GO-Compass: Visual Navigation of Multiple Lists of GO terms

T. Harbig[1] ⓘ, M. Witte Paz[1] ⓘ and K. Nieselt[1] ⓘ

[1] University of Tübingen, Institute for Bioinformatics and Medical Informatics, Tübingen, Germany

Abstract

Analysis pipelines in genomics, transcriptomics, and proteomics commonly produce lists of genes, e.g., differentially expressed genes. Often these lists overlap only partly or not at all and contain too many genes for manual comparison. However, using background knowledge, such as the functional annotations of the genes, the lists can be abstracted to functional terms. One approach is to run Gene Ontology (GO) enrichment analyses to determine over- and/or underrepresented functions for every list of genes. Due to the hierarchical structure of the Gene Ontology, lists of enriched GO terms can contain many closely related terms, rendering the lists still long, redundant, and difficult to interpret for researchers.

In this paper, we present GO-Compass *(Gene Ontology list comparison using Semantic Similarity), a visual analytics tool for the dispensability reduction and visual comparison of lists of GO terms. For dispensability reduction, we adapted the* RE-VIGO *algorithm, a summarization method based on the semantic similarity of GO terms, to perform hierarchical dispensability clustering on multiple lists. In an interactive dashboard,* GO-Compass *offers several visualizations for the comparison and improved interpretability of GO terms lists. The hierarchical dispensability clustering is visualized as a tree, where users can interactively filter out dispensable GO terms and create flat clusters by cutting the tree at a chosen dispensability. The flat clusters are visualized in animated treemaps and are compared using a correlation heatmap, UpSet plots, and bar charts.*

With two use cases on published datasets from different omics domains, we demonstrate the general applicability and effectiveness of our approach. In the first use case, we show how the tool can be used to compare lists of differentially expressed genes from a transcriptomics pipeline and incorporate gene information into the analysis. In the second use case using genomics data, we show how GO-Compass *facilitates the analysis of many hundreds of GO terms. For qualitative evaluation of the tool, we conducted feedback sessions with five domain experts and received positive comments.* GO-Compass *is part of the Tue-Vis Visualization Server as a web application available at* https://go-compass-tuevis.cs.uni-tuebingen.de/

CCS Concepts

• Human-centered computing → Scientific visualization; • Applied computing → Bioinformatics;

1. Introduction

Lists of genes are one of the most common analysis results in bioinformatics. Often, multiple lists of genes are produced in large-scale molecular biology experiments, so-called "omics" studies. These lists are interpreted and compared by bioinformaticians and biologists to gain insight into the structure, function, and dynamics of an organism.

In genomics, mutations or single nucleotide polymorphisms (SNPs) located in genes are of interest since they can alter the function of genes. For example, when comparing cancerous to non-cancerous cells, researchers aim to identify all genes containing mutations and study how these mutations may affect the function of the cell. Moreover, in microbiological studies, bacterial strains are compared which have specific properties, such as antibiotic resistance. Studying lists of genes showing SNPs between the strains can enhance insights into resistance mechanisms [ASJ*16].

In transcriptomics and proteomics the expression of genes is studied, i.e., the intensity with which genes are transcribed into mRNA and translated into proteins. Usually, the expression of altered organisms (with induced mutations) and/or the expression at different conditions, such as timepoints or different environmental conditions, is studied. Multiple biological replicates are created for each condition, which allows calculating lists of significantly differentially expressed genes associated with fold changes between conditions. Moreover, genes can be clustered by their behavior across conditions resulting in a list of genes per cluster.

Comparing lists of genes is a non-trivial task. While sometimes genes can be compared individually, for example, if a gene is contained in multiple lists of differentially expressed genes, usually the lists overlap only partly or not at all. Therefore, gene lists are often compared in terms of their functional composition using underlying knowledge such as pathways or gene ontology terms associated with the genes [HCI*04].

The gene ontology (GO) is a graph-based, species-agnostic representation of information about the functions of genes, with respect to three domains: molecular function, biological process, and cellular component. It is organized in single GO terms encoding for specific functions. Every gene can be annotated with one or multiple GO terms. Each domain of the ontology is organized roughly hierarchically in a directed acyclic graph, which means that terms higher up in the hierarchy are more general terms of function than terms further down the hierarchy (see Suppl. Figure S1 for an example). This structure is both machine-readable and human-readable, which allows applying a range of methods that make use of the ontology and produce easily interpretable results. GO terms can be used to summarize the overall functions present in a list of genes, which is often done using gene ontology enrichment, a method to find statistically overrepresented GO terms [TCK*03, KZP*18].

Not every gene is annotated with the same level of specificity. While some genes are annotated with very specific GO terms, others are annotated with very general terms. Therefore, the GO annotation of genes is usually propagated, which means that a gene is not only annotated with the specific GO term but also with all its parents to allow finding enriched GO terms even if not all the annotations are very specific. In practice, this means that a GO enrichment often results in long lists of GO terms containing many, possibly nearly identical terms. For example, if a very specific term, such as *riboflavin metabolic process* (GO:0006771, Suppl. Figure S1), is enriched, its parent terms including *water-soluble vitamin metabolic process* (GO:0006767), *vitamin metabolic process* (GO:0006766), and *small molecule metabolic process* (GO:0044281) might be present as well, but do not add much information.

While the structure of the GO graph causes this redundancy, it can also be used to combat it. The graph can be used to calculate the semantic similarity of GO terms which serves as a measurement to filter out highly similar terms [LSBG03]. One of the most simple similarity measures is the number of edges in the shortest path between two GO terms, but usually, more advanced methods are preferred (see Section 3.2). Our tool `GO-Compass` uses semantic similarity to hierarchically cluster GO terms. The tool lets users decide on the desired level of redundancy by visualizing the clustering result and allowing them to decide on a cutoff. The filtered GO term lists are then visualized in an interactive dashboard that contains comparative visualizations including an animated treemap, a correlation heatmap, an UpSet plot [LGS*14], and a comparative bar chart. We demonstrate how `GO-Compass` provides an efficient approach applicable in the omics domains by showing two use cases on published datasets. Moreover, in structured expert feedback sessions, we qualitatively evaluated our tool.

2. Related Work

For the downstream analysis and comparison of lists of GO terms various packages in R or Python have been developed. The Python package `goatools` implements functionalities such as statistical testing, summarizing gene lists, GO enrichment, and semantic similarity calculation [KZP*18]. The R package `Viseago` provides a comparative heatmap visualization of multiple sets of GO terms

while clustering them by their semantic similarity [BJHA19]. However, the initial list of GO terms is not reduced, which can lead to very large heatmaps. `GOsummaries` is an R package that focuses on creating word clouds of GO terms [KV15]. It allows the processing of the data using filtering, clustering, and dimension reduction. While these packages are extremely powerful, since they can be easily extended and combined with other packages, the requirement of programming skills limits the size of the user base. However, they can serve as the basis for the backends of other tools.

Furthermore, tools with a graphical interface have been developed, which can be accessed by a broader audience. `REVIGO` is a tool for reducing the redundancy in a list of GO terms and for visualizing the results [SBŠŠ11]. The tool is based on a clustering algorithm, which ranks the GO terms by their dispensability based on their semantic similarity, their frequency, their *p*-values, and their relationship to each other in the GO graph. The visualizations consist of a treemap, a scatterplot, and a table. `REVIGO` is one of the most frequently used tools for reducing the redundancy and visualizing lists of GO-terms as reflected by its citation count. Moreover, the clustering result is used as an input for other tools, such as `CirGO`, a visualization tool that uses the `REVIGO` clustering results for visualization with sunburst charts instead of treemaps [KLS*19].

