lncRNA_Mdeep: an alignment-free predictor for long non-coding RNAs identification by multimodal deep learning

CURRENT STATUS: UNDER REVISION

BMC Bioinformatics BMC Series

Xiao-Nan Fan
Northwestern Polytechnical University

✉ fanxn0223@163.com *Corresponding Author*

Shao-Wu Zhang
Northwestern Polytechnical University

Song-Yao Zhang
Northwestern Polytechnical University

Jin-Jie Ni
Northwestern Polytechnical University

DOI:

10.21203/rs.2.16792/v1

SUBJECT AREAS

Bioinformatics

KEYWORDS

Long noncoding RNA, alignment-free, multimodal learning, deep learning

Abstract

Background: Long non-coding RNAs (lncRNAs) play crucial roles in diverse biological processes and human complex diseases. Distinguishing lncRNAs from protein-coding transcripts is a fundamental step for analyzing lncRNA functional mechanism. However, the experimental identification of lncRNAs is expensive and time-consuming.

Results: In this study, we present an alignment-free multimodal deep learning framework (namely lncRNA_Mdeep) to distinguish lncRNAs from protein-coding transcripts. lncRNA_Mdeep incorporates three different input modalities (i.e. OFH modality, k-mer modality, and sequence modality), then a multimodal deep learning framework is built for learning the high-level abstract representations and predicting the probability whether a transcript is lncRNA or not.

Conclusions: lncRNA_Mdeep achieves 98.73% prediction accuracy in 10-fold cross-validation test on human. Compared with other eight state-of-the-art methods, lncRNA_Mdeep shows 93.12% prediction accuracy independent test on human, which is 0.94%~15.41% higher than that of other eight methods. In addition, the results on 11 cross-species datasets show that lncRNA_Mdeep is a powerful predictor for identifying lncRNAs. The source code can be downloaded from https://github.com/NWPU-903PR/lncRNA_Mdeep.

Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed.

However, the manuscript can be downloaded and accessed as a PDF.

Figures

Figure 1

Overview of IncRNA_Mdeep

Figure 2

Results of k-mer_DNN and One-hot_CNN with different parameters. (A) Accuracy of k-mer_DNN with different k value. (B) Accuracy of One-hot_CNN with different maxlen value.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

[Additional file 1.tif](#)

[Additional file 2.docx](#)