

февраль 2022 г.

студенческая школа-
конференция ИТИС(б)

РАЗРАБОТКА МЕТОДОВ АНАЛИЗА ДАННЫХ NGS В ЭКСПЕРИМЕНТАХ ПО ЛЕНТИВИРУСНОЙ ТРАНСДУКЦИИ МЕЗЕНХИМАЛЬНЫХ СТВОЛОВЫХ КЛЕТОК

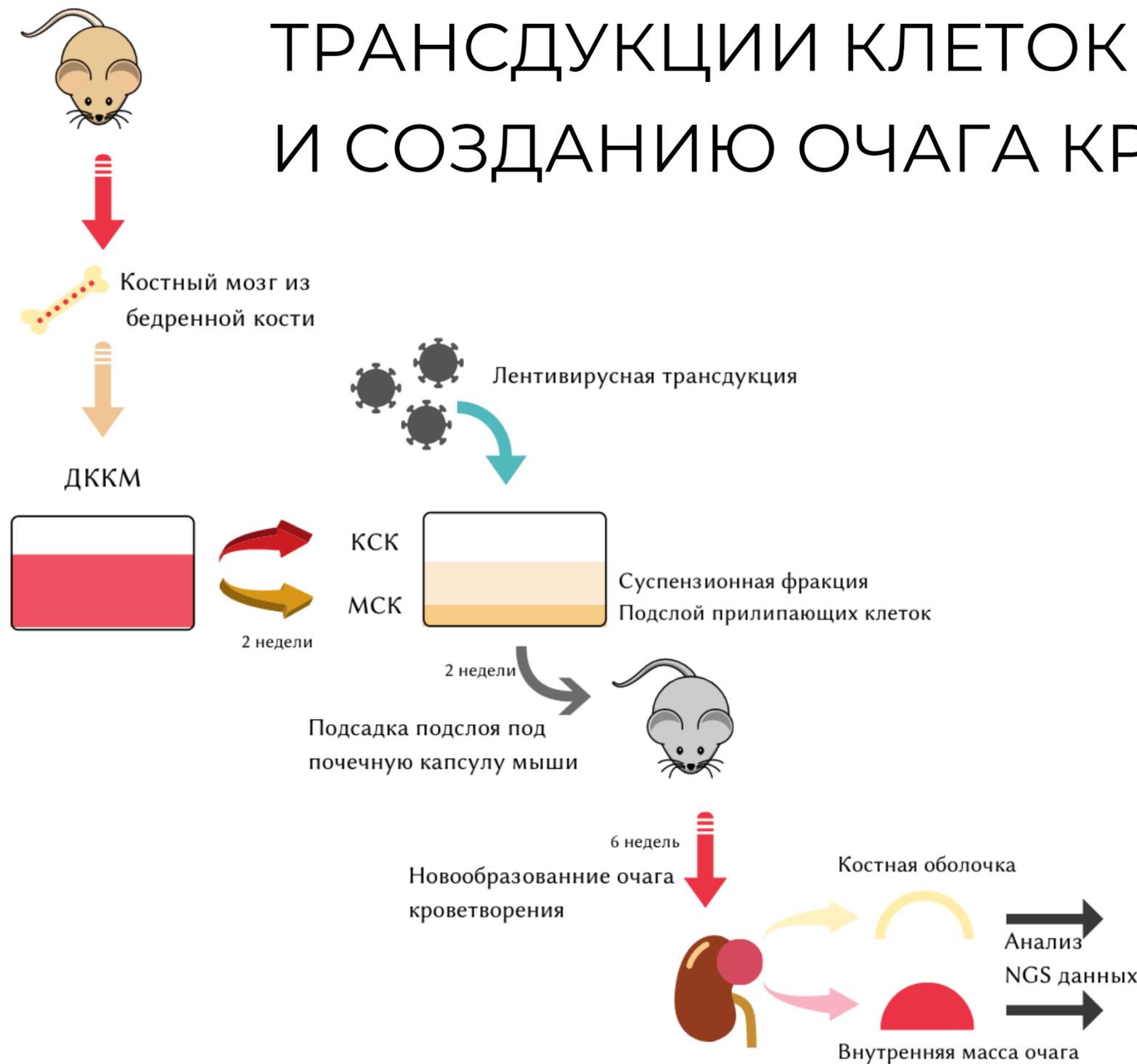
**Д.Д.Кожевникова,
А.К. Шайтан**

Биологический факультет, МГУ им М.В.Ломоносова;
Международная лаборатория биоинформатики, Факультет
Компьютерных Наук, Высшая Школа Экономики

**Д.В.Карпенко,
А.Е.Бигильдеев**

Лаборатория физиологии кроветворения ФГБУ “НМИЦ
Гематологии” Минздрава России

СХЕМА ЭКСПЕРИМЕНТА ПО ЛЕНТИВИРУСНОЙ ТРАНСДУКЦИИ КЛЕТОК КОСТНОГО МОЗГА МЫШИ И СОЗДАНИЮ ОЧАГА КРОВЕТВОРЕНИЯ



МСК - мезенхимные стволовые клетки

В новообразованном эктопическом очаге кроветворения стромальные клетки очага происходят от клеток донора, кроветворные клетки принадлежат реципиенту.

Цель лентивирусной трансдукции состоит в индивидуальном маркировании стромальных клеток в новообразованном очаге кроветворения. Маркировкой служит сайт интеграции лентивирусного вектора в геном клетки. За счет совпадения сайтов интеграции у клеточных клонов есть возможность оценить размер клонов.

Работа посвящена попытке исследования пролиферативного и дифференцировочного потенциала популяции мезенхимных стволовых клеток (МСК) в красном костном мозге мышей.

Цель:

Разработка методов для оценки и оценка количества мезенхимальных стволовых клеток в костном мозге мышей с помощью анализа сайтов интеграции лентивирусного вектора в геномной ДНК клеток представленных в NGS данных.

