

J-Express Pro

tutorial documents

Loading Affymetrix txt data

Pre phase

Two types of affymetrix output files are supported by default in the J-Express Pro framework. One is the tab-delimited text files normally with the *.txt* suffix. The other is the cel files combined with a cdf (Chip Definition File) file. This document describes loading tabular data files (*.txt*) and not the cel files.

Cel files can be loaded using the RMA interface (under the File menu).

Getting Started

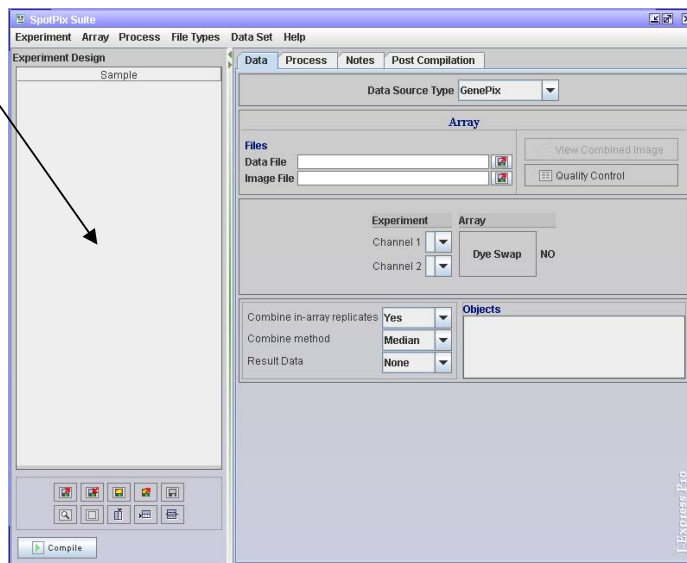
Start the J-Express application and open the SpotPix Suite by selecting *load raw data (SpotPix Suite)* from the *File* menu. This is the main framework for loading and preparing data directly from image analysis systems. See also documentation available by clicking *F1* while the SpotPix Suite window is activated.

Setting up the experiment

There are several ways of setting up the experiment select one of the following:

1. Drag and Drop: Drag the files from your file explorer onto the experiment table. Double-click in the first column to change the names of the samples.

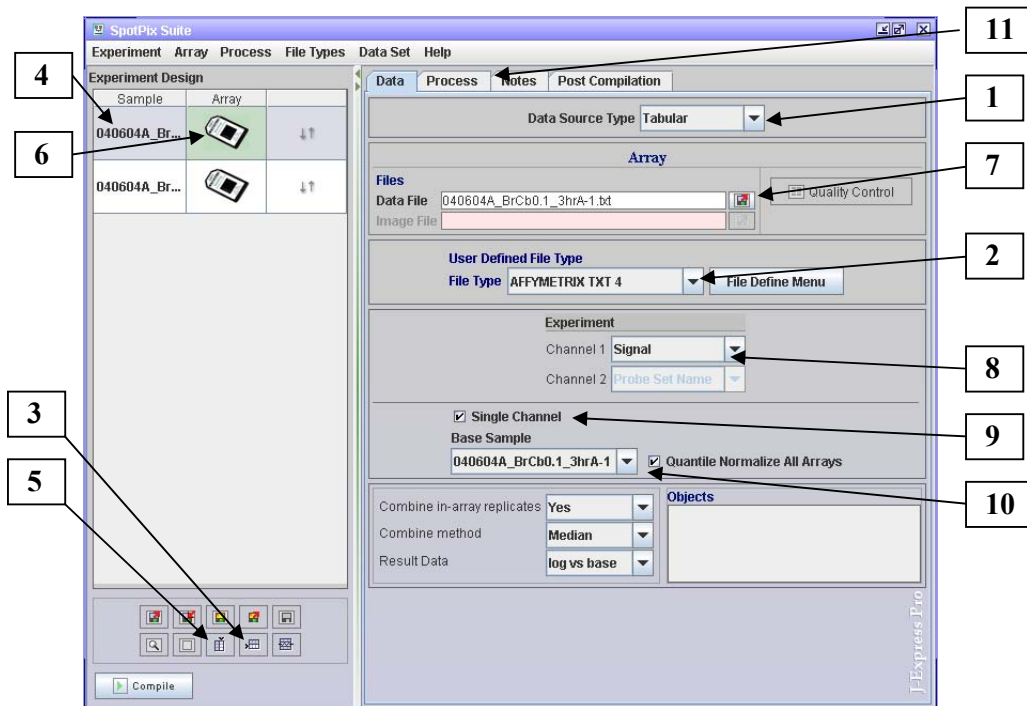
Drop here



2. Load Experiment From File List: select this option from the *experiment* menu. This opens a file browser and enables the selection of multiple files. Hold the

ctrl button down on your keyboard to select more than one file. Make sure all files are affymetrix data files. Double-click in the first column to change the names of the samples.

3. Manually set up the experiment by first choosing the *Tabular* format in the Data Source Type Combo-box (1). Set the File Type to *AFFYMETRIX TEXT* (2). Start adding the data files by clicking the *Add Sample* (3) button or select from the *Experiment* menu. Double click in the first column (4) to set the names of the samples in your experiment. Click the *Add Replicate Column* button (5) or select from the *Experiment* menu. Create an abstract array by right-clicking in a row on the column called *Array* and selecting *Add Array* (6). Set the file reference by clicking on the array-icon (6) and opening the file chooser by clicking on the choose data file button next to the data file field (7).



repeat this procedure for all your data files.

Setting up the experiment

Set the correct channels by clicking the *channel 1* combobox (8). Set this to *Avg Diff* (If the data column is called *Avg Diff*). Click the *Single Channel* Checkbox (9) and select a base sample (10) (to normalize against).

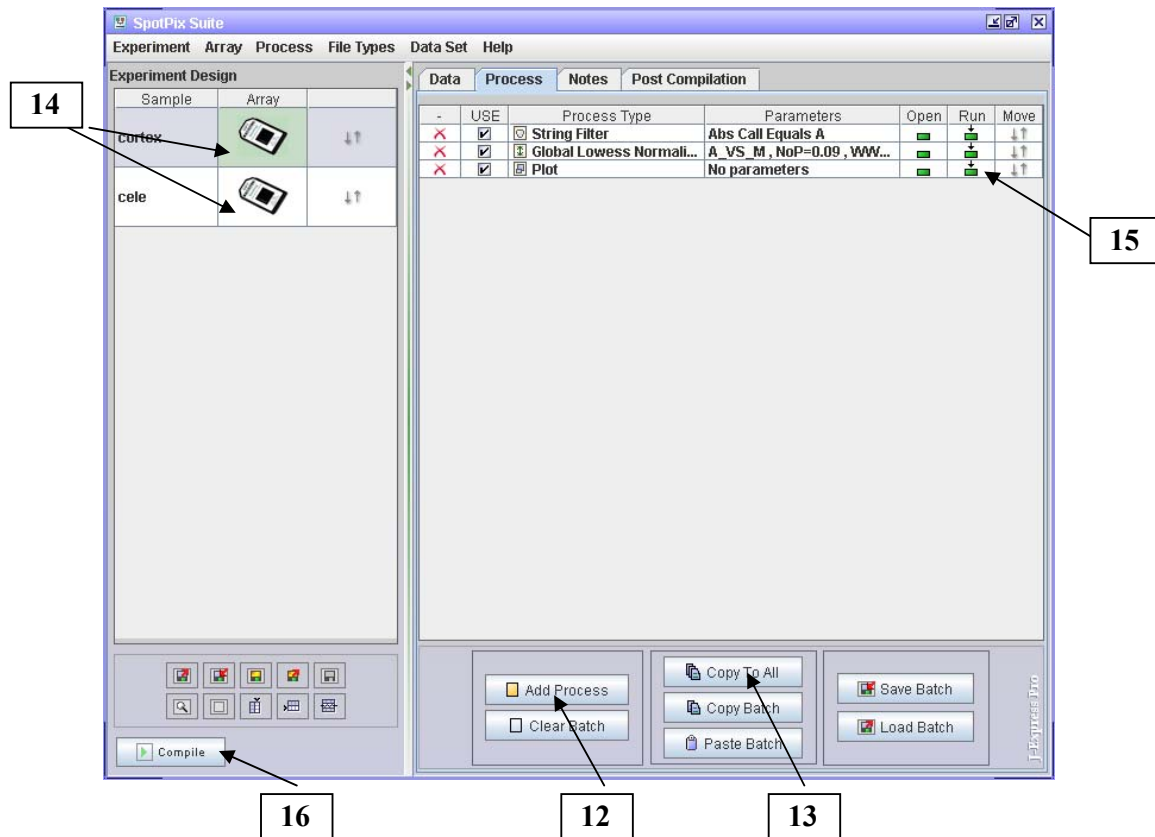
Now choose if you would like to combine replicates, select the combine method you like and set the Result data to *Log vs Base* (if this is what you want). *Log vs Base* will give you log ratio values for the sample and the base sample. If you want the absolute values, select *NONE* in the Result Data combobox.

