#### **CG920 Genomics**

#### Lesson 11

**Systems Biology** 

Jan Hejátko

**Functional Genomics and Proteomics of Plants**,

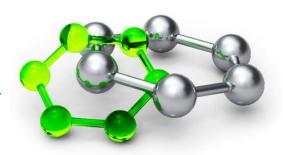
CEITEC - Central European Institute of Technology
And

**National Centre for Bimolecular Research,** 

Faculty of Science,

Masaryk University, Brno <a href="mailto:hejatko@sci.muni.cz">hejatko@sci.muni.cz</a>, <a href="mailto:www.ceitec.eu">www.ceitec.eu</a>





#### Literature

#### Literature sources for Chapter 12:

- Wilt, F.H., and Hake, S. (2004). Principles of Developmental Biology. (New York; London: W. W. Norton)
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.
- The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796-815.
- Benitez, M. and Hejatko, J. Dynamics of cell-fate determination and patterning in the vascular bundles of Arabidopsis thaliana (submitted)
- de Luis Balaguer MA, Fisher AP, Clark NM, Fernandez-Espinosa MG, Moller BK, Weijers D, Lohmann JU, Williams C, Lorenzo O, Sozzani R. 2017. Predicting gene regulatory networks by combining spatial and temporal gene expression data in Arabidopsis root stem cells. Proc Natl Acad Sci U S A 114(36): E7632-E7640.



#### **Outline**

- Definition of Systems Biology
- Tools
  - Gene Ontology Analysis
  - Bayesian Networks
  - Molecular/Gene Regulatory Networks Modeling
  - Inferring Gene Regulatory Networks from Large Omics
     Datasets



#### **Definition**

Systems biology is the computational and mathematical analysis and modeling of complex biological systems. It is a biology-based interdisciplinary field of study that focuses on complex interactions within biological systems, using a holistic approach (holism instead of the more traditional reductionism) to biological research (Wikipedia).



#### **Definition**

Systems biology is the study of biological systems whose behaviour cannot be reduced to the linear sum of their parts' functions. Systems biology does not necessarily involve large numbers of components or vast datasets, as in genomics or connectomics, but often requires quantitative modelling methods borrowed from physics (Nature).



#### **Definition**

Nice explanatory video by Dr. Nathan Price,

associate director of the Institute for Systems Biology at <a href="https://www.youtube.com/watch?v=OrXRI\_8UFHU">https://www.youtube.com/watch?v=OrXRI\_8UFHU</a>.





### **Outline**

Definition of Systems Biology

- Tools
  - Gene Ontology analysis



### Results of -omics Studies vs **Biologically Relevant Conclusions**

Results of -omics studies represent huge amount of data, e.g. genes with differential expression. But how to get any biologically relevant conclusions out of it?

Ddii et al.. unpublished

gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat		-	significant
AT1G07795	1:2414285-2414967	WT	MT	ОК	0	1,1804	1.79769e+308	1.79769e+3 08	6.88885e-05	0,00039180	) 1 yes
HRS1	1:4556891-4558708	WT	MT	OK	0	0,696583	1.79769e+308	1.79769e+3 08	6.61994e-06	4.67708e- 05	yes
ATMLO14	1:9227472-9232296	WT	MT	OK	0	0,514609	1.79769e+308	1.79769e+3 08	9.74219e-05	0,00053505	5 5 yes
NRT1.6	1:9400663-9403789	WT	MT	OK	0	0,877865	1.79769e+308	1.79769e+3 08	3.2692e-08	3.50131e- 07	yes
AT1G27570	1:9575425-9582376	WT	MT	ОК	0	2,0829	1.79769e+308	1.79769e+3 08	9.76039e-06	6.647e-05	yes
AT1G60095	1:22159735-22162419	WT	MT	OK	0	0,688588	1.79769e+308	1.79769e+3 08	9.95901e-08	9.84992e- 07	yes
AT1G03020	1:698206-698515	WT	MT	ОК	0	1,78859	1.79769e+308	1.79769e+3 08		0,0277958	3 yes
AT1G13609	1:4662720-4663471	WT	MT	ОК	0	3,55814	1.79769e+308	1.79769e+3 08		0,00108079	) yes
AT1G21550	1:7553100-7553876	WT	MT	ОК	0	0,562868	1.79769e+308	1.79769e+3 08		0,00471497	ves
AT1G22120	1:7806308-7809632	WT	MT	ОК	0	0.617354	1.79769e+308	1.79769e+3 08		1.91089e-	ves
AT1G31370	1:11238297-11239363	WT	MT	OK	0	1,46254	1.79769e+308	1.79769e+3 08	4.83523e-05	0,00028514	1 Byes
APUM10	1:13253397-13255570	WT	MT	ОК	0	0,581031	1.79769e+308	1.79769e+3 08	7.87855e-06	5.46603e- 05	yes
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0,556525	1.79769e+308	1.79769e+3		0,00037473	3 Syes
AT1G59077	1:21746209-21833195	WT	MT	OK	0	138,886	1.79769e+308	1.79769e+3 08		0,00496816	
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0.370087	1.79769e+308	1.79769e+3 08			
AT4G15242	4:8705786-8706997			OK	0,00930712				1.05673e-05		
AT5G33251	5:12499071-12500433	WT	MT	OK	0,0498375	52,2837	10,0349	-9,8119	0		0 yes
AT4G12520	4:7421055-7421738			OK	0,0195111	15,8516		-3,90043	9.60217e-05		
AT1G60020 AT5G15360	1:22100651-22105276 5:4987235-4989182			OK OK	0,0118377 0,0988273	7,18823 56,4834		-7,50382 -10,4392	6.19504e-14 0		,
A13G13300	5:4987235-4989182	VVI	IVI I	UK	0,0988273	56,4834	9,1587	-10,4392			0 yes

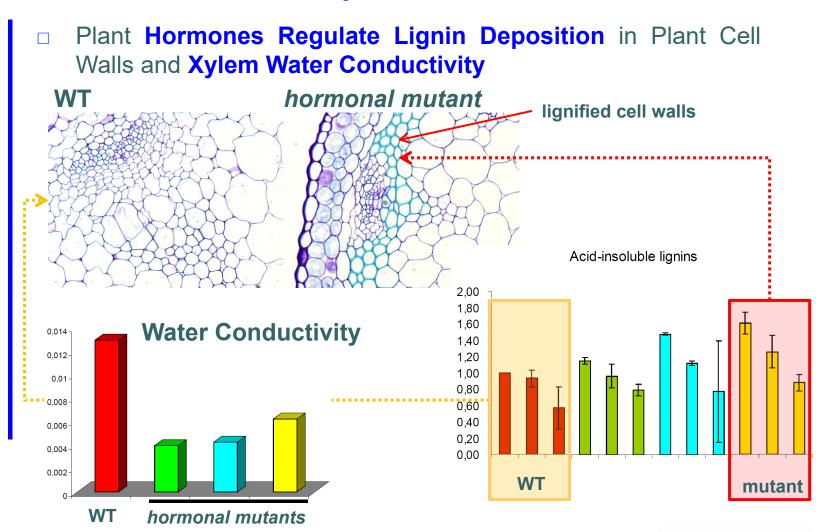
# Plant Vascular Tissue Development

Procambium Xylem

Vascular tissue as a developmental model for GO analysis and MRN modeling (a) (b) Cotyledon (c) Photosyntheti Lateral nutrients Primary root Root cap Lehesranta etal., Trends in Plant Sci (2010)