`BACA` is a tool for the comparative visualization of GO terms [FAG15]. It does not only visualize GO terms, but it also includes information about gene expression. Each row in the visualization corresponds to a GO term and each column to a condition. For each condition and term, two circles are drawn corresponding to up- and downregulated genes. The size of the circles corresponds to the number of genes. The tool itself does not decrease the redundancy of lists of GO terms. Moreover, the number of genes might be a misleading value, as the size of sets of genes associated with different GO terms varies immensely due to the hierarchical structure of the GO graph.

3. Method

`GO-Compass` is a domain-specific tool for the comparison of lists of GO terms that arise from the analysis of large-scale omics experiments. The tool implements semantic similarity clustering of GO terms to reduce the redundancy in the lists. In a visualization dashboard, the clustering result is visualized and multiple visualizations are offered for the comparison of the reduced GO term lists.

`GO-Compass` makes use of several packages and methods introduced in Section 2. For GO enrichment and semantic similarity calculation, the `goatools` package is used. Moreover, `GO-Compass` includes an adapted re-implementation of the clustering algorithm of `REVIGO` that can handle multiple lists of GO terms. Similar to `REVIGO`, it implements a treemap for the visualization of the results of each list with an animated transition between treemaps of different lists. Gene information is incorporated using a glyph in the treemap. However, our visualization is not limited to gene lists from differential expression experiments, it can also be used for lists of GO terms and genes from other sources.

3.1. Data Input and Preprocessing

`GO-Compass` offers an internal enrichment method that can be used by submitting lists of genes. The genes can optionally be associated with numerical values (e.g., fold changes when calculating differential expression) that are used for visualization. Moreover, users can submit lists of GO terms with associated *p*-values arising from other enrichment methods. The GO term lists are arranged in a table where each column contains the *p*-values of one of the lists. In this case, the lists of genes are optional.

Furthermore, `GO-Compass` requires one or multiple background files containing the association of genes with GO terms for the enrichment and the clustering of GO terms. The lists of genes, GO terms, and backgrounds can originate from different species to enable inter-species comparisons or from different types of omics data (such as transcriptomic and proteomics) for multi-omics comparisons. `GO-Compass` propagates the backgrounds, which means that for every GO term, all parent GO terms are associated with its set of genes. Propagation can be disabled by the user if the lists have been previously propagated.

With a *p*-value filter, users can define their own significance level to filter out GO terms. Thus, only the GO terms are kept whose *p*-value is lower than the specified filter in at least one list. Moreover, a method for calculating the semantic similarity of GO terms can be selected, which is described in Section 3.2.

If lists of genes but no custom enrichment results have been submitted, `GO-Compass` performs GO enrichment with the Python package `goatools` [KZP*18] using the propagated background. For every list, the enrichment is calculated using Fisher's exact test. The resulting *p*-values are corrected using Benjamini-Hochberg's *False Discovery Rate* (FDR) [BH95].

3.2. Semantic Similarity

`GO-Compass` clusters GO terms based on their semantic similarity. There are various definitions of semantic similarity. The simplest measure is the shortest path between two GO terms *A* and *B* in the GO graph (Edge-distance measurement). A problem with this approach is that it assumes uniform distances in the graph. Since the GO graph is highly imbalanced, `GO-Compass` offers three other commonly used measures, the Resnik, Lin, and Wang semantic similarity [Res99, LO98, WDP*07]. Resnik semantic similarity is a node-based measure, defined as the information content (IC) of the lowest common subsumer of two GO terms, also called the *most informative common ancestor* (MICA) [Res99]. Let $S(A,B)$ denote the set of common ancestors of terms *A* and *B* and $p(t)$ the probability of term *t*. Then the Resnik similarity of *A* and *B* is defined as:

$$sim_{Resnik}(A,B) = IC(MICA) = \max_{t \in S(A,B)} (-\log(p(t))) \quad (1)$$

The Lin semantic similarity represents a normalized version of the Resnik similarity [LO98]:

$$sim_{Lin}(A,B) = \frac{2IC(MICA)}{IC(A) + IC(B)} \quad (2)$$

Wang semantic similarity is a graph-based method, which considers the contribution of all ancestors to the semantics of a term, with closer terms contributing more than more distant terms [WDP*07]. The contribution of a term *t* to the semantics of term *A* is defined by the *S*-value:

$$S_A(t) = \begin{cases} 1 & \text{if } t = A \\ \max\{w_e \times S_A(t') \| t' \in \text{children of t}\} & \text{else} \end{cases} \quad (3)$$

w_e is the contribution factor of edge *e* linking *t* and *t'*. The semantic value is then the sum of all *S*-values of the set T_A which contains all ancestors of *A*:

$$SV(A) = \sum_{t \in T_A} S_A(t) \quad (4)$$

The Wang similarity between two terms *A* and *B* is then defined as:

$$sim_{Wang}(A,B) = \frac{\sum_{t \in T_A \cap T_B} (S_A(t) + S_B(t))}{SV(A) + SV(B)} \quad (5)$$

In this equation, the semantic contributions of the shared ancestors of *A* and *B* are summed up and divided by the total contribution of each term's ancestors.

`GO-Compass` uses `goatools` for the calculation of semantic similarity with these four measures [KZP*18]. At our request, the authors of `goatools` have also added Wang's method to their repertoire.

3.3. Algorithm

`GO-Compass` applies an adapted version of the `REVIGO` algorithm to perform hierarchical dispensability clustering separately for each GO domain. For every pair of GO terms, semantic similarity is computed and the pairs are sorted. Starting with the most similar pair *A* and *B*, a term is rejected, i.e., set as a child of the other term if one of several criteria, is fulfilled. The criteria are based on the terms' frequencies, their significance, and their relation in the graph. More concretely, the rejection criteria for term *A* are (applied in this order, and vice versa for *B*):

1. *A* has a frequency in the background data > 5% and the frequency of *A* is bigger than *B*.
2. The majority of *p*-values of *A* is less significant (difference >5% of the *p*-value range in the data).
3. *A* is the parent of *B* and contains more than 75% of its genes.
4. *A* is a child of *B*.
5. (Pseudo) random rejection: A term is rejected at random using the numbers of the GO ID corresponding to term *A* as a seed to guarantee that the result is consistent when the tool is run multiple times.

Criterion 3 guarantees that if two GO terms in a parent-child relationship are associated with a similar set of genes, the one that is less specific, i.e., the parent is rejected, while criterion 4 rejects the child if the sets of genes are different.

The semantic similarity at which a term is rejected corresponds to the term's dispensability. The term's uniqueness is calculated in the following way, where *T* corresponds to the set of all present GO terms:

$$uniqueness(A) = 1 - \frac{\sum_{t \in T} sim(A,t)}{|T|} \quad (6)$$

While REVIGO offers the Resnik, Lin, and two further IC-based methods, our adapted algorithm uses additionally and by default the Wang method, which takes the semantics of all ancestors of the GO terms into account to calculate their similarity. Moreover, the Wang method does not rely on the frequency of GO terms in the annotation of a species and is thus especially suited for comparisons using multiple different annotation backgrounds.

