

Babelomics

Microarray Data Analysis Differential Gene Expression

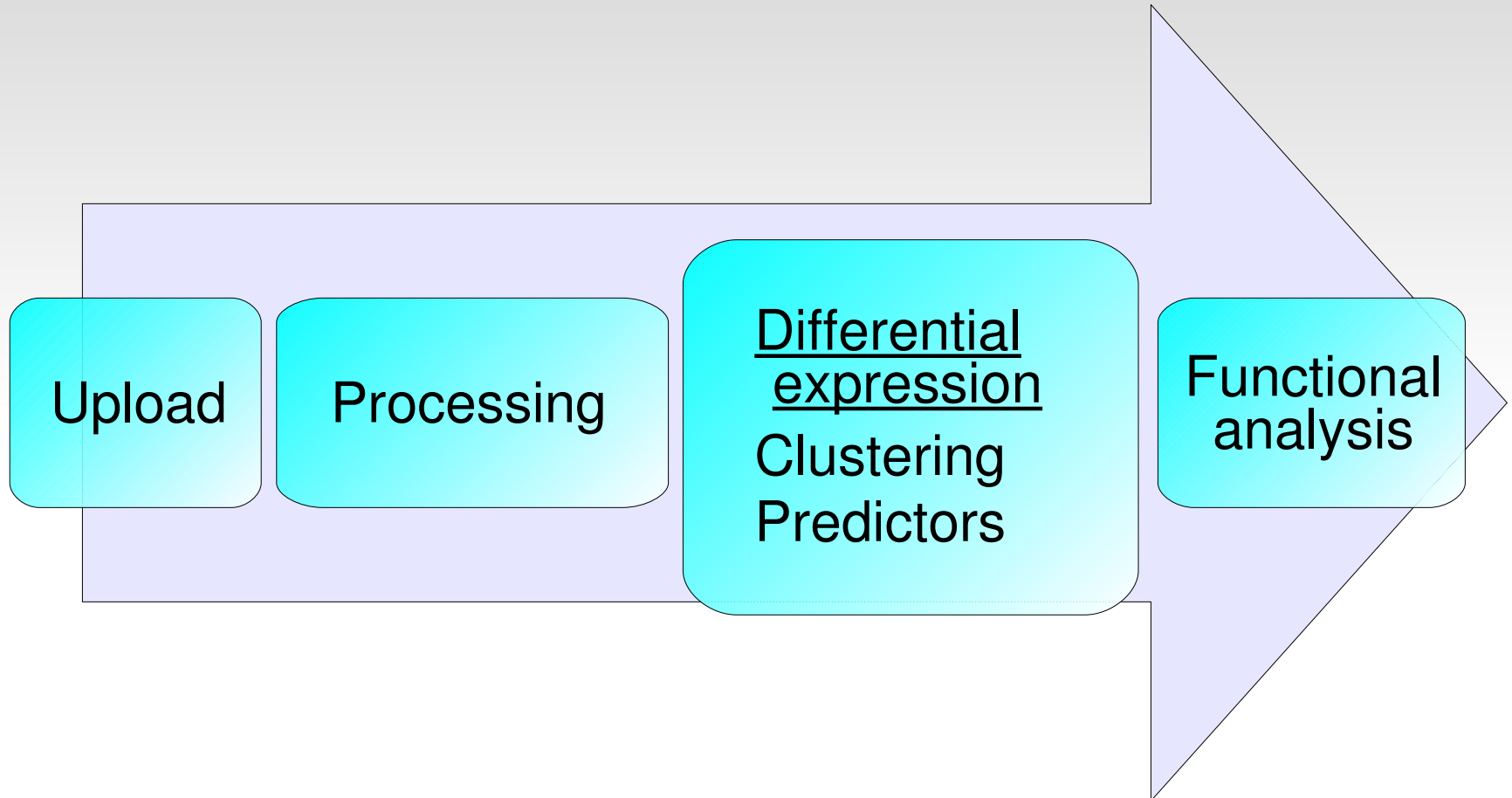
Valencia, June 2010

Francisco García, fgarcia@cipf.es

Bioinformatics and Genomics Department
Centro de Investigación Principe Felipe (CIPF)
(Valencia, Spain)



Data analysis workflow



Differential Expression

Class comparison

Correlation

Survival

Time / dosage
series

Input

1. After normalization, we will add information relative to experimental design

BABELOMICS 4
gene expression and functional profiling analysis suite

Upload data | Preprocessing | Expression | Genomic | Functional analysis | Utilities

testing@cipf.es working on project default 475.71 Mb of 1.00 Gb (46.46%)

Welcome to the new Babelomics 4, you can still use Babelomics 3 at: <http://babe>

Preprocessing data

- Edit
 - Allow modify your uploaded data.
- Normalize
 - Expression
 - Affymetrix
 - One-channel

Affymetrix methods are those implemented in the `affy` package

Edit data

Select your data

browse multiclassses
Data tags compatibles: [\[datamatrix\]](#)

+ Create new variable ↻ Undo changes

Variable name	Type	Values	Description
conditions	CATEGORICAL	A B C	
classes	CATEGORICAL	con mut	

-classes:

