[BIOINFO 2025]

CIPHER introduces translational clarity as a novel metric to identify actively translating open reading frames

Dong Jun Min [1,*] , Dahee Jeong [1] , Dulguun Sumberzul [1] , Acquah Emmanuel [1] , and Jun Cho [1,*]

[1]Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea

*Corresponding author: dongjun1010@gm.gist.ac.kr, juncho@gist.ac.kr

Micropeptides encoded by small open reading frames (smORFs) have emerged as functional regulators in diverse biological processes and as promising therapeutic targets in diseases. Despite advancements on bioinformatics tools leveraging ribosome profiling (Ribo-seq) data, prevailing approaches frequently generate false positive predictions due to the metrics that suboptimally capture the consistency of translation across ORFs. Here, we present CIPHER, an analytical framework that utilizes translational clarity, a novel metric based on cosine similarity, for precise identification of actively translating ORFs from Ribo-seq data. CIPHER achieved high predictive accuracy, minimized false positives from inconsistent three nucleotide periodicity, and accentuated biologically relevant and actively translating smORFs, including AW112010 lncRNA in M1 macrophages. Moreover, changes in translational clarity, or Δ translational clarity, captured condition-specific translational activity of Ppp1r15a uORF during M1 macrophage polarization. CIPHER offers a robust and broadly applicable approach for accurate ORF prediction, enabling deeper exploration of context-specific smORF translation across diverse biological settings.