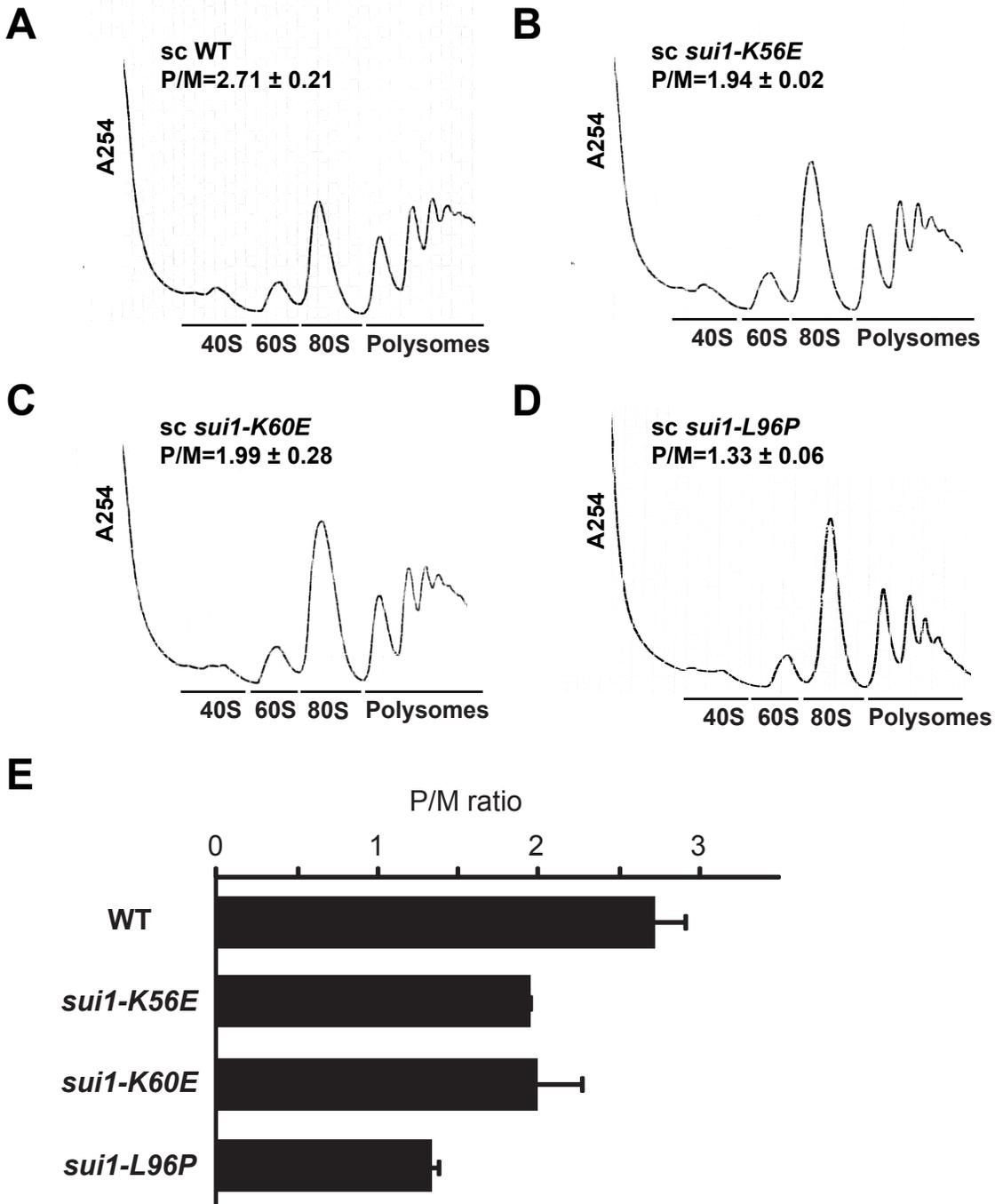


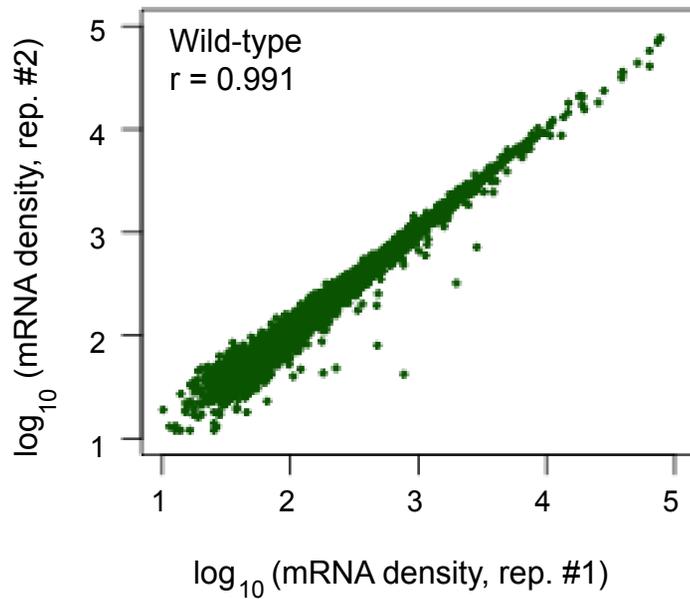


**Fig. S2**

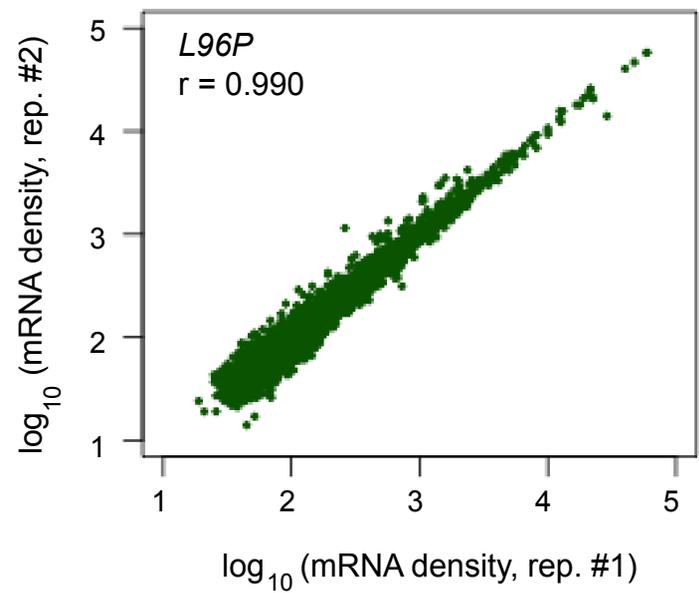


**Fig. S3**

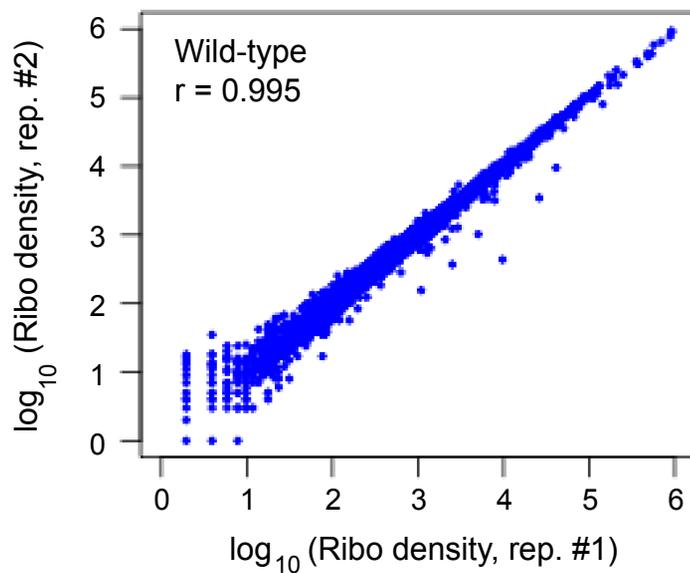
**A**



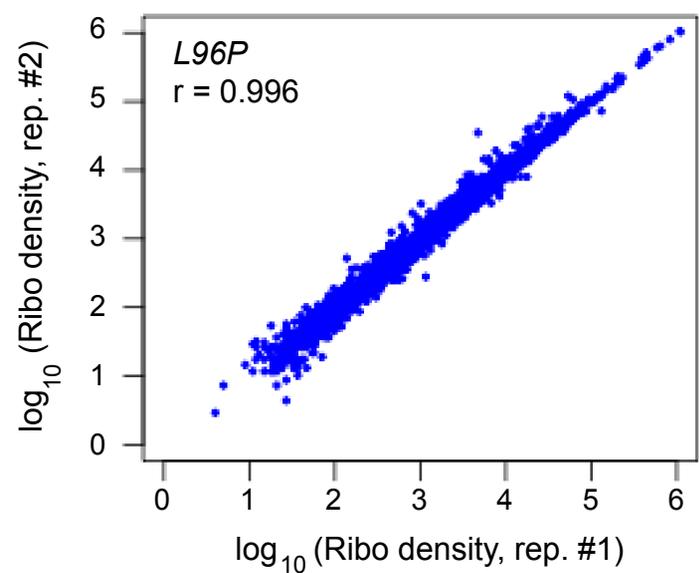
**B**

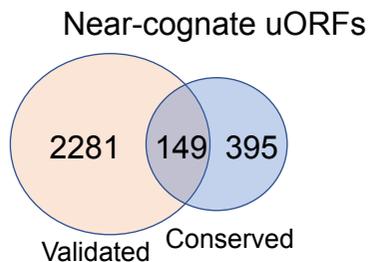
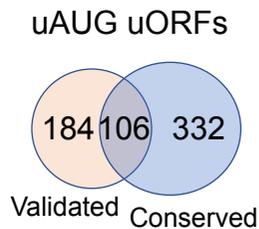
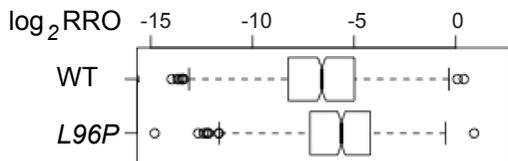
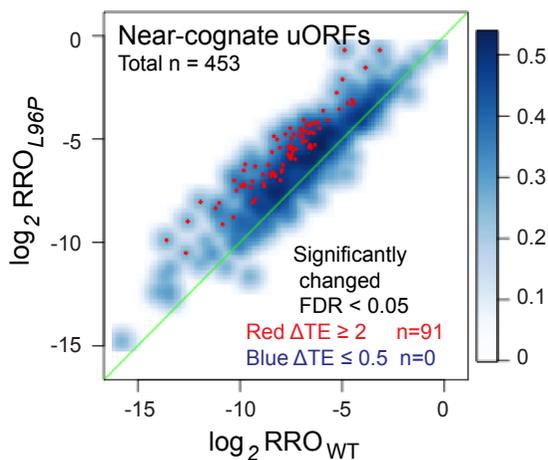
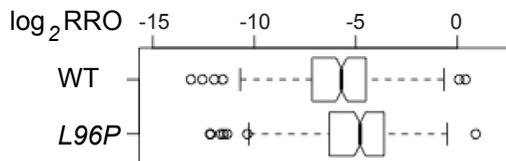
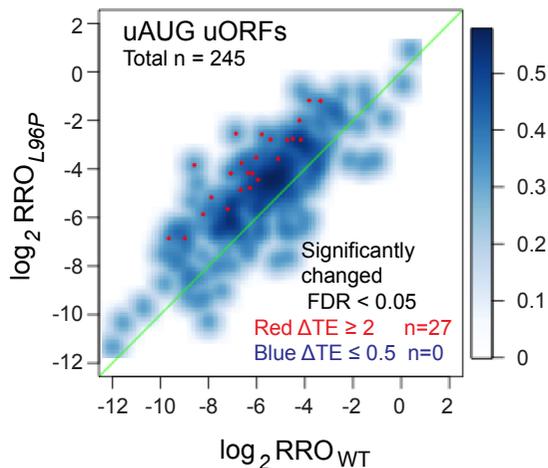


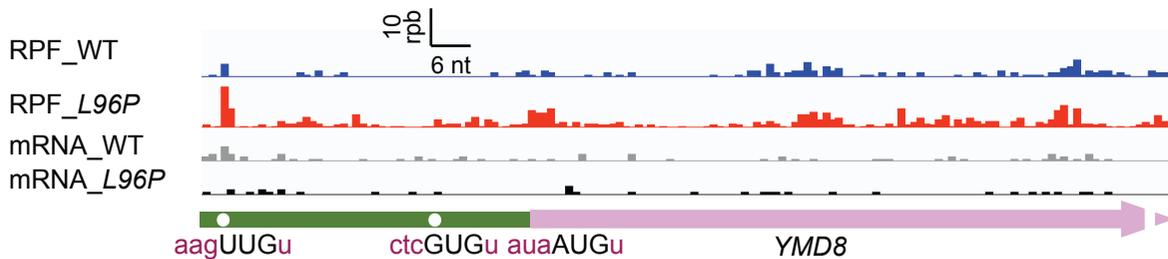
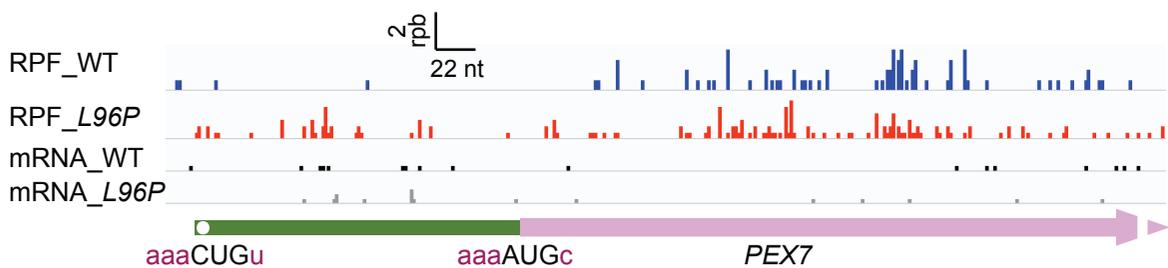
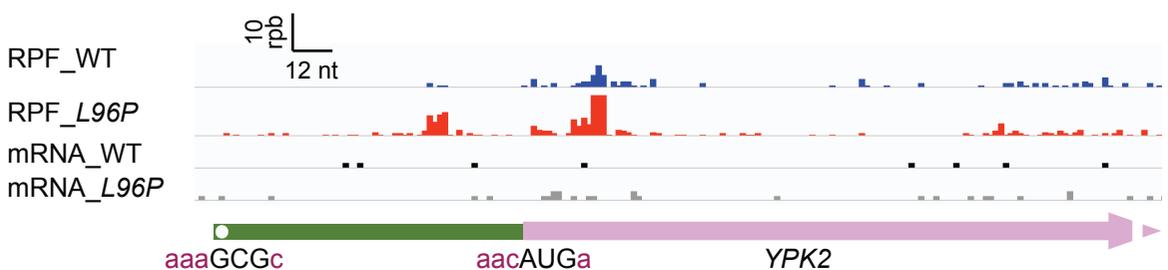
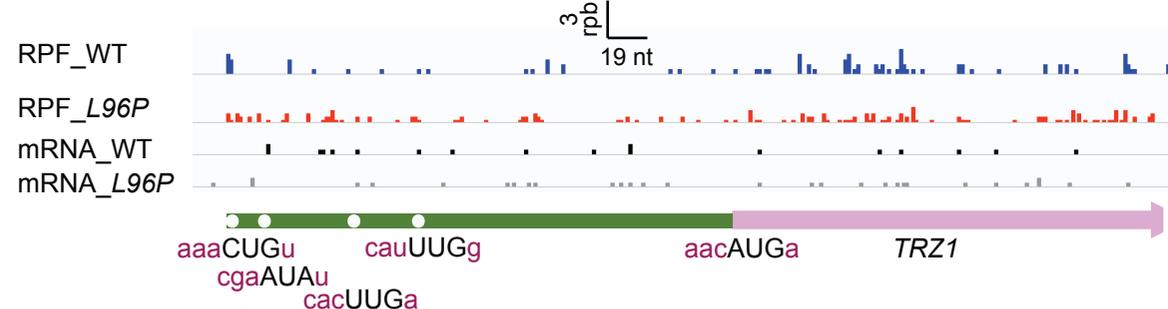
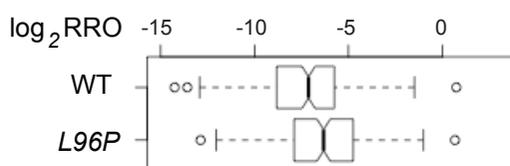
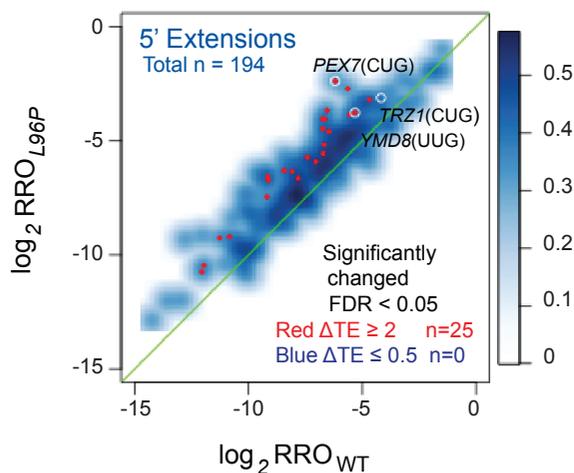
**C**

