

ROSE: a deep learning based framework for predicting ribosome stalling

Sai Zhang^{1,†}, Hailin Hu^{2,†}, Jingtian Zhou^{2,†}, Xuan He¹,
Tao Jiang^{3,4,5} and Jianyang Zeng¹

¹ Institute for Interdisciplinary Information Sciences, Tsinghua University,
Beijing, China.

zengjy321@tsinghua.edu.cn

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.

Abstract. Translation elongation plays a crucial role in multiple aspects of protein biogenesis, e.g., differential expression, cotranslational folding and secretion. However, our current understanding on the regulatory mechanisms underlying translation elongation dynamics and the functional roles of ribosome stalling in protein synthesis still remains largely limited. Here, we present a deep learning based framework, called ROSE, to effectively predict ribosome stalling events in translation elongation from coding sequences. Our validation results on both human and yeast datasets demonstrate superior performance of ROSE over conventional prediction models. With high prediction accuracy and robustness across different datasets, ROSE shall provide an effective index to estimate the translational pause tendency at codon resolution. We also show that the ribosome stalling score (RSS) output by ROSE correlates with diverse putative regulatory factors of ribosome stalling, e.g., codon usage bias, codon cooccurrence bias, proline codons and N⁶-methyladenosine (m⁶A) modification, which validates the physiological relevance of our approach. In addition, our comprehensive genome-wide *in silico* studies of ribosome stalling based on ROSE recover several notable functional interplays between elongation dynamics and cotranslational events in protein biogenesis, including protein targeting by the signal recognition particle (SRP) and protein secondary structure formation. Furthermore, our intergenic analysis suggests that the enriched ribosome stalling events at the 5' ends of coding sequences may be involved in the modulation of translation efficiency. These findings indicate that ROSE can provide a useful index to estimate the probability of ribosome stalling and offer a powerful tool to analyze the large-scale ribosome profiling data, which will further expand our understanding on translation elongation dynamics. The full version of this work can be found as a preprint at <https://doi.org/10.1101/067108>.

Acknowledgements: This work was supported in part by the National Basic Research Program of China Grant 2011CBA00300 and 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61361136003 and 61472205, the US National Science Foundation Grant DBI-1262107 and IIS-1646333, the China's Youth 1000-Talent Program, and the Beijing Advanced Innovation Center for Structural Biology.