Omics-based biomarkers detection

Francisco García García
Bioinformatics and Biostatistics Unit, CIPF

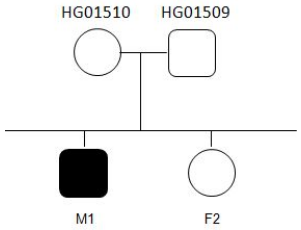
24 Oct 2018

Unidad de Bioinformática y Bioestadística

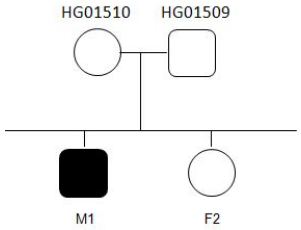
PRINCIPE FELIPE
CENTRO DE INVESTIGACION

WODA
WEB-BASED OMICS DATA ANALYSIS

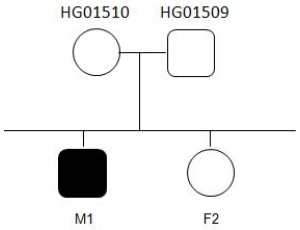
Detecting biomarkers

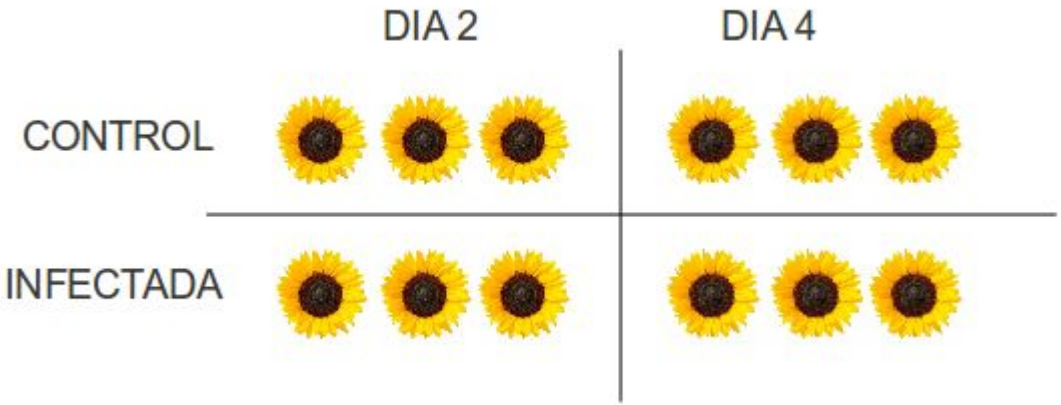


Detecting biomarkers

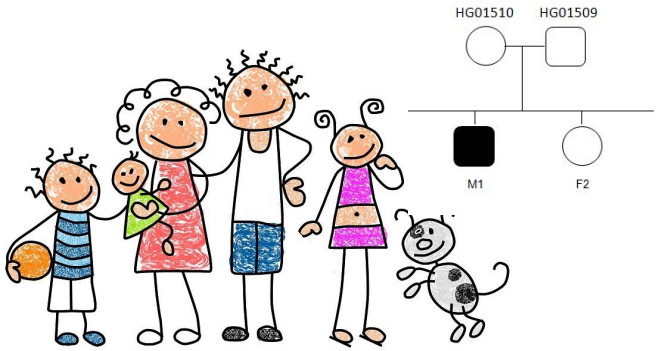


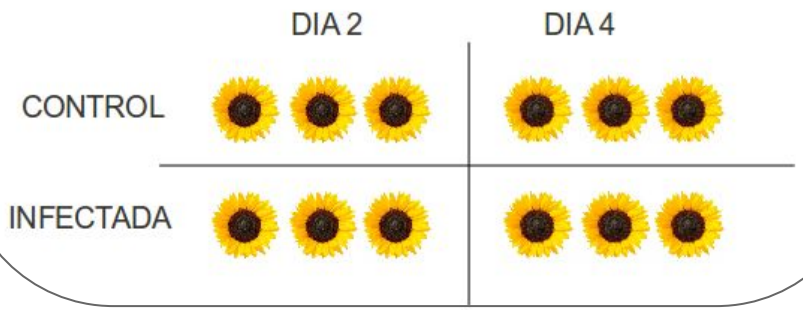
Detecting biomarkers

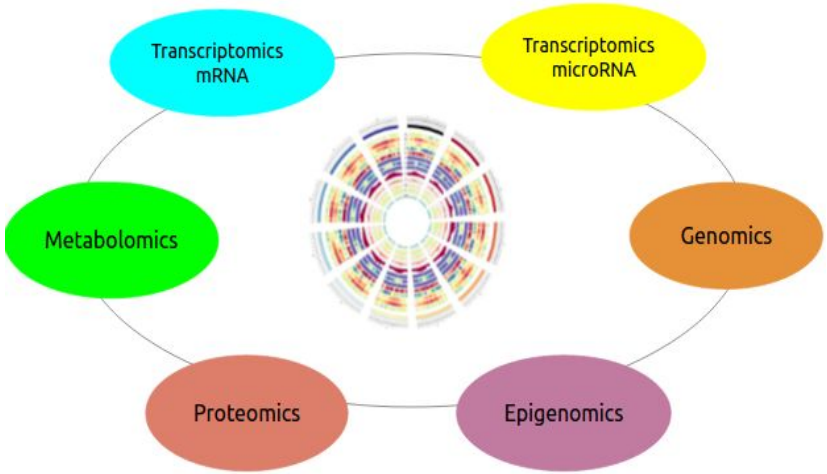




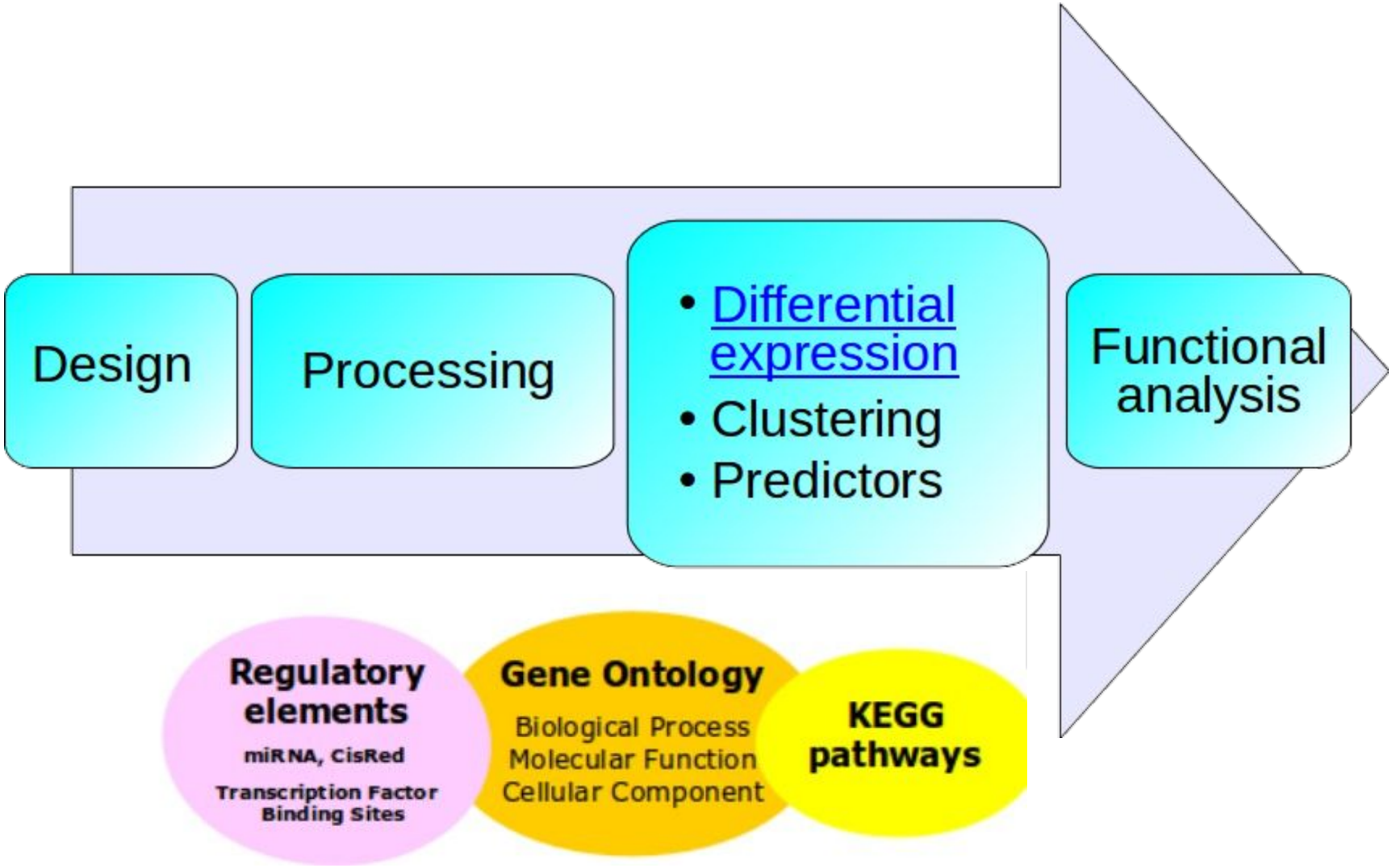
