

# CG920 Genomika

## Lekce 10

### Systemová biologie

Jan Hejátko

**Funkční genomika a proteomika rostlin,**  
Mendlovo centrum genomiky a proteomiky rostlin,  
Středoevropský technologický institut (CEITEC), Masarykova univerzita, Brno  
[hejatko@sci.muni.cz](mailto:hejatko@sci.muni.cz), [www.ceitec.muni.cz](http://www.ceitec.muni.cz)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Literatura

## ▪ Literární zdroje ke kapitole 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.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Osnova

- Definice **systemové biologie**
- **Nástroje**
  - Analýza **genové ontologie**
  - **Bayesovské sítě**
  - Modelování **molekulárních/genových regulačních sítí**
  - Odvození **genových regulačních sítí z velkých -omických dat**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

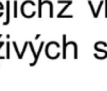
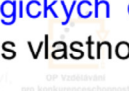
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Definice

**Systémová biologie** je vědecký směr v biologii využívající přístupy dalších věd, především biochemie, chemie, informatiky a matematiky. Zabývá se **studiem biologických funkcí a mechanismů** vyskytujících se v biologických systémech jako důsledek **komplexních interakcí**.

Základní myšlenkou je **komplexní pohled**, opak redukcionismu (který je převládajícím paradigmatem například v molekulární biologii), tedy předpoklad, že **systém je víc než pouhý součet jeho částí**.

Systémová biologie často **pracuje s modely**, které jsou vytvářeny **matematickými a informatickými přístupy** na **základě biologických dat**, jejichž vlastnosti jsou posléze porovnávány s vlastnostmi živých systémů (**Wikipedia**).



# Definice

Systémová biologie se zabývá studiem biologických systémů, jejichž chování nelze redukovat na lineární součet funkcí jejich částí. Systémová biologie nemusí nutně zahrnovat velké množství komponent nebo rozsáhlých datových souborů, jako je tomu v genomice nebo konektomice, ale často vyžaduje metody kvantitativního modelování původně vyvinuté fyziky (Nature).



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Definice

Názorně vysvětluje na videu Dr. Nathan Price,  
zástupce ředitele Ústavu pro systémovou biologii na  
[https://www.youtube.com/watch?v=OrXRI\\_8UFHU](https://www.youtube.com/watch?v=OrXRI_8UFHU).



ZVOJE VZDĚLÁVÁNÍ  
mentace je spolufinancována  
evropským sociálním fondem  
rozpočtem České republiky

# Osnova

- Definice systémové biologie
- Nástroje
  - Analýza genové ontologie



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Výsledky –omických studií vs. biologicky relevantní závěry

- Výsledky –omických studií reprezentují **enormní množství dat**, např. geny s rozdílnou expresí. Ale jak z nich získat **biologicky relevantní závěry**?

Ddil et al., unpublished

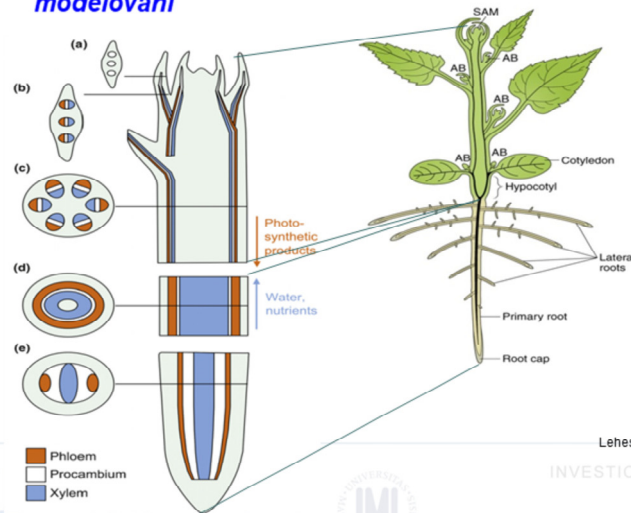
gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1.18041.79769e+308		1.73769e+308	6.88855e-05	0.00039180	1 yes
HRS1	1:4566891-4568708	WT	MT	OK	0	0.6966831.79769e+308		1.73769e+308	6.61994e-06	4.67708e-05	yes
ATMLO14	1:9227472-9232296	WT	MT	OK	0	0.5148091.79769e+308		1.73769e+308	9.74219e-05	0.00053505	yes
NRT1.6	1:9400883-9403789	WT	MT	OK	0	0.8778861.79769e+308		1.73769e+308	3.2892e-08	3.50131e-07	yes
AT1G27070	1:9075425-9082370	WT	MT	OK	0	2.06291.79769e+308		1.73769e+308	9.70039e-00	0.047e-05	yes
AT1G60095	1:22159735-22102419	WT	MT	OK	0	0.6885881.79769e+308		1.73769e+308	9.94992e-08	9.94992e-08	yes
AT1G03020	1:696200-998515	WT	MT	OK	0	1.788591.79769e+308		1.73769e+308	0.00913915	0.0277958	yes
AT1G13009	1:4062720-4063471	WT	MT	OK	0	3.558141.79769e+308		1.73769e+308	0.00021683	0.00108079	yes
AT1G21550	1:7553100-7553876	WT	MT	OK	0	0.5829881.79769e+308		1.73769e+308	0.00115582	0.00471497	yes
AT1G22120	1:7808308-7809632	WT	MT	OK	0	0.6173541.79769e+308		1.73769e+308	2.48392e-06	1.91089e-05	yes
AT1G31370	1:11238297-11239393	WT	MT	OK	0	1.462541.79769e+308		1.73769e+308	4.83523e-05	0.00028514	yes
APUM10	1:13253387-13255570	WT	MT	OK	0	0.5810311.79769e+308		1.73769e+308	7.87855e-06	5.48803e-05	yes
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0.5565251.79769e+308		1.73769e+308	6.53917e-05	0.00037473	yes
AT1G59077	1:21748208-21833195	WT	MT	OK	0	138.8881.79769e+308		1.73769e+308	0.00122789	0.00498810	yes
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0.3700871.79769e+308		1.73769e+308	0.00117953	0.0048001	yes
AT4G15242	4:8705786-8708987	WT	MT	OK	0.00930712	17.9056	10.9098	-4.405231.05673e-05	7.13983e-05	yes	
AT5G33251	5:12496071-12500433	WT	MT	OK	0.0496375	52.2837	10.0349	-9.8119	0	0 yes	
AT4G12820	4:7421055-7421738	WT	MT	OK	0.0195111	16.6516	8.88812	-3.900439.80217e-05	0.000528904	yes	
AT1G60020	1:22100851-22105276	WT	MT	OK	0.0116377	7.16623	9.24811	-7.503328.19504e-14	1.4088e-12	yes	
AT5G15380	5:4987235-4989182	WT	MT	OK	0.0988273	56.4834	9.1587	-10.4392	0	0 yes	

Example of an output of transcriptional profiling study using Illumina sequencing performed in our lab. Shown is just a tiny fragment of the complete list, comprising about 7K genes revealing differential expression in the studied mutant.



# Vývoj rostlinných vodivých pletiv

- **Vodivé pletivo** jako vývojový model pro **GO analýzu** a **MRN modelování**



Lehesranta et al., *Trends in Plant Sci* (2010)



MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY  
OP Vzdělávání pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

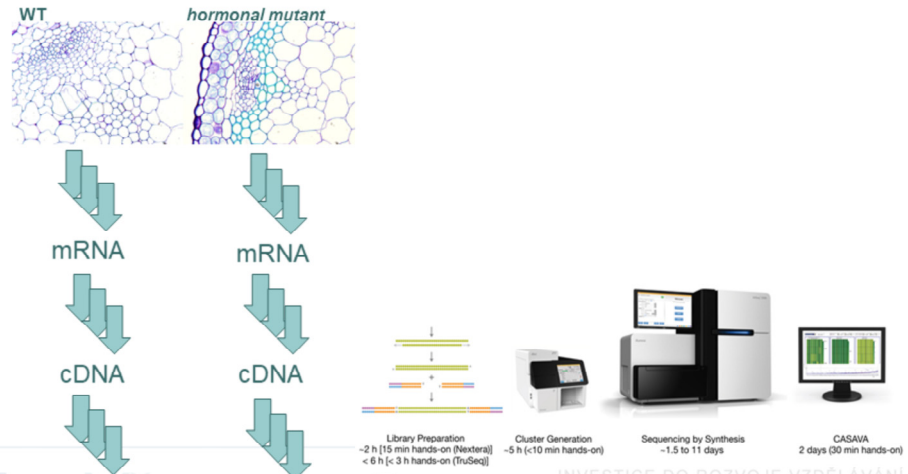
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

Molecular Regulatory Networks (MRN)



# Hormonální regulace vývoje rostlinných vodivých pletiv

- **Transkripční profilování pomocí sekvenování RNA**



Sekvenování společností Illumina a určení počtu transkriptů

EVROPSKÁ UNIE **est** Mládež a zaměstnanost pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

# Výsledky –omických studií vs. biologicky relevantní závěry

- Transkripční profilování identifikovalo víc než **9K odlišně regulovaných genů**...