Preparing the data

Filtering and normalization is normally a very important step of microarray analysis. For affymetrix data however, it seems like filtering is the most crucial step (The dye

effect from two-channel arrays is not present). Click the *Process* tab (**11**) to create a batch of a simple normalization, a filter and a plot to control the layout of your data:

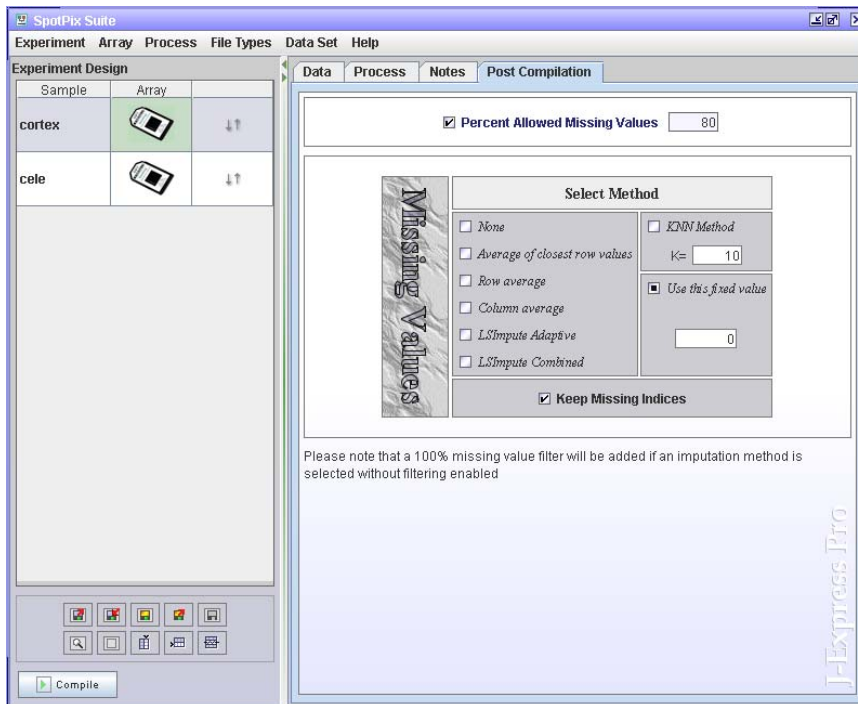
1. Click the *Add Process* button (**12**)
2. Click the *String Filter* button
3. Set the first combo box saying experiment to *Abs Call*
4. Set an 'A' in the text field saying regular expression and click filter. Control that any ABSENT features are filtered. Click *OK*.
5. Click the *Add Process* again.
6. Click the Normalization button to the left.
7. Select the *Median* method. Click Plot to see the normalization result (If the selected array is the base array, all spots will be located on the diagonal).
8. Click ok.
9. Click the *Add Process* again.
10. Click on the *Other* button to the left.
11. Select Plot and click OK.
12. Click the *Copy To All* button (**13**) to copy this process batch to all your arrays.
13. Control your arrays by selecting the different arrays (**14**) and clicking the run column on the plot row (**15**) and the *plot* button in the scatter plot window.