Furthermore, the *p*-value rejection criterion has been adapted to handle multiple *p*-values and thus to cluster terms from many lists of GO terms. Instead of rejecting the less significant *p*-value, the term for which the majority of p-values is less significant is rejected. The other rejection criteria have been implemented as described in the original REVIGO publication. Moreover, while the original algorithm stops clustering at a dispensability cutoff value provided by the user, GO-Compass always clusters all terms. The resulting structure is a tree with rejected terms as child nodes of accepted terms. The closer a term is to the root, the less dispensable it is.

4. Visualization

Based on our survey of related work we created multiple objectives for the development of GO-Compass:

1. Show all clustered GO terms to increase the explainability of the results. Users should see how their choice on a dispensability cutoff affects the final list of GO terms. Moreover, make terms that did not pass the filter accessible.
2. Provide visualizations that have a high acceptance and are commonly used in the field, like bar charts and REVIGO treemaps.
3. Summarize the overall similarity of the lists using, e.g., correlation values or the number of common enriched GO terms.
4. If available, also incorporate and visualize gene information.

Moreover, we refined the visualizations after testing the first version of the tool with domain experts in expert feedback sessions (see Section 6). In the following, we describe the implementation of our objectives in the current version of GO-Compass.

4.1. Dashboard Components

GO-Compass provides multiple visualizations organized in a dashboard for the comparison of the lists of GO terms based on the dispensability clustering. The dashboard consists of five components (Figure 1, original screenshot in Suppl. Figure S2). Two components provide an overview of the dataset by showing the clustering results (Figure 1A) or by summarizing the comparison of different GO term lists into single values (Figure 1C). The other components (Figure 1B and D) provide more detailed views of the individual GO term lists. All visualizations are interactive, provide linked highlighting, and are enhanced using tooltips displaying the full GO term names and further information. All GO terms together with their associated genes, *p*-values, dispensabilty, and uniqueness are collected in a table below the visualization components of the dashboard (see Suppl. Figure S3 for an example).

Using the controls menu of the dashboard, users can switch the GO domain displayed and define the significance threshold for the visualizations. Every visualization as well as the final filtered list

of GO terms together with their *p*-values and dispensability can be exported.

4.1.1. Hierarchical Clustering Visualization and Cutoff Selection

To improve the explainability of the dispensability clustering results, created with the modified REVIGO algorithm (Objective 1), the clustering tree is visualized using an indented tree layout (Figure 1A). The nodes in the tree represent GO terms, positioned by their dispensability indicated on the *x*-axis. For interactive cutoff selection, two sliders in the tree can be used for selecting the data for the other visualizations. With the right slider, redundant terms can be filtered out. The left slider cuts the tree at a specific dispensability to produce flat clusters, which are colored using a set of categorical colors. These colors are used as the background color of the flat clusters in the tree, as well as in the other components of the dashboard. Next to each slider a number is shown indicating the number of flat clusters and the total number of GO terms currently visualized. Numbers next to the GO terms indicate the number of direct descendants of this term in the hierarchical clustering that are currently filtered out. By allowing an interactive selection of the cutoffs, users have full control over the number of clusters and GO terms that are visualized.

To facilitate the creation of flat clusters in the tree, child terms are always placed below parent terms in decreasing sort order by dispensability. Thereby, when a new flat cluster is created, all GO terms belonging to this cluster are placed in a block. Moreover, this layout prevents crossing edges and terms from changing positions when the filter slider is moved.

The *p*-values of the GO terms across different lists are visualized in a heatmap, which is aligned horizontally to the tree using guiding lines as a visual aid. The heatmap shows the negative logarithm of the *p*-values in the different lists using a color scale from white (high *p*-value) to red (low *p*-value). The *p*-values that pass the significance threshold defined by the user are indicated using a black dot in the center of the heatmap cell.

To deal with trees larger than the designated component, the tree and the aligned heatmap are vertically scrollable. A custom scroll bar that serves as a summary visualization on the left side shows the current position and size of the visible section of the full tree relative to the currently created flat clusters (Suppl. Figure S4).

4.1.2. Treemaps

For list-centered overviews, the significance of GO terms at each condition is visualized using treemaps, similar to *REVIGO* treemaps (Objective 2, Figure 1B). The size of a rectangle corresponds to the negative logarithm of the *p*-value. Rectangles representing GO terms with significant *p*-values receive a full color fill, and non-significant GO terms are indicated with a striped fill. One list is selected to be shown as the main treemap (Figure 1B.i), while the others are visualized in small multiples in a scrollable container (Figure 1B.ii). By clicking on a small multiple, the corresponding list is selected and triggers an animation where the rectangles dynamically change their size. This helps find especially prominent changes between lists. For smooth transitions, the
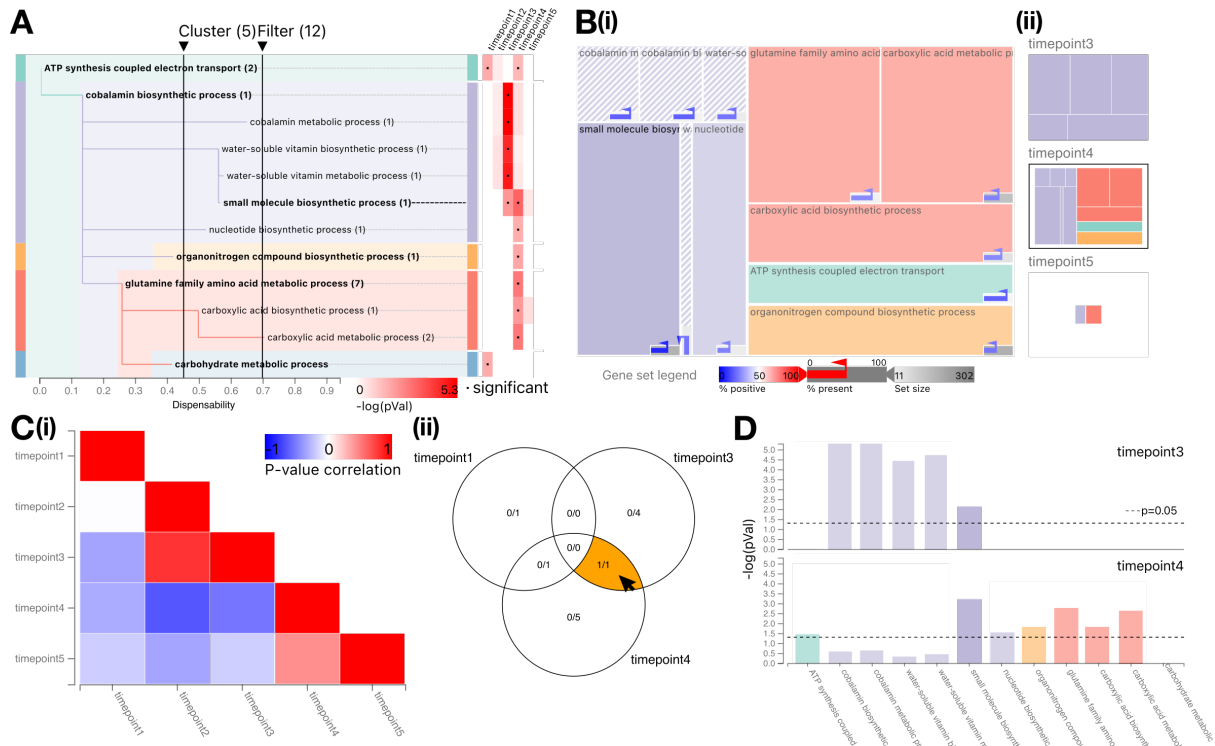
Figure 1: Overview of the dashboard components of GO-Compass using a dataset from a time-series gene expression experiment [SKvD*20]. (A) The REVIGO clustering results are visualized in an indented tree layout. Two sliders enable the interactive filtering of GO terms and the selection of a cutoff for the creation of flat clusters at chosen dispensability. The flat clusters are visualized using a set of categorical colors. The numbers next to the left and the right slider indicate the number of currently created flat clusters and the number of GO terms passing the filter cutoff. The number next to the GO terms indicates the number of direct descendants in the clustering that are currently filtered out. To visualize the *p*-values of each GO term, a heatmap is aligned to the tree on the right side. Guiding lines connect the GO terms in the tree with the heatmap. On the left side of the tree, an overview component is located which shows the current position in the tree is larger than the currently visible part. (B.i) A treemap shows the significance of GO terms at a selected list and encodes the flat clusters depending on the selected cutoff value. (B.ii) Using small multiples, different conditions can be selected. (C) A Venn diagram visualizes the intersection of significant GO terms across the different lists. The number of each subset represents the current selection (left) and the total size of each set (right). The subsets can be hovered (indicated by an orange highlight) for a more detailed comparison using bar charts showing the −log(*p*-values) in the hovered lists. (D) The dotted lines represent the chosen *p*-value threshold.
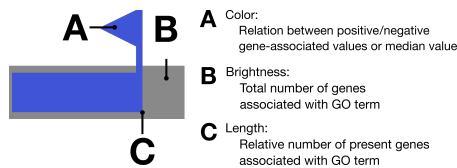