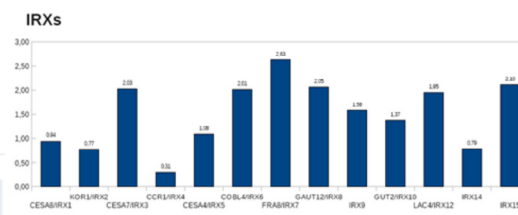
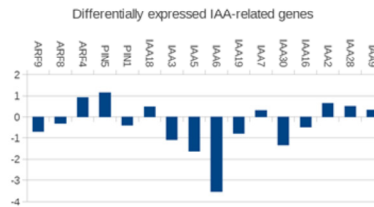
Ddil et al., unpublished

gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1.18041.79769e+308		1.73769e+308	6.88885e-05	0.00039180	1 yes
HRS1	1:4566891-4568708	WT	MT	OK	0	0.6966831.79769e+308		1.73769e+308	6.61994e-06	0.00053505	4 yes
ATMLO14	1:9227472-9232296	WT	MT	OK	0	0.5148091.79769e+308		1.73769e+308	9.74219e-05	0.00053505	6 yes
NRT1.6	1:9400883-9403789	WT	MT	OK	0	0.8778861.79769e+308		1.73769e+308	3.2892e-08	0.00053505	7 yes
AT1G27070	1:9075425-9082370	WT	MT	OK	0	2.06291.79769e+308		1.73769e+308	9.70039e-00	0.047e-05	yes
AT1G60095	1:22159735-22102419	WT	MT	OK	0	0.6885881.79769e+308		1.73769e+308	9.94992e-08	0.00053505	7 yes
AT1G03020	1:696200-998515	WT	MT	OK	0	1.788591.79769e+308		1.73769e+308	0.00913915	0.0277958	yes
AT1G13009	1:4062720-4063471	WT	MT	OK	0	3.556141.79769e+308		1.73769e+308	0.00021683	0.00108079	yes
AT1G21550	1:7553100-7553876	WT	MT	OK	0	0.5829881.79769e+308		1.73769e+308	0.00115582	0.00471497	yes
AT1G22120	1:7808308-7809632	WT	MT	OK	0	0.6173541.79769e+308		1.73769e+308	2.48392e-06	0.00053505	yes
AT1G31370	1:11238297-11239393	WT	MT	OK	0	1.462541.79769e+308		1.73769e+308	4.83523e-05	0.00028514	3 yes
APUM10	1:13253387-13255570	WT	MT	OK	0	0.5810311.79769e+308		1.73769e+308	7.87855e-06	0.00053505	yes
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0.5565251.79769e+308		1.73769e+308	6.53917e-05	0.00037473	6 yes
AT1G59077	1:21748208-21833195	WT	MT	OK	0	138.8881.79769e+308		1.73769e+308	0.00122789	0.00498810	yes
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0.3700871.79769e+308		1.73769e+308	0.00117953	0.0048001	yes
AT4G15242	4:8705786-8706997	WT	MT	OK	0.00930712	17.9056	10.9098	-4.40523	1.05673e-05	7.13983e-05	yes
AT5G33251	5:12496071-12500433	WT	MT	OK	0.0496375	52.2837	10.0349	-9.8119	0	0	yes
AT4G12820	4:7421055-7421738	WT	MT	OK	0.0195111	16.6516	8.88812	-8.90043	9.80217e-05	0.000528904	yes
AT1G60020	1:22100851-22105276	WT	MT	OK	0.0116377	7.16623	9.24811	-7.503028	1.0504e-14	1.4088e-12	yes
AT5G15380	5:4987235-4989182	WT	MT	OK	0.0988273	56.4834	9.1587	-10.4392	0	0	yes

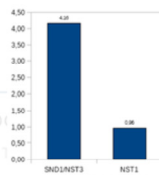
Example of an output of transcriptional profiling study using Illumina sequencing performed in our lab. Shown is just a tiny fragment of the complete list, comprising about 7K genes revealing differential expression in the studied mutant.

# Analýza genové ontologie

- Jedním z možných přístupů je studium **genové ontologie**, tj. **klasifikace genů** podle kontrolovaných **biologických procesů**



XYLEM MARKERS



Ddii et al., unpublished

OP Vzdělávání pro konkurenceschopnost

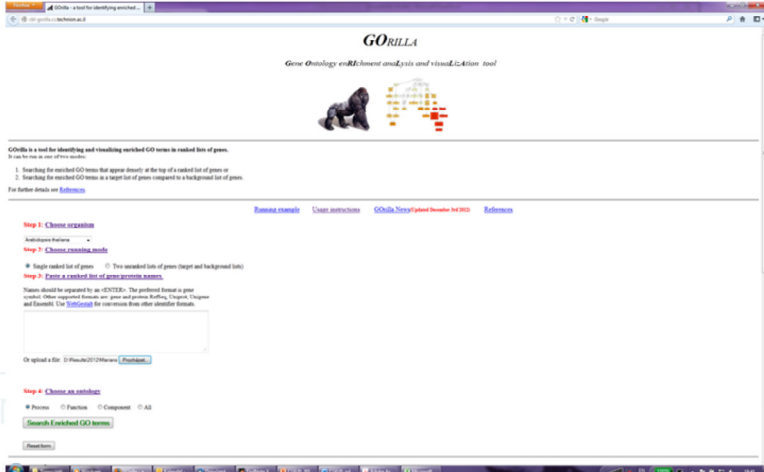


INVESTICE DO

OVÁNÍ  
ancována  
n fondem  
a státním rozpočtem České republiky

# Analýza genové ontologie

- Nástroje umožňující **statisticky vyhodnotit obohacení seznamu diferenciálně regulovaných genů o geny asociované se specifickými procesy**

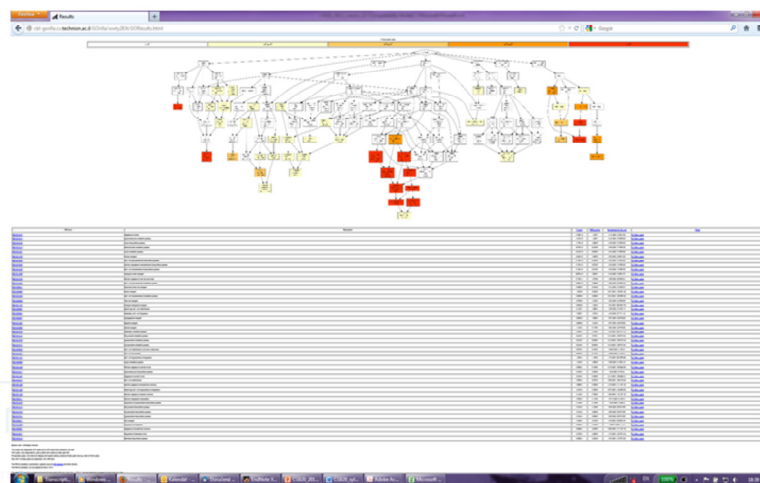


The screenshot shows the GORILLA web interface. At the top, it says "GORILLA Gene Ontology enrichment analysis and visualization tool". Below this, there is a small graphic of a gorilla and a network diagram. The main content area contains instructions for using the tool, such as "GORILLA is a tool for identifying and visualizing enriched GO terms in ranked lists of genes." and "It can be run in one of two modes: 1. Searching for enriched GO terms that appear directly at the top of a ranked list of genes or 2. Searching for enriched GO terms in a target list of genes compared to a background list of genes." There are also navigation links like "Home", "Users", "FAQ", "References", and "Help". At the bottom, there are logos for the European Union and ESF, and a citation: "Eden et al., BMC Bioinformatics (2009)".

One of such recent and very useful tools is Gorilla software, freely available at <http://cbl-gorilla.cs.technion.ac.il/>.

# Analýza genové ontologie

- Nástroje umožňující **statisticky vyhodnotit obohacení** seznamu diferenciálně regulovaných genů o geny **asociované se specifickými procesy**



MINISTERSTVO  
SKOLY

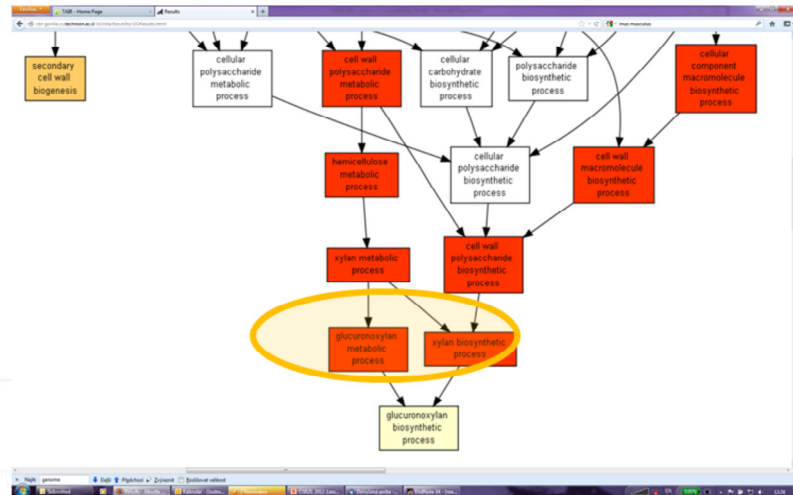
AVÁNÍ  
rokována  
fondem  
publiky





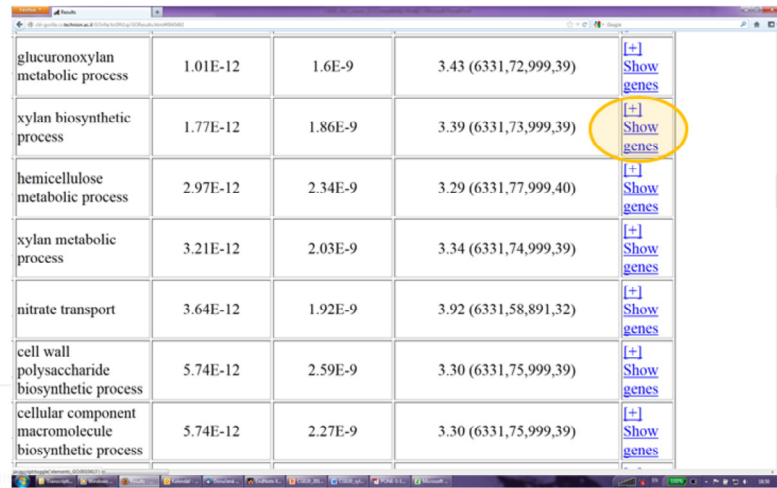
# Analýza genové ontologie

- Nástroje umožňující **statisticky vyhodnotit obohacení** seznamu diferenciálně regulovaných genů o geny **asociované se specifickými procesy**



# Analýza genové ontologie

- Nástroje umožňující **statisticky vyhodnotit obohacení** seznamu diferenciálně regulovaných genů o geny **asociované se specifickými procesy**



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

# Analýza genové ontologie

- Nástroje umožňující **statisticky vyhodnotit obohacení** seznamu diferenciálně regulovaných genů o **geny asociované se specifickými procesy**

Description	P-value	EDR q-value	Enrichment (S, B, n, b)	Genes
response to nitrate	4.76E-13	1.5E-9	4.13 (6331,55,891,32)	[+] Show genes
glucuronoylan metabolic process	1.01E-12	1.6E-9	3.43 (6331,72,999,39)	[+] Show genes
xylian biosynthetic process	1.77E-12	1.86E-9	3.39 (6331,73,999,39)	

The screenshot also shows a list of genes associated with the highlighted processes, including GO terms like GO:0009765, GO:0009766, and GO:0009767, along with their corresponding gene names and accession numbers.



MINISTERSTVO  
MLÁDEŽI  
A TĚLOVÝCHOVY

# Osnova

- Definice systémové biologie
- Nástroje
  - Analýza genové ontologie
  - Bayesovské sítě

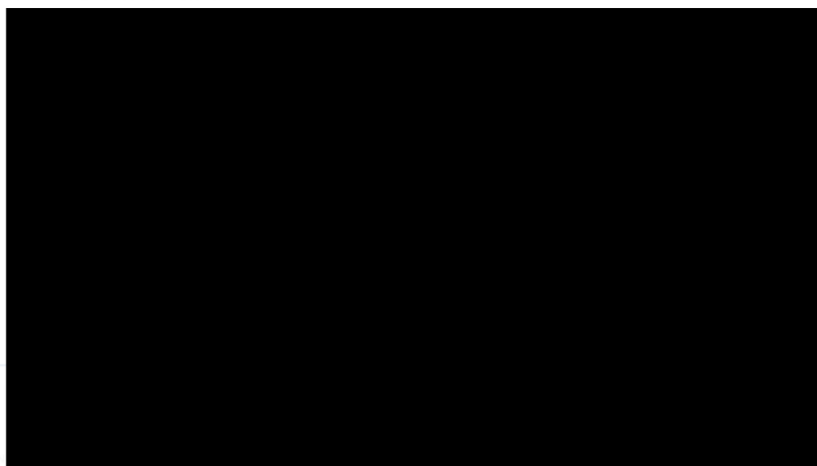


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Bayesovské sítě

- Co je Bayesovská síť?
  - **Pravděpodobnostní grafický model**, který se používá k **vytváření modelů z dat a/nebo názorů odborníků**



<https://www.youtube.com/watch?v=4fcqzVJwHM>

# Bayesovské sítě

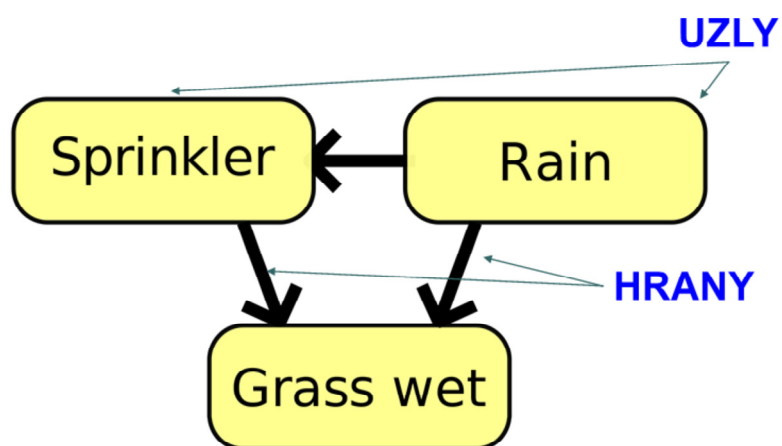
- Co je **Bayesovská síť**?
  - **Pravděpodobnostní grafický model**, který se používá k **vytváření modelů z dat a/nebo názoru odborníka**
  - může být využit v široké škále úkolů včetně **predikcí, detekcí anomálií, diagnostiky, automatického objasňování, zdůvodňování, predikce vývoje v čase a rozhodování za nejistoty**
- **UZLY**
  - každý uzel představuje **proměnnou**, jako je výška, věk nebo pohlaví. Proměnná může být diskrétní, jako například Pohlaví = {samíči, samčí}, nebo spojitá, jako např. věk
- **HRANY**
  - spojnice popisující **závislosti mezi uzly**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

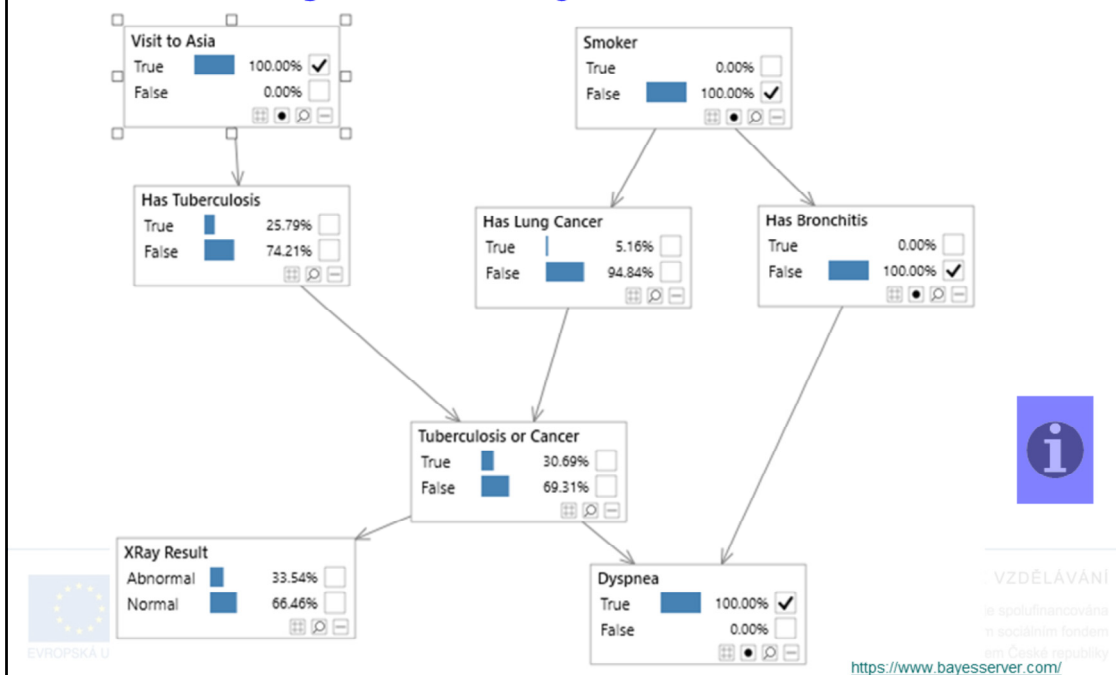
# Bayesovské síť



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Asijská Bayesovská síť





# Osnova

- Definice systémové biologie
- Nástroje
  - Analýza genové ontologie
  - Bayesovské sítě
  - Modelování molekulárních/genových regulačních sítí

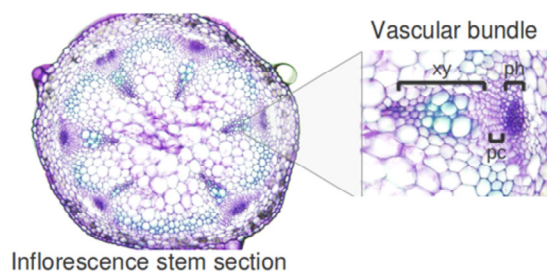


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Modelování molekulárních regulačních sítí

- **Vodivé pletivo** jako vývojový model pro **Molecular Regulatory Network (MRN) modelování**



