GEM2Net a module of CATdb from Gene Expression Modeling to Networks in Plants

IPS2 Institute of Plant Science Paris Saclay
V. Brunaud
Team Genomic Networks (ML Martin-Magniette)

Bioinformatic and Transcriptomic PF

POPS
Plateforme transcriptOmique
IPS2

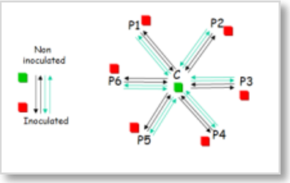
LLOYD'S REGISTER QUALITY ASSURANCE · ISO9001

CATdb
a Complete Arabidopsis Transcriptome database

GEO Gene Expression Omnibus
NCBI

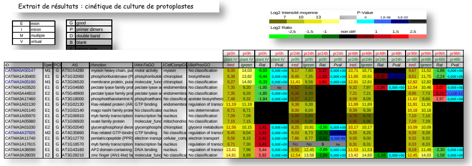
Biological question

Experimental design



Libraries-Sequencing
Hybridizations

Bioinformatic and Statistical analyses



Extrait de résultats : cinétique de culture de protoplastes

E	exon
I	intron
M	multiple
V	virtual

G	good
P	primer dimers
D	double band
B	blank
	wrong size

Log2 Intensité moyenne 7 10 13
P-Value 0 1,E-08 5,E-02
Log2 Ratio -2,5 -1,5 -1 non diff 1 1,5 2,5

ID	Spec	PCF	At1	1fonction	1Mol FoGO	1CellCompG	1BioProcGO
CATMA5A50147	M1	G	AT5G54280	myosin heavy chain, put	motor activity	myosin	No classification
CATMA1A30400	E1	G	AT1G32060	phosphoribulokinase (P	phosphoribulok	chloroplast	biosynthesis
CATMA2A05190	M1	G	AT2G06520	membrane protein, puta	molecular_func	chloroplast	No classification
CATMA1A03520	E1	G	AT1G04680	pectate lyase family prot	pectate lyase a	endomembra	No classification
CATMA5A44810	E2	G	AT5G48900	pectate lyase family prot	pectate lyase a	endomembra	No classification
CATMA3A17460	E1	G	AT3G18000	phosphoethanolamine N	phosphoethand	No classifica	acetate biosynthesis
CATMA1A01130	E1	G	AT1G02130	Ras-related protein (AR	GTP binding	endomembra	regulation of transcr
CATMA1A01140	E1	G	AT1G02140	mago nashi family prote	No classificatio	nucleus	sex determination
CATMA1A05970	E1	G	AT1G06910	myb family transcription	transcription fa	nucleus	No classification
CATMA1A05980	E1	G	AT1G06920	ovate family protein	molecular_func	mitochondrio	No classification
CATMA3A01030	E2	G	AT3G02040	glycerophosphoryl diest	glycerophosph	chloroplast	glycerol metabolism
CATMA4A37505	M2	G	AT4G35860	Ras-related GTP-bindin	GTP binding	No classifica	regulation of transcr
CATMA3A11740	E1	G	AT3G12770	pentatricopeptide (PPR)	electron transp	cellular_comp	electron transport
CATMA1A17615	E1	G	AT1G18570	myb family transcription	transcription fa	nucleus	regulation of transcr
CATMA1A36590	E1	G	AT1G43160	AP2 domain-containing	DNA binding	nucleus	regulation of transcr
CATMA3A28035	E2	G	AT3G28210	zinc finger (AN1-like) fa	molecular_func	No classifica	No classification

	pr0h	pr0h	pr0h	pr0h	pr24h	pr24h	pr24h	pr24h	pr48h	pr48h	pr48h	pr48h	pr96h	pr96h	pr96h	pr96h
	plant IV	plant IV	plant IV	plant IV	pr0h	pr0h	pr0h	pr0h	pr24h	pr24h	pr24h	pr24h	pr48h	pr48h	pr48h	pr48h
ID	Ired	Igreen	Rat	Pval	Ired	Igreen	Rat	Pval	Ired	Igreen	Rat	Pval	Ired	Igreen	Rat	Pval
CATMA5A50147	8,00	14,03	-6,03	0,00E+00	9,17	7,80	1,37	0,00E+00	10,71	10,71			9,67	11,52	-1,86	0,00E+00
CATMA1A30400	8,38	13,82	-5,44	0,00E+00	8,45	7,35	1,10	0,00E+00	11,65	10,91	0,74	1,57E-03	9,51	11,75	-2,24	0,00E+00
CATMA2A05190	9,27	14,60	-5,33	0,00E+00	11,41	9,56	1,84	0,00E+00	12,81	12,81			12,82	12,82		
CATMA1A03520	7,35	9,20	-1,85	No	6,62	6,62			9,32	7,97	1,34	0,00E+00	12,54	10,46	2,07	0,00E+00
CATMA5A44810	7,35	9,20	-1,85	0,00E+00	6,30	6,30			7,21	7,21			8,95	7,73	1,21	1,11E-11
CATMA3A17460	7,48	9,32	-1,84	0,00E+00	6,27	6,27			8,09	8,09			11,62	8,66	2,97	0,00E+00
CATMA1A01130	11,19	11,19			9,38	9,38			11,89	11,89			11,91	11,91		
CATMA1A01140	8,71	8,71			8,00	8,00			9,39	9,39			10,18	10,18		
CATMA1A05970	7,09	7,09			5,89	5,89			6,78	6,78			7,10	7,10		
CATMA1A05980	7,15	7,15			6,06	6,06			6,96	6,96			7,29	7,29		
CATMA3A01030	11,56	10,15	1,41	0,00E+00	8,25	10,81	-2,56	0,00E+00	10,17	10,17			10,09	10,09		
CATMA4A37505	9,45	8,04	1,41	0,00E+00	8,03	8,73	-0,70	5,77E-04	9,25	9,25			9,39	9,39		
CATMA3A11740	9,56	8,15	1,41	0,00E+00	7,86	8,77	-0,92	6,52E-09	9,00	9,00			9,32	9,32		
CATMA1A17615	8,71	7,30	1,41	0,00E+00	No	No	$	No	8,31	8,31			8,33	8,33		
CATMA1A36590	13,41	7,96	5,44	0,00E+00	9,55	12,45	-2,89	0,00E+00	11,53	11,53			9,19	11,49	-2,30	0,00E+00
CATMA3A28035	14,81	8,89	5,92	0,00E+00	12,54	13,58	-1,04	0,00E+00	13,42	14,60	-1,18	0,00E+00	12,45	14,03	-1,58	0,00E+00

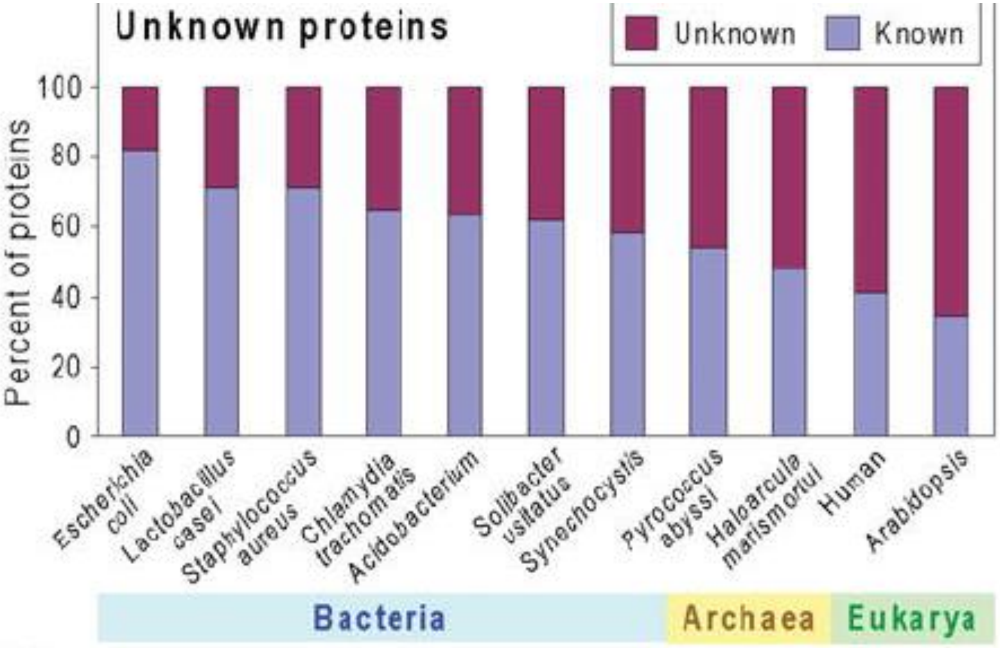
Differential Gene expression

Which gene is regulated?
Information on gene function

Functional Annotation

A huge "orphan of function" gene space

- orphan genes = genes without homologs with a known function (Fukushi et al., 2003).
- 20% to 40% of the predicted genes for completely sequenced eukaryotic organisms have no assigned function (Hanson et *al*., 2010).