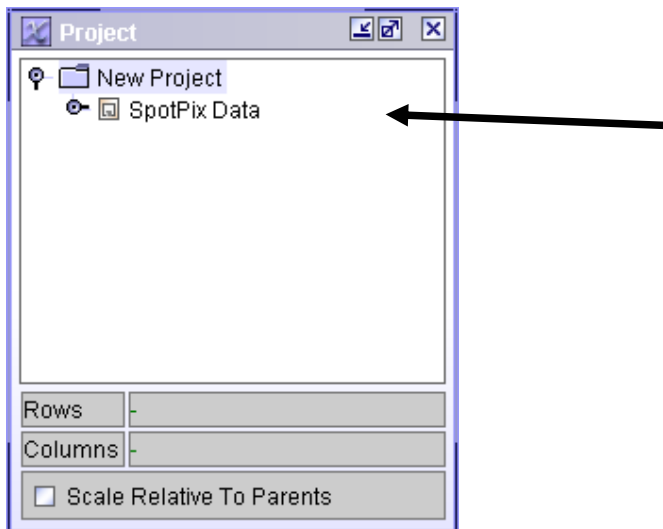
You have now created a simple filter, normalization method and a plot to process your arrays

1. Click the *Post Compilation* tab.
2. Let the Percent allowed missing values be 80 (This will filter all genes having 80% missing values or more.).

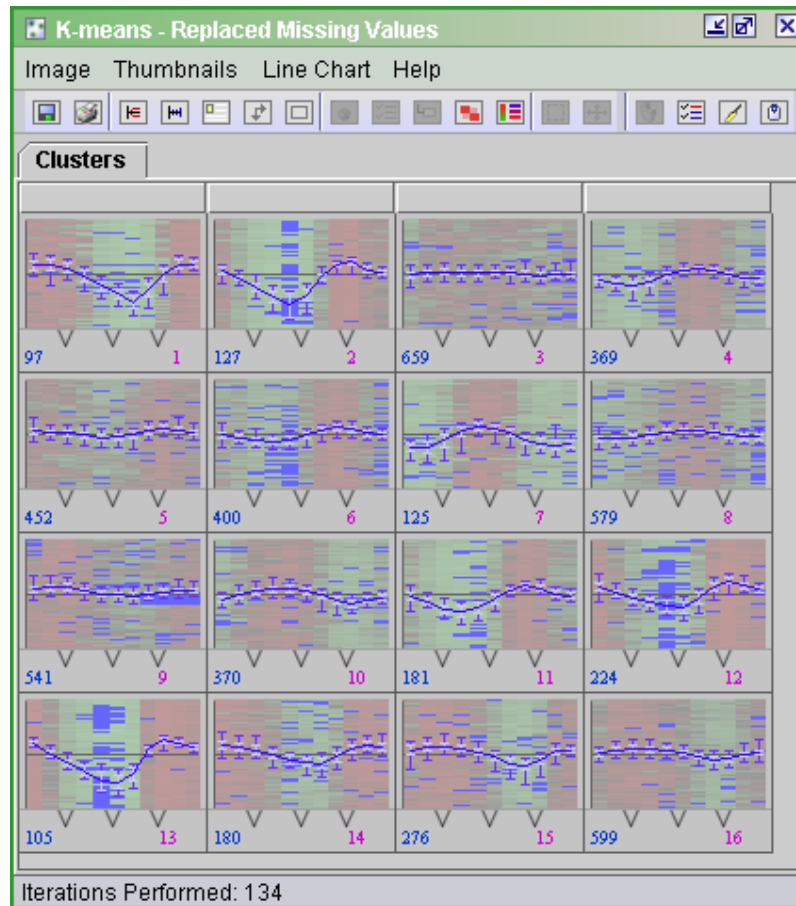
Set the missing values imputation method to be “Use this fixed value” and let the value be 0. Have the keep missing indices selected. This will set all filtered values to 0.0. (See documentation for information about the other imputation methods). The imputed values will be labelled so that you know this in later analysis.



Click the *compile* button (16) to create a gene expression matrix of the experiment.



3. Click on the new dataset (SpotPix Data).
4. Select K-Means from the Methods menu.
5. Click Ok in the K-Means clustering dialog.
6. Click on one of the clusters and select the new tab.



The Data are now grouped in regards to expression profiles.

7. Continue on your own with downstream analysis. Good luck☺