Vol.2018-MPS-118 No.49
Vol.2018-BIO-54 No.49
2018/6/15

Prediction and Characterization of m6A-contained Sequences using Deep Neural Network

Yiqian Zhang[1,a)] Michiaki Hamada[1,b)]

Abstract: N6-methyladensine (m6A) is one of the most common and abundant RNA methylation modifications found in various species. As a mode of post-transcriptional methylation, m6A plays an important role in diverse RNA activities such as alternative splicing, interplay with microRNAs and translation efficiency. Though existing tools such as SRAMP could predict m6A at single-base resolution based on the data of miCLIP-Seq, a sequencing technology to pin exact m6A sites in genomes, the biological features of m6A-contained sequences are still unclear. We apply deep neural network to explore the biological information in the m6A-contained sequences. Our model is built on two layers of Convolution Neural Network (CNN), one layer of Bidirectional Long Short-Term Memory (BLSTM) and one fully-connected layer. We built separate models for human, mouse and zebrafish miCLIP-Seq data. Our deep learning model achieves better performance compared to other algorithms such as Random Forest, Logistic Regression and SVM. Moreover, independent test on the real MeRIP-Seq data shows our model achieves better prediction power than SRAMP. The learned motifs from the model correspond to known m6A readers like HNRNPG. Interestingly, our model also identifies a newly recognized m6A reader FMR1. In conclusion, we develop a useful tool to predict and characterize m6A-contained sequences and hope to provide more insights for m6A study.

[1] Department of Electrical Engineering and Bioscience, Waseda University
[a)] yiqian.zhang@fuji.waseda.jp
[b)] mhamada@waseda.jp