

Молекулярная генетика и геномика

Лекция 6

Дифференциальная экспрессия генов Вычитающая
гибридизация+SAGE

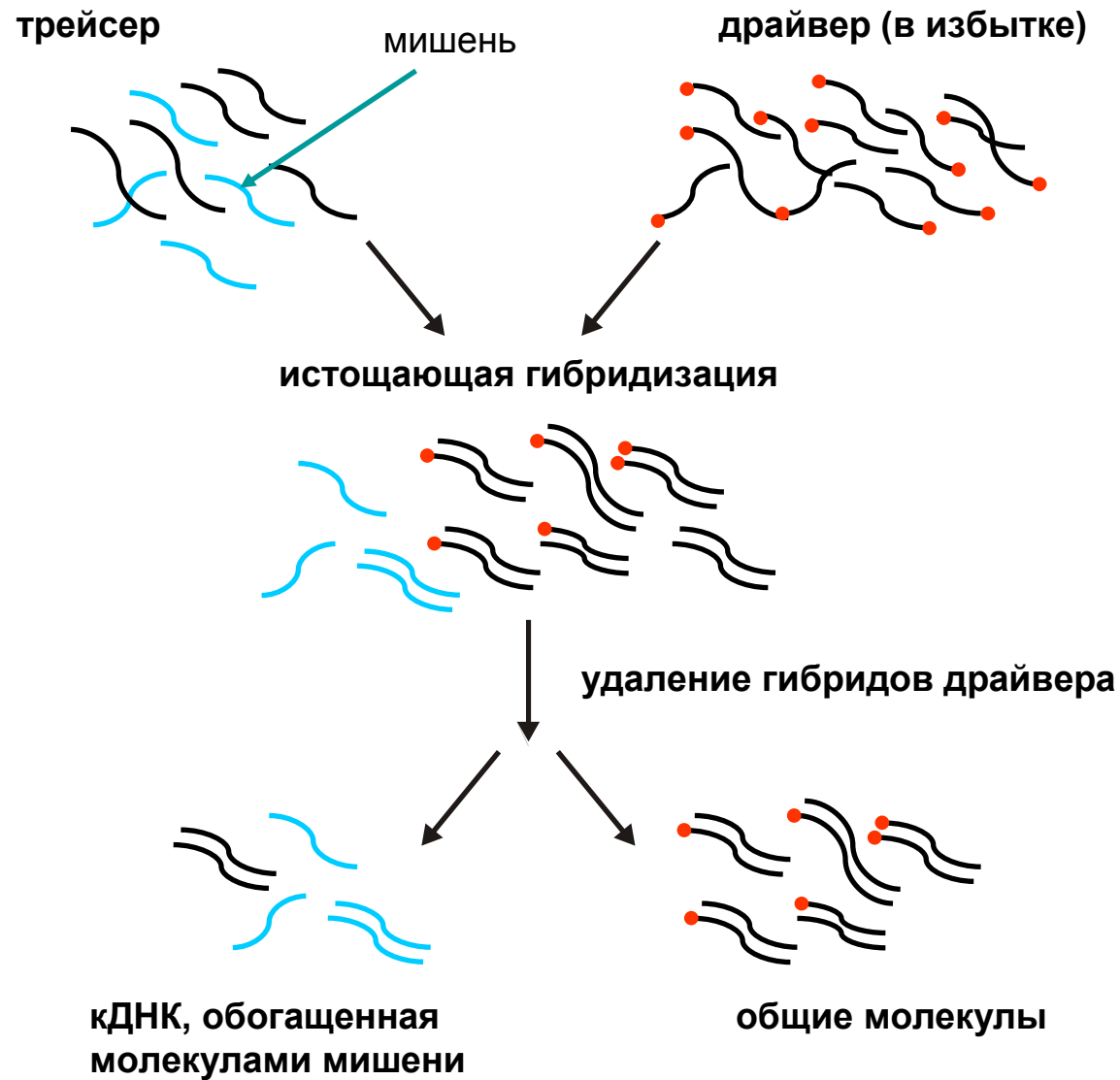
Сергей Лукьянов

Институт биоорганической химии имени академиков М.М. Шемякина и Ю.А.
Овчинникова РАН

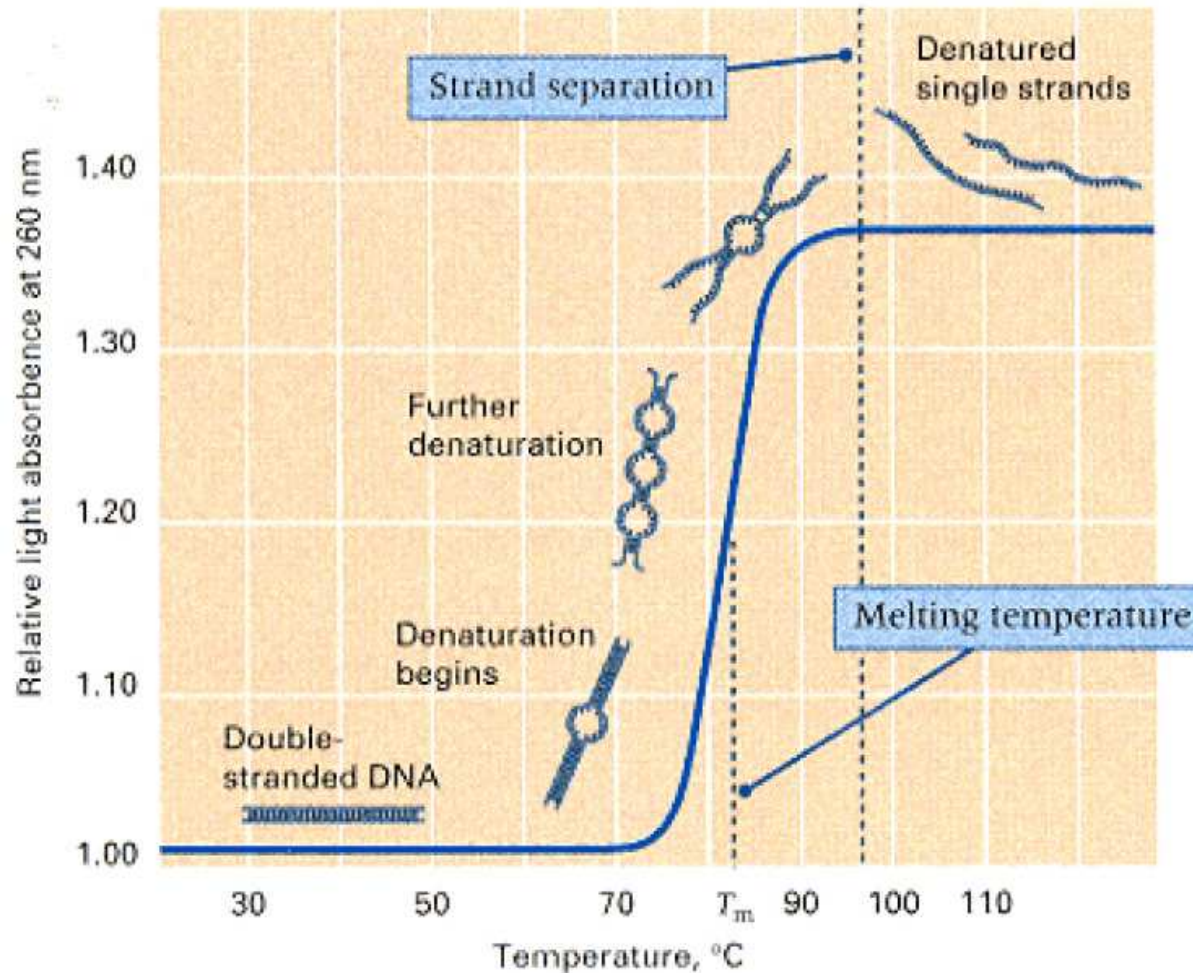


Вычитающая гибридизация

Стратегия вычитающей гибридизации



Плавление (денатурация) ДНК



Температура плавления ДНК (T_m) – это температура, при которой цепи ДНК диссоциированы наполовину

Факторы, влияющие на T_m :

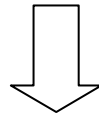
- pH
- Ионная сила
- Органические растворители
- **Наличие неспаренных оснований:** 1% неспаренных оснований снижает T_m на 1°C

Реассоциации нуклеиновых кислот – реакция второго порядка

C = концентрация одноцепочечной (ss) ДНК во время t (моль нуклеотидов/литр).

Скорость уменьшения C в ходе гибридизации описывается формулой для реакций второго порядка:

$$\frac{-dC}{dt} = kC^2 \quad \text{or} \quad \frac{dC}{C^2} = -kdt$$



$$\boxed{\frac{C}{C_0} = \frac{1}{1 + kC_0 t}}$$

$$\int x^{-2} dx = \frac{x^{-1}}{-2+1}$$

$$-\frac{dC}{C^2} = k dt$$

$$-\int_{C_0}^C \frac{dC}{C^2} = \int_0^t k dt$$

$$\frac{1}{C} - \frac{1}{C_0} = kt$$

$$\frac{C_0}{C} - 1 = C_0 kt$$

$$\frac{C_0}{C} = C_0 kt + 1$$

$$\frac{C}{C_0} = \frac{1}{C_0 kt + 1}$$

Время полуреассоциации обратно пропорционально константе скорости

$$\frac{C}{C_0} = \frac{1}{1 + kC_0t}$$

C = концентрация одноцепочечной (ss) ДНК во время t (моль нуклеотидов/литр).

