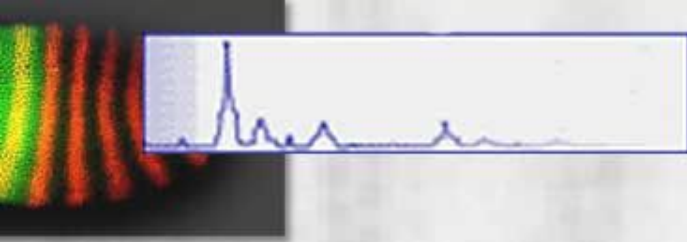


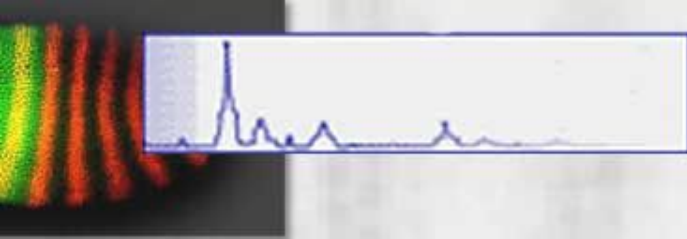
On tentative research projects

Maria Samsonova
*St.Petersburg State Polytechnical
University, Russia*

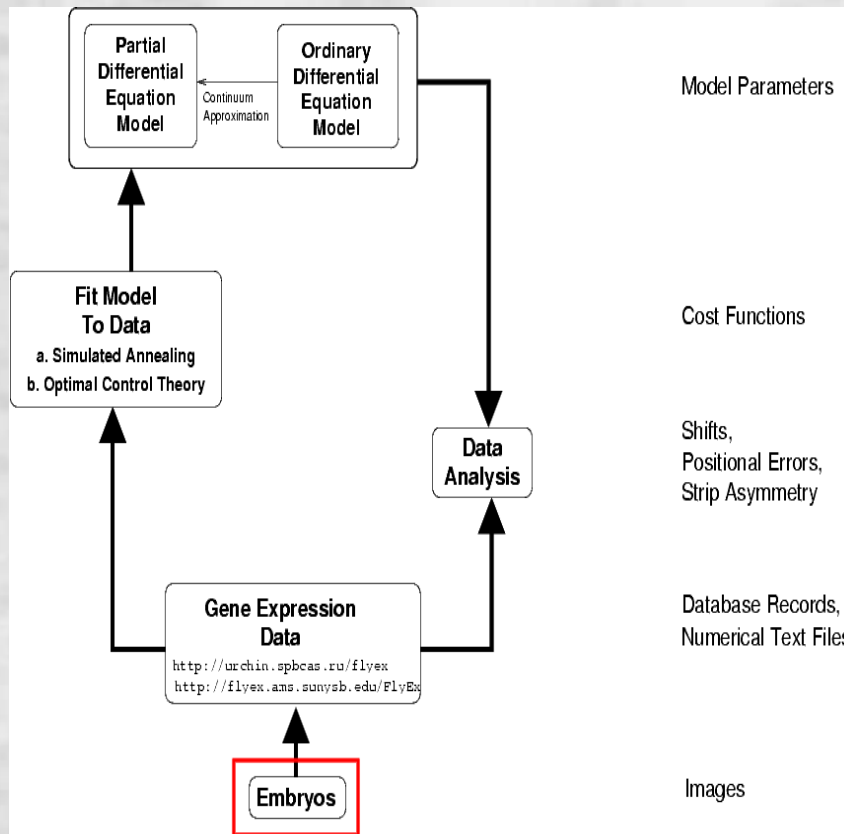


Research focus

- methods and tools for acquisition of high-precision data;
- mathematical methods for data mining and system modeling;
- methods for data integration and information extraction.



