

NPO MUNI MSMT-16606/2022

Analýza genové exprese a nástroje systémové biologie

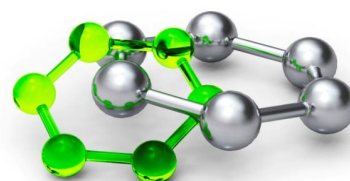
Jan Hejátko

Funkční genomika a proteomika rostlin,
Středoevropský technologický institut (CEITEC)

a
Národní centrum pro výzkum biomolekul,
Přírodovědecká fakulta,

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

M U N I
S C I



Zdrojová literatura

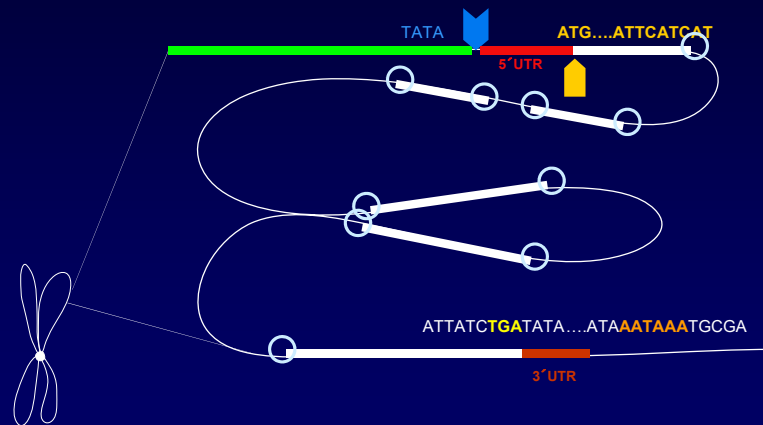
- Benitez, M. and Hejatko, J. Dynamics of cell-fate determination and patterning in the vascular bundles of *Arabidopsis thaliana* (submitted)
- Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. 2007. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318, 801-806.
- 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
- 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.
- Fucile G, Di Biase D, Nahal H, La G, Khodabandeh S, Chen Y, Easley K, Christendat D, Kelley L, Provart NJ. 2011. ePlant and the 3D data display initiative: integrative systems biology on the world wide web. *PLoS One* 6, e15237.
- Karaikos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP (2017) The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 358, 194-199
- Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 131, 174-187.
- Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., and Amster-Choder, O. (2011). Translation-independent localization of mRNA in *E. coli*. *Science* 331, 1081-1084

Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - **Kvalitativní analýza exprese genů**
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - **Tkáňově a buněčně specifická** analýza genové exprese
 - **Kvantitativní analýza exprese**
 - **DNA čipy**
 - **Next gen** transkripční profilování
- **Systémová biologie**
 - **Definice**
 - **Nástroje**
 - **Genová ontologie**
 - **Bayesovské sítě**
 - Modelování **molekulárních/genových regulačních sítí**

Struktura genů

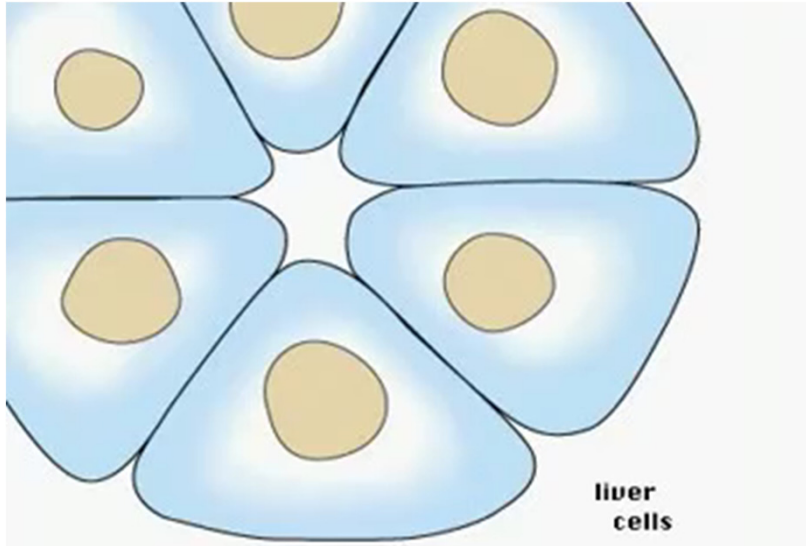
- promotor
- počátek transkripce
- 5' UTR
- počátek translace
- místa sestřihu
- stop kodon
- 3' UTR
- polyadenylační signál



Osnova

- Struktura genů
- Metody analýzy **genové exprese**

Časoprostorová specifita genové exprese



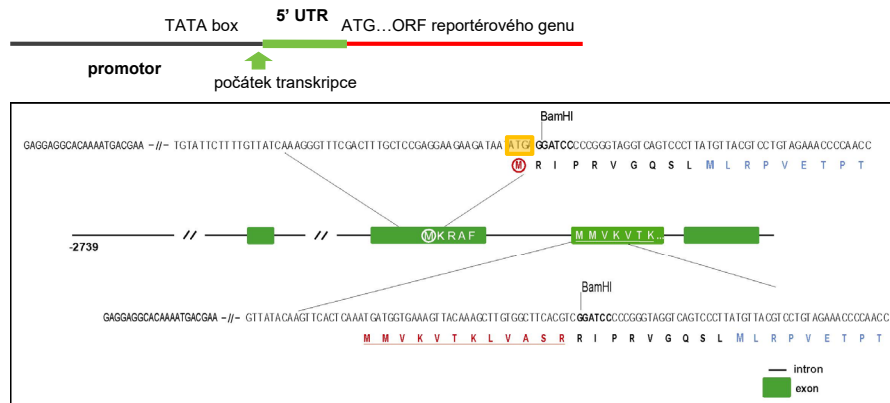
Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - **Kvalitativní** analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)

Transkripční fúze

- ☐ **Příprava transkripční fúze promotoru s reportérem**

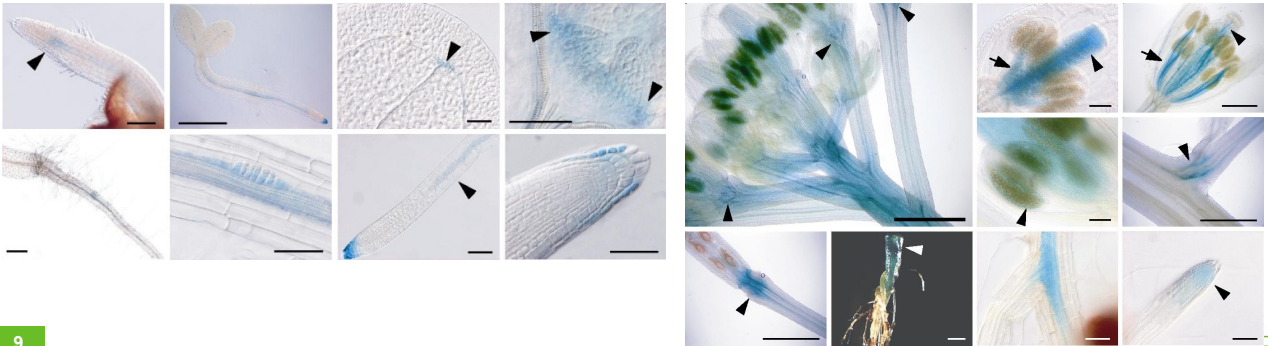
- Identifikace a klonování promotorové oblasti genu
- příprava rekombinantní DNA nesoucí promotor a reportérový gen (uidA, GFP)



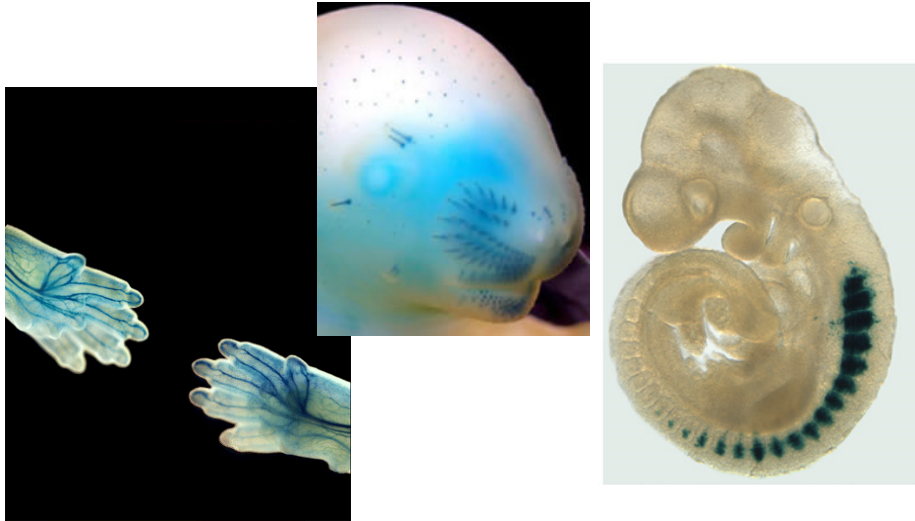
Transkripční fúze

□ Příprava transkripční fúze promotoru s reportérem

- Identifikace a klonování promotorové oblasti genu
- příprava rekombinantní DNA nesoucí promotor a reportérový gen (uidA, GFP)
- příprava transgenních organismů nesoucích tuto rekombinantní DNA a jejich histologická analýza



LacZ reporter in mouse embryos

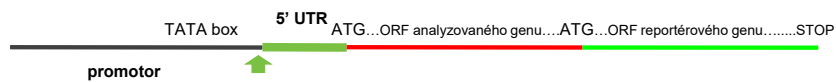


Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem

Translační fúze

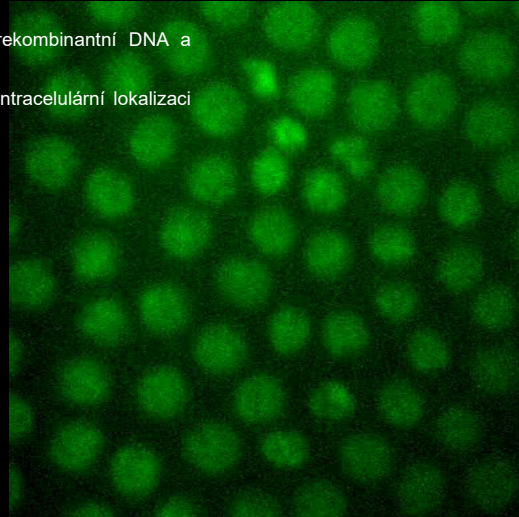
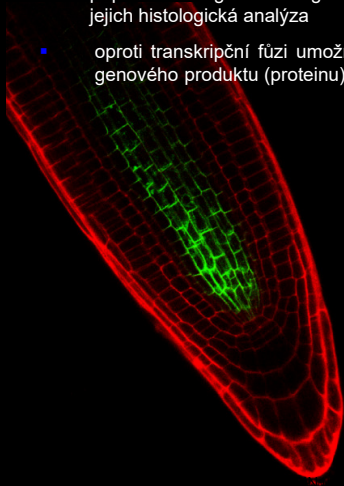
- **Příprava translační fúze kódující oblasti analyzovaného genu s reportérovým genem**
 - Identifikace a klonování **promotorové** a **kódující** oblasti analyzovaného genu
 - příprava **rekombinantní DNA** nesoucí **promotor** a **kódující sekvenci** **studovaného genu** ve **fúzi** s **reportérovým genem** (uidA, GFP)