Samples names	Values
Sample1	control
Sample2	
Sample3	
Sample4	

2. Assigning values of variables to each array

Input

Array names

Arrays

Tab separated file

#NAMES	col1	col2	col3	col4	col5	col6	col7
YGR138C	-1.23	-0.81	1.79	0.78	-0.42	-0.69	0.58
YPR156C	-1.76	-0.94	1.16	0.36	0.41	-0.35	1.12
YOR230W	-2.19	0.13	0.65	-0.51	0.52	1.04	0.36
YAL018C	-1.22	-0.98	0.79	-0.76	-0.29	1.54	0.93
YBR287W	-1.47	-0.83	0.85	0.07	-0.81	1.53	0.65
YCL075W	-1.04	-1.11	0.87	-0.14	-0.80	1.74	0.48
YDR055w	-1.57	-1.17	1.29	0.23	-0.20	1.17	0.26
YOR358W	-1.53	-1.25	0.59	-0.30	0.32	1.41	0.77
YBR006W	-1.76	-0.72	0.13	-0.01	-0.23	1.30	1.28
YBR241C	-1.39	-0.42	-0.08	-0.29	-0.65	1.85	0.98
YCR021c	-1.52	-0.99	0.26	0.04	-0.42	1.43	1.19
YCR061W	-1.57	-0.39	0.33	-0.54	-0.51	1.59	1.09
YDL024c	-1.27	-1.14	0.57	-0.30	-0.47	1.46	1.14
YDR298C	-1.49	-0.87	0.41	-0.47	-0.25	1.38	1.29
YER141w	-1.69	-0.60	0.00	0.41	-0.62	1.45	1.05

.....

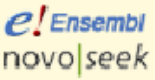
Results

#NAMEs	statistic	p-value	adj. p-value
201322_at	5.708008670390058	9.706907759422906E-6	0.09337500977560666
217826_s_at	5.36267554627671	3.939115118155811E-5	0.09337500977560666
200730_s_at	5.103924738211194	6.864136450202152E-5	0.09337500977560666
202318_s_at	5.017681592834031	1.020788250419645E-4	0.09337500977560666
201447_at	4.925445907931246	7.932015741900609E-5	0.09337500977560666
218363_at	4.910910447680612	8.263799736261134E-5	0.09337500977560666
217916_s_at	4.880574980202908	7.505311094157185E-5	0.09337500977560666
204240_s_at	4.840959025619068	8.200429350286117E-5	0.09337500977560666

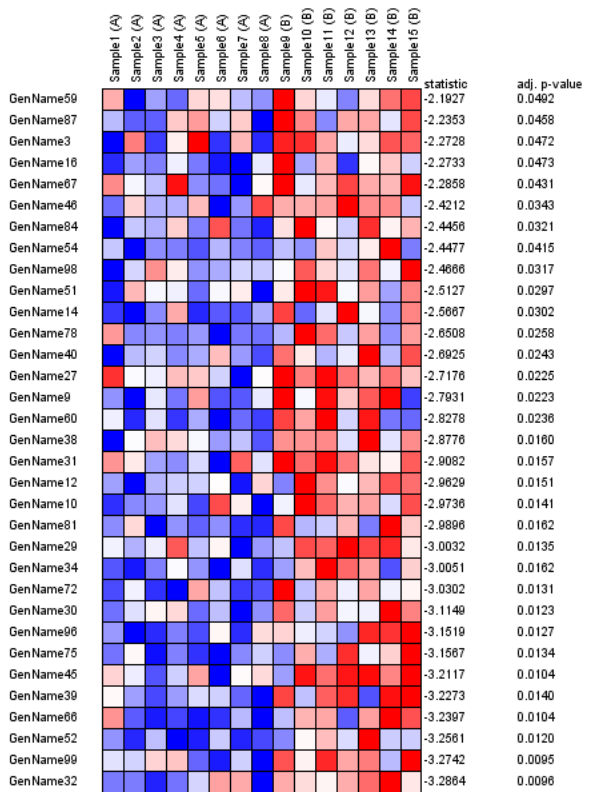
name	statistic	p-value	adj. p-value
200067_x_at	5.5382	0.0000049746	0.00024376
200052_s_at	5.2111	0.00001452	0.00047431
200054_at	5.1028	0.000042635	0.0010445
200009_at	4.2093	0.00019599	0.0027557
200017_at	4.0805	0.00022496	0.0027557
1053_at	3.9461	0.00060822	0.0059605
200013_at	3.767	0.00070427	0.0062744
200071_at	3.518	0.0014872	0.012146
200076_s_at	3.1376	0.0039127	0.024703
177_at	3.0053	0.0061375	0.030074

Search the term 200067_x_at

General databases





Other info



Results

▼ Continue processing

- Send ranked list to FatiScan tool :
 **Send to FatiScan tool...**
- Send significant results to FatiGO tool :
 **Send to FatiGO tool...**

▼ Other actions

- **Open input form**

Different experimental designs

Class comparison

Select your data

no data selected.
Or go to Upload Data form: [Upload \[datamatrix\]](#)

1. data

Select the class to analyse

Class name:

Specify the class values to test:

2. classes

Select test

One-class (for log ratios)

Limma

Two-classes

T-test
 Limma
 Fold-change

Multi-classes

Anova
 Limma

3. methods

4. adj. p-value

Select multiple-test correction

Benjamini and Hochberg (BH), FDR
 Benjamini and Yekutieli (BY)
 Hochberg
 Holm
 Bonferroni

Select adjusted p-value

Adj. p-value (0.0-1.0)

Select fold-change value

Fold-change value

Job

Job name:

Job description

5. job name

Two-classes form

Select your data

twoclasses.txt (demo data)
Data tags compatibles: [\[datamatrix\]](#)

Select the class to analyse

Class name:

Specify the class values to test:

Select test

One-class

Limma

Two-classes

T-test
 Limma
 Fold-change

Methods:

Limma, t-test:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Fold-change:

$$\log_2 (\tilde{y}_1 / \tilde{y}_2)$$

$$\tilde{y}_1 - \tilde{y}_2$$

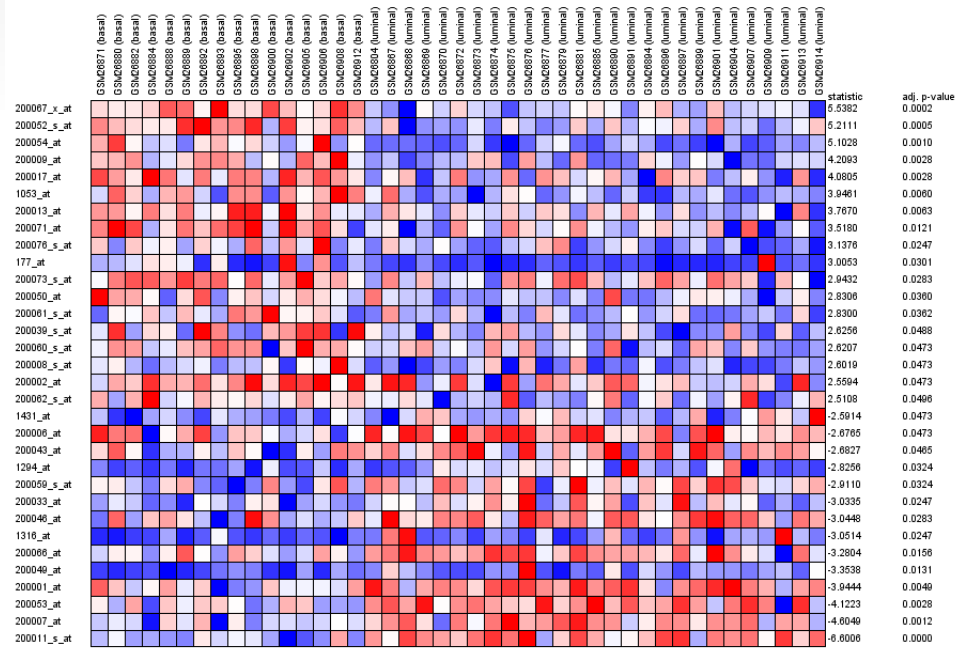
Two-classes results

Limma, t-test

Fold-change



#NAMES	log	diff
AFFX-BioB-5_at	0.023216278176490163	0.07865499999999948
AFFX-BioB-M_at	-0.01743916189488063	-0.06712000000000007
AFFX-BioB-3_at	0.011365357483625202	0.03912499999999941
AFFX-BioC-5_at	-0.014803025848131719	-0.06477500000000092
AFFX-BioC-3_at	0.012163222943743255	0.05631999999999948
AFFX-BioDn-5_at	-0.03199050255238944	-0.16258999999999998
AFFX-BioDn-3_at	-0.04557267653715912	-0.26996499999999991
AFFX-CreX-5_at	0.005800532734088386	0.042909999999999116
AFFX-CreX-3_at	0.005947534544836981	0.045390000000000115

#NAMES	statistic	p-value	adj. p-value
200067_x_at	5.538204986516149	4.974604961551534E-6	2.4375564311602516E-4
200052_s_at	5.2110706353314535	1.4519552681244469E-5	4.743053875873193E-4
200054_at	5.102797363044661	4.263454480923057E-5	0.001044546347826149
200009_at	4.209329258084638	1.9598818661190837E-4	0.002755732412247896
200017_at	4.0805286865632855	2.2495774793860376E-4	0.002755732412247896
1053_at	3.9460740578057503	6.082189146003286E-4	0.005960545363083221
200013_at	3.767033234598989	7.042746674112254E-4	0.0062744470369936371
200071_at	3.5180398564848283	0.0014872364080140634	0.012145763998781516
200076_s_at	3.137574787036864	0.003912733450155826	0.02470303941792398
177_at	3.0053355520231624	0.0061374669029413305	0.030073587824412523
200073_s_at	2.9431619702299616	0.005421086687530431	0.028266292930286572




Two-classes results

Continue processing

- Send top list vs bottom list to FatiGO tool :
 [Send to FatiGO tool...](#)
- Send top list vs genome to FatiGO tool :
 [Send to FatiGO tool...](#)

Significant results

- Number of significant results (adj. p-value = 0.05) : **32**
- Significant values dataset (adj. p-value = 0.05) : **t_significant_dataset.txt**
 [Send to Clustering tool...](#)

Gene set analysis

- [Online examples \(test the form with example data\)](#)

Select your ranked list

[browse server](#) t_ranked_list.txt (from job trained_ttest)

Or go to Upload Data form: [Upload \[idlist:ranked\]](#)

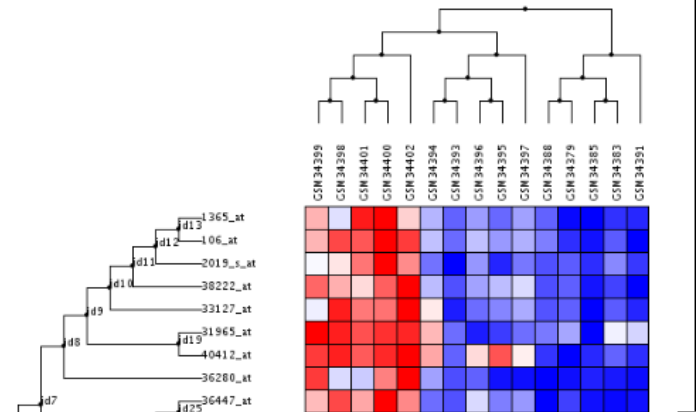
Options

Logistic model

Fatican

Fisher exact test

Remove duplicates?



