Anni 2.0 Tutorial: Analyzing a microarrray dataset

Version 1.0

The input file

In this example, we will use a set of genes that are significantly up or down regulated following agonistic stimulation of the androgen receptor in a prostate cancer cell line. The androgen receptor is a transcription factor, activated by the androgens testosterone and dihydrotestosterone. The androgen receptor is responsible for development and maintenance of the function of the normal prostate and for growth of early stage prostate cancer.

Illustration 1 shows the input file in a spreadsheet program. In this case, the input file contains three columns, with various information about the genes. Please note that this format is not compulsory! You can specify any columns you want, as long as at least one column contains identifiers, such as names, or in this case Entrez-Gene identifiers, that can be used to identify the concepts (in this case all concepts are genes).

The file is saved as a tab-delimited text file, and is available on the Biosemantics.org website.

	A	B	C
1	**Entrez Gene**	**Up or down regulated**	**Gene Name**
2	51182	down	HSPA14
3	8495	up	PPFIBP2
4	64207	down	C14orf4
5	3977	up	LIFR
6	5899	down	RALB
7	9967	up	THRAP3
8	23534	down	TNPO3
9	219404	down	MGC9850
10	79029	up	SPATA5L1

Illustration 1: The input file, shown in a spreadsheet program.

Importing the file into Anni

Click the *import concepts* button as shown in Illustration 2 to open the import dialog.

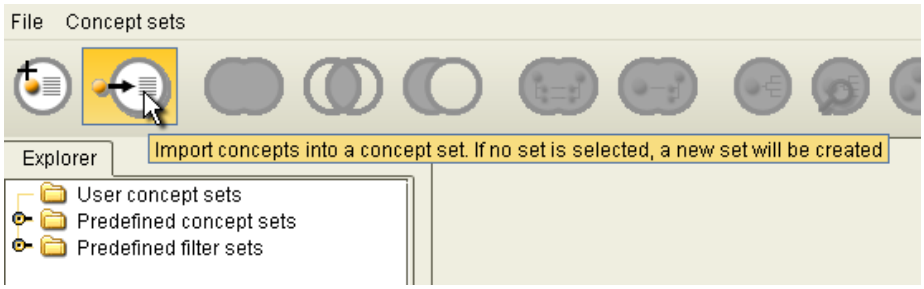

Illustration 2: Click the import button to load the file in Anni

Next, click the *load* button, select the file and click *open*. The file is now loaded into the table. Anni now needs to know which concepts belong to each line in the table. We need to map the entries in the table to concepts. To do this, we first specify which column is the *identifier column*, in this case the column with the title *Entrez Gene*. Next,we specify what type of identifiers is in this column, as shown in Illustration 3. A wide variety of identifiers is available. For genes, we highly recommend not using the gene names as identifiers, because these are highly ambiguous! Many genes have the same name, and this will require you to specifiy which concept you mean before you can proceed.

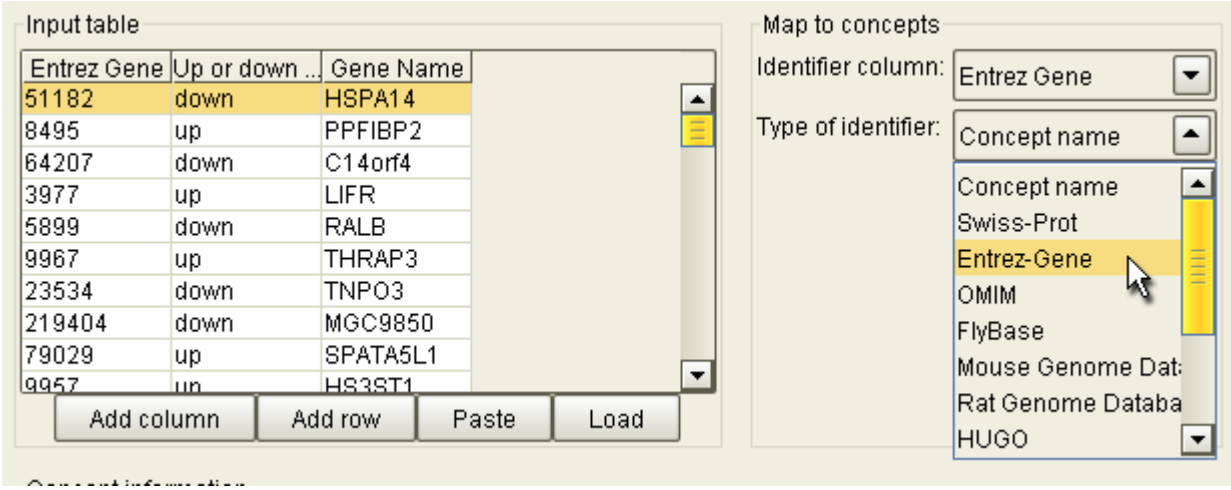

Illustration 3: The concept import dialog. The column specifying the identifiers is selected, and the type of identifiers it contains is specified.

After specifying the identifier column and type, click the *map to concepts* button. Anni will now retrieve the concepts, and add a column to the table. If you now click on one of the lines in the table, the concept information panel at the bottom of the dialog will show the available information about that concept. Click *Ok* to close the concept import dialog.

In the concept set explorer at the left side of the screen, you now see a concept set called "New concept set" in the *User concept sets* folder. If you like, you can rename the concept to a more meaningful name. Right click on the concept set and select *Rename item* from the menu as shown in Illustration 4. Type the new name, for instance "AR stimulated", and press enter.

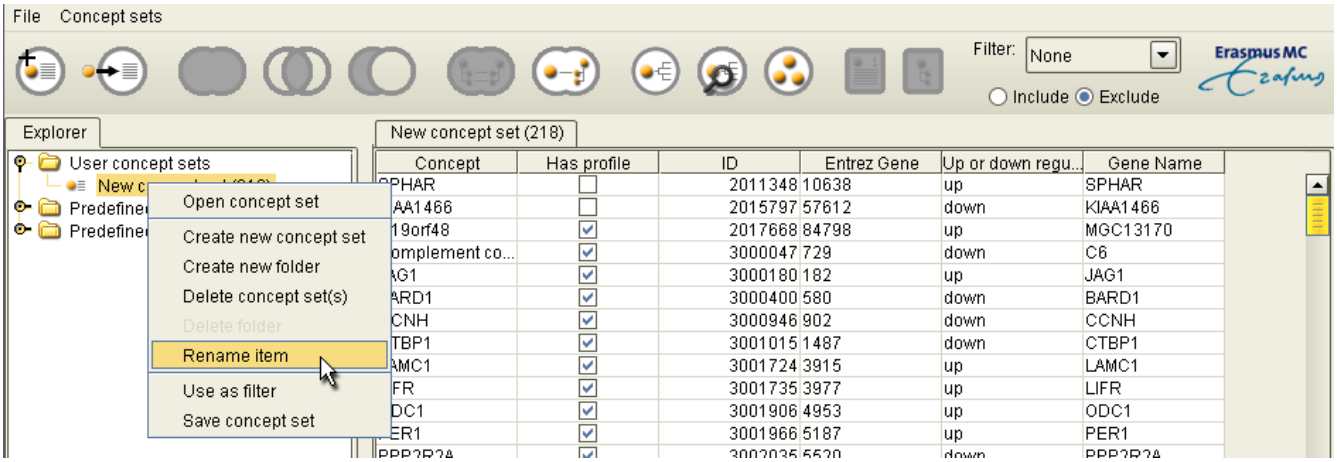

Illustration 4: Renaming the new concept set

Clustering the concepts

To understand what genes are semantically similar, we can now perform a cluster analysis. Select the new concept set in the concept explorer (it probably still is selected), and click the *Cluster concept profiles* button as shown in Illustration 5.

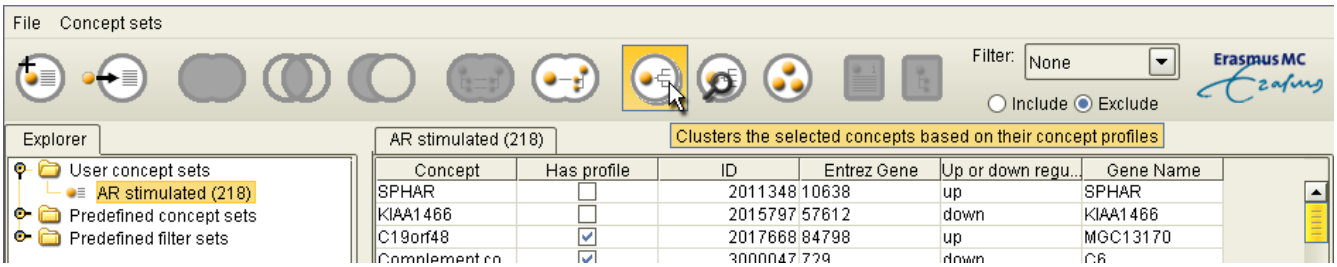

Illustration 5: The new concept set is selected, and we now click the cluster button.

