

# Bioinformatics Analysis in R

## Gene Expression Analysis

Ivan G. Costa, Joseph Kuo

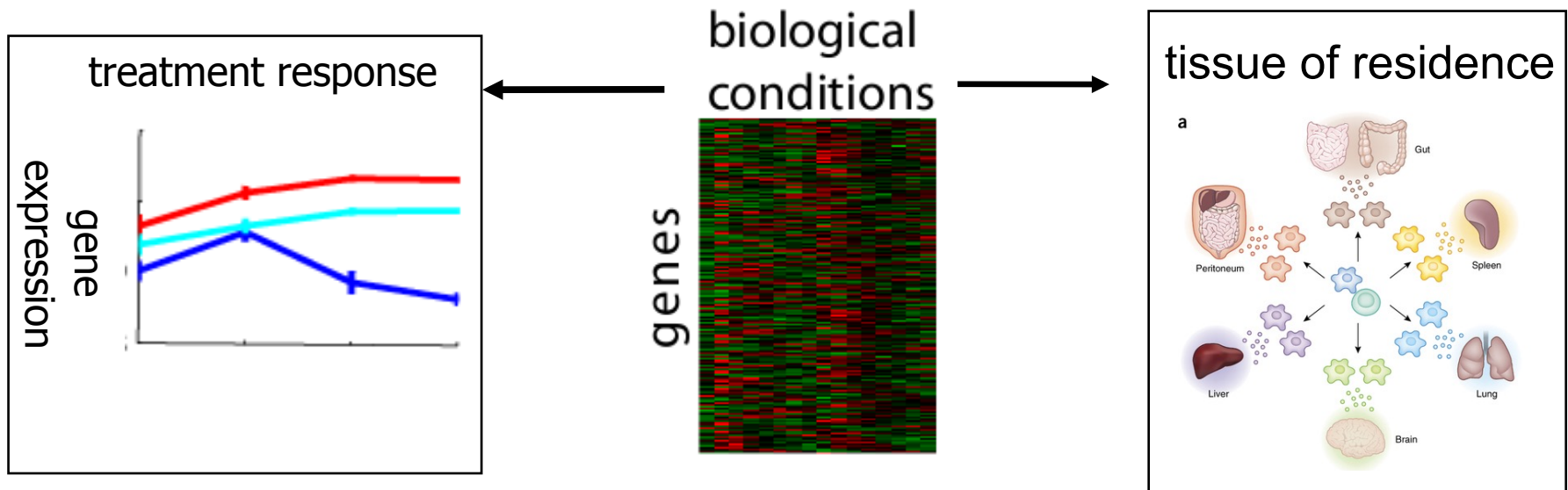
Institute for Computational Genomics  
RWTH University Hospital  
[www.costalab.org](http://www.costalab.org)

# Objective of the course

---

- 1 - Give you a overview on the use of R/bioconductor tools for gene expression analysis
- 2 - Show a real example with all steps necessary for gene expression analysis (based on arrays and RNA-seq)

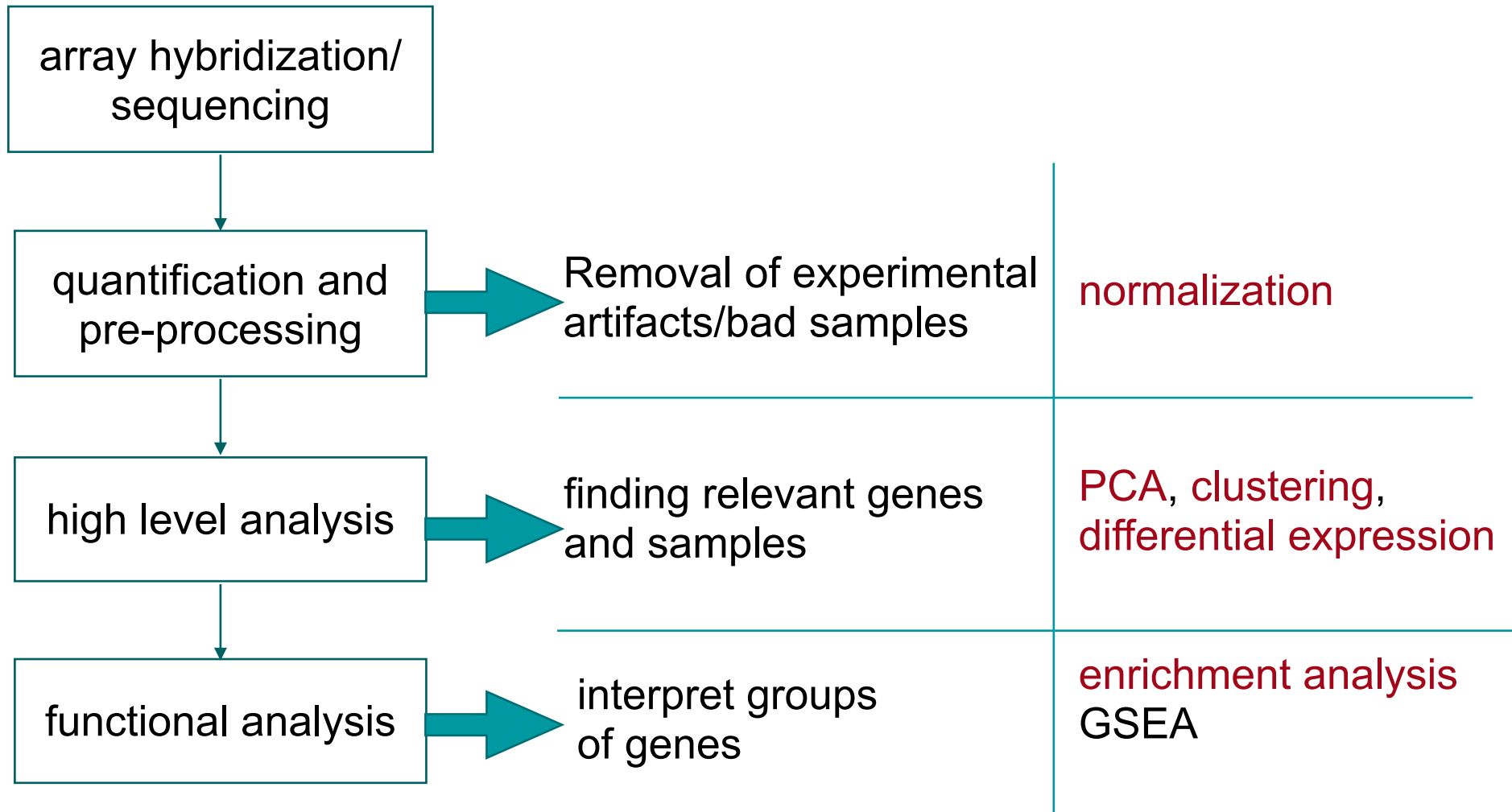
# Analysis of Gene Expression



adapted from: Amit et al. 2016

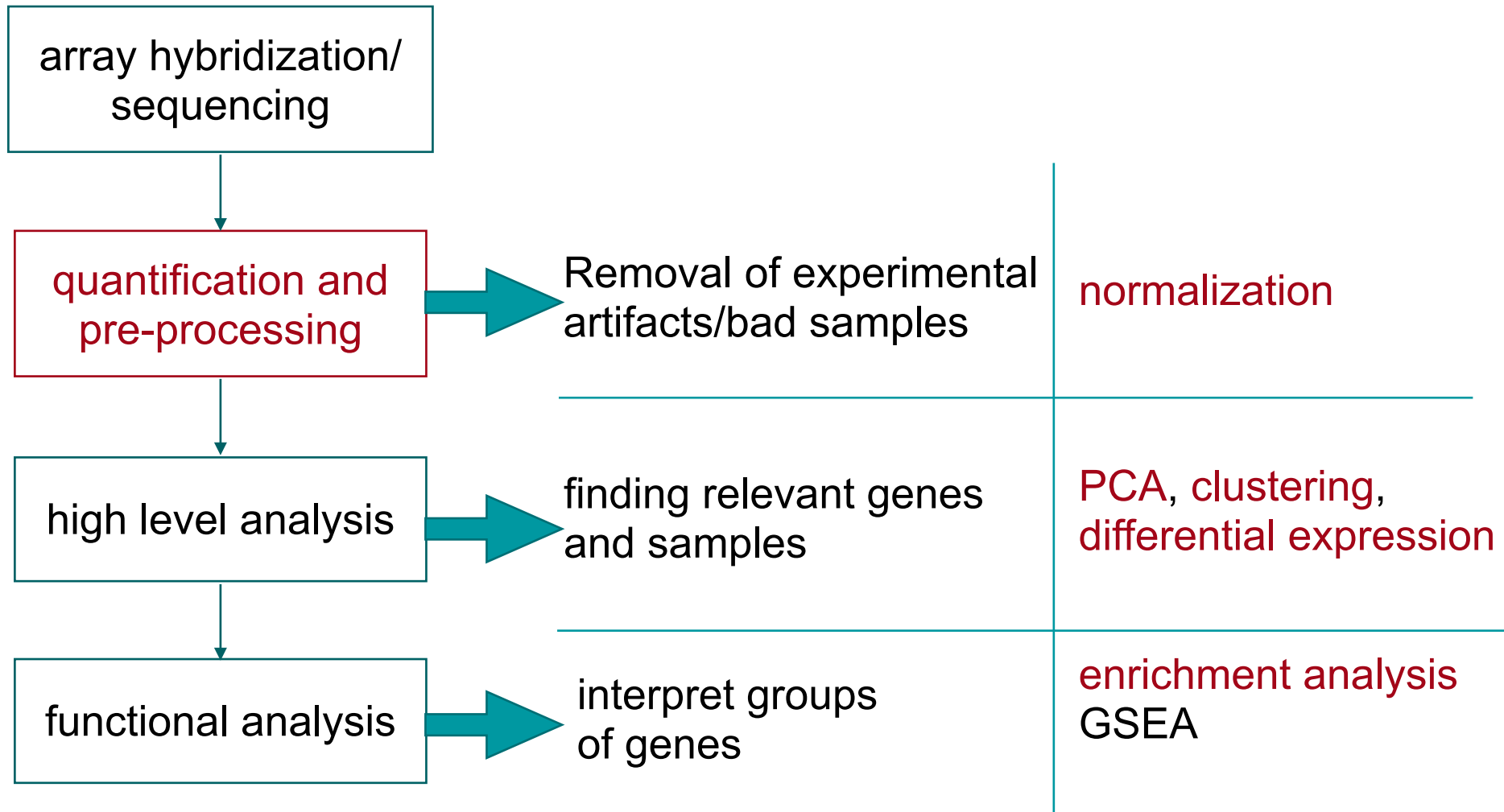
- 1- Which genes are up/down regulated after treatment?
  - differential analysis / clustering genes
- 2 - Which cells are more similar?
  - clustering samples / PCA
- 3 - How to interpret large lists of genes?
  - gene ontology enrichment /gene set enrichment analysis (GSEA)

# Bioinformatics - Gene Expression Analysis



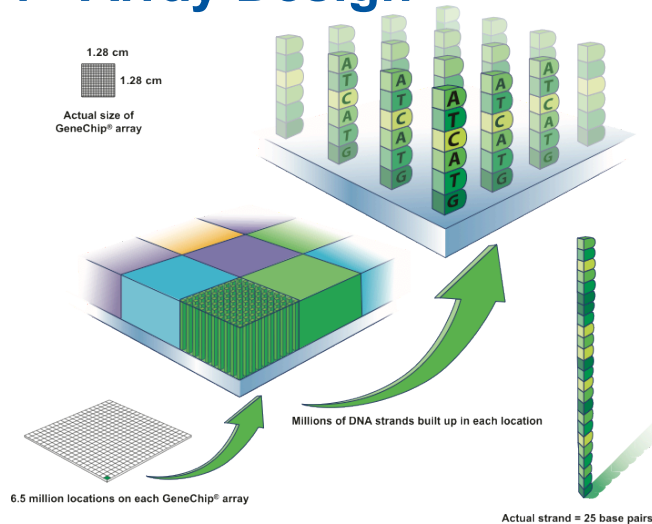


# Bioinformatics - Gene Expression Analysis

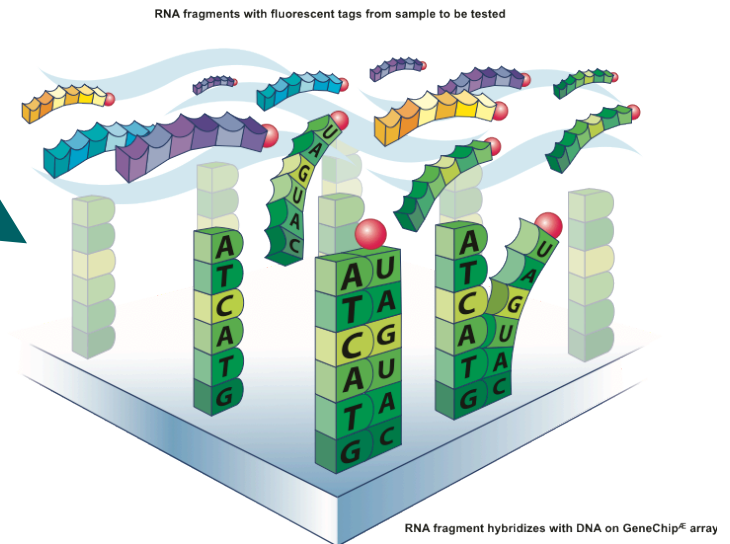


# Affymetrix Arrays - Example

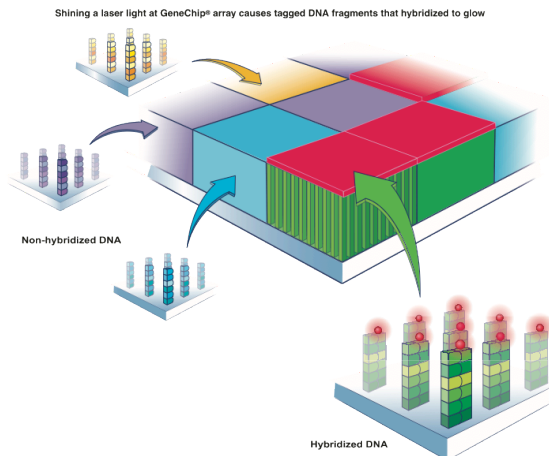
## 1 - Array Design



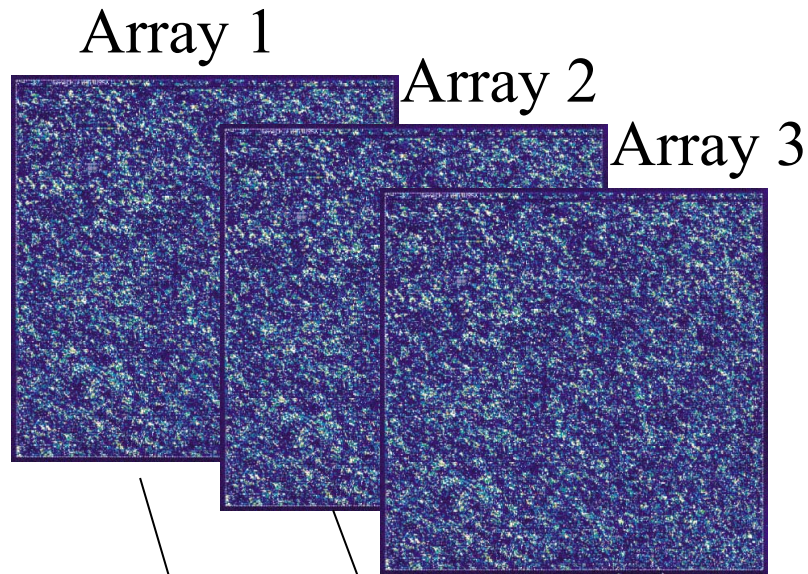
## 2 - cDNA Hybridization



## 3 -Quantification



# Quantification/Pre-processing



- 1 - Quantify gene expression values
- 2 - Quality Control
  - remove bad samples
- 3 - Correct for Experimental artifacts
  - normalization

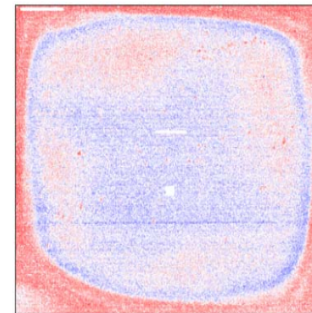
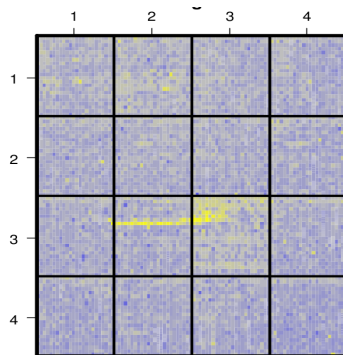
	Array 1	Array 2	Array 3
Gene 1	100	200	500
Gene 2	3000	5000	10000
Gene 3	50	10	100
...	...	...	

# Why is QC / Normalization important?

---

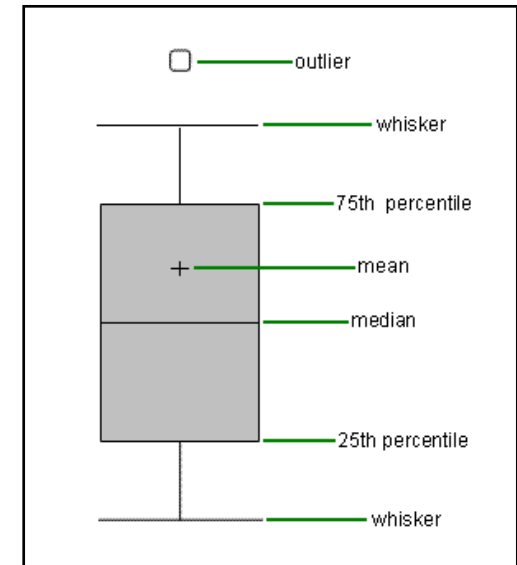
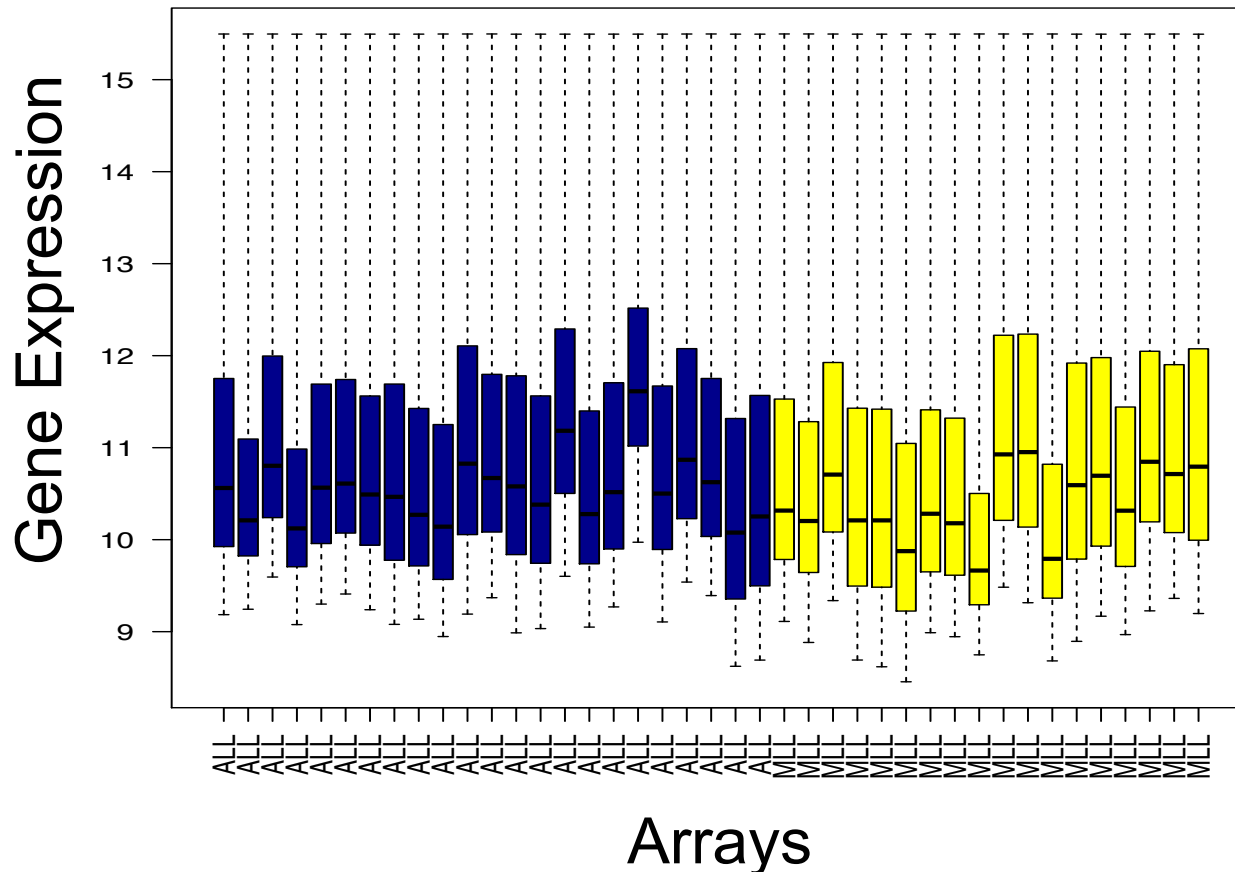
- Systematic errors (array wise)
  - labeling efficiency, scanning parameters, reverse transcriptase, batch effects
- Stochastic errors
  - cross-hybridization, image processing failure, error on probe sequence (manufacturer defect) (gene wise)
  - dust in array, hybridization problems (array wise)

