

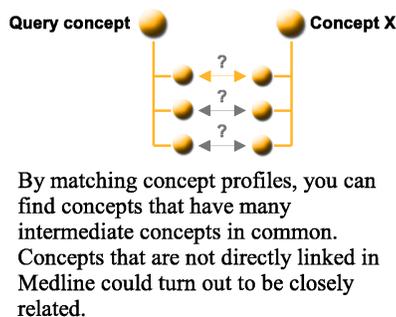
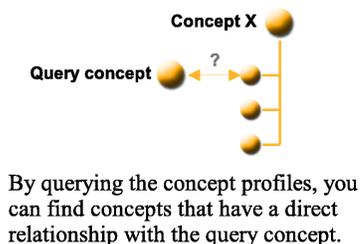
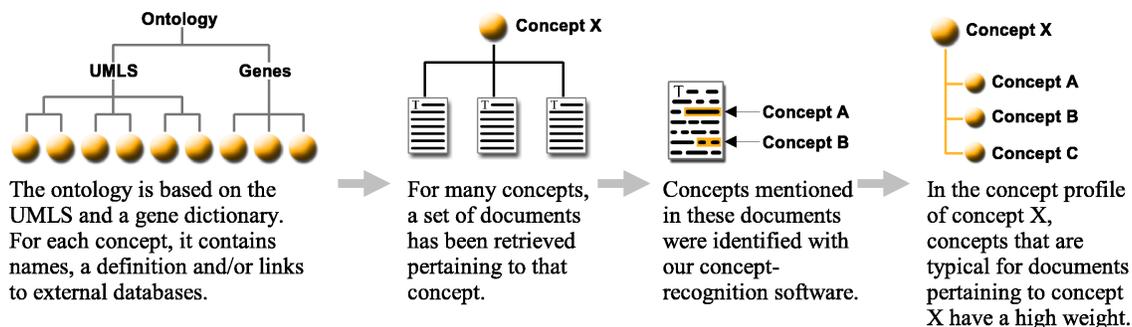
Anni 2.0 Manual

Anni provides an ontology-based interface to Medline and retrieves documents and associations for several classes of biomedical concepts, including genes, drugs and diseases, with proven text-mining technology. In Anni, texts associated to a concept are characterized by a so-called concept profile. A concept profile consists of a list of related concepts and each concept in the profile has a weight to signify its importance. Concept profiles can be used to retrieve associations between concepts in two ways (Figure 1): 1. Querying the concept weights in the concept profiles: e.g. a query with the concept "prostate cancer" on the set of all genes will retrieve the genes mentioned together with this concept in abstracts, sorted by strength of association as given by the weights in the concept profiles. 2. Concept profiles can be matched to identify similarities between concept profiles, for instance to identify genes associated with similar biological processes.

Anni does not restrict the user's actions, but some of the options can be very memory intensive and computationally expensive, especially when large numbers of concept profiles are queried, matched, projected or clustered. To exploit the full potential of Anni it is advisable to have a fast internet connection, and an up-to-date computer with more than a gigabyte of free memory.

If you encounter a bug in Anni or if you feel changing or adding a feature would improve Anni, please let us know by email: a.veldhoven@erasmusmc.nl

Figure 1. The technology behind Anni at a glance. Yellow balls indicate ontology concepts.

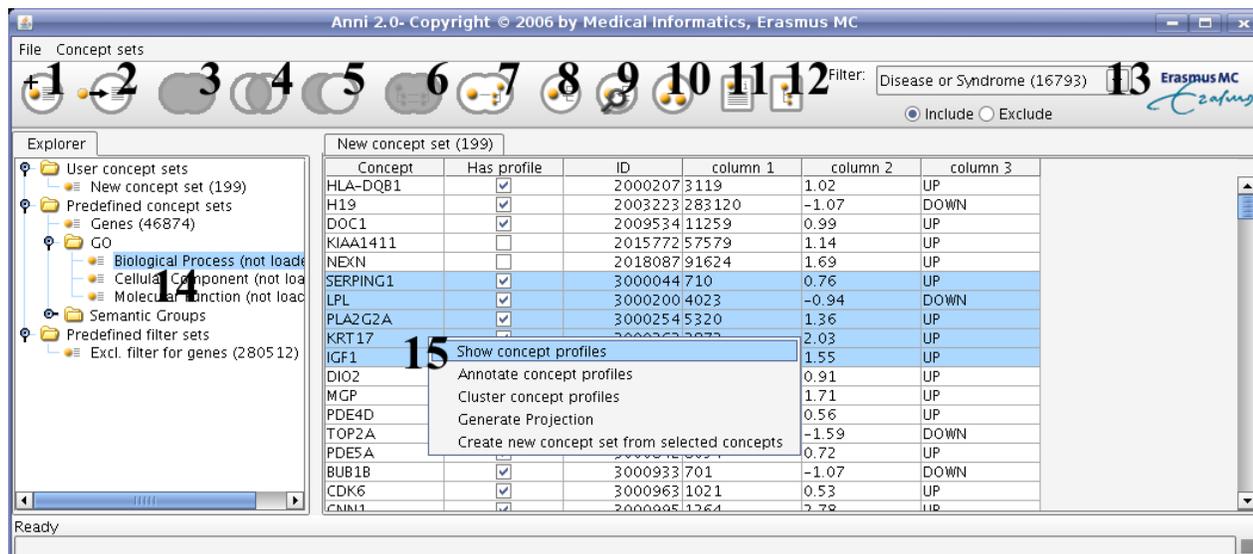


1. Concept sets

In Anni concepts are organized into concept sets. The available concept sets are organized in a tree structure (see Figure 2, #14). Upon start-up a range of predefined concept sets are loaded: the three branches of the Gene Ontology, the genes from our gene thesaurus (mouse, rat and human genes with homologs mapped to each other), and the concepts from the Unified Medical Language System (UMLS) organized into semantic types. The user can define his own concept sets in three ways: First, through manipulating existing concept sets, through the set operators (union, intersect and subtract; Figure 2, #3,4,5). Second, by selecting concepts in the concept set view (or any other view showing concepts) and transfer them to a new concept set (#15). Third, by adding new concepts to a concept set through an input screen (see following section).

Figure 2. Anni screenshot.

1. Make a new concept set.
2. Import concepts to the selected concept set.
3. Union, merge two concept sets into a new concept set.
4. Take the intersection of two concept sets as a new concept sets.
5. Subtract a concept set from a concept set to produce a new concept set.
6. Match two concept sets.
7. Query for concept associations in a concept set.
8. Cluster selected concepts.
9. Annotate the selected concepts.
10. Project the selected concepts.
11. Show the concept information for the selected concept.
12. Show the concept profile for the selected concept.
13. The filter drop down box: the set concept set functions as a filter for matching concept profiles, either exclusive or inclusive.
14. The concept set tree, including several predefined concept sets.
15. The main view pane showing a concept set. The concept set view shows whether a concept profile is available for a concept, in addition to the concept ID and user defined meta data. After selecting concepts right-clicking the table reveals the shown menu.

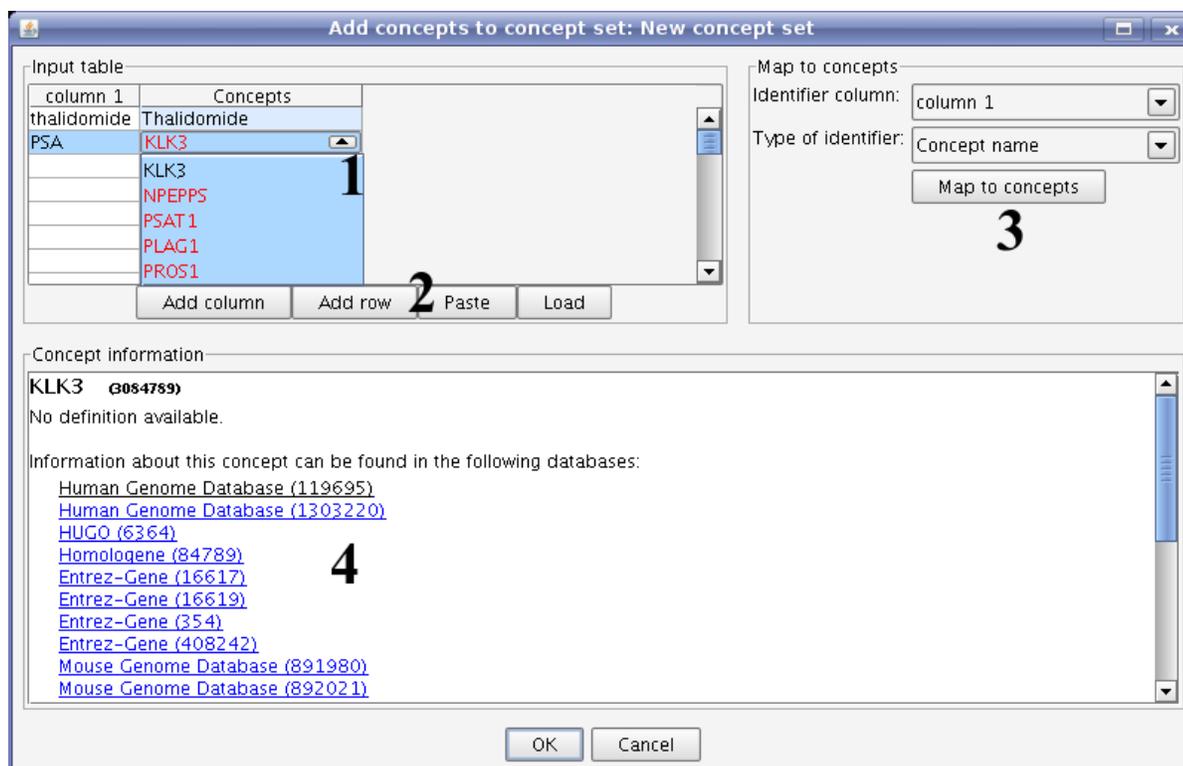


2. Finding concepts

Several types of concept identifiers can be mapped to the ontology concepts in the concept input panel (see Figure 3). Concept names can be used as concept identifiers and will be mapped to concepts by our concept recognition program Peregrine. Alternatively, database IDs can be supplied such as from Entrez Gene, mouse genome database, rat genome database or Swissprot (Entrez Gene IDs have the best coverage). Concept IDs can be typed, pasted or loaded from a tab-delimited text file. In this concept input panel the user can also supply additional information to be associated with the concepts (new columns can be added; Figure 3, #2). This information is visible in the application for the user's convenience and can be used for the color legend in the projection view. After entering the concept information, the ids have to be mapped to concepts (Figure 3, #3). The mapping should be checked and ambiguities in the concept ids can be resolved by the user.

Figure 3. The concept input panel.

1. The user's input is shown here. If a concept is ambiguous, such as PSA in this example, the suggested concept appears in red and the user is requested to disambiguate.
2. Input can be provided through typing, but also by pasting and loading a tab-delimited text-file. In addition, the user can provide additional columns to be associated to the concepts and shown in the program.
3. In this section the column containing the concept identifiers and the type of identifier can be selected.
4. Concept information of concepts selected in panel 1 is shown here. In the KLK3 example shown here, hyperlinks to online databases are available.



3. Querying a concept set

The concept profiles in a concept set can be queried for the association strengths with one or more concepts that are defined in another concept set (the query concept set). To query a concept set, first select the concept set that you want to query and the concept set you want to use for querying and press the query button (#7 in Figure 1). Alternatively the query can be defined on the fly in the query panel (Figure 4). The default result view is shown in Figure 5. In the resulting view concepts are ranked by their query score, which is the sum of query concepts weights in the concept profiles. Queries can also be weighted to modulate the contribution of individual concepts to the eventual query score. Apart from the query score column, a column is provided that indicates if all concepts of the query are found in the concept profile. The query result view can also indicate the actual co-occurrence counts between the query and queried concepts and provides a link out to supporting documents for the given scores.

