CG020 Genomika

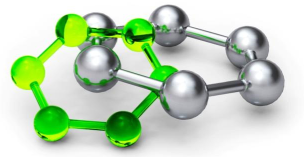
Přednáška 11

Systémová biologie

Jan Hejátko

Funkční genomika a proteomika rostlin,
Středoevropský technologický institut (CEITEC)
a
Národní centrum pro výzkum biomolekul,
Přírodovědecká fakulta,

Masarykova univerzita, Brno
hejatko@sci.muni.cz, www.ceitec.eu



Literatura

- Literární zdroje ke kapitole 11:
  - 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.

Osnova

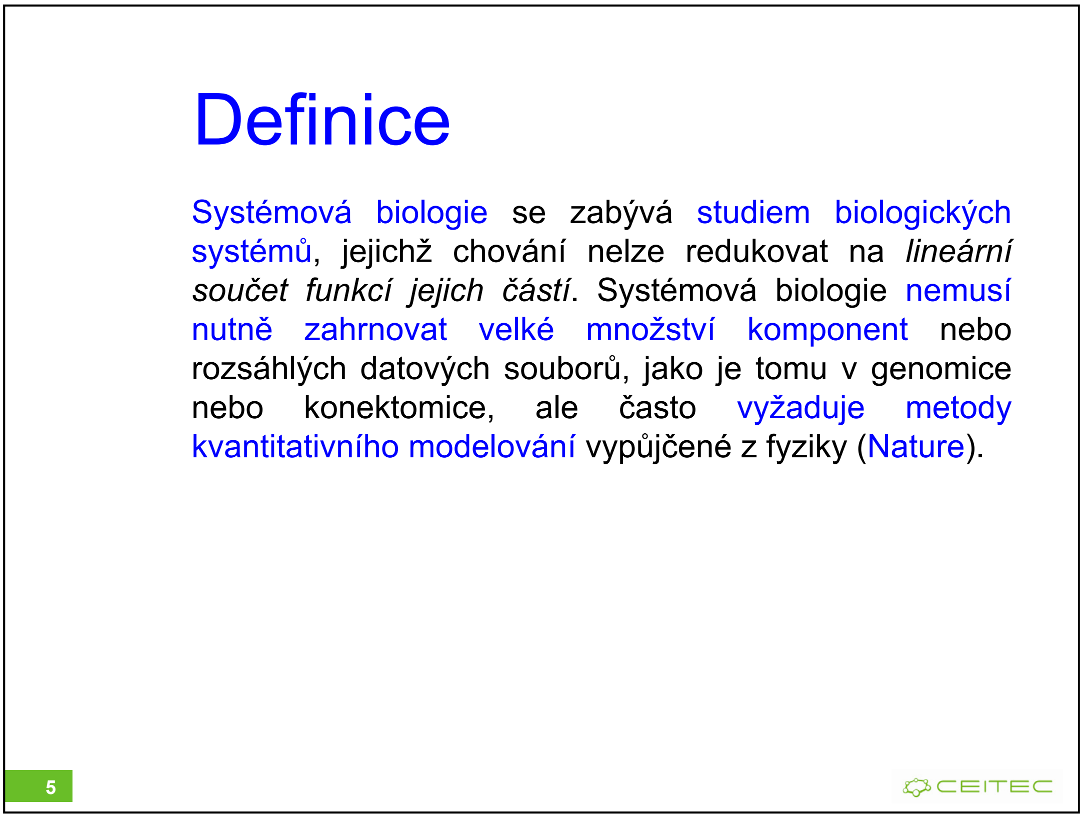
- Definice Systémové biologie
- Nástroje
  - Genová ontologie
  - Bayesovské sítě
  - Modelování molekulárních/genových regulačních sítí
  - Odvození genových regulačních sítí z velkých omických datových sad

Definice

Systémová biologie je vědecký směr v biologii využívající přístupy dalších věd, především biochemie, chemie, informatiky a matematiky. Zabývá se studiem biologických funkcí a mechanizmů vzniklých následkem komplexních interakcí v biologických systémech.

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

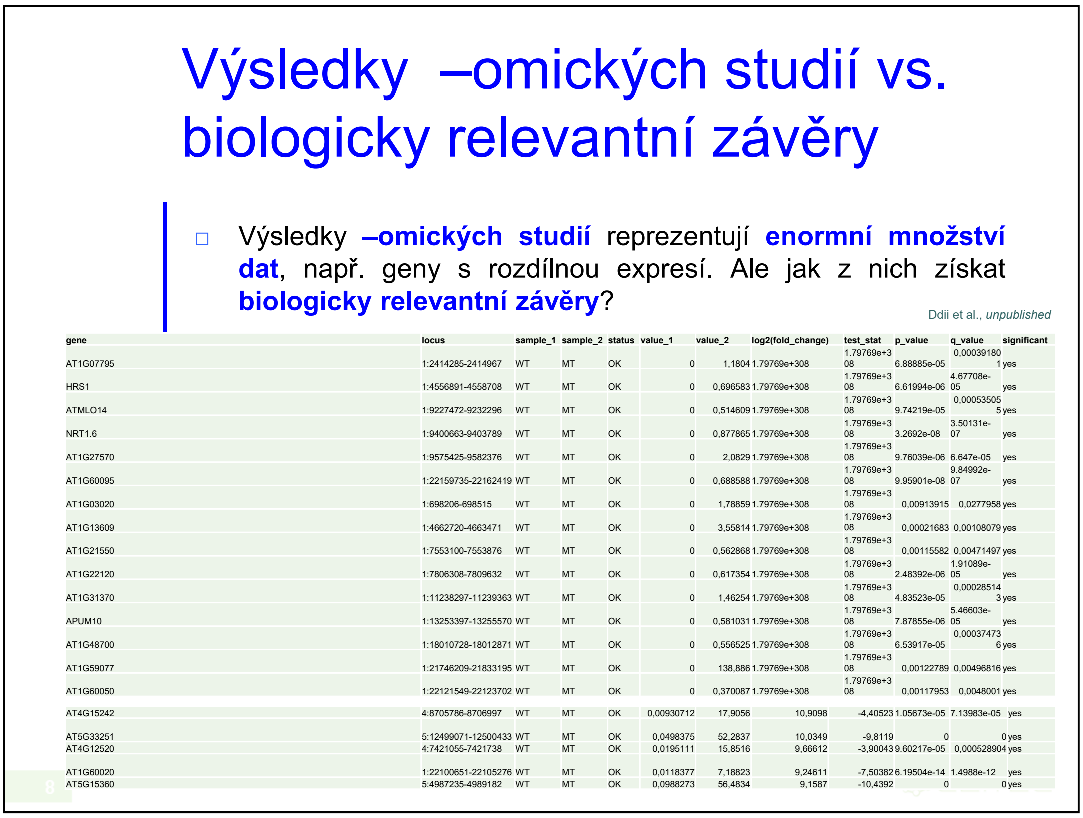
Systémová biologie se zabývá studiem biologických systémů, jejichž chování nelze redukovat na *lineární součet funkcí jejich částí*. Systémová biologie nemusí nutně zahrnovat velké množství komponent nebo rozsáhlých datových souborů, jako je tomu v genomice nebo konektomice, ale často vyžaduje metody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

Definice

Názorně vysvětluje video Dr. Nathana Price, zástupce ředitele Ústavu pro systémovou biologii na https://www.youtube.com/watch?v=OrXRI_8UFHU.

Osnova

- Definition of Systems Biology
- Tools
  - Gene Ontology analysis



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**?

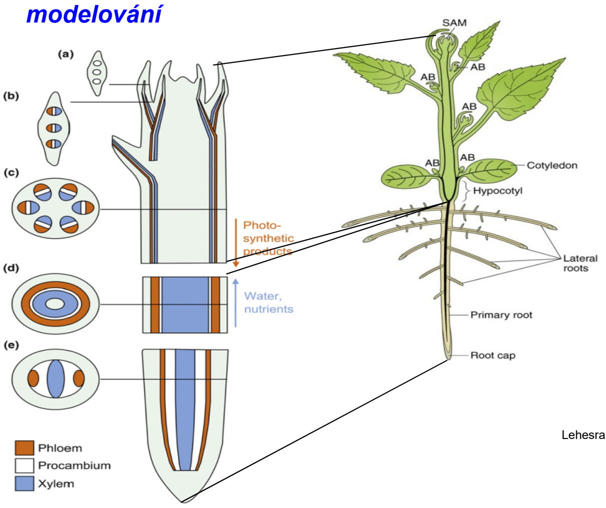
Ddii et al., *unpublished*

gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1,1804	1.79769e+308	1.79769e+308	6.88885e-05	0,000391801	yes
HRS1	1:4556891-4558708	WT	MT	OK	0	0,696583	1.79769e+308	1.79769e+308	6.61994e-06	4.67708e-05	yes
ATMLO14	1:9227472-9232296	WT	MT	OK	0	0,514609	1.79769e+308	1.79769e+308	9.74219e-05	0,000535055	yes
NRT1.6	1:9400663-9403789	WT	MT	OK	0	0,877865	1.79769e+308	1.79769e+308	3.2692e-08	3.50131e-07	yes
AT1G27570	1:9575425-9582376	WT	MT	OK	0	2,0829	1.79769e+308	1.79769e+308	9.76039e-06	6.647e-05	yes
AT1G60095	1:22159735-22162419	WT	MT	OK	0	0,688588	1.79769e+308	1.79769e+308	9.95901e-08	9.84992e-07	yes
AT1G03020	1:698206-698515	WT	MT	OK	0	1,78859	1.79769e+308	1.79769e+308	0,00913915	0,0277958	yes
AT1G13609	1:4662720-4663471	WT	MT	OK	0	3,55814	1.79769e+308	1.79769e+308	0,00021683	0,00108079	yes
AT1G21550	1:7553100-7553876	WT	MT	OK	0	0,562868	1.79769e+308	1.79769e+308	0,00115582	0,00471497	yes
AT1G22120	1:7806308-7809632	WT	MT	OK	0	0,617354	1.79769e+308	1.79769e+308	2.48392e-06	1.91089e-05	yes
AT1G31370	1:11238297-11239363	WT	MT	OK	0	1,46254	1.79769e+308	1.79769e+308	4.83523e-05	0,000285143	yes
APUM10	1:13253397-13255570	WT	MT	OK	0	0,581031	1.79769e+308	1.79769e+308	7.87855e-06	5.46603e-05	yes
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0,556525	1.79769e+308	1.79769e+308	6.53917e-05	0,000374736	yes
AT1G59077	1:21746209-21833195	WT	MT	OK	0	138,886	1.79769e+308	1.79769e+308	0,00122789	0,00496816	yes
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0,370087	1.79769e+308	1.79769e+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	5:12499071-12500433	WT	MT	OK	0,0498375	52,2837	10,0349	-9,8119	0	0	yes
AT4G12520	4:7421055-7421738	WT	MT	OK	0,0195111	15,8516	9,66612	-3,90043	9.60217e-05	0,000528904	yes
AT1G60020	1:22100651-22105276	WT	MT	OK	0,0118377	7,18823	9,24611	-7,50382	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