Figure 2: Glyph for the visualization of gene information. (A) The color of the bar either encodes the median value in a set of genes or the proportion of positive/negative values on a color scale from blue to white to red, where blue and red shades correspond to the majority of values being negative and positive, respectively. A white shade encodes an equal number of positive and negative values. (B) The brightness of the background in the glyph corresponds to the set size of the GO term, where smaller sets receive darker fills than bigger sets. (C) The length of the bar encodes for the relative number of present genes associated with the GO term.

treemap is implemented using a modified version of the squarify algorithm [BHvW00], where nodes only change size but not positions when the animation is triggered.

A glyph is used to additionally show gene information if available (Figure 2, Objective 4). The glyph consists of a bar showing the proportion of genes associated with a GO term that are present in the corresponding genes. The color shade of the bar is used to encode additional numerical information if the genes are associated with numbers (e.g., fold changes in differential expression). Users can choose to either show the median value or the proportions of negative to positive values on a color scale from blue to red. A small flag-like attachment on the right end of the bar ensures that the color is still recognizable even if the bar is short. If no information about fold changes is available, the bar is colored black. Furthermore, the background of the glyph encodes the gene set size of a GO term using a grayscale color scale, where dark values encode for larger sets than bright values. If no gene information has

been uploaded, only the gene set size is shown, which is always available since it is calculated using the uploaded background lists.

When hovering over a rectangle in the treemap the *p*-values of the corresponding GO term in the different lists are displayed in a tooltip with bar charts or line charts. Moreover, the full GO-term name, as well as the gene information is shown in the tooltip.

4.1.3. Summary Visualizations and Bar Charts

In order to allow a quick comparison of the lists (Objective 3), two summary visualizations show the similarity of GO terms between the lists (Figure 1C). The user can choose between visualizing the correlation of the *p*-values of the GO terms (Figure 1C.i) or the intersecting sets of significant GO terms across the lists (Figure 1C.ii). A correlation heatmap visualizes Pearson's correlation of the *p*-values of GO terms between the lists regardless of the significance threshold (as seen in the use case in Figure 3B) using a blue (negative correlation) to red (positive correlation) color scale. The set summary visualization shows the overlap of significantly enriched GO terms between the conditions. For less than four lists, a Venn diagram is used (Figure 1C.ii). For more than three lists an UpSet plot is visualized instead (as shown in the use case in Figure 3D) [LGS*14]. UpSet plots show set intersections as a matrix, where rows correspond to the sets and columns to the intersections.

For detailed comparison of subsets of lists, cells in the correlation heatmap or intersections of the Venn diagram or UpSet plot can be hovered and the GO terms can be compared in vertically juxtaposed bar charts (Figure 1D), where each bar chart corresponds to one of the hovered lists. All GO terms present in the hierarchy after filtering are shown on the horizontal axis.

4.2. Implementation

GO-Compass is implemented as a client-server application using Python for the enrichment calculation and dispensability clustering with goatools in the backend and JavaScript for the visualization in the front end. The Python package Flask [fla10] is used for client-server communication. The frontend is implemented using a combination of React [Fac13] for the component structure of the application, mobx [mob15] for storing the application's state, and D3 [BOH11] for the visualizations, i.e., scales, animations, and the treemap layout.

5. Use Cases

To demonstrate the general applicability of GO-Compass in the omics field, we present two use cases. The first use case shows how GO-Compass allows the exploration of many GO term lists from a transcriptomics experiment and also includes further important gene information. In the second use case, we present the results of an enrichment analysis of genomics data and demonstrate how GO-Compass facilitates the analysis of a few hundred of GO terms.

5.1. Use Case 1: Functional Enrichment of Antibiotic Response in the Mouse Transcriptome

In the study by Lavelle et al. [LHP*19], the authors addressed the question how antibiotics change the transcriptome response across

a time window of 18 days in two mice. For this, they sampled RNA at four different time points (Day 0, 8, 11, and 18), analyzed the expression levels using microarrays, and computed differentially expressed genes between each time point. Of the six pairwise comparisons, one did not return significant differential expression (day 18 compared to day 11, in short D18vsD11). To acquire a functional overview of the differences, the authors ran an enrichment analysis using goana [RPW*15] on the five lists of genes. Independently for each comparison, they extracted the 100 most significant GO terms and reduced the datasets to those terms with a dispensability lower than 0.4 by using REVIGO for each of the lists. For visualization of the results they created a tailored visualization using custom R scripts. Figure 3C in the original article visualizes the results using bar charts encoding for the negative logarithm of the *p*-value [LHP*19].

We reproduced the results with GO-Compass to show how it is able to provide the same information and furthermore include exploratory as well as other informative features. The preprocessing was reproduced from the data and scripts provided by Lavelle et al., resulting in five lists of differentially expressed genes for each comparison (referred to as D8vsD0, etc. in the following) and the significant GO terms identified by goana. To make the results comparable, we also reduced the dataset of each comparison to the 100 most significant terms as done by the authors. The five lists of genes with their $\log_2$-fold changes, the results of the GO enrichment, and the used GO background were loaded into GO-Compass for the dispensability computation and visualization. Wang's similarity measurement and a *p*-value filtering of 5×10^{-5} were selected as further parameters.

