Tutorial: Microarray-based expression analysis part I: Getting started

December 15, 2009

CLC bio
Finlandsgade 10-12 · 8200 Aarhus N · Denmark
Telephone: +45 70 22 55 09 · Fax: +45 70 22 55 19
www.clcbio.com · info@clcbio.com

Tutorial: Microarray-based expression analysis part I: Getting started

This tutorial is the first part of a series of tutorials about expression analysis. Expression analysis often requires advanced skills in statistics, but this tutorial is intended to show a straight-forward example of how to identify and interpret the differentially expressed genes in samples from two different tissues. If you are familiar with the statistical concepts and issues within expression analysis, you may find this tutorial too simplistic, but we have favored a simple and quick introduction over an exhaustive and more "correct" explanation.

The data comes from a study of gene expression in tissues from cardiac left ventricle and diaphragm muscle of rats [van Lunteren et al., 2008]. During this series of tutorials, you will see how to import and set up the data in an experiment with two groups (part I), to perform quality checks on the data (part II), to perform statistics and clustering to identify and visualize differentially expressed genes (part III), and finally to use annotations to categorize and interpret patterns among the differentially expressed genes in a biological context (part IV).

Importing array data

First, import the data set which can be downloaded from the Gene Expression Omnibus (GEO) database at NCBI: `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6943&targ=gsm&form=text&view=data`. After download, click **Import** ( ) in the Tool bar and select the file. You will now have 12 arrays in your **Navigation Area** as shown in figure 1.

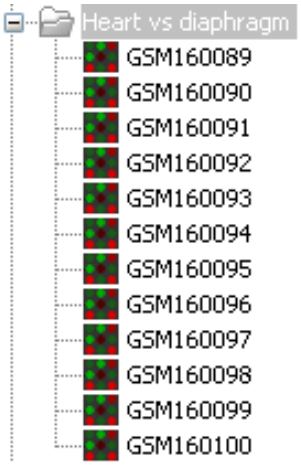


Figure 1: *12 microarrays have been imported.*

Grouping the samples

The next step is to tell the *CLC Main Workbench* how the 12 samples are related.

This is done be setting up an **Experiment** ( ). An **Experiment** is the central data type when analyzing expression data in the *CLC Main Workbench*. It includes a set of samples and information about how the samples are related (which groups they belong to). The **Experiment** is also used to accumulate calculations like t-tests and clustering.

First step is to set up the experiment:

Toolbox | **Expression Analysis ( )** | **Set Up Experiment ( )**

Select the 12 arrays that you have imported (see figure 2).

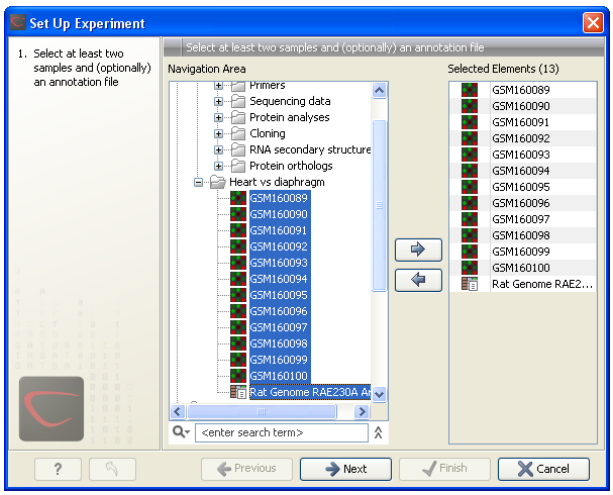


Figure 2: *Select the 12 microarrays that have been imported.*

Note that we use "samples" as the general term for both microarray-based expression values and sequencing-based expression values. Clicking **Next** shows the dialog in figure 3.

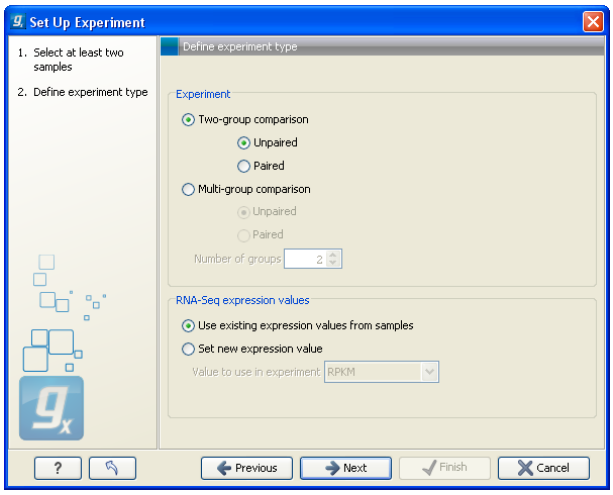


Figure 3: *Defining the number of groups.*

Here you define the number of groups in the experiment. Since we compare heart tissue with diaphragm tissue, we use a two-group comparison. Leave it as **Unpaired**. Clicking **Next** shows the dialog in figure 4.

Name the first group **Heart** and the second group **Diaphragm** and click **Next** (see figure 5).

Here you see a list of all the samples you chose in figure 2. Now select the first 6 samples (by clicking the first sample, holding down the mouse button while you drag and select the other five samples), right-click and select **Heart**. Select the last 6 samples, right-click and select **Diaphragm**. In this way you define which group each sample belongs to.

Click **Finish** and the experiment will be created. Note that the information from samples located in the **Navigation Area** is copied into the experiment, so they now exist independently of each other.

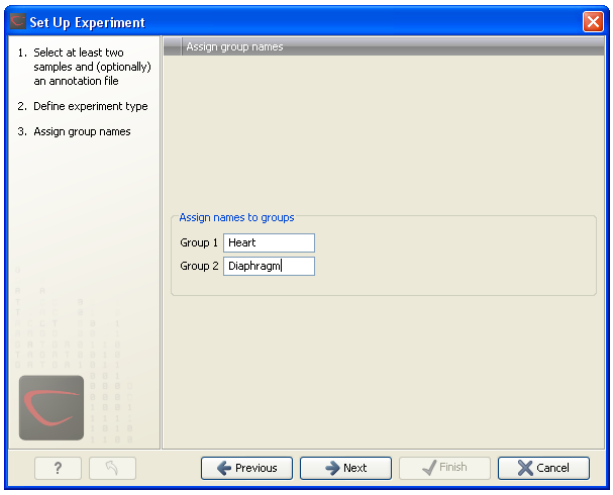


Figure 4: *Naming the groups.*

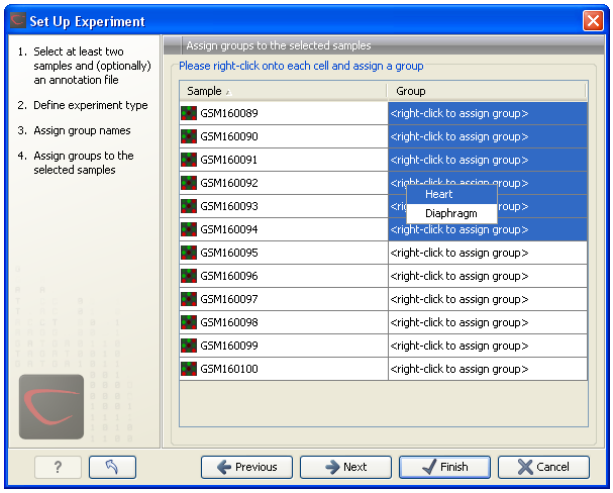


Figure 5: *Assigning the samples to groups.*

The experiment table

Once it is created, the experiment will be opened in a table as shown in figure 6.

The table includes the expression values for each sample and in addition a few extra values have been calculated such as the group mean, the IQR (Interquartile Range) values and the present counts for the whole experiment and the individual groups (note that absent/present calls are not available on all kinds of data).

Save the experiment and you are ready to proceed to the expression analysis tutorial part II.

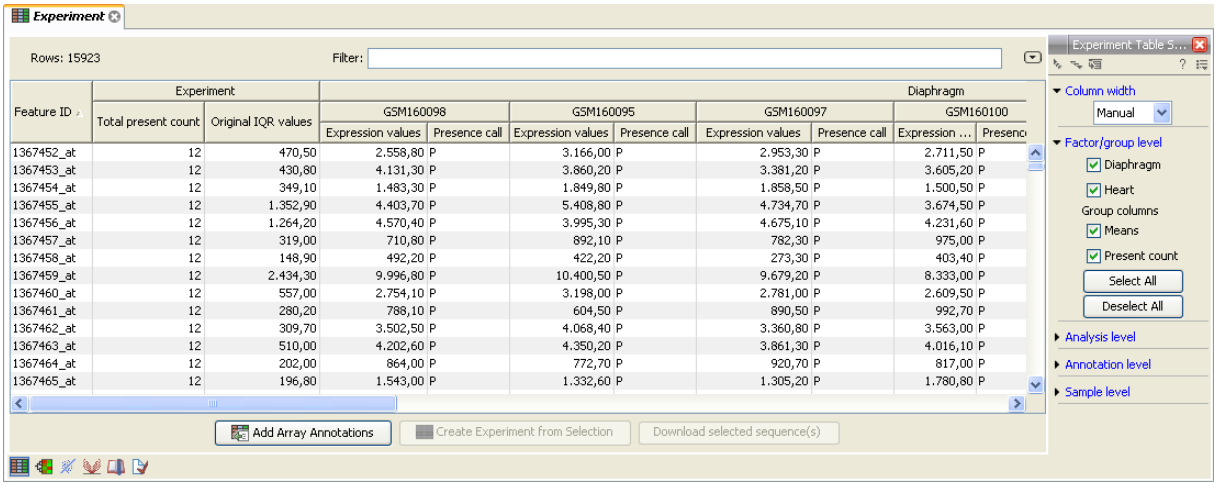


Figure 6: *The experiment table.*

Bibliography

[van Lunteren et al., 2008] van Lunteren, E., Spiegler, S., and Moyer, M. (2008). Contrast between cardiac left ventricle and diaphragm muscle in expression of genes involved in carbohydrate and lipid metabolism. *Respir Physiol Neurobiol*, 161(1):41–53.