Detecting biomarkers



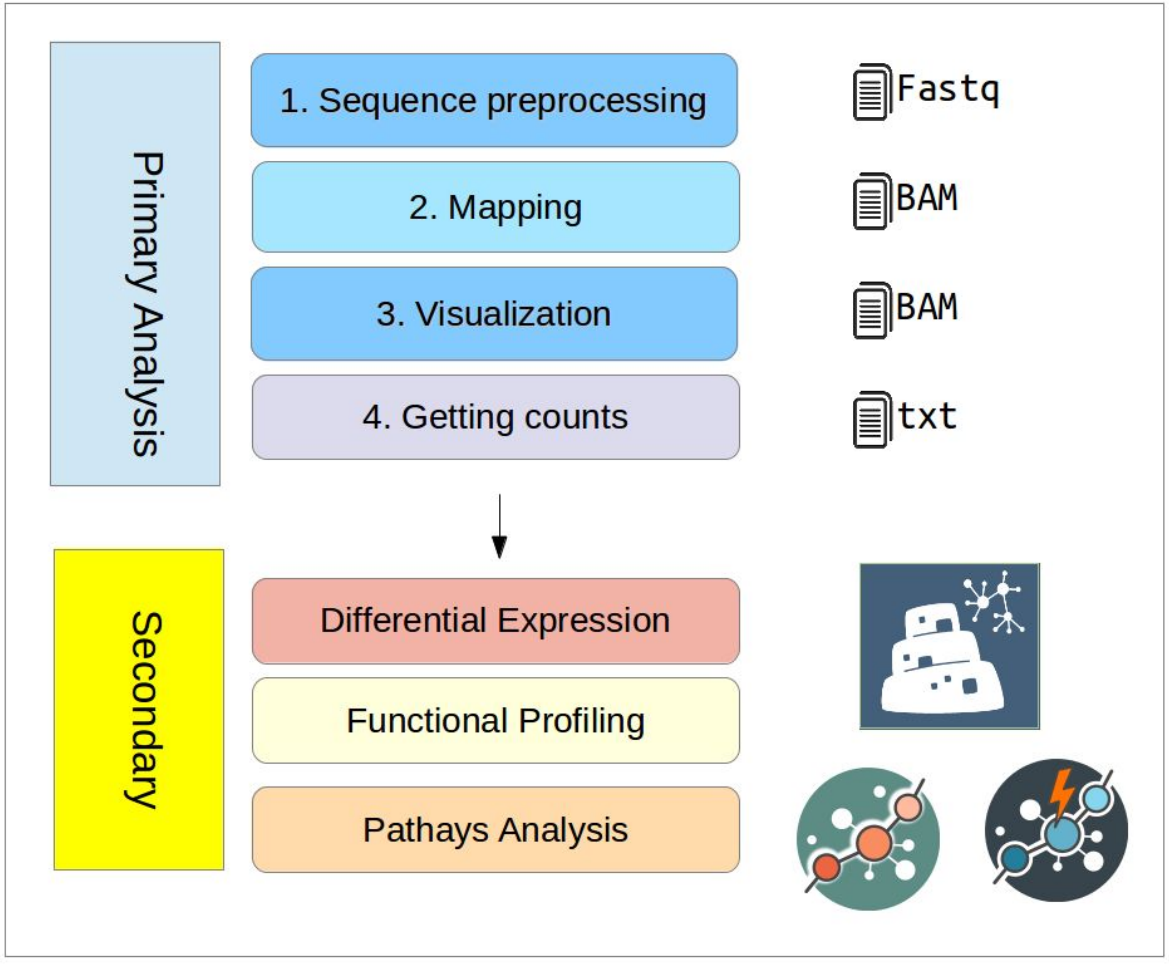




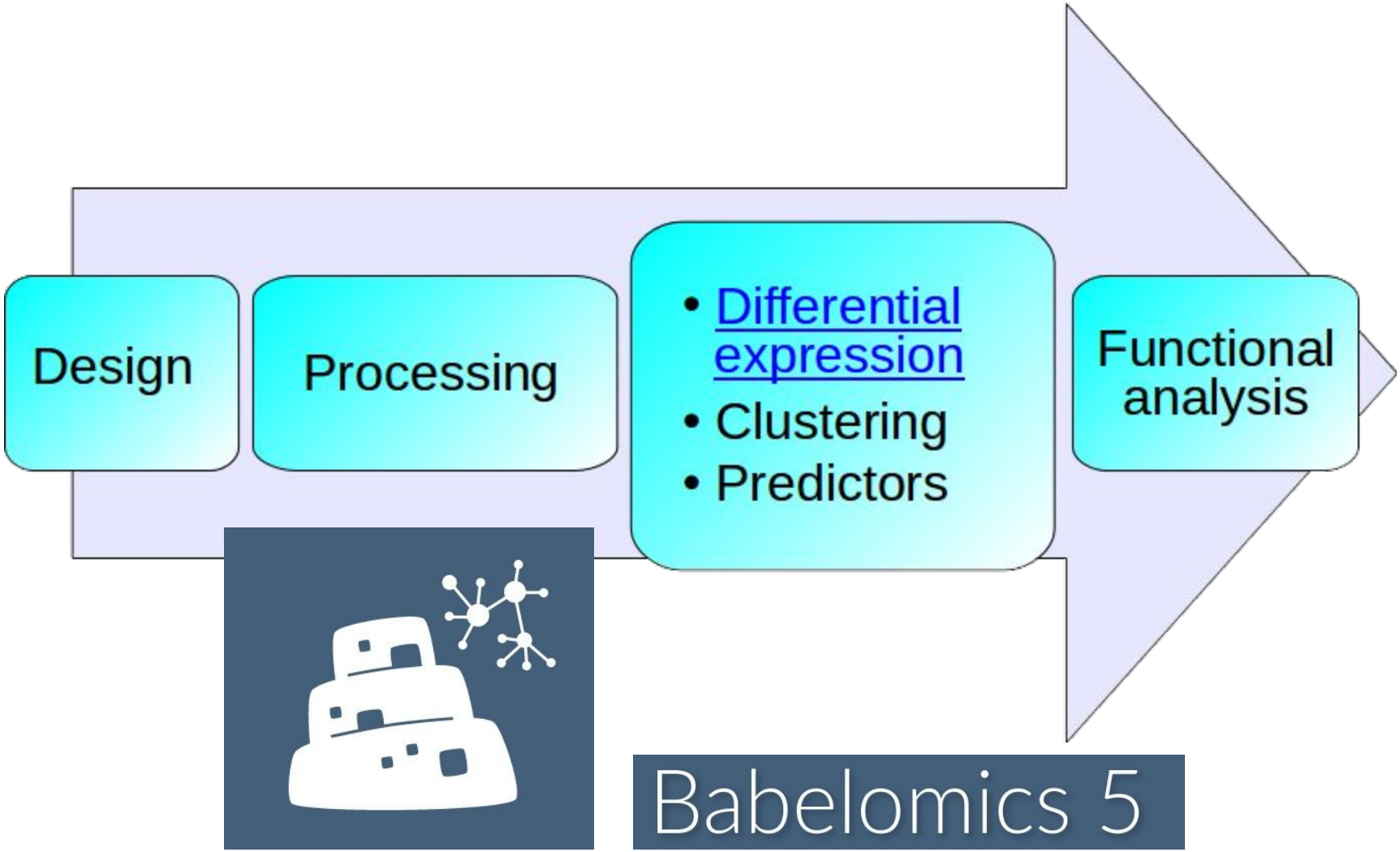
Data analysis workflow



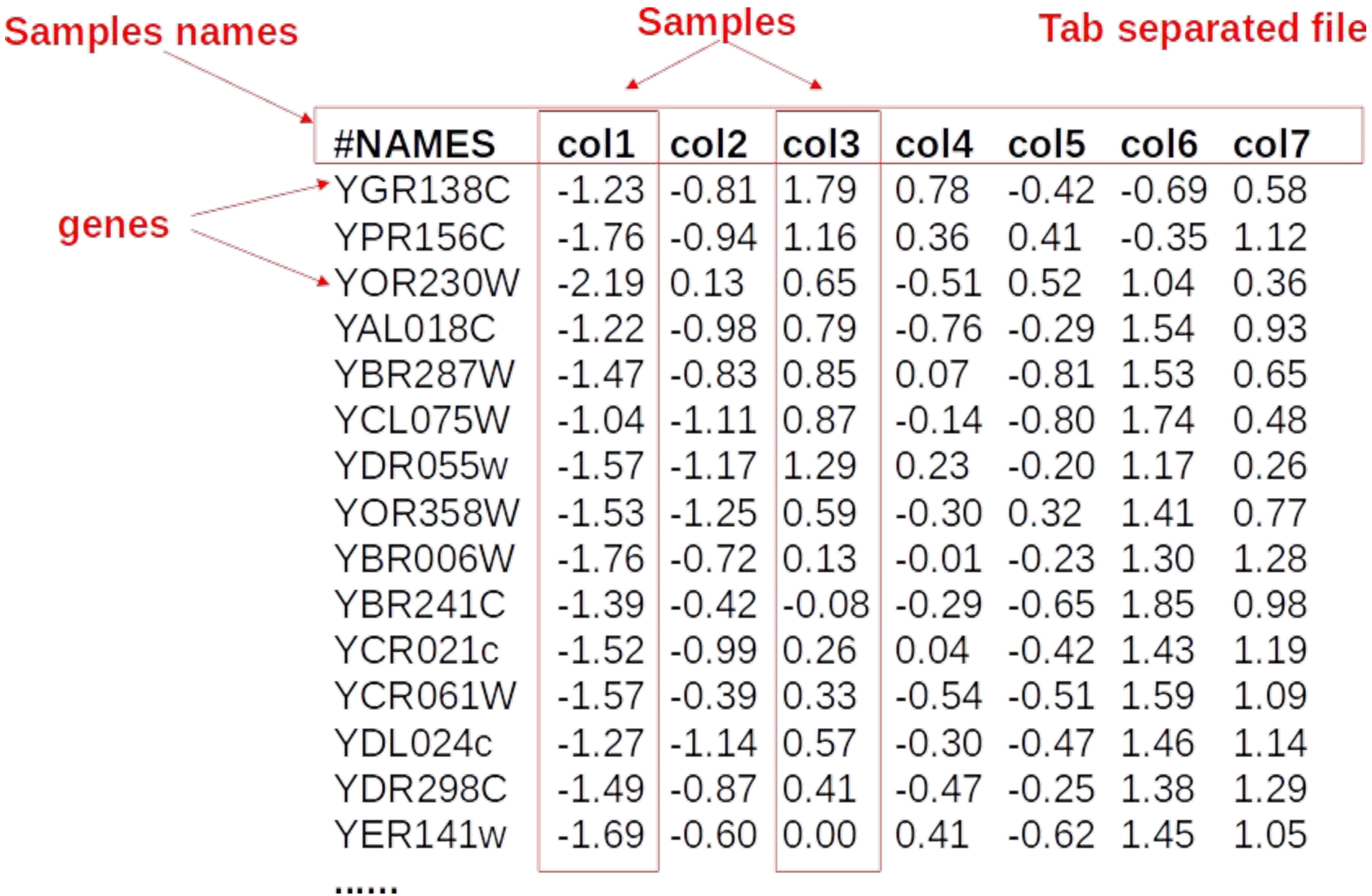
Data analysis workflow



Data analysis workflow



Input

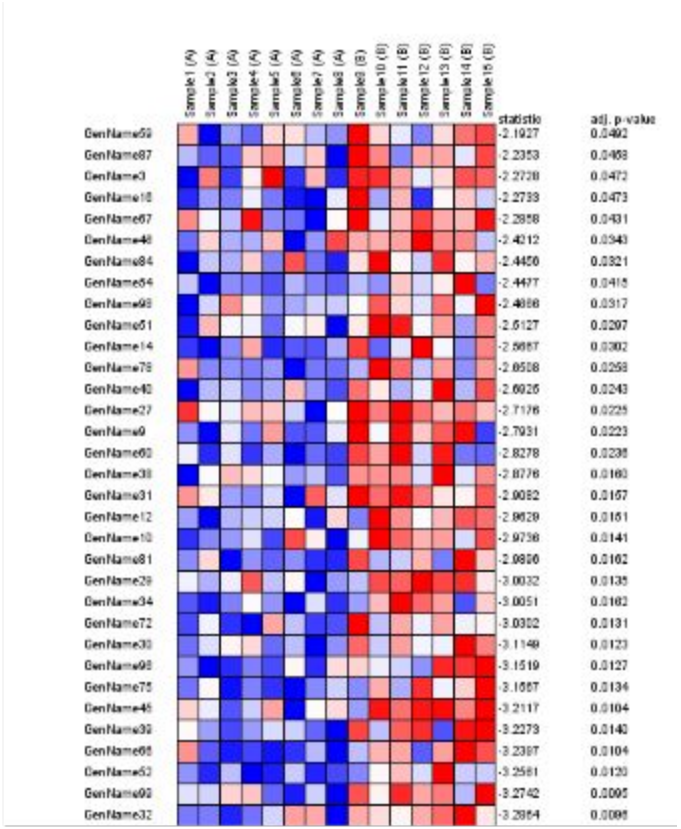


#NAMES	col1	col2	col3	col4	col5	col6	col7
YGR138C	-1.23	-0.81	1.79	0.78	-0.42	-0.69	0.58
YPR156C	-1.76	-0.94	1.16	0.36	0.41	-0.35	1.12
YOR230W	-2.19	0.13	0.65	-0.51	0.52	1.04	0.36
YAL018C	-1.22	-0.98	0.79	-0.76	-0.29	1.54	0.93
YBR287W	-1.47	-0.83	0.85	0.07	-0.81	1.53	0.65
YCL075W	-1.04	-1.11	0.87	-0.14	-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
.....							

