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..)

Ex: more or less active than it should be

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 - fields of bio? Ex: genomics – data types from these omics

Gene set - set of related genes

- by pathway, location, type, etc.

Gene list of interest - input for pathway enrichment analysis

- derived from omics experiment

Ranked gene list - rank list of interest by a score [ex: level of diff expression]

- pathways enriched in genes at top score higher

- ranked by different tests \*

Pathway - group of genes involved in a particular biological process

Pathway enrichment analysis - identify pathways significantly rep’d in a gene list of interest

Enriched pathway - pathway that is under/overactive compared to usual?

Enrichment score - rep’s enrichment of pathway

- how much the pathway tends towards top/bottom of ranked list

Multiple testing correction - when individually testing pathways, significant values could appear by chance

- technique to correct p-vals from individual enrichment tests to reduce chance of false postive

Leading-edge gene - subset of genes found in ranking at/just before the maximal ES

- often accounts for a pathway being defined as enriched

- ? so these are the genes most significantly causing the pathway to be enriched?

[Hypergeometric testing (ORA) and GSEA were used in functional analysis workshop]

GSEA is a type of pathway enrichment analysis

Recommended when all/most genes can be ranked

Does not prefilter – considers all genes

Increases/dec ES based on position of genes of pathway in list

Hypergeometric test is often used for pathway enrichment analysis

Used on unranked list

Uses a threshold to create significant and background set

Compares ratio of pathway genes in significant set with background set

Advantages of pathway enrichment analysis:

- more reliable analysis

- aggregates mutation count in genes over entire process

- reduces dimensionality

- from tens of thousands of genes to smaller number of pathways

- map different types of data to same pathways

- various omics (genomics, transcriptomics, etc.)

Disadvantages:

- not all pathways treated equally

- larger/well known given more significance

- if controlled by few genes or not by gene expression, given less significance

- pathway boundaries are sometimes arbitrary

- multifunctional genes throw off results