA, $tRNA^{Ser,UGA}$) indicate that these RNAs are depleted to levels far below the level of ac^4C in mRNAs **(Figure 1)**. We also took advantage of the fact that tRNAs and rRNAs are highly modified molecules to assess our mRNA sample purity in the context of ac^4C. We reasoned that if tRNA and rRNA degradation fragments containing ac^4C are present in our mRNA samples, then modifications located in close proximity to ac^4C on the same tRNA or rRNA should also be retained in our mRNA samples. We evaluated the retention levels of all of the modifications present in tRNAs where ac^4C are found **(Figure 3)** and paid particular attention to those modifications in close proximity (2-10 nucleotides) to ac^4C. We find that the modifications present in tRNAs containing ac^4C are not present above background in our mRNA samples **(Figure 3)**. Additionally, we also find that a modification only found once in cells in 18S rRNA (1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine ($m^1acp^3\Psi$)) is not present above background **(Figure 3)**.

Because we retained four modifications present in relatively low abundance in our mRNAs (m^6A, ac^4C, f^5C, and mnm^5U), we considered the possibility that our purification artificially enriched for tRNA modifications present at low to moderate levels. However, our identification of mRNAs modifications appears to be specific. While N6-isopentyladenosine (i6A), 3-uridine (m^3U), 1-methylinosine (m^1I) and 2-thiocytidine (s^2C) levels in total RNA are comparable or lower than those of m^6A, ac^4C, mnm^5U and f^5C, we do not detect any of these modifications above background in our mRNAs.

Our systematic analyses of our data are particularly important in light of the current dialogue in the burgeoning field of epitranscriptomics regarding prevalence of several mRNA modifications. While the groundbreaking studies that mapped m^5C and m^1A to the transcriptome seemed to indicate that these modifications are present at thousands of sites, recent follow-up reports suggest that these modifications may only be present in a handful of mRNAs[21-23]. For example, Lyko and co-workers have raised concerns about the presence of m^5C in mRNAs, reporting that when mRNA is sufficiently purified and the bisulfite sequencing data is alternatively modeled, m^5C is seldom present[22]. Our results support these findings and suggest that m^5C, if present, is not common in yeast mRNAs.

However, we also did not detect the levels of some other previously identified mRNA modifications above background – indicating that our approach may only be able to identify modifications that are abundant in mRNAs relative to other RNA species. This may explain why we find Ψ and the 2' OMe modified nucleosides present just below background in our mRNA samples (e.g. Ψ is estimated to occur in only ~ 5% mRNA sequences, but is present in all tRNAs and rRNAs[11]), but we detect appreciable levels of m^6A and m^7G in mRNAs (discussed in **Supporting Information**). While m^6A and m^7G are both in non-coding and coding RNAs, the preponderance of these modifications are confined to mRNA species.

Nucleoside modification levels are dynamically regulated in eukaryotic non-coding RNAs[24]. tRNA modifications, for example, can mediate the cellular response to stress by controlling the selective, codon-biased translation of particular mRNAs[24], and modulations in rRNA modification patterns have been observed in response to environmental changes, during development, and in disease[25]. Akin to this, it is entirely possible that mRNA modifications also contribute to the rapid cellular responses to environmental changes. The modification of mRNA has the potential to quickly fine-tune metabolism by directly regulating protein production and folding[26-28]. We assessed how mRNA modification levels change in response to stress using UHLPC-MS/MS. We analyzed modified nucleoside levels in mRNA samples collected from yeast grown under three stress-conditions: oxidative stress, heat-shock, and glucose starvation. We find that the levels of 9 mRNA modifications demonstrate statistically significant variations from basal conditions ($p < 0.005$, $p < 0.01$, and $p < 0.05$) in response to environmental challenges **(Figure 4, Supplementary Figure S2 and Supplementary Table S5)**. The changes in mRNA modification levels under different stress conditions suggests that such nucleoside modifications could potentially alter mRNA function during environmental changes.

Three of the modifications we investigated here are of particular interest: m^6A, f^5C, and ac^4C. m^6A is widely conserved **(Figure 1A)** [29-3129-3128-3028-30]and recent studies demonstrate that it has an important role in regulating various biological processes including heat-shock response[32], circadian rhythm[33] and stem cell differentiation[34]. Our analysis revealed that m^6A exhibited statistically significant fluctuations in levels ($p < 0.005$) under heat shock (-24%), and glucose starvation conditions (+34%) **(Figure 4 and**

Supplementary Table S5). Given that the RNA N6-methyladenosine transferase complex (MIS) responsible for incorporating m^6A in yeast is triggered by nutrient limitation, it may make sense that the largest m^6A levels are observed under glucose starvation[35], and we might expect to find even higher levels under meiosis or sporulation. N4-acetylcytidine (ac^4C) is present in total RNA in all of our samples but only detectable in the mRNA-samples grown under oxidative stress **(Figure 4)**. Our results indicate that ac^4C is one of the most prevalent modified nucleoside species under oxidative stress. This observation, coupled with the conservation of ac^4C across phylogeny **(Figure 1A)** lead us to propose that ac^4C may have a possible role in regulating the cellular response to oxidative stress in yeast. Like m^7G, m^6A and ac^4C, f^5C is also found in all domains of life. f^5C is an *in vivo* oxidation product of the m^5C, via hm^5C, and has been observed in total RNA from all domains of life and in polyA-enriched RNA fractions from mammalian cells[36]. Our findings confirm the speculation that f^5C is in mRNAs[37] and suggest that f^5C may be quite common.

In sum, we investigated the mRNA post-transcriptional modification landscape in *S. cerevisiae* by examining its epitranscriptome profile under conditions of oxidative stress, heat-shock and glucose starvation. Our results expand the pool of known mRNA modifications to include N4-acetylcytidine and 5-formylcytidine and support the idea that mRNA modification may provide a means for cellular adaptive mechanisms as a rapid response to environmental stressors.

Materials and Methods

Growth Conditions and Stress Experiments

Saccharomyces cerevisiae (BY4741) 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). Before exposing cells to different stress conditions, cells grown in YPD medium (OD_{600} = 0.6) were collected and used as a control. Stress conditions were as follows: 1) oxidative stress: incubation of cells with 0.25 mM H_2O_2 (30 minutes, 30 °C), 2) heat-shock: cells grown at 37°C for 45 minutes, 3) glucose-starvation: cells grown in SC (-) glucose media (30°C, 60 minutes). Details for each growth condition can be found in **Supporting Information**.

Total RNA Extraction, mRNA Enrichment and qRT-PCR

Total RNA was extracted using hot acid phenol. Total RNA samples were treated with RNase-free DNase I, and mRNA was isolated in two sequential steps: oligo-dT magnetic beads (Dynabeads) were first used to selectively isolate poly-adenylated RNAs, followed by a commercial rRNA depletion kit (RiboZero Gold) to remove residual rRNAs. The purity of the isolated mRNA was evaluated using Bioanalyzer RNA 6000 Pico Kit, qRT PCR, and analysis of modification levels **(Figure 1, Supplementary Figure S1, Supplementary Table S2 and Supporting Information)**.

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 (**Supplementary Figure S3**). qRT-PCRs were performed with Luminaris HiGreen qRT-PCR Master Mix using gene-specific primers (**Supplementary Table S6**), with *ACT1* as the internal reference gene.

UHPLC-MS/MS Analysis

RNA samples (100ng/10μL each) were analyzed as previously described[17]. We used a Waters XEVO TQ-STEM triple quadrapole instrument with sensitivity down to 23.01 femtograms, 64.09 attomoles, in large excess of the sensitivity required to analyze and characterize RNA modifications at the single molecular level. Moreover, our samples were separated by a high resolution UHPLC prior to being detected by tandem mass spectrometry instrument, thus further enhancing the selectivity and sensitivity. Details of the technique and data analyses are given in **Supporting Information**.

Data Processing and Analysis

Each of our reported values **(Figures 1-4, Supplementary Table S3)** 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 comparision we first internally normalized the levels of each nucleoside by dividing its molar concentration to the total value of its corresponding modified and unmodified nucleoside molar concentration (e.g. $m^6A_{normalized}$

= $[m^6A]/([A] + [A_{all_modified_nucleosides}])$). We determined if a nucleoside as present above background in our mRNA sample by dividing the level of nucleoside in our mRNA-enriched sample by the level of the nucleoside in our total RNA sample for each condition (e.g. ($m^6A_{normalized,mRNA}/m^6A_{normalized,totalRNA}$)) (**Figures 1, 3** and **Supplementary Table S2**). Fold change of each nucleoside under different stress conditions was calculated by dividing the level of nucleoside in stress exposed mRNA-enriched sample by the level of the nucleoside in control mRNA-enriched sample (no-stress) (e.g. ($m^6A_{normalized,mRNA,\ stress\ condition}/m^6A_{normalized,mRNA,\ no\text{-}stress\ control}$)) (**Figure 4** and **Supplementary Figure S2**).

The number of each nucleoside per mRNA was calculated as follows. We estimated the average mRNA length for the entire yeast genome as 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^6A nucleoside = ($m^6A_{normalized}$)*(average mRNA length)*(frequency of adenine nucleotide in yeast transcriptome)). The number of mRNAs per nucleoside is expressed as 1/frequency of nucleoside (**Supplementary Table S4**).

Author Information

Corresponding author

Kristin Koutmou, 930 N University, Ann Arbor, MI 48105, kkoutmou@umich.edu, (734) 764-5650

ORCID

0000-0002-7763-9262 (K. Koutmou)

Author Contributions

Mehmet Tardu designed, performed and analyzed all RNA isolation experiments, analyzed the mass-spectrometry data, and wrote the manuscript. Qishin Lin performed the UHPLC/mass-spectrometry experiments and analyzed data. Kristin Koutmou designed experiments, analyzed data, and wrote the manuscript.

Funding

This work was funded by the University of Michigan start-up funds (to KSK).

Notes

The authors declare no competing financial interests.

Acknowledgements

We would like to acknowledge Dr. Daniel Eyler for his helpful discussion of data analysis. We thank Dr. Markos Koutmos, Dr. Ryan Bailey and Dr. Anna Mapp for their thoughtful reading of the paper and comments.

References

1. Davis, F. F.; Allen, F. W., Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem* **1957,** *227* (2), 907-15.
2. Holley, R. W.; Everett, G. A.; Madison, J. T.; Zamir, A., Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. *J Biol Chem* **1965,** *240*, 2122-8.
3. Gilbert, W. V.; Bell, T. A.; Schaening, C., Messenger RNA modifications: Form, distribution, and function. *Science* **2016,** *352* (6292), 1408-12.
4. Nachtergaele, S.; He, C., The emerging biology of RNA post-transcriptional modifications. *RNA Biol* **2017,** *14* (2), 156-163.
5. Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; Ren, B.; Pan, T.; He, C., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014,** *505* (7481), 117-20.
6. Wang, X.; Zhao, B. S.; Roundtree, I. A.; Lu, Z.; Han, D.; Ma, H.; Weng, X.; Chen, K.; Shi, H.; He, C., N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **2015,** *161* (6), 1388-99.
7. Cantara, W. A.; Crain, P. F.; Rozenski, J.; McCloskey, J. A.; Harris, K. A.; Zhang, X.; Vendeix, F. A.; Fabris, D.; Agris, P. F., The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* **2011,** *39* (Database issue), D195-201.
8. Machnicka, M. A.; Milanowska, K.; Osman Oglou, O.; Purta, E.; Kurkowska, M.; Olchowik, A.; Januszewski, W.; Kalinowski, S.; Dunin-Horkawicz, S.; Rother, K. M.; Helm, M.; Bujnicki, J. M.; Grosjean, H., MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res* **2013,** *41* (Database issue), D262-7.
9. Dominissini, D.; Nachtergaele, S.; Moshitch-Moshkovitz, S.; Peer, E.; Kol, N.; Ben-Haim, M. S.; Dai, Q.; Di Segni, A.; Salmon-Divon, M.; Clark, W. C.; Zheng, G.; Pan, T.; Solomon, O.; Eyal, E.; Hershkovitz, V.; Han, D.; Dore, L. C.; Amariglio, N.; Rechavi, G.; He, C., The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **2016,** *530* (7591), 441-6.
10. Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C., Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **2017,** *169* (7), 1187-1200.
11. Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V., Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **2014,** *515* (7525), 143-6.
12. Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; Leon-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; Fink, G.; Regev, A., Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014,** *159* (1), 148-162.
13. Xuan, J. J.; Sun, W. J.; Lin, P. H.; Zhou, K. R.; Liu, S.; Zheng, L. L.; Qu, L. H.; Yang, J. H., RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res* **2017**.
14. George, H.; Ule, J.; Hussain, S., Illustrating the Epitranscriptome at Nucleotide Resolution Using Methylation-iCLIP (miCLIP). *Methods Mol Biol* **2017,** *1562*, 91-106.
15. Harcourt, E. M.; Kietrys, A. M.; Kool, E. T., Chemical and structural effects of base modifications in messenger RNA. *Nature* **2017,** *541* (7637), 339-346.
16. Frye, M.; Jaffrey, S. R.; Pan, T.; Rechavi, G.; Suzuki, T., RNA modifications: what have we learned and where are we headed? *Nat Rev Genet* **2016,** *17* (6), 365-72.

17. Basanta-Sanchez, M.; Temple, S.; Ansari, S. A.; D'Amico, A.; Agris, P. F., Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res* **2016,** *44* (3), e26.
18. Chu JM, Y. T., Ma CJ, Lan MD, Liu T, Yuan BF, Feng YQ., Existence of Internal N7-Methylguanosine Modification in mRNA Determined by Differential Enzyme Treatment Coupled with Mass Spectrometry Analysis. *ACS Chemical Biology* **2018,** (Epub, ahead of print).
19. Fu, L.; Guerrero, C. R.; Zhong, N.; Amato, N. J.; Liu, Y.; Liu, S.; Cai, Q.; Ji, D.; Jin, S. G.; Niedernhofer, L. J.; Pfeifer, G. P.; Xu, G. L.; Wang, Y., Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J Am Chem Soc* **2014,** *136* (33), 11582-5.
20. Ito S, A. Y., Noma A, Kimura S, Miyauchi K, Ikeuchi Y, Suzuki T, Suzuki T, A single acetylation of 18 S rRNA is essential for biogenesis of the small ribosomal subunit in Saccharomyces cerevisiae. *Journal of Biological Chemistry* **2014,** *289*, 26201-26212.
21. Safra, M.; Sas-Chen, A.; Nir, R.; Winkler, R.; Nachshon, A.; Bar-Yaacov, D.; Erlacher, M.; Rossmanith, W.; Stern-Ginossar, N.; Schwartz, S., The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **2017,** *551* (7679), 251-255.
22. Legrand, C.; Tuorto, F.; Hartmann, M.; Liebers, R.; Jacob, D.; Helm, M.; Lyko, F., Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res* **2017,** *27* (9), 1589-1596.
23. Edelheit, S.; Schwartz, S.; Mumbach, M. R.; Wurtzel, O.; Sorek, R., Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* **2013,** *9* (6), e1003602.
24. Maraia, R. J.; Arimbasseri, A. G., Factors That Shape Eukaryotic tRNAomes: Processing, Modification and Anticodon-Codon Use. *Biomolecules* **2017,** *7* (1).
25. Sloan, K. E.; Warda, A. S.; Sharma, S.; Entian, K. D.; Lafontaine, D. L. J.; Bohnsack, M. T., Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **2017,** *14* (9), 1138-1152.
26. Hoernes, T. P.; Huttenhofer, A.; Erlacher, M. D., mRNA modifications: Dynamic regulators of gene expression? *RNA Biol* **2016,** *13* (9), 760-5.
27. Helm, M., Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res* **2006,** *34* (2), 721-33.
28. Faure, G.; Ogurtsov, A. Y.; Shabalina, S. A.; Koonin, E. V., Role of mRNA structure in the control of protein folding. *Nucleic Acids Res* **2016,** *44* (22), 10898-10911.
29. Kennedy, T. D.; Lane, B. G., Wheat embryo ribonucleates. XIII. Methyl-substituted nucleoside constituents and 5'-terminal dinucleotide sequences in bulk poly(AR)-rich RNA from imbibing wheat embryos. *Can J Biochem* **1979,** *57* (6), 927-31.
30. Deng, X.; Chen, K.; Luo, G. Z.; Weng, X.; Ji, Q.; Zhou, T.; He, C., Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res* **2015,** *43* (13), 6557-67.
31. Horowitz, S.; Horowitz, A.; Nilsen, T. W.; Munns, T. W.; Rottman, F. M., Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci U S A* **1984,** *81* (18), 5667-71.

32. Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S. R.; Qian, S. B., Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **2015,** *526* (7574), 591-4.
33. Fustin, J. M.; Doi, M.; Yamaguchi, Y.; Hida, H.; Nishimura, S.; Yoshida, M.; Isagawa, T.; Morioka, M. S.; Kakeya, H.; Manabe, I.; Okamura, H., RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* **2013,** *155* (4), 793-806.
34. Batista, P. J.; Molinie, B.; Wang, J.; Qu, K.; Zhang, J.; Li, L.; Bouley, D. M.; Lujan, E.; Haddad, B.; Daneshvar, K.; Carter, A. C.; Flynn, R. A.; Zhou, C.; Lim, K. S.; Dedon, P.; Wernig, M.; Mullen, A. C.; Xing, Y.; Giallourakis, C. C.; Chang, H. Y., m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **2014,** *15* (6), 707-19.
35. Agarwala, S. D.; Blitzblau, H. G.; Hochwagen, A.; Fink, G. R., RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS Genet* **2012,** *8* (6), e1002732.
36. Huber, S. M.; van Delft, P.; Mendil, L.; Bachman, M.; Smollett, K.; Werner, F.; Miska, E. A.; Balasubramanian, S., Formation and abundance of 5-hydroxymethylcytosine in RNA. *Chembiochem* **2015,** *16* (5), 752-5.
37. Wang R, L. Z., He K, Delaney MO, Chen D, Sheng J, Base pairing and structural insights into the 5-formylcytosine in RNA duplex. *Nucleic Acids Research* **2016,** *44* (10), 4968-4977.

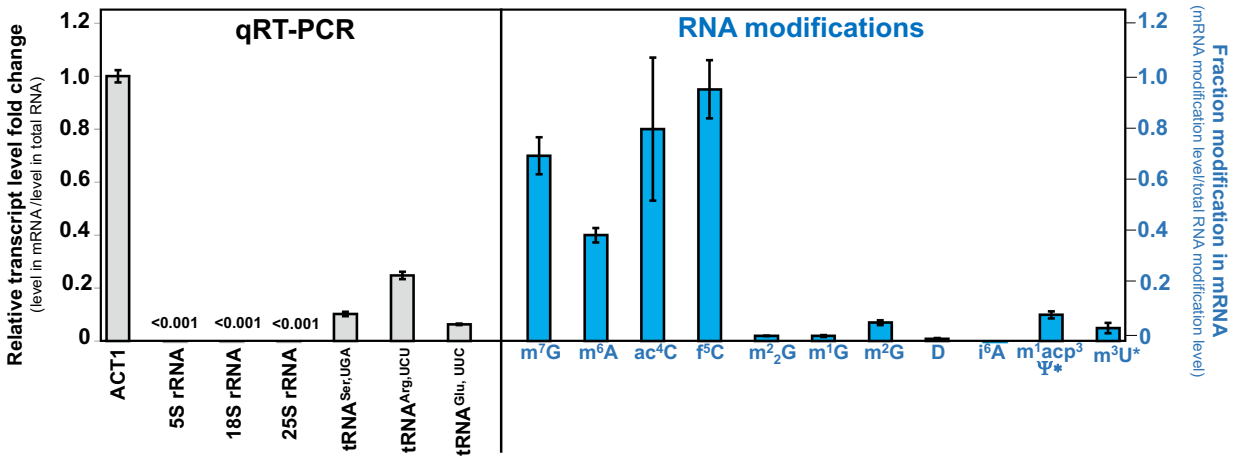


Figure 1: Non-coding RNAs are depleted in mRNA samples, and mRNA modifications are retained. qRT-PCR demonstrates that the levels of rRNAs and tRNAs in our mRNA samples are very low (left panel - gray bars, black y-axis). UHPLC-MS/MS reveals that the levels of m^7G, m^6A, ac^4C and ^{f5}C are high in mRNAs, while most others are depleted (blue bars and y-axis). *indicates modifications that are only present in rRNA: m^1acp^3Y (18S rRNA) and m^3U (25S rRNA).

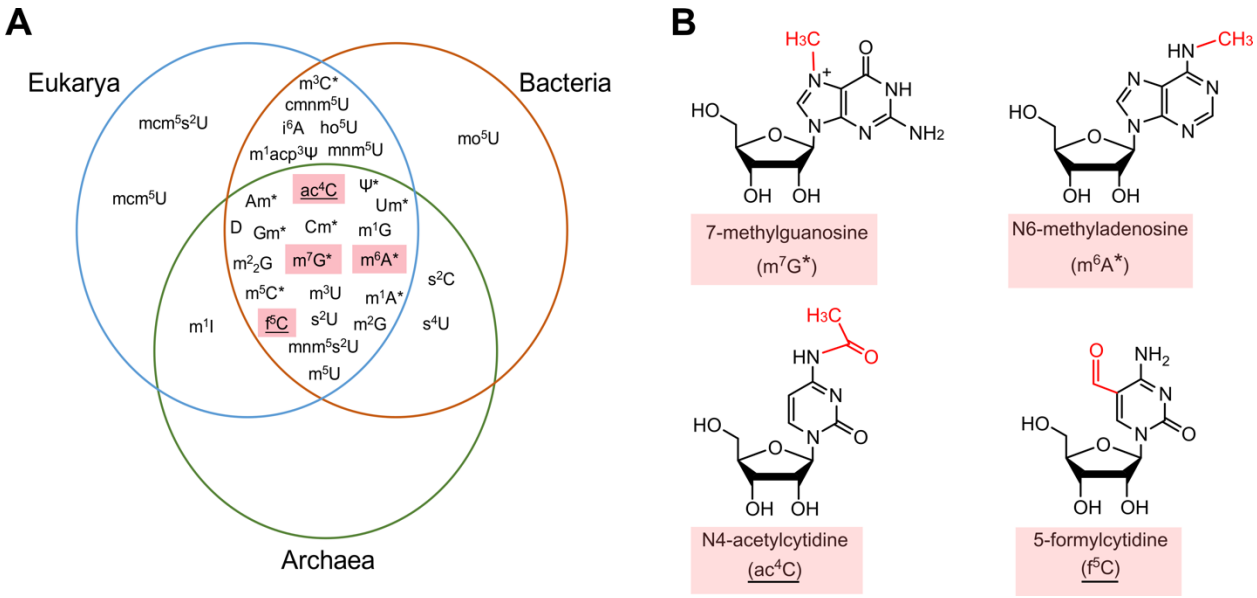


Figure 2. Identification of modifications in mRNAs. A) Phylogenetic distribution of modified nucleosides observed in yeast RNAs in our study. Nucleosides we found in mRNAs are highlighted in red. Asterisks indicate previously known mRNA modifications, and underlining denotes new mRNA modifications identified found in our study. **B)** Structures of modifications we identify in mRNAs.

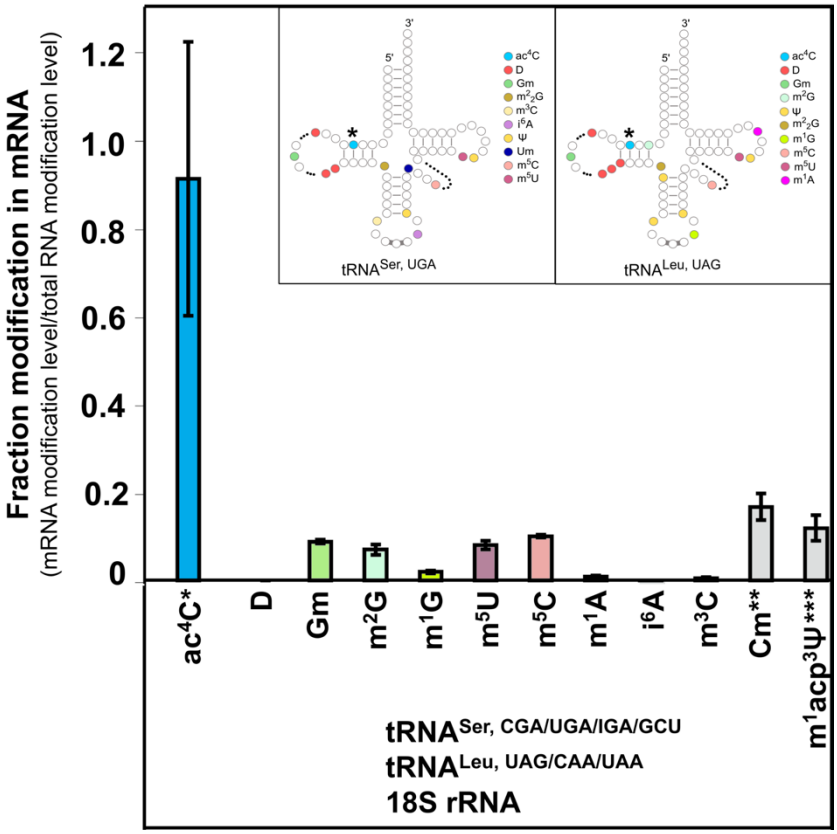


Figure 3: Other modifications in non-coding RNAs where ac^{4}C is found are not retained in mRNAs. We plot the level of modified nucleoside in our mRNA sample relative to the modification concentration present in total RNA. The plot shows the levels of ac^{4}C (blue) and modifications on non-coding RNAs that contain ac^{4}C (gray) retained in our mRNA samples. The calculation of the 'fraction modification in mRNA' is described in materials and methods. Together, our qRT-PCR and mass-spectrometry analyses support the idea that our mRNA is sufficiently pure, and we have not isolated tRNAs or tRNA degradation fragments. Please note that Gm is present also within nine nucleotides of ac^{4}C in 18S rRNA. * denotes the location of ac^{4}C on tRNA$^{Ser,\ UGA}$ and tRNA$^{Leu,\ UAG}$. **Cm is found once in tRNA^{Leu,UAA} and three times in 18S rRNA. ***m^{1}acp$^3\Psi$ is not present in tRNAs but is found once in 18S rRNA.

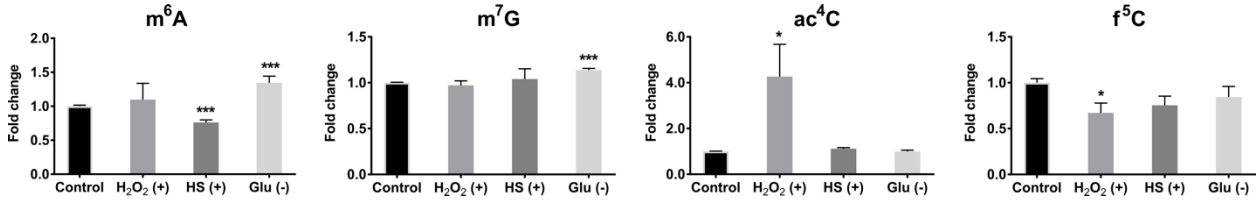


Figure 4: **Post-transcriptional modifications in mRNA of *S. cerevisiae* exhibit differential responses to environmental stressors.** The fold change for each modification upon stress induction was calculated as the ratio of nucleoside level in the mRNA sample of stress-treated condition ((H2O2 (+), heat-shock (HS) (+) and glucose (Glu) (-)) and nucleoside level in the mRNA-enrichment 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.

Supporting Information

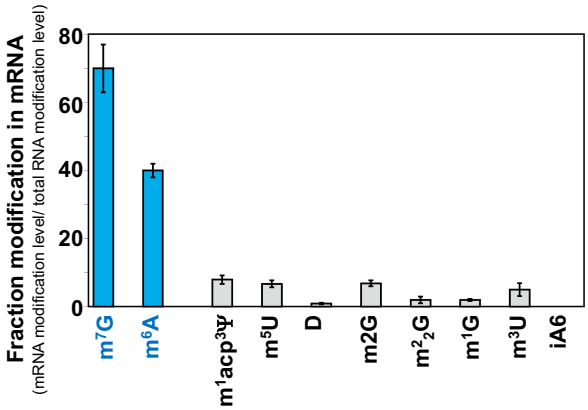
Establishing the cut-off for designating modifications as present in mRNA

We set our cut-off for mRNA modification retention and background level based on four criteria: 1) retention of non-coding modifications in mRNA samples, 2) BioAnalyzer data, 3) concentration of modification, and 4) qRT-PCR for rRNAs and tRNAs.

Criterion 1: Retention of modifications in mRNA samples

We determined the fraction of each modification retained in our mRNA samples collected from all experimental conditions. The fraction of nucleoside retained in mRNAs = $\text{nucleoside}_{\text{normalized,mRNA}}/\text{nucleoside}_{\text{normalized,totalRNA}}$, as described in materials and methods. The error on each of these fractions was propagated, using the technical replicates of all biological replicates (error = ((standard error on mRNA measurements)^2+(standard error on total RNA measurements)^2)$^{1/2}$).

We then used the fraction of modification retained in mRNAs of eight nucleosides to quantify the upper limit for signal derived from non-coding species (**Figure 1B and Supporting Information Figure** below): 3-methyluridine (m^3U), 5-methyluridine (m^5U), 2,2-dimethylguanosine ($m^2{}_2G$), 1methyl-inosine (^1mI), dihydrouridine (D), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine ($m^1acp^3\Psi$), N6-isopentyladenosine (iA6), and 2-methylguanosine (m^2G). This assay takes into account signal that originates from modifications in full-length RNAs, processing intermediates, or degradation products. We selected these eight nucleosides because they had not been reported in mRNAs and they represent a broad variety of modifications in terms of distribution and concentration in *S. cerevisiae*: m^3U (two in 25S rRNA), $m^1acp^3\Psi$ (one in 18S rRNA) $m^2{}_2G$ (one in $tRNA^{Phe,GAA}$), ^1mI ($tRNA^{Ala,ACG}$), D (in all tRNAs), ^5mU (in all tRNAs), iA6 (in 8 tRNAs), and m^2G (present once in the



Supporting Information Figure. Fraction of known mRNA (blue) and 8 non-coding (gray) modifications retained in mRNA samples from total RNA.

majority of tRNAs)[1]. This analysis indicates the background level of signal is 0-10%.

Criterion 2: BioAnalyzer

BioAnalyzer results (**Supplementary Figure S1**) indicate that our samples lack small RNAs (snRNAs and tRNAs), and that rRNAs are retained up to 2-5% in our mRNA samples.

Criterion 3: Concentration of modification in mRNA

We calculated the concentration of each modification per nucleoside and per mRNA (**Supplementary Table S4**, and materials and methods). We only evaluated the levels of modifications that we identified as retained mRNAs using criteria 1-2 under all experimental conditions. There were two clear groups of modifications: those whose concentration was within 2-fold of the well-established m^6A and m^7G modifications, and a second group that as present at concentrations 100-fold less than m^6A and m^7G. We only annotated the modifications in the first group (f^5C, mnm^5U, and ac^4C), as being present above background levels.

Criterion 4: qRT-PCR

We performed qRT-PCR to check the levels of $tRNA^{Arg,UCU}$, $tRNA^{Glu,UUC}$, $tRNA^{Ser,UGA}$, 5S rRNA, 18S rRNA, and 25S rRNA using the primers shown in Supplementary Table S4. These data show that < 0.1% of rRNAs are retained in our mRNA samples.

We find low levels of $tRNA^{Ser,UGA}$, $tRNA^{Glu,UUC}$, $tRNA^{Arg,UCU}$ present in our samples (10%, 6% and 25%, respectively, of their levels in total RNA) **(Figure 1)**. These data are consistent with our BioAnalyzer data suggesting that the samples are very depleted for tRNAs in general.

Criteria 1-3 suggest that cut-off of 10% retention in mRNA should be a sufficient to categorize a modification as present in mRNA. Indeed, if we use 10% cut-off we would capture nearly all of the modifications previously identified in mRNAs: 2'Omethyl modifications (Cm, Um, Gm, Am), m^5C, m^6A, m^7G. Ψ is present at levels just below 10%. We do not see evidence for m^3C and m^1A, as these modifications are retained < 2 % in

mRNAs. It is of note that we also would have annotated s^2C as an mRNA modification using these three criteria.

Criteria 1-3 were combined with the qRT PCR data (criterion 4) to designate a 20% retention of a modification in mRNA AND an estimated occurrence of the modifications once per every 1-10 mRNAs. Please note, that even with a 25% retention of $tRNA^{Arg,\ UCU}$, **criteria 1-3 already would have excluded us from designating any of the modifications in $tRNA^{Arg,\ UCU}$ as present in mRNAs** (m^1G, m^2G, Ψ, D, mcm^5U, m^5U, m^1A). Note that while $tRNA^{Arg,UCU}$ is depleted to 25% of its level in total RNA, the levels of its known modified nucleotides are at (m^1G, m^2G, Ψ, D, mcm^5U, m^5U, m^1A are retained at 0.9-7% and excluded by criteria 1-2, and mcm^5U is retained at 80%, but found in trace concentrations less than once per every 250 mRNAs – thereby excluding it by criterion 3). The final 20% cut-off excludes all previously identified mRNA modifications except m^6A and m^7G. Since our filters identify only modifications present at the same level as the cap nucleotide m^7G and the abundant m6A, we have high confidence that the novel modifications ac^4C and f^5C are genuine. We speculate that however that some genuine mRNA modifications are excluded by setting the bar so high (e.g. 2'OMe modifications, Ψ, and s^2C).

Growth conditions and stress experiments

Saccharomyces cerevisiae (BY4741) 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.

Total RNA Extraction, mRNA Enrichment and qRT-PCR

Total RNA was extracted from 10 ml of yeast cells (OD_{600} = 0.6) using standard hot acidic phenol method[53]. 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 28S 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. 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.

qRT-PCR was performed to measure the mRNA levels of *CCT1*, *HSP30* and *HXT2* genes at different time points for H_2O_2 (+), heat-shock (+) and glucose (−) conditions, respectively to verify that stress was induced under each condition (**Supplementary Figure S3**). The qRT-PCRs were performed with Luminaris HiGreen qRT-PCR Master Mix (Thermo Scientific, USA) on StepOnePlus Real-Time PCR System (Applied Biosystems, USA) using gene-specific primers (**Supplementary Table S6**), with *ACT1* (YFL039C) 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, $\Delta\Delta Ct = \Delta Ct$ (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 = $2^{-(\Delta\Delta Ct)}$).

