



МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М.В.ЛОМОНОСОВА

Выпускная квалификационная работа бакалавра

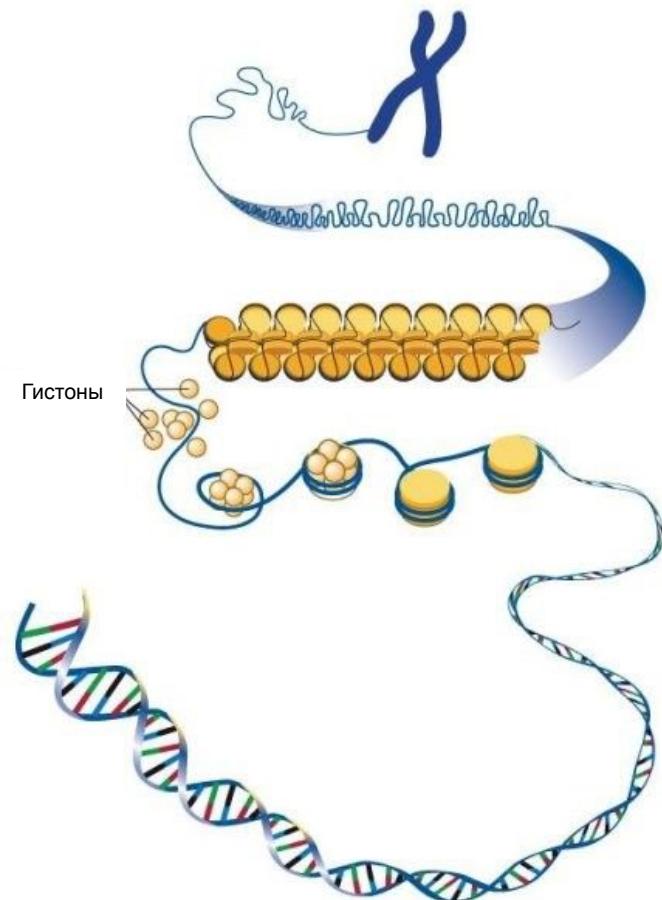
Определение распределения позиций сайтов связывания пионерных транскрипционных факторов относительно положений нуклеосом по данным MNase-Seq-ExoIII

