

# ***Classification of real-time PCR data***

Mikael Kubista

Freising, Germany

**TATAA Biocenter**

# Clustering

---

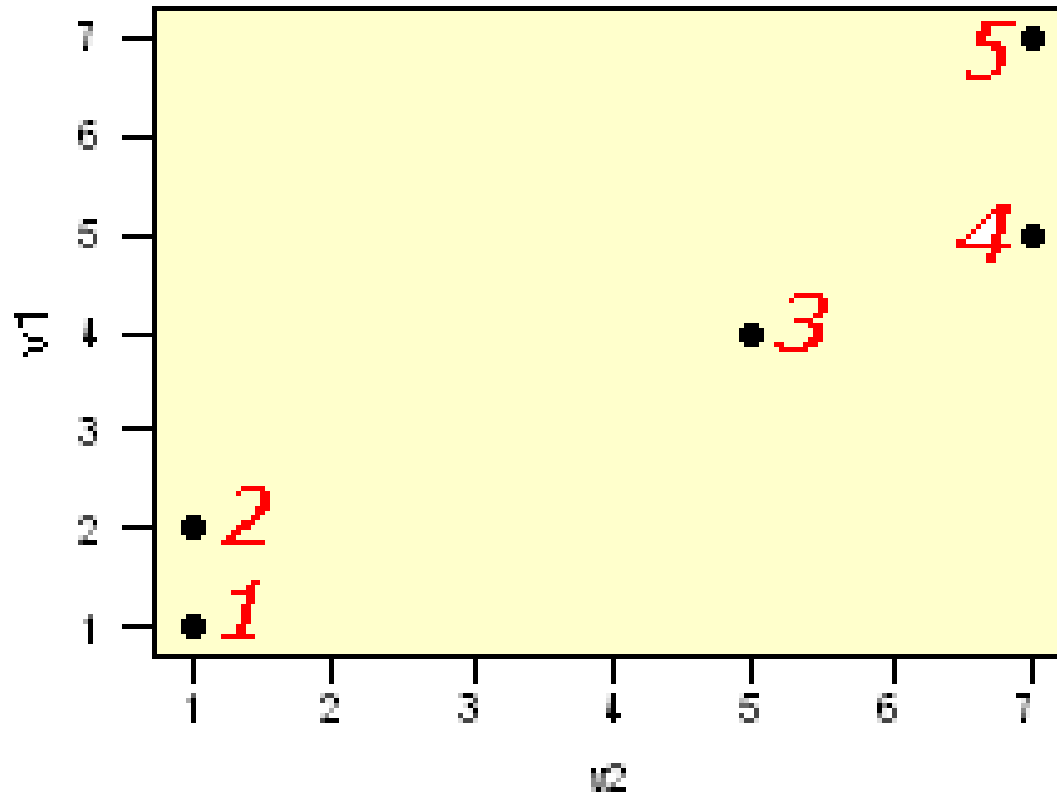
Genes/samples that behave similarly are identified by their expression patterns

# Example

---

<b>Sample</b>	<b>Gene 1</b>	<b>Gene2</b>
1	1	1
2	2	1
3	4	5
4	7	7
5	5	7

# Graphical representation



---

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>1</b>	0.0				
<b>2</b>	1.0	0.0			
<b>3</b>	5.0	4.5	0.0		
<b>4</b>	8.5	7.8	3.6	0.0	
<b>5</b>	7.2	6.7	2.2	2.0	0.0

---

	1	2	3	4	5
1	0.0				
2	1.0	0.0			
3	5.0	4.5	0.0		
4	8.5	7.8	3.6	0.0	
5	7.2	6.7	2.2	2.0	0.0

---

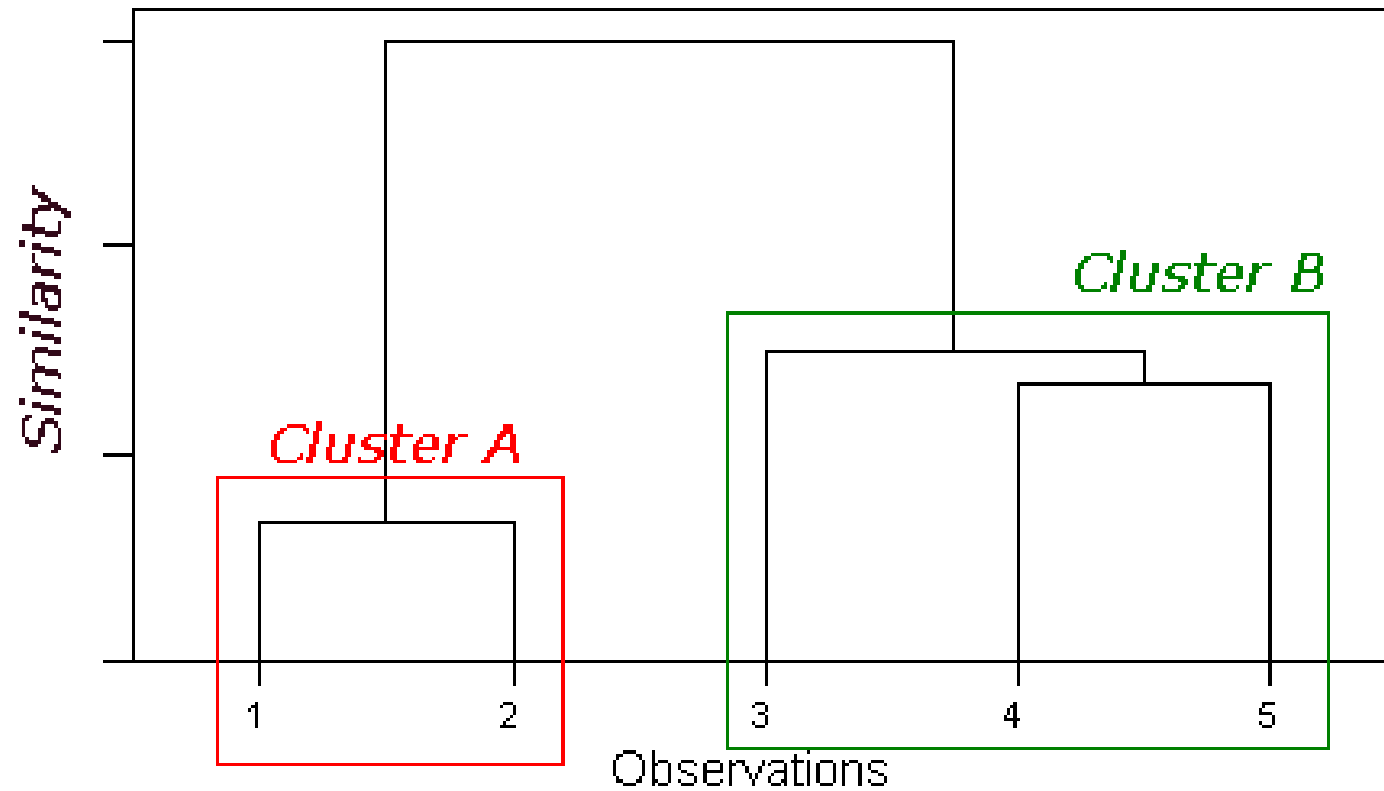
	A	3	4	5
A	0.0			
3	4.7	0.0		
4	8.1	3.6	0.0	
5	6.9	2.2	2.0	0.0