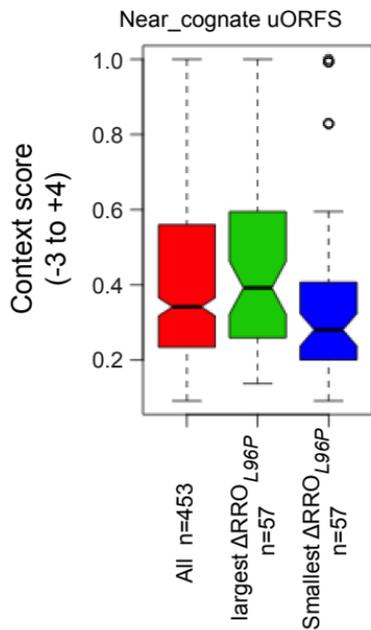
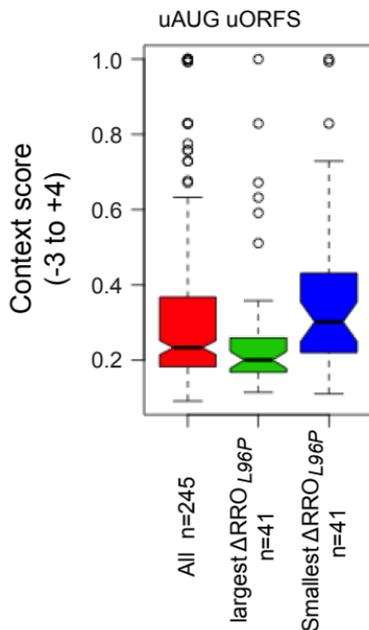
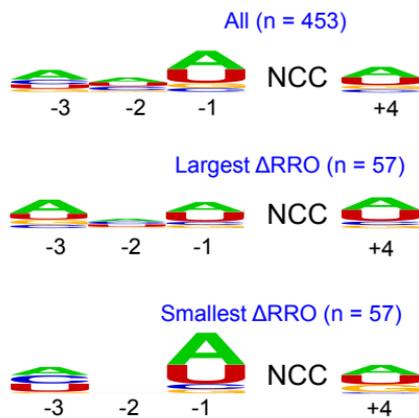
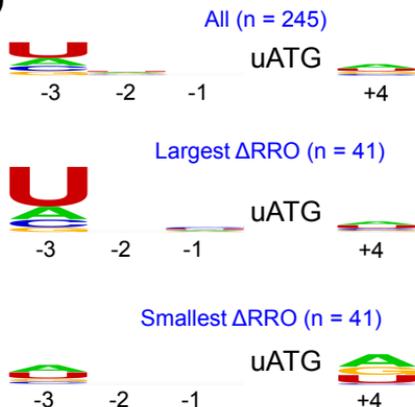


**D**



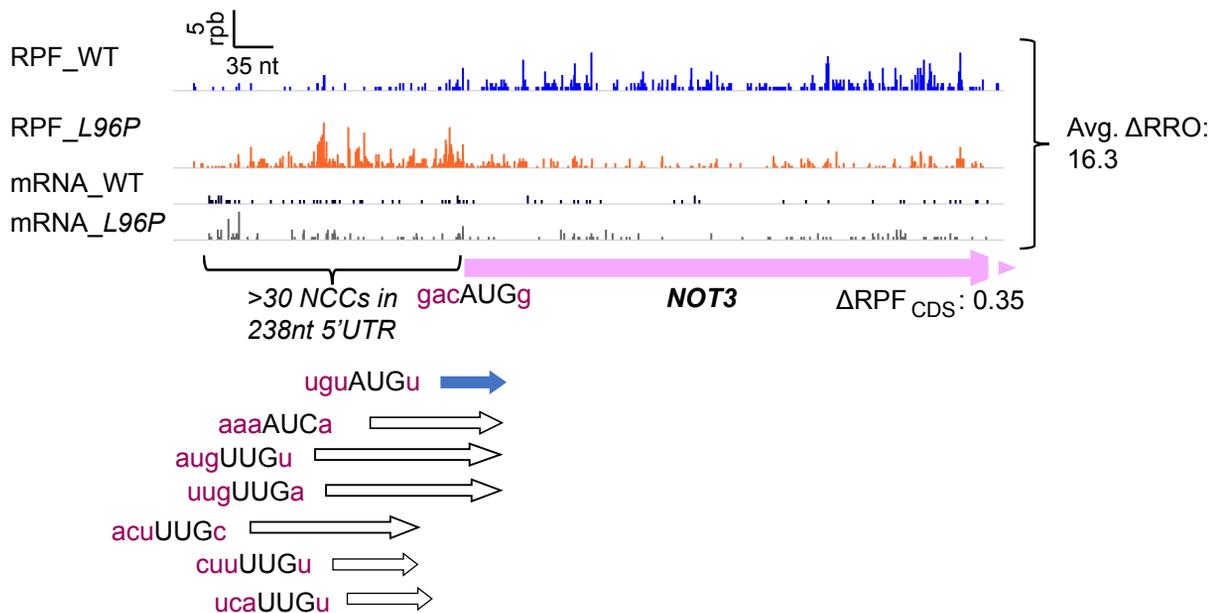
**Fig. S4****A****B****C****D**

**Fig. S5****A****B****C****D****E**

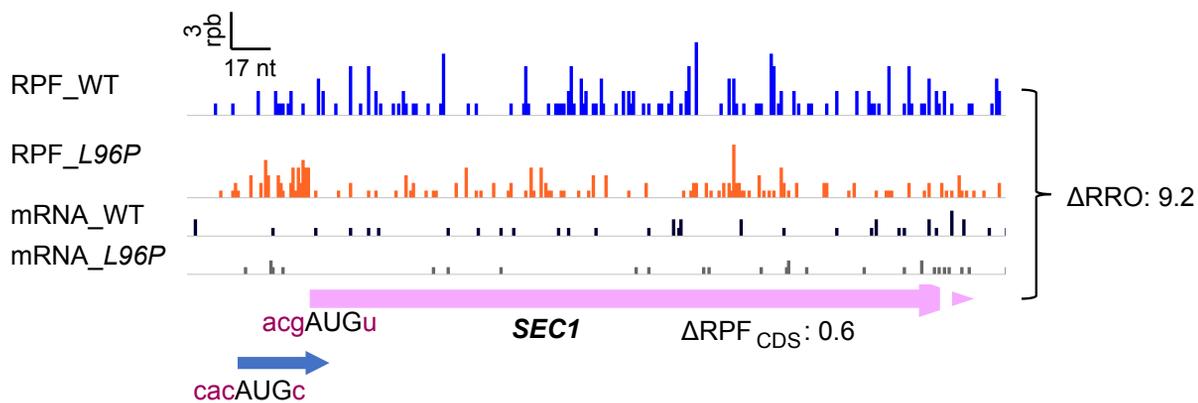
**Fig. S6****A****B****C****D**

# Fig. S7

## A



## B



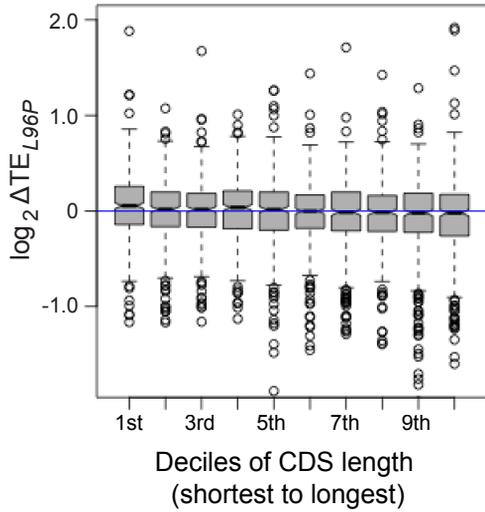
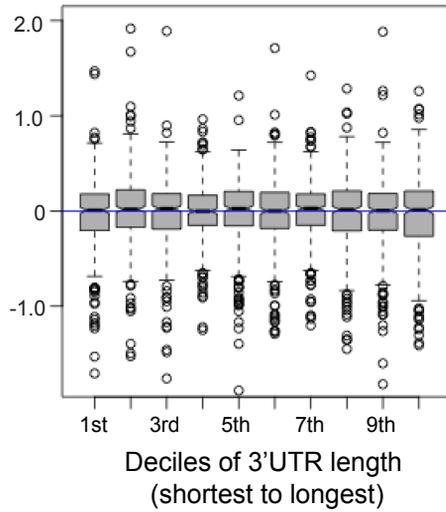
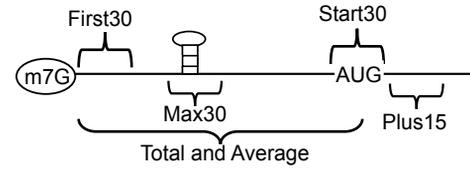
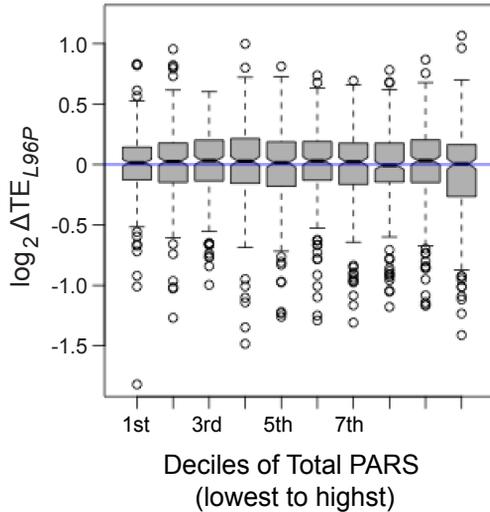
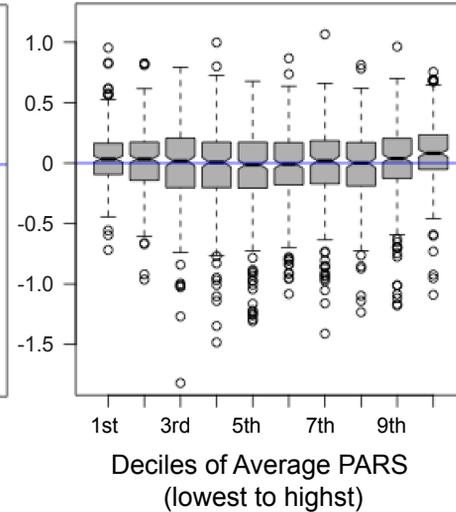
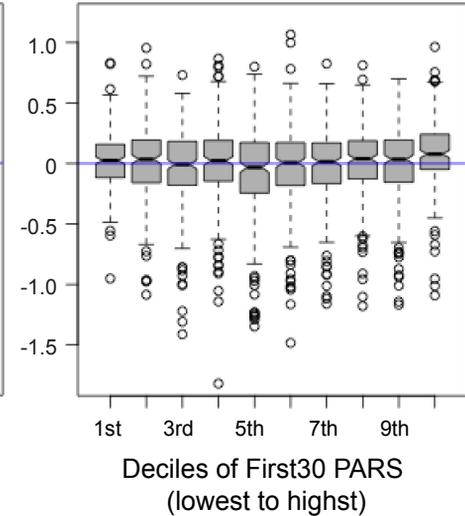
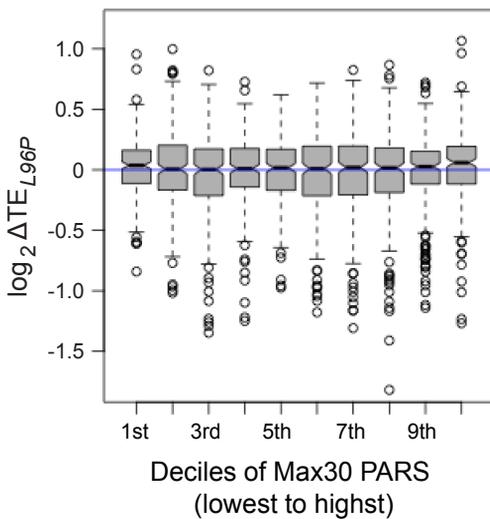
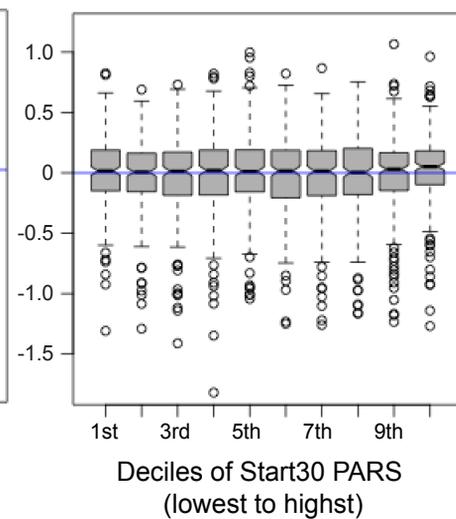
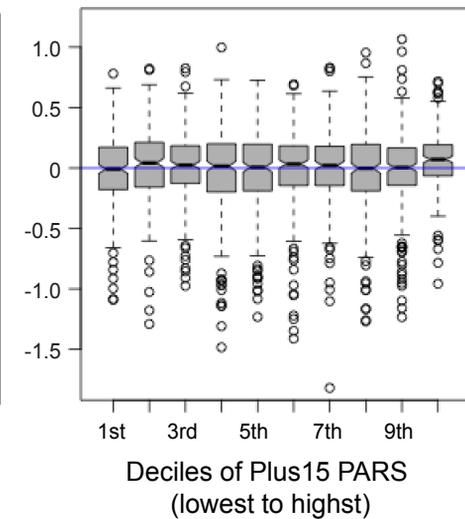
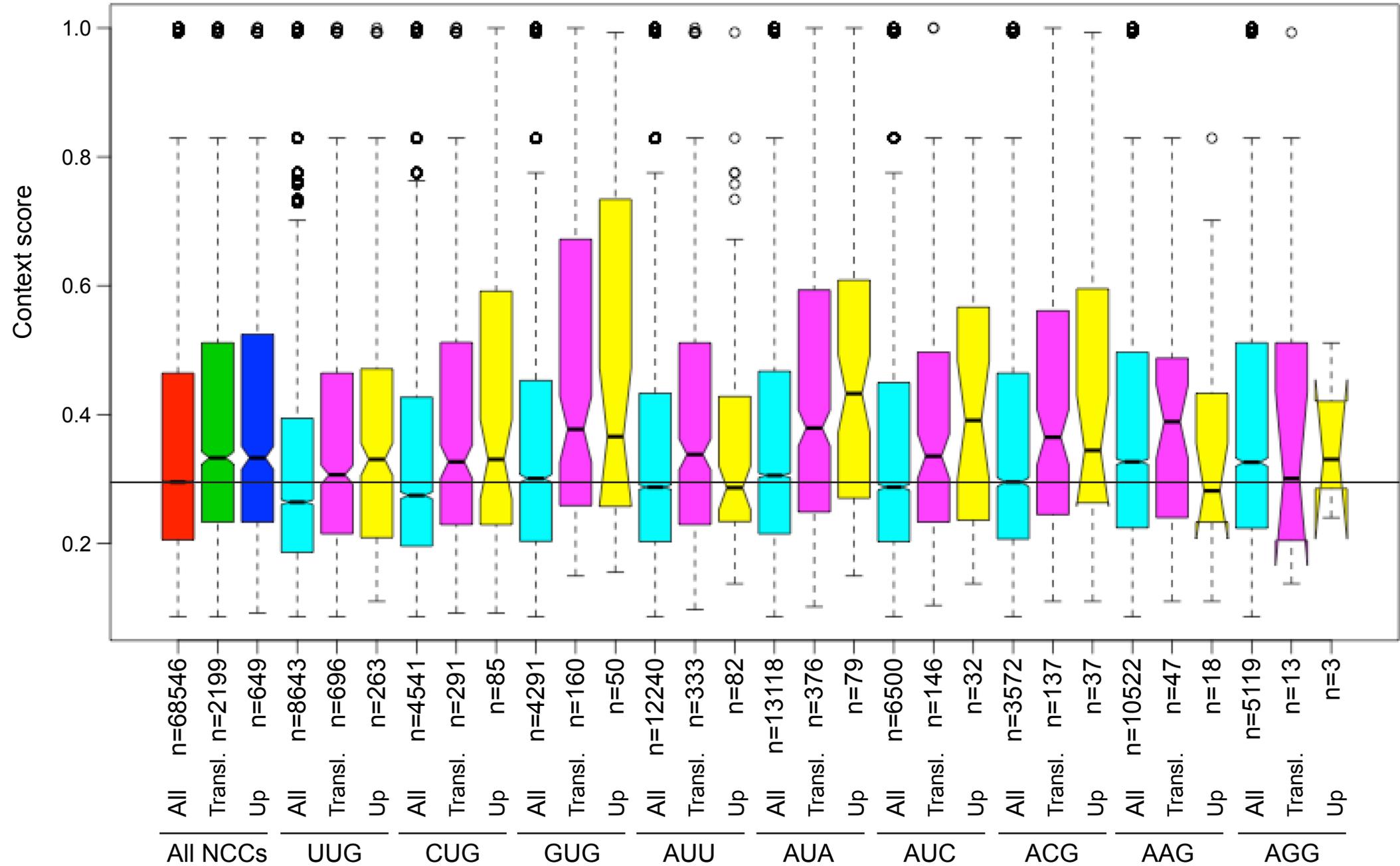
**Fig. S8****A****B****C****D****E****F****G****H****I**