Выполнил студент
Рябов Дмитрий Михайлович

Научный руководитель:
к.ф.-м.н. Армееев Григорий Алексеевич

Москва
2024

Нуклеосома – базовая единица хроматина

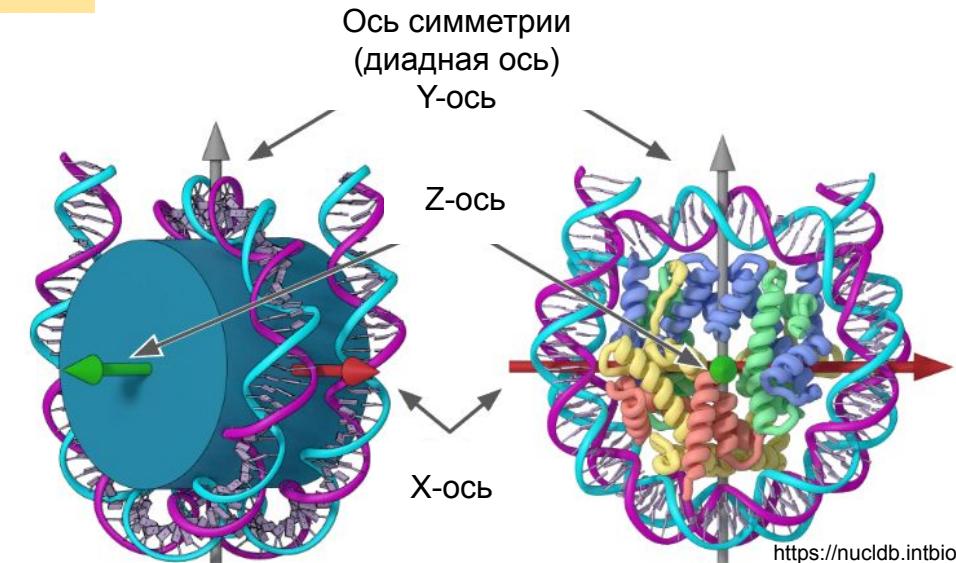


Хромосома

фибрилла

Хроматин в
форме бусин
на нити

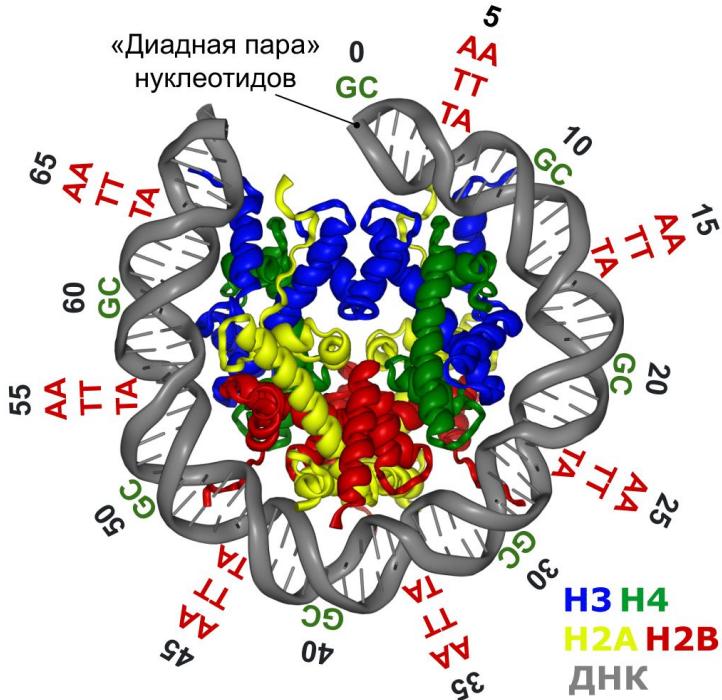
ДНК



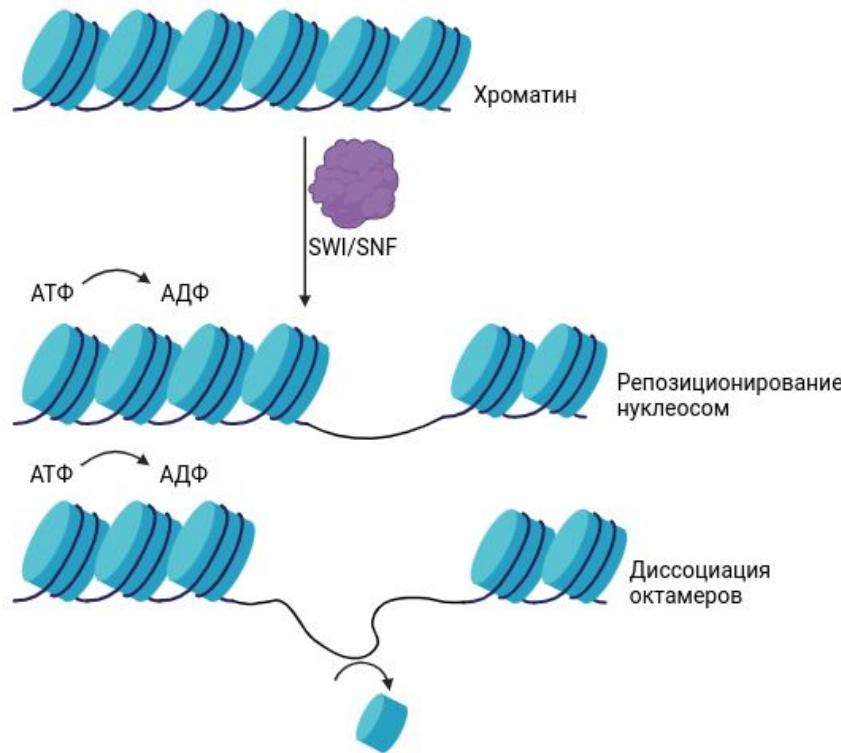
<https://nucldb.intbio.org>

- Диада - ось симметрии нуклеосомной ДНК
- Хроматин - комплекс ДНК и белков

Особенности позиционирования нуклеосом



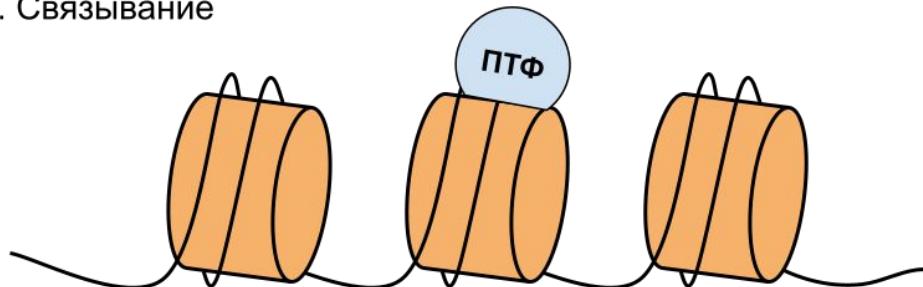
- Особенности чередования частот встречаемости нуклеотидных шагов относятся к пассивным механизмам позиционирования
WW = AT, TA, TT, AA
SS = GC, GG, CC, CG



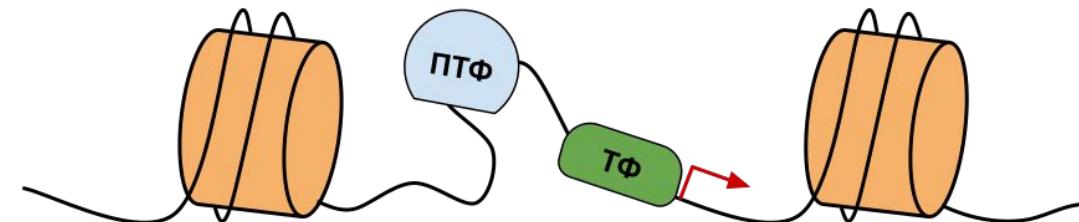
- К активным механизмам позиционирования нуклеосом можно отнести работу ремоделлеров, происходящую с затратой АТФ

Пионерные транскрипционные факторы (ПТФ)

1. Связывание



2. Реорганизация



- ПТФ - белки, связывающиеся первыми с гетерохроматином, вызывая его перестройки
- Есть ли взаимосвязь между позициями нуклеосом и расположением сайтов ПТФ?

Цели и задачи

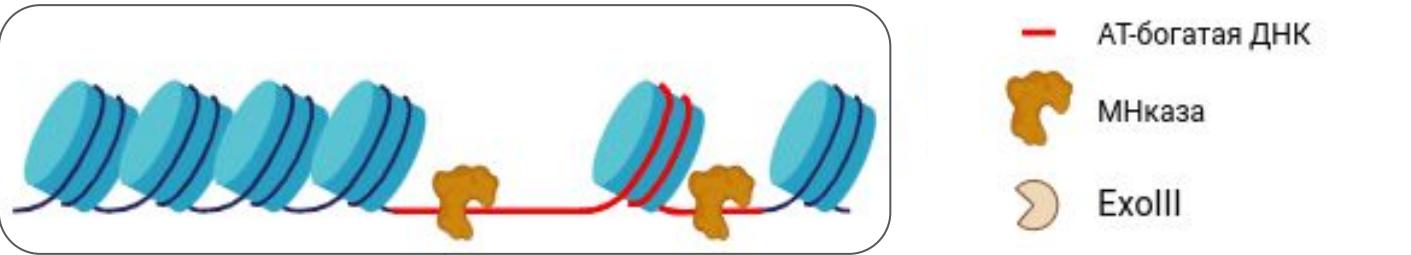
Цель работы:

Определить распределения сайтов связывания ПТФ относительно диад нуклеосом

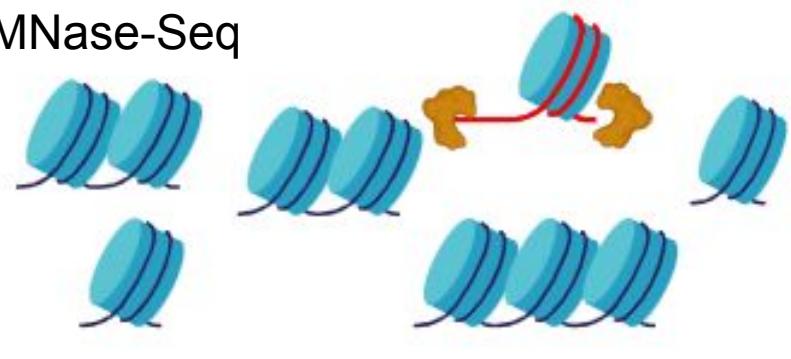
Задачи работы:

- Картировать данные MNase-Seq-ExoIII на геном *S.cerevisiae*
- Разработать метод определения позиций диадной пары нуклеотидов по данным MNase-Seq-ExoIII секвенирования
- Провести сравнение разработанного метода с существующими подходами

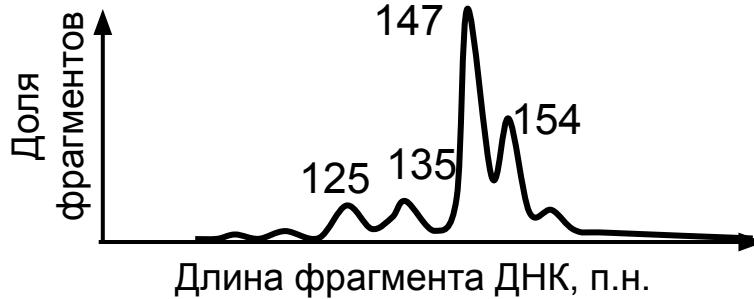
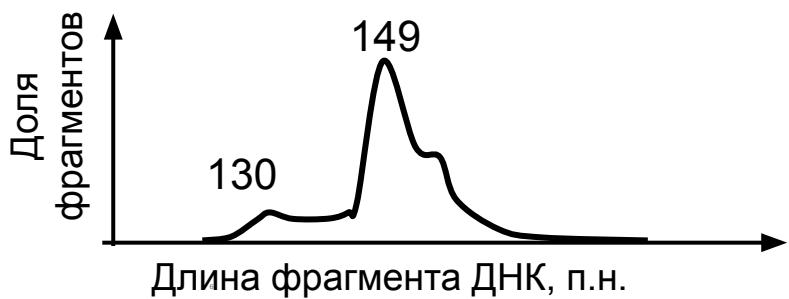
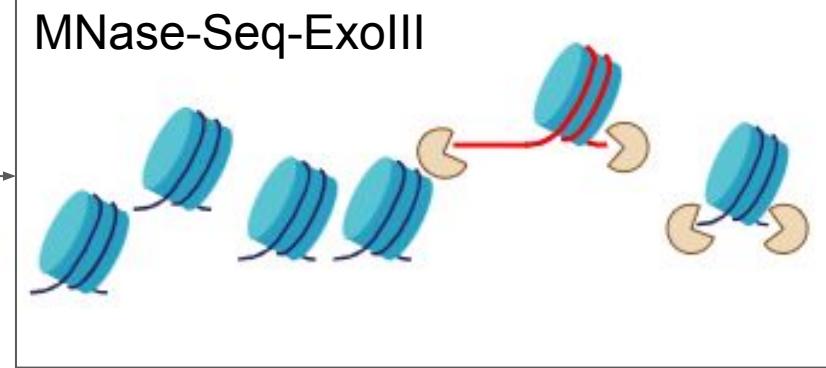
Сравнение MNase-Seq и MNase-Seq-ExoIