

Введение в транскриптомный анализ

Анастасия Жарикова

1 декабря 2020

azharikova89@gmail.com

Нужны ли нам дополнительные занятия?

Задача

Секвенировать транскриптом

Типы РНК в клетке

- **Тотальная РНК**
- **полиА**
- **Без фракции рРНК**
- **По размеру:**
 - Малые РНК:
 - микроРНК
 - малые ядерные РНК
 - малые ядрышковые РНК
 - малые интерферирующие РНК
 - пиРНК
 - «Длинные» РНК:
 - мРНК
 - длинные некодирующие РНК
- **По внутриклеточной локализации:**
 - Ядерные
 - Цитоплазматические
- ...

Процесс

1. Подготовка нужной фракции РНК
2. Проверка качества РНК
3. Обратная транскрипция => кДНК
4. Фрагментация (~ 200-300 нк)
5. Секвенирование (чем глубже, тем лучше)

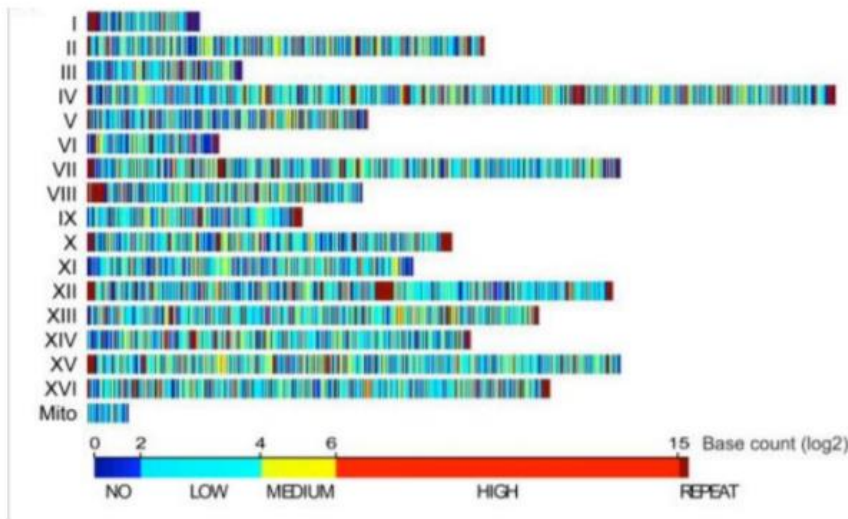
Технические реплики – повторный анализ одного и того же образца
Биологические реплики – повторное взятие образца и анализ

Задачи

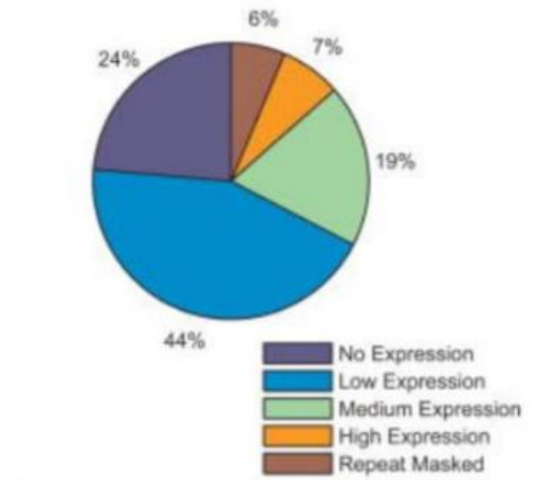
1. Определение концентрации РНК, сравнение в нескольких образцах (виды, ткани, исследования типа «случай-контроль», анализ раковых транскриптомов)
2. Выявление однонуклеотидных полиморфизмов (SNP)
3. Детекция мест альтернативного сплайсинга
4. Поиск некодирующих РНК
5. Редактирование РНК
6. Анализ транскриптомов единичной клетки

Анализ транскриптомов

Первые работы по секвенированию транскриптомов появились в 2008 году

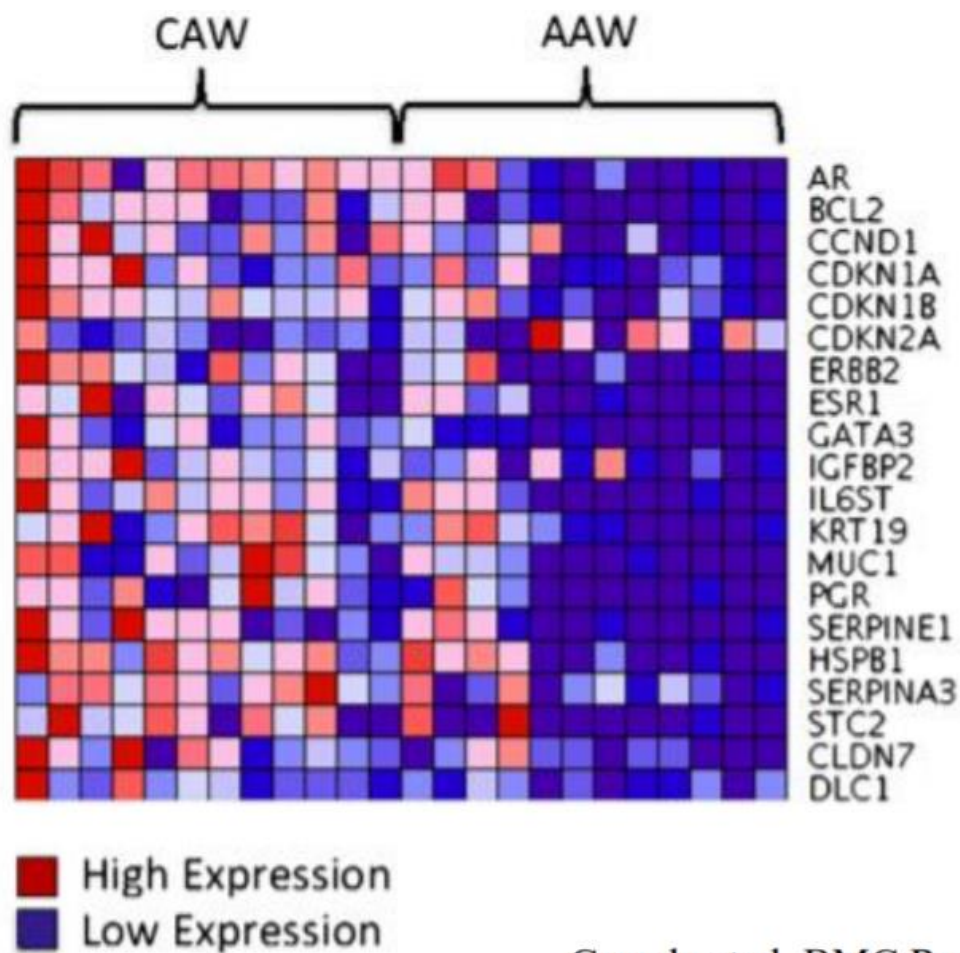


дрожжи



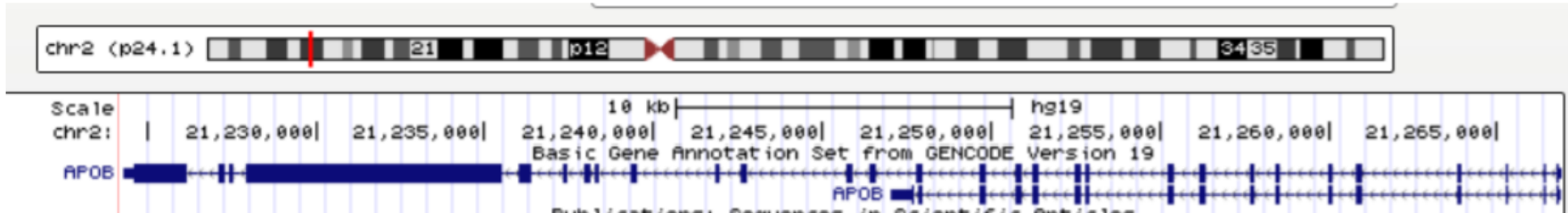
Science. 2008 Jun 6; 320(5881): 1344–1349.

Дифференциальная экспрессия



Grunda et al. BMC Research Notes 2012, 5:248

Альтернативный сплайсинг

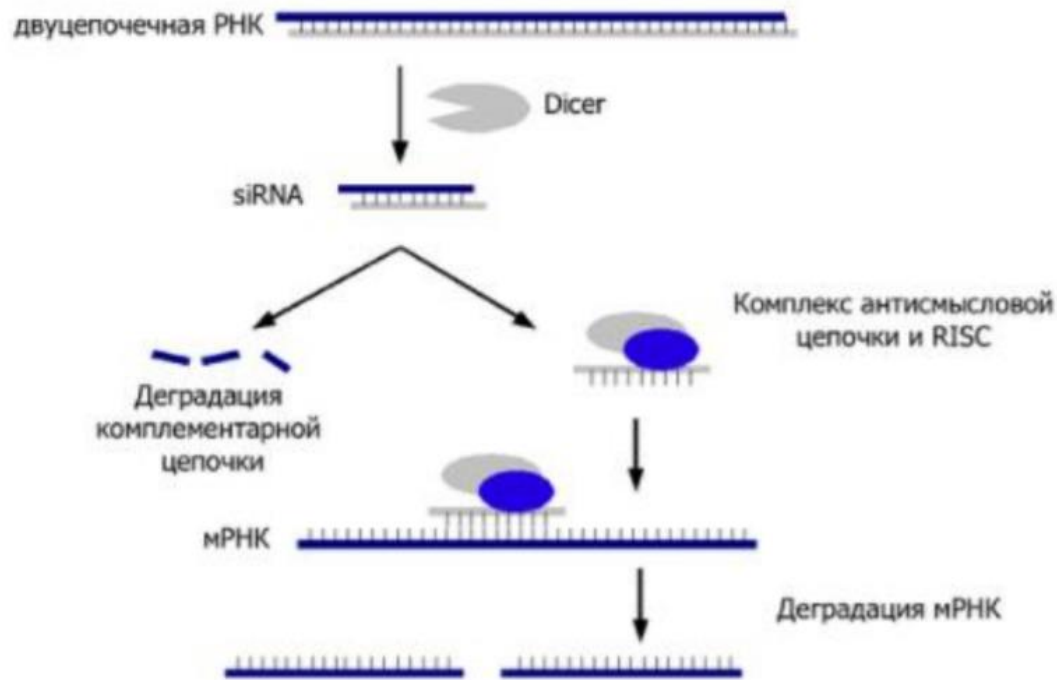


АпоВ-100 – длинный транскрипт – синтезируется в печени.

АпоВ-48 – короткий транскрипт – синтезируется в кишечнике.

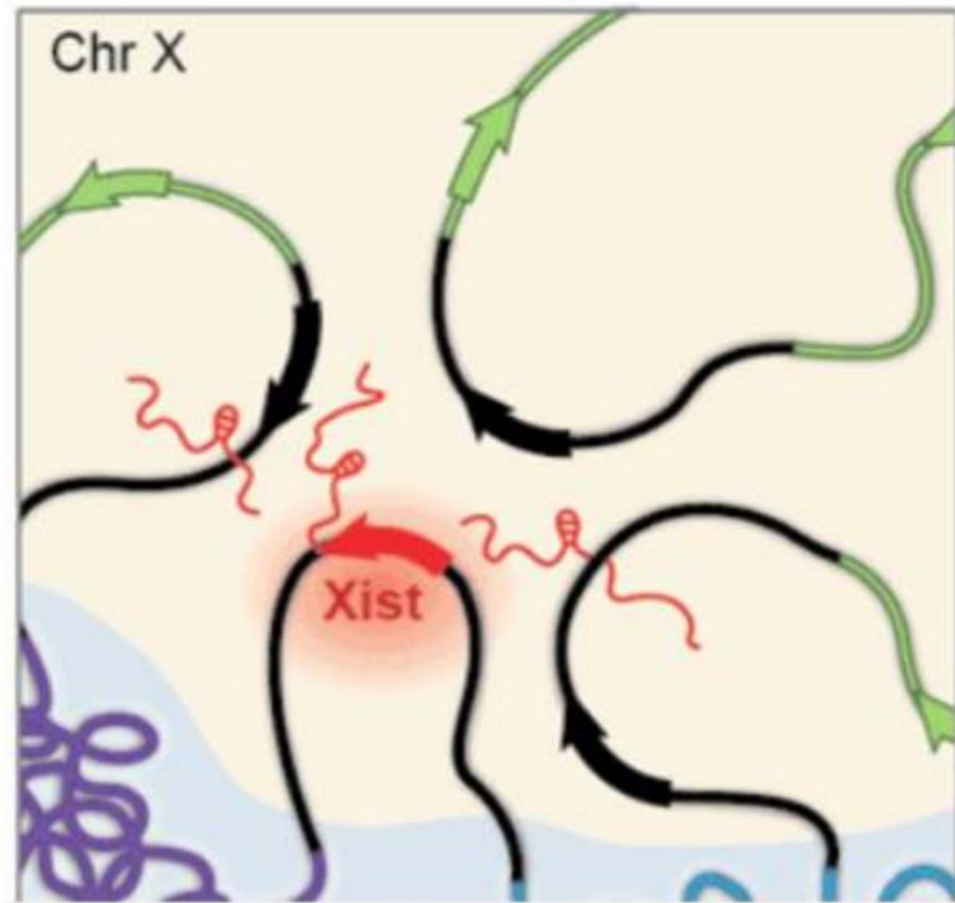
Синтезирующиеся белки входят в состав разных групп липопротеинов, которые в последующем идут каждый своим путем метаболизма

Некодирующие РНК



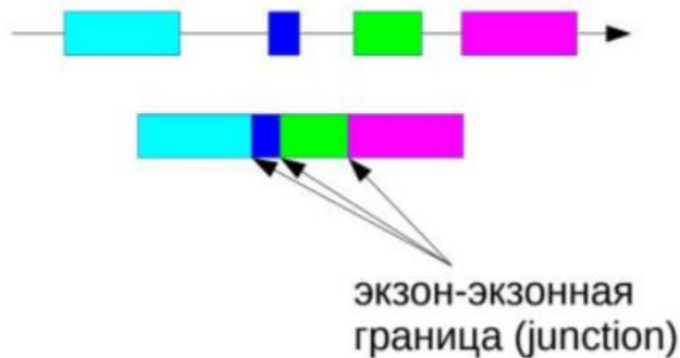
За открытие механизма РНК-интерференции в 2006 году присуждена Нобелевская премия по медицине

Некодирующие РНК



Science, 2013

Картирование

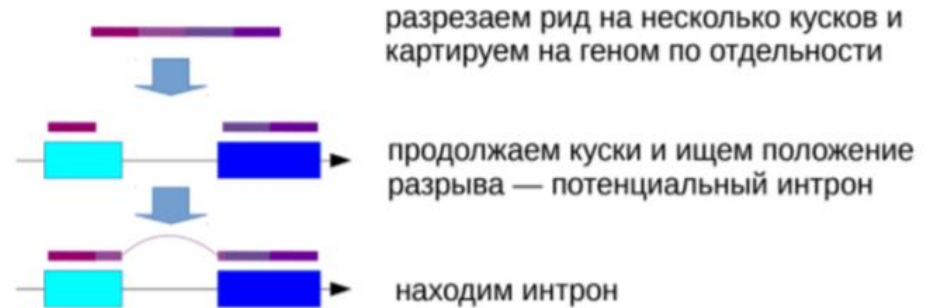


Использование аннотации:

- Только аннотированные экзон-экзонные границы
- Все возможные экзон-экзонные границы



Предсказание аннотации из данных:

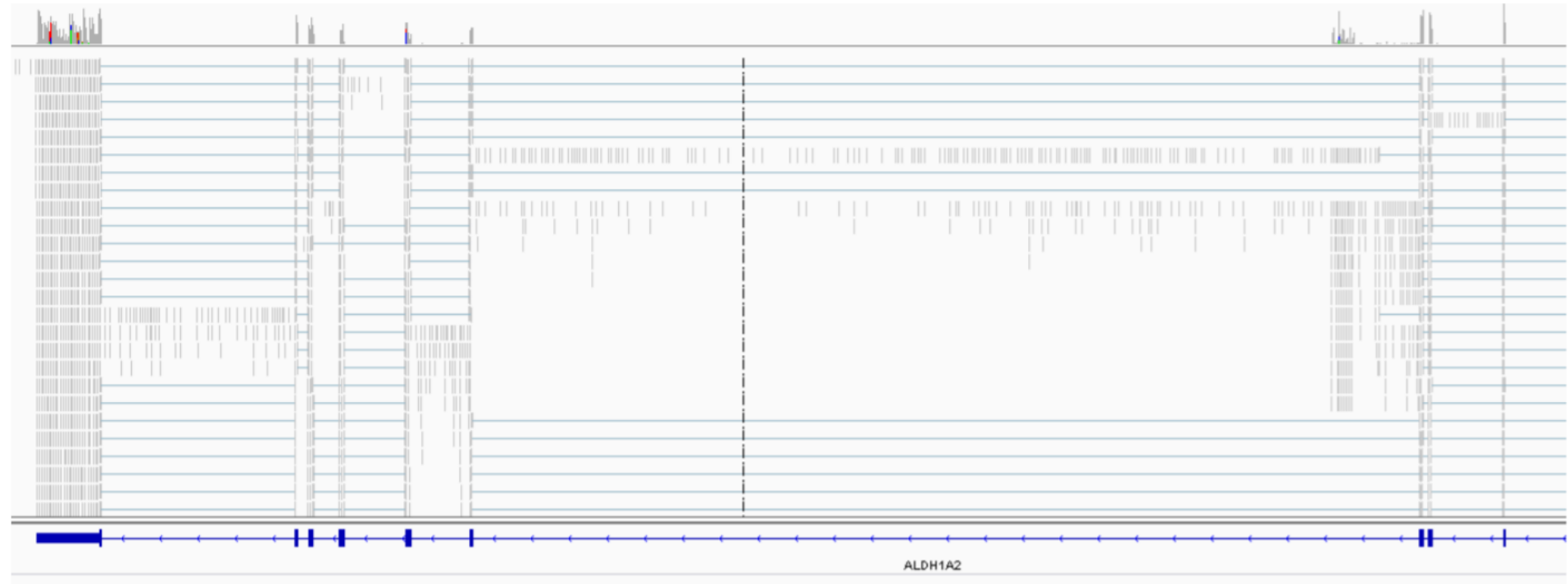


Индексирование генома

1. Аналогично задаче картирования экзомного секвенирования
2. Индексирование с учетом разметки (.gtf)
 - Экстракция из аннотации экзонов
 - Экстракция из аннотации сайтов сплайсинга
 - Индексирование с использованием списка экзонов и сайтов сплайсинга

hisat2

IGV



Глубина покрытия важна

В клетке присутствуют 2 варианта транскриптов одного гена:



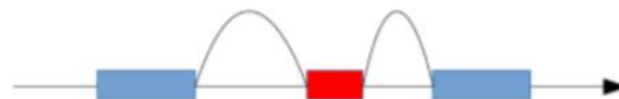
Секвенировали два образца: один получился с хорошим покрытием, другой – с плохим

Видим:

Хорошо покрытый образец: все экзон-экзонные границы нашлись



Плохо покрытый образец: один вариант потерялся



Дифференциальный альтернативный сплайсинг?

Дифференциальная экспрессия?

Нужно нормировать на размер библиотеки и оценивать нормировочные коэффициенты!

Подсчет чтений – Htseq-count

	union	intersection_strict	intersection_nonempty
	gene_A	gene_A	gene_A
	gene_A	no_feature	gene_A
	gene_A	no_feature	gene_A
	gene_A	gene_A	gene_A
	gene_A	gene_A	gene_A
	ambiguous (both genes with --nonunique all)	gene_A	gene_A
	ambiguous (both genes with --nonunique all)		
	alignment_not_unique (both genes with --nonunique all)		

```

ENSG00000000005.5      0
ENSG000000000419.8    23
ENSG000000000457.9    397
ENSG000000000460.12   239
ENSG000000000938.8     0
ENSG000000000971.11    13
ENSG00000001036.9     19
ENSG00000001084.6      1
ENSG00000001167.10    12
ENSG00000001460.13    23
    
```

```

__no_feature      13696354
__ambiguous       66168
__too_low_aQual  0
__not_aligned    0
__alignment_not_unique  0
    
```


Что можно сделать дальше?

- Оценить самосогласованность образцов
- Сколько чтений легло в границы разметки?
- Подсчет дифференциальной экспрессии
 - GO аннотация
 - Анализ альтернативного сплайсинга
 - Сборка аннотации
 - ...

