

Соотнесение устойчивых сочетаний геномных и эпигенетических маркеров с границами топологически ассоциированных доменов хроматина

Сидоров Святослав Игоревич

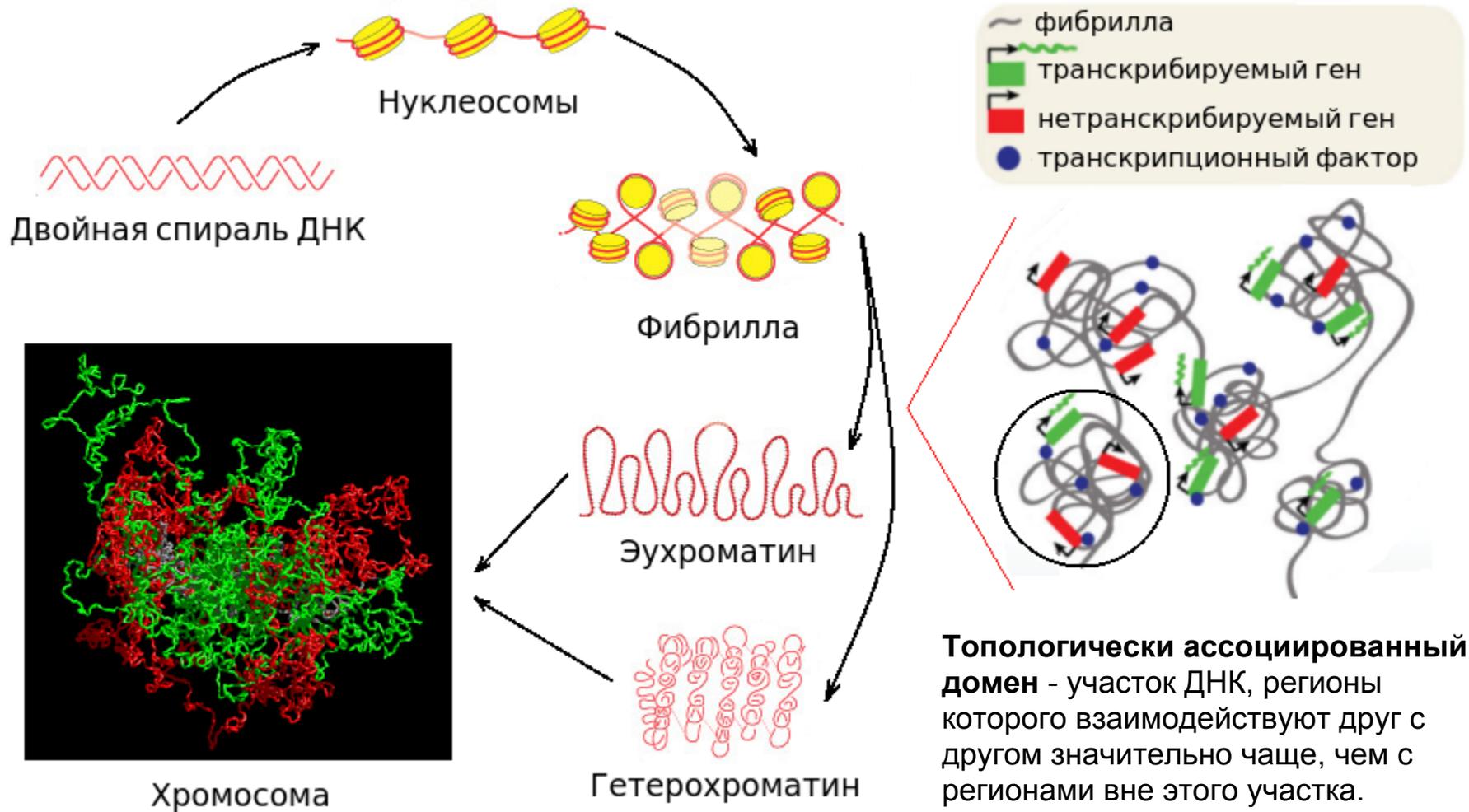
Научный руководитель:

Предеус Александр Владимирович, PhD, постдок

Санкт-Петербургский Академический Университет

2015

Топологически ассоциированные домены (ТАДы)

