

## Visualizing GO Annotations

Fran Supek and Nives Škunca

### Abstract

Contemporary techniques in biology produce readouts for large numbers of genes simultaneously, the typical example being differential gene expression measurements. Moreover, those genes are often richly annotated using GO terms that describe gene function and that can be used to summarize the results of the genome-scale experiments. However, making sense of such GO enrichment analyses may be challenging. For instance, overrepresented GO functions in a set of differentially expressed genes are typically output as a flat list, a format not adequate to capture the complexities of the hierarchical structure of the GO annotation labels.

In this chapter, we survey various methods to visualize large, difficult-to-interpret lists of GO terms. We catalog their availability—Web-based or standalone, the main principles they employ in summarizing large lists of GO terms, and the visualization styles they support. These brief commentaries on each software are intended as a helpful inventory, rather than comprehensive descriptions of the underlying algorithms. Instead, we show examples of their use and suggest that the choice of an appropriate visualization tool may be crucial to the utility of GO in biological discovery.

**Key words** Gene Ontology, Visualization, Interpretation, Redundancy, Enrichment, Tools

---

### 1 Introduction

We have entered the era of massive data sets in biology. A variety of experimental and computational techniques can produce readouts for many genes—or whole genomes—simultaneously. Moreover, we can also assign rich functional annotations to most of the genes of interest. Such a wealth of data is accompanied with challenges in interpretation.

In this chapter, we focus on methods that visualize long lists of Gene Ontology (GO) terms [1]. The methods we survey take as input a flat list of GO terms, often accompanied by some user-supplied measure of statistical significance or importance. Visualization methods summarize such lists to distil the most relevant information. Finally, these methods produce various styles of visualization that can aid interpretation.

First, we examine the challenges related to understanding large lists of GO terms; second, we provide a systematic overview of the published methods that address these challenges; third, we discuss different visualization styles these methods use; and fourth, we give usage examples for a selection of these tools.

---

## 2 Understanding Large Lists of Genes and Their Gene Ontology Labels

A classical example of a large biological dataset are gene expression measurements by RNA-Seq, which monitor the genome-wide changes in transcriptional regulation between experimental conditions. Typically, tens or hundreds of genes will be upregulated or downregulated in response to a particular treatment. This indicates that a systems-level change in the experimental model has occurred, which may be described by examining the common properties of the genes whose expression was altered. Do these genes participate in the same metabolic or signaling pathways? Do they perform similar biochemical functions? Do their protein products colocalize in the cell? Formally, such sets of genes are subjected to statistical tests for enrichment for various functional categories [2]. The gene functions tested are typically described by Gene Ontology (GO) terms [3], although alternatives such as KEGG Pathways or CORUM protein complexes can be used.

Of note, such GO enrichment analyses are by no means restricted to experiments measuring changes in gene expression, nor to experimental data in general. Any list of genes for which interpretation is sought can be described using enriched GO terms and it could, for instance, derive from comparative genomics. In particular, one could perform an evolutionary analysis to look at biological roles of gene families that have expanded in a certain eukaryotic lineage, e.g., [4]. Similarly, a researcher may wish to describe the overall functional repertoire in a newly sequenced genome, while comparing to existing genomes of related organisms.

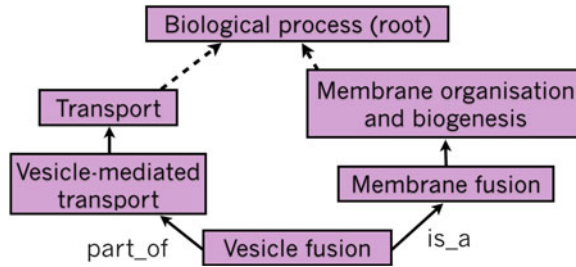
### 2.1 Challenges in Interpreting Lists of Enriched GO Terms

As Chap. 3 [5] describes, the GO is a hierarchical structure, wherein the individual terms can have not only multiple descendants, but also multiple parents; more formally, GO is a directed graph; the basic version of the GO is also a directed *acyclic* graph (Chap. 3<sup>1</sup> [5]; Fig. 1). This complex structure, along with its large size—the GO has thousands of nodes—make it challenging to display the part(s) of the GO of interest. For instance, a list of GO terms found to be enriched in a gene expression experiment could be concentrated in one part of the GO graph.

A further complication is that such lists of interesting GO terms tend to be large, meaning that many different biological processes or molecular functions may appear to be affected in the experiment.

---

<sup>1</sup><http://geneontology.org/page/download-ontology>



**Fig. 1** A subset of the Gene Ontology Directed Acyclic Graph (DAG) for the GO term “vesicle fusion” (GO:0006906). The GO is a DAG: terms are nodes, while the relations are edges. Two main relation types between terms are “is\_a” and “part\_of.” More specific terms are found deeper in the graph. Thus, if a gene product is annotated with a GO term, it is by definition also annotated with all the parent terms of that GO term

One reason for this is that the GO itself is designed and developed to describe nuances in gene function as exhaustively as possible; consequently, many of the GO terms will be partially redundant. For instance, many of the genes participating in “translation” (GO:0006412) are also structurally a part of “ribosome” (GO:0005840).