Systems biology of segmentation in *Drosophila*



letters to nature

Dynamic control of positional information in the early *Drosophila* embryo

Johannes Jaeger¹, Svetlana Surkova², Maxim Blagov², Hilde Janssens¹, David Kosman³, Koestantin M. Kazlov², Manu¹, Ekaterina Myasnikova², Carlos E. Vazario-Alonso^{1,4}, Maria Samsonova², David H. Sharp⁵ & John Reinitz¹

¹Department of Applied Mathematics and Statistics, and Center for Developmental Genetics, Stony Brook University, Stony Brook, New York 11794-3600, USA

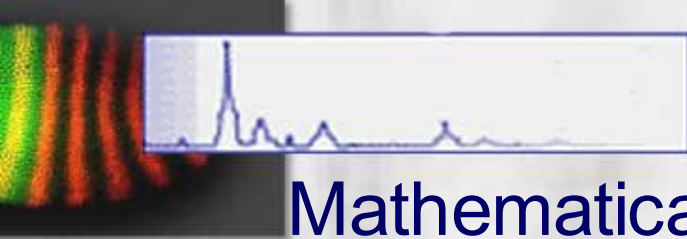
²Department of Computational Biology, Center for Advanced Studies, St Petersburg State Polytechnic University, St Petersburg, 195251 Russia

³Department of Biology, University of California, San Diego, California 92093, USA

⁴Universidade Federal do Rio de Janeiro, Instituto de Biofísica Carlos Chagas Filho, Rio de Janeiro, Rio de Janeiro 21949-900, Brazil

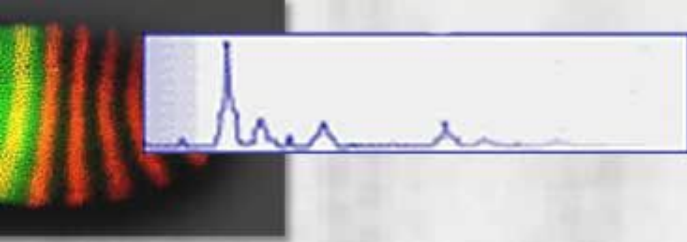
⁵Applied Physics Division, and Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Morphogen gradients contribute to pattern formation by determining positional information in morphogenetic fields^{1,2}. Interpretation of positional information is thought to rely on direct, concentration-threshold-dependent mechanisms for establishing multiple differential domains of target gene expression^{1,3,4}. In *Drosophila*, maternal gradients establish the initial position of boundaries for zygotic gap gene expression, which in turn convey positional information to pair-rule and segment-polarity genes, the latter forming a segmental pre-pattern by the onset of gastrulation⁵⁻⁷. Here we report, on the basis of quantitative gene expression data, substantial anterior shifts in the position of gap domains after their initial establishment. Using a data-driven mathematical modelling approach⁸⁻¹¹, we show that these shifts are based on a regulatory mechanism that relies on asymmetric gap-gap cross-repression and does not require the diffusion of gap proteins. Our analysis implies that the threshold-dependent interpretation of maternal morphogen concentration is not sufficient to determine shifting gap domain boundary positions, and suggests that establishing and interpreting positional information are not independent processes in the *Drosophila* blastoderm.



Mathematical modeling and statistical analysis of data

- Various statistical and machine learning methods are in use for data mining;
- Models based on ordinary and partial differential reaction-diffusion equations;
- Different optimization methods are available to find the system parameters:
 - Simulated Annealing
 - Optimal Steepest Descent Algorithm
 - Tunneling
 - Parallel Differential Evolution.



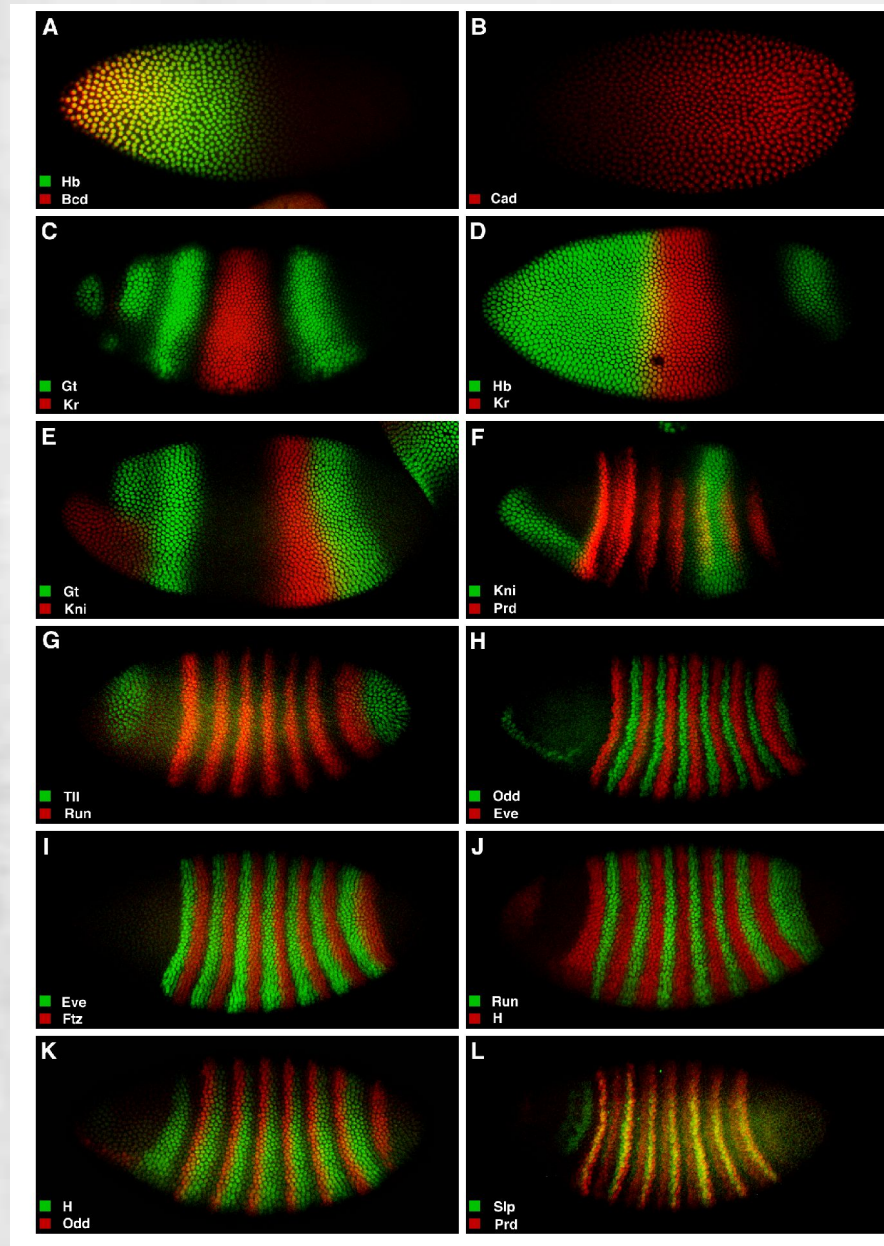
Open questions

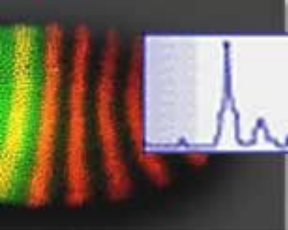
- Construction of models of biological systems and processes which operate on different time scales and/or complexity levels
- Approaches towards validation of existing models;
- Adequate representation of diffusion in models;
- Logics behind selection of relevant models to uncover the structure and dynamics of a particular biological system.



Acquisition of high-precision quantitative data: confocal scans of gene expression patterns

Myasnikova et al (2001), Bioinformatics 17:3-12

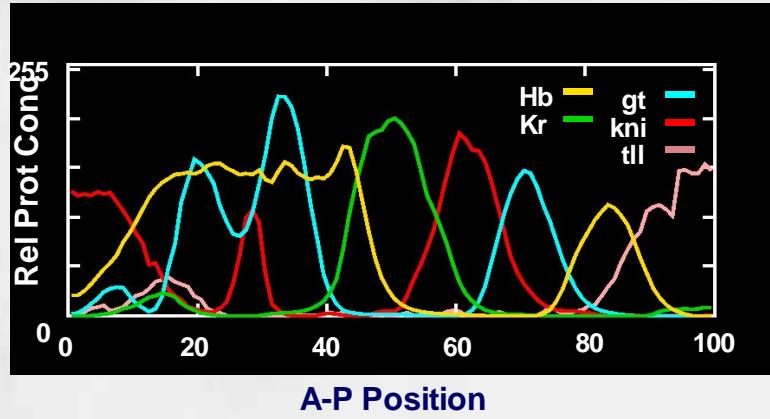
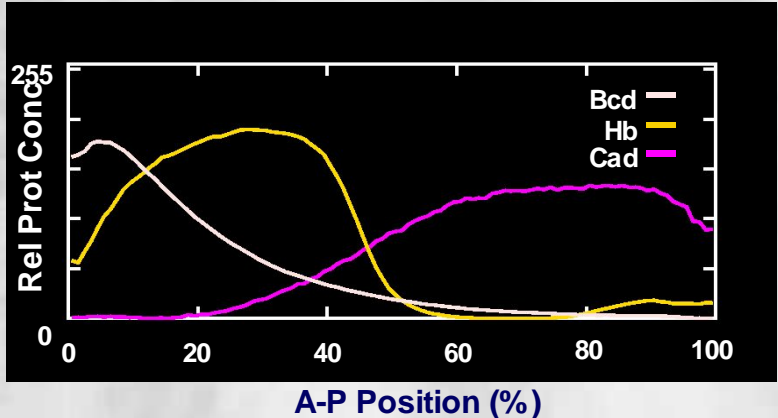
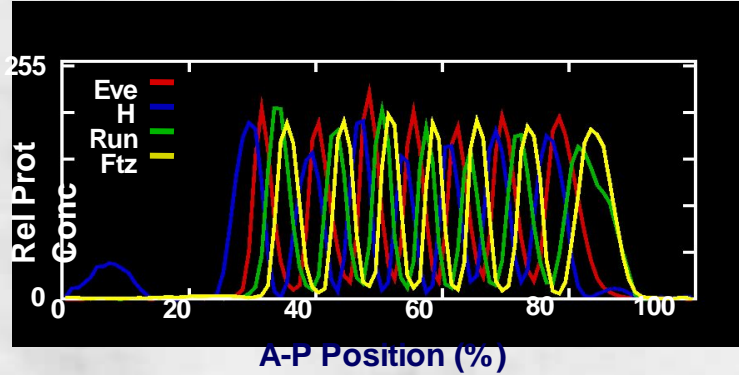


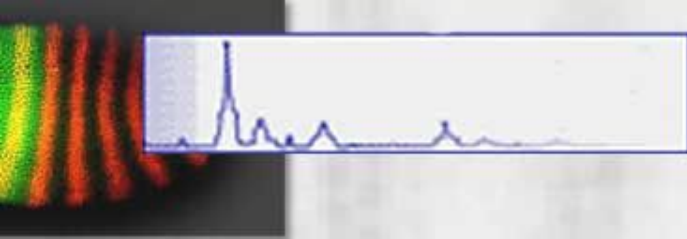


Quantitative data on segmentation gene expression with cellular resolution in space and temporal resolution of 6.5 minutes of development

Nuclear Coordinate **Fluorescence intensities**

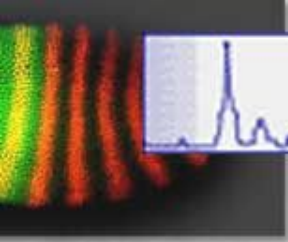
S	Coordinate	Intensity 1	Intensity 2	Intensity 3	Intensity 4
0	3.67346	41.5401	53.6951	25.939	
1	4.05099	45.3917	53.8919	27.0811	
2	4.00196	37.3767	54.7	22.85	
3	5.01298	39.8973	59.4688	26.5625	
4	5.07766	52.1475	58.8267	29.04	
5	5.44521	43.4772	62.25	30.5125	
6	5.52642	47.3609	56.7215	28.9494	
7	5.78564	36.9171	56.8689	26.2623	
8	6.26299	50.786	64.3521	30.6056	
9	6.62735	40.1422	59.3086	32.0123	
10	6.69946	54.7947	60.0795	29.5227	





Why the effective image processing packages are necessary to develop?

- we need high-precision quantitative data with high spatial resolution;
- different sophisticated microscope techniques are currently available;
- software packages provided by microscope manufactures are designed to produce a high-quality image. Usually do not support effective processing and analysis of images in batch.



Pro-stack package

(“Prostack” means “a straight man” in Russian ;-))

Functionality:

- **Processing of *stacks*** of images,
- More than 40 image processing methods, from thresholding to the calculation of object characteristics.
- The number of methods is growing.

Technical Data:

- Implemented in ANSI C as methods library libparus and command line interface prostack.
- All methods are available in distributed computing environment iSIMBioS via wrappers written in Perl.
- Can work with images in TIFF format with 8 bits per pixel using libtiff library for I/O operations.

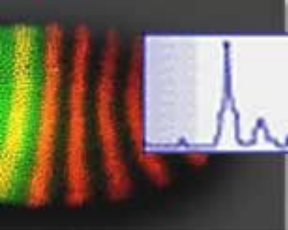


Image processing scenario

C:\PIPE\pegas_public\wksp\c3na\rotate.ap

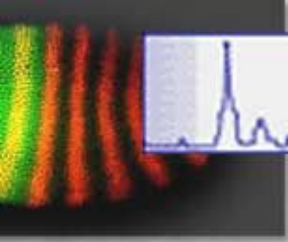
File Edit Element Execute View Tools Help

Open Save msIE preView LinkType Options Font PAM

mal-laptop.spbcas.ru:7778/
freya.spbcas.ru:7776/
convert
prostak
max
avg
rotate
turn
crop
geometry
blob
test
freya.spbcas.ru:7776/

rmask.tif → : PAM : rotate >
i_E.tif → : PAM : turn >
i_l.tif → : PAM : turn >
i_A.tif → : PAM : turn >

The screenshot displays the PAM (Picture Archiving and Management) software interface. The main workspace contains a workflow diagram with four input files (rmask.tif, i_E.tif, i_l.tif, i_A.tif) and three processing steps (PAM: rotate, PAM: turn, PAM: turn). The left sidebar lists various processing tools. Below the main workspace are two image preview windows: the left one shows a white oval mask on a black background, and the right one shows a grayscale image of a textured oval. The Windows taskbar at the bottom shows the system tray with the time 9:24.



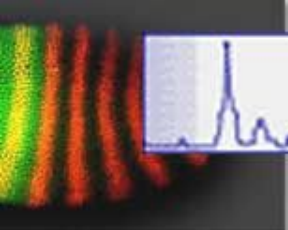
Data integration

- We have developed a method for integration of databases with common subject domain (<http://urchin.spbcas.ru/NLP.htm>).
- Now we propose to extend our approach to design an information management system for collaboration within distributed working environment.