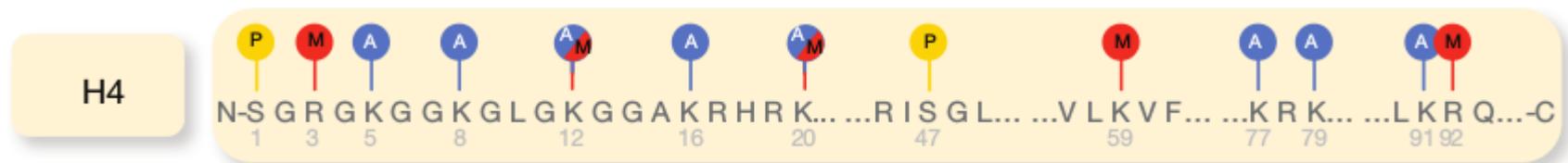


Геномные и эпигенетические маркеры

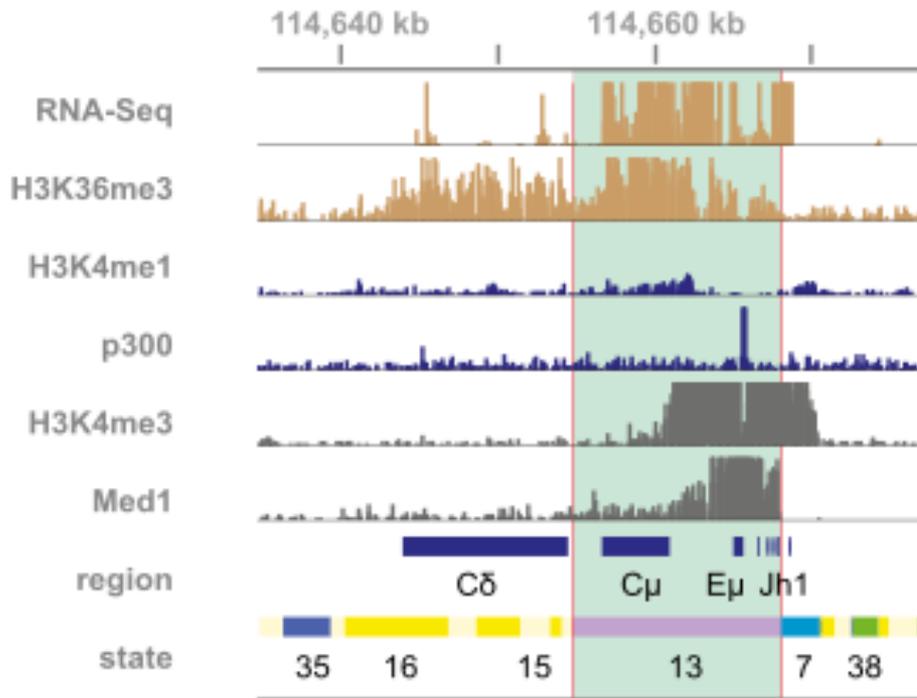
Геномные маркеры - участки связывания белков.



Эпигенетические маркеры - эпигенетические модификации гистонов.



Устойчивые сочетания маркеров



Маркеры

Устойчивые сочетания маркеров (состояния СММ)

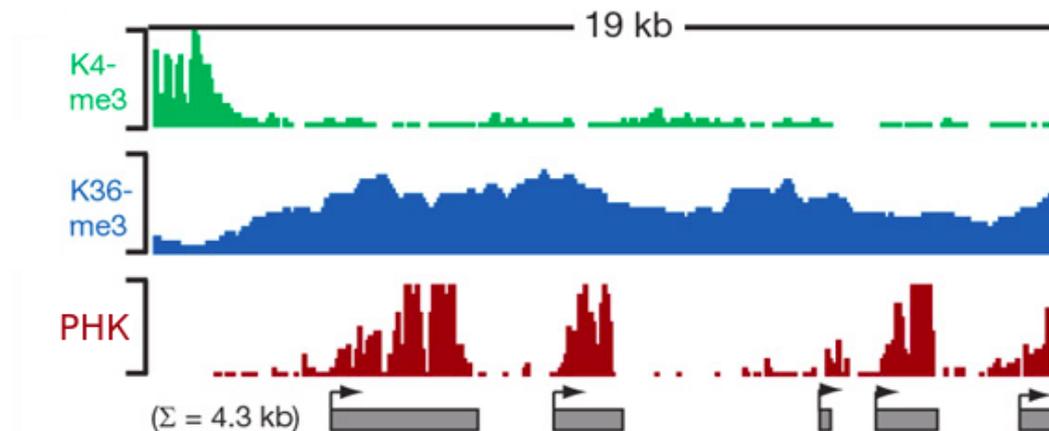
Цель работы

Сопоставить границы ТАДов с устойчивыми сочетаниями маркеров для заданных клеточных линий, разработав и применив соответствующее программное обеспечение с использованием существующих технологий.

Актуальность работы

Указанное сопоставление в будущем позволит аннотировать границы между ТАДами *de novo* (определять наличие в них определенных элементов генома).