Время полуренатурации

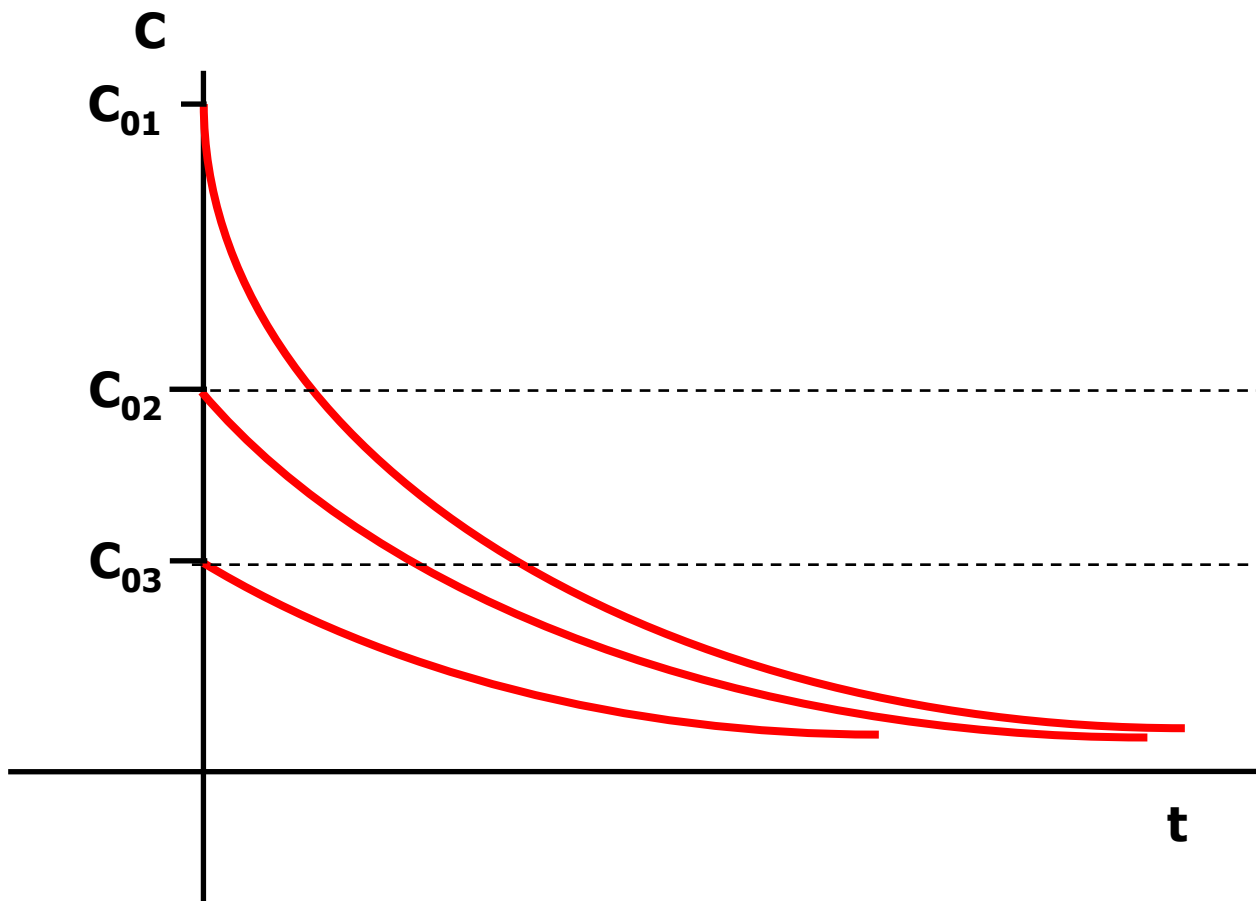
$$\frac{C}{C_0} = 0.5, \text{ and } t = t_{1/2}$$

$$C_0 t_{1/2} = \frac{1}{k}$$

k литры*(моль нт)⁻¹ сек⁻¹

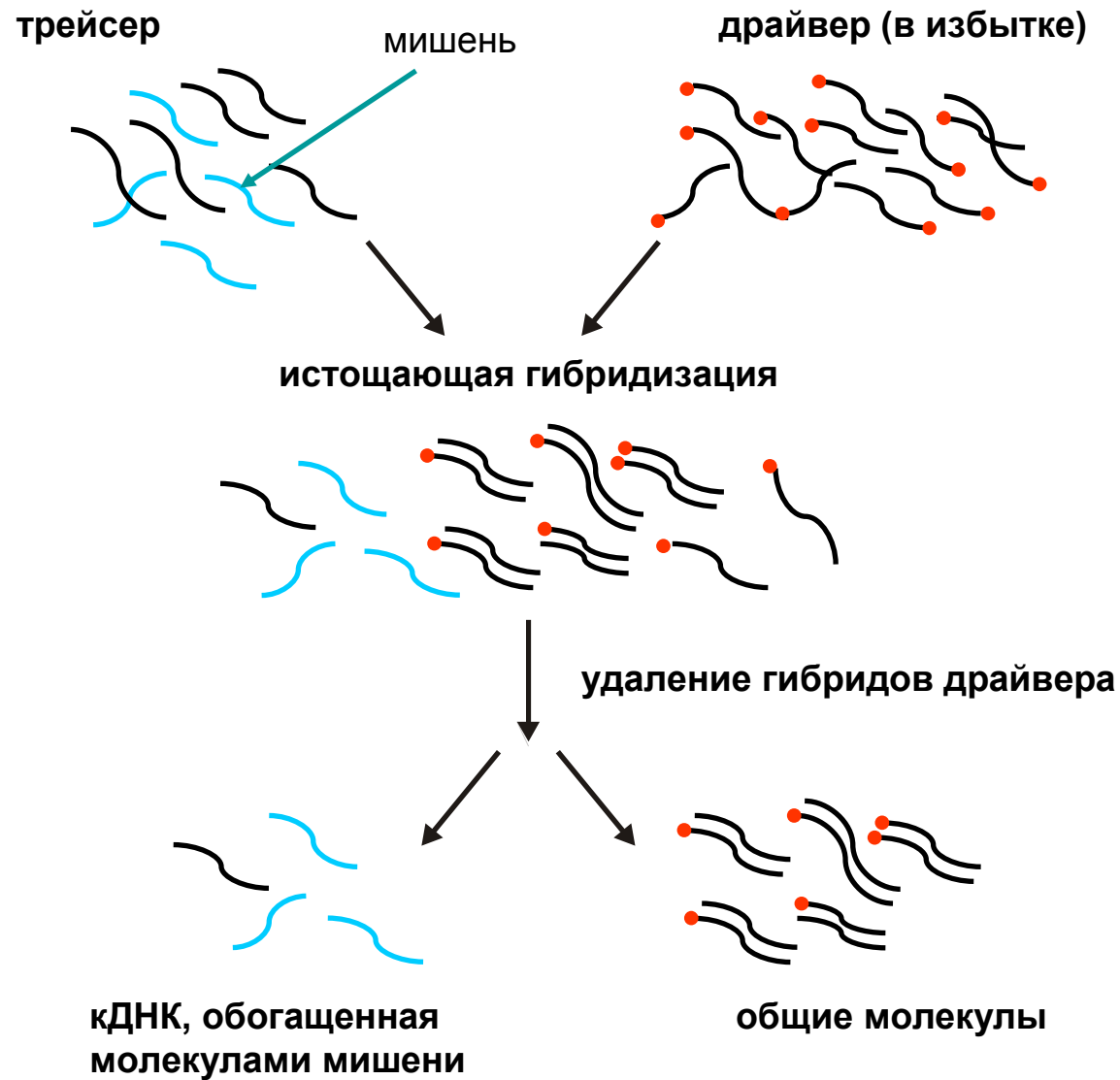
$$t_{1/2} = \frac{1}{kC_0}$$

Реассоциация индивидуальной последовательности ДНК



C_0 - концентрация ДНК (нуклеотиды моль/л), t - время (сек)

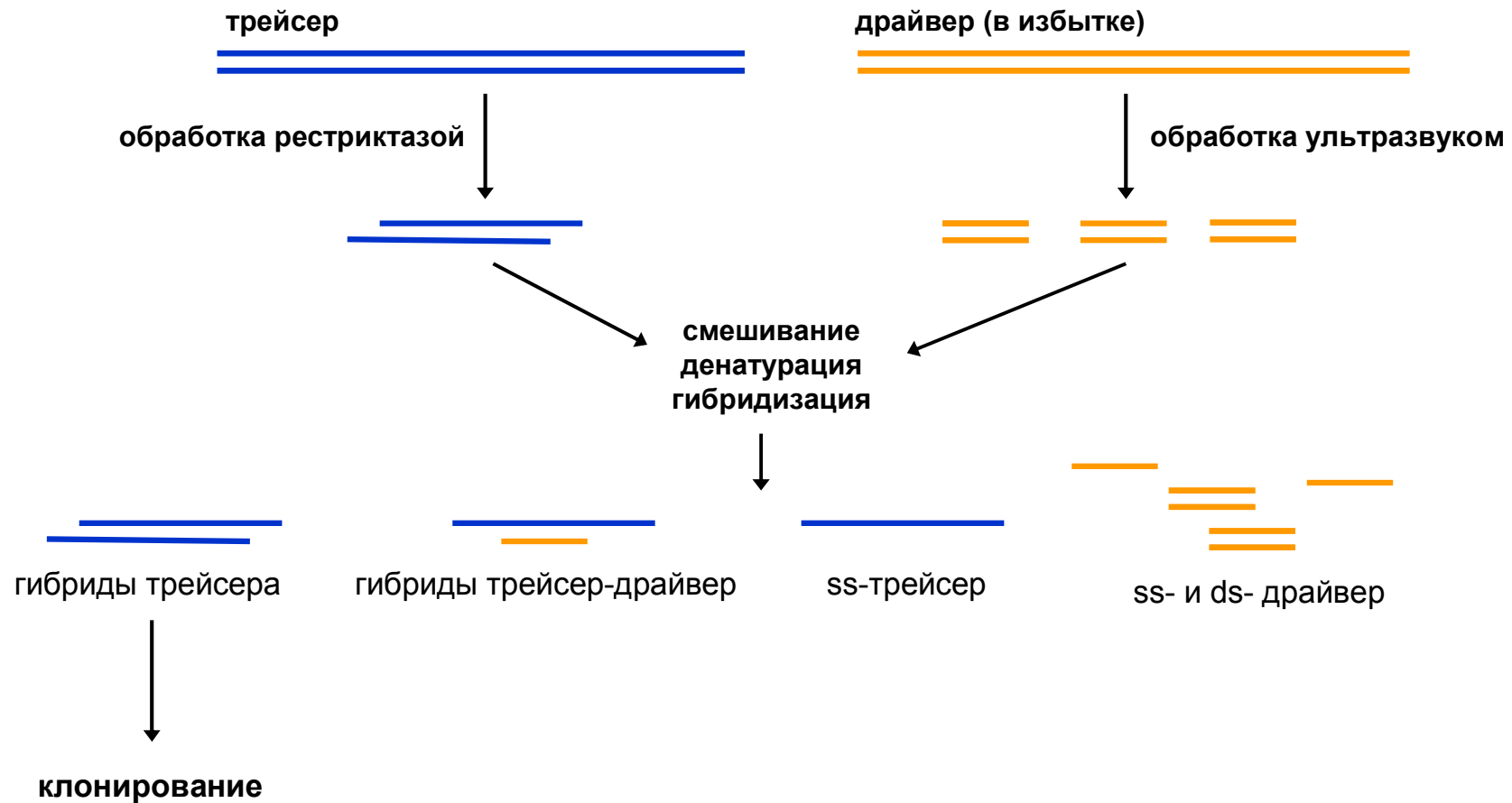
Стратегия вычитающей гибридизации



Расчет обогащения для вычитающей гибридации на основе ss ДНК

- ОБОГАЩЕНИЕ:
- $E = C\text{-target}/C_0\text{-target}$
- или
 $E = \% \text{ target после вычитания} / \% \text{ targ в исходнов образце ДНК}$
- При существенном избытке драйвера обогащение происходит в линейной зависимости от времени и не имеет теоретического предела
т.е. C-target стремится к 100% !
- Реально обогащение всегда ниже, поскольку НЕВОЗМОЖНО:
 1. приготовить чистую ss фракцию ДНК
 2. Провести разделение гибридов с чистотой выше 95%

