See.. > Nitin > Computational Tools

<https://www.nature.com/articles/s41596-018-0103-9>

January 23, 2023

Notes on paper/terminology

**Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap**

Pathway enrichment analysis: interpret results from RNA-seq experiments wrt processes that are over/underrepresented

Input: (sometimes ranked) list of genes and expression/activity data

Output: list of pathways and statistical measures indicating activity wrt expected (p,q-val..)

can then be visualised/interpreted in different ways

3 steps:

- define gene lists from the data [large amounts of data]

- determine enriched pathways

- visualise the results

-

Terminology:

Omics:

Enriched pathways:

Pathway. Genes that work together to carry out a biological process.

Gene set. A set of related genes. A ‘pathway gene set’ includes all genes in a pathway. Gene sets can be based on various relationships between genes, such as cellular localization (e.g., nuclear genes) or enzymatic function (e.g., protein kinases). Details such as protein interactions are not included. Gene list of interest. The list of genes derived from an omics experiment that is input to pathway enrichment analysis.

Ranked gene list. In many omics data (e.g., that from RNA-seq for gene expression), genes can be ranked according to some score (e.g., level of differential expression) to provide more information for pathway enrichment analysis. Pathways enriched in genes clustered at the top of a ranked list would score higher than if the pathway genes are randomly scattered across the ranked list.

Pathway enrichment analysis. A statistical technique to identify pathways that are significantly represented in a gene list or ranked gene list of interest.

Multiple testing correction. Thousands of pathways may be individually tested for enrichment, and this could lead to significant enrichment P values appearing by chance alone. Multiple testing correction is a statistical technique to correct the P values from individual enrichment tests to address this problem and reduce the chance of falsepositive enrichment (Box 3).

Leading-edge gene. A subset of genes found in the ranking at or just before the maximal ES in a GSEA analysis. This subset of genes often accounts for a pathway being defined as enriched.