It will take Anni several seconds to compute the clustering. The clustering view shows a dendogram and a heatmap. The intensity in the heatmap shows the similarity between concept profiles of the concepts. Concepts with highly similar concept profiles will cluster together. You can zoom in and out, and increase or decrease the sensitivity of the dendogram and heatmap.

By clicking on a line in the dendogram, you can select a cluster as shown in Illustration 6.

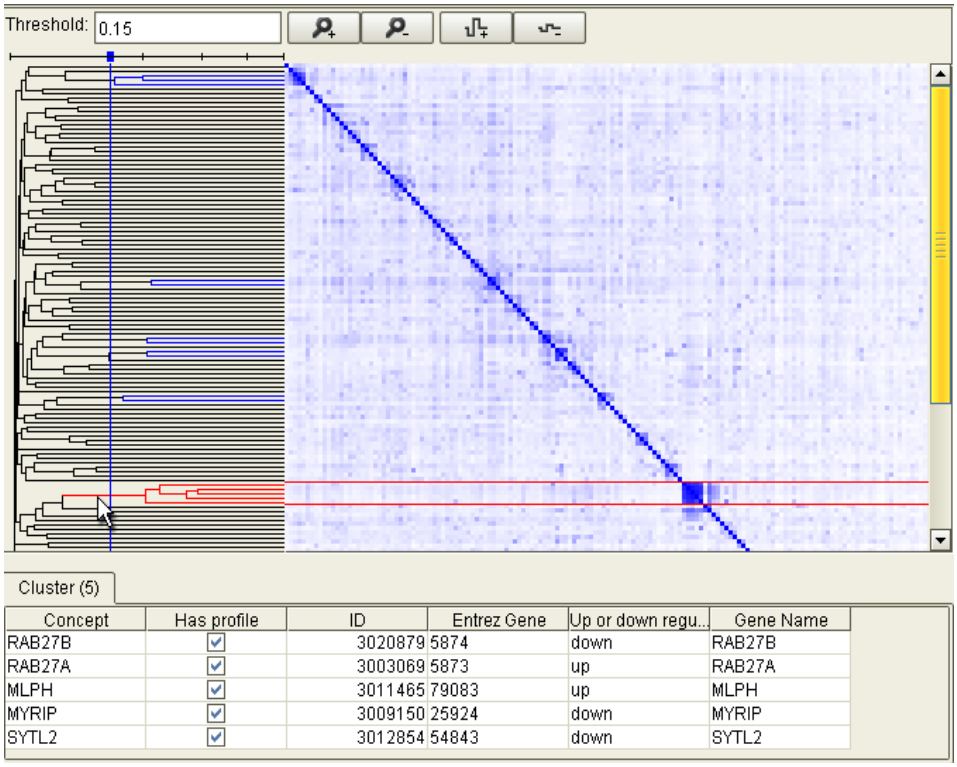

Illustration 6: Selection of a cluster in the cluster view. The concepts in the selected cluster are shown in the bottom panel.

Annotating the cluster

In this example, one of the strongest clusters contains 5 genes. If we want to understand what this cluster is about, we select those genes (click on the first gene, hold the shift button and click on the last gene), and click on the *Annotate selected concepts* button as shown in Illustration 7.

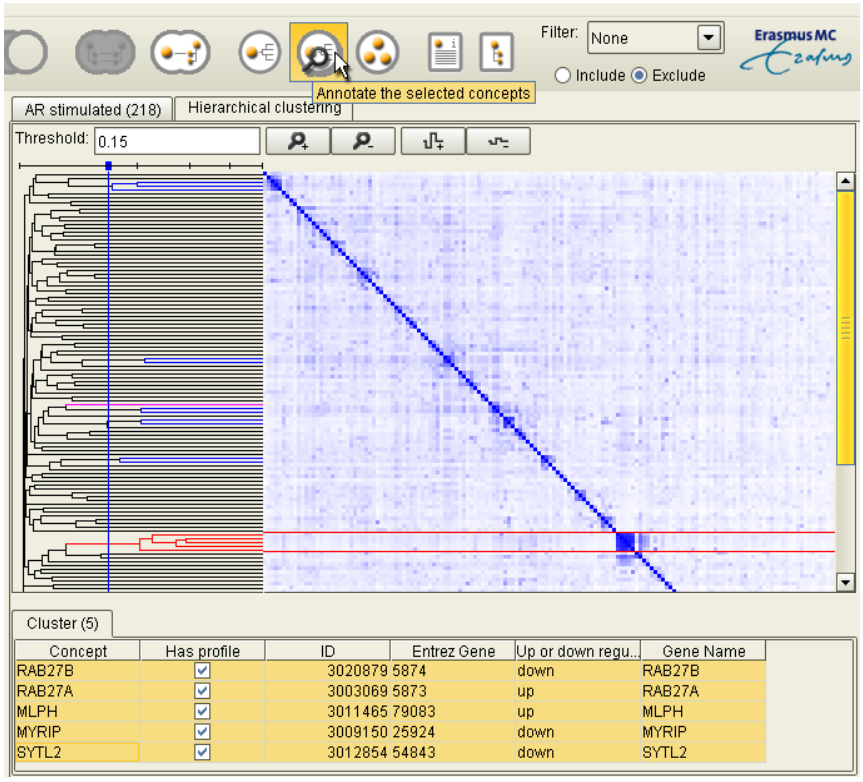

Illustration 7: Selecting the genes in the cluster and clicking on the annotation button.

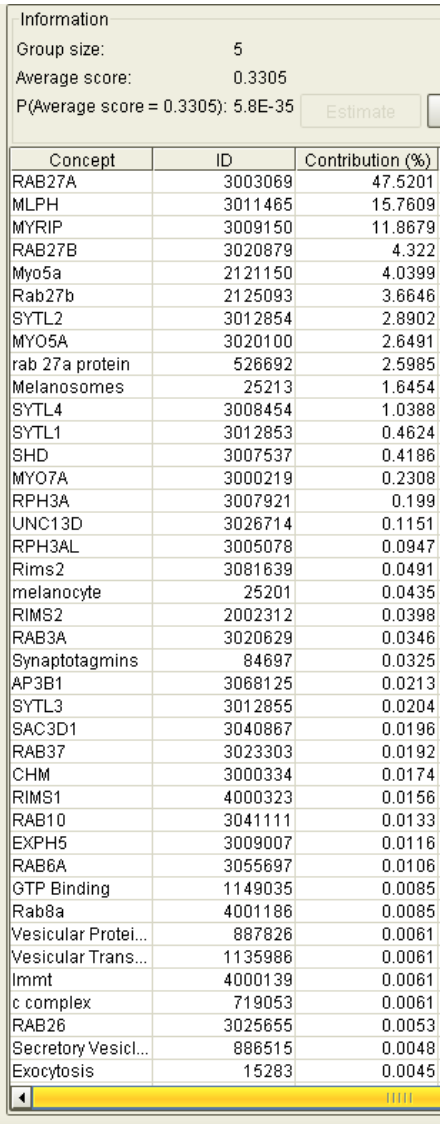

Illustration 8: Annotation of a cluster

The annotation view is shown in Illustration 8. In the top, we see a P-value that specifies how likely it is that we have selected a cluster at random. The first column shows which concepts contributed most to the similarity between the concept profiles in the cluster. The *Contribution(%)* column shows how much this concept contributed to the cluster.

In this example, we see that the genes themselves occur quite high in the list. This indicates that the genes are often mentioned together in the literature, and therefore are strongly related to each other. Many of the other concepts contributing to the cluster are also genes. We also find concepts such as “Melanosomes”, “melanocyte” and “Exocytosis”, which may give us a hint about the functions of the genes in the cluster.

Understanding the annotation

If we are unfamiliar with a concept that appears in the annotation, or any other concept we encounter in Anni, we can select that concept and click the *Concept information* button as shown in Illustration 9.

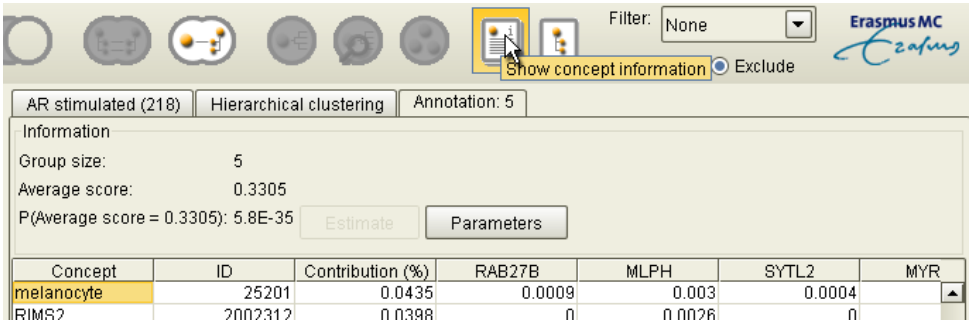

Illustration 9: When a concept is selected, you can click on the concept information button to find out more about that concept.

In this example, we could learn (if we didn't already know) what a melanocyte is, as shown in Illustration 10. For other concepts, such as genes or Gene Ontology terms, you will also find links to external databases about the concept.

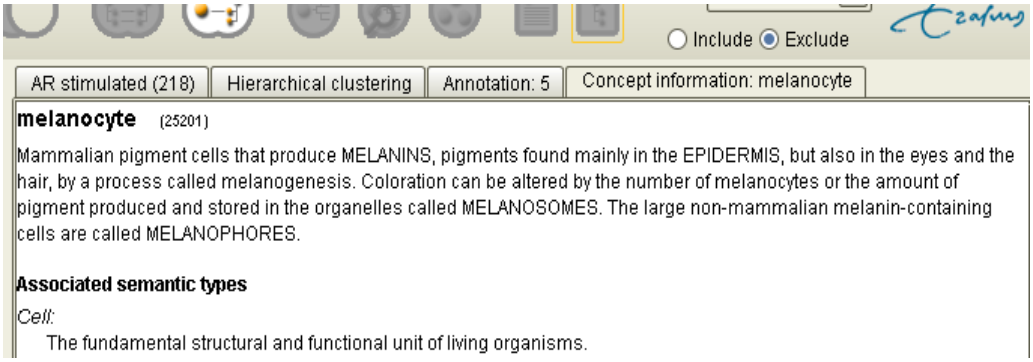

Illustration 10: Information on the concept 'melanocyte'

Finding the documents that support the annotation

In the end, if we found an interesting association, we would like to view the papers on which the association is based. In this case, one of the concepts that we are interested in is melanocyte, which is one of the concepts that binds the genes in our cluster together. If we want to know the relationship between melanocyte and one of the genes, for instance MLPH, we select the cell in the table that specifies the strength of the association between melanocyte and MLPH, as shown in Illustration 11. We then click on the right mouse button, and select *Find supporting documents*.

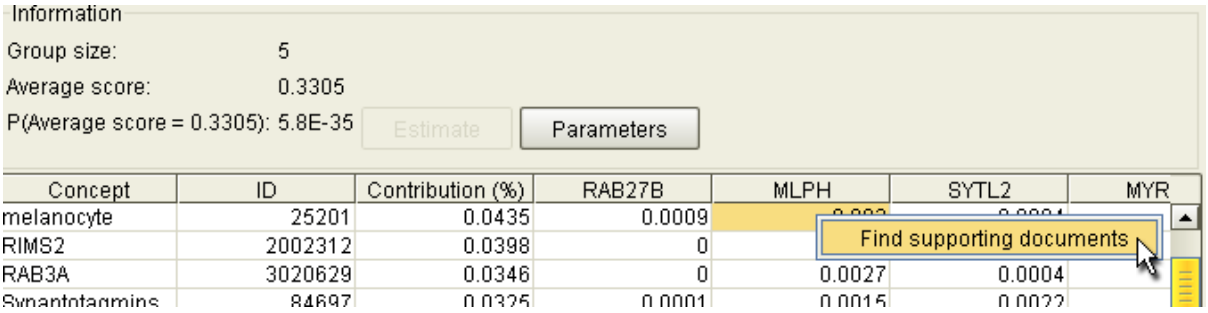

Illustration 11: Selecting an association in the annotation view and retrieving those documents that support the association.

We will then see a page that shows that there are 31 papers in which melanocyte and MLPH are mentioned together. If we click on the link on this page, we will be redirected to PubMed, where these 31 papers are shown, as you can see in Illustration 12.

When going through these documents, please remember that we have also included synonyms in our search for concepts. The concept may be mentioned by a completely different name than the preferred name that is shown in Anni!

Illustration 12: PubMed, showing the 31 articles in which melanocyte and MLPH are mentioned together.