Decoding ribosomal RNA modification dynamics at single molecule resolution

Oguzhan Begik[1,2,3,#], Morghan C Lucas[1,4,#], Jose Miguel Ramirez[1], Ivan Milenkovic[1,4], Sonia Cruciani[1,4], Helaine Graziele Santos Vieira[1], Rebeca Medina[1], Huanle Liu[1], Aldema Sas-Chen[5], John S Mattick[3], Schraga Schwartz[5] and Eva Maria Novoa[1,2,3,4,6*]

[1]*Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain*

[2]*Garvan Institute of Medical Research, Darlinghurst, NSW, 2010, Australia*

[3]*UNSW Sydney, Darlinghurst, NSW, 2052, Australia*

[4]*Universitat Pompeu Fabra (UPF), Barcelona, Spain*

[5] *Weizmann Institute of Science, Rehovot, IL*

[6] *Lead Contact*

[#] *These authors contributed equally*

* Correspondence to: Eva Maria Novoa (eva.novoa@crg.eu)

SUMMARY

A broad diversity of modifications decorate RNA molecules. Originally conceived as static components, evidence is accumulating that some RNA modifications may be dynamic, contributing to cellular responses to external signals and environmental circumstances. A major difficulty in studying these modifications, however, is the need of tailored protocols to map each modification individually. Here, we present a new approach that uses direct RNA nanopore sequencing to identify diverse RNA modification types present in native RNA molecules, using rRNA as the exemplar, and show that each RNA modification type results in distinct and characteristic base-calling 'error' signatures. We demonstrate the value of these signatures for *de novo* prediction of pseudouridine (Y) modifications transcriptome-wide, confirming known Y modifications in rRNAs, snRNAs and mRNAs, and uncovering a novel Pus4-dependent Y modification in yeast mitochondrial rRNA. Using a machine learning classifier, we show that the stoichiometry of modified sites can be quantified by identifying current intensity alterations in individual RNA reads. Finally, we explore the dynamics of pseudouridylation across a battery of environmental stresses, revealing novel heat-sensitive Y-modified sites in both snRNAs and snoRNAs. Altogether, our work demonstrates that Y RNA modifications can be predicted *de novo* and in a quantitative manner using native RNA nanopore sequencing.

Keywords : Ribosomal RNA, Epitranscriptome, RNA Modifications, Pseudouridylation, Nanopore, Direct RNA Sequencing, *Saccharomyces cerevisiae*

INTRODUCTION

RNA modifications are chemical moieties that decorate RNA molecules, expanding their lexicon. By coupling antibody immunoprecipitation or chemical probing with next-generation sequencing (NGS), transcriptome-wide maps of several RNA modifications have been constructed, including N6-methyladenosine (m^6A) (Dominissini et al., 2012; Meyer et al., 2012), pseudouridine (Y) (Carlile et al., 2014; Li et al., 2015; Lovejoy et al., 2014; Schwartz et al., 2014), 5-methylcytosine (m^5C) (Huang et al., 2019; Hussain et al., 2013), 5-hydroxymethylcytosine (hm^5C) (Delatte et al., 2016), 1-methyladenosine (m^1A) (Li et al., 2017; Safra et al., 2017), N3-methylcytosine (m^3C) (Marchand et al., 2018), N4-acetylcytosine (ac^4C) (Arango et al., 2018; Sas-Chen et al., 2020) and 7-methylguanosine (m^7G) (Pandolfini et al., 2019; Zhang et al., 2019). These studies have revealed that RNA modifications play a pivotal role in a large variety of cellular processes, including regulation of cellular fate (Delaunay and Frye, 2019), sex determination (Haussmann et al., 2016) and cellular differentiation (Vu et al., 2017), among others.

Despite these advances, a fundamental challenge in the field is the lack of a generic approach for mapping diverse RNA modification types simultaneously (Anreiter et al., 2020; Li et al., 2016; Motorin and Helm, 2019; Novoa et al., 2017). Currently, customized protocols must be individually set up and optimized for each RNA modification type, leading to experimental designs in which the RNA modification type to be studied is chosen beforehand, hindering the ability to characterize the plasticity of the epitranscriptome in a systematic and unbiased manner in response to different conditions. Moreover, even in those cases where a selective antibody or chemical is available, NGS-based methods are often not quantitative (i.e. cannot solve the 'stoichiometry' problem), are unable to produce maps for highly repetitive regions, cannot provide information regarding the co-occurrence of distant modifications in same transcripts, do not provide isoform-specific information, and require multiple ligations steps and extensive PCR amplification during the library preparation, introducing undesired biases in the sequencing data (Lahens et al., 2014).

A promising alternative to NGS-based technologies that can, in principle, overcome these limitations is the direct RNA sequencing platform developed by Oxford Nanopore Technologies (ONT), which has the potential to detect virtually any given RNA modification in native RNA molecules (Garalde et al., 2018; Jonkhout et al., 2017; Novoa et al., 2017). Algorithms to detect RNA modifications have been made available in the last few months (Leger et al., 2019; Liu et al., 2019; Parker et al., 2020), many of which rely on the use of systematic base-calling 'errors' caused by the presence of RNA modifications. However, to date the vast majority of efforts have been devoted to the detection of m^6A modifications (Liu et al., 2019; Parker et al., 2020; Pratanwanich et al., 2020; Price et al., 2019). Thus, it is largely unknown whether other modifications of RNA bases may be distinguishable from their unmodified counterparts using this technology.

In contrast to mRNAs, ribosomal RNA (rRNA) molecules are extensively modified, stabilizing the secondary and tertiary structure of the rRNA scaffold (Polikanov et al., 2015). It has been proposed

that differential rRNA modifications may constitute a source of ribosomal heterogeneity (Bellodi et al., 2010; Jack et al., 2011; Yoon et al., 2006), leading to fine tuning of the ribosomal function and ultimately proteome output. Previous studies have shown that temperature changes affect rRNA pseudouridylation levels at specific sites, suggesting that cells may be able to generate compositionally distinct ribosomes in response to environmental cues (Schwartz et al., 2014; Taoka et al., 2015; Wu et al., 2011). Similarly, alterations in the stoichiometry of 2'-O-methylation (Am, Cm, Gm, Um) (Basu et al., 2011; Belin et al., 2009; Marcel et al., 2013) and pseudouridylation (Y) (Bellodi et al., 2010; Jack et al., 2011; Yoon et al., 2006) can affect translation initiation of mRNAs containing internal ribosome entry sites (IRES) (Buchhaupt et al., 2014; Chen et al., 2020). While these studies have provided valuable insights into the function and dynamics of rRNA modifications, a systematic, multiplexed and unbiased approach that can map and quantify all rRNA modifications simultaneously in full-length molecules is currently lacking.

Here, we map the *S. cerevisiae* ribosomal epitranscriptome at single molecule resolution using native RNA nanopore sequencing. We find that most RNA modifications are characterized by systematic base-calling errors, and that the signature of these base-calling 'errors' can be used to identify the underlying RNA modification type. For example, pseudouridine typically appears in the form of U-to-C mismatches, whereas m^5C modifications appear in the form of insertions. We then exploit the identified signatures to *de novo* predict RNA modifications in rRNAs, finding a previously undescribed Y854 site in mitochondrial rRNA, which we confirmed using CMC-probing coupled to nanopore sequencing (nanoCMC-seq). We demonstrate that the modification at this novel site is placed by the enzyme Pus4, which was previously thought to pseudouridylate only mRNAs and tRNAs (Schwartz et al., 2014). Moreover, we show that once the Y RNA modifications have been accurately predicted using base-calling 'errors', the stoichiometry of a given Y-modified site can be estimated by clustering per-read current intensities of the modified region.

Finally, we characterize the epitranscriptome dynamics in rRNAs, snRNAs and snoRNAs upon a battery of environmental cues, translational repertoires and genetic strains. Contrary to expectations, we find that none of the environmental stresses tested lead to significant changes in the ribosomal epitranscriptome. By contrast, our method does recapitulate previously reported heat-dependent Y snRNA modifications, as well as identifies novel heat-sensitive sites. Altogether, our work establishes a framework for the study of RNA modifications using direct RNA sequencing, opening novel avenues to study the plasticity of the epitranscriptome at single molecule resolution.

RESULTS

Detection of RNA modifications in direct RNA sequencing data is strongly dependent on base-calling and mapping algorithms

Previous work has shown that N6-methyladenosine (m^6A) RNA modifications can be detected in the form of non-random base-calling 'errors' in direct RNA sequencing datasets (Liu et al., 2019; Parker et al., 2020; Price et al., 2019). However, it is unclear how these 'errors' vary with the choice of base-calling and mapping algorithms, and consequently, affect the ability to detect and identify RNA modifications. To systematically determine the accuracy of commonly used algorithms for direct RNA base-calling, as well as to assess their ability to detect RNA modifications in the form of base-calling 'errors' (Liu et al., 2019), we compared their performance on *in vitro* transcribed RNA sequences which contained all possible combinations of 5-mers, referred to as 'curlcakes' (CCs) (Liu et al., 2019), that included: (i) unmodified nucleosides (UNM), (ii) N6-methyladenosine (m^6A), (iii) pseudouridine (Y), (iv) N5-methylcytosine (m^5C), and (v) N5-hydroxymethylcytosine (hm^5C) (**Figure 1A**). In addition, a sixth dataset containing unmodified short RNAs (UNM-S), with median length of 200 nucleotides, was included in the analysis to assess the effect of input sequence length in base-calling (see *Methods*). Each dataset was base-called with two distinct algorithms (*Albacore* and *Guppy*), and using two different versions for each of them, namely: (i) *Albacore* version 2.1.7 (AL 2.1.7); (ii) its latest version, *Albacore* 2.3.4 (AL 2.3.4); (iii) *Guppy* 2.3.1 (GU 2.3.1); and (iv) a more recent version of the latter base-caller, *Guppy* 3.0.3 (GU 3.0.3), which employs a flip-flop algorithm. We found that the latest version of *Albacore* (2.3.4) base-called 100% of sequenced reads in all 6 datasets, whereas its previous version did not (average of 90.8%) (**Figure 1B**). In contrast, both versions of *Guppy* (2.3.1 and 3.0.3) produced similar results in terms of percentage of base-called reads (98.71% and 98.75%, respectively) (**Table S1**).

We then assessed whether the choice of mapper might affect the ability to detect RNA modifications. To this end, we employed two commonly used long-read mappers, *minimap2* (Li, 2018) and *GraphMap* (Sović et al., 2016), using either 'default' or 'sensitive' parameter settings (see *Methods* for detailed parameter settings used). Strikingly, we found that the choice of mapper, as well as the parameters used, severely affected the final number of mapped reads for each dataset (**Figure 1C,** see also **Table S1**). The most extreme case was observed with the Y-modified dataset, where *minimap2* was unable to map the majority of the reads (0-0.3% mapped reads) (**Figure 1C,D,** see also **Figure S1**). By contrast, *GraphMap* 'sensitive' was able to map 35.5% of Y-modified base-called reads, proving to be a more appropriate choice for highly modified datasets. To ascertain whether an increase in the number of base-called and mapped reads was at the expense of decreased accuracy, we assessed the sequence identity percentage (as a read-out of accuracy), finding that *GraphMap* outperforms *minimap2* with only a minor loss in accuracy (3%) (**Figure 1E,** see also **Table S2**).

Base-calling 'error' signatures can be used to predict RNA modification type, but these may vary with the sequence context

Base-calling 'errors' can be used to identify m^6A RNA modified sites (Liu et al., 2019; Parker et al., 2020; Wongsurawat et al., 2018). However, whether this approach is applicable for the detection of other RNA modifications is largely unknown. To this end, we systematically characterized the base-calling errors caused by the presence of m^6A, Y, m^5C and hm^5C. We found that, regardless of the base-caller and mapper settings used, modified RNA sequences presented higher mismatch frequencies (**Figure 2A**) and decreased quality scores (**Figure S2A,B**). In addition, these differences were more prominent in Y-modified datasets. Principal component analysis of base-calling 'errors' of each modified dataset (m^6A,Y, m^5C and hm^5C) -relative to unmodified- showed that this difference was greatest in Y-modified datasets (**Figure 2B**), and maximized in datasets that were base-called with GU 3.0.3. Thus, we find that all four RNA modifications can be detected in direct RNA sequencing data; however, their detection is severely affected by the choice of both base-calling and mapping algorithms, and varies depending on the RNA modification type.

We then examined whether the base-called 'errors' observed in modified and unmodified datasets occured in the modified position. We found that both m^6A and Y modifications led to increased mismatch frequencies at the modified site (**Figure 2C**), mainly in the form of U-to-C mismatches in the case of Y modifications (**Figure S2C**). By contrast, m^5C and hm^5C modifications did not appear in the form of increased mismatch frequencies at the modified site; rather, these modifications appeared in the form of increased mismatch frequencies in the neighboring residues (position -1 and +1 in the case of m^5C modifications; position +1 in hm^5C) (**Figure 2C**) Moreover, the observed base-called 'error' signatures of m^5C and hm^5C were also dependent on the sequence context (**Figure S2D**). Altogether, we find that all four RNA modifications studied (m^6A, m^5C, hm^5C and Y) lead to altered base-called features, and that their 'error' patterns vary depending both on the RNA modification type as well as on the sequence context.

Y modifications can be detected *in vivo,* in the form of U-to-C mismatches and with single nucleotide resolution

We then examined whether the results obtained using *in vitro* transcribed constructs would be applicable to *in vivo* RNA sequences. To this end, total RNA from *S. cerevisiae* was poly(A)-tailed to allow for ligation between the RNA molecules and the commercial ONT adapters, and then prepared for direct RNA sequencing. The reverse transcription step was performed using Maxima reverse transcriptase instead of SuperScript III, which can be used at higher temperatures, thus ensuring proper linearization of the RNA molecules (see *Methods*). Visual inspection of the mapped reads revealed that our approach captured a high proportion of full length rRNA molecules, with a high proportion of base-calling errors present in 25s and 18s rRNAs, as could be expected from sequences that are highly enriched in RNA modifications (**Figure 3A**). By contrast, 5s and 5.8s rRNAs did not show such base-calling errors, in agreement with their low level of modification.

Then, we systematically analyzed base-called features (mismatch, deletion, insertion and per-base qualities) in rRNAs, comparing the features from rRNA modified sites relative to unmodified ones (**Figure 3B**). We found that all rRNA modification types consistently led to decreased per-base qualities at modified sites, suggesting that per-base qualities can be employed to identify RNA modifications, but not the underlying RNA modification type. Moreover, we found that Y modifications caused significant variations in mismatch frequencies, in agreement with our observations using *in vitro* constructs. By contrast, other RNA modifications, such as 2'-O-methylcytidine (Cm) or 5-methylcytosine (m^5C) did not appear in the form of increased mismatch frequencies at modified sites, but rather, in the form of increased insertions. In addition, Y modifications typically appeared in the form of U-to-C mismatches (**Figure 3C,** see also **Figure S3**), in agreement with our *in vitro* observations, whereas other RNA modifications such as 2'-O-methyladenosine (Am) did not cause mismatches with unique directionality. Thus, we conclude that distinct rRNA modification types can be detected in the form altered base-called features *in vivo*, and that their base-calling 'error' signature is dependent on the RNA modification type.

To confirm that the detected signal (U-to-C mismatches) in Y positions is caused by the presence of the Y modification, we compared ribosomal RNA modification profiles from wild type *S. cerevisiae* to those from snoRNA-knockout strains (snR3, snR34 and snR36), which lack Y modifications at known rRNA positions (**Figure 4A,** see also **Table S3**). Our results show that changes in rRNA modification profiles were consistently and exclusively observed in those positions reported as targets of each snoRNA. Moreover, the remaining Y-modified positions were not significantly altered by the lack of Y modifications guided by snR3, snR34 or snR36 (**Figure 4B**), suggesting that the modification status of Y sites is largely independent from other Y sites, at least in the cases examined.

2'-O-methylations can be detected *in vivo* in the form of systematic base-calling 'errors', but their signatures varies across sites

We then sequenced 3 additional *S. cerevisiae* strains depleted from snoRNAs (snR60, snR61 and snR62 knockouts) guiding 2'-O-methylation (Nm) at specific positions. In contrast to Y modifications, we found that 2'-O-methylations often caused increased mismatch and deletion signatures at neighboring positions (**Figure 4C,** see also **Figure S4A**), in a similar fashion to what we had previously observed with m^5C and hm^5C modifications (**Figure 2C**). These errors disappeared in the knockout strain, suggesting that neighboring base-calling errors were indeed caused by the 2'-O-methylation (**Figure 4C**, see also **Table S3**). In contrast to Y modifications, which mainly affected mismatch frequency, we observed that Nm modifications often affected several base-called 'error' features (mismatch, insertion and deletion frequency) (**Figure S4B**). Thus, we reasoned that combining all three features might improve the signal-to-noise ratio for the detection of 2'-O-methylated sites (**Figure 4D**), and found that the combination of features led to improved detection of Nm-modified sites, relative to each individual feature. We should note that position 25s:Gm908 was poorly detected both wild type and snoRNA-depleted strains (**Figure S4A,B**) regardless of the feature

combination used, likely due to the sequence context in which the site is embedded -an homopolymeric GGGG sequence-, which is often troublesome for nanopore base-calling algorithms.

Current intensity variations can be used to detect Y and Nm RNA modifications, but do not allow accurate prediction of modified site

We then wondered whether Y and Nm sites would also be detected at the level of current intensity changes. We observed that certain Nm and Y-modified sites, such as 25s:Y2880 and 25s:Am817, showed drastic alterations of their current intensity values in the snoRNA-depleted strain (**Figure 4E,F**, upper panels). By contrast, the distributions of current intensities in other sites did not significantly change in the knockout strain (18s:Y1187) or did not differ in their mean (25s:Y2133).

We hypothesized that deviations current intensity alterations might not always be maximal in the modified site, but might sometimes appear in neighbouring sites. To test this, we examined the difference in current intensity values along the rRNA molecules for each wild type-knockout pair (**Figure 5A**, see also **Figure S5A**). As expected, we found that the depletion of snR3 led to two regions with altered current intensity values along the 25s rRNA - one comprising the 25s:Y2129 and 25s:Y2133 sites, and the second comprising the 25s:Y2264 site-. However, the highest deviations in current intensity were not observed at the modified site (**Figure 5A,** lower panel)**.** From all 6 Y sites that were depleted in the 3 knockout strains studied, only 2 of them (25s:Y2826 and 25s:Y2880) showed a maximal deviation in current intensity in the modified site (**Figure 5B**). Similarly, depletion of Nm sites led to changes in current intensity values, but the largest deviations were not observed at the modified site (**Figure 5C**). Thus, we conclude that current intensity-based methods can detect both Y and Nm RNA modifications; however, base-calling errors are a better choice to achieve single nucleotide resolution, at least in the case of Y RNA modifications.

