SUPPLEMENTARY FIGURES

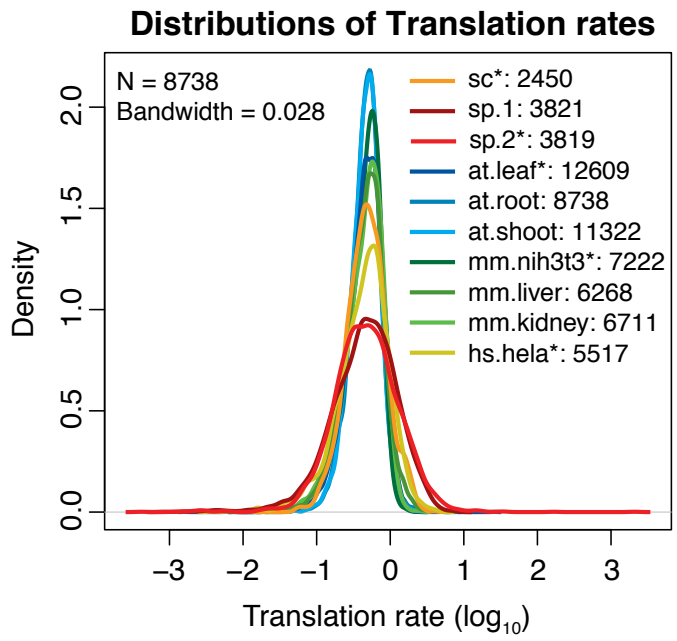

Figure S1. The distributions of translation rates. The $\log_{10}$ transformed ribosome profiling data have been scaled to have the same median while retaining their original variance. The species is denoted by a two letter code. The number of genes in each dataset is indicated. The example datasets are sc; sp.2; at.leaf; mm.nih3t3; and hs.hela (*).

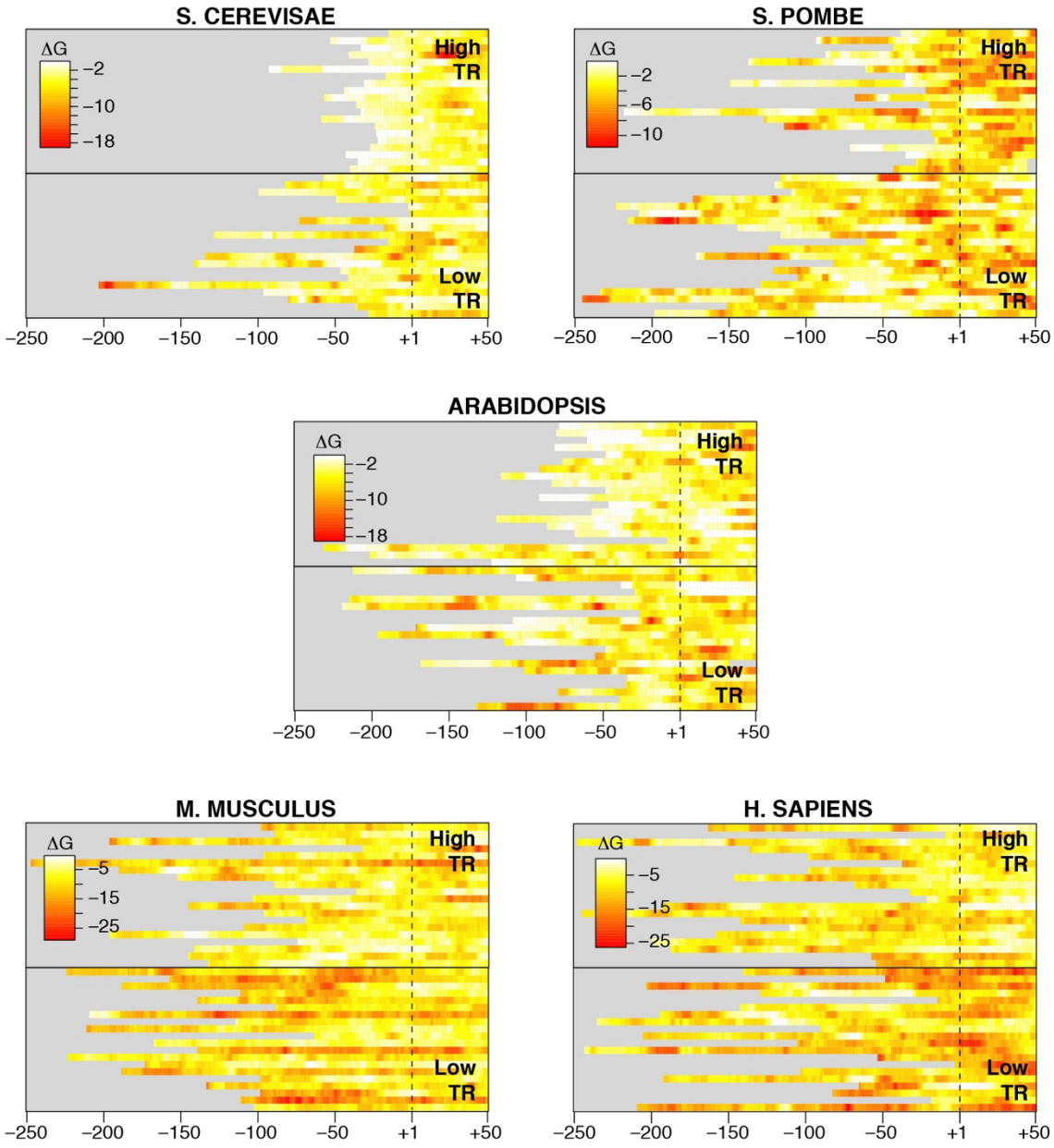

Figure S2. The locations of secondary structures in individual mRNA are highly variable. The heat maps show the free energy of RNA folding (ΔG kcal/mol) of 35 nucleotide windows for the 20 most highly translated mRNAs with 5'UTRs ≤250 nucleotides (high TR) and for the 20 most poorly translated mRNAs with the same length constraint (low TR). Windows representing every one nucleotide offset were calculated (*x*-axis). The mRNAs are aligned at the iAUG (dashed line). While the most strongly folded regions tend to be found in low TR mRNAs, the locations of the most folded regions within individual mRNAs are highly variable.