## Example of Hybridization Problems



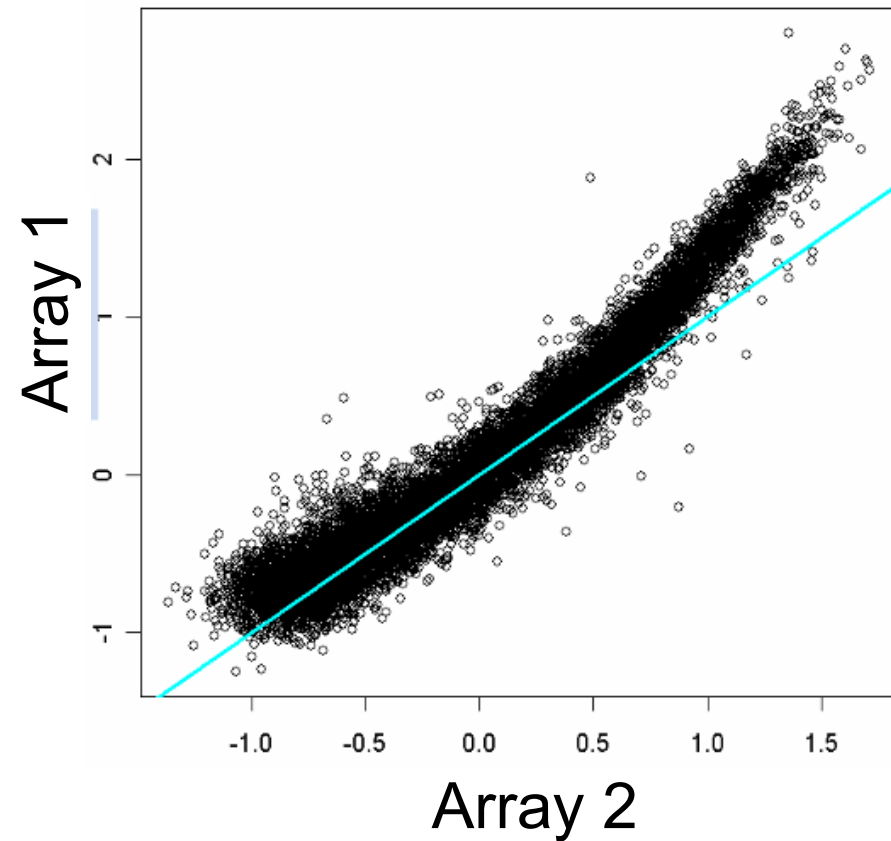
# Normalization Principles

- 1 - Most genes don't change expression -> small/same variance
- 2 - Arrays are hybridized with the same amount of DNA -> same mean

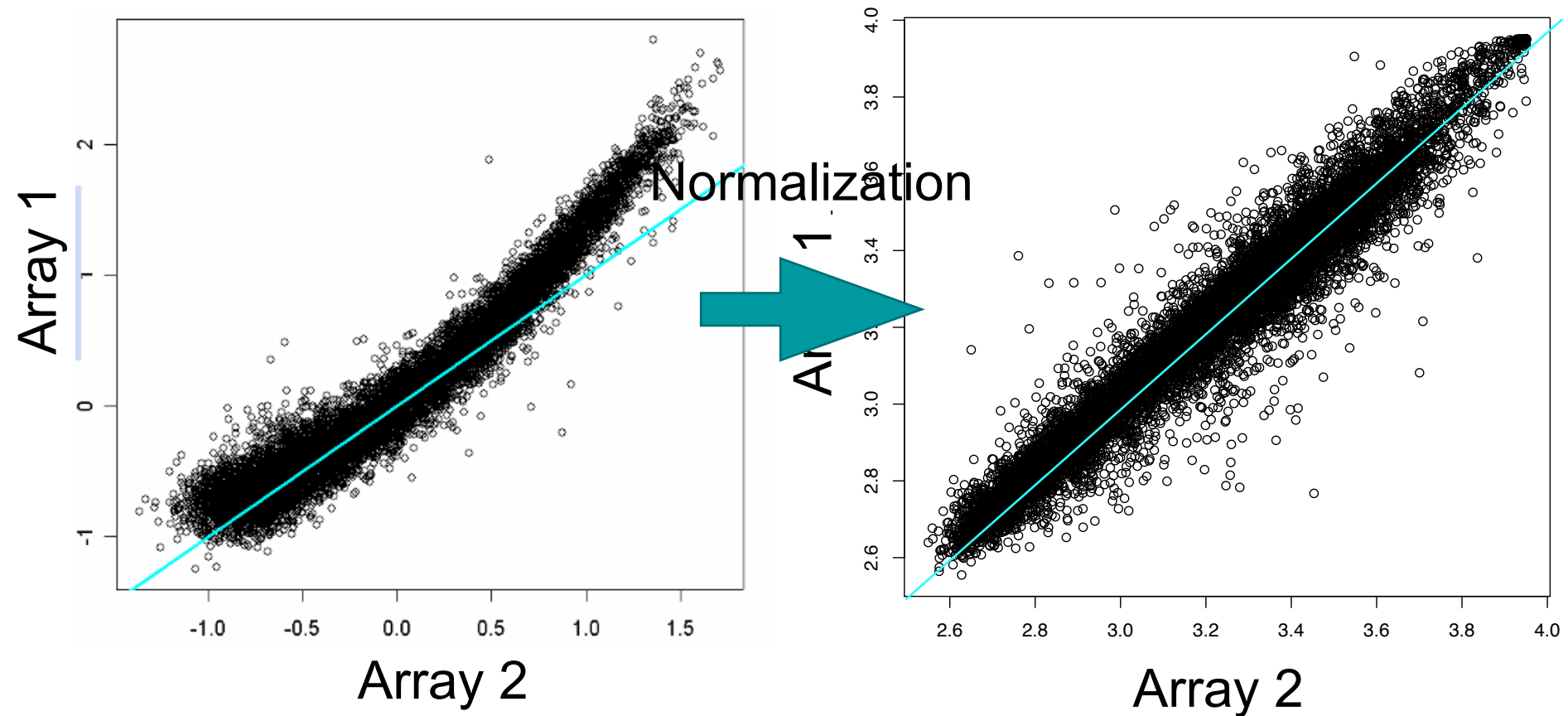


# Scatter Plots - Comparing 2 arrays

---

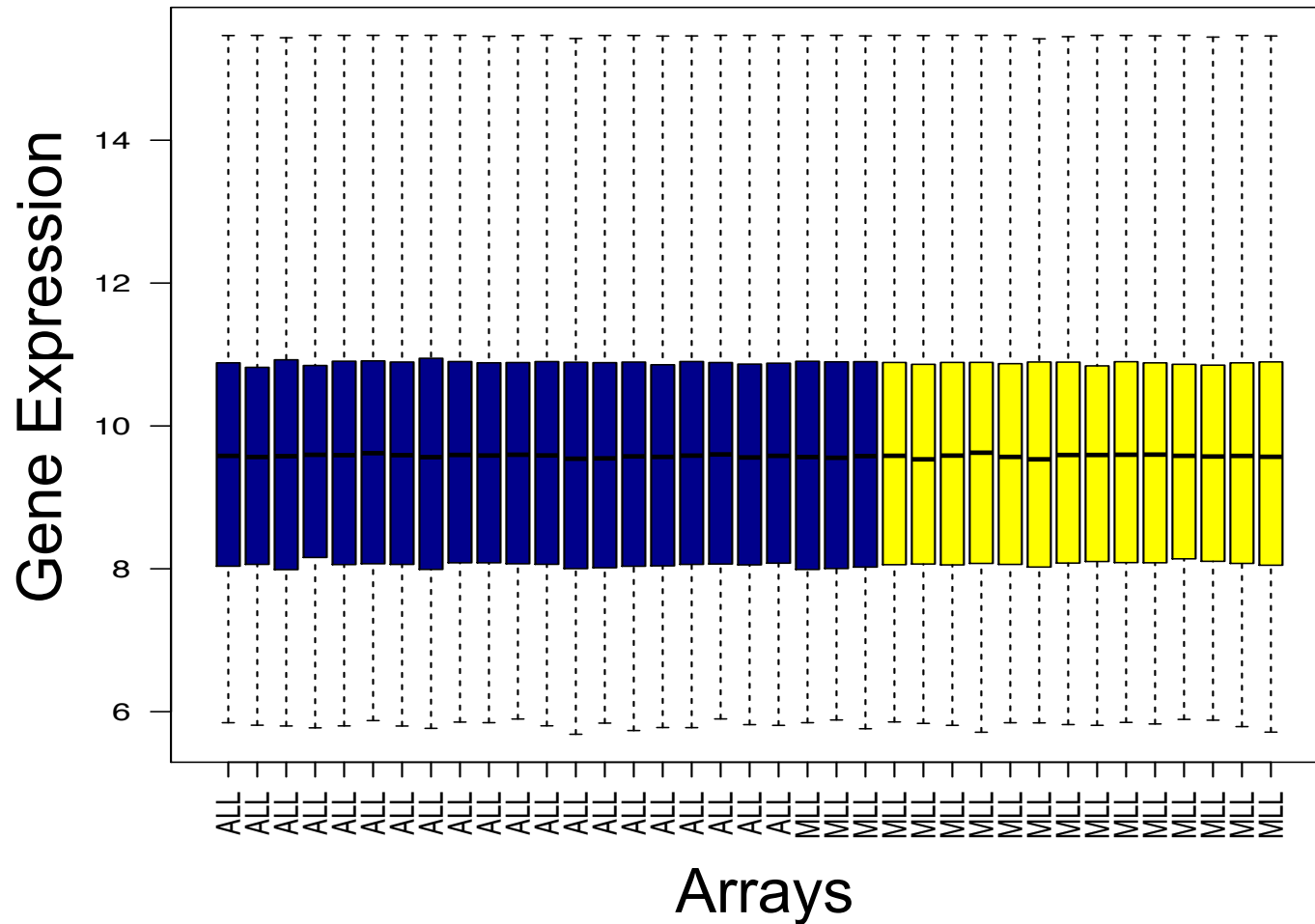


# Scatter Plots - Comparing 2 arrays



# Normalization Results

Application of BetweenArray normalization from limma package





# MA Plots

---

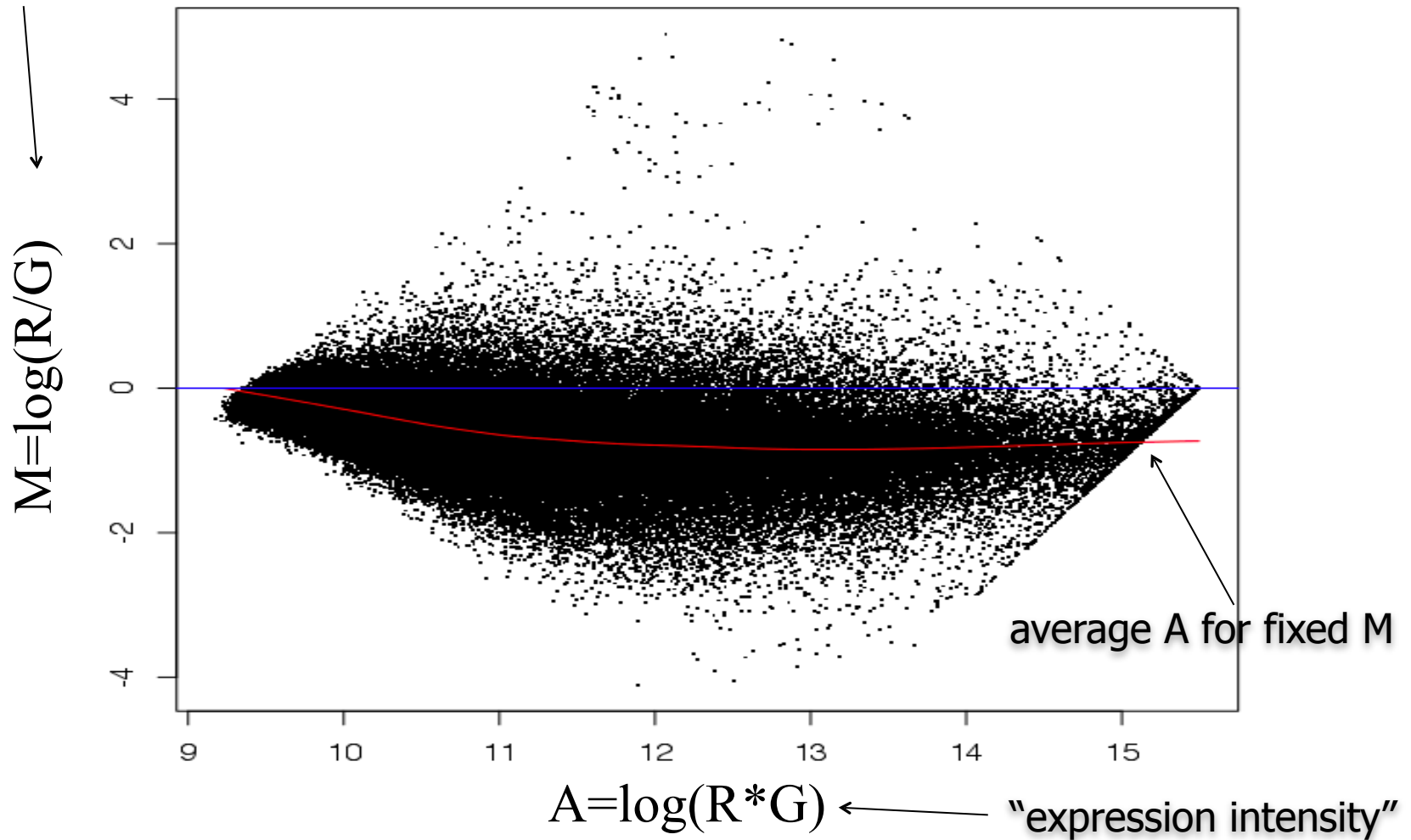
Shows systematic dependence between fluorescence intensities between arrays.

- $M = \log R/G$
- $A = \log \sqrt{R \cdot G} (= 1/2 [\log(R) + \log(G)])$

For Affymetrix/single channel arrays, R is the intensity of the microarray experiment of interest and the G is the intensity of median values of all the arrays

# MA Plots

"relative expression" (fold change)

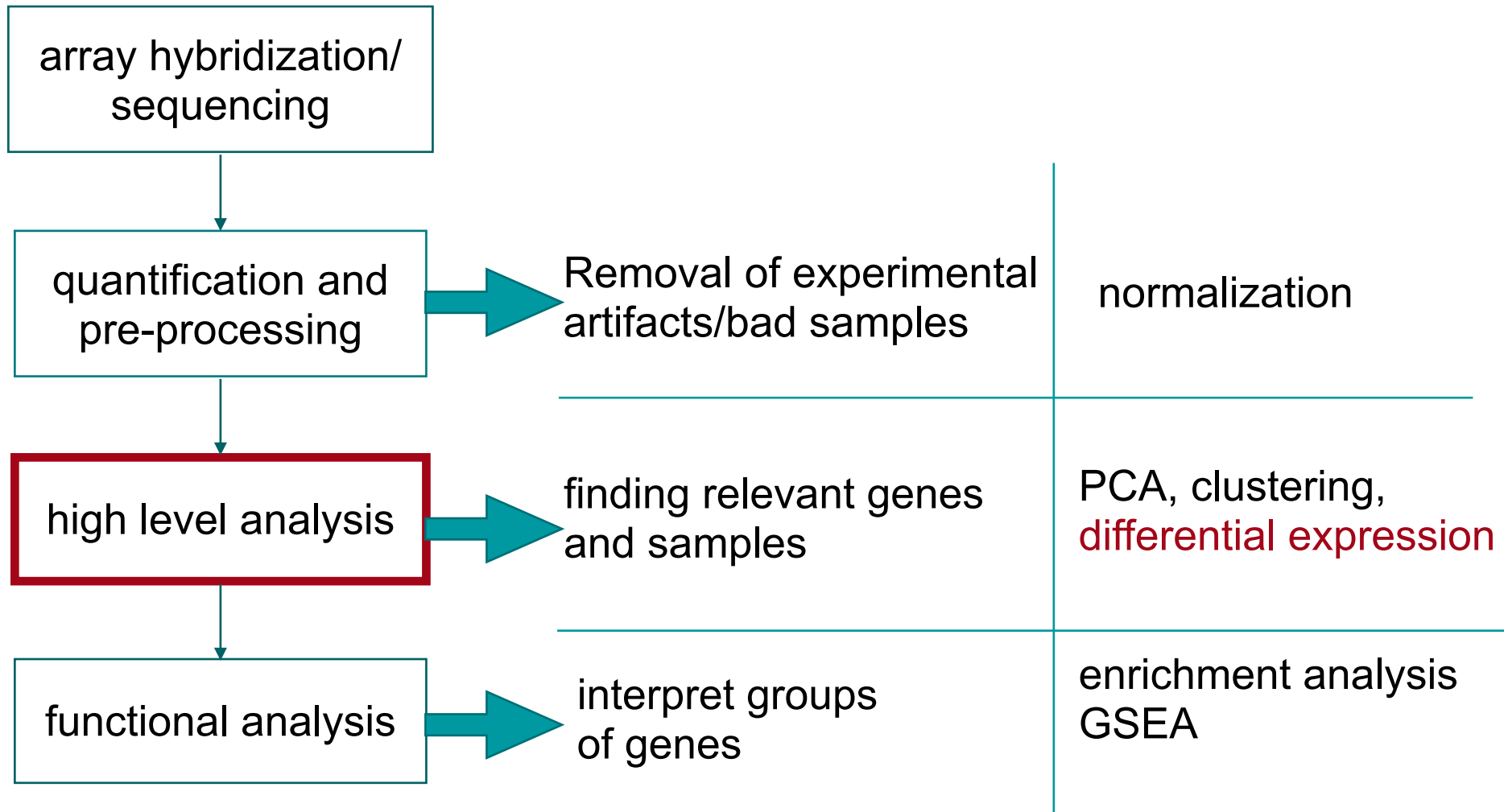


# Quantification/Pre-processing - Resume

---

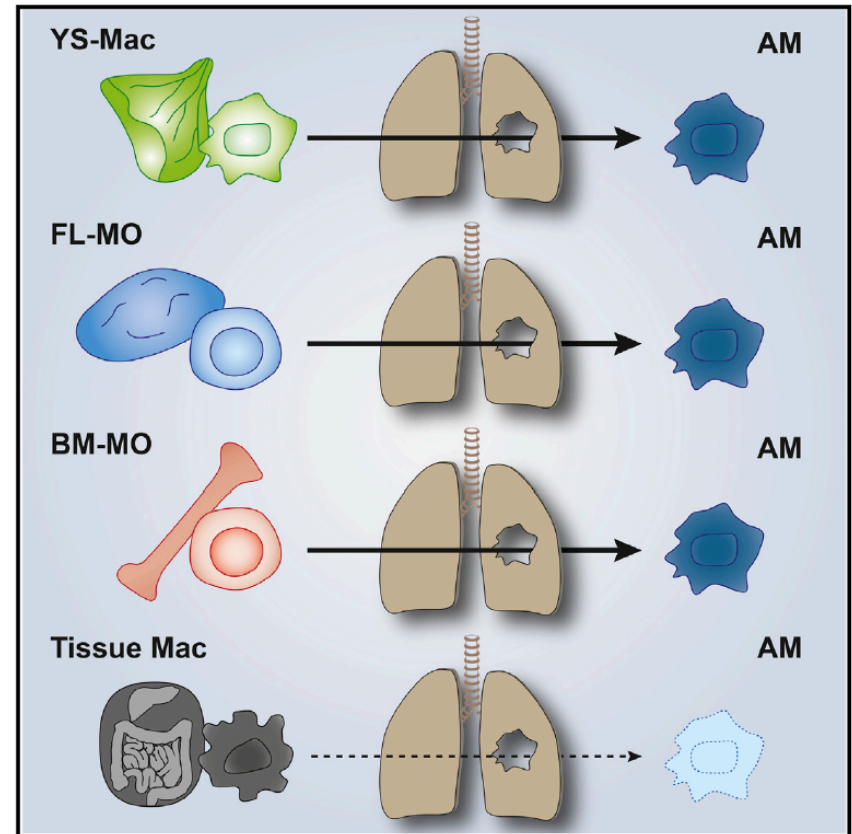
- Normalization is important to confirm the quality and consistency of data
- Boxplots should also be performed after all steps to assure data standards
- Exclusion of “bad samples” has positive effect on downstream analysis
- **In doubt, consult a bioinformatician!**

