

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

hejatko@sci.muni.cz, www.ceitec.eu

M U N I
S C I



Zdrojová literatura

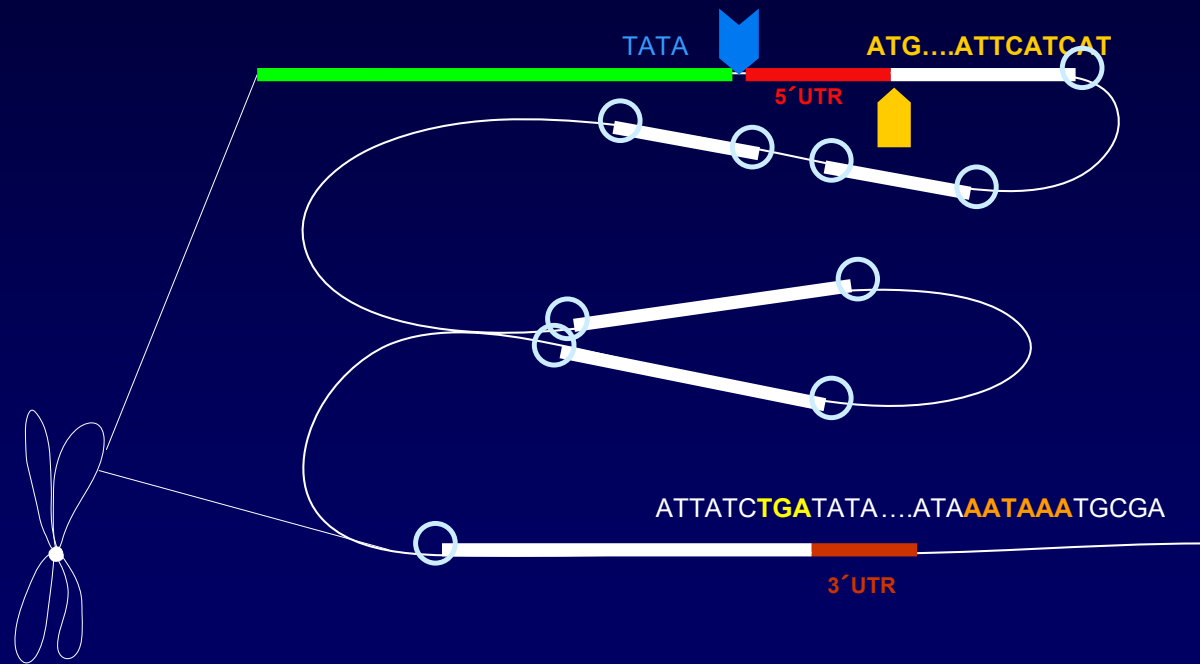
- Benitez, M. and Hejatko, J. Dynamics of cell-fate determination and patterning in the vascular bundles of *Arabidopsis thaliana* (submitted)
- Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. 2007. A high-resolution root spatiotemporal map reveals dominant expression 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.
- Karaiskos 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

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

Struktura genů

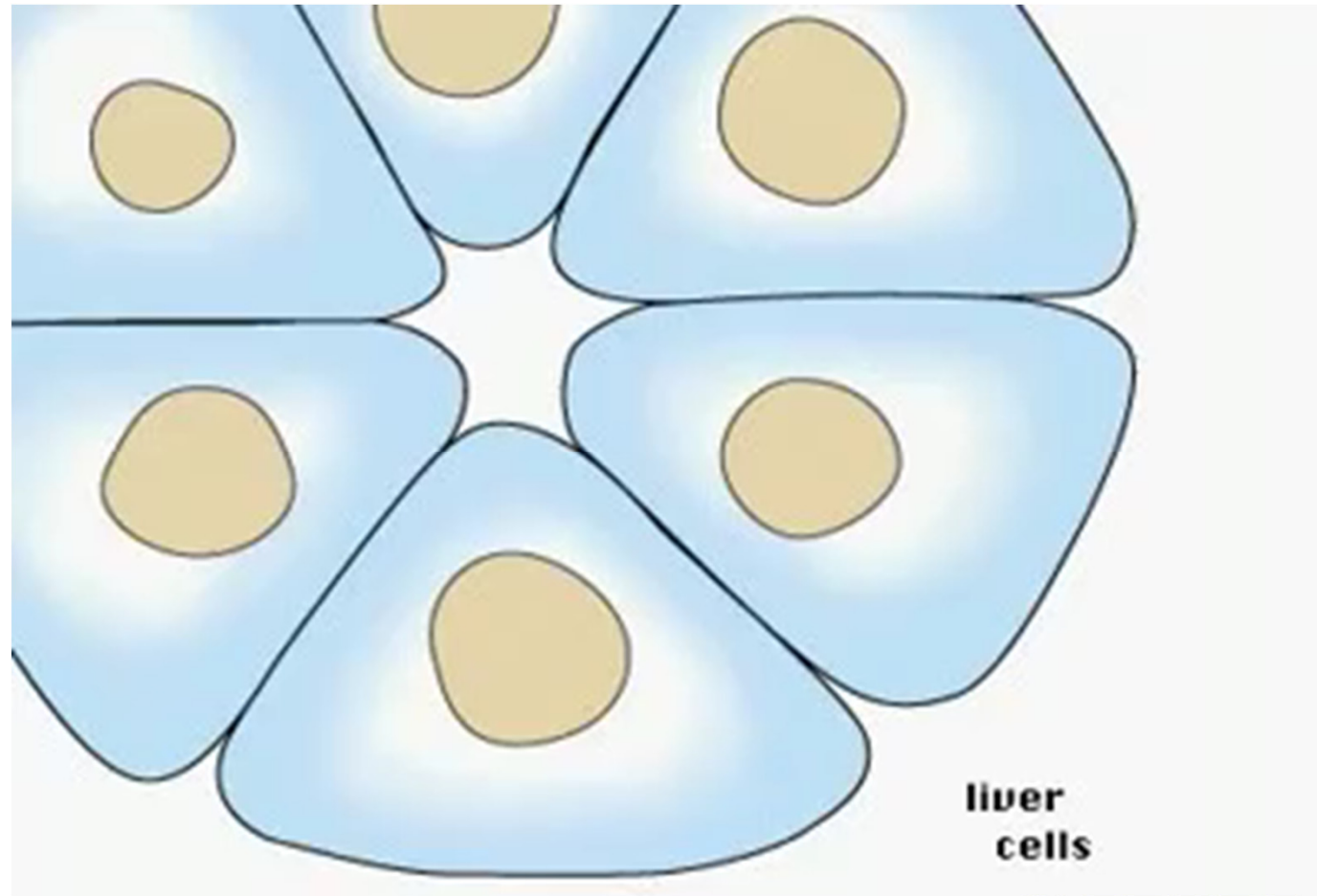
- promotor
- počátek transkripce
- 5' UTR
- počátek translace
- místa sestřihu
- stop kodon
- 3' UTR
- polyadenylační signál



Osnova

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

Časoprostorová specifita genové exprese



Alberts et al., Molecular Biology of the Cell

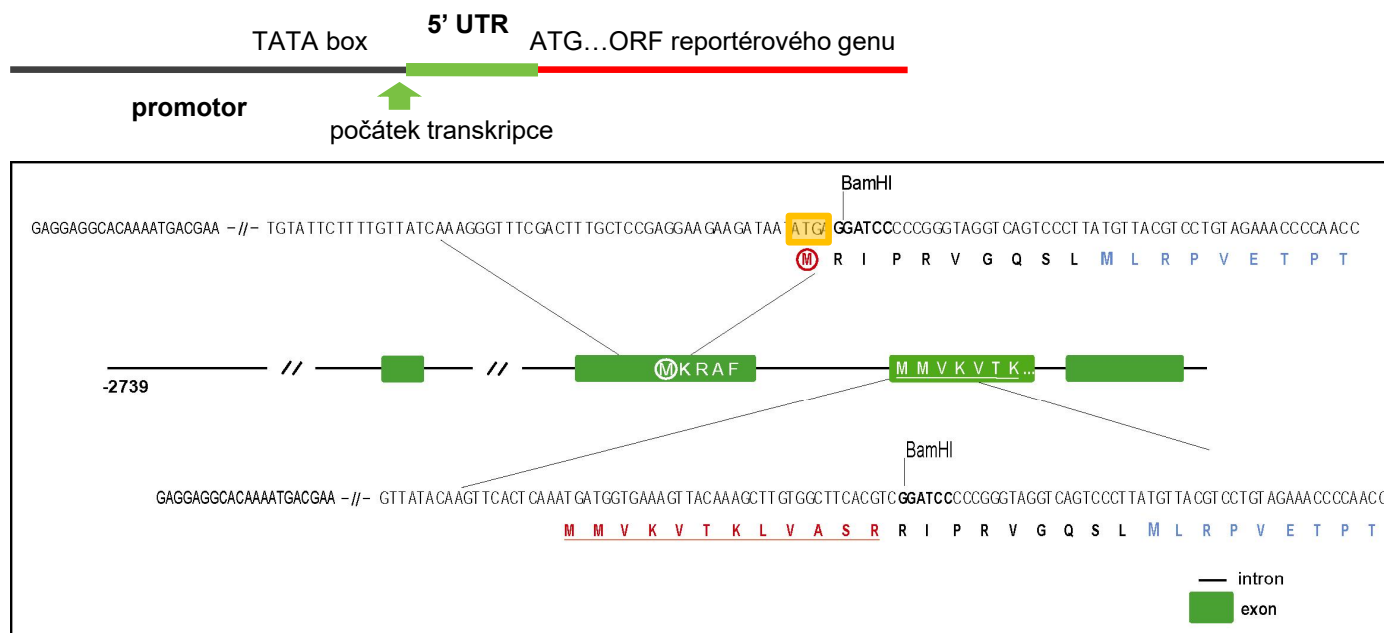
CEITEC

Osnova

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

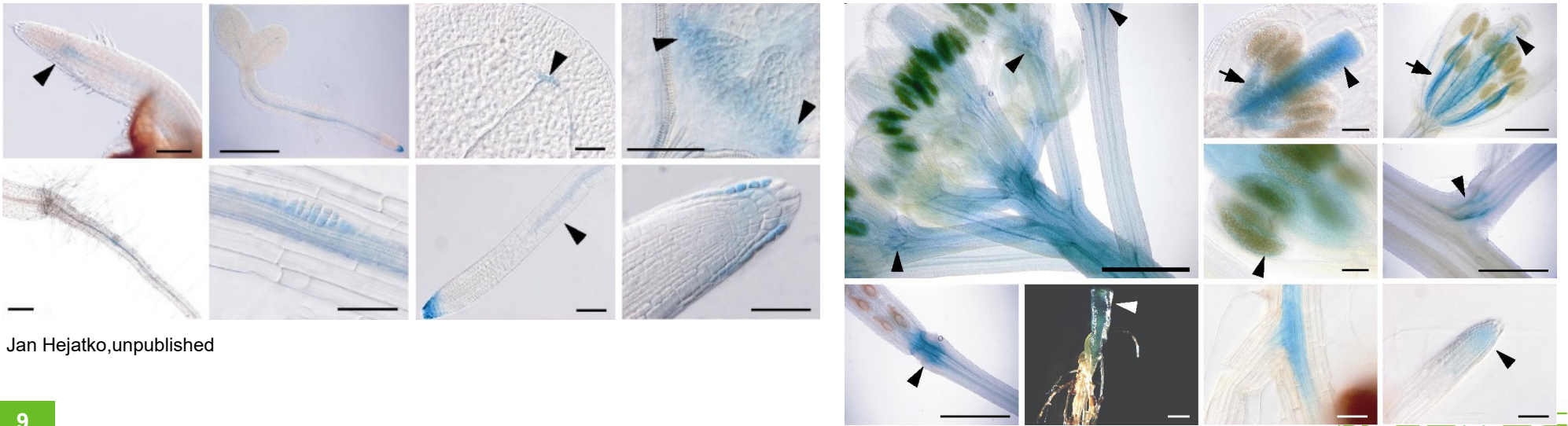
Transkripční fúze

- **Příprava transkripční fúze promotoru s reportérem**
 - Identifikace a klonování promotorové oblasti genu
 - příprava rekombinantní DNA nesoucí promotor a reportérový gen (uidA, GFP)

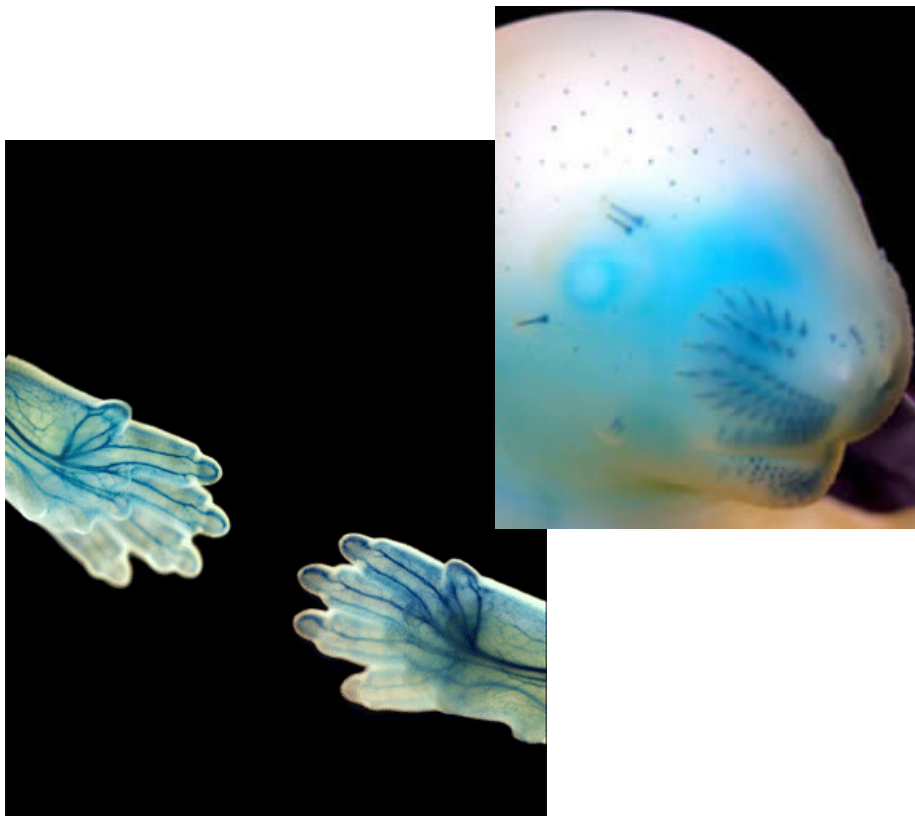


Transkripční fúze

- **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

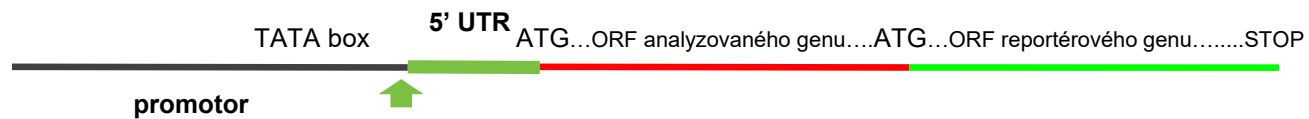


Osnova

- 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

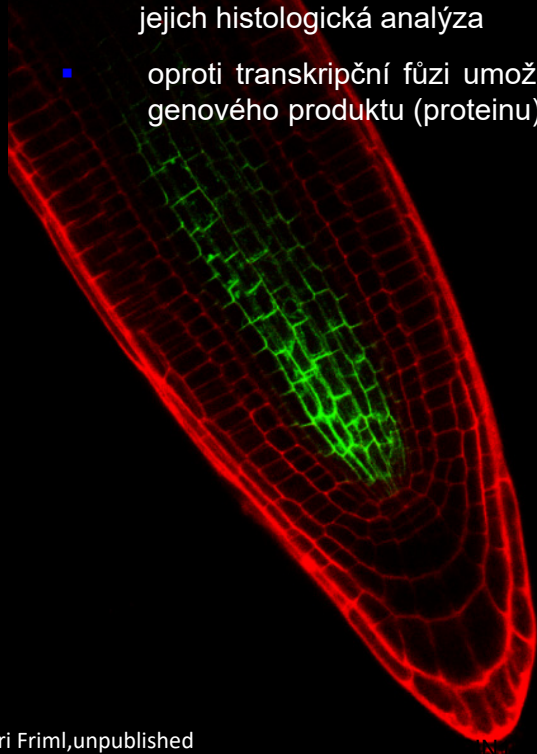
- **Příprava translační fúze kódující oblasti analyzovaného genu s reportérovým genem**
 - Identifikace a klonování **promotorové** a **kódující** oblasti analyzovaného genu
 - příprava **rekombinantní DNA** nesoucí **promotor** a **kódující** sekvenci studovaného genu ve fúzi s **reportérovým genem**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



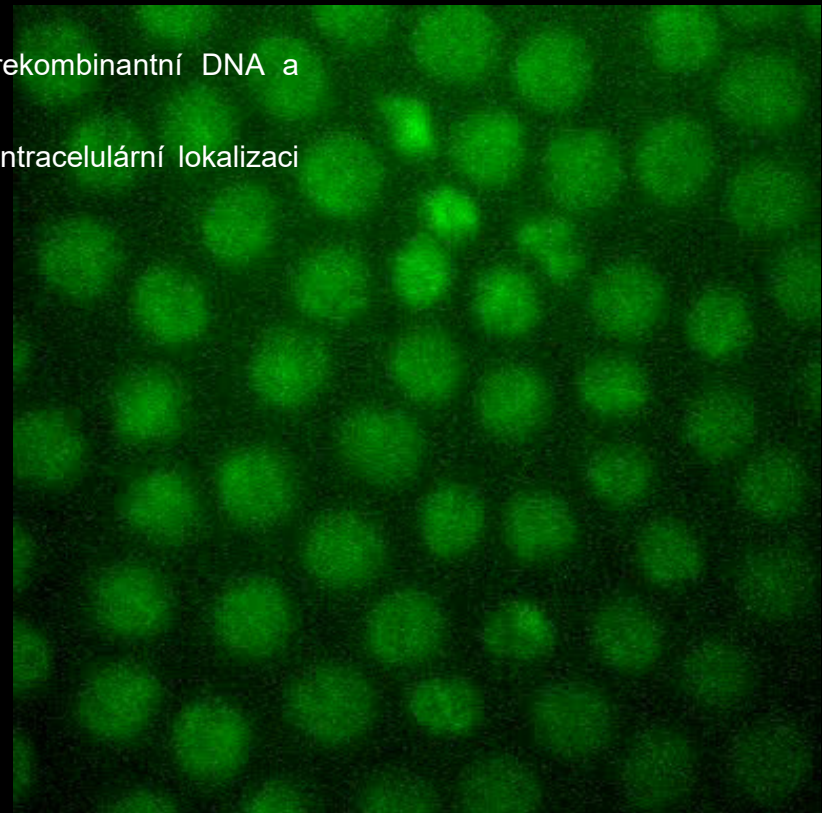
Translační fúze

□ 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



Jiri Friml, unpublished



EMBO GFP Course, 2000

Translační fúze

- Translační fúze kódující oblasti analyzovaného genu s repotérovým genem

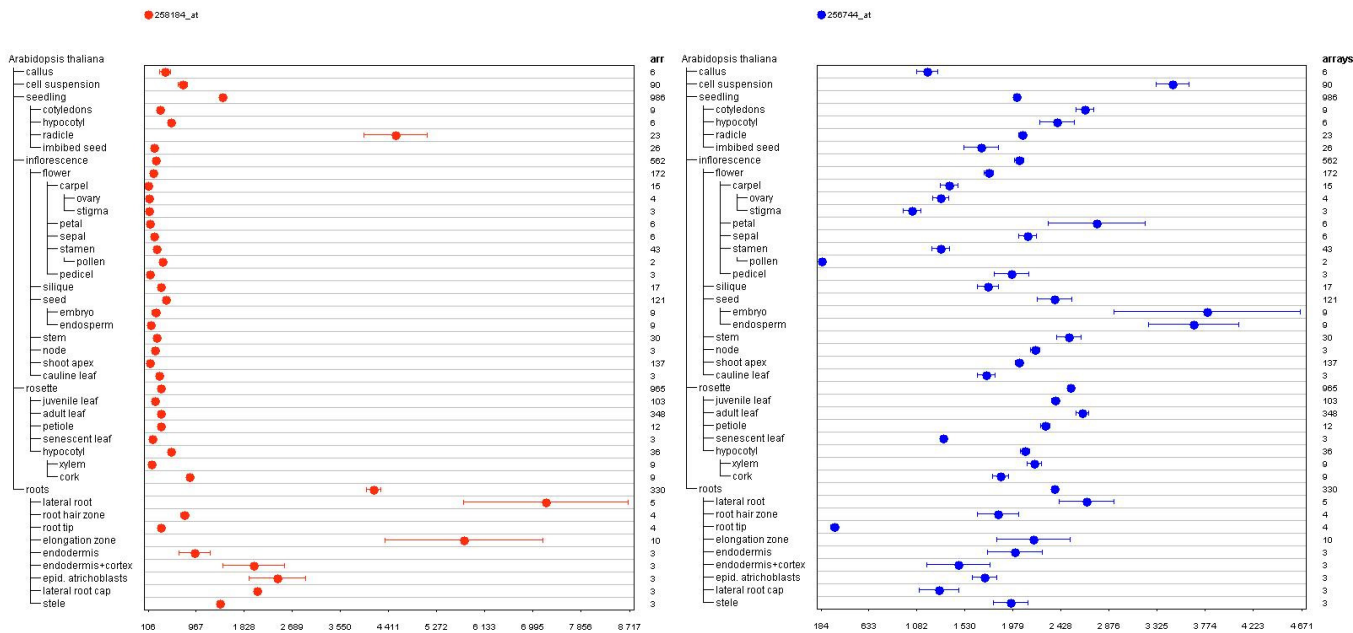


Osnova

- 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

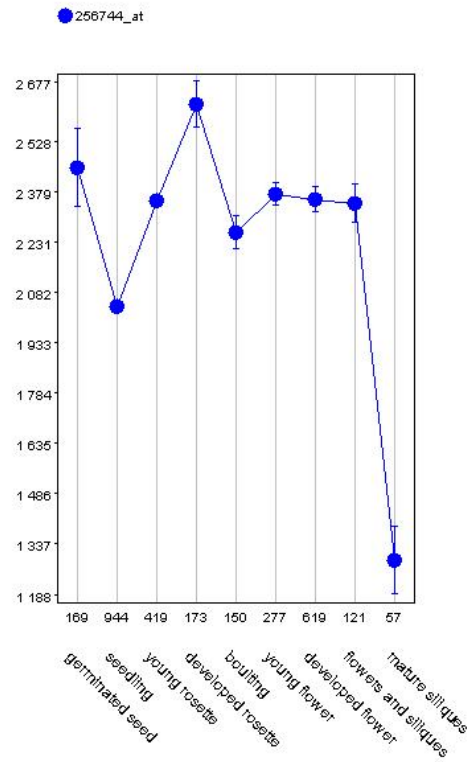
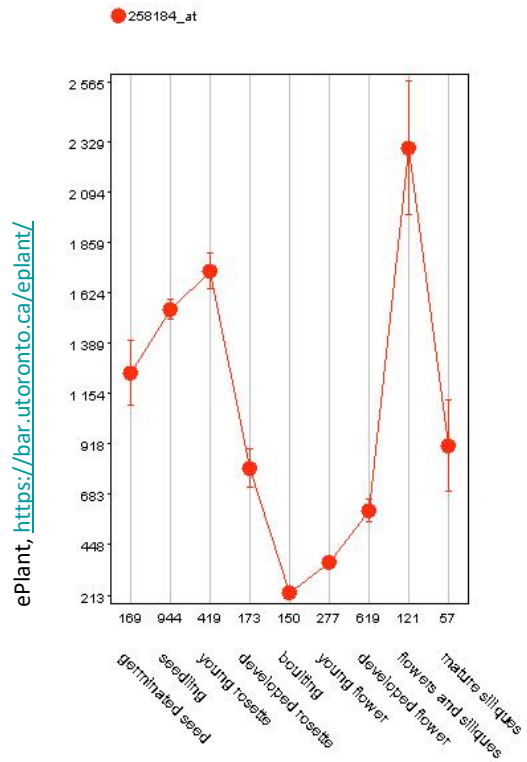
- **Analýza exprese pomocí Genevestigator (AHP1 a AHP2, Arabidopsis, Affymetrix ATH 22K Array)**



ePlant, <https://bar.utoronto.ca/eplant/>

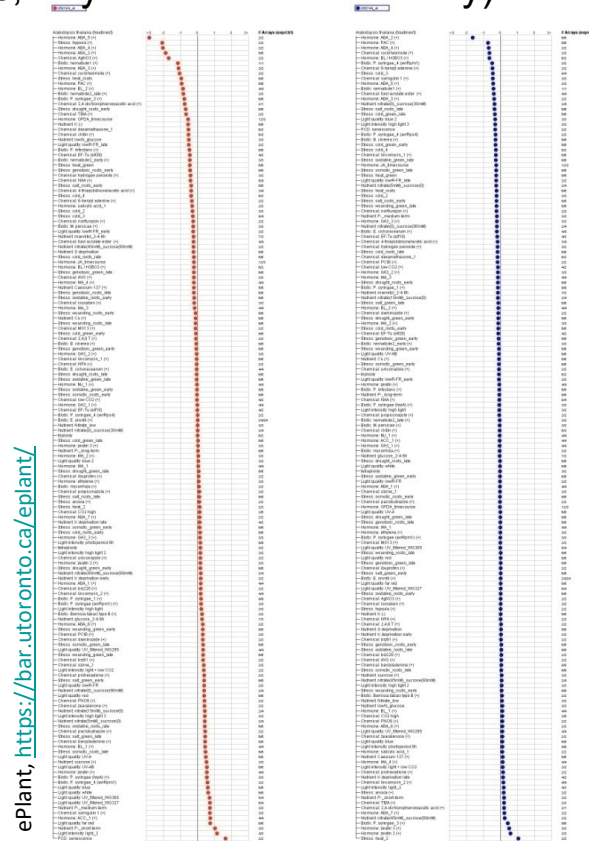
Databáze

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



Databáze

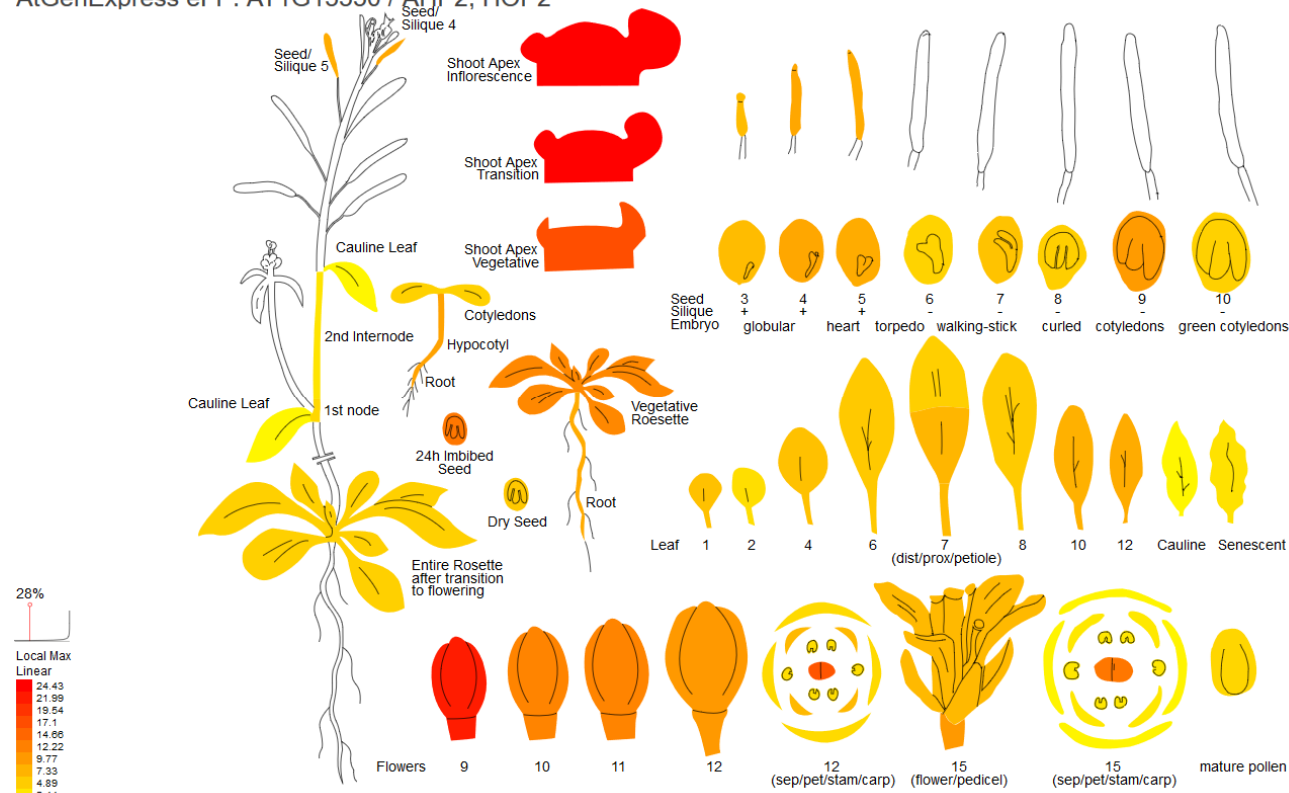
- **Analýza exprese pomocí Genevestigator** (**AHP1** a **AHP2**, *Arabidopsis*, Affymetrix ATH 22K Array)



Databáze

□ Analýza exprese pomocí ePlant

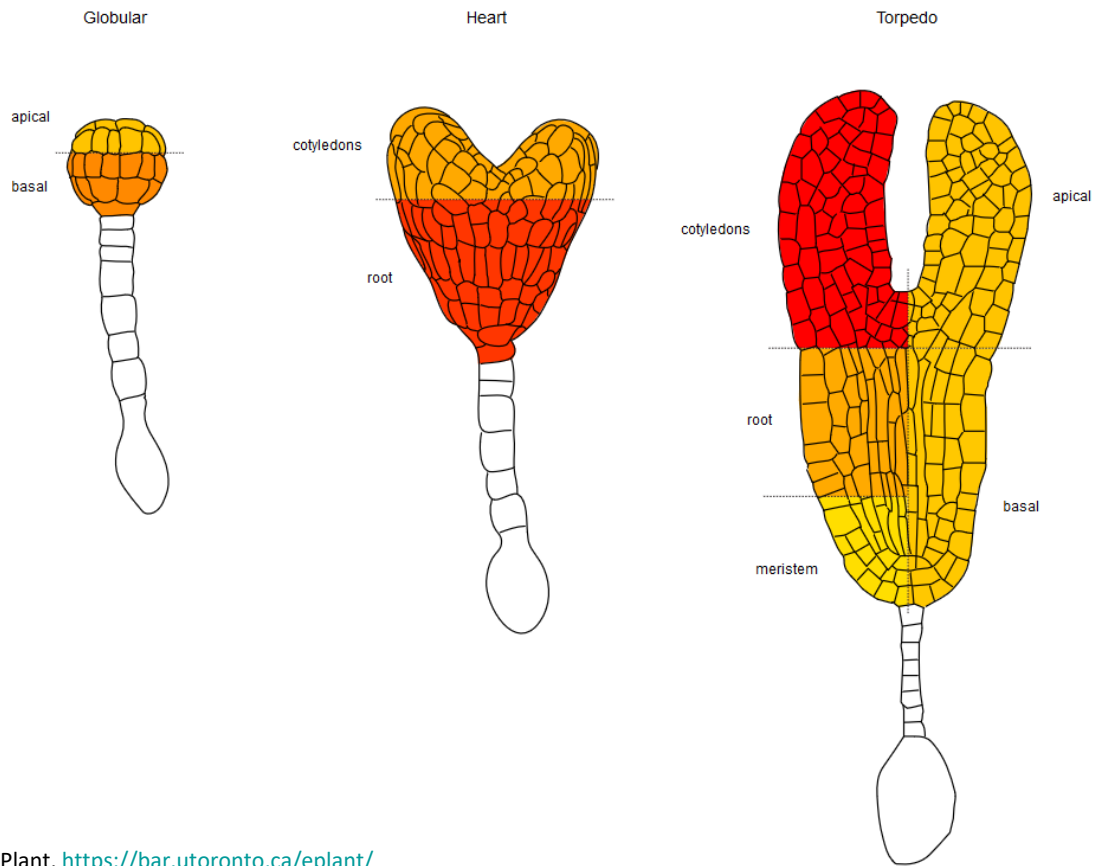
AtGenExpress eFP: AT1G13330 / AHP2, HOP2



ePlant, <https://bar.utoronto.ca/eplant/>

Databáze

□ **Analýza exprese pomocí ePlant**

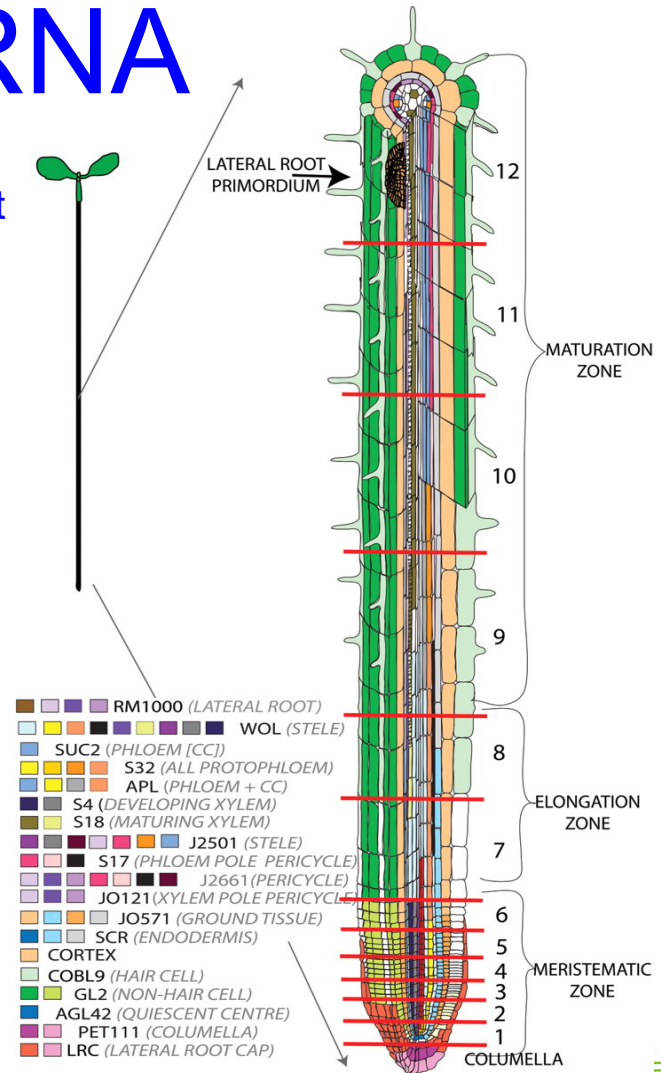
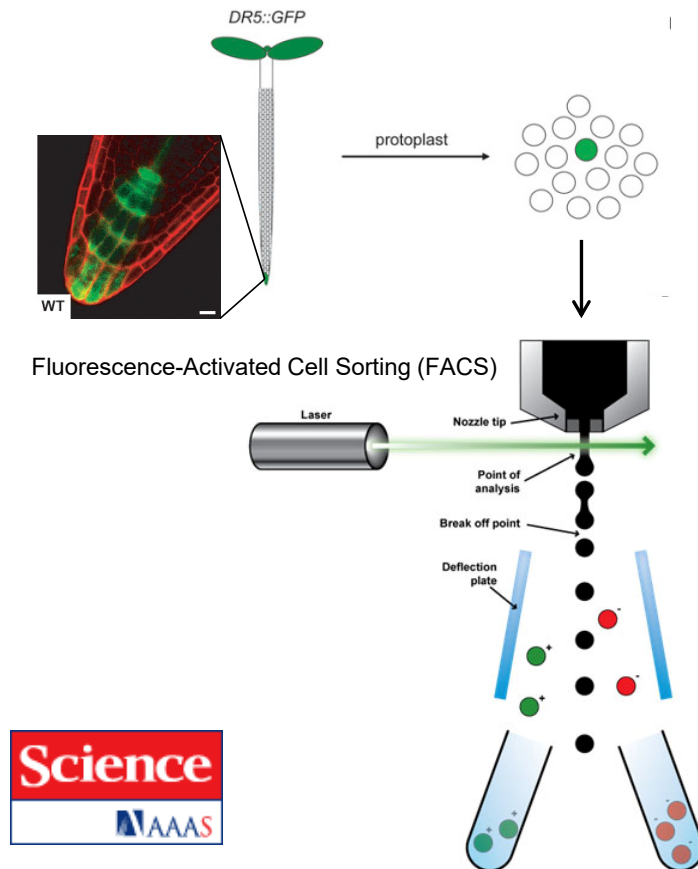


Osnova

- 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

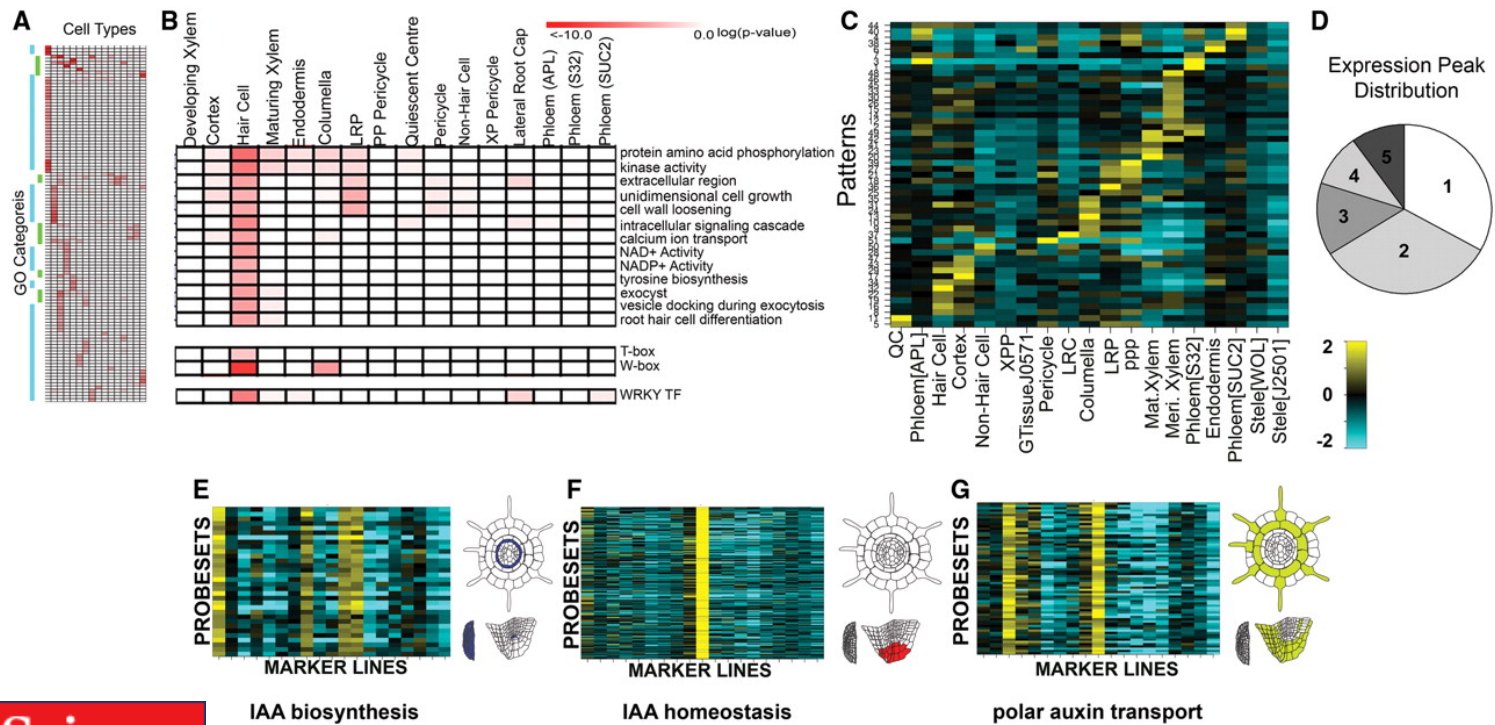
Expression Maps - RNA

High-Resolution Expression Map in Arabidopsis Root



Expression Maps - RNA

High-Resolution Expression Map in Arabidopsis Root



IAA biosynthesis

IAA homeostasis

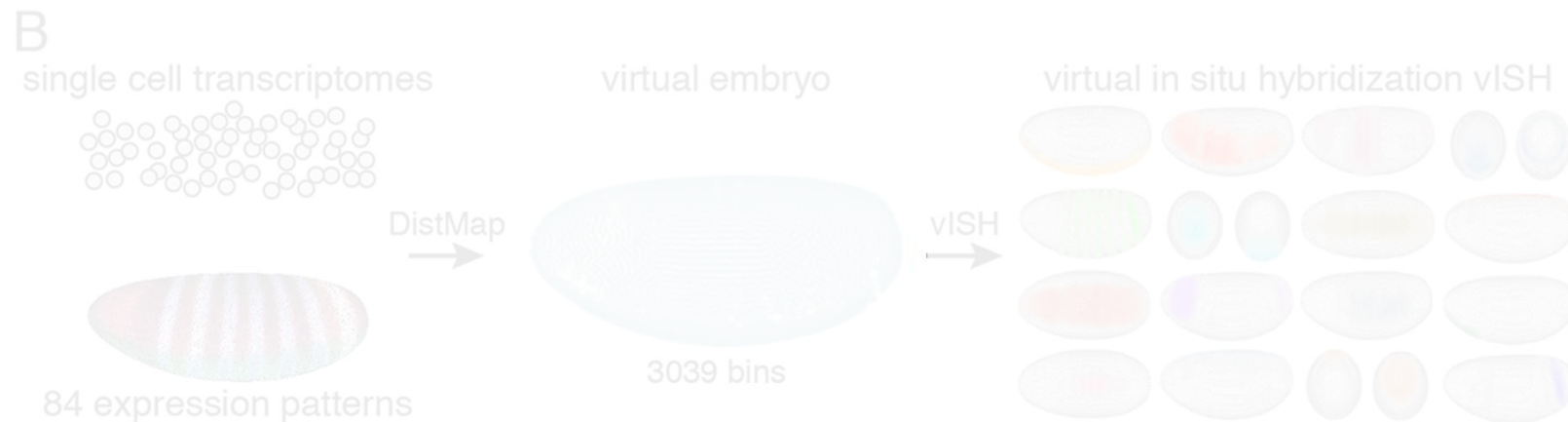
polar auxin transport

Brady et al., *Science*, 2007



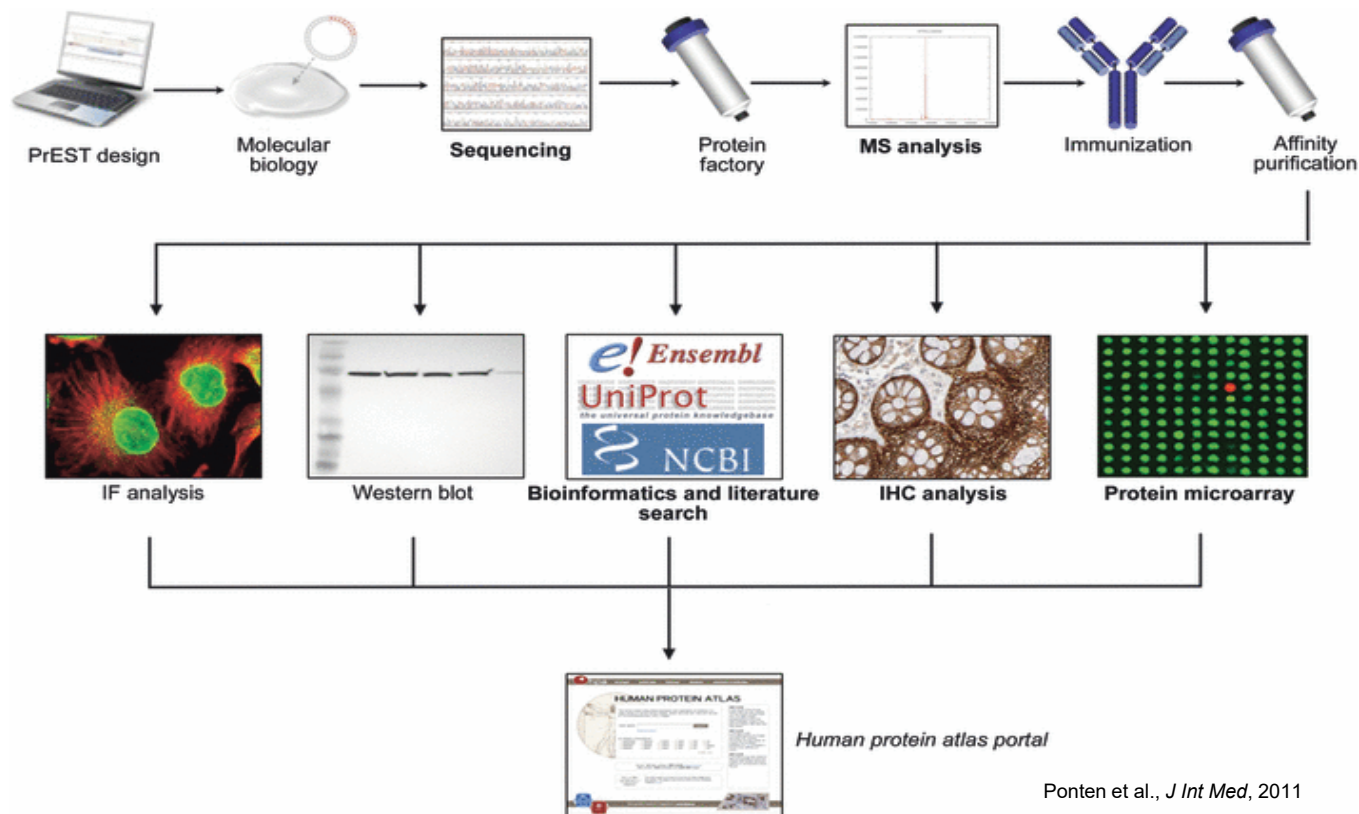
Expression Maps - RNA

- High-Resolution Expression Map in Drosophilla



Expression Maps - Proteins

- Human Protein Atlas



Expression Maps - Proteins

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

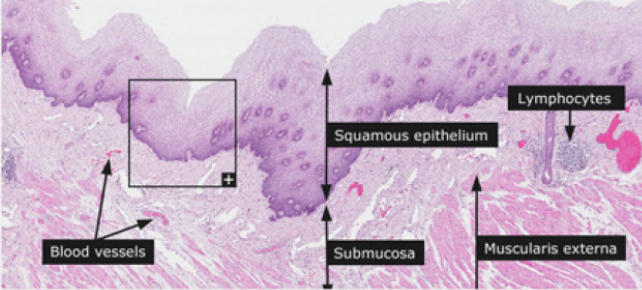
THE HUMAN PROTEIN ATLAS

ABOUT & HELP

SEARCH ? >>

[Fields >](#)

e.g. [CD44](#), [ELF3](#), [KLK3](#), or use Fields to search specific fields such as [protein_class:Transcription factors](#) or [chromosome:X](#)



News

Protein evidence according to [Fagerberg et al](#) is summarized in the [chromosome progress diagram](#).

Version: **11.0**
Atlas updated: 2013-03-11
[release history](#)

15156 genes with protein expression profiles based on **18707** antibodies.

dictionary: histology of esophagus

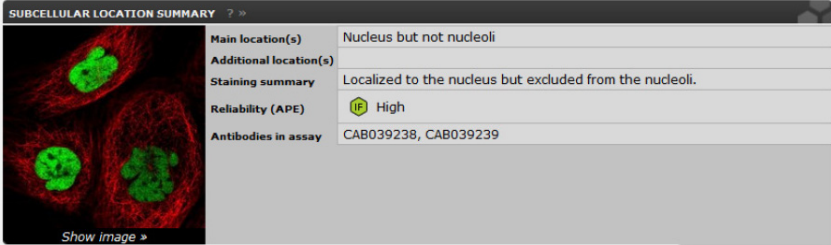
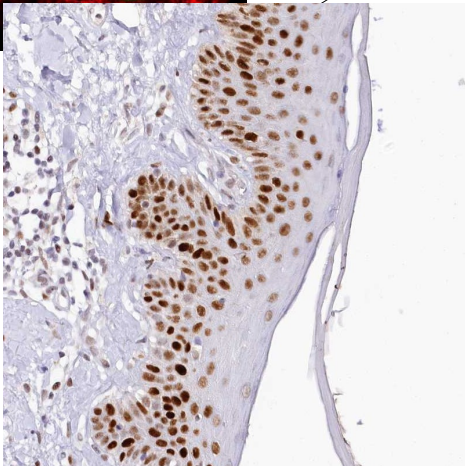
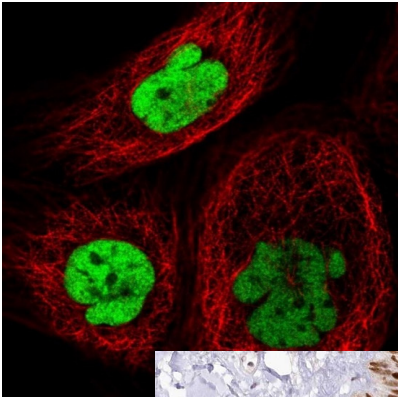
*Knut och Alice
Wallenbergs
Stiftelse*

The Human Protein Atlas project is funded by the Knut & Alice Wallenberg foundation.



Expression Maps - Proteins

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



SUBCELLULAR LOCATION SUMMARY

Main location(s)	Nucleus but not nucleoli
Additional location(s)	
Staining summary	Localized to the nucleus but excluded from the nucleoli.
Reliability (APE)	High
Antibodies in assay	CAB039238, CAB039239

[Show image >](#)

NORMAL TISSUE & ORGAN SUMMARY

Expression summary: Fractions of cells showed weak nuclear and/or cytoplasmic expression.
Tissue specificity: Expressed in 11 out of 82 cell types
Reliability (APE): High
Antibodies in assay: CAB002973, CAB039238, CAB039239

Organ	No of cell types	Protein expression
CNS (brain)	11	<div style="width: 100%;"></div>
Hematopoietic (blood)	8	<div style="width: 100%;"></div>
Liver and pancreas	5	<div style="width: 100%;"></div>
Digestive (GI-tract)	13	<div style="width: 100%;"></div>
Respiratory (lung)	4	<div style="width: 100%;"></div>
Cardiovascular	1	<div style="width: 100%;"></div>
Female tissues	13	<div style="width: 100%;"></div>
Placenta	2	<div style="width: 100%;"></div>
Male tissues	5	<div style="width: 100%;"></div>
Urinary tract (kidney)	3	<div style="width: 100%;"></div>
Skin and soft tissues	14	<div style="width: 100%;"></div>
Endocrine tissues	3	<div style="width: 100%;"></div>

[Show image >](#)

MORE SUBCELL DATA

MORE TISSUE DATA

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**

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)



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

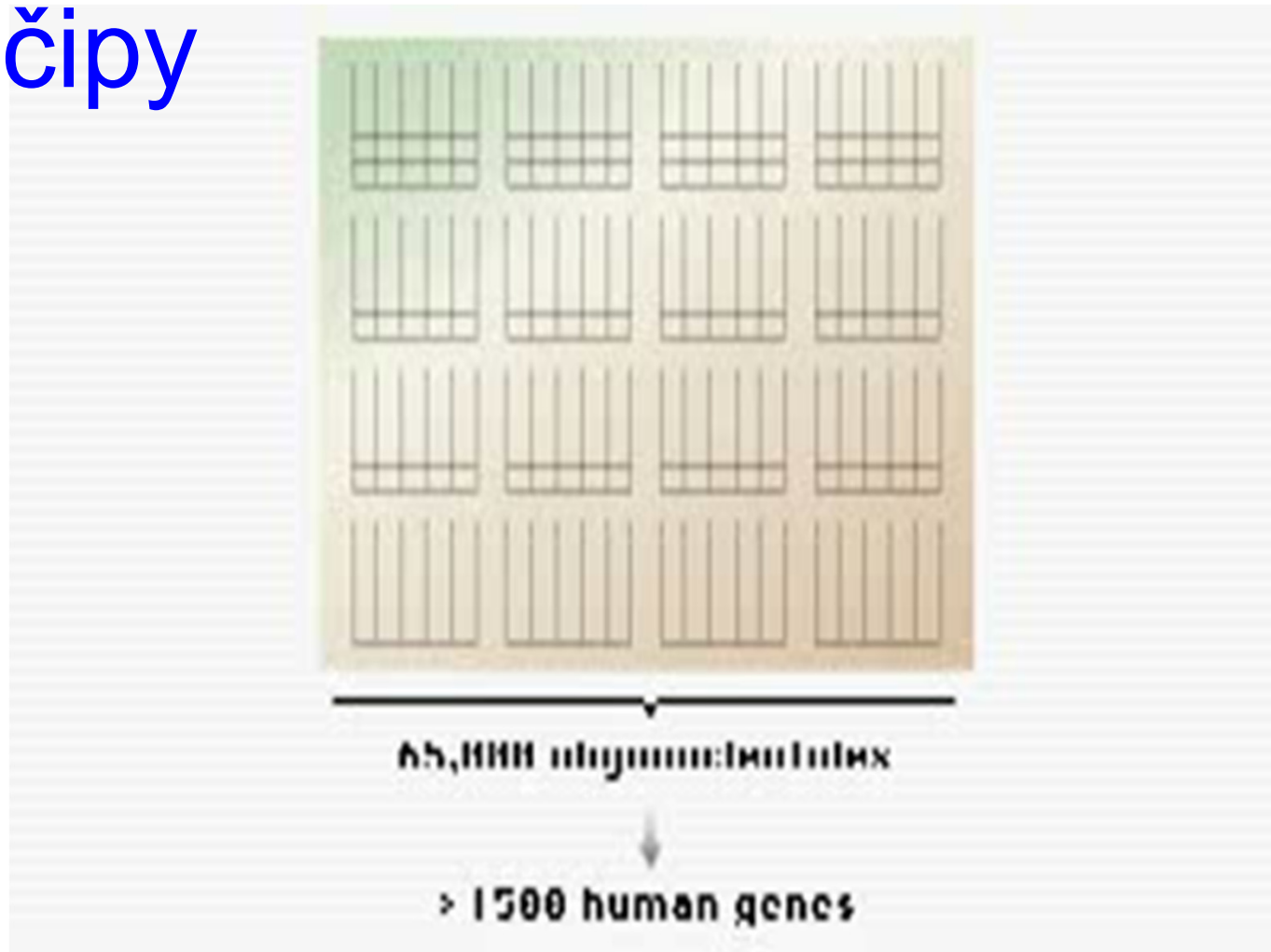


Affymetrix ATH1 *Arabidopsis* genome array

Critical Specifications	
Number of arrays	One
Number of sequence represented	>24,000 gene sequences
Feature size	18 µm
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*

