

# Базовая биоинформатика

Лекция 6

Оценки экспрессии генов

05 / 11 / 2020

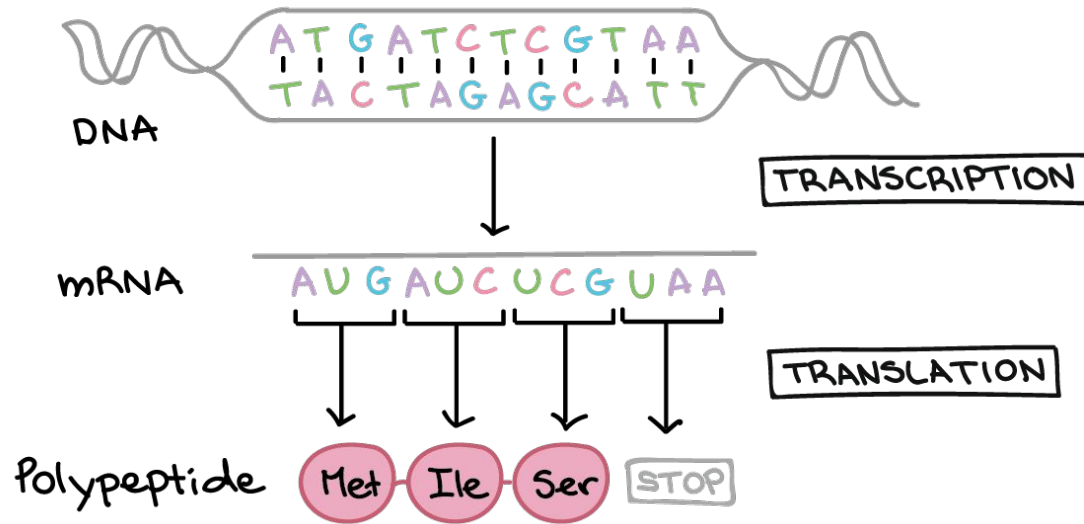


- Выпускаюсь с ФББ МГУ (сейчас на 6 курсе)
- И. о. младшего научного сотрудника в ИППИ РАН
- Преподаю (частично) практическую биоинформатику и NGS на ФББ МГУ
- Преподаю биоинформатику школьникам
- Раньше занимался сравнительной геномикой разных беспозвоночных зверушек и метагеномикой
- Сейчас занимаюсь анализом scRNA-Seq данных человека в BostonGene

- Экспрессия гена
- Основы биологии РНК
- Микрочиповые технологии
- Процедура RNA-Seq

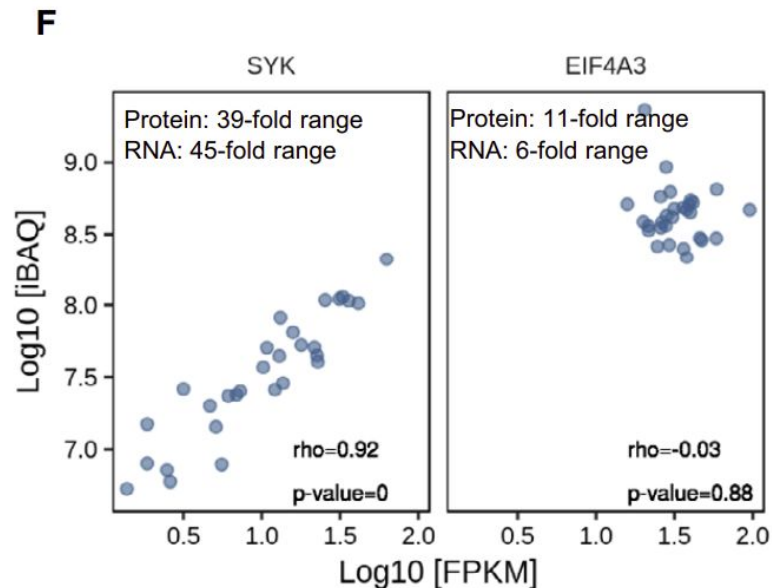
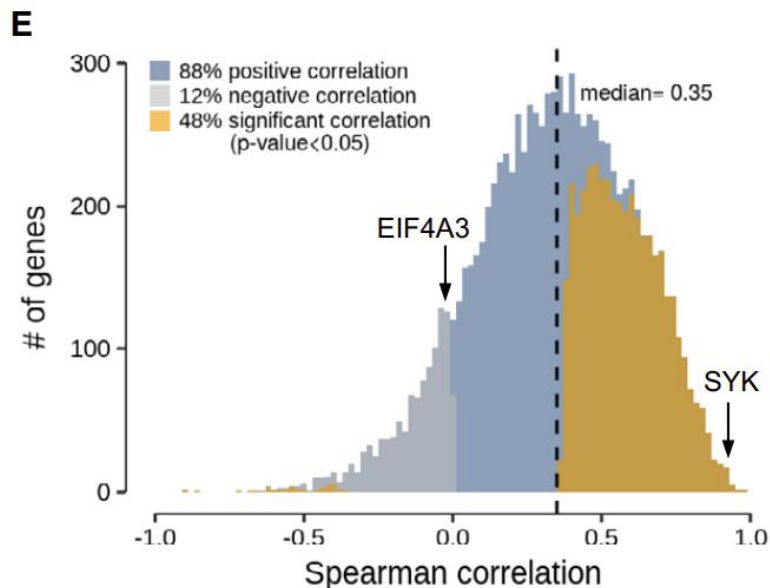
**Что мы хотим?**

# Зачем мы изучаем РНК?

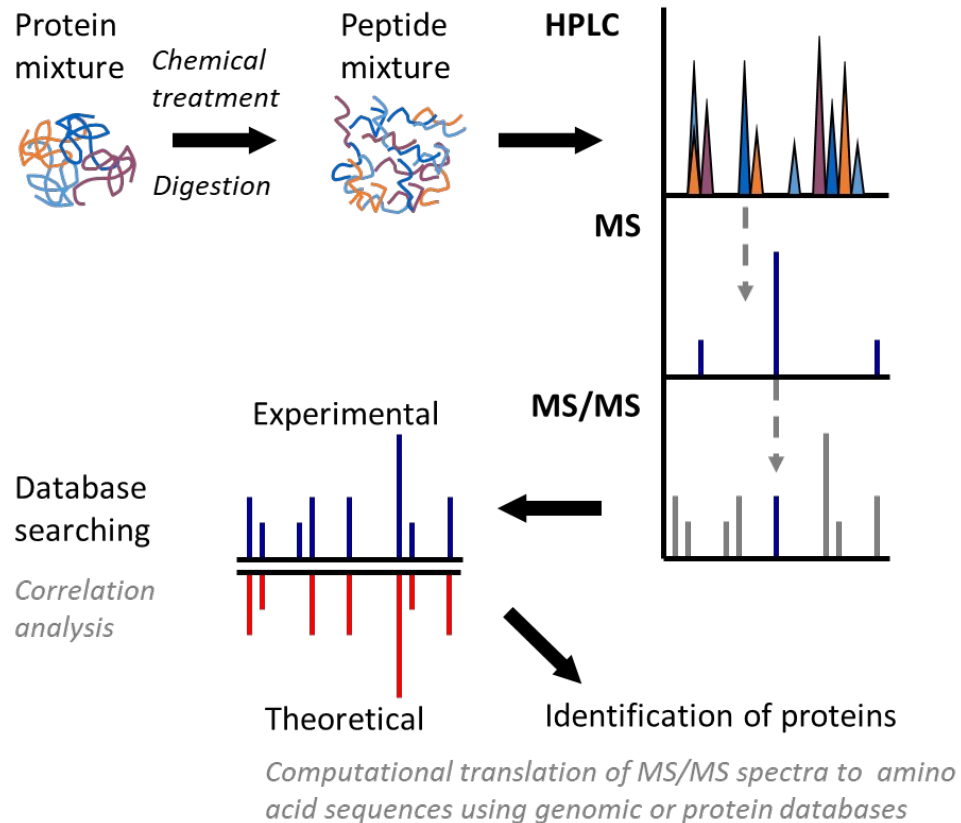


Фенотип ≈ Белки ≈ Транскрипты

- Представленность белка не всегда строго коррелирует с тем, какие транскрипты находятся в клетке



- Можно напрямую оценивать, какие белки и в каком количестве находились в клетке.
- Это дорого и не очень точно.

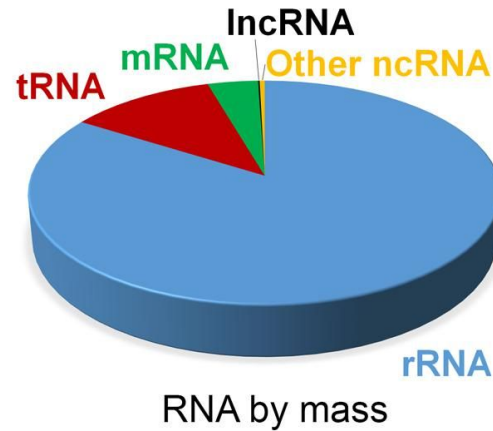


# Биология РНК

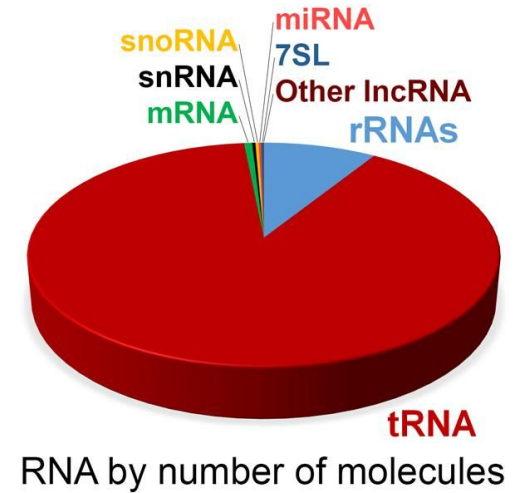


1. Малые РНК:
  - а. микроРНК
  - б. мяРНК
  - с. мякРНК
  - д. миРНК
  - е. пиРНК
  - ф. тРНК
2. “Длинные” РНК:
  - а. мРНК**
  - б. рРНК
  - с. днкРНК

A

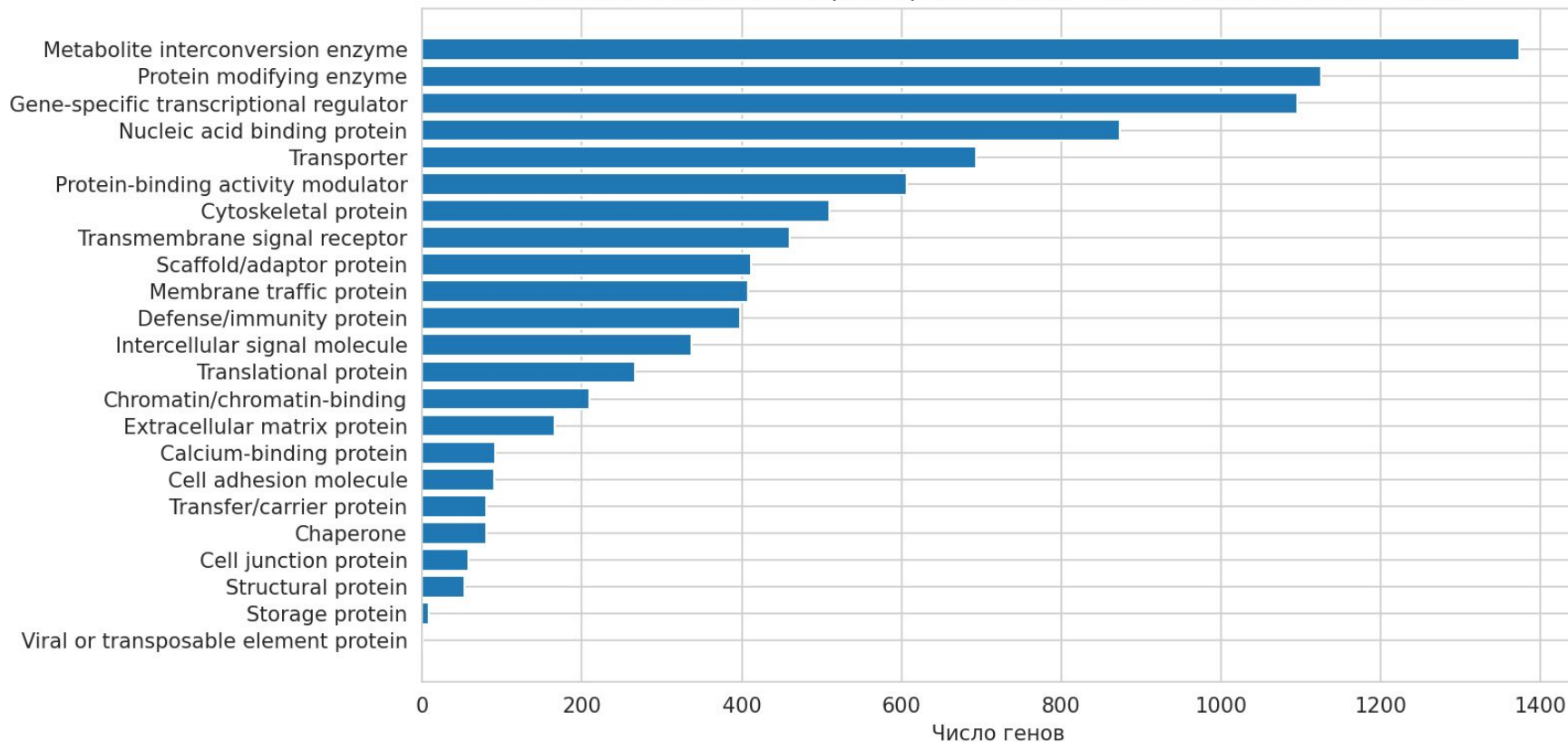


B



# Разнообразие белок-кодирующих генов

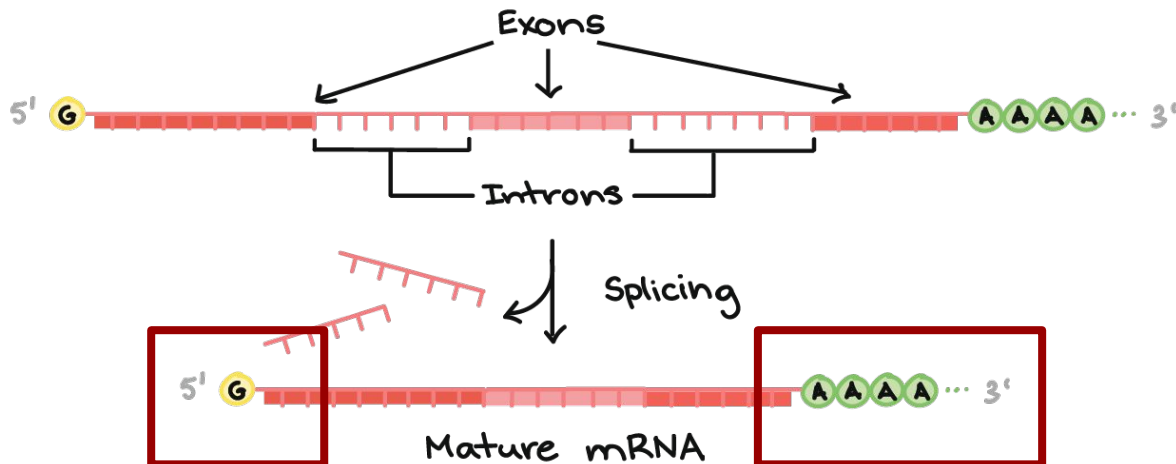
Классы белков, к которым принадлежат гены по Panther (9397 генов)



- Сначала пре-мРНК **кэпируется** и **полиаденилируется** (у эукариот)

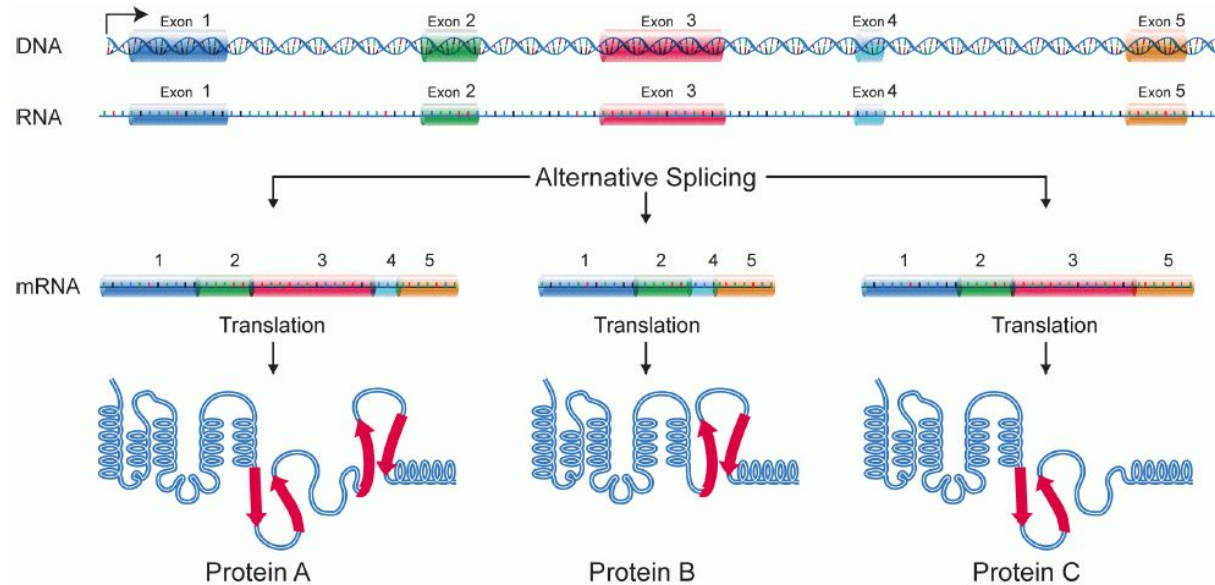


- Затем из мРНК вырезаются интроны (**сплайсинг**)



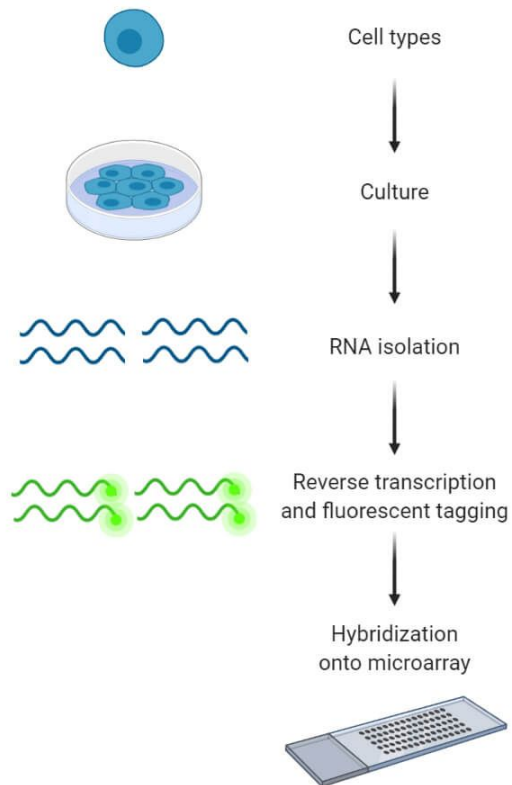
# Альтернативный сплайсинг

- У одного и того же гена в ходе сплайсинга могут встраиваться различные экзоны
- Примерно 95% генов подвергаются альтернативному сплайсингу



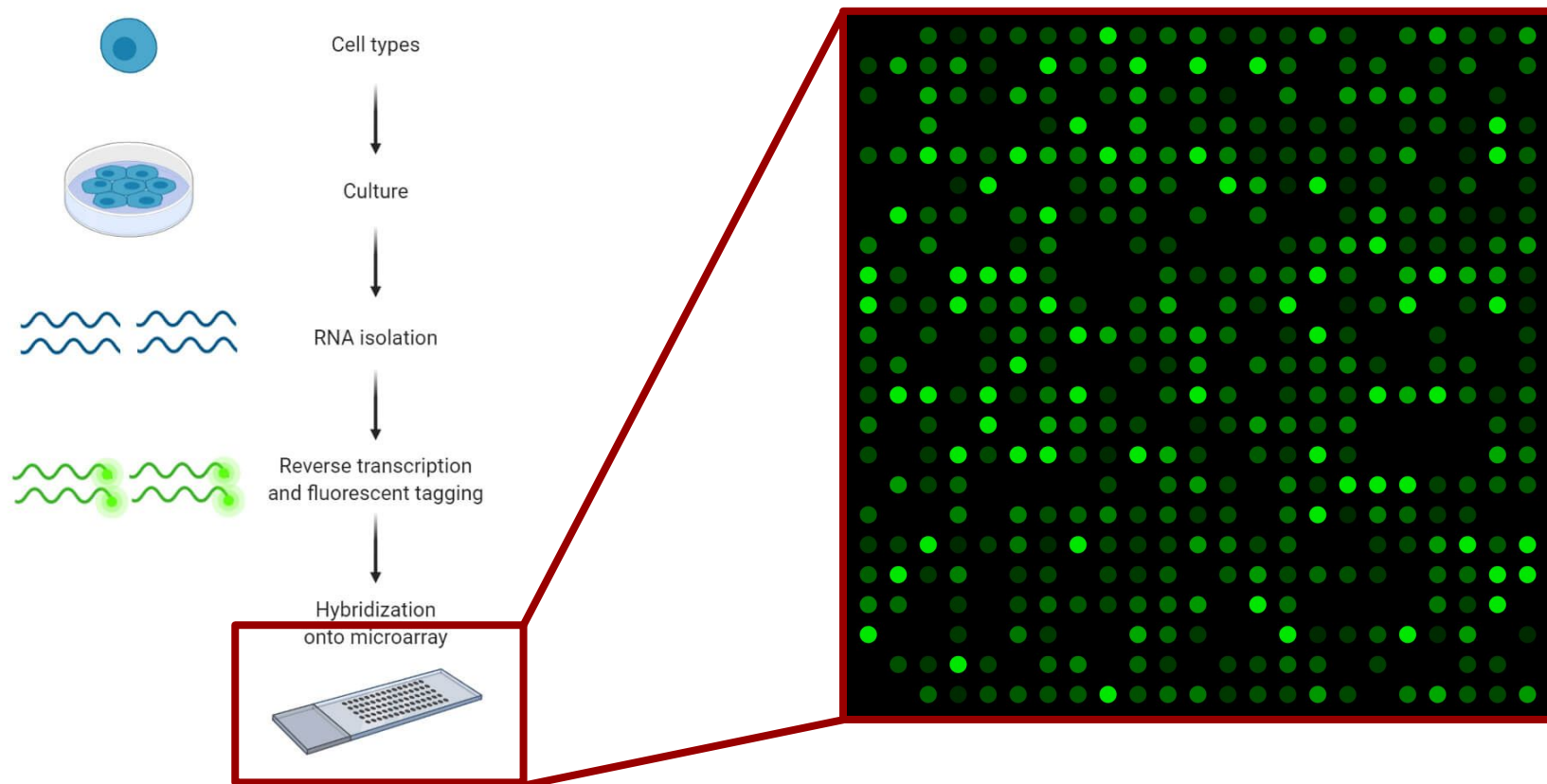
- Сейчас практически не существует методов прямой работы с РНК. Нам всегда необходимо из РНК делать кДНК (её “копию” в виде ДНК)
- Синтезировать ДНК на матрице РНК умеет **обратная транскриптаза** (РНК-зависимая-ДНК-полимераза). Чаще всего про неё можно услышать в контексте ретровирусов
- Важно, что, как и для обычной ДНК-полимеразы, ей для синтеза нужна затравка (**праймер**). Ретровирусы используют для этих целей тРНК клетки-хозяина

# Микрочипы

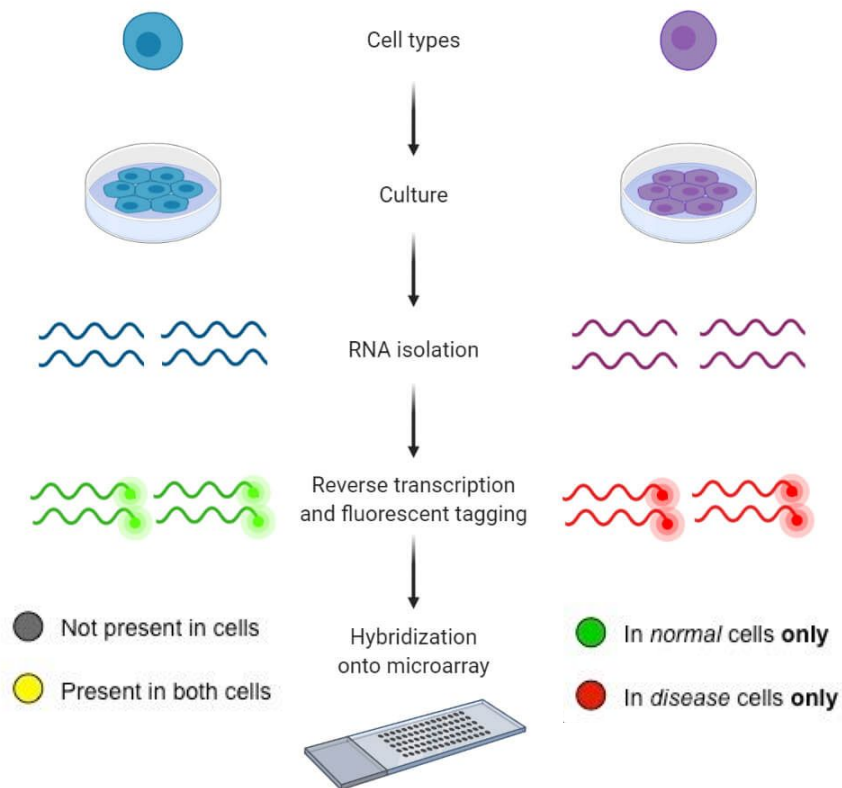


- Сначала выделяют РНК и получаем из неё кДНК при помощи обратной транскрипции
- Затем к кДНК пришивается флуоресцентная метка
- кДНК с меткой гибридизуется со специальным **микрочипом**, к которому “пришиты” комплементарные транскриптам последовательности
- Затем при помощи лазера детектируется свечение на микрочипе. **Интенсивность свечения ~ число транскриптов**

# Одноканальные микрочипы

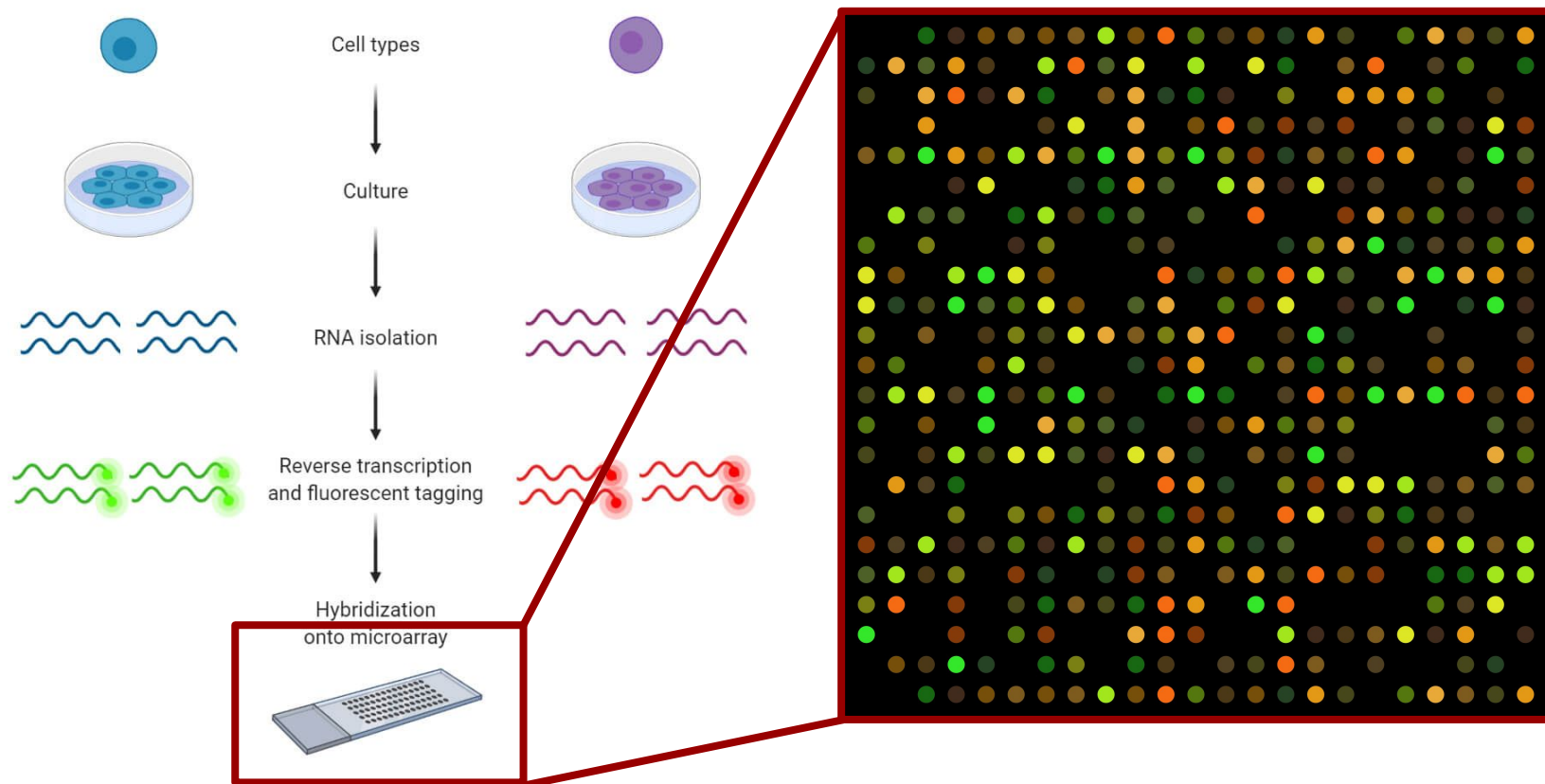






- Также существует вариация метода, которая позволяет сразу же **сравнить экспрессию** генов в двух образцах
- Для этого в образцах метят кДНК разными флуоресцентными метками и измеряют детекцию сразу в двух спектрах лазера
- Такие чипы называются **двуканальными**

# Двуканальные микрочипы



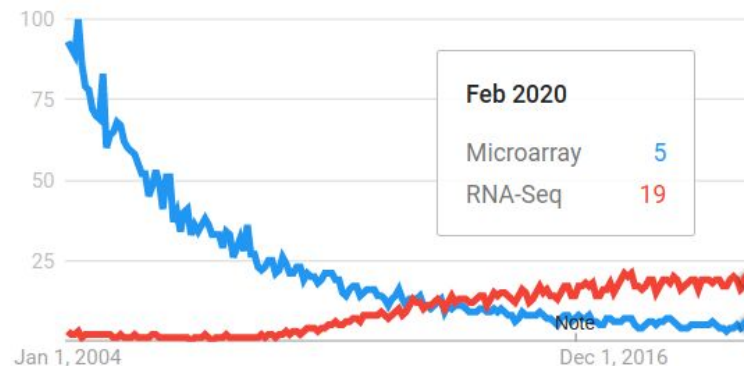
# Микрочипы vs. RNA-Seq

## Плюсы **микрочипов**:

- Это дешевле, чем RNA-Seq
- Проще обработка и анализ

## Плюсы **RNA-Seq**:

- Не нужны априорные знания о последовательности
- Получаем абсолютные значения экспрессии
- Можно идентифицировать новые транскрипты и события сплайсинга
- Нужны меньшие концентрации исходной РНК (1ng против 1μg)