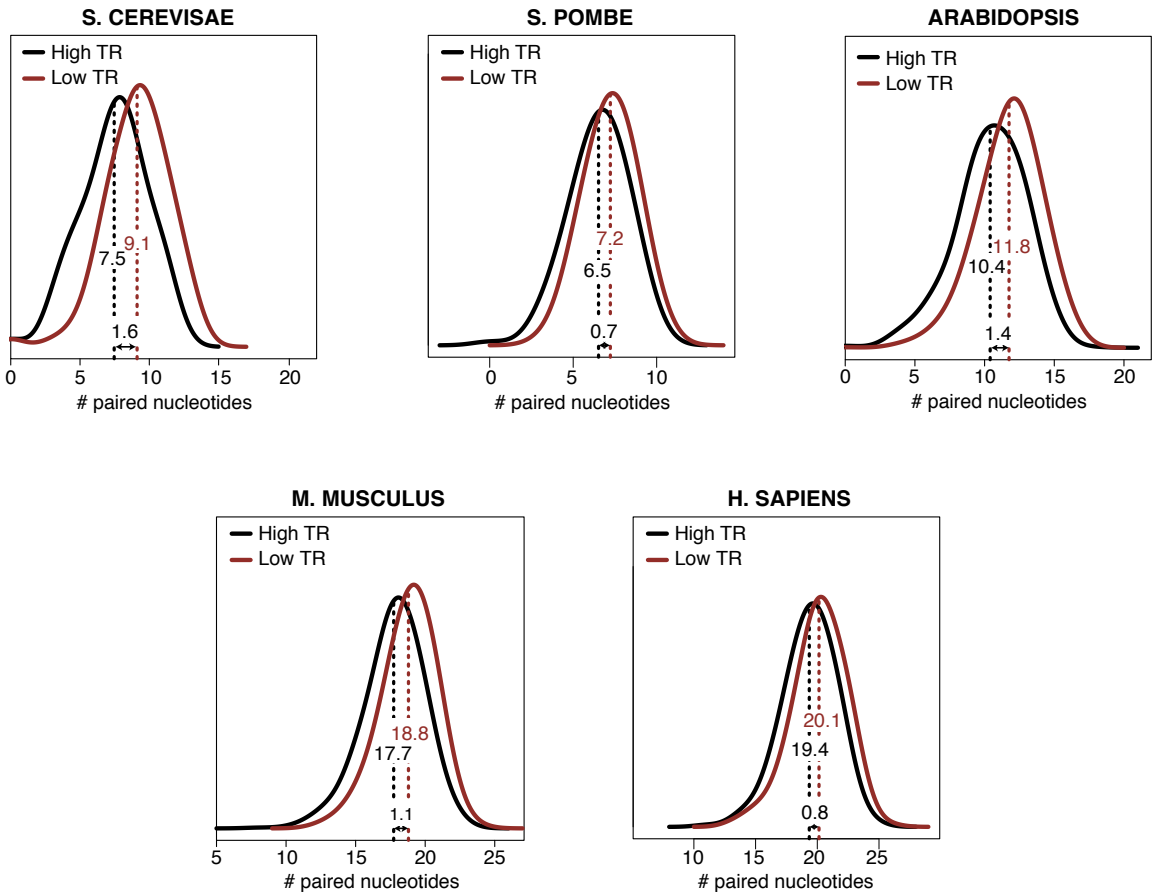

Figure S3. The distributions of paired nucleotide number in the most folded windows of highly and poorly translated mRNAs. The most folded ("min") windows were ranked based on the translation rate (TR) of the mRNA, and the 1st (high) and 10th (low) deciles identified. The distributions of the number of paired nucleotides in the most folded windows in each of the two cohorts are plotted. The vertical dotted line indicates the mean for each cohort. The result shows while the low TR cohort contains on average more paired nucleotides than the high TR cohort, there is a considerable overlap in structures between the two cohorts.

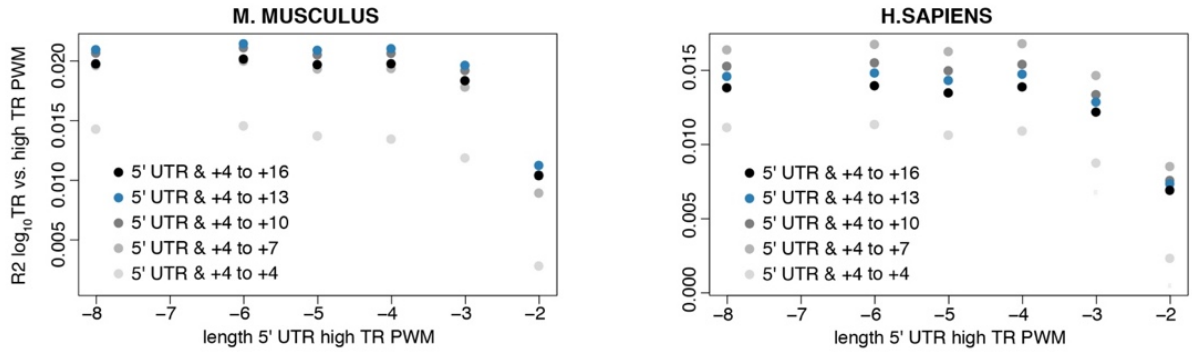

Figure S4. Fine mapping the 5' and 3' boundaries of APEs in *M. musculus* and *H. sapiens*. The R^2 coefficients of determination between $\log_{10}$ translation rates (TR) and PWM scores. The figure is as described in Figure 7 except that PWMs were chosen to more precisely map APE boundaries. The results show the APEs extend from -6 to +13 in *M. musculus* and -6 to +7 in *H. sapiens*.

