Omics-based biomarkers detection

Francisco García García
Bioinformatics and Biostatistics Unit, CIPF
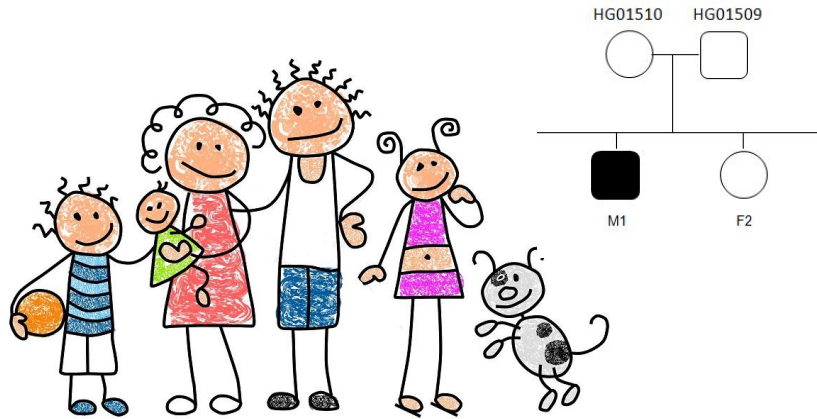
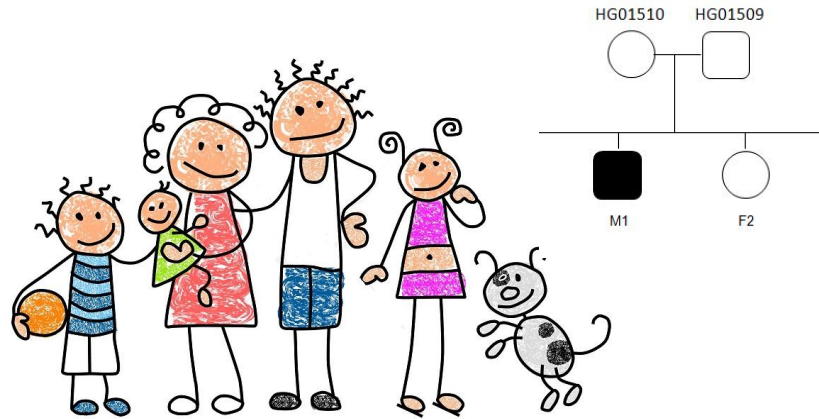24 Oct 2018

