



# Gene Expression Data Analysis

## PreP+07

### Guided Exercise

**PreP+07 : an integrated software platform to improve gene expression data quality**

Victoria Martin-Requena<sup>(1)</sup>, Antonio Muñoz-Mérida<sup>(1)</sup>, Gonzalo Claros<sup>(2)</sup> and Oswaldo Trelles<sup>(1,\*)</sup>

(1) Computer architecture department, University of Malaga, Spain

(2) Molecular biology and biochemistry department; University of Malaga, Spain

(\*) to whom correspondence should be addressed ([ots@ac.uma.es](mailto:ots@ac.uma.es))

[Submitted manuscript]

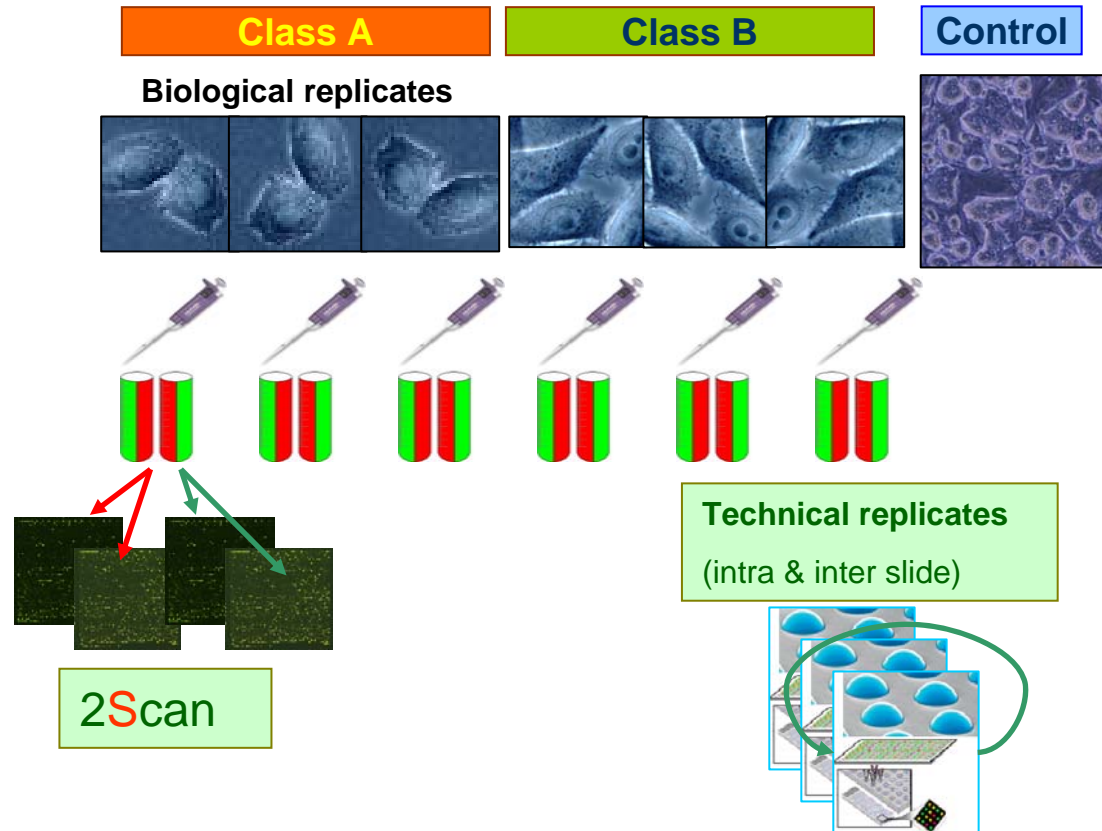
Malaga, July 2008



# The Experiment

## Design

- 2 Experiments (Class A and class B)
- 3 biological replicates per experiment
- 2 technical replicates per each biological replicate
- One technical replicate in dye Swap
- Data acquisition in Double Scan