Translační fúze

□ Translační fúze kódující oblasti analyzovaného genu s repotérovým genem

- příprava transgenních organismů nesoucích tuto rekombinantní DNA a jejich histologická analýza
- oproti transkripční fúzi umožňuje analyzovat např. intracelulární lokalizaci genového produktu (proteinu) nebo jeho dynamiku



Translační fúze

- Translační fúze kódující oblasti analyzovaného genu s repotérovým genem

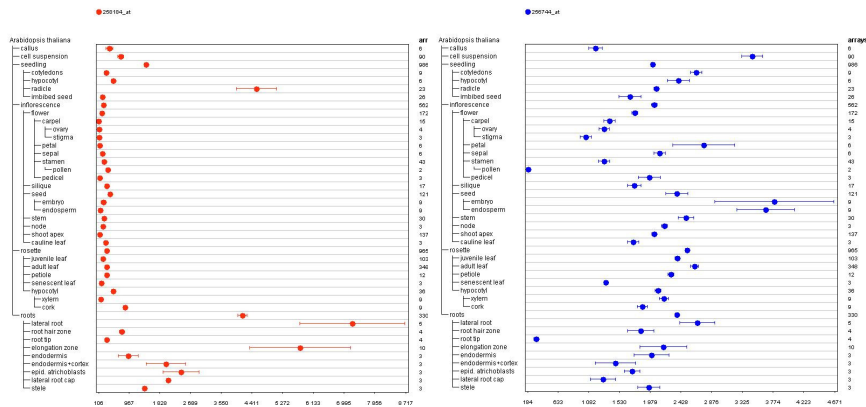


Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**

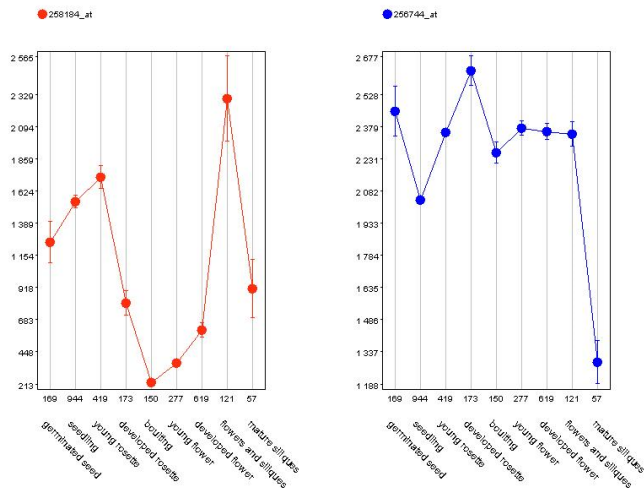
Databáze

- **Analýza exprese pomocí Genevestigator (*AHP1* a *AHP2*, *Arabidopsis*, Affymetrix ATH 22K Array)**



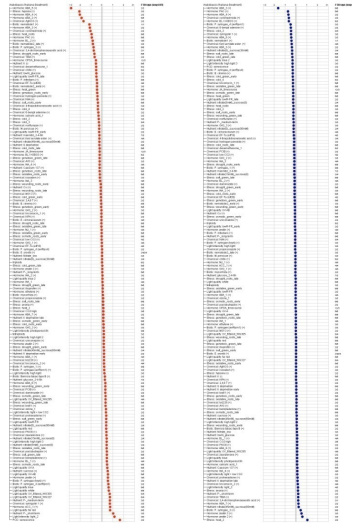
Databáze

- **Analýza exprese pomocí Genevestigator** (**AHP1** a **AHP2**, *Arabidopsis*, Affymetrix ATH 22K Array)



Databáze

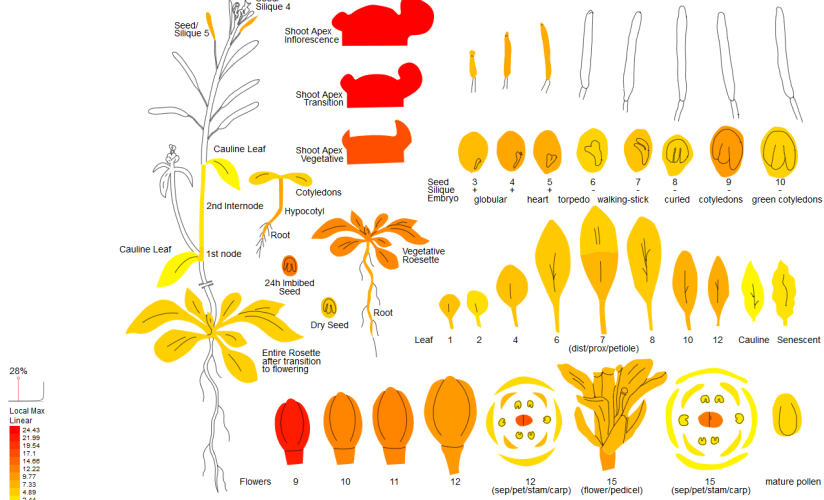
- **Analýza exprese pomocí Genevestigator** (**AHP1** a **AHP2**, *Arabidopsis*, Affymetrix ATH 22K Array)



Databáze

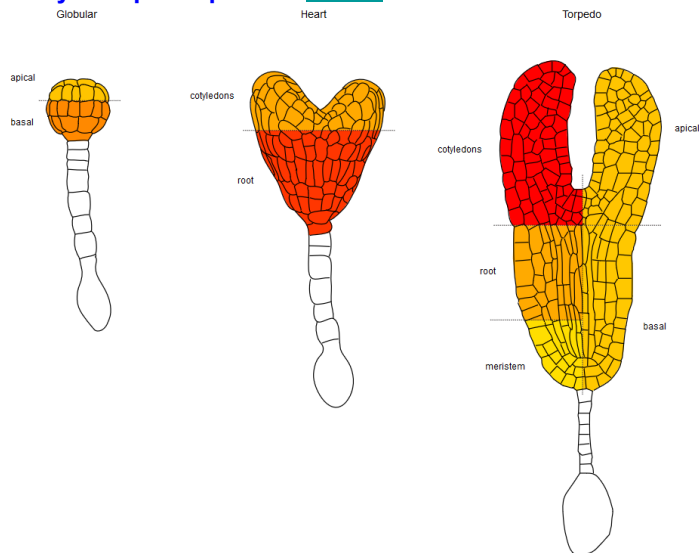
Analýza exprese pomocí ePlant

AtGenExpress eFP: AT1G13330 / AHP2, HOP2



Databáze

Analýza exprese pomocí [ePlant](#)

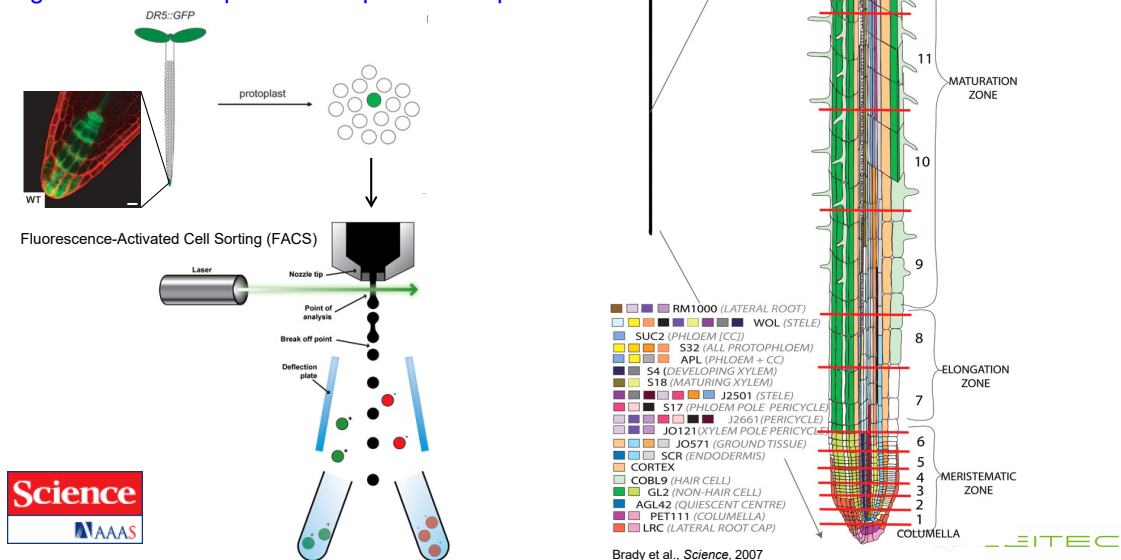


Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - **Tkáňově a buněčně specifická** analýza genové exprese

Expression Maps - RNA

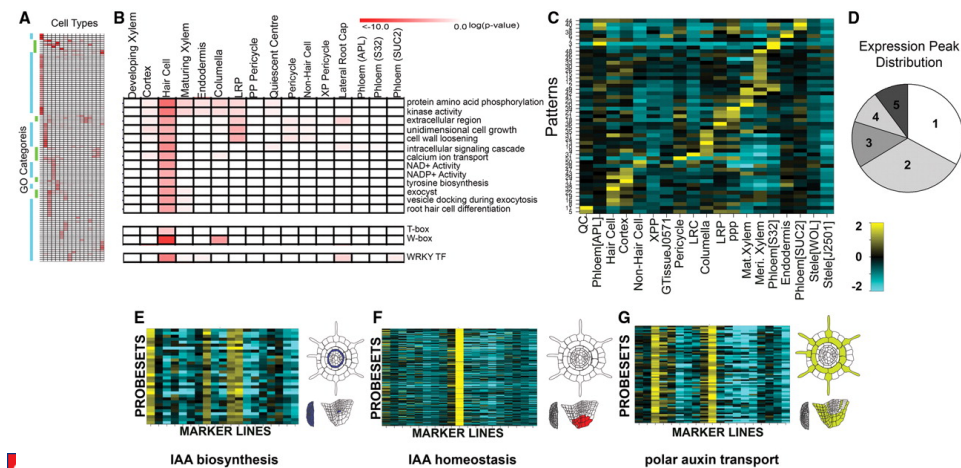
High-Resolution Expression Map in Arabidopsis Root



Microarray expression profiles of 19 fluorescently sorted GFP-marked lines were analyzed (3–9, 23, 24). The colors associated with each marker line reflect the developmental stage and cell types sampled. Thirteen transverse sections were sampled along the root's longitudinal axis (red lines) (10). CC, companion cells.

Expression Maps - RNA

High-Resolution Expression Map in Arabidopsis Root



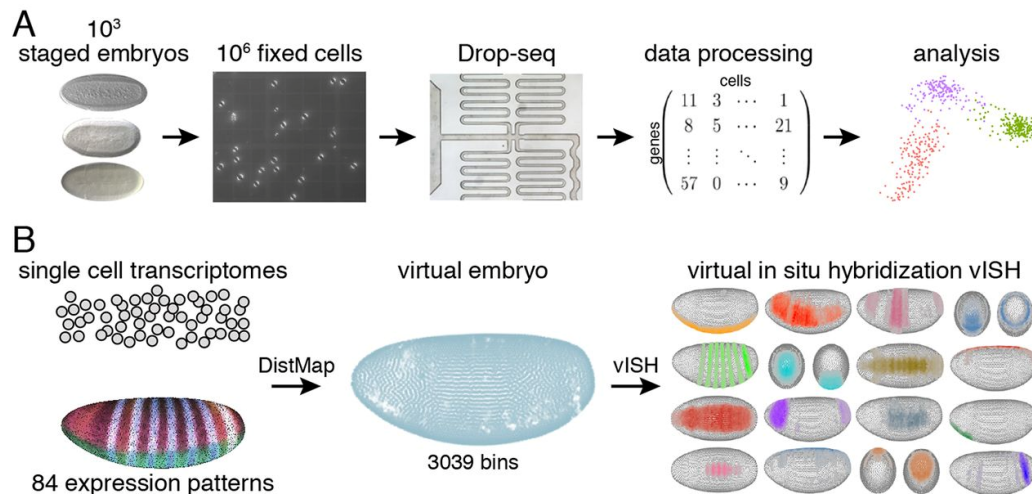
Brady et al., Science, 2007



Microarray expression profiles of 19 fluorescently sorted GFP-marked lines were analyzed (3–9, 23, 24). The colors associated with each marker line reflect the developmental stage and cell types sampled. Thirteen transverse sections were sampled along the root's longitudinal axis (red lines) (10). CC, companion cells.

Expression Maps - RNA

□ High-Resolution Expression Map in Drosophilla



24

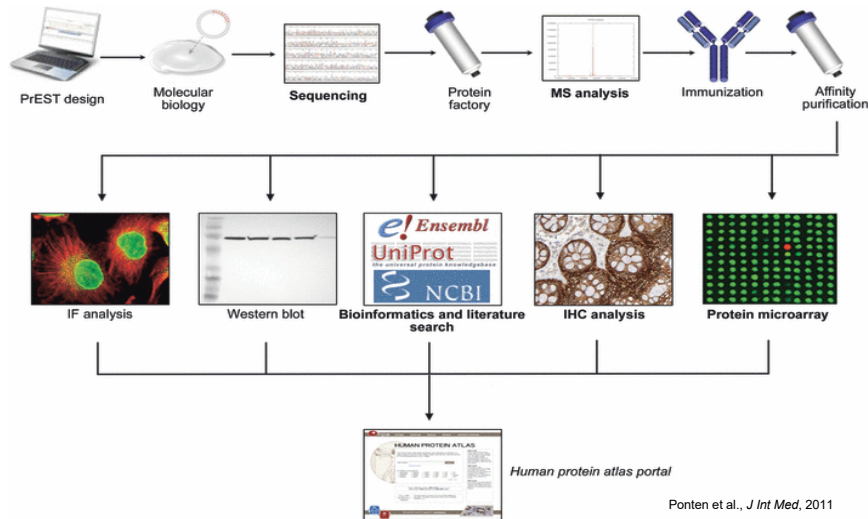
Nikos Karaïskos et al. Science 2017;science.aan3235

CEITEC

Microarray expression profiles of 19 fluorescently sorted GFP-marked lines were analyzed (3–9, 23, 24). The colors associated with each marker line reflect the developmental stage and cell types sampled. Thirteen transverse sections were sampled along the root's longitudinal axis (red lines) (10). CC, companion cells.

Expression Maps - Proteins

Human Protein Atlas



25

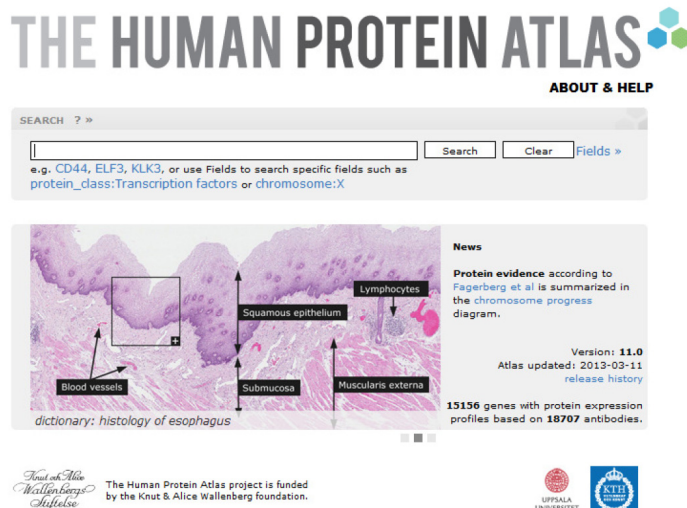
Ponten et al., J Int Med, 2011

CEITEC

Microarray expression profiles of 19 fluorescently sorted GFP-marked lines were analyzed (3–9, 23, 24). The colors associated with each marker line reflect the developmental stage and cell types sampled. Thirteen transverse sections were sampled along the root's longitudinal axis (red lines) (10). CC, companion cells.

Expression Maps - Proteins

- Human Protein Atlas (<http://www.proteinatlas.org/>)

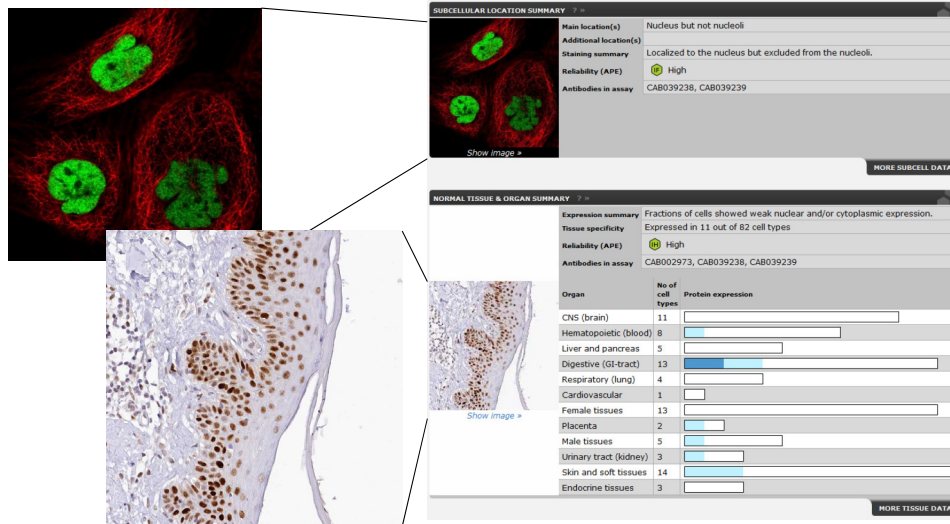


26

Microarray expression profiles of 19 fluorescently sorted GFP-marked lines were analyzed (3–9, 23, 24). The colors associated with each marker line reflect the developmental stage and cell types sampled. Thirteen transverse sections were sampled along the root's longitudinal axis (red lines) (10). CC, companion cells.

Expression Maps - Proteins

- Human Protein Atlas (<http://www.proteinatlas.org/>)



27

CEITEC

Microarray expression profiles of 19 fluorescently sorted GFP-marked lines were analyzed (3–9, 23, 24). The colors associated with each marker line reflect the developmental stage and cell types sampled. Thirteen transverse sections were sampled along the root's longitudinal axis (red lines) (10). CC, companion cells.

Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - **Tkáňové** a **buněčné specifická** analýza genové exprese
 - **Kvantitativní** analýza exprese
 - **DNA čipy**

DNA čipy

- metoda umožňující rychlé porovnání velkého množství genů/proteinů mezi testovaným vzorkem a kontrolou
- nejčastěji jsou používány oligo DNA čipy
 - k dispozici komerčně dostupné sady pro celý genom
 - firma Operon (Qiagen), 29.110 70-mer oligonukleotidů reprezentujících 26.173 genů kódujících proteiny, 28.964 transkriptů a 87 microRNA genů *Arabidopsis thaliana*
 - možnost používat pro přípravu čipů fotolitografické techniky-usnadnění syntézy oligonukleotidů např. pro celý genom člověka (cca $3,1 \times 10^9$ bp) je touto technikou možno připravit 25-mery v pouze 100 krocích

Affymetrix ATH1 *Arabidopsis* genome array



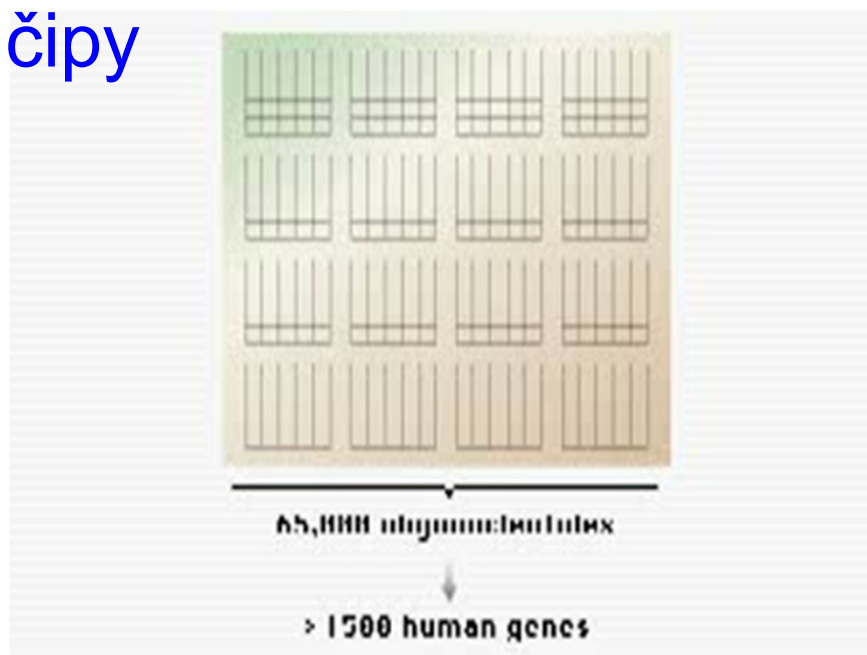
čipy nejen pro analýzu exprese, ale např. i genotypování (SNP polymorfizmy, sekvenování pomocí čipů, ...)

Critical Specifications

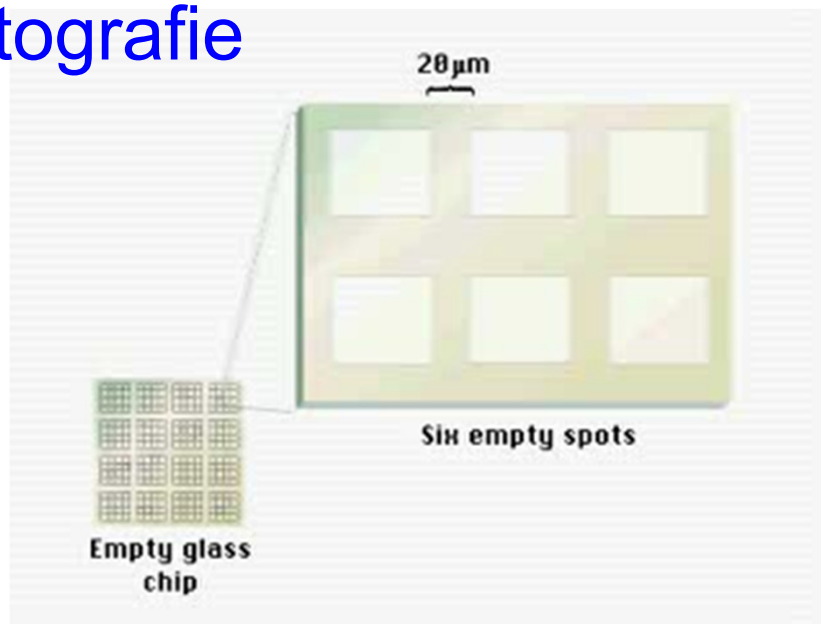
Number of arrays	One
Number of sequence represented	>24,000 gene sequences
Feature size	18 μ m
Oligonucleotide probe length	25-mer
Probe pairs/sequence	11
Control sequences	<i>E. coli</i> genes <i>bioB</i> , <i>bioC</i> , <i>bioD</i> , <i>B. subtilis</i> gene <i>lysA</i> , Phage P1 <i>cre</i> gene, <i>Arabidopsis</i> maintenance genes GAPDH, Ubiquitin, and Actin
Detection sensitivity	1:100,000*

*As measured by detection in comparative analysis between a complex target containing spiked control transcriptions and a complex target with no spikes.

DNA čipy



Fotolitografie



DNA čipy

- pro správnou interpretaci výsledků je nutná dobrá znalost pokročilých statistických metod
- je nutné zahrnout dostatečný počet kontrol i opakování

- kontrola na **přesnost měření** (opakované měření na několika čípech se stejným vzorkem, vynesení stejných vzorků analyzovaných na různých čípech proti sobě)
- kontrola **reproducibility měření** (opakované měření s různými vzorky, izolovanými za stejných podmínek na stejném čipu-stejně podmínky proti sobě)
- identifikace **hranice spolehlivého měření**
- konečně **vynesení experimentu proti kontrole** nebo **různých podmínek** proti sobě – vlastní výsledek

Expression of 195M677 in response to chemical treatment

Home | About TAIR | Sitemap | Contact | Help | Order | Login

Search | Tools | Arabidopsis Info | News | Links | FTP | Stocks

Gene

Experiment: Aluminum Stress

Experiment Summary | Samples | Slides & Datasets | Array Design | View All

Slide (name & description)	External ID	Replicate (id & name)	Reverse replicate	Sample	Experimental variables	Label	Get Data
HoekengaS7 [1] - Aluminum Stress 1 (strong spatial bias)	AFGC 7304	63: Aluminum Stress	technical	7304_Cy3.7305_Cy5	no treatment pool of 3, 6, and 24 hours	Cy3	Download
				7304_Cy5.7305_Cy3	Aluminum (50 μM AlCl ₃ pool of 3, 6, and 24 hours)	Cy5	Download
HoekengaS7 - Aluminum Stress 2 (strong spatial bias)	AFGC 7305	64: Aluminum Stress	technical	7304_Cy5.7305_Cy3	Aluminum (50 μM AlCl ₃ pool of 3, 6, and 24 hours)	Cy3	Download
				7304_Cy3.7305_Cy5	no treatment pool of 3, 6, and 24 hours	Cy5	Download

Che et al., 2002

- v současnosti je již velké množství výsledků různých experimentů lokalizovaných ve veřejně přístupných databázích

Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - **Tkáňové** a **buněčně specifická** analýza genové exprese
 - Kvantitativní analýza exprese
 - DNA a **proteinové čipy**
 - **Next gen** transkripční profilování
 - Viz přístupy systémové biologie níže

Klíčové koncepty – analýza genové exprese

- Promotory a další regulační elementy řídí genovou expresi s různou mírou časoprostorové specificity
- Genovou expresi lze studovat pomocí kvalitativních nebo kvantitativních přístupů
- Cenné údaje o specifitě genové exprese lze získat i ve veřejně dostupných databázích a to i na úrovni jednotlivých buněk

Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - **Tkáňově a buněčně specifická** analýza genové exprese
 - Kvantitativní analýza exprese
 - **DNA čipy**
 - **Next gen** transkripční profilování
 - Viz přístupy systémové biologie níže
- **Systémová biologie**
 - **Definice**

Systémová biologie - 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](#) 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](#)).

Systémová biologie - 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*).

Systemová biologie - definice

Definice dle Dr. Nathana Price,

zástupce ředitele Ústavu pro systémovou biologii, Seattle, USA; https://www.youtube.com/watch?v=OrXRI_8UFHU.



38

CEITEC

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**?

Ddli, Arnaud et al., *bioRxiv* 10.1101/2023.07.26.550726

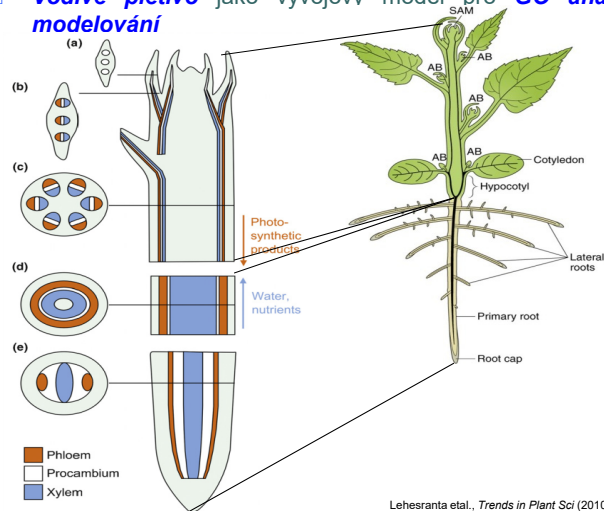
gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1.1804 1.79769e+308	1.79769e+308	08	6.88885e-05	0.00039180	1 yes
HRS1	1:4556891-4558708	WT	MT	OK	0	0.896583 1.79769e+308	1.79769e+308	08	6.61994e-06	0.00053505	5 yes
ATMLD14	1:9227472-9232296	WT	MT	OK	0	0.514809 1.79769e+308	1.79769e+308	08	9.74219e-05	0.00053505	5 yes
NRT1.6	1:9400663-9403789	WT	MT	OK	0	0.877865 1.79769e+308	1.79769e+308	08	3.2692e-08	0.00053505	5 yes
AT1G27570	1:9575425-9582376	WT	MT	OK	0	2.0829 1.79769e+308	1.79769e+308	08	9.76039e-06	6.647e-05	yes
AT1G60095	1:22159735-22162419	WT	MT	OK	0	0.688588 1.79769e+308	1.79769e+308	08	9.95901e-08	0.00053505	5 yes
AT1G03020	1:686206-686515	WT	MT	OK	0	1.78859 1.79769e+308	1.79769e+308	08	0.00913915	0.0277958	yes
AT1G13609	1:4662720-4663471	WT	MT	OK	0	3.55814 1.79769e+308	1.79769e+308	08	0.00021683	0.00108079	yes
AT1G21550	1:7553100-7553876	WT	MT	OK	0	0.562868 1.79769e+308	1.79769e+308	08	0.00115582	0.00471497	yes
AT1G22120	1:7806308-7809632	WT	MT	OK	0	0.617354 1.79769e+308	1.79769e+308	08	2.48392e-06	1.91089e-05	yes
AT1G31370	1:11238297-11239363	WT	MT	OK	0	1.46254 1.79769e+308	1.79769e+308	08	4.83523e-05	0.00028514	3 yes
APUM10	1:13253397-13255570	WT	MT	OK	0	0.581031 1.79769e+308	1.79769e+308	08	7.87855e-06	5.46603e-05	yes
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0.556525 1.79769e+308	1.79769e+308	08	6.53917e-05	0.00037473	0 yes
AT1G59077	1:21746209-21833195	WT	MT	OK	0	138.886 1.79769e+308	1.79769e+308	08	0.00122789	0.00496816	yes
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0.370087 1.79769e+308	1.79769e+308	08	0.00117993	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	0 yes
AT4G12520	4:7421055-7421738	WT	MT	OK	0.0195111	15.8516	9.68612	-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	1.49888e-12	yes
AT5G15360	5:4987235-4989162	WT	MT	OK	0.0886273	56.4834	9.1587	-10.4392	0	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

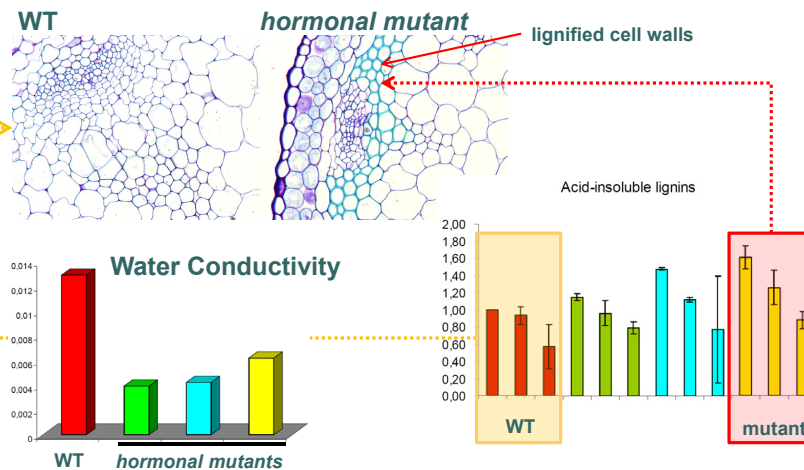
- **Vodivé pletivo** jako vývojový model pro **GO analýzu** a **MRN modelování**



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

Hormonální regulace vývoje rostlinných vodivých pletiv

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

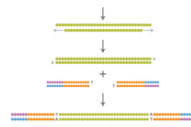
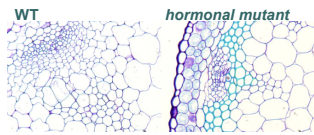


41

CEITEC

Hormonální regulace vývoje rostlinných vodivých pletiv

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



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



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

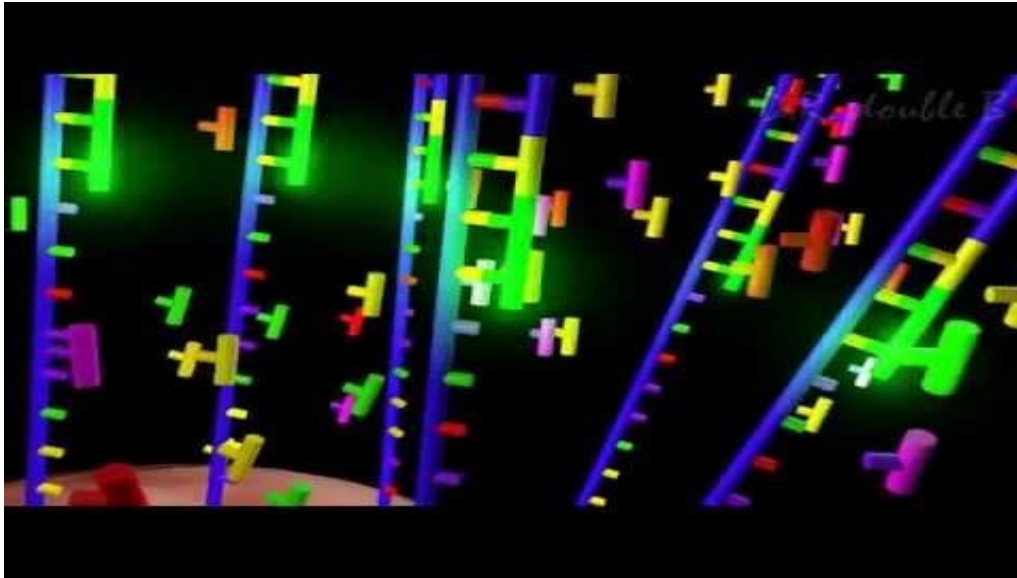


Sequencing by Synthesis
~1.5 to 11 days



CASAVA
2 days (30 min hands-on)

Sekvenování DNA metodami NGS



Original video at <https://www.youtube.com/watch?v=-7GK1HXwCtE>.

For more detailed description see e.g.

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>.

Výsledky –omických studií vs. biologicky relevantní závěry

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

Ddli, Arnaud et al., *bioRxiv* 10.1101/2023.07.26.550726

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	0	1.79769e+308	6.88885e-05	0.00039180	yes
HRS1	1:4556891-4558708	WT	MT	OK	0	0.696583 1.79769e+308	0	1.79769e+308	6.61994e-05	4.67708e-05	yes
ATML014	1:9227472-9232296	WT	MT	OK	0	0.514609 1.79769e+308	0	1.79769e+308	9.74219e-05	0.00053505	5 yes
NRT1.6	1:9400663-9403789	WT	MT	OK	0	0.877865 1.79769e+308	0	1.79769e+308	3.2692e-08	3.50131e-07	yes
AT1G27570	1:9575425-9582376	WT	MT	OK	0	2.0829 1.79769e+308	0	1.79769e+308	9.76039e-06	6.647e-05	yes
AT1G60095	1:22159735-22162419	WT	MT	OK	0	0.888588 1.79769e+308	0	1.79769e+308	9.95901e-08	9.94992e-07	yes
AT1G03020	1:696206-696515	WT	MT	OK	0	1.79859 1.79769e+308	0	1.79769e+308	0.00913915	0.0277568	yes
AT1G13809	1:4662720-4663471	WT	MT	OK	0	3.55814 1.79769e+308	0	1.79769e+308	0.00021683	0.00108079	yes
AT1G21550	1:7553100-7553876	WT	MT	OK	0	0.562868 1.79769e+308	0	1.79769e+308	0.00115582	0.00471497	yes
AT1G22120	1:7806308-7809632	WT	MT	OK	0	0.617354 1.79769e+308	0	1.79769e+308	2.48392e-05	1.91089e-05	yes
AT1G31370	1:11238297-11239363	WT	MT	OK	0	1.46254 1.79769e+308	0	1.79769e+308	4.83523e-05	0.00028514	3 yes
APUM10	1:13253397-13255570	WT	MT	OK	0	0.581031 1.79769e+308	0	1.79769e+308	7.87855e-06	5.46603e-05	yes
AT1G48700	1:18010728-18012871	WT	MT	OK	0	0.556525 1.79769e+308	0	1.79769e+308	6.53917e-05	0.00037473	6 yes
AT1G59077	1:21746209-21833195	WT	MT	OK	0	138.886 1.79769e+308	0	1.79769e+308	0.00122789	0.00496816	yes
AT1G60050	1:22121549-22123702	WT	MT	OK	0	0.370087 1.79769e+308	0	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



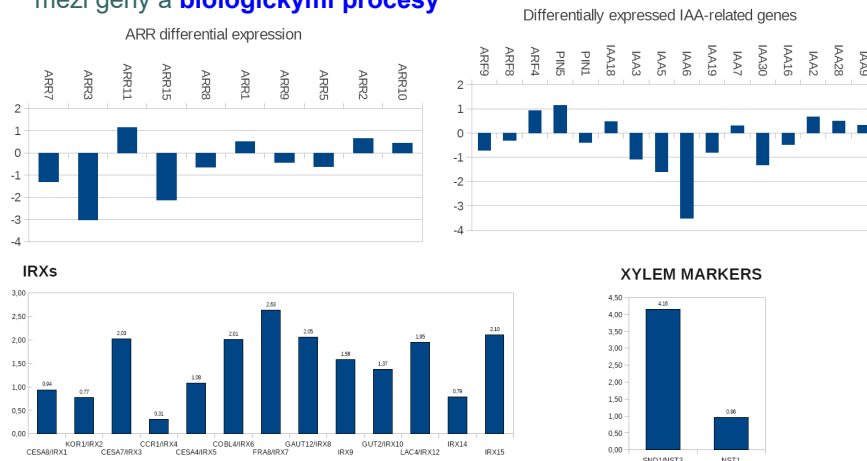
Example of an output of transcriptional profiling study using Illumina sequencing performed in our lab. Shown is just a tiny fragment of the complete list, comprising about 7K genes revealing differential expression in the studied mutant.

Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - Tkáňově a **buněčně specifická** analýza genové exprese
 - Kvantitativní analýza exprese
 - DNA čipy
 - Next gen transkripční profilování
- **Systémová biologie**
 - Definice
 - **Nástroje**
 - **Genová ontologie**