Пример аннотации, использующей устойчивые сочетания маркеров: открытие большого разнообразия длинных некодирующих РНК (lincRNA):



Задачи

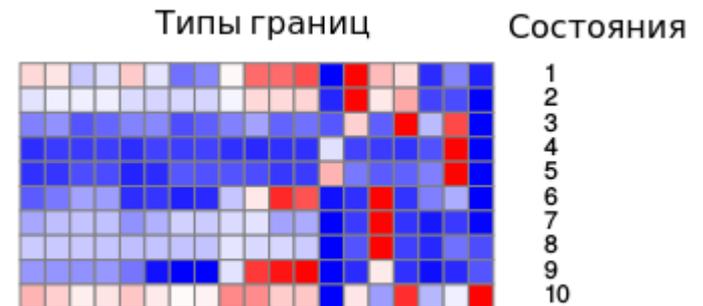
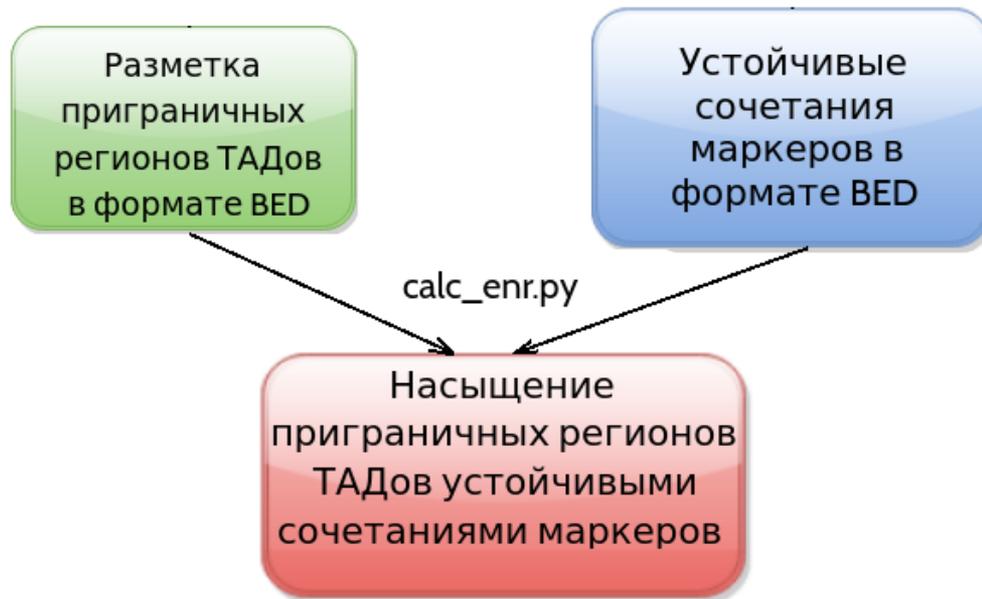
1. Автоматизировать выделение границ ТАДов по данным Hi-C и нахождение устойчивых сочетаний маркеров по данным ChIP-Seq.
2. Разработать и реализовать метод сопоставления границ ТАДов с устойчивыми сочетаниями маркеров.
3. Использовать разработанные средства для обработки данных по различным типам клеток.

Обработка данных ChIP-Seq



Использованные технологии: Python, bowtie, SAMtools, IGVtools, BED-Tools, SICER, MACS, ChromHMM.

Заключительный этап обработки



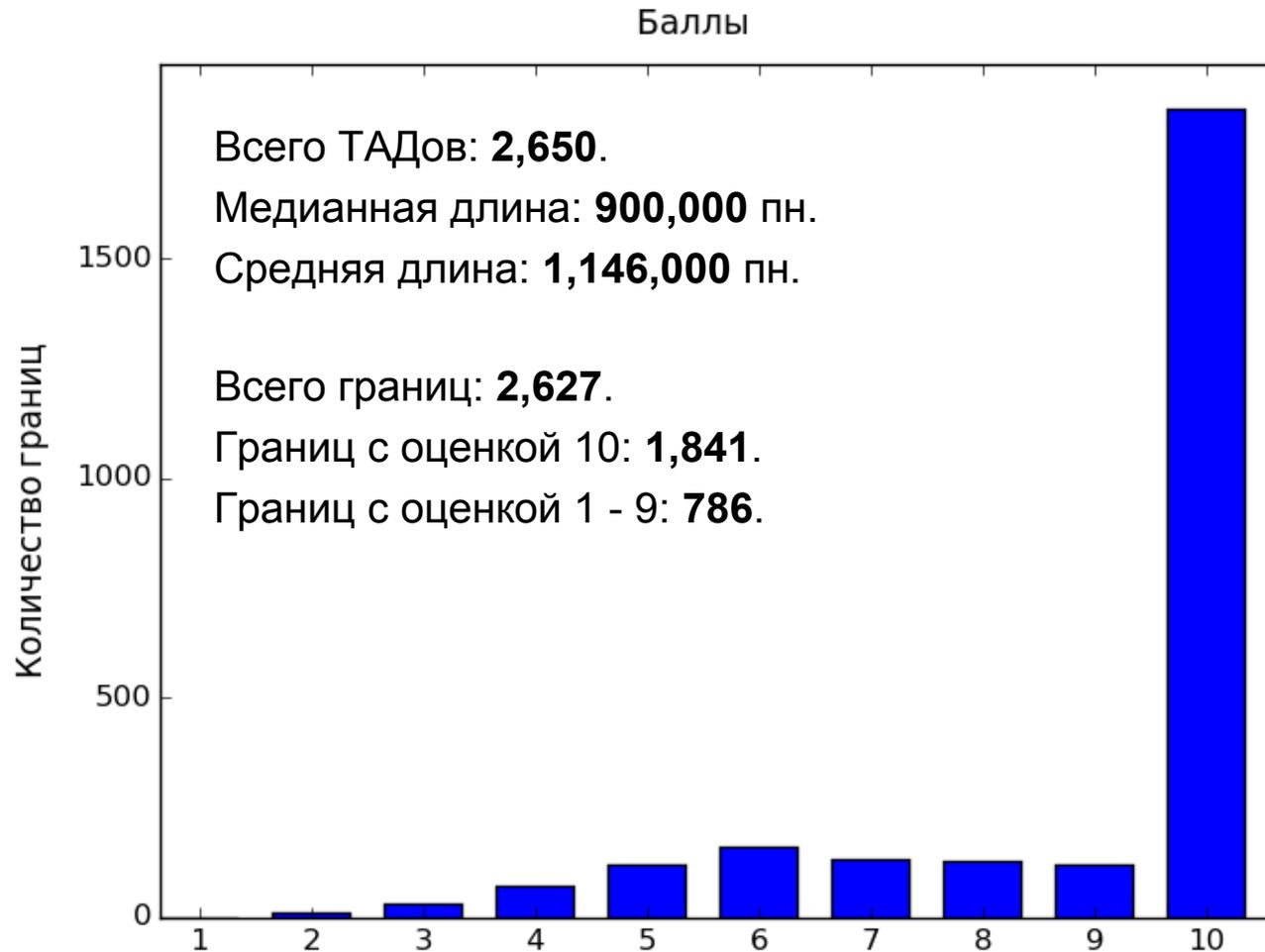
Скрипты доступны в репозитории <https://github.com/sidorov-si/TADStates> (лицензия GPL версии 2 и выше).

Данные

