

C6215 Advanced Biochemistry and its Methods

Lesson 1

Introduction into Genomics

Jan Hejátko

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

a

Národní centrum pro výzkum biomolekul,
Přírodovědecká fakulta,

Masarykova univerzita, Brno
hejatko@sci.muni.cz, www.ceitec.eu



M U N I
S C I

Outline

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

Outline

- Definition Of Genomics

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**

Forward Genetics



3

:



?

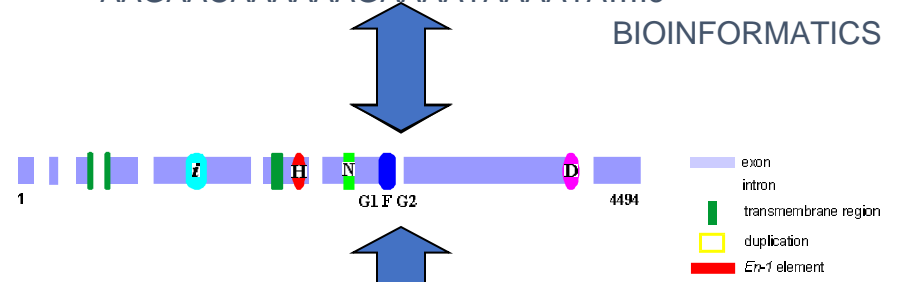


1

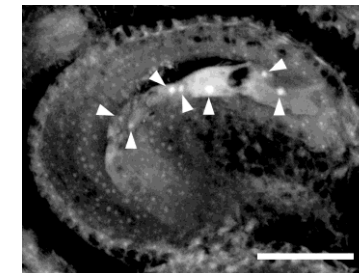
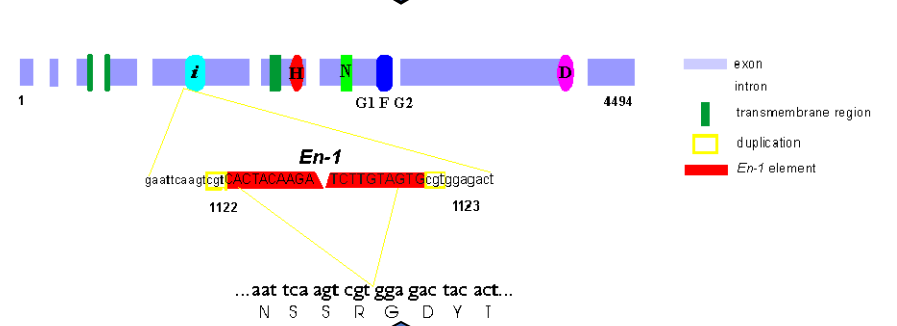
Reverse Genetics

5'TTATATATATATATATTTAAAAAATAAAATAA
AAGAACAAAAAGAAAAATAAATA....3'

BIOINFORMATICS



FUNCTIONAL GENOMICS

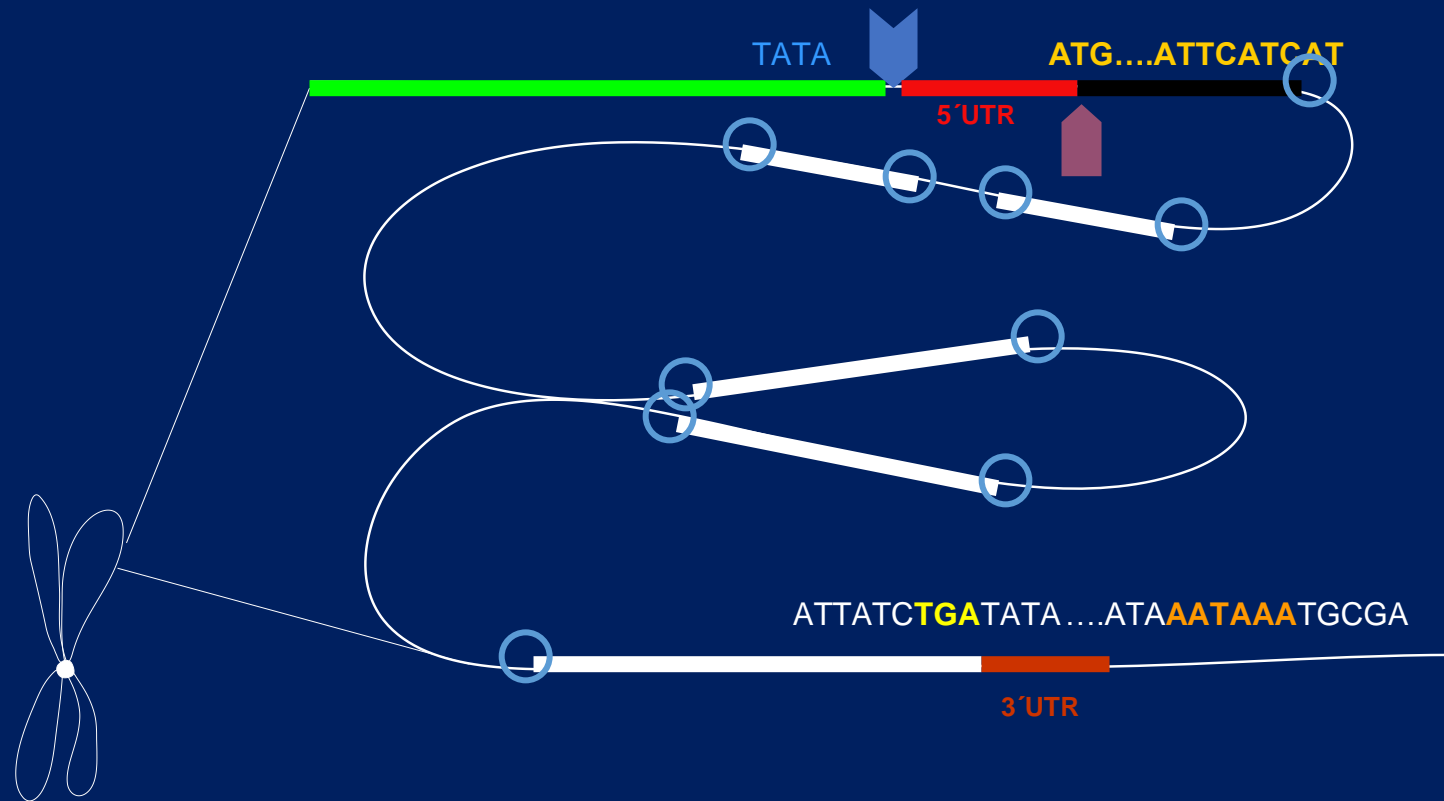


Outline

- Definition Of Genomics
- Forward vs Reverse Genetics
- **Gene Structure and Identification**

Gene 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
- 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

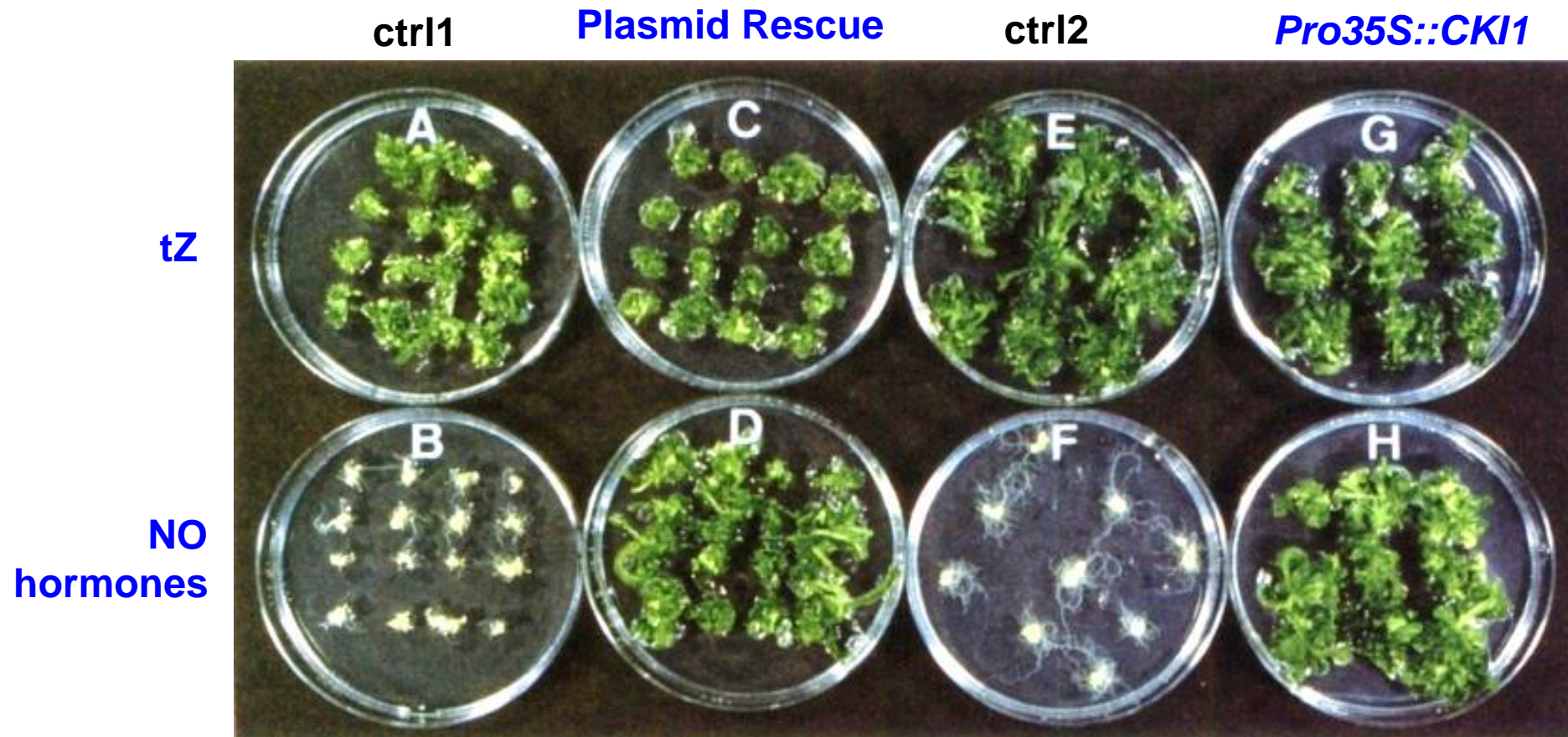
- Principles of experimental identification of genes using forward and reverse genetics
 - Alteration of phenotype after mutagenesis
 - **Forward genetics**
 - Identification of sequence-specific mutant and analysis of its phenotype
 - **Reverse genetics**
 - Analysis of expression of a particular gene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

Experimental Gene Identification

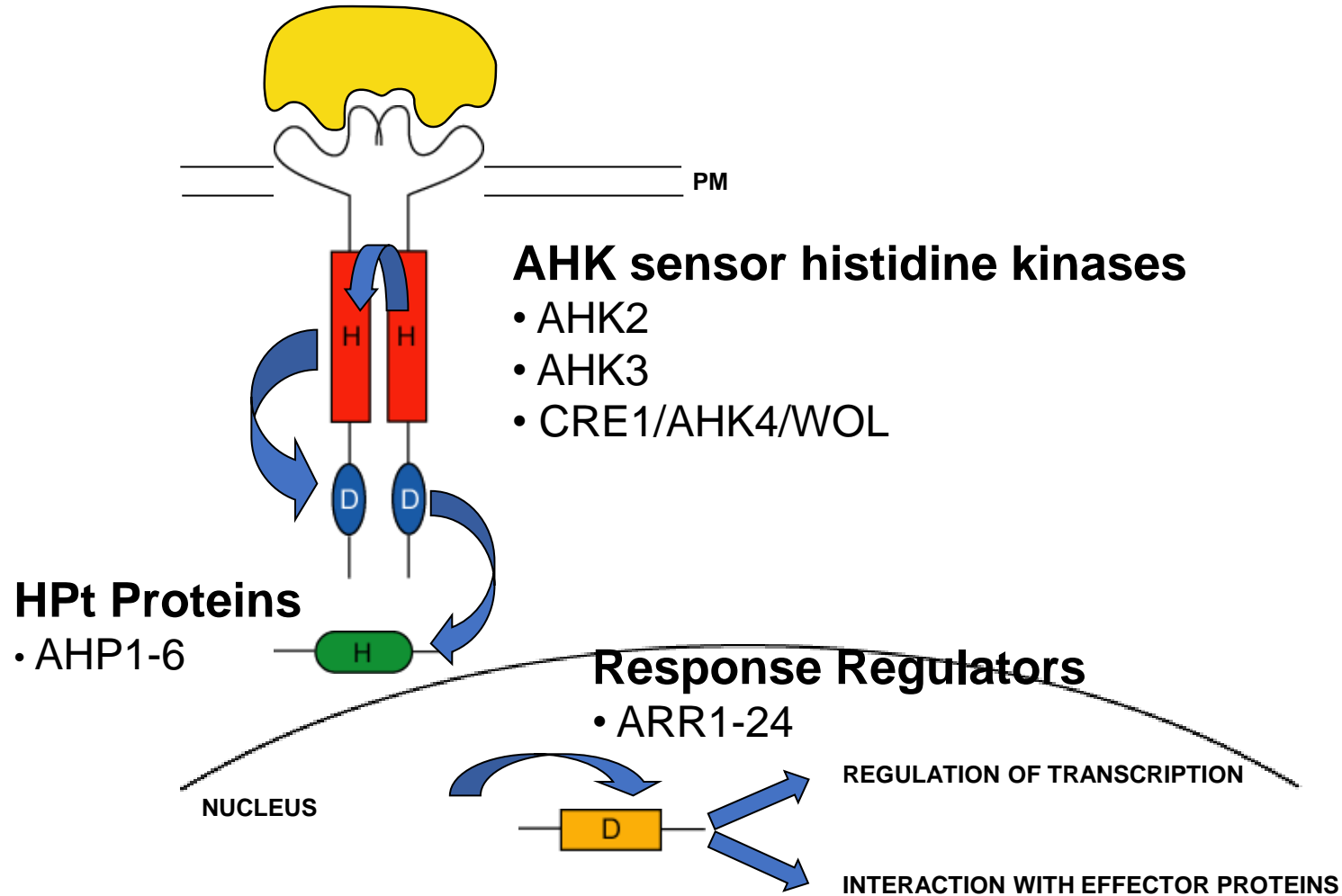
- Principles of experimental identification of genes using forward and reverse genetics
 - Alteration of phenotype after mutagenesis
 - **Forward genetics**

Identification of *CKI1* via Activation Mutagenesis

- *CKI1* overexpression mimics cytokinin response



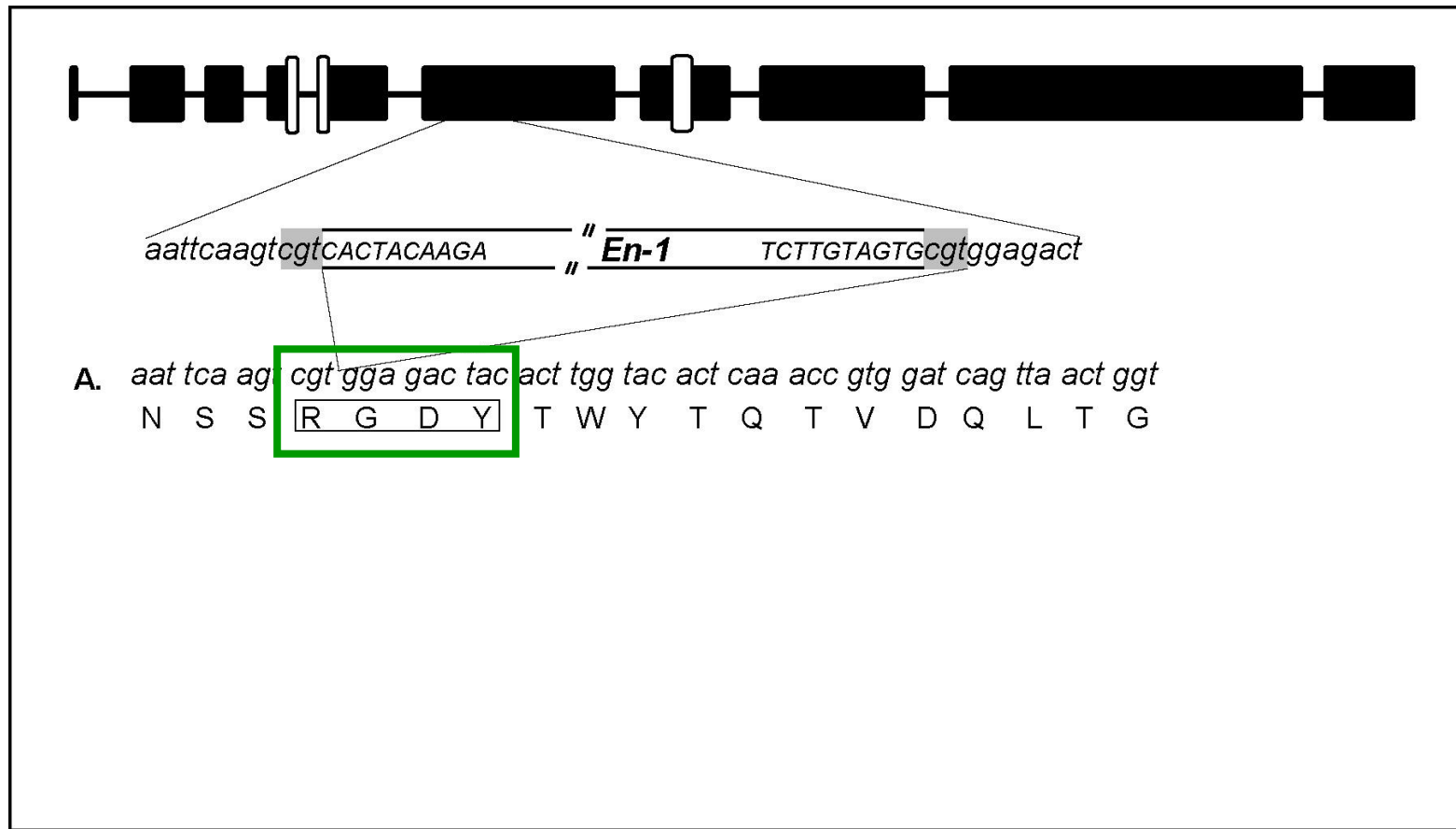
Signal Transduction via MSP



Reverse Genetics

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

Identification of insertional *cki1* mutant

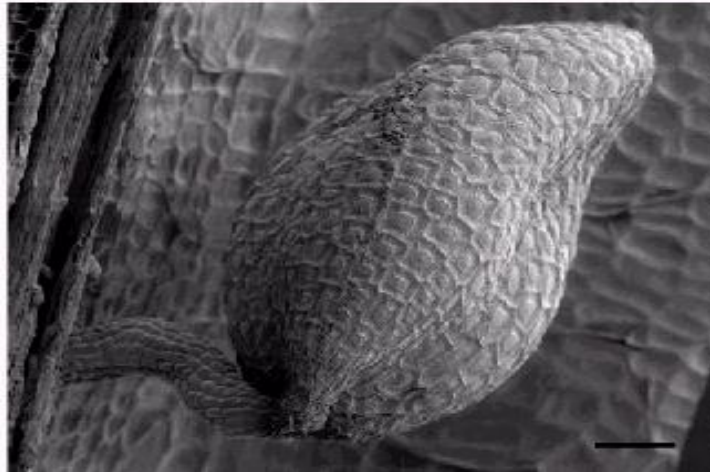
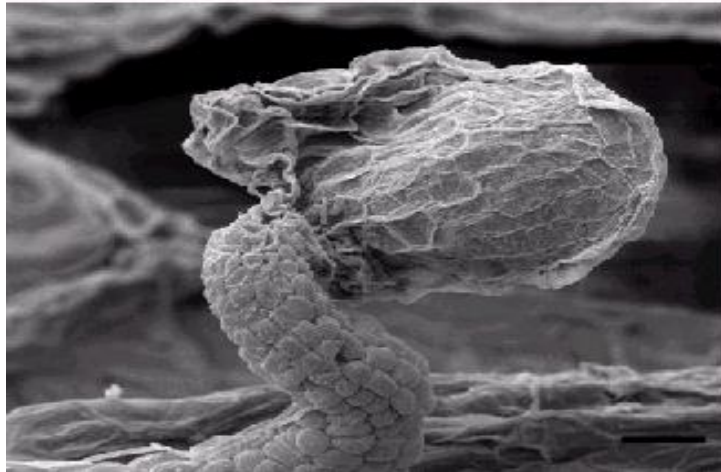


CKI1 Regulates Female Gametophyte Development

CKI1/cki1-i



CKI1/CKI1





cki1-i reveals non-Mendelian inheritance

P *CKI1/cki1-i*

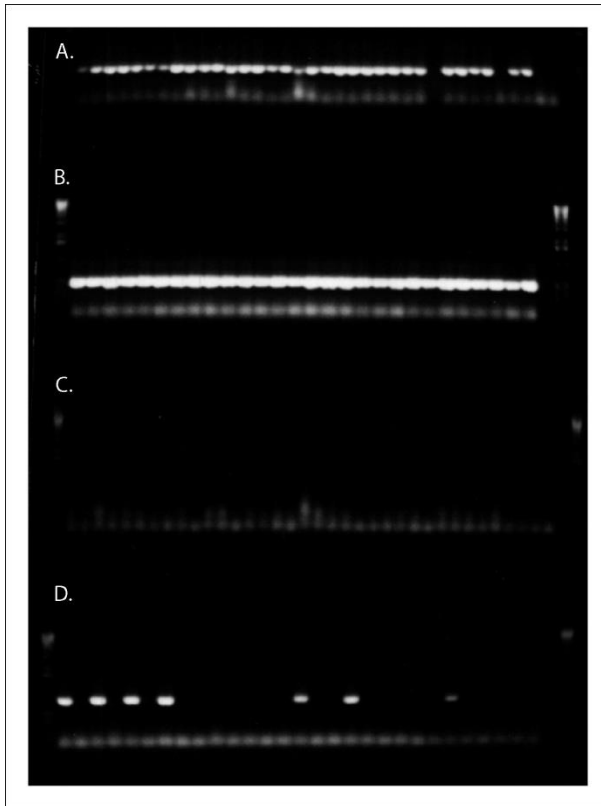
F1 Anticipated: 1 *CKI1* : 2 *CKI1/cki1-i* : 1 *cki1-i*

Observed: 1 *CKI1* : 1 *CKI1/cki1-i*

 	<i>CKI1</i>	<i>cki1-i</i>
<i>CKI1</i>	<i>CKI1/CKI1</i>	<i>CKI1/cki1-i</i>
<i>cki1-i</i>	<i>CKI1/cki1-i</i>	<i>cki1-i/cki1-i</i>

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*

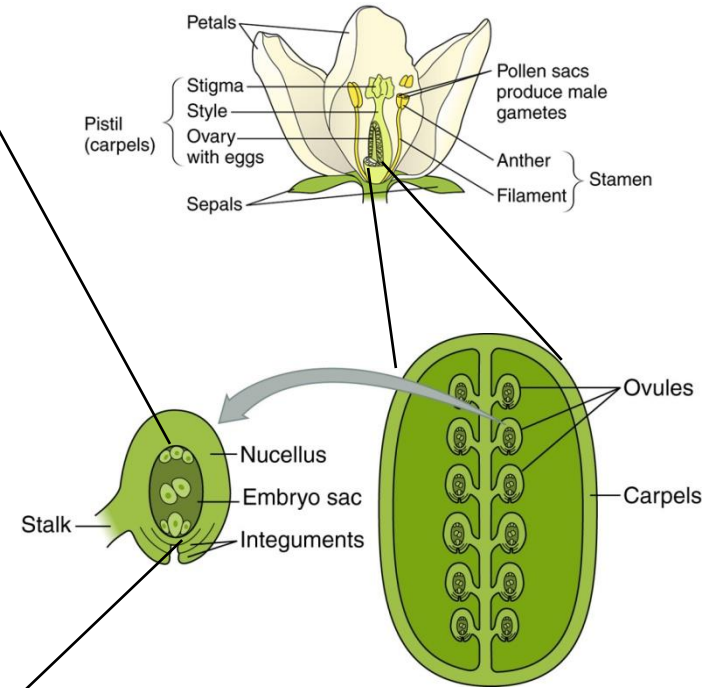
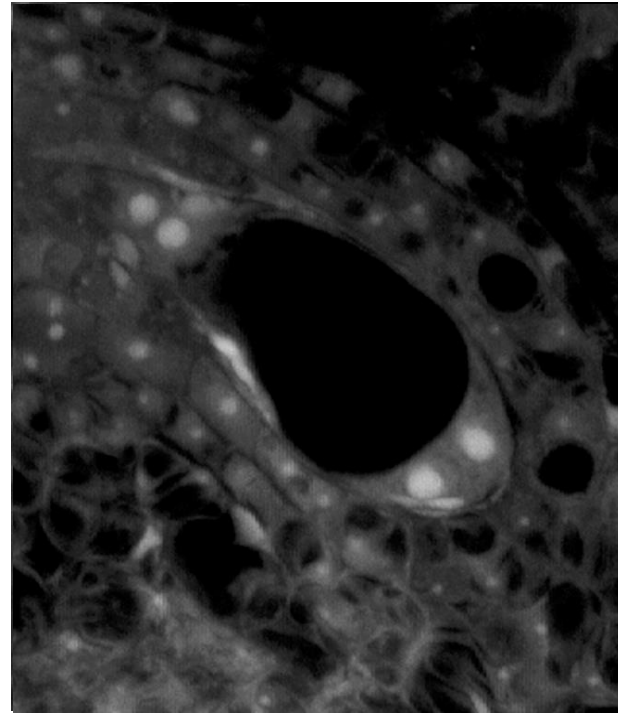


cki1-i specific primers

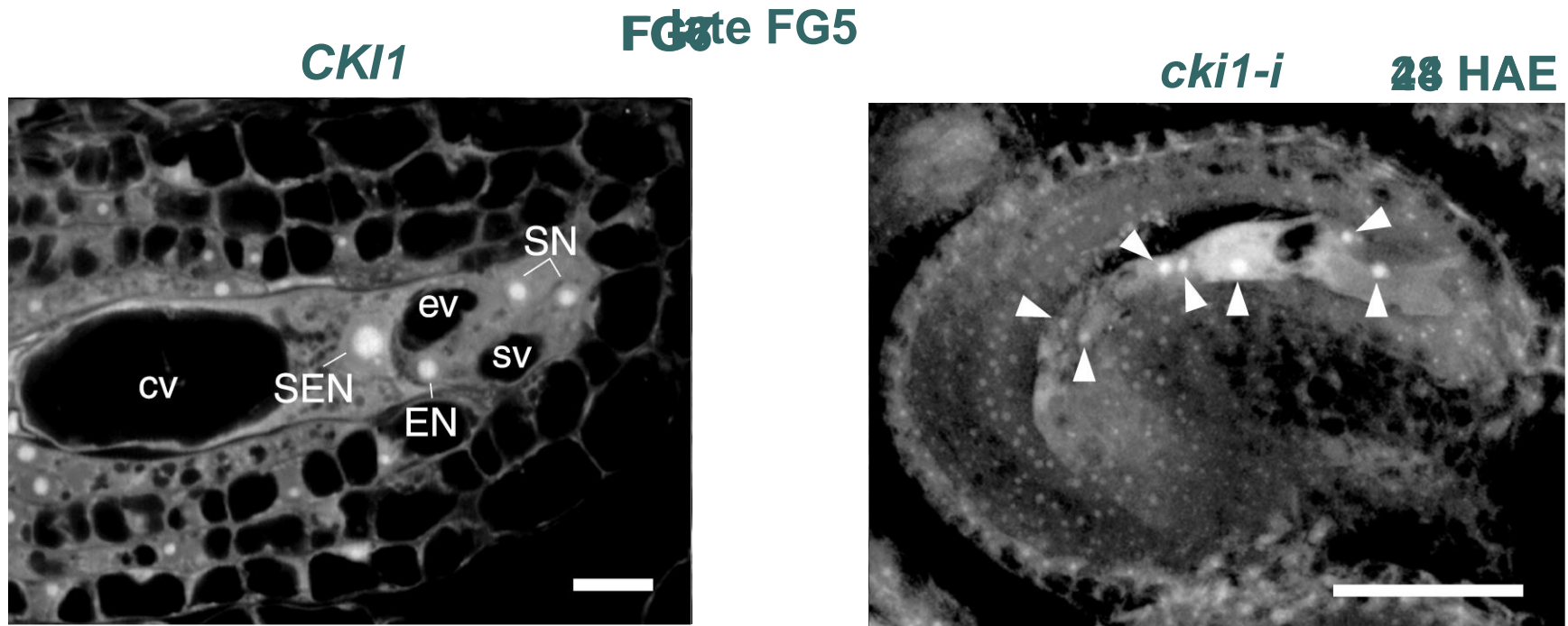
D. ♂ *CKI1/cki1-i* x ♀ wt

CKI1 and Megagametogenesis

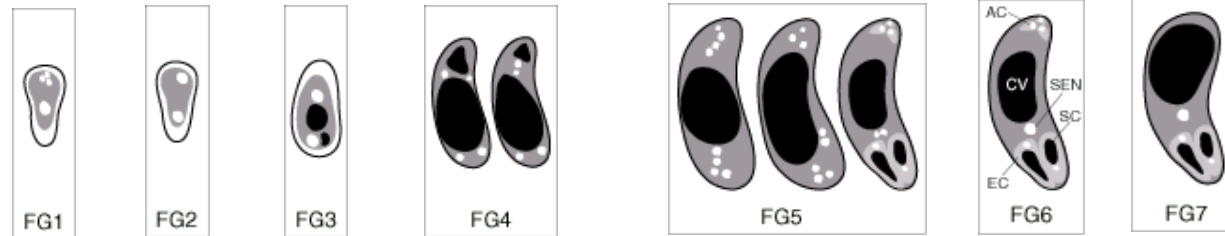
FG 



CKI1 and Megagametogenesis



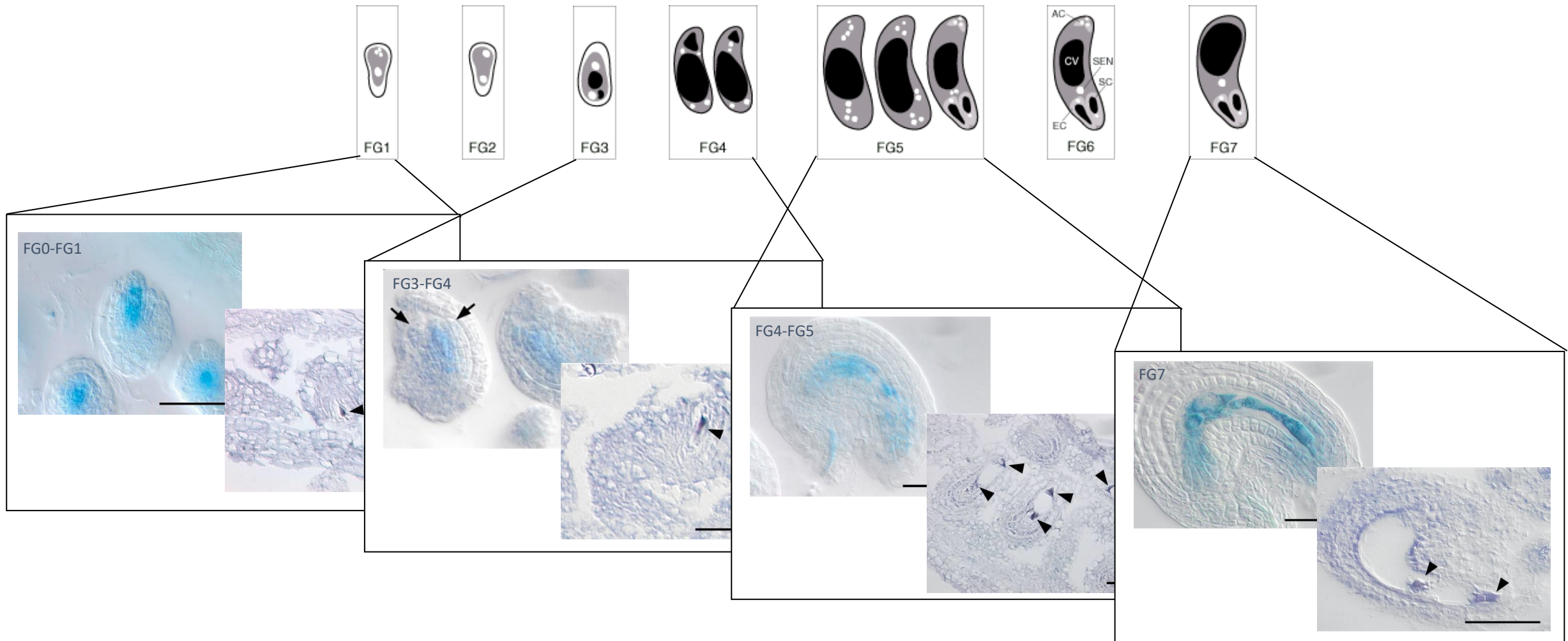
Hejátko et al., *Mol Genet Genomics* (2003)



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

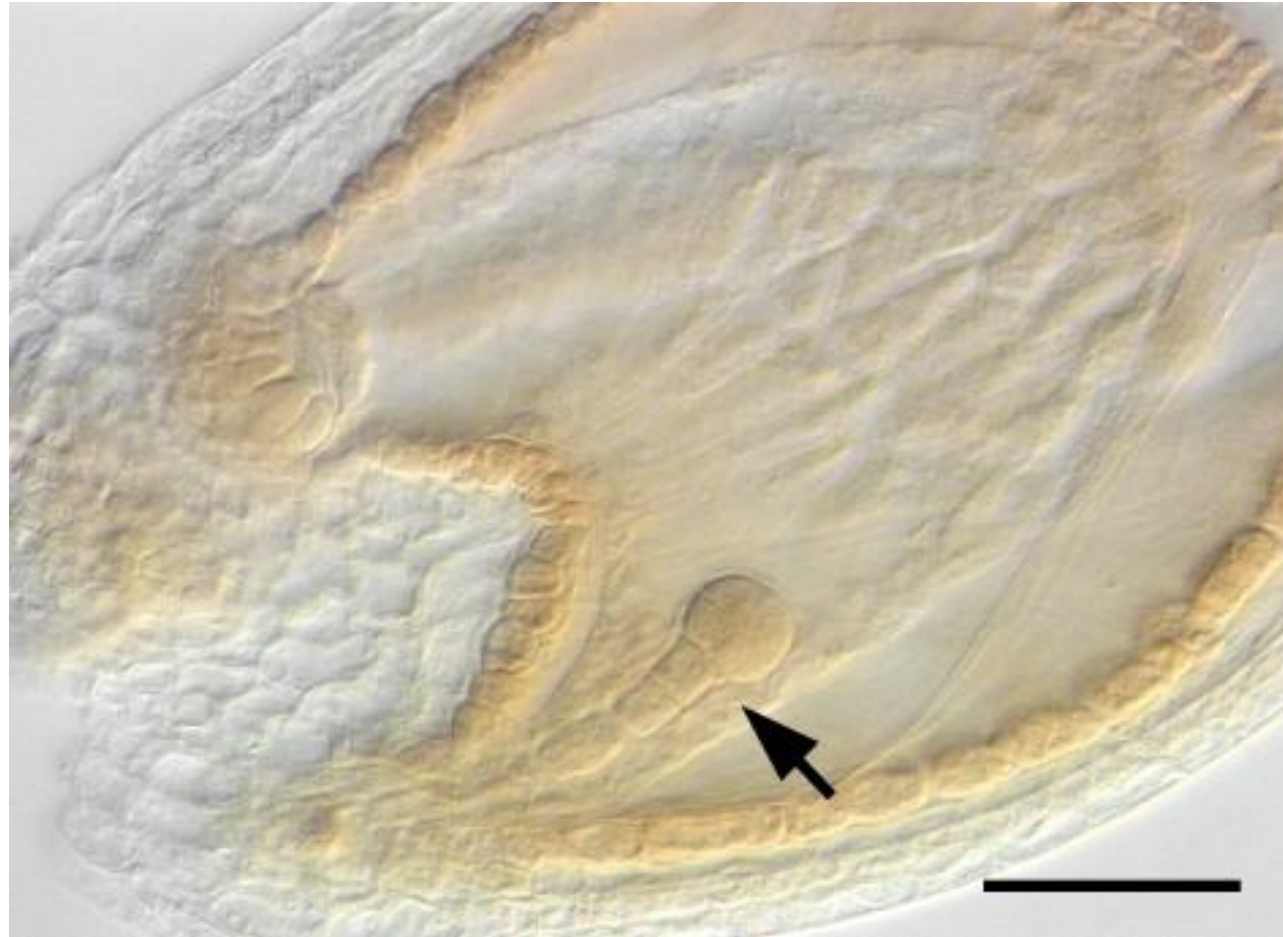
CKI1 and Megagametogenesis



Paternal *CKI1* is Expressed Early after Fertilization

♀ wt x ♂ Pro*CKI1*:*GUS*

22 HAP
(hours
after
pollination)



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

Outline

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

Sanger Sequencing

Frederick Sanger

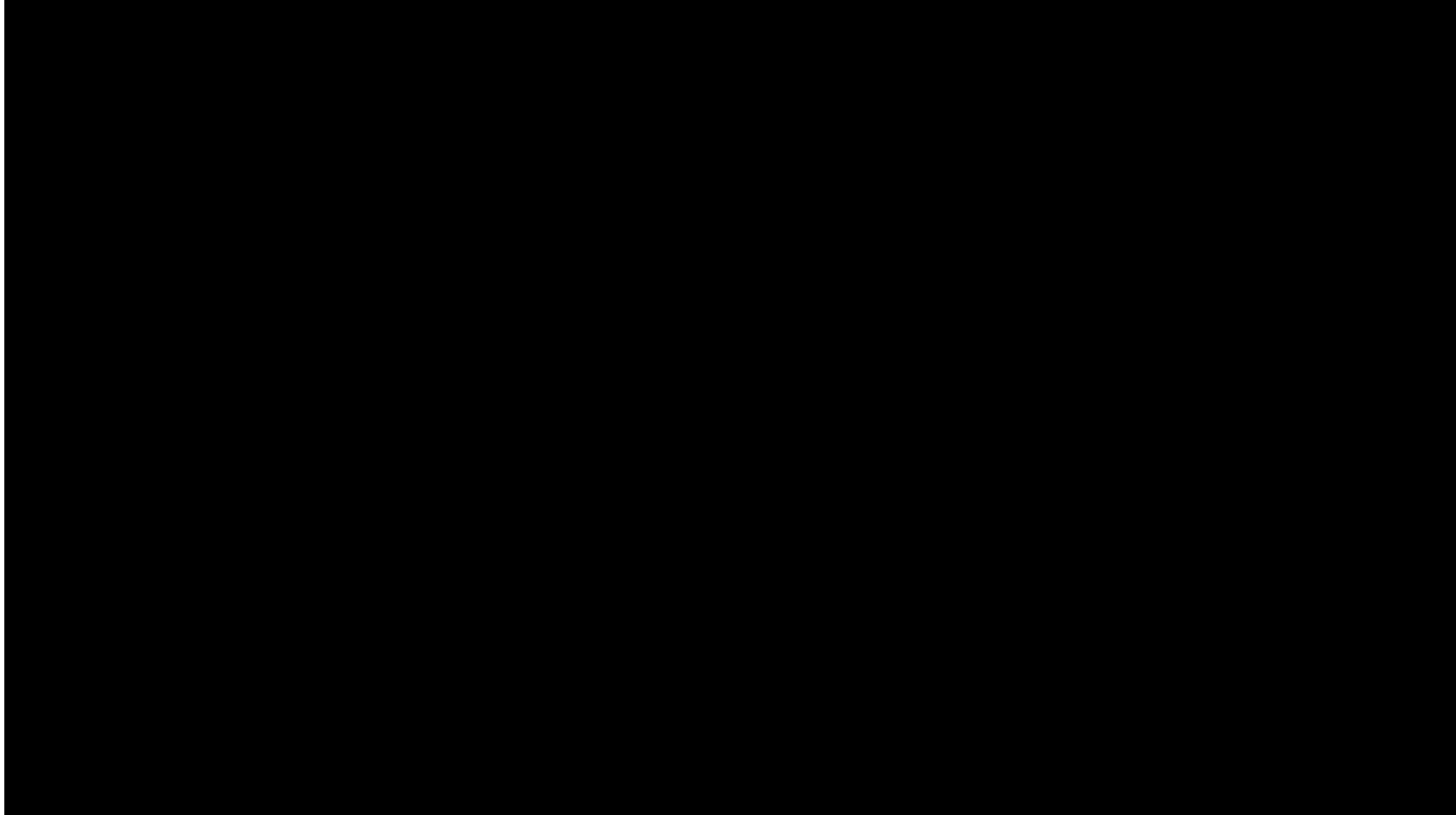
1958 – Nobel prize – insulin structure

1975 - Dideoxy sequencing method

1980 – second Nobel prize for NA sequencing



Sanger Sequencing



NGS Sequencing



Outline

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

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

Expression Assays

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

DNA Chips

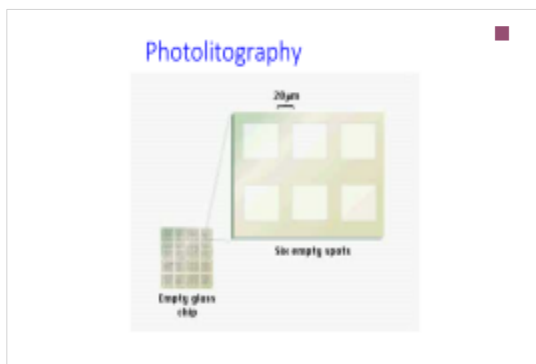
■ DNA čipy

- metoda umožňující rychlé porovnání velkého množství genů/proteinů mezi testovaným vzorkem a kontrolou
- nejčastěji jsou používány oligo DNA čipy

- k dispozici komerčně dostupné sady pro celý genom

- firma Operon (Qiagen), 29.110 70-mer oligonukleotidů reprezentujících 26.173 genů kódujících proteiny, 28.964 transkriptů a 87 microRNA genů *Arabidopsis thaliana*

- možnost používat pro přípravu čipů fotolitografické techniky-usnadnění syntézy oligonukleotidů např. pro celý genom člověka (cca $3,1 \times 10^9$ bp) je touto technikou možno připravit 25-mery v pouze 100 krocích)



čipy nejen pro analýzu exprese, ale např. i genotypování (SNPs – jednonukleotidové polymorfizmy, sekvenování pomocí čipů, ...)

Affymetrix ATH1 *Arabidopsis* genome array

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

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

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

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 [] Search

Experiment: Aluminum Stress

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

Slide Details

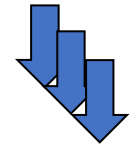
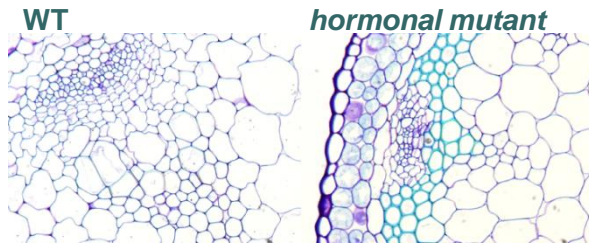
Slide (name ? : description)	External ID ?	Replicate (id ? : name)	Replicate type ?	Reverse replicate ?	Sample ?	Experimental variables	Label ?	Get Data ?
HoekengaS7 ["]: Aluminum Stress 1 [strong spatial bias]	AFGC: 7304	63: Aluminum Stress	technical		7304_Cy3.7305_Cy5	no treatment (pool of 3, 8, and 24 hours)	Cy3	Download
					7304_Cy5.7305_Cy3	Aluminum (50 5M AlCl3, pool of 3, 8, and 24 hours)	Cy5	
HoekengaS8 Aluminum Stress 2 [strong spatial bias]	AFGC: 7305	64: Aluminum Stress	technical	63	7304_Cy5.7305_Cy3	Aluminum (50 5M AlCl3, pool of 3, 8, and 24 hours)	Cy3	Download
					7304_Cy3.7305_Cy5	no treatment (pool of 3, 8, and 24 hours)	Cy5	

Gene Expression Assays

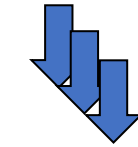
- Methods of gene expression analysis
 - Quantitative analysis of gene expression
 - DNA chips
 - Next generation transcriptional profiling

Next Gen Transcriptional Profiling

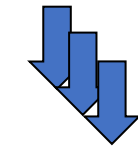
- *Transcriptional profiling* via *RNA sequencing*



mRNA

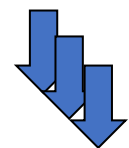


cDNA

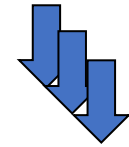


Sequencing by Illumina and

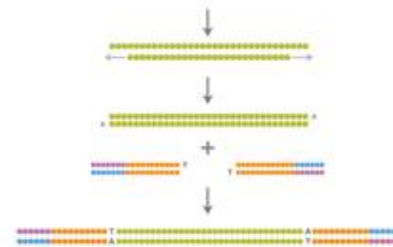
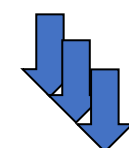
number of transcripts determination



mRNA



cDNA



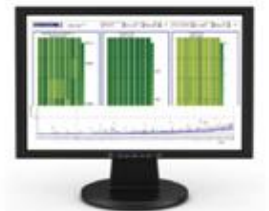
Library Preparation
~2 h [15 min hands-on (Nextera)]
< 6 h [< 3 h hands-on (TruSeq)]



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



Sequencing by Synthesis
~1.5 to 11 days



CASAVA
2 days (30 min hands-on)

Results of –omics Studies vs Biologically Relevant Conclusions

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

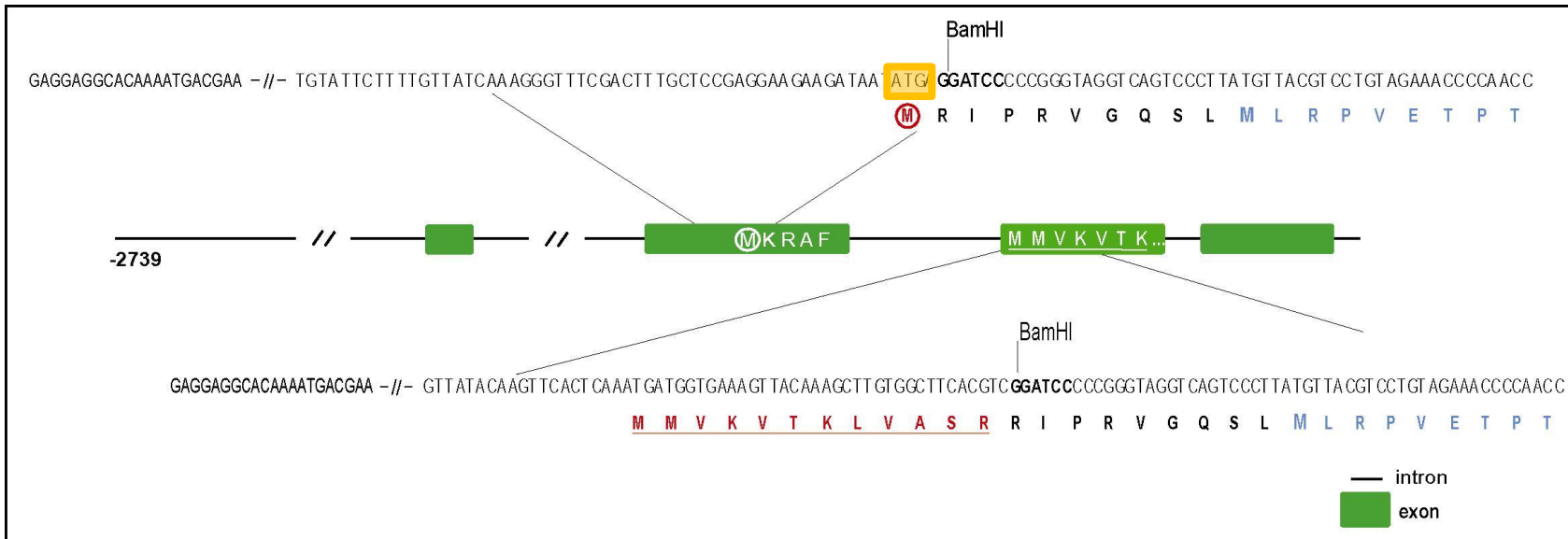
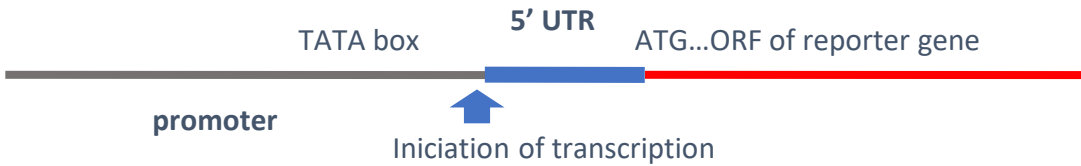
gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
AT1G07795	1:2414285-2414967	WT	MT	OK	0	1,1804	1.79769e+308	1.79769e+308	6.88885e-05	0,000391801	yes
HRS1	1:4556891-4558708	WT	MT	OK	0	0,696583	1.79769e+308	1.79769e+308	6.61994e-06	4.67708e-05	yes
ATMLO14	1:9227472-9232296	WT	MT	OK	0	0,514609	1.79769e+308	1.79769e+308	9.74219e-05	0,000535055	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	0,000285143	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	0,000374736	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

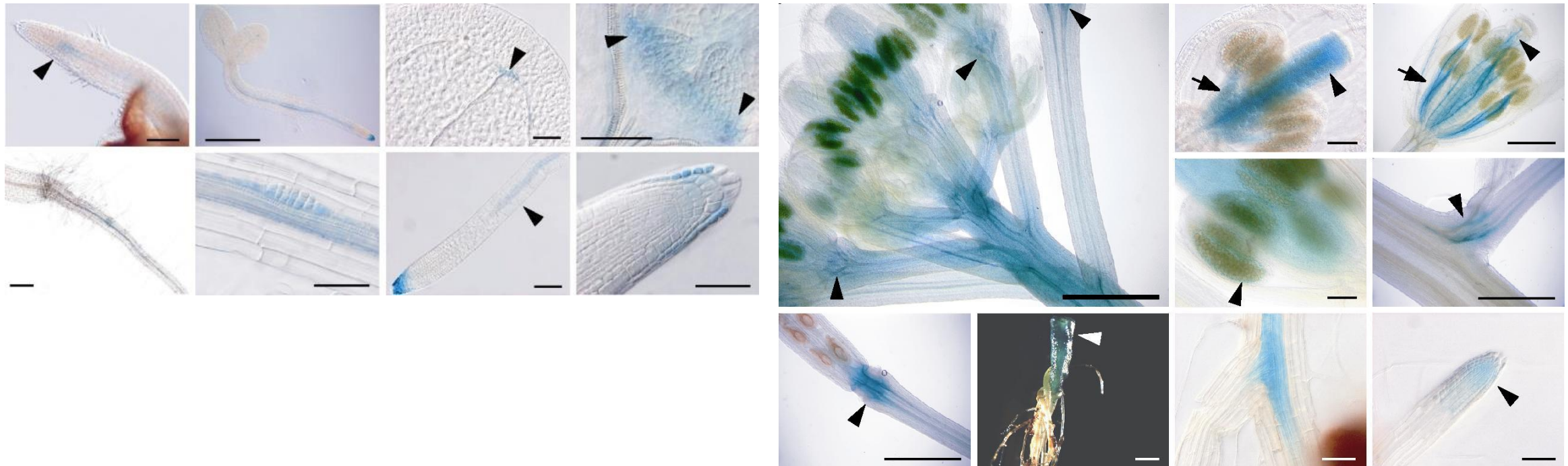
Transcriptional Fusion

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

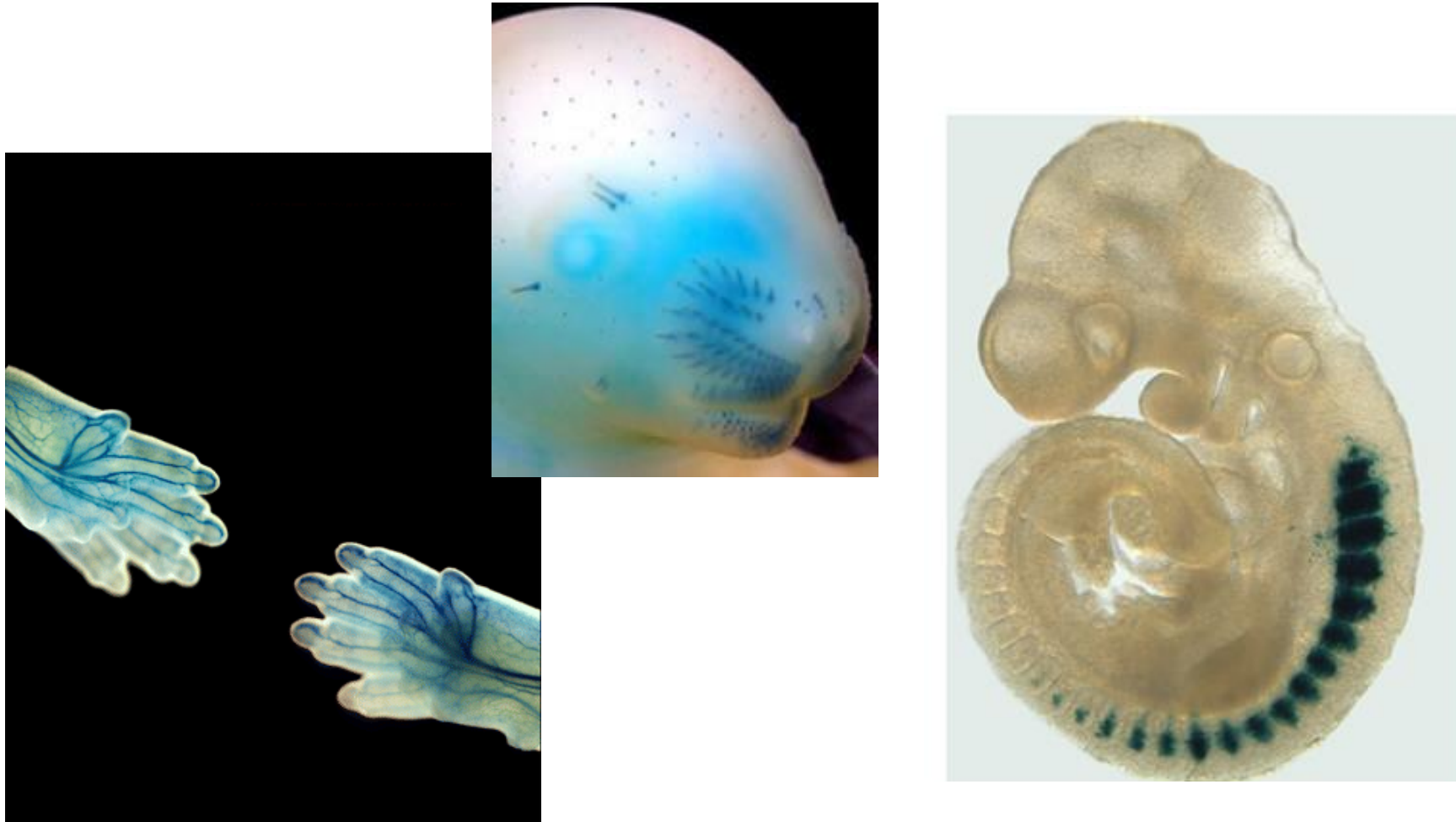


Transcriptional Fusion

- 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



LacZ Reporter in Mouse Embryos

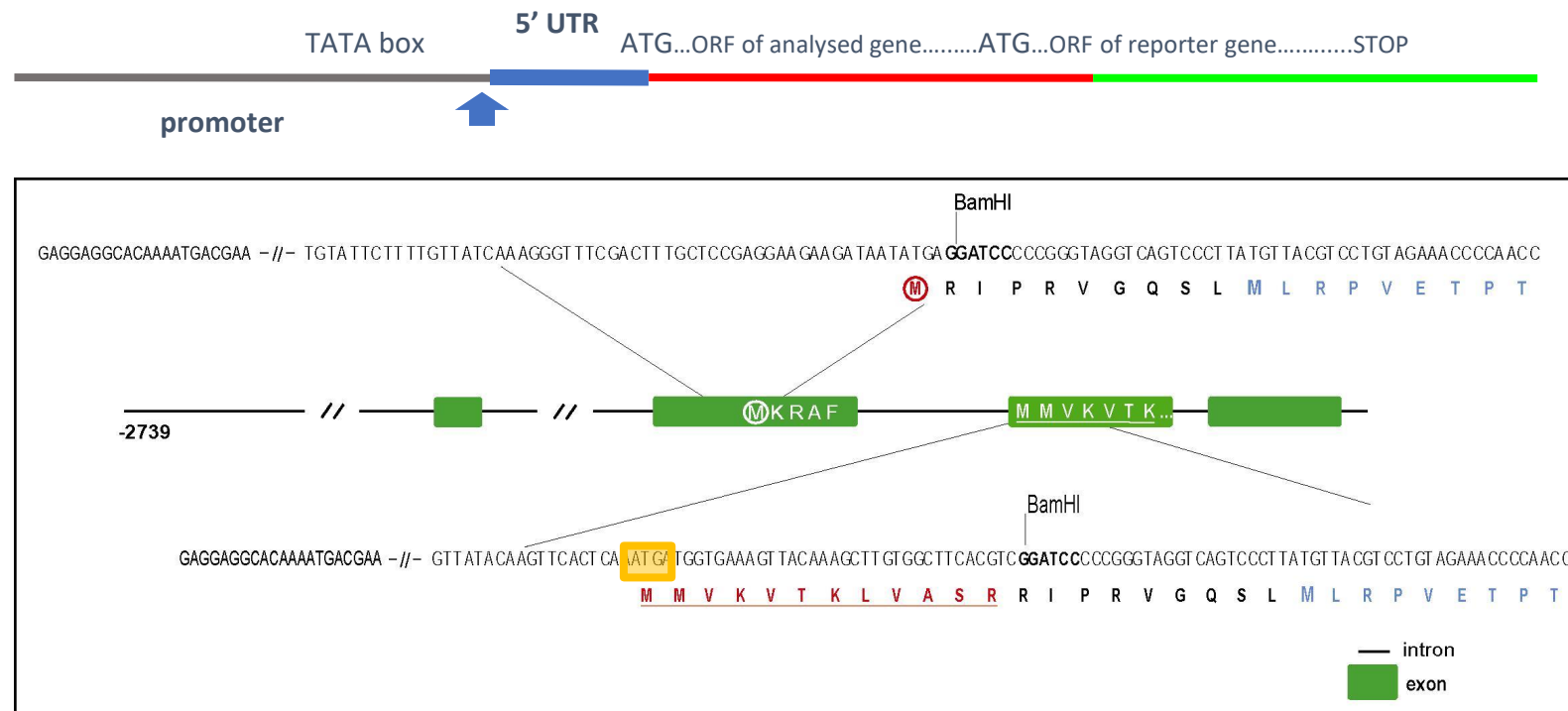


Gene Expression Assays

- 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

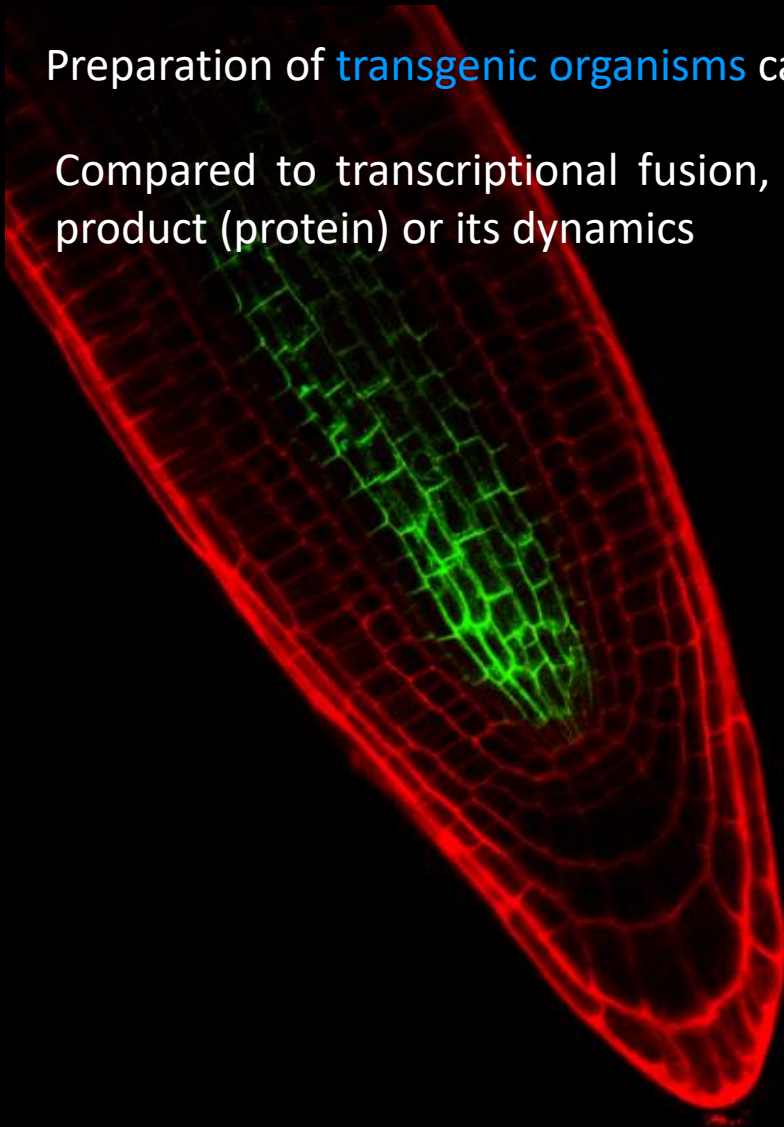
Translational Fusion

- 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)

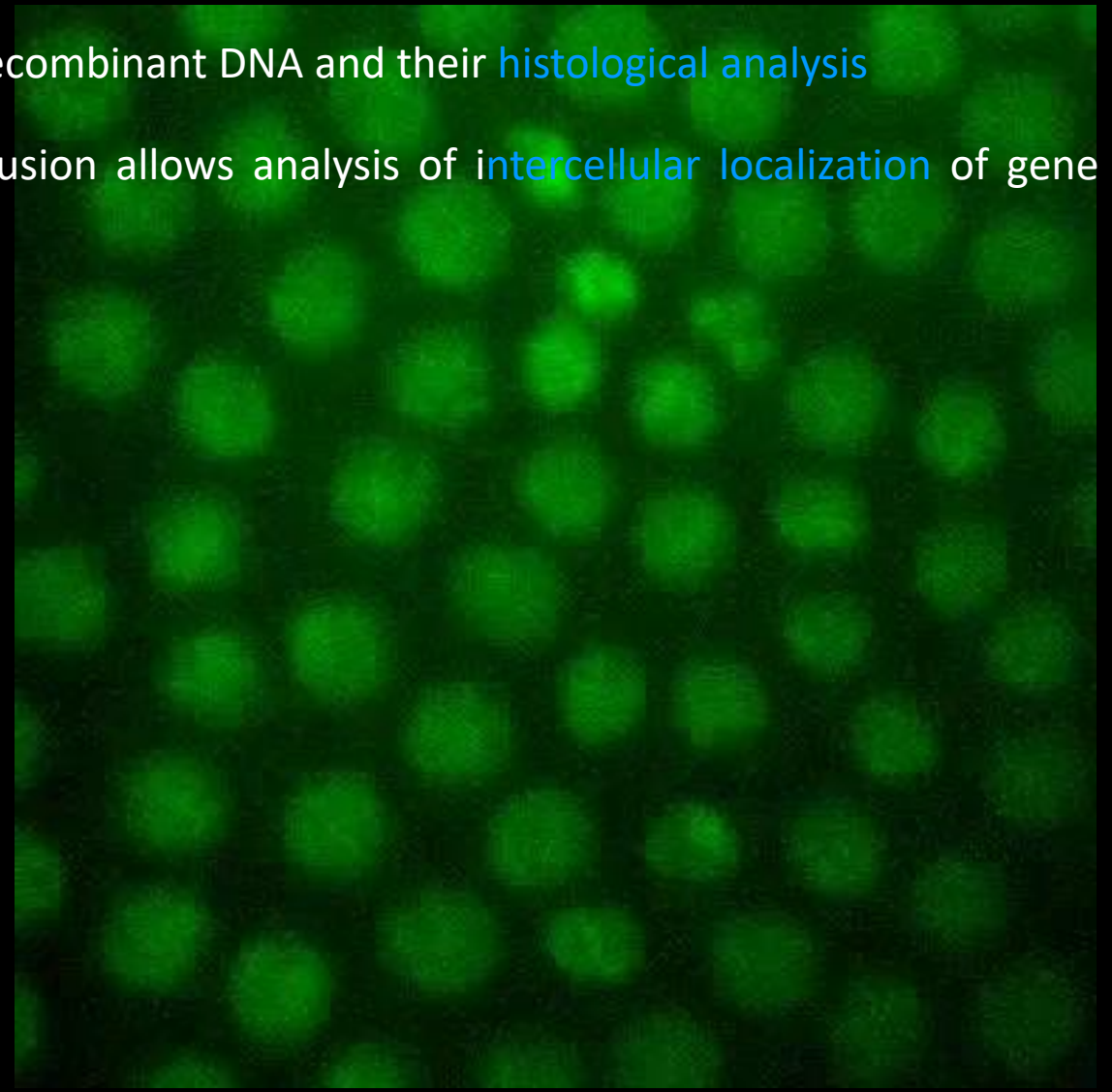


Translational Fusion

- 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

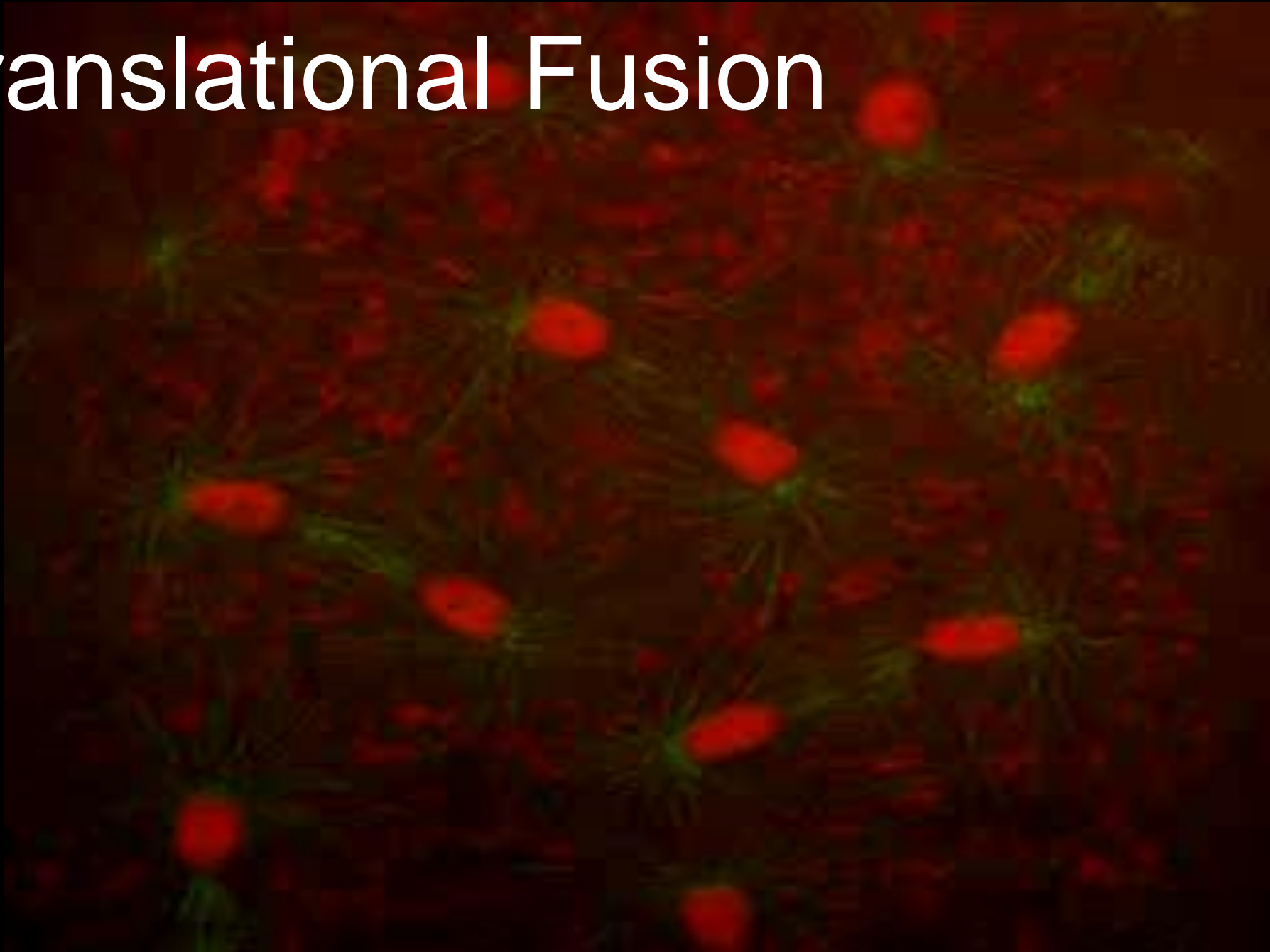


PIN1-GFP in *Arabidopsis*



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

Translational Fusion

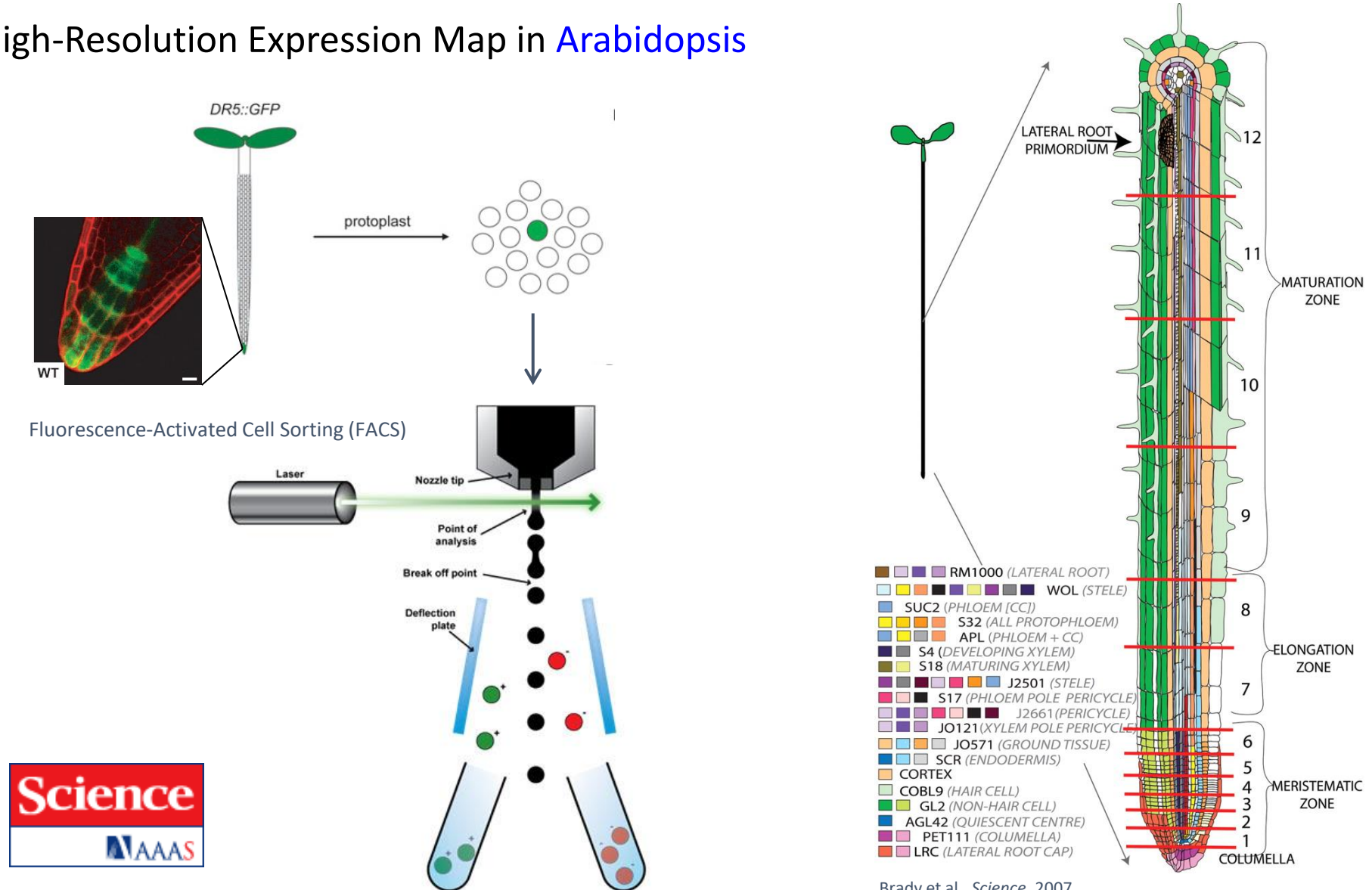


Gene Expression Assays

- 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
 - Use of the data available in public databases
 - Tissue- and cell-specific gene expression analysis

Gene Expression Assays

- High-Resolution Expression Map in *Arabidopsis*



BAR ePlant

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

BAR ePlant

Welcome Screen

Enter a gene name

Example: ABI3 or AT5G60200

Expression Angler

Mutant Phenotype Selector

RSVP OFF

1 gene / gene product currently loaded

- 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

Data visualization tools for multiple levels of plant data.

environment

natural variation

DNA sequence (gene & promoter)

ncRNA & conserved regions outside genes

methylation

transcription

RNA transcript (sequence, abundance & alternative splicing)

signaling & signal transduction cascades

protein networks

translation

protein sequence

3D structure

metabolism

secondary metabolism

subcellular localization

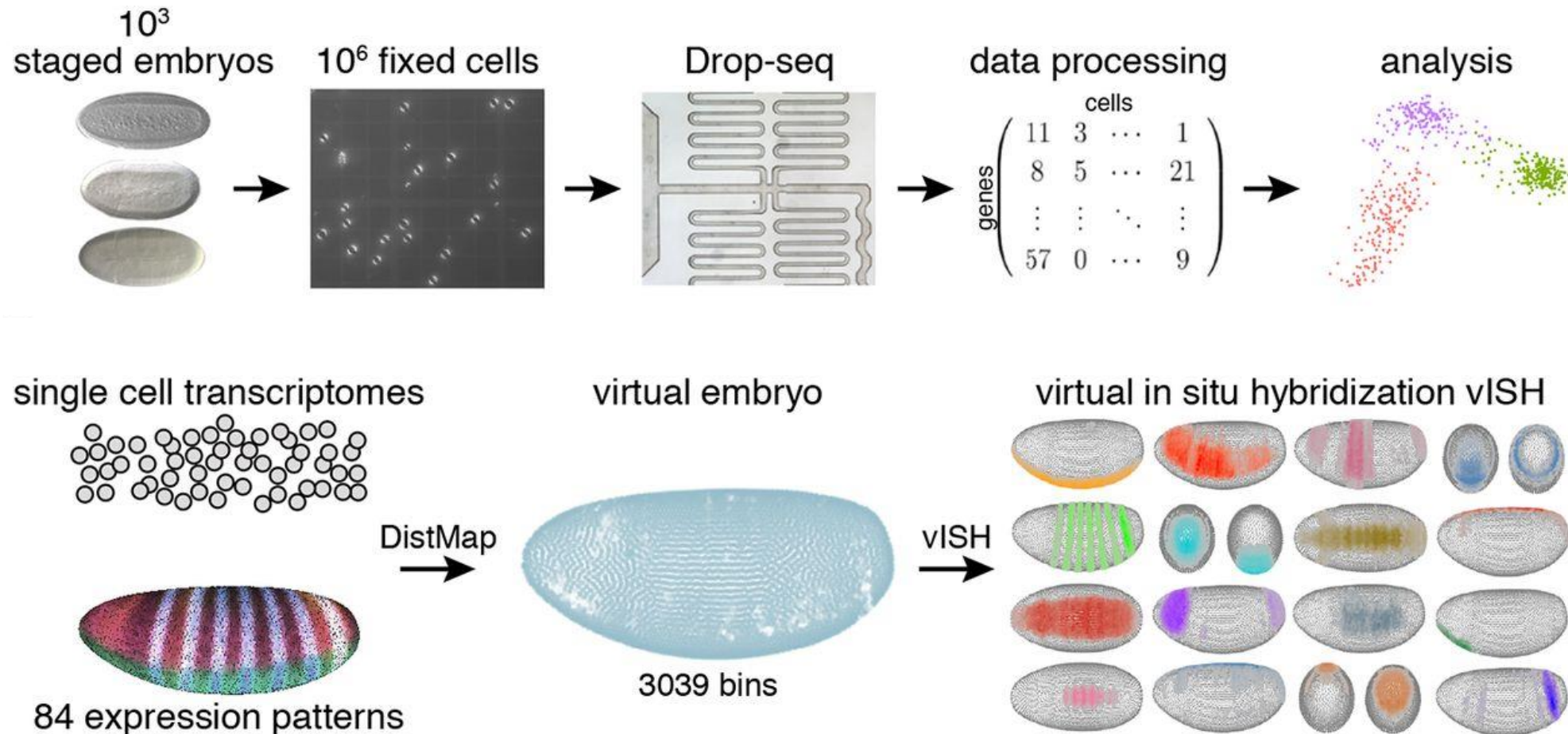
spatio-temporal distribution / perturbation response

primary metabolism

phenotype / response

Expression Maps - RNA

- High-Resolution Expression Map in *Drosophilla*



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

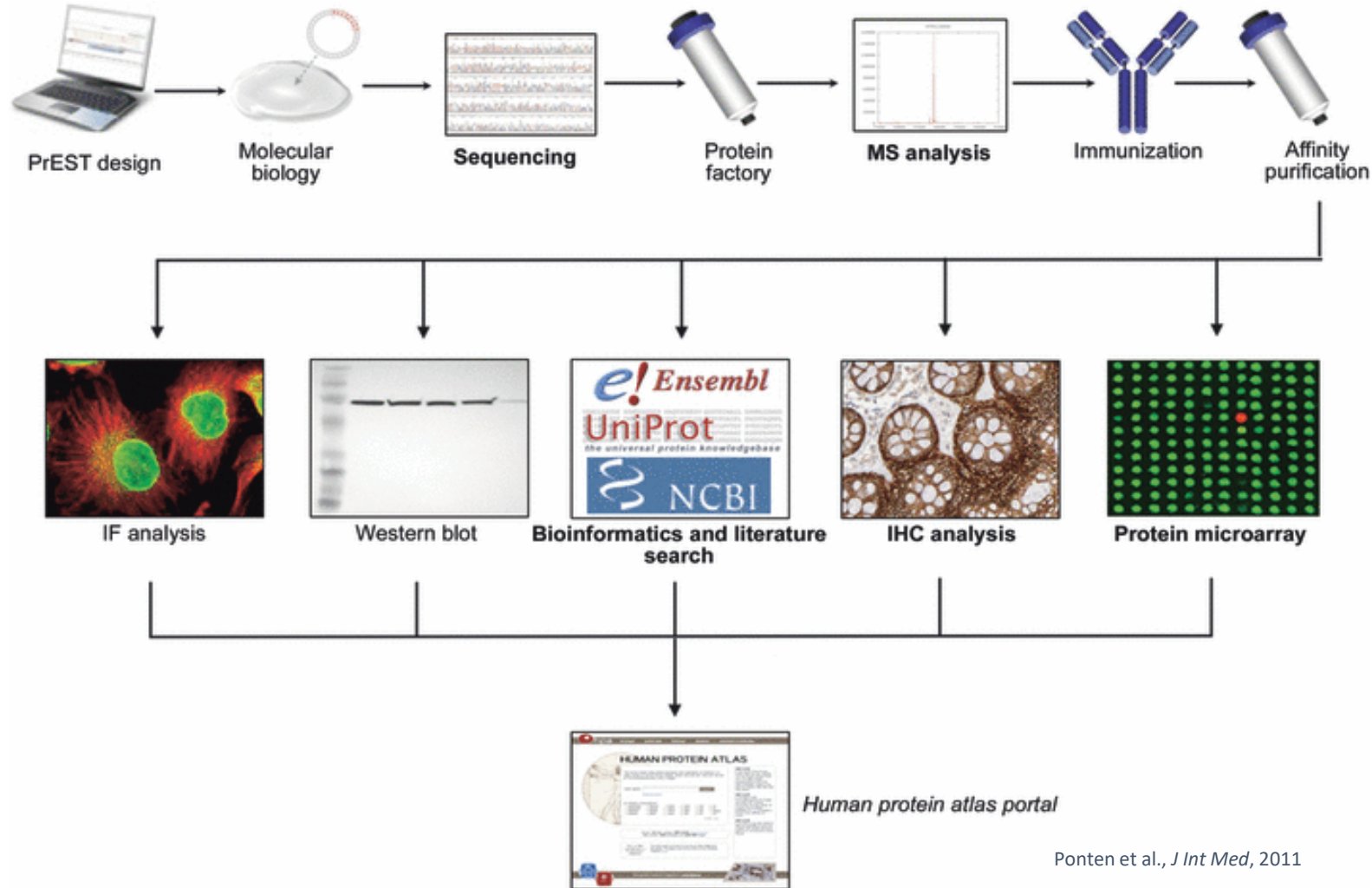
^bSystems biology of neural tissue differentiation, BIMSB, MDC

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.

Expression Maps - Proteins

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



Expression Maps - Proteins

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

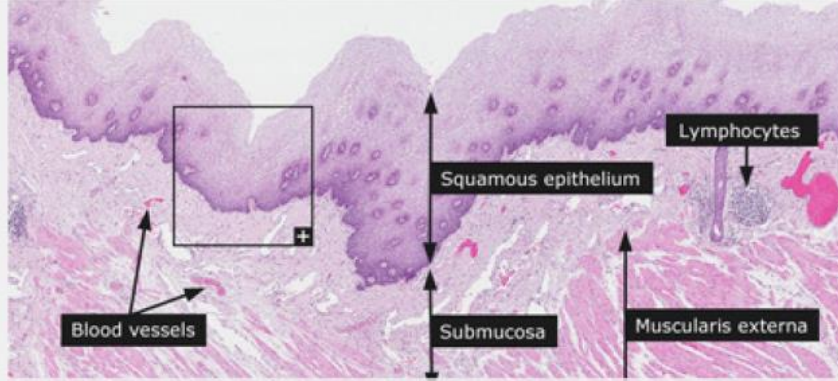
THE HUMAN PROTEIN ATLAS

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

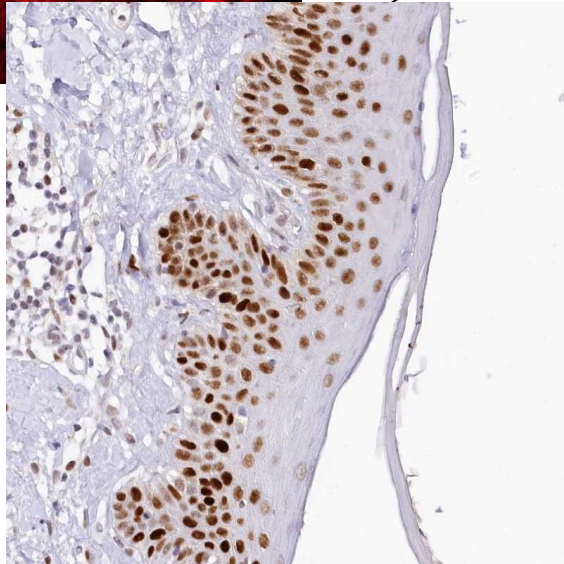
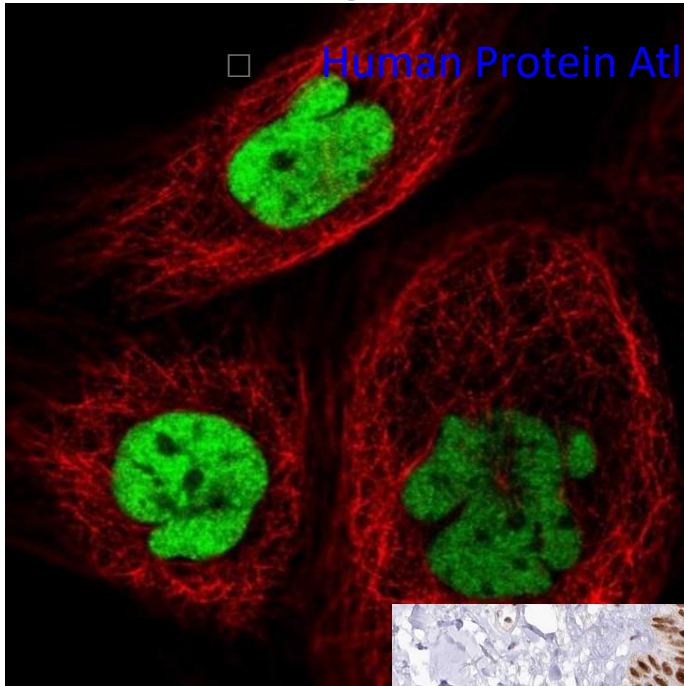


The Human Protein Atlas project is funded by the Knut & Alice Wallenberg foundation.



Expression Maps - Proteins

□ [Human Protein Atlas](http://www.proteinatlas.org/) (<http://www.proteinatlas.org/>)



SUBCELLULAR LOCATION SUMMARY ? »

Main location(s) Nucleus but not nucleoli

Additional location(s)

Staining summary Localized to the nucleus but excluded from the nucleoli.

Reliability (APE) High

Antibodies in assay CAB039238, CAB039239

[Show image >](#)

NORMAL TISSUE & ORGAN SUMMARY ? »

Expression summary Fractions of cells showed weak nuclear and/or cytoplasmic expression.

Tissue specificity Expressed in 11 out of 82 cell types

Reliability (APE) High

Antibodies in assay CAB002973, CAB039238, CAB039239

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

[Show image >](#)

MORE SUBCELL DATA

MORE TISSUE DATA

Summary

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

Discussion