

Tools for MultiOMICS data integration

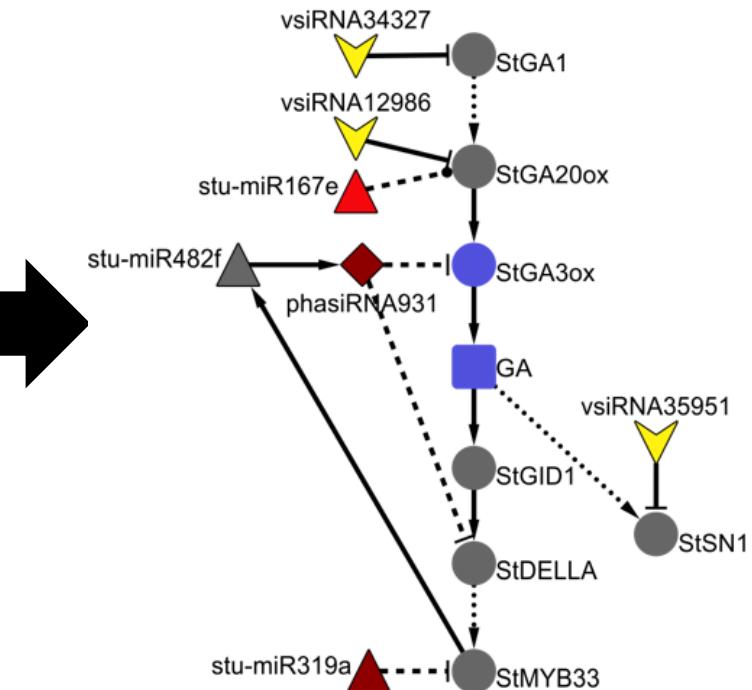
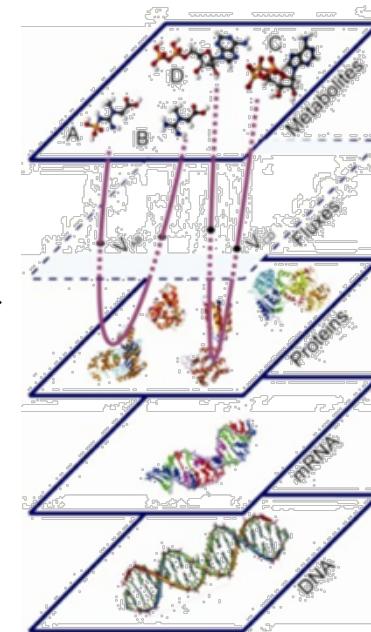
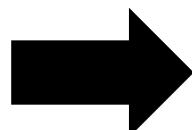
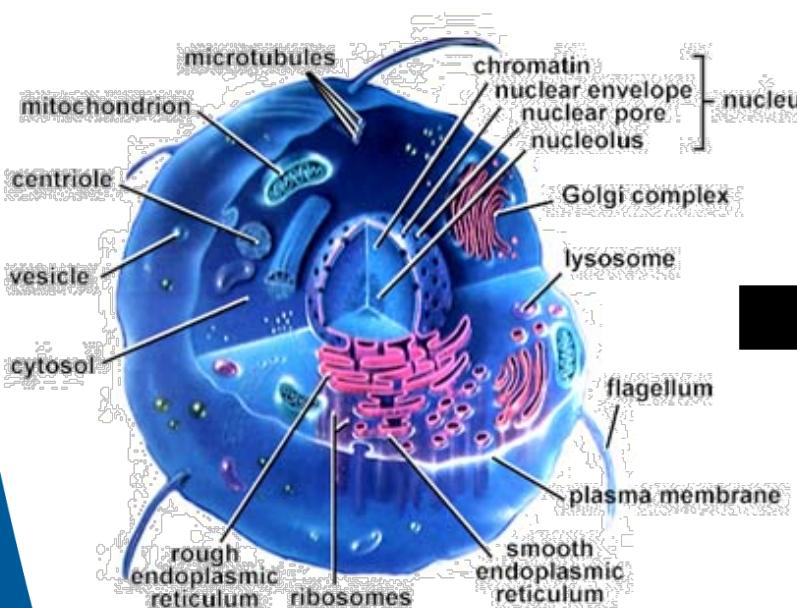
Kristina Gruden

Department of Biotechnology and Systems Biology,
National Institute of Biology, Ljubljana, Slovenia



NACIONALNI INŠTITUT ZA BIOLOGIJO
NATIONAL INSTITUTE OF BIOLOGY

Multilevel integrative analysis



Multilevel integrative analysis

How to approach it?

- Early data integration:
 - All datasets into one matrix -> analysis
 - Problems: normalisation, scale of measurements, intensity of responses
 - Can be done: smart mathematics, using vectors,... but delicate
- Late data integration

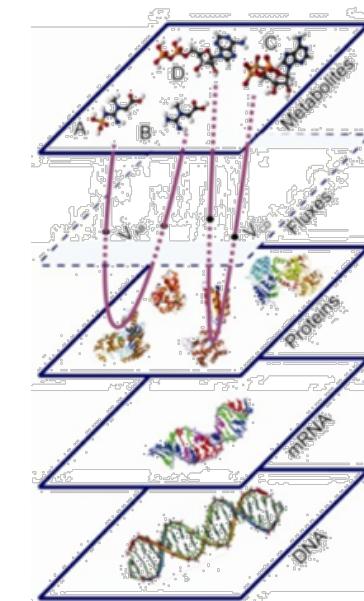


NACIONALNI INŠITITUT ZA BIOLOGIJO
NATIONAL INSTITUTE OF BIOLOGY

Multilevel integrative analysis

How to approach it?

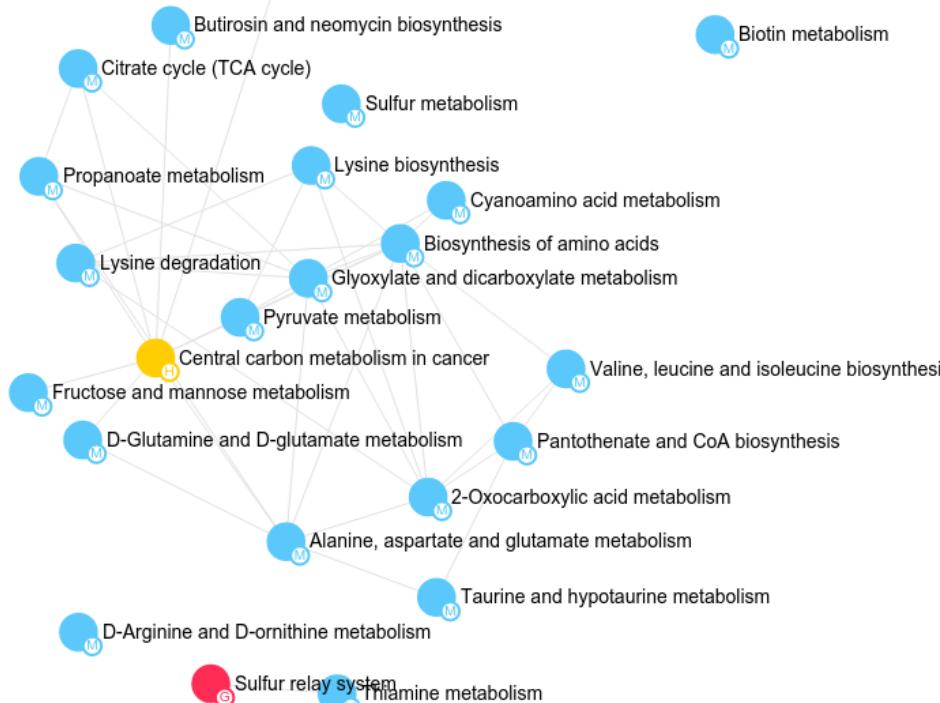
- Late data integration:
 - Normalisation and data analysis on separate datasets
-> integration
- Late data integration linked to biological knowledge



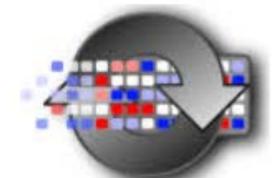
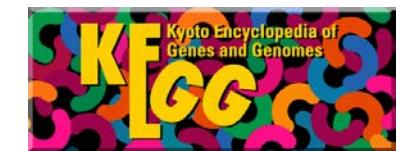
NACIONALNI INŠITUT ZA BIOLOGIJO
NATIONAL INSTITUTE OF BIOLOGY

Functional ontologies

From the level of **individual components** (genes/proteins/metabolites...) to the level of **biological process**



- Gene ontologies (GO)
- KEGG
- BioCyc
- MapMan



MapMan ontology

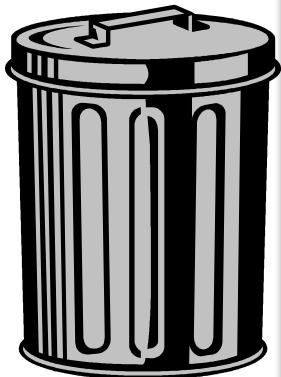
Components assigned to BINs

Combination of GO concepts for
Biological process & Molecular function

Specifically designed for plants

Manual curation

Developed within MapMan tools group



BIN code	BIN name
0	control genes
1	photosynthesis
2	major carbohydrates
3	minor carbohydrates
4	glycolysis
5	fermentation
6	gluconeogenesis / glyoxylate cycle
7	oxidative pentose phosphate pathway
8	TCA cycle / organic acid transformations
9	mitochondrial electron transport / ATP synthesis
10	cell wall
11	lipid metabolism
12	nitrogen metabolism
13	amino acid metabolism
14	sulphur assimilation
15	metal handling
16	secondary metabolism
17	hormone metabolism
18	cofactor and vitamin synthesis
19	tetrapyrrole synthesis
20	stress
21	redox
22	polyamine metabolism
23	nucleotide metabolism
24	biodegradation of xenobiotics
25	C1 metabolism
26	miscellaneous enzyme families
27	RNA
28	DNA
29	protein
30	signalling
31	cell
33	development
34	transport
35	not assigned

