CG920 Genomics

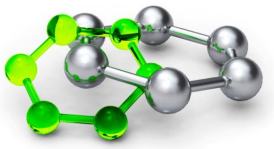
Lesson 6

Gene Expression and Chemical Genetics

Jan Hejátko

Functional Genomics and Proteomics of Plants,
CEITEC - Central European Institute of Technology
And
National Centre for Bimolecular Research,
Faculty of Science,

Masaryk University, Brno
hejatko@sci.muni.cz, www.ceitec.eu



Literature

- Literature resources for Lesson 06

- Brady, S. M. et al. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science.* **318** (5851), 801-806 (2007).
- Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP (2017) The Drosophila embryo at single-cell transcriptome resolution. *Science* **358**: 194-199
- 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
- Schonberger, J., Hammes, U.Z., and Dresselhaus, T. (2012). In vivo visualization of RNA in plants cells using the lambdaN(22) system and a GATEWAY-compatible vector series for candidate RNAs. *The Plant Journal* **71**, 173-181.
- Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science.* **353** (6294), 78-82 (2016).
- Xia, K. et al. The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in arabidopsis leaves. *Dev Cell.* **57** (10), 1299-1310 e1294 (2022)

CEITEC

Outline

- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases
    - Tissue- and cell-specific gene expression analysis
    - Spatial trascriptomics
  - Quantitative analysis of gene expression
    - DNA and protein chips
    - Next generation transcriptional profiling
- Regulation of gene expression in the identification of gene function by gain-of-function approaches
  - T-DNA activation mutagenesis
  - Ectopic expression and regulated gene expression systems

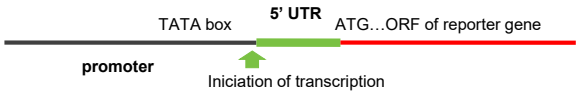
CEITEC

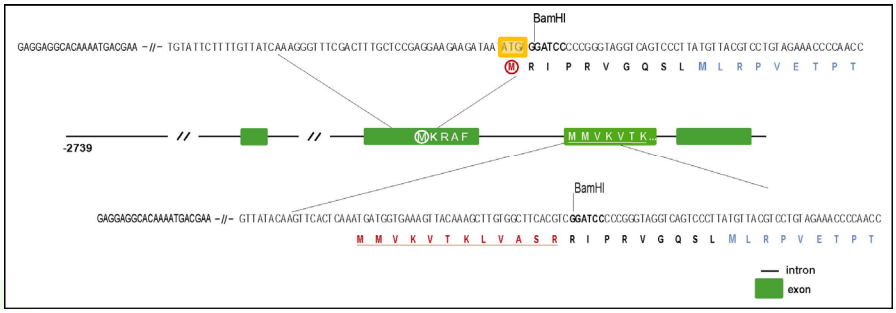
Outline

- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene

Transcriptional Fusion

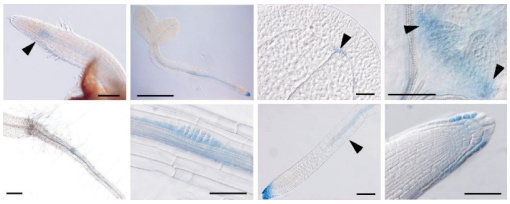
- Identification and cloning of the promoter region of the gene
- Preparation of recombinant DNA carrying the promoter and the reporter gene (uidA, GFP)

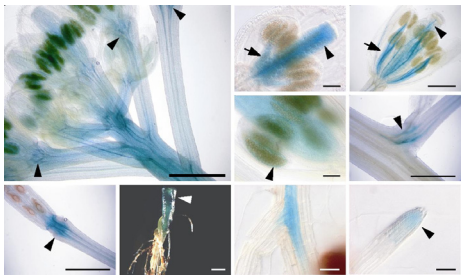




Transcriptional Fusion

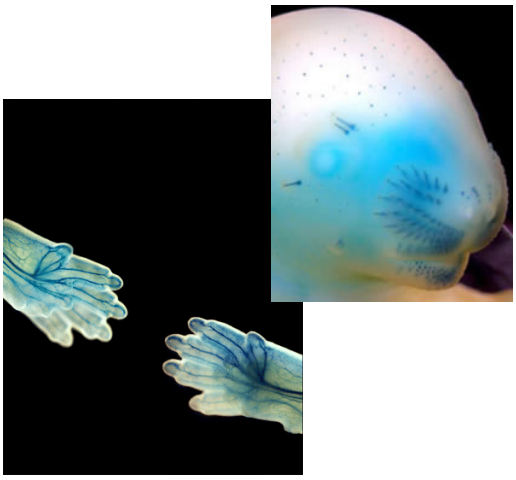
- Identification and cloning of the promoter region of the gene
- Preparation of recombinant DNA carrying the promoter and the reporter gene (uidA, GFP)
- Preparation of transgenic organisms carrying this recombinant DNA and their histological analysis





CEITEC

GUS Reporter in Mouse Embryos



CEITEC

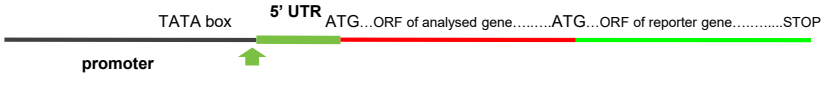
Outline

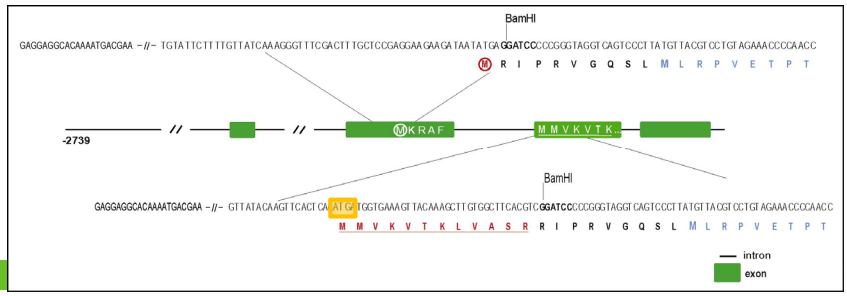
- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene

CEITEC

Translational Fusion

- Identification and cloning of the promoter and coding region of the analyzed gene
- Preparation of a recombinant DNA carrying the promoter and the coding sequence of the studied gene in a fusion with the reporter gene (uidA, GFP)

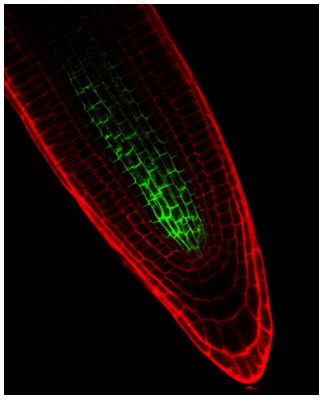




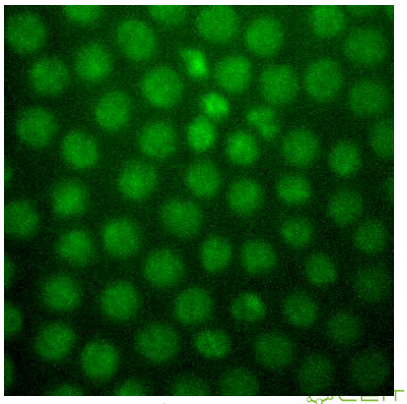
CEITEC

Translational Fusion

- Preparation of transgenic organisms carrying the recombinant DNA and their histological analysis
- Compared to transcriptional fusion, translation fusion allows analysis of intercellular localization of gene product (protein) or its dynamics

