

PreP+07 protocol for Supervised Lowess Ratio Shift Correction

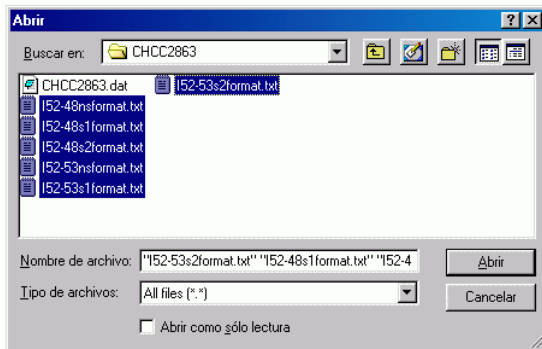
INTRODUCTION

Supervised Lowess Ratio Shift Correction is implemented in PreP+07. Now PreP+07 (manuscript in preparation) can load DNA microarray including an additional filter-column to specify to either use the measurement (gene) for the Lowess procedure or not. If a gene is not used for the lowess procedure, then its change in LogRatio is obtained by interpolation from the Lowess Fit.

Gene selection is based on a parametric matching criteria: a certain keyword or phrase contained in the filter column. This will do lowess procedure based only to those genes and then interpolate the others ones with corrected values in function of their weighted distance.

OPERATION

1.- Loading data



Use the button “add slide” to include new data sets.

Be sure if you want to do supervised Lowess that the slide has a column with genes to use.

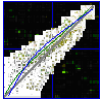
PreP - [PreParation 1]	
File Edit View Grouping Calculations	
Spots	7776
Name	i48_23_pru...
raw intensit...	-
raw intensit...	-
background ...	CONTROL BKG
background ...	TARGET BKG
Net intensit...	-
Net intensit...	-
Row	X
Column	Y
Grid	GRID
Description	Description
Use	Use
Ignored cells	Ignored cells

Assign functionality (describe the meaning of each column in the data set)

Assign grid distribution.

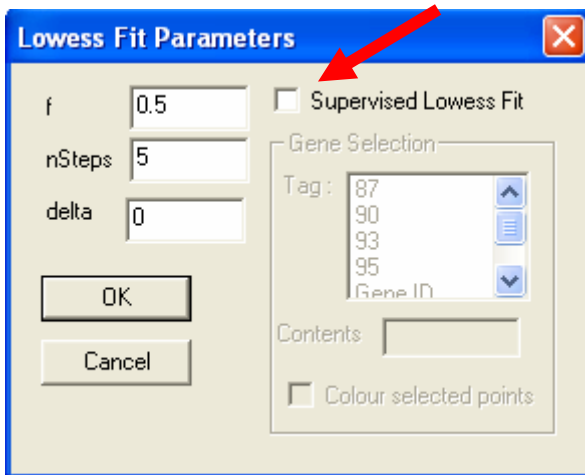
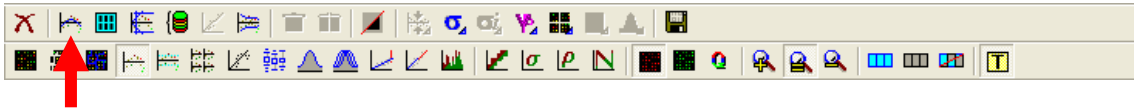
In this example, the column “Use” will be used for filtering genes.

At this point the data have been loaded, and they are ready for pre-processing.

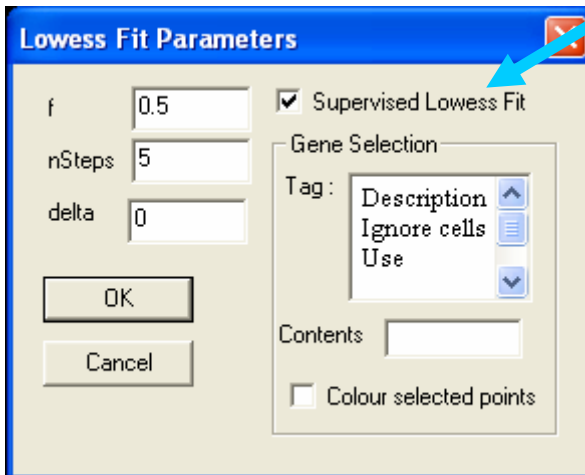


2. - Selecting Genes and Lowess Fit on them.

Supervised Lowess has been implemented in the same button than Lowess, doing Supervised Lowess or Lowess is an option in a dialog box that appears when you click on that button:

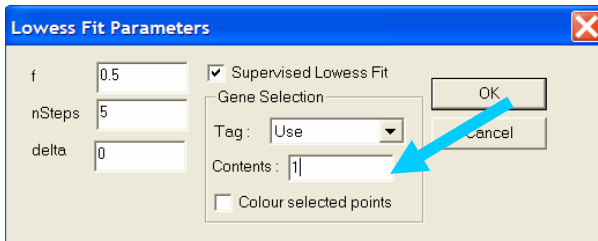


First of all you must select the Supervised-Lowess option by clicking the appropriated check-button



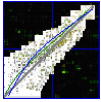
The “Gene Selection zone” becomes active. Then, you must specify the filter-column (the column to be uses to select the genes on which we are going to do Lowess).

