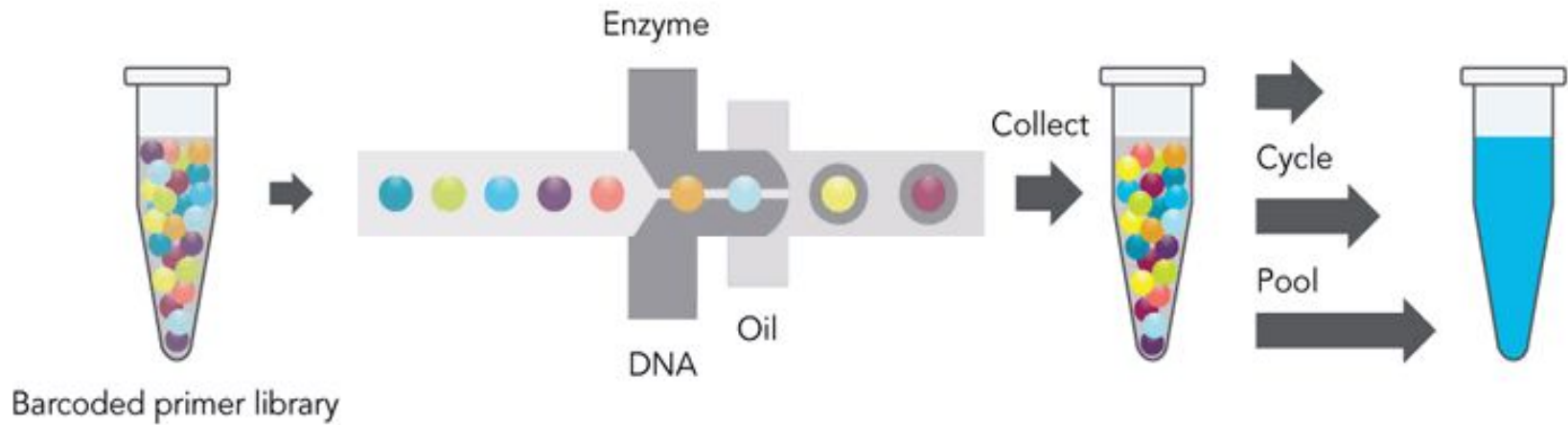


# 10xGenomics data

Иван Толстогоанов

Руководитель: Антон Банкевич,  
ЦАБ СПбГУ

# Технология



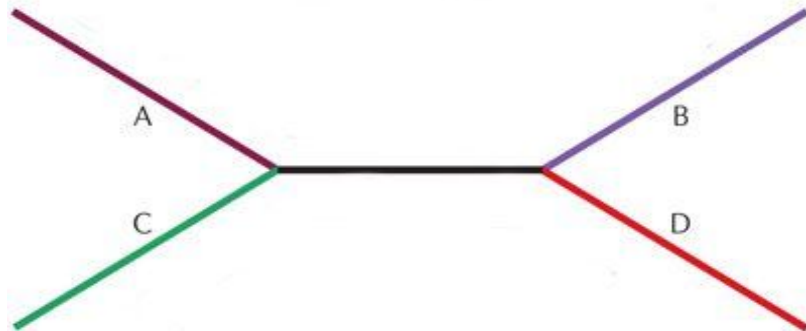
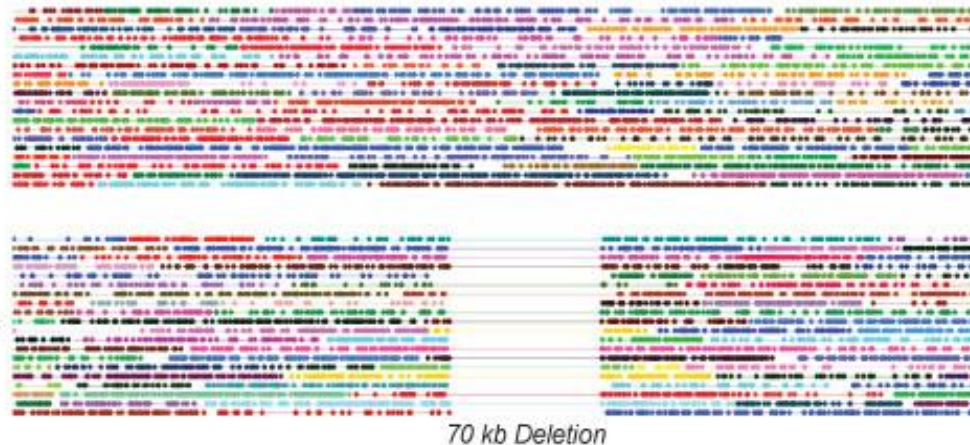
# Применения

- Поиск структурных вариаций
- Восстановление гаплотипа
- Обнаружение новых клеточных популяций

## Публикации:

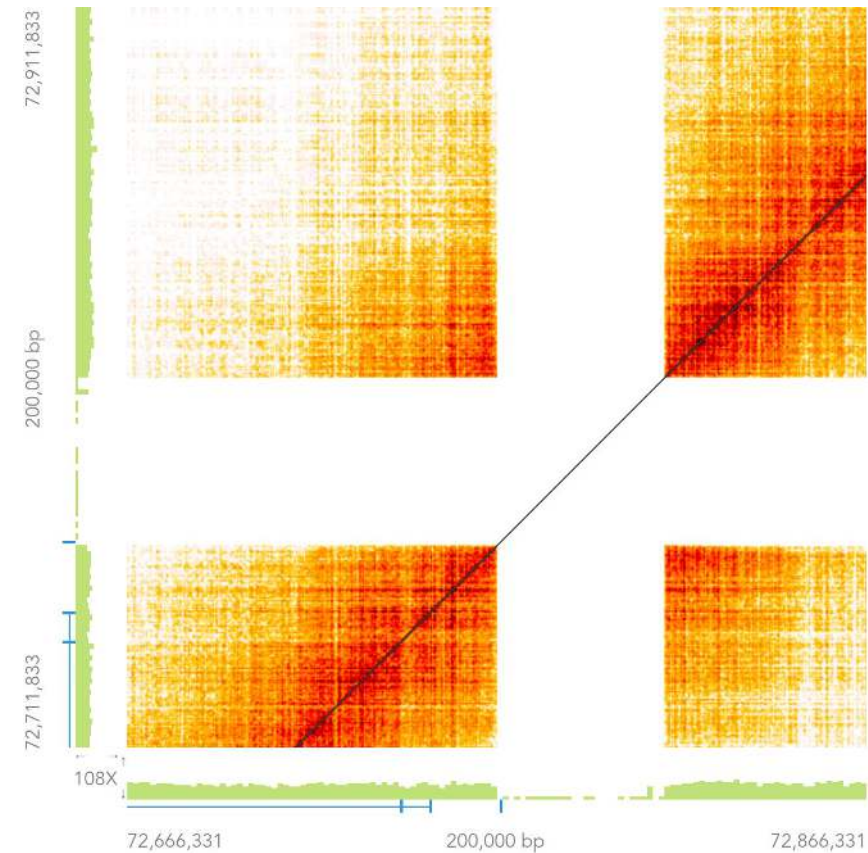
- [Haplotyping germline and cancer genomes with high-throughput linked-read sequencing](#)
- [A hybrid approach for \*de novo\* human genome sequence assembly and phasing](#)

# Применение к сборке



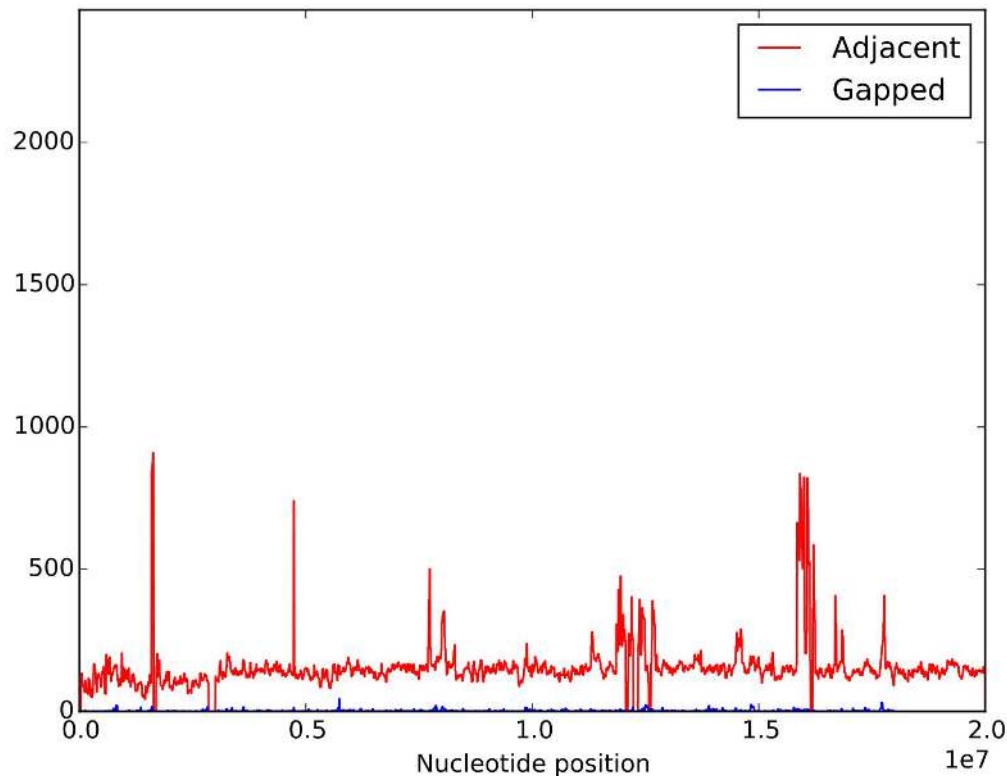
- Хотим понять, в какое ребро в графе де Брюйна надо идти из повтора
- У близких ребер в геноме должен быть похожий набор цветов

# Почему это должно работать?



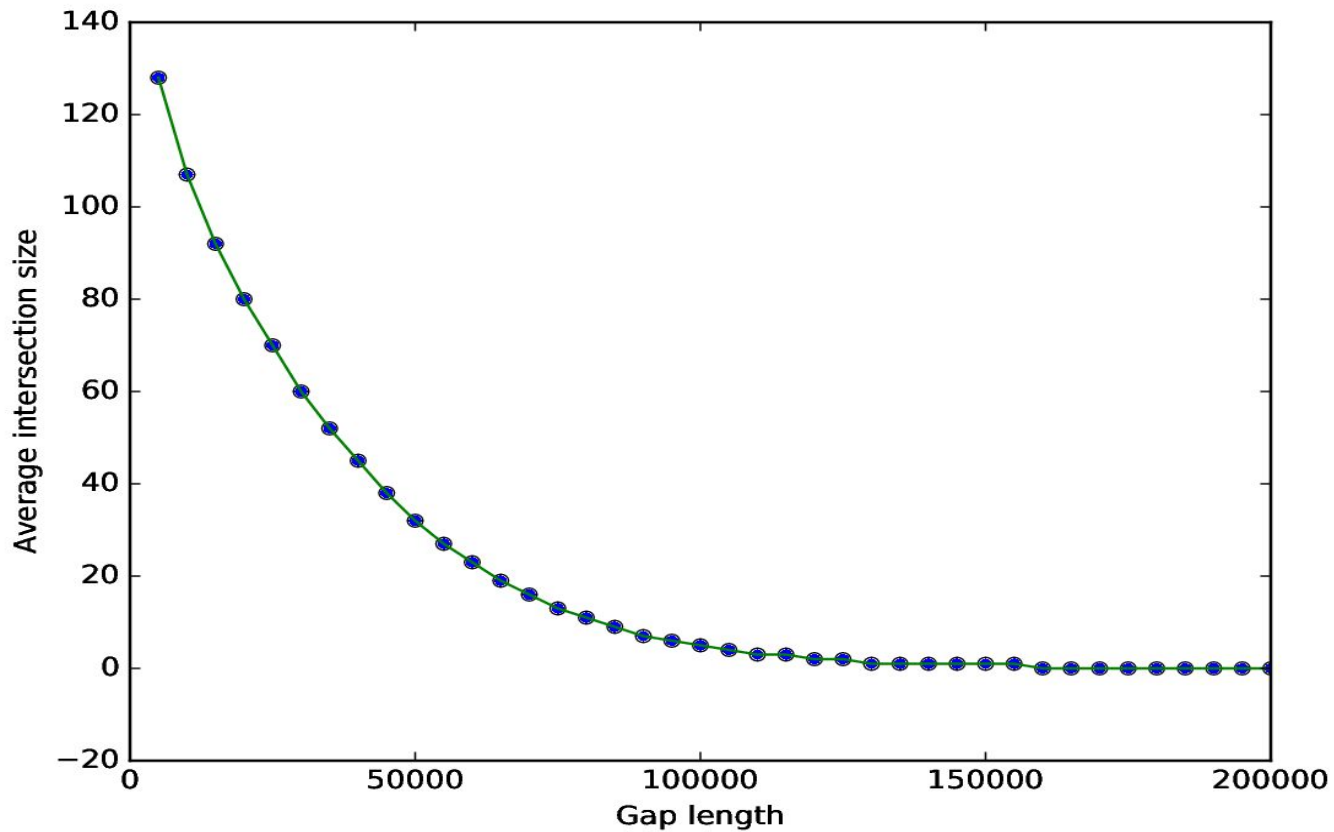
- Редко встречается ситуация, когда у расположенных далеко друг от друга в референсе фрагментов мало общих баркодов
- Это может свидетельствовать о наличии вариации

# Распределение баркодов по референсу



- Среднее значение совпавших баркодов для подряд идущих рамок: 150
- Для рамок из разных частей референса: 2.4

# Непрерывность распределения баркодов

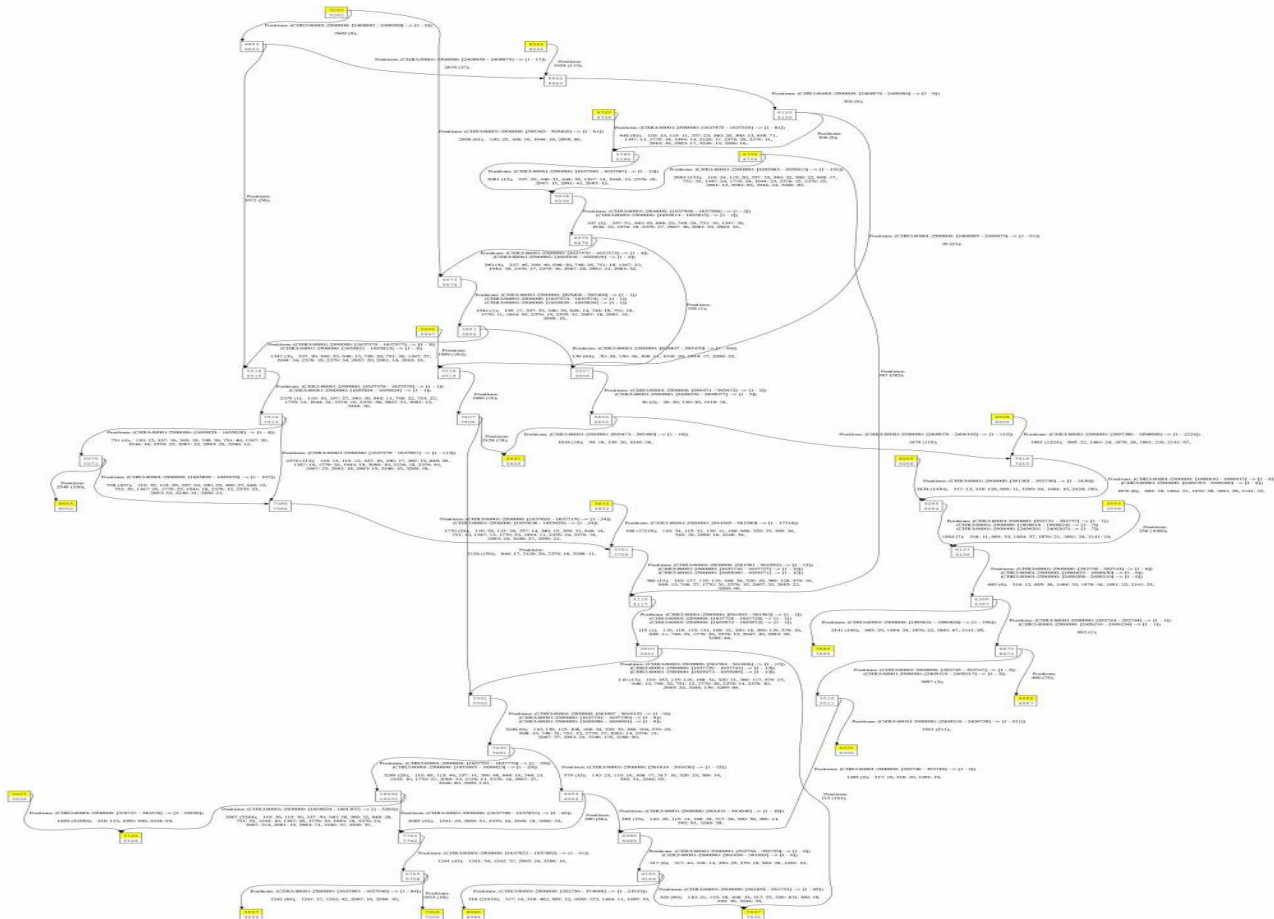


# Как должно работать

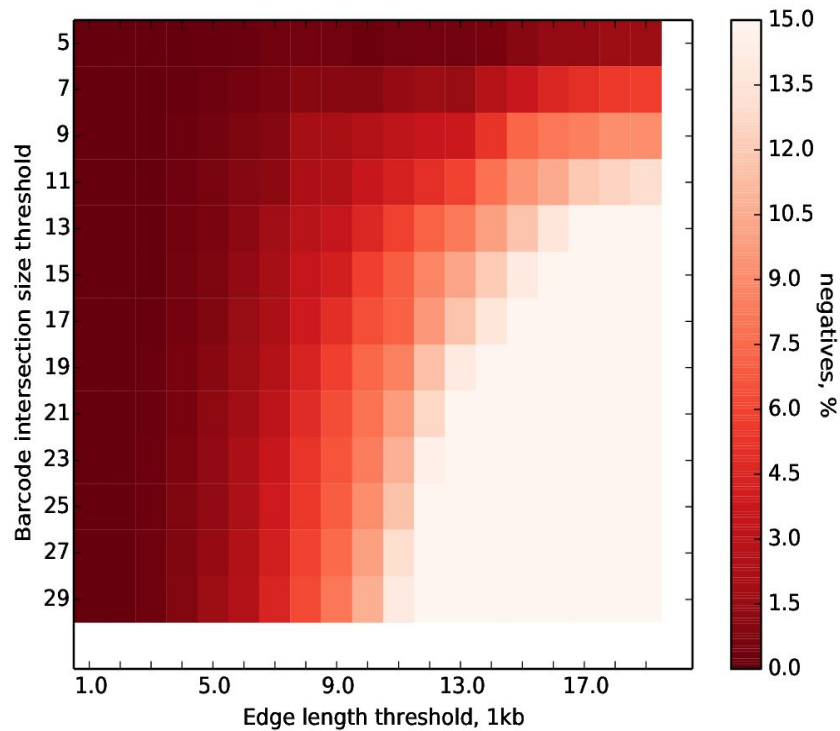
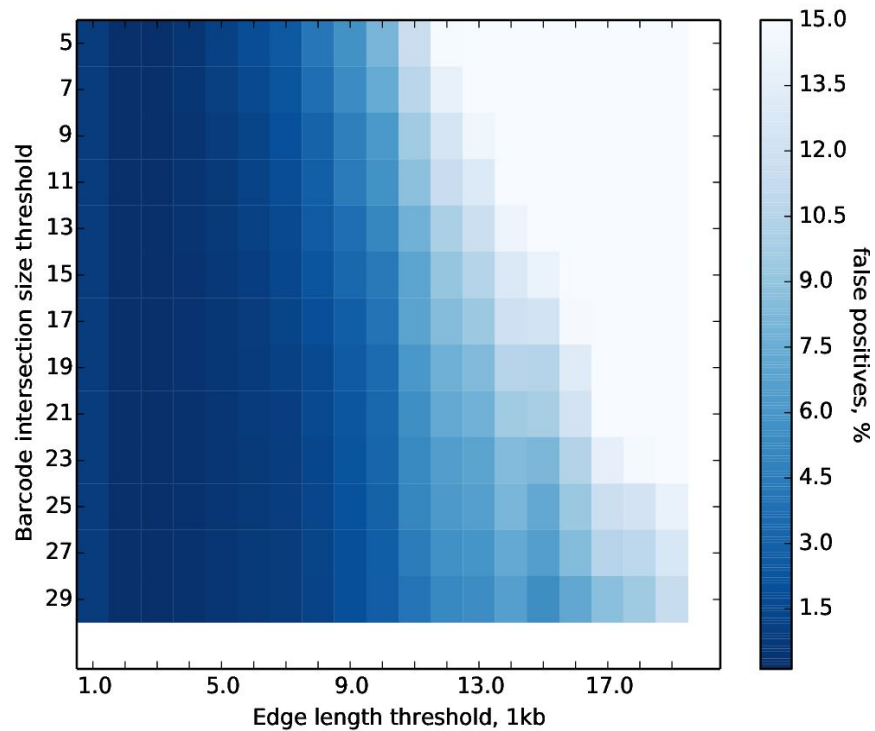
- Постепенно увеличиваем правильный путь в графе
- Считаем, что у следующего в геноме ребра пересечение с текущим должно быть выше определенного порога
- Бесполезно смотреть на пересечение по баркодам со слишком короткими ребрами, поэтому смотрим на достаточно длинные ребра



# Как это выглядит



# Подбор параметров



# Результаты

- Проведен первичный анализ данных секвенирования
- Реализована структура данных для отображения информации о баркодах
- Изучены возможности технологии применительно к разрешению повторов
- **Итог: на основе данных 10xGenomics можно разработать алгоритмы для разрешения повторов в сложных геномах**

## Что дальше?

- Реализация более продвинутого repeat resolver'a
- Использование данных Illumina TruSeq Synthetic Long-Read

**Спасибо!**