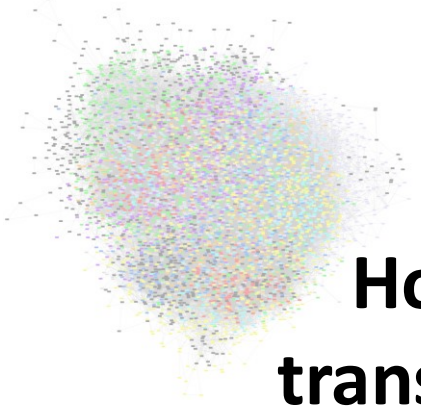
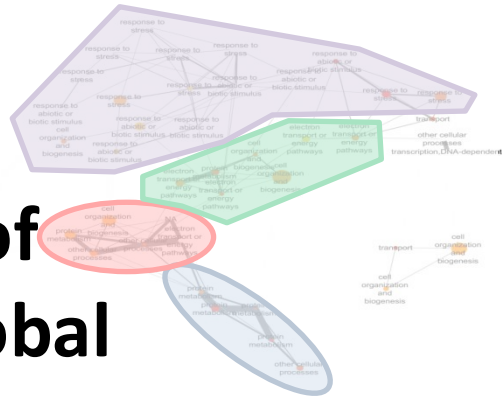
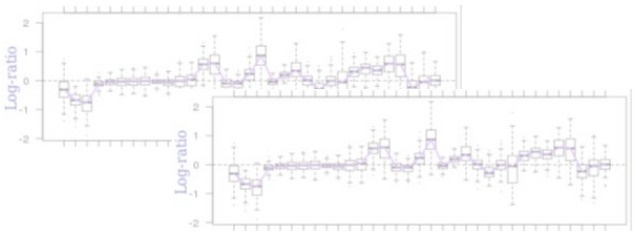


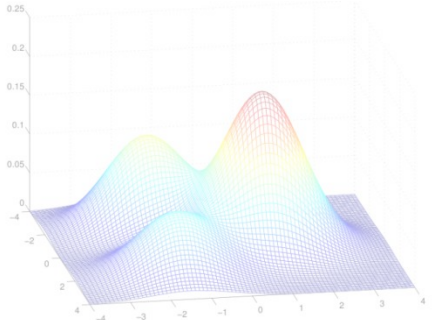


How a statistical meta-analysis of transcriptomic data identifies a global response to stresses

Marie-Laure Martin-Magniette

Group leader of the team Genomic networks at IPS2
Member of the team Statistique and Genome of MIA-Paris

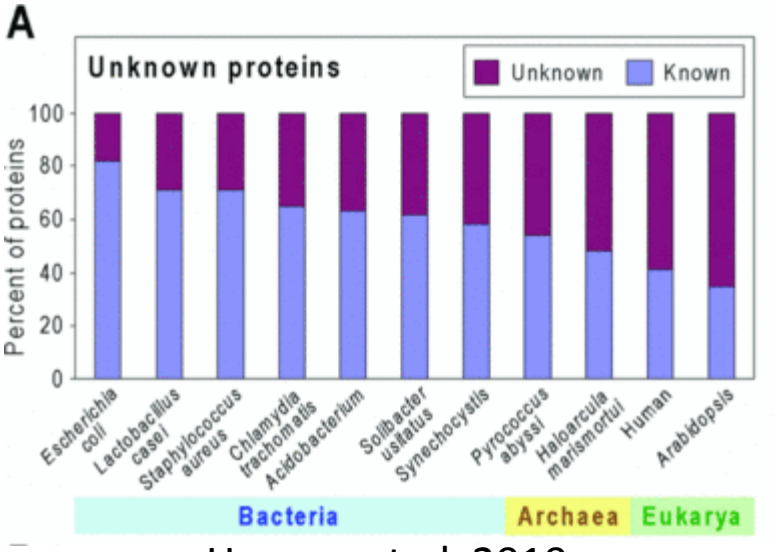




INRA
SCIENCE & IMPACT

IPS2
Institute of Plant Sciences
Paris - Saclay

Functional annotation is really a challenge



Hanson et al. 2010

	2015	2018	2019
Nb of genes	30 251	30 255	30 828
% of genes with a GOSlim annotation and an experimental validation	38 %	42,6%	42,8%

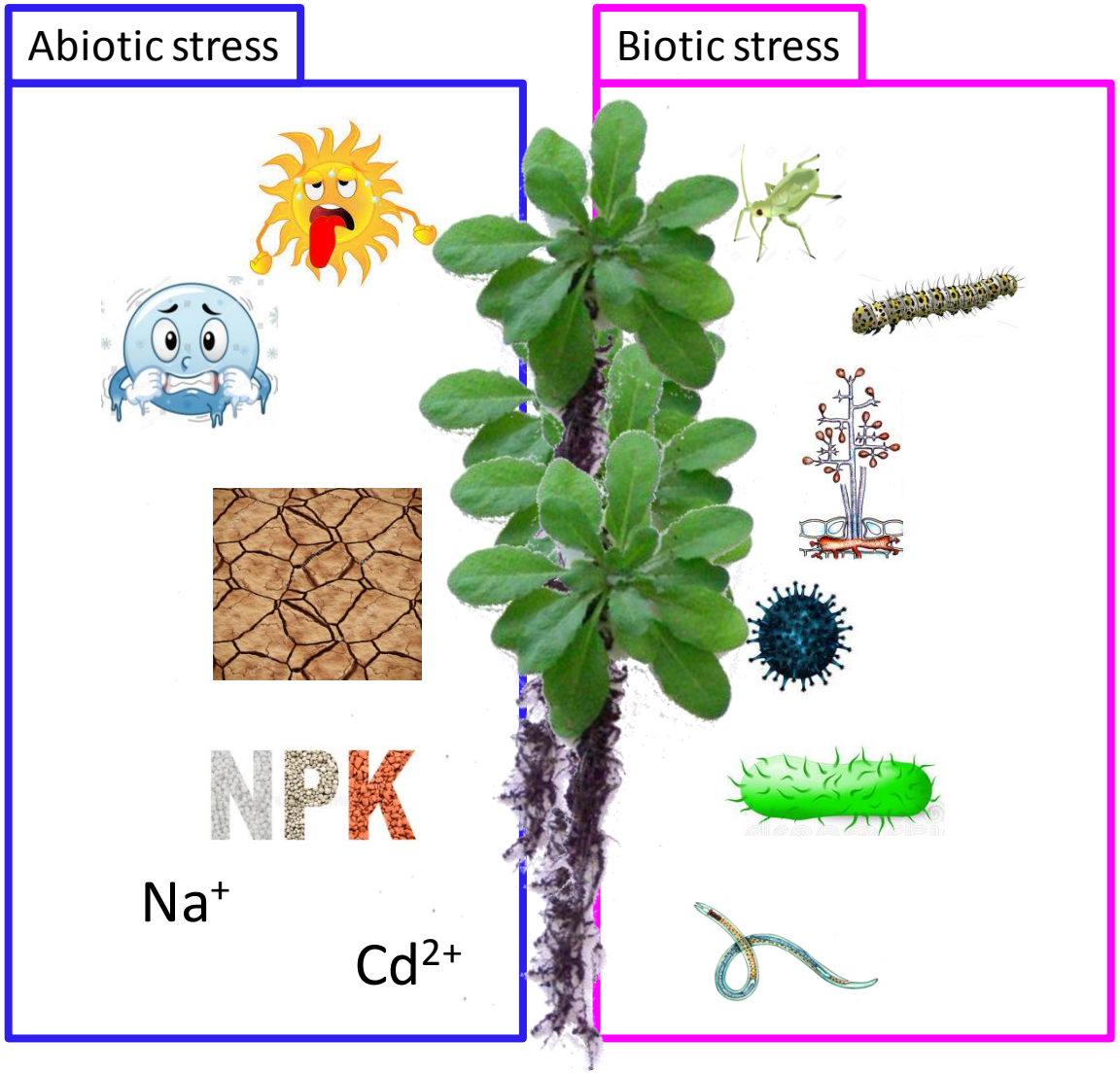
Even for Arabidopsis, few information exists on functional annotation