# Bioinformatics - Gene Expression Analysis



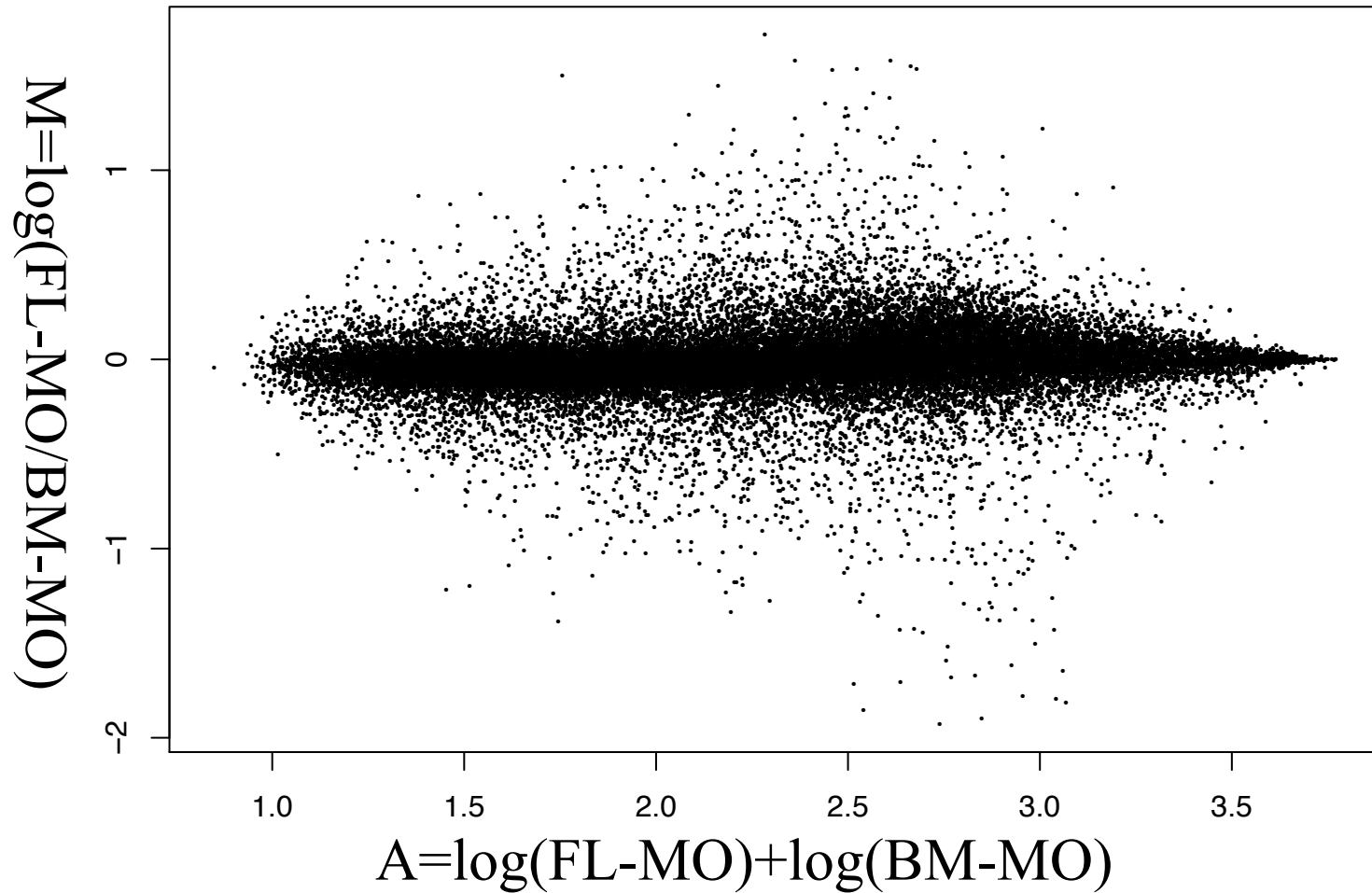
# Differential Expression Analysis

- Identify genes related to a particular condition
  - example - van de Laar, et al. 2016, Immunity, 2016.
- We will consider:
  - You Sac Macrophages (YS-Mac)
  - Fetal Liver Monocytes (FL-MO)
  - Bone Marrow Monocytes (BM-MO)
  - 4 replicates per condition



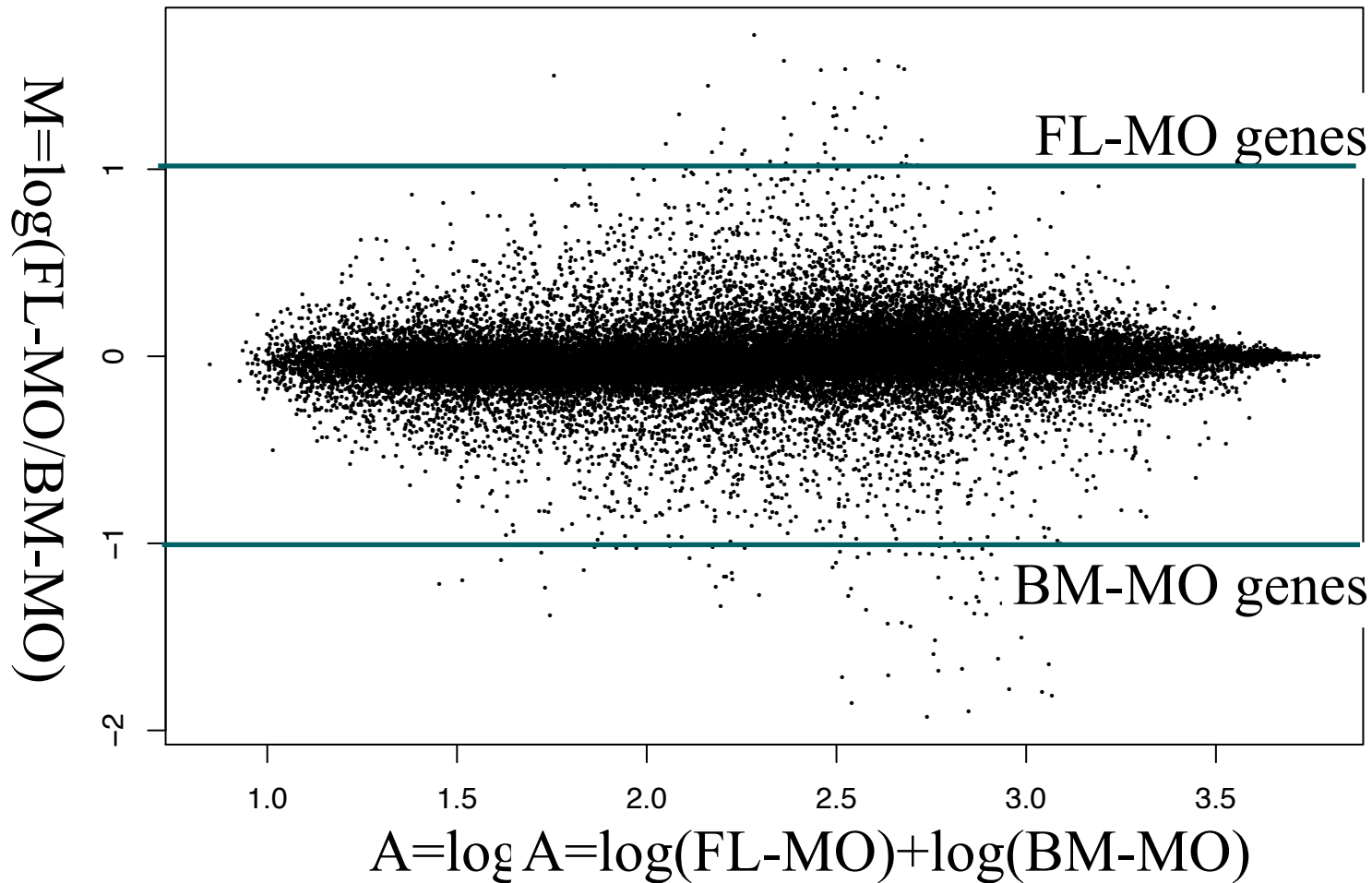
Source: van de Laar, et al. 2016, Immunity, 2016.

# Differential Expression - Example



# Differential Expression - Example

- Fold change analysis - change  $> |\log_2(2)|$



# Problems - Fold change

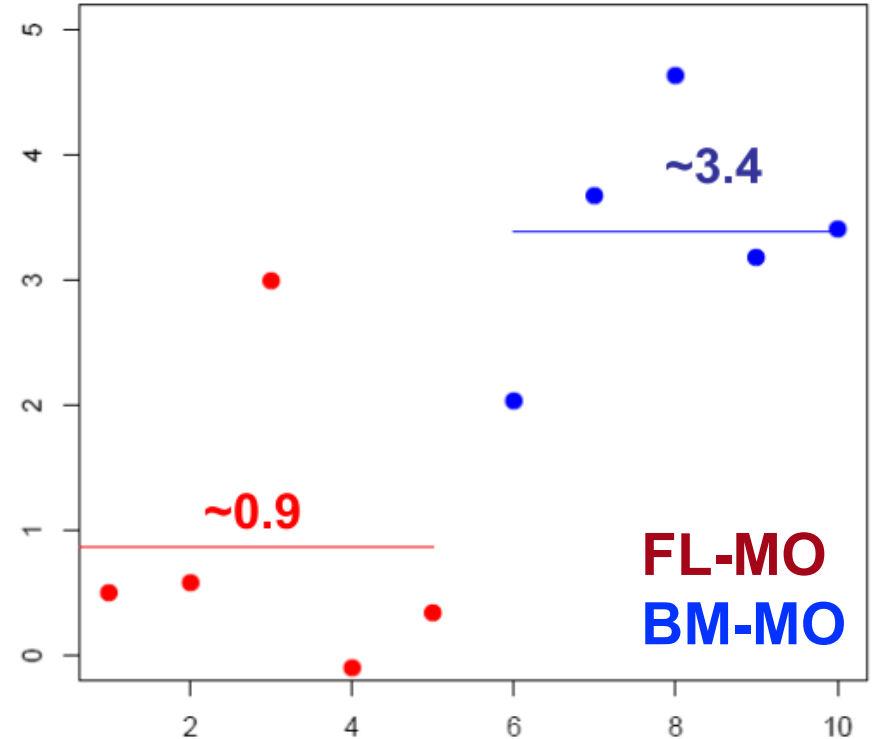
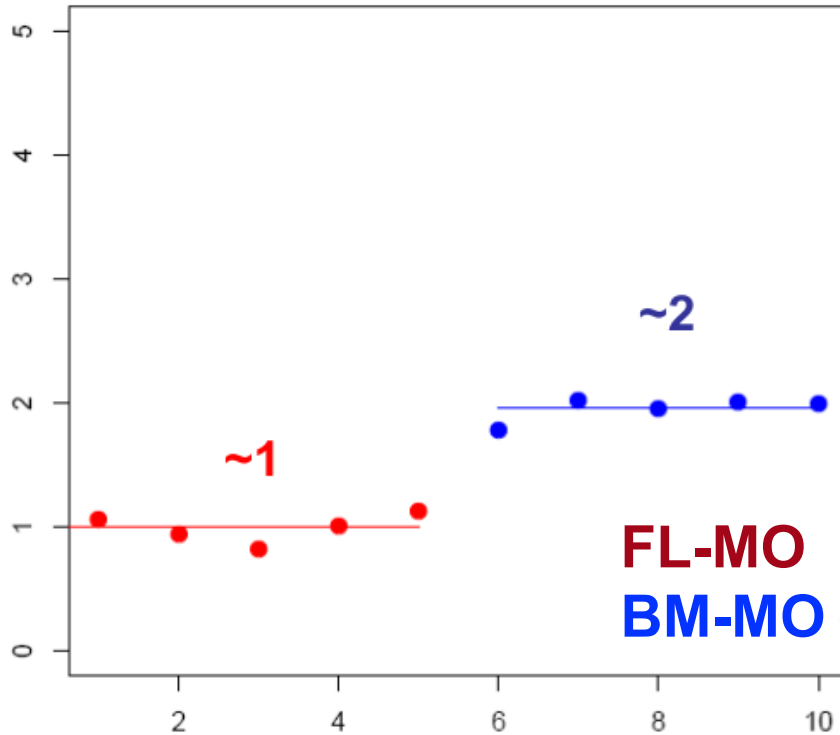
---

- Low expression genes are treated equally as high expression genes
- We lose information about the variance from genes
- No statistical significance
- Is the only alternative when no replicate samples are available (**not recommended!**)



# Basic Concepts

## Mean vs. variability



# T-test

---

We can use the t-statistic as an indication of differential expression

$$t = \frac{\bar{X} - \bar{Y}}{SE},$$

difference between means

variance

$$SE = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \quad \text{and} \quad s_X^2 = \frac{1}{n_X - 1} \sum_{i=1}^{n_X} (X_i - \bar{X})^2.$$

where  $\bar{X}$  and  $\bar{Y}$  are the mean (log) expression values of a gene in each group sample and  $n_X$  and  $n_Y$  are the number of samples on these groups

# Student T-test

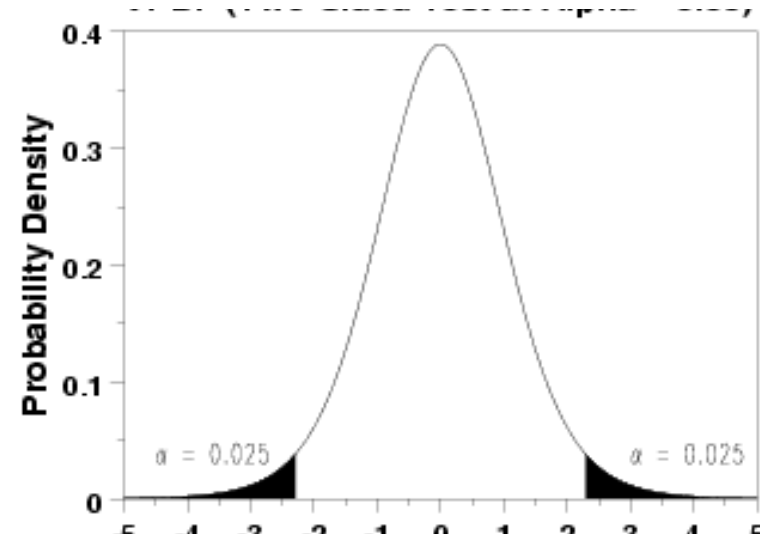
Test the hypothesis  $H_0 : X - Y = 0$

$$H_1 : X - Y \neq 0$$

We can use the t-student distribution to estimate for which t-statistic values the null hypothesis is rejected.

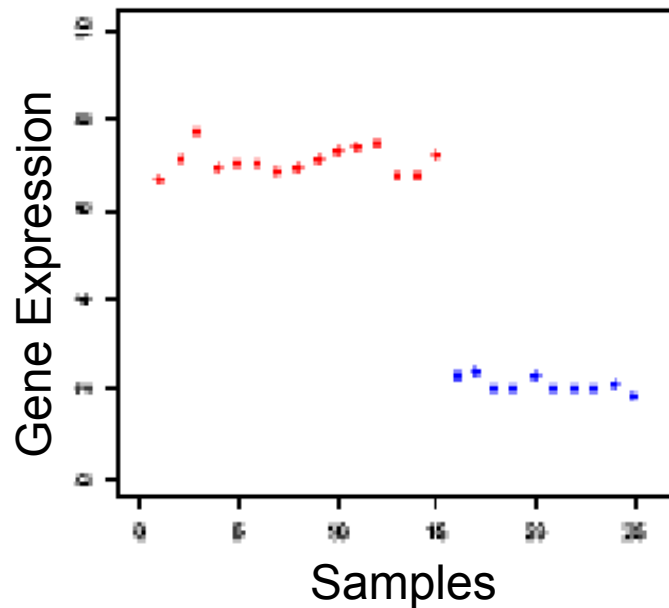
$P\text{-value} = \Pr(t \text{ as extreme or more} | H_0),$

t student pdf – p-value = 0.05



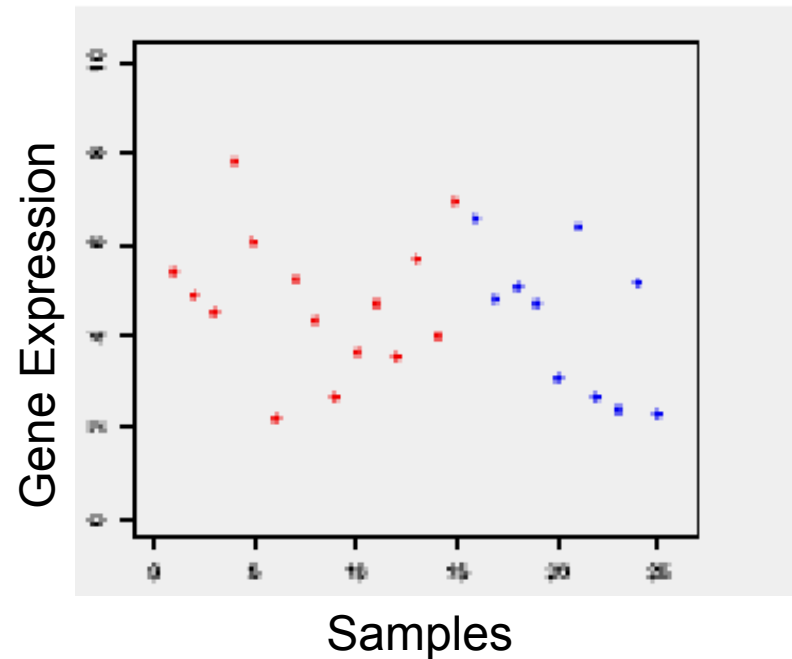
# Examples

**Change: HIGH**  
**Variance: SMALL**



**T huge**

**Change: SMALL**  
**Variance: HIGH**



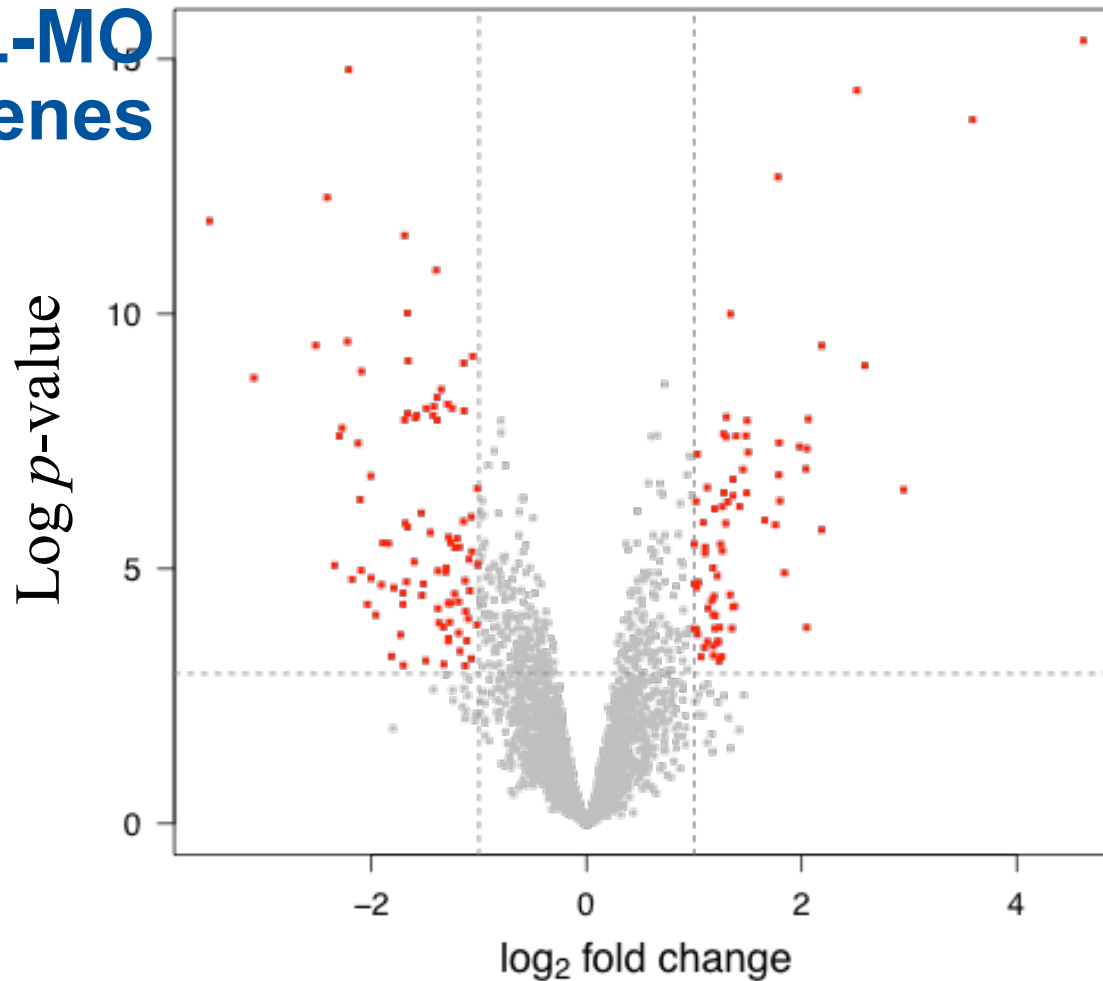
**T ~ 0**

# Results - FL-MO vs. BM-MO

Volcano Plot - combine p-value and fold change

**FL-MO  
genes**

**BM-MO  
genes**



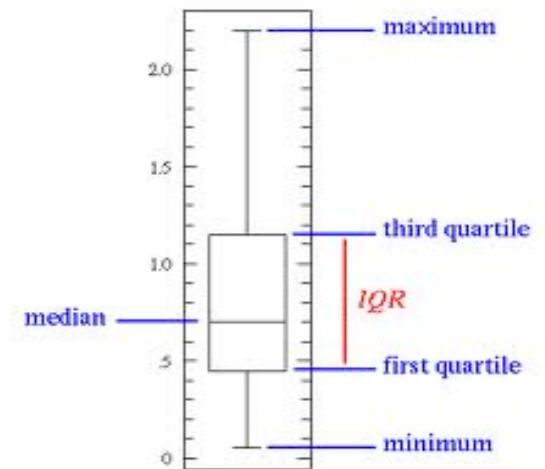
# Multiple Test Correction

---

- With a p-value of 0.01, we expect to make one mistake every 100 tests
- We have 12.626 genes, therefore 126 mistaken from 1046 DE genes.
- To solve this, a multiple test correction method is necessary (i.e. Benjamini-Hochberg)
  - It is based on the false discovery rate, i.e. the proportion of false DE genes in your list of DE genes

