BioCyc
Database Collection

Metabolomics Analysis with Groups

Video Script

Associated with "06_BioCyc-Groups_Metabolomics_120913.mov"

Note: Set of human compounds to use as an example:
http://biocyc.org/group?id=Biocyc13-61-3584666788

The file, "enrichment.txt" can also be used as a file import to get the compounds]

This webinar will discuss the analysis of metabolomics datasets with Web Groups.

We assume that you have created a compound group containing a set of metabolites of interest from your experiment, such as a set of metabolites that an external statistical analysis program has identified as significantly up- or down-regulated compared to a control. The group should be created with the BioCyc current organism set to the organism for which you have obtained the metabolomics data. As our example we will use a synthetically generated set of human metabolites.

We present three analysis approaches here: mapping the metabolites to the complete set of metabolic pathways they are present in; visualization of a metabolite set on an organism's metabolic map diagram; and statistical enrichment analysis of the metabolite set.

Mapping Metabolites to Pathways:

This mapping can be accomplished via a transformation. Select the transformation "Pathways of compound", then click the "+" in the last column to create a group containing the set of pathways in that column. We can see that *Coenzyme A* is present in several pathways, but several of the metabolites are not involved in any pathways.

Visualizing Metabolites on a Cellular Overview:

We call an organism's metabolic map diagram in BioCyc the cellular overview. To display a set of metabolites on the cellular overview, we'll go back to our original group, go to the Paint Data menu, and select "paint data on cellular overview." That brings up the cellular overview with the compounds that are in our group visually highlighted so that we can see their relationships within the pathway map. If we zoom in far enough, it will start to show you the names of some of these pathways. So for instance, here's a pathway that contains a few highlighted metabolites.

If we click a reaction or metabolite, we'll see a tooltip that identifies that entity. If we click "Keep Open," the tooltip window will stay open, and we can move it around

and create additional windows for use in a publication. The "E, R, P" buttons within reaction tooltips let us control how much information is presented.

The control panel at the right lets us remove or add back highlighting for multiple sets of genes that we might have highlighted at one time.

Quantitative concentrations values for multiple time points can also be painted onto this diagram. For example, if we go back to our group and select the concentration values column, we can then choose "Paint Data >"
"on Cellular Overview Omics Viewer." Make sure *Concentrations* is selected, and click, "Go." The cellular overview now shows the relative concentration of the compounds based on the Concentration column values.

Enrichment Analysis:

Enrichment analysis is a statistical technique for finding what a set of metabolites or other objects have in common with one another. That is, do some of the metabolites in our group belong to specific metabolic pathways at statistically significant levels? For example, enrichment analysis could tell you that a set of metabolites contains more metabolites involved in, say, fatty acid biosynthesis than you would expect to find by chance.

To perform an enrichment analysis, we make sure the Compounds column is selected, click on the Enrichments menu, and select "Compounds enriched for pathways." This pops up a menu with options that control the details of the statistical computation.

We'll just take defaults, although multiple options are available for the statistical test to use and for correction of multiple-hypothesis testing.

This analysis will create a new group of pathways. Some things to note about this group: the first column contains both individual pathways and pathway classes, which are distinguished by the use of uppercase initial letters. The second column is the p-value meaning the statistical significance, and the column is ordered by p-values, so the pathways at the top are the ones that are most over represented. The last column shows you the compounds from the original set that are present in this pathway. Let's delete that match column just so we can see the pathways a little better.

We see that these are all the pathways that the enrichment analysis has produced and we can see that a lot of them relate to cholesterol biosynthesis, so that's something we've learned. In addition, a number of the metabolites are involved in different human amino-acid biosynthetic pathways.