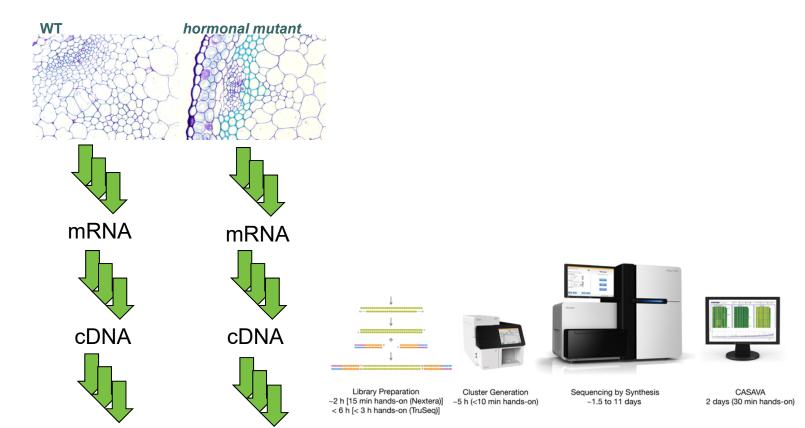


# Hormonal Control Over Vascular Tissue Development



# Hormonal Control Over Vascular Tissue Development

Transcriptional profiling via RNA sequencing



Sequencing by Illumina and

**number of transcripts** determination



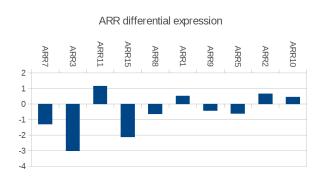
## Results of –omics Studies vs Biologically Relevant Conclusions

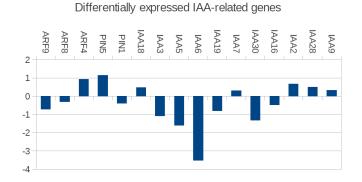
Transcriptional profiling yielded more then 9K differentially regulated genes...

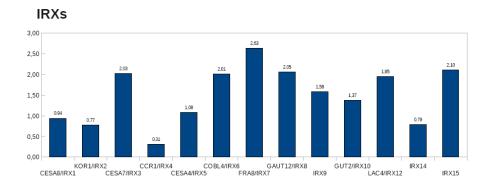
Ddii et al.. unpublished

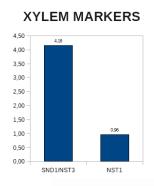
gene	locus	sample 1	sample 2	status	value 1	value 2	log2(fold_change)	test stat	p_value	q value	significant
		_			_	_	3 ( 1 1 2 1 3 1 )	1.79769e+3		0,00039180	
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1,1804	1.79769e+308		6.88885e-05	1	l yes
UDO4	4 4550004 4550700			014		0.000500	4.70700 .000	1.79769e+3		4.67708e-	
HRS1	1:4556891-4558708	WT	MT	OK	0	0,696583	1.79769e+308	08 1.79769e+3	6.61994e-06		yes
ATMLO14	1:9227472-9232296	WT	МТ	OK	0	0.514609	1.79769e+308		9.74219e-05	0,00053505	yes
					-	2,2 : : : : :		1.79769e+3		3.50131e-	,, ==
NRT1.6	1:9400663-9403789	WT	MT	OK	0	0,877865	1.79769e+308		3.2692e-08	07	yes
474007570	4 0575405 0500070			014		0.000	4.70700 .000	1.79769e+3	0.70000 00	0.047.05	
AT1G27570	1:9575425-9582376	WT	MT	OK	0	2,0829	1.79769e+308	08 1.79769e+3	9.76039e-06	6.647e-05 9.84992e-	yes
AT1G60095	1:22159735-22162419	WT	MT	ок	0	0.688588	1.79769e+308		9.95901e-08		yes
						,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		1.79769e+3			ĺ
AT1G03020	1:698206-698515	WT	MT	OK	0	1,78859	1.79769e+308	08	0,00913915	0,0277958	3 yes
AT4040000	4.4000700 4000474	.A/T		OK	•	0.55044	4 70700 - 1000	1.79769e+3	0.00004000	0.00400070	
AT1G13609	1:4662720-4663471	WT	MT	OK	0	3,55814	1.79769e+308	08 1.79769e+3	0,00021683	0,00108079	yes
AT1G21550	1:7553100-7553876	WT	MT	ок	0	0.562868	1.79769e+308	08	0,00115582	0.00471497	ves
						.,		1.79769e+3		1.91089e-	
AT1G22120	1:7806308-7809632	WT	MT	OK	0	0,617354	1.79769e+308		2.48392e-06		yes
474004070	4 44000007 44000000			014		4 40054	4.70700	1.79769e+3		0,00028514	
AT1G31370	1:11238297-11239363	VV I	MT	OK	0	1,46254	1.79769e+308	08 1.79769e+3	4.83523e-05	5.46603e-	3 yes
APUM10	1:13253397-13255570	WT	MT	ок	0	0,581031	1.79769e+308		7.87855e-06		yes
								1.79769e+3		0,00037473	3
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0,556525	1.79769e+308		6.53917e-05	6	3 yes
AT1G59077	4.24746200 24822405 1	M/T	MT	ОК	0	420.000	1.79769e+308	1.79769e+3 08	0,00122789	0.00406046	2
A11G59077	1:21746209-21833195	VV I	IVI I	UK	0	130,000	1.797090+300	1.79769e+3	0,00122769	0,00490610	yes
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0,370087	1.79769e+308	08	0,00117953	0,0048001	l yes
AT4G15242	4:8705786-8706997	WT	MT	OK	0,00930712	17,9056	10,9098	-4,40523	1.05673e-05	7.13983e-05	5 yes
AT5G33251	5:12499071-12500433 \	W/T	MT	OK	0,0498375	52,2837	10,0349	-9,8119	0		0 yes
AT4G12520	4:7421055-7421738			OK	0,0496373				9.60217e-05		
	2.333 7421100				0,0.00111	.0,0010	0,00012	0,00040		2,00002000	. ,
AT1G60020	1:22100651-22105276			OK	0,0118377	7,18823			6.19504e-14	1.4988e-12	yes
AT5G15360	5:4987235-4989182	WT	MT	OK	0,0988273	56,4834	9,1587	-10,4392	0		0 yes