One gene-one enzyme hypothesis is too naive

Co-expressed genes are good candidates to be involved in a same biological process (Eisen et al, 1998)

Genes are involved in contextual functional modules: for a gene, the module membership changes so that plant achieves or maintains an equilibrium

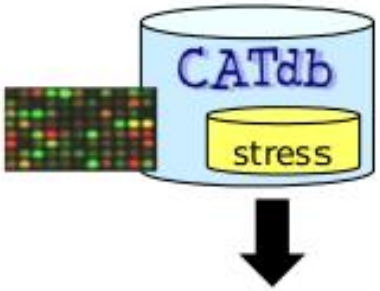
Context under study : stress environement

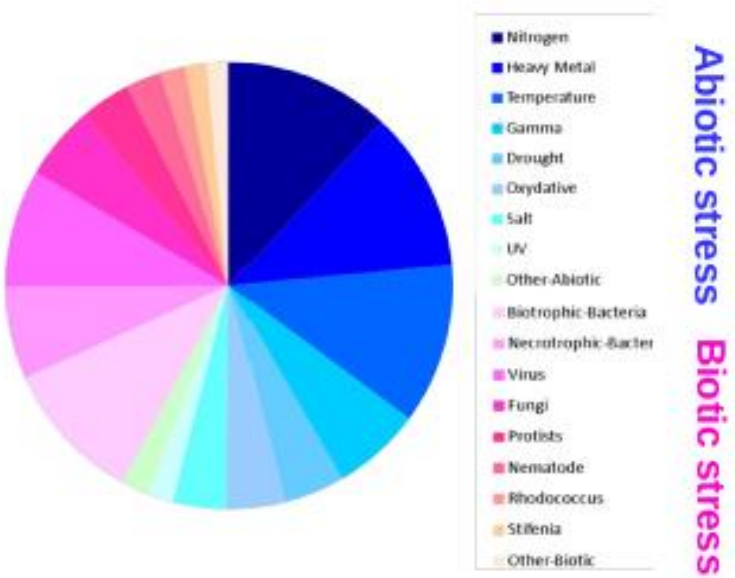


Does a coordinated response to stresses exist ?

A dedicated transcriptomic dataset

Data come from the same plaform
to control some batch effects





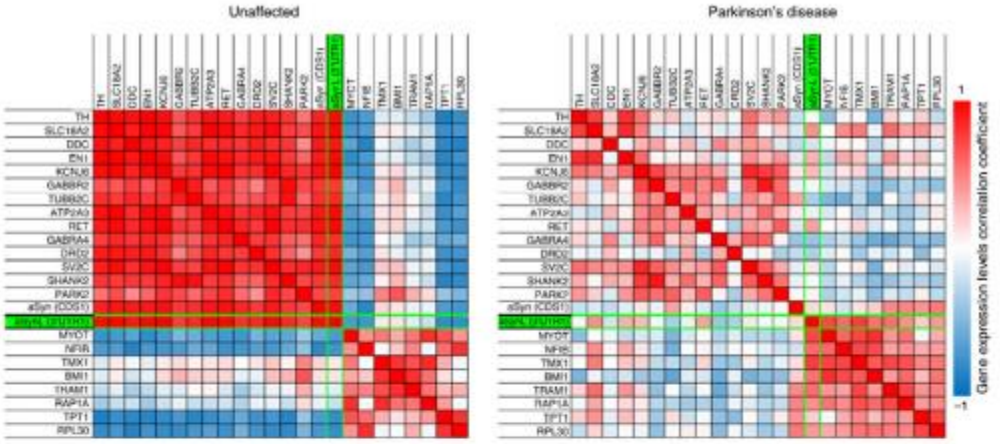
~400 microarray experiments measuring the expression difference between a stress condition and a control condition

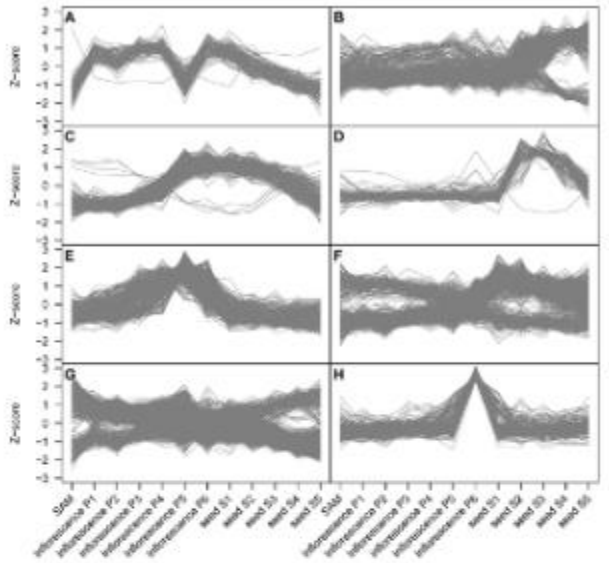
60% of the genes coding proteins are impacted in their transcription

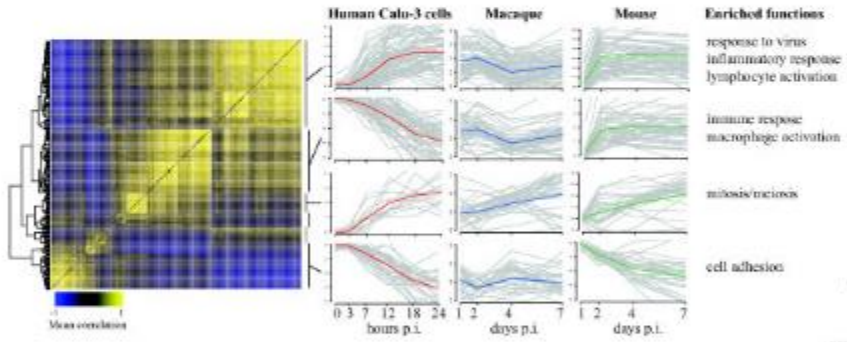
A large number of genes impacted by both types of stress

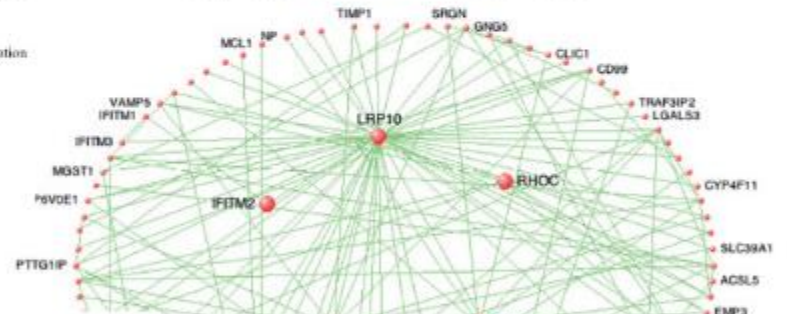
Gene co-expression

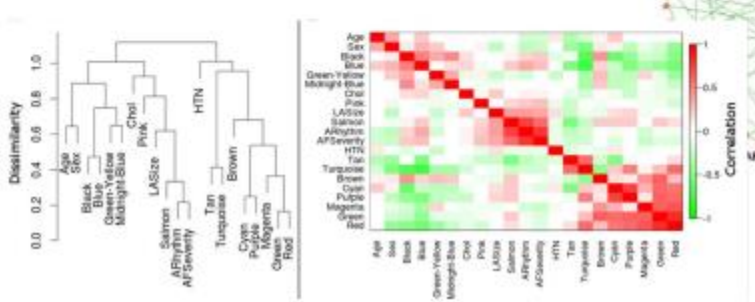
Study patterns of relative gene expression across several conditions



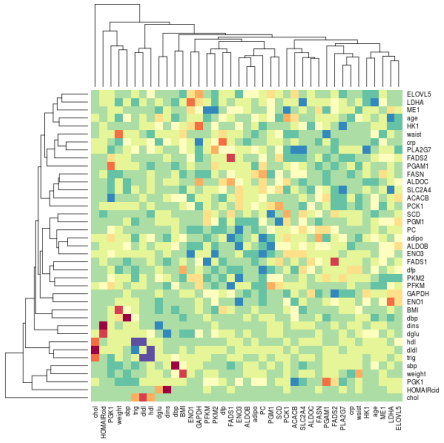




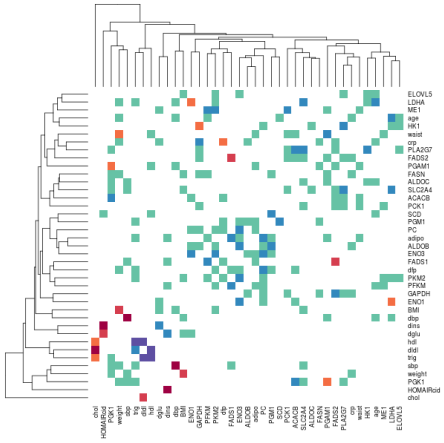




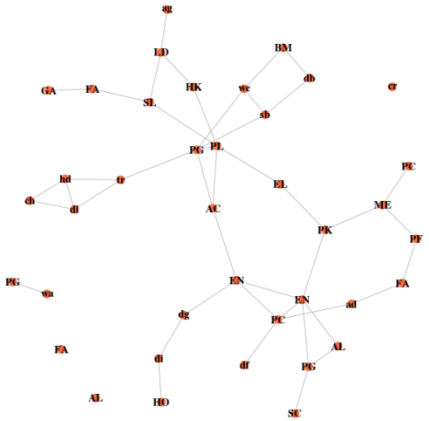
Co-expression based on correlation



Correlation

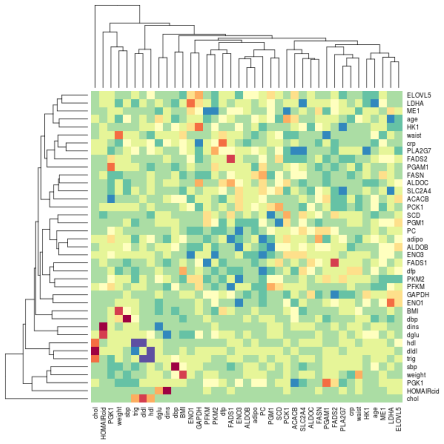


Threshold

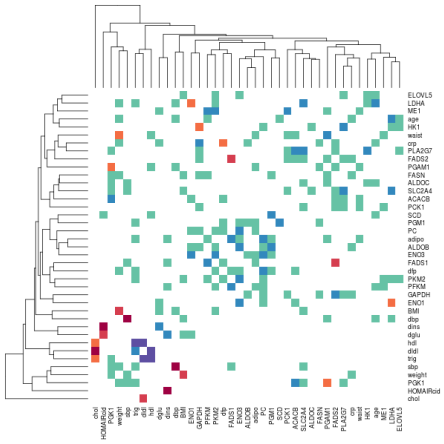


Network

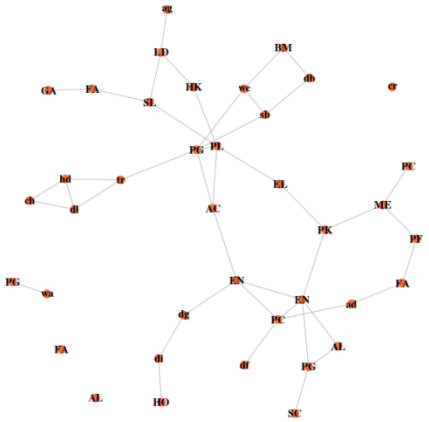
Co-expression based on correlation



Correlation



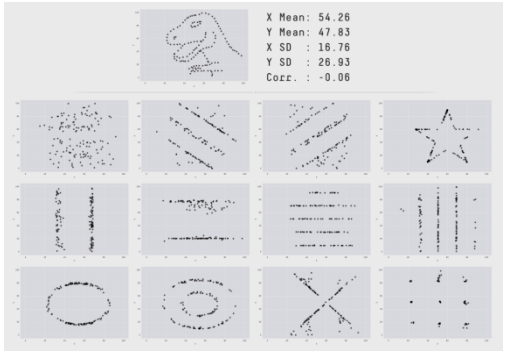
Threshold



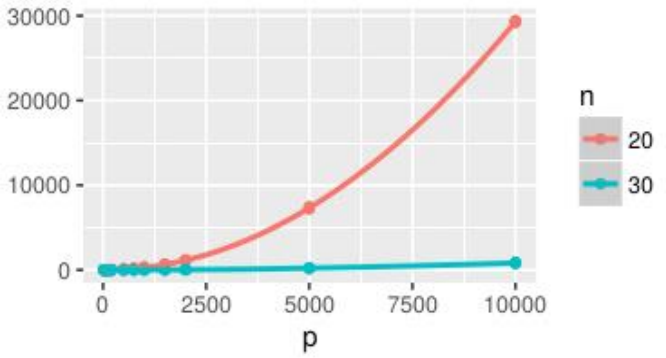
Network

Simple indicators and arbitrarily thresholds can mislead the interpretation

Figures with the same indicators



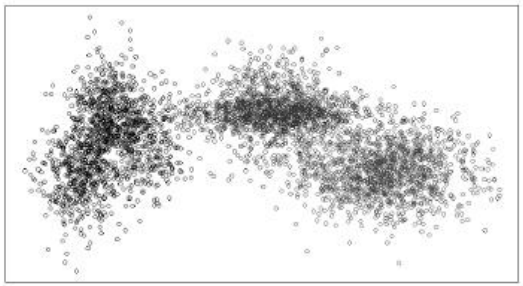
Nb of correlations above a threshold of 0.7



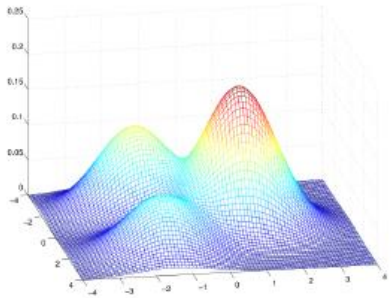
Mixture models

- Global modelling of the whole population
- Individuals are assumed to come from different subpopulations

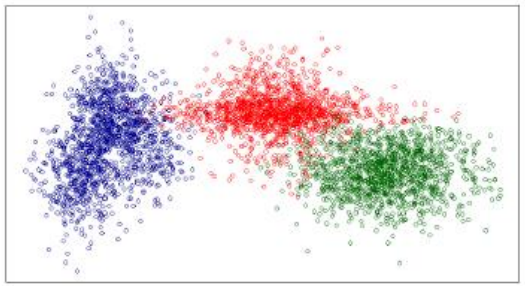
what we observe



the model



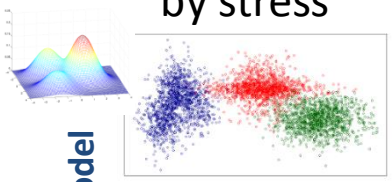
the expected results



- Rigorous framework for choosing a right number of subpopulations
- Each individual is assigned to a subpopulation with a probability.

Coexpression analyses of 18 stress responses

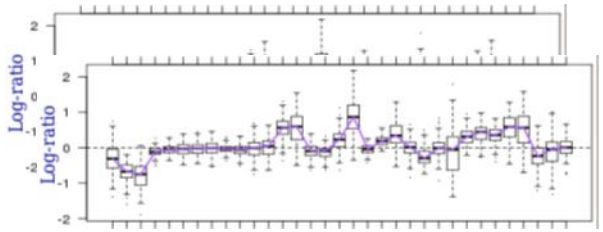
Matrix
{ genes x log-ratios}
by stress



Gaussian Mixture Model

Mathematical Criterion to select the cluster number (BIC)

Classification rule based on conditional probabilities



Stress category	Sample_nb	Gene_nb	Cluster_nb
Drought	17	8167	34
Gamma ray	25	5419	32
Heavy metals	45	10533	57
Nitrogen	46	13807	60
Oxidative stress	16	10027	52
Salt	15	5786	30
Temperature	45	11199	34
UV	7	7903	37
Other abiotic	8	3944	24
Fungi	21	9705	51
Biotrophic bacteria	40	11817	56
Necrotrophic bacteria	26	11030	50
Nematodes	10	7487	29
Oomycetes	14	5591	31
Rhodococcus	7	1965	13
Stifenia	6	1565	17
Virus	33	11685	54
Other biotic	6	3803	20

~700 Clusters of coexpressed genes

Zaag R *et al.,* Nucleic Acids Res, 2015

GEM2Net

http://tools.ips2.u-psud.fr//GEM2NET/

Stress category: VIRUS

# Total genes	# Clusters	Classification rule	# Classified genes	# CATdb projects
11685	54	MFDR	6046	5 >>

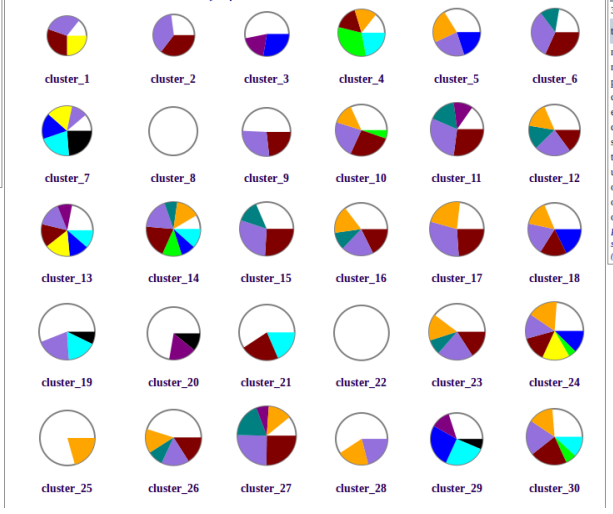
Clustering | **Biological process** | Cellular component | Molecular function | Subcell | Bibliostress | Orphan | Transcription factor | Hormone | Interactome | Networks

The GO Biological process was used to characterize the clusters for the stress category VIRUS. Results of gene set enrichment analyses are displayed as one pie chart per cluster, its size reflecting the total number of genes in the cluster.
While the mouse hovers over a pie chart, the total number of genes in cluster appears in a popup and in the 'Biological process' frame on the right side. As well, the number of genes annotated with a GO term is displayed and the hypergeometric test p-value is mentioned when statistical significance is achieved.

Legends

- DNA_or_RNA_metabolism
- cell_organization_and_biogenesis
- developmental_processes
- electron_transport_or_energy_pathways
- protein_metabolism
- response_to_abiotic_or_biotic_stimulus
- response_to_stress
- signal_transduction
- transcription_DNA_dependent
- transport

click on a Cluster name to see all Functional Analyses for the cluster // click inside a circle to see the Clustered Gene list



Biological process

31 genes in cluster_2

term name	nb genes	p-value	Ref nb
response_to_abiotic_or_biotic_stimulus	17	3.64e-9	3758
response_to_stress	18	1.57e-9	4117
protein_metabolism	1		
cell_organization_and_biogenesis	2		
electron_transport_or_energy_pathways	2		
developmental_processes	4		
signal_transduction	4		
transport	6		
unknown_biological_processes	6		
other_metabolic_processes	15		
other_cellular_processes	16		
other_biological_processes	20		

Ref nb: number of genes annotated with the term in the reference set
(see documentation)

Stress category: VIRUS

# Total genes	# Clusters	Classification rule	# Classified genes	# CATdb projects
11685	54	MFDR	6046	5 >>

cluster_49

# Protein-protein interactions	# TF-target interactions
42	0

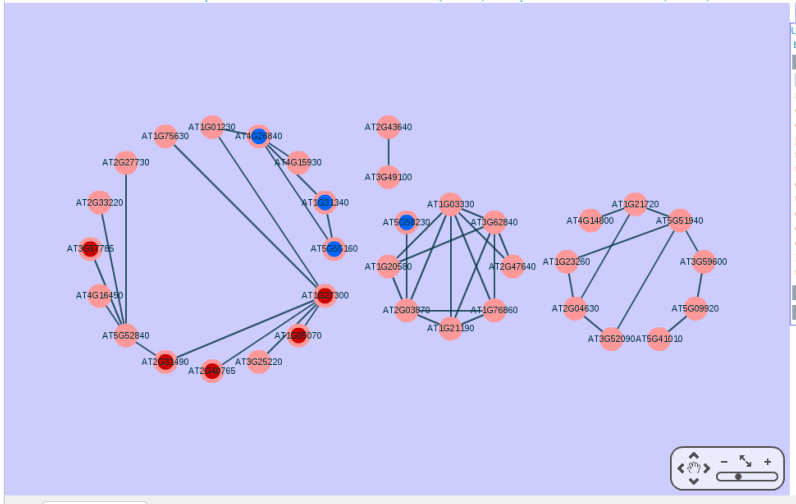
Clustering | Biological process | Cellular component | Molecular function | Subcell | Bibliostress | Orphan | Transcription factor | Hormone | Interactome | **Networks**

Networks of Protein-protein interactions or Target genes of Transcription factors (TFs) are shown for a selected cluster. By default, all protein interactions (experimental and predicted interactomes), as well as confirmed links of TFs to their targets are displayed for gene accessions inside the selected cluster. Out-cluster interactions can be seen on option. Functional annotation is available to charaterize nodes. On the right frame, Filters are provided to view only nodes of the selected term(s). Additional information is available on the bottom side by clicking on a node or an edge.
Notice that this is a beta-test version

Select a cluster: cluster_49 | *FUNCTIONAL ANNOTATION:* Transcription Factor | Hormone: All Hormone | Orphan

PROTEIN INTERACTOMES — *in-Cluster interactions:* All interactions | Confirmed interactions — *options:* Self interactions | out-Cluster interactions (confirmed)

TARGETS of TRANSCRIPTION FACTORS — *in-Cluster interactions:* Confirmed interactions — *option:* out-Cluster interactions (confirmed)



Filter | Search | Save

Use filters to view nodes of the selected item(s)
Filter by GO terms

BIOLOGICAL PROCESS

Terms	Pvalue
response_to_stress	
other_cellular_processes	
other_metabolic_processes	
protein_metabolism	
response_to_abiotic_or_biotic_stimulus	
unknown_biological_processes	
cell_organization_and_biogenesis	
transcription_DNA_dependent	
developmental_processes	
electron_transport_or_energy_pathways	
other_biological_processes	
DNA_or_RNA_metabolism	
transport	

CELLULAR COMPONENT

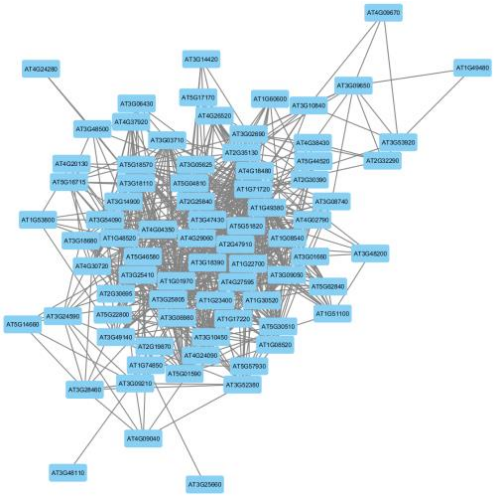
MOLECULAR FUNCTION

Layout: Circle | Node Labels

From stress coexpression clusters to stress coregulation gene network

Integration

Edge = occurrence of coexpression
{1, …, 18}

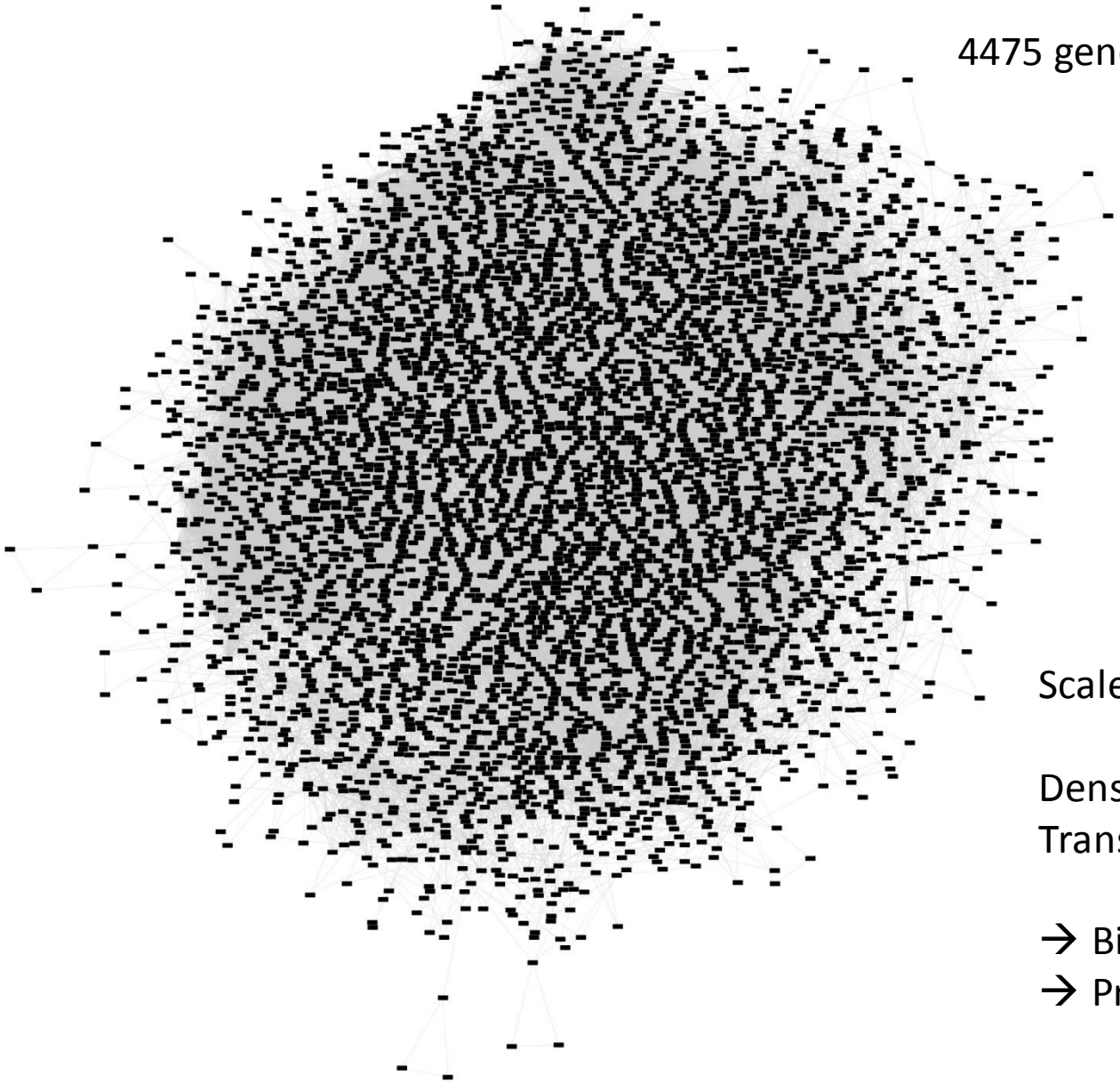


Clean-up

Edge = occurrence of coexpression
{**3**, …, 18}

Coregulation network

Arabidopsis stress co-regulation network



4475 genes and 56487 co-regulation links

86% of the co-regulation links are supported by both types of stress

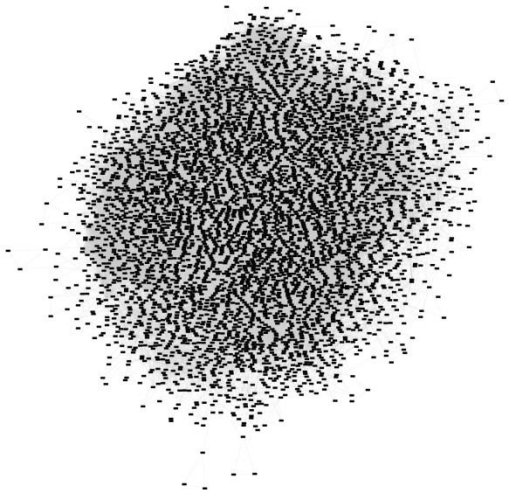
Scale-free network

Density = 0.006
Transitivity= 0.54

→ Biological network
→ Presence of gene clusters

The backbone of plant stress response

4475 genes

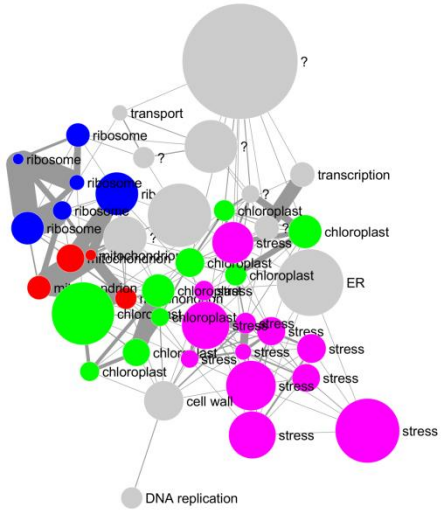


SBM

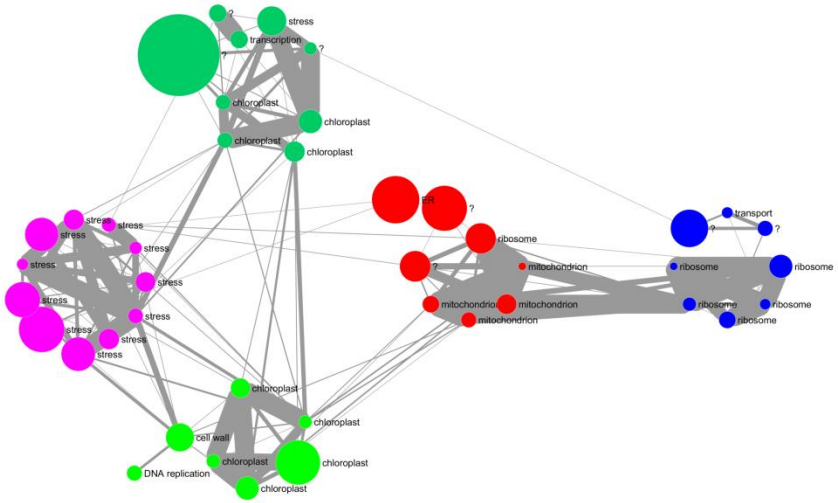
+ cluster annotation

3 months of intensive calculations

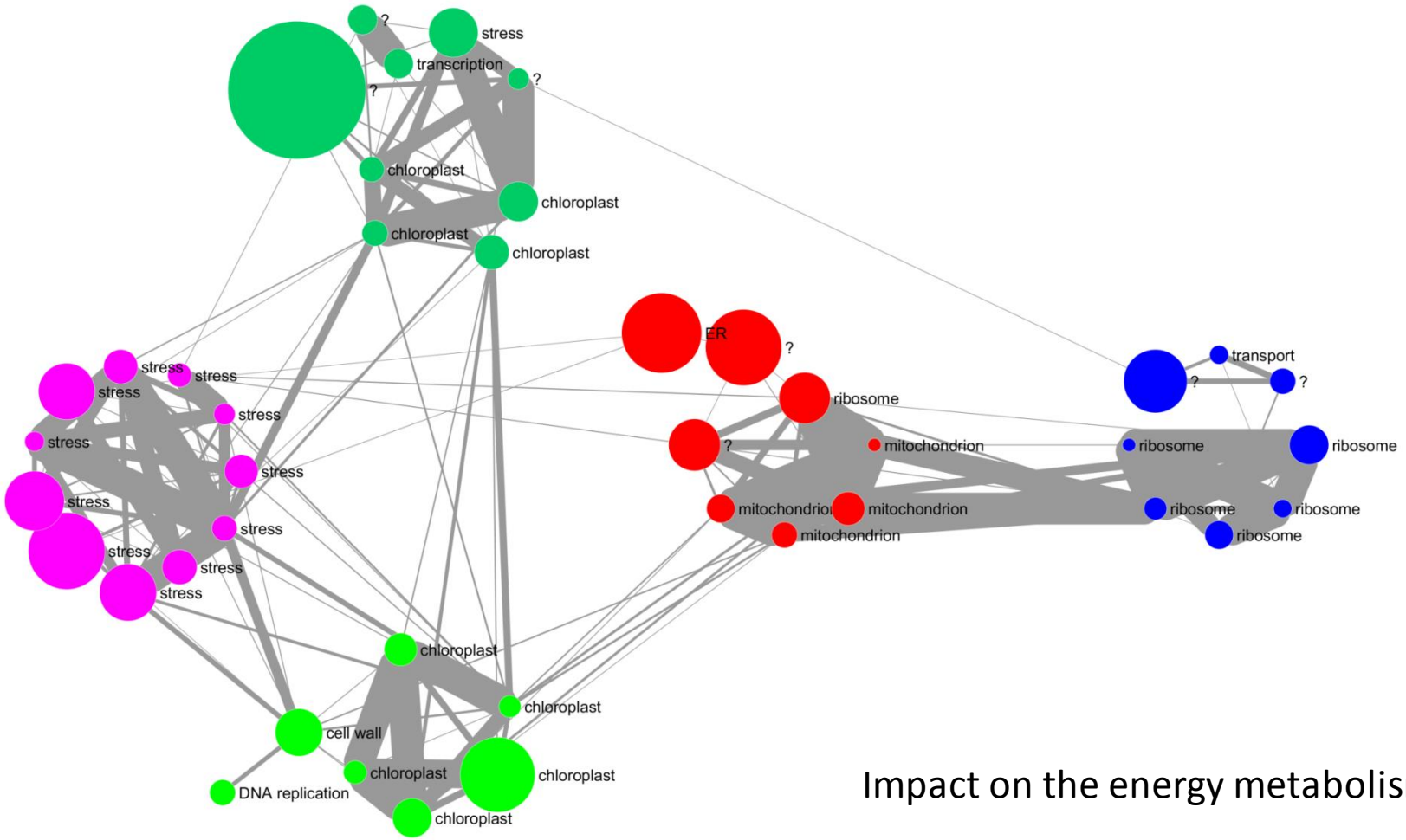
2674 genes in 43 stable clusters



Ward clustering based on the shortest paths



The backbone of plant stress response



Impact on the energy metabolism

ENERGY

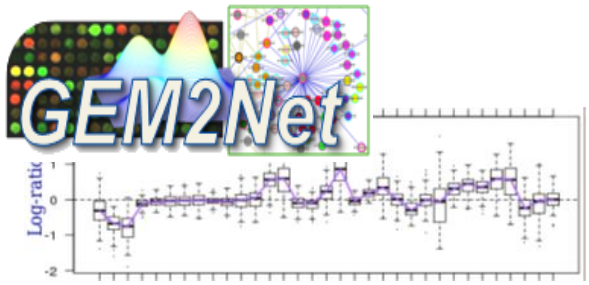
signal

source

Conversion
ATP synthesis

sink

Conclusions



Transcriptome meta-analysis

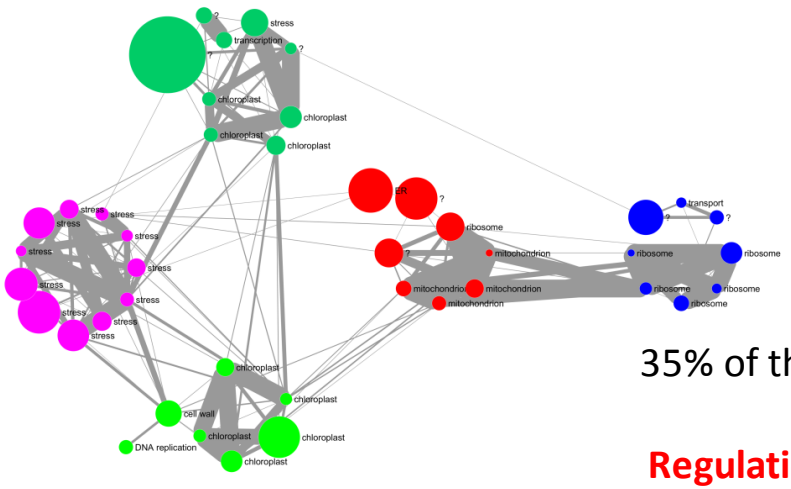
STRESS

PHOTOSYNTHESIS

ATP SYNTHESIS

TRANSLATION

- Identification of a common response to stress
- Involvement of the energy organelles and translation
- Hierarchical organization along the energy gradient
- A new genomic resource to predict gene functions



