What to do with a gene list of interest. An R-Shiny application.

Costas Bouyioukos, team EDCD, UMR7216 Epigenetics and cell fate.

Projet long Proposal – M2BI 2019

Goal

To develop an R-Shiny application for the downstream treatment and analysis of a list of "genes of interest". The program will integrate input from an array of tools for gene-list, gene-enrichment, ontologies and pathways analysis, with a goal to extract as much useful biological knowledge possible.

Introduction

All high throughput modern biological techniques applied to any *omics experiment, including genomics, transcriptomics, translatomics and proteomics have as a result a list of genes of interest. There exist numerous approaches, namely all the family of enrichment tests, which aim at extracting some biological knowledge out of a list of genes, hover with the exception of Gene Ontology (GO) analyses they are all either obscured or excluded by most of studies because of luck of usability and/or interpretability.

The enhancement of the results of any kind of genome-wise high-throughput *omics experiment at a level that goes beyond the finding of a single gene list will provide a valuable tool in the hands of many researchers in biosciences. The visualisation and interpretation of biological knowledge that can be extracted from such an approach will additionally motivate and facilitate the search more more mechanistic explanations for the biological systems of study.

Project

The project will be based on the developed competences and expertise developed in the team from the computational analysis of a wide variety of multi-omics datasets. It will consolidate and integrate in a single framework the methods, tools and techniques that were developed in-house and will aspire to deliver a useful application for any bioscientist working with genome-wise high-throughput *omics data.

The application will use the well developed and mature code base of R-bioconductor for the analysis of *omics data and the R-shiny package to create interactive applications in the R programming language.

The students will have the opportunity to learn how to develop a whole application entirely within R by applying some of the competences they have already developed during the "Scientific Software Carpentry" course with python. Will have a first-hand experience of the

packages, methods and tools that are routinely used for an everyday bioinformatics analysis but, and more importantly, they will have to develop their own expertise and methodological approach to solve problems related to data visualisation, statistical testing and software engineering.

References

[1] Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, Merico D, Bader GD.
Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap.

Nat Protoc. 2019 Feb;14(2):482-517. doi: 10.1038/s41596-018-0103-9.

[2] Huang da W, Sherman BT, Lempicki RA.
Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.

Nat Protoc. 2009;4(1):44-57. doi: 10.1038/nprot.2008.211.

[3] R Shiny, R Studio.

https://shiny.rstudio.com/