Supplementary Materials for

ROSE: a deep learning based framework for predicting ribosome stalling

Sai Zhang[1,†], Hailin Hu[2,†], Jingtian Zhou[2,†], Xuan He[1], Tao Jiang[3,4,5] and Jianyang Zeng[1,*]

1 Supplementary Notes

1.1 Functional analysis of ribosome stalling in the ramp sequences

From previous studies in the literature [1, 2, 3], we expected a gene with a higher interRSS in it ramp region to be more efficiently translated. To verify this hypothesis, we performed a large-scale study to investigate the relation between the interRSS of the ramp region and the translation efficiency of the corresponding gene. We used the logarithm of the protein expression level divided by the corresponding mRNA expression level to measure the translation efficiency (TE) of each gene. In our analysis, the mRNA and protein expression data of human (lymphoblastoid cell lines, LCLs) and yeast (*S. cerevisiae*) were obtained from [4] and [5, 6], respectively. As short genes may introduce bias to our analysis of the ramp regions, here we only focused on those genes with more than 300 codons, which in total resulted in 12,734 and 3,590 genes for human and yeast, respectively. We then divided these obtained genes into two classes: those with the highest 10% mean interRSSes of the ramp regions were denoted by "ramp+", while the remaining ones were denoted by "ramp-". Our comparison showed that the genes in ramp+ owned stringently higher translation efficiency than those in ramp- (Supplementary Figs. 4(a) and 4(b); $P = 3.85 \times 10^{-7}$ for human and $P = 2.7 \times 10^{-3}$ for yeast, one-sided Wilcoxon rank-sum test).

Furthermore, we investigated the enriched functional categories of those genes with high ribosome stalling potential in their ramp sequences. Specifically, we only focused on the genes with high translation efficiency (i.e., with top 50% TE) as they may be crucial for supporting the fundamental cellular activities, and used all the expressed genes in the dataset as the background for the gene ontology (GO) analysis (Supplementary Table 2). Interestingly, we found that in yeast, those genes in ramp+ were significantly enriched with many housekeeping GO terms, such as translation regulation and amino acid biogenesis, while the others in ramp- showed much weaker

[1] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China.
[2] School of Medicine, Tsinghua University, Beijing, China.
[3] Department of Computer Science and Engineering, University of California, Riverside, CA, USA.
[4] MOE Key Lab of Bioinformatics and Bioinformatics Division, TNLIST/Department of Computer Science and Technology, Tsinghua University, Beijing, China.
[5] Institute of Integrative Genome Biology, University of California, Riverside, CA, USA.
[†] These authors contributed equally to this work.
[*] To whom correspondence should be addressed. Email: zengjy321@tsinghua.edu.cn.

enrichment or even depletion of these GO categories (Supplementary Fig. 4(c) and Supplementary Table 3). For human, such a GO enrichment was not significantly observed after correcting the P values (Supplementary Table 3; $P > 0.05$). This may be due to the intrinsic expression property of LCLs, as we also failed to observe a significant GO enrichment even when simply focusing on those genes of high TE without differentiating their interRSS levels of the ramp sequences (Supplementary Table 3; $P > 0.05$). Nevertheless, for the cell cycle related GO terms, the corrected P values of genes in ramp+ with high TE were expectedly smaller than those of genes in ramp- (Supplementary Table 3). Taken together, these results suggested that interRSS can provide a good indicator for studying the regulatory functions of the ramp sequences, which may modulate the translation efficiency of important genes at the elongation level.

1.2 Supplementary analysis results on associations between diverse putative factors and ribosome stalling

tRNA adaptation In general, the codons recognized by abundant cognate tRNAs have short decoding time [7]. However, previous analyses of ribosome profiling data often led to inconsistent conclusions on the effects of this feature, which may be attributed to experimental bias, methodological difference or unknown coregulation factors [8, 9, 10, 11, 12]. In addition to tRNA concentration, the strength of a wobble pairing interaction may also influence the elongation rate [13]. The tRNA adaptation index (tAI) has been proposed to consider both the tRNA concentration (approximated by the copy number of the corresponding tRNA gene) and the strength of codon-anticodon pairing (computed according to the Crick wobble rules) [14]. In fact, it has been found that codon usage bias is often correlated with tRNA abundance [15, 16, 17, 1]. Our analysis also observed a certain positive correlation between cAI (also %MinMax) and tAI for both human and yeast datasets (Supplementary Table 1). We reexamined the relation between the tRNA adaptation and ribosome stalling using ROSE, in which the ribosome occupancy sites were quantified by their tAI scores. In particular, we compared the intraRSSes of the ribosome A- and P-sites enriched with low tAI scores to those of the background, and our comparative analysis on both yeast and human datasets supported the conclusion that lower tRNA concentrations and weaker wobble pairing interactions correlate closely with higher ribosome stalling tendency (Figs. 3(a) and 3(b), Supplementary Figs. 5(a) and 5(b); $P < 10^{-60}$ by one-sided Wilcoxon rank-sum test).

mRNA secondary structure During the translation elongation process, the ribosome should first unwind the locally folded mRNA secondary structures (e.g., stem-hairpin or stem-internal loops) to move forward [18, 19]. This indicates that in a highly double-stranded region, translation elongation can be slowed down, which thus increases the probability of ribosome stalling [20, 2, 11, 13, 9, 21, 22, 23, 24]. To verify this hypothesis, we first ran RNAfold [25] to predict the secondary structures of all mRNA sequences in the background dataset, which contained 10,000 randomly-selected ribosome occupancy sites from the genome. Here, the mRNA sequences covering the codon sites of interest were 183 nucleotides long, as we extended each putative ribosome occupancy site by 30 codons both upward and downward as input to ROSE. We then measured the folding level of each sequence by computing its double-stranded ratio (denoted by ds%) in the local region of a ribosome A-site, and regarded the top 5,000 mRNA sequences with the highest ds% scores as highly folded. Next, we compared the intraRSSes of highly-folded mRNAs to those of the remaining sequences. For both human and yeast, we observed an expected increase of intraRSS for the mRNAs with highly-folded structures (Supplementary Fig. 5(e); $P < 10^{-8}$ by one-sided Wilcoxon rank-sum test), which provided another evidence on the association between mRNA secondary structure and ribosome stalling.

RNA-binding proteins Recently, RNA-binding proteins (RBPs) have received broad interests for their crucial roles in post-transcriptional and translational regulation [26, 27]. By affecting the stability and the translation process of their target mRNAs, RBPs can act as important regulatory factors to control gene expression [28, 29]. As a specific RBP, the fragile X mental retardation protein (FMRP) is essential for neuronal translation regulation, whose transcriptional inactivation has been known to be involved in many diseases, such as fragile X syndrome and autism [30]. It has been found that FMRP can associate with polyribosomes and impede the elongation of a peptide chain [31]. Structural studies have also indicated that FMRP may directly bind to both RNAs and ribosomal proteins to prohibit ribosome movement (Supplementary Fig. 6(a)) [32]. Therefore, a region with downstream FMRP binding in the CDS is expected to have a higher chance of ribosome stalling.

Here, we estimated the FMRP binding affinity of the region downstream the ribosome A-site based on the known FMRP binding sites identified by the PAR-CLIP experiment [30]. In particular, suppose that we index the codon position at the ribosome A-site as zero. Then the downstream region covering positions from +1 to +3 is still protected by the ribosome (Supplementary Fig. 1). We were particularly interested in estimating the binding affinity of FMRP in the region of next ten codons after the ribosome protected fragment (i.e., codons from +4 to +13), which was denoted by R, and then investigating the correlation between this estimated binding affinity score and ribosome stalling. We mainly used the abundance of the mapped reads of FMRP binding sites identified by PAR-CLIP [30] to estimate its binding affinity. Specifically, if there were N reads identified in region $[i, i+x]$, then for any site $s \in [i, i+x]$, its FMRP binding affinity, denoted by $\text{aff}(s)$, was estimated by $\text{aff}(s) = N/x$. After that, the overall binding affinity of the region R right after the ribosome protected fragment was calculated by $\text{aff}(R) = \sum_{s \in R} \text{aff}(s)$. Here we only considered the binding sites whose lengths were within one standard deviation from the mean calculated based on the length distribution of FMRP binding sites, as the extremely long regions may introduce bias to our analysis. We then calculated the intraRSSes of the codon sites enriched with FMRP binding downstream, and found that these sites had significantly higher intraRSSes than the background (Supplementary Fig. 6(b); $P = 5.49 \times 10^{-15}$ by one-sided Wilcoxon rank-sum test), which confirmed the effectiveness of our model to capture the ribosome stalling events regulated by this specific RNA-binding protein.

For general RBPs, the estimation of their binding affinity was similar to that for FMRP except that now we used the E- and Z-scores provided by the CISBP-RNA database [27] instead of the mapped PAR-CLIP reads. In particular, given a region R and an RBP binding motif set M, for any 7-mer m, we defined $\text{aff-max}(R) = \max_{m \in R}(\max_{m \in M}(m))$ for the max-score estimation, and $\text{aff-mean}(R) = \text{mean}_{m \in R}(\max_{m \in M}(m))$ for the mean-score estimation, where $\max_{m \in M}(m)$ returns the maximum E- or Z-score of the 7-mer m within the set M. We hypothesized that it would be highly probable for ribosomes to stall in an mRNA region enriched with RBP binding motifs downstream, as the bound RBPs may obstruct ribosome movement. To verify this hypothesis, we analyzed the intraRSSes of the codon sites with strong RBP binding propensity downstream. We found that indeed these sites exhibited significantly higher intraRSSes than the background (Figs. 3(a) and 3(b); max E-score estimation, $P = 9.36 \times 10^{-82}$ for human and $P = 1.90 \times 10^{-4}$ for yeast, one-sided Wilcoxon rank-sum test). The increases were significant for human with respect to all four criteria used to estimate the RBP binding affinity, i.e., max/average E- and Z-scores (Fig. 3(a) and Supplementary Fig. 5(a)). However, for yeast, the increase was only significant for the max E-score (Fig. 3(b) and Supplementary Fig. 5(b)), which may be attributed to the limited number of RBP binding motifs available in yeast (3 motifs in yeast vs. 91 motifs in human). Overall, our analysis suggested that RBP binding may act as another factor in influencing ribosome stalling. Of course, we cannot rule out other unknown factors associated with RBP binding mo-

tifs that virtually control ribosome stalling, which will certainly require additional experimental studies and further investigation.

Positively-charged amino acids We also studied the correlation between the amino acid charge and ribosome stalling, which still remains controversial and unclear. Several studies claimed that the positively-charged amino acids can slow down the formation of a peptide chain by interacting with the negatively-charged ribosomal exist tunnel [33, 23, 34, 9, 22]. However, others found no such correlation by arguing the quality of experimental data and the methodological limitations in the previous studies [35]. Here, we reexamined this problem based on our method. Indeed, for those codon sites enriched with the positively-charged amino acids upstream (i.e., with the 10,000 highest ratios of the positively-charged amino acids in the upstream 30 codons) in the genome, we did not observe a significant increase of intraRSS compared to the background. To probe this problem in more detail, we further looked into the specific positively-charged amino acids, including histidine, lysine and arginine, and asked whether any particular amino acid can associate with ribosome stalling. However, although significant difference of intraRSS was observed between the sites enriched with a particular amino acid upstream and the background, we cannot reach a consistent conclusion over human and yeast datasets (Supplementary Fig. 7).

2 Supplementary Tables

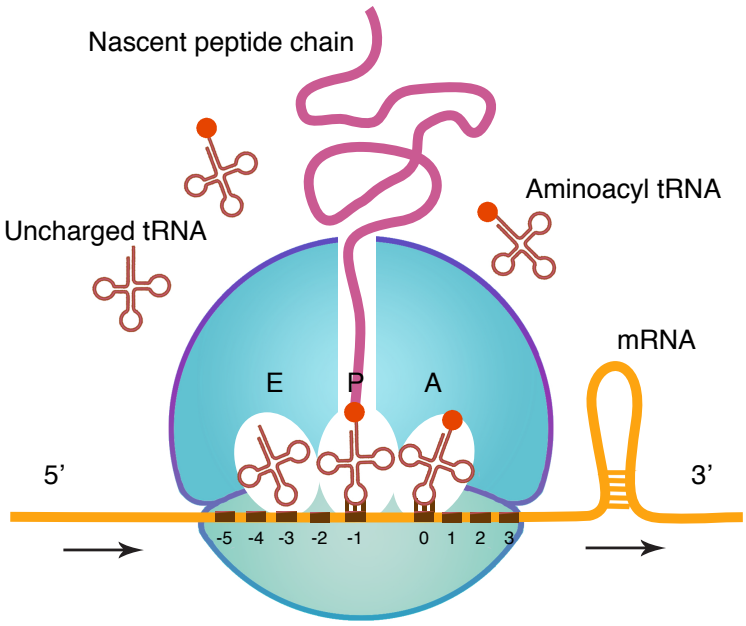
Supplementary Table 1. Spearman correlation coefficients between different putative factors that may affect ribosome stalling for both human and yeast datasets.

Supplementary Table 2. All expressed genes in LCLs and *S. cerevisiae* that were used as the background for GO analysis for both human and yeast. The Ensembl gene ID was used as the gene identifier.

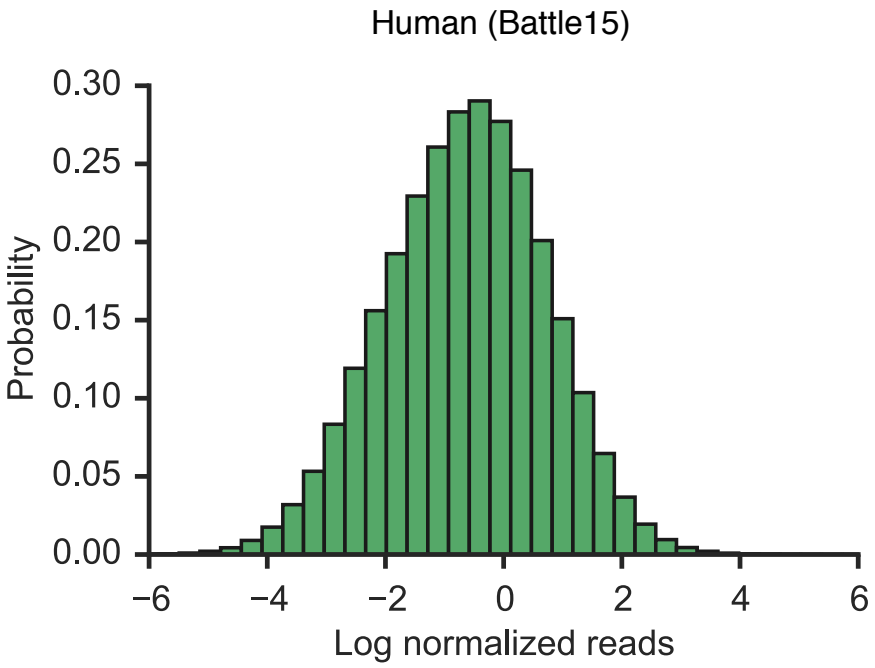
Supplementary Table 3. The enriched GO terms for genes in ramp+ and ramp- in human and yeast. The enriched GO terms for genes with high translation efficiency were also provided.

Supplementary Table 4. The hyperparameter values of ROSE calibrated using our antomatic one-way model selection strategy for both human (Battle15) and yeast (Pop14) datasets.

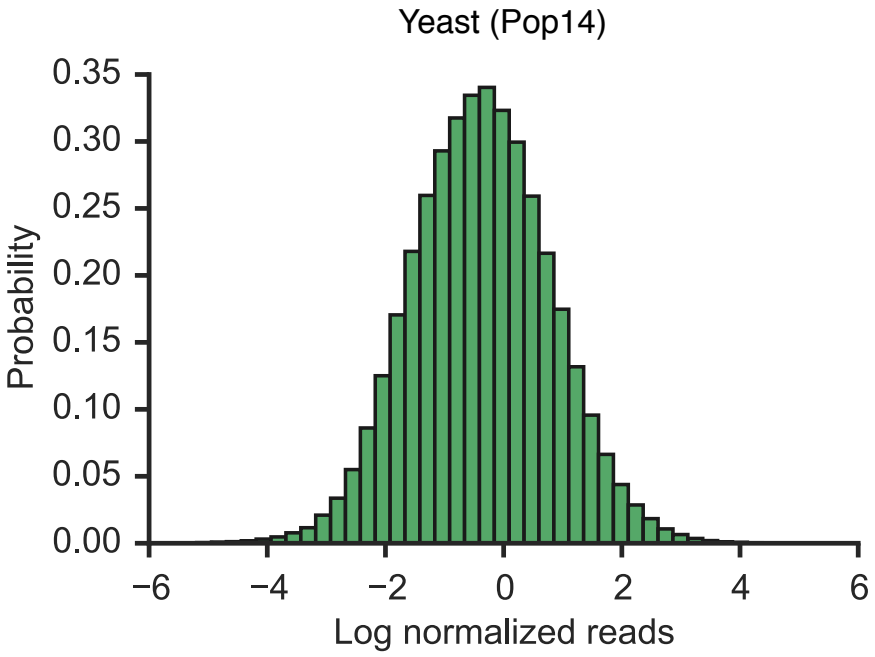
3 Supplementary Figures



Supplementary Figure 1: A schematic illustration of the translation elongation process. During translation elongation, the ribosome travels along the mRNA and gradually grows the nascent peptide chain, in which each codon along the mRNA is read and an aminoacyl tRNA is brought into the ribosome A-site to match the corresponding codon. Next, the bond between the peptide and the aminoacyl tRNA at the ribosome P-site is broken, and a new bond between this peptide and the amino acid that is just introduced at the ribosome A-site is formed. Then the ribosome moves forward to the next codon, while the uncharged tRNA is released from the ribosome E-site. In general, a ribosome covers about 27 nts (i.e., nine codons) of an mRNA [36]. Here, the position of the codon at the ribosome A-site is indexed as zero.

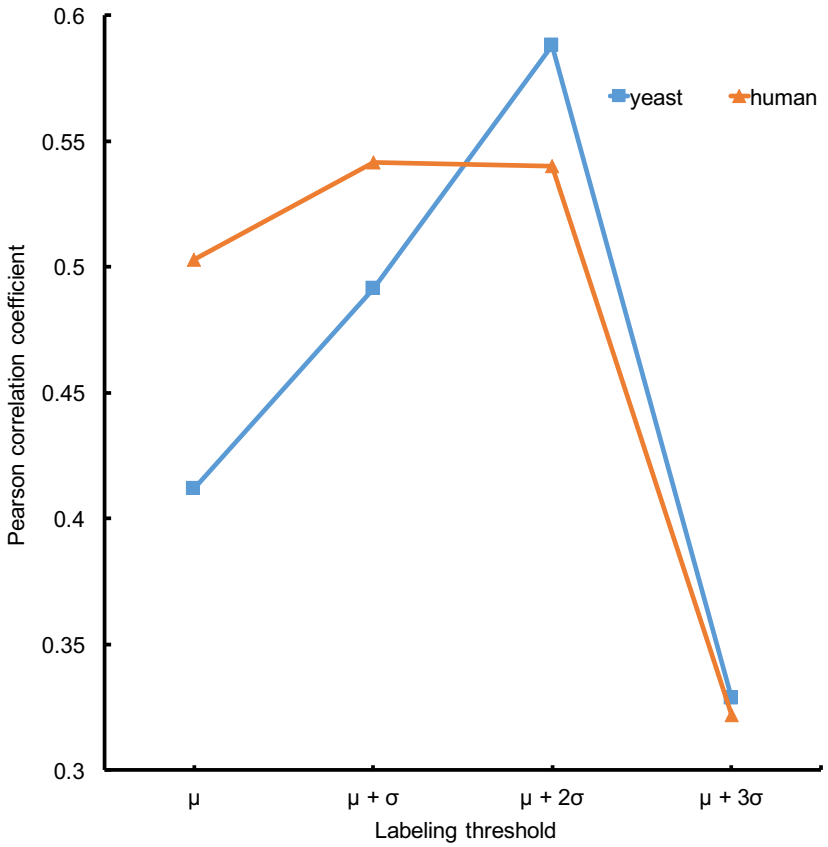


(a)

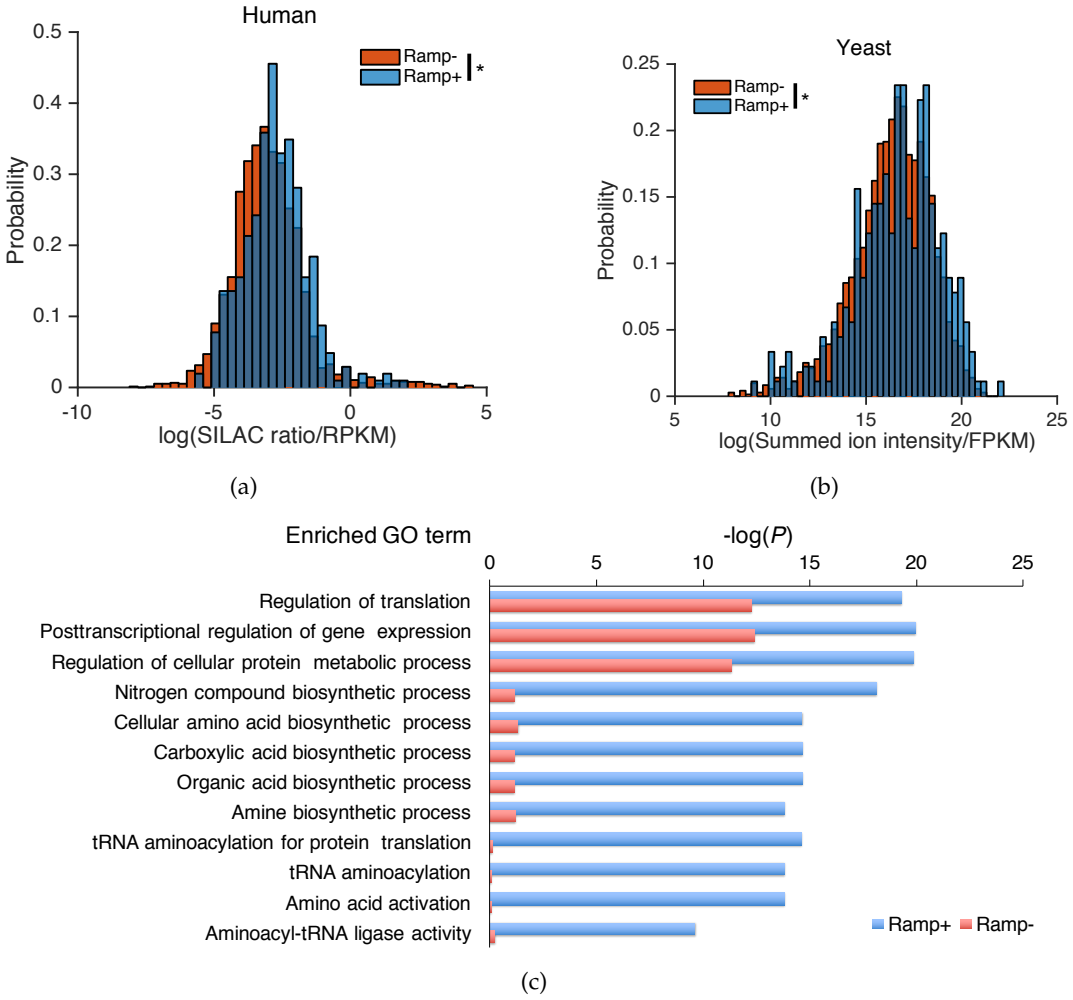


(b)

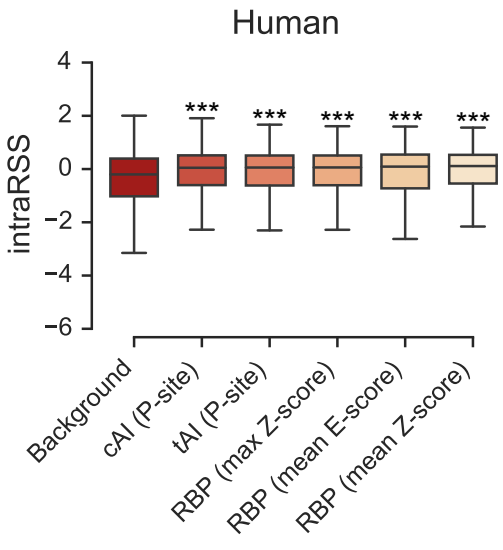
Supplementary Figure 2: The distributions of the normalized ribosome footprint density for (a) human (Battle15) and (b) yeast (Pop14) datasets.



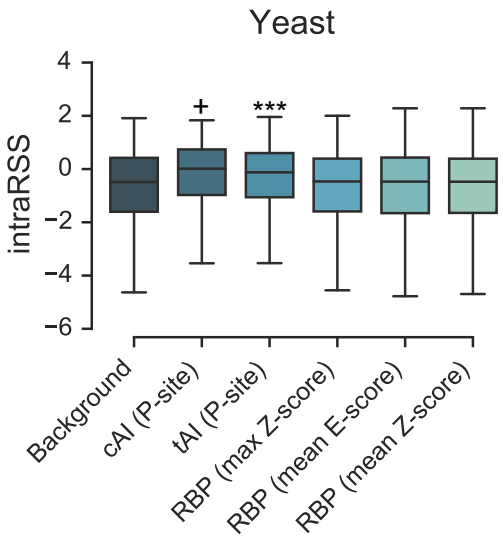
Supplementary Figure 3: Comparison of different thresholds for labeling samples to train the CNN classifier. The mean and the standard deviation of the ribosome footprint density distribution are denoted by μ and σ, respectively. Pearson correlation coefficient between the normalized footprint densities and the predicted ribosome stalling probabilities of samples in a separate validation dataset was calculated for each choice of threshold. More details can be found in Methods.



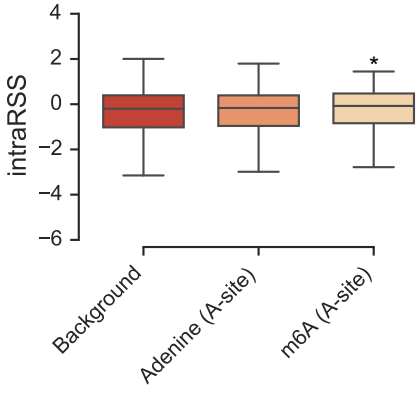
Supplementary Figure 4: The intergenic RSS landscape reveals the difference of translation efficiency and enriched functional categories between genes with high ribosome stalling probabilities in their ramp sequences and the background (see Supplementary Notes for more details). (a) The translation efficiency distributions of genes in ramp+ and ramp- in human. The mRNA levels measured by RNA-Seq were quantified by reads per kilobase transcript per million mapped reads (RPKM), while the protein levels measured by mass spectrometry were quantified by the SILAC ratio [4]. (b) The translation efficiency distributions of genes in ramp+ and ramp- in yeast. The mRNA levels measured by RNA-Seq were quantified by fragments per kilobase of transcript per million mapped reads (FPKM) [6], while the protein levels measured by mass spectrometry were quantified by the summed ion intensity [5]. The translation efficiency was estimated by the logarithm of the protein expression level divided by the corresponding mRNA expression level. (c) The comparison of the enriched GO terms between genes in ramp+ and ramp- in yeast. GO enrichment analysis was performed using DAVID [37, 38]. Here, the P values were computed after multiple testing correction according to the Benjamini-Hochberg procedure. For a full list of the functional annotation clustering, see Supplementary Table 3. *: $5 \times 10^{-25} < P < 5 \times 10^{-2}$; one-sided Wilcoxon rank-sum test.


Human
intraRSS
4
2
0
−2
−4
−6
***
***
***
***
***
Background
cAI (P-site)
tAI (P-site)
RBP (max Z-score)
RBP (mean E-score)
RBP (mean Z-score)

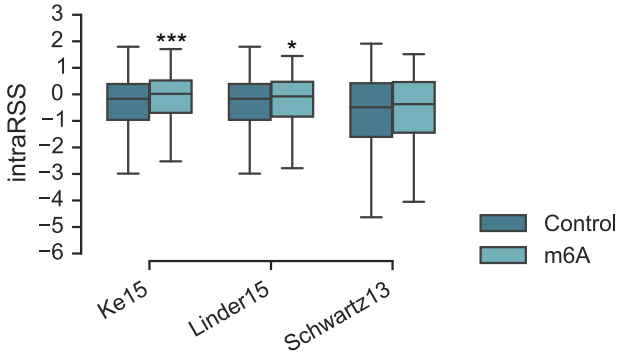
(a)


Yeast
intraRSS
4
2
0
−2
−4
−6
+
***
Background
cAI (P-site)
tAI (P-site)
RBP (max Z-score)
RBP (mean E-score)
RBP (mean Z-score)

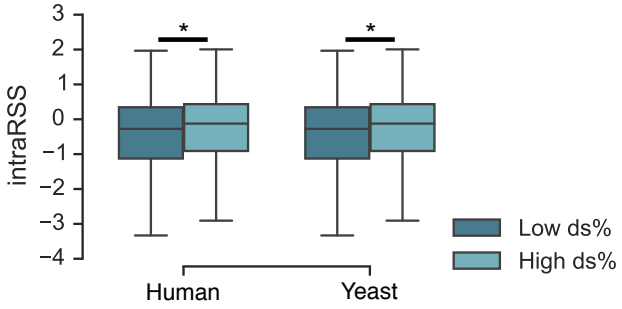
(b)


intraRSS
4
2
0
−2
−4
−6
*
Background
Adenine (A-site)
m6A (A-site)

(c)

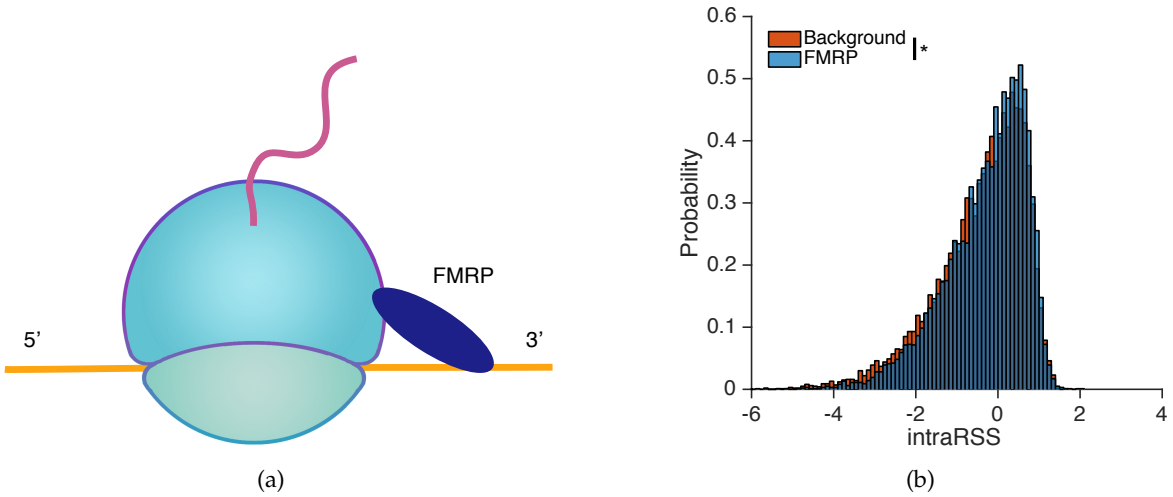

intraRSS
3
2
1
0
−1
−2
−3
−4
−5
−6
***
*
Ke15
Linder15
Schwartz13
Control
m6A

(d)


intraRSS
3
2
1
0
−1
−2
−3
−4
*
*
Human
Yeast
Low ds%
High ds%

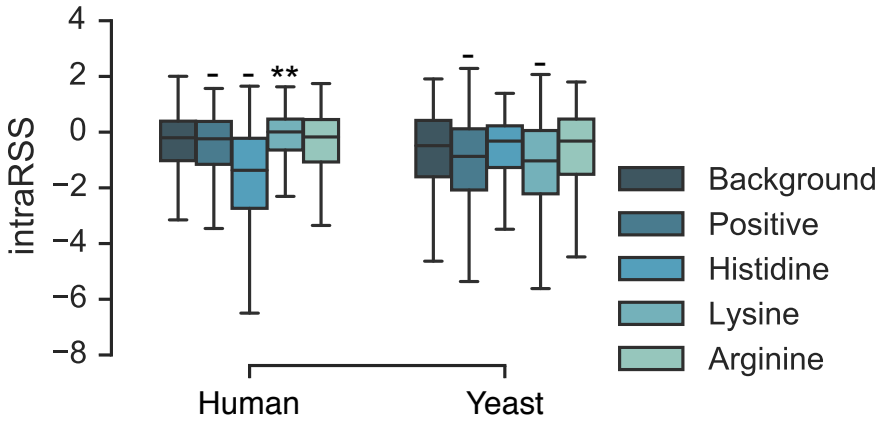
(e)

Supplementary Figure 5: A comprehensive reexamination on the relations between diverse putative regulatory factors and ribosome stalling using ROSE (supplementary to Fig. 3 in the main text). (a) and (b) The comparisons of intraRSS between the codon sites enriched with individual factors and the background for human and yeast, respectively. (c) The comparisons of intraRSS of the background vs. the codon sites with m^6A modification derived from [39] as well as a control dataset, which consisted of 10,000 randomly-selected codon sites containing adenine but without m^6A modification. (d) The comparisons of intraRSS between the control datasets and the codon sites with m^6A modification derived from different sources, including the Ke15 [40] and Linder15 [39] datasets of human, and the Schwartz13 [41] dataset of yeast. (e) The comparisons of intraRSS between highly and weakly double-stranded regions for both human and yeast. *: $5 \times 10^{-25} < P < 1 \times 10^{-2}$; **: $5 \times 10^{-50} < P \leq 5 \times 10^{-25}$; ***: $5 \times 10^{-100} < P \leq 5 \times 10^{-50}$; +: $P \leq 5 \times 10^{-100}$; one-sided Wilcoxon rank-sum test.

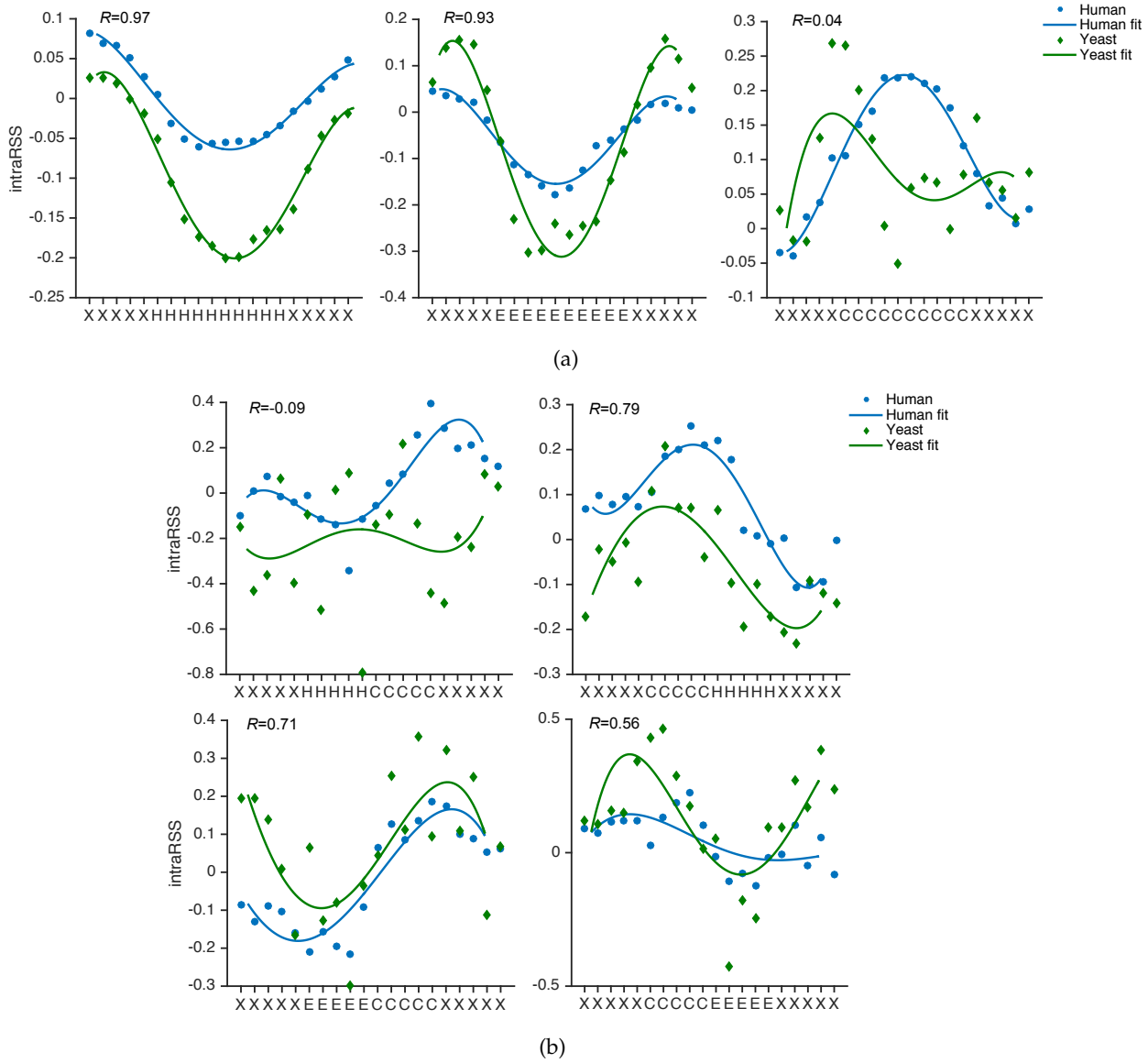


(a) (b)

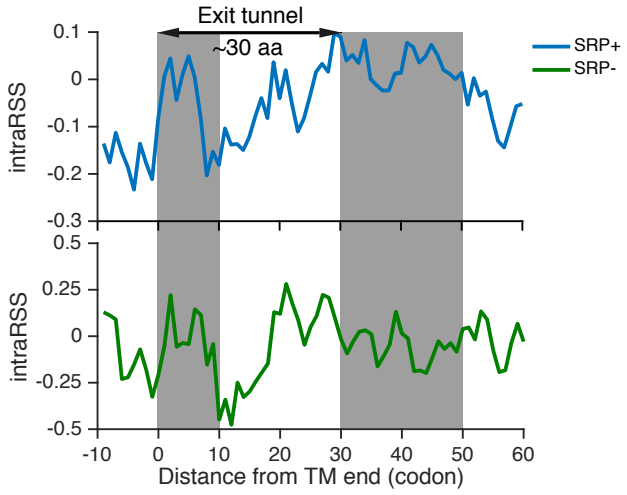
Supplementary Figure 6: A comprehensive reexamination on the relations between diverse putative regulatory factors and ribosome stalling using ROSE (supplementary to Supplementary Fig. 5 and Fig. 3 in the main text). (a) A schematic illustration of the FMRP binding to impede ribosome movement. (b) The comparison of intraRSS between the FMRP target regions and the background. *: $5 \times 10^{-25} < P < 1 \times 10^{-2}$; one-sided Wilcoxon rank-sum test.



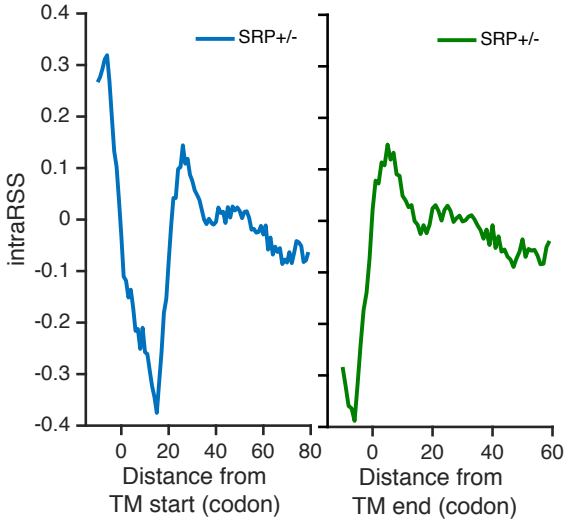
Supplementary Figure 7: The comparisons of intraRSS between the background and the codon sites enriched with the positively-charged amino acids in the upstream 30 codons for both human and yeast. Here, we considered both general positively-charged amino acids (which include histidine, lysine and arginine) and particular amino acids. "-" means that the difference of intraRSS between the codon sites of interest and the background was significant ($P < 1 \times 10^{-2}$) but contrary to our expectation (i.e., a higher ratio of positively-charged amino acids should yield a larger intraRSS). **: $5 \times 10^{-50} < P \leq 5 \times 10^{-25}$; one-sided Wilcoxon rank-sum test.



Supplementary Figure 8: The intragenic RSS landscapes of different protein secondary structure patterns with window size ten (supplementary to Fig. 4 in the main text). (a) The intraRSS landscapes of the alpha helix, beta strand and random coil regions. (b) The intraRSS landscapes of the SSE transition regions. "H", "E" and "C" stand for alpha helix, beta strand and random coil, respectively, while "X" stands for any SSE type in the flanking regions on both sides. Polynomial curve fitting of degree four was used to show the general intraRSS tendency. The Spearman correlation coefficients between human and yeast intraRSS tendencies were calculated.

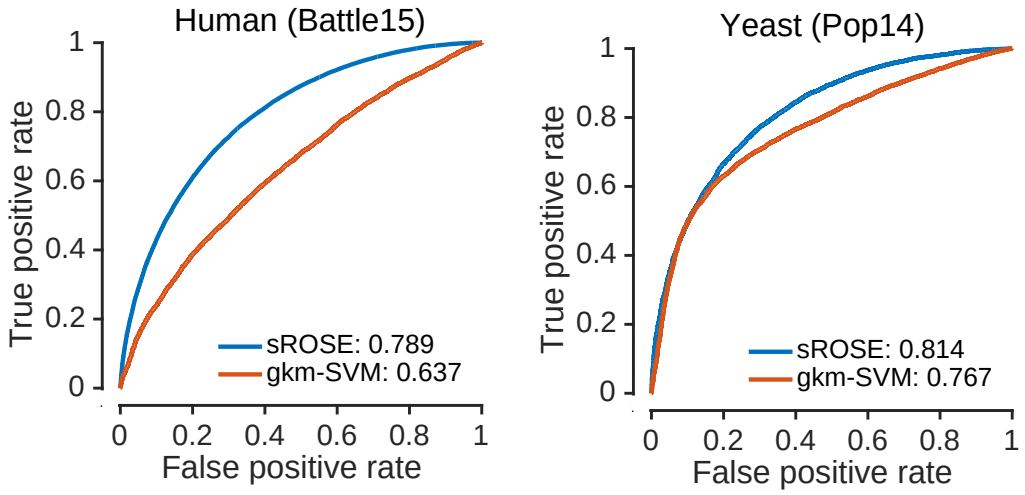


(a)



(b)

Supplementary Figure 9: The intragenic RSS landscape recovers that the ribosome stalling tendency associates with the SRP binding of TM segments (supplementary to Fig. 5 in the main text). (a) The comparison of the intraRSS tendency between the TM segments with and without SRP binding in yeast, in which all the protein sequences were aligned with regard to the end of the TM segment, which was indexed as zero. For those TM segments without SRP binding (termed SRP-), the peak in positions 30–50 downstream the TM segment, corresponding to the peak in positions 50–70 in Fig. 5(b) of the main text, was significantly diminished compared to that of the TM segments with SRP binding (termed SRP+; $P = 9.8 \times 10^{-5}$ by one-sided Wilcoxon rank-sum test). The grey rectangles represent two intraRSS peaks downstream the TM segment. (b) The intraRSS tendency of the TM segments in human, in which either the TM start or the TM end position was used as the initial alignment location. A mixed set of TM segments with and without SRP binding (termed SRP+/-) was considered.



(a) (b)

Supplementary Figure 10: The comparison of the three-fold CV performance between ROSE and gkm-SVM. (a) and (b) The ROC curves and the corresponding AUROC scores of CV on the human (Battle15) and yeast (Pop14) training data, respectively. "sROSE" stands for the single version of ROSE (i.e., with one CNN model).

References

[1] T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, and Y. Pilpel, "An evolutionarily conserved mechanism for controlling the efficiency of protein translation," *Cell*, vol. 141, no. 2, pp. 344–354, 2010.

[2] T. Quax, N. Claassens, D. Söll, and J. van der Oost, "Codon bias as a means to fine-tune gene expression," *Molecular Cell*, vol. 59, no. 2, pp. 149–161, 2015.

[3] T. Tuller and H. Zur, "Multiple roles of the coding sequence 5′ end in gene expression regulation," *Nucleic Acids Research*, 2014.

[4] A. Battle, Z. Khan, S. H. Wang, A. Mitrano, M. J. Ford, J. K. Pritchard, and Y. Gilad, "Impact of regulatory variation from RNA to protein," *Science*, vol. 347, no. 6222, pp. 664–667, 2015.

[5] L. M. F. de Godoy, J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Frohlich, T. C. Walther, and M. Mann, "Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast," *Nature*, vol. 455, pp. 1251–1254, Oct. 2008.

[6] I. Nookaew, M. Papini, N. Pornputtpong, G. Scalcinati, L. Fagerberg, M. Uhlén, and J. Nielsen, "A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: A case study in *Saccharomyces cerevisiae*," *Nucleic Acids Research*, 2012.

[7] E. P. Rocha, "Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization," *Genome Research*, vol. 14, no. 11, pp. 2279–2286, 2004.

[8] A. Dana and T. Tuller, "The effect of tRNA levels on decoding times of mRNA codons," *Nucleic Acids Research*, 2014.

[9] C. Pop, S. Rouskin, N. T. Ingolia, L. Han, E. M. Phizicky, J. S. Weissman, and D. Koller, "Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation," *Molecular Systems Biology*, vol. 10, no. 12, 2014.

[10] J. Gardin, R. Yeasmin, A. Yurovsky, Y. Cai, S. Skiena, and B. Futcher, "Measurement of average decoding rates of the 61 sense codons *in vivo*," *eLife*, vol. 3, p. e03735, Oct. 2014.

[11] N. T. Ingolia, "Ribosome profiling: New views of translation, from single codons to genome scale," *Nat Rev Genet*, vol. 15, pp. 205–213, Mar. 2014.

[12] J. A. Hussmann, S. Patchett, A. Johnson, S. Sawyer, and W. H. Press, "Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast," *PLoS Genet*, vol. 11, pp. 1–25, Dec. 2015.

[13] J. L. Chaney and P. L. Clark, "Roles for synonymous codon usage in protein biogenesis," *Annual Review of Biophysics*, vol. 44, no. 1, pp. 143–166, 2015.

[14] M. d. Reis, R. Savva, and L. Wernisch, "Solving the riddle of codon usage preferences: A test for translational selection," *Nucleic Acids Research*, vol. 32, no. 17, pp. 5036–5044, 2004.

[15] T. Ikemura, "Codon usage and tRNA content in unicellular and multicellular organisms," *Molecular Biology and Evolution*, vol. 2, no. 1, pp. 13–34, 1985.

[16] E. M. Novoa and L. Ribas de Pouplana, "Speeding with control: Codon usage, tRNAs, and ribosomes," *Trends in Genetics*, vol. 28, pp. 574–581, Nov. 2012.

[17] H. Dong, L. Nilsson, and C. G. Kurland, "Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates," *Journal of Molecular Biology*, vol. 260, pp. 649–663, Aug. 1996.

[18] C. Chen, H. Zhang, S. L. Broitman, M. Reiche, I. Farrell, B. S. Cooperman, and Y. E. Goldman, "Dynamics of translation by single ribosomes through mRNA secondary structures," *Nat Struct Mol Biol*, vol. 20, pp. 582–588, May 2013.

[19] J.-D. Wen, L. Lancaster, C. Hodges, A.-C. Zeri, S. H. Yoshimura, H. F. Noller, C. Bustamante, and I. Tinoco, "Following translation by single ribosomes one codon at a time," *Nature*, vol. 452, pp. 598–603, Apr. 2008.

[20] T. E. Gorochowski, Z. Ignatova, R. A. Bovenberg, and J. A. Roubos, "Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate," *Nucleic Acids Research*, 2015.

[21] J.-R. Yang, X. Chen, and J. Zhang, "Codon-by-codon modulation of translational speed and accuracy via mRNA folding," *PLoS Biol*, vol. 12, p. e1001910, July 2014.

[22] T. Tuller, I. Veksler-Lublinsky, N. Gazit, M. Kupiec, E. Ruppin, and M. Ziv-Ukelson, "Composite effects of gene determinants on the translation speed and density of ribosomes," *Genome Biology*, vol. 12, no. 11, pp. 1–18, 2011.

[23] C. A. Charneski and L. D. Hurst, "Positively charged residues are the major determinants of ribosomal velocity," *PLoS Biol*, vol. 11, pp. 1–20, Mar. 2013.

[24] T. Tuller, Y. Y. Waldman, M. Kupiec, and E. Ruppin, "Translation efficiency is determined by both codon bias and folding energy," *Proceedings of the National Academy of Sciences*, vol. 107, no. 8, pp. 3645–3650, 2010.

[25] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, "ViennaRNA package 2.0," *Algorithms for Molecular Biology*, vol. 6, no. 1, pp. 1–14, 2011.

[26] S. Gerstberger, M. Hafner, and T. Tuschl, "A census of human RNA-binding proteins," *Nature Reviews Genetics*, vol. 15, pp. 829–845, Nov. 2014.

[27] D. Ray, H. Kazan, K. Cook, M. Weirauch, H. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. Matzat, R. Dale, S. Smith, C. Yarosh, S. Kelly, B. Nabet, D. Mecenas, W. Li, R. Laishram, M. Qiao, H. Lipshitz, F. Piano, A. Corbett, R. Carstens, B. Frey, R. Anderson, K. Lynch, L. Penalva, E. Lei, A. Fraser, B. Blencowe, Q. Morris, and T. Hughes, "A compendium of RNA-binding motifs for decoding gene regulation," *Nature*, vol. 499, pp. 172–177, July 2013.

[28] A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B. Beckmann, C. Strein, N. Davey, D. Humphreys, T. Preiss, L. Steinmetz, J. Krijgsveld, and M. Hentze, "Insights into RNA biology from an atlas of mammalian mRNA-binding proteins," *Cell*, vol. 149, no. 6, pp. 1393–1406, 2012.

[29] A. Baltz, M. Munschauer, B. Schwanhäusser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, and M. Landthaler, "The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts," *Molecular Cell*, vol. 46, no. 5, pp. 674–690, 2012.

[30] M. Ascano, N. Mukherjee, P. Bandaru, J. B. Miller, J. D. Nusbaum, D. L. Corcoran, C. Langlois, M. Munschauer, S. Dewell, M. Hafner, Z. Williams, U. Ohler, and T. Tuschl, "FMRP targets distinct mRNA sequence elements to regulate protein expression," *Nature*, vol. 492, pp. 382–386, Dec. 2012.

[31] J. C. Darnell, S. J. Van Driesche, C. Zhang, K. Y. S. Y. Hung, A. Mele, C. E. Fraser, E. F. Stone, C. Chen, J. J. Fak, S. W. W. Chi, D. D. Licatalosi, J. D. Richter, and R. B. Darnell, "FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism," *Cell*, vol. 146, pp. 247–261, July 2011.

[32] E. Chen, M. R. Sharma, X. Shi, R. K. Agrawal, and S. Joseph, "Fragile X mental retardation protein regulates translation by binding directly to the ribosome," *Molecular cell*, vol. 54, pp. 407–417, May 2014.

[33] J. Lu and C. Deutsch, "Electrostatics in the ribosomal tunnel modulate chain elongation rates," *Journal of Molecular Biology*, vol. 384, pp. 73–86, Dec. 2008.

[34] J. Lu, W. R. Kobertz, and C. Deutsch, "Mapping the electrostatic potential within the ribosomal exit tunnel," *Journal of Molecular Biology*, vol. 371, pp. 1378–1391, Aug. 2007.

[35] C. G. Artieri and H. B. Fraser, "Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation," *Genome Research*, 2014.

[36] N. T. Ingolia, "Chapter 6 – Genome-wide translational profiling by ribosome footprinting," in *Guide to Yeast Genetics: Functional Genomics, Proteomics, and Other Systems Analysis*, vol. Volume 470, pp. 119–142, Academic Press, 2010.

[37] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Research*, vol. 37, no. 1, pp. 1–13, 2009.

[38] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat. Protocols*, vol. 4, pp. 44–57, Dec. 2008.

[39] B. Linder, A. V. Grozhik, A. O. Olarerin-George, C. Meydan, C. E. Mason, and S. R. Jaffrey, "Single-nucleotide-resolution mapping of m^6A and m^6Am throughout the transcriptome," *Nat Meth*, vol. 12, pp. 767–772, Aug. 2015.

[40] S. Ke, E. A. Alemu, C. Mertens, E. C. Gantman, J. J. Fak, A. Mele, B. Haripal, I. Zucker-Scharff, M. J. Moore, C. Y. Park, C. B. Vågbø, A. Kusśnierczyk, A. Klungland, J. E. Darnell, and R. B. Darnell, "A majority of m^6A residues are in the last exons, allowing the potential for 3' UTR regulation," *Genes & Development*, vol. 29, no. 19, pp. 2037–2053, 2015.

[41] S. Schwartz, S. Agarwala, M. Mumbach, M. Jovanovic, P. Mertins, A. Shishkin, Y. Tabach, T. Mikkelsen, R. Satija, G. Ruvkun, S. Carr, E. Lander, G. Fink, and A. Regev, "High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis," *Cell*, vol. 155, pp. 1409–1421, Dec. 2013.