Задачи:

- Дизайн принципиальной схемы анализа данных на основе протоколов эксперимента.
- Оценка качества данных, построение модели накопления ошибок в данных.
- Предварительный анализ данных: определение количества ридов в полученных данных и степени соответствия ридов ожидаемой структуре.
- Разработка алгоритмов фильтрации и кластеризации данных.
- Анализ и сравнение технических повторов (оценка разнообразия и повторяемости IS в технических повторах).
- Разработка различных подходов к оценке количества уникальных сайтов интеграции.
- Разработка различных подходов к оценке количества уникальных клонов.

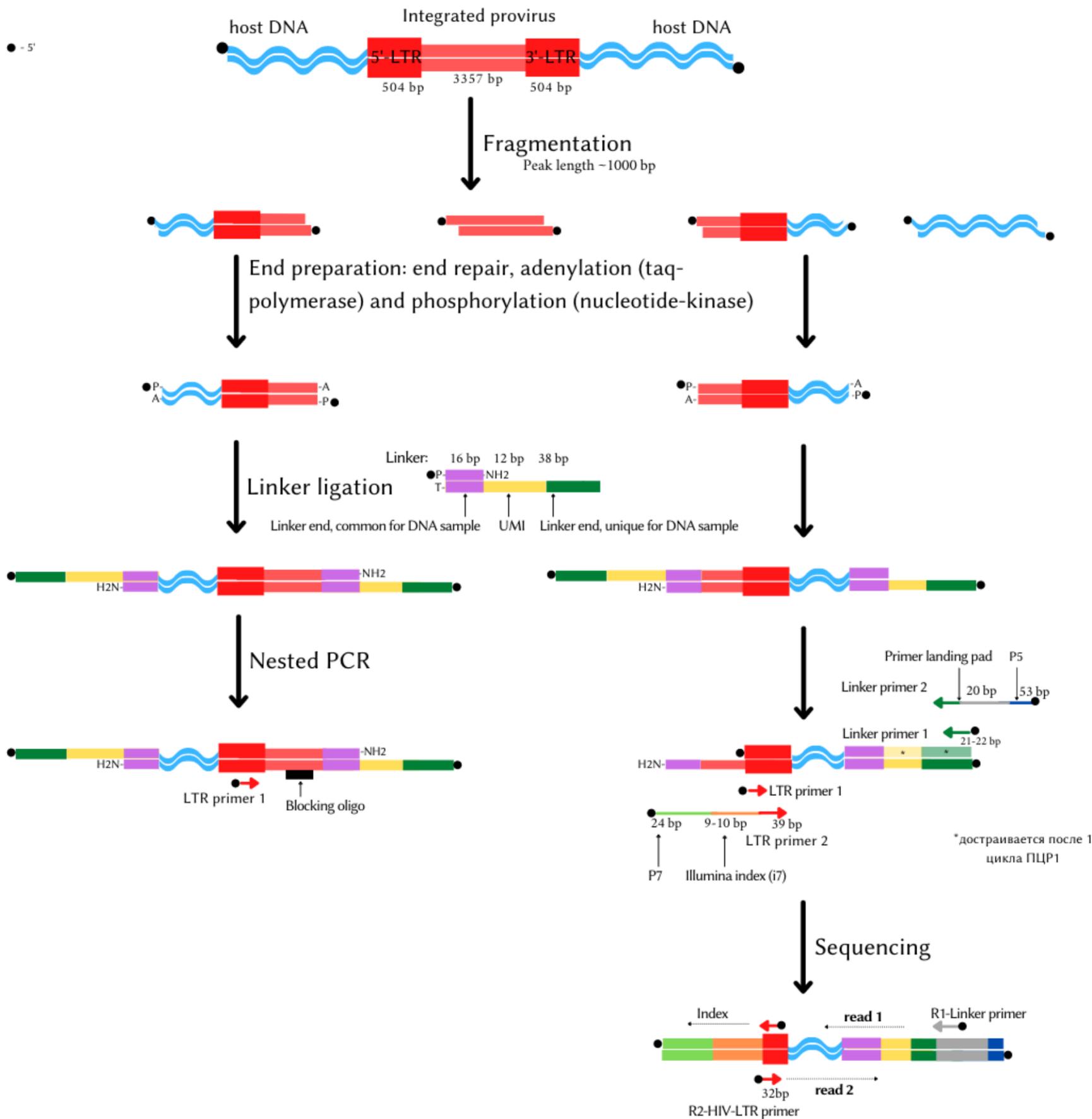
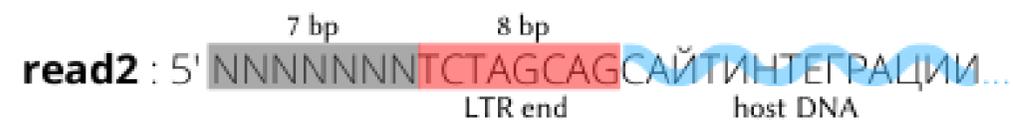
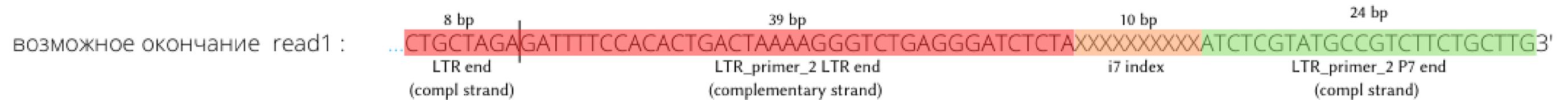
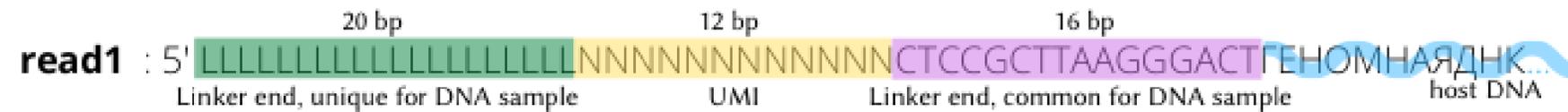


СХЕМА ПРОБОПОДГОТОВКИ И СЕКВЕНИРОВАНИЯ

В линкере содержится 3 типа последовательностей:

- последовательность уникальная для образца
- последовательность общая для всех образцов
- UMI - unique molecular identifier - случайная последовательность из 12 оснований, используемая в качестве баркода для каждой молекулы ДНК, поступившей в пробоподготовку на этапе лигирования линкеров.

ПРЕДПОЛАГАЕМАЯ СТРУКТУРА РИДОВ



возможное окончание read2 :

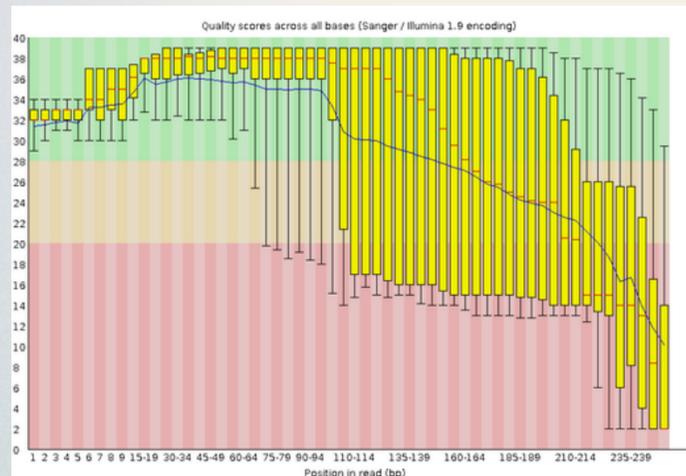


РЕСУРСЫ ИСПОЛЬЗОВАННЫЕ ДЛЯ НАПИСАНИЯ КОДА

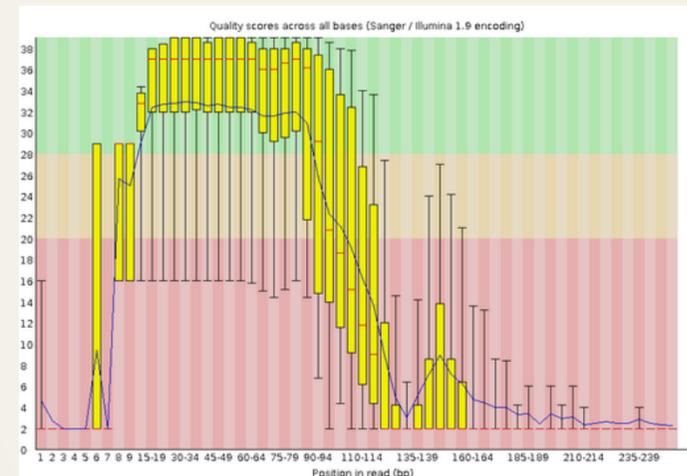
ОЦЕНКА РАЗМЕРА И КАЧЕСТВА
ИСХОДНЫХ ДАННЫХ

- с помощью программы **FastQC**

Read1 phred score

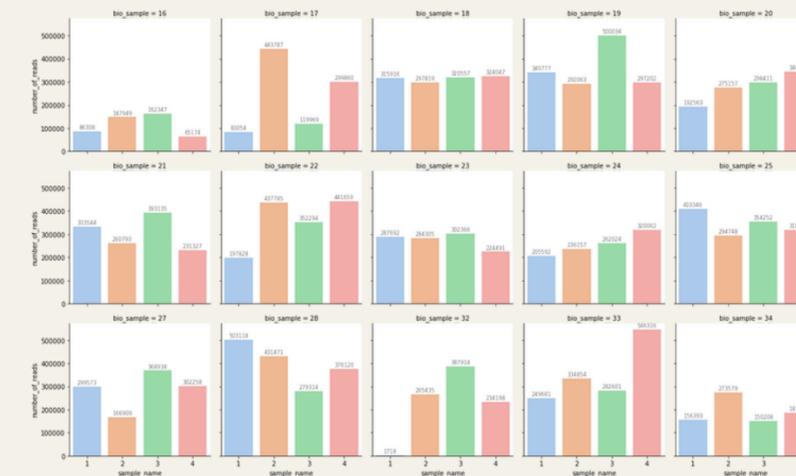


Read2 phred score



Кол-ва ридов

в 15-ти экспериментах, содержащих по 4 повтора



ВЫРАВНИВАНИЕ РИДОВ

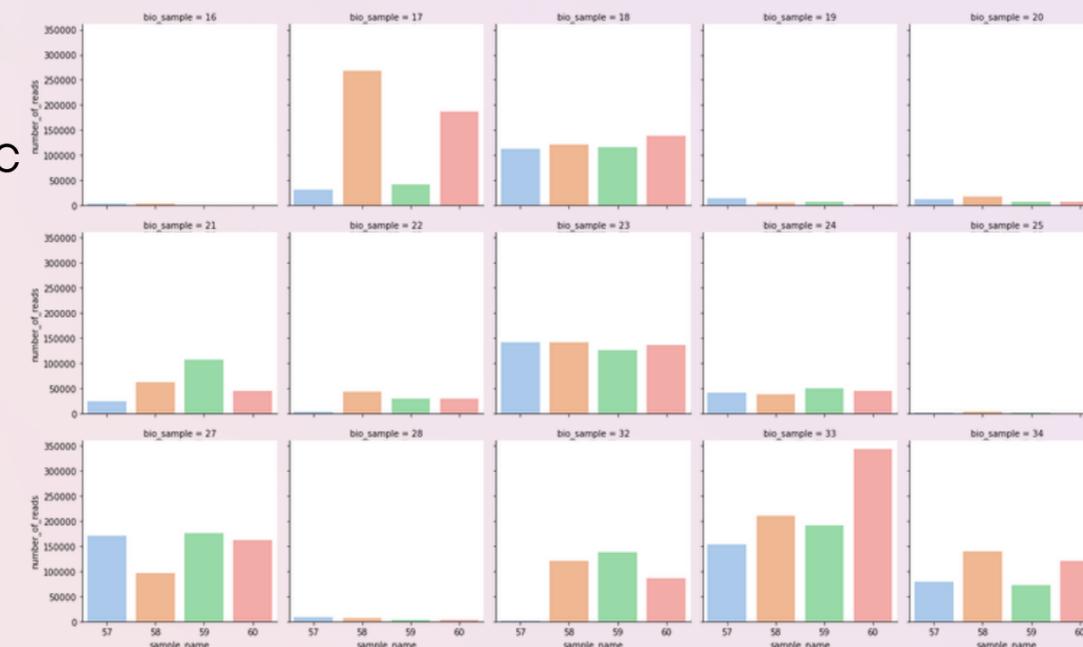
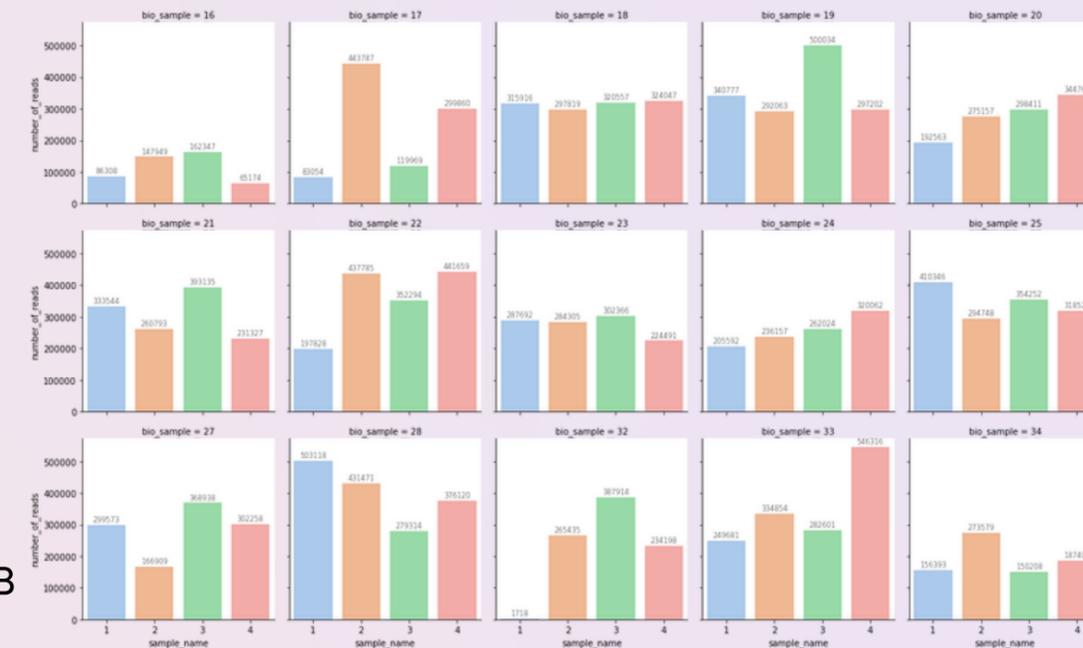
- в программе **BWA**

АЛГОРИТМЫ ДАЛЬНЕЙШЕЙ
ОБРАБОТКИ ДАННЫХ

- написаны на языке Python с использованием библиотек **BioPython** и **HTSeq**

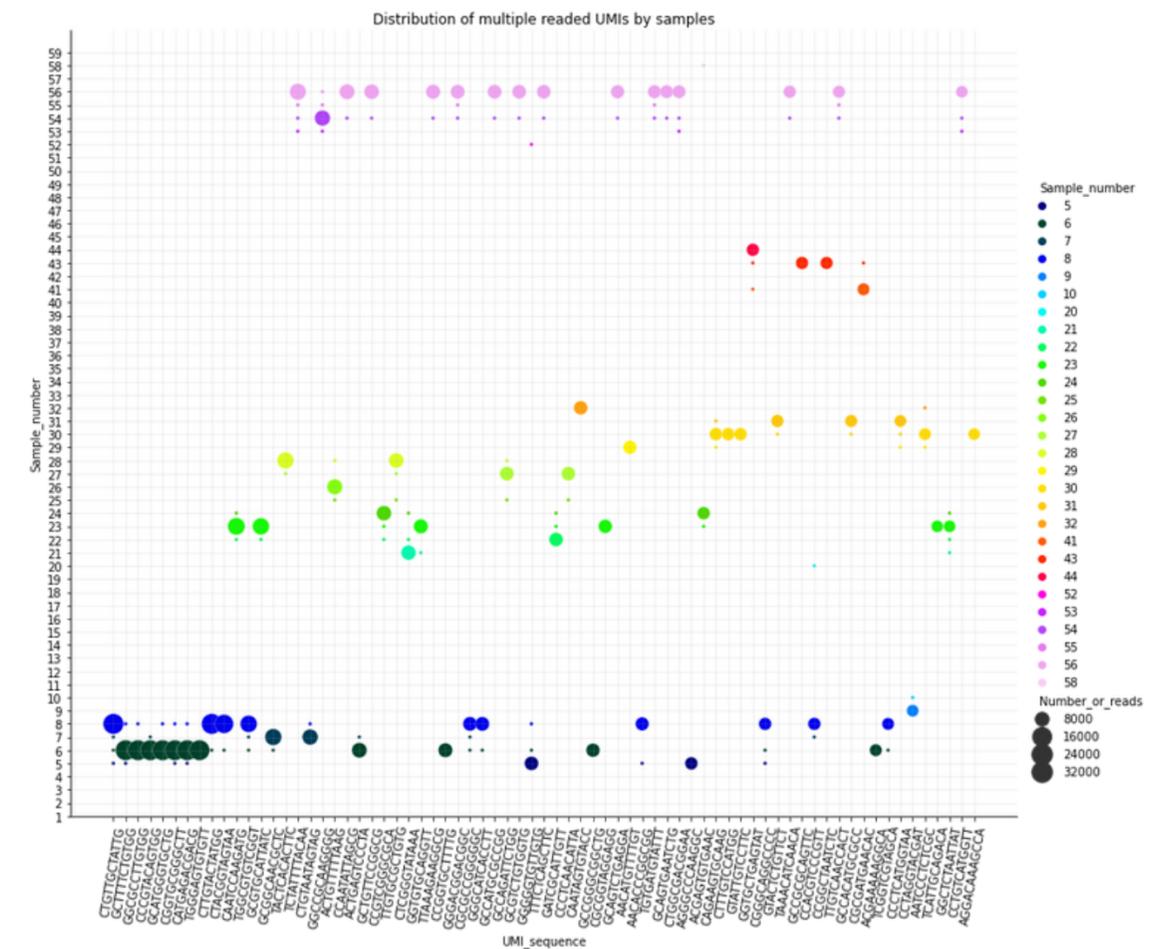
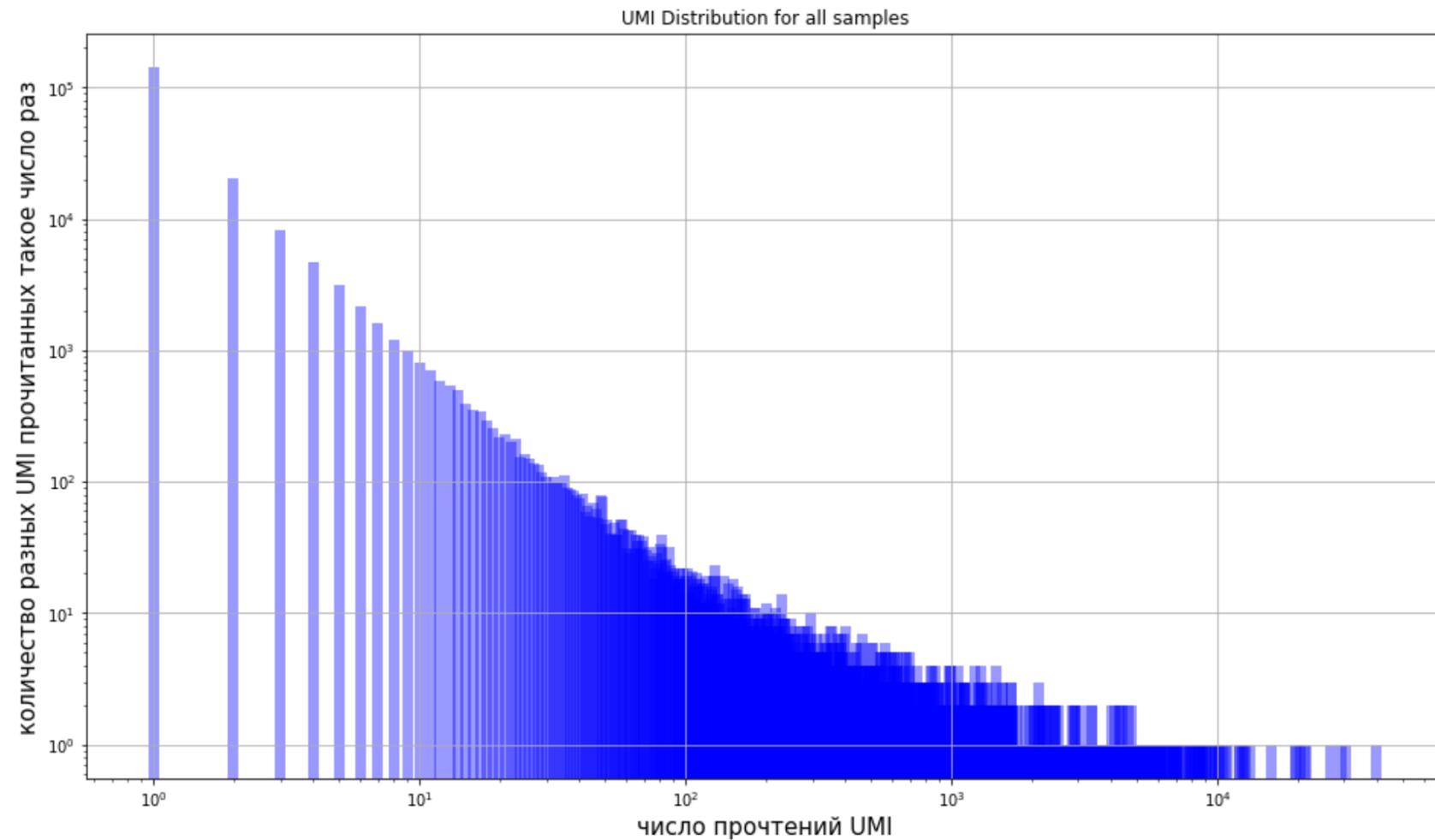
Фильтрация данных:

- Фильтрация ридов 2 по наличаю фрагмента LTR провируса
- Фильтрация ридов 1 по наличаю фрагмента, уникального для биологического образца
- Фильтрация ридов 1 по наличаю фрагмента, общего для всех образцов
- Отбор правильно выравненных парных ридов, в которых:
 - оба рида в паре выравнены на разные цепи ДНК
 - оба рида выравнены на одну хромосому
 - расстояние между началом выравнивания рида 2 и концом выравнивания рида 1 не превышает 3000 пн
- Отбор тех ридов, в которых в рида 2 начало выравнивания совпадает с последовательностью LTR
- Отфильтровка нежелательного продукта амплификации - ридов 2, содержащих внутренний фрагмент провируса
- Отфильтровка по длине выравнивания на референсный геном - выбирались рида, геномные фрагменты в которых > 19 оснований