# Filtering

- Higher level analysis are eased by filtering of non-specific genes
  - genes that show no expression changes between arrays
  - i.e. filter genes with low IQR (interquantile range)
- Affymetrix chips has spike-in control probes
  - Should be removed after normalization



# Differential Analysis - Conclusions

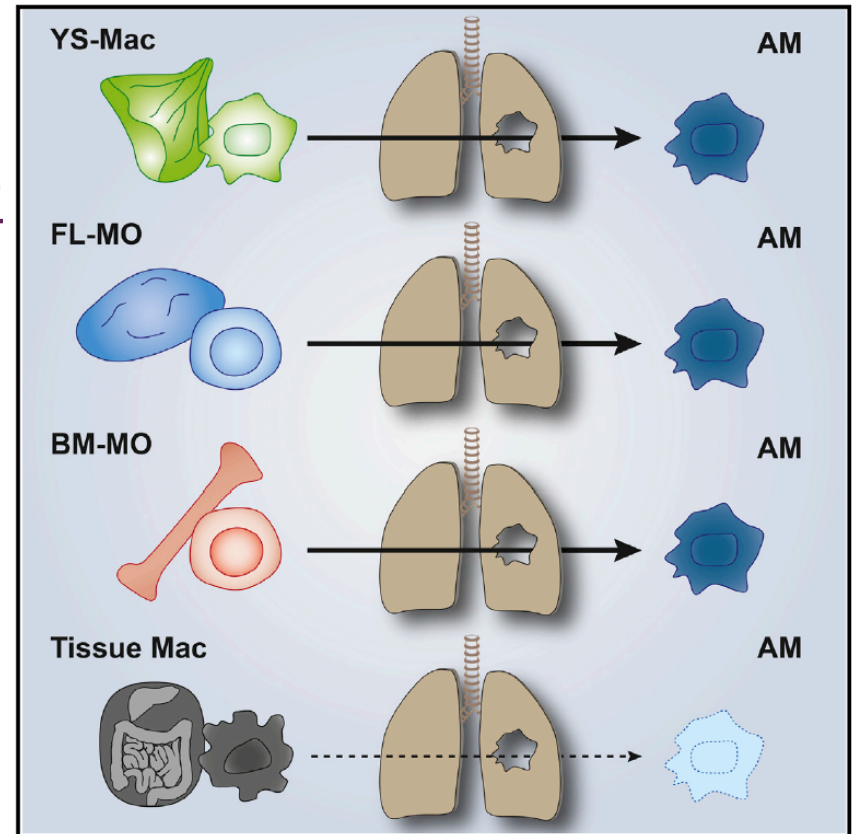
---

- Fold-change (alone) -> should be avoided
- For patient samples
  - high number of replicates are necessary ( $>30$ )
  - otherwise - low DE genes replicability
- For model (mouse) experiments
  - at least 3 samples (and moderated t-test)
  - we can not tell the variance without measuring it!
- All correct for multiple testing! Also, non-specific filtering can help if low number of DE genes is found.



# Practical Example

- This data is deposited in the public repository GEO under accession [GSE76999](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76999)
- This can be found at the materials and methods of papers.
- GEO - public database with raw, pre-processed data and experimental details of expression (and other omics) experiments.



Source: van de Laar, et al. 2016, Immunity, 2016.

# GEO - van de Laar, et al. 2016

Series GSE76999

Query DataSets for GSE76999

GEO ID

Status	Public on Mar 01, 2016
Title	Capacity of yolk sac macrophages, fetal liver and adult monocytes to colonize an empty niche and develop into functional tissue resident macrophages
Organism	<a href="#">Mus musculus</a>
Experiment type	Expression profiling by array
Summary	<p>Tissue-resident macrophages can derive from yolk sac macrophages, fetal liver monocytes or adult bone marrow monocytes. Whether these precursors can give rise to transcriptionally identical alveolar macrophages is unknown. Here, we transferred traceable yolk sac macrophages, fetal liver monocytes, adult bone marrow monocytes or adult alveolar macrophages as a control, into the empty alveolar macrophage niche of neonatal <i>Csf2rb</i><sup>-/-</sup> mice. All precursors efficiently colonized the alveolar niche and generated alveolar macrophages that were transcriptionally almost identical, with only 22 genes that could be linked to their origin. Underlining the physiological relevance of our findings, all transfer-derived alveolar macrophages self-maintained within the lungs for up to 1 year and durably prevented alveolar proteinosis. Thus, precursor origin does not affect the development of functional self-maintaining tissue-resident macrophages.</p>
Overall design	<p>CD45.1+CD45.2+ yolk sac macrophages, fetal liver monocytes, adult bone marrow monocytes or adult alveolar macrophages from the bronchoalveolar lavage were sorted from wild type CD45.1+CD45.2+ mice of indicated ages. From part of these samples RNA was isolated. The other part was transferred intranasally into the lungs of neonate <i>Csf2rb</i><sup>-/-</sup> mice. 6 weeks post-transfer, transfer-derived CD45.1+CD45.2+ alveolar macrophages were sorted from the bronchoalveolar lavage. Wild type CD45.1+CD45.2 alveolar macrophages from the bronchoalveolar lavage of 6 week old mice were sorted as control. 36 samples (arrays) in total. RNA was isolated, amplified with Nugene pico kit, converted to cDNA and then hybridised on Affymetrix GeneChip Mouse Gene 1.0 ST Arrays.</p>
Contributor(s)	<a href="#">van de Laar L</a> , <a href="#">Saelens W</a> , <a href="#">De Prijck S</a> , <a href="#">Martens L</a> , <a href="#">Scott CL</a> , <a href="#">Van Isterdael G</a> , <a href="#">Hoffmann E</a> , <a href="#">Beyaert R</a> , <a href="#">Saeys Y</a> , <a href="#">Lambrecht BN</a> , <a href="#">Guilliams M</a>
Citation(s)	<p>van de Laar L, Saelens W, De Prijck S, Martens L et al. Yolk Sac Macrophages, Fetal Liver, and Adult Monocytes Can Colonize an Empty Niche and Develop into Functional Tissue-Resident Macrophages. <i>Immunity</i> 2016 Apr 19;44(4):755-68. PMID: <a href="#">26992565</a></p>

Information  
about the study

# GEO - van de Laar, et al. 2016

Submission date Jan 20, 2016  
Last update date Jul 13, 2018  
Contact name Martin Guilliams  
Organization name VIB-University of Ghent  
Department VIB Inflammation Research Center  
Street address Technologiepark 927  
City Ghent  
ZIP/Postal code 9000  
Country Belgium

Platforms (1) [GPL6246](#) [MoGene-1\_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]

Samples (36) [GSM2042244](#) Monocyte extracted from adult (wk6-12) Bone Marrow, biological replicate 1  
[More...](#)

[GSM2042245](#) Monocyte extracted from adult (wk6-12) Bone Marrow, biological replicate 2

[GSM2042246](#) Monocyte extracted from adult (wk6-12) Bone Marrow, biological replicate 3

## Relations

BioProject [PRJNA309234](#)

Analyze with GEO2R

## Download family

[SOFT formatted family file\(s\)](#)

[MINiML formatted family file\(s\)](#)

[Series Matrix File\(s\)](#)

## Format

SOFT [?](#)

MINiML [?](#)

TXT [?](#)

Supplementary file	Size	Download	File type/resource
GSE76999_RAW.tar	135.3 Mb	<a href="#">(http)(custom)</a>	TAR (of CEL)

Raw data provided as supplementary file

array used

single  
experiments

raw data



# GEO - van de Laar, et al. 2016

**Sample GSM2042244**

[Query DataSets for GSM2042244](#)

Status	Public on Mar 01, 2016
Title	Monocyte extracted from adult (wk6-12) Bone Marrow, biological replicate 1
Sample type	RNA
Source name	Monocyte, extracted from Bone Marrow (BM)
Organism	<a href="#">Mus musculus</a>
Characteristics	strain: C57BL/6 tissue: Bone Marrow age: wk6-12
Treatment protocol	not applicable
Growth protocol	Tissues were isolated from the mice at the indicated ages.
Extracted molecule	total RNA
Extraction protocol	Single cell suspensions were prepared by organ digestion (yolk sac and fetal liver) with 1 mg/ml collagenase A and 10 U/ml DNA (30 and 5 minutes at 37oC), crushing (bones) or flushing of the lungs (broncholaveolar lavage). 2x10 <sup>4</sup> cells were FACS purified into RLT buffer (Qiagen) containing 10 ml/ml 2-mercaptoethanol. RNA was isolated using the RNA isolation kit micro (Qiagen no74034).
Label	biotin
Label protocol	Affymetrix WT Terminal Labeling Kit
Hybridization protocol	Standard Affymetrix protocol. cDNA was hybrised on Affymetrix GeneChip Mouse Gene 1.0 ST Arrays (GPL6246).
Scan protocol	Affymetrix Gene ChIP Scanner 3000 7G
Description	Monocyte extracted from Bone Marrow
Data processing	Data were processed using Bioconductor. Normalisation was done by RMA. MoGene-1_0-st-v1.r4.pgf MoGene-1_0-st-v1.r4.mps

ID of array

name of condition

details

---

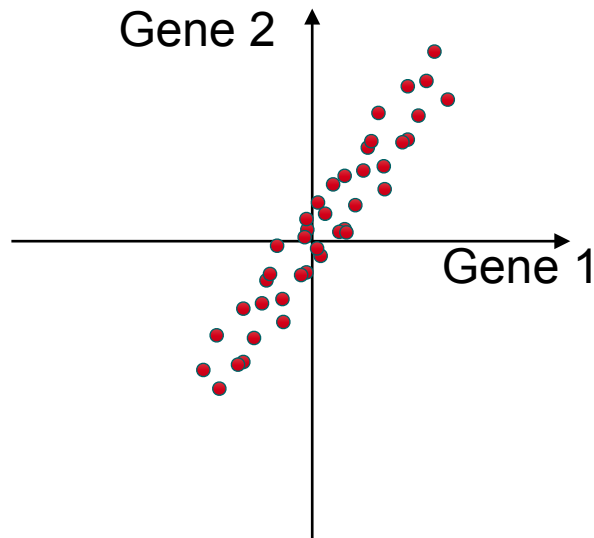
# Hands on!

Handout Steps 1,2 and 3

# Principal Component Analysis

---

- **method for dimension reduction**
  - **find combination of genes explaining cells with distinct expression**
- **finding directions with highest variance**

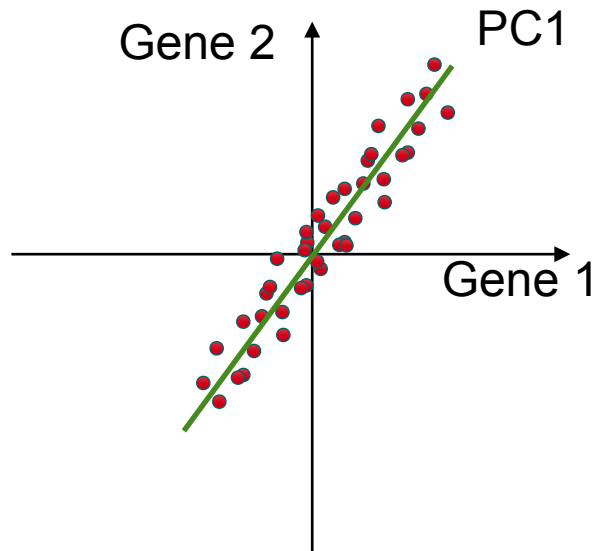


Recommended reading:  
Ringner M., *Nature Biotechnology* 26, 303 - 304 (2008)

# Principal Component Analysis

---

- **method for dimension reduction**
  - **find combination of genes explaining cells with distinct expression**
- **finding directions with highest variance**

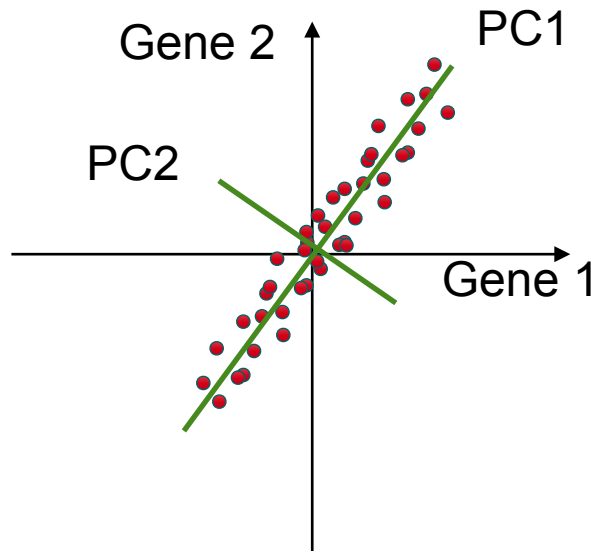


Recommended reading:  
Ringner M., *Nature Biotechnology* 26, 303 - 304 (2008)

# Principal Component Analysis

---

- **method for dimension reduction**
  - **find combination of genes explaining cells with distinct expression**
- **finding directions with highest variance**

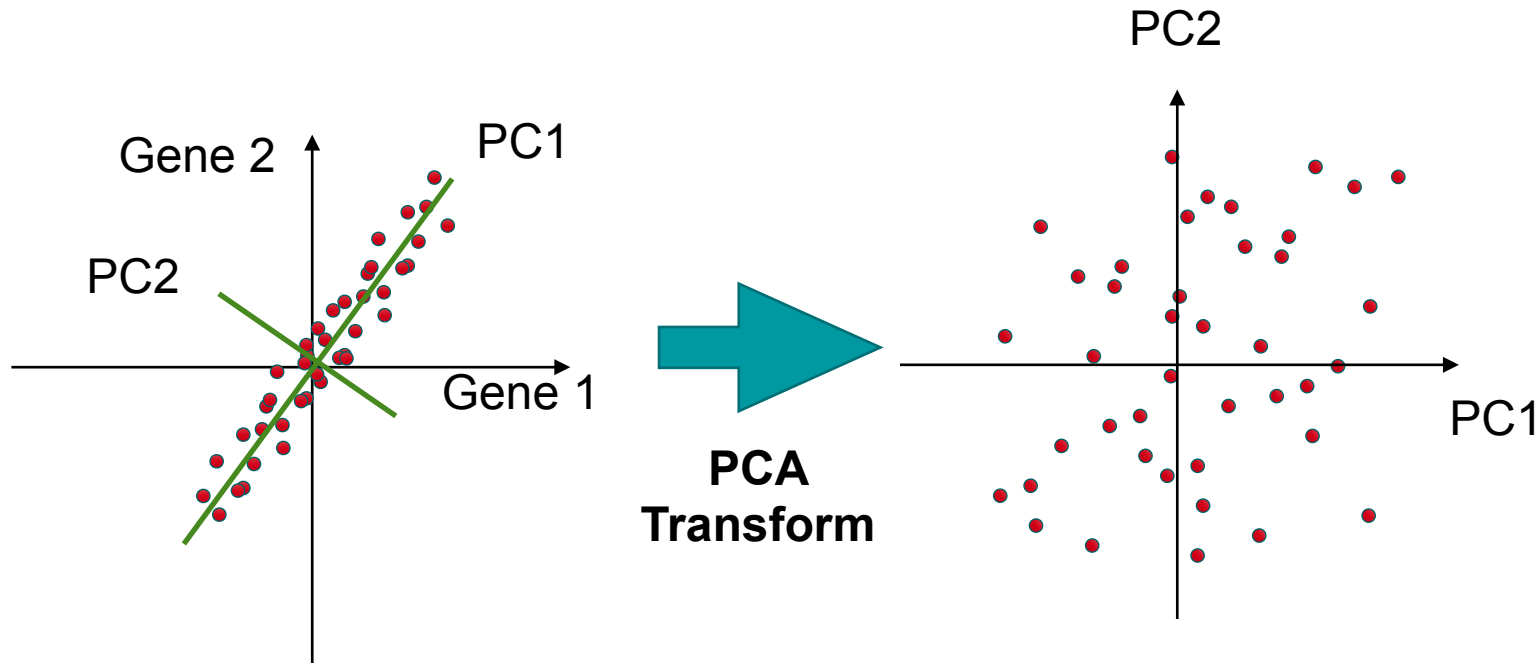


Recommended reading:  
Ringner M., *Nature Biotechnology* 26, 303 - 304 (2008)



# Principal Component Analysis

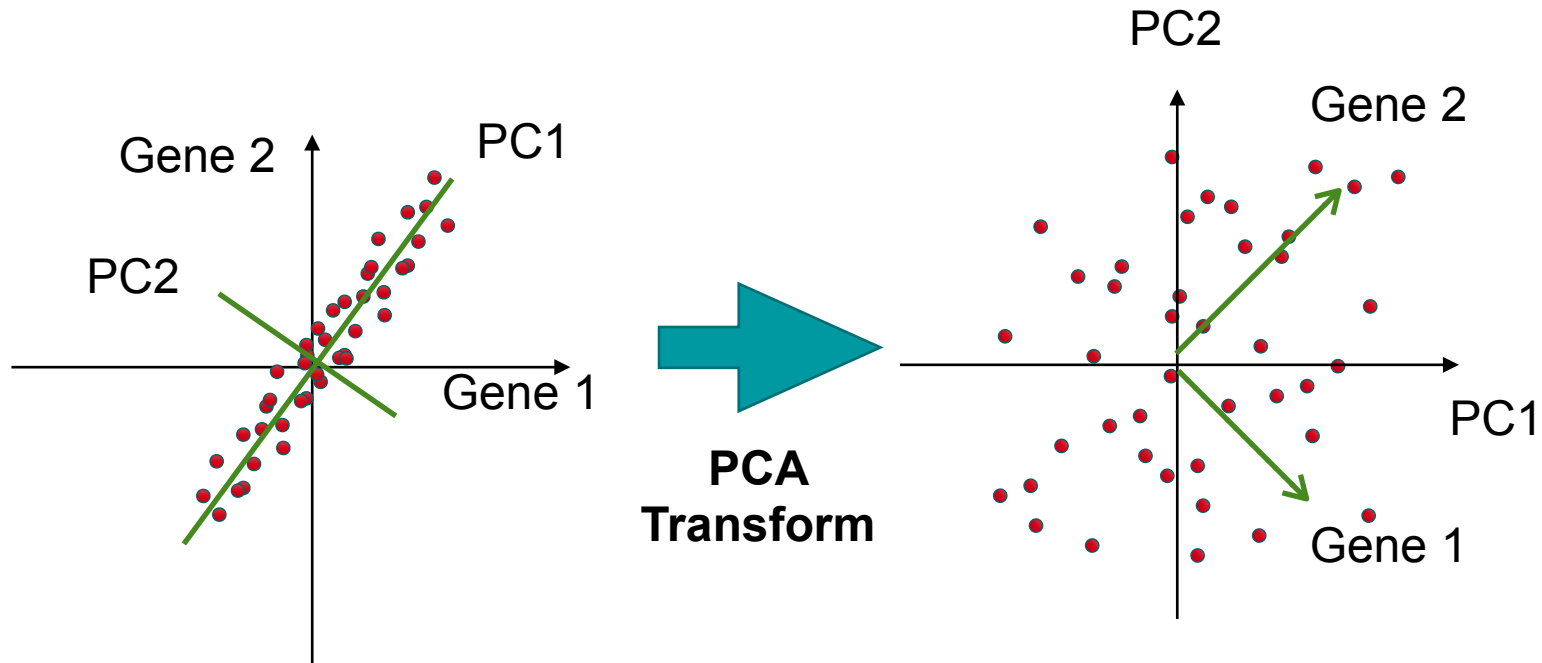
- **method for dimension reduction**
  - find combination of genes explaining cells with distinct expression
- **finding directions with highest variance**



Recommended reading:  
Ringner M., *Nature Biotechnology* 26, 303 - 304 (2008)

# Principal Component Analysis

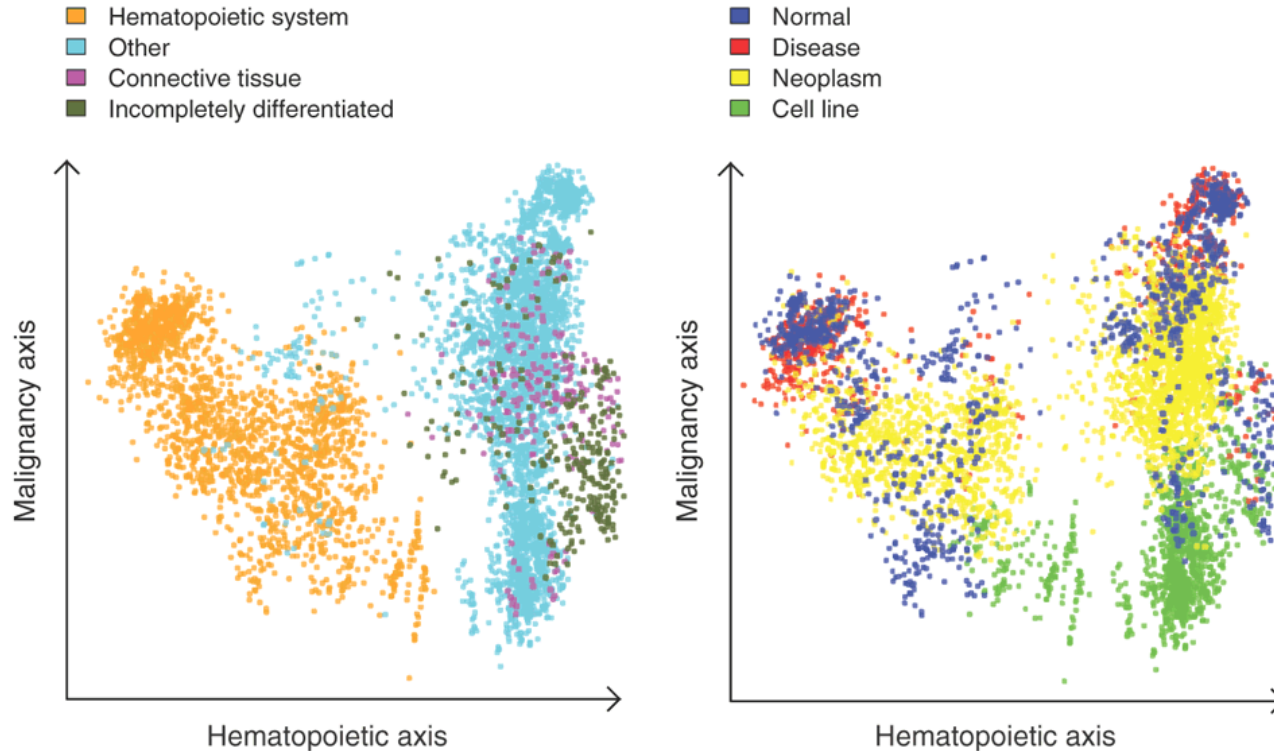
- **method for dimension reduction**
  - find combination of genes explaining cells with distinct expression
- **finding directions with highest variance**



Recommended reading:  
Ringner M., *Nature Biotechnology* 26, 303 - 304 (2008)

# Gene Expression - PCA Example 1

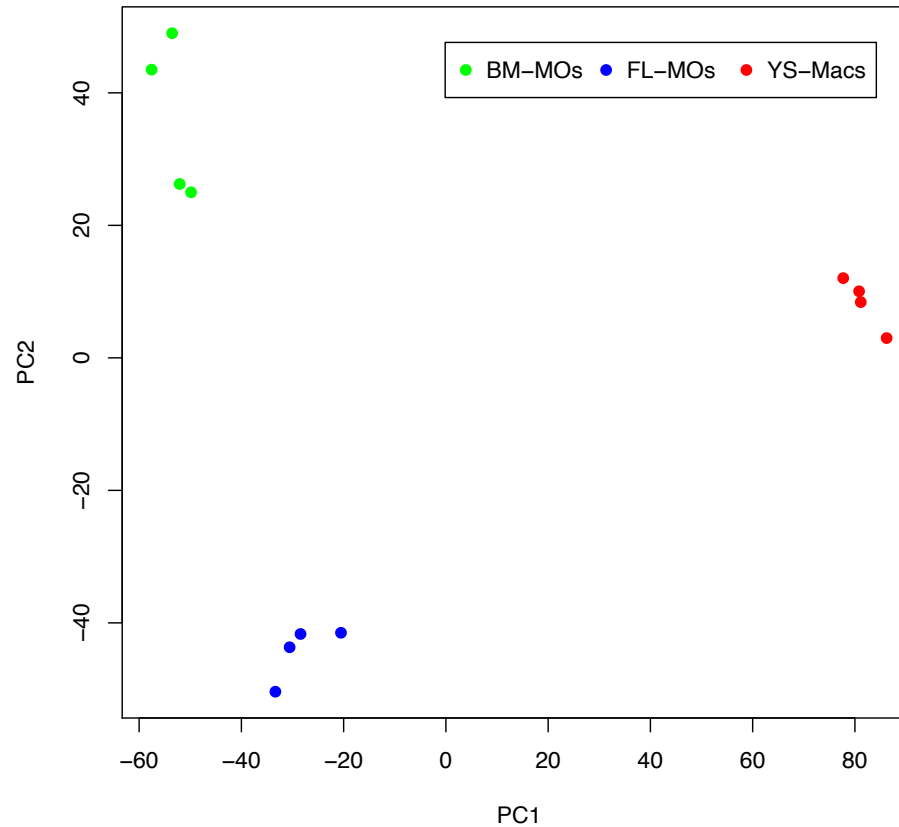
Can be interpreted as a computational FACS sorting (without knowing the markers)



First 2 PCs on the analysis of 5000 samples from Array Express/EBI

# Gene Expression - PCA Example 2

## PCA Analysis of van de Leer, 2016 data



First 2 PCs van de Leer, 2016 data

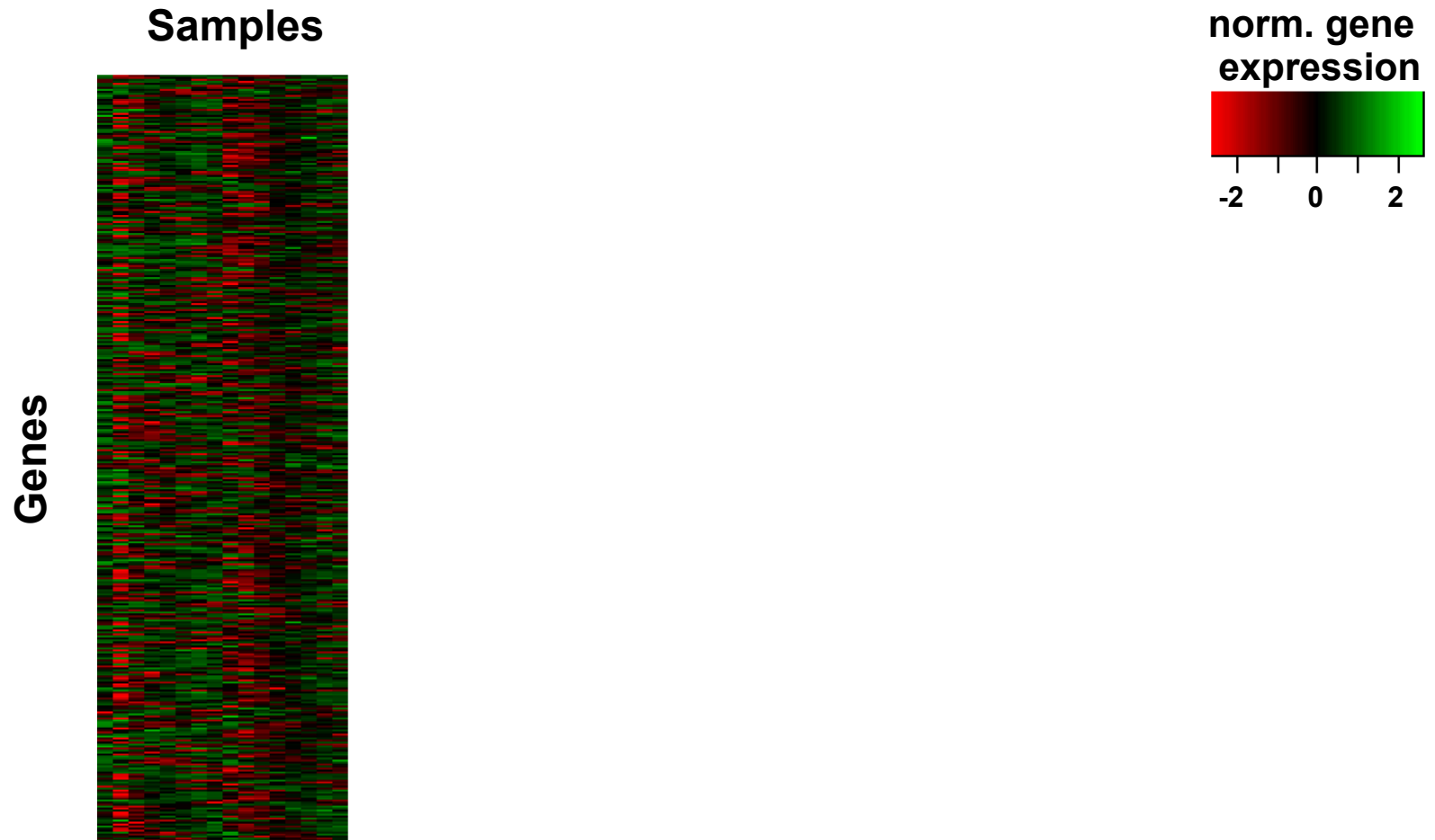
# PCA Analysis - Conclusions

---

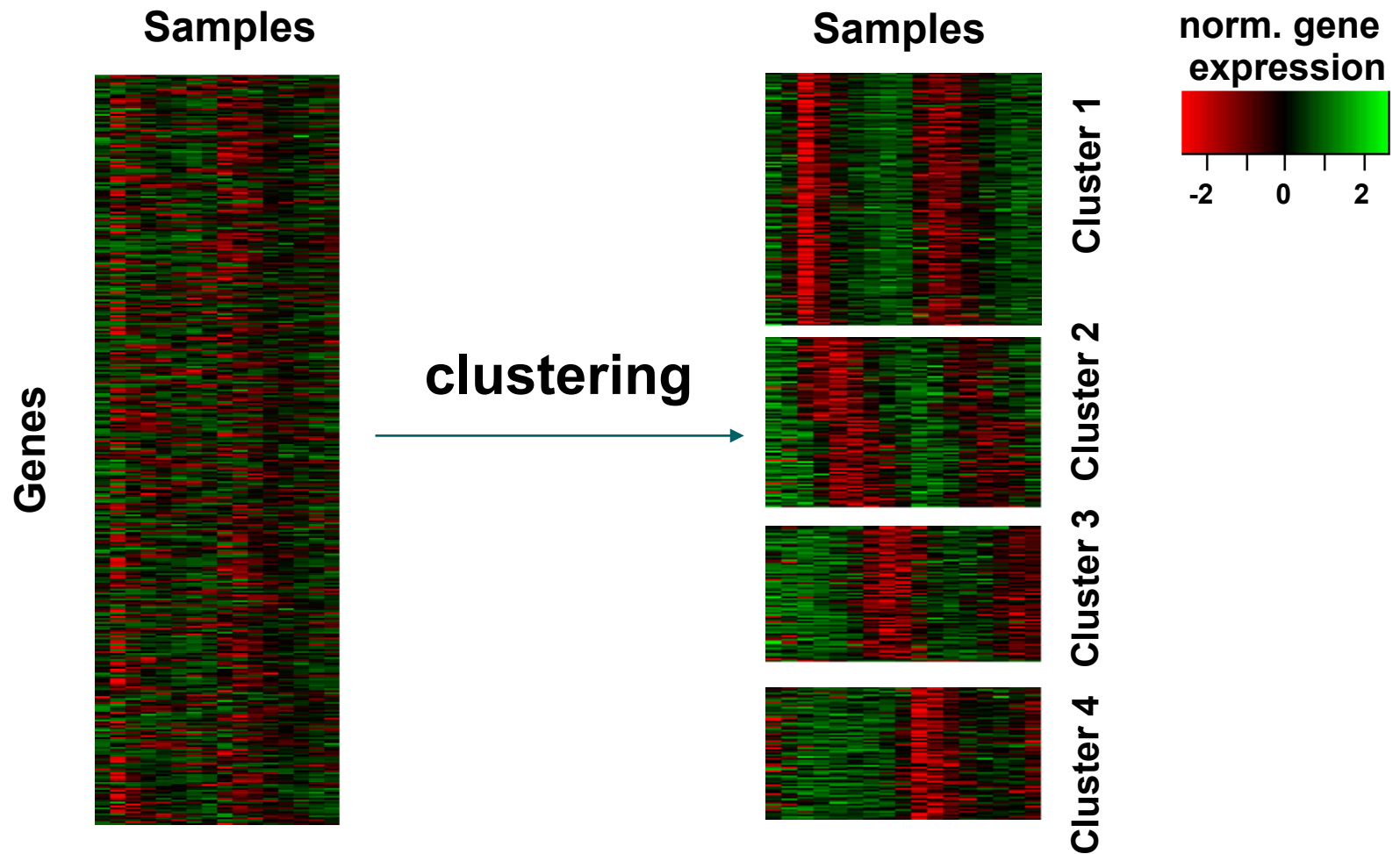
- **PCA allows an “blind” cell sorting**
  - only works if variant directions split the groups
  - is complementary to clustering
- **Weights allow interpretation of relevant variables**
- **Can also be used for quality check**
  - samples not fitting to groups
- **Alternatives to PCA:**
  - **tSNE - very commonly used in single cell RNA-seq**

# Clustering / Heatmaps

---



# Clustering / Heatmaps



clustering methods: k-means, **hierarchical clustering**, ...

# Distance

---

For a expression matrix  $X$  (genes vs. arrays), measure the distance between expression values of two genes ( $x_i$  e  $x_j$ )

- Euclidean distance (sensitive to scale)

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^L (x_{il} - x_{jl})^2}$$

- Pearson correlation (not sensitive to scale / similarity measure)

$$d(x_i, x_j) = \frac{\sum_{l=1}^L (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j)}{\sigma_i^2 \sigma_j^2}$$

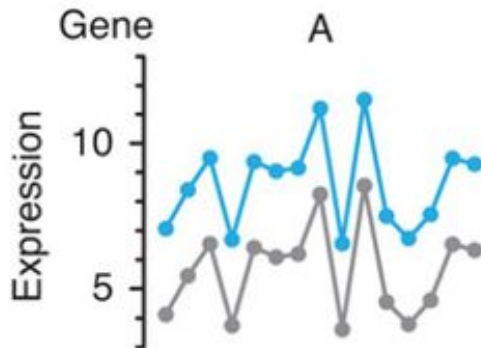


# Distance

Which distance for gene expression?

- example of two genes for 15 cancer patients

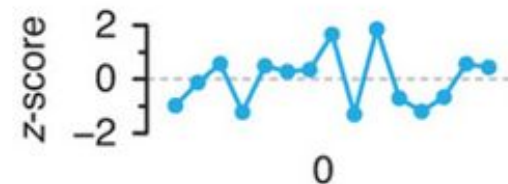
absolute expression



Euclidean - not similar

Correlation - similar

z-score normalised expression

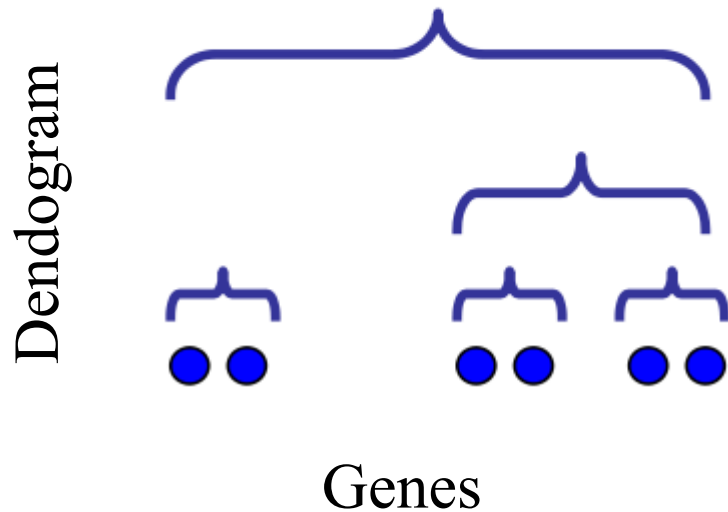


$$z = \frac{x_{ij} - \mu_i}{\sigma_i}$$

Euclidean - similar

Correlation - similar

# Hierarchical Clustering

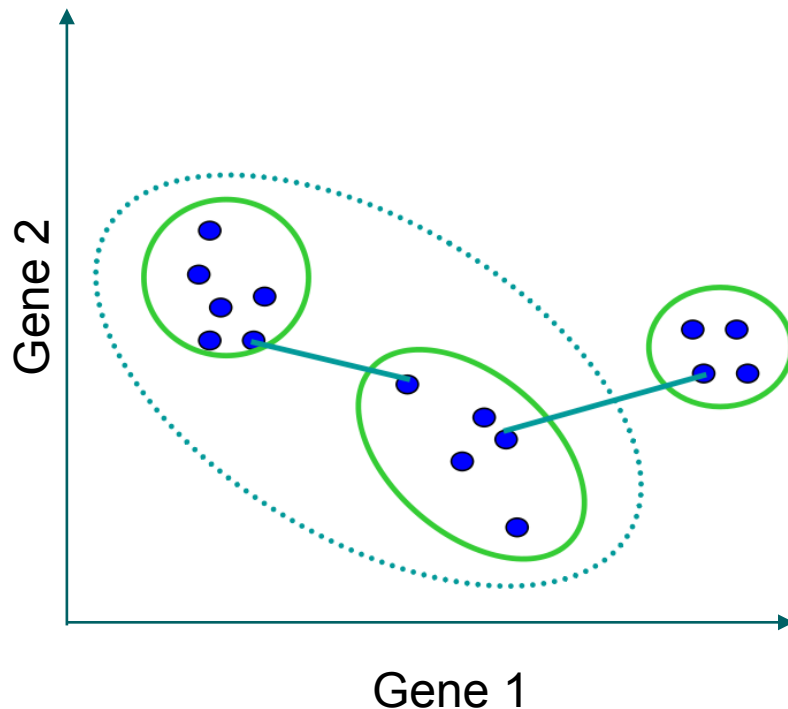


- Bottom up method
- Starting with a distance (similarity) matrix and each object as a group
- Repeat:
  - Joint two most similar groups
- Until the dendrogram has only one group

# Hierarchical Clustering

## Single-Linkage

- Join two groups where two examples are close
- Find groups with linear shapes



# Hierarchical Clustering

---

## Distance Matrix

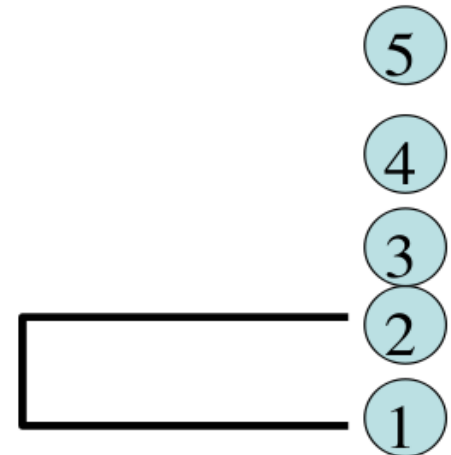
	1	2	3	4	5
1	0				
2	2	0			
3	6	3	0		
4	10	9	7	0	
5	9	8	5	4	0



# Hierarchical Clustering

## Distance Matrix

	1	2	3	4	5
1	0				
2	2	0			
3	6	3	0		
4	10	9	7	0	
5	9	8	5	4	0



# Hierarchical Clustering

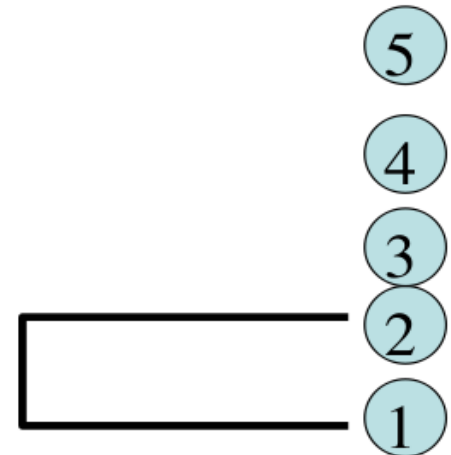
## Distance Matrix

$$\begin{array}{c} 1 \ 2 \ 3 \ 4 \ 5 \\ \begin{bmatrix} 0 & & & & \\ 2 & 0 & & & \\ 6 & 3 & 0 & & \\ 10 & 9 & 7 & 0 & \\ 9 & 8 & 5 & 4 & 0 \end{bmatrix} \end{array} \quad \rightarrow \quad \begin{array}{c} (1,2) \ 3 \ 4 \ 5 \\ \begin{bmatrix} 0 & & & \\ 3 & 0 & & \\ 9 & 7 & 0 & \\ 8 & 5 & 4 & 0 \end{bmatrix} \end{array}$$

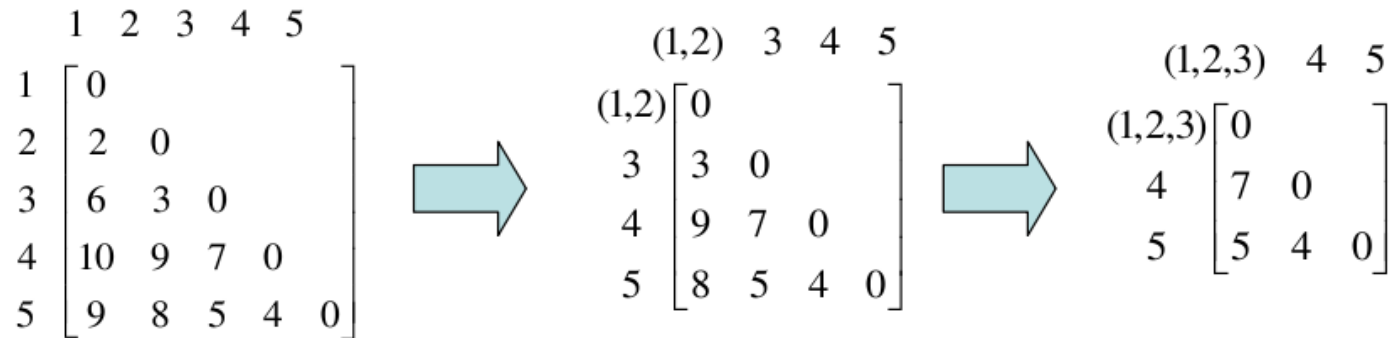
$$d_{(1,2),3} = \min\{d_{1,3}, d_{2,3}\} = \min\{6, 3\} = 3$$

$$d_{(1,2),4} = \min\{d_{1,4}, d_{2,4}\} = \min\{10, 9\} = 9$$

$$d_{(1,2),5} = \min\{d_{1,5}, d_{2,5}\} = \min\{9, 8\} = 8$$

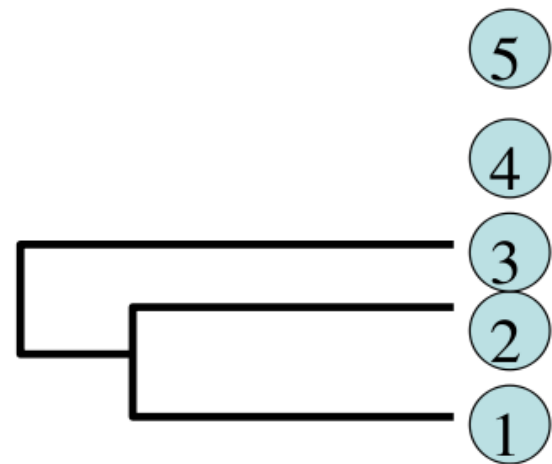


# Hierarchical Clustering

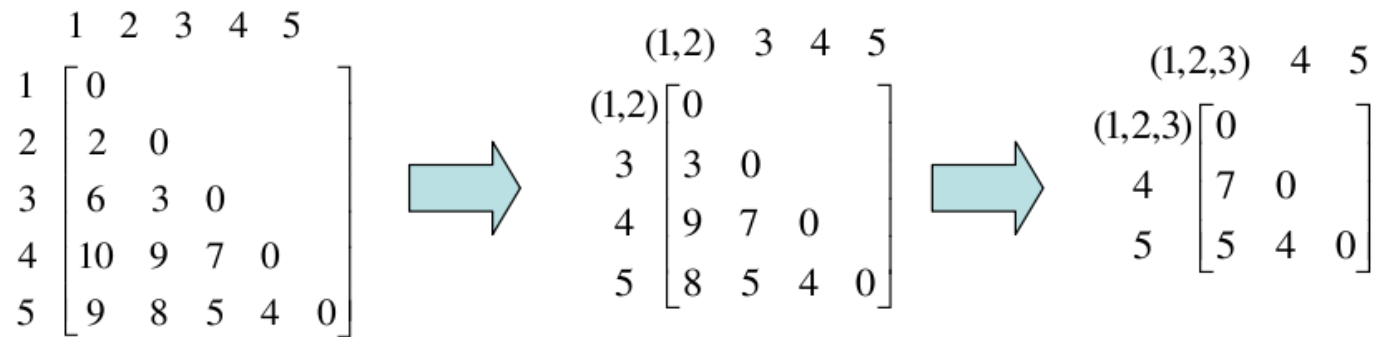


$$d_{(1,2,3),4} = \min\{d_{(1,2),4}, d_{3,4}\} = \min\{9, 7\} = 7$$

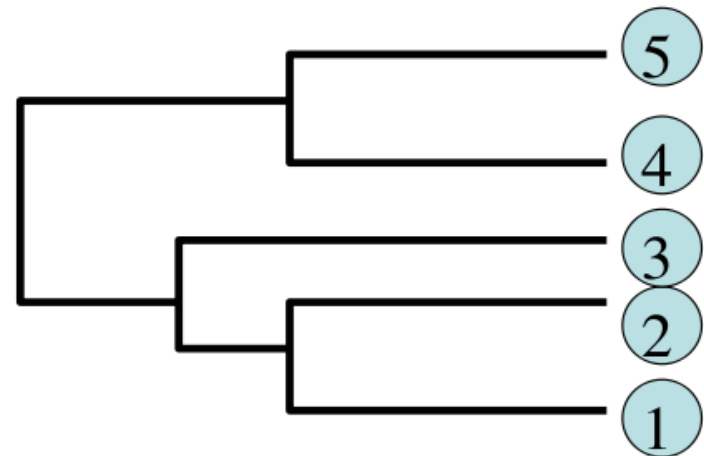
$$d_{(1,2,3),5} = \min\{d_{(1,2),5}, d_{3,5}\} = \min\{8, 5\} = 5$$



# Hierarchical Clustering

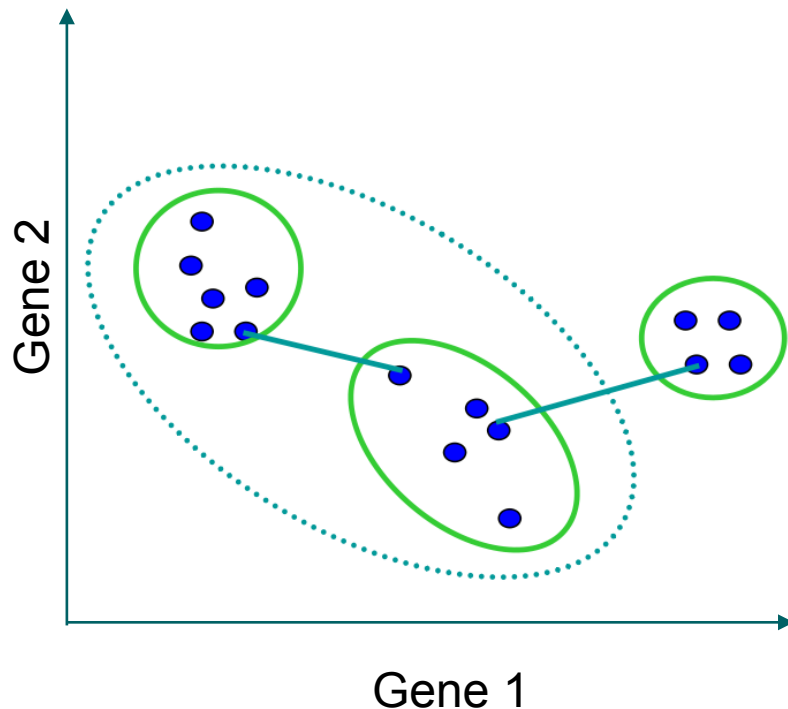


$$d_{(1,2,3),(4,5)} = \min\{d_{(1,2,3),4}, d_{(1,2,3),5}\} = 5$$





# Hierarchical Clustering



## Single-Linkage

- Groups with closest genes
- linear shapes

## Complete-Linkage

- Closest groups with more far genes
- Compact clusters

## Average Linkage

- Groups with closest centroids (middle)
- Outlier robust

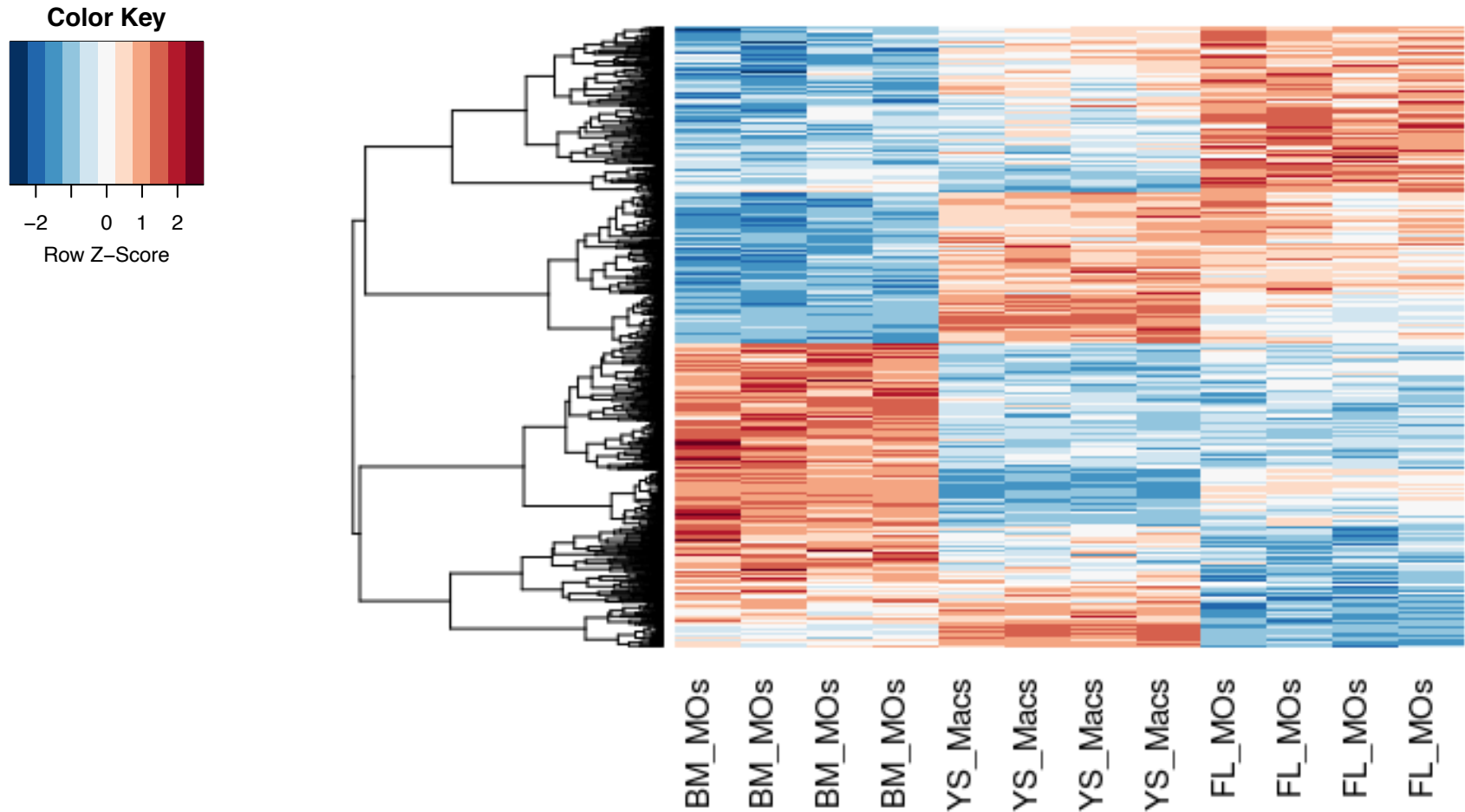
# Hierarchical Clustering

---

Which linkage?

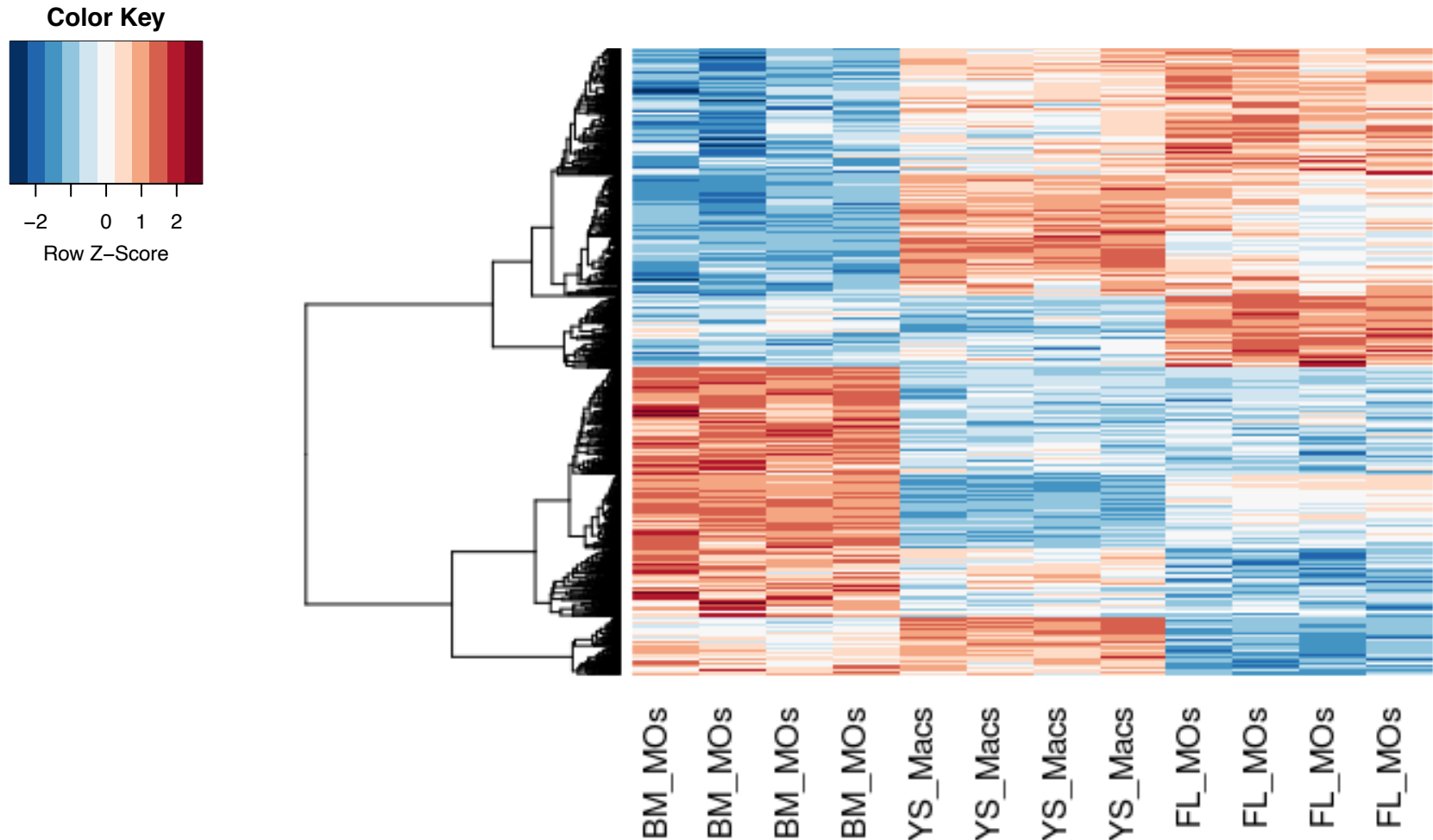
Which distance?

