SUPPLEMENTAL METHODS

5'-end sequencing in *Saccharomyces cerevisiae* offers new insights into 5'-ends of tRNAHis and snoRNAs.

Samantha Dodbele, Blythe Moreland, Spencer M. Gardner, Ralf Bundschuh, Jane E. Jackman

Plasmid shuffle assay

Plasmids for galactose-inducible expression of BtTLP [*CEN LEU2* P_{GAL}*−BtTLP*] along with analogous yeast *THG1*-containing and empty vector control were transformed into yeast strain JJY20 (relevant genotype *MATα thg1Δ::kanMX his3-1 leu2Δ met15Δ ura3Δ* [*CEN URA3* P_{THG1}*–THG1*])[1,2]. Cells were grown at 30 °C on minimal synthetic dextrose (SD) drop out media plates (SD-ura-leu) to select for the presence of both plasmids. Positive transformants were grown at 30 °C on synthetic galactose (SGal) media containing 5-flouroortic acid (5FOA), and uracil (SGal-leu+5FOA media). 5FOA acts as a toxin to cells containing the *URA3* gene, therefore only cells that lost the [*CEN URA3* P_{THG1}*–THG1*] covering plasmid can be recovered.

Library preparation and sequencing

Size selected RNAs (MiRvana) were eluted in ddH_2O and quantified by NanoDrop. To remove any 5'-triphosphorylated ends, size selected RNAs were incubated with Tobacco Acid Pyrophosphatase (NEB). Library preparation for high throughput sequencing was carried out using the NEBNext Small RNA Library Prep Set for Illumina (NEB) according to the manufacturer's instructions. RNAs were analyzed for purity at each step with the Agilent 2100 Bioanalyzer (Agilent) using an Agilent RNA Nano 6000 Kit and Agilent High Sensitivity DNA Kit. The resulting libraries were sequenced on an Illumina HiSeq 2500 instrument through the OSUCCC James Genomics Shared Resource.

RNA-Seq quality filtering and alignment

FASTQ files of the paired-end reads were obtained following sequencing and underwent adapter trimming and quality filtering by the program Skewer v0.2.2 [3]. Minimum adapter overlap was 5 bp. Paired reads were accepted if both ends had an error ratio ≤ 0.2, an indel error ratio ≤ 0.030, and if the mean base quality score was ≥ 20. The minimum read length allowed after trimming was 25 nt. Of the 9,814,324 paired-end reads from the Thg1 WT sample and 12,372,531 paired-end reads from the BtTLP-complemented sample, this left 9,054,435 and 11,228,570 paired-end reads, respectively, for further processing.

After quality filtering, reads were aligned to the *S. cerevisiae* reference genome (sacCer3) downloaded from the UCSC Genome Browser with the fast RNA alignment software STAR (Spliced Transcripts Alignment to a Reference) version 2.4.2a [4]. Only alignments with the same score as the highest score attained for that paired-end read were accepted. In the end, 99.5% and 99.3% of reads, from the ScThg1 and BtTLP-complemented samples, respectively, were aligned to the reference genome with 1.99% and 2.65% unique alignments.

Importantly, STAR considers alignments to the genome with soft-clipped bases at the 5' and 3' ends. An alignment that includes soft-clipped bases signifies that the bases at the 5' or 3' ends do not match the reference genome at that location. Thus, if an enzyme had added a nucleotide post-transcription that did not match the genetically encoded base, STAR could still locate the site of the original transcription but with the first bases considered "soft-clipped".

BEDgraph files of 5'-end accumulation

A read's alignment was considered further if the number of soft-clipped bases at the 5'-end was less than or equal to 10 nt. The 5'-end of a read was determined with respect to the strand that the read aligned to, and the position by the full projection of the read (including soft-clipped bases) onto the forward strand. Information from all considered alignments was accumulated into BEDgraph file entries that count the number of read alignments with 5'-end starting at that genomic location on the given strand. In addition, the base corresponding to that 5'-end was counted and accumulated into the descriptive fields (**Figure 2B, main text**). Alignments of 3'-ends were similarly filtered – the maximum number of clipped bases now applying to the 3'-end of the paired-read – and accumulated into BEDgraph files.

Detecting and classifying peaks in alignments around snoRNA annotations

To explore deviations from annotated snoRNA starts, we looked through every snoRNA in the sgdOther database for peaks in 5'-end alignments (as filtered in the BEDgraph files) within 10 nt of the annotated 5'-end. For genomic locations in this [-10,10] interval, the location, x, with the highest number of 5'-end alignments, N_x, was determined to be a "peak" if $N_x \geq 50$, and furthermore classified as a "distinct peak" if $N_{x\pm1} < (2/3)N_x$, or as an "ambiguous peak" otherwise.

