



Исследование качественных и количественных характеристик транскриптома

Катохин Алексей Вадимович, к.б.н.

Кафедра трансформационной 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.

Функции генома:

- содержать/кодировать генетическую информацию и генетические программы развития,
- сохранять их в процессах жизнедеятельности одной особи (как в процессе одного клеточного цикла - репликация, репарация и т.д., так и в процессе пролиферации клеток - передача через ряд митотических делений),
- передавать их в ряду поколений (мейоз, рекомбинация),
- создавать 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

Задачи транскриптомики:

- исследование (1) структуры транскриптов и (2) их дифференциального временного и пространственного распределения в клетках и организмах.

Применение методов биоинформатики позволяет:

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

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
- возможность одновременного измерения соотношения транскриптов как можно большого количества генов
- возможность детекции транскриптов очень слабо или очень специализированно экспрессирующихся генов (достаточно широкий динамический диапазон)
- высокая производительность и эффективность для производства достаточно большого массива данных.



Методы исследования отдельных элементов транскриптома – транскриптов.

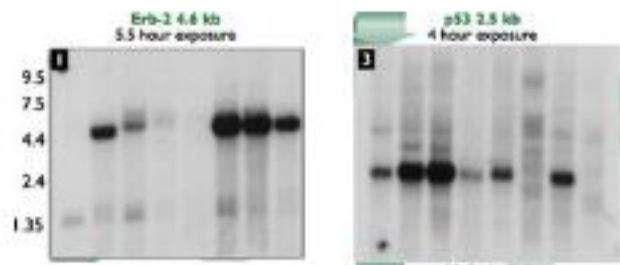
Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

Электрофорез в агарозном геле образцов РНК



• Прямая детекция РНК

•• Нозерн-блот-гибридизация (Northern blot hybridization)

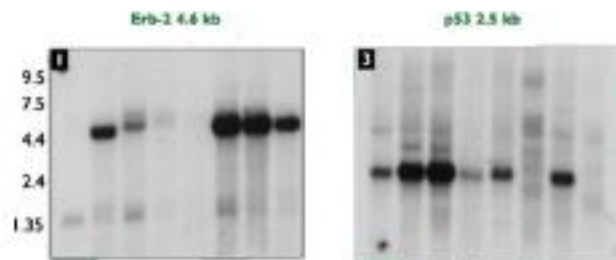




Методы исследования отдельных элементов транскриптома – транскриптов.

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

Электрофорез в агарозном геле образцов РНК



- Прямая детекция РНК

- Нортен-блот-гибридизация (Northern blot hybridization)

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

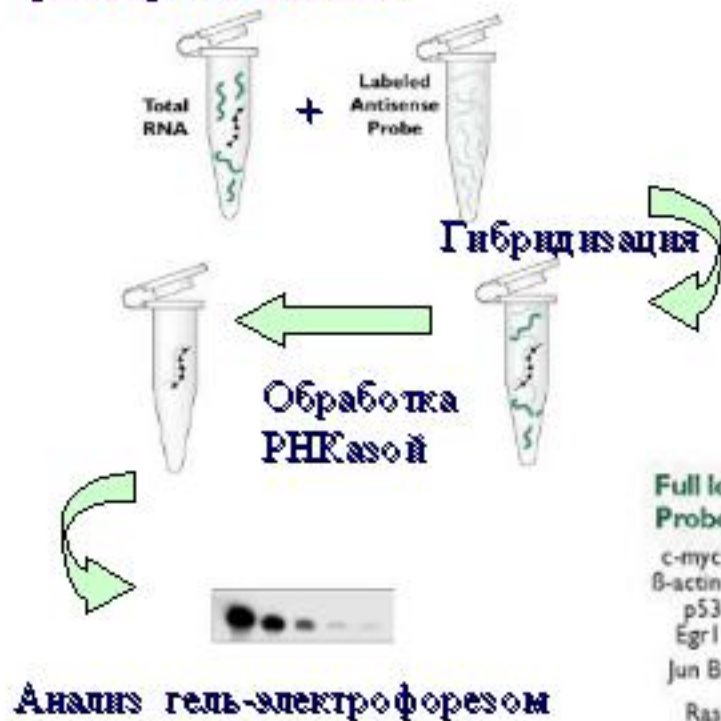
Достоинства: позволяет определить размер транскриптов и выявлять изоформы, образованные в результате альтернативных процессов транскрипции и сплайсинга; отражает реальное соотношение количеств мРНК.

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



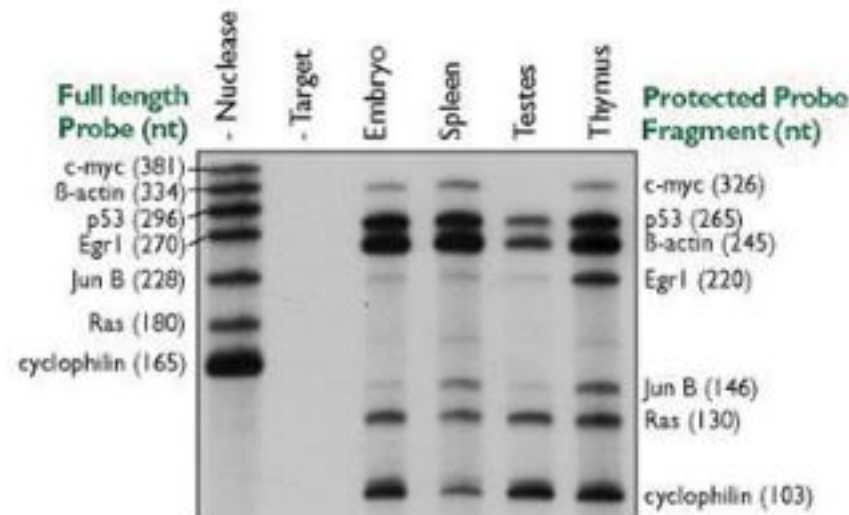
Методы исследования отдельных элементов транскриптома – транскриптов.

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



• Прямая детекция РНК

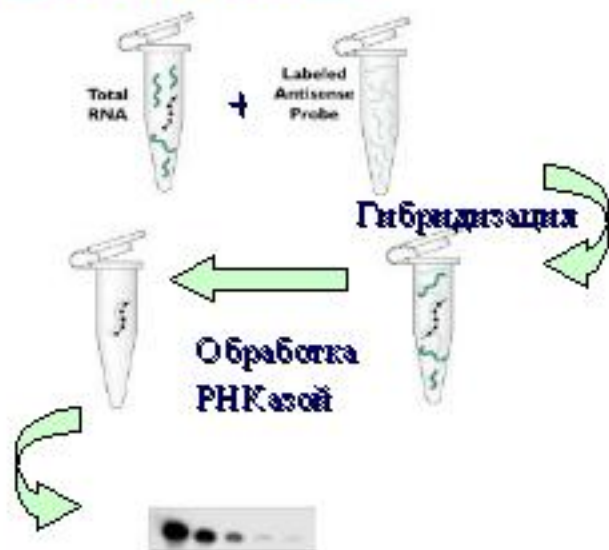
•• Анализ с помощью защиты от рибонуклеазы (Ribonuclease protection assay)



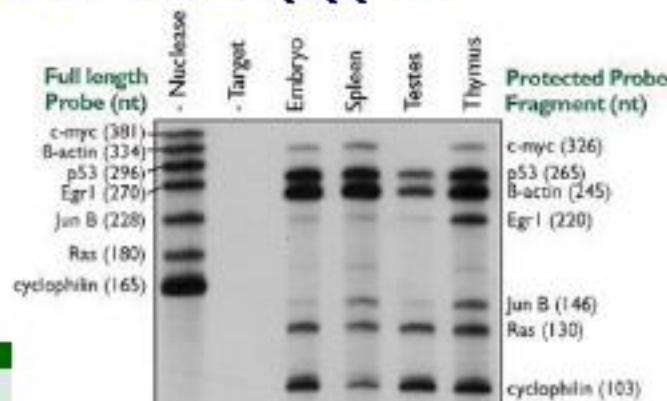


Методы исследования отдельных элементов транскриптома – транскриптов.

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



Анализ гель-электрофорезом



- 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 количества мРНК, высокочувствительный; позволяет точно картировать 5'- и 3'- окончания транскриптов и экзон-интронные стыки

Недостатки: метод еще более трудоемкий, долгий и малопродуктивный



Методы исследования отдельных элементов транскриптома – транскриптов.

