

RTranscriptogram: a tool for biological data integration

Alex Augusto Biazotti¹, Túlio Moreira Fernandes¹, André Luiz Molan², Agnes Alessandra Sekijima Takeda¹, Jose Rybarczyk-Filho²,

1 Instituto de Biociências de Botucatu - UNESP

2 UNESP

Abstract

Every day, new technologies are emerging that make it possible the large-scale study of RNAs transcribed by an organism under specific conditions, providing a huge amount of information. However, the traditional methodologies are not able to efficiently analyze these data due the use of pre-defined cut-offs, thus eliminating a large number of genes not considered differentially expressed, and consequently reducing precision and accuracy of the study. This work proposes the tool called RTranscriptogram which performs an overall analysis of an organism, integrating protein networks biological, processes and expression genes.. This tool clusterize the network, extract the group and their respective biological information and project the expression genes on the network. To test the tool, we used the STRING database version 10 to prospect for a protein network for Homo sapiens, Gene Ontology provided the biological processes (BPs) for this study. Gene expression data were prospected from Gene Expression Omnibus (GEO): GSE19804 - lung samples from Taiwanese female nonsmokers with and without cancer. GSE10072: - lung samples from Italian female and male smokers, former-smokers, non-smokers with or without cancer. These databases were integrated by the transcriptograma technique to obtain expression profiles correlated with protein networks and ontologies. The analyzes presented 2 up-regulated biological processes and 30 down-regulated biological processes that were similar in the comparisons made with people with cancer, in addition the transcriptional activity for Taiwanese and Italians presented a similar profile. Smokers with cancer presented 175 altered BPs when compared with non-smokers. Despite different habits among populations, lung cancer has a high similarity in transcriptional activity.

Funding: CNPq processes 473789/2013-2