

«Анализ данных NGS»

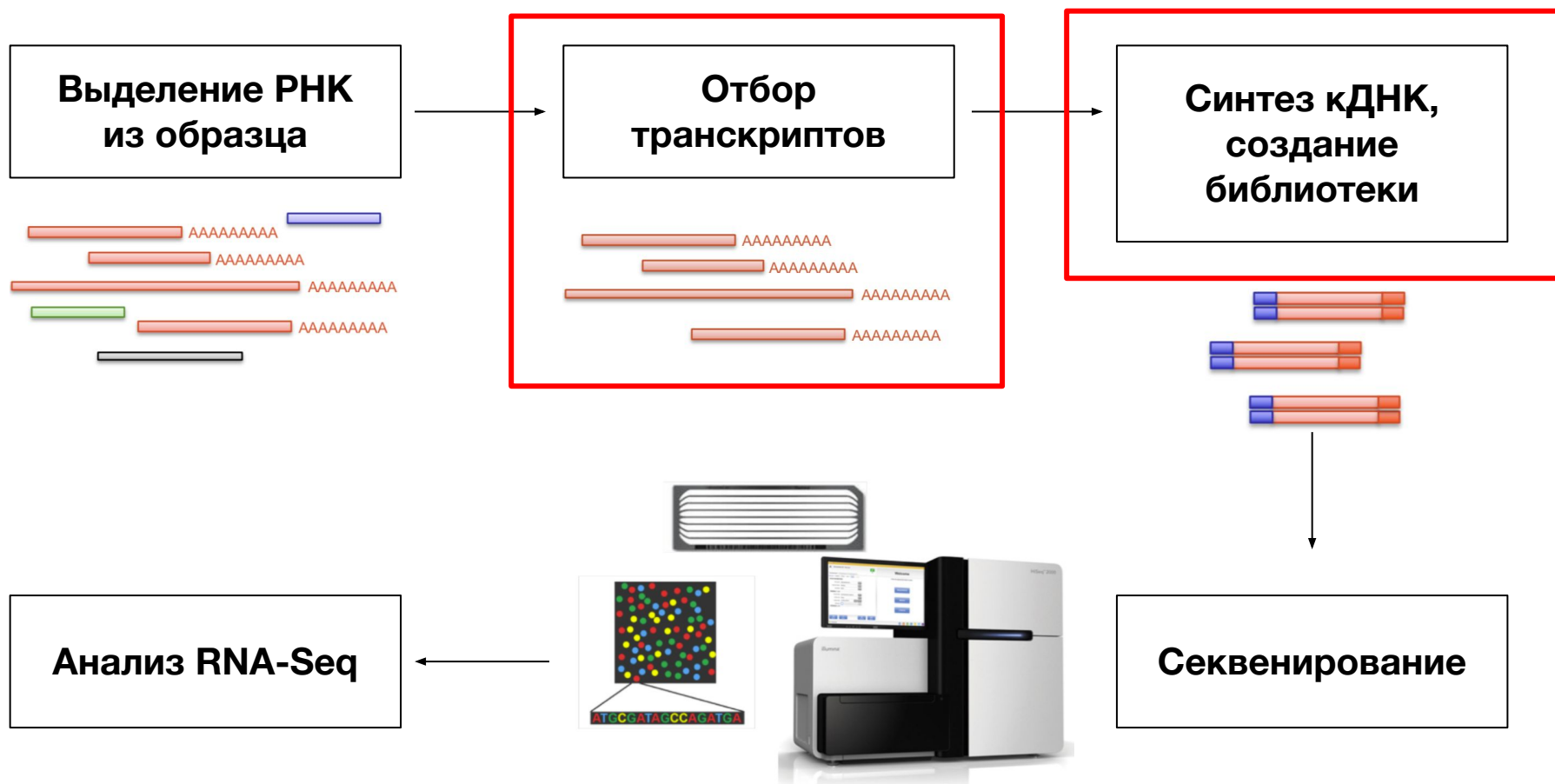
Лекция #7

Методы, связанные с RNA-Seq

Серёжа Исаев

аспирант **MedUni Vienna**

Какие можно внести изменения в протокол?



BS-RNA-Seq

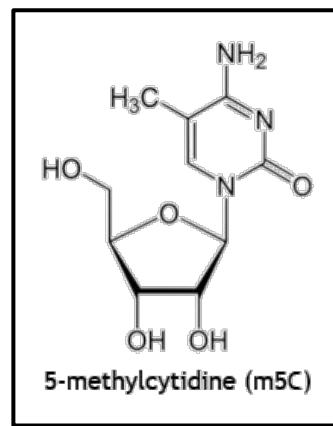
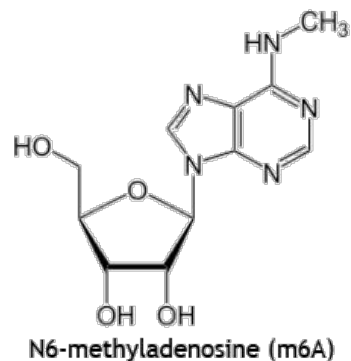
Метилирование ДНК

Важно различать **метилирование гистонов** и **метилирование нуклеиновых кислот**

Метилирование ДНК ассоциировано с так называемыми **CpG островами**, расположенных в основном рядом с промоторными участками генов. Метилирование ДНК в этих островах ассоциировано с репрессией экспрессии генов

Однако мы с вами говорим про метилирование РНК, которое выполняет иную функцию

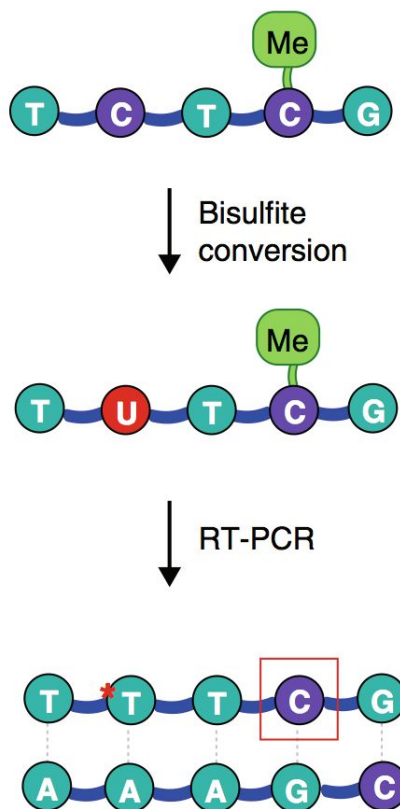
Метилирование РНК



В молекулах РНК встречается два основных типа метилирования азотистых оснований: m6A и m5C, мы фокусируемся сейчас на m5C-метилировании

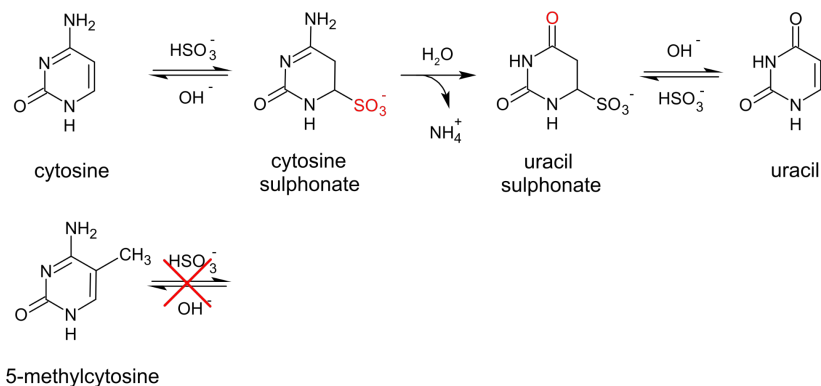
Оно встречается как в ncRNA (tRNA, mt-rRNA), так и в обычных mRNA. По всей видимости, m5C повышает конформационную стабильность молекул РНК

Бисульфитная конверсия

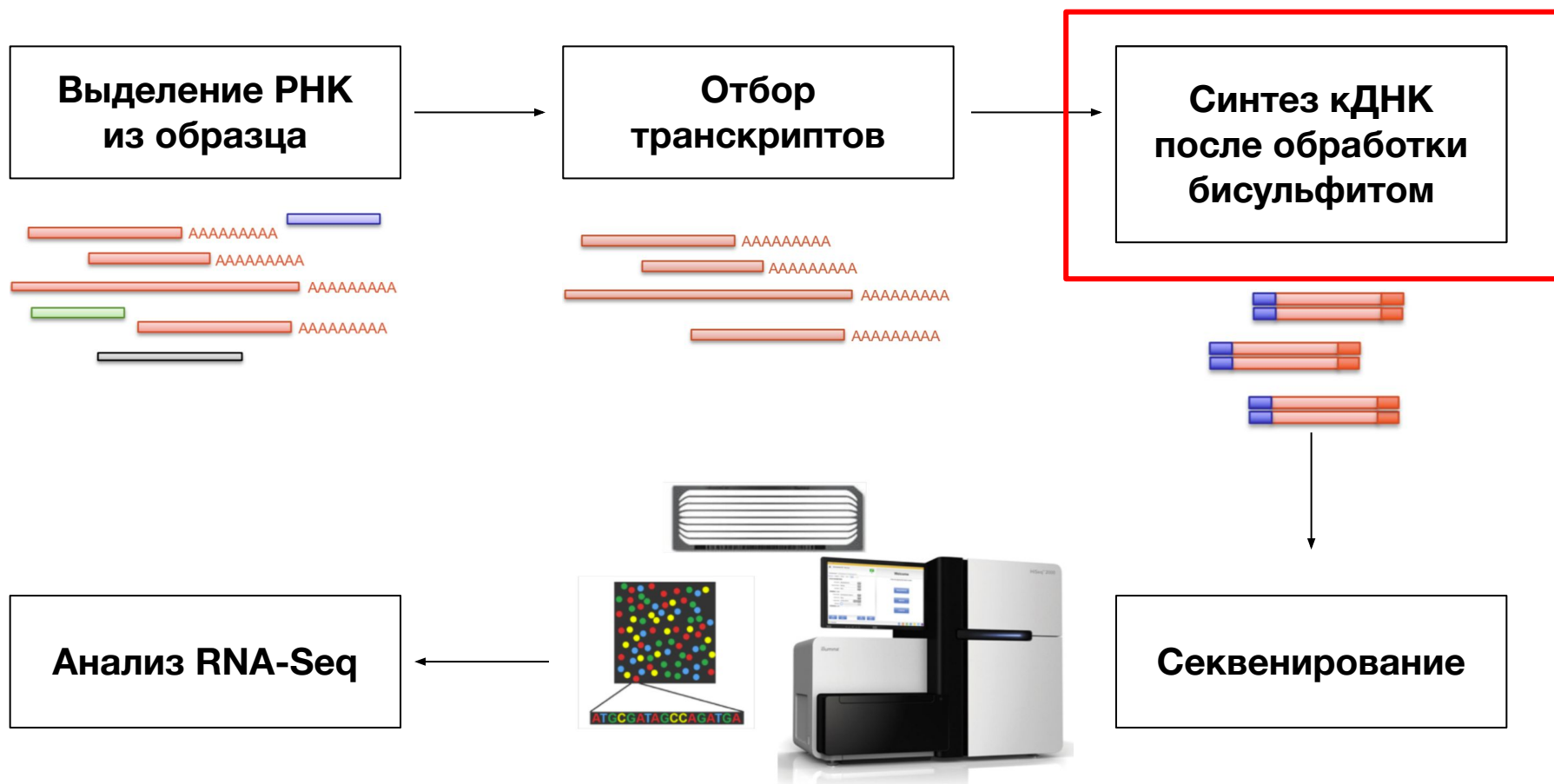


Library preparation/sequencing

В результате бисульфитной конверсии (обработки образца РНК при помощи HSO_3^-) цитозины химически превращаются в урацил, а метилированные цитозины — нет



Какие можно внести изменения в протокол?