One-class

Select your data

rma.summary
Data tags compatibles: [\[datamatrix\]](#)

Select the class to analyse

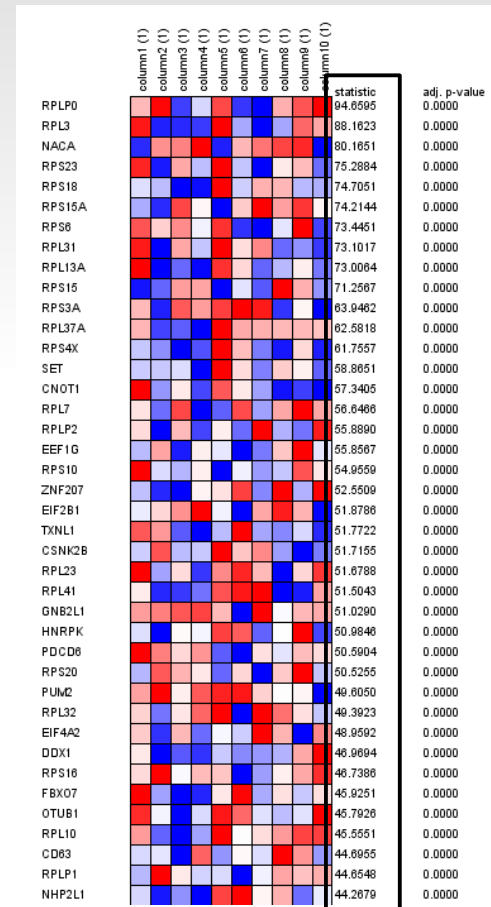
Class name:
Specify the class values to test:

Select test

One-class

Limma

$H_0: \mu = 0$
 $H_a: \mu \neq 0$



Genes ordered by statistic

Multi-classes

Select your data

correlation_data

Data tags compatibles: [\[datamatrix\]](#)

Select the class to analyse

Class name:

Specify the class values to test:

Select test

One-class

Limma

Two-classes

T-test

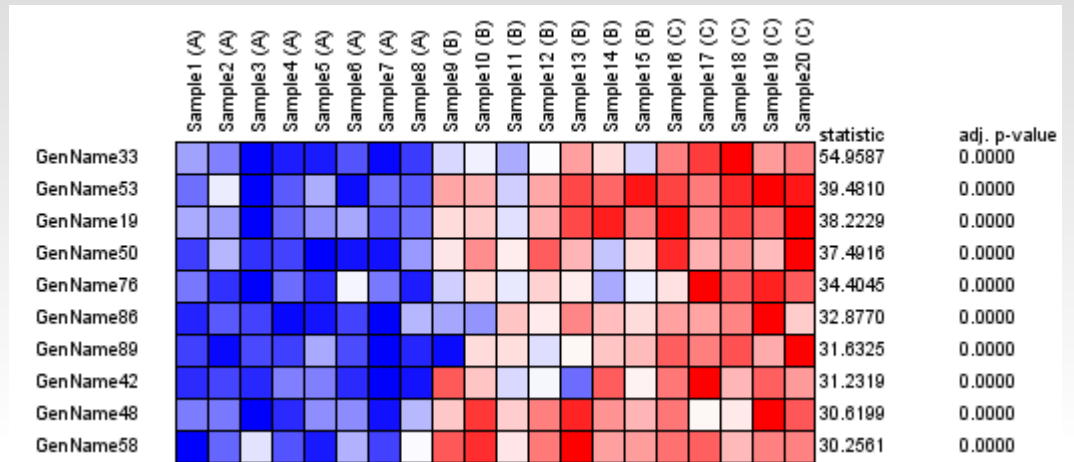
Limma

Fold-change

Multi-classes

Anova

Limma



#NAMES	statistic	p-value	adj. p-value
GenName33	54.95873793763448	3.7921342843638683E-8	3.7921342843638683E-6
GenName53	39.48102621845048	4.0829103398198896E-7	1.4629620390116393E-5
GenName19	38.22289111283603	5.117505867957561E-7	1.4629620390116393E-5
GenName50	37.491605144160076	5.851848156046557E-7	1.4629620390116393E-5
GenName76	34.40451478863487	1.0562954578796635E-6	2.112590915759327E-5
GenName86	32.87695372059555	1.4375186997428457E-6	2.39586449957141E-5
GenName89	31.63250021159884	1.8635584850823506E-6	2.507104595927956E-5
GenName42	31.231879988875637	2.0294501638096207E-6	2.507104595927956E-5
GenName48	30.619945995150157	2.3156652191325122E-6	2.507104595927956E-5
GenName58	30.256077313275878	2.507104595927956E-6	2.507104595927956E-5

$H_0: \mu_1 = \mu_2 = \dots = \mu_n$
 $H_a: \text{not } H_0$

Gene expression related to a continuous variable, form

Select your data

no data selected.

