

# Электронная библиотека AbOvo, содержащая информацию о генных сетях, контролирующих эмбриональное развитие

М.Г. Самсонова, С.Ю. Суркова, В.Н. Серов, К.А. Соболев,  
Р.М. Шулаков, А.В. Веремьев, Е.Г. Пустельникова, Л.С. Юденич

Институт высокопроизводительных вычислений и баз данных,  
Санкт-Петербург, Россия

## 1 Введение

Изучение эмбриогенеза имеет первостепенное значение для понимания молекулярных механизмов функционирования клеток и организмов и возникновения болезней, а также для разработки эффективных методов лечения различных болезней, начиная с гиперплазии и кончая дегенеративными болезнями. Быстрый прогресс в этой области науки привел к накоплению огромного объема информации, которая однако носит разобщенный характер и содержится в небольших специализированных базах данных, индивидуальных архивах и больших базах данных геномов отдельных организмов. С целью объединения этой информации мы приступили к созданию цифровой библиотеки AbOvo. В данной работе мы предполагаем обсудить структуру этой электронной библиотеки, типы содержащихся в ней данных, а также способы доступа к хранящейся информации.

## 2 Структура электронной библиотеки AbOvo

В настоящее время электронная библиотека AbOvo обеспечивает доступ к трем специализированным базам данных GeNet [1], UrchiNet и FlyEx, разработанным в лаборатории Биоинформационных систем Института Высокопроизводительных вычислений и баз данных, а также к ряду электронных ресурсов сети Интернет - к виртуальной библиотеке, поддерживаемой Обществом биологии развития, и к базам данных геномов организмов. GeNet - это полнотекстовая база данных о генных сетях, контролирующих раннее эмбриональное развитие у модельного биологического объекта - плодовой мушки дрозофилы. UrchiNet - это полнотекстовая база данных, содержащая информацию о механизмах эмбриогенеза у другого модельного биологического объекта - морского ежа. FlyEx - это реляционная база данных, содержащая количественные данные об экспрессии генов плодовой мушки дрозофилы, контролирующих ранние стадии эмбриогенеза. Виртуальная библиотека Общества биологии развития обеспечивает доступ к информации, содержащейся в индивидуальных архивах и специализированных

базах данных по эмбриогенезу, таких, как The Interactive fly (<http://sdb.bio.purdue.edu/fly/aimain/1aahome.htm>), FlyNets ([http://gifts.univmrs.fr/FlyNets.FlyNets\\_home\\_page.html](http://gifts.univmrs.fr/FlyNets.FlyNets_home_page.html)), HoxPro (<http://academic.mssm.edu/molbio/hoxpro/>) и т.д. Разрабатываемое нами средство E-gene позволит осуществлять поиск в больших базах данных геномов организмов информации об ортологах генов, данные о которых содержатся в базах данных GeNet, UrchiNet, и FlyEx.

AbOvo доступна по адресу  
<http://www.csa.ru/AbOvo/index.html>

Зеркальная копия некоторых компонент AbOvo, а именно базы данных GeNet, установлена на сервере Школы медицины Маунт-Синай (США) по адресу [www.mssm.edu/molbio/genet/genet.htm](http://www.mssm.edu/molbio/genet/genet.htm).

## 3 Типы данных

AbOvo содержит различные типы данных, соответствующие 3-м взаимосвязанным уровням организации эмбриона: молекулярному, клеточному и организма в целом. Информация на молекулярном уровне представлена данными о генных сетях, данными о компонентах генных сетей (генах, регуляторных районах, белках и т.д.) и об их молекулярных взаимодействиях, а также цифровыми изображениями картин экспрессии генов сетей и количественными данными об экспрессии этих генов. Информация на клеточном уровне представлена данными о клеточных линиях и сигналах, определяющих раннюю спецификацию эмбриональных территорий. Информация на уровне эмбриона в целом представлена данными об эмбриональных территориях, тканях и органах эмбриона, а также трехмерными моделями эмбрионов. По своей сути эти данные представляют собою текстовую информацию, цифровые изображения, карты, анимацию, и численные данные. Рассмотрим некоторые типы данных более подробно.

### 3.1 Карты генных сетей

Для представления карт генных сетей использована технология JAVA апплетов ([http://www.csa.ru:81/Inst/gorb\\_dep/inbios/genet/Urchin/Petri/s\\_purp\\_w.htm](http://www.csa.ru:81/Inst/gorb_dep/inbios/genet/Urchin/Petri/s_purp_w.htm)) [2]. (рис.1)

Первая Всероссийская научная конференция  
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:  
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,  
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ  
19 - 21 октября 1999 г., Санкт-Петербург

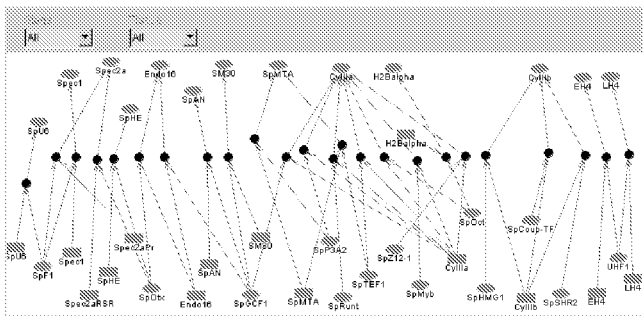


Рис. 1: Представление карт генных сетей с использованием технологии JAVA апплетов: карта генной сети морского ежа *Strongylocentrotus purpuratus*.

Компоненты генной сети - регуляторные районы генов, белки, молекулы РНК - изображены в виде фигур различной формы - прямоугольников, овалов, треугольников. Стрелки, изображающие взаимодействие компонент, направлены от молекул, участвующих во взаимодействии (или реакции) к молекулам, образующимся в результате взаимодействия элементов.

### 3.2 Данные о механизмах взаимодействия элементов генной сети

Информация о механизмах взаимодействия элементов геной сети представлена данными 3-х типов: данными о молекулах, вступающих во взаимодействие (генах: белках, и т.п.), данными о регуляторных элементах генов ([http://www.csa.ru/Inst/gorb\\_dep/inbios/genet/Urchin/u2span.htm](http://www.csa.ru/Inst/gorb_dep/inbios/genet/Urchin/u2span.htm)) и собственно данными о взаимодействиях.

**SpAN gene regulatory region**

LOCUS PR\_SPAN 421 bp INV 02-APR-1998

DEFINITION SpAN, serine-type protease, promoter

RESOURCE GENBANK: name 882594, accession 882594

KEYWORDS

SOURCE sea urchin 12-hr embryo.

ORGANISM Echinoidae  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Echinodermata;  
Echinozoa.

REFERENCE 1 (bases 1 to 421)

AUTHORS Kozlowski, D.J., Gagnon, M.L., Marchant, J.K., Reynolds, S.D., Angerer, L.M. and Angerer, R.C.

TITLE Characterization of a SpAN promoter sufficient to mediate correct spatial regulation along the animal-vegetal axis of the sea urchin embryo

JOURNAL Dev. Biol. 176 (1), 95-107 (1996)

MEDLINE 882594

REMARK GenNet staff at the INCBP created this entry from the original journal article. This sequence comes from Fig. 5 and 1.

COMMENTS Sequences within 300 bp upstream of the transcription start site are sufficient for high-level activity and correct non-vegetal expression of SpAN.CAT transgenes. However this region is not sufficient to shut down the transcription from the endogenous SpAN gene shortly after very

Рис. 2: Пример представления данных о регуляторном элементе гена

**SpP3A2 - CyIIIa interaction**

- REGULATOR: SpP3A2
- TARGET: CyIIIa
- INTERACTION TYPE: protein-DNA
- INTERACTION RESULT: repression
- TISSUE: oral ectoderm
- STAGE: gastrula
- TARGET SITE: (CT)N(C)TGGCC(A)T
- PROGS: whole-mount in situ hybridization of the fusion constructs
- REFERENCES: Kirchhamer and Davidson, 1996
- COMMENTS: This interaction accounts for the absence of CyIIIa expression in oral ectoderm. Interaction at two P3A sites of the proximal module are required to confine the CyIIIa expression to aboral ectoderm.

Рис. 3: Данные о взаимодействии элементов генной сети

Записи, посвященные взаимодействию элементов сети ([http://www.csa.ru/Inst/gorb\\_dep/inbios/genet/Urchin/u5pbcs.htm](http://www.csa.ru/Inst/gorb_dep/inbios/genet/Urchin/u5pbcs.htm)), содержат информацию о типе взаимодействия элементов сети, а также об экспериментальных данных, подтверждающих факт подобного взаимодействия.

