CG920 Genomics

Lesson 10
Systems Biology

Jan Hejátko

Functional Genomics and Proteomics of Plants,
Mendel Centre for Plant Genomics and Proteomics,
Central European Institute of Technology (CEITEC), Masaryk University, Brno
hejatko@sci.muni.cz, www.ceitec.muni.cz

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

EVROPSKÁ UNIE

1

Literature

- Literature sources for Chapter 12:

- Wilt, F.H., and Hake, S. (2004). Principles of Developmental Biology. (New York ; London: W. W. Norton)

- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.

- The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796-815.

- Benitez, M. and Hejatko, J. Dynamics of cell-fate determination and patterning in the vascular bundles of Arabidopsis thaliana (submitted)

- de Luis Balaguer MA, Fisher AP, Clark NM, Fernandez-Espinosa MG, Moller BK, Weijers D, Lohmann JU, Williams C, Lorenzo O, Sozzani R. 2017. Predicting gene regulatory networks by combining spatial and temporal gene expression data in Arabidopsis root stem cells. Proc Natl Acad Sci U S A 114(36): E7632-E7640.

Outline

- Definition of Systems Biology

- Tools
  - Gene Ontology Analysis
  - Bayesian Networks
  - Molecular/Gene Regulatory Networks Modeling
  - Inferring Gene Regulatory Networks from Large Omics Datasets

Definition

Systems biology is the computational and mathematical analysis and modeling of complex biological systems. It is a biology-based interdisciplinary field of study that focuses on complex interactions within biological systems, using a holistic approach (holism instead of the more traditional reductionism) to biological research (Wikipedia).

Definition

Systems biology is the study of biological systems whose behaviour cannot be reduced to the linear sum of their parts' functions. Systems biology does not necessarily involve large numbers of components or vast datasets, as in genomics or connectomics, but often requires quantitative modelling methods borrowed from physics (Nature).

Definition

Nice explanatory video by Dr. Nathan Price, associate director of the Institute for Systems Biology at https://www.youtube.com/watch?v=OrXRI_8UFHU.

Outline

- ~~Definition of Systems Biology~~

- Tools
  - Gene Ontology analysis

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

Plant Vascular Tissue Development

☐ **Vascular tissue** as a developmental model for *GO analysis* and *MRN modeling*

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

Hormonal Control Over Vascular Tissue Development

☐ *Transcriptional profiling* via *RNA sequencing*

WT | hormonal mutant

mRNA | mRNA

cDNA | cDNA

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

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

Sequencing by Synthesis
~1.5 to 11 days

CASAVA
2 days (30 min hands-on)

Sequencing by Illumina and
number of transcripts determination

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

Gene Ontology Analysis

☐ One of the possible approaches is to study *gene ontology*, i.e. previously demonstrated **association** of genes to **biological processes**

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

Gene Ontology Analysis

☐ Several tools allow **statistical evaluation** of **enrichment** for genes **associated with specific processes**

Gene Ontology Analysis

☐ Several tools allow **statistical evaluation** of **enrichment** for genes **associated with specific processes**

Gene Ontology Analysis

- Several tools allow **statistical evaluation** of **enrichment** for genes **associated with specific processes**

Gene Ontology Analysis

☐ Several tools allow **statistical evaluation** of **enrichment** for genes **associated with specific processes**

Gene Ontology Analysis

☐ Several tools allow **statistical evaluation** of **enrichment** for genes **associated with specific processes**

Outline

- Definition of Systems Biology

- Tools
  - Gene Ontology analysis
  - Bayesian Networks

Bayesian Networks

- What are Bayesian networks?
  - **Probabilistic Graphical Model that can be used to build models from data and/or expert opinion**

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

Bayesian Networks

- ## What are **Bayesian Networks**?
  - **Probabilistic Graphical Model** that can be used to **build models from data** and/or **expert opinion**
  - can be used for a wide range of tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction** and **decision making under uncertainty**

- ## NODES
  - each node represents a **variable** such as someone's height, age or gender. A variable might be discrete, such as Gender = {Female, Male} or might be continuous such as someone's age

- ## LINKS
  - added **between nodes** to indicate that **one node** directly **influences** the **other**

Asia Bayesian Network

Visit to Asia
True	100.00%	✓
False	0.00%	

Smoker
True	0.00%	
False	100.00%	✓

Has Tuberculosis
True	25.79%	
False	74.21%	

Has Lung Cancer
True	5.16%	
False	94.84%	

Has Bronchitis
True	0.00%	
False	100.00%	✓

Tuberculosis or Cancer
True	30.69%	
False	69.31%	

XRay Result
Abnormal	33.54%	
Normal	66.46%	

Dyspnea
True	100.00%	✓
False	0.00%	

https://www.bayesserver.com/
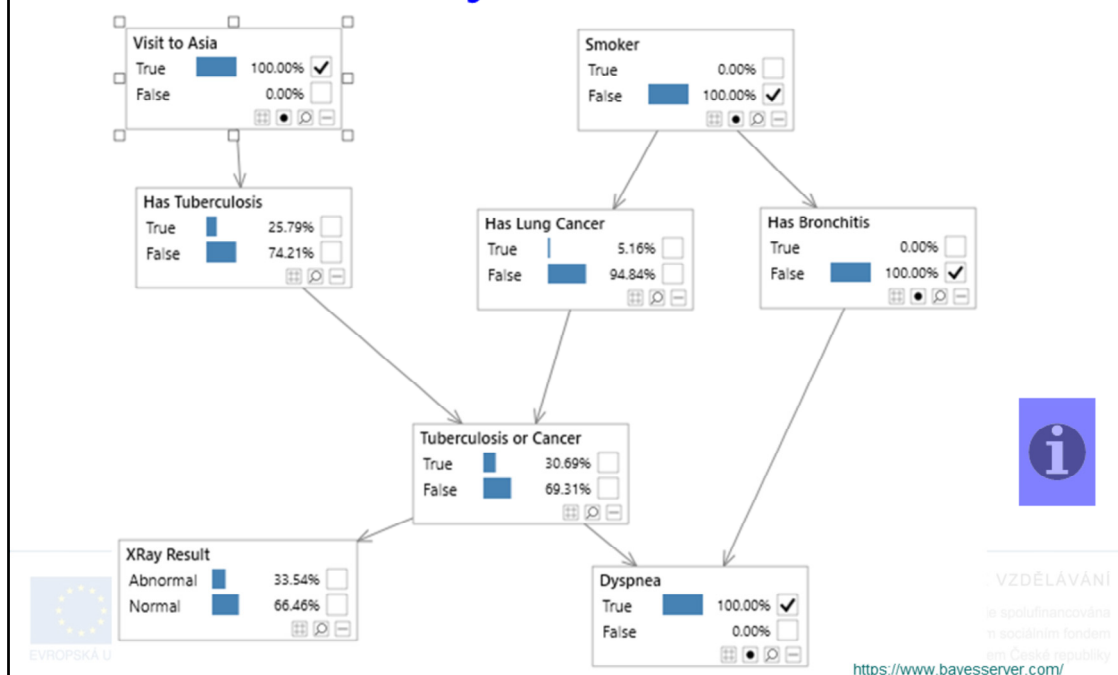
Outline

- Definition of Systems Biology

- Tools
  - Gene Ontology analysis
  - Bayesian Networks
  - **Molecular/Gene Regulatory Networks Modeling**

Molecular Regulatory Networks Modeling

□ *Vascular tissue* as a developmental model for *MRN modeling*
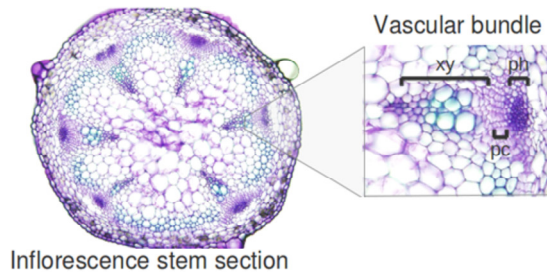
Vascular bundle

Inflorescence stem section

Benitez and Hejatko, *PLoS One,* 2013

Molecular Regulatory Networks Modeling

☐ **Literature search** for published data and **creating small database**

Interaction	Evidence	References
A-ARRs –| CK signaling	Double and higher order type-A ARR mutants show increased sensitivity to CK.	[27]
	Spatial patterns of A-type ARR gene expression and CK response are consistent with partially redundant function of these genes in CK signaling.	[27]
	A-type ARRs decreases B-type ARR6-LUC.	[13]
	Note: In certain contexts, however, some A-ARRs appear to have effects antagonistic to other A-ARRs.	[27]
AHP6 –| AHP	ahp6 partially recovers the mutant phenotype of the CK receptor WOL.	[9]
	Using an in vitro phosphotransfer system, it was shown that, unlike the AHPs, native AHP6 was unable to accept a phosphoryl group. Nevertheless, AHP6 is able to inhibit phosphotransfer from other AHPs to ARRs.	[9]

Benitez and Hejatko, *PLoS One,* 2013

Molecular Regulatory Networks Modeling

☐ Formulating *logical rules* defining the *model dynamics*

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

Benitez and Hejatko, *PLoS One*, 2013

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

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

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

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

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

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

Molecular Regulatory Networks Modeling

☐ Specifying of missing interactions via *informed predictions*

Interaction	Evidence	References
CK → PIN7 radial localization	Predicted interaction (could be direct or indirect)	
	Informed by the following data:	
	During the specification of root vascular cells in Arabidopsis thaliana, CK regulates the radial localization of PIN7.	[18]
	Expression of PIN7:GFP and PIN7::GUS is upregulated by CK with no significant influence of ethylene.	[18,20]
	In the root, CK signaling is required for the CK regulation of PIN1, PIN3, and PIN7. Their expression is altered in wol, cre1, ahk3 and ahp6 mutants.	[19]
CK→APL	Predicted interaction (could be direct or indirect)	
	Consistent with the fact that APL overexpression prevents or delays xylem cell differentiation, as does CKs.	[21]
	Partially supported by microarray data and phloem-specific expression patterns of CK response factors.	(TAIR, ExpressionSet:1 005823559, [22])
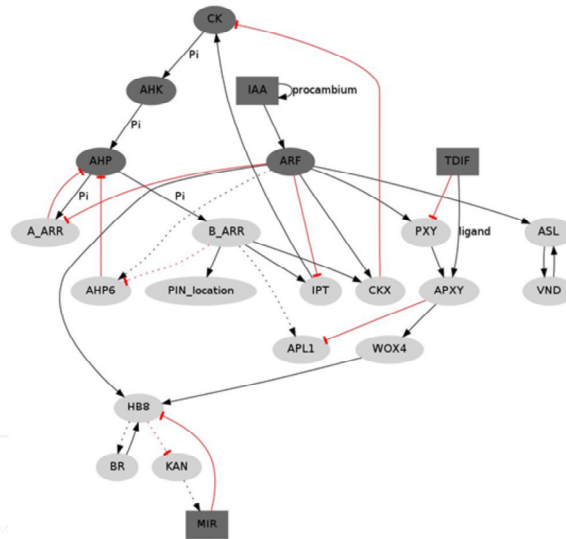
In comparison to the model shown on slide 21, the final version of the model contains the predicted interactions (dashed lines).

Molecular Regulatory Networks Modeling
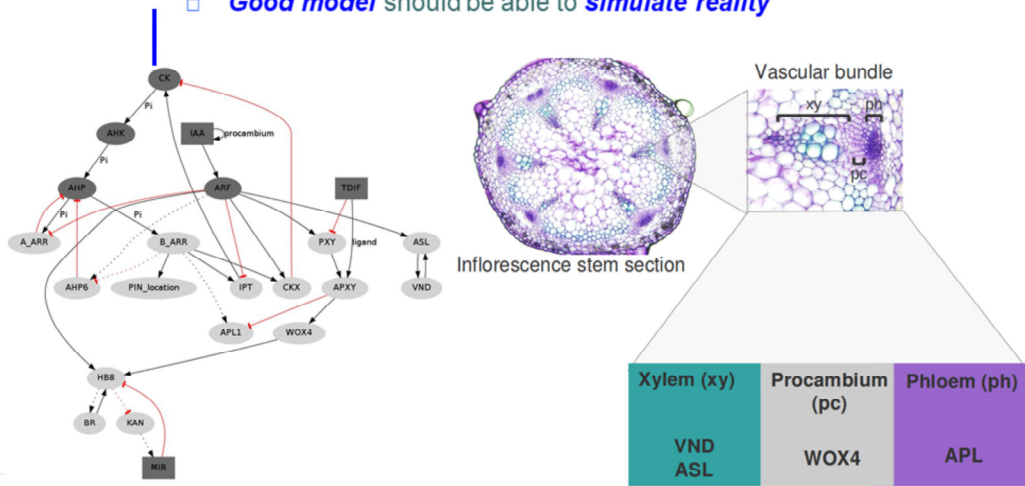
□ *Good model* should be able to *simulate reality*

Vascular bundle

Inflorescence stem section

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

Benitez and Hejatko, *PLoS One*, 2013

Molecular Regulatory Networks Modeling

☐ Formulating *equations* describing the *relationships* in the model
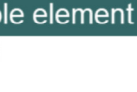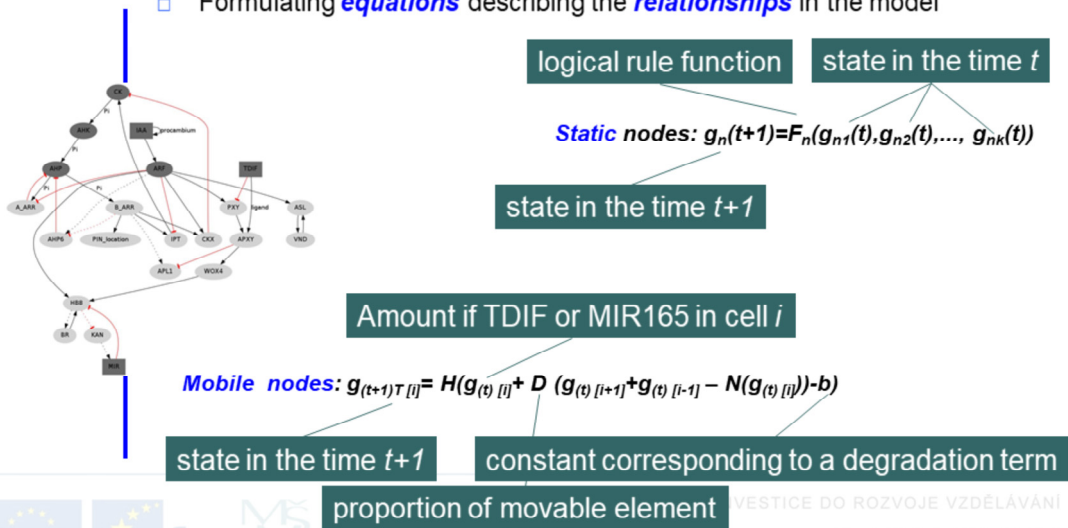
| logical rule function | state in the time *t* |

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

| 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 |

Molecular Regulatory Networks Modeling

□ **Good model** should be able to **simulate reality**

Static nodes: $g_n(t+1)=F_n(g_{n1}(t),g_{n2}(t),...,g_{nK}(t))$

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

Xylem · Procambium · Phloem

Initial condition · time · Steady state

Auxin · BR · CK

Benítez and Hejátko, *submitted*
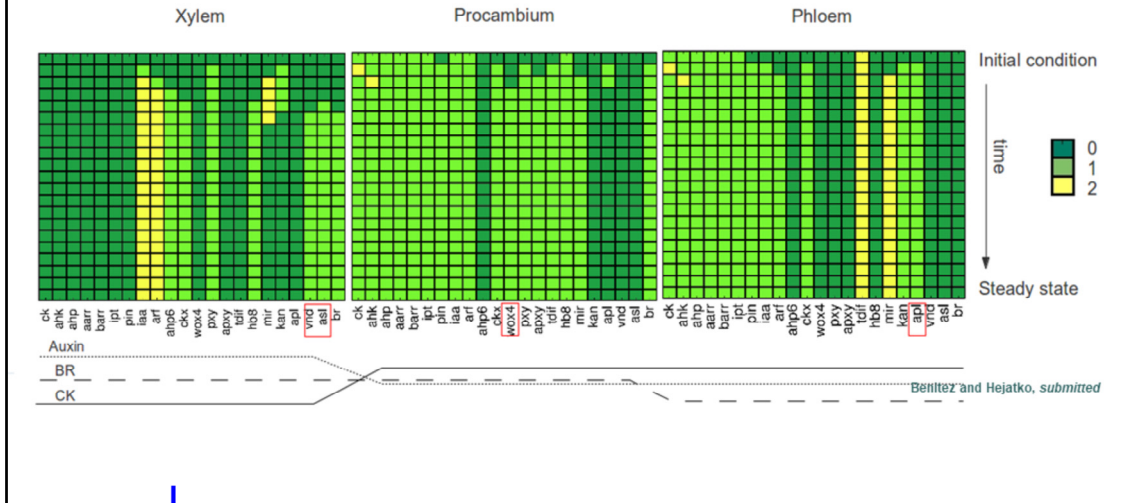
The initial conditions specify the initial state of some of the network elements (figure above) and are the following :

I) In the procambial position (central compartment), CK is initially available and there is an initial and sustained IAA input or self-upregulation. This condition is supported by several lines of evidence. Also *HB8*, a marker of early vascular development that has been found in preprocambial cells, is assumed to be initially present at this position. These conditions are not fixed, however. After the initial configuration, all the members of the CK and IAA signaling pathways, as well as *HB8*, can change their states according to the logical rules.

II) In the xylem and phloem positions, it is assumed that no element is initially active except for the CK signaling pathway and TDIF, both in the phloem position.The level of expression for a given node is represented by a discrete variable *g* and its value at a time *t+1* depends on the state of other components of the network (*g1, g2, ..., gN*) at a previous time unit. The state of every gene *g* therefore changes according to:

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

In this equation, *gn1, gn2,..., gnk* are the regulators of gene *gn* and *Fn* is a discrete function known as a logical rule (logical rules are grounded in available experimental data, for example see slide 20). Given the logical rules, it is possible to follow the dynamics of the network for any given initial configuration of the nodes expression state. One of the most important traits of dynamic models is the existence of steady states in which the entire network enters into a selfsustained configuration of the nodes state. It is thought that in developmental systems such self-sustained states correspond to particular cell types.

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

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

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

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

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

Using the logical rules, equations 1–3, and a broad range of parameter values (not shown here), it is possible fully to reproduce the results and analyses reported in the following sections (see the figure above for the simulation time course).
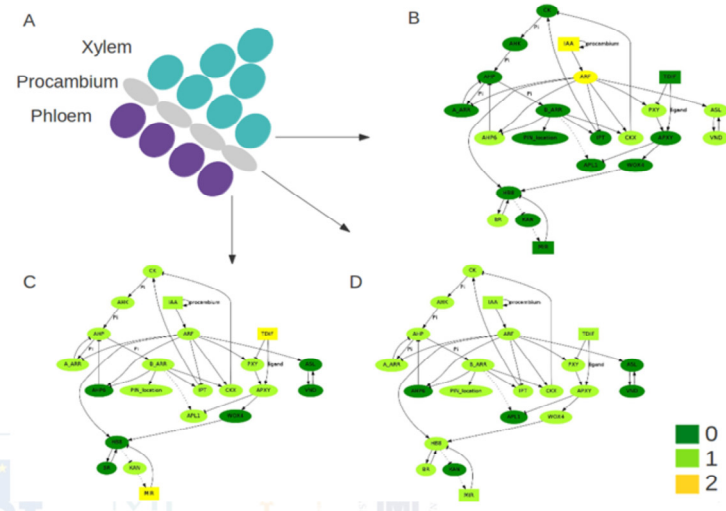
Another representation of the distinct expression profiles in the individual vascular bundle compartments (phloem, procambium and xylem).

Molecular Regulatory Networks Modeling

- Simulation of *mutants*
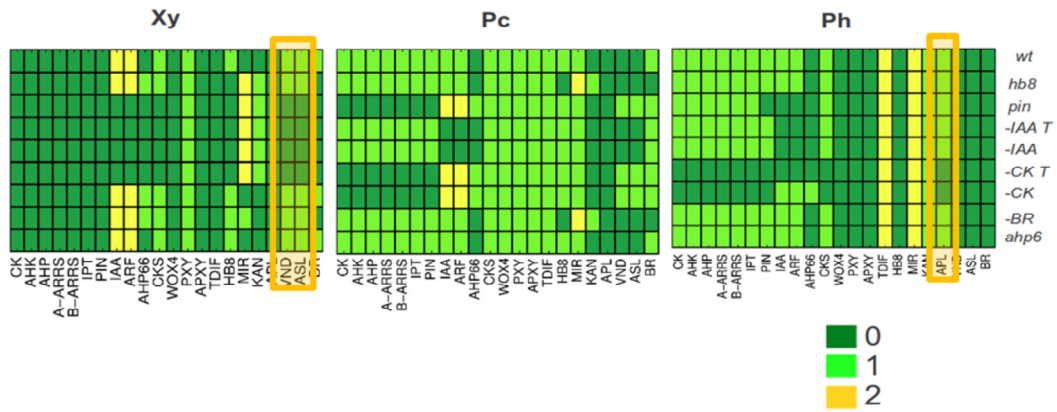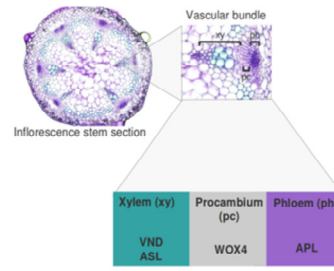
Outline

- Definition of Systems Biology

- Tools
    - Gene Ontology analysis
    - Bayesian Networks
    - Molecular/Gene Regulatory Networks Modeling
    - Inferring Gene Regulatory Networks from Large Omics Datasets

Systems Biology in Cancer Research

miRNA/mRNA Profiling

Guo et al., *Mol Med Reports*, 2017
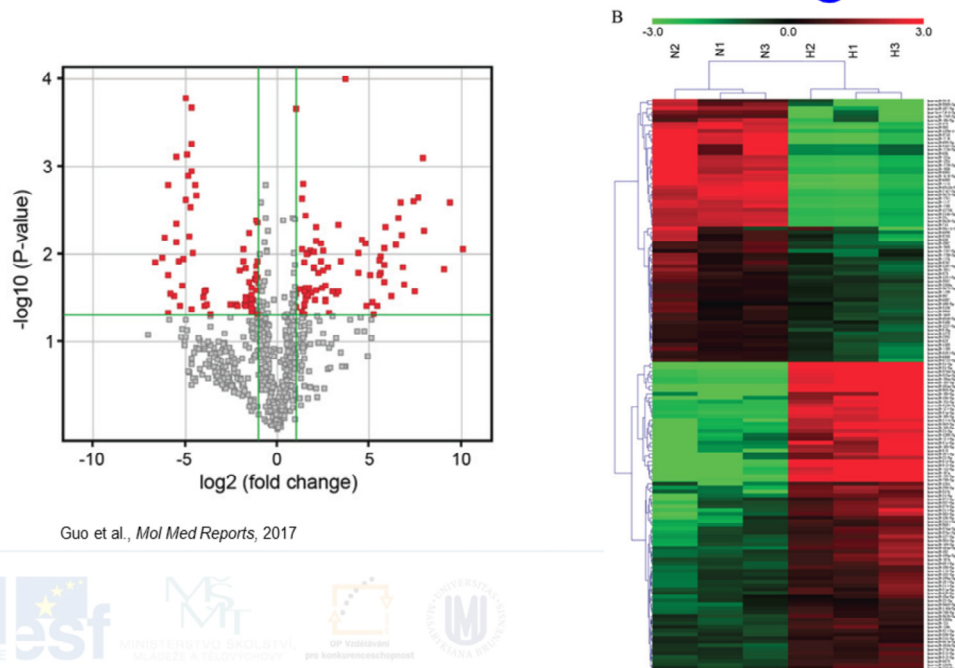
Hypertrophic scars (HS) area fibroproliferative disorder of the skin, which causes aesthetic and functional impairment. However, the molecular pathogenesis of this disease remains largely unknown and currently no efficient treatment exists. MicroRNAs (miRNAs) are involved in a variety of pathophy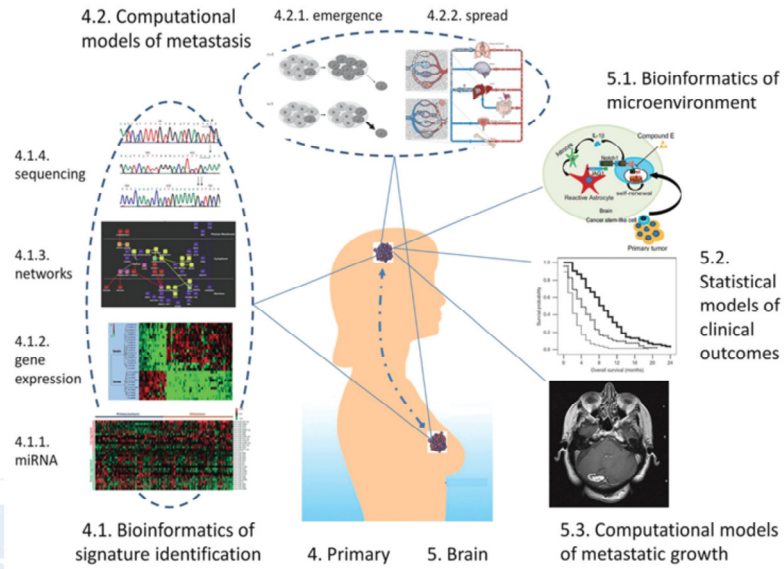siological processes, however the role of miRNAs in HS development remains unclear. To investigate the miRNA expression signature of HS, microarray analysis was performed and 152 miRNAs were observed to be differentially expressed in HS tissue compared with normal skin tissues. Of the miRNAs identified, miRNA-21 (miR-21) was significantly increased in HS tissues and hypertrophic scar fibroblasts (HSFBs) as determined by reverse transcription-quantitative polymerase chain reaction analysis. It was also observed that, when miR-21 in HSFBs was blocked through use of an antagomir, the phenotype of fibrotic fibroblasts *in vitro* was reversed, as demonstrated by growth inhibition, induction of apoptosis and suppressed expression of fibrosis-associated genes collagen type I α 1 chain (COL1A1), COL1A2 and fibronectin. Furthermore, miR-21 antagomir administration significantly reduced the severity of HS formation and decreased collagen deposition in a rabbit ear HS model. The total scar area and scar elevation index were calculated and were demonstrated to be significantly decreased in the treatment group compared with control rabbits. These results indicated that the miR-21 antagomir has a therapeutic effect on HS and suggests that targeting miRNAs may be a successful and novel therapeutic strategy in the treatment of fibrotic diseases that are difficult to treat with existing methods.

miRNA expression signature profiling in hypertrophic scars (HS). (A) Volcano plot presenting differentially expressed miRNAs between HS and paired (non-scar, obtained from donor sites during scar resection) NS tissue. miRNA microarray expression profiling from three paired HS and NS tissues was performed. Differentially expressed miRNAs were identified by fold change and a P-value calculated using Student's *t*-test. The threshold set to identify up and downregulated genes was a fold change ≥2 and P<0.05. Red dots indicate points-of-interest that exhibit large-magnitude fold-changes (x-axis; log2 of the fold change) and high statistical significance (y-axis; -log10 of the P-value). (B) Hierarchical clustering showing differentially expressed miRNAs from HS samples compared with paired NS tissues. Each row represents one miRNA and each column represents one tissue sample. The relative miRNA expression is depicted according to the color scale. Red indicates upregulation and green indicates downregulation. N1-3 represents NS tissue samples, whereas H1-3 represents HS tissue samples. The differentially expressed miRNAs were clearly separated into clusters. miRNA, microRNA; hsa-miR, human microRNA; HS, hypertrophic scar; NS, normal skin.
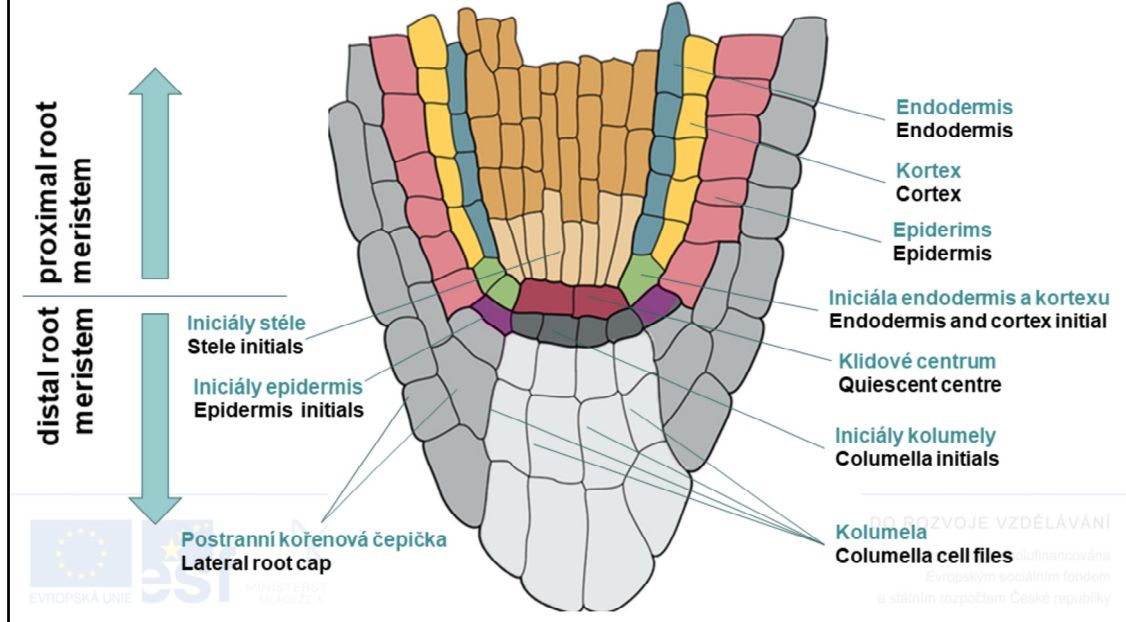
Inferring Gene Regulatory Networks
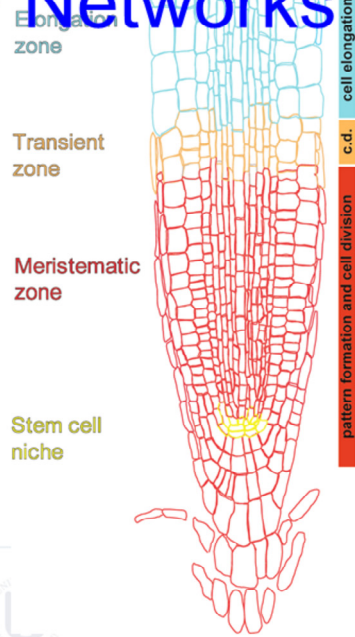
Benkova and Hejatko, *Plant Mol Biol* (2008)

41

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

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

Gene Regulatory Networks - GENIST

- Inferring GRNs via **GENIST**
  - **GE**ne regulatory **N**etwork **I**nference from **S**patio**T**emporal data algorithm
  - Combining spatial- and time-specific gene expression profiles
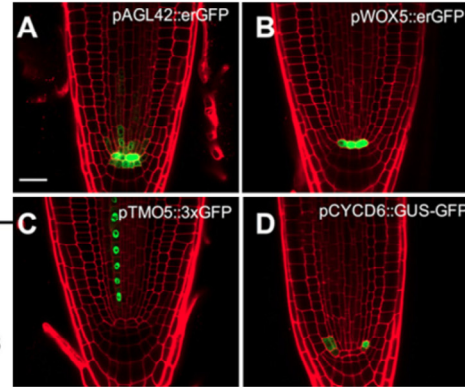
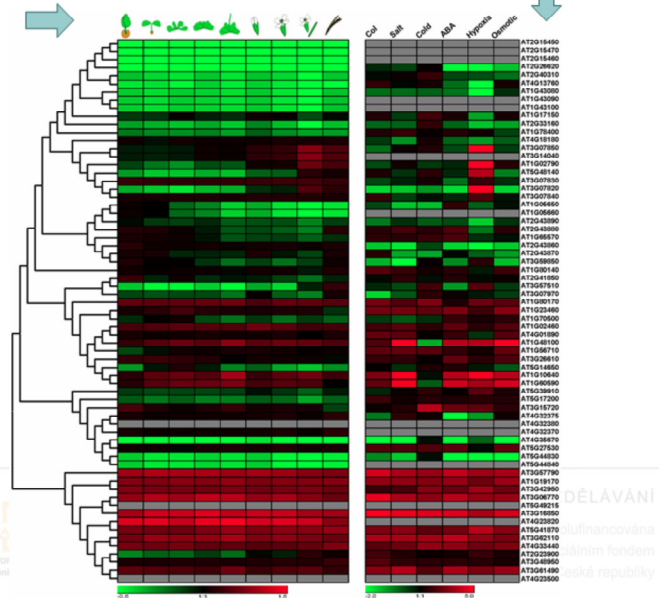de Luis Balaguer et al., *PNAS*, 2017

Birnbaum et al., *Science*, 2003

Combining Large Omics Datasets

TISSUE/TIME

GENES

44

GENIST algorithm

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

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

GENIST algorithm

The MATLAB source code for GENIST is publically available at https://github.com/madeluis/GENIST.
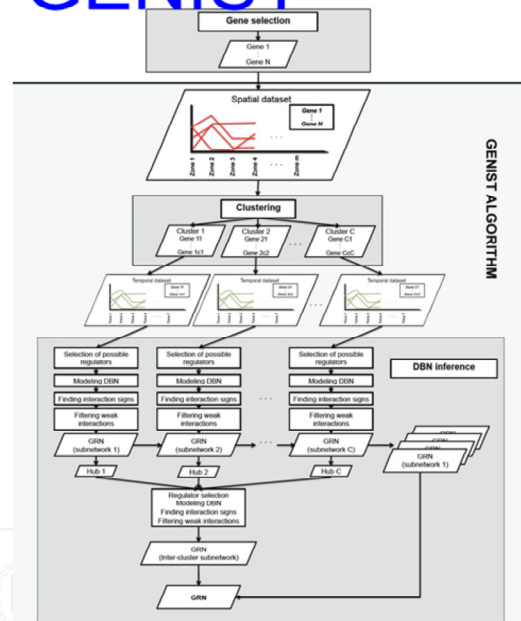
For the detailed description of the procedure, see de Luis Balaguer et al., 2017, SI (https://www.pnas.org/content/114/36/E7632/tab-figures-data)

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

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

Gene Regulatory Networks - GENIST
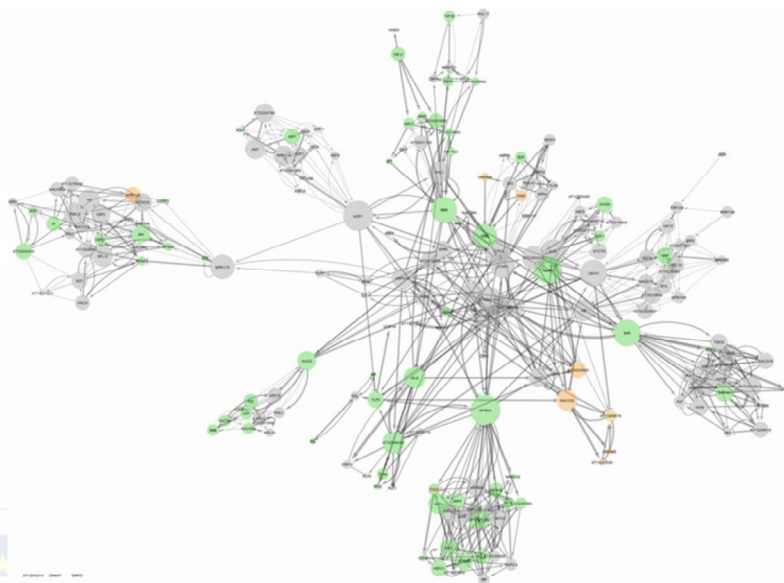
de Luis Balaguer et al., *PNAS*, 2017

