SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Strains, Plasmids, and Oligonucleotides

The yeast strain BY4741 (MATa *his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*) was the parent strain for all linear transformations, except for constructs with the *Renilla* luciferase-codon insert-GFP fusion reporter where the parent strain was AW765 (BY4741, *nmd2Δ::kanMX*)(Wolf and Grayhack, 2015). Plasmids used in this study are reported in Supplementary Table S7. The plasmid vector pEAW315, in which P_{GAL1} controls transcription of a *GLN4* NTD-GFP fusion, is a variation on plasmids previously described (Wolf and Grayhack, 2015). The $GLN4_{(1\text{-}99)}$ fragment was obtained by PCR amplification of pJE1012a (Grant et al., 2012) using oligonucleotides OW377 and OW378, and was then inserted into the Pac1 site of pEKD1024 (Dean and Grayhack, 2012) using LIC cloning, which regenerated the Pac1 site as well as a LIC site. Thus, all insertions of codon sequences into the *GLN4*-GFP fusion protein were performed via LIC cloning of two annealed oligos into pEAW315. Oligonucleotides employed in this study are reported in Supplementary Table S6.

Library Construction and Fluorescence-Activated Cell Sorting

The RNA-ID GFP construct is a fusion protein encoding a site for 3C protease, an HA epitope, and His6, followed by superfolder GFP. The $(NNN)_3$ and $(VNN)_3$ libraries of GFP variants in *E. coli* were derived from those made by Dean and Grayhack (Dean and Grayhack, 2012). DNA was obtained from *E. coli* libraries (1 ml frozen aliquots) grown at 37 °C to saturation in 100 ml LB+amp, using a QiaFilter kit (Qiagen). For the $(NNN)_3$ and $(VNN)_3$ libraries, 1900 ng and 1950 ng of StuI cut and gel-purified linear DNA were transformed into 10 ml BY4741 yeast cells, which were plated on four plates of selective media (SD-met) to obtain 73,068 [$(NNN)_3$ Library 1], 50,532 [$(NNN)_3$ Library 2] and 77,200 [$(VNN)_3$ Library] yeast transformants, as described (Guy et al., 2014). For each transformation, transformants from four plates were pooled by scraping the plates into YP + 2% raffinose + 8% DMSO and saved at -80 °C. Aliquots were grown for 3.5 generations in selective media (S-met + 2% raffinose) at 30 °C. Then cells were diluted into 25 ml YP + 80 mg/L Ade + 2% raffinose + 2% galactose media at a starting OD_{600} of 0.08 and grown for 3.5 generations at 30 °C, diluted into another 50 ml at a starting OD_{600} of 0.05, grown overnight for 5 generations, and diluted into 10 ml at a starting OD_{600} of 0.3, followed by another 4h of growth at 30 °C. Fluorescence-activated cell sorting (FACS) of ~3 million to ~9.5 million cells was performed as previously described (Dean and Grayhack, 2012).

Quality Filtering of GFP Library Sequences

We used PRINSEQ (Schmieder and Edwards, 2011) to trim reads and require that each of the 9 variable base calls had a quality score of at least Q30. We also applied a read depth cutoff based on the maximum number of possible variants in each bin (the total number of cells sorted).

To ensure a dataset of only the highest-quality variant expression scores, we applied a minimum threshold for total number of variant read counts across FACS bin samples. For the $(NNN)_3$ libraries, we determined these thresholds empirically, based on the drop in stop codon-containing variants with a high proportion of spurious, high-expression bin counts. This threshold was 30 read counts for $(NNN)_3$ Library 1 variants, and 60 read counts for $(NNN)_3$ Library 2 variants. We then used stop codon-containing variants with reads above these thresholds to estimate the average degree of spread into distant fluorescent bins. We removed variants with bimodal-like distributions, where the variant had more than an average spread in both the background (no expression) and the high expression bin. These bimodal-like variants constituted 4% to 5% of each library. For the $(VNN)_3$ Library, we imposed the stricter of each $(NNN)_3$ Library's threshold values. Furthermore, we removed any $(VNN)_3$ Library variants with a "T" base call in the first nucleotide position of a codon. Variants with ≥ 75% of reads in the background bin were removed from all libraries.