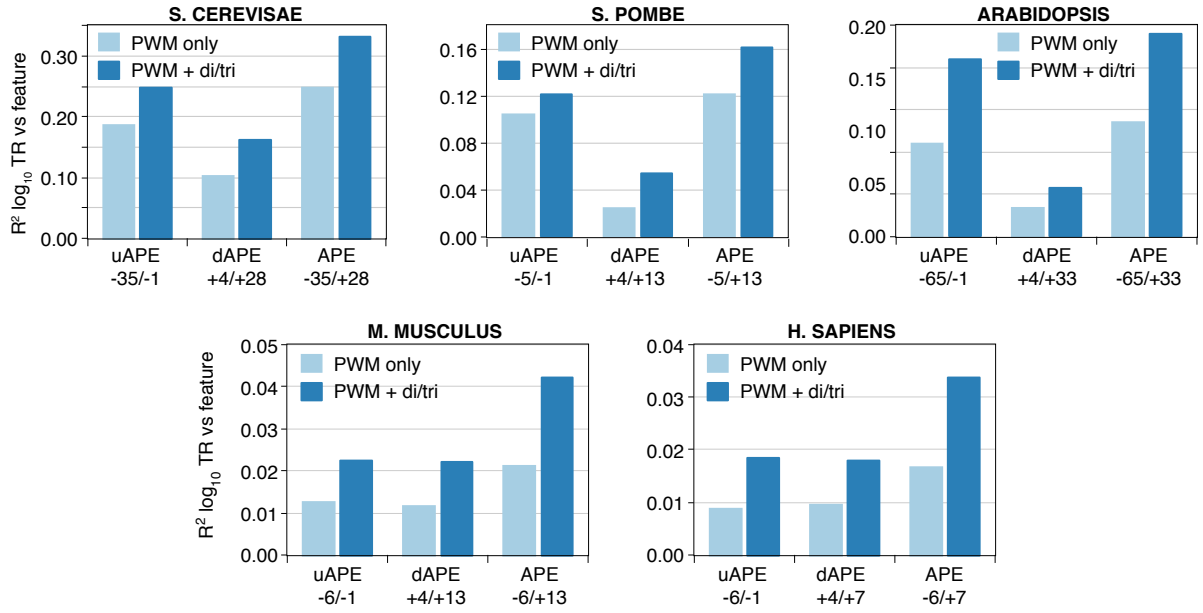

Figure S5. AUG proximal elements (APEs) comprise sequences upstream and downstream of the iAUG and are best described by Position Weight Matrices (PWMs) and di- and tri-nucleotide frequencies. The R^2 coefficients of determination between $\log_{10}$ translation rate (TR) and feature(s) describing the iAUG upstream portion of the APE (uAPE); the iAUG downstream portion of the APE (dAPE); and the complete APE. Results are shown for a model employing only the PWM score and for a multivariate model combining the PWM score with a BIC selected subset of di and tri-nucleotide frequencies. The models are described in Supplementary file 3.