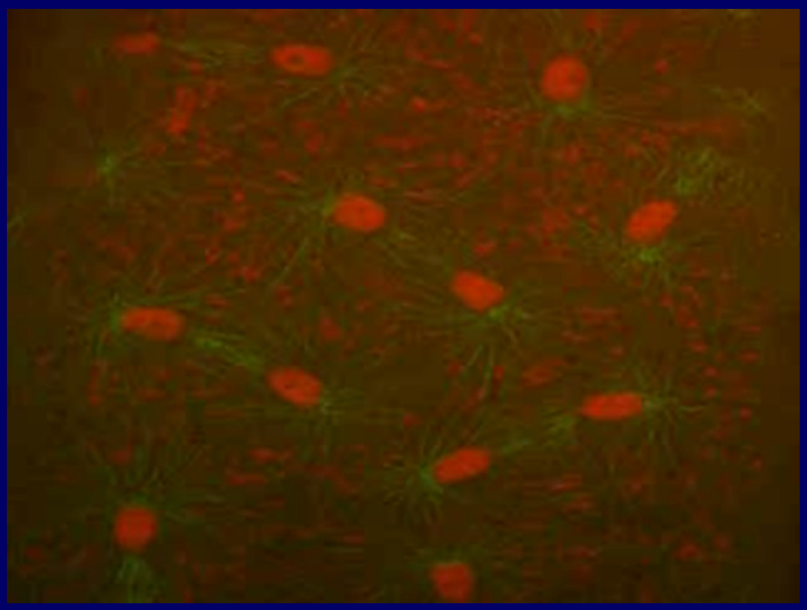


PIN1-GFP in *Arabidopsis*



Histone 2A-GFP in *Drosophila* embryo by PAM

Translational Fusion

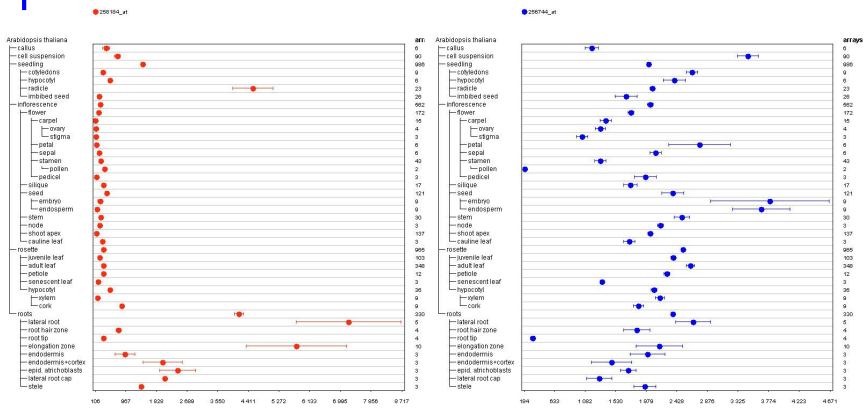


Outline

- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases

Databases

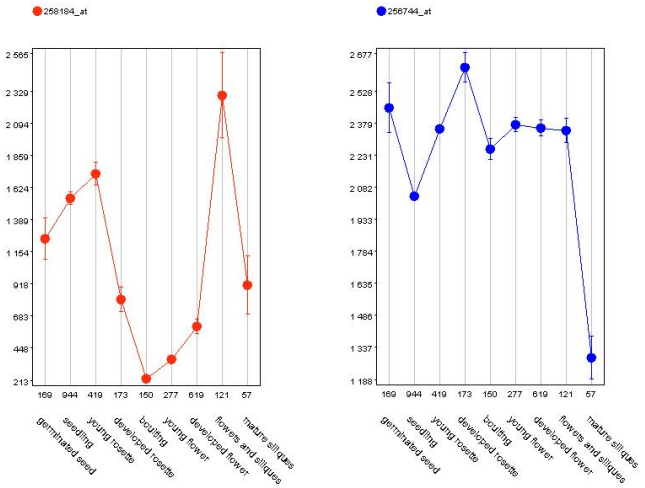
- **Analysis of expression using Genevestigator** (***AHP1*** and ***AHP2***, *Arabidopsis*, Affymetrix ATH 22K Array)



CEITEC

Databases

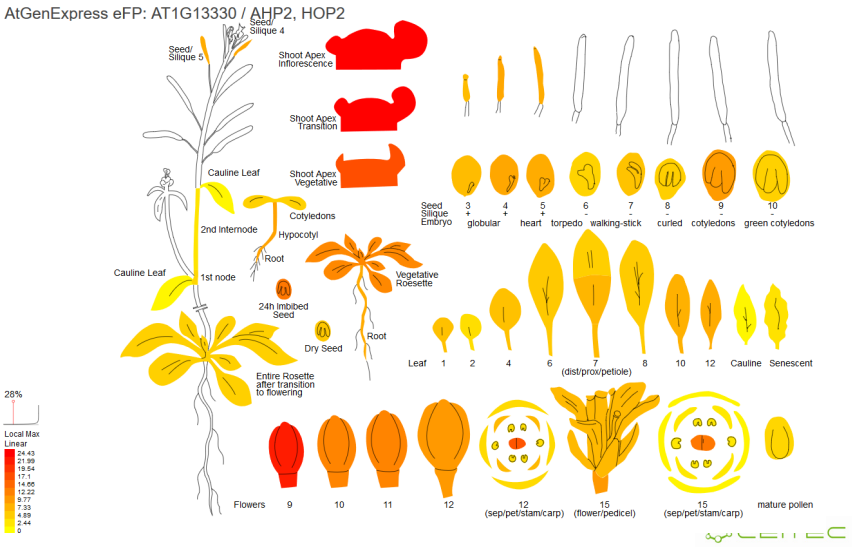
- **Analysis of expression using Genevestigator** (***AHP1*** and ***AHP2***, *Arabidopsis*, Affymetrix ATH 22K Array)



CEITEC

Databases

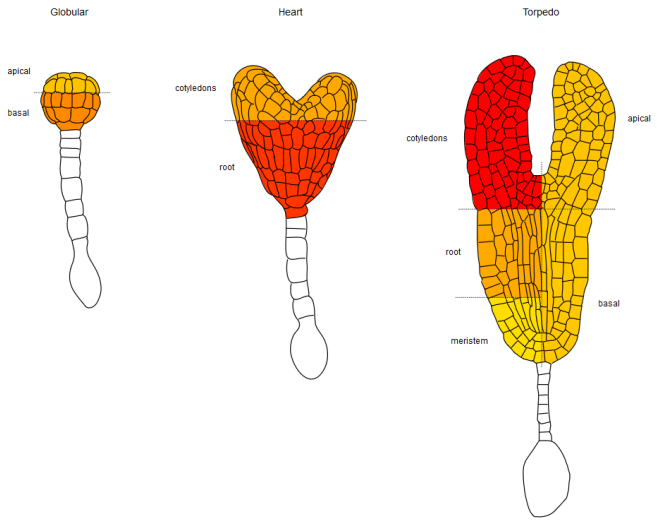
- **Analysis of expression using ePlant**



15

Databases

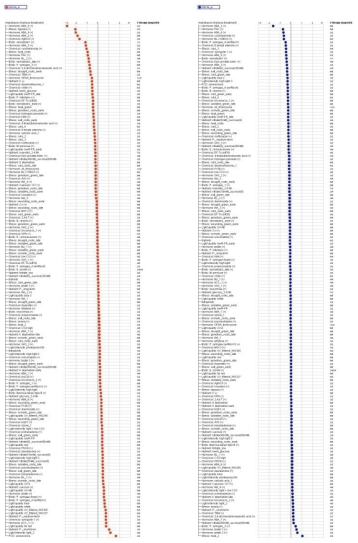
- **Analysis of expression using ePlant**



CEITEC

Databases

- **Analysis of expression using Genevestigator** (***AHP1*** and ***AHP2***, *Arabidopsis*, Affymetrix ATH 22K Array)

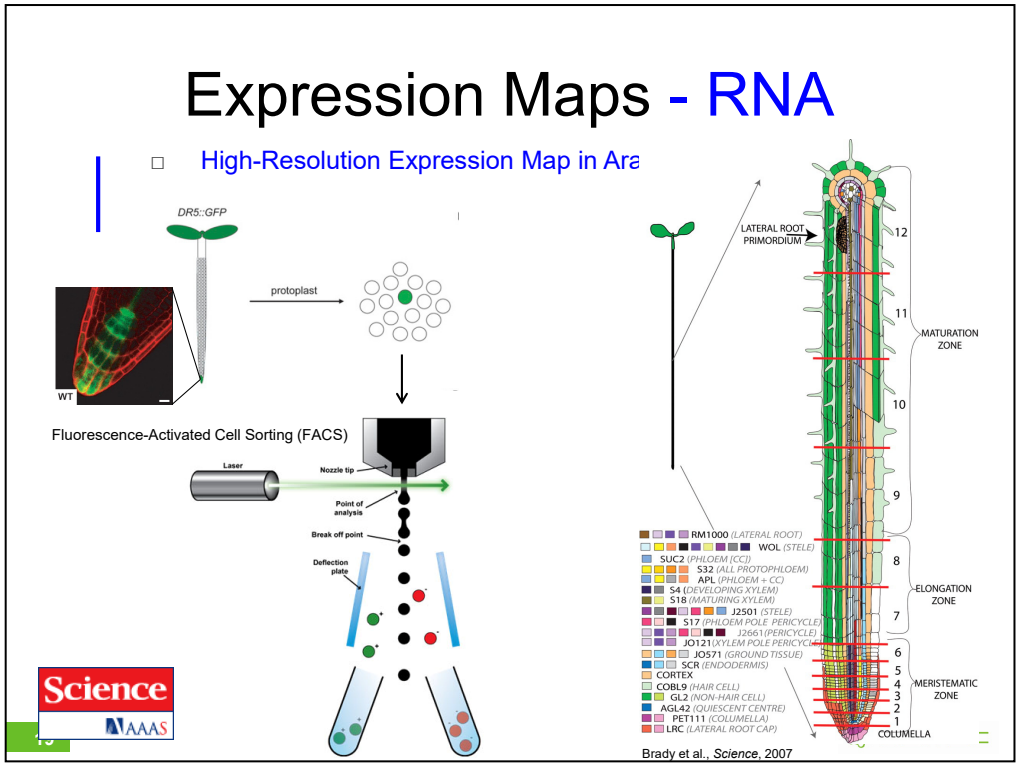


CEITEC

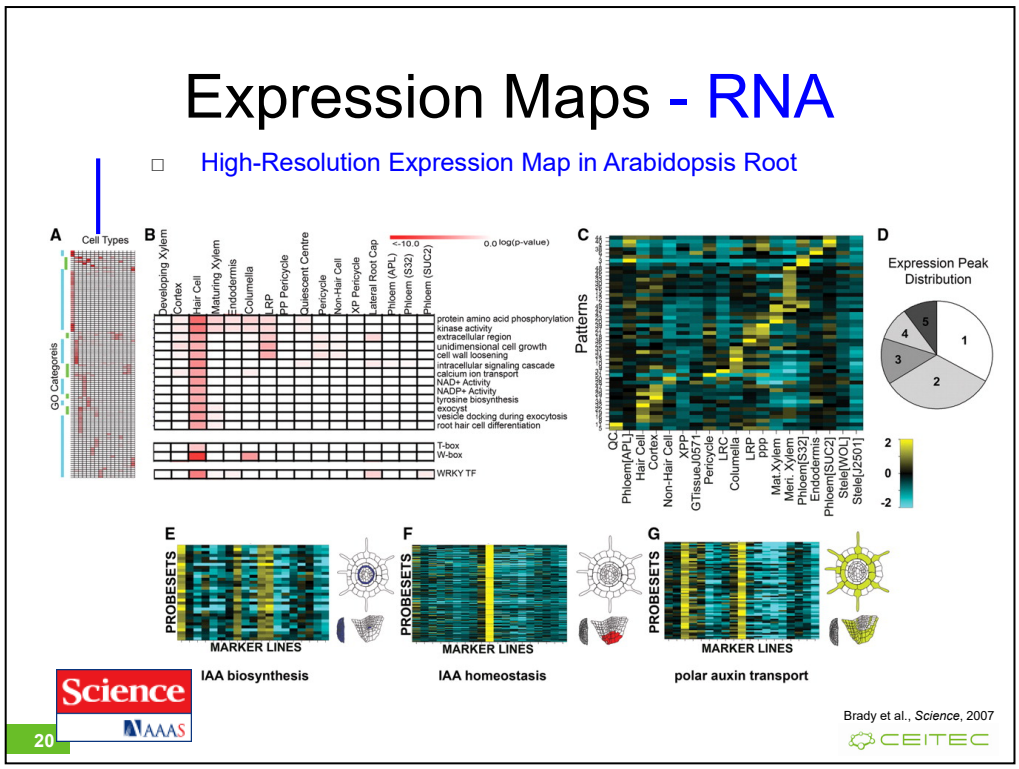
Outline

- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases
    - Tissue- and cell-specific gene expression analysis

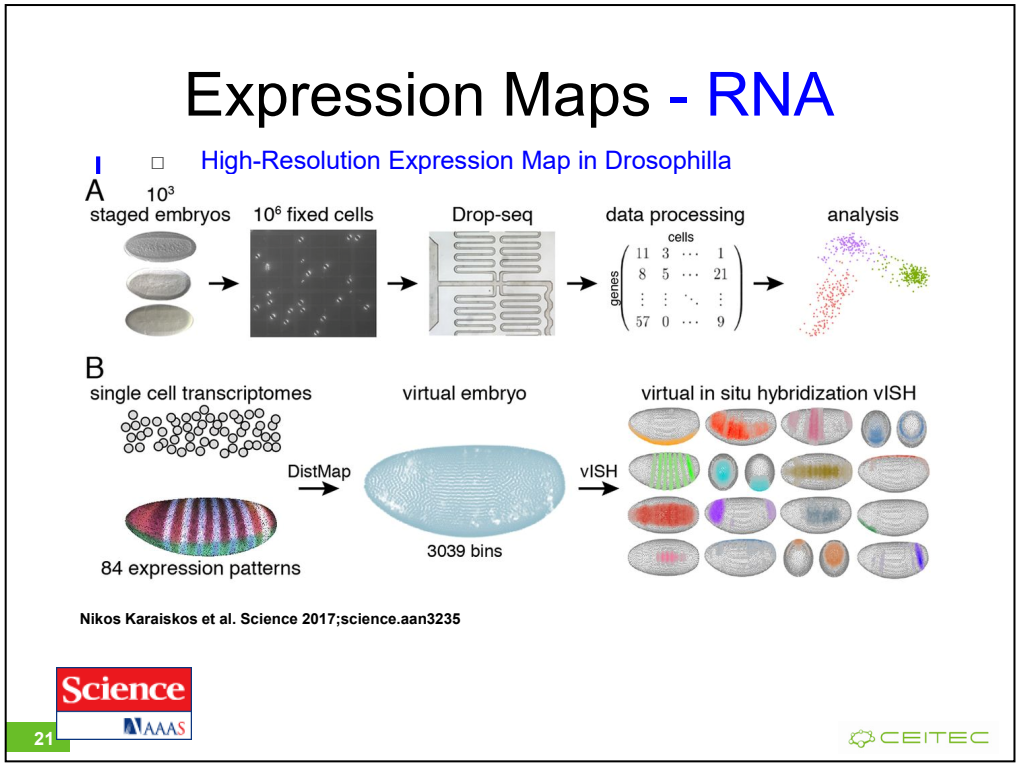
CEITEC



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



(**A**) The majority of enriched GO terms (hierarchically clustered) are associated with individual cell types (blue bar). A smaller number are present across multiple cell types (green bar). (fig. S2) (**B**) GO category enrichment for hair cells confirms a previous report (*15*). Enriched cis-elements and an enriched TF family were also identified. (**C**) From the top 50% of varying probe sets, 51 dominant radial patterns were identified. Pattern expression values were mean-normalized (rows) and $\log_2$ transformed to yield relative expression indices for each marker line (columns). Marker line order is the same for all figures; see table S1 for marker line abbreviations. (**D**) Pattern expression peaks were found across one to five cell types. (**E** to **G**) Patterns where expression is enriched in single and multiple cell types support transcriptional regulation of auxin flux and synthesis. In all heat maps with probe sets, expression values were mean-normalized and $\log_2$ transformed. Expression is false-colored in representations of a root transverse section, a cut-away of a root tip, and in a lateral root primordium. (E) Auxin biosynthetic genes (*CYP79B2, CYP79B3, SUPERROOT1*, and *SUPERROOT2*) are transcriptionally enriched in the QC, lateral root primordia, pericycle, and phloem-pole pericycle (P = 1.99E^{-11}, pattern 5). All AGI identifiers and TAIR descriptions are found in table S14. (F) Auxin amido-synthases *GH3.6* and *GH3.17* that play a role in auxin homeostasis show enriched expression in the columella, just below the predicted auxin biosynthetic center of the QC (P = 8.82E^{-4}, pattern 13). (G) The expression of the auxin transporter, *PIN-FORMED2*, and auxin transport regulators (*PINOID, WAG1*) are enriched in the columella, hair cells, and cortex (P = 1.03E^{-4}, pattern 31).



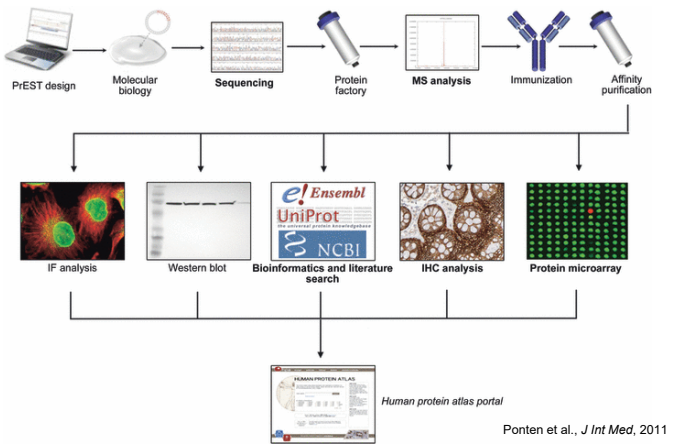
Deconstructing and reconstructing the embryo by single-cell transcriptomics combined with spatial mapping.

(A) Single-cell sequencing of the Drosophila embryo: ~1000 handpicked stage 6 fly embryos are dissociated per Drop-seq replicate, cells are fixed and counted, single cells are combined with barcoded capture beads, and libraries are prepared and sequenced. Finally, single-cell transcriptomes are deconvolved, resulting in a digital gene expression matrix for further analysis.

(B) Mapping cells back to the embryo: Single-cell transcriptomes are correlated with high-resolution gene expression patterns across 84 marker genes, cells are mapped to positions within a virtual embryo, and expression patterns are computed by combining the mapping probabilities with the expression levels (virtual in situ hybridization).

Expression Maps - Proteins

- Human Protein Atlas



Ponten et al., *J Int Med*, 2011

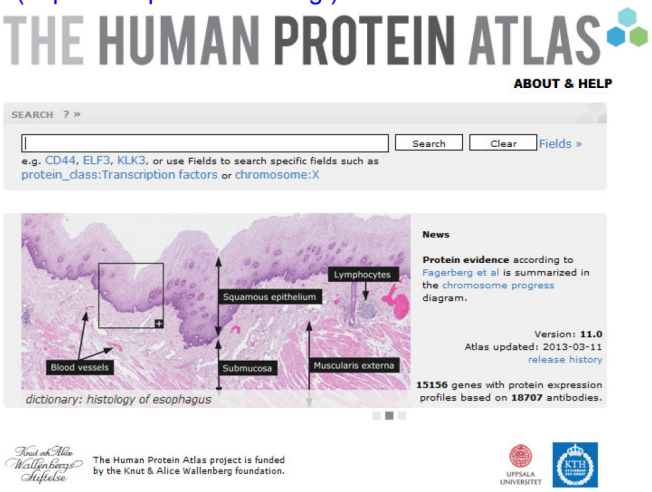
22

CEITEC

Schematic flowchart of the Human Protein Atlas. For each gene, a signature sequence (PrEST) is defined from the human genome sequence, and following RT-PCR, cloning and production of recombinant protein fragments, subsequent immunization and affinity purification of antisera results inmunospecific antibodies. The produced antibodies are tested and validated in various immunoassays. Approved antibodies are used for protein profiling in cells (immunofluorescence) and tissues (immunohistochemistry) to generate the images and protein expression data that are presented in the Human Protein Atlas (Ponten et al., *J Int Med*, 2011).

Expression Maps - Proteins

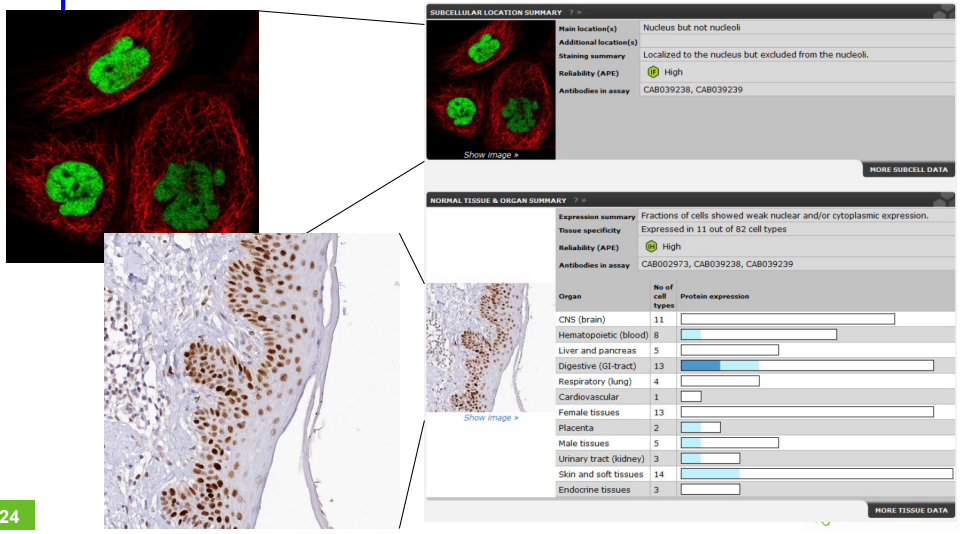
- Human Protein Atlas (http://www.proteinatlas.org/)



CEITEC

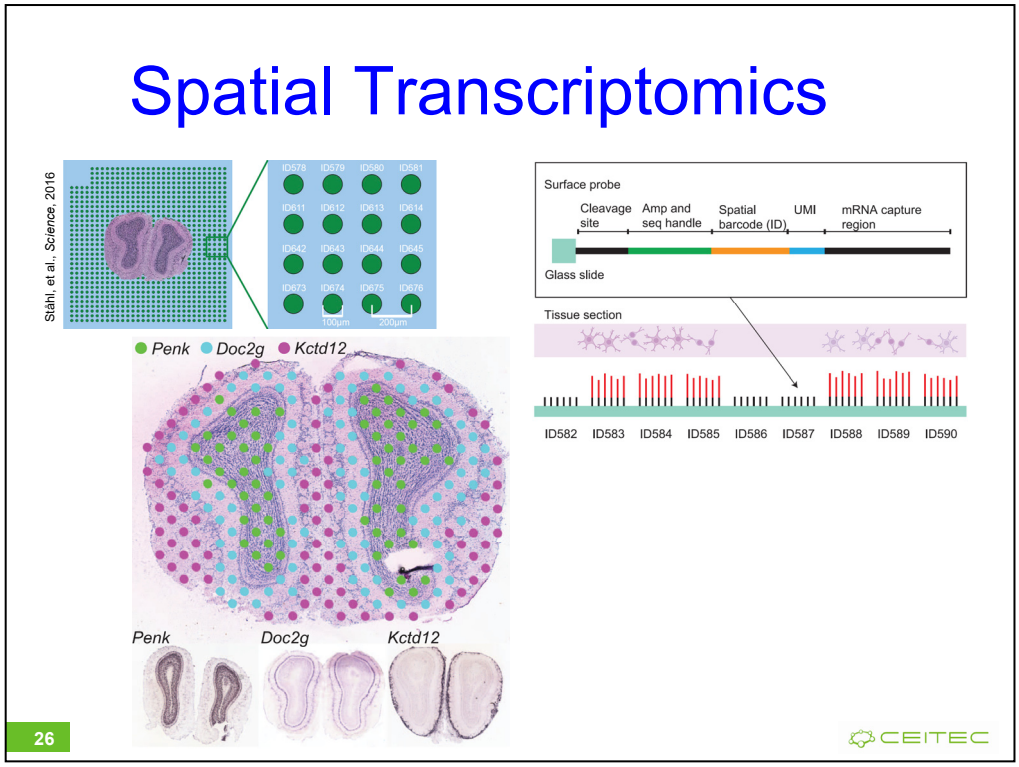
Expression Maps - Proteins

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



Outline

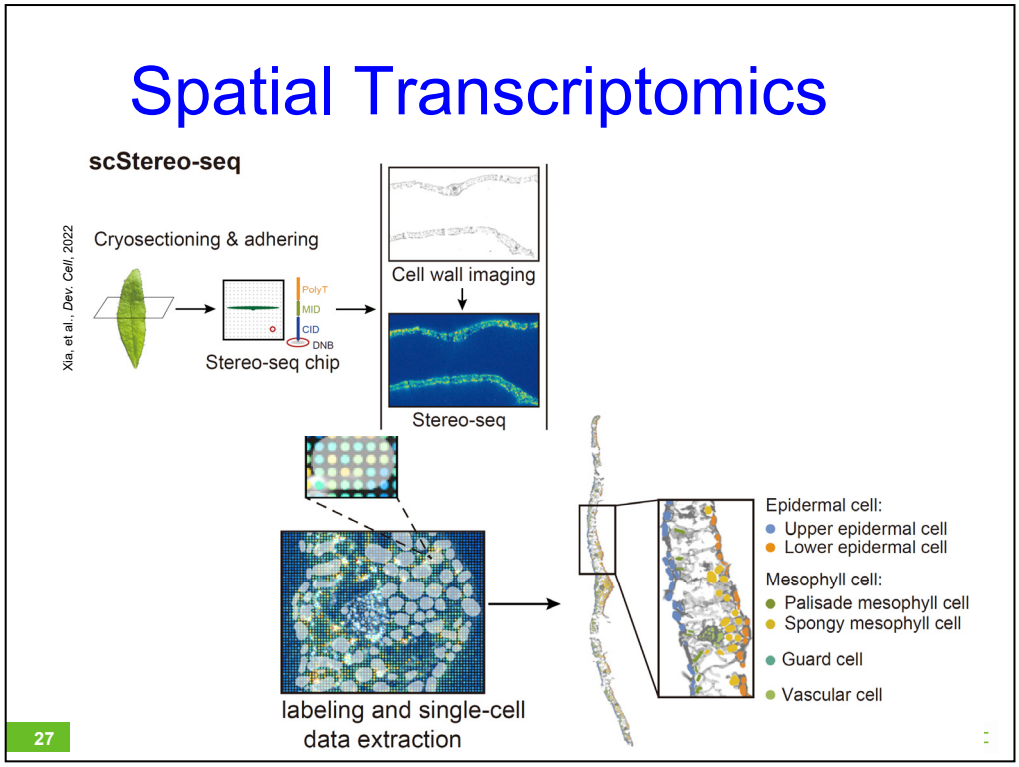
- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases
    - Tissue- and cell-specific gene expression analysis
    - Spatial trascriptomics



Spatially resolved gene expression. (A) Each array feature contains uniqueDNA-barcodedprobes containingacleavagesite,aT7amplificationand sequencing handle, a spatial barcode, a unique molecular identifier (UMI), and an oligo(dT) VN-capture region, where V is anything but Tand where N is any nucleotide. cDNA (red) is generated from captured mRNA by reverse transcription.

(B) Visualization of the expression of three genes by spatial transcriptomics (top) and in situ hybridization (bottom). Penk and Kctd12 in situ images are from the Allen Institute. Cutoff normalized counts, Penk, 8; Doc2g,

13; and Kctd12, 19.



Cauline leaves were cryo-sectioned and positioned on top of four separate chips (or sections) with seven leaf samples on each chip. On the surface of the chip, a DNA nanoball (DNB) is docked in a

grid-patterned array of spots. Each spot is 220 nm in diameter and the center-to-center distance between neighboring spots is 500 nm (Figure 1A). The DNB contains random barcoded sequences, the coordinate identity (CID), molecular identifiers (MIDs), and polyT sequence-containing oligonucleotides designed to capture mRNAs. After cell wall staining and imaging, the chips were used for Stereo-seq library construction and data acquirement. In short, mRNA was released from tissue cells through permeabilization and was captured by polyT in the DNB. The released mRNA was then reverse-transcribed and amplified into cDNA, which

was used for PCR amplification and library sequencing. The sequencing data were visualized in the STOmics visualization system (https://stereomap.cngb.org/) and processed using a series of Stereo-seq exclusive tools, including SAW (https://github.com/BGIResearch/SAW) and stereopy (https://github.com/BGIResearch/stereopy). In total, we selected data from 26 leaf samples with good morphology in these four chips for further analyses.

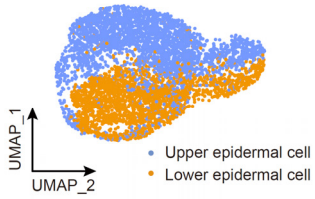
(A) Schematic representation of the single-cell Stereo-seq procedure. Arabidopsis thaliana cauline leaves are cryo-sectioned and positioned on top of the chip surface with DNA nanoball (DNB) docked in a grid-patterned array of spots, and the capture probes contaiidentifiers), and PloyT oligos to enable the recordation of the spatial coordinates, the identification of unique transcripts per gene, and the capture of mRNAs.

After cell wall staining and imaging, the same section is sequenced with Stereo-seq. Through the combined high-resolution image and MIDs, a single-cell level of MID distribution is achieved. A robust extraction method is built to be used in extracting single cells and in the identification of major cell types in cauline leaves. Using spatial single-cell data, several cell subtypes are distinguished (i). Next, the leaf is divided into four distinct parts, and spatial gene expression pattern (ii) and spatial developmental trajectory (iii) are determined.

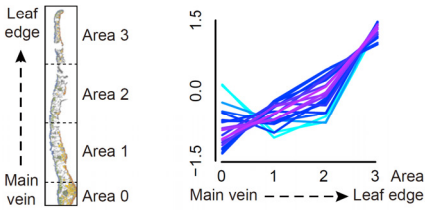
Spatial Transcriptomics

Spatial single-cell transcriptome analysis

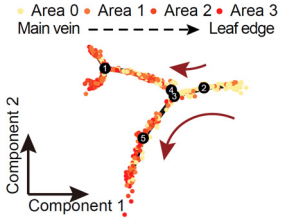
(i) Cell sub-type identification



(ii) Spatial gene expression pattern



(iii) Spatial developmental trajectory



CEITEC

Outline

- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases
    - Tissue- and cell-specific gene expression analysis
    - Spatial trascriptomics
  - Quantitative analysis of gene expression
    - DNA and protein chips

DNA Chips

- Method, which provides quick comparison of a large number of genes/proteins between the test sample and control
- Oligo DNA chips are used the most

- There are commercialy available kits for the whole genome
  - company Operon (Qiagen), 29.110 of 70-mer oligonucleotides representing 26.173 genes coding proteins, 28.964 transcripts and 87 microRNA genes of *Arabidopsis thaliana*
  - Possibility of use for the preparation of photolithography chips – facilitation of oligonucletide synthesis e.g. for the whole human genome (about 3,1 x 10^9 bp) jit is possible to prepare 25-mers in only 100 steps, by this technique

- Chips not only for the analysis of gene expression, but also for e.g. Genotyping (SNPs, sequencing with chips, …)

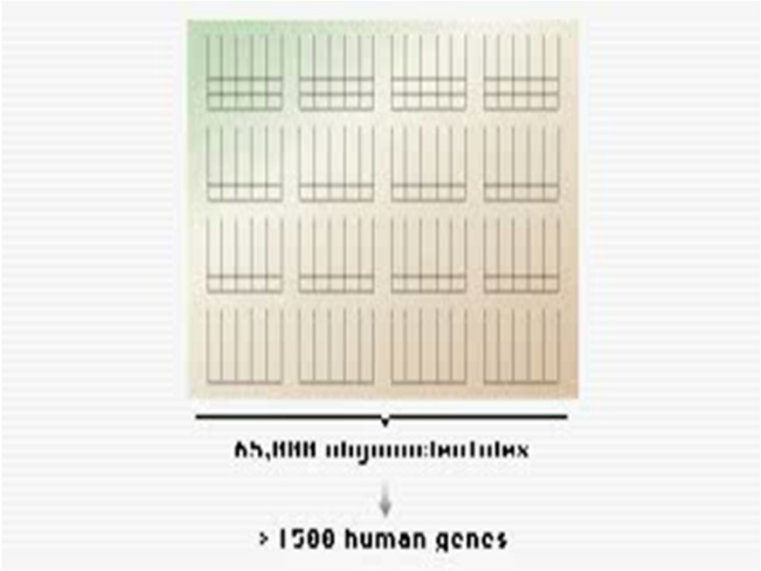
Affymetrix ATH1 *Arabidopsis* genome array

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

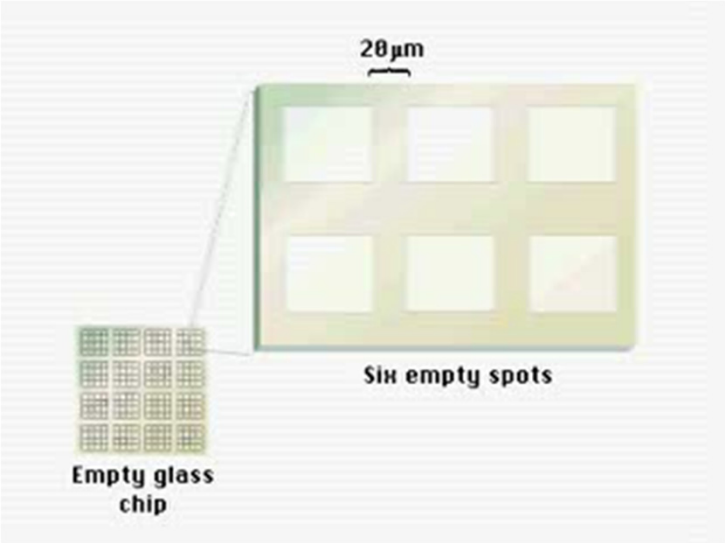
*As measured by detection in comparative analysis between a complex target containing spiked control transcriptions and a complex target with no spikes.

30

DNA Chips

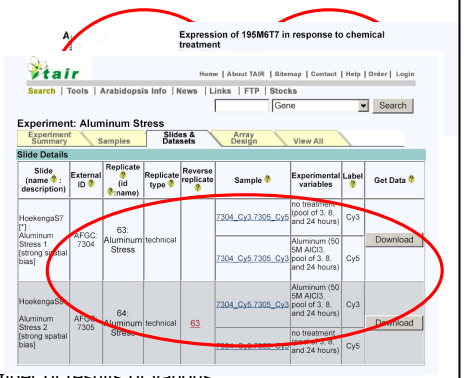


Photolitography



DNA Chips

- For the correct interpretation of the results, good knowledge of advanced statistical methods is required
- It is necessary to include a sufficient number of controls and repeats
  - Control of accuracy of the measurement (repeated measurements on several chips with the same sample, comparing the same samples analysed on different chips with each other)
  - Control of reproducibility of measurements (repeated measurements with different samples isolated under the same conditions on the same chip – comparing with each other)
  - Identification of reliable measurement treshold
  - Finally comparing the experiment with the control or comparing different conditions with each other -> the result



- Currently there's been a great number of results of various experiments in publicly accessible databases

Che et al., 2002

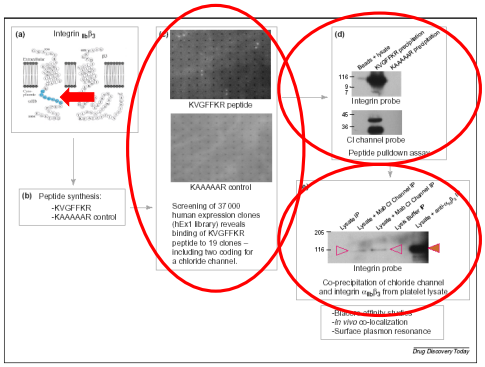
CEITEC

Protein Chips

- Protein chips
  - Chips with high density containing 10^4 proteins
  - Analysis of protein-protein interactions, kinase substrates and interactions with small molecules
  - Possibility of using antibodies – more stable than proteins

Protein Chips

- Identification of proteins interacting with integrin $\alpha_{IIb}\beta_3$ cytoplasmic domain of platelets
  - Expression of cytoplasmic part as a fusion peptide biotin-KVGFFKR
  - Analysis of binding to the protein chip containing 37.000 clones of *E. coli* expressing human recombinant proteins
  - Confirmation of interaction by pull-down analysis of peptides and by coprecipitation of whole proteins as well (e.g. chloride channel Icln)
  - Other use: e.g. in the identification of kinase substrates, when substrates are bound to the chip and exposed to kinases in the presense of radiolabeled ATP (786 purified proteins of barely, of which 21 were identified as CK2α kinase substrates; Kramer et al., 2004)



Lueking et al., 2005

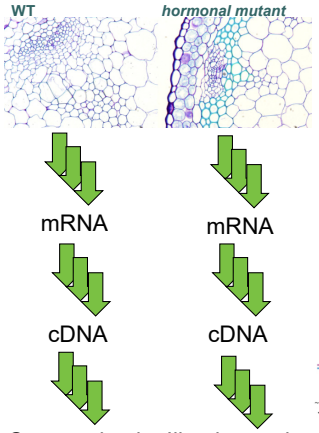
CEITEC

Outline

- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases
    - Tissue- and cell-specific gene expression analysis
    - Spatial trascriptomics
  - Quantitative analysis of gene expression
    - DNA and protein chips
    - Next generation transcriptional profiling

Next Gen Transcriptional Profiling

- ***Transcriptional profiling*** via ***RNA sequencing***



Sequencing by Illumina and **number of transcripts** determination

CEITEC

Results of –omics Studies vs Biologically Relevant Conclusions

- Transcriptional profiling yielded more then **7K differentially regulated genes**…

Ddii et al., *unpublished*

gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1,1804	1.79769e+308	1.79769e+308	6.88885e-05	0,00039180 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	0,0049681 6	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

38

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.

Outline

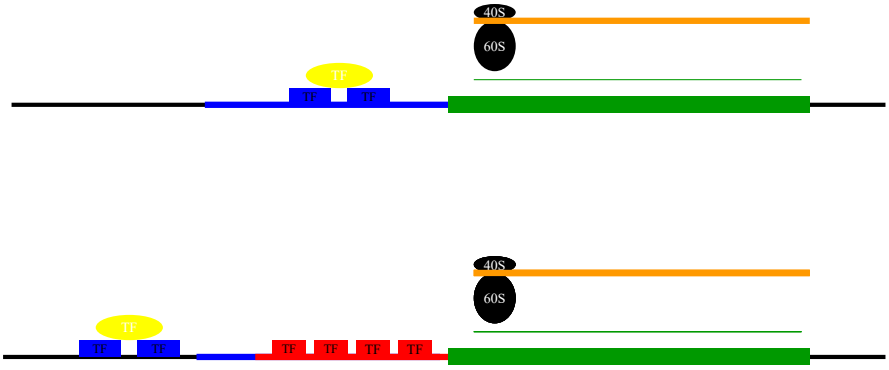
- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases
    - Tissue- and cell-specific gene expression analysis
    - Spatial trascriptomics
  - Quantitative analysis of gene expression
    - DNA and protein chips
    - Next generation transcriptional profiling
- Regulation of gene expression in the identification of gene function by gain-of-function approaches
  - T-DNA activation mutagenesis

CEITEC

Gain-of-Function Approaches

- Methods for identification of gene function using gain-of-function approaches
  - T-DNA activation mutagenesis
    - Method enabling isolation of dominant mutants by random insertion of constitutive promoter, resulting in overexpression of the gene and therefore in corresponding phenotypic changes
    - First step: preparation of mutant library prepared by tansformation of a strong constitutive promoter or enhancer
    - Next step: search of interesting phenotypes
    - Identification of the affected gene, e.g. by plasmid-rescue

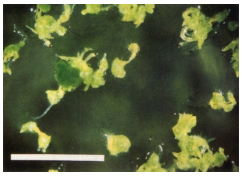
Activation Mutagenesis



CEITEC

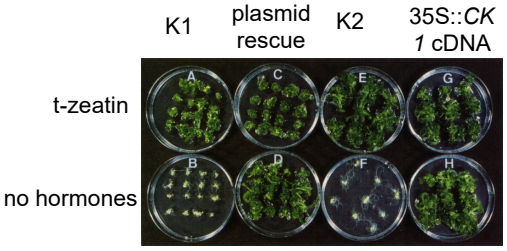
Isolation of *CKI1* Gene

- Tatsuo Kakimoto, *Science* 274 (1996), 982-985 *
- Isolation of the gene using activation mutagenesis



*

- Mutant phenotype is a phenocopy of exogenous application of cytokinins (*CKI1*, ***C****YTO****K****ININ* ***I****NDEPENDENT 1*)



*

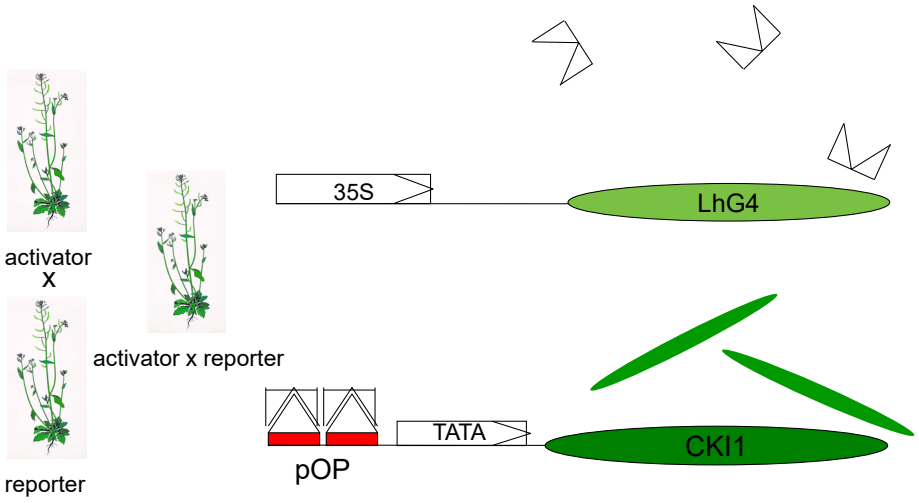
CEITEC

Outline

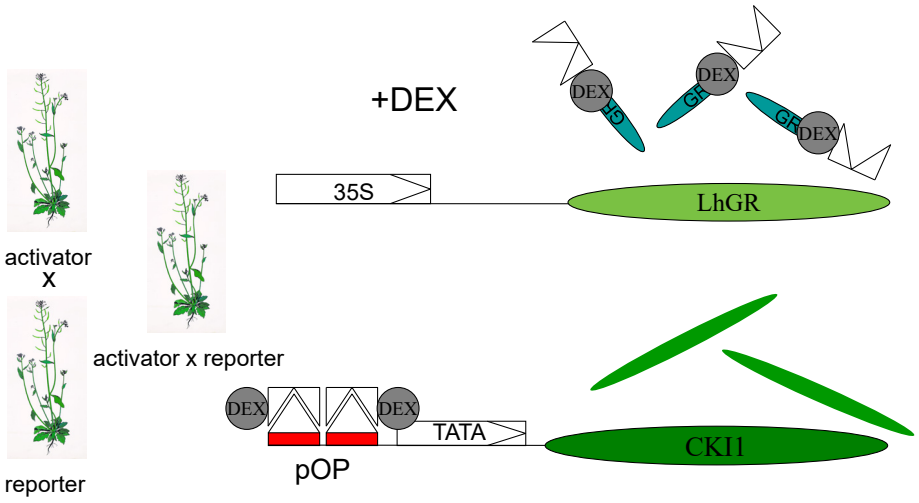
- Methods of gene expression analysis
  - Qualitative analysis of gene expression
    - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene
    - Preparation of translational fusion of the coding region of the analysed gene with reporter gene
    - Use of the data available in public databases
    - Tissue- and cell-specific gene expression analysis
    - Spatial trascriptomics
  - Quantitative analysis of gene expression
    - DNA and protein chips
    - Next generation transcriptional profiling
- Regulation of gene expression in the identification of gene function by gain-of-function approaches
  - T-DNA activation mutagenesis
  - Ectopic expression and regulated gene expression systems

CEITEC

Regulated Expression Systems

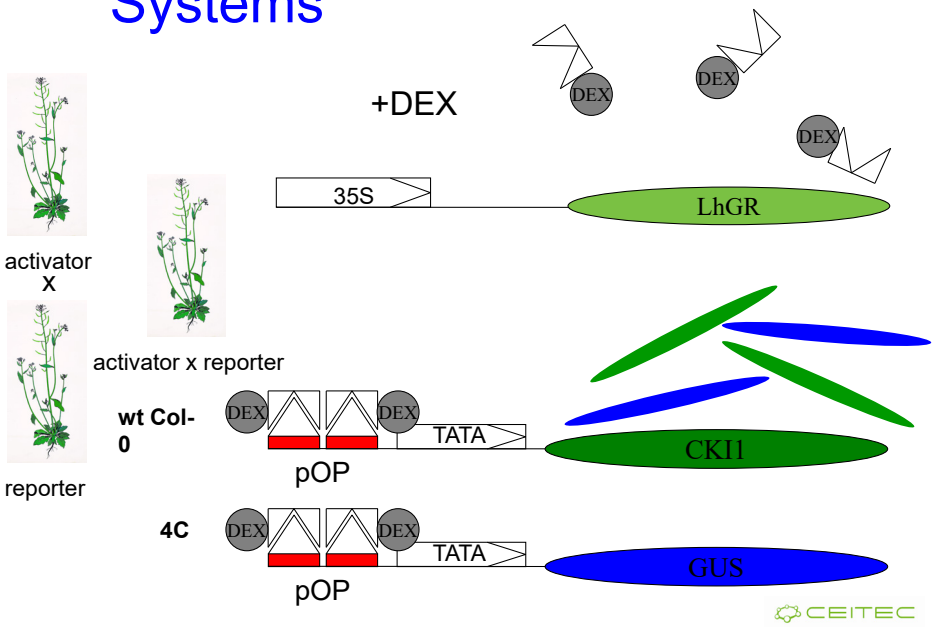


Regulated Expression Systems



CEITEC

Regulated Expression Systems



CEITEC

Regulated Expression Systems

- Regulated transgene expression systems
  - Allow time- or site-specific regulation of gene expression, leading to a change in phenotype and thereby identification of the natural function of the gene
    - pOP system
    - UAS system

CEITEC

UAS System

http://www.plantsci.cam.ac.uk/Haseloff/

Key Concepts

- Gene expression has spatiotemporal specificity
  - Analysis of spatiotemporal specificity of gene expression using
    - Transcriptional fusion of the promoter of analyzed gene with reporter gene
    - Translational fusion of coding region of teh assayed gene with reporter gene
    - Publicly accessible databases frequently with s cellular resolution
  - Quantitative analysis of gene expression
    - DNA and proteinové chips
    - Next gen transcriptional profiling
- Via regulating gene expression it is possible to identify gene function – gain of function approaches

Discussion