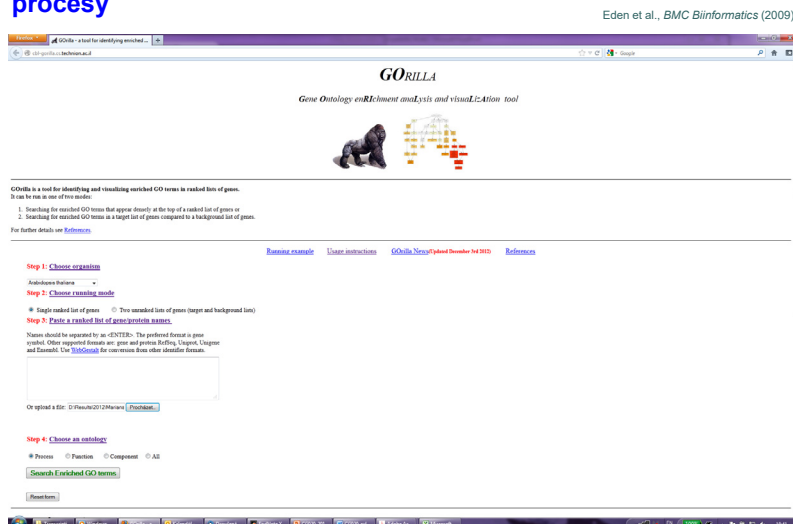
Analýza genové ontologie

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



Analýza genové ontologie

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

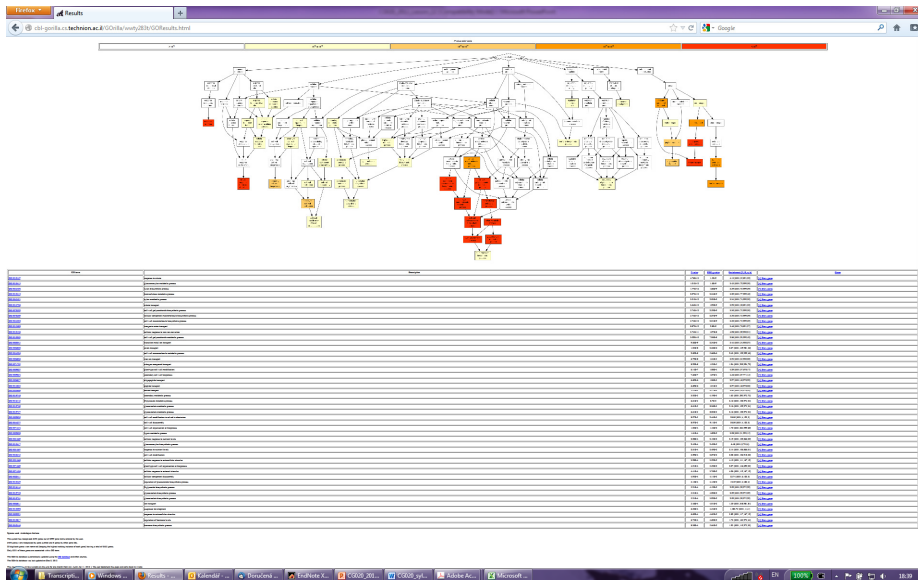


47

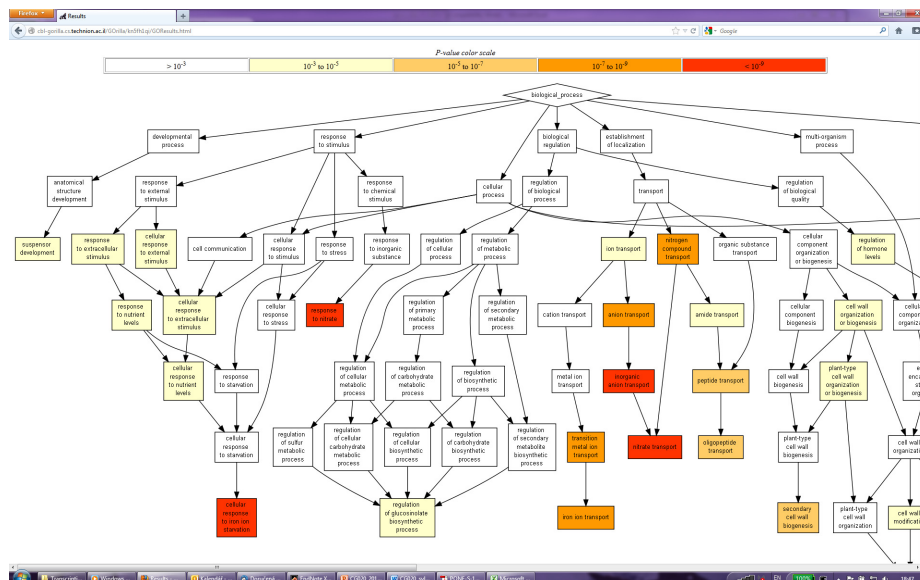
CEITEC

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

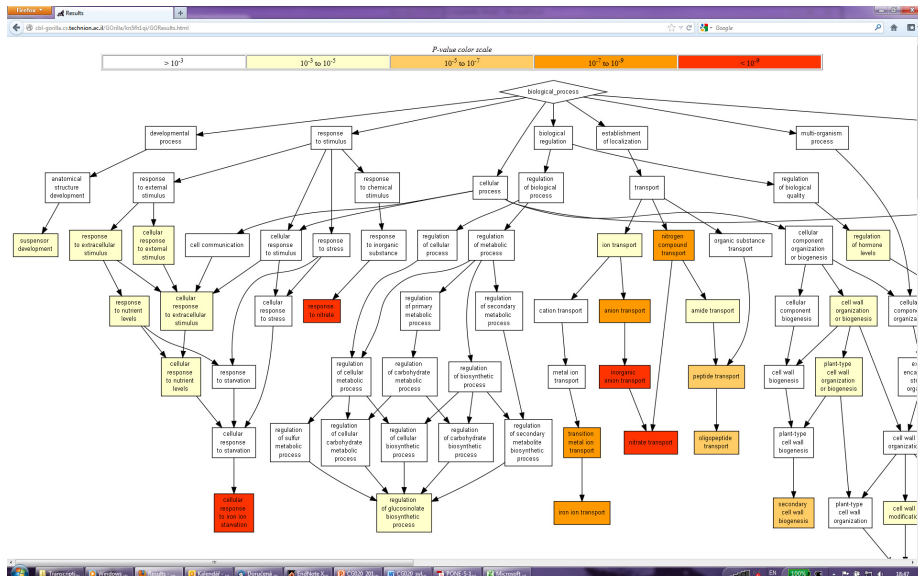
GORILLA



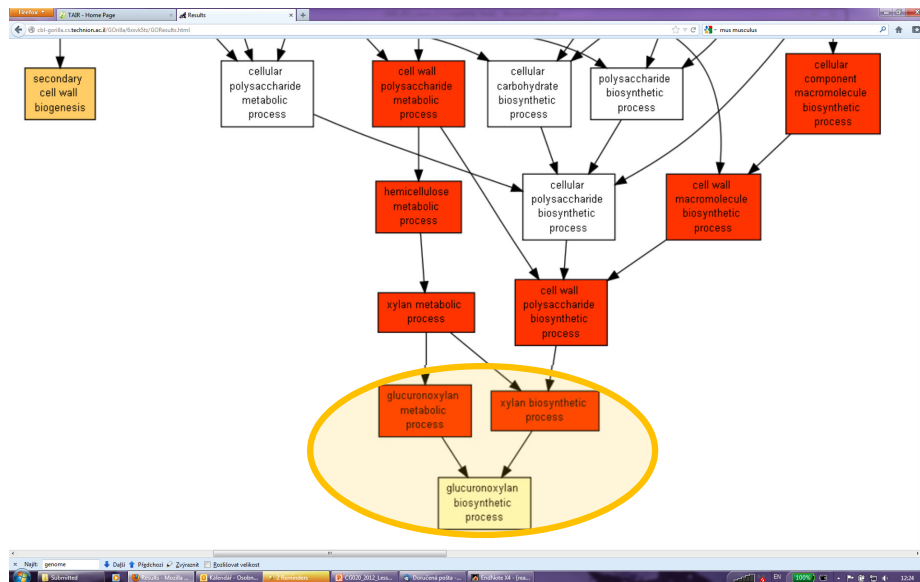
GORILLA



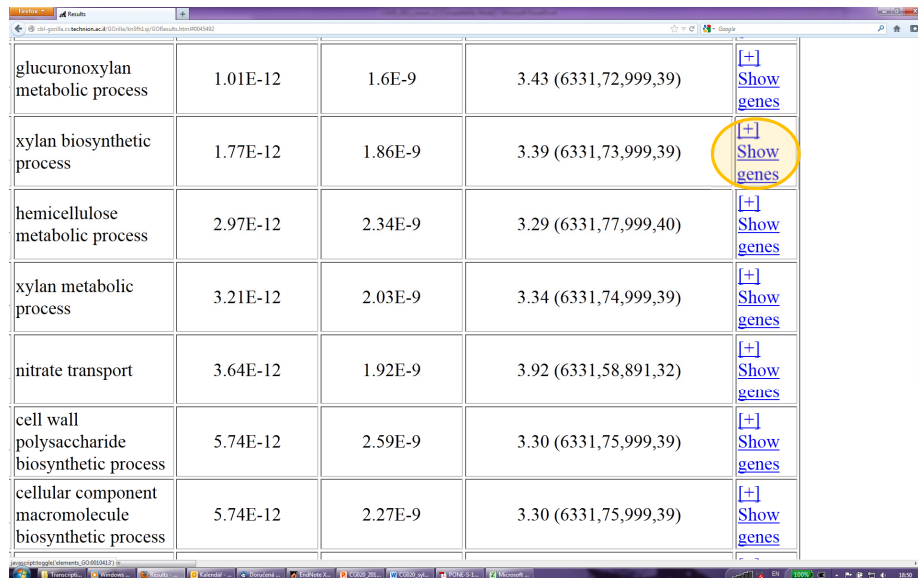
GORILLA



GORILLA



GORILLA



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

GORILLA

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 PGBP3 - plant glycogen-like starch initiation protein 3 FRA8 - oxostom-like protein GAUT12 - alpha-1,4-galacturonosyltransferase AT4G22460 - functional subunit of lipid transfer protein used storage 2a albumin-like protein AT5G42180 - peroxidase 64 AT4G10910 - ring-h2 finger protein atf72 LAC17 - lactase 17 KNA17 - homodimer protein located 1-like 7 NAC012 - nac domain-containing protein 12 RXN - nucleoside-diphosphate-sugar transferases-like protein AT1G70500 - protein lyase-like protein CESA4 - cellulose synthase a catalytic subunit 4 [udp-forming] AT1G08340 - rho gtpase activating protein with pak-box p21-binding domain CTL2 - chitinase-like protein 2 RX6 - cohen-like protein 4 MYB63 - myb domain protein 63 PGBP1 - plant glycogen-like starch initiation protein 1 AT5G46340 - putative o-acetyltransferase AT3G21710 - hypothetical protein AT2G03200 - aspartyl protease-like protein AT1G09440 - protein kinase family protein AT5G46020 - pathogenesis-related thumatin-like protein AT3G21090 - targeting protein for xcp2-like protein AT5G66210 - hypothetical protein AT1G56230 - htp-pec domain-containing protein AT2G19190 - hypothetical protein JP630 - putative polygalacturonase non-catalytic subunit jp630 MAP70-5 - microtubule-associated proteins 70-5 AT5G50220 - hypothetical protein AGL44 - protein agnosia-like 44 RX12 - lactase-4 NAC073 - nac domain containing protein 73 RX1 - cellulose synthase a catalytic subunit 7 [udp-forming] AT4G27435 - hypothetical protein MYB46 - transcription factor myb46 AT1G72220 - ring-h2 finger protein atf54 FRD1 - nate efflux family protein AT1G33800 - hypothetical protein
hemiacetalase metabolic process	2.07E-12	2.24E-9	2.20 (6321,77,000,40)	[+] Show genes

Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - Tkáňové a **buněčně specifická** analýza genové exprese
 - Kvantitativní analýza exprese
 - DNA čipy
 - Next gen transkripční profilování
- **Systémová biologie**
 - Definice
 - Nástroje
 - Genová ontologie
 - **Bayesovské sítě**

Bayesovské sítě

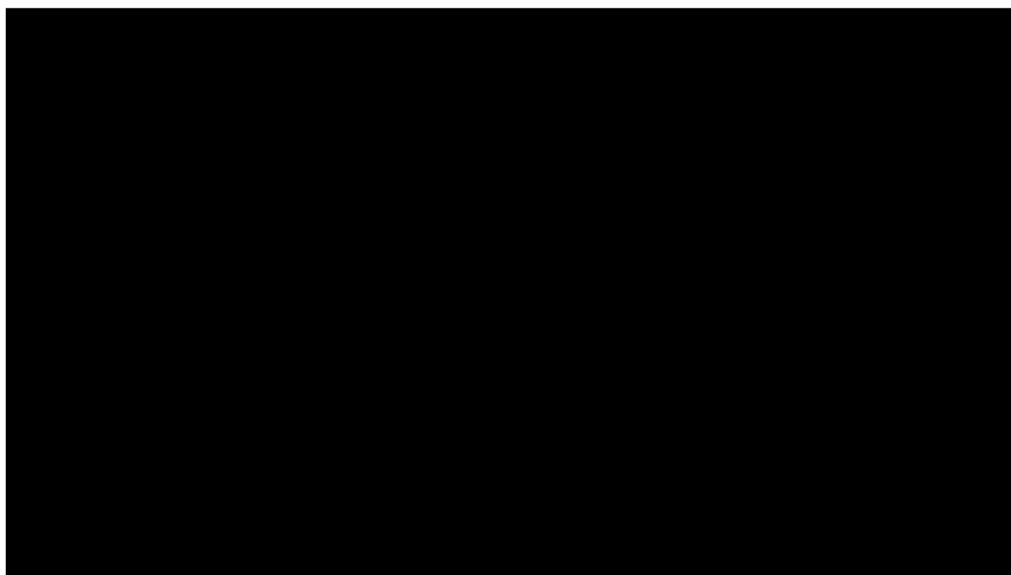
Co je Bayesovská síť?

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



Bayesovské sítě

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



Bayesovské sítě

Co je Bayesovská síť?

- Praviděpodobný 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

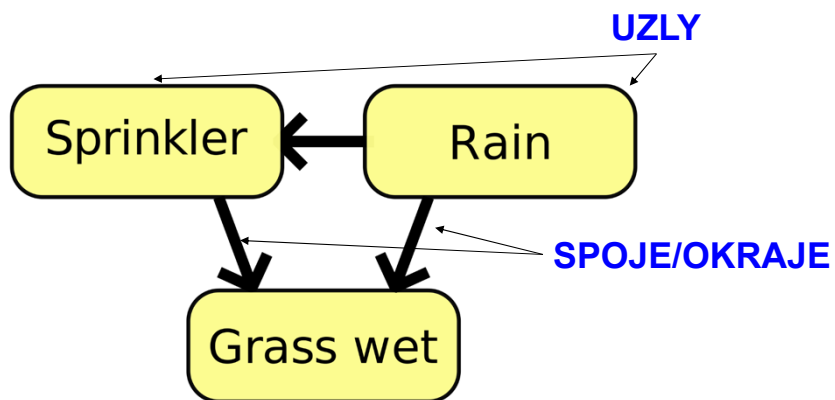
▪ UZLY

- 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



Asijská Bayesovská síť



<https://www.bayesserver.com/>

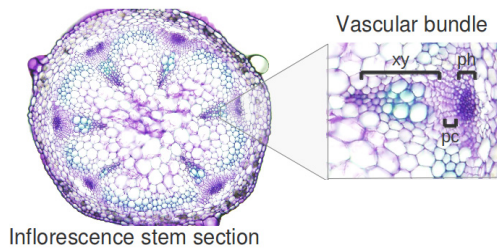
CEITEC

Osnova

- Struktura genů
- Metody analýzy **genové exprese**
 - Kvalitativní analýza exprese genů
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných dat ve **veřejných databázích**
 - **Tkáňově a buněčně specifická** analýza genové exprese
 - Kvantitativní analýza exprese
 - DNA čipy
 - Next gen transkripční profilování
- **Systémová biologie**
 - Definice
 - Nástroje
 - Genová ontologie
 - Bayesovské sítě
 - Modelování **molekulárních/genových regulačních sítí**

Modelování molekulárních regulačních sítí

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



Benitez and Hejatk, *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 ARR 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]

Modelování molekulárních regulačních sítí

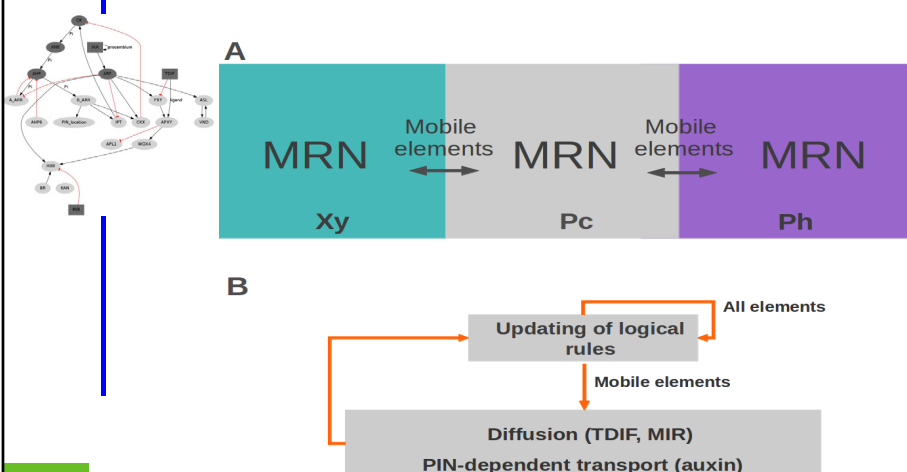
- Formulace *logických pravidel* definujících *dynamiku modelu*

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 Hejatk, *PLoS One*, 2013

Modelování molekulárních regulačních sítí

- Specifikace *mobilních prvků* a jejich chování v modelu



64

CEITEC

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:

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

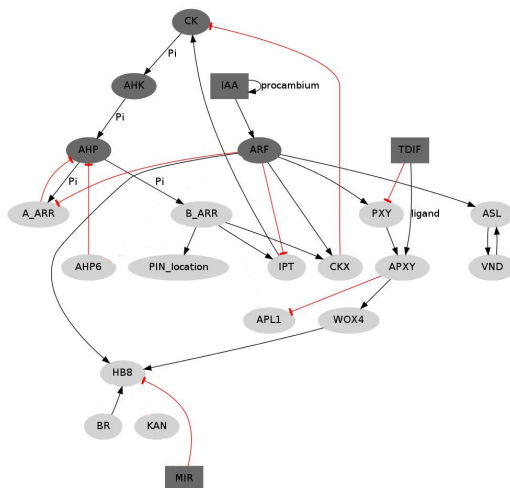
where $g(t)T[i]$ is the total amount of TDIF or MIR165 in cell (i). D is a parameter that determines the proportion of g that can move from any cell to neighboring ones and is correlated to the diffusion rate of g . b is a constant corresponding to a degradation term. H is a step function that converts 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*



65

CEITEC

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.

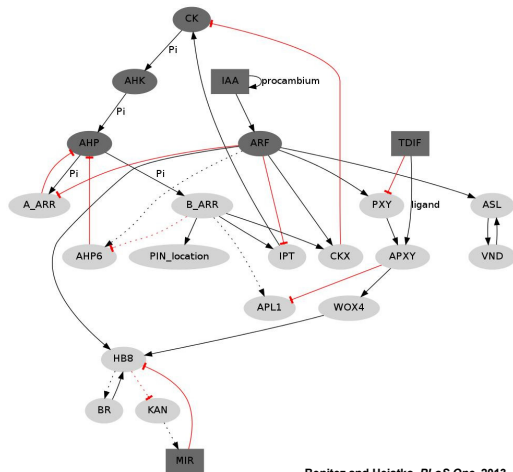
Modelování molekulárních regulačních sítí

- 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 <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, ExpressionSet:1005823559, [22])

Modelování molekulárních regulačních sítí

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



67

Benitez and Hejatk, PLoS One, 2013

CEITEC

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

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



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

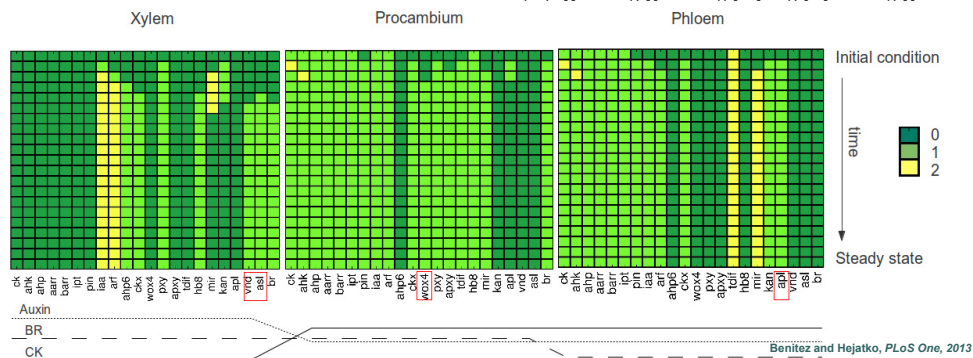


Modelování molekulárních regulačních sítí

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

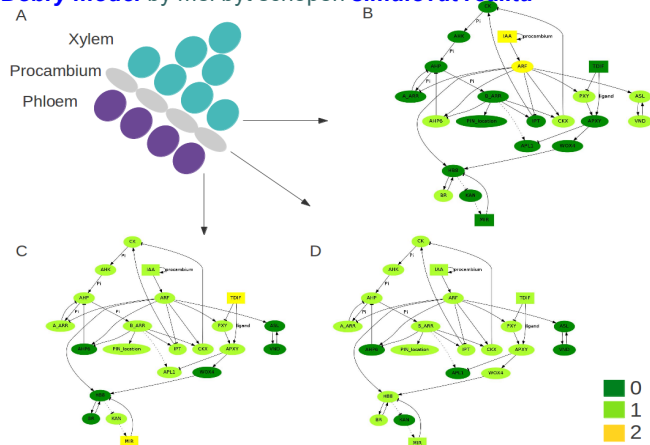
Static nodes: $g_n(t+1) = F_n(g_{n1}(t), g_{n2}(t), \dots, g_{nk}(t))$

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



Modelování molekulárních regulačních sítí

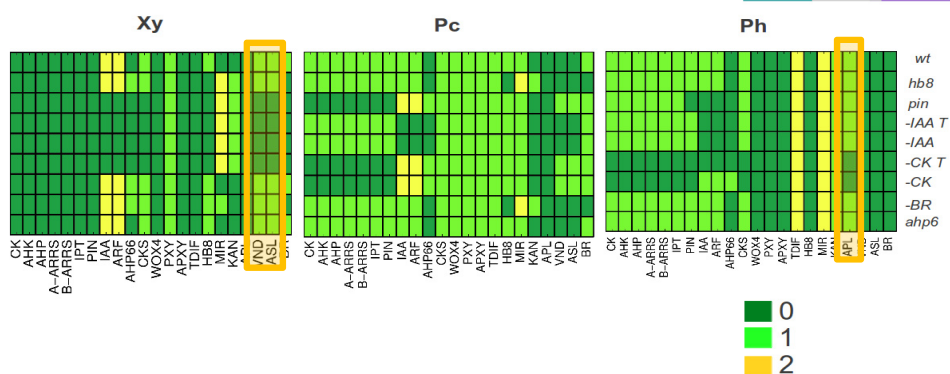
- **Dobry model** by měl být schopen **simulovat realitu**



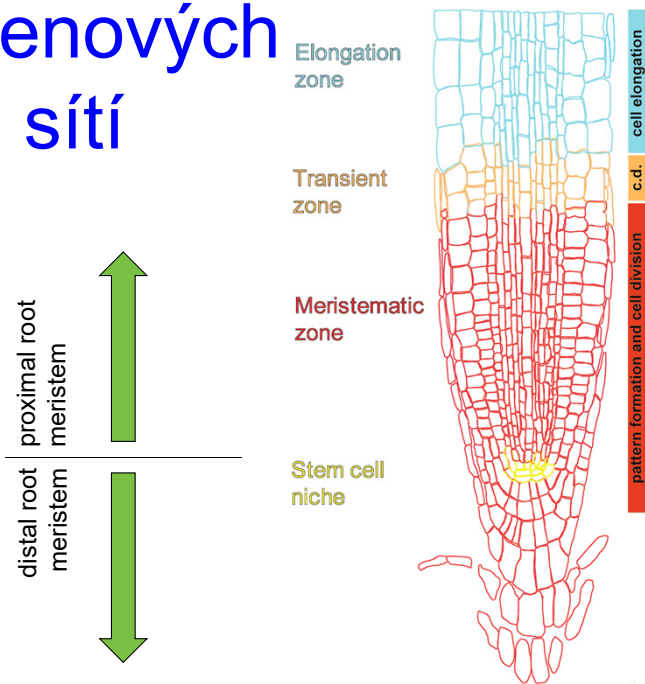
Benitez and Hejatk, submitted

Modelování molekulárních regulačních sítí

- Simulace *mutantů*



Odvození genových regulačních sítí

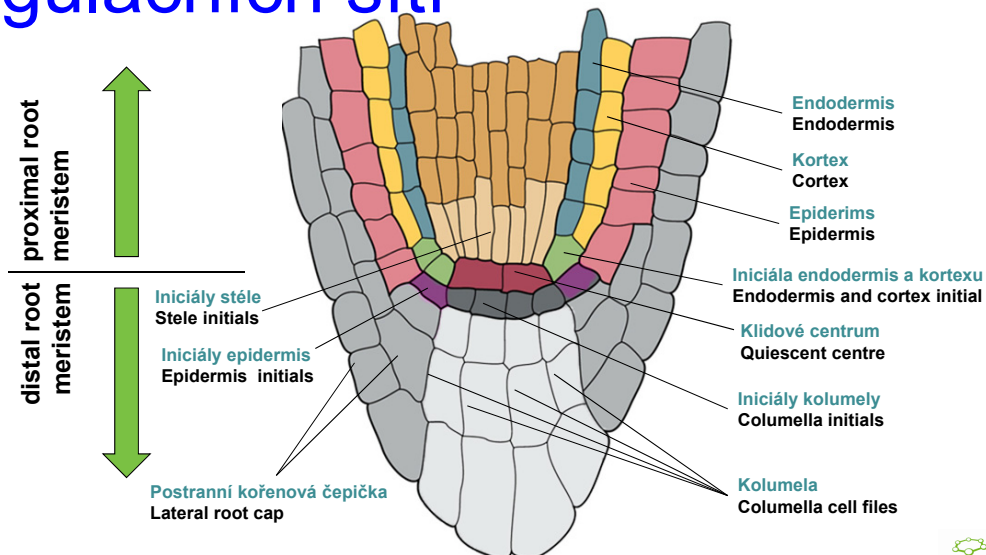


73

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

CEITEC

Odvození genových regulačních sítí



74

CEITEC

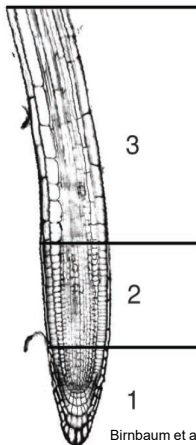
In the root, several functional and anatomical units could be recognized.

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.

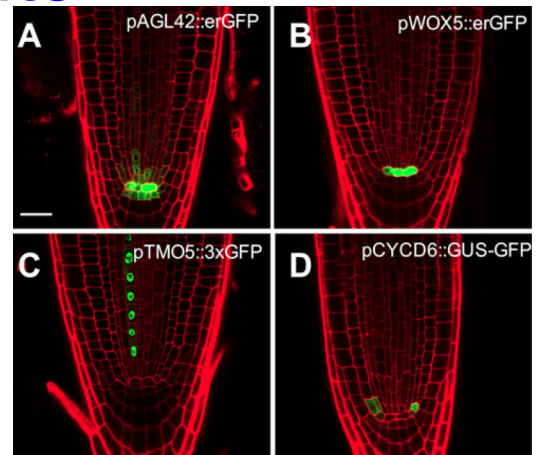
Genové regulační sítě - GENIST

- Odvození GR sítí přes **GENIST**

- **GE**ne regulatory **NE**twork **I**nference from **S**patio**T**emporal data algorithm
- Kombinace prostorových- a časově- specifických profilů exprese genů



Birnbaum et al., *Science*, 2003



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

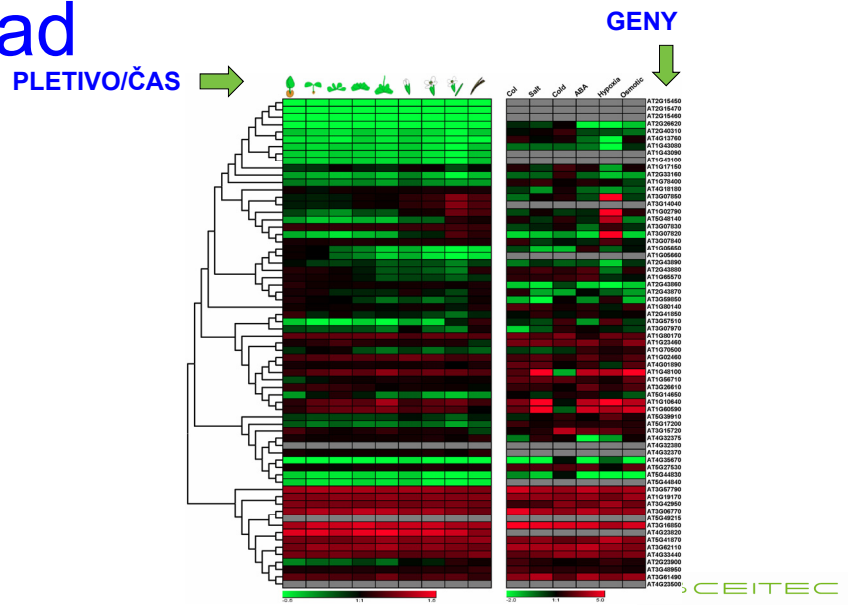
75

CEITEC

Specific subpopulations from the stem cell niche (SCN) were isolated via protoplasting the root (removing the cell wall enzymatically allowing to release the individual cells) of several specific reporter lines (A-D in the figure on the right) and GFP-positive cells were isolated using cell sorter. The mRNA was isolated and transcriptional; profiling via NGS was performed.

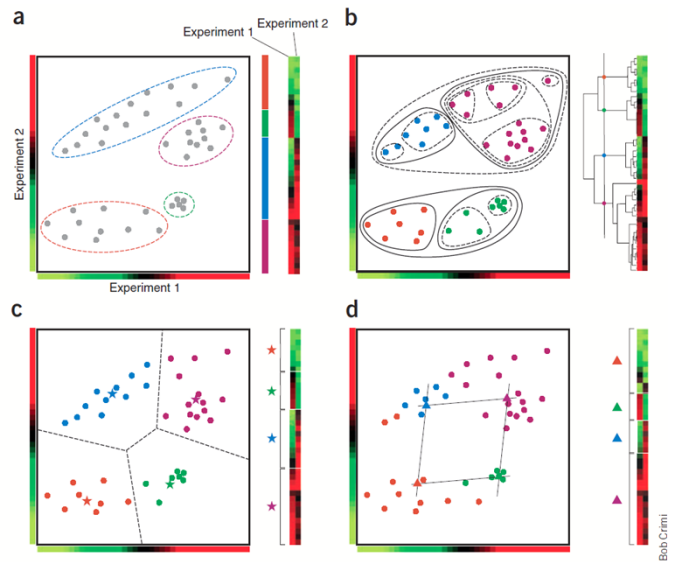
By comparing the cell type-specific transcriptomes with developmental-specific root transcriptomes (isolating mRNA from meristematic (1, the figure on the left), elongation (2) and differentiation (3) zones, the stem cell-specific transcriptomes were identified.

Kombinace velkých omických datových sad



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**



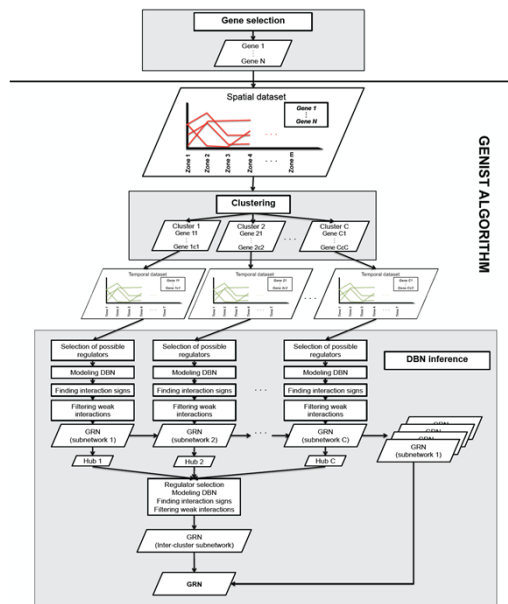
GENIST algorithm

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** potenciálních **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

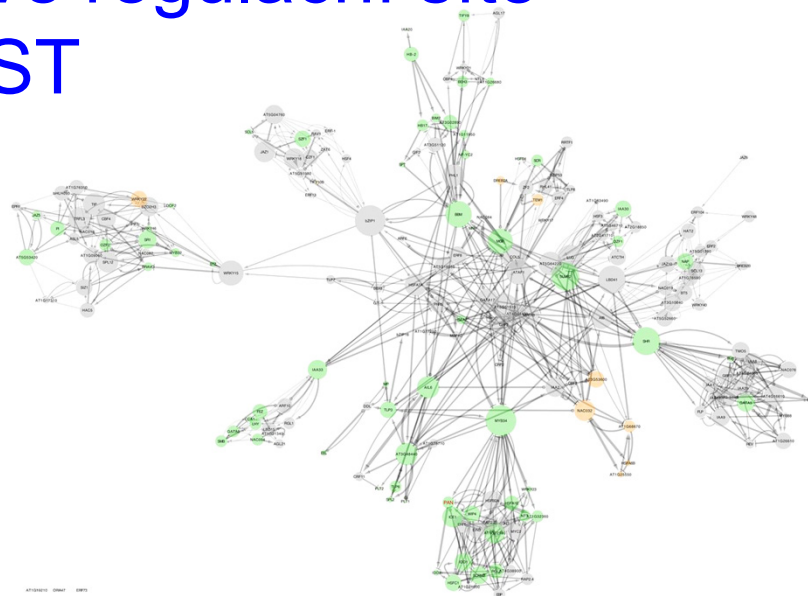
CEITEC

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

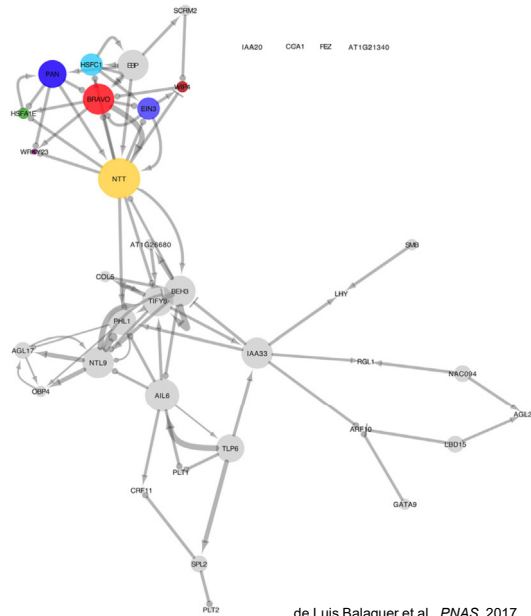


79

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

Network of the 201 TFs enriched in the SCN, inferred with the 12 developmental time points of the Arabidopsis root. Clusters of nodes indicate groups of TFs functionally related or functioning in the same cell type. Node sizes indicate importance of the nodes in terms of the number of TFs that they regulate. The highly connected groups of genes or subnetworks correspond to the dynamic Bayesian network (DBN) inferred for each cluster. Green (orange) nodes represent factors that are differentially down-regulated (up-regulated) in the pan mutant with respect to Col-0 wild type. Blue represents the PAN node.

Genové regulační sítě - GENIST

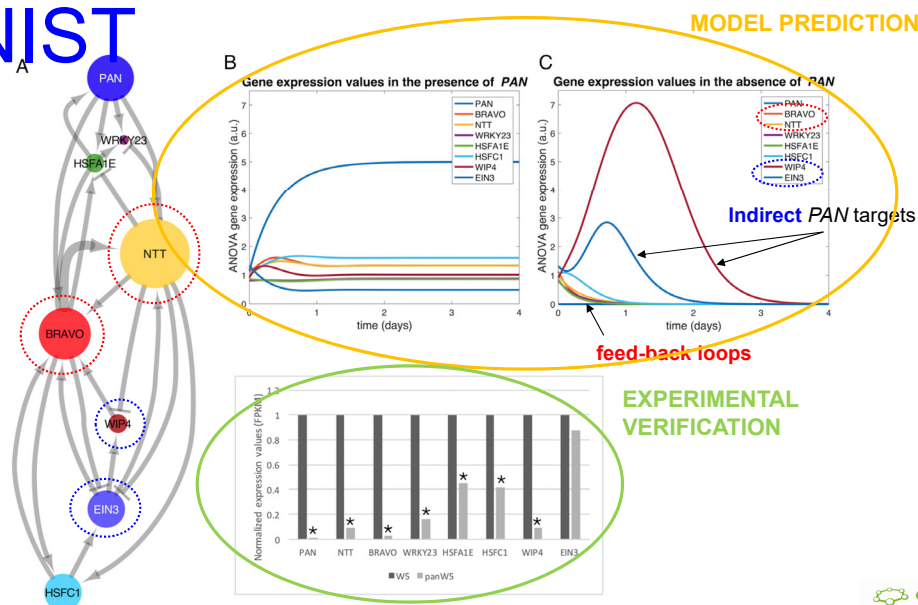


de Luis Balaguer et al., *PNAS*, 2017

CEITEC

Network of QC-enriched TFs. (A) Network among the QC-enriched TFs inferred with the 12 developmental time points of the Arabidopsis root. Node sizes indicate importance of the nodes in terms of the number of TFs that they regulate. Color-coded nodes represent genes downstream of PERIANTHA (PAN) that were used for the mathematical model and experimental confirmations.

Genové regulační sítě - GENIST



81

CEITEC

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.

With the exception of indirect target EIN3, the qRT PCR-based gene expression quantification confirmed the predicted expression values.

Klíčové koncepty – systémová biologie

- 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 učení („umělá inteligence“)