One of the possible approaches is to study **gene ontology**, i.e. previously demonstrated **association** of genes to **biological processes** 



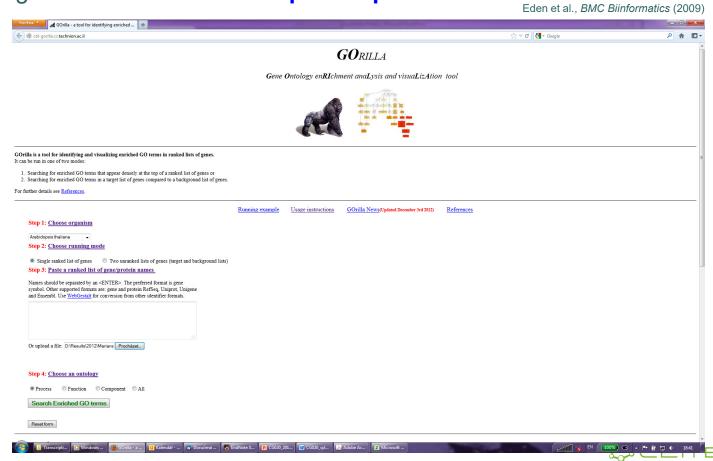


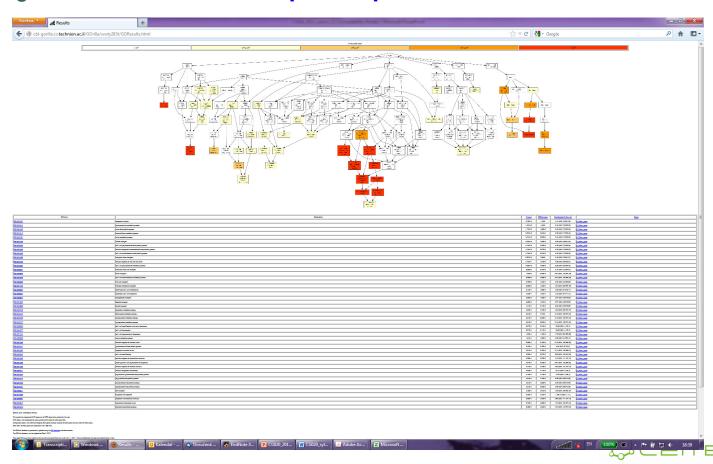


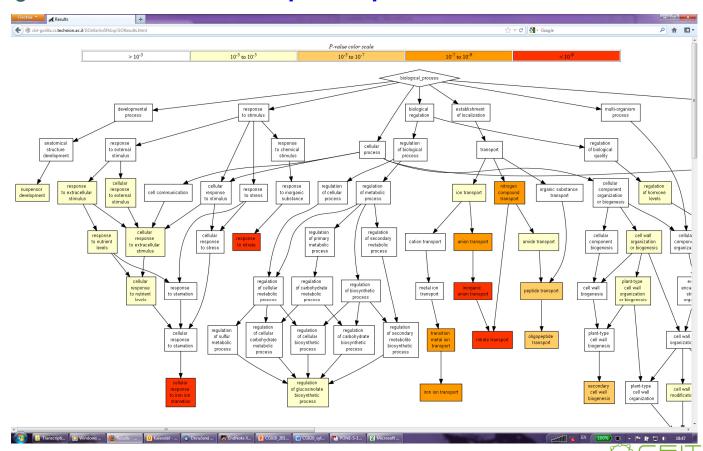


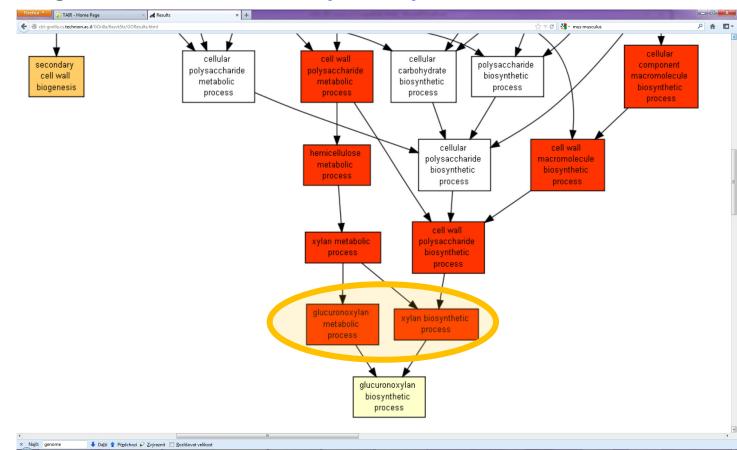
CEITEC



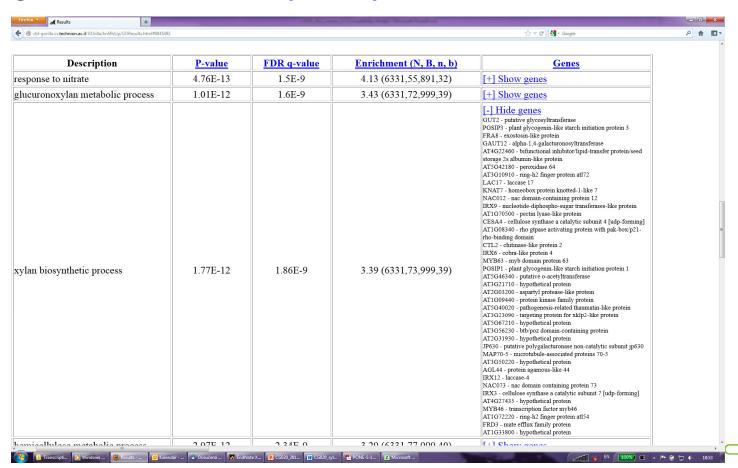








Results  Cbl-gorilla.cs.technion.ac.il/GOrilla/kn5fh1qi/GOResult	.html#0045492	(MAX) (ive	∱ ∀ €   <mark>% •</mark> •	ioogle	<i>P</i> ♠ <b>E</b>
glucuronoxylan metabolic process	1.01E-12	1.6E-9	3.43 (6331,72,999,39)	[+] Show genes	
xylan biosynthetic process	1.77E-12	1.86E-9	3.39 (6331,73,999,39)	[+] Show genes	
hemicellulose metabolic process	2.97E-12	2.34E-9	3.29 (6331,77,999,40)	[+] Show genes	
xylan metabolic process	3.21E-12	2.03E-9	3.34 (6331,74,999,39)	[+] Show genes	
nitrate transport	3.64E-12	1.92E-9	3.92 (6331,58,891,32)	[+] Show genes	
cell wall polysaccharide biosynthetic process	5.74E-12	2.59E-9	3.30 (6331,75,999,39)	[+] Show genes	
cellular component macromolecule biosynthetic process	5.74E-12	2.27E-9	3.30 (6331,75,999,39)	[+] Show genes	



#### **Outline**

- **Definition of Systems Biology**
- - Gene Ontology analysisBayesian Networks



# **Bayesian Networks**

- What are Bayesian networks?
  - Probabilistic Graphical Model that can be used to build models from data and/or expert opinion



# **Bayesian Networks**

#### What are Bayesian Networks?

- Probabilistic Graphical Model that can be used to build models from data and/or expert opinion
- can be used for a wide range of tasks including prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction and decision making under uncertainty

#### NODES

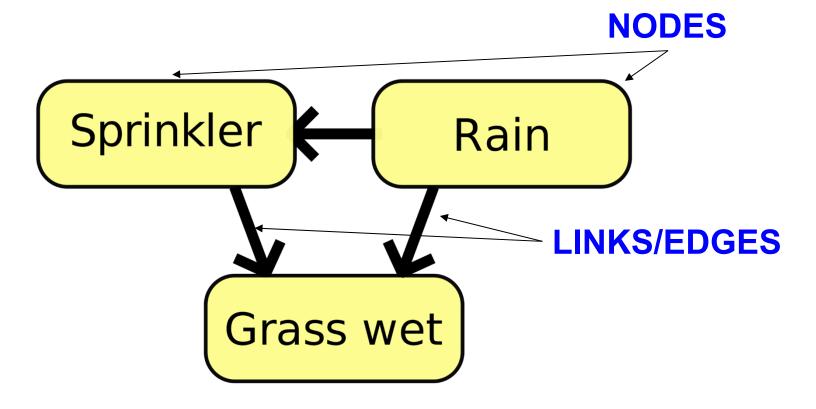
each node represents a variable such as someone's height, age or gender.
 A variable might be discrete, such as Gender = {Female, Male} or might be continuous such as someone's age

#### LINKS

added between nodes to indicate that one node directly influences the other

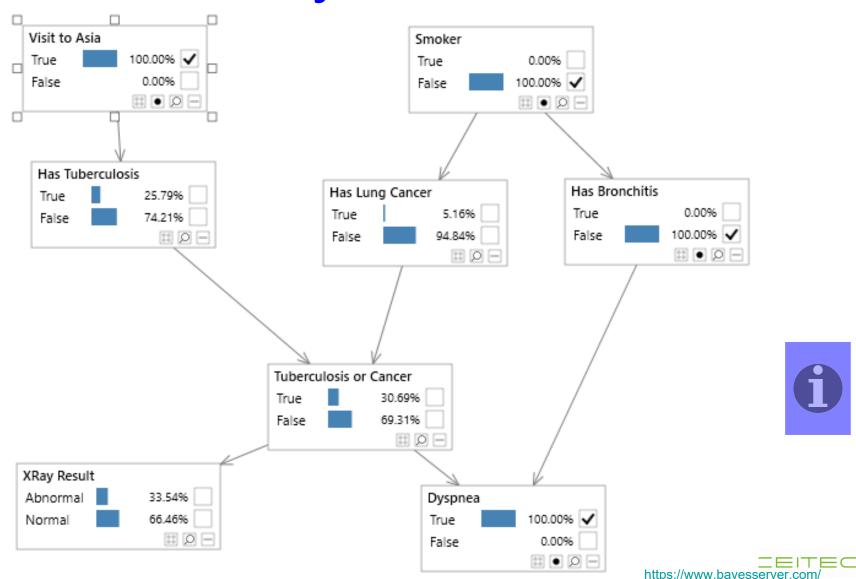


# **Bayesian Networks**





# Asia Bayesian Network



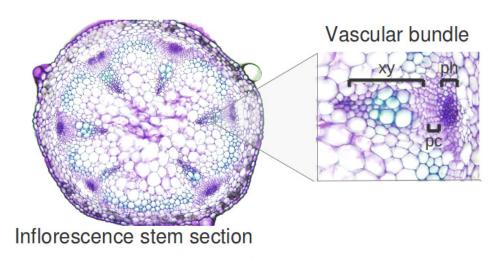
#### **Outline**

- **Definition of Systems Biology**
- Gene Ontology analysisBayesian Networks

  - Molecular/Gene Regulatory Networks Modeling



Vascular tissue as a developmental model for MRN modeling



Benitez and Hejatko, PLoS One, 2013



#### Literature search for published data and creating small database

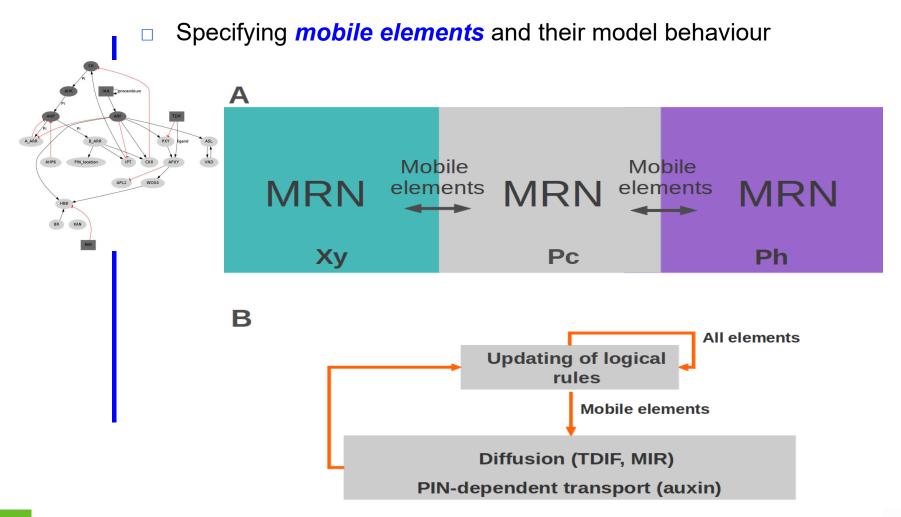
Interaction	Evidence	References
A-ARRs –  CK signaling	Double and higher order type-A ARR mutants show increased sensitivity to CK.	[27]
	Spatial patterns of A-type ARR gene expression and CK response are consistent with partially redundant function of these genes in CK signaling.	[27]
	A-type ARRs decreases B-type ARR6-LUC.	[13]
	Note: In certain contexts, however, some A-ARRs appear to have effects antagonistic to other A-ARRs.	[27]
AHP6 –  AHP	ahp6 partially recovers the mutant phenotype of the CK receptor WOL.	[9]
Benitez and Hejatko, <i>PLoS One</i>	Using an in vitro phosphotransfer system, it was shown that, unlike the AHPs, native AHP6 was unable to accept a phosphoryl group. Nevertheless, AHP6 is able to inhibit phosphotransfer from other p.AHPs to ARRs.	[9]
		~~~~

Formulating *logical rules* defining the *model dynamics* 

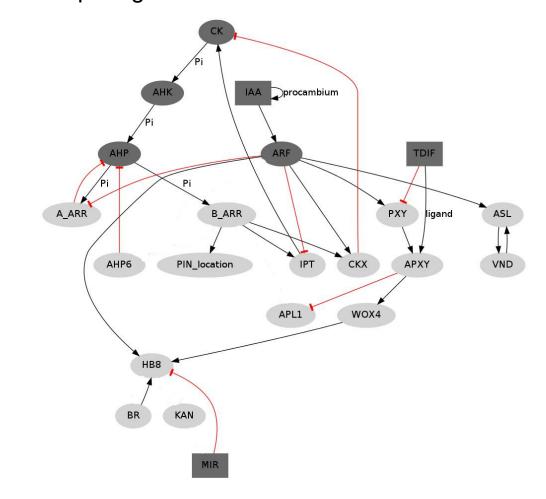
Network node	Dynamical rule
CK	2 If ipt=1 and ckx=0 1 If ipt=1 and ckx=1 0 else
CKX	1 If barr>0 or arf=2 0 else
AHKs	ahk=ck
AHPs	2 If ahk=2 and ahp6=0 and aarr=0 1 If ahk=2 and (ahp6+aarr<2) 1 If ahk=1 and ahp6<1 0 else
B-Type ARRs	1 If ahp>0 0 else
A-Type ARRs	1 If arf<2 and ahp>0 0 else







Preparing the *first version* of the model and its *testing* 

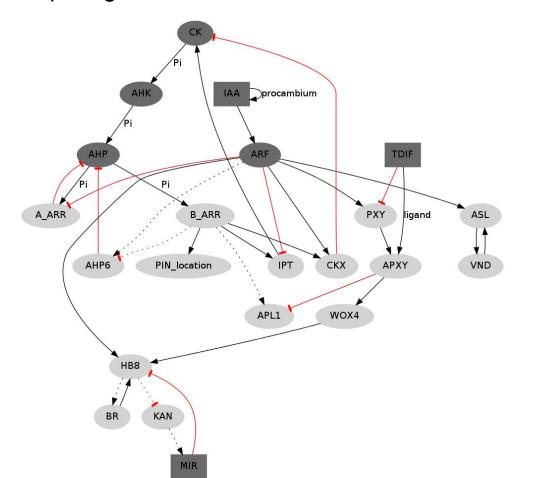




#### Specifying of missing interactions via informed predictions

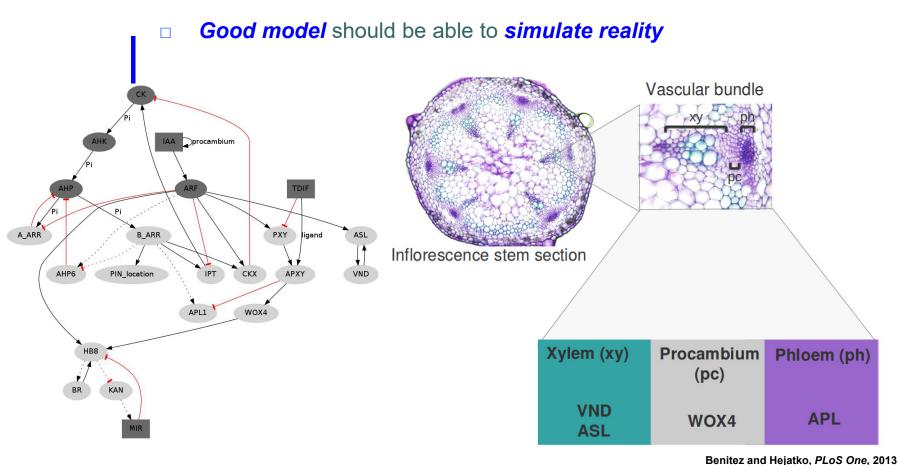
Interaction	Evidence	References
CK → PIN7 radial localization	Predicted interaction (could be direct or indirect)	
	Informed by the following data:	
	During the specification of root vascular cells in Arabidopsis thaliana, CK regulates the radial localization of PIN7.	[18]
	Expression of PIN7:GFP and PIN7::GUS is upregulated by CK with no significant influence of ethylene.	[18,20]
	In the root, CK signaling is required for the CK regulation of PIN1, PIN3, and PIN7. Their expression is altered in wol, cre1, ahk3 and ahp6 mutants.	[19]
$CK \rightarrow APL$	Predicted interaction (could be direct or indirect)	
	Consistent with the fact that APL overexpression prevents or delays xylem cell differentiation, as does CKs.	[21]
		(TAIR,
	Partially supported by microarray data and phloem-specific expression patterns of CK response factors.	ExpressionSet: 1005823559, [22])

Preparing the next version of the model and its testing

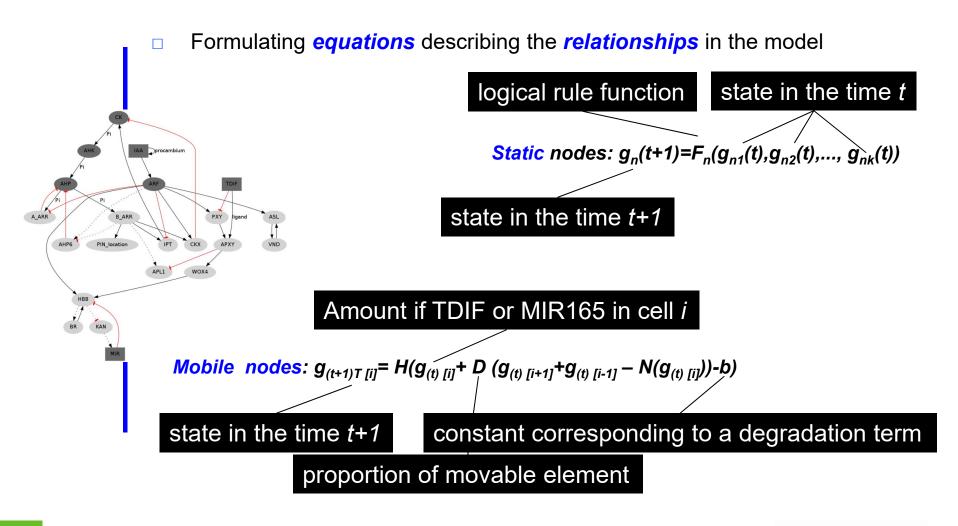


Benitez and Hejatko, PLoS One, 2013

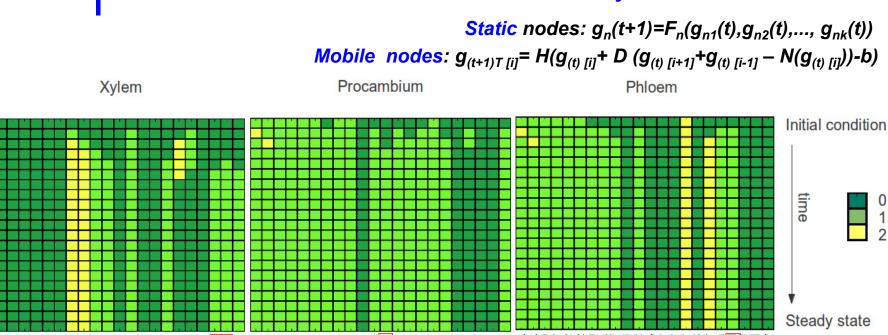








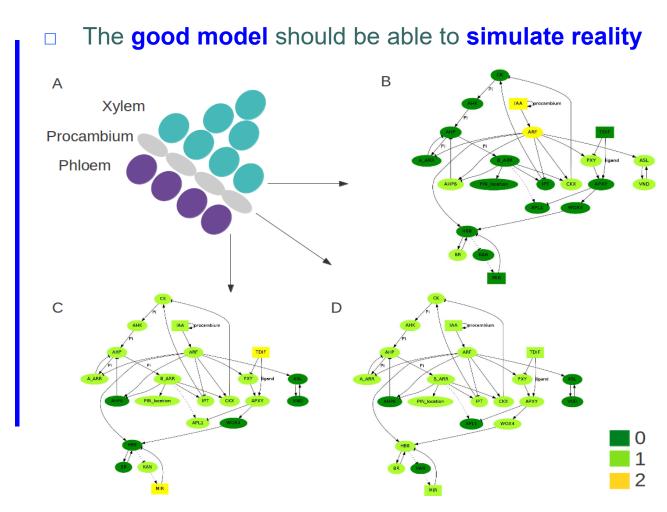
Good model should be able to simulate reality



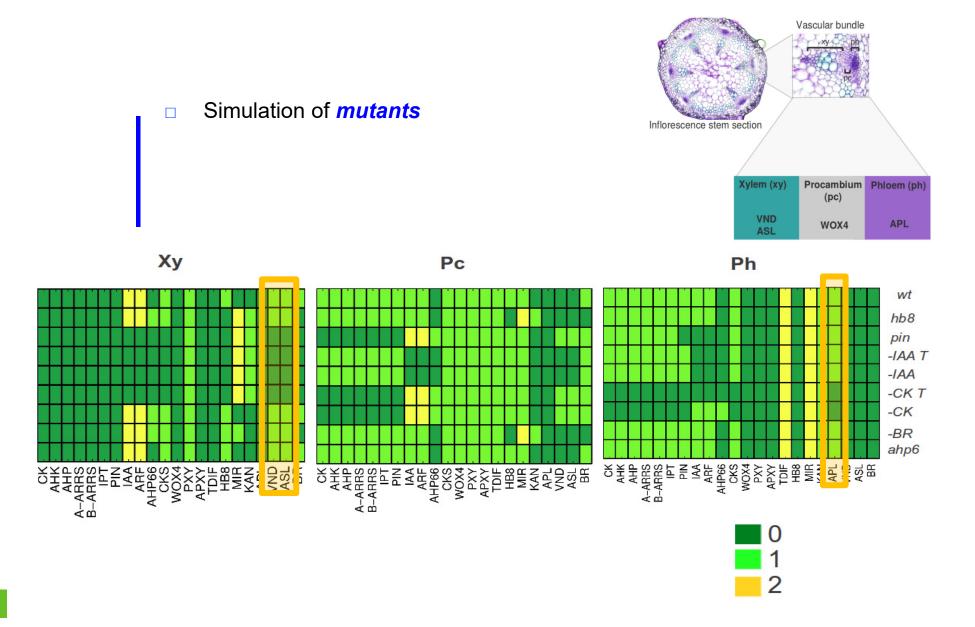
Auxin

BR\_ CK

......Benitez and Hejatko, submitted



#### Molecular Regulatory Networks Modeling

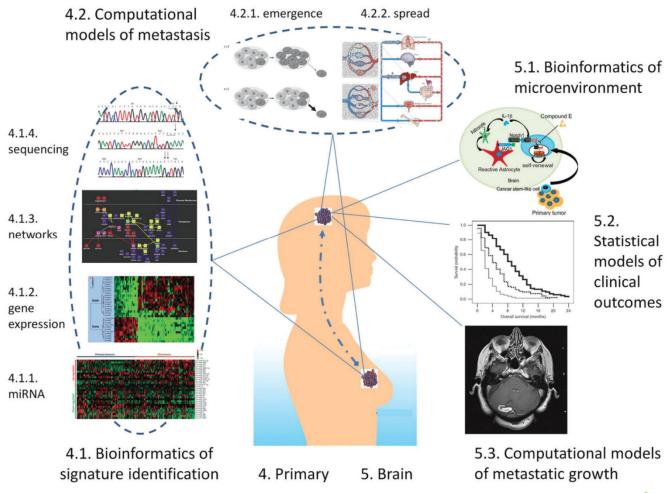


#### **Outline**

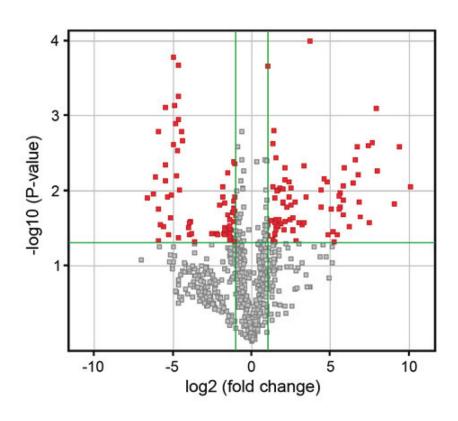
- Definition of Systems Biology
- Tools
  - Gene Ontology analysis
  - Bayesian Networks
  - Molecular/Gene Regulatory Networks Modeling
  - Inferring Gene Regulatory Networks from Large Omics
     Datasets



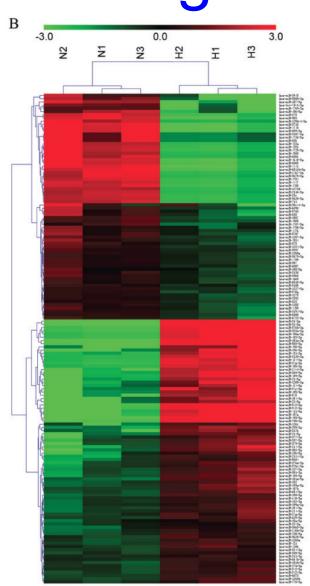
## Systems Biology in Cancer Research

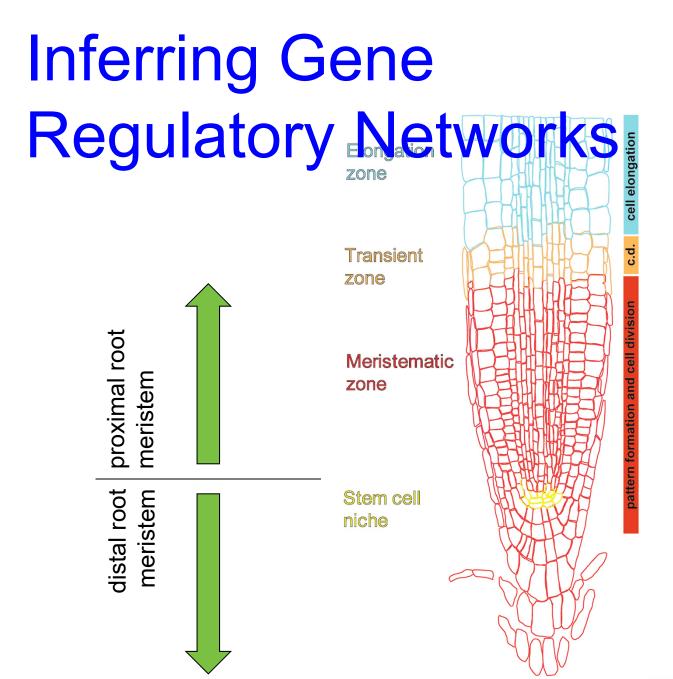


### miRNA/mRNA Profiling

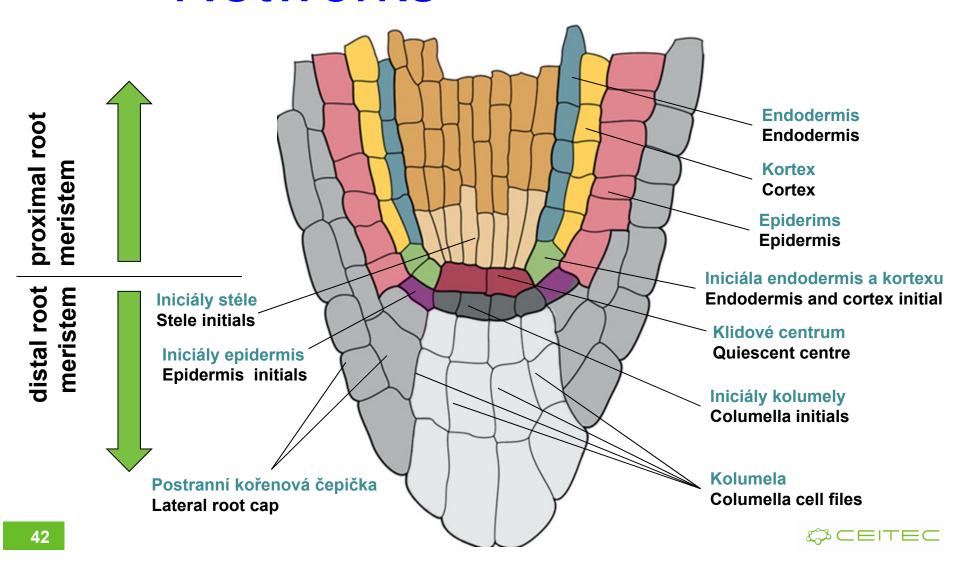


Guo et al., Mol Med Reports, 2017





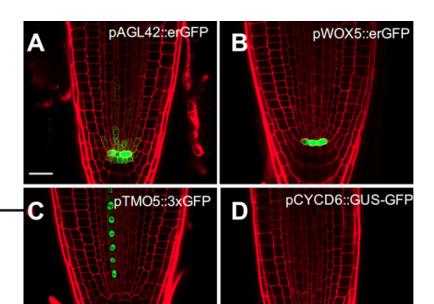
#### Gene Regulatory Networks



Inferring GRNs via GENIST

 GEne regulatory Network Inference from SpatioTemporal data algorithm

 Combining spatial- and timespecific gene expression profiles

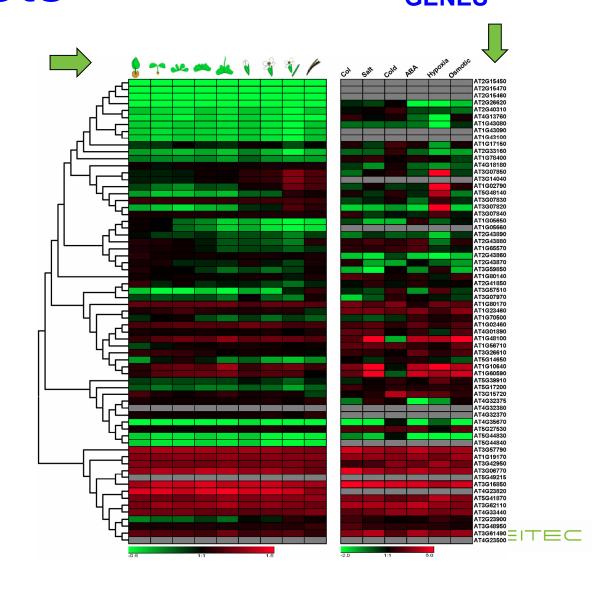


Birnbaum et al., Science, 2003

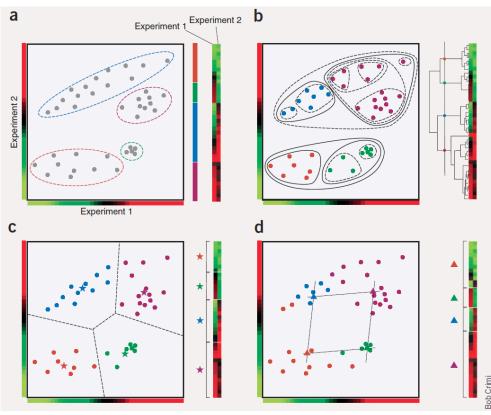
de Luis Balaguer et al., PNAS, 2017

# Combining Large Omics Datasets GENES

**TISSUE/TIME** 



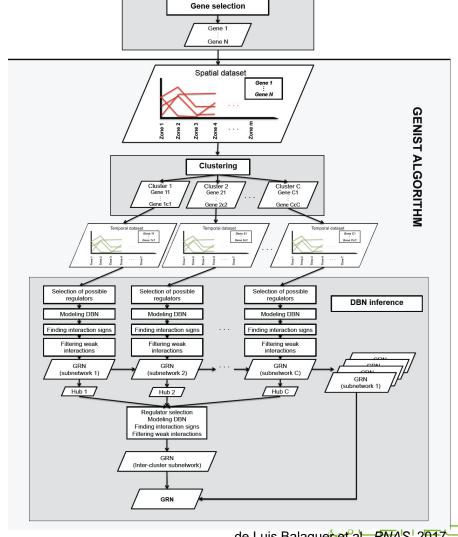
- Inferring GRNs via GENIST
  - Clustering of genes
    - Expression similarity under various conditions/genetic backgrounds, time points, ...
  - Inferring intra-cluster connections
    - Selection of potential regulators and co-regulators