Stoichiometry of Y- and Nm-modified sites can be estimated from per-read current intensity alterations

Direct RNA sequencing produces current intensity measurements for each individual native RNA molecule. Thus, native RNA sequencing can in principle estimate modification stoichiometries by identifying the proportion of reads that contain alterations in their current intensity at a given site. However, this requires identifying RNA modifications in individual RNA molecules, which has proven so far a challenging task, largely due the low signal-to-noise ratio that single molecule nanopore reads have, compared to other sequencing platforms.

Rather than performing prediction of RNA modifications transcriptome-wide at a per-read level, we reasoned that a targeted approach might prove more useful to improve the signal-to-noise ratio. Considering that Y RNA modifications can be detected in the form of base-calling ‘errors’ with single nucleotide resolution, we reasoned that stoichiometry could be predicted at per-read level by using the current intensity information of the 15-mer sequences surrounding the modified sites. We first examined the per-read current intensity values of wild type and knockout strains at the depleted sites.

Although we found that there was a significant variability across reads -even when 100% of the positions are unmodified- we were able to observe robust differences in current intensities across strains at the per-read level (**Figure 5D,E,** see also **Figure S5B**). As a control, we performed the same analysis in Y sites unaffected by snoRNA depletion, finding no differences between wild type and knockout strains at these positions (**Figure 5D**, bottom panel).

We then performed Principal Component Analysis (PCA) of the current intensity values corresponding to the 15-mer regions that contained the modified site (**Figure 5D,E,** right panels), for all snoRNA-depleted strains affecting Y modifications (snR3, snR34, snR36) and Nm modifications (snR60, snR61, snR62) as well as for the wild type strain. We observed that the reads clustered into two distinct populations: the first cluster mainly comprised unmodified reads from the snoRNA-depleted strain, whereas the second comprised reads from the 3 other strains, which are mostly modified. Thus, the presence or absence of Y or Nm modification in a read was a sufficiently robust and distinct feature at the level of current intensity to bin the reads into two separate populations: modified and unmodified. To follow these findings, we built a k-nearest neighbor (KNN) classifier to predict the modification status of a given read, previously trained on wild type and knockout-strains (**Figure 5F,** see also **Table S4**). The algorithm was able to classify reads of both Y and Nm-modified positions with a global accuracy of 87% (**Figure S5C**). Alternative classification algorithms were also tested, however they led to decreased performance compared to KNN (**Table S4).** Overall, our results show that stoichiometry of modification can be predicted in individual RNA reads using direct RNA sequencing by classifying the current intensity of the 15-mers surrounding the modified site, once the site of interest has been first identified in the form of base-calling errors.

De novo **prediction of Y modifications reveals a novel Pus4-dependent mitochondrial rRNA modification**

The identification of RNA modification-specific signatures allows us to perform *de novo* prediction of Y RNA modifications transcriptome-wide using direct RNA sequencing. In this regard, *S. cerevisiae* mitochondrial rRNAs remain much less characterized than cytosolic rRNAs, with only 3 modified sites identified so far in *S.cerevisiae* LSU (21s) (Pintard et al., 2002)**,** and none in SSU (15s) rRNAs. Thus, we hypothesized that direct RNA might reveal previously uncharacterized sites in mitochondrial rRNAs. Firstly, we performed mismatch signature analysis in mitochondrial rRNAs, finding one candidate Y site (15s:854) that displayed both high modification frequency and U-to-C mismatch signature (**Figure 6A**). This site was located in the P-site of the 15s rRNA (**Figure 6B**), and embedded in a similar sequence context and structure as the t-arm of tRNAs, which contains a pseudouridylated (Y55) position placed by Pus4 (**Figure 6C**). Given the resemblance between these two sequences and structures, we hypothesized that Pus4 might be responsible for this modification. To validate our hypothesis as well as the candidate Y854 site, we sequenced rRNA from a *S. cerevisiae* Pus4 knockout strain. Mismatch frequency analysis showed that 15s:U854 position loses its mismatch signature upon deleting Pus4 gene without altering the base-called feature of any other position on the ribosomal RNAs (**Figure 6D,** see also **Figure S6A)**. IGV coverage track snapshots

also show that the signature at 15s:U854 position is completely removed upon the deletion of Pus4, whereas at a control site (18s 221) no change is observed (**Figure 6E)**. Additionally, we observed that the Pus4 target sites reported (TEF1:239,TEF2:239) (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014) completely lost their mismatch signature in Pus4 knockout cells **(Figure S6B,C)**.

To further confirm that this position is pseudouridylated, we developed a novel protocol to identify Y modifications by using CMC probing coupled with nanopore sequencing, which we term nanoCMC-seq. This method allows capturing reverse-transcription drop-off information by sequencing only the first-strand cDNA molecules of CMC-probed RNAs using direct cDNA sequencing with modified steps during the library preparation (**Figure 6F,** see also *Methods*). NanoCMC-seq captured known sites with a very high signal-to-noise ratio (**Figure 6G,** see also **Figure S6D**), and confirmed the existence of Y site in position 854, validating our de novo predictions using direct RNA sequencing. **(Figure 6G**).

rRNA modification profiles do not vary upon exposure to oxidative or thermal stress, whereas Y modification levels in several snRNAs and snoRNAs drastically change upon heat exposure

Ribosomal RNA modifications have recently emerged as an additional regulatory layer to fine-tune translation initiation, efficiency and fidelity (reviewed in (Erales et al., 2017; Esguerra et al., 2008; Jack et al., 2011; Lafontaine, 2015; Yoon et al., 2006)). Indeed, it has been shown that some pseudouridylated and 2′-O-methylated rRNA sites are only partially modified, and that their stoichiometry is cell-type dependent, suggesting that rRNAs modifications may be an important source of ribosomal heterogeneity (Birkedal et al., 2015; Buchhaupt et al., 2014; Krogh et al., 2016; Natchiar et al., 2017; Taoka et al., 2016, 2018). However, a systematic and comprehensive analysis of which environmental cues may lead to changes in rRNA modification stoichiometries, and which RNA modifications may be subject to this tuning, is largely missing.

To assess whether rRNA modification profiles change in response to environmental stimuli, as well as which rRNA modifications may vary with each different exposure, we treated *S. cerevisiae* cells with diverse environmental cues (oxidative, cold and heat stress) and sequenced their RNA, in biological duplicates. We find that none of the stresses tested lead to significant changes in the base-calling ‘errors’ observed in rRNAs (**Figure 7A**). By contrast, we recapitulate previously reported changes in snRNA Y modifications whose stoichiometry has been reported to change upon exposure to environmental cues (**Figure 7B**) (Schwartz et al., 2014), as well as 24 additional Y modification sites in snRNAs and snoRNAs whose stoichiometry varies upon heat exposure, which had not been previously described (**Figure 7C-D,** see also **Figure S7A and Table S5**) (Carlile et al., 2014; van der Feltz et al., 2018; Schwartz et al., 2014; Wu et al., 2011). Overall, our approach confirms previous reports and predicts novel sites whose Y modification levels vary upon heat shock exposure (**Figure 7B-D**), but does not identify any rRNA modified site to be varying in its stoichiometry upon any of the tested stress conditions.

Next we questioned whether pseudouridylation changes during translation may be more nuanced, in that Y levels may differ between rRNAs from different translational fractions, which would not be detected when examining ribosome-bound rRNAs as a whole. To test this, we sequenced both total (input) and polysomal rRNAs from untreated and H_2O_2-treated yeast cells (**Figure 7D**). However, we observed no significant changes in Y rRNA modification profiles when comparing rRNAs from actively translating ribosomes from untreated and H_2O_2-treated cells (**Figure 7E**). Finally, in an attempt to further dissect the different translational repertoires into a higher number of rRNA pools, we sequenced: i) rRNAs from unassembled free rRNA fractions (F1), ii) rRNAs from 40s and 60s subunits (F2), iii) rRNAs from monosomal fractions (F3) and iv) rRNAs from polysomal fractions (F4) (**Figure 7F**). While two positions showed slightly decreased levels of Y (5.8s:Y73 and 25s:YY776) in the free rRNA fraction (F1) compared to assembled ribosomes, no significant changes were observed across the other translational fractions (**Figure 7G,** see also **Figure S7B**). Globally, these results indicate that differential rRNA modification is likely not a mechanism employed by yeast cells to adapt to environmental stress conditions, in agreement with previous observations measuring Y rRNA modification changes (Carlile et al., 2014).

DISCUSSION

RNA modifications are key regulators of a wide range of biological processes (Frye et al., 2018; Li and Mason, 2014; Roundtree et al., 2017). They can modulate the fate of RNA molecules, such as mRNA splicing (Louloupi et al., 2018; Wang et al., 2020; Zhou et al., 2019) or mRNA decay (Guo et al., 2017; Lee et al., 2020), as well as affect major cell and organism-level decisions, including cellular differentiation (Geula et al., 2015; Weng et al., 2018) and sex determination (Haussmann et al., 2016; Kan et al., 2017; Lence et al., 2016). While the biological relevance of RNA modifications is out of question, a major difficulty in studying them is the need for tailored protocols to map each modification individually (Novoa et al., 2017; Schaefer et al., 2017). The lack of a systematic, quantitative and multiplexed approach that can map and quantify all RNA modifications simultaneously is limiting our ability to functionally dissect the role that RNA modifications play in diverse biological conditions and in disease contexts.

Ribosomal RNA are extensively modified as part of their normal maturation, and their modification landscape is relatively well-defined for a series of organisms (Fischer et al., 2015; Sergeeva et al., 2015; Sharma and Lafontaine, 2015; Taoka et al., 2015, 2018), Typically placed by either stand-alone enzymes or snoRNA-guided mechanisms, rRNA modifications tend to cluster in functionally important sites of the ribosome, stabilizing its structure and fine-tuning its decoding capacities (Sloan et al., 2017). Despite the central role that rRNA molecules play in protein translation, recent evidence has shown that rRNA modifications are in fact dynamically regulated (Hebras et al., 2019; Higa-Nakamine et al., 2012), and that their alterations can lead to disease states (Belin et al., 2009; Bortolin-Cavaille et al., 2012; Erales et al., 2017; Heiss et al., 1998; Knight et al., 1999; Liao et al., 2010; Marcel et al., 2013; Mei et al., 2012; Sahoo et al., 2008). However, the extent of the regulatory potential of rRNA

modifications upon environmental exposures has been unclear. For example, previous studies have shown that yeast cells maintain a consistent pattern and level of rRNA modifications regardless of the logarithmic and stationary growth phases (Taoka et al., 2016), the ploidy state (Taoka et al., 2016) or upon post-diauxic growth (Carlile et al., 2014). By contrast, recent studies have shown that rRNA modification levels can be drastically upregulated in hyperthermophilic archaeal species upon temperature stress (Sas-Chen et al., 2020). Thus, whether rRNA modification plasticity is a generalized cellular response to confronting external cues, or whether it is only observed under very specific stress types, species and RNA modifications is yet to be deciphered.

To address these questions and overcome these limitations, we employed direct RNA nanopore sequencing to obtain comprehensive maps of RNA modifications with single molecule resolution, in full-length rRNA molecules. First, we performed systematic analyses of base-calling 'errors' both in *in vitro* transcribed constructs (**Figure 1** and **2**) and *in vivo* rRNA sequences (**Figure 3** and **4**), showing that each RNA modification type studied results in distinct 'error' signatures, which can be employed to distinguish different RNA modification types without the need of coupling wild type and knockout strains. Second, we exploited these findings to predict Y RNA modifications *de novo*, revealing a novel mitochondrial Y rRNA modification in *S. cerevisiae* that was not reported to date, as well as confirming previously reported Y-modified sites in rRNAs, snRNAs and mRNAs (**Figure 6**). Moreover, we showed that Pus4, previously thought to modify only tRNAs and mRNAs, is the enzyme responsible for the Y854 modification in mitochondrial rRNA. Third, we characterized the dynamics of Y modifications in rRNAs, snRNAs and snoRNAs upon a battery of environmental cues and translational fractions. We confirmed previously heat-sensitive snRNAs modifications, as well as revealed novel snRNA and snoRNA modifications that appear upon heat stress exposure (**Figure 7**). By contrast, we found no dynamic changes in the ribosomal epitranscriptome, suggesting that ribosome heterogeneity mediated via rRNA modification dynamics is not a generalized response to stress, neither globally, nor in the pool of rRNAs that are present in actively translating ribosomes. Fourth, we developed a novel method, which we termed nanoCMC-seq, which allowed us to orthogonally validate the novel predicted Y-modified sites. NanoCMC-seq provides a cost-effective strategy to validate Y modifications transcriptome-wide, and is not affected by PCR bias. Fifth, we developed a novel approach to predict stoichiometry of modification using direct RNA sequencing data at the level of individual reads, and showed that it can be applied both to pseudouridylated as well as 2'-O-methylated sites.

Previous studies employing direct RNA sequencing to map RNA modifications transcriptome-wide have been largely focused on m^6A modifications, and have mainly approached the problem of RNA modification detection by comparing wild type and knockout conditions (Leger et al., 2019; Liu et al., 2019; Parker et al., 2020; Price et al., 2019). However, here we show that different RNA modification types (e.g. Y versus m^5C) produce distinct base-calling 'error' signatures (**Figure 2B,C**). Consequently, base-calling errors can be used not only to predict whether a given site is modified or not, but also to identify the underlying RNA modification type, without the need of coupling the direct

RNA sequencing run to a knockout strain. While we should note that base-calling signatures depend to some extent on the surrounding sequence context, we find that Y modifications lead to robust U-to-C mismatch signatures, which can be exploited for *de novo* prediction of Y modifications transcriptome-wide (**Figure 6**). Through this approach, we identified a novel Y modification in the P-site of yeast 15s mitochondrial rRNA (15s:Y854), and show that Pus4 is the enzyme responsible for placing this modification. Finally, we validate our findings using nanoCMC-seq, a novel orthogonal method that can detect Y modifications with single nucleotide resolution by coupling CMC probing to nanopore cDNA sequencing (**Figure 6F,G**).

While we find that Y modifications can be detected both in the form of base-calling 'errors' and altered current intensities (**Figure 4**), we observe that the latter does not provide single nucleotide resolution, and that the maximal current intensity shifts are often seen a few nucleotides away from the real modified site. Thus, we propose that the combination of base-called features and current intensity values can provide stoichiometric information of Y-modified sites with single nucleotide resolution. Specifically, we show that once the site has been located using base-calling error features, binning the per-read current intensity values of the 15-mer regions surrounding Y-modified site into 2 separate clusters (Y-modified and unmodified), provides reasonable estimates of Y modification stoichiometry in all 5 sites analyzed (**Figure 5F**). We should note that while the KNN algorithm shows higher accuracy than the K-means algorithm in terms of stoichiometry prediction (**Figure S5D**), the latter does not require previous training of the model, and thus could be, in principle, applicable to any given site that contains two populations of reads (modified and unmodified).

NGS-based technologies can be used to detect RNA modifications that affect the Watson-Crick base-pairing, such as m^1A, m^3C or $m^{2,2}G$, which become apparent in the form of increased mismatch frequencies at the modified site (Novoa et al., 2020; Ryvkin et al., 2013; Safra et al., 2017). These mismatches reflect nucleotide misincorporations that occur during the reverse transcription step, and are thought to be indicative of the stoichiometry of the modified site. By contrast, here we find that in nanopore direct RNA sequencing data, mismatch frequencies do not reflect the stoichiometry of the site (**Figure 4**). Rather, we propose that the mismatch frequency is in fact a consequence of the deviation of the current intensity of the modified k-mer relative to unmodified counterparts. For example, in the case of Y, the current intensity distribution of the Y-centered k-mers is shifted towards C-centered k-mers, and consequently, it is typically detected in the form of U-to-C mismatch signatures (**Figure S7C**). However, the shift will vary depending on the sequence context, leading to differences in mismatch frequencies (e.g. 25s:Y2826 compared to 25s:Y2880), despite having similar stoichiometries (**Table S3**).

We should note that while our approach allows studying the ribosomal epitranscriptome in a comprehensive manner, there are caveats and limitations, leaving open room for future improvements. First, not all RNA modifications lead to strong alterations in the base-calling features and/or current intensity patterns, such as 2'-O-methylcytosine (Cm), which is poorly detected in direct

RNA sequencing datasets, compared to other RNA modifications (**Figure 3B**). Newer versions of protein nanopores, which are actively being developed, might lead to increased differences in current intensities when these RNA modifications pass through the nanopores. Second, the detection of RNA modifications is dependent on the sequence context; for example, we were unable to detect 18s:Gm908 (**Figure S4**). Similarly, some Y-modified sites, such as 18s:Y1187, caused weaker alterations in base-calling features than other Y-modified positions (**Figure 4**). Third, not all RNA modifications lead to base-calling errors with single nucleotide resolution -such as pseudouridine-. For example, 2'-O-methylations often affect neighboring bases (**Figure 4C** and **S4A**), making it challenging to *de novo* predict modified sites without any prior information. Fourth, stoichiometry could not be accurately predicted in all sites. Specifically, we were unable to predict stoichiometry in 25s:Y2264 due to the low number of reads that the Nanopolish algorithm was able to resquiggle (**Figure S5B**). Future algorithms that improve the current intensity-to-base relationship will be greatly needed to maximize our ability to extract information from direct RNA nanopore sequencing datasets.

Despite these challenges and limitations, our work provides a novel framework for the systematic and comprehensive analysis of the ribosomal epitranscriptome with single molecule resolution, showing that direct RNA sequencing can be employed to estimate modification stoichiometry as well as to *de novo* predict Y RNA modifications transcriptome-wide, without the need of coupling the analysis to a knockout strain. Future work will be needed to functionally dissect the biological roles and dynamics of RNA modifications across further biological conditions and in disease states, to better comprehend how and when the epitranscriptome is tuned to regulate protein translation.

ACKNOWLEDGEMENTS

We thank all the members of the Novoa lab for their valuable insights and discussion. OB is supported by a UNSW International PhD fellowship. MCL is supported by an FPI Severo-Ochoa fellowship by the Spanish Ministry of Economy, Industry and Competitiveness (MEIC). IM and SC are supported by ”la Caixa” Inphinit PhD fellowships (LCF/BQ/DI18/11660028 and LCF/BQ/DI19/11730036, respectively). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 713673. This work was supported by the Australian Research Council (DP180103571 to EMN) and the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) (PGC2018-098152-A-100 to EMN). We acknowledge the support of the MEIC to the EMBL partnership, Centro de Excelencia Severo Ochoa and CERCA Programme / Generalitat de Catalunya.

