

C6215 Advanced Biochemistry and its Methods

Lesson 1

Introduction into Genomics

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Outline

- Definition Of Genomics
- Forward vs Reverse Genetics
- Genes Structure and Identification
- Nucleic Acid Sequencing
- Analysis of Gene Expression

Outline

- Definition Of Genomics



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

GENOMICS – What is it?

- *Sensu lato* (in the broad sense) – it is interested in **STRUCTURE and FUNCTION** of genomes
 - Necessary prerequisite: knowledge of the genome (sequence) – work with databases
- *Sensu stricto* (in the narrow sense) – it is interested in **FUNCTION** of **INDIVIDUAL GENES** – **FUNCTIONAL GENOMICS**
 - It uses mainly the reverse genetics approaches

Outline

- Definition Of Genomics
- **Forward vs Reverse Genetics**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

GENOMICS – What is it?

The role of BIOINFORMATICS in FUNCTIONAL GENOMICS

Forward („classical“) Genetics Approaches

Reverse Genetics Approaches

5'TTATATATATATATTAATAAAATAAAATAAAA
GAACAAAAAGAAAATAAAATA....3'



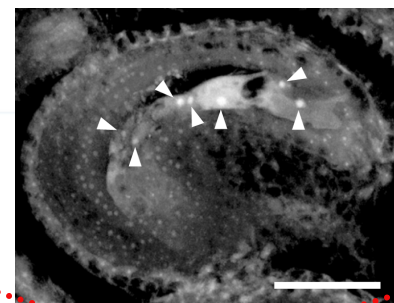
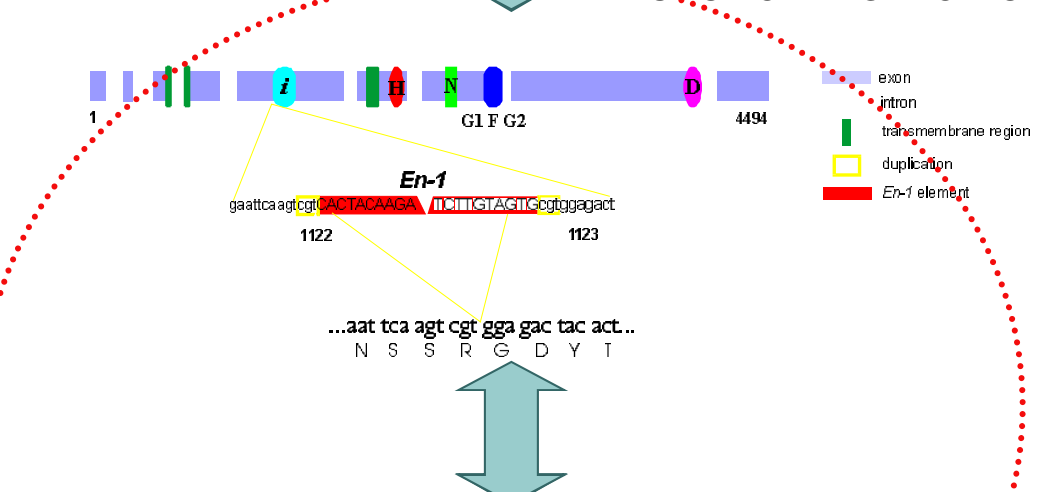
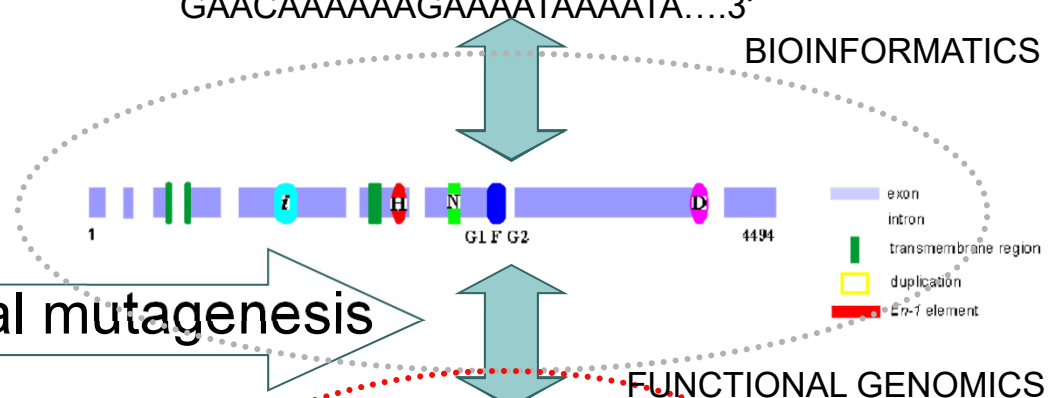
3

:

1



?



EVROPSKÁ UNIE



MLÁDEŽE A TĚLOVÝCHOVY

pro konkurenceschopnost



pro konkurenceschopnost

VOJE VZDĚAVÁNÍ

ntace je spolufinancována

agovými sociálními fondem

zpočetm České republiky

Outline

- Definition Of Genomics
- Forward vs Reverse Genetics
- Genes Structure and Identification

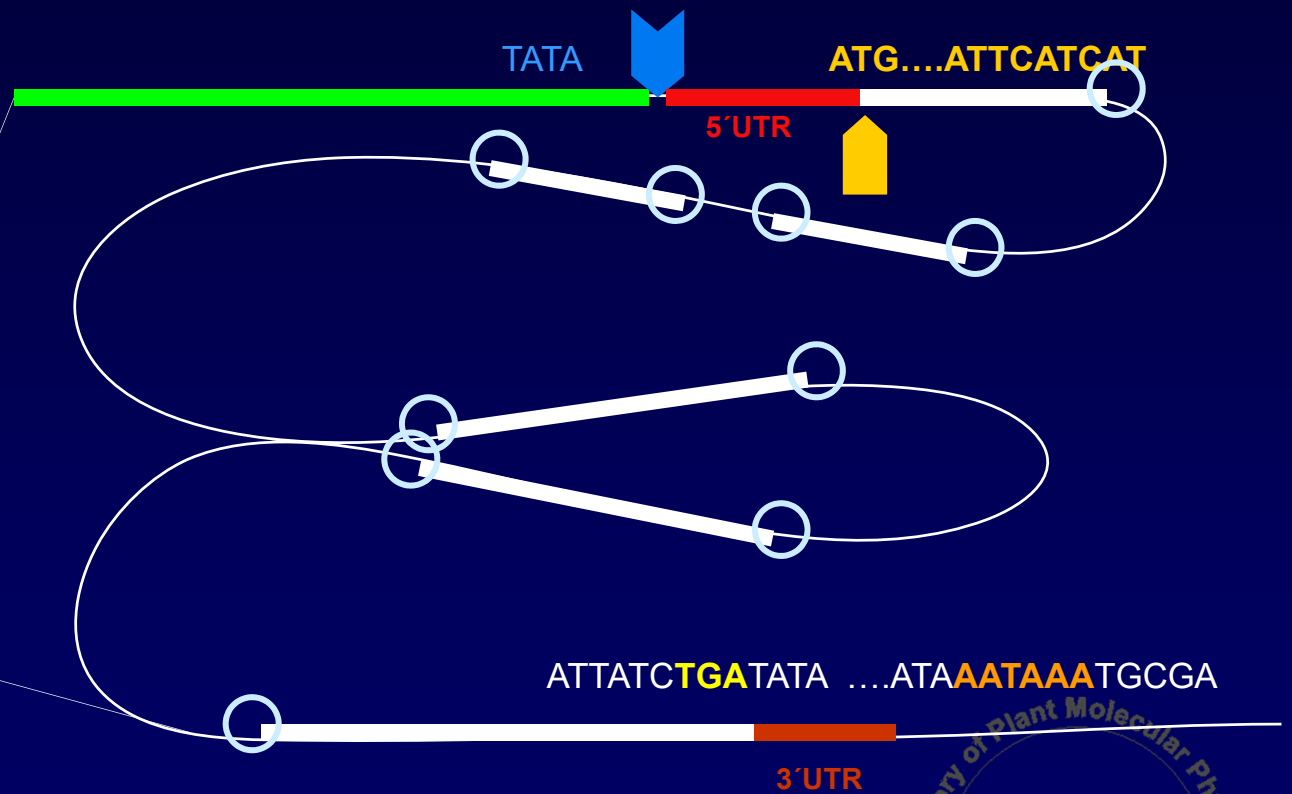


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Genes Structure

- Promoter
- Transcriptional start
- 5'UTR
- Translational start
- Splicing sites
- Stop codon
- 3'UTR
- Polyadenylation signal



Identification of Genes *Ab Initio*

- Omitting 5' and 3' UTR
- Identification of translation start (ATG) and stop codon (TAG, TAA, TGA)
- Finding donor (typically GT) and acceptor (AG) splicing sites
- Many ORFs are not real coding sequences – in *Arabidopsis*, there are on average approximately 350 milion ORFs in every 900 bp of sequence(!)
- Using various statistic models (e.g. Hidden Markov Model – HMM, see recommended literature, Majoros *et al.*, 2003) to evaluate and score the weight of identified donor and acceptor sites

Experimental Gene Identification

- Principles of experimental identification of genes using forward and reverse genetics
 - Alteration of phenotype after mutagenesis
 - **Forward genetics**
 - Identification of sequence-specific mutant and analysis of its phenotype
 - **Reverse genetics**
 - Analysis of expression of a particular gene and its spatiotemporal specificity

Forward Genetics

- Principles of experimental identification of genes using forward and reverse genetics
 - Alteration of phenotype after mutagenesis
 - **Forward genetics**

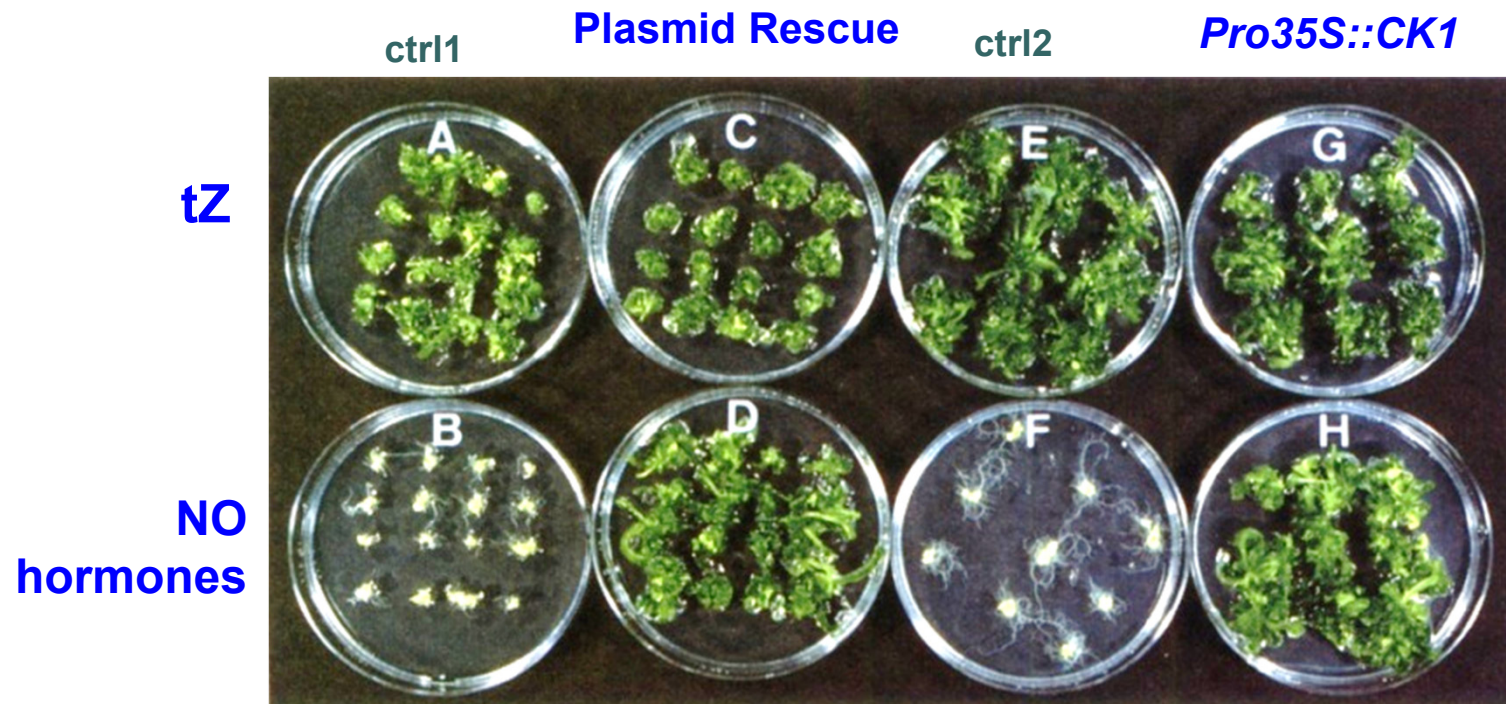


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

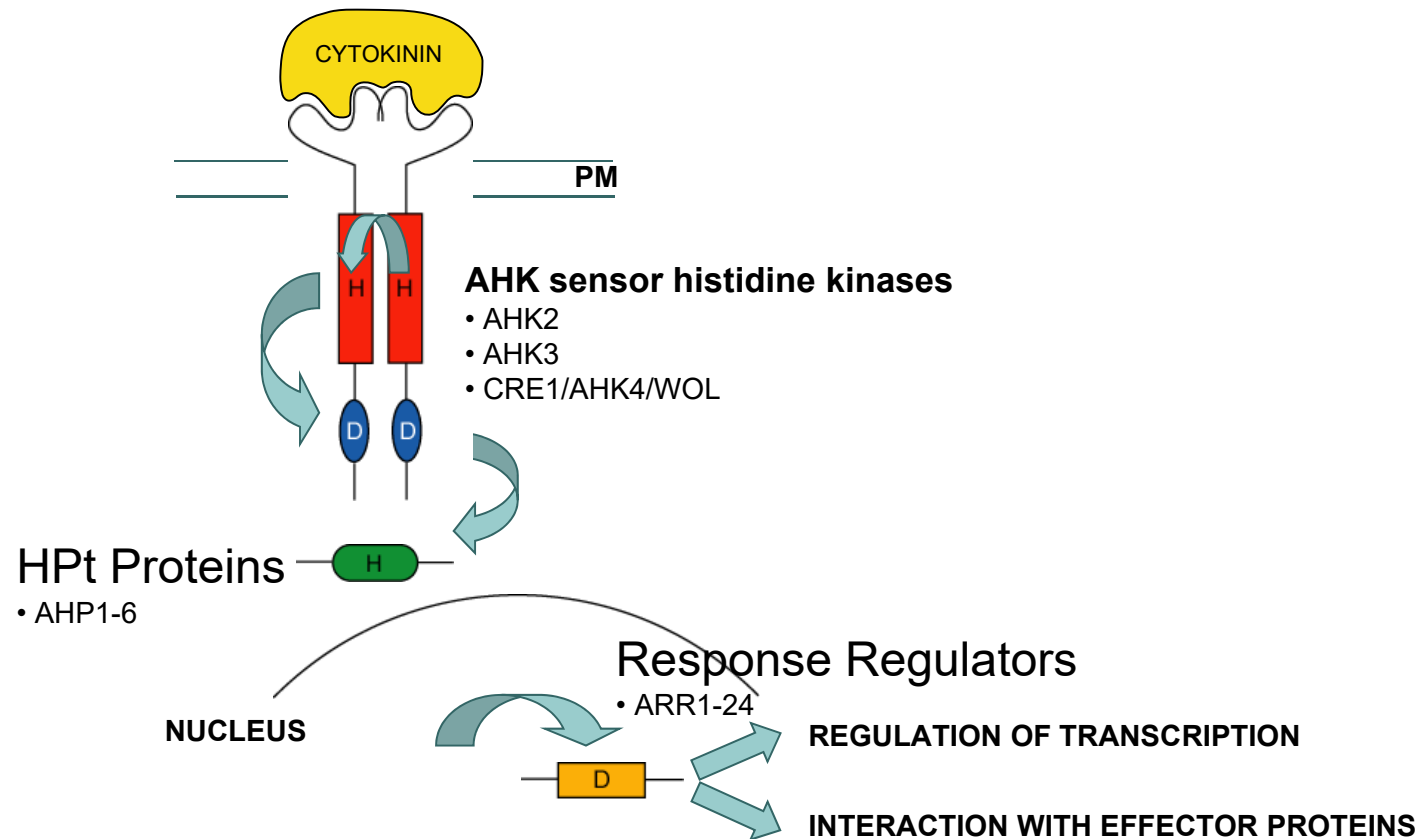
Identification of *CK11* via Activation Mutagenesis

- *CK11* overexpression mimics cytokinin response



Kakimoto, *Science*, 1996

Signal Transduction via MSP



Reverse Genetics

- Principles of experimental identification of genes using forward and reverse genetics
 - Alteration of phenotype after mutagenesis
 - **Forward genetics**
 - Identification of insertional mutant and analysis of its phenotype
 - **Reverse genetics**

Identification of insertional *cki1* mutant allele



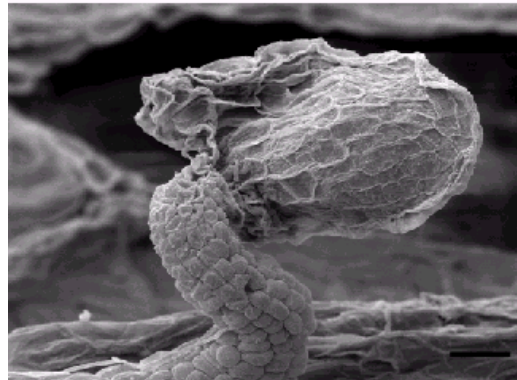
aattcaagtcgctCACTACAAGA " *En-1* TCTTGTAGTGCgtggagact

A. aat tca agt **cg t gga gac tac** act tgg tac act caa acc gtg gat cag tta act ggt
N S S **R G D Y** T W Y T Q T V D Q L T G

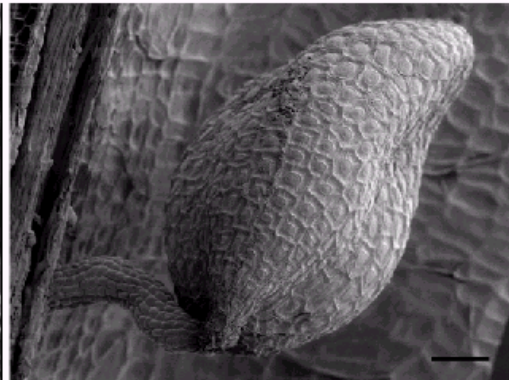
CKI1 Regulates Female Gametophyte Development

- CKI1 is necessary for proper megagametogenesis in *Arabidopsis*

CKI1/cki1-i



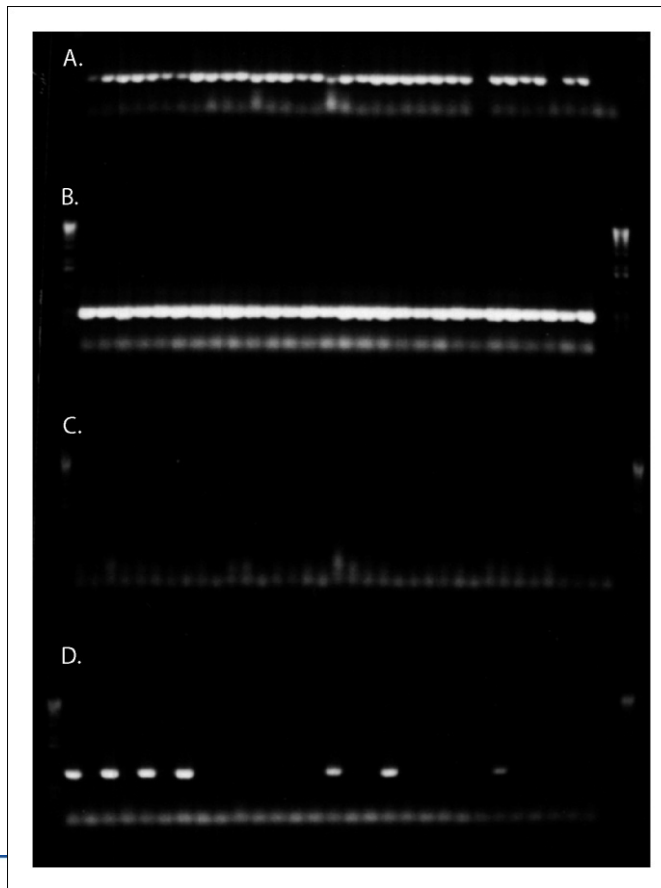
CKI1/CKI1



Hejátko et al., *Mol Genet Genomics* (2003)

CKI1 and Megagametogenesis

- *cki1-i* is not transmitted through the female gametophyte



A. ♂ wt x ♀ **CKI1/cki1-i**



CKI1 specific primers (PCR positive control)

B. ♂ **CKI1/cki1-i** x ♀ wt

C. ♂ wt x ♀ **CKI1/cki1-i**

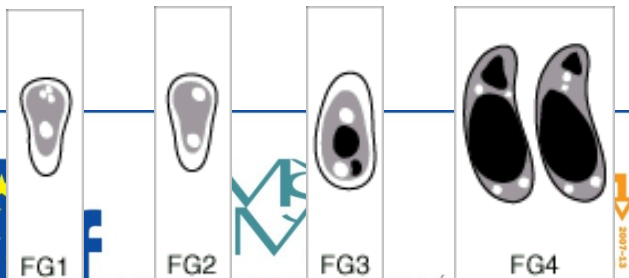
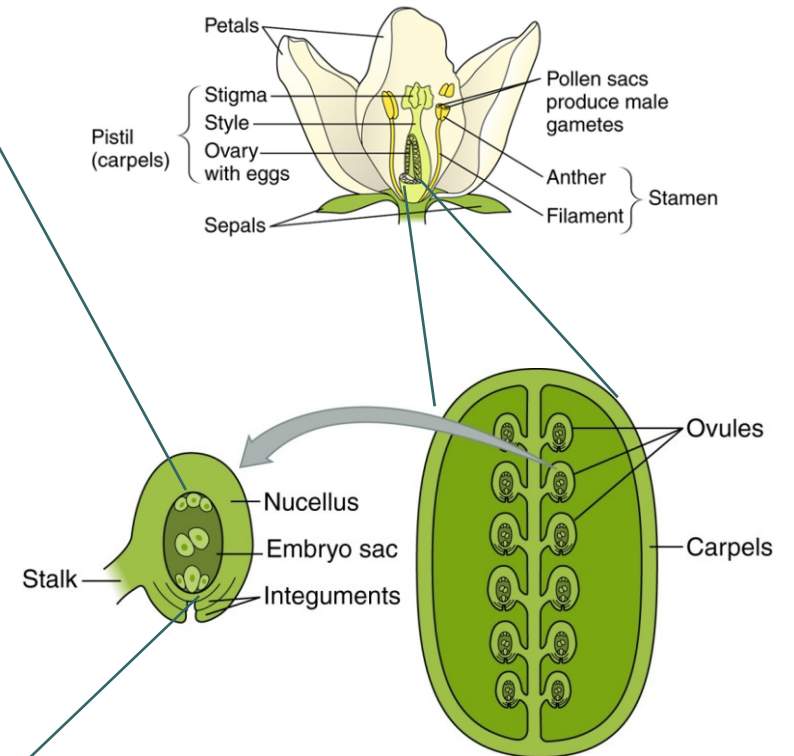
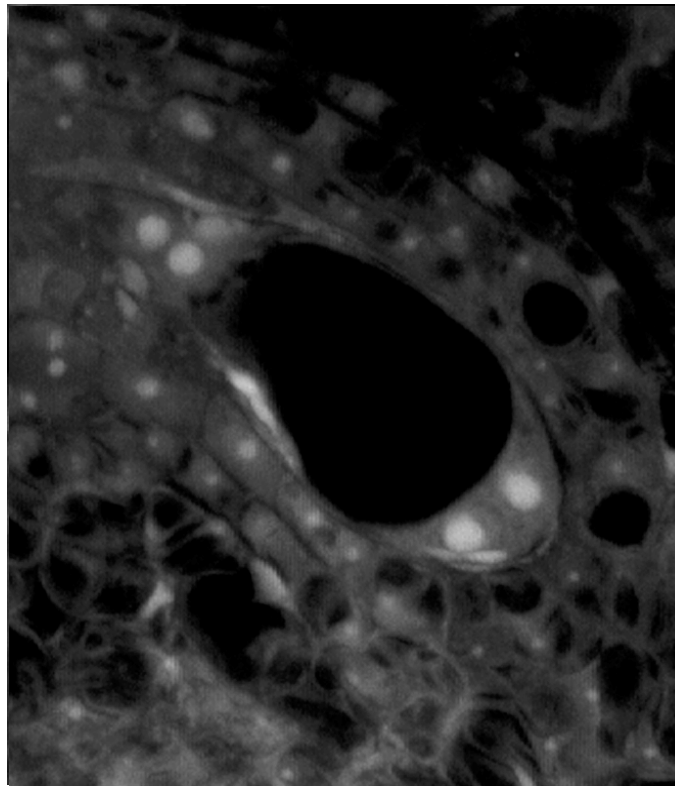


cki1-i specific primers

D. ♂ **CKI1/cki1-i** x ♀ wt

CKI1 and Megagametogenesis

FG 4



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

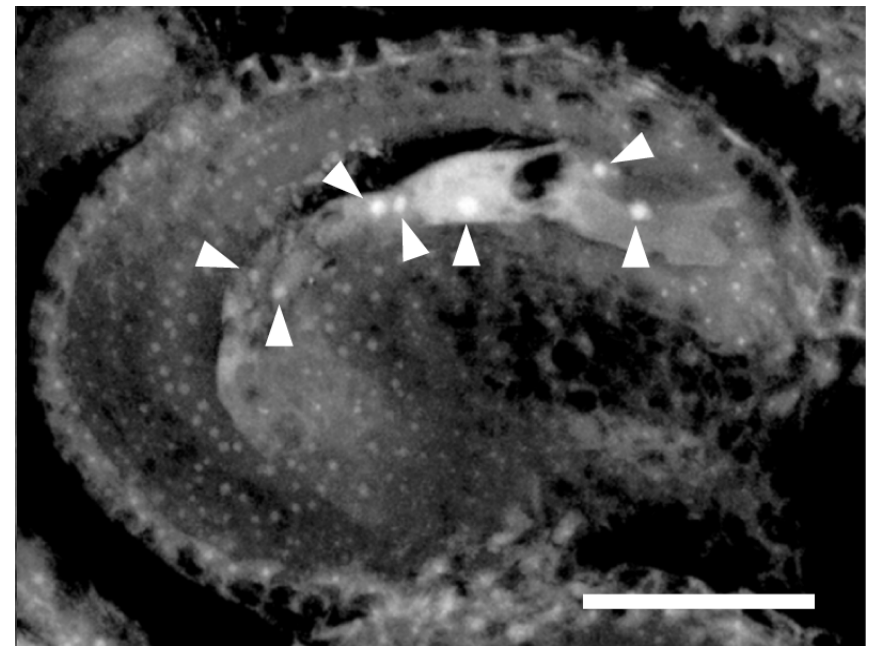
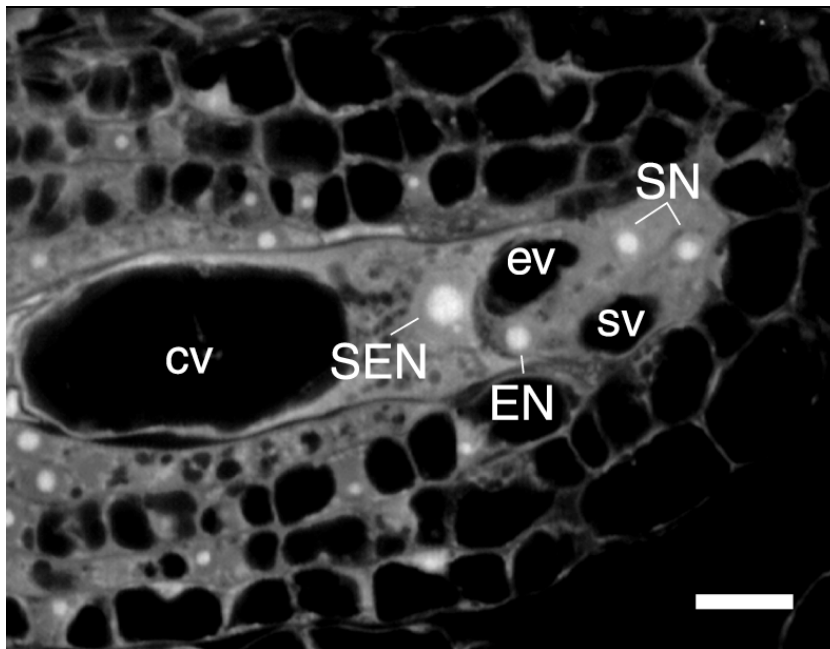
CKI1 and Megagametogenesis

CKI1

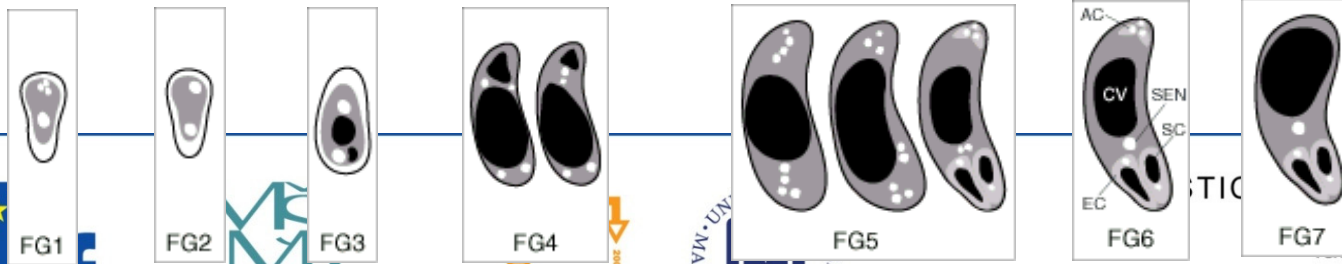
FG4 to FG5

cki1-i

28 HAE



Hejátko et al., *Mol Genet Genomics* (2003)



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání
pro konkurenceschopnost

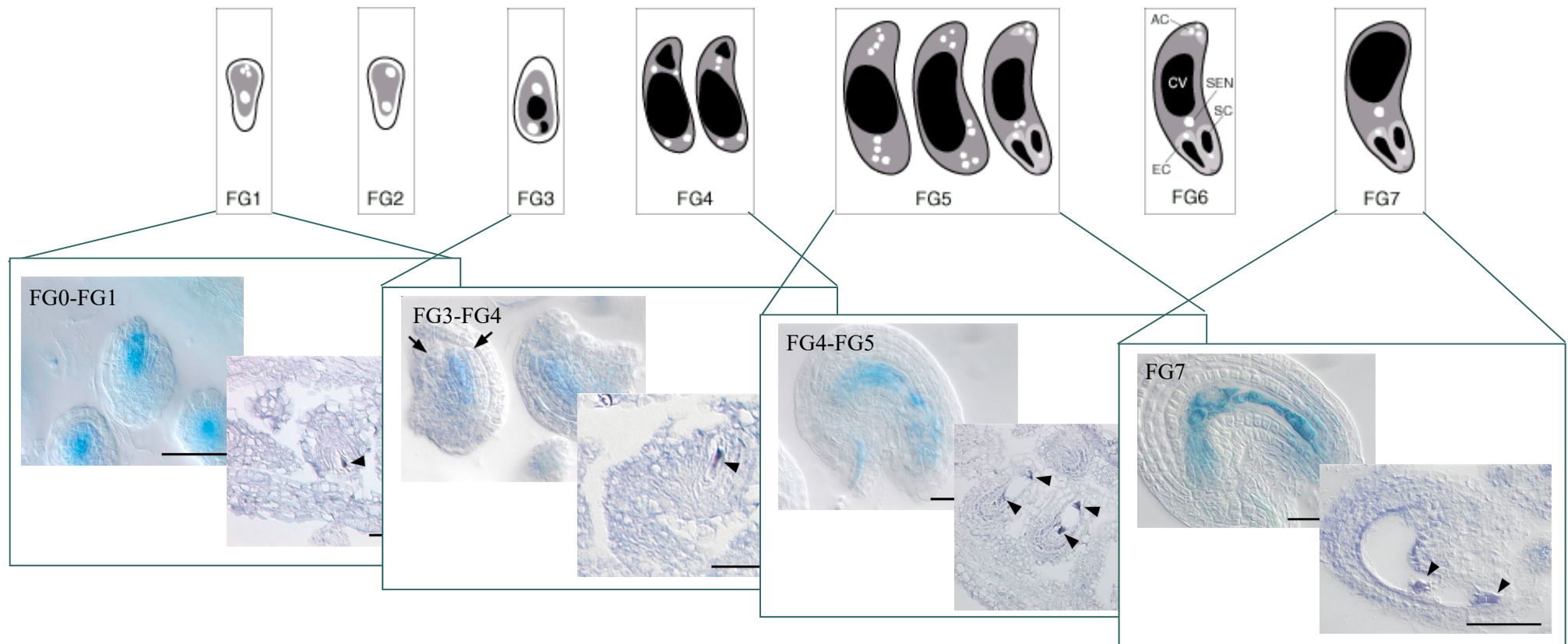


TIC ROZVOJE VZDĚLÁVÁNÍ
prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Forward and Reverse Genetics

- Principles of experimental identification of genes using forward and reverse genetics
 - Alteration of phenotype after mutagenesis
 - **Forward genetics**
 - Identification of insertional mutant and analysis of its phenotype
 - **Reverse genetics**
 - Analysis of expression of a particular gene and its spatiotemporal specificity

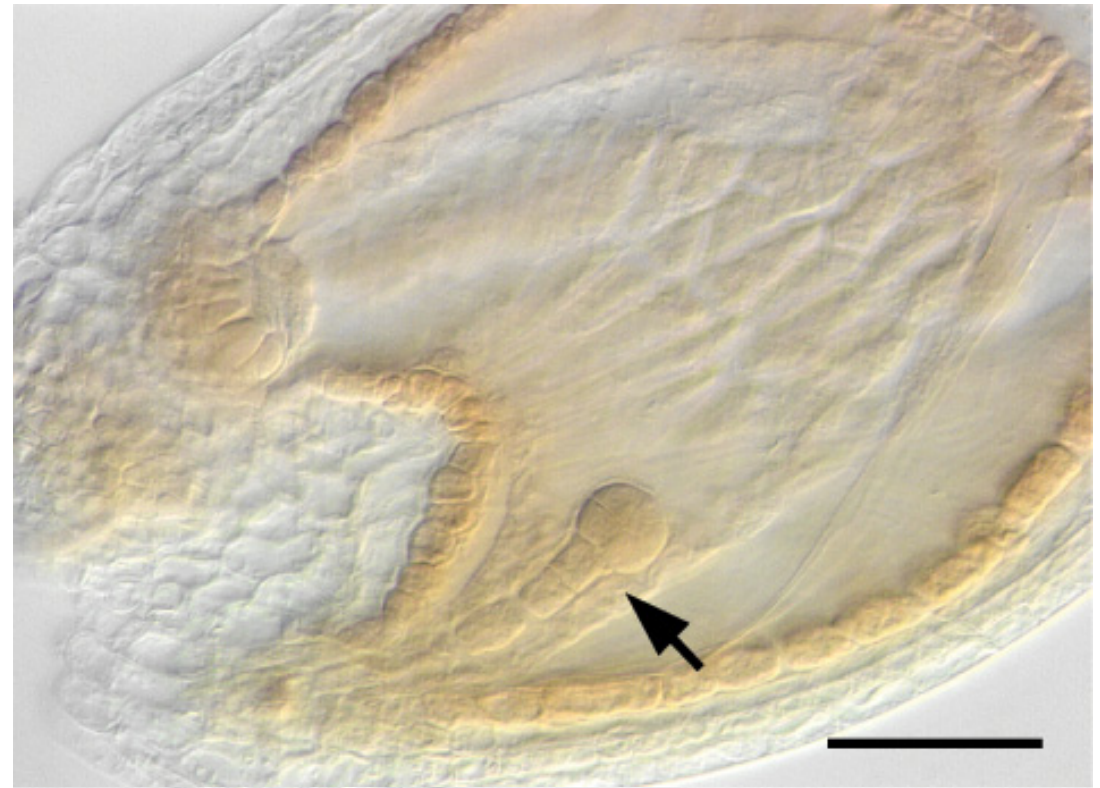
CKI1 is Expressed During Megagametogenesis



Paternal *CKI1* is Expressed in the *Arabidopsis* Sporophyte Early after Fertilization

♀ wt x ♂ Pro*CKI1*:*GUS*

28 HAP
(hours
after
pollination)



Hejátko et al., *Mol Genet Genomics* (2003)

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Outline

- Definition Of Genomics
- Forward vs Reverse Genetics
- Genes Structure and Identification
- Nucleic Acid Sequencing



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

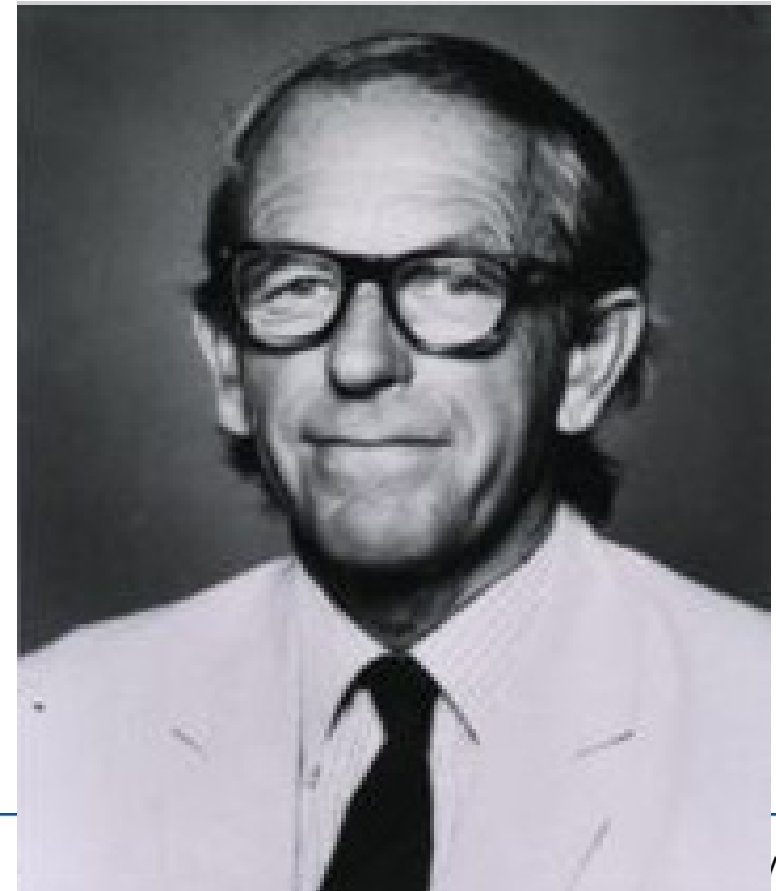
Sanger Sequencing

Frederick Sanger

1958 – Nobel prize – insuline structure

1975 - Dideoxy sequencing method

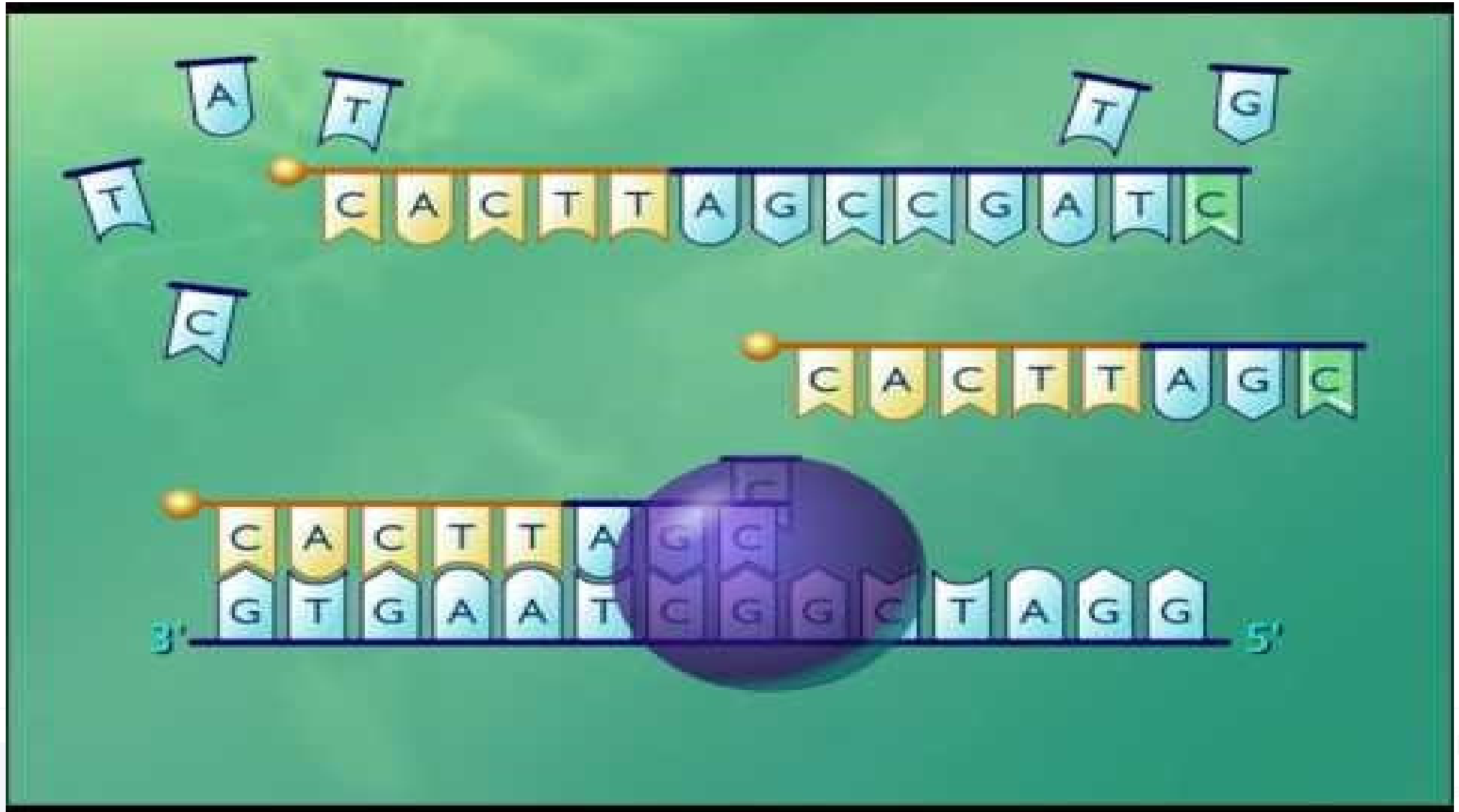
1980 – second Nobel prize for NA sequencing



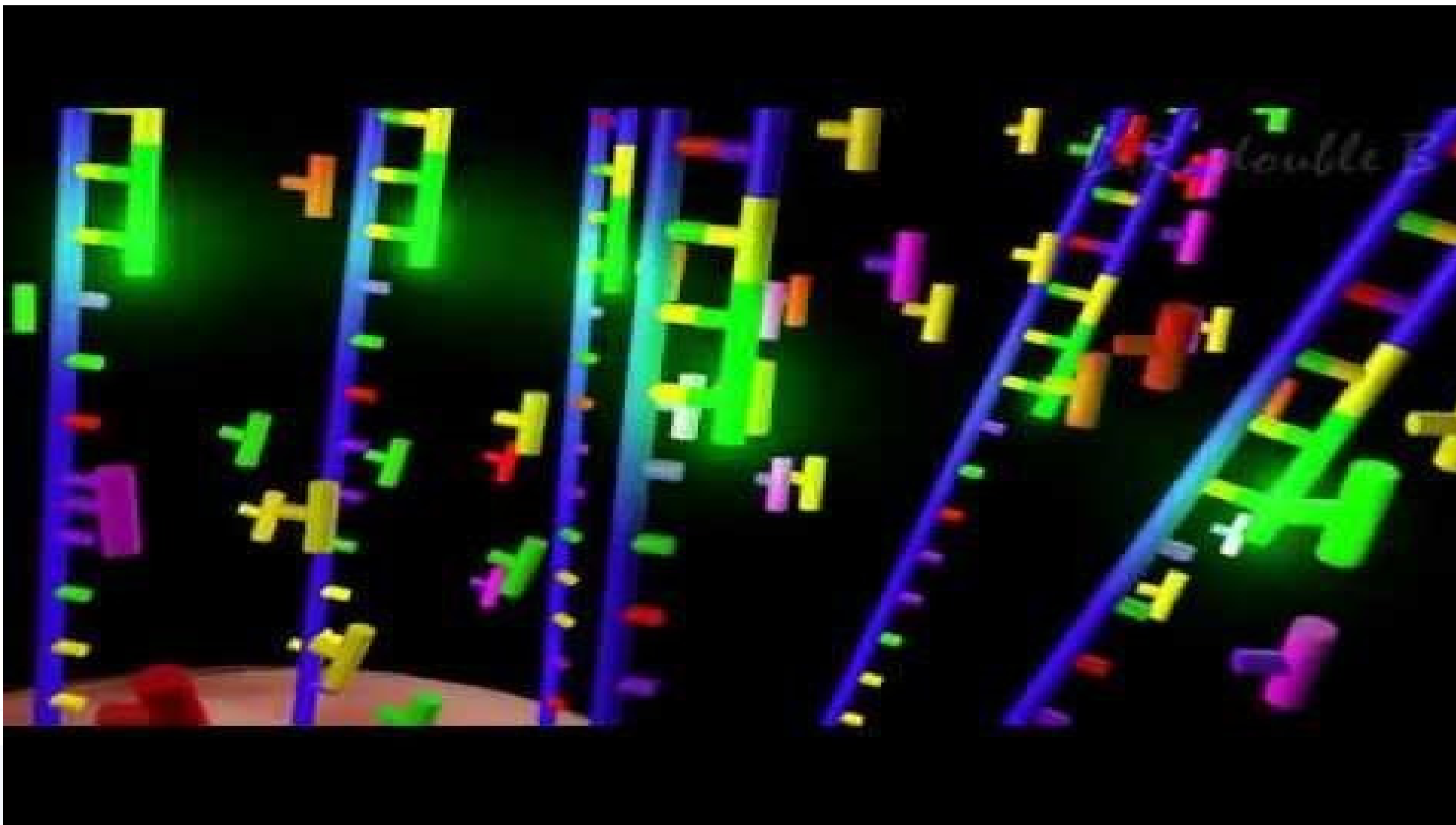
ÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Sanger Sequencing



NGS Sequencing



Outline

- Definition Of Genomics
- Forward vs Reverse Genetics
- Genes Structure and Identification
- Nucleic Acid Sequencing
- Analysis of Gene Expression



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Gene Expression Assays

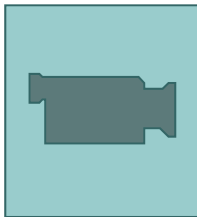
- **Methods of gene expression analysis**
 - **Quantitative analysis of gene expression**
 - DNA chips
 - Next generation transcriptional profiling
 - **Qualitative analysis of gene expression**
 - Preparation of **transcriptional fusion** of **promoter** of analysed gene with a **reporter gene**
 - 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

Gene Expression Assays

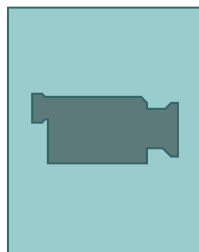
- Methods of gene expression analysis
 - Quantitative analysis of gene expression
 - DNA chips

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 commercially 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 oligonucleotide synthesis e.g. for the whole human genome (about 3,1 x 10⁹ bp) it 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	<i>E. coli</i> genes <i>bioB</i> , <i>bioC</i> , <i>bioD</i> . <i>B. subtilis</i> gene <i>lysA</i> . Phage P1 <i>cre</i> gene. <i>Arabidopsis</i> maintenance genes GAPDH, Ubiquitin, and Actin
Detection sensitivity	1:100,000*

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

DNA 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 threshold
- Finally comparing the experiment with the control or comparing different conditions with each other -> the result

Expression of 195M6T7 in response to chemical treatment

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

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

Gene

Experiment: Aluminum Stress

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

Slide Details

Slide (name ? : description)	External ID ?	Replicate (id ? : name)	Replicate type ?	Reverse replicate ?	Sample ?	Experimental variables	Label ?	Get Data ?
HoekengaS7 (*) : Aluminum Stress 1 [strong spatial bias]	AFGC: 7304	63: Aluminum Stress	technical		7304_Cy3.7305_Cy5	no treatment (pool of 3, 8, and 24 hours)	Cy3	Download
					7304_Cy5.7305_Cy3	Aluminum (50 5M AlCl3, pool of 3, 8, and 24 hours)	Cy5	
HoekengaSc Aluminum Stress 2 [strong spatial bias]	AFGC: 7305	64: Aluminum Stress	technical	63	7304_Cy5.7305_Cy3	Aluminum (50 5M AlCl3, pool of 3, 8, and 24 hours)	Cy3	Download
					7304_Cy3.7305_Cy5	no treatment (pool of 3, 8, and 24 hours)	Cy5	

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

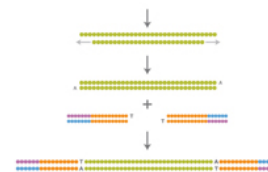
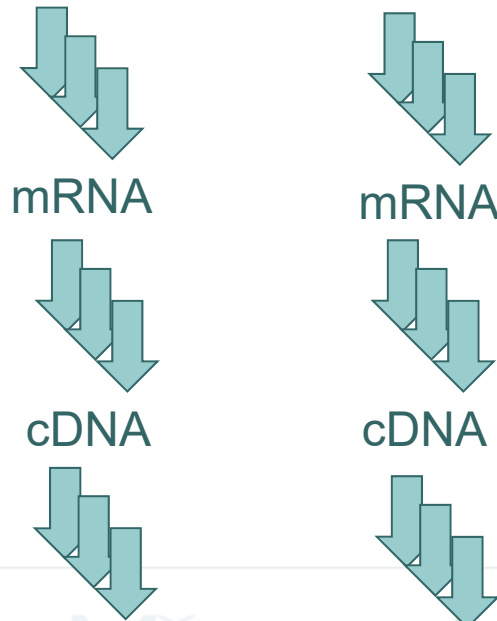
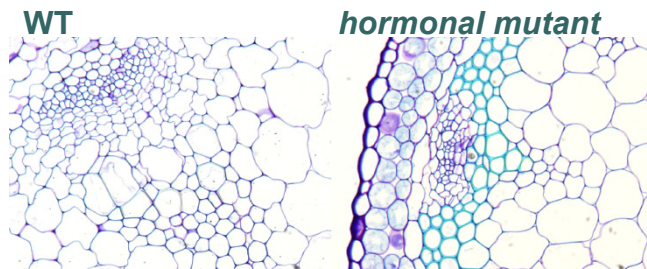
Che et al., 2002

Gene Expression Assays

- Methods of gene expression analysis
 - Quantitative analysis of gene expression
 - DNA chips
 - Next generation transcriptional profiling

Next Gen Transcriptional Profiling

- *Transcriptional profiling* via *RNA sequencing*



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

Results of –omics Studies vs Biologically Relevant Conclusions

- Transcriptional profiling yielded more than **7K differentially regulated genes...**

Ddii et al., unpublished

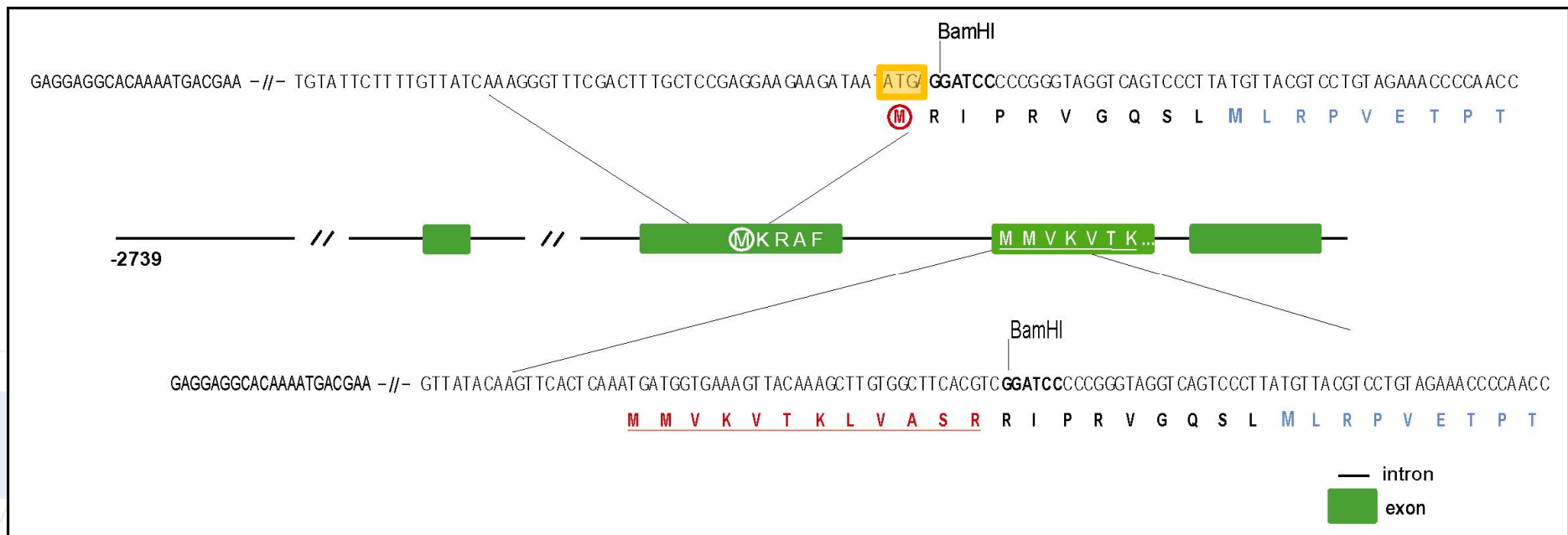
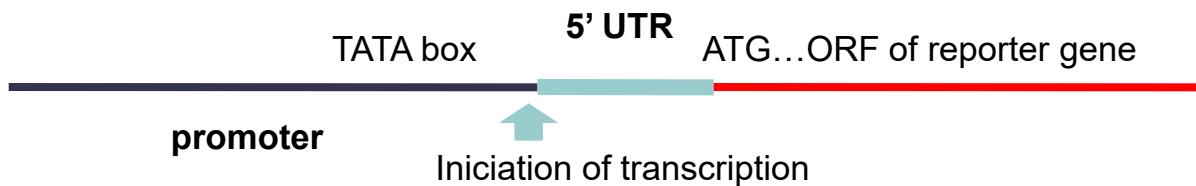
gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1,1804	1.79769e+308	1.79769e+308	6.88885e-05	0,00039180	1 yes
HRS1	1:4556891-4558708	WT	MT	OK	0	0,696583	1.79769e+308	1.79769e+308	6.61994e-06	4.67708e-05	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	0,00496816	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

Gene Expression Assays

- Methods of gene expression analysis
 - Quantitative analysis of gene expression
 - DNA chips
 - Next generation transcriptional profiling
 - Qualitative analysis of gene expression
 - Preparation of transcriptional fusion of promoter of analysed gene with a reporter gene

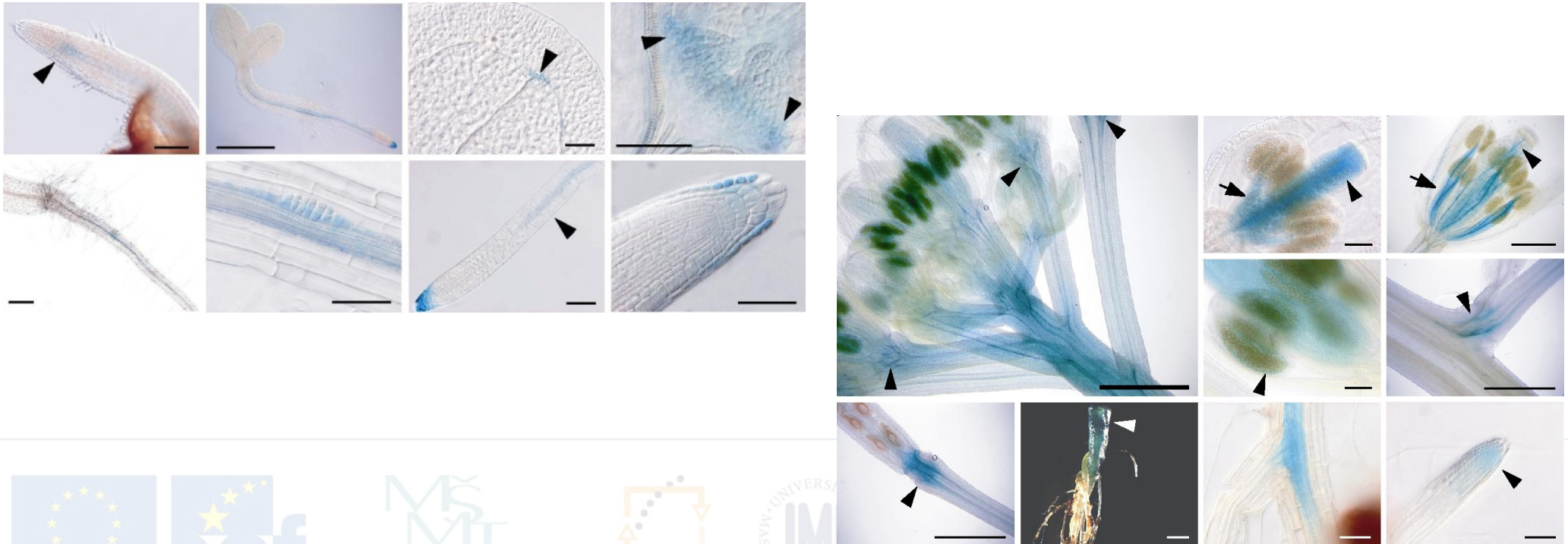
Transcriptional Fusion

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

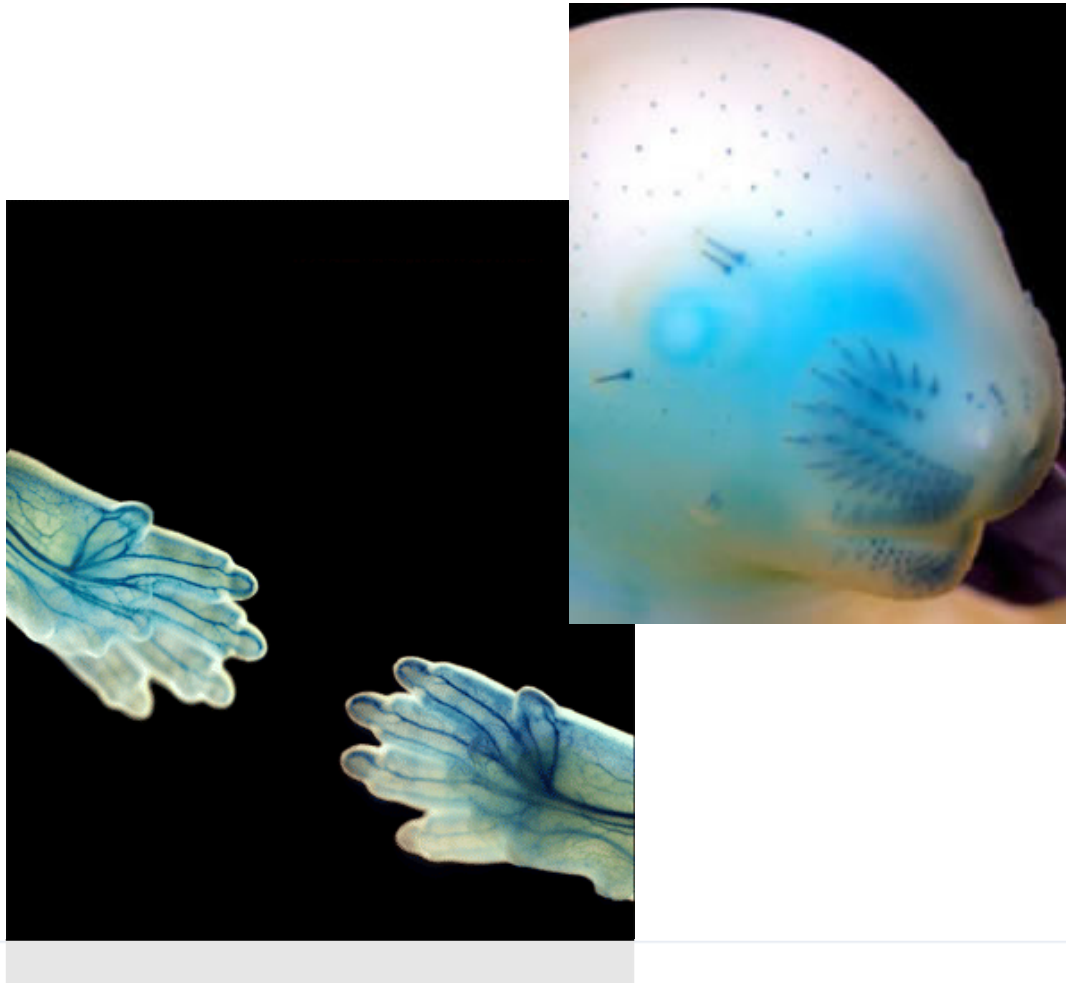


Transcriptional Fusion

- 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



GUS Reporter in Mouse Embryos

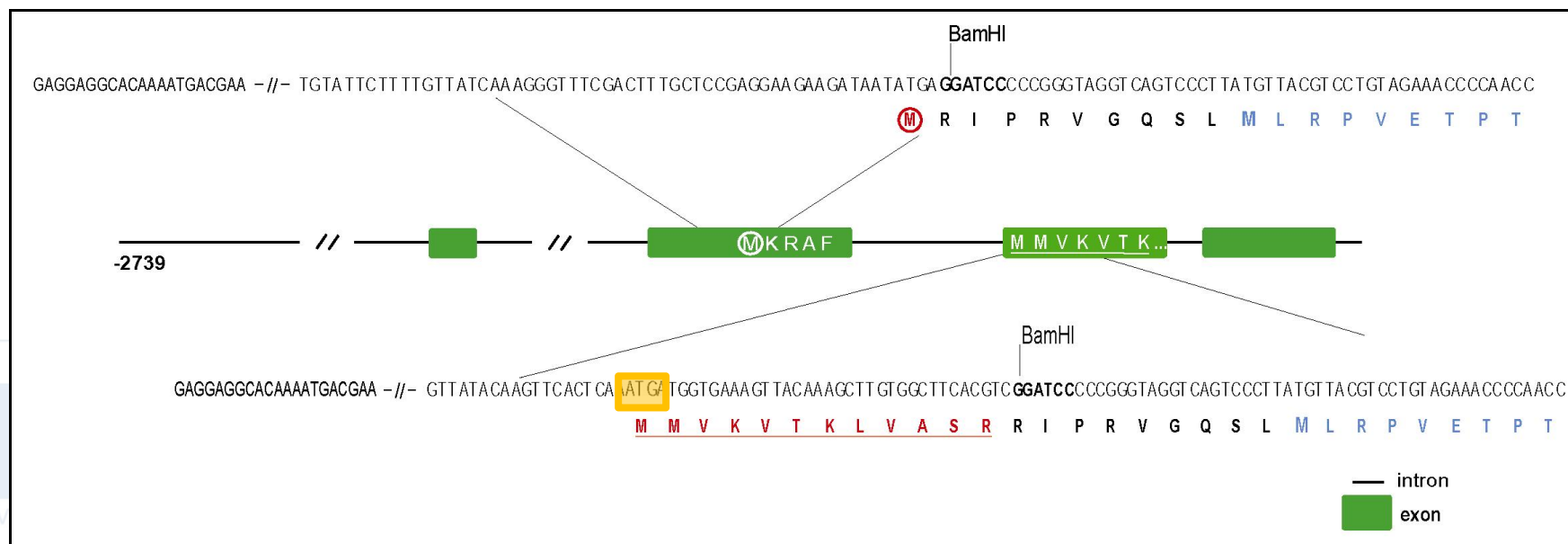
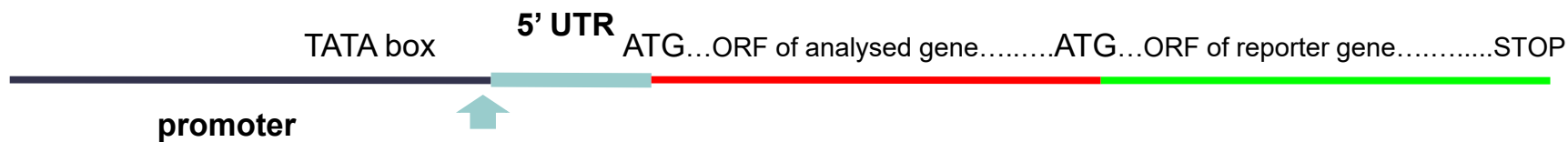


Gene Expression Assays

- **Methods of gene expression analysis**
 - **Quantitative analysis of gene expression**
 - DNA chips
 - Next generation transcriptional profiling
 - **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**

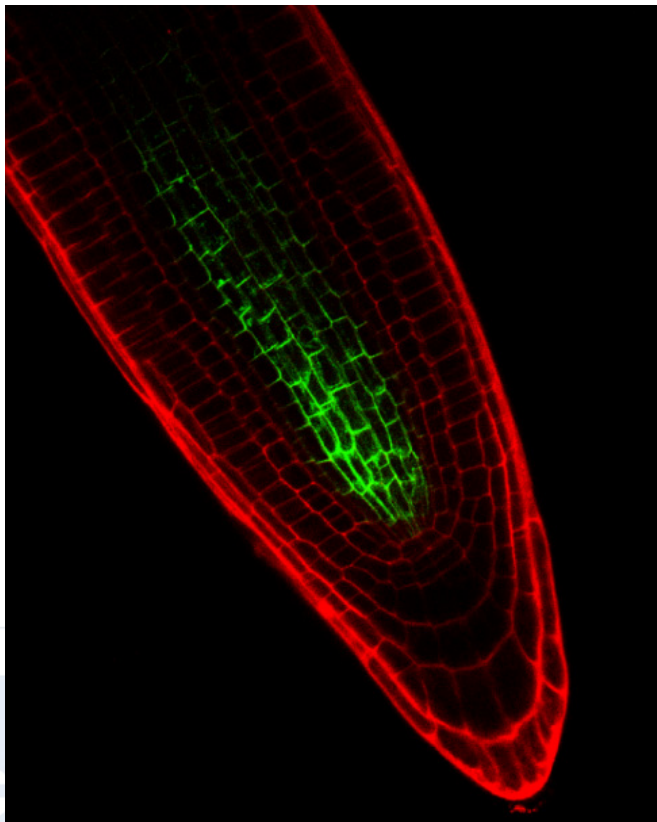
Translational Fusion

- 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)

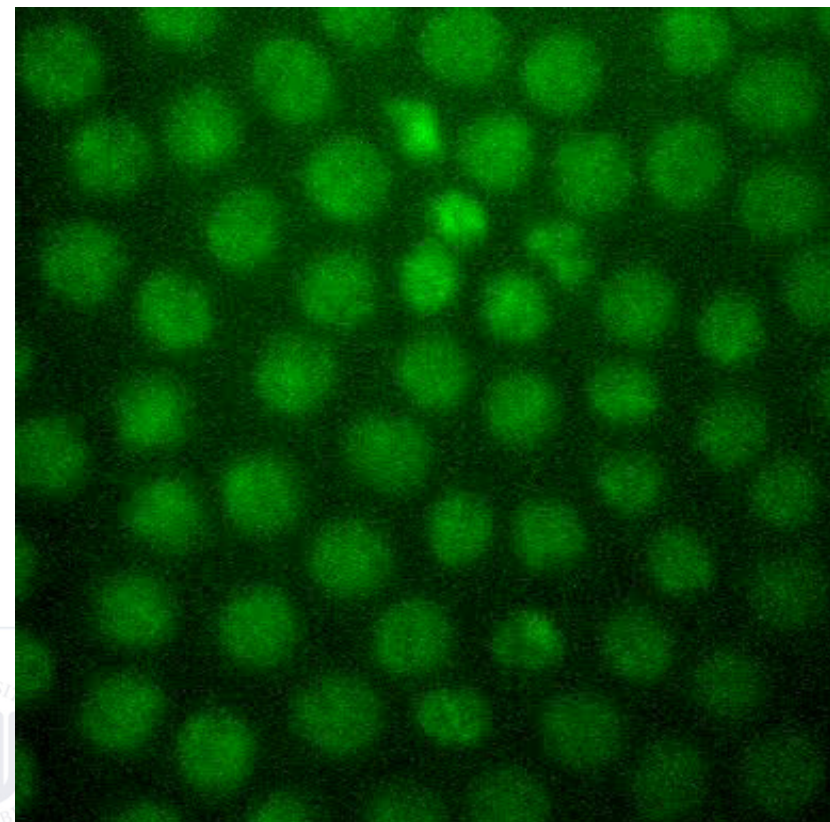


Translational Fusion

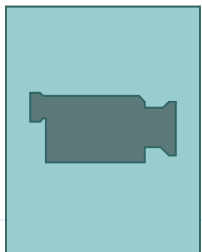
- Preparation of transgenic organisms carrying the recombinant DNA and their histological analysis
- Compared to transcriptional fusion, translation fusion allows analysis of intercellular localization of gene product (protein) or its dynamics



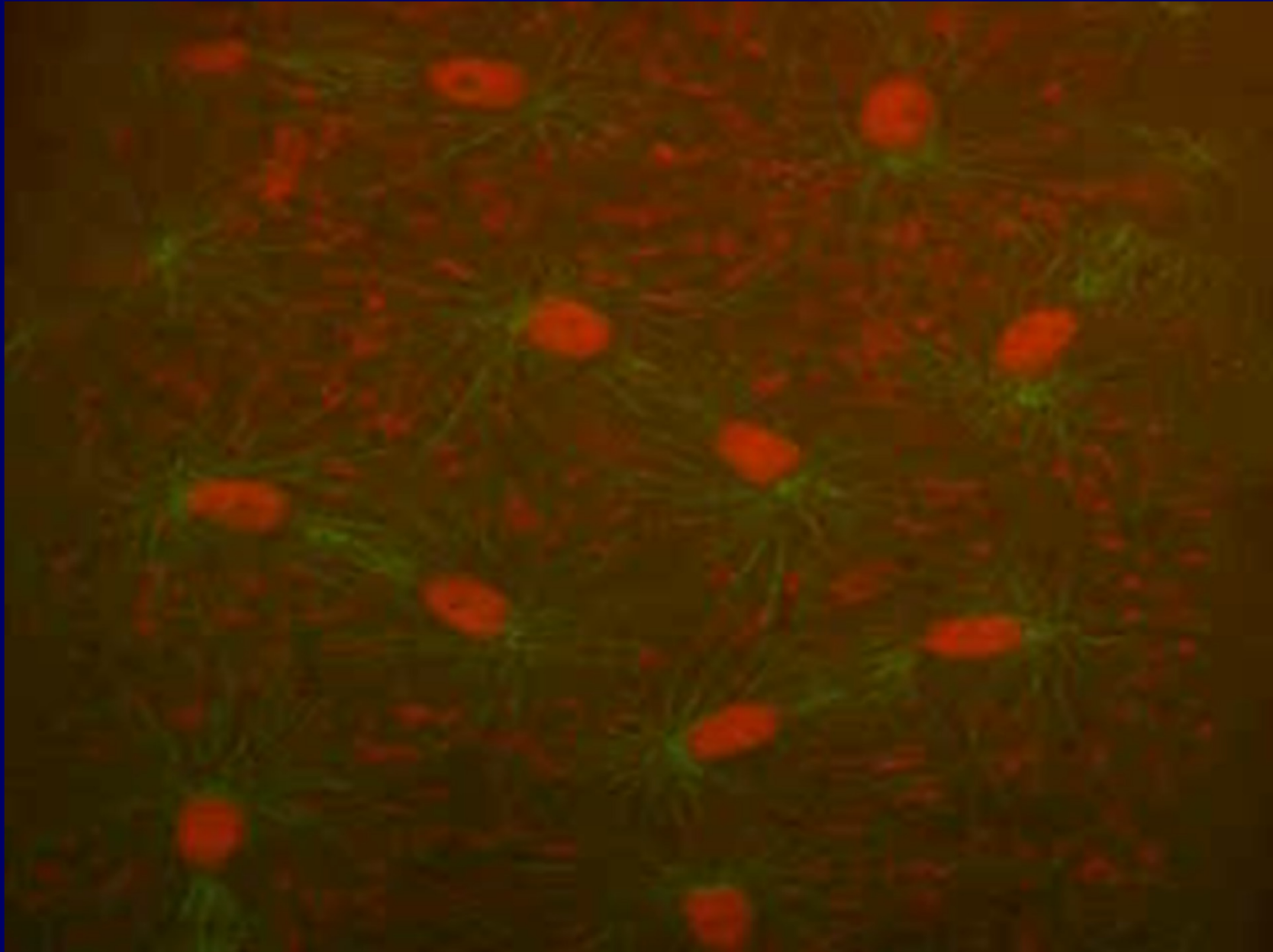
PIN1-GFP in *Arabidopsis*



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



Translational Fusion

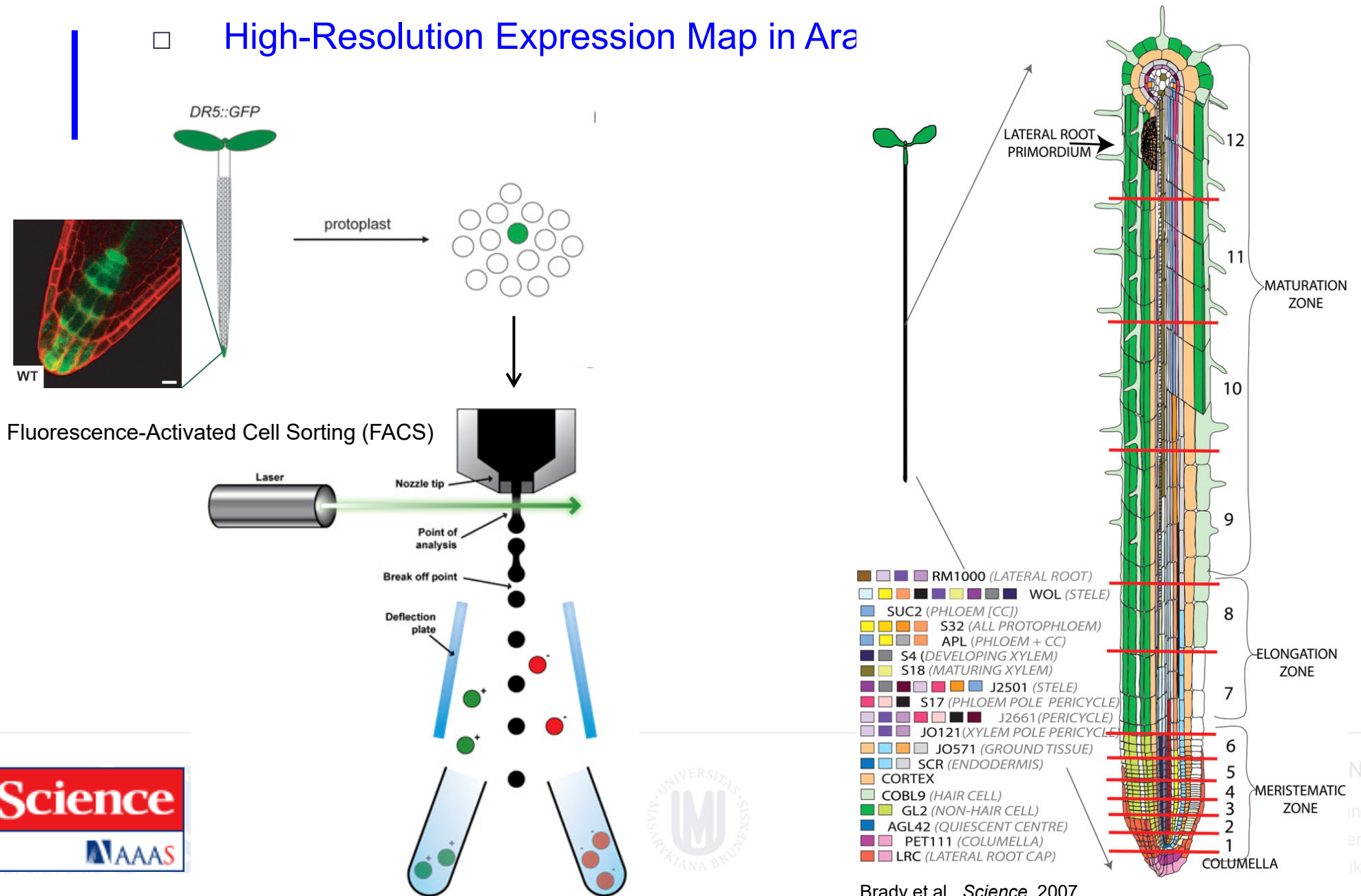


Gene Expression Assays

- Methods of gene expression analysis
 - Quantitative analysis of gene expression
 - DNA chips
 - Next generation transcriptional profiling
 - 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
 - Tissue- and cell-specific gene expression analysis

Expression Maps - RNA

High-Resolution Expression Map in Ara



BAR ePlant

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

BAR ePlant

Enter a gene name

Expression Angler

Mutant Phenotype Selector

RSVP OFF

1 gene / gene product currently loaded

- Home
- Info
- Layers
- Plant eFP
- Tissue & Experiment eFP
- Cell eFP
- Chromosome Viewer
- Interaction Viewer
- Molecule Viewer
- Sequence Browser

Welcome Screen

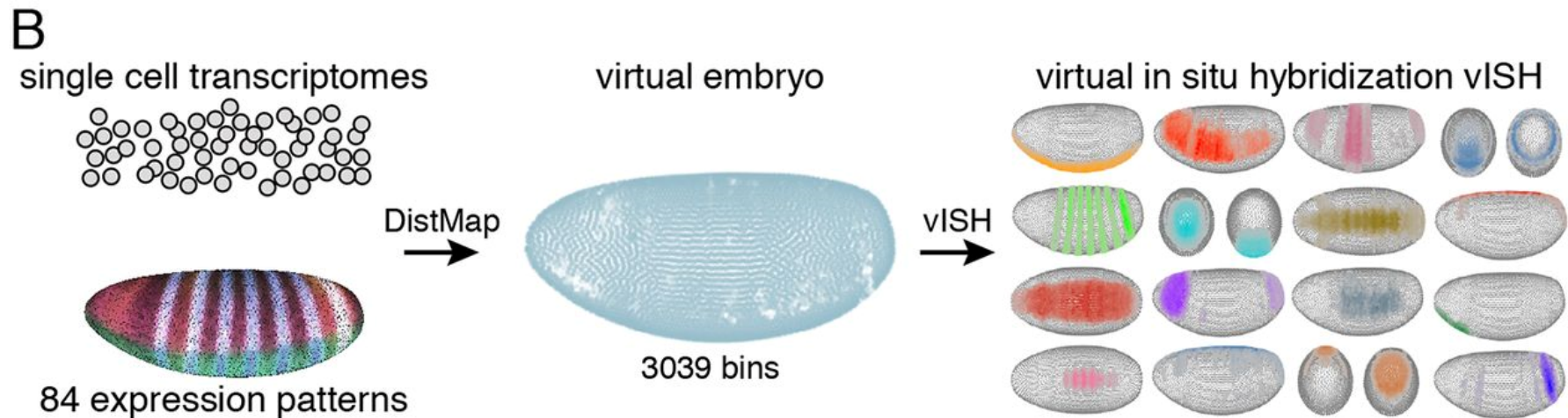
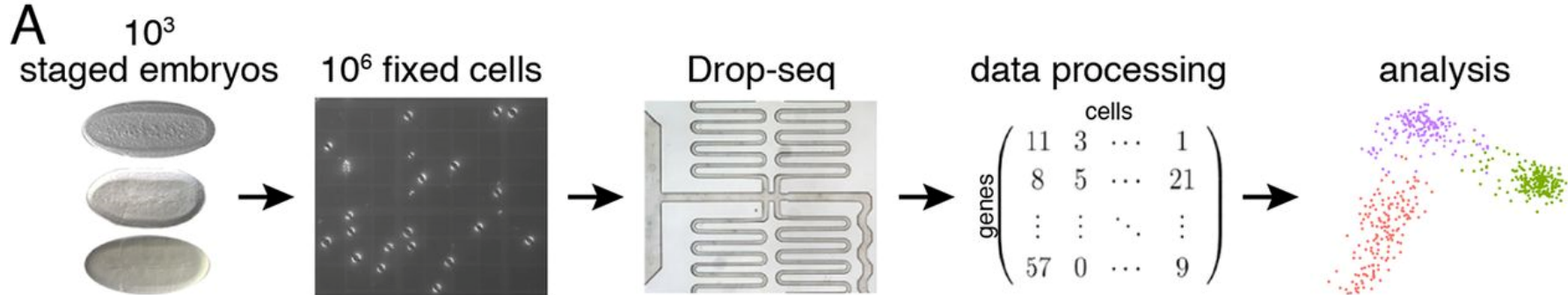
Data visualization tools for multiple levels of plant data.

- World eFP
- Plant eFP
- Tissue & Experiment eFP
- Cell eFP
- Chromosome Viewer
- Interaction Viewer
- Molecule Viewer
- Sequence Browser

----- influences
----- leads to
----- "central dogma" of molecular biology

Expression Maps - RNA

A High-Resolution Expression Map in Drosophilla



Nikos Karaïskos et al. Science 2017;science.aan3235

Drosophila Virtual Expression eXplorer

<https://shiny.mdc-berlin.de/DVEX/>

DVEX t-SNE vISH vISHs vISH - D. vir. Gradients Archetypes Download About



Loading data ... (DVEX is currently better supported on Linux / Mac OS X)

Data loaded for 8924 genes.

Drosophila Virtual Expression eXplorer

DVEX is an online resource tool which offers an easy way to explore the transcriptome of the stage 6 *Drosophila* embryo at the single cell level. It is part of the collaboration between the Rajewsky and Zinzen labs in the Berlin Institute of Medical Systems Biology of the Max Delbrueck Center in Berlin. DVEX accompanies the following publication:

Science 358 (6360), 194-199

The *Drosophila* Embryo at Single Cell Transcriptome Resolution

Nikos Karaiskos^a, Philipp Wahle^b, Jonathan Alles^a, Anastasiya Boltengagen^a, Salah Ayoub^a, Claudia Kipar^b, Christine Kocks^a, Nikolaus Rajewsky^a, Robert Zinzen^b,

^aSystems biology of gene regulatory elements, BIMSB, MDC

^bSystems biology of neural tissue differentiation, BIMSB, MDC

Correspondence regarding the publication: Nikolaus Rajewsky and Robert P. Zinzen.

DVEX is created and maintained by Nikos Karaiskos. Contact the author for questions, or troubleshooting.

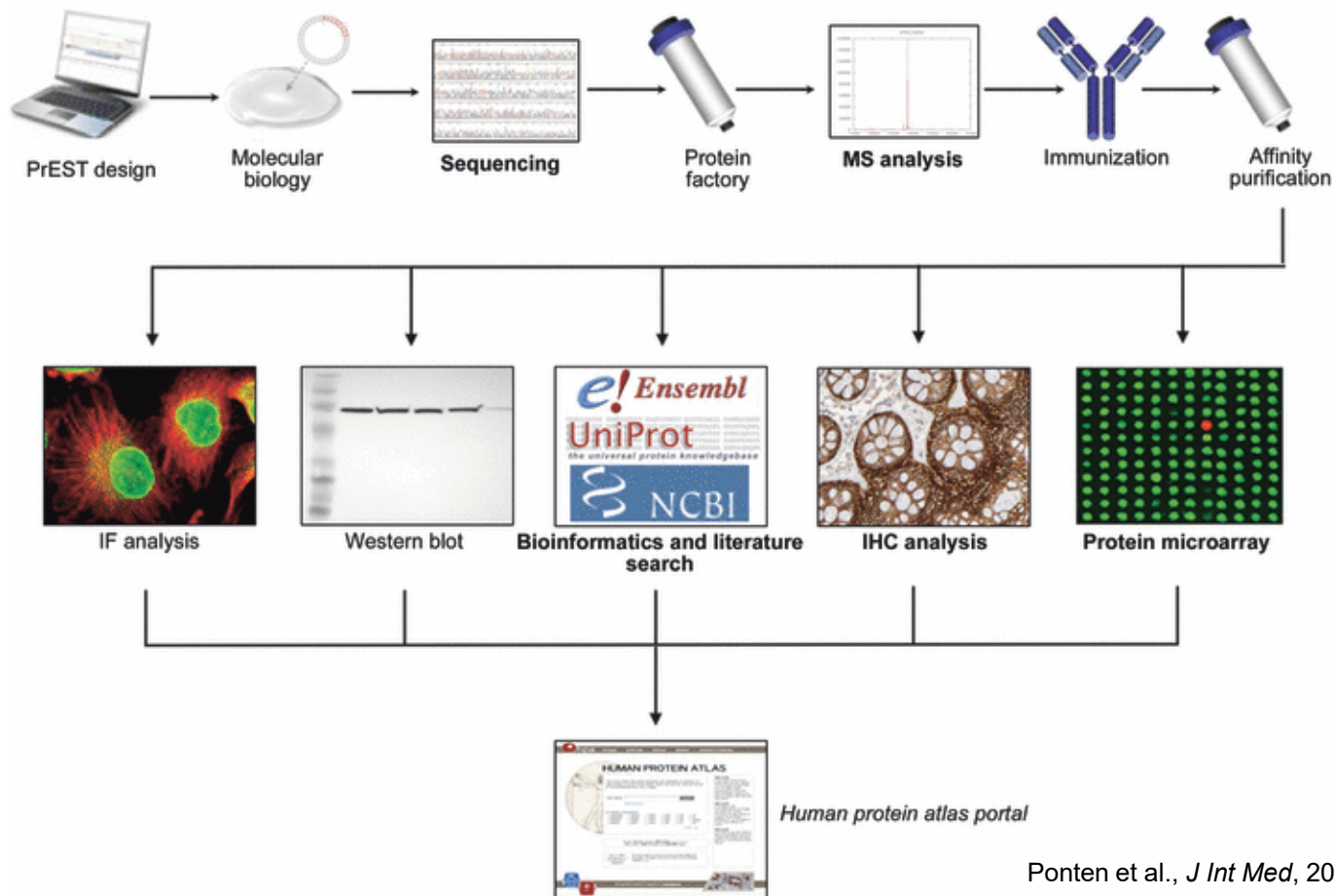


DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Expression Maps - Proteins

□ Human Protein Atlas



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

Expression Maps - Proteins

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

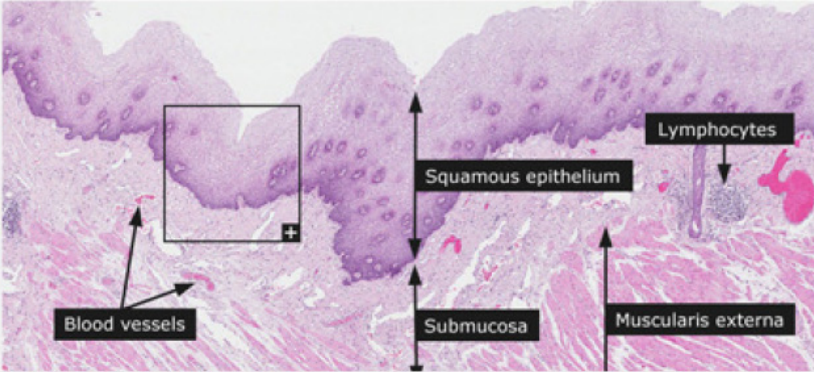
THE HUMAN PROTEIN ATLAS

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](#)



News

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

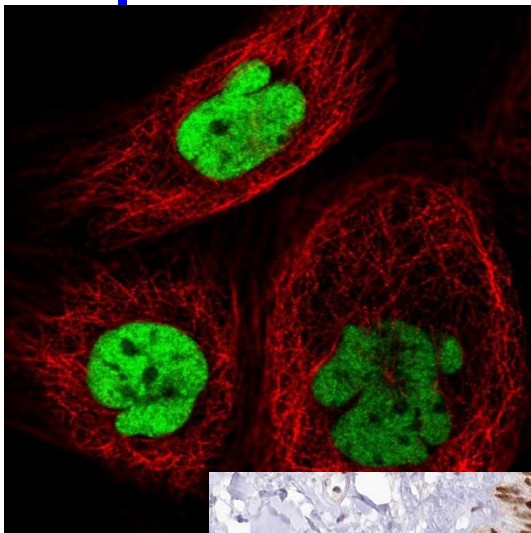
Version: **11.0**
Atlas updated: 2013-03-11
[release history](#)

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

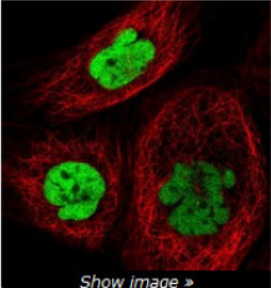
dictionary: histology of esophagus


Expression Maps - Proteins

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



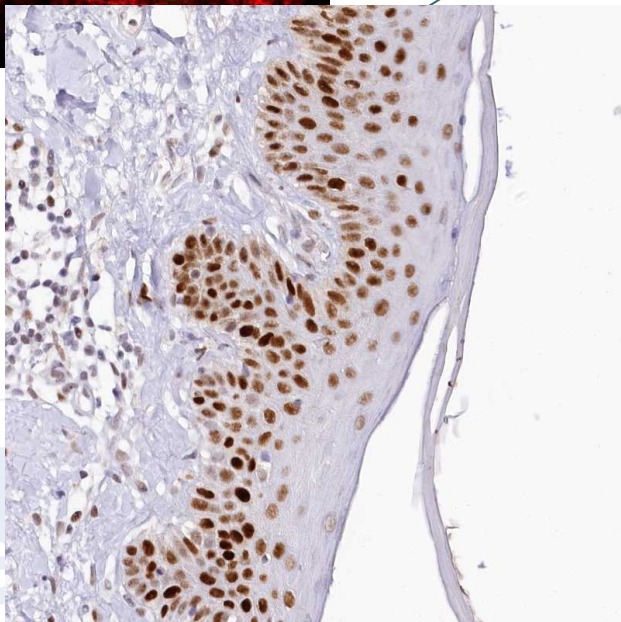
SUBCELLULAR LOCATION SUMMARY ? »



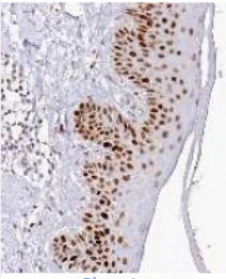
Main location(s) Nucleus but not nucleoli
Additional location(s)
Staining summary Localized to the nucleus but excluded from the nucleoli.
Reliability (APE)  High
Antibodies in assay CAB039238, CAB039239


[Show image »](#)













MORE SUBCELL DATA



NORMAL TISSUE & ORGAN SUMMARY ? »



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

Organ	No of cell types	Protein expression
CNS (brain)	11	
Hematopoietic (blood)	8	
Liver and pancreas	5	
Digestive (GI-tract)	13	
Respiratory (lung)	4	
Cardiovascular	1	
Female tissues	13	
Placenta	2	
Male tissues	5	
Urinary tract (kidney)	3	
Skin and soft tissues	14	
Endocrine tissues	3	

[Show image »](#)

MORE TISSUE DATA

Summary

- Definition Of Genomics
- Forward vs Reverse Genetics
- Genes Structure and Identification
- Nucleic Acid Sequencing
- Analysis of Gene Expression

Discussion



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky