# Concerted modification of nucleotides at functional centers of the ribosome revealed by single-molecule RNA modification profiling

Andrew D. Bailey IV[1,6], Jason Talkish[2,6], Hongxu Ding[1,3], Haller Igel[2], Alejandra Durán[4], Shreya Mantripragada[5], Benedict Paten[1] and Manuel Ares, Jr.[2]

[1]Department of Biomolecular Engineering and Santa Cruz Genomics Institute, UC Santa Cruz, Santa Cruz, California, USA.
[2]RNA Center and Department of Molecular, Cell & Developmental Biology, UC Santa Cruz, Santa Cruz, California, USA.
[3]Department of Pharmacy Practice & Science, College of Pharmacy, University of Arizona, Tucson, Arizona USA.
[4]Colegio Santa Francisca Romana, Bogotá, Colombia
[5]Monta Vista High School, Cupertino, California, USA.
[6]These authors contributed equally to this work.

*Correspondence should be addressed to A.B. (andbaile@ucsc.edu), B.P. (bpaten@ucsc.edu) or M.A. (ares@ucsc.edu).

Highlights

- A method enabling single-molecule profiling of RNA modifications is developed and reveals heterogeneous classes of modified ribosomes.
- Most rRNA 2´O methylation and pseudouridylation modifications are independent of each other.
- Nucleotides in functional centers of the ribosome are modified in a concerted fashion.
- Loss of function for RNA helicases Dbp3 and Prp43 produce discrete overlapping subpopulations of incompletely modified ribosomes.
- RNA modification profiles are resilient to changing nutrient conditions and perturbation of translation

**ABSTRACT**

Nucleotides in RNA and DNA are chemically modified by numerous enzymes that alter their function. Eukaryotic ribosomal RNA (rRNA) is modified at more than 100 locations, particularly at highly conserved and functionally important nucleotides. During ribosome biogenesis, modifications are added at various stages of assembly. The existence of differently modified classes of ribosomes in normal cells is unknown because no method exists to simultaneously evaluate the modification status at all sites, within a single rRNA molecule. Using a combination of yeast genetics and nanopore direct RNA sequencing, we developed a reliable method to track the modification status of single rRNA molecules at 37 sites in 18S rRNA and 73 sites in 25S rRNA. We use our method to characterize patterns of modification heterogeneity and identify concerted modification of nucleotides found near functional centers of the ribosome. Distinct undermodified subpopulations of rRNAs accumulate upon loss of Dbp3 or Prp43 RNA helicases, suggesting overlapping roles in ribosome biogenesis. Modification profiles are surprisingly resistant to change in response to many genetic and environmental conditions that affect translation, ribosome biogenesis, and pre-mRNA splicing. The ability to capture single molecule RNA modification profiles provides new insights into the roles of nucleotide modifications in RNA function.

## INTRODUCTION

In addition to the four standard nucleotides, there are more than 160 distinctly modified ribonucleotides and more than 50 distinctly modified deoxyribonucleotides found in the RNA and DNA of cells (Boccaletto et al., 2018; Jonkhout et al., 2017; Sood et al., 2019). Many of these modified nucleotides provide extra regulatory information and are crucial for cell function. Irregular DNA methylation patterns are linked to several cancers, neurological disorders and autoimmune diseases (Portela and Esteller, 2010; Raiber et al., 2017). Aberrant RNA modification has been linked to the development of cognitive functions, neurological defects, breast cancer, genetic birth defects and diabetes (Bednářová et al., 2017; Benrahma et al., 2014; Delaunay et al., 2016; Delaunay and Frye, 2019; Hideyama et al., 2012; Wu et al., 2005). In ribosomal RNA (rRNA), extensive and highly conserved modifications are vital for correct ribosome structure and function (Polikanov et al., 2015; Sloan et al., 2017). Modifications on rRNA have been generally considered to be constitutive in support of fine tuning function (Karijolich et al., 2010) rather than mediating specific regulatory changes in ribosome function. However, the fraction of rRNA molecules modified at individual positions can change in response to the environment, disease and developmental state (Birkedal et al., 2014; Sloan et al., 2017; Taoka et al., 2016), and evidence is accumulating that ribosome heterogeneity may influence translation (Gay et al., 2021; K. Liu et al., 2021). It seems possible that modification status throughout locations in the ribosome could control translation by creating functional heterogeneity in the cell's pool of ribosomes.

One technical challenge in analyzing the effect of modification on the function of RNA is that no method is available to capture the modification state at all positions of a single RNA molecule. Traditional approaches examine ensembles of molecules and estimate the fraction modified at individual sites independently. For example, non-sequencing based techniques such as liquid chromatography-tandem mass spectrometry (LC-MS/MS)(Taoka et al., 2009) and cryogenic electron microscopy (cryo-EM)(Natchiar et al., 2017) can identify the presence of all types of modified nucleotides. Some methods such as immunoprecipitation-seq(Dominissini et al., 2012) or mismatch-seq(Ramaswami et al., 2013), aggregate information from several reads to detect modifications at a specific site, but do not capture associations between modification status at distant sites in large RNA molecules. Other approaches such as bisulfite-seq (Frommer et al., 1992), Ψ-seq ("psi-seq") (Schwartz et al., 2014), and RiboMeth-seq (Birkedal et al., 2014) are highly specific for a single type of modified nucleotide, but also require fragmentation of RNA, preventing capture of modification status at multiple distant sites in an RNA. Such whole molecule information is necessary to assess the relationship between function and modification status of individual ribosomal subunits.

New advances in direct single-molecule sequencing of RNA using nanopore technology may circumvent many of these limitations. Direct nanopore sequencing of full-length RNA molecules (Deamer et al., 2016; Garalde et al., 2016) has the potential to report modification status across entire RNA molecules without chemical treatment or amplification steps. Modified nucleotides produce changes in electrical current distinct from canonical nucleotides, permitting modification detection algorithms to identify modifications in both DNA and RNA. Given enough training data, basecalling algorithms can predict both canonical nucleotides and modifications directly from the signal (Wick et al., 2019). However, training data for most modifications is limited, and thus many algorithms rely on aligning reads to a reference sequence to identify modified nucleotides (Furlan et al., 2021). Current signals can be modeled using secondary features like quality scores and base miscalls (Begik et al., 2021; H. Liu et al., 2021) or directly using the underlying signal (Rand et al., 2017; Simpson et al., 2017; Stoiber et al., 2017). However, no currently available method captures combinations of distinctly modified nucleotides at multiple distant sites in RNA.

Here we demonstrate accurate, single molecule modification profiling of 13 distinct types of modified nucleotides at 110 positions across full transcripts of 18S and 25S rRNA from *S. cerevisiae*. We preserve long-range associations between modification status at distant positions on single RNA molecules, allowing us to identify highly correlated positions and explore heterogeneity in ribosomal RNA modification. Clustering analysis separates populations of distinctly modified ribosomes in wild type yeast, as well as in yeast deficient in gene functions required for modification including Cbf5, Nop58, several snRNAs, RNA helicases Dbp3, Prp43 and their G-patch protein partners. These studies provide evidence that groups of nucleotides are modified in a concerted manner, especially at functional centers of the ribosome. Further application of single molecule modification profiling will enable dissection of the contributions of nucleotide modification to the function of large RNAs.

## RESULTS

### *Profiling rRNA modifications at single molecule resolution*

To investigate the overall modification status of yeast rRNA on a single molecule level, we used nanopore current traces from Oxford Nanopore flow cells (see Methods) of complete rRNA transcripts to capture modification status at every modified position along individual molecules. To create these single molecule profiles, we trained signalAlign (Rand et al., 2017) by modeling wild type rRNA reads as modified and in vitro transcribed (IVT) reads as unmodified to detect all 110 annotated modifications in

*S. cerevisiae* 18S and 25S rRNA (Stoiber et al., 2017; Taoka et al., 2016) (See Methods, Suppl. Fig. S1, Suppl. File S1). For each rRNA read, the model estimates the probability of modification, regardless of modification type, at each annotated position and outputs a list of modification probabilities for each full length rRNA read (Suppl. Fig. S2A).

We wondered whether different genetic and environmental conditions would alter modification profiles to reveal distinct subpopulations of ribosomes. To test the ability of the trained model to capture single molecule modification profiles in yeast suffering catastrophic depletion of rRNA modifications, we isolated RNA under conditions in which either of two main classes of snoRNA-guided modifications is inhibited. In *S. cerevisiae*, 34 of the 37 18S and 66 of the 73 25S rRNA annotated modifications are guided by the C/D box (2´O-methylation) and H/ACA box (pseudouridylation) snoRNPs (Piekna-Przybylska et al., 2007; Tollervey and Kiss, 1997; Yang et al., 2016). To ablate these modifications *en masse* we used strains in which Nop58 (core component of C/D snoRNPs) or Cbf5 (H/ACA snoRNP pseudouridylase) can be depleted using a *GAL1* promoter (Lafontaine and Tollervey, 1999; Watkins et al., 1998). Thus, in galactose-grown cells shifted to glucose medium, Nop58 (or Cbf5) expression will be repressed, leading to loss of C/D box (or H/ACA box) snoRNPs, blocking modification (Lafontaine et al., 1998; Lafontaine and Tollervey, 1999). Under these conditions, single molecule modification profiles produced by our model reveal accumulation of large numbers of rRNA molecules lacking most 2´O-methyl (Nm) (Nop58-depleted) or pseudouridine (Ψ) (Cbf5-depleted) modifications at snoRNA-guided positions in 18S (Suppl. Fig. S3) or 25S rRNA (Fig. 1).

To examine subpopulations of modified rRNA molecules in these cells, we performed hierarchical clustering. After pooling profiles from the separate samples and clustering them, we observe clear separation of subpopulations representing wild type rRNA, IVTs, and molecules arising from the Nop58 or Cbf5-depleted cells (Fig. 1A, Suppl. Fig. S3A). We used dimension reduction UMAP visualization (McInnes et al., 2018) of 18S and 25S rRNA modification profiles to confirm the presence of these distinct molecular populations (Fig. 1C, Suppl. Fig. S3C). Comparing clusters 1 and 3 derived from the snoRNP-depleted cells (Fig. 1A and C) suggests that 2´O methylation is largely independent from pseudouridylation. Some molecules from snoRNP-depleted cells appear in cluster 2 with wild type rRNA (Fig. 1A and C), more often for 18S rRNA than for 25S rRNA (compare to Suppl. Fig. S3). This difference may reflect a more severe impact of modification loss on accumulation of 18S rRNA (Lafontaine et al., 1998; Lafontaine and Tollervey, 1999), leading to a higher proportion of residual ribosomes modified before depletion was complete. We conclude that clustering of single molecule rRNA modification profiles reveals two large but distinct classes of

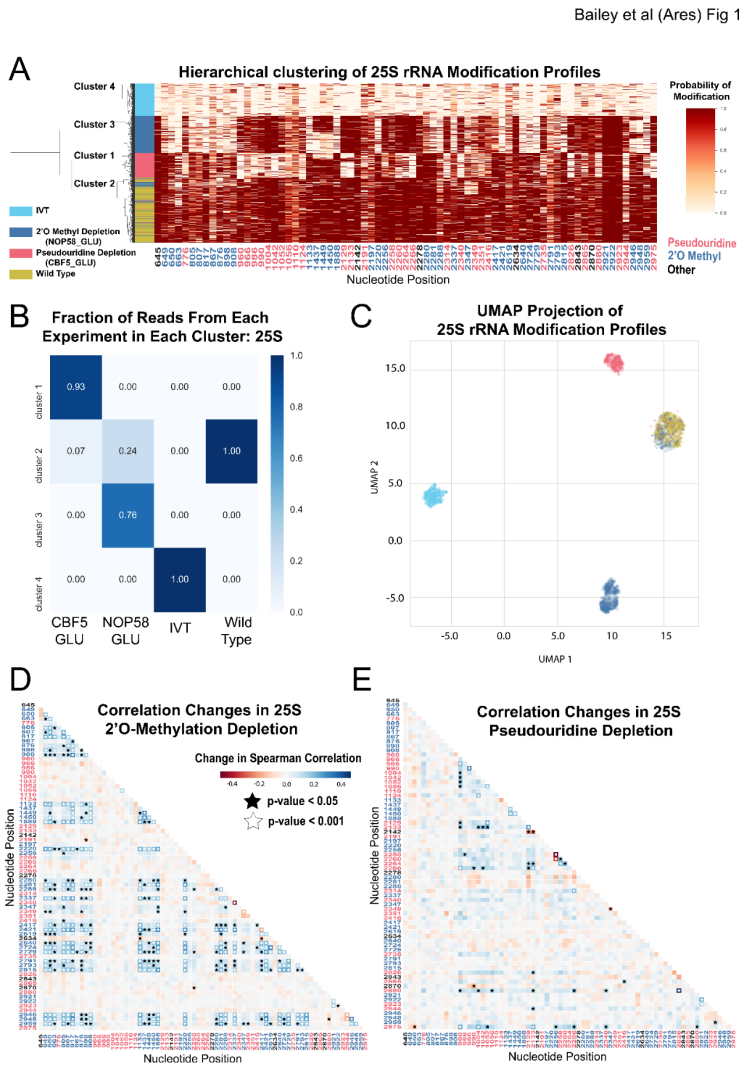


**Figure 1: Clustering and correlation analysis of depletion experiment modification profiles in 25S.** (A) Hierarchical clustering of 25S yeast rRNA modification profiles of IVT, wild type, and both pseudouridine and 2´O methyl depletion experiments (1000 reads/experiment). Each row is a full length molecule, each column is a modified nucleotide and the color represents modification probability, see scale. (B) Fraction of reads from IVT, wild type and both depletion experiments in each cluster of 25S rRNA. (C) UMAP visualization of 25S yeast rRNA modification profiles of IVT, wild type, and both pseudouridine and 2´O methyl depletion experiments. UMAP color scheme is the same as the labels in panel A. (D,E) Change in Spearman correlations of 25S reads in 2´O methyl depletion (D) and pseudouridine depletion (E) when compared to wild type. Stars represent significant changes when compared to wild type correlation (see Methods). Nucleotide positions are blue for 2´O-methyl, red for pseudouridine, and black for other (neither 2´O-methyl nor pseudouridine).

undermodified rRNA molecules induced by depletion of each of the two main classes of snoRNPs.

A powerful advantage of single-molecule modification profiling is the ability to measure concerted changes between modifications within the same RNA. To test this, we measured the change in Spearman correlation between pairs of modified sites, in the absence of Cbf5 or Nop58, compared to wild type rRNA (Fisher z-transform test and Brown's method, see Methods, Fig. 1D and E, Suppl. Fig. S3E-F, Suppl. Table S1). In the absence of Cbf5 and Nop58, we observe that the correlated loss of pseudouridine (p-value=5.5e-05, Brown's method) and 2´O-methyl (p-value=1.5e-16, Brown's method) modifications is highly significant, respectively. (Fig. 1D-E, Suppl. Fig. S2C-D). Surprisingly, a comparison of the pairwise tests for all combinations of modified positions confirmed that, to a large extent, 2´O-methylation is independent of pseudouridylation in yeast rRNA (Fig. 1D and 1E).

Several subpopulations of molecules displayed unusual patterns of modification. The sites of methylation guided by the C/D box snoRNA U24, Cm1437, Am1449, Gm1450 within the polypeptide exit tunnel (PET) in 25S rRNA, appear to be modified in a concerted fashion. Approximately 40% of the molecules from the Nop58 depletion remain methylated at all three sites despite being unmethylated at most other sites, (Fig. 1A), suggesting that either the modification of this triad is more efficient, or that rRNAs lacking modification here are more unstable. Furthermore about 13% of molecules in the cluster formed by depletion of the pseudouridylase Cbf5 are also unmethylated at all three sites, suggesting that concerted methylation at these positions may be partly dependent on pseudouridine modification. Particularly striking is the highly concerted modification between 25S rRNA positions Um2921 with Gm2922 and Ψ2923 in the peptidyl transfer center (PTC). These appear refractory to loss of modification in both depletion experiments, remaining modified on a large fraction of molecules otherwise lacking multiple other modifications (Fig. 1A). Modification of Um2921 is guided by C/D box snoRNA snR52, and Gm2922 is modified by the non-snoRNP methyltransferase Spb1, which can also methylate Um2921 in the absence of snR52 (Lapeyre and Purushothaman, 2004). The extremely low numbers of ribosomes unmodified at these important positions suggests their modification may be essential for rRNA stability.

We also observe a concerted change in loss of modification between the two N4-acetylcytidines (ac4C1280 and ac4C1773) in 18S and loss of 2´O-methylation at many positions in the Nop58 depletion (p-value = 2.3e-05, Brown's method) (Suppl. Fig. S3A and E). N4-acetylcytidine modification depends on the C/D box snoRNAs snR4 and snR45, which do not guide methylation, but instead bring the cytidine acetylase

Kre33 to positions C1280 and C1773, respectively (Sharma et al., 2017; van Nues et al., 2011). These atypical C/D box snRNAs also require Nop58, explaining the coordinate loss of cytidine acetylation and 2´O-methylation. We confirmed that our model recognizes ac4C modified sites by knocking out each snoRNA (Suppl. Fig. S4A, B). Also, the N1-methyl-N3-aminocarboxypropyl-pseudouridine (m1acp3Ψ1191) residue in 18S is significantly correlated with pseudouridine positions in the Cbf5 depletion (p-value = 5.4e-07, Brown's method, Suppl. Fig. S3A and F), as expected given that snR35-guided pseudouridylation of U1191 is the first step to generate this complex modification (Brand et al., 1978).

The concerted modifications in the PTC (Um2921, Gm2922, Ψ2923) and the PET (Am1449, Gm1450) occur on sequential nucleotides at each location. Given that our model incorporates information from sequential 5-mers that contain the position under consideration, pairs of modifications close in sequence share 5-mers presenting a challenge to the prediction algorithm. To examine the extent to which the model may be affected by closely spaced modifications in these regions, we examined the underlying current signal directly, by clustering the single molecule signal means (see Methods) (Ding et al., 2020). This test reveals that the concerted pattern of modifications predicted by the model agree with the underlying event means clustering, with the exception that modification of Ψ2923 may be slightly overestimated (Suppl. Fig. S5, see Suppl. Note S1). We conclude that single molecule modification profiling allows identification of subpopulations of individual rRNA molecules and captures concerted patterns of modification at functionally important sites in the ribosome.

***Resolving subpopulations of ribosomes that differ at a single modified site***

The global loss of modification by depletion of snoRNPs creates grossly undermodified rRNA molecules that are easily distinguished by profiling. To test the sensitivity of the method to resolve classes of ribosomes that differ by a single modification, we first estimated variation in wild type rRNA profiles that could arise from a combination of experimental noise, behavior of the model, or biological variation in modification. We calculated the variance of predictions for each annotated modification for three wild type replicates. Based on the largest variance (position 562 in 18S, ~9%), we chose a conservative cutoff of a 10% change in modification to call any site affected by a given experimental perturbation (mutation or treatment, Suppl. Fig. S4B), despite the fact that many sites were never observed to vary more than a few percent in any experiment. We also compare the predicted modification frequency at a given site in an experiment to its predicted frequency in wild type using a chi-square test (see Methods, Suppl. Table S2). We sequenced rRNAs from strains containing individual snoRNA knockouts (snR80, snR83, snR87, snR4, and snR45) previously shown to completely

lack modification at one or a few annotated sites in each case (Piekna-Przybylska et al., 2007), demonstrating significant decreases in modification at the appropriate site for each (p-values<1e-04, chi-square test, Suppl. Fig. S4B). As expected, signal distributions from snoRNA knockout kmers match the model's canonical unmodified kmer distributions (Suppl. Fig. S6). This experiment confirms our ability to identify loss of modification at single locations with high confidence.

To test the limits of our ability to deconvolute a heterogeneous mixture of ribosomes differing by single modifications, we pooled equal amounts of total RNA from three snoRNA knockout strains (snR80, snR83 and snR87) and wild type, and acquired single molecule modification profiles from the mixture (Fig. 2). Hierarchical clustering of the profiles obtained from the mixture reveals four similarly sized main clusters of differently modified 18S rRNA (Fig. 2A and C, see Methods). We observe positive correlation changes between positions Ψ1290 and Ψ1415 (Fig. 2B, p-value = 6.4e-16, Fisher z-transform test), that likely arise from their shared dependence on snR83. Moreover, we observe negative correlation changes between all independent pairs of modification sites at Am436, Ψ759, and [Ψ1290+Ψ1415] (Fig. 2B, p-value=3.5e-08, Brown's method). Since loss of modification at these positions in this reconstructed mixture arose independently, the population would be extremely unlikely to have molecules with simultaneous loss of modification at any pair of these three locations, creating the negative correlation change. Together, this analysis establishes the ability of the method to resolve minority subpopulations of ribosomes with subtly different modification profiles.

***Correlated modification at distant sites on wild type yeast ribosomes***

By ensemble methods, most rRNA positions are almost completely (>95%) modified and cluster around functional centers, with a few that are partially modified (Suppl. Fig S7)(Birkedal et al., 2014; Marchand et al., 2016; Taoka et al., 2016; Yang et al., 2016). Taken one position at a time, our model arrives at very similar estimates (Suppl. Fig. S8). However, ensemble methods cannot detect heterogeneity in the modification pattern of whole molecules. To identify modification heterogeneity in wild type yeast ribosomes, we searched for subpopulations of rRNA with single molecule modification profiles not expected by chance. Hierarchical clustering of wild type profiles shows no large classes of distinctly modified ribosomes. However some smaller (<10% of total) subpopulations are apparent that have correlated pairs of unmodified nucleotides (Fig. 2E, Suppl. Fig. S9A), confirmed by comparing wild type to IVT (Fig. 2F, Suppl. Fig. S9B, See Methods). One pair of significantly correlated positions in 18S, Ψ632 and Ψ766 (p-value=1.3e-04, Fisher z-transform test) are guided by the same snoRNA, snR161, possibly explaining the basis for this correlation.

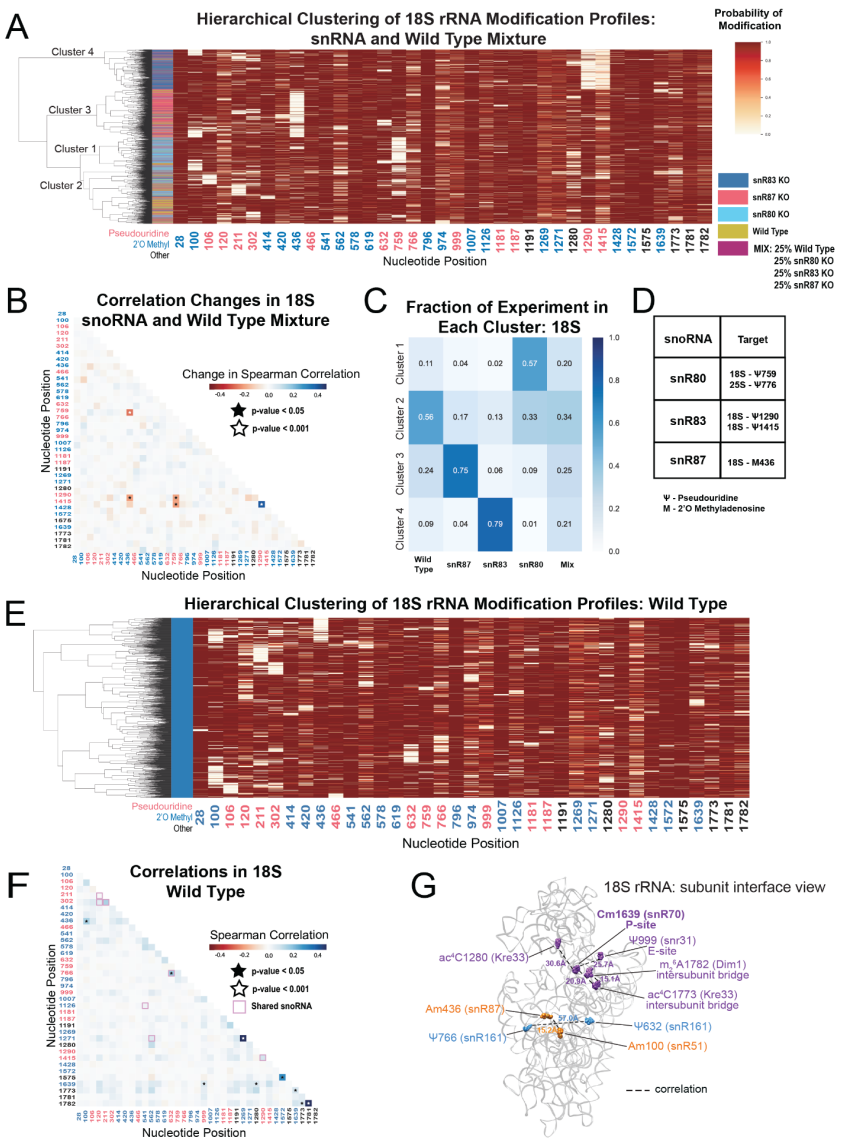


**Figure 2: Clustering of 18S rRNA modification profiles and correlation analysis from the mixture experiment and wild type rRNA.** (A) Hierarchical clustering of 18S modification of profiles from pooled wild type, snR80Δ, snR83Δ, and snR87Δ RNA (500 reads in each experiment). Each row is a full length molecule, each column is a modified nucleotide and the color represents modification probability, see scale. (B) Change in Spearman correlations of 18S reads in the mixture experiment when compared to wild type. Stars represent significant changes when compared to wild type correlation and significantly different from zero correlation (see Methods). (C) Fraction of wild type, snR80Δ, snR83Δ, and snR87Δ profiles in each cluster of 18S rRNA. (D) Table of snoRNAs knocked down with the corresponding expected knocked down modifications. (E) Hierarchical clustering of 18S yeast rRNA modification profiles from wild type yeast (1000 reads). (F) Wild type Spearman correlation of 18S wild type reads. Stars represent significantly different to IVT correlations and significantly different from zero. correlation. (G) Crystal structure model of wild type *S. cerevisiae* 18S rRNA highlighting significant correlated positions. PDB: 4V88 (Ben-Shem et al., 2011).