Or go to Upload Data form: [Upload \[datamatrix\]](#)

Select the class to analyse

Class name:

Select test

Pearson's correlation

Spearman correlation

Regression

Select multiple-test correction

Benjamini and Hochberg (BH), FDR

Benjamini and Yekutieli (BY)

Hochberg

Holm

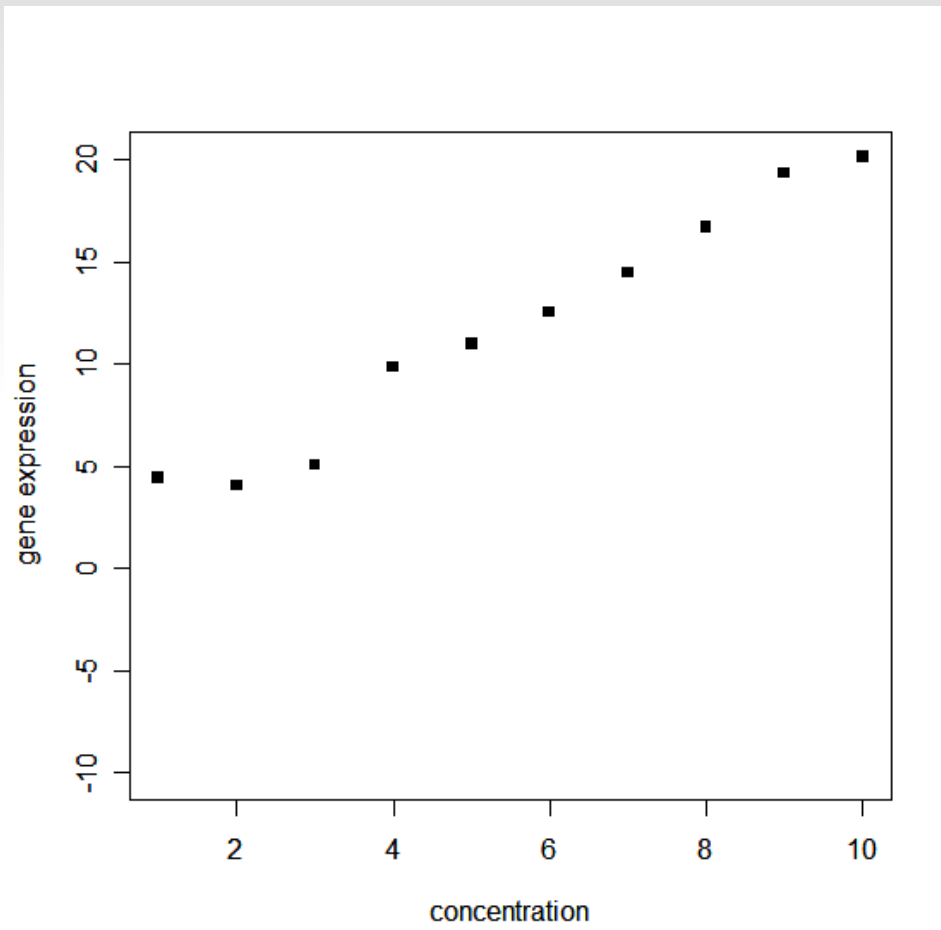
Bonferroni

methods

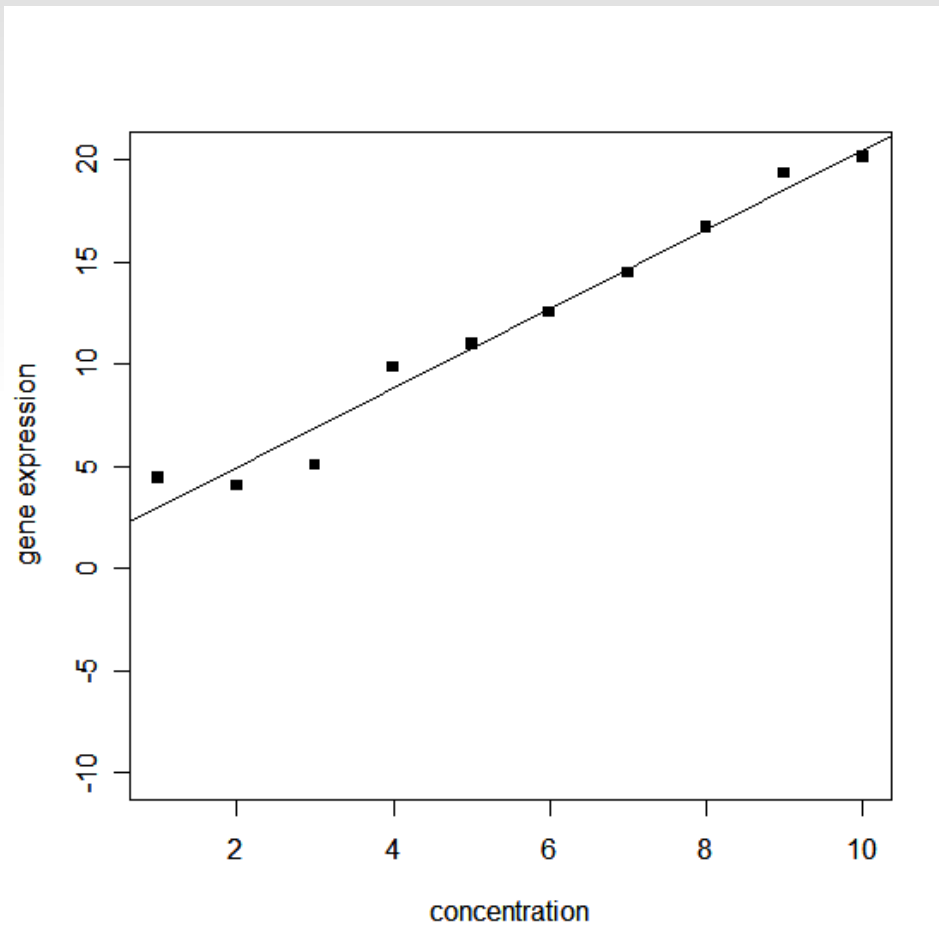
$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Regression

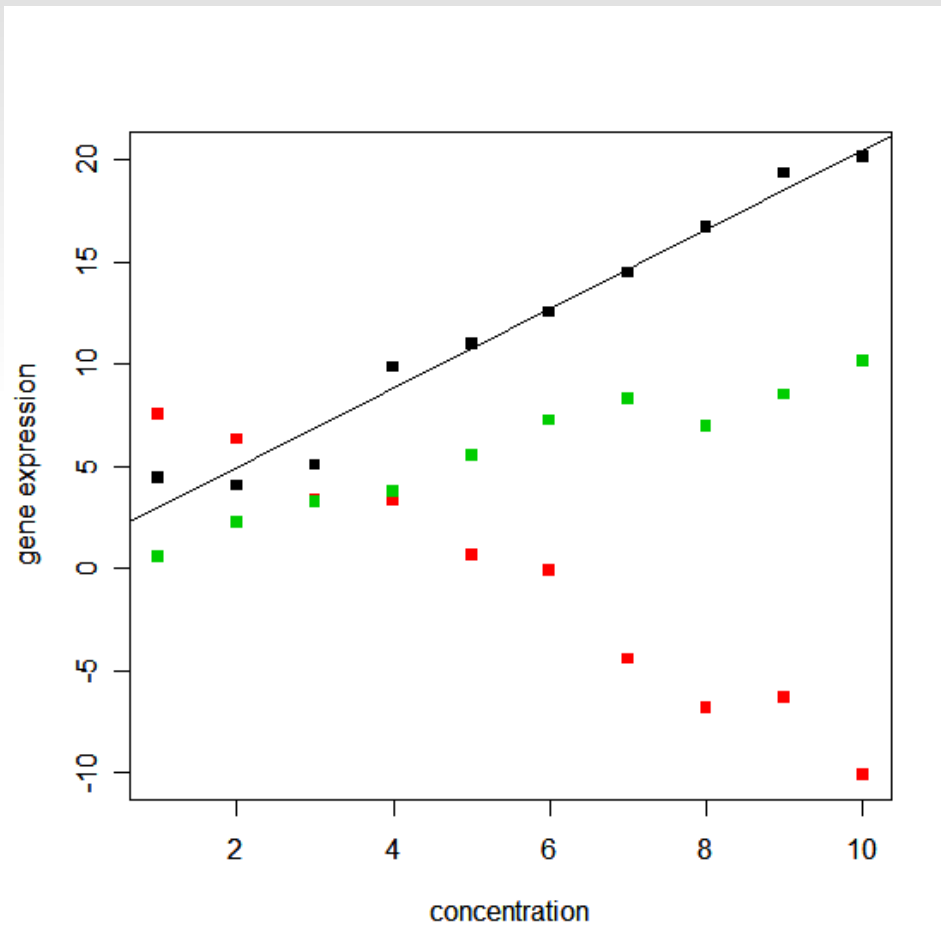


Regression



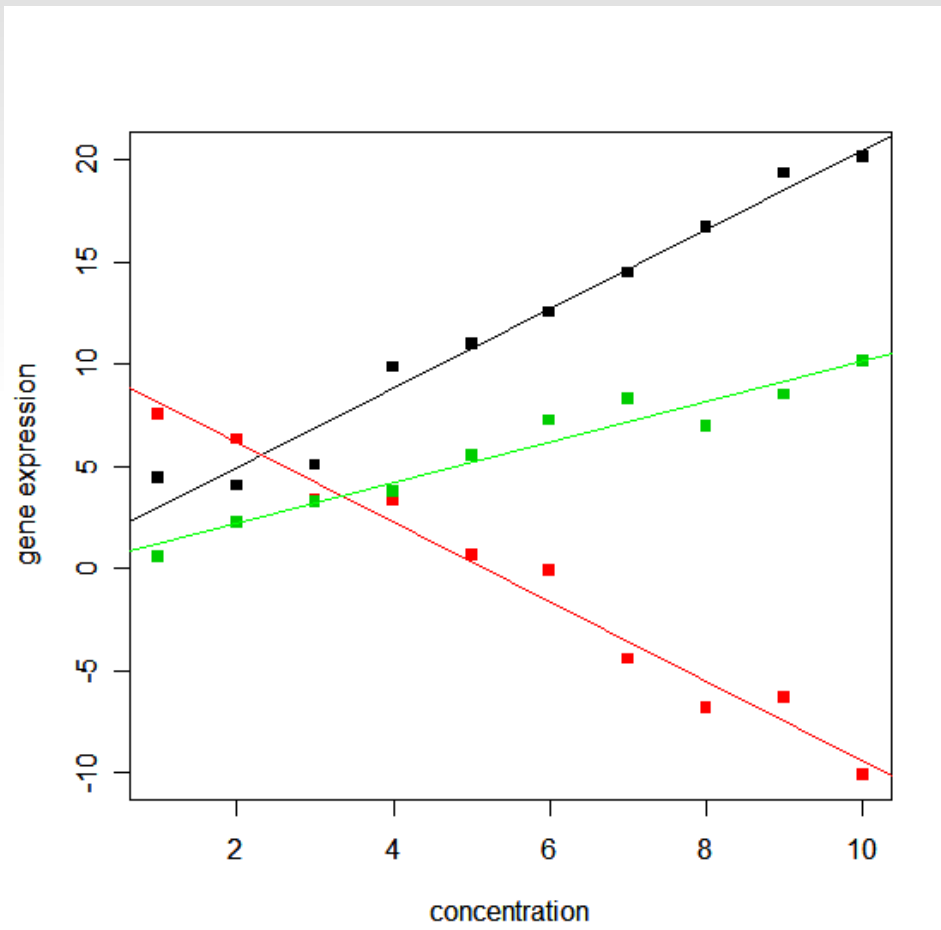
gene1 slope

Regression



gene1 slope

Regression



gene1 slope
gene2 slope
gene3 slope
...

Survival *form*

Select your data

no data selected.
Or go to Upload Data form: [Upload \[datamatrix\]](#)

Select time and series variables

Time variable name:

Censored variable name:

Select test

Cox

Select multiple-test correction

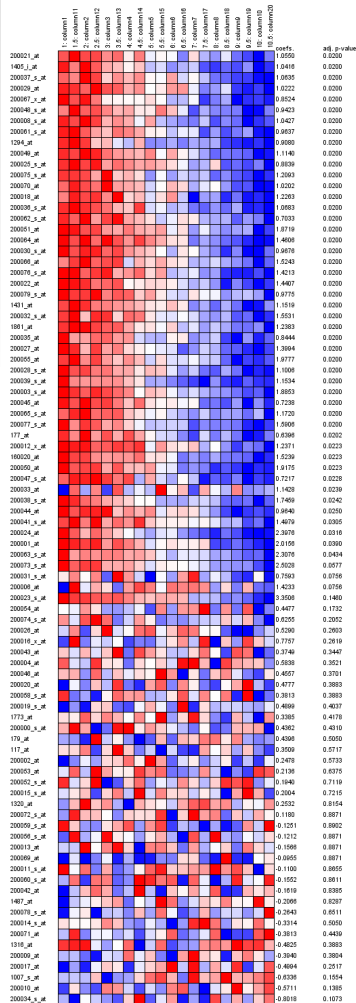
Benjamini and Hochberg (BH), FDR
 Benjamini and Yekutieli (BY)
 Hochberg
 Holm
 Bonferroni

Select adjusted p-value

Adj. p-value (0.0-1.0)

- ← 1. Expression
- ← 2. Survival times & censoring variable
- ← 3. Cox proportional hazards regression model
- ← 4. Multiple-test correction
- ← 5. adjusted p-value

Survival results



← Arrays ranked according to the survival time

↙ Genes ranked by their relationship with survival time

- Cox model coefficients
- Estimate for the statistics
- p-values

Time course analysis / Dose analysis, form

Select your data

no data selected.
Or go to Upload Data form: [Upload \[datamatrix\]](#)

Series variables

Name of Continuous Variable:

Name of Variable defining Series:

Options

Polynomial degree