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

Vývoj rostlinných vodivých pletiv

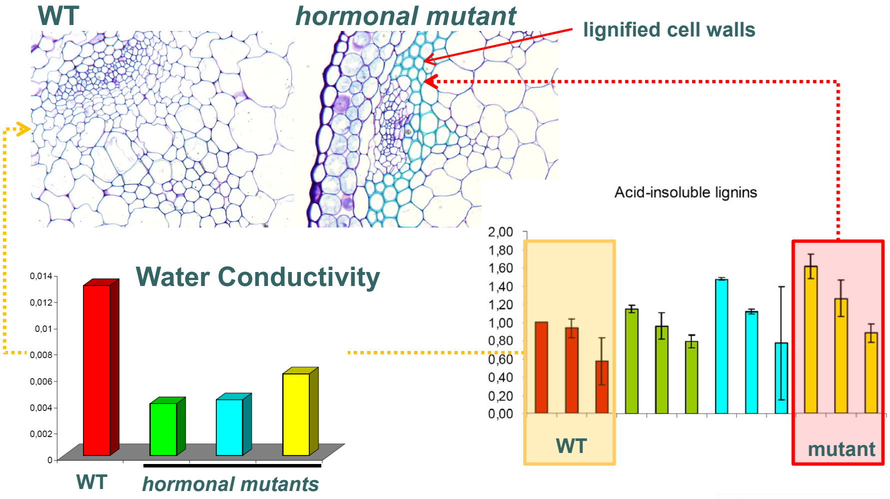
- ***Vodivé pletivo*** jako vývojový model pro ***GO analýzu*** a ***MRN modelování***



Lehesranta etal., *Trends in Plant Sci* (2010)

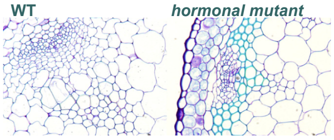
Hormonální regulace vývoje rostlinných vodivých pletiv

- Rostlinné **hormony regulují ukládání ligninu** v buněčných stěnách a **transport vody xylemem**



Hormonální regulace vývoje rostlinných vodivých pletiv

- ☐ ***Transkripční profilování*** pomocí ***sekvenování RNA***



mRNA mRNA

cDNA cDNA

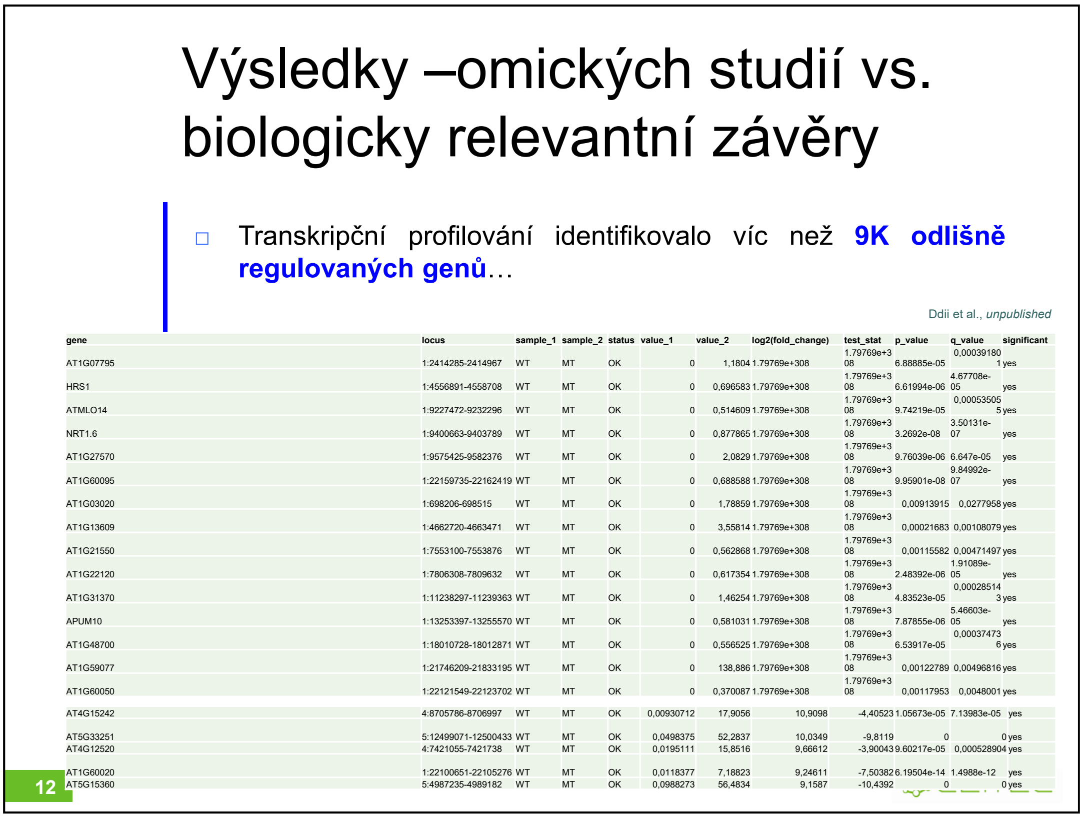

Library Preparation
~2 h [15 min hands-on (Nextera)]
< 6 h [< 3 h hands-on (TruSeq)]

Cluster Generation
~5 h (<10 min hands-on)

Sequencing by Synthesis
~1.5 to 11 days

CASAVA
2 days (30 min hands-on)

Sekvenování společností Illumina a určení **počtu transkriptů**



Výsledky –omických studií vs. biologicky relevantní závěry

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

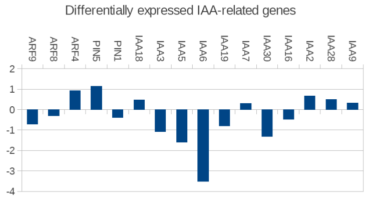
Ddii et al., *unpublished*

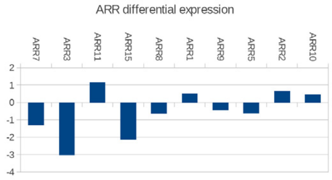
gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1,1804	1.79769e+308	1.79769e+308	6.88885e-05	0,000391801	yes
HRS1	1:4556891-4558708	WT	MT	OK	0	0,696583	1.79769e+308	1.79769e+308	6.61994e-06	4.67708e-05	yes
ATMLO14	1:9227472-9232296	WT	MT	OK	0	0,514609	1.79769e+308	1.79769e+308	9.74219e-05	0,000535055	yes
NRT1.6	1:9400663-9403789	WT	MT	OK	0	0,877865	1.79769e+308	1.79769e+308	3.2692e-08	3.50131e-07	yes
AT1G27570	1:9575425-9582376	WT	MT	OK	0	2,0829	1.79769e+308	1.79769e+308	9.76039e-06	6.647e-05	yes
AT1G60095	1:22159735-22162419	WT	MT	OK	0	0,688588	1.79769e+308	1.79769e+308	9.95901e-08	9.84992e-07	yes
AT1G03020	1:698206-698515	WT	MT	OK	0	1,78859	1.79769e+308	1.79769e+308	0,00913915	0,0277958	yes
AT1G13609	1:4662720-4663471	WT	MT	OK	0	3,55814	1.79769e+308	1.79769e+308	0,00021683	0,00108079	yes
AT1G21550	1:7553100-7553876	WT	MT	OK	0	0,562868	1.79769e+308	1.79769e+308	0,00115582	0,00471497	yes
AT1G22120	1:7806308-7809632	WT	MT	OK	0	0,617354	1.79769e+308	1.79769e+308	2.48392e-06	1.91089e-05	yes
AT1G31370	1:11238297-11239363	WT	MT	OK	0	1,46254	1.79769e+308	1.79769e+308	4.83523e-05	0,000285143	yes
APUM10	1:13253397-13255570	WT	MT	OK	0	0,581031	1.79769e+308	1.79769e+308	7.87855e-06	5.46603e-05	yes
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0,556525	1.79769e+308	1.79769e+308	6.53917e-05	0,000374736	yes
AT1G59077	1:21746209-21833195	WT	MT	OK	0	138,886	1.79769e+308	1.79769e+308	0,00122789	0,00496816	yes
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0,370087	1.79769e+308	1.79769e+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	5:12499071-12500433	WT	MT	OK	0,0498375	52,2837	10,0349	-9,8119	0	0	yes
AT4G12520	4:7421055-7421738	WT	MT	OK	0,0195111	15,8516	9,66612	-3,90043	9.60217e-05	0,000528904	yes
AT1G60020	1:22100651-22105276	WT	MT	OK	0,0118377	7,18823	9,24611	-7,50382	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

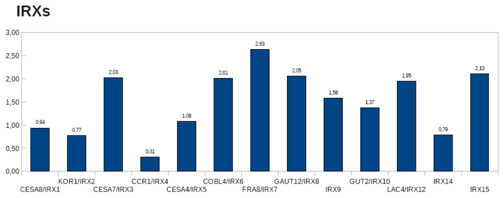
Excample of an output of transcriptional profiling study using Illumina sequencing performed in our lab. Shown is just a tiny fragment of the complete list, copmprising about 7K genes revealing differential expression in the studied mutant.

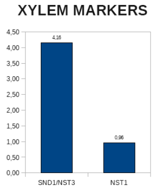
Genová ontologie

- ☐ Jedním z možných přístupů je studium ***genové ontologie***, tj. dříve prokázané **spojitosti** mezi geny a **biologickými procesy**


Differentially expressed IAA-related genes


ARR differential expression



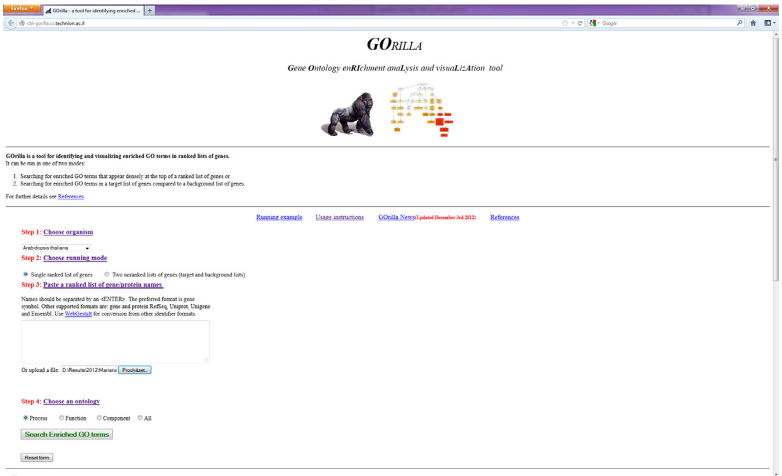


Ddii et al., *unpublished*

Genová ontologie

- Několik nástrojů umožňuje **statisticky vyhodnotit obohacení** o geny **spojené se specifickými procesy**

Eden et al., *BMC Biinformatics* (2009)



GOrilla

Gene Ontology enRIchment anaLysis and visuaLizAtion tool

GOrilla is a tool for identifying and visualizing enriched GO terms in ranked lists of genes.
It can be run in one of two modes:

1. Searching for enriched GO terms that appear densely 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.

For further details see References.

Running example Usage instructions GOrilla News(Updated December 3rd 2012) References

Step 1: Choose organism

Arabidopsis thaliana

Step 2: Choose running mode

● Single ranked list of genes ○ Two unranked lists of genes (target and background lists)

Step 3: Paste a ranked list of gene/protein names

Names should be separated by an <ENTER>. The preferred format is gene symbol. Other supported formats are: gene and protein RefSeq, Uniprot, Unigene and Ensembl. Use WebGestalt for conversion from other identifier formats.

Or upload a file: D:\Results\2012\Mariana Procházet...

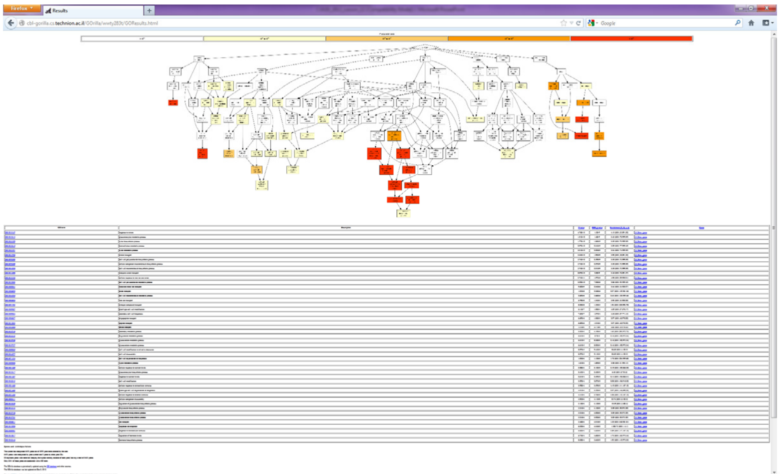
Step 4: Choose an ontology

● Process ○ Function ○ Component ○ All

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

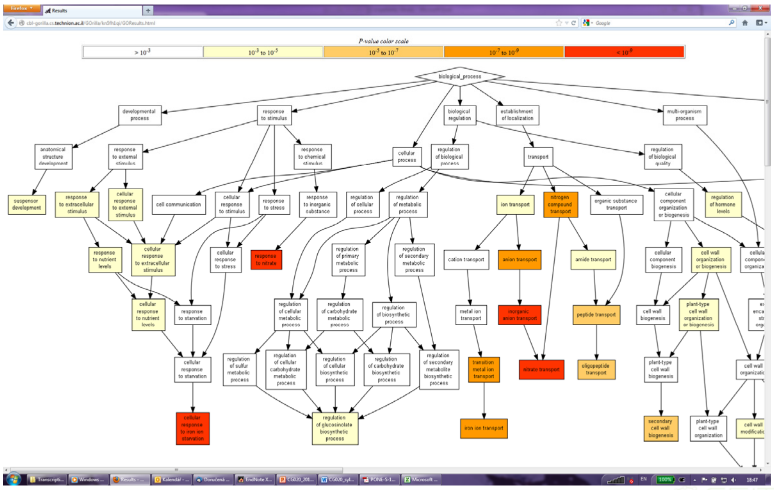
Genová ontologie

- Několik nástrojů umožňuje **statisticky vyhodnotit obohacení** o geny **spojené se specifickými procesy**



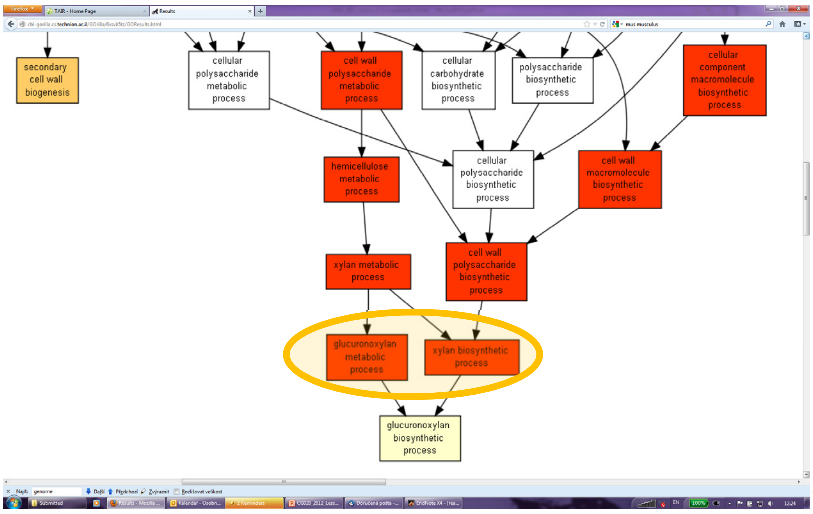
Genová ontologie

- Několik nástrojů umožňuje **statisticky vyhodnotit obohacení** o geny **spojené se specifickými procesy**



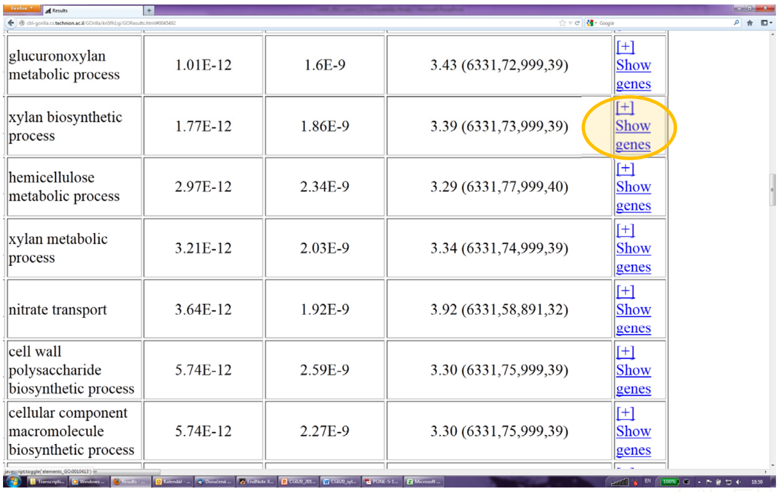
Genová ontologie

- □ Několik nástrojů umožňuje **statisticky vyhodnotit obohacení** o geny **spojené se specifickými procesy**



Genová ontologie

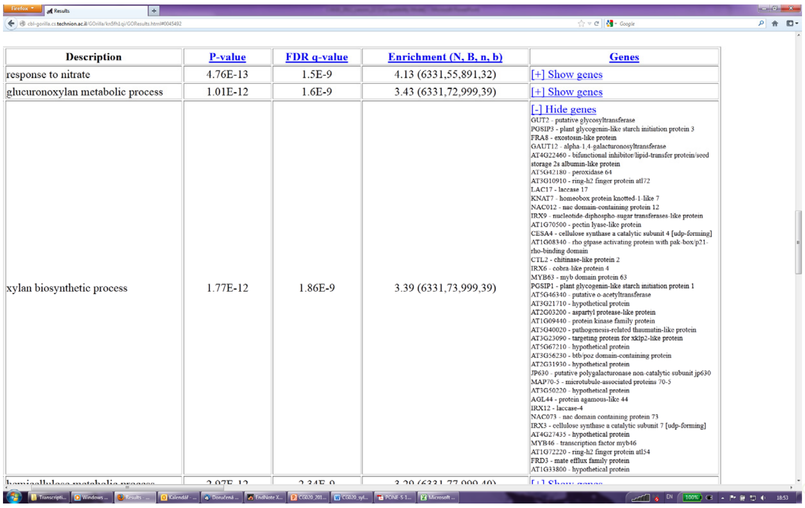
- Několik nástrojů umožňuje **statisticky vyhodnotit obohacení** o geny **spojené se specifickými procesy**



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

Genová ontologie

- Několik nástrojů umožňuje **statisticky vyhodnotit obohacení** o geny **spojené se specifickými procesy**



Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
response to nitrate	4.76E-13	1.5E-9	4.13 (6331,55,891,32)	[+] Show genes
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)	[-] Hide genes GUT2 - putative glycosyltransferase PGSIP3 - plant glycogenin-like starch initiation protein 3 FRA8 - exostosin-like protein GAUT12 - alpha-1,4-galacturonosyltransferase AT4G22460 - bifunctional inhibitor/lipid-transfer protein/seed storage 2s albumin-like protein AT5G42180 - peroxidase 64 AT3G10910 - ring-h2 finger protein atl72 LAC17 - laccase 17 KNAT7 - homeobox protein knotted-1-like 7 NAC012 - nac domain-containing protein 12 IRX9 - nucleotide-diphospho-sugar transferases-like protein AT1G70500 - pectin lyase-like protein CESA4 - cellulose synthase a catalytic subunit 4 [udp-forming] AT1G08340 - rho gtpase activating protein with pak-box/p21-rho-binding domain CTL2 - chitinase-like protein 2 IRX6 - cobra-like protein 4 MYB63 - myb domain protein 63 PGSIP1 - plant glycogenin-like starch initiation protein 1 AT5G46340 - putative o-acetyltransferase AT3G21710 - hypothetical protein AT2G03200 - aspartyl protease-like protein AT1G09440 - protein kinase family protein AT5G40020 - pathogenesis-related thaumatin-like protein AT3G23090 - targeting protein for xklp2-like protein AT5G67210 - hypothetical protein AT3G56230 - btb/poz domain-containing protein AT2G31930 - hypothetical protein JP630 - putative polygalacturonase non-catalytic subunit jp630 MAP70-5 - microtubule-associated proteins 70-5 AT3G50220 - hypothetical protein AGL44 - protein agamous-like 44 IRX12 - laccase-4 NAC073 - nac domain containing protein 73 IRX3 - cellulose synthase a catalytic subunit 7 [udp-forming] AT4G27435 - hypothetical protein MYB46 - transcription factor myb46 AT1G72220 - ring-h2 finger protein atl54 FRD3 - mate efflux family protein AT1G33800 - hypothetical protein
hemicellulose metabolic process	2.97E-12	2.34E-9	3.29 (6331,77,999,40)	[+] Show genes

Osnova

- Definice Systémové biologie
- Nástroje
  - Genová ontologie
  - Bayesovské sítě

Bayesovské sítě

Co je Bayesovská sít´?

- **Pravděpodobný grafický model,** který se používá k **vytváření modelů z dat a/nebo názoru odborníka**

Bayesovské sítě

22

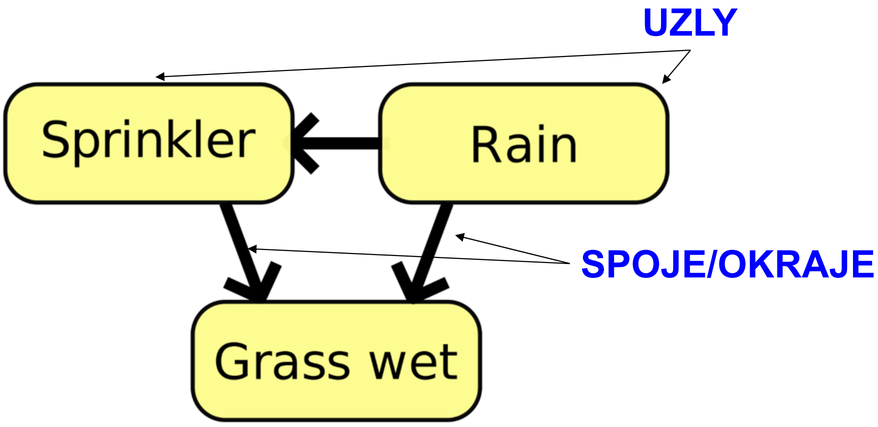
Bayesovské sítě

https://www.youtube.com/watch?v=4fcqyzVJwHM

Bayesovské sítě

- Co je **Bayesovská síť**?
  - **Pravděpodobný grafický model, který se používá k vytváření modelů z dat a/nebo názoru odborníka**
  - může být využit v široké škále úkolů včetně **predikce, detekce anomálie, diagnostiky, automatického pohledu na věc, uvažování, predikce časové řady** a **rozhodování za nejistoty**
- **UZLY**
  - každý uzel představuje **proměnnou**, jako je výška, věk nebo pohlaví. Proměnná může být *diskrétní*, jako například pohlaví = {samičí, samčí}, nebo *spojitá*, jako např. věk
- **SPOJE**
  - přidány **mezi uzly,** aby ukazovaly, že **jeden uzel má** přímý **vliv** na **druhý**

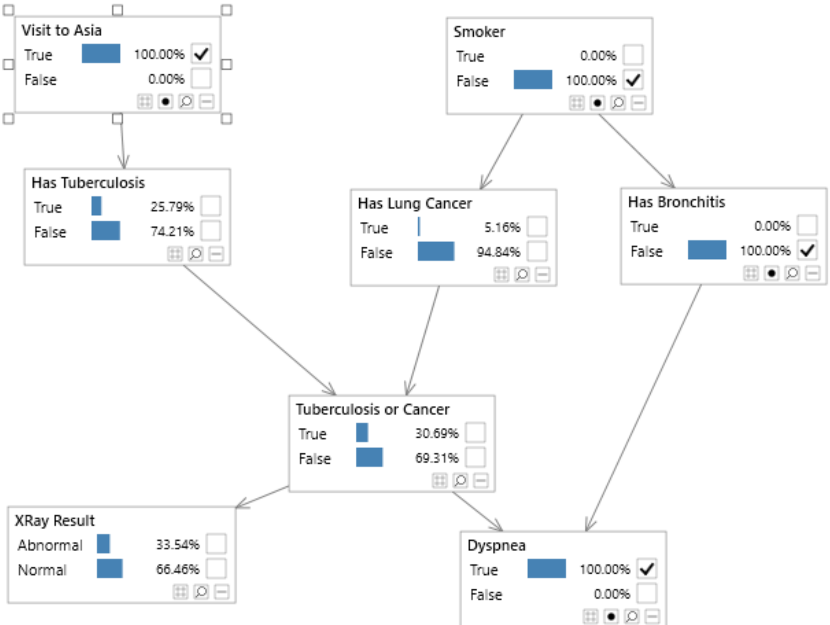
Bayesovské sítě



UZLY

SPOJE/OKRAJE

Asijská Bayesovská síť



Visit to Asia
True 100.00% ☑
False 0.00% ☐

Smoker
True 0.00% ☐
False 100.00% ☑

Has Tuberculosis
True 25.79% ☐
False 74.21% ☐

Has Lung Cancer
True 5.16% ☐
False 94.84% ☐

Has Bronchitis
True 0.00% ☐
False 100.00% ☑

Tuberculosis or Cancer
True 30.69% ☐
False 69.31% ☐

XRay Result
Abnormal 33.54% ☐
Normal 66.46% ☐

Dyspnea
True 100.00% ☑
False 0.00% ☐

https://www.bayesserver.com/

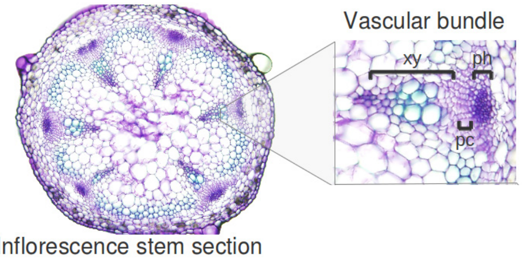
Osnova

- Definice Systémové biologie
- Nástroje
    - Genová ontologie
    - Bayesovské sítě
    - Modelování molekulárních/genových regulačních sítí

CEITEC

Modelování molekulárních regulačních sítí

- ***Vodivé pletivo*** jako vývojový model pro ***MRN modelování***



Benitez and Hejatko, *PLoS One*, 2013

Modelování molekulárních regulačních sítí

- **Vyhledávání** publikovaných dat a **vytvoření malé databáze**

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]
	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 ARRs.	[9]

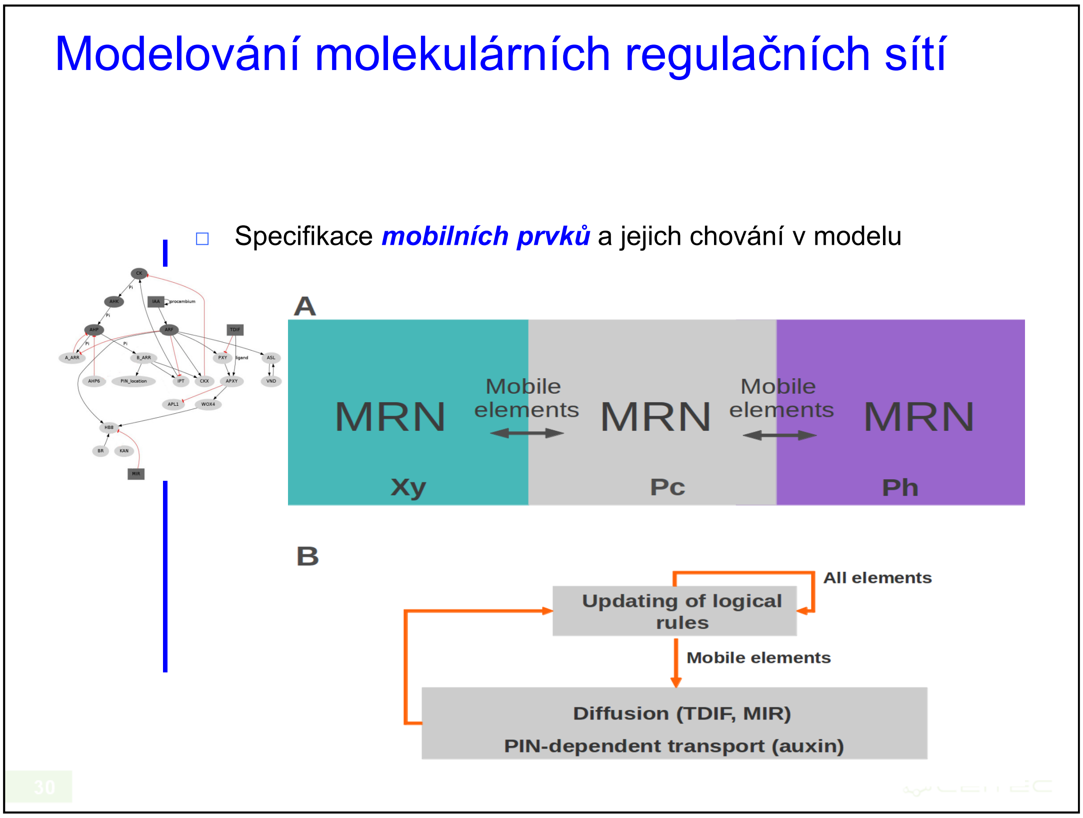
Benitez and Hejatko, ***PLoS One*, 2013**

Modelování molekulárních regulačních sítí

- Formulace ***logických pravidel*** definujících ***dynamiku modelu***

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

Benitez and Hejatko, *PLoS One*, 2013



Modelování molekulárních regulačních sítí

- ◻ Specifikace ***mobilních prvků*** a jejich chování v modelu

A

MRN | Mobile elements | MRN | Mobile elements | MRN

Xy | Pc | Ph

B

Updating of logical rules

All elements

Mobile elements

Diffusion (TDIF, MIR)

PIN-dependent transport (auxin)

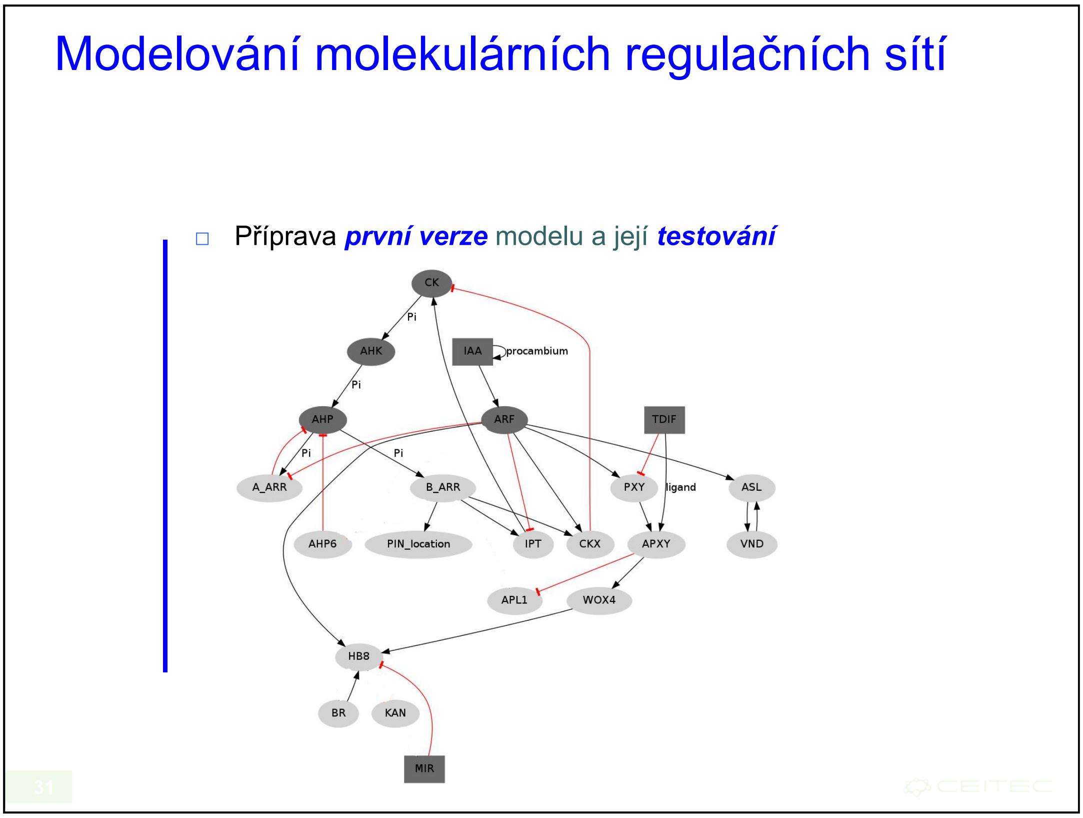
According to experimental evidence for the system under study, the hormone IAA, the peptide TDIF, and the microRNA MIR165/6 are able to move among the cells. In the case of TDIF and MIR165/6, the mobility is defined as diffusion and is given by the following equation:

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

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:

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

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 $biaa$ 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***

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, PLoS One, 2013.

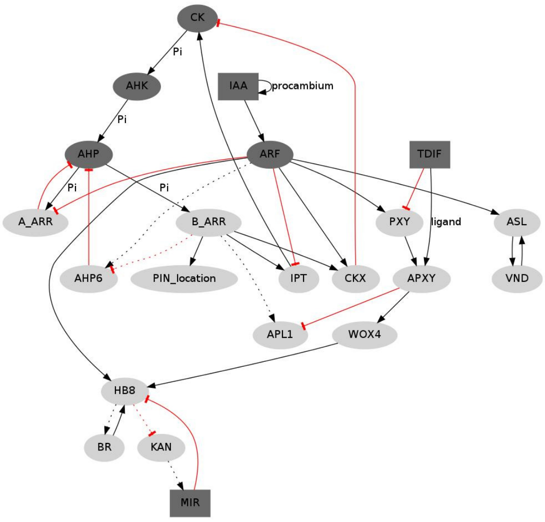
Modelování molekulárních regulačních sítí

- 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 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→ 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]
	Partially supported by microarray data and phloem-specific expression patterns of CK response factors.	(TAIR, ExpressionSet: 1005823559, [22])

Modelování molekulárních regulačních sítí

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

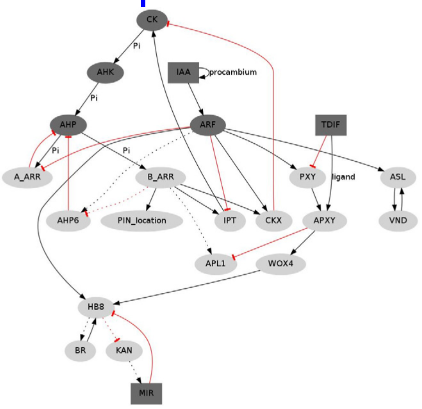


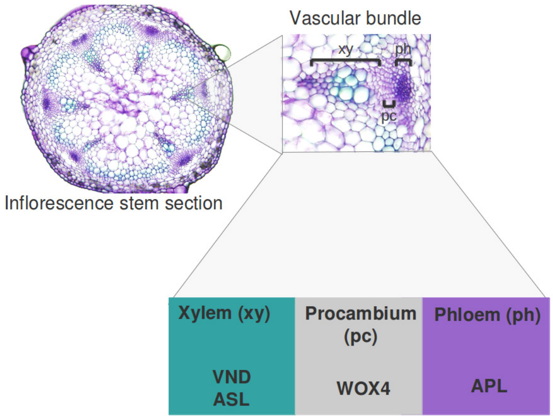
Benitez and Hejatko, *PLoS One*, 2013

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

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



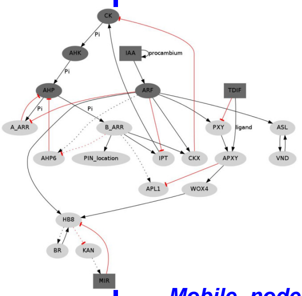


Xylem (xy)	Procambium (pc)	Phloem (ph)
VND ASL	WOX4	APL

Benitez and Hejatko, *PLoS One*, 2013

Modelování molekulárních regulačních sítí

- Formulace ***rovnic*** popisujících ***vzájemné vztahy*** v modelu



logical rule function

state in the time *t*

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

state in the time *t+1*

Amount if TDIF or MIR165 in cell *i*

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

state in the time *t+1*

constant corresponding to a degradation term

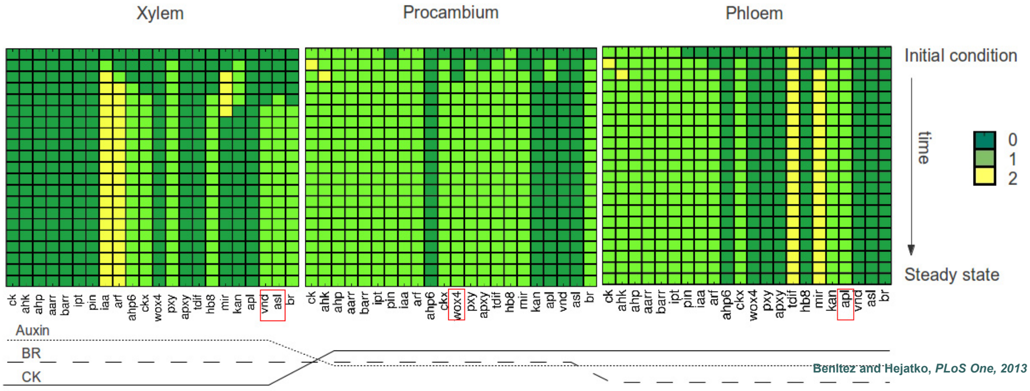
proportion of movable element

Modelování molekulárních regulačních sítí

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

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

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



Benitez and Hejatko, *PLoS One, 2013*

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 *HB8*, 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 *HB8*, 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 (*g1*, *g2*, ..., *gN*) at a previous time unit. The state of every gene *g* therefore changes according to:

$gn(t+1)=Fn(gn1(t),gn2(t),\dots, gnk(t))$ (1).

In this equation, *gn1*, *gn2*,…, *gnk* are the regulators of gene *gn* and *Fn* 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 selfsustained 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:

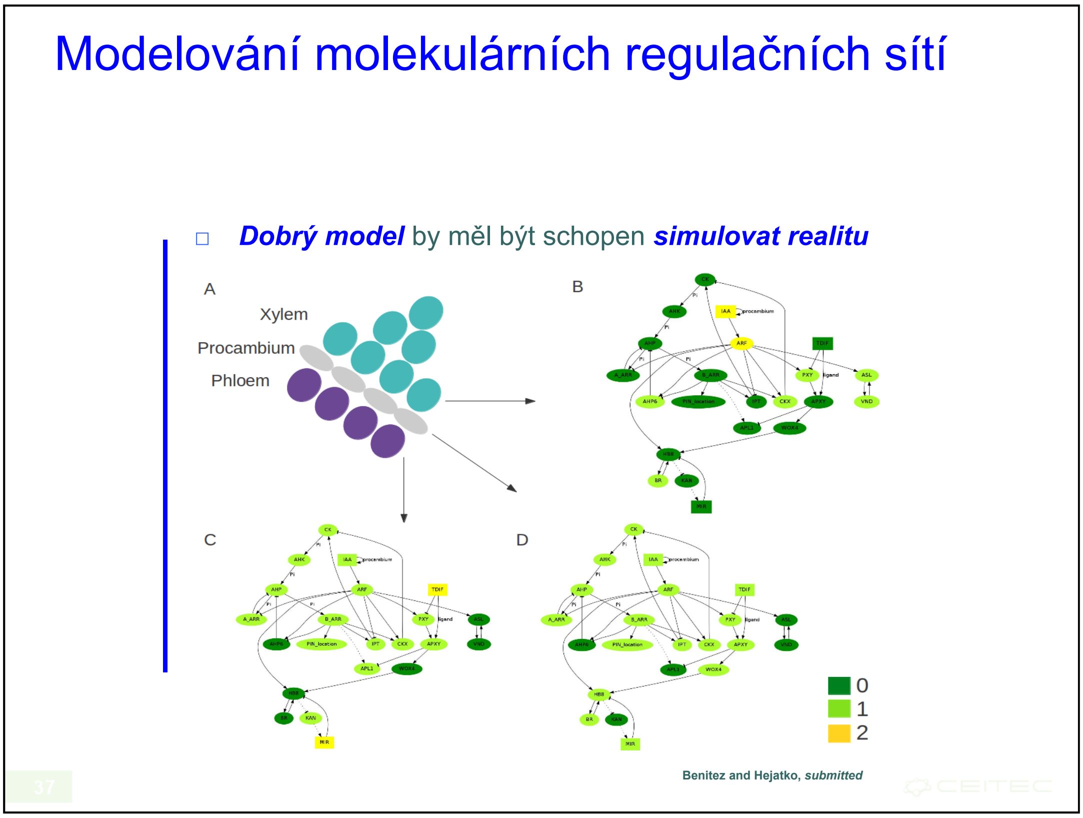
$g(t+1)T[i]= H(g(t)[i]+ D\,(g(t)[i+1]+g(t)[i-1] - N(g(t)[i]))-b)$ (2),

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:

$iaa(t+1)T[i]=Hiaa(iaa(t)[i]+Diaa(pin(t)[i+1])(iaa(t)[i+1])+Diaa(pin(t)[i-1])(iaa(t)[i-1])-N(Diaa)(pin(t)[i])(iaa(t)[i])-biaa)$ (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 *biaa* 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

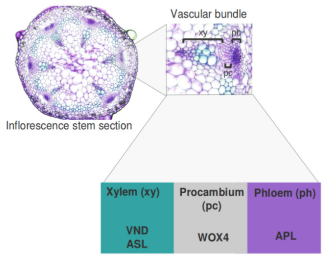
- ☐ ***Dobrý model*** by měl být schopen ***simulovat realitu***

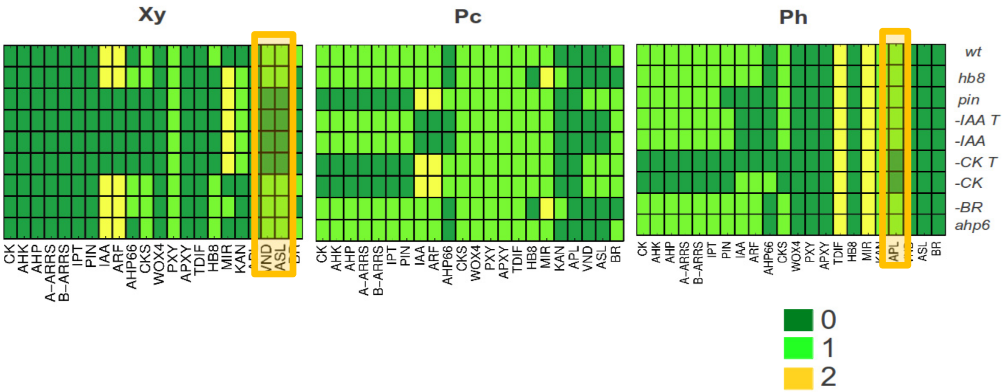
Benitez and Hejatko, *submitted*

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

Modelování molekulárních regulačních sítí

- Simulace ***mutantů***

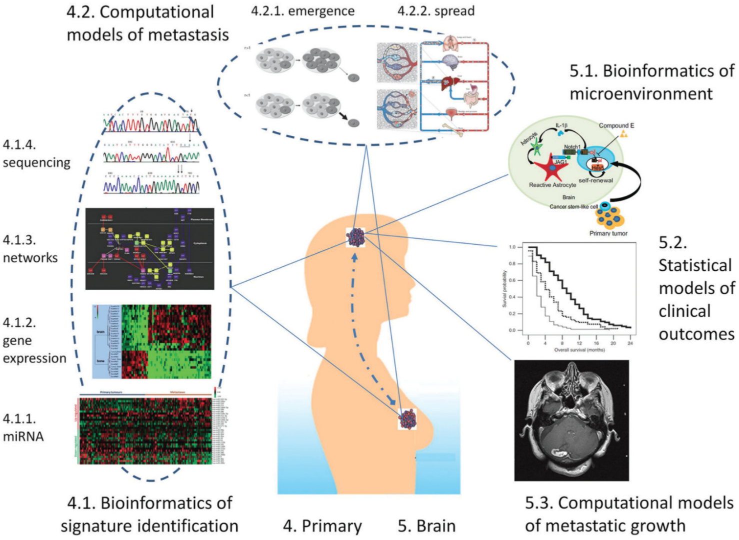




Osnova

- Definice Systémové biologie
- Nástroje
  - Genová ontologie
  - Bayesovské sítě
  - Modelování molekulárních/genových regulačních sítí
  - Odvození genových regulačních sítí z velkých omických datových sad

Systémová biologie ve výzkumu rakoviny



4.2. Computational models of metastasis

4.2.1. emergence

4.2.2. spread

5.1. Bioinformatics of microenvironment

4.1.4. sequencing

4.1.3. networks

4.1.2. gene expression

4.1.1. miRNA

5.2. Statistical models of clinical outcomes

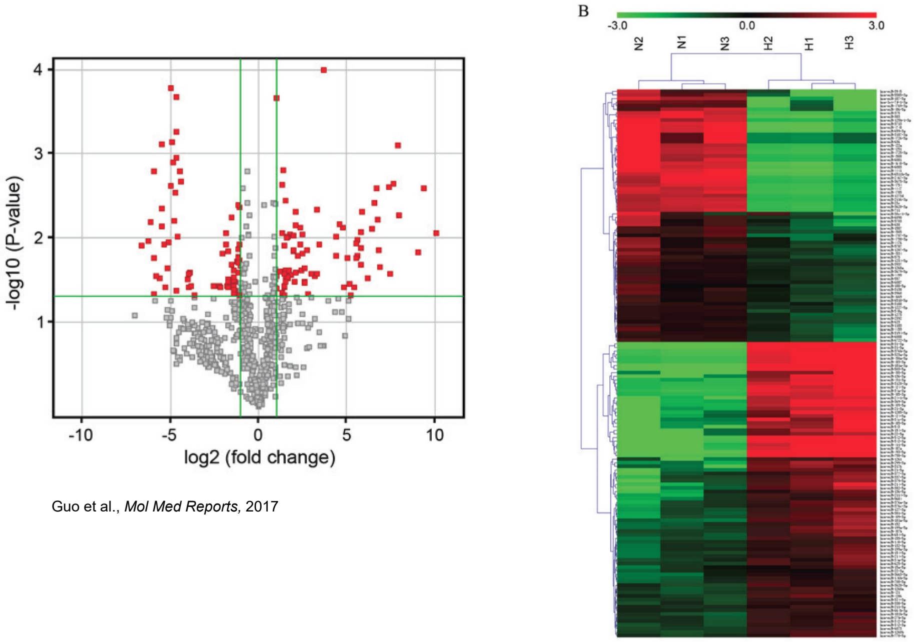
4.1. Bioinformatics of signature identification

4. Primary

5. Brain

5.3. Computational models of metastatic growth

miRNA/mRNA profilování

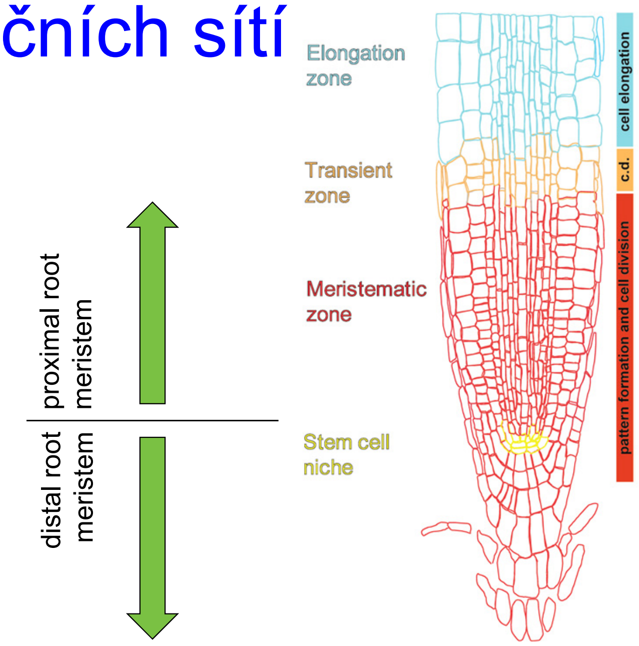


Guo et al., *Mol Med Reports*, 2017

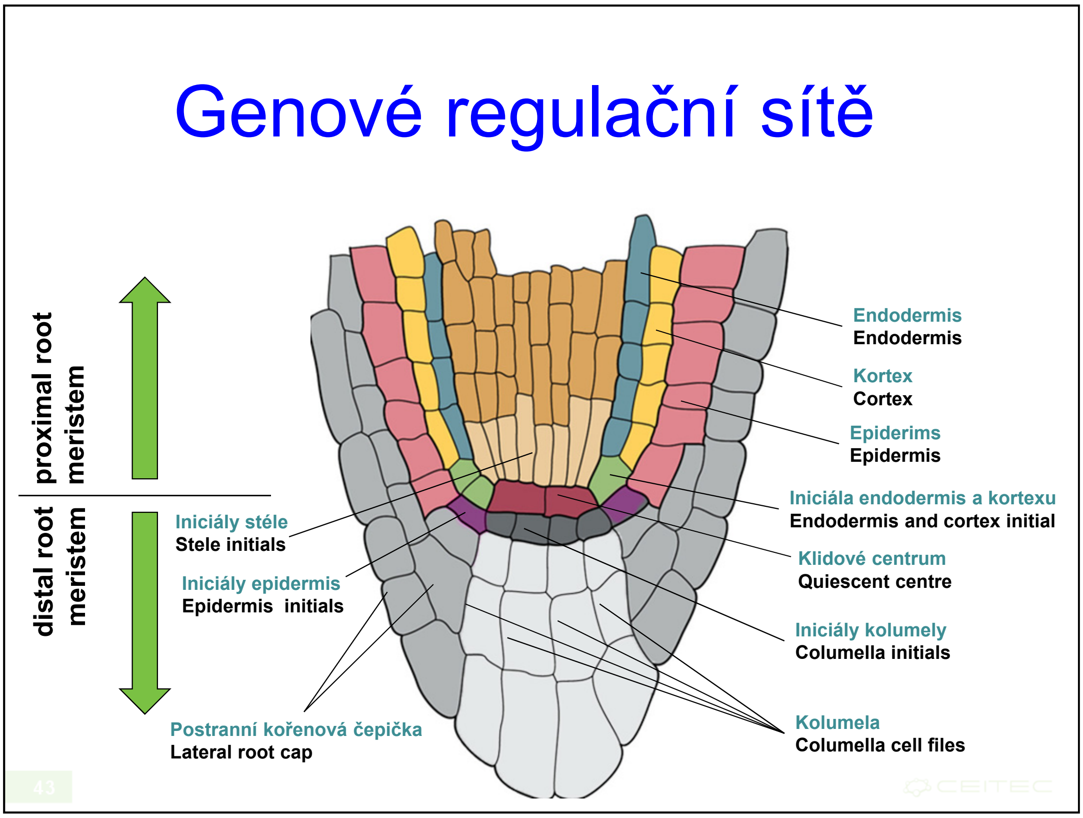
Hypertrophic scars (HS) area fibroproliferative disorder of the skin, which causes aesthetic and functional impairment. However, the molecular pathogenesis of this disease remains largely unknown and currently no efficient treatment exists. MicroRNAs (miRNAs) are involved in a variety of pathophysiological 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 α 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 ≥2 and P<0.05. Red dots indicate points-of-interest that exhibit large-magnitude fold-changes (x-axis; log2 of the fold change) and high statistical significance (y-axis; -log10 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



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



Genové regulační sítě

proximal root meristem

distal root meristem

Inciály stéle
Stele initials

Iniciály epidermis
Epidermis initials

Postranní kořenová čepička
Lateral root cap

Endodermis
Endodermis

Kortex
Cortex

Epiderims
Epidermis

Iniciála endodermis a kortexu
Endodermis and cortex initial

Klidové centrum
Quiescent centre

Iniciály kolumely
Columella initials

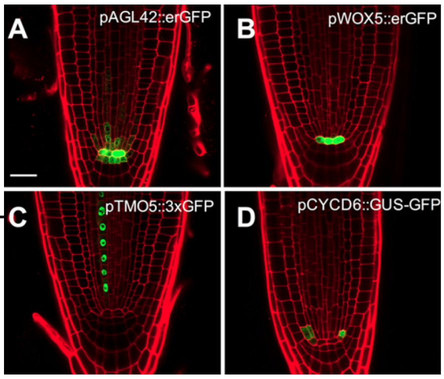
Kolumela
Columella cell files

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

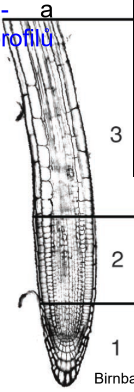
Along the longitudinal axis, the root meristem forms a distal root tip, including stem cell niche, columella and lateral root cap, proximal meristem with a population of rapidly dividing cells and elongation zone where cells leaving the root meristem undergo rapid elongation and mature.

Genové regulační sítě - GENIST

- Odvození GR sítí přes **GENIST**
  - **GE**ne regulatory **N**etwork **I**nference from **S**patio**T**emporal data algorithm
  - Kombinace prostorových- a časově- specifických profilů exprese genů



