ROSE: a deep learning based framework for predicting ribosome stalling

Sai Zhang[1,6], Hailin Hu[2,6], Jingtian Zhou[2,6], Xuan He[1], Tao Jiang[3,4,5] and Jianyang Zeng[1]*

[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*

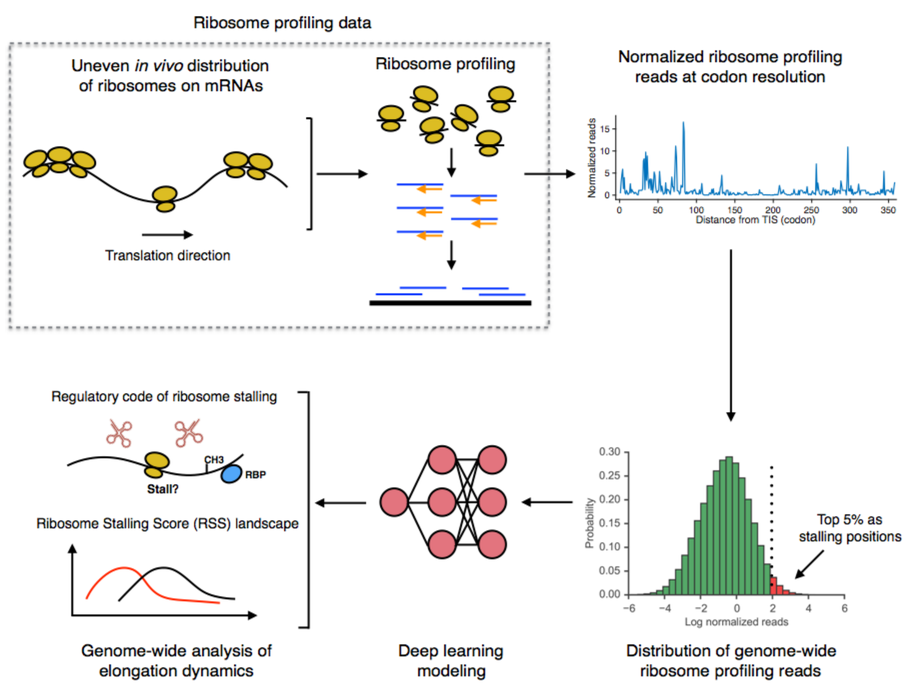
[5]*Institute of Integrative Genome Biology, University of California, Riverside, CA, USA*

[6]*These authors contributed equally to this work*

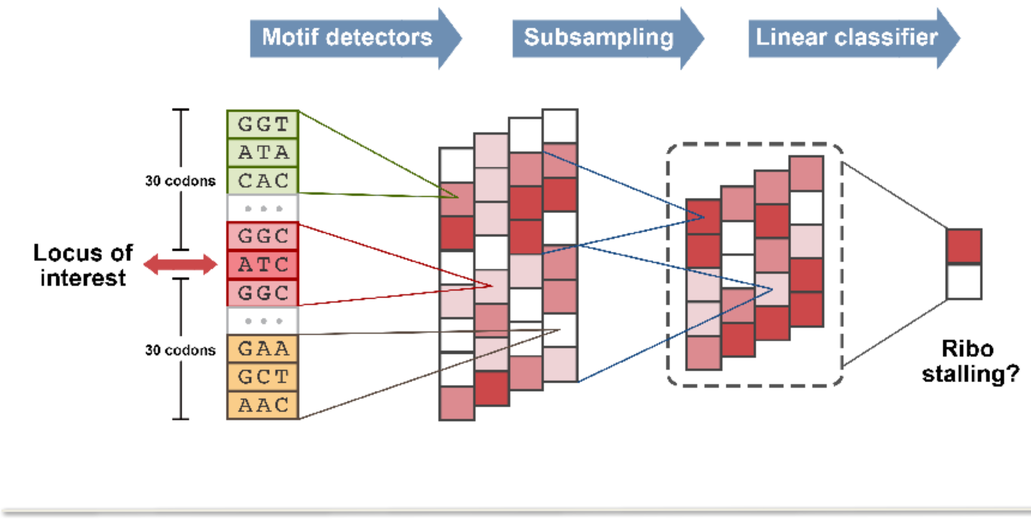
Abstract

We present a deep learning based framework, called ROSE, to accurately predict ribosome stalling events in translation elongation from coding sequences based on high-throughput ribosome profiling data. Our validation results demonstrate the superior performance of ROSE over conventional prediction models. ROSE provides an effective index to estimate the likelihood of translational pausing at codon resolution and understand diverse putative regulatory factors of ribosome stalling. Also, the ribosome stalling landscape computed by ROSE can recover the functional interplay between ribosome stalling and cotranslational events in protein biogenesis, including protein targeting by the signal recognition particle (SRP) and protein secondary structure formation.

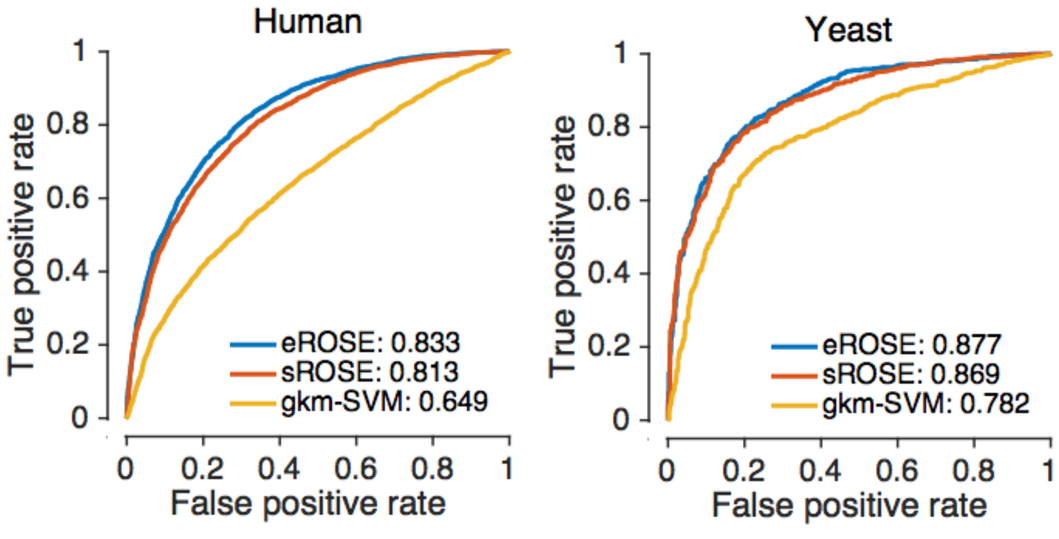
ROSE pipeline and CNN model



$$p(s) = \mathrm{sigm}(\mathrm{concat}_{i=1,2,3}(\mathrm{pool}^i(\mathrm{ReLU}^i(\mathrm{conv}^i(\mathrm{encode}(s))))))$$

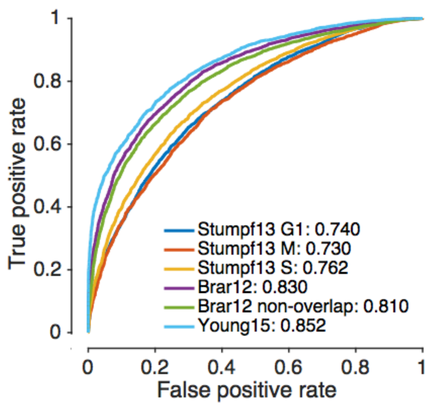


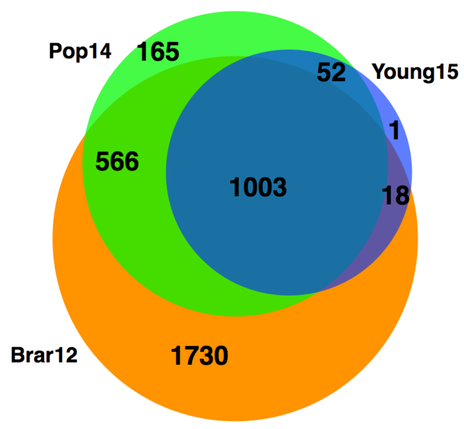
Predicting performance of ROSE



The receiver operating characteristic (ROC) curves and the area under the corresponding ROC curve (AUROC) scores on the human (Battle15) and yeast (Pop14) test datasets, respectively. "sROSE" and "eROSE" stand for the ROSE frameworks with one (single) and 64 (ensemble) CNNs, respectively.

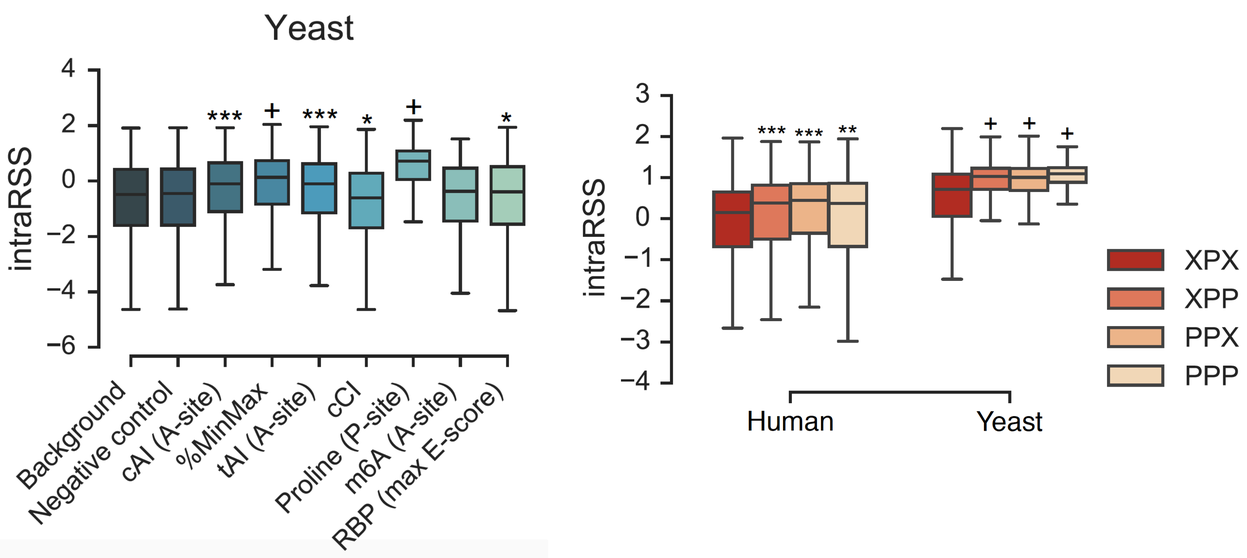
Robustness of ROSE





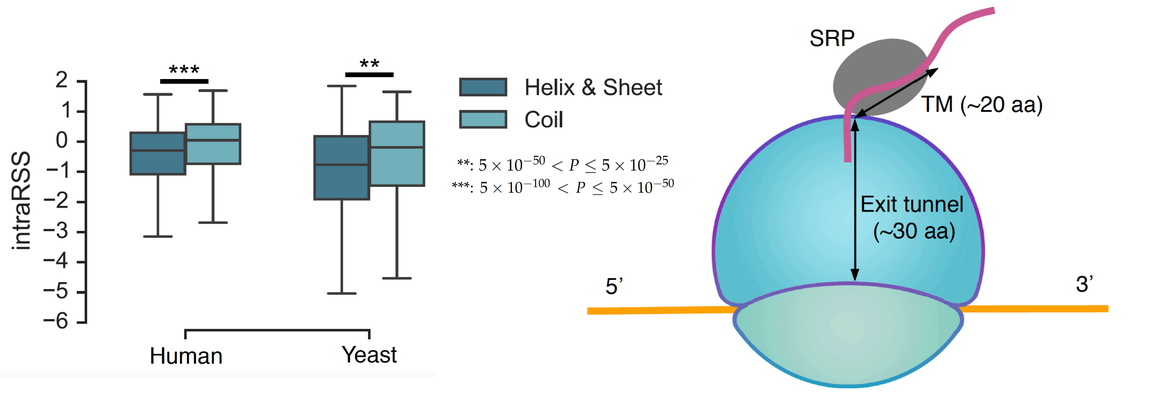
The ROC curves and AUROC scores of the cross-study tests on additional human (Stumpf13 G1, M and S) and yeast (Brar12 and Young15) datasets, respectively. The Brar12 non-overlapping dataset includes 1,748 genes with sufficient (over 60%) ribosome profiling coverage in the Brar12 dataset but not in the Pop14 dataset.

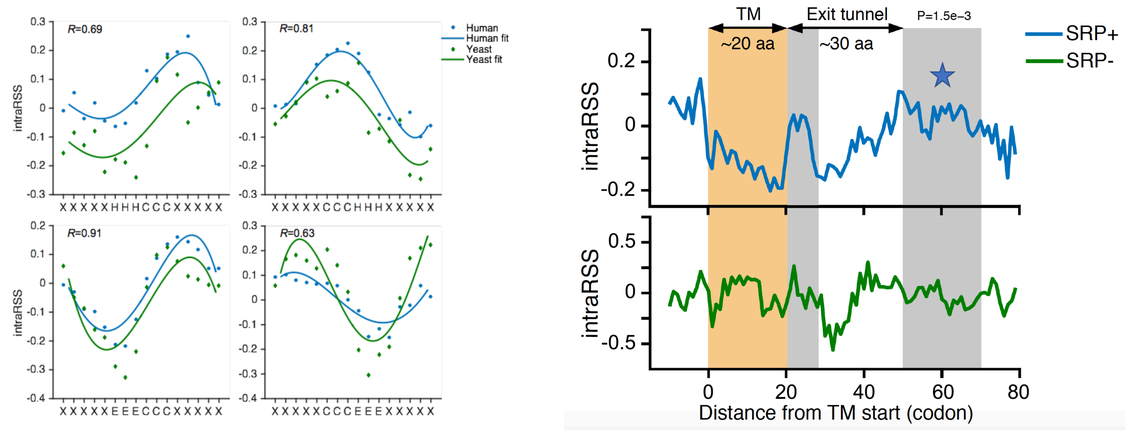
Regulatory factors of stalling



A comprehensive reexamination on the relations between diverse putative regulatory factors and ribosome stalling using ROSE.

Ribosome stalling Landscape





Ribosome stalling associates with protein secondary structure and the SRP binding of transmembrane segments.

Reference

[1] Zhang *et al.* Analysis of ribosome stalling and translation elongation dynamics by deep learning. *Cell Systems* (2017) In press. https://doi.org/10.1101/067108

This work was supported in part by the National Natural Science Foundation of China Grant 61472205, China's Youth 1000-Talent Program, and the Beijing Advanced Innovation Center for Structural Biology. The ISMB 2017 travel funding was provided by ISCB.