

# Comparison of bioinformatics approaches to evaluate altered GO processes in in vivo and in vitro studies of antineoplastics of OPEN TG-GATEs online database

Giordano Bruno<sup>1</sup>, André Luiz Molan<sup>1</sup>, Jose Rybarczyk-Filho<sup>1</sup>,

*1 UNESP*

## Abstract

Toxicogenomics is a promising field that has been continuously developed in the last decades. It's main goal is to study the toxic phenotypes of various chemicals in biologic systems with the aid of technologies used by omics sciences. The in silico component of a toxicogenomics study, still presents many challenges in regards to what methods should be used, for example, to determine what biological functions are significantly altered using gene expression data as a starting point. In this work we propose the comparison of two bioinformatics methodologies for the determination of biological processes (BP), molecular functions (MF) and celular components (CC). Gene expression data of three antineoplastics, Cyclophosphamide, Etoposide and Lomustine from the OPEN TG-GATEs online database will be used. The studies that will be used for comparison are the High dose and 24 hours cases of Homo sapiens in vitro, Rattus norvegicus in vitro and Rattus norvegicus in vivo. The first methodology will be functional enrichment of differentially expressed genes (DEGs) and the second method is an application of Shannon's normalized function of entropy to determine the relative activity and diversity of groups of functionally associated genes (GFAGs), this method is applied by the EntropyClusterGenes R package. The gene expression data used as input for both methodologies will be normalized using the same protocol, the RMA (Robust-Multiarrray Average) which is a widely used normalization procedure for affymetrix Genechip microarrays data. In the first method genes that present  $p\text{-value} < 0.05$  and  $\text{LogFC} > 1$  or  $\text{LogFC} < -1$  will be considered as DEGs, Bioconductor's R package 'topGO' will then use these DEGs as input to determine p-values for Gene Ontology (GO) processes (BP, MF and CC). The second method will group genes according to their GO and then use the groups total expression levels along with a bootstrapping statistic and FDR of 0,05 to determine adjusted p-values for said groups. In both methods, GO processes that presented  $p\text{-values} < 0,001$  were considered as significantly altered. In general, the GFAG method detected more process than DEGs method, but there certain cases where the DEGs method detected more processes. However both methodologies showed similar behavior in regards to the number of detected processes across the drugs. The methodologies also were concordant in some of the results. In the etoposide case, for example, processes related to DNA replication and cell division were found to be significantly down regulated in both methodologies.

Funding: CNPq processo134467 / 2016-7