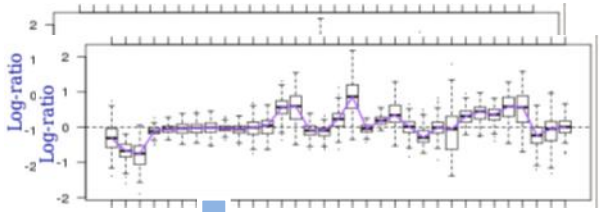
35% of the genes annotated

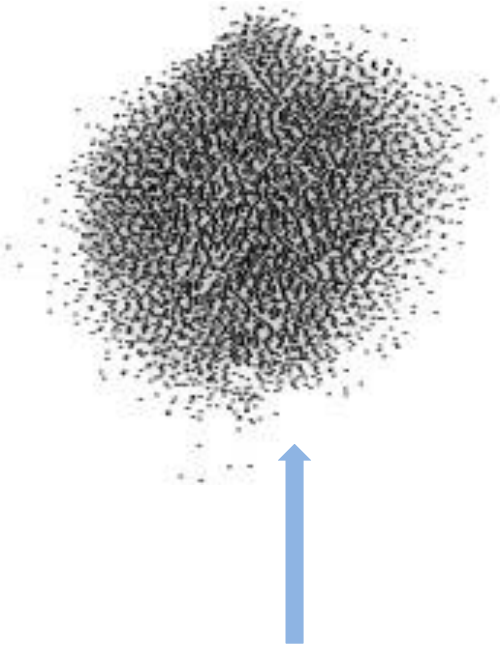
ribosome

Regulations of translation?

New genomic ressource in constant evolution

Co-expression based on microarray

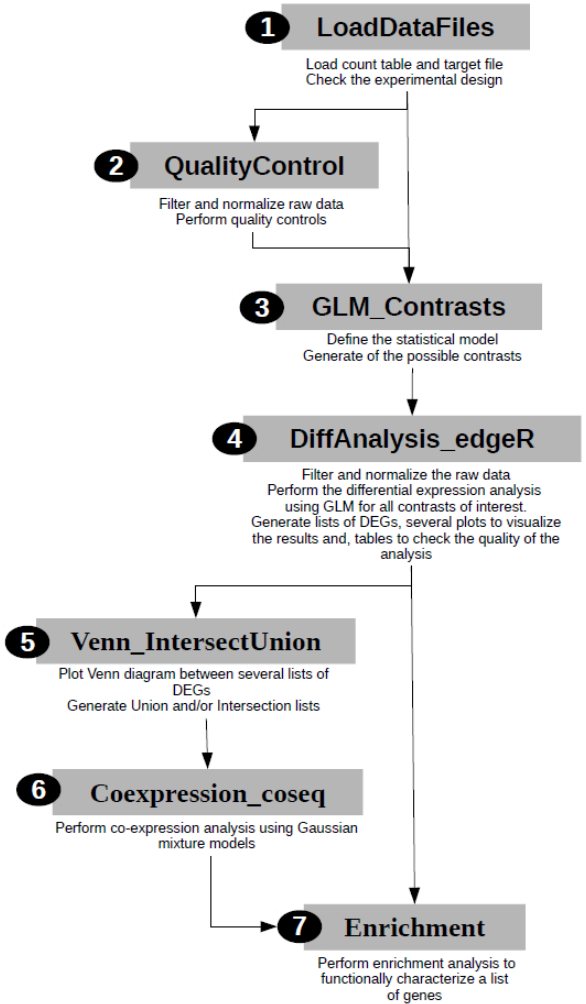




Co-expression based on RNA-seq data

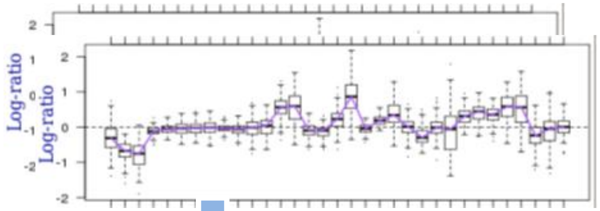
DiCoExpress for
multifactorial RNAseq experiments

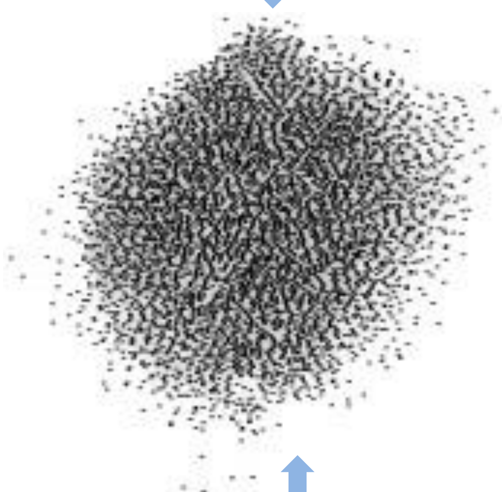
(PhD of Ilana Lambert)



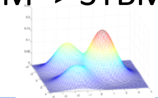
New genomic ressource in constant evolution

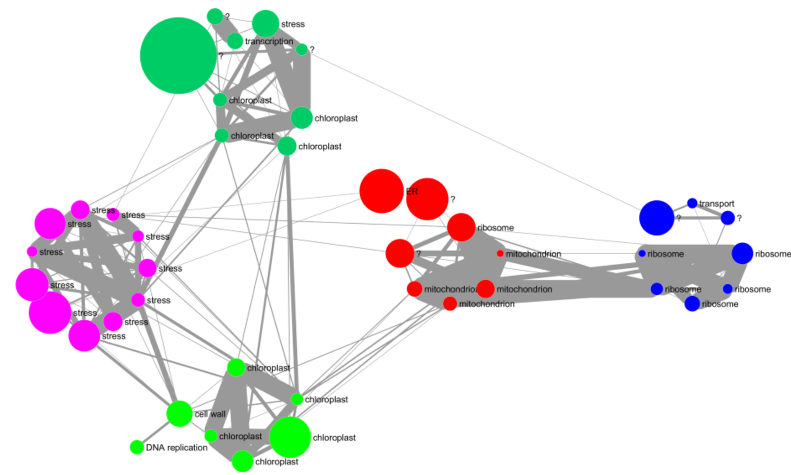
Co-expression based on microarray





SBM -> STBM





Model taking edge information into account to refine the modules (collab. P. Latouche)

Co-expression based on RNA-seq data

Acknowledgement

Bioinfo

Rim Zaag
Jean-Philippe Tamby
Cecile Guichard
Zakia Tariq
Véronique Brunaud
Ilana Lambert
Stefano Colella

Biology

The transcriptomic platform

Sebastien Aubourg
Jean-Pierre Renou

Etienne Delannoy
(co-coordinator of this project)

Stat

Gilles Celeux
Cathy-Maugis-Rabusseau
Andrea Rau
Tristan Mary-Huard
Pierre Latouche
Guillem Rigaill