# Процедура RNA-Seq

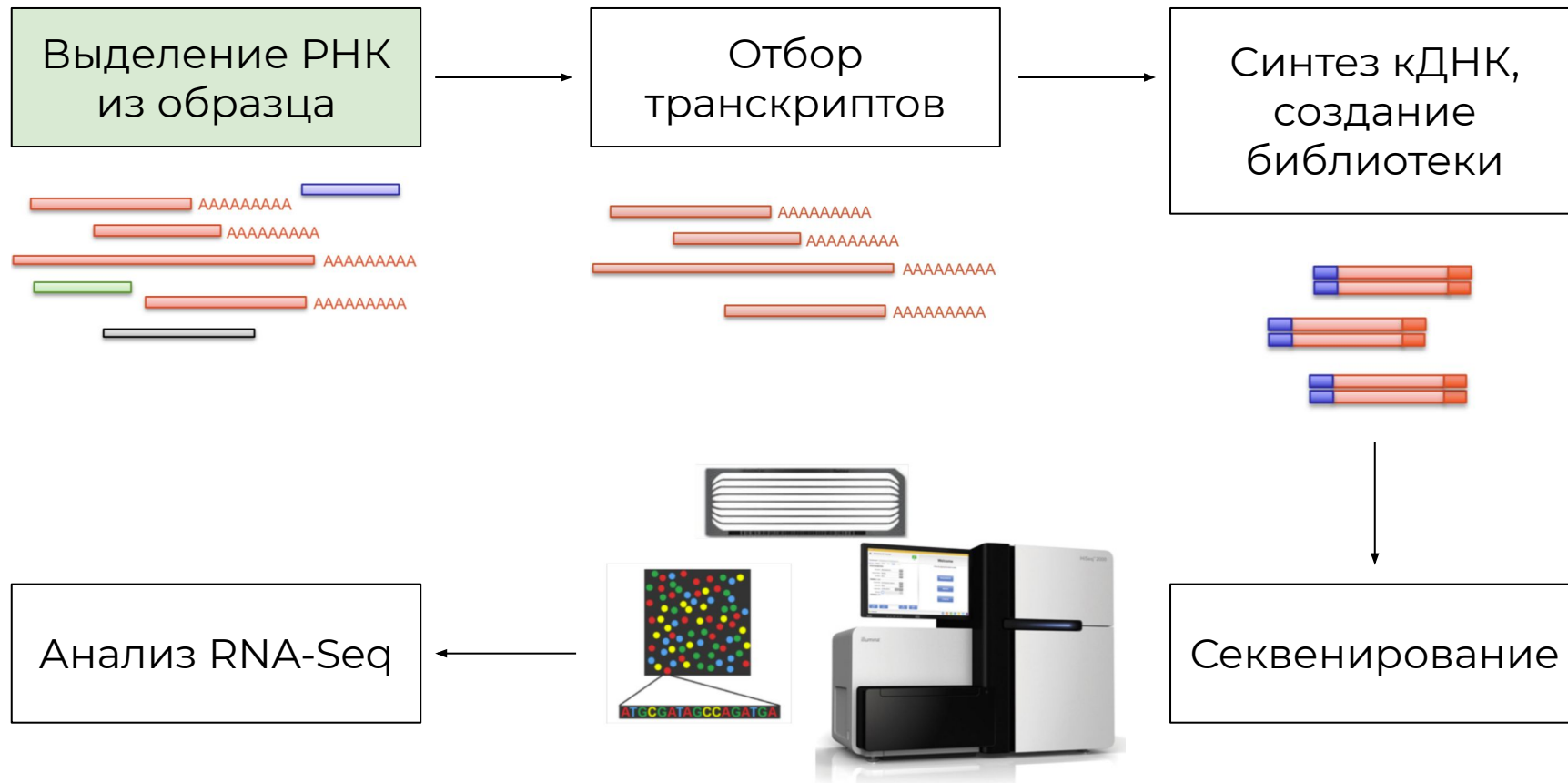
# Процедура RNA-Seq

BostonGene



# Процедура RNA-Seq

BostonGene



# Как выглядит образец?

**Формалин и парафин (FFPE —**  
formalin-fixed, paraffin-embedded)

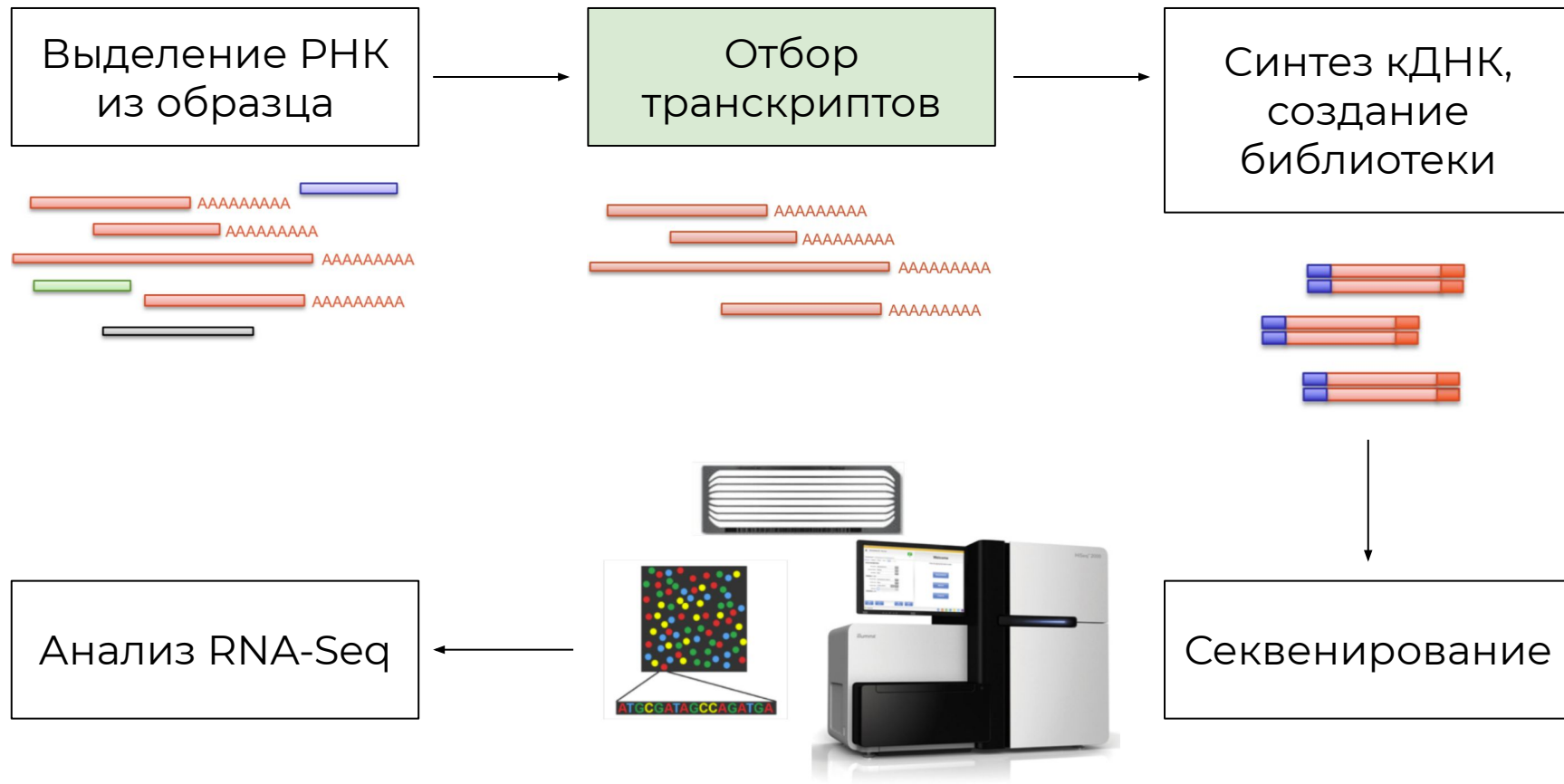


**Заморозка в жидком азоте (FF —**  
fresh frozen)



# Процедура RNA-Seq

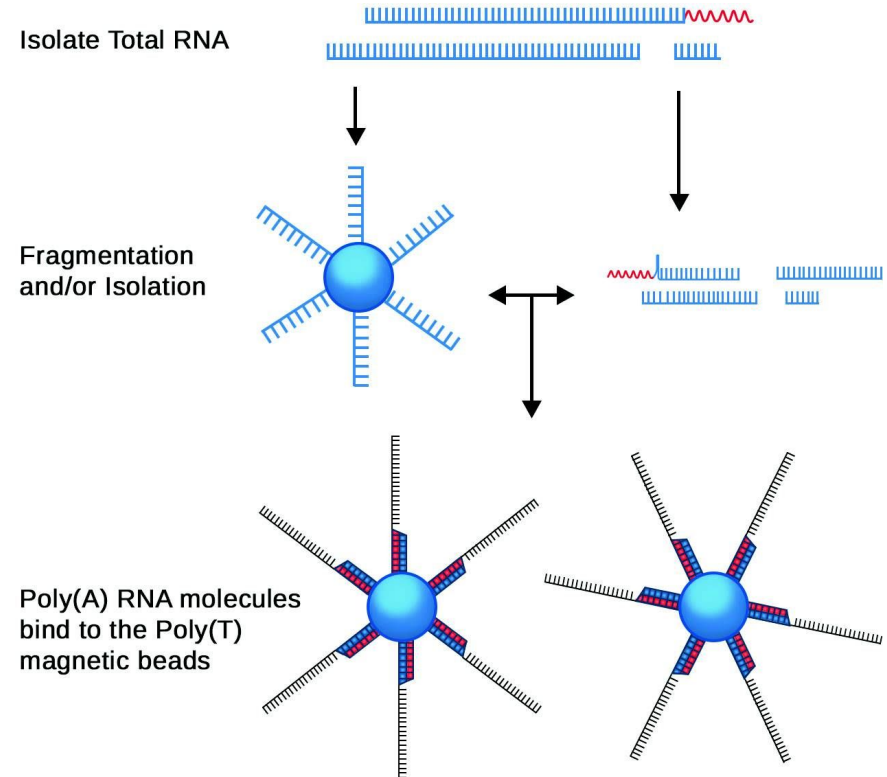
BostonGene





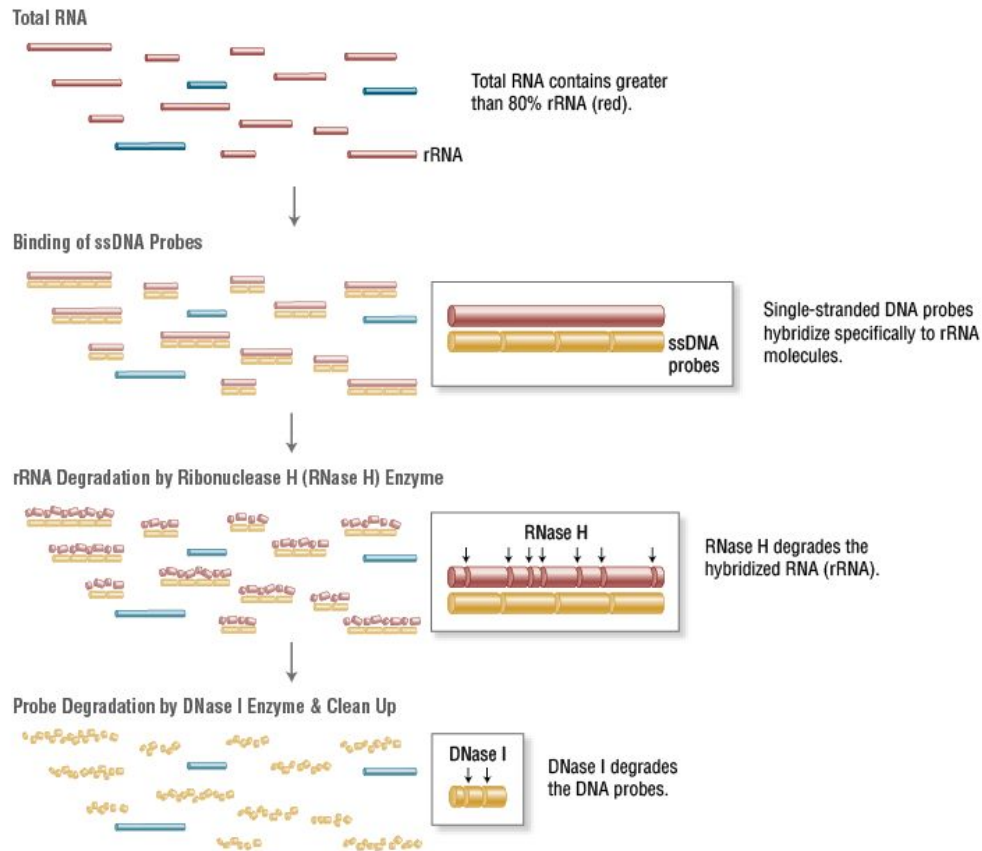
# oligo-(dT)-гибридизация

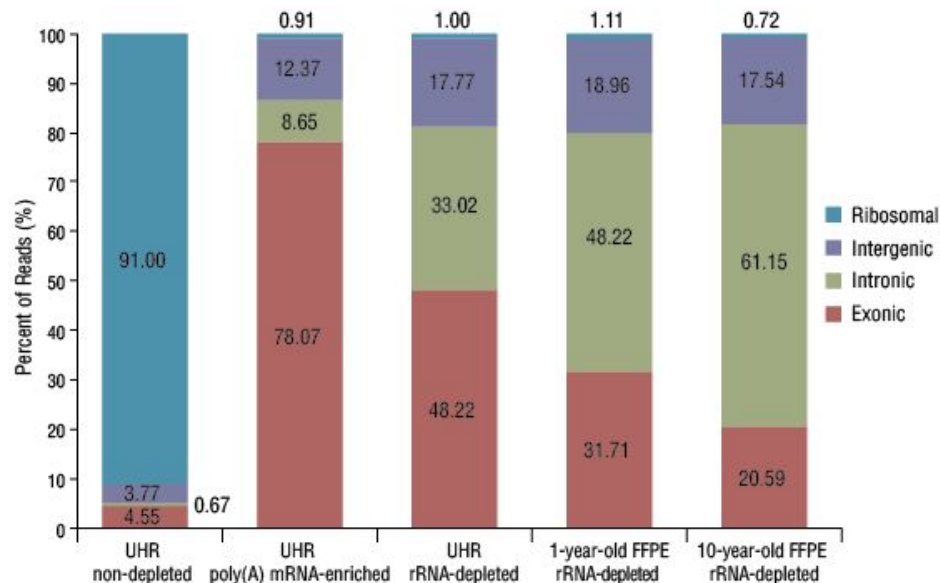
- В этом случае мы отбираем только транскрипты с poly(A)
- Имеются магнитные шарики, к которым присоединены oligo-(dT)
- С этими шариками гибридизуется зрелая мРНК
- Затем эти шарики вместе с мРНК вытаскиваются за магнит
- мРНК отмывается от шариков



# Деплеция рРНК

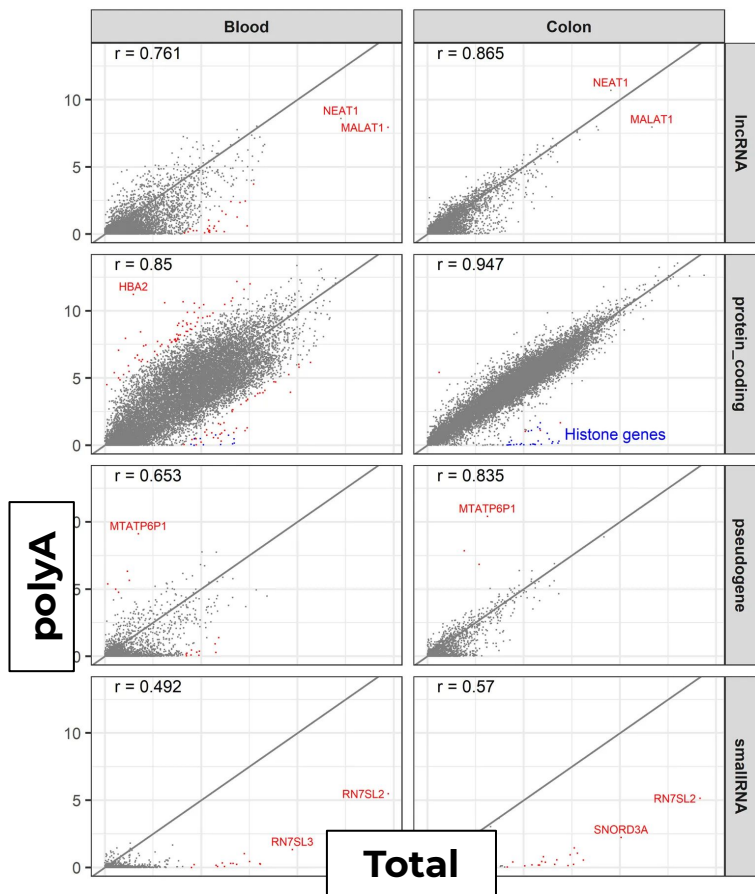
- Добавляем ДНК, комплементарную к рРНК
- Образуются дуплексы, которые впоследствии режутся при помощи РНКазы Н
- В пробе остаются все транскрипты, кроме рРНК



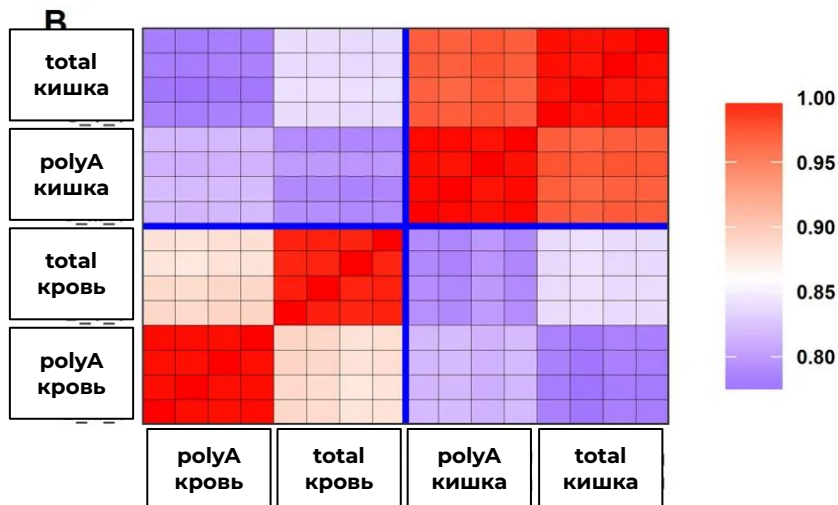


- Без выделения целевых транскриптов больше всего прочтений приходится на рРНК
- При poly(A)-подготовке больше всего экзонных прочтений
- С течением времени FFPE-фиксированные образцы постепенно деградируют

# Сравнение методов



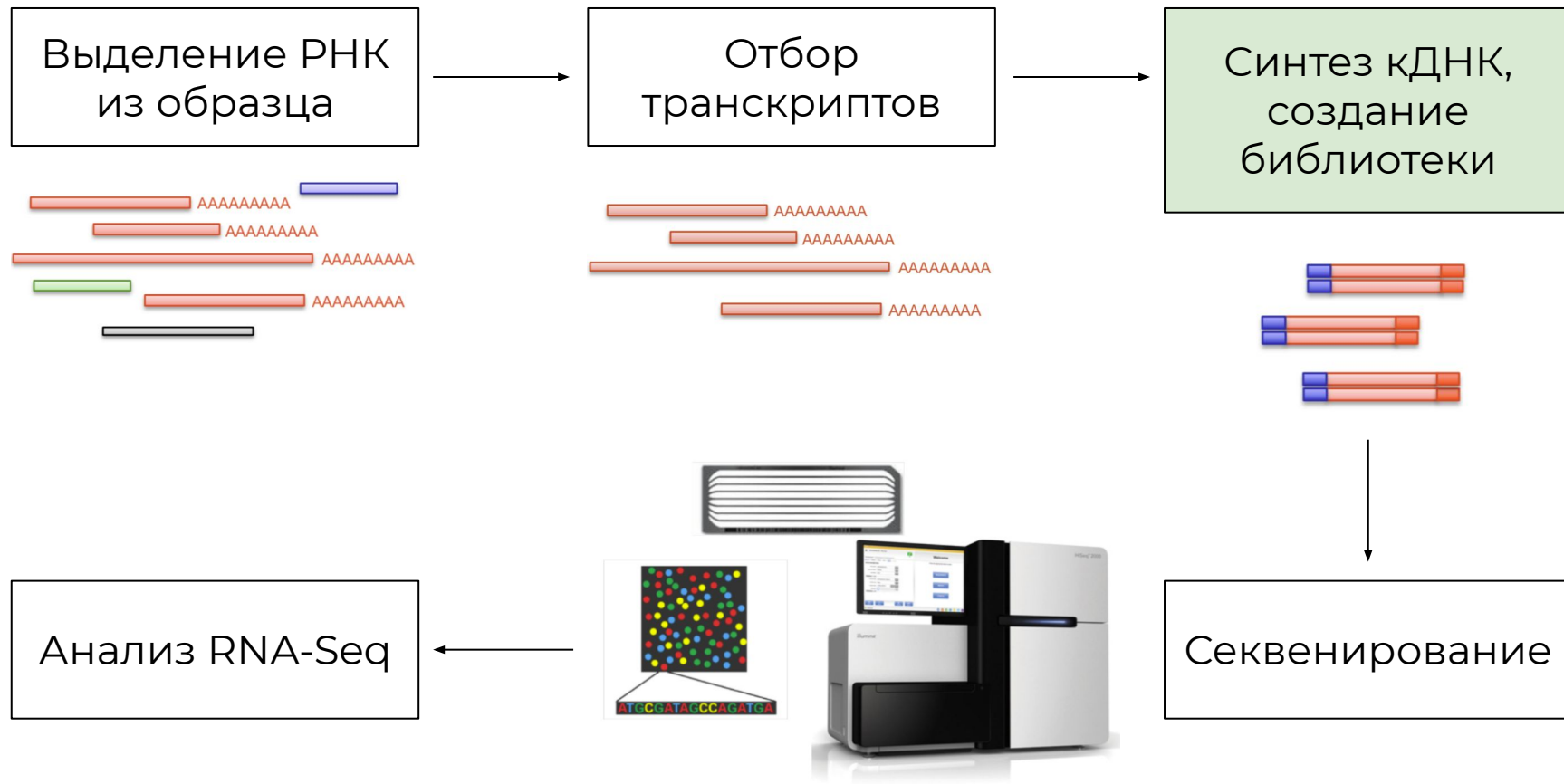
- В целом оценки экспрессий достаточно хорошо скоррелированы при разных способах подготовки библиотек
- Малые РНК лучше изучать с отбором целевых транскриптом методом Total



- Методы скоррелированы лучше внутри себя
- Однако перепутать кишку и кровь у нас не получится

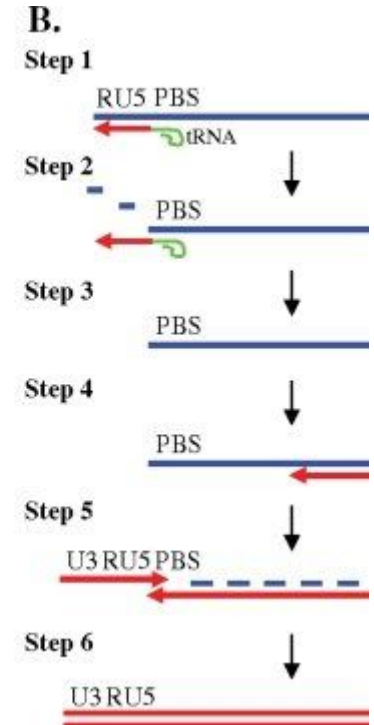
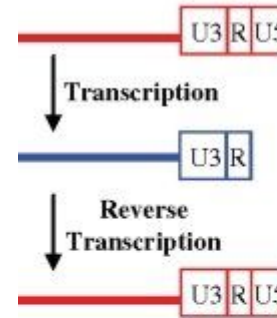
# Процедура RNA-Seq

BostonGene



# Синтез ДНК у ретровирусов

- В геномной РНК ретровирусов есть участки, комплементарные концевым участкам тРНК клетки-хозяина
- В итоге в качестве затравки выступает именно эта тРНК
- А как поступить нам?





