



Московский государственный университет имени М.В.Ломоносова
Биологический факультет
Кафедра биоинженерии
Группа интегративной биологии



Выпускная квалификационная работа
бакалавра

Разработка подходов интегративного моделирования структуры и динамики нуклеосомных фибрилл на основе анализа экспериментальных данных 3D-геномики субнуклеосомного разрешения

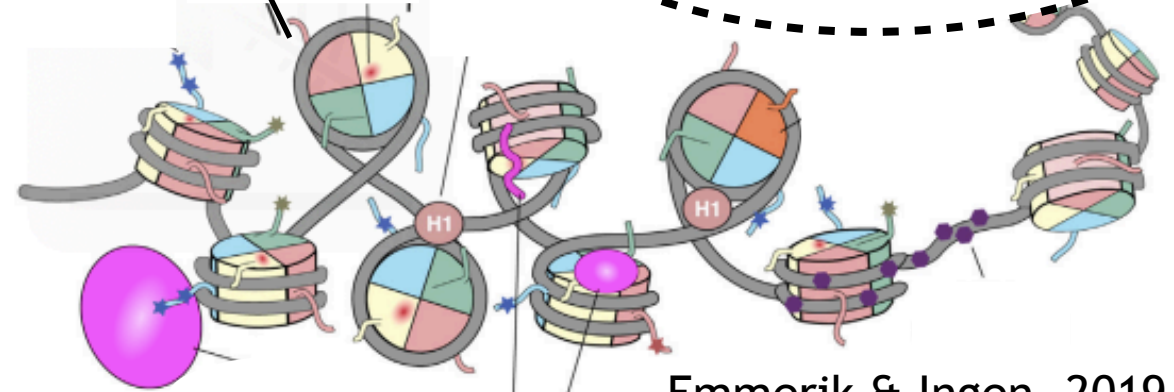
Тимохин Григорий Сергеевич,
студент 426 группы
Научный руководитель:
д. ф.-м. н. Шайтан Алексей Константинович.