Tip: When viewing a table, clicking the headers will sort the table contents on that column. Clicking again will reverse the sorting order.

Figure 4. The query input panel.

1. The query set and the set to be queried can be switched here.
2. The query concept set is shown here. The user can define the query on the fly, or use an existing concept set.
3. The query concepts can be given weights to fine tune the query.
4. If the user defines a query on the fly, concepts can be mapped through this section of the panel.

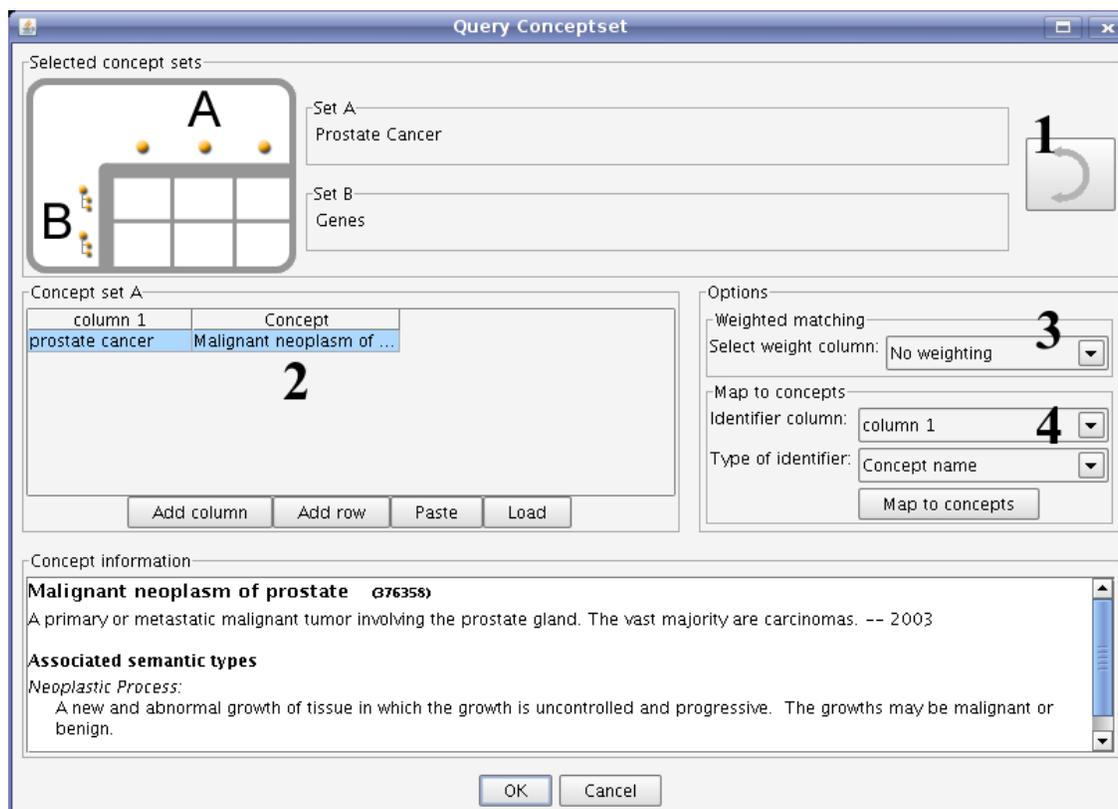
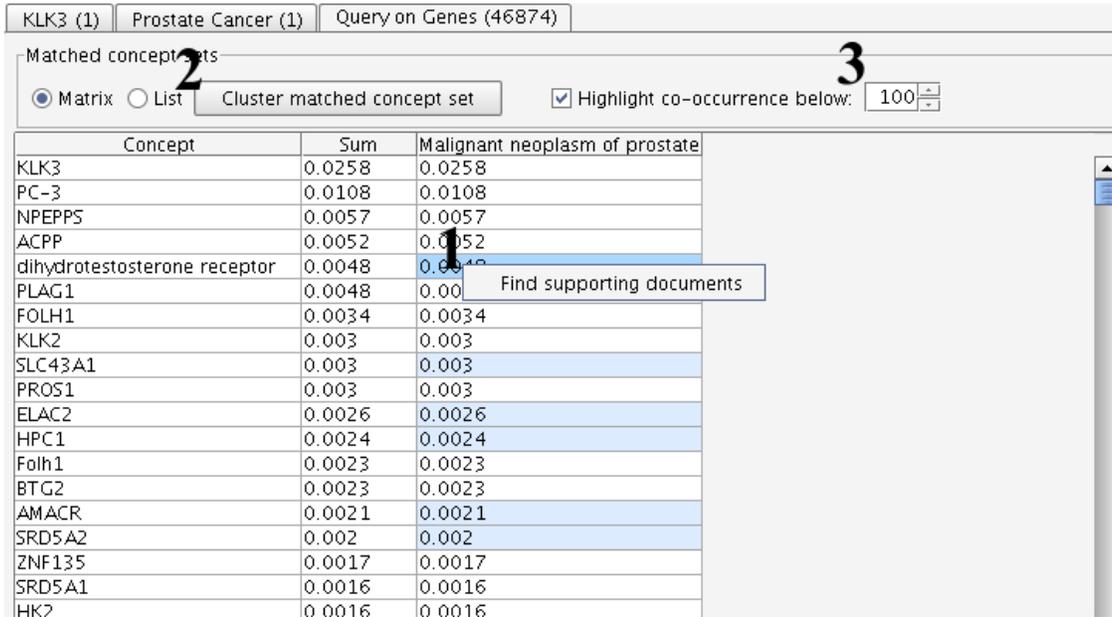


Figure 5. The query result view.

1. Concepts are shown with their query score. The columns show a break up of the score over the query concepts. Right clicking these scores will allow you to see the supporting documents.
2. There are different views available on the data, including a list view and a clustered heat-map view.
3. In the matrix view, concepts that only co-occur seldom or never can be highlighted. In the list view, the actual number of co-occurrences in Medline are shown.



4. Matching concept profiles and exploring concept associations

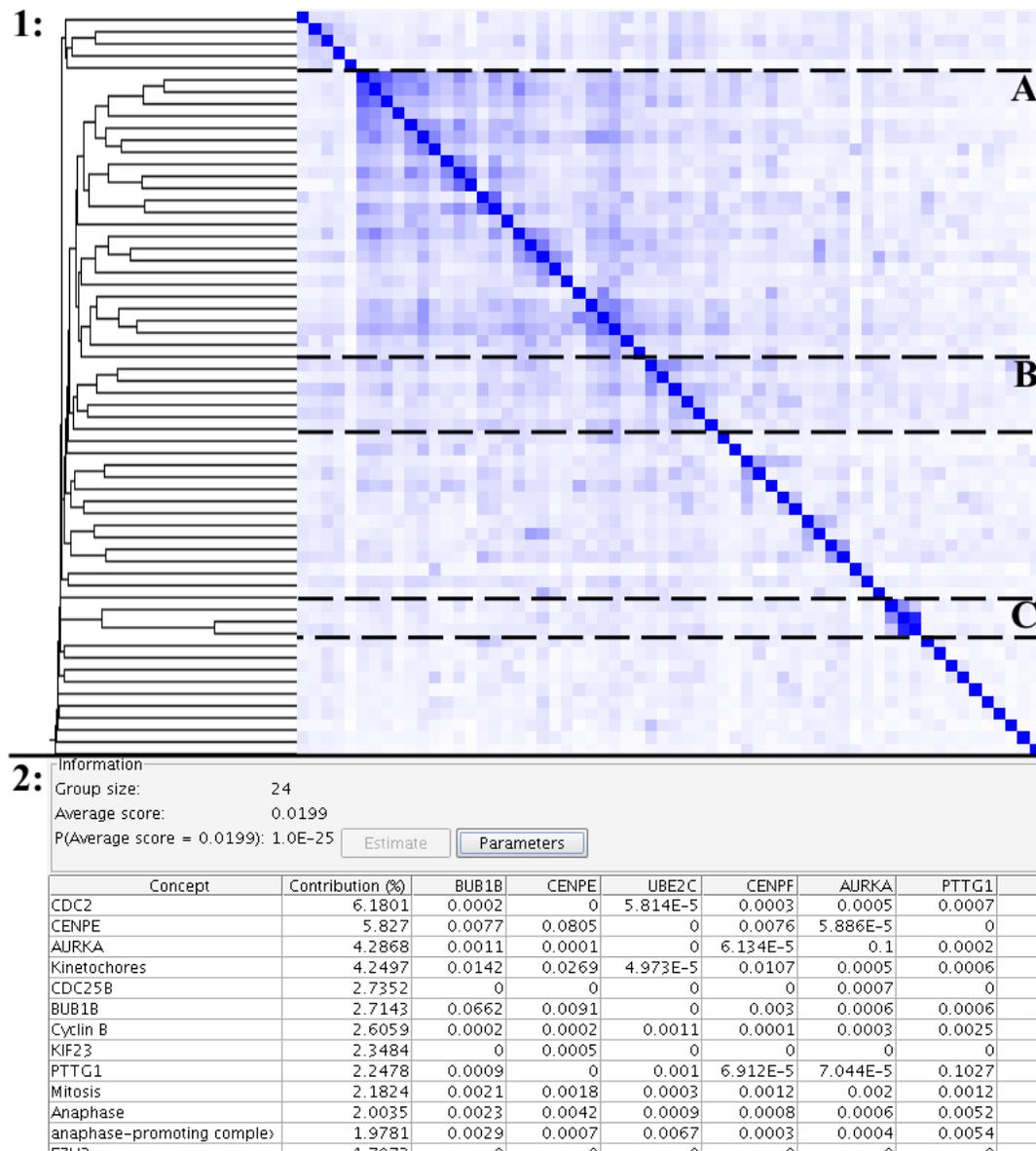
The classical way to use Anni is to explore associations between concepts in a set by matching the concept profiles. This allows you to identify concepts that have many intermediate concepts in common, but may never have been mentioned together in the literature. The concepts that are taken into account during matching can be controlled through a filter (Figure 2, #13). Concept sets can be used as a filter in either an inclusive or exclusive manner. Concept sets can be selected as a filter through the drop down list (Figure 2, #13) or through right clicking the set in the tree.

4.1. Clustering a concept set

Associations between concepts in a concept set as a whole can be explored through hierarchical clustering (Figure 6, accessible through icon #8 in Figure 2). Alternatively, in tables where the rows depict concepts, a number of concepts can be selected and clustered or projected through the appropriate icons or the right mouse click menus. If you find an interesting cluster of associated genes you can retrieve the similarities between the concept profiles in the Annotation view (Figure 6, #2). The annotation view provides a score for the group cohesion. The value of this score is dependent on the used filter for matching. To facilitate interpretation of this score, a p-value is given, which indicates how often a group with the same or higher cohesion score would be drawn at random. In the view the concept associations can be traced back to the underlying documents in Medline by right-clicking the association score and selecting 'supporting documents'.

Figure 6. Clustered heat-map of the associations in a concept set and annotation of a concept cluster.

1. A symmetrical clustered heat-map of matched concept profiles. The intensity-values represent the matching-score between the concept profiles. Three clusters are indicated. Clusters can be selected by clicking in the tree or in the heat-map.
2. The annotation of concept cluster A. This view is accessible through the right mouse click menu on selected concepts or icon Figure 1, #9. The annotation view displays for a cluster a cohesion score with a p-value, and a list of concepts with their contribution to the cohesion score as a percentage. The actual weights of the concepts in the concept profiles are also shown. The distribution underlying the p-value can be modulated through the parameters menu; the number of simulations can be adapted as well as the concept set used for sampling.