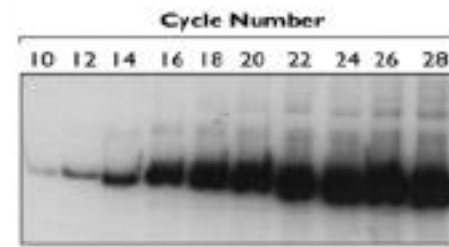
Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

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

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



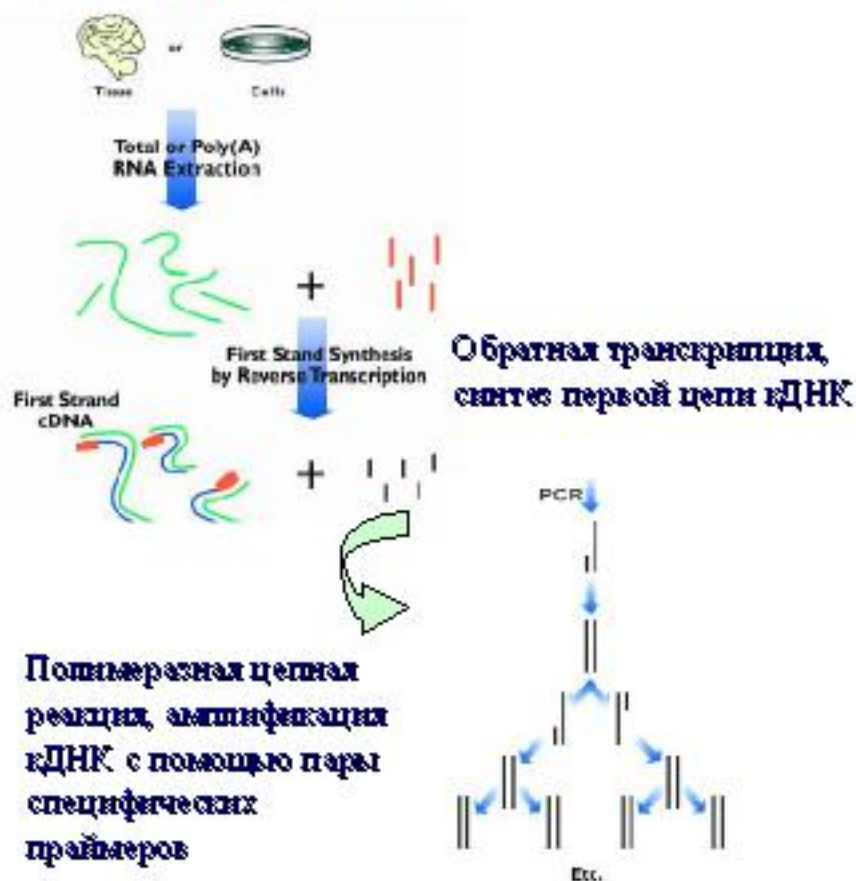
Анализ гель-электрофореза





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.

Недостатки: метод страдает от неспецифичности, 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.

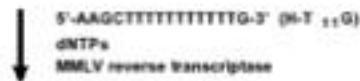
Качественные и полуколичественные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

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

mRNA preparation



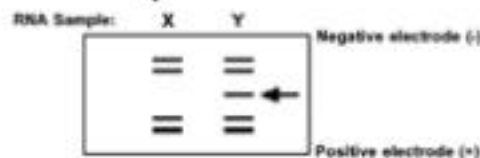
reverse ranscription



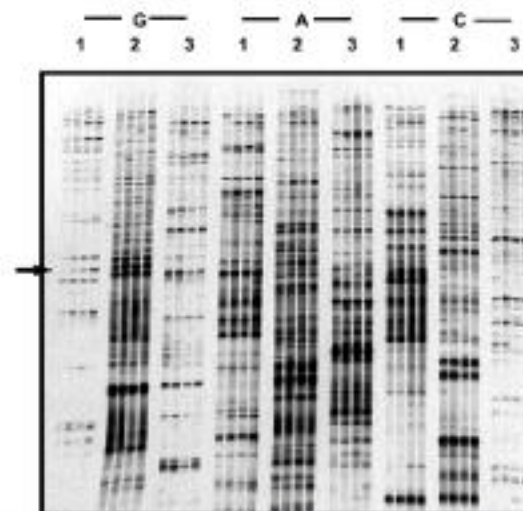
PCR amplification



denaturing polyacrylamide gel



•• Дифференциальный дисплей (Differential display)

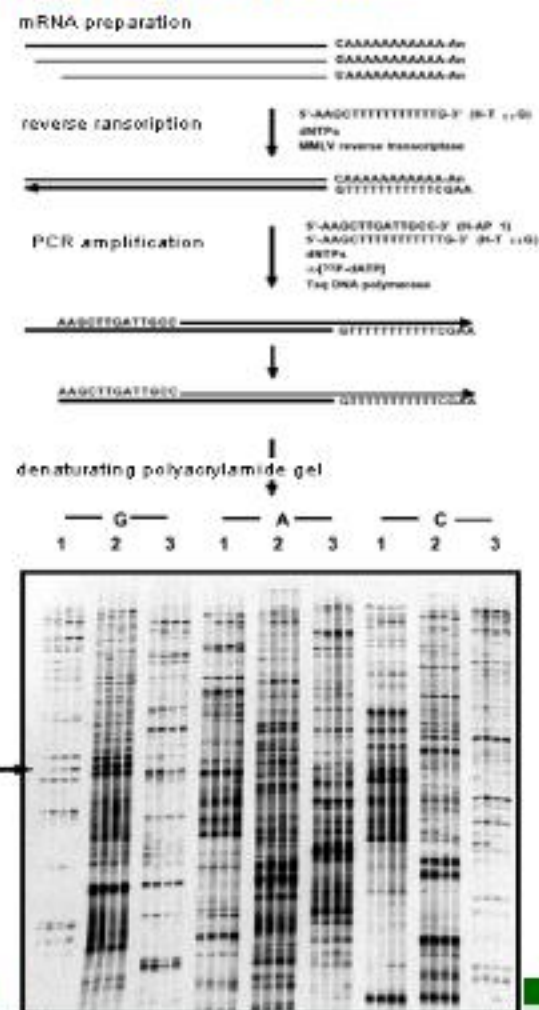


Клонирование и идентификация дифференциально экспрессированных фрагментов



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 распространённый метод для быстрого, точного и чувствительного выявления различий в относительном количестве мРНК во многих образцах.

Достоинства: высокая чувствительность; высокая производительность.

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

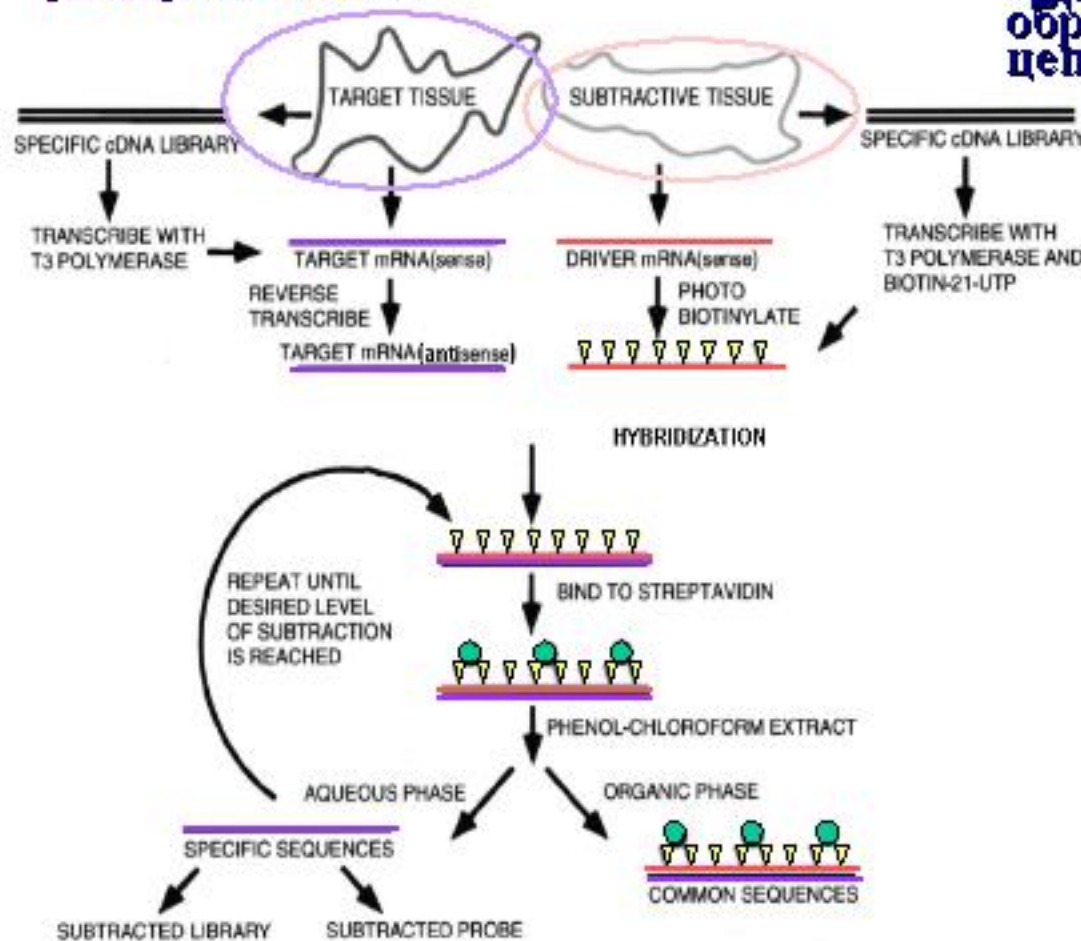


Методы исследования отдельных элементов транскриптома – транскриптов.

Качественные и полуквантитативные 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)





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.

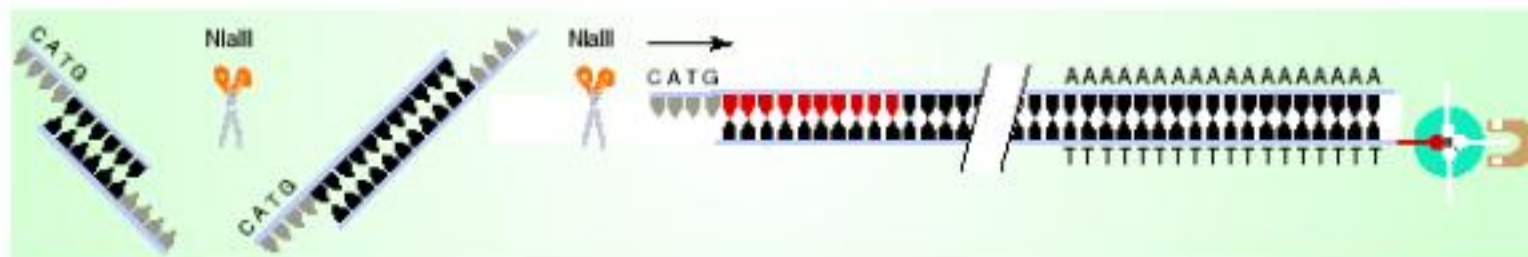
Высокопродуктивные и высокоинформативные методы исследования структуры транскриптов и их дифференциального временного и пространственного распределения

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

1. Синтез биотинилированной двуцепочечной ДНК.



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



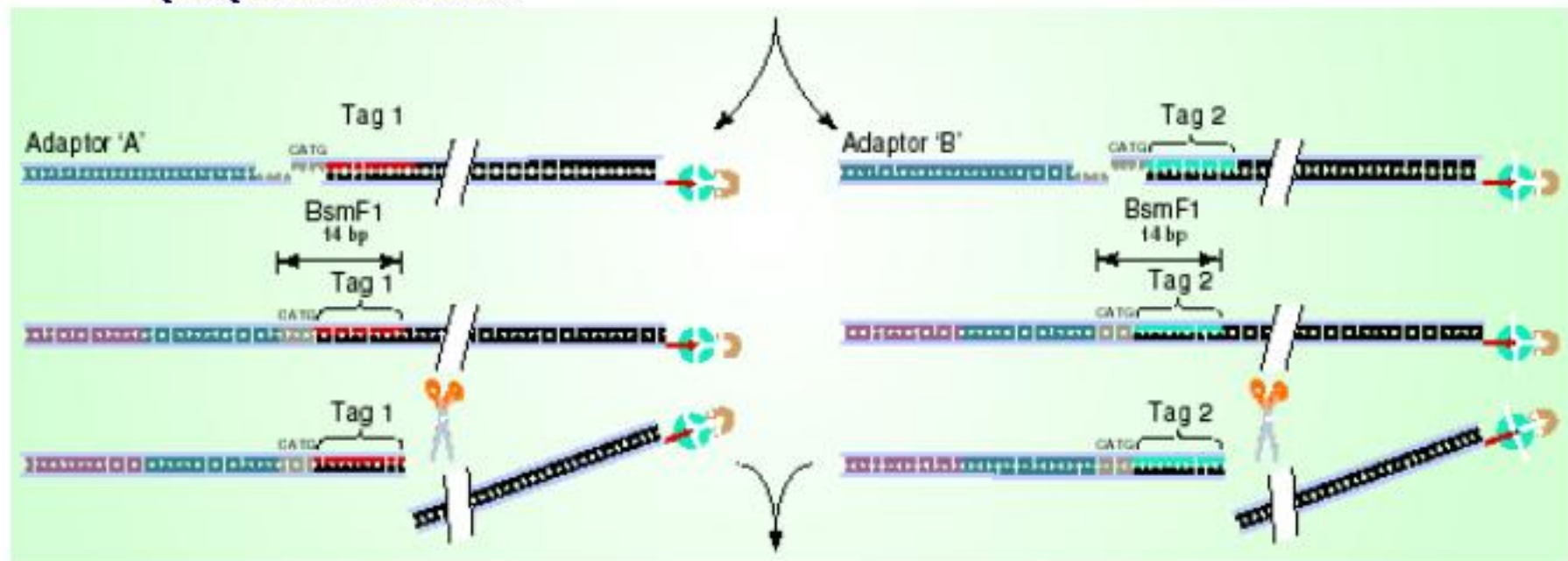


Методы массового исследования транскриптов.

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, тегов (tag), длиной 10 п.о. с помощью рестриктазы Ps типа.



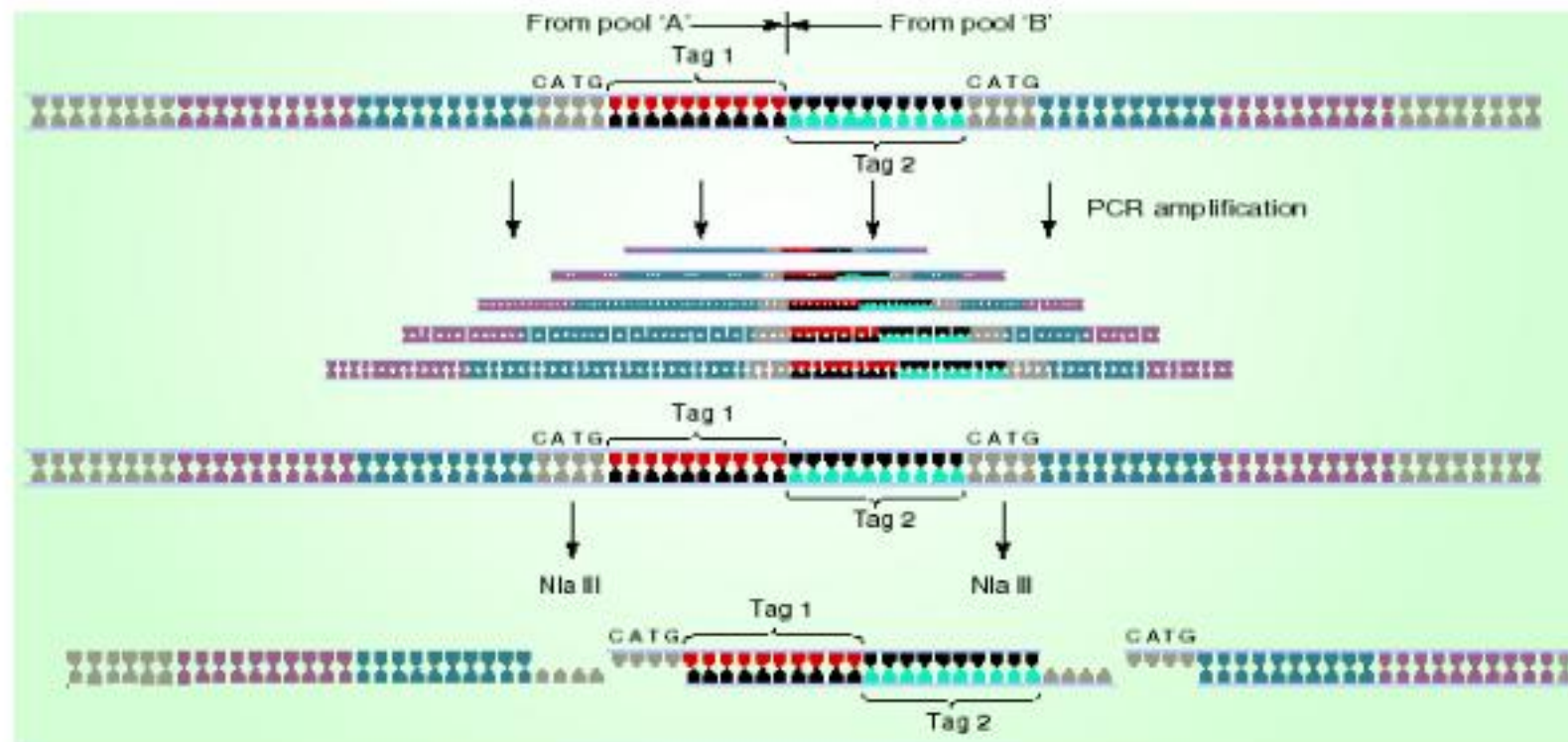


Методы массового исследования транскриптов.

