

Bioinformatic assessment of factors affecting the correlation between protein abundance and elongation efficiency in prokaryotes

Korenskaia A.E.^{1, 2*}, Matushkin Yu.G.³, Lashin S.A.^{1, 2}, Klimenko A.I.¹

¹ Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* korenskaia@bionet.nsc.ru

Key words: translation elongation efficiency, protein abundance prediction

Motivation and Aim: As proteins are the key elements that provide cell function, developing the protein abundance prediction methods based on a gene sequence is the subject of scientific interest. One of the steps of protein abundance prediction is the evaluation of translation elongation efficiency based on mRNA sequence. The efficiency of translation elongation depends on the codon bias, amount and energy of secondary structures in mRNA. EloE (Elongation Efficiency) [1] tool uses these parameters in several combinations classifying a genome under analysis into one of the five elongation efficiency optimization types and to sort protein-coding genes according to the predicted efficiency of the translation elongation stage estimated by the corresponding Elongation Efficiency Index (EEI). The goal of this study is to evaluate the correlation coefficients between experimentally measured protein abundance and predicted elongation efficiency calculated by EloE for prokaryotes belonging to a wide range of taxonomic groups that have not been evaluated before.

Methods and Algorithms: The Spearman correlation coefficients between EEI indices and protein abundance, which was obtained by LC-MS/MS spectroscopy, were calculated for 26 prokaryotes. The dependence of this parameter on other factors, such as EEI type, minimal doubling time, mean (M) and standard deviation (R) EEI for ribosomal protein-coding genes, was estimated by the Pearson's correlation coefficient test with bootstrapping.

Results and Conclusion: It has been found that the obtained correlation values are not high and vary greatly among the organisms under study. Several factors associated with the correlation level have been revealed. The first factor is the elongation efficiency type – the organisms that belong to the translation elongation efficiency type that is taking into account codon bias only, have significantly higher correlation coefficients. The second factor is taxonomical identity – bacteria that belong to the most studied classes (Bacilli and Gammaproteobacteria) tend to have higher correlation coefficients. The third one is a growth rate – the organisms with higher correlation between EEI and protein abundance show lower minimal doubling time. Finally, we have demonstrated that the initial approach used by EloE that relies on assessing the ranks of ribosomal proteins in the gene list sorted by the base EEI value is adequate to the experimental data of the organisms under study and can be used in further developments of the algorithms. The obtained results are useful for further improvement of protein abundance prediction

methods including those that would take into account not only translation elongation but also other stages that affect the level of gene expression.

Acknowledgements: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

References

1. Sokolov V., Zuraev B., Lashin S., Matushkin Y. Web application for automatic prediction of gene translation elongation efficiency. *J Integrative Bioinformatics*. 2015;12(1):256.