Supplementary Materials for

Modeling translation elongation dynamics by deep learning reveals new insights into the landscape of ribosome stalling

Sai Zhang[1,†], Hailin Hu[2,†], Jingtian Zhou[2,†], Xuan He[1], and Jianyang Zeng[1,*]

August 1, 2016

Supplementary Tables

Supplementary Table 1. Spearman correlation coefficients between different factors affecting ribosome stalling for both human and yeast datasets.

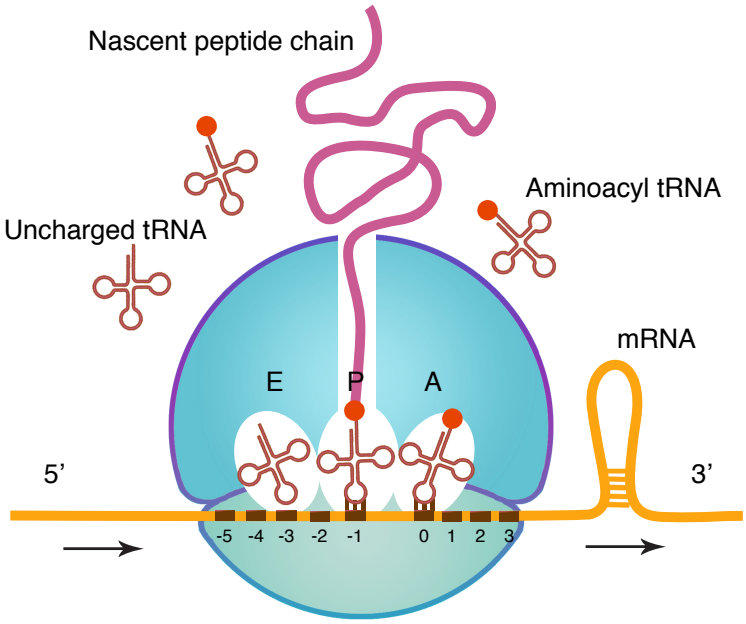
Supplementary Table 2. All expressed genes in LCLs and *S. cerevisiae* that were used as the background for GO analysis for both human and yeast. The Ensembl gene ID was used as the gene identifier.

Supplementary Table 3. The enriched GO terms for both ramp and non-ramp genes in human and yeast. The enriched GO terms for those genes with high translation efficiency are also provided.

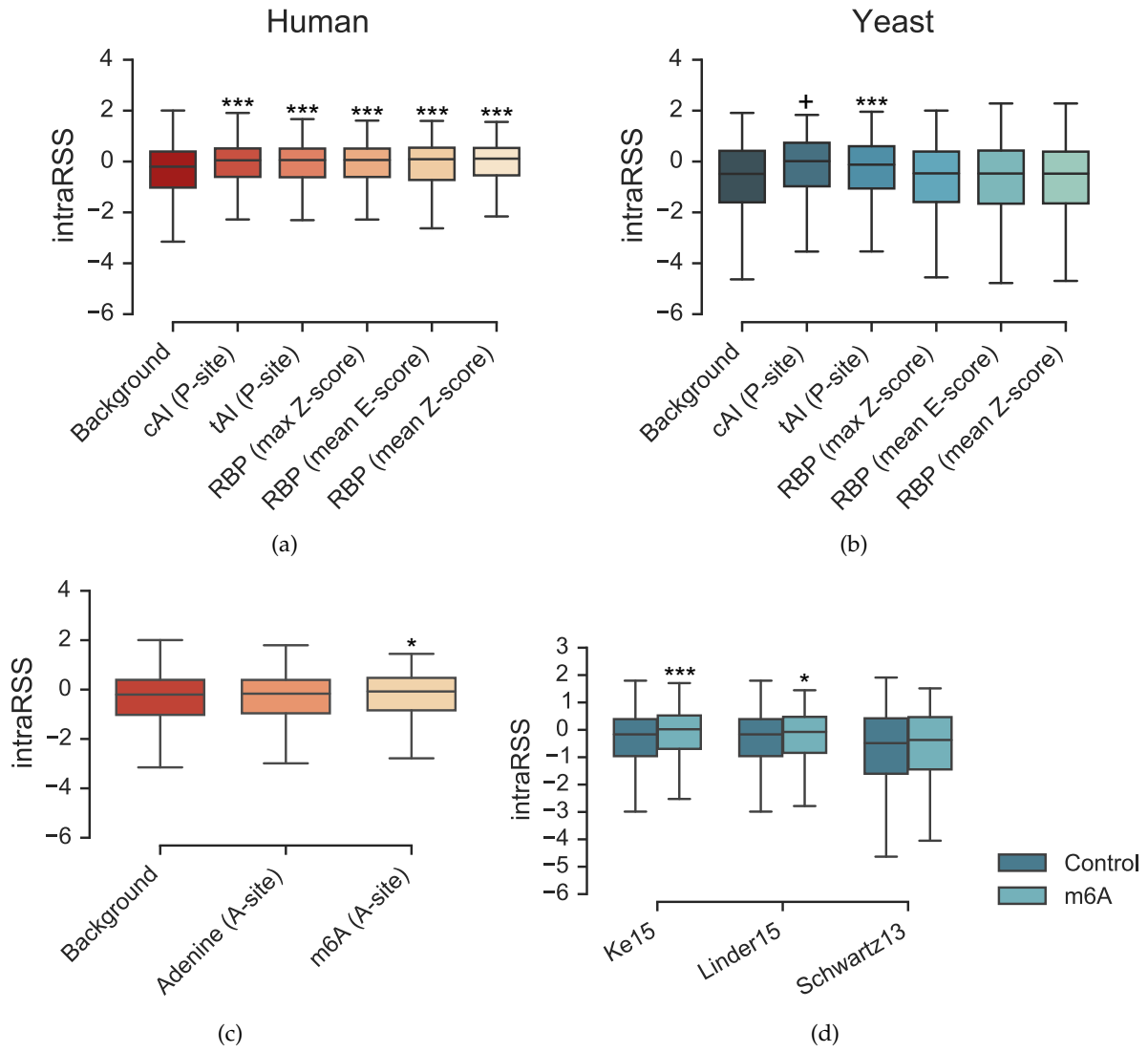
Supplementary Table 4. The hyperparameter values of ROSE calibrated using our antomatic one-way model selection strategy for both human (Battle15) and yeast (Pop14) datasets.

[1] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China.
[2] School of Pharmaceutical Sciences, Tsinghua University, Beijing, China.
[†] These authors contributed equally to this work.
[*] To whom correspondence should be addressed. Email: zengjy321@tsinghua.edu.cn.

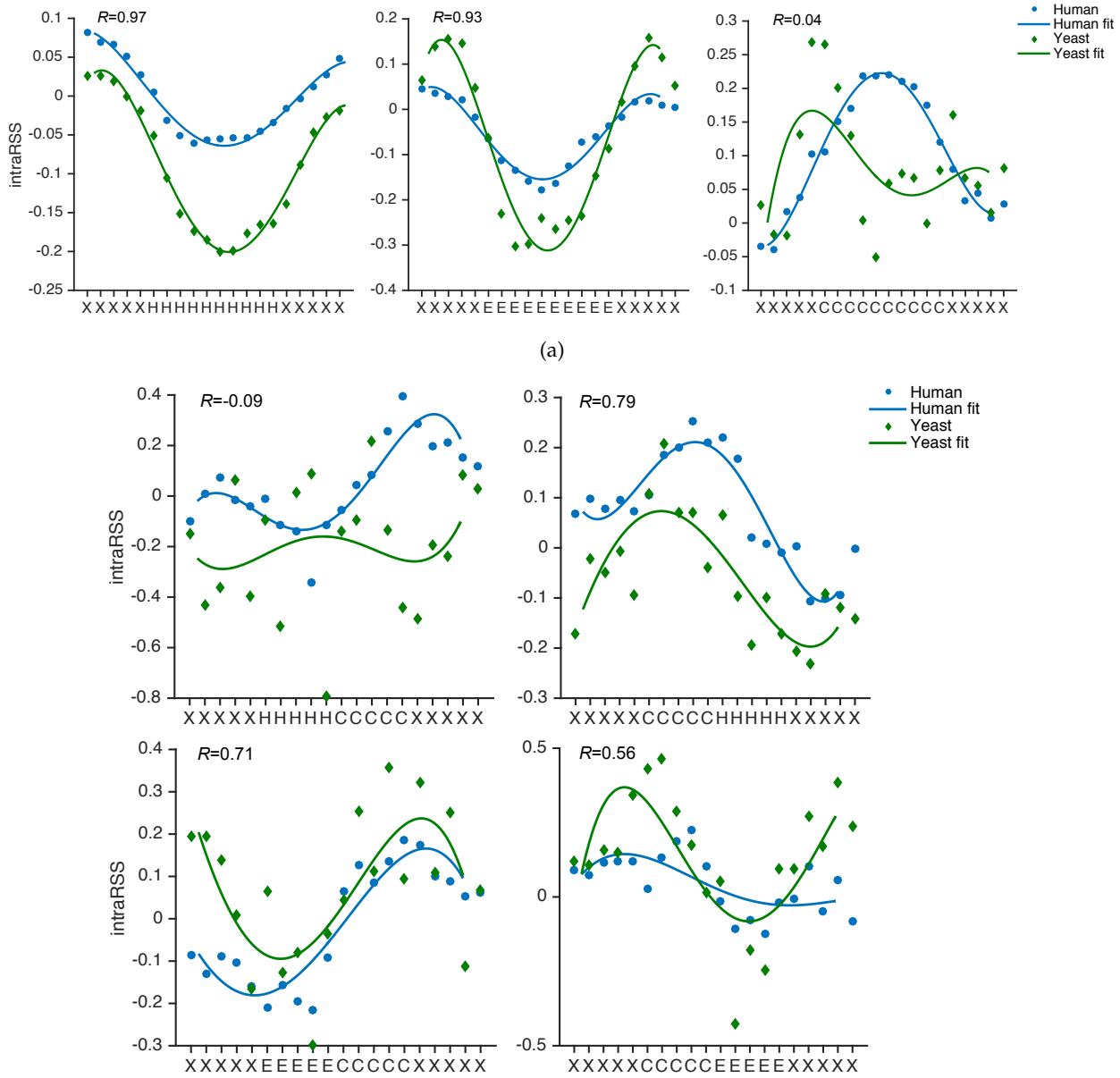
Supplementary Figures



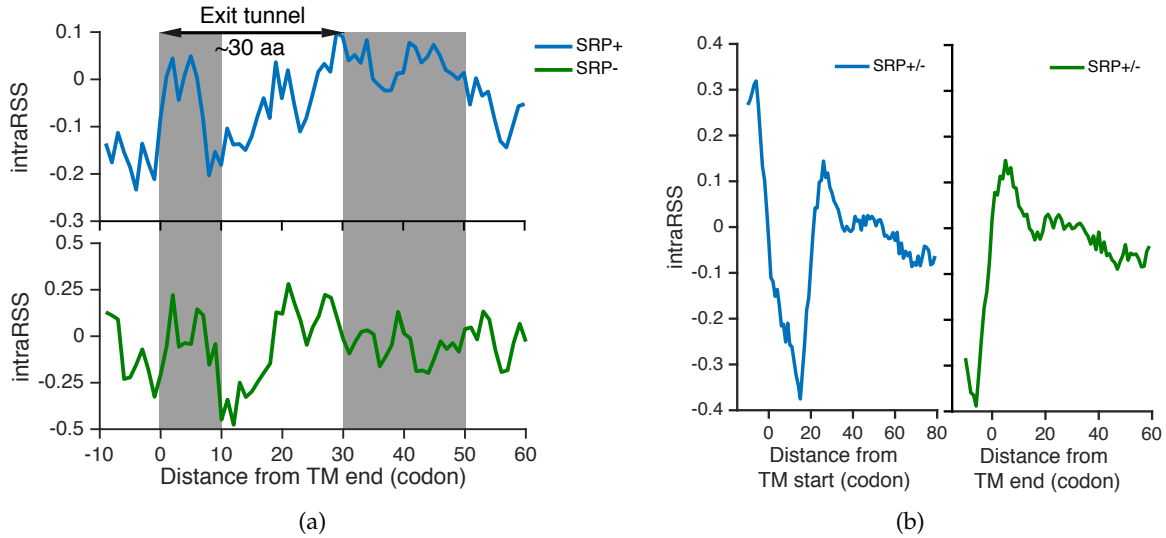
Supplementary Figure 1: A schematic illustration of the translation elongation process. During translation elongation, the ribosome travels along the mRNA and gradually grows the nascent peptide chain, in which each codon along the mRNA is read and an aminoacyl tRNA is brought into the ribosome A-site 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 nt (i.e., nine codons) of an mRNA [1]. Here, the position of the codon at the ribosome A-site is indexed as zero.



Supplementary Figure 2: A comprehensive investigation on the regulatory effects of different factors on 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 controlling 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 [2] 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 [3] and Linder15 [2] datasets of human, and the Schwartz13 [4] dataset of 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.

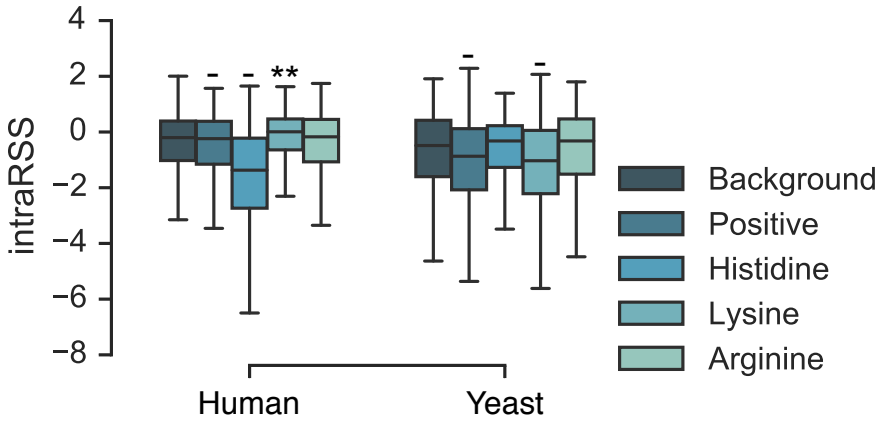


Supplementary Figure 3: The intraRSS landscapes of different protein secondary structure patterns with window size ten (supplementary to Fig. 4 in the main text). (a) The intraRSS landscapes of the alpha helix, beta strand and random coil regions. (b) The intraRSS landscapes of the SSE transition regions. "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.

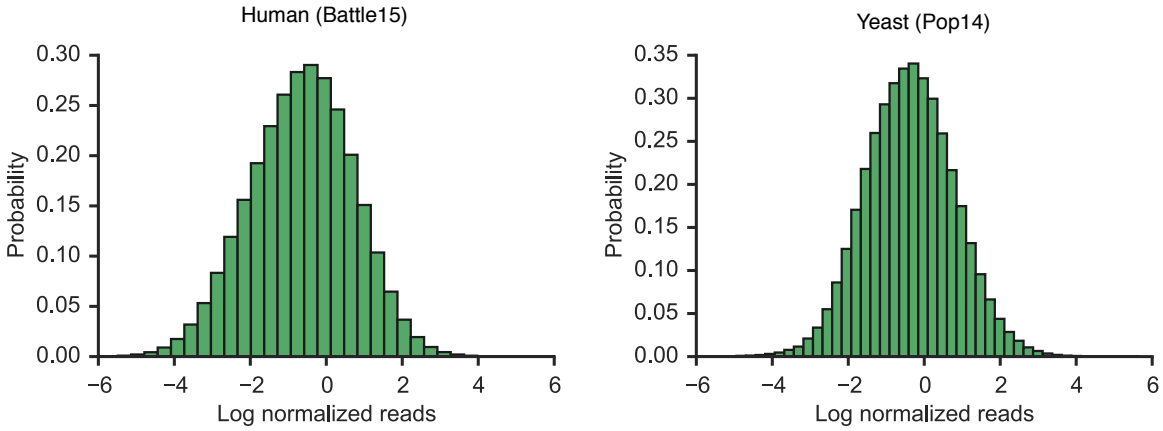


(a) (b)

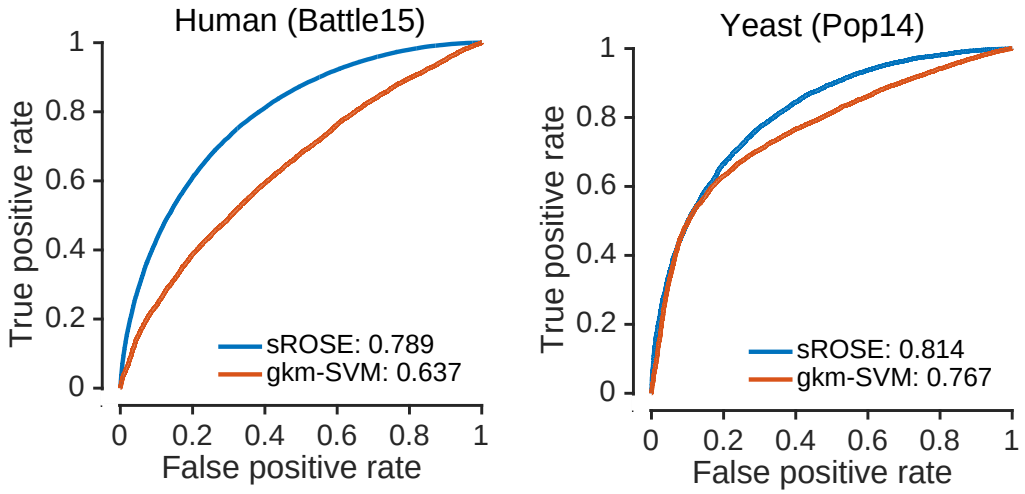
Supplementary Figure 4: The intragenic RSS landscape reveals that translation elongation dynamics promotes SRP binding in the TM segments (supplementary to Fig. 5 in the main text). (a) The comparison of the intraRSS tendency between the TM segments with and without SRP binding in yeast, in which all 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. 5b 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 regarded as the initial alignment location. A mixed set of both TM segments with and without SRP binding (termed SRP+/-) was considered.



Supplementary Figure 5: The comparisons of intraRSS between the background and those 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 6: The distributions of the normalized ribosome profiling reads for human (a) and yeast (b) datasets.



(a) (b)

Supplementary Figure 7: The comparisons 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] 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.

[2] 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.

[3] 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.

[4] 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.