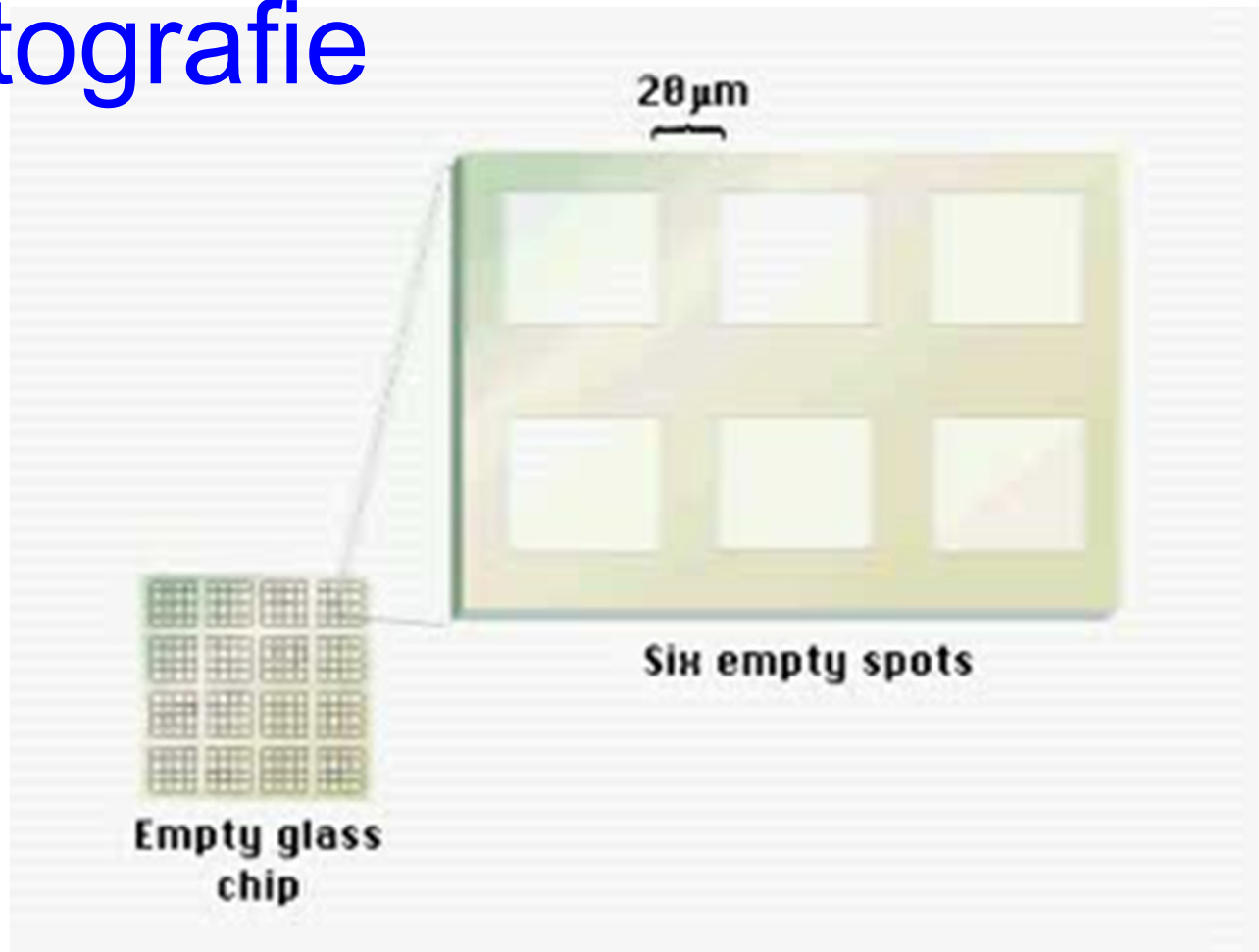
*As measured by detection in comparative analysis between a complex target containing spiked control transcriptions and a complex target with no spikes.

DNA čipy



Alberts et al., Molecular Biology of the Cell

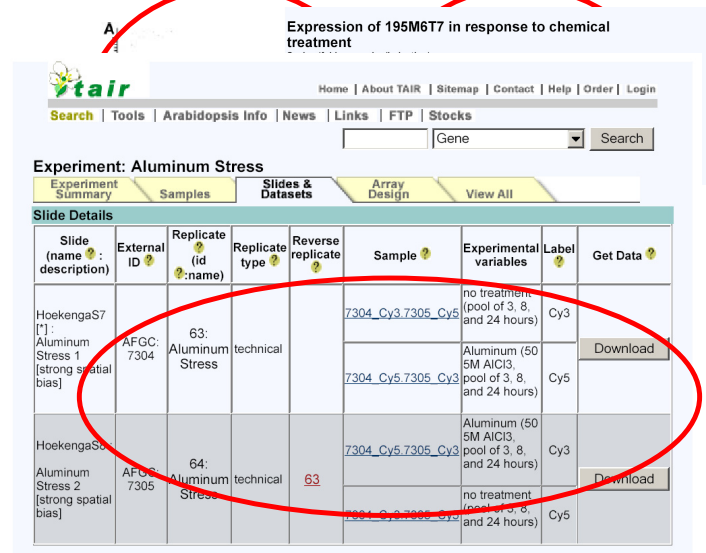
Fotolitografie



Alberts et al., Molecular Biology of the Cell

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



The screenshot shows the TAIR database interface for the experiment 'Aluminum Stress'. The table below is a simplified version of the 'Slide Details' table shown in the image.

Slide (name & description)	External ID	Replicate (id & name)	Replicate type	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, 8, and 24 hours)	Cy3	Download
					7304_Cy5.7305_Cy3	Aluminum (50 5M AlCl3, pool of 3, 8, and 24 hours)	Cy5	
HoekengaS8: Aluminum Stress 2 [strong spatial bias]	AFGC: 7305	64: Aluminum Stress	technical	63	7304_Cy5.7305_Cy3	Aluminum (50 5M AlCl3, pool of 3, 8, and 24 hours)	Cy3	Download
					7304_Cy3.7305_Cy5	no treatment (pool of 3, 8, and 24 hours)	Cy5	

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

Che et al., 2002

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 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**

Systemová biologie - definice

Systemová 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.

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

Systemová biologie - definice

Systemová biologie se zabývá studiem biologických systémů, jejichž chování nelze redukovat na *lineární součet funkcí jejich částí*. Systemová 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.



Výsledky –omických studií vs. biologicky relevantní závěry

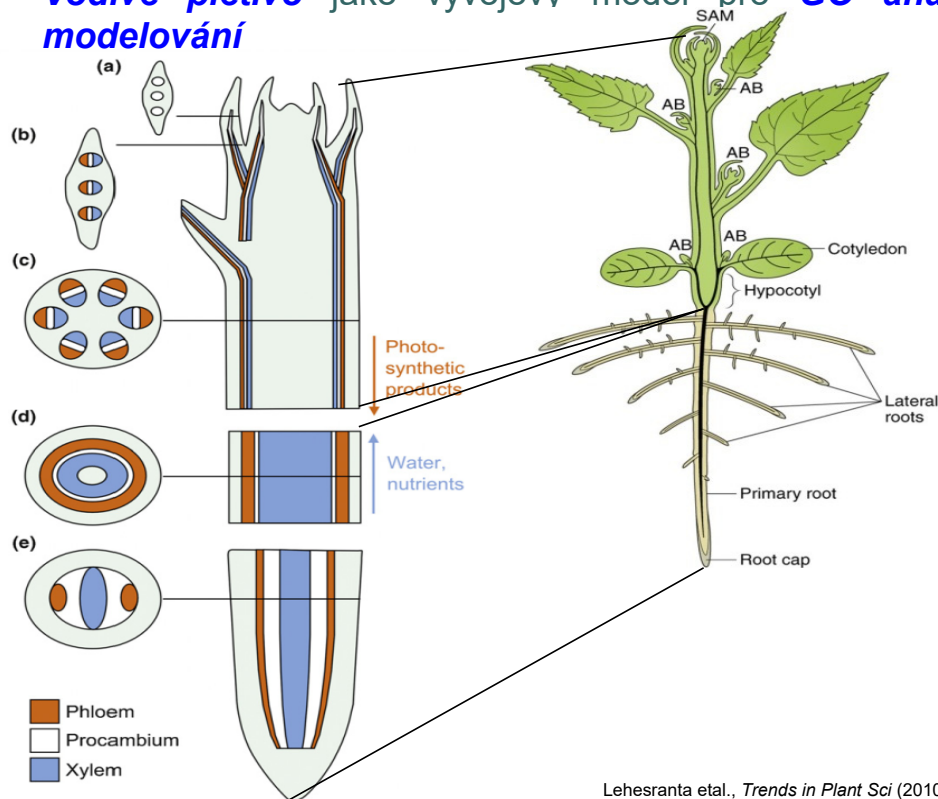
- Výsledky –omických studií reprezentují **enormní množství dat**, např. geny s rozdílnou expresí. Ale jak z nich získat **biologicky relevantní závěry**?

Djii, 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	6.88885e-05	0,00039180	1 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	0,00053505	5 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	0,00028514	3 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	0,00037473	6 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

Vývoj rostlinných vodivých pletiv

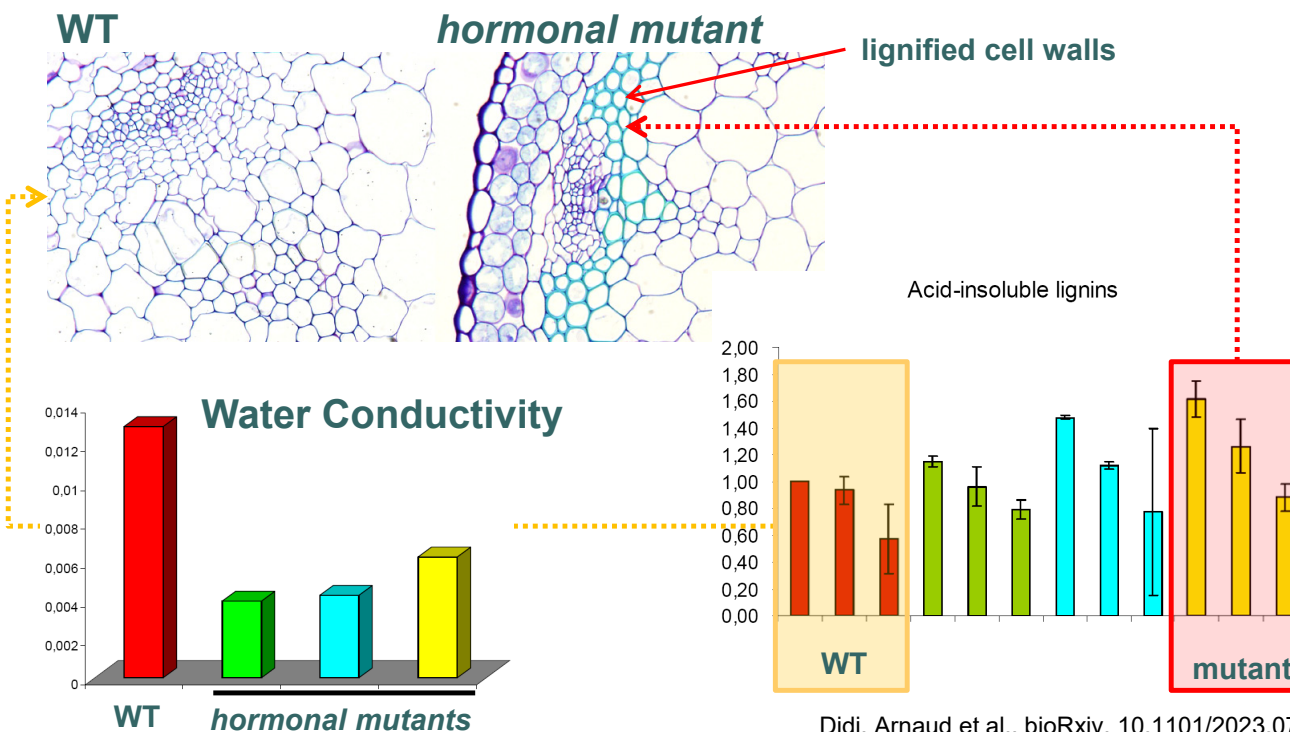
- **Vodivé pletivo** jako vývojový model pro **GO analýzu** a **MRN modelování**



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

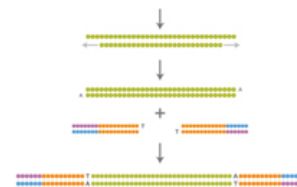
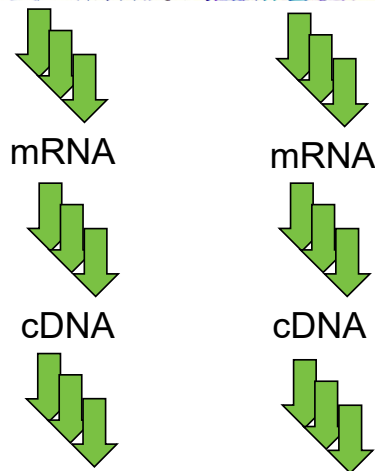
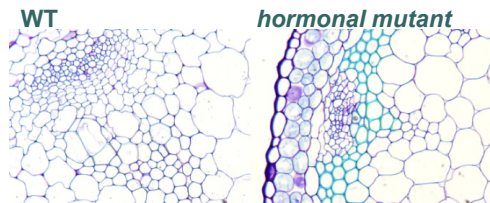
Hormonální regulace vývoje rostlinných vodivých pletiv

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



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)



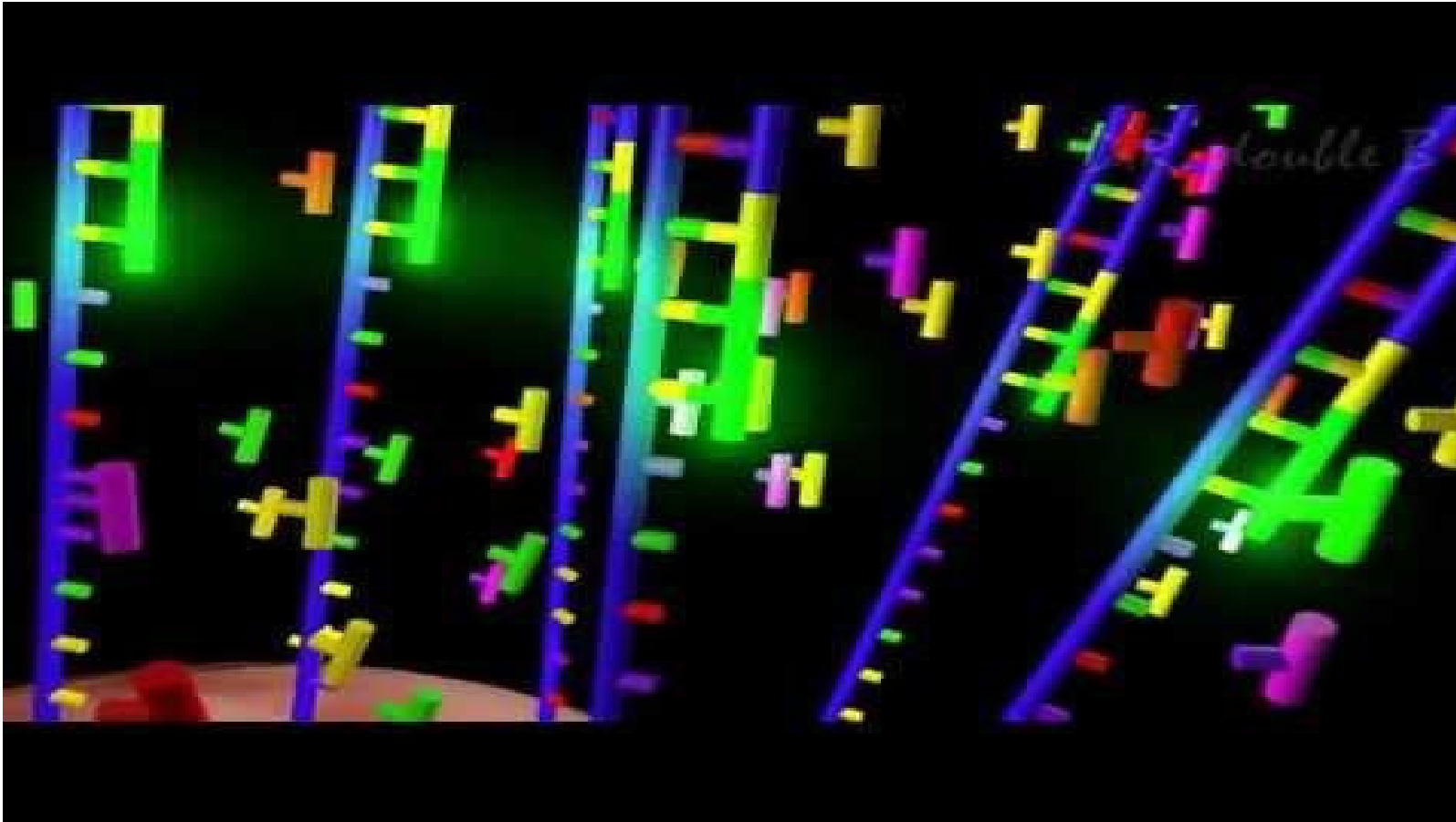
Sequencing by Synthesis
~1.5 to 11 days



CASAVA
2 days (30 min hands-on)
Illumina.com

Sekvenování pomocí Illumina a určení přesného **počtu transkriptů**

Sekvenování DNA metodami NGS



Výsledky –omických studií vs. biologicky relevantní závěry

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

Ddii, 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	6.88885e-05	0,00039180	1 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	0,00053505	5 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	0,00028514	3 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	0,00037473	6 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

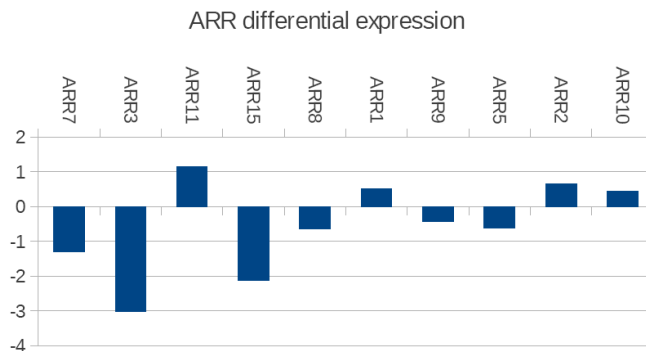
Ddii, Arnaud et al., *bioRxiv*, 10.1101/2023.07.26.550726 and unpublished

Osnova

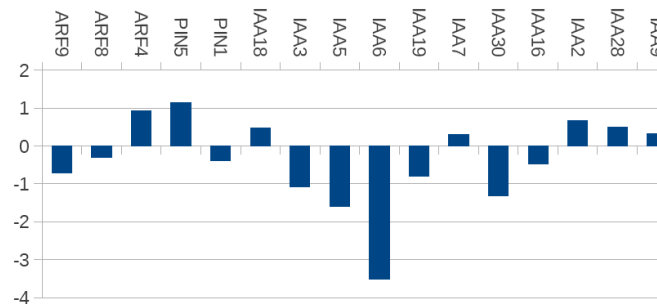
- Struktura genů
- Metody analýzy **genové exprese**
 - **Kvalitativní analýza exprese genů**
 - Příprava **transkripční fúze** promotoru analyzovaného genu s reporterovým genem (**gen zpravodaj**)
 - Příprava **translační fúze** kódující oblasti analyzovaného genu s reporterovým genem
 - Využití dostupných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

Analýza genové ontologie

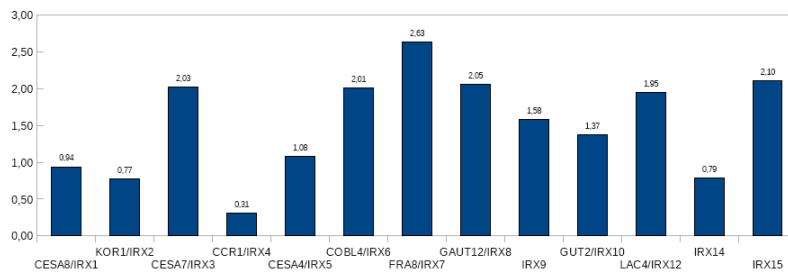
- Jedním z možných přístupů je studium **genové ontologie**, tj. dříve prokázané **spojitosti** mezi geny a **biologickými procesy**



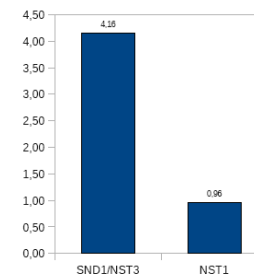
Differentially expressed IAA-related genes



IRXs



XYLEM MARKERS



Analýza genové ontologie

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

Eden et al., *BMC Bioinformatics* (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
Arabisidopsis 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 [WebGeneIt](#) for conversion from other identifier formats.

Or upload a file: D:\Results\2012\Mariane\Prochazka

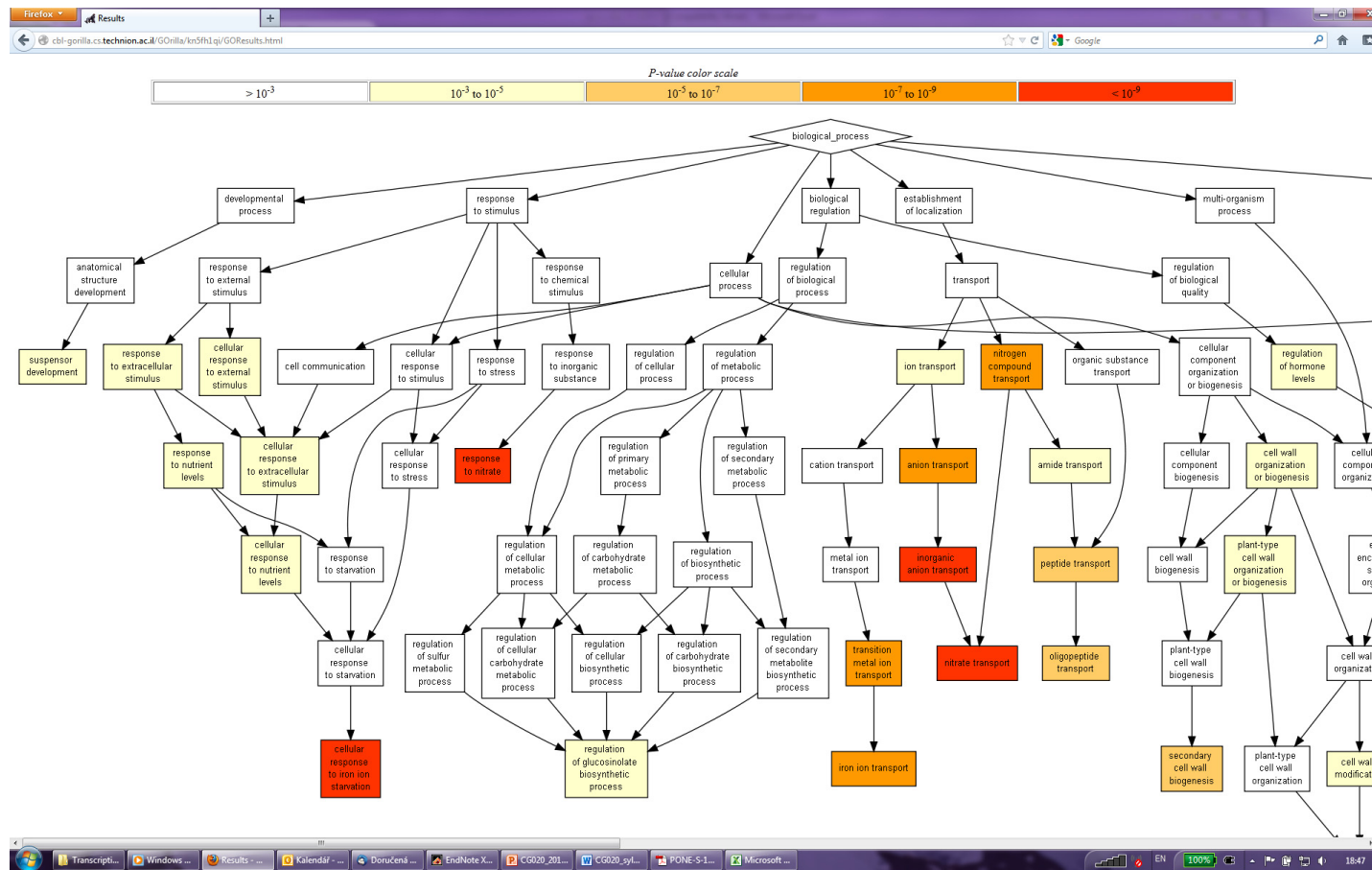
Step 4: Choose an ontology
 Process Function Component All

<http://cbl-gorilla.cs.technion.ac.il/>

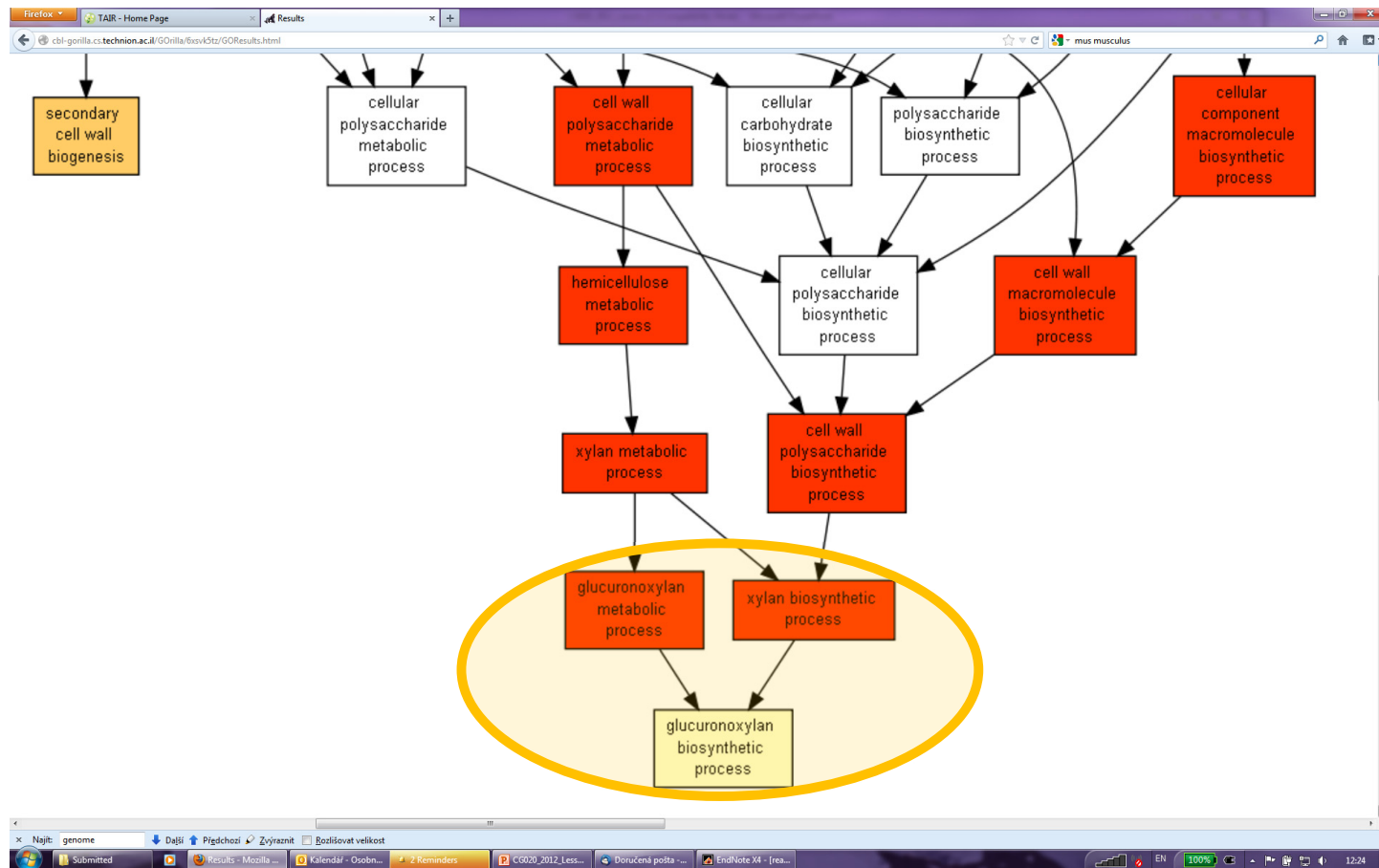
GORILLA

The screenshot shows a web browser window displaying the GORILLA interface. The top part of the page features a complex network diagram with numerous nodes and connections, representing a hierarchical or dependency structure. Below the diagram is a large table with columns for 'ID', 'Name', 'Status', 'Date', and 'Size'. The table contains a list of entries, likely representing the results of a simulation or analysis. The browser's address bar shows the URL 'cbl-gorilla.cs.technion.ac.il/Gorilla/wty283r/GORResults.html'. The browser's taskbar at the bottom shows various open applications, including 'Transcript...', 'Windows ...', 'Results ...', 'Kalendář ...', 'Doručená ...', 'EndNote X...', 'CGO20_201...', 'CGO20_sy...', 'Adobe Ac...', and 'Microsoft ...'. The system tray shows the date and time as '18:39'.

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 PG5IP3 - plant glycoengin-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 atf72 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 PG5IP1 - plant glycoengin-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 atf54 FRD3 - mate efflux family protein AT1G33800 - hypothetical protein
hemiacetaldehyde metabolic process	2.07E-12	2.24E-9	3.30 (6331,77,999,40)	[+] Show genes

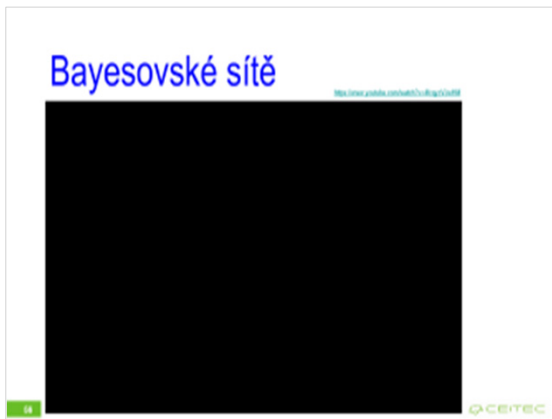
Osnova

- 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
 - Tkáňově 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

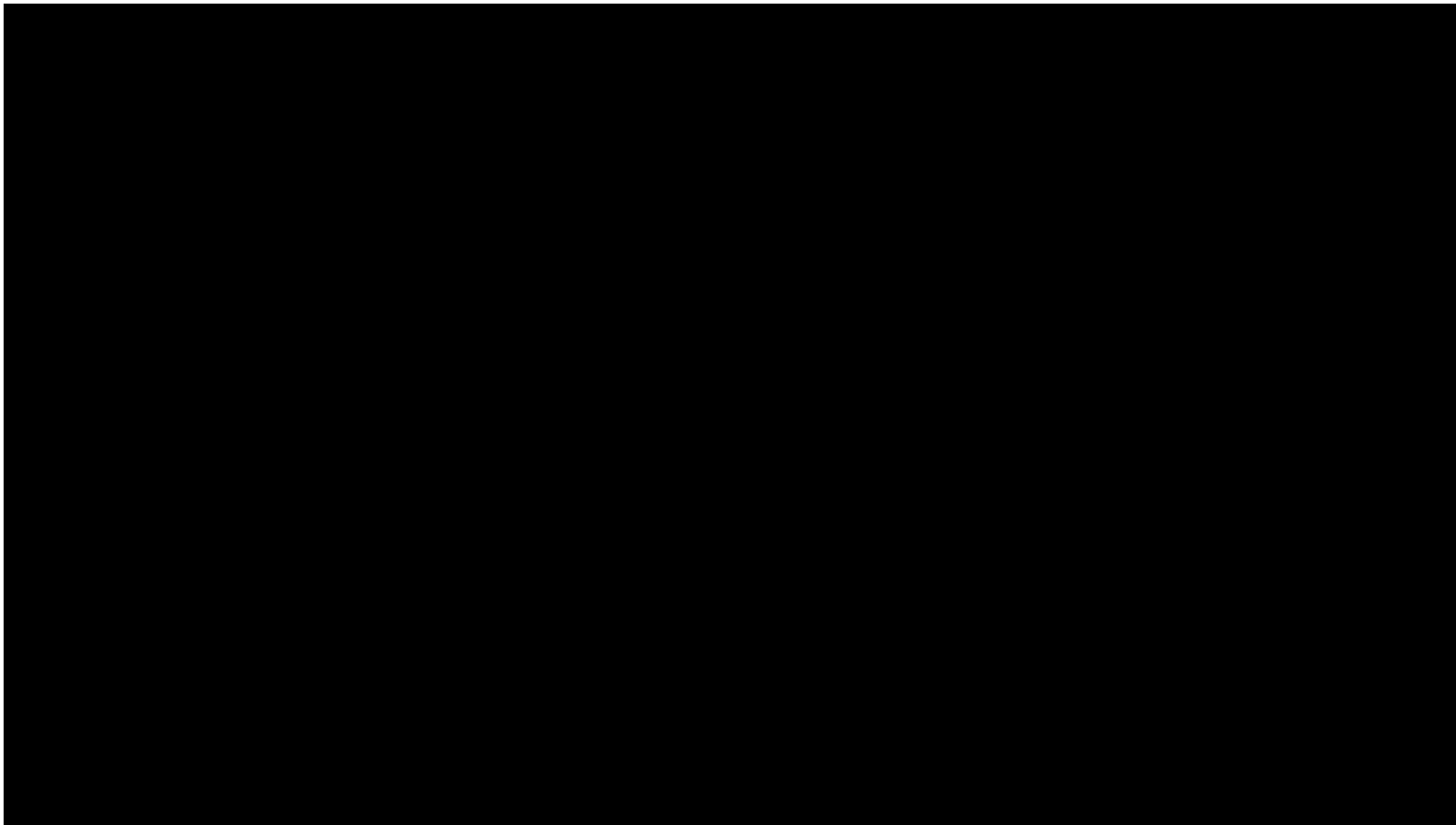
Co je Bayesovská síť?

- **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?

- **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**

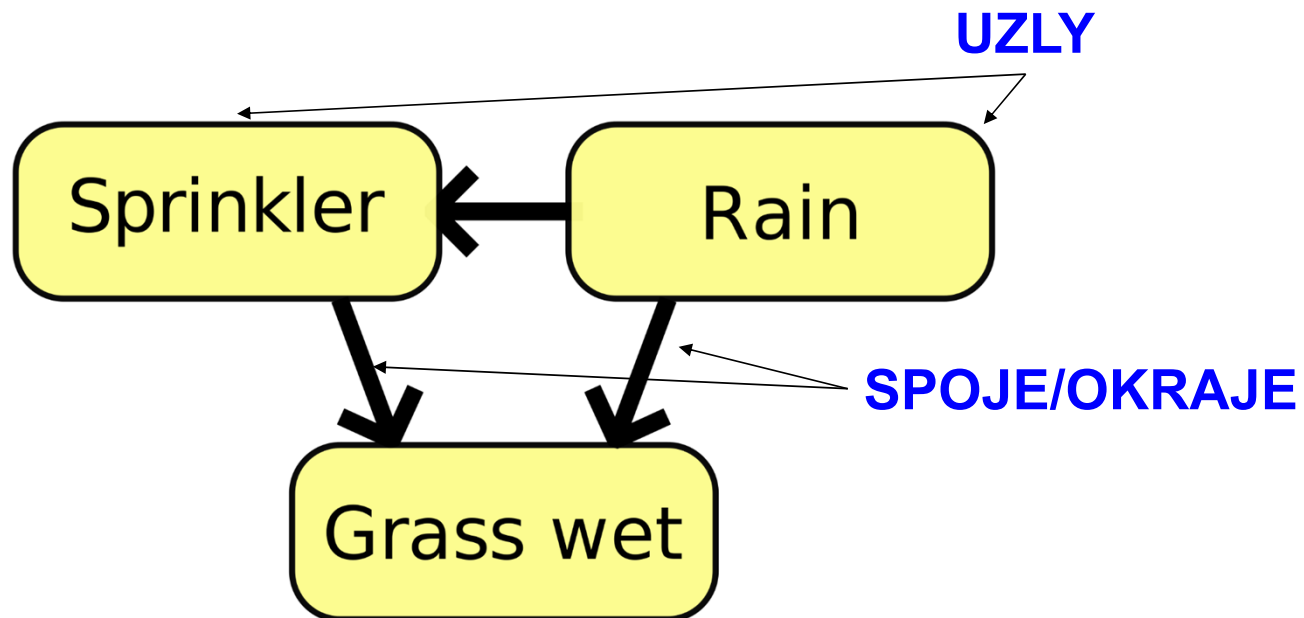
▪ UZLY

- každý uzel představuje **proměnnou**, jako je výška, věk nebo pohlaví. Proměnná může být *diskrétní*, jako například pohlaví = {samičí, samčí}, nebo *spojitá*, jako např. věk

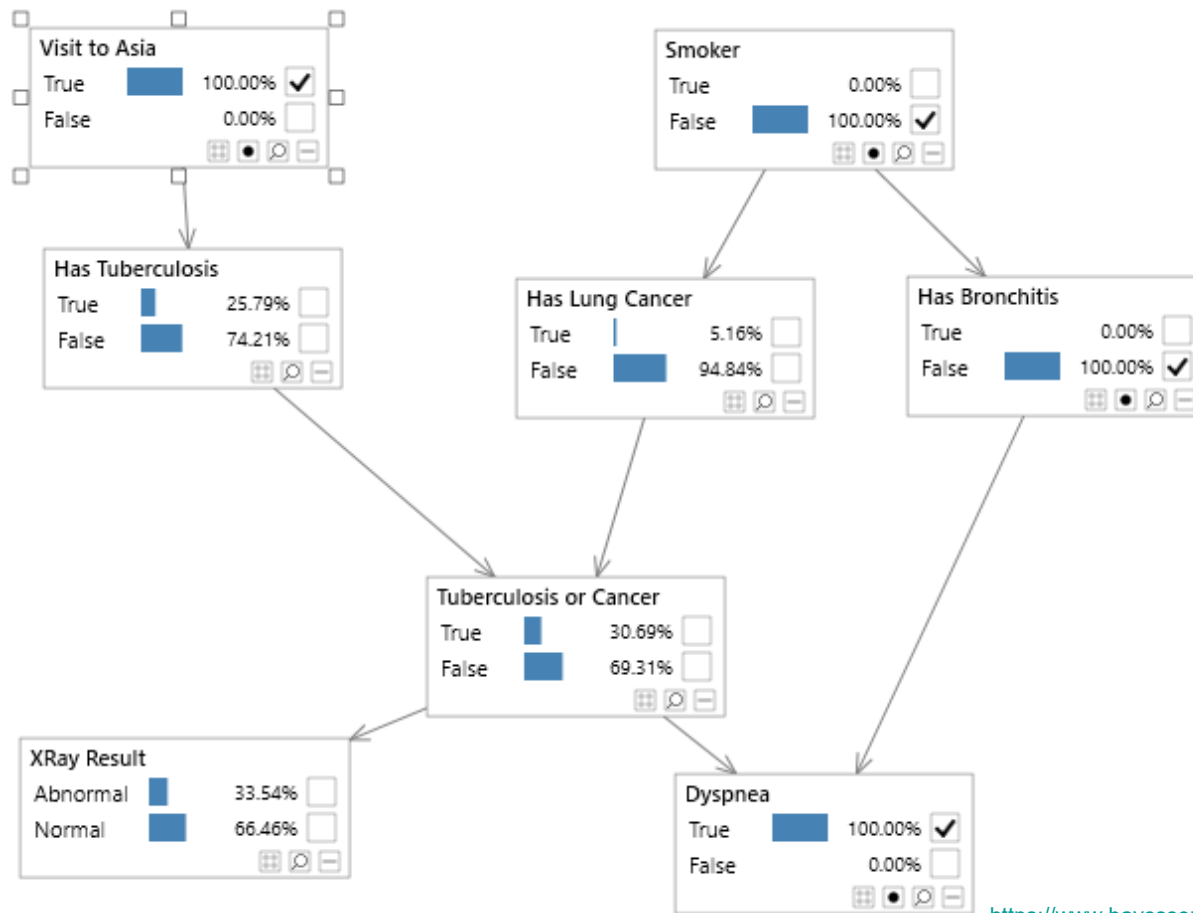
▪ SPOJE

- přidány **mezi uzly**, aby ukazovaly, že **jeden uzel má** přímý **vliv** na **druhý**

Bayesovské sítě



Asijská Bayesovská síť



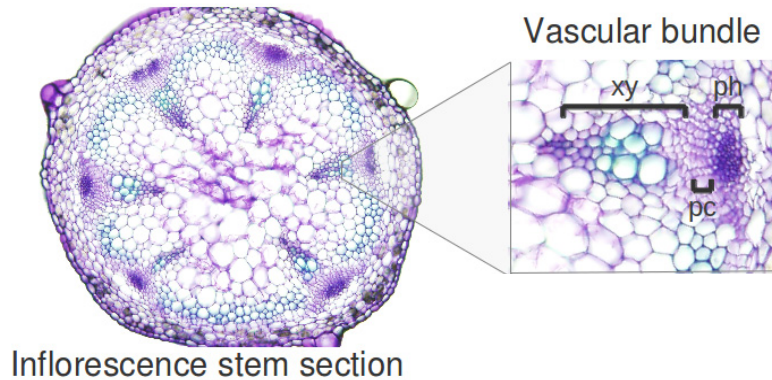
<https://www.bayesserver.com/>

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
- **Systemová 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 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 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 ARR6.	[9]

Modelování molekulárních regulačních sítí

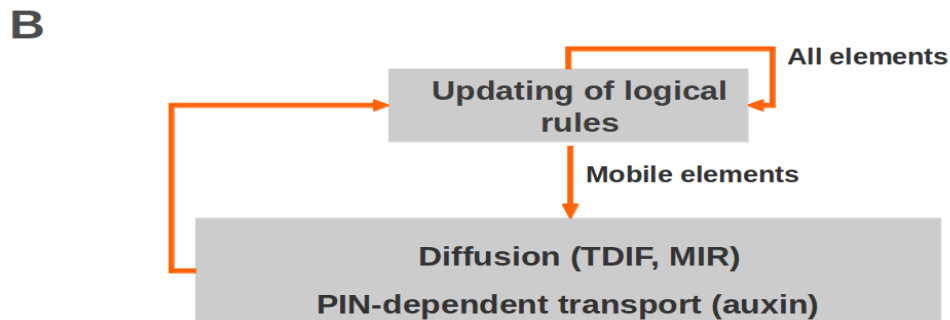
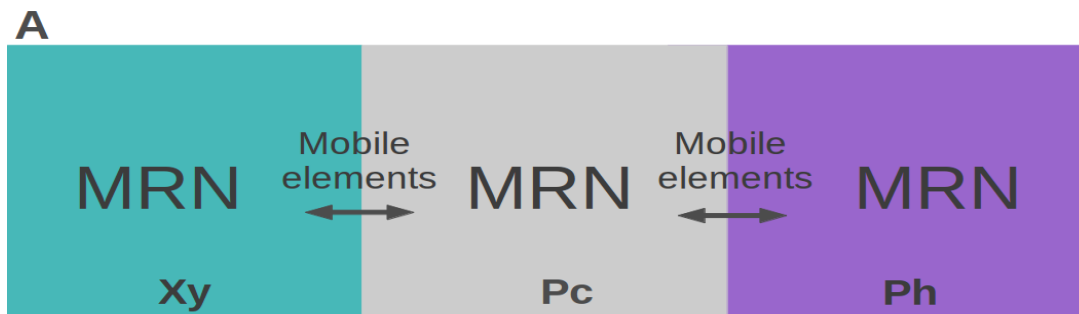
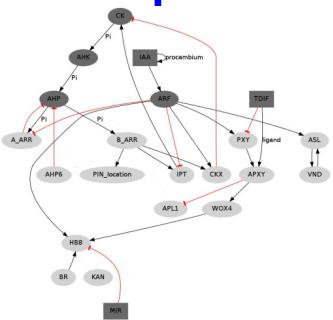
- Formulace *logických pravidel* definujících *dynamiku modelu*

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



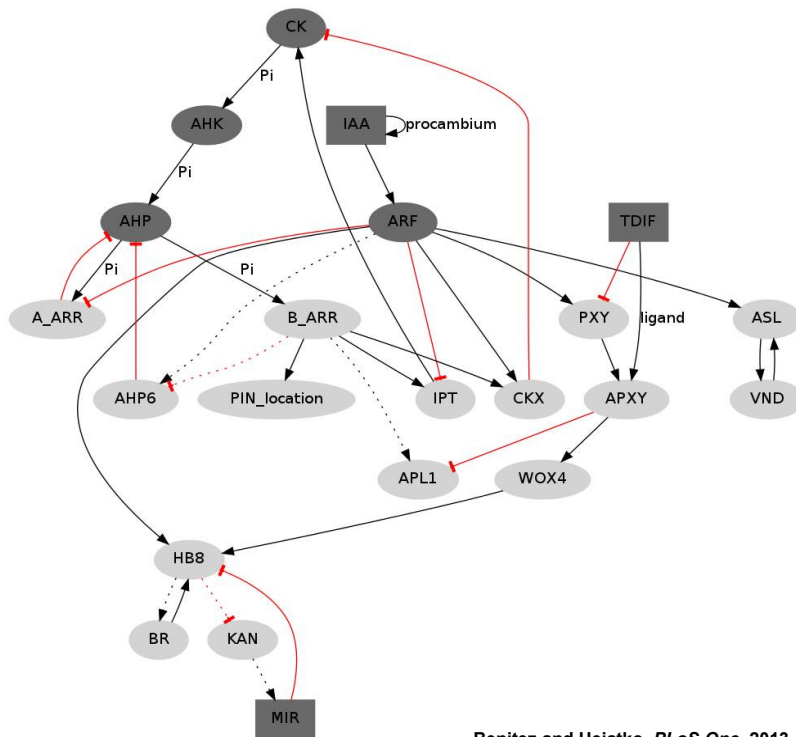
Modelování molekulárních regulačních sítí

- Specifikace chybějících interakcí ze *známých predikcí*

Interaction	Evidence	References
CK → PIN7 radial localization	<p>Predicted interaction (could be direct or indirect)</p> <p>Informed by the following data:</p> <p>During the specification of root vascular cells in <i>Arabidopsis thaliana</i>, CK regulates the radial localization of PIN7.</p> <p>Expression of PIN7:GFP and PIN7::GUS is upregulated by CK with no significant influence of ethylene.</p> <p>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.</p>	<p>[18]</p> <p>[18,20]</p> <p>[19]</p>
CK→ APL	<p>Predicted interaction (could be direct or indirect)</p> <p>Consistent with the fact that APL overexpression prevents or delays xylem cell differentiation, as does CKs.</p> <p>Partially supported by microarray data and phloem-specific expression patterns of CK response factors.</p>	<p>[21]</p> <p>(TAIR, ExpressionSet:1005823559, [22])</p>

Modelování molekulárních regulačních sítí

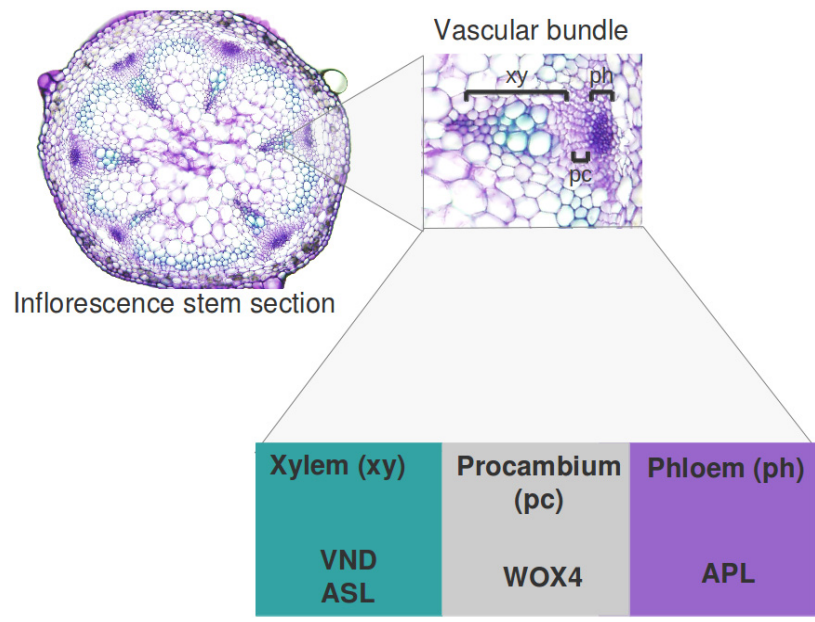
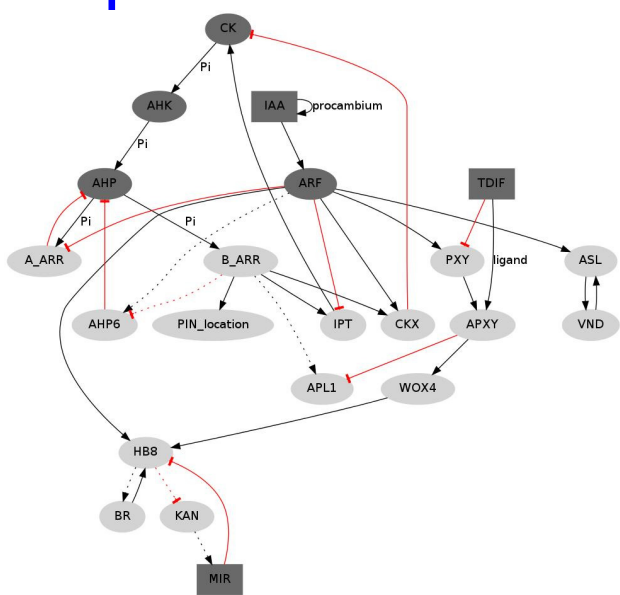
- Příprava *další verze* modelu a její *testování*



Benitez and Hejatko, *PLoS One*, 2013

Modelování molekulárních regulačních sítí

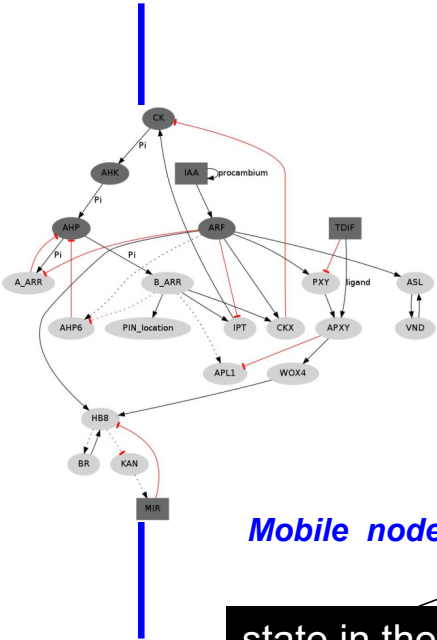
- *Dobry model* by měl být schopen *simulovat realitu*



Benitez and Hejatko, *PLoS One*, 2013

Modelování molekulárních regulačních sítí

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



logical rule function

state in the time t

$$\text{Static nodes: } g_n(t+1) = F_n(g_{n_1}(t), g_{n_2}(t), \dots, g_{n_k}(t))$$

state in the time $t+1$

Amount of TDIF or MIR165 in cell i

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

state in the time $t+1$

constant corresponding to a degradation term

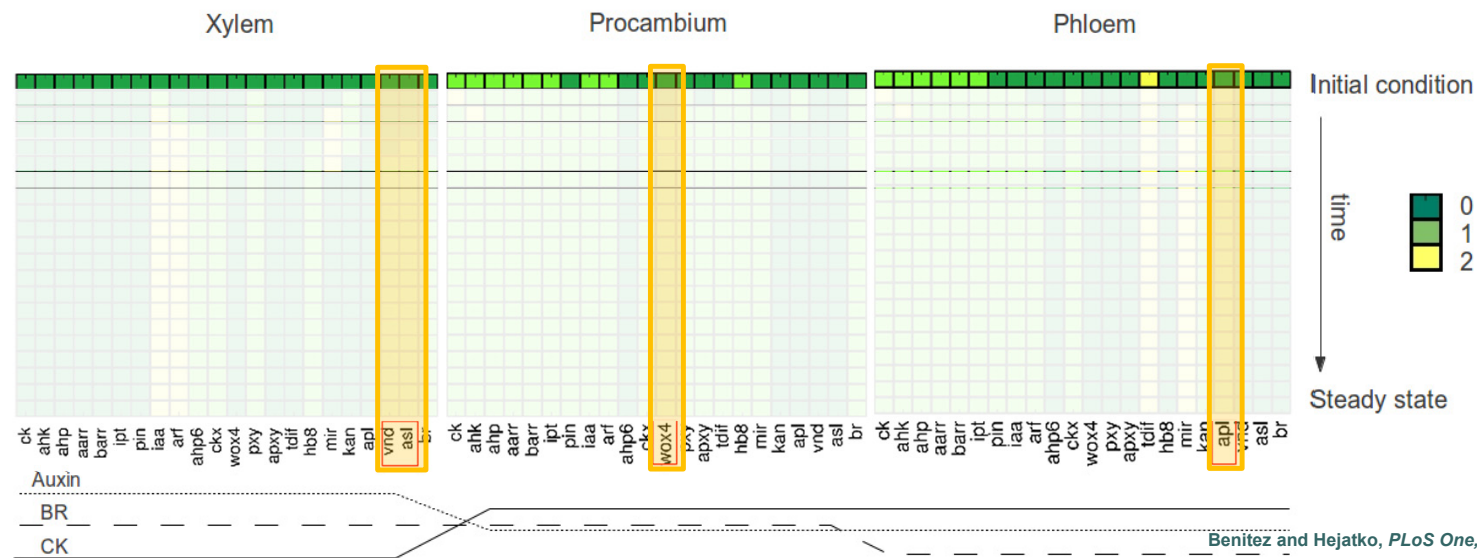
proportion of movable element

Modelování molekulárních regulačních sítí

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

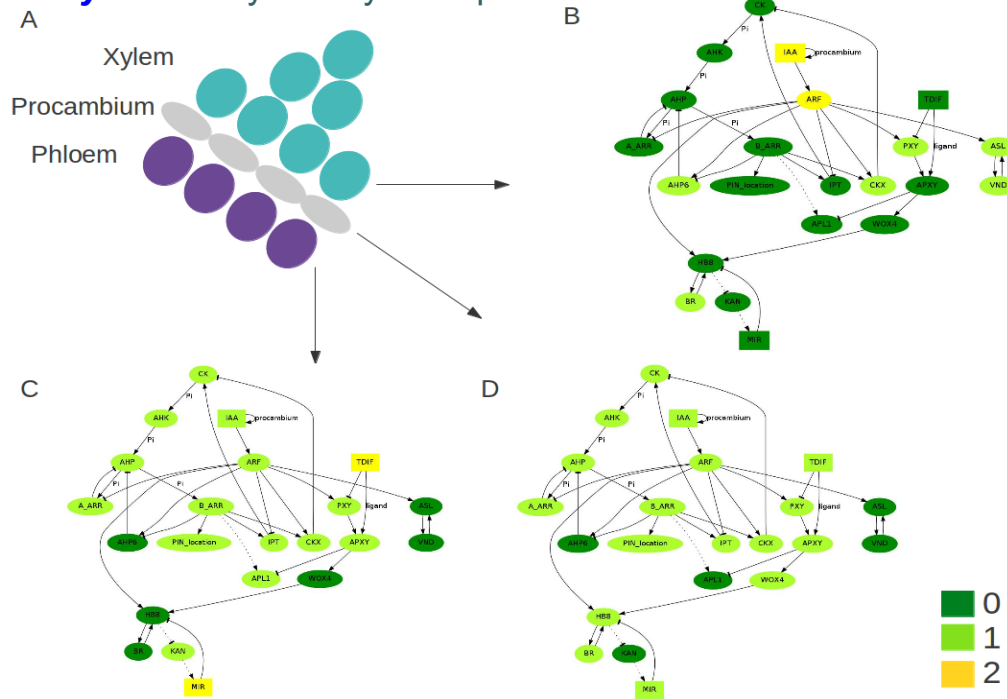
Static nodes: $g_n(t+1) = F_n(g_{n1}(t), g_{n2}(t), \dots, g_{nk}(t))$