Figure S9



## FIGURE LEGENDS FOR SUPPLEMENTAL FIGURES

**Figure S1. Model describing known conformational rearrangements of the PIC during scanning and start codon recognition.** (i) eIF1 and the scanning enhancers (SE) in the C-terminal tail (CTT) of eIF1A (blue ovals) stabilize an open conformation of the 40S subunit to which TC rapidly binds. (ii) The 43S PIC in the open conformation scans the mRNA for the start codon with Met-tRNA<sub>i</sub><sup>Met</sup> bound in the P<sub>OUT</sub> state. eIF2 can hydrolyze GTP to GDP•P<sub>i</sub>, but release of P<sub>i</sub> is blocked by eIF1. The N-terminal tail (NTT) of eIF1A interacts with the eIF5-CTD. (iii) On AUG recognition, Met-tRNA<sub>i</sub><sup>Met</sup> moves from the P<sub>OUT</sub> to P<sub>IN</sub> state, clashing with eIF1 and the CTT of eIF1A, provoking displacement of the eIF1A CTT from the P site, dissociation of eIF1 from the 40S subunit, and P<sub>i</sub> release from eIF2. The NTT of eIF2β interacts with the eIF5-CTD, and the eIF1A-NTT, harboring scanning inhibitor (SI) elements, interacts with the codon:anticodon helix. (Above) Arrows summarize that eIF1 and the eIF1A SE elements promote P<sub>OUT</sub> and impede transition to P<sub>IN</sub> state, whereas the eIF1A SI element in the NTT stabilizes the P<sub>IN</sub> state. (Adapted from (Hinnebusch 2014)).

**Figure S2. *SUII* Sui<sup>r</sup> mutations reduce polysome assembly. (A-E)** Polysome profiles of *sui1Δ* strains harboring the indicated *sui1* alleles on plasmids. The indicated strains were cultured in synthetic complete medium lacking leucine (SC-Leu) at 30°C to A<sub>600</sub> of ≈ 1, with cycloheximide added 10 min prior to harvesting. WCEs were separated by sucrose density gradient centrifugation and scanned at 254 nm. Mean polysome:monosome ratios (P/M) ± S.E.M.s were determined from two biological replicates, as summarized in the histogram in (E).

**Figure S3. Genome-wide ribosome footprint and mRNA reads for WT and *sui1-L96P* strains. (A–D)** Scatterplots of mRNA density (total mRNA reads normalized to mRNA length (A–B) and RPF density (total RPF reads normalized to CDS length) (C–D) from WT and *sui1-L96P* strains, for all genes with ≥128

total mRNA reads or  $\geq 40$  RPF reads in the four samples combined (two strains and their two replicates). Pearson correlation coefficients ( $r$ ) were calculated for each plot.

**Figure S4. *sui1-L96P* increases the utilization of NCC and uAUG start codons for evolutionarily conserved uORFs. (A-B).** Overlaps between the sets of translated uORFs analyzed in Fig. 1D-F identified previously (Validated) (Martin-Marcos et al. 2017) and those described by Spealman et al. showing evolutionary conservation (Conserved) (Spealman et al. 2018) **(C-D).** Scatter plots of  $\log_2$  RRO values in *L96P* versus WT cells for the conserved 453 NCC-uORFs (C) and 245 uAUG\_ uORFs (D) described in panels A-B from genes with  $\geq 32$  RPF reads in the main CDS and  $\geq 2$  RPF reads in the 5'UTR or uORF in the averaged reads from 4 samples, two replicates of WT and two replicates of *sui1-L96P*). uORFs with significantly changed TE values in *L96P* vs. WT cells of  $\geq 2$ -fold (at FDR  $< 0.05$ ) are indicated in red or dark blue. Notched box-plots below each scatterplot compare the distributions of  $\log_2$  RRO values in WT vs *L96P* cells for the complete sets of NCC\_uORFs (C) or AUG\_uORFs (D).

**Figure S5. *sui1-L96P* increases the utilization of NCC and uAUG start codons for synthesizing proteins with N-terminal extended isoforms. (A-D)** Ribosome protected fragments (RPFs) and mRNA reads on the *YMD8*, *PEX7*, *YPK2* and *TRZ1* genes in WT and *L96P* cells in units of rpb (reads per billion mapped reads), showing schematically the position of the main CDS (pink), the 5' N-terminal extensions (green), and the -3 to -1 and +4 context nucleotides surrounding the near-cognate codon (NCC) (in brick red). *YMD8*, *PEX7* and *TRZ1* are among the 2226 NCC uORFs we described in Figure 1E. **(E)** Scatter plots comparing  $\log_2$  RRO (relative ribosome occupancy) values in *L96P* versus WT cells for 194 from genes with  $\geq 32$  RPF reads in the main CDS and  $\geq 2$  RPF reads in the 5' extension in the averaged reads from 4 samples, two replicates of WT and two replicates of *sui1-L96P*. 5' extensions showing significantly changed RRO values in *L96P* vs. WT cells of  $\geq 2$ -fold (at FDR  $< 0.05$ ) are indicated in red or dark blue, and the numbers (n) of NTEs plotted are indicated. The expression of NTEs for *YMD8* (UUG), *PEX7* (CUG) and *TRZ1* (CUG) are denoted by white circles. (Note that the increased translation of the NTE of *TRZ1* in

*L96P* cells did not satisfy the criteria of  $\Delta TE \geq 2$  at FDR  $< 0.05$  for highlighting its dot in red in Fig. S5E, and that the NTE of *YPK2* was not included in our list of 2226 NCC uORFs identified for analysis (described in Fig. 1E) and, thus, is not represented as a dot in Fig. S5E.) Notched box-plots below each scatterplot compare the distributions of  $\log_2$  RRO values in WT vs *L96P* cells for the complete sets of expressed 5' extensions. The 194 expressed 5' extensions are among 2226 NCC-uORFs, and 205 uAUG uORFs described in Figure 1E-F.

**Figure S6. *sui1-L96P* has opposing effects on NCC and uAUG start codons in discriminating against poor Kozak context.** (A) Box plot analysis of the distribution and median context scores of a subset of the conserved NCC uORFs described in Fig. 1-S4A that were translated in our strains (453 NCC\_uORFs), and for the sets of 57 uORFs comprising the top (Largest  $\Delta RRO_{L96P}$ ) or bottom (Smallest  $\Delta RRO_{L96P}$ ) 1/8ths of these 453 NCC\_uORFs ranked by  $\Delta RRO_{L96P}$  values from largest to smallest. A fraction of outliers were omitted from the plots to expand the y-axis scale. Note that the group representing 1/8<sup>th</sup> of all such uORFs with the largest increases in  $RRO_{L96P}$  values depicted in col. 2 exhibit significantly lower context scores compared to the 1/8<sup>th</sup> with the smallest increases in  $RRO_{L96P}$  values shown in col. 3; and that the median context scores for these two groups of 0.20 and 0.32, respectively, are very similar to the values determined for the corresponding quartiles of AUG\_uORFs analyzed in Fig. 2C. (B) Same analysis as in (A) but for the subset of conserved AUG\_uORFs described in Fig. 1-S4B translated in our strains (245 uORFs), and divided into 6 equal groups (C-D) Sequence logos for the -3 to -1 and +4 positions for the same groups of NCC\_uORFs and AUG\_uORFs described in panels A & B, respectively. Considering the AUG\_uORFs analyzed in (D), note that thymidine is the most prevalent base encoding the -3 position for the quartile of AUG\_uORFs showing the largest increases in  $RRO_{L96P}$  values (row 2); whereas a small bias for the preferred A at -3 was observed for the quartile with the smallest increases in  $RRO_{L96P}$  values (row 3), in keeping with results shown for AUG\_uORFs in Fig. 2D.

**Figure S7. Examples of mRNAs with elevated 5'UTR ribosome occupancies coupled with reduced RPFs in downstream CDS. (A-B)** RPFs and mRNA reads for the *NOT3* (A) and *SEC1* genes (B), depicted as in Fig. 1A, except that  $\Delta$ RRRO values in *L96P* vs. WT cells are indicated separately for an AUG\_uORF and a collection of NCC\_uORFs in the *NOT3* 5'UTR in (A).

**Figure S8. The translation efficiency change in *sui1-L96P* was not affected by the lengths of CDS and 3'UTR, or the structures in the 5'UTR or around main AUGs. (A-B)** Box-plot analyses of CDS  $\Delta$ TE<sub>*L96P*</sub> values for 4242 mRNAs (A) and 3956 mRNAs (B) divided into deciles of equal size binned according to CDS and 3'UTR lengths, respectively. (C) Schematic showing mRNA 5' UTR and CDS intervals assigned for calculating cumulative PARS scores for each gene. (D-I) Box-plot analysis of CDS  $\Delta$ TE<sub>*L96P*</sub> values for 2497 mRNAs (panels D, E, F, & I), 2361 mRNAs (H), or 2289 mRNAs (G) divided into deciles of equal size binned according to the indicated categories of cumulative PARS scores, as follows. As shown schematically in panel C, for each transcript we tabulated the total PARS score over the entire 5'UTR (Total PARS), the average PARS score per nucleotide for the entire 5'UTR (Average PARS), and the sum of PARS scores for the following intervals: (i) the first 30nt from the mRNA 5'end (First30 PARS), (ii) the 30nt window with the highest cumulative PARS scores (Max30 PARS), (iii) the 30nt surrounding the start codon (Start 30 PARS, excluding genes with a 5' UTR of <16nt), and (iv) the 30nt centered on the +15 nucleotides downstream of the AUG (Plus15 PARS). For each of these PARS features, the 2679 annotated transcripts were assigned to deciles of increasing PARS score and each decile was interrogated for  $\Delta$ TE<sub>*L96P*</sub> values. Examining the results in panels D-I revealed that the only significant differences were observed for decile 10 (containing the highest PARS scores), which showed an increase in TE conferred by *L96P* for all PARS features except the Total PARS score. While it is not obvious why the 10% of all mRNAs with the most highly structured 5'UTRs would exhibit a moderate TE increase in *L96P* cells, it is noteworthy that this group tends to show the greatest reductions in TE in response to mutations in Ded1, eIF4B, and eIF4A

(Sen et al. 2015; Sen et al. 2016). These findings indicate that eIF1 generally does not cooperate with Ded1, eIF4B, and eIF4A to overcome structural impediments to PIC attachment or scanning.

**Figure S9. Context scores for groups of NCC uORF start codons.** The Kozak context scores for each of the 9 NCCs located  $\geq 3$  nt from the 5' end of the mRNA were calculated, including 68546 of 74353 NCC uORFs encoded in the genome, 2199 of the 2226 translated NCC\_uORFs (shown in Fig. 5B) and 649 of the 659 NCC\_uORFs whose translation was upregulated in the *L96P* mutant (Fig. 5A). The resulting scores are presented in a notched box-plot for All NCC\_uORFs (cols. 1-3) and for each of the 9 types of NCC\_uORFs (remaining columns), with results for all encoded uORFs (All), all translated uORFs (Transl.), and uORFs translationally upregulated in *L96P* cells (Up), presented in successive columns in shades of red, green and blue (for All NCC\_uORFs) or cyan, magenta and yellow (for each individual NCC\_uORF), respectively.