



Исследование функциональных характеристик генома (1)

Катохин Алексей Вадимович, к.б.н.

Кафедра информационной биологии ФЕН НГУ



Функции генома состоят в том, чтобы содержать в виде кодов или сигналов генетическую информацию:

- для обеспечения сохранения этой информации в процессах жизнедеятельности одной клетки в течение одного клеточного цикла и в процессе пролиферации клеток при передаче через митоз (репликация, репарация, упаковка в хроматин, и т.д.);
- для обеспечения сохранения качества этой информации в ряду поколений особей при мейозе (сегрегация хромосом, рекомбинация);
- для обеспечения развертывания генетических программ развития, т.е. определять условия для формирования транскриптома, предопределенного генетическими программами развития для клеток определенных типов, т.е. обеспечение дифференциальной экспрессии генов в результате процессов транскрипции и сплайсинга.



Транскриптом – совокупность всех транскриптов всех генов, экспрессирующихся в какой-либо клетке, или группе однотипных клеток – т.е. ткани, или во всех клетках организма в определенные моменты функционирования и/или развития. Транскриптом – первый уровень фенотипа, т.е. первый уровень развертывания и реализации генетической информации, заключенной в геноме. Исследование транскриптома – одна из задач функциональной геномики и основная задача транскриптомики.

Задачи транскриптомики - исследование:

- (1) структуры транскриптов
- (2) их дифференциального временного и пространственного распределения в клеточных типах или тканях.

Применение методов биоинформатики в транскриптомике позволяет:

- реконструировать структуру генома и коды, заключенные в геноме (взаимодействие с геномикой)
- выявлять информацию в виде сигналов и кодов, необходимых для формирования протеома (взаимодействие с протеомикой)



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома



Требования к методам исследования структуры транскриптов и их дифференциального временного и пространственного распределения в клетках и организмах:

- возможность измерения относительного и абсолютного содержания транскриптов определенного гена в клетках разных типов, т.е. возможность сравнения результатов разных экспериментов при разных модификациях методов;
- возможность одновременного измерения для как можно большего количества генов соотношения транскриптов
- возможность детекции транскриптов очень слабо или очень специализированно экспрессирующихся генов (чувствительность и достаточно широкий динамический диапазон)
- высокая производительность и эффективность для производства достаточно большого массива данных.



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома



Характеристика методов детекции и оценки количества транскриптов

метод	измерение относит. и абсолют., сравнимость	количество генов	чувствительность и динамический диапазон	производительность и эффективность
Прямая детекция мРНК гибридизацией транскриптов с одноцеп. ДНК- или РНК-пробами				
Нозерн-блот гибридизация	+	(1-5) x (5- 20)	- -	- -
Защита от рибонуклеазы	++	(10-15) x (5- 20)	+	- -
Детекция кДНК, полученной ОТ-ПЦР из мРНК				
Количественная ОТ-ПЦР, ОТ-ПЦР с дет. в реал. времени	++	десятки	++	++
Дифференц. дисплей	+	десятки	++	++
Олиго/кДНК микрочипы	++	тысячи	+++	+++
Компьютерный анализ сиквенсов кДНК , « <i>in silico</i> гибридизация» транскриптов				
EST	+	тысячи	+++	+++
SAGE	+++	тысячи	+++	+++
MPSS	++++	тысячи	++++	++++



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



Высокопродуктивные и высокоинформативные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)
- Серийный анализ экспрессии генов (Serial analysis of gene expression - SAGE)
- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)
- Экспрессионные ДНК-биочипы (Expression microarrays, biochips)



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Прочитанные фрагменты экспрессированных последовательностей
(Expressed Sequence Tags - ESTs)

Мощный метод для глобального компьютерного анализа структуры транскриптов, реконструкции структуры генов и измерения уровней экспрессии генов.

Предложен:

Принцип - Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science, 252:1651-1656

Количественное приложение - Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nature Genetics, 2:173-179



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)



- Прочитанные фрагменты экспрессированных последовательностей
(Expressed Sequence Tags - ESTs)

ESTs - это короткие (обычно около 300-500 bp), прочитанные за один раз (single-pass reads) фрагменты кДНК.

Они представляют собой «слепок», «отпечаток» (snapshot) с транскриптома, проэкспрессированного в определенной ткани или на определенной стадии развития. Они являются метками (tags) экспрессии генов для определенной библиотеки кДНК.

Основные принципы метода ESTs:

- анализ совокупности прочитанных EST, ассоциированных с каким-либо геном, позволяет реконструировать структуру транскриптов гена, а после сравнения ее со структурой соответствующего геномного района – структуру самого гена – генную модель;
- для каждой библиотеки кДНК частотное распределение кДНК клонов в целом соответствует исходному распределению транскриптов в популяции мРНК, из которой была приготовлена библиотека;
- чтобы получить достоверную информацию о структуре генов и паттерне его экспрессии EST должно быть получено очень много;
- получение многочисленных EST должно быть относительно дешевым.



Исследование функциональных характеристик генома

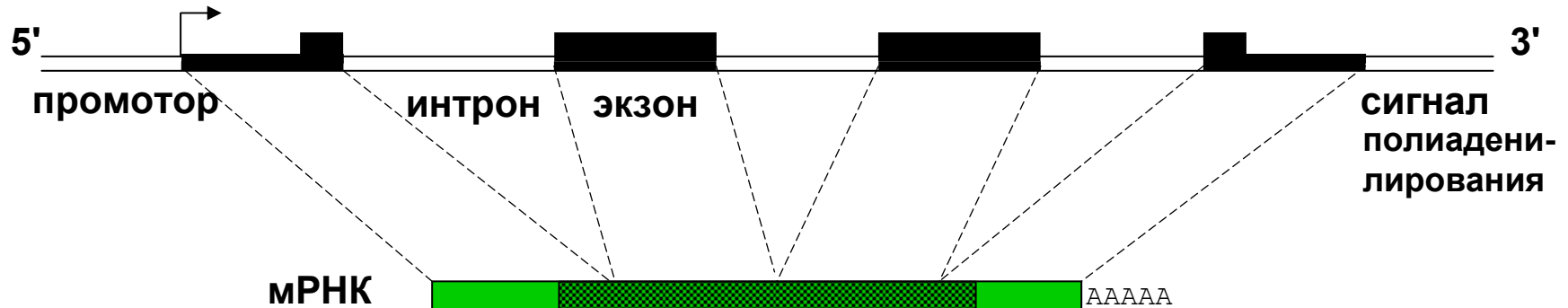
Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

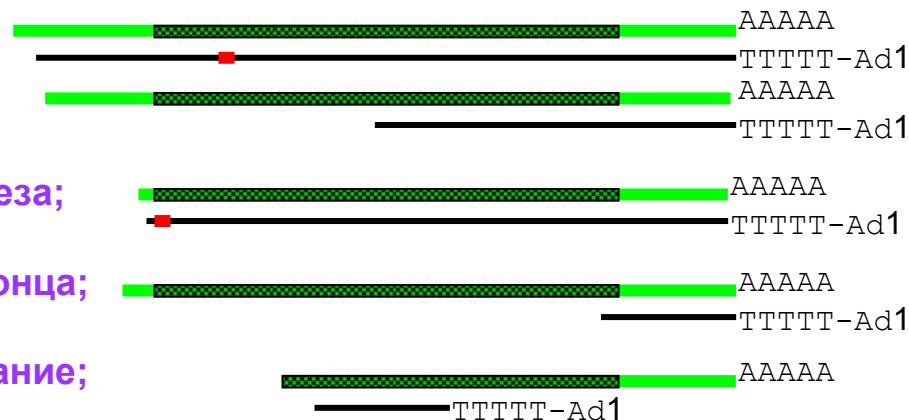
А. Экспериментальная фаза. Приготовление кДНК библиотеки



СИНТЕЗ кДНК ЦЕПИ

Источники отличий от исходных природных молекул:

- случайный обрыв синтеза;
- деградация мРНК с 5'-конца;
- внутреннее праймирование;



Источники ошибок:

-ошибки обратной транскриптазы



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

А. Экспериментальная фаза. Приготовление кДНК библиотеки

Подготовка к клонированию

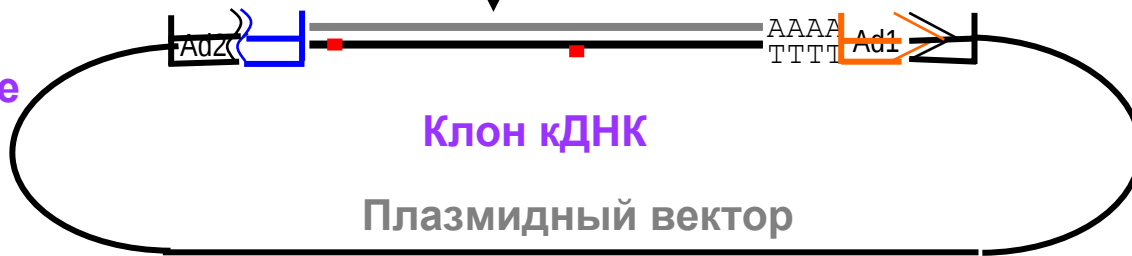
- присоединение Ad2



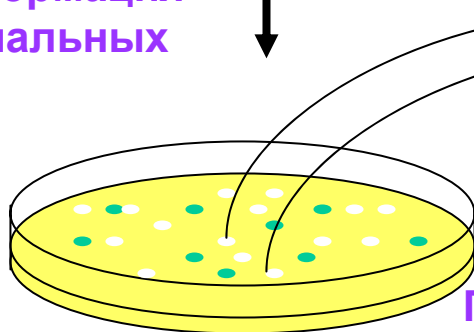
- замещение РНК второй цепью кДНК



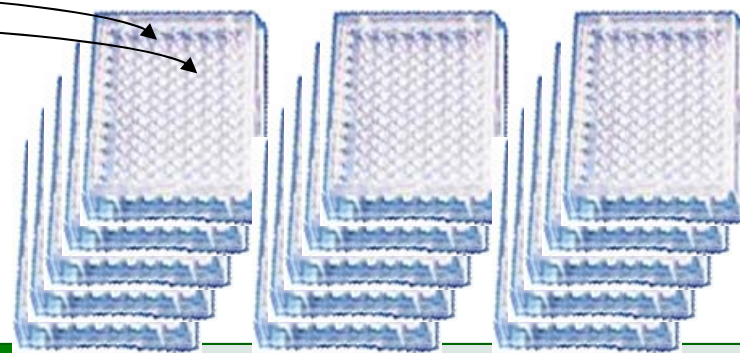
Клонирование кДНК



Трансформация бактериальных клеток



Пересев клонов



кДНК библиотека

Источники отличий от исходных природных молекул:

- образование артефактных химерных молекул;

Источники ошибок:
- ошибки Taq полимеразы



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



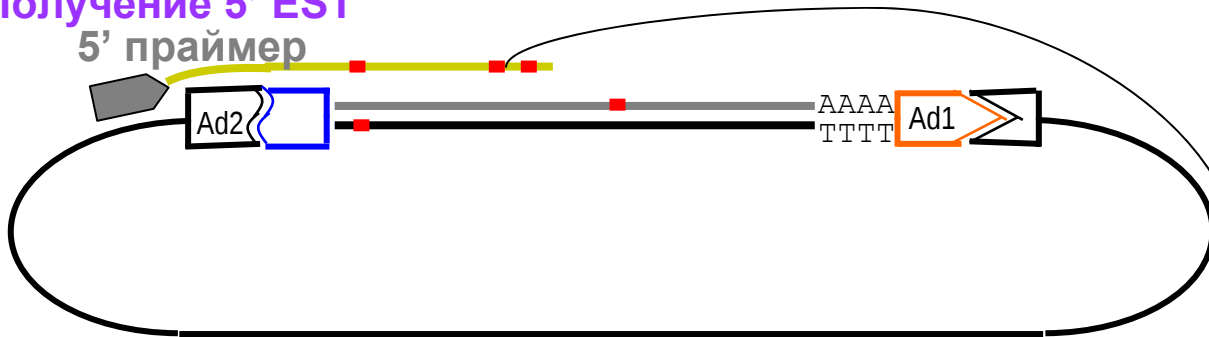
- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

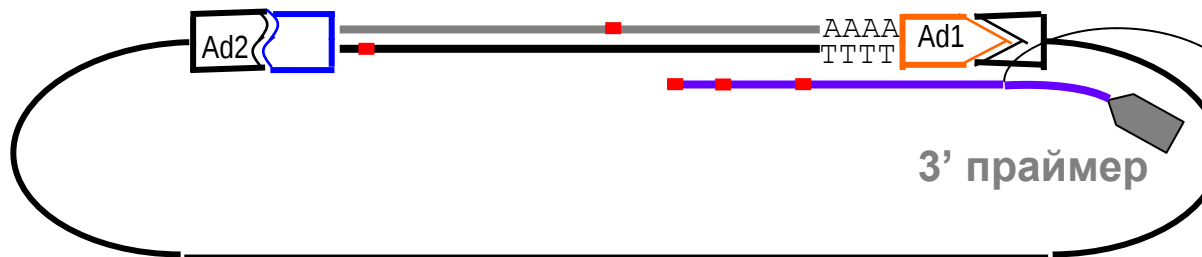
А. Экспериментальная фаза. Приготовление кДНК библиотеки

Использование специфических праймеров к сайтам вектора, окаймляющим встройку кДНК

Получение 5' EST



Получение 3' EST



Источники ошибок:

- ошибки Taq полимеразы или секвеназы
- ошибка в обозначении идентификатора клона (id_clone identification error)
- ошибка в обозначении направления секвенирования клона (5'/3' identification error)
- смещение дорожек (виртуальные химеры) (lane slipping)

В подраздел GenBank – dbEST (или EMBL bank)



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



• Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

dbEST: database of "Expressed Sequence Tags"

dbEST release 111105

November 11, 2005

Homo sapiens (human)	7,057,754
Mus musculus + domesticus (mouse)	4,688,047
Xenopus tropicalis	1,038,272
Rattus sp. (rat)	704,494
Bos taurus (cattle)	702,645
Danio rerio (zebrafish)	689,581
Ciona intestinalis	686,396
Zea mays (maize)	656,945
Triticum aestivum (wheat)	600,039
Gallus gallus (chicken)	578,445
Sus scrofa (pig)	502,501
Xenopus laevis (African clawed frog)	473,792
Arabidopsis thaliana (thale cress)	420,789
Oryza sativa (rice)	406,790
Hordeum vulgare + subsp. vulgare (barley)	395,019
Drosophila melanogaster (fruit fly)	383,407
Glycine max (soybean)	355,978
Canis familiaris (dog)	349,306
Pinus taeda (loblolly pine)	329,469
Caenorhabditis elegans (nematode)	302,080

dbEST release 111706

November 17, 2006

Homo sapiens (human)	7,895,572
Mus musculus + domesticus (mouse)	4,722,069
Oryza sativa (rice)	1,211,064
Zea mays (maize)	1,160,485
Danio rerio (zebrafish)	1,152,269
Bos taurus (cattle)	1,141,099
Xenopus tropicalis	1,039,143
Rattus norvegicus + sp. (rat)	871,144
Triticum aestivum (wheat)	855,067
Arabidopsis thaliana (thale cress)	734,275
Ciona intestinalis	686,396
Sus scrofa (pig)	640,034
Gallus gallus (chicken)	599,171
Xenopus laevis (African clawed frog)	542,288
Drosophila melanogaster (fruit fly)	514,613
Hordeum vulgare + subsp. vulgare (barley)	437,321
Salmo salar (Atlantic salmon)	428,803
Canis familiaris (dog)	365,909
Glycine max (soybean)	359,402
Caenorhabditis elegans (nematode)	346,064
Pinus taeda (loblolly pine)	329,469
Vitis vinifera (wine grape)	316,756
Oryzias latipes (Japanese medaka)	309,868



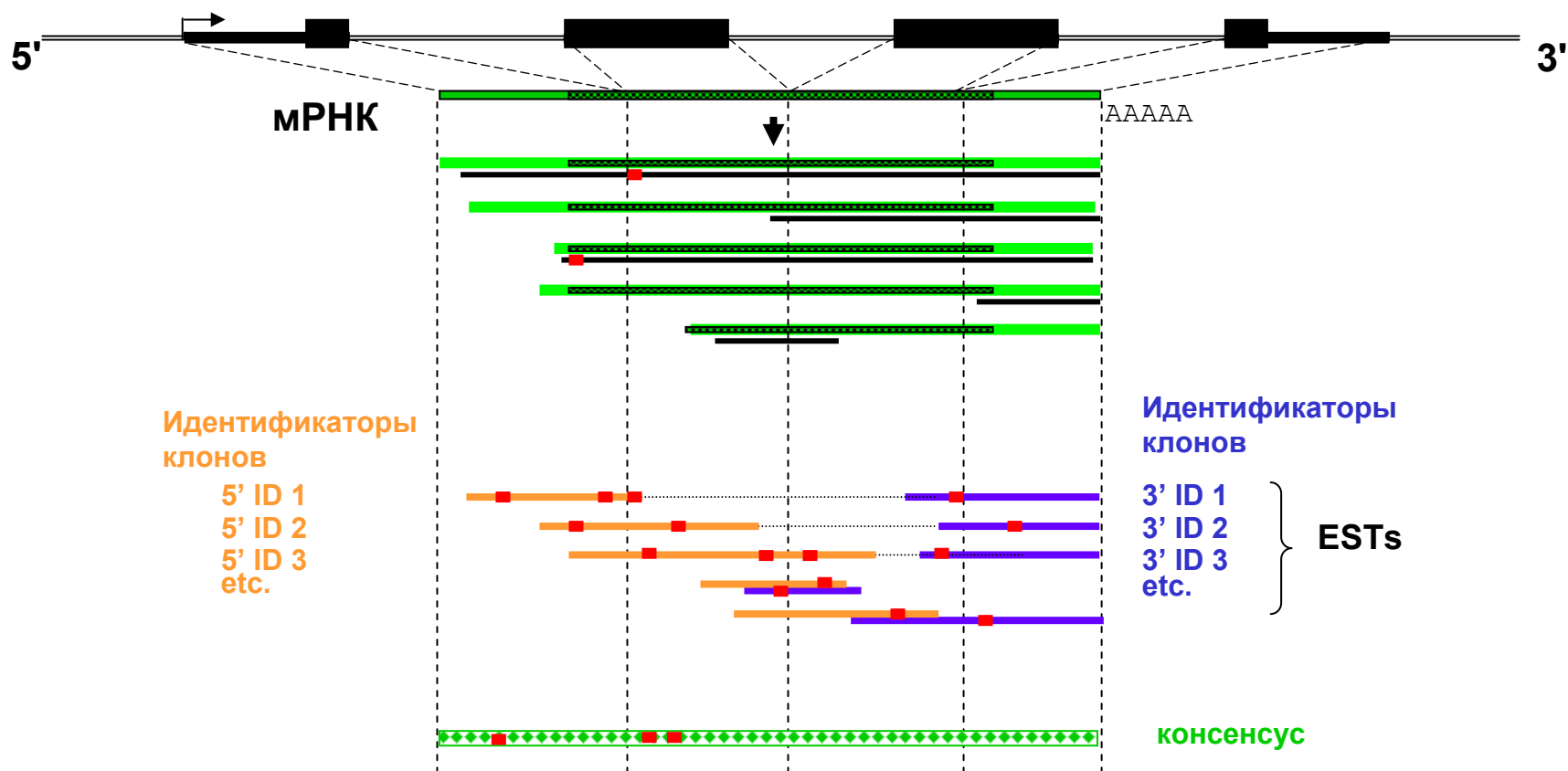
Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

Б. Компьютерная фаза. Выравнивание и кластеризация EST





Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

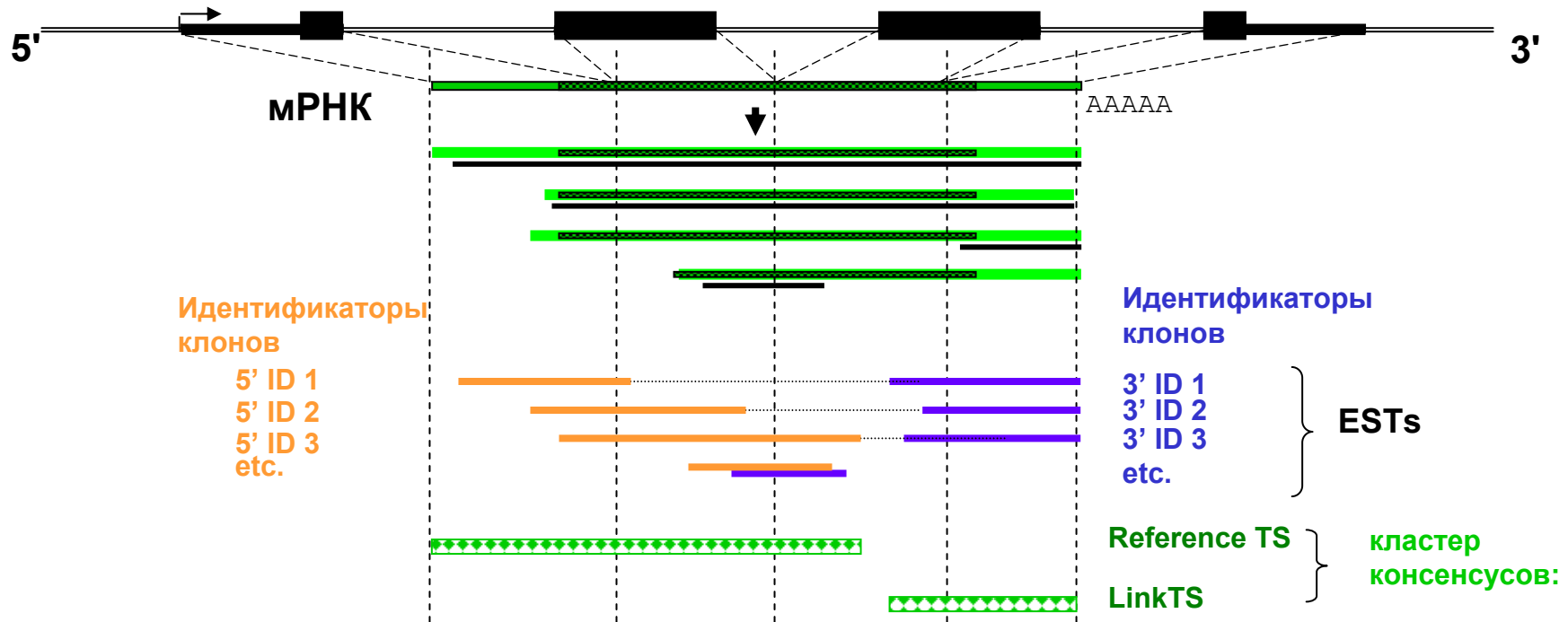


- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

Б. Компьютерная фаза.

Б1. Выравнивание и кластеризация EST



Недостаточное «перекрытие» всей мРНК посредством EST приводит к появлению т.н. LinkTS, т.е. сборок, в которых есть хоть одна EST, идентификатор клона которой совпадает с таковым для какой-либо EST, входящей в основную (ReferenceTS) сборку для гена.



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



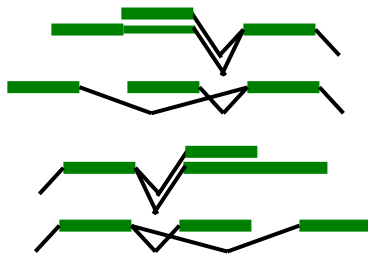
- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

Б. Компьютерная фаза

Б2. Анализ структуры кластера

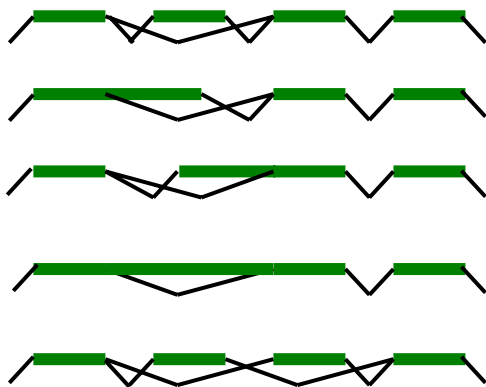
Альтернативная транскрипция



- альтернативные старты

- альтернативные сигналы полиаденилирования

Альтернативный сплайсинг внутренних экзонов



- пропуск экзона (exon skipping)

- удлинение экзона в 3'-направлении (exon 3'-extention)

- удлинение экзона в 5'-направлении (exon 5'-extention)

- удержание интрона (intron retention) или несплайсированная незрелая форма транскрипта (premature mRNA)

- альтернативное включение экзонов (alt. exon usage, cassette exons)



