

## **Воркшоп в рамках дисциплины "Прикладная биоинформатика: 3С-технологии".**

### **Отчет по анализу данных Hi-C**

#### **Введение**

Hi-C (High-throughput Chromosome Conformation Capture) позволяет изучать пространственную организацию хромосом с более высоким разрешением и в более крупном масштабе.

Этот отчет предоставляет детальное описание процесса анализа данных Hi-C, выполненного в рамках изучаемой дисциплины. В процессе работы использовались различные инструменты и пакеты для подготовки, обработки и визуализации данных. Основной целью было воспроизведение анализа, описанного в руководствах и документации, а также представление результатов в виде контактных матриц и их дальнейшего анализа.

#### **Настройка окружения**

Для обеспечения изолированного и воспроизводимого анализа на локальной машине был установлен Docker. Использовался Docker-образ `duplexa/4dn-hic:v42`, включающий все необходимые инструменты и зависимости для работы с данными Hi-C.

Также была установлена Conda для управления пакетами и создания виртуального окружения. В данном окружении были установлены необходимые пакеты, включая Jupyter Notebook и Hiclass-manager.

#### **Получение и подготовка данных**

Данные были загружены в домашний каталог `~/data/` для дальнейшей обработки. Для создания файла формата `.cool` использовался инструмент `cooler`, который агрегировал список пар ридов.

#### **Анализ данных**

##### **Построение контактной матрицы**

Геном был разбит на бины фиксированного размера 1 Мб с помощью команды `cooler makebins`.

Построена контактная матрица для данных GM12878 на разрешении 1 Мб с помощью команды `cooler cload` (агрегирует парные чтения и записывает их в формат `.cool`).

##### **Нормализация матрицы**

Для нормализации контактной матрицы использовался инструмент `cooler balance`, что позволило достичь равномерных сумм строк и столбцов в глобальной контактной матрице.

##### **Визуализация данных**

Для визуализации контактных матриц использовался `cooler show`, который позволяет отображать контактные матрицы с использованием библиотеки `matplotlib`.

В итоге результаты представлены в виде интерактивных контактных карт, позволяющих исследовать пространственную организацию генома.

## Описание результатов визуализации в HiGlass

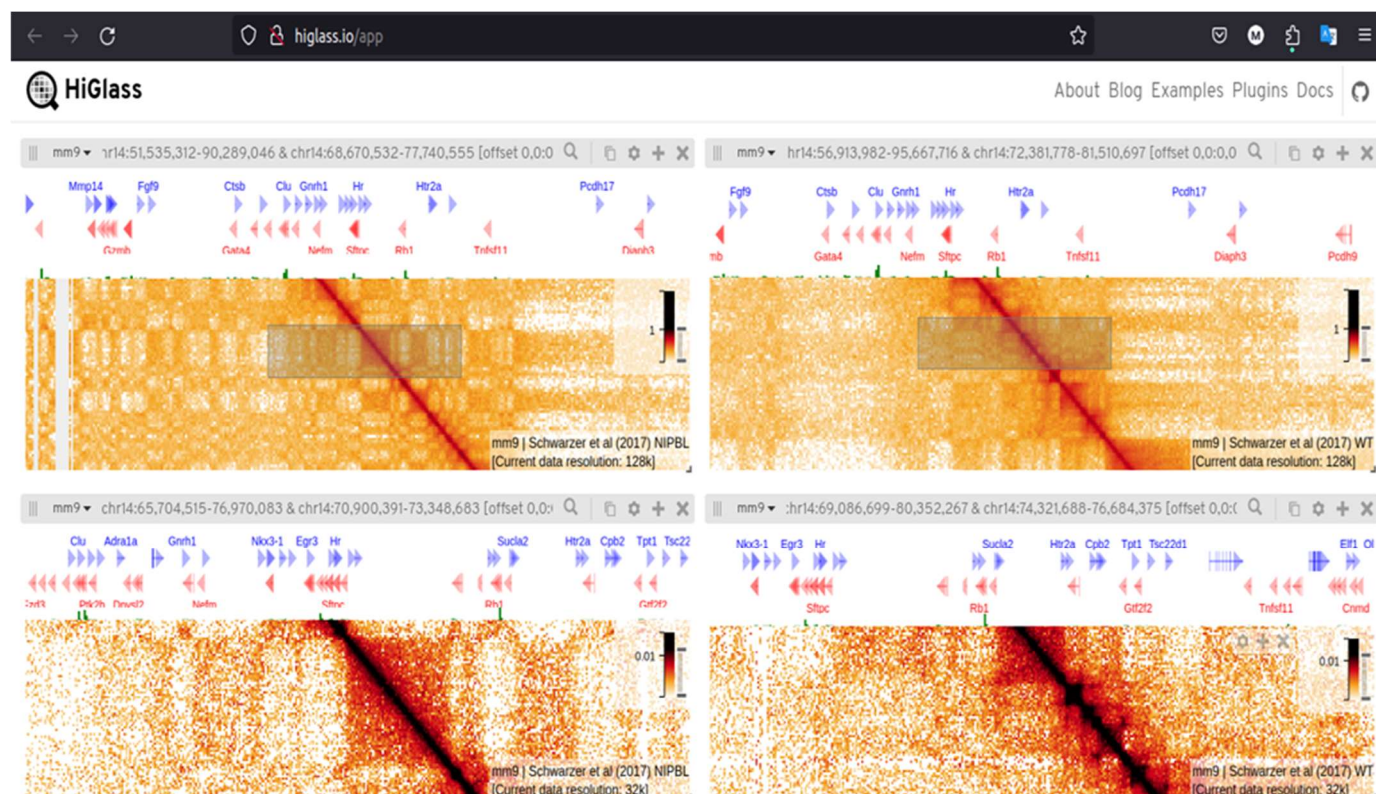


Рисунок №1. Результаты визуализации в HiGlass  
(<http://higlass.io/l/?d=cCq4OQjLSmi8hholmLFssA>)

На предоставленном рисунке №1 показаны результаты визуализации данных Hi-C с использованием HiGlass. HiGlass является мощным инструментом для визуализации и анализа данных геномных контактов, таких как данные Hi-C.

Изображение показывает различные участки контактной матрицы генома мыши (mm9) на разных уровнях разрешения.

### Области контактных матриц:

Верхние две панели показывают контактные матрицы для двух различных областей хромосомы с использованием данных Schwarzer et al. (2017) NIPBL и WT при разрешении 128k.

Нижние две панели показывают контактные матрицы для этих же областей хромосомы на более низком разрешении (32k).

### Геномная аннотация:

Над каждой матрицей отображены гены и другие важные элементы генома. Синие и красные стрелки указывают направление генов.

Цветовая шкала указывает на плотность контактов: темно-красные области указывают на высокую плотность контактов (больше взаимодействий между этими регионами), тогда как светло-желтые и белые области указывают на низкую плотность контактов.

Диагональные линии на контактных картах указывают на внутривнутрихромосомные взаимодействия.

Контактные карты позволяют визуализировать и анализировать взаимодействия между различными участками генома.

## Заключение

В данном отчете был выполнен анализ данных Hi-C с использованием инструментов и методов, описанных в руководствах и документации. Были установлены и настроены необходимые окружения, выполнена загрузка и подготовка данных, построение и нормализация контактных матриц, а также их визуализация. Все этапы анализа выполнены с использованием контейнеризации для обеспечения воспроизводимости результатов.