### 3.3 Цифровые изображения картин экспрессии генов сети

AbOvo содержит данные о картинах экспрессии генов, контролирующих ранние этапы эмбриогенеза у плодовой мушки дрозофилы ([http://www.csa.ru/Inst/gorb\\_dep/inbios/genet/s7ruha.htm](http://www.csa.ru/Inst/gorb_dep/inbios/genet/s7ruha.htm)).

Эта информация представлена цифровыми изображениями локализации продуктов этих генов, полученными методом конфокальной микроскопии. В настоящее время AbOvo содержит около 1500 таких изображений.

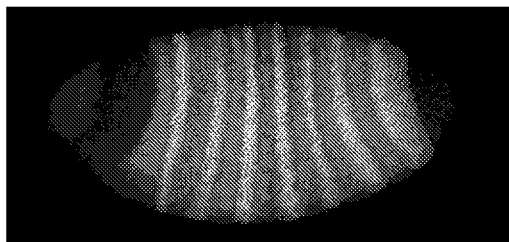
### 3.4 Количественные данные об экспрессии генов

AbOvo содержит количественные данные ([http://www.csa.ru/Inst/gorb\\_dep/inbios/genet/FlyEx/fig5.JPG](http://www.csa.ru/Inst/gorb_dep/inbios/genet/FlyEx/fig5.JPG)) об экспрессии генов, контролирующих ранние стадии эмбриогенеза у дрозофилы. Эти количественные данные были получены нами в сотрудничестве с доктором Рейнитцем из Школы медицины Маунт Синай в результате процессинга изображений картин экспрессии генов (см. раздел 2.3). Нами выполнена также реконструкция экспрессии этих генов как в различных частях эмбриона, так и во времени [3]. Эти данные также содержатся в AbOvo.

Апплет, представленный на рисунке позволяет отобразить графически количественные данные по экспрессии генов содержащиеся в AbOvo. При выборе конкретного гена апплет запускает на сервере cgi-script, который в свою очередь, обращается к AbOvo посредством



Gene expression patterns in *Drosophila*: runt and hairy



runt (green) and hairy (red) expression patterns at cleavage cycle 14.

- runt
- hairy
- gene interactions:
  - [run-hair genetic network](#)
  - [hairy regulatory connectives](#)

Seed/Dr: [Drosophila melanogaster](#) Gene expression patterns in *Drosophila*

Рис. 4: Изображения картин экспрессии генов, контролирующих ранние этапы эмбриогенеза дрозофилы, полученные методом конфокальной микроскопии.

SQL. Панель на которую выводится информация и панель управления разработаны с помощью классов пакета java.awt. и java.awt.Graphics.

Для удобства работы с графиками функций имеющих множественные особенности реализована функция масштабирования. Апплет позволяет строить до 10 графиков на одной панели. При добавлении и удалении графиков с панели автоматически меняется легенда, позволяющая установить соответствие между рядом данных и цветом графика. При позиционировании курсора в точке панели можно просмотреть координаты этой точки.

### 3.5 3-х мерные модели строения эмбриона

В настоящее время с помощью экспериментальных методов детально прослежены ранние стадии эмбрионального развития ряда организмов. Эта информация позволяет осуществить компьютерную реконструкцию организации эмбрионов на ранних стадиях развития. Следует подчеркнуть, что на ранних стадиях развития происходят решающие процессы морфогенеза - детерминация осей строения организма и судьбы клеток, а также закладка зародышевых слоев. Целью создания таких 3-х мерных моделей является визуализация этой информации и ее представление в удобной для пользователя форме.

В качестве примера рассмотрим 3-хмерную модель эмбриона морского ежа на стадии 400клеточной бластулы (что соответствует по времени 20 часам - с момента оплодотворения). Эта модель ([http://www.csa.ru/Inst/gorb\\_dep/inbios/genet/Urchin/u0blast.htm](http://www.csa.ru/Inst/gorb_dep/inbios/genet/Urchin/u0blast.htm))

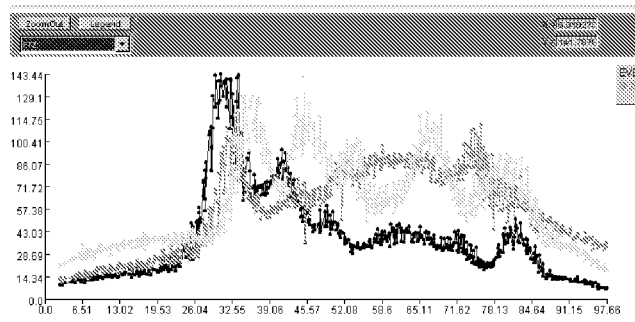
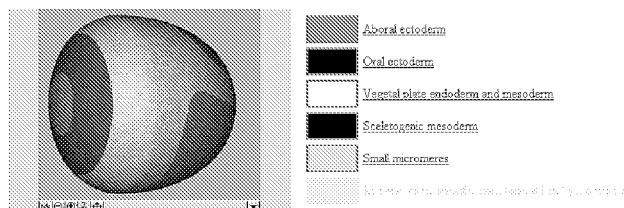


Рис. 5: JAVA апплет, представляющий результат пространственной реконструкции экспрессии генов Bcd, eve и ftz плодовой мушки дрозофилы в 14-м цикле дробления ядер эмбриона.

### Specification of territories in blastula-stage sea urchin embryo

Proposed fate map of territories in *Strongylocentrotus purpuratus* embryo, projected as an external image of the hatched blastula of about 400 cells (~20 hours postfertilization) (Davidson et al., 1998)



Use mouse and arrow keys (up and down) to navigate the model.

The model is designed by [Andrei Gorb](#), [Sergey Gorbunov](#) of [CSA](#).

Рис. 6: 3-х мерная модель 400-клеточного зародыша морского ежа на стадии бластулы ( 20 часов развития)

создана средствами пакета трехмерной графики Maya фирмы Alias/ Wavefront. Средствами этого же пакета была выполнена визуализация модели в виде последовательности нескольких кадров, каждый из которых позволяет наблюдать модель с разных сторон. Полученная последовательность кадров была сконvertирована в объект QuickTime VR. В цифровой библиотеке AbOvo данная модель служит для удобной визуализации данных об эмбриональных территориях морского ежа.

## 4 Способы доступа к информации

### 4.1 Пользовательские интерфейсы

Разработанные нами пользовательские интерфейсы базируются в основном на технике Java апплетов [2]. что позволяет запустить их на любой платформе, содержащей Sun Java VM. При разработке использовалась IBM VisualAge for Java 2.0 и JDK 1.1.7. Такие интерфейсы обеспечивают богатые возможности для взаимодействия пользователя с информационной системой. В качестве примера укажем интерфейс к информации об эмбриональных территориях, клеточных линиях и клетках эмбриона морского ежа на стадии 60-и клеток ([http://www.csa.ru/Inst/gorb\\_dep/inbios/genet/Urchin/Model/viewer/viewer.html](http://www.csa.ru/Inst/gorb_dep/inbios/genet/Urchin/Model/viewer/viewer.html))

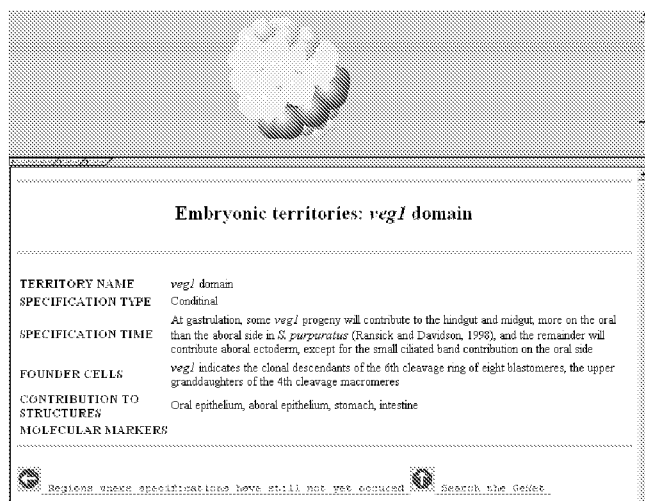


Рис. 7: Пример пользовательского интерфейса электронной библиотеки AbOvo, обеспечивающего доступ к информации об эмбриональных территориях, клеточных линиях и отдельных клетках эмбриона.

