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 the various methods to visualize large, difficult-to-interpret lists of GO terms. We catalogue 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.

Keywords: 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 distill the most unique and/or relevant information. Finally, these methods produce various styles of visualization that can aid in the 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 examples of use 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 up- or down-regulated in response to a particular treatment. This indicates that a systems-level change in the experimental model has occurred, and it 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 co-localize 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 could 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 chapter “Primer on the GO” 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\(^1\) (Figure 1). This complex structure, along with its large size—the GO has thousands of nodes—make it rather challenging to display the part(s) of the GO of interest. Such parts could be, for instance, derived from a list of GO terms found to be enriched in a gene expression experiment.

\(^1\) http://geneontology.org/page/download-ontology
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. 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 errors, 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, hopefully 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 interrelationships between the GO terms alongside the statistical support for each in the experimental data, one would help avoid focusing on outlying—and perhaps spurious—results. Similarly, one could be made aware of the common pitfall where one GO term is chosen, while other somewhat similar and equally 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 will systematize and describe the currently available tools for visualizing sets of GO annotations. Additionally, we will 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 (5) and QuickGO (6). Of note, OLSVis (7) 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 weren’t designed specifically for GO, they tend to be more complicated to use. See Cytoscape (8), Gephi (9), and Pajek (10).
   a. Of note, there are Cytoscape plugins specialized for handling groups of GO terms: EnrichmentMap (11) and BINGO (12).

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 (13), GRYFUN (14), GOFFA (15), and SimCT (16).
   a. In addition to the GO, similar tools are available which can highlight the individual members in displayed KEGG pathways (17); the pathways can also be shown in a KEGG BRITE functional hierarchy with FuncTree (18).

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 (19) and RedundancyMiner (20).
   a. Some provisions for this are made in g:Profiler (21), which collapses similar GO terms.
   b. The Ontologizer (22) can perform a statistical test for enrichment that accounts for the parent-child redundancy (23) 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 coloured tiles), as in REVIGO (19) or GOSummaries (24). Additionally, several tools now support the display of multiple GO enrichment analyses side-by-side; see BACA (25) or GOSummaries. SimCT (16) can display subtrees of other biomedical ontologies in addition to the GO.
3.1 Case studies with selected tools

**GOrilla (13)** 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. The lower the displayed p-value, the more probable it is for the 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 (26), obtained from the Expression Atlas (27) 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 Figure 2. Based on the colour of the boxes, the user can clearly visualize which GO terms are enriched, and the connecting lines describe their relationship to other terms in the GO graph.

**REVIGO (19)** analyzes—the often large—list of significant GO terms and removes the redundant terms, to further narrow the search to a set of non-redundant and highly significant GO terms. Briefly, REVIGO creates clusters of GO terms that are semantically similar, and selects one representative for each cluster.

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: 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; interactive graph that connects the user-supplied set of GO terms based on the structure of the GO hierarchy; a TreeMap where terms are clustered by colour; and a word cloud highlights the most frequent words 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. The results are shown in Figures 3 and 4. The various styles of the visualization highlight the GO terms that are enriched in the input dataset.

**RedundancyMiner (20)** is another tool to focus on non-redundant terms in a large list of enriched GO terms, producing a Clustered Image Map (CIM) as a result. It is 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 analysed 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, labelled 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 (28), available at http://discover.nci.nih.gov/cimminer/home.do, in single matrix mode.

Results of our example analysis are shown in Figure 5. Even with the stringent threshold of requiring the log2 fold change greater than 5, similar trends in significant GO terms are visible as shown with the remaining two tools.

4 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, while avoiding biases; 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 (19), a versatile visualization tool, which implements several of the approaches listed below.

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 may 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 (Figure 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. (29); tools like Cytoscape (8) 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., Figure 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 much more “shallow” than the Biological Process ontology—and we therefore recommend the use of the information content (IC) measure (30) 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 (31).

2. **Semantic similarity space.** Various mathematical methods measure the semantic similarity between pairs of GO terms, such as SimRel (32); see (33) 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 (Figure 3, panel A).

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 (Figure 4, panel B).

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 perhaps generality of a GO term (see information content above). This visualization method is implemented in GOSummaries and REVIGO (Figure 4, panel C).

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 (Figure 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 may 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 (34) 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 (35). REVIGO will draw bubble charts where the GO terms are displayed in a semantic similarity space (19), 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.

5 Concluding remarks and outlook

In summary, we have outlined 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 (19, 20) 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, eg. BACA (25), 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's far from the only one (36)—over 100 biomedical ontologies exist that describe e.g., environments, phenotypes, and chemical entities (see chapter “Beyond the GO”). We foresee substantial developments in the tools that can summarize and visualize results of various biological experiments in the context of such emerging ontologies.

