HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2019 October 15.

Published in final edited form as:

Nat Methods. 2019 May ; 16(5): 437–445. doi:10.1038/s41592-019-0370-6.

Unbiased screen of RNA tailing activities reveals a poly(UG) polymerase

Melanie A. Preston[1,5], **Douglas F. Porter**[1,6], **Fan Chen**[2], **Natascha Buter**[1,5], **Christopher P. Lapointe**[1,7], **Sunduz Keles**[2,3], **Judith Kimble**[1,4], and **Marvin Wickens**[1,*]

[1]Department of Biochemistry, University of Wisconsin, Madison, WI USA

[2]Department of Statistics, University of Wisconsin, Madison, WI USA

[3]Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI USA

[4]Howard Hughes Medical Institute, University of Wisconsin, Madison, WI USA

[5]Present address: Promega Corporation, Madison, WI USA

[6]Present address: Program in Epithelial Biology, Stanford University Medical School, Stanford, CA USA

[7]Present address: Department of Structural Biology, Stanford University, Stanford CA USA

Abstract

Ribonucleotidyl transferases (rNTases) add non-templated ribonucleotides to diverse RNAs. We developed TRAID-Seq, a screening strategy in *S. cerevisiae* to identify sequences added to a reporter RNA at single-nucleotide resolution by overexpressing candidate enzymes from different organisms. The rNTase activities of 22 previously unexplored enzymes were determined. In addition to poly(A)- and poly(U)-adding enzymes, we identified a C-adding enzyme that is likely part of a two-enzyme system that adds CCA to tRNAs in a eukaryote; a nucleotidyl transferase that adds nucleotides to RNA without apparent nucleotide preference; and a poly(UG) polymerase, *C. elegans* MUT-2, which adds alternating U and G nucleotides to form poly(UG) tails. MUT-2 is known to be required for certain forms of RNA silencing, and mutations in the enzyme that are

*Corresponding author, wickens@biochem.wisc.edu.

AUTHOR CONTRIBUTIONS

M.A.P, J.K., and M.W. designed experiments; M.A.P. performed the experiments and analyzed data unless otherwise noted. D.F.P. wrote the PuppyTails program used to analyze TRAID-Seq data, including “tail-o-grams.” F.C. and S.K. performed statistical analyses of tail sequence motifs. N.B. prepared *N. crassa and C. albicans* TRAID-Seq samples. C.P.L wrote the script used to calculate total nucleotide incorporation. M.A.P. and M.W. wrote the paper, with contributions from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing interests.

Data availability

All sequencing data that support the findings of this study have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and are accessible through the GEO Series accession number GSE123478.

Code availability

All custom scripts have been made available at https://github.com/melanieapreston/PuppyTails.

Life Sciences Reporting Summary

Further information on experimental design is available in the Life Sciences Reporting Summary.

defective in silencing fail to add poly(UG) tails in our assay. We propose that MUT-2 poly(UG) polymerase activity is required to promote genome integrity and RNA silencing.

INTRODUCTION

Covalent modifications pervade biological regulation. RNAs are extensively modified: 5′ termini often are capped, internal positions are altered both on ribose rings and bases, and 3′ termini receive untemplated nucleotides, referred to as “tails”. In eukaryotes, tails occur on most classes of RNA and control their processing, transport, stability, and function. Tails and enzymes that add them are critical in biology. Uridylation is implicated in tumorigenesis, proliferation, stem cell maintenance, and immunity[1–3], and polyadenylation in early development, cancer, and memory[4–6]. Global approaches are needed to uncover previously undetected tailing systems.

The DNA polymerase β-like superfamily of nucleotidyl transferases append nucleotides to divergent substrates, including RNAs, nucleotides, and antibiotics[7,8]. Ribonucleotidyl transferases (rNTases) add nucleotides to RNAs without using a template, and include poly(A) polymerases (PAPs), poly(U) polymerases (PUPs, aka TUTases), and CCA-adding enzymes that act on tRNAs[9]. PAPs and PUPs cannot be distinguished definitively by their protein sequences.

We suspected that other types of rNTases and tails exist but have escaped detection. Studies *in vitro* and in *Xenopus* oocytes identified many rNTase activities[10,11], but are incompatible with genome-wide analyses. Similarly, powerful sequencing methods can identify tails on cellular RNAs without bias[12,13]. Tails could be missed if added at specific times, expressed in certain cell types, exist only transiently, or termini are blocked.. The challenge is to uncover all forms of tails, and identify enzymes responsible, at a genome-wide scale.

We developed an approach to identify enzymes that add non-templated nucleotides to RNAs. Candidate rNTases were tethered to a reporter RNA in *S. cerevisiae*, and the number and identity of nucleotides added were determined at single-nucleotide resolution. The approach revealed previously undetected enzymes and tails, including a eukaryotic system with separate enzymes that add CC and A to tRNAs and a poly(UG) polymerase that adds alternating U and G residues. Mutations in the gene encoding this poly(UG) polymerase elevate transposition frequency[14,15], disrupt RNA silencing[16–18], and impair RNA interference[19–22]. The poly(UG) polymerase and poly(UG) tails likely are required for these events.

RESULTS

An *in cellulo* tethering assay identifies rNTase activities

To identify rNTase activities we developed TRAID-Seq (tethered rNTase activity identified by high-throughput sequencing). Enzymes were fused to MS2 coat protein (MS2), and co-expressed in yeast with a reporter RNA bearing high-affinity MS2 binding sites. The interaction of MS2 with its binding sites tethered the fusion protein to the RNA[23], and circumvented proteins that might bring the rNTase to its endogenous substrates.

To develop a reporter RNA, we first expressed an RNase P-derived RNA bearing two MS2 binding sites in cells containing MS2-PUP fusions (*C. elegans* PUP-2 or *S. pombe* Cid1, Supplementary Fig. 1a). RT-PCR analysis designed to detect U or A tails revealed that U tails were added and required a functional active site; however, high levels of endogenous polyadenylation in the absence of expressed rNTases complicated analysis (Supplementary Fig. 1b, c). We constructed an alternative RNA substrate based on *S. cerevisiae* $tRNA^{Ser(AGA)}$, in which its four-base pair variable arm was replaced with one MS2 binding site (Fig. 1a). This reporter tRNA had virtually undetectable background polyadenylation, judged by gel analysis of reaction products (Fig. 1b), and enabled unambiguous classification of rNTase activities.

The assay accurately recapitulated activities of well-characterized rNTases. As proof-of-principle, we analyzed *C. elegans* PUP-2[11] (*Ce*PUP-2), *S. pombe* Cid1[11,24] (*Sp*Cid1), and a known PAP, *C. elegans* GLD-2[25] (*Ce*GLD-2). In RT-PCR assays, a U-specific primer yielded products with *Ce*PUP-2 and *Sp*Cid1, while an A-specific primer yielded products with *Ce*GLD-2 (Fig. 1b). Tails were not detected using catalytically-inactive mutants of *Ce*PUP-2 and *Sp*Cid1, nor in cells that lacked the proteins or reporter RNA (Fig. 1b).

To identify tails of any nucleotide composition and length, we used high-throughput sequencing (Fig. 1c). Total RNA from each sample was ligated to a DNA adapter to attach a known sequence to the 3′ ends of all RNAs. The adapter enabled sequencing of added tails and introduced a seven-nucleotide randomized sequence (random heptamer) to facilitate computational removal of PCR duplicates. Following reverse transcription, samples were PCR-amplified with primers specific for the reporter tRNA and 3′ adapter. Gel-purified products were subjected to Illumina paired-end sequencing.

We computationally extracted added tails, defined as nucleotides between the reporter tRNA 3′ end (including the CCA) and the random heptamer. After removing PCR duplicates, we quantitated and plotted the number of unique tails, tail length, and the nucleotide composition at each detected tail length in “tail-o-grams”. In these plots, each tail length was assessed as a population to determine the percent of each nucleotide added among all tails of that length, and proportions were color-coded by nucleotide. Numbers of reads at each tail length were normalized to the number of unique random heptamers (TPMH, tails per million heptamers) and displayed on a log scale.

The assay was accurate and sensitive. *Ce*PUP-2 and *Sp*Cid1 added tails primarily of uridines, and *Ce*GLD-2 added tails of adenosines (Fig. 1d–e, Supplementary Fig. 2a), consistent with their known specificities. Furthermore, the high sensitivity enabled detection of secondary nucleotide addition preferences. For example, *Sp*Cid1 added uridine tails with 8.6% adenosine (Supplementary Fig. 2a), consistent with its ability to add both A and U *in vitro*[24,26].

TRAID-Seq circumvents the need for purified enzymes and precisely identifies thousands of independently added tails, enabling sensitive determination of their sequences and relative abundances.

PUPs, PAPs, and CCA-adding enzymes

We used TRAID-Seq to analyze nucleotide specificities of characterized and previously untested rNTases. We tested 40 proteins from seven species: *Homo sapiens* (*Hs*, Fig. 2a), *Caenorhabditis elegans* (*Ce*, Fig. 2b), *Aspergillus nidulans* (*An*), *Candida albicans* (*Ca*), *Neurospora crassa* (*Nc*), *Schizosaccharomyces pombe* (*Sp*), and *Saccharomyces cerevisiae* (*Sc*) (Fig. 2c). Candidate rNTases were identified by the presence of a characteristic G(G/S) $X_{7–13}$ DhDh motif and a downstream third aspartate[9]. To focus on noncanonical rNTases, we included putative rNTases with at least a partial type II nucleotide recognition motif (NRM)[8,9], and excluded canonical PAPs, which are distinguished by a type I NRM[9].

Nucleotide addition activities were classified by the nucleotide composition of all tails added to the reporter tRNA (Fig. 2, Table 1). For example, if added tails consisted of primarily uridines, then the rNTase was classified as a PUP. By this criterion, we uncovered two PUPs and 12 previously uncharacterized PAPs. We also identified likely CCA-adding enzymes in *N. crassa, C. albicans* and *C. elegans*, consistent with homology predictions. Tails added by these enzymes, *Ce*HPO-31, *Ca*Cca1, and *Nc*NCU08022, are primarily composed of C and A (Fig. 2b, c) and show an enrichment for the repeating CCA pattern. The *p*-values of CCA occurrence among tails added by each enzyme, determined using a one-sided Wald's test, are highly significant (adjusted *p*-values less than 1.6×10^{-22}).

Enabled by the sensitivity of TRAID-Seq, we confirmed nucleotide specificities of previously characterized rNTases[10,11,24,25,27–32] and identified surprising secondary preferences in certain enzymes. *Sp*Cid13 and *Sp*Cid14 were both previously identified as PAPs[27] but also added other nucleotides in TRAID-Seq. *Sp*Cid13 added 90.3% adenosine (s.d. 0.3%; n=4), yet also added 6.0% cytosine (s.d. 0.3%; n=4; Fig. 2c, Supplementary Fig. 2b). *Sp*Cid14 added 77.9% adenosine (s.d. 1.2%; n=3) and 19.7% guanosine (s.d. 0.8%; n=3; Fig. 2c, Supplementary Fig. 2c). Analysis of the patterns of nucleotides added by enzymes with secondary preferences revealed no specific sequence motifs within the tails.

Application of TRAID-Seq enabled us to identify new PAPs, PUPs, and CCA-adding enzymes (Fig. 2d) and reveal enzymes with previously unknown activities, as described below.

C tails and a eukaryotic two-enzyme CCA-adding system

We identified a *S. pombe* rNTase that adds primarily cytidines. Based on sequence similarity, *S. pombe* SPAC1093.04 (*Sp*SPAC1093.04) is predicted to be a CCA-adding enzyme, a highly conserved rNTase subfamily that adds CCA to tRNA 3′ termini. In TRAID-Seq, *Sp*SPAC1093.04 yielded tails predominantly of oligo(C) or oligo(A) on reporter tRNAs with a CCA 3′ end (Fig. 2c, Fig. 3a; cytosine=46.0%, s.d. 6.0%; adenosine = 52.8%, s.d. 5.9%; n=5). The oligo(A) may have been added by endogenous PAPs in the TRAMP complex[33], perhaps in competition with the tethered enzyme. Furthermore, reporters with CC 3′ termini received almost exclusively oligo(C) (Fig. 3a). Tails added by *Sp*SPAC1093.04 and the *S. cerevisiae* CCA-adding enzyme (*Sc*Cca1) were distinct (Fig. 3a, b). Most tails added by *Sc*Cca1 contained multiple CCA motifs. In contrast, *Sp*SPAC1093.04 added long cytosine stretches of up to 19 cytosines. Computational

analyses of sequence motifs in tails added by *Sp*SPAC1093.04 and *Sc*Cca1 confirmed differences in their activities: the trinucleotide CCA was highly enriched with *Sc*Cca1 but not *Sp*SPAC1093.04 (Fig. 3c). Both enzymes added tails significantly enriched for CC dinucleotides, as expected (Fig. 3c; Supplementary Fig. 3). We conclude that *Sp*SPAC1093.04 possesses a distinctive C-addition activity.

The *S. pombe* genome encodes a second enzyme (*Sp*SPCC645.10) in the CCA-adding enzyme subfamily, which yielded tails almost entirely of adenosines (Fig. 2c, 96.3%, s.d. 0.7%). Thus, we thought that *Sp*SPAC1093.04 and *Sp*SPCC645.04 might act sequentially to add CCA to tRNAs, with *Sp*SPAC1093.04 adding two C's and *Sp*SPCC645.10 adding the terminal A.

To test this idea, we asked whether the two *S. pombe genes* rescued lethality due to loss of CCA-adding activity in *S. cerevisiae*. We used a *cca1–1* mutant strain bearing a temperature-sensitive (*ts*) allele of the essential *CCA1* gene, which encodes the single protein that adds CCA to tRNAs in *S. cerevisiae*[34]. SPAC1093.04 and SPCC645.10 were expressed in the *cca1–1* strain using the *CCA1* promoter and terminator sequences on single-copy plasmids. Effects on temperature sensitivity were assessed in strains expressing the *S. pombe* proteins either individually or together, and with empty vector controls (Fig. 3d, Supplementary Fig. 4).

Coexpression of both *S. pombe* enzymes rescued absence of endogenous CCA-addition activity in *S. cerevisiae*. *cca1–1* temperature sensitivity at 37°C was fully rescued by co-expression of SPAC1093.04 and SPCC645.10, and by the wild-type *CCA1* positive control. Expression of SPAC1093.04 alone partially suppressed the *cca1–1 ts* phenotype. Expression of SPCC645.10 alone or catalytic-inactive versions of SPAC1093.04 and SPCC645.10 failed to rescue temperature sensitivity. We suggest that SPAC1093.04 and SPCC645.10 cooperate to add CCA to tRNAs to rescue the *cca1–1 ts* phenotype, and that this collaboration is also necessary for CCA addition to tRNAs in *S. pombe.* Both enzymes are essential in *S. pombe*[35] and members of the CCA-adding enzyme subfamily. We propose this is the first identified dual-enzyme CCA-addition system in a eukaryote; our data are supported by a recent report[36,37].

An enzyme with broad specificity

C. elegans F31C3.2 displayed a uniquely broad nucleotide specificity (Fig. 2b, Supplementary Fig. 5a). The majority of nucleotides added were adenosines and uridines, but guanosines and cytosines also were prominent. The nucleotide composition of tails paralleled intracellular ribonucleotide concentrations in *S. cerevisiae* (Supplementary Fig. 5b). The added tails yielded no discernible pattern or sequence motif, and computational analysis of all 16 possible dinucleotide sequences revealed no statistically significant enrichment among the added tails (Supplementary Fig. 5c). We suggest that *Ce*F31C3.2 is relatively indiscriminate in nucleotide preference and so provisionally refer to it as "nucleotide polymerase-1" (*Ce*NPOL-1).

A poly(UG) polymerase required for RNA silencing

C. elegans MUT-2 added tails with a 1:1 ratio of uridines and guanosines (Fig. 2b, Fig. 4a, b). *Ce*MUT-2 tails consisted of striking, polymeric sequences of alternating U and G (Fig. 4b). Computational analysis confirmed repetitive UG addition, and revealed that tails began with either uridine or guanosine We refer to *Ce*MUT-2 as a poly(UG) polymerase.

To determine whether this unusual activity was influenced by the reporter tRNA, we used a different reporter RNA, derived from RNase P RNA (Supplementary Fig. 6a). This RNA had neither a CCA 3′ end nor similarity to tRNAs. *Ce*MUT-2 again added tandem UG repeats, as demonstrated by representative sequences from three biological replicates (Supplementary Fig. 6b), and did so on multiple termini formed on RNase P reporter RNA prior to tailing.

We tested *Ce*MUT-2 in a different organism and cell type to further examine whether UG addition was intrinsic to the protein. *Ce*MUT-2-MS2 protein was expressed in *Xenopus laevis* oocytes via mRNA injection. We then injected a reporter RNA with MS2 binding sites and a sequence distinct from the yeast reporters. Tails were detected on 35–37% of reporter RNA molecules from two biological replicates. Replicate 1 resulted in 43 independently cloned reporter sequences, 16 of which had added tails. Replicate 2 resulted in 31 independently cloned reporter sequences, 11 of which had added tails. All sequenced tails contained tandem UG repeats (Fig. 4c). Short uridine stretches also were observed, perhaps due to *Xenopus* TUT4/TUT7 activity.

We evaluated sequences of *Ce*MUT-2-catalyzed tails from all TRAID-Seq experiments to quantify enrichment of each of the possible 16 dinucleotide pairs (Fig. 4d). 5′-GU-3′ and 5′-UG-3′ were highly enriched, with $-\log_{10}$ (*p*-values) of 7.3 and 6.2. The UG repeats are essentially perfectly repeated throughout the tails added, a remarkable pattern not observed in tails added by known nucleotidyl transferases.

We also assayed a construct corresponding to another predicted splicing isoform of *Ce*MUT-2 (mut-2b; https://wormbase.org/species/c_elegans/gene/WBGene00003499#0-1-3). Only *Ce*MUT-2a exhibited UG-addition activity (Fig. 5a, b).

The poly(UG) polymerase activity of *Ce*MUT-2 (aka RDE-3) likely is critical for RNAi. *Ce*MUT-2 was first identified genetically, in screens for elevated Tc1 transposition in *C. elegans*[14] (hence MUT-2 for “mutator”). Later, the same gene emerged from a screen for genes critical for RNAi (hence RDE-3 for “RNAi-defective”)[19]. The RNAi-defective screen yielded six independent *mut-2* alleles with mutations in regions likely important for catalytic activity (Fig. 5a). We assayed *Ce*MUT-2 proteins bearing each of these mutations, and a *Ce*MUT-2 protein engineered to be inactive (DADA, Fig. 5b). All mutant *Ce*MUT-2 proteins lacked UG-addition activity, and the nucleotide compositions of the few tails detected resembled the catalytically inactive DADA enzyme and vector controls. Since *C. elegans* mutants harboring these alleles are defective for suppression of transposition, RNAi interference, and RNA silencing, we propose that poly(UG) polymerase activity is important in those events.

DISCUSSION

With TRAID-seq we tested proteins identified through sequence similarity to rNTases, although the approach could be applied to enzymes that catalyze any RNA modification detectable through sequencing, including certain base modifications. Despite its sensitivity and ability to assay activity without purification of the modified RNA or candidate rNTase, limitations of TRAID-Seq arise from the use of an artificial substrate to which the enzyme is tethered, and from measuring activity in a foreign cell.

The active site regions of the 17 PAPs and PUPs we identified bear on how U and A are distinguished by rNTases. A histidine in rNTase active sites can dictate nucleotide preference for U[38–41]. Yet three PUPs we uncovered (*Sp*Cid16, *Ce*PUP-3 and *Ce*F43E2.1) lack that histidine (Table 1). Similarly, *Ce*NPOL-1 and *Ce*MUT-2 can add purines, yet possess the active-site histidine. These findings emphasize the complexity of nucleotide preferences among rNTases and the need for further structural analyses.

TRAID-Seq may miss effects of the natural RNA substrates, co-factors or protein partners of rNTases. For example, in mammalian cells, *Hs*TUT1/Star-PAP adds U's to U6-snRNA[29], but adds A's to a variety of mRNAs[30]. In TRAID-Seq, we detected a strong preference for A (adenosine 89.5%, s.d. 1.4%), and only low levels of incorporation of other nucleotides (U=3.2%, s.d. 0.7%; C= 2.5%, s.d. 0.4%; G=4.8%, s.d. 0.6%). A specific phosphoinositide enhances *Hs*TUT1/Star-PAP A addition activity *in vitro*[30] and may underlie these differences. In addition, *Aspergillus An*CutA adds CU-rich tails to RNAs *in vivo* and prefers CTP *in vitro*[42,43]. In TRAID-Seq, *An*CutA predominantly added A (91.8%, s.d. 0.4%) vs. C (5.9%, s.d. 0.2%) or U (1.9%, s.d. 0.3%). In *Aspergillus, An*CutB collaborates with *An*CutA to form CU-rich tails[42] but also added A's in TRAID-Seq (98.7%, s.d. 0.2%) vs. C (0.4%, s.d. 0.05%) or U (0.3%, s.d. 0.07%). These findings emphasize TRAID-Seq as a starting point for further studies.

The sensitivity of TRAID-Seq revealed previously undetected nucleotide addition capabilities that may underlie the addition of *in vivo* tails that have been enigmatic. For example, three human PAPs (*Hs*TENT2, *Hs*TENT4b, and *Hs*TUT1) are capable of G addition, albeit at a low level in our system (Fig. 2a), and could contribute to G-addition on mRNAs in human cells[12]. Indeed, *Hs*TENT4a and *Hs*TENT4b were recently implicated in G-addition to mRNAs, which protects them from deadenylation[44]. The discovery of other human PAPs that add mixed tails might indicate that other classes of RNAs are subject to regulation by G-addition. The abilities of *Sp*Cid13 and *Sp*Cid14 to add C and G, respectively, in addition to A, might suggest an alternate mechanism of RNA regulation in *S. pombe*. We suspect that the nature and roles of tails are more varied than previously realized.

C. elegans NPOL-1 added tails composed of random combinations of all four nucleotides. The levels of incorporation mirror intracellular ribonucleotide concentrations, which may influence the proportions of nucleotides added. *Ce*NPOL-1 diverges in sequence from other enzymes that can catalyze template-independent addition of any nucleotide tail[45], and belongs to a different subfamily of nucleotidyl transferases[8]. Addition of random nucleotides

within, or at the end of, homopolymeric tails could interfere with RNA function[44]. It will be of interest to test the roles of *Ce*NPOL-1 *in vivo*.

We propose that *Sp*SPAC1093.04 and *Sp*SPCC645.10 constitute a two-enzyme system that catalyzes CCA addition to tRNAs in *S. pombe*. This is strongly suggested by their nucleotide specificities and ability to jointly complement a yeast strain lacking a functional CCA-adding enzyme, consistent with a recent report[36,37]. These studies are the first observations of a two-enzyme system in a eukaryote, and await verification in *S. pombe*.

*Ce*MUT-2, the poly(UG) polymerase, is remarkable both in enzyme activity and its roles in RNA biology. We detected tails of up to 18 perfect UG repeats; indeed, longer tails were likely added but went undetected due to sequencing read limitations. The number of UG's added *in vivo* is not yet known. Alternating U and G addition bears comparison to CCA addition via the single active site of CCA-adding enzyme[46,47]. Perhaps *Ce*MUT-2 promotes consecutive rounds of UG-addition by similarly repositioning the 3′-most UG relative to the active site.

The diverse roles of *Ce*MUT-2 – preserving genome integrity by suppressing transposition[14,15], silencing transgenes[16–18] and promoting RNAi due to exogenous dsRNA[19–22] – all likely reflect the same underlying molecular mechanism. *Ce*MUT-2 biological functions likely hinge on its poly(UG) polymerase activity, since mutations identified in mutator and RNAi-defective *mut-2* mutants abrogate its enzymatic activity (Fig. 5a, b). *Ce*MUT-2 increases the abundance of secondary small RNAs during RNAi[19,21], suggesting that UG tails are important in RdRP-based secondary small RNA synthesis or stabilization. In one model, *Ce*MUT-2 adds poly(UG) to the 3′ end of sliced RNAs generated in an Ago-dependent process. The poly(UG) tails would provide a distinctive mark to recruit RdRP, either directly or via a separate UG-binding protein (Fig. 5c). In either scenario, the tail could be single-stranded, or form a more complex structure (depicted simply as UG pairing in Fig. 5c). By recruiting RdRPs to amplify small RNA pools, and perhaps by directly stabilizing sliced RNAs, poly(UG) tails could promote long-term gene silencing known to occur in *C. elegans*[48–50]. Suppression of transposition by *Ce*MUT-2 implies that it acts on an RNA vital in that process. Identification of the RNA targets of *Ce*MUT-2 should provide an entree into roles of poly(UG) polymerases and the tails they add.

ONLINE METHODS

Plasmid Construction

To enable overexpression of rNTases as MS2 coat protein fusions in *S. cerevisiae*, the MAP72 MS2 cassette vector was constructed. YEplac 181 (*LEU2* 2*μ*)[51] was digested with *Hind*III and *Xho*I. Then each portion of the MS2 cassette was subcloned with unique restrictions sites, resulting in the following insert: *S. cerevisiae TEF1* promoter, MS2 coat protein, a multiple cloning site to insert the rNTase to test (consisting of *Bam*HI, *Xma*I/*Sma*I, *Not*I, *Xba*I, *Pst*I, and *Kpn*I sites), SV40 nuclear localization signal, RGS(H_6) sequence to verify rNTase expression by Western blotting, and *S. cerevisiae ADH1* terminator sequence.

Each rNTase tested was cloned into MAP72 by amplifying the genes indicated in Supplementary Table 1 using the primers listed. All inserts were sequenced to confirm identity and lack of mutations. Site-directed mutations were made using standard methods with oligomers corresponding to the mutated sequences.

The MAP80A tRNA reporter vector was constructed using a $tRNA^{His}$ expression cassette, MAB812A[52]. $tRNA^{His}$ sequence was removed by digestion with *Xho*I and *Bgl*II. Then DNA corresponding to the tRNA reporter sequence was inserted by annealing overlapping oligomers to construct both strands of the DNA sequence. The tRNA reporter is *S. cerevisiae* $tRNA^{Ser(AGA)}$ altered to contain one MS2 stem loop sequence (underlined) in place of the endogenous $tRNA^{Ser(AGA)}$ variable arm (5′-GGCAACTTGGCCGAGTGGTTAAGGCGAAAGATTAGAAATCTTTACATGAGGATCACCCATGTCGCAGGTTCGAGTCCTGCAGTTGTCG-3′).

A *CCA1* cassette vector was constructed using YCplac 111 (*LEU2 CEN*)[51] in order to express *CCA1*, SPAC1093.04, or SPCC645.10 with the same promoter and C-terminal epitope tag [$RGS(H)_6$]. BY4741 yeast genomic DNA was used as a template to generate an amplicon consisting of *LEU2 CEN* vector sequence at the 5′ end, the *CCA1* promoter sequence, and a 3′ terminal sequence corresponding to the multiple cloning site of MAP72 using 5′-GAAACAGCTATGACCATGATTACGCCAAGCTTACTAGTAGCTACTTCAGGGACAAGCAAC-3′, and 5′-ACCCTGCAGTCTAGAAGGCGGCCGCGTGGATCCACACAAAAAAAGCCCTTATAACCCACG-3′. MAP72 was used as a template to generate an amplicon consisting of the multiple cloning site, $RGS(H_6)$ sequence, *ADH1* terminator sequence of MAP72, and *LEU2 CEN* vector sequence at the 3′ end using 5′-GGATCCACGCGGCCGCCTTCTAGACTGCAGGGTACCAGAGGTTCTCACCACCACCACCAC-3′ and 5′-CCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCCTCGAGCGGTAGAGGTGTGGTCA-3′. These two amplicons were combined with *LEU2 CEN* vector (YCplac111) linearized with *Pst*I/*Sac*I and assembled by Gibson cloning[53]. The *CCA1* cassette sequence was confirmed by Sanger sequencing. *CCA1*, SPAC1093.04, or SPCC645.10 sequences were subcloned from their respective MAP72-based constructs into the *CCA1* cassette for expression in *cca1-1* yeast.

To construct the MAP136 MUT-2 oocyte expression vector (pCS2 3HA MS2-MUT-2 WT), MUT-2a was PCR-amplified from its MAP72-based vector using 5′-CTACCATGGATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGAC-3′ and 5′-ACTCTCGAGTTAGTGGTGGTGGTGGTGGTGAGAACCTCTGGTACCCTGCAGTACAAATGA-3′ and then cloned into the *Nco*I/*Xho*I site of pCS2 3HA MS2. MUT-2 DNA sequence was verified prior to oocyte injections.

All constructs used in this study are available from the authors upon request.

Yeast Growth

BY4741 yeast were co-transformed using standard methods[54] with a plasmid expressing the reporter RNA and a plasmid expressing the rNTase of interest, or vector controls, and selected on synthetic yeast medium lacking uracil and leucine (SD-Ura-Leu). Cultures were inoculated independently with single colonies to produce biological replicates, grown to saturation, and then diluted to 0.1 OD_{600}/mL and grown to log phase (0.8–1 OD_{600}/mL). Cells were spun down in pellets of 25 OD_{600} (approximately 5×10^8 cells) and stored at −80°C until RNA extraction or protein expression analysis. We performed Western blotting with mouse anti-RGS-His Antibody (1:2500 dilution, 5PRIME/Qiagen). Only those samples with clear expression of the rNTase fusion protein were analyzed by high-throughput sequencing.

cca1–1 yeast were co-transformed with vectors as listed in Fig. 3 and Supplementary Fig. 4 using standard methods[54], and selected on SD-Ura-Leu plates at room temperature. Colonies were selected and grown to saturation in SD-Ura-Leu liquid media. Cultures were diluted to 0.5 OD/mL followed by three 10-fold serial dilutions, spotted on SD-Ura-Leu plates, and incubated at room temperature (23°C) for 4 days or 37°C for 3 days.

RNA Extraction

RNA was extracted from 25 OD of yeast corresponding to each sample by modification of a previously described method[55]. To each sample, 0.5 g of 0.5 mm acid washed beads (Sigma-Aldrich), 0.5 mL of RNA ISO buffer (500 mM NaCl, 200 mM Tris-Cl pH 7.5, 10 mM EDTA, 1 % SDS) and 0.5 mL of phenol-chloroform-isoamyl alcohol pH 6.7 (PCA, Fisher Scientific) was added. Samples were lysed with 10 cycles that each consisted of vortexing for 20 seconds and incubation on ice for 30 seconds. 1.5 volumes (relative to starting amount of RNA ISO Buffer) of RNA ISO Buffer and of PCA were added, and samples were centrifuged at 4°C to separate phases. The aqueous layer was transferred to a pre-spun phase-lock gel (heavy) tube (5PRIME/Quantabio); an equal volume of PCA was added and mixed prior to centrifugation at room temperature to separate phases. The aqueous layer was transferred to 2 new tubes for ethanol precipitation with 2 volumes of 100% ethanol followed incubation at −80°C for 1 hour to overnight. Precipitated RNA was pelleted by centrifugation at 4°C. Each pellet was dissolved in 25 μL nuclease-free water and combined into 1 tube per sample. Co-purifying DNA was digested with 20 U of Turbo DNase (Invitrogen) at 37°C for 4 hours. RNA was cleaned up with the GeneJET RNA Purification Kit (Thermo Fisher Scientific) and eluted with 50 μL of DEPC-treated water.

RT-PCR Experiments

RT-PCR experiments to detect A tails or U tails on an RNase P RNA reporter[56,57] (see Supplementary Fig. 1) were performed by using a tail-specific reverse transcription step with 5 pmol of a T_{33} or A_{33} DNA primer and 100 ng of total RNA using ImProm-II Reverse Transcriptase (Promega Corporation). Then the resulting reactions were PCR-amplified using reporter-specific primers (5′-TCGAGCCCGGGCAGCTTGCATGC-3′ and 5′-GGGAATTCCGATCCTCTAGAGTC-3′). If a tail was added to the RNase P RNA reporter, then the RT reaction would produce cDNA, and the PCR would result in an amplicon.

RT-PCR experiments to detect tails added to the tRNA reporter were performed as described with the RNase P RNA reporter but with the following modifications. PCR amplification was performed with a forward primer specific to the 5′ end of the tRNA (5′-GGCAACTTGGCCGAGTGGTTAAGG-3′) and a reverse primer specific to the 3′ end of the tRNA with an A tail or U tail, respectively: 5′-AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGGCGACAACTGC-3′ or 5′-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGCGACAACTGC-3′. If a tail was added to the tRNA reporter, then the RT reaction would produce cDNA, and the tail-specific PCR would result in an amplicon.

TRAID-Seq Library Preparation

Total RNA (100 ng) was ligated with 20 pmol of a PAGE gel-purified 5′ adenylated primer containing a 7-nucleotide random DNA sequence (random heptamer), Illumina TruSeq adapter sequence and a 3′ dideoxycytidine (5′-A(pp) NNNNNNN TGGAATTCTCGGGTGCCAAGG ddC-3′) using 200 U of T4 RNA ligase 2, truncated KQ (New England BioLabs) in a 20μL reaction with 16°C overnight incubation. This ligation added the random heptamer and Illumina TruSeq adapter sequence to the 3′ end of the RNAs in the sample.

Half of the ligation reaction (10 μL) was reverse transcribed using 5 pmol of Illumina RNA RT primer (5′-GCCTTGGCACCCGAGAATTCCA-3′) and ImProm-II Reverse Transcriptase (Promega Corporation) with 1.5 mM $MgCl_2$ and 0.5 mM dNTPs, according to manufacturer's instructions.

Samples were PCR-amplified with a forward primer consisting of Illumina-specific sequences and tRNA reporter-specific sequences (underlined)(5′-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATC GAGGATCACCCATGTCGCAG-3′) and a reverse Illumina RNA PCR Primer with various indices used for multiplexing, using GoTaq Green PCR Master Mix (Promega Corporation). PCR products were run on an 8% polyacrylamide 8M urea gel and gel extracted. Resulting samples for each sequencing run were combined in equimolar amounts and run on an Illumina HiSeq2000 or HiSeq2500 (2×50 bp or 2×100 bp), to produce approximately 1×10^6 reads per sample.

Experiments with the RNase P RNA reporter were performed essentially as described above but with one modification. For TRAID-Seq, the 5′ primer used for PCR amplification was specific for the RNase P RNA reporter (5′-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATCG TCTGCAGGTCGACTCTAGAAA-3′).

TRAID-Seq Data Analysis

Reads resulting from sequencing of TRAID-Seq samples were analyzed using a group of Python scripts that we call the "PuppyTails" program. Briefly, PuppyTails identifies sequences corresponding to the tRNA reporter, CCA end of the tRNA, and added tail in read 1. In read 2, the program identifies the random heptamer sequence, added tail sequence, and, if read length allows, the CCA end and tRNA reporter sequence. Reads were collapsed into

unique ligation events using the random heptamer and then compared to identify and remove sequences resulting from PCR amplification (PCR duplicates). The number of unique times that each tail sequence is observed is counted. Tail sequences are sorted by length to calculate the nucleotide composition at each tail length and the number of tails per million heptamers (TPMH) measured for each tail length; these data are plotted as tail-o-grams (for example, Fig. 1d, e). Tails of all lengths were used for analyses of nucleotide composition, but for all tail-o-grams, tails of 5 nucleotides or greater are shown for clarity. Tails of 1–4 nucleotides included poly(A) sequences (likely added by endogenous PAPs in *S. cerevisiae*) and also nucleotides not explained by the activity of the rNTase. The sequences at these tail lengths were the same in the absence of expressed enzyme and with catalytically inactive versions of the enzymes tested (see Fig. 5b).

A Perl script was used to calculate the overall nucleotide composition of tails added by a given rNTase for each of its biological replicates. All tail lengths were assessed as one population. The abundance of observed tail sequences were factored into calculations of nucleotide composition. Nucleotide addition percentages reported in this study were generated using this analysis approach. Data shown in Fig. 2a–c were generated using this type of analysis.

Protocol Accessibility

A step by step version of the TRAID-Seq workflow is available online on Protocol Exchange (DOI: 10.1038/protex.2019.016)

Computational Analyses of Sequence Motifs

To analyze tail sequences, a general feature screening with a random forest application[58] was performed at the replicate level. We first quantified the number occurrences of all oligonucleotides (k=1, 2, 3, 4) within each tail sequence and utilized the resulting set of 340 features, as well as the length of the tail. The variable importance, defined as the percent mean decrease in accuracy (with 500 trees, 113 candidate variables at each split, minimum node size of 5), were estimated for all features. We define the selected features as those whose importance measures are greater than 4% across replicates. We fitted a Poisson regression model in which the response variable was tail sequence counts.

Tails added by S. cerevisiae Cca1, S. pombe SPAC1093.04, and predicted CCA-adding enzymes. The above selected features were used as covariates. *P*-values from individual replicates, calculated from one-sided Wald's test, were aggregated using Fisher's (n<4) or Wilkinson's (n>=4) method, followed by multiplicity correction with the Bonferroni procedure. This process identified oligonucleotides that differ between tails added by *S. cerevisiae* Cca1 and *S. pombe* SPAC1093.04 at level 0.05.

Tails added by C. elegans MUT-2. We evaluated the impacts of 16 dinucleotides by formally testing for their effects by a comparison of a null model without each dinucleotide and the alternative model deduced from random forest filtered set of features plus other dinucleotides. This procedure identified UG and GU as the most significant dinucleotides.

In vitro Transcription

pCS2 3HA MS2-MUT-2 (MAP136) was linearized with *Sac*II, and 3 μg of linearized plasmid was transcribed with Ampliscribe SP6 High Yield Transcription Kit (Epicentre), according to manufacturer's instructions. pLGMS2-luc (RNA with three MS2-binding sites)[10,59] was linearized with *Bgl*II, and 1 μg of linearized plasmid was transcribed with T7 Flash In Vitro Transcription Kit (Epicentre), according to manufacturer's instructions. Transcription reactions included $m^7G(5')ppp(5')G$ RNA Cap Structure Analog (New England Biolabs).

Tethered Function Assays and Oocyte RNA Extraction

Xenopus laevis oocyte manipulations and injections were performed as in previous studies[10,59,60].

Tethered function assays were conducted essentially as previously described[11]. Briefly, Stage VI oocytes were injected with 50 nL of 600 ng/μL capped mRNA encoding MS2-HA-MUT-2 protein. After 6 hours, the same oocytes were injected with 50 nL of 3 ng/μL pLGMS2-luc reporter mRNA. After 16 hours, oocytes were collected, lysed, and assayed. Three oocytes were used to confirm protein expression. Total RNA was extracted from oocytes using TRI reagent (Sigma-Aldrich), as described previously[11], then treated with 8 U of Turbo DNase (Invitrogen) at 37°C for 1 hour, and cleaned up with the GeneJET RNA Purification Kit (Thermo Fisher Scientific).

Oocyte RNA Analysis and Tail Sequencing

Oocyte total RNA (100 ng) was ligated with 20 pmol of the 5′ adenylated primer as described above. This ligation added the random heptamer sequence and a known sequence to the 3′ ends of RNAs in the sample for tail sequence-independent analyses. Half of the ligation reaction (10 μL) was reverse transcribed as described above.

Samples were PCR-amplified with a forward primer specific to the RNA reporter (5′-CTCTGCAGTCGATAAAGAAAACATGAG-3′) and a reverse primer specific to the known sequence added to the 3′ end of the RNA (5′- GCCTTGGCACCCGAGAATTCCA-3′), using GoTaq Green PCR Master Mix (Promega Corporation). PCR products were run on a 1.5% agarose gel, and purified with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). Non-templated A overhangs were added by treating the purified PCR products with 10 U of TaqPlus Precision Polymerase Mixture (Agilent Genomics) in TaqPlus Precision buffer supplemented with 0.2 mM dATP at 70°C for 30 minutes. PCR products were then cloned with the TOPO TA Cloning Kit for Subcloning (Thermo Fisher Scientific) as follows: 6% of the A addition reaction volume (2.4 μL) was combined with 0.6 μL of Salt Solution and 0.7 μL of TOPO Vector and incubated at room temperature for 30 minutes. Reactions were diluted 1 in 4 with water, transformed into DH5α competent cells, and selected on LB agar with 100 μg/mL ampicillin and 75 μg/mL X-Gal for blue/white screening. White colonies were selected, plasmids were extracted, and inserts were sequenced to identify tails added to the reporter. All reporter sequences with added tails are reported in Fig. 4c.

NRM sequence analysis

Protein sequences of known[10,11,24,25,27–33,61–63] and new rNTases tested (excluding CCA-adding enzymes) were aligned using ClustalX 2.1 software to identify the nucleotide recognition motif (NRM). The amino acid for each rNTase reported in Table 1 corresponds to histidine 336 of *S. pombe* Cid1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank members of the Wickens, Kimble, Kennedy and Anderson labs for advice throughout the work. We are particularly grateful to S. Kennedy for discussions concerning MUT-2, and A. Hopper and E. Phizicky for advice on CCA-adding enzymes. We thank A. Shukla and J. Yan for comments on the manuscript. We acknowledge University of Wisconsin Biotechnology Center DNA Sequencing Facility, especially M. Adams and M. Sussmann, for providing Illumina sequencing facilities and services, and M. Harte for assistance with cloning *S. pombe* rNTases. We also thank S. Sasaki for harvesting *Xenopus* oocytes and advice on injections. We are grateful to L. Vanderploeg of the UW Biochemistry Media Laboratory for help with figures. This work was supported by a Ruth Kirschstein National Research Service Award (1F32GM103130–01A1) to M.A.P. and NIH grants GM50942 to M.W. and HG003747 to S.K. J.K. is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Blahna MT, Jones MR, Quinton LJ, Matsuura KY & Mizgerd JP Terminal uridyltransferase enzyme Zcchc11 promotes cell proliferation independent of its uridyltransferase activity. J Biol Chem 286, 42381–42389, doi:10.1074/jbc.M111.259689 (2011). [PubMed: 22006926]
2. Hagan JP, Piskounova E & Gregory RI Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol 16, 1021–1025, doi:10.1038/nsmb.1676 (2009). [PubMed: 19713958]
3. Jones MR et al. Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nat Cell Biol 11, 1157–1163, doi:10.1038/ncb1931 (2009). [PubMed: 19701194]
4. A PS & Laishram RS Nuclear Phosphatidylinositol-Phosphate Type I Kinase alpha-Coupled Star-PAP Polyadenylation Regulates Cell Invasion. Mol Cell Biol 38, doi:10.1128/MCB.00457-17 (2018).
5. Benoit P, Papin C, Kwak JE, Wickens M & Simonelig M PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in Drosophila. Development 135, 1969–1979, doi:10.1242/dev.021444 (2008). [PubMed: 18434412]
6. Kwak JE et al. GLD2 poly(A) polymerase is required for long-term memory. Proc Natl Acad Sci U S A 105, 14644–14649, doi:10.1073/pnas.0803185105 (2008). [PubMed: 18780789]
7. Yue D, Maizels N & Weiner AM CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. RNA 2, 895–908 (1996). [PubMed: 8809016]
8. Aravind L & Koonin EV DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. Nucleic Acids Res 27, 1609–1618, doi:10.1093/nar/27.7.1609 (1999). [PubMed: 10075991]
9. Martin G & Keller W RNA-specific ribonucleotidyl transferases. RNA 13, 1834–1849, doi:10.1261/rna.652807 (2007). [PubMed: 17872511]
10. Kwak JE, Wang L, Ballantyne S, Kimble J & Wickens M Mammalian GLD-2 homologs are poly(A) polymerases. Proc Natl Acad Sci U S A 101, 4407–4412, doi:10.1073/pnas.0400779101 (2004). [PubMed: 15070731]
11. Kwak JE & Wickens M A family of poly(U) polymerases. RNA 13, 860–867, doi:10.1261/rna.514007 (2007). [PubMed: 17449726]

12. Chang H, Lim J, Ha M & Kim VN TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Mol Cell 53, 1044–1052, doi:10.1016/j.molcel.2014.02.007 (2014). [PubMed: 24582499]
13. Gazestani VH, Hampton M, Abrahante JE, Salavati R & Zimmer SL circTAIL-seq, a targeted method for deep analysis of RNA 3' tails, reveals transcript-specific differences by multiple metrics. RNA 22, 477–486, doi:10.1261/rna.054494.115 (2016). [PubMed: 26759453]
14. Collins J, Saari B & Anderson P Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. Nature 328, 726–728, doi:10.1038/328726a0 (1987). [PubMed: 3039378]
15. Yuan JY, Finney M, Tsung N & Horvitz HR Tc4, a *Caenorhabditis elegans* transposable element with an unusual fold-back structure. Proc Natl Acad Sci U S A 88, 3334–3338, doi:10.1073/pnas.88.8.3334 (1991). [PubMed: 1849651]
16. Tabara H et al. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell 99, 123–132, doi:10.1016/S0092-8674(00)81644-X (1999). [PubMed: 10535731]
17. Ketting RF & Plasterk RH A genetic link between co-suppression and RNA interference in *C. elegans*. Nature 404, 296–298, doi:10.1038/35005113 (2000). [PubMed: 10749214]
18. Robert VJ, Sijen T, van Wolfswinkel J & Plasterk RH Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. Genes Dev 19, 782–787, doi:10.1101/gad.332305 (2005). [PubMed: 15774721]
19. Chen CC et al. A member of the polymerase beta nucleotidyltransferase superfamily is required for RNA interference in *C. elegans*. Curr Biol 15, 378–383, doi:10.1016/j.cub.2005.01.009 (2005). [PubMed: 15723801]
20. Lee RC, Hammell CM & Ambros V Interacting endogenous and exogenous RNAi pathways in *Caenorhabditis elegans*. RNA 12, 589–597, doi:10.1261/rna.2231506 (2006). [PubMed: 16489184]
21. Gu W et al. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. Mol Cell 36, 231–244, doi:10.1016/j.molcel.2009.09.020 (2009). [PubMed: 19800275]
22. Jose AM, Garcia GA & Hunter CP Two classes of silencing RNAs move between *Caenorhabditis elegans* tissues. Nat Struct Mol Biol 18, 1184–1188, doi:10.1038/nsmb.2134 (2011). [PubMed: 21984186]
23. Coller JM, Gray NK & Wickens MP mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. Genes Dev 12, 3226–3235, doi: 10.1101/gad.12.20.3226 (1998). [PubMed: 9784497]
24. Rissland OS, Mikulasova A & Norbury CJ Efficient RNA polyuridylation by noncanonical poly(A) polymerases. Mol Cell Biol 27, 3612–3624, doi:10.1128/MCB.02209-06 (2007). [PubMed: 17353264]
25. Wang L, Eckmann CR, Kadyk LC, Wickens M & Kimble J A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. Nature 419, 312–316, doi:10.1038/nature01039 (2002). [PubMed: 12239571]
26. Read RL, Martinho RG, Wang SW, Carr AM & Norbury CJ Cytoplasmic poly(A) polymerases mediate cellular responses to S phase arrest. Proc Natl Acad Sci U S A 99, 12079–12084, doi: 10.1073/pnas.192467799 (2002). [PubMed: 12218190]
27. Saitoh S et al. Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. Cell 109, 563–573, doi:10.1016/S0092-8674(02)00753-5 (2002). [PubMed: 12062100]
28. Tomecki R, Dmochowska A, Gewartowski K, Dziembowski A & Stepien PP Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. Nucleic Acids Res 32, 6001–6014, doi:10.1093/nar/gkh923 (2004). [PubMed: 15547249]
29. Trippe R et al. Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. RNA 12, 1494–1504, doi:10.1261/rna.87706 (2006). [PubMed: 16790842]
30. Mellman DL et al. A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs. Nature 451, 1013–1017, doi:10.1038/nature06666 (2008). [PubMed: 18288197]

31. Schmid M, Kuchler B & Eckmann CR Two conserved regulatory cytoplasmic poly(A) polymerases, GLD-4 and GLD-2, regulate meiotic progression in *C. elegans*. Genes Dev 23, 824–836, doi:10.1101/gad.494009 (2009). [PubMed: 19339688]
32. Rammelt C, Bilen B, Zavolan M & Keller W PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. RNA 17, 1737–1746, doi:10.1261/rna.2787011 (2011). [PubMed: 21788334]
33. Kadaba S, Wang X & Anderson JT Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. RNA 12, 508–521, doi:10.1261/rna.2305406 (2006). [PubMed: 16431988]
34. Aebi M et al. Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*. J Biol Chem 265, 16216–16220 (1990). [PubMed: 2204621]
35. Kim DU et al. Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. Nat Biotechnol 28, 617–623, doi:10.1038/nbt.1628 (2010). [PubMed: 20473289]
36. Preston MA, Porter DF, Chen F, Buter N, Lapointe CP, Keles S, Kimble J, Wickens M Unbiased screen of RNA tailing enzymes at single-nucleotide resolution reveals a poly(UG) polymerase required for genome integrity and RNA silencing. Preprint at bioRxiv https://doi.org/10.110¼22972 (2018).
37. Reid NE, Ngou JS & Joyce PBM *Schizosaccharomyces pombe* contains separate CC- and A-adding tRNA nucleotidyltransferases. Biochem Biophys Res Commun 508, 785–790, doi:10.1016/j.bbrc.2018.11.131 (2019). [PubMed: 30528393]
38. Chung CZ, Jo DH & Heinemann IU Nucleotide specificity of the human terminal nucleotidyltransferase Gld2 (TUT2). RNA 22, 1239–1249, doi:10.1261/rna.056077.116 (2016). [PubMed: 27284165]
39. Munoz-Tello P, Gabus C & Thore S Functional implications from the Cid1 poly(U) polymerase crystal structure. Structure 20, 977–986, doi:10.1016/j.str.2012.04.006 (2012). [PubMed: 22608966]
40. Lunde BM, Magler I & Meinhart A Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. Nucleic Acids Res 40, 9815–9824 doi:10.1093/nar/gks740 (2012). [PubMed: 22885303]
41. Yates LA et al. Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. Nat Struct Mol Biol 19, 782–787, doi:10.1038/nsmb.2329 (2012). [PubMed: 22751018]
42. Morozov IY et al. mRNA 3' tagging is induced by nonsense-mediated decay and promotes ribosome dissociation. Mol Cell Biol 32, 2585–2595, doi:10.1128/MCB.00316-12 (2012). [PubMed: 22547684]
43. Kobylecki K, Kuchta K, Dziembowski A, Ginalski K & Tomecki R Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity toward CTP and increased tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails. RNA 23, 1902–1926, doi:10.1261/rna.061010.117 (2017). [PubMed: 28947555]
44. Lim J et al. Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. Science 361, 701–704, doi:10.1126/science.aam5794 (2018). [PubMed: 30026317]
45. Boule JB, Rougeon F & Papanicolaou C Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. J Biol Chem 276, 31388–31393, doi:10.1074/jbc.M105272200 (2001). [PubMed: 11406636]
46. Xiong Y & Steitz TA Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template. Nature 430, 640–645, doi:10.1038/nature02711 (2004). [PubMed: 15295590]
47. Cho HD, Verlinde CL & Weiner AM Archaeal CCA-adding enzymes: central role of a highly conserved beta-turn motif in RNA polymerization without translocation. J Biol Chem 280, 9555–9566, doi:10.1074/jbc.M412594200 (2005). [PubMed: 15590678]
48. Fire A et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans.* Nature 391, 806–811, doi:10.1038/35888 (1998). [PubMed: 9486653]

49. Vastenhouw NL et al. Gene expression: long-term gene silencing by RNAi. Nature 442, 882, doi: 10.1038/442882a (2006). [PubMed: 16929289]

50. Buckley BA et al. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. Nature 489, 447–451, doi:10.1038/nature11352 (2012). [PubMed: 22810588]

METHODS-ONLY REFERENCES

51. Gietz RD & Sugino A New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. Gene 74, 527–534, doi: 10.1016/0378-1119(88)90185-0 (1988). [PubMed: 3073106]

52. Whipple JM, Lane EA, Chernyakov I, D'Silva S & Phizicky EM The yeast rapid tRNA decay pathway primarily monitors the structural integrity of the acceptor and T-stems of mature tRNA. Genes Dev 25, 1173–1184, doi:10.1101/gad.2050711.

53. Gibson DG et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343–345, doi:10.1038/nmeth.1318 (2009). [PubMed: 19363495]

54. Sherman F Getting started with yeast. Methods Enzymol 350, 3–41, doi:10.1016/S0076-6879(02)50954-X (2002). [PubMed: 12073320]

55. Preston MA, D'Silva S, Kon Y & Phizicky EM $tRNA^{His}$ 5-methylcytidine levels increase in response to several growth arrest conditions in *Saccharomyces cerevisiae*. RNA 19, 243–256, doi: 10.1261/rna.035808.112 (2013). [PubMed: 23249748]

56. SenGupta DJ et al. A three-hybrid system to detect RNA-protein interactions in vivo. Proc Natl Acad Sci U S A 93, 8496–8501, doi:10.1073/pnas.93.16.8496 (1996). [PubMed: 8710898]

57. Stumpf CR, Opperman L & Wickens M Chapter 14. Analysis of RNA-protein interactions using a yeast three-hybrid system. Methods Enzymol 449, 295–315, doi:10.1016/S0076-6879(08)02414-2 (2008). [PubMed: 19215764]

58. Liaw A & Wiener M Classification and Regression by randomForest. R News 2, 18–22 (2002).

59. Dickson KS, Thompson SR, Gray NK & Wickens M Poly(A) polymerase and the regulation of cytoplasmic polyadenylation. J Biol Chem 276, 41810–41816, doi:10.1074/jbc.M103030200 (2001). [PubMed: 11551905]

60. Gray NK, Coller JM, Dickson KS & Wickens M Multiple portions of poly(A)-binding protein stimulate translation in vivo. EMBO J 19, 4723–4733, doi:10.1093/emboj/19.17.4723 (2000). [PubMed: 10970864]

61. Pisacane P & Halic M Tailing and degradation of Argonaute-bound small RNAs protect the genome from uncontrolled RNAi. Nat Commun 8, 15332, doi:10.1038/ncomms15332 (2017). [PubMed: 28541282]

62. Haracska L, Johnson RE, Prakash L & Prakash S Trf4 and Trf5 proteins of *Saccharomyces cerevisiae* exhibit poly(A) RNA polymerase activity but no DNA polymerase activity. Mol Cell Biol 25, 10183–10189, doi:10.1128/MCB.25.22.10183-10189.2005 (2005). [PubMed: 16260630]

63. Ruegger S, Miki TS, Hess D & Grosshans H The ribonucleotidyl transferase USIP-1 acts with SART3 to promote U6 snRNA recycling. Nucleic Acids Res 43, 3344–3357, doi:10.1093/nar/gkv196 (2015). [PubMed: 25753661]

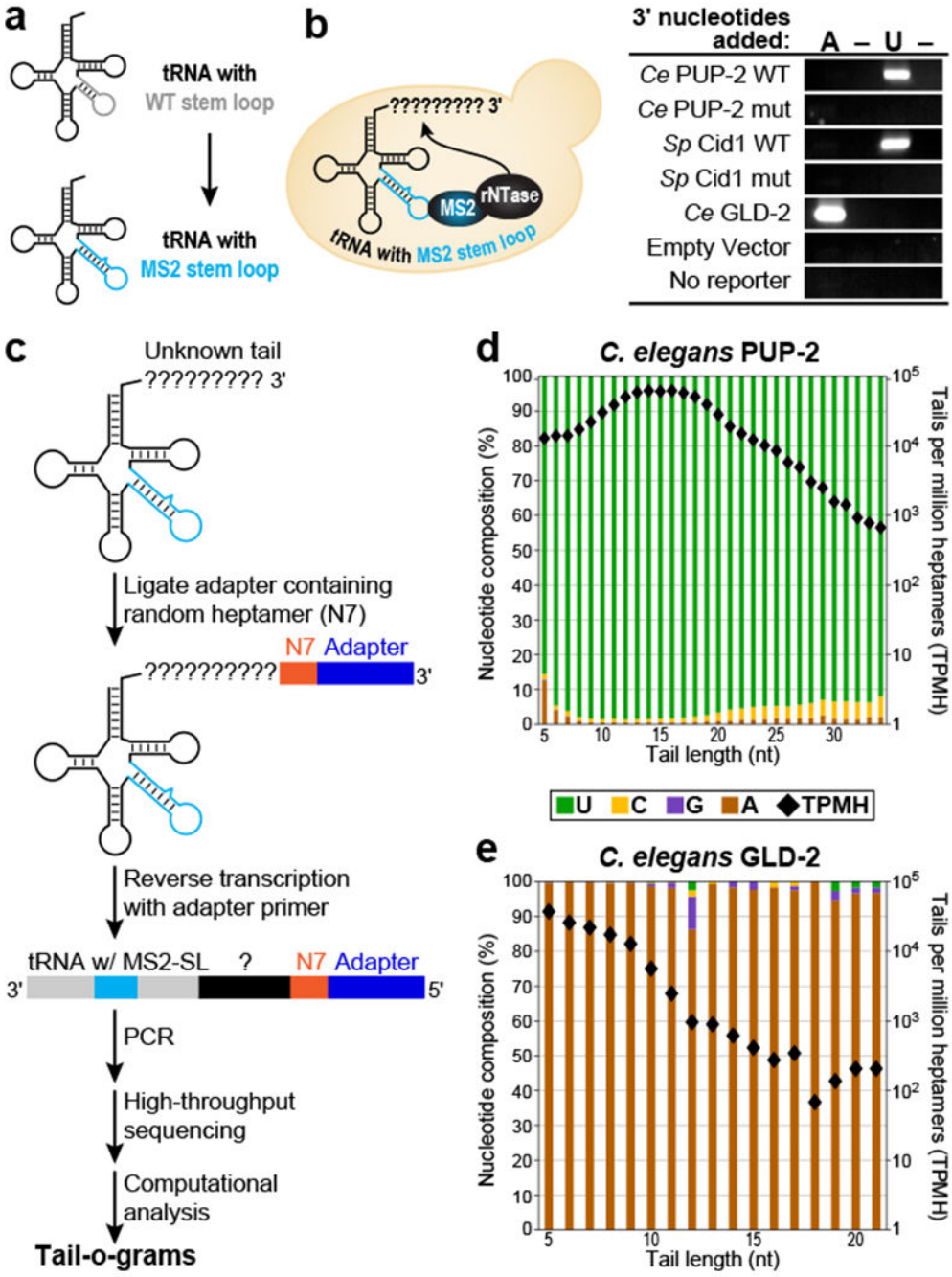


Figure 1. TRAID-Seq assay measures nucleotide addition activity *in vivo*.

(a, b) TRAID-Seq strategy. **(a)** $tRNA^{Ser(AGA)}$ variable arm (gray) is mutated to an MS2 stem loop (cyan) to form the tRNA reporter. **(b)** Left, tRNA reporter is co-expressed with an MS2 coat protein-rNTase fusion in *S. cerevisiae*. The tethered rNTase adds nucleotides to the 3′ end of the tRNA. Right, RT-PCR analysis to detect A tails or U tails added by control rNTases, relative to empty vector or a no-reporter control. Lanes marked with a dash indicate reactions performed without reverse transcriptase. Representative gel image from four independent experiments. **(c)** Schematic of sample processing. **(d, e)** Tail-o-grams of

nucleotides added by control rNTases, *C. elegans* PUP-2 (d) and *C. elegans* GLD-2 (e). Percent of each nucleotide at each tail length is color-coded and plotted on the left y-axis; U (green), C (yellow), G (purple), A (brown). Tails lengths of five nucleotides or greater are shown for clarity (see Online Methods). The number of tails detected per million heptamers (TPMH) are indicated by black diamonds and correspond to the log scale on the right y-axis.

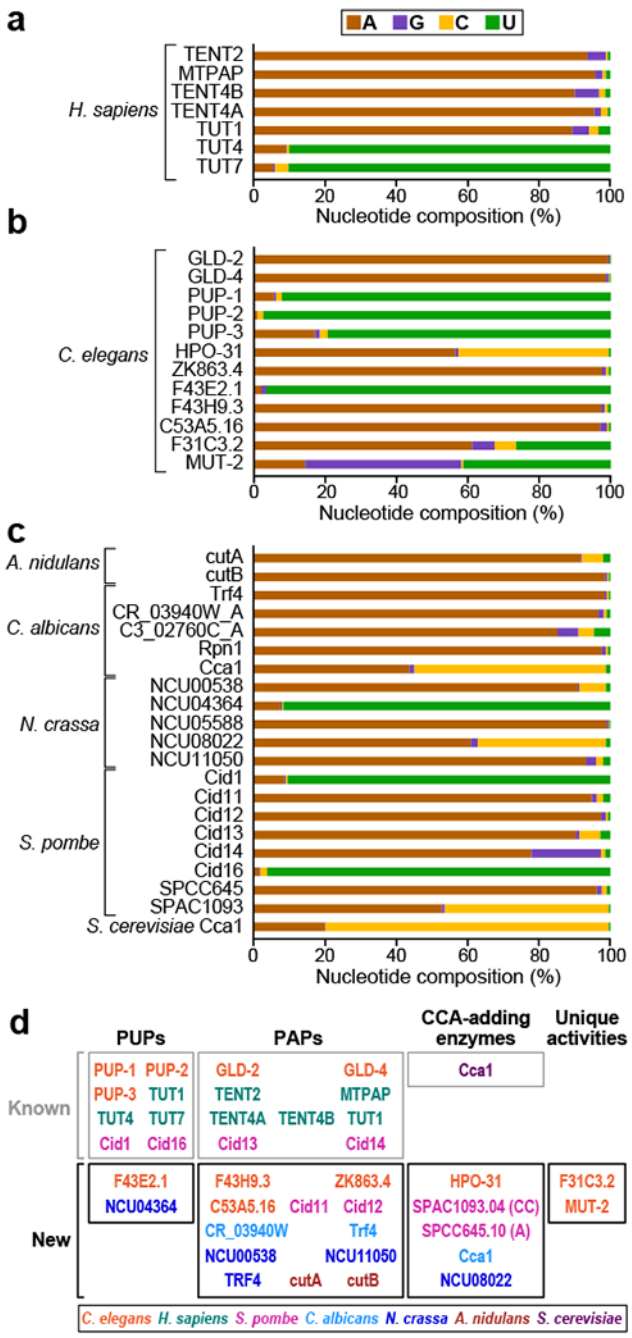
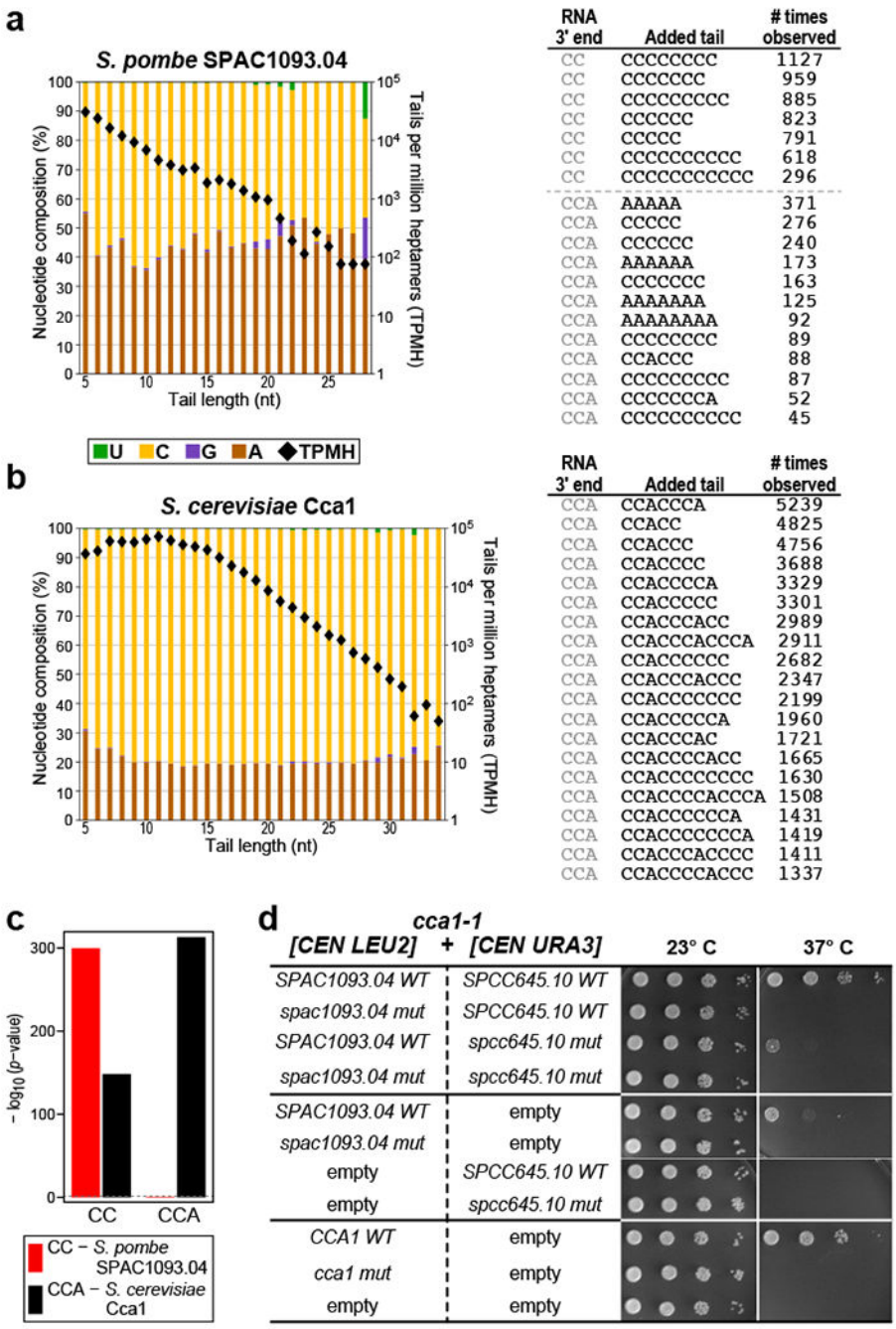


Figure 2. Analyses of nucleotide addition activities of 40 noncanonical rNTases from seven species.

Overall percentages of each nucleotide added by **(a)** *H. sapiens*, **(b)** *C. elegans*, and **(c)** fungal rNTases. **(d)** Categorization of rNTases as PUPs, PAPs, CCA-adding enzymes, or those with unique activities. rNTases are color-coded by organism. Gray boxes (top) indicate previously characterized (known) enzymes, and black boxes (bottom) indicate enzyme activities identified in this study (new).



RNA 3' end	Added tail	# times observed
CC	CCCCCCCC	1127
CC	CCCCCCC	959
CC	CCCCCCCCC	885
CC	CCCCCC	823
CC	CCCCC	791
CC	CCCCCCCCCC	618
CC	CCCCCCCCCCC	296
CCA	AAAAA	371
CCA	CCCCC	276
CCA	CCCCCC	240
CCA	AAAAAA	173
CCA	CCCCCCC	163
CCA	AAAAAAA	125
CCA	AAAAAAAA	92
CCA	CCCCCCCC	89
CCA	CCACCC	88
CCA	CCCCCCCCC	87
CCA	CCCCCCCA	52
CCA	CCCCCCCCCC	45

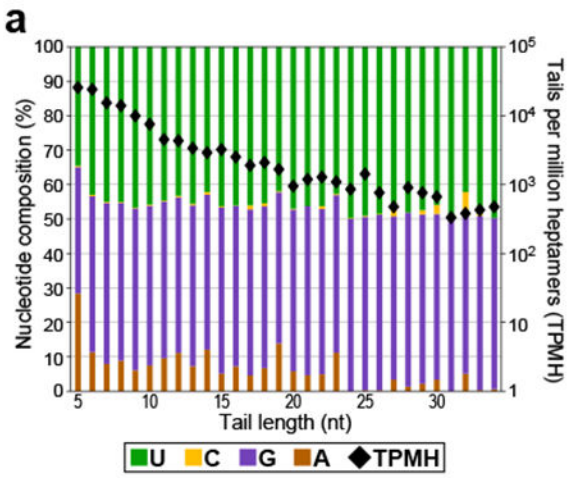
RNA 3' end	Added tail	# times observed
CCA	CCACCCA	5239
CCA	CCACC	4825
CCA	CCACCC	4756
CCA	CCACCCC	3688
CCA	CCACCCCA	3329
CCA	CCACCCCC	3301
CCA	CCACCCACC	2989
CCA	CCACCCACCCA	2911
CCA	CCACCCCCC	2682
CCA	CCACCCACCC	2347
CCA	CCACCCCCCC	2199
CCA	CCACCCCCA	1960
CCA	CCACCCAC	1721
CCA	CCACCCCACC	1665
CCA	CCACCCCCCCC	1630
CCA	CCACCCCACCCA	1508
CCA	CCACCCCCCA	1431
CCA	CCACCCCCCCA	1419
CCA	CCACCCACCCC	1411
CCA	CCACCCCACCC	1337

[CEN LEU2]	[CEN URA3]
SPAC1093.04 WT	SPCC645.10 WT
spac1093.04 mut	SPCC645.10 WT
SPAC1093.04 WT	spcc645.10 mut
spac1093.04 mut	spcc645.10 mut
SPAC1093.04 WT	empty
spac1093.04 mut	empty
empty	SPCC645.10 WT
empty	spcc645.10 mut
CCA1 WT	empty
cca1 mut	empty
empty	empty

Figure 3. Nucleotide addition activity of *S. pombe* SPAC1093.04 and *S. cerevisiae* Cca1.

(a) Left, tail-o-gram depicting nucleotide composition in each added tail length added by *S. pombe* SPAC1093.04 and number of tails normalized to unique heptamer sequences. Right, most abundant tail sequences added to tRNA reporter containing a 3′ CC, or 3′ CCA end. **(b)** Left, tail-o-gram depicting nucleotide composition in each added tail length added by *S. cerevisiae* Cca1 and number of tails normalized to unique heptamer sequences. Right, most abundant tail sequences added to tRNA reporter containing a 3′ CCA end. **(c)** Sequence motif effect analysis of tails added by *Sp*SPAC1093.04 (red, n=5) and *Sc*Cca1 (black, n=3).

Each adjusted *p*-value quantifies the significance of contribution of the indicated oligonucleotide to the variation in tail sequence read counts. Significances for dinucleotide (CC) and trinucleotides (CCA) after multiplicity correction with the Bonferroni procedure are shown. A dashed line indicates significance level 0.05. The $-\log_{10}$ *p*-values from left to right in the figure are 300, 148, 0.87, and 313. **(d)**. *cca1-1* mutant strains containing *CEN* plasmids expressing indicated plasmids were serially diluted, spotted on SD-Ura-Leu media and grown at 37°C for 3 days or 23°C for 4 days. This experiment was repeated twice with similar results.



b

MUT-2 replicate #1

Added tail	# times observed
UGUGUG	214
UGUGUGUG	180
UGUGU	165
GUGUG	151
GUGUGU	151
AAAAA	132
UGUGUGU	126
GUGUGUG	121
UGUGUGUGU	84
UGUGUGUGUG	84
GUGUGUGUG	69
AAAAAA	41
UGUGUGUGUGUG	41
GUGUGUGU	37
GUGUGUGUGUGUGUG	29
UGUGUGUGUGU	29
GUGUGA	28
GUGUGUGUGU	26
GUGUGUGUGUG	24
UGUGUGUGUGUGUG	23

MUT-2 replicate #2

Added tail	# times observed
GUGUGU	588
GUGUG	565
GUGUGUG	370
UGUGU	331
UGUGUG	302
UGUGUGUGUG	269
UGUGUGU	255
UGUGUGUG	245
GUGUGUGU	199
AAAAA	173
GUGUGUGUG	143
UGUGUGUGU	127
GUGUGA	101
UGUGUGUGUGUG	95
GUGUGAA	77
GUGUGUGUGUG	73
UGUGUGUGUGU	69
GUGUGUGUGU	64
GUGUGUGUGUGUG	61
AAAAAA	59

c

MUT-2 replicate #1

UGU
GUUGUGUGUGUUGUUU
UGUGUGUGUGUUGUGUGUGUGUGUGUGUGUGUG
UGUGUGUGUGUGUGUGUAAUUAG
AUGUGUGUGUGUGUGU
UGU
UGUGUGUGUUUGUG
UGUGUGUUUAUCUGUGUAUGUGUG
UGUGUGUGUUUUUGU
UGUGUGUGUGUGUGUGUGUGUGUGUGU
UGUGUGUGUGUGUGUG
UUGUGUGUG
GUGUGUGUUUUUGUGAUGU
UGUGUUUUGUGUGUGUGUG
GUGUGAUUUUUUGUGUGUG
GUGUGUGUGUGUGUUU

MUT-2 replicate #2

GU
UGUUUG
UGUUUUUUUUU
UGUGUGUGUGUGUG
UGUGUGUGUGUGUGUGUGUGUGU
UGUGUUGUGUGUGU
UGUA
GAUUUUGUUU
UGUGUGUGUGUGUGUGUGUGUGUGUGU
UUUUUGU
UGUGUGUGUGUGUGU

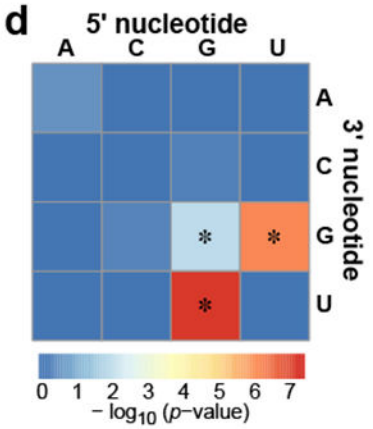


Figure 4. *Ce*MUT-2 is a poly(UG) polymerase.

(a) Tail-o-gram depicting *Ce*MUT-2 nucleotide addition activity in TRAID-Seq. **(b)** The most abundant tail sequences identified in two biological replicates of *Ce*MUT-2 TRAID-Seq assays. **(c)** UG tail sequences from two biological replicates of *Ce*MUT-2 activity in *X. laevis* oocytes. **(d)** Analysis of all possible dinucleotides in the tails added by *Ce*MUT-2 from 8 independent biological replicates. A heatmap of -$\log_{10}$ *p*-values for individual dinucleotides is shown. Each *p*-value quantifies the significance of adjusted contribution of each dinucleotide to the variation in tail sequence read counts. Dinucleotides with a significant effect after multiplicity correction at significance level 0.05 are marked with an asterisk (*). Details of statistical analyses performed are provided in the Online Methods (Computational Analyses of Sequence Motifs).

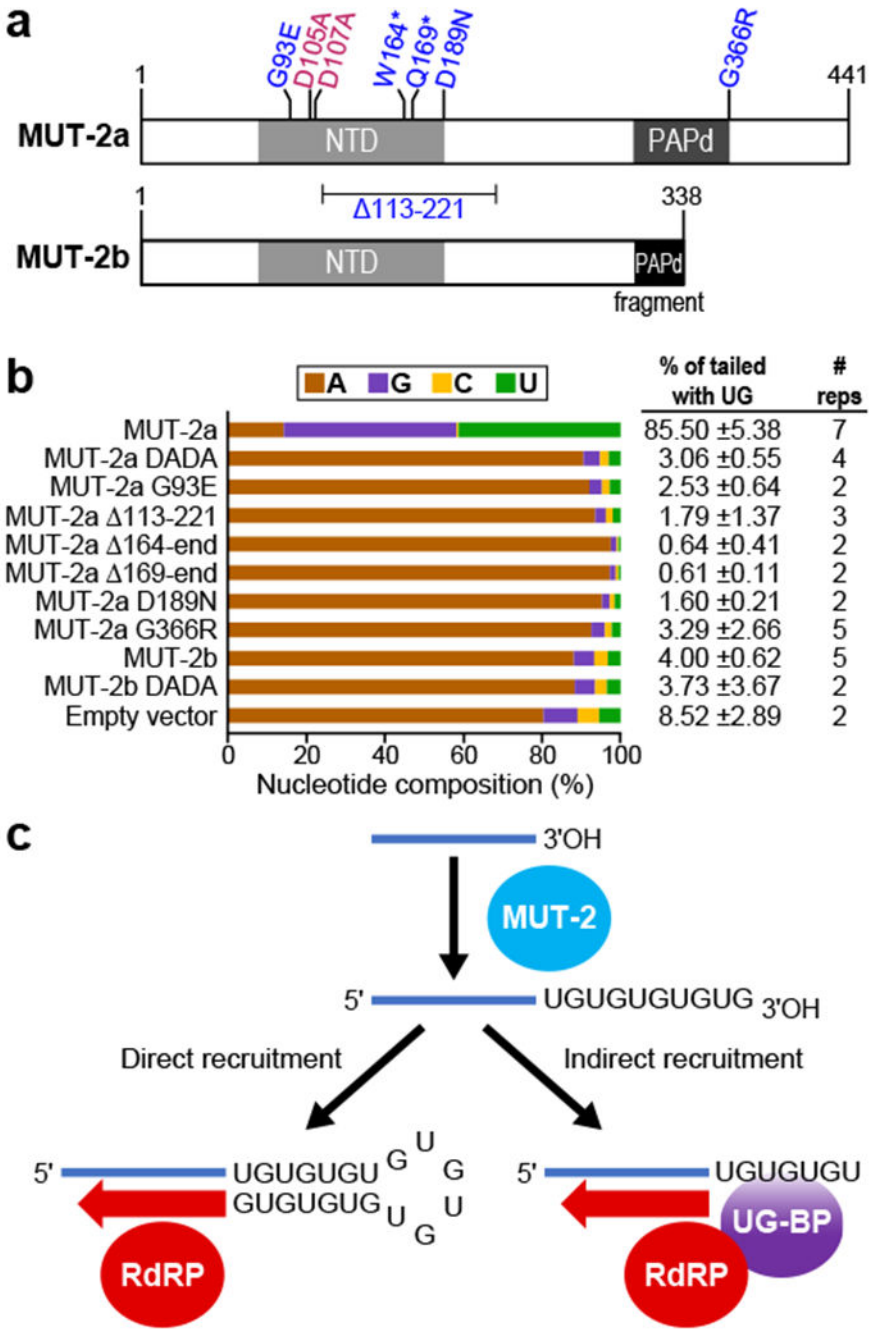


Figure 5. *Ce*MUT-2 mutants defective for RNAi lack poly(UG) polymerase activity.

(a) Schematic of *Ce*MUT-2 isoforms and tested mutations, known catalytic mutants (pink), and mutants identified in forward genetic screen[19] (blue). NTD, Nucleotidyl transferase domain; PAPd, Poly(A) polymerase-associated domain. * indicates that a truncated version of *Ce*MUT-2 was made to recapitulate this nonsense mutant. **(b)** Percent of nucleotides added by each *Ce*MUT-2 enzyme variant. Percent of tails containing UG repeats, standard deviation, and number of biological replicates are indicated. **(c)** Model depicting potential roles of poly(UG) tails in small RNA amplification in *C. elegans*. Poly(UG) tails could

directly recruit RNA-dependent RNA polymerase (RdRP) (left). Alternatively, poly(UG) tails could be identified by a poly(UG) binding protein (UG-BP), which then recruits RdRP (right). In both cases, the UG tails could be single-stranded or form a higher-order structure.

Table 1:

Summary of nucleotide addition preferences of tested rNTases and NRM analysis

rNTase	Species	Nucleotide Preference	Histidine in NRM
Cid1	*S. pombe*	U	Yes
PUP-1	*C. elegans*	U	Yes
PUP-2	*C. elegans*	U	Yes
TUT4	*H. sapiens*	U	Yes
TUT7	*H. sapiens*	U	Yes
NCU04364.7	*N. crassa*	U	Yes
Cid16	***S. pombe***	**U**	**No (Lys)**
PUP-3	***C. elegans***	**U**	**No (Arg)**
F43E2.1	***C. elegans***	**U**	**No (Arg)**
F31C3.2	***C. elegans***	**A, U, C, G (indiscriminate)**	**Yes**
MUT-2	***C. elegans***	**U, G**	**Yes**
TUT1 (Star-PAP)	***H. sapiens***	**A**	**Yes**
ZK863.4	***C. elegans***	**A**	**Yes**
cutB	*A. nidulans*	A	No (Asn)
Trf4	*S. cerevisiae*	A	No (Asn)
Trf5	*S. cerevisiae*	A	No (Asn)
Trf4	*C. albicans*	A	No (Asn)
CR_03940W_A	*C. albicans*	A	No (Asn)
GLD-2	*C. elegans*	A	No (Asn)
GLD-4	*C. elegans*	A	No (Asn)
TENT2 (GLD2)	*H. sapiens*	A	No (Asn)
TENT4B (PAPD5)	*H. sapiens*	A	No (Asn)
TENT4A	*H. sapiens*	A	No (Asn)
TRF4	*N. crassa*	A	No (Asn)
Cid12	*S. pombe*	A	No (Asn)
Cid14	*S. pombe*	A	No (Asn)
cutA	*A. nidulans*	A	No (Arg)
NCU00538.7	*N. crassa*	A	No (Arg)
Cid11	*S. pombe*	A	No (Arg)
Cid13	*S. pombe*	A	No (Arg)
MTPAP	*H. sapiens*	A	No (Leu)
NCU11050.7	*N. crassa*	A	No (Leu)
C53A5.16	*C. elegans*	A	No (Pro)
F43H9.3	*C. elegans*	A	No (not aligned)
RPN1	*C. albicans*	A	No (Glu)

"Histidine in NRM" indicates the amino acid in each rNTase that corresponds to histidine 336 of SpCid1. Bold text indicates PUPs that do not have histidine in the NRM, and rNTases that can add nucleotides other than uridine and have histidine in the NRM.