

The ICannoLncRNA pipeline for a Long-Non-coding-RNA identification and Annotation in Transcriptomic Data

Pronozin A.Y., Afonnikov D.A.

Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

E-mail: pronozinartem95@gmail.com



Long non-coding RNAs (lncRNAs) are molecules of more than 200 nucleotides in length that do not encode proteins. lncRNAs play essential roles in different regulatory processes in animals and plants. Existing pipelines often develop lncRNA recognition models using well-annotated data from a limited number of model organisms. This makes it difficult to apply this pipelines for growing transcriptomic data. In the present work, we developed automatic bioinformatics pipeline ICannoLncRNA for identification and analysis of lncRNA sequences in assembled transcriptomic sequences. The pipeline includes: Pre-processing, lncRNA filtering and lncRNA annotation. We analysed 15 maize transcriptome libraries from different plant tissues/organs. We identified 61,004 (9%) lncRNAs. Homologous search for lncRNAs identified in study among sequences from databases revealed - 643 maize lncRNA transcripts similar with lncRNAs in other plants. The largest numbers of homologues found in *O.sativa*. Analysis of the expression of newly detected lncRNAs points to high expression of them in the tassels of maize. The work was funded by the Kurchatov Genome Center of the FRC-IC&G SB RAS, № 075-15-2019-1662.

Take-home message:

The proposed automatic pipeline made it possible to identify 61,004 new lncRNAs in maize genome, classified into classes depending on their localization in the genome, annotate them and evaluate their structural features.