References

1. M. Ashburner, C.A. Ball, J.A. Blake, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nature genetics. 25, 25–29.
2. I. Rivals, L. Personnaz, L. Taing, et al. (2007) Enrichment or depletion of a GO category within a class of genes: which test?, Bioinformatics . 23, 401–407.
3. L. du Plessis, N. Skunca, and C. Dessimoz (2011) The what, where, how and why of gene ontology--a primer for bioinformaticians, Briefings in bioinformatics. 12, 723–735.
4. O. Lespinet, Y.I. Wolf, E.V. Koonin, et al. (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes, Genome research. 12, 1048–1059.
5. S. Carbon, A. Ireland, C.J. Mungall, et al. (2009) AmiGO: online access to ontology and annotation data, Bioinformatics . 25, 288–289.
6. D. Binns, E. Dimmer, R. Huntley, et al. (2009) QuickGO: a web-based tool for Gene Ontology searching, Bioinformatics . 25, 3045–3046.
7. S. Vercruysse, A. Venkatesan, and M. Kuiper (2012) OLSVis: an animated, interactive visual browser for bio-ontologies, BMC bioinformatics. 13, 116.
8. M.E. Smoot, K. Ono, J. Ruscheinski, et al. (2011) Cytoscape 2.8: new features for data integration and network visualization, Bioinformatics . 27, 431–432.
9. M. Bastian, S. Heymann, and M. Jacomy (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks, In: Third International AAAI Conference on Weblogs and Social Media .
10. V. Batagelj (2011) Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences), Cambridge University Press.
11. D. Merico, R. Isserlin, and G.D. Bader (2011) Visualizing gene-set enrichment results using the Cytoscape plug-in enrichment map, Methods in molecular biology . 781, 257–277.
12. S. Maere, K. Heymans, and M. Kuiper (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks, Bioinformatics . 21, 3448–3449.
13. E. Eden, R. Navon, I. Steinfeld, et al. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists, BMC bioinformatics. 10, 48.
14. H.P. Bastos, L. Sousa, L.A. Clarke, et al. (2015) GRYFUN: a web application for GO term annotation visualization and analysis in protein sets, PloS one. 10, e0119631.
15. H. Sun, H. Fang, T. Chen, 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.
16. C. Herrmann, S. Bérard, and L. Tichit (2009) SimCT: a generic tool to visualize ontology-based relationships for biological objects, Bioinformatics. 25, 3197–3198.
17. KEGG Mapper, http://www.genome.jp/kegg/tool/map_pathway2.html.
18. T. Uchiyama, M. Irie, H. Mori, et al. (2015) FuncTree: Functional Analysis and Visualization for Large-Scale Omics Data, PloS one. 10, e0119631.
19. F. Supek, M. Bošnjak, N. Škunca, et al. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms, PloS one. 6, e21800.
20. B.R. Zeeberg, H. Liu, A.B. Kahn, et al. (2011) RedundancyMiner: De-replication of redundant GO categories in microarray and proteomics analysis, BMC bioinformatics. 12, 52.
21. J. Reimand, T. Arak, and J. Vilo (2011) g:Profiler—a web server for functional interpretation of gene lists (2011 update), Nucleic acids research. 39, W307–W315.
22. S. Bauer, S. Grossmann, M. Vingron, et al. (2008) Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration, Bioinformatics. 24, 1650–1651.
23. S. Bauer, J. Gagneur, and P.N. Robinson (2010) GOing Bayesian: model-based gene set analysis of genome-scale data, Nucleic acids research. 38, 3523–3532.
24. R. Kolde and J. Vilo (2015) GOsummaries: an R Package for Visual Functional Annotation of Experimental Data, F1000Research. 4, 574.
25. V. Fortino, H. Alenius, and D. Greco (2015) BACA: bubble chArt to compare annotations, BMC bioinformatics. 16, 37.
26. S.D. Kobayashi, K.R. Braughton, A.M. Palazzolo-Ballance, et al. (2010) Rapid neutrophil destruction following phagocytosis of Staphylococcus aureus, Journal of innate immunity. 2, 560–575.
27. R. Petryszak, T. Burdett, B. Fiorelli, et al. (2014) Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments, Nucleic acids research. 42, D926–D932.
28. J.N. Weinstein, T.G. Myers, P.M. O’Connor, et al. (1997) An information-intensive approach to the molecular pharmacology of cancer, Science. 275, 343–349.
29. D. Merico, D. Gfeller, and G.D. Bader (2009) How to visually interpret biological data using networks, Nature biotechnology. 27, 921–924.
30. P.W. Lord, R.D. Stevens, A. Brass, et al. (2003) Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation, Bioinformatics. 19, 1275–1283.
31. E.C. Dimmer, R.P. Huntley, Y. Alam-Faruque, et al. (2012) The UniProt-GO Annotation database in 2011, Nucleic acids research. 40, D565–D70.
32. A. Schlicker, F.S. Domingues, J. Rahnenführer, et al. (2006) A new measure for functional similarity of gene products based on Gene Ontology, BMC bioinformatics. 7, 302.
33. G.K. Mazandu and N.J. Mulder (2013) Information content-based gene ontology semantic similarity approaches: toward a unified framework theory, BioMed research international. 2013, 292063.
34. H. Wickham (2009) ggplot2: elegant graphics for data analysis.
35. E.H. Baehrecke, N. Dang, K. Babaria, et al. (2004) Visualization and analysis of microarray and gene ontology data with treemaps, BMC bioinformatics. 5, 84.
36. B. Smith, M. Ashburner, C. Rosse, et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, Nature biotechnology. 25, 1251–1255.
Figure 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 annotated with all the parent terms of that GO term.

Figure 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”.
Figure 3. Visualizations of Biological Process GO annotations using REVIGO: scatterplot and table views. The dataset used was imported from GOrilla (see legend of Figure 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 (19). The colour of the bubble reflects the p-value obtained in the GOrilla analysis, 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.
Figure 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 Figure 2). We used the default settings of the tool.
Figure 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_2 fold change greater than 5.