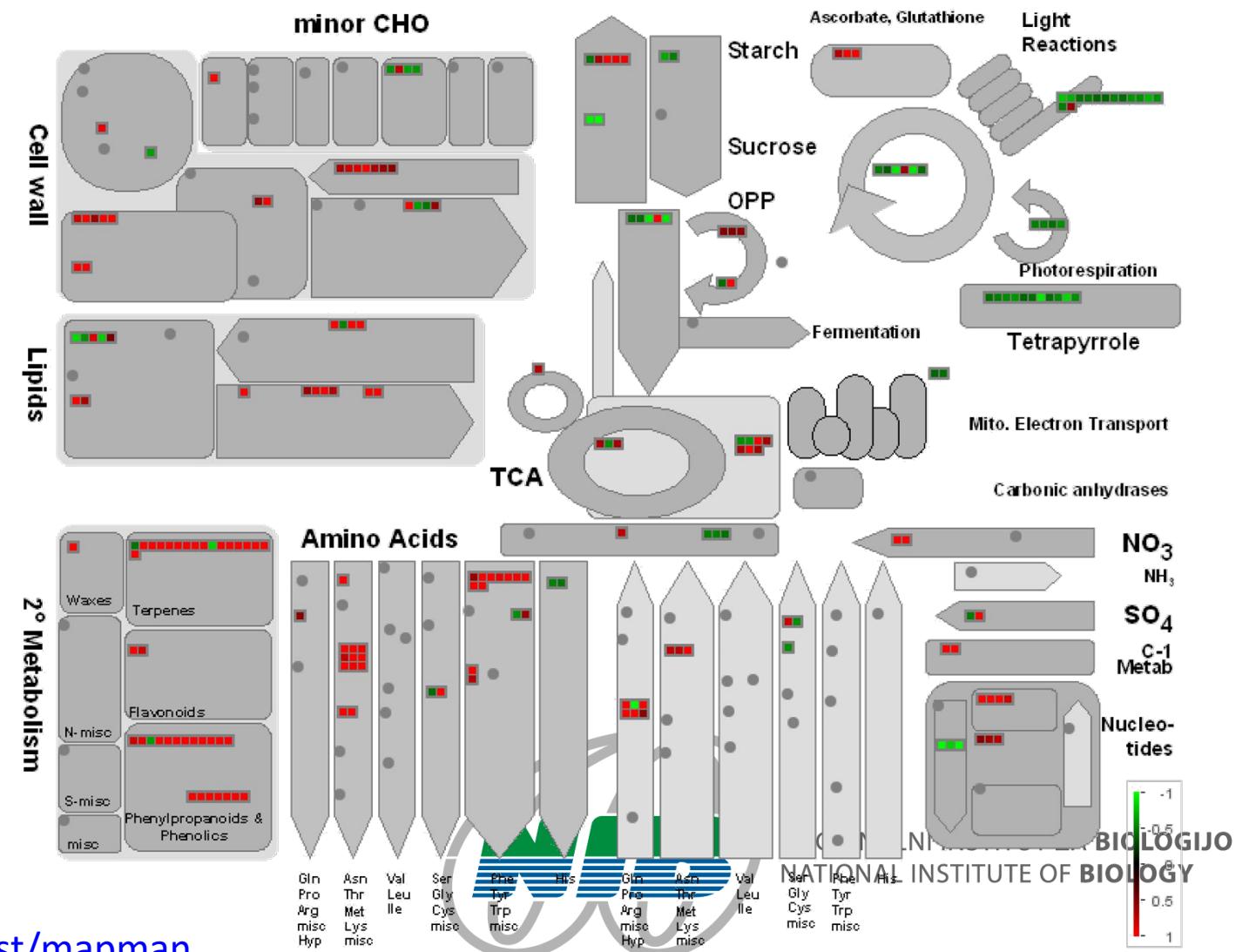
- ▶ 1 | Photosynthesis
- ▶ 2 | Cellular respiration
- ▶ 3 | Carbohydrate metabolism
- ▶ 4 | Amino acid metabolism
- ▶ 5 | Lipid metabolism
- ▶ 6 | Nucleotide metabolism
- ▶ 7 | Coenzyme metabolism
- ▶ 8 | Polyamine metabolism
- ▶ 9 | Secondary metabolism
- ▶ 10 | Redox homeostasis
- ▶ 11 | Phytohormone action
- ▶ 12 | Chromatin organisation
- ▶ 13 | Cell cycle organisation
- ▶ 14 | DNA damage response
- ▶ 15 | RNA biosynthesis
- ▶ 16 | RNA processing
- ▶ 17 | Protein biosynthesis
- ▶ 18 | Protein modification
- ▶ 19 | Protein homeostasis
- ▶ 20 | Cytoskeleton organisation
- ▶ 21 | Cell wall organisation
- ▶ 22 | Vesicle trafficking
- ▶ 23 | Protein translocation
- ▶ 24 | Solute transport
- ▶ 25 | Nutrient uptake
- ▶ 26 | External stimuli response
- ▶ 27 | Multi-process regulation
- ▶ 50 | Enzyme classification

V4

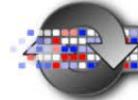
IOLOGIJO
OLOGY

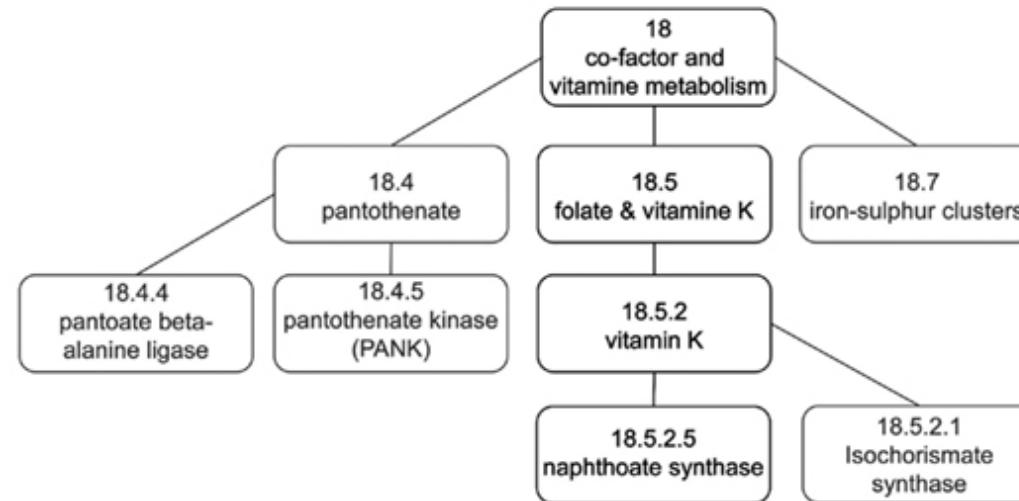
MapMan tool: Data visualization & interpretation

- Experimental data
- Ontology mappings
- Pathway diagrams



MapMan ontology vs GO

Hierarchical tree 
 (components can be assigned to multiple BINs)



/ Directed acyclic graph 

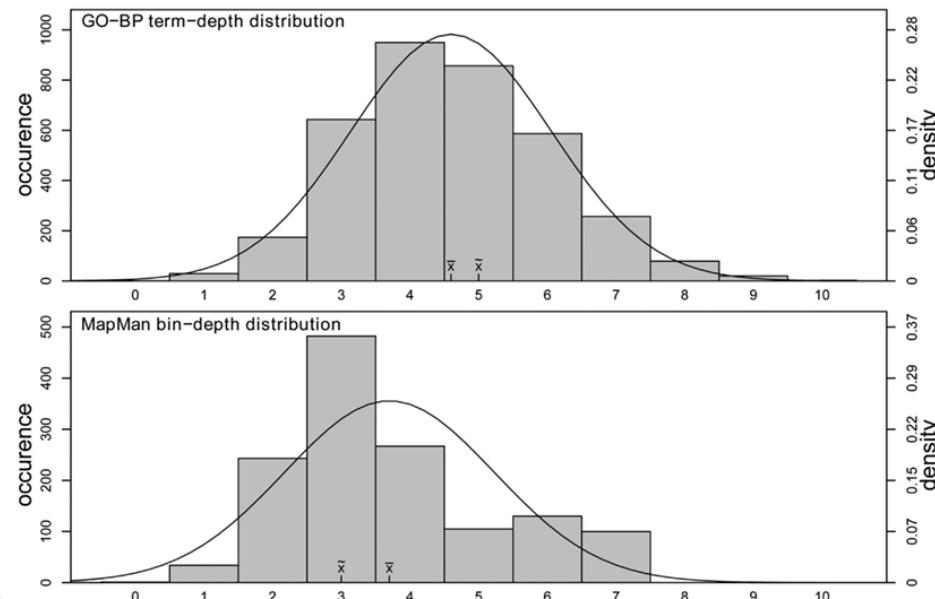


MapMan ontology vs GO

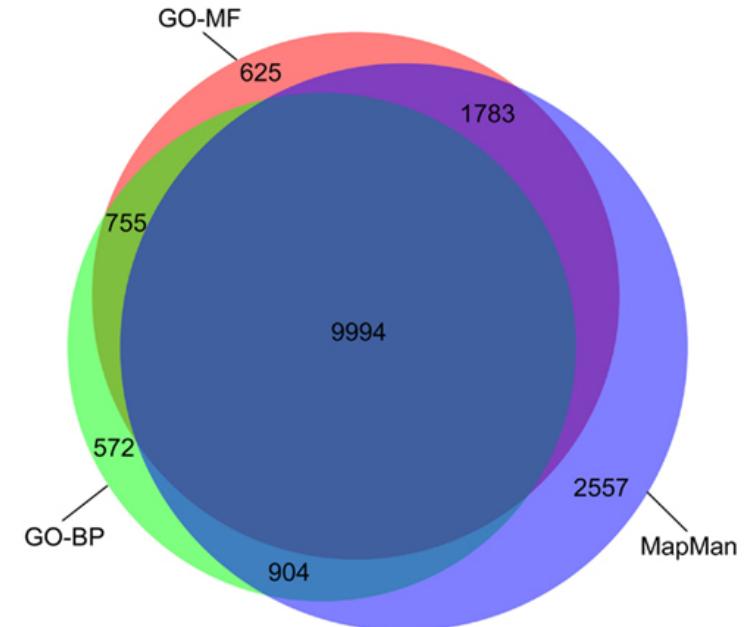
Comparative evaluation performed for *Arabidopsis thaliana*

- Overlap of gene annotations:
caa 50% of genes annotated

- Information content higher: concept depth lower



Overview of gene annotation by MapMan and GO



Klie and Nikoloski, 2012



NACIONALNI INŠITUT ZA BIOLOGIJO
NATIONAL INSTITUTE OF BIOLOGY

GoMapMan: database of MapMan gene annotations



Easy Browse, Search, Curate, Export functions implemented
Consolidation of gene annotations across species

1 PS

2 major CHO metabolism

3 minor CHO metabolism

4 glycolysis

5 fermentation

6 gluconeogenesis / glyoxylate cycle

7 OPP

8 TCA / organic transformation

8.1 TCA / organic transformation.TCA

8.1.1 TCA / organic transformation.TCA.pyruvate DH

8.1.1.1 TCA / organic transformation.TCA.pyruvate DH.E1

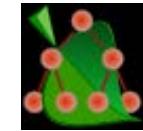
8.1.1.2 TCA / organic transformation.TCA.pyruvate DH.E2

8.1.1.3 TCA / organic transformation.TCA.pyruvate DH.E3

ORTHO03D003203	AT1G48030	Sotub05g025970.1.1	LOC_Os01g22520	Solyc05g053300.3.1
	AT3G17240	Sotub12g031040.1.1	LOC_Os05g06750	Solyc12g099100.2.1



GoMapMan: database of MapMan gene annotations



Easy Browse, Search, Curate, Export functions implemented

Link to other data sources: GO, PFAM, ENTREZ, EC, KEGG,...