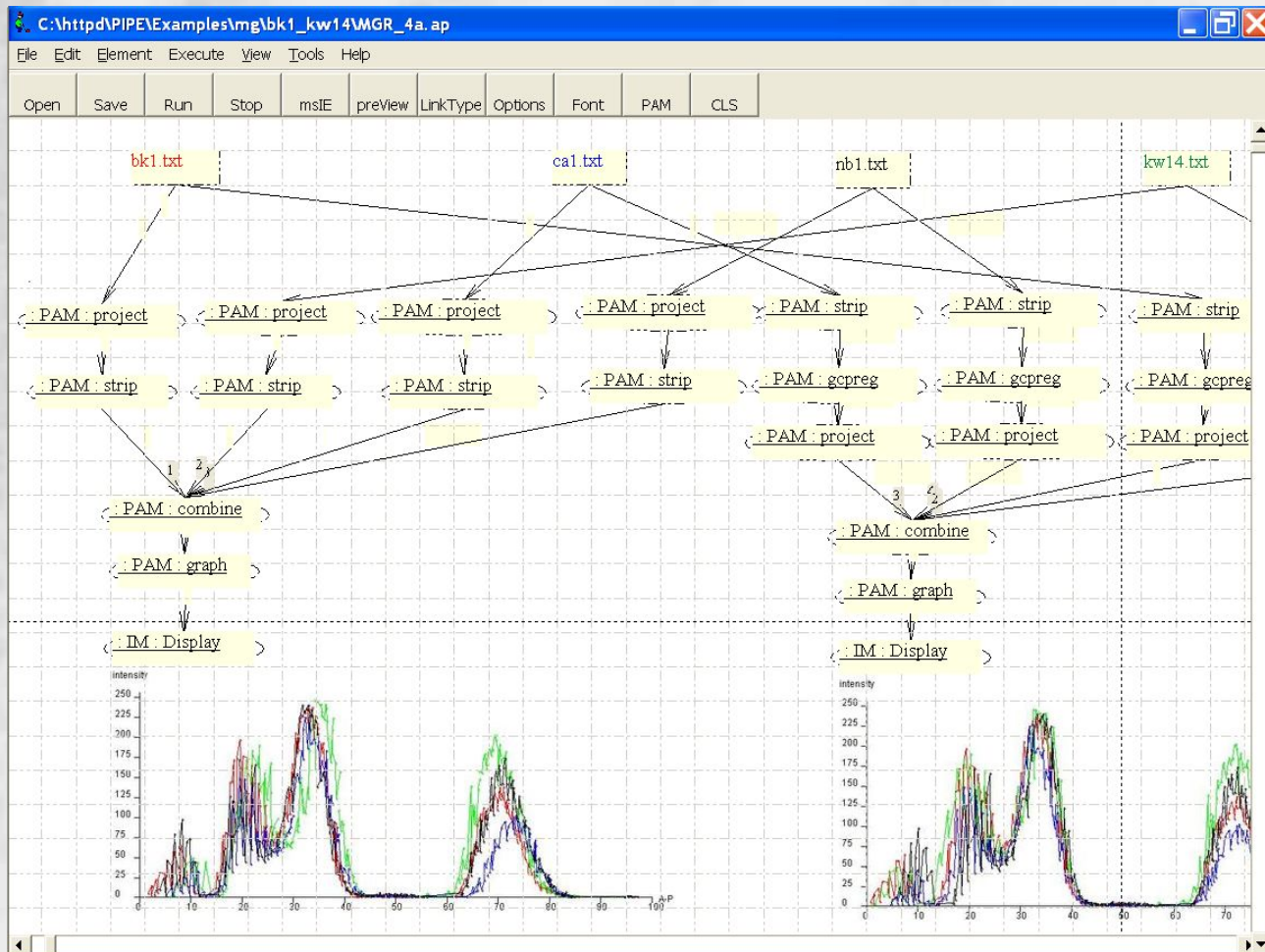


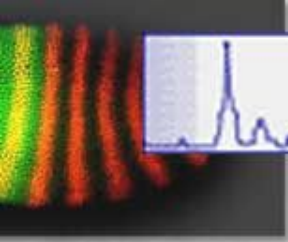
iSIMBioS

- ✓ fast and convenient access to data;
- ✓ data can be processed and analyzed by combining programs and services (modules) into workflows;
- ✓ workflow modules and data can be distributed over network;
- ✓ workflows can be constructed visually;
- ✓ simultaneous access of multiple users to shared data and methods;
- ✓ extendable, scalable and flexible in specification and modification of analysis methods;
- ✓ failure resistant, portable;
- ✓ provides access through firewall and proxy servers;
- ✓ dissemination of data and programs.