Москва
2021

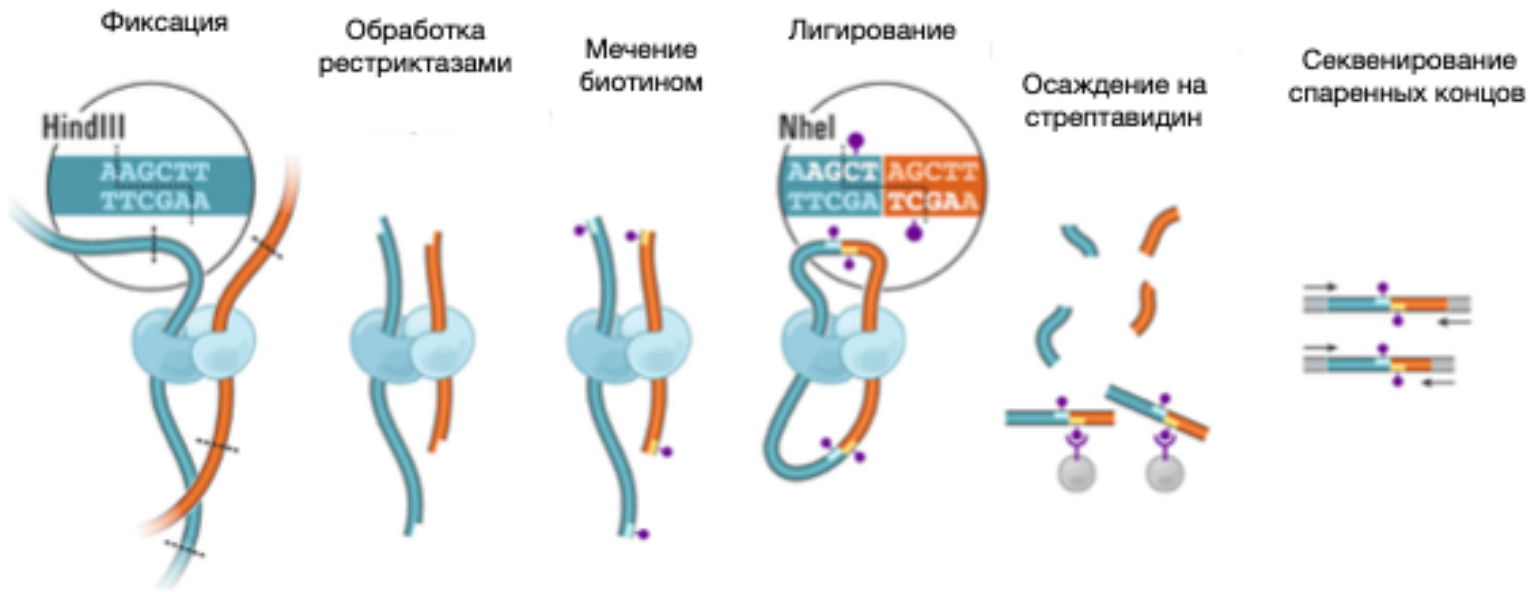
Уровни организации хроматина

Хромосомные
территории

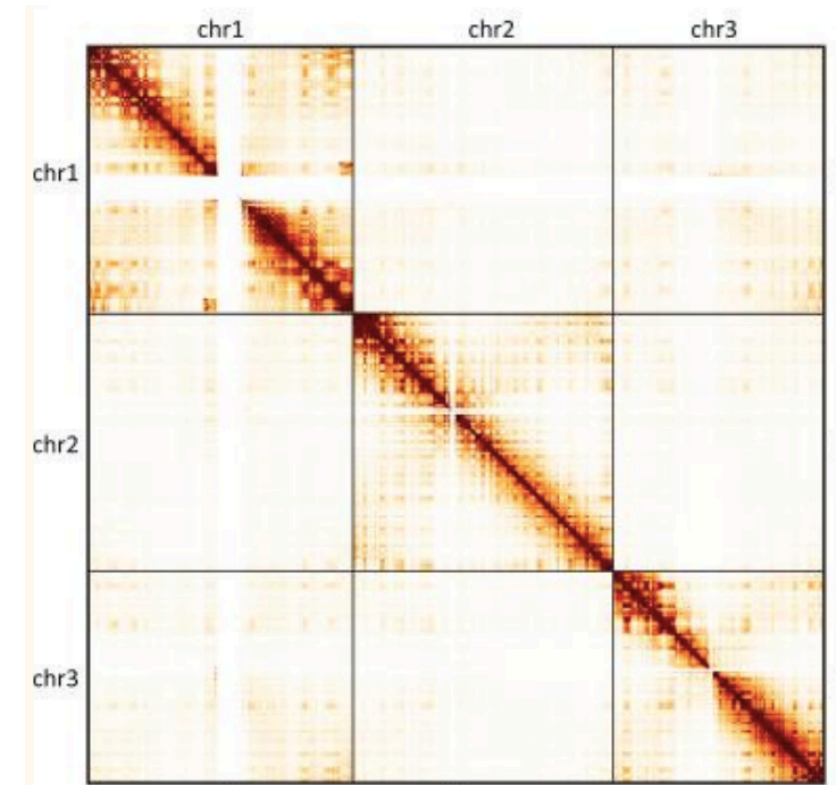
Нуклеосомная
фибрилла



Метод Hi-C

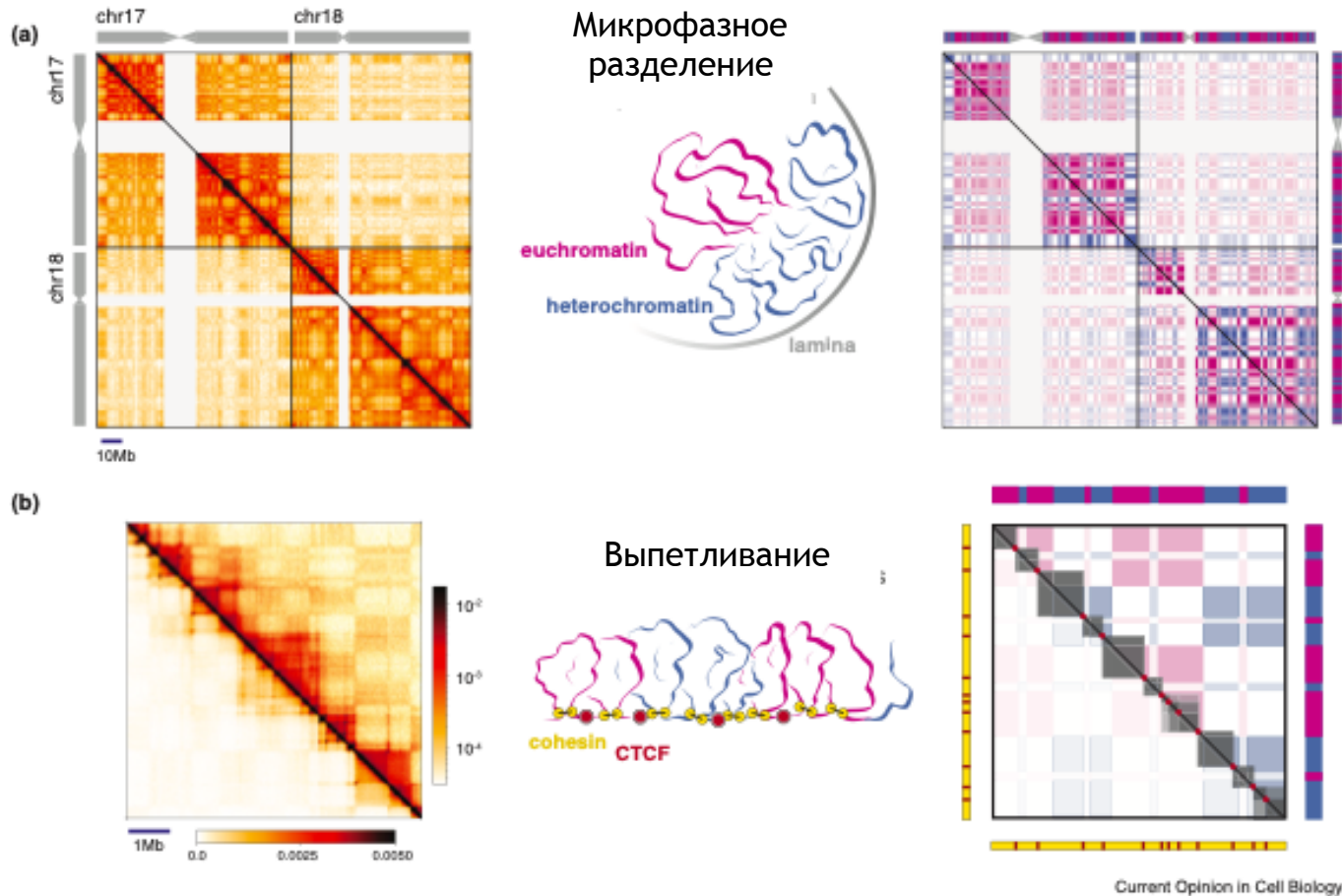


Lieberman-Aiden et al., 2009

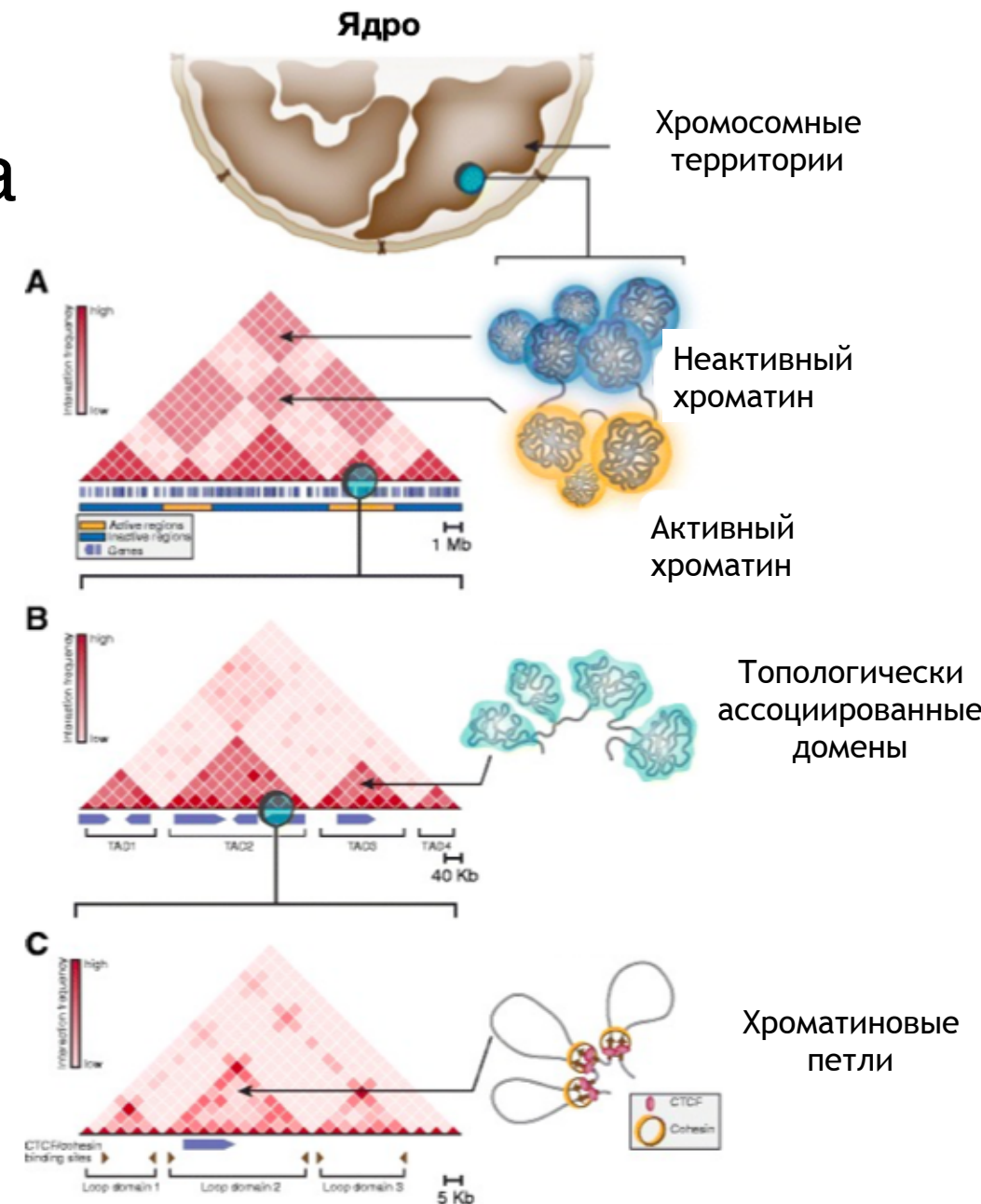


Lajoie et al., 2014

Контактные карты Hi-C и структурная организация хроматина

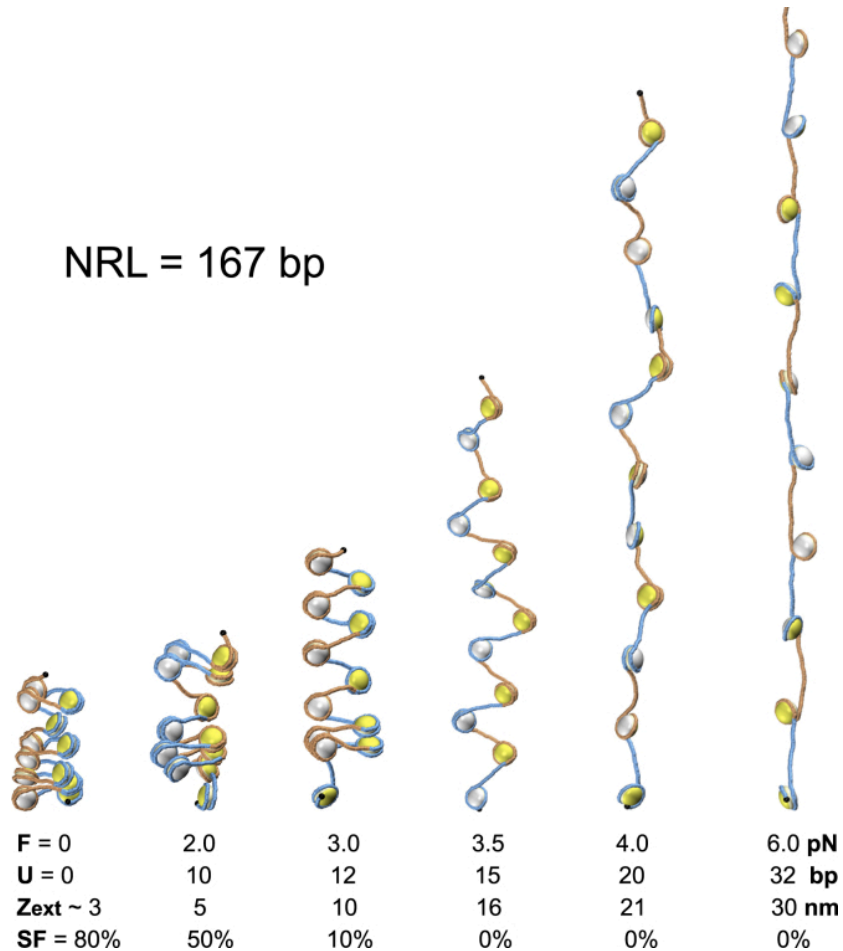


Mirny et al., 2019

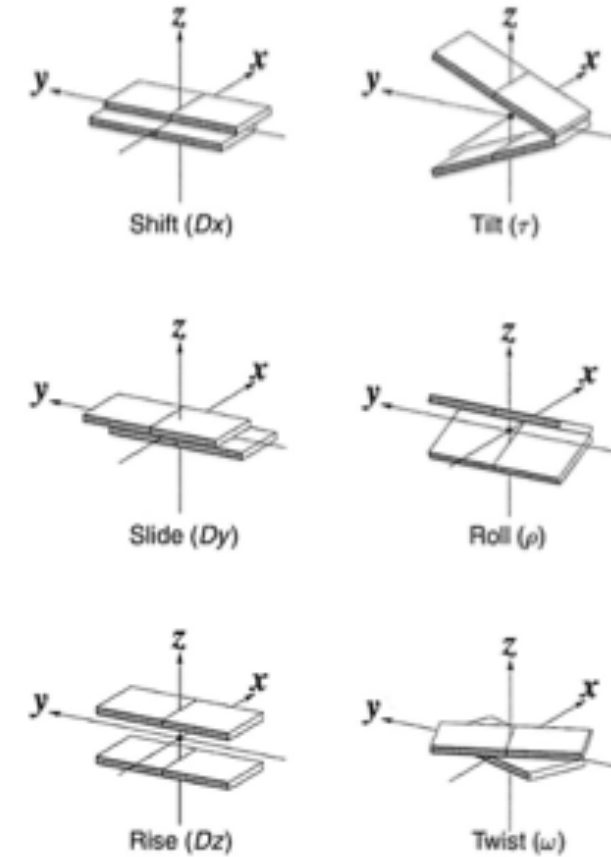


Razin et al., 2017

Методы моделирования супрануклеосомной структуры хроматина

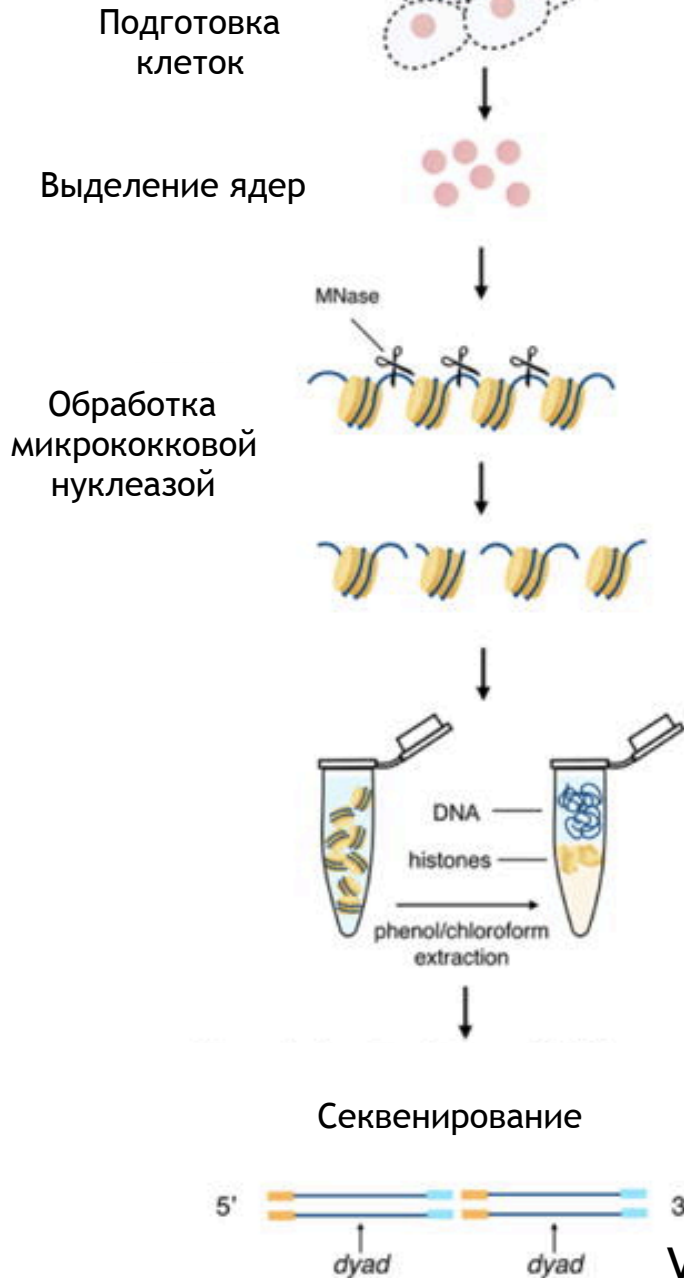


Norouzi & Zhurkin, 2018

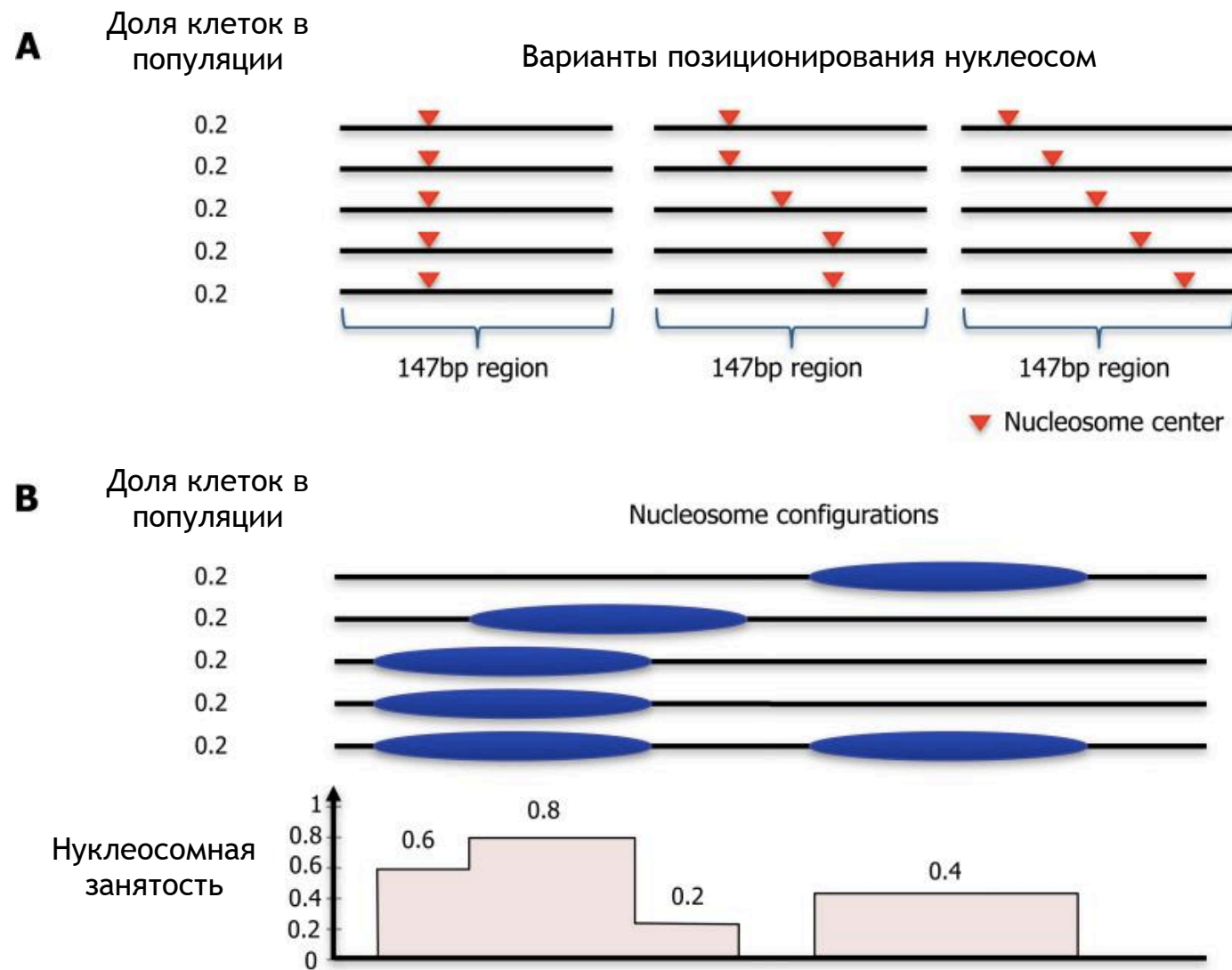


Xiang-Jun Lu et al., 2003

Метод MNase-seq

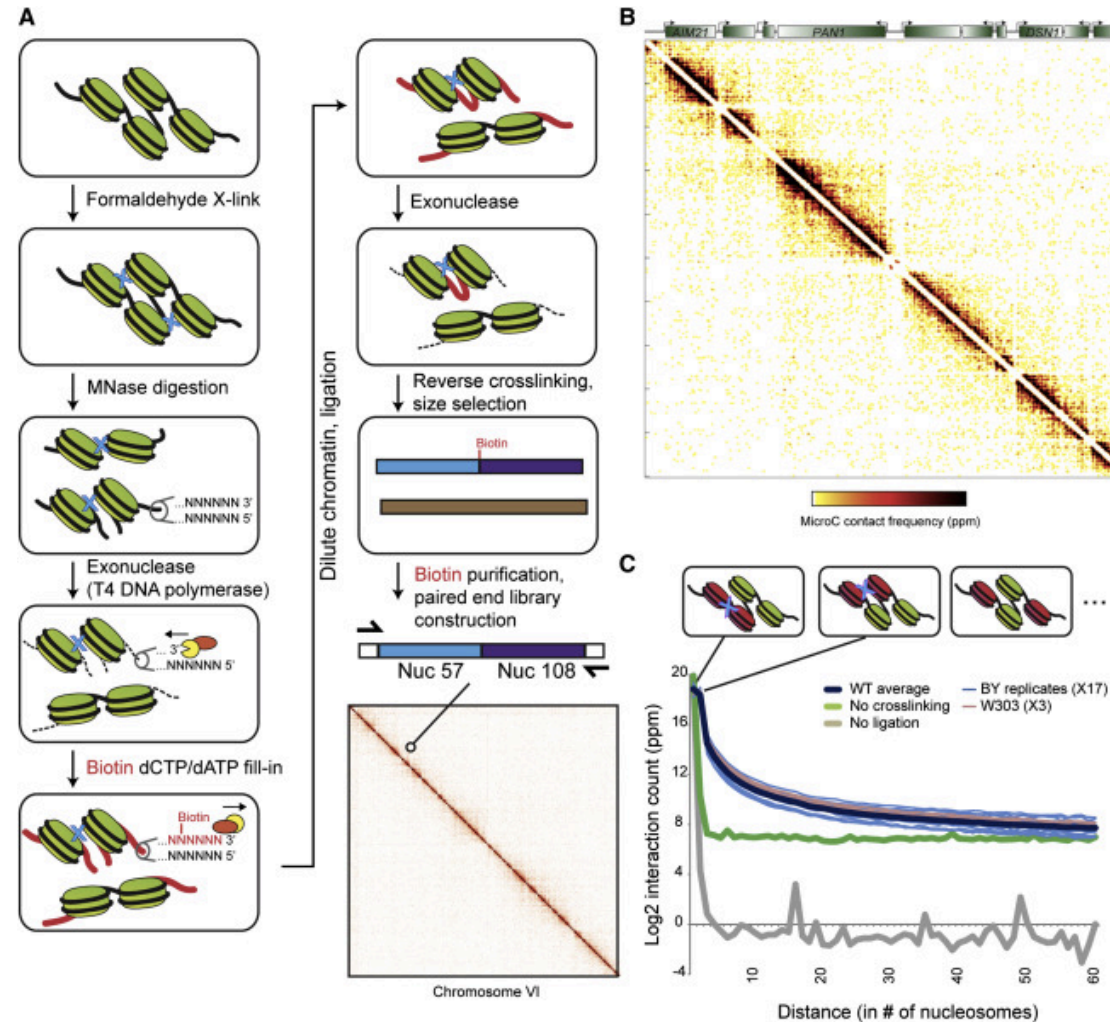


Voong et al., 2017



Struhl et al., 2013

Micro-C как метод изучения супрануклеосомной структуры хроматина



Цели и задачи

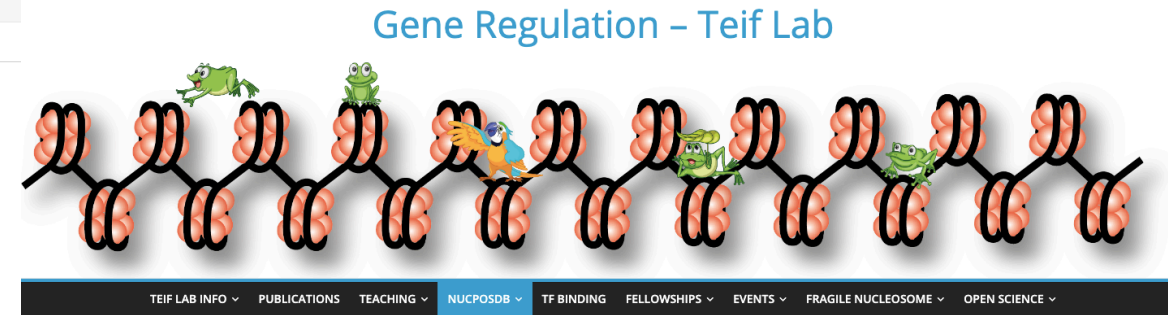
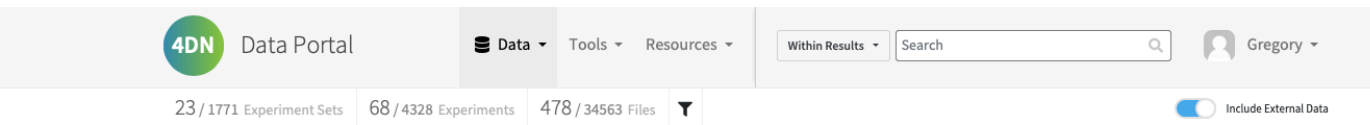
Цель работы - разработка и тестирование методов реконструкции супрануклеосомной структуры и динамики хроматина на основе экспериментальных данных 3D-геномики субнуклеосомного разрешения.