# Hierarchical Clustering - Complete Linkage



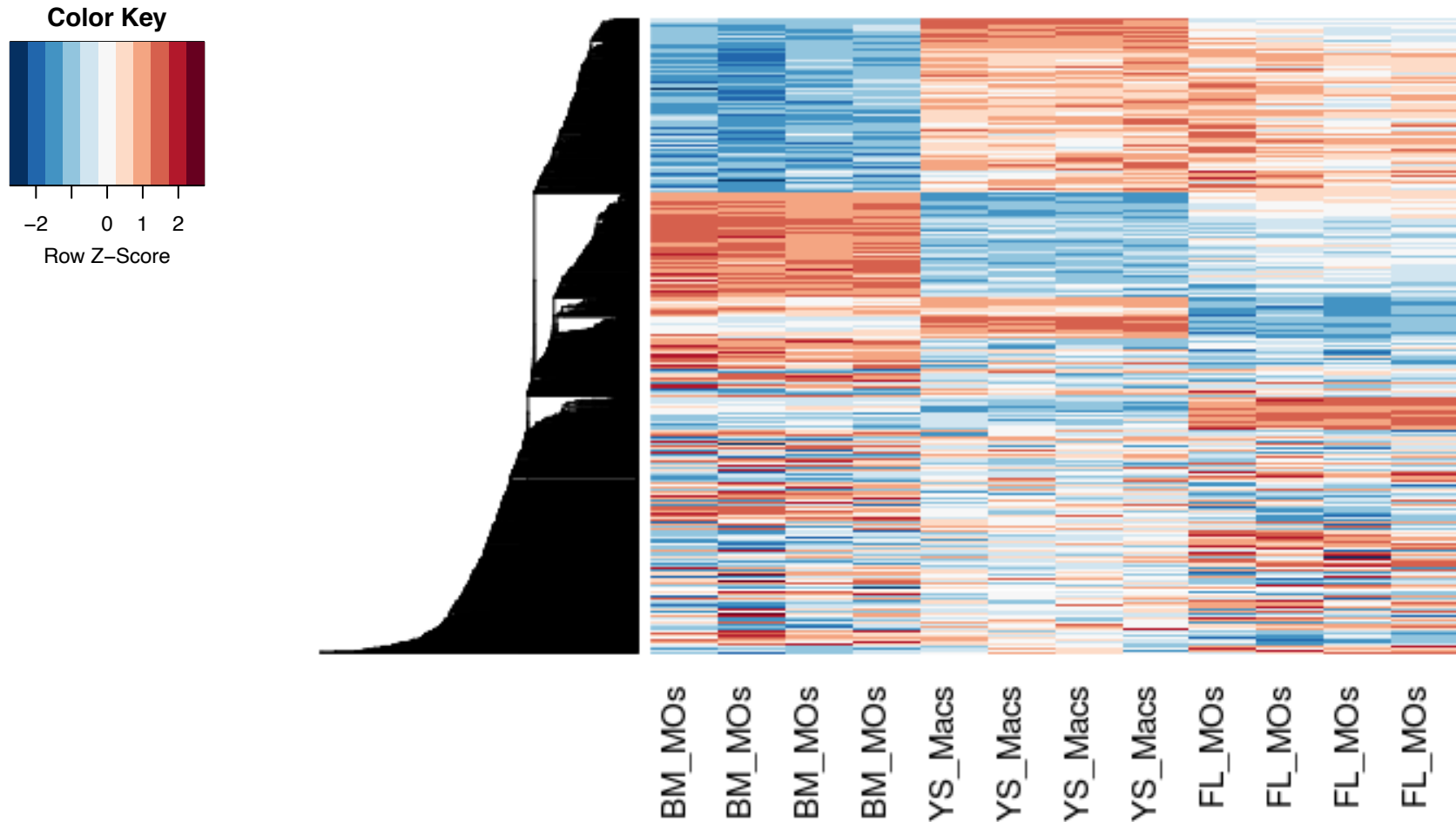
**metric - Pearson correlation (or Euclidean + z transform)**

# Hierarchical Clustering - Average Linkage



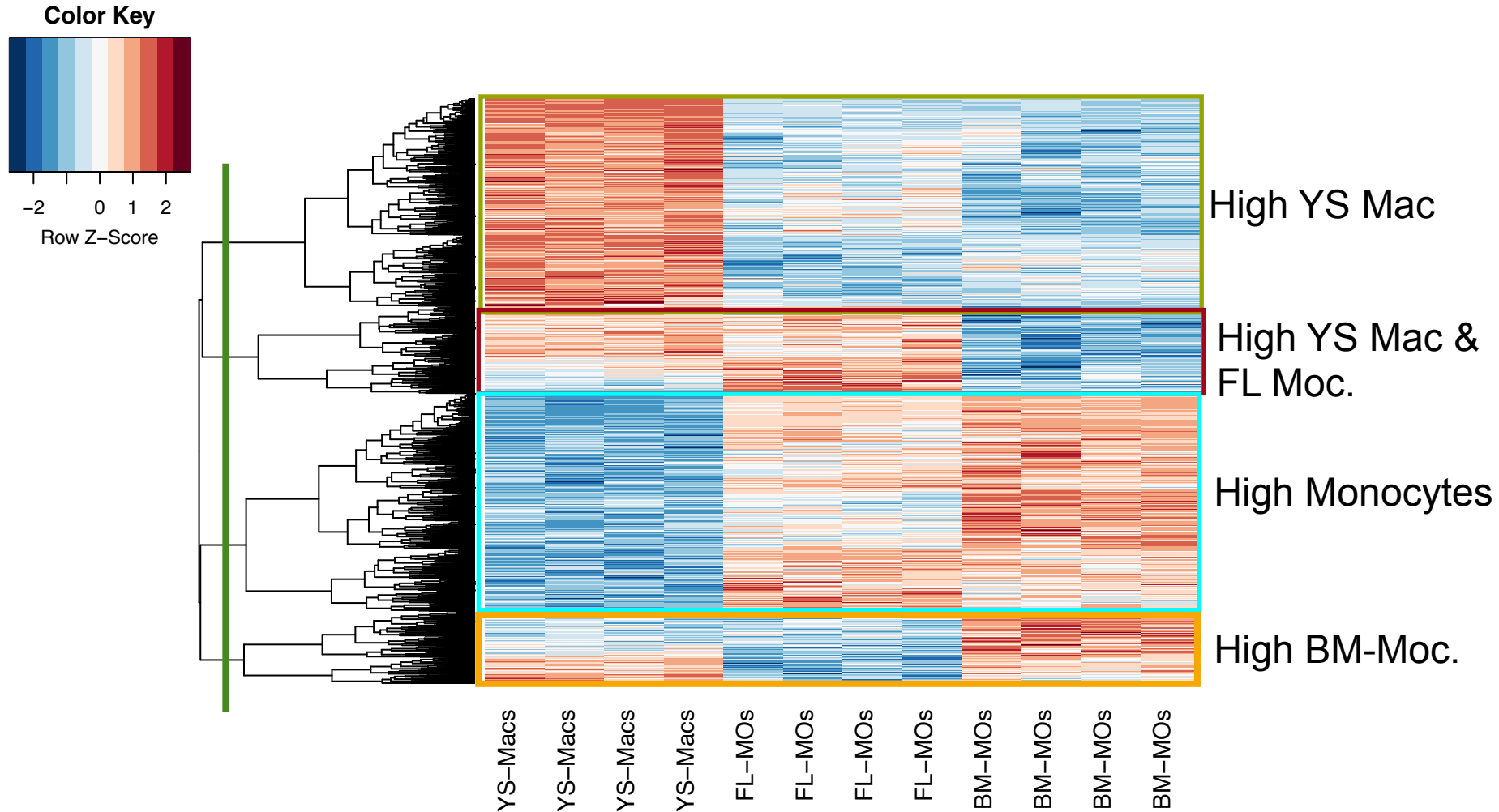
**metric - Pearson correlation (or Euclidean + z transform)**

# Hierarchical Clustering - Single Linkage



**metric - Pearson correlation (or Euclidean + z transform)**

# Hierarchical Clustering - Final Results



**distance metric - Pearson correlation recommended**

# Clustering - Resume

---

- Clustering allow detection of unknown groups in the data
- Classical methods (hierarchical or k-mean) work well in general
- How to choose distance and linkage?
  - Pearson or Euclidean (followed by z-transform)
  - Heatmaps usually only like nice with z-transform
- How to find number of groups?
  - No simple solution!

---

# Hands on!

Handout Step 4 and 5



# Functional Analysis

---

Clustering/Differential Expression (DE) returns lists of hundreds of genes How to functionally characterize these?

**Solution 1** - Look at each gene individually

**Solution 2** - Relate these genes to annotations from databases

- Gene Ontology, pathways, gene sets, disease ontology, ...

# Databases

Manually or automatic curated annotation of genes

## Pathways



## Experimental



**MSigDB**  
Molecular Signatures  
Database

## Ontologies



# Gene Ontology

---

Controlled vocabulary to describe gene and gene product attributes in any organism

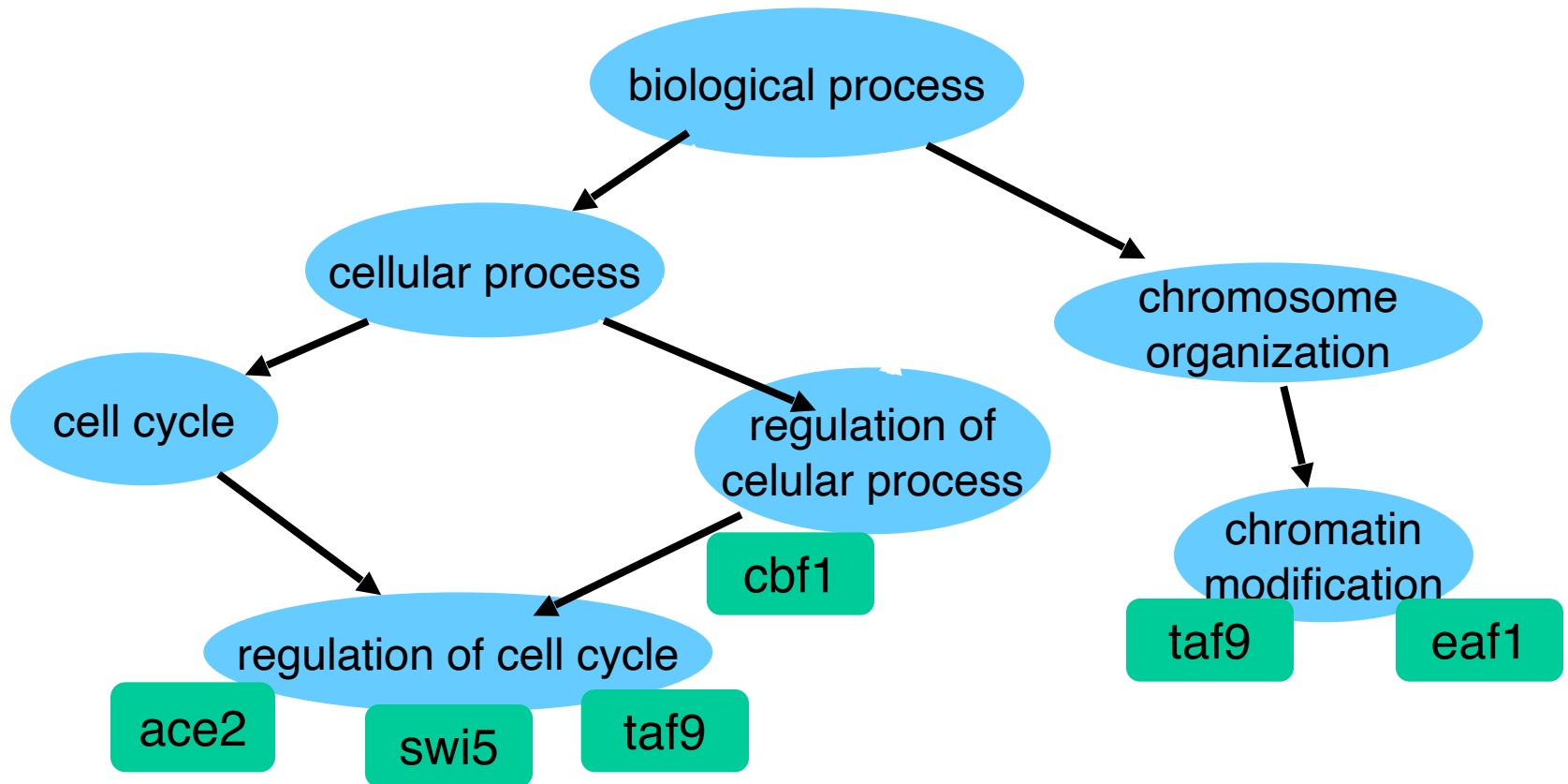
Formed by three ontologies

1. Biological Process (BP)
2. Molecular Function (MF)
3. Cellular Component (CC)

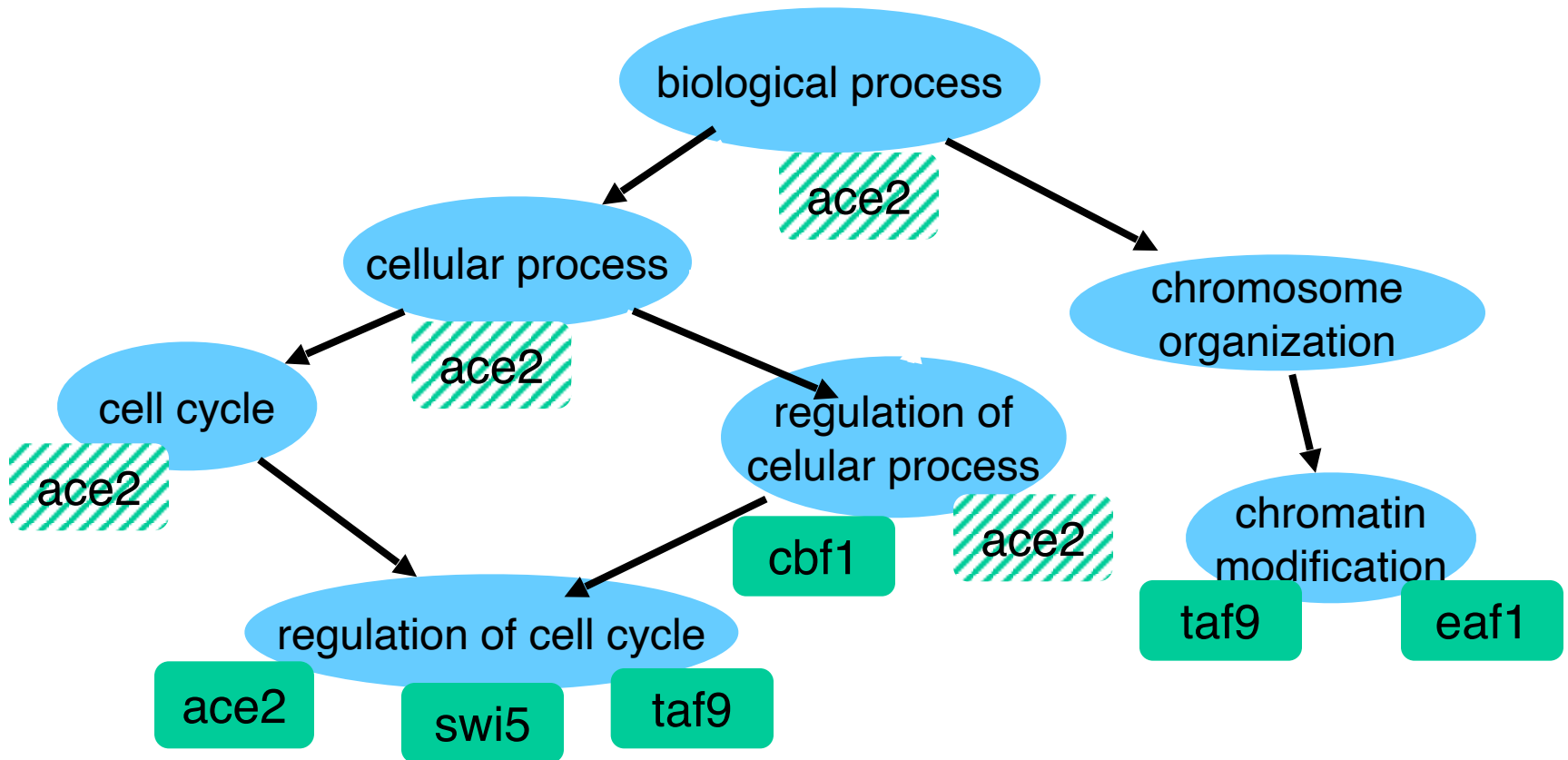
Annotation (Organism depend)

- genes are associated to terms manually (literature) or automatically (sequence homology)

# Gene Ontology



# Gene Ontology



inheritance property

# GO Enrichment Analysis

## DE analysis results

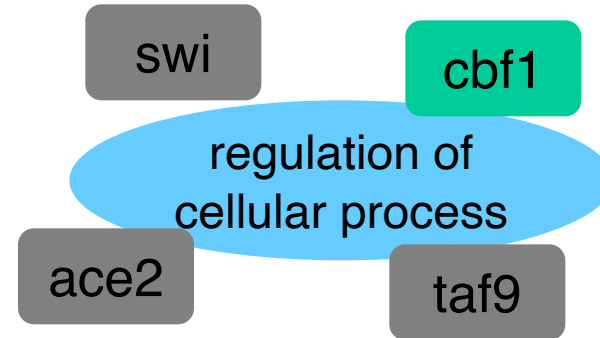
up regulated genes

SWI  
ACE2  
CBF1  
YJL099W  
YDL198C  
YCR085W  
YCR043C  
YDR825C

all other genes

YDL093W  
YER016W  
YNL126W  
YKL053W  
YJL099W  
YDL198C  
YCR085W  
YBR043C  
YDR325W  
YCR085W  
YBR043C  
...

## GO Term



How probable is that 3 up regulated genes are annotated to the GO term?

# GO Enrichment Analysis

## DE analysis results

up regulated genes

SWI  
ACE2  
CBF1  
YJL099W  
YDL198C  
YCR085W  
YCR043C  
YDR825C

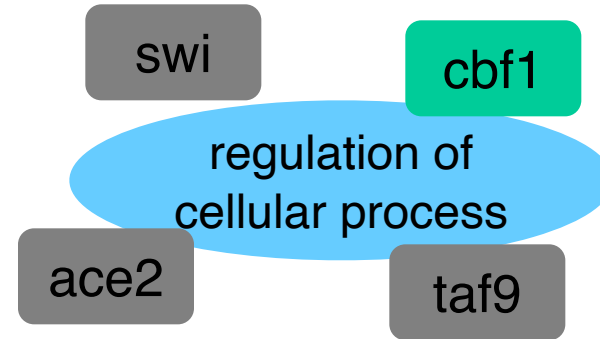
all other genes

YDL093W  
YER016W  
YNL126W  
YKL053W  
YJL099W  
YDL198C  
YCR085W  
YBR043C  
YDR325W  
YCR085W  
YBR043C  
...

## Statistics:

Fisher's Exact Test

## GO Term



## GO Term Annotation

	GO Term Annotation	
	YES.	NO
Up-regulated	YES	3
	NO	6421

# Enrichment Analysis Tools

---

For a given gene list:

1. evaluate the the overlap of the list vs. all gene sets  
i.e. GO terms, pathways, ...
2. Estimate p-value (corrected by multiple testing)
3. Rank gene sets by lowest p-value



We interface for enrichment analysis with:  
Gene Ontology, KEGG Pathway and TF binding

<http://biit.cs.ut.ee/gprofiler/index.cgi>

Check the results for my favorite genes:

Irf8 Id2 Spi1 Klf4 Runx2 Egr1

# Gene Set Enrichment Analysis

---

Perform a functional evaluation of ranking of genes

- i.e all genes ranked by fold change cond. A vs. B

Advantages over “Normal” enrichment analysis:

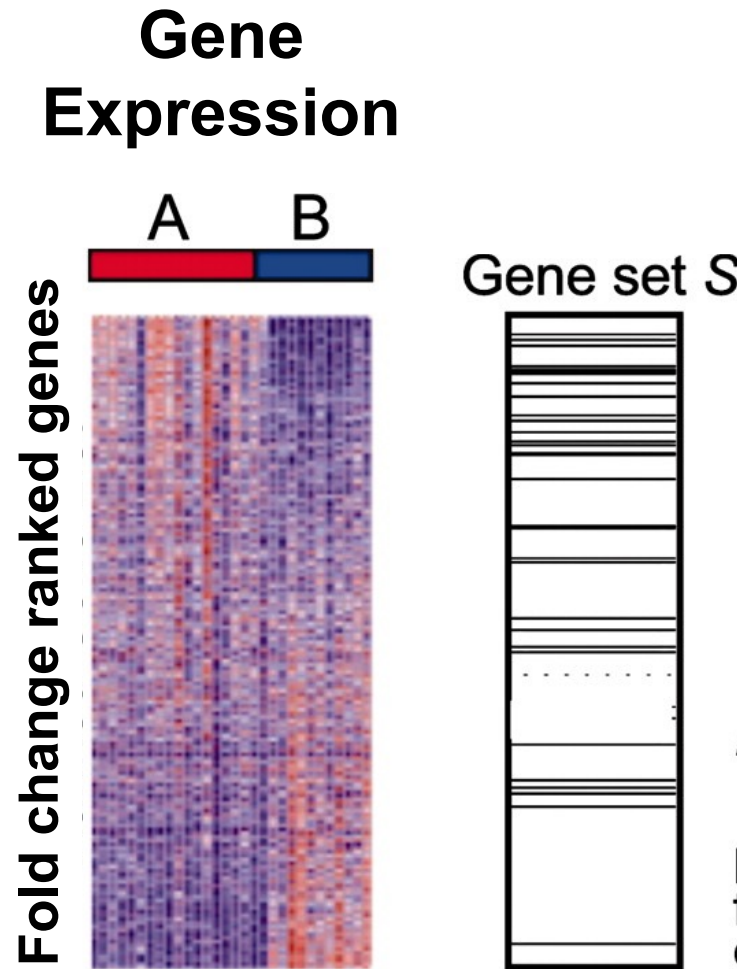
- do not require previous DE analysis
- works when effects of the experiment are low

