Contributions of sequence features to the variance in translation rates among endogenous mRNAs

Sequence feature(s)	R^2 feature vs TR
S. cerevisiae, correlation to initiation rate modelled from RP data (Shah et al., 2013)	
-4 to + 37 folding energy	0.02
CDS length	0.35
S. cerevisiae, correlation to initiation rate modelled from RP data (Weinberg et al., 2016)	
5' most 70 nucleotides folding energy	0.14
CDS length	0.24
5' UTR folding energy, 5' UTR %GC, 5' UTR #uAUGs, 5'UTR length, CDS length	0.39
S. cerevisiae, correlation to initiation rate modelled from RP data (Li et al., 2017)	
5' UTR length	0.05
5' UTR #uORFs	0.14
5' UTR folding energy	0.19
-35/+28 motif	0.33
CDS length	0.32
codon frequency	0.60
all 5' UTR features, -35/+28 motif, CDS length	0.58
all 5' UTR features, -35/+28 motif, CDS length, codon frequency	0.80
mouse ES cell line (Chew et al., 2016)	
5' UTR density of uAUGs	0.04
5' UTR length	0.01
5' UTR mean folding energy	0.01
-25 to +10 folding energy	0.01
+1 to + 35 folding energy	0.01
CDS mean folding energy	0.06
-10 to +13 motif	0.02
all above seven features	0.14
mouse liver (Janich et al., 2015)	
5' UTR length	0.05
CDS length	0.16
3' UTR length	0.02

* R^2 coefficient of determination estimated as the square of a spearman rank correlation coefficient.

Contributions of sequence features/motifs to the variance in translation rates across a set of heterologous mRNAs that have varying 5' UTRs fused upstream of a reporter CDS

Sequence feature(s)	R^2 feature vs TR
S. cerevisiae, in vitro, 18 natural 5' UTRs (Rojas-Duran and Gilbert, 2012)	
5' UTR length	0.18
5' UTR folding energy	0.25
S. cerevisiae, in vivo, 141 random mutations in region –4 to +37 (Shah et al., 2013)	
-4 to +37 folding energy	0.22
tRNA index	0.02
S. cerevisiae, in vivo, 2,041 random mutations in region -10 to -1 (Dvir et al., 2013)	
-15 to +50 folding energy	0.18
-10 to -1 uAUGs	0.06
-3 to -1 sequence motif	0.29
-10 to -1 kmer frequency motif	0.19
folding energy and sequence motifs (13 features)	0.68
S. cerevisiae, in vivo, 383 random mutations in -14 to -1 or in the CDS (Ben-Yehezkel et al., 2015)	
-14 to +39 mean folding energy	0.12*
-7 to +33 folding energy	0.20*
A at -3	0.12*
T at -3	0.12*
AUG context motif	0.21*
S. cerevisae, in vivo, 500,000 random mutations in region -50 to -1 (Cuperus et al., 2017)*	
-50 to -1 folding energy	0.08
-50 to -1 >100 13 mer PWMs / Neural network model	0.47-0.62
S. cerevisae, in vivo, 11,856 50 nucleotide segments of natural 5' UTRs (Cuperus et al., 2017)*	
>100 13 mer PWMs / Neural network model trained on 500,000 random mutants	0.60
H. sapiens, in vivo, 300,000 random mutations in region -50 to -1 (Sample et al., 2018)	
-50 to -1 folding energy	0.19
-50 to -1 >100 8 mer PWMs / Neural network model	0.93
H. sapiens, in vivo, sequences -50 to -1 for 35,212 natural 5' UTRs (Sample et al., 2018)	
-50 to -1 >100 8 mer PWMs / Neural network model	0.81
H. sapiens, in vivo, 65,536 random mutations in region -6 to +5 (Noderer et al., 2014)	
PWM and di-nucleotide motif	0.83

* R^2 coefficient of determination estimated as the square of a spearman rank correlation coefficient.
** An indirect measure of translation based on competative growth selection which could also capture effects on mRNA stability and transcription.

Ben-Yehezkel, T., Atar, S., Zur, H., Diament, A., Goz, E., Marx, T., Cohen, R., Dana, A., Feldman, A., Shapiro, E.*, et al.* (2015). Rationally designed, heterologous S. cerevisiae transcripts expose novel expression determinants. RNA Biol *12*, 972-984.

Chew, G.L., Pauli, A., and Schier, A.F. (2016). Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. Nat Commun *7*, 11663.

Cuperus, J.T., Groves, B., Kuchina, A., Rosenberg, A.B., Jojic, N., Fields, S., and Seelig, G. (2017). Deep learning of the regulatory grammar of yeast 5' untranslated regions from 500,000 random sequences. Genome Res. *27*, 2015-2024.

Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L.B., Weinberger, A., and Segal, E. (2013). Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. Proc. Natl. Acad. Sci. USA *110*, E2792-2801.

Janich, P., Arpat, A.B., Castelo-Szekely, V., Lopes, M., and Gatfield, D. (2015). Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res. *25*, 1848-1859.

Li, J.J., Chew, G.L., and Biggin, M.D. (2017). Quantitating translational control: mRNA abundance-dependent and independent contributions and the mRNA sequences that specify them. Nucleic Acids Res. *45*, 11821-11836.

Noderer, W.L., Flockhart, R.J., Bhaduri, A., Diaz de Arce, A.J., Zhang, J., Khavari, P.A., and Wang, C.L. (2014). Quantitative analysis of mammalian translation initiation sites by FACS-seq. Mol Syst Biol *10*, 748.

Rojas-Duran, M.F., and Gilbert, W.V. (2012). Alternative transcription start site selection leads to large differences in translation activity in yeast. RNA *18*, 2299-2305.

Sample, P.J., Wang, B., Reid, D.W., Presnyak, V., McFadyen, I., Morris, D.R., and Seelig, G. (2018). Human 5' UTR design and variant effect prediction from a massively parallel translation assay. BioRxiv.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. Cell *153*, 1589-1601.

Weinberg, D., Shah, P., Eichhorn, S., Hussmann, J., Plotkin, J., and Bartel, D. (2016). Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Reports *14*, 1787-1799.