

# Поиск новых V и J сегментов в данных секвенирования репертуара антител и рецепторов T клеток

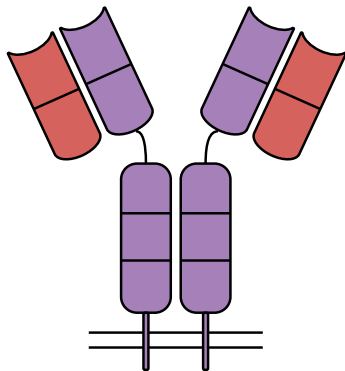
Проданов Тимофей Петрович

научный руководитель: к.ф.-м.н. Сафонова Яна Юрьевна

СПбАУ

15 июня 2017 г.

- В и Т клетки играют гигантскую роль в адаптивной иммунной системе
- Производят белки: антитела и рецепторы Т клеток
- Занимаются обнаружением и нейтрализацией антигенов

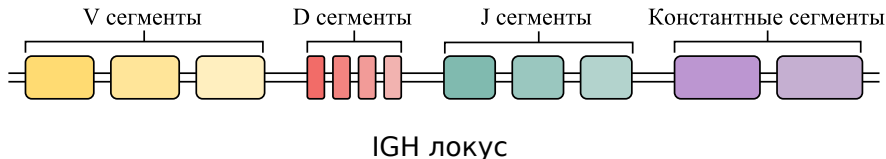


Антитело

Иммунные локусы содержат три семейства повторов:

- V — Variant
- D — Diversity
- J — Joining

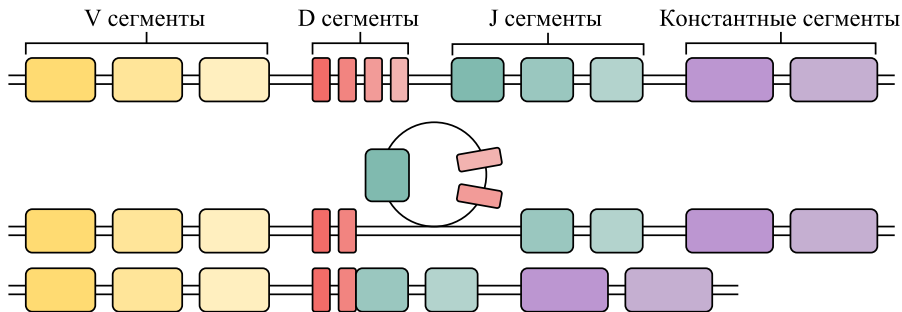
А также константные сегменты.



# V(D)J рекомбинация. I

В и Т клетки проходят стадию V(D)J рекомбинации:

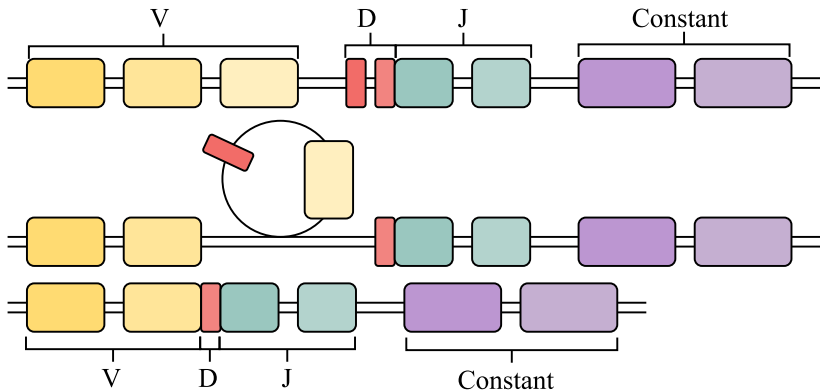
- Выбираются случайные D, J сегменты
- Геном между ними вырезается



# V(D)J рекомбинация. II

В и Т клетки проходят стадию V(D)J рекомбинации:

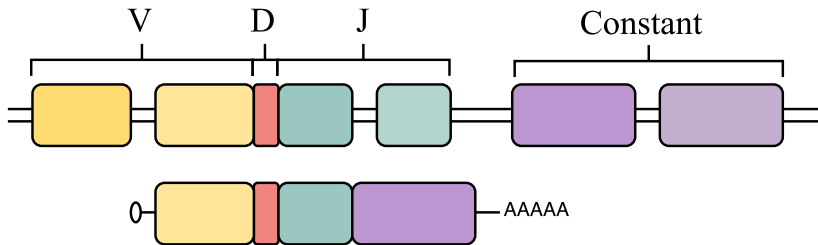
- Затем выбирается случайный V сегмент
- Геном между ним и оставшимися D сегментами вырезается