- Click on the right arrow of “Tag” to pop-ups the available column Selection List
- Select the column which got the selecting values.

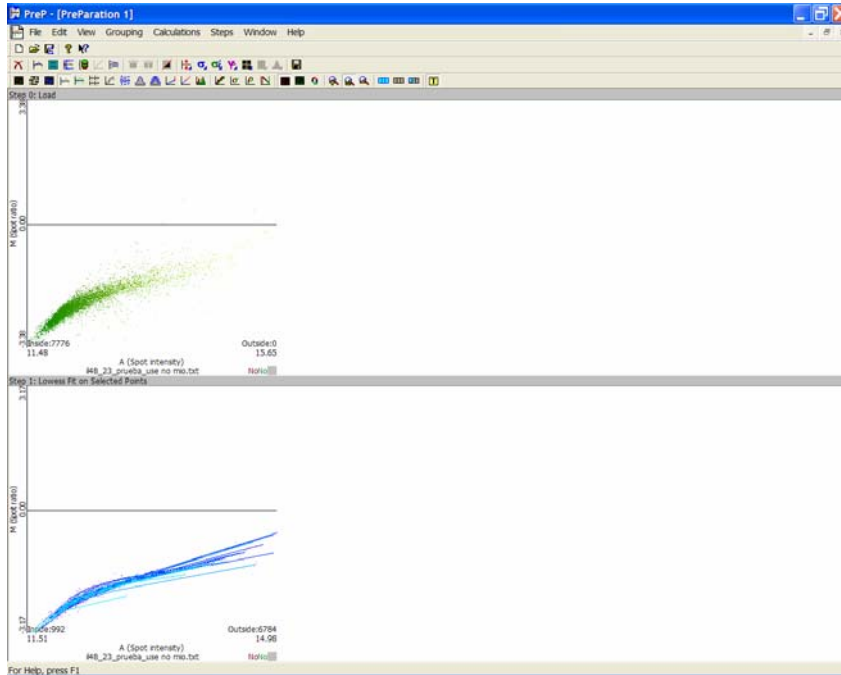


Additionally, you must specify the “selection” value. Genes will be used in the Supervised-Lowess if they contains a given value

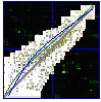
- Write in “Contents”, the value which select the genes (1).



- If you want to use a different color to differentiate the selected genes click on “Color Selected Points”, that will color those genes in purple.
- Write the F, nSteps and delta values for lowess on Selected genes.
- Click on the “OK” button.

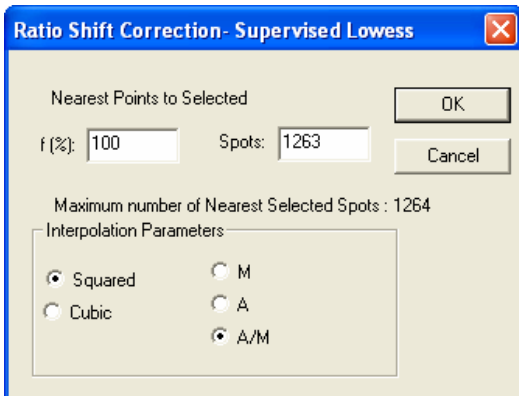
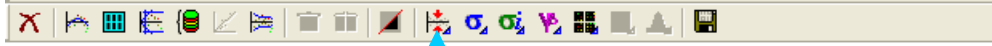


Now selected spots are filtered and we have done Lowess Fit on them, you can see the Fit Lines and selected genes in purple.



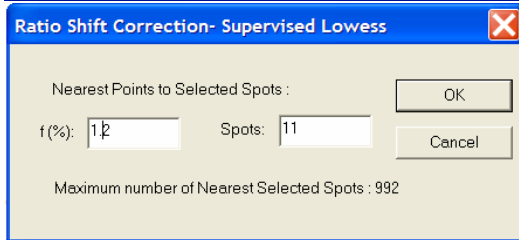
3. - Ratio Shift Correction.

The same button is used for Lowess and for Supervised Lowess procedures. PreP+07 is able to recognize whether a Supervised fitting was performed (or not) as previous step.

