Functional Profiling

Rubén Grillo Risco
Bioinformatics & Biostatistics Unit. CIPF

WODA

WEB-BASED OMICS DATA ANALYSIS

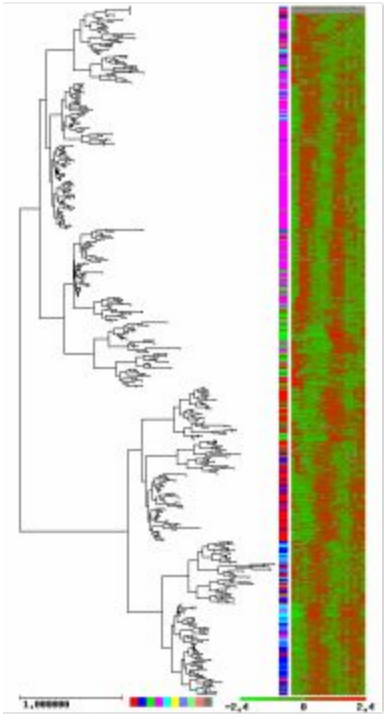
Unidad de
Bioinformática y
Bioestadística

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

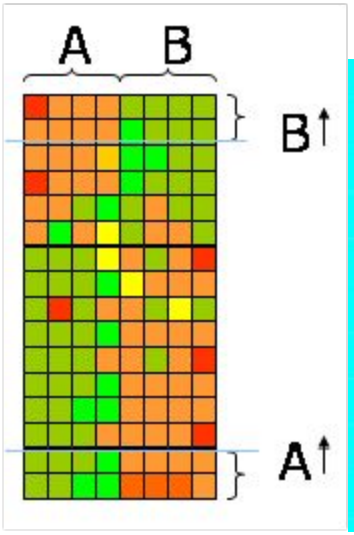
Outline

- Introduction
- Over-Representation Analysis (ORA)
- Gene Set Analysis (GSA)
- Network Analysis (NA)

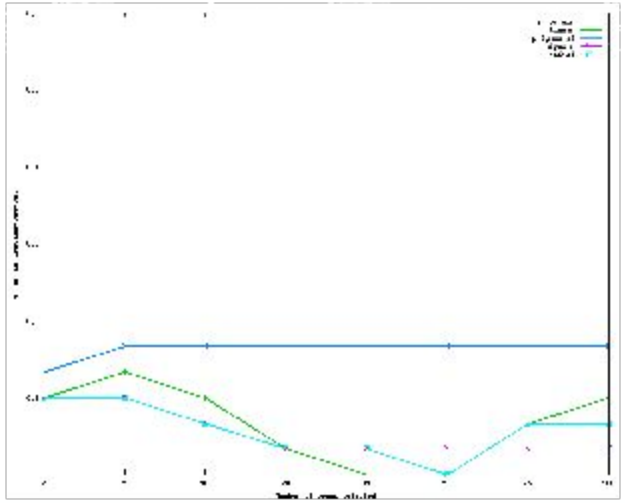
Genome-scale experiment output

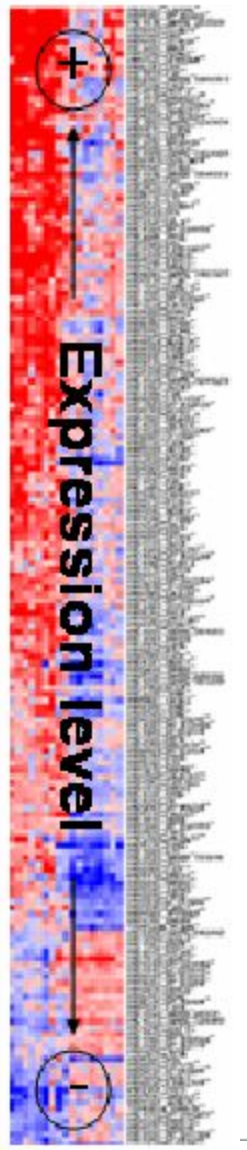


BEST1
BRCA2
FIT
BRCA1



1007_s_at	12.4
1053_at	11.5
117_at	10.3
121_at	10.2
1255_g_at	9.9
1294_at	9.3
1316_at	8.2
1320_at	8.1
1405_i_at	7.7
1431_at	7.4