---

	A	B	3
A	0.0		
B	7.5	0.0	
3	4.7	2.8	0.0

---

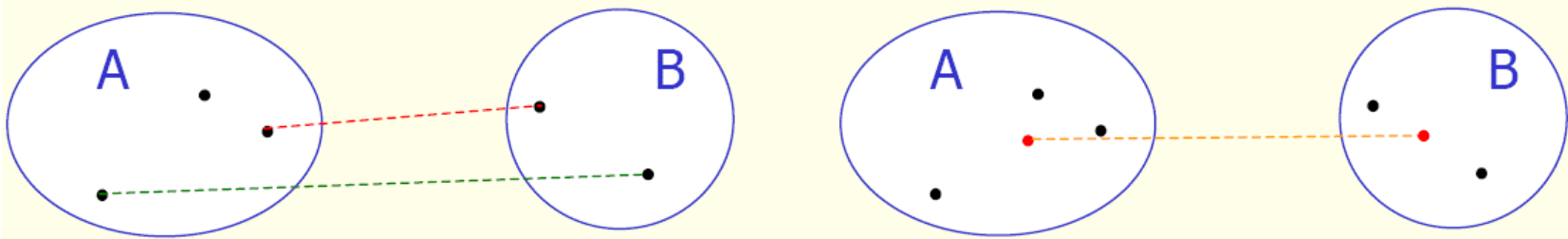
	A	B
A	0.0	
B	6.4	0.0



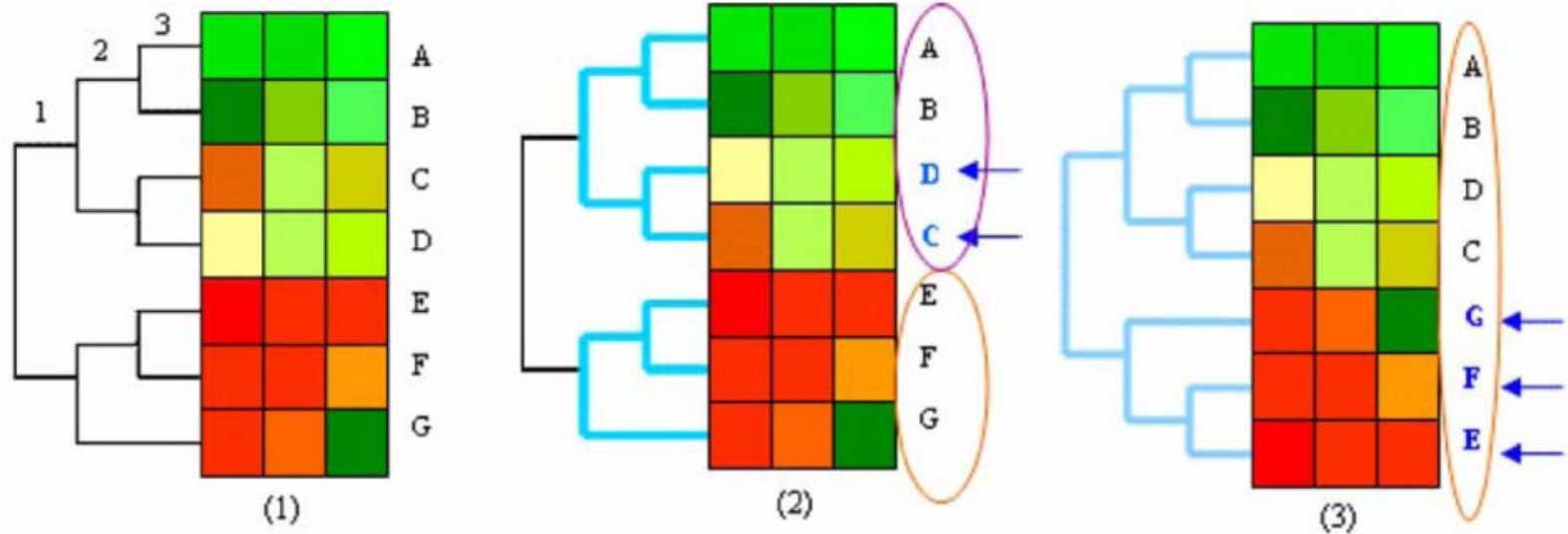
# Distance measures

- Single linkage
- Complete linkage
- Average linkage
- Centroid linkage

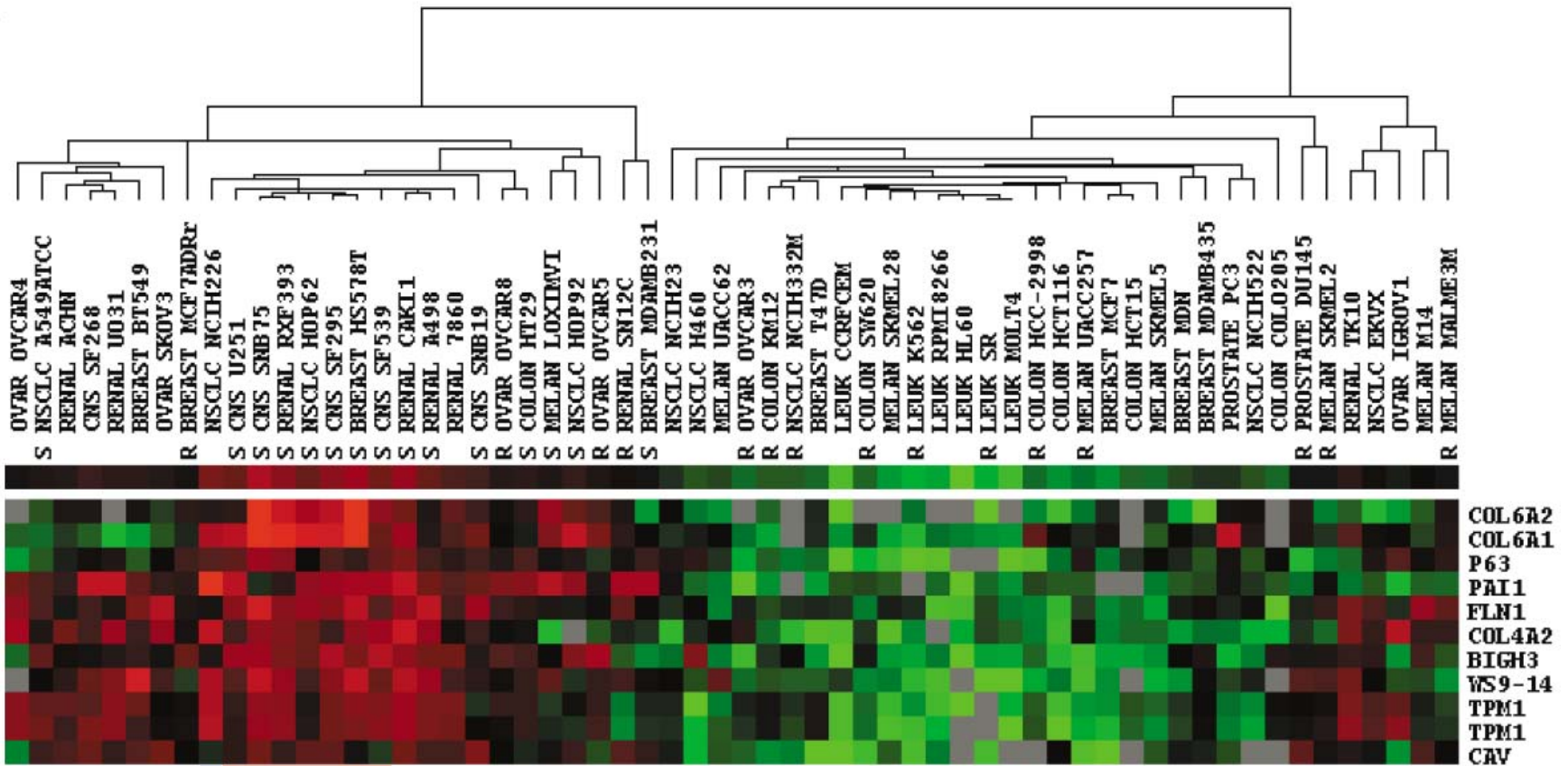
- “Closeness”: *single linkage* (minimum distance) *complete linkage* (maximum distance), *mean distance* and *distance between centroids*.