For Arabidopsis genome

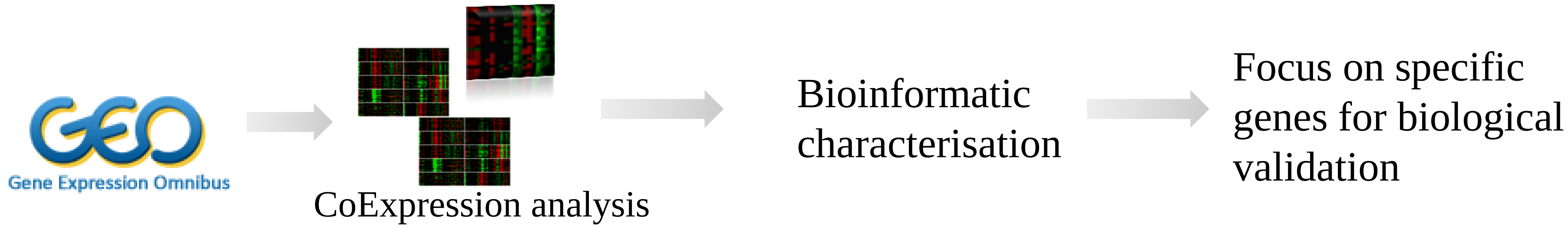
- **Only 16% of genes have a validated function**
- **More than 5000 Arabidopsis genes are still orphans (TAIRv10)**

Classical analyses by sequence similarity

Based on a comparison of protein sequences to identify structural similarities

- Some sequences with a low similarity may share a same function (Galperin et al, 1998)
- A high similarity does not guarantee a functional similarity (Tian et al, 2003)
- Protein sequence comparison gives information about the biochimical function (Nehrt et al, 2011) → common proteic motif/domain

Classical analyses of co-expression



Gene Expression Omnibus

CoExpression analysis

Bioinformatic characterisation

Focus on specific genes for biological validation

Transcriptomic studies seem relevant approaches since co-expressed genes are often involved in a same biological process

- Data are generally extracted from international repositories
- With heterogeneous data in terms of acquisition and preprocessing

Co-expression usually measure a correlation value

- Comparison by Pearson correlation based on “gene pairs”
- Difficult to interpret since the number of gene pairs is large

Local point of view of a complex question since genes work in modules

Integration of omics data

Integrating various resources of omics data improves the success of prediction (Radiovojac et al, 2013) ‘Guilt by association’

But various sources of heterogeneity exist

- Data are qualitative or quantitative
- Available information describes the biological entities or their relationships
- Observations are obtained with various techniques / workflows

Not obvious links between these various data

e.g. Protein Interactome (PPI) by Arabidopsis Interactome Mapping consortium (2011)

- only 10% have been studied
- high rate of false positive

GEM2Net: Goals

Goals: from Gene Expression Modeling TO gene Network to investigate Arabidopsis thaliana stress responses

→ Agronomic challenge to characterize genetic variants resistant to biotic/abiotic attacks

CATdb GEM2Net

GEM2Net: Goals & methods

Goals: From gene expression modeling to gene network to investigate Arabidopsis thaliana stress response

CATdb GEM2Net

Methods: original features

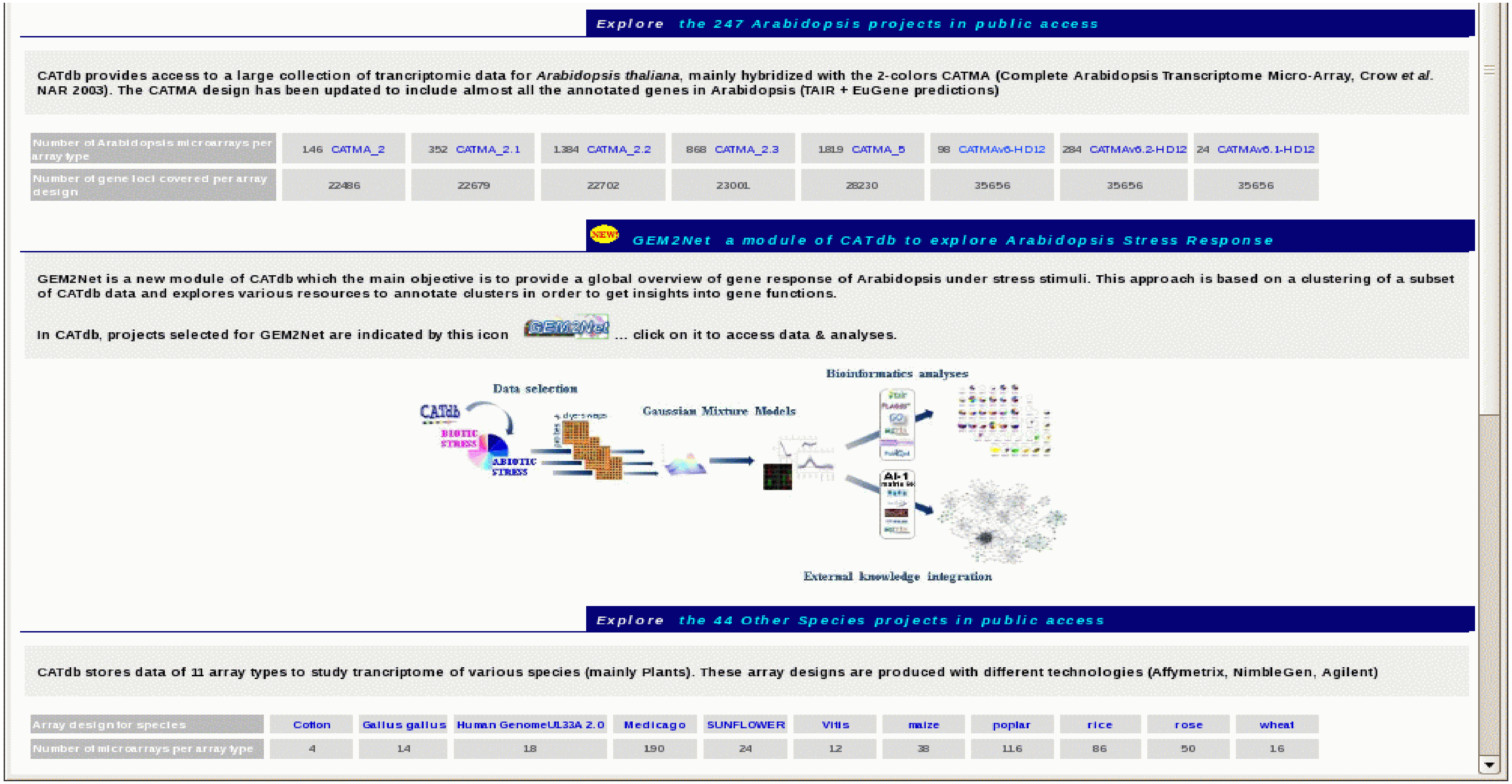
- homogeneous and dedicated transcriptomic data in CATdb
- classification of several CATdb projects in biotic/abiotic stress
- method of clustering: model based method
- integration of various resources to improve the functional inference

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

4613 diff. samples on 25 species

→ 1657 treated (183 exp. Factors)

→ 32 organs, 132 development stages



Explore the 247 Arabidopsis projects in public access

CATdb provides access to a large collection of trancriptomic data for *Arabidopsis thaliana*, mainly hybridized with the 2-colors CATMA (Complete Arabidopsis Transcriptome Micro-Array, Crow *et al.* NAR 2003). The CATMA design has been updated to include almost all the annotated genes in Arabidopsis (TAIR + EuGene predictions)

Number of Arabidopsis microarrays per array type	146 CATMA_2	352 CATMA_2.1	1384 CATMA_2.2	868 CATMA_2.3	1819 CATMA_5	98 CATMAv6-HD12	284 CATMAv6.2-HD12	24 CATMAv6.1-HD12
Number of gene loci covered per array design	22486	22679	22702	23001	28230	35656	35656	35656

NEW! GEM2Net a module of CATdb to explore Arabidopsis Stress Response

GEM2Net is a new module of CATdb which the main objective is to provide a global overview of gene response of Arabidopsis under stress stimuli. This approach is based on a clustering of a subset of CATdb data and explores various resources to annotate clusters in order to get insights into gene functions.

In CATdb, projects selected for GEM2Net are indicated by this icon GEM2Net ... click on it to access data & analyses.

Explore the 44 Other Species projects in public access

CATdb stores data of 11 array types to study trancriptome of various species (mainly Plants). These array designs are produced with different technologies (Affymetrix, NimbleGen, Agilent)

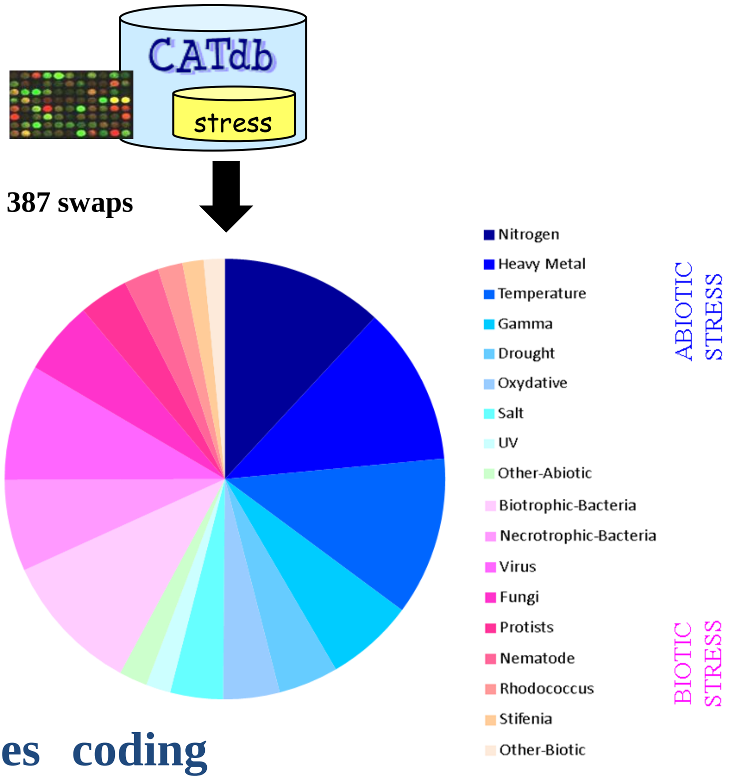
Array design for species	Cotton	Gallus gallus	Human GenomeU133A 2.0	Medicago	SUNFLOWER	Vitis	maize	poplar	rice	rose	wheat
Number of microarrays per array type	4	14	18	190	24	12	38	116	86	50	16

Transcriptomic Dataset

CATdb : experiments implicated into stresses

☐ Large and homogeneous transcriptome resource generated by the POPS platform and available in CATdb (Gagnot et al., NAR 2008).

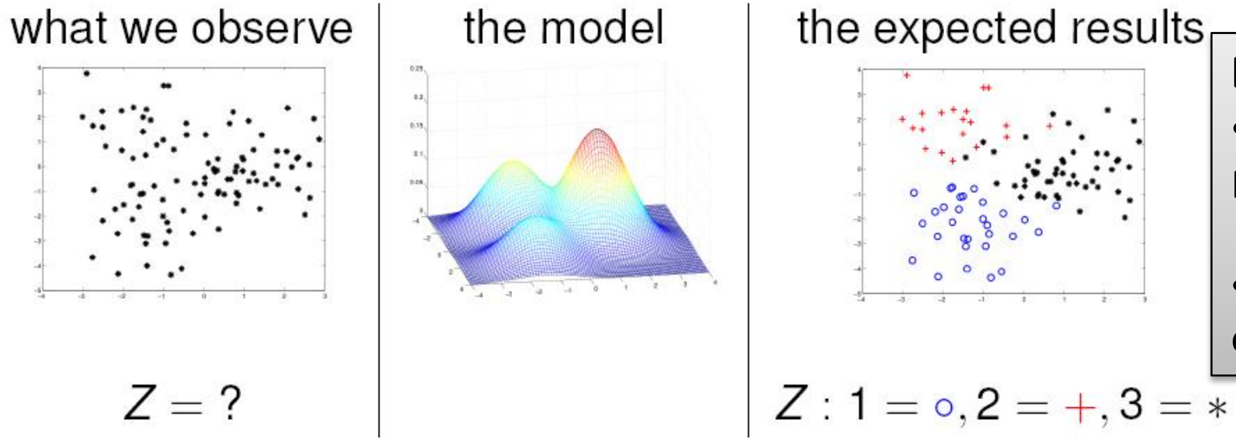
☐ ~ 400 conditions dedicated to stresses
9 biotic and 9 abiotic stress categories



- Based on differential analyses, **60% of the genes coding proteins** have their transcription impacted directly or not by a stress

- Large overlap of impacted genes between biotic and abiotic stresses

Co-expression analysis by mixture models



Data-driven method

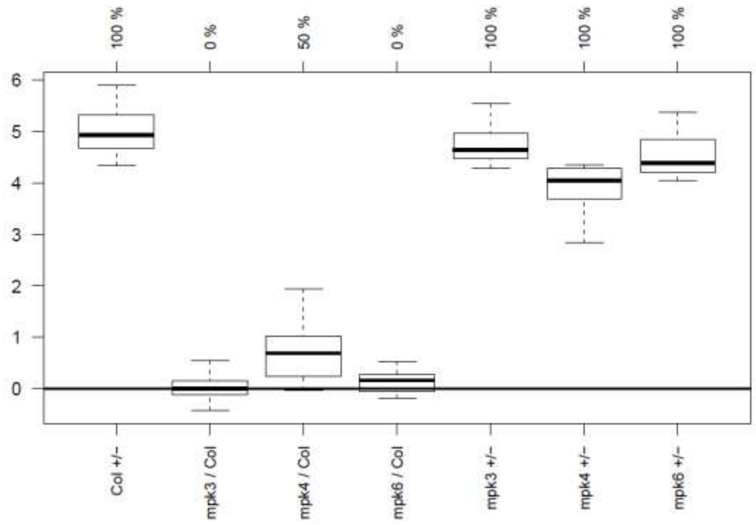
- number of cluster chosen by BIC Bayesian Information Criterion

- gene classification based on the conditional probabilities

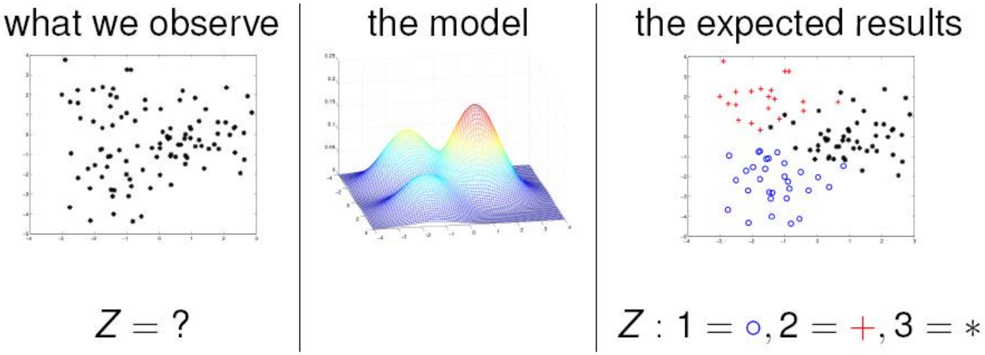
Matrix by stress
{ genes x log-ratios}

Gaussian mixture

clusters of
co-expressed genes
(groups of genes with same profile)



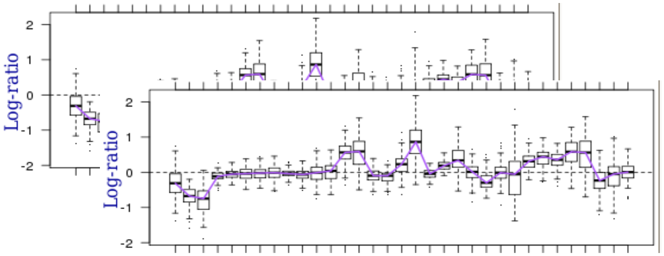
Co-expression analysis by mixture models



Matrix by stress
{ genes x log-ratios}

Gaussian mixture

~ 700 clusters of genes

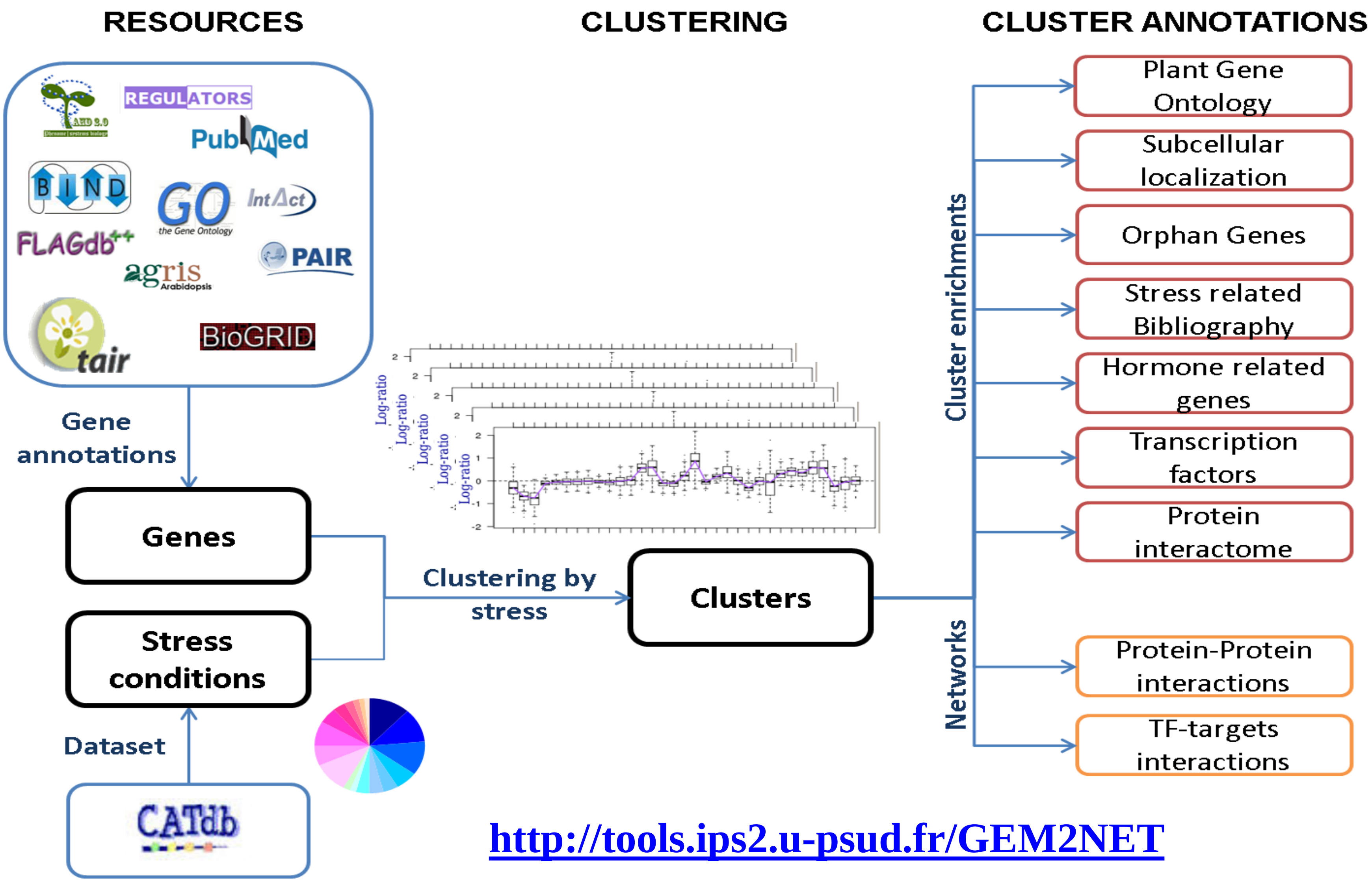


Stress category	Gene_nb	Cluster_nb
Nitrogen	13 495	59
Temperature	11 365	34
Drought	8 143	34
Salt	5 729	30
Heavy metal	10 617	57
UV	7 894	37
Gamma	5 350	32
Oxydative stress	10 127	52
Nectrophic bacteria	11 220	50
Biotrophic bacteria	12 023	56
Fungi	9 773	51
Rhodococcus	1 900	13
Oomycete	5 508	31
Nematode	7 413	27
Stifenia	1 525	17
Virus	11 832	54

