Геномика и протеомика

Лекция 5. Транскриптомика

(бакалавры)

Составитель: проф. М.Р. Шарипова

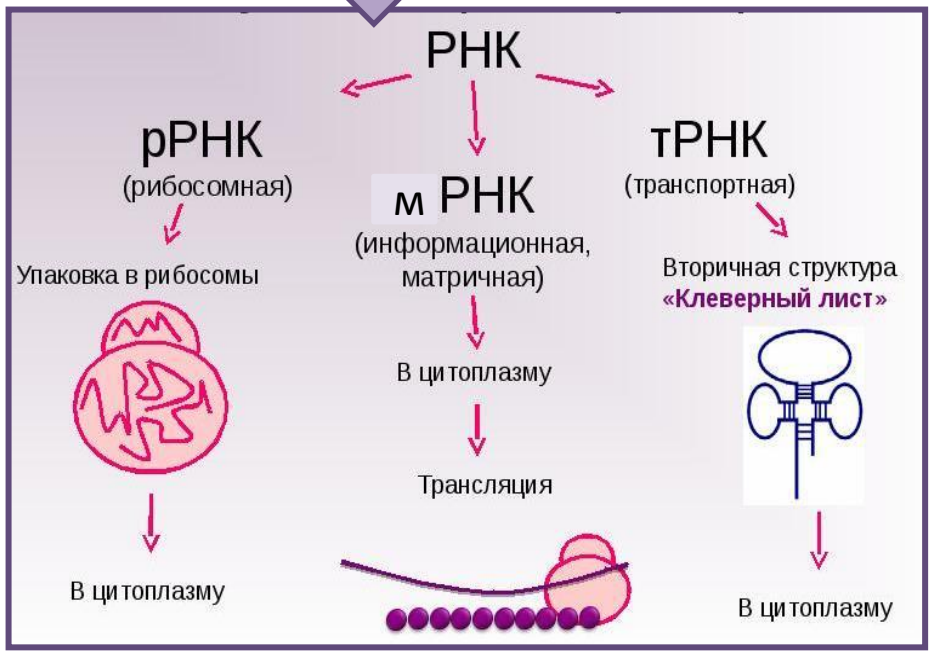
- **Транскриптомика** - раздел функциональной геномики, изучающий **транскриптомы** - набор РНК, присутствующих в организме
- **Транскриптомика** – набор инструментов и подходов (омиксные технологии) для глобального анализа формирования транскриптов
- **Цель транскриптомики** – исследование полноразмерной структуры и механизмов формирования транскриптов

Объект изучения транскриптомики – это совокупность всех транскриптов, которые образуются в клетке

РНК

Кодирующая РНК

Некодирующая РНК



Большая часть генома – dark genome - транскрибируется в некодирующую РНК

- Протяженность транскриптома не коррелирует с размером генома

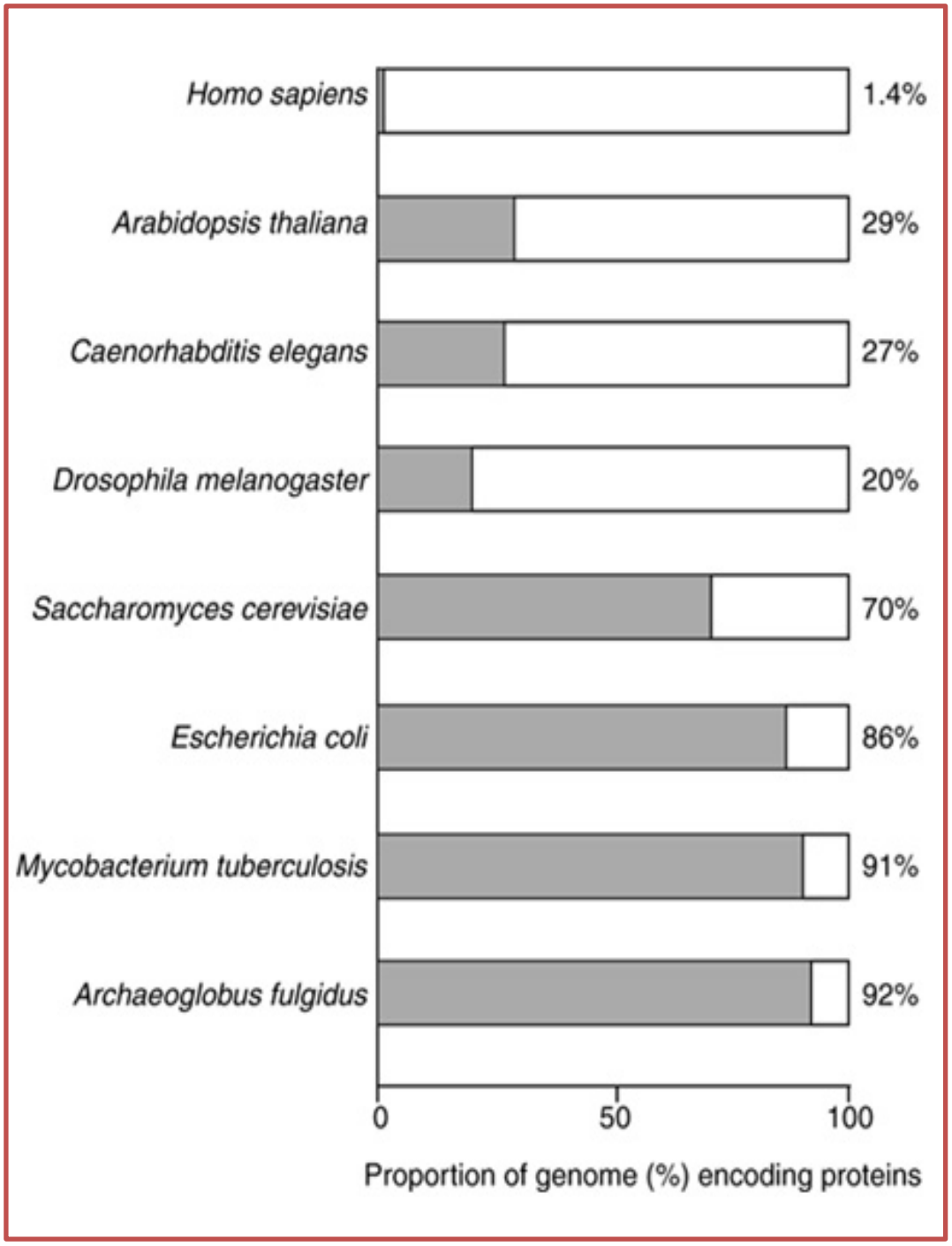
- Белок-кодирующие гены описывают как «вложенный» ген/рамка, «перекрывающийся» ген/рамка или более понятным термином «ген-матрёшка»

Задачи транскриптомики –

- **Выявление всех транскрипционных единиц – транскриптонов и нкРНК**
- **Анализ транскриптонов: сайтов инициации транскрипции, 5’ и 3’- концов генов**
- **Анализ паттернов (образцов) сплайсинга и других посттранскрипционных модификаций**
- **Выявление информации в виде сигналов и кодов, необходимой для формирования протеома**

Определение транскриптома

- **Транскриптом** – это совокупность всех транскриптов, набор всех РНК, которые экспрессируются в клетках организма
- **Транскриптом** – это первый уровень реализации генетической информации, заключенной в геноме, т.е. первый уровень фенотипа
- **Транскриптон** - участок ДНК, ограниченный промотором и терминатором, представляет собой единицу транскрипции в структуре транскриптома

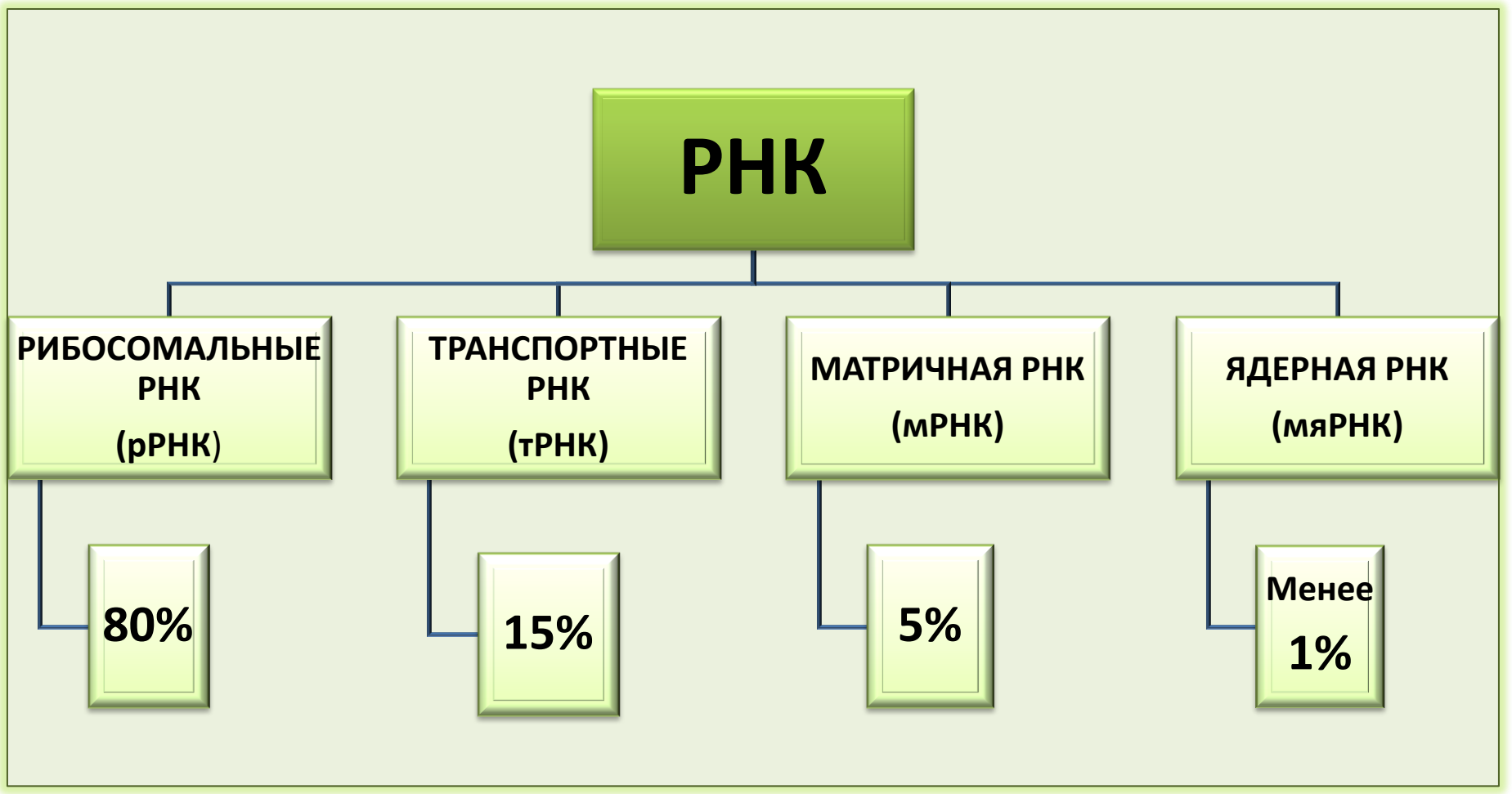


Доля (%) белок-кодирующих последовательностей в эукариотических и бактериальных геномах