Выравнивание при помощи Bismark

Выравнивать полученные прочтения достаточно проблематично по нескольким причинам:

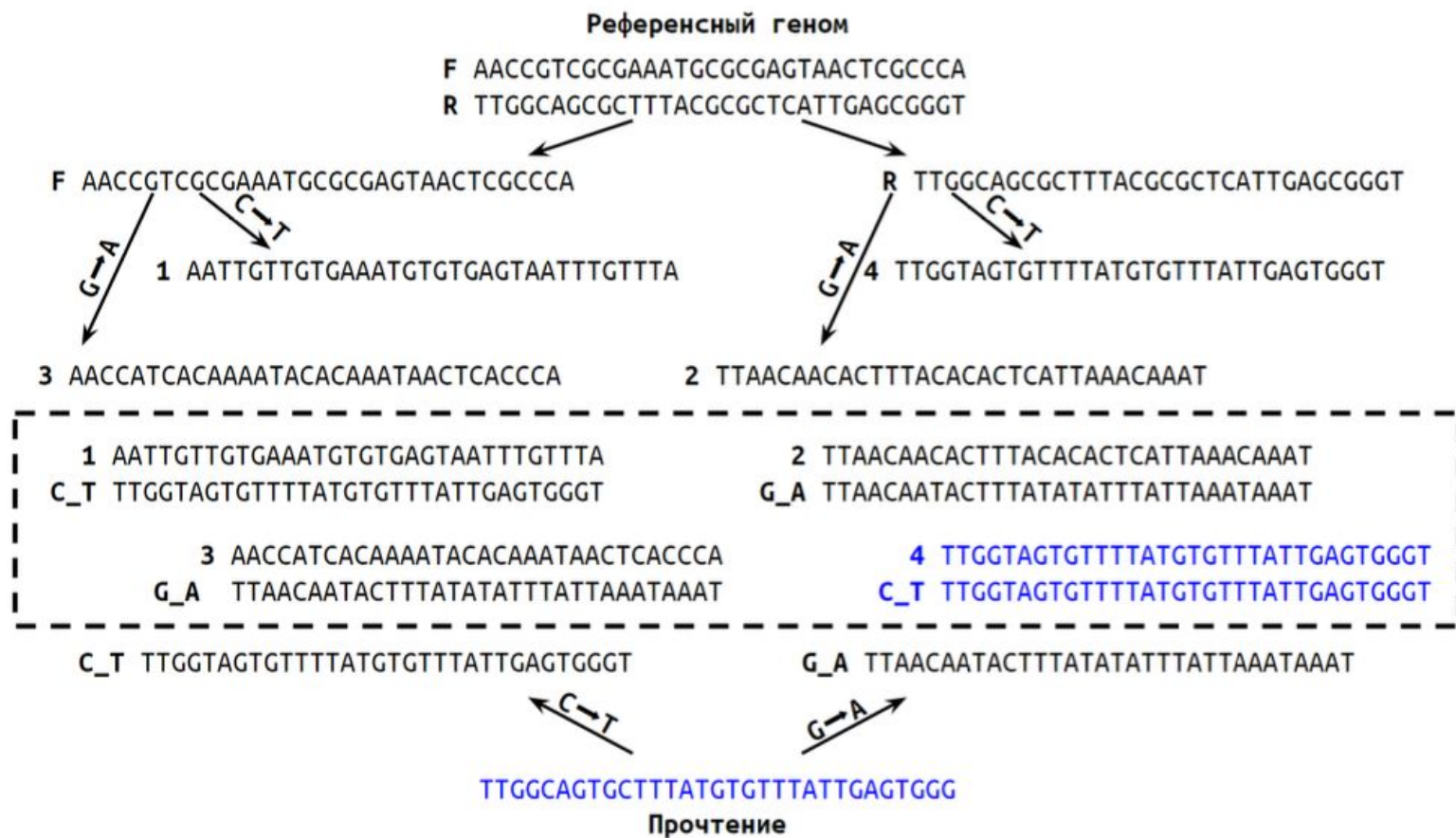
1. Обычно картировщики не устойчивы к настолько большому количеству мисмэтчей между ридом и референсом,

⇒ можно подготовить референс, в котором заведомо будут проведены необходимые замены (большая часть цитозинов всё-таки не метилирована)

2. Обычно картировщики не очень устойчивы даже к тому количеству мисмэтчей, которое возникнет в результате метилирования цитозинов

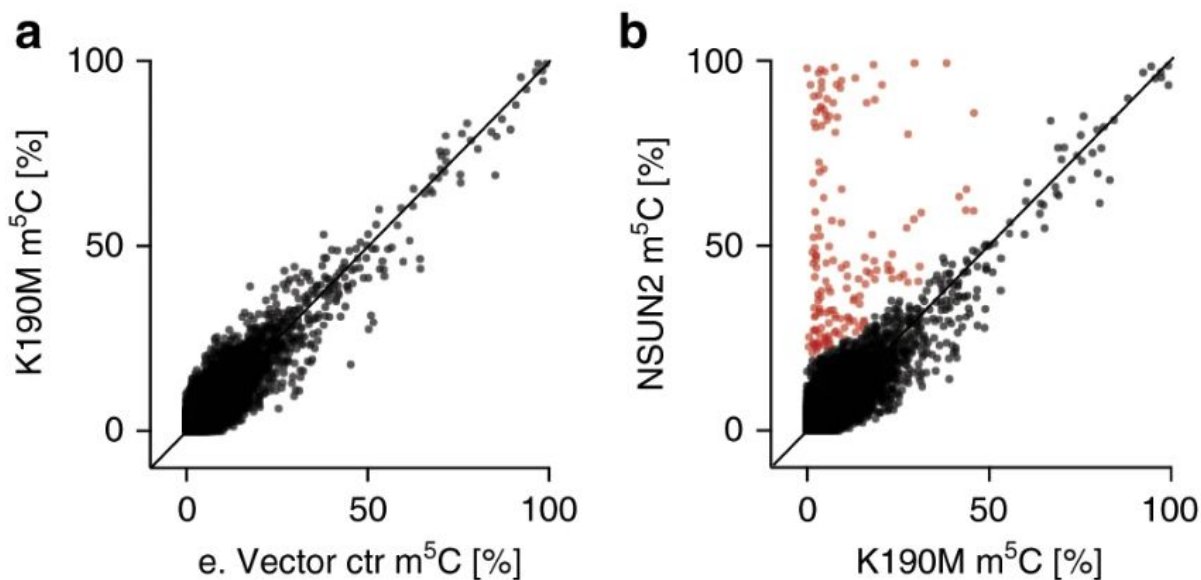
⇒ можно конвертировать в том числе и риды и запомнить, где были неконвертированные позиции

Выравнивание при помощи Bismark



Дифференциальное метилирование

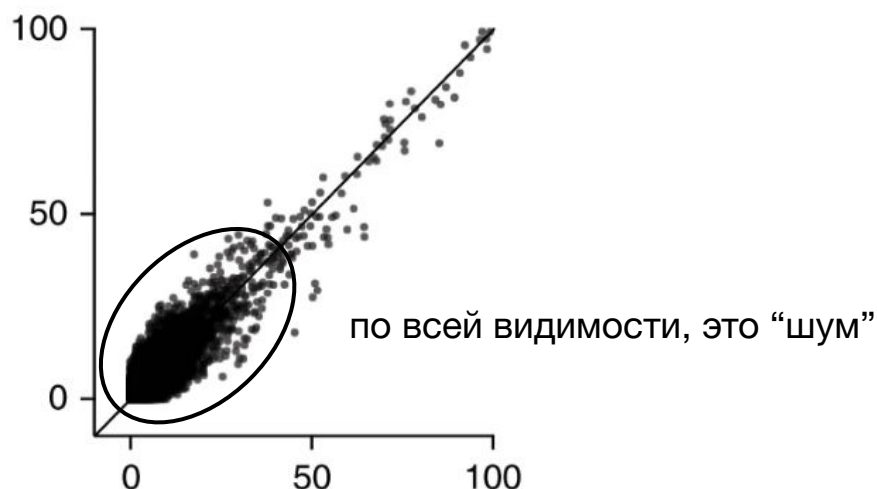
Для определения дифференциального метилирования используется несколько эвристик, однако полезно посмотреть просто “глазами” на то, как распределены пропорции конвертированных позиций в зависимости от условия



Поиск метилирования

Иногда не представляется возможным сделать эксперимент со сравнением метилирования в разных условиях и хочется просто определить сайты, которые достоверно метилированы

Для этого можно использовать пакет BisRNA, который оценивает вероятность недоконверсии по Пуассону, и после этого оценивает значимость отличий от этого технического распределения

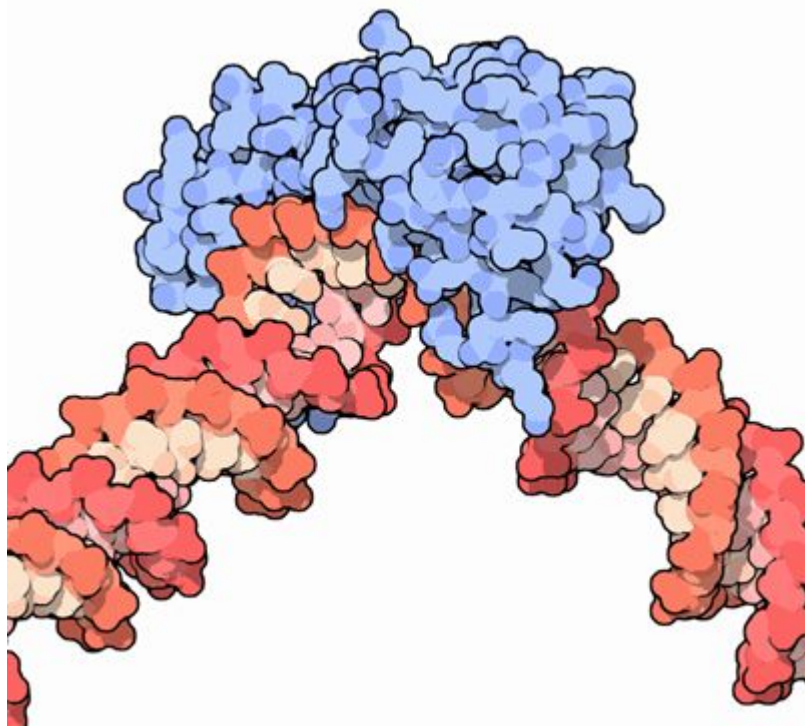


Проблемы поиска метилирования РНК

1. Может быть так, что *in vivo* метилирование происходит не во всех клетках, поэтому необходимо таргетно выбирать клетки интереса
2. Количество молекул РНК для разных генов разнится очень сильно, поэтому для обнаружения метилирования в молекулах с низкой экспрессией необходима очень высокая глубина секвенирования (больше, чем для классического RNA-Seq)

CLIP-Seq

ДНК-связывающие белки

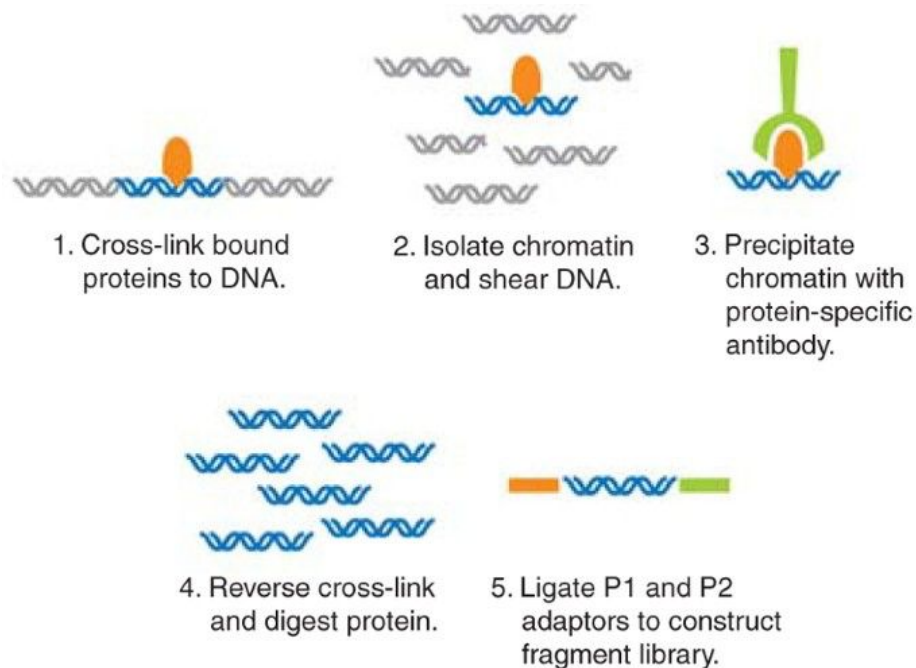


Обнаружение и анализ сайтов связывания белков с ДНК — это важнейшая задача для определения регуляции экспрессии генов

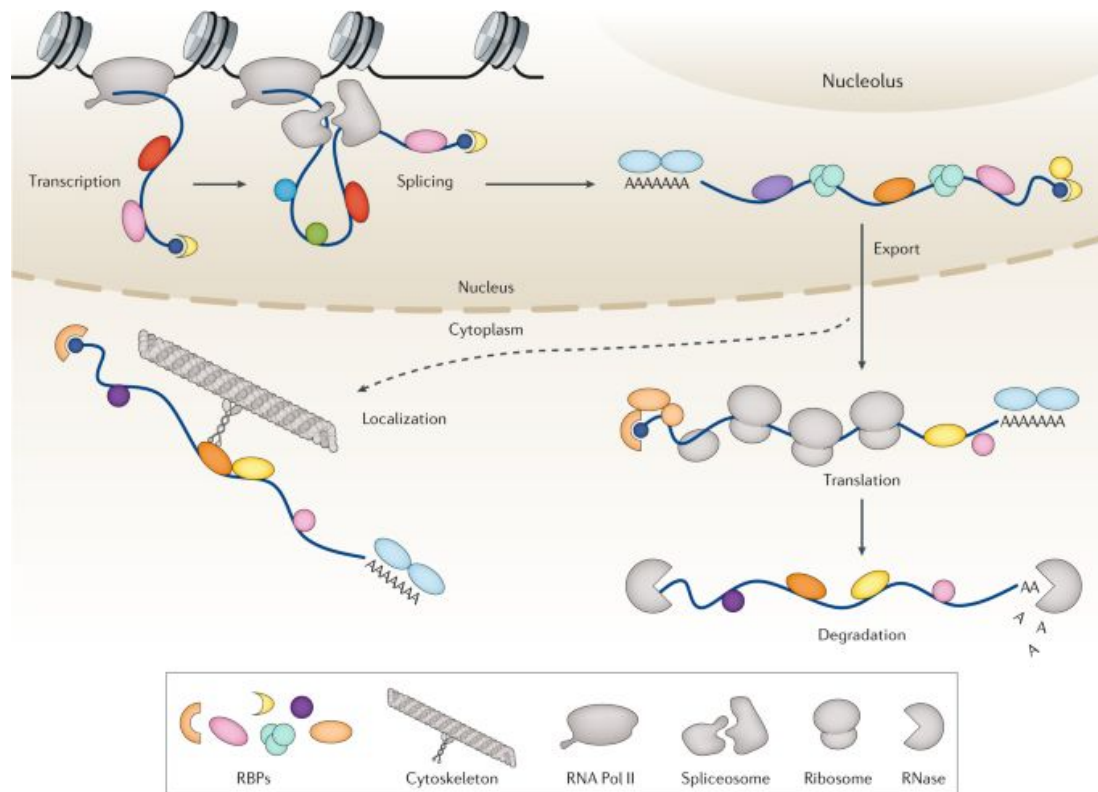
Речь может идти про транскрипционные факторы, гистоны, архитектурные белки и проч.

ChIP-Seq

Исторически первым возник метод, который позволяет определять сайты связывания какого-то конкретного белка и **ДНК** — ChIP-Seq

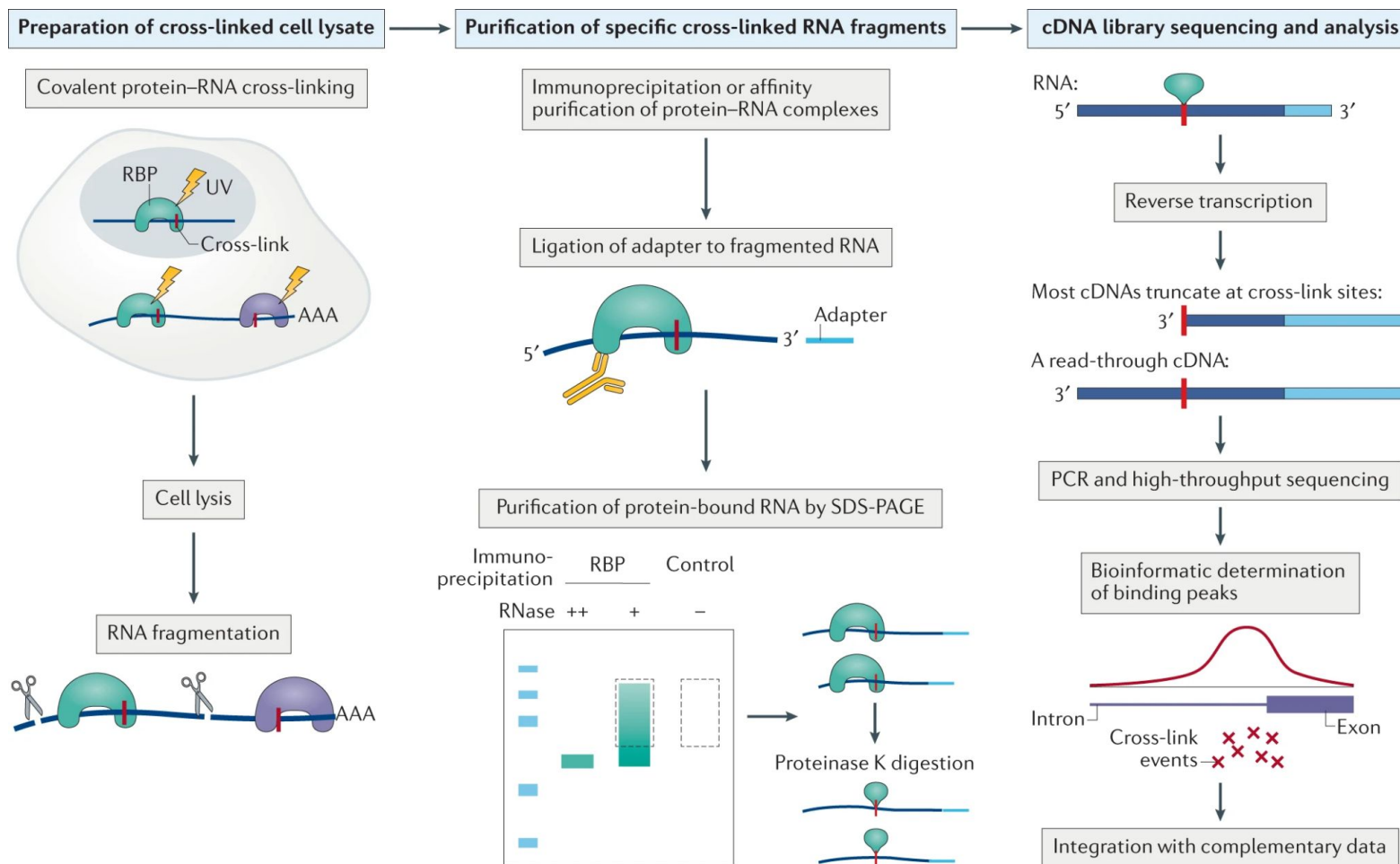


РНК-связывающие белки (RBP)

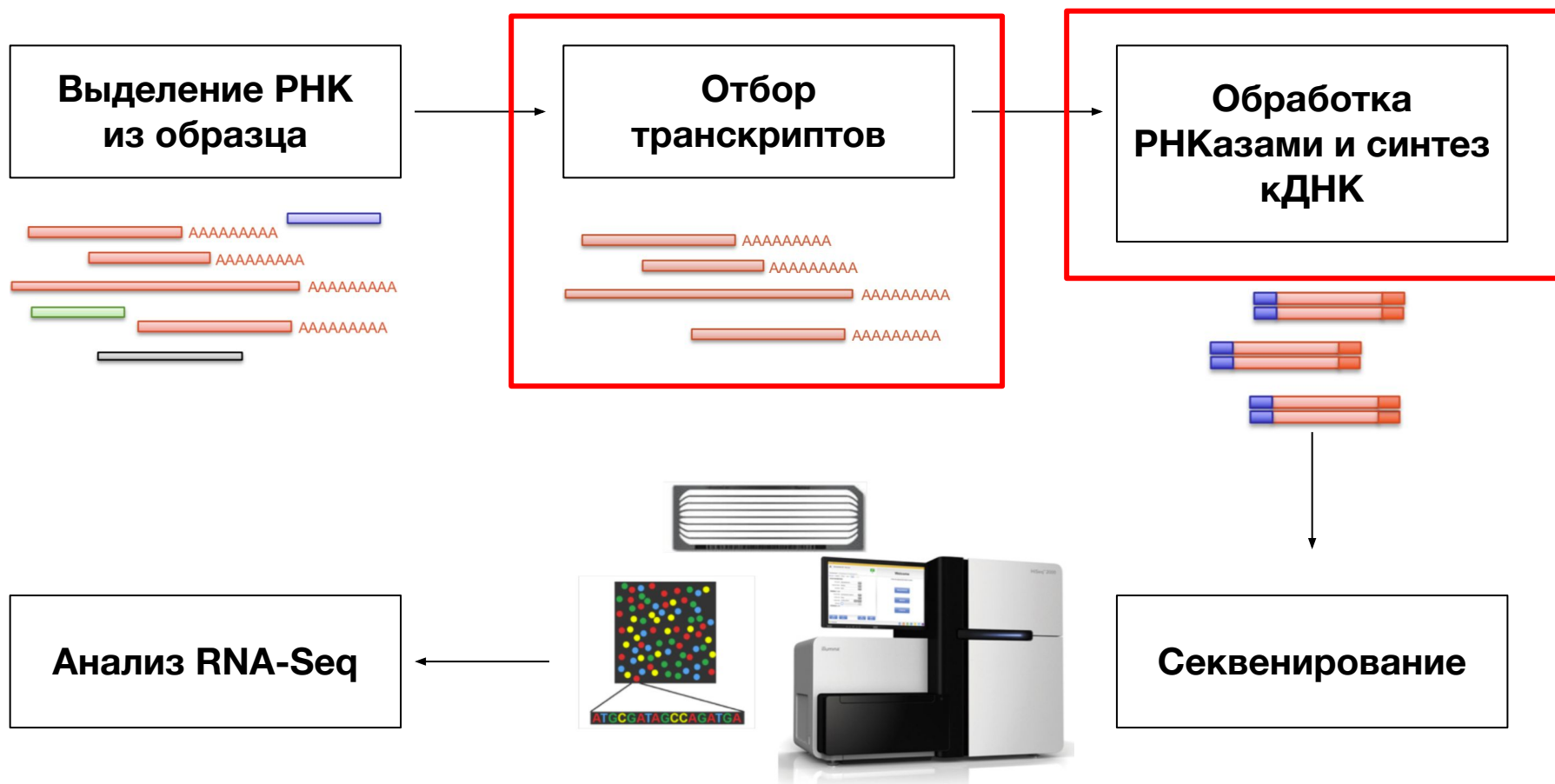


RBP также выполняют множество регуляторных функций

Общая схема CLIP-Seq



Какие можно внести изменения в протокол?



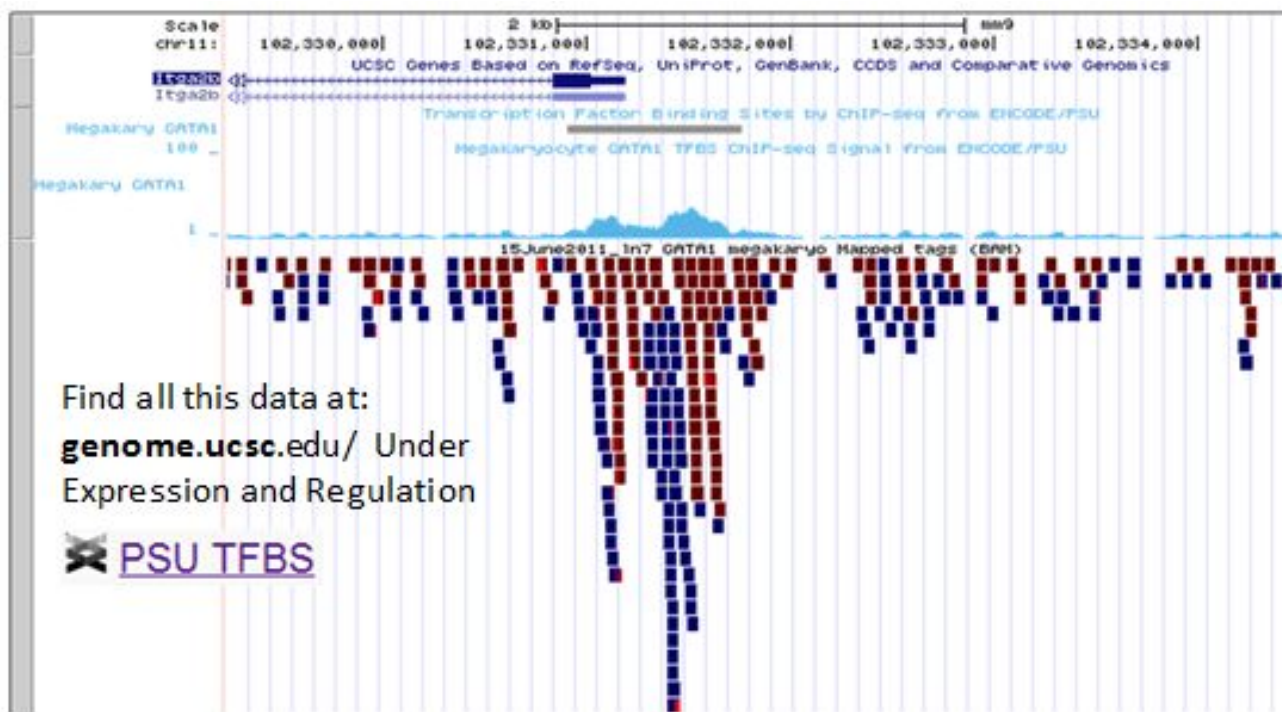
Анализ CLIP-Seq

После процедуры выравнивания присутствуют две важные стадии:

1. коллинг пиков (мест в геноме, на которых находятся скопления прочтений) и
2. определение дифференциально доступных пиков между контролем и таргетным экспериментом

Peak calling

Коллинг пиков — это важная стадия, для которой существует несколько моделей и алгоритмов



Peak calling

1. **MACS2** — алгоритм, классически используемый для определения пиков в ChIP-Seq и ATAC-Seq, использует модель Пуассона распределения каунтов при верной нулевой модели
2. **Piranha** — аналог, специально созданный для CLIP-Seq, использует NB-распределение при задании нулевой модели

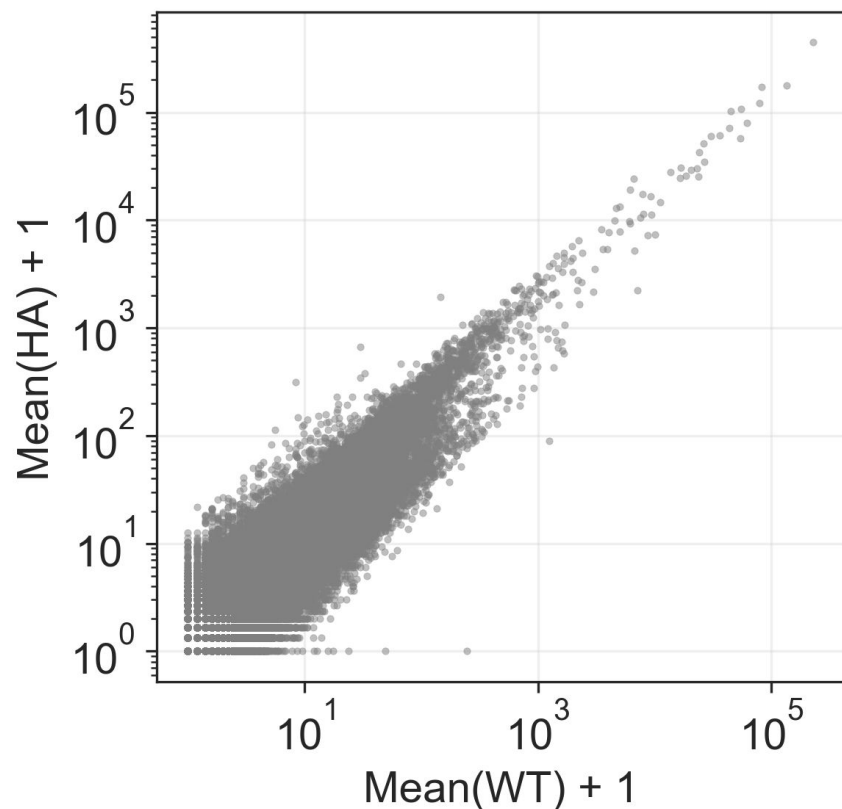
Как коллить пики при нескольких повторностях?

1. Коллить пики на индивидуальных репликах — проблемы с границами пика
2. Коллить пики на всех репликах совместно — проблемы с наличием / отсутствием пика на повторностях

Как делаю обычно я:

1. Коллим пики на всех образцах совместно
2. Выделяем границы пиков и преобразуем в .gtf-файл
3. При помощи featureCounts подсчитываем число каунтов для каждого пика в каждом из образцов

Зачем нужны контроли?



Чаще всего большинство пиков, которые вы увидите, будут обусловлены контаминацией РНК, поэтому необходимо использовать контроли для определения дифференциально доступных пиков

Какую информацию можно извлечь потом?

Если белок связывается селективно:

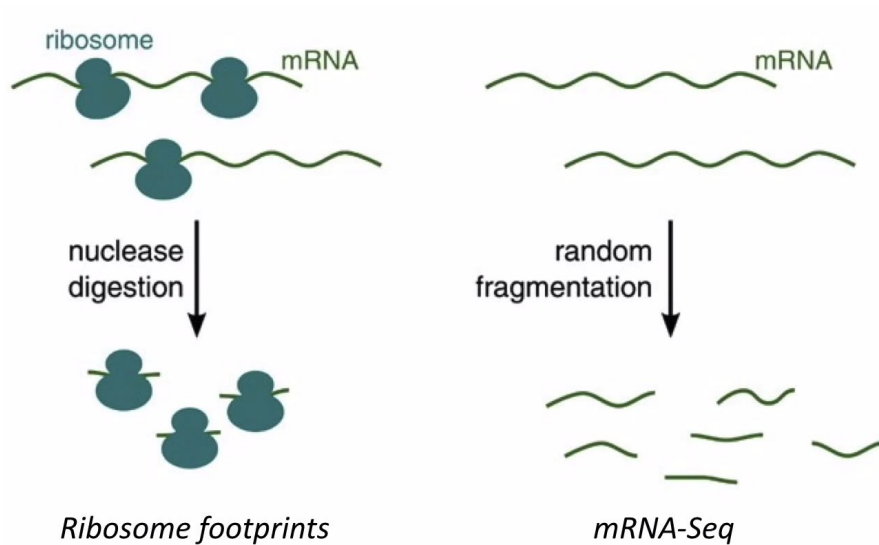
1. С какими именно молекулами РНК связывается белок и в каком месте?

Если белок связывается неселективно:

1. В каких участках транскрипта (экзоны, 3' UTR, 5' UTR, интроны) связывается белок?
2. Какой сайт связывания у изучаемого белка (*MEME*, *ChIP-Munk*)?
3. Функциональная значимость транскриптов, с которыми связывается изучаемый белок (тРНК, рРНК, функциональные группы мРНК)

Ribo-seq

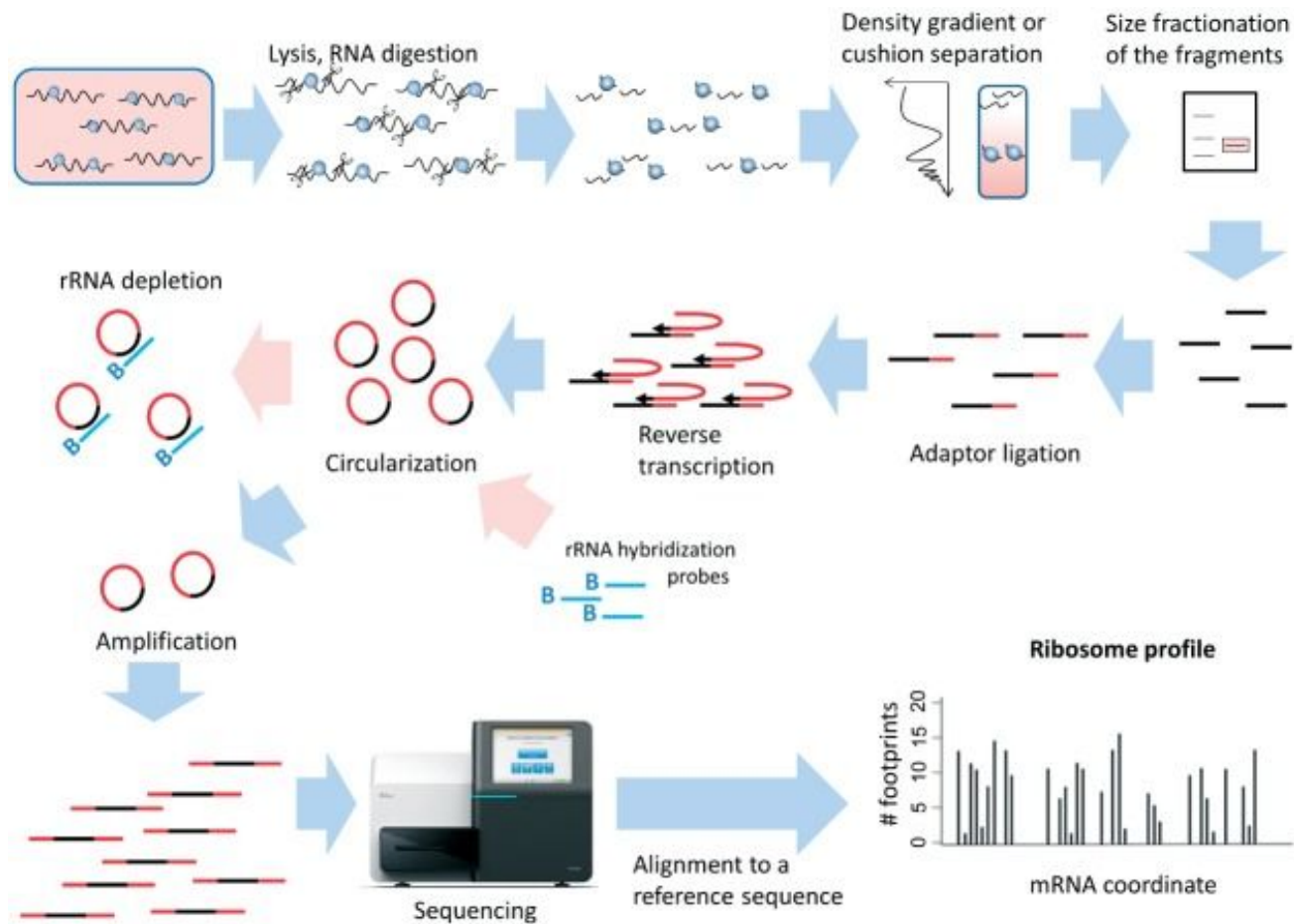
“CLIP-Seq” на рибосому



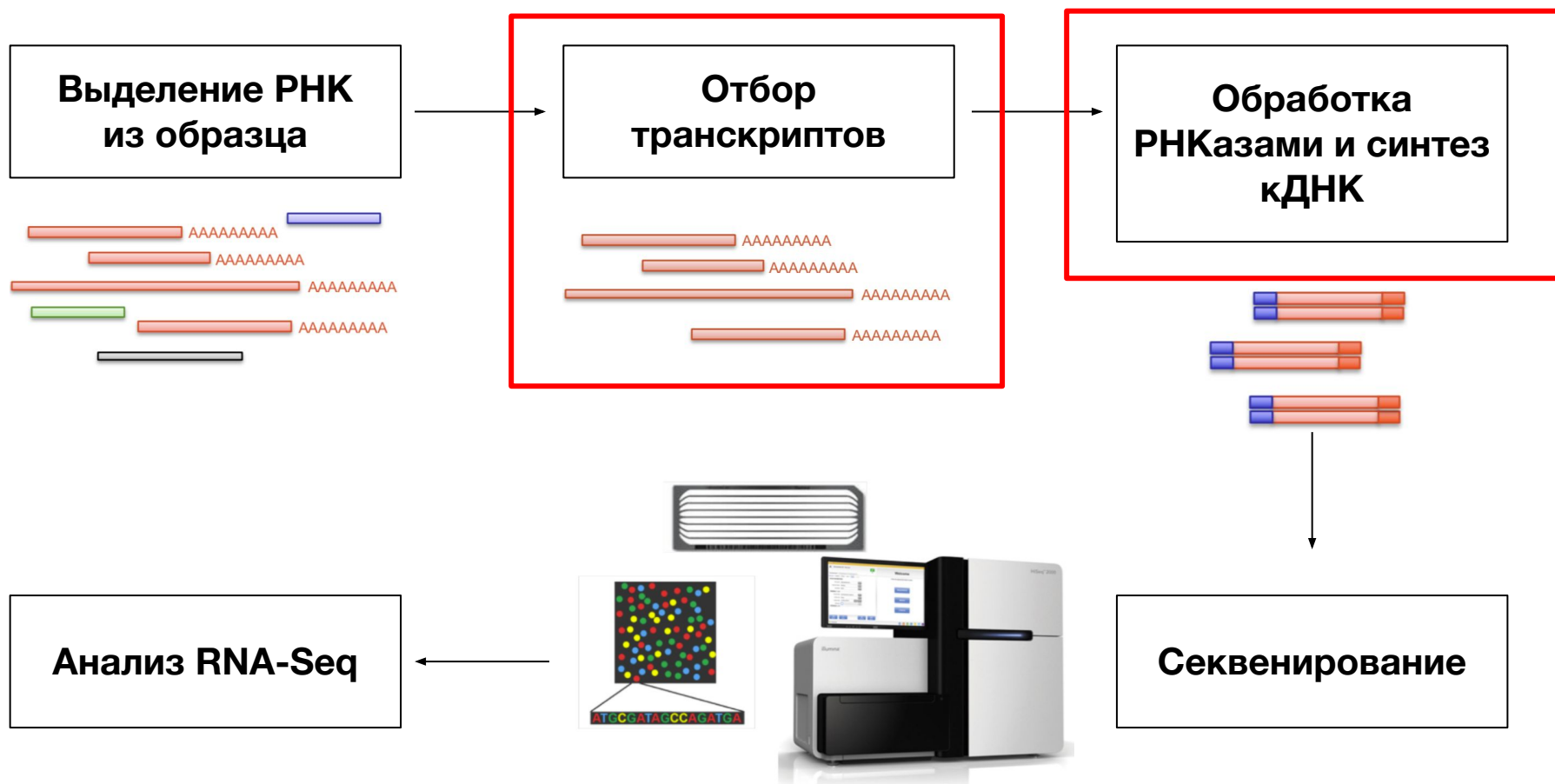
Ingolia NT et al., Science (2009)

Что будет, если мы попытаемся мысленно экстраполировать CLIP-Seq не на белок, но на рибосому?

Протокол Ribo-seq

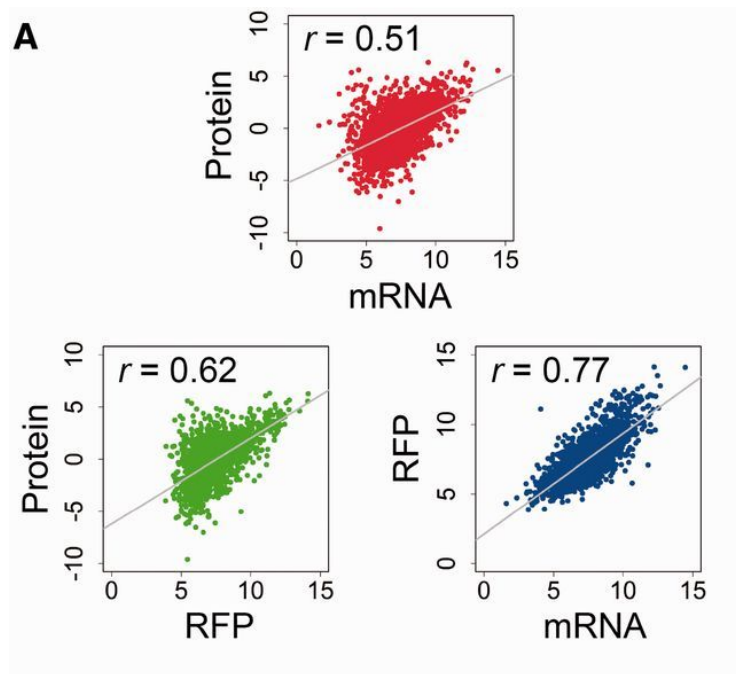


Какие можно внести изменения в протокол?

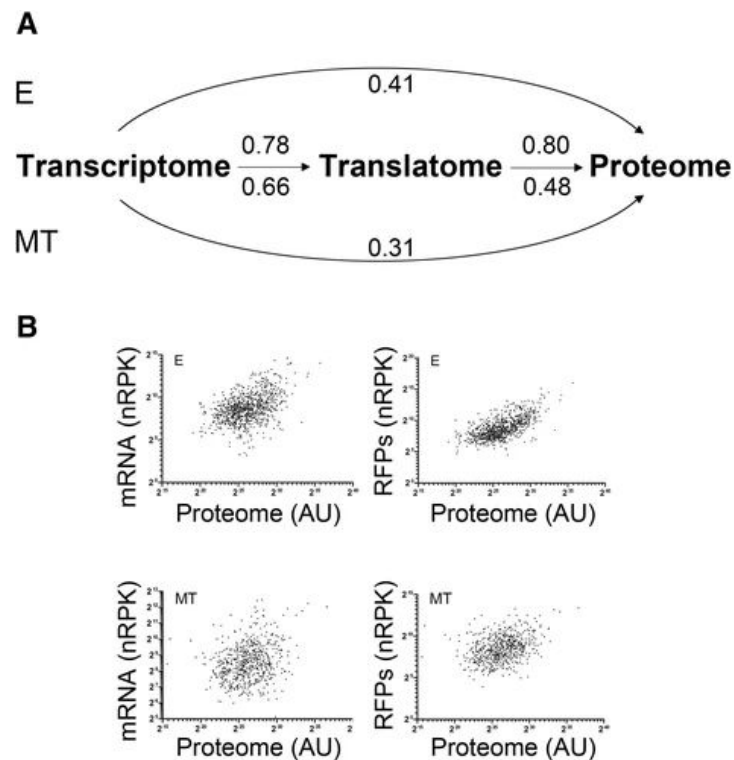


Зачем нам нужен Ribo-seq?

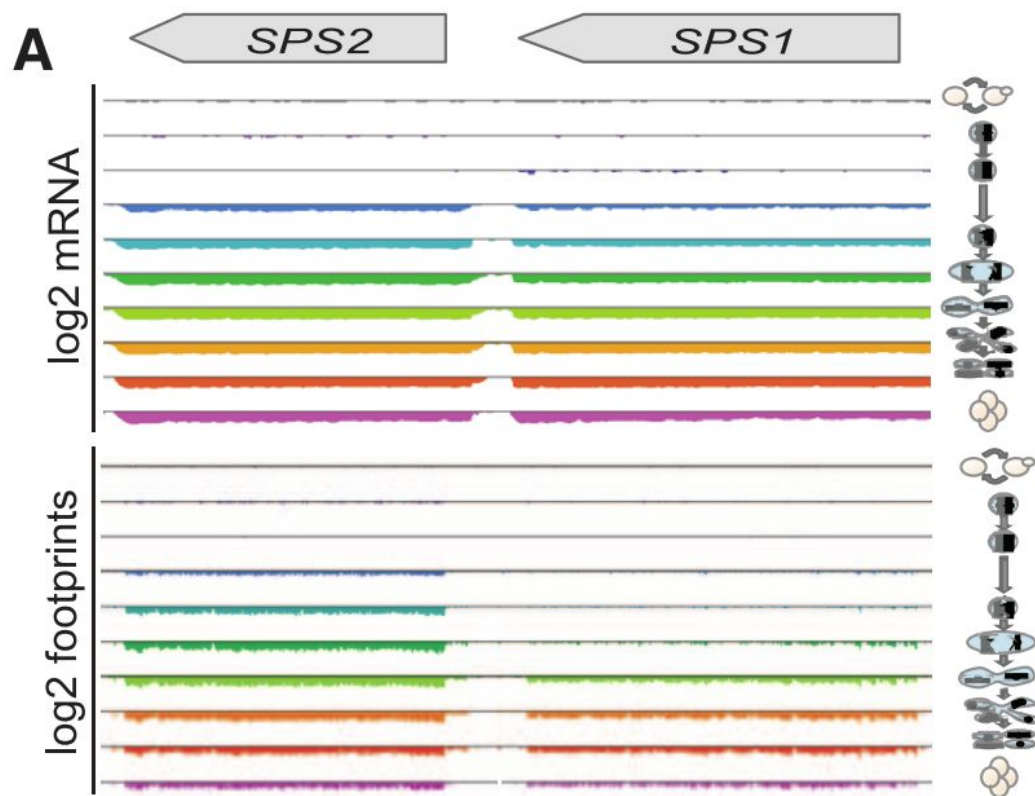
Транслятом лучше коррелирует с протеомом, чем транскриптом



Wang et al., 2015

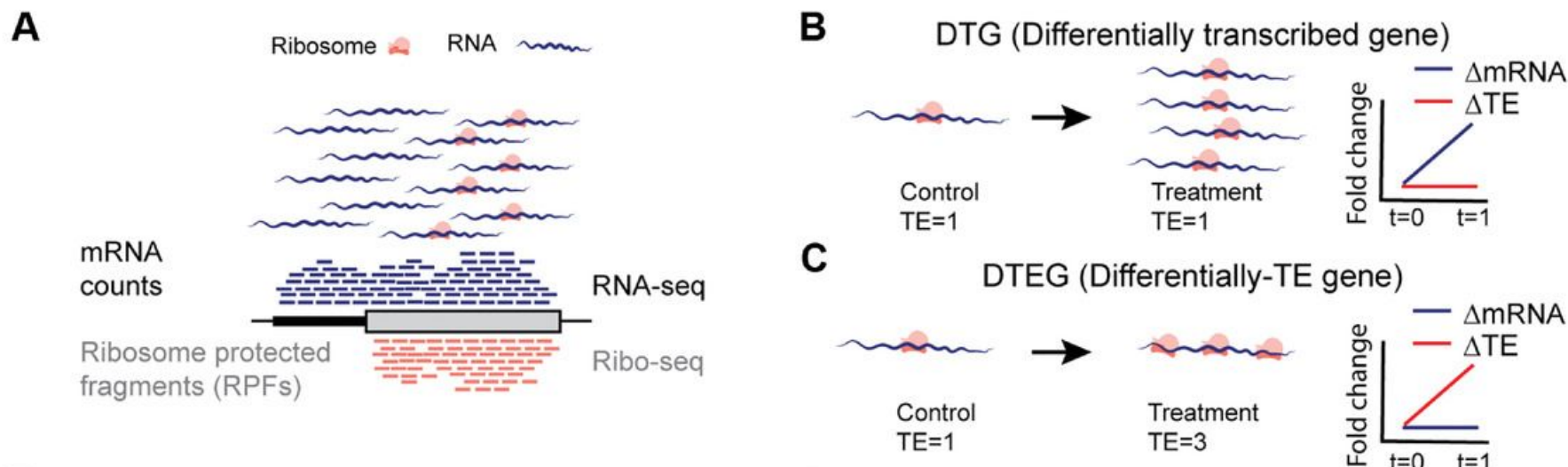


Дифференциальная эффективность трансляции



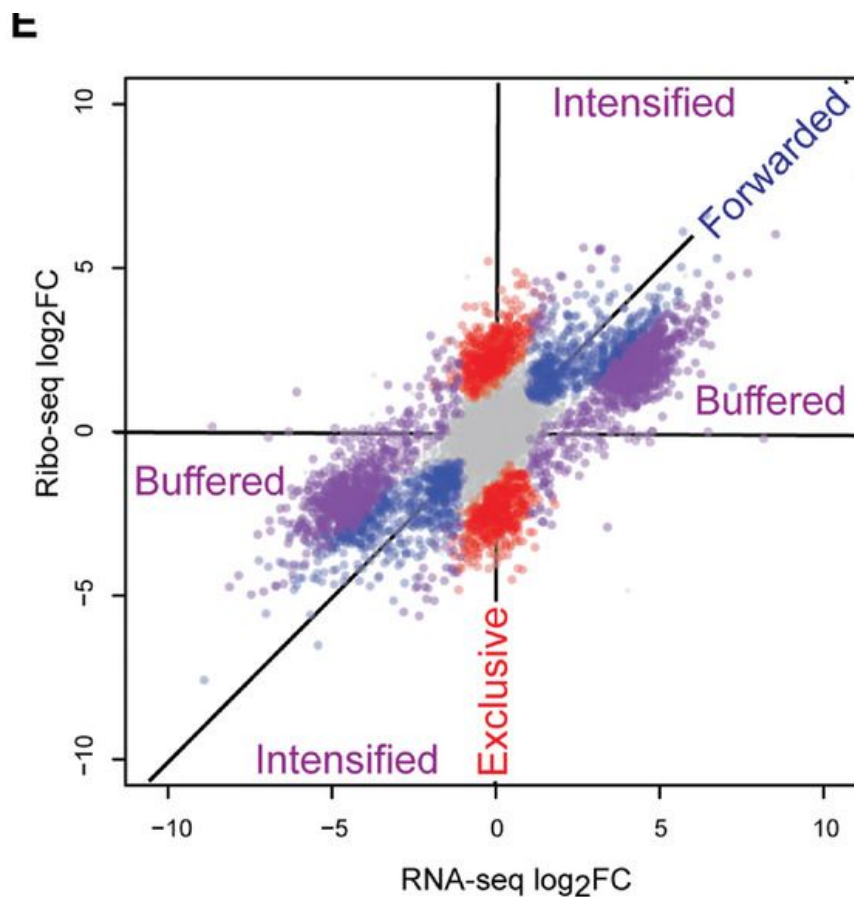
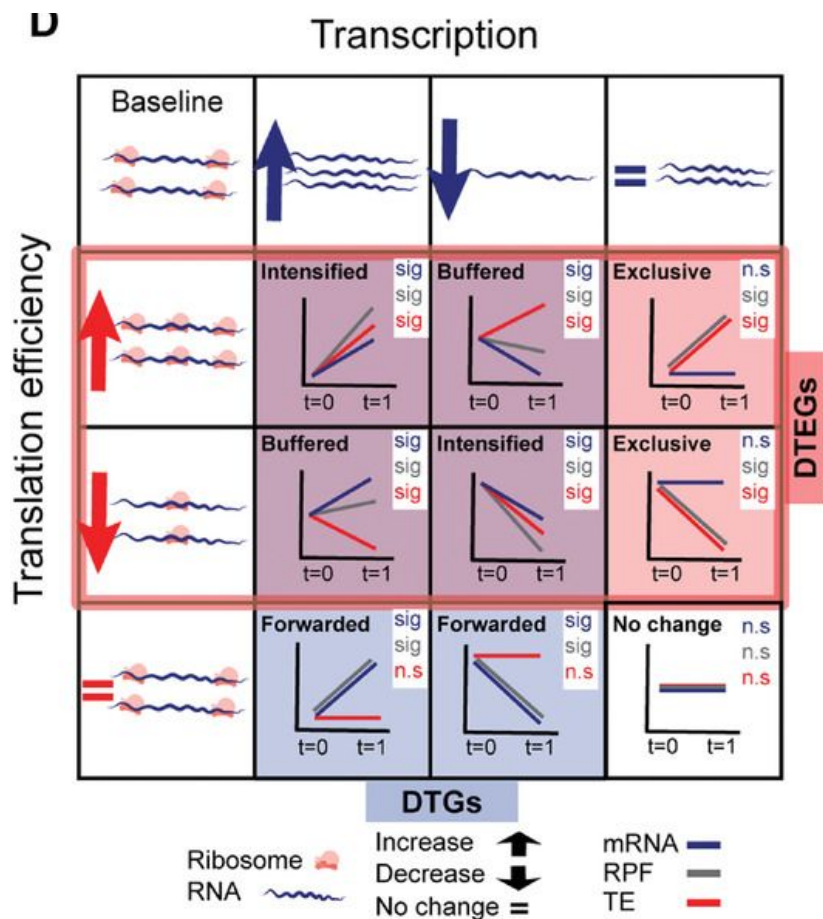
Экспрессия гена может регулироваться на стадии после транскрипции

Дифференциальная эффективность трансляции

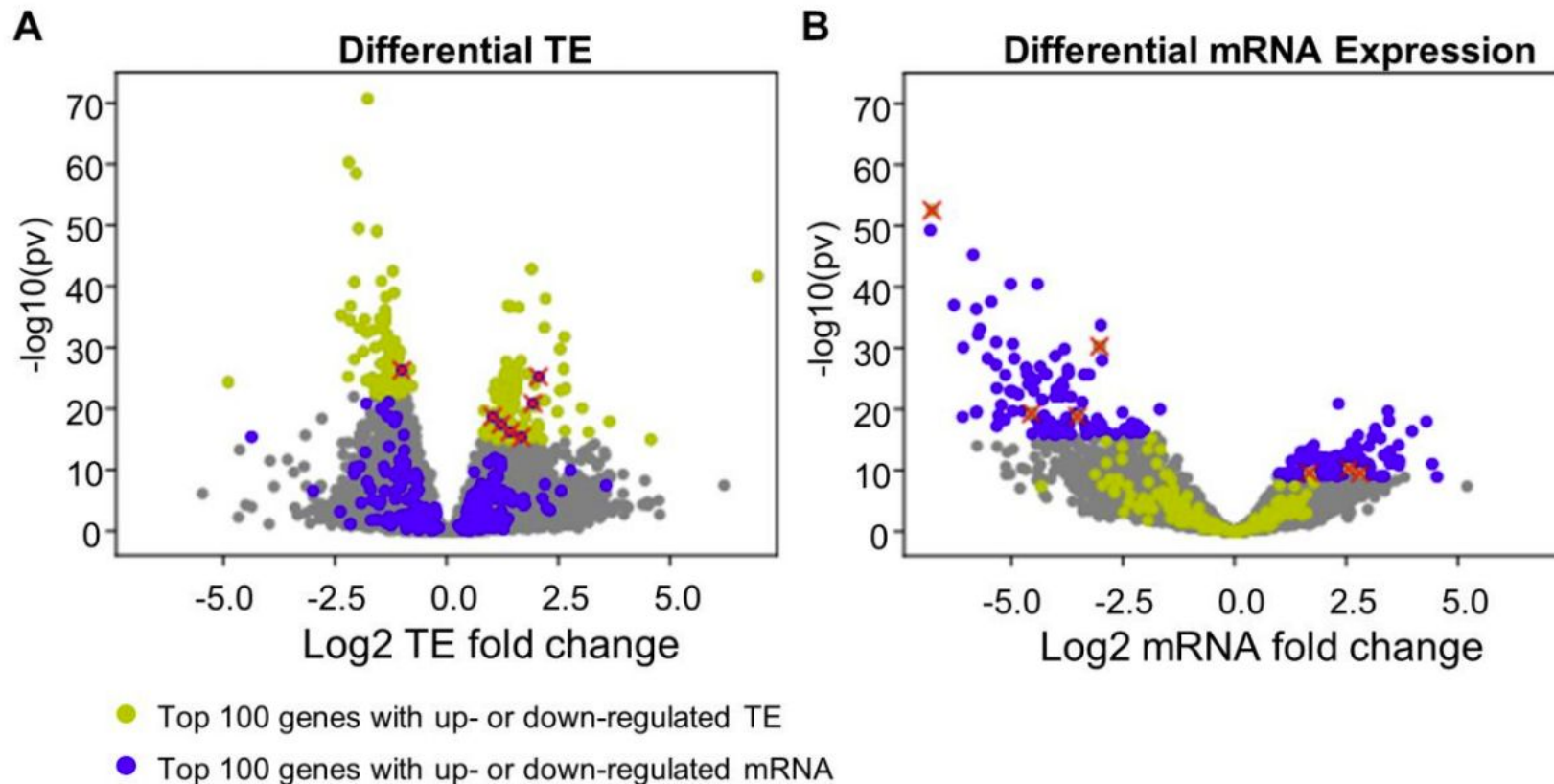


Эффективность трансляции — это отношение каунтов футпринтов к числу каунтов мРНК

Дифференциальная эффективность трансляции

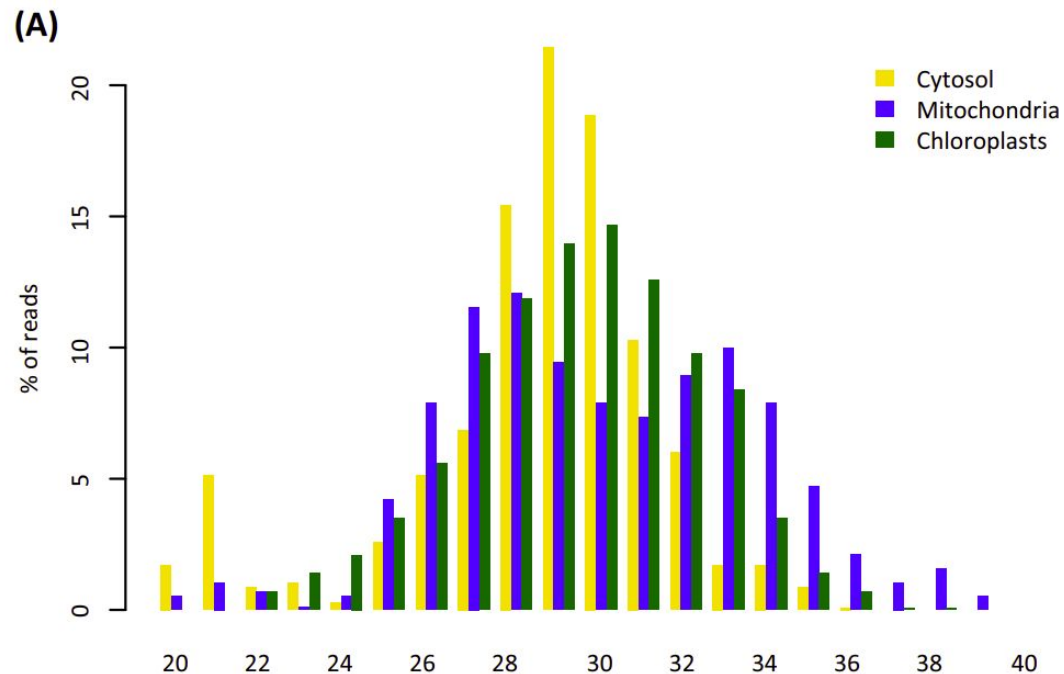


Дифференциальная эффективность трансляции

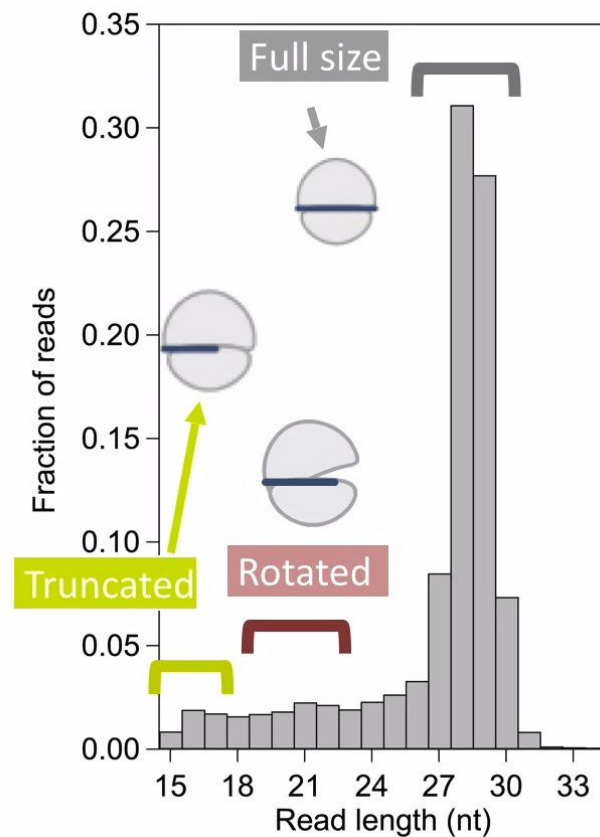
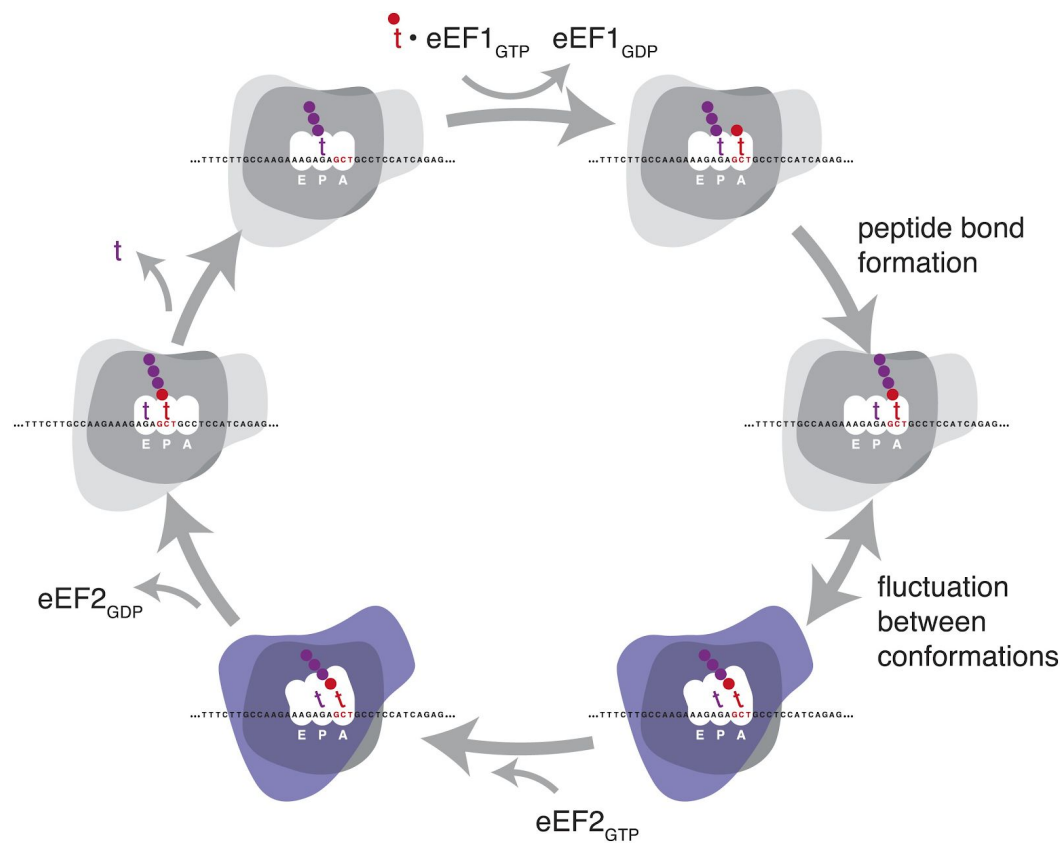


Что ещё мы можем вычлениить из этих данных?

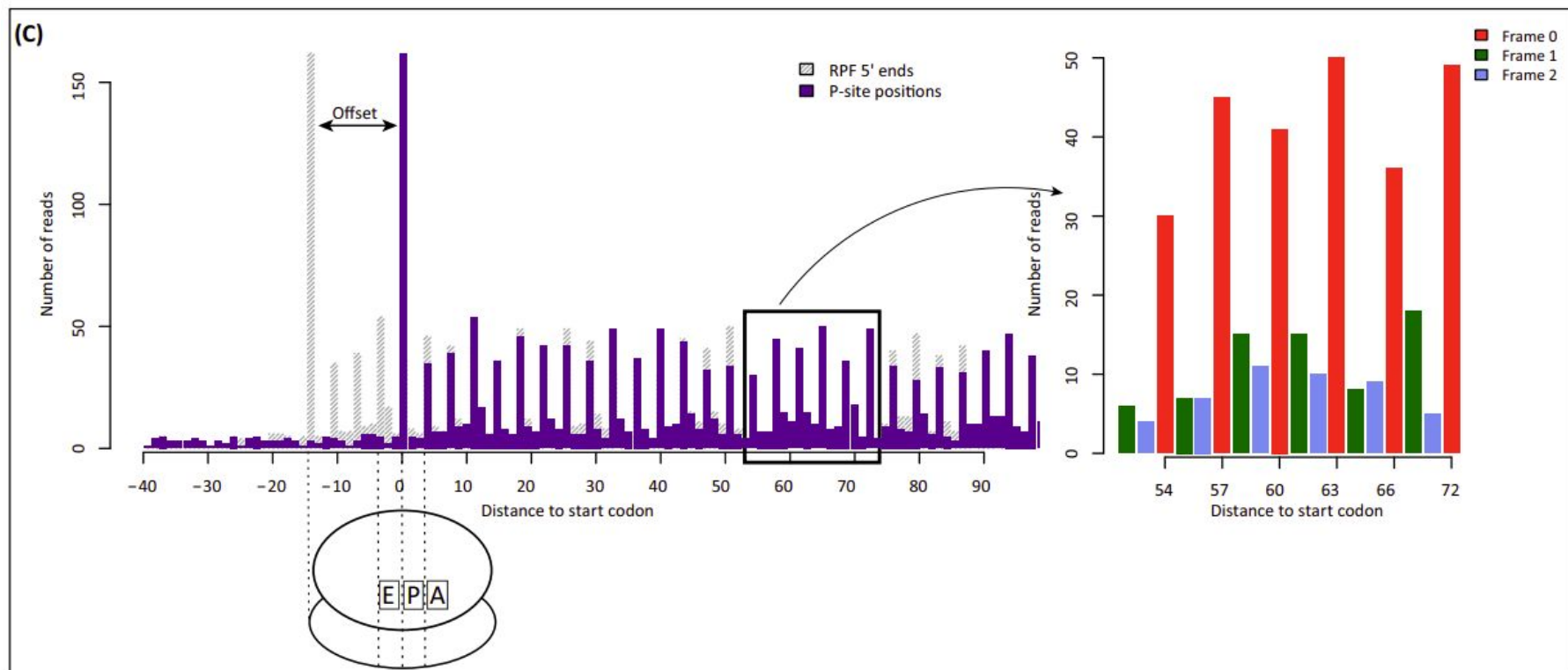
От чего зависит длина получившегося рида?



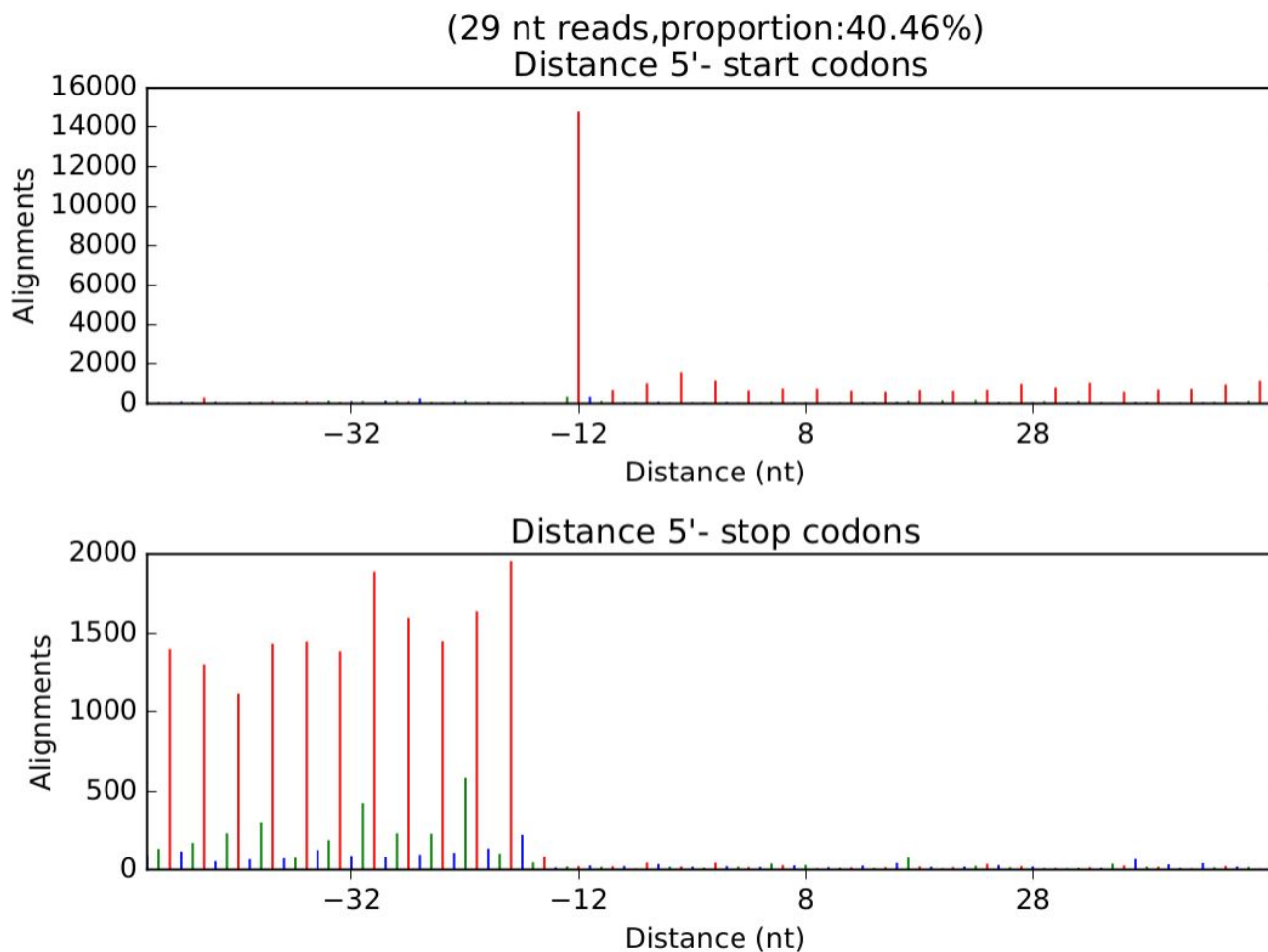
Длина прочтения и конформация рибосомы



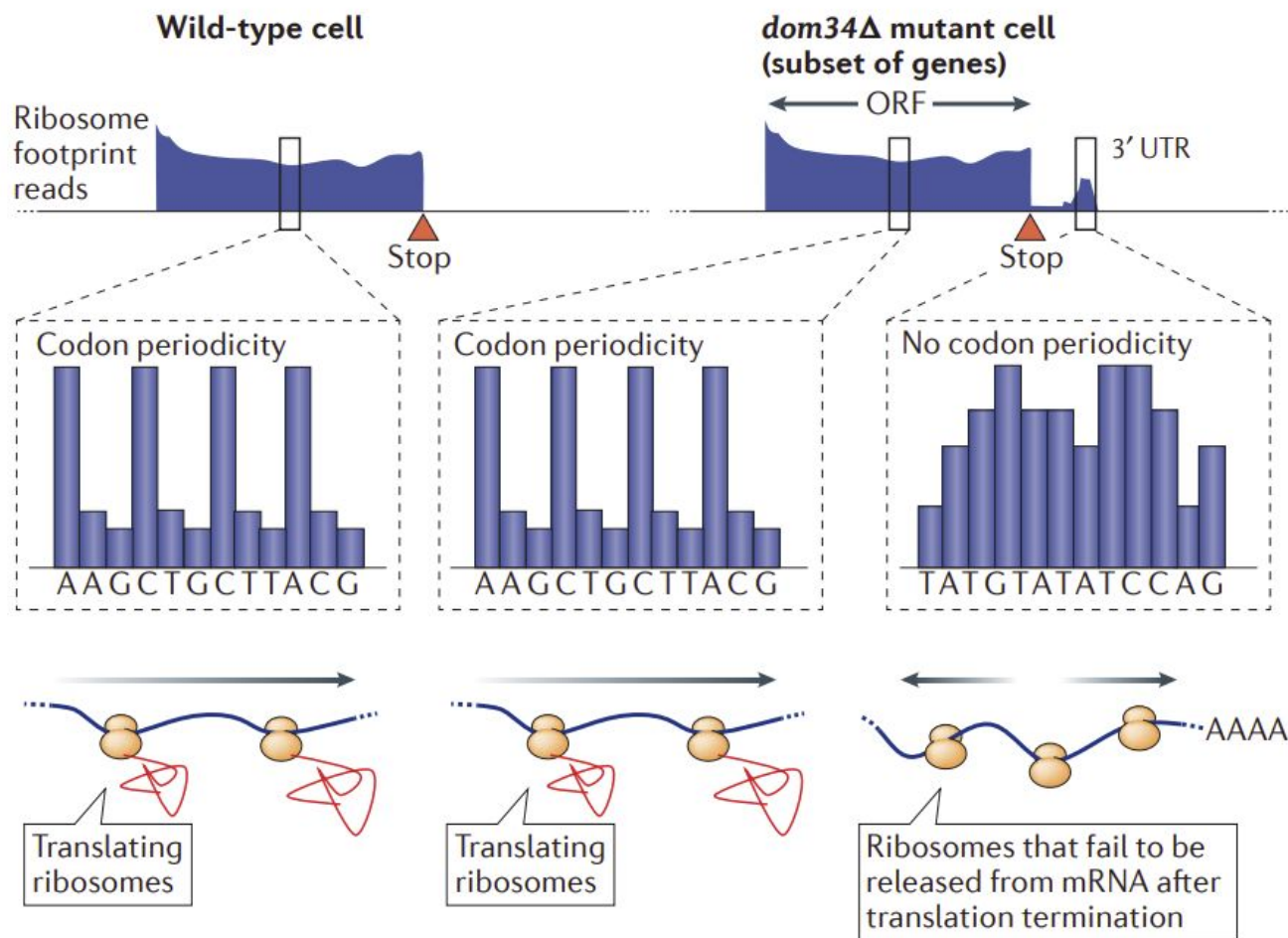
Фазирование рибосомы



Фазирование рибосомы

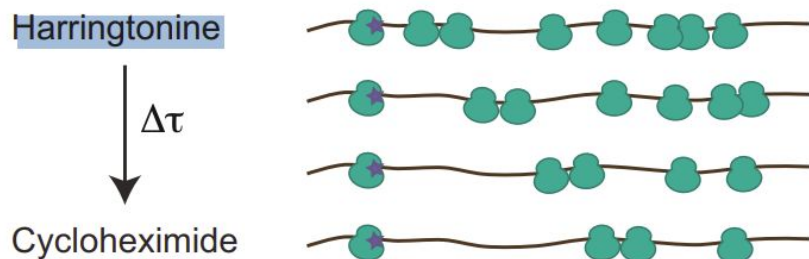


Изучение механизмов трансляции

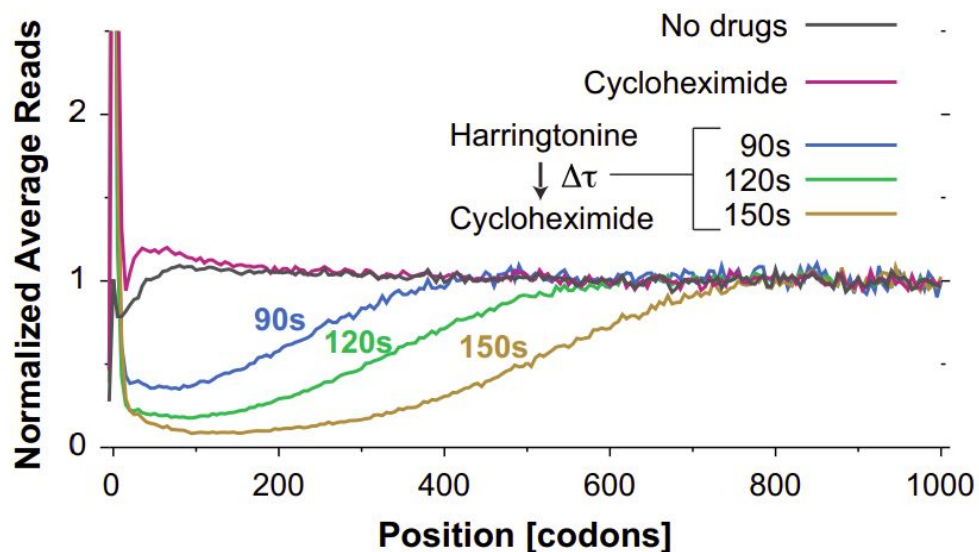


Изучение кинетики трансляции

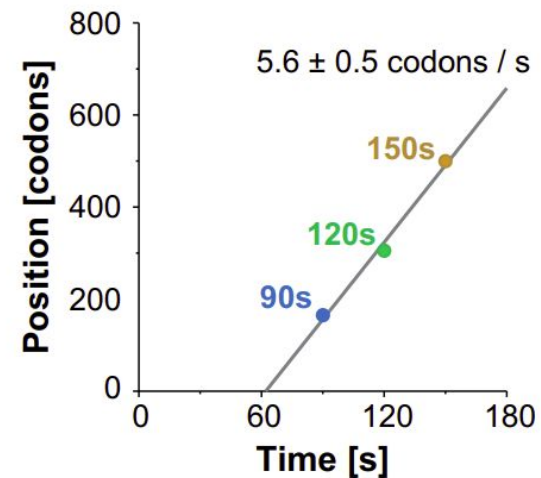
A



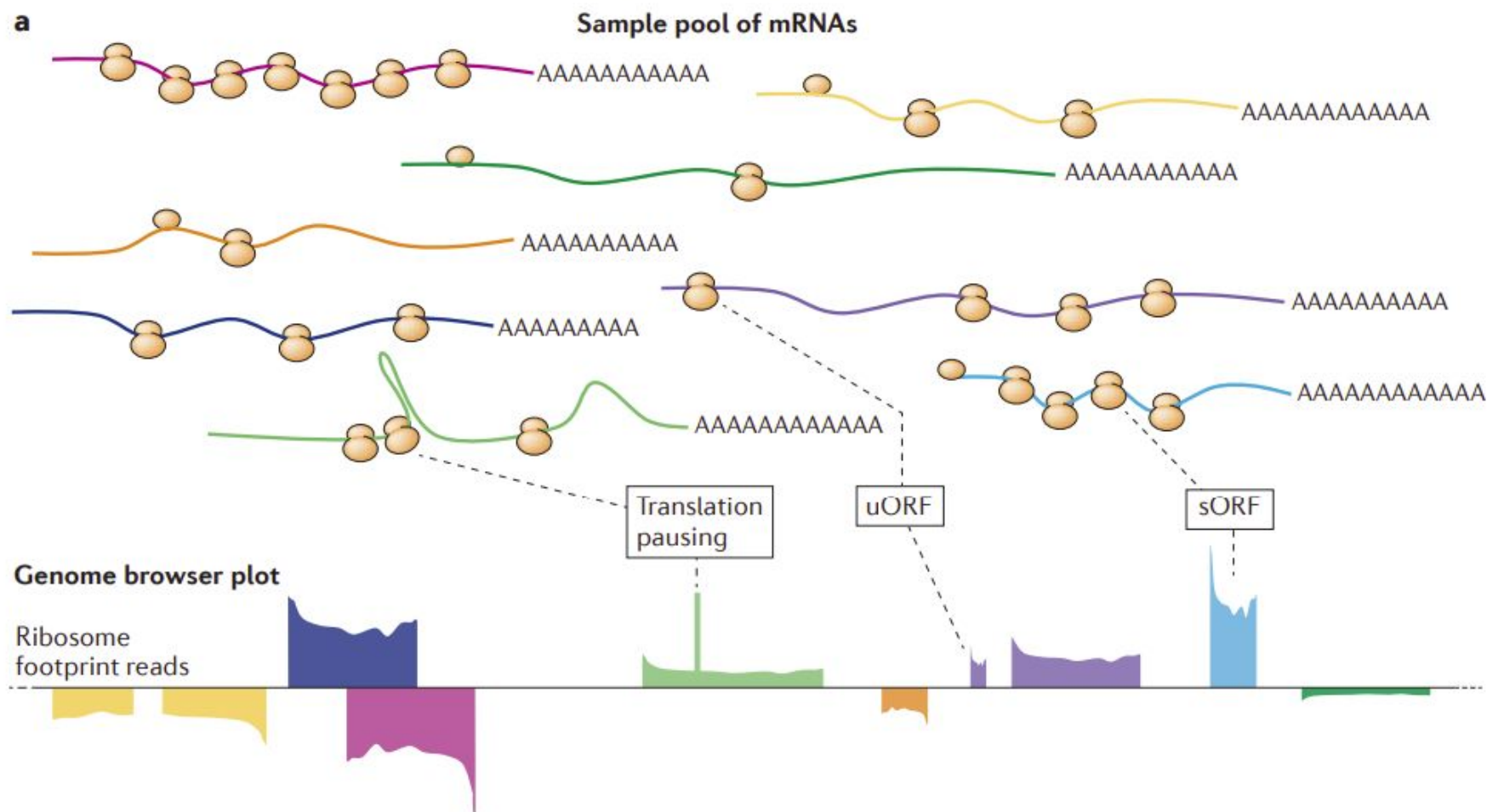
B



C



Поиск новых ORF'ов



QC в Ribo-seq