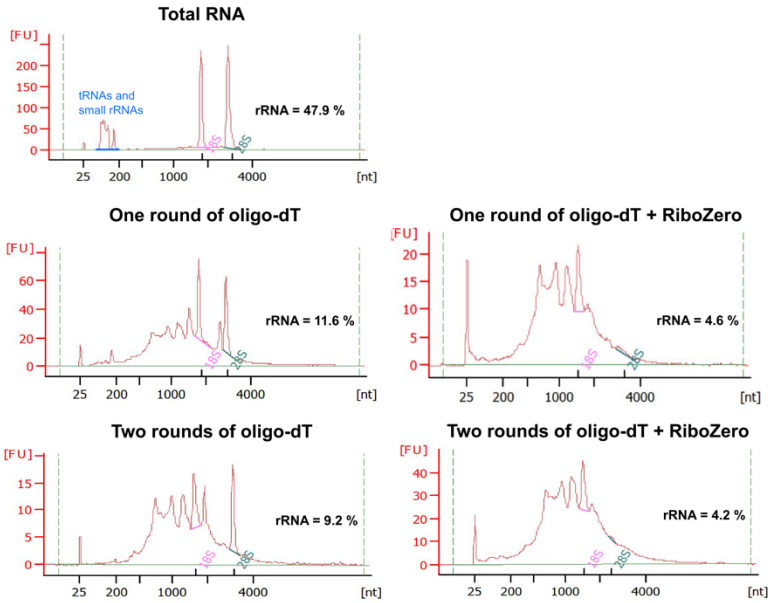
UHPLC-MS/MS Analysis

Briefly, RNAs were first hydrolyzed to the composite mononucleosides via a two-step enzymatic hydrolysis. 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. To quantify RNA modified nucleosides, calibration curves were prepared for 42 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 for quantification of RNA modifications (quantifications given in **Supplementary Table S3)**. Python script / Production of calibration curves as well as quantification from samples was produced in Originlab software suite 2017.

Supplementary Figures (Figure S1, Figure S2 and Figure S3)

A


Total RNA
tRNAs and small rRNAs
rRNA = 47.9 %
One round of oligo-dT
rRNA = 11.6 %
One round of oligo-dT + RiboZero
rRNA = 4.6 %
Two rounds of oligo-dT
rRNA = 9.2 %
Two rounds of oligo-dT + RiboZero
rRNA = 4.2 %
[FU]
[nt]
25
200
1000
4000
18S
28S

B

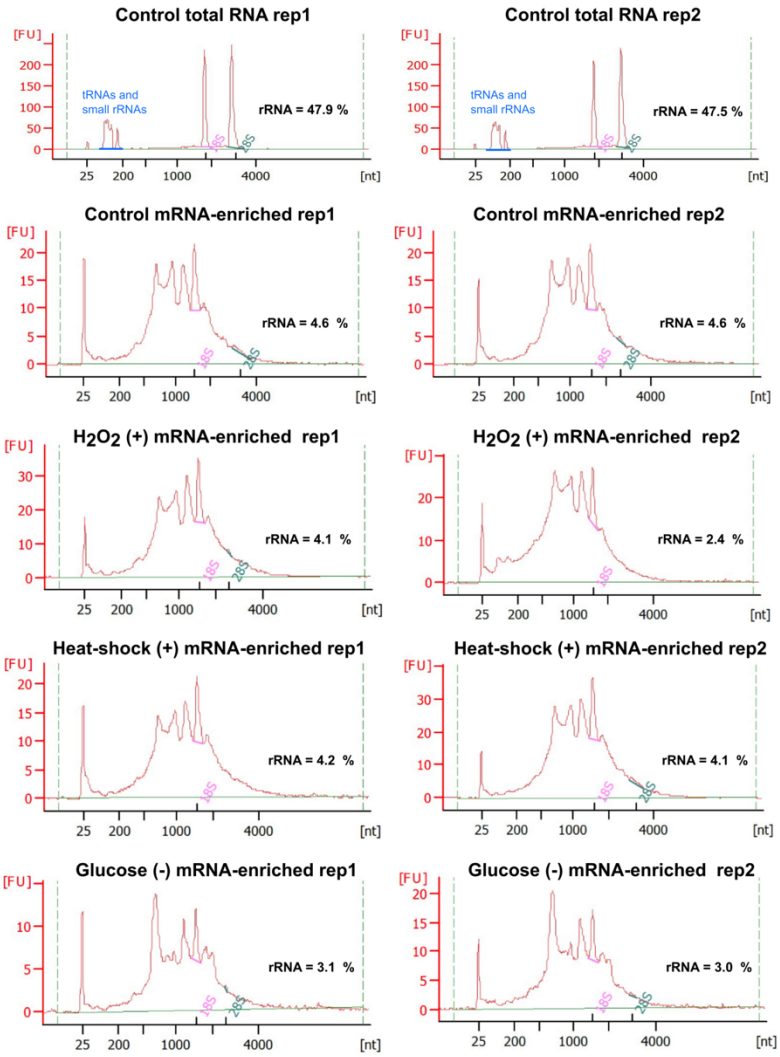

Control total RNA rep1
tRNAs and small rRNAs
rRNA = 47.9 %
Control total RNA rep2
tRNAs and small rRNAs
rRNA = 47.5 %
Control mRNA-enriched rep1
rRNA = 4.6 %
Control mRNA-enriched rep2
rRNA = 4.6 %
H2O2 (+) mRNA-enriched rep1
rRNA = 4.1 %
H2O2 (+) mRNA-enriched rep2
rRNA = 2.4 %
Heat-shock (+) mRNA-enriched rep1
rRNA = 4.2 %
Heat-shock (+) mRNA-enriched rep2
rRNA = 4.1 %
Glucose (-) mRNA-enriched rep1
rRNA = 3.1 %
Glucose (-) mRNA-enriched rep2
rRNA = 3.0 %
[FU]
[nt]
25
200
1000
4000
18S
28S

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-downa 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 28S rRNAs decreased from ~200 FU to ~20 FU after mRNA enrichment protocol. The rRNA contamination percentage was calculated using Bioanalyzer software and they are included in plots for each individual sample.

