



Исследование качественных и количественных характеристик транскриптома

Катохин Алексей Вадимович, к.б.н.

Кафедра информационной биологии ФЕННГУ



Исследование качественных и количественных характеристик транскриптома

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.

Функции генома:

- содержать/кодировать генетическую информацию и генетические программы развития,
- сохранять их в процессах жизнедеятельности одной клетки (в процессе одного клеточного цикла - репликация, репарация и т.д.) и в процессе пролиферации клеток - передача через ряд митотических делений),
- передавать их в ряду поколений особей (мейоз, рекомбинация),
- определять условия для формирования транскриптома, характерного для клеток определенных типов и предопределенного генетическими программами развития, в процессах транскрипции и сплайсинга.



Исследование качественных и количественных характеристик транскриптома

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.

Задачи транскриптомики:

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

Применение методов биоинформатики позволяет:

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



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.



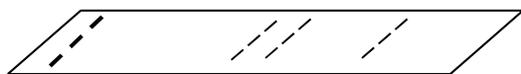
Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

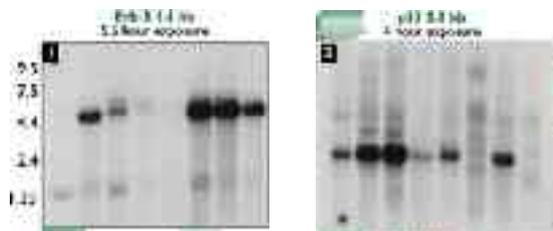
Электрофорез в агарозном геле образцов РНК



Перенос фракций РНК на мембрану



Гибридизация с индивидуальными мечеными ДНК-или РНК-пробами



- Прямая детекция РНК

- Нозерн-блот-гибридизация (Northern blot hybridization)



Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

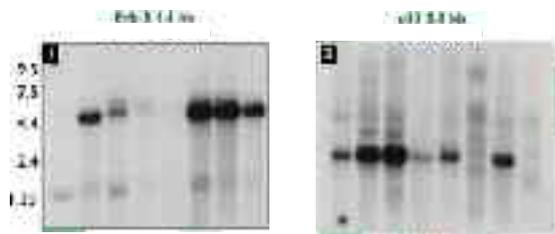
Электрофорез в агарозном геле образцов РНК



Перенос фракций РНК на мембрану



Гибридизация с индивидуальными мечеными ДНК-или РНК-пробами



- Прямая детекция РНК

- Нозерн-блот-гибридизация (Northern blot hybridization)

Стандартный метод для детекции отдельных мРНК и количественной оценки их относительного содержания в 5-20 образцах.

Достоинства: позволяет определять размер транскриптов и выявлять изоформы, образованные в результате альтернативных процессов транскрипции и сплайсинга; отражает реальное соотношение количеств мРНК.

Ограничения: не дает данных об абсолютном содержании транскриптов, метод малочувствительный, очень трудоемкий, долгий и малопродуктивный



Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

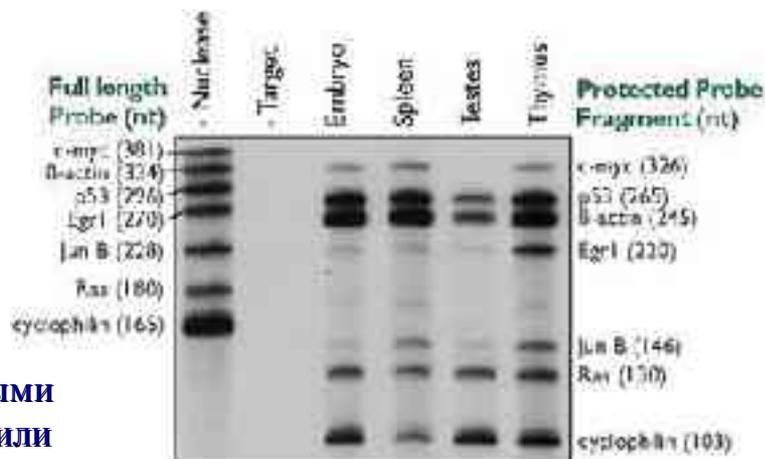


Анализ гель-электрофорезом

Гибридизация с индивидуальными или наборами меченымх ДНК-или РНК-проб

- Прямая детекция РНК

- Анализ с помощью защиты от рибонуклеазы (Ribonuclease protection assay)



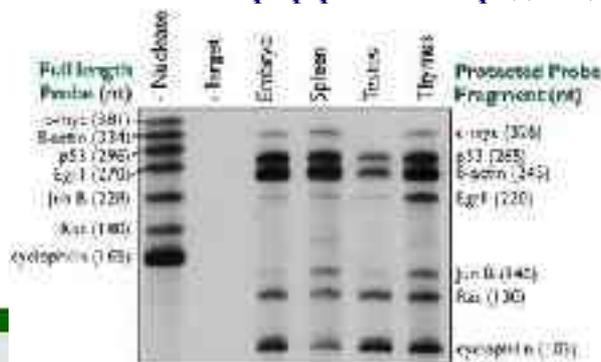


Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения



Анализ гель-электрофорезом и гибридизацией



- Прямая детекция молекул РНК
- Анализ с помощью защиты от рибонуклеазы (Ribonuclease protection assay)

Метод для одновременной детекции 10-15-ти мРНК и количественной оценки их содержания в 5-20-ти образцах.

Достоинства: отражает реальное соотношение количеств мРНК, высокочувствительный; позволяет точно картировать 5'- и 3'- окончания транскриптов и экзон-интронные стыки.

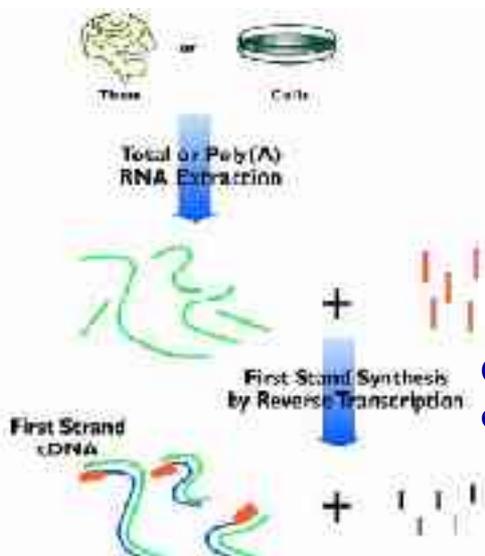
Ограничения : метод еще более трудоемкий, долгий и малопродуктивный.



Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

- Детекция молекул РНК с применением обратной транскрипции и полимеразной цепной реакции



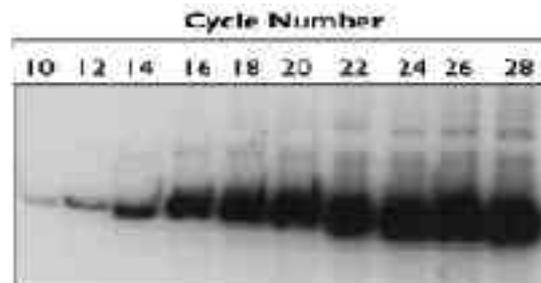
Обратная транскрипция, синтез первой цепи кДНК

- количественная ОТ-ПЦР (Обратная Транскрипция+Полимеразная Цепная Реакция) (quantitative RT-PCR, qRT-PCR)

Полимеразная цепная реакция, амплификация кДНК с помощью пары специфических праймеров



Анализ гель-электрофорезом





Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения



- Детекция молекул РНК с применением обратной транскрипции и полимеразной цепной реакции

- количественная ОТ-ПЦР (quantitative RT-PCR, qRT-PCR)

Самый чувствительный метод для детекции мРНК и количественной оценки их содержания в образцах.

Достоинства: высокая чувствительность; высокая производительность; может выявлять несколько разных транскриптов одновременно.

Ограничения: метод страдает от неспецифики, чувствителен к загрязнениям геномной ДНК, оценки соотношений между разными продуктами реакции страдают от нелинейных искажений, количество одновременно выявляемых транскриптов ограничено.

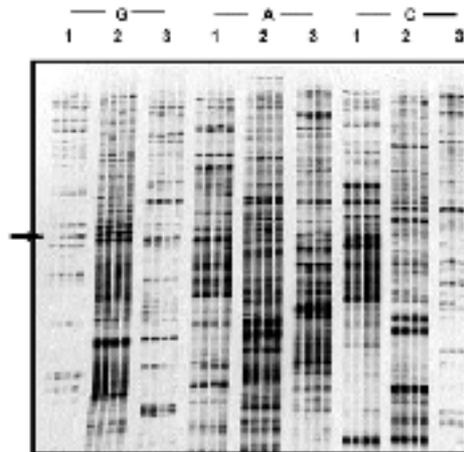
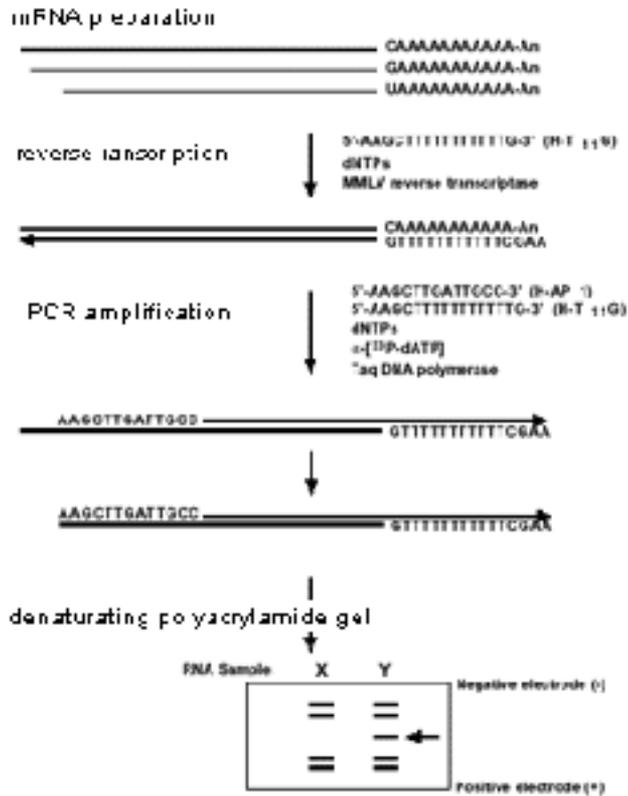


Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

- Детекция молекул РНК с применением обратной транскрипции и полимеразной цепной реакции

- Дифференциальный дисплей (Differential display)



Клонирование и идентификация дифференциально экспрессированных фрагментов



Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

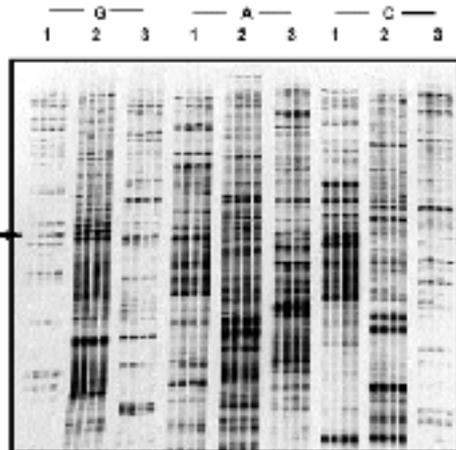
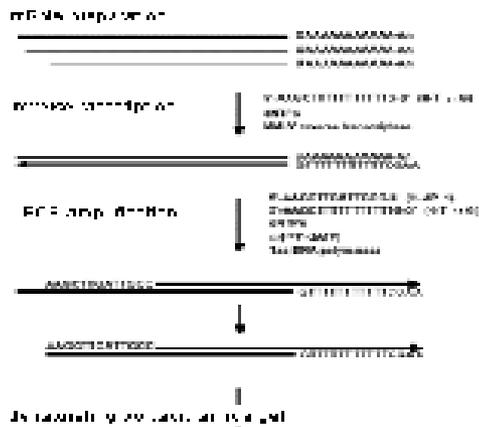
- Детекция молекул РНК с применением обратной транскрипции и полимеразной цепной реакции

- Дифференциальный дисплей (Differential display)

С 1992 самый распространенный метод для быстрого, точного и чувствительного выявления различий в относительном количестве мРНК во многих образцах.

Достоинства: высокая чувствительность; высокая производительность.

Ограничения: анонимность выявляемых полос и необходимость дальнейшей идентификации полос; много фальшивых позитивных полос; можно одновременно анализировать не очень большое количество образцов.