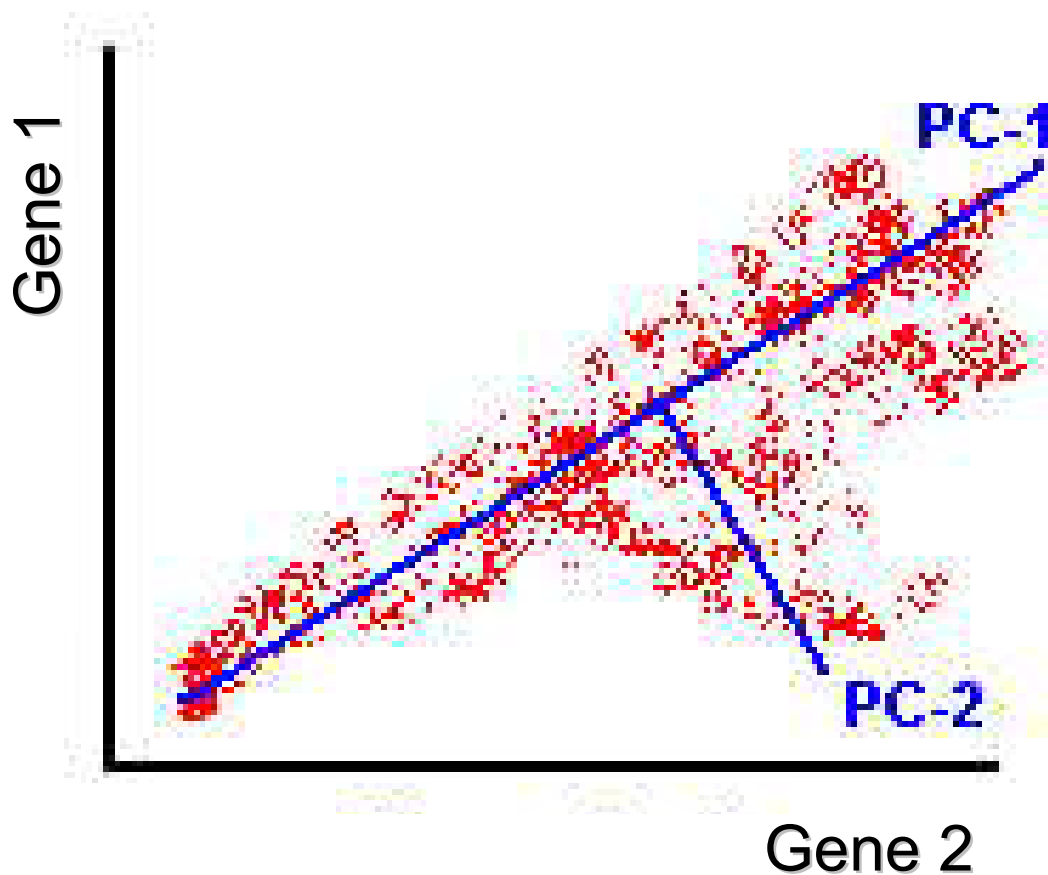
# Color coding and rearranging



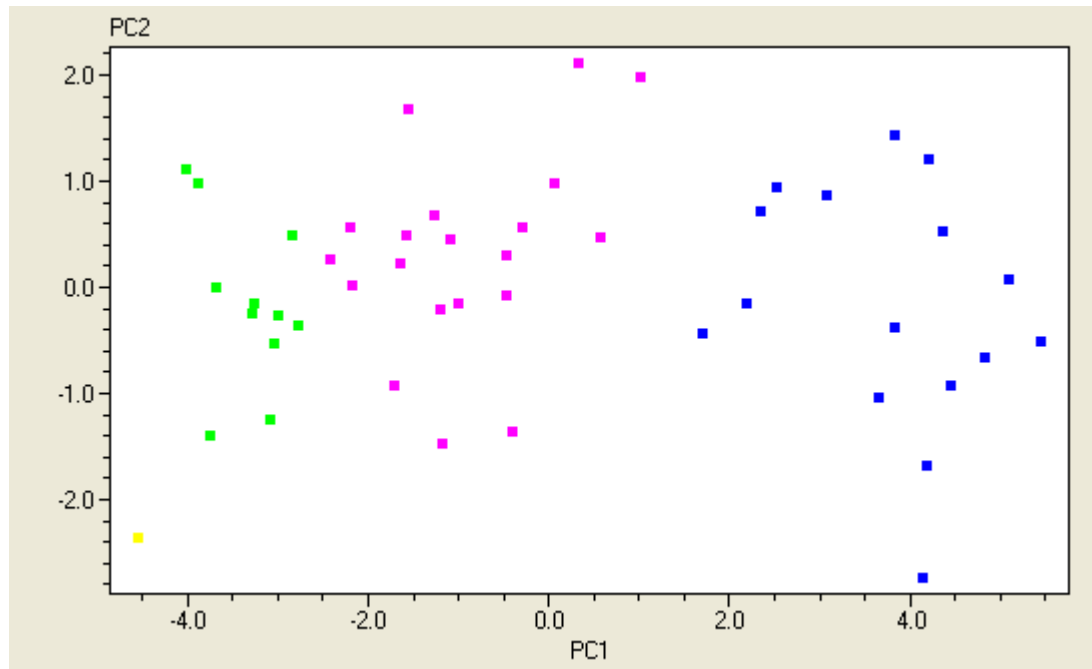
# Response to cytochalasin-D in 60 cancer cell lines



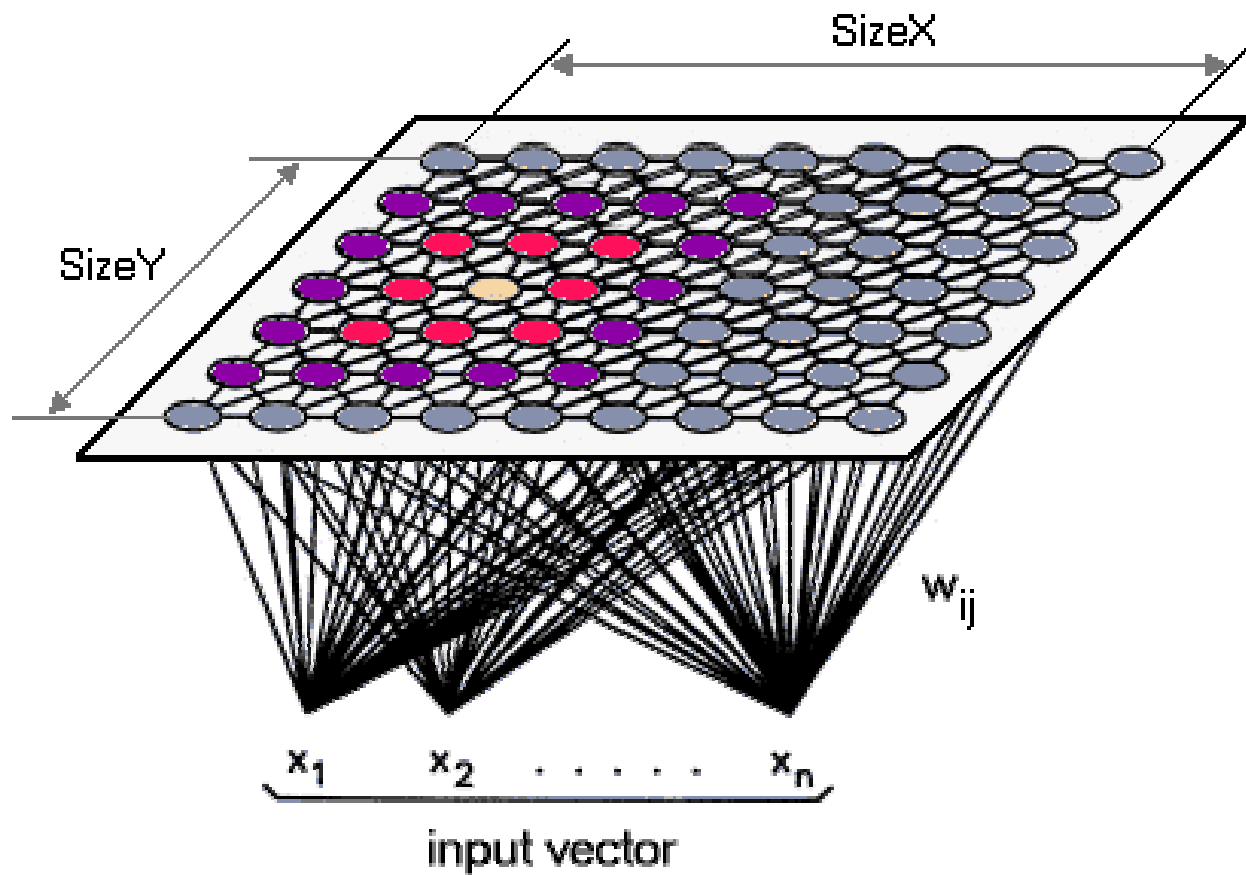
# Principle components



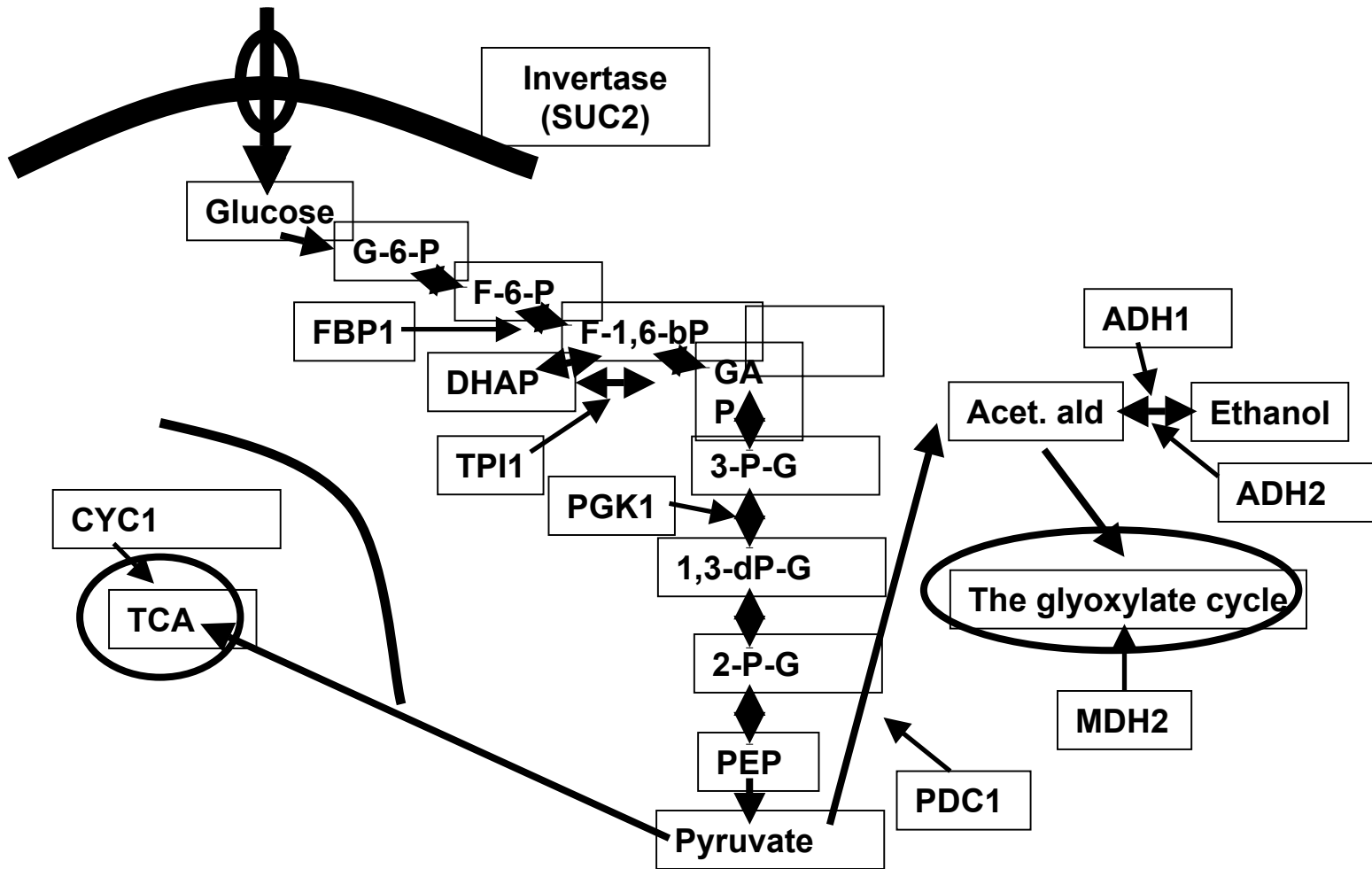
# Scatter plot (samples)



# Self organizing map



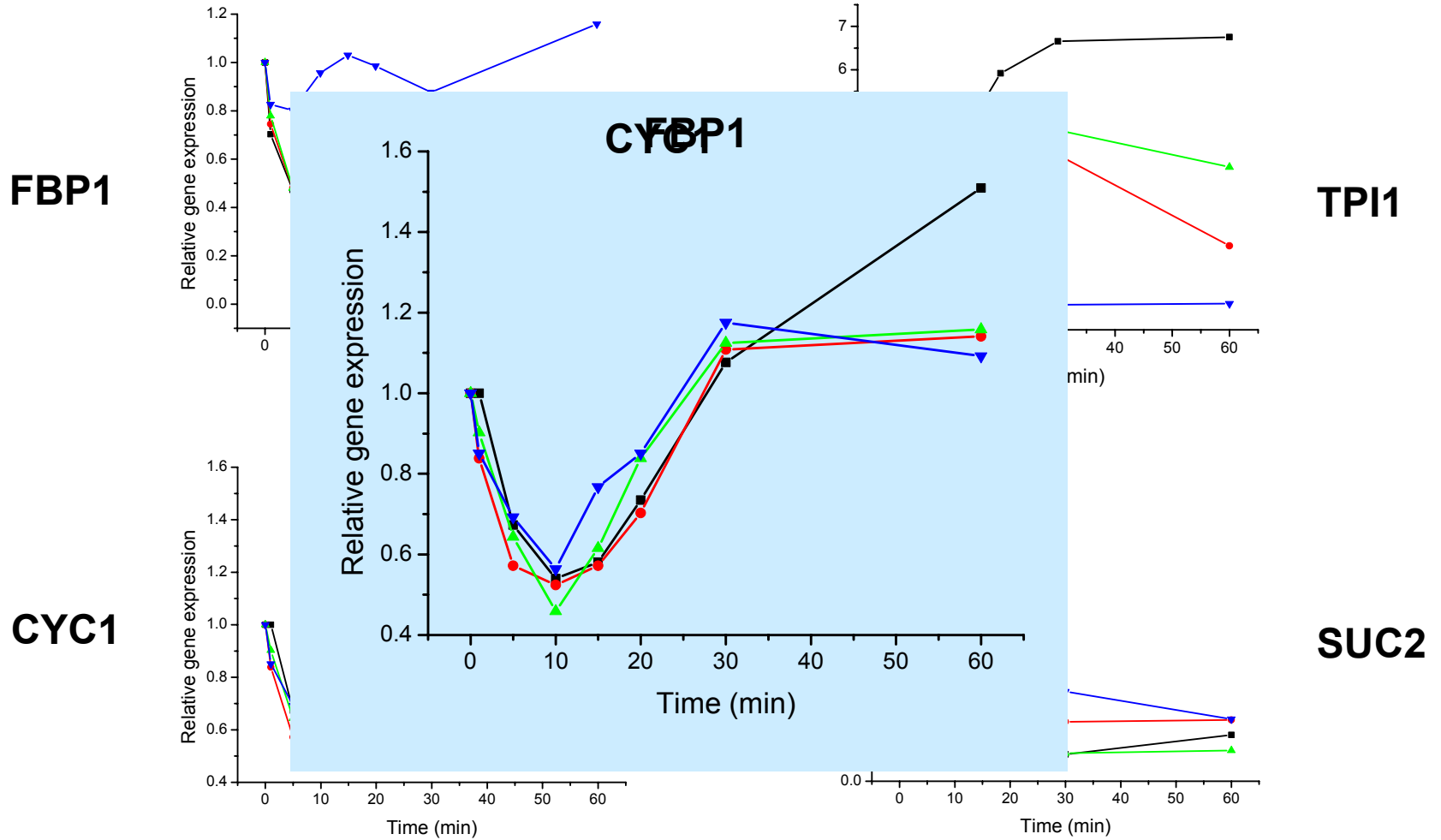
# Yeast metabolic pathway



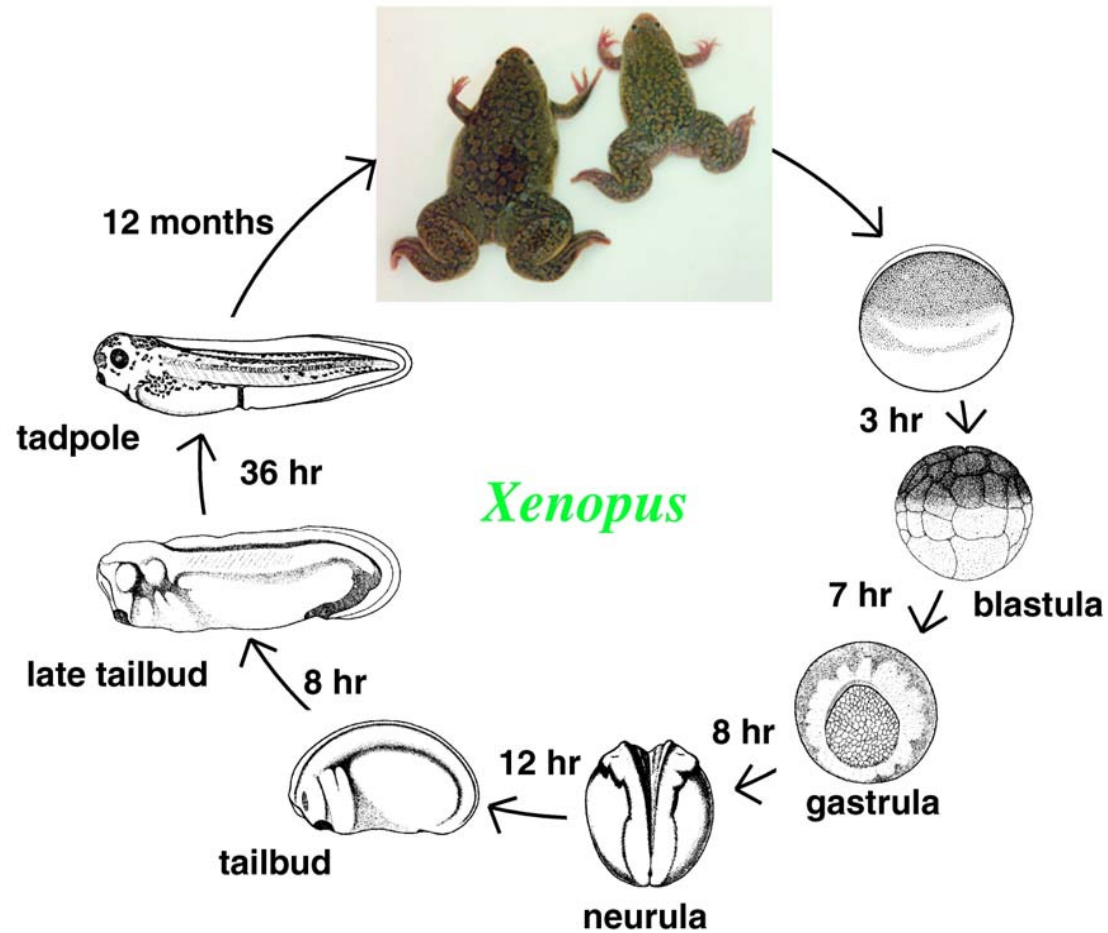
# Experimental design

- Four strains of yeast: Wt, Hxt7, Tm6 and Null
- Expression over time after glucose addition: 0 – 60 min
- Expression of genes:
- Genes:
  - » Ref: ACT, IPPI, PDA
  - » Glycolysis: TPI, PGK, PDC, ADH1
  - » Glycogenesis: FBP, MDH2, SUC2, ADH2
  - » Unknown: ADH3, ADH4, ADH5, ADH6
  - » HSP = heat shock
  - » CYC = Cytochrom c

# Gene expression after glucose stimuli



# Embryonic Development



# Normalization

Measure expression of  $n$  genes in  $m$  samples  $\Rightarrow n \times m$  data matrix.

wtcopy_new	ACT	IPPI	PGK	MDH2	PDA	PDC
0	7200000	5220000	232277.3263	5390000	1810000	953377.1536
1	8190000	5220000	290784.0484	4900000	1990000	1540000
5	10600000	6130000	608327.1781	2930000	2750000	4740000
10	9010000	6130000	1750000	1080000	2130000	9310000
15	7680000	5390000	3230000	210956.921	1990000	13200000
20	6540000	4040000	2340000	51396.77346	1270000	13700000
30	6130000	3550000	1930000	32795.05287	1230000	15600000
60	5220000	4900000	1750000	23040.60374	1400000	17700000

Normalize genes or samples or both?

- Normalize samples to total RNA
- Autoscale expression (subtract mean and divide with standard deviation)

# CT or copy number or....?

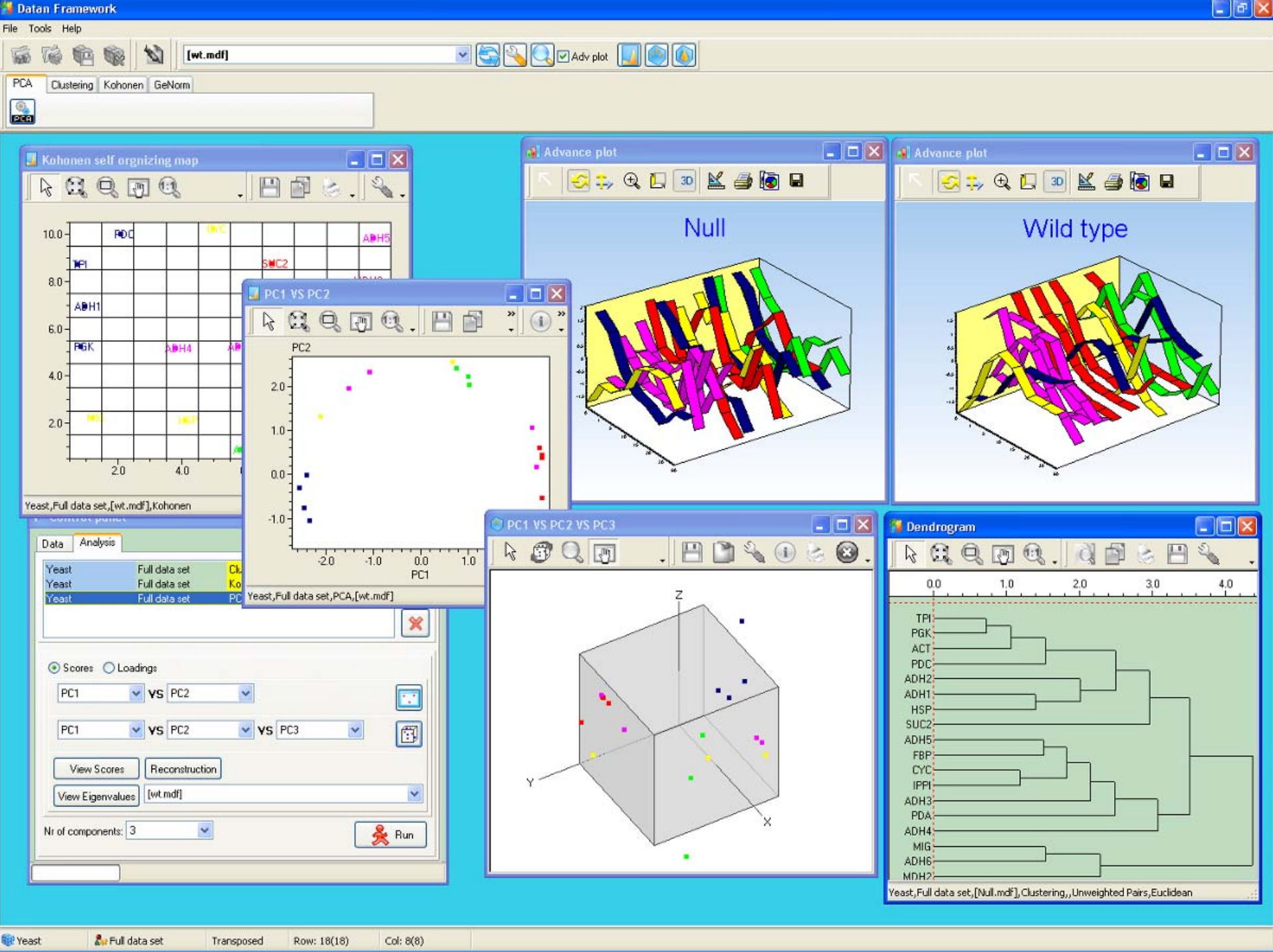
wild-type	ACT	IPPI	PGK	MDH2	PDA	PDC
0	13.137183240...	13.683175960...	18.979299724...	13.628575207...	15.484946347...	16.576933978...
1	12.918786358...	13.683175960...	18.597105174...	13.792374700...	15.321149997...	15.757948032...
5	12.481985953...	13.410178658...	17.341323079...	14.665959540...	14.775160715...	13.846974473...
10	12.754989288...	13.410178658...	15.539545162...	16.358530032...	15.211955441...	12.700389173...
15	13.027983700...	13.628575207...	14.502161146...	19.143097389...	15.321149997...	12.099801368...
20	13.300981702...	14.119971011...	15.048154420...	21.545463134...	16.085547661...	12.045197567...
30	13.410178658...	14.338364059...	15.375747145...	22.309852235...	16.140141756...	11.826804881...
60	13.683175960...	13.792374700...	15.539545162...	22.910443672...	15.921748317...	11.608408111...

wild-type	ACT	IPPI	PGK	MDH2	PDA	PDC
0	30768541.017...	21672522.713...	723762.19819...	22445517.517...	6818039.0306...	3382704.0140...
1	35398578.367...	21672522.713...	924983.53779...	20205515.019...	7573878.0054...	5722150.5564...
5	46853874.120...	25823104.192...	2071036.6496...	11533337.603...	10752632.031...	19509675.293...
10	39322837.066...	25823104.192...	6583242.7112...	3891749.5613...	8123752.4560...	40725349.719...
15	33002486.975...	22445517.517...	12811925.270...	651533.43506...	7573878.0054...	59879868.221...
20	27697941.408...	16373867.560...	9024368.1858...	139401.87175...	4637027.8502...	62015727.403...
30	25823104.192...	14232245.438...	7313060.2925...	85347.811748...	4477353.6973...	71347636.263...
60	21672522.713...	20205515.019...	6583242.7112...	58046.409991...	5151092.6485...	82083993.220...

- Most statistics is developed for normal distributed data.
- Sampling theorem: test samples taken from the same population will exhibit normal distribution in copy number.
- Many effects in nature are logarithmic!

Compare test samples drawn from the same population in copy number!

Compare test samples from different populations based in log(copy number)!



# Acknowledgement

- Amin Forootan, MultiD Analyses AB
- Björn Sjögren, MultiD Analyses AB
- José Manuel Andrade, University of a Coruña
- Anders Stålberg, TATAA Biocenter
- Karin Elbing, Chalmers University
- Radek Sindelka, Institute of molecular genetics, Prague.