Significance level for gene selection (0.0-1.0)

Multiple testing adjustment

Benjamini and Hochberg (BH), FDR
 Benjamini and Yekutieli (BY)
 Hochberg
 Holm
 Bonferroni
 Hommel

Significance Level for model variable(0.0-0.9)

Cluster method Hierarchical clustering
 K-Means

Number of clusters (k-value)

← Expression data

← Time variable and series classification

← Complexity of model

← Significance level

← Multiple testing

← Clustering

Time course analysis / Dose analysis, *example*

Arabidopsis.

4 series: **control and 3 treatments**
(cold, salt, heat).

3 time points.

3 replicates.

What do we want?

Time course analysis / Dose analysis, results

Control

#genes	cluster
STMEQ29	1
STMID05	2
STMGB57	1
STMEY09	3
STMHY68	4
STMGI03	5
STMCU02	1
STMGB35	6
STMDI90	1
STMJI76	7
STMJO83	8
STMCS44	1
STMIA31	3
STMJF53	1
STMIQ37	4
STMJC14	9
STMCM86	2
STMGQ83	3
STMCK87	1
STMCU87	1
STMHN19	9
STMED61	3
STMIC27	5
STMCH79	6
STMDU84	8
STMIO93	9
STMEG09	1
STMIX47	9
STMIP63	6
STMEV77	4

Cold vs Control

#genes	cluster
STMJH42	1
STMDE66	2
STMHZ45	1
STMGL58	3
STMIF71	1
STMEG62	4
STMFB37	5
STMEQ29	1
STMDW06	6
STMEL85	5
STMEG74	1
STMCO26	7
STMHX33	3
STMDV94	3
STMID12	1
STMCV66	2
STMGH56	2
STMEJ16	5
STMCD46	1
STMIT95	1

Salt vs Control

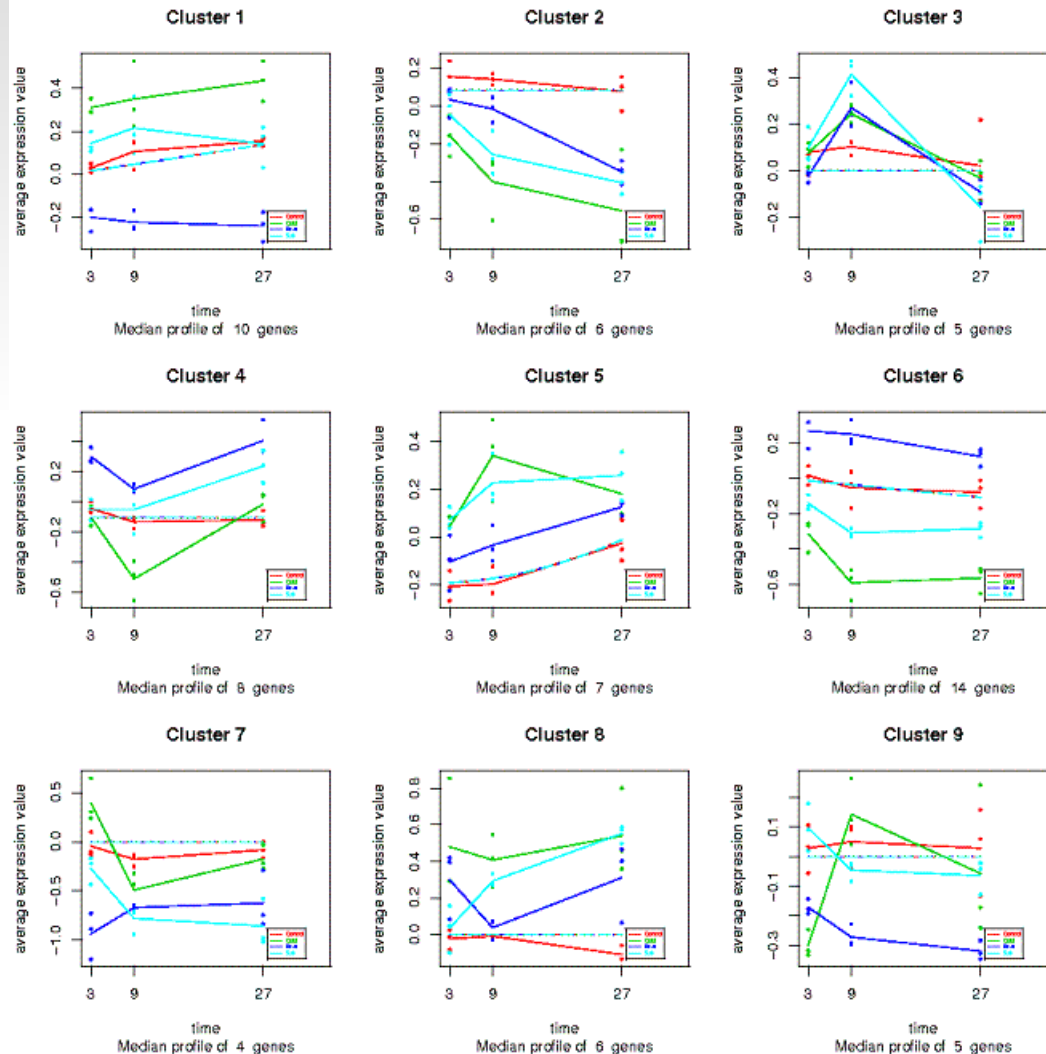
#genes	cluster
STMJH42	1
STMDE66	2
STMHZ45	1
STMGL58	3
STMIF71	1
STMEG62	4
STMEQ29	1
STMDW06	5
STMEG74	1
STMCO26	6
STMID05	5
STMID12	1
STMDH27	7
STMEJ16	1
STMCD46	1
STMIT95	3
STMHJ39	1
STMGB57	3
STMIT31	1
STMEZ42	1
STMIM44	5
STMHN16	5
STMEY09	2
STMCE01	1
STMIY82	2
STMEU24	8
STMHH10	1
STMGQ20	5
STMGI03	9
STMCI10	5
STMHV34	5
STMHY91	2
STMJN05	1
STMEF65	1

Heat vs Control

#genes	cluster
STMJH42	1
STMDE66	2
STMHZ45	1
STMGL58	3
STMIF71	1
STMEG62	4
STMFB37	5
STMEQ29	1
STMDW06	6
STMEL85	5
STMEG74	1
STMCO26	7
STMHX33	3
STMDV94	3
STMID12	1
STMCV66	2
STMGH56	2
STMEJ16	5
STMCD46	1
STMIT95	1
STMJE19	2
STMHJ39	1
STMGU26	5

Time course analysis / Dose analysis, results

#genes	cluster
STMFB γ	1
STMDW \cdot 1	2
STMEL λ 0	1
STMCO γ 1	3
STMHX γ 3	1
STMHQ γ 1	1
STMCV γ 1	4
STMDH γ 7	4
STMIT γ 0	0
STMJE γ 1	6
STMHU γ 9	0
STMGU γ 1	1
STMEZ ϵ 7	0
STMCV γ 1	4
STMIM ϵ 4	6
STMEM γ 9	0
STMHY γ 1	4
STMGH λ 0	1
STMGQ γ 0	2
STMGI \cdot 2	7
STMJN \cdot 0	8
STMDB γ 0	7
STMFB γ 1	4
STMDJ \cdot 2	9
STMGB γ 0	0
STMDU γ 9	9
STMDE σ 9	4
STMCF \cdot 1	6
STMHK ϵ 4	7
STMJI γ 1	3
STMEM λ 0	6
STMIA γ 9	6
STMIO γ 0	8



significant genes

Let's practise!

<http://babelomics.bioinfo.cipf.es/>

Go to the tutorial