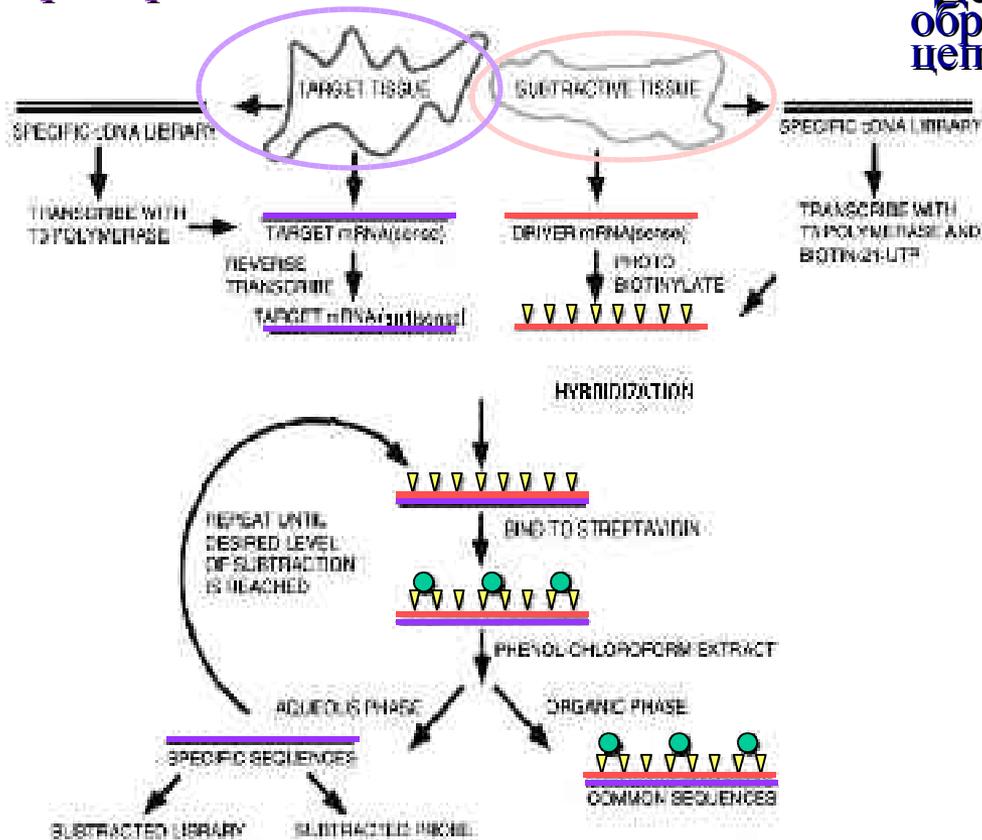


Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

- Детекция молекул РНК с применением обратной транскрипции и полимеразной цепной реакции

- Вычитательная гибридизация (Subtractive Hybridization)





Методы массового исследования транскриптов.

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

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

Мощный метод для глобального компьютерного анализа уровней экспрессии генов.

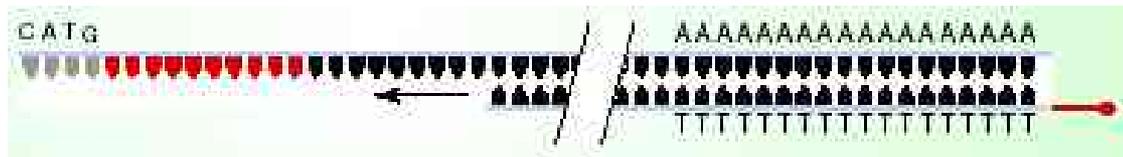
Предложен Victor E. Velculescu в 1995
(Velculescu et al. Science 270 (5235) : 484-487)

Методы массового исследования транскриптов.

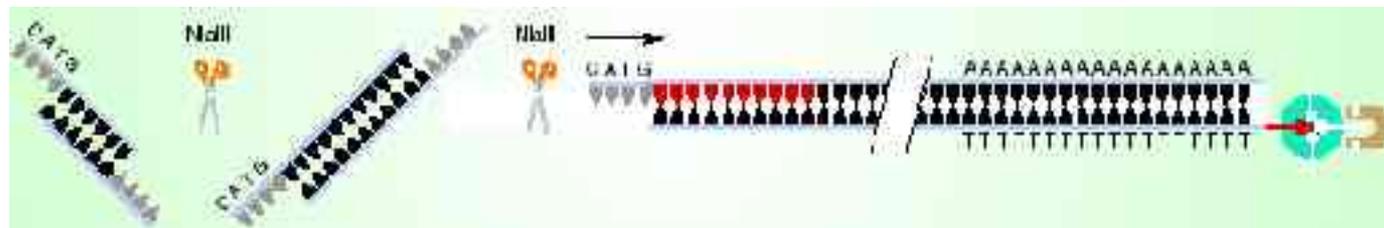
Высокопродуктивные и высокоинформативные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

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

1. Синтез биотинилированной двухцепочечной кДНК.



2. Рестрикция по *Nla*III сайтам и отделение самых 3'-крайних фрагментов.