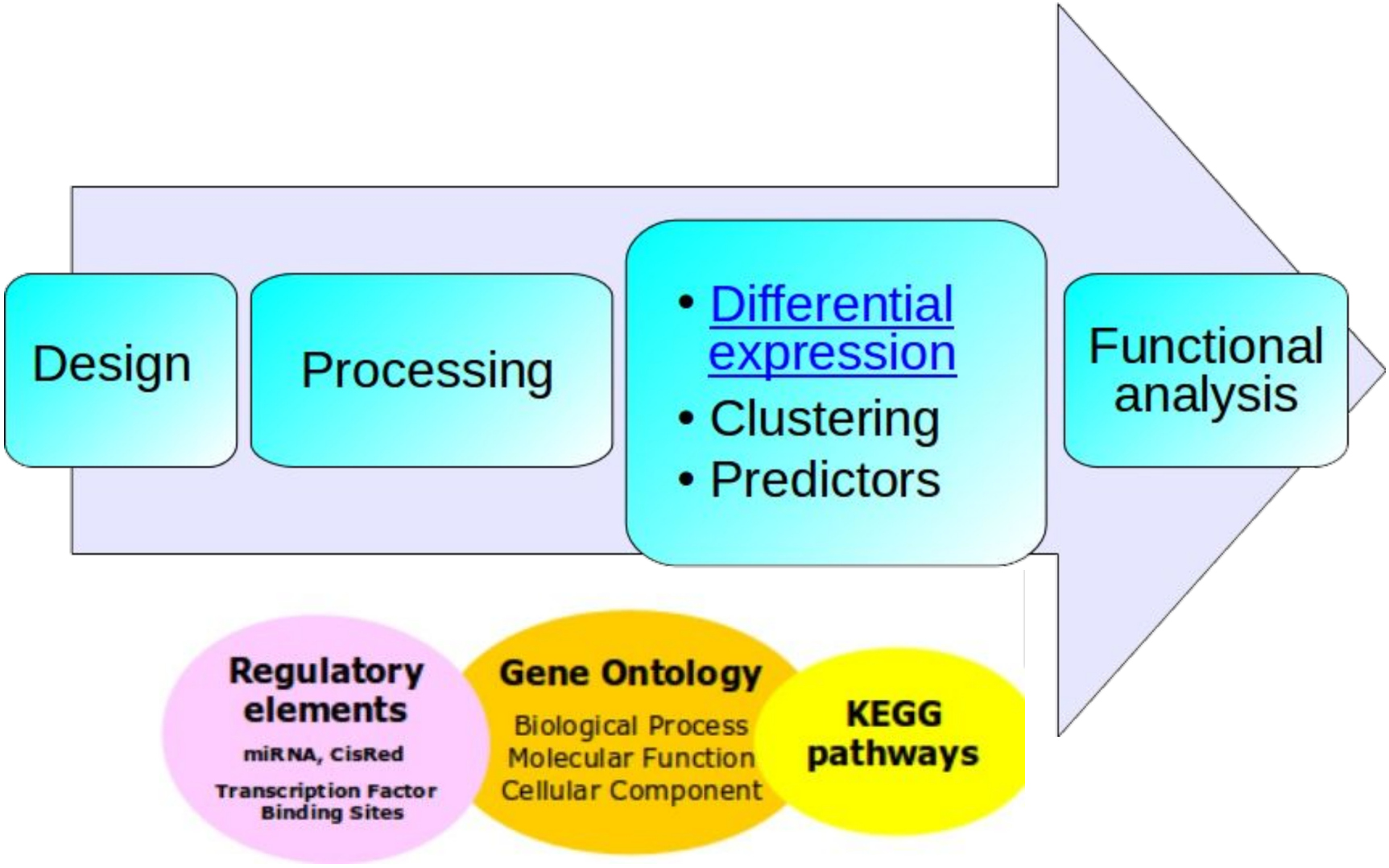




Questions we try to answer

- Is there any significant functional enrichment in my gene list / gene sets?
- Are these genes involved in common pathways?
- Do they share specific regulation?
- Are they involved in the same disease?

Omics data analysis pipeline



What do we need?

Annotation file + **Statistical tests**

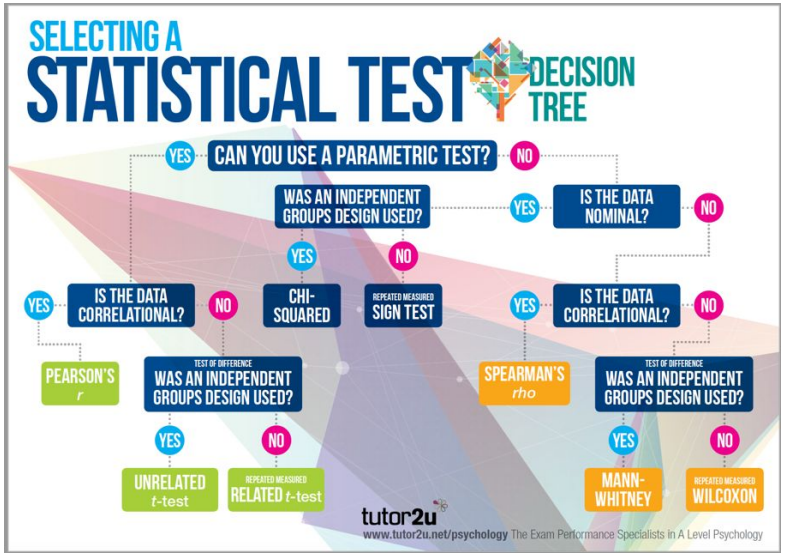
gene → function

FunctionA: Gene1, Gene2, Gene3…
FunctionB: Gene4, Gene2, Gene5...

GENEONTOLOGY
Unifying Biology

KEGG Kyoto Encyclopedia of Genes and Genomes

reactome



Functional databases









Homo sapiens | *Mus musculus* | *Rattus norvegicus* | *Gallus gallus* | *Danio rerio* | *Drosophila melanogaster* | *C. elegans* | *Saccharmoyces cerevisae* | *Arabidopsis thaliana*

UniProt/Swiss-Prot
UniProtKB/TrEMBL
Ensembl IDs
EntrezGene
Affymetrix
Agilent

Genes IDs

HGNC symbol
EMBL acc
RefSeq
PDB
Protein Id
IPI....

Biological databases

KEGG pathways

Biocarta pathways

Keywords Swissprot

Gene Ontology

Biological Process
Molecular Function
Cellular Component

Gene Expression in tissues

Regulatory elements

MiRNA, CisRed

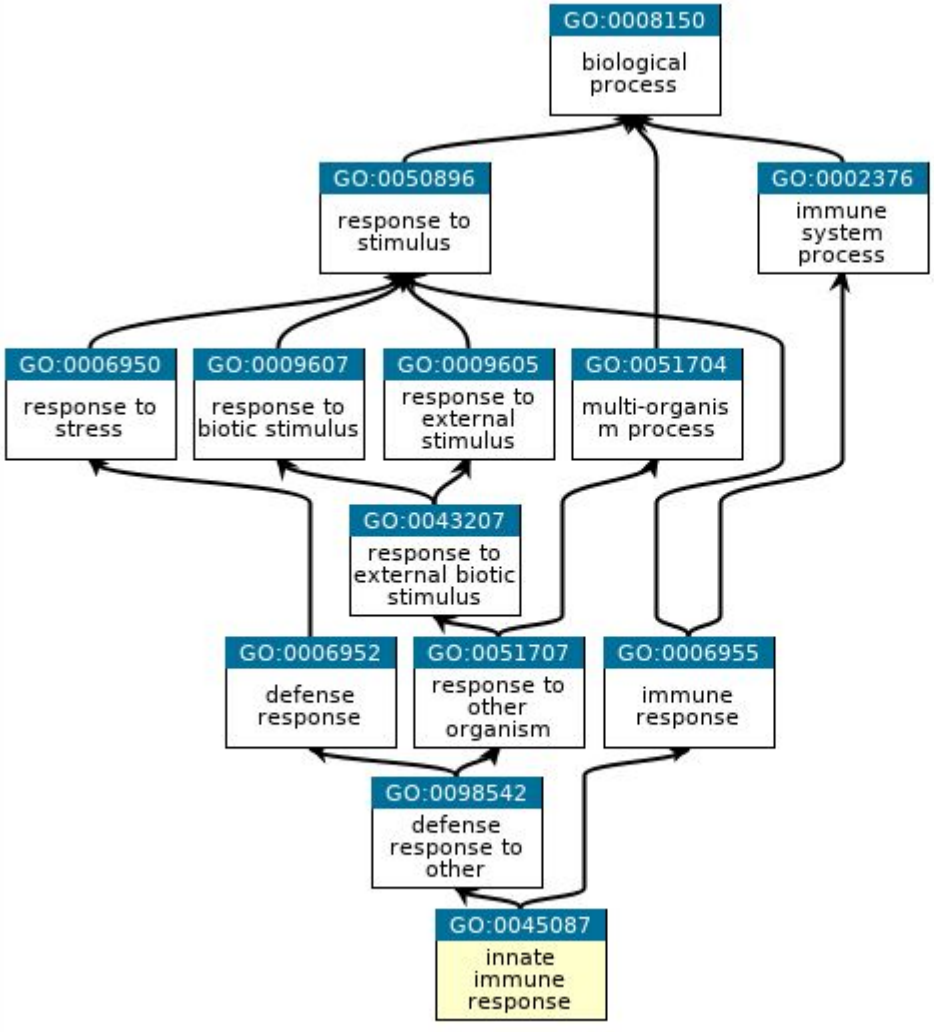
Transcription Factor Binding Site

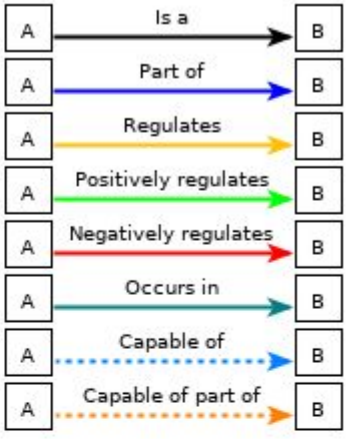
Bioentities from literature:

