



Swiss Institute of  
Bioinformatics

# Enrichment analysis

**Geert van Geest**

Slide courtesy: Tania Wyss

# Many differentially expressed genes. What's next?

Manually check  $> 100$  (often  $> 1000$ ) genes



Check for enriched sets of genes based on e.g. pathway, function, location

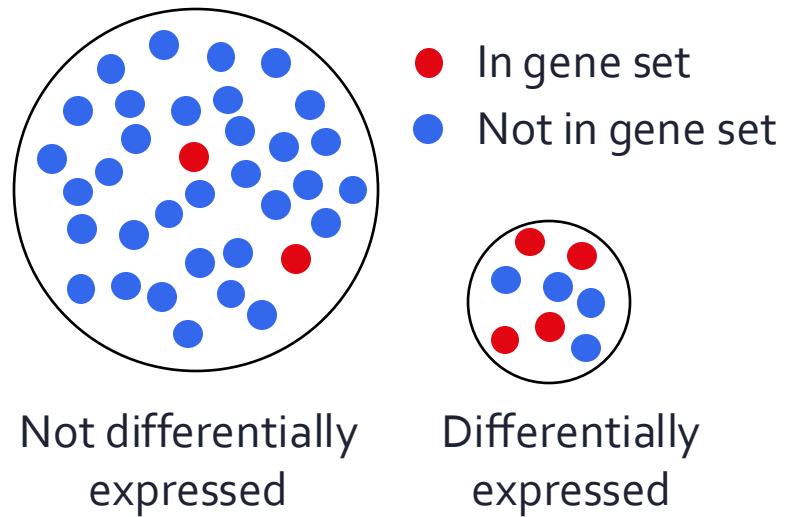


# Gene sets can be anything

- »» All genes part of a pathway
- »» Genes of which the proteins co-occur in a cellular location
- »» Proteins that are regulated by the same transcription factor
- »» Genes that show co-expression
- »» Genes that can carry mutations associated with a disease
- »» etc.

# The two most used methods

## Overrepresentation analysis (ORA; Fisher's test)

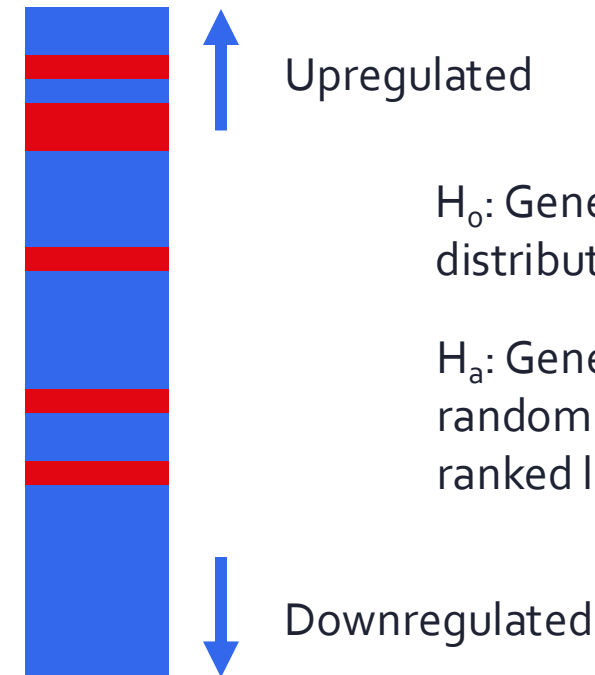


$H_0$ : The proportion of genes in the gene set is the same for both groups

$H_a$ : The proportion of genes in the gene set is higher in the differentially expressed group

## Gene set enrichment analysis (GSEA)

Genes ranked by  
test statistic



$H_0$ : Genes in set are randomly distributed over ranked list

$H_a$ : Genes in set are not randomly distributed over the ranked list

# Sources of gene sets

- ⌘ GeneOntology <https://geneontology.org>. Available in R through [OrgDb packages](#)
- ⌘ [MSigDB](#): database containing several types of gene set lists
  - ⌘ hallmark
  - ⌘ published gene sets
- ⌘ [KEGG](#)
- ⌘ [Reactome](#)
- ⌘ [WikiPathways](#)

# GSEA and ORA in R

- ❖ Most used package: clusterProfiler
- ❖ For ORA:
  - ❖ enrichGO : GO terms
  - ❖ enricher : other gene sets
- ❖ For GSEA:
  - ❖ gseGO : GO terms
  - ❖ gseKEGG : KEGG
  - ❖ GSEA : other gene sets

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
GO:0007059	chromosome segregation	105/809	424/18870	0	0	0
GO:0098813	nuclear chromosome segregation	79/809	312/18870	0	0	0
GO:0000070	mitotic sister chromatid segregation	62/809	184/18870	0	0	0
GO:0000280	nuclear division	88/809	441/18870	0	0	0
GO:0044772	mitotic cell cycle phase transition	77/809	470/18870	0	0	0
GO:0051983	regulation of chromosome segregation	40/809	131/18870	0	0	0

# Visualizations

```
enrichGO(de) |>  
pairwise_termsim() |>  
emapplot()
```