We first focus on those terms related to the domain *Biological process*. For this domain, GO-Compass returns a visualization with 179 GO terms, which can be reduced to 25 terms by filtering terms with a dispensability higher than 0.4 (Figure 3A), as done in the original analysis. To gain more insight into the different GO terms, the cluster cutoff of GO-Compass was set to 0.2, returning seven flat clusters (Figure 3). The visualization of many aspects of the lists within the graphical interface allows simultaneous exploration. For example, the heatmap aligned next to the tree shows that all but the first list of GO terms contain terms of the cluster rooted at *extracellular matrix organization*. This dissimilarity can be confirmed by looking at the correlation heatmap (Figure 3B) that shows how D8vsD0 is negatively correlated with all but one comparison at this dispensability level (D18vsD8, Pearson correlation $\rho \approx 0.08$).

By using the treemaps, users are able to compare the flat clusters between these two lists and account for further data, such as the $\log_2$-fold change of the genes. The animated transition facilitates the finding of the common terms between the lists and tracking major changes (Figure 3C). The visualization of the median shows that the expression of genes corresponding to the lists is not similar. While the genes related to the terms *rhythmic process* and *regulation of circadian rhythm* in D8vsD0 are upregulated, the genes in D18vsD8 are strongly downregulated (see glyphs in Figure 3C).

The UpSet plot can be used to compare the terms that are significant in more than one list (Figure 3D). For example, the GO terms *extracellular matrix organization*, *blood vessel development* and
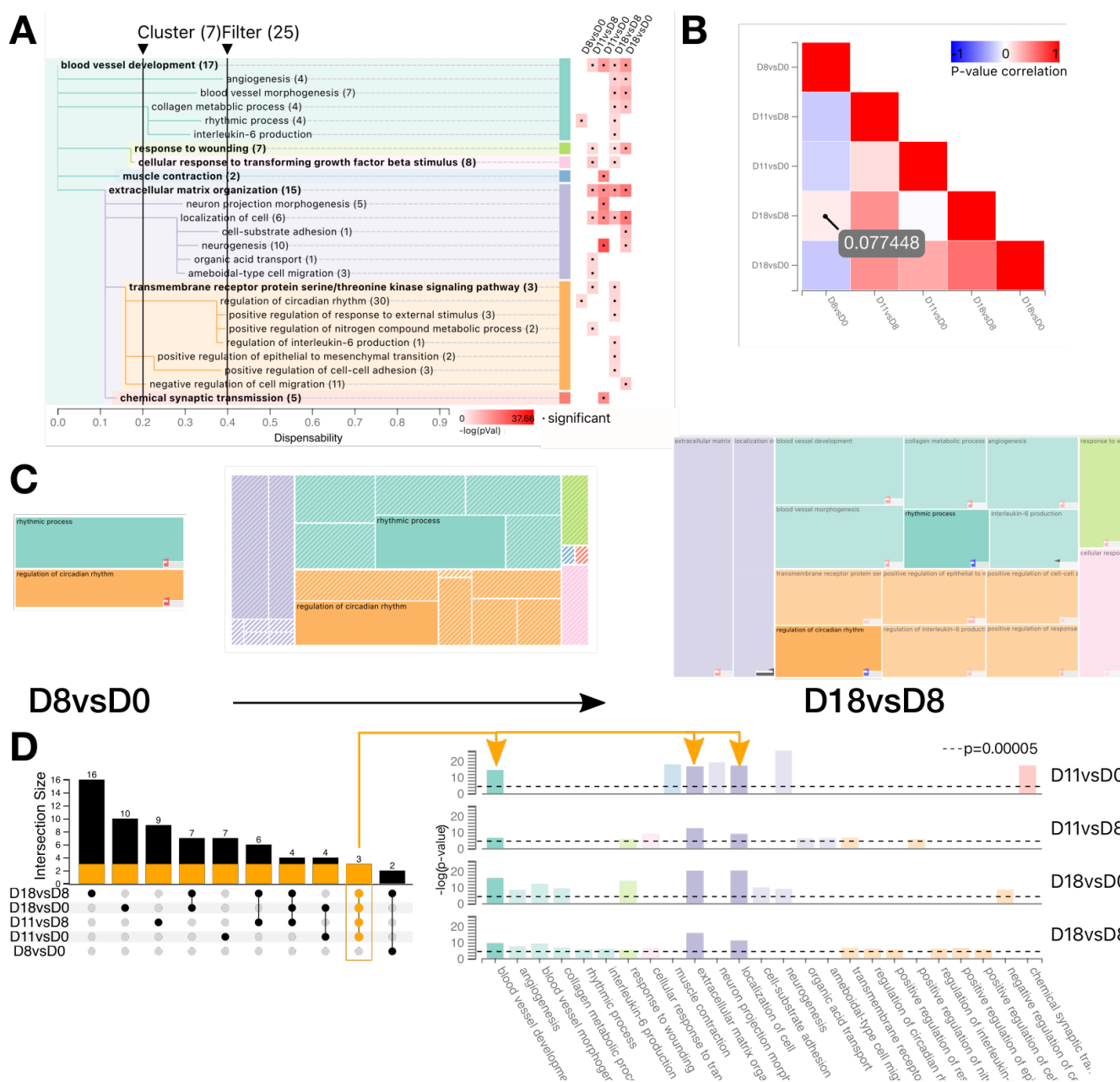
Figure 3: Use case for `GO-Compass` with the transcriptome data from the study by Lavelle et al. (2019). All subfigures visualize terms from the ontology domain *Biological process*. (A) Visualization of the clustering results. The clusters were cut at a dispensability level of 0.2 using the left slider. Terms with a dispensability lower than 0.4 have been filtered out using the right slider. (B) Correlation heatmap of all *p*-values of the terms contained in the five lists. The list comparing day 8 with day 0 (D8vsD0) shows an anticorrelation of GO terms with all other lists, except for a minimal positive correlation with the list D18vsD8. (C) Treemap visualizations for the conditions D8vsD0 and D18vsD8. The transition animation (middle treemap) helps to follow the changes between the two treemaps. The glyph shows the median $\log_2$-fold change of the genes linked to each GO term. (D) UpSet plot comparing the specific terms shared across conditions. Three GO terms are found significant in all four lists. The juxtaposed bar charts containing all terms of these four lists are shown, and the bars for the common terms are highlighted.
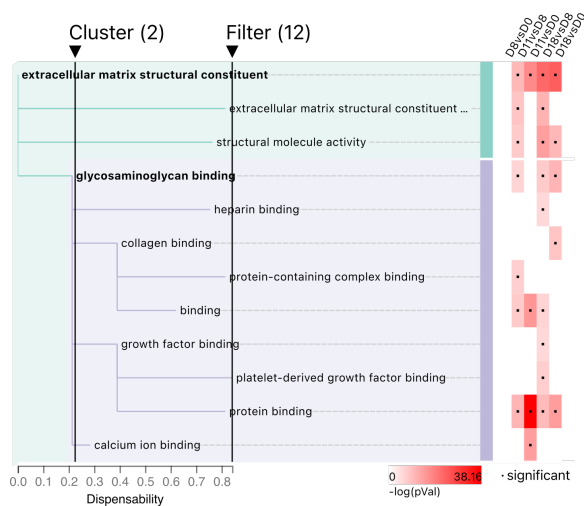
Figure 4: Tree showing the clustering results for all terms related to the ontology domain *Molecular function* for the study by Lavelle et al. (2019). The cluster cutoff was defined at a dispensability level close to 0.2 to visualize a cluster rooted at *glycosaminoglycan binding* (purple cluster).

localization of cell are found significant in four lists and their corresponding *p*-values can be easily compared using the bar charts. Interestingly, in the dispensability reduction of the original analysis, the term *extracellular matrix organization* appears only in three of the five lists. However, this term is actually significant in four of the lists, which is evident in the analysis with GO-Compass.

Lavelle et al. also analyzed significant GO terms for the ontology domain *Molecular function*, for which they identified ten different GO terms, all related to *binding*. With the visualization of GO-Compass, the user is able to visualize the hierarchy of the clusters and identify the most important terms regarding their dispensability (Figure 4). In this case, by setting a cluster cutoff at 0.2 a meaningful flat cluster is created containing all terms related to *binding* rooted at *glycosaminoglycan binding*. Moreover, by visualizing the whole clustering and interactively selecting the filter cutoff, the user can decide whether general terms such as *binding* should be filtered out to reduce the redundancy of the results for further analysis. This cutoff is independent of the filter cutoffs selected for the other GO domains, while in the original analysis with REVIGO a single cutoff value for all domains had to be chosen.

5.2. Use Case 2: Genomic Variability in the Syphilis Agent, *Treponema pallidum*

We demonstrate the wide applicability of GO-Compass for studying the genetic variability of *Treponema pallidum*, the bacterium responsible for syphilis [SR99]. For this, the data published by Pla-Díaz et al. [PDSBG*22] was analyzed using the tool Evidente [WPHN22] to identify enriched GO terms for genes associated with SNPs in phylogenetic clades. A clade is a set of organisms that share a common ancestor in a phylogenetic tree. The data of Pla-Díaz et al. consists of genomic samples from four main strains of *T. pallidum*: ten samples from the strain *Nichols*
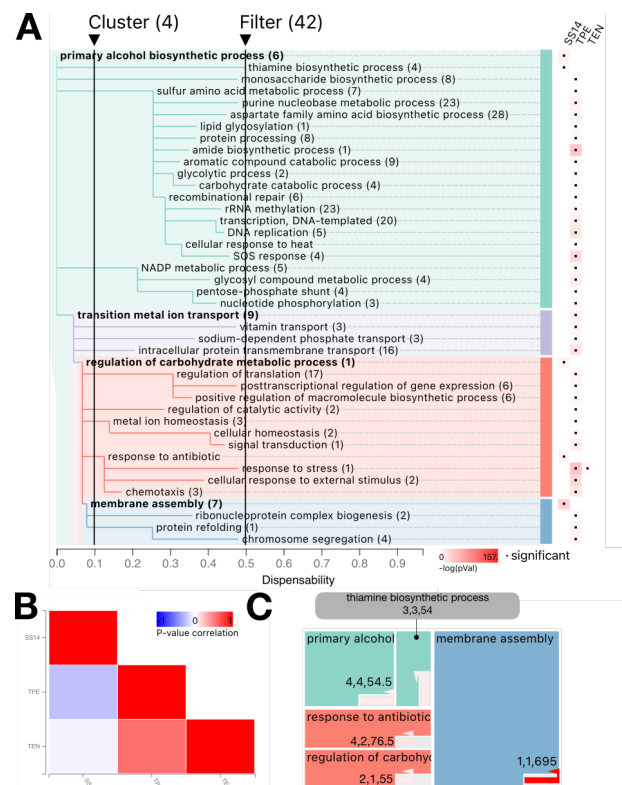

Figure 5: Use case for GO-Compass resulting from the GO enrichment of phylogenetic clades of *Treponema pallidum*. All subfigures visualize terms from the ontology domain *Biological process* with a dispensability lower than 0.5. (A) Visualization of the clustering results. The cluster cutoff was set at a dispensability level of 0.1. (B) Correlation heatmap of all *p*-values of the terms contained in the three clades. (C) Treemap visualization for the SS14 clade. The numbers above or beside the gene information glyph encode the total number of genes associated with the GO term, the genes associated with the GO term containing SNPs, and the median number of SNPs found within these genes. The same values are encoded in the gene information glyph using the background color, the length of the bar and the color of the bar, respectively.

and 58 samples from the strain *SS14* (both belong to subspecies *T. pallidum pallidum, TPA*), seven samples from subspecies *pertenue* (TPE) and one sample from subspecies *endemicum* (TEN). All samples were compared to one *Nichols* reference genome to identify single-nucleotide polymorphisms (SNPs). Using Evidente we identified GO terms that are enriched for genes affected by mutations within one clade compared to the rest of the phylogenetic tree. A GO enrichment was calculated for four main phylogenetic clades, producing a list of significantly enriched GO terms per clade (*p*-value < 0.05). All non-reference clades showed enriched GO terms: 61 within the SS14 clade, 70 within TPE and 514 within the TEN clade. Since Evidente does not contain features for direct comparison of GO terms, these three lists, containing a total of 576 unique significant GO terms, were loaded into GO-Compass. Furthermore, for each clade, a list containing the number of SNPs

per gene across all strains and positions was uploaded as further gene information. Wang's semantic similarity measure and *p*-value < 0.05 were used as parameters for the redundancy calculation.

All three GO domains (biological process, molecular function, and cellular component) show high correlation and GO term similarity across the TPE and TEN lists, while the SS14 clade shows distinct GO terms over all three domains. We further reduced the data by filtering out GO terms with a dispensability larger than 0.5. In the case of the domain *Biological process*, the dataset of 306 GO terms was reduced to 41 terms which still reflects the similarities between TPE and TEN as well as their differences with SS14 (Figure 5A/B). Interestingly, the GO terms specific for the SS14 clade include low dispensability terms such as *membrane assembly* and *response to antibiotic* (Figure 5A). The low dispensability and the position of the terms in the tree indicate that they are specific to the SS14 clade and are only very distantly related to terms enriched in TPE and TEN. This is in agreement with the information presented by Pla-Díaz et al., where they discuss the changes in the membrane of SS14 strains and the antibiotic-resistant potential across SS14 samples when compared to strains from the Nichols clade. With the glyph visualization in the treemap, we were able to identify that a single gene is linked to the term *membrane assembly*. The red color of the bar indicates that there is a high number of SNPs (695) associated with this gene across all samples and positions of the gene (Figure 5C). This specific gene (*TPANIC_RS01590*) can be identified using the tabular view for further analysis, for example, to study the effect of the SNPs on the translated protein.

6. Qualitative Evaluation

For qualitative evaluation of our tool, we performed semi-structured interviews with five domain experts including two bioinformaticians, two biologists, and a clinical doctor who is one of the authors of the original publication discussed in the first use case (see Section 5.1). None of the experts was involved in the development of the tool or had used our tool before. The interviews were conducted in an online video session and took around 1h each. First, the experts were interviewed about their experiences using GO terms. Depending on their level of expertise the concept of Gene Ontology was recapitulated and the central ideas of the tool were introduced. Apart from the dispensability clustering tree, none of the visualizations were explained to get an unbiased view of the tool's intuitiveness. After the session, the experts were asked to fill out a short survey that includes a selected subset of the system usability scale questionnaire (see Suppl. Figure S5 for the questions) [Bro96].

For the testing, the data from the first use case was used. During the session, the experts were presented with five tasks, each related to a different visualization component (Suppl. Table S1). The first task (T1) was to reduce the size of the dispensability tree and simplify the visualization. In the second task (T2) users were asked which lists are the most/least similar ones. For the third task (T3) they were asked which three GO terms are the most significant ones shared between the two most similar lists. The fourth task (T4) was about gaining an overview of a single list and the experts were asked to summarize the biological processes happening in the list D11vsD0. The last task (T5) was made up of subtasks for compar-

ing the gene sets between two GO terms in the treemap based on the underlying gene information.