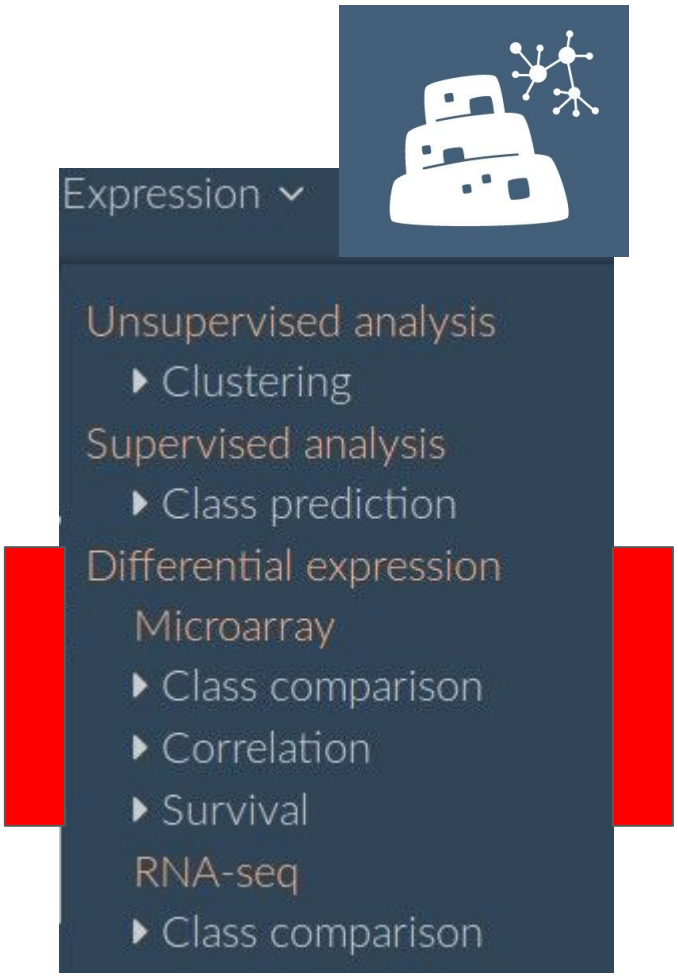
Results

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

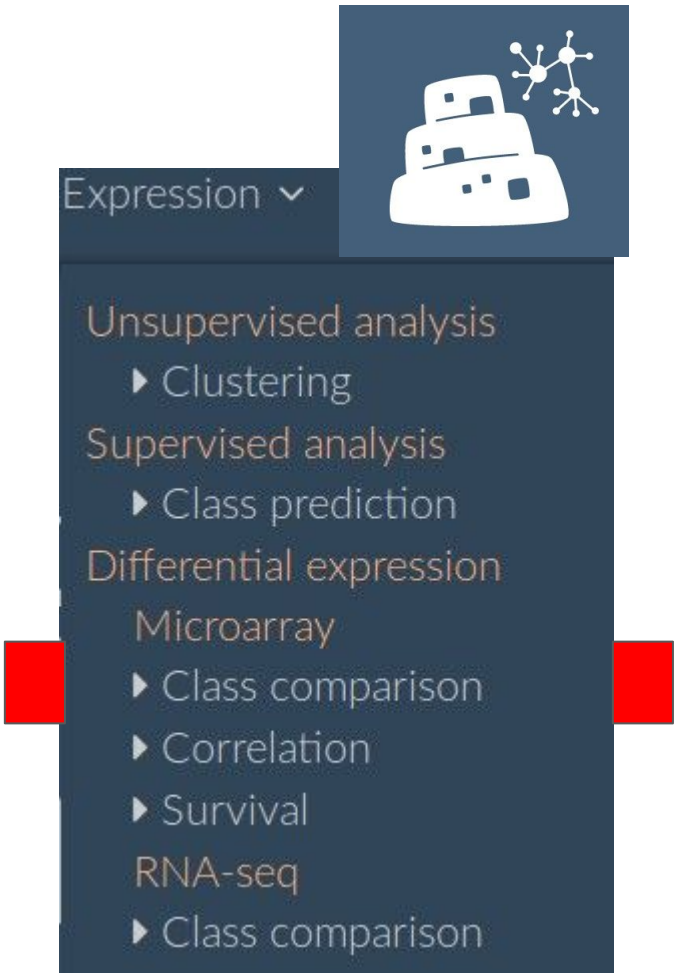


Detecting biomarkers



A. **Continuous** variables:
- Metabolomics
- Proteomics
- Transcriptomics arrays
- Experimental data

Different experimental designs



Class comparison

Methods:

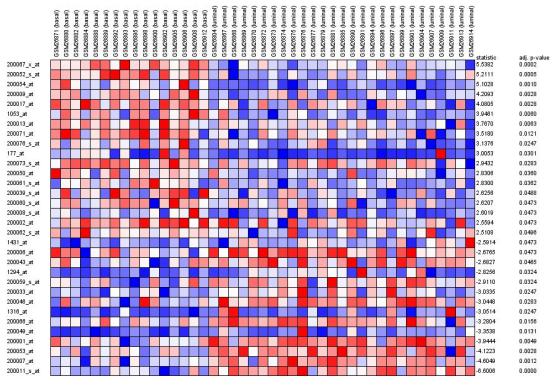
Limma, t-test:

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

Fold-change:

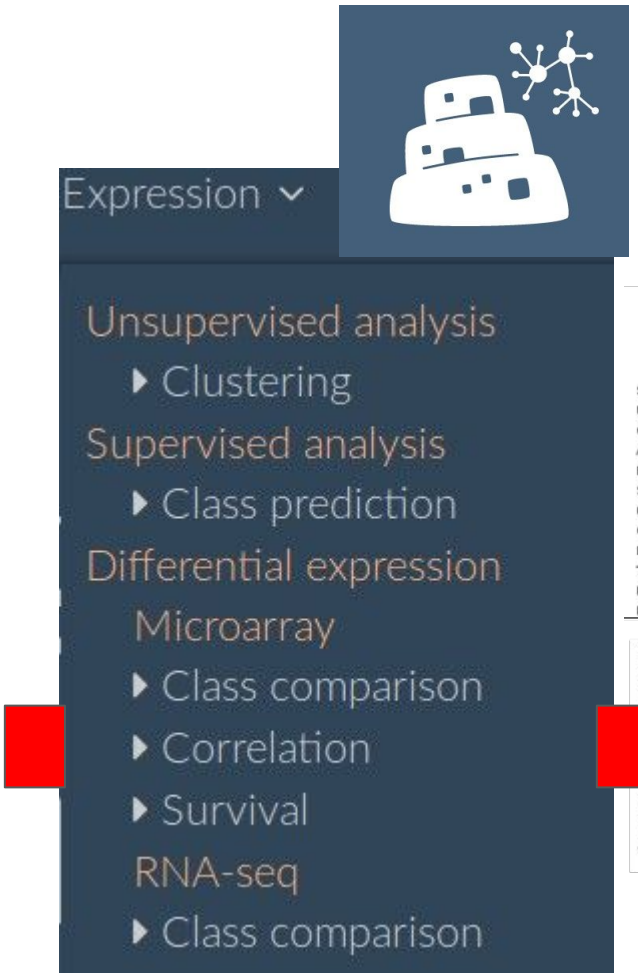
$\text{Log}_2 (\overline{y_1} / \overline{y_2})$

$\overline{y_1} - \overline{y_2}$

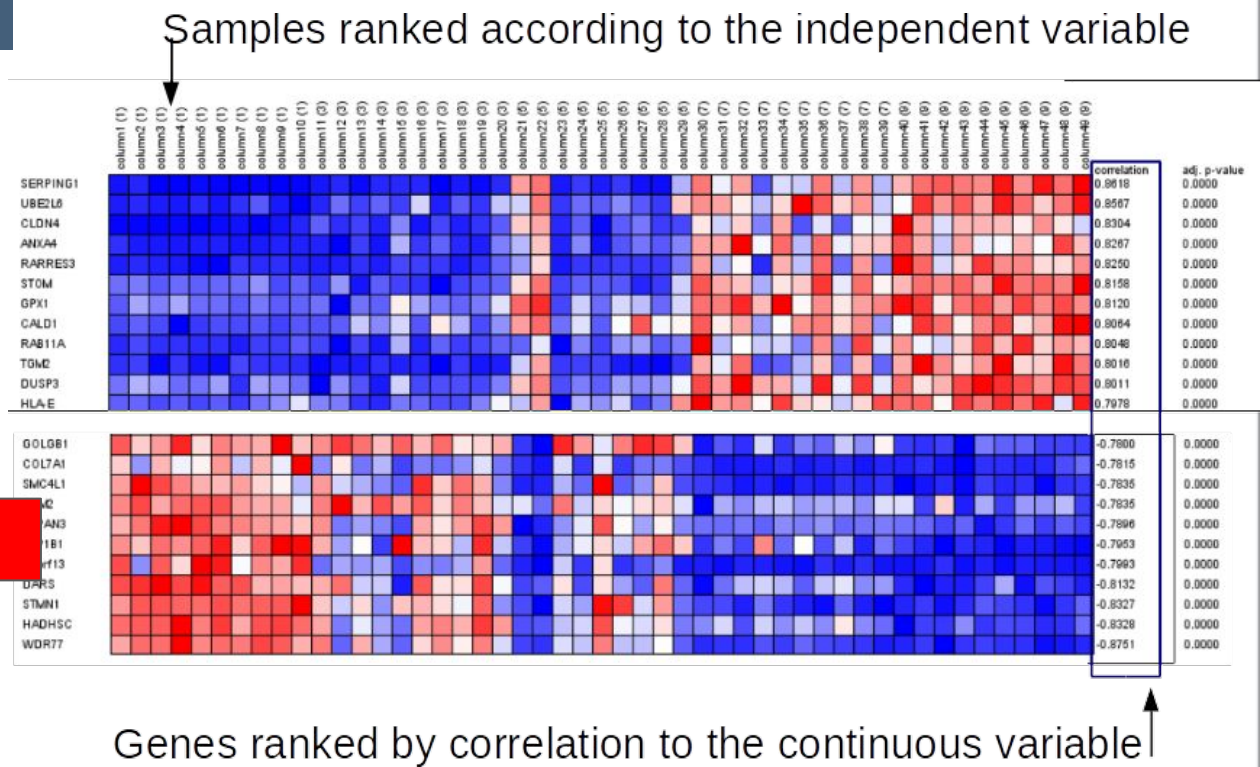


$H_0: \mu_1 = \mu_2 = \ldots = \mu_n$
$H_a:$ not H_0

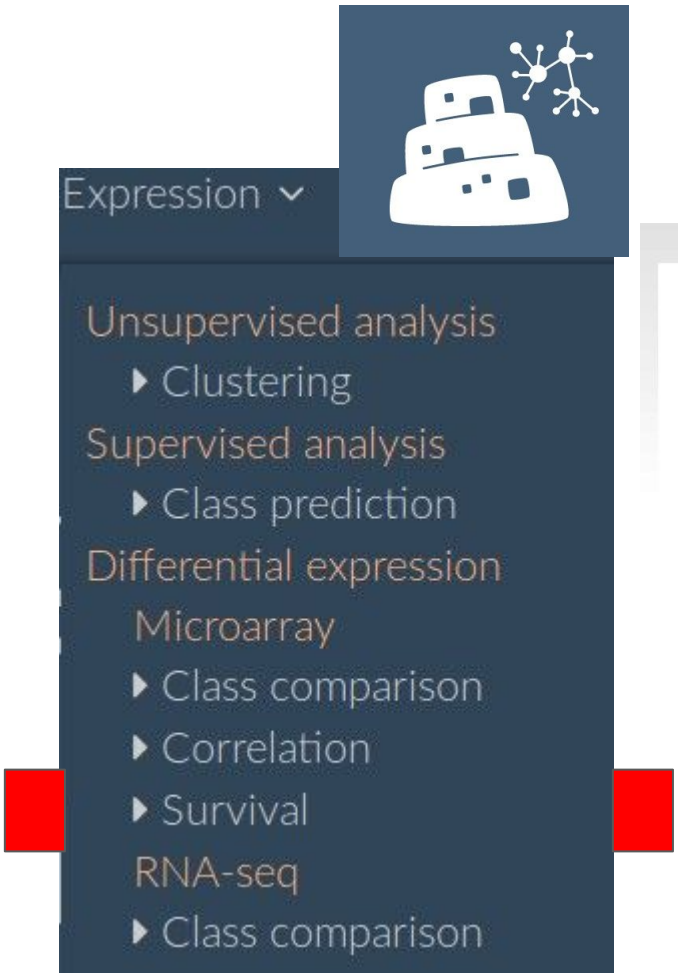
Different experimental designs



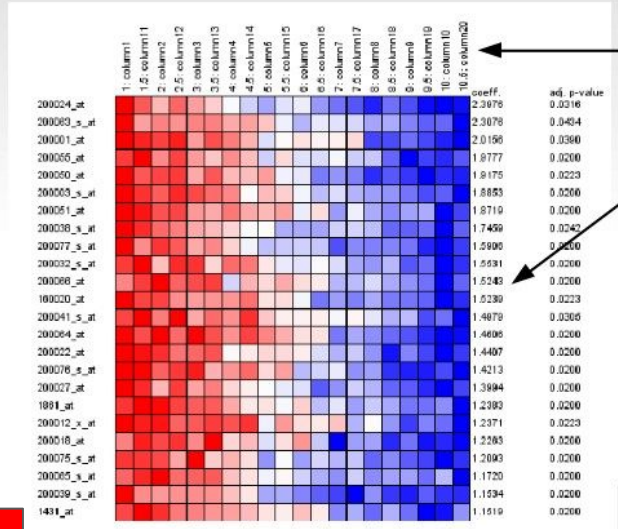
Correlation



Different experimental designs



Survival

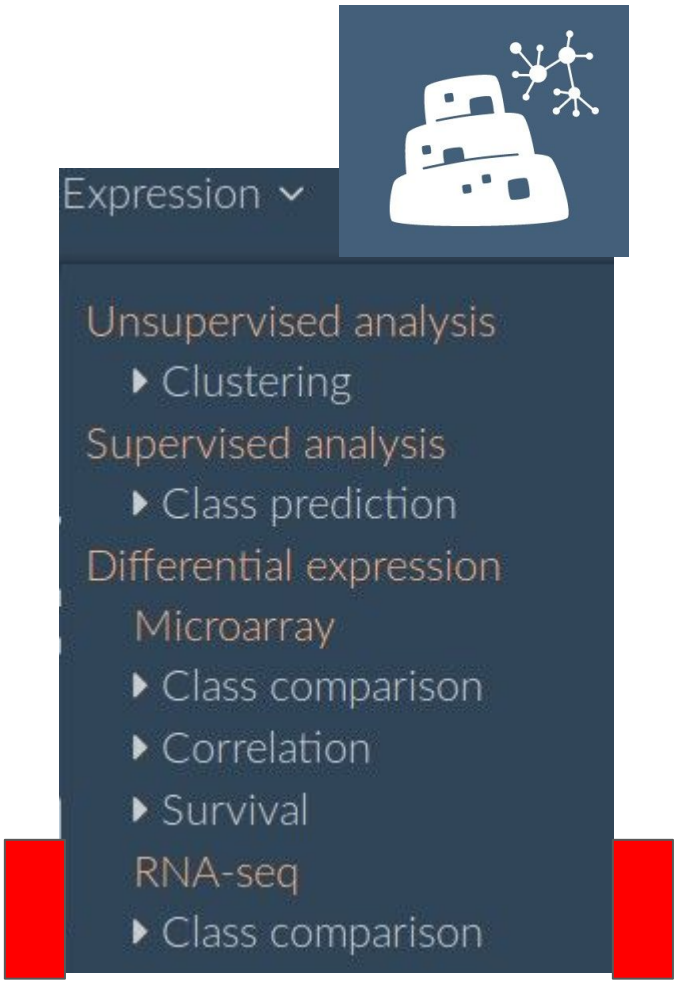


Samples ranked according to the survival time

Genes ranked by their relationship with survival time

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

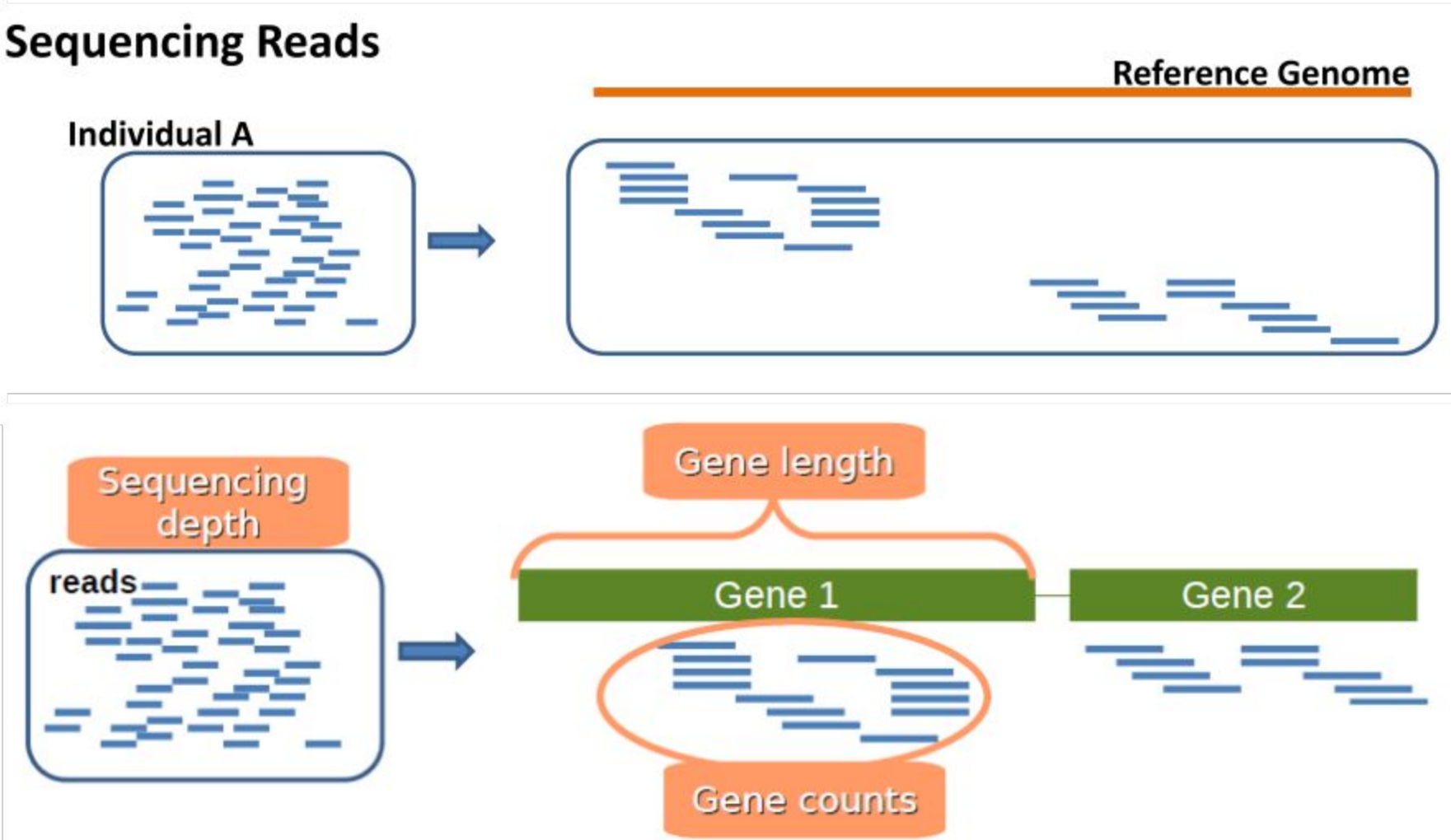
Detecting biomarkers



B. **Discrete** variables:

- RNA-Seq
- Experimental data

General context



Count Normalization

- **Transcript length**: *within* library
- **Library size**: *between* libraries
- Many **other biases** ...
  - Differences on the read count distribution among samples.
  - GC content of the gene affects the detection of that gene (Illumina)
  - sequence-specific bias is introduced during the library preparation

Count Normalization

- **RPKM**: Reads Per Kilobase of the transcript per Million mapped reads

$$RPKM = 10^9 \times \frac{C}{N*L}$$

- **C** is the number of mappable reads mapped onto the gene's exons.
- **N** is the total number of mappable reads in the experiment.
- **L** is the total length of the exons in base pairs.
- Fragments Per Kilobase of exon per Million fragments mapped (FPKM),

RNA-Seq Data Analysis Pipeline

Primary

1. Sequence preprocessing
2. Mapping
3. Quantification

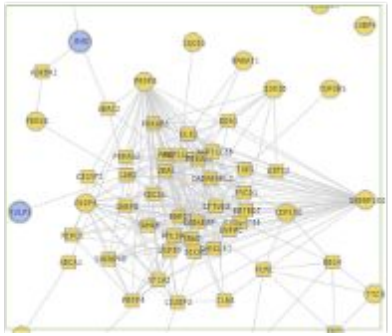
Secondary

4. Normalization
5. Differential expression
6. Functional Profiling

Working in Babelomics

UPLOAD DATA → **EDIT DATA** → **NORMALIZATION + DIFFERENTIAL EXPRESSION** → **FUNCTIONAL PROFILING**

#NAMES	k1	k2	k3	k4	k5	l1	l2	l3	l4	l5
TSPAN6	203	198	194	176	202	157	190	200	201	208
TNMD	0	0	0	1	0	0	0	0	0	0
DPM1	66	85	89	82	80	37	50	50	47	40
SCYL3	21	30	31	27	31	28	31	37	15	21
C1orf112	10	12	8	11	18	17	22	12	12	19
FGR	19	28	18	20	10	47	50	43	49	48
FUCA2	240	272	261	256	211	76	82	85	68	83
GCLC	98	100	84	94	86	354	362	373	369	326
NFYA	59	61	53	56	59	59	66	63	66	62
STPG1	34	43	41	31	48	6	7	7	8	7



Any question?