Задачи:

1. Разработать и настроить программные конвейеры для обработки, анализа и интеграции экспериментальных данных Micro-C и MNase-seq
2. С использованием методов огрубленного моделирования ДНК разработать подходы моделирования супрануклеосомной структуры хроматина на основе интегрированных данных
3. С помощью разработанных методов проанализировать супрануклеосомную структуру хроматина в локусе Igf2-H19 для различных клеточных линий *M.musculus* и *H.sapiens*

Методы. Используемые базы данных.

Выбранные для анализа экспериментальные данные.



Nucleosome positioning maps *in vivo*



Below is the repository of nucleosome positioning datasets *in vivo*. This collection is part of NucPosDB, a manually curated database of experimental nucleosome maps *in vivo*, cDNA and computational tools related to nucleosome positioning

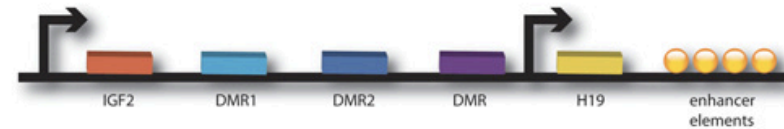
Для тестирования разработанного программного обеспечения использовались данные Micro-C и MNase-seq для клеточных линий *H.sapiens*: H1-hESC, HeLa-S3 и *M.musculus*: JM8.N4 (mESC)

Для получения Micro-C данных использовался репозиторий 4DNucleome
Для получения MNase-seq данных — репозиторий NucPosDB (агрегатор ссылок на GEO)
Также в работе использовались GEO, Ensembl

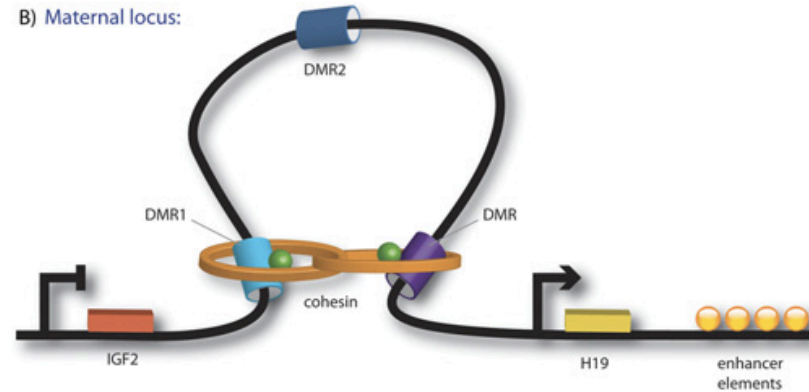
Локус Igf2-H19

Геномные координаты локуса:
M.musculus – 7:142,460,685-142,720,685
H.sapiens – 11:1,911,877-2,181,877

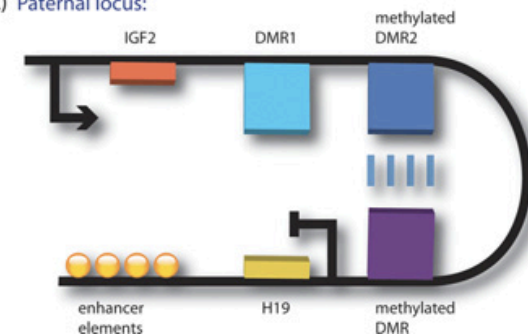
A) Structure of the locus:



B) Maternal locus:



C) Paternal locus:



Методы. Использованное программное обеспечение

Все вычисления велись на кластере Newton

Программное обеспечение разрабатывалось в интегрированной среде разработки JupyterLab

Использованные языки программирования: python 3, R, bash

Использованные для обработки Micro-C данных биоинформатические библиотеки: pairtools, cooltools, HICExplorer, HICcompare, GenomicInteractions

Для молекулярного моделирования использовалась библиотека PyNaMod, разрабатываемая в группе интегративной биологии — в библиотеке используются принципы Монте-Карло и динуклеотидное приближение ДНК

Результаты. Программное обеспечение, разработанное для обработки и интеграции данных MNase-seq и Micro-C

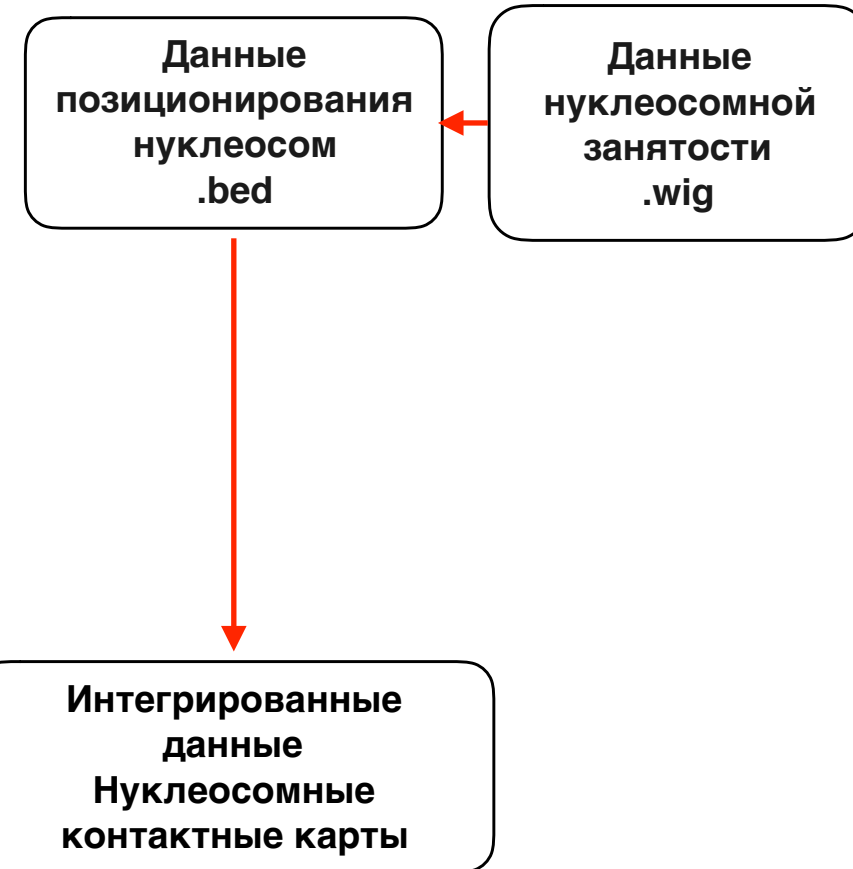
В ходе выполнения работы было:

- 1) Настроено ПО для обработки и анализа Micro-C данных
- 2) Разработано ПО для конвертирования данных нуклеосомной занятости в данные позиционирования нуклеосом (путем выявления пиков и нахождения центров их оснований)
- 3) Разработано ПО для интеграции данных позиционирования нуклеосом и данных Micro-C — с получением в результате нуклеосомных контактных карт (путем картирования центров нуклеосом на оси контактных карт)

Micro-C данные



MNase-seq данные



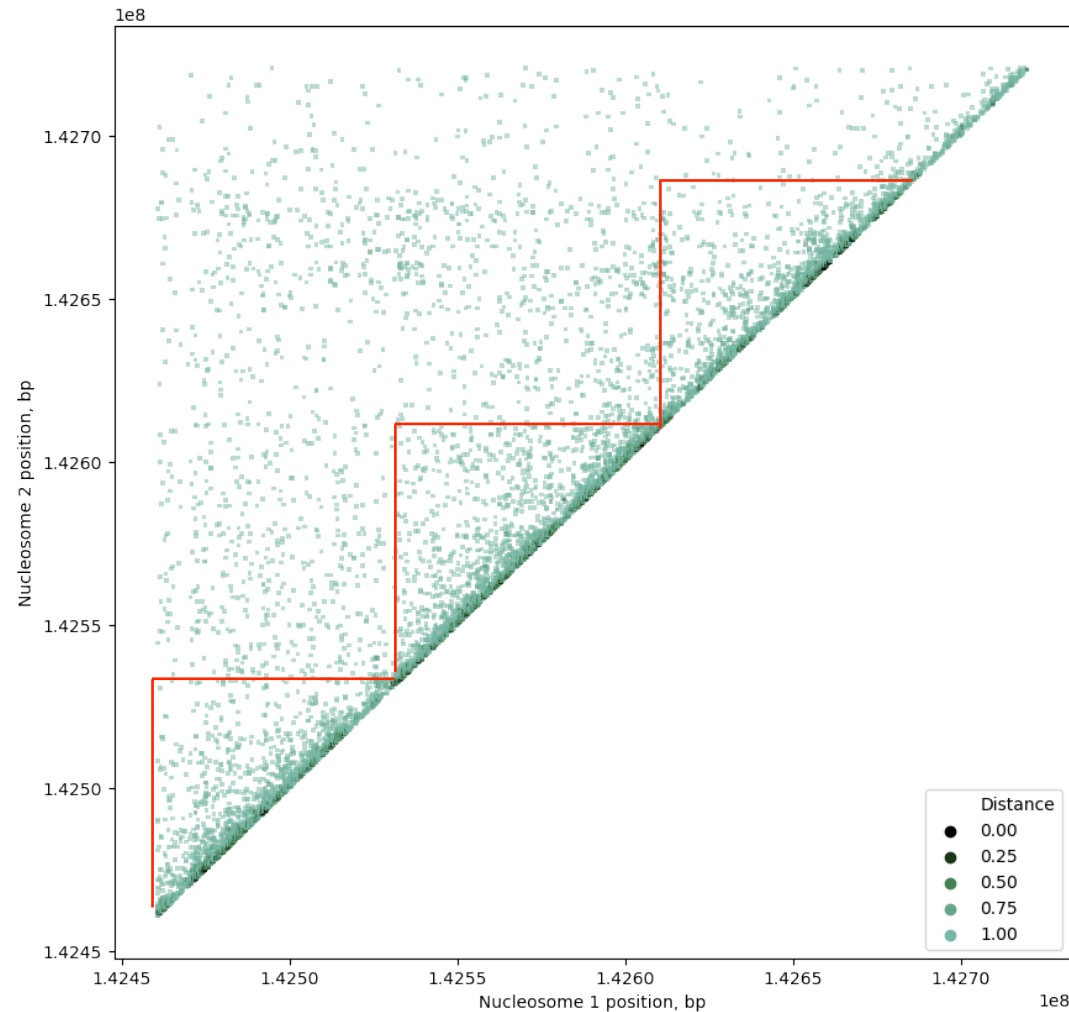
Результаты. Разработанные подходы моделирования на основании интегрированных данных

В ходе работы было:

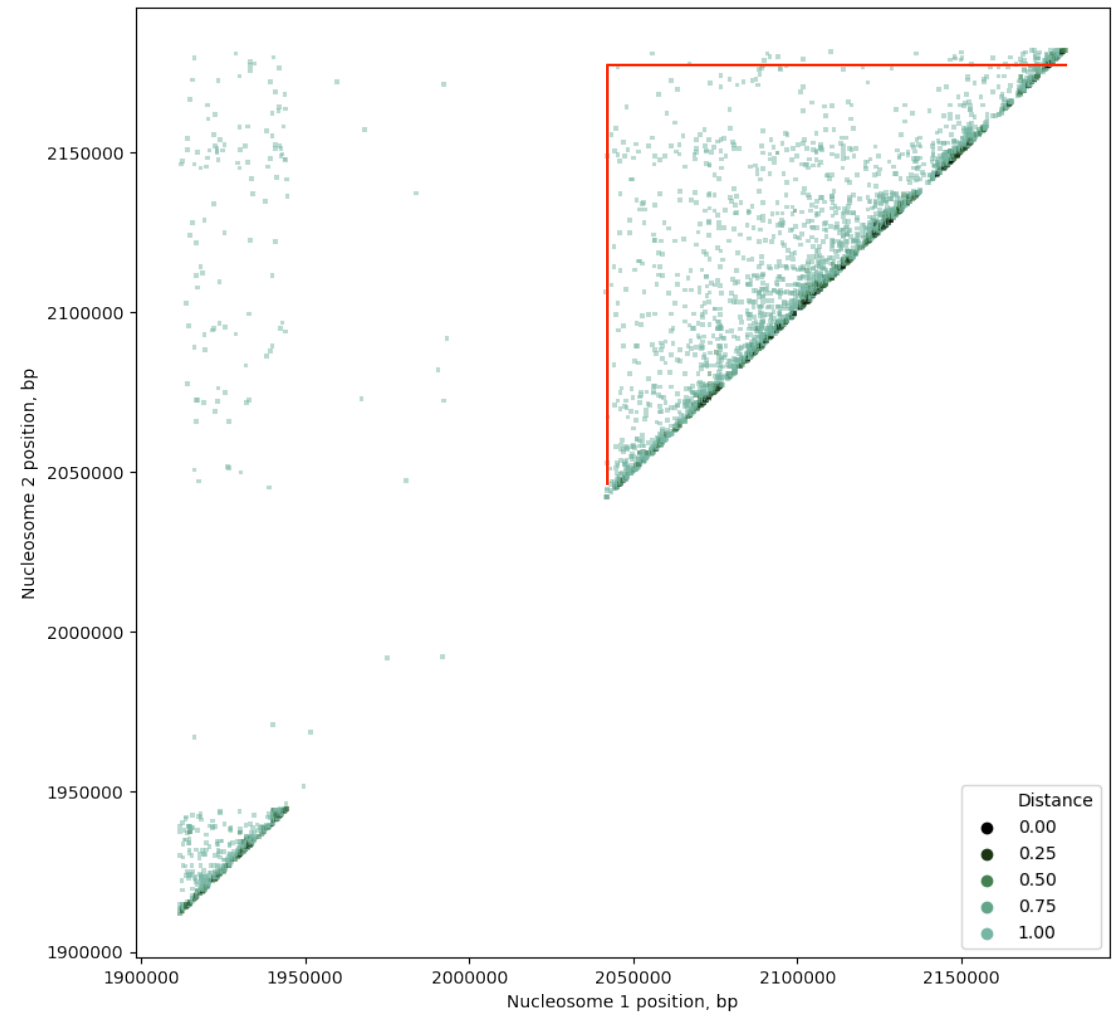
- 1) С использованием основанных на расстоянии (distance based) принципов перевода из данных о контактах в данные о расстоянии настроено ПО, позволяющее проводить с помощью библиотеки PyNaMod моделирование на основании интегрированных нуклеосомных карт
- 2) Настроено ПО для моделирования помощью библиотеки PyNaMod супрануклеосомной структуры хроматина



Результаты анализа экспериментальных данных. Реконструкция супрануклеосомной структуры хроматина



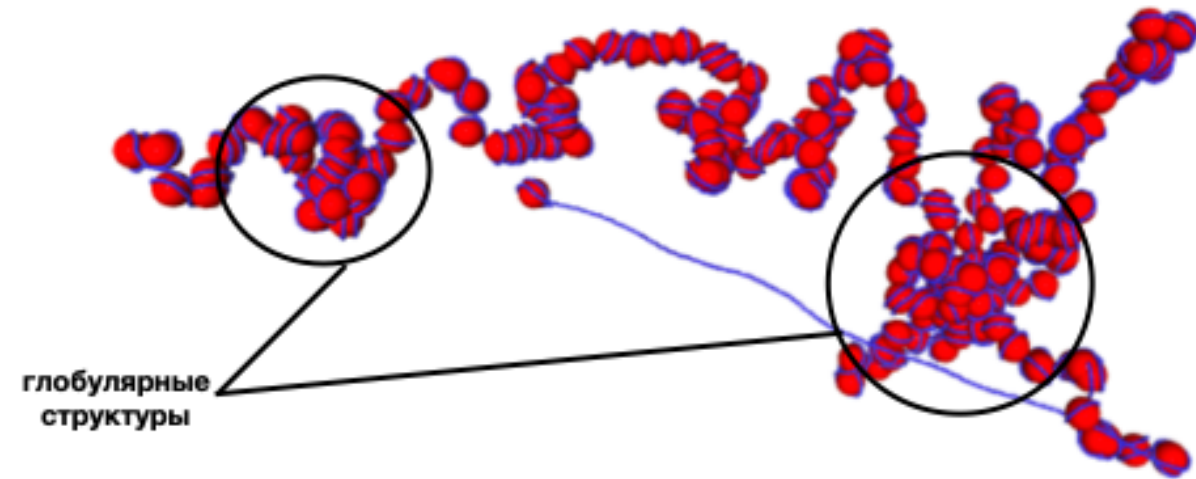
Контактная нуклеосомная карта локуса Igf2-H19
M.musculus, клеточная линия JM8.N4 (mESC)
Визуально выявляются ТАД-подобные паттерны



Контактная нуклеосомная карта локуса Igf2-H19
H.sapiens, клеточная линия H1 (hESC)
Визуально выявляются ТАД-подобные паттерны

Результаты анализа экспериментальных данных. Реконструкция супрануклеосомной структуры хроматина

В ходе работы был реконструирован участок супрануклеосомной структуры хроматина с помощью библиотеки PyNaMod — реконструированный участок представляет собой периодически образующую глобулярные структуры нерегулярно организованную фибриллу



Реконструкция супрануклеосомной структуры хроматина в участке локуса Igf2-H19 *M.musculus*, клеточная линия JM8.N4 (mESC)

Выводы

1. В ходе работы были разработаны и настроены программные конвейеры, обрабатывающие данные Micro-C и интегрирующие их с данными MNase-seq для получения нуклеосомных контактных карт
2. Были созданы протоколы моделирования, позволяющие на основе подходов Монте-Карло и моделирования ДНК в динуклеотидном приближении моделировать на основании нуклеосомных карт супрануклеосомную структуру хроматина
3. На основании данных Micro-C и MNase-seq для клеточных линий *H.sapiens*: H1-hESC, HeLa-S3 и *M.musculus*: JM8.N4 (mESC) были вычислены нуклеосомные контактные карты локуса Igf2-H19
4. В ходе визуального анализа полученных нуклеосомных карт выявлено наличие ТАД-подобных паттернов
5. Сравнительный анализ полученных нуклеосомных контактных карт локуса Igf2-H19 для клеточных линий H1-hESC, HeLa-S показал их сходство
6. Сравнительный визуальный анализ полученных нуклеосомных контактных карт локуса Igf2-H19 *H.sapiens* и *M.musculus* показал отсутствие значимого сходства между ними
7. Реконструкция участка супрануклеосомной структуры хроматина в локусе Igf2-H19 *M.musculus* представляет собой нерегулярно организованную, локально образующую глобулярные структуры, фибриллу