6.1. Expert Feedback

All participants were able to successfully reduce the size of the tree (T1) by using the filter slider. Two participants were initially confused by the cluster slider but were able to figure out its function when being told to observe the effect of moving the slider in the other visualizations. Moreover, all participants were able to compare lists using the correlation heatmap (T2). While they were not able to tell which values are correlated (*p*-values), they had the intuition that similar lists were indicated by a red color (see Figure 1C,i). All participants were able to compare the lists in detail using the bar chart (T3) and found the linkage between the correlation heatmap and the bar charts useful.

For the fourth task, all users were able to navigate to the list in question in the treemap visualization and could summarize the biological processes. Similar to the second task, users were not able to explicitly tell what is encoded, but still were able to solve the task. The glyph visualization (T5) received mixed feedback from the users. It is important to note, that the version tested did not include a legend for the gene set glyph. Two users wrongly assumed that the blue-to-red color scale of the glyph is related to the correlation displayed in the correlation heatmap. Moreover, initially, two users could not tell that the background of the glyph was colored using different shades of gray. In the tested version big gene sets received brighter fills than small gene sets, which was considered unintuitive by all users. We, therefore, decided to reverse the color encoding of the gene set sizes for the final published version of GO-Compass. Eventually, all users were able to solve the subtasks and considered the glyph useful after the explanation.

Overall, the tool received positive feedback from all experts. One expert noted that although she could not tell why it was very intuitive for her to know where to click and hover. While the visualization dashboard was first considered slightly overwhelming by some users, all found it useful and consider using it for their own research or for teaching. One expert noted, that the tool obviously shows a lot of information. However, they did not consider it to be a flaw as the underlying data contributes a lot to the complexity and the tool already reduces the mental load for analyzing the data. One expert noted that he "can't think of anything that beats this" when being asked for an overall evaluation of the tool.

From the feedback session, we deducted several action items and features for implementing the tool that we then included in the current version. Many of these features relate to improving the labeling and legends of the visualizations as well as solving minor usability issues. Legends were included for the glyph and a label now indicates that the correlation heatmap shows the correlation of p-values. Moreover, the scrolling behavior of the dispensability tree was improved by making the overview visualization draggable. Furthermore, a user suggested a more detailed visualization of gene information as well as a simplified version of the dispensability tree which we consider enhancements for future versions of the tool.

The answers collected from the survey after the interview supported the positive impression we had during the interviews (Suppl.

Figure S5). While the tool was considered complex, all users found it useful and would like to use it frequently for their own research.

7. Discussion

We presented `GO-Compass`, an application with a unique set of features enabling the interactive visual comparison of lists of gene ontology terms. The tool extends the `REVIGO` algorithm and the treemap visualization for the analysis of multiple lists of GO terms and has a wide field of applicability for lists created in different omics studies due to the simple input formats it requires. With its tree visualization, it increases the explainability of the underlying hierarchical clustering and gives users full control over the desired level of dispensability in the lists of GO terms. Furthermore, it effectively combines well-known visualization techniques, for the direct comparison of sets such as UpSet plots and correlation heatmaps with detailed-focused visualizations like bar charts and treemaps for a comprehensive analysis. Moreover, with the glyph in the treemaps gene information is included, if available.

`GO-Compass` is based on the structure of the gene ontology, semantic similarity calculations, and on GO enrichment. Since the GO graph is created using expert knowledge, which differs for the different sub-domains of the graph, it lacks uniform density and uniform relatedness of parent-child relationships. The edges of the GO graph are not weighted according to relatedness, therefore it is not possible to take this factor into account for semantic similarity calculation. `GO-Compass` counteracts these factors by allowing interactive selection of the filter cutoff and therefore gives users the opportunity to decide on a desired level of redundancy with computational support. Furthermore, different dispensability cutoffs can be chosen for each gene ontology domain. Thus, with our interactive dispensability filtering we compensate for the bias that all GO terms are treated the same regardless of their relatedness in the GO graph. When reducing the dispensability in lists separately, for example using `REVIGO`, it can happen that a term is filtered out in some of the lists despite being significant in all of them. Our simultaneous analysis of all lists avoids losing information, as seen with the term *extracellular matrix organization* in the first use case, which was filtered out in one of the lists in the original analysis.

`GO-Compass` clusters the complete lists of GO terms instead of stopping at a fixed dispensability cutoff. This gives the users insights into the underlying algorithm and increases their control over the desired level of redundancy. This can be seen especially prominently in the first use case, where dispensable terms could easily be identified and filtered out using the visualization. The indented layout of the tree is designed to facilitate the process of filtering and the creation of flat clusters, as it ensures that all children of a GO term in the hierarchical clustering are organized in a block without the need of rearranging the tree after moving the sliders.

`GO-Compass` implements treemaps for visualizing the functional composition of a single list. An animation is triggered to show the changes when the condition is switched. While it is difficult to compare the areas of rectangles of different sizes and follow multiple concurrent animations, the most drastic changes are underlined using the animation. A more detailed comparison can be done using the bar chart appearing in a tooltip when hovering over a

GO term. Moreover, with the designed glyph any type of numerical information associated with genes can be included in the analysis. Moreover, by including linked summarizing charts (the correlation heatmap and the UpSet plot) and detailed views (treemap and juxtaposed bar charts), users can access different levels of information of the datasets. This helps to compare the lists using many relevant aspects for the domain, such as further gene information or shared GO terms.

In our use cases, we demonstrated the applicability of `GO-Compass` for two different omics data types. In our first use case, we presented a use case for functional analysis in transcriptomics studies, which is one of the most common use cases where GO enrichment is applied. In the second use case, we show how the tool can be applied to genomics data and handle data up to a few hundred GO terms. However, `GO-Compass` can be applied in many other contexts, such as the integration of lists originating from different omics layers in multi-omics experiments, as well as lists originating from different species.

`GO-Compass` compares lists of genes regarding their functional composition using GO terms. However, also other annotation systems, such as KEGG Pathways or protein families, are frequently used by researchers for functional analysis. Compared to these annotations, the advantage of GO terms is the inherent structure of the GO graph, which allows the simple calculation of semantic similarity to serve as the basis for dispensability clustering. Nevertheless, we plan to extend our tool to be able to deal with other types of enrichments using custom algorithms for dispensability calculation.

To conclude, as omics experiments increasingly consist of the comparison of many conditions, semantic-similarity-based reduction for generating visualizations of multiple GO term lists has become more and more popular and useful. With `GO-Compass` we provide an easy-to-use visualization tool for this task with the ultimate aim to help researchers interpret the biology of the systems studied.