# Loading Data



Spots	13056	13056	13056	13056
"Name"	Exp1_RB1...	Exp1_RB1...	Exp1_RB1...	Exp1_RB1...
"Block"	GRID	GRID	GRID	GRID
"Column"	X	X	X	X
"Row"	Y	Y	Y	Y
"Name"	"Name"	"Name"	"Name"	"Name"
"ID"	-	-	-	-
"X"	-	-	-	-
"Y"	-	-	-	-
"Dia."	-	-	-	-
"F635 Median"	CONTROL SIG	CONTROL SIG	CONTROL SIG	CONTROL SIG
"F635 Mean"	-	-	-	-
"F635 SD"	-	-	-	-
"F635 CV"	-	-	-	-
"B635"	-	-	-	-
"B635 Median"	CONTROL BKG	CONTROL BKG	CONTROL BKG	CONTROL BKG
"B635 Mean"	-	-	-	-
"B635 SD"	-	-	-	-
"B635 CV"	-	-	-	-
"% > B635..."	-	-	-	-
"% > B635..."	-	-	-	-
"F532 % Sat."	-	-	-	-
"F532 Median"	TARGET SIG	TARGET SIG	TARGET SIG	TARGET SIG
"F532 Mean"	-	-	-	-
"F532 SD"	-	-	-	-
"F532 CV"	-	-	-	-
"B532"	-	-	-	-
"B532 Median"	-	-	-	-
"B532 Mean"	-	-	-	-
"B532 SD"	-	-	-	-
"B532 CV"	-	-	-	-
"% > B532..."	-	-	-	-
"F532 % Sat"	-	-	-	-
"Ratio of Me"	-	-	-	-
"Ratio of Me"	-	-	-	-
"Median of R"	-	-	-	-
"Mean of Ra"	-	-	-	-
"Ratios SD (	-	-	-	-
"Rgn Ratio (	-	-	-	-
"Rgn R2 (63	-	-	-	-
"F Pixels"	-	-	-	-
"B Pixels"	-	-	-	-
"Circularity"	-	-	-	-
"Sum of Me"	-	-	-	-
"Sum of Me"	-	-	-	-
"Log Ratio (...	-	-	-	-
"F635 Media"	-	-	-	-

header		
data		
function x	X	
function y	Y	
function grid	Block	
function control_sig		F635 Median
function control_bkg		B635 Median
function target_sig		F532 Median
function target_bkg		B532 Median
function tag	Name	Flags

- 1- Using a format File
- 2- Mapping COL functionality



# 1. Using a format file

- Get the format file: GenePix.fmt

## - Data load step

load (add) a new format file

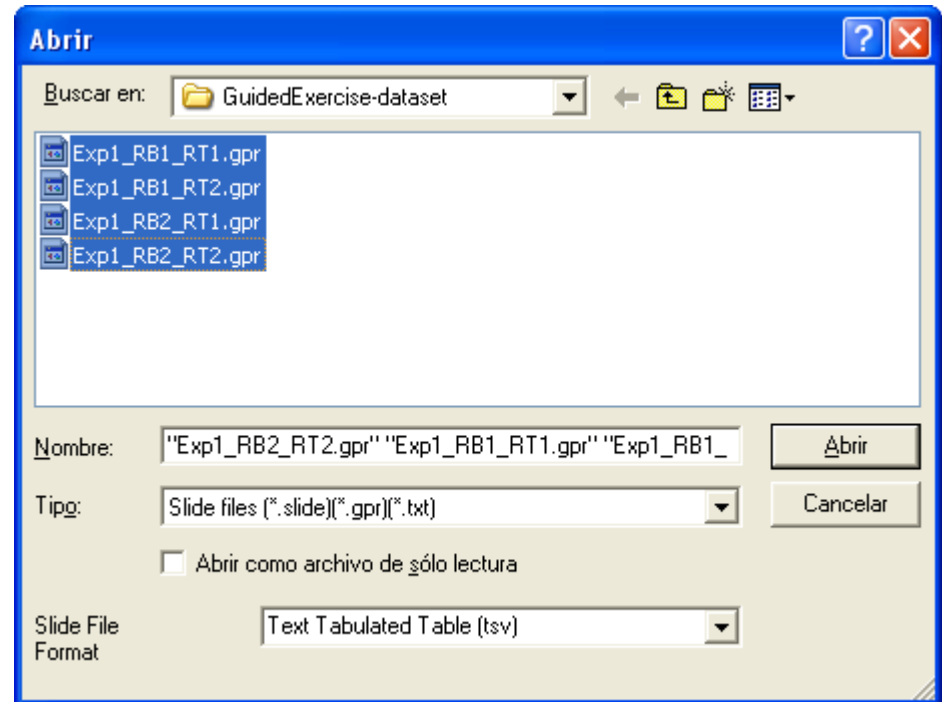


load (add) the selected files



data is loaded with the functionality assigned

“**continue**”  button became active,  
Click and continue this tutorial





# 2. Mapping COL functionality

## - Data load step

load (add) a new slide



Once the minimal functionality has been defined, the "continue" button became active

## -Control/target selection

- Control (red 635)
- Target (green 532)

Spots	13056	13056	13056	13056
Name	Exp1_RB1_...	Exp1_RB1_...	Exp1_RB1_...	Exp1_RB1_...
"Block"	GRID	GRID	GRID	GRID
"Column"	X	X	X	X
"Row"	Y	Y	Y	Y
"Name"	"Name"	"Name"	"Name"	"Name"
"ID"	-	-	-	-
"X"	-	-	-	-
"Y"	-	-	-	-
"Dia."	-	-	-	-
"F635 Median"	CONTROL SIG	CONTROL SIG	CONTROL SIG	CONTROL SIG
"F635 Mean"	-	-	-	-
"F635 SD"	-	-	-	-
"F635 CV"	-	-	-	-
"B635"	-	-	-	-
"B635 Median"	CONTROL BKG	CONTROL BKG	CONTROL BKG	CONTROL BKG
"B635 Mean"	-	-	-	-
"B635 SD"	-	-	-	-
"B635 CV"	-	-	-	-
"% > B635..."	-	-	-	-
"% > B635..."	-	-	-	-
"F635 % Sat."	-	-	-	-
"F532 Median"	TARGET SIG	TARGET SIG	TARGET SIG	TARGET SIG
"F532 Mean"	-	-	-	-
"F532 SD"	-	-	-	-
"F532 CV"	-	-	-	-
"B532"	-	-	-	-
"B532 Median"	Unused	-	-	-
"B532 Mean"	Tag...	-	-	-
"B532 SD"	Description Tag	-	-	-
"B532 CV"	Target	-	-	-
"% > B532..."	Control	-	-	-
"% > B532..."	Target Signal	-	-	-
"F532 % Sat"	Control Signal	-	-	-
"Ratio of Me"	Target Background	-	-	-
"Ratio of Me"	Control Background	-	-	-
"Median of ."	Target	-	-	-
"Mean of Ra"	Control	-	-	-
"Ratios SD (...)"	X (Column)	-	-	-
"Rgn Ratio (...)"	Y (Row)	-	-	-
"Rgn R2 (63..."	Grid	-	-	-
"F Pixels"	Meta X (Grid X)	-	-	-
"B Pixels"	Meta Y (Grid Y)	-	-	-
"Circularity"	Quality	-	-	-
"Sum of Me..."	-	-	-	-
"Sum of Me..."	-	-	-	-
"Log Ratio (...)"	-	-	-	-
"F635 Media"	-	-	-	-

Minimal functionality

Spot coordinates

- Control Signal

- Control background

- Target Signal

- Target background

(Use right button or hot keys)

You can also assign:

-Target (net signal)

-Control (net signal)



# Gene Expression Data Analysis



## using PreP+07

- Choose the block distribution into the chip

The dialog box is titled "PreP+07" and has a close button (X) in the top right corner. It is divided into two main sections: "Spot Range" and "Grid Distribution".

**Spot Range:**

- Base index of X coordinate (Min: 0 Max: 1):
- Base index of Y coordinate (Min: 0 Max: 1):
- Base index of grid number (Min: 0 Max: 1):
- Base index of meta X number (Min: 0 Max: 0):
- Base index of meta Y number (Min: 0 Max: 0):

**Grid Distribution:**

- Row Major  Column Major
- Width (Row Major) or Height (Column Major) (Min: 1 Max: 49):
- Apply structure to all slides
- Include Step History in the Slide Name when Saving
- Buttons: OK, Cancel

**Row Major** needs the number of blocks in each row

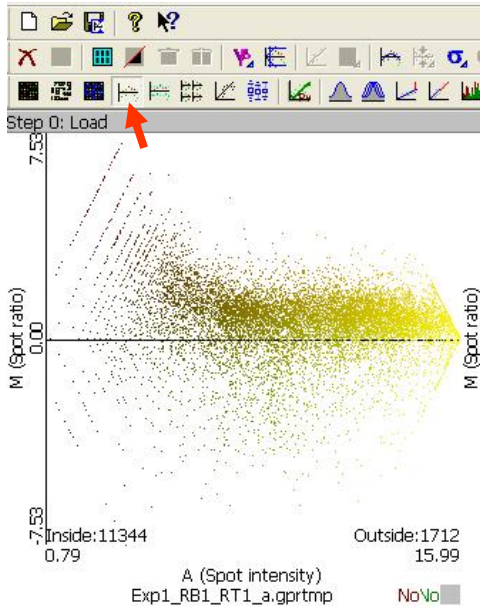
**Column Major** needs the number of blocks in each column



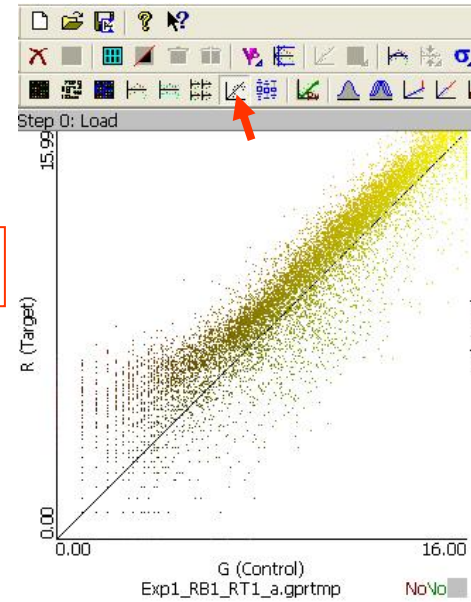
# Data visualization



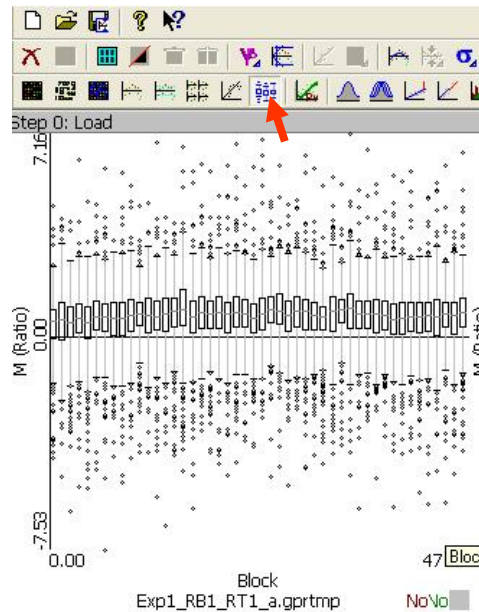
**MA Graph**



**RG Graph**



**Box-plot Graph**



**Use the view toolbar  
To see all the available  
views**



# 2Scan resolution

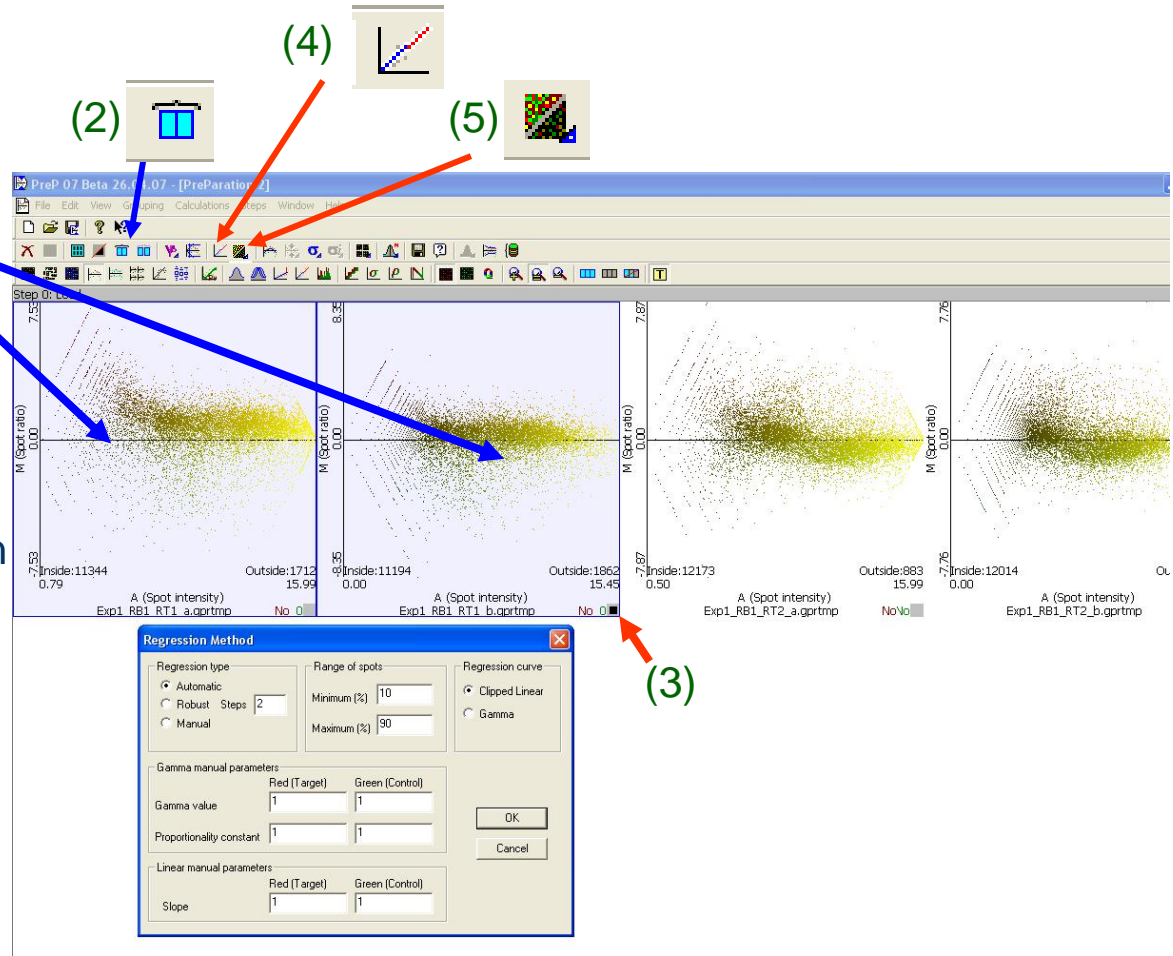
(1) Select slides corresponding to the same chip with different intensities (shift+leftClick)

(2) Group both slides

(3) Mark low-intensity slide

(4) Perform the 2Scan regression

(5) Solve the model



**NOTE:** In the example files there are two 2Scan files, so this Procedure must be done twice, in two steps one by pair.



# Lowess



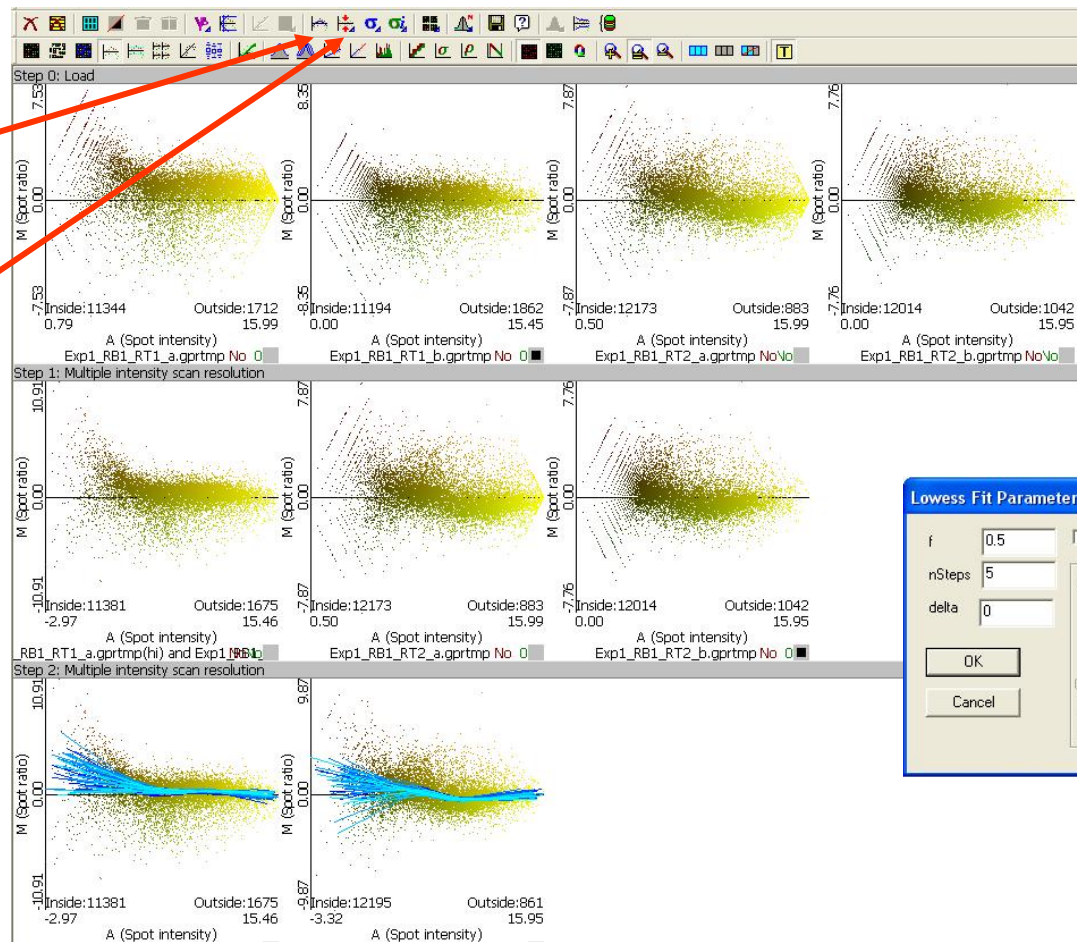
-This procedure is used to adjust the distribution to the 0 value assuming that most of the genes don't change their expression along the experiment, to minimize the difference in detection of the dyes solving signal detection problems

Steps:

(1) Perform fitting



(2) Solve adjust



**Lowess Fit Parameters**

f: 0.5  Supervised Lowess Fit

nSteps: 5

delta: 0

Gene Selection  
Tag: "Name"

Contents:

Colour selected points

OK Cancel



# Technical replicates resolution



- Allows the user to obtain a more faithful data set from different chips
- Dye-swap resolution balance the asymmetry between dyes

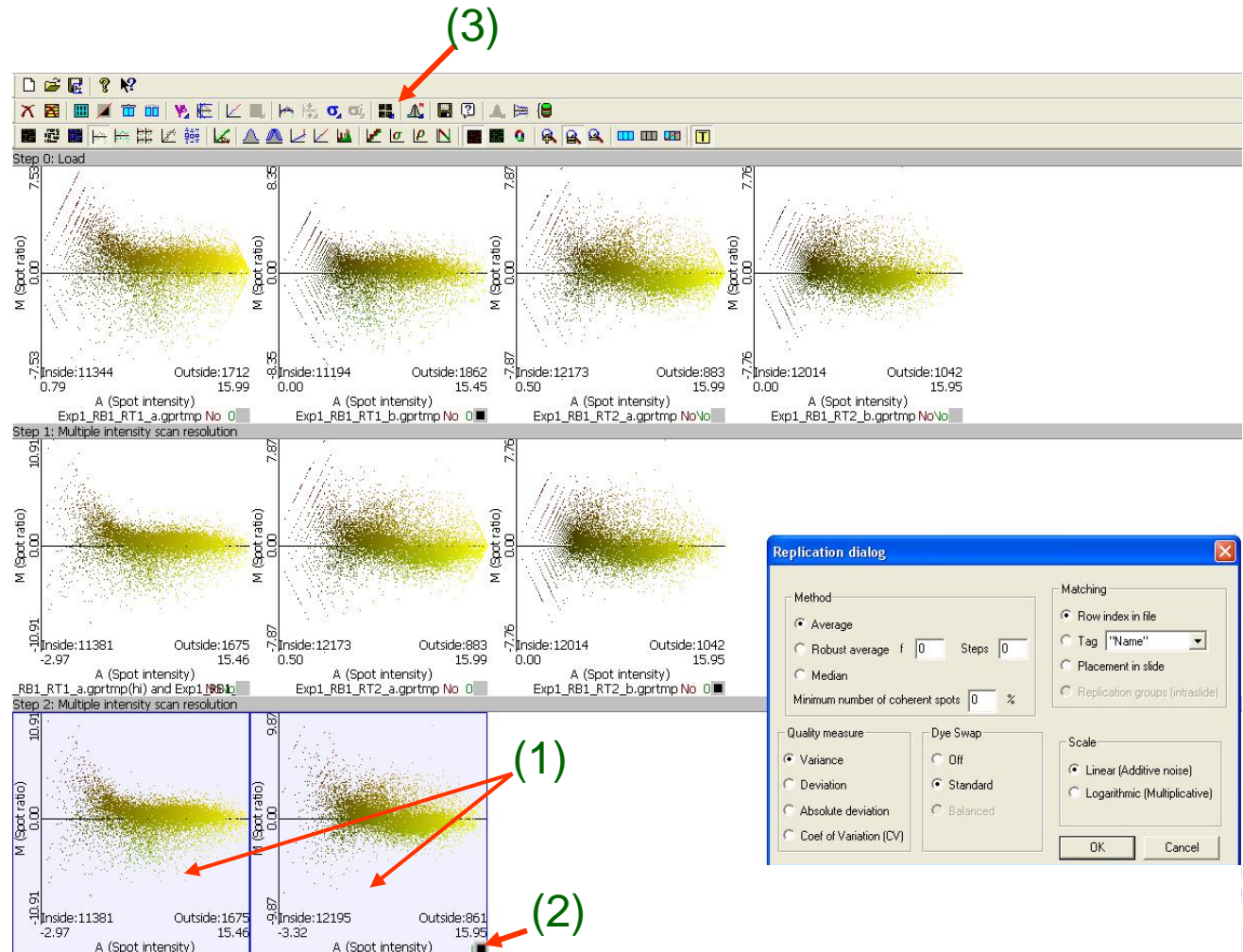
Steps:

(1) Select replicates

(shift+leftClick)

(2) Mark dye swap (if any)

(3) Solve replication



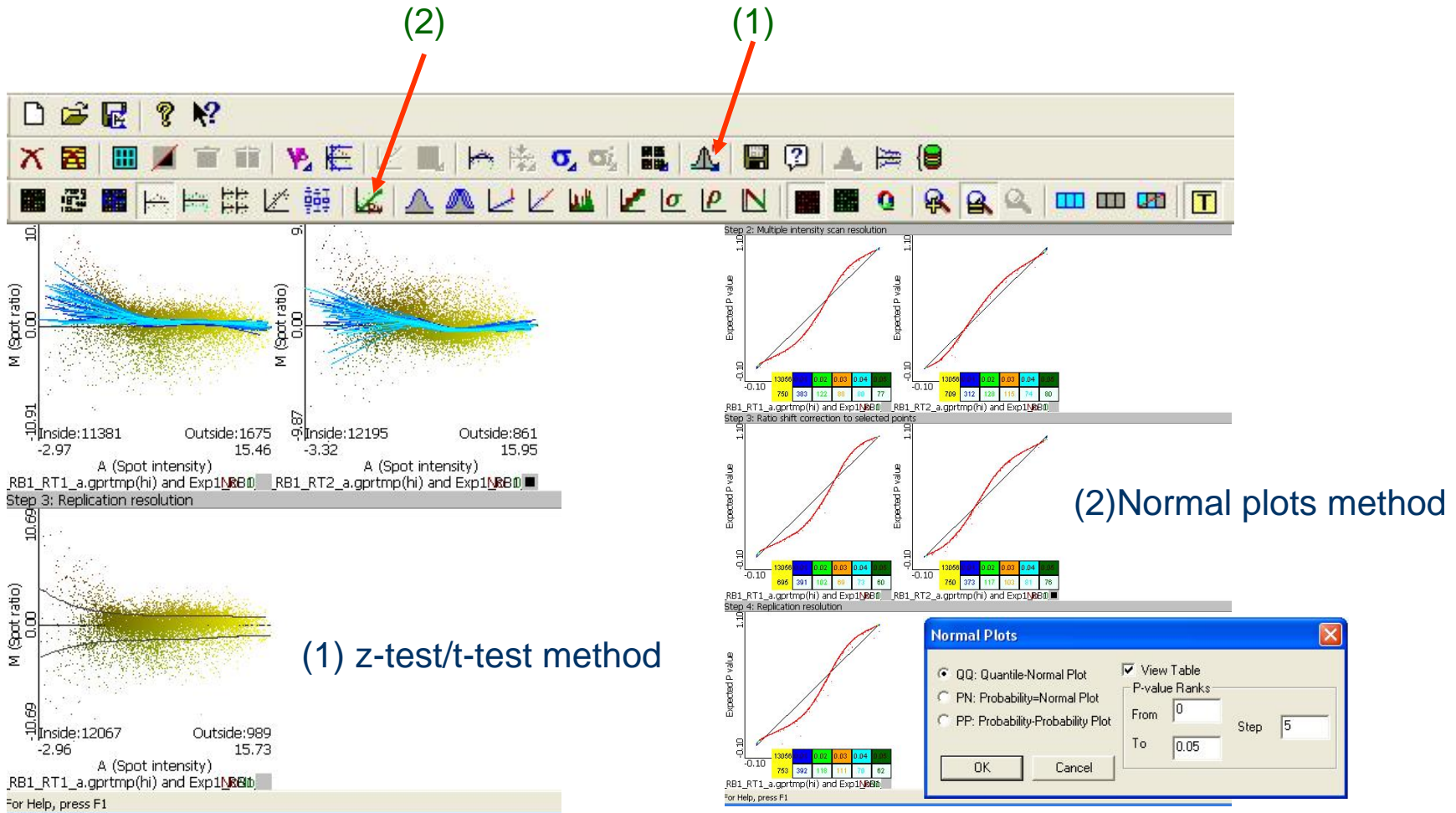
(1)

(2)



# Z-Score and p-Value calculation

Zscore provides a standard value of the intensity change between normalized target and control. P-value is a measure of how significant the z-score is





# Saving the data



Individual/Multiple saving

Matching criteria

Grouping data

Output file format

Data to save

Final results

Id	0: Exp1_RB1_RT1. gprtmp -- LogRatio	0: Exp1_RB1_RT1. gprtmp -- p-Value	0: Exp1_RB1_RT1. gprtmp -- Z-score
0	0,488632	0,475048	0,71429
1	-0,0216231	0,998016	0,00248612
2	0,246281	0,706759	0,376212
3	-0,594269	0,425826	-0,796354
4	0,720347	0,299487	1,03753
5	-0,00992098	0,984992	0,0188105
6	1,13794	0,105217	1,62007
7	-0,189529	0,816738	-0,231743
8	0,490116	0,473768	0,716362
9	Empty	Empty	Empty
10	-1,26775	0,0825893	-1,73586
11	-3,39419	2,57E-06	-4,70224
12	0,362874	0,589984	0,538859
13	0,646196	0,350256	0,934092

**PreP+07** allows saving intermediate results, as well as storing the whole state of the analysis and resume again from this state. A "project" file has the ".prep" extension