AUTHOR CONTRIBUTIONS

OB and MCL performed the majority of wet lab experiments, including RNA extraction and nanopore library preparation. OB performed bioinformatic analysis of the data, together with JMR and EMN. OB conceived and performed nanoCMC-Seq experiments. MCL produced the *in vitro* transcribed sequences with modifications and their corresponding nanopore libraries. JMR performed bioinformatic analyses on in vitro transcribed constructs and compared base-calling and mapping

algorithms. IM built polysome gradients and helped with their corresponding nanopore libraries. SC and IM prepared and sequenced the 2'-O-methylation mutant strains. HGSV and RM cultured the *S. cerevisiae* strains under different stress conditions. HL contributed with code for the analysis of current intensity values. ASC cultured all snoRNA-depleted yeast mutant strains and extracted their total RNA. EMN conceived the project. EMN supervised the work, with the assistance of SS and JSM. MCL, OB and EMN built the figures. OB, MCL and EMN wrote the paper, with contributions from all authors.

DECLARATIONS OF INTERESTS

The authors declare that they have no competing interests.

MATERIALS AND METHODS

Yeast culturing

Saccharomyces cerevisiae (strain BY4741) was grown at 30ºC in standard YPD medium (1% yeast extract, 2% Bacto Peptone and 2% dextrose). The deletion strains snR3Δ, snR34Δ and snR36Δ were generated on the background of the BY4741 strain by replacing the genomic snoRNA sequence with a *kanMX4* cassette as detailed in Parker et al. (Parker et al., 2017). After 10 minutes incubation, cells were then quickly transferred into 50 mL pre-chilled falcon tubes, and centrifuged for 5 minutes at 3,000 g in a 4ºC pre-chilled centrifuge. Supernatant was discarded, and cells were flash frozen. For thermal stress, *Saccharomyces cerevisiae* BY4741 cultures were grown in 4 mL of YPD overnight at 30ºC. The next day, cultures were diluted to 0.0001 OD600 in 200 mL of YPD and grown overnight at 30ºC shaking (250 rpm). When the cultures reached an OD600 of 0.4-0.5, the cultures were divided into 3 x 50 mL subcultures, which were then incubated at 30ºC (control), 45ºC (heat shock) or 4ºC (cold shock) for 1 hour. Cells were collected by pelleting and snap freezing. For the analysis of rRNAs modifications across polysomal fractions, yeast BY4741 starter cultures were grown in 6 mL YPD medium at 30ºC with shaking (250 rpm) overnight. 100 mL of fresh YPD medium was inoculated with 10 µL of the stationary culture in a 250 mL erlenmeyer flask, in biological duplicates. Cells were incubated at 30ºC with shaking (250 rpm) until the cultures reached mid-exponential growth phase ($O.D_{660}$.~ 0.4-0.6). Yeast cells were then treated with 1 mM H_20_2 or left without treatment (control) for 30 minutes. 1 mL of cycloheximide stock solution (10 mg/mL) was added to each culture. Pus4 knockout strains (BY4741 MATa pus4::KAN) and its parental strain were obtained from the Yeast Knockout Collection (Dharmacon) and grown under standard conditions in YPD (1% [w/v] yeast extract, 2% [w/v] peptone supplemented with 2% glucose) at 30°C unless stated otherwise.

Total RNA extraction from yeast cultures

Saccharomyces cerevisiae BY4741 cells (strains: snR3Δ, snR34Δ snR36Δ, snR60Δ, snR61Δ, snR62Δ and WT) were harvested via centrifugation at 3000 rpm for 1 minute, followed by two washes with water. RNA was purified from pelleted cells using a MasterPure Yeast RNA extraction kit (Lucigen, MPY03100), according to manufacturer's instructions. Total RNA was then treated with

Turbo DNase (Thermo, #AM2238) with a subsequent RNAClean XP bead cleanup prior to starting the library preparation. For stress conditions and Pus4KO strain, flash frozen pellets were resuspended in 700 µl Trizol with 350 µl acid washed and autoclaved glass beads (425-600 µm, Sigma G8772). The cells were disrupted using a vortex on top speed for 7 cycles of 15 seconds (the samples were chilled on ice for 30 seconds between cycles). Afterwards, the samples were incubated at room temperature for 5 minutes and 200 µll chloroform was added. After briefly vortexing the suspension, the samples were incubated for 5 minutes at room temperature. Then they were centrifuged at 14,000 g for 15 minutes at 4ºC and subsequently the upper aqueous phase was transferred to a new tube. RNA was precipitated with 2X volume Molecular Grade Absolute ethanol and 0.1X volume Sodium Acetate. The samples were then incubated for 1 hour at -20ºC and centrifuged at 14,000 g for 15 minutes at 4ºC. Then the pellet was washed with 70% ethanol and resuspended with nuclease-free water after air drying for 5 minutes on the benchtop. Purity of the total RNA was measured with the NanoDrop 2000 Spectrophotometer. Total RNA was then treated with Turbo DNase (Thermo, #AM2238) with a subsequent RNAClean XP bead cleanup.

Polysome gradient fractionation and rRNA extraction

Yeast pellets from 100 mL cultures were washed with 6 mL of ice-cold Polysome Extraction Buffer (PEB), which contained 20 mM Tris-HCl pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mg/mL cycloheximide and 100 U/mL RNAse inhibitors (RNaseOUT, Invitrogen, #18080051). Cells were centrifuged for 5 minutes at 3,000 g at 4ºC. Washing was repeated by adding 6 mL of ice-cold PEB, followed by centrifugation. Cells were then resuspended in 700 µl of ice-cold PEB, and transferred into pre-chilled 2 ml Eppendorf tubes containing 450 µl of pre-chilled RNAse-free 425-600 µm diameter glass beads (Sigma G8772). Cells were lysed by vortexing at maximum speed for 5 minutes at 4ºC, followed by centrifugation also at maximum speed at bench centrifuge for 5 minutes at 4ºC. 10% of the supernatant was aliquoted into Trizol for total RNA isolation, and kept at -80ºC, which was later used as input. The remaining volume, corresponding approximately to 8 x 10^8 cells, was subsequently loaded onto the sucrose gradient. Linear sucrose gradients of 10-50% were prepared using the Gradient Station (BioComp). Briefly, SW41 centrifugation tubes (Beckman, Ultra-ClearTM 344059) were filled with Gradient Solution 1 (GS1), which consisted of 20 mM Tris-HCl pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mg/mL cycloheximide and 10% w/v RNAse-free sucrose. Solutions GS1 and GS2 were prepared with RNase-DNase free UltraPure water and filtered with a 0.22 µM filter. The tube was then filled with 6.3 mL of Gradient Solution 2 (GS2) layered at the bottom of the tube, which consisted of 20 mM Tris-HCl pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mg/mL cycloheximide and 50% w/v RNAse-free sucrose. The linear gradient was formed using the tilted methodology, with the Gradient Station Maker (Biocomp). Once the gradients were formed, 350 µL of each lysate was carefully loaded on top of the gradients, and tubes were balanced in pairs, placed into pre-chilled SW41Ti buckets and centrifuged at 4ºC for 150 minutes at 35,000 rpm. Gradients were then immediately fractionated using the Gradient Station, and 20 x 500 µl fractions were collected in 1.5 mL Eppendorf tubes, while absorbance was monitored at 260 nm continuously. Fractions were combined in the following way: the free rRNA (F1, fractions 1 and 2), the unassembled

subunits (F2, fractions 3-6), the lowly-translating monosomes (F3, fractions 7-10) and the highly-translating polysomes (F4, fractions 12-17). The pooled fractions were then concentrated using Amicon-Ultra 100K columns (Millipore), and washed two times with cold PEB. The final volume was brought down to 200 µL, and RNA was extracted using TRIzol reagent. Purity of the RNA was measured with NanoDrop 2000 Spectrophotometer.

In vitro transcription of modified and unmodified RNAs

The synthetic 'curlcake' sequences (Liu et al., 2019) used in this study are designed to include all possible 5-mers while minimizing the secondary RNA structure, and consist in 4 *in vitro* transcribed constructs: (i) Curlcake 1, 2244 bp; (ii) Curlcake 2, 2459 bp; (iii) Curlcake 3, 2595 bp, and (iv) Curlcake 4, 2709. The 'curlcake' constructs were *in vitro* transcribed using Ampliscribe™ T7-Flash™ Transcription Kit (Lucigen-ASF3507) with either unmodified rNTPs (UNM), N6-methyladenosine triphosphate (m^6ATP), 5-methylcytosine triphosphate (m^5CTP), 5-hydroxymethylcytosine triphosphate (hm^5CTP) or pseudouridine triphosphate (YTP). All modified NTPs were purchased from TriLink. The sequences included in the short unmodified dataset (UNM-S), which included *B. subtilis* guanine riboswitch, *B. subtilis* lysine riboswitch and *Tetrahymena* ribozyme were also produced by *in vitro* transcription using Ampliscribe™ T7-Flash™ Transcription Kit (Lucigen-ASF3507). All constructs were 5' capped using vaccinia capping enzyme (NEB-M2080S) and polyadenylated using E. coli Poly(A) Polymerase (NEB-M0276S). Poly(A)-tailed RNAs were purified using RNAClean XP beads, and the addition of poly(A)-tail was confirmed using Agilent 4200 Tapestation. Concentration was determined using Qubit Fluorometric Quantitation. Purity of the IVT product was measured with NanoDrop 2000 Spectrophotometer.

Direct RNA library preparation and sequencing of in vitro transcribed constructs

The RNA libraries for direct RNA Sequencing (SQK-RNA001) were prepared following the ONT Direct RNA Sequencing protocol version DRS_9026_v1_revP_15Dec2016, which corresponds to the flowcell FLO-MIN106. Briefly, 800 nanograms of Poly(A)-tailed and capped RNA – 200 ng per construct – was ligated to ONT RT Adaptor (RTA) using concentrated T4 DNA Ligase (NEB-M0202T), and was reverse transcribed using SuperScript III RT (Thermo Fisher Scientific-18080044). The products were purified using 1.8X Agencourt RNAClean XP beads (Fisher Scientific-NC0068576), washing with 70% freshly prepared ethanol. RNA Adapter (RMX) was ligated onto the RNA:DNA hybrid, and the mix was purified using 1X Agencourt RNAClean XP beads, washing with Wash buffer (WSB) twice. The sample was then eluted in Elution Buffer (ELB) and mixed with RNA running buffer (RRB) prior to loading onto a primed R9.4.1 flowcell, and ran on a MinION sequencer with MinKNOW acquisition software version 1.15.1. The sequencing was performed in independent days and using a different flowcell for each sample (UNM, m^6A, m^5C, hm^5C, Y, UNM-S).

Direct RNA library preparation and sequencing of yeast total RNAs

Yeast total RNA was first treated with T4 Polynucleotide Kinase (PNK) (NEB M0201S) in order to remove possible phosphorylated 3'ends before polyA tailing. Briefly, we mixed 1 µl PNK enzyme with

RNAs in a 50 µl reaction with T4 PNK Buffer (10X) and incubated at 37°C for 30 minutes. The products were purified using 1.8X Agencourt RNAClean XP beads (Fisher Scientific-NC0068576), washing with 70% freshly prepared ethanol. Dephosphorylated RNAs were then polyadenylated using E.coli Poly(A) Polymerase, following the commercial protocol. Four different direct RNA libraries were barcoded according to the recent protocol that our lab has recently published (Smith et al., 2019). Custom RT adaptors (IDT) were annealed using following conditions: custom Oligo A and B (**Table S6**) were mixed in annealing buffer (0.01M Tris-Cl pH 7.5, 0.05M NaCl) to the final concentration of 1.4 µM each in a total volume of 75 µL. The mixture was incubated at 94°C for 5 minutes and slowly cooled down (-0.1°C/s) to room temperature. RNA library for direct RNA Sequencing (SQK-RNA002) was prepared following the ONT Direct RNA Sequencing protocol version DRS_9080_v2_revI_14Aug2019 with half reaction for each library until the RNA Adapter (RMX) ligation step. Per reaction (half), 250 ng total of polyA tailed yeast RNAs were ligated to pre-annealed custom RT adaptors (IDT) (Smith et al., 2019) using concentrated T4 DNA Ligase (NEB-M0202T), and was reverse transcribed using Maxima H Minus RT (Thermo Scientific, EP0752), without heat inactivation step. The products were purified using 1.8X Agencourt RNAClean XP beads (Fisher Scientific-NC0068576), washing with 70% freshly prepared ethanol. 50 ng of reverse transcribed RNA from each reaction was pooled and RMX adapter, composed of sequencing adapters with motor protein, was ligated onto the RNA:DNA hybrid and the mix was purified using 1X Agencourt RNAClean XP beads, washing with Wash Buffer (WSB) twice. The sample was then eluted in Elution Buffer (EB) and mixed with RNA Running Buffer (RRB) prior to loading onto a primed R9.4.1 flowcell, and ran on a MinION sequencer with MinKNOW acquisition software version v.3.5.5.

nanoCMC-seq

CMC treatment was adapted from Schwartz et al (Schwartz et al., 2014) with minor changes. Briefly, 20 ug total RNA was incubated in NEBNext® Magnesium RNA Fragmentation Module at 94°C for 1.5 minutes. The fragmented RNA was then incubated with either 0.3M CMC dissolved in 100 µl TEU buffer (50 mM Tris pH8.5, 4 mM EDTA, 7M Urea) or 100 µl TEU buffer (no CMC) for 20 minutes at 37°C. Reaction was stopped with 100 µL of Buffer A (0.3M NaOAc and 0.1mM EDTA, pH 5.6), 700 µL absolute ethanol, and 1 µl GlycoBlue (Thermo Scientific, AM9515). RNA in the stop solution was chilled on dry ice for 5 minutes, and then centrifuged at maximum speed for 15 minutes at 4°C. Supernatant was removed and the pellet was washed with 70% ethanol. After air drying for a few minutes, the pellet was dissolved in 100 µl Buffer A and mixed with 300 µl absolute ethanol and 1 µl GlycoBlue. After chilling on dry ice for 5 minutes, the solution was then centrifuged at maximum speed for 15 minutes at 4°C. Supernatant was removed, and the pellet was washed with 70% ethanol. After washing, the pellet was air dried, and resuspended in 40 µl of 50 mM sodium bicarbonate, pH 10.4, and incubated at 37°C for 3 hours. Furthermore, RNA was mixed with 100 µl Buffer A, 700 µl ethanol, and 1 µl Glycoblue overnight at -20°C. The next day, the solution was centrifuged at maximum speed for 15 minutes at 4°C and the pellet was washed with 70% ethanol and dissolved in the appropriate amount of water after air drying. Unprobed and probed RNAs were treated with T4 Polynucleotide

Kinase (PNK) (NEB, M0201S) as described above before proceeding with ONT Direct cDNA sequencing.

Before starting the library preparation, 9 µl of 100 µM Reverse-transcription primer (Original ONT VNP: 5' /5Phos/ACTTGCCTGTCGCTCTATCTTCTTTTTTTTTTTTTTTTTTTTVN 3') and 9 µl of 100 µM complementary oligo (CompA: 5' GAAGATAGAGCGACAGGCAAGTA 3') were mixed with 1 µ 0.2 M Tris pH 7.5 and 1 µl 1 M NaCl. The mix was incubated at 94°C for 1 minute and the temperature was ramped down to 25°C (-0.1°C/s) in order to pre-anneal the oligos. Then, 100 ng polyA-tailed RNA was mixed with 1 µl pre-annealed VNP+CompA, 1 µl 10 mM dNTP mix, 4 µl 5X RT Buffer, 1 µl RNasin® Ribonuclease Inhibitor (Promega, N2511), 1 µl Maxima H Minus RT (Thermo Scientific. EP0742) and nuclease-free water up to 20 µl. The reverse-transcription mix was incubated at 60°C for 60 minutes and inactivated by heating at 85°C for 5 minutes before moving into the ice. Furthermore, RNAse Cocktail (Thermo Scientific, AM2286) was added to the mix in order to digest the RNA and the mix was incubated at 37°C for 10 minutes. Then the reaction was cleaned up using 1.2X AMPure XP Beads (Agencourt, A63881). In order to be able to ligate the sequencing adapters the the first strand, 1 µl 100 µM CompA was again annealed to the 15 µl cDNA in a tube with 2.25 µl 0.1 M Tris pH 7.5, 2.25 µl 0.5 M NaCl and 2 µ nuclease-free water. The mix was incubated at 94°C for 1 minute and the temperature was ramped down to 25 °C (-0.1°C/s) in order to anneal the complementary to the first strand cDNA. Furthermore, 22.5 µl first strand cDNA was mixed with 2.5 µl Native Barcode (EXP-NBD104) and 25 µl Blunt/TA Ligase Mix (NEB, M0367S) and incubated in room temperature for 10 minutes. The reaction was cleaned up using 1X AMPure XP beads and the libraries were pooled into one tube that finally contains 200 fmol library. The pooled library was then ligated to the sequencing adapter (AMII) using Quick T4 DNA Ligase (NEB, M2200S) in room temperature for 10 minutes, followed with 0.65X AMPure XP Bead cleanup using ABB Buffer for washing. The sample was then eluted in Elution Buffer (EB) and mixed with Sequencing Buffer (SQB) and Loading Beads (LB) prior to loading onto a primed R9.4.1 flowcell, and ran on a MinION sequencer with MinKNOW acquisition software version v.3.5.5.

Analysis of nanoCMC-seq

Reads were base-called with stand-alone Guppy version 3.6.1 with default parameters running in GPU, with built-in demultiplexing tool of Guppy. Unclassified reads were then demultiplexed further using Porechop with --barcode_threshold 50 option (https://github.com/rrwick/Porechop). Then all the merged classified reads were mapped to cytosolic and mitochondrial ribosomal RNA sequences in S. cerevisiae using minimap2 default. Furthermore, a custom script was used to extract RT-drop signatures and the RT-drop scores were plotted using ggplot2. All scripts used to process nanoCMC-seq data with RT-Drop information have been made available in GitHub (https://github.com/novoalab/yeast_RNA_Mod). Notably, due to the 5'end truncation of the nanopore sequencing reads by ~13 nt, RT-drop positions were shifted by 13 nt to accurately determine the exact RT-drop positions.

Demultiplexing direct RNA sequencing

Demultiplexing of the barcoded direct RNA sequencing libraries was performed using DeePlexiCon with default parameters (Smith et al., 2019). Reads with demultiplexing confidence scores greater than 0.95 were kept for downstream analyses . We used a lower score in the case of polysomal fractions (0.8), due to the low read coverage of some fractions. We should note that the dataset was also analyzed using 0.95 threshold, and results and conclusions of the analysis did not change, compared to those obtained using 0.80 threshold.