We observe a significant correlation between Am100 and Am436 (p-value=3.1e-04, Fisher z-transform test) as well as between Cm1639 and $ac^4C1773$ in 18S (p-value=4.5e-06, Fisher z-transform test, Fig. 2F), neither of which share a snoRNA or modification enzyme. Cm1639 (snR70) in the P-site is also correlated with Ψ999 (snR31, E-site, p-value=1.7e-03, Fisher z-transform test), and $ac^4C1280$ (Kre33, p-value=3.9e-05, Fisher z-transform test)(Fig. 2). Furthermore, $ac^4C1773$ and $m_2^6A1782$ (Dim1, intersubunit bridge, p-value=1.7e-04, Fisher z-transform test) are correlated. Some correlated pairs lie close to each other (15-22 Å) in the three-dimensional structure of the mature ribosome (Fig. 2G), suggesting a structural or functional basis for their coordinated modification status. In 25S, positions Cm1437, Am1449 and Gm1450 are all significantly more correlated with each other than expected (Suppl. Fig. S9, p-value=1.0e-44, Brown's method), echoing our observations in the depletion experiment. Several of the significant long range correlations in wild type show up in other experiments (see below), indicating that concerted modification at those positions is a feature of normal yeast ribosomes.

***Loss of different RNA helicase-related functions result in distinct subpopulations of differently modified rRNA molecules***

Previous studies have connected helicase activities required for ribosome biogenesis with changes in 2´O methylation at single positions in ensembles of rRNA molecules (Martin et al., 2013). To explore how ATP-dependent RNA helicase functions implicated in ribosome biogenesis and snoRNP dynamics may influence single molecule patterns of rRNA modification, we profiled cells deficient in two proteins with distinct roles in ribosome biogenesis; Dbp3 (Weaver et al., 1997) and Prp43 (Combs et al., 2006; Leeds et al., 2006). We observe loss of 2´O methylation at specific locations in 18S and 25S rRNAs using a Dbp3 knockout strain (*dbp3*Δ) or a cold-sensitive Prp43 Q423N mutant (*prp43-cs*) grown at nonpermissive temperature (Fig. 3A, Suppl. Fig. S10A) consistent with previous ensemble studies (Suppl. Fig. S7C-D) (Aquino et al., 2021). Despite the numerous locations at which modification is compromised, hierarchical clustering of 25S rRNA single molecule profiles reveals that just 2-3 distinct but related sets of modification profiles describe the vast majority of ribosomes in both experiments (Fig. 3B). The triad of 2´O methylations guided by the snoRNA U24 at 25S positions 1437, 1449 and 1450 surrounding the PET are often left unmodified in a highly concerted manner (Fig. 3B), as observed in a minority of wild type 25S rRNA molecules (Suppl. Fig. S9), and in the snoRNP depletion experiments (Fig. 1). The pairwise correlations within this triad are significantly higher in both the *dbp3*Δ and the *prp43-cs* mutants relative to wild type (Fig. 4, Suppl. Fig. S11D-E, *dbp3*Δ p-value = 3.6e-68, *prp43-cs* 3.5e-11, Brown's method). To confirm the highly concerted modification of the

Bailey et al (Ares) Fig 3

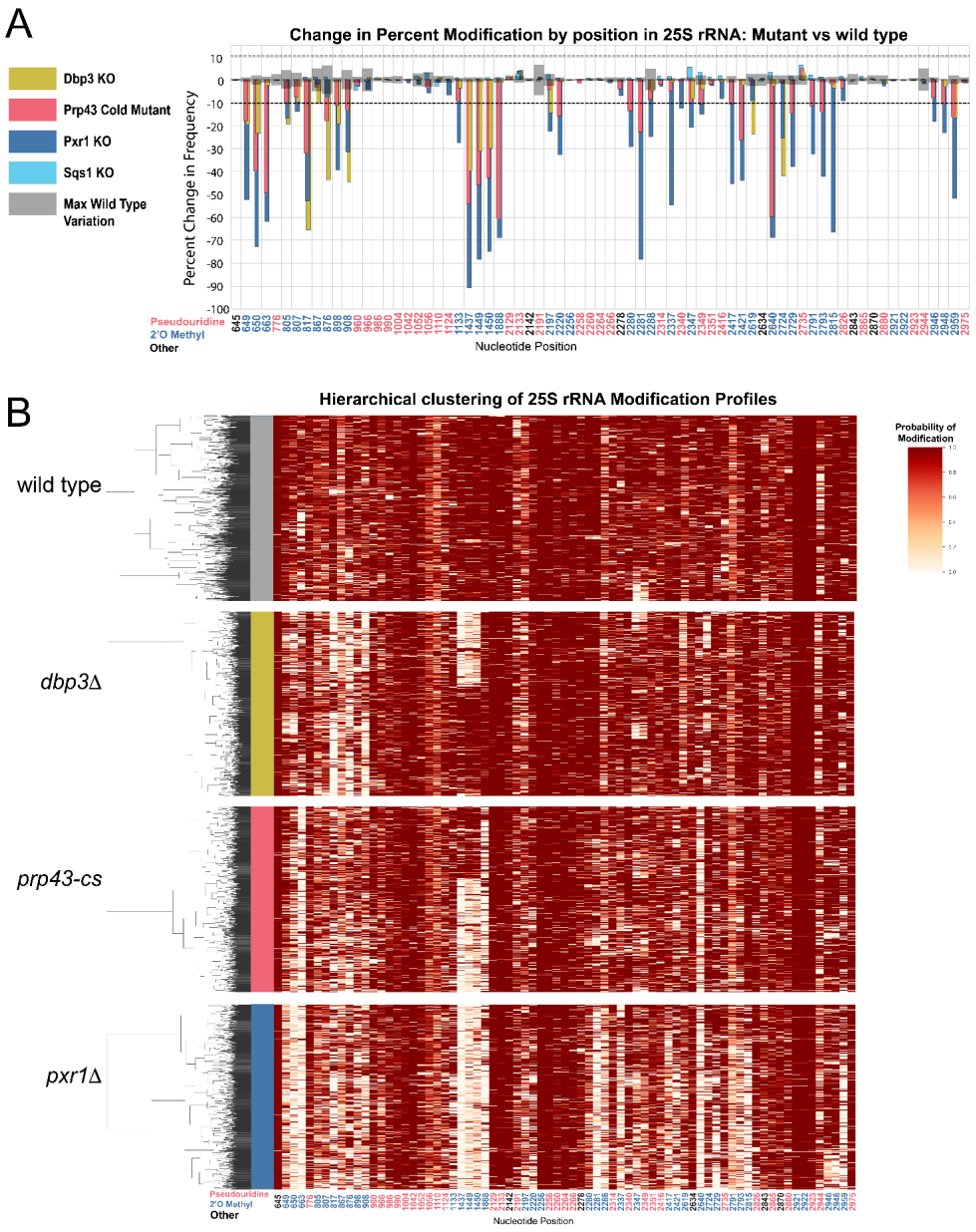


**Figure 3: Clustering of 25S rRNA modification profiles and percent change in modification frequency of mutant helicases Dbp3 and Prp43 and G-patch proteins Pxr1 and Sqs1.** (A) Barplots of the difference between wild type modification frequency and *dbp3*Δ, *prp43-cs*, *pxr1*Δ, and *sqs1*Δ modification frequencies in 25S yeast rRNA. Gray bars indicate the variance of wild type rRNA modification at each position and the black dotted lines represent the maximum variance observed at any site. (B) Hierarchical clustering of 25S yeast rRNA modification profiles from wild type, *dbp3*Δ, *prp43-cs*, and *pxr1*Δ (1000 reads in each experiment). Each row is a full length molecule, each column is a modified nucleotide and the color represents modification probability, see scale.

closely spaced U24-dependent positions, we clustered the underlying raw signal event means from the *dbp3*Δ and *prp43-cs* mutants as above (see Methods)(Ding et al., 2020). This reveals two clear subpopulations of reads distinguished by the signal means at positions C1437, A1449, and G1450 in both the *dbp3*Δ and *prp43-cs* mutants (Suppl. Fig. S12). Therefore, if the U24 snoRNP is unable to guide modification of any of these positions, then all 3 positions are left unmodified. Furthermore both Dbp3 and Prp43 functions are required for the concerted modification of these nucleotides near the PET.

Prp43 interacts with a number of G-patch proteins that direct it to either the ribosome or the spliceosome (Chen et al., 2014; Heininger et al., 2016; Martin et al., 2002; Pertschy et al., 2009; Tanaka et al., 2007; Tsai et al., 2005). Two of these, Pxr1 and Sqs1, are important for correct pre-rRNA processing (Banerjee et al., 2015; Guglielmi and Werner, 2002; Pertschy et al., 2009). To discover the role of Pxr1 and Sqs1 in rRNA modification, we obtained modification profiles from strains deleted for each. Interestingly, deletion of Sqs1 had little effect on modification (see below). In contrast, loss of Pxr1 produced modification profiles similar in pattern to the *prp43-cs* mutant, but with more extreme modification deficiencies (Fig. 3A). All modifications affected by *prp43-cs* and all but two 18S 2´O methylations affected by *dbp3*Δ (Suppl. Fig. 10A) are also affected in *pxr1*Δ. This suggests that loss of Prp43 activity guided by Pxr1, but not that guided by Sqs1, is responsible for the concerted changes in modification pattern observed in the *prp43-cs* strain.

Despite the similarities in modification patterns in the different mutants (Fig. 4, Suppl. Fig. 11), they are not identical. For example in 25S, positions Am817 and Gm908 (both guided by snR60) and Gm2619 and Um2724 (both guided by snR67) are significantly more correlated in both *pxr1*Δ and *dbp3*Δ relative to wild type (817-908 p-value = 4.7e-17 and 2.4e-33, 2619-2724 p-value = 7.2 e-04 and 3.5e-22, Fisher z-transform tests, Suppl. Fig. S11D). However, they are not significantly more correlated in the *prp43-cs* mutant relative to wild type (817-908 p-value=0.34, 2619-2724 p-value=0.919, Fisher z-transform test). Pxr1 and Dbp3 both appear to promote concerted modification of positions guided by snR60 and snR67, however the contribution of Prp43 is less clear. It is possible that the conditional Prp43 mutation is not severe enough to produce a strong block to modification at those sites, or alternatively that Pxr1 has functions in addition to its role in support of Prp43.

Together our data reveal distinct classes of ribosomes that result from a concerted network of modification losses, many of which reside in the functional centers of the ribosome (Fig 4). For example, loss of Prp43 and Pxr1 induce a concerted loss of modification of a set of nucleotides in the decoding site of the small subunit (green circles, Fig. 4B). Loss of Pxr1 leads to concerted loss of a set of modifications in the

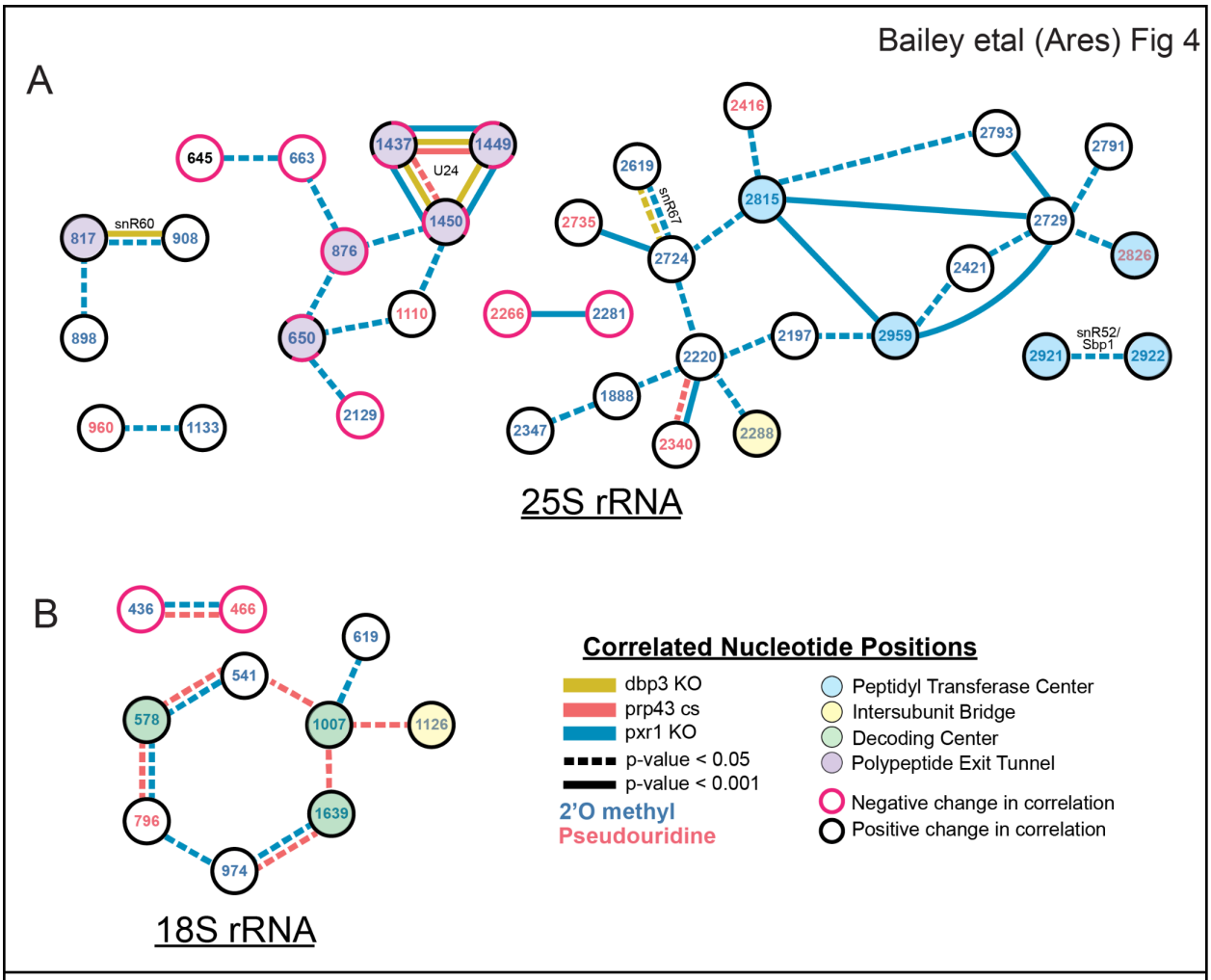


**Figure 4: Changes in correlated nucleotide positions in *dbp3*Δ, *prp43-cs*, or *pxr1*Δ mutants.** Pairs of correlated nucleotide changes (nodes) are shown for each mutant (edges) relative to wild type yeast 25S rRNA (A) and 18S rRNA (B). In cases where correlated pairs show differential changes in correlation in different mutants (eg. U24 modifications), node color rings are fragmented with the appropriate mutant edge connecting to either the magenta (negative change in correlation) or black (positive change in correlation) portion of the node.

peptidyl transfer center of the large subunit (blue circles, Fig. 4A). And all three mutants create a complex set of correlated modification changes in the triad Cm1437, Am1449, and Gm1450 near the protein exit tunnel of the large subunit (purple circles, Fig. 4A). Concerted modification of this triad is observed in wild type ribosomes (Suppl. Fig. S9) as well as in the snoRNP depletion experiments (Fig. 1). As discussed above, a shared snoRNP (e. g. snR60, snR67) may explain part of the concerted modification phenomenon, however in the majority of cases the elements that underlie concerted modification are not obvious.

***Resilience of rRNA modification profiles to splicing perturbations and environmental treatments***

Ribosome biogenesis and splicing are connected processes in yeast. For example, Prp43 also mediates disassembly of spliceosomes (Combs et al., 2006; Leeds et al., 2006; Martin et al., 2002) aided by the G-patch protein Spp382 (also called Ntr1)(Christian et al., 2014; Fourmann et al., 2016; Pandit et al., 2006; Tanaka et al., 2007; Tsai et al., 2005). In addition, a number of snoRNAs are encoded within introns of genes important in ribosome biogenesis and translation, and their synthesis can be compromised by mutations that affect splicing (Ooi et al., 1998; Petfalski et al., 1998; Piekna-Przybylska et al., 2007; Vincenti et al., 2007). To determine if loss of splicing functions affect rRNA modification, we acquired single molecule modification profiles for ribosomes from an Spp382/Ntr1 mutant, a cold sensitive mutant of Prp16 (*prp16-302*) (Madhani and Guthrie, 1994; Tseng et al., 2011) that accumulates splicing intermediates, and a deletion of Dbr1 that prevents debranching of the intron lariat (Chapman and Boeke, 1991), a reaction that promotes processing of some intronic snoRNAs, in particular U24 (Ooi et al., 1998).

Using the threshold of >10% change in modification relative to wild type established above (Fig. 3A, Suppl. Fig. S5), we examined splicing-related perturbations for effects on rRNA modification that might be mediated through loss of one or more intronic snoRNAs (Fig. 5). We observe a 36.8% reduction in modification for 18S Ψ106 (guided by snR44 from intron 2 of *RPS22B*) and an 11.0% reduction in modification frequency for 18S Am974 (guided by snR54 from intron 1 of *IMD4*) as a consequence of the loss of Dbr1. There are hints that other intron-encoded snoRNAs may be affected, for example U24, however none of these reached our conservative threshold (Fig. 5). Neither the *prp16-302* nor *spp382-1* mutant produced a modification defect that passed the threshold (Fig 5), suggesting that deficiencies in ongoing splicing do not create noticeable changes in rRNA modification pattern. Both the *prp16-302* and the *spp382-1* mutant alleles are viable hypomorphs of essential genes, and might not be sufficiently severe to trigger steady state loss of snoRNAs or other perturbations we could detect. In the case of the Dbr1 null strain, there are alternative snoRNA maturation pathways that are independent of splicing, for example via Rnt1 (Grzechnik et al., 2018), and incompletely processed U24 still can guide modifications at its corresponding locations (Aquino et al., 2021; Ooi et al., 1998; Petfalski et al., 1998). These results argue that the modification defects observed for prp43-cs arise primarily from its deficiencies in promoting ribosome biogenesis, and not from the loss of its contributions to splicing.

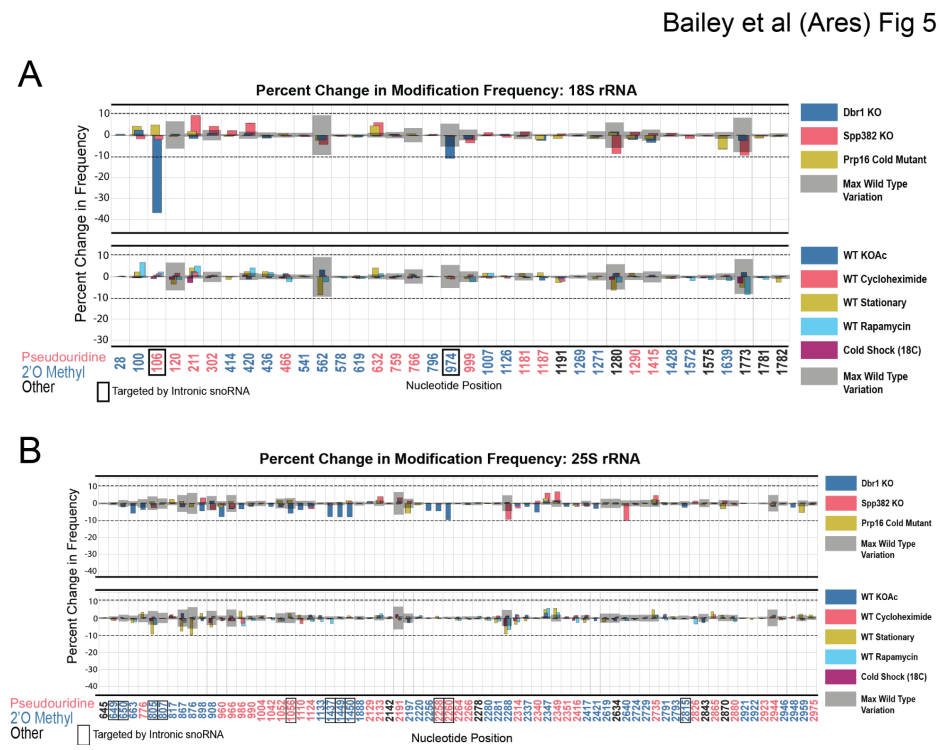


**Figure 5: Resilience of yeast rRNA modifications to a variety of splicing mutants and experimental conditions.** Barplots of the difference between wild type modification frequency and *dbr1*Δ, *spp382-1*, *prp16-302*, KOAc treated, cycloheximide treated, stationary, rapamycin treated and cold shock yeast modification frequencies in yeast 18S (A) and 25S (B) rRNA. Gray bars indicate the variance of wild type rRNA modification at each position and the black dotted lines represent the maximum variance across sites.

To test whether single molecule modification profiles were altered by environmental and growth conditions known to affect ribosome function and biogenesis (Gasch, 2000; Warner, 1999), we isolated and acquired profiles of rRNA isolated from cells at stationary phase (Ju and Warner, 1994; Talkish et al., 2016), after a 1 hour shift to potassium acetate to induce starvation, treated with rapamycin (TOR kinase inhibitor) for 1 h to block nutrient signaling (Cardenas et al., 1999; Hardwick et al., 1999; Heitman et al., 1991; Honma et al., 2006; Powers and Walter, 1999; Talkish et al., 2012; Vanrobays et al., 2008), treated with cycloheximide to block translational elongation (Obrig et al., 1971) and create ribosome collisions (Simms et al., 2017), or after cold shock. In none of these treatments did we detect substantial changes in modification profile (Fig. 5, Suppl. Tables S1 and S2).

Together, these observations indicate that most annotated modification patterns on rRNA are refractory to rapid alterations by dramatic changes in the physiological

conditions we tested. Because pre-existing snoRNPs and ribosomes are stable over time, rapid changes in modification in acute conditions would likely rely on enzymes that reverse the modification at pre-existing modified sites. At this time, there are no known enzymes that would reverse either pseudouridylation or 2´O methylation of ribose in RNA. However, these results do not exclude the possibility that chronic genetic or environmental conditions regulate rRNA modification and ribosome function.

## DISCUSSION

To discover relationships and dependencies between distant modifications in rRNA, we captured single molecule modification profiles of *S. cerevisiae* 18S and 25S rRNA. By depletion of the two main classes of snoRNPs responsible for the bulk of rRNA pseudouridylation and 2´O methylation, we validated the framework for our method and found that these two classes of snoRNP-dependent modification are largely independent of each other (Fig. 1). We resolved populations of ribosomes that differ by a single modification, and identified instances of concerted modification of sets of nucleotides in the wild type ribosome population (Fig. 2). Analysis of the single molecule modification profiles resulting from loss of two distinct helicase activities provided by Dbp3 and Prp43 (and its G-patch protein Pxr1) reveals a complex set of long-range concerted effects on modification, with implications for ribosome biogenesis and function (Figs. 3, 4 and 6). Finally, we found that most annotated modifications are refractory to changes in physiological conditions, inhibitors of ribosome function, or stress on splicing, which is critical to ongoing ribosome biogenesis (Fig. 5) (Warner, 1999). These results provide new perspectives on ribosome heterogeneity as represented by RNA modification patterns, and open a path to single molecule analysis of RNA modification for other classes of RNA.

### *Concerted modification in the polypeptide exit tunnel*

Distinct subsets of 25S rRNAs collectively missing 2´O-methyl modification at positions Cm1437, Am1449, and Gm1450 accumulate together and form a major subset of ribosomes in Pxr1, Prp43, and Dbp3 mutants (Figs. 3 and 4). Importantly, wild type ribosomes show a similar pattern of concerted modification at these residues (Suppl. Fig. S9), suggesting that these are not exclusively a result of the mutations.The polypeptide exit tunnel (PET), acquires modification in a concerted fashion (Figs. 3, 4, and Suppl. Fig. S9). These three nucleotides line the PET, where interactions with the nascent polypeptide chain can influence protein folding (Choi et al., 2018), within a few angstroms of conserved loops of ribosomal proteins uL4 (rpL4) and uL22 (rpL17) (Ben-Shem et al., 2011). These elements insert into the PET to form the constriction site

and may act as an “exit gate” (Fig. 6A-B) (Nakatogawa and Ito, 2002; Wilson et al., 2020; Zhang et al., 2013). Furthermore, these three positions are in domain 0 of ribosomal rRNA, which acts as a central hub around which the other six 25S rRNA domains fold (Klinge and Woolford, 2019; Petrov et al., 2014). In the absence of Cm1437, Am1449, and Gm1450, the rRNA and the loops of uL4 and uL22 may not be properly positioned, affecting the structure and chemistry of the PET, translation, and protein folding. Future work will be necessary to assess the functional differences between the two populations of ribosomes we observe that differ by concerted modification of these PET nucleotides.

***Distinct and overlapping helicase functions during ribosomes biogenesis***

At least 21 putative RNA helicases drive yeast ribosome biogenesis forward by enabling pre-rRNA rearrangements and releasing snoRNPs (Martin et al., 2013). Recent work (Aquino et al., 2021) supports a model in which Dbp3 and Prp43 release certain snoRNPs to allow subsequent modification of adjacent sites by a second snoRNP. Here, we identify concerted changes in modifications over much longer distances in rRNA when the activity of Dbp3 or Prp43 is compromised, suggesting the regulation of these helicases is not only important for locally overlapping modified positions, but also for modifications separated by large stretches of sequence (Fig. 3, Suppl. Figs. S10 and S11). Furthermore, our work shows that Pxr1, but not Sqs1, is the main G-patch protein important for Prp43 function during rRNA modification.

In the absence of ribosomal helicase-related functions (Prp43, Pxr1, and Dbp3) we observe a large set of overlapping but not identical changes in modified positions. Single molecule profiling reveals distinct hubs of concerted modifications, many of which reside in functional centers of the ribosome (Fig. 3, Fig. 4, Suppl. Fig. S10). In the case of 18S rRNA, a hub of nucleotides are left unmodified in a concerted fashion when Prp43 and Pxr1 activity are compromised, creating a distinct subset of undermodified subunits. Previous work from (Aquino et al., 2021; Bohnsack et al., 2009) showed that when Prp43 is absent from pre-ribosomes, the snR55 snoRNP is retained on pre-ribosomes but its target site, 18S rRNA Um1269, remains modified. This suggests snR55 is able to bind and direct modification, but then is not efficiently released from the pre-ribosome. Consistent with this, the Prp43-Q423N protein is more tightly associated with snR55 and other snoRNAs than wild type Prp43 (Leeds et al., 2006).

Here we show that the Prp43-Q423N mutant protein permits modification of Um1269, but causes concerted loss of modification at nucleotides that surround the Prp43 binding site in the base of helix 44 (Fig. 6C)(Bohnsack et al., 2009). However, there are also a number of nucleotides nearby in three-dimensional space that are

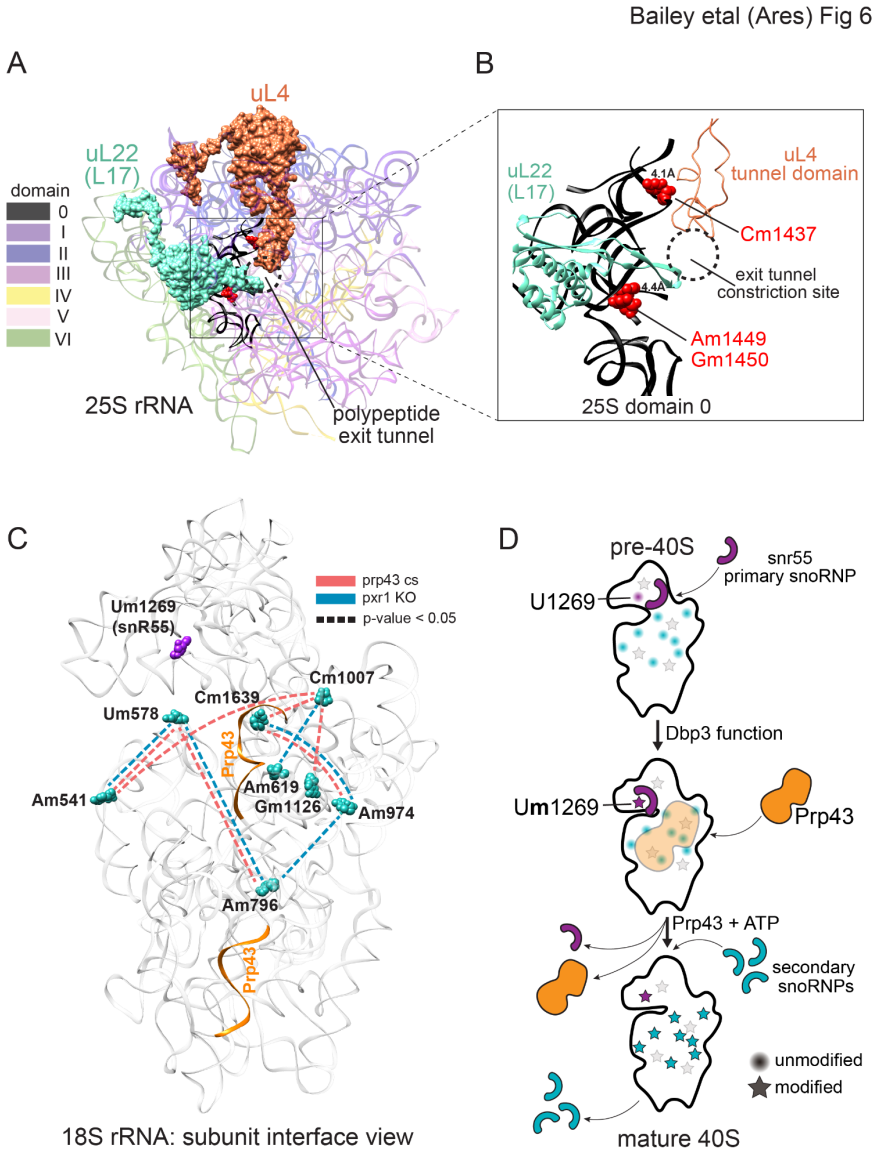


**Figure 6: Models describing the functional basis for concerted modification at specific locations in the ribosome.** (A) Crystal structure model of yeast 25S rRNA and ribosomal proteins uL4 and uL22 in surface view (PDB:4V88)(Ben-Shem et al., 2011). rRNA domains are color coded according to the RiboVision Suite (Bernier et al., 2014). The distal end of the polypeptide exit tunnel is indicated. U24-guided modified nucleotides Cm1437, Am1449, and Gm1450 are shown in blue. (B) Focused view of the uL4 tunnel domain and the internal loop of uL22 forming the exit tunnel constriction sites. 25S rRNA domain 0 is shown in black. 4.1 Å and 4.4 Å represent the distance between the indicated, modified nucleotide and the internal loops of uL4 and uL22 respectively. (C) Crystal structure model of yeast 18S rRNA showing changes in concerted loss of modifications in *prp43-cs*, or *pxr1*Δ mutants when snR55 (Um1269, purple) is retained on the pre-ribosome in the absence of Prp43 (Bohnsack et al., 2009). Prp43 crosslinking sites are indicated in orange. (D) Model showing the hierarchical function of Dbp3 and Prp43, to promote concerted modification of 18S rRNA nucleotides, during ribosome biogenesis. Circles and stars represent unmodified and modified positions, respectively. Modified nucleotides in the 5´ end of 18S rRNA, proposed to occur early and co-transcriptionally, are represented as light grey stars. Color scheme is the same as panel C.

undermodified, but not in concert with the others. These non-concerted nucleotides are in the 5' end of 18S rRNA, and likely to be modified co-transcriptionally (Kos and Tollervey, 2010) prior to snR55 binding. Interestingly, snR55 also remains associated with pre-ribosomes in the absence of Dbp3, but modification of the snR55 target nucleotide U1269 does not occur (Suppl. Fig. S10A)(Aquino et al., 2021). Furthermore, the positions left undermodified in the Prp43 mutant (where the snR55-dependent modification occurs) are modified in the Dbp3 deletion (where the snR55-dependent modification does not occur) (Suppl. Fig. S10) (Aquino et al., 2021). Together, this suggests a hierarchical model for the action of Dbp3 and Prp43/Pxr1 during co-transcriptional ribosome biogenesis (Fig. 6D). We propose that modifications at 5' end of 18S occur early and co-transcriptionally, independently of later modifications. As transcription and ribosome assembly progresses, Dbp3 is necessary for modification of Um1269 by snR55 at 18S Um1269, possibly by promoting a conformational change in the pre-ribosome or snR55. However, the absence of Dbp3 does not prevent snR55 association with the pre-ribosome, nor does it cause snR55 to block modification at positions surrounding the Prp43 binding site (Suppl. Fig. S10). After snr55-mediated modification of U1269, Prp43 binding promotes snR55 release, allowing other snoRNPs to modify positions around the Prp43 binding site. Because the Prp43-Q423N protein is slow to dissociate, snR55 and mutant Prp43 are retained on the pre-ribosome (Bohnsack et al., 2009; Leeds et al., 2006), occluding the other snoRNPs, resulting in concerted loss of those modifications (Fig. 6C and D).

Prp43 is present in multiple, consecutive pre-ribosomal particles and is thought to bind to a number of different positions along the pre-rRNA (Bohnsack et al., 2009; Lebaron et al., 2005), providing opportunity for hierarchical relationships in modification at other locations. During its function in the termination of pre-mRNA splicing, Prp43 is proposed to bind the 3´ end of U6 snRNA and translocate along it, resulting in removal of U6 from the spliceosome (He et al., 2017; Toroney et al., 2019). Because many biochemical principles are shared between the spliceosome and the ribosome (Staley and Woolford, 2009), we propose that Prp43 acts in an analogous manner to release certain snoRNPs from the pre-ribosome upon completion of modification. Since other modification sites in 25S rRNA are affected by loss of Prp43 activity, it seems likely that Prp43 binding at multiple locations promotes release of snoRNPs from the pre-ribosome. Thus, hubs of concerted loss of modification in the Prp43 mutant might reflect critical points in ribosome biogenesis where Prp43 removes key snoRNAs.

***Is U24 snoRNP-dependent methylation distributive or processive?***

The exact mechanism by which snoRNPs bind to and are released from the pre-ribosomal particles during the course of assembly is still unknown. Models must

account for the fact that many snoRNP binding sites overlap in both the linear rRNA sequence and the three-dimensional spaces of the assembling ribosome. In this work we document concerted modification at several sets of locations, some distant locations on the sequence that are near in space in the mature ribosome, and others that may or may not share a snoRNA guide (Figs. 1-4, Suppl. Figs. S3, S9, and S11). The most striking example of concerted modifications are the U24-guided modifications at Cm1437, Am1449, and Gm1450 in the PET discussed above. The depletion experiments (Fig. 1) provide clear evidence that most ribosomes are either fully modified or completely unmodified at those three sites, and the loss of modification in the Dbp3 and Prp43 mutants for this group is similar (Fig. 3, Suppl. Fig. S11). A stochastic mechanism dependent on a single snoRNP that binds its target and dissociates with each round of modification (i. e. a distributive mechanism) should produce a set of rRNAs with partial modifications in different combinations at these positions, and this is not observed (Figs. 1 and 3 and Suppl. Figs. S9 and S11).

The diameter of a snoRNP may approach 100 Å (Duan et al., 2009; Lin et al., 2011), whereas the yeast ribosome is approximately 250 Å in diameter. Thus if two different snoRNPs have overlapping binding sites that guide two closely spaced modifications, they almost certainly must act sequentially. In this instance, concerted loss of modification could occur if the pre-rRNA substrate fails to adopt a conformation that renders the local region containing both sites accessible. However, this may be inadequate to explain concerted modification of closely spaced nucleotides by the same snoRNP, such as the U24-guided modifications at Cm1437, Am1449, and Gm1450 (Fig. 3, Suppl. Fig. S9). In this case if the site is inaccessible, all three positions would be unmodified. However if the site is accessible and modification proceeds distributively, then intermediates representing partial modification should be observed, in particular under depletion of C/D box snoRNPs (Fig. 1) or loss of Dbp3 (Fig. 3)(Aquino et al., 2021). Instead two main populations of ribosomes are observed, all modified or all unmodified at these locations, opening the possibility that modification of these sites by U24 snoRNP is highly processive. In this view, when the site is accessible all three sites become methylated without dissociation of the U24 snoRNP from the pre-ribosome. In Dbp3 mutants and wild type cells, most U24 is associated with the pre-ribosome (Aquino et al., 2021), suggesting that either (1) there are three high affinity binding sites at which release and rebinding of U24 occurs distributively at a rate that prevents detection of partially modified intermediates, or (2) there is a single binding site at which the target nucleotides are bound and methylated processively after a single U24 snoRNP binding event.

### *Prospects and Limitations*

We developed a hidden Markov model-based approach that allows 1) single molecule profiling and clustering of RNAs to visualize high-level relationships between features in a population of complete molecules, 2) the ability to test for changes in correlations between any given pair of modifications on the same molecule, and 3) a way to estimate the fraction of modification of each site. We used wild type rRNA as our fully modified training example, with the clear understanding that not all wild type molecules are fully modified (Taoka et al., 2016). In several instances we confirmed that this assumption had little effect on performance of the model (Suppl. Fig. S2). A second limitation arises when the training samples do not have enough information to resolve dense clusters of modifications. Where this was a concern, we validated patterns of modification by clustering the raw signal means (Suppl. Fig. S5 and S12). While there is some evidence that unknown modified kmer distributions can be estimated using known kmer distributions (Ding et al., 2021), generating more specific modification training data sets that contain all combinations of partially modified closely spaced clusters of nucleotides may be required to produce more accurate and general modification detection algorithms. This is especially true if de-novo detection of modifications within complex sequences is the goal (Leger et al., 2019; Stoiber et al., 2017).

While our analysis focuses on mature rRNA, the capture of and profiling of pre-rRNA intermediates, or chromatin-associated pre-rRNA will certainly further resolve and enhance our understanding of how the concerted placement and timing of rRNA modifications occurs during ribosome biogenesis. Importantly, the model training paradigm we have developed to profile modifications can easily be applied to other nucleic acids of interest. Recent advancements in nanopore sequencing of full length human rRNAs (Jain et al., 2021) and *E. coli* tRNAs (Thomas et al., 2021) make these heavily modified molecules ripe for single molecule profiling to understand how RNA modifications regulate translation (Erales et al., 2017; Krogh et al., 2016; Schimmel, 2018). Furthermore, it is becoming increasingly clear that modifications of pre-mRNAs and the snRNAs within the splicing machinery are important for regulating alternative splicing and misregulation of modifications are associated with disease (Bartosovic et al., 2017; Lin et al., 2016; Zheng et al., 2013). Thus, the ability to monitor the modification status at each position across an mRNA molecule, along with its alternative splicing pattern (Drexler et al., 2020; Tang et al., 2020), will become increasingly important for basic biological understanding and discovery, as well as providing biomarkers for identifying and understanding human disease.

## METHODS

### *Growth of yeast strains*

Yeast strains *GAL-NOP58* and *GAL-CBF5* are described in (Lafontaine et al., 1998; Lafontaine and Tollervey, 1999). Cells were grown at 30 °C in YEPgal liquid medium (2% galactose, 2% peptone, 1% yeast extract) or shifted to liquid YEPD (2% dextrose, 2% peptone, 1% yeast extract) to mid-log phase ($OD_{600}$ = 0.25-0.5) for 16 hours to repress expression of Nop58 or Cbf5. Cells were harvested by centrifugation and RNA was isolated. Unless indicated, all other strains were grown in YEPD at 30 °C to mid-log phase. Cells exposed to various environmental conditions were treated as follows: 1% KOAc (1 hr, 30 °C), cycloheximide (1 ug/ml for 1 hour), or rapamycin (200 ng/ml for 1 or 5 hours). Stationary phase cells were grown to an $OD_{600}$ = 10. Strains carrying *prp16-302* (Madhani and Guthrie, 1994) and *prp43-Q423N* (Leeds et al., 2006) mutations, and wild type (Schattner et al., 2004) were grown to mid log phase at 30 °C and shifted to 18 °C for 1 hour by addition of an equal volume of 6 °C YEPD. The *spp382-1* strain is described in (Pandit et al., 2006). The *pxr1Δ* is described in (Banerjee et al., 2015). The strains deleted for the *SNR80* (YWD448a), *SNR83* (YWD451a) or *SNR87* (YWD452a) genes are described in (Schattner et al., 2004). Yeast strains deleted for the *SNR4* and *SNR45* genes are described in (Parker et al., 2018). The *DBR1*, *DBP3,* and *SQS1* deletion strains were obtained from Open BioSystems. All yeast strains and genotypes can be found in Suppl. Table S4.

### *RNA isolation*

RNA was extracted from approximately five total $OD_{600}$ of cells (usually 10 ml culture at $OD_{600}$ = 0.5 for mid log cells, 0.5 ml of stationary cells at $OD_{600}$ = 10) using a hot phenol protocol 1 described in (Ares, 2012).

### *In vitro synthesis of 18S and 25S rRNA*

Unmodified yeast 18S and 25S rRNAs were transcribed *in vitro* from plasmids encoding T7-18S and T7-25S sequences using T7 RNA polymerase. PCR products encoding 18S and 25S rDNA were amplified from the plasmid pWL155 which contains the *RDN1-1* gene fused with the *GAL* promoter at the 5´ end ((Liang and Fournier, 1997) a kind gift from Jelena Jakovlievic) and cloned into a T7 promoter-containing plasmid digested with EcoRI and HindIII using Gibson Assembly (NEB). The resulting plasmids were then digested with HindIII and run-off transcription was performed using the MEGAscript T7 kit (Invitrogen) following the manufacturer's instructions. T7-18S and -25S *in vitro* transcription reactions were evaluated by gel electrophoresis for bands of

correct size that correspond to 18S and 25S rRNAs. Transcription reactions were extracted and purified with phenol:chloroform:isoamyl alcohol (25:24:1), ethanol precipitated and resuspended in nuclease-free H2O. Purified T7-18S and -25S rRNA transcripts were then quantified on a NanoDrop spectrophotometer and pooled in equimolar ratios for sequencing library preparation. The T7 run-off transcription reactions terminate in a 3´ end generated by HindIII digestion and thus include an additional AAGCU sequence not present in endogenous 18S and 25S rRNAs. Therefore, T7-18S and T7-25S splint oligonucleotides were used to capture the 3´ end of T7 transcribed rRNAs (see below, Suppl. Table S5).

***Sequencing library preparation***

Direct RNA sequencing libraries were constructed using the SQK-RNA002 (Oxford Nanopore Technologies) kit following the manufacturer's protocol with the following modifications. Briefly, 750 ng of total yeast RNA was used as input material. To facilitate ligation of sequencing adapters to endogenous yeast 18S and 25S rRNA, 1 ul of 10 pmol/ul custom oligonucleotide duplexes complementary to the 3´ ends of 18S and 25S rRNA and the 5´ end of the ONT RMX sequencing adapter were used instead of the kit provided RTA adapter (Suppl. Table S5). To create duplexes, 100 pmol of either 18S or 25S splint oligo was incubated with 100 pmol of sequencing adapter and nuclease free H20 in a total volume of 10 ul. Reactions were heated to 95 °C for 2 minutes and gradually cooled at 65 °C for 10 minutes, 48 °C for 10 minutes, room temperature for 10 minutes and then placed on ice. Annealed oligonucleotide duplexes targeting 18S and 25S rRNAs were then pooled in equimolar ratio and 1 ul of the pool was used for sequencing library preparation. In the case of T7 rRNA sequencing libraries, T7-18S splint and T7-25S splint oligos were used to capture the 3´ end generated by HindIII digestion and run-off transcription. To enhance ligation efficiency during library preparation, the first and second ligation steps were increased from 10 minutes to 15 minutes and performed at room temperature. Reverse transcription was omitted. Sequencing-adapted libraries were eluted in 21 ul of elution buffer.

***Nanopore sequencing***

RNAs extracted from *GAL-NOP58* and *GAL-CBF5* strains, and *in vitro* transcribed RNA were sequenced on the MinION Mk1B sequencer using MinION FLO-MIN106D R9.4.1 flow cells (Oxford Nanopore Technologies) following the manufacturer's instructions. 20 ul of Sequencing libraries was mixed with 17.5 ul of $H_20$ and 37.5 ul of RRB buffer. 75 ul of the prepared sequencing library was loaded onto a flushed and primed flow cell and sequenced for 12-48 hour depending on the lifetime of active pores. RNAs extracted from all other strains and growth conditions were

sequenced on the MinION Mk1B sequencer using Flongle FLO-FLG001 R9.4.1 flow cells. Flongle flow cells were flushed and primed with 120 ul of flush buffer mix (117 ul FLB and 3 ul FLT). 30 ul of prepared sequencing library (described above) was loaded onto the flow cell and sequenced for 8-24 hours. Sequencing experiments were controlled using the MinKNOW software (Oxford Nanopore Technologies).

### *Data preprocessing*

Basecalling was done using the RNA model from Guppy v3.1.5+781ed57. To analyze specific subsets of reads more efficiently, we split the multi-fast5 reads into individual reads using the `multi_to_single_fast5` command from https://github.com/nanoporetech/ont_fast5_api. We then created an index file matching a fast5 to a fastq entry using `nanopolish index` from https://github.com/jts/nanopolish (Simpson et al., 2017). The reference sequence for the *S. cerevisiae* 18S and 25S rRNA came from (Engel et al., 2014). Initial basecalled sequence to reference alignment was done via minimap2 version 2.17-r943-dirty from https://github.com/lh3/minimap2 using the --MD flag which speeds up processing of signalAlign (Li, 2018). Alignment files were sorted and filtered using samtools version 1.9 by flag `-F 2308` which filters out unmapped reads, non-primary alignment reads and supplemental alignment reads (Li et al., 2009). Given that nanopore sequencing with RNA is 3´-5´, in order to filter for 'full length' reads we used `samtools view` to select for reads that covered the first 15 bases of either 18S or 25S rRNAs (Li et al., 2009). Read information and quality control metrics in Suppl. Table S6 were gathered using pycoQC version v2.5.0.23 (Leger and Leonardi, 2019).

### *SignalAlign Pipeline*

#### Model Definition

We initialized the transition probabilities from previous signalAlign r9.4 models. The initialized kmer distributions were defined in `r9.4_180mv_70bps_5mer_RNA` from ONT https://github.com/nanoporetech/kmer_models. Unlike previous kmer model modification detection algorithms, we chose to model modifications independently from other modifications of the same class in order to maintain the same informational inputs to each modification position. So, we iteratively redefined shared kmers with unused kmers from the model until all modifications were covered by unique kmers (see Code availability). For all kmers outside of modification branch points, we used the default RNA kmer distributions from ONT (r9.4_180mv_70bps_5mer_RNA).

## Training Configuration

SignalAlign v1.0.0 uses a variable-order hidden Markov model (HMM) which allows the number of paths through the HMM to be correctly constructed when ambiguous positions are defined (see Code availability) (Rand et al., 2017). Recent updates to signalAlign allow for relatively easy model definition and variant site selection which allows a user to define modifications, set prediction site locations and train a model. We defined all modified positions in the IVT sample as canonical and all positions in the wild type as modified. The locations of modified positions were determined by (Taoka et al., 2016) via mass spectrometry. For supervised training, all potential modified positions were defined as either canonical (IVT) or modified (wild type). Modifications within the first 15 nucleotides from the 5´ end of 18S or 25S cannot be called because the signal from that part of the sequence is not captured by standard nanopore sequencing. This limitation does not affect our experiments because there are no modifications in yeast which are within the first 15 nucleotides from the 5´ end of 18S or 25S.

We used 500 reads from both wild type and IVT reads and ran 30 rounds of training. For each round of training, we generated alignments between events and the reference sequence. Then, we generated new event Gaussian distributions for all kmers covering modified positions. The mean of the gaussian distribution was defined as the median of the empirical kmer distribution and the standard deviation was defined as the median absolute deviation of the empirical kmer distribution. Similar to another study, we have seen that the median is less susceptible to being influenced by outliers(Ding et al., 2020). To train the model, we used `trainModels.py` from signalAlign (see Code availability).

## Inference and Accuracy Metrics

In order to test our results, we used ‘runSignalAlign.py’ and a trained model to predict modification status on all positions of 500 hold out IVT reads and 500 hold out wild type reads (see Code availability). SignalAlign produced posterior probabilities of event alignments to both canonical kmers and modified kmers. We used `embed_main sa2bed` to decode the posterior probabilities from the signalAlign output into the probability of a position being modified (Rand et al., 2017). These probabilities are used for the receiver operating characteristic curve, precision-recall curve, calibration curve of Suppl. Fig. S2. A probability cutoff of 0.5 is used for the confusion matrix as well as the quantification of percent modified for any position.

We also compared accuracy on our test set to several snoRNA knockouts. Again, assuming snoRNA knockouts completely ablate target modifications and modifications are 100% present at all other positions, the average balanced accuracy over the

snoRNA knockout positions is 85.1% and the expected balanced accuracy is 87.1% (Suppl. Table S2-3). Average balanced accuracy is calculated by getting the average of all balanced accuracies across all snoRNA knockout positions. Balanced accuracy for one position is calculated by adding the specificity to the sensitivity and dividing by two.

### Percent Modification Change

For every experiment and each modification position, we perform several chi-square two sample tests comparing the three wild type replicates modification frequencies to all experiment repeats' modification frequency(Pearson, 1900). We then select the highest p-value and correct for multiple tests using the Benjamani-Hochberg procedure (Benjamini and Hochberg 1995). We also control for batch effects by filtering out positions which fall below the maximum change in modification frequency between the replicates of the wild type. All sample modification percentages, p-values from chi-square two sample tests compared to wild type replicates and IVT, and Benjamani-Hochberg corrected p-values can be found in Suppl. Table S2.

## *Hierarchical Clustering Analysis*

### Dendrogram creation procedure

In order to determine any subclusters of reads based on a modification profile, we used hierarchical clustering on the modification profiles we generated from the inference step (Pedregosa et al., 2011; Waskom et al., 2020). We generated the dendrogram using Ward's method and euclidean distance as the distance metric(Ward, 1963). Before clustering analysis, we filter out reads which do not cover every modification site. UMAP dimension reduction was done using the umap python package and visualization using matplotlib (Hunter, 2007; McInnes et al., 2018).

### Cluster Partitioning

To determine the number of reads in a set of N clusters we simply cut the dendrogram to create N subclusters and calculated the fraction of reads within each branch.

## *Modification Correlations*

To calculate correlations between modified positions, we first filter out reads which did not cover all modifications and select the set of probabilities associated with

each position. We then calculate the Spearman rank correlation between all pairwise combinations of modification positions on the same molecule. We use Spearman rank correlation to study the relationship between two probabilities because pairs of probabilities do not follow a bivariate normal distribution. P-values were calculated using a two sided t-test and multiple tests corrected via the Benjamani-Hochberg procedure (Benjamini and Hochberg, 1995; Student, 1908).

To compare correlations between experiments, we used Fisher's z-transformation to convert correlations into z-scores and then performed a z-test to obtain p-values (Fieller et al., 1957; Fieller and Pearson, 1961; Fisher, 1915; Zar, 2014). We then correct for multiple tests using the Benjamani-Hochberg procedure (Benjamini and Hochberg, 1995). These p-values represent the confidence that, between two samples, there is a significant difference between the two correlations. All correlation plots have stars for positions which are both significantly different from a comparison experiment (wild type or IVT) and are significantly different from zero (p-value < 0.05). To account for variation in experimental repeats, we plot the minimum difference and highest corrected p-value for all pairwise comparisons between experimental repeats and wild type repeats.

For higher order claims which require aggregating information from several hypothesis tests we use Empirical Brown's method (Brown, 1975; Poole et al., 2016). The Empirical Brown's method uses empirical data to calculate the covariance matrix which is used to extended Fisher's method to the dependent case by using a re-scaled $\chi^2$ distribution (see Code availability).

Spearman correlation values, original two sided t-test p-values, corrected two sided t-test p-values, Fisher z-transform test comparison p-values, and corrected Fisher z-transform tests p-values can be found in Suppl. Table S1.

### *Event Visualization*

Using a similar procedure outlined in a previous study (Ding et al., 2020), we gather the kmer to reference mapping generated from signalAlign and extract the most probable event to kmer alignment path using the maximum expected accuracy alignment (Durbin et al., 1998; Rand et al., 2017). For each read, we standardize the raw signal and calculate event means. Prior to clustering and visualization, we combine all reads together and standardize events by column. We generate the dendrogram using the same procedure as hierarchical clustering of modification profiles, Ward's method and euclidean distance (Ward, 1963).

For kmer distributions seen in Suppl. Figure S5, we plot the kernel density estimate of all events aligning to the corresponding kmer with a probability greater than 0.5. We then simply plot the corresponding kmer distributions from the final trained kmer model.

### *Sample Compare Site Detection*

#### Tombo Pipeline

Using Tombo version 1.5.1, initial embedding of fastq data into the raw fast5s was done with the `tombo preprocess annotate_raw_with_fastqs` and signal to reference alignment with `tombo resquiggle` (Stoiber et al., 2017). Finally, `tombo detect_modifications level_sample_compare` was used to generate windowed means of individual position Kolmogorov–Smirnov tests comparing the IVT sample position signal distributions to the wild type sample (WT_YPD) position signal distributions (Stoiber et al., 2017). For a given position i, the windowed mean D-statistic is $w_i = \frac{\sum_{i-1}^{i+1} d_i}{3}$ where d is the D-statistic for a given position and w is the final reported statistic plotted in Suppl. Fig. S1A and D.

#### Accuracy of Modification Site Prediction

Using the D-statistic generated from Tombo, we calculated the per-position modification location detection AUROC (Area Under the Receiver Operating Characteristic) for yeast 18S (0.924) and 25S (0.934) rRNA (Suppl. Fig. S1B and E, see Supplemental Methods).

We also wanted to see if peaks corresponded with modifications, not just raw D-statistic values. So, we identified peaks and considered a peak within 2 nucleotides of a modification as a true positive. We then calculated the AUROC for this less stringent method on yeast 18S (0.984) and 25S (0.986) rRNA (Suppl. Fig. S1C and F, see Supplemental Methods).

### *Modification Labels and Frequency*

Underlying labels for modification and frequency for the *S. cerevisiae* 18S and 25S rRNA came from (Taoka et al., 2016). Expected changes in modification frequency in the Dbp3 deletion experiment came from (Aquino et al., 2021). SnoRNA modification sites on yeast rRNA come from the UMASS Amherst Yeast snoRNA database (Piekna-Przybylska et al., 2007).

### *Data availability*

Fastq files from all direct RNA sequencing runs and signalAlign modification calls are publicly available in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE186634 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186634). Fast5 files for all direct RNA sequencing are available in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB48183 (https://www.ebi.ac.uk/ena/browser/view/PRJEB48183). A detailed description of the datasets used and sequenced in this work with their corresponding ENA, GEO, and SRA IDs can be found in Supplementary Table S7.

### *Code availability*

Documentation, install requirements, and analysis scripts for all work specific to this paper can be found at https://github.com/adbailey4/yeast_rrna_modification_detection. SignalAlign v1.0.0 can be found at https://github.com/UCSC-nanopore-cgl/signalAlign and embed_fast5 1.0.0 can be found https://github.com/adbailey4/embed_fast5.

## ACKNOWLEDGEMENTS

We would like to thank our colleagues for graciously sharing strains and plasmids: Skip Fournier and Wayne Decatur (snR80, snR83, and snR87 deletions), David Tollervey (*GAL-CBF5* and *GAL-NOP58*), Raymond O'Keefe (snR4 and snR45 deletions), Jon Staley (*prp43-Q423N*), Brian Rymond (*spp382-1, pxr1Δ*), Hiten Madhani (*prp16-302*), Jelena Micic and John Woolford (rDNA plasmid originally provided by Skip Fournier). We also thank Nicholas Forino and the Michael Stone lab for initial use of the flongle adaptor. We also thank Jordan Eizenga, John Paul Donohue, Miten Jain, and Logan Mulroney for technical assistance and advice, and to colleagues who provided scientific discussion and feedback on earlier versions of this manuscript: Gina Mawla, Guillaume Chanfreau, Jelena Micic, Idil Ulengin-Talkish, Jen Quick-Cleveland.

This research was supported by NIH grants R01 HG010053 (to M. Akeson, Ares and Paten Co-PIs) and R01 GM040478 (M. Ares), as well as U41HG010972, R01HG010485, U01HG010961, OT3HL142481, OT2OD026682, U01HL137183, and 2U41HG007234 to B. Paten.

## AUTHOR CONTRIBUTIONS

All authors contributed to conception and experimental design. JT, MA and HI grew yeast, extracted RNA, prepared libraries, ran the nanopore sequencer and performed base calling. AB performed most of the nanopore data analysis, including model building, clustering, and correlation measurements. HD provided feedback and advice throughout the development and analysis process. AD and SM built some of the clustering analysis scripts. AD and SM's participation in this research took place under the auspices of the Science Internship Program at the University of California Santa Cruz. MA and BP supervised the project. All authors assisted in manuscript preparation.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

## REFERENCES


Aquino GRR, Krogh N, Hackert P, Martin R, Gallesio JD, van Nues RW, Schneider C, Watkins NJ, Nielsen H, Bohnsack KE, Bohnsack MT. 2021. RNA helicase-mediated regulation of snoRNP dynamics on pre-ribosomes and rRNA 2′-O-methylation. *Nucleic Acids Res* **49**:4066–4084.

Ares M. 2012. Isolation of total RNA from yeast cell cultures. *Cold Spring Harb Protoc* **2012**:1082–1086.

Banerjee D, McDaniel PM, Rymond BC. 2015. Limited portability of G-patch domains in regulators of the Prp43 RNA helicase required for pre-mRNA splicing and ribosomal RNA maturation in Saccharomyces cerevisiae. *Genetics* **200**:135–147.

Bartosovic M, Molares HC, Gregorova P, Hrossova D, Kudla G, Vanacova S. 2017. N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3'-end processing. *Nucleic Acids Res* **45**:11356–11370.

Bednářová A, Hanna M, Durham I, VanCleave T, England A, Chaudhuri A, Krishnan N. 2017. Lost in Translation: Defects in Transfer RNA Modifications and Neurological Disorders. *Front Mol Neurosci* **10**:1–8.

Begik O, Lucas MC, Pryszcz LP, Ramirez JM, Medina R, Milenkovic I, Cruciani S, Liu H, Vieira HGS, Sas-Chen A, Mattick JS, Schwartz S, Novoa EM. 2021. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol*. doi:10.1038/s41587-021-00915-6

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol* **57**:289–300.

Benrahma H, Charoute H, Lasram K, Boulouiz R, Atig RK-B, Fakiri M, Rouba H, Abdelhak S, Barakat A. 2014. Association Analysis of IGF2BP2, KCNJ11, and CDKAL1 Polymorphisms with Type 2 Diabetes Mellitus in a Moroccan Population: A Case–Control Study and Meta-analysis. *Biochem Genet* **52**:430–442.

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. 2011. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**:1524–1529.

Bernier CR, Petrov AS, Waterbury CC, Jett J, Li F, Freil LE, Xiong X, Wang L, Migliozzi BLR, Hershkovits E, Xue Y, Hsiao C, Bowman JC, Harvey SC, Grover MA, Wartell ZJ, Williams LD. 2014. RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss* **169**:195–207.

Birkedal U, Christensen-Dalsgaard M, Krogh N, Sabarinathan R, Gorodkin J, Nielsen H. 2014. Profiling of Ribose Methylations in RNA by High-Throughput Sequencing. *Angew Chem Int Ed* **54**. doi:10.1002/anie.201408362

Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Bagiński B, Wirecki TK, de Crécy-Lagard V, Ross R, Limbach PA, Kotter A, Helm M, Bujnicki JM. 2018. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**:D303–D307.

Bohnsack MT, Martin R, Granneman S, Ruprecht M, Schleiff E, Tollervey D. 2009. Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol Cell* **36**:583–592.

Brand RC, Klootwijk J, Planta RJ, Maden BE. 1978. Biosynthesis of a hypermodified nucleotide in Saccharomyces carlsbergensis 17S and HeLa-cell 18S ribosomal ribonucleic acid. *Biochem J* **169**:71–77.

Brown MB. 1975. 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics* **31**:987–992.

Cardenas ME, Cutler NS, Lorenz MC, Di Como CJ, Heitman J. 1999. The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Dev* **13**:3271–3279.

Chapman KB, Boeke JD. 1991. Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* **65**:483–492.

Chen Y-L, Capeyrou R, Humbert O, Mouffok S, Kadri YA, Lebaron S, Henras AK, Henry Y. 2014. The telomerase inhibitor Gno1p/PINX1 activates the helicase Prp43p during ribosome biogenesis. *Nucleic Acids Res* **42**:7330–7345.

Choi J, Grosely R, Prabhakar A, Lapointe CP, Wang J, Puglisi JD. 2018. How Messenger RNA and Nascent Chain Sequences Regulate Translation Elongation. *Annu Rev Biochem* **87**:421–449.

Christian H, Hofele RV, Urlaub H, Ficner R. 2014. Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. *Nucleic Acids Res* **42**:1162–1179.

Combs DJ, Nagel RJ, Ares M Jr, Stevens SW. 2006. Prp43p is a DEAH-box spliceosome disassembly factor essential for ribosome biogenesis. *Mol Cell Biol* **26**:523–534.

Deamer D, Akeson M, Branton D. 2016. Three decades of nanopore sequencing. *Nat Biotechnol* **34**:518–524.

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

Delaunay S, Rapino F, Tharun L, Zhou Z, Heukamp L, Termathe M, Shostak K, Klevernic I, Florin A, Desmecht H, Desmet CJ, Nguyen L, Leidel SA, Willis AE, Büttner R, Chariot A, Close P. 2016. Elp3 links tRNA modification to IRES-dependent translation of LEF1 to sustain metastasis in breast cancer. *J Exp Med* **213**:2503–2523.

Ding H, Anastopoulos I, Bailey AD, Stuart J, Paten B. 2021. Towards inferring nanopore sequencing ionic currents from nucleotide chemical structures. *Nat Commun* **12**:1–9.

Ding H, Bailey AD, Jain M, Olsen H, Paten B. 2020. Gaussian mixture model-based unsupervised nucleotide modification number detection using nanopore-sequencing readouts. *Bioinformatics* 1–7.

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

Drexler HL, Choquet K, Churchman LS. 2020. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol Cell* **77**:985–998.e8.

Duan J, Li L, Lu J, Wang W, Ye K. 2009. Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell* **34**:427–439.

Durbin R, Eddy SR, Krogh A, Mitchison G. 1998. Biological Sequence Analysis: Probabilistic

Models of Proteins and Nucleic Acids. Cambridge University Press.
Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, Weng S, Wong ED, Lloyd P, Skrzypek MS, Miyasato SR, Simison M, Cherry JM. 2014. The reference genome sequence of Saccharomyces cerevisiae: then and now. *G3* **4**:389–398.
Erales J, Marchand V, Panthu B, Gillot S, Belin S, Ghayad SE, Garcia M, Laforêts F, Marcel V, Baudin-Baillieu A, Bertin P, Couté Y, Adrait A, Meyer M, Therizols G, Yusupov M, Namy O, Ohlmann T, Motorin Y, Catez F, Diaz J-J. 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.
Fieller EC, Hartley HO, Pearson ES. 1957. Tests for Rank Correlation Coefficients. I. *Biometrika* **44**:470–481.
Fieller EC, Pearson ES. 1961. Tests for Rank Correlation Coefficients: II. *Biometrika* **48**:29–40.
Fisher RA. 1915. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika* **10**:507–521.
Fourmann J-B, Dybkov O, Agafonov DE, Tauchert MJ, Urlaub H, Ficner R, Fabrizio P, Lührmann R. 2016. The target of the DEAH-box NTP triphosphatase Prp43 in Saccharomyces cerevisiae spliceosomes is the U2 snRNP-intron interaction. *Elife* **5**. doi:10.7554/eLife.15564
Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. 1992. A genomic sequencing protocol that yields a positive display of 5- methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **89**:1827–1831.
Furlan M, Delgado-Tejedor A, Mulroney L, Pelizzola M, Novoa EM, Leonardi T. 2021. Computational methods for RNA modification detection from nanopore direct RNA sequencing data. *RNA Biol* 1–10.
Garalde DR, Snell EA, Jachimowicz D, Heron AJ, Bruce M, Lloyd J, Warland A, Pantic N, Admassu T, Ciccone J, Serra S, Keenan J, Martin S, McNeill L, Wallace J, Jayasinghe L, Wright C, Blasco J, Sipos B, Young S, Juul S, Clarke J, Turner DJ. 2016. Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv* 068809.
Gasch AP. 2000. Genomic Expression Programs in the Yeast Saccharomyces Cerevisiae Responding to Diverse Environmental Changes. Stanford University.
Gay DM, Lund AH, Jansson MD. 2021. Translational control through ribosome heterogeneity and functional specialization. *Trends Biochem Sci*. doi:10.1016/j.tibs.2021.07.001
Grzechnik P, Szczepaniak SA, Dhir S, Pastucha A, Parslow H, Matuszek Z, Mischo HE, Kufel J, Proudfoot NJ. 2018. Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage. *Nat Commun* **9**:1783.
Guglielmi B, Werner M. 2002. The yeast homolog of human PinX1 is involved in rRNA and small nucleolar RNA maturation, not in telomere elongation inhibition. *J Biol Chem* **277**:35712–35719.
Hardwick JS, Kuruvilla FG, Tong JK, Shamji AF, Schreiber SL. 1999. Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proc Natl Acad Sci U S A* **96**:14866–14870.
Heininger AU, Hackert P, Andreou AZ, Boon K-L, Memet I, Prior M, Clancy A, Schmidt B, Urlaub H, Schleiff E, Sloan KE, Deckers M, Lührmann R, Enderlein J, Klostermeier D, Rehling P, Bohnsack MT. 2016. Protein cofactor competition regulates the action of a multifunctional RNA helicase in different pathways. *RNA Biol* **13**:320–330.
Heitman J, Movva NR, Hall MN. 1991. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science* **253**:905–909.
He Y, Staley JP, Andersen GR, Nielsen KH. 2017. Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. *RNA* **23**:1110–1124.