18 stress categories

GEM2Net Flowchart

This project has been implemented as a new CATdb module: GEM2Net associated with a user-friendly Interface (Zaag et al., 2015)



RESOURCES

CLUSTERING

CLUSTER ANNOTATIONS

Gene annotations

Genes

Stress conditions

Dataset

Clustering by stress

Clusters

Cluster enrichments

- Plant Gene Ontology
- Subcellular localization
- Orphan Genes
- Stress related Bibliography
- Hormone related genes
- Transcription factors
- Protein interactome

Networks

- Protein-Protein interactions
- TF-targets interactions

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

Cluster Annotations

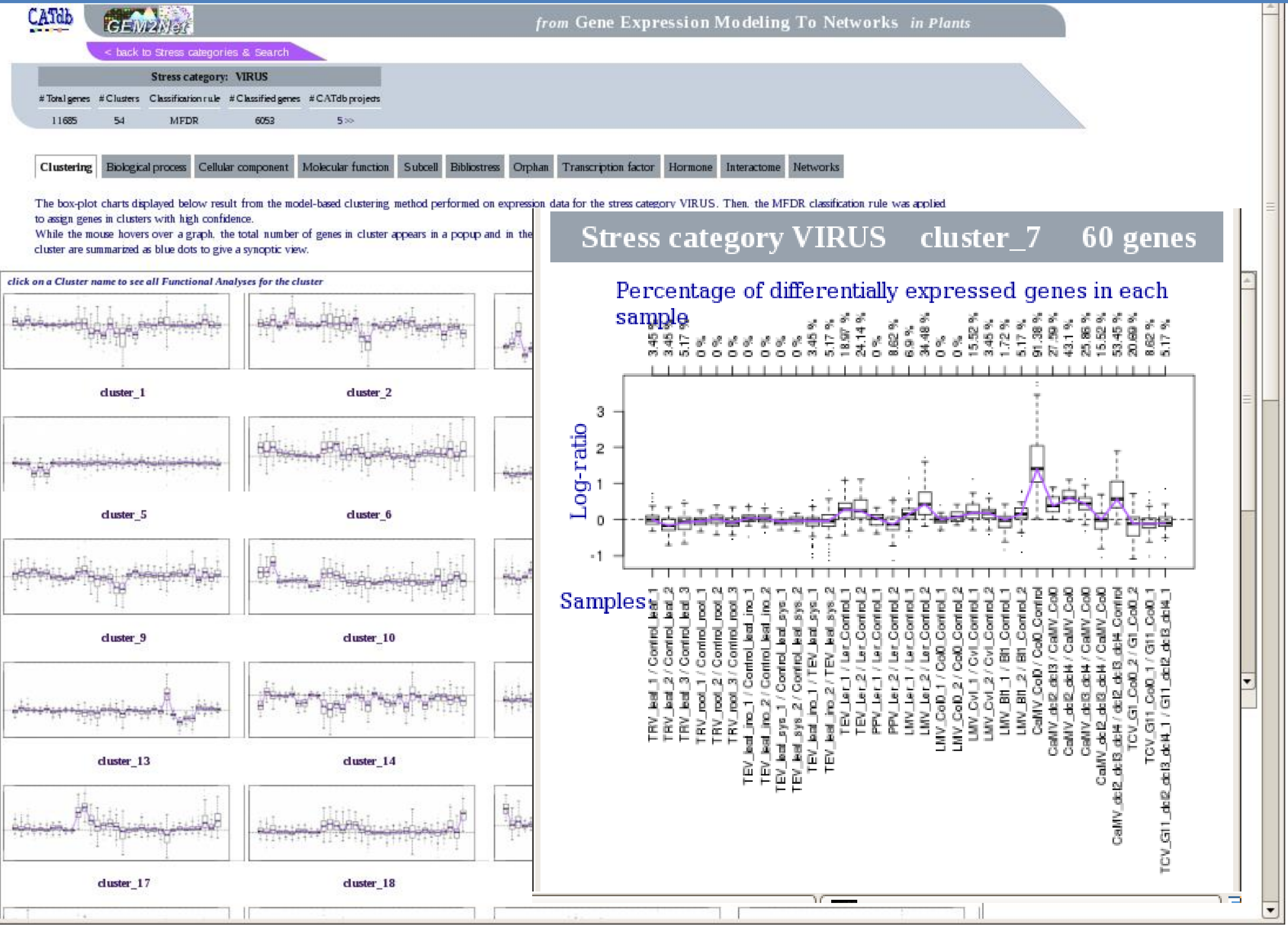
681 coexpression clusters covering 17264 genes

- 98% of clusters have a functional bias in a term of GO
- 80% of clusters are associated to a Stress term

GEM2Net gene set

	Total	Orphan	BP stress	Bibliostress	TF	Hormone
Arabidopsis Reference	34042	5105 (15%)	5106 (15%)	2580 (7.5%)	2260 (6.5%)	695 (2%)
GEM2Net dataset	17264	2165 (13%)	**4003 (23%)**	**2064 (12%)**	**1578 (9%)**	487 (3%)

GEM2Net Snapshot:



Stress category: VIRUS

# Total genes	# Clusters	Classification rule	# Classified genes	# CATdb projects
11685	54	MFDR	6053	5 >>

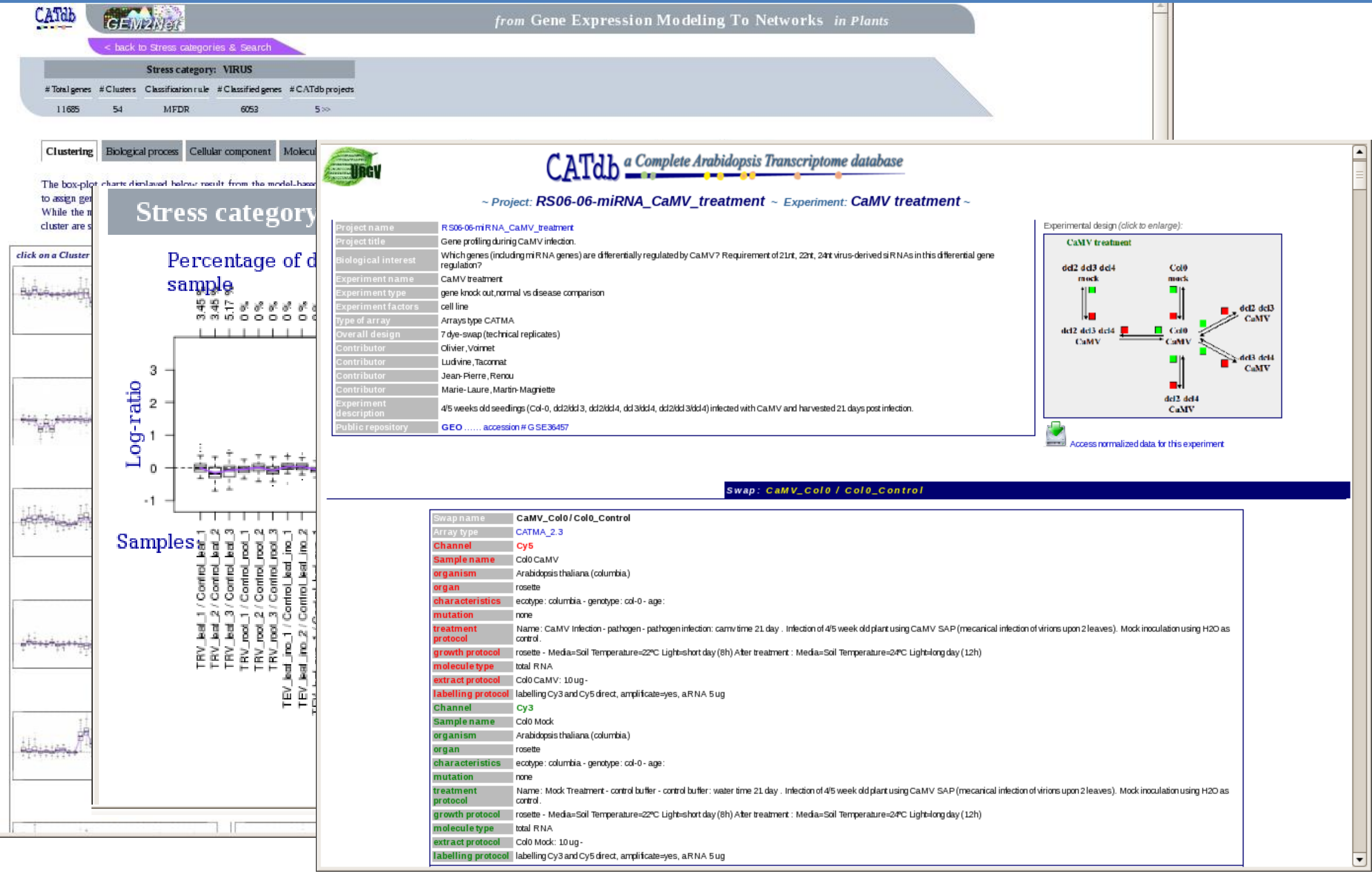
The box-plot charts displayed below result from the model-based clustering method performed on expression data for the stress category VIRUS. Then, the MFDR classification rule was applied to assign genes in clusters with high confidence.
While the mouse hovers over a graph, the total number of genes in cluster appears in a popup and in the
cluster are summarized as blue dots to give a synoptic view.

click on a Cluster name to see all Functional Analyses for the cluster

Stress category VIRUS cluster_7 60 genes

Percentage of differentially expressed genes in each sample

GEM2Net Snapshot:



Stress category: VIRUS

# Total genes	# Clusters	Classification rule	# Classified genes	# CATdb projects
11685	54	MFDR	6053	5 >>

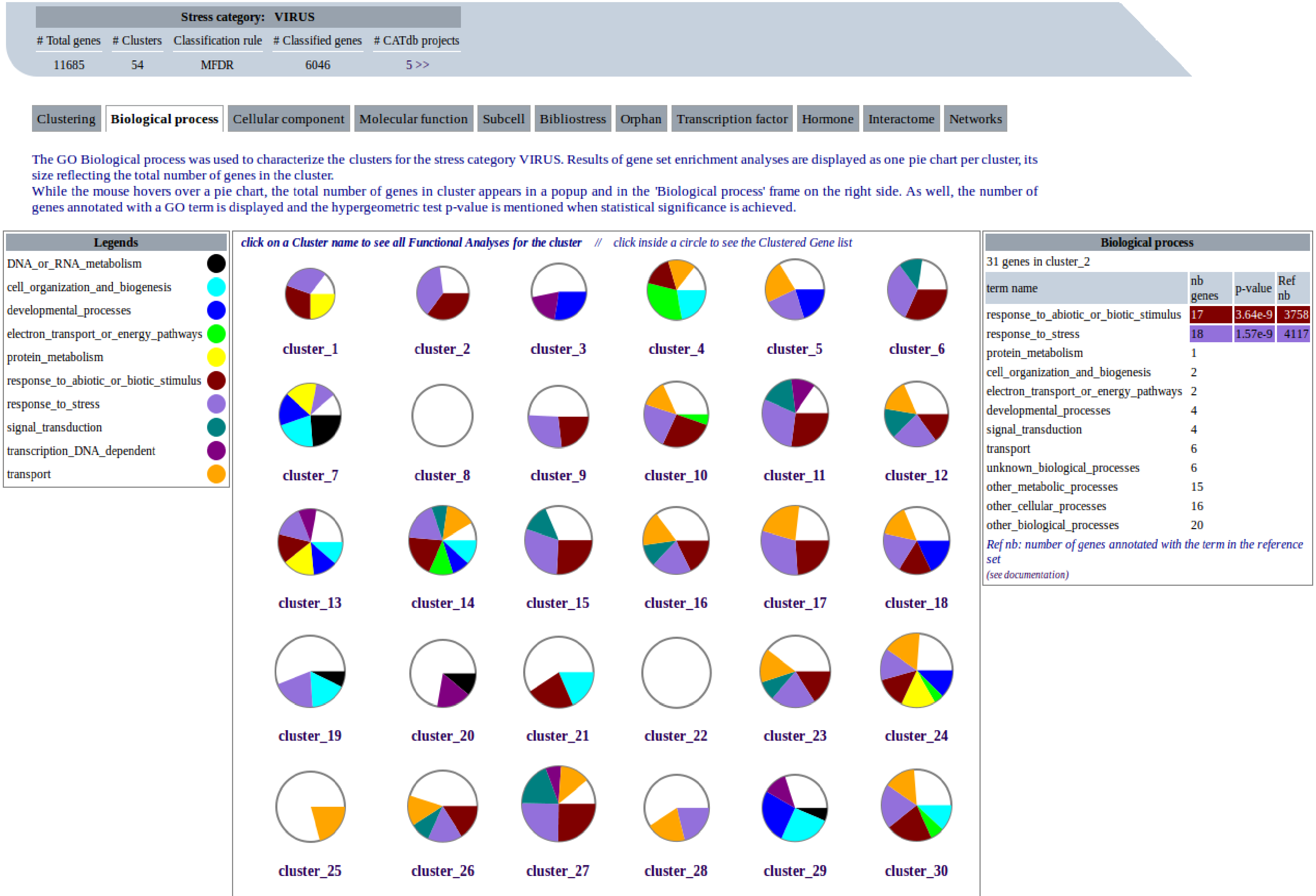
~ Project: **RS06-06-miRNA_CaMV_treatment** ~ Experiment: **CaMV treatment** ~

Project name	RS06-06-miRNA_CaMV_treatment
Project title	Gene profiling durinig CaMV infection.
Biological interest	Which genes (including miRNA genes) are differentially regulated by CaMV? Requirement of 21nt, 22nt, 24nt virus-derived siRNAs in this differential gene regulation?
Experiment name	CaMV treatment
Experiment type	gene knock out,normal vs disease comparison
Experiment factors	cell line
Type of array	Arrays type CATMA
Overall design	7 dye-swap (technical replicates)
Contributor	Olivier, Voinnet
Contributor	Ludivine, Taconnat
Contributor	Jean-Pierre, Renou
Contributor	Marie-Laure, Martin-Magniette
Experiment description	4/5 weeks old seedlings (Col-0, dcl2/dcl3, dcl2/dcl4, dcl3/dcl4, dcl2/dcl3/dcl4) infected with CaMV and harvested 21 days post infection.
Public repository	GEO ...... accession # GSE36457

Swap: CaMV_Col0 / Col0_Control

Swap name	**CaMV_Col0 / Col0_Control**
Array type	CATMA_2.3
Channel	**Cy5**
Sample name	Col0 CaMV
organism	Arabidopsis thaliana (columbia)
organ	rosette
characteristics	ecotype: columbia - genotype: col-0 - age:
mutation	none
treatment protocol	Name: CaMV Infection - pathogen - pathogen infection: camv time 21 day . Infection of 4/5 week old plant using CaMV SAP (mecanical infection of virions upon 2 leaves). Mock inoculation using H2O as control.
growth protocol	rosette - Media=Soil Temperature=22°C Light=short day (8h) After treatment : Media=Soil Temperature=24°C Light=long day (12h)
molecule type	total RNA
extract protocol	Col0 CaMV: 10 ug -
labelling protocol	labelling Cy3 and Cy5 direct, amplificate=yes, aRNA 5 ug
Channel	**Cy3**
Sample name	Col0 Mock
organism	Arabidopsis thaliana (columbia)
organ	rosette
characteristics	ecotype: columbia - genotype: col-0 - age:
mutation	none
treatment protocol	Name: Mock Treatment - control buffer - control buffer: water time 21 day . Infection of 4/5 week old plant using CaMV SAP (mecanical infection of virions upon 2 leaves). Mock inoculation using H2O as control.
growth protocol	rosette - Media=Soil Temperature=22°C Light=short day (8h) After treatment : Media=Soil Temperature=24°C Light=long day (12h)
molecule type	total RNA
extract protocol	Col0 Mock: 10 ug -
labelling protocol	labelling Cy3 and Cy5 direct, amplificate=yes, aRNA 5 ug

GEM2Net Snapshot:



Stress category: VIRUS				
# Total genes	# Clusters	Classification rule	# Classified genes	# CATdb projects
11685	54	MFDR	6046	5 >>

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*

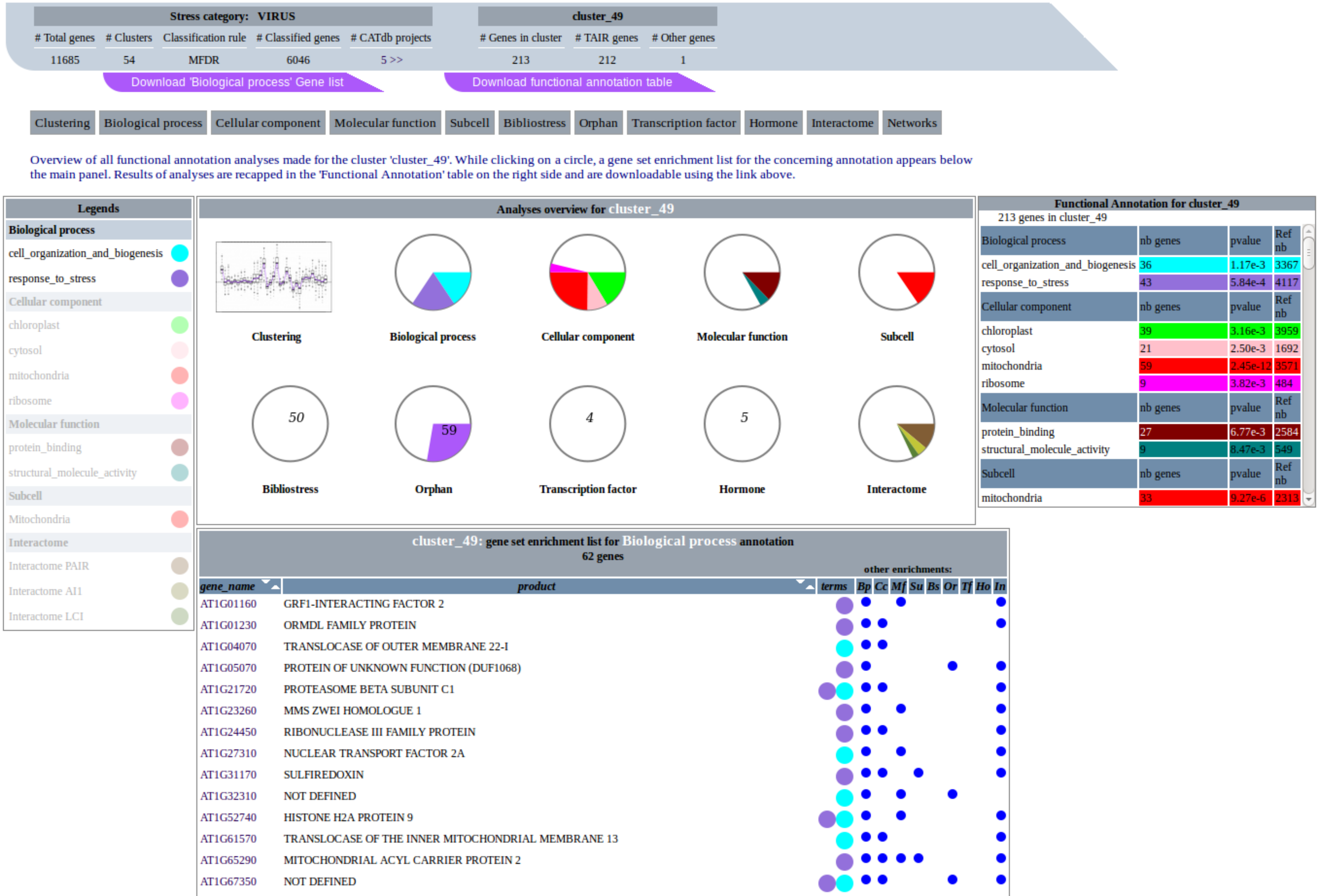
Biological process

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)

GEM2Net Snapshot:



Stress category: VIRUS				
# Total genes	# Clusters	Classification rule	# Classified genes	# CATdb projects
11685	54	MFDR	6046	5 >>

cluster_49		
# Genes in cluster	# TAIR genes	# Other genes
213	212	1

Overview of all functional annotation analyses made for the cluster 'cluster_49'. While clicking on a circle, a gene set enrichment list for the concerning annotation appears below the main panel. Results of analyses are recapped in the 'Functional Annotation' table on the right side and are downloadable using the link above.

Legends

- **Biological process**
  - cell_organization_and_biogenesis ●
  - response_to_stress ●
- Cellular component
  - chloroplast ●
  - cytosol ●
  - mitochondria ●
  - ribosome ●
- Molecular function
  - protein_binding ●
  - structural_molecule_activity ●
- Subcell
  - Mitochondria ●
- Interactome
  - Interactome PAIR ●
  - Interactome AI1 ●
  - Interactome LCI ●

Analyses overview for cluster_49

Functional Annotation for cluster_49

213 genes in cluster_49

Biological process	nb genes	pvalue	Ref nb
cell_organization_and_biogenesis	36	1.17e-3	3367
response_to_stress	43	5.84e-4	4117
Cellular component	**nb genes**	**pvalue**	**Ref nb**
chloroplast	39	3.16e-3	3959
cytosol	21	2.50e-3	1692
mitochondria	59	2.45e-12	3571
ribosome	9	3.82e-3	484
Molecular function	**nb genes**	**pvalue**	**Ref nb**
protein_binding	27	6.77e-3	2584
structural_molecule_activity	9	8.47e-3	549
Subcell	**nb genes**	**pvalue**	**Ref nb**
mitochondria	33	9.27e-6	2313

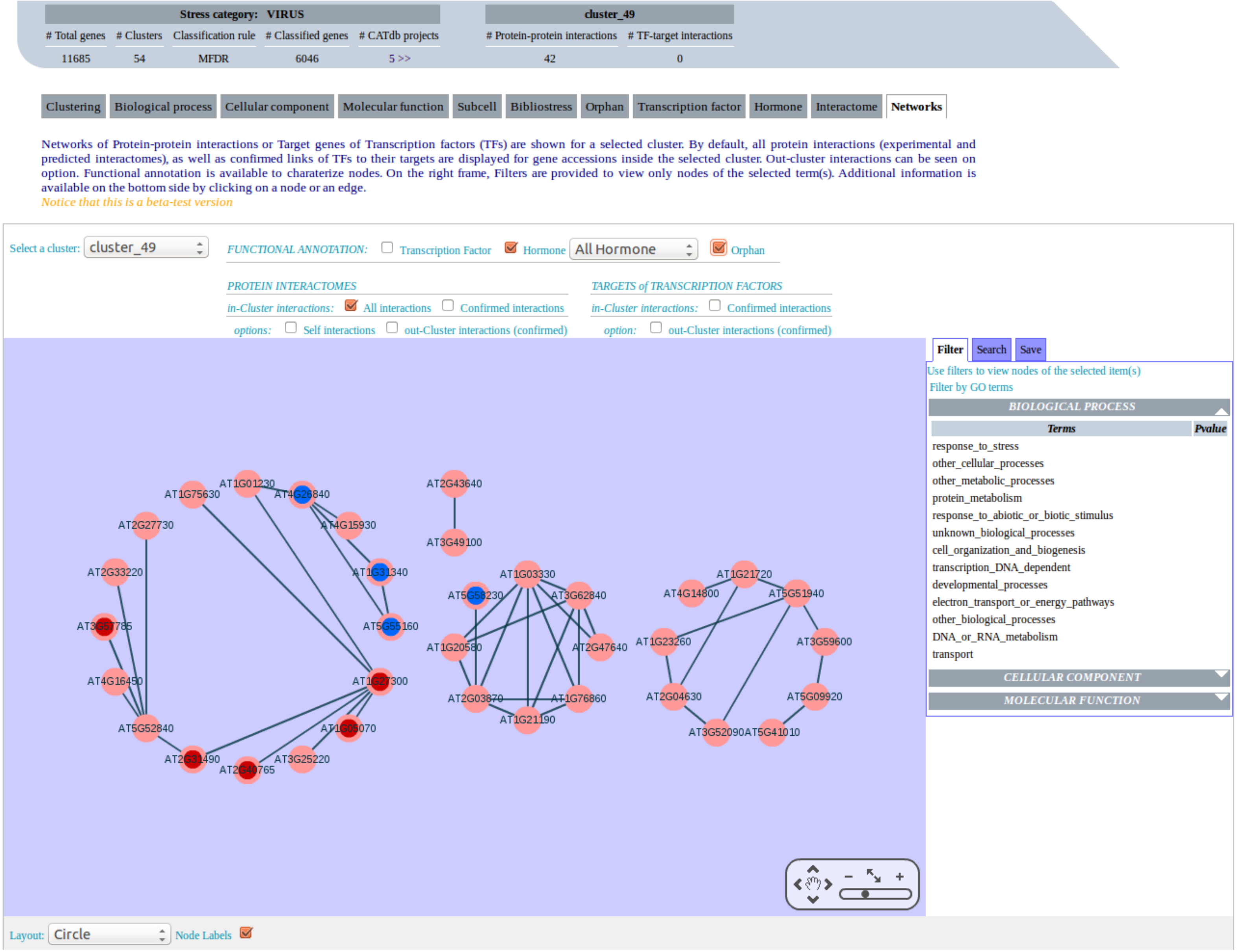
cluster_49: gene set enrichment list for Biological process annotation

62 genes

other enrichments:

gene_name	product	terms	Bp	Cc	Mf	Su	Bs	Or	Tf	Ho	In
AT1G01160	GRF1-INTERACTING FACTOR 2	●	●		●						●
AT1G01230	ORMDL FAMILY PROTEIN	●	●	●							●
AT1G04070	TRANSLOCASE OF OUTER MEMBRANE 22-I	●	●	●							
AT1G05070	PROTEIN OF UNKNOWN FUNCTION (DUF1068)	●	●					●			●
AT1G21720	PROTEASOME BETA SUBUNIT C1	●●	●	●							●
AT1G23260	MMS ZWEI HOMOLOGUE 1	●	●		●						●
AT1G24450	RIBONUCLEASE III FAMILY PROTEIN	●	●	●							●
AT1G27310	NUCLEAR TRANSPORT FACTOR 2A	●	●		●						●
AT1G31170	SULFIREDOXIN	●	●	●		●					●
AT1G32310	NOT DEFINED	●	●		●			●			
AT1G52740	HISTONE H2A PROTEIN 9	●●	●		●						●
AT1G61570	TRANSLOCASE OF THE INNER MITOCHONDRIAL MEMBRANE 13	●	●	●							●
AT1G65290	MITOCHONDRIAL ACYL CARRIER PROTEIN 2	●	●	●	●	●					●
AT1G67350	NOT DEFINED	●●	●	●				●			●

GEM2Net Snapshot:



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)

Use filters to view nodes of the selected item(s)

Filter by GO terms

BIOLOGICAL PROCESS

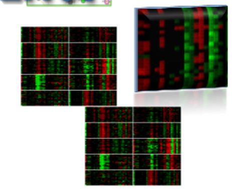
Terms	*Pvalue*
response_to_stress	
other_cellular_processes	
other_metabolic_processes	
protein_metabolism	
response_to_abiotic_or_biotic_stimulus	
unknown_biological_processes	
cell_organization_and_biogenesis	
transcription_DNA_dependent	
developmental_processes	
electron_transport_or_energy_pathways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

CELLULAR COMPONENT

MOLECULAR FUNCTION

Layout: Circle Node Labels ☑

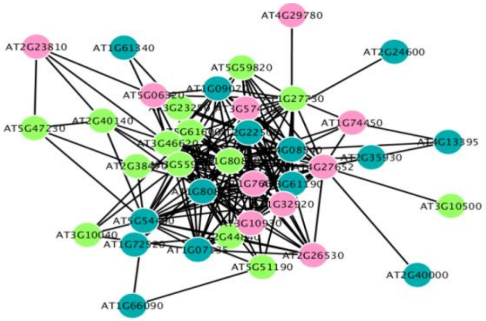
From co-expression to co-regulation



Clusters of co-expression
By stress

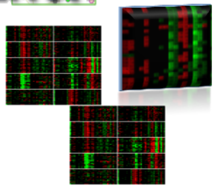
Horizontal integration

across the 18 stresses



Co-regulation networks

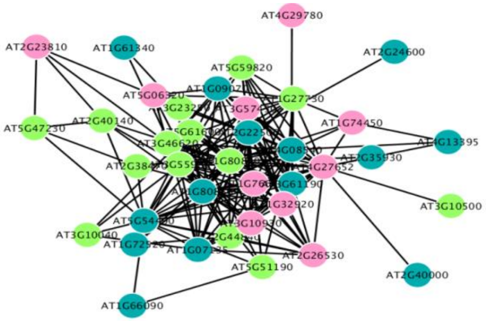
From co-expression to co-regulation



Clusters of co-expression
By stress

Horizontal integration

across the 18 stresses



Co-regulation networks

- Method : Identify pairs of genes co-expressed in different stress catgories compared to random networks (resampling test)

#stress	**2+**	**3+**	**4+**	**5+**	**6+**	**7+**	**8+**	**9+**	**10+**	**11+**	**12+**	**13+**	**14+**
#paires	295105	57833	19087	7997	3782	1908	903	395	190	83	36	9	2
# Paires aléatoires	68338	733	5	~0	0	0	0	0	0	0	0	0	0
Taux d'erreur	23,15%	1,2%	0,026 %	0	0	0	0	0	0	0	0	0	0

- a pair observed more than 3 times is statistically significant

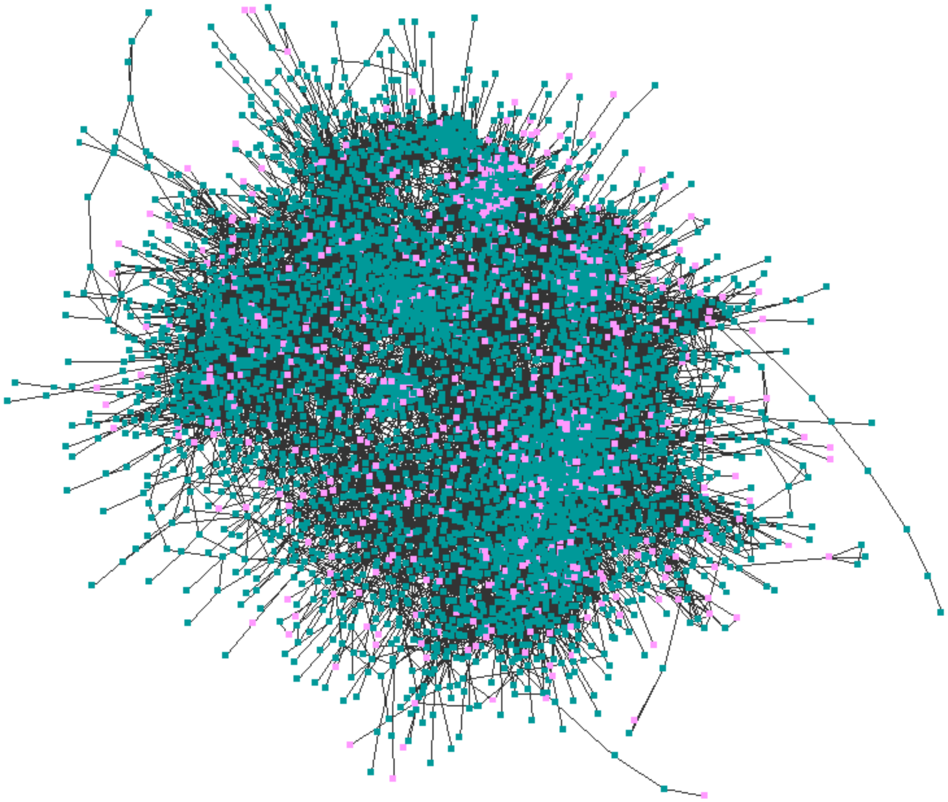
Co-Regulation (R. Zaag PhD)

At Network « 3+ »
stresses scale

Identification of Coregulated genes

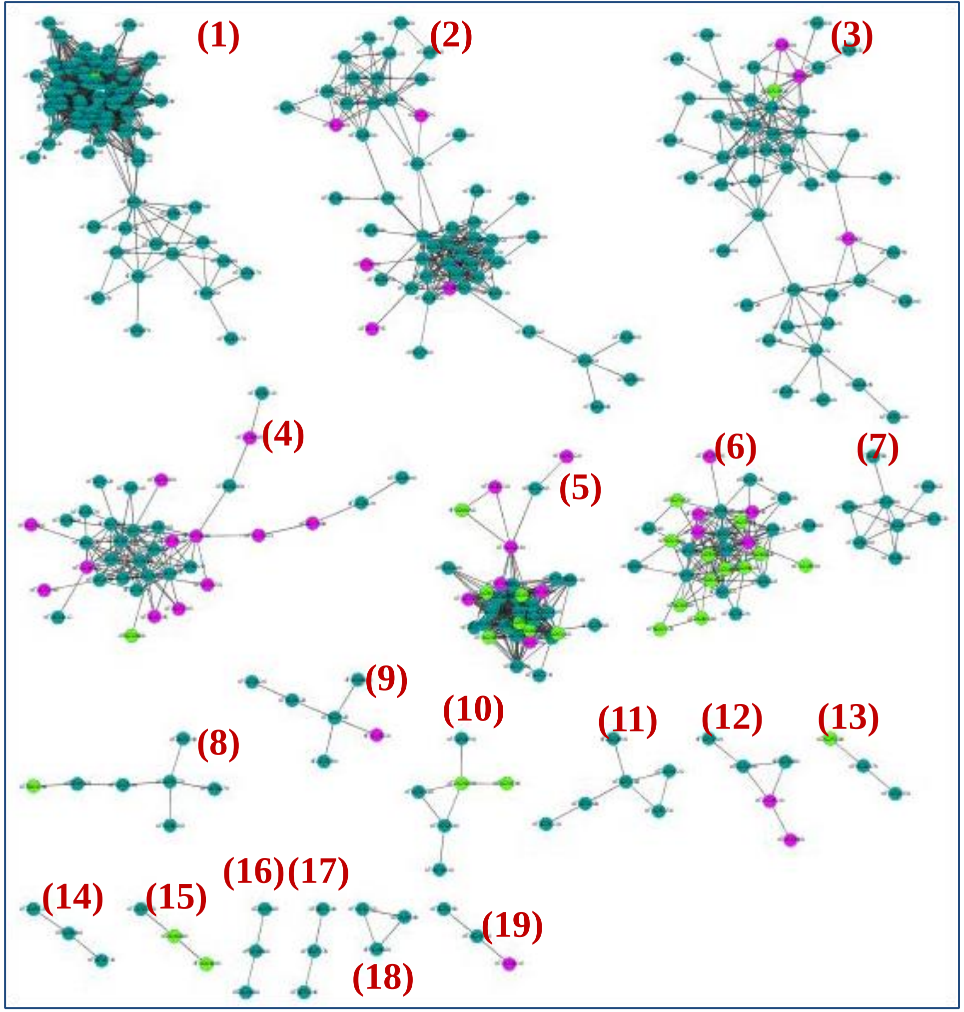
Describe groups of functional partners

Annotation of orphan genes



5626 nodes 57 833 edges/pairs
→ 713 orphans

Co-regulation Network (7 stresses)



Legend

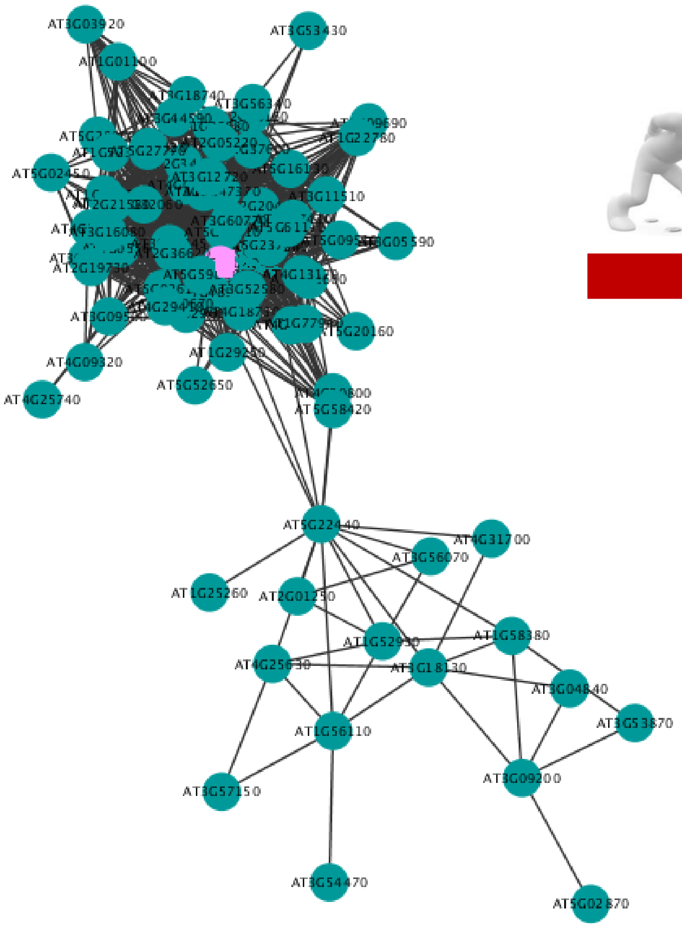
● Coregulated genes
● Orphan genes

The network with gene pairs conserved in at least 7 stresses is the first network showing connected components

415 genes with 41 orphan genes, 1908 interactions

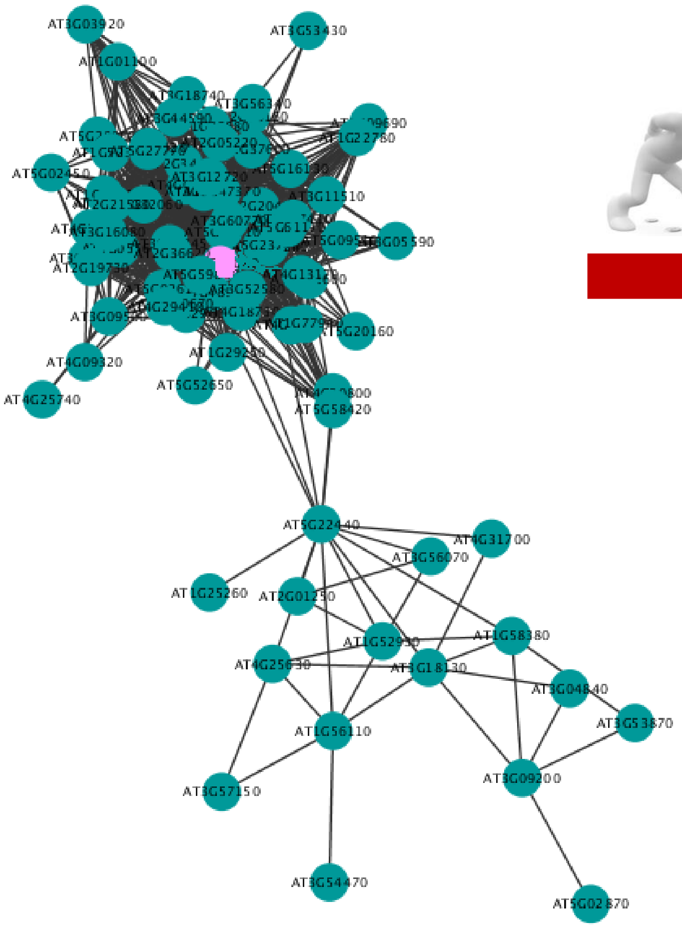
Example of an orphan identification within a module

