

NPO MUNI MSMT-16606/2022

## Analýza genové exprese a nástroje systémové biologie

Jan Hejátko

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

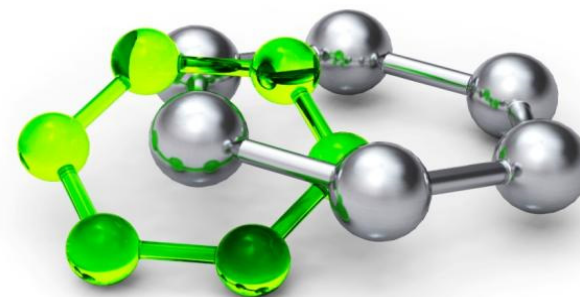
a

**Národní centrum pro výzkum biomolekul,**  
Přírodovědecká fakulta,

Masarykova univerzita, Brno

[hejatk@sci.muni.cz](mailto:hejatk@sci.muni.cz), [www.ceitec.eu](http://www.ceitec.eu)

**M U N I**  
**S C I**



# Zdrojová literatura

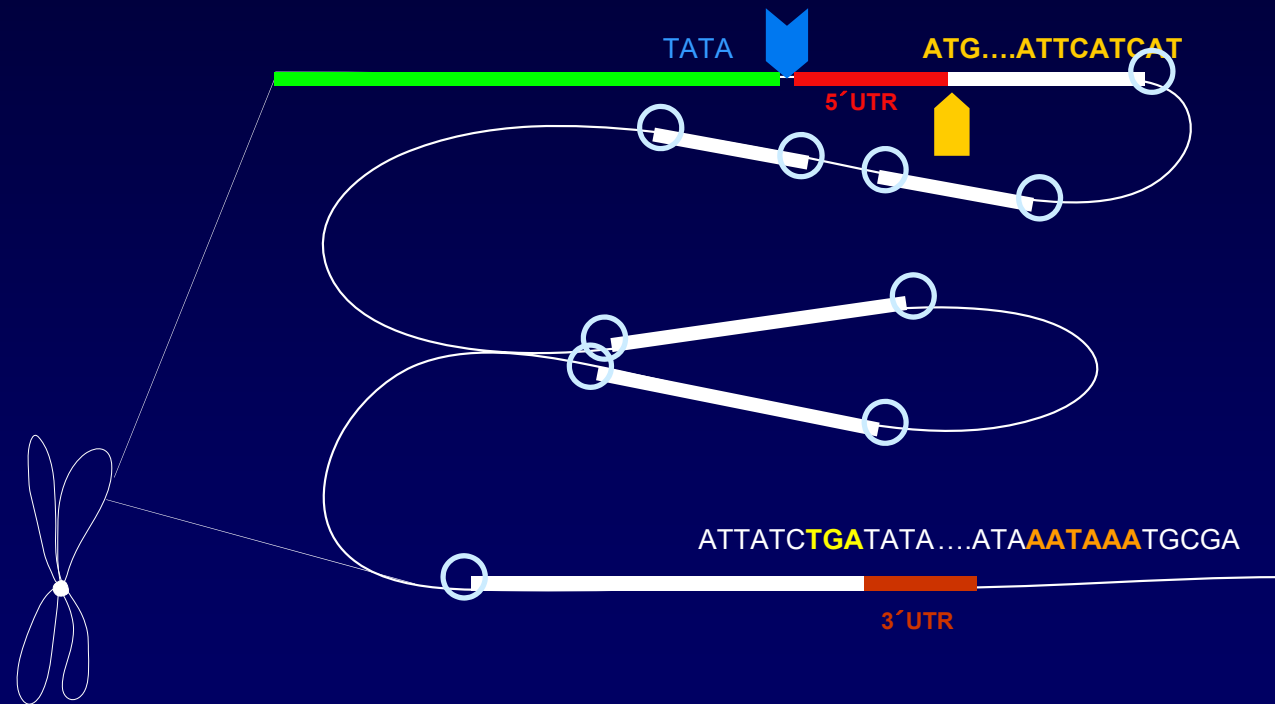
- Benitez, M. and Hejatko, J. Dynamics of cell-fate determination and patterning in the vascular bundles of *Arabidopsis thaliana* (submitted)
- Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. 2007. A high-resolution root spatiotemporal map reveals dominant expression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

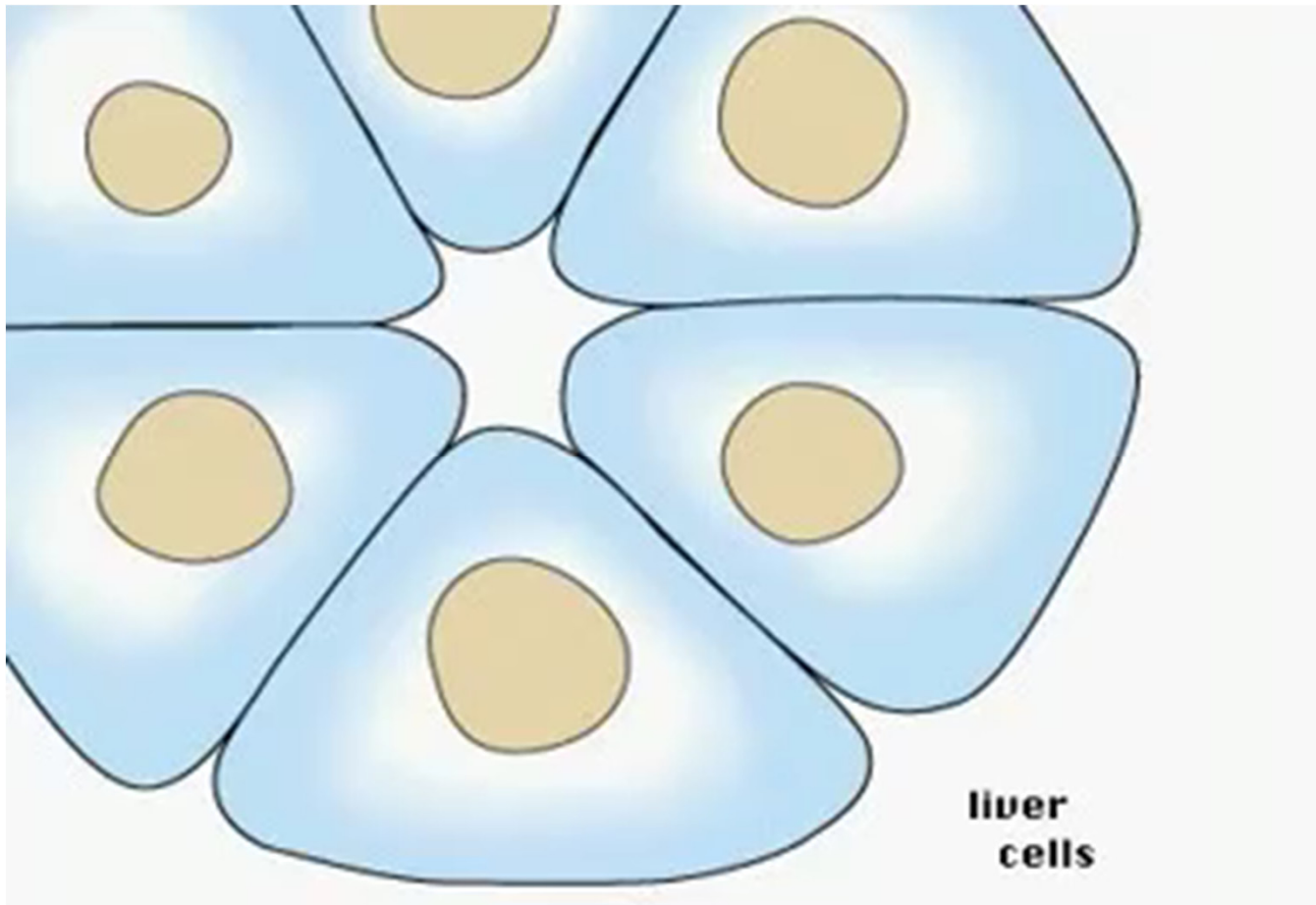
- promotor
- počátek transkripce
- 5' UTR
- počátek translace
- místa sestřihu
- stop kodon
- 3' UTR
- polyadenylační signál



# Osnova

- Struktura genů
- Metody analýzy genové exprese

# Časoprostorová specifita genové exprese

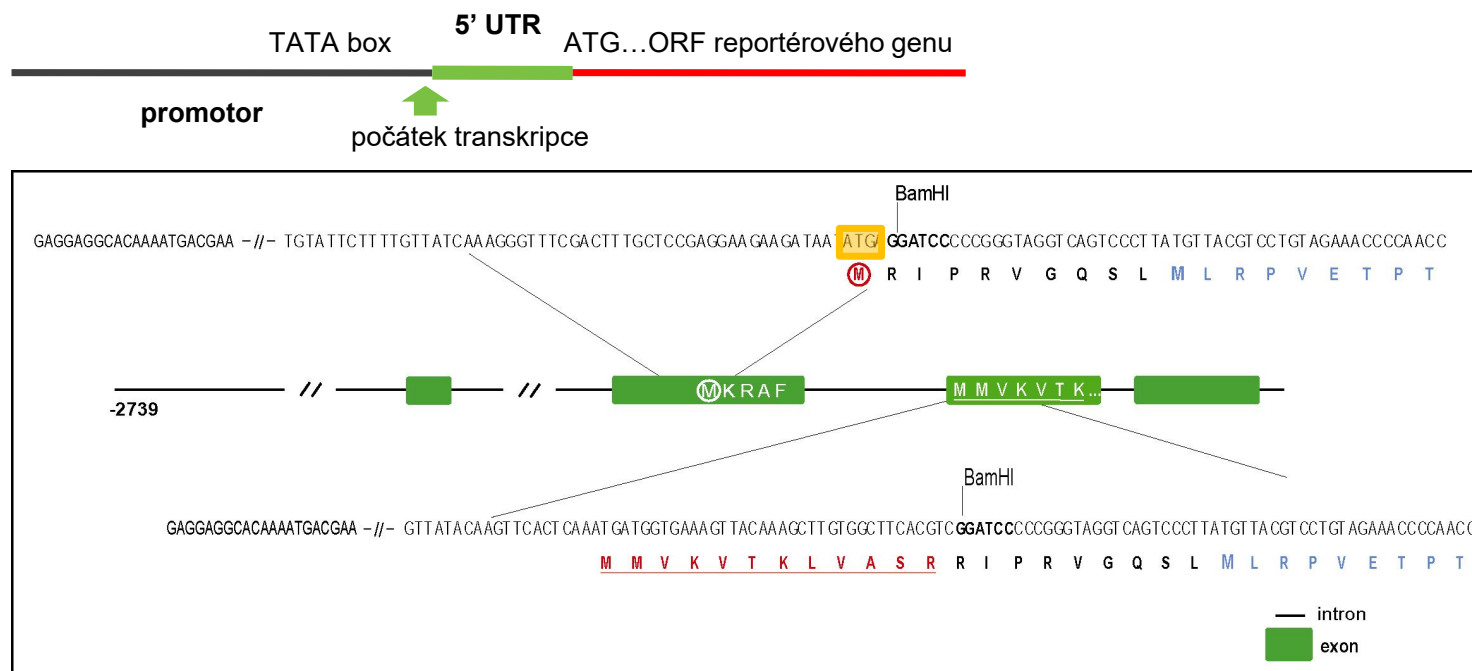


# Osnova

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

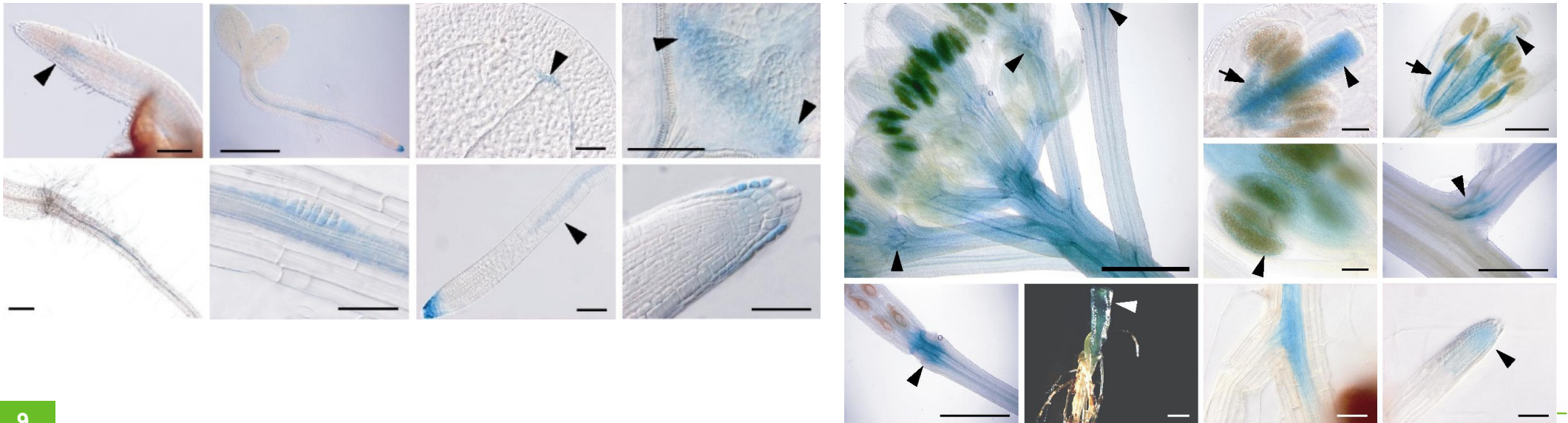
# Transkripční fúze

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

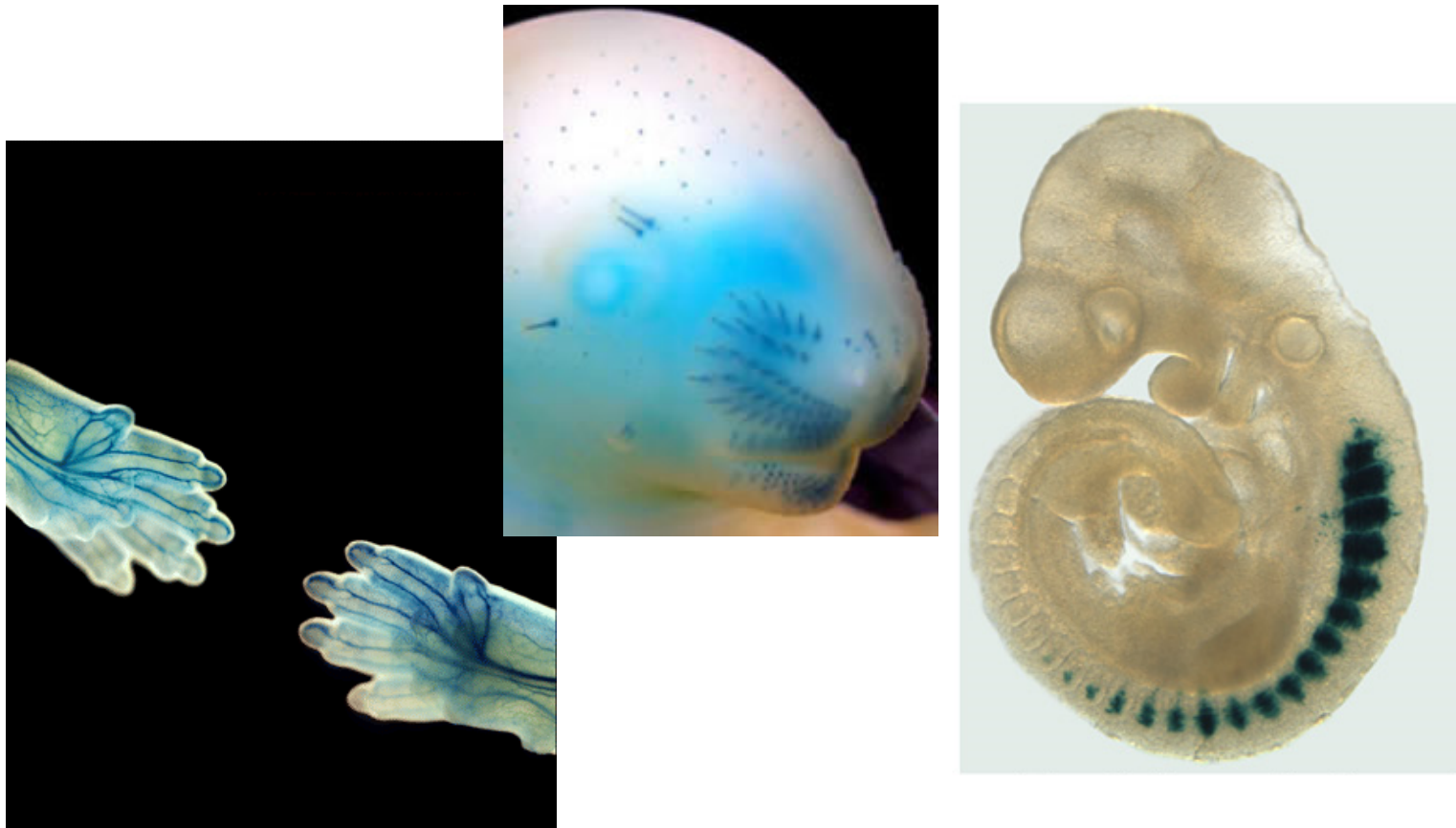


# Transkripční fúze

- **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

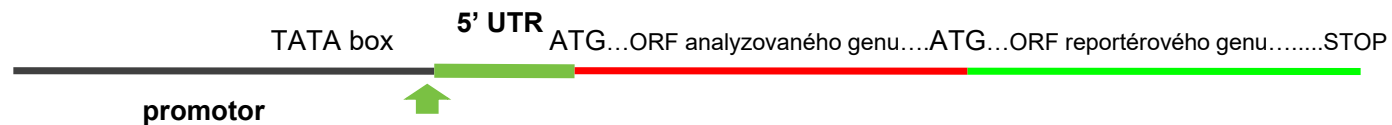


# Osnova

- 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

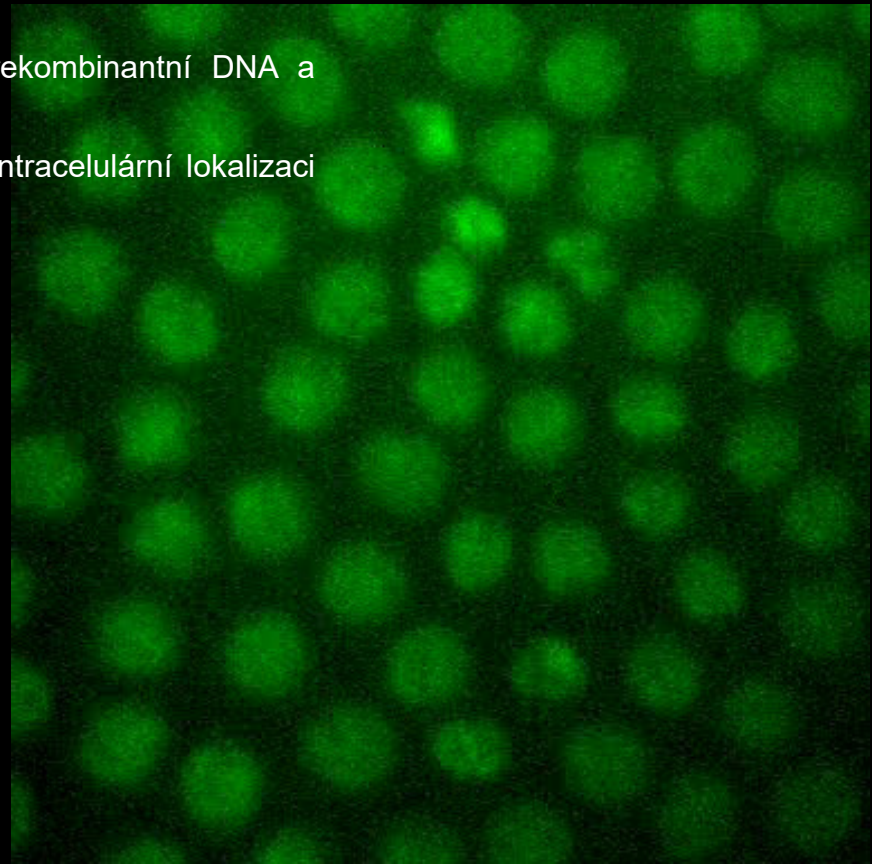
- **Příprava translační fúze kódující oblasti analyzovaného genu s reportérovým genem**
  - Identifikace a klonování **promotorové** a **kódující** oblasti analyzovaného genu
  - příprava **rekombinantní DNA** nesoucí **promotor** a **kódující** sekvenci **studovaného genu** ve fúzi s **reportérovým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

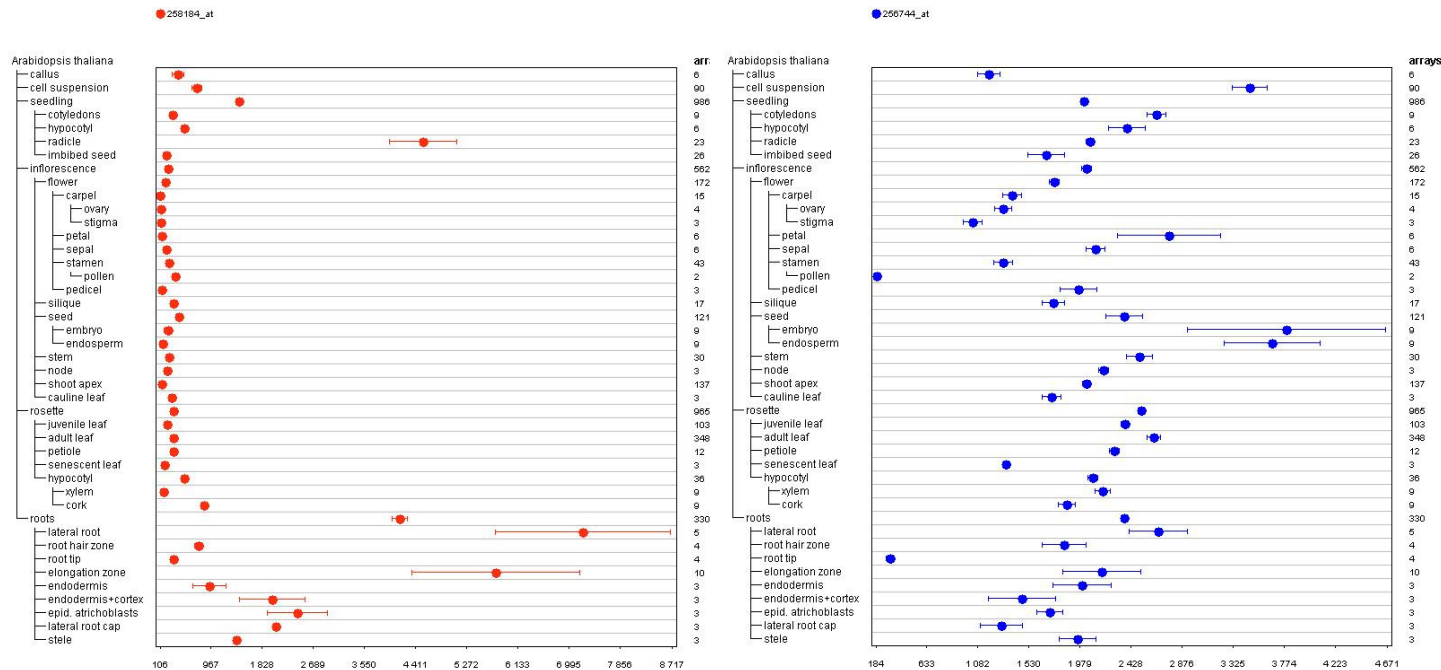


# Osnova

- 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**

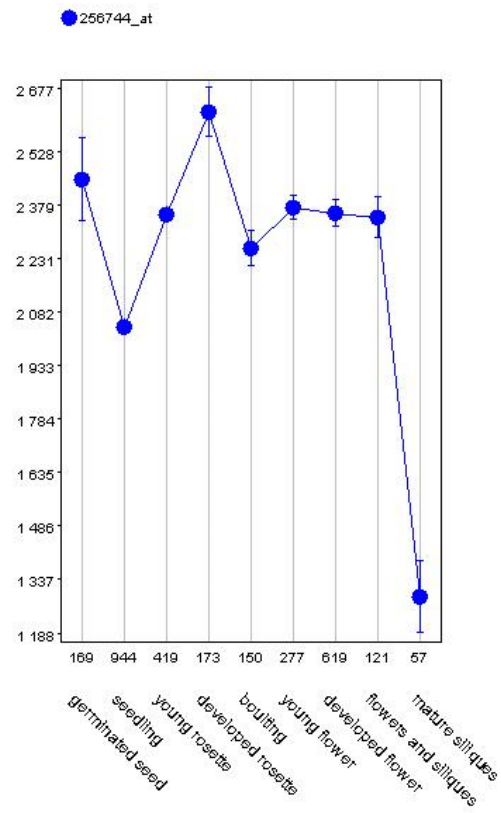
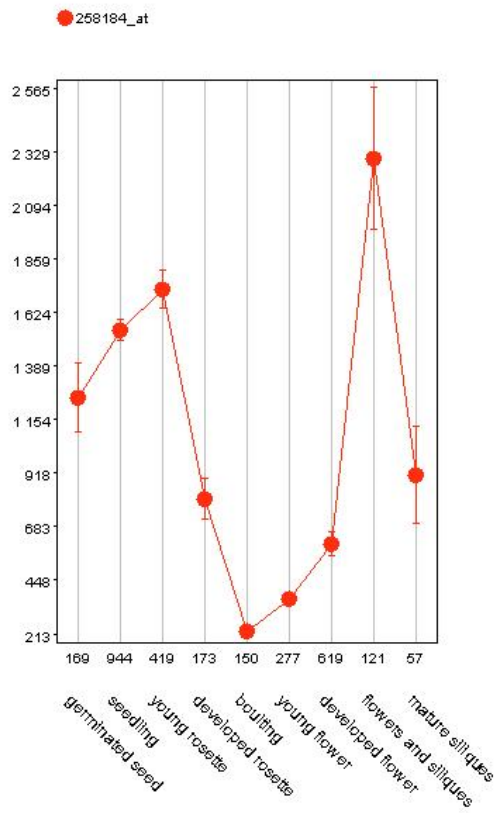
# Databáze

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



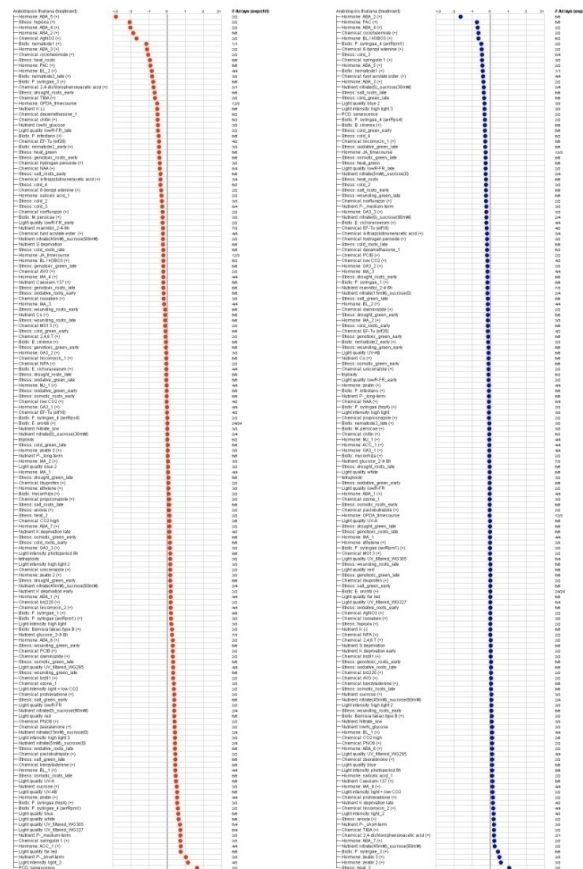
# Databáze

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



# Databáze

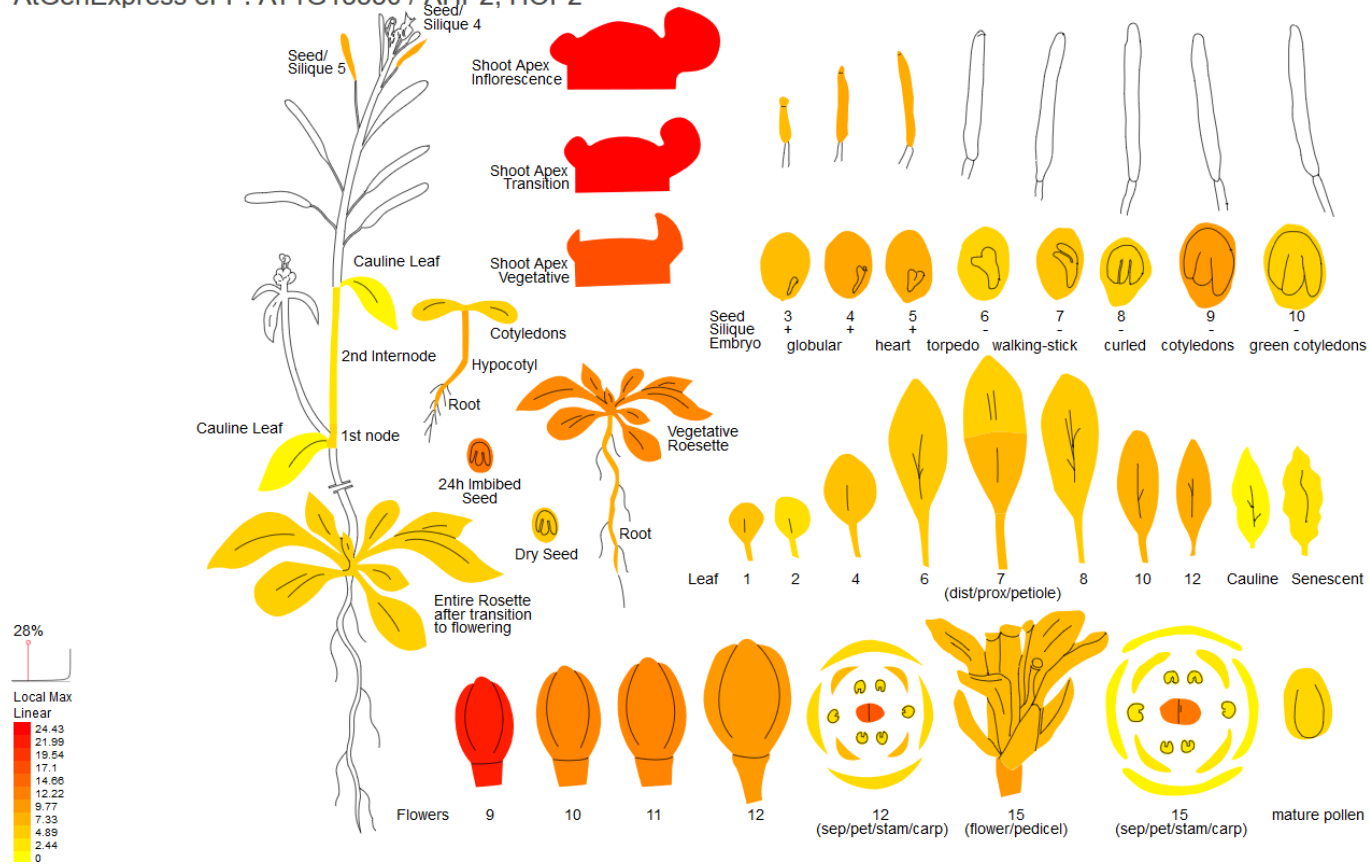
- Analýza exprese pomocí Genevestigator (**AHP1** a **AHP2**, *Arabidopsis*, Affymetrix ATH 22K Array)



# Databáze

- **Analýza exprese pomocí ePlant**

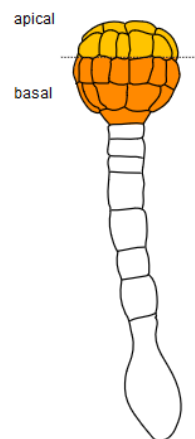
AtGenExpress eFP: AT1G13330 / AHP2, HOP2



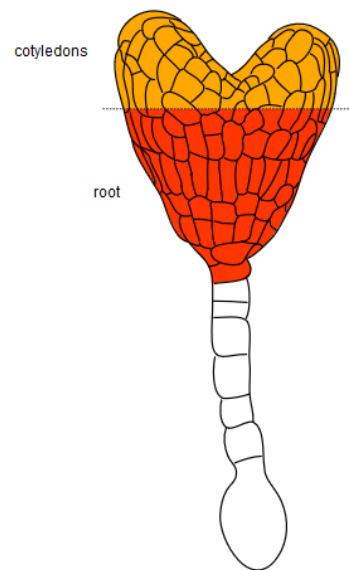
# Databáze

## □ Analýza exprese pomocí [ePlant](#)

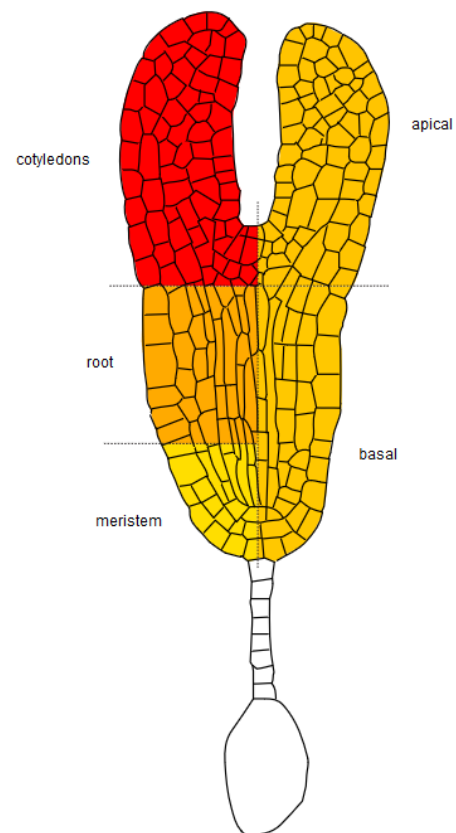
Globular



Heart



Torpedo

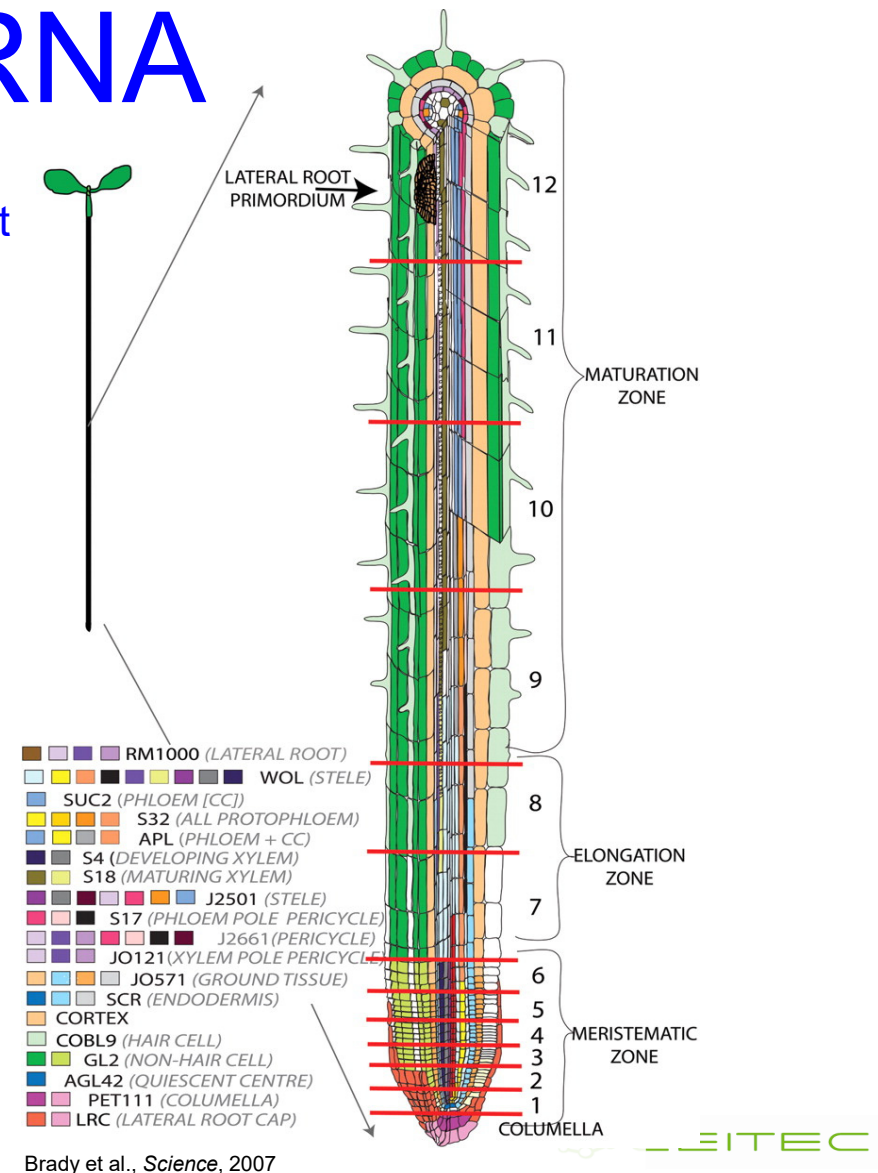
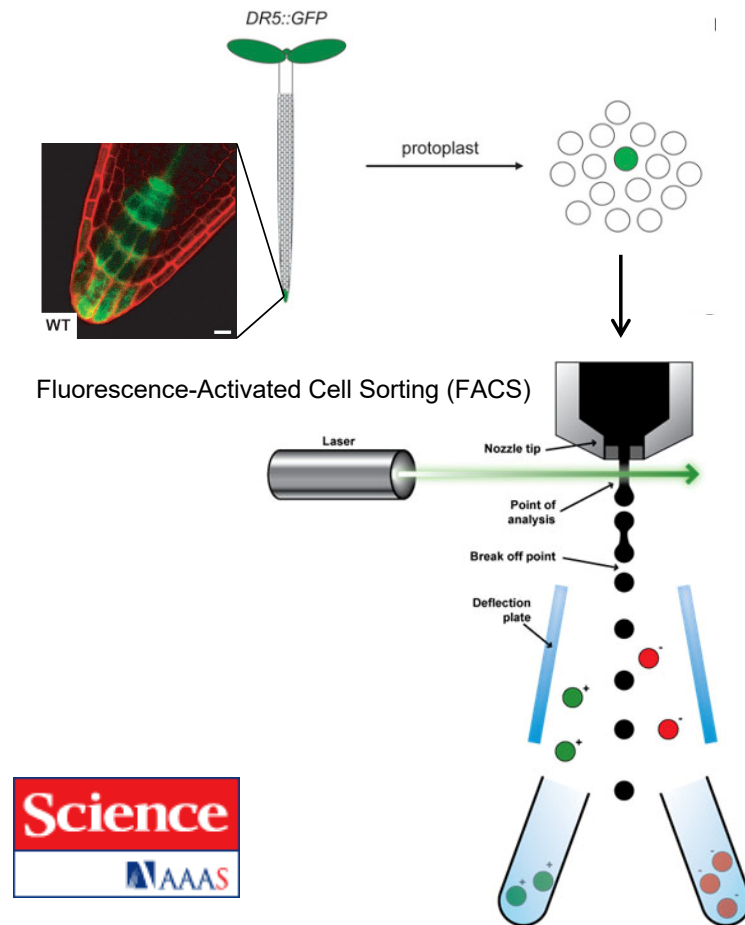


# Osnova

- **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

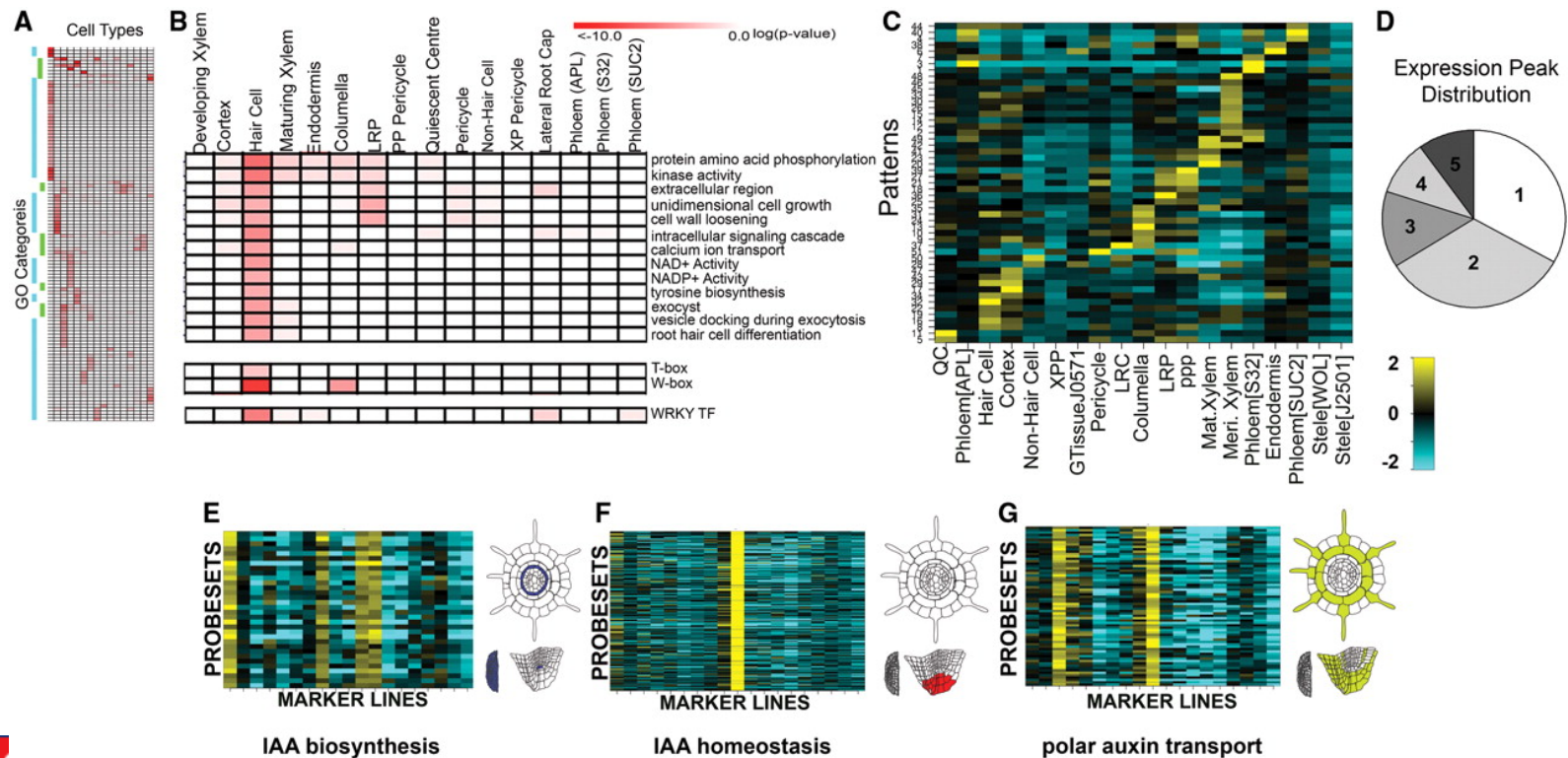
# Expression Maps - RNA

## □ High-Resolution Expression Map in Arabidopsis Root



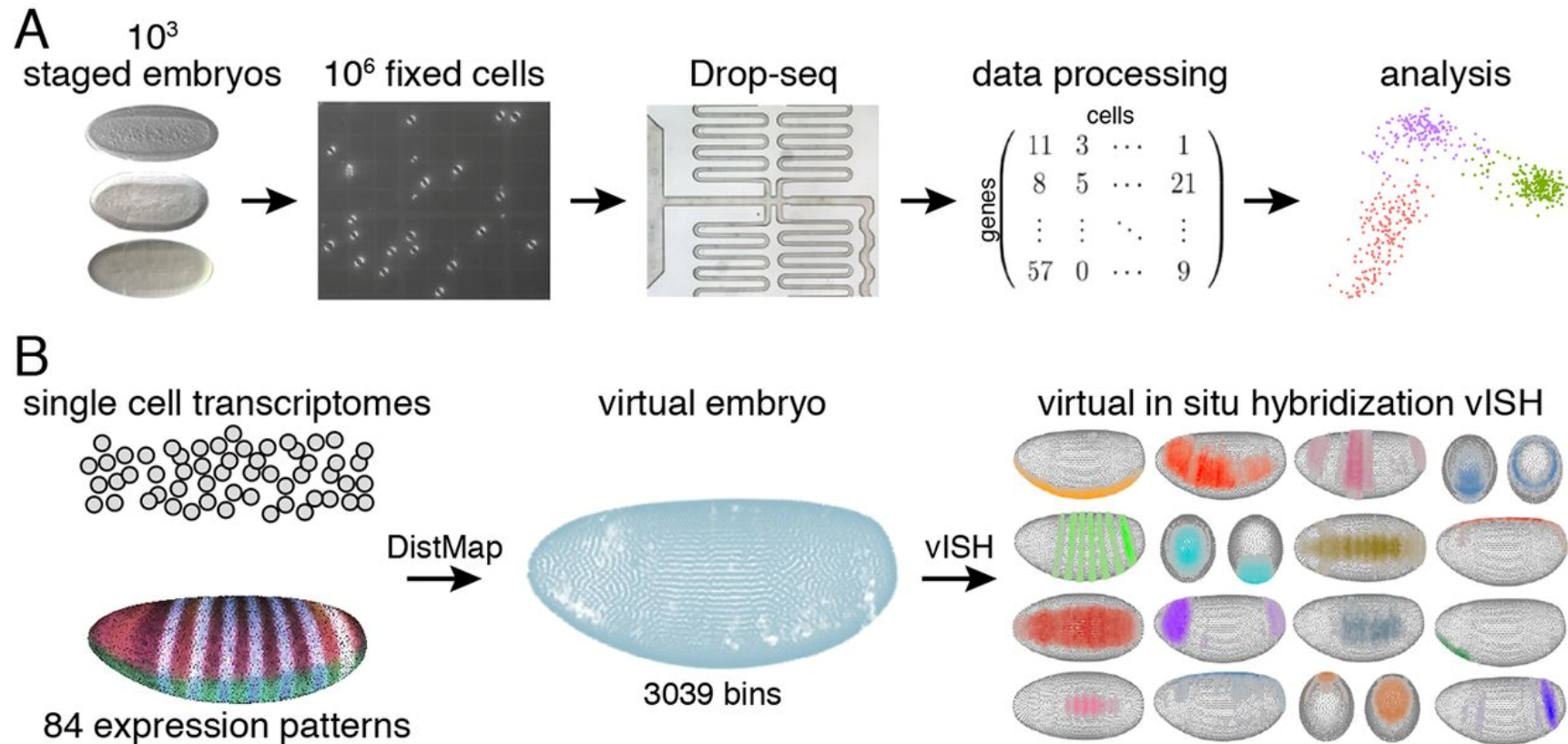
# Expression Maps - RNA

## High-Resolution Expression Map in Arabidopsis Root



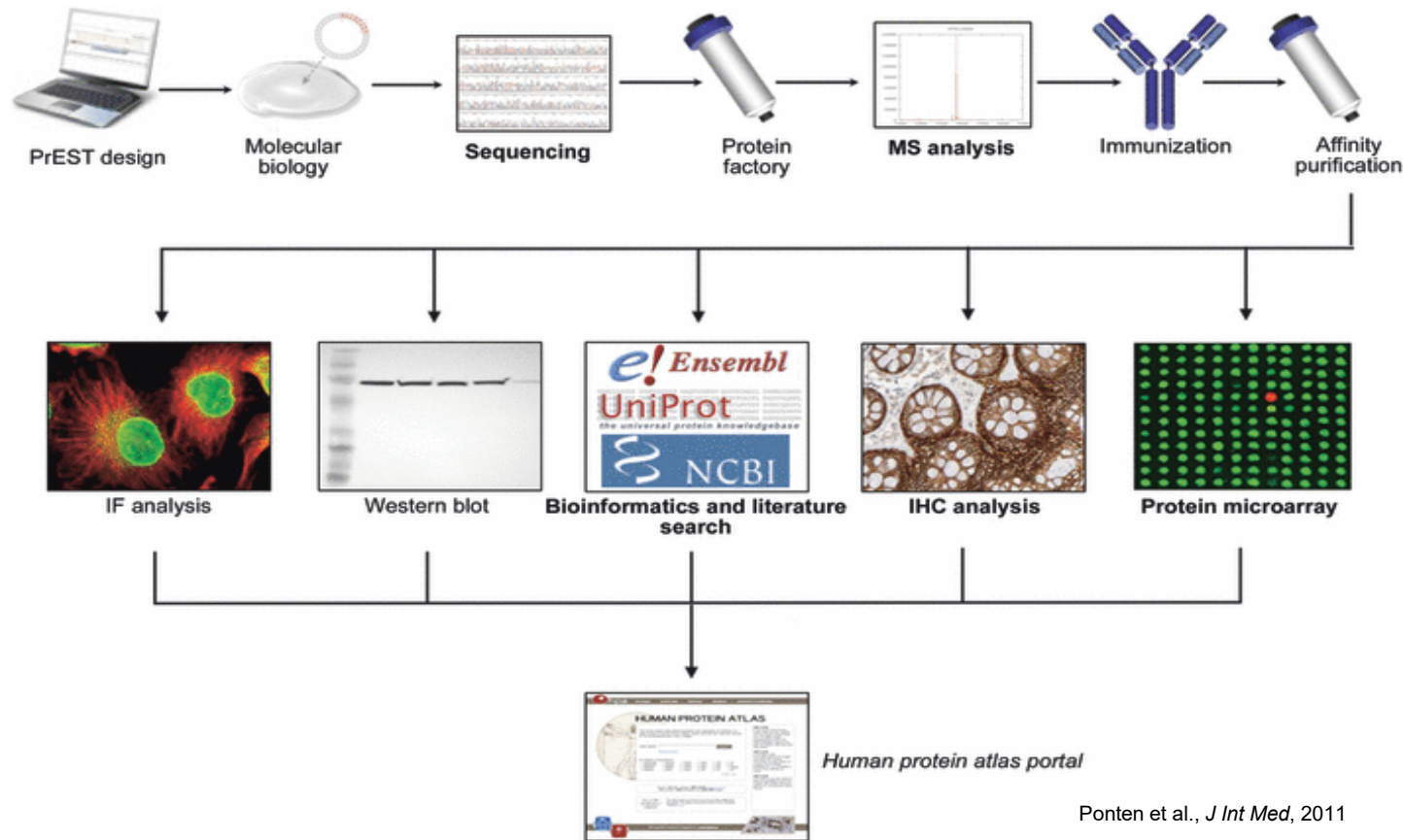
# Expression Maps - RNA

## □ High-Resolution Expression Map in Drosophilla



# Expression Maps - Proteins

## □ Human Protein Atlas



# Expression Maps - Proteins

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

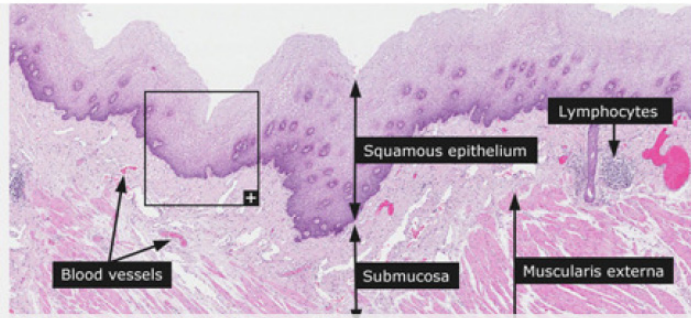
**THE HUMAN PROTEIN ATLAS**

ABOUT & HELP

SEARCH ? »

Search Clear Fields »

e.g. CD44, ELF3, KLK3, or use Fields to search specific fields such as protein\_class:Transcription factors or chromosome:X



dictionary: histology of esophagus

**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.

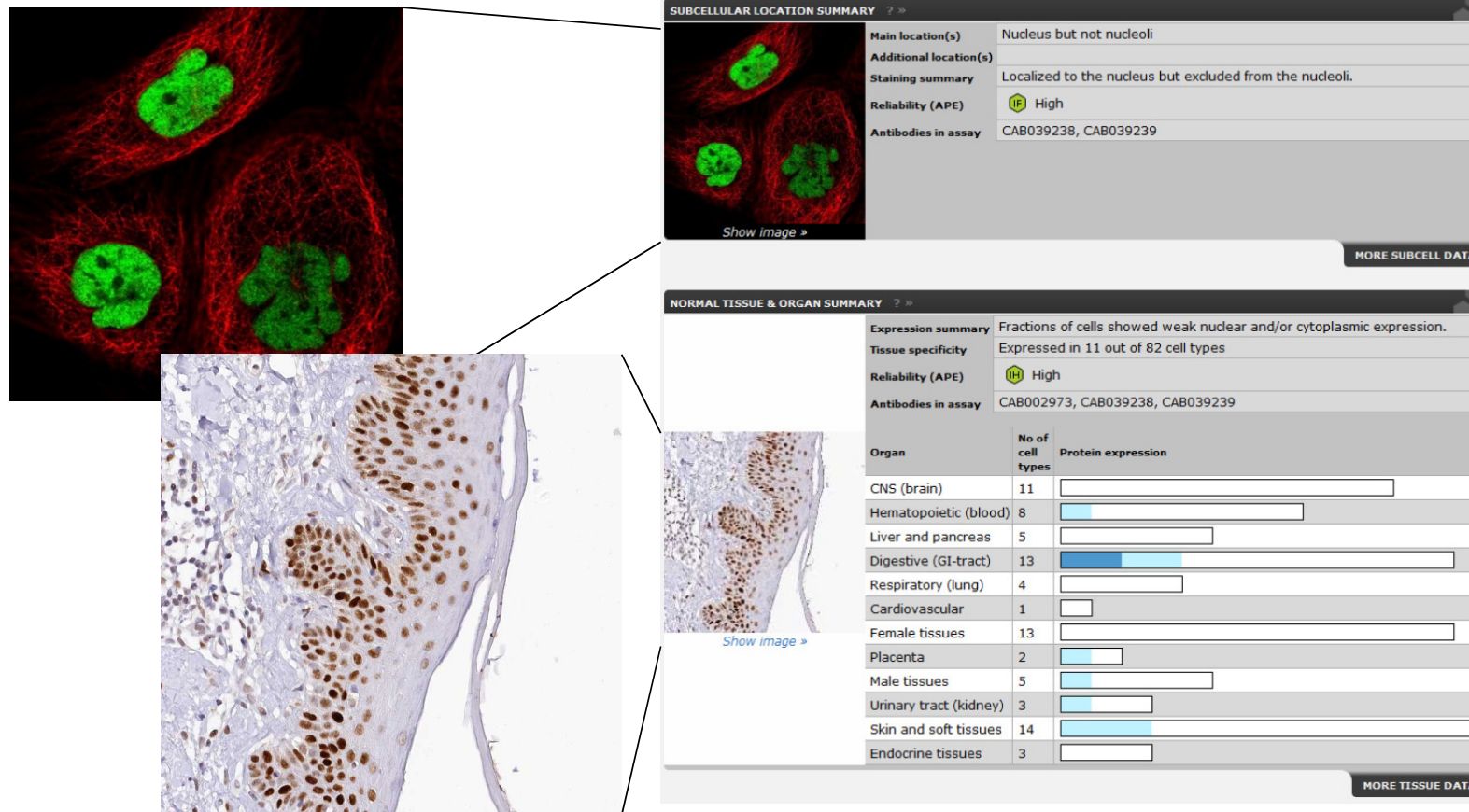
Knut och Alice  
Wallenberg  
Stiftelse

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



# Expression Maps - Proteins

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



# 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)

Affymetrix ATH1 *Arabidopsis* genome array

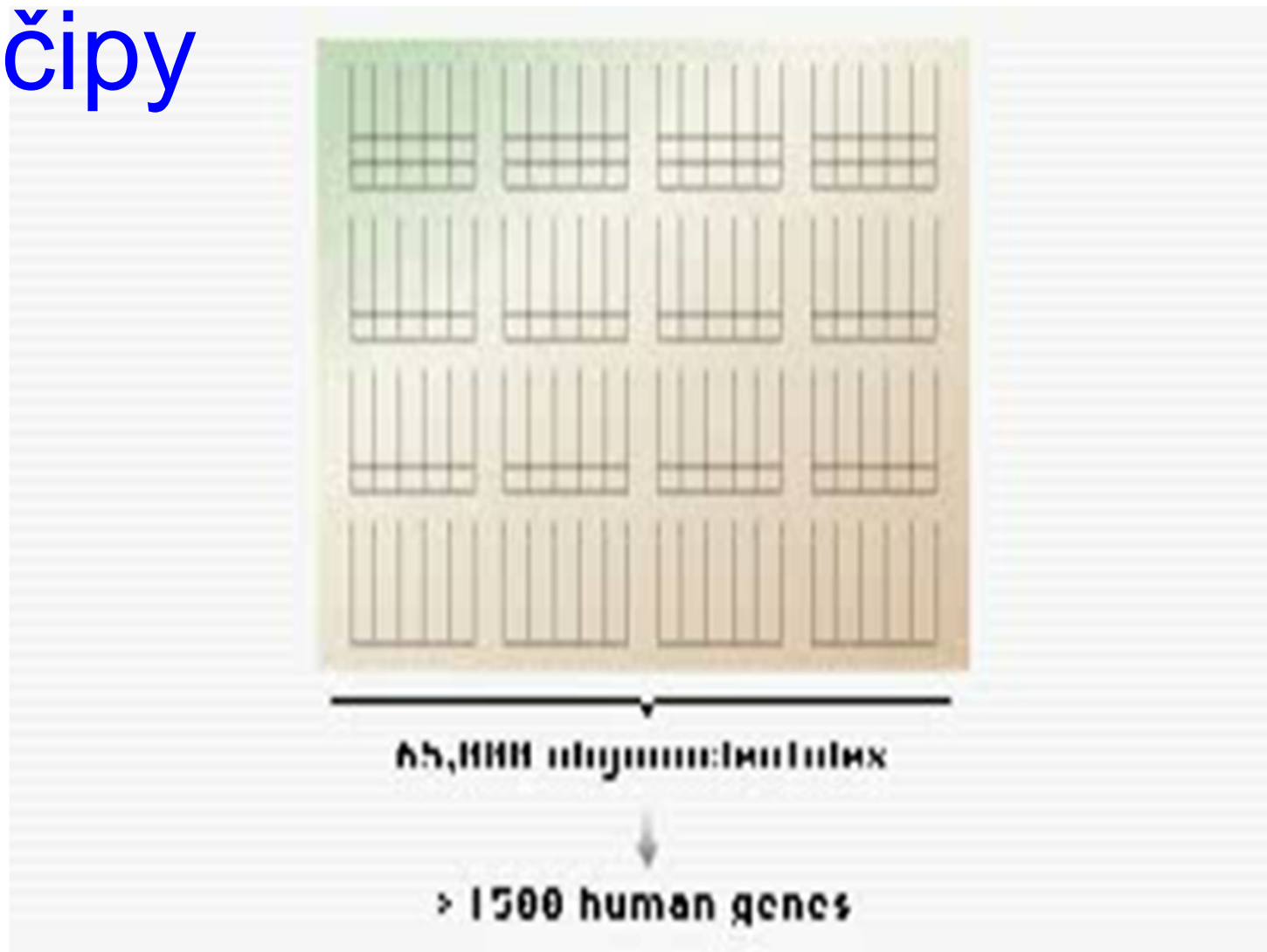


- čipy nejen pro analýzu exprese, ale např. i genotypování (SNP polymorfizmy, sekvenování pomocí čipů, ...)

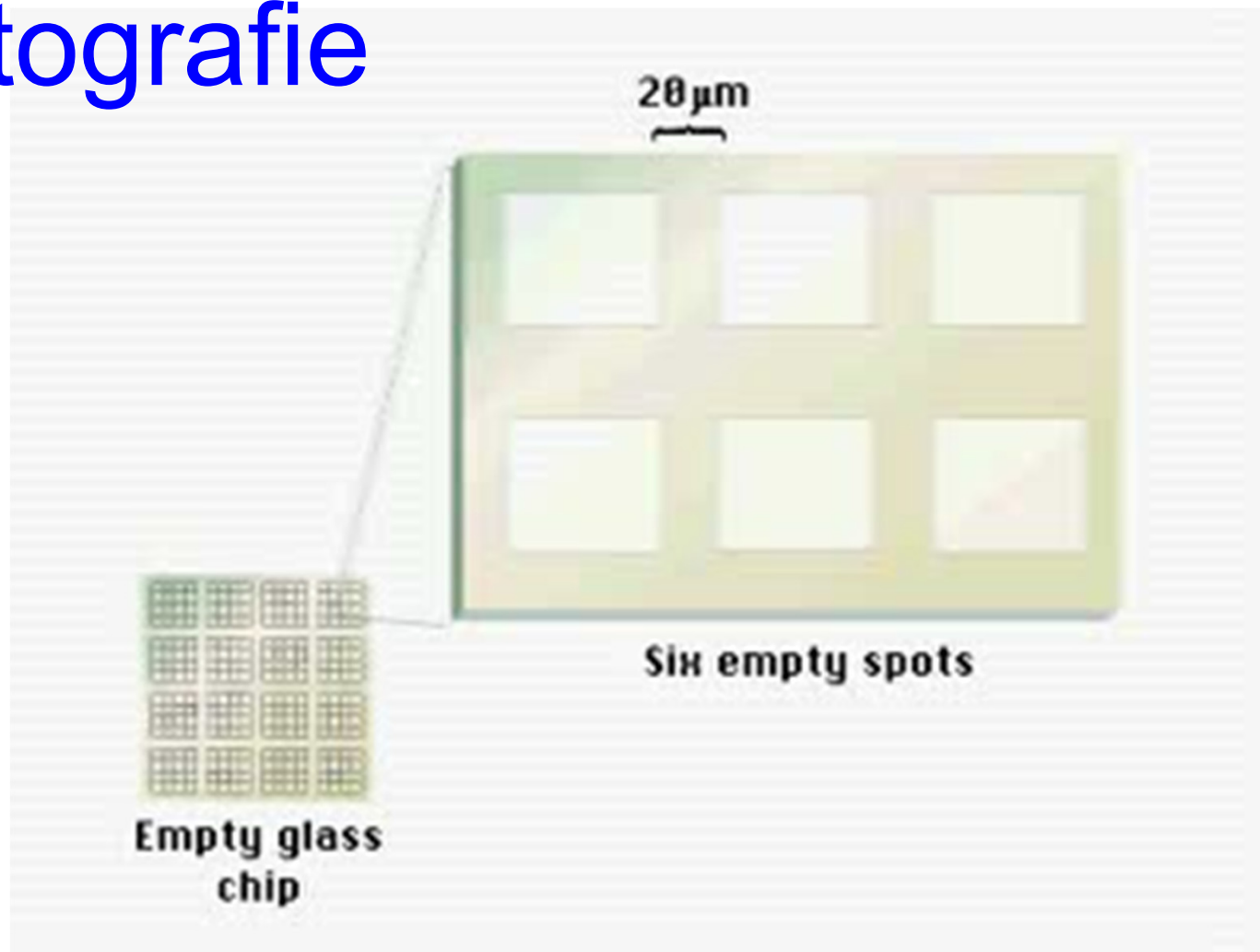


| Critical Specifications                                                                                                                                  |                                                                                                                                                                                                       |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Number of arrays                                                                                                                                         | One                                                                                                                                                                                                   |
| Number of sequence represented                                                                                                                           | >24,000 gene sequences                                                                                                                                                                                |
| Feature size                                                                                                                                             | 18 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> .<br><i>B. subtilis</i> gene <i>lysA</i> . Phage P1 <i>cre</i> gene.<br><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 195M6T7 in response to chemical treatment

TAIR 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 Details

| Slide (name & description)                                 | External ID | Replicate (id & name) | Replicate type | Reverse replicate | Sample            | Experimental variables                             | Label | Get Data |
|------------------------------------------------------------|-------------|-----------------------|----------------|-------------------|-------------------|----------------------------------------------------|-------|----------|
| HoekengaS7<br>["]: 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 AICl3, pool of 3, 8, and 24 hours) | Cy5   | Download |
| HoekengaS8<br>Aluminum Stress 2 (strong spatial bias)      | AFGC: 7305  | 64: Aluminum Stress   | technical      | 63                | 7304_Cy5 7305_Cy3 | Aluminum (50 5M AICl3, pool of 3, 8, and 24 hours) | Cy3   | Download |
|                                                            |             |                       |                |                   | 7304_Cy3 7305_Cy5 | no treatment (pool of 3, 8, and 24 hours)          | Cy5   | Download |

- 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**

# 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*).

# Systémová biologie - definice

Definice dle [Dr. Nathana Price](#),

zástupce ředitele Ústavu pro systémovou biologii, Seattle, USA; [https://www.youtube.com/watch?v=OrXRI\\_8UFHU](https://www.youtube.com/watch?v=OrXRI_8UFHU).



# Výsledky –omických studií vs. biologicky relevantní závěry

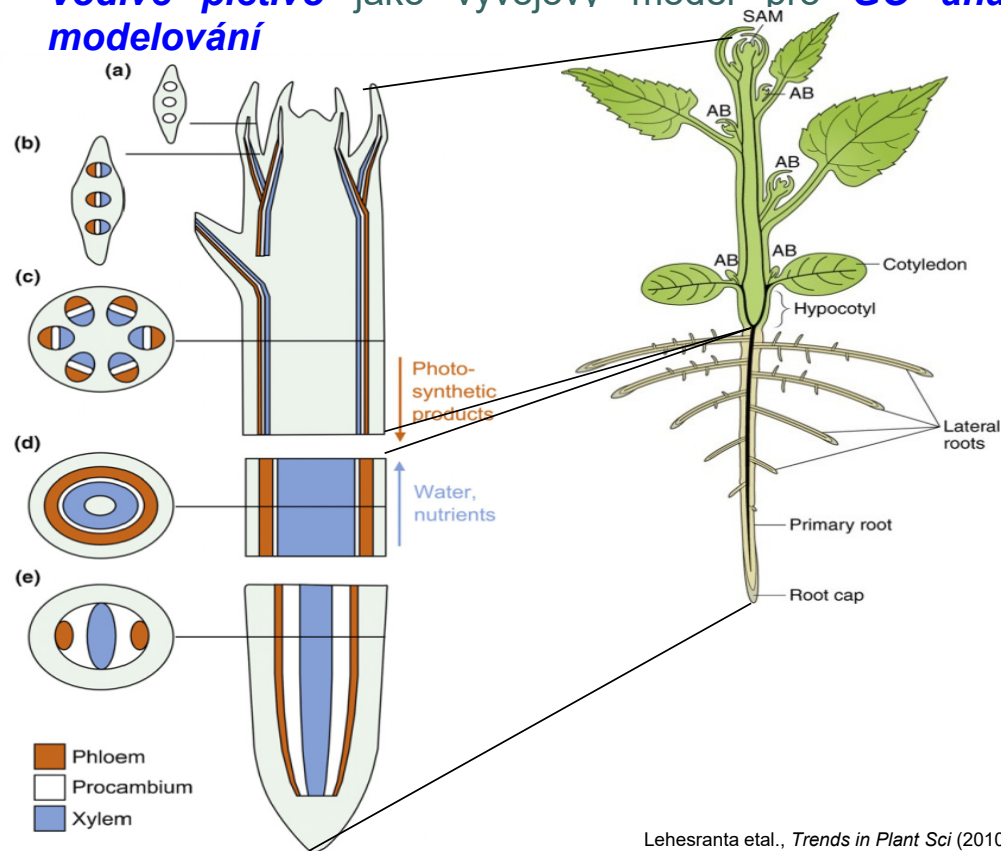
- Výsledky –omických studií reprezentují **enormní množství dat**, např. geny s rozdílnou expresí. Ale jak z nich získat **biologicky relevantní závěry**?

Dii, 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**



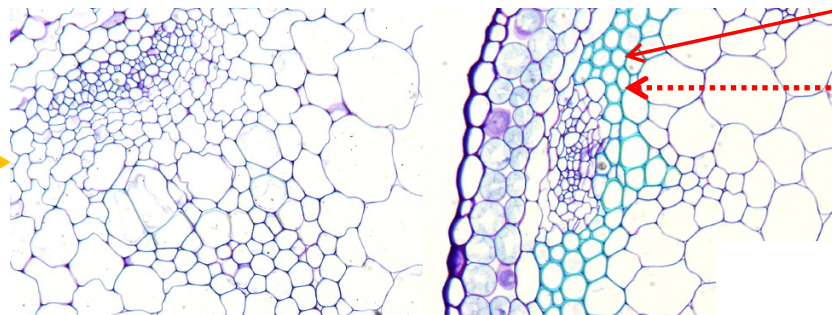
Lehesranta et al., *Trends in Plant Sci* (2010)

# Hormonální regulace vývoje rostlinných vodivých pletiv

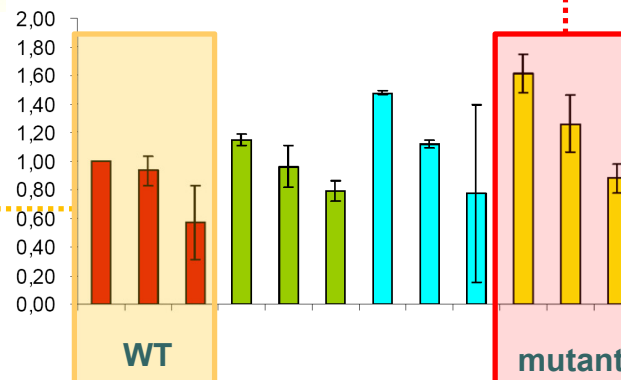
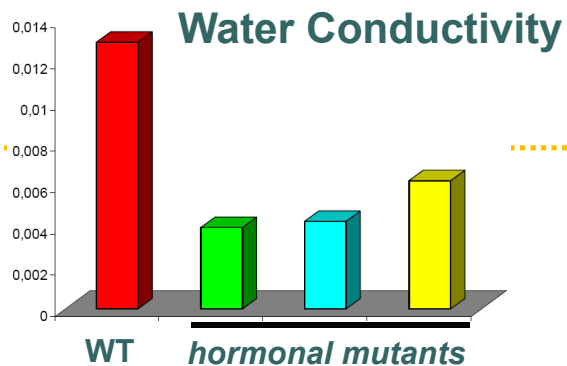
- Rostlinné hormony regulují ukládání ligninu v buněčných stěnách a transport vody xylemem

WT

*hormonal mutant*

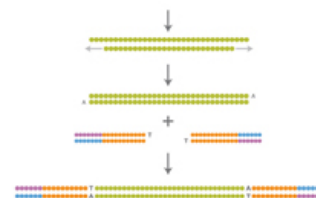
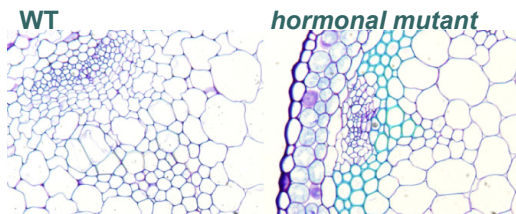


Acid-insoluble lignins



# 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)



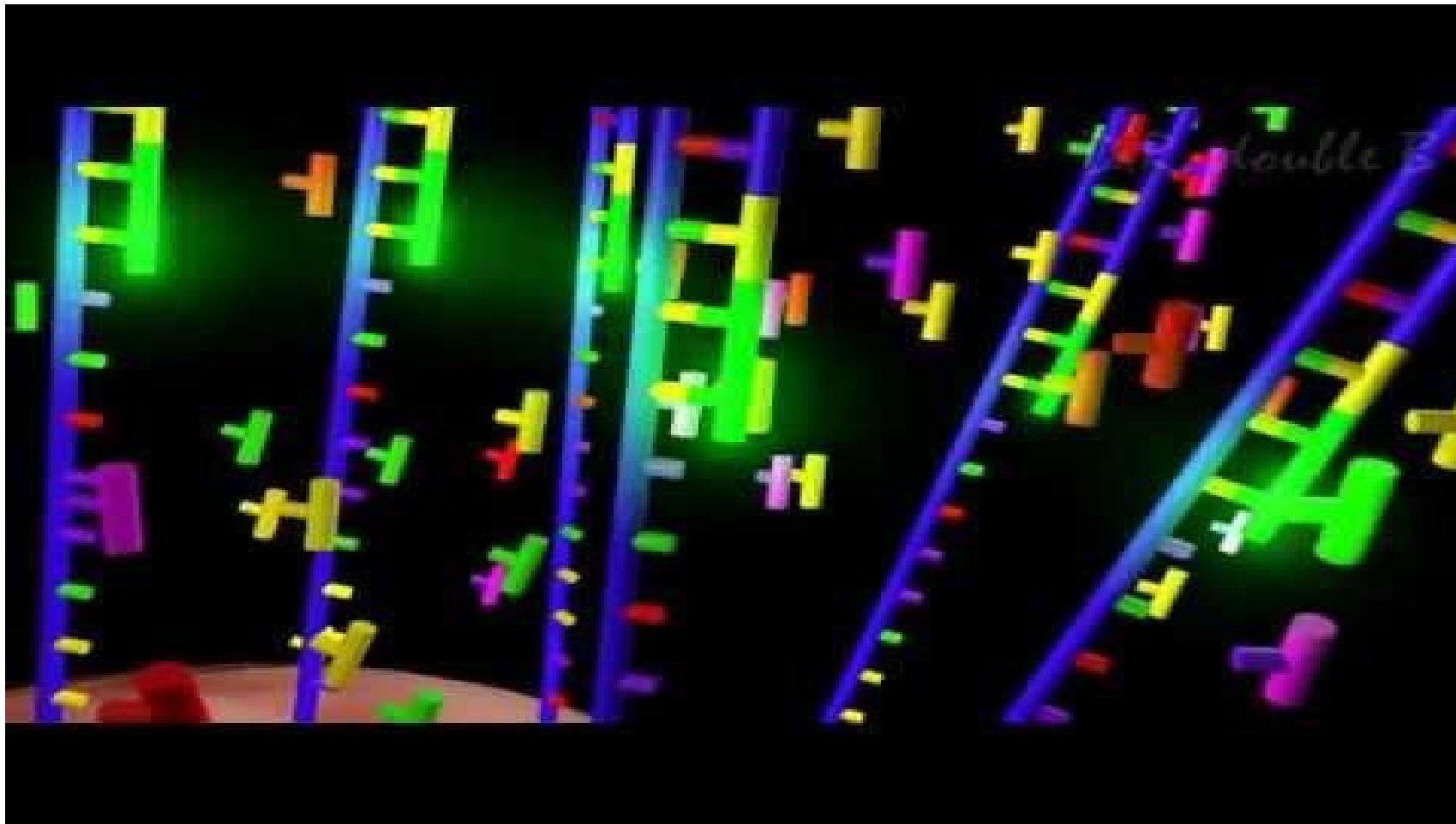
Sequencing by Synthesis  
~1.5 to 11 days



CASAVA  
2 days (30 min hands-on)

Sekvenování 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         |

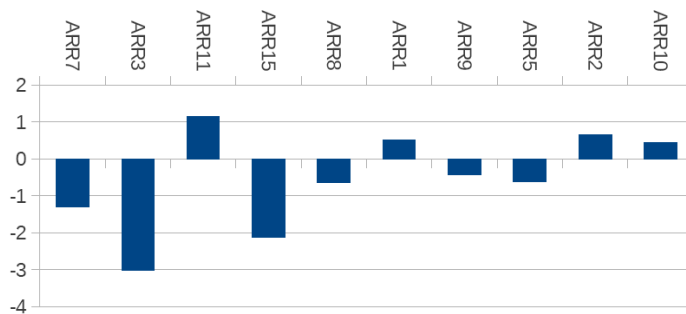
# Osnova

- **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**

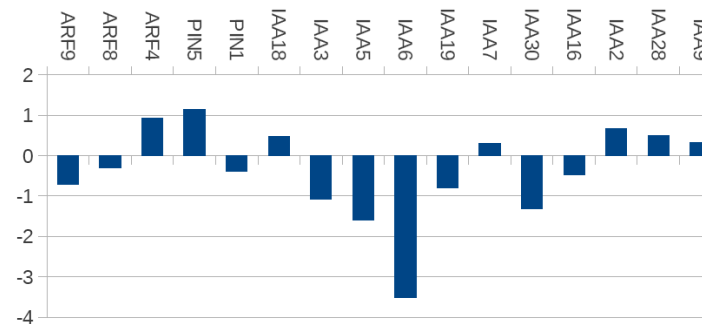
# Analýza genové ontologie

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

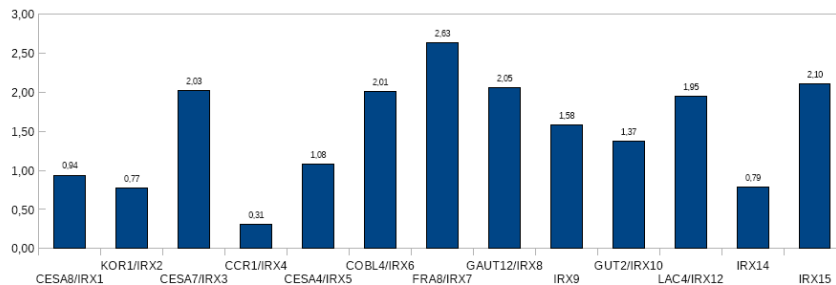
ARR differential expression



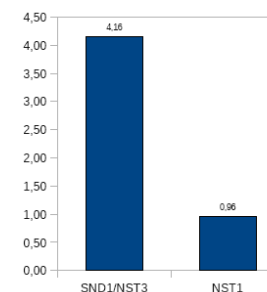
Differentially expressed IAA-related genes