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

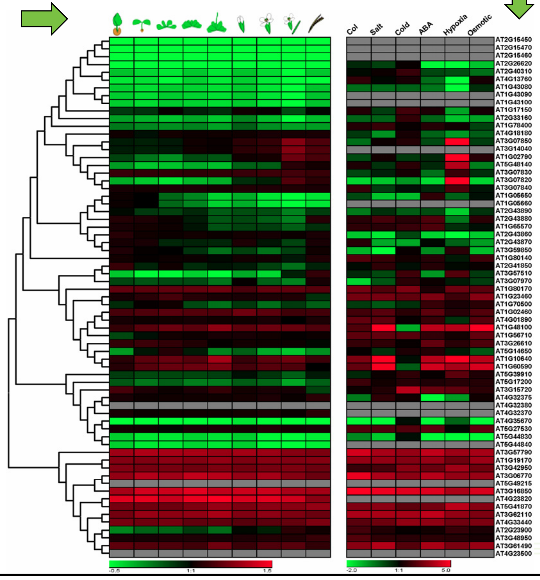


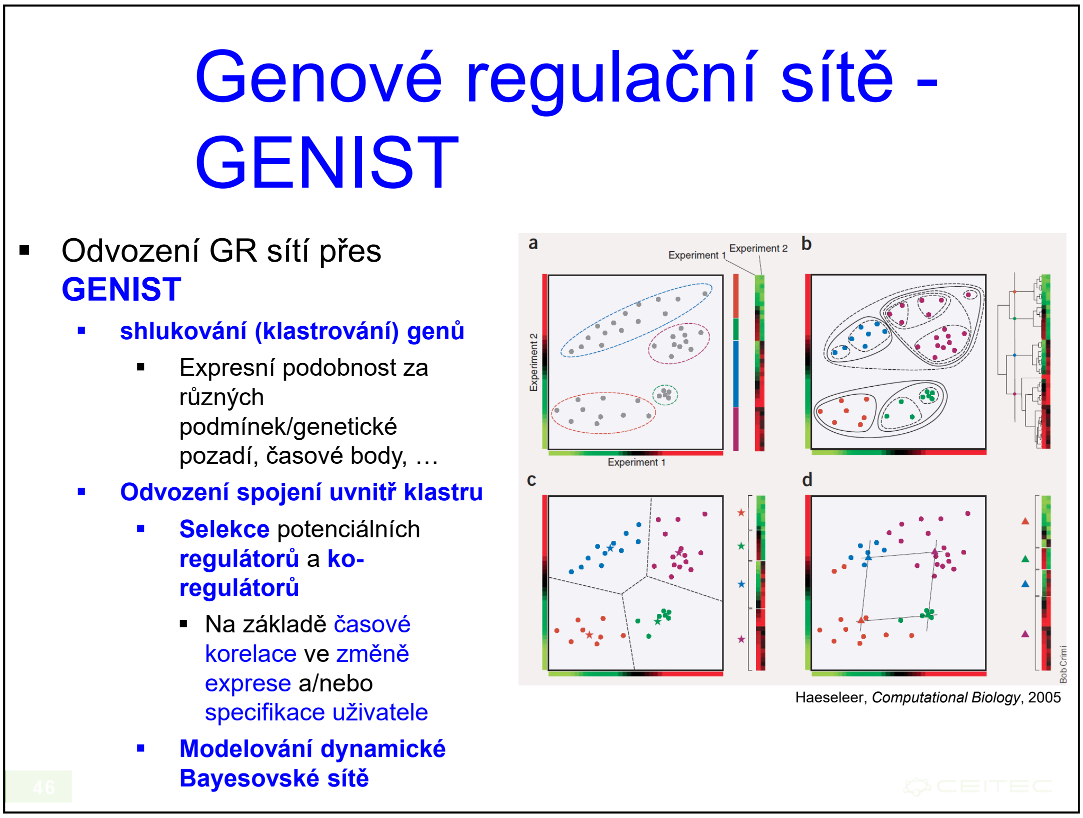
Birnbaum et al., *Science*, 2003

Kombinace velkých omických datových sad

GENY

PLETIVO/ČAS





Genové regulační sítě - GENIST

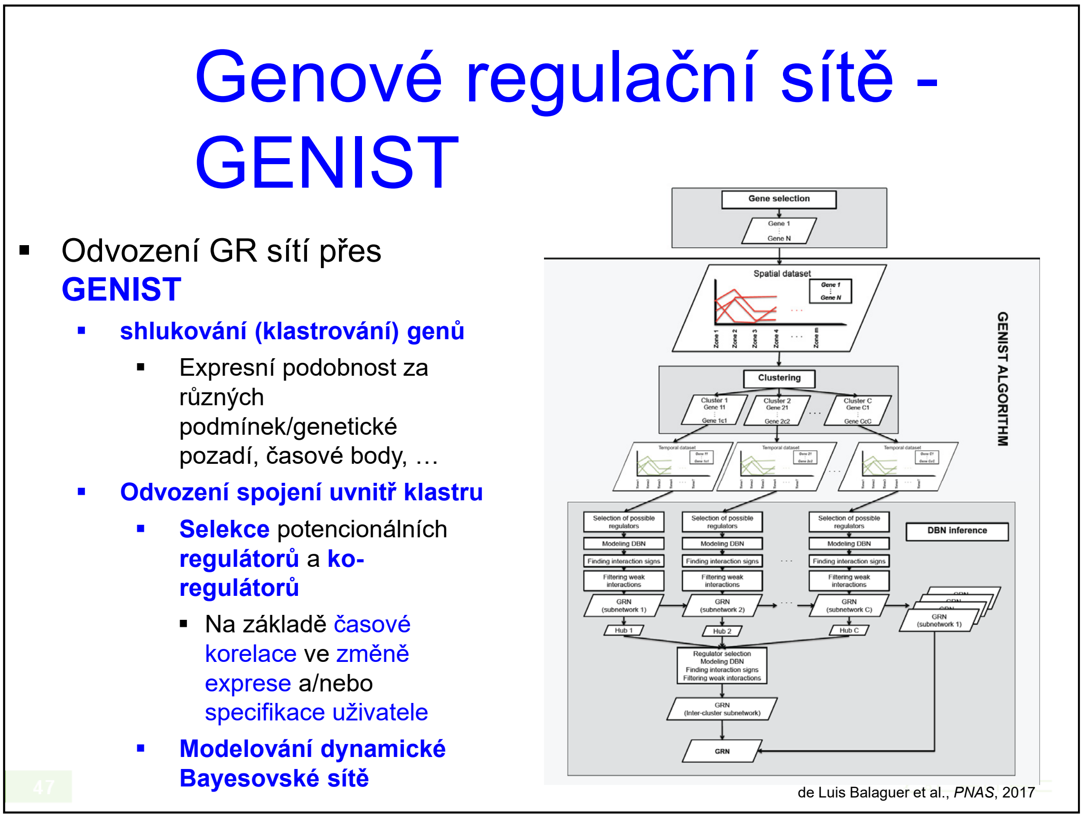
- Odvození GR sítí přes **GENIST**
  - **shlukování (klastrování) genů**
    - Expresní podobnost za různých podmínek/genetické pozadí, časové body, …
  - **Odvození spojení uvnitř klastru**
    - **Selekce** potenciálních **regulátorů** a **ko-regulátorů**
      - Na základě časové korelace ve změně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

Haeseleer, *Computational Biology*, 2005

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, …
  - **Odvození spojení uvnitř klastru**
    - **Selekce** potencionálních **regulátorů** a **ko-regulátorů**
      - Na základě časové korelace ve změně exprese a/nebo specifikace uživatele
    - **Modelování dynamické Bayesovské sítě**

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

GENIST algorithm

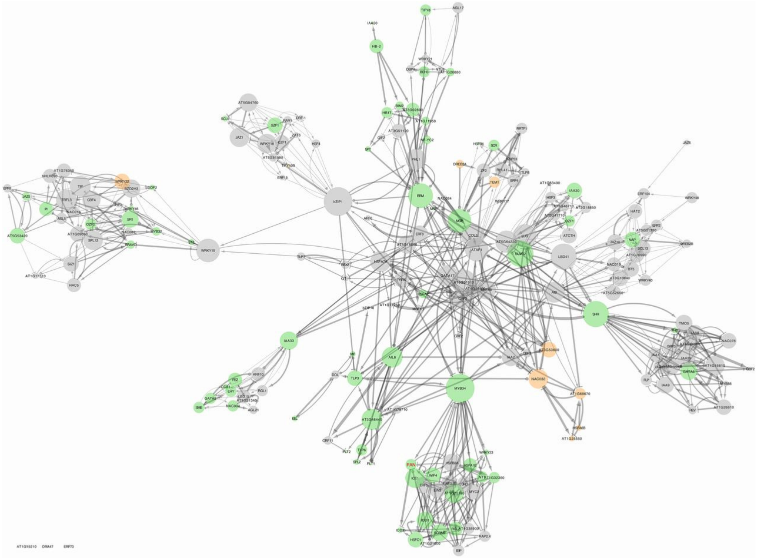
The MATLAB source code for GENIST is publically available at https://github.com/madeluis/GENIST.

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

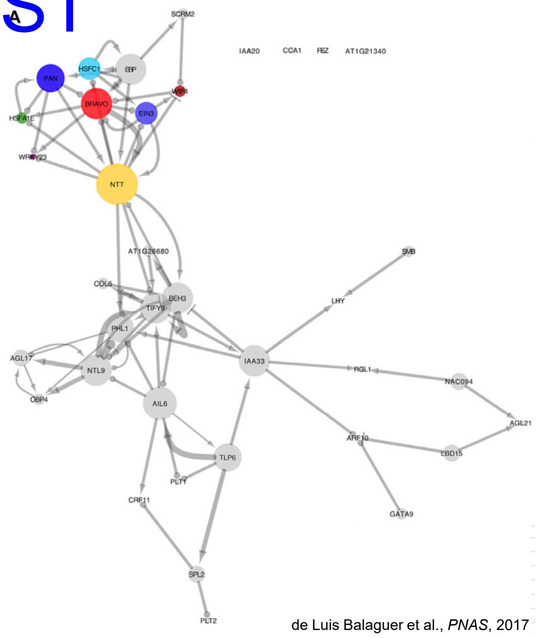
independently processed by the GRN inference step, based on a temporal dataset.

Genové regulační sítě - GENIST

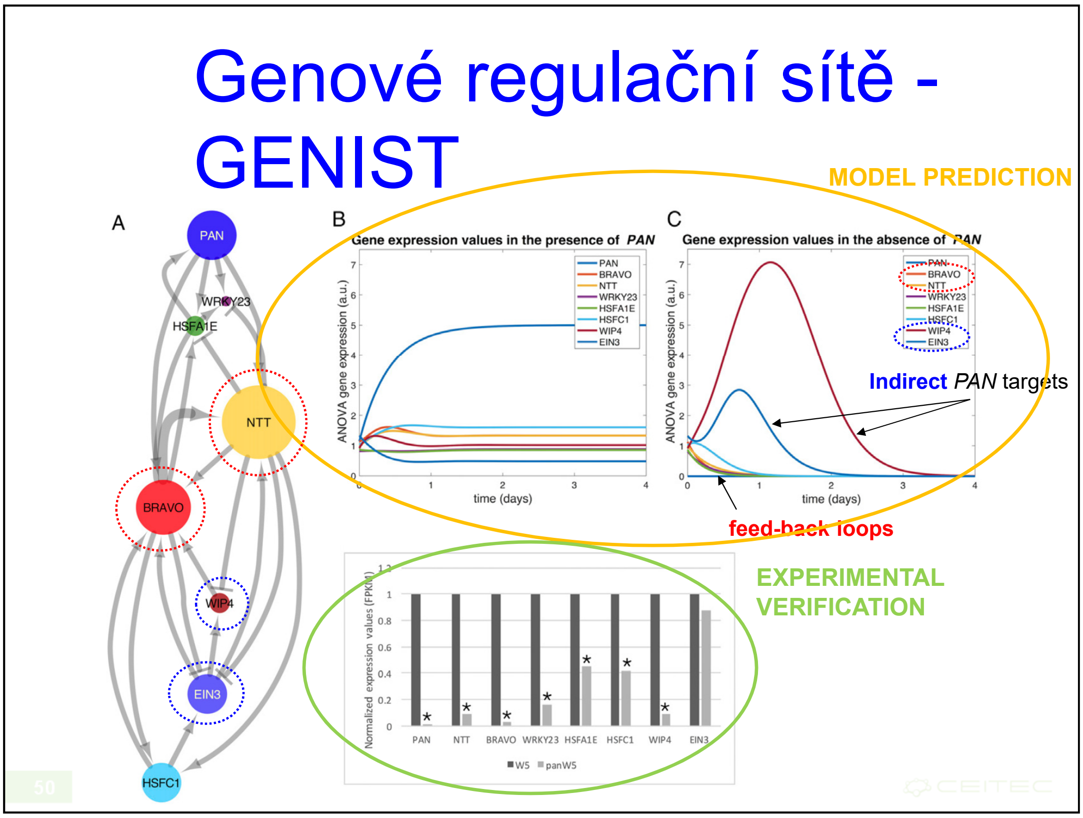


de Luis Balaguer et al., *PNAS*, 2017

Genové regulační sítě - GENIST



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



Genové regulační sítě - GENIST

MODEL PREDICTION

Indirect *PAN* targets

feed-back loops

EXPERIMENTAL VERIFICATION

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.

Klíčové koncepty

- Systémová biologie se pokouší identifikovat nové vlastnosti/chování skupin funkčních podjednotek (regulátorů/molekul), které nejsou prostým součtem vlastností jednotlivých podjednotek, ale jsou novou vlastností závislou na způsobu jejich vzájemné interakce
- Využívá matematické modely, často Bayesovské sítě
- Genové regulační sítě lze identifikovat i pomocí (semi)automatických nástrojů z velkých datových sad (např. genové exprese na úrovni celého genomu)
- Využití metod strojového účení („umělá inteligence“)

Diskuse