iSIMBioS workflow: estimation of the registration accuracy





Information extraction

- We have developed a method for processing of natural language queries to a relational database (*Samsonova et al. (2003) Bioinformatics 19, suppl. 1, i241-249*).
- Now we propose to use NLP for information extraction from biomedical literature.

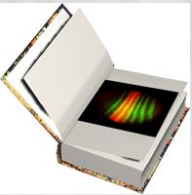


FlyEx database



Contains


- Images of segmentation gene expression in individual embryos.
- Quantitative data on segmentation gene expression in each nucleus of an individual embryo.
- Averaged data on expression of each segmentation gene at each time point and at cellular resolution.

Natural language interface is available at
<http://urchin.spbcas.ru/NLP/NLP.html>



Russian Version of the Query Form

Адрес  http://urchin.spbcas.ru/NLP/NLrus.html  [Переход](#) [Ссылки](#)

 **Естественно-языковой интерфейс**

Запрос:

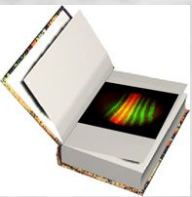
Макс. число возвращаемых строк (по умолчанию - все):

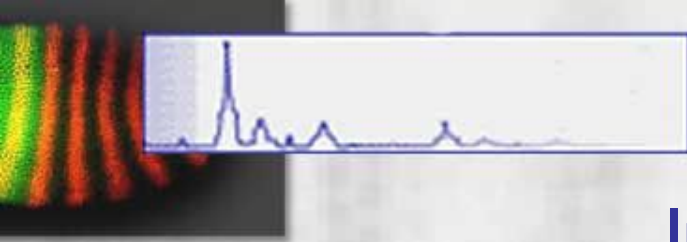
Примеры запросов:

- эмбрионы ранние классы Kruppel
- картина экспрессии эмбрионы класс 5
- количественные данные sq7 таблица
- картина экспрессии bd1 bicoid
- гены bd1
- Сколько эмбрионов, принадлежащих к ранним временным классам, сканировано для выявления экспрессии гена even-skipped?

[Переключиться на английский](#)

[Описание](#) [Концептуальная схема](#) [Логическая схема](#)





Information extraction

- In collaboration with Prof. Rubashkin (St.Petersburg State University) and Prof. Kolchanov (Institute for Cytology and Genetics of the RAS) groups.
- System prototype to automatically extract information about gene expression in *Arabidopsis thaliana*
 - text analysis in both automatic and interactive modes;
 - interactive training of the system by an expert;
 - use of text recognition and analysis algorithms;
 - relational database to store extracted information.



Acknowledgments

St.Petersburg

Alexander Samsonov

Vitaly Gursky

Konstantin Kozlov

Ekaterina Myasnikova

Andrei Pisarev

Ekaterina Poustelnikova

Svetlana Surkova

Stony Brook

John Reinitz

Jean Cadet

King-Wai Chu

Yuefan Deng

Hilde Janssens

Johannes Jaeger

Manu

San Diego

Dave Kosman

Los

Alamos

Dave Sharp

Bielefeld

Ralf Hoffstadt

Novosibirsk

Nikolai Kolchanov.

Nadezhda Omelyanchuk

<http://urchin.spbcas.ru/flyex>

<http://flyex.ams.sunysb.edu/flyex>