References

[ASJ*16] ARORA N., SCHUENEMANN V. J., JÄGER G., PELTZER A., SEITZ A., HERBIG A., STROUHAL M., GRILLOVÁ L., SÁNCHEZ-BUSÓ L., KÜHNERT D., BOS K. I., DAVIS L. R., MIKALOVÁ L., BRUISTEN S., KOMERICKI P., FRENCH P., GRANT P. R., PANDO M. A., VAULET L. G., FERMEPIN M. R., MARTINEZ A., CENTURION LARA A., GIACANI L., NORRIS S. J., ŠMAJS D., BOSSHARD P. P., GONZÁLEZ-CANDELAS F., NIESELT K., KRAUSE J., BAGHERI H. C.: Origin of modern syphilis and emergence of a pandemic treponema pallidum cluster. *Nat Microbiol 2* (Dec. 2016), 16245. 1

[BH95] BENJAMINI Y., HOCHBERG Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological) 57*, 1 (1995), 289–300. 3

[BHvW00] BRULS M., HUIZING K., VAN WIJK J. J.: Squarified treemaps. In *Data Visualization 2000* (2000), Springer Vienna, pp. 33–42. 5

[BJHA19] BRIONNE A., JUANCHICH A., HENNEQUET-ANTIER C.: ViSEAGO: a bioconductor package for clustering biological functions using gene ontology and semantic similarity. *BioData Min. 12* (Aug. 2019), 16. 2

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2301–2309. 6

[Bro96] BROOKE J.: SUS -A quick and dirty usability scale. http://www.tbistafftraining.info/smartphones/documents/b5_during_the_trial_usability_scale_v1_09aug11.pdf, 1996. Accessed: 2023-3-15. 9

[Fac13] FACEBOOK: React - A JavaScript library for building user interfaces. https://reactjs.org/, 2013. Accessed on: 02-12-2022. 6

[FAG15] FORTINO V., ALENIUS H., GRECO D.: BACA: bubble chart to compare annotations. *BMC Bioinformatics 16* (Feb. 2015), 37. 2

[fla10] Flask — A micro web framework written in Python. https://flask.palletsprojects.com/, 2010. Accessed on: 02-12-2022. 6

[HCI*04] HARRIS M. A., CLARK J., IRELAND A., LOMAX J., ASHBURNER M., FOULGER R., EILBECK K., LEWIS S., MARSHALL B., MUNGALL C., RICHTER J., RUBIN G. M., BLAKE J. A., BULT C., DOLAN M., DRABKIN H., EPPIG J. T., HILL D. P., NI L., RINGWALD M., BALAKRISHNAN R., CHERRY J. M., CHRISTIE K. R., COSTANZO M. C., DWIGHT S. S., ENGEL S., FISK D. G., HIRSCHMAN J. E., HONG E. L., NASH R. S., SETHURAMAN A., THEESFELD C. L., BOTSTEIN D., DOLINSKI K., FEIERBACH B., BERARDINI T., MUNDODI S., RHEE S. Y., APWEILER R., BARRELL D., CAMON E., DIMMER E., LEE V., CHISHOLM R., GAUDET P., KIBBE W., KISHORE R., SCHWARZ E. M., STERNBERG P., GWINN M., HANNICK L., WORTMAN J., BERRIMAN M., WOOD V., DE LA CRUZ N., TONELLATO P., JAISWAL P., SEIGFRIED T., WHITE R., GENE ONTOLOGY CONSORTIUM: The gene ontology (GO) database and informatics resource. *Nucleic Acids Res. 32*, Database issue (Jan. 2004), D258–61. 1

[KLS*19] KUZNETSOVA I., LUGMAYR A., SIIRA S. J., RACKHAM O., FILIPOVSKA A.: CirGO: an alternative circular way of visualising gene ontology terms. *BMC Bioinformatics 20*, 1 (Feb. 2019), 84. 2

[KV15] KOLDE R., VILO J.: GOsummaries: an R package for visual functional annotation of experimental data. *F1000Res. 4* (Aug. 2015), 574. 2

[KZP*18] KLOPFENSTEIN D. V., ZHANG L., PEDERSEN B. S., RAMÍREZ F., WARWICK VESZTROCY A., NALDI A., MUNGALL C. J., YUNES J. M., BOTVINNIK O., WEIGEL M., DAMPIER W., DESSIMOZ C., FLICK P., TANG H.: GOATOOLS: A python library for gene ontology analyses. *Sci. Rep. 8*, 1 (July 2018), 10872. 2, 3

[LGS*14] LEX A., GEHLENBORG N., STROBELT H., VUILLEMOT R., PFISTER H.: UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (Dec. 2014), 1983–1992. 2, 6

[LHP*19] LAVELLE A., HOFFMANN T. W., PHAM H.-P., LANGELLA P., GUÉDON E., SOKOL H.: Baseline microbiota composition modulates antibiotic-mediated effects on the gut microbiota and host. *Microbiome 7*, 1 (2019), 1–13. 6

[LO98] LIN D., OTHERS: An information-theoretic definition of similarity. In *Icml* (1998), vol. 98, pdfs.semanticscholar.org, pp. 296–304. 3

[LSBG03] LORD P. W., STEVENS R. D., BRASS A., GOBLE C. A.: Investigating semantic similarity measures across the gene ontology: the relationship between sequence and annotation. *Bioinformatics 19*, 10 (July 2003), 1275–1283. 2

[mob15] Mobx — simple, scalable state management. https://mobx.js.org/, 2015. Accessed on: 02-12-2022. 6

[PDSBG*22] PLA-DÍAZ M., SÁNCHEZ-BUSÓ L., GIACANI L., ŠMAJS D., BOSSHARD P. P., BAGHERI H. C., SCHUENEMANN V. J., NIESELT K., ARORA N., GONZÁLEZ-CANDELAS F.: Evolutionary processes in the emergence and recent spread of the syphilis agent, Treponema pallidum. *Molecular biology and evolution 39*, 1 (2022), msab318. 8

[Res99] RESNIK P.: Semantic similarity in a taxonomy: An Information-Based measure and its application to problems of ambiguity in natural language. *jair 11* (July 1999), 95–130. 3

[RPW*15] RITCHIE M. E., PHIPSON B., WU D., HU Y., LAW C. W., SHI W., SMYTH G. K.: limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research 43*, 7 (2015), e47–e47. 6

[SBŠŠ11] SUPEK F., BOŠNJAK M., ŠKUNCA N., ŠMUC T.: REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One 6*, 7 (July 2011), e21800. 2

[SKvD*20] SULHEIM S., KUMELJ T., VAN DISSEL D., SALEHZADEH-YAZDI A., DU C., VAN WEZEL G. P., NIESELT K., ALMAAS E., WENTZEL A., KERKHOVEN E. J.: Enzyme-Constrained models and omics analysis of streptomyces coelicolor reveal metabolic changes that enhance heterologous production. *iScience 23*, 9 (Sept. 2020), 101525. 5

[SR99] SINGH A. E., ROMANOWSKI B.: Syphilis: review with emphasis on clinical, epidemiologic, and some biologic features. *Clinical microbiology reviews 12*, 2 (1999), 187–209. 8

[TCK*03] THOMAS P. D., CAMPBELL M. J., KEJARIWAL A., MI H., KARLAK B., DAVERMAN R., DIEMER K., MURUGANUJAN A., NARECHANIA A.: PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res. 13*, 9 (Sept. 2003), 2129–2141. 2

[WDP*07] WANG J. Z., DU Z., PAYATTAKOOL R., YU P. S., CHEN C.-F.: A new method to measure the semantic similarity of GO terms. *Bioinformatics 23*, 10 (May 2007), 1274–1281. 3

[WPHN22] WITTE PAZ M., HARBIG T. A., NIESELT K.: Evidente—a visual analytics tool for data enrichment in SNP-based phylogenetic trees. *Bioinformatics Advances 2*, 1 (2022), vbac075. 8