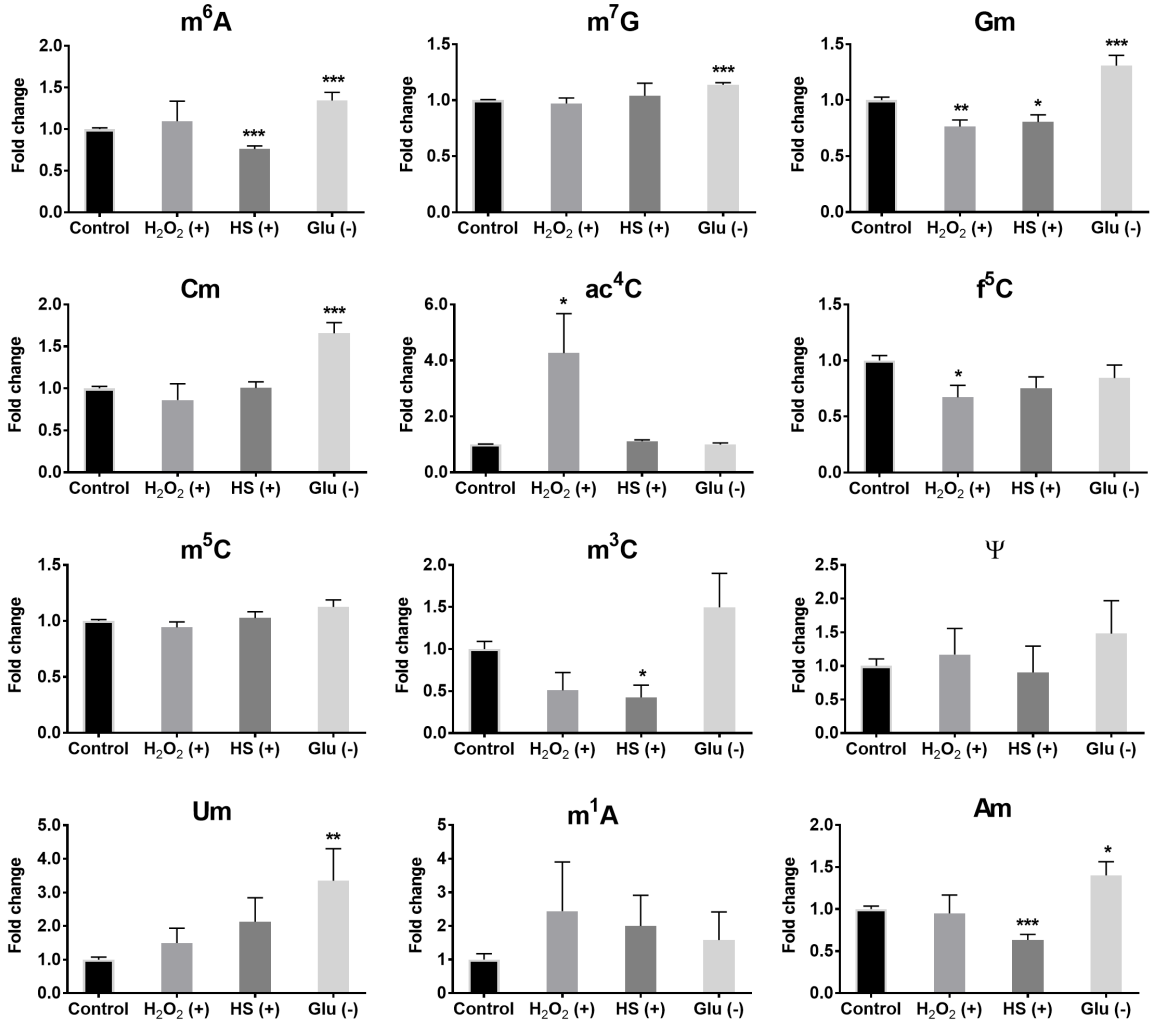


Figure S2: **Post-transcriptional modifications in mRNA of *S. cerevisiae* exhibit differential responses to environmental stressors.** The fold change for all modifications identified either here or previously upon stress induction was calculated as the ratio of nucleoside level in the mRNA-enriched sample of stress-treated condition ((H2O2 (+), heat-shock (HS) (+) and glucose (Glu) (-)) and nucleoside level in the mRNA-enrichment 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.

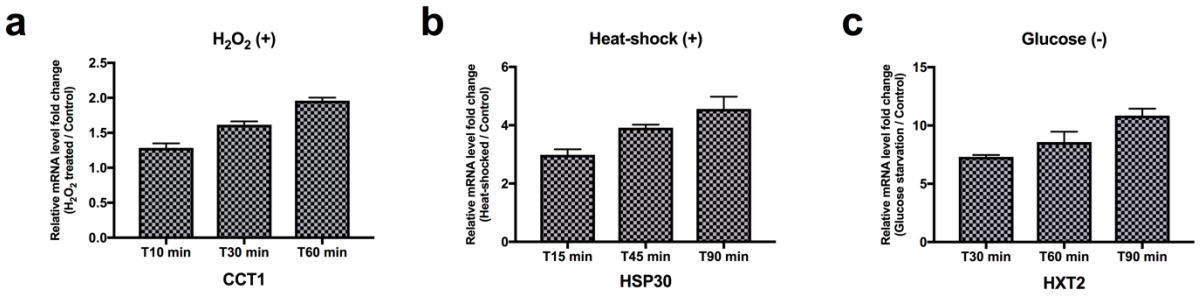


Figure S3: Validation of stress induction protocols by qRT-PCR on known stress response genes. Stress induction for **A)** oxidative-stress **B)** heat-shock stress and **C)** glucose-starvation conditions was confirmed by measuring the mRNA levels of stress-specific *CCT1*, *HSP30* and *HXT2* genes, 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.