Snoscan analysis of proposed snoRNAs

We used the program scan-yeast, the yeast version of snoscan (v. 0.9.1) [5], to predict features of the C/D box snoRNAs as currently annotated and features of the snoRNAs as

adjusted by our 5' and 3' sequencing peaks. Snoscan searches for snoRNAs among the query inputs using known yeast RNA methyl sites and yeast rRNA sequences and looks for rRNA targets for these candidates. It also predicts a secondary structure for the candidate snoRNA that includes a terminal stem feature of varying length. To evaluate the adjusted C/D box snoRNA sequences, RNAs with strong peaks in alignments at the 5'-end were used. For each snoRNA, the sequence corresponding to a transcript starting at the genomic location of the peak in the 5'-end alignments and ending at the genomic location of the peak in the 3'-end alignments was then used as input. For the 10 C/D box snoRNAs the corresponding peak in 3'-end alignments was distinct except for snR78. For this snoRNA we found the most common 5'-3' trajectory to choose a representative 3'-end.

SUPPLEMENTAL REFERENCES

1. Jackman, J. E. & Phizicky, E. M. Identification of critical residues for G-1 addition and substrate recognition by tRNA(His) guanylyltransferase. *Biochemistry* **47,** 4817–25 (2008).
2. Rao, B. S., Maris, E. L. & Jackman, J. E. tRNA 5'-end repair activities of tRNAHis guanylyltransferase (Thg1)-like proteins from Bacteria and Archaea. *Nucleic Acids Res.* **39,** 1833–42 (2011).
3. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15,** 182 (2014).
4. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
5. Lowe, T. M. A Computational Screen for Methylation Guide snoRNAs in Yeast. *Science (80-.).* **283,** 1168–1171 (1999).

Supplemental Table 1. Primer sequences for primer extension analysis.

snoRNA	Primer Sequence (5' to 3')
snR75	CAGACTCGTCATCTATAAATATCTC
snR45	ATCATCGTATTCTTTTAAAAACTTGG
snR78	TCTAGTTTGTAAAAATTTATTTTGGTC
snR39	GTATTGGGATACAATGGATTTTCATC
snR70	GCCCGGTCGACAACAAATCATC
snR56	ACAGGTCTGTGTTAATATATTTTTTC
snR61	AGCAGAACTGAATAGTAAATTTTATC

Supplemental Table 2. Other ncRNA with statistically significant changes in upstream/downstream distributions. Percentage of reads aligning in the upstream (top) or downstream (bottom) distributions for ncRNA with significant differences between Thg1-WT and BtTLP-complemented strains evaluated at adjusted p-value < 0.05. These ncRNAs were not investigated further because the effect size as indicated by the RNA-seq data was not large enough (by either fractional or absolute difference).

RNA	Sample	-3	-2	A_{-1}	U_{-1}	C_{-1}	G_{-1}	+1	Counts
5.8S rRNA	ScThg1	0.8%	0.8%	8.4%	<0.1%	<0.1%	<0.1%	89.8%	$6.1x10^6$
	BtTLP	0.8%	0.8%	7.8%	0.1%	<0.1%	<0.1%	90.4%	$7.1x10^6$
5S rRNA	ScThg1	0.4%	<0.1%	<0.1%	<0.1%	0.8%	<0.1%	98.7%	$9.4x10^3$
	BtTLP	0.3%	0.2%	<0.1%	0.2%	0.8%	0.1%	98.3%	$1.0x10^4$

RNA	Sample	+1	+2	+3	+4	Counts
5.8S rRNA	ScThg1	98.9%	0.8%	0.1%	0.1%	$5.5x10^6$
	BtTLP	98.8%	0.8%	0.2%	0.2%	$6.5x10^6$
$mt\text{-}tRNA^{Asp}$	ScThg1	94.6%	0.3%	5.1%	<0.1%	$3.0x10^3$
	BtTLP	98.1%	0.3%	1.5%	<0.1%	$2.5x10^3$
snR66	ScThg1	34.1%	65.7%	0.2%	<0.1%	$3.3x10^3$
	BtTLP	39.6%	60.0%	0.2%	0.2%	$1.0x10^4$
$tRNA^{Gly}_{CCC}$	ScThg1	89%	<1%	11%	<1%	271
	BtTLP	98%	<1%	2%	<1%	190
$tRNA^{Val}_{AAC}$	ScThg1	67%	<1%	26%	8%	172
	BtTLP	85%	<1%	12%	3%	143