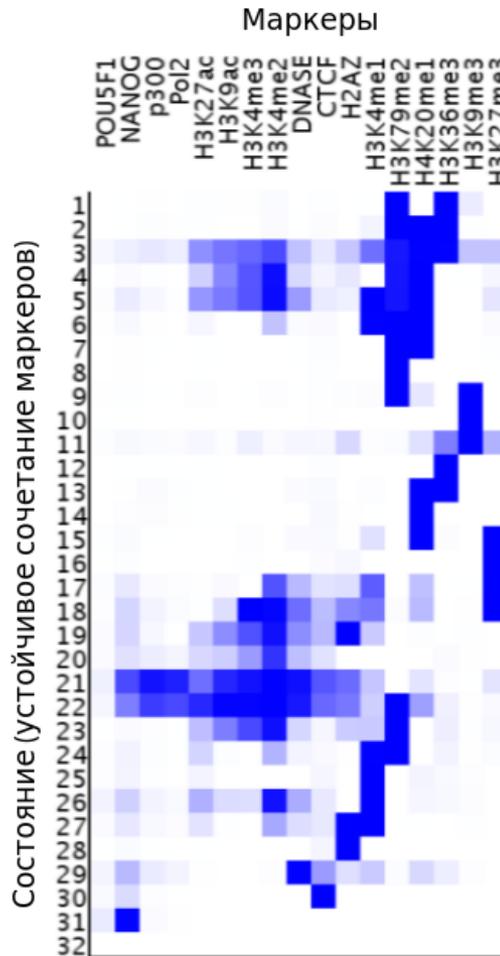
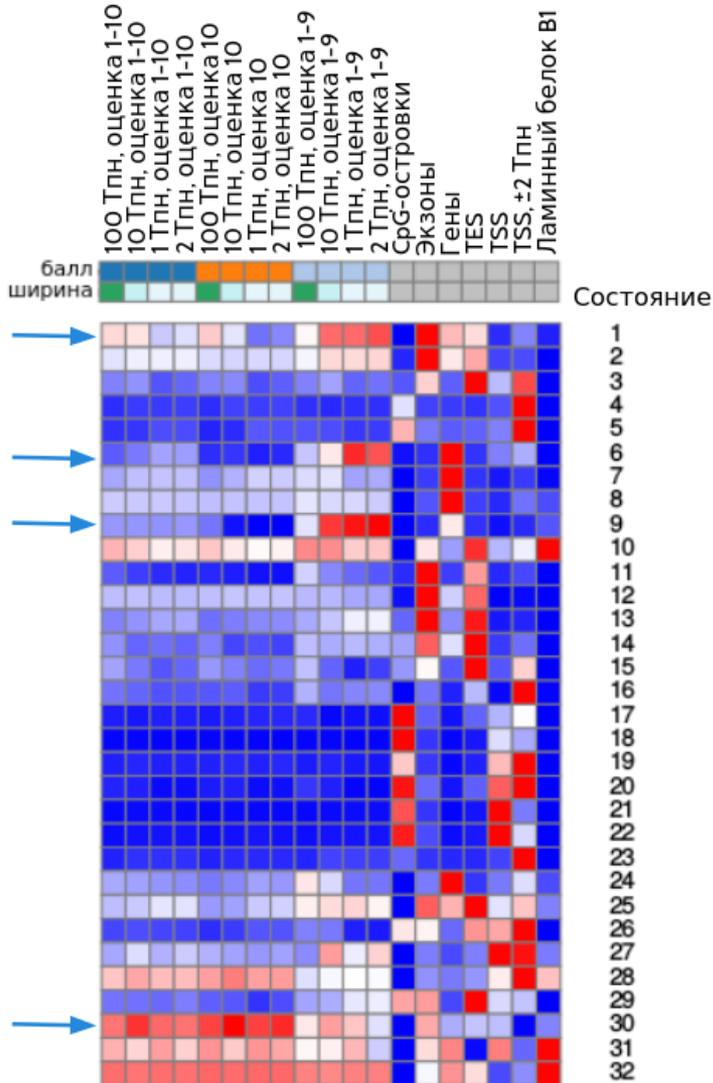
Клеточная линия hESC-H1 - эмбриональные стволовые клетки человека. Для них выбраны 17 маркеров из БД ENCODE.

Цель исследования: попытаться понять природу границ между ТАДами.

Результаты I



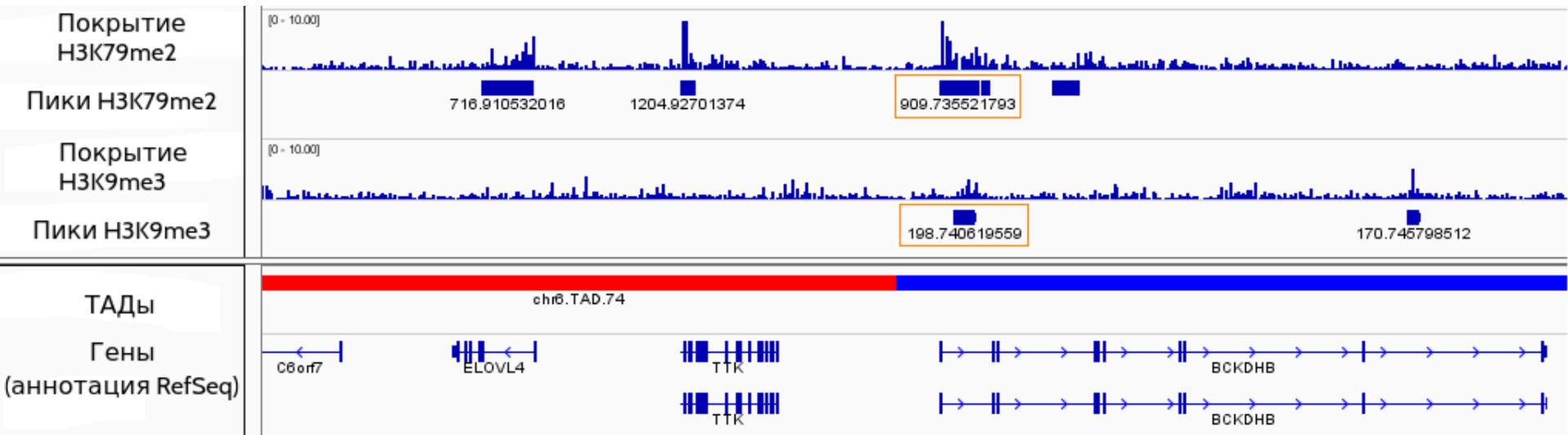
Результаты II



Состояние, характерное для сильных границ (балл 10):
30: CTCF

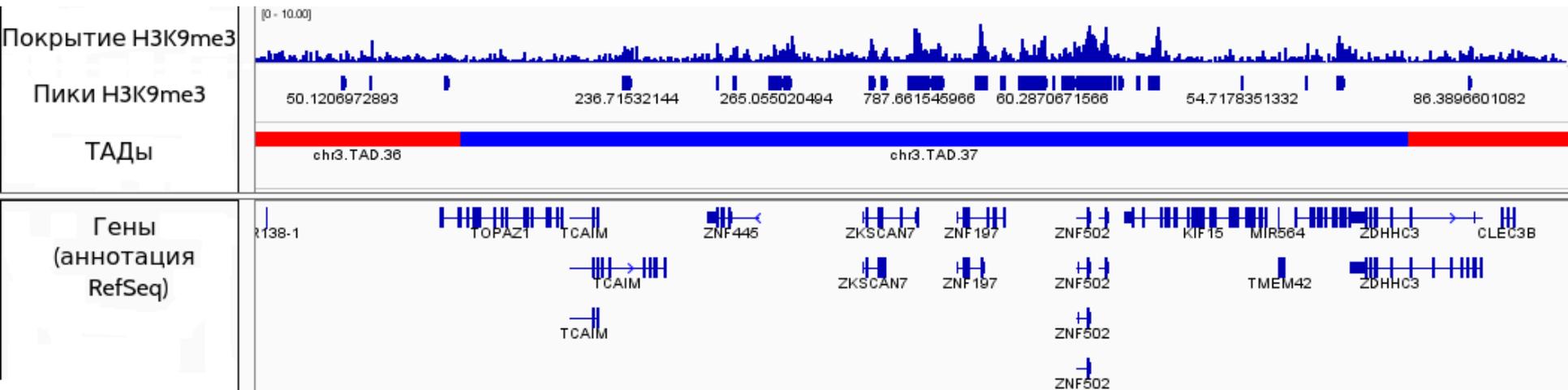
Состояния, характерные для более слабых границ (балл 1 - 9):
1: H3K79me2, H3K36me3
6: H3K4me1, H3K79me2, H4K20
9: H3K79me2, H3K9me3

Результаты III



Необычное сочетание маркеров H3K79me2 и H3K9me3 на границе двух ТАДов (пики, выделенные рамками).

Результаты IV



Пример ТАДа, предположительно находящегося в гетерохроматине: значительное количество пиков маркера H3K9me3.

Выводы

1. Автоматизированы выделение границ ТАДов и нахождение устойчивых сочетаний геномных и эпигенетических маркеров.
2. Разработан и реализован метод сопоставления границ ТАДов с устойчивыми сочетаниями маркеров.
3. Проанализированы приграничные регионы ТАДов в эмбриональных стволовых клетках человека (клеточная линия hESC-H1).

Благодарности

Washington University in St. Louis (США)

Александр Предеус

Максим Артёмов

Centro nacional de análisis genómico (Испания)

Франсуа Серра

Центр геномной биоинформатики им. Ф. Г. Добржанского, СПбГУ:

Алексей Комиссаров

Институт биоинформатики

Николай Вяххи

Екатерина Чайкина

Санкт-Петербургский Академический Университет

Александр Шлемов

JetBrains

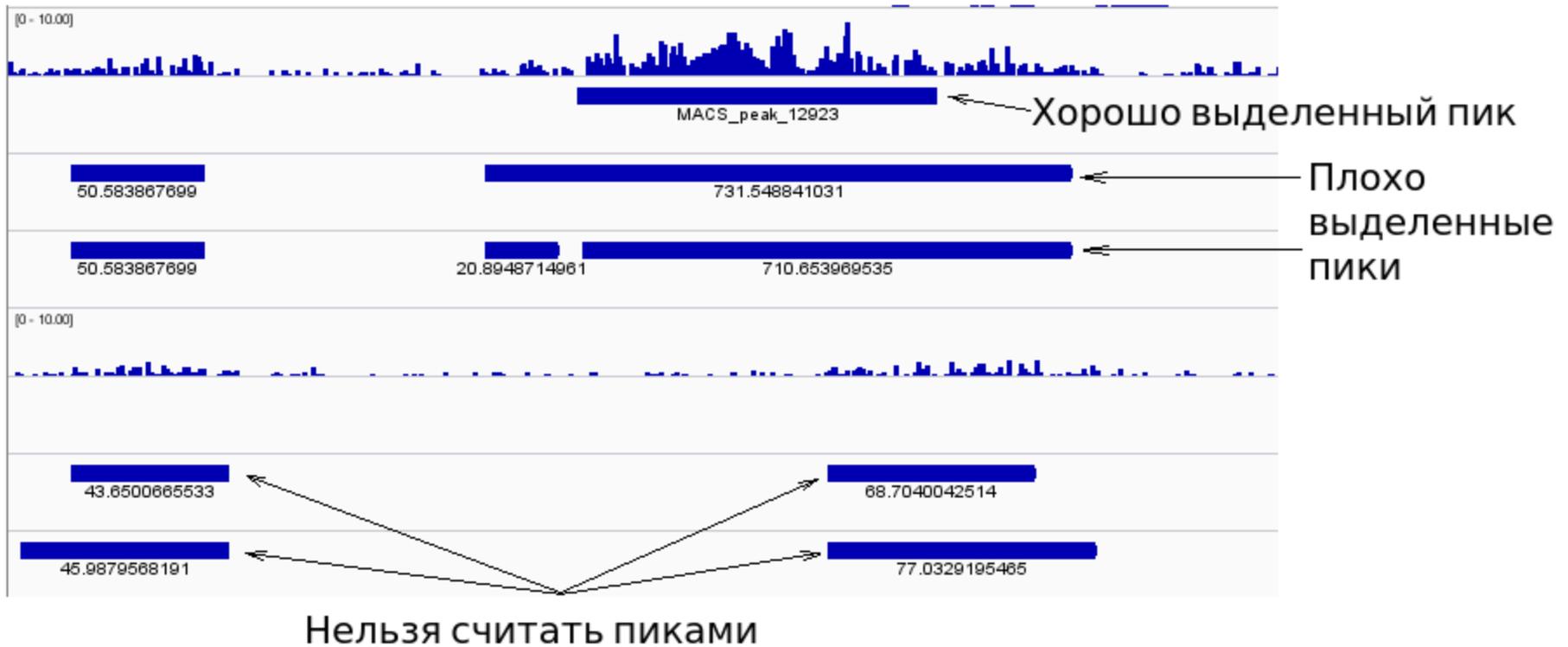
Сергей Лебедев

Олег Шпынов

Спасибо за внимание!

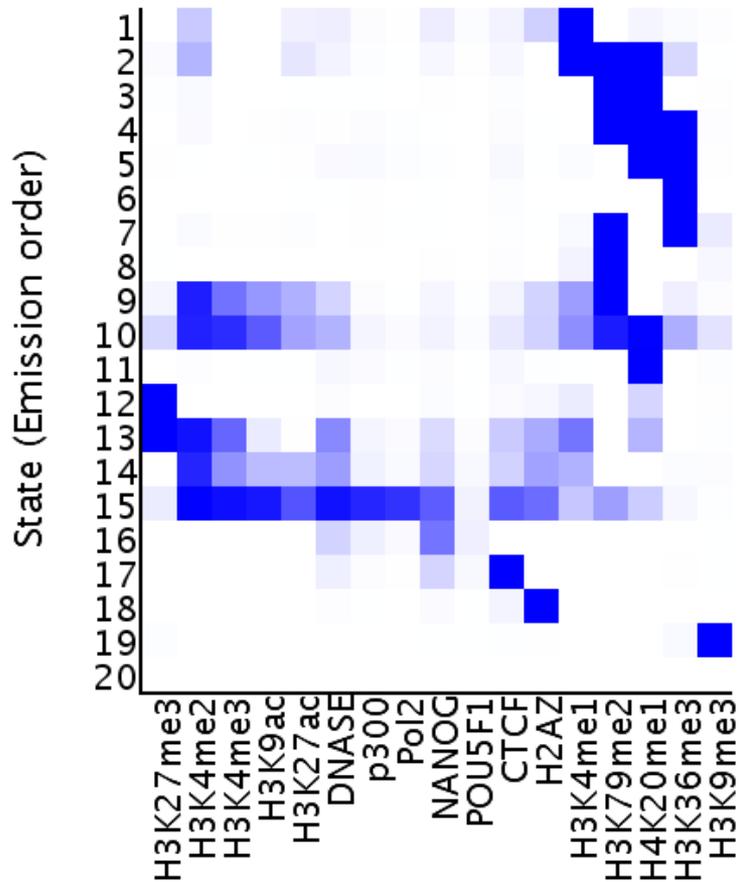
Дополнительные слайды

Режимы поиска пиков; фильтрация пиков



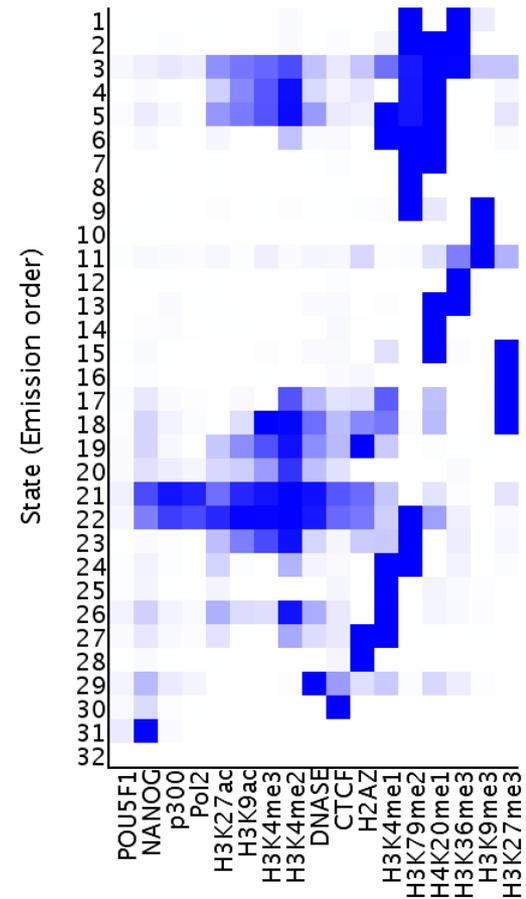
20 состояний СММ против 32 состояний СММ

Emission Parameters



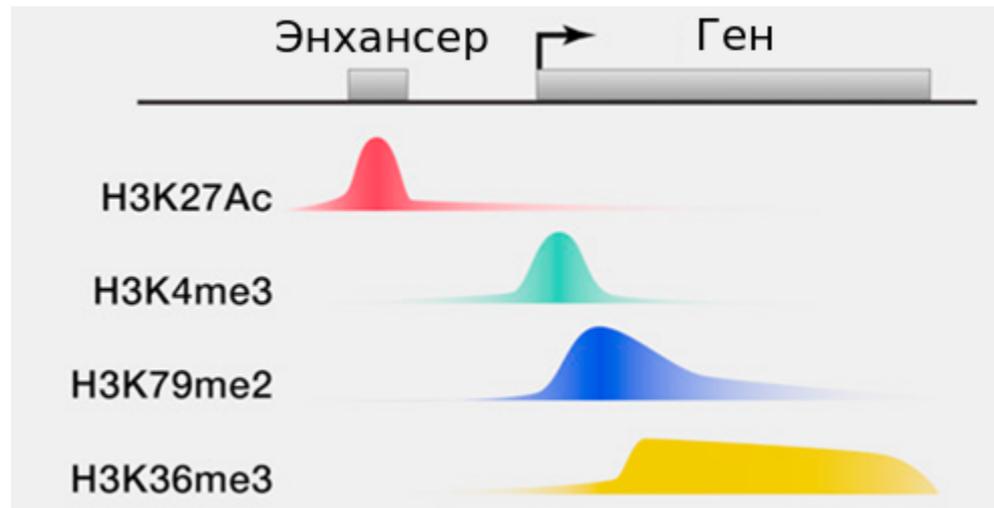
20 состояний

Emission Parameters



32 состояния

Некоторые эпигенетические маркеры



По Lee T. I. et al. Cell, 2013, 152(6):1237–1251.

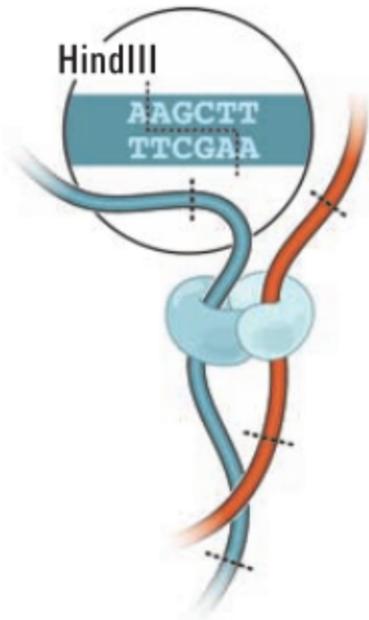
Дополнительные данные

Клеточная линия T47D-MTVL - клетки рака груди. Для них есть 4 маркера и данные RNA-seq, взятые до и после обработки клеток гормоном.

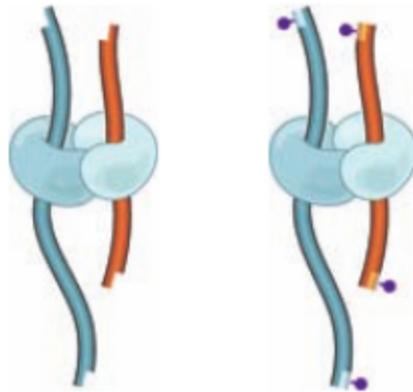
Цель исследования: выявить изменение устойчивых сочетаний маркеров в ТАДах до и после обработки.

Метод Hi-C

Закрепление взаимодействий

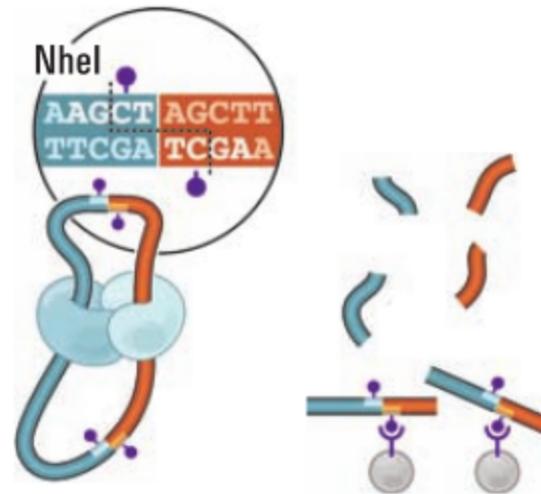


Разрезание ДНК рестриктазой



Восстановление нуклеотидов на концах и пометка концов биотинином

Лигирование концов



Очистка и разрезание ДНК; выделение участков с биотином

Секвенирование с получением парных ридов

