FoldGO for functional annotation of transcriptome data to identify fold-change-specific GO categories

D.S. Wiebe[1, 2], A.M. Mukhin[1, 2], N.A. Omelyanchuk[1, 2], V.V. Mironova[1, 2]*

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

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

* e-mail: victoria.v.mironova@gmail.com

Key words: functional annotation, Gene Ontology, RNA-Seq, microarray

Motivation and Aim: Scientists extensively generate transcriptome data to study genome response to external factors or internal conditions, thousands datasets are available online. As a rule, the transcriptome data contain much more information than has been extracted and published. Typical scenario of transcriptome data analysis is identification of differentially expressed genes (DEGs), those with significant changes in the number of transcripts, and functional enrichment analysis of the gene set using Gene Ontology [1]. This procedure provides for biological processes, molecular functions and cellular components GO terms that are significantly enriched in up- or down-regulated genes. In this analysis, one ignores the continuous distribution of expression changes and uses a binary classification of genes (DEGs and not DEGs) instead.

Methods and Algorithms: We developed a new method FoldGO, which classifies DEGs according to their fold change, performs functional annotation of the multiple gene lists generated on the basis of this classification (with overlaps) and provides for the GO terms, significantly overrepresented for the genes responded to the factor within a narrow fold-change-interval. The application implemented as an R package and as a web service.

Results: First, the method has been applied to the transcriptome responded to plant hormone auxin in *Arabidopsis thaliana* roots. In addition to the GO categories, which were significantly enriched for the DEGs, we found many others, more significant if we consider a narrower fold-change-interval [2]. For example, we found that among auxin-responsive genes associated with translation, 90 % only slightly increased their expression in auxin response. We also tested the method on a random set of transcriptomes generated for different organisms. We found fold-change-specific GO categories for most tested datasets.

Conclusion: FoldGO allows better describing the effect of an external factor on a gene network and identifying the groups of coordinatively expressed and functionally-related genes.

References
1. Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Sherlock G. (2000) Gene Ontology: tool for the unification of biology. Nat. Genet. 25(1):25-29.
2. Omelyanchuk N.A. et al. (2017) Auxin regulates functional gene groups in a fold-change-specific manner in *Arabidopsis thaliana* roots. Nat. Sci. Rep. 7(1):2489.