Схема метода вычитающей гибридизации по Виглеру

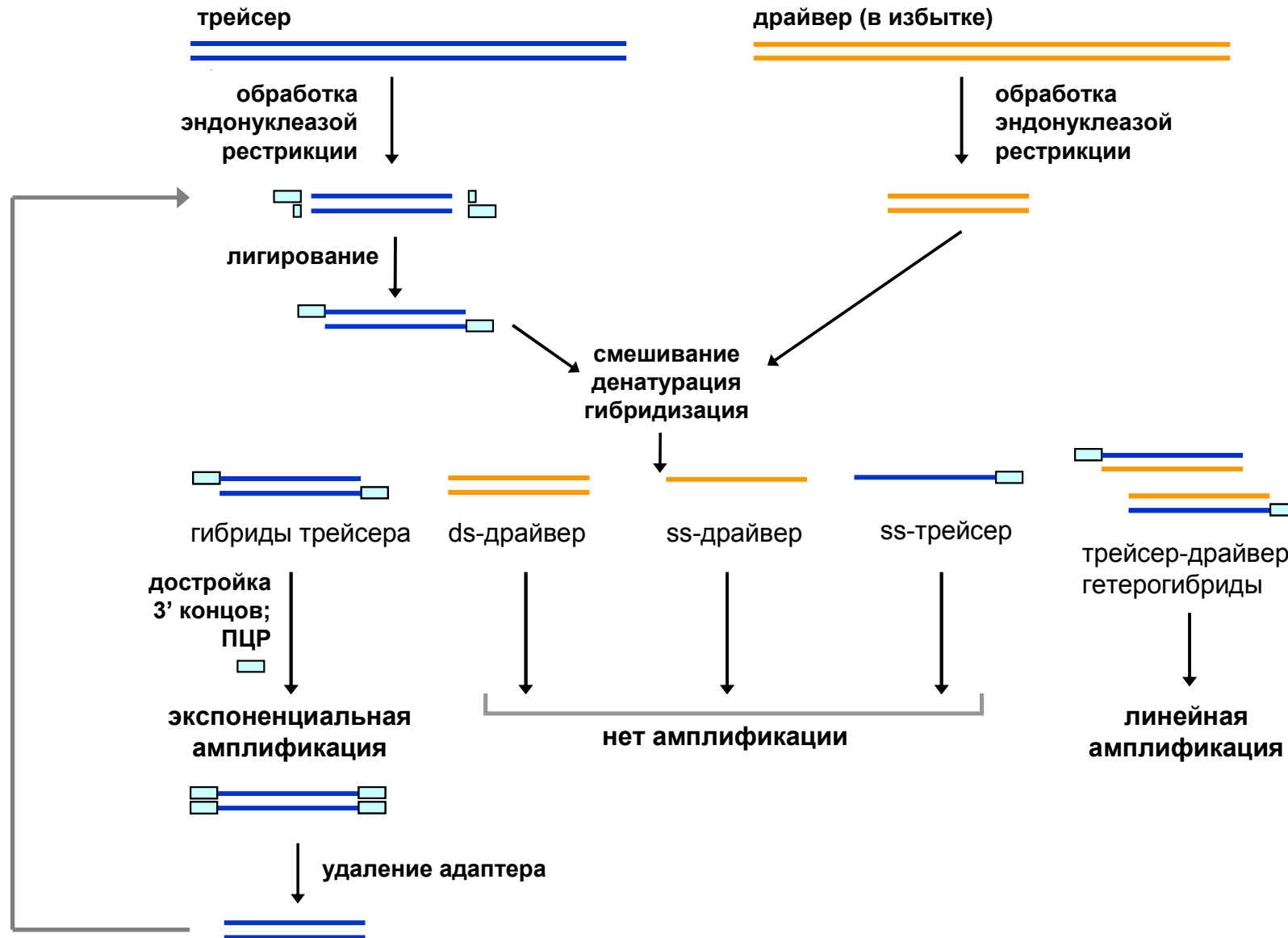




Расчет обогащения для вычитающей гибридизации на основе ds ДНК

- Условный расчет исходя из прохождения реакции гибридизации на 100%
- $E = C\text{-драйвер} / C\text{-трайсер}$
- Реально обогащение всегда ниже, поскольку гибридизация не проходит на 100% НИКОГДА!


Схема метода вычитающей гибридизации RDA





Расчет обогащения для 2-х циклов вычитающей гибридизации ДНК RDA

- Условный расчет исходя из прохождении реакции гибридизации на 100% (Реально обогащение всегда ниже, поскольку гибридизация не проходит на 100% НИКОГДА!)
- 1-ый цикл:
- $E1 = \text{Сдрайвер} / \text{Страйсер} = 50 / 1 = \mathbf{50}$
- 2-ой цикл:
- $E2 = E1 \times (\text{Сдр} / \text{Стр}) = 50 \times (50 / 0.02) = 50 \times 2500 = \mathbf{125\ 000}$



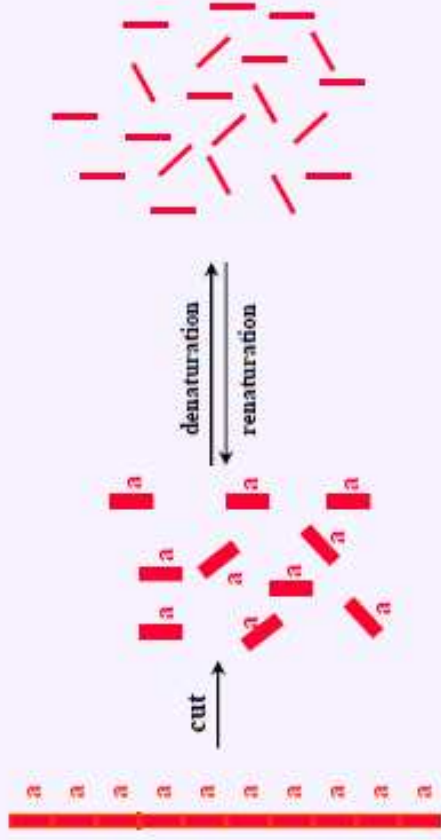
Особенности ВЫЧИТАЮЩЕЙ ГИБРИДИЗАЦИИ кДНК – необходимость выравнивания (нормализации) концентраций различных генетических последовательностей

- Проведение нормализации невозможно до процедуры «вычитания», поскольку теряется дифференциальность
- Проведение нормализации невозможно после «вычитания», поскольку «фоновые» молекулы в ходе нормализации восстановят свою концентрацию в образце кДНК

Explanation of the term “complexity of the genome”

DNA with repetitive sequence.

This DNA will renature 10 times faster than the DNA shown below. Its complexity is one tenth of that of the DNA with unique sequence.



DNA with unique sequence. Its complexity is high.

Note that the size of the genome by itself does not determine its complexity. The two hypothetical genomes have the same size.



Время полуреассоциации обратно пропорционально константе скорости

$$\frac{C}{C_0} = \frac{1}{1 + kC_0t}$$

C = концентрация одноцепочечной (ss) ДНК во время t (моль нуклеотидов/литр).

Время полуренатурации $\frac{C}{C_0} = 0.5$, and $t = t_{1/2}$

$$C_0 t_{1/2} = \frac{1}{k}$$

k литры*(моль нт)⁻¹ сек⁻¹

$$t_{1/2} = \frac{1}{kC_0}$$

Константа скорости k обратно пропорциональна сложности ДНК

$$k \propto \frac{\sqrt{L}}{N} \quad L = \text{длина}; N = \text{сложность}$$

Эмпирически k можно подсчитать как

$$k = 3 \times 10^5 \frac{\sqrt{L}}{N}$$

$$t_{1/2} = \frac{1}{kC_0} = \frac{N}{\sqrt{LC_0}}$$

в 1.0 M Na⁺ при $T = T_m - 25^\circ\text{C}$

Время полуреассоции прямо пропорционально сложности ДНК

$$C_0 t_{1/2} \propto \frac{N}{\sqrt{L}}$$

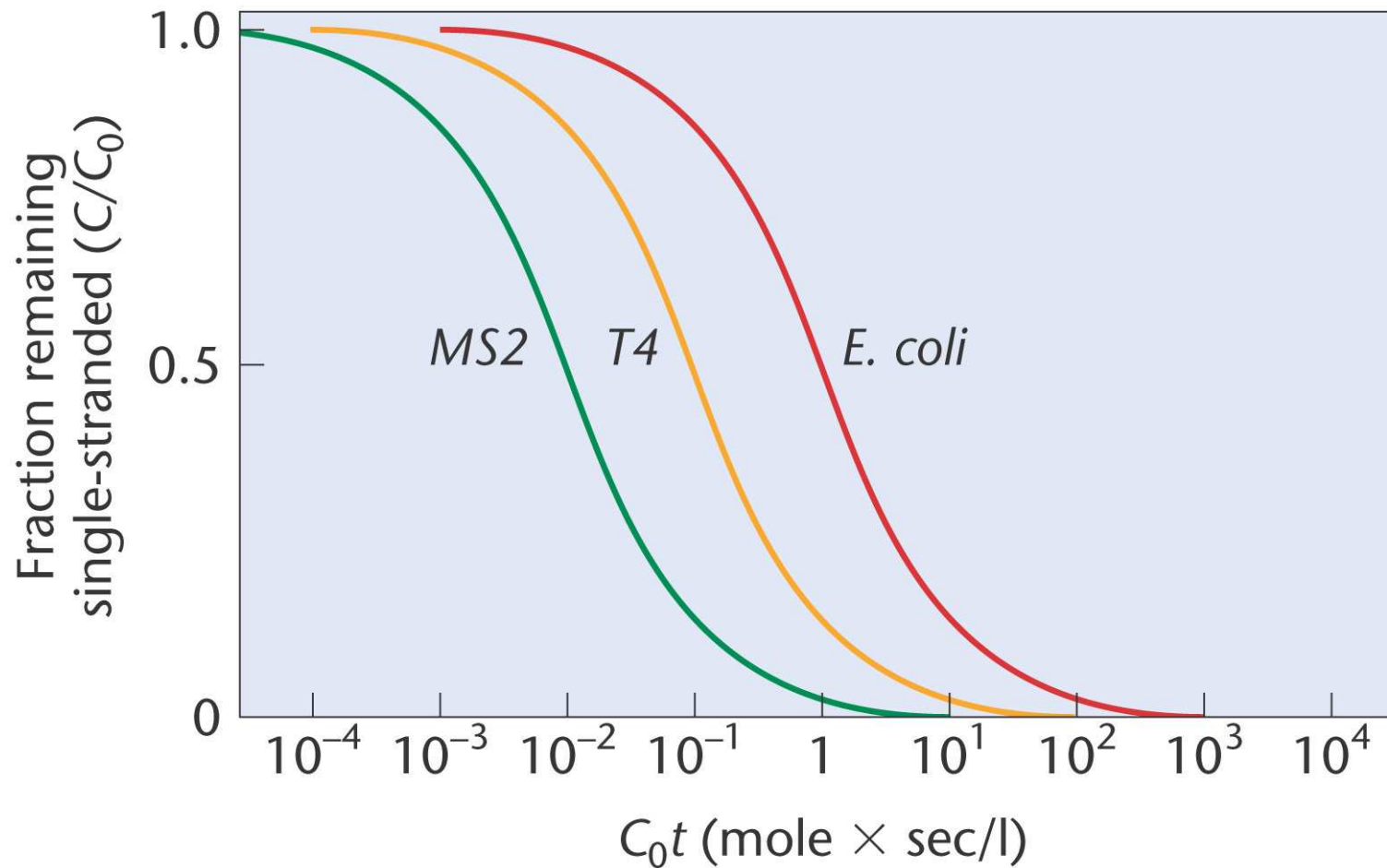
Для анализа сложности ДНК можно использовать фрагменты фиксированной длины (например, 400 bp). Тогда:

$$C_0 t_{1/2} \propto N$$

$$\frac{N^{unknown}}{N^{standard}} = \frac{C_0 t_{1/2}^{unknown}}{C_0 t_{1/2}^{standard}}$$

Для *E. coli* $N = 4.639 \times 10^6$ bp

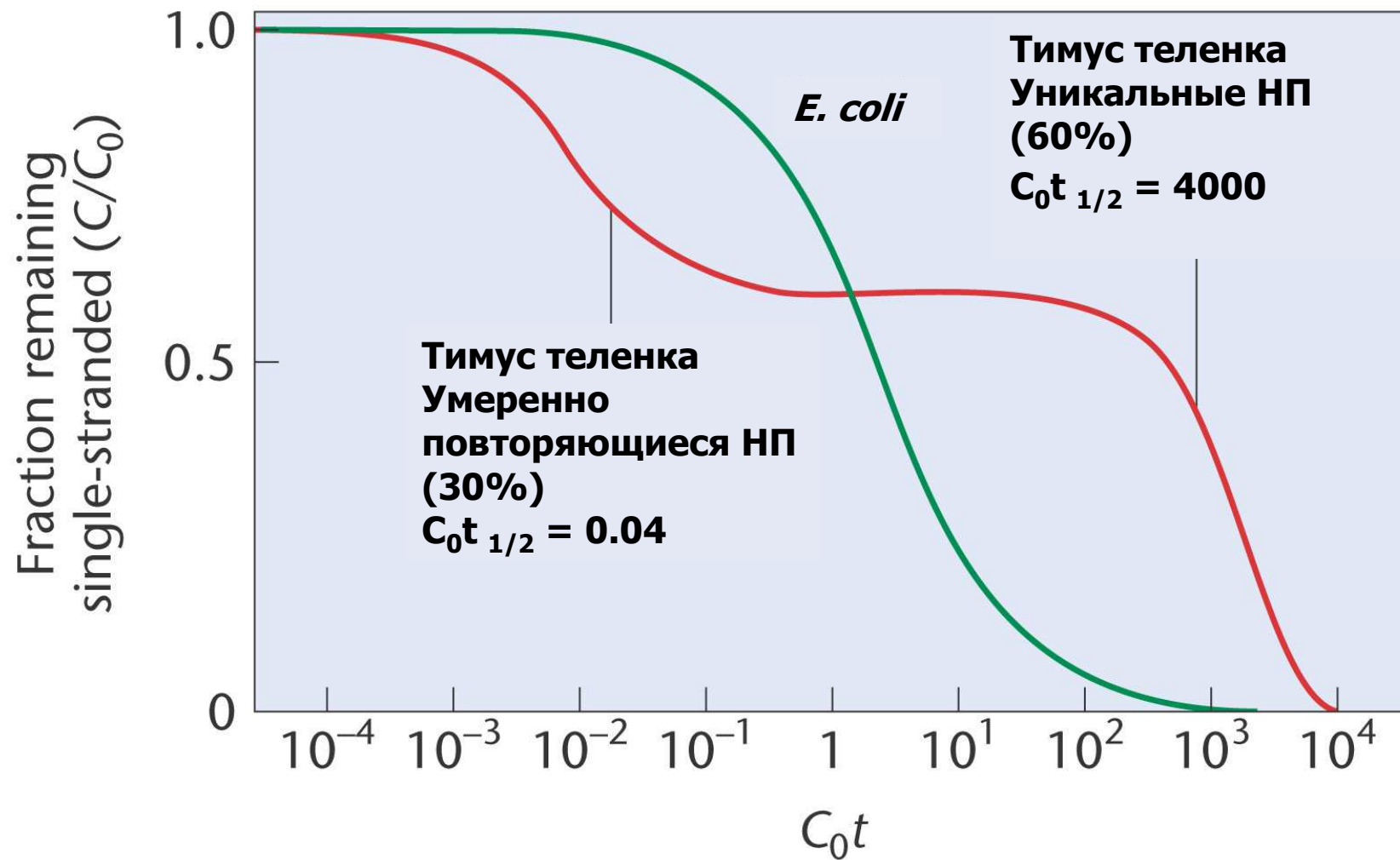
Скорость реассоциации зависит от сложности ДНК



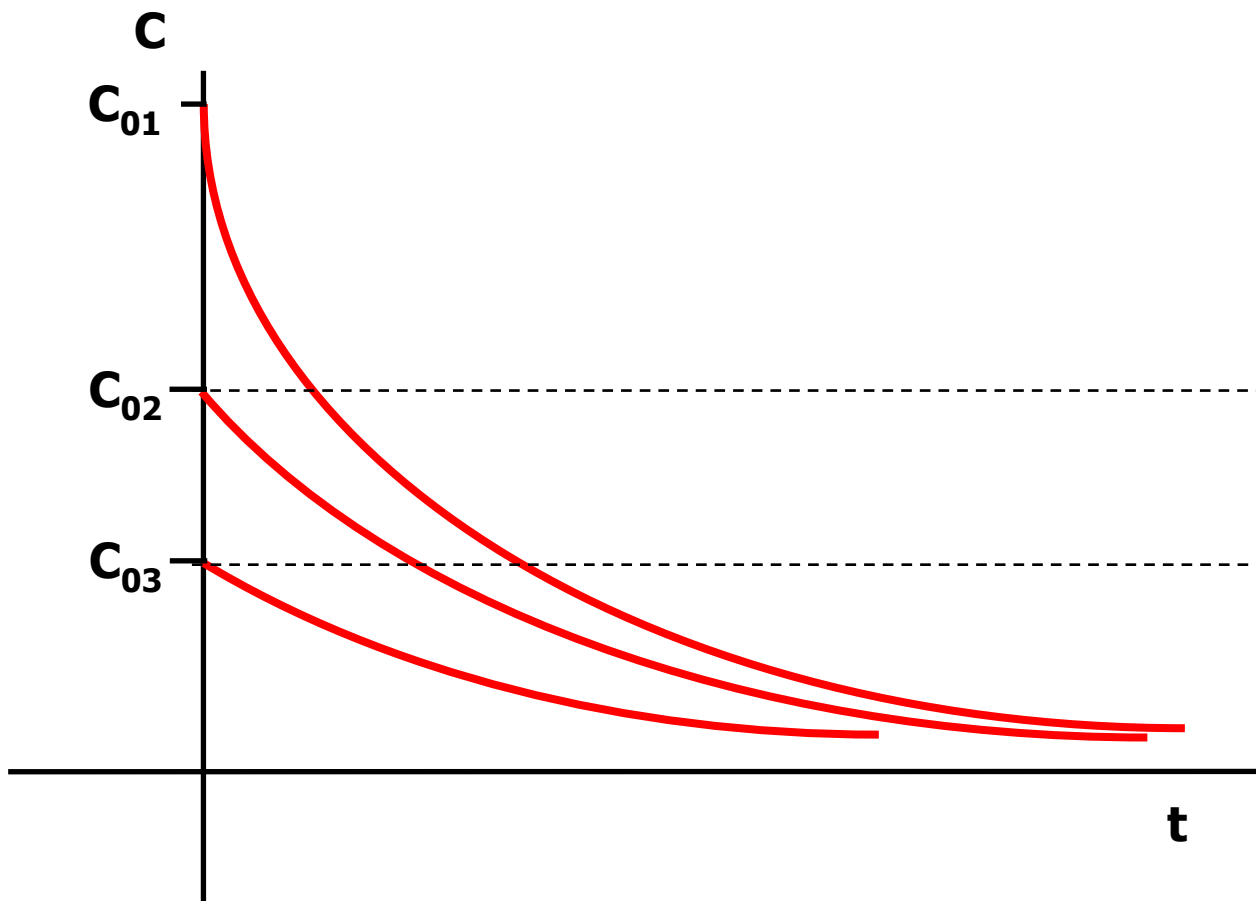
Copyright © 2006 Pearson Prentice Hall, Inc.

C_0 - концентрация ДНК (нуклеотиды моль/л), t - время (сек)

Кривая реассоциации ДНК, содержащей повторяющиеся элементы



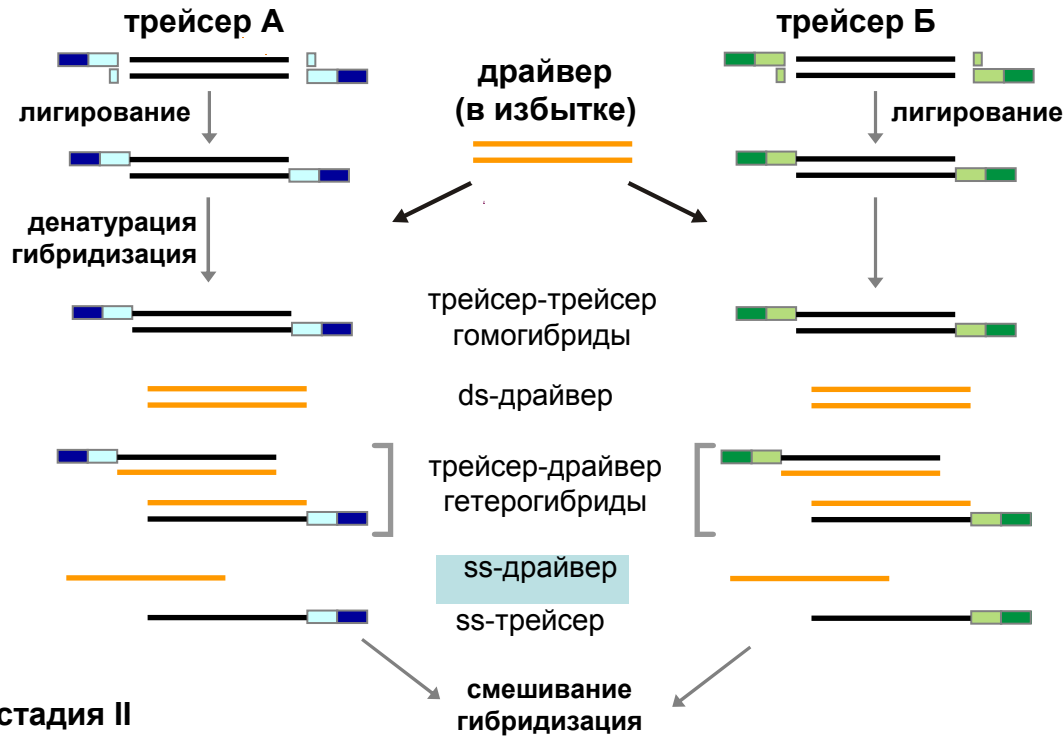
Реассоциация индивидуальной последовательности ДНК



C_0 - концентрация ДНК (нуклеотиды моль/л), t - время (сек)

Схема метода вычитающей гибридизации SSH (suppression subtractive hybridization)

стадия I



стадия II

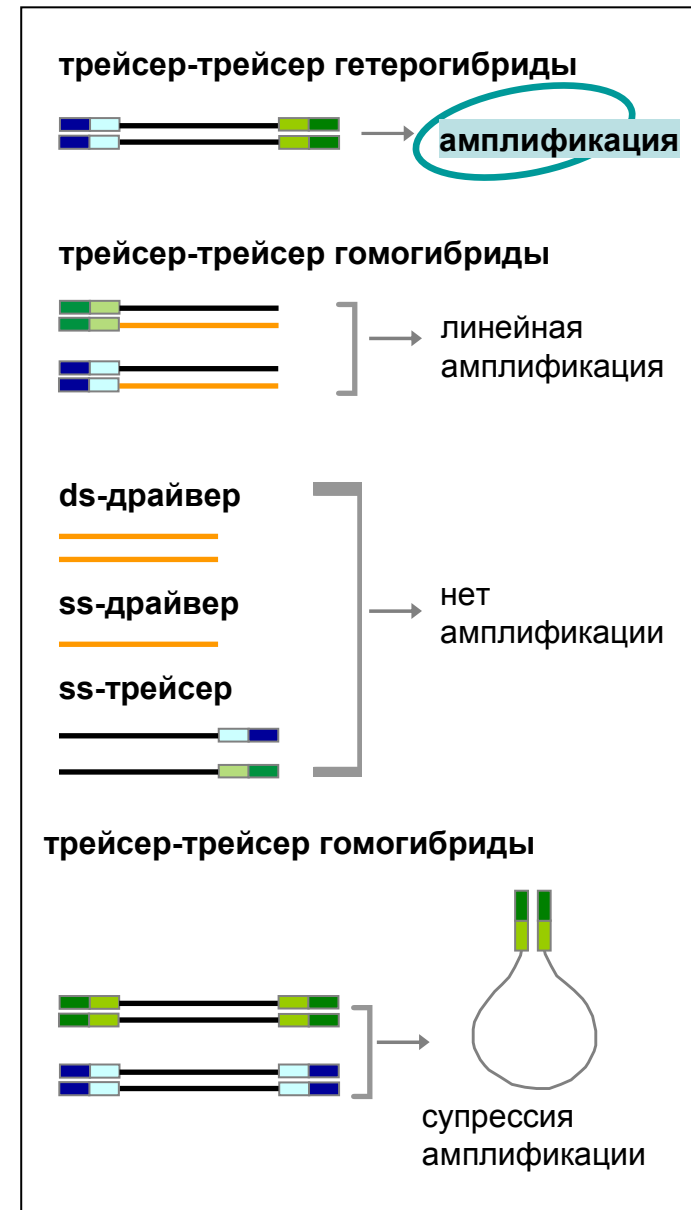
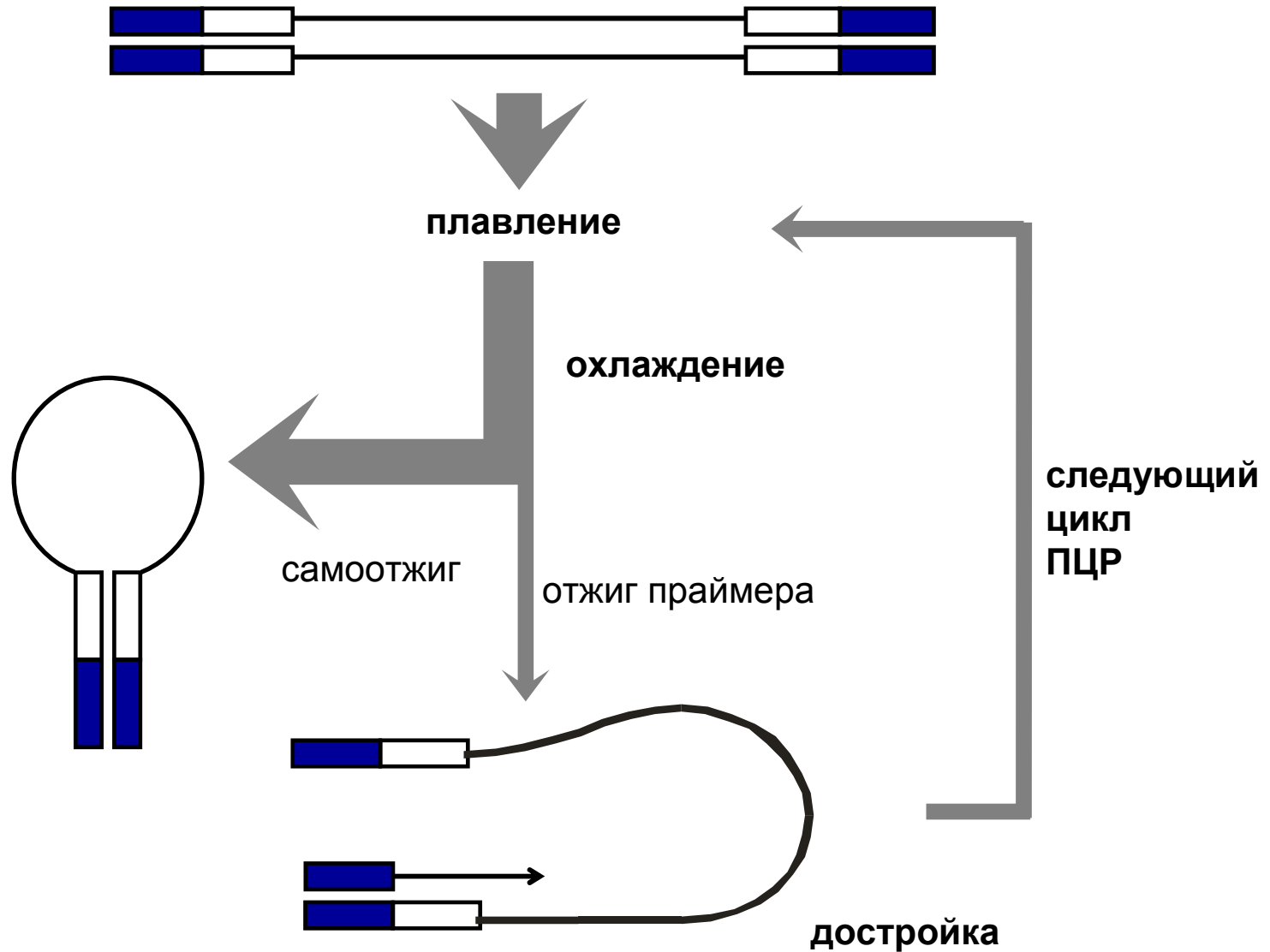


Схема селективной супрессии ПЦР

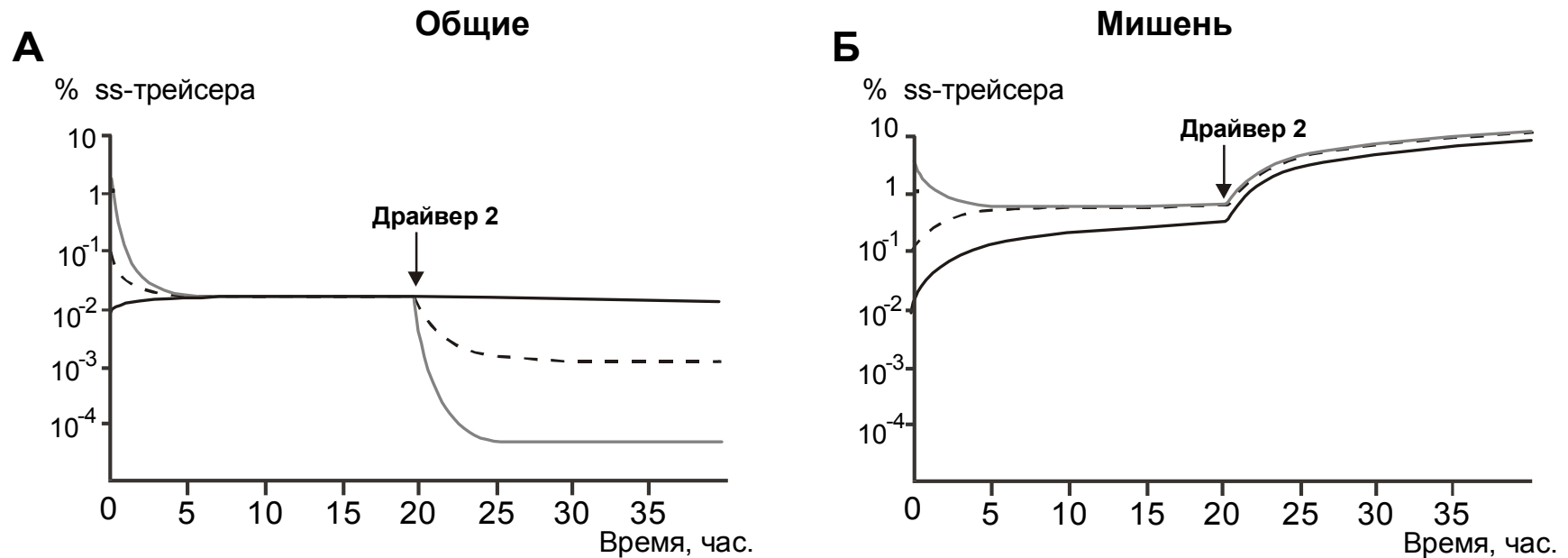




Расчет обогащения для вычитающей гибридации SSH

- Условный расчет исходя из предположения о прохождении реакции гибридации на 99% для всех молекул (в реальности всегда ниже!)
- Стадия I – в каждом из образцов А и Б в ss-фракции
 $E1 = \text{Сдрайвер} / \text{Страйсер} = 50 / 1 = \mathbf{50}$
- Стадия II – обогащение в ds-фракции с разными адаптерами на концах
 $E2 = E1A \times E1B = 50 \times 50 = \mathbf{2\ 500}$

Изменения концентрации различных типов молекул во фракции одноцепочечной ДНК в ходе SSH-гибридизации

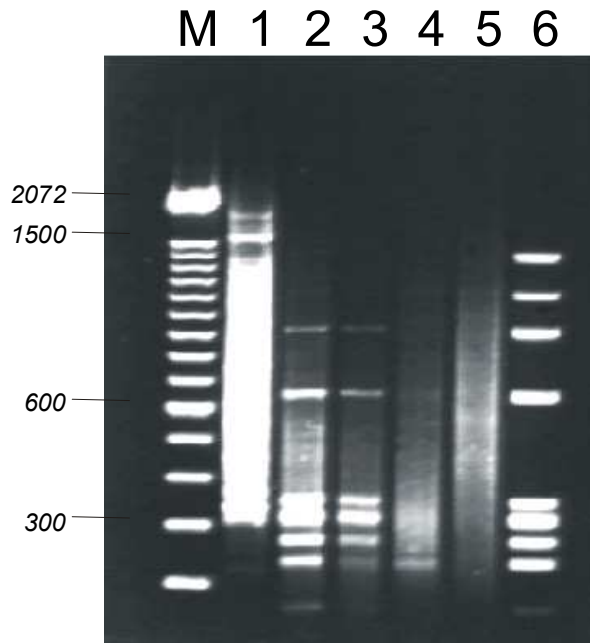


Концентрации оценивали с помощью программы Subtract. Жирная кривая – редкие транскрипты, пунктирная – средние, серая – высокопредставленные. Процесс начинается с первого добавления драйвера.

$$\frac{-dC}{dt} = kC^2$$

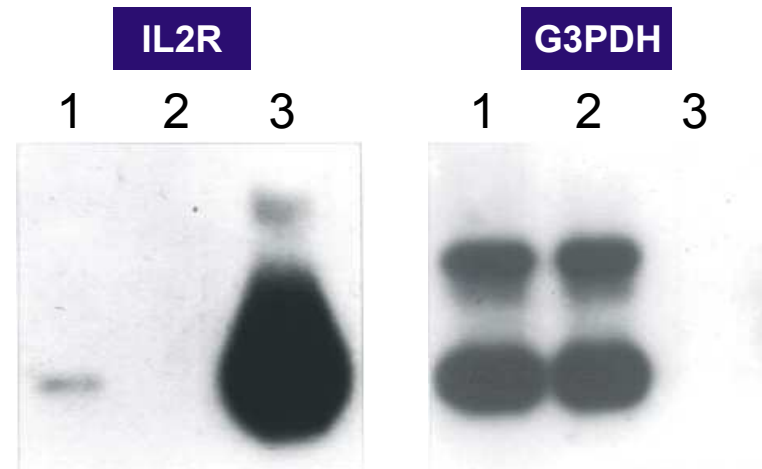
Вычитающая гибридизация SSH в модельных системах

Результат вычитающей гибридизации модельной мишени - ДНК phiX174/HaeIII, добавленной к кДНК в различных количествах.



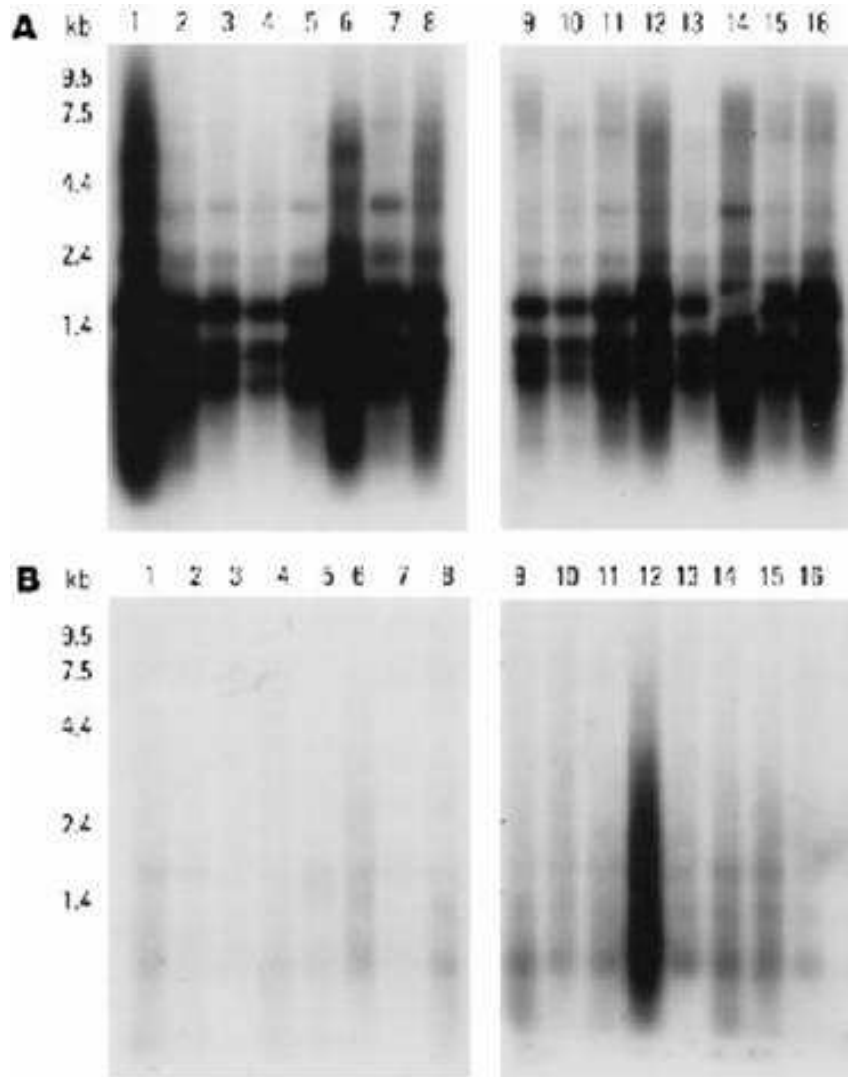
Гель-электрофорез продуктов ПЦР, проведенной с содержащей мишень кДНК до вычитания (1); после вычитания с кДНК, содержащей мишень в количествах 0.1%, 0.01%, 0.001% и 0.0001% (2, 3, 4 и 5, соответственно). На дорожке 6 - продукт ПЦР с phiX174/HaeIII; М - маркер длин 100bp ladder.

Саузерн-блот гибридизация амплифицированной кДНК трейсера (активированные Jurkat Т-клетки), кДНК драйвера (неактивированные Jurkat Т-клетки) и образца кДНК после вычитающей гибридизации SSH



Саузерн-блот гибридизация амплифицированной кДНК трейсера (активированные Jurkat Т-клетки), кДНК драйвера (неактивированные Jurkat Т-клетки) и образца кДНК после SSH с IL2R - кДНК гена рецептора интерлейкина 2 (активируется при активации иммунного ответа в культуре иммунокомпетентных клеток) и G3PDH - кДНК гена глицерид-3-фосфат-дегидрогеназы (общая последовательность). 1 - кДНК трейсера; 2 - кДНК драйвера; 3 - вычитанный образец кДНК.

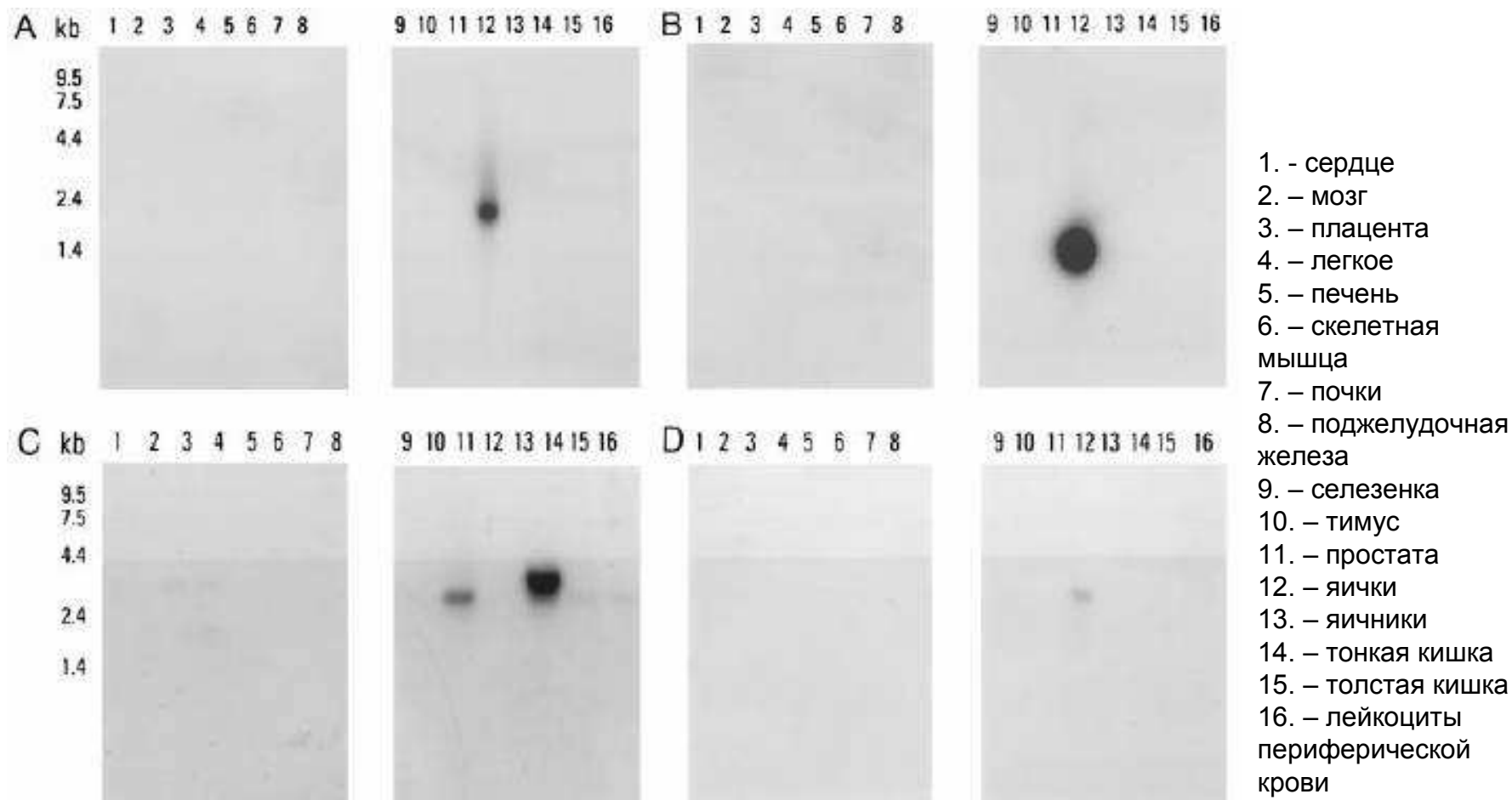
Вычитающая гибридизация: кДНК из яичек – 10 других тканей человека



Результат Нозерн-блот гибридизации невычитанной (А) и вычитанной (Б) кДНК на панели poly(A)+ РНК тканей человека (по 2 мкг на дорожку):

1. - сердце
2. - мозг
3. - плацента
4. - легкое
5. - печень
6. - скелетная мышца
7. - почки
8. - поджелудочная железа
9. - селезенка
10. - тимус
11. - простата
12. - яички
13. - яичники
14. - тонкая кишка
15. - толстая кишка
16. - лейкоциты периферической крови

Нозерн-блот гибридизация дифференциальных клонов на панели poly(A)+ РНК тканей человека



Зеркально-ориентированная селекция

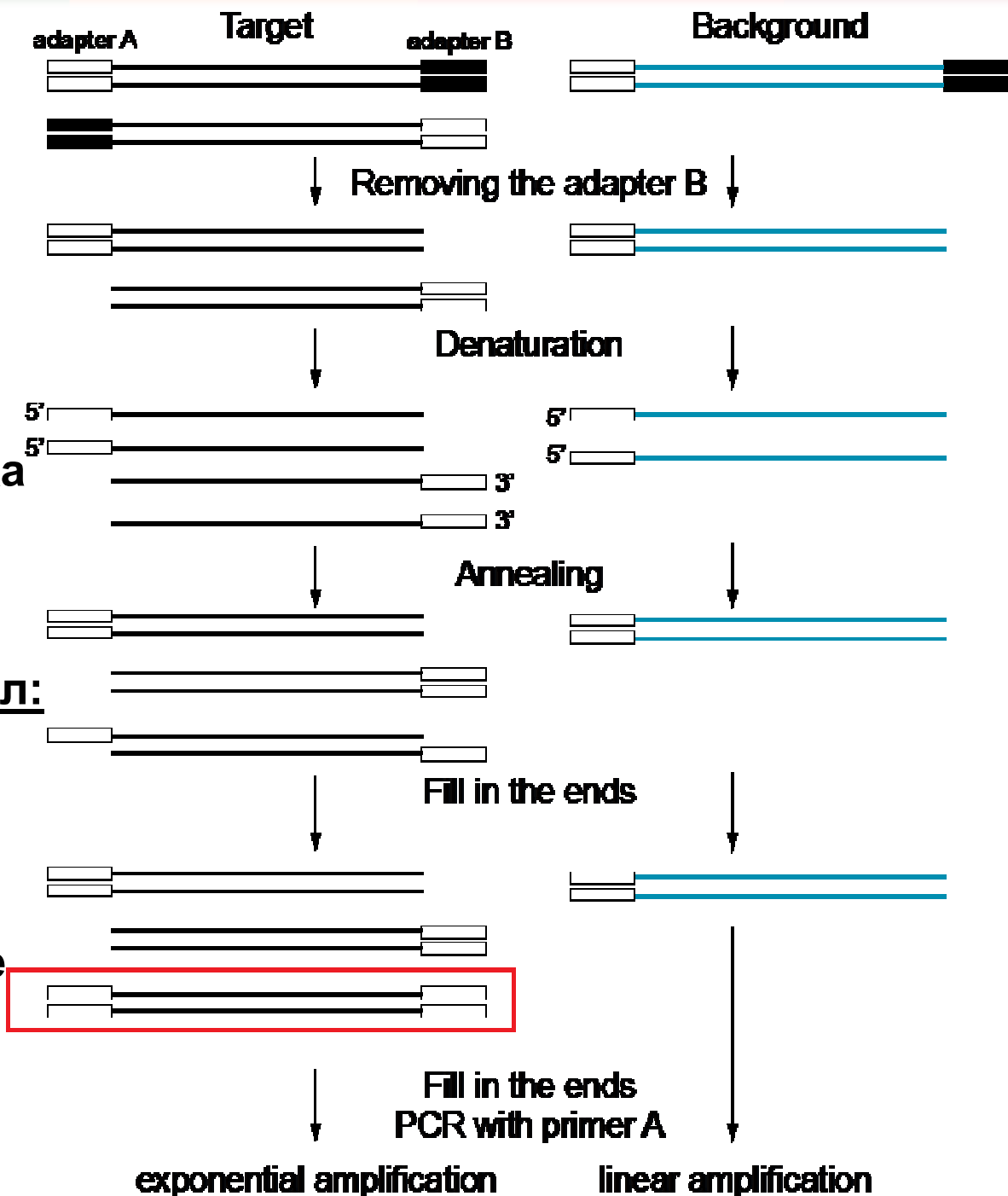
(MOS: Mirror Orientation Selection) –

дополнительная очистка продуктов SSH от фоновых молекул

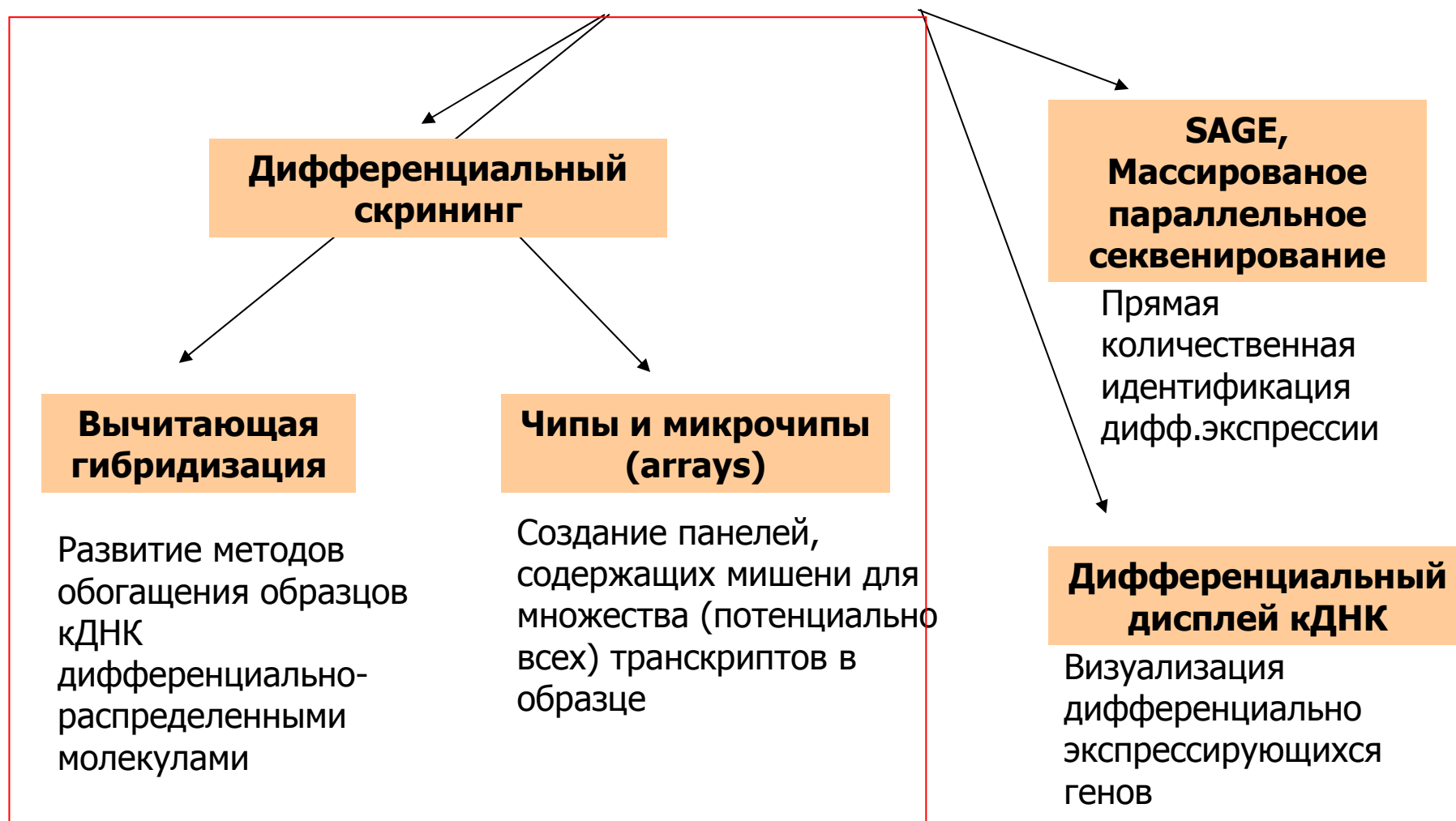
2 типа фоновых молекул:

1. in vitro cloning

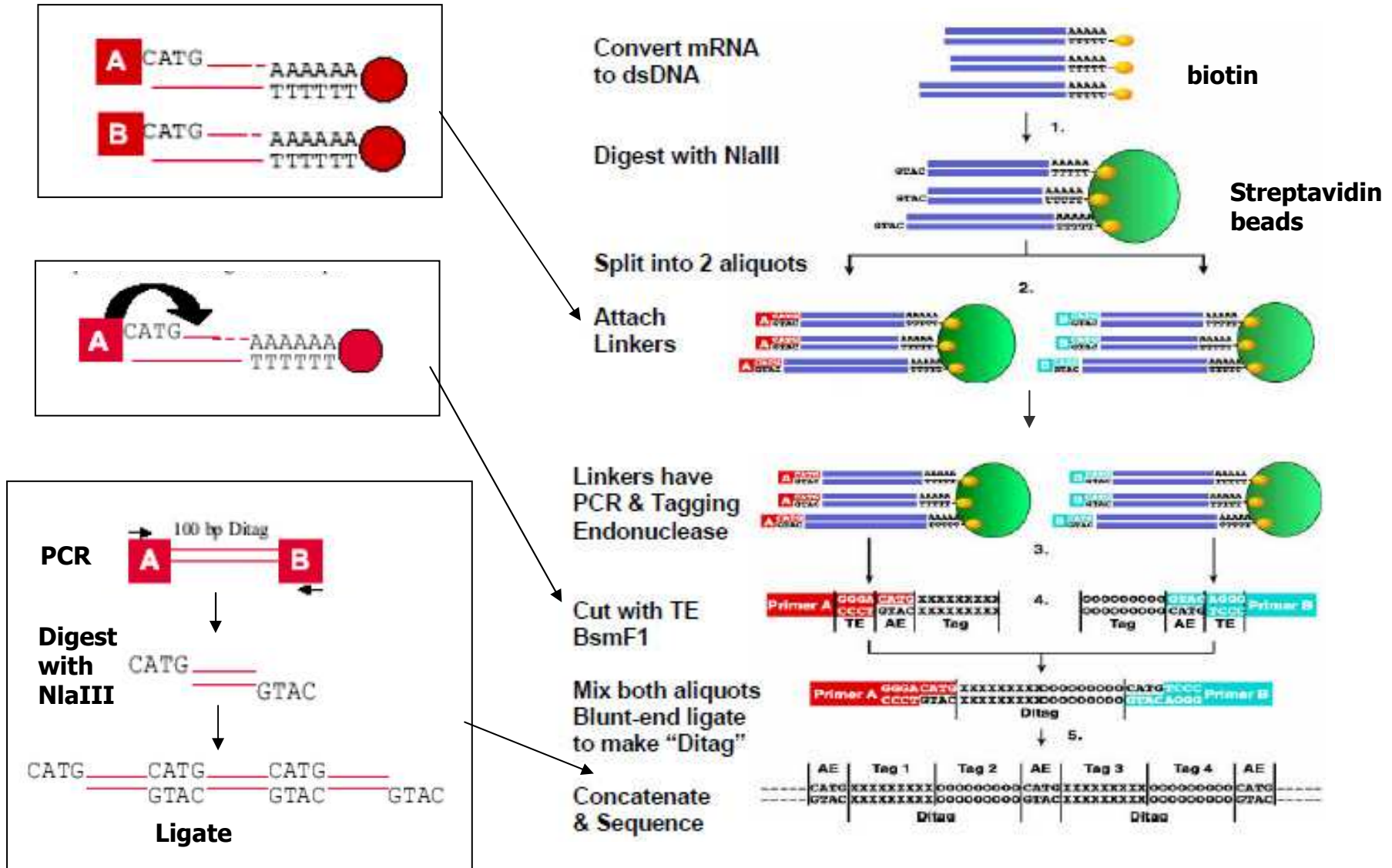
2. отжиг праймера на похожий сиквенс в ходе SSH



Поиск и исследование дифференциально экспрессирующихся генов на уровне мРНК

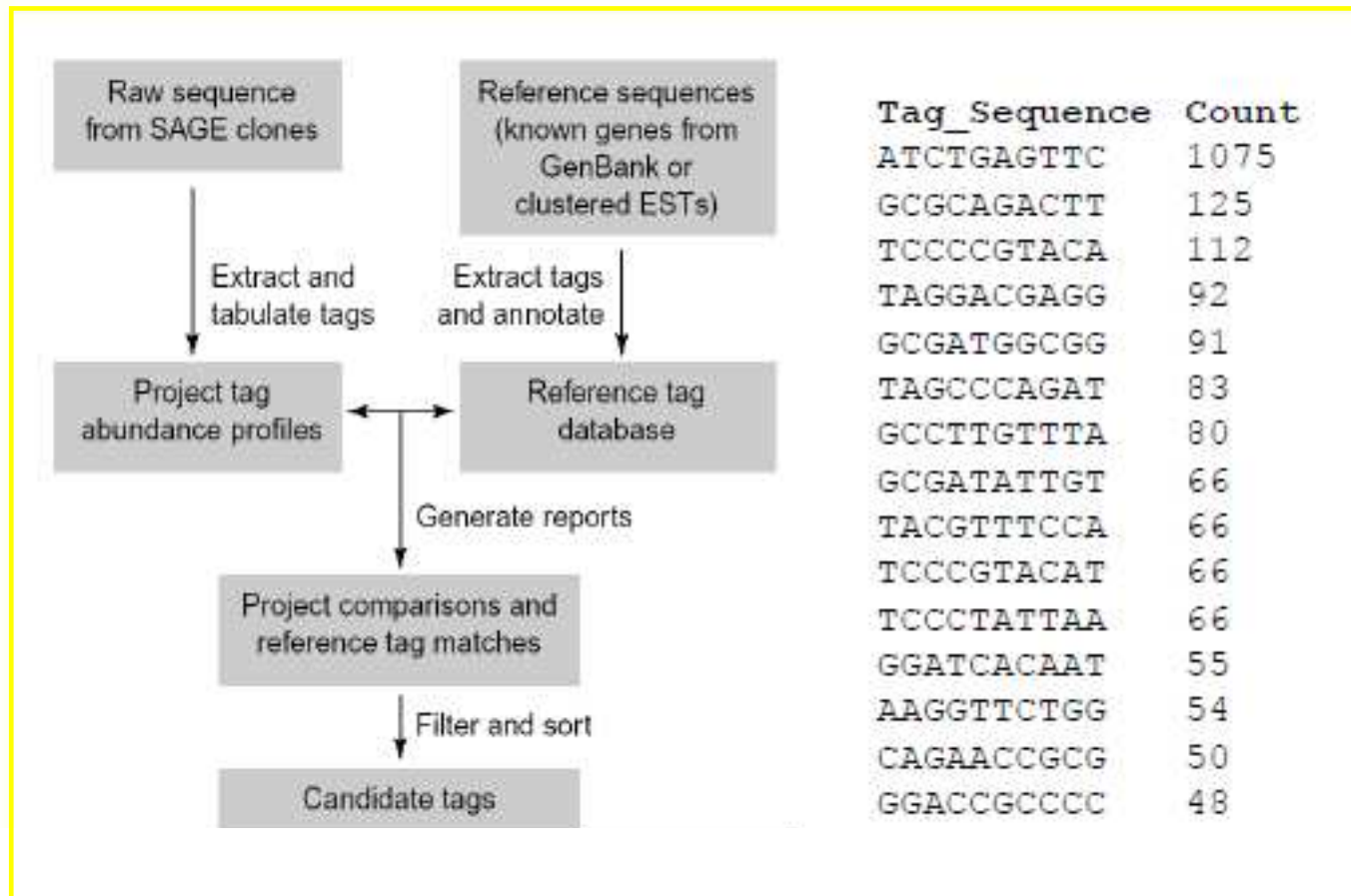


Serial Analysis of Gene Expression (SAGE)



Serial Analysis of Gene Expression (SAGE)

A computer program generates a list of tags and tells how many times each one has been found in the cell:

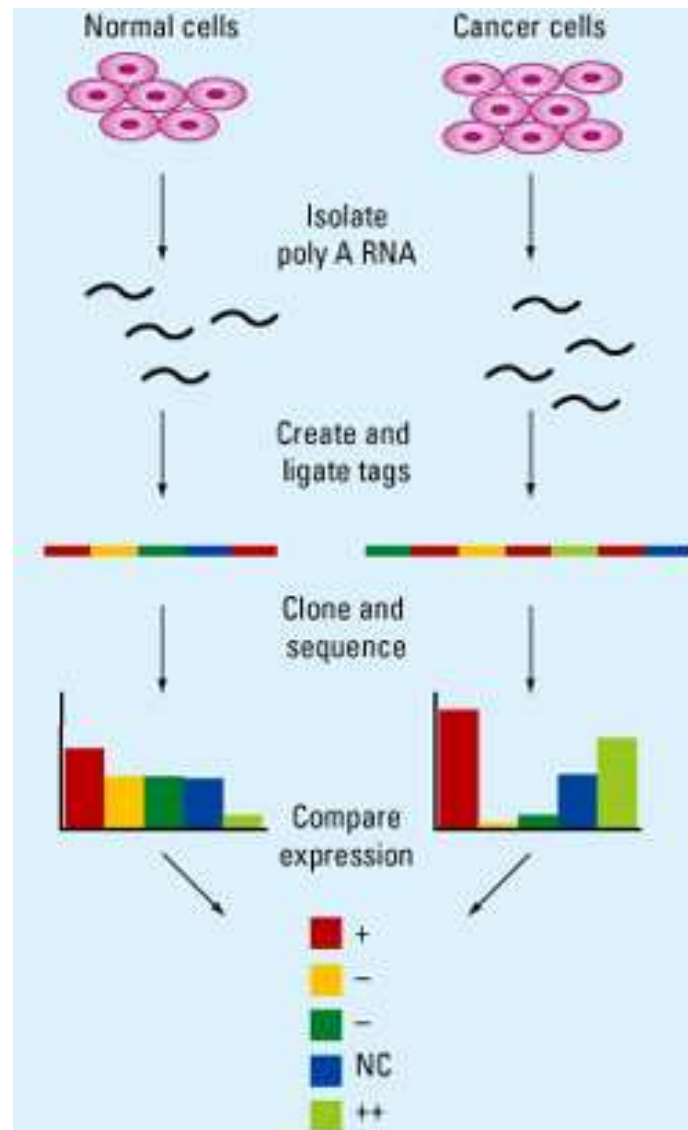


Serial Analysis of Gene Expression (SAGE)

To identify the RNA and the gene that produced each of the tags. This is done by comparing the tags to a database containing all known genes from the organism.

Tag_Sequence	Count	Gene Name
ATATTGTCAA	5	translation elongation factor 1 gamma
AAATCGGAAT	2	T-complex protein 1, z-subunit
ACCGCCTTCG	1	no match
GCCTTGTTTA	81	rpa1 mRNA fragment for r ribosomal protein
GTTAACCATC	45	ubiquitin 52-AA extension protein
CCGCCGTGGG	9	SF1 protein (SF1 gene)
TTTTTGTTAA	99	NADH dehydrogenase 3 (ND3) gene
GCAAAACCGG	63	rpL21
GGAGCCCGCC	45	ribosomal protein L18a
GCCCGCAACA	34	ribosomal protein S31
GCCGAAGTTG	50	ribosomal protein S5 homolog (M(1)15D)
TAACGACCGC	4	BcDNA.GM12270

SAGE позволяет сравнивать популяции мРНК из разных источников





Serial Analysis of Gene Expression (SAGE)

Distribution of tags

- **303,706 total tags**
- **48,471 distinct tags**
- **Distribution**
 - **85.9% seen up to 5 times (25% of mass)**
 - **12.7% between 5 and 50 times (30%)**
 - **0.1% between 50 and 500 times (26%)**
 - **0.1% more than 500 times (19%)**

From tags to genes

- **Ideal situation:**
 - **one gene = one tag**
- **True situation**
 - **one gene = many tags (alternative splicing; alternative polyadenylation)**
 - **one tag = many genes (conserved 3' regions)**