Benítez and Hejatko, *PLoS One*, 2013



ZDĚLÁVÁNÍ  
polufinancováno  
sociálním fondem  
České republiky

# Modelování molekulárních regulačních sítí

## □ Vyhledávání publikovaných dat a vytvoření malé databáze

Interaction	Evidence	References
A-ARRs –  CK signaling	Double and higher order type-AARR 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 ARR6 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]
	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 AHPs to ARR6.	[9]



Benitez and Hejatko, *PLoS One*, 2013

OP Vzdělávání pro konkurenceschopnost



EVROPSKÁ UNIE

DĚLÁVÁNÍ

financována

Evropským sociálním fondem a státním rozpočtem České republiky

# Modelování molekulárních regulačních sítí

- Formulování **logických pravidel** definujících **dynamiku modelu**

Network node	Dynamical rule
CK	2 If ipt=1 and cxx=0 1 If ipt=1 and cxx=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

Benitez and Hejatkó, *PLoS One*, 2013

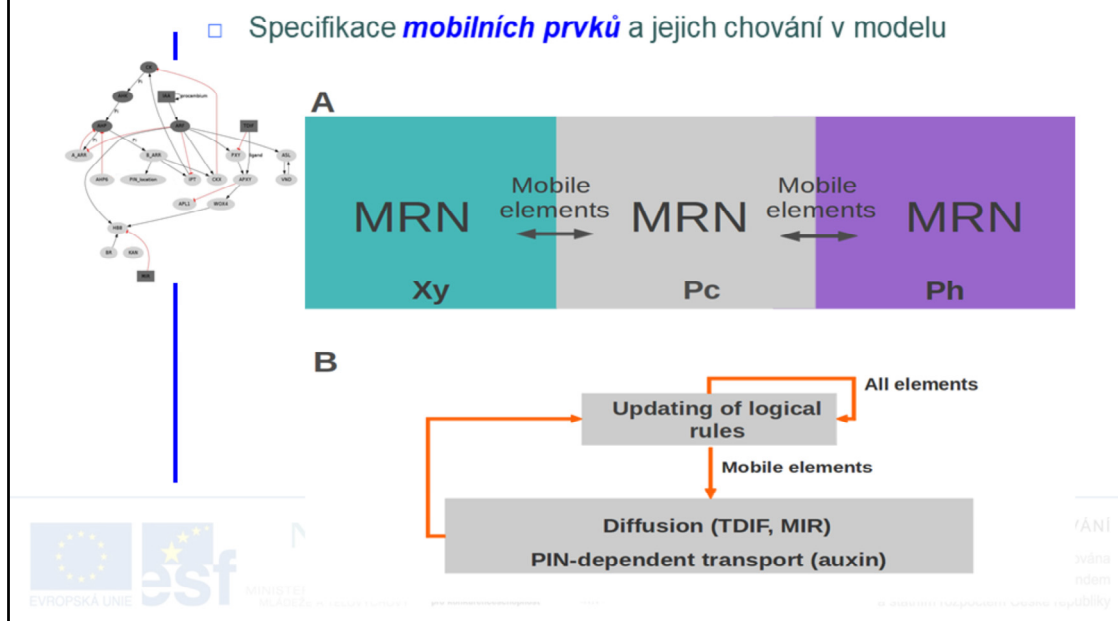


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Modelování molekulárních regulačních sítí

- Specifikace *mobilních prvků* a jejich chování v modelu



According to experimental evidence for the system under study, the hormone IAA, the peptide TDIF, and the microRNA MIR165/6 are able to move among the cells. In the case of TDIF and MIR165/6, the mobility is defined as diffusion and is given by the following equation:

$$g(t+1)T[i] = H(g(t)[i] + D(g(t)[i+1] + g(t)[i-1] - N(g(t)[i]) - b)) \quad (2),$$

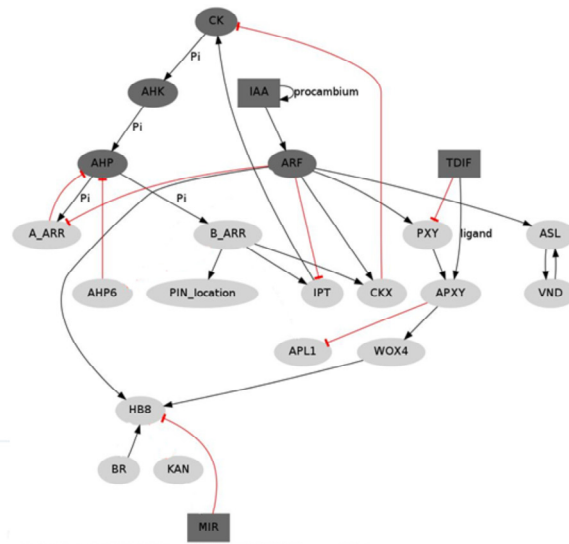
where  $g(t)T[i]$  is the total amount of TDIF or MIR165 in cell ( $i$ ).  $D$  is a parameter that determines the proportion of  $g$  that can move from any cell to neighboring ones and is correlated to the diffusion rate of  $g$ .  $b$  is a constant corresponding to a degradation term.  $H$  is a step function that converts the continuous values of  $g$  into a discrete variable that may attain values of 0, 1 or 2.  $N$  stands for the number of neighbors in each cell. Boundary conditions are zero-flux. In the case of IAA, the mobility is defined as active transport dependent on the radial localization of the PIN efflux transporters and is defined by the equation:

$$iaa(t+1)T[i] = H(iaa(t)[i] + Diaa(pin(t)[i+1])(iaa(t)[i+1]) + Diaa(pin(t)[i-1])(iaa(t)[i-1]) - N(Diaa)(pin(t)[i])(iaa(t)[i]) - b) \quad (3),$$

where  $Diaa$  is a parameter that determines the proportion of IAA that can be transported among cells. The transport depends on the presence of IAA and PIN in the cells and  $b$  corresponds to a degradation term. As in equation 2,  $H$  is a step function that converts the continuous values to discrete ones and  $N$  stands for the number of neighbors in each cell. Boundary conditions for IAA motion are also zero-flux.

# Modelování molekulárních regulačních sítí

- Příprava *první verze* modelu a její *testování*



TICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

The proposed model considers data that we identified and evaluated through an extensive search (up to January 2012). It takes into account molecular interactions, hormonal and expression patterns, and cell-to-cell communication processes that have been reported to affect vascular patterning in the bundles of Arabidopsis. The model components and interactions are graphically presented in the figure above. In the network model, nodes stand for molecular elements regulating one another's activities. Most of the nodes can take only 1 or 0 values (light gray nodes in the figure), corresponding to "present" or "not present," respectively. Since the formation of gradients of hormones and diffusible elements may have important consequences in pattern formation, mobile elements TDIF and MIR, as well as members of the CK and IAA signaling systems, can take 0, 1 or 2 values (dark gray nodes in the figure above) Benitez and Hejatko, submitted.

# Modelování molekulárních regulačních sítí

## □ Specifikace chybějících interakcí ze *známých predikcí*

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 <i>Arabidopsis thaliana</i> , 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]
CK → APL	In the root, CK signaling is required for the CK regulation of PIN1, PIN3, and PIN7. Their expression is altered in <i>wol</i> , <i>cre1</i> , <i>ahk3</i> and <i>ahp6</i> mutants.	[19]
	Predicted interaction (could be direct or indirect)	
	Consistent with the fact that APL overexpression prevents or delays xylem cell differentiation, as does CKs.	[21]
	Partially supported by microarray data and phloem-specific expression patterns of CK response factors.	(TAIR, ExpressionSet1 005823559, [22])



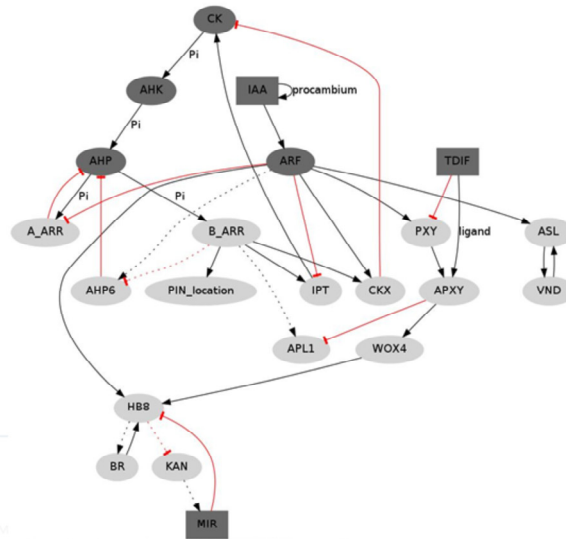
EVROPSKÁ UNIE

MŠMT

EVROPSKÁ UNIE

# Modelování molekulárních regulačních sítí

- Příprava **další verze** modelu a její **testování**



Benitez and Hejatko, *PLoS One*, 2013

OPROJEKT ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

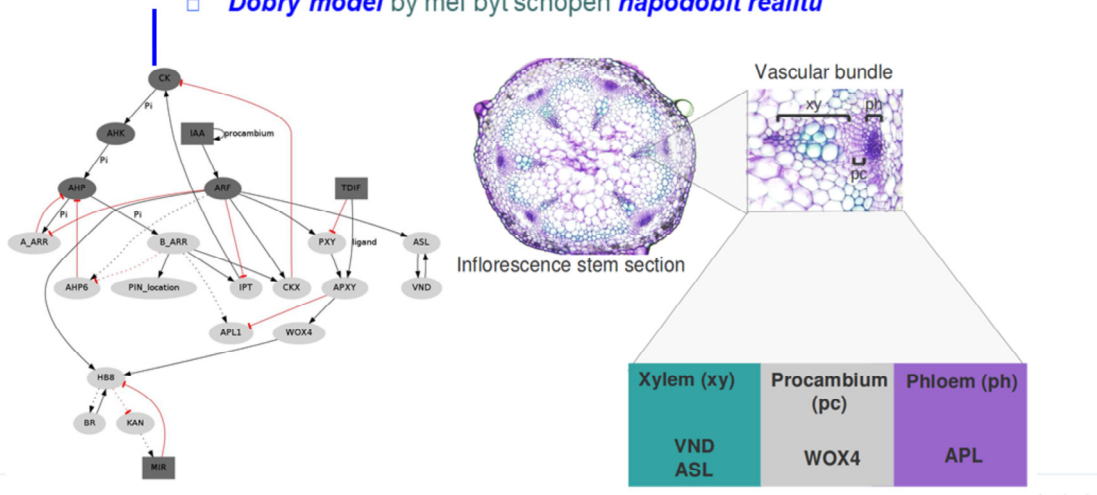


In comparison to the model shown on slide 21, the final version of the model contains the predicted interactions (dashed lines).



# Modelování molekulárních regulačních sítí

- *Dobry model* by měl být schopen *napodobit realitu*

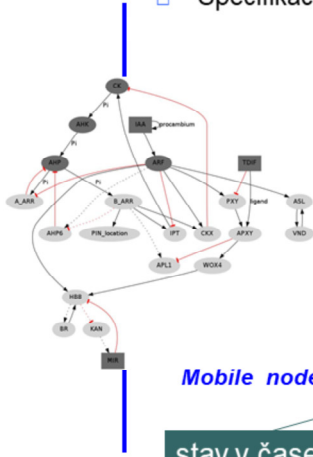


Benitez and Hejatkó, *PLoS One*, 2013

Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

# Modelování molekulárních regulačních sítí

- Specifikace *rovníc* popisujících *vzájemné vztahy* v modelu



funkce popisující logická pravidla

stav v čase  $t$

$$\text{Static nodes: } g_n(t+1) = F_n(g_{n1}(t), g_{n2}(t), \dots, g_{nk}(t))$$

stav v čase  $t+1$

množství TDIF nebo MIR165 v buňce  $i$

$$\text{Mobile nodes: } g_{(t+1)T[i]} = H(g_{(t)[i]} + D(g_{(t)[i+1]} + g_{(t)[i-1]} - N(g_{(t)[i]})) - b)$$

stav v čase  $t+1$

konstanta popisující dobu degradace

podíl mobilních elementů



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

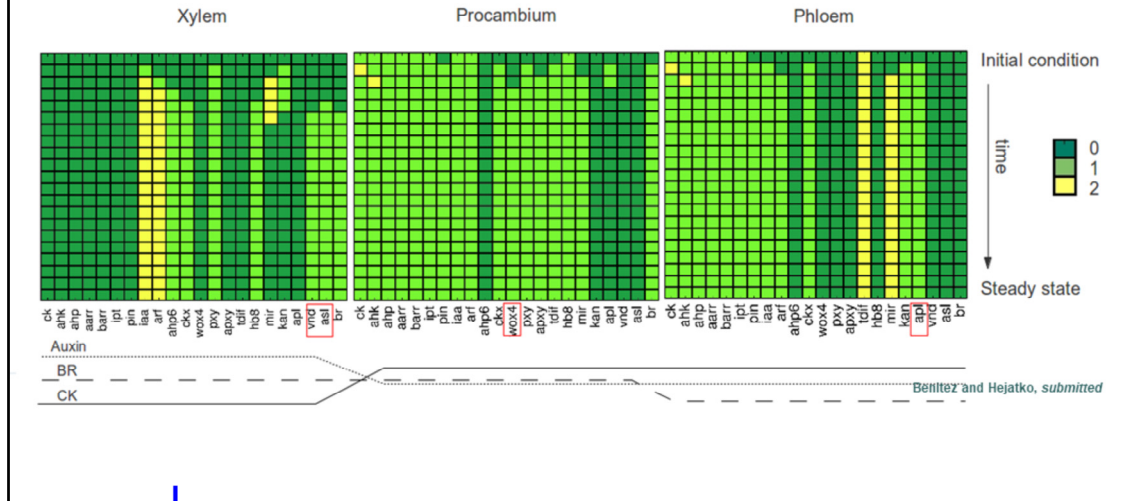
Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Modelování molekulárních regulačních sítí

- **Dobrý model** by měl být schopen **simulovat realitu**

**Statické uzly:**  $g_n(t+1) = F_n(g_{n1}(t), g_{n2}(t), \dots, g_{nk}(t))$

**Mobilní uzly:**  $g_{(t+1)T[i]} = H(g_{(t)[i]} + D(g_{(t)[i+1]} + g_{(t)[i-1]} - N(g_{(t)[i]})) - b)$



The initial conditions specify the initial state of some of the network elements (figure above) and are the following :

I) In the procambial position (central compartment), CK is initially available and there is an initial and sustained IAA input or self-upregulation. This condition is supported by several lines of evidence. Also *HBB*, a marker of early vascular development that has been found in preprocambial cells, is assumed to be initially present at this position. These conditions are not fixed, however. After the initial configuration, all the members of the CK and IAA signaling pathways, as well as *HBB*, can change their states according to the logical rules.

II) In the xylem and phloem positions, it is assumed that no element is initially active except for the CK signaling pathway and TDIF, both in the phloem position. The level of expression for a given node is represented by a discrete variable  $g$  and its value at a time  $t+1$  depends on the state of other components of the network ( $g_1, g_2, \dots, g_N$ ) at a previous time unit. The state of every gene  $g$  therefore changes according to:

$$g_n(t+1) = F_n(g_{n1}(t), g_{n2}(t), \dots, g_{nk}(t)) \quad (1)$$

In this equation,  $g_{n1}, g_{n2}, \dots, g_{nk}$  are the regulators of gene  $g_n$  and  $F_n$  is a discrete function known as a logical rule (logical rules are grounded in available experimental data, for example see slide 20). Given the logical rules, it is possible to follow the dynamics of the network for any given initial configuration of the nodes expression state. One of the most important traits of dynamic models is the existence of steady states in which the entire network enters into a self-sustained configuration of the nodes state. It is thought that in developmental systems such self-sustained states correspond to particular cell types.

According to experimental evidence for the system under study, the hormone IAA, the peptide TDIF, and the microRNA MIR165/6 are able to move among the cells. In the case of TDIF and MIR165/6, the mobility is defined as diffusion and is given by the following equation:

$$g_{(t+1)T[i]} = H(g_{(t)[i]} + D(g_{(t)[i+1]} + g_{(t)[i-1]} - N(g_{(t)[i]})) - b) \quad (2)$$

where  $g_{(t)T[i]}$  is the total amount of TDIF or MIR165 in cell  $(i)$ .  $D$  is a parameter that determines the proportion of  $g$  that can move from any cell to neighboring ones and is correlated to the diffusion rate of  $g$ .  $b$  is a constant corresponding to a degradation term.  $H$  is a step function that converts the continuous values of  $g$  into a discrete variable that may attain values of 0, 1 or 2.  $N$  stands for the number of neighbors in each cell. Boundary conditions are zero-flux. In the case of IAA, the mobility is defined as active transport dependent on the radial localization of the PIN efflux transporters and is defined by the equation:

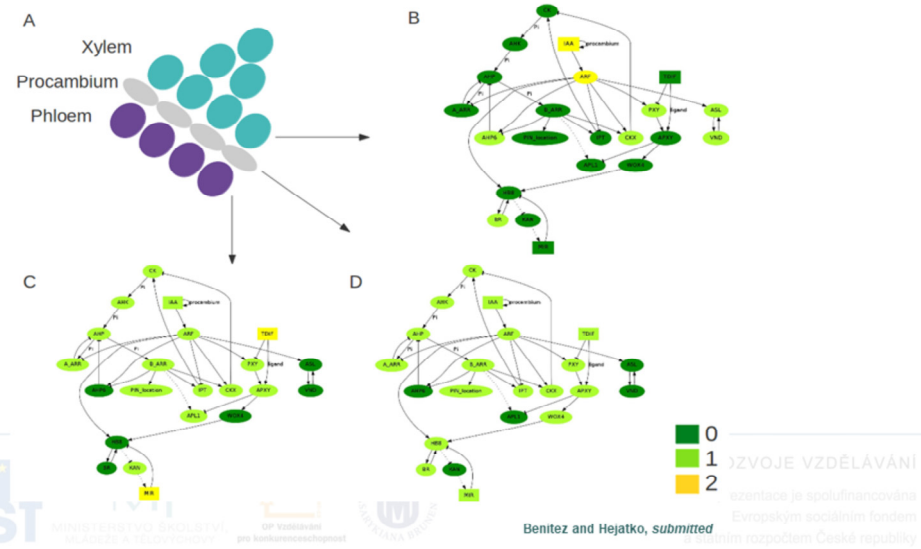
$$iaa_{(t+1)T[i]} = H(iaa_{(t)[i]} + Diaa(pin_{(t)[i+1]})(iaa_{(t)[i+1]}) + Diaa(pin_{(t)[i-1]})(iaa_{(t)[i-1]}) - N(Diaa)(pin_{(t)[i]})(iaa_{(t)[i]}) - b) \quad (3)$$

where  $Diaa$  is a parameter that determines the proportion of IAA that can be transported among cells. The transport depends on the presence of IAA and PIN in the cells and  $b$  corresponds to a degradation term. As in equation 2,  $H$  is a step function that converts the continuous values to discrete ones and  $N$  stands for the number of neighbors in each cell. Boundary conditions for IAA motion are also zero-flux.

Using the logical rules, equations 1–3, and a broad range of parameter values (not shown here), it is possible fully to reproduce the results and analyses reported in the following sections (see the figure above for the simulation time course).

# Modelování molekulárních regulačních sítí

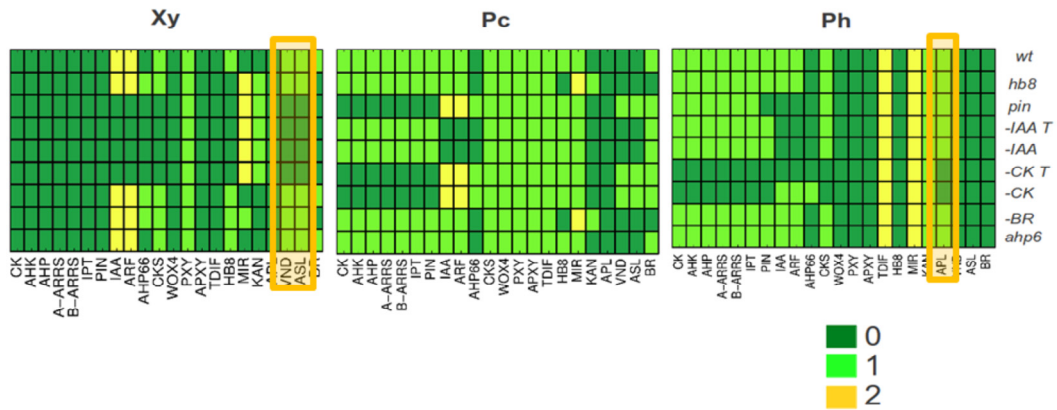
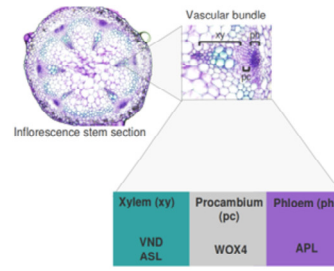
□ **Dobrý model** by měl být schopen *simulovat realitu*



Another representation of the distinct expression profiles in the individual vascular bundle compartments (phloem, procambium and xylem).

# Modelování molekulárních regulačních sítí

□ Simulace *mutantů*



# Osnova

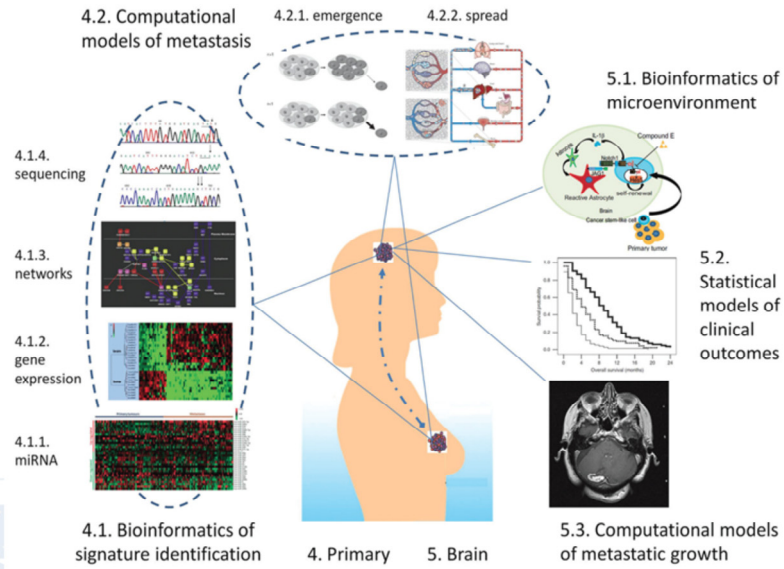
- Definice systémové biologie
- Nástroje
  - Analýza genové ontologie
  - Bayesovské sítě
  - Modelování molekulárních/genových regulačních sítí
- Odvození genových regulačních sítí z velkých -omických dat



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

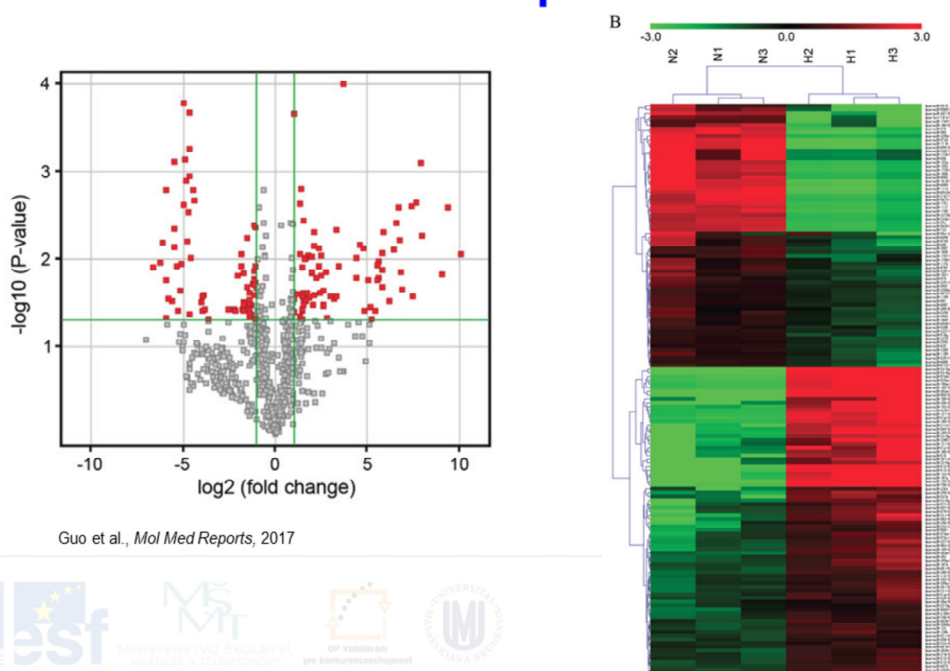
# Systemová biologie ve výzkumu rakoviny



EVROPSKÁ UNIE    es    Mládeží a inovativnosti    pro konkurenceschopnost    TMA 9\*

ZDĚLÁVÁNÍ  
spolufinancováno  
sociálním fondem  
a státním rozpočtem České republiky

# miRNA/mRNA profilování



Hypertrophic scars (HS) are a fibroproliferative disorder of the skin, which causes aesthetic and functional impairment. However, the molecular pathogenesis of this disease remains largely unknown and currently no efficient treatment exists. MicroRNAs (miRNAs) are involved in a variety of pathophysiological processes, however the role of miRNAs in HS development remains unclear. To investigate the miRNA expression signature of HS, microarray analysis was performed and 152 miRNAs were observed to be differentially expressed in HS tissue compared with normal skin tissues. Of the miRNAs identified, miRNA-21 (miR-21) was significantly increased in HS tissues and hypertrophic scar fibroblasts (HSFBs) as determined by reverse transcription-quantitative polymerase chain reaction analysis. It was also observed that, when miR-21 in HSFBs was blocked through use of an antagomir, the phenotype of fibrotic fibroblasts *in vitro* was reversed, as demonstrated by growth inhibition, induction of apoptosis and suppressed expression of fibrosis-associated genes collagen type I  $\alpha$  1 chain (COL1A1), COL1A2 and fibronectin. Furthermore, miR-21 antagomir administration significantly reduced the severity of HS formation and decreased collagen deposition in a rabbit ear HS model. The total scar area and scar elevation index were calculated and were demonstrated to be significantly decreased in the treatment group compared with control rabbits. These results indicated that the miR-21 antagomir has a therapeutic effect on HS and suggests that targeting miRNAs may be a successful and novel therapeutic strategy in the treatment of fibrotic diseases that are difficult to treat with existing methods.

miRNA expression signature profiling in hypertrophic scars (HS). (A) Volcano plot presenting differentially expressed miRNAs between HS and paired (non-scar, obtained from donor sites during scar resection) NS tissue. miRNA microarray expression profiling from three paired HS and NS tissues was performed. Differentially expressed miRNAs were identified by fold change and a P-value calculated using Student's *t*-test. The threshold set to identify up and downregulated genes was a fold change  $\geq 2$  and  $P < 0.05$ . Red dots indicate points-of-interest that exhibit large-magnitude fold-changes (x-axis;  $\log_2$  of the fold change) and high statistical significance (y-axis;  $-\log_{10}$  of the P-value). (B) Hierarchical clustering showing differentially expressed miRNAs from HS samples compared with paired NS tissues. Each row represents one miRNA and each column represents one tissue sample. The relative miRNA expression is depicted according to the color scale. Red indicates upregulation and green indicates downregulation. N1-3 represents NS tissue samples, whereas H1-3 represents HS tissue samples. The differentially expressed miRNAs were clearly separated into clusters. miRNA, microRNA; hsa-miR, human microRNA; HS, hypertrophic scar; NS, normal skin.