ВИЗУАЛИЗАЦИЯ АМЛИФИЦИРОВАННОСТИ ПОСЛЕДОВАТЕЛЬНОСТЕЙ UMI

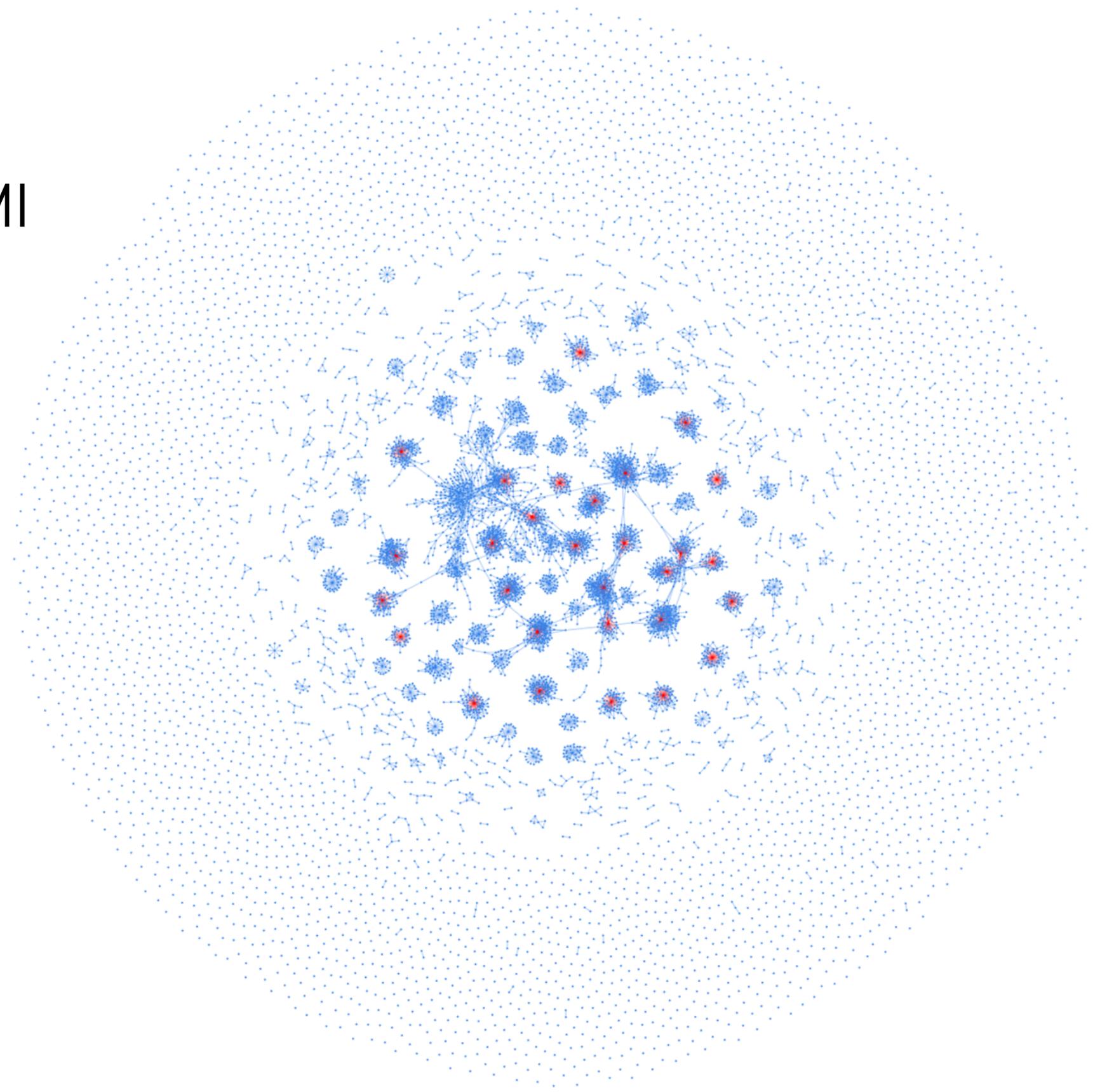
UMI (unique molecular identifier) – последовательность, содержащаяся в линекере, маркирующая каждую уникальную молекулу, поступившую на этапе пробоподготовки. Каждая уникальная последовательность UMI соответствует индивидуальной клетке.



КЛАСТЕРИЗАЦИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ UMI

UMI (unique molecular identifier) – последовательность, содержащаяся в линекере, маркирующая каждую уникальную молекулу, поступившую на этапе пробоподготовки. Каждая уникальная последовательность UMI соответствует индивидуальной клетке.

Для кластеризации были построены графы, вершины которого соответствуют уникальным последовательностям UMI, ребра соединяют те UMI, расстояние Левенштейна между которыми равняется 1, красным цветом покрашены вершины, соответствующие UMI, которые были прочитаны >5000 раз.

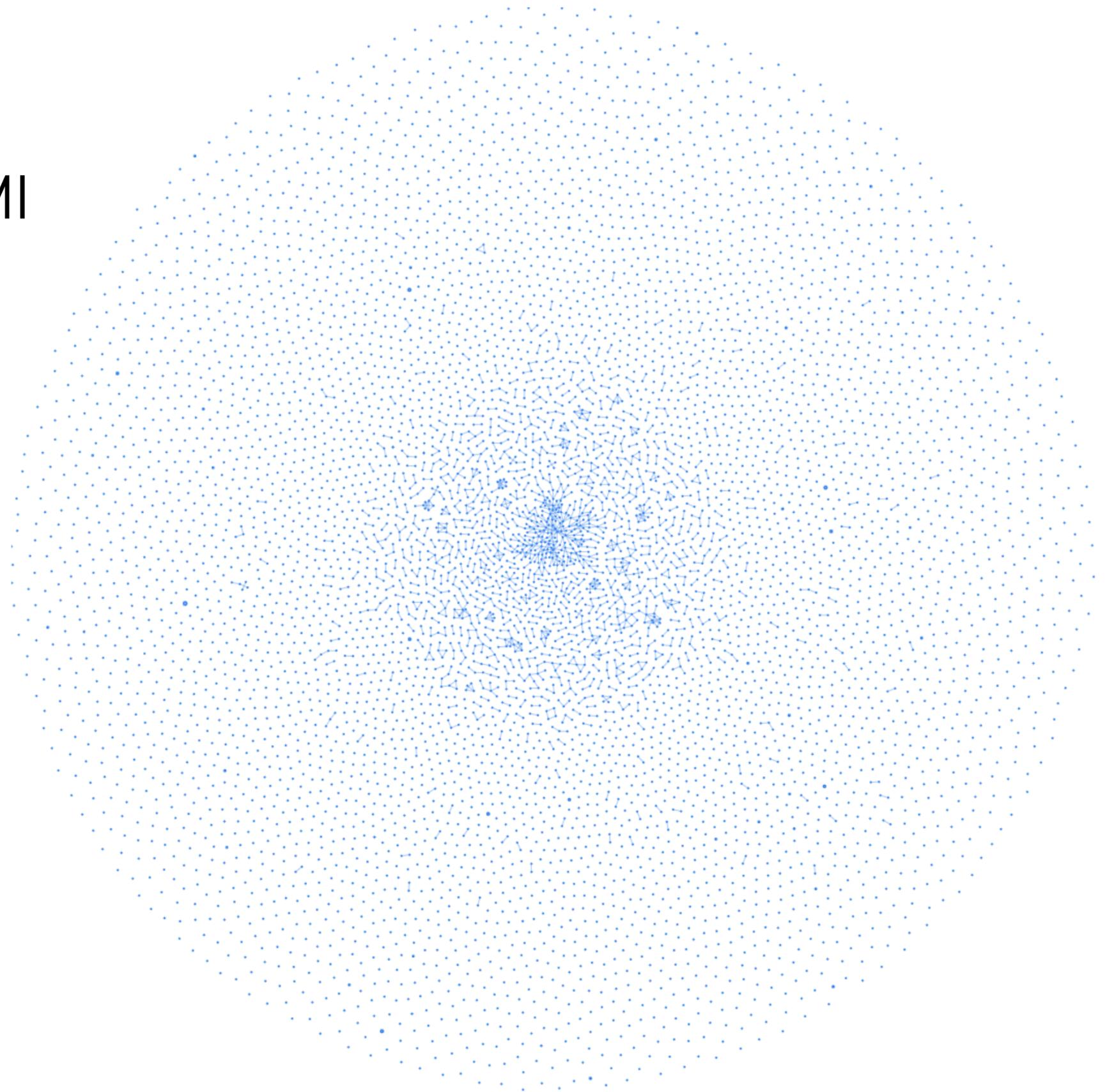


Граф, визуализирующий пространство UMI для 17 биологического образца

КЛАСТЕРИЗАЦИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ UMI

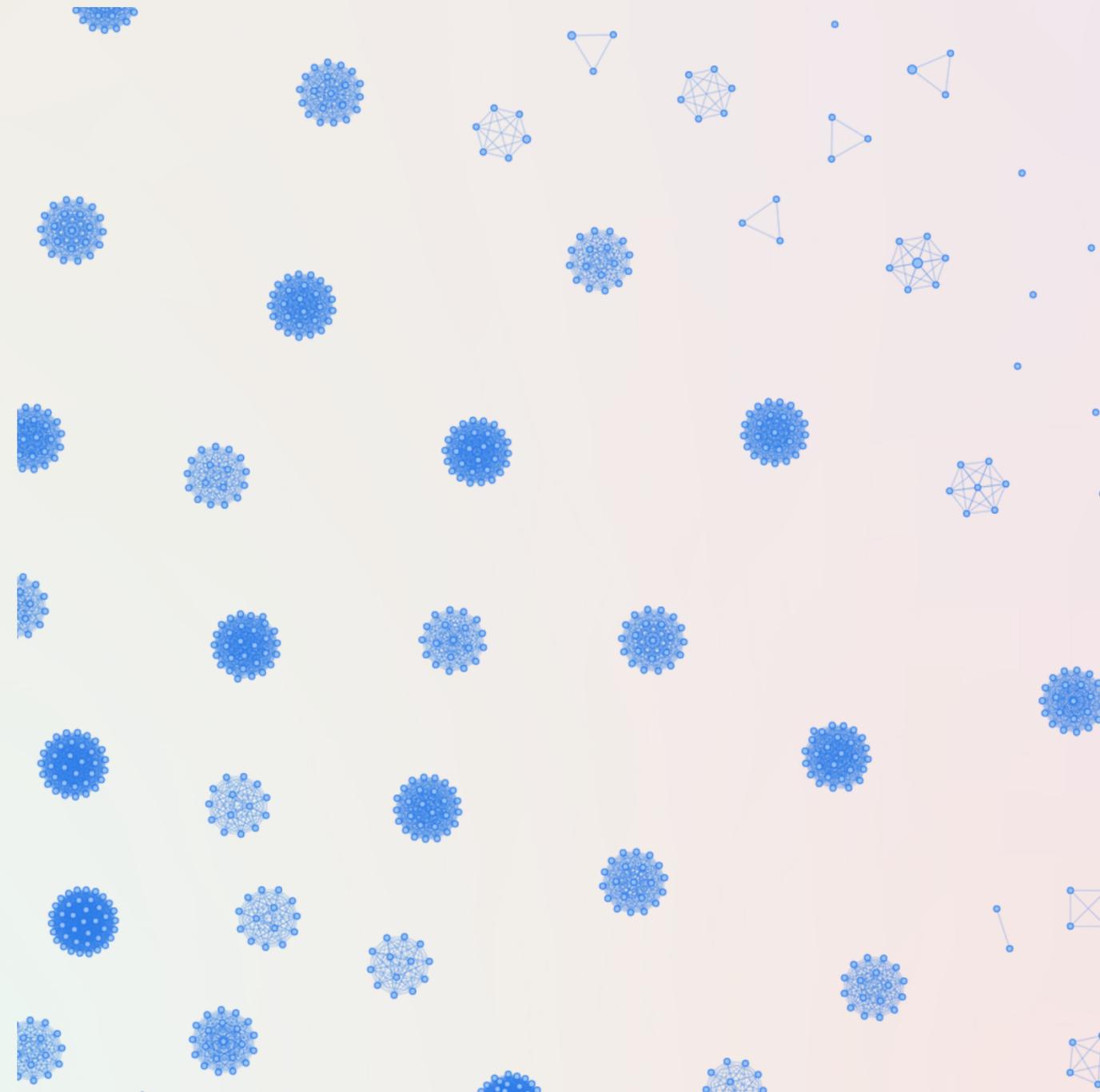
Алгоритм кластеризации UMI был основан на задании направления в графе, ребро направлено в сторону вершины соответствующей последовательности UMI, прочитанной большее количество раз. Для этого вершинам графа приписывались веса равные числу прочтений UMI. "Схлопывание" кластеров происходило в сторону вершины с наибольшим весом.

Критерий отличия исходно различных последовательностей от последовательностей с накопленными ошибками был задан в виде минимального веса центра кластера равного 30.

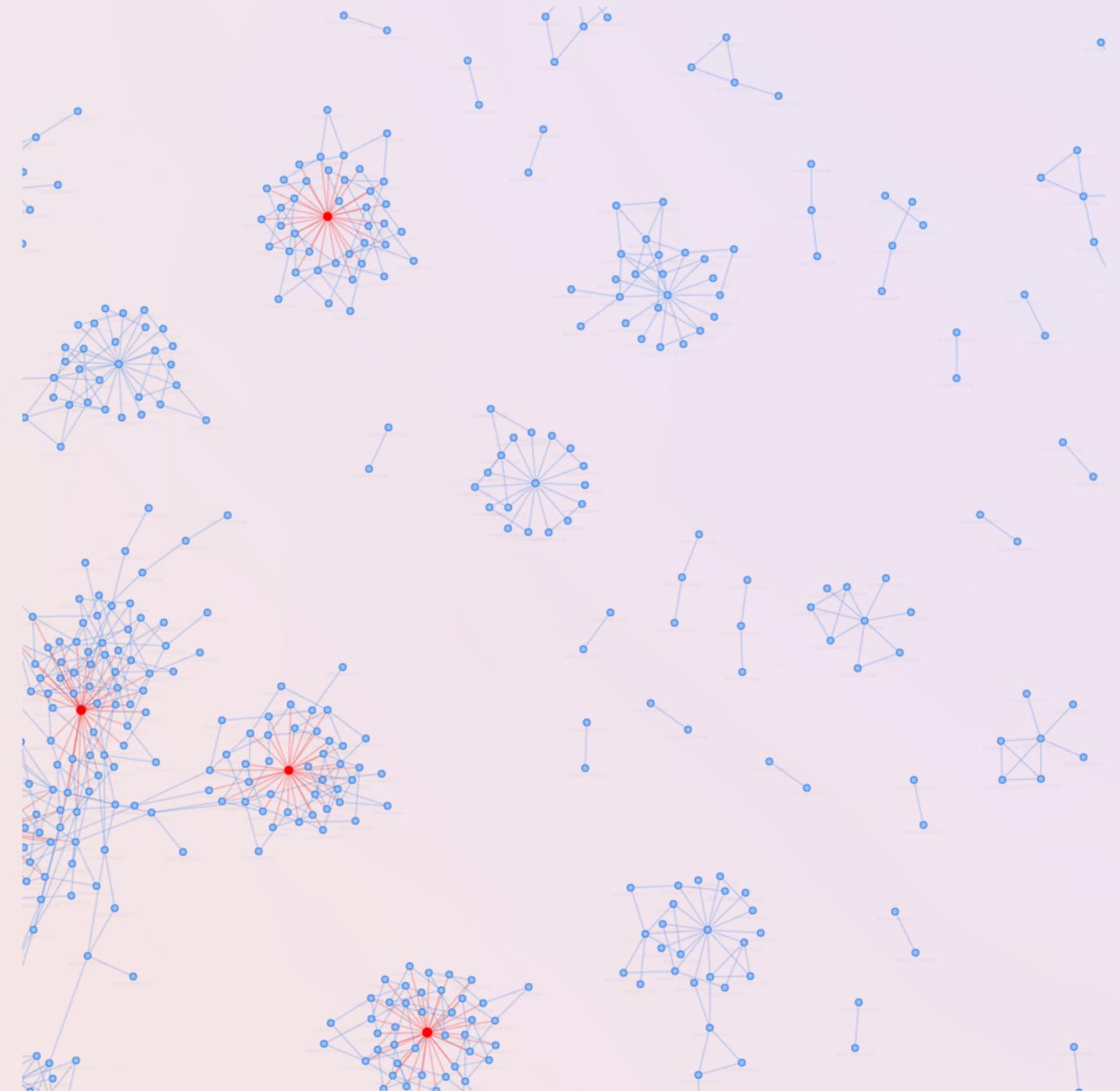


Граф, визуализирующий пространство UMI для 17 биологического образца после кластеризации

КЛАСТЕРИЗАЦИЯ САЙТОВ ИНТЕГРАЦИИ



УСОВЕРШЕНСТВОВАНИЕ ПРОЦЕДУРЫ КЛАСТЕРИЗАЦИИ



РЕЗУЛЬТАТЫ И ВЫВОДЫ

Был разработан алгоритм анализа данных

После уточнений процедуры фильтрации и удаления контаминаций и невалидных данных было выявлено слишком мало сайтов интеграции - до 1000 на образцы из косточек и еще меньше для бразцов из внутренней клеточной массы. Коллеги сделали цифровую капельную ПЦР и КОЕф полученных из очагов кроветворения и ее результаты показывают, что промаркированы лишь единичные клетки.

Ответить на поставленные вопросы:

1. сколько МСК строит очаг и экстраполировать это на концентрацию МСК в костном мозге мыши.
2. Сделать выводы о дифференцировочном потенциале клеток, строящих очаг, сравнивая сайты интеграции в клетках внутренней массы и косточки очага.

не представилось возможным, так как в проведенном эксперименте не удалось эффективно маркировать МСК.

После получения предварительных результатов была сформулирована еще одна цель:

1. Сравнение клеточных клонов в облученных и необлученных реципиентах: посмотреть соотношение umi/is для необл и обл (ожидается выше у обл)
2. Оценка вклада нестволовых клеток в формирование кроветворной территории de novo.