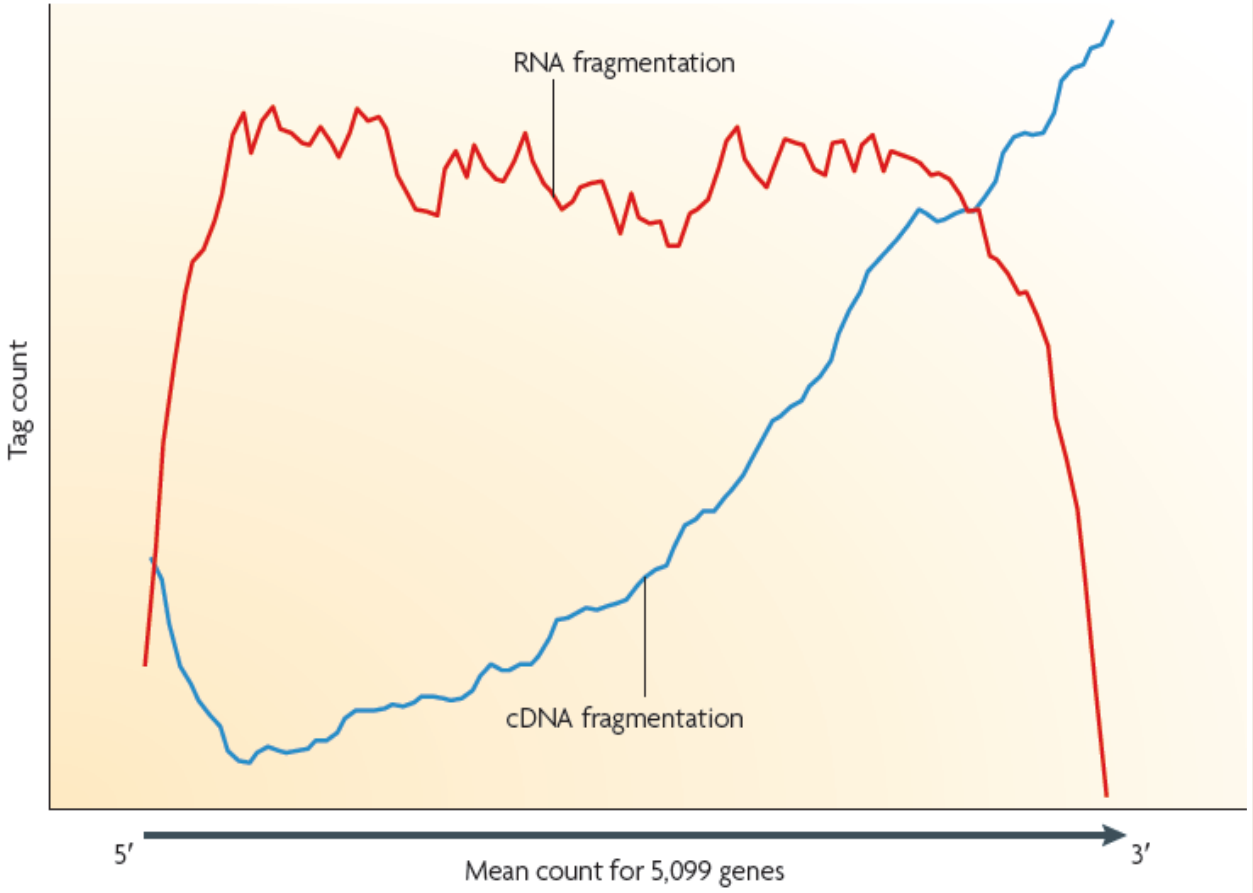
- С повышением сложности организма вклад белок-кодирующего генома снижается
- **Большая часть транскриптома (90%) млекопитающих состоит из некодирующей РНК**

- белок-кодирующие последовательности геномов

Соотношение стандартных типов РНК в первичном транскрипте



De novo секвенирование транскриптома

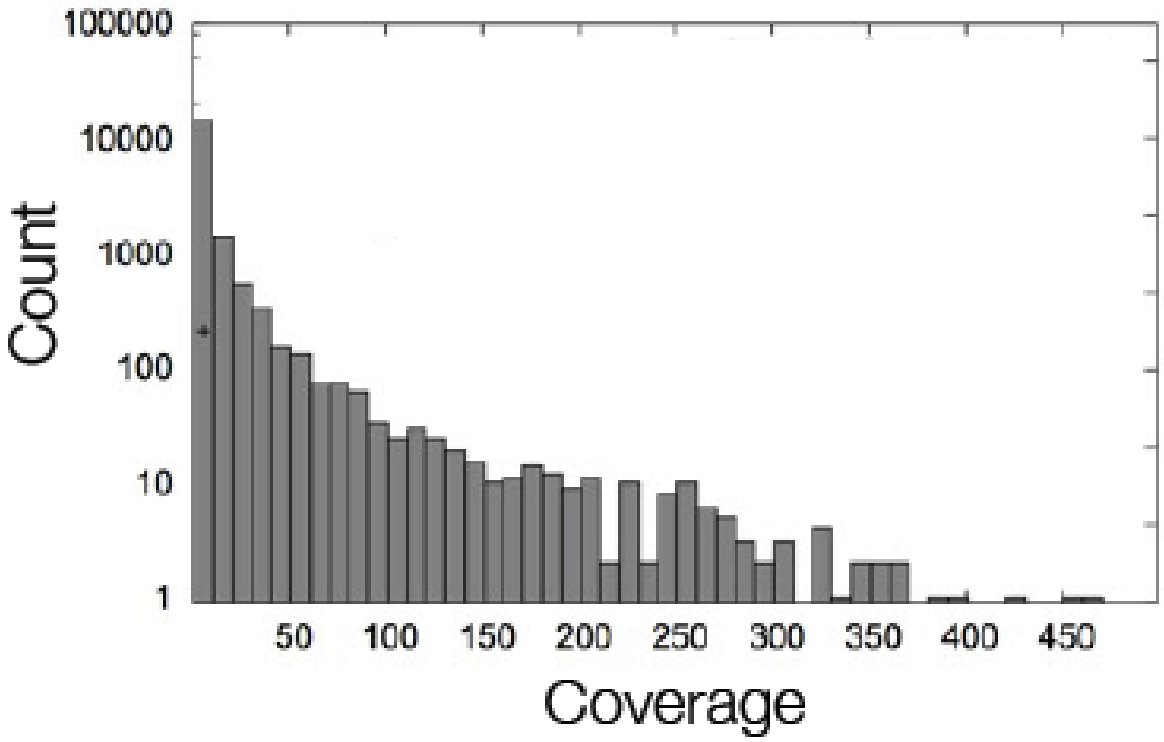


Транскрибируется значительно большая часть генома, чем предполагали ранее:

1) **выявлены новые транскрипты с неизвестной функцией**
2) **различие транскриптов в разных тканях**
3) **Резкая перестройка после стрессов**

Секвенирование транскриптомов показало неравномерность покрытия транскриптов:

- Транскриптомы тканей в разные промежутки времени могут очень сильно отличаться
- 20% генов дают 80% ридов (reads)



ТРАНСКРИПТОМЫ ДИНАМИЧНЫ И ИМЕЮТ СЛОЖНУЮ СТРУКТУРУ

Динамичность транскриптома

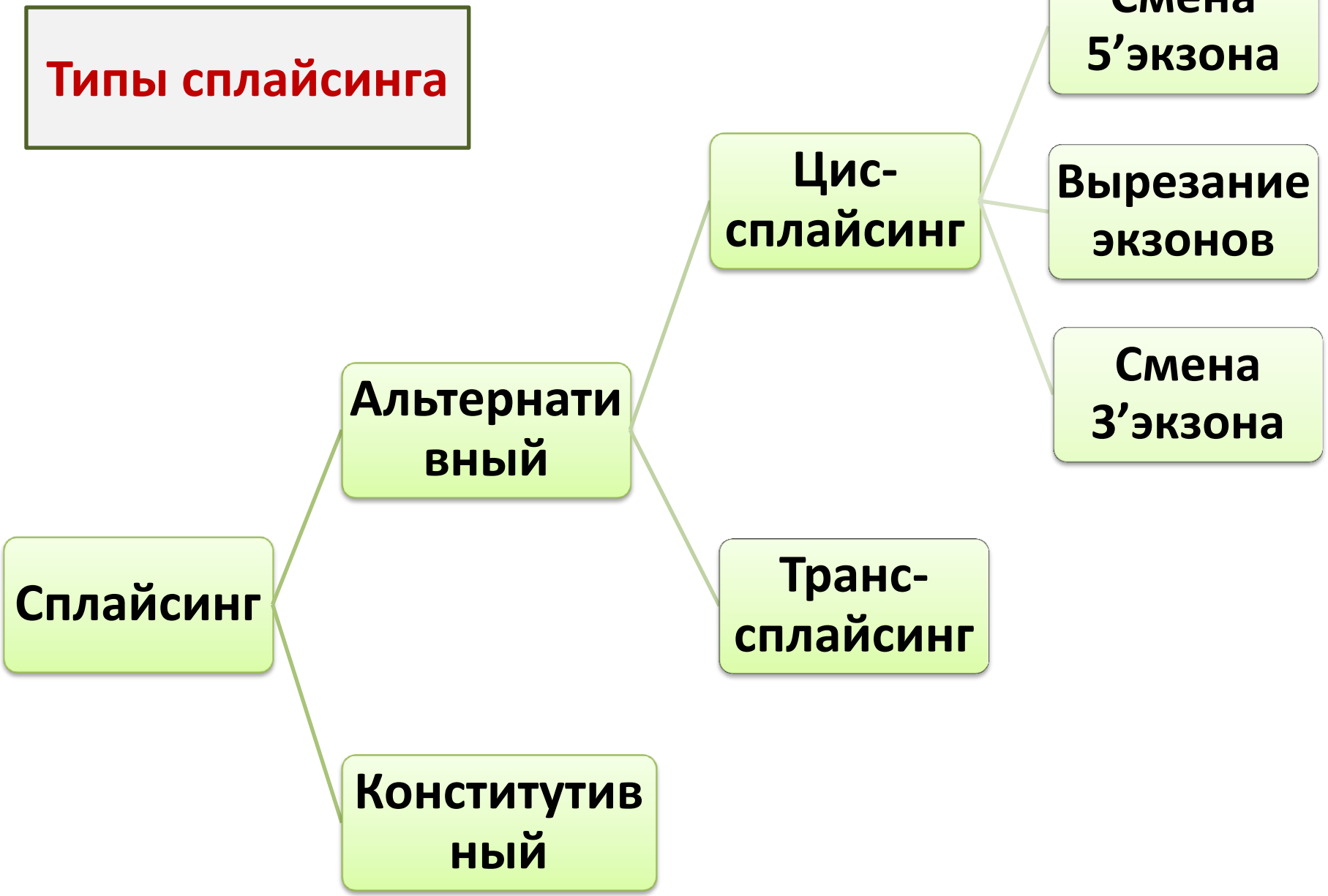
- **Динамичность транскриптома** зависит:
  1) - от стадии клеточного цикла
  2) - от типа клеток и тканей
  3) - от стадий развития всего организма
  4) - от наличия внешних сигналов важных для транскрипции генов