Hideyama T, Yamashita T, Aizawa H, Tsuji S, Kakita A, Takahashi H, Kwak S. 2012. Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiol Dis* **45**:1121–1128.
Honma Y, Kitamura A, Shioda R, Maruyama H, Ozaki K, Oda Y, Mini T, Jenö P, Maki Y, Yonezawa K, Hurt E, Ueno M, Uritani M, Hall MN, Ushimaru T. 2006. TOR regulates late steps of ribosome maturation in the nucleoplasm via Nog1 in response to nutrients. *EMBO J* **25**:3832–3842.
Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* **9**:90–95.
Jain M, Olsen HE, Akeson M, Abu-Shumays R. 2021. Adaptation of Human Ribosomal RNA for Nanopore Sequencing of Canonical and Modified Nucleotides. *Methods Mol Biol* **2298**:53–74.
Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. 2017. The RNA modification landscape in human disease. *RNA* **23**:1754–1769.
Ju Q, Warner JR. 1994. Ribosome synthesis during the growth cycle of Saccharomyces cerevisiae. *Yeast* **10**:151–157.
Karijolich J, Kantartzis A, Yu Y-T. 2010. RNA Modifications: A Mechanism that Modulates Gene ExpressionMethods in Molecular Biology (Clifton, N.J.). Springer US. pp. 1–19.
Klinge S, Woolford JL Jr. 2019. Ribosome assembly coming into focus. *Nat Rev Mol Cell Biol* **20**:116–131.
Kos M, Tollervey D. 2010. Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell* **37**:809–820.
Krogh N, Jansson MD, Häfner SJ, Tehler D, Birkedal U, Christensen-Dalsgaard M, Lund AH, 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 Res* **44**:7884–7895.
Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D. 1998. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev* **12**:527–537.
Lafontaine DL, Tollervey D. 1999. Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA* **5**:455–467.
Lapeyre B, Purushothaman SK. 2004. Spb1p-directed formation of Gm2922 in the ribosome catalytic center occurs at a late processing stage. *Mol Cell* **16**:663–669.
Lebaron S, Froment C, Fromont-Racine M, Rain J-C, Monsarrat B, Caizergues-Ferrer M, Henry Y. 2005. The splicing ATPase prp43p is a component of multiple preribosomal particles. *Mol Cell Biol* **25**:9269–9282.
Leeds NB, Small EC, Hiley SL, Hughes TR, Staley JP. 2006. The splicing factor Prp43p, a DEAH box ATPase, functions in ribosome biogenesis. *Mol Cell Biol* **26**:513–522.
Leger A, Amaral PP, Pandolfini L, Capitanchik C, Capraro F, Barbieri I, Migliori V, Luscombe NM, Enright AJ, Tzelepis K, Ule J, Fitzgerald T, Birney E, Leonardi T, Kouzarides T. 2019. RNA modifications detection by comparative Nanopore direct RNA sequencing. *bioRxiv* 843136.
Leger A, Leonardi T. 2019. pycoQC, interactive quality control for Oxford Nanopore Sequencing. *JOSS* **4**:1236.
Liang WQ, Fournier MJ. 1997. Synthesis of functional eukaryotic ribosomal RNAs in trans: development of a novel in vivo rDNA system for dissecting ribosome biogenesis. *Proc Natl Acad Sci U S A* **94**:2864–2868.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079.

Lin J, Lai S, Jia R, Xu A, Zhang L, Lu J, Ye K. 2011. Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **469**:559–563.
Lin S, Choe J, Du P, Triboulet R, Gregory RI. 2016. The m(6)A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol Cell* **62**:335–345.
Liu H, Begik O, Novoa EM. 2021. EpiNano: Detection of m6A RNA Modifications Using Oxford Nanopore Direct RNA Sequencing. *Methods Mol Biol* **2298**:31–52.
Liu K, Santos DA, Hussmann JA, Wang Y, Sutter BM, Weissman JS, Tu BP. 2021. Regulation of translation by methylation multiplicity of 18S rRNA. *Cell Rep* **34**:108825.
Madhani HD, Guthrie C. 1994. Genetic interactions between the yeast RNA helicase homolog Prp16 and spliceosomal snRNAs identify candidate ligands for the Prp16 RNA-dependent ATPase. *Genetics* **137**:677–687.
Marchand V, Blanloeil-Oillo F, Helm M, Motorin Y. 2016. Illumina-based RiboMethSeq approach for mapping of 2′-O-Me residues in RNA. *Nucleic Acids Res* **44**:e135–e135.
Martin A, Schneider S, Schwer B. 2002. Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J Biol Chem* **277**:17743–17750.
Martin R, Straub AU, Doebele C, Bohnsack MT. 2013. DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol* **10**:4–18.
McInnes L, Healy J, Melville J. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [statML]*.
Nakatogawa H, Ito K. 2002. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**:629–636.
Natchiar SK, Myasnikov AG, Kratzat H, Hazemann I, Klaholz BP. 2017. Visualization of chemical modifications in the human 80S ribosome structure. *Nature* **551**:472–477.
Obrig TG, Culp WJ, McKeehan WL, Hardesty B. 1971. The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes. *J Biol Chem* **246**:174–181.
Ooi SL, Samarsky DA, Fournier MJ, Boeke JD. 1998. Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA* **4**:1096–1110.
Pandit S, Lynn B, Rymond BC. 2006. Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc Natl Acad Sci U S A* **103**:13700–13705.
Parker S, Fraczek MG, Wu J, Shamsah S, Manousaki A, Dungrattanalert K, de Almeida RA, Invernizzi E, Burgis T, Omara W, Griffiths-Jones S, Delneri D, O'Keefe RT. 2018. Large-scale profiling of noncoding RNA function in yeast. *PLoS Genet* **14**:e1007253.
Pearson K. 1900. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **50**:157–175.
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Others. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**:2825–2830.
Pertschy B, Schneider C, Gnädig M, Schäfer T, Tollervey D, Hurt E. 2009. RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1. *J Biol Chem* **284**:35079–35091.
Petfalski E, Dandekar T, Henry Y, Tollervey D. 1998. Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol Cell Biol* **18**:1181–1189.
Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovits E, Hsiao C, Harvey SC, Hud NV, Fox GE, Wartell RM, Williams LD. 2014. Secondary structures of rRNAs from all three domains of life. *PLoS One* **9**:e88222.
Piekna-Przybylska D, Decatur WA, Fournier MJ. 2007. New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA* **13**:305–312.

Polikanov YS, Melnikov SV, Söll D, Steitz TA. 2015. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat Struct Mol Biol* **22**:342–344.
Poole W, Gibbs DL, Shmulevich I, Bernard B, Knijnenburg TA. 2016. Combining dependent P-values with an empirical adaptation of Brown's method. *Bioinformatics* **32**:i430–i436.
Portela A, Esteller M. 2010. Epigenetic modifications and human disease. *Nat Biotechnol* **28**:1057–1068.
Powers T, Walter P. 1999. Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in Saccharomyces cerevisiae. *Mol Biol Cell* **10**:987–1000.
Raiber E-A, Hardisty R, van Delft P, Balasubramanian S. 2017. Mapping and elucidating the function of modified bases in DNA. *Nature Reviews Chemistry* **1**. doi:10.1038/s41570-017-0069
Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB. 2013. Identifying RNA editing sites using RNA sequencing data alone. *Nat Methods* **10**:128–132.
Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B. 2017. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods*. doi:10.1038/nmeth.4189
Schattner P, Decatur WA, Davis CA, Ares M Jr, Fournier MJ, Lowe TM. 2004. Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome. *Nucleic Acids Res* **32**:4281–4296.
Schimmel P. 2018. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol* **19**:45–58.
Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, Fink G, Regev A. 2014. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**:148–162.
Sharma S, Yang J, van Nues R, Watzinger P, Kötter P, Lafontaine DLJ, Granneman S, Entian K-D. 2017. Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genet* **13**:e1006804.
Simms CL, Yan LL, Zaher HS. 2017. Ribosome Collision Is Critical for Quality Control during No-Go Decay. *Mol Cell* **68**:361–373.e5.
Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting dna cytosine methylation using nanopore sequencing. *Nat Methods*. doi:10.1038/nmeth.4184
Sloan KE, Warda AS, Sharma S, Entian K-D, Lafontaine DLJ, Bohnsack MT. 2017. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**:1138–1152.
Sood AJ, Viner C, Hoffman MM. 2019. DNAmod: the DNA modification database. *J Cheminform* **11**:30.
Staley JP, Woolford JL. 2009. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr Opin Cell Biol* **21**:109–118.
Stoiber M, Quick J, Egan R, Lee JE, Celniker S, Neely RK, Loman N, Pennacchio LA, Brown J. 2017. De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *bioRxiv*. doi:10.1101/094672
Student. 1908. The Probable Error of a Mean. *Biometrika* **6**:1–25.
Talkish J, Biedka S, Jakovljevic J, Zhang J, Tang L, Strahler JR, Andrews PC, Maddock JR, Woolford JL Jr. 2016. Disruption of ribosome assembly in yeast blocks cotranscriptional pre-rRNA processing and affects the global hierarchy of ribosome biogenesis. *RNA* **22**:852–866.
Talkish J, Zhang J, Jakovljevic J, Horsey EW, Woolford JL Jr. 2012. Hierarchical recruitment into nascent ribosomes of assembly factors required for 27SB pre-rRNA processing in Saccharomyces cerevisiae. *Nucleic Acids Res* **40**:8646–8661.
Tanaka N, Aronova A, Schwer B. 2007. Ntr1 activates the Prp43 helicase to trigger release of

lariat-intron from the spliceosome. *Genes Dev* **21**:2312–2325.
Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, Brooks AN. 2020. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun* **11**:1438.
Taoka M, Nobe Y, Yamaki Y, Yamauchi Y, Ishikawa H, Takahashi N, Nakayama H, 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, Yamauchi Y, Nobe Y, Masaki S, Nakayama H, Ishikawa H, Takahashi N, Isobe T. 2009. An analytical platform for mass spectrometry-based identification and chemical analysis of RNA in ribonucleoprotein complexes. *Nucleic Acids Res* **37**:e140–e140.
Thomas NK, Poodari VC, Jain M, Olsen HE, Akeson M, Abu-Shumays RL. 2021. Direct Nanopore Sequencing of Individual Full Length tRNA Strands. *ACS Nano* **15**:16642–16653.
Tollervey D, Kiss T. 1997. Function and synthesis of small nucleolar RNAs. *Curr Opin Cell Biol* **9**:337–342.
Toroney R, Nielsen KH, Staley JP. 2019. Termination of pre-mRNA splicing requires that the ATPase and RNA unwindase Prp43p acts on the catalytic snRNA U6. *Genes Dev* **33**:1555–1574.
Tsai R-T, Fu R-H, Yeh F-L, Tseng C-K, Lin Y-C, Huang Y-H, Cheng S-C. 2005. Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev* **19**:2991–3003.
Tseng C-K, Liu H-L, Cheng S-C. 2011. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**:145–154.
van Nues RW, Granneman S, Kudla G, Sloan KE, Chicken M, Tollervey D, Watkins NJ. 2011. Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. *EMBO J* **30**:2420–2430.
Vanrobays E, Leplus A, Osheim YN, Beyer AL, Wacheul L, Lafontaine DLJ. 2008. TOR regulates the subcellular distribution of DIM2, a KH domain protein required for cotranscriptional ribosome assembly and pre-40S ribosome export. *RNA* **14**:2061–2073.
Vincenti S, De Chiara V, Bozzoni I, Presutti C. 2007. The position of yeast snoRNA-coding regions within host introns is essential for their biosynthesis and for efficient splicing of the host pre-mRNA. *RNA* **13**:138–150.
Ward JH. 1963. Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* **58**:236–244.
Warner JR. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**:437–440.
Waskom M, Gelbart M, Botvinnik O, Ostblom J, Hobson P, Lukauskas S, Gemperline DC, Augspurger T, Halchenko Y, Warmenhoven J, Cole JB, de Ruiter J, Vanderplas J, Hoyer S, Pye C, Miles A, Swain C, Meyer K, Martin M, Bachant P, Quintero E, Kunter G, Villalba S, Brian, Fitzgerald C, Evans C, Williams ML, O'Kane D, Yarkoni T, Brunner T. 2020. mwaskom/seaborn: v0.11.1 (December 2020). doi:10.5281/zenodo.4379347
Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Lührmann R. 1998. Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* **4**:1549–1568.
Weaver PL, Sun C, Chang TH. 1997. Dbp3p, a putative RNA helicase in Saccharomyces cerevisiae, is required for efficient pre-rRNA processing predominantly at site A3. *Mol Cell Biol* **17**:1354–1365.
Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20**:129.
Wilson DM, Li Y, LaPeruta A, Gamalinda M, Gao N, Woolford JL Jr. 2020. Structural insights

into assembly of the ribosomal nascent polypeptide exit tunnel. *Nat Commun* **11**:5111.
Wu Q, Niebuhr E, Yang H, Hansen L. 2005. Determination of the “critical region” for cat-like cry of Cri-du-chat syndrome and analysis of candidate genes by quantitative PCR. *Eur J Hum Genet* **13**:475–485.
Yang J, Sharma S, Watzinger P, Hartmann JD, Kötter P, Entian K-D. 2016. Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS One* **11**:e0168873.
Zar JH. 2014. Spearman rank correlation: Overview. *Wiley StatsRef: Statistics Reference Online*. doi:10.1002/9781118445112.stat05964
Zhang Y, Wölfle T, Rospert S. 2013. Interaction of nascent chains with the ribosomal tunnel proteins Rpl4, Rpl17, and Rpl39 of Saccharomyces cerevisiae. *J Biol Chem* **288**:33697–33707.
Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang C-M, Li CJ, Vågbø CB, Shi Y, Wang W-L, Song S-H, Lu Z, Bosmans RPG, Dai Q, Hao Y-J, Yang X, Zhao W-M, Tong W-M, Wang X-J, Bogdan F, Furu K, Fu Y, Jia G, Zhao X, Liu J, Krokan HE, Klungland A, Yang Y-G, He C. 2013. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**:18–29.