Gene Regulatory Networks - GENIST
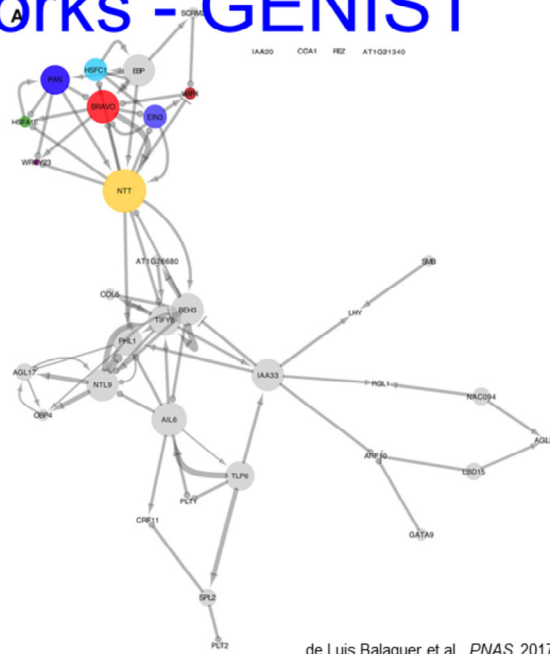
de Luis Balaguer et al., *PNAS*, 2017
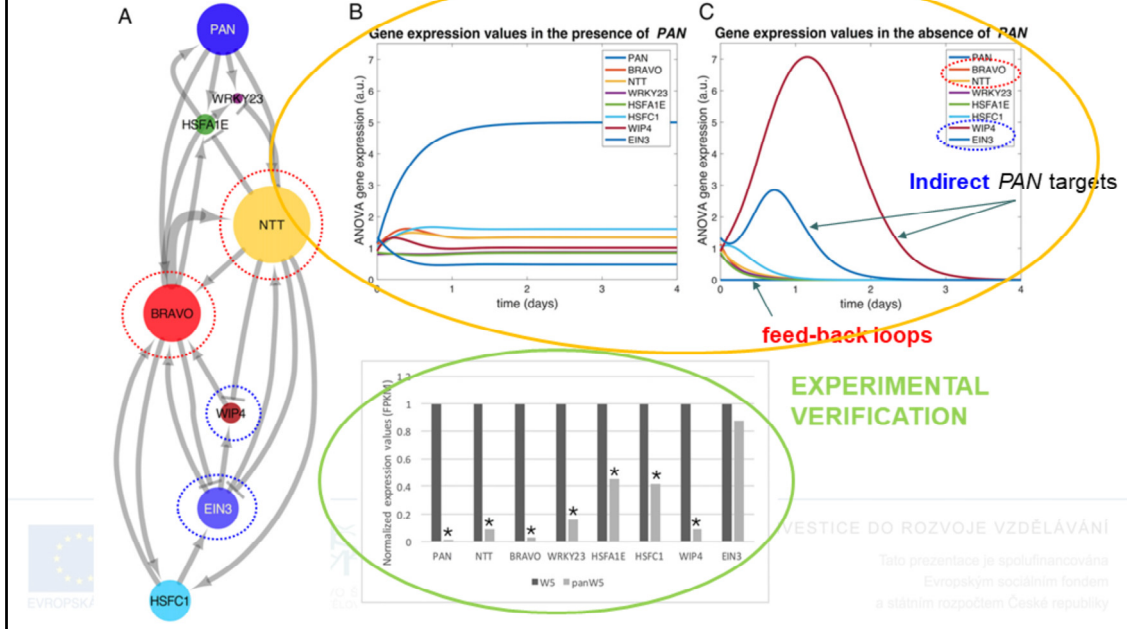
PAN subnetwork in the QC inferred with the 12 developmental time points of the Arabidopsis root. (A) Optimal configuration (combination of signs— activation or repression— of the regulations that were inferred with undefined signs, which best fits the data in the simulations of the equations) of the subnetwork of PAN and its downstream targets. (B and C) Resulting expression values of PAN and its downstream targets, over time, after simulating the optimal configuration of the model. Simulations were run for 5 d and plots are shown until all factors reached steady states in the WT and pan mutant simulations. (B)Model simulated with the fitted equation parameters. (C)Model simulated with the PAN-associated parameters set to zero to simulate a pan mutant situation. (D) Normalized expression values of PAN and its predicted downstream targets in Col-0 wild type and in pan mutant. Statistically significant changes of expression between the mutant and the wild type, *q < 0.05.

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

49

Summary

- Definition of Systems Biology

- Tools
  - Gene Ontology analysis
  - Bayesian Networks
  - Molecular/Gene Regulatory Networks Modeling
  - Inferring Gene Regulatory Networks from Large Omics Datasets

Discussion