Base-calling direct RNA sequencing

Reads were base-called with stand-alone Albacore versions 2.1.7 and 2.3.4 with the --disable_filtering parameter, and stand-alone Guppy versions 2.3.1 and 3.0.3 with default parameters running in CPU. *In-house* scripts were used for computing the number of unique and common base-called reads between the different approaches, as well as to compare the tendency of each base-caller regarding read lengths and qualities. Both Albacore and Guppy are available to ONT customers via their community site (https://community.nanoporetech.com/). Differences between the base-called features using distinct base-callers were determined using Kruskal-Wallis test with Bonferroni correction for pairwise comparisons, whereas differences between unmodified and modified sites were assessed using Mann-Whitney-Wilcoxon test.

Mapping algorithms and parameters

Reads were mapped using either *Minimap2* (Li, 2018) or *GraphMap* (Sović et al., 2016). *Minimap2* version 2.14 was run with two different parameter settings: (i) minimap2 -ax map-ont, which is the recommended setting for direct RNA sequencing mapping, and thus we refer to as 'default', and (ii) minimap2 -ax map-ont -k 5, which we refer to as 'sensitive'. *GraphMap* version 0.5.2 was also run with two different parameter settings, for comparison, (i) graphmap align, using default parameters, and (ii) graphmap align --rebuild-index -v 1 --double-index --mapq -1 -x sensitive -z 1 -K fastq --min-read-len 0 -A 7 -k 5, which is expected to increase the tolerance to errors that may occur under the presence of RNA modifications, and thus we refer to as 'sensitive'. For the yeast RNA runs, we used graphmap with default settings to map the reads to different references based on the analysis (ribosomal RNAs, non-coding RNAs and genome supplemented with ribosomal RNAs), which can be found in the github repository https://github.com/novoalab/yeast_RNA_Mod.

Analysis of base-called features in curlcakes

Sam files were transformed into bam files using Samtools version 1.9 (Li et al., 2009), and were then sorted and indexed in order to visualize the data using the Integrative Genomics Viewer (IGV) version 2.4.16 (Robinson et al., 2011). For extracting the mismatch frequencies we first used samtools mpileup and then *in-house* scripts, available in https://github.com/novoalab/Best_Practices. Principal Component Analysis (PCA) was used to reduce the dimensionality of the base-calling error data to visually inspect for base-calling differences, using as input the base-called features (mismatch

frequency, deletion frequency and per-base quality) from all 5 positions of each k-mer. Only k-mers that contained a given modification once in the 5-mer were included in the analysis.

Analysis of base-called features in yeast RNAs

Sam files were transformed into bam files using Samtools version 1.9 (Li et al., 2009), and we then sorted and indexed them in order to visualize the data using the Integrative Genomics Viewer (IGV) version 2.4.16 (Robinson et al., 2011). Base-called features were extracted using Epinano version 1.1. R package ggplot2 was used for data visualization (scripts available in https://github.com/novoalab/yeast_RNA_Mod). Sequencing statistics can be found in **Table S7**.

Analysis of current intensity

Nanopolish (Loman et al., 2015) was used to extract the aligned current intensity values per read and position, using the option *--scale-events*. Mean current intensity per-position was computed by summing the current intensities of all reads aligned to the same position, divided by the total number of reads mapping at a given position. All scripts used to process Nanopolish event align output, including scripts to display mean current intensity values along transcripts have been made available in GitHub (https://github.com/novoalab/yeast_RNA_Mod).

Per-read classification prediction of Y stoichiometry

Nanopolish output was processed to extract the current intensity values corresponding to the 15-mer regions centered in the modified pseudouridine sites, for the 6 sites for which knockout data was available (25s:2133, 25s:2129; 25s:2826, 25s:2880, 25s:2264, 18s:1187), and for all 4 sequencing datasets (wild type, snR3-KO, snR34-KO, snR36-KO). Reads with empty values in the 15-mer region were omitted from the analysis. The reads were then subdivided into training (50%) and testing (50%), and the training set reads for a given pair of wild type-knockout were fed to the K-nearest neighbor algorithm (*caTools* package in R). The performance of the algorithm was assessed both in the test set reads that were not used in the training step, as well as on reads from independent datasets (the other two snoRNA-depleted strains). As an alternative to KNN, k-means clustering was also tested using the built-in *kmeans* function in R, both on scaled current intensity values (KMEANS) as well as on the two first principal component loadings of the current intensity values of the 15-mer (KMEANS_PCA). Predictions by each clustering algorithm, and for each individual site, are shown in **Table S4**.

DATA AVAILABILITY

For *in vitro* transcribed datasets, Fast5 files used in this work were already publicly available (UNM and m^6A: PRJNA521324), or have been made publicly available in SRA (m^5C:PRJNA563591; hm^5C: PRJNA548268; Y:PRJNA511582, UNM-S: PRJNA575545). Base-called and demultiplexed FASTQ from all yeast total RNA direct RNA sequencing data runs have been made publicly available in GEO, under the accession number GSE148603, including processed EpiNano outputs and Nanopolish

outputs. FAST5 files for yeast total RNA direct RNA sequencing are available in ENA under accession PRJEB37798.

CODE AVAILABILITY

All scripts and code used in this work are available at: https://github.com/novoalab/Best_Practices (analysis of *in vitro* curlcake datasets) and https://github.com/novoalab/yeast_RNA_Mod (analysis of *in vivo* datasets). The reference fasta sequences used in this work can be found in https://github.com/novoalab/Best_Practices/reference_fasta and . https://github.com/novoalab/yeast_RNA_Mod/reference_fasta, respectively.

REFERENCES

Anreiter, I., Mir, Q., Simpson, J.T., Janga, S.C., and Soller, M. (2020). New Twists in Detecting mRNA Modification Dynamics. Trends Biotechnol. *0*.

Arango, D., Sturgill, D., Alhusaini, N., Dillman, A.A., Sweet, T.J., Hanson, G., Hosogane, M., Sinclair, W.R., Nanan, K.K., Mandler, M.D., et al. (2018). Acetylation of Cytidine in mRNA Promotes Translation Efficiency. Cell *175*, 1872–1886.e24.

Basu, A., Das, P., Chaudhuri, S., Bevilacqua, E., Andrews, J., Barik, S., Hatzoglou, M., Komar, A.A., and Mazumder, B. (2011). Requirement of rRNA methylation for 80S ribosome assembly on a cohort of cellular internal ribosome entry sites. Mol. Cell. Biol. *31*, 4482–4499.

Belin, S., Beghin, A., Solano-Gonzàlez, E., Bezin, L., Brunet-Manquat, S., Textoris, J., Prats, A.-C., Mertani, H.C., Dumontet, C., and Diaz, J.-J. (2009). Dysregulation of Ribosome Biogenesis and Translational Capacity Is Associated with Tumor Progression of Human Breast Cancer Cells. PLoS ONE *4*, e7147.

Bellodi, C., Krasnykh, O., Haynes, N., Theodoropoulou, M., Peng, G., Montanaro, L., and Ruggero, D. (2010). Loss of Function of the Tumor Suppressor DKC1 Perturbs p27 Translation Control and Contributes to Pituitary Tumorigenesis. Cancer Research *70*, 6026–6035.

Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of ribose methylations in RNA by high-throughput sequencing. Angew. Chem. Int. Ed Engl. *54*, 451–455.

Bortolin-Cavaille, M.-L., -L. Bortolin-Cavaille, M., and Cavaille, J. (2012). The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. Nucleic Acids Research *40*, 6800–6807.

Buchhaupt, M., Sharma, S., Kellner, S., Oswald, S., Paetzold, M., Peifer, C., Watzinger, P., Schrader, J., Helm, M., and Entian, K.-D. (2014). Partial methylation at Am100 in 18S rRNA of baker's yeast reveals ribosome heterogeneity on the level of eukaryotic rRNA modification. PLoS One *9*, e89640.

Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W.V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature *515*, 143–146.

Chen, H., Liu, Q., Yu, D., Natchiar, K., Zhou, C., Hsu, C.-H., Hsu, P.-H., Zhang, X., Klaholz, B., Gregory, R.I., et al. (2020). METTL5, an 18S rRNA-specific m6A methyltransferase, modulates expression of stress response genes. BioRxiv.

Delatte, B., Wang, F., Ngoc, L.V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S., et al. (2016). RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science *351*, 282–285.

Delaunay, S., and Frye, M. (2019). RNA modifications regulating cell fate in cancer. Nat. Cell Biol. *21*, 552–559.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Erales, J., Marchand, V., Panthu, B., Gillot, S., Belin, S., Ghayad, S.E., Garcia, M., Laforêts, F., Marcel, V., Baudin-Baillieu, A., et al. (2017). Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. Proc. Natl. Acad. Sci. U. S. A. *114*, 12934–12939.

Esguerra, J., Warringer, J., and Blomberg, A. (2008). Functional importance of individual rRNA 2'-O-ribose methylations revealed by high-resolution phenotyping. RNA *14*, 649–656.

van der Feltz, C., DeHaven, A.C., and Hoskins, A.A. (2018). Stress-induced Pseudouridylation Alters the Structural Equilibrium of Yeast U2 snRNA Stem II. J. Mol. Biol. *430*, 524–536.

Fischer, N., Neumann, P., Konevega, A.L., Bock, L.V., Ficner, R., Rodnina, M.V., and Stark, H. (2015). Structure of the E. coli ribosome--EF-Tu complex at< 3 Å resolution by C s-corrected cryo-EM. Nature *520*, 567–570.

Frye, M., Harada, B.T., Behm, M., and He, C. (2018). RNA modifications modulate gene expression during development. Science *361*, 1346–1349.

Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods *15*, 201–206.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science *347*, 1002–1006.

Guo, M., Liu, X., Zheng, X., Huang, Y., and Chen, X. (2017). m6A RNA Modification Determines Cell Fate by Regulating mRNA Degradation. Cellular Reprogramming *19*, 225–231.

Haussmann, I.U., Bodi, Z., Sanchez-Moran, E., Mongan, N.P., Archer, N., Fray, R.G., and Soller, M. (2016). m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. Nature *540*, 301–304.

Hebras, J., Krogh, N., Marty, V., Nielsen, H., and Cavaillé, J. (2019). Developmental changes of rRNA ribose methylations in the mouse. RNA Biol. 1–15.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat. Genet. *19*, 32–38.

Higa-Nakamine, S., Suzuki, T., Uechi, T., Chakraborty, A., Nakajima, Y., Nakamura, M., Hirano, N., Suzuki, T., and Kenmochi, N. (2012). Loss of ribosomal RNA modification causes developmental defects in zebrafish. Nucleic Acids Res. *40*, 391–398.

Huang, T., Chen, W., Liu, J., Gu, N., and Zhang, R. (2019). Genome-wide identification of mRNA 5-methylcytosine in mammals. Nat. Struct. Mol. Biol. *26*, 380–388.

Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013). Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol. *14*, 215.

Jack, K., Bellodi, C., Landry, D.M., Niederer, R.O., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, A.M., Thompson, S.R., et al. (2011). rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. Mol. Cell *44*, 660–666.

Jonkhout, N., Tran, J., Smith, M.A., Schonrock, N., Mattick, J.S., and Novoa, E.M. (2017). The RNA

modification landscape in human disease. RNA *23*, 1754–1769.

Kan, L., Grozhik, A.V., Vedanayagam, J., Patil, D.P., Pang, N., Lim, K.-S., Huang, Y.-C., Joseph, B., Lin, C.-J., Despic, V., et al. (2017). The m6A pathway facilitates sex determination in Drosophila. Nat. Commun. *8*, 15737.

Knight, S.W., Heiss, N.S., Vulliamy, T.J., Greschner, S., Stavrides, G., Pai, G.S., Lestringant, G., Varma, N., Mason, P.J., Dokal, I., et al. (1999). X-Linked Dyskeratosis Congenita Is Predominantly Caused by Missense Mutations in the DKC1 Gene. The American Journal of Human Genetics *65*, 50–58.

Krogh, N., Jansson, M.D., Häfner, S.J., Tehler, D., Birkedal, U., Christensen-Dalsgaard, M., Lund, A.H., and Nielsen, H. (2016). Profiling of 2′-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. Nucleic Acids Research *44*, 7884–7895.

Lafontaine, D.L.J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. Nat. Struct. Mol. Biol. *22*, 11–19.

Lahens, N.F., Kavakli, I.H., Zhang, R., Hayer, K., Black, M.B., Dueck, H., Pizarro, A., Kim, J., Irizarry, R., Thomas, R.S., et al. (2014). IVT-seq reveals extreme bias in RNA sequencing. Genome Biol. *15*, R86.

Lee, Y., Choe, J., Park, O.H., and Kim, Y.K. (2020). Molecular Mechanisms Driving mRNA Degradation by m6A Modification. Trends Genet. *36*, 177–188.

Leger, A., Amaral, P.P., Pandolfini, L., and Capitanchik, C. (2019). RNA modifications detection by comparative Nanopore direct RNA sequencing. BioRxiv.

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C.H., Kreim, N., Andrade-Navarro, M.A., Poeck, B., Helm, M., et al. (2016). m6A modulates neuronal functions and sex determination in Drosophila. Nature *540*, 242–247.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

Li, S., and Mason, C.E. (2014). The pivotal regulatory landscape of RNA modifications. Annu. Rev. Genomics Hum. Genet. *15*, 127–150.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., and Yi, C. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat. Chem. Biol. *11*, 592–597.

Li, X., Xiong, X., and Yi, C. (2016). Epitranscriptome sequencing technologies: decoding RNA modifications. Nat. Methods *14*, 23–31.

Li, X., Xiong, X., Zhang, M., Wang, K., Chen, Y., Zhou, J., Mao, Y., Lv, J., Yi, D., Chen, X.-W., et al. (2017). Base-Resolution Mapping Reveals Distinct mA Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. Mol. Cell *68*, 993–1005.e9.

Liao, J., Yu, L., Mei, Y., Guarnera, M., Shen, J., Li, R., Liu, Z., and Jiang, F. (2010). Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Mol. Cancer *9*, 198.

Liu, H., Begik, O., Lucas, M.C., Ramirez, J.M., Mason, C.E., Wiener, D., Schwartz, S., Mattick, J.S., Smith, M.A., and Novoa, E.M. (2019). Accurate detection of m6A RNA modifications in native RNA sequences. Nat. Commun. *10*, 4079.

Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Methods *12*, 733–735.

Louloupi, A., Ntini, E., Conrad, T., and Ørom, U.A.V. (2018). Transient N-6-Methyladenosine

Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. Cell Reports *23*, 3429–3437.

Lovejoy, A.F., Riordan, D.P., and Brown, P.O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One *9*, e110799.

Marcel, V., Ghayad, S.E., Belin, S., Therizols, G., Morel, A.-P., Solano-Gonzàlez, E., Vendrell, J.A., Hacot, S., Mertani, H.C., Albaret, M.A., et al. (2013). p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. Cancer Cell *24*, 318–330.

Marchand, V., Ayadi, L., Ernst, F.G.M., Hertler, J., Bourguignon-Igel, V., Galvanin, A., Kotter, A., Helm, M., Lafontaine, D.L.J., and Motorin, Y. (2018). AlkAniline-Seq: Profiling of m7G and m3C RNA Modifications at Single Nucleotide Resolution. Angew. Chem. Int. Ed. *57*, 16785–16790.

Mei, Y.-P., Liao, J.-P., Shen, J., Yu, L., Liu, B.-L., Liu, L., Li, R.-Y., Ji, L., Dorsey, S.G., Jiang, Z.-R., et al. (2012). Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. Oncogene *31*, 2794–2804.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. Cell *149*, 1635–1646.

Motorin, Y., and Helm, M. (2019). Methods for RNA Modification Mapping Using Deep Sequencing: Established and New Emerging Technologies. Genes *10*.

Natchiar, S.K., Myasnikov, A.G., Kratzat, H., Hazemann, I., and Klaholz, B.P. (2017). Visualization of chemical modifications in the human 80S ribosome structure. Nature *551*, 472–477.

Novoa, E.M., Mason, C.E., and Mattick, J.S. (2017). Charting the unknown epitranscriptome. Nat. Rev. Mol. Cell Biol. *18*, 339–340.

Novoa, E.M., Beaudoin, J.-D., Giraldez, A.J., Mattick, J.S., and Kellis, M. (2020). Best practices for genome-wide RNA structure analysis: combination of mutational profiles and drop-off information. BioRxiv.

Pandolfini, L., Barbieri, I., Bannister, A.J., Hendrick, A., Andrews, B., Webster, N., Murat, P., Mach, P., Brandi, R., Robson, S.C., et al. (2019). METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. Mol. Cell *74*, 1278–1290.e9.

Parker, M.T., Knop, K., Sherwood, A.V., Schurch, N.J., Mackinnon, K., Gould, P.D., Hall, A.J.W., Barton, G.J., and Simpson, G.G. (2020). Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. eLife *9*.

Parker, S., Fraczek, M.G., Wu, J., Shamsah, S., Manousaki, A., Dungrattanalert, K., de Almeida, R.A., Estrada-Rivadeneyra, D., Omara, W., Delneri, D., et al. (2017). A resource for functional profiling of noncoding RNA in the yeast Saccharomyces cerevisiae. RNA *23*, 1166–1171.

Pintard, L., Bujnicki, J.M., Lapeyre, B., and Bonnerot, C. (2002). MRM2 encodes a novel yeast mitochondrial 21S rRNA methyltransferase. EMBO J. *21*, 1139–1147.

Polikanov, Y.S., Melnikov, S.V., Söll, D., and Steitz, T.A. (2015). Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. Nat. Struct. Mol. Biol. *22*, 342–344.

Pratanwanich, P.N., Yao, F., Chen, Y., Koh, C.W.Q., Hendra, C., Poon, P., Goh, Y.T., Yap, P.M.L., Yuan, C.J., Chng, W.J., et al. (2020). Detection of differential RNA modifications from direct RNA sequencing of human cell lines. BioRxiv.

Price, A.M., Hayer, K.E., McIntyre, A.B.R., Gokhale, N.S., Della Fera, A.N., Mason, C.E., Horner, S.M., Wilson, A.C., Depledge, D.P., and Weitzman, M.D. (2019). Direct RNA sequencing reveals m6A modifications on adenovirus RNA are necessary for efficient splicing. BioRxiv.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187–1200.

Ryvkin, P., Leung, Y.Y., Silverman, I.M., Childress, M., Valladares, O., Dragomir, I., Gregory, B.D., and Wang, L.-S. (2013). HAMR: high-throughput annotation of modified ribonucleotides. RNA *19*, 1684–1692.

Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmanith, W., Stern-Ginossar, N., and Schwartz, S. (2017). The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. Nature *551*, 251–255.

Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W., and Beaudet, A.L. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat. Genet. *40*, 719–721.

Sas-Chen, A., Thomas, J.M., Matzov, D., Taoka, M., Nance, K.D., Nir, R., Bryson, K.M., Shachar, R., Liman, G.L.S., Burkhart, B.W., et al. (2020). Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. Nature.

Schaefer, M., Kapoor, U., and Jantsch, M.F. (2017). Understanding RNA modifications: the promises and technological bottlenecks of the “epitranscriptome.” Open Biology *7*, 170077.

Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., et al. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell *159*, 148–162.

Sergeeva, O.V., Bogdanov, A.A., and Sergiev, P.V. (2015). What do we know about ribosomal RNA methylation in Escherichia coli? Biochimie *117*, 110–118.

Sharma, S., and Lafontaine, D.L.J. (2015). “View From A Bridge”: A New Perspective on Eukaryotic rRNA Base Modification. Trends Biochem. Sci. *40*, 560–575.

Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.-D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol. *14*, 1138–1152.

Smith, M.A., Ersavas, T., Ferguson, J.M., Liu, H., Lucas, M.C., Begik, O., Bojarski, L., Barton, K., and Novoa, E.M. (2019). Barcoding and demultiplexing Oxford Nanopore native RNA sequencing reads with deep residual learning. BioRxiv.

Sović, I., Šikić, M., Wilm, A., Fenlon, S.N., Chen, S., and Nagarajan, N. (2016). Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat. Commun. *7*, 11307.

Taoka, M., Nobe, Y., Hori, M., and Takeuchi, A. (2015). A mass spectrometry-based method for comprehensive quantitative determination of post-transcriptional RNA modifications: the complete chemical structure of …. Nucleic Acids.

Taoka, M., Nobe, Y., Yamaki, Y., Yamauchi, Y., Ishikawa, H., Takahashi, N., Nakayama, H., and Isobe, T. (2016). The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. Nucleic Acids Res. *44*, 8951–8961.

Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N., et al. (2018). Landscape of the complete RNA chemical modifications in the human 80S ribosome. Nucleic Acids Res. *46*, 9289–9298.

Vu, L.P., Pickering, B.F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., MacKay, M., et al. (2017). The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. Nature Medicine *23*, 1369–1376.

Wang, X., Li, Z.-T., Yan, Y., Lin, P., Tang, W., Hasler, D., Meduri, R., Li, Y., Hua, M.-M., Qi, H.-T., et al. (2020). LARP7-Mediated U6 snRNA Modification Ensures Splicing Fidelity and Spermatogenesis in Mice. Mol. Cell *77*, 999–1013.e6.

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B.S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C., et al. (2018). METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m6A Modification. Cell Stem Cell *22*, 191–205.e9.

Wongsurawat, T., Jenjaroenpun, P., Wassenaar, T.M., Wadley, T.D., Wanchai, V., Akel, N.S., Franco, A.T., Jennings, M.L., Ussery, D.W., and Nookaew, I. (2018). Decoding the Epitranscriptional Landscape from Native RNA Sequences. BioRxiv.

Wu, G., Xiao, M., Yang, C., and Yu, Y.-T. (2011). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. EMBO J. *30*, 79–89.

Yoon, A., Peng, G., Brandenburger, Y., Zollo, O., Xu, W., Rego, E., and Ruggero, D. (2006). Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. Science *312*, 902–906.

Zhang, L.-S., Liu, C., Ma, H., Dai, Q., Sun, H.-L., Luo, G., Zhang, Z., Zhang, L., Hu, L., Dong, X., et al. (2019). Transcriptome-wide Mapping of Internal N7-Methylguanosine Methylome in Mammalian mRNA. Mol. Cell *74*, 1304–1316.e8.

Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Ż., Pan, J.N., He, C., Parisien, M., and Pan, T. (2019). Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG. Mol. Cell *76*, 70–81.e9.

FIGURES

Figure 1. **Systematic analysis of base-calling and mapping algorithms for the detection of RNA modifications in direct RNA sequencing datasets (A)** Overview of the synthetic constructs used to benchmark the algorithms, which included both unmodified (UNM and UNM-S) and modified (m^6A, m^5C, hm^5C and Y) sequences. For each dataset, we performed: i) comparison of base-calling algorithms, ii) comparison of mapping algorithms, iii) detection of RNA modifications using base-called features and iv) comparative analysis of features to distinguish similar RNA modifications. **(B)** Barplots comparing the percentage of base-called reads using 4 different base-calling algorithms in 6 different unmodified and modified datasets. **(C)** Relative proportion of base-called and mapped reads using all possible combinations (16) of base-callers and mappers included in this study, for each of the 6 datasets analyzed. **(D)** IGV snapshots illustrating the differences in mapping for 3 distinct datasets: UNM, m^6A-modified and Y-modified when base-called with GU 3.0.3. **(E)** Mean sequence identity of different combinations of base-calling and mapping algorithms, for each of the 6 datasets analyzed.

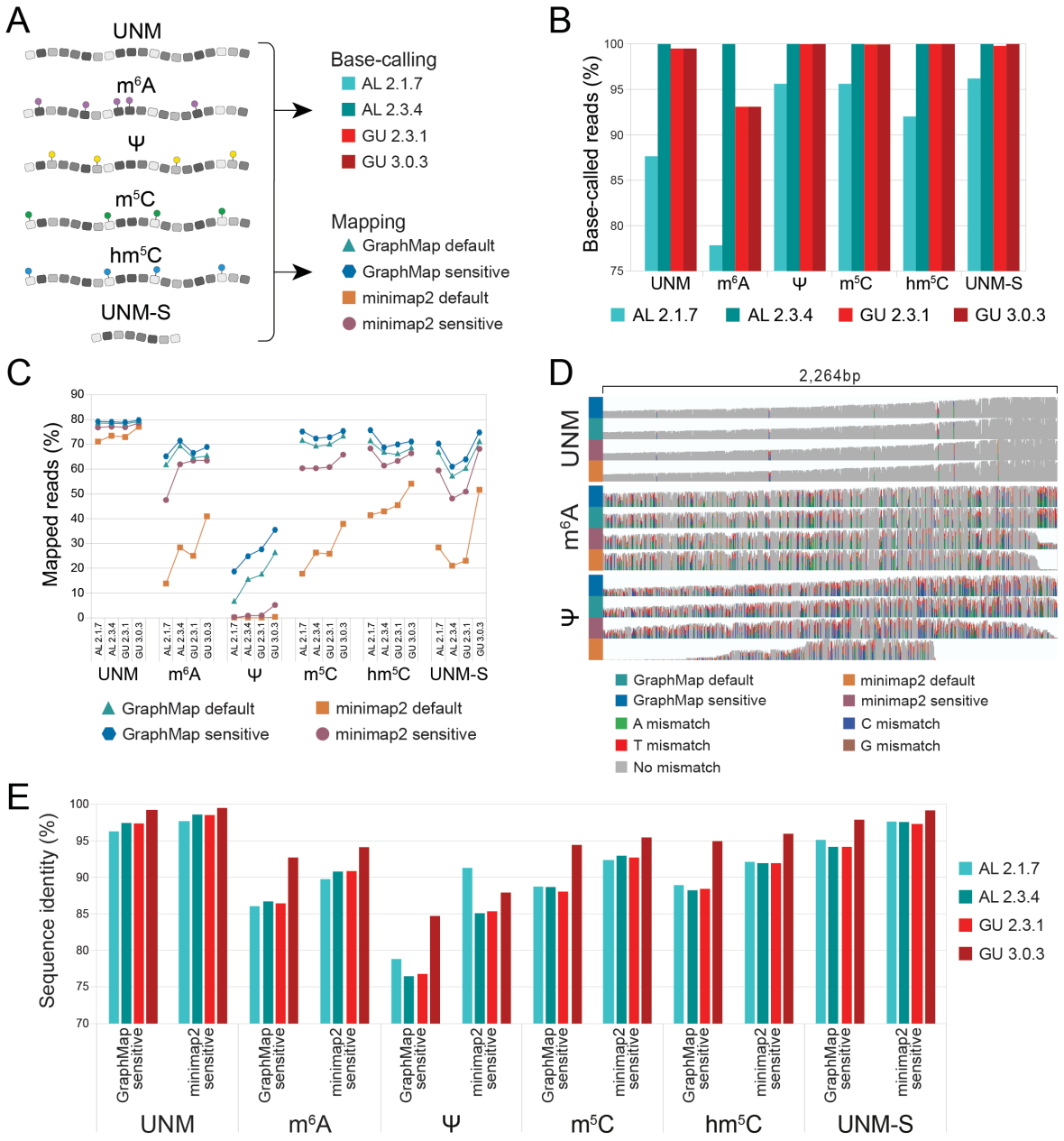


Figure 2. RNA modifications can be detected in the form of systematic base-calling 'errors'. (A) Comparison of global mismatch frequencies using different base-calling algorithms, for the 6 datasets analyzed. **(B)** Principal Component Analysis (PCA) using as input the base-calling error features of quality, mismatch frequency and deletion frequency in positions -2, -1, 0, 1 and 2, for all datasets base-called with GU 3.0.3 and AL 2.1.7 and mapped with GraphMap and minimap2 on sensitive settings. Only k-mers that contained a modification at position 0 were included in the analysis, and the equivalent set of unmodified k-mers was used as a control. **(C)** Per-nucleotide mismatch frequency illustrating which is the reference base where the increased base-calling errors are occurring, for each dataset and base-caller mapper (GraphMap sensitive or minimap2 sensitive).

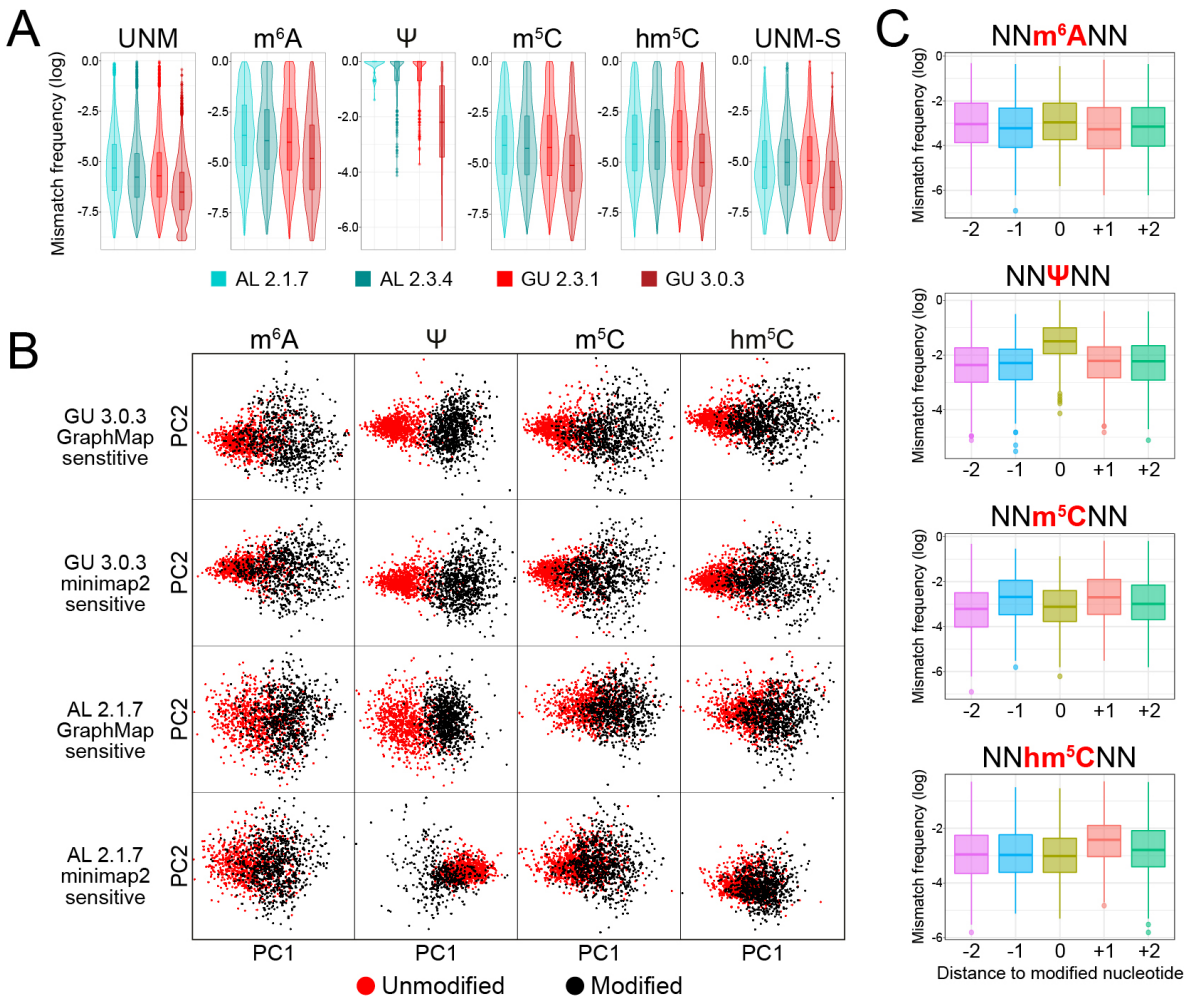


Figure 3. RNA modifications can be detected in yeast ribosomal RNA using base-calling errors. (A) IGV snapshots of yeast ribosomal subunits 5s, 5.8s, 18s and 25s. Known modification sites are indicated below each snapshot and nucleotides with mismatch frequencies greater than >0.1 have been colored. **(B)** Comparison of base-calling features (base quality, mismatch, deletion and insertion frequency) from distinct RNA modification types present in yeast ribosomal RNA. The most descriptive base-calling error per modification is outlined in red. Only RNA modification sites without additional neighboring RNA modifications in the 5-mer were included in the analysis: Y (n=37), Am (n=14), Cm (n=8), Gm (n=8), Um (n=7), ac^4C (n=2), m^1A (n=2), m^3U (n=2), m^5C (n=2), m^1acp^3Y (n=1), m^5U (n=1), m^7G (n=1). (**C**) Ternary plots and barplots depicting the mismatch directionality for selected rRNA modifications (Y, Am, Cm, Gm). Y rRNA modifications tend towards U-to-C mismatches while Am, Cm and Gm modifications did not show specific mismatch directionality patterns.

Figure 3

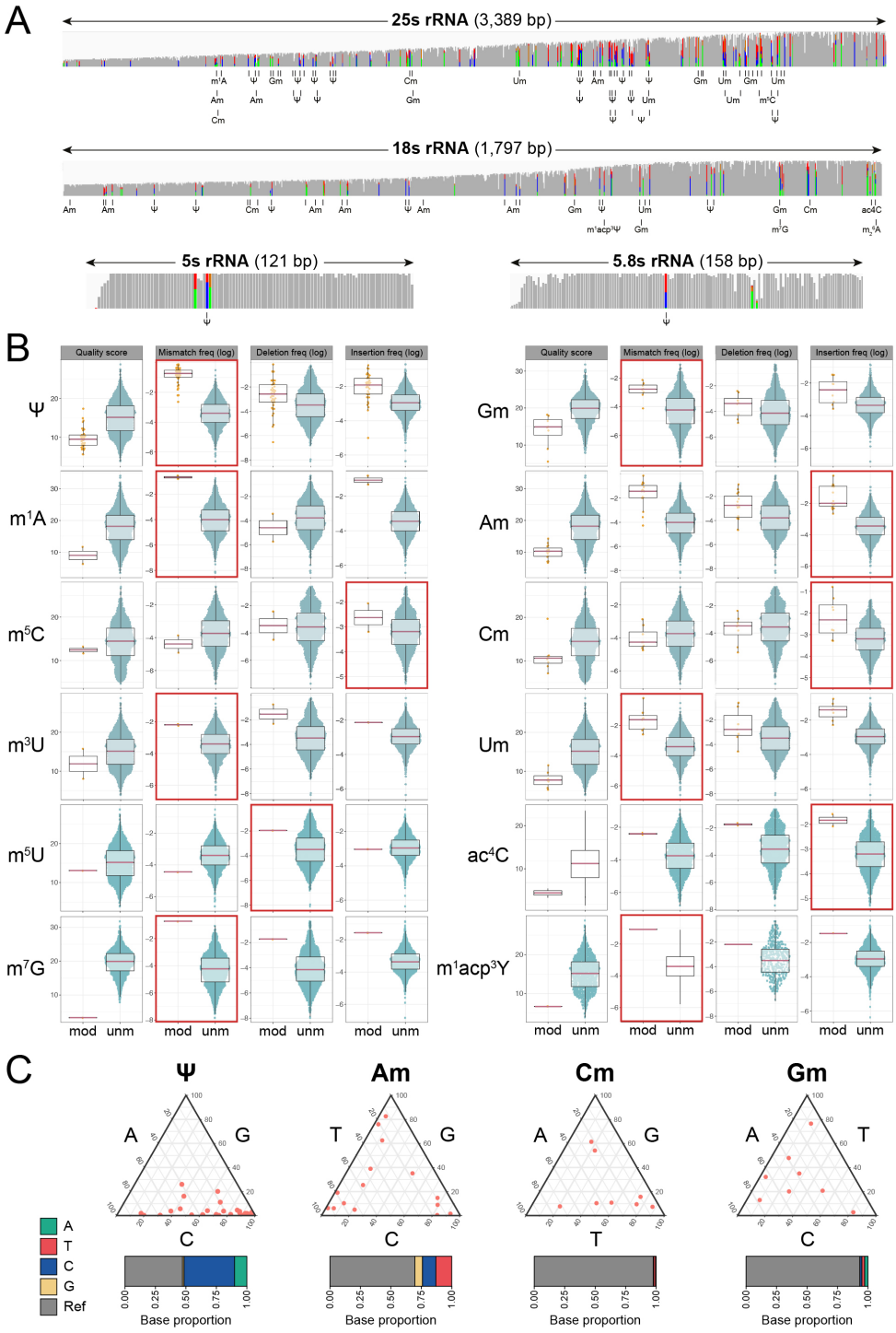

A
25s rRNA (3,389 bp)
18s rRNA (1,797 bp)
5s rRNA (121 bp)
5.8s rRNA (158 bp)
B
Quality score
Mismatch freq (log)
Deletion freq (log)
Insertion freq (log)
Ψ
m1A
m5C
m3U
m5U
m7G
Gm
Am
Cm
Um
ac4C
m1acp3Y
mod
unm
C
Ψ
Am
Cm
Gm
A
T
C
G
Ref
Base proportion

Figure 4. Pseudouridylation and 2'-O-methylations cause systematic base-calling 'errors' as well as altered current intensities, and their signature disappears upon depletion of snoRNAs guiding the modification. (A) IGV snapshots of wild type and three snoRNA-depleted strains with zoomed subsets depicting the site-specific loss of base-called errors at known Y target positions (indicated by asterisks). Nucleotides with mismatch frequencies greater than >0.1 have been colored. **(B)** Comparison of snoRNA knockout mismatch frequencies for each base, relative to wild type, with snoRNA targets sites indicated in red, and non-target sites in gray. **(C)** IGV snapshots of wild type and three snoRNA knockout yeast strains depicting the site-specific loss of base-called errors at known Nm target positions. Nucleotides with mismatch frequencies greater than >0.1 have been colored. **(D)** Comparison of snoRNA knockout summed error frequencies for each base, relative to wild type, with snoRNA targets sites indicated in red, neighboring sites in blue and non-target sites in gray. **(E)** Changes in per-read current intensity distributions at known Y-modified sites were altered upon deletion of specific snoRNAs relative to wild type and control Y-modified sites that were not targeted by the depleted snoRNAs. **(F)** Distributions of per-read current intensity at known 2'-O-methylated sites were altered upon deletion of specific snoRNAs relative to wild type and control 2-O-methyl-modified sites that were not targeted by the depleted snoRNAs.