- Случайные праймеры — это набор гексамеров из случайных нуклеотидов
- Так мы надеемся, что попадём на все транскрипты
- Oligo(dT)-праймеры дают смещённое покрытие



# Когда фрагментировать библиотеку?

Перед  
созданием кДНК

Oligo(dT) primers

Random primers

После создания  
кДНК

Standard



Anchored

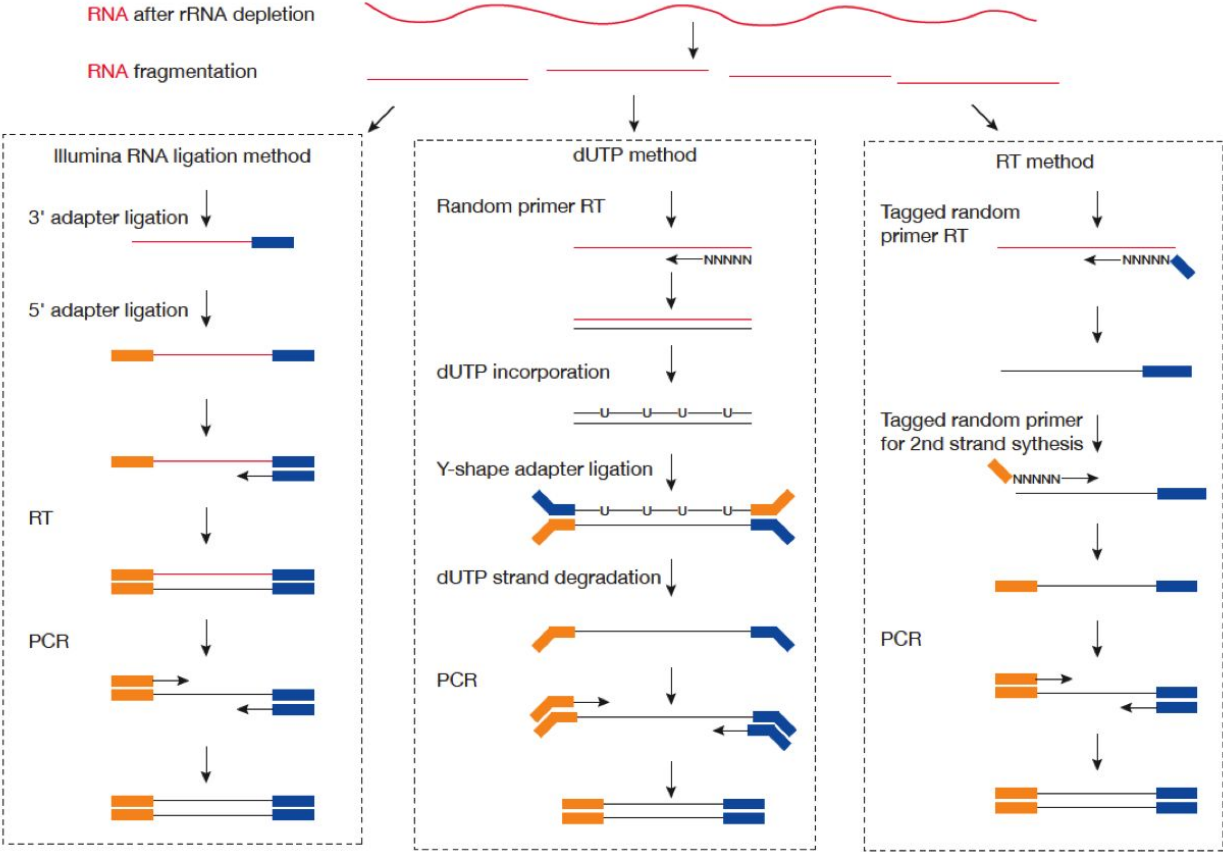


mRNA

cDNA

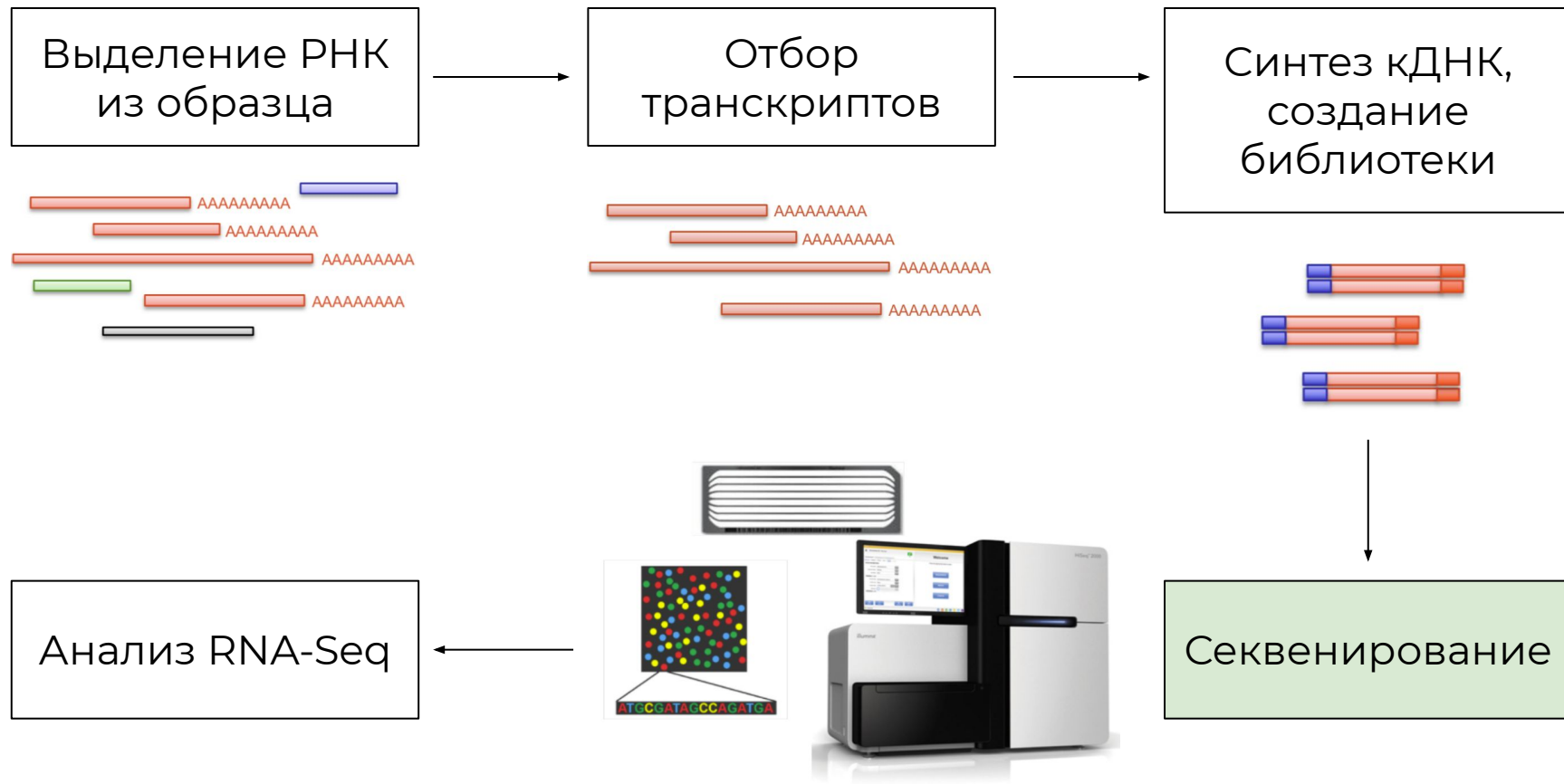
NVATGC Primer

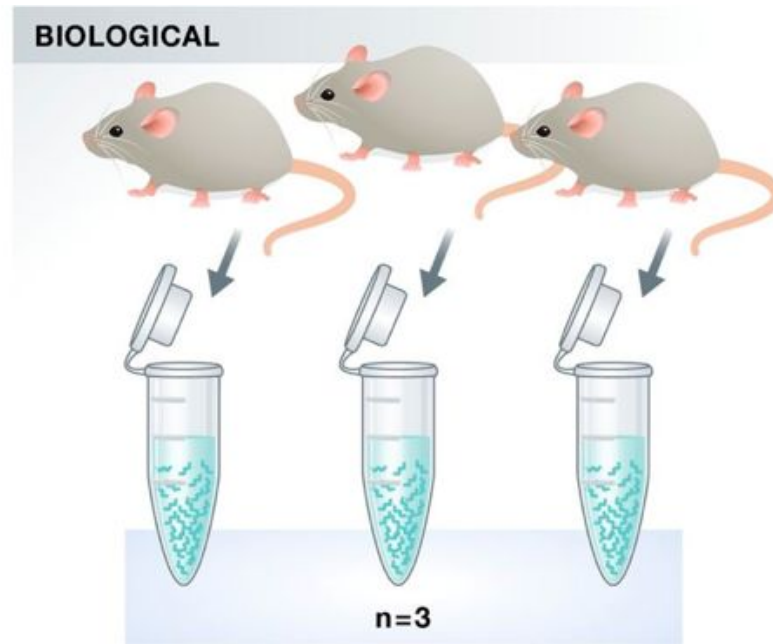
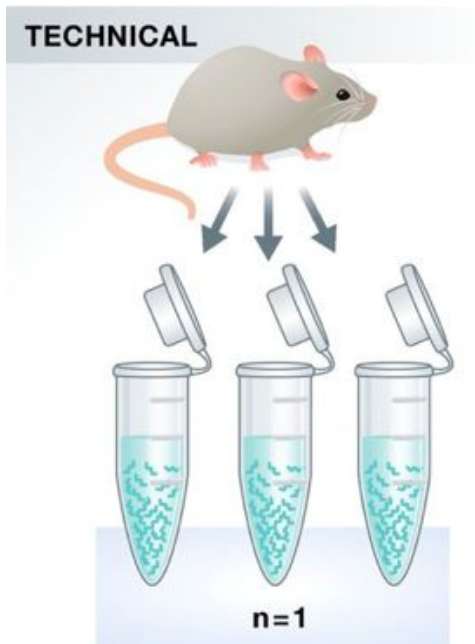
# Stranded RNA-Seq



# Процедура RNA-Seq

BostonGene





- Повторности эксперимента нужны для того, чтобы набрать статистические данные
- Из-за того, что цель эксперимента — это оценка экспрессии (одно секвенирование = одно число на один ген), нам необходимы выборки, которые дали бы нам какую-то выборку
- Существуют **технические** реплики (выделили из одной мышки, несколько раз отсеквенировали) = для определения технического шума
- Также есть **биологические** реплики (несколько мышек, несколько секвенирований) = для определения биологического “шума”

## Series GSE133521

[Query DataSets for GSE133521](#)

Status Public on Jul 31, 2019  
 Title RNA-seq analysis of A498 cell line treated with siSETD2 or si-NC  
 Organism [Homo sapiens](#)  
 Experiment type Expression profiling by high throughput sequencing  
 Summary the coordinated expression of SETD2-miRNAs-MAPK/JNK may be predictive of poor prognostic in patients with RCC. Our findings also emphasize the therapeutic potential of MAP4K4 in RCC therapy and support the development of an effective therapeutic strategy to target MAP4K4 by molecularly targeted approaches.

Overall design six samples. Including three siSETD2 samples and three control samples

Contributor(s) [Yan L, Liu H](#)  
 Citation missing *Has this study been published? Please [login](#) to update or [notify GEO](#).*  
 Submission date Jun 28, 2019  
 Last update date Jul 31, 2019  
 Contact name Libin Yan  