Система состоит из двух независимых частей - конструктора и просмотрщика (вьюера) моделей. Общими элементами являются понятия атома (одного шарика) и модели - набора атомов различных типов с фиксированным взаимным расположением. При этом каждый атом представляет собой просто плоскую картинку трехмерного шарика. Он "знает" только собственный цвет и "умеет" нарисовать себя в заданной точке и заданного размера. Вся информация о трехмерных координатах атомов хранится в модели. Модель - главная часть, ответственная за создание иллюзии трехмерности. Здесь также хранятся функции по добавлению и удалению атомов, работе с группами и обратной связи с пользователем. Кроме того, для вьюера нарисованы производные классы, содержащие элементы, необходимые для работы с URL и интернетом. Вся отрисовка графики сделана стандартными средствами Java (java.awt.Graphics, java.awt.Image), без привлечения Java3D или связок с OpenGL или VRML. Это упрощает аплет и повышает его мобильность, хотя на больших моделях скорость отрисовки оставляет желать лучшего. Интерфейсная часть написана с использованием AWT. При создании активно использовался Visual Composition Editor.

Отметим, что важной особенностью этого интерфейса является способность последовательной детализации получаемой пользователем информации.

## 4.2 Способы поиска информации

В настоящее время разрабатывается система поиска информации по всему массиву данных электронной библиотеки. На данный момент в AbOvo реализованы две различные поисковые системы:

Первая позволяет проводить поиск по ключевым словам в базах данных GeNet и UrchiNet. Алгоритм [5] производит поиск по всему массиву данных электронной библиотеки и выдает ответ в виде html-файла. Для осуществления поиска пользователь в начале выбирает базу данных и тип необходимой информации, а затем вводит ключевые слова. В результате обработки запроса программа генерирует html-файл с гипертекстовыми отсыл-

ками на те записи указанной базы, которые соответствуют заданным критериям поиска.

Вторая поисковая система предназначена для поиска в базе данных FlyEx и реализована средствами СУБД DB2, на платформе которой размещается данная база.

## 5 Планы по развитию электронной библиотеки

В дальнейшем мы планируем развивать электронную библиотеку AbOvo как предметно-ориентированную электронную библиотеку, обеспечивающую поддержку работы пользователя на большинстве этапов решения научной проблемы.

Для этого мы считаем необходимым:

- \* усовершенствовать механизмы поиска, обработки и интерпретации информации,

- \* совершенствовать способы представления и визуализации информации.

## 6 Выводы

Электронная библиотека AbOvo обеспечивает доступ к электронным ресурсам сети Интернет, содержащим информацию о механизмах эмбрионального развития.

Важной чертой этой электронной библиотеки является присутствие в ней информации о механизмах эмбриогенеза на 3-х взаимосвязанных уровнях организации эмбриона: молекулярном, клеточном и уровне организма в целом. Подобная структура электронной библиотеки позволяет пользователю получать исчерпывающую информацию о механизмах эмбрионального развития.

## Библиография

- [1] A.V.Spirov, M.G.Samsonova GeNet database as a tool for analysis of regulatory genetic networks (1998) Information Processing in Cells and Tissues, M.Holcombe and R.Paton (Eds) Plenum Press, 285-294.
- [2] V.N.Serov, A.V.Spirov, M.G.Samsonova Graphical interface to the genetic network database GeNet (1998) Bioinformatics, 14, 6, 546-547
- [3] E.M.Myasnikova, D.Kosman, J.Reinitz and M.Samsonova Spatio-temporal registration of the expression patterns of Drosophila segmentation genes (1999) Proceedings of the ISMB 99, in press.
- [4] M.G.Samsonova, V.N.Serov NetWork: an interactive interface to the tools for the analysis of genetic network structure and dynamics (1999) Proceedings of Pacific Symposium on Biocomputing, 4, 102-111
- [5] V.N.Serov, E.G.Savostyanova, S.Yu.Surkova, O.V.Kirillova, A.V.Spirov, J.Reinitz, M.G.Samsonova A database of genetic networks controlling embryogenesis. CSTB Bulletin, Spring 1998. <http://www.mar.dfo-mpo.gc.ca/science/mesd/he/science/cstb/papers/serov>

## Список авторов доклада

М.Г.Самсонова, зав. лаб., ИВВиБД,  
samson@fn.csa.ru

С.Ю.Суркова, научн. сотр., ИВВиБД,  
surkova@fn.csa.ru

В.Н.Серов, старший программист, ИВВиБД,  
serov@fn.csa.ru

К.А.Соболев, стажер-исследователь, ИВВиБД

Р.М.Шулаков, ИВВиБД, rshul@fn.csa.ru

А.В.Веремьев, ведущий программист, ИВВиБД,  
erg@fn.csa.ru

Е.Г.Пустельникова, системный аналитик, ИВВиБД,  
cawa@fn.csa.ru

Л.С.Юденич, научн. сотр., ИВВиБД,  
judenich@fn.csa.ru