-> incorporation of MapMan4 ontology in progress

www.gomapman.org

Description	acyl-CoA synthetase 5
Short Name	4CLL1
Synonym	4CLL1 ACOS5
Genomic Context	chr1:23310532-23312795 FORWARD
Source	Araport11 (Araport11 official release from Arabidopsis Information Portal (2016))
Last Modified	25 May 2018

Ontology annotations

[Click on bin to open it in ontology tree.](#)

Bin Code	Bin Name	Evidence Code
16.2	secondary metabolism.phenylpropanoids	IC

Orthologues

PGSCoMCL09271	iTAG_RSD02151	iTAGoMCL10146	wc05000	HOM03M000192	ORTHO03M042428
HOM03D000189	ORTHO03D007223	OCD_all_010100	OCD_PLAZATAG_010011	OCD_PLAZA_008684	

Transcripts or microarray features for the gene

(no existing connections)

Annotations

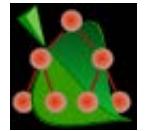
EC	6.2.1.- Acid-thiol ligases. 6.2.1.3 Long-chain-fatty-acid-CoA ligase.
ENTREZ	842596
GO	GO:0003824 catalytic activity GO:0004467 long-chain fatty acid-CoA ligase activity GO:0005737 cytoplasm GO:0008152 metabolic process GO:0010584 pollen exine formation GO:0016207 4-coumarate-CoA ligase activity GO:0031956 medium-chain fatty acid-CoA ligase activity GO:0046949 fatty-acyl-CoA biosynthetic process GO:0080110 sporopollenin biosynthetic process
KEGG	ath00130 , ath00360 , ath00940 , ath01100 , ath01110
PFAM	PF00501 AMP-binding enzyme
PO	PO:0001078 E expanded cotyledon stage PO:0007611 petal differentiation and expansion stage PO:0009010 seed PO:0009031 sepal PO:0009046 flower PO:0009071 anther wall tapetum PO:0025022 collective leaf structure

GoMapMan: database of MapMan gene annotations



Consolidation of gene annotations across species: using orthologue information: 13 species currently included

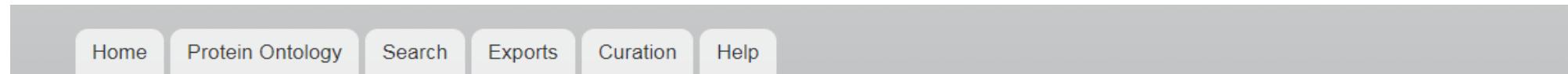




GoMapMan: database of MapMan gene annotations

Easy Browse, Search, Curate, **Export functions** implemented

Enabling use of different tools: MapMan, pathway enrichment tools (GSEA), BioMine, any others (generic export)



Exports

Exports of gene identifiers and microarray features for individual species are available for import to [BioMine](#), [GSEA](#), [MapMan](#) and [SEGS](#). Date of the [last export](#) creation is **25 May 2018**.

Listing: [Export](#) » [Current](#) » [Mapman](#)

- [vvi_Genoscope-12x-Vo_2018-05-25_mapping.txt.gz](#)
- [vvi_Genoscope-12x-V1_2018-05-25_mapping.txt.gz](#)
- [vvi_Grapevine-V1-Gene_2018-05-25_mapping.txt.gz](#)
- [vvi_Grapevine-V2-Gene_2018-05-25_mapping.txt.gz](#)
- [vvi_Nimblegen_gene_2018-05-25_mapping.txt.gz](#)
- [vvi_Nimblegen_probe_2018-05-25_mapping.txt.gz](#)
- [vvi_Nimblegen_probe_old_2018-05-25_mapping.txt.gz](#)
- [vvi_genoscope-v1_2018-05-25_mapping.txt.gz](#)

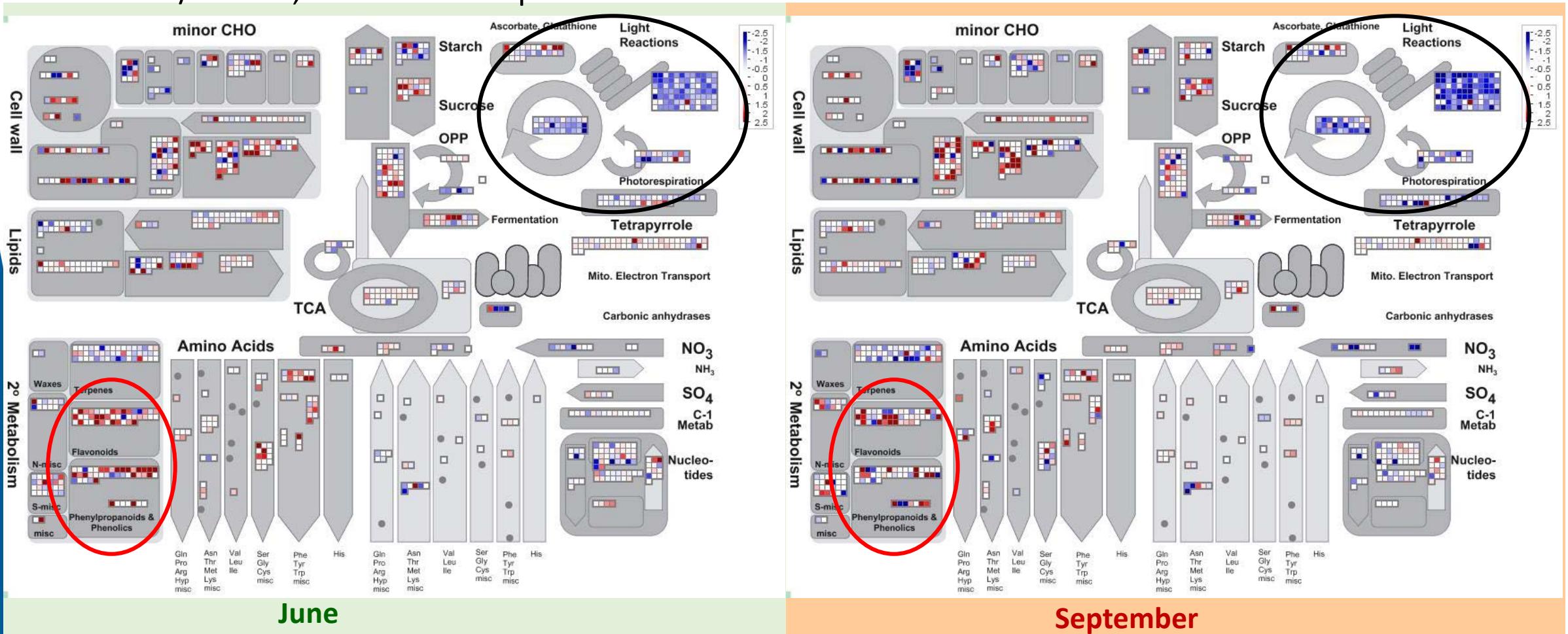


NACIONALNI INŠITUT ZA BIOLOGIJO
NATIONAL INSTITUTE OF BIOLOGY

www.gomapman.org

MapMan visualisation:

Use case: Mechanisms of interaction between grapevine and phytoplasma infected/control, FDR corrected p<0.05



Changes on the level of biological process

Use case:

Mechanisms of interaction between grapevine and phytoplasma

- GSEA: secondary metabolites

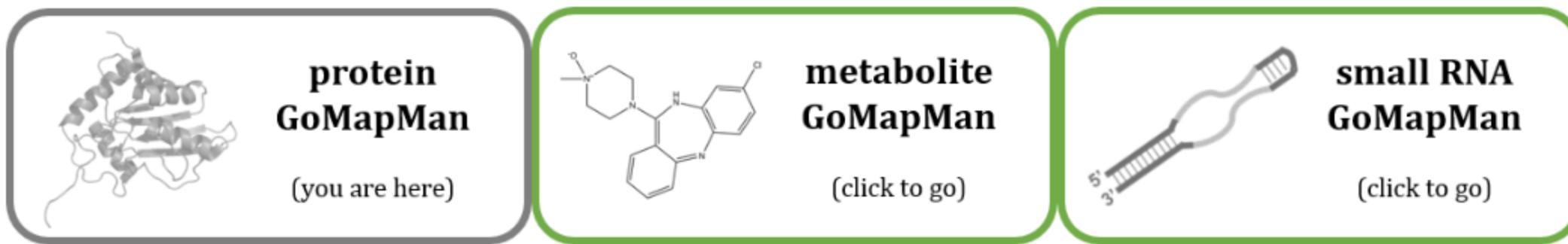
bin	name	elements no	june	september
16	secondary metabolism	176	1	
16.1	secondary metabolism. isoprenoids	48	1	
16.1.4	secondary metabolism.isoprenoids.carotenoids	13	1	
16.2	secondary metabolism. phenylpropanoids	37	1	
16.2.1	secondary metabolism.phenylpropanoids.lignin biosynthesis	31	1	
16.2.1.1	secondary metabolism.phenylpropanoids.lignin biosynthesis.PAL	11	1	
16.2.1.2	secondary metabolism.phenylpropanoids.lignin biosynthesis.C4H	2	1	
16.2.1.7	secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1	3	1	
16.8	secondary metabolism. flavonoids	44	1	
16.8.2.2	secondary metabolism.flavonoids.chalcones.chalcone isomerase	2	1	
16.8.3	secondary metabolism.flavonoids.dihydroflavonols	21	1	
16.8.3.2	secondary metabolism.flavonoids.dihydroflavonols.flavanone 3-hydroxylase	2	1	
16.8.5	secondary metabolism.flavonoids.isoflavonols	5	1	
16.8.5.1	secondary metabolism.flavonoids.isoflavonols.isoflavone reductase	5	1	