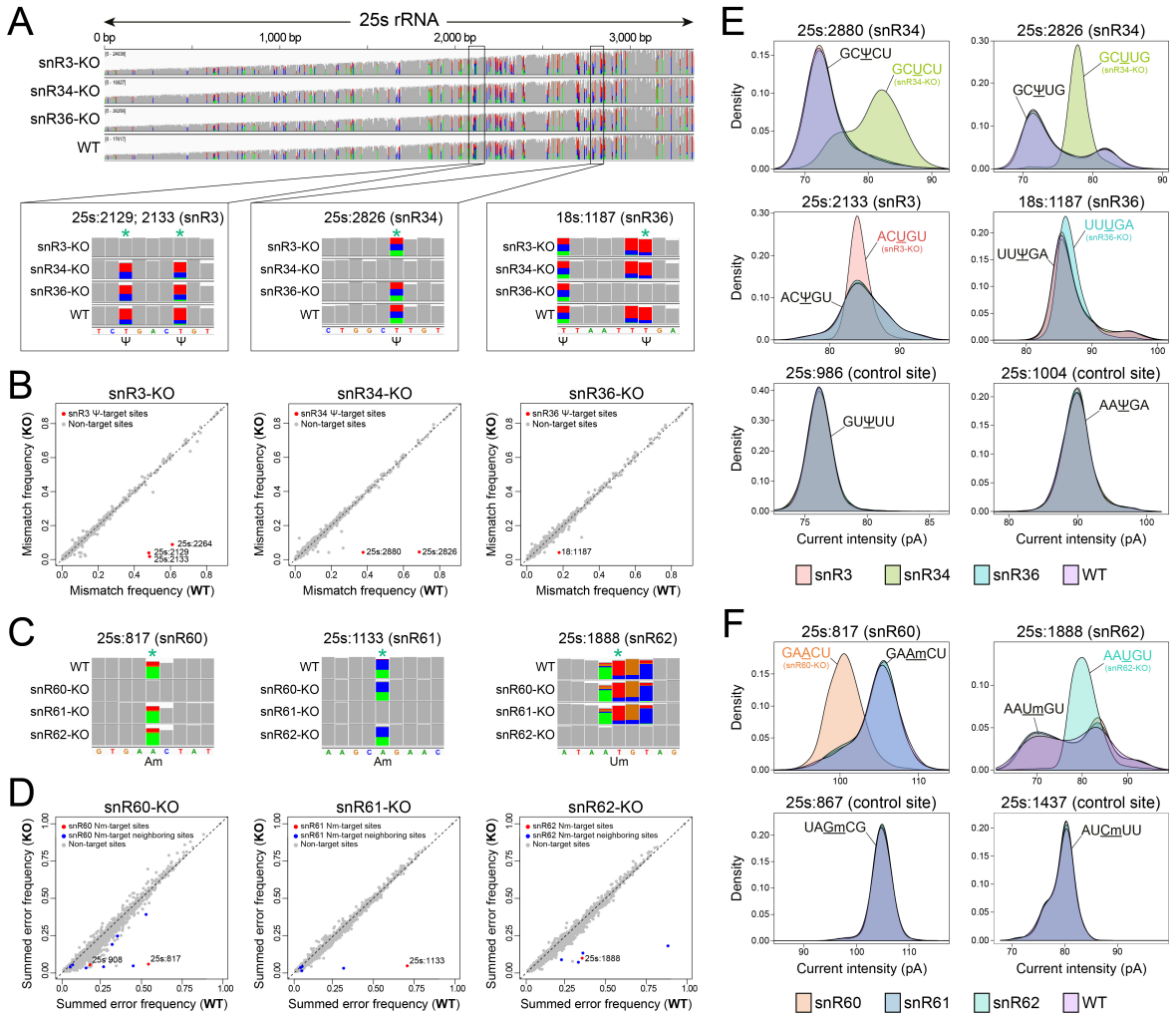


Figure 5. Loss of specific Y rRNA modifications causes deviations in current intensity in regions surrounding the Y sites. (A) Current intensity changes along the 25s rRNA molecule upon snR3 depletion, relative to the wild type strain. In the bottom panel, a zoomed version focusing on the two regions with the most significant current intensity deviations is shown; the first one comprising the 25s:Y2129 and 25s:Y2133 sites, and the second one comprising the 25s:Y2264 site. **(B)** Comparison of current intensity changes for Y knockout sites across each of the snoRNA knockout strains, and two control sites (25s:Y986 and 25s:Y1004), which are not affected by snoRNA depletion. The dotted line indicates the modification position. Of the 6 Y sites that were targeted in the 3 knockout strains studied, only 2 (25s:Y2826 and 25s:Y2880) show a significant deviation in current intensity in the expected position relative to wild type. **(C)** Comparison of current intensity changes for 2'-O-methyl knockout sites across each of the snoRNA knockout strains, as well as of a control site (25s:Am867), which is not affected by snoRNA depletion. The dotted line indicates the modification position. **(D)** Per-read analysis of current intensities centered at the Y-modified sites targeted by each of the 3 knockout strains. In each panel, the per-read current intensities centered in the modified site are shown, both for the wild type (purple) and knockout strain (red: snR3, first panel; green: snR34, second panel; cyan: snR36, third panel). As a control, the same analysis was performed at a control site (25s:Y1004), using reads from wild type (purple) and snR34 knockout strain (green) showing no differences between the read populations. For each site, Principal Component Analysis was performed using 15-mer current intensity values, and the corresponding scatterplot of the two first principal components (PC1 and PC2) is shown on the right, using as input the same read populations as in the left panels. Each dot corresponds to a different read, and is colored according to the strain (snR3:red; snR34:green; snR36:cyan;wt:purple) . **(E)** Per-read analysis of current intensities centered at 2'-O-methylated sites targeted by snR60 and snR61. In each panel, the per-read current intensities centered in the modified site are shown, both for the wild type (purple) and knockout strain (snR60:salmon; snR61:navy; snR62: light blue). On the right side of each current intensity panel, scatterplots of the two first principal components (PC1 and PC2) of the 15-mer current intensity values are shown. Each dot corresponds to a different read, and is colored according to the strain. **(F)** Predicted stoichiometry of Y- and Nm-modified sites using a k-nearest neighbors (KNN) classification algorithm to cluster reads into Y-modified and unmodified.

Figure 5

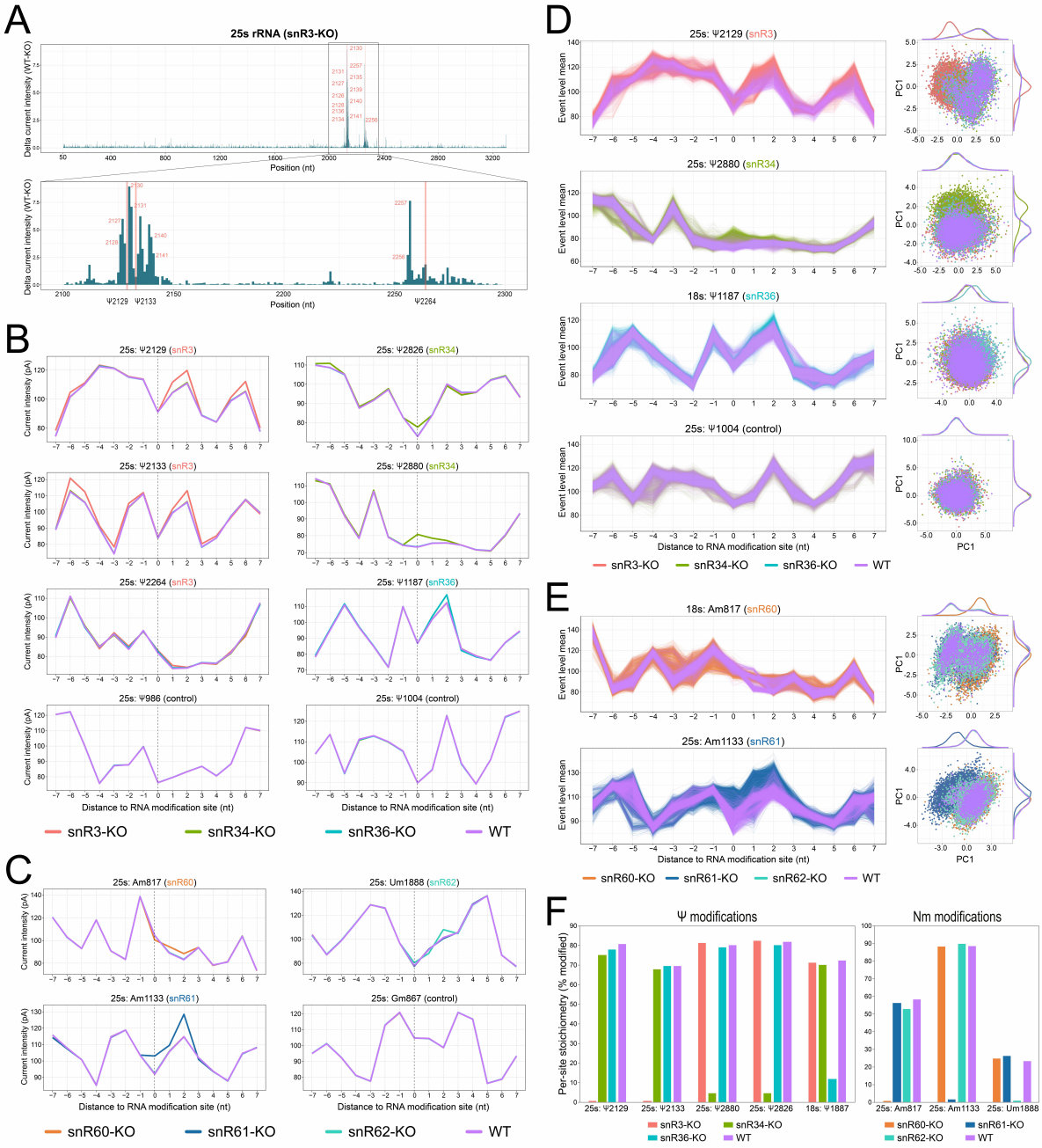

A
25s rRNA (snR3-KO)
Delta current intensity (WT-KO)
Position (nt)
Ψ2129 Ψ2133
Ψ2264
B
25s: Ψ2129 (snR3)
25s: Ψ2826 (snR34)
25s: Ψ2133 (snR3)
25s: Ψ2880 (snR34)
25s: Ψ2264 (snR3)
25s: Ψ1187 (snR36)
25s: Ψ986 (control)
25s: Ψ1004 (control)
Current intensity (pA)
Distance to RNA modification site (nt)
snR3-KO
snR34-KO
snR36-KO
WT
C
25s: Am817 (snR60)
25s: Um1888 (snR62)
25s: Am1133 (snR61)
25s: Gm867 (control)
Current intensity (pA)
Distance to RNA modification site (nt)
snR60-KO
snR61-KO
snR62-KO
WT
D
25s: Ψ2129 (snR3)
25s: Ψ2880 (snR34)
18s: Ψ1187 (snR36)
25s: Ψ1004 (control)
Event level mean
PC1
Distance to RNA modification site (nt)
snR3-KO
snR34-KO
snR36-KO
WT
E
18s: Am817 (snR60)
25s: Am1133 (snR61)
Event level mean
PC1
Distance to RNA modification site (nt)
snR60-KO
snR61-KO
snR62-KO
WT
F
Ψ modifications
Nm modifications
Per-site stoichiometry (% modified)
25s: Ψ2129
25s: Ψ2133
25s: Ψ2880
25s: Ψ2826
18s: Ψ1887
25s: Am817
25s: Am1133
25s: Um1888
snR3-KO
snR34-KO
snR36-KO
WT
snR60-KO
snR61-KO
snR62-KO
WT

Figure 6. ***De novo* prediction of Y modifications reveals a novel Pus4-dependent mitochondrial rRNA modification**. **(A)** Ternary plot of mismatch signatures of U positions in mitochondrial 15s rRNA. Each position has been colored according to its mismatch frequency. Position 15s:854 shows high mismatch frequency and a U-to-C mismatch signature, which is highly similar to the signatures generated by Y modifications. In the right panel, IGV coverage tracks of the 15s mitochondrial rRNA, including a zoomed version showing the tracks centered at the 15s:854 site, in two biological replicates. **(B)** Location of the putative Y854 modified site in the yeast mitochondrial ribosome. The LSU has been colored in cyan, whereas the SSU has been colored in gray. The tRNA is located in the P-site of the ribosome. The PDB structure shown corresponds to 5MRC. **(C)** The candidate Y site is located at the 852-860 loop of the 15s rRNA, which resembles the t-arm of the tRNAs that is modified by Pus4. The binding motif of Pus4 (RRUUCNA) matches with the motif surrounding the 854U site (Schwartz et al., 2014). **(D)** IGV coverage tracks showing that Pus4 knockout leads to depletion of the mismatch signature in the 15s:854 position. **(E)** Scatterplot of mismatch frequencies in WT and Pus4KO cells, showing that the only significant position affected by the knockout of Pus4 is 15sU854. **(F)** Validation of the 15s:Y854 with nanoCMC-Seq, which combines CMC treatment with Nanopore cDNA sequencing in order to capture RT-drops that occur at Y-modified sites upon CMC probing. RT-drops are defined by counting the number of reads ending (3') at a given position. CMC-probed samples will cause accumulation of reads with same 3' ends at positions neighboring the Y site (red), whereas untreated samples will show random distribution of 3' ends of their reads (teal) **(G)** 854U position in 15s rRNA gives the most significant RT-stop signal in CMC-probed conditions (red), whereas this peak is not observed in untreated samples (teal). As a control, we analyzed the nanoCMC-seq results in other rRNA molecules, finding that all detected RT-stop peaks correspond to known Y rRNA modification sites.

Figure 6

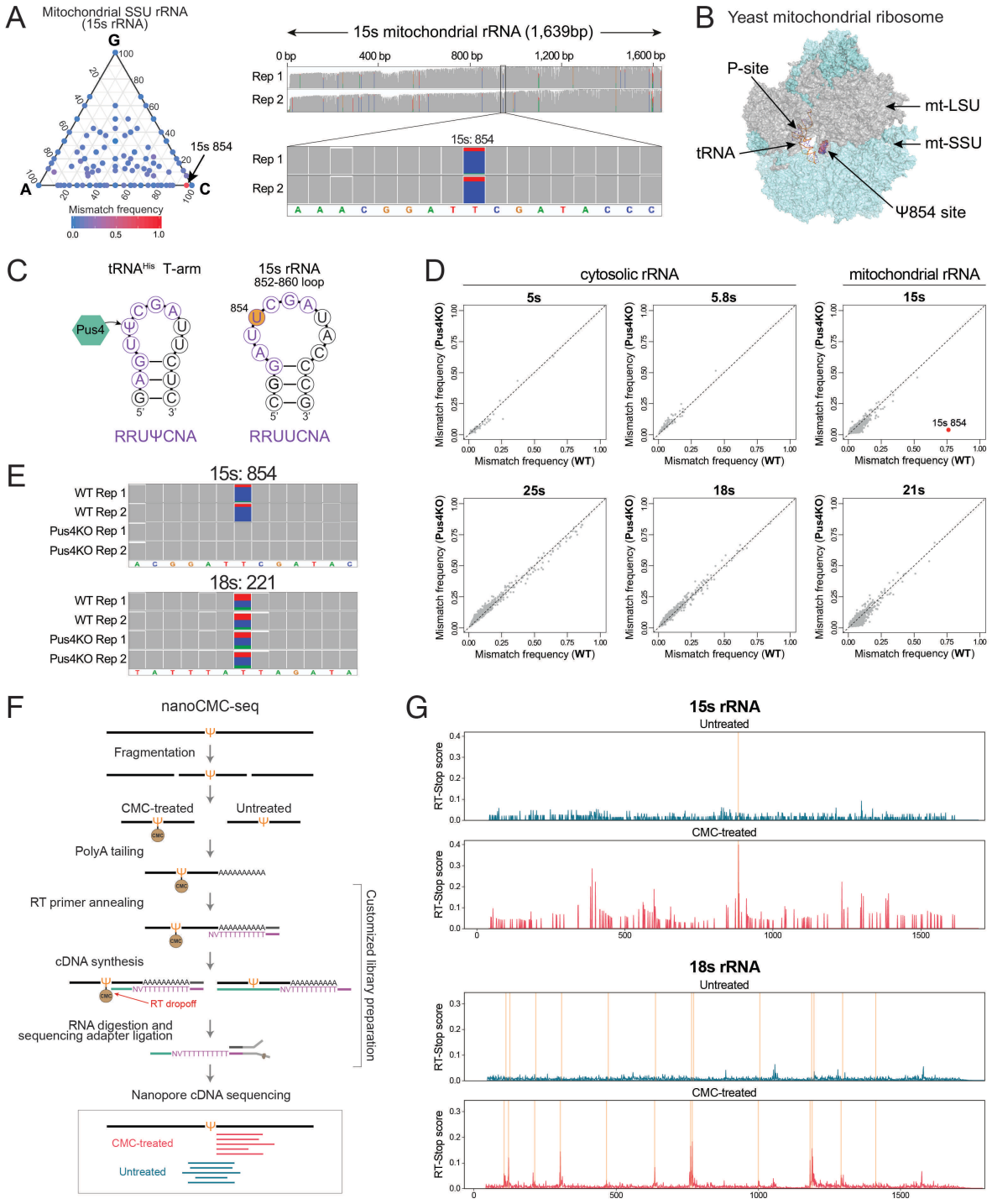
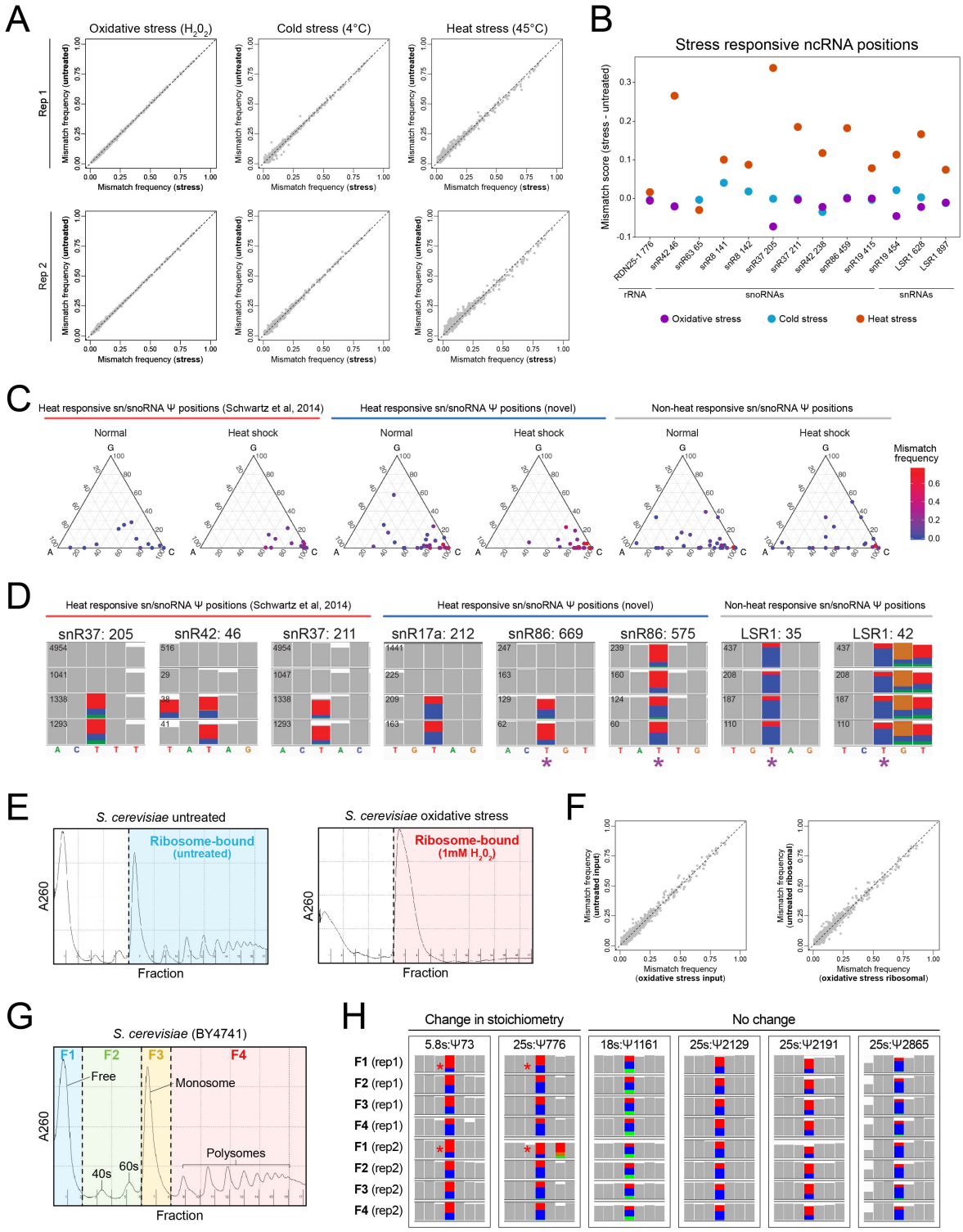

A
Mitochondrial SSU rRNA (15s rRNA)
G
A
C
15s 854
Mismatch frequency
0.0
0.5
1.0
15s mitochondrial rRNA (1,639bp)
0 bp
400 bp
800 bp
1,200 bp
1,600 bp
Rep 1
Rep 2
15s: 854
B
Yeast mitochondrial ribosome
P-site
mt-LSU
mt-SSU
tRNA
Ψ854 site
C
tRNA^His T-arm
15s rRNA 852-860 loop
Pus4
854
RRUΨCNA
RRUUCNA
D
cytosolic rRNA
mitochondrial rRNA
5s
5.8s
15s
25s
18s
21s
Mismatch frequency (Pus4KO)
Mismatch frequency (WT)
15s 854
E
15s: 854
18s: 221
WT Rep 1
WT Rep 2
Pus4KO Rep 1
Pus4KO Rep 2
F
nanoCMC-seq
Fragmentation
CMC-treated
Untreated
PolyA tailing
RT primer annealing
cDNA synthesis
RT dropoff
RNA digestion and sequencing adapter ligation
Customized library preparation
Nanopore cDNA sequencing
CMC-treated
Untreated
G
15s rRNA
Untreated
CMC-treated
18s rRNA
Untreated
CMC-treated
RT-Stop score

Figure 7. **Comparative analysis of yeast rRNA and snRNA Y modifications upon distinct environmental stresses identifies previously known and novel heat-sensitive snRNA and snoRNA Y-modified sites. (A)** Comparison of mismatch frequencies for all rRNA bases from untreated or yeast exposed to oxidative stress (H_2O_2, left panel), cold stress (4ºC middle panel) or heat stress (45ºC, right panel), across two biological replicates. Each dot represents a base. All rRNA bases from cytosolic rRNAs are included in the analysis and plots. **(B)** Untreated vs stressed mismatch scores, for all three stress conditions, in previously reported ncRNA Y sites (Schwartz et al., 2014). **(C)** Ternary plots showing the mismatch frequency and distribution of the known sn/snoRNA pseudouridylation positions in normal condition and heat shock. Positions are grouped in three, positions reported to be heat responsive (Schwartz et al, 2014); novel positions discovered to be heat responsive; non-heat responsive positions. **(D)** IGV snapshots of normal condition (rep1 and rep2) and heat shock condition (rep1 and rep2) yeast strains zoomed into the known sn/snoRNA pseudouridylation positions (Labeled with a blue star). Nucleotides with mismatch frequencies greater than >0.1 have been colored. Coverage for each position/condition is given on the top left of each row. **(E)** Polysome profiles of ribosomal-bound RNA fractions isolated from untreated and stressed H_2O_2-treated yeast cells. **(F)** Comparison of mismatch frequency for untreated vs H_20_2-treated input RNA (left) and untreated vs H_2O_2-treated ribosome-bound RNA (right). **(G)** Profiles of ribosomal fractions isolated from yeast grown under normal conditions, using polysome gradient fractionation, including free rRNAs which are not part of ribosomal subunit (F1), rRNAs from 40s and 60s subunits (F2), rRNAs extracted from monosomal fractions (F3) and polysome fractions (F4). **(H)** IGV snapshots of the two Y sites that change stoichiometry between translational fractions and four representative Y sites that show no significant change. Nucleotides with mismatch frequencies greater than >0.1 have been colored.

Figure 7


A
Oxidative stress (H2O2)
Cold stress (4°C)
Heat stress (45°C)
Rep 1
Rep 2
Mismatch frequency (untreated)
Mismatch frequency (stress)
B
Stress responsive ncRNA positions
Mismatch score (stress - untreated)
RDN25-1 776
snR42 46
snR63 65
snR8 141
snR8 142
snR37 205
snR37 211
snR42 238
snR86 459
snR19 415
snR19 454
LSR1 628
LSR1 897
rRNA
snoRNAs
snRNAs
Oxidative stress
Cold stress
Heat stress
C
Heat responsive sn/snoRNA Ψ positions (Schwartz et al, 2014)
Heat responsive sn/snoRNA Ψ positions (novel)
Non-heat responsive sn/snoRNA Ψ positions
Normal
Heat shock
Mismatch frequency
D
Heat responsive sn/snoRNA Ψ positions (Schwartz et al, 2014)
Heat responsive sn/snoRNA Ψ positions (novel)
Non-heat responsive sn/snoRNA Ψ positions
snR37: 205
snR42: 46
snR37: 211
snR17a: 212
snR86: 669
snR86: 575
LSR1: 35
LSR1: 42
E
S. cerevisiae untreated
Ribosome-bound (untreated)
S. cerevisiae oxidative stress
Ribosome-bound (1mM H2O2)
A260
Fraction
F
Mismatch frequency (untreated input)
Mismatch frequency (oxidative stress input)
Mismatch frequency (untreated ribosomal)
Mismatch frequency (oxidative stress ribosomal)
G
S. cerevisiae (BY4741)
F1
F2
F3
F4
Free
Monosome
Polysomes
40s
60s
A260
Fraction
H
Change in stoichiometry
No change
5.8s:Ψ73
25s:Ψ776
18s:Ψ1161
25s:Ψ2129
25s:Ψ2191
25s:Ψ2865
F1 (rep1)
F2 (rep1)
F3 (rep1)
F4 (rep1)
F1 (rep2)
F2 (rep2)
F3 (rep2)
F4 (rep2)

SUPPLEMENTARY FIGURES

Figure S1. IGV snapshots illustrate the differences in mapping results when using distinct base-calling and mapping algorithm combinations. The algorithms have been benchmarked using four *in vitro* transcribed sequences ('curlcakes'), which were either unmodified (UNM), m^6A-modified (m^6A), m^5C-modified (m^5C), hm^5C-modified (hm^5C) or Ψ-modified (Ψ). Reads were base-called using either Albacore 2.1.7 (AL 2.1.7) or Guppy 3.0.3 (GU 3.0.3), and then mapped using minimap2 or GraphMap in 'sensitive' mode.

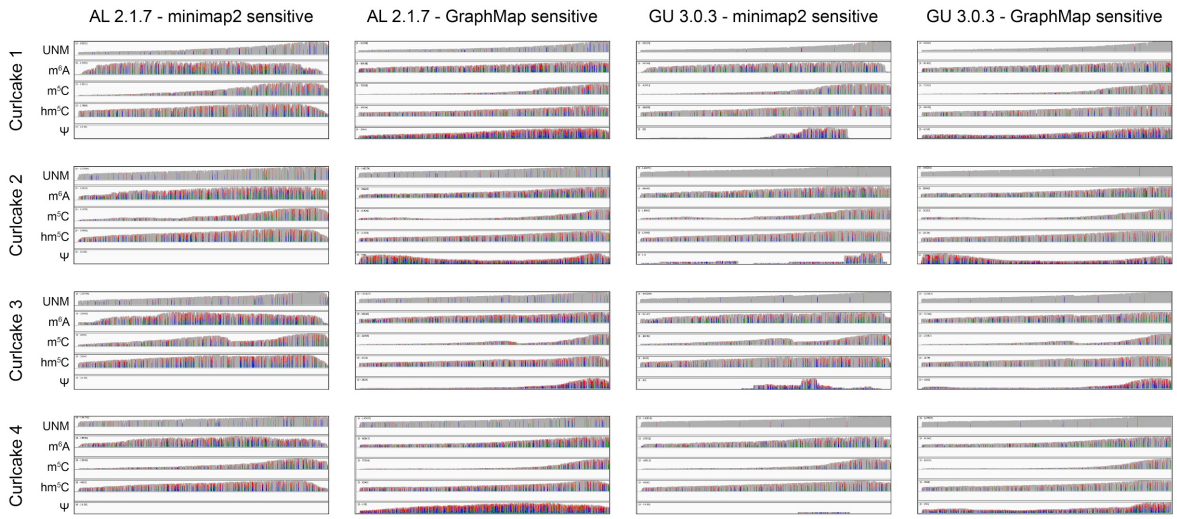


Figure S2. RNA modifications have distinct base-calling 'error' signatures and are sequence context-dependent. (A) Comparison of read lengths and mean quality scores in different *in vitro* transcribed datasets (UNM, m^6A, Ψ, m^5C, hm^5C and UNM-S) when base-called using different algorithms (AL 2.1.7, AL 2.3.4, GU 2.3.1 or GU 3.0.3). Results show that read lengths do not largely vary across base-callers. By contrast, per-base quality strongly varies depending on the choice of base-calling algorithm. **(B)** Barplots of mean per-base quality show that per-base qualities are slightly decreased in all modified datasets, relative to unmodified ones, being this different most evident when the data has been base-called with GU 3.0.3. **(C)** Ternary plots depicting the mismatch distribution of the unmodified (left) and modified (right) positions colored by log coverage, in 5 different *in vitro* datasets: unmodified (all left panels), m^6A-modified (m^6A), Ψ-modified (Ψ), m^5C-modified (m^5C), hm^5C-modified (hm^5C). Only modified nucleotides, and their relative unmodified counterparts in the UNM dataset, are shown. Each dot represents a different nucleotide in the reference. **(D)** Logo representations of the mismatch signatures generated by m^5C and hm^5C. Results show that the signatures are different depending on the modification; however, these also vary depending on the 5-mer sequence. Mismatch mostly occurs in neighboring positions.

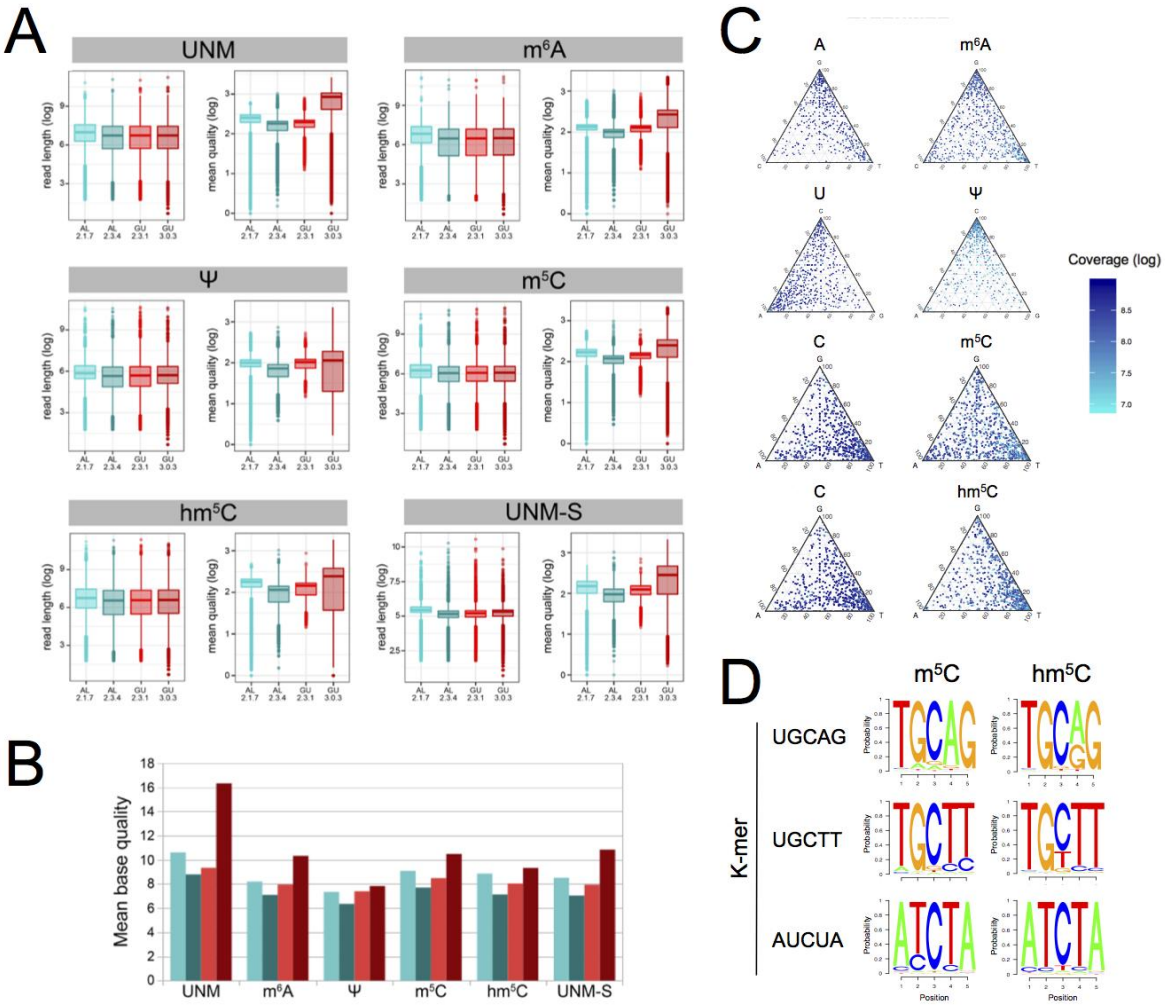


Figure S3. Ribosomal RNA modifications present in yeast rRNAs show distinct base-calling 'error' signatures. (A) IGV snapshots centered in distinct yeast ribosomal RNA modifications. Known rRNA modification sites are indicated below each snapshot. Nucleotides with mismatch frequencies greater than >0.15 have been colored. **(B)** Dotplots of base-calling errors (deletion frequency, insertion frequency, mismatch Frequency, and per-base quality) observed in modified 5-mers, centered in the modified position. Each dot corresponds to a different 5-mer. The total number of 5-mers included in the analysis varies depending on the abundance of each rRNA modification type in yeast rRNAs: Ψ (n=46), Am (n=14), Cm (n=10), Gm (n=15) and Um (n=9). 5-mers that contain more than one modification in the 5-mer region were excluded from the analysis.