Diseases terms
Chemical terms

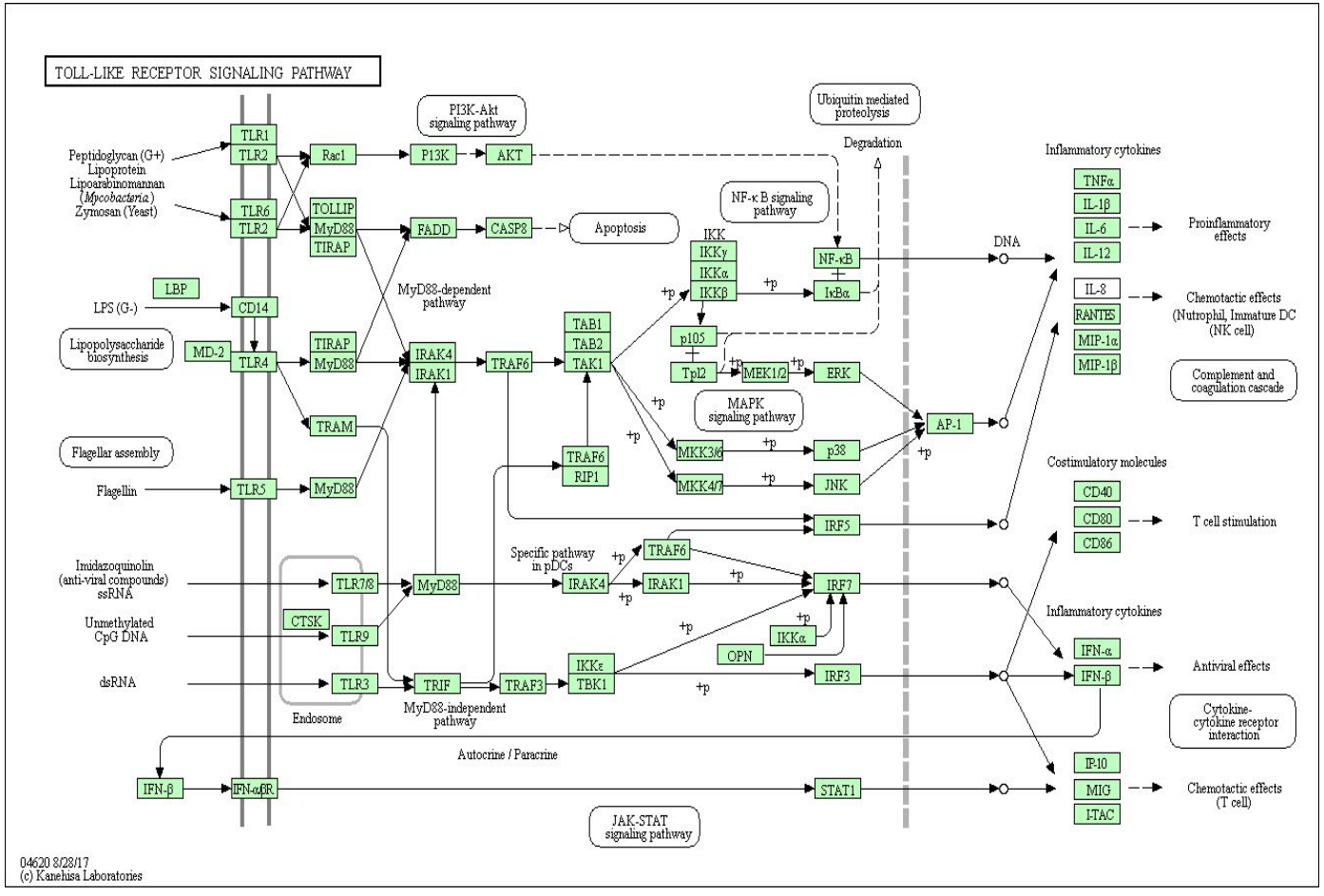
Gene Ontology

GENEONTOLOGY
Unifying Biology

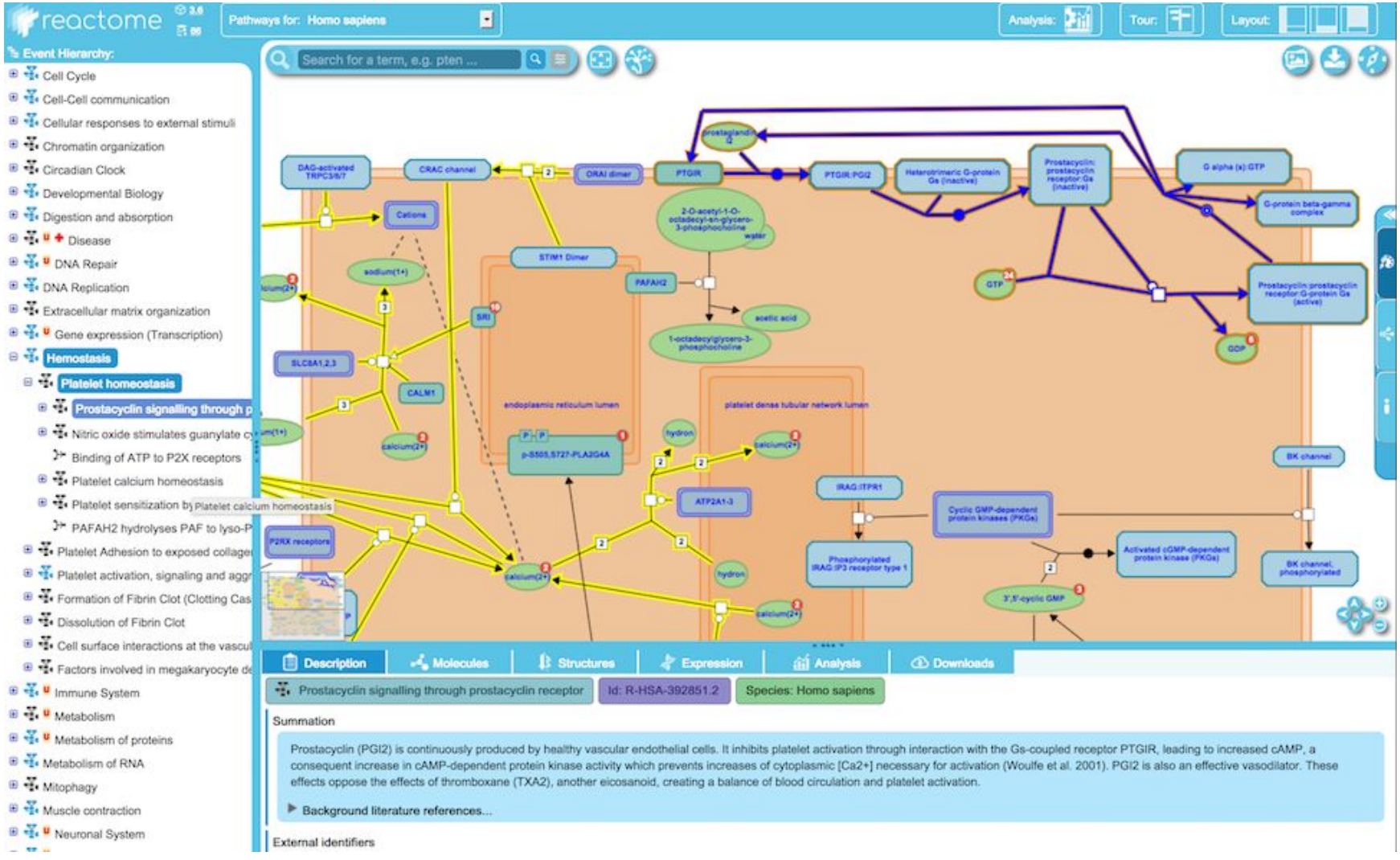




KEGG


TOLL-LIKE RECEPTOR SIGNALING PATHWAY
Peptidoglycan (G+)
Lipoprotein
Lipoarabinomannan (Mycobacteria)
Zymosan (Yeast)
LPS (G-)
Lipopolysaccharide biosynthesis
Flagellar assembly
Flagellin
Imidazoquinolin (anti-viral compounds) ssRNA
Unmethylated CpG DNA
dsRNA
TLR1
TLR2
TLR6
TLR2
LBP
CD14
MD-2
TLR4
TLR5
TLR7/8
CTSK
TLR9
TLR3
Endosome
TOLLIP
MyD88
TIRAP
TRAM
Rac1
PI3K
AKT
PI3K-Akt signaling pathway
FADD
CASP8
Apoptosis
MyD88-dependent pathway
IRAK4
IRAK1
TRAF6
TRIF
TRAF3
MyD88-independent pathway
TAB1
TAB2
TAK1
TRAF6
RIP1
IKK
IKKγ
IKKα
IKKβ
NF-κB signaling pathway
NF-κB
IκBα
Ubiquitin mediated proteolysis
Degradation
p105
Tpl2
MEK1/2
ERK
MAPK signaling pathway
MKK3/6
MKK4/7
p38
JNK
AP-1
DNA
Specific pathway in pDCs
TRAF6
IRAK4
IRAK1
IRF7
IRF5
IKKε
TBK1
IKKα
OPN
IRF3
Inflammatory cytokines
TNFα
IL-1β
IL-6
IL-12
Proinflammatory effects
IL-8
RANTES
MIP-1α
MIP-1β
Chemotactic effects (Nutrophil, Immature DC (NK cell)
Complement and coagulation cascade
Costimulatory molecules
CD40
CD80
CD86
T cell stimulation
Inflammatory cytokines
IFN-α
IFN-β
Antiviral effects
Cytokine-cytokine receptor interaction
Autocrine / Paracrine
IFN-β
IFN-αβR
STAT1
IP-10
MIG
I-TAC
Chemotactic effects (T cell)
JAK-STAT signaling pathway
04620 8/28/17
(c) Kanehisa Laboratories

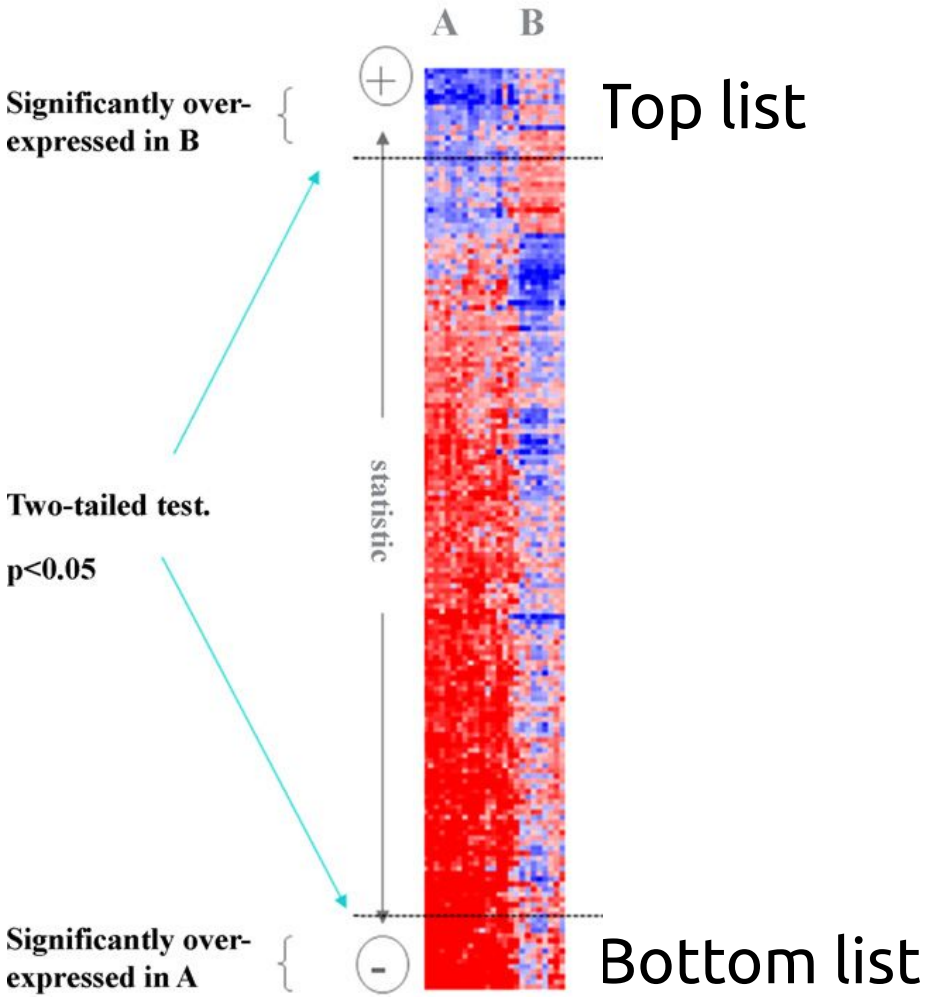
Reactome


reactome
Pathways for: Homo sapiens
Analysis:
Tour:
Layout:
Event Hierarchy:
Cell Cycle
Cell-Cell communication
Cellular responses to external stimuli
Chromatin organization
Circadian Clock
Developmental Biology
Digestion and absorption
Disease
DNA Repair
DNA Replication
Extracellular matrix organization
Gene expression (Transcription)
Hemostasis
Platelet homeostasis
Prostacyclin signalling through p
Nitric oxide stimulates guanylate cy
Binding of ATP to P2X receptors
Platelet calcium homeostasis
Platelet sensitization by
PAFAH2 hydrolyses PAF to lyso-P
Platelet Adhesion to exposed collage
Platelet activation, signaling and aggr
Formation of Fibrin Clot (Clotting Cas
Dissolution of Fibrin Clot
Cell surface interactions at the vascul
Factors involved in megakaryocyte de
Immune System
Metabolism
Metabolism of proteins
Metabolism of RNA
Mitophagy
Muscle contraction
Neuronal System
Search for a term, e.g. pten ...
DAG-activated TRPC3/6/7
CRAC channel
ORAI dimer
PTGIR
PTGIR:PGI2
Heterotrimeric G-protein Gs (inactive)
Prostacyclin; prostacyclin receptor:Gs (inactive)
G alpha (s):GTP
G-protein beta-gamma complex
Cations
STIM1 Dimer
PAFAH2
acetic acid
GTP
GDP
Prostacyclin:prostacyclin receptor:G-protein Gs (active)
SLC8A1,2,3
CALM1
endoplasmic reticulum lumen
platelet dense tubular network lumen
p-S505,S727-PLA2G4A
ATP2A1-3
IRAG:ITPR1
Cyclic GMP-dependent protein kinases (PKGs)
Activated cGMP-dependent protein kinase (PKGs)
BK channel
BK channel, phosphorylated
Phosphorylated IRAG:IP3 receptor type 1
3',5'-cyclic GMP
P2RX receptors
Description
Molecules
Structures
Expression
Analysis
Downloads
Prostacyclin signalling through prostacyclin receptor
Id: R-HSA-392851.2
Species: Homo sapiens
Summation
Prostacyclin (PGI2) is continuously produced by healthy vascular endothelial cells. It inhibits platelet activation through interaction with the Gs-coupled receptor PTGIR, leading to increased cAMP, a consequent increase in cAMP-dependent protein kinase activity which prevents increases of cytoplasmic [Ca2+] necessary for activation (Woulfe et al. 2001). PGI2 is also an effective vasodilator. These effects oppose the effects of thromboxane (TXA2), another eicosanoid, creating a balance of blood circulation and platelet activation.
Background literature references...
External identifiers

Outline

- Introduction
- Over-Representation Analysis (ORA)
- Gene Set Analysis (GSA)
- Network Analysis (NA)

Over-Representation Analysis



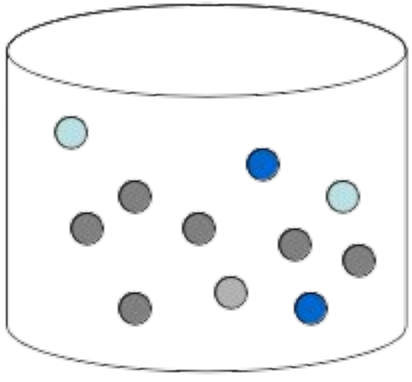
1. List is created using a certain threshold or criteria.
2. For each pathway, input genes that are part of the pathway are counted.
3. This process is repeated for an appropriate background list of genes
4. Every pathway is tested for over- or under-representation in the list of input genes.

Over-Representation Analysis

FatiGO test

Finding significant associ**ati**ons of **G**ene **O**ntology

One Gene List (A)

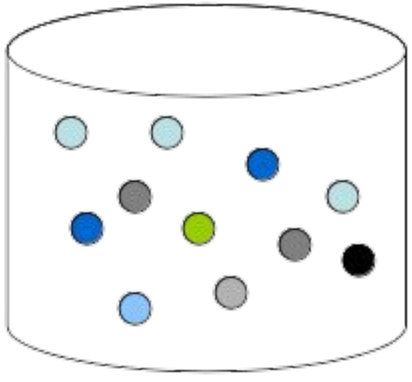


Biosynthesis 60%

Sporulation 20%

Are this two groups of genes carrying out different biological roles?

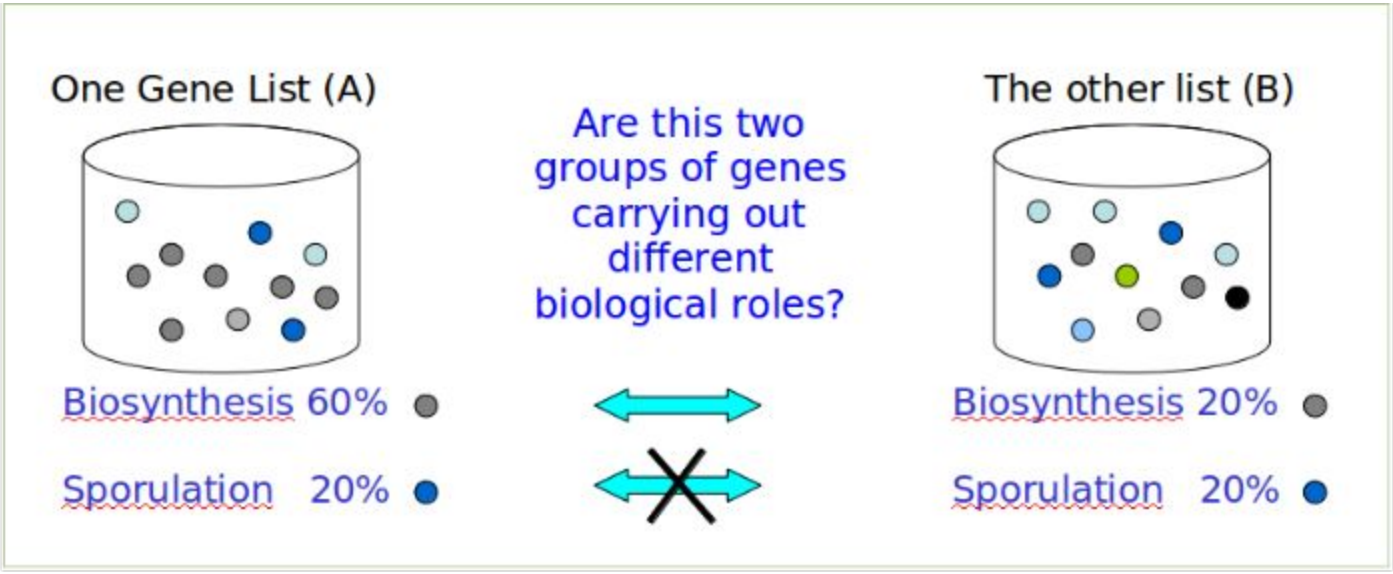
The other list (B)



Biosynthesis 20%

Sporulation 20%

Over-Representation Analysis



Genes in group A have significantly to do with biosynthesis, but not with sporulation.

	A	B
Biosynthesis	**6**	**2**
No biosynthesis	**4**	**8**

We do this for each term (GO, miRNA, Interpro , ...)
Thousand of terms, so Multiple Test Correction is needed!!!

Web tools for ORA

- Babelomics (FatiGO): http://babelomics.bioinfo.cipf.es/
- Panther: http://www.pantherdb.org/
- DAVID: https://david.ncifcrf.gov/
- Reactome: https://reactome.org/
- g:Profiler: https://biit.cs.ut.ee/gprofiler/gost
- FuncAssoacite: http://llama.mshri.on.ca/funcassociate/
- WebGestal: http://www.webgestalt.org/

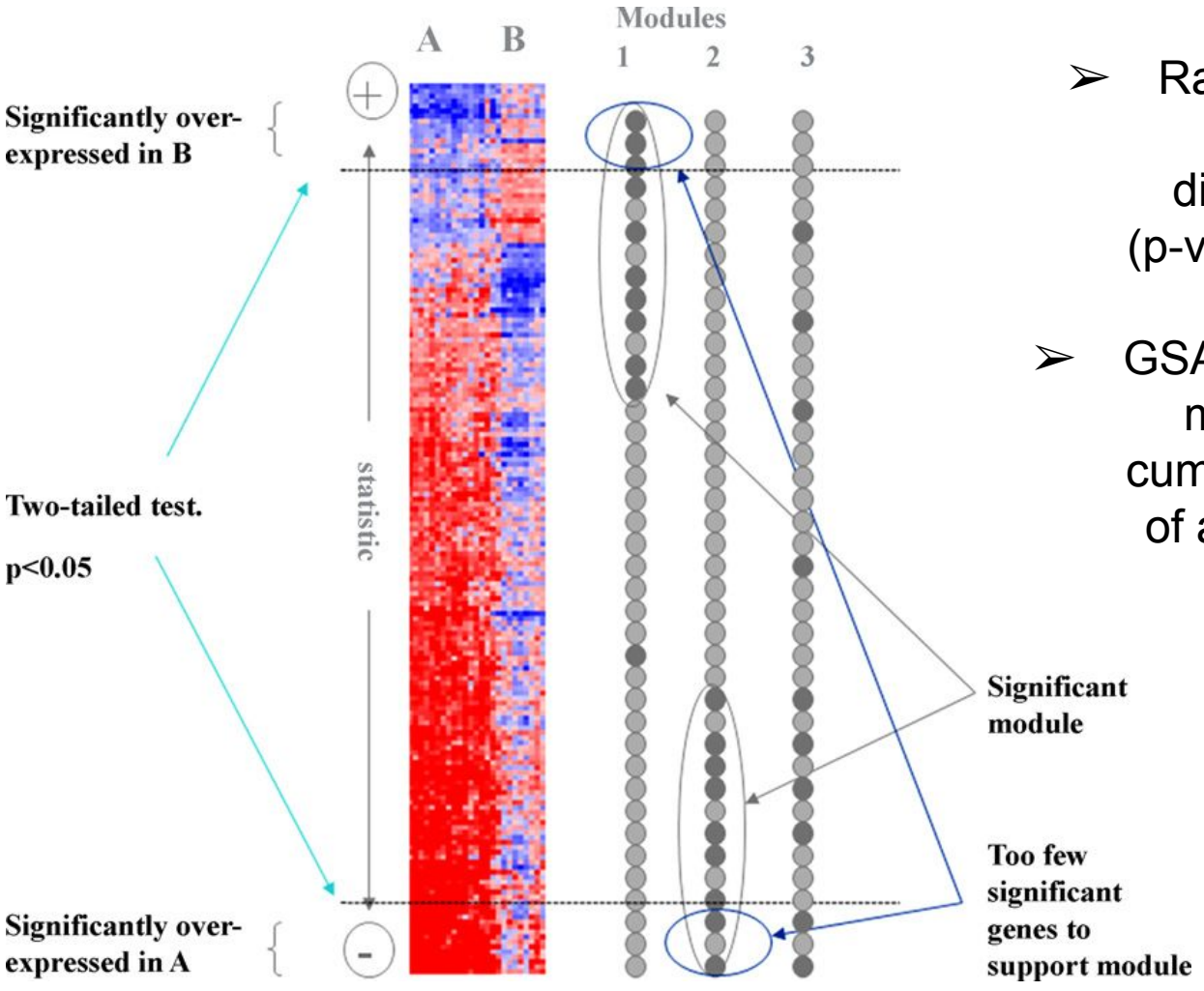
Outline

- Introduction
- Over-Representation Analysis (ORA)
- Gene Set Analysis (GSA)
- Network Analysis (NA)

