Computer program for visualization of gene ontology results based on transcriptomics data

A.A. Zaporozhchenko[1], A.G. Galieva[1], Y.L. Orlov[1, 2]*

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

* *e-mail: orlov@bionet.nsc.ru*

Motivation and Aim: Analysis of gene ontology results based on selected gene lists presents technical problem for processing of transcriptomics data. New annotation categories, such as presence of miRNA, non-coding RNA, transcripts demand development of new mathematical functional for data processing [1]. Previously, program complex ICGenomics has been designed for storage, mining, and analysis of high-throughput sequencing experiments [2]. We sought extend potential of the tools by new functional properties for results presentation and visualization.

Methods and Algorithms: The developed program complex able fulfill the following tasks: ChIP-seq analysis; functional annotation of gene regulatory regions in nucleotide sequences; prediction of nucleosome positioning; and structural and functional annotation of proteins, including prediction of allergenicity parameters, as well as estimates of evolution changes in protein families. We developed set of new computer scripts for integration into the program complex. It implements both standard and modern methods for processing, analyzing, and visualizing sequencing data and functional annotation of genome regions.

Results and Conclusion: The program complex allows process nucleotide sequences from next generation sequencing data, contract gene expression tables, calculate percentage of genes related to defined categories (ontology terms, parameters of expression in different tissues – tissue-specificity). Correlation of any parameters of gene groups could be visualized as heat map. The work has to be extended to the analysis of transcription factor binding sites in eukaryotic genomes.

References
1. Wang J., Meng X., Dobrovolskaya O.B., Orlov Y.L., Chen M. (2017) Non-coding RNAs and their roles in stress response in plants. Genomics Proteomics Bioinformatics. 15(5):301-312.
2. Orlov Y.L. et al. (2012) ICGenomics: program complex for symbol sequence analysis in genomics, Vavilov journal of genetics and breeding. 16(4/1):732-741. (In Russian).