 E-mail(s) [chosenyan@126.com](mailto:chosenyan@126.com)  
 Phone 15527269390  
 Organization name Tongji Hospital, Tongji Medical College, Huazhong University of Science Technology  
 Department Urology  
 Street address No.1095, Jiefang Avenue  
 City Wuhan  
 State/province Hubei  
 ZIP/Postal code 430000  
 Country China

Platforms (1) [GPL11154](#) Illumina HiSeq 2000 (Homo sapiens)

Samples (6) [GSM3911248](#) siSETD2-1  
[More...](#) [GSM3911249](#) siSETD2-2  
[GSM3911250](#) siSETD2-3

## Relations

BioProject [PRJNA551678](#)  
 SRA [SRP212624](#)

[SRX6384114](#): GSM3911253: si-NC-3; Homo sapiens; RNA-Seq

1 ILLUMINA (Illumina HiSeq 2000) run: 23.5M spots, 6.7G bases, 2.4Gb downloads

Submitted by: NCBI (GEO)

Study: RNA-seq analysis of A498 cell line treated with siSETD2 or si-NC

[PRJNA551678](#) • [SRP212624](#) • [All experiments](#) • [All runs](#)  
[show Abstract](#)

Sample: si-NC-3

[SAMN12162889](#) • SRS5043632 • [All experiments](#) • [All runs](#)  
 Organism: [Homo sapiens](#)

## Library:

*Instrument:* Illumina HiSeq 2000  
*Strategy:* RNA-Seq  
*Source:* TRANSCRIPTOMIC  
*Selection:* cDNA  
*Layout:* PAIRED  
*Construction protocol:* none provided by submitter

## Experiment attributes:

*GEO Accession:* GSM3911253

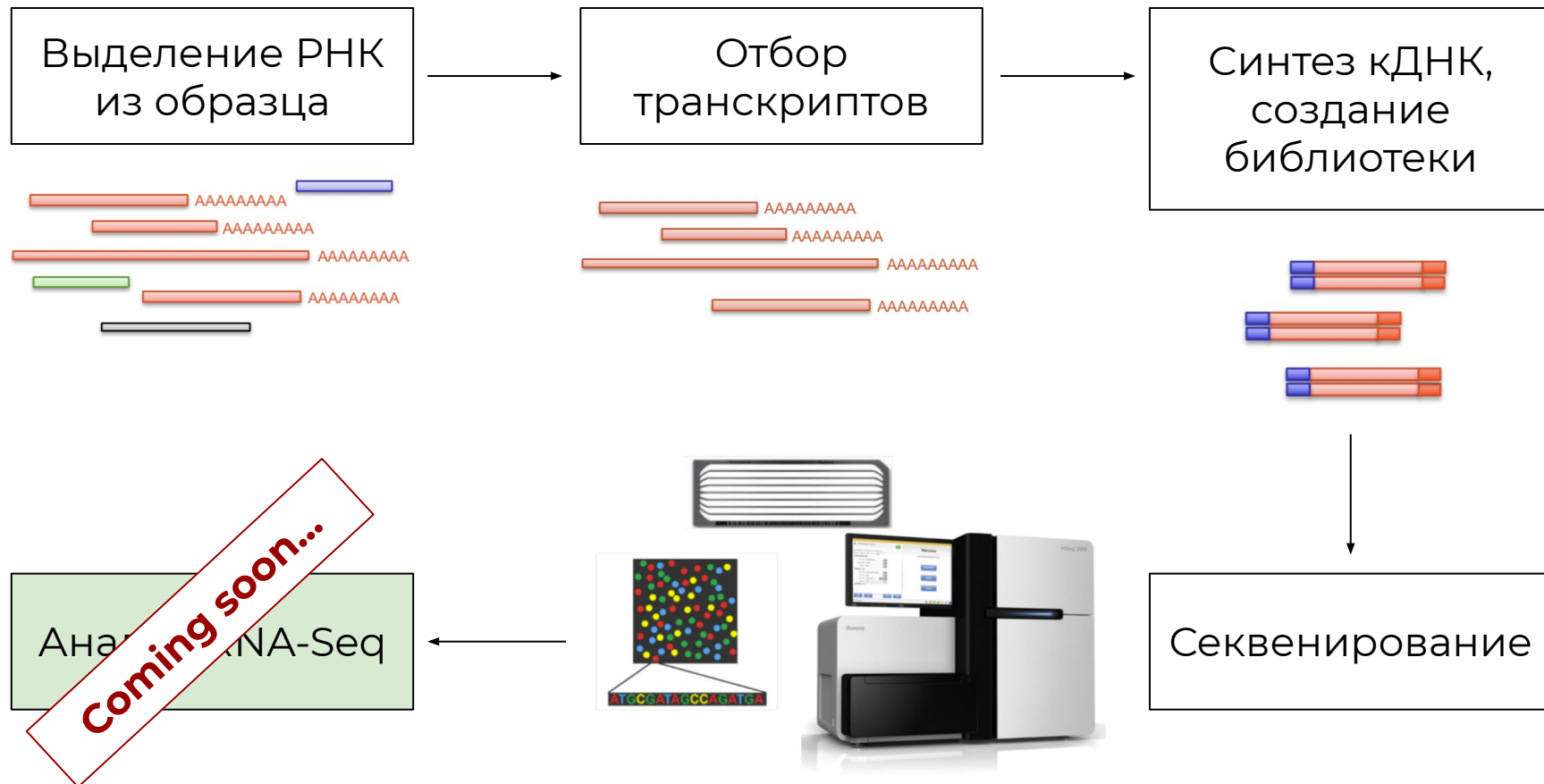
## Links:

Runs: 1 run, 23.5M spots, 6.7G bases, [2.4Gb](#)

Run	# of Spots	# of Bases	Size	Published
<a href="#">SRR9621174</a>	23,514,348	6.7G	2.4Gb	2019-07-31

# Процедура RNA-Seq

BostonGene



## Контакты



**Оля** — вопросы по биоинформатике & Feedback по курсу:



@olya\_kudryashova



olga.kudryashova@bostongene.com

**Серёжа** — вопросы насчёт блока по транскриптомике:



@sergisa



sergei.isaev@bostongene.com

**Телеграмм-группа курса** иммунологии и биоинформатики от BostonGene:



<https://t.me/joinchat/B1VA6B1Qe1zGiBeZuTC2vQ>

**Катя Титова** (HR) — вопросы о стажировке в BostonGene летом 2021:



[vk.com/titovakate](https://vk.com/titovakate)



ekaterina.titova@bostongene.com