

BITS :: Call for Abstracts 2021 - Poster

<i>Type</i>	Poster
<i>Session</i>	Gene regulation, transcriptomics and epigenomics
<i>Title</i>	Reproducible, Robust and High Quality Ontology Enrichment Visualizations via Auto-GO
<i>All Authors</i>	Eleonora Sperandio (1), Isabella Grassucci (1), Lorenzo D'Ambrosio (2), Matteo Pallocca (1)*

Affiliation

(1) UOSD Biostatistics, Bioinformatics and Clinical Trial Center, IRCSS Regina Elena National Cancer Institute, Rome

(2) UOSD Immunology and Tumor Immunotherapy, IRCSS Regina Elena National Cancer Institute, Rome

Motivation

Almost every Molecular Biology paper contains a Gene Ontology (GO) or a Pathway enrichment analysis, and the level of standardization and reproducibility is worryingly low for such a popular technique. The level of concern is twofold: first, since in the GO analysis lies one of crucial aspect of a manuscript, namely the list of significant putative functions related to a gene list. Secondly, since it has been shown that current approaches in the -omics era suffer from (1) Fishing expeditions, i.e. naïve approaches that fail to address the basics of the phenomena (2) Cherry-picking, i.e. selecting specific functions that are not top scored by GO algorithms as cosmetic or supporting data. Most of the GO analyses currently published rely on web applications, that on one hand enable non-bioinformaticians to conduct exploratory analyses, on another opened another huge concern for non-reproducibility of results, being a manual step of data processing strongly contrasting the rules for reproducible bioinformatics [Sandve 2013, Kulkarni 2017]. Virtualization techniques such as Docker and Singularity help encapsulate software in order to enable total reproducibility, while additional layers such as Nextflow and Snakemake enable to build complex virtualized pipelines and run them in High Performance Computing Clusters. Unfortunately, what is presented on a life science paper is not primary output matrices, but functional enrichments, that currently do not benefit from such technology advancements. All these issues must be tackled by a framework that enables (1) reproducible GO analyses (2) high-quality automated visualizations (3) proposes a high-level visualization for complex experimental designs with multiple conditions. Here we present Auto-GO, a framework that addresses all these issues.

Methods

The Auto-GO protocol core is an atomic function that enriches a gene list over a list of selected databases or libraries as defined by the Enrichr naming system. Every combination produced a high-quality barplot with the top N terms enriched, with dynamic resizing in order to accommodate long terms naming (e.g. long term here) in the final plot. The gene list can derive from several Genomics applications, such as (1) modulated genes from RNA-sequencing (2) differential epigenetically active regions from ChIP-seq, ATAC-seq or any other tag-seq (3) a list of significantly mutated genes from Whole Genome or Whole Exome Sequencing. The framework is built with the possibility of being generalized with other enrichment functions other than EnrichR, with the only constrain of having a gene list as input and a tuple matrix with as output.

Results

To provide a proof-of-concept application of our project, we will employ RNA-seq data on several silencing (iRNA) conditions on Acute Myeloid Leukemia cell line K562, from the ENCODE project. This setting represents a controlled in-vitro environment with multiple replicated conditions. As an in-vivo casuistry, we will employ TCGA-LUAD RNA-seq data on Lung Adenocarcinoma FFPE samples, separating groups according to the two mostly significant mutated oncogenes, namely KRAS and TP53. Finally, to show how Auto-GO applies to Epigenetic screening as well, we will dissect the multi-GO branch of our framework on multiple ChIP-seq Transcriptional Factors assays on K562. An example application of the multi-GO Heatmap is represented in Figure 1, where several comparison terms can be displayed in one heatmap, tackling one of the typical issues of differential expression analysis that enables only 1vs1 comparison visualizations. The overall code repository is available at <https://gitlab.com/bioinfo-ire-release/auto-go/>.

Info

If the BITS congress needs it, this abstract/poster can be turned into an oral presentation. Additional work is being put in place to deploy easily the package into a Docker container and/or a conda package.

Figure

<i>Availability</i>	https://gitlab.com/bioinfo-ire-release/auto-go/
Corresponding Author	
<i>Name, Surname</i>	Matteo, Pallocca
<i>Email</i>	matteo.pallocca@ifo.gov.it
<i>Submitted on</i>	08.05.2021

Società Italiana di Bioinformatica

C.F. / P.IVA 97319460586

E-mail bits@bioinformatics.it

Sede legale Viale G. Mazzini, 114/B - 00195 Roma

Website bioinformatics.it

message generated by sciencedev.com for bioinformatics.it 11:03:46 08.05.2021