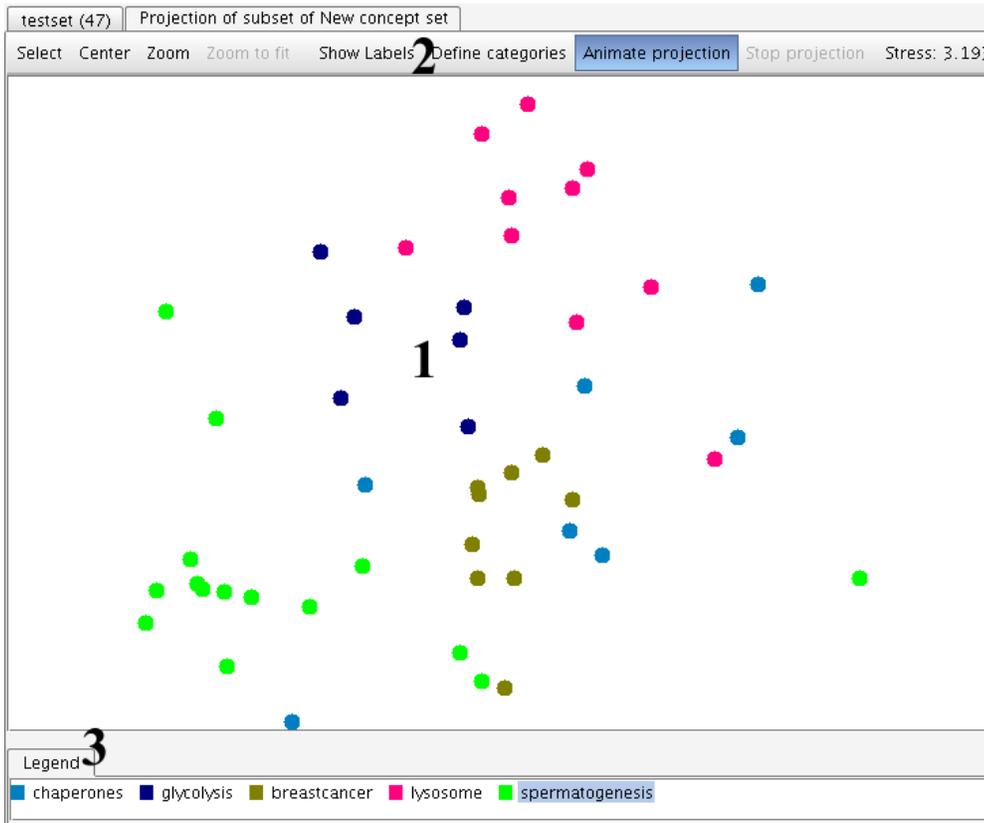


4.2. Projecting concepts

Concept associations can also be explored through a projection of the concepts with multidimensional scaling (MDS, Figure 7). In this projection, the distance between concepts reflects the matching value between the concept profiles belonging to those concepts. Similar to

Figure 7. MDS projection of a test-set of 47 genes consisting of 5 groups sharing a biological association.

1. The projection pane shows the concepts projected in a two-dimensional space. It is possible to select concepts in the projection pane.
2. The toolbar shows several options; you can select whether a mouse-click will result in the selection of a concept, centering the view around a concept, or zooming in or out. You can also define categories based on one of the meta data fields.
3. The legend shows the different categories. By clicking on a category, you can change its color.



the hierarchical clustering, groups of concepts can be selected and the similarities between their concept profiles can be seen in the Annotation view. Projected concepts nodes can be colored in groups defined in the meta data of a concept set.

4.3. Matching concept sets

Concept sets can be matched by selecting two concept sets in the tree and clicking the appropriate icon (Figure 2 #6). The resulting matrix can be viewed directly (similar to Figure 5), transformed to a list or can be explored through a 2d-clustered heat-map (Figure 8).

Tip: Many actions in Anni can be accessed through right mouse click drop down menus.

Figure 8. Clustered heat-map view of two matched concept sets.

1. These buttons allow zooming in and out and modulating the color intensities used in the heat-map.
2. Sections of the heat-map can be selected or the tree can be clicked to show corresponding subsections of the matrix.
3. A subsection of the depicted matrix.