In addition to the inherent redundancy of the GO, responses of biological systems to experimental perturbation often genuinely involve coordinated activity of many related and/or overlapping subsystems. For example, replicating cells facing DNA damage may upregulate “nucleotide-excision repair” (GO:0006289) to help fix the lesions, but at the same time resorting to “error-prone translesion synthesis” (GO:0042276) to ensure DNA replication finishes.

## 2.2 Visualizing the GO to Facilitate Insight and Avoid Biases

GO term enrichment analyses often result in lists of significant GO terms that are both long and redundant, hampering interpretation. Various methods to visualize such lists may help investigators spot dominant trends in the data, leading to novel biological insight. Such visualizations mostly operate by different ways of grouping and displaying similar GO terms together, wherein the structure of the GO defines what is similar and what is not (see *semantic similarity analysis* below). In its simplest form, this involves displaying a part of the GO hierarchy with the GO terms of interest highlighted and their parent–child relationships shown. Displaying also the user-supplied experimental data may help prioritize which GO terms, among many similar ones, are of higher interest.

We suggest that having an unbiased way to algorithmically organize GO terms derived from experimental data helps prevent unintentional biases in interpretation. If unaware of the overall semantic structure in the set of significant GO terms, the investigator may pick one or two GO terms in the list that “make sense,” in terms of fitting with their expectations. By visualizing the interre-

relationships between the GO terms alongside the statistical support for each in the experimental data could help avoid focusing on outlying—and perhaps spurious—results. In addition, one could be made aware of the common pitfall where one GO term is chosen, while other similarly statistically supported terms are ignored. Finally, and very importantly, a good visualization is also an effective means of presenting summaries of scientific results, whether in papers, presentations or posters.

---

### 3 Overview of the GO Visualization-Related Tools

Here we systematize and describe the currently available tools for visualizing sets of GO annotations. Additionally, we highlight three of these tools in more detail. The tools and the underlying methods they implement can be classified thusly:

1. *Interactive GO browsers.* Tools for interactively browsing the entire GO and also the genes known to be annotated with chosen GO terms. Importantly, these do not take into account a user-supplied set of annotations of interest, e.g., derived from an enrichment analysis of experimental data. Visualization is typically not emphasized and not configurable. See AmiGO [6] and QuickGO [7]. Of note, OLSVis [8] can display other biomedical ontologies in addition to the GO.
2. *Network visualization tools.* These are not particular to the GO, but can display any kind of graph, including the GO or a part thereof. The visualization options are highly configurable; however, since these tools were not designed specifically for GO, they tend to be more complicated to use. See Cytoscape [9], Gephi [10], and Pajek [11].
  - (a) Of note, there are Cytoscape plugins specialized for handling groups of GO terms: EnrichmentMap [12] and BINGO [13].
3. *GO visual overlays.* Tools that can visualize an interesting subset of the GO, and display some additional data about each shown GO term. Typically, this involves coloring the GO terms by the enrichments or p-values determined from user-supplied gene lists (these tools tend to also perform the GO enrichment analysis). They display the terms arranged by parent–child relationships, in a tree-like visual layout. Examples include GOrilla [14], GRYFUN [15], GOFFA [16], and SimCT [17].
  - (a) In addition to the GO, similar tools are available which can highlight the individual members in displayed KEGG pathways [18]; the pathways can also be shown in a KEGG BRTE functional hierarchy with FuncTree [19].
4. *Semantic similarity analysis.* Tools that examine the semantic similarity (redundancy) between various GO terms, including

those that are not linked by direct parent–child relationships. The similarities are used to organize a set of interesting GO terms into clusters and/or graphs, while simultaneously allowing highly redundant terms to be filtered out. The user can supply enrichments or p-values to prioritize results. Implemented in REVIGO [20] and RedundancyMiner [21].

- (a) Some provisions for this are made in g:Profiler [22], which collapses similar GO terms.
  - (b) The Ontologizer [23] can perform a statistical test for enrichment that accounts for the parent–child redundancy [24] prior to visualizing results.
5. *Emerging methods.* These may involve display of the trends underlying a group of GO terms in a so-called “tag cloud” (with text in various colors and sizes), or in a tree map (a hierarchical organization of colored tiles), as in REVIGO [20] or GOSummaries [25]. Additionally, several tools now support the display of multiple GO enrichment analyses side-by-side; see BACA [26] or GOSummaries. SimCT [17] can display subtrees of other biomedical ontologies in addition to the GO.

---

## 4 Case Studies with Selected Tools

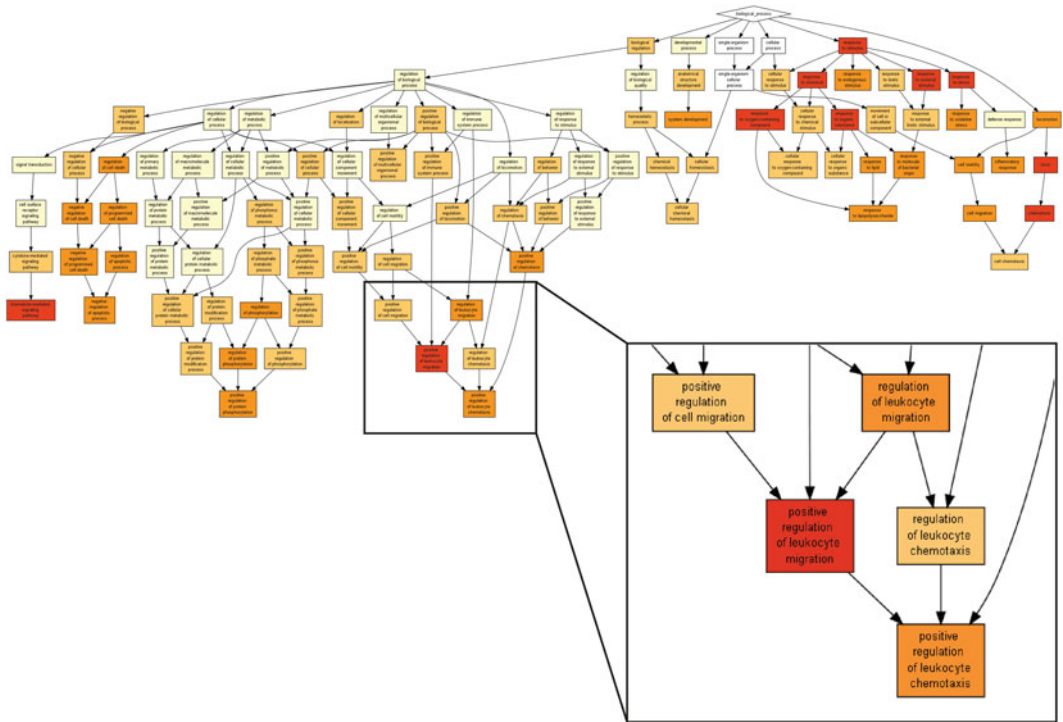
**GORilla** [14] is a Web-based tool that can take two types of input: either a ranked list of genes or two lists, one with the target genes and the other with the background genes. As output, GORilla produces a visualization that indicates which terms are significantly enriched.

We focus here on the enrichment analysis that takes a ranked list of genes. Briefly, the null hypothesis is that the occurrences of a GO term at various points in the ranked list are equiprobable. Lower *p*-values indicate a higher confidence for a GO term to be enriched towards the top of the list.

As an example analysis, we downloaded a dataset of transcription profiling by microarray of human peripheral blood mononuclear cells after a treatment with *Staphylococcus aureus* and incubation for different lengths of time [27], obtained from the Gene Expression Atlas [28] at <http://www.ebi.ac.uk/gxa/experiments/E-GEOD-16837>. In the GORilla Web interface (<http://cbl-gorilla.cs.technion.ac.il/>), we set the p-value threshold to  $10^{-3}$ , and the remaining settings were the defaults in the tool.

The display is shown in Fig. 2. Based on the color of the boxes, the user can visualize which GO terms are enriched, and the connecting lines describe their relationship to other terms in the GO graph.

**REVIGO** [20] analyzes large lists of significant GO terms and removes the redundant terms, in order to further narrow the search to a set of nonredundant and highly significant GO terms. Briefly, REVIGO creates clusters of GO terms that are semantically similar, and selects one representative for each cluster.



**Fig. 2** A visualization of the Biological Process Gene Ontology annotations using GOrilla. The dataset used is a microarray transcription profiling of human peripheral blood mononuclear cells after treatment with *Staphylococcus aureus* (Expression Atlas dataset ID E-GEOD-16837). The GOrilla settings were left at default values:  $p$ -value threshold of  $p < 10^{-3}$ , organism *Homo sapiens* and running mode “single ranked list”

One possible input for REVIGO is a list of GO terms with the associated  $p$ -values, such as the output list from GOrilla. Alternatively, REVIGO can take as input any other list of GO terms, with or without associated numerical values, and provide various styles of visualization. First, a scatterplot that distributes the GO terms, represented as bubbles, in a 2D space that will put two GO terms closer together if they are more semantically similar. Second, an interactive graph that connects the user-supplied set of GO terms based on the structure of the GO hierarchy. Third, a TreeMap where terms are clustered and clusters displayed as colored tiles. Fourth, REVIGO provides a word cloud that highlights the most frequent keywords in the names and descriptions of the GO terms.

To perform the analysis, we used the setting in GOrilla to automatically forward its GO term enrichment results as a query to the REVIGO tool. In REVIGO, we used the default settings.

**Fig. 3** (continued) while its size reflects the generality of the GO term in the UniProt-GOA database. **(b)** The table view shows the list of all the input GO terms: those shown in the scatterplot are written in regular font, while those labeled as redundant by REVIGO are shown in *gray italics*

# a REVIGO

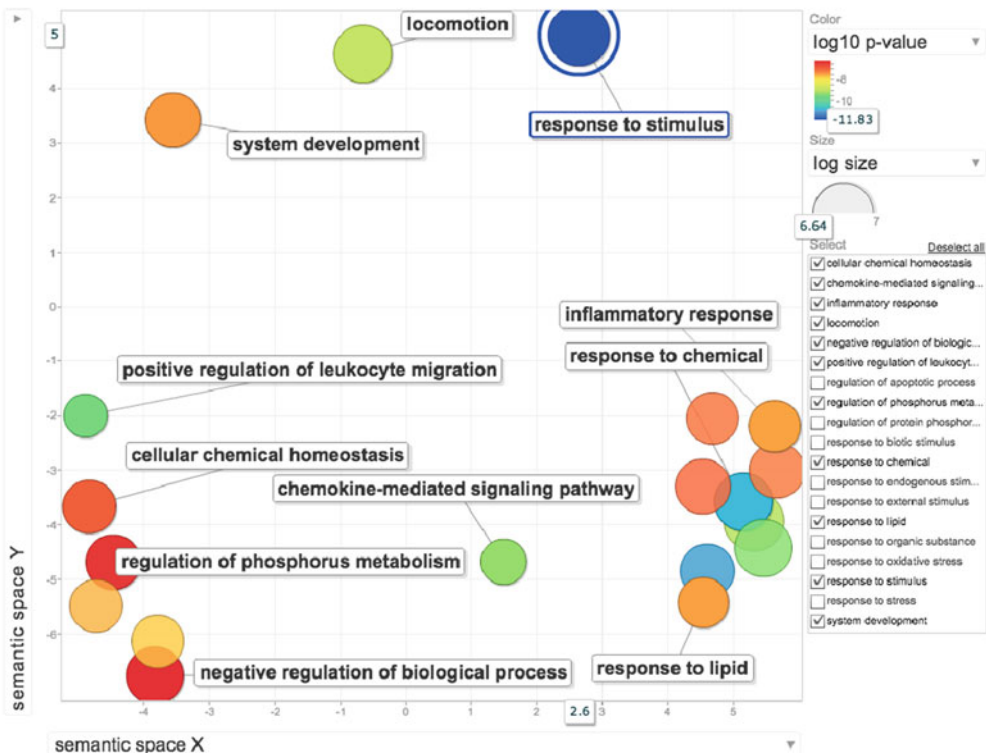
Biological Process (54)

Tag Clouds

Scatterplot & Table

Interactive Graph

TreeMap



## b

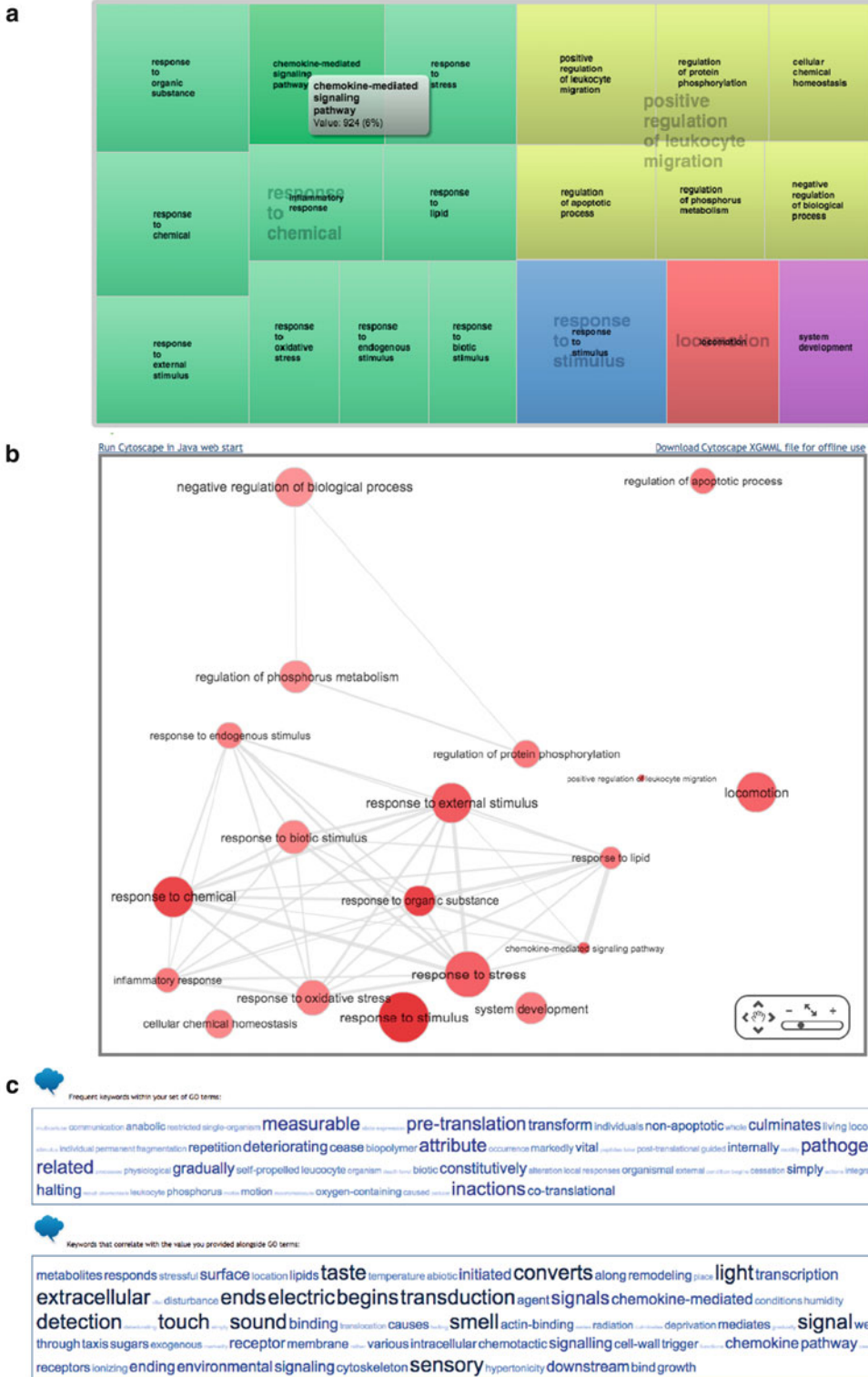
Hide/show dispensable GO terms

Export results to text table (CSV)

Make R script for plotting

term ID	description	frequency	pin?	log <sub>10</sub> p-value	uniqueness	dispensability
GO:0007687	positive regulation of leukocyte migration	0.004 %		-9.5017	0.42	0.00
-GO:2009147	positive regulation of cell motility	0.018 %	[-]	-6.3019	0.40	0.99
-GO:0050921	positive regulation of chemotaxis	0.005 %	[-]	-7.1884	0.28	0.86
-GO:0048520	positive regulation of behavior	0.006 %	[-]	-6.7570	0.41	0.87
-GO:0050920	regulation of chemotaxis	0.017 %	[-]	-6.3575	0.32	0.92
-GO:0020325	positive regulation of cell migration	0.017 %	[-]	-6.2830	0.39	0.93
-GO:0002690	positive regulation of leukocyte chemotaxis	0.003 %	[-]	-7.5361	0.25	0.96
-GO:0002688	regulation of leukocyte chemotaxis	0.004 %	[-]	-6.9747	0.37	0.96
-GO:0002685	regulation of leukocyte migration	0.006 %	[-]	-8.0820	0.50	0.90
-GO:0051272	positive regulation of cellular component movement	0.018 %	[-]	-6.3019	0.46	0.83
GO:0040011	locomotion	1.437 %		-8.8125	0.96	0.00
GO:0042221	response to chemical	1.603 %		-10.7167	0.63	0.00
GO:0050896	response to stimulus	8.818 %		-11.8297	0.97	0.00
GO:0048731	system development	0.354 %		-7.3546	0.91	0.04
-GO:0048522	positive regulation of cellular process	0.429 %	[-]	-6.0625	0.52	0.82
-GO:0051240	positive regulation of multicellular organismal process	0.131 %	[-]	-6.3307	0.53	0.79
GO:0042981	regulation of apoptotic process	0.129 %		-7.7282	0.61	0.21
-GO:0060548	negative regulation of cell death	0.051 %	[-]	-7.6440	0.63	0.93
-GO:0043049	negative regulation of programmed cell death	0.049 %	[-]	-7.2434	0.62	0.99
-GO:0043066	negative regulation of apoptotic process	0.047 %	[-]	-7.3706	0.62	0.92
-GO:0043067	regulation of programmed cell death	0.134 %	[-]	-7.6819	0.61	0.99
-GO:0010941	regulation of cell death	0.137 %	[-]	-8.2765	0.62	0.89
GO:0055082	cellular chemical homeostasis	0.197 %		-6.8239	0.72	0.25
GO:0001932	regulation of protein phosphorylation	0.166 %		-7.5467	0.58	0.26
-GO:0001934	positive regulation of protein phosphorylation	0.039 %	[-]	-7.3893	0.49	0.88
-GO:0010562	positive regulation of phosphorus metabolic process	0.053 %	[-]	-6.7670	0.50	0.79
-GO:0042325	regulation of phosphorylation	0.184 %	[-]	-7.4522	0.60	0.87
-GO:0042327	positive regulation of phosphorylation	0.047 %	[-]	-6.9586	0.49	0.97
-GO:0019220	regulation of phosphate metabolic process	0.391 %	[-]	-6.4168	0.59	0.93
-GO:0045937	positive regulation of phosphate metabolic process	0.053 %	[-]	-6.7670	0.49	0.99
-GO:0021461	positive regulation of protein modification process	0.048 %	[-]	-6.0942	0.52	0.90

**Fig. 3** Visualizations of Biological Process GO annotations using REVIGO: scatterplot and table views. The data-set used was imported from GOrilla (see legend of Fig. 2). We used the default settings of the REVIGO tool. (a) The scatterplot view visualizes the GO terms in a “semantic space” where the more similar terms are positioned closer together [20]. The color of the bubble reflects the p-value obtained in the GOrilla analysis,



**Fig. 4** Visualizations of Biological Process GO annotations using REVIGO: TreeMap (a), interactive graph (b) and word cloud views (c). The dataset used was imported from GOrilla (see legend of Fig. 2). We used the default settings of the tool



The results are shown in Figs. 3 and 4. The various visualization styles highlight the GO terms that are enriched in the input dataset.

**RedundancyMiner** [21] is another tool that focuses on non-redundant terms in a large list of enriched GO terms, producing a Clustered Image Map (CIM) as a result. It is a part of a larger pipeline: RedundancyMiner relies on GOMiner input and on CIM miner for visualization. In particular, RedundancyMiner performs Fisher's exact tests for each pair of GO terms in the datasets, calculating whether the two sets of genes annotated with these GO terms are overlapping. A symmetrical matrix of these p-values is subsequently analyzed to arrive to a set of GO terms that are most independent, and therefore least redundant.

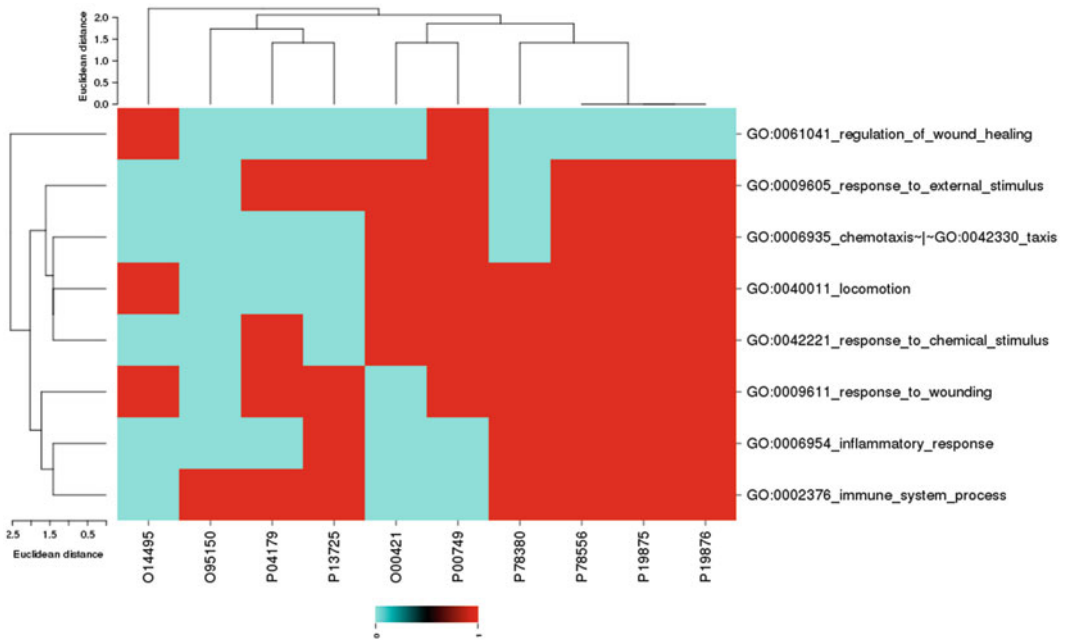
To perform the analysis, we started with the same file as for the two tools described above. First, we generated two files using a custom Python script: (1) a file containing all the genes in the array and (2) a file containing the genes that are over or underexpressed, labeled with "1" or "-1," respectively. Of note, Python is not necessary for RedundancyMiner and these files could be generated otherwise. Second, we put these files as input for the GOMiner tool (<http://discover.nci.nih.gov/gominer/GoCommandWebInterface.jsp>). We selected the databases that contain *Homo sapiens* data and as the organism we set *H. sapiens*. The remaining parameters were the defaults in the tool. Third, we used the resulting folder as the working folder for RedundancyMiner and we ran the analysis in default mode. Finally, we visualized the resulting CIM file using cimMiner [29], available at <http://discover.nci.nih.gov/cimminer/home.do>, in single matrix mode.

Results of our example analysis are shown in Fig. 5. Even with the stringent threshold of requiring the  $\log_2$  fold change greater than 5, similar trends in significant GO terms are visible as shown with the remaining two tools.

---

## 5 Choice of Visualization

Above, we have outlined some of the currently available software tools that can visualize a set of GO terms. We have also argued that a good visualization is an effective means of discovering underlying trends in the data in an unbiased fashion; an appropriate visual display is also imperative when communicating the results to others. The question of which software tool to apply should be addressed keeping these goals in mind. A related yet distinct question is which specific visualization method to choose. Here, we give a summary of the available options. Of note, the authors of this text are also the developers of REVIGO [20], a versatile visualization tool, which implements several of the approaches listed below.



**Fig. 5** A visualization of a set of Biological Process GO annotations using RedundancyMiner. The dataset used is a microarray transcription profiling of human peripheral blood mononuclear cells after treatment with *Staphylococcus aureus*. For this visualization, we focus on genes that had log<sub>2</sub> fold change greater than 5

1. *Graphs/networks*. The GO graph consists of *nodes* (here, Gene Ontology terms) and *edges* (here, parent–child relationships), which connect the nodes and which have directionality. Nodes and edges can have multiple attributes that can be visualized. For instance, the enrichment of a GO term in a user’s experiment may be shown as a color of a node (Fig. 2). Importantly, the spatial arrangement of the nodes on the final plot is called a *layout*, and is often created to suggest related clusters of nodes by placing similar nodes closer together. Such approaches are reviewed and demonstrated by Merico et al. [30]; tools like Cytoscape [9] support a variety of visual layouts.
  - (a) A special case of a layout is a tree-like display that highlights the ‘levels’ in the Gene Ontology and the parent–child relationships between terms (e.g., Fig. 2). These levels (determining the *depth* of a node in the graph) are often used as a measure for how general the GO term is. However, this may be misleading in some instances—for example, the Molecular Function ontology is more shallow than the Biological Process ontology—and we therefore recommend the use of the *information content* (IC) measure [31] for this purpose. This is defined as the negative

logarithm of the relative frequency of the respective term annotations in some underlying database, such as the UniProt-GOA [32].

2. *Semantic similarity space*. Various mathematical methods measure the semantic similarity between pairs of GO terms, such as SimRel [33]; see ref. 34 for a review. If the first term in a pair is a direct parent, child or sibling of the second term, their semantic similarity will be very high. However, also the more distantly related terms will show some degree of similarity, as long as they reside in a common branch of the GO tree structure. Many such pairwise similarities within a group of GO terms can be processed by a projection technique, such as *principal components analysis* (PCA) or *multidimensional scaling*. The resulting plots preserve as much of the original pairwise distances as possible, while showing all supplied GO terms in a two-dimensional plane. The main visualization in REVIGO is based on this approach (Fig. 3a).
3. *Treemaps*. Hierarchical diagrams consisting of tiles subdivided into smaller tiles. Treemaps are good for interactive exploration, as they can be ‘zoomed in’ by clicking a tile and revealing finer levels of subdivisions. Here, tiles can be GO terms and the subdivisions their child terms. The tile sizes may correspond to some measure of importance of GO terms to the user, such as enrichment or p-values. REVIGO has an implementation of this visualization approach (Fig. 4b).
4. *Word clouds*. A display with text shown in various sizes and possibly colors. Here, the individual words or short phrases may be the names of the GO terms or some keywords associated to the GO terms. The text size/color may convey the importance to the user (enrichment), or in some instances generality of a GO term (see *information content* above). This visualization method is implemented in GOSummaries and REVIGO (Fig. 4c).
5. *Clustered Heatmaps*. Two-dimensional grids of values, wherein the rows and/or columns are clustered to reveal the ‘block structure’ in the data. Clustered heatmaps are often used for showing high-dimensional data in biology, but rarely so for GO terms. In fact, this could be done to show the GO terms’ similarity based on what genes are annotated to them, or on the terms’ semantic similarity (which is defined by the structure of the GO graph). An example implementation can be found in RedundancyMiner (Fig. 5).

In addition to the above, many of the tools specializing in GO enrichment testing (or in other analyses of large-scale biological data) often come bundled with visualizations that include GO as an important context. Examples include the Bioconductor packages GOexpress, GOfunction and GOSim. In addition, it is often

possible to customize such displays in more detail by manually passing the GO data to a dedicated visualization software, such as the *ggplot2* package [35] in R, or to *gnuplot* software. For example, a specialized software to draw treemaps can be made to display GO enrichments from a biological experiment via a script that prepares the data in a correct format [36]. REVIGO will draw bubble charts where the GO terms are displayed in a semantic similarity space [20], and it can export a *ggplot2* script which is further customizable for e.g., font sizes, colors, and line styles; it can similarly export a graph to be further customized in Cytoscape.

---

## 6 Concluding Remarks and Outlook

In summary, we outline several tools that biologists can use to visualize sets of Gene Ontology terms and uncover novel and interesting trends in their experimental data. We anticipate that the future will bring even more massive biological data sets, which will have several consequences. First, the lists of interesting GO terms will grow in length, as larger sample sizes afford more statistical power to detect associations. Therefore, refinements of the existing approaches that address redundant GO terms [20, 21] will come in useful. Second, the visualization software will need to deal with more than a single list of enriched GO terms. While some current tools can display such results from multiple experiments side-by-side, e.g. BACA [26], tools will be needed that can integrate such lists and extract patterns across them. Finally, while GO is a prominent example of an ontology used by biologists, it is far from the only one [37]—over 100 biomedical ontologies exist that describe environments, phenotypes, and chemical entities (*see* Chap. 19) [38]. We foresee substantial developments in the tools that can summarize and visualize results of various biological experiments in the context of such emerging ontologies.

---

## Acknowledgements

We acknowledge the support of the European Commission via grants MAESTRA (ICT-2013-612944) and InnoMol (FP7-REGPOT-2012-2013-1-316289), and of the Croatian Science Foundation grant MultiCaST (# 5660). Open Access charges were funded by the University College London Library, the Swiss Institute of Bioinformatics, the Agassiz Foundation, and the Foundation for the University of Lausanne.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use,

duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

## References

1. Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29
2. Rivals I, Personnaz L, Taing L et al (2007) Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* 23:401–407
3. du Plessis L, Škunca N, Dessimoz C (2011) The what, where, how and why of gene ontology--a primer for bioinformaticians. *Brief Bioinform* 12:723–735
4. Lespinet O, Wolf YI, Koonin EV et al (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res* 12:1048–1059
5. Gaudet P, Škunca N, Hu JC, Dessimoz C (2016) Primer on the gene ontology. In: Dessimoz C, Škunca N (eds) *The gene ontology handbook. Methods in molecular biology*, vol 1446. Humana Press. Chapter 3
6. Carbon S, Ireland A, Mungall CJ et al (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics* 25:288–289
7. Binns D, Dimmer E, Huntley R et al (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25:3045–3046
8. Vercruyse S, Venkatesan A, Kuiper M (2012) OLSVis: an animated, interactive visual browser for bio-ontologies. *BMC Bioinformatics* 13:116
9. Smoot ME, Ono K, Ruschinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27:431–432
10. Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. In: *Third international AAAI conference on weblogs and social media*
11. Batagelj V (2011) *Exploratory social network analysis with Pajek (Structural analysis in the social sciences)*. Cambridge University Press, Cambridge
12. Merico D, Isserlin R, Bader GD (2011) Visualizing gene-set enrichment results using the Cytoscape plug-in enrichment map. *Methods Mol Biol* 781:257–277
13. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449
14. Eden E, Navon R, Steinfeld I et al (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48
15. Bastos HP, Sousa L, Clarke LA et al (2015) GRYFUN: a web application for GO term annotation visualization and analysis in protein sets. *PLoS One* 10:e0119631
16. Sun H, Fang H, Chen T et al (2006) GOFFA: gene ontology for functional analysis--a FDA gene ontology tool for analysis of genomic and proteomic data. *BMC Bioinformatics* 7(Suppl 2):S23
17. Herrmann C, Bérard S, Tichit L (2009) SimCT: a generic tool to visualize ontology-based relationships for biological objects. *Bioinformatics* 25:3197–3198
18. *KEGG Mapper*. [http://www.genome.jp/kegg/tool/map\\_pathway2.html](http://www.genome.jp/kegg/tool/map_pathway2.html)
19. Uchiyama T, Irie M, Mori H et al (2015) FuncTree: functional analysis and visualization for large-scale omics data. *PLoS One* 10:e0126967
20. Supek F, Bošnjak M, Škunca N et al (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800
21. Zeeberg BR, Liu H, Kahn AB et al (2011) RedundancyMiner: de-replication of redun-

- dant GO categories in microarray and proteomics analysis. *BMC Bioinformatics* 12:52
22. Reimand J, Arak T, Vilo J (2011) g:Profiler—a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* 39:W307–W315
  23. Bauer S, Grossmann S, Vingron M et al (2008) Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* 24:1650–1651
  24. Bauer S, Gagneur J, Robinson PN (2010) GOing Bayesian: model-based gene set analysis of genome-scale data. *Nucleic Acids Res* 38:3523–3532
  25. Kolde R, Vilo J (2015) GOSummaries: an R Package for visual functional annotation of experimental data. *F1000Research* 4:574
  26. Fortino V, Alenius H, Greco D (2015) BACA: bubble chArt to compare annotations. *BMC Bioinformatics* 16:37
  27. Kobayashi SD, Braughton KR, Palazzolo-Ballance AM et al (2010) Rapid neutrophil destruction following phagocytosis of *Staphylococcus aureus*. *J Innate Immun* 2:560–575
  28. Petryszak R, Burdett T, Fiorelli B et al (2014) Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res* 42:D926–D932
  29. Weinstein JN, Myers TG, O'Connor PM et al (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* 275:343–349
  30. Merico D, Gfeller D, Bader GD (2009) How to visually interpret biological data using networks. *Nat Biotechnol* 27:921–924
  31. Lord PW, Stevens RD, Brass A et al (2003) Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics* 19:1275–1283
  32. Dimmer EC, Huntley RP, Alam-Faruque Y et al (2012) The UniProt-GO Annotation database in 2011. *Nucleic Acids Res* 40:D565–D570
  33. Schlicker A, Domingues FS, Rahnenführer J et al (2006) A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* 7:302
  34. Mazandu GK, Mulder NJ (2013) Information content-based gene ontology semantic similarity approaches: toward a unified framework theory. *BioMed Res Int* 2013:292063
  35. Wickham H (2009) ggplot2: elegant graphics for data analysis
  36. Baehrecke EH, Dang N, Babaria K et al (2004) Visualization and analysis of microarray and gene ontology data with treemaps. *BMC Bioinformatics* 5:84
  37. Smith B, Ashburner M, Rosse C et al (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 25:1251–1255
  38. Furnham N (2016) Complementary sources of protein functional information: the far side of GO. In: Dessimoz C, Škunca N (eds) *The gene ontology handbook. Methods in molecular biology*, vol 1446. Humana Press. Chapter 19