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))



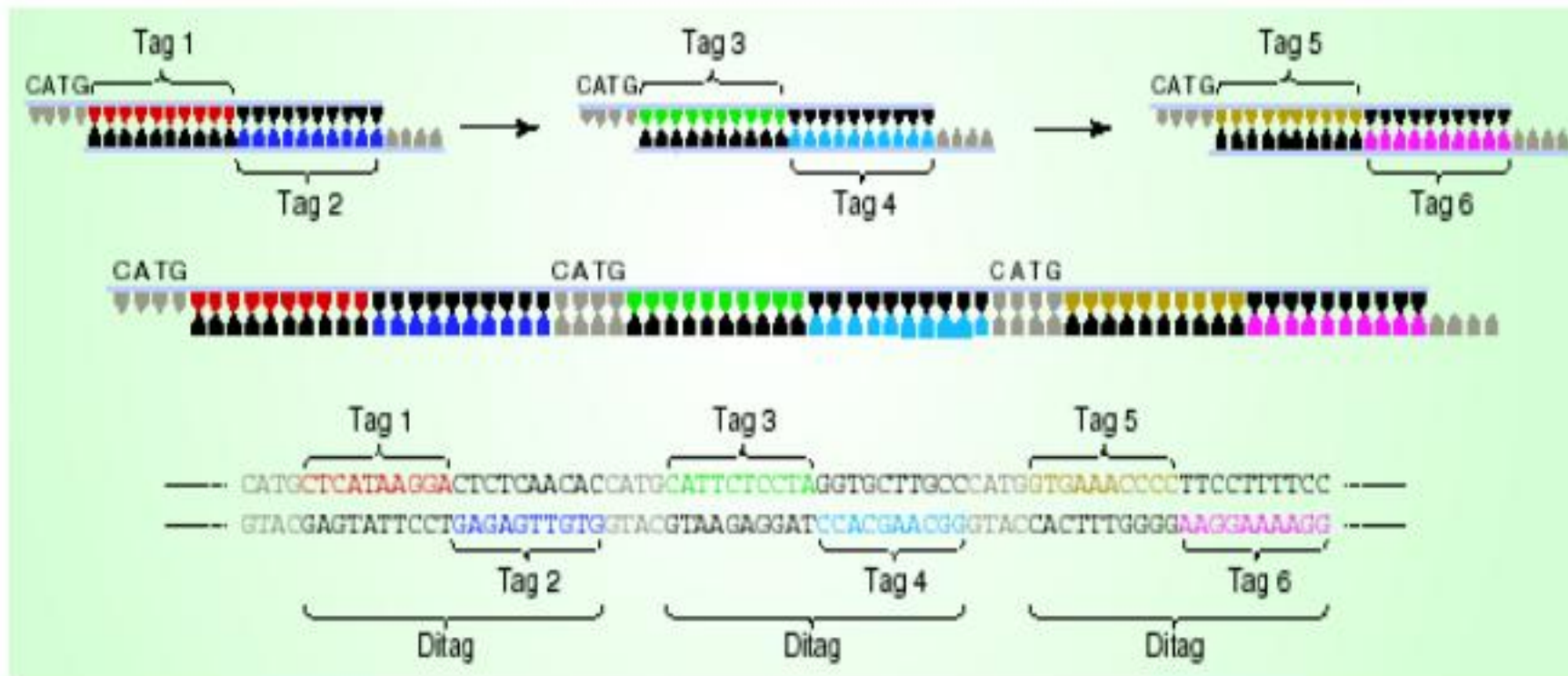


Методы массового исследования транскриптов.

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))





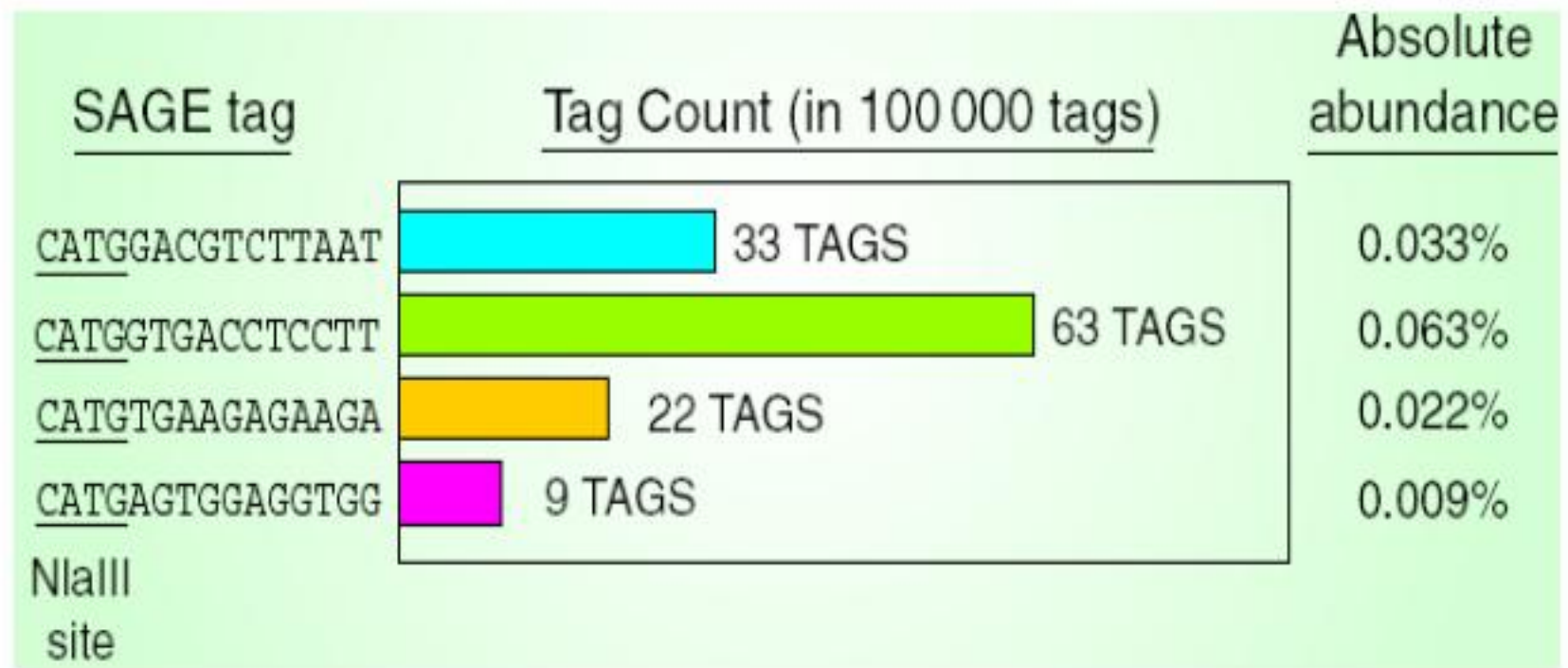
Методы массового исследования транскриптов.



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 тегов)
3. Минимизация связанных с амплификацией искажений в соотношении тегов (ампликоны почти равны по размеру ~ 100 п.о. и составу – на 70% состоят из адаптеров)



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

- **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,

Недостатки:

- 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))



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.

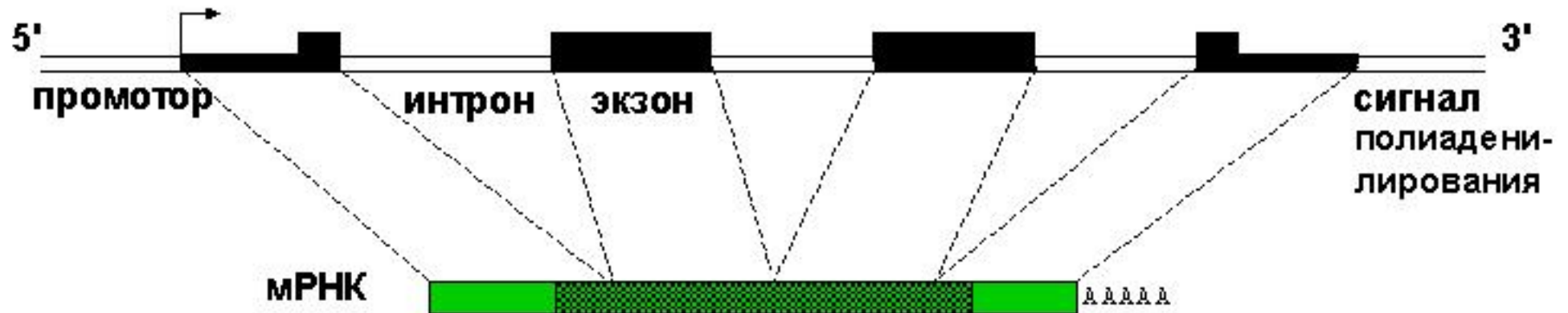
Принципы метода ESTs:

- 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 подготовленной библиотеки кДНК частотное распределение кДНК клонов в целом соответствует исходному распределению транскриптов в популяции мРНК, из которой приготовлена библиотека;
- прочитанных фрагментов должно быть очень много;
- прочтение фрагментов должно быть как можно дешевле.



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

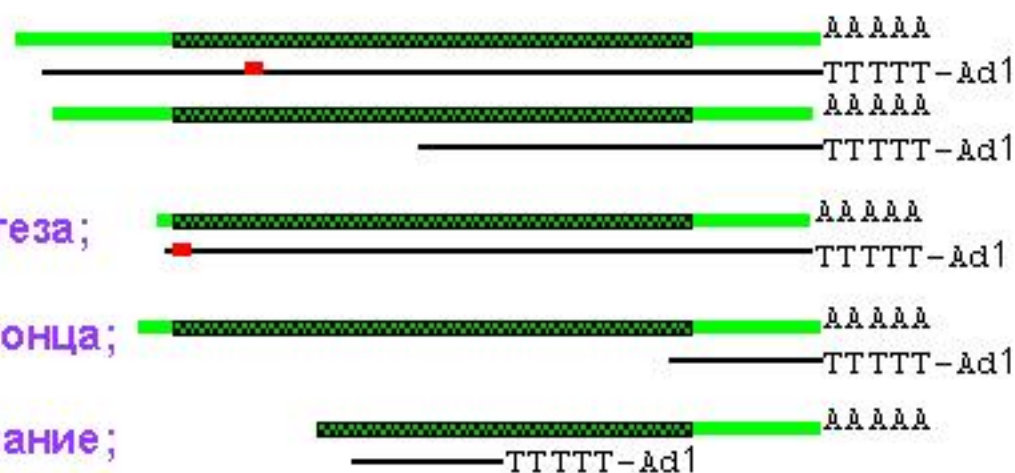
РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА А. Экспериментальная фаза А1. Приготовление кДНК библиотеки



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

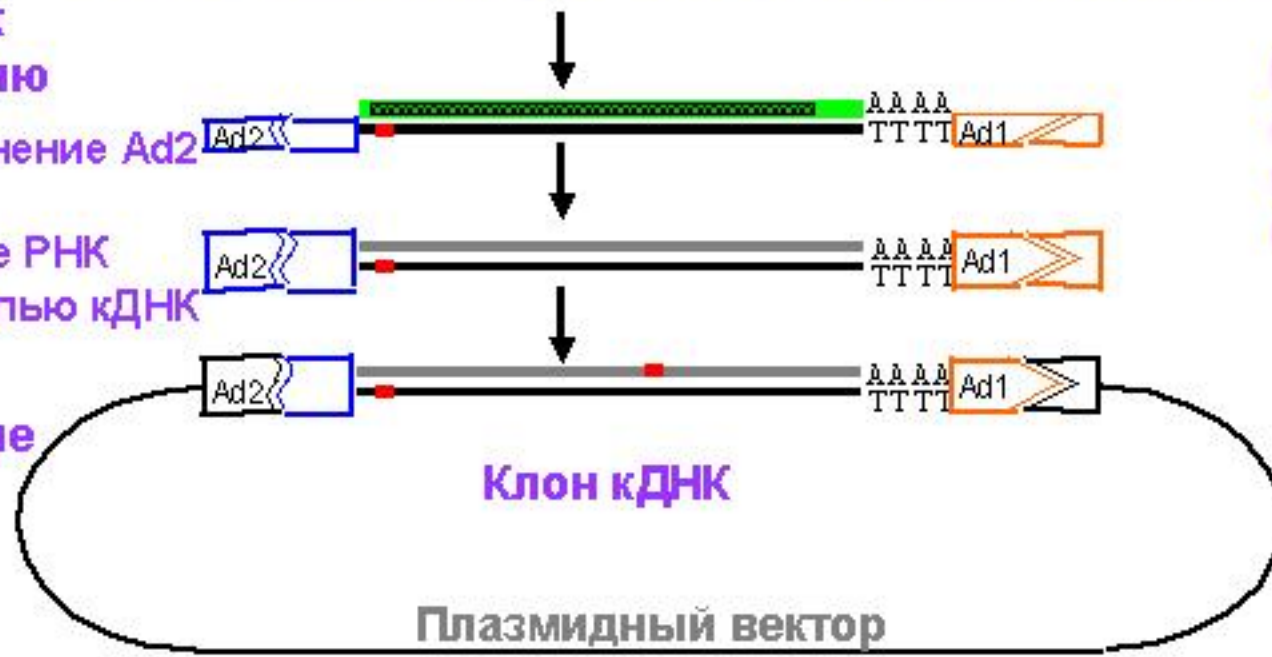
Подготовка к клонированию

- присоединение Ad2
- замещение РНК второй цепью кДНК

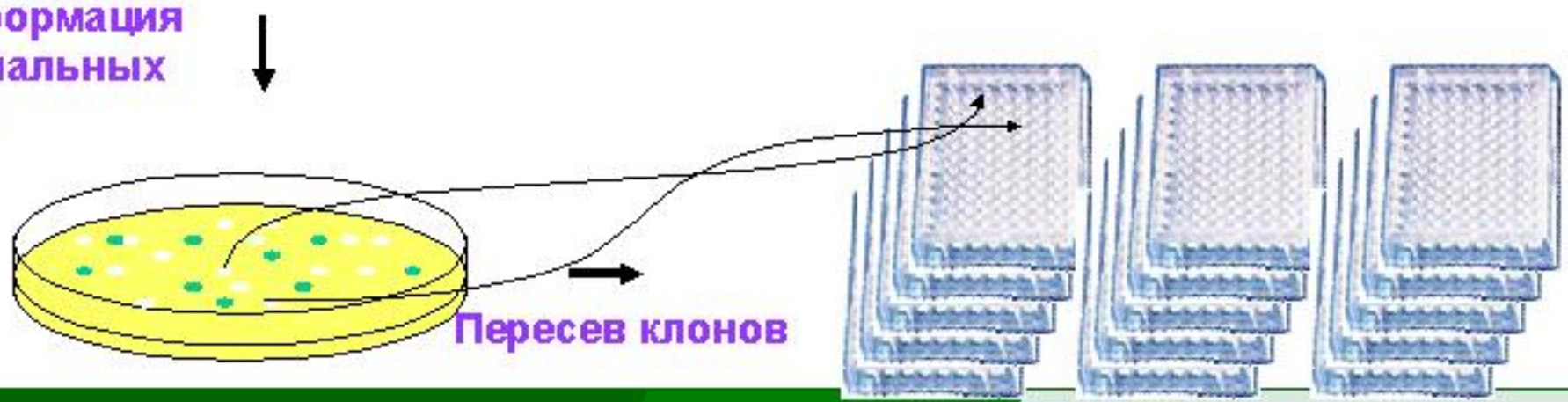
Источник отличий от исходных природных молекул:
- образование артефактных химерных молекул;

Источник ошибок:
- ошибки Таq полимеразы

Клонирование кДНК



Трансформация бактериальных клеток



кДНК библиотека

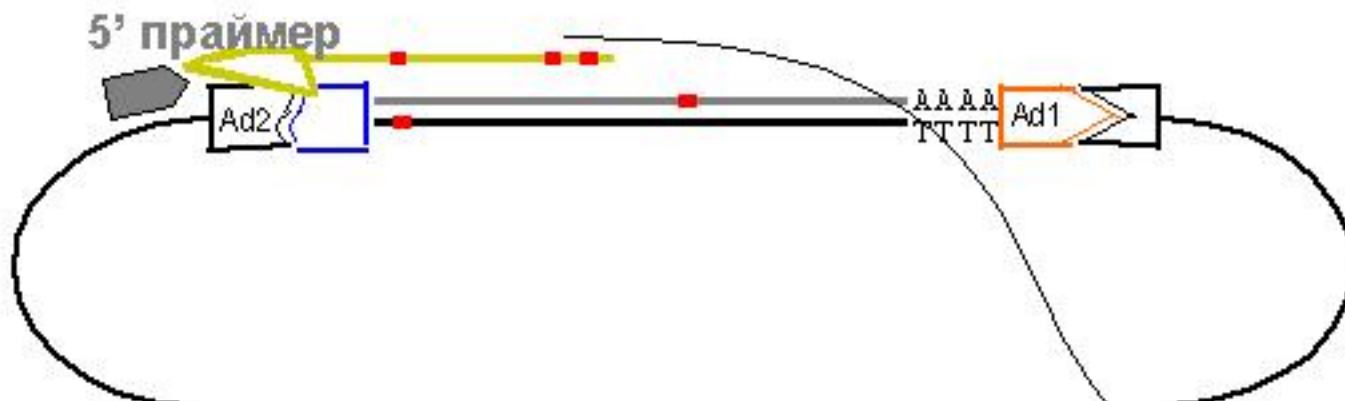


Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

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

Использование специфических праймеров к сайтам вектора,
окаймляющим встройку кДНК

Получение 5' EST



Получение 3' EST



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)

В подраздел GenBank – dbEST (или EMBL bank)



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)



Number of public entries:

18,814,550	
<i>Homo sapiens</i> (human)	5,426,047
<i>Mus musculus + domesticus</i> (mouse)	3,903,755
<i>Rattus</i> sp. (rat)	538,073
<i>Triticum aestivum</i> (wheat)	500,898
<i>Ciona intestinalis</i>	492,488
<i>Gallus gallus</i> (chicken)	451,565
<i>Zea mays</i> (maize)	383,738
<i>Danio rerio</i> (zebrafish)	362,445
<i>Hordeum vulgare + subsp. vulgare</i> (barley)	348,233
<i>Xenopus laevis</i> (African clawed frog)	344,695
<i>Glycine max</i> (soybean)	341,574
<i>Bos taurus</i> (cattle)	322,081
<i>Drosophila melanogaster</i> (fruit fly)	261,414
<i>Oryza sativa</i> (rice)	260,892
<i>Saccharum officinarum</i>	246,301
<i>Caenorhabditis elegans</i> (nematode)	215,200
<i>Silurana tropicalis</i>	203,960
<i>Arabidopsis thaliana</i> (thale cress)	190,683

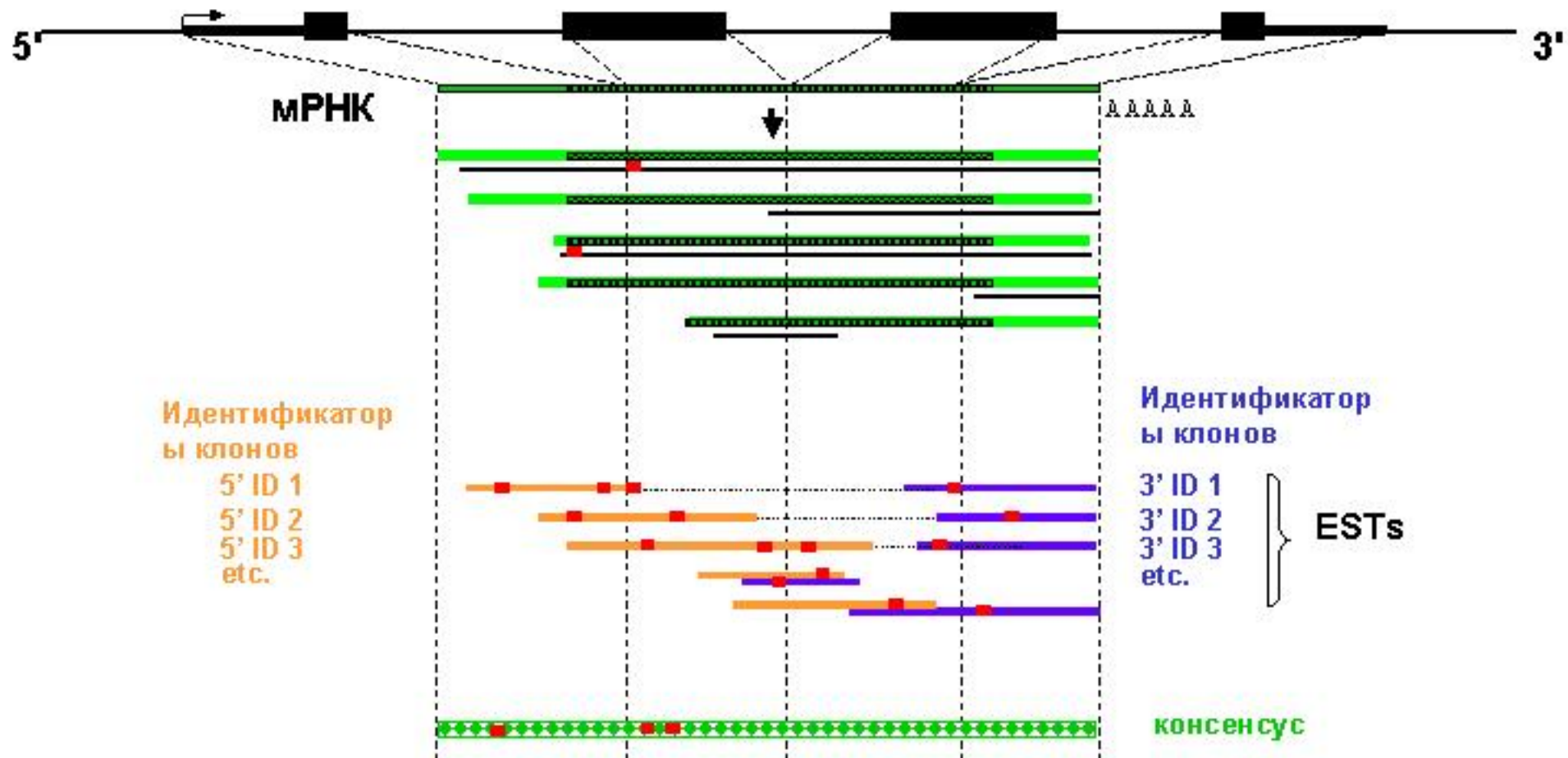
dbEST: database of
"Expressed Sequence Tags"

dbEST release 101003
Summary by Organism -
October 10, 2003



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

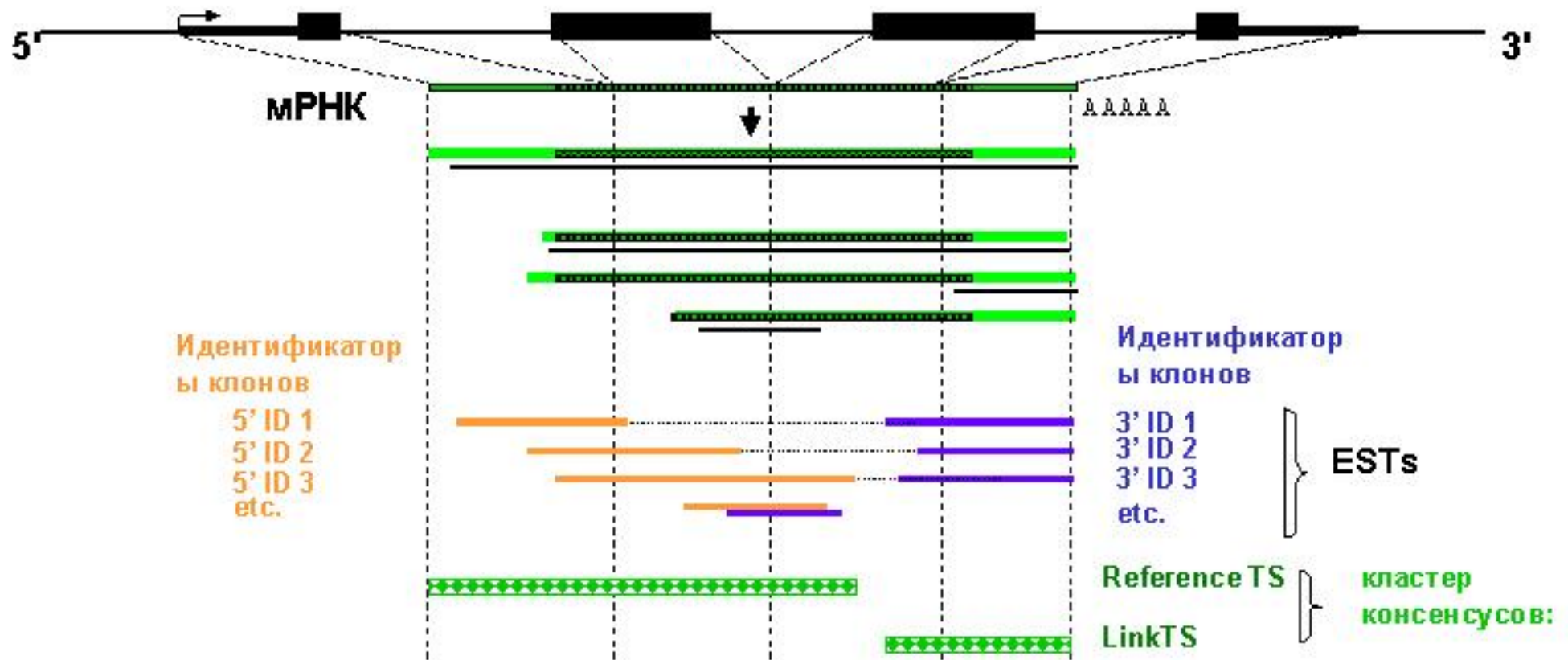
РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА Б. Компьютерная фаза Б1. Выравнивание и кластеризация EST





Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

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



Недостаточное «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.

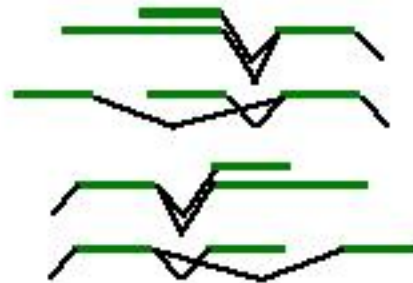


Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА

Б. Компьютерная фаза Б2. Анализ структуры кластера

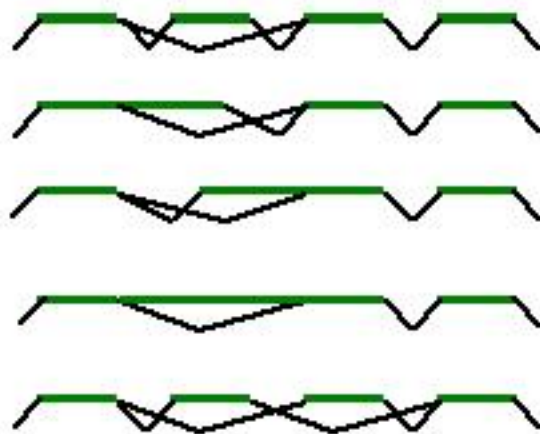
Альтернативная транскрипция



- 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)

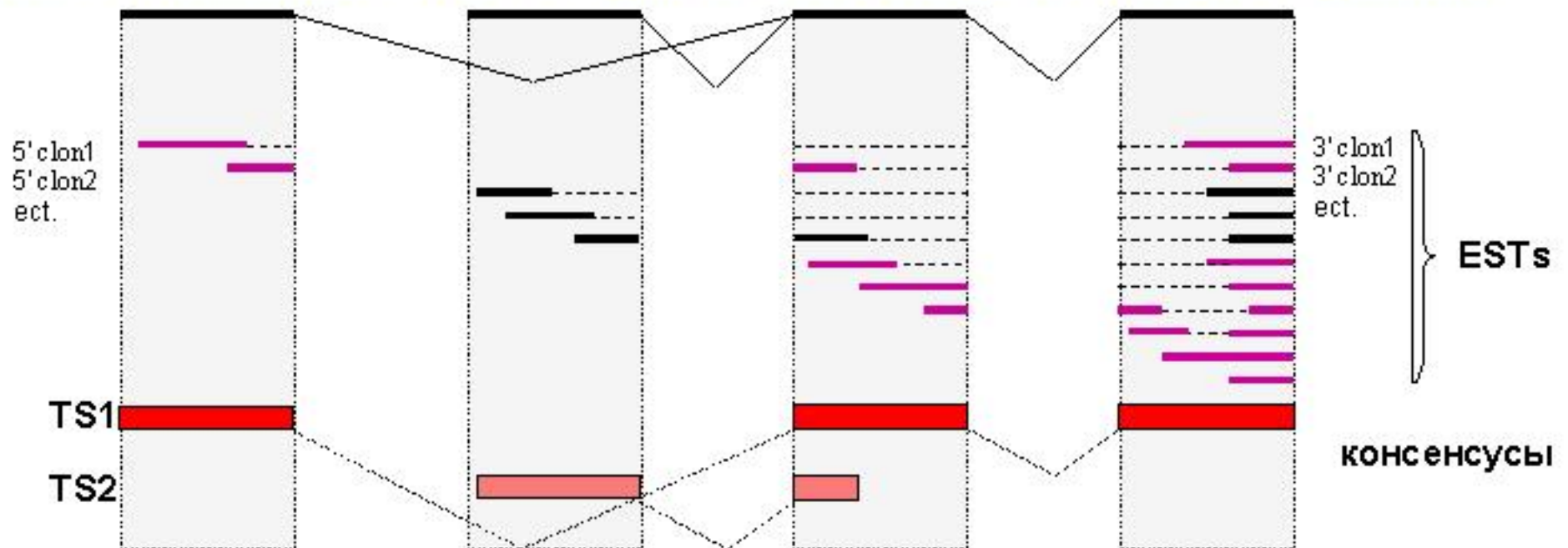
- альтернативное включение экзонов (alt. exon usage, cassette exons)



Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

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

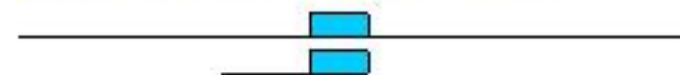
АЛЬТЕРНАТИВНЫЕ СТАРТЫ ТРАНСКРИПЦИИ В АЛЬТЕРНАТИВНЫХ 5'-ЭКЗОНАХ



DOTS BLASTViewer



NCBI "2 Seq BLAST" result

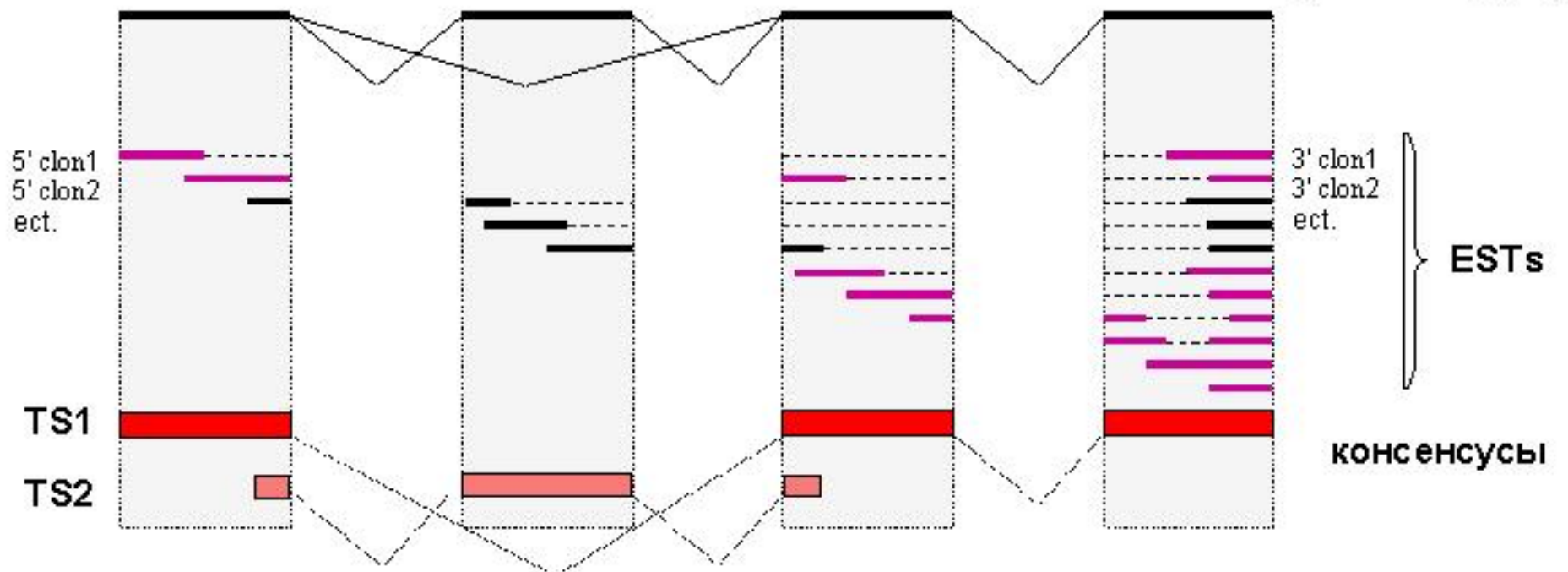




Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

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

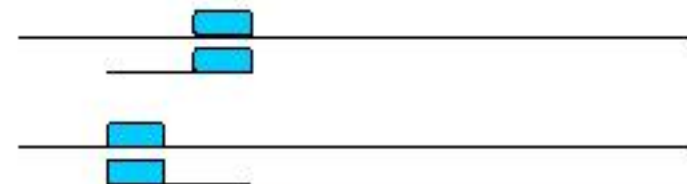
АЛЬТЕРНАТИВНЫЙ СПЛАЙСИНГ ВНУТРЕННИХ ЭКЗОНОВ: пропуск экзона (exon skipping)



DOTS BLASTViewer



NCBI "2 Seq BLAST" result



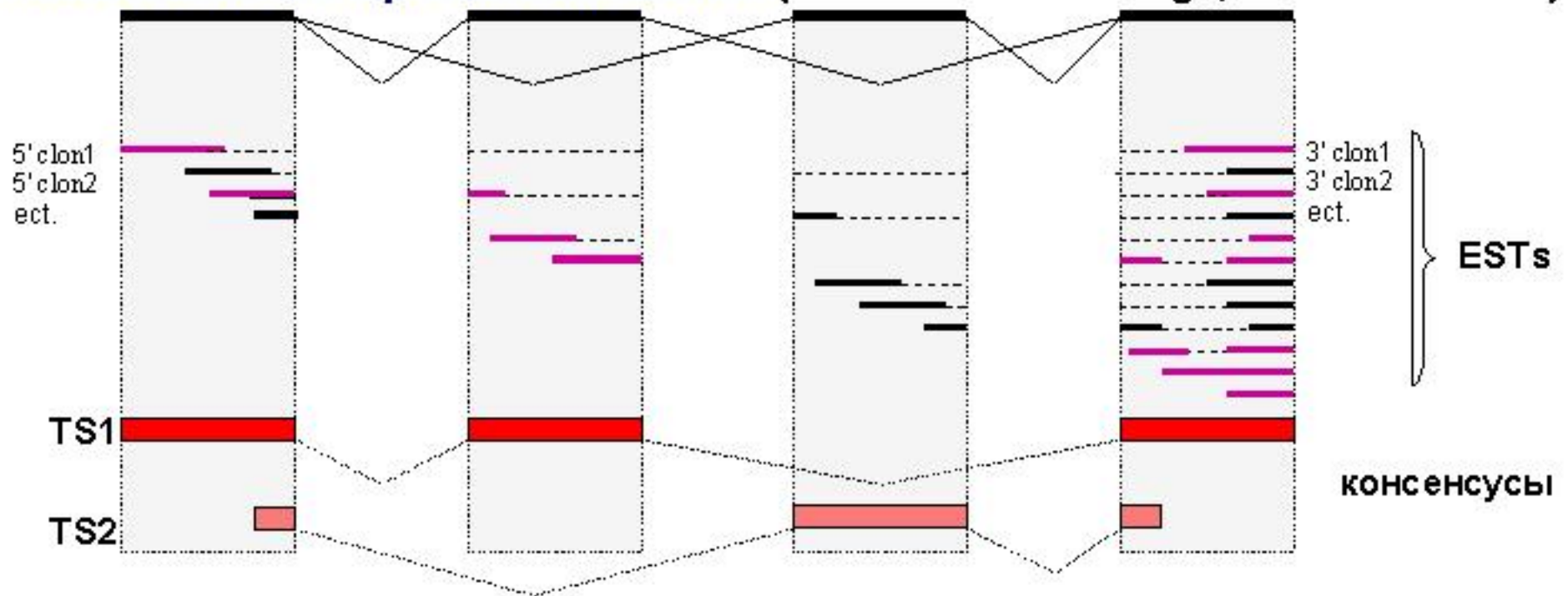


Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

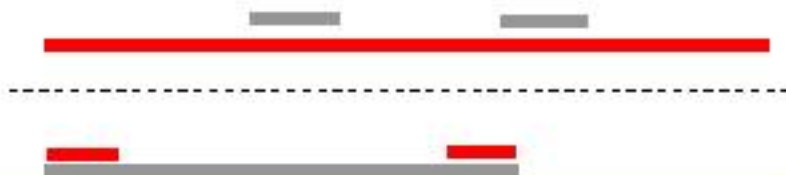
РЕКОНСТРУКЦИЯ СТРУКТУРЫ ТРАНСКРИПТОВ ГЕНА Б. Компьютерная фаза Б2. Анализ структуры кластера

АЛЬТЕРНАТИВНЫЙ СПЛАЙСИНГ ВНУТРЕННИХ ЭКЗОНОВ:

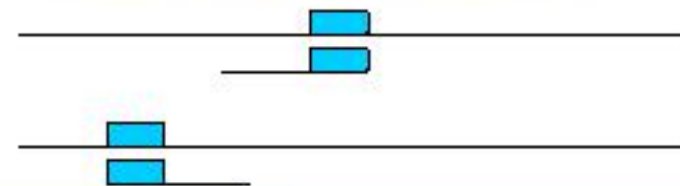
использование альтернативных экзонов (alternative exon usage, cassette exons)



DOTS BLASTViewer



NCBI "2 Seq BLAST" result

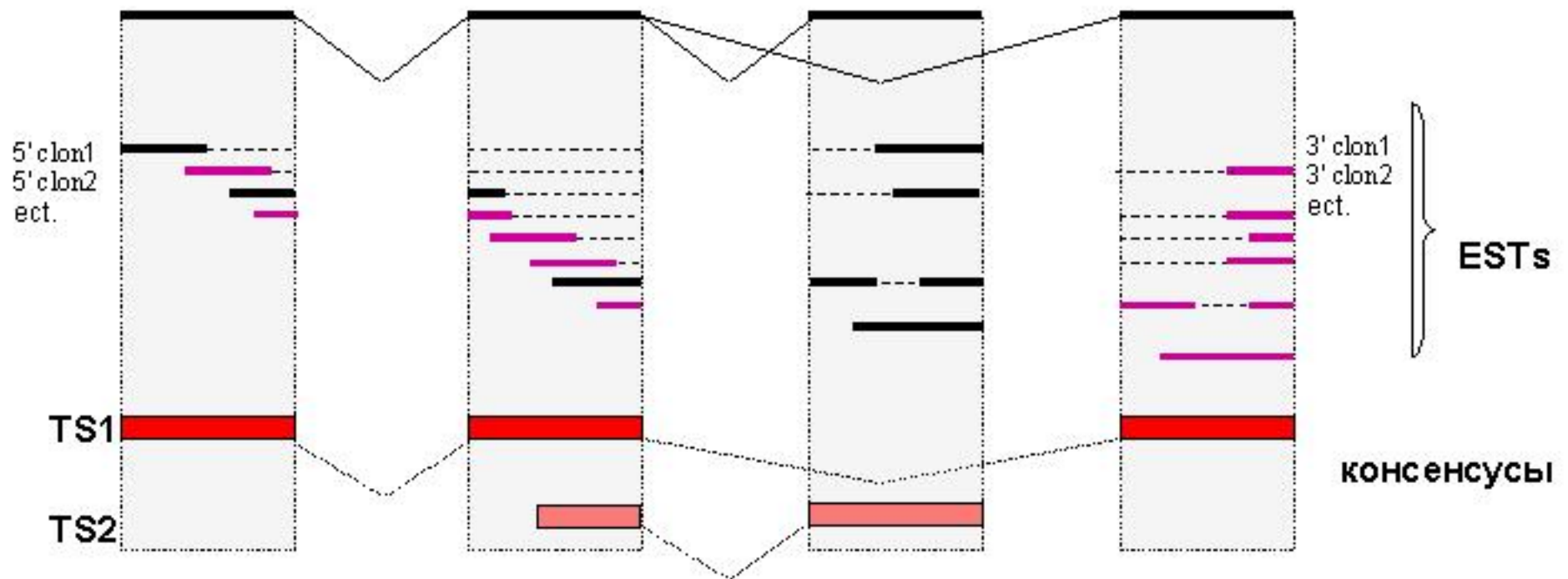




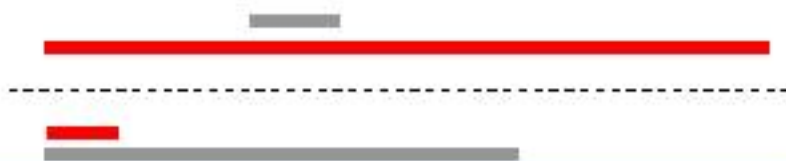
Методы массового исследования транскриптов. Expressed Sequence Tags (ESTs)

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

АЛЬТЕРНАТИВНЫЕ 3'-ЭКЗОНЫ



DOTS BLASTViewer



NCBI "2 Seq BLAST" result





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/>

Alternative Splice Forms from EST and mRNA Sequences ?

Reverse Read File Operating System:

Max. Number of lines returned:

Search for ...

Current List:

Integrated Genes - Map Service Last updated on August 01, 2002

<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

J. Drabiyuk, M. Brudno, M. S. Gelfand, M. Zorn, and I. Dubchak (2000) Nucleic Acids Research 28(1): 290-297.

M. S. Gelfand, I. Dubchak, J. Drabiyuk and M. Zorn (1999) Nucleic Acids Research, 27(1), 501.