CAP-RNAseq: An online tool for Clustering, Annotation and Prioritization of RNAseq data

Merve Vural Özdeniz	mrvvrl.mv@gmail.com	Turkey	Bilkent University
Kubra Calisir	kubra.calisir@bilkent.edu.tr	Turkey	bilkent university
Burçin Irem Arıcı	irem.arici@bilkent.edu.tr	Turkey	Bilkent University
Rana Acar	ranaacar.00@gmail.com	Turkey	Bilkent University
Seniye Targen	Seniye.Targen@unice.fr	Turkey	Bilkent University
ErtugrulDalgic	ertuda@gmail.com	Turkey	Zonguldak Bulent Ecevit University
Ozlen Konu	konu@fen.bilkent.edu.tr	Turkey	Bilkent University

Transcriptome analysis has been an effective high throughput method for examining the regulation and function of genomes. However, analysis of RNAseq experiments with more than two groups/factors is complex and can benefit from clustering followed by annotation and prioritization of genes. While a few advanced analysis pipelines are available for such data sets, none of them make the gene selection process automated and hence easier for validation experiments. To fill this gap, we have developed a web tool named CAP-RNAseq, which creates co-expressed gene clusters using logCPM (voom transformed) values, and then annotates and prioritizes genes in the selected cluster. Users can upload their raw count RNAseq data, followed by application of ANOVA to filter/reduce the data to obtain genes showing significant changes between any two groups. k-means clustering is then applied with the user-specified cluster numbers; and the emerging clusters can be visualized and enriched functionally with mSigDB, GO and KEGG databases. CAP-RNAseq also suggests genes for future validation by qRT-PCR and Western Blotting. The algorithm used for this prioritization implements distance correlation (multivariate independence), which measures the correlation between the consensus profile of a cluster and profile of each gene in that cluster. Then, top correlated genes are shown in a table from which the user can select a gene for primers to be designed. Moreover, the Human Protein Atlas data has been integrated into CAP-RNAseq to visualize the protein levels of suggested or user-selected genes in selected cancer types. CAP-RNAseq has been designed and implemented for the R-Shiny platform and is user-friendly. We present several case studies for its use in breast cancer and hormone replacement therapy based on our own RNAseq datasets.

This study has been funded by TUSEB (Grant No. 4405).