# Odvození genových regulačních sítí

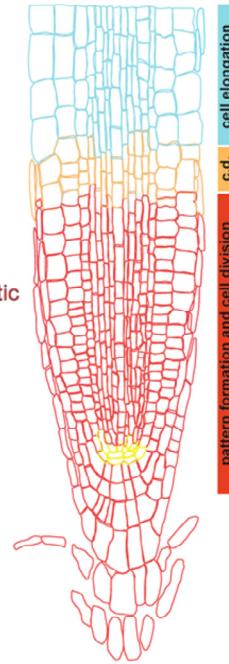


Elongation zone

Transient zone

Meristematic zone

Stem cell niche



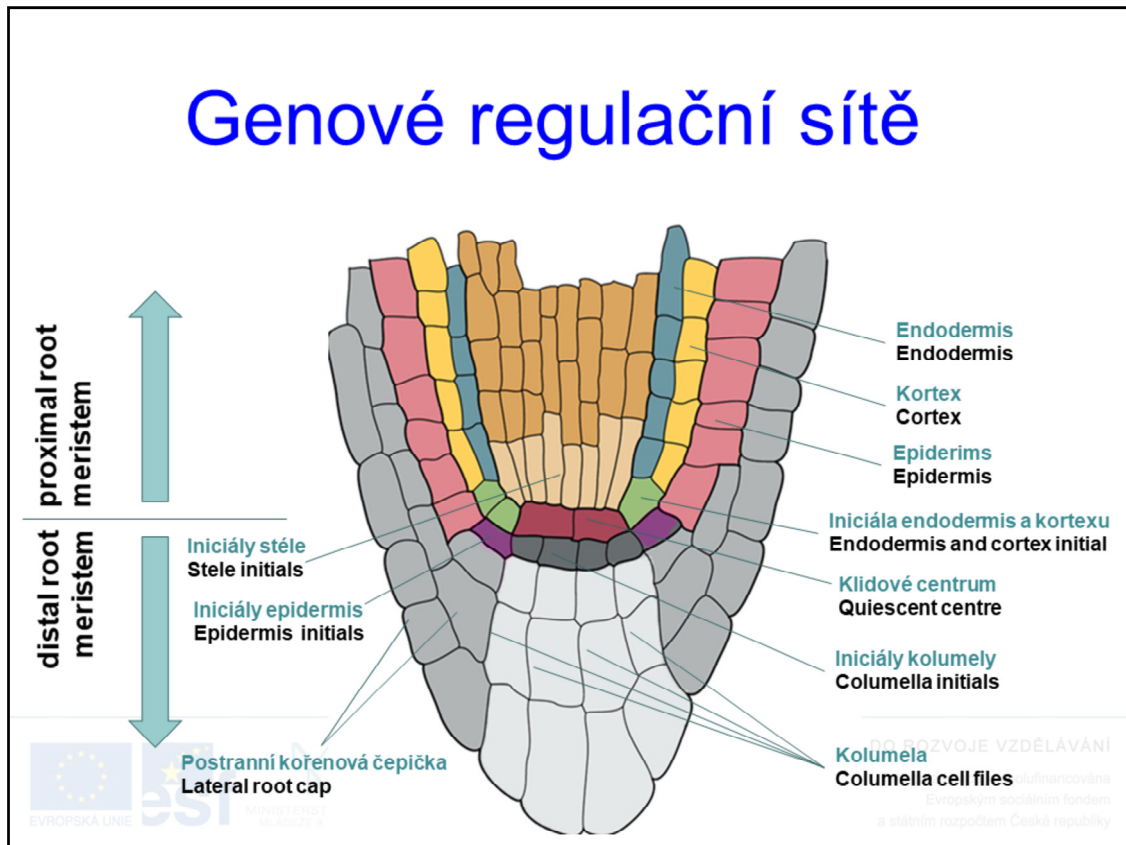
Benkova and Hejatkó, *Plant Mol Biol* (2008)



VZVOJE VZDĚLÁVÁNÍ

organizace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Genové regulační sítě

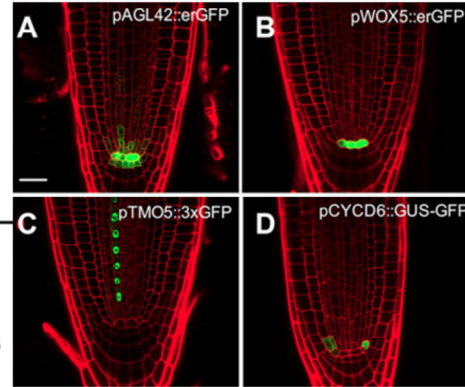


In the root, several functional and anatomical units could be recognized.

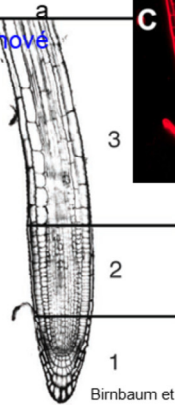
Along the longitudinal axis, the root meristem forms a distal root tip, including stem cell niche, columella and lateral root cap, proximal meristem with a population of rapidly dividing cells and elongation zone where cells leaving the root meristem undergo rapid elongation and mature.

# Genové regulační sítě - GENIST

- Odvození GR sítí přes **GENIST**
  - **GENE** regulatory **Network** Inference from **SpatioTemporal** data algorithm
  - Kombinace **prostorové** vývojové **specifity** **genové** exprese



de Luis Balaguer et al., *PNAS*, 2017



Birnbaum et al., *Science*, 2003

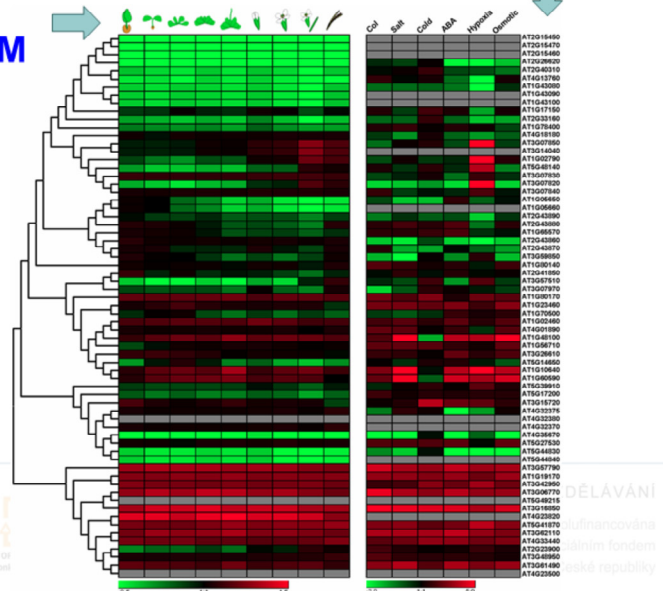
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Kombinace velkých omických datových sad

PLETIVO/  
VÝVOJOVÉ STADIUM

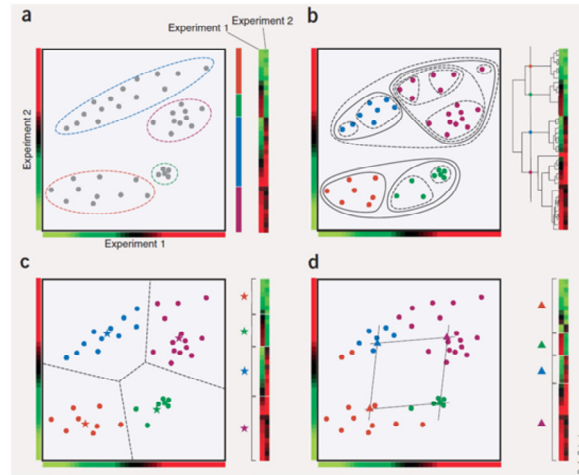
GENY



DĚLÁVÁNÍ  
ulínancována  
ním fondem  
ské republiky

# Genové regulační sítě - GENIST

- Odvození GR sítí přes **GENIST**
  - **shlukování (klastrování) genů**
    - Expresní podobnost za různých podmínek/genetické pozadí, časové body, ...
  - **Identifikace spojení uvnitř klastru**
    - **Selekce** potenciálních **regulátorů** a **ko-regulátorů**
      - Na základě **časové korelace** ve **změně exprese** a/nebo manuální **specifikace uživatelem**
  - **Modelování dynamické Bayesovské sítě**



Haeseleer, *Computational Biology*, 2005

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

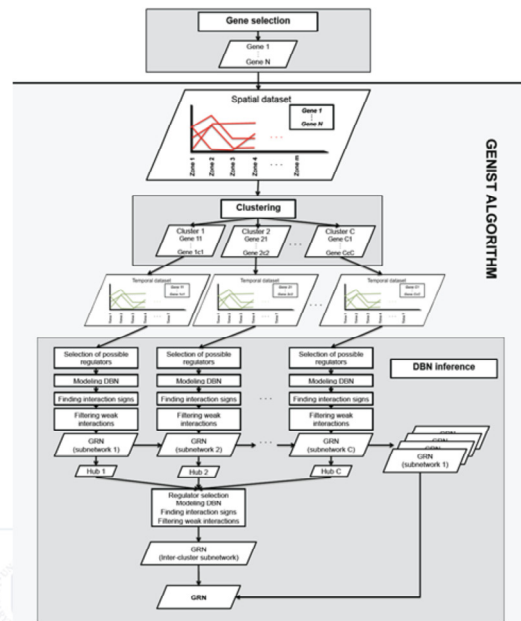
## *GENIST algorithm*

The MATLAB source code for GENIST is publically available at <https://github.com/madeluis/GENIST>.