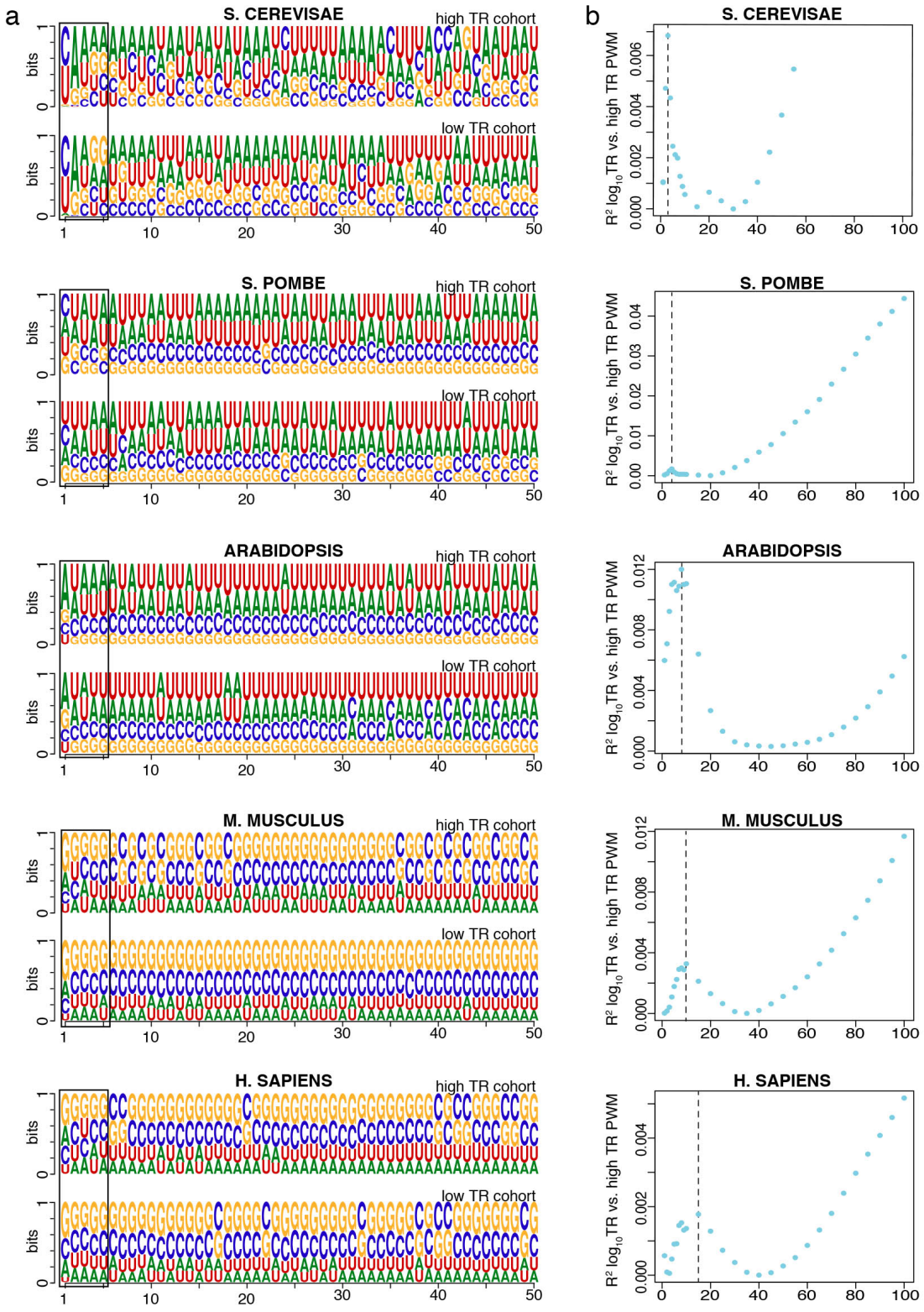

Figure S6. A 5' cap element. (a) PWMs for the 10% of mRNAs with the highest translation rate (high TR cohort) and the 10% with the lowest rate (low TR cohort). Sequence logos show the frequency of each nucleotide at each position relative to the first nucleotide of the transcript (i.e.

the 5' cap). Only 5'UTR sequences that lie 5' of the APE were included in the analysis. (b) The R^2 coefficients of determination between $\log_{10}$ TR and PWM scores. PWMs of varying lengths were built from the sequences of the high TR cohort. Log odds scores were then calculated for all mRNAs that completely contained a given PWM. PWMs extending 3' from the 5' cap in 1 nucleotide or 5 nucleotide increments were tested (x-axis, right to left). A local maxima in R^2 values is seen between 3 to 15 nucleotides from the 5' cap, depending on the species. Because this 5'cap element only controls less than 1.2% of the variance in TR and to simplify our models, the PWM score of the first 5 nucleotides was used in a model for the 5'ofAPE region for all species, together with a BIC selected subset of di and tri-nucleotides from the entire 5'ofAPE region.

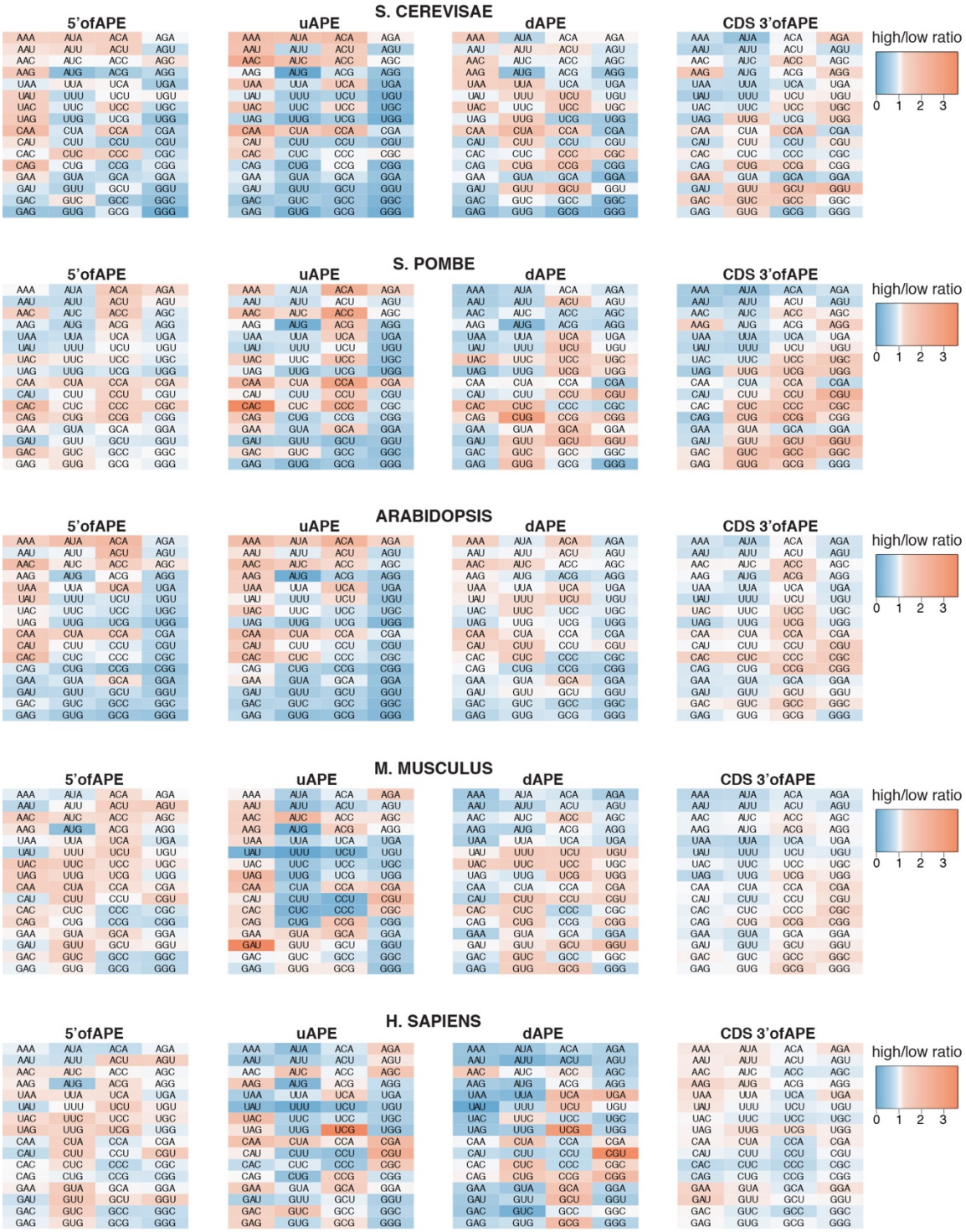

Figure S7. Tri-nucleotides regulating translation. The heat maps show the frequency of each tri-nucleotide in the most highly translated 10% of genes divided by its frequency in the most poorly translated 10% of genes ((TR high / TR low) ratios). Results are presented for four mRNA regions of each species.

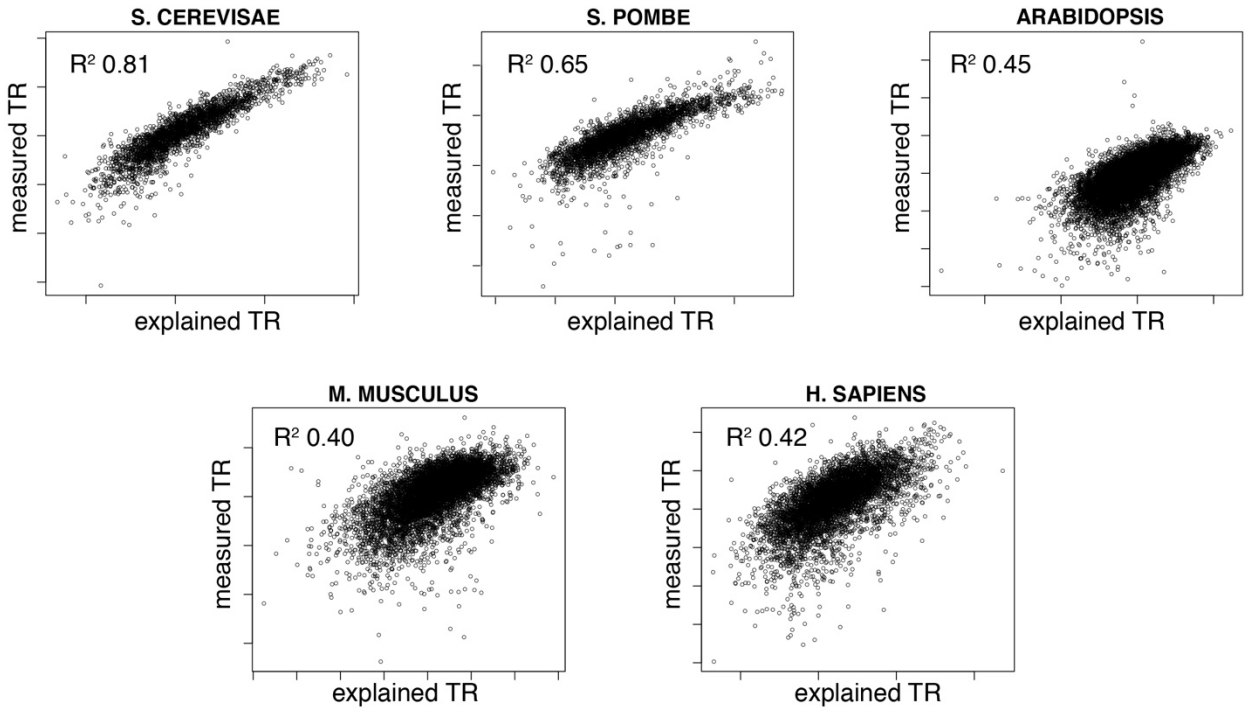

Figure S8. Multivariate models for the five general features. Scatter plots show the relationship between measured translation rates (*y*-axis) and translation rates explained by multivariate models for the five general features (*x*-axis). The R2 coefficients of determination are given.

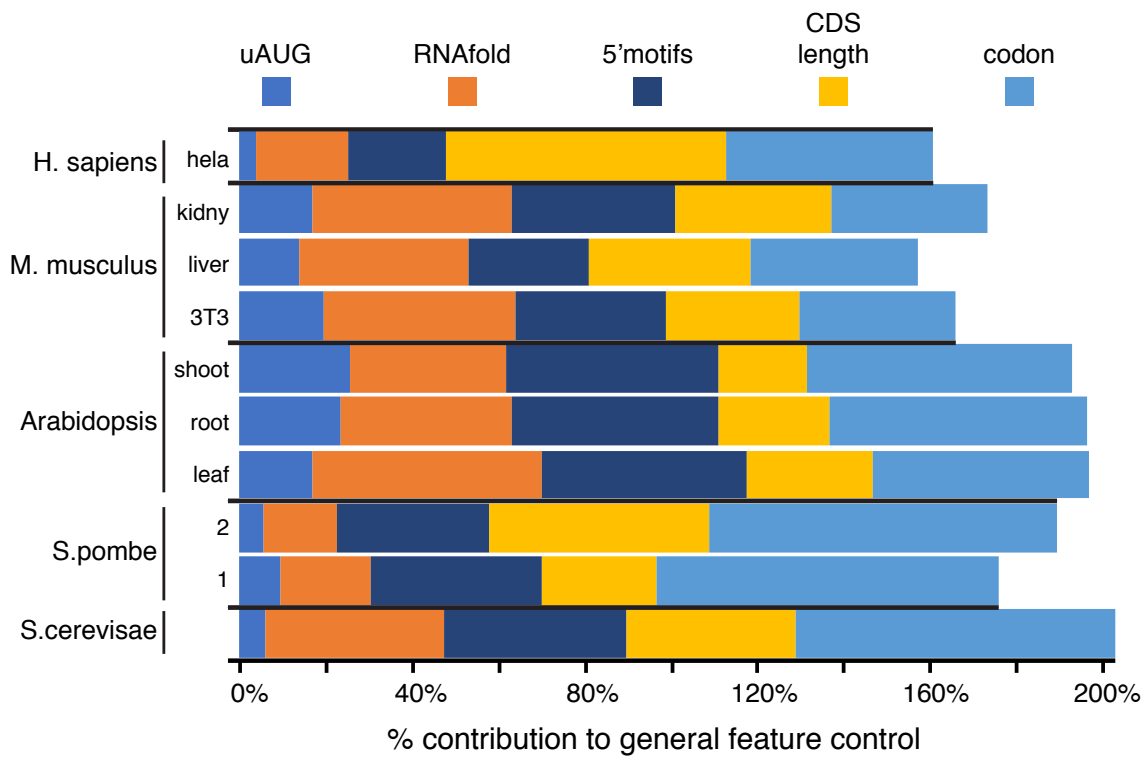

Figure S9. (a) For each feature separately, its R^2 coefficient of determination vs $\log_{10}$ translation rate is given as a percent the R^2 coefficient for a linear multivariate model for all five features. The result shows that the sum of these percent contributions is much greater than the variance in translation explained by the multivariate five-feature model. The R^2 values and the values plotted for each feature and the combined model are given in Additional file 8.

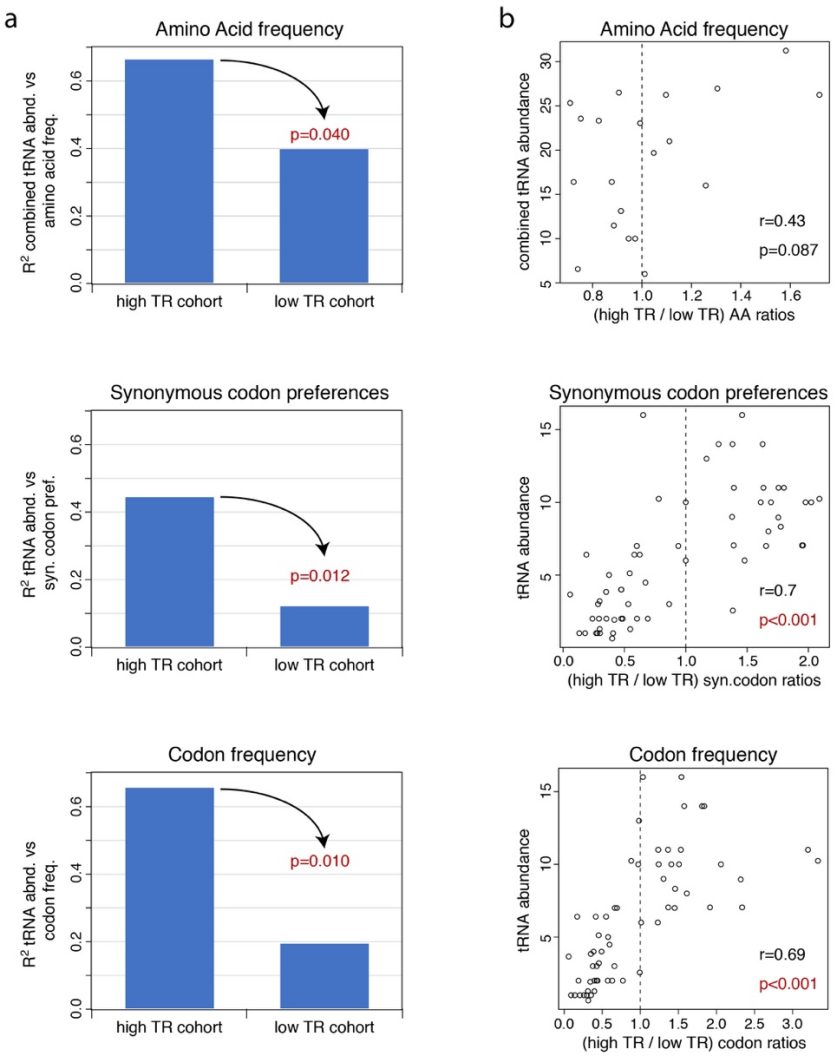

Figure S10. Control by amino acid frequencies and synonymous codon preferences correlates with tRNA abundances. The relationship between tRNA abundances and control by amino acid content and synonymous codon preferences was tested in *S. cerevisiae* because estimates of effective tRNA abundances are particularly well established for this species. The frequencies of amino acids (AA), or codons (codon), or the preferences for synonymous codon (syn.codon) were determined separately for the most highly translated 10% of genes (high TR) and for the most poorly translated 10% of genes (low TR). Genes not encoding all 20 amino acids were excluded from the analysis. (a) The coefficient of determination (R^2) for the high TR cohort or low TR cohort AA, syn.codon or codon frequencies vs their cognate tRNA abundances. For AA, the frequencies of all cognate tRNAs for each amino acid were summed to give a combined

tRNA abundance. *p*-values testing if the correlation of tRNA abundance with the high TR cohort is greater than that with the low TR cohort are given, with significant *p*-values shown in red. The High TR mean frequencies correlate more strongly with tRNA abundances than do the low TR frequencies, indicating that translation of high TR mRNAs uses the cellular population of amino acylated tRNAs more efficiently than translation of low TR mRNAs. (b) The ratios between AA, syn.codon or codon frequencies in the high TR cohort divided by those in the low TR cohort were determined. Ratios > 1 thus indicate a larger frequency in high TR cohorts than in low TR cohorts. Scatter plots are shown between these (high TR /low TR) ratios and tRNA abundance along with the Pearson correlation coefficients (r) and *p*-values testing if the correlations are significant (significant *p*-values in red). Dashed vertical lines indicate a ratio of 1. The Pearson correlation coefficients range from +0.43 – +0.70, establishing that codons for high abundance tRNAs are more prevalent in highly translated mRNAs, whereas codons for low abundance tRNAs are more prevalent in poorly translated mRNAs.