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



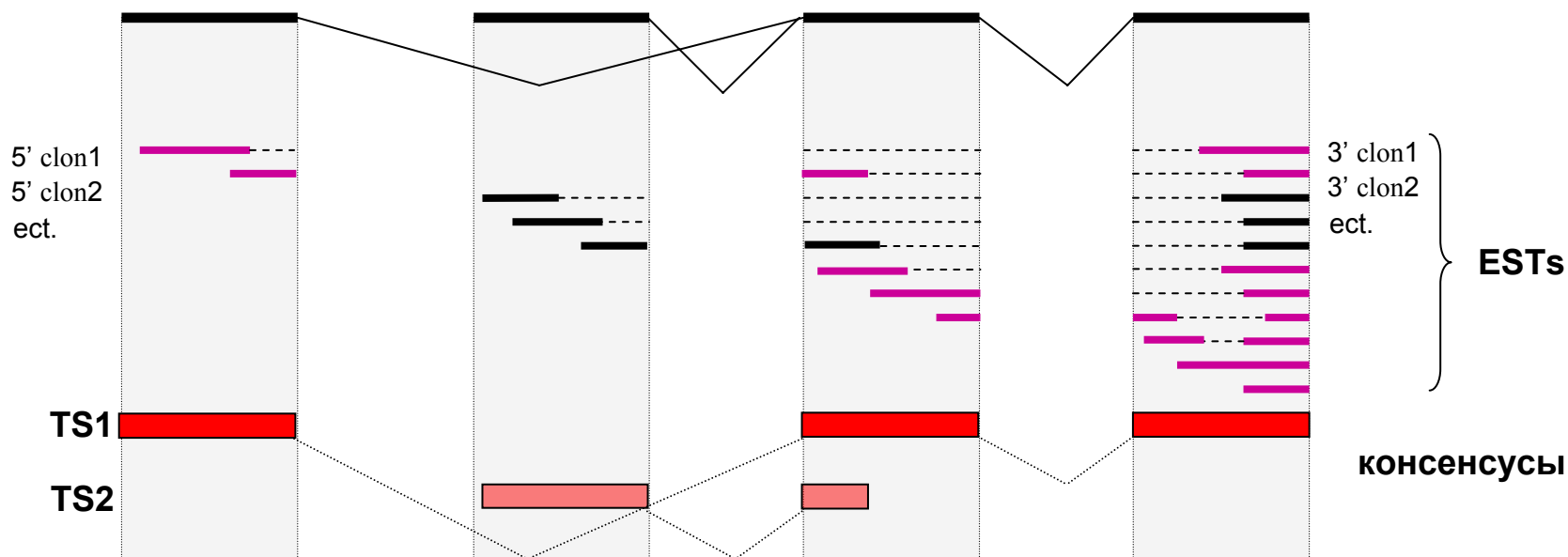
- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

Б. Компьютерная фаза

Б2. Анализ структуры кластера

АЛЬТЕРНАТИВНЫЕ СТАРТЫ ТРАНСКРИПЦИИ В АЛЬТЕРНАТИВНЫХ 5'-ЭКЗОНАХ





Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

Б. Компьютерная фаза Б2. Анализ структуры кластера

Alternative Splice Forms from EST and mRNA Sequences ?

Retrieve Reset Form Operating System :

Max. Number of Items returned : ☐ 10 ☐ 100 ☐ 1000

Search for ...

Current List :

Integrated Genetic Map Service Last updated on August 01, 2002

<http://medseq.bioinf.mdc-berlin.de/imap/splicelib/>

<http://devnull.lbl.gov:8888/alt/>

Alternative Splicing DB

[DB CONTENT](#) | [HOW TO USE](#) | [FURTHER WORK](#) | [SEARCH](#)



References to the Alternative Splicing Database:

ASDB: database of alternatively spliced genes

I. Dralyuk, M. Brudno, M. S. Gelfand, M. Zorn, and I. Dubchak (2000) *Nucleic Acids Research* 28(1), 296-297.

M. S. Gelfand, I. Dubchak, I. Dralyuk and M. Zorn (1999) *Nucleic Acids Research*, 27(1), 301.



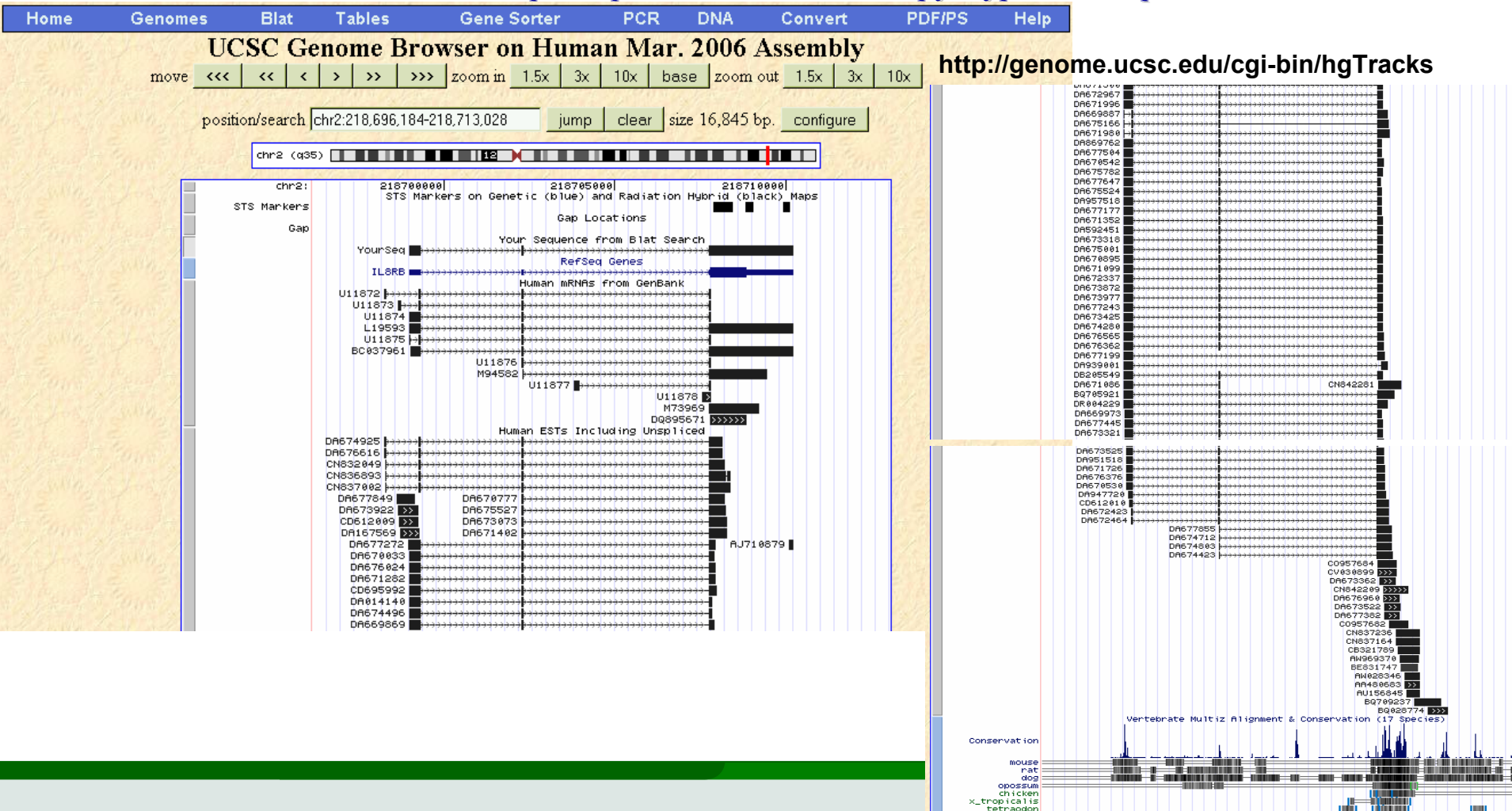
Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

Б. Компьютерная фаза Б2. Анализ структуры кластера





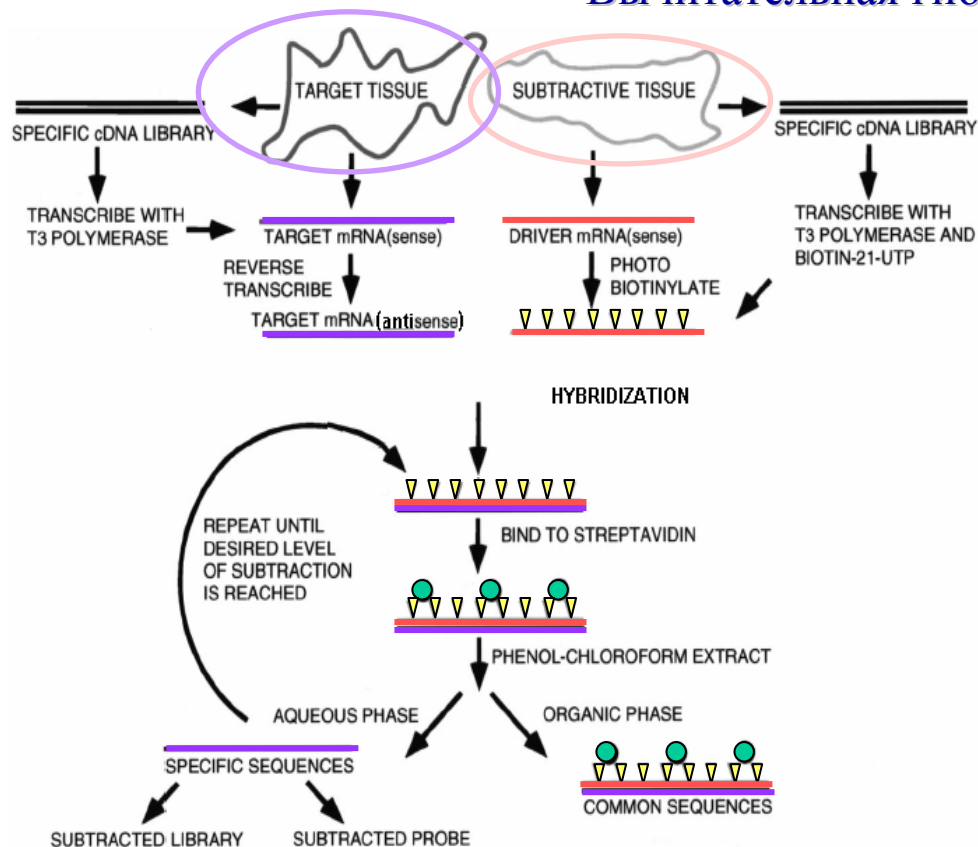
Исследование качественных и количественных характеристик транскриптома



- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

Особенности приготовления библиотек клонов кДНК

Вычитательная гибридизация



Позволяет выявлять очень редкие ткане- или стадия-специфичные транскрипты,

но изменяет соотношение между транскриптами



Исследование качественных и количественных характеристик транскриптома



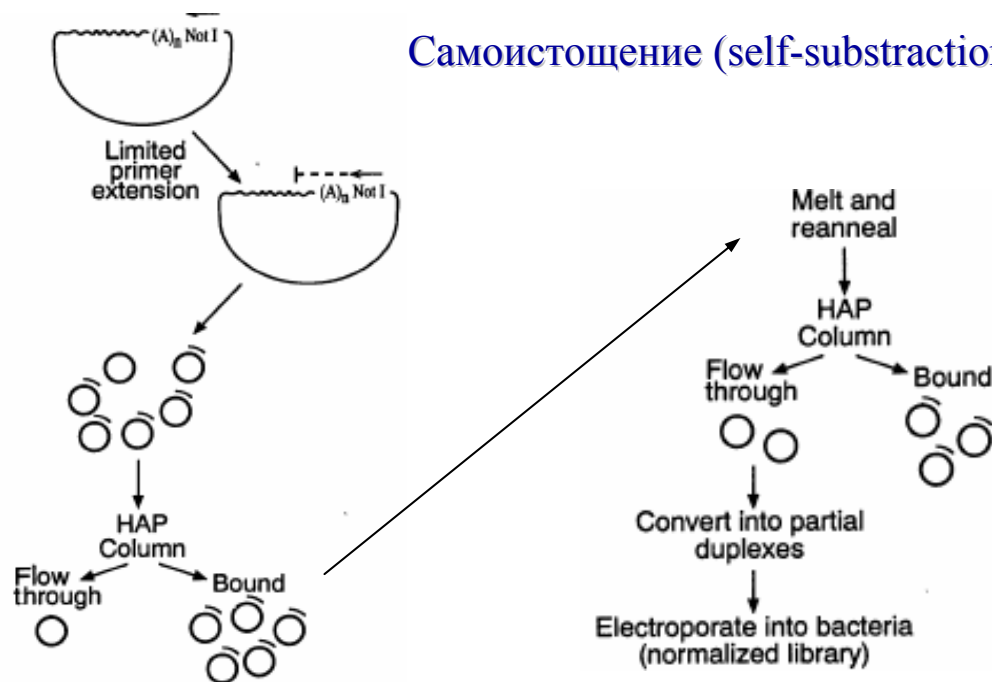
- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)
Особенности приготовления библиотек клонов кДНК

Нормированные библиотеки кДНК клонов

Самоистощение (self-substraction)

Позволяет выявлять очень редкие транскрипты,

но изменяет соотношение между транскриптами



Soares M.B. *et al.*, 1994,
Construction and characterization of a
normalized cDNA library. *Proc Natl
Acad Sci U S A.*;91(20):9228-32.



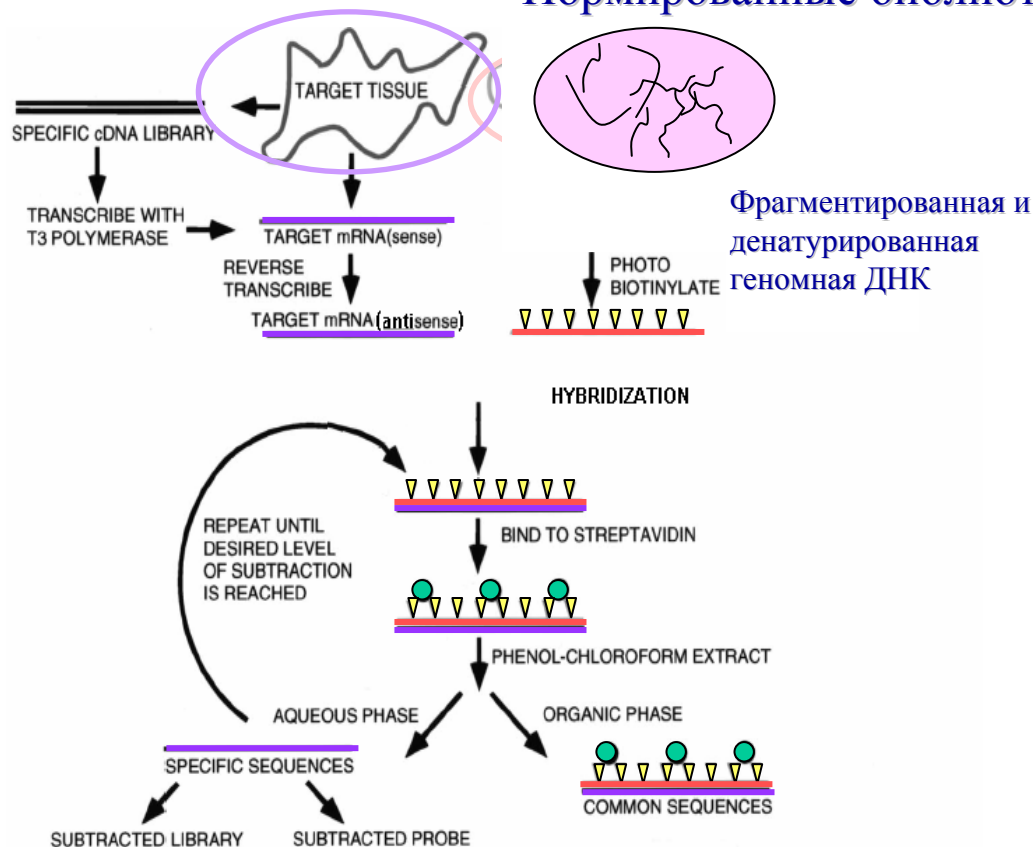
Исследование качественных и количественных характеристик транскриптома



- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

Особенности приготовления библиотек клонов кДНК

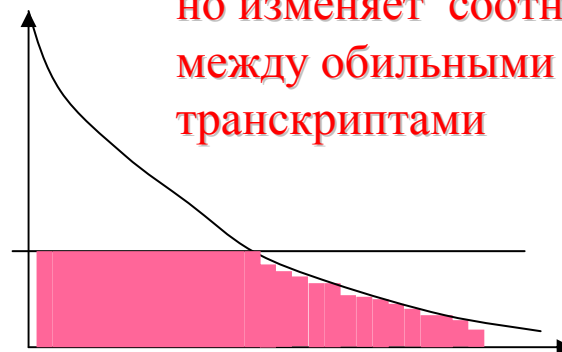
Нормированные библиотеки кДНК клонов

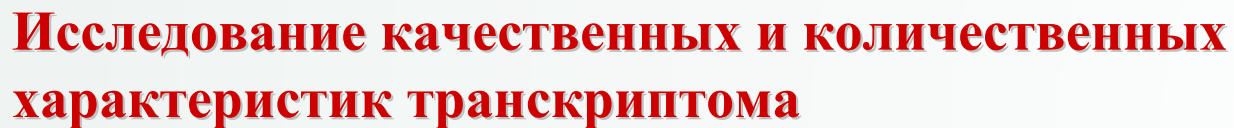


Истощение геномными последовательностями

Позволяет выявлять очень редкие транскрипты,

но изменяет соотношение между обильными и редкими транскриптами

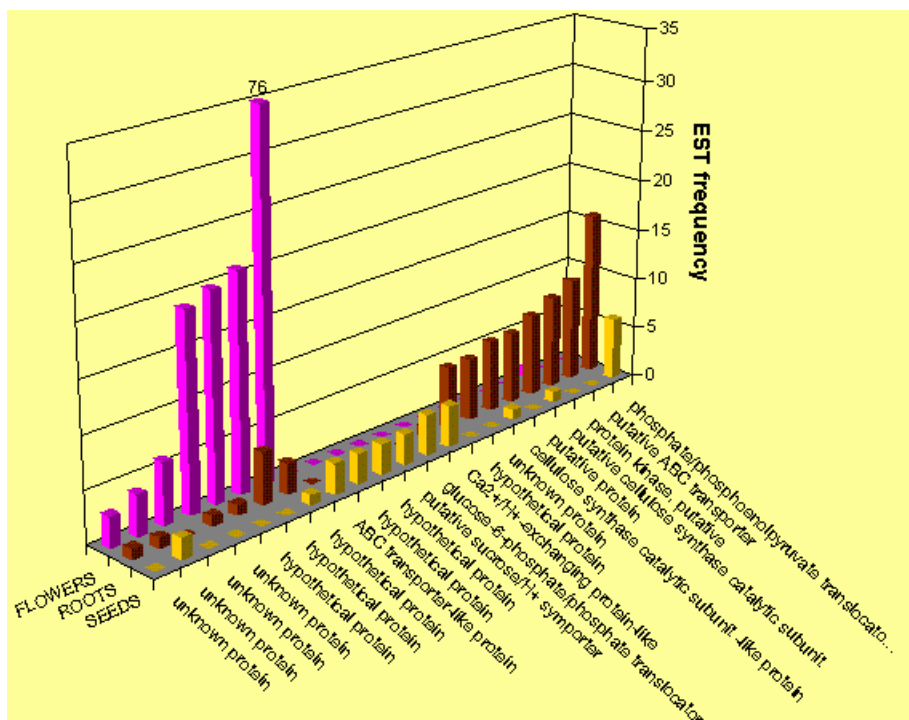




- ## Приложение для количественно оценки содержания транскриптов

Цифровой дифференциальный дисплей Digital Differential Display

Audic and Claverie
(<http://igs-server.cnrs-mrs.fr>)



Mégý K, Audic S, Claverie JM. 2002; Heart-specific genes revealed by expressed sequence tag (EST) sampling. *Genome Biol.* 3(12): research0074.1-research0074.11.



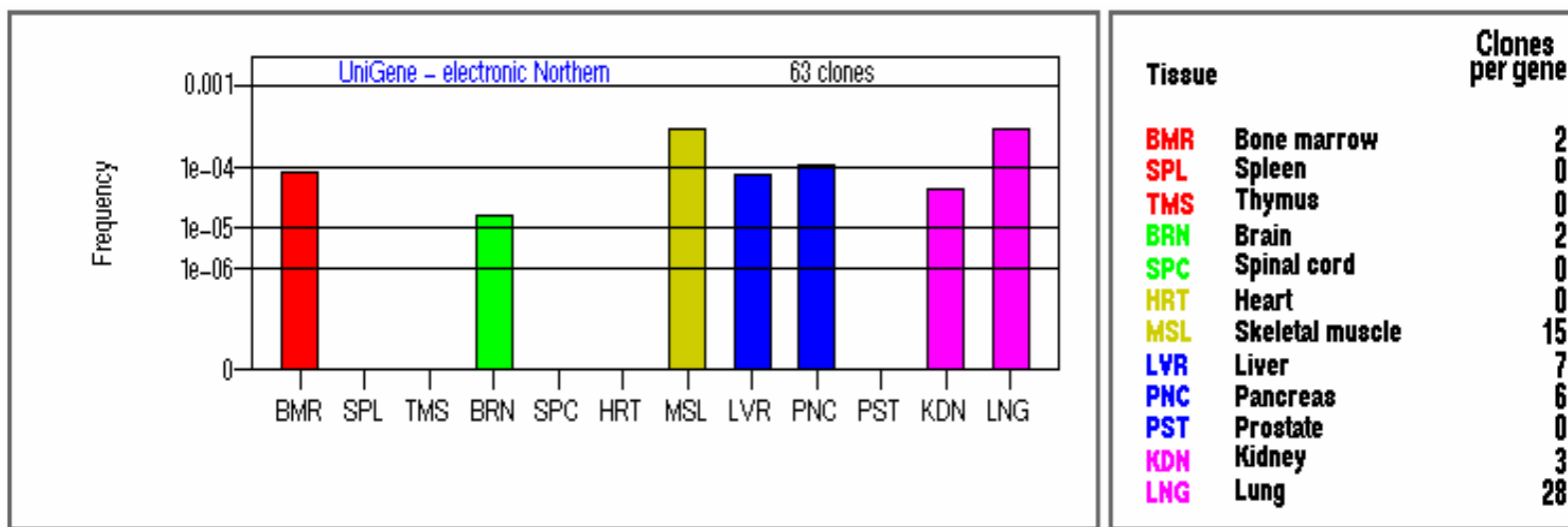
Исследование качественных и количественных характеристик транскриптома



- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

Приложение для количественно оценки содержания транскриптов

Электронный или цифровой Нозерн



NCBI

UniGene

ORGANISED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

NCBI

UniGene Homepage

Digital Differential Display (DDD)

Begin DDD for: Homo sapiens

<http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi?ORG=Hs>



Исследование качественных и количественных характеристик транскриптома



- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

Приложение для количественно оценки содержания транскриптов

Электронный или цифровой нозерн

[allgenes.org](http://www.allgenes.org)

[news](#) [statistics](#) [manip](#) [EST](#) [query](#) [boolean](#) [history](#)

Human DoTS Transcript DT.95345342

RNA last modified Feb 09, 2004 03:49:25PM

100% identity to 100% of Mannose-6-phosphate isomerase (Phosphomannose isomerase) (PMI) (Phosphohexomutase)

2194 bp RNA containing 61 sequence(s) review status=unreviewed

ON THIS
PAGE: [estAnat](#)

OTHER [Summary](#) | [Transcript features](#) | [Transcript seq](#) | [Assembly alignment](#) | [Input seqs](#) | [EST libs](#) | [EST anatomy](#) | [Motifs](#) | [NRDB](#)

VIEWS: [hits](#) | [Protein features](#) | [Protein seq](#) |

Source anatomy terms for ESTs in the assembly

Anatomy Term	Description	ESTs	Percentage of ESTs
brain	encephalon	10	19%
placenta		7	13%
cell line		2	3%
cell line	HeLa	2	3%
colon		3	5%
retina		3	5%
cell line		2	3%
cell line		2	3%
leukocyte	white blood cell	2	3%

<http://www.allgenes.org/allgenes/servlet?page=rna&id=DT.95345342&view=estAnat>



Исследование качественных и количественных характеристик транскриптома

- Прочитанные фрагменты экспрессированных последовательностей (Expressed Sequence Tags - ESTs)

Интернет-ресурсы

The EST Machine

<http://www.tigem.it/ESTmachine.html>

Directory of links to selected bioinformatic resources for Expressed Sequence Tags analysis

EST databases	I.M.A.G.E. Consortium / WashU-Merck-HHMI Project
• dbEST at NCBI (information on the current dbEST release) • Search dbEST using SRS6 at HGMP • UniGene at NCBI (<i>Homo sapiens</i>, <i>Mus musculus</i>, <i>Rattus norvegicus</i>). See also the SRS implementations at Infobiogen, HGMP, and Sanger • Library Browser with EST expression information from UniGene • Sequence Tag Alignment and Consensus Knowledgebase at SANBI • TIGR Database (TDB) and TIGR Human Gene Index (see also the TIGR Mouse Gene Index) • -The Drosophila EST project • C.elegans EST Database (Japan) • Muscle TRAIT homepage (University of Padua) • Plant EST Analysis: similarity analysis of different plant cDNA/EST sequences at U. Minnesota	• The I.M.A.G.E. Consortium homepage • WashU-Merck Human EST Project (cDNA Library and Vector Information) • The IMAGEne database (I.M.A.G.E. has clustered all their cDNA clones to all known human genes) • The Cancer Genome Anatomy Project (CGAP, the National Cancer Institute EST project) • Oncogenes and Tumor Suppressor genes currently found in dbEST • WashU-HHMI Mouse EST Project • WashU Zebrafish EST Project • View a Sequence Trace (either for a human or a mouse EST) • The LENS (Linking ESTs and their associated Name Space) database • I.M.A.G.E. Consortium Clone Distributors

<http://www.ncbi.nlm.nih.gov/dbEST/>

<http://www.tigem.it/ESTmachine.html>

<http://cgap.nci.nih.gov/>

http://industry.ebi.ac.uk/~mulu/EST/EST_links.html

<http://image.llnl.gov/>

<http://www.cs.jhu.edu/labs/compbio/morgan.html>

<http://hercules.tigem.it/ASSEMBLY/assemble.html>



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

Предложен Victor E. Velculescu в 1995
(Velculescu et al. Science 270 (5235) : 484-487)

Основные принципы метода SAGE:

1. Небольшой фрагмент из определенного места транскрипта достаточно информативен для идентификации этого транскрипта.
(10 п.о. $\Rightarrow 4^{10} = 1\,048\,576$ транскриптов)
2. Конкатенация и параллельный анализ повышают продуктивность
(секвенирование 96 колоний за один прогон $\Rightarrow 2400$ прочитанных тагов)
3. Минимизация связанных с амплификацией искажений в соотношении тагов
(ампликоны почти равны по размеру ~ 100 п.о. и составу – на 70% состоят из адаптеров)

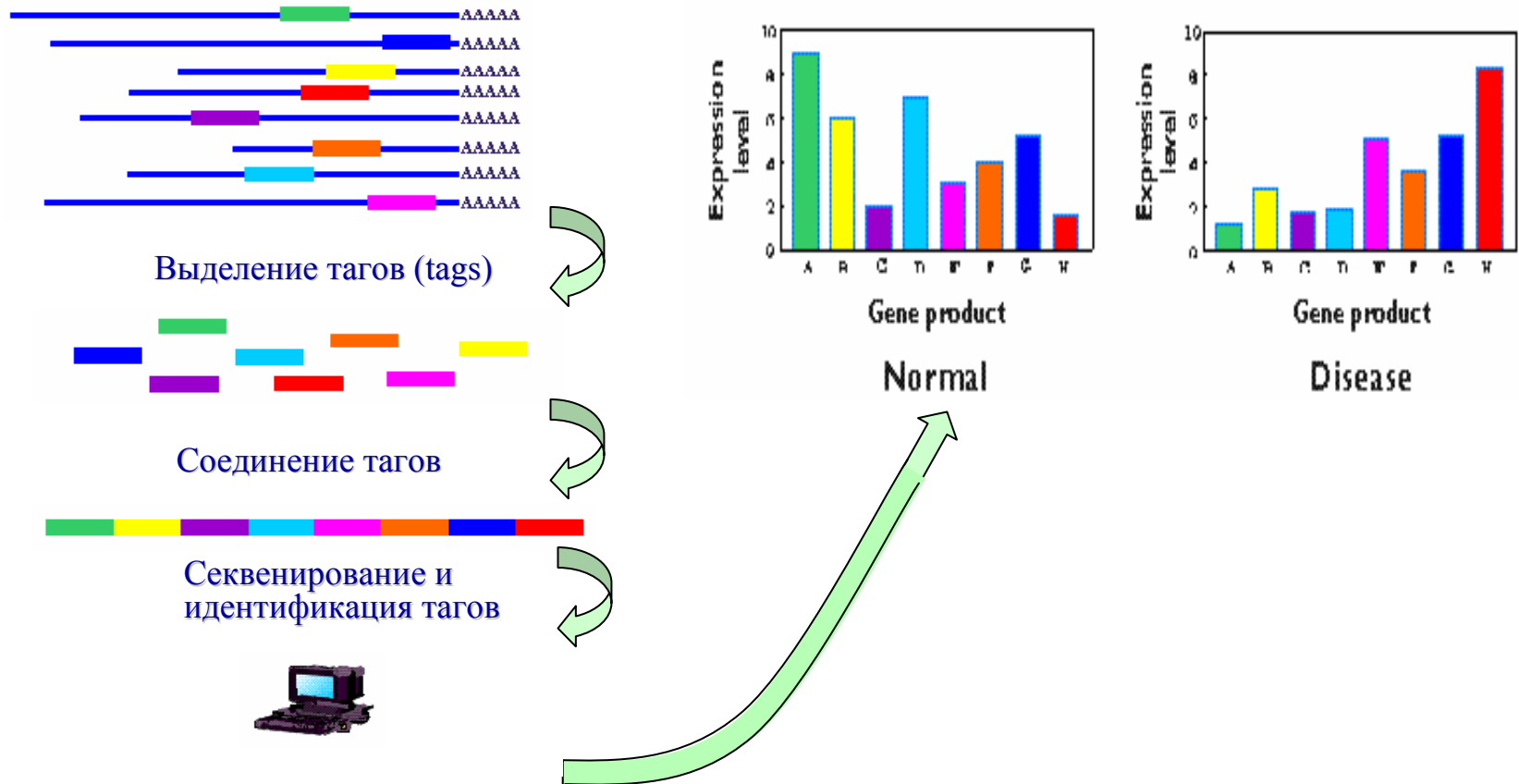


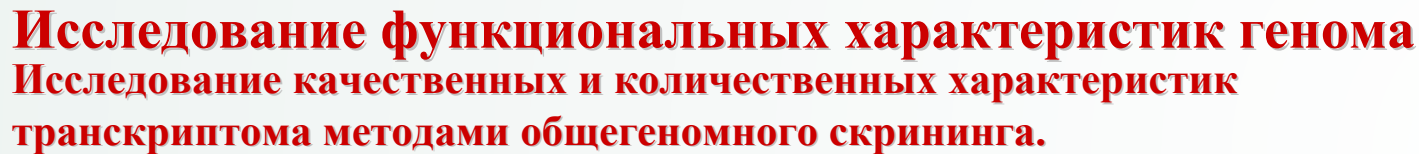
Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



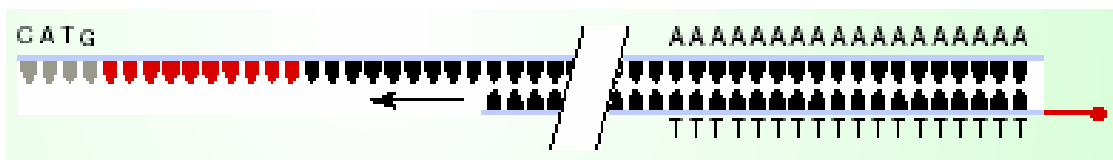
- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)



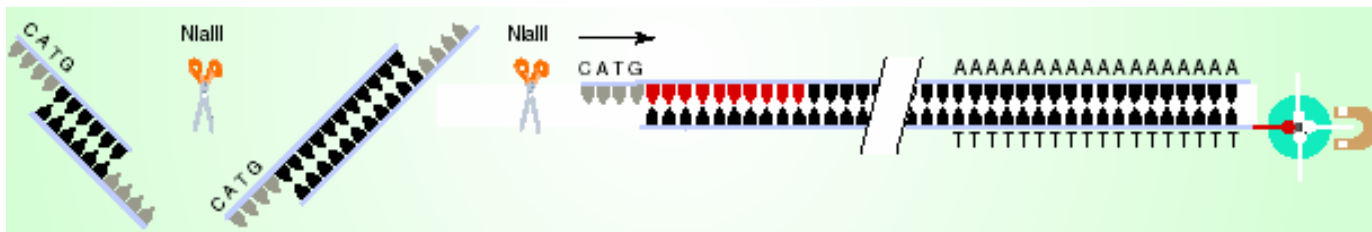


- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

1. Синтез биотинилированной двухцепочечной кДНК.



2. Рестрикция по NlaIII сайтам и отделение фрагментов, примыкающих к 3'-концу смысловой цепи кДНК.





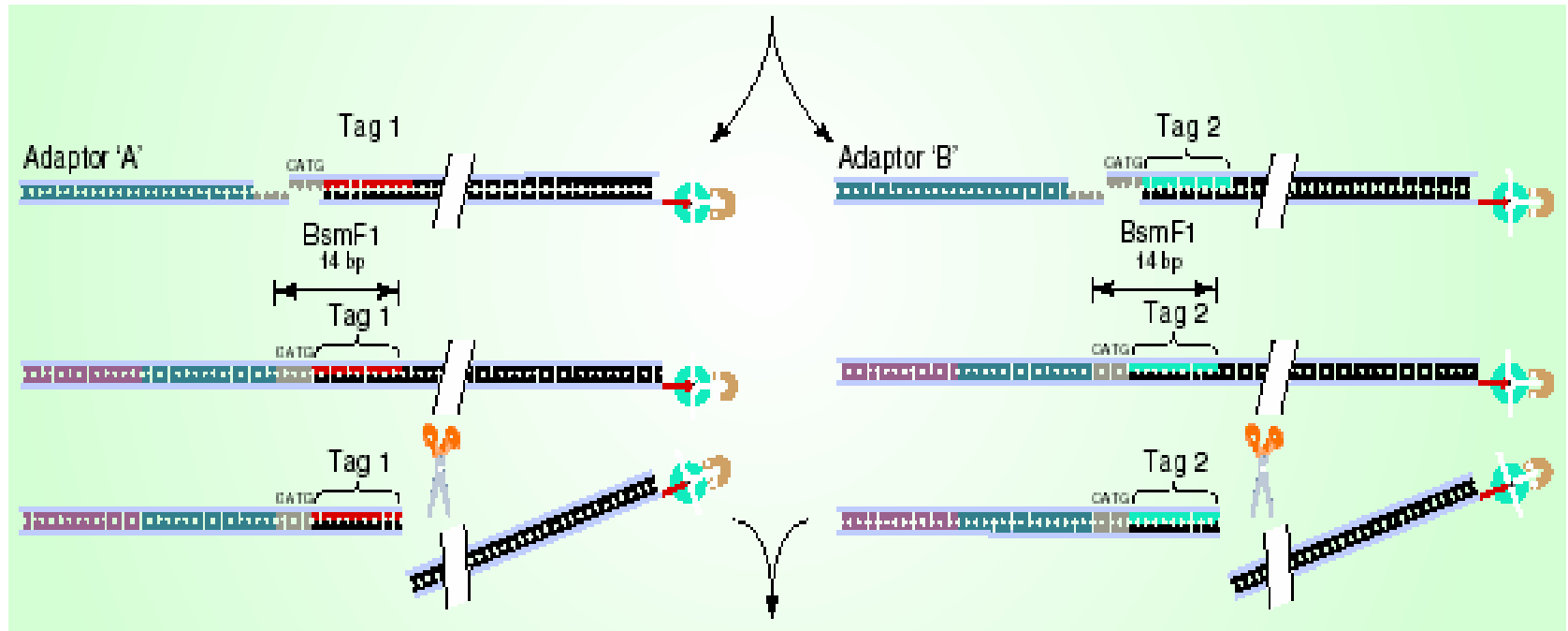
Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

3. Присоединение специфичных адаптеров и вырезание меток, тагов (tag), длиной 10 п.о. с помощью эндонуклеазы рестрикции *Pst* типа.





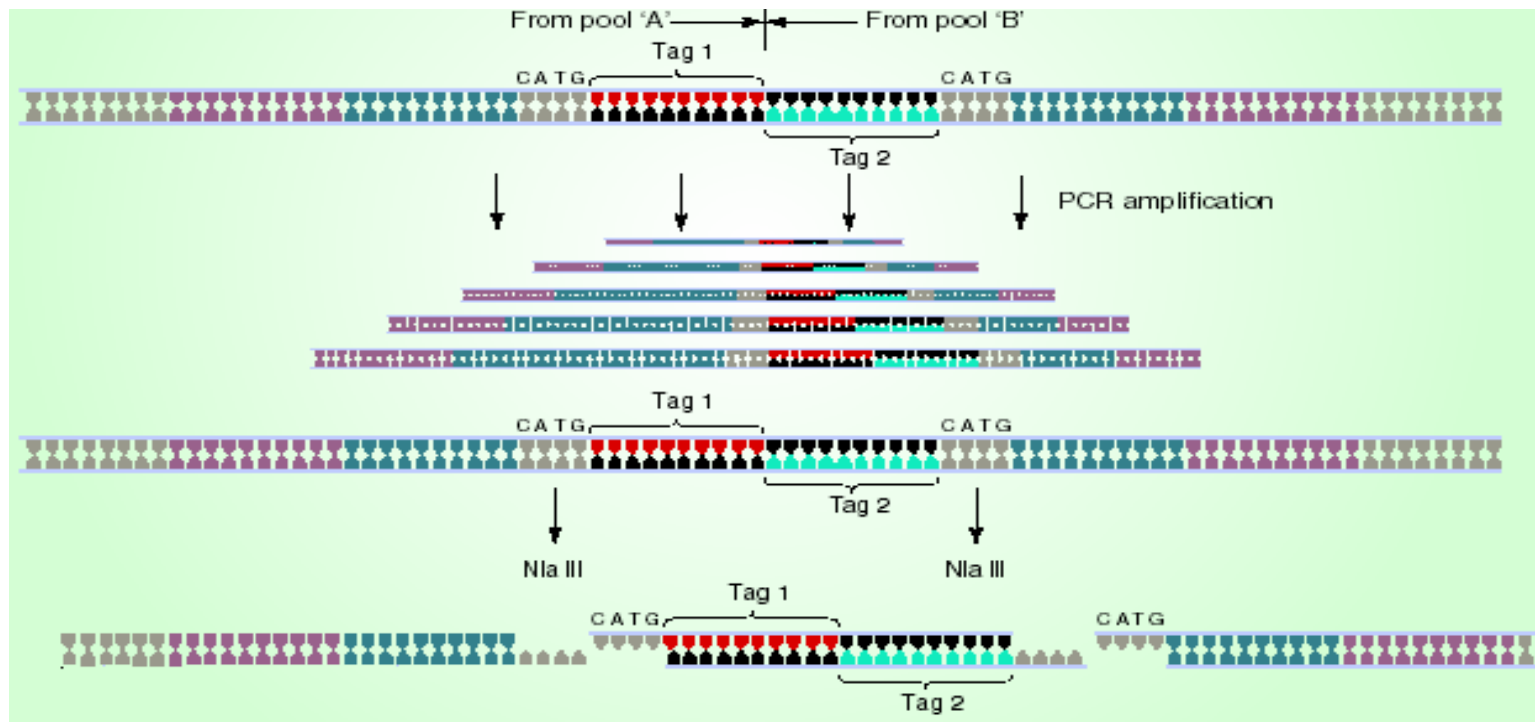
Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

4. Образование дитагов, их амплификация и удаление адаптеров.