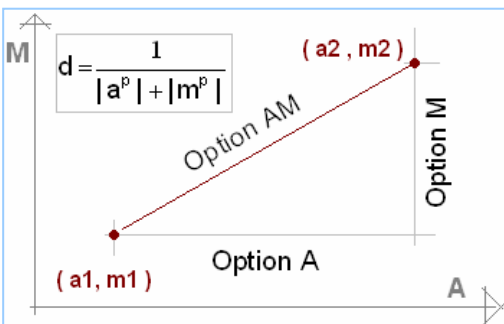


A dialog-box is opened.

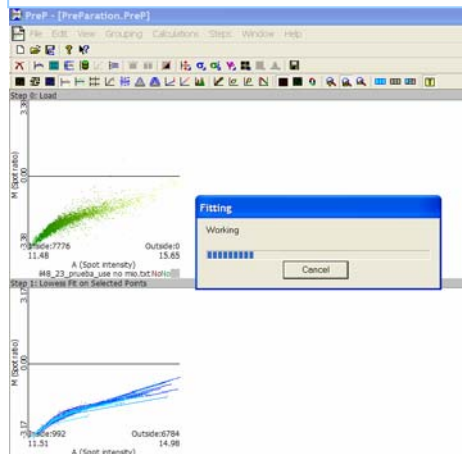
The interpolated ratio shift for 'not selected genes' is computed based on the nearest 'Selected Genes' to each spot. PreP+07 need the percentage of the Selected Spots or the number of Spots (both values are synchronized) in such a way that if we change the percentage (f%) the number of Spots in the other field is automatically updated and vice-versa.



The interpolation parameters are used to control the Interpolation method and distance. On the left a distance scheme is shown. The default distance is Cartesian, but can also be based only by the A or M values. Square or cubic exponent can be used to compute distance.

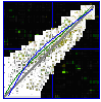


When ready, click OK, and the progress bar is active while the program is working.

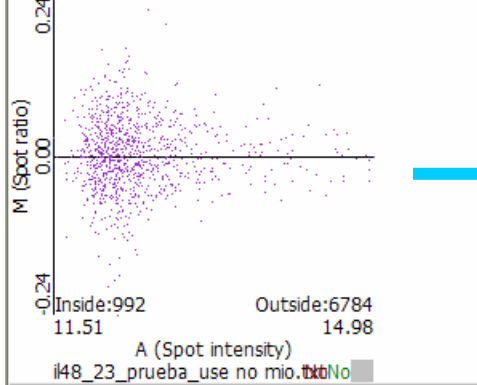


When the process finished, two new steps are displayed, the first of them has the Ratio Shift Correction on selected Spots and the second the Ratio Shift Correction interpolated.

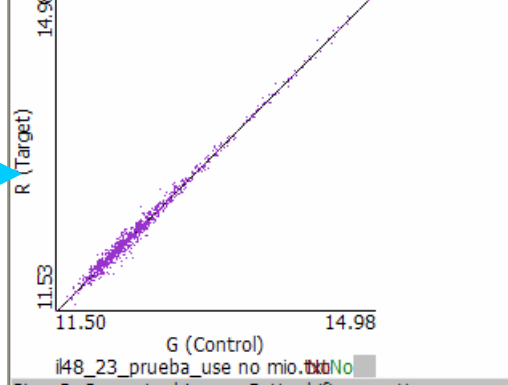
Supervised Lowess is done now.



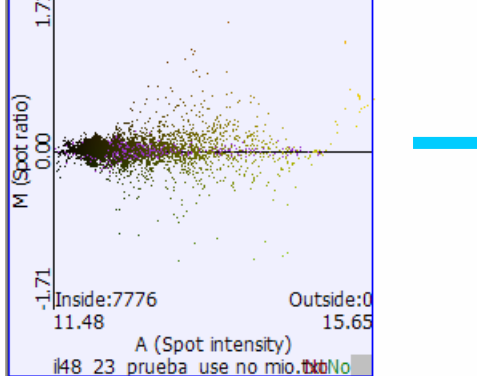
Step 2: Ratio shift correction to selected points



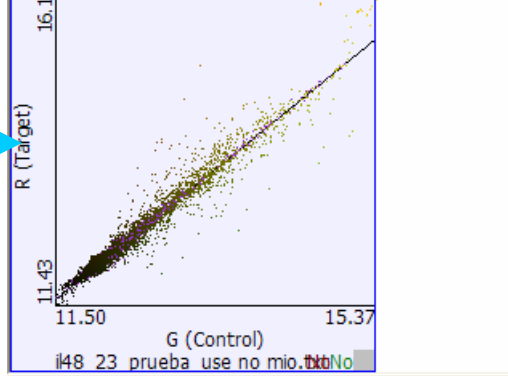
Step 2: Ratio shift correction to selected points

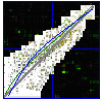


Step 3: Supervised LowessRatio shift correction



Step 3: Supervised LowessRatio shift correction





PreP+07 Guided Exercise for Supervised Lowess Ratio Shift Correction

This exercise is aimed to show the procedure used for Supervised Lowess (SL) ratio shift correction.