Таким образом, транскриптом характеризуется пространственной и временной дифференциацией

Основа динамичности транскриптома

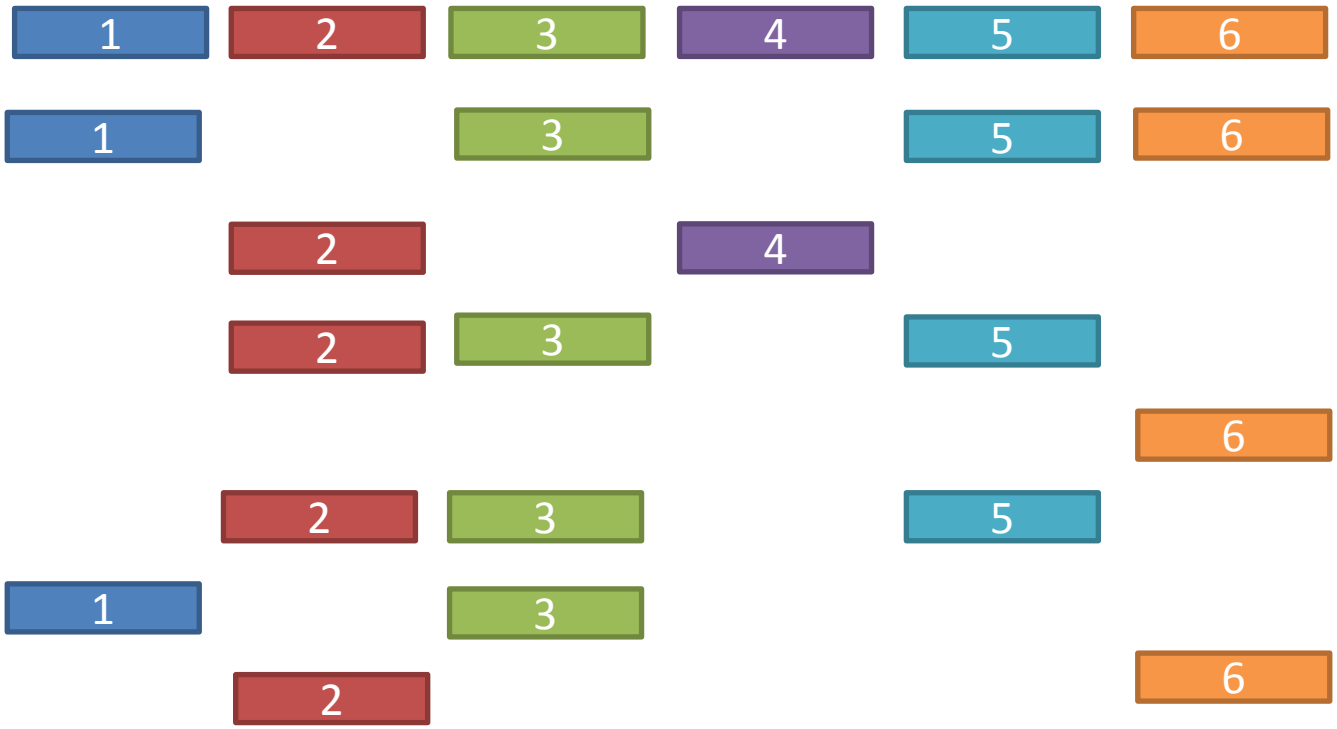
- **(1) регуляторные сайты транскрипции:** промоторы, энхансеры, сайленсеры, терминаторы транскрипции и др.
- **(2) сплайсинг**: регуляторы сплайсинга, альтернативный сплайсинг, транс-сплайсинг
- **(3) взаимодействие с РНК – регуляторами**

Типы сплайсинга



Частота использования экзонов

- **Выбор экзона для реакции сплайсинга и получения зрелой мРНК является контролируемым и зависит от набора регуляторов, которые участвуют в этом процессе**
- **э к з о н ы**



- **Альтернативный сплайсинг лежит в основе многообразия транскриптов**

Структура транскриптов

мРНК

рРНК

тРНК

мяРНК

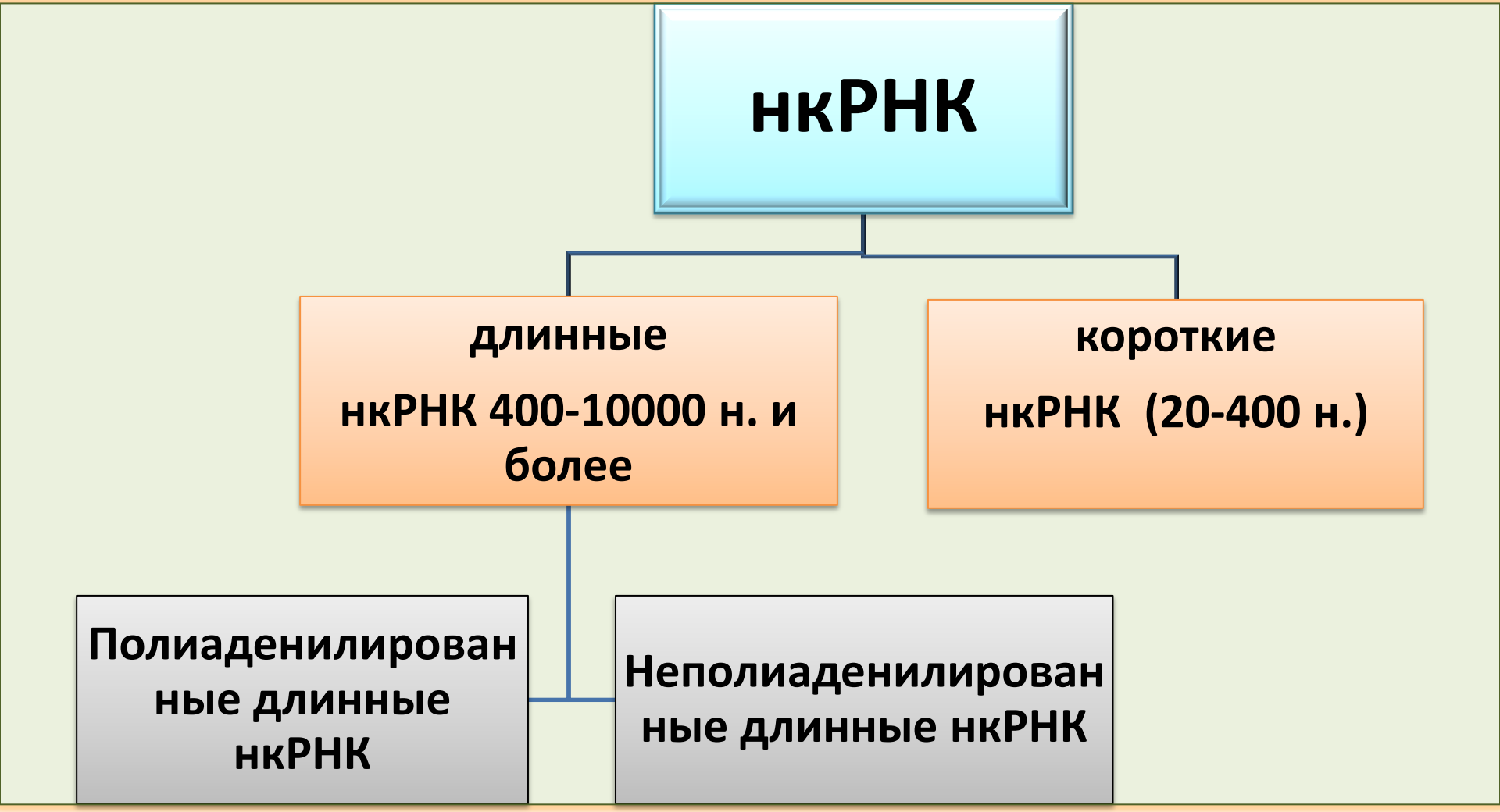
нкРНК

Некодирующие РНК – это РНК-регуляторы экспрессии генов

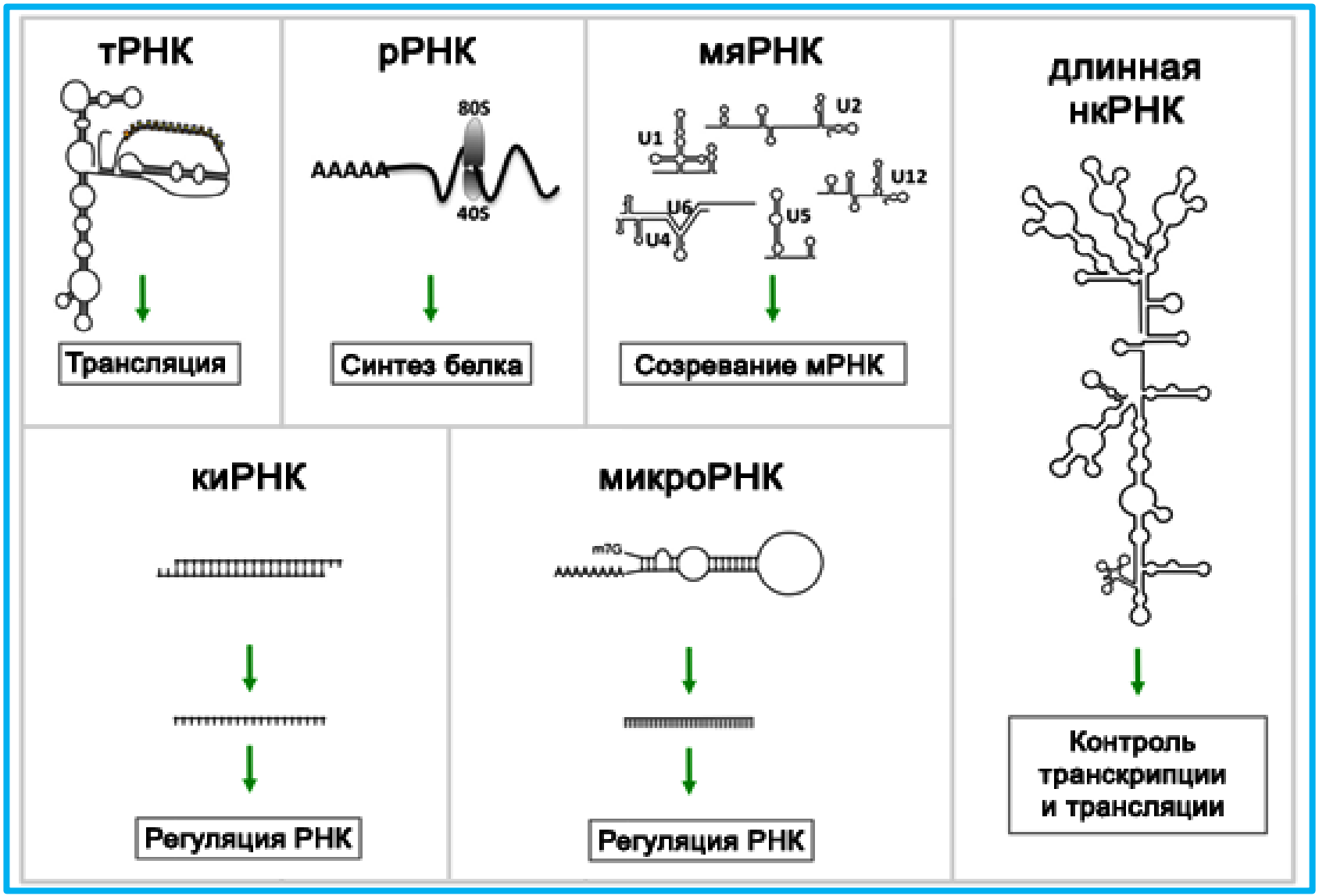
Геноинформационный анализ:

- Секвенирование транскриптома позволяет получать полный объем РНК-транскриптов клеточной популяции
- Основную часть структуры транскриптома составляют нкРНК, что указывает на новую систему регуляции генов
- Создан Web-сайт NONCODE для сбора данных о функциональных ncРНК по всем организмам
  http://noncode.bioinfo.org.cn/

Состав некодирующей РНК



Структура транскриптома



киРНК – это siРНК – короткие интерферирующие РНК

Длинные нкРНК

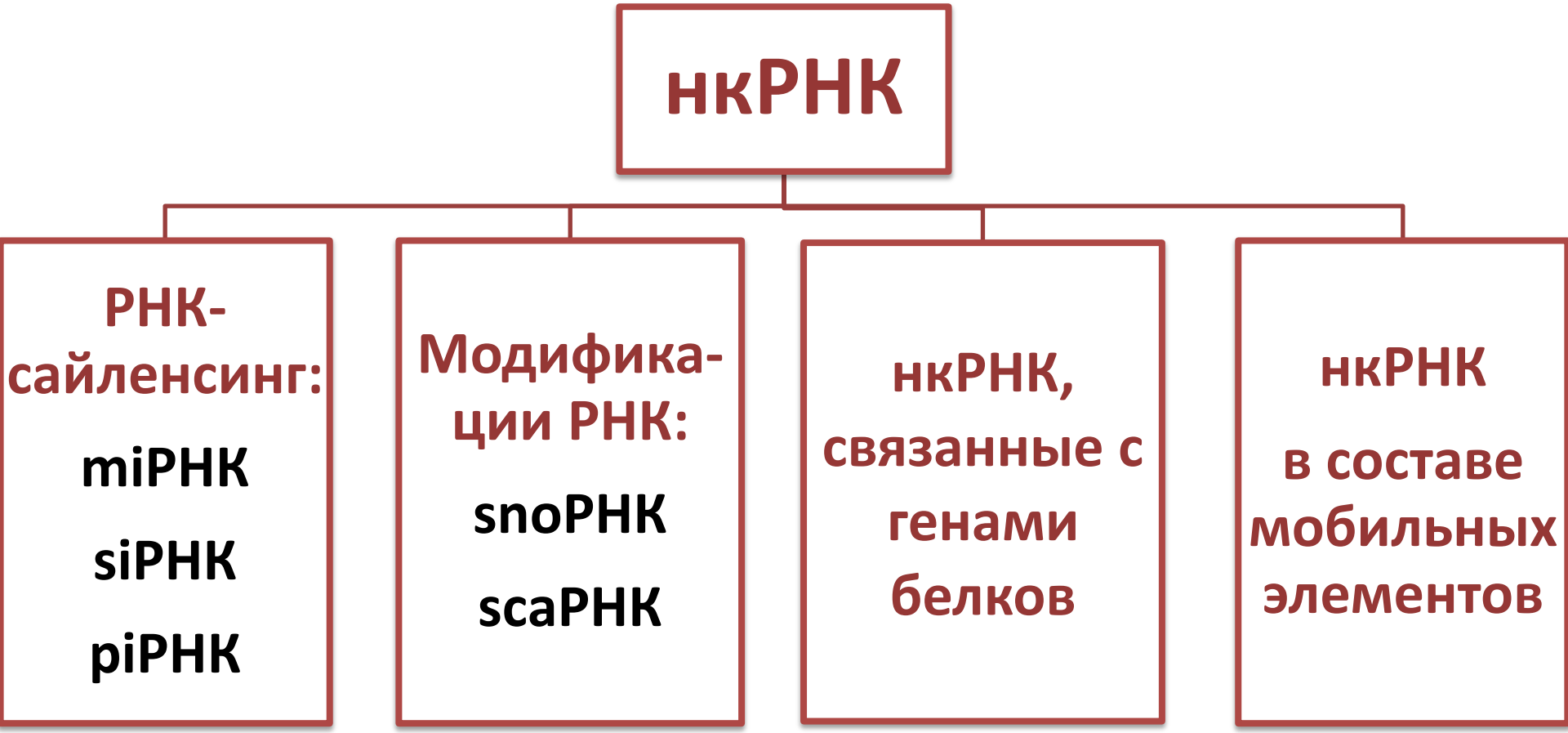
В геноме человека и других модельных организмов обнаружены десятки тысяч длинных нкРНК

- Длинные нкРНК часто расположены в тех же локусах, что и белки, но на противоположных цепях, регулируют экспрессию своего партнера
- Большинство нкРНК обладает тканеспецифичной экспрессией и транскрибируется на определенной стадии развития
- Многие длинные РНК служат предшественниками коротких РНК
- Некоторые длинные нкРНК являются ко-активаторами транскрипции генов

Регуляция импринтинга

- Показано, что нкРНК ассоциированы с кластерами импринтированных генов (экспрессия блокирована)
- Обнаружены нкРНК млекопитающих, которые регулируют активность хроматина и длинные нкРНК, которые участвуют в инактивации (в импринтинге) Х-хромосомы
- Мутации указывают на регуляторную роль нкРНК в сайленсинге импринтированных кластеров генов и отдельных генов

Короткие нкРНК



miРНК

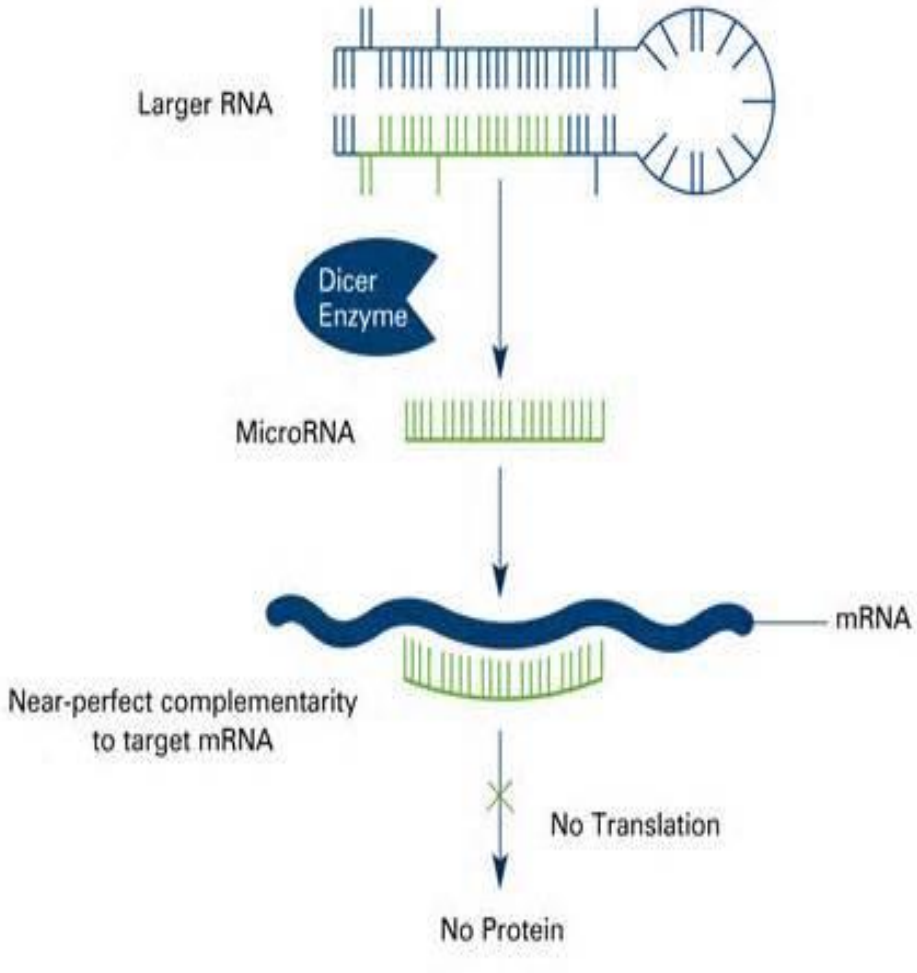
- 1) Большинство генов микроРНК являются межгенными, транскрибируются как независимые единицы либо вместе с геном-мишенью, что обеспечивает их взаимную регуляцию
- Около половины генов miРНК (22 н.) расположены в интронах генов белков и генах длинных нкРНК
- 2) Некоторые гены miРНК расположены в 3’НТО генов, регулируют клеточный уровень многих неродственных мРНК
- 2) В геноме человека выявлено 300 miРНК, предсказано более 1000
- 4) До 30% генов белков регулируются miРНК