      - Based on the time correlation in the change of expression and/or user specification
    - Dynamic Bayesian Network modeling

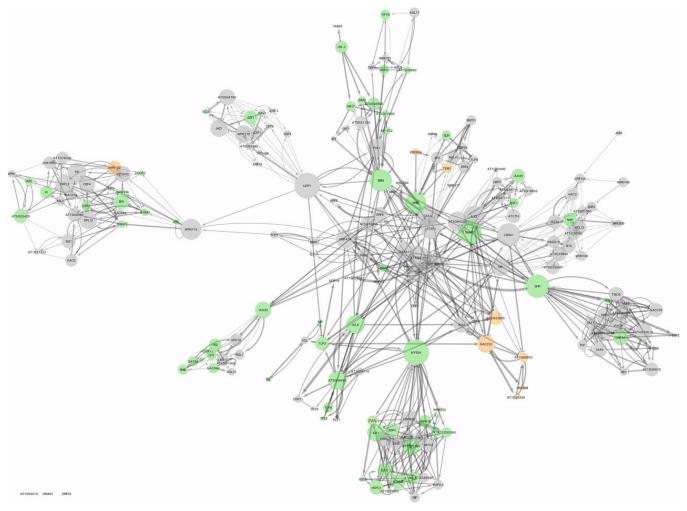


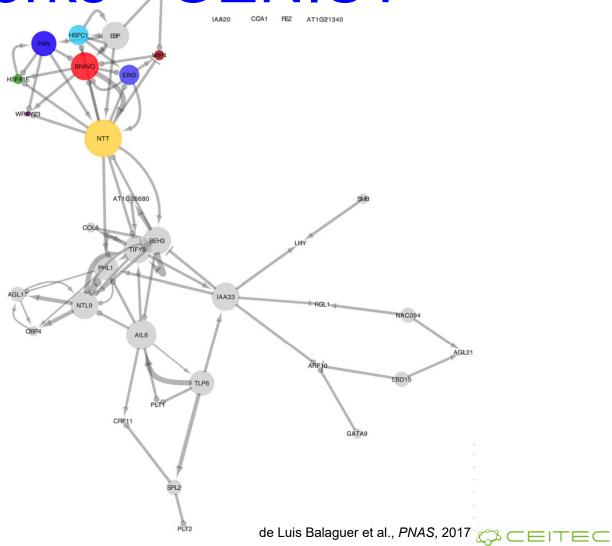
Haeseleer, Computational Biology, 2005



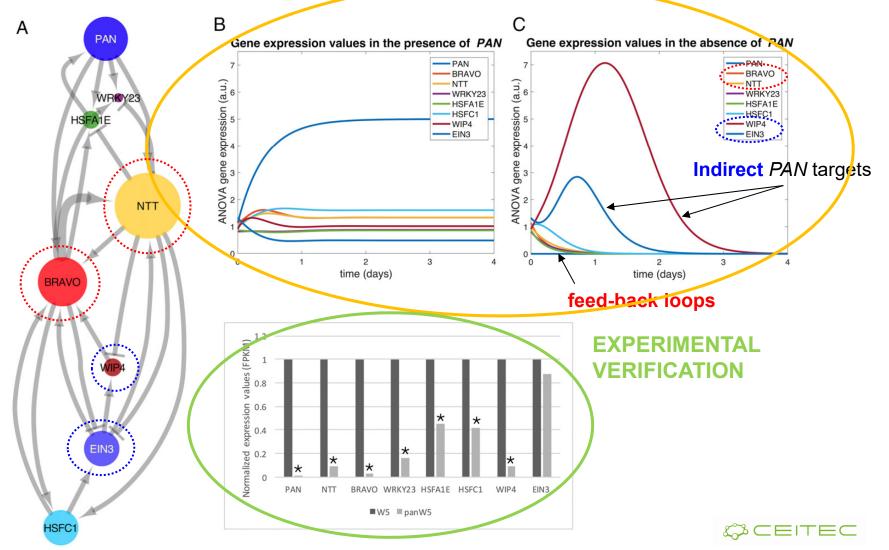
- Inferring GRNs via **GENIST** 
  - **Clustering of genes** 
    - Expression similarity under various conditions/genetic backgrounds, time points, ...
  - Inferring intra-cluster connections
    - Selection of potential regulators and COregulators
      - Based the on time correlation in the change of expression and/or user specification
    - **Dynamic Bayesian Network modeling**







# Gene Regulatory Networks - GENIST MODEL PREDICTION



#### **Key Concepts**

- Systems biology aims to identify new properties/behavior of groups of functional subunits (regulators, molecules) that are not simple addition of individual subunits, but represent a new feature dependent on the way of their mutual interaction
- It uses mathematical models, often Bayesian networks
- Gene regulatory networks can be identified also with the help of (semi)automated tools using large datasets (e.g. genome-wide expression profiles)
- Machine learning (AI) approaches are frequently used



#### Summary

- Definition of Systems Biology
- Tools
  - Gene Ontology analysis
  - Bayesian Networks
  - Molecular/Gene Regulatory Networks Modeling
  - Inferring Gene Regulatory Networks from Large Omics
     Datasets



#### Discussion