Multilevel integrative analysis

GoMapMan + visualisation in MapMan

Mappings for genes: RNA, **proteins**; **sRNAs**: miRNAs+tasiRNAs; **metabolites**

GoMapMan

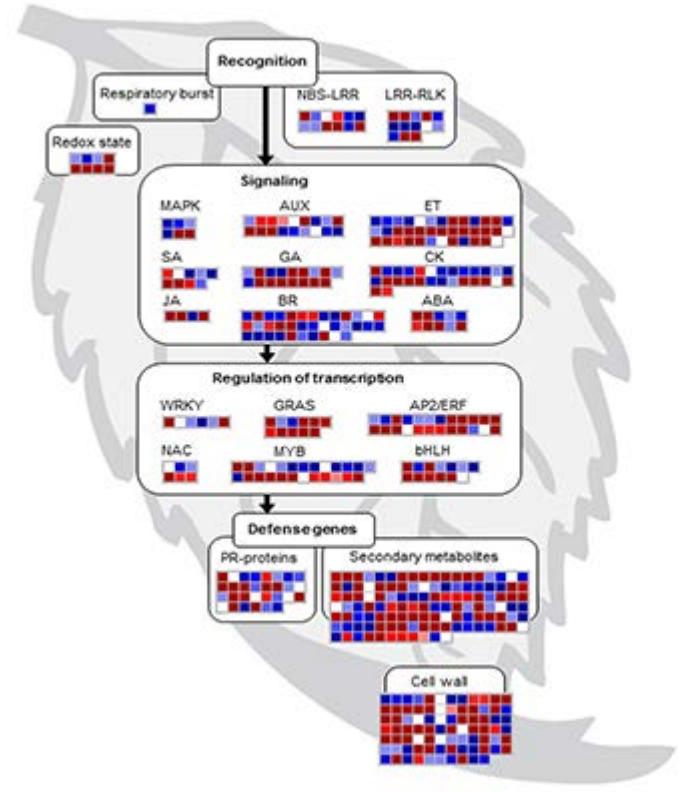
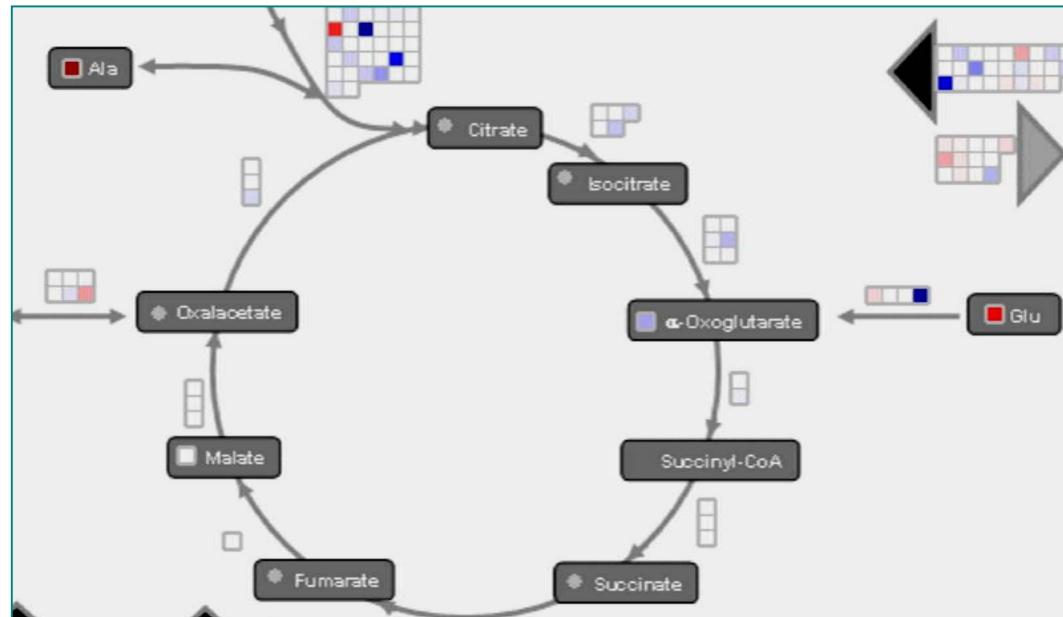


GoMapMan is an open web-accessible resource for gene functional annotations in the plant sciences. It was developed to facilitate improvement, consolidation and visualization of gene annotations across several plant species. GoMapMan is based on the plant specific [MapMan ontology](#), organized in the form of a hierarchical tree of biological concepts, which describe gene functions.

Multilevel integrative analysis

GoMapMan + visualisation in MapMan

Two levels in parallel at max: protein/transcript/miRNA+metabolites



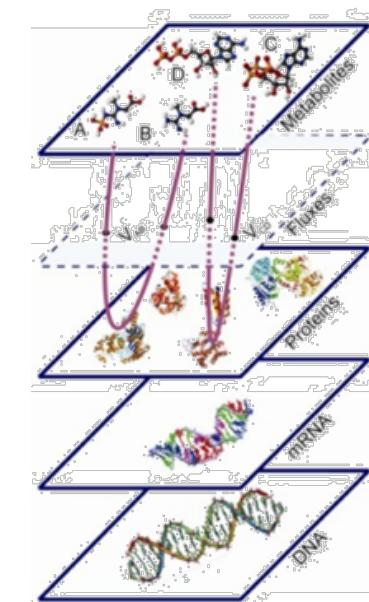
Cytoscape miRNA+transcripts (Omics Analysis Collection plugin)

Multilevel integrative analysis

How to approach it?

- Late data integration:
 - Normalisation and data analysis on separate datasets
-> integration

- Late data integration linked to biological knowledge
 - Link to ontologies (MapMan)
 - Construction of knowledge networks



NACIONALNI INŠITITUT ZA BIOLOGIJO
NATIONAL INSTITUTE OF BIOLOGY

Comprehensive knowledge network

Integration of dispersed knowledge sources (mainly *Arabidopsis*):

Metabolic pathways: KEGG

Protein Protein Interactions

STRING

High-throughput PPI datasets (*Arabidopsis Interactome Mapping C.*, 2011, PPI immunity, PPI membrane)

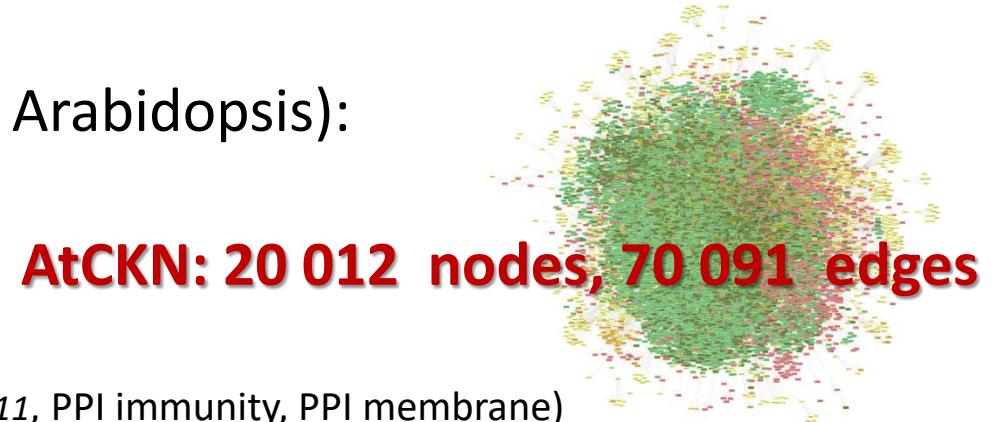
Transcriptional regulation network

agris DB (*AtTFDB = AtRegNet*)

Literature EIN3....

miRNA regulatory network

miRTarBase, plant microRNA DB

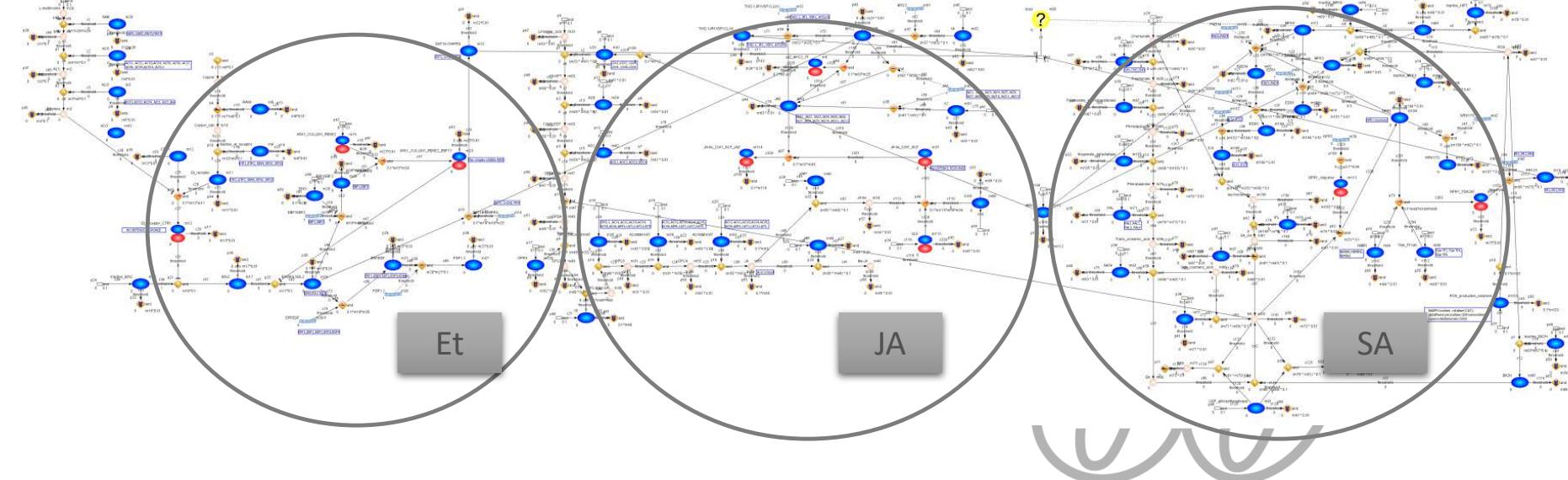


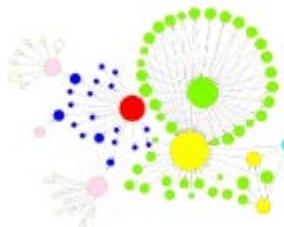
Detailed curated knowledge networks

Bottom-up -> plant immune signaling model

- Pathway databases
- Literature mining

→ AtPIS.v1: 175 components, 524 reactions





DiNAR Input

Background knowledge

- Network format: nodes and edges
- Node position coordinates



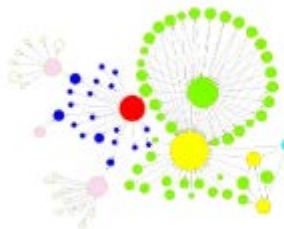
Experimental data

- Any type: transcriptomics, miRNAomics, epigenomics, metabolomics
- IDs should match the ID of the nodes in the network
- Best time series, can be also any other multiconditional setting