Data file: SLtest.txt (available in <http://chirimoyo.ac.uma.es/prep/index.html>).

File content.

This file contains the measurements on a gene-expression experiment with 4484 rows and the following columns (see Figure 1)

- *Gene ID*: gene identifier. As can be observed several replicates are available for each gene (table has been ordered by Gene ID column)
- Row, column and Grid: spot coordinates.
- Raw {A}, Raw {B}: Control and target raw intensities
- Bg {A}, Bg {B}: Background intensities for control and target respectively.
- Identity (%): Percentage of sequence identity for the corresponding sequences (control and target samples) obtained by blasting the sequences.
- 95, 93, 90, 87: flag used to mark genes with a given degree of similarity (values have been taken as integers, this is to say the column “95” contains a 1 if the similarity is greater than 95 (96 or more). This flags can be set with any other criterion, and can contain numeric values as well as alphanumeric values.

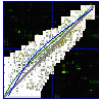
Gene ID	Row	Column	Grid	Raw {A}	Raw {B}	Bg {A}	Bg {B}	identity (%)	95	93	90	87
IL1403_accA	12	4	4	21618,0282	12333,8451	18123,7234	6381,85106	100	1	1	1	1
IL1403_accA	3	4	4	19793,0704	12832,5634	17070,617	6566,59575	100	1	1	1	1
IL1403_accB	5	4	12	24933,9859	10307,5493	15949,8298	6657,85106	100	1	1	1	1
IL1403_accB	14	4	12	24358,0423	10263,8451	17074,8085	6616,06383	100	1	1	1	1
IL1403_accC	17	4	8	32049,0704	14286,5211	19638,234	6964,6383	100	1	1	1	1
IL1403_accC	8	4	8	33752,4366	14112,6056	18845,6596	7067,17021	100	1	1	1	1
IL1403_accD	12	4	3	27870,2535	11454,6901	17069,5106	6772,34043	99,492	1	1	1	1
IL1403_accD	3	4	3	32280,4789	10908,6338	19904,7234	6685,57447	99,492	1	1	1	1
IL1403_ackA1	5	4	11	35869,4085	18202,9859	19035,1702	8298,21277	95,408	0	1	1	1
IL1403_ackA1	14	4	11	32127,9437	16111,3803	19833,7234	8107,76596	95,408	0	1	1	1
IL1403_ackA2	17	4	7	31266,3662	15521,4648	16978,5958	7139,31915	90,863	0	0	1	1
IL1403_ackA2	8	4	7	30989,9859	15661,4507	17986,766	7119,57447	90,863	0	0	1	1
IL1403_acmA	3	4	2	27700,3944	23976,8732	16129,3192	8591,93617	90,306	0	0	1	1
IL1403_acmA	12	4	2	27464,4648	23701,831	16656,7021	7752,19149	90,306	0	0	1	1
IL1403_acmB	14	4	10	23759	21923,1972	17096,2553	7437,46809	85,714	0	0	0	0
IL1403_acmB	5	4	10	24258,5775	21931,6197	15933,2979	7601,19149	85,714	0	0	0	0
IL1403_acmC	8	4	6	23130,7324	15802,1127	18334,7234	7637,57447	81,633	0	0	0	0
IL1403_acmC	17	4	6	23811,0845	17781,2676	17589,2979	8246,82979	81,633	0	0	0	0
IL1403_acmD	12	4	1	20479,7747	16426,169	14173,6596	6836,95745	89,286	0	0	0	1
IL1403_acmD	3	4	1	20077,5634	17383,9155	14941,7872	7021,29787	89,286	0	0	0	1

Figure 1.- First 19 rows in the test data file. Additional to the traditional columns with slide coordinates for each gene, on the right it has been included information about sequence similarity.



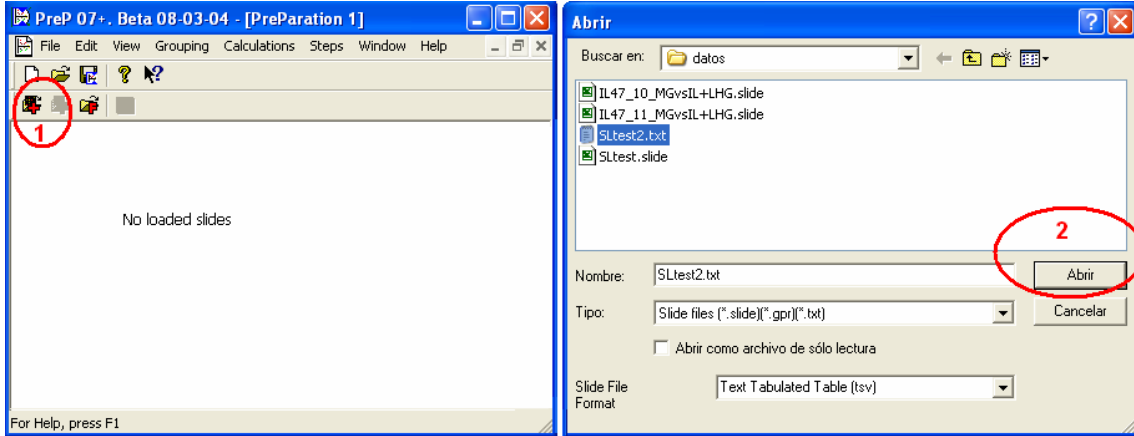
Click on PreP+07 icon to launch the program