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



Modelování molekulárních regulačních sítí

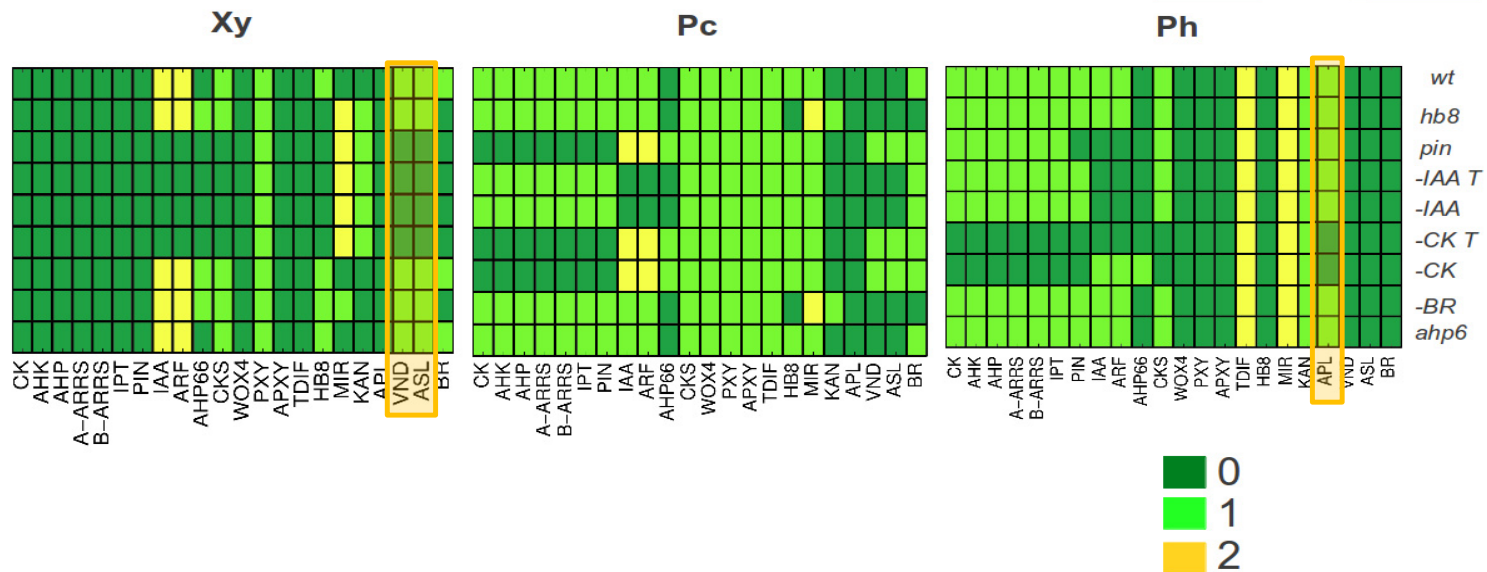
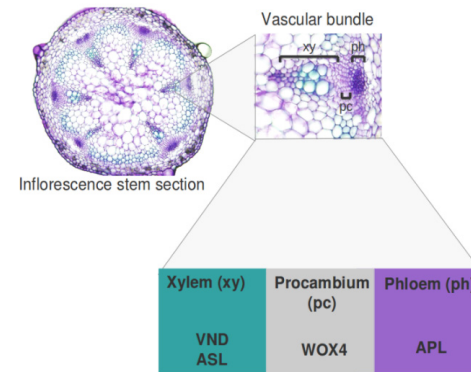
- **Dobrý model** by měl být schopen *simulovat realitu*