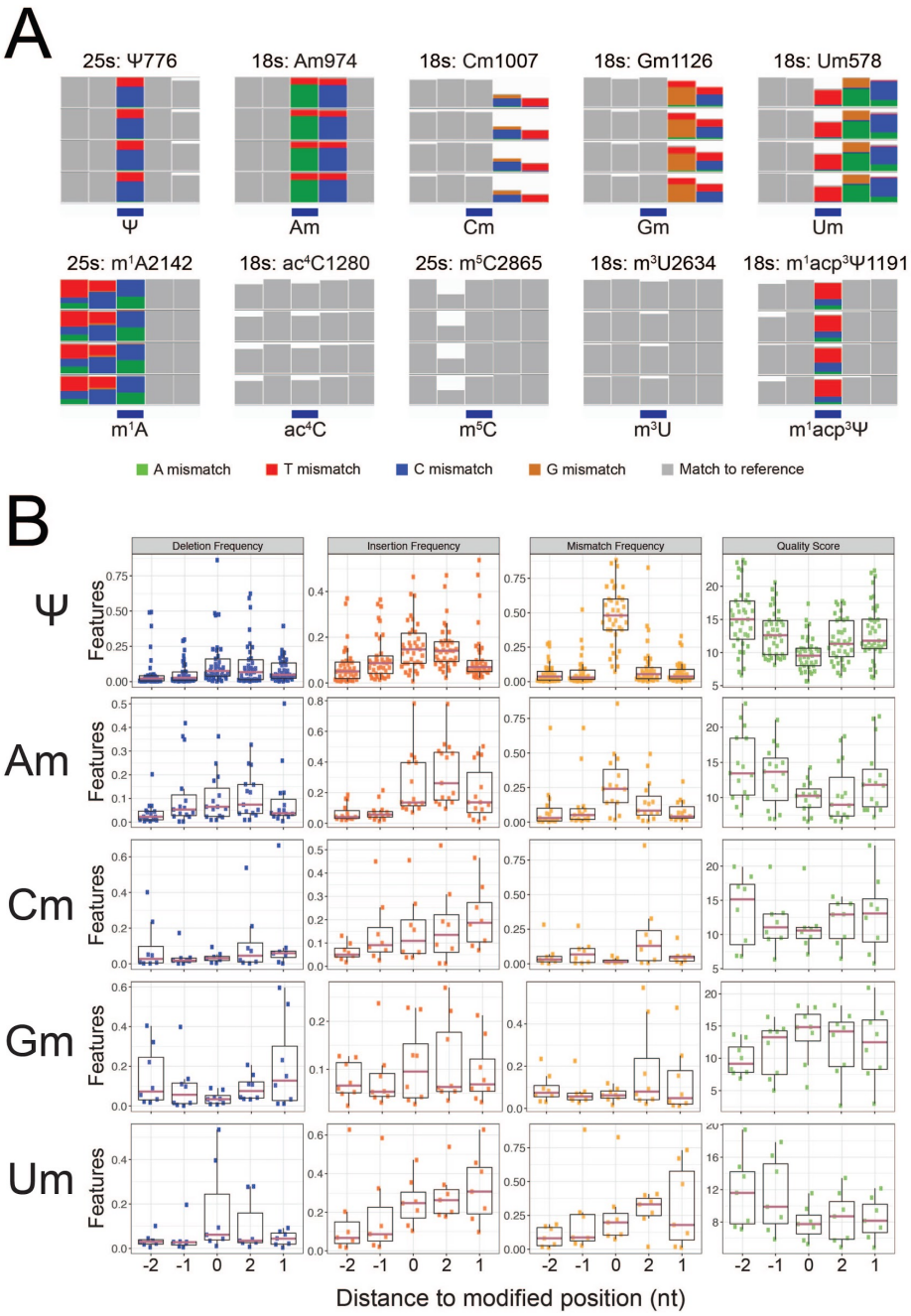


Figure S4. Base-calling signature of 2'-O-methylations often alter the neighboring positions, whereas Ψ modifications mainly affect the modified site. (A) IGV snapshots centered in distinct yeast rRNA modified sites: Ψ-modified sites are shown in the upper panels, whereas 2'-O-methylated sites are shown in the bottom panels. **(B)** Comparison of base-calling 'errors' (mismatch, deletion and insertion frequency) observed in snoRNA-depleted strains (snR60, top panels; snR61, middle panels, snR62, bottom panels) relative to wild type, with snoRNA targets sites indicated in red, neighboring sites indicated in blue and non-target sites in gray.

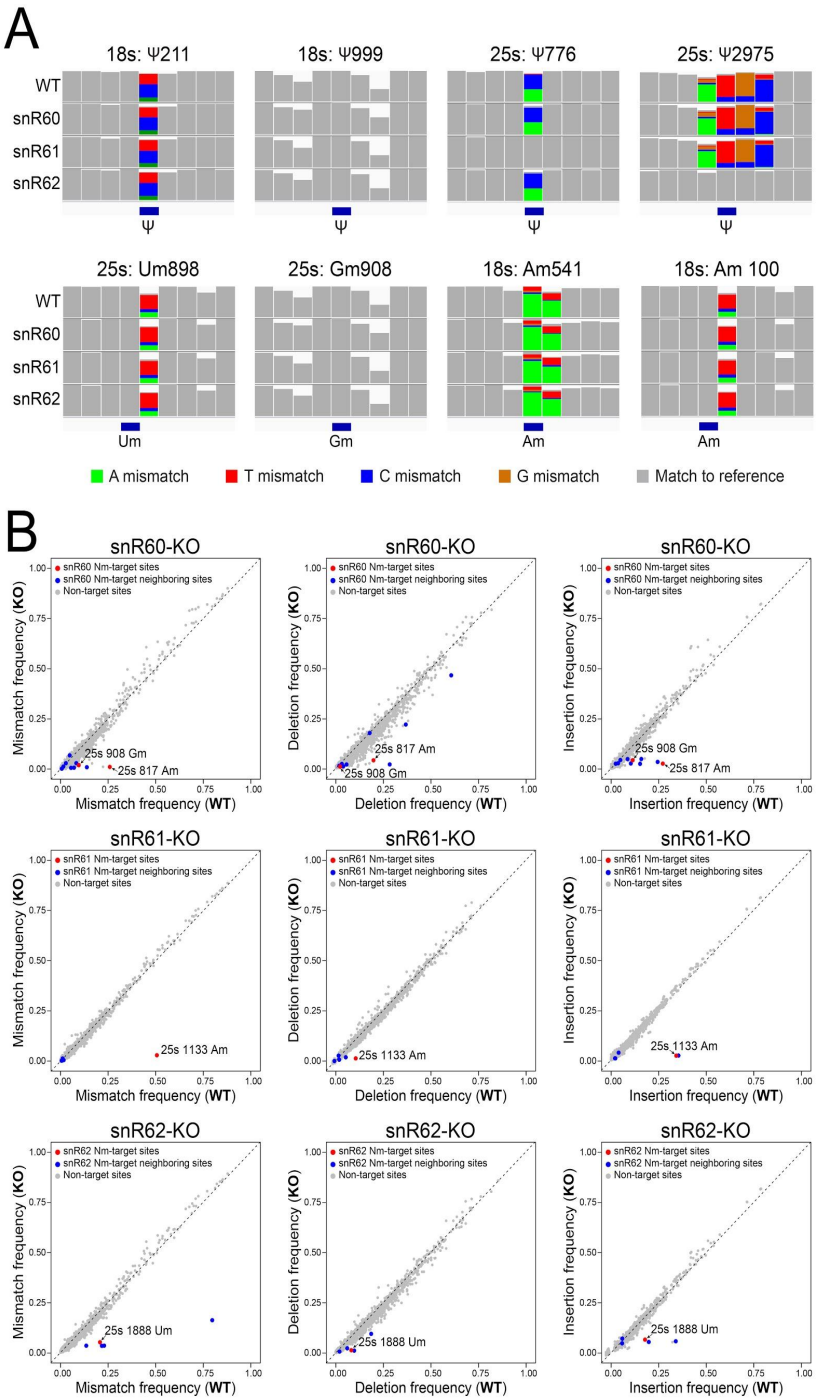


Figure S5. Pseudouridylations and 2'-O-methylations can be detected and quantified in the form of altered current intensities. (A) Absolute differences in current intensity along the 25s rRNA and 18s rRNAs upon depletion of snR34 and snR36, respectively, relative to the wild type strain. Red vertical lines indicate the KO pseudouridylation positions. **(B)** Per-read analysis of current intensities centered at the Ψ modified sites targeted by each of the 3 knockout strains. In each panel, the per-read current intensities centered in the modified site are shown, both for the wild type (purple) and knockout strain (red: snR3; green: snR34; cyan: snR36). As a control, the same analysis was performed at two control sites (25s:Ψ986), using reads from wild-type (purple) and snR34 knockout strain (green) showing no differences between the read populations. Each line indicates a single read. **(C)** Accuracy barplots of stoichiometry predictions using different algorithms at pseudouridylation and 2'-O-methylation positions.

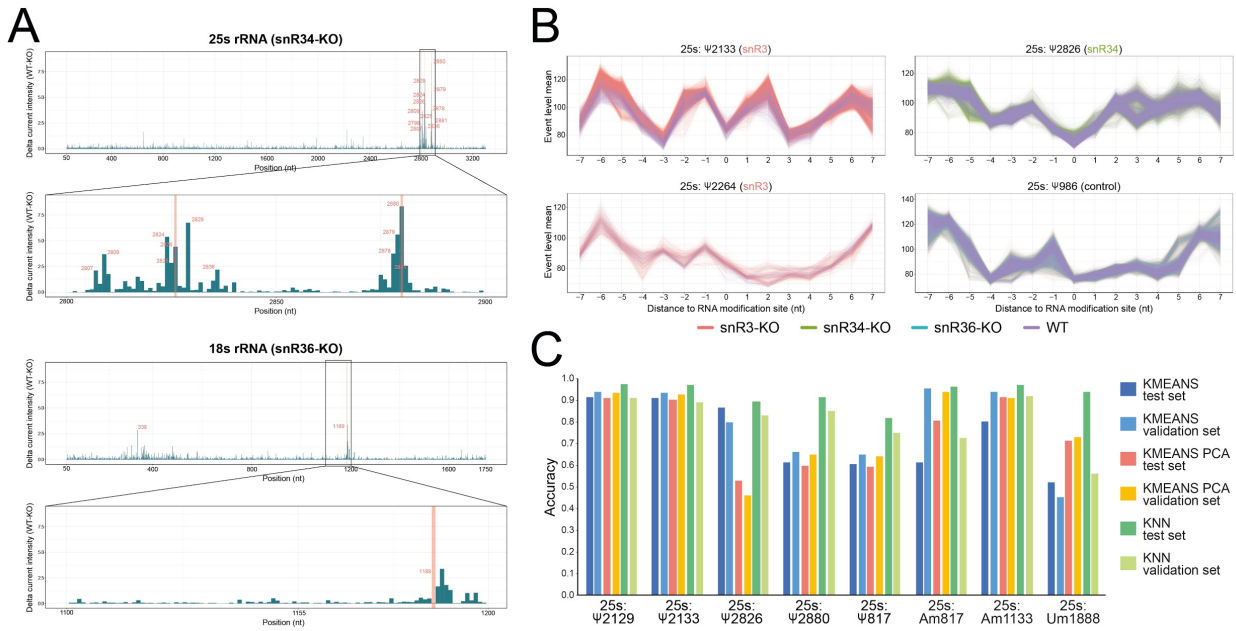


Figure S6. *De novo* prediction of Ψ modifications reveals a novel Pus4-dependent modification (15s:Ψ854) in yeast mitochondrial rRNAs, and captures previously reported Pus4-dependent mRNA modifications. (A) Comparison of base-calling errors (deletion and Insertion frequency) in Pus4-depleted strains, relative to wild type, both in cytosolic (left panels) and mitochondrial rRNAs (right panels). See also main Figure 6 for comparison of mismatch frequencies between the strains. **(B)** Comparison of mismatch frequency for each base, relative to wild type in positions mapped to yeast genome and rRNA. Top plot is replicate 1 and bottom plot is replicate 2. **(C)** IGV snapshots of wild type (rep1 and rep2) and Pus4 knockout (rep1 and rep2) yeast strains with zoomed subsets depicting the site-specific loss of mismatch at known target positions. Nucleotides with mismatch frequencies greater than >0.1 have been colored. **(D)** RT-stop signal on 25s rRNA in CMC-probed conditions (green), whereas this peak is not observed in untreated samples (red). All detected RT-stop peaks correspond to known Ψ rRNA modification sites indicated with a dashed line.

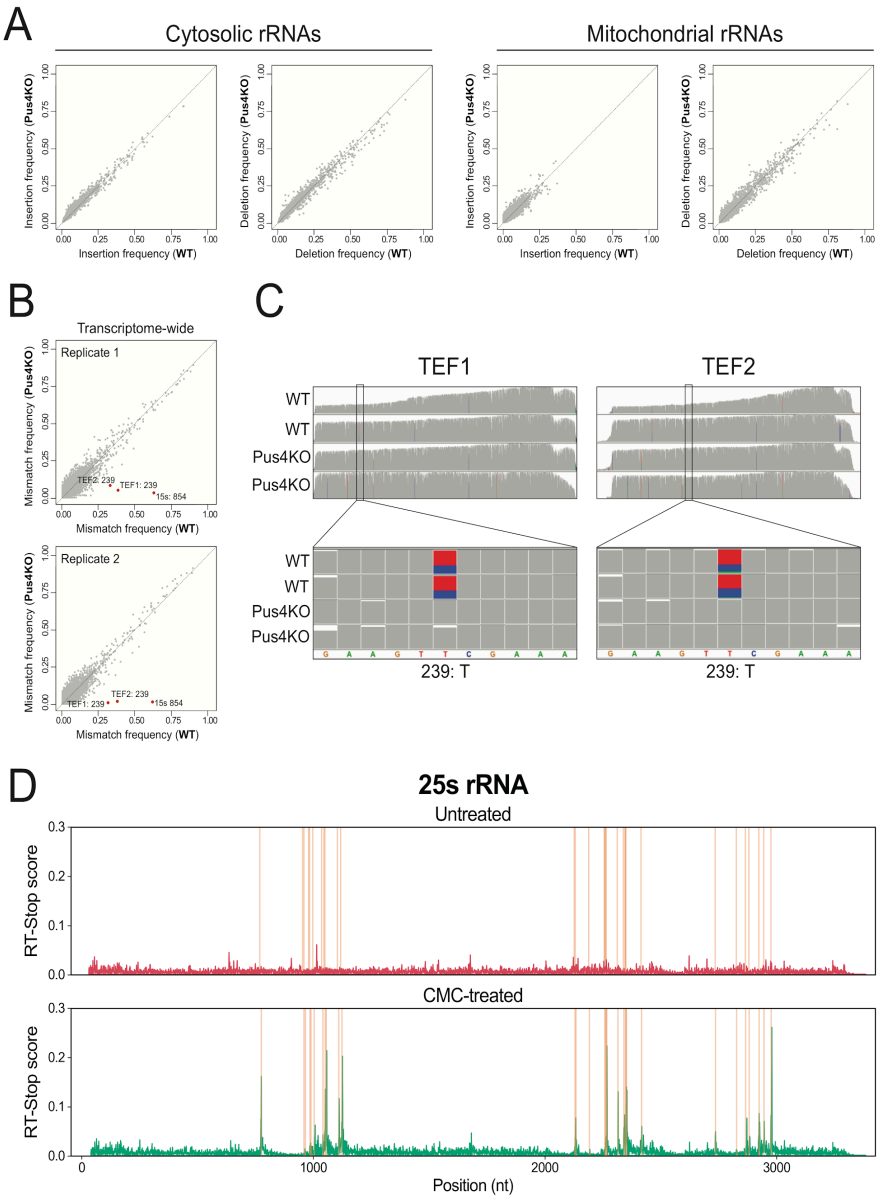
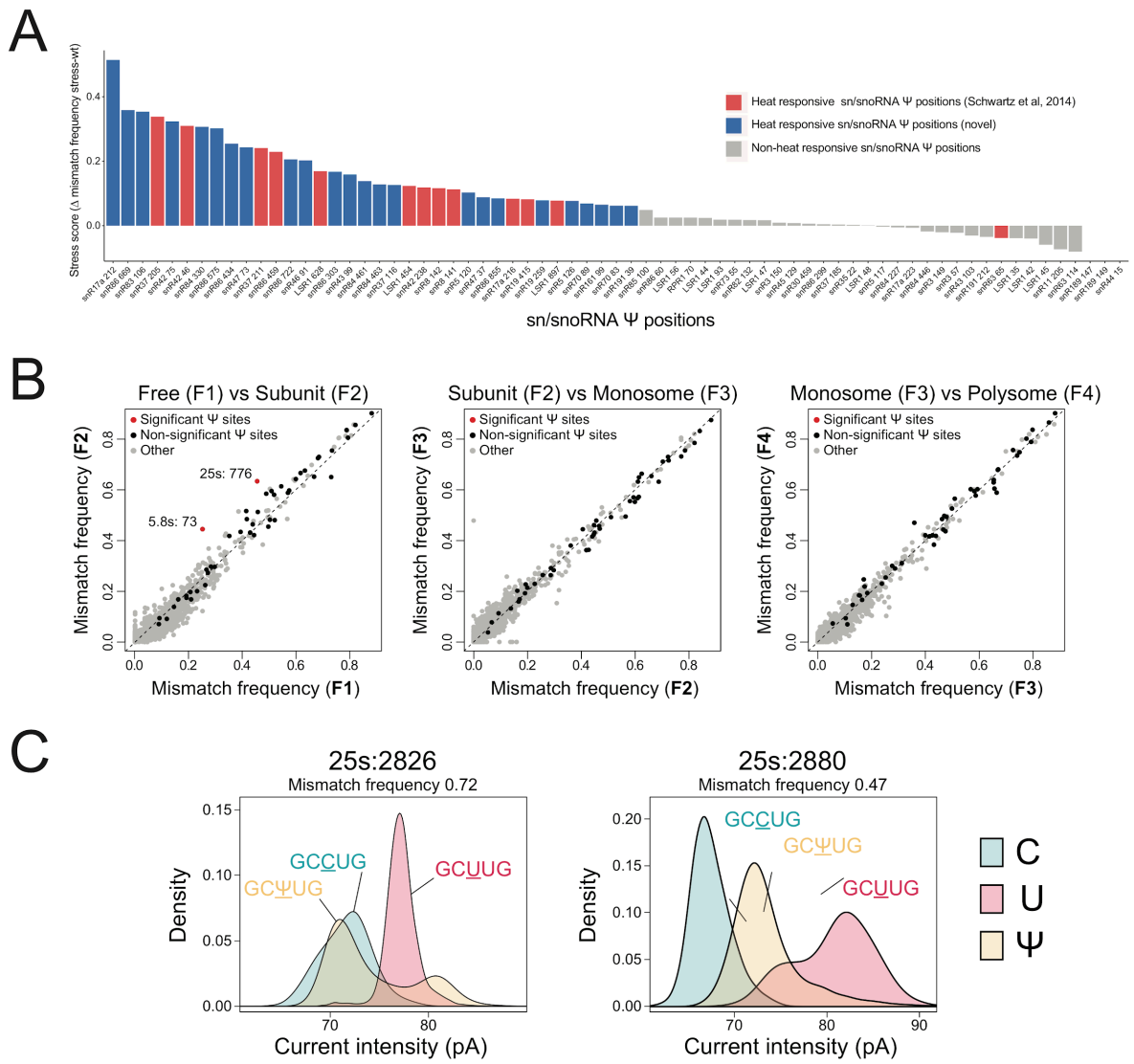



Figure S7. Comparative analysis of mismatch frequencies can be employed to identify differentially modified Ψ sites across strains and conditions, but the mismatch frequencies at Ψ positions do not correspond to Ψ modification stoichiometries. (A) Mismatch scores in individual known sn/snoRNA pseudouridylation positions calculated by delta mismatch frequency between heat shock and normal conditions. **(B)** Comparison of mismatch frequencies for different fractions of ribosomal RNAs (F1: Free, F2: Subunit, F3: Monosome, F4: Polysome). Each dot represents a base, with significantly altered Ψ sites highlighted in red. The remaining Ψ sites are shown in black and the rest of the sites in gray. All rRNA bases from cytosolic rRNAs are included in the analysis and plots. **(C)** Current intensity density plots showing altered current intensity distribution in positions 25s:Ψ2826 (left) and 25s:Ψ2880 (right), in the wild-type-strain strain (which will correspond to a Ψ-centered k-mers) and snR34 knockout strain (which will correspond to U-centered k-mers). Current intensity distribution of the equivalent 5-mers with C in the middle position are shown in blue (C).