

Comparative Visualization of Long Lists of Gene Ontology Terms

T. Harbig¹ and K. Nieselt¹

¹University of Tübingen, Germany

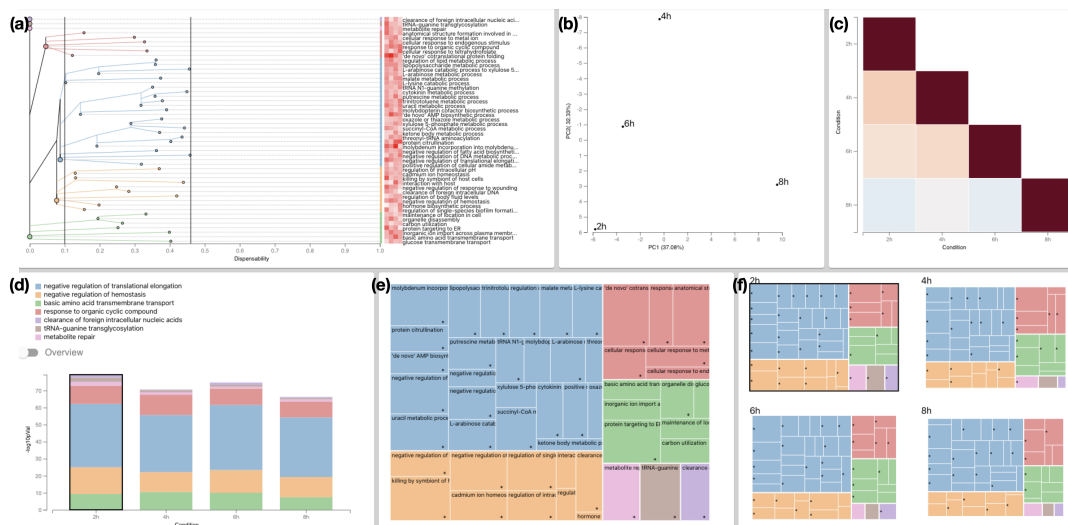


Figure 1: Dashboard for comparing GO term lists. (a) GO terms are visualized in a clustered heatmap as well as in (d) bar charts and (e,f) treemaps for each condition. The overall similarity of the conditions is visualized in (b) a PCA plot and (c) a correlation heatmap.

Abstract

Analysis pipelines for RNA sequencing data often produce long lists of differentially expressed genes. For functional analysis, these lists can be analyzed using Gene Ontology (GO) [ABB*00] enrichment analysis tools such as PANTHER [MME*19] and DAVID [STC*07] resulting in lists of enriched GO terms with associated p-values. Since GO terms are organized roughly hierarchical, an enrichment of a child term propagates to enrichments of parent terms which can lead to redundancy in the list. Moreover, when GO enrichments are computed for multiple experimental conditions, it can be of interest to compare and summarize the results. With our tool we introduce a technique to reduce the redundancy in multiple lists of GO terms and provide an interactive visualization for list comparisons (see Figure 1).

We developed a web application that takes multiple lists of GO terms or genes as input. For lists of genes GO term enrichment is computed using the PANTHER API resulting in lists of GO terms. The GO terms are clustered hierarchically using a version of the REVIGO algorithm adapted to use multiple lists of GO terms simultaneously [SBSŠ11]. REVIGO clusters GO terms based on their semantic similarity, p-values, and relatedness resulting in a hierarchical clustering where less dispensable terms are placed closer to the root. The clustering is visualized in a clustered heatmap showing the p-values for each term. With sliders users can filter dispensable GO terms and select a cutoff for the extraction of non-hierarchical clusters resulting in a set of GO terms with reduced redundancy and a meaningful grouping. In order to compare the clusters a treemap is visualized for each condition. Moreover, the treemaps can be compared by animation which facilitates viewing changes between conditions and bar charts for detailed comparisons of single terms. In order to provide a global overview, the overall similarity of the lists is visualized in a PCA plot and a correlation heatmap. A detailed table shows further information about the GO terms.

This tool will enable researchers to compare the functional analysis of multiple experimental conditions. Moreover, it can be extended to compare GO enrichment results of any other application where gene lists are produced, such as the analysis of clusters in gene co-expression networks or the comparison of differentially expressed genes in multiple species.

References

- [ABB*00] ASHBURNER M., BALL C. A., BLAKE J. A., BOTSTEIN D., BUTLER H., CHERRY J. M., DAVIS A. P., DOLINSKI K., DWIGHT S. S., EPPIG J. T., ET AL.: Gene ontology: tool for the unification of biology. *Nature genetics* 25, 1 (2000), 25–29. [1](#)
- [MME*19] MI H., MURUGANUJAN A., EBERT D., HUANG X., THOMAS P. D.: Panther version 14: more genomes, a new panther go-slim and improvements in enrichment analysis tools. *Nucleic acids research* 47, D1 (2019), D419–D426. [1](#)
- [SBŠŠ11] SUPEK F., BOŠNJAK M., ŠKUNCA N., ŠMUC T.: Revigo summarizes and visualizes long lists of gene ontology terms. *PloS one* 6, 7 (2011), e21800. [1](#)
- [STC*07] SHERMAN B. T., TAN Q., COLLINS J. R., ALVORD W. G., ROAYAEI J., STEPHENS R., BASELER M. W., LANE H. C., LEMPICKI R. A., ET AL.: The david gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology* 8, 9 (2007), R183. [1](#)