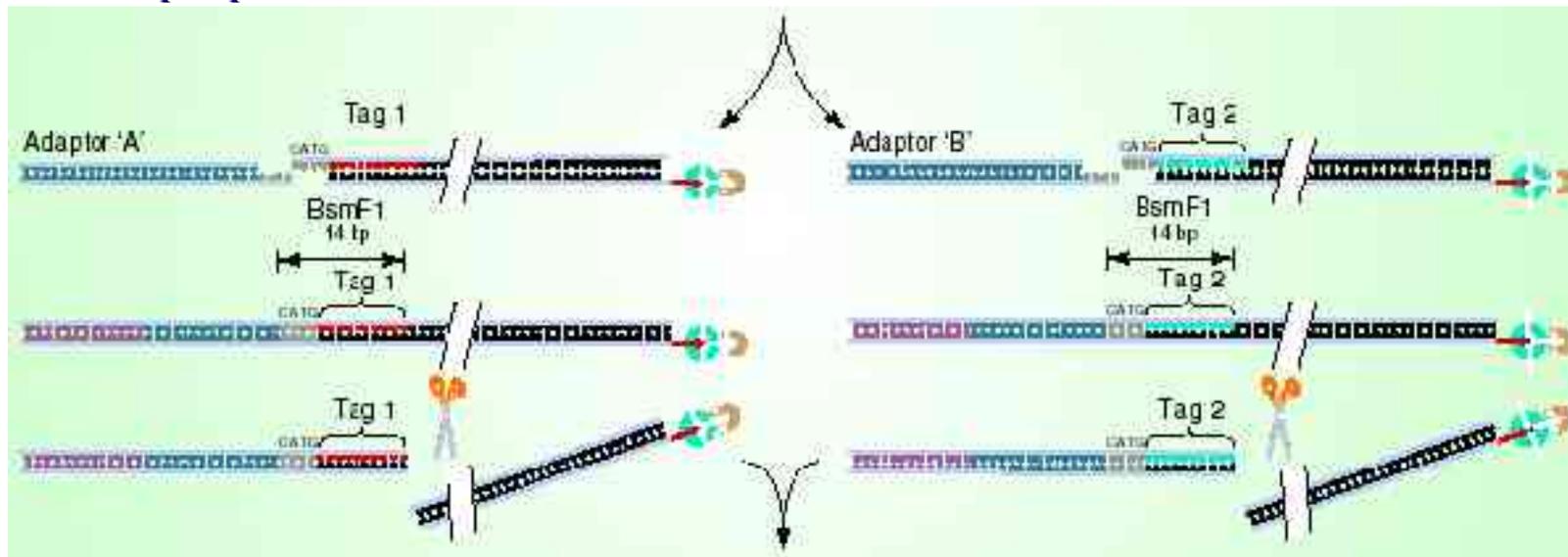


Методы массового исследования транскриптов.

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

- 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 рестриктазы *BsmF1* типа.

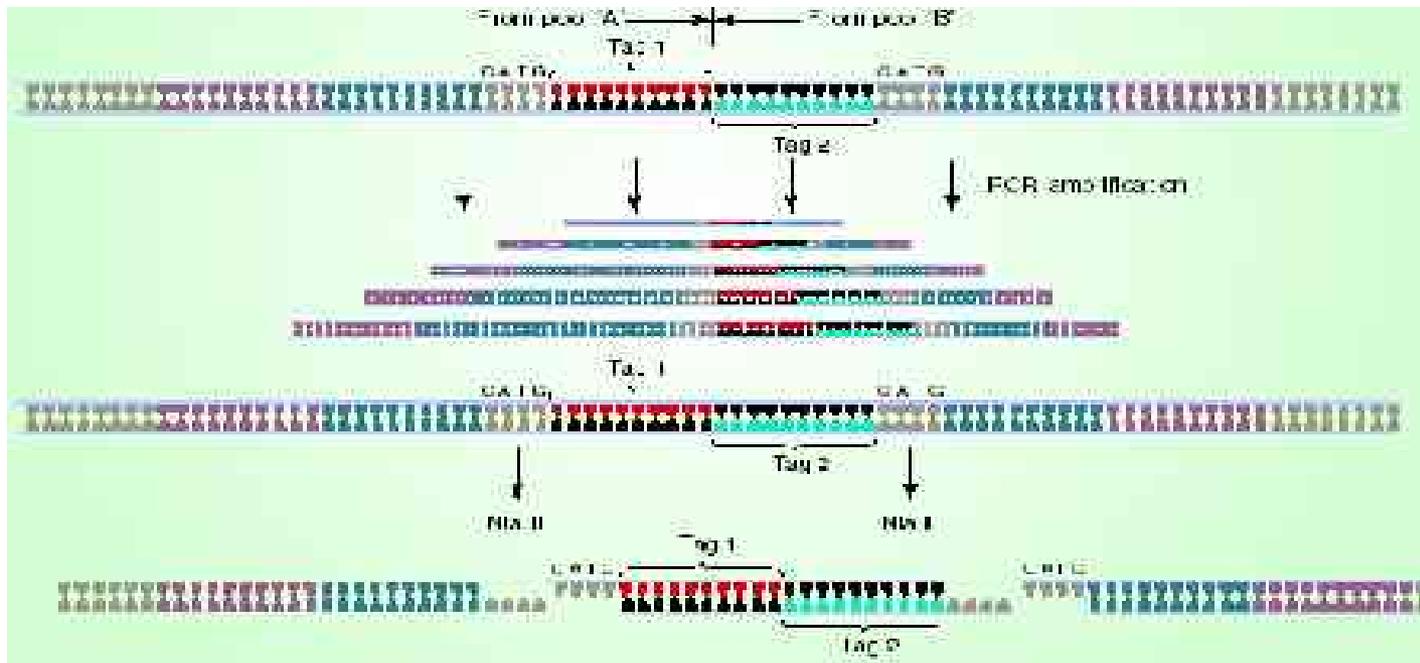


Методы массового исследования транскриптов.

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

4. Образование дитагов, их амплификация и удаление адаптеров.

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

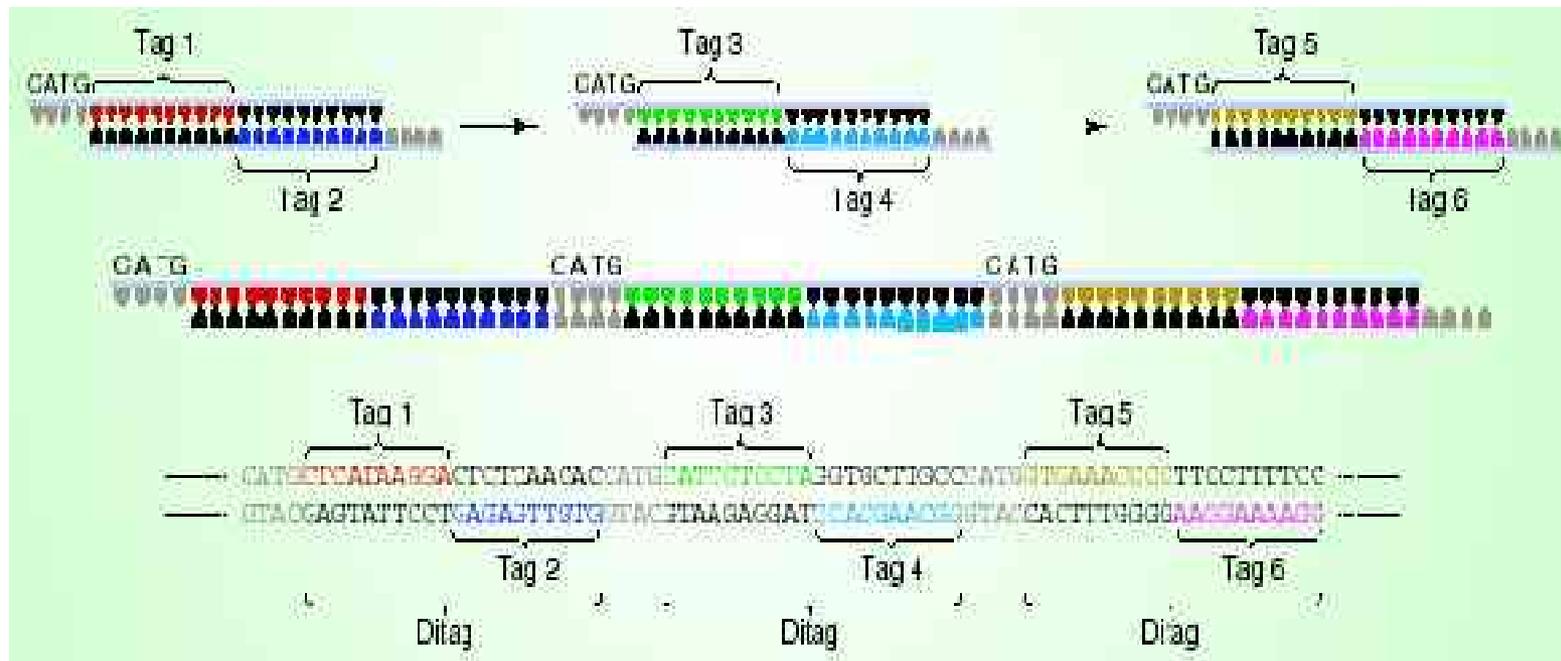


Методы массового исследования транскриптов.

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

5. Формирование конкатемеров и их секвенирование

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





Методы массового исследования транскриптов.

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

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

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





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

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

Основные принципы метода SAGE:

1. Небольшой фрагмент из определенного места транскрипта достаточно информативен для идентификации этого транскрипта.
(9 п.о. => $4^9=262\ 144$ транскрипта)
2. Конкатенация и параллельный анализ повышают продуктивность (секвенирование 96 колоний за один прогон => 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)



Методы массового исследования транскриптов.

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

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,

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

Ограничения:

- зависимость от предварительного знания последовательности значительной части генов организма
- невыявление сплайсированных форм
- 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.



Методы массового исследования транскриптов.

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

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



Методы массового исследования транскриптов.

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

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. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression
Nature Genetics 1992, 2:173-179



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

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

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

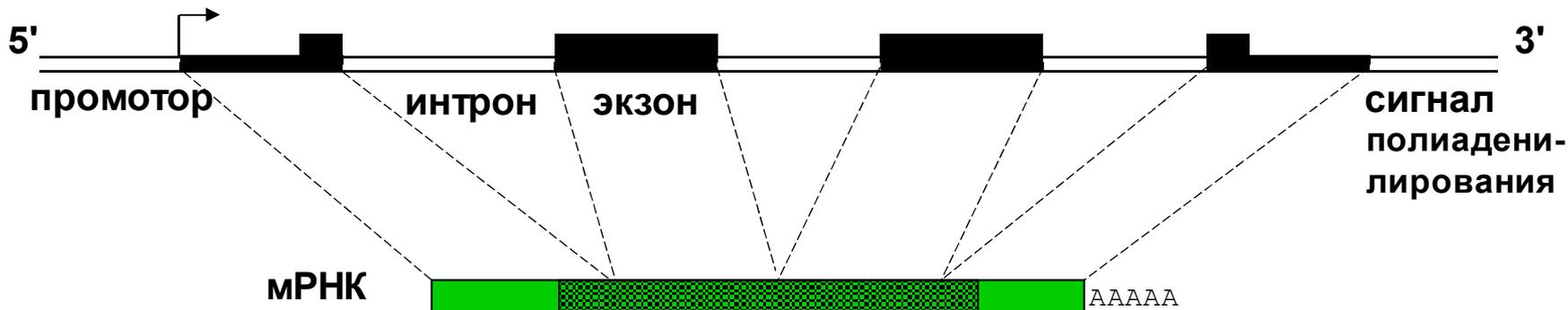
Принципы метода ESTs:

- анализ совокупности прочитанных фрагментов, ассоциированных с каким-либо геном, позволяет реконструировать структуру транскриптов гена, а после сравнения ее со структурой геномного района – структуру самого гена;
- в пределах определенным образом приготовленной библиотеки кДНК частотное распределение кДНК клонов в целом соответствует исходному распределению транскриптов в популяции мРНК, из которой приготовлена библиотека;
- многочисленность прочитанных фрагментов;
- относительная дешевизна прочтения фрагментов .



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

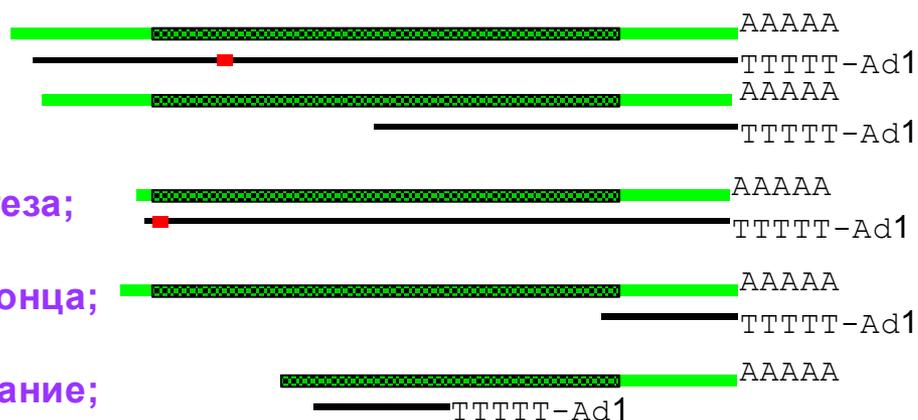
РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА
А. Экспериментальная фаза А1. Приготовление кДНК библиотеки



СИНТЕЗ кДНК ЦЕПИ

Источники отличий от исходных природных молекул:

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



Источник ошибок:

-ошибки обратной транскриптазы

Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА А. Экспериментальная фаза А1. Приготовление кДНК библиотеки

Подготовка к клонированию

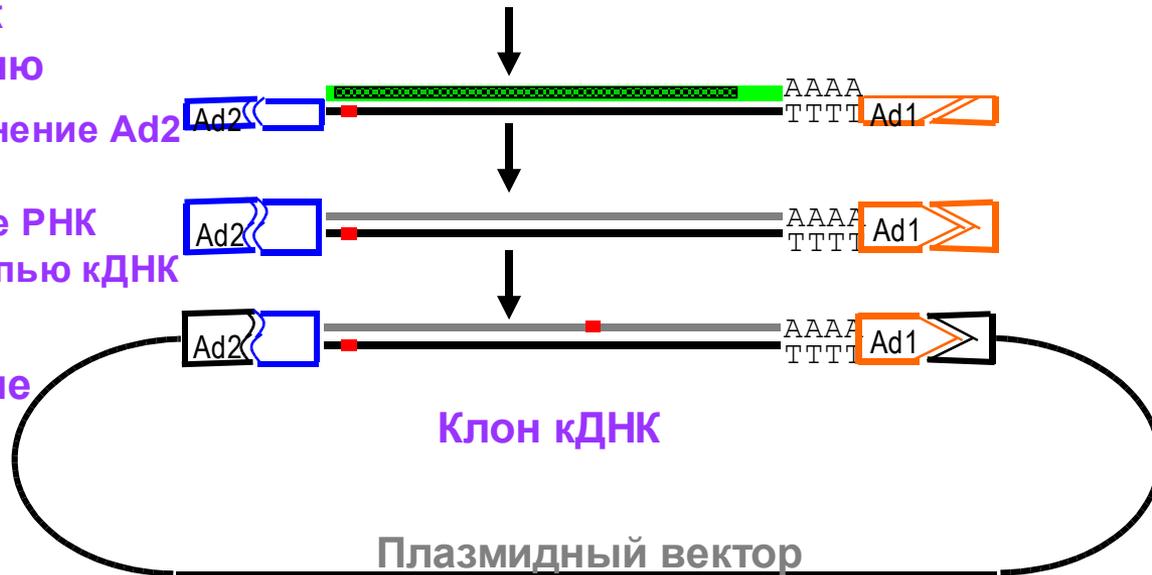
- присоединение Ad2
- замещение РНК второй цепью кДНК

Источник отличий от исходных природных молекул:

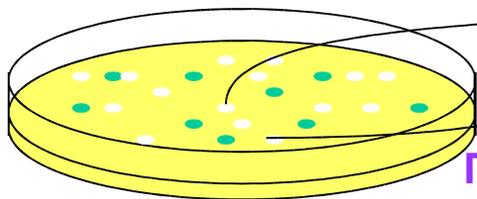
- образование артефактных химерных молекул;

Источник ошибок:
- ошибки Taq полимеразы

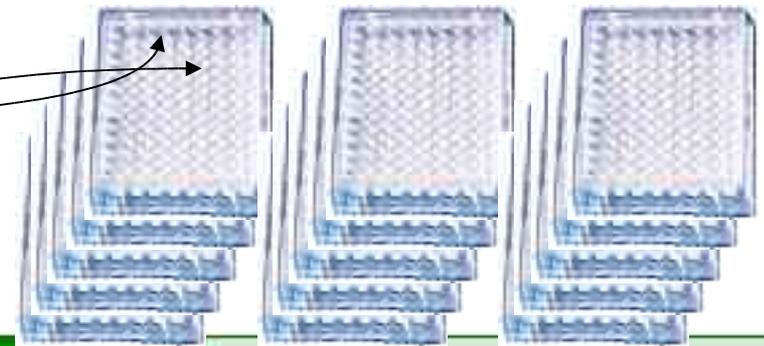
Клонирование кДНК



Трансформация бактериальных клеток



Пересев клонов



кДНК библиотека



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

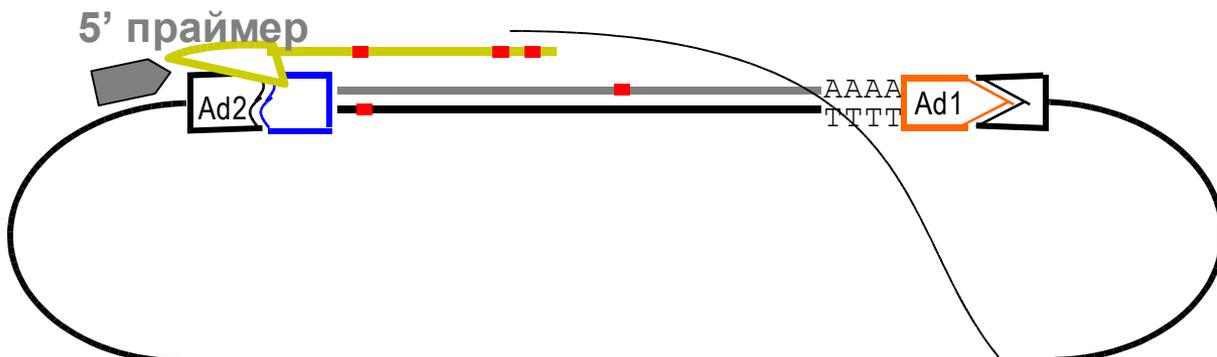
РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА
А. Экспериментальная фаза А1. Приготовление кДНК библиотеки

Использование специфических праймеров к сайтам вектора,
окаймляющим встройку кДНК

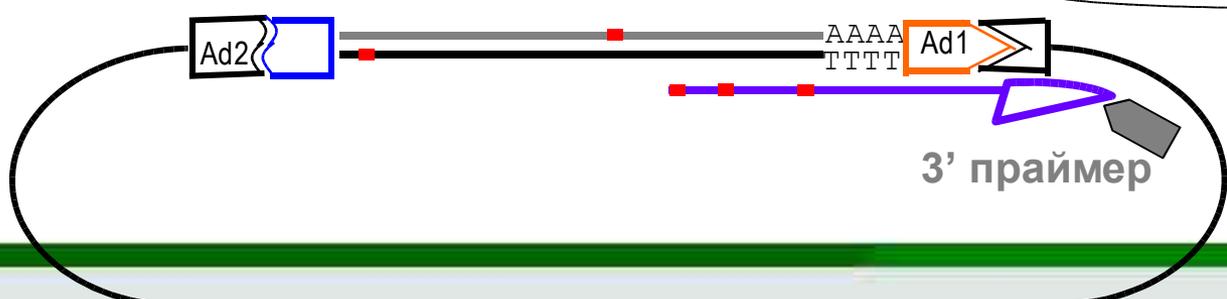
Источник ошибок:

- 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)

Получение 5' EST



Получение 3' EST



В подраздел GenBank – dbEST (или EMBL bank)



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)



Number of public entries:

31,307,034

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: database of
"Expressed Sequence Tags"

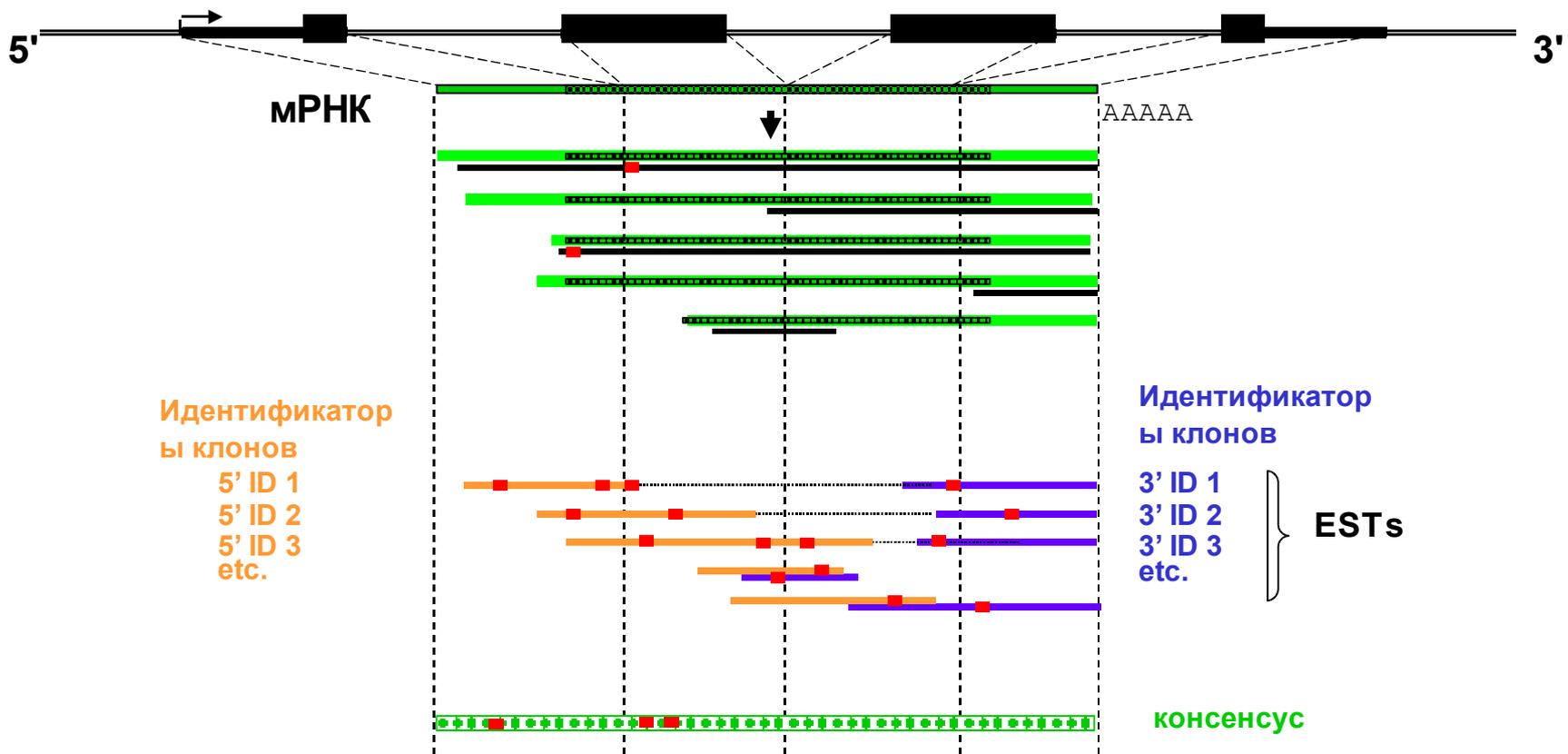
dbEST release 111105

Summary by Organism -
November 11, 2005



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

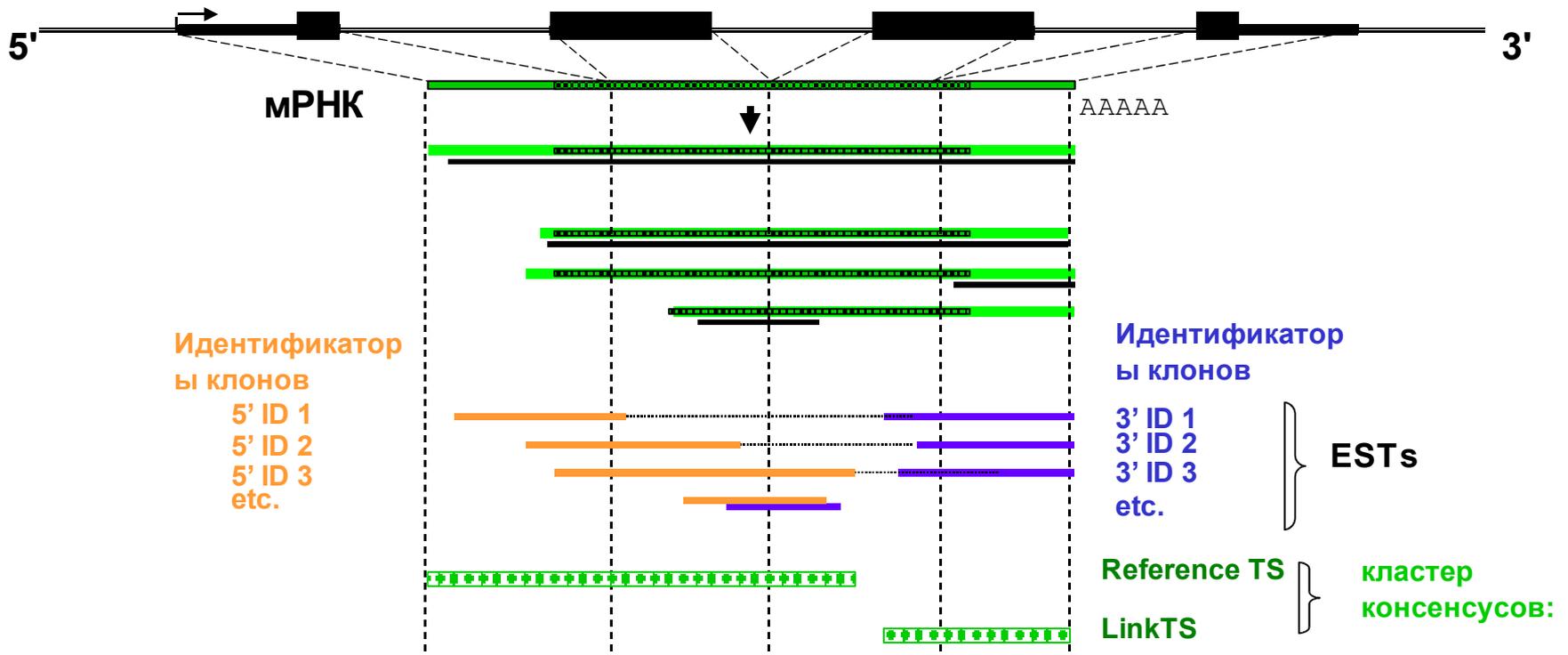
РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА Б. Компьютерная фаза Б1. Выравнивание и кластеризация EST





Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА Б. Компьютерная фаза Б1. Выравнивание и кластеризация EST



Недостаточное «перекрывание» 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.



Методы массового исследования транскриптов. 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 (premature mRNA)



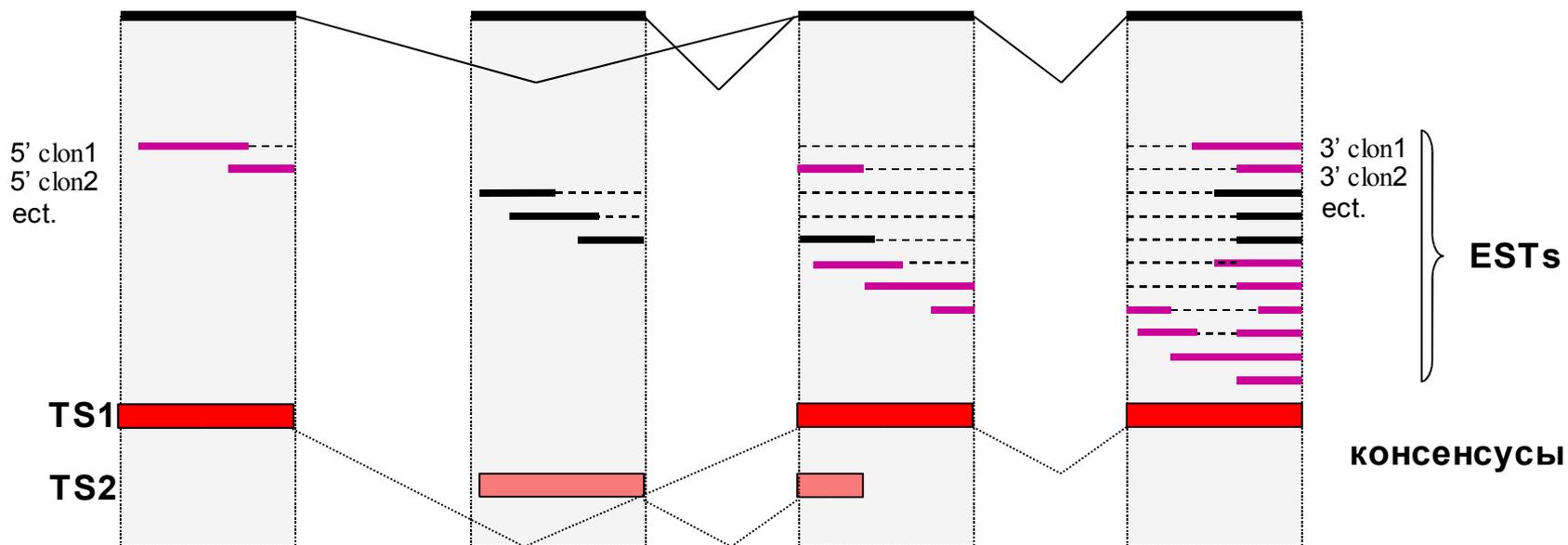
- альтернативное включение экзонов (alt. exon usage, cassette exons)



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА Б. Компьютерная фаза Б2. Анализ структуры кластера

АЛЬТЕРНАТИВНЫЕ СТАРТЫ ТРАНСКРИПЦИИ В АЛЬТЕРНАТИВНЫХ 5'-ЭКЗОНАХ





Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА Б. Компьютерная фаза Б2. Анализ структуры кластера

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



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

Alternative Splicing DB

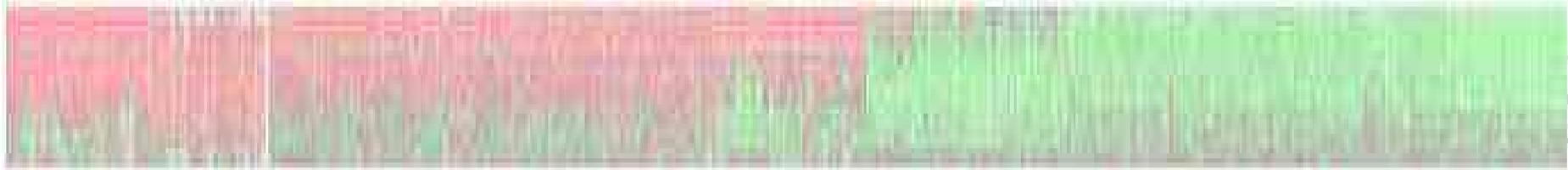
DB CONTENT | HOW TO USE | FURTHER WORK | SEARCH

References to the Alternative Splicing Database:

- ASDB - Database of Alternative Splicing Genes
- Li, Q. and Zhang, J. (2003) ESTs, Exons, Intron, and Splicing: A New Paradigm for Gene Structure Analysis. *Genome Research* 13: 100-108.
- Li, Q., Zhang, J., Sun, Y., Zhang, J. and Zhang, J. (2003) A New Paradigm for Gene Structure Analysis. *Genome Research* 13: 100-108.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Спасибо за внимание.