Note: this document is only concerned with SL procedure. Additional documentation about Getting started and general directives to use PreP are available in the corresponding documents (see the web page)

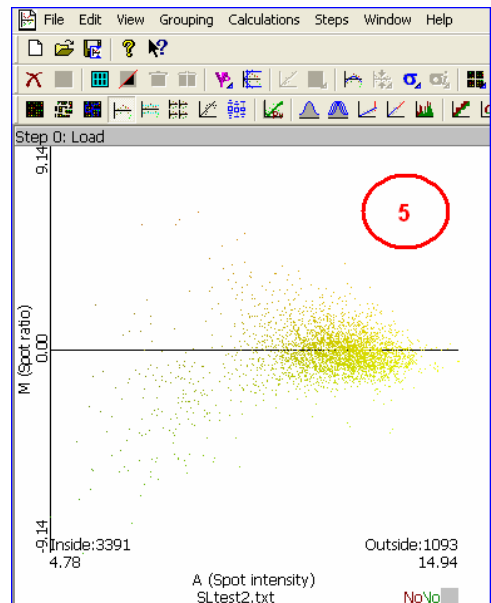
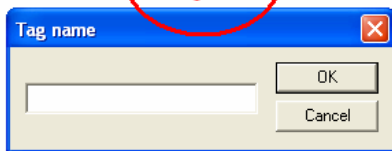
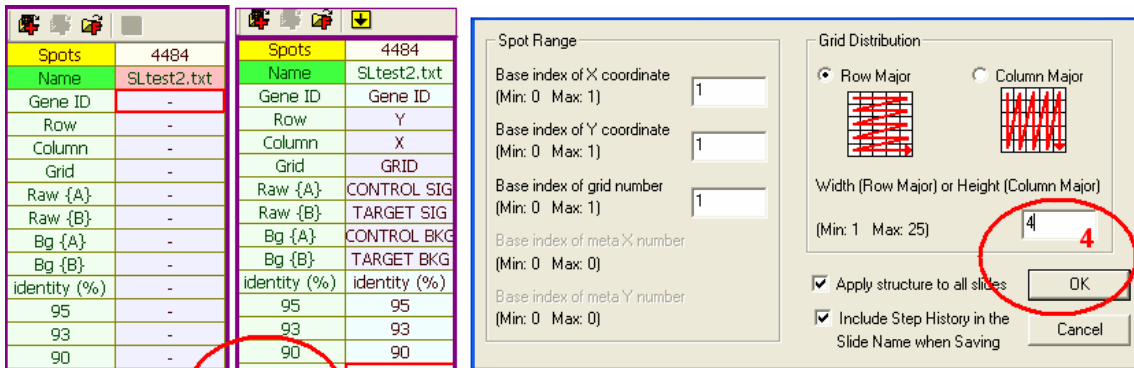


1. - Load step.

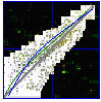
Use the “Add” button [1] for browsing [2] and up-load a new file for data analysis.



Use the right button over a given field to assign functionality [3: before and after functionality assignment]. To define “descriptive” columns (such as similarity or GeneID) press “Enter” to open a text-box. “Enter” again to define the column with the original label. Finally, define the Grid organization. In this case we use Row mayor and a 4x4 Grid design.

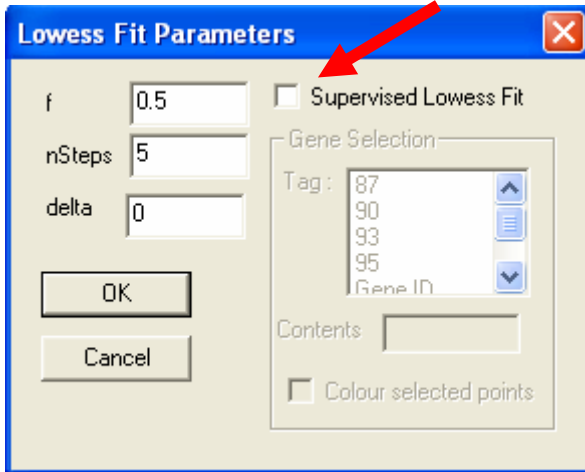
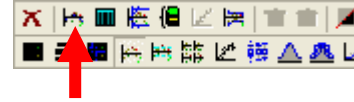


After complete the load procedure an image in AM graph fashion [5] will be displayed. PreP is ready to proceed.