Схема расположения функциональных участков в мРНК

Связывание miРНК с РНК-транскриптами осуществляется в кодирующей и некодирующей области

miРНК

- **miРНК** регулируют экспрессию на уровне транскрипции и посттранскрипционном уровне
- На посттранскрипционном уровне **miРНК** взаимодействуют с мРНК и подавляют трансляцию или приводят к деградации мРНК
- Между **miРНК** и её мРНК-мишенью нет однозначного соответствия: **miРНК** может иметь несколько мРНК-мишеней, и мРНК может иметь несколько соответствующих ей **miРНК**
- Мутации в **miРНК** ослабляют или усиливают экспрессию мРНК-мишеней



siРНК

siРНК — это молекулы небольшой длины (21-25 н.)

siРНК специфично связываются с мРНК по принципу комплиментарности и блокируют трансляцию **(РНК-интерференция).**

siРНК способны подавлять транскрипцию через модификацию гистонов

Функции siРНК :

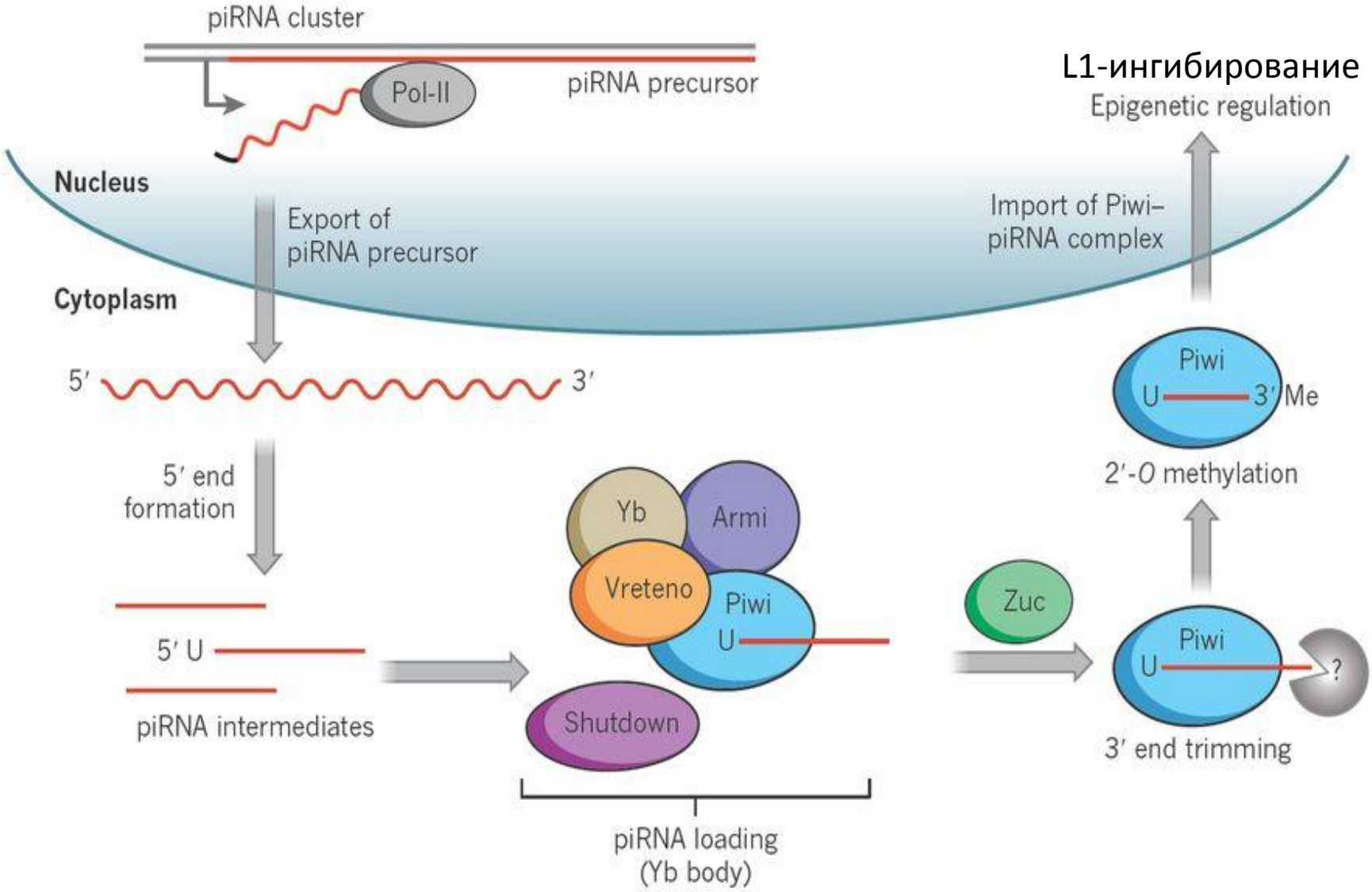
- защита от вирусов
- подавление транспозиции
- репрессия трансгеноза
- формирование гетерохроматина.

Источники происхождения **siРНК**: вирусные РНК, МЭ, шпильки, искусственные конструкции, двунаправленная транскрипция

piRNA

- **piRNA** - многочисленная группа риборегуляторов (26 - 32 нуклеотидов)
- Число **piРНК** составляет около 50 тысяч у млекопитающих и 13 тысяч у дрозофилы
- **piРНК** собраны в геноме в кластеры, в 90% расположены в участках, не содержащих аннотированные гены
- **piRNA** экспрессируются в клетках зародышевой линии
- **piRNA** связывается с PIWI-белком (**p**iwi-**i**nteracting RNAs), комплекс **piRNA-PIWI** участвует в репрессии ретротранспозонов и других генетических элементов
- **piRNA** могут иметь эпигенетические эффекты

L1-ингибирование

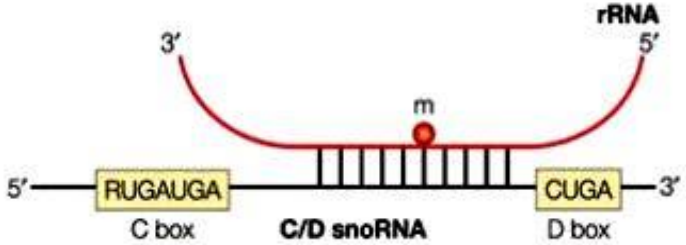


snoРНК и scaРНК

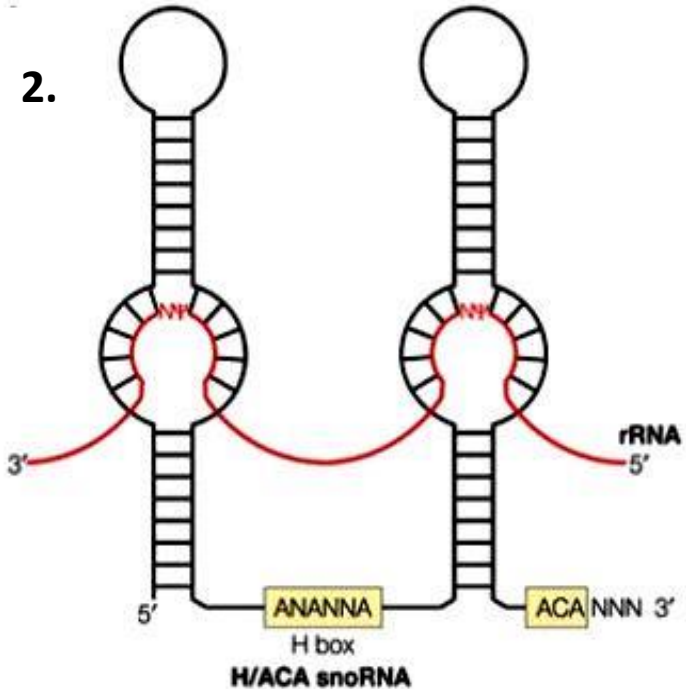
- **snoРНК** - малые ядрышковые РНК, **scaРНК** – малые РНК, локализованы в тельцах Кахаля, являются подгруппой малых ядерных РНК
- Количество **snoРНК** больше 200 для млекопитающих, длина 100-150 н.
- **snoРНК** и **scaРНК** участвуют в химических модификациях (метилировании и псевдоуридилировании) пре-рРНК и мяРНК, соответственно
- гены **snoРНК** локализованы в интронах других генов, которые называют генами-хозяевами, один интрон содержит только один ген **snoРНК**
- **snoРНК** процессируют во время сплайсинга пре-мРНК либо вырезаются из интронов эндонуклеазами. Гены-хозяева кодируют белки или нкРНК

Характеристика snoРНК

1. C/D-семейство (box C/D snoRNAs)



2.



H/ACA-семейство (box H/ACA snoRNAs)

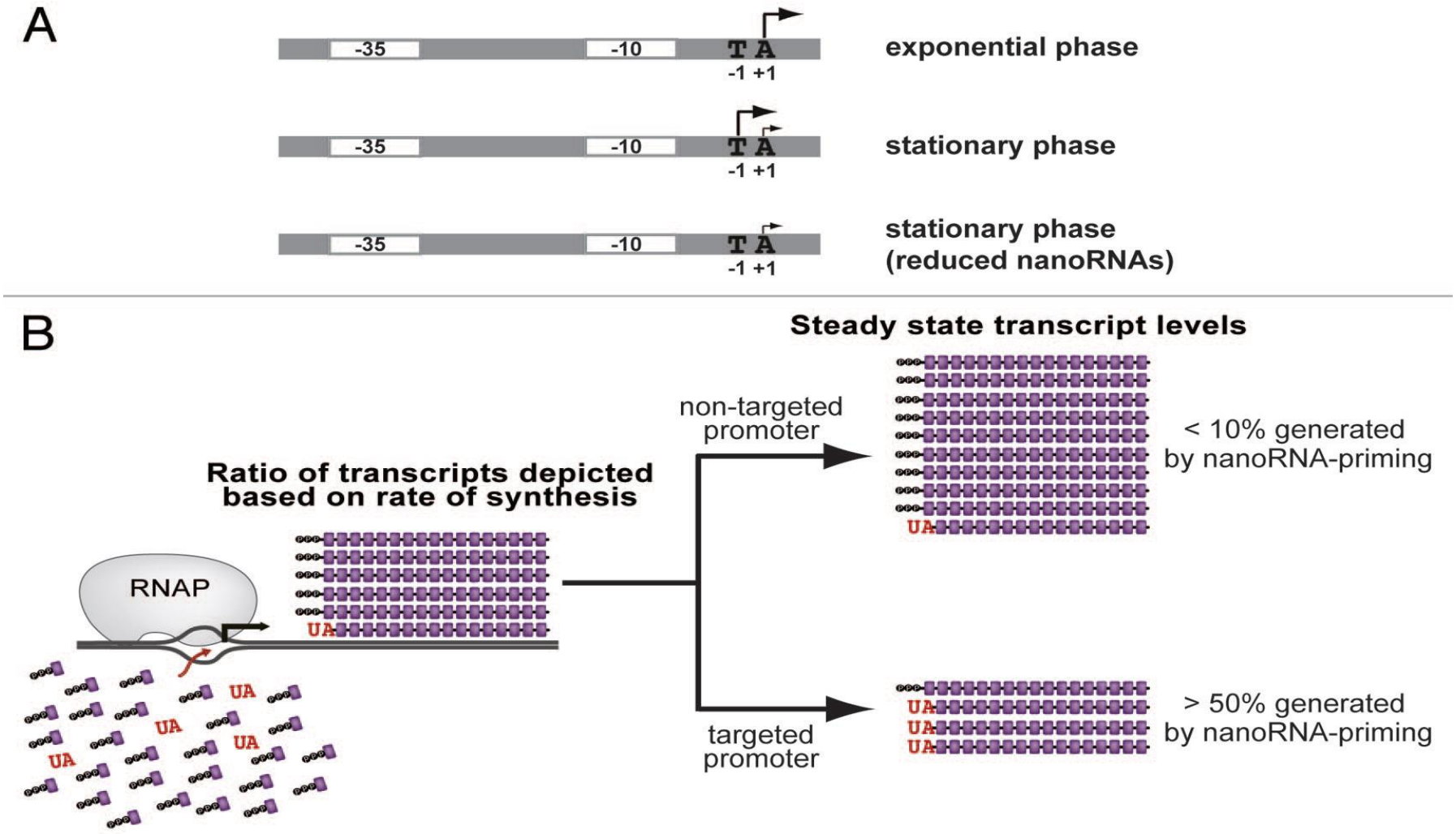
Основные функции snoРНК связаны с определением сайтов модификации рРНК

SnoRNAs

- **Large Family**
- **Intron-encoded**
- **Guide RNA Modification**

метилирование и псевдоуридилирование

Типы РНК-регуляторов: nanoRNA



nanoRNA изменяют сайт начала транскрипции (A). Стабильность модифицированных транскриптов значительно выше, чем стабильность транскриптов, инициированных De Novo (B).

нкРНК

Выявление новых нкРНК изменило представление о роли РНК в клетке: нкРНК выполняют множество функций с использованием неизвестных ранее механизмов:

- участвуют в регуляции транскрипции генов и сплайсинге
- вовлечены в регуляцию трансляции и деградацию мРНК
- вовлечены в процессинг и модификацию рРНК
- предохраняют от вирусных инфекций и транспозиции мобильных элементов
- нкРНК, как и белки, необходимы для обеспечения жизнедеятельности клеток

Описание транскриптома позволило осознать значимость нкРНК, но оценить их вклад в функционирование генома пока трудно

Особенности в сборке транскриптома

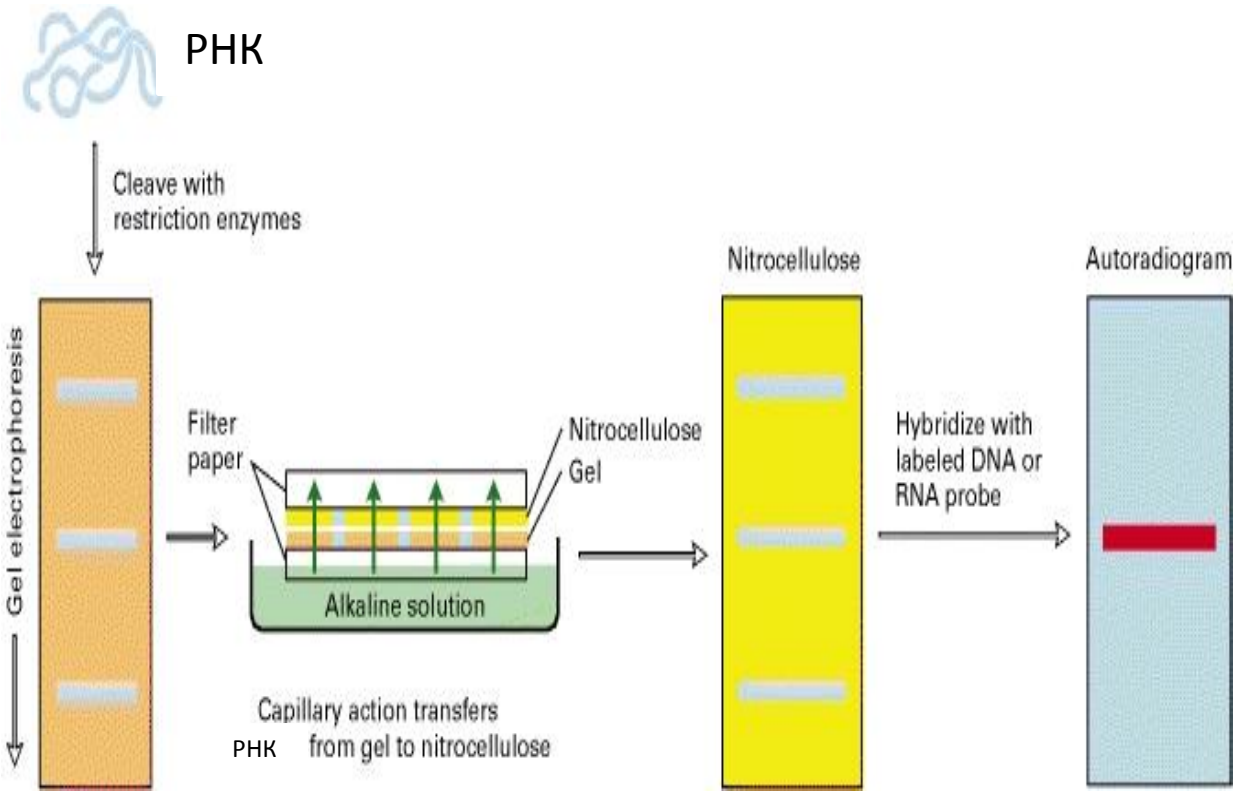
- 1. При сборке генома стояла задача собрать одну длинную молекулу ДНК, при сборке транскриптома стоит задача восстановить множество нуклеотидных последовательностей, имеющих различную длину
- 2. При секвенировании генома покрытие различных участков генома равномерно. При секвенировании транскриптома покрытие двух разных транскриптов отличается, что определяет разный уровень экспрессии генов в заданный момент
- 3. На одной молекуле ДНК формируются **различные** транскрипты как результат альтернативного сплайсинга

Сборка транскриптома

- Сборку транскриптома проводят de-novo или на референсном транскриптоме или геноме
- Более длинные риды (считывания) упрощают сборку и обеспечивают глубину секвенирования транскриптома:
  1) **определение экзонов**
  2) **идентификация паттернов сплайсинга - всех вариантов транскриптов**
  3) **определение стыков сайтов сплайсинга**

Методы исследования транскриптов

Для исследования транскриптов используют метод **прямой детекция молекул РНК -** метод Нозерн-блот-гибридизации:



1. РНК фракционируют на электрофорезе

2. РНК переносят (блоттинг) на мембрану и иммобилизируют на ней

3. Проводят гибридизацию со специфическими мечеными РНК-зондами.

4. Получают полосы на дорожке, соответствующие длине транскрипта гена.

5. Используя различные пробы можно выявить несколько наборов полос для нескольких генов

Нозерн-блот-гибридизация

Нозерн-блот-гибридизация является стандартным методом для выявления мРНК и оценки их относительного содержания в 5-20 образцах

- **Достоинства метода:**

  1) позволяет определять размер транскриптов и выявлять изоформы, образованные в результате альтернативного сплайсинга

  2) дает данные об относительном содержании разных РНК в образце

- **Ограничения:**

  1) не дает данные об абсолютном содержании РНК в образце

  2) не выявляет транскрипты при низком уровне экспрессии генов

  3) трудоемкий и малопроизводительный метод

«Анализ с помощью защиты от рибонуклеазы»
(Ribonuclease protection assay)

Принцип метода:

- с образцами РНК проводят гибридизацию в растворе со специфической антисмысловой меченой РНК-пробой;
- продукты гибридизации обрабатывают РНКазой, которая деградирует все одноцепочечные молекулы РНК, но оставляет в растворе РНК-дуплексы;
- оставшиеся РНК-дуплексы анализируют на гель-электрофорезе

✓ Получают полосы на дорожках, соответствующие длине дуплексов.

✓ Используя различные РНК-пробы можно выявлять продукты гибридизации одновременно для нескольких генов.

✓ Метод позволяет получить данные о содержании разновидностей транскрипта в образце.

«Анализ с помощью защиты от рибонуклеазы»
(Ribonuclease protection assay)

- Метод пригоден для определения 10-15-ти мРНК и оценки их содержания в 5-20-ти образцах одновременно
- **Достоинства:**

позволяет картировать 5’- и 3’- окончания транскриптов и экзон-интронные стыки

- **Ограничения:**

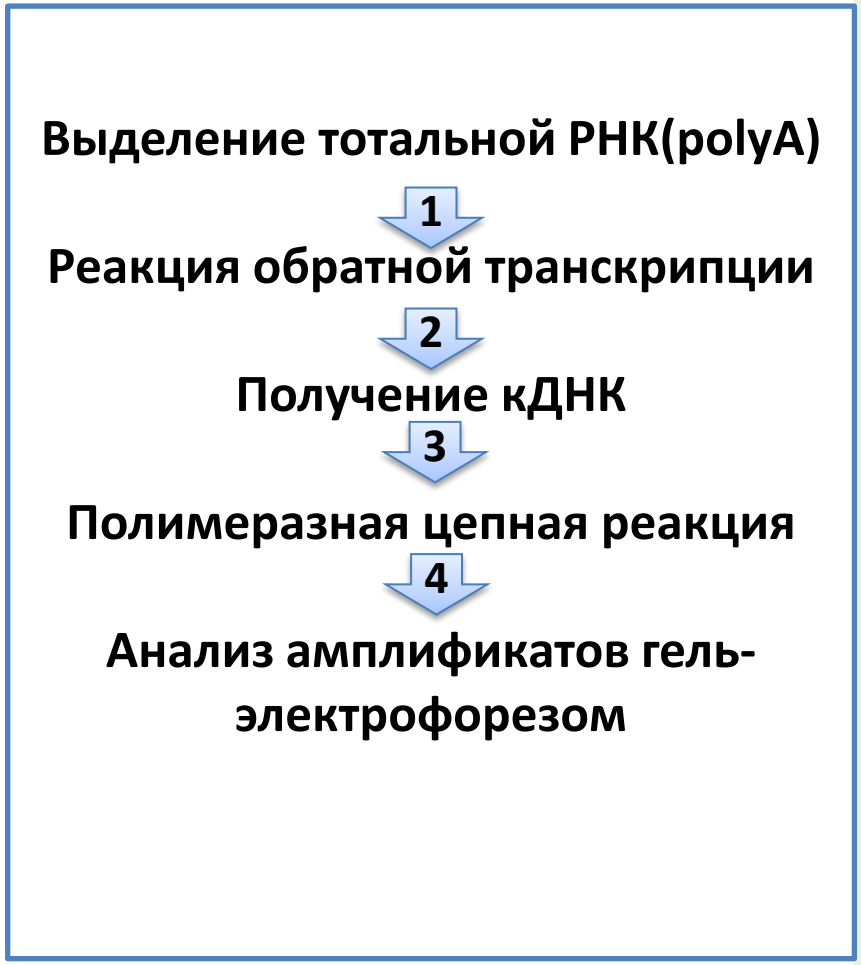
метод трудоемкий и малопроизводительный

Количественная ОТ-ПЦР (quantitative RT-PCR)

Детекция молекул РНК с применением **обратной транскрипции и ПЦР:**

- Обратная транскрипция существенно изменила стратегию исследования транскриптома
- ОТ позволила переносить информацию от нестабильных молекул оцРНК к более стабильным молекулам дцДНК, что обеспечило:
  1) **хранение информации о транскриптах**
  2) **возможность амплифицировать ДНК до необходимых количеств в случае низкой экспрессии**

Количественная ОТ-ПЦР (quantitative RT-PCR)



Принцип метода:

- 1) с образцами РНК проводят реакцию обратной транскрипции
- 2) получение кДНК
- 3) ПЦР-амплификация кДНК
- 4) электрофорез амплификатов

- Количество продукта определяют с помощью стандартного транскрипта, точная концентрация которого определена

Количественная ОТ-ПЦР (quantitative RT-PCR)

Метод количественной ОТ-ПЦР предназначен для детекции мРНК и количественной оценки их содержания в образцах.

- **Достоинства:**

  1) Можно идентифицировать и анализировать несколько разных транскриптов одновременно

- **Ограничения метода :**

  риск неспецифической амплификации

Дифференциальный дисплей (Differential display)

Метод дифференциального дисплея не нуждается в знании структуры и служит для сравнения транскриптов

Выделение РНК(polyA)

1

Обратная транскрипция

2

ПЦР на кДНК

3

Получают транскрипты, содержащие сайт для второго праймера

4

Гель-электрофорез

5

1) проводят реакцию обратной транскрипции

2) проводят ПЦР на кДНК с мечеными нуклеотидами, используя праймеры :

- олигоТ + (N)
- праймер случайного состава размером 10-12н

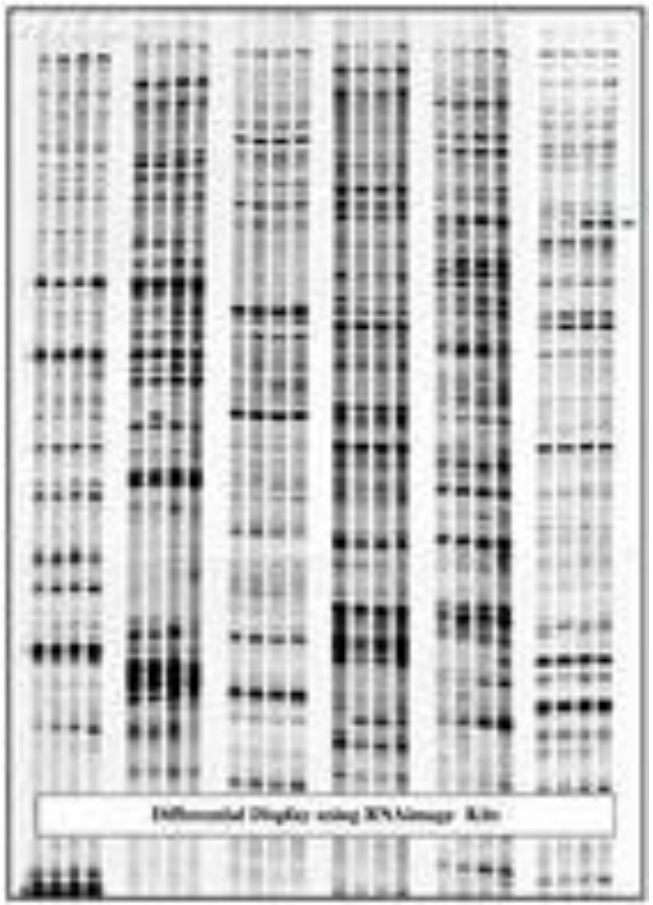
Использование трех вариаций олигоТ-праймера+(N) и набора случайных олигонуклеотидов позволяет вовлечь в сравнение транскриптов большое число генов

3) Получают наборы меченых транскриптов генов, содержащих фрагмент, гомологичный для второго праймера

4) **меченые продукты амплификации анализируют гель-электрофорезом и сравнивают образцы по интенсивности свечения**

Дифференциальный дисплей (Differential display)

5



5) Извлечение из геля отдельных фракций позволяет клонировать и идентифицировать ген

Клонирование и идентификация дифференциально экспрессированных генов

Дифференциальный дисплей (Differential display)

Метод дифференциального дисплея полезен для выявления различий в относительном количестве множества видов мРНК для многих образцов

- **Достоинства:**

❑высокая чувствительность;

❑высокая производительность.

- **Ограничения:**

❑анонимность выявляемых полос и необходимость их дальнейшей идентификации

❑присутствие фальшивых позитивных полос

Серийный анализ экспрессии генов
Serial Analysis of Gene Expression (SAGE)

Серийный анализ экспрессии генов позволяет количественно и качественно охарактеризовать экспрессию тысяч различных генов путем анализа их транскриптов

Основной принцип метода:

1) Выделяют РНК и получают меченые дц кДНК

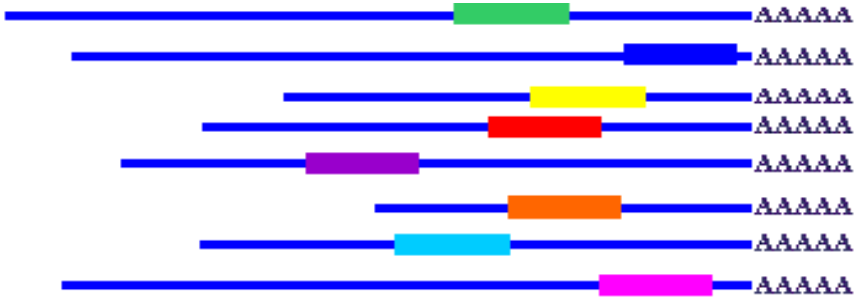
2)получают короткие нуклеотидные последовательности для кДНК («ярлыки») (10—14 п.н.), достаточные для идентификации индивидуальных генных продуктов

3) связывание их друг с другом в одну последовательность (конкатемеризация ярлыков)

4) Секвенирование

5) Анализ

SAGE-метод



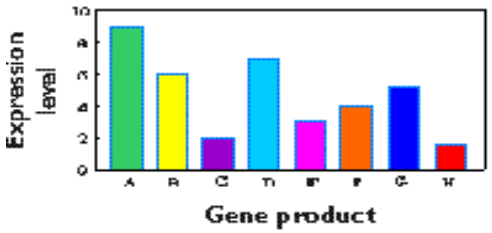
кДНК библиотека

Изолируют таги

Лигируют таги

Секвенируют конкатемер

Определяют экспрессию



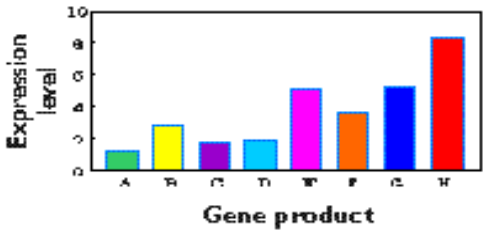


Схема анализа:

1. Получение кДНК
2. Иммобилизация кДНК на мембране
3. Обработка рестриктазой
4. Обработка эндонуклеазой
5. Лигирование
6. Конкатемер-секвенирование
7. Анализ экспрессии

Серийный анализ экспрессии генов

Достоинства метода:

- определение уровней экспрессии генов
- возможность выявлять слабо экспрессирующиеся гены
- возможность выявлять новые гены
- информативность

Ограничения:

- применяют только к полиаденилированным транскриптам

Метод учета маркерных экспрессирующихся последовательностей (Expressed Sequence Tags)

EST - метод исследования транскриптов и их дифференциального распределения

Принцип метода EST:

- суть метода заключается в систематическом секвенировании библиотек кДНК
- секвенированию подвергается фрагмент последовательности каждой молекулы кДНК (300-500 п.о.)
- такой подход позволил ускорить идентификацию новых транскриптов

Метод учета маркерных экспрессирующихся последовательностей (Expressed Sequence Tags)

- **Expressed Sequence Tags (EST)-** короткие (300-500 bp), прочитанные за один раз фрагменты кДНК
- **EST представляют «отпечаток» с продуктов гена в определенной ткани на определенной стадии развития**
- **EST являются маркерами экспрессии гена для определенной библиотеки кДНК.**
- **EST служат для выравнивания генома и транскриптома**

TABLE

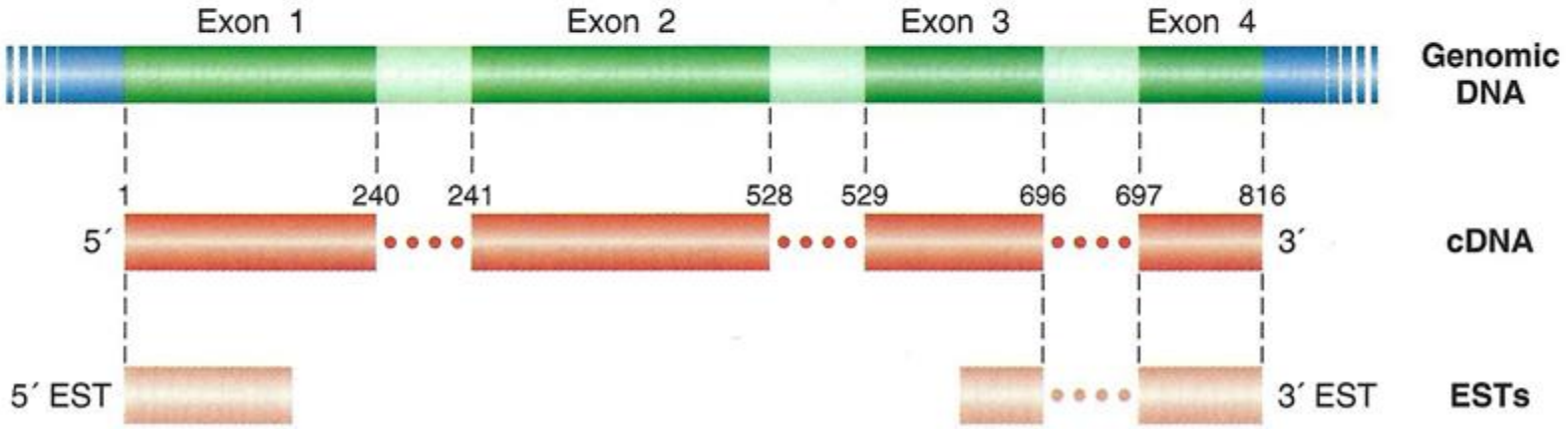
Most abundant amastigote expressed sequence tags (ESTs)

Putative product	Number of ESTs
Histone H3	18
Ribosomal protein S9	15
Cytochrome-c oxidase chain III	13
Heat shock protein 90	8
Cytochrome b	7
GP85	7
Histone H2A	6
Amastin	6
Ribosomal protein S12	5
Elongation factor 1-alpha	5
Mucins	5
Casein kinase 1 isoform 2	5

- **EST используют для изучения размеров, разнообразия и транскрипционной активности экспрессирующихся в организме генов**

Метод учета маркерных экспрессирующихся последовательностей (Expressed Sequence Tags)

Выравнивание сиквенсов кДНК и EST-последовательностей по отношению к геному:



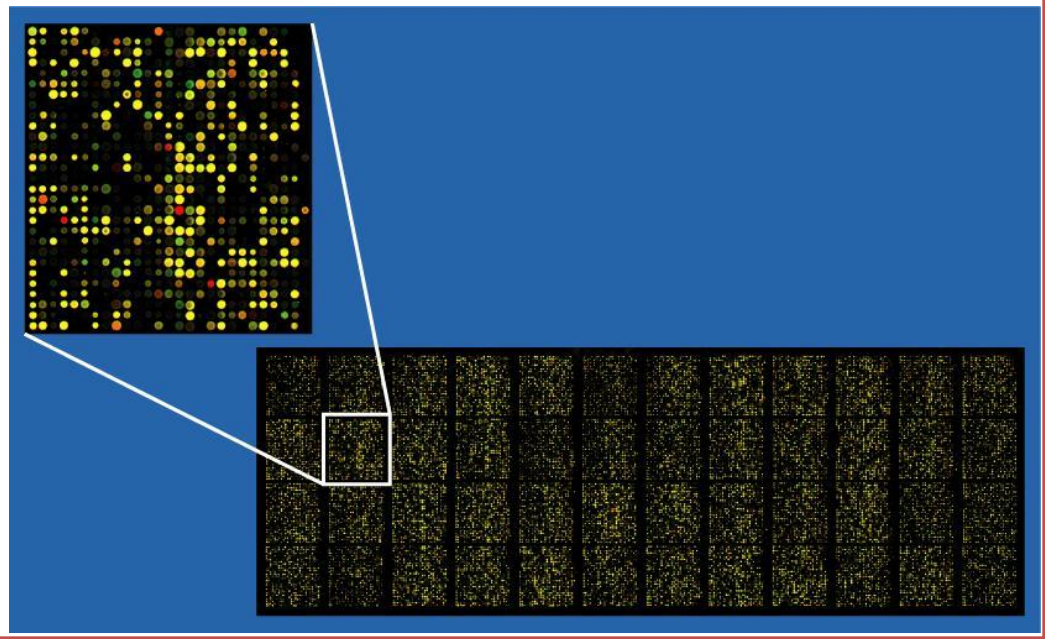
dbEST: database of "Expressed Sequence Tags"

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
• Danio rerio (zebrafish)	689,581
• Zea mays (maize)	656,945
• Triticum aestivum (wheat)	600,039
• Gallus gallus (chicken)	578,445
• Sus scrofa (pig)	502,501
• Arabidopsis thaliana (thale cress)	420,789
• Oryza sativa (rice)	406,790
• Hordeum vulgare + subsp. vulgare (barley)	395,019
• Drosophila melanogaster (fruit fly)	383,407
• Pinus taeda (loblolly pine)	329,469
• Caenorhabditis elegans (nematode)	302,080

Анализ транскриптома на основе микрочипов

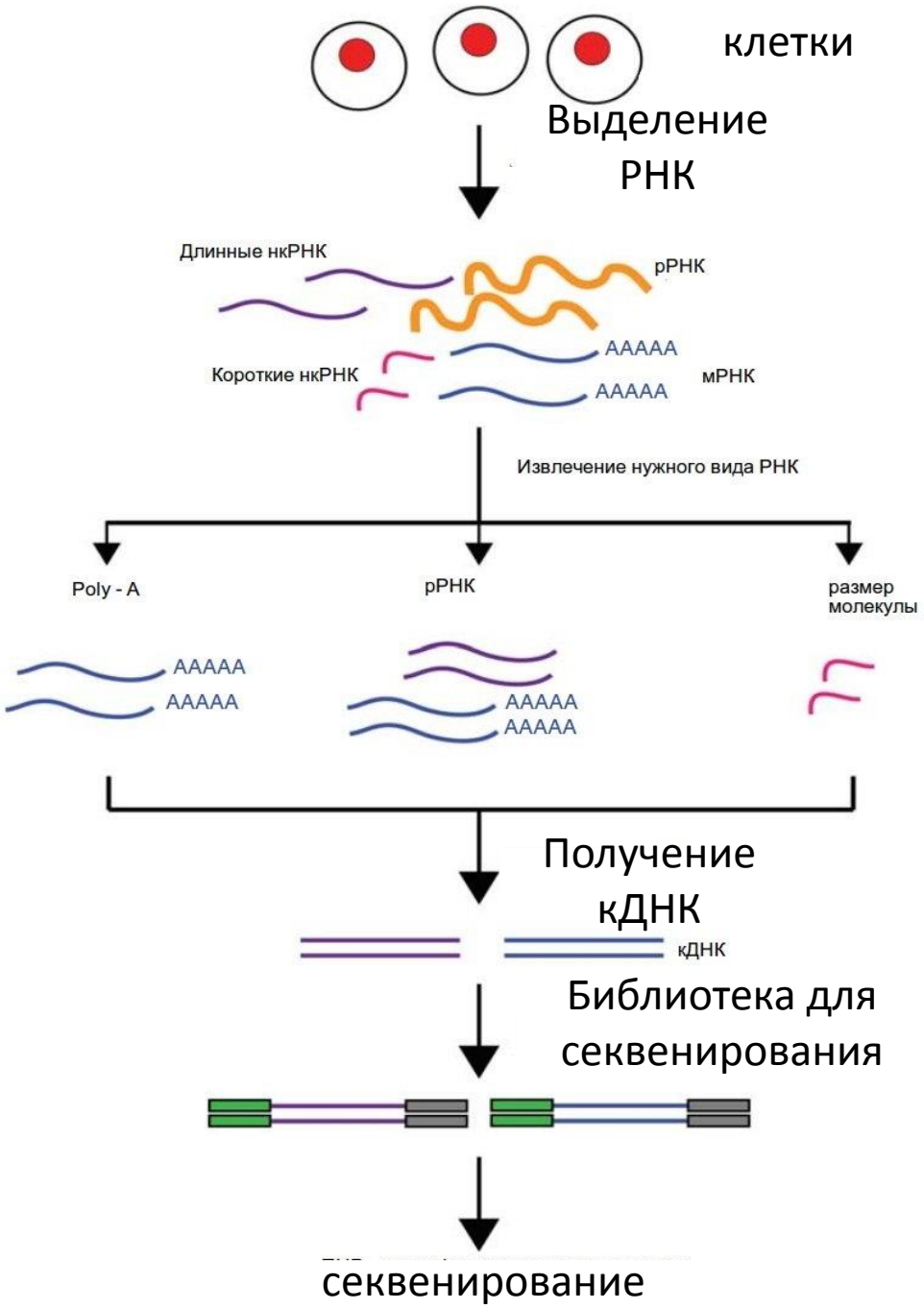
- **В основе метода** - **гибридизационные подходы**, основанные на комплементарном взаимодействии меченых кДНК или РНК с собственными ДНК-микрочипами
- **Экспериментальные стадии анализа :**

1. Дизайн или покупка чипов
2. Выделение РНК
3. Обратная транскрипция (с мечением) = кДНК
4. ПЦР
5. Гибридизация на чипе
6. Отмывка
7. Сканирование



РНК-секвенирование (RNA-Seq)

- Технологией полногеномной транскриптомики является сиквенсный подход - **RNA-Seq**
- **Принцип RNA-Seq заключается в секвенировании кДНК на высоко производительных платформах нового поколения**
- В результате секвенирования получаются риды длиной 300 нуклеотидов. Риды выравнивают на референсном геноме или референсных транскриптах, либо осуществляют сборку de novo без референсной последовательности
- В результате получают полногеномные транскрипционные карты, включающие качественные (структурные) и количественные характеристики экспрессии каждого гена



Секвенирование РНК (RNA-Seq)

Высокопроизводительные технологии секвенирования произвели революцию и в транскриптомике

Технология RNA-Seq состоит из выделения РНК, преобразования ее в комплементарную ДНК (кДНК), построения библиотеки секвенирования и ее последовательности на платформе NGS:

Секвенатор Illumina HiSeq 2000

- 2 flow cells (can be run independently)
- Up to 320Gb mapped sequence per FC
- 64Gb sequence per day (2 flow cells)

Длина ридов до 500 н

Платформа **Illumina HiSeq** является стандартом технологии секвенирования следующего поколения для RNA-Seq
Платформа имеет две проточные ячейки, каждая обеспечивает восемь отдельных полос для реакции секвенирования, которое может занять до 12 дней.

Illumina выпустила MiSeq - секвенатор с более низким выходом ридов, но более высокой скоростью работы (30 миллионов парных ридов за 24 часа). Упрощенный рабочий цикл секвенатора MiSeq обеспечивает быстрое время секвенирования транскриптомов в меньших масштабах.

Секвенатор PacBio RS II

Длина ридов более 10000 пн

Одномолекулярная платформа PacBio – мономолекулярное секвенирование в режиме реального времени. Используется ДНК-полимераза для синтеза ДНК с флуоресцентно меченными нуклеозидами, каждое основание включается в растущую цепь ДНК и в режиме реального времени регистрируется характерный импульс флуоресценции.

Преимущества:

1) не включает этап амплификации, избегая ошибок амплификации и улучшая равномерное покрытие транскриптома

2) генерирует риды более 10 000 bp, что значительно улучшает обнаружение новых транскриптов

Заключение

- **Транскриптом** занимает промежуточное положение между статичным геномом и динамичным протеомом
- Такая позиция делает **транскриптом** удобным для изучения взаимодействия между ними и расшифровки механизмов регуляции
- Информация об экспрессии генов (тип ткани или клеток, стадия развития, факторы воздействии и др.), степени ее интенсивности служит фундаментом для изучения **протеома**