For the detailed description of the procedure, see de Luis Balaguer et al., 2017, SI (<https://www.pnas.org/content/114/36/E7632/tab-figures-data>)

# Genové regulační sítě - GENIST

- Odvození GR sítí přes **GENIST**
  - **shlukování (klastrování) genů**
    - Expresní podobnost za různých podmínek/genetické pozadí, časové body, ...
  - **Identifikace spojení uvnitř klastru**
    - **Selekce** potenciálních **regulátorů** a **ko-regulátorů**
      - Na základě **časové korelace** ve změně **exprese** a/nebo manuální **specifikace** uživatelem
    - **Modelování dynamické Bayesovské sítě**



de Luis Balaguer et al., *PNAS*, 2017

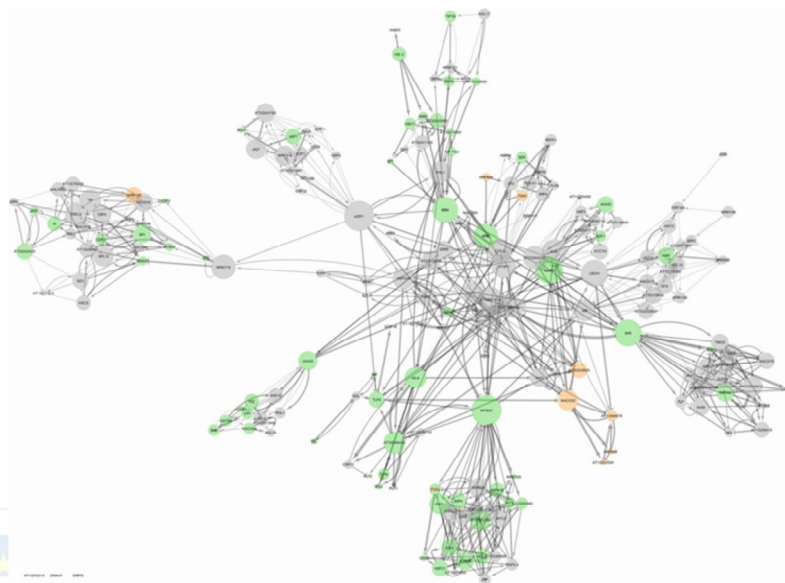
## GENIST algorithm

The MATLAB source code for GENIST is publically available at <https://github.com/madeluis/GENIST>.

For the detailed description of the procedure, see de Luis Balaguer et al., 2017, SI (<https://www.pnas.org/content/114/36/E7632/tab-figures-data>)

GENIST block diagram. GENIST is implemented in MATLAB, and is composed of two consecutive steps, clustering and GRN inference. Clustering is performed based on a spatial dataset. Each resulting cluster is independently processed by the GRN inference step, based on a temporal dataset.

# Genové regulační sítě - GENIST



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání  
pro konkurenceschopnost



OP Vzdělávání

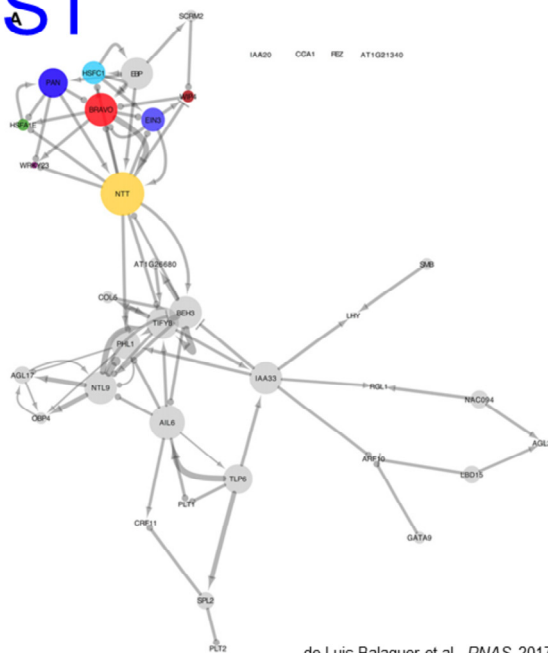
je spolufinancována

Evropským sociálním fondem

de Luis Balaguer et al., *PNAS*, 2017

čl. 173 odst. 1 písm. c) zák. č. 139/2004 Sb., o vysoké škole, ve znění pozdějších předpisů

# Genové regulační sítě - GENIST



MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

70% JE VZDĚLÁVÁNÍ

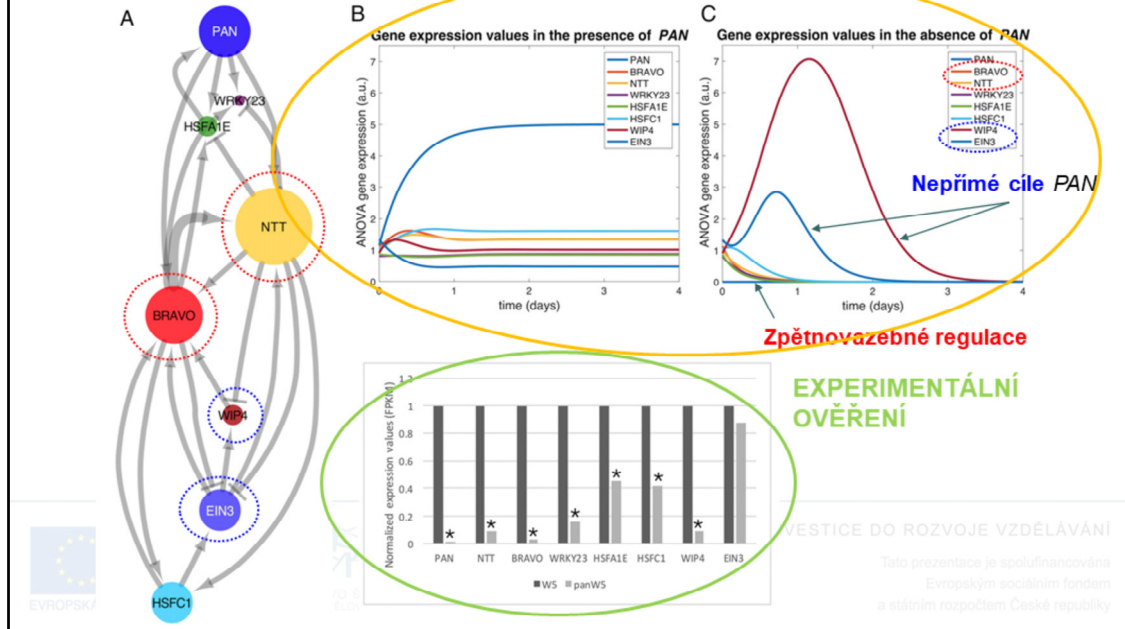
„níctace je spolufinancována  
evropským sociálním fondem  
z rozpočtu České republiky

de Luis Balaguer et al., *PNAS*, 2017



# Genové regulační sítě - GENIST

PŘEDPOVĚDÍ MODELU



PAN subnetwork in the QC inferred with the 12 developmental time points of the Arabidopsis root. (A) Optimal configuration (combination of signs— activation or repression—of the regulations that were inferred with undefined signs, which best fits the data in the simulations of the equations) of the subnetwork of PAN and its downstream targets. (B and C) Resulting expression values of PAN and its downstream targets, over time, after simulating the optimal configuration of the model. Simulations were run for 5 d and plots are shown until all factors reached steady states in the WT and pan mutant simulations. (B) Model simulated with the fitted equation parameters. (C) Model simulated with the PAN-associated parameters set to zero to simulate a pan mutant situation. (D) Normalized expression values of PAN and its predicted downstream targets in Col-0 wild type and in pan mutant. Statistically significant changes of expression between the mutant and the wild type, \* $q < 0.05$ .

In the WT simulation, all targets reached steady states by day 1 with subtle changes of expression during the transients (time length until expression values reach their steady states). On the contrary, the pan mutant simulation showed that EIN3 and WIP4 presented high expression values during the transients and reached steady states at later stages (days 3 and 4, respectively). These delayed responses and initial activations of EIN3 and WIP4 reflect the prediction that these genes are indirectly affected by PAN. Further, the dynamics of our simulations depict that BRAVO, NTT, and WIP4 are, in our equations, connected through feedback loops. During the transient phase of the mutant simulation, NTT and BRAVO show an exponential decay, which is consistent with the prediction that they activate each other in the absence of PAN. However, their steady states are not immediately reached since they are activated by WIP4 and EIN3. Conversely, WIP4, which is repressed by a decaying NTT, shows high levels of expression.

# Souhrn

- Definice **systemové biologie**
- **Nástroje**
  - Analýza **genové ontologie**
  - **Bayesovské sítě**
  - Modelování **molekulárních/genových regulačních sítí**
  - Odvození **genových regulačních sítí z velkých -omických dat**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Diskuse



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky