[BIOINFO 2025]

DuET: A Unified Deep Learning Framework for Predicting mRNA Translation Efficiency Across Human Cell Types

Sungho Lee[1], Jae-Won Lee[1], Jinhee Hong[1], Sangheon Lee[1,*], and Hyunjin Shin[1,*]

[1]*MOGAM Institute for Biomedical Research*

**Corresponding author: sangheon.lee@mogam.re.kr; hyunjin.shin@mogam.re.kr*

Translation efficiency (TE) of mRNA is mainly regulated by cis-regulatory elements of the nucleotide sequence. However, the quantitative understanding of how these elements function and interact to impact TE remains elusive. Therefore, developing an artificial intelligence (AI) model to analyze the potential association between mRNA sequence and TE will accelerate the research in this area. In our study, we introduce DuET (DUal Encoder model for Translation efficiency prediction), a deep convolutional neural network that predicts TE using the 5' untranslated region (UTR) and coding sequence (CDS) of an mRNA. For extensive and robust model training and evaluation, we compiled a large-scale dataset by aggregating multiple data sets from 2,830 experiments, encompassing 934,586 mRNA sequence-TE pairs across 65 human cell types provided by public databases. Across diverse cell types together, DuET demonstrated outstanding prediction performance, achieving a mean Spearman's correlation of 0.819. Furthermore, our analysis of the sequence representations obtained by DuET revealed the relative contribution of the 5′ UTR and CDS to TE prediction and demonstrated that end-to-end deep learning architectures can discover intricate patterns that surpass the scope of human-curated feature engineering. Finally, our efforts for model interpretation confirmed the importance of the translation initiation site (TIS)-proximal region and other 5′ UTR sequence motifs to TE prediction, offering further insights into the regulatory mechanism of mRNA translation.