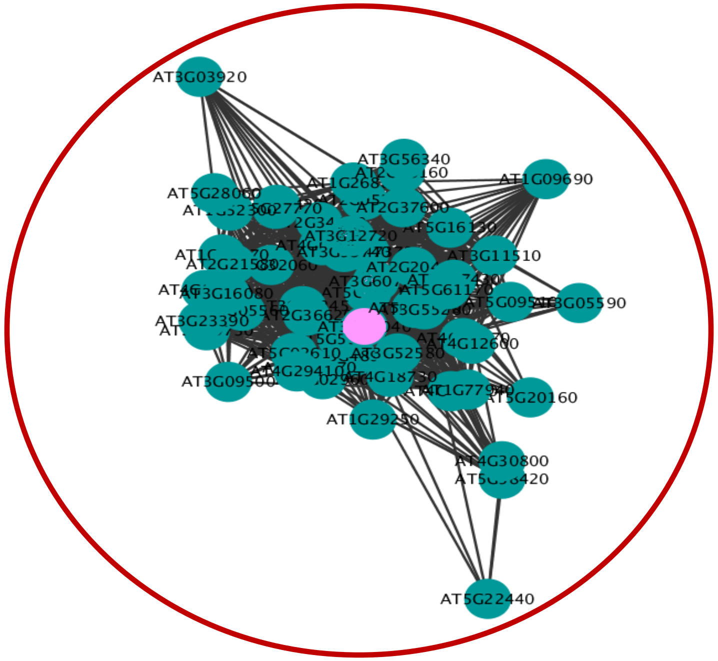
81 genes, 945 pairs, 1 orphan



Example of an orphan identification within a module

81 genes, 945 pairs, 1 orphan



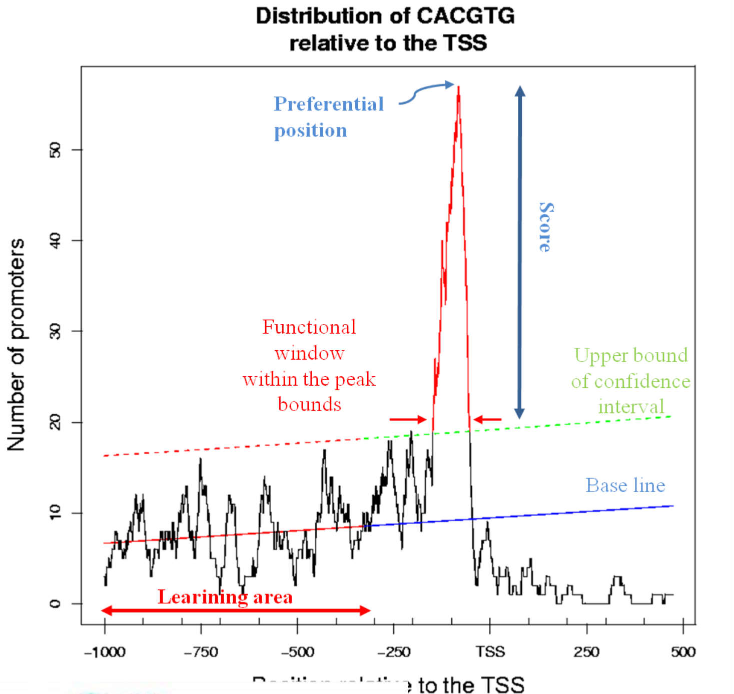


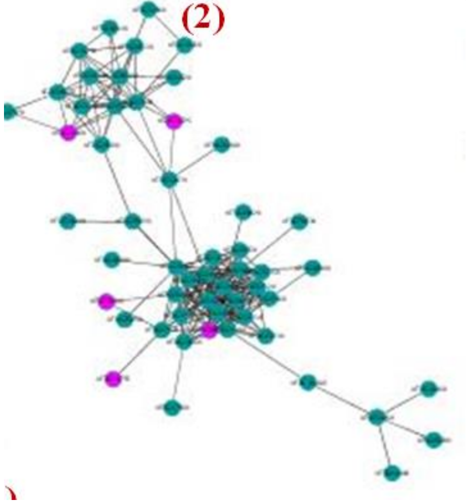
First Neighbors of the orphan gene

48 of the 55 neighbors are annotated as « Structural constituent of ribosome »

This orphan gene most likely codes for a ribosomal protein

Cis-regulatory motifs enrichment

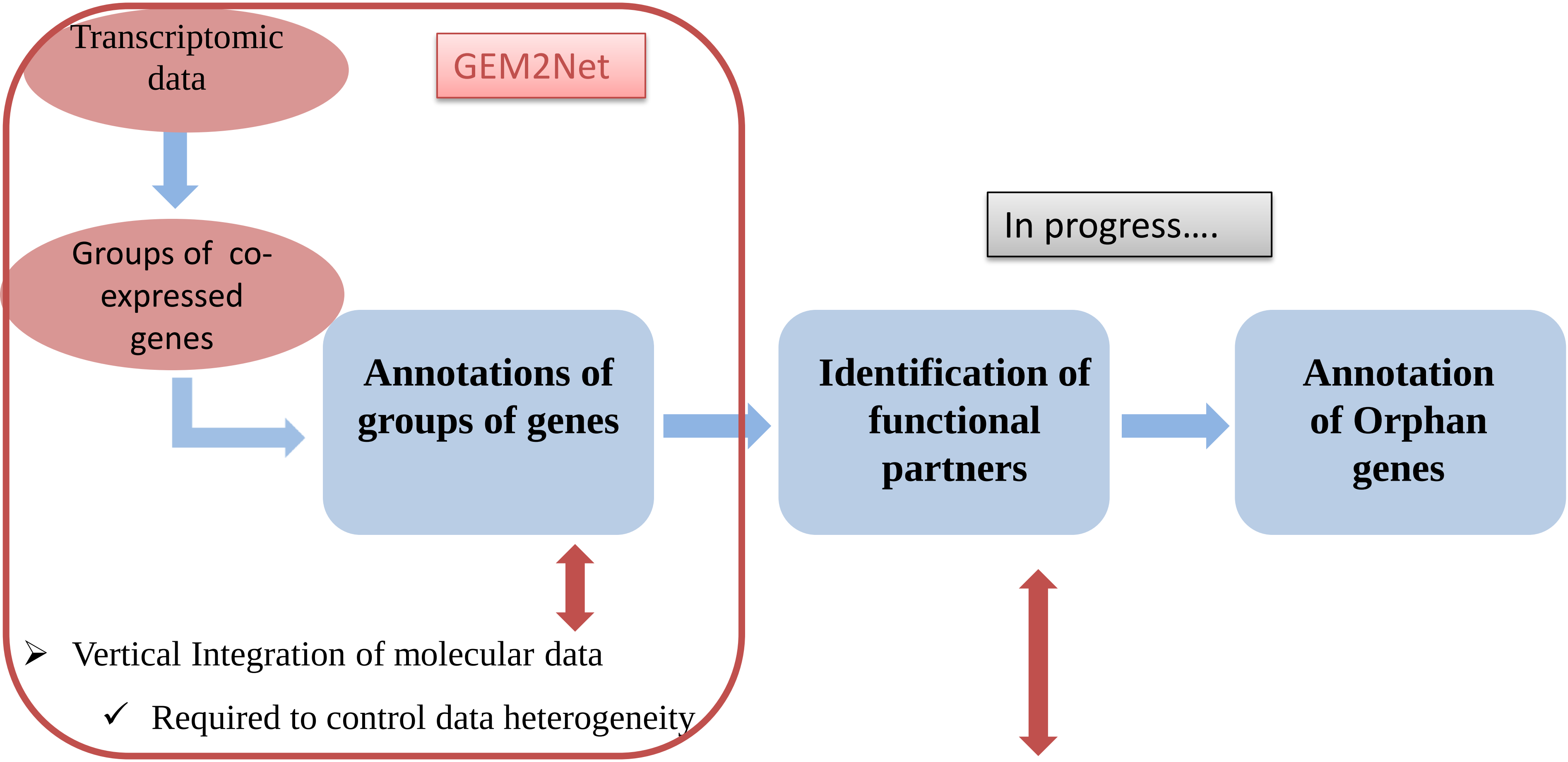




cis-regulatory motifs found by PLMdetect (Bernard et.,2010)

- ➢ 10 components are enriched in cis-regulatory motifs
- ➢ For 4 of them, the motif is present in over 80% of the gene promotors
- ➢ Component 2 has 5 motifs related to the light regulation, present at most in 50% of gene promoters

From Gene Expression Modeling to gene Network



Transcriptomic data

GEM2Net

Groups of co-expressed genes

Annotations of groups of genes

In progress….

Identification of functional partners

Annotation of Orphan genes

- ➢ Vertical Integration of molecular data
  - ✓ Required to control data heterogeneity

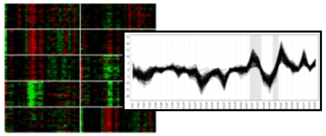
- ➢ Co-regulation = Horizontal integration of co-expressed data
  - ✓ Required to control stress heterogeneity

Actors in GEM2Net project

Bioinformatics

ML Martin-Magniette (team GNet)
R. Zaag (PhD)
V. Brunaud
J.-P. Tamby
C.Guichard
Z.Tariq
G. Rigaill
S. Aubourg

Statistics



G. Celeux
C. Maugis
T. Mary-Huard

Biology

E. Delannoy
J-P Renou
S. Balzergue

Funding :