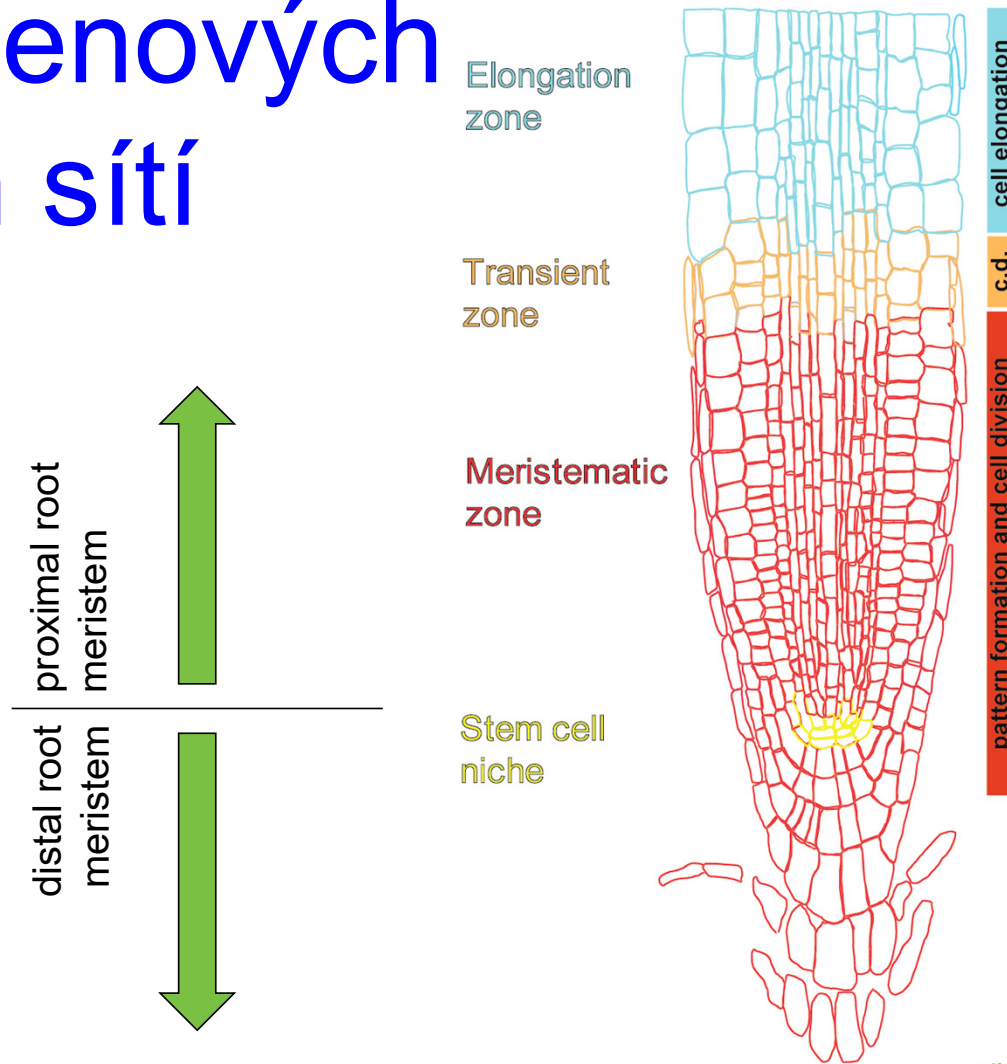
Benitez and Hejatko, *submitted*

Modelování molekulárních regulačních sítí

Simulace *mutantů*

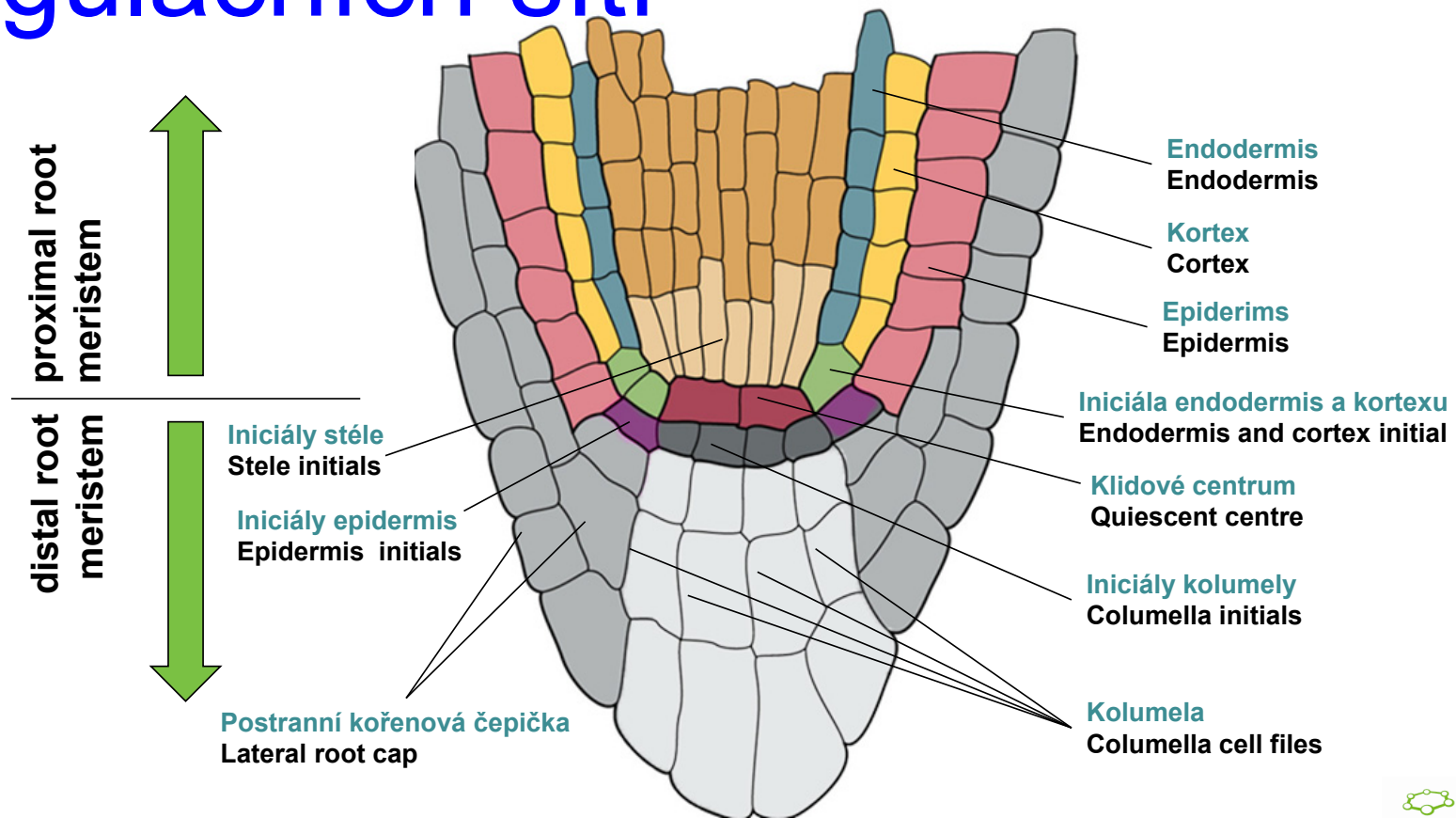


Odvození genových regulačních sítí



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

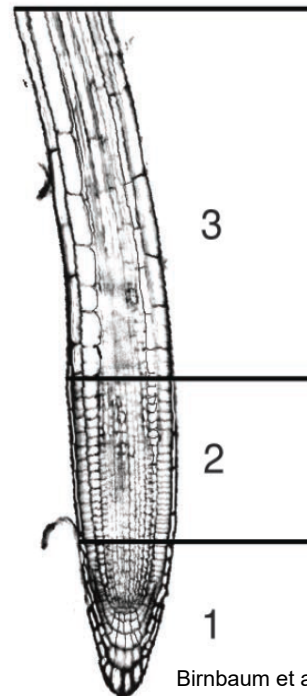
Odvození genových regulačních sítí



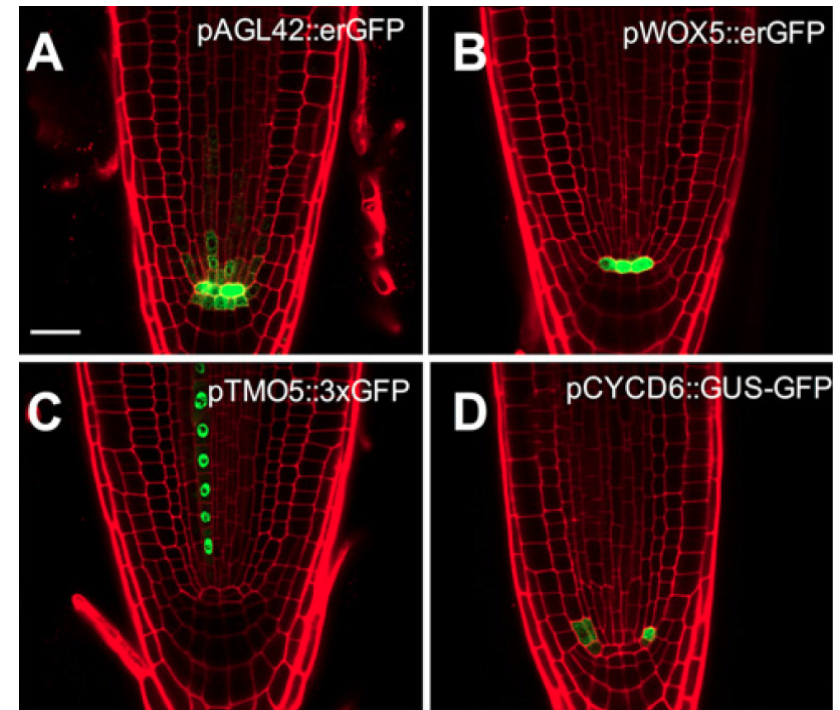
Genové regulační sítě - GENIST

- Odvození GR sítí přes
GENIST

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



Birnbaum et al., *Science*, 2003



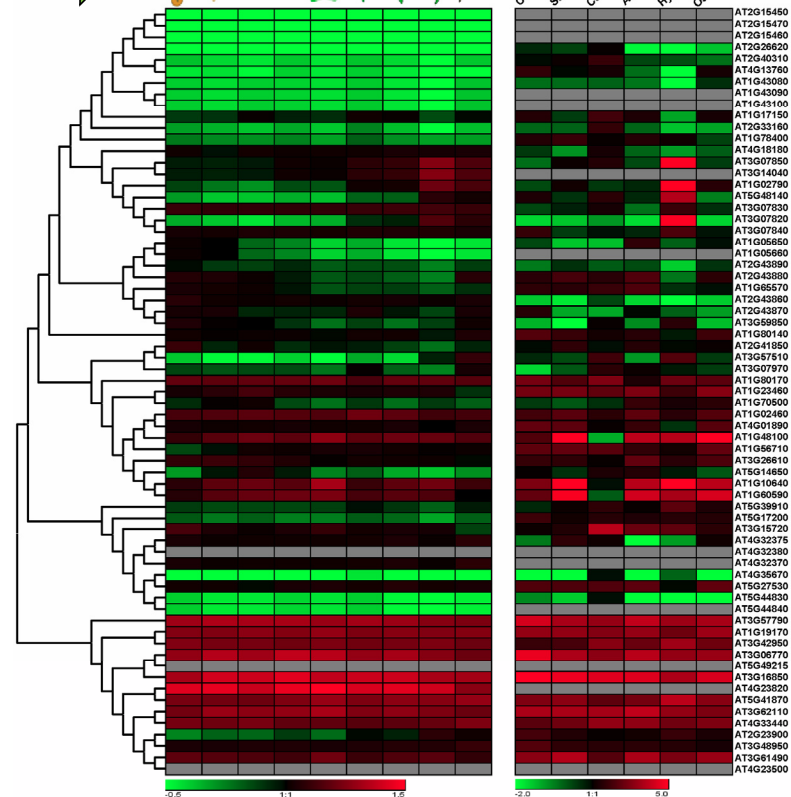
de Luis Balaguer et al., *PNAS*, 2017

Kombinace velkých omických datových sad

PLETIVO/ČAS

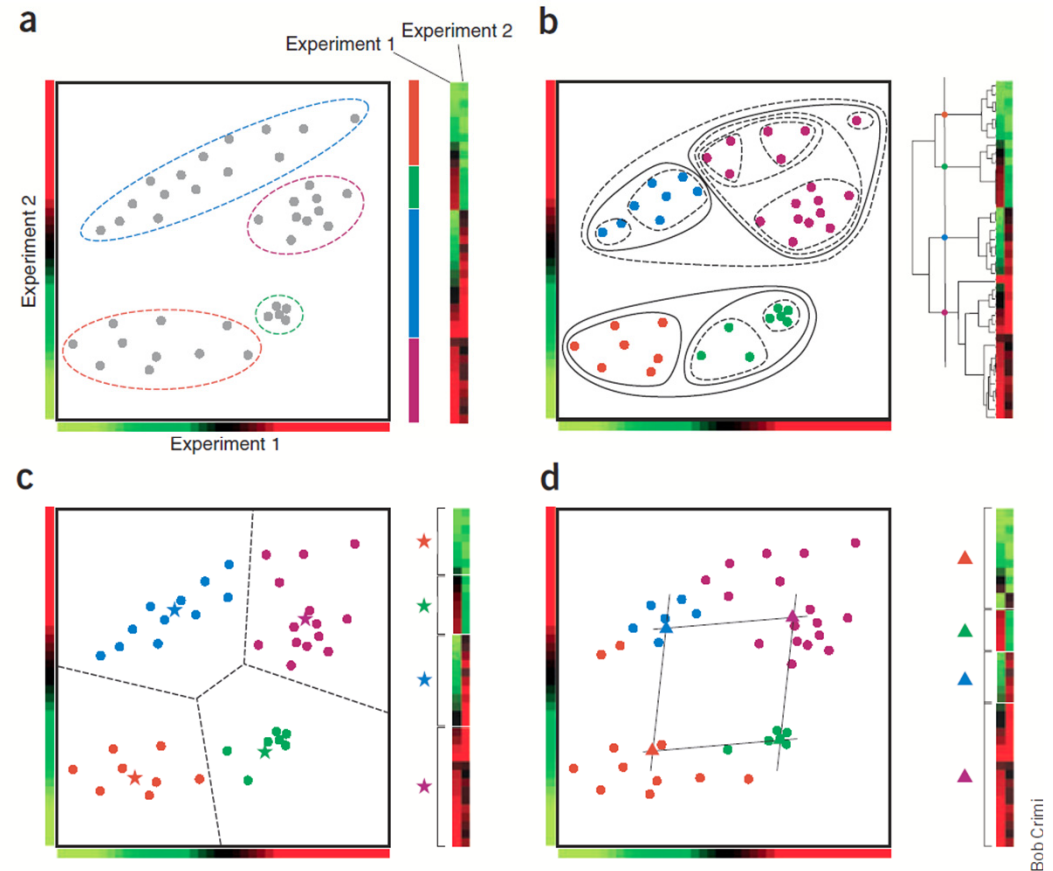


GENY



Genové regulační sítě - GENIST

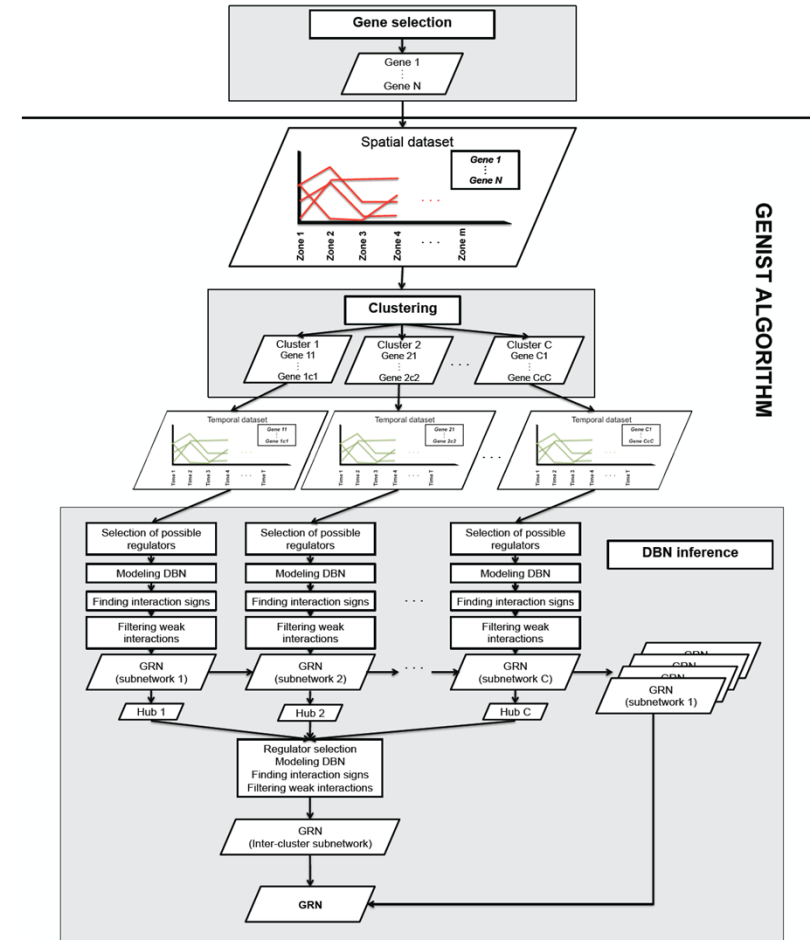
- Odvození GR sítí přes **GENIST**
 - **shlukování (klastrování) genů**
 - Expresní podobnost za různých podmínek/genetické pozadí, časové body, ...
 - **Odvození spojení uvnitř klastru**
 - **Selekce** potenciálních **regulátorů** a **ko-regulátorů**
 - Na základě časové korelace ve změně exprese a/nebo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



Haeseleer, *Computational Biology*, 2005

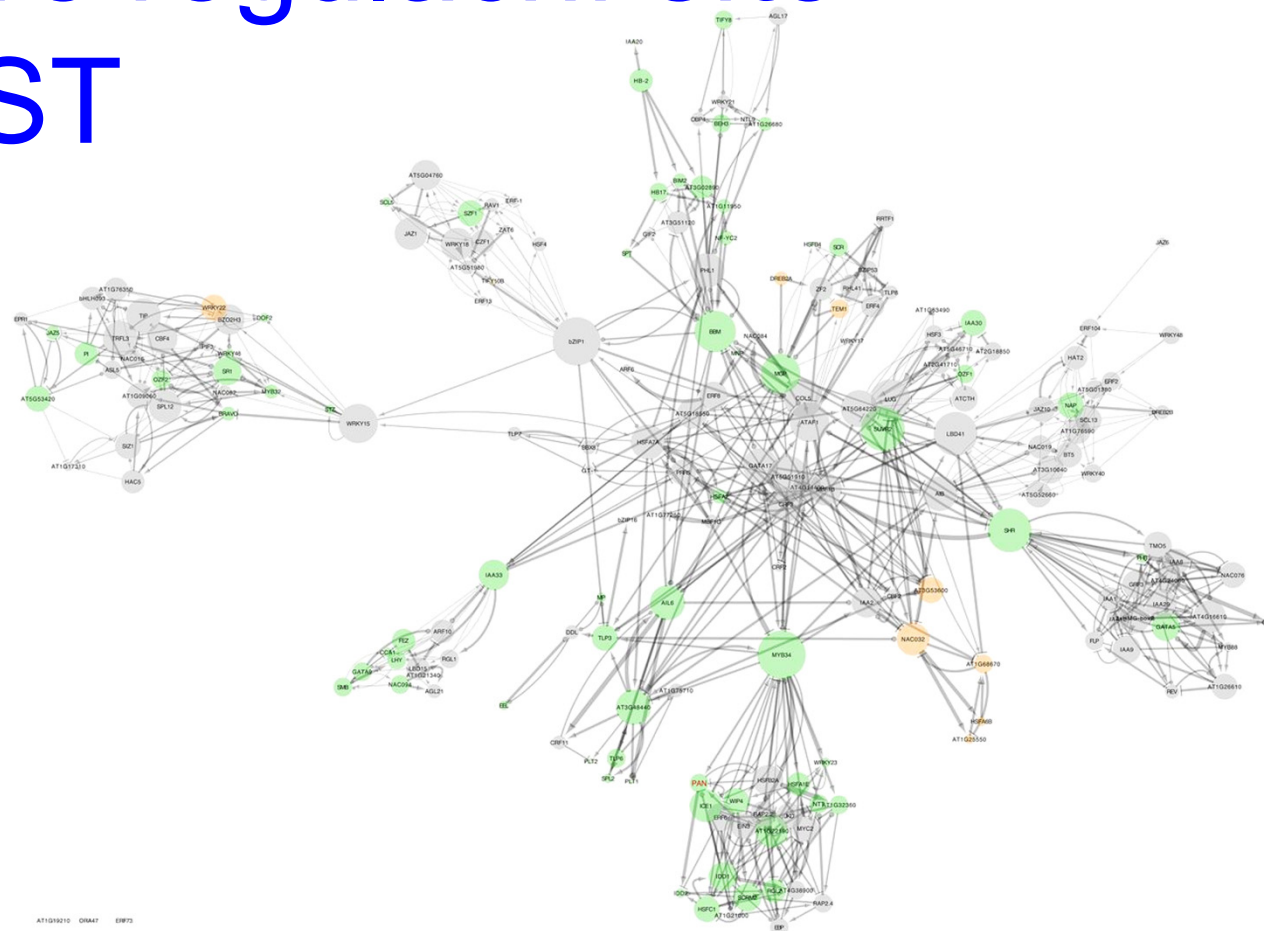
Genové regulační sítě - GENIST

- 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**



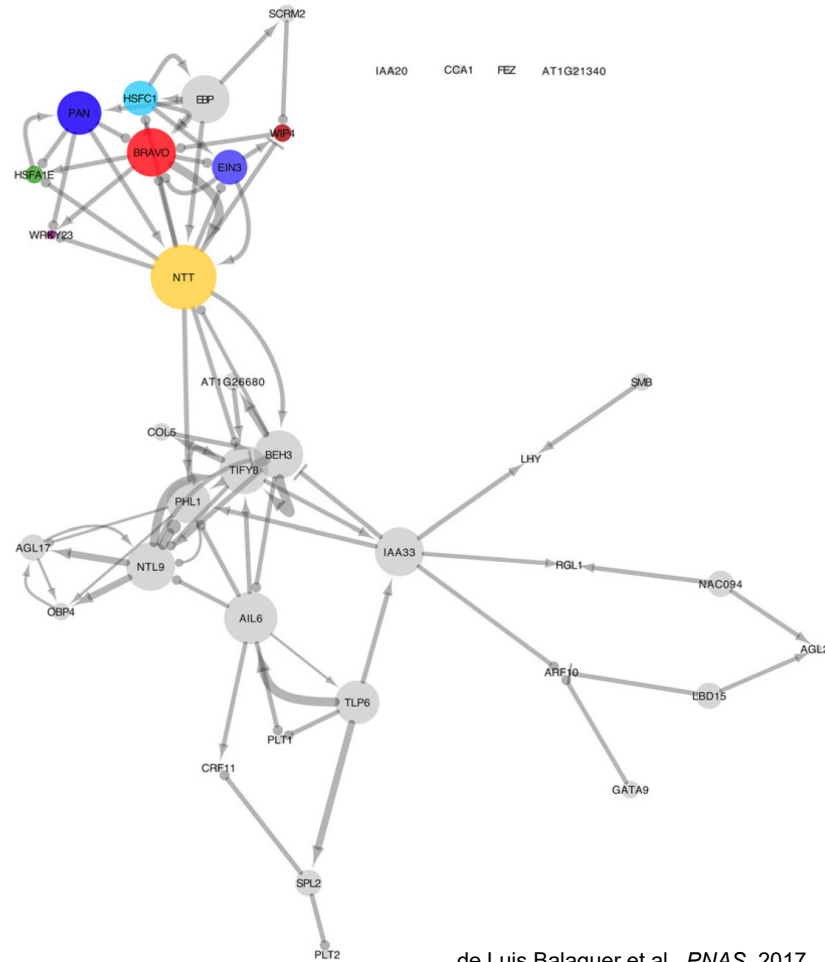
de Luis Balaguer et al., PNAS, 2017

Genové regulační sítě - GENIST



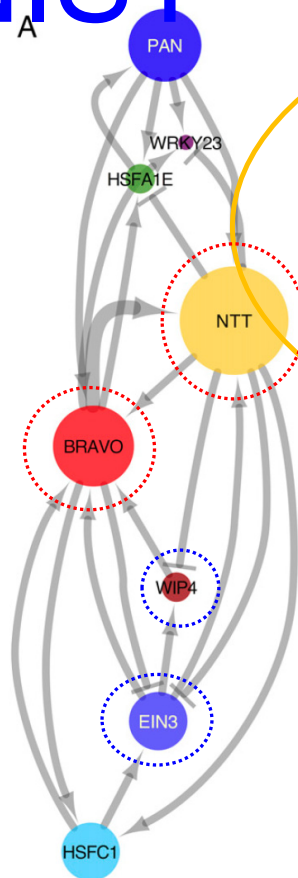
de Luis Balaguer et al., *PNAS*, 2017

Genové regulační sítě - GENIST

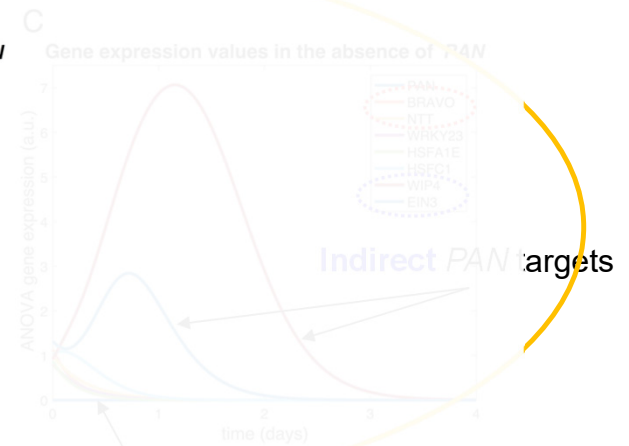
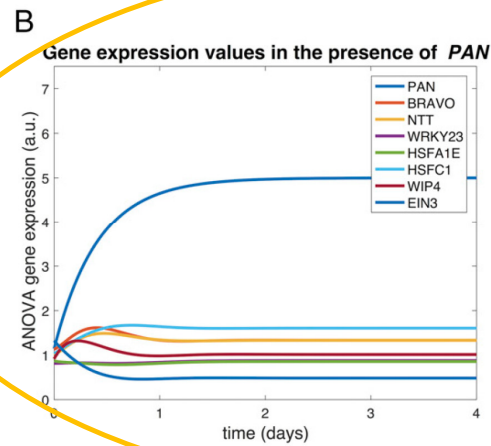


de Luis Balaguer et al., *PNAS*, 2017

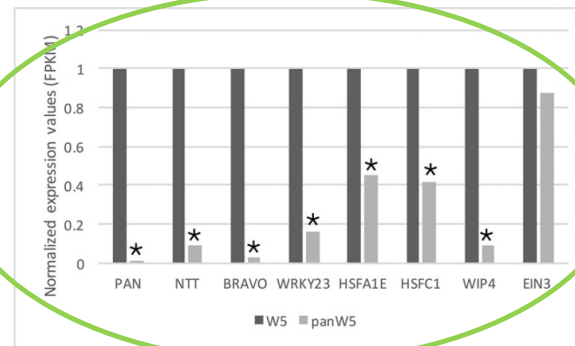
Genové regulační sítě - GENIST



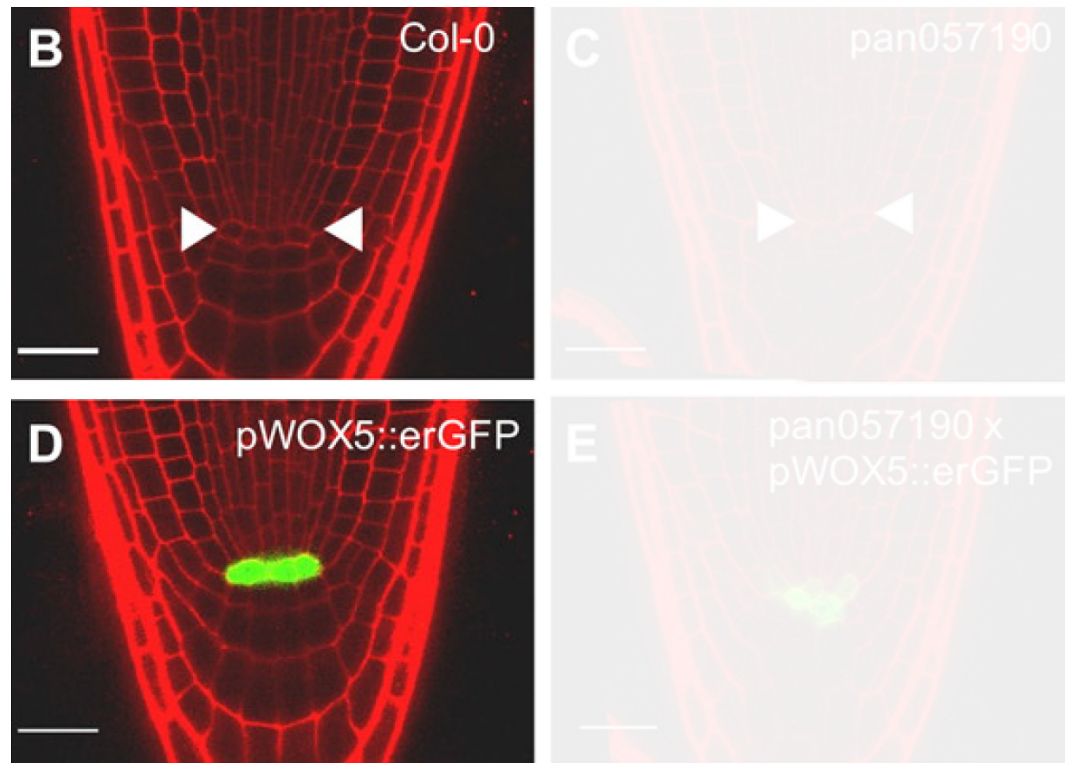
MODEL PREDICTION



feed-back loops



Identifikace PAN jako důležitého regulátoru identity QC



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

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“)