# V(D)J рекомбинация. Транскрипция

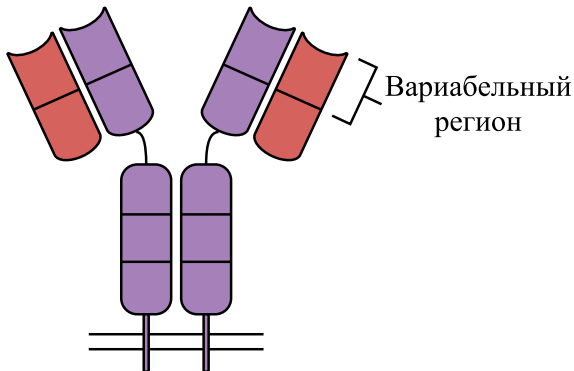
При транскрипции региона вырезаются:

- Лишние V и J сегменты
- Расстояние до константного сегмента
- Интроны в константном сегменте

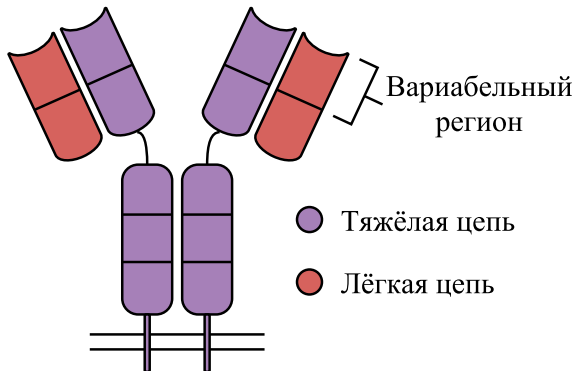


# Варибельный регион

- VDJ регион транслируется в *варибельный* регион антитела
- Константный сегмент — всё остальное



- Антитело состоит из тяжёлой и лёгкой цепи
- Тяжёлая цепь — **VDJ** рекомбинация
- Лёгкая цепь — **VJ** рекомбинация
- Локусы разные, сегменты разные



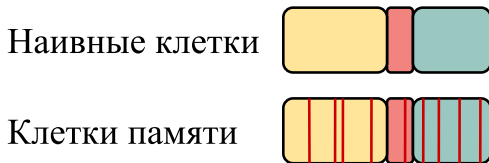


# Рецептор Т клетки

- Два типа:  $\alpha:\beta$  и  $\gamma:\delta$
- $\beta, \gamma$  — **VDJ** рекомбинация
- $\alpha, \delta$  — **VJ** рекомбинация



- Большинство В клеток проходят стадию *гипермутагенеза*
- **Не** прошли стадию — наивные клетки
- Пройли стадию — клетки памяти, плазма клетки
- Т клетки не проходят стадию гипермутагенеза



- Семейство технологий секвенирования репертуара антител (*Rep-seq*)
- В основном: РНК-секвенирование переменного региона
- Парные прочтения, достаточная длина

- Существует база данных иммунных сегментов IMGT
- Была сделана с помощью сборки локуса
- Сборка локуса — тяжёлая задача из-за высокой повторяемости

## Проблема

Предполагается, что база данных IMGT не полна.

## Мотивация

- Изучение В и Т клеток жизненно необходимо:
  - Разработка лекарств
  - Изучение иммунной системы
- Большинство методов ориентируется на знание иммунных сегментов
- Rep-seq данные не использовались для расширения базы данных
- Предпосылка для дальнейшего популяционного анализа

## Цель

Дополнить базу данных иммунных сегментов IMGT с помощью данных Rep-seq

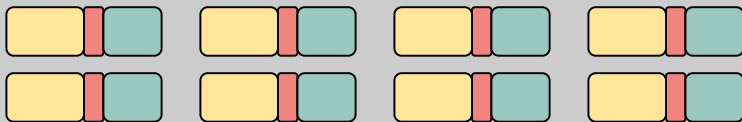
## Задачи

- Разработать метод поиска неучтённых иммунных сегментов
- Применить метод к данным Rep-seq наивных В клеток и Т клеток
- Валидировать сегменты используя данные Rep-seq В клеток памяти

# Входные и выходные данные

## Входные данные

Rep-seq



IMGT

V1

V2

V3

## Выходные данные

V1

V2

V3

V4

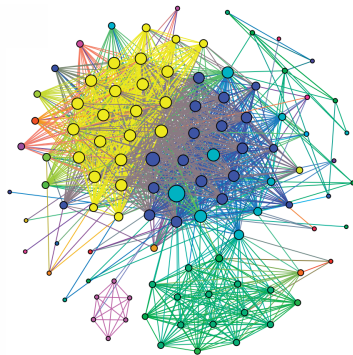
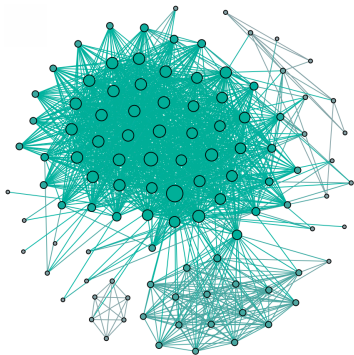
V5

Следует отличить неучтённый сегмент от

- PCR-артефактов
- Клеток памяти

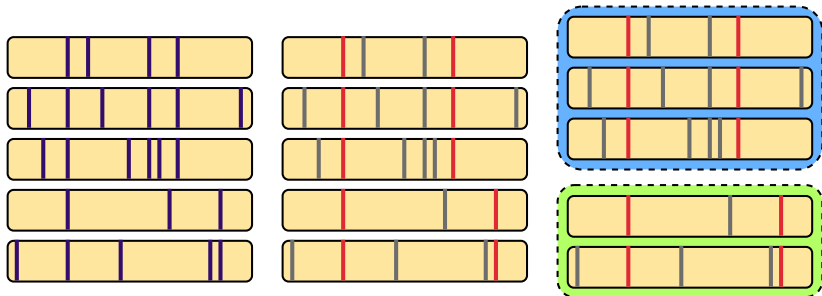


- На входных данных запускается Ig Repertoire Constructor (*IgReC*)
- На Rep-seq прочтениях строится граф Хэмминга
- Почти полные клики склеиваются в одно прочтение
- Цель: избавиться от PCR-артефактов



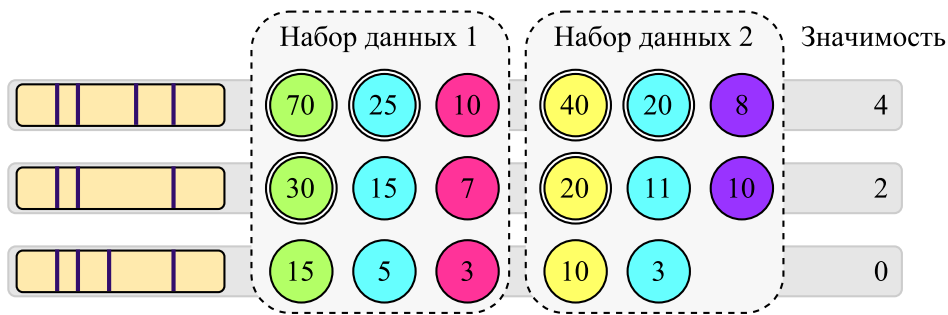
В прочтениях, выровнявшихся на один и тот же сегмент:

- Поиск слишком частых полиморфизмов
- Комбинирование по совместной встречаемости
- Генерация новых сегментов



Считаем значимость нового V сегмента:

- Со сколькими J сегментами встретился
- J сегменты в разных наборах данных считаются разными
- PCR-артефакты и клетки памяти будут иметь маленькую значимость



Наборы данных:

- Наивные В клетки: 19 наборов данных
- Т клетки: 30 наборов данных

Наивные В клетки (тяжёлые цепи):

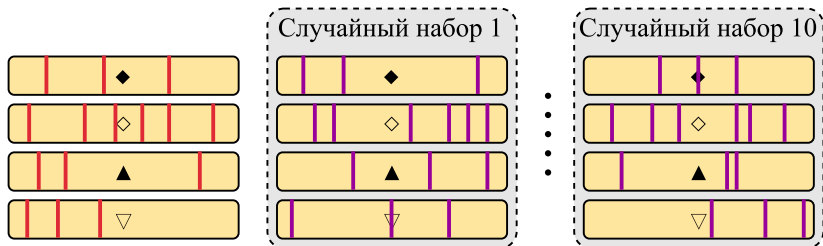
- 20 новых V сегментов  
Если разрешить небольшое отличие: 45
- 4 новых J сегментов

T клетки:

- 20 новых V сегментов  
19 -  $\beta$ , 1 -  $\alpha$
- 2 новых J сегментов  
1 -  $\beta$ , 1 -  $\alpha$

# Валидация В клетками памяти. I

- Используется много наборов данных Rep-seq клеток памяти
- Наборы данных не связаны с наивными наборами данных
- В прочтениях, выровнявшихся на новые сегменты замеряется доля полиморфизмов в ключевых позициях
- Для сравнения — случайно сгенерированные новые сегменты



39 наборов данных В клеток памяти.

Валидацию проходят:

- 7 V сегментов из 20
- 0.2 из 20 случайных (в среднем)
  
- 3 J сегмента из 4
- 0.3 из 4 случайных (в среднем)

Все сегменты не будут провалидированы:

- Могут не присутствовать в клетках памяти
- Могут не присутствовать в других людях

- Разработан метод поиска неучтённых иммунных сегментов
- Найдены новые V и J сегменты в наивных В клетках и Т клетках
- Некоторые новые сегменты провалидированы с помощью В клеток памяти