Web tools for GSA

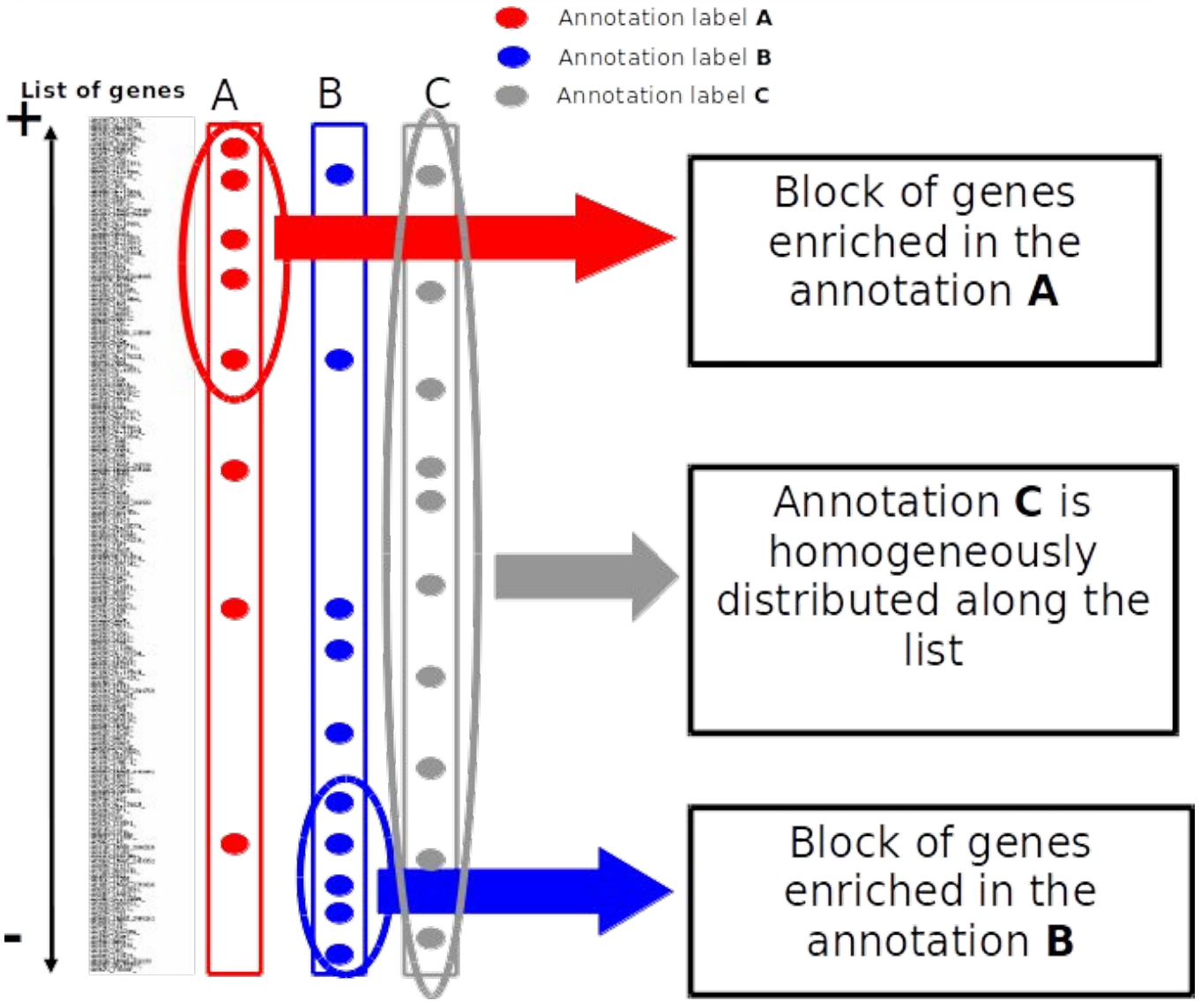
- GSEA (Gene Set Enrichent Analysis): http://software.broadinstitute.org/gsea/index.jsp
- Babelomics (logistic model): http://babelomics.bioinfo.cipf.es/
- Panther: http://www.pantherdb.org/
- FuncAssoacite: http://llama.mshri.on.ca/funcassociate/
- WebGestal: http://www.webgestalt.org/

Gene Set Analysis



- Ranked list of all genes according to their differential expression (p-value, t-statistic, logFC)
- GSA directly tests for gene modules significantly cumulated in the extremes of a ranked list of genes.

Gene Set Analysis



Any question?

Activities

1. Over-representation and GSA exercises:
   http://bioinfo.cipf.es/WODA19/doku.php/bbdd

2. Protein-protein interaction exercises:
   http://bioinfo.cipf.es/WODA19/doku.php/ex_ppi