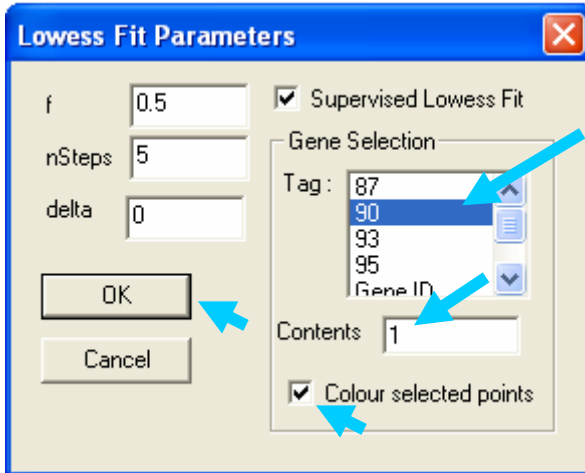


2. - Supervised Lowess.

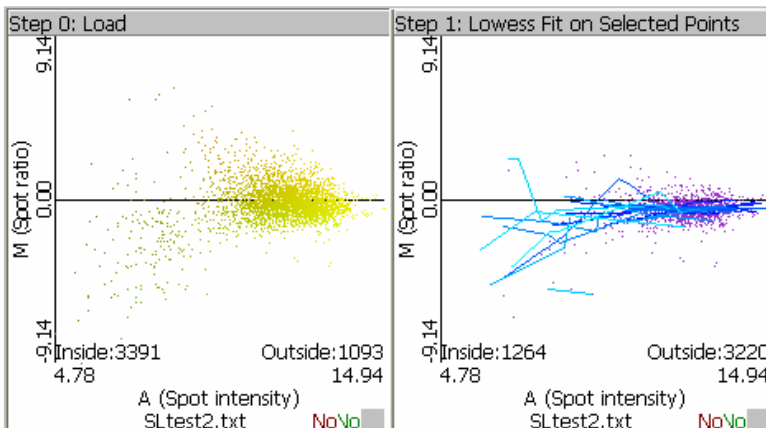
Select "Supervised Lowess" procedure:



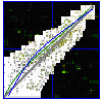
First of all you must select the Supervised-Lowess option by clicking the appropriated check-button



- The "Gene Selection" zone becomes active. Mark the "90" (similarity percentage) as the filter-column and define "1" as matching criterion. Click on "Color Selected Points", to color those genes in purple.
- Use the default values for lowess parameters (f=0,5, nSteps=5 and delta value=0.
- Click on the OK button.



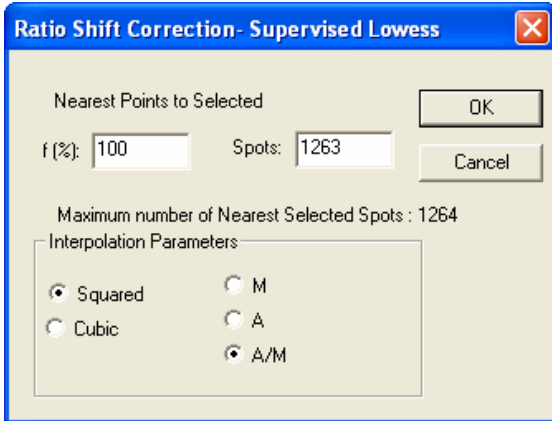
Now selected spots are filtered and we have done Lowess Fit on them, you can see the Fit Lines and selected genes in purple.



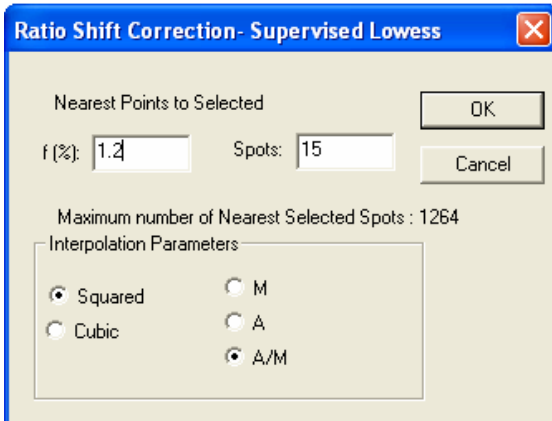
3. - Ratio Shift Correction.



The “Ratio Shift Correction” procedure is able to recognize the way lowess procedure was performed (supervised or traditional lowess). For the first case (as in this example)



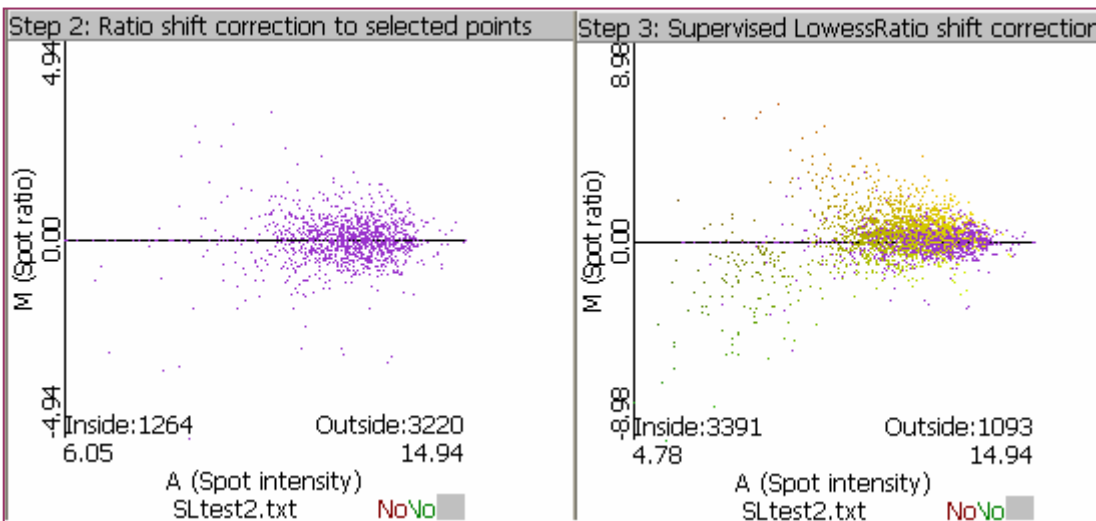
To interpolate ratio shift to ‘not selected genes’ we need the nearest ‘Selected Genes’ to them, so you can select a percentage (f%) of the Selected Spots or the number of “Spots”.

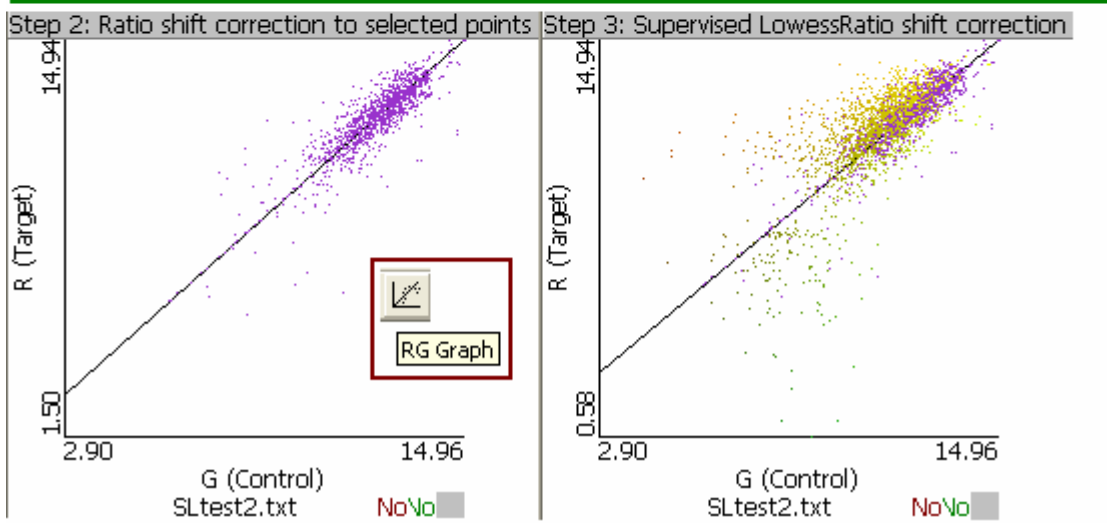
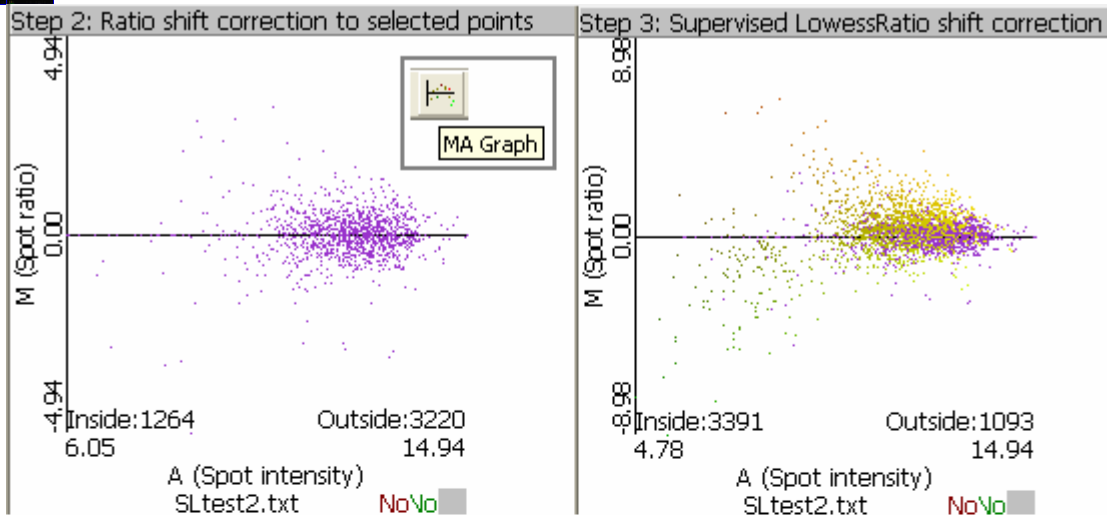
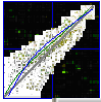