Unidad de Bioinformática y Bioestadística
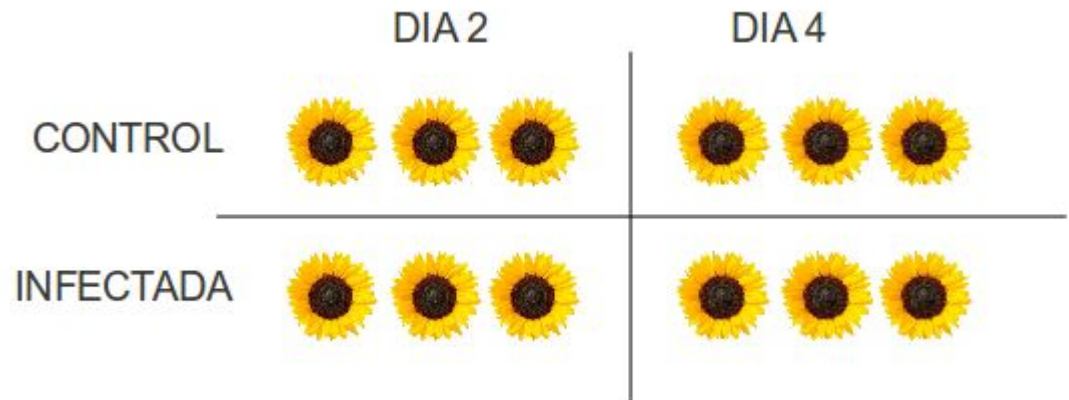
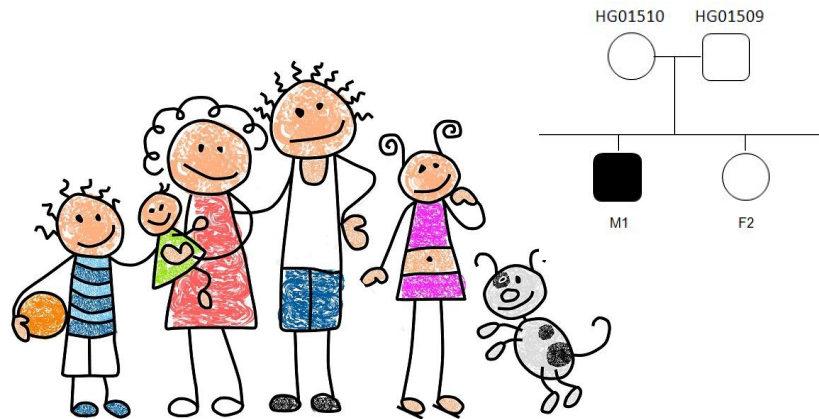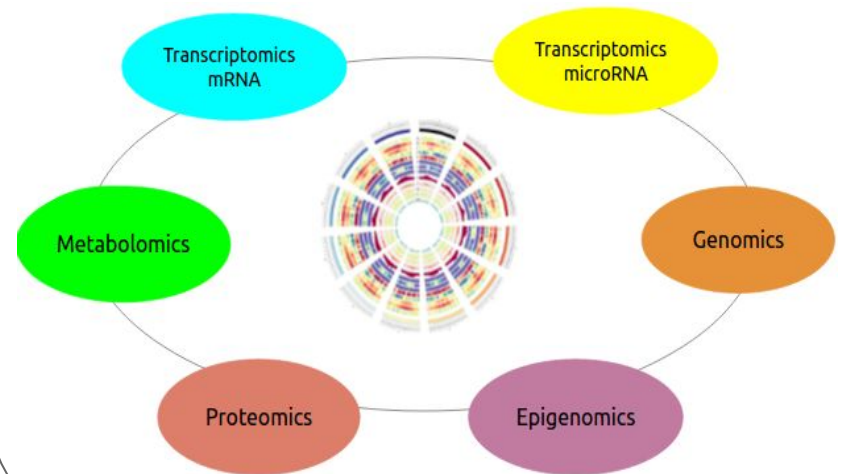PRINCIPE FELIPE
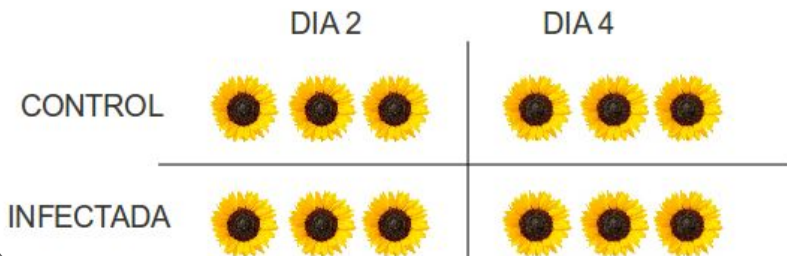CENTRO DE INVESTIGACION

WODA
WEB-BASED OMICS DATA ANALYSIS

Detecting biomarkers

Detecting biomarkers

HG01510 HG01509

M1   F2

Detecting biomarkers

HG01510   HG01509

M1   F2

DIA 2   DIA 4

CONTROL

INFECTADA

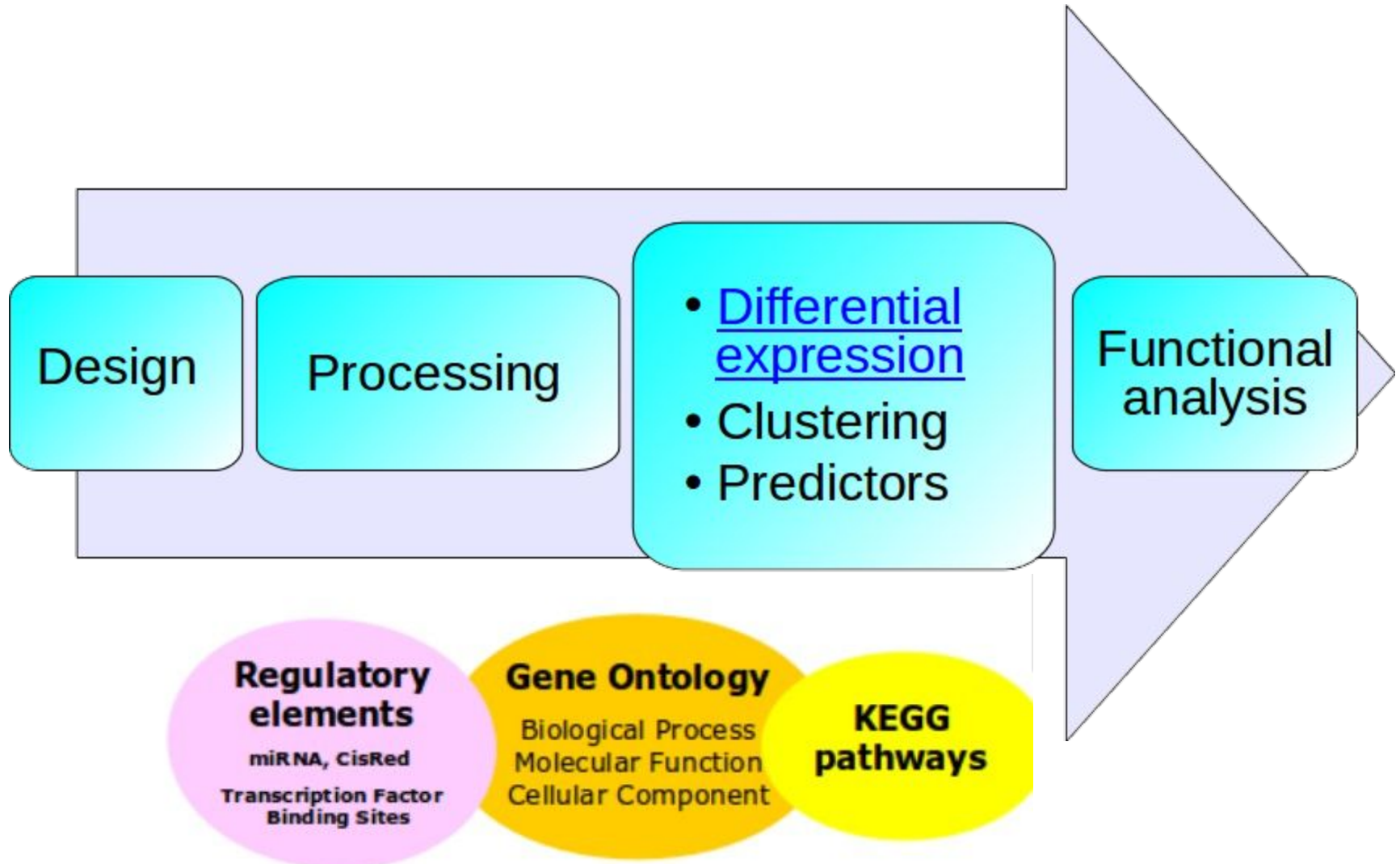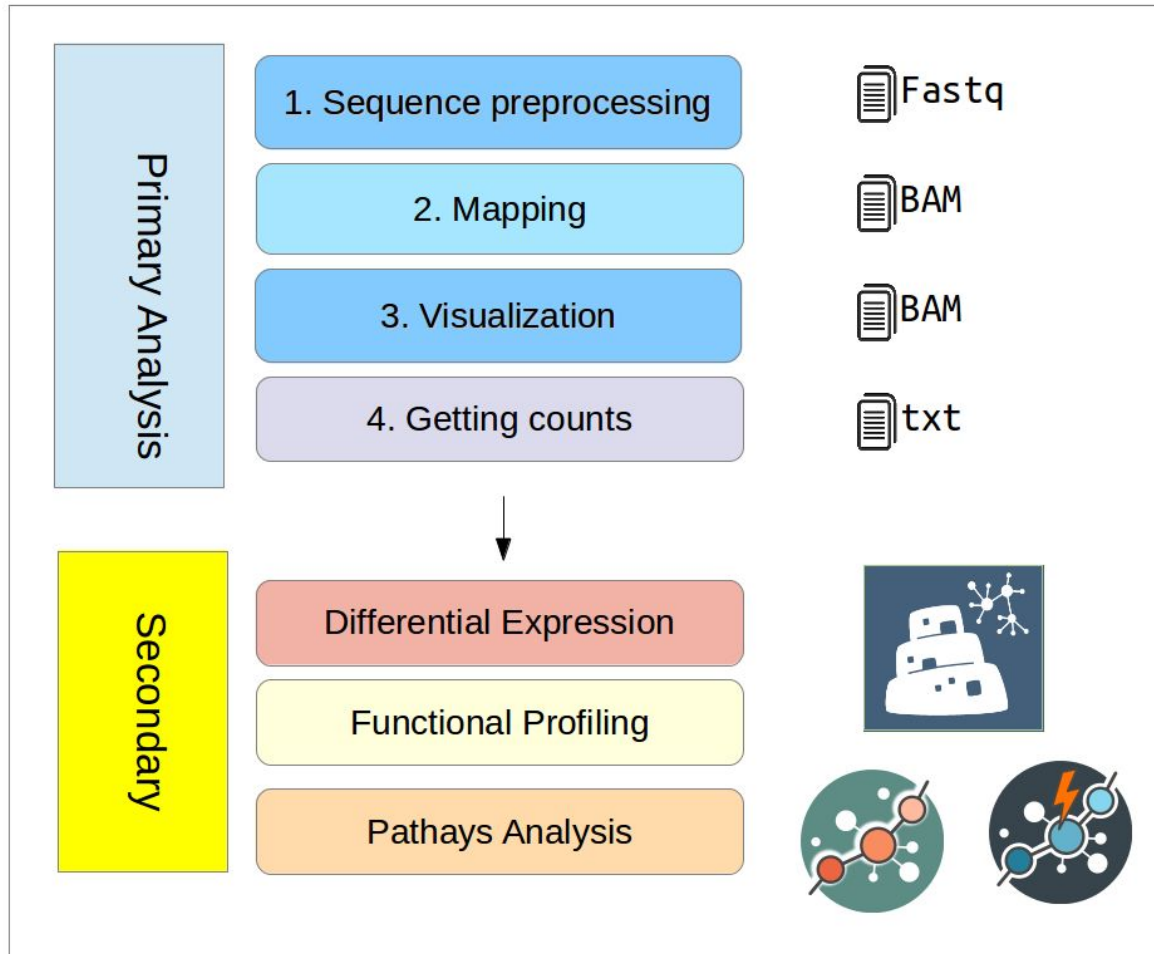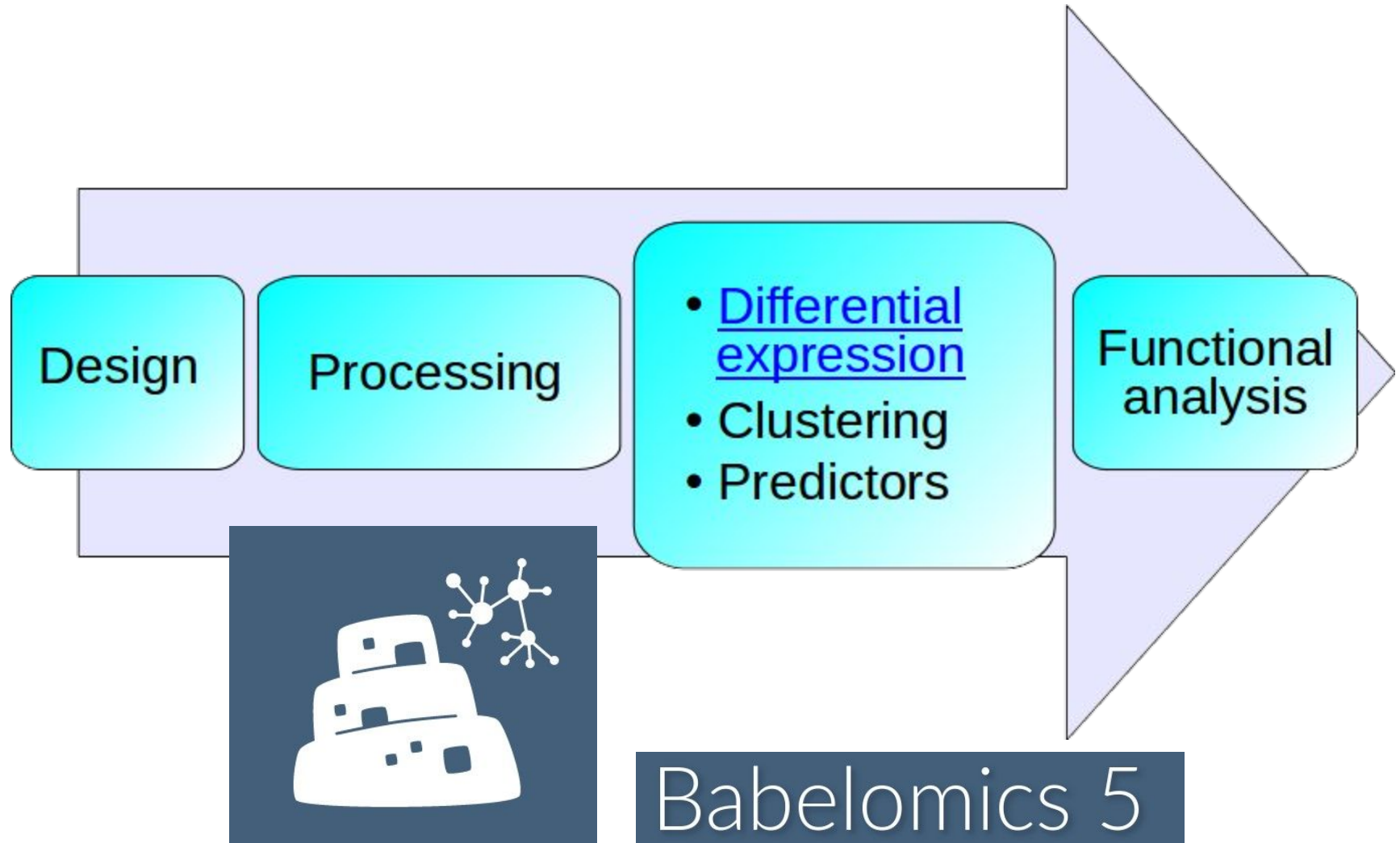# Detecting biomarkers
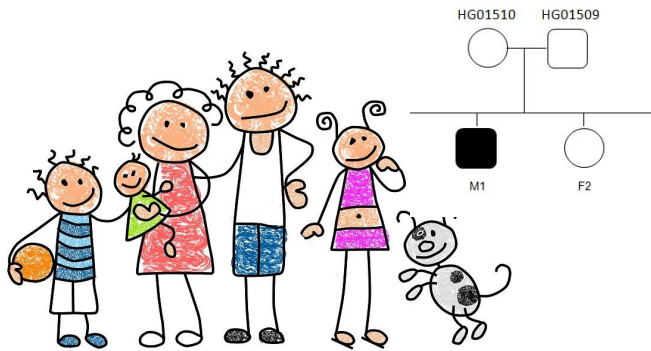
Data analysis workflow

Data analysis workflow

Data analysis workflow

Input

Samples names
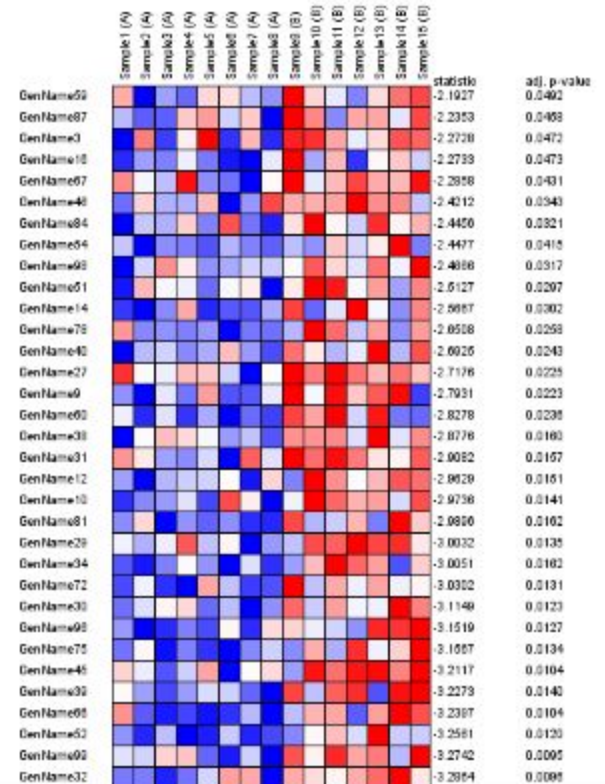
Samples

Tab separated file

genes

#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
200067_x_at	5.538204986516149	4.974604961551534E-6	2.4375564311602516E-4
200052_s_at	5.21107063353314535	1.4519552681244469E-5	4.743053875873193E-4
200054_at	5.102797363044661	4.263454480923057E-5	0.0010445463478826149
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	0.0059605453363083221
200013_at	3.767033234598989	7.042746674112254E-4	0.0062744470369363371
200071_at	3.5180398564848283	0.0014872364080140634	0.012145763998781516
200076_s_at	3.137574787036864	0.003912733450155826	0.02470303941792398
177_at	3.0053355520231624	0.00613746690294133305	0.030073587824412523

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

prev next page 1 of 10

Detecting biomarkers

Continuous variables:
- ➢ Metabolomics
- ➢ Proteomics
- ➢ Transcriptomics arrays
- ➢ Experimental data

Detecting biomarkers

Expression ∨

Unsupervised analysis
▸ Clustering
Supervised analysis
▸ Class prediction
Differential expression
Microarray
▸ Class comparison
▸ Correlation
▸ Survival
RNA-seq
▸ Class comparison

Discrete variables:
➢ RNA-Seq
➢ Experimental data

Different experimental designs

Class comparison

Methods:

Limma, t-test:

$$H_0: \mu_1 = \mu_2$$
$$H_a: \mu_1 \neq \mu_2$$

Fold-change:

$$Log_2 \left(\overline{y}_1 / \overline{y}_2 \right)$$

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

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_n$$
$$H_a: \text{not } H_0$$
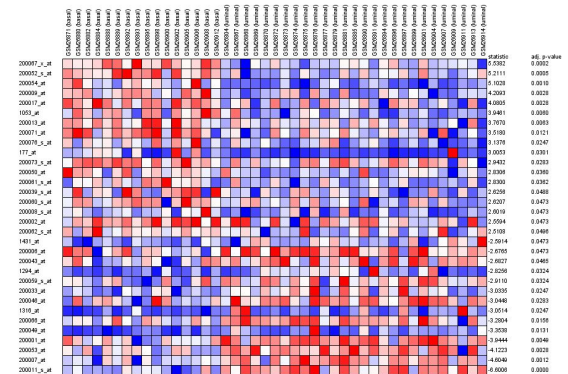
Expression ▾

Unsupervised analysis
  ▸ Clustering
Supervised analysis
  ▸ Class prediction
Differential expression
  Microarray
  ▸ Class comparison
  ▸ Correlation
  ▸ Survival
  RNA-seq
  ▸ Class comparison

Different experimental designs

Correlation

Samples ranked according to the independent variable

Genes ranked by correlation to the continuous variable

Different experimental designs

Survival

Expression ∨

Unsupervised analysis
- ▶ Clustering

Supervised analysis
- ▶ Class prediction
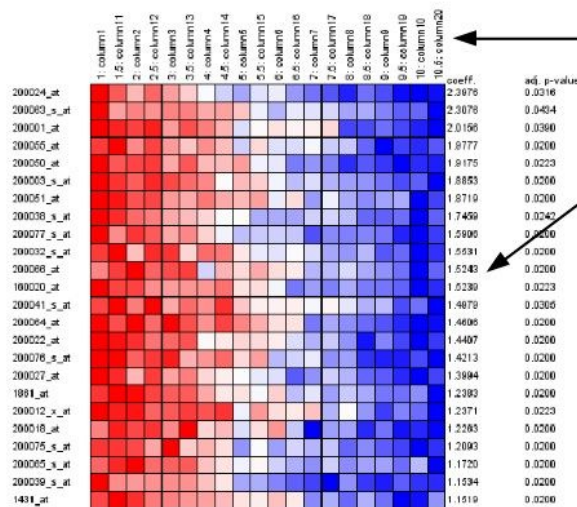
Differential expression

Microarray
- ▶ Class comparison
- ▶ Correlation
- ▶ Survival

RNA-seq
- ▶ Class comparison

Samples ranked according to the survival time

Genes ranked by their relationship with survival time

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

Any question?