IRXs



XYLEM MARKERS



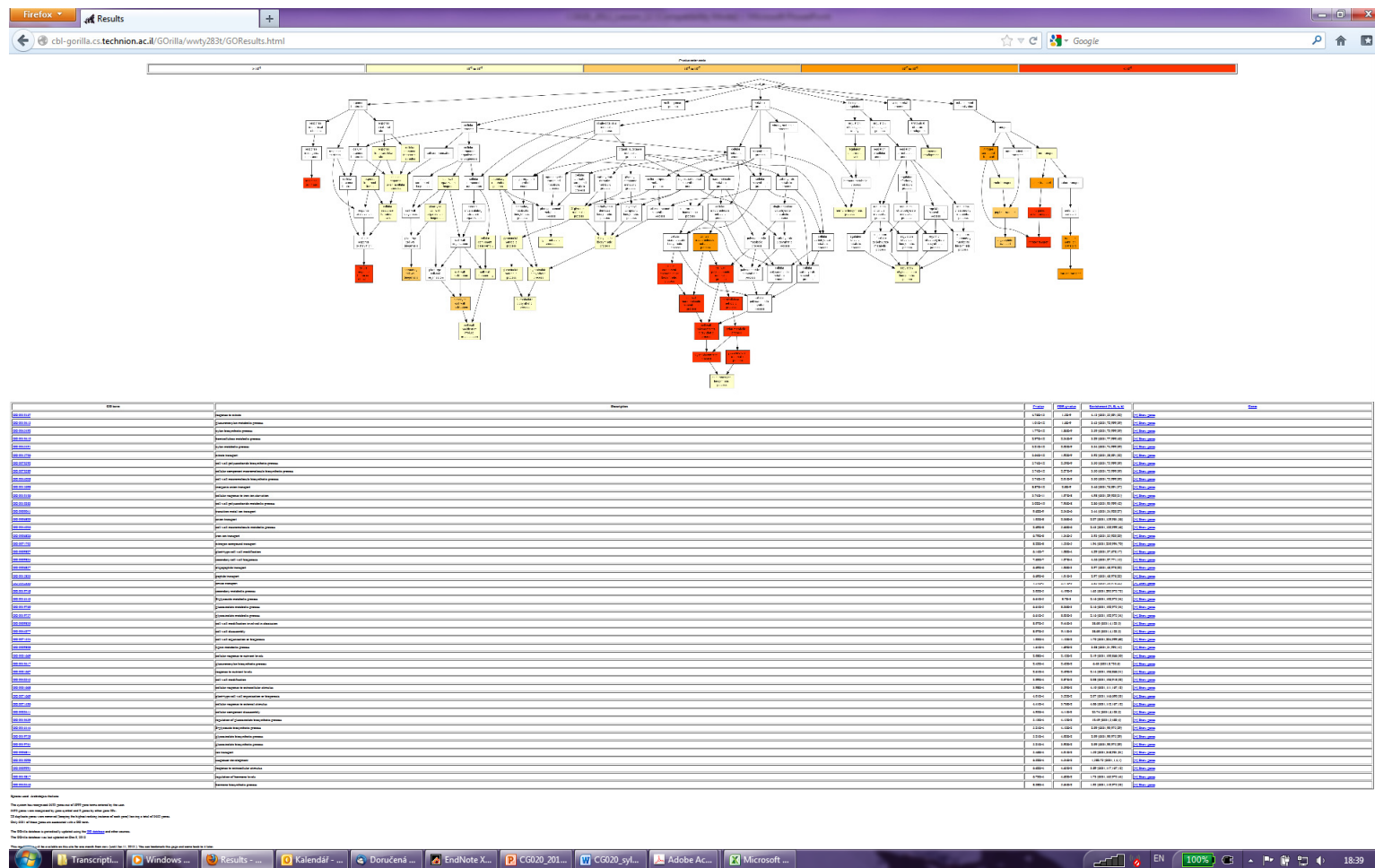
# Analýza genové ontologie

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

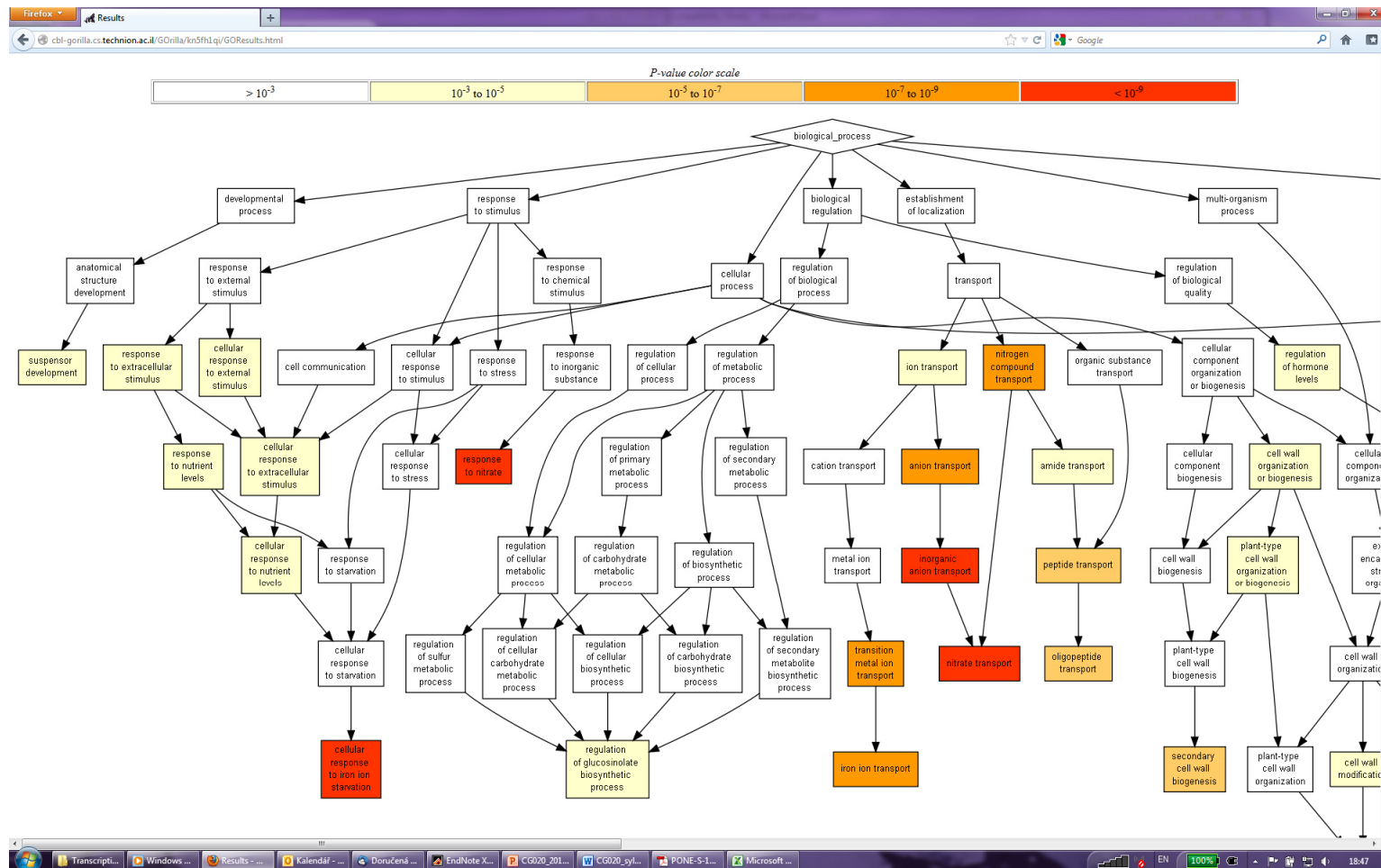
Eden et al., *BMC Bioinformatics* (2009)

The screenshot shows the GOrilla web interface in a Firefox browser window. The page title is "GOrilla - a tool for identifying enriched ...". The URL is "cbl-gorilla.cs.technion.ac.il". The main heading is "GOrilla" with the subtitle "Gene Ontology enRichment anaLysis and visualiZation tool". Below the heading is a logo featuring a gorilla and a network diagram. The text describes GOrilla as a tool for identifying and visualizing enriched GO terms in ranked lists of genes, and lists 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. It also provides links for "Running example", "Usage instructions", "GOrilla News (Updated December 3rd 2012)", and "References". The interface is divided into four steps: Step 1: Choose organism (Arabis thaliana selected), Step 2: Choose running mode (Single ranked list of genes selected), Step 3: Paste a ranked list of gene/protein names (with a text area and an upload button), and Step 4: Choose an ontology (Process selected). A "Search Enriched GO terms" button is at the bottom, along with a "Reset form" button. The browser's taskbar at the bottom shows various open applications like Transcript..., Windows..., GOrilla..., Kalendar..., DoruCent..., EndNote X..., C020\_201..., C020\_sy..., Adobe AC..., and Microsoft...

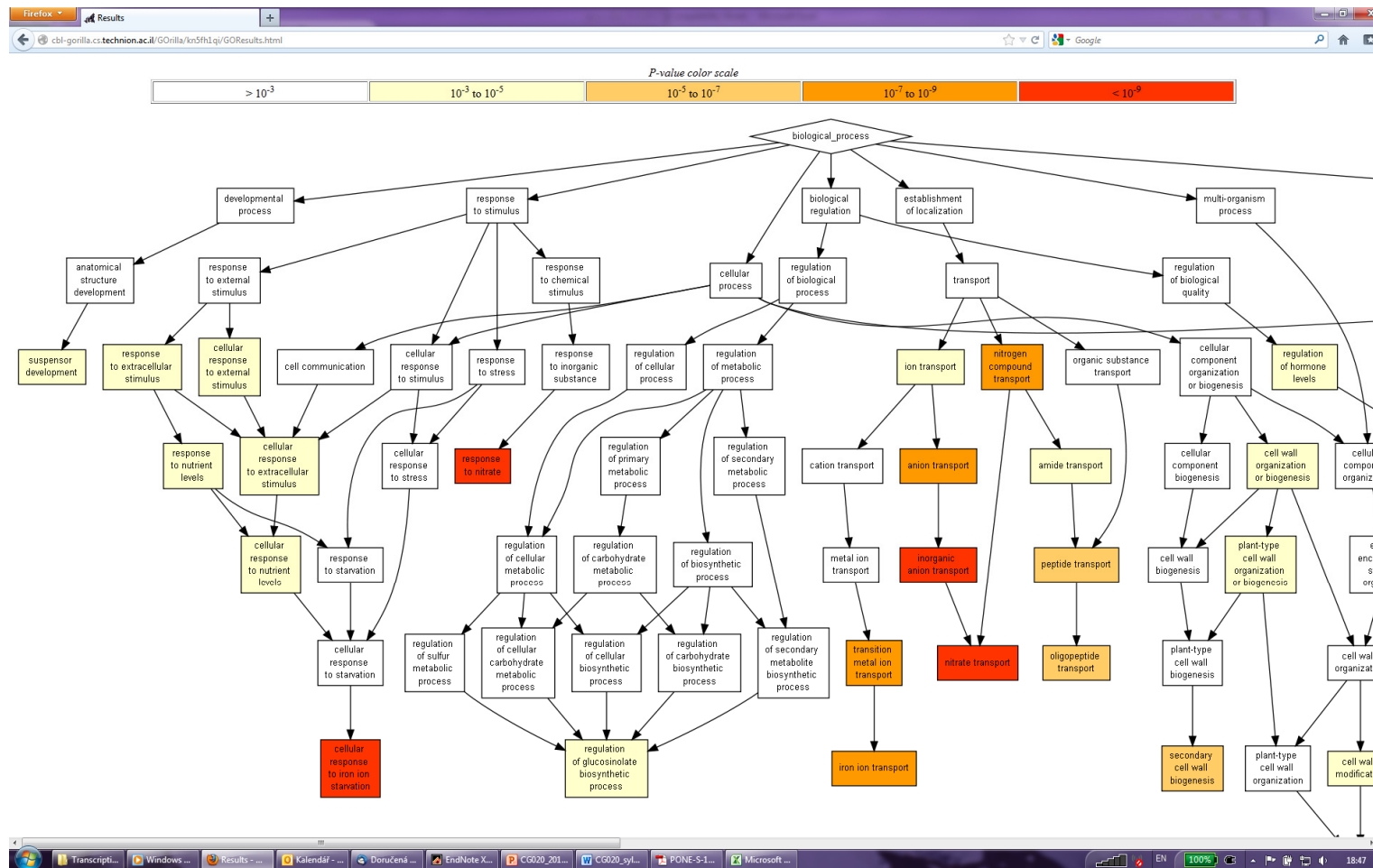
# GORILLA



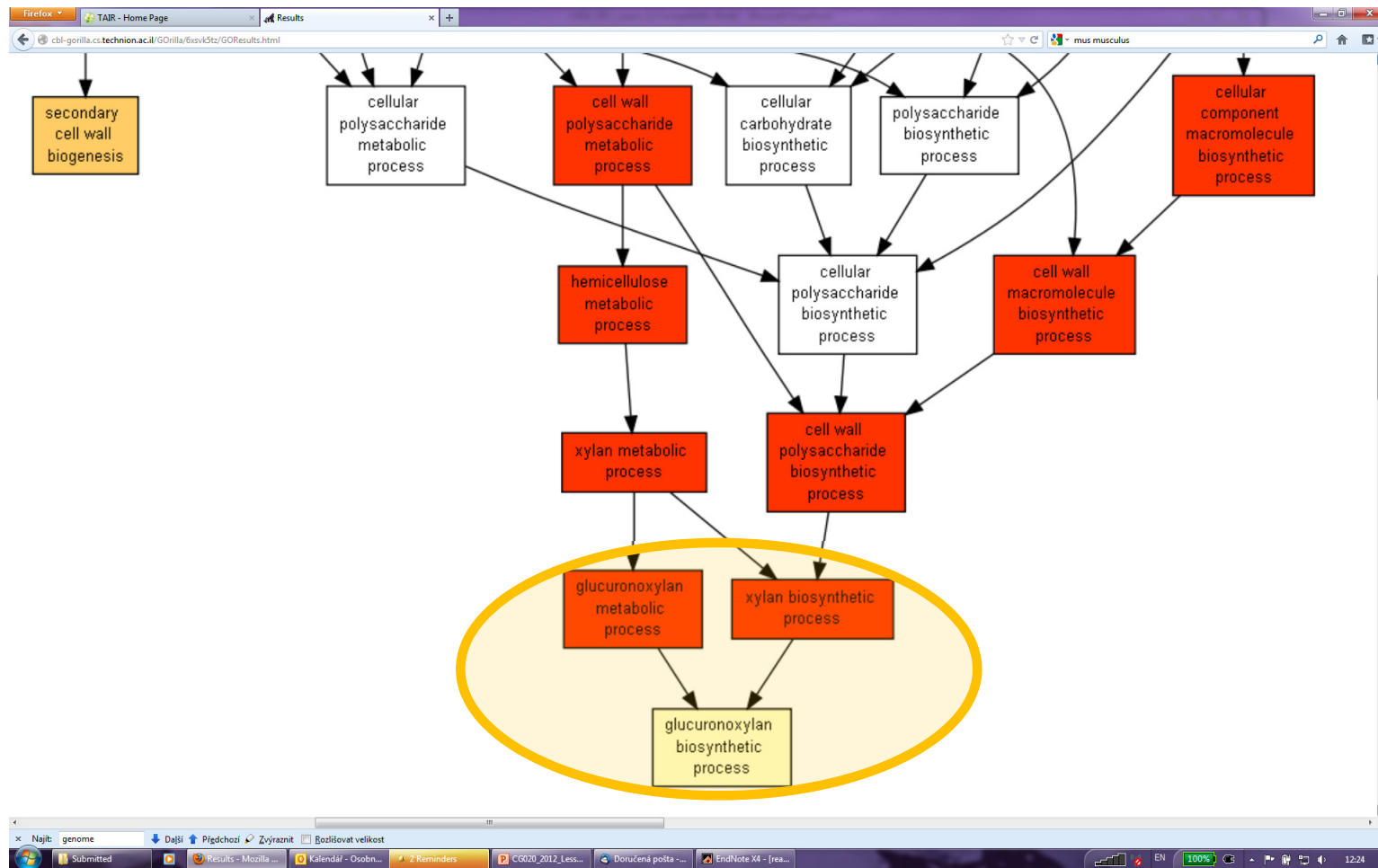
# GORILLA



# GORILLA



# GORILLA



# GORILLA

|                                                       |          |         |                       |                                |
|-------------------------------------------------------|----------|---------|-----------------------|--------------------------------|
| glucuronoxylan metabolic process                      | 1.01E-12 | 1.6E-9  | 3.43 (6331,72,999,39) | [+] <a href="#">Show genes</a> |
| xylan biosynthetic process                            | 1.77E-12 | 1.86E-9 | 3.39 (6331,73,999,39) | [+] <a href="#">Show genes</a> |
| hemicellulose metabolic process                       | 2.97E-12 | 2.34E-9 | 3.29 (6331,77,999,40) | [+] <a href="#">Show genes</a> |
| xylan metabolic process                               | 3.21E-12 | 2.03E-9 | 3.34 (6331,74,999,39) | [+] <a href="#">Show genes</a> |
| nitrate transport                                     | 3.64E-12 | 1.92E-9 | 3.92 (6331,58,891,32) | [+] <a href="#">Show genes</a> |
| cell wall polysaccharide biosynthetic process         | 5.74E-12 | 2.59E-9 | 3.30 (6331,75,999,39) | [+] <a href="#">Show genes</a> |
| cellular component macromolecule biosynthetic process | 5.74E-12 | 2.27E-9 | 3.30 (6331,75,999,39) | [+] <a href="#">Show genes</a> |

# 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)   | <a href="#">[+] Show genes</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| glucuronoxylan metabolic process | 1.01E-12 | 1.6E-9      | 3.43 (6331,72,999,39)   | <a href="#">[+] Show genes</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| xylan biosynthetic process       | 1.77E-12 | 1.86E-9     | 3.39 (6331,73,999,39)   | <a href="#">[-] Hide genes</a><br>GUT2 - putative glycosyltransferase<br>PGSIP3 - plant glycogenin-like starch initiation protein 3<br>FRA8 - exostosin-like protein<br>GAUT12 - alpha-1,4-galacturonosyltransferase<br>AT4G22460 - bifunctional inhibitor/lipid-transfer protein/seed storage 2s albumin-like protein<br>AT5G42180 - peroxidase 64<br>AT3G10910 - ring-h2 finger protein at172<br>LAC17 - laccase 17<br>KNAT7 - homeobox protein knotted-1-like 7<br>NAC012 - nac domain-containing protein 12<br>IRX9 - nucleotide-diphospho-sugar transferases-like protein<br>AT1G70500 - pectin lyase-like protein<br>CESA4 - cellulose synthase a catalytic subunit 4 [udp-forming]<br>AT1G08340 - rho gtpase activating protein with pak-box/p21-rho-binding domain<br>CTL2 - chitinase-like protein 2<br>IRX6 - cobra-like protein 4<br>MYB63 - myb domain protein 63<br>PGSIP1 - plant glycogenin-like starch initiation protein 1<br>AT5G46340 - putative o-acetyltransferase<br>AT3G21710 - hypothetical protein<br>AT2G03200 - aspartyl protease-like protein<br>AT1G09440 - protein kinase family protein<br>AT5G40020 - pathogenesis-related thaumatin-like protein<br>AT3G23090 - targeting protein for xklp2-like protein<br>AT5G67210 - hypothetical protein<br>AT3G56230 - btb/poz domain-containing protein<br>AT2G31930 - hypothetical protein<br>JP630 - putative polygalacturonase non-catalytic subunit jp630<br>MAP70-5 - microtubule-associated proteins 70-5<br>AT3G50220 - hypothetical protein<br>AGL44 - protein agamous-like 44<br>IRX12 - laccase-4<br>NAC073 - nac domain containing protein 73<br>IRX3 - cellulose synthase a catalytic subunit 7 [udp-forming]<br>AT4G27435 - hypothetical protein<br>MYB46 - transcription factor myb46<br>AT1G72220 - ring-h2 finger protein at154<br>FRD3 - mate efflux family protein<br>AT1G33800 - hypothetical protein |
| hemicellulase metabolic process  | 2.07E-12 | 2.24E-9     | 3.20 (6331,77,000,40)   | <a href="#">[+] Show genes</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

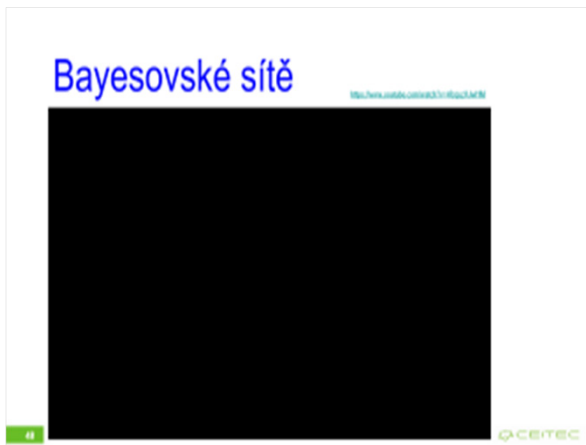
# Osnova

- **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

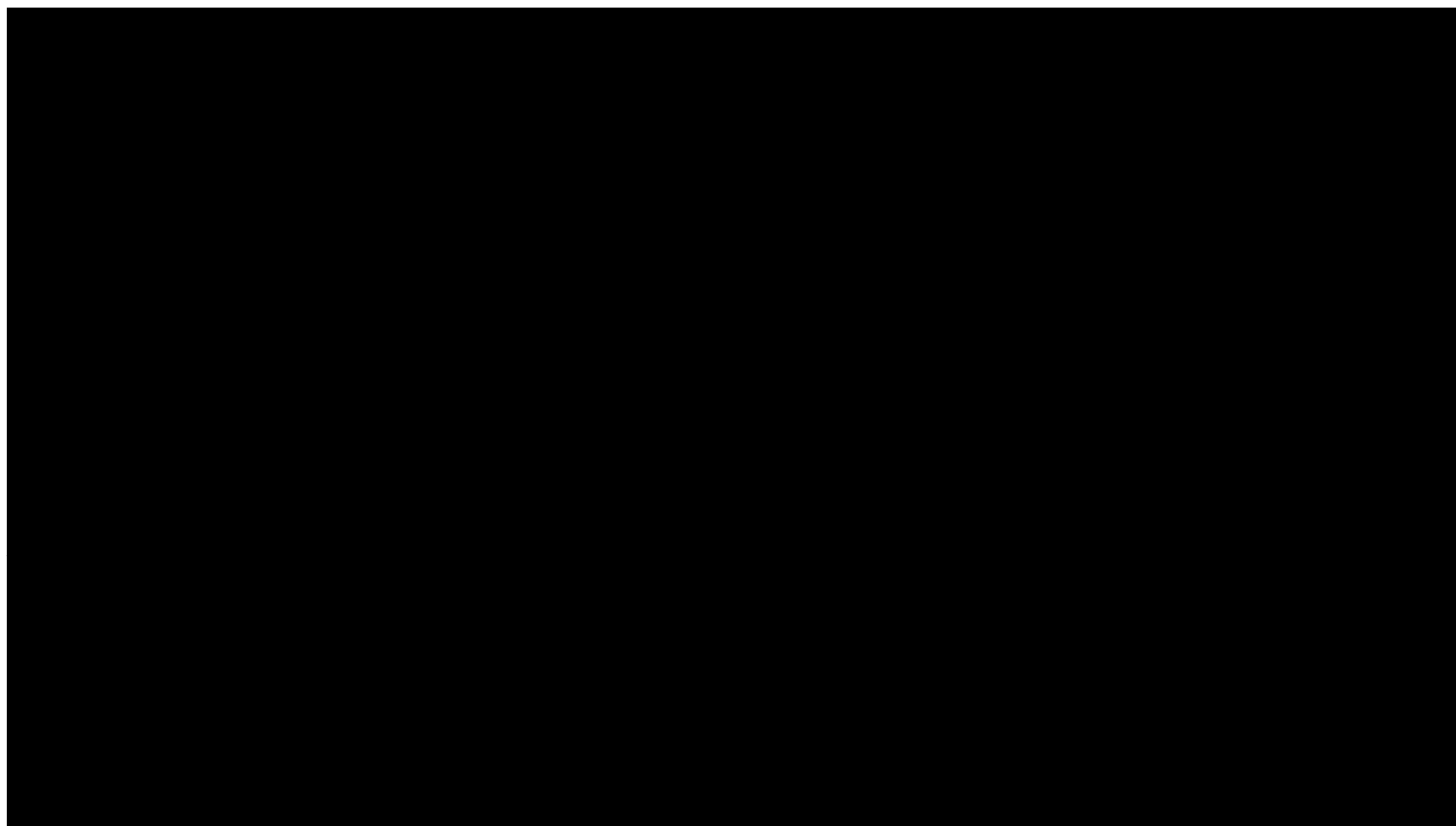
Co je Bayesovská síť?

- **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**

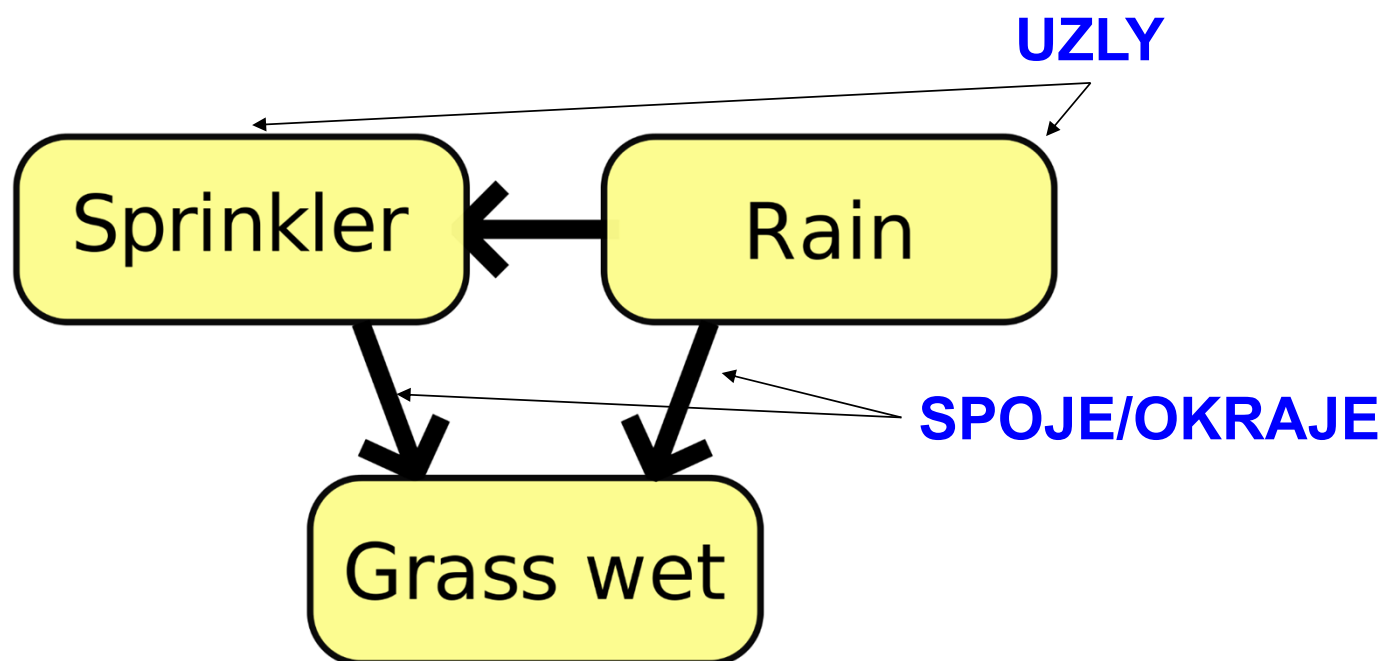
## ▪ UZLY

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

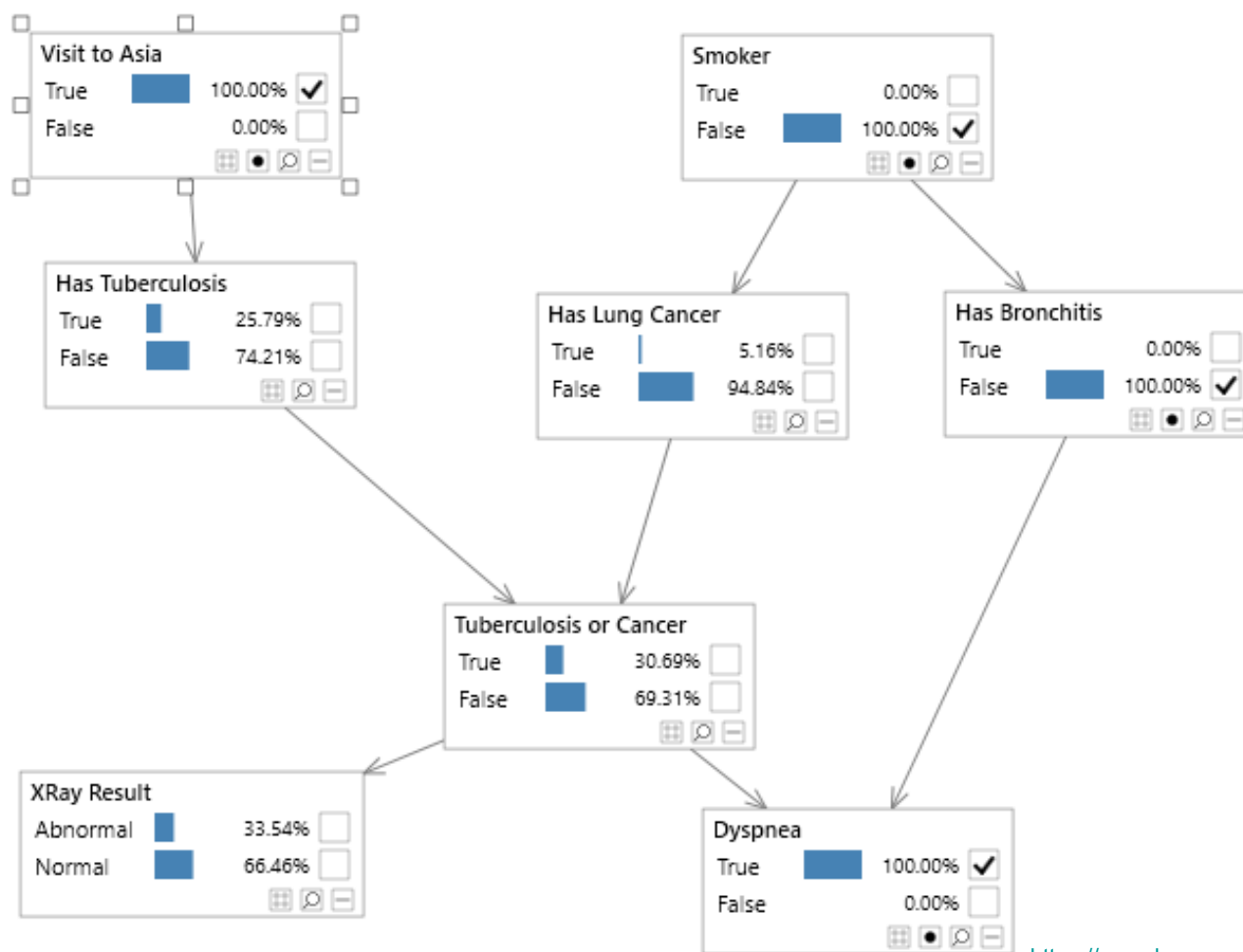
## ▪ SPOJE

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

# Bayesovské sítě



# Asijská Bayesovská síť

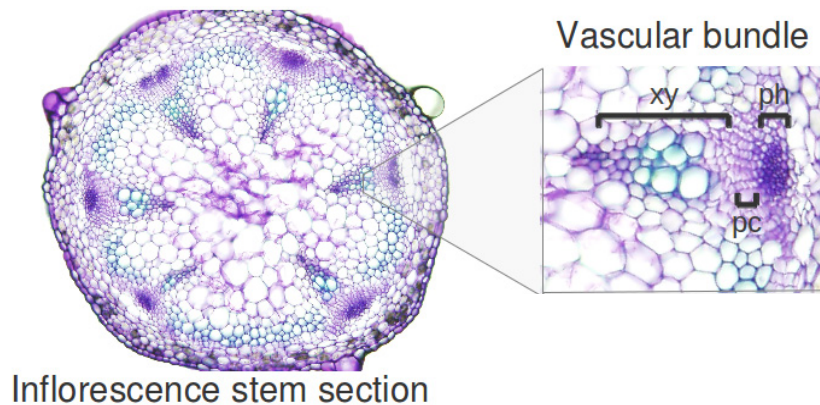


# Osnova

- **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 ARR. | [9]        |

Benitez and Hejatko, *PLoS One*, 2013

# Modelování molekulárních regulačních sítí

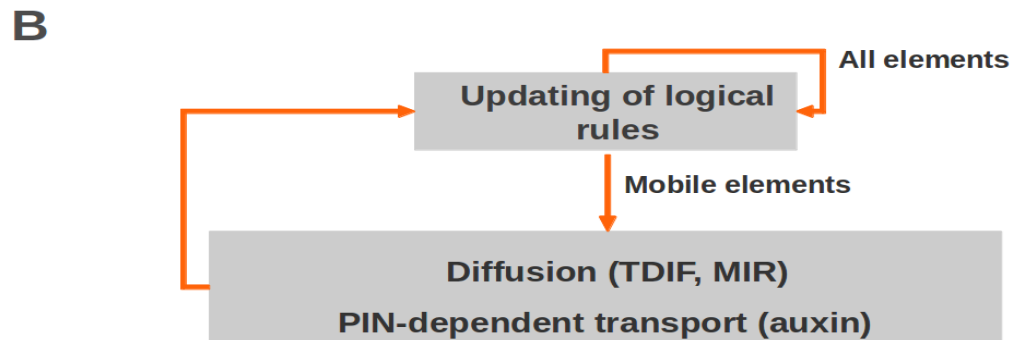
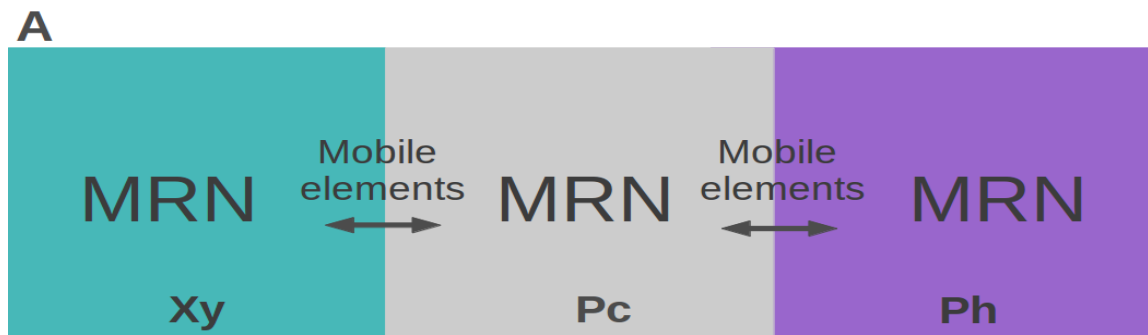
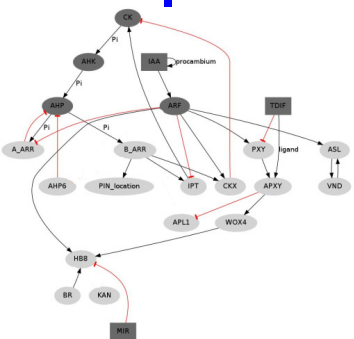
- Formulace *logických pravidel* definujících *dynamiku modelu*

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

Benitez and Hejatkó, *PLoS One*, 2013

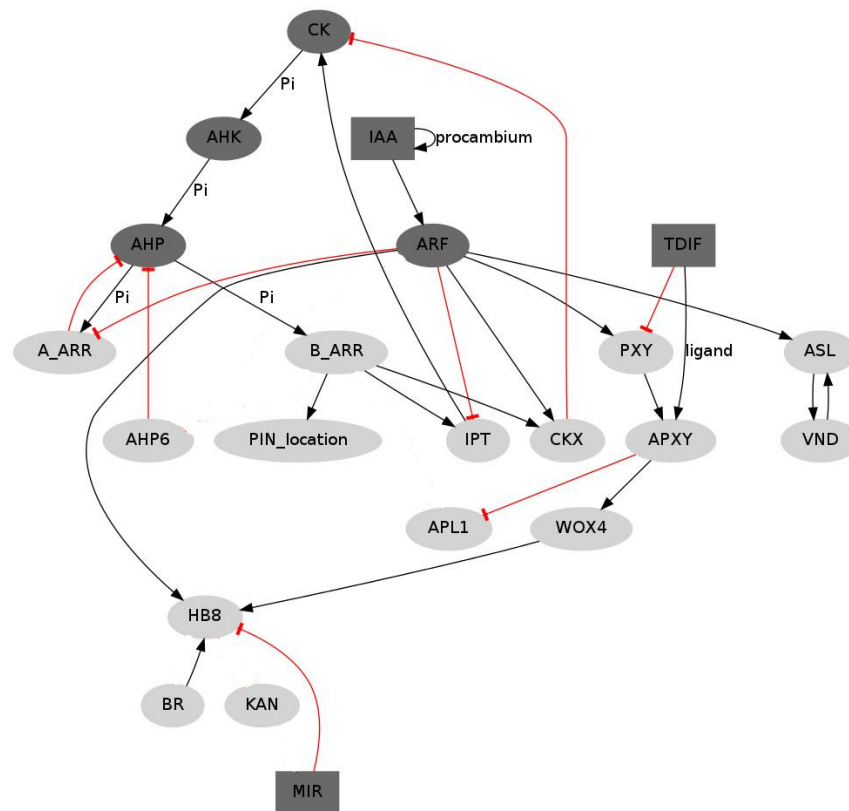
# Modelování molekulárních regulačních sítí

- Specifikace *mobilních prvků* a jejich chování v modelu



# Modelování molekulárních regulačních sítí

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



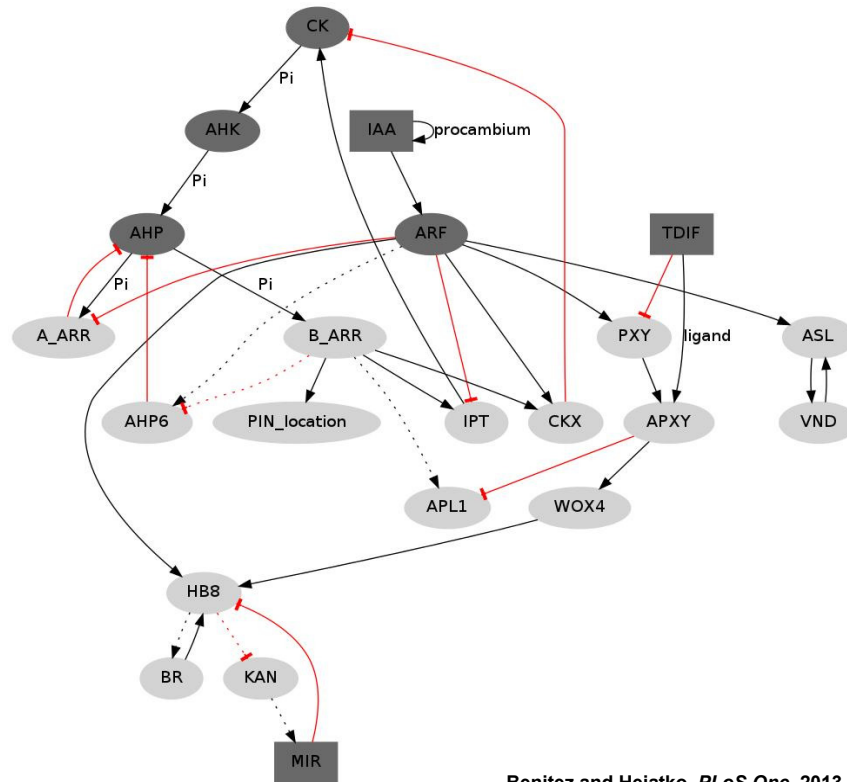
# Modelování molekulárních regulačních sítí

- 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

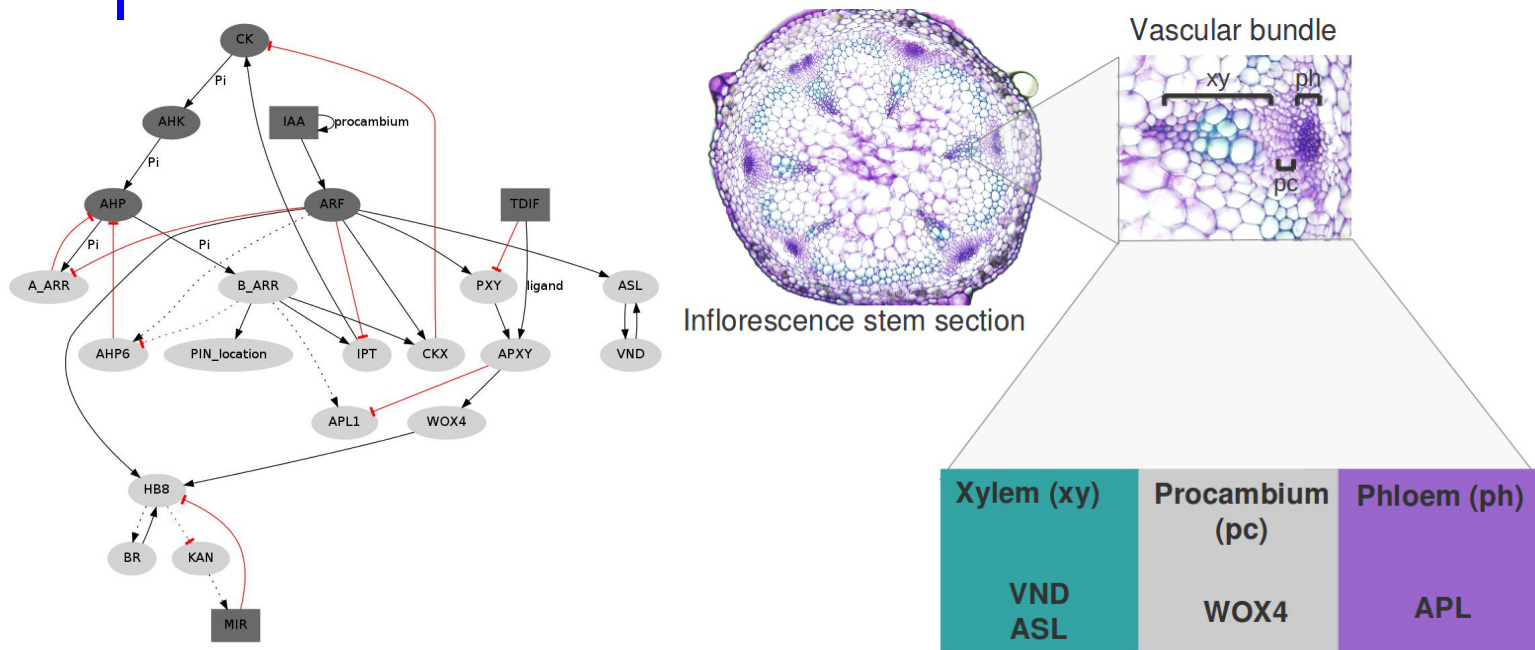
- Příprava *další verze* modelu a její *testování*



Benítez and Hejatko, *PLoS One*, 2013

# Modelování molekulárních regulačních sítí

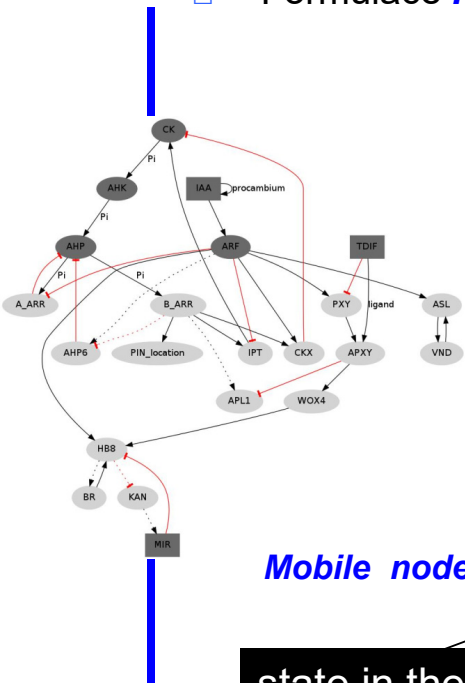
- *Dobrá model* by měl být schopen *simulovat realitu*



Benitez and Hejatko, *PLoS One*, 2013

# Modelování molekulárních regulačních sítí

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



logical rule function

state in the time  $t$

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

state in the time  $t+1$

Amount if TDIF or MIR165 in cell  $i$

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

state in the time  $t+1$

constant corresponding to a degradation term

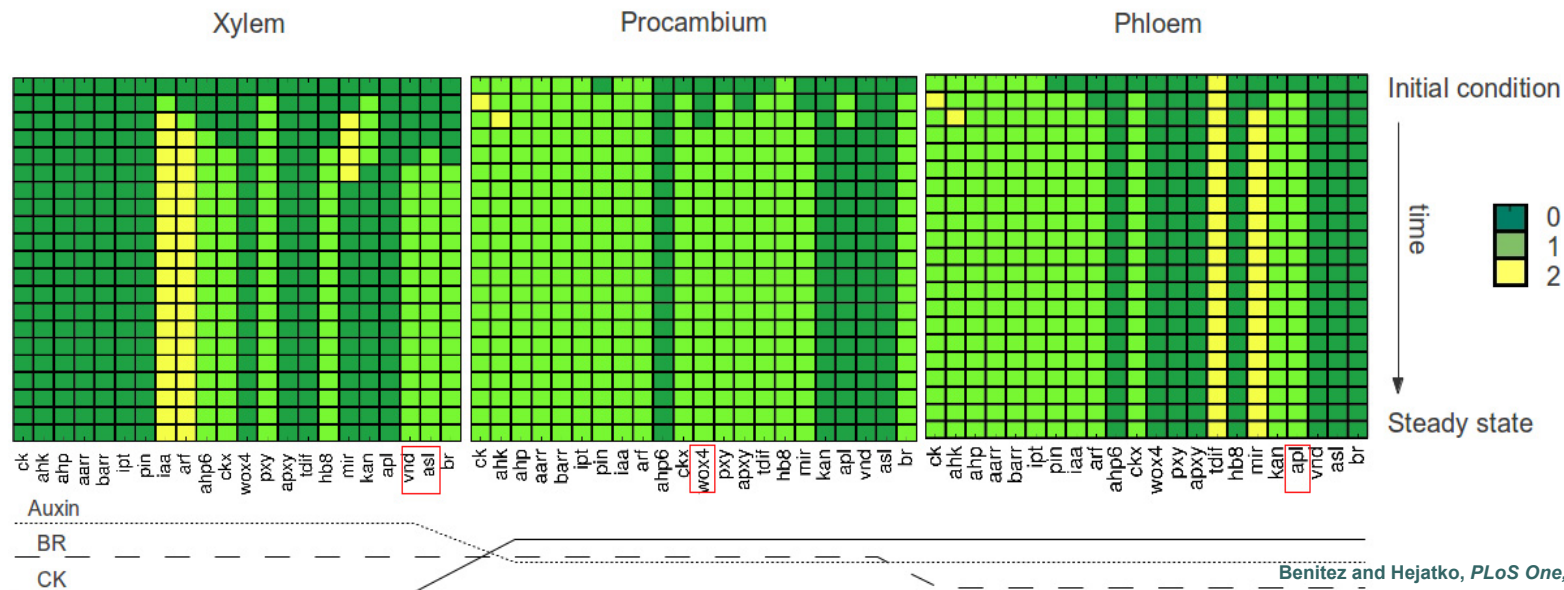
proportion of movable element

# Modelování molekulárních regulačních sítí

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

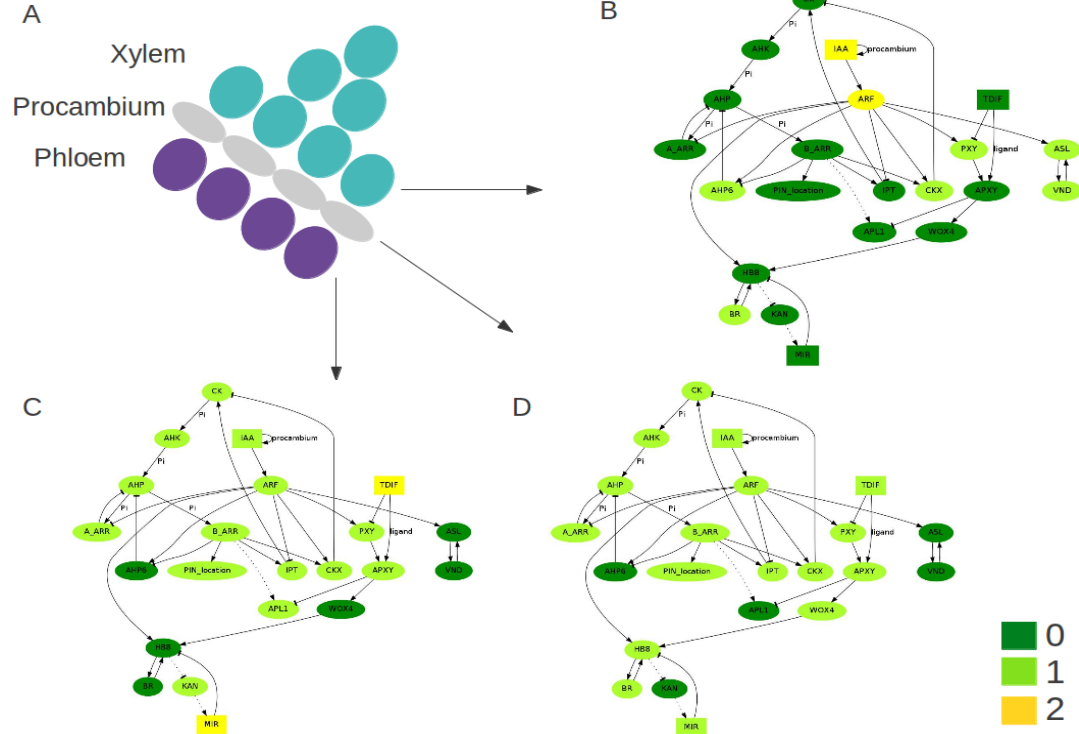
**Static nodes:**  $g_n(t+1) = F_n(g_{n1}(t), g_{n2}(t), \dots, g_{nk}(t))$

**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

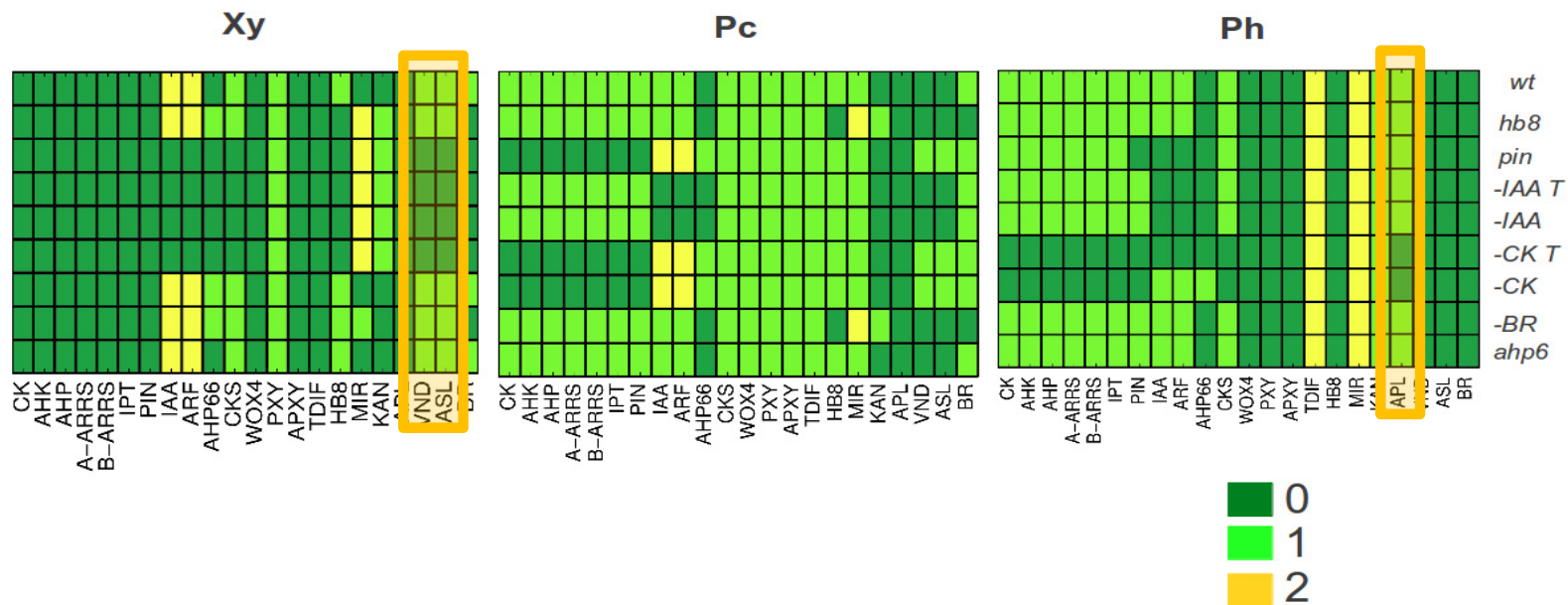
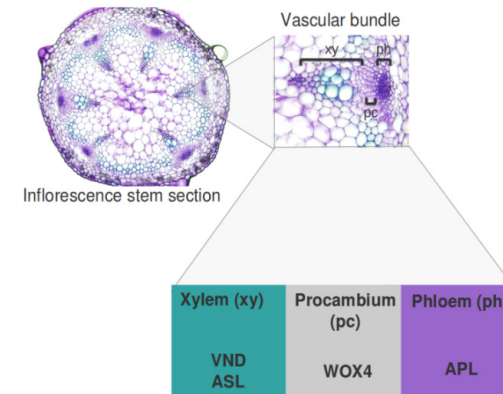
- **Dobrá model** by měl být schopen *simulovat realitu*



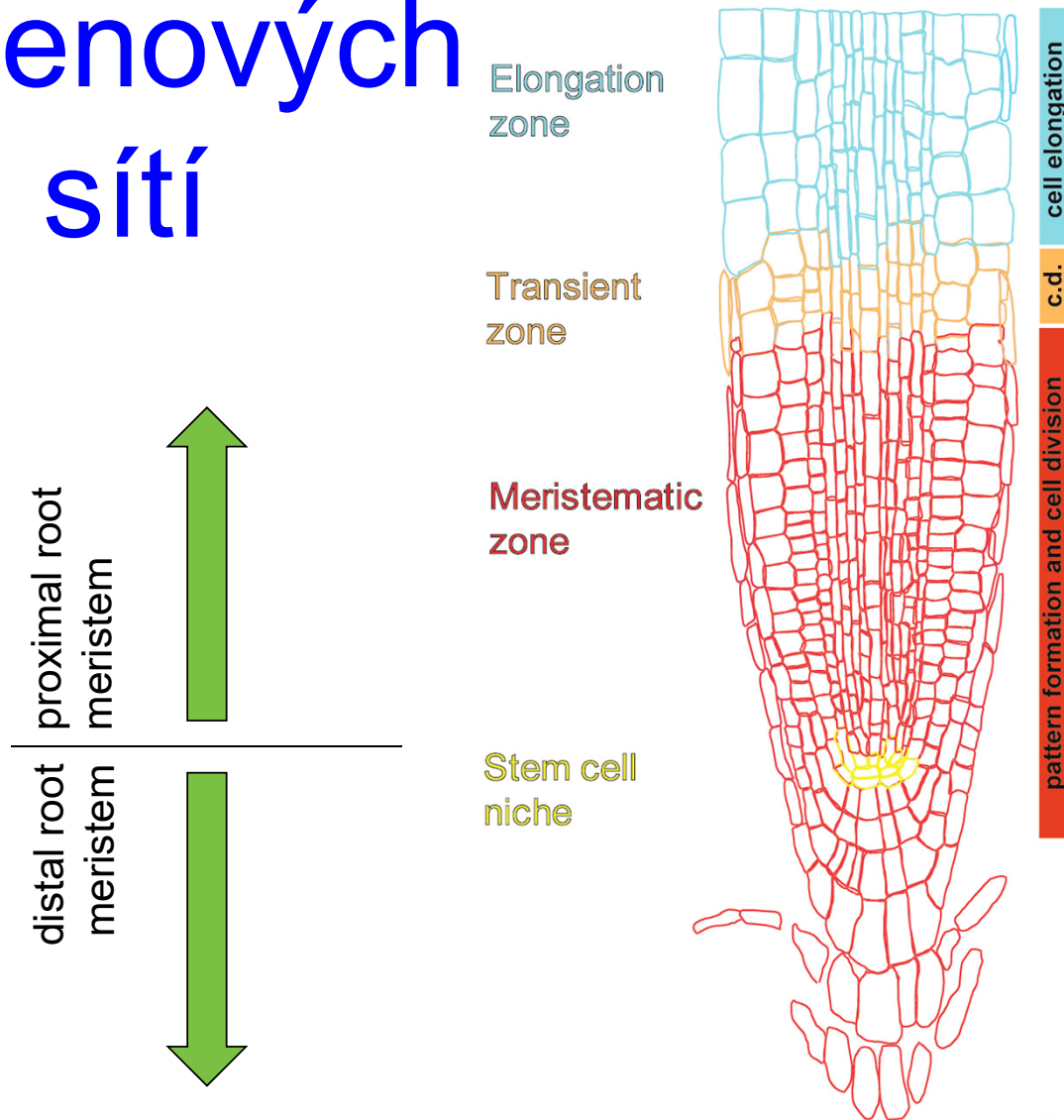
Benítez and Hejatkó, *submitted*

# Modelování molekulárních regulačních sítí

- Simulace *mutantů*

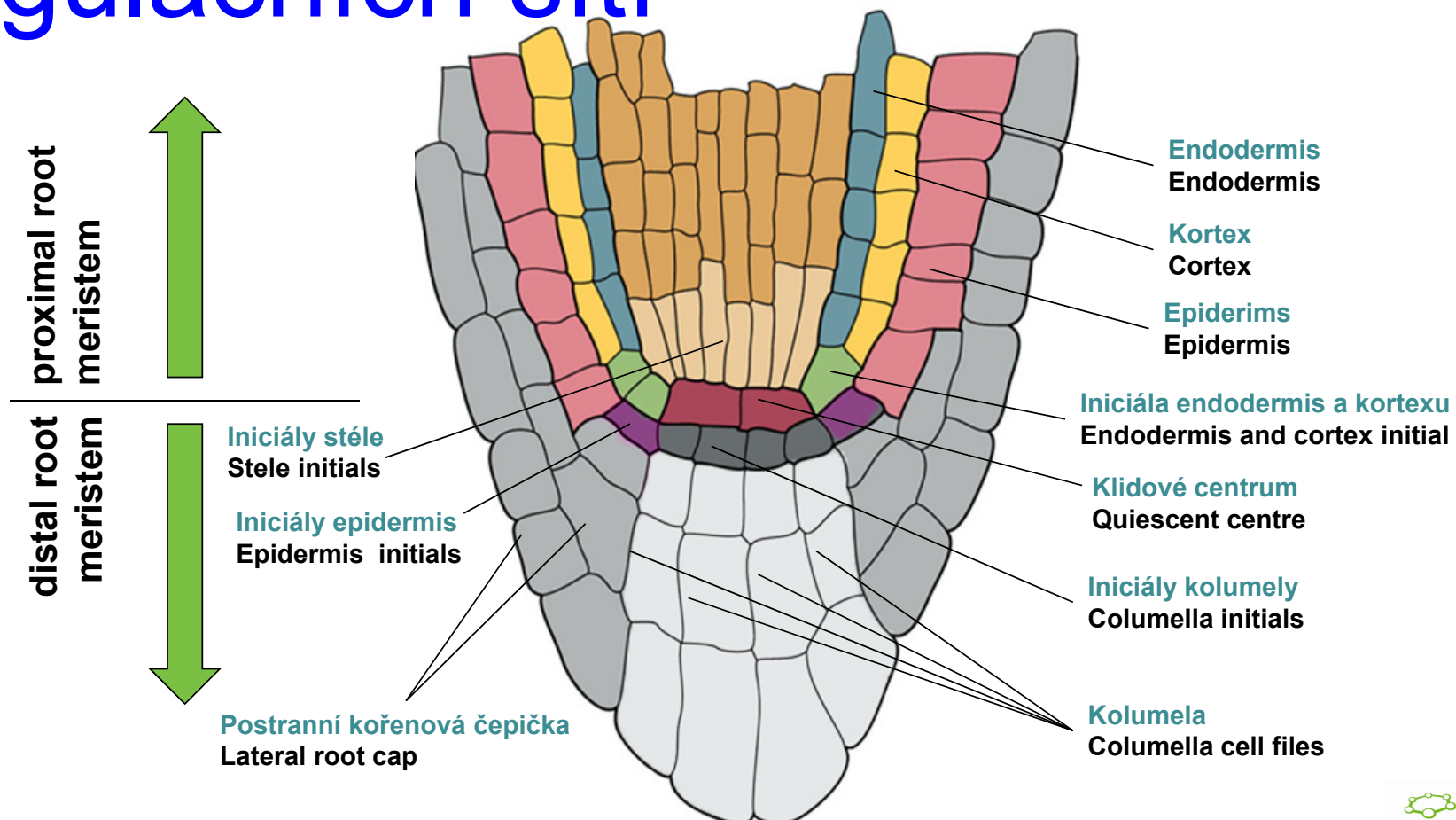


# Odvození genových regulačních sítí



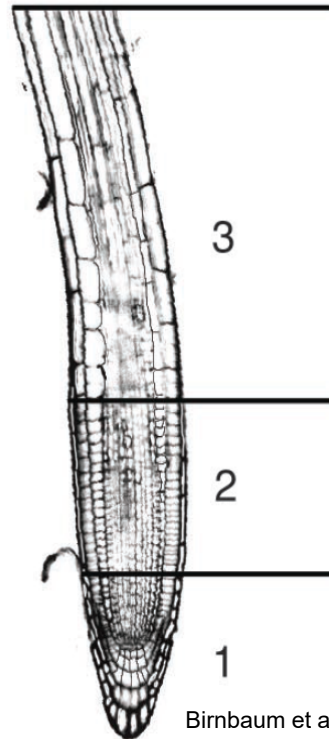
Benkova and Hejatko, *Plant Mol Biol* (2008)

# Odvození genových regulačních sítí

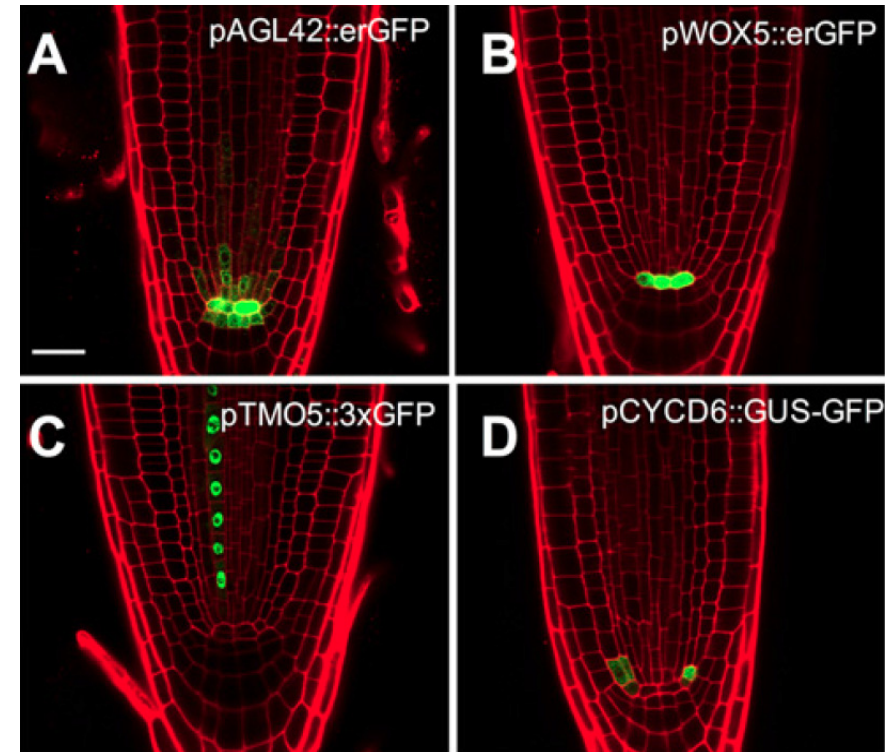


# Genové regulační sítě - GENIST

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



Birnbaum et al., *Science*, 2003



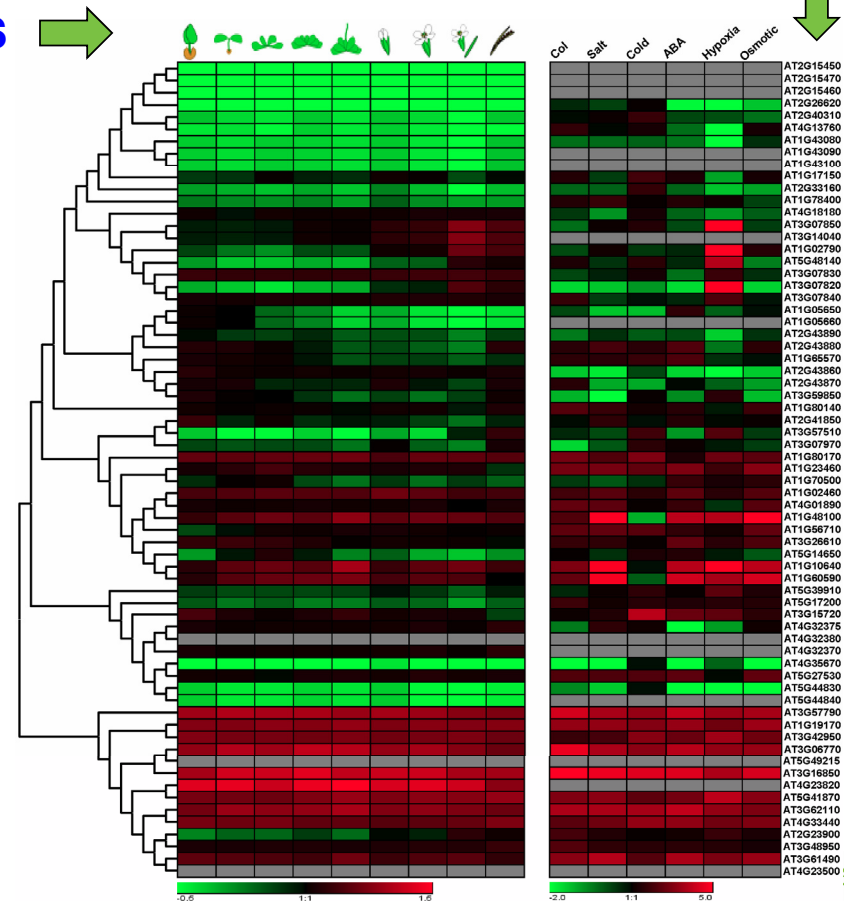
de Luis Balaguer et al., *PNAS*, 2017

# Kombinace velkých omických datových sad

PLETIVO/ČAS

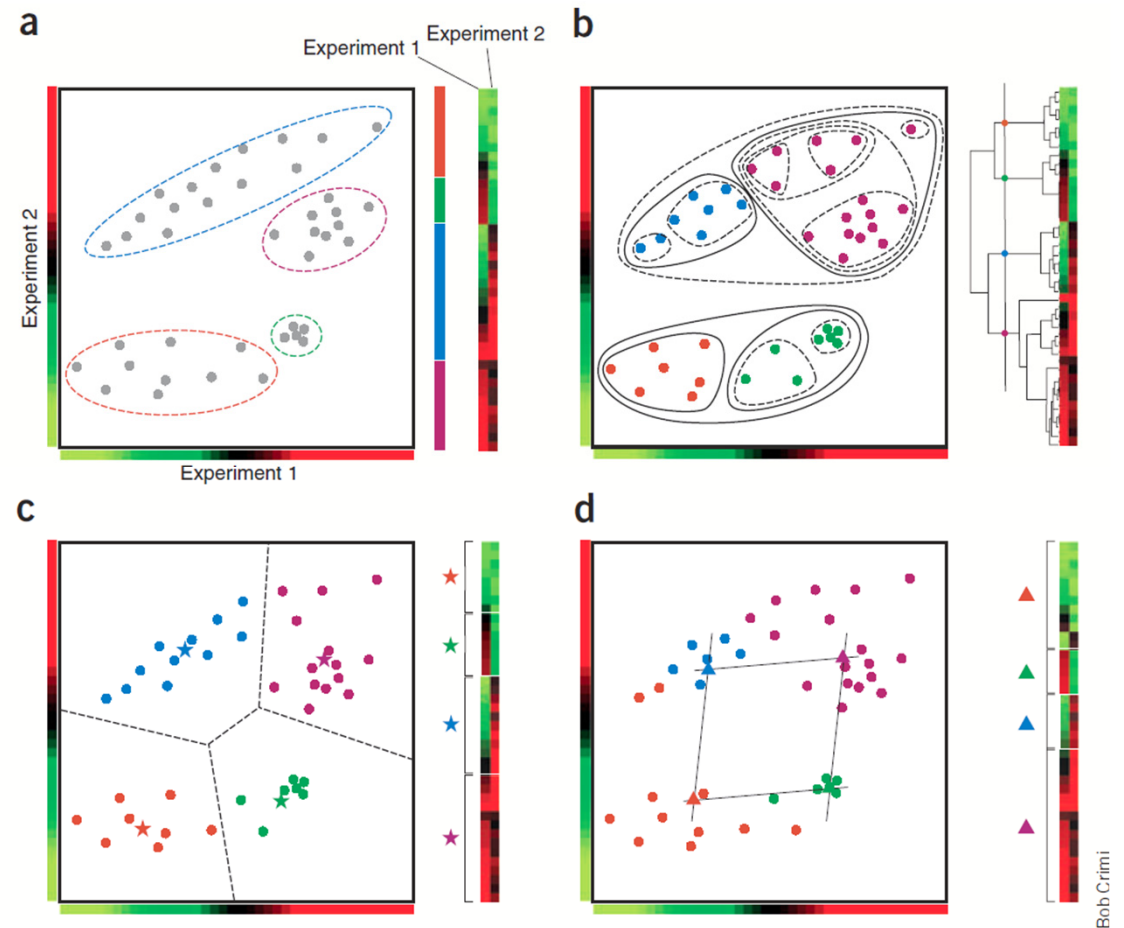


GENY



# Genové regulační sítě - GENIST

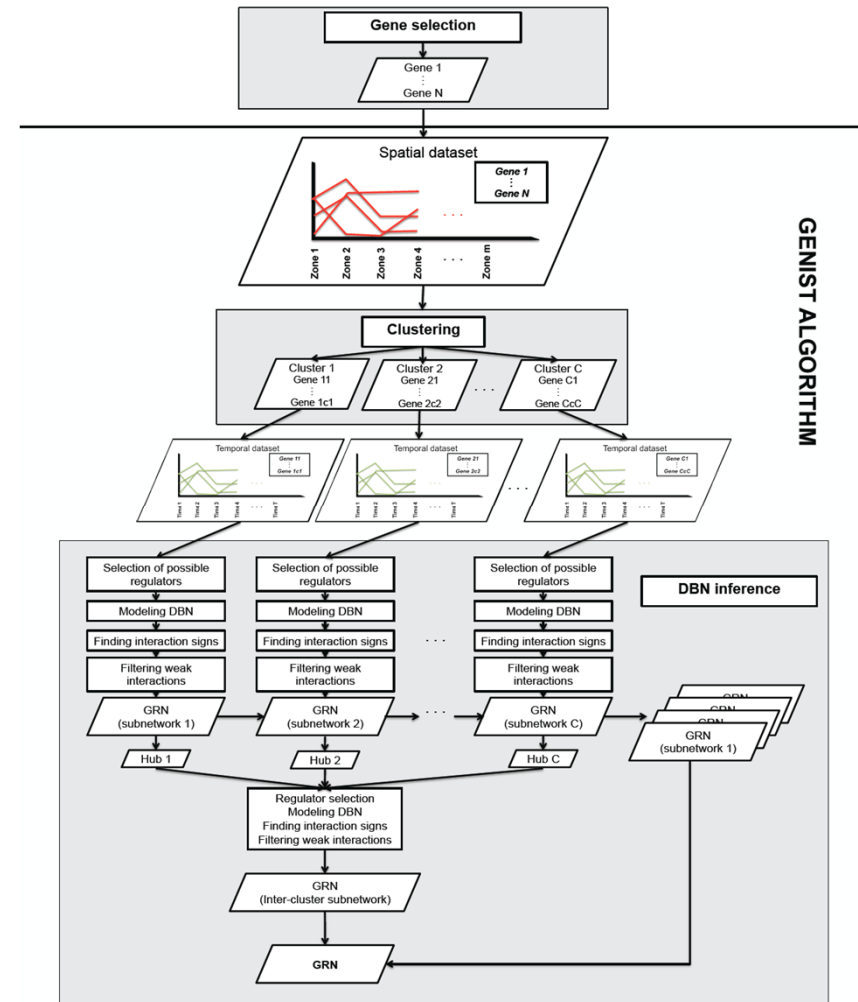
- Odvození GR sítí přes **GENIST**
  - **shlukování (klastrování) genů**
    - Expresní podobnost za různých podmínek/genetické pozadí, časové body, ...
  - **Odvození spojení uvnitř klastru**
    - **Selekce** potenciálních **regulátorů** a **ko-regulátorů**
      - Na základě časové korelace ve změně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



Haeseleer, *Computational Biology*, 2005

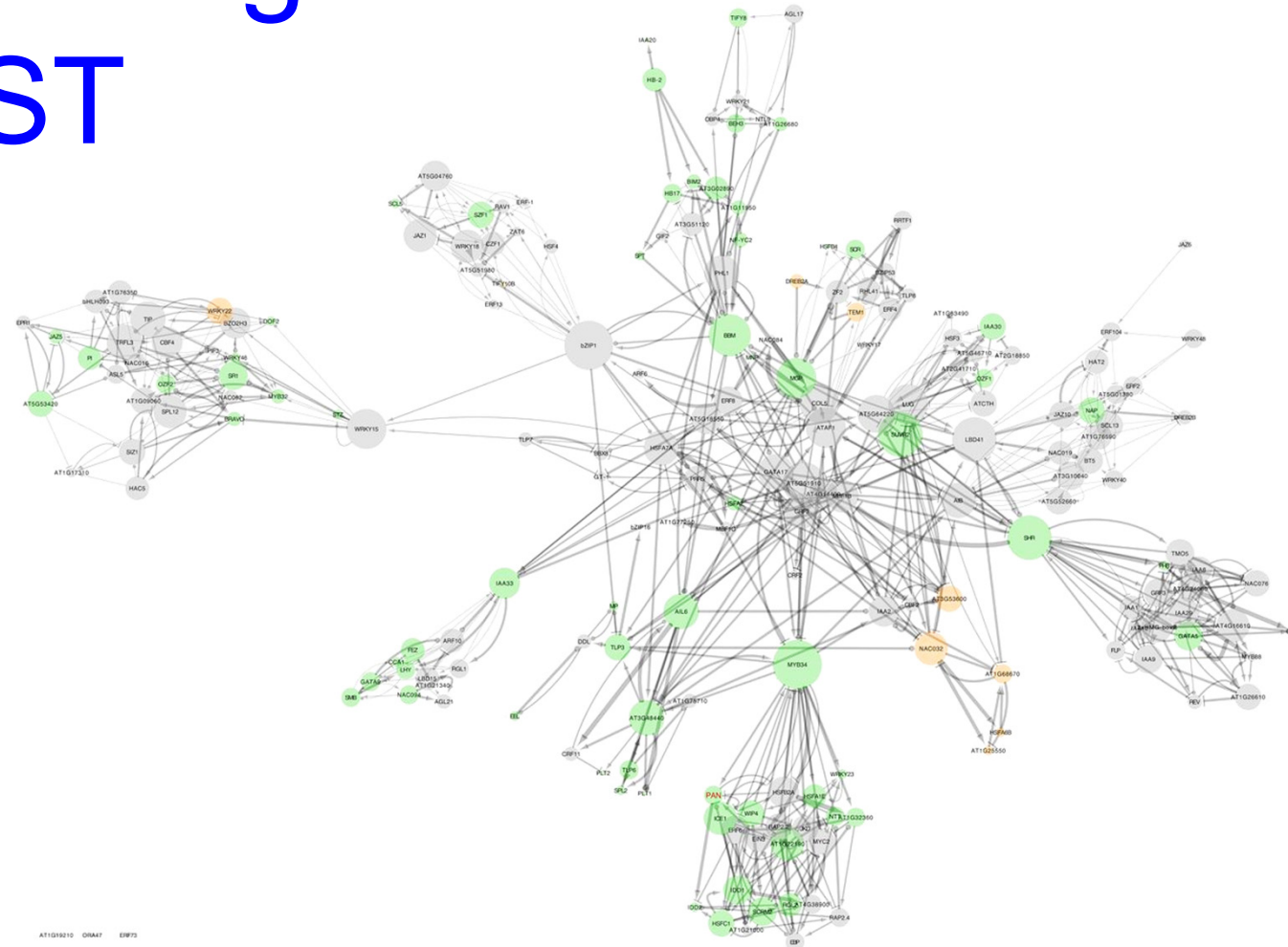
# Genové regulační sítě - GENIST

- 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**

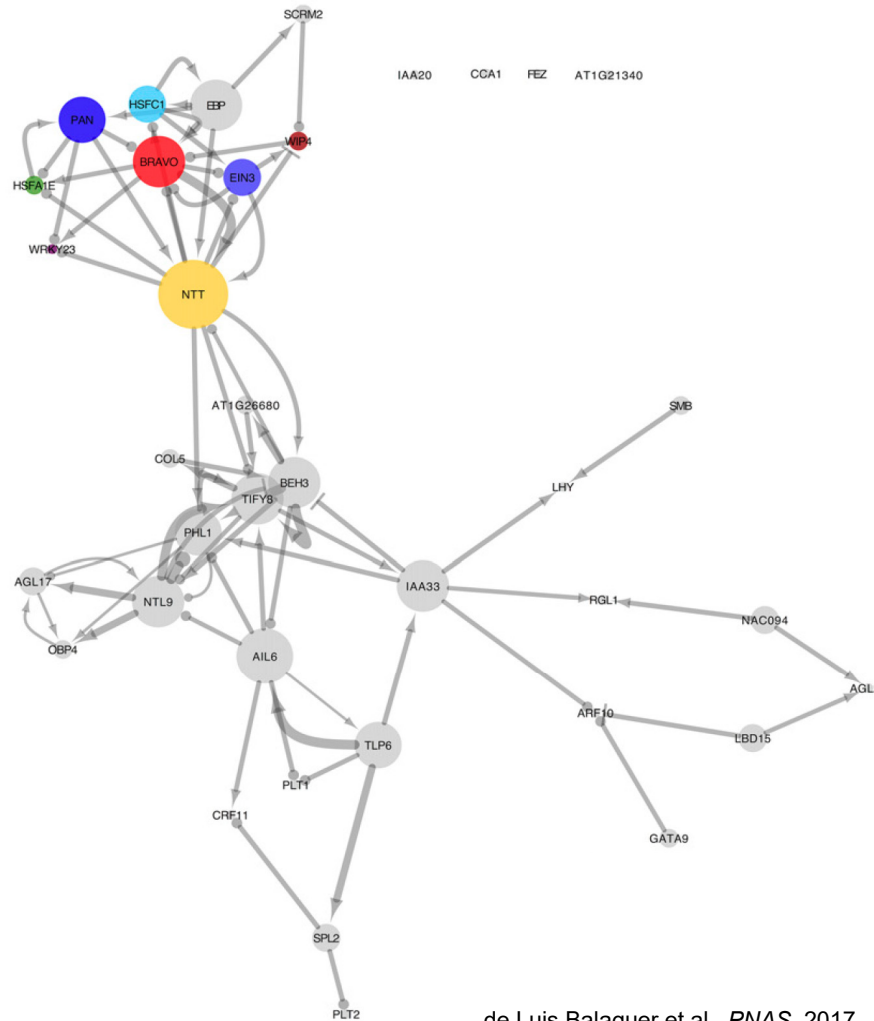


de Luis Balaguer et al., PNAS, 2017

# Genové regulační sítě - GENIST

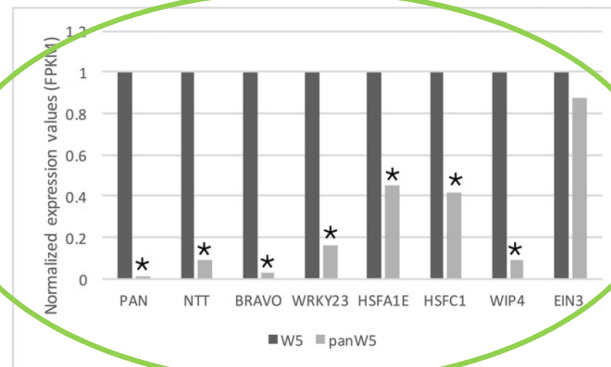
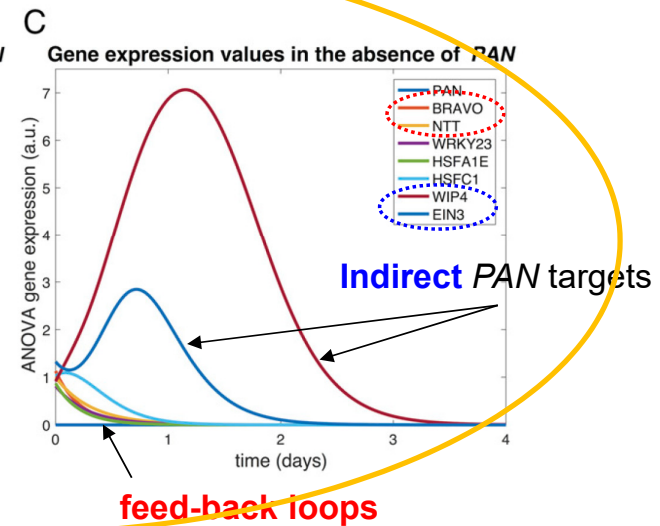
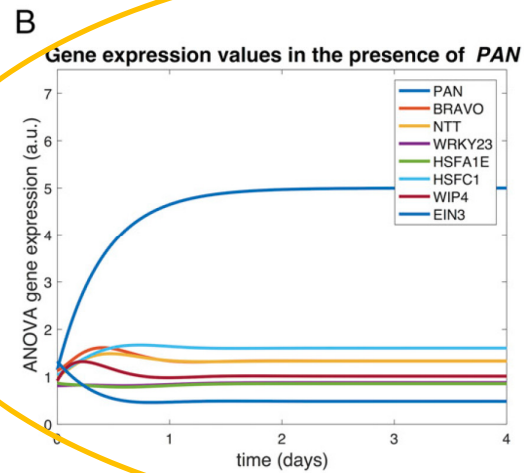
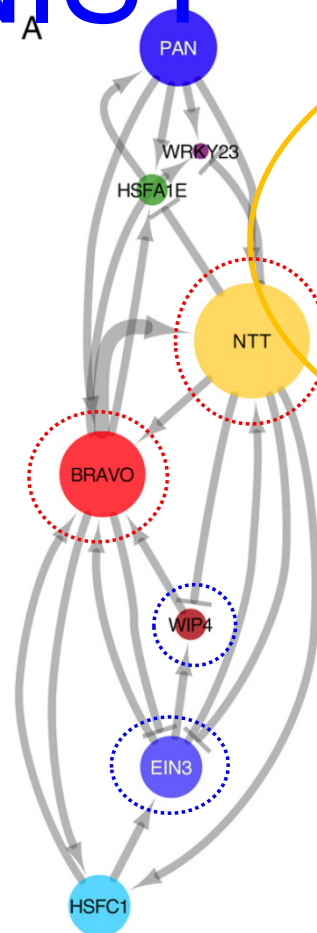


# Genové regulační sítě - GENIST



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

# Genové regulační sítě - GENIST



# 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“)