Let’s use 1.2 % of the nearest spots (15) for interpolation (Note: dot is used for decimal values), and the default values for interpolation parameters

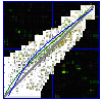
Click OK to proceed.

Two new steps are created; the first of them has the Ratio Shift Correction on selected Spots and the second the Ratio Shift Correction interpolated. Supervised Lowess is done now.





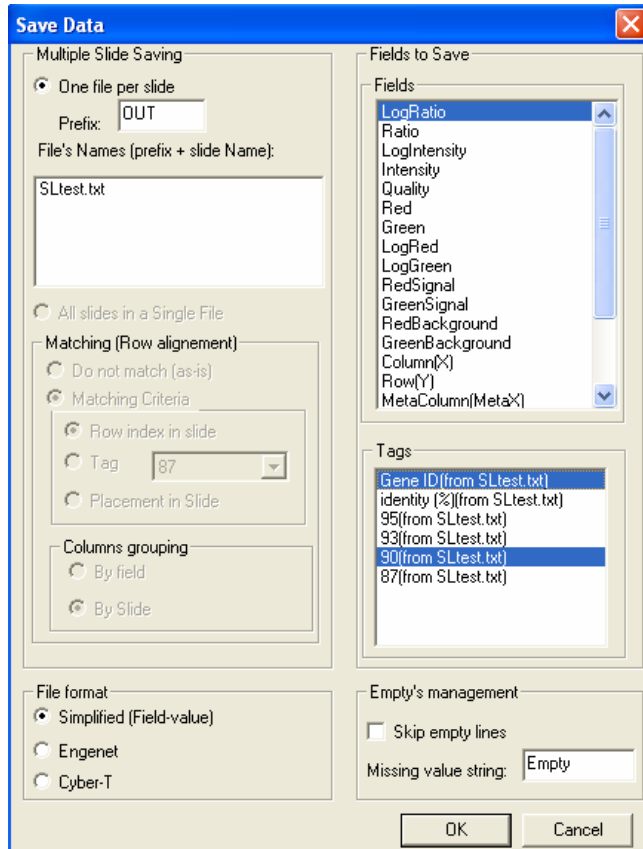
Other views of data are also available (e.g. MA and RG representations)



4. - Saving Data.



Click on “Save Data” button in the toolbar. Select the Fields and Tags you wish to save and press the OK button.



“OUT” is used as prefix for the output datafile (OUT-SLtest.txt)

Additional to the internal index, the LogRatio (computed field) and the Gene_ID and 90% tags will be stored

Important Note: observe in the output file that “dot” is used as decimal separator. Thus, if you are going to use a program like “MS-Excel” to visualize, you must use the advanced option to modify “,” as by-default decimal separator.

SLtest.txt-- Gene ID	SLtest.txt-- 90	Index	0: SLtest.txt-- LogRatio
IL1403_accA	1	0	1,08945
IL1403_accA	1	1	1,52155
IL1403_accB	1	2	-0,761535
IL1403_accB	1	3	-0,470771
IL1403_accC	1	4	-0,284654
IL1403_accC	1	5	-0,581015
IL1403_accD	1	6	-0,514645
IL1403_accD	1	7	-0,873419
IL1403_ackA1	1	8	-0,0575788
IL1403_ackA1	1	9	0,0227554
IL1403_ackA2	1	10	-0,205843
IL1403_ackA2	1	11	-0,0488938
IL1403_acmA	1	12	0,779293
IL1403_acmA	1	13	0,945922
IL1403_acmB	0	14	1,16877
IL1403_acmB	0	15	0,853745

In the table the first 15 rows of the OUT-SLtest.txt file are displayed

More info is available at:

<http://chirimoyo.ac.uma.es/prep/>

Requests and error reporting:

Victoria Martin-Requena vickymr@ac.uma.es
Oswaldo Trelles ots@ac.uma.es