Tab delimited format

<https://nib-si.shinyapps.io/DiNAR/>

Zagorščak et al. Plant Methods 2018



DiNAR Features

- Static visualisation of networks
- Dynamic visualisation

animatorR package: homotopy to interpolate node and edge weight value between discrete conditions/time-points

visNetwork package: javascript package for dynamic visualisation

Exports:

Analysis settings: reproducibility, Simple and interactive images, Interactive videos

DiNAR can also be run locally in R or hosted on a local RStudio Shiny Server:

R and **Shiny** package for interactive web applications

<https://nib-si.shinyapps.io/DiNAR/>

Zagorščak et al. Plant Methods 2018



Differential Network Analysis in *R*

Dynamic Time/Condition

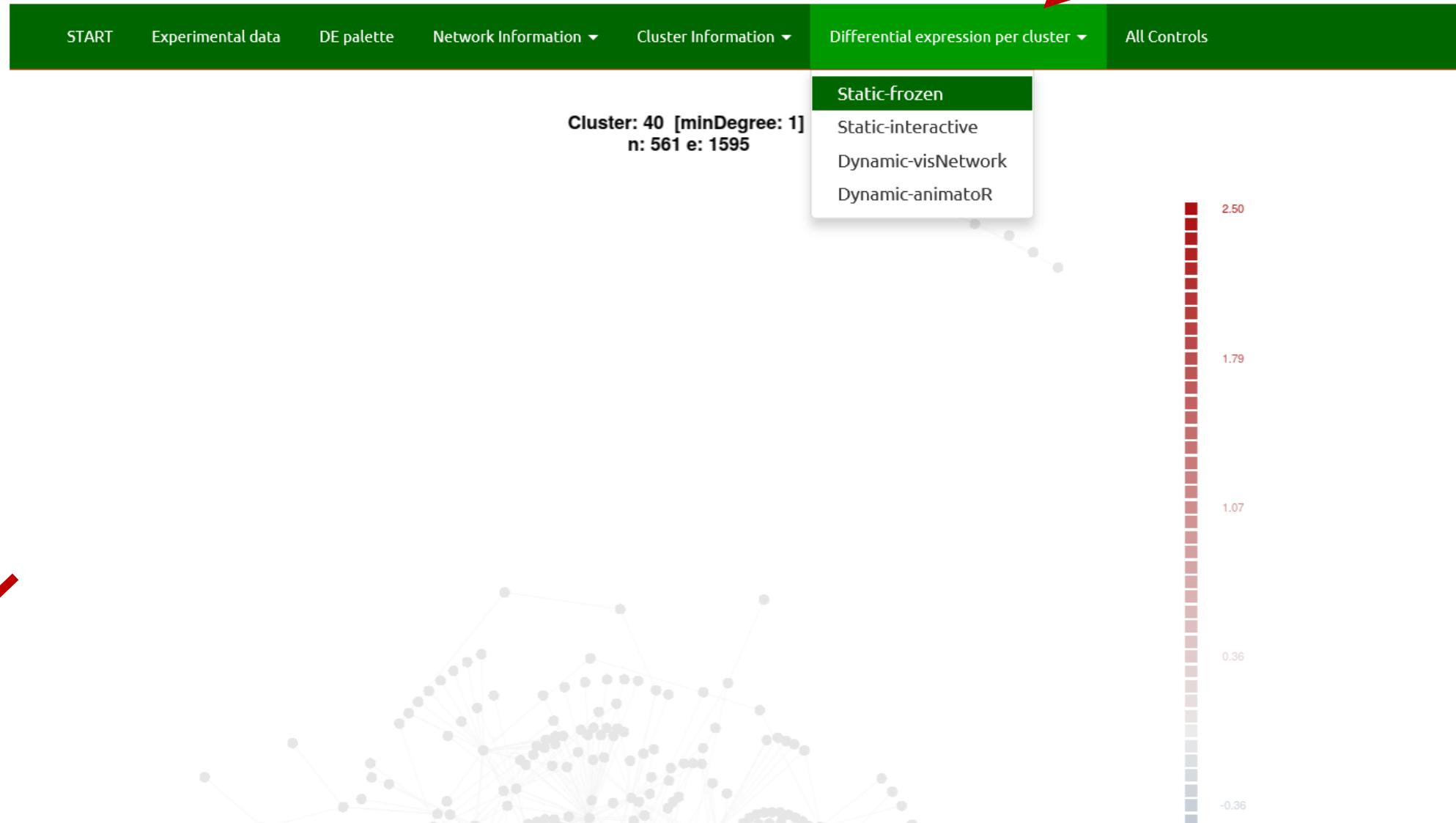
Time/Condition
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5

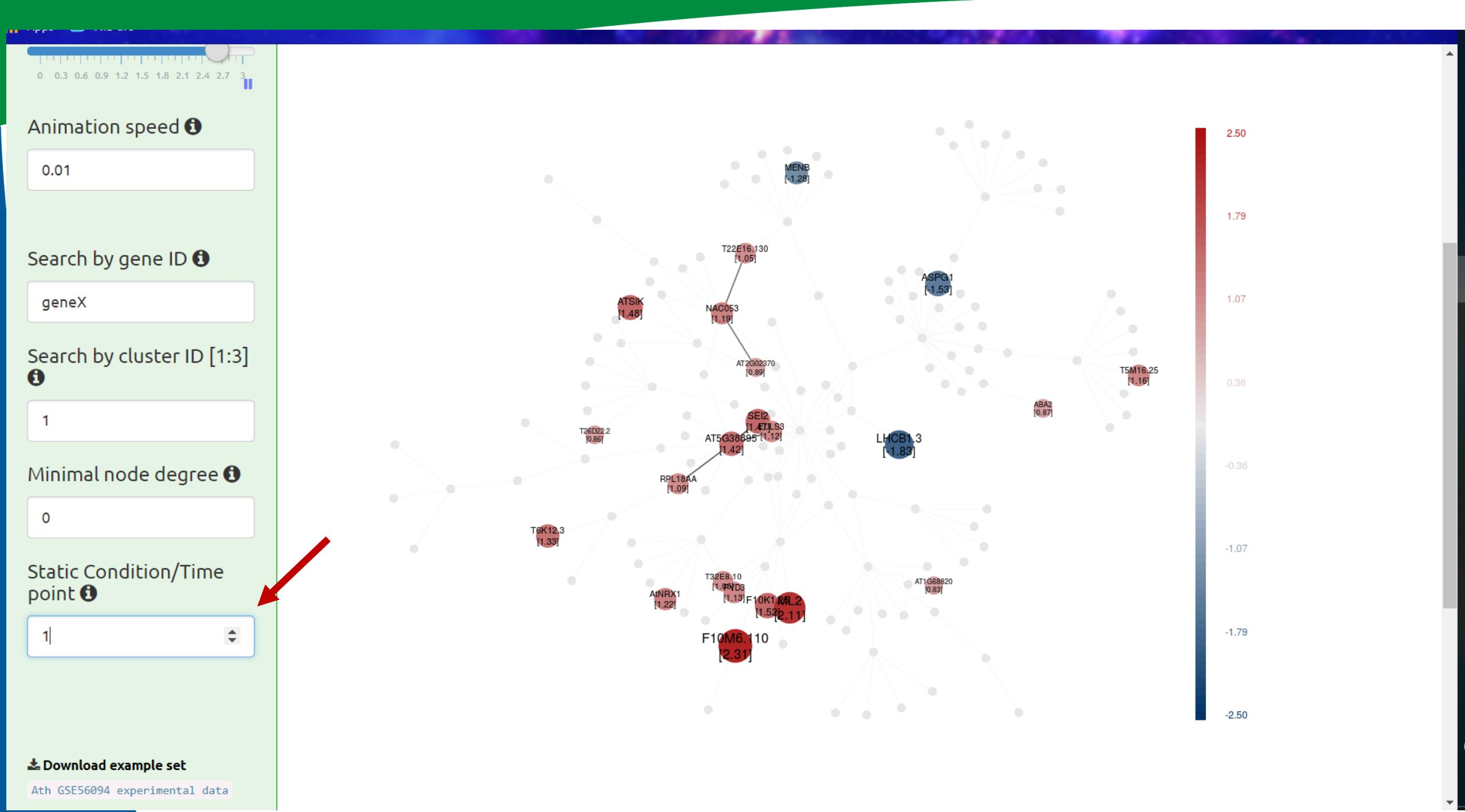
Animation speed
0.1

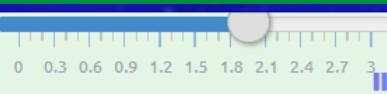
Search by gene ID
geneX

Search by cluster ID [1:48]
40

Minimal node degree
1







Animation speed i

0.01

Search by gene ID i

geneX

Search by cluster ID [1:3] i

1

Minimal node degree i

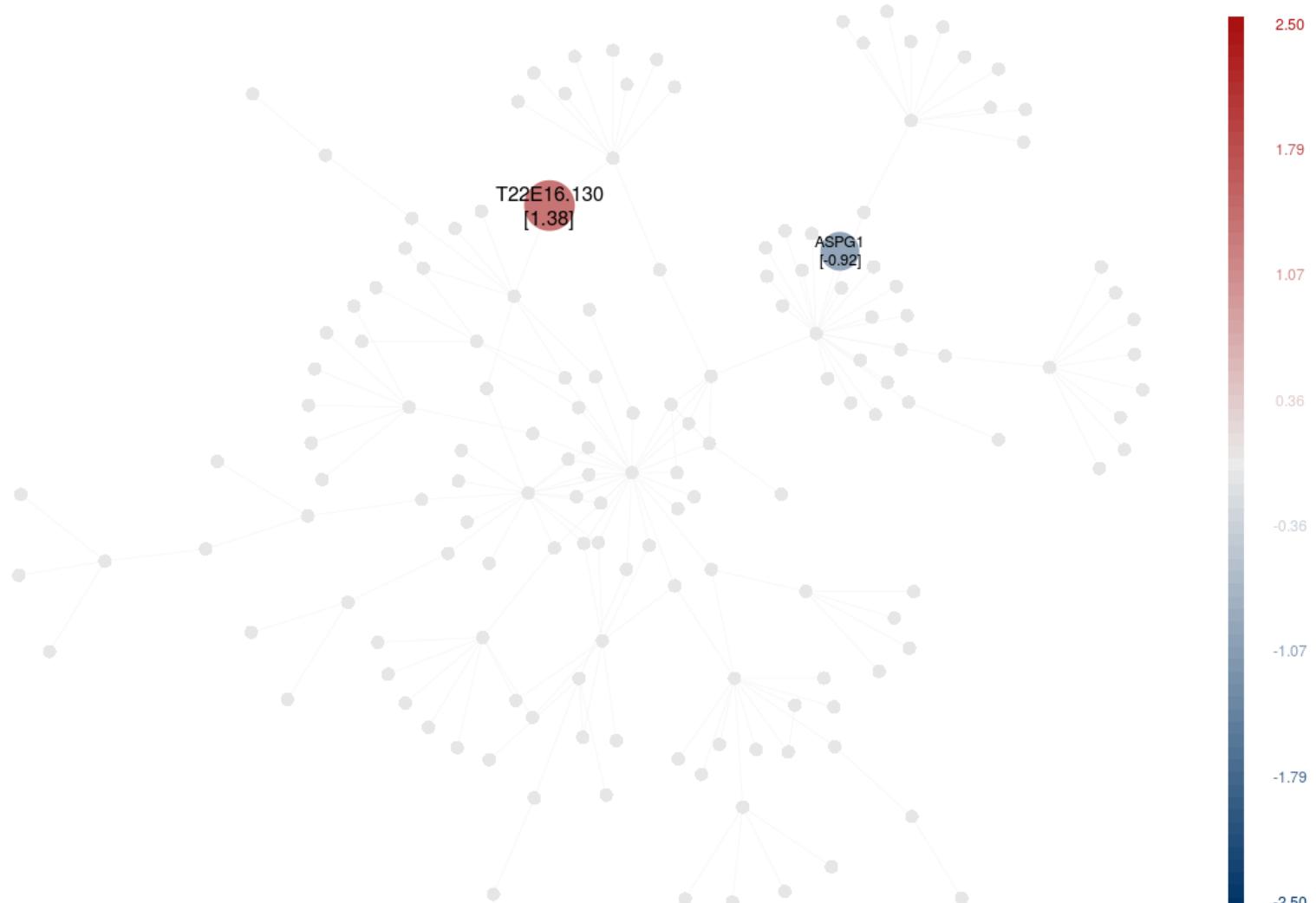
0

Static Condition/Time point i

2

Download example set

[Ath GSE56094 experimental data](#)





Animation speed i

0.01

Search by gene ID i

geneX

Search by cluster ID [1:3] i

1

Minimal node degree i

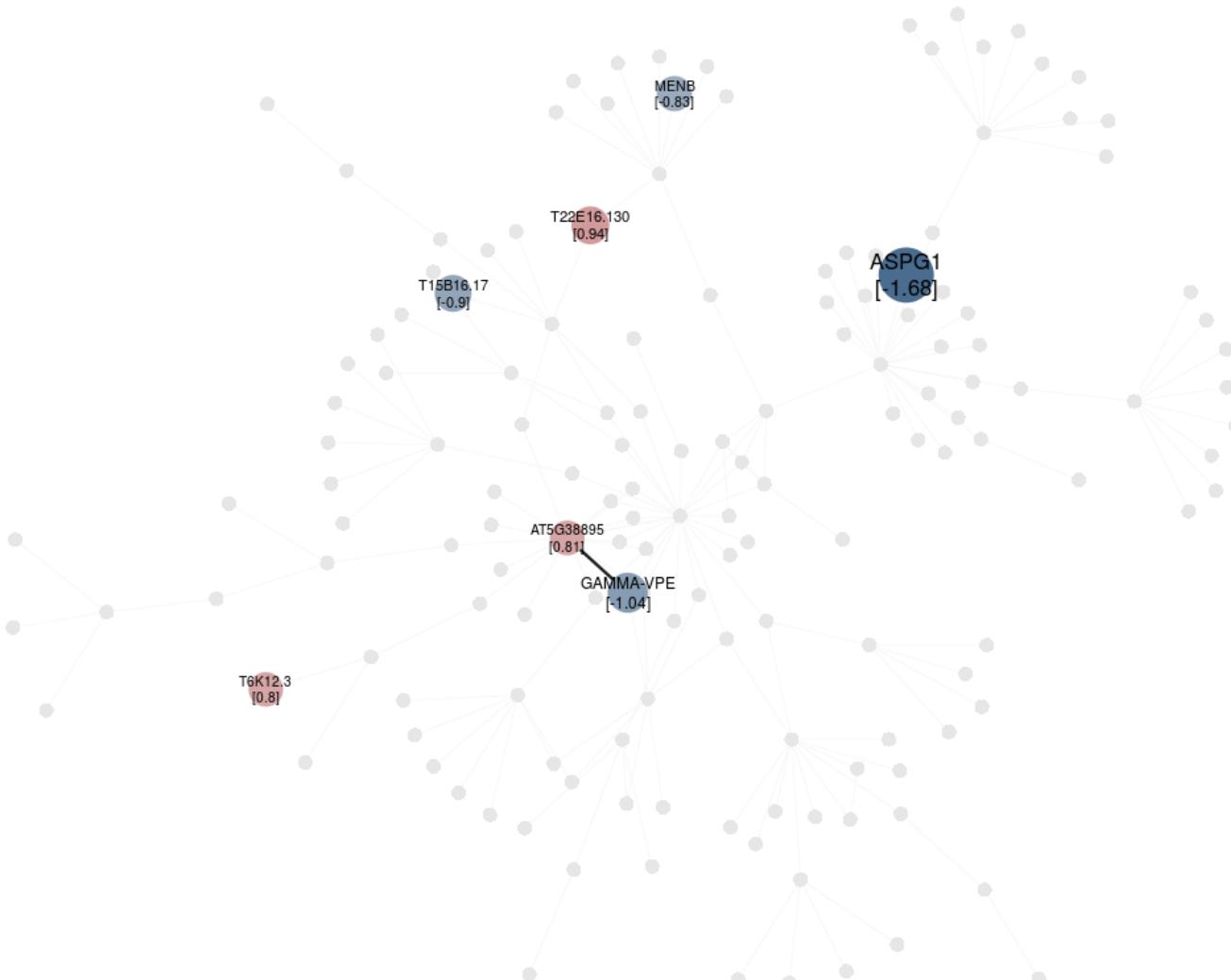
0

Static Condition/Time point i

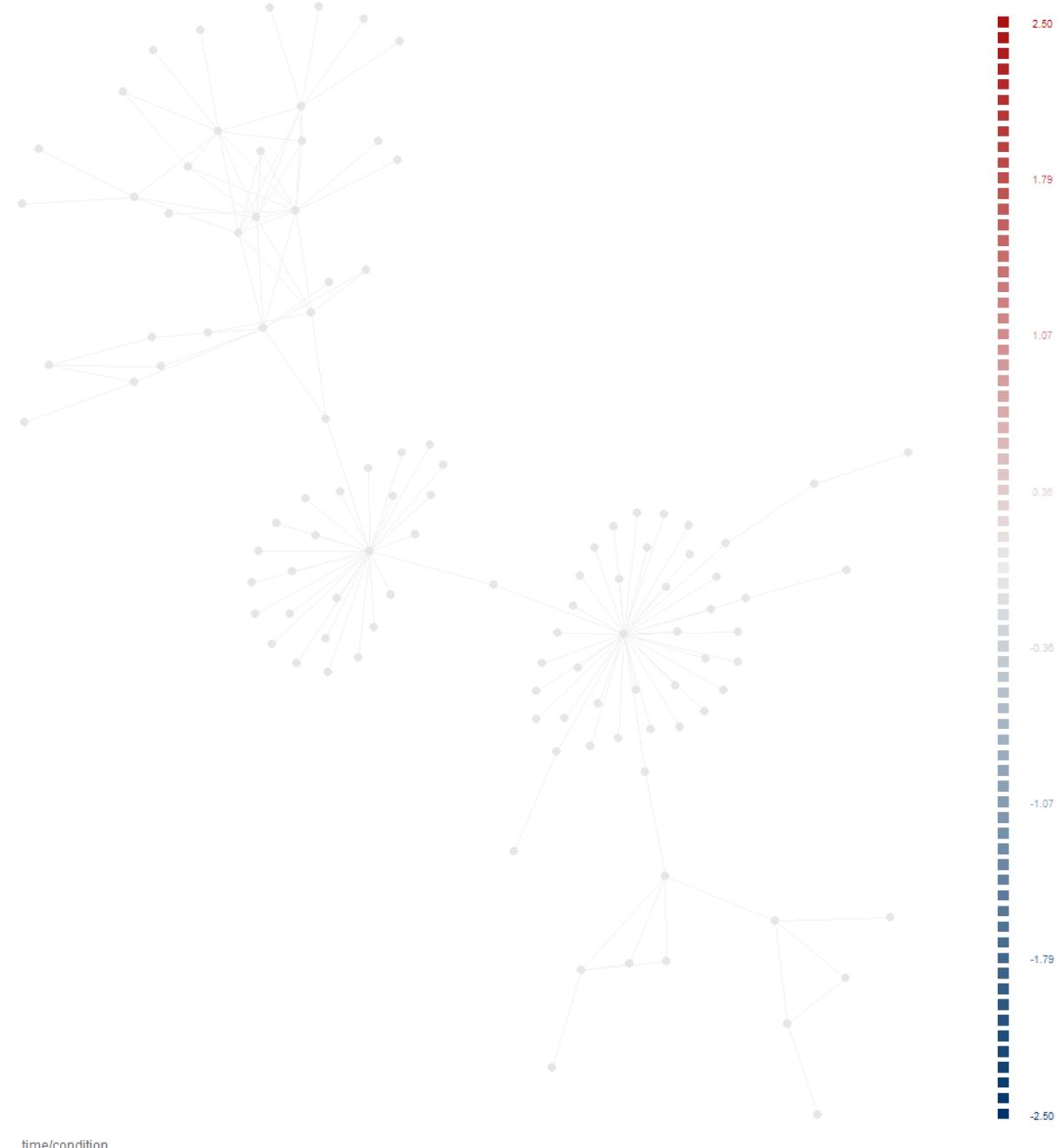
3

Download example set

[Ath GSE56094 experimental data](#)

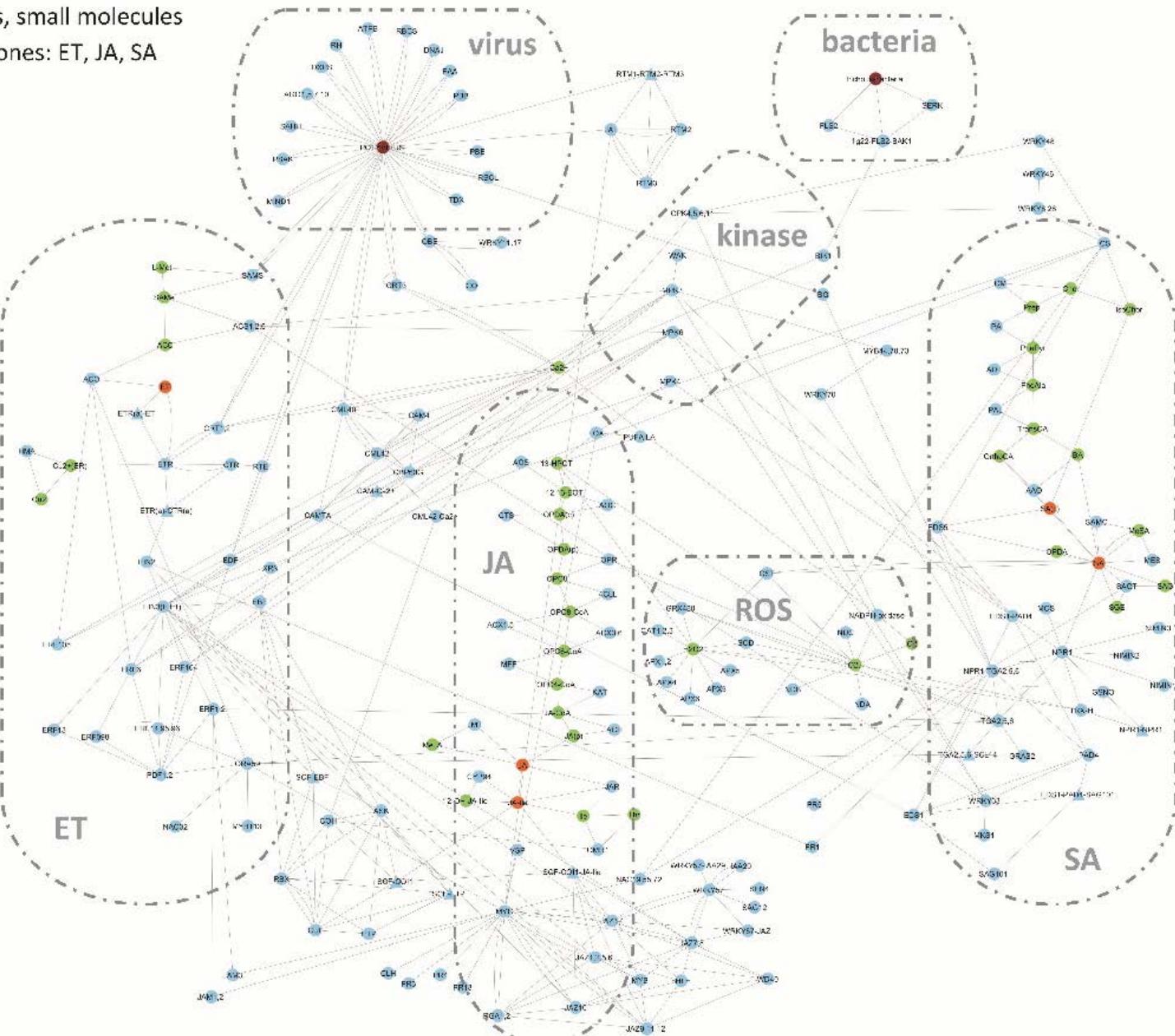


Cluster: 17 [minDegree: 1]
n: 109 e: 191

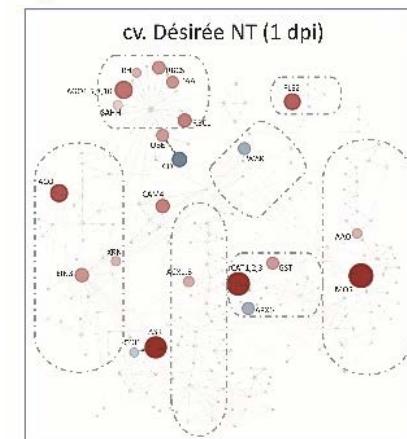


Items

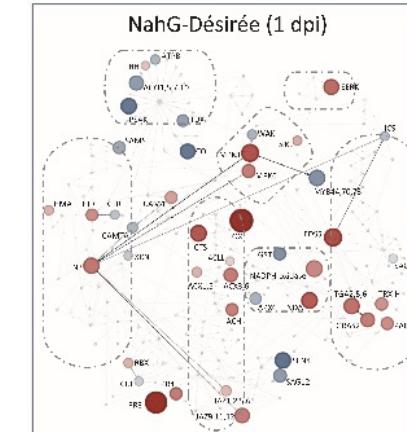
tes, small molecules
mones: ET, JA, SA

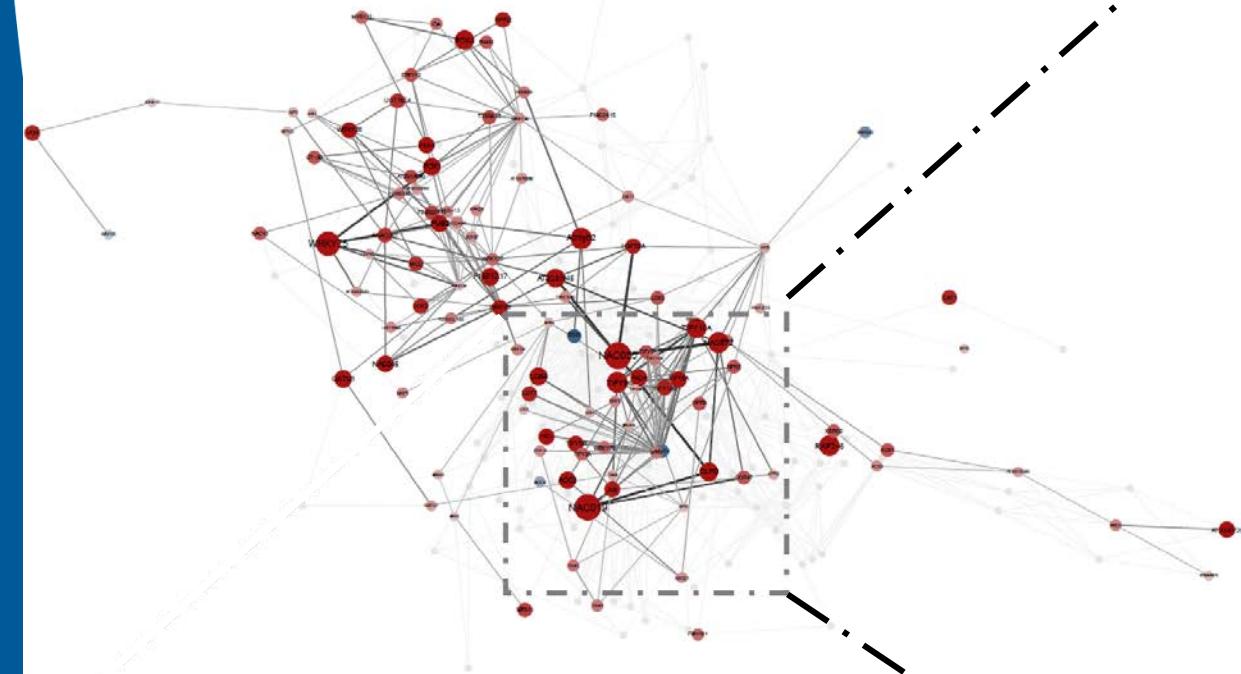
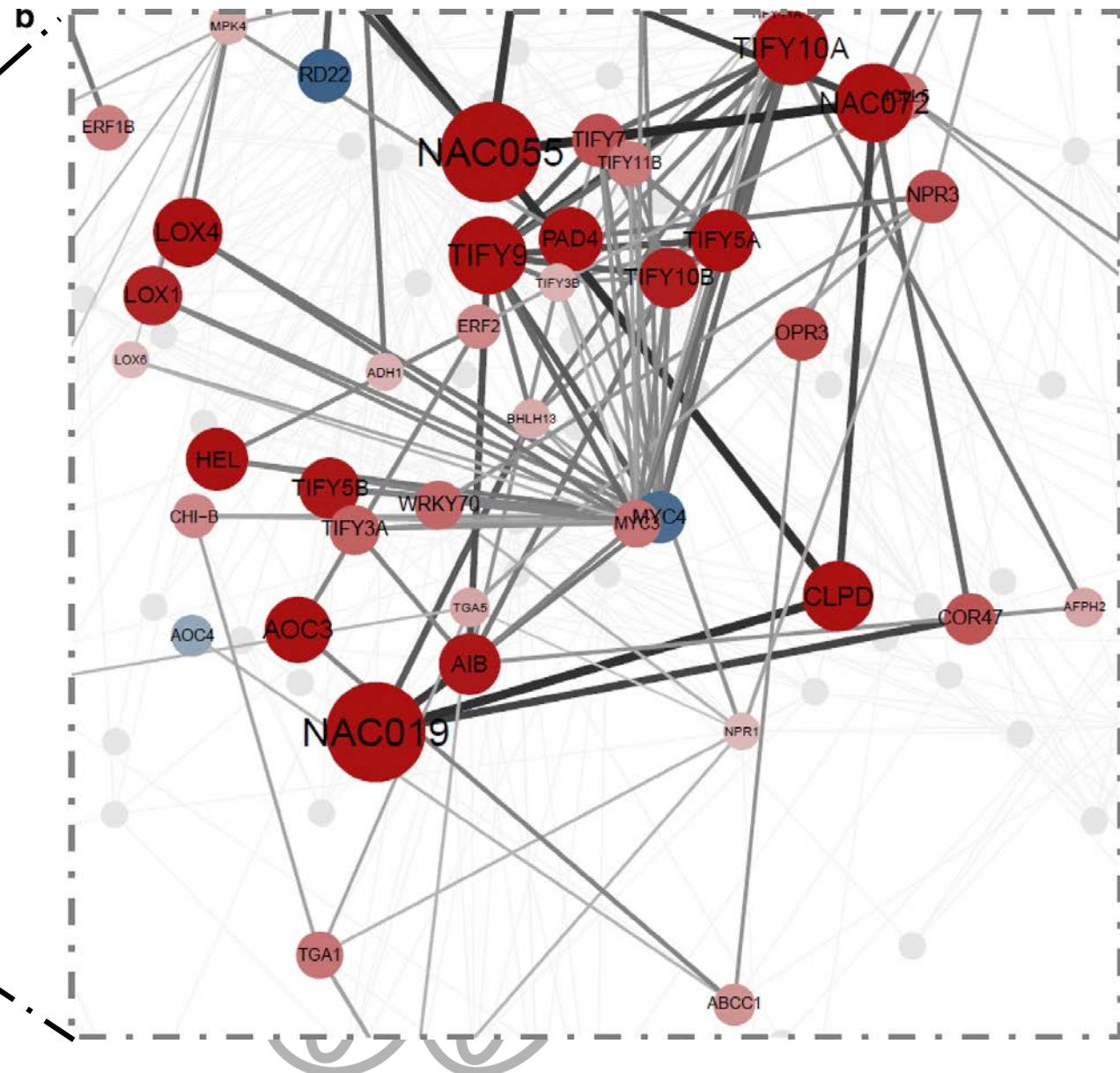