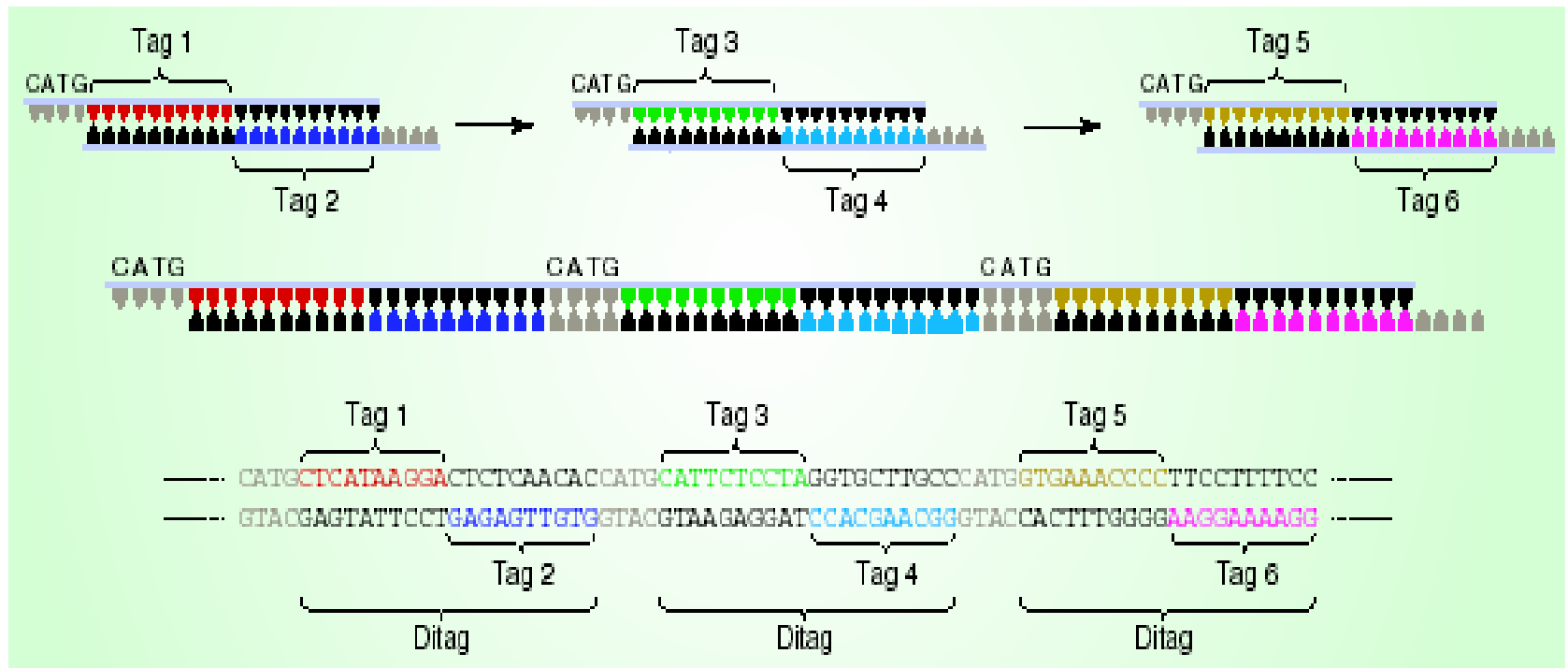
Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

5. Формирование конкатемеров и их секвенирование





Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

6. Вычисление уровня экспрессии транскрипта исходя из числа встречаемости определенного тага.

<u>SAGE tag</u>	<u>Tag Count (in 100 000 tags)</u>	<u>Absolute abundance</u>
<u>CATGGACGTCTTAAT</u>	 33 TAGS	0.033%
<u>CATGGTGACCTCCTT</u>	 63 TAGS	0.063%
<u>CATGTGAAGAGAAGA</u>	 22 TAGS	0.022%
<u>CATGAGTGGAGGTGG</u>	 9 TAGS	0.009%
<u>NlaIII</u> site		



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



• Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

Достоинства:

- высокая продуктивность
- возможность определить соотношение транскриптов тысяч генов
- почти отсутствует наведенные амплификацией количественные искажения
- возможность определять абсолютные значения уровней экспрессии генов
- возможность сравнивать результаты разных экспериментов
- возможность выявлять слабоэкспрессирующиеся гены
- возможность выявлять новые экспрессирующиеся гены,

Ограничения:

- зависимость от предварительного знания последовательности значительной части генов организма
- невыявление сплайсированных и других изоформ
- зависимость от качества синтеза первой цепи кДНК и от расстояния между сайтом рестрикции и полиА-трактом
- применимость только полиаденилированным транскриптам эукариот
- требует большое количество РНК и поэтому не позволяет профилировать экспрессию генов с высоким разрешением.

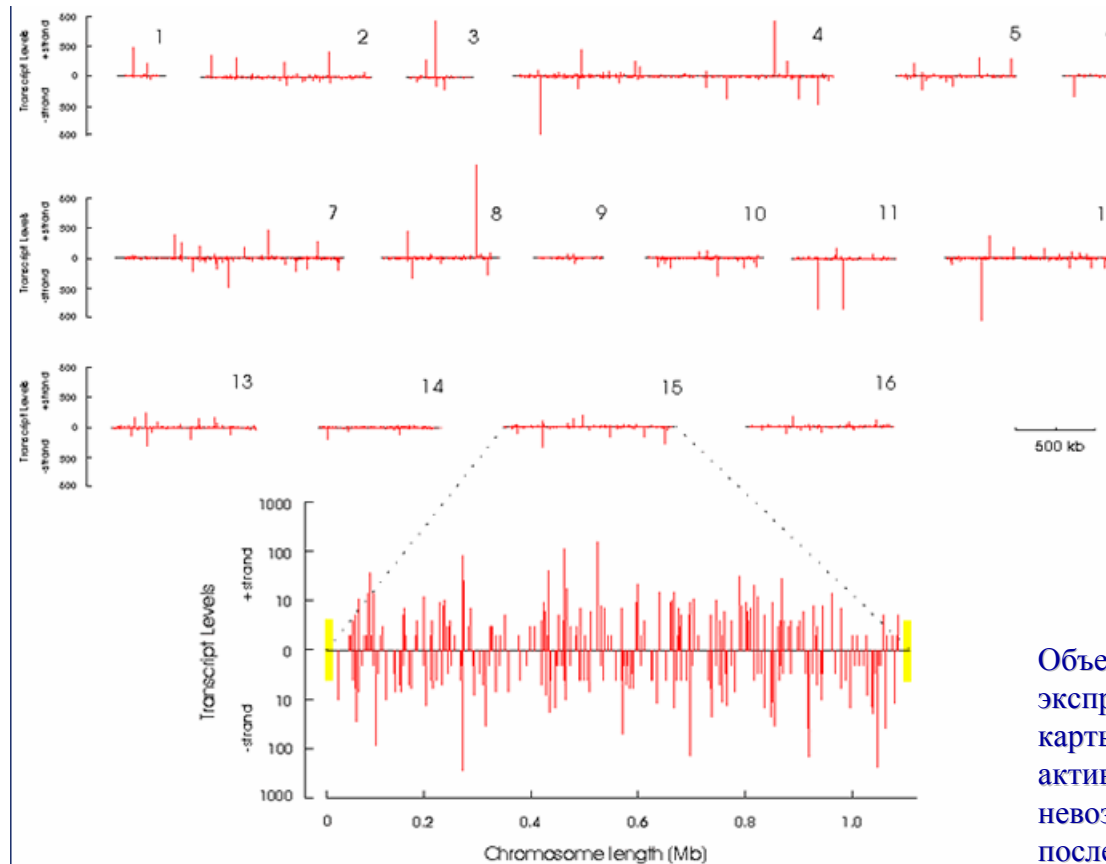


Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)
Пример приложения для общегеномного анализа



Velculescu V.E. *et al.*, 1997
Analysis of Yeast Transcriptome
// Cell 88: 243-251

Анализ транскриптов из дрожжевых клеток трех состояний (лог-фаза, задержанные в S- и G2/M- фазах), всего 60,633 тагов, выявил 4,665 генов с уровнями экспрессии от 0.3 до более 200 транскриптов на клетку.

Число SAGE-тагов определено исходя из требований достижения определенного уровня достоверности выявления редких мРНК. При оценке 15,000 молекул мРНК на клетку секвенирование 20,000 тагов должно дать 72% вероятности обнаружить одну молекулу мРНК на клетку.

Объединение позиционной информации с данными об экспрессии генов позволило построить хромосомные карты экспрессии с районами транскрипционной активности и выявить гены, существование которых было невозможно предсказать, исходя из знания только о последовательности геномного района.



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)
Развитие технологии



Plant Physiology 134:890-897 (2004)
© 2004 American Society of Plant Biologists

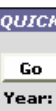
BREAKTHROUGH TECHNOLOGIES

Robust-LongSAGE (RL-SAGE): A Substantially Improved LongSAGE Method for Gene Discovery and Transcriptome Analysis¹

Malali Gowda, Chatchawan Jantasuriyarat, Ralph A. Dean and Guo-Liang Wang^{*}

Department of Plant Pathology, Ohio State University, Columbus, Ohio 43210 (M.G., C.J., G.-L.W.); and Fungal Genomics Laboratory, Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27695 (R.A.D.)

Serial analysis of gene expression (SAGE) is a widely used technique for large-scale transcriptome analysis in mammalian systems. Recently, a modified version called LongSAGE (S. Saha, A.B. Sparks, C. Rago, V. Akmaev, C.J.



This Article | Info for Authors | Subscribe | About

PNAS

Proceedings of the National Academy of Sciences of the United States of America

Note: Performing your original search, **LongSAGE**, in PubMed Central will retrieve **33 citations**.

Journal List > Proc Natl Acad Sci U S A > v.101(32); Aug 10, 2004

Proc Natl Acad Sci U S A. 2004 August 10; 101(32): 11701–11706.
Published online 2004 July 22. doi: 10.1073/pnas.0403514101.

Copyright © 2004, The National Academy of Sciences

Genetics

5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation

Chia-Lin Wei,^{*†} Patrick Ng,^{*†} Kuo Ping Chiu,^{*} Chee Hong Wong,[†] Chin Chin Ang,^{*} Leonard Lipovich,^{*} Edison T. Liu,^{*} and Yijun Ruan^{*§}

^{*}Genome Institute of Singapore, 60 Biopolis Street, Genome 02-01, Singapore 138672; and [†]Bioinformatics Institute, 30 Biopolis Street, Matrix 08-01, Singapore 138671

SADE

Serial microanalysis (<http://www-dsv.cea.fr/thema/get/sade.html>)
a **SAGE** Adaptation for **Downsized Extracts**



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)

Интернет-ресурсы

SAGE
Serial Analysis of Gene Expression

ligation to a restriction enzyme cleavage site
PCR amplification
sequencing

FINDINGS RESOURCES PUBLICATIONS CONFERENCES CONTACT US

What's New?

Serial analysis of gene expression (SAGE) is a powerful tool that allows the analysis of overall gene expression patterns with digital analysis. Because SAGE does not require a preexisting clone, it can be used to identify and quantitate new genes as well as known genes.

[Learn more...](#)

<http://www.sagenet.org/>

Resources

SAGE Data

Human Colorectal Tissue

Library Name	Description	Tags Analyzed
NC1	Normal colon epithelium	51,637
NC2	Normal colon epithelium	52,938
CACO2	CACO2 colon cancer cells	64,110
HCT116	HCT116 colon cancer cells	59,572
SW837	SW837 colon cancer cells	65,158
RKO	RKO colon cancer cells	56,972
TU102	primary colon cancer	62,104
TU98	primary colon cancer	46,571
AD-A	adenoma A	52,801
AD-B	adenoma B	42,844

Human Pancreatic Cancer

Library Name	Description	Tags Analyzed
PL45	PL45 pancreatic cancer cells	31,886
ASPC	ASPC pancreatic cancer cells	32,300
91-16113	Primary pancreatic cancer	33,180
96-6252	Primary pancreatic cancer	34,416

Mouse Fibroblast

Library Name	Description	Tags Analyzed
3T3	3T3 cells (mouse)	28,531

Yeast

Library Name	Description	Tags Analyzed
Yasyn	Log phase yeast	20,832
YHU	Hydroxyurea treated yeast	21,360
Ynoco	Nocodazole treated yeast	20,773



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Серийный анализ экспрессии генов - Serial analysis of gene expression (SAGE)
Интернет-ресурсы



Serial Analysis of Gene Expression
Tag to Gene Mapping



HOMESEARCHSITE MAP

SAGE DatasetsSAGEmap PaperEmail SAGEmap

NCBI > SAGE

In order to support the public use and dissemination of serial analysis of gene expression (SAGE) data, NCBI has recently refurbished this website. SAGEmap is a SAGE data resource for the query and retrieval and analysis of SAGE data from any organism. All of the data present on this website has been accessioned in the [Gene Expression Omnibus repository](#).

SAGEmap is undergoing restructuring and revised tag to gene mappings. Our current mappings are not up to date.
We apologize for the inconvenience and hope to be up again soon.

Recent news

May 18, 2005
A new utility is now available to map the user's SAGE library to gene identifiers and UniGenes. To utilize this tool, follow the "[Annotate your library](#)" link on the SAGEmap main page.

Site contents

Information

- Briefs
- About mappings
- Current holdings
- Retrieving data
- GEO
- SAGEmap paper

Retrieve

- ...by tag
- ...by sequence
- ...by gene

<http://www.ncbi.nlm.nih.gov/SAGE/>



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)

Предложен Brenner S, Johnson M, Bridgham J, et al.,
(Gene expression analysis by massively parallel signature sequencing
(MPSS) on microbead arrays. Nat Biotechnol. 2000; 18(6):630-634.)

MPSS-технология позволяет идентифицировать почти все кДНК-фрагменты в образце.

- Продуцирует количественный профиль экспрессии генов на основе подсчета всех мРНК в образце, и эти цифровые данные удобны для создания реляционных баз данных
- Обладает высокой чувствительностью, способностью детектировать редкие транскрипты. В каждый MPSS-анализ вовлечены более миллиона транскриптов, что обеспечивает исключительный динамический диапазон: от менее чем 10 транскриптов на миллион до 50,000 транскриптов на миллион.
- Не нуждается в предварительных молекулярно-генетических знаниях об организме

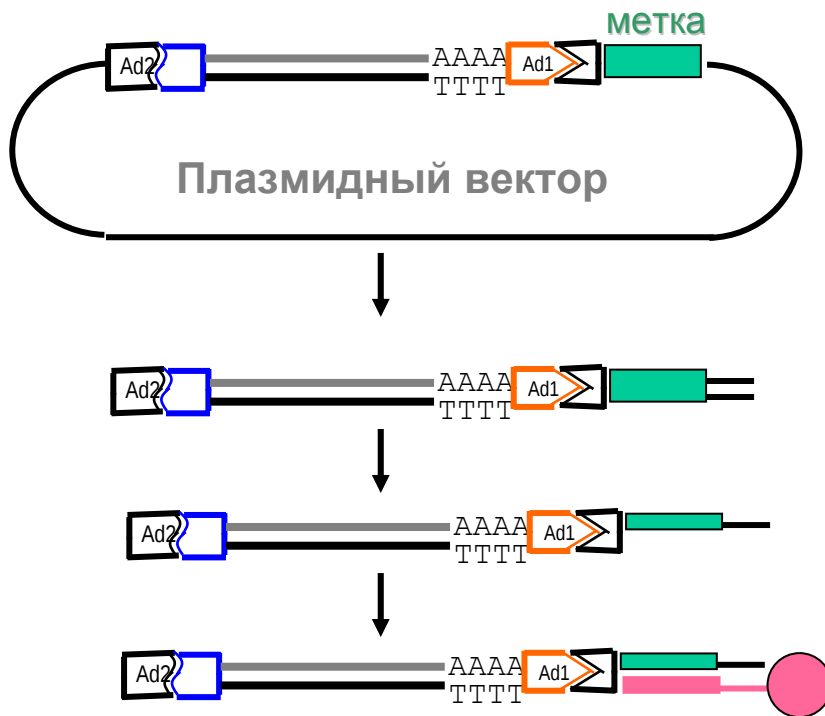


Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)



1. Приготовление кДНК из мРНК и клонирование в вектор, содержащий набор из 1.67×10^7 разных 32-мерных олигонуклеотидных меток (тагов). Пул кДНК матриц, представляющих $3-4 \times 10^4$ разных транскриптов, после клонирования формирует пул из $5-7 \times 10^{11}$ конъюгатов. Далее берется образец, включающий только 1% ($\sim 1.6 \times 10^5$) от общего числа меток.

Каждый транскрипт, таким образом, конъюгирован с уникальной меткой. Даже единичный транскрипт представлен с вероятностью более 99%.

Образец из конъюгатов амплифицируется в ПЦР, метки делают одноцепочечными. Конъюгаты гибридизуются с популяцией микробусин, каждая снабжена каким-либо одним компонентом. 1% микробусин будет связан с конъюгатами. Только эти будут отсортированы с помощью FACS (Fluorescence-Activated Cell Sorter).



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



• Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)

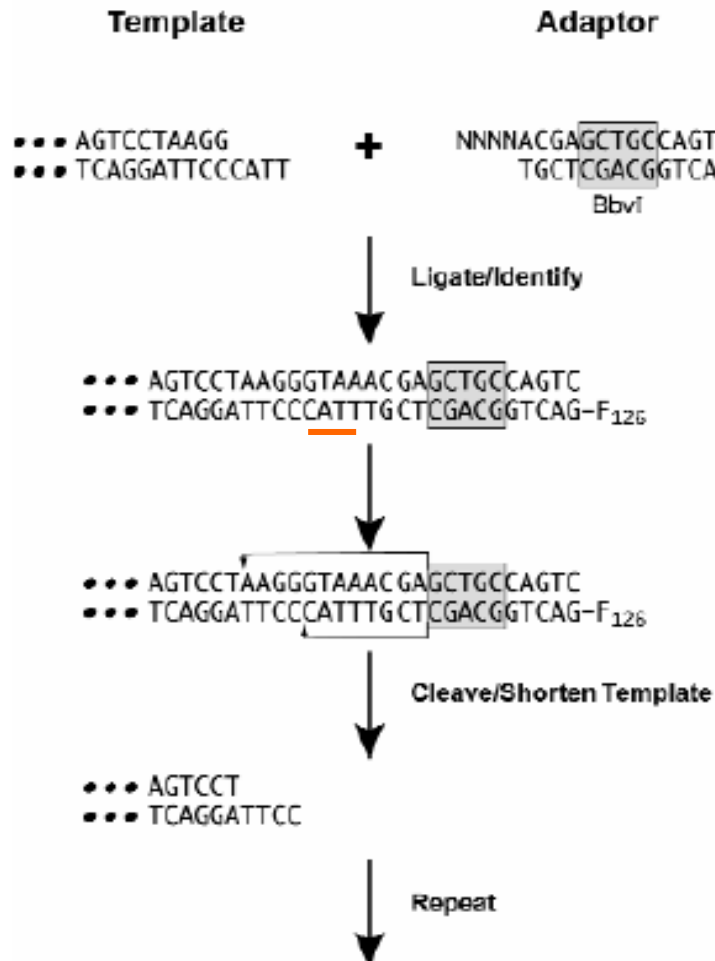


Table 1. Sequences of encoded adaptors^a

Common strand:

5'-GACTGGCAGCTCGT

Encoded adaptors for detecting base 1:

5'-NNNAACGAGCTGCCAGTCcatttaggcg

5'-NNNGACGAGCTGCCAGTCctgattaccg

5'-NNNCACGAGCTGCCAGTCaccaatacgg

5'-NNNTACGAGCTGCCAGTCcgctttgtag

Encoded adaptors for detecting base 2:

5'-NNANACGAGCTGCCAGTCggaacctgaa

5'-NNGNACGAGCTGCCAGTCgtgtcgtgat

5'-NNCNACGAGCTGCCAGTCaccgacattc

5'-NNTNACGAGCTGCCAGTCattctctctc

Encoded adaptors for detecting base 3:

5'-NANNACGAGCTGCCAGTCcgaagaagtc

5'-NGNNACGAGCTGCCAGTCtggtctctct

5'-NCNNACGAGCTGCCAGTCtagcggactt

5'-NTNNACGAGCTGCCAGTCggcgataact

Encoded adaptors for detecting base 4:

5'-ANNNACGAGCTGCCAGTCgcatccatct

5'-GNNNACGAGCTGCCAGTCcaactcgta

5'-CNNNACGAGCTGCCAGTCcacagcaaca

5'-TNNNACGAGCTGCCAGTCgccagtgta

- Смесь адаптеров, включающих все возможные четырехбуквенные выступы, отождествляется к целевой последовательности так, чтобы лигировался только тот, который имеет совершенный комплементарный выступ.

- Каждый из этих 256 адаптеров имеет уникальный мотив, F_n, который может быть обнаружен после лигирования.

- Следующий цикл начинается с расщепления BbvI, чтобы высвободить следующие четыре основания матрицы

Brenner S, Johnson M, Bridgham J, et al., Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol. 2000; 18(6):630-634.

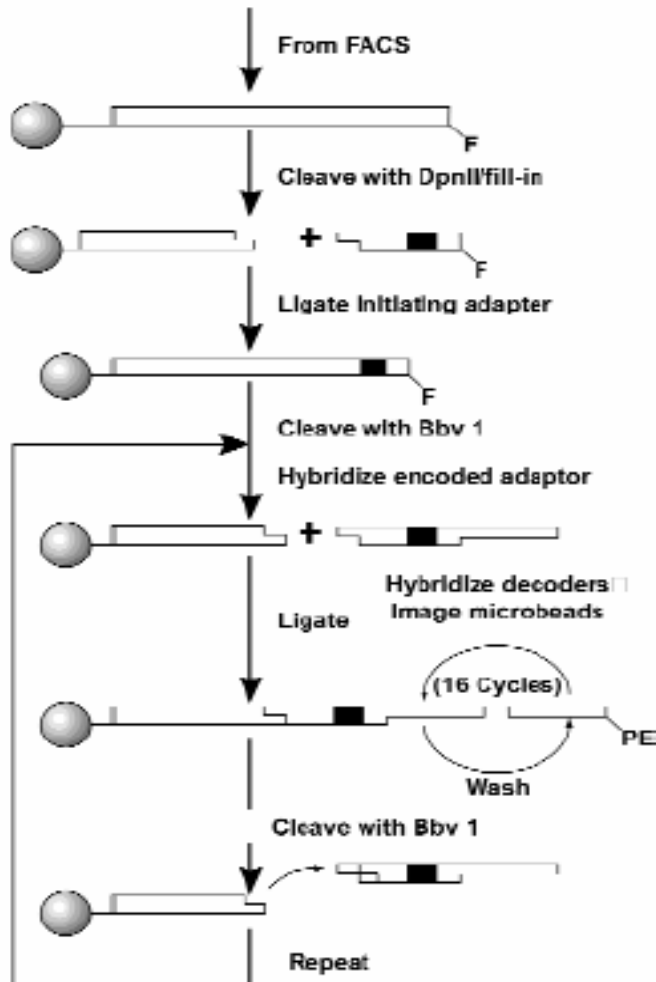


Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



• Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)



- После того как микробусины, загруженные флуоресцентно помеченными (F) кДНК, изолированы с помощью FACS, кДНК разрезаются рестриктазой *DpnII*, чтобы высвободить четыре основания в виде выступа. Этот выступ достраивается до выступа с тремя основаниями.
- Флуоресцентно помеченные (F) иницирующие адаптеры, содержащие *BbvI*-сайты, лигируются к кДНК в отдельной реакции, после чего микробусины загружаются в специальные ячейки.
- Затем кДНК разрезаются *BbvI*, гибридизуются кодирующие адаптеры, те что прогибридизовались с ними - лигируются.
- Шестнадцать помеченных фикоэритрином (PE) зонда по отдельности гибридизуются с декодирующими мотивами кодирующих адаптеров. После каждой гибридизации ячейка сканируется, изображение рядов микробусин анализируется для идентификации оснований выступов.
- Кодирующие адаптеры снова обрабатываются *BbvI*, который расщепляет кДНК и высвобождает четыре новых основания для следующего цикла лигирования и расщепления.

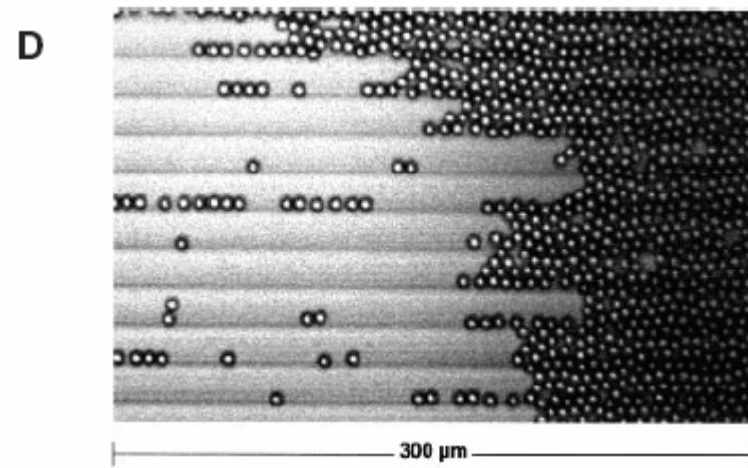
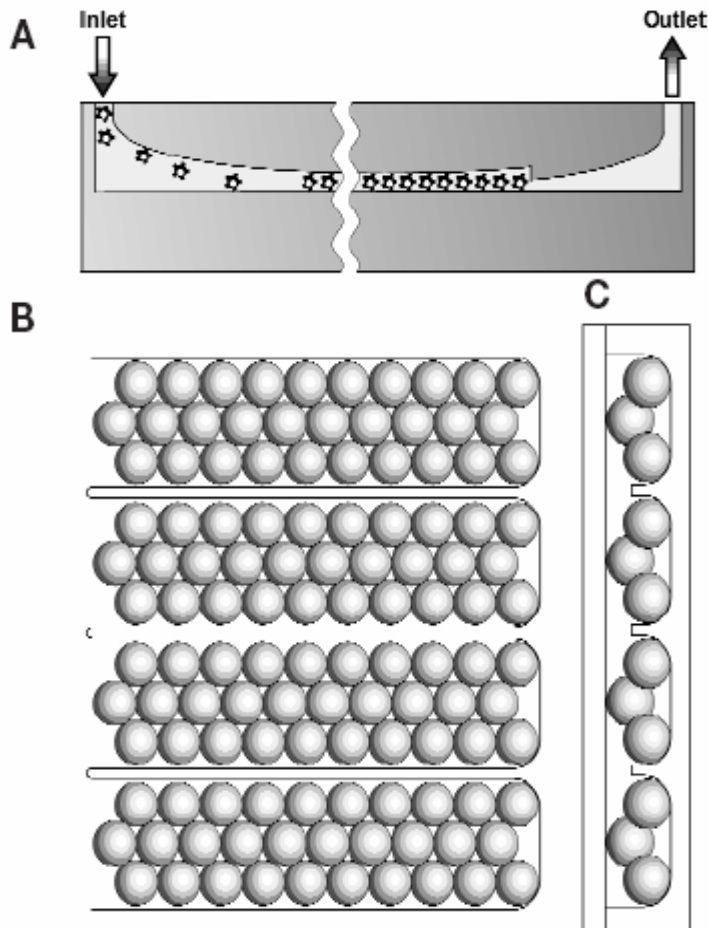
Brenner S, Johnson M, Bridgham J, et al., Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol. 2000; 18(6):630-634.



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)



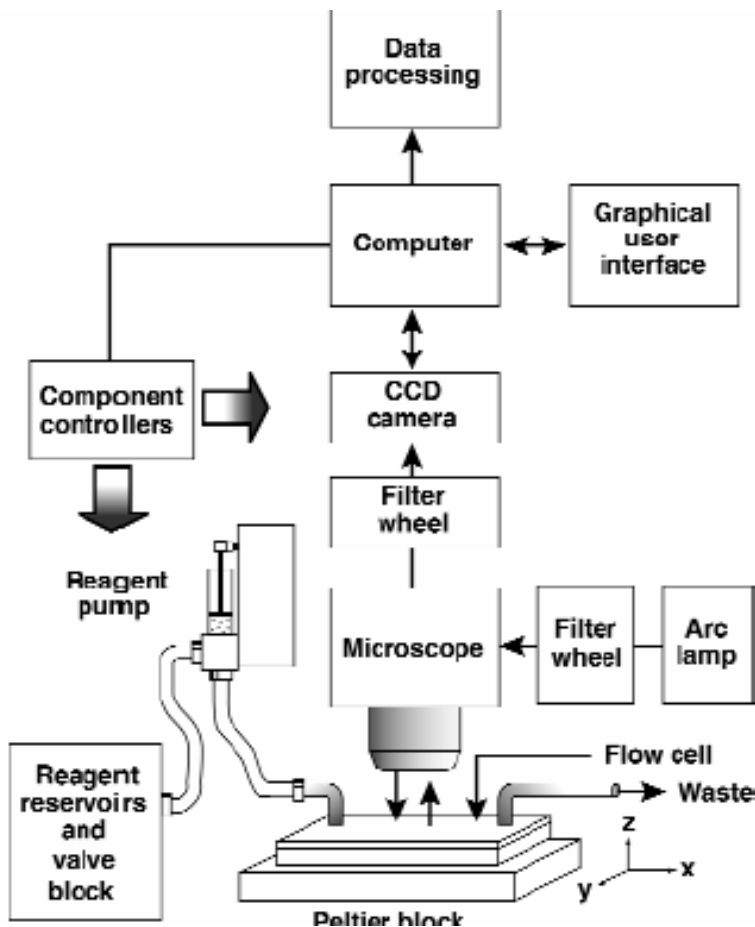
Слева сверху – продольный разрез ячейки. Она сделана так, чтобы удерживать плоские ряды из микробусин. Загрузка микробусин и реагентов идет через входное отверстие. Микробусины задерживаются специальным барьером, а растворы проходят насквозь. Небольшие перемещения микробусин, зафиксированные сканером, учитываются при анализе изображений.



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)



Проточная ячейка монтируется на терморегулируемом блоке Пельтье в конфокальном микроскопе, снабженном ксеноновой лампой, фильтрами и CCD-камерой.

Для сканирования изображения ряды микробусин разделены на 18 секций с 62,000 микробусин в каждой. Специальное программное обеспечение обрабатывает изображения, выделяя центр каждой микробусины и суммируя флуоресцентный сигнал с определенных пикселей изображения.

Как правило, удастся получить надежные данные о последовательности 16-20 нуклеотидов кДНК-матрицы. Далее начинается сказываться «шум» из-за ошибок рестрикции и лигирования.

Полученные 16-20-нуклеотидные мотивы потом преобразуются в сигнатуры (идентифицирующие фрагменты) с помощью специальных программ.

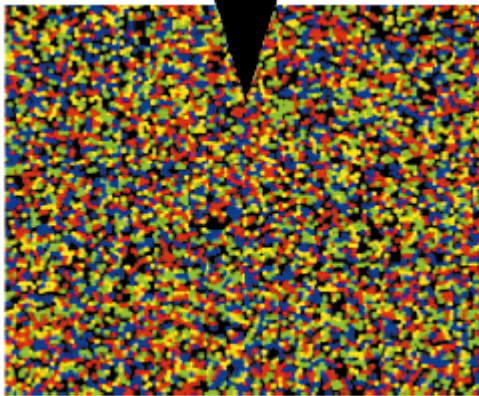
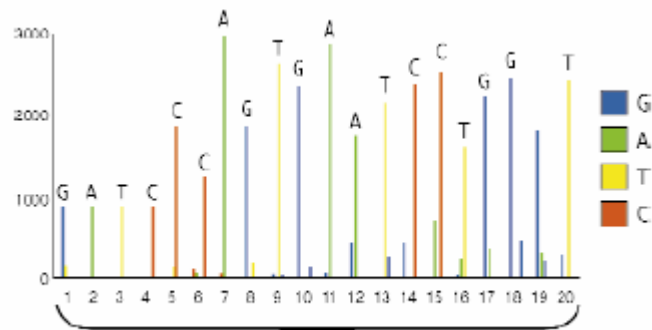


Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)



Так выглядит псевдоцветное компьютерное изображение некоторого количества микробусин и изображение гистограммы значений флуоресцентного сигнала с расшифровкой последовательности для некоей микробусины.



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.



- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)

Интернет-ресурсы

About the Plant MPSS Databases

These data have been generated as part of a project funded by the National Science Foundation Plant Genome Research Project. The database and the web pages that you see were produced by [the Meyers lab](#), with plant tissue generated within the lab or produced by collaborating laboratories. The Meyers lab is located in the [Delaware Biotechnology Institute](#). All of the data was produced under contract by [Solexa, Inc.](#) The data are freely available, but we ask that you cite this web page or related publications (in the works); we would also like to know that you have found our data to be useful, so please let us know if you use data from this page. Please note, however, that the availability of this data and information does not constitute scientific publication; we request that you respect our right to publish on the complete data set.

<http://mpss.udel.edu/>



Home
Basic Queries
Bulk Query
Query by Sequence
Word Analysis
Library Analysis
Library Information
Tutorial & Facts
FAQs & Help
Download Data
Publications
Links
Credits
Comments

Meyers Laboratory
Small RNA Collaborators:

Arabidopsis MPSS Database

This data set was the first generated by this project under NSF F award #0110528, between October, 2001 and October, 2003. The libraries that comprise more than 43 million total signatures. Thes [this link](#) or click on the button below.

[Go to Arabidopsis database](#)

Rice MPSS Database

This data set was generated under NSF Plant Genome Research F October, 2003 and the present time. This project is a collaboratic [Wang at Ohio State](#). There will be at least 60 rice libraries that c signatures. These data are available if you select [this link](#) or click

[Go to rice database](#)

Grape MPSS Database

Basic Queries

First choose a database for the mRNA data (17 bp data is available for MPSS small RNAs and full length for 454 sequences):

☒ 17 bp signatures ☐ 20 bp signatures

Next in the box below, enter specific information to retrieve MPSS map or data:

Protein entry code (e.g. At3g62980)		Get Data
BAC clone name (e.g. F27M3)		Get Data
Sequence of mRNA MPSS signature (e.g. GATCCAGAGGATGTACA)		Get Data
Sequence of small RNA MPSS signature (e.g. AGAATCTTGATGATGCT)		Get Data
Keyword for predicted protein function (e.g. actin)		Get Data

Query by Chromosome Position

First choose a database:

☒ 17 bp signatures ☐ 20 bp signatures



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

- Массовое одновременное секвенирование идентифицирующих фрагментов (Massively Parallel Signature Sequencing - MPSS)

Интернет-ресурсы



http://www.lynxgen.com/wt/tert.php3?page_name=mpss

Massively Parallel Signature Sequencing Technology

Our MPSS™ technology addresses the need to generate **sequence** information from millions of **DNA** fragments. At this extremely large scale, our MPSS™ technology approach eliminates the need for individual **sequencing** reactions and the physical separation of DNA fragments required by conventional sequencing methods.

MPSS™ technology enables the simultaneous identification of nearly all the DNA molecules in a sample. With one million or more Megaclone™ micro-beads fixed in a single layer array in a flow cell, solvents and reagents can be washed over the micro-beads in each cycle of the process. Our proprietary protocol elicits sequence-dependent fluorescent responses from the micro-beads, which are recorded by a CCD camera after each cycle. Short 17- to 20-base signature, or identifying.



LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

[About](#)

[Search](#)

[Download raw
MPSS data](#)

MPSS data

We have used massively parallel signature sequencing (MPSS) to sample the transcriptomes of 32 normal human tissues to an unprecedented depth, thus documenting the patterns of expression of almost 20,000 genes with high sensitivity and specificity. The data confirm the widely held belief that differences between cell and tissue types are largely determined by the expression of a limited number of tissue-specific genes, rather than by combinations of more promiscuously expressed genes. Expression of a little over half of all known human genes seems to account for both the common requirements and the specific functions of the tissues sampled. A classification of tissues based on patterns of gene expression largely reproduces classifications based on anatomical and biochemical properties. The unbiased sampling of the human transcriptome achieved by MPSS supports the notion that most human genes have been mapped, if not functionally characterised. This dataset should

<http://mpss.licr.org/>



Исследование функциональных характеристик генома

Исследование качественных и количественных характеристик транскриптома методами общегеномного скрининга.