## GSEA Gene Sets

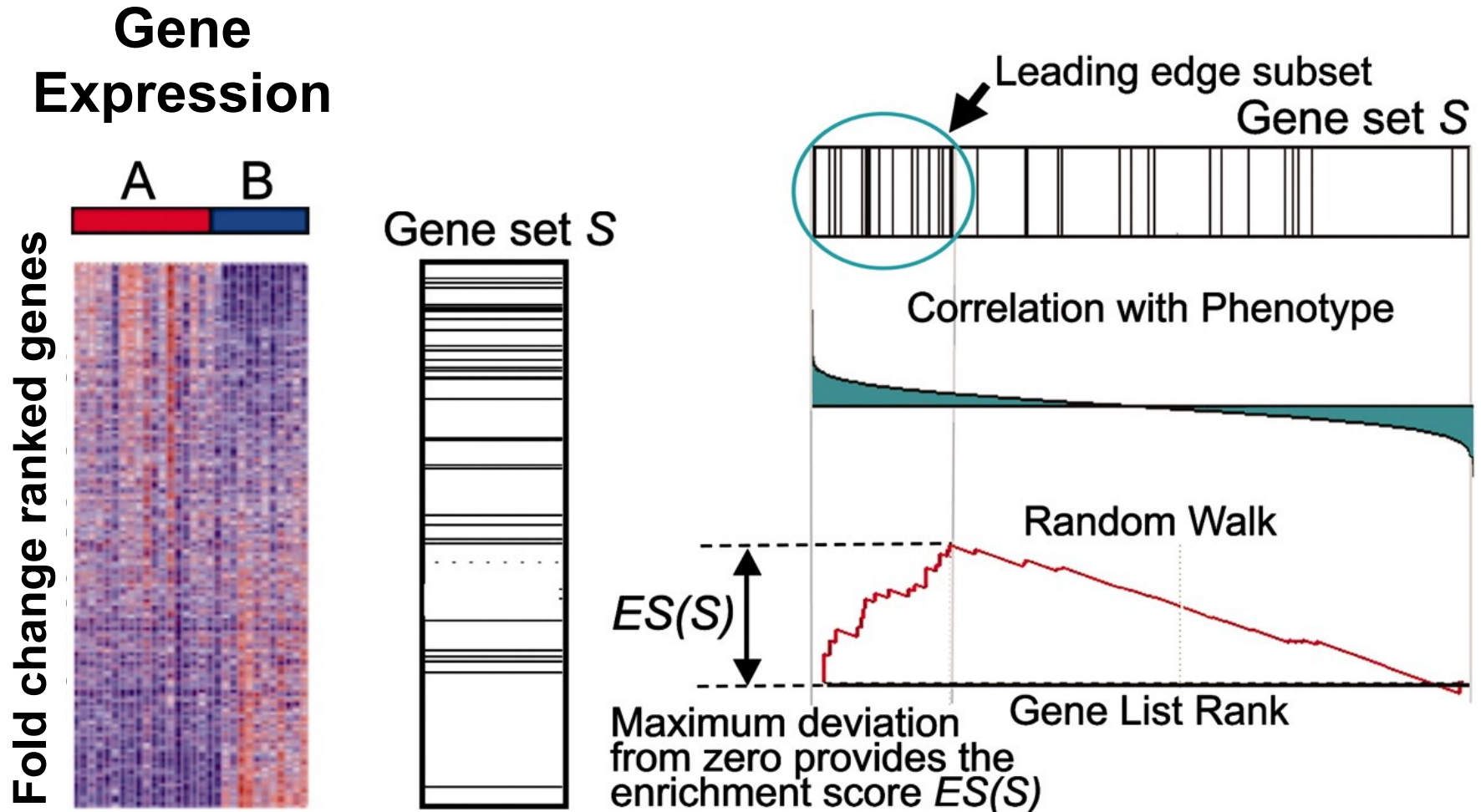
- GO Terms, KEGG Pathways
- experimentally derived Gene Sets
  - DE genes from microarray studies from GEO
  - Can be obtained at mysigdb  
([software.broadinstitute.org/gsea/msigdb/](http://software.broadinstitute.org/gsea/msigdb/))

# Gene Set Enrichment Analysis

---



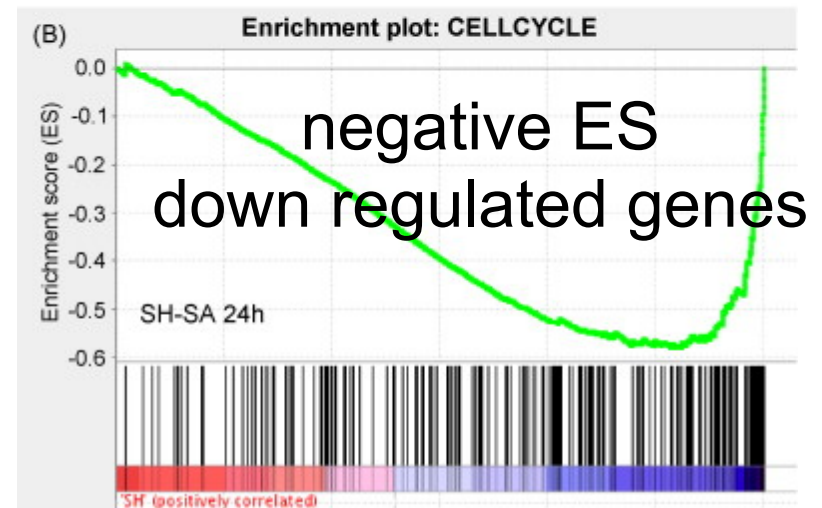
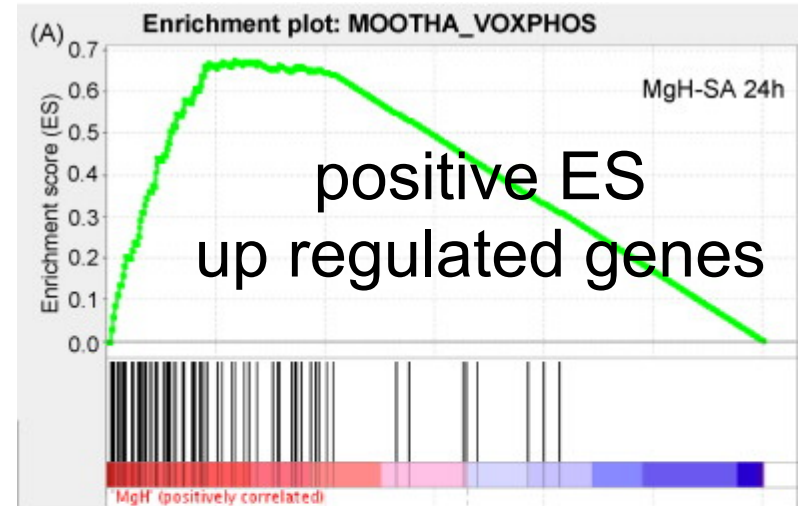
# Gene Set Enrichment Analysis



# Gene Set Enrichment Analysis

For a given gene ranking:

1. evaluate ES score for all gene sets
2. estimate p-value(corrected)
3. rank gene sets by lowest p-value



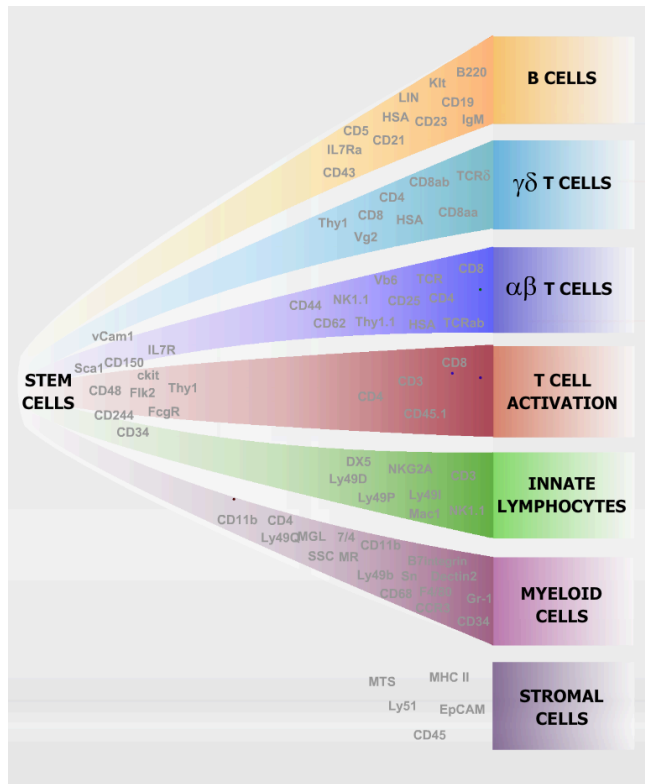
---

# Hands on!

Handout Step 6 and 7

# Integrative Analysis - ImmGen

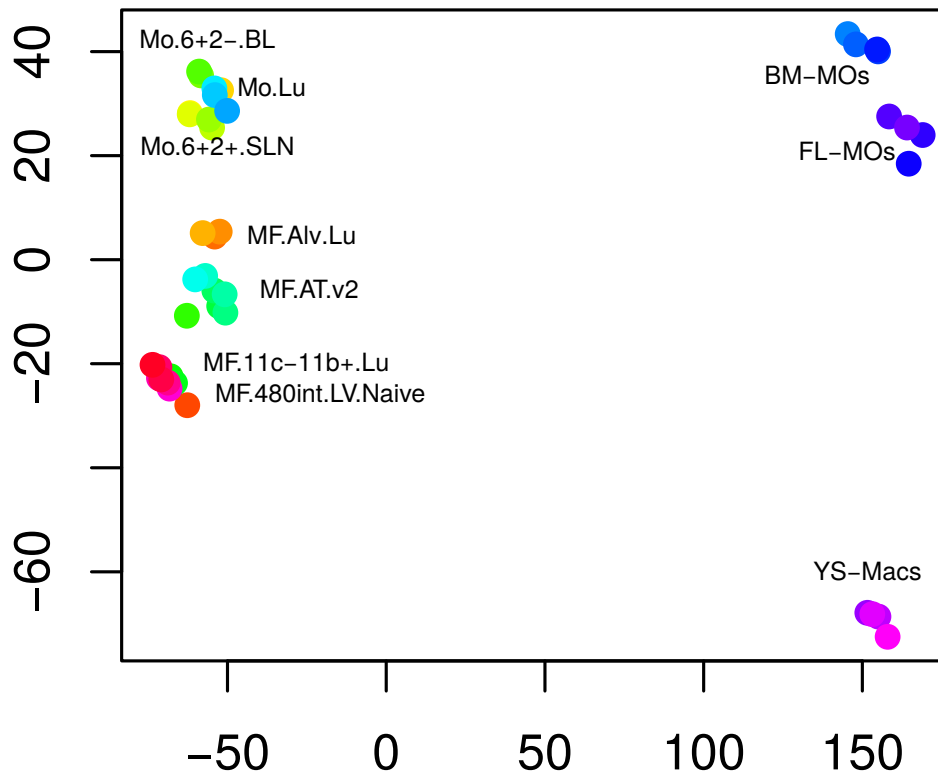
- ImmGen - expression data of immune cells under standardized conditions



- How do cells from **van de Leer, 2016** compares to monocyte/macrophages from ImmGenn?
- we obtained/pre-processed ImmGen data (v1) from GEO (GSE15907)

# Integrative Analysis - Problem

- Batch Effects - Arrays from distinct lab tends to cluster together



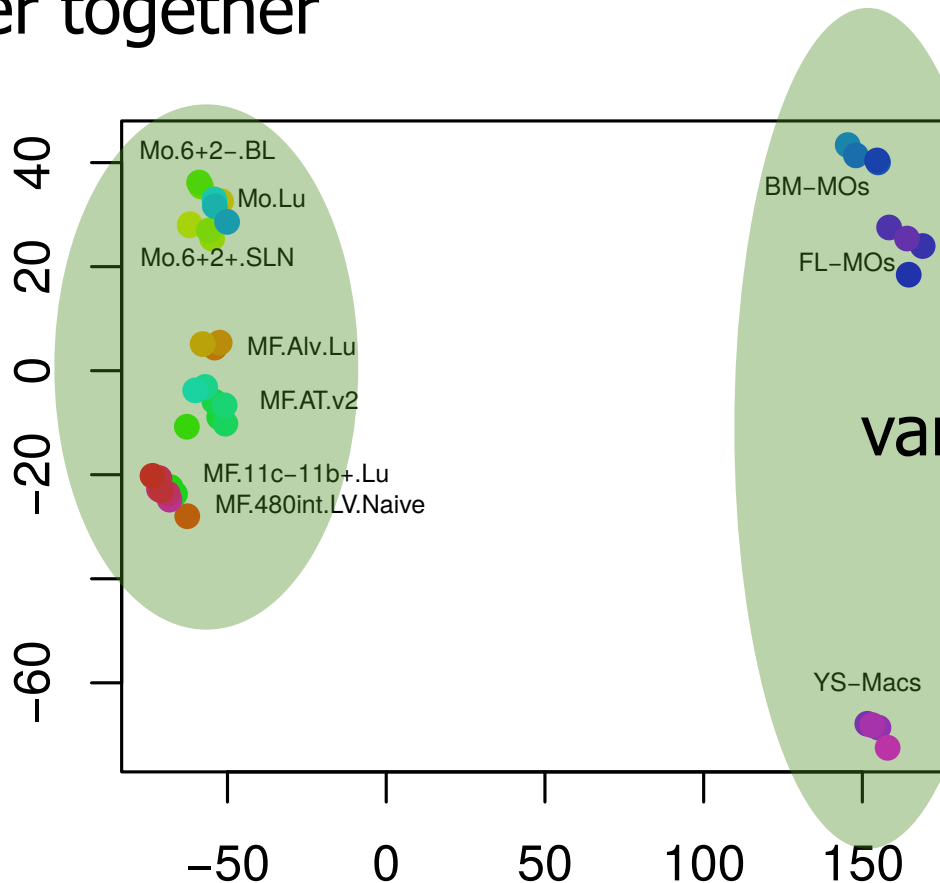
See: Leek JT,.... (2016). sva: Surrogate Variable Analysis. R package version 3.22.0.



# Integrative Analysis - Problem

- Batch Effects - Arrays from distinct lab tends to cluster together

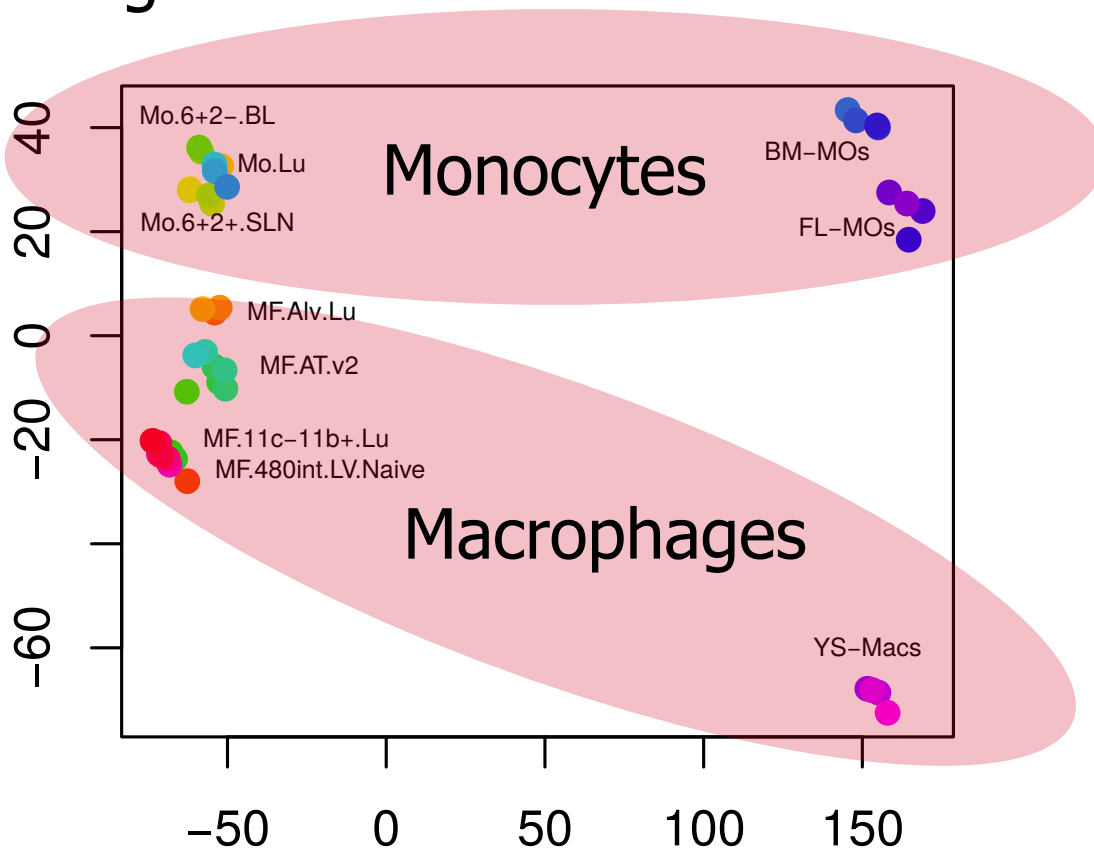
ImmGenn



See: Leek JT,.... (2016). sva: Surrogate Variable Analysis. R package version 3.22.0.

# Integrative Analysis - Problem

- Batch Effects - Arrays from distinct lab tends to cluster together



See: Leek JT,.... (2016). sva: Surrogate Variable Analysis. R package version 3.22.0.

# Integrative Analysis - PCA After Combat

---

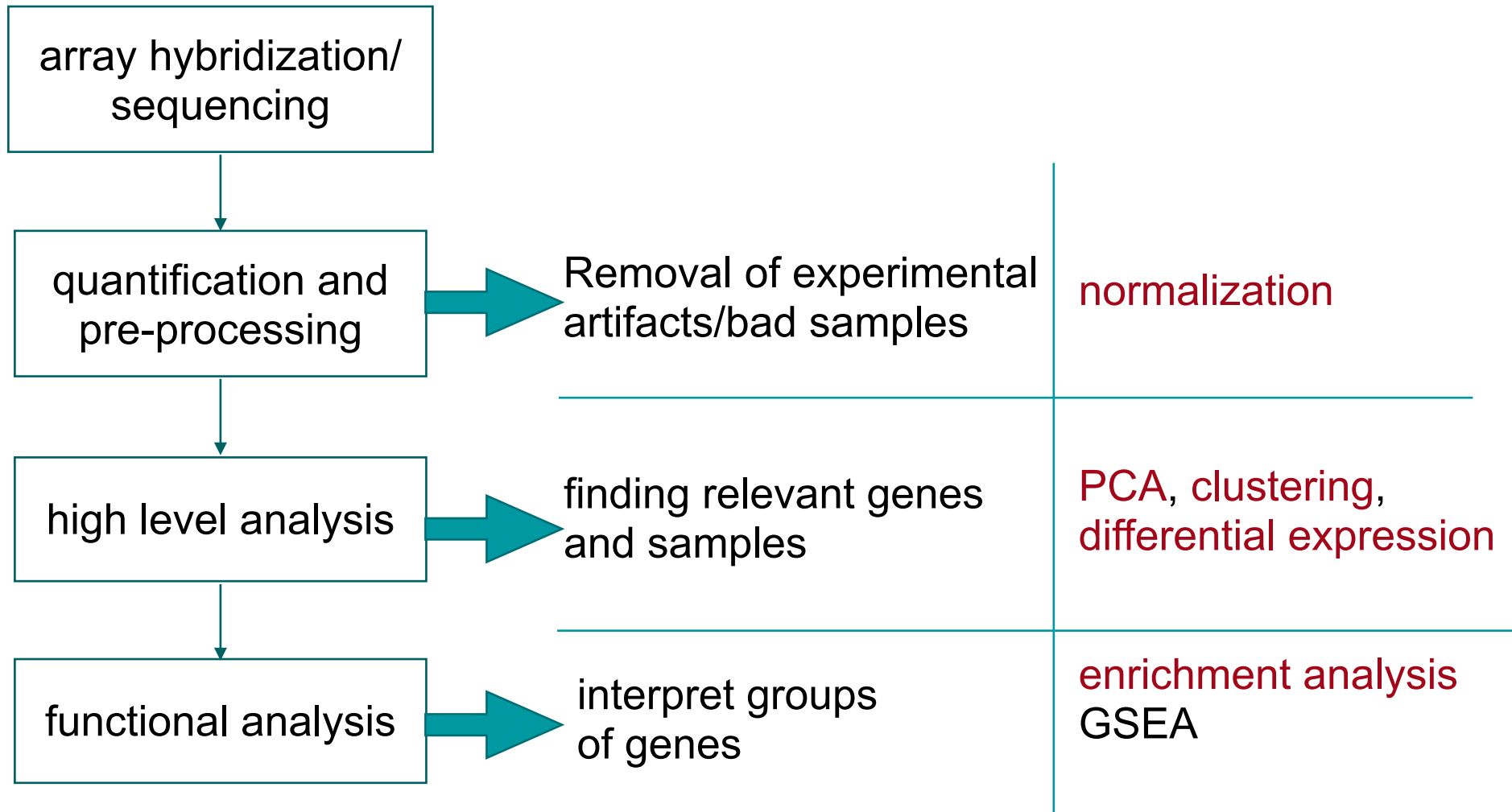
- Solution - Batch effect removal with COMBAT
  - annotation of your data: tissue of origin, cell type, experimental batches

## Hands on!

Handout Step 7

See: Leek JT,.... (2016). sva: Surrogate Variable Analysis. R package version 3.22.0.

# Bioinformatics - Gene Expression Analysis



# Afternoon Exercise

---

- Analyse gene expression data (steps 1-7 of handout) of the following paper:
  - Spence JR, Mayhew CN, Rankin SA, Kuhar MF et al. Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. Nature 2011 Feb 3;470(7332):105-9.
- Try to get answers to the following questions with your analysis:
  - Are the stem cells and induced pluripotent cells the same?
  - If not, what are the reasons?