Potato + virus PVY



NahG-Désirée (1 dpi)



a**b**



Immune signaling network

Transcriptomics data
potato+virus PVY

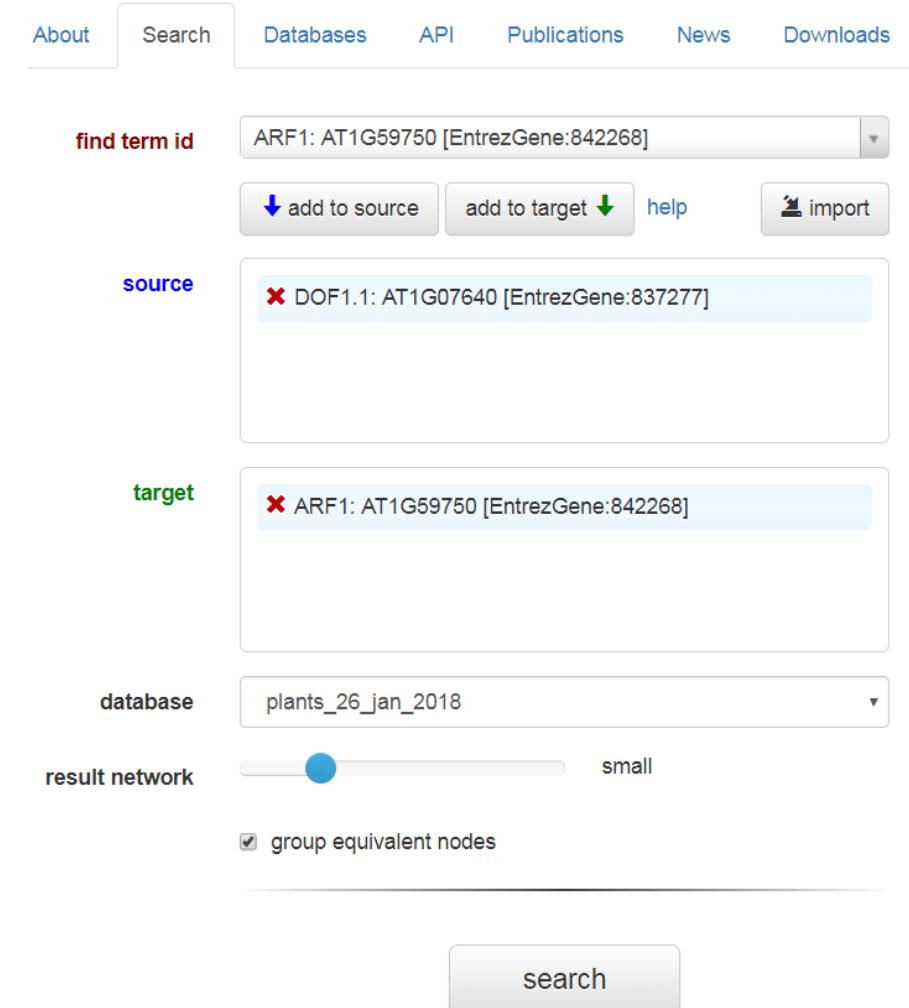


NACIONALNI INŠITUT ZA BIOLOGIJO
NATIONAL INSTITUTE OF BIOLOGY

Detailed understanding of mechanisms: Plant BioMine

Knowledge database:

- PubMed, STRING, GoMapMan, Plaza orthologues, CKN (Ramšak et al, Plant Physiol 2018)



The screenshot shows the Plant BioMine web application interface. At the top, there is a navigation bar with links for About, Search, Databases, API, Publications, News, and Downloads. Below the navigation bar, there is a search input field labeled "find term id" containing "ARF1: AT1G59750 [EntrezGene:842268]". Below the search input are several buttons: "add to source" (with a blue arrow), "add to target" (with a green arrow), "help", and "import".

The interface is divided into two main sections: "source" and "target". The "source" section contains a list box with one item: "DOF1.1: AT1G07640 [EntrezGene:837277]". The "target" section also contains a list box with one item: "ARF1: AT1G59750 [EntrezGene:842268]".

Below these sections are dropdown menus for "database" (set to "plants_26_jan_2018") and "result network" (set to "small"). There is also a checked checkbox for "group equivalent nodes". At the bottom right is a large "search" button.

Link discovery function:

- find the connection between ARF TFs and DOF TFs

Podpečan et al, Bioinformatics 2019

<http://biomine.ijs.si/>

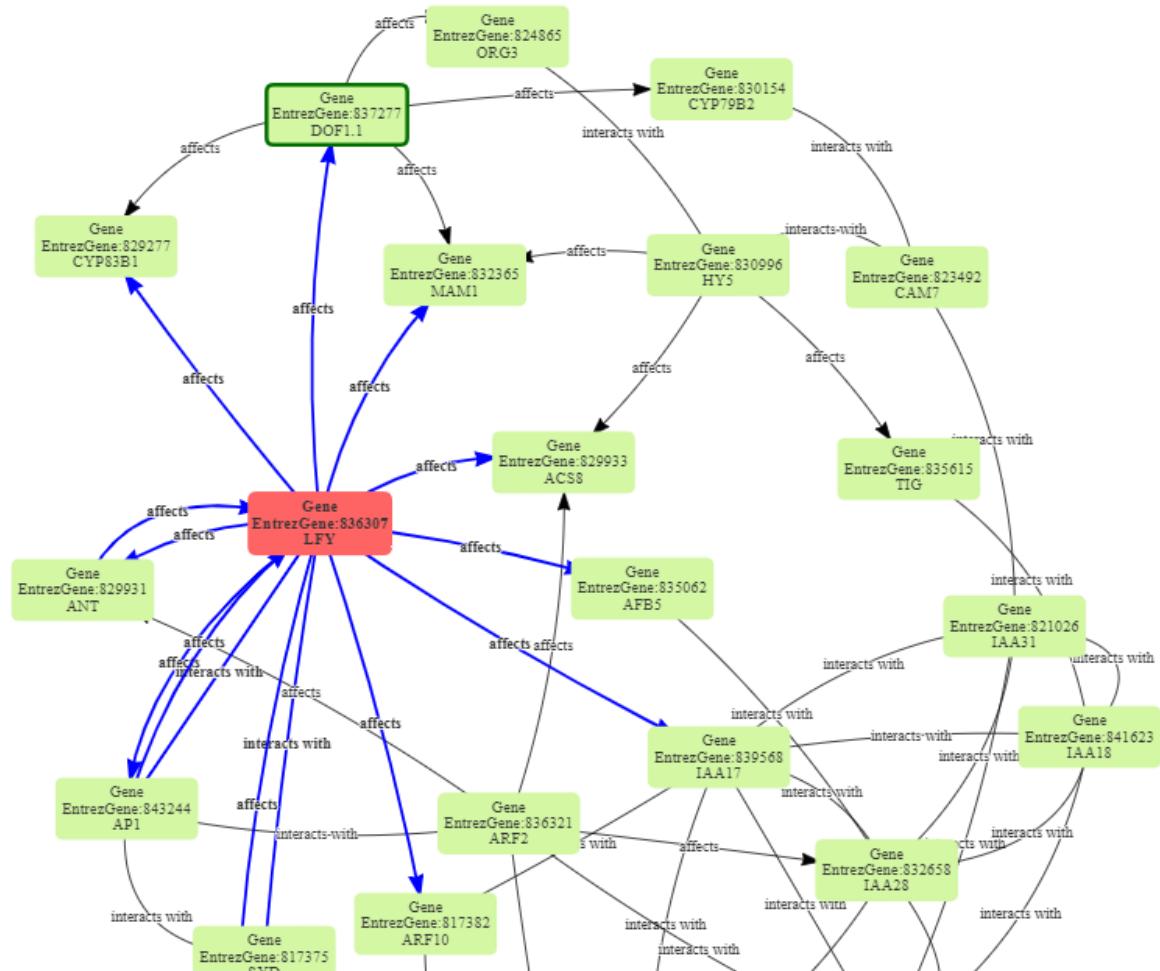
Plant BioMine

Knowledge database:

- PubMed, STRING, GoMapMan, Plaza orthologues, CKN (Ramšak et al, Plant Physiol 2018)

Link discovery function:

- find the connection between ARF TFs and DOF TFs



Novel transcription regulation subnetwork identified: DOF1-LFY/ANT-IAA-ARF

Podpečan et al, Bioinformatics 2019
<http://biomine.ijs.si/>

Acknowledgments

MapMan4, GoMapMan

www.plabipd.de/portal/web/guest/mercator4
www.gomapman.org

Analysis of multiconditional multilevel experiments: **DiNAR**

nib-si.shinyapps.io/DiNAR

Analysis of multilevel experiments

PaintOmics

www.paintomics.org

Detailed mechanistic exploration

PlantBiomine

biomine.ijs.si

Živa Ramšak, Špela Baebler, NIB

Usadel's lab: RWTH Aachen University, Germany



**RWTHAACHEN
UNIVERSITY**

Maja Zagorščak, Andrej Blejec, NIB




PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Marko Petek, NIB

Conesa's lab, CIPF, Spain

Živa Ramšak, NIB

Nada Lavrač, Vid Podpečan, IJS, Slovenia


Jožef Stefan
Institute
Ljubljana, Slovenija