GFP^{SEQ} and syn-GFP^{SEQ} Expression Scores

We used the median GFP fluorescence value of each FACS bin and the proportion of variant read counts in each bin to calculate a mean expression score (GFP^{SEQ}) for each 9-base sequence variant, relative to 100% high bin expression. After checking for a high degree of correlation between libraries (Figure S1A), we combined the variant data from each library. For identical sequences measured in separate FACS libraries, we treated each library's measurement as a biological replicate and took the average. To obtain syn-GFP^{SEQ} scores we normalized to the highest GFP^{SEQ} score among a set of synonymous variants. If there were no

synonymous variants, or if the highest GFP^{SEQ} fell below the mean of all scores (0.9547), then these sequences were not included in the downstream analysis.

Heatmap Generation

The heatmap shows permutation p-values for enrichment of 6-mers in low expression variants at each of the four possible 6-mer positions in the 9-base variable region. It includes all 6-mers that have a permutation p-value ≤ 0.001 at one or more of the four positions and numeric p-values at all four positions. Fifty-seven significant sequences with missing data at one or more of the positions are not plotted (Table S3). Most of these 6-mers (79%) had a 3-nucleotide stop codon sequence in one of the reading frames. We clustered and plotted the data using pheatmap package in R.

RNA Structure Prediction and Reduced Structure Subset

To assess RNA structure across all 35,811 sequence variants of our library, we found the global free energy (ΔG) for each variant and compared local, paired nucleotide probabilities between synonymous sequences in a manner similar to Goodman et al. (Goodman et al., 2013). We used NUPACK pairs to make ΔG and paired nucleotide predictions (Zadeh et al., 2011). We ran these calculations at 30 ºC for a 175 base window, from the transcription start site (-74 from the start codon) through the 101st base in the open reading frame. NUPACK computes the pair probability between each nucleotide pair combination across the length of the sequence window. Then it estimates the probability that a given position will be unpaired. Taking NUPACK's unpaired estimate for each position in a variant, we calculated the mean probability that a position would be unpaired within sliding windows of 10 bases and subtracted from 1 to arrive at mean paired nucleotide probability for each window.

We then identified windows where the mean paired nucleotide probabilities differed between synonymous sequences. To identify these windows, for each window position and each variant, we took the ratio of probabilities between the window in the variant sequence and the same window position in a synonymous reference sequence. This ratio yielded the relative paired nucleotide probability for each window. To evaluate whether structure may contribute to synonymous expression differences, for each window position we found the Spearman rank correlation between relative paired nucleotide probabilities and syn-GFP^{SEQ} (Figure S1C).

To identify a subset of variants with similar degrees of structure, we first identified all window positions, after the translational start, where relative paired nucleotide probabilities had a significant negative correlation with expression (p-value < 0.001; positions: 1-10, 17-35, 51-54, 57-59, 78-87, and 94-96). Then we removed variants with probabilities more than 1 standard deviation away from the high category mean. We also removed variants for which the global free energy (as measured by NUPACK), for the region running from the transcription start site to +102, fell more than 1 standard deviation away from the high variant category mean. This reduced structure subset had 13,061 variants in total, with 183 low variants, 1,133 intermediate variants, and 6,597 non-reference high variants.

Analysis of Individual Variants by Flow Cytometry

Variant sequences in GFP (at amino acids 6 to 8), in *GLN4*$_{(1-99)}$-GFP (beginning at amino acid 100), and in *Renilla* luciferase-GFP (beginning at amino acid 318) were inserted using LIC cloning as previously described (Dean and Grayhack, 2012; Wolf and Grayhack, 2015). Synonymous optimal codons were chosen based on CAI. If two synonymous codons had a similar CAI, the more A-U rich codon was chosen. Strains to be analyzed by flow cytometry were grown overnight in YP + 80 mg/L Ade + 2% raffinose + 2% galactose media at 30 °C, followed by dilution to OD_{600} between 0.1-0.2 and grown for 4-6 hours in the same media to OD_{600} of ~0.8. Analytical flow cytometry and downstream analysis were performed for 4 independent isolates of each strain (sometimes 3) as previously described (Dean and Grayhack, 2012).

To examine the effects of expressing various tRNA genes from 2μ vectors, strains were grown using the same protocol in S minimal media lacking leucine + 2% raffinose + 2% galactose media (Sherman, 1986).

Quantitative RT-PCR.

Strains were grown as described above for flow cytometry. We measured GFP and RFP for each strain and harvested cell pellets from the same culture. Bulk RNA was prepared using glass beads and treated with DNase (Promega); a total of 62.5 ng RNA was used to synthesize cDNA using Superscript II Reverse Transcriptase (Invitrogen). As a negative control, the reverse transcription (RT) reaction was also prepared

without enzyme. Primers used to amplify GFP cDNA and the internal controls, Actin cDNA and RFP cDNA, are shown in Table S6. cDNA was amplified using Fast SYBR Green Master Mix (Applied Biosystems), detected using the 7500 Fast Real-Time PCR system, and analyzed with the 7500 Software v2.3 (Applied Biosystems).

Strain Growth for *leu2*-d Selection in tRNA Suppression Studies

Strains transformed with the 2μ *leu2-d* vectors, pECB1118 and pECB1406, were grown for ~18 hours at 30 °C in 5 ml S-ura + 2% raffinose + 2% galactose + 80 mg/L Ade media, followed by dilution to OD_{600} of 0.01 in 5 ml S-ura-leu + 2% raffinose + 2% galactose + 80 mg/L Ade media and grown overnight at 30 °C. Approximately 4 hours before flow cytometry analysis, the strains were diluted to OD_{600} of 0.25 in 5 ml S-ura-leu + 2% raffinose + 2% galactose + 80 mg/L Ade media (Whipple et al., 2011).

Yeast Translation Efficiency Data Sources and ORF Comparisons

We downloaded *S. cerevisiae* coding ORFs from SGD (www.yeastgenome.org). ORF mRNA levels were downloaded from the RNA sequencing work of Presnyak (Presnyak et al., 2015). From this work, we used the mRNA levels from rRNA-depleted, whole-cell RNA at the initial expression level and took the average of two runs. Protein copy number estimates were downloaded from the mass spectrometry work of Kulak (Kulak et al., 2014)

The combined dataset of mRNA and protein measurements included 4,489 yeast ORFs. We applied a minimum mRNA threshold of 1. For each ORF we calculated the mean CAI of its codons. Mean CAI for ORFs with an inhibitory pair fell predominately in the range from 0.4 to 0.6. Across this range, we grouped ORFs into 8 CAI bins of size 0.025. Six CAI bins from 0.425 through 0.575 had at least 30 ORFs in each category. For these bins, we ran t-tests between the estimated translation efficiencies ($\log_2$(protein/mRNA)) of ORFs with at least one of the 17 inhibitory codon pairs and ORFs without any identified pairs.

For a subset of 12 pairs (excluding AUA-CGA, CGA-AUA; CUG-CGA, CGA-CUG; and CGA-CGA) we compared ORFs with at least one of the 12 pairs to ORFs with at least one of these pairs in the reverse codon order. These categories were mutually exclusive, such that if an ORF had an inhibitory pair and a reverse pair, it was excluded from the analysis. (There were 614 ORFs with both inhibitory and reverse pairs). All t-test p-values were corrected using the Holms-Bonferroni procedure.

Footprint Counts and Ribosome Occupancy

We used A-site codon footprint counts derived from the Jan et al. (Jan et al., 2014) sec63mVenusBirA_-CHX_1minBiotin_input dataset, a whole cell ribosome profiling experiment with no cycloheximide treatment. For each codon pair, we aligned all ORFs with the pair, relative to the pair site. Then we took the sum of A-site footprint counts at aligned codon positions, thereby obtaining a joint footprint count at the pair and at each codon distance up to 49 codons away from the pair (49 positions 5' of the pair and 49 positions 3' of the pair). We calculated ribosome occupancy at each position as the joint count relative to total joint counts across the 100-codon position window. For Fisher's exact test comparisons, we used the sum of joint counts at positions with the pair in the ribosomal P, A-sites and E, P-sites versus the joint counts sum across all surrounding window positions (except those with the pair at A-site, -1 and +1, E-site locations).

From Lareau et al. 2014 (Lareau et al., 2014), we took the tallied ribosome footprints at inferred A-site codons in untreated cells. We used the total of both short (20-22 nucleotides) and long (28-30 nucleotides) mRNA fragment sizes. These datasets excluded ORFs with fewer than ten footprints. The authors also excluded footprints for the first 50 codons of each open reading frame (ORF) due to the scarcity of footprints from these regions. ORF coverage (footprints/codon) varied widely across the three biological replicates. Overall, Replicate 1 (GSM1406453) averaged 0.65 footprints per codon (from the pooled dataset), whereas Replicate 2 (GSM1406454) averaged 0.54 and Replicate 3 (GSM1406455) only 0.16. Inhibitory pairs identified in our GFP assay occurred relatively infrequently in the sampled transcriptome and often in transcripts with relatively low expression (where footprints were rarer). We considered sufficient coverage of many ORFs in the low expression range critical to our analysis. Thus, we pooled footprint counts from each replicate, and then follow the same analysis procedure as with the Jan et al. dataset.

Statistical Methods

For the permutation p-values of 6-mer sequence enrichment in low expression variants, we calculated Benjamini-Hochberg false discovery rates (FDR) to control for the number of false positives. The 28 candidate pairs reached significance in the full dataset at a FDR of 3%, while the revised list of 20 candidate pairs reached significance in the reduced structure set of variants at a FDR of 7%. In evaluating the reduced structure set, we determined permutation p-values based on occurrences at the combination of s1 and s4 positions; as opposed to at each position independently.

We ran one-sided Wilcoxon rank sum tests to compare syn-GFP^{SEQ} distributions. Wilcoxon rank sum tests were carried out in R. For each of the 20 candidate pairs, we compared the distribution of variants with an inhibitory pair to variants with the 6-mer sequence in an out-of-frame position as well as to variants with the two codons present but separated. For 12 inhibitory pairs we compared the distribution of variants with the inhibitory pair to the distribution of variants with the codons in reverse order. We corrected Wilcoxon p-values for 52 tests using the Holms-Bonferroni procedure.

In the ribosome profiling analysis, permutation analysis p-values were corrected for 17 inhibitory codon pair tests using the Holms-Bonferroni procedure. To directly compare the significance of ribosome occupancy differences between related, comparison pairs, we used a one-sided Fisher's exact test. For each pair, we took the footprint count with the pair occupying 2 ribosomal site positions and footprint count in the remainder of the 100 codon distance window. We evaluated whether the inhibitory codon pair had a higher proportion of footprints at ribosomal sites than the comparison pair. We ran these Fisher's exact tests in R, and we compared the footprint counts of 17 inhibitory pairs to two synonymous sequences. We also compared the footprint counts to the counts for a pair with the same codons in reverse order (except for the CGA-CGA pair). We corrected Fisher's exact p-values for 50 tests using the Holms-Bonferroni procedure. From each Fisher's exact test we also obtained the odds ratio (calculated by conditional maximum likelihood estimation) and confidence interval.

SUPPLEMENTAL TABLES AND LEGENDS

Codon	Amino Acid	Frequency	CAI	Wobble
CGA	R	0.092	0.002	I•A
CUG	L	0.077	0.003	U•G
CGG	R	0.071	0.002	-
AGG	R	0.060	0.003	-
CUU	L	0.047	0.006	U•U
GUA	V	0.043	0.002	-
GUG	V	0.041	0.018	-
CCG	P	0.035	0.002	U•G
AUA	I	0.034	0.003	-
CUC	L	0.032	0.003	-

Table S2. Frequency of Codon Use in Low Variant Insertions, Related to Figure 1
The top 10 frequencies are shown. The expected frequency due to chance is 1/61 or 0.016.

Candidate	Median	IQR	n	Construct	Type	GFP^{FLOW}	Inhib./Opt.
AGG-CGA	0.48	0.31	30	AGGCGAAAT	Inhibitory	58.0 ±2.0	0.42 ±0.02
				AGAAGAAAT	Optimal	138.1 ±3.5	
AGG-CGG	0.82	0.46	36	AGGCGGCAC	Inhibitory	58.1 ±2.7	0.52 ±0.03
				AGAAGACAC	Optimal	112.0 ±3.8	
AUA-CGA	0.58	0.30	11	ATACGAGAT	Inhibitory	55.7 ±1.0	0.39 ±0.01
				ATTAGAGAT	Optimal	143.1 ±2.6	
AUA-CGG	0.65	0.43	27	ATACGGACG	Inhibitory	56.8 ±0.5	0.64 ±0.02
				ATTAGAACG	Optimal	89.0 ±3.3	
CGA-AUA	0.51	0.29	27	CGAATACAT	Inhibitory	38.6 ±1.7	0.34 ±0.02
				AGAATTCAT	Optimal	112.9 ±0.01	
CGA-CCG	0.44	0.05	22	CGACCGAGC	Inhibitory	13.0 ±0.5	0.16 ±0.01
				AGACCAAGC	Optimal	82.6 ±2.1	
CGA-CGA	0.44	0.06	25	CGACGAACT	Inhibitory	24.5 ±1.0	0.19 ±0.01
				AGAAGAACT	Optimal	126.7 ±4.1	
CGA-CGG	0.48	0.13	38	CGACGGAGC	Inhibitory	42.2 ±0.7	0.35 ±0.01
				AGAAGAAGC	Optimal	121.6 ±2.3	
CGA-CUG	0.47	0.31	21	AACCGACTG	Inhibitory	71.6 ±0.9	0.46 ±0.01
				AACAGATTG	Optimal	154.8 ±4.2	
CGA-GCG	0.44	0.03	30	CGAGCGAGT	Inhibitory	35.2 ±0.5	0.26 ±0.01
				AGAGCTAGT	Optimal	136.0 ±3.1	
CUC-AUA	0.70	0.45	12	CTCATAACG	Inhibitory	60.4 ±6.9	0.47 ±0.05
				TTGATTACG	Optimal	129.1 ±1.7	
CUC-CCG	0.44	0.04	15	CTCCCGACT	Inhibitory	17.8 ±0.8	0.14 ±0.01
				TTGCCAACT	Optimal	126.9 ±1.3	
CUG-AUA	0.71	0.30	22	CTGATAATG	Inhibitory	58.0 ±3.0	0.61 ±0.06
				TTGATTATG	Optimal	94.3 ±7.8	
CUG-CCG	0.49	0.39	30	CTGCCGACC	Inhibitory	49.9 ±0.4	0.39 ±0.01
				TTGCCAACC	Optimal	127.7 ±1.6	
CUG-CGA	0.50	0.48	25	CTGCGAAGT	Inhibitory	46.0 ±1.2	0.37 ±0.01
				TTGAGAAGT	Optimal	124.1 ±1.5	
CUG-CUG	0.66	0.31	25	CTGCTGACA	Inhibitory	62.4 ±3.4	0.76 ±0.08
				TTGTTGACA	Optimal	82.1 ±8.0	
CUU-CUG	0.74	0.32	27	CTTCTGACG	Inhibitory	65.6 ±4.4	0.66 ±0.06
				TTGTTGACG	Optimal	99.1 ±5.6	
GUA-CCG	0.80	0.42	25	GTACCGAGT	Inhibitory	59.9 ±2.0	0.42 ±0.03
				GTTAGAAGT	Optimal	142.0 ±7.5	
GUA-CGA	0.53	0.21	36	GTACGACAA	Inhibitory	35.7 ±3.4	0.39 ±0.04
				GTTAGACAA	Optimal	91.5 ±1.9	
GUG-CGA	0.60	0.33	30	GTGCGAACT	Inhibitory	50.0 ±0.6	0.43 ±0.01
				GTTAGAACT	Optimal	115.8 ±1.7	

Table S4. Candidate Inhibitory Pairs and GFP^{FLOW} Comparisons, Related to Figure 1. For each candidate, the syn-GFP^{SEQ} median, interquartile range (IQR), and number of library variants (n) is shown with the individual construct sequences used in validating reduced expression. GFP^{FLOW} (GFP*100/RFP) are the mean of 3 or 4 independent isolates ±SD.

SUPPLEMENTAL REFERENCES

Grant, T.D., Snell, E.H., Luft, J.R., Quartley, E., Corretore, S., Wolfley, J.R., Snell, M.E., Hadd, A., Perona, J.J., Phizicky, E.M.*, et al.* (2012). Structural conservation of an ancient tRNA sensor in eukaryotic glutaminyl-tRNA synthetase. Nucleic Acids Res *40*, 3723-3731.

Guy, M.P., Young, D.L., Payea, M.J., Zhang, X., Kon, Y., Dean, K.M., Grayhack, E.J., Mathews, D.H., Fields, S., and Phizicky, E.M. (2014). Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. Genes Dev *28*, 1721-1732.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863-864.

Sherman, F., Fink,G., and Hicks,J.B. (1986). In Methods in Yeast Genetics (New York: Cold Spring Harbor Laboratory Press), pp. 145-149.

Whipple, J.M., Lane, E.A., Chernyakov, I., D'Silva, S., and Phizicky, E.M. (2011). The yeast rapid tRNA decay pathway primarily monitors the structural integrity of the acceptor and T-stems of mature tRNA. Genes & Development *25*, 1173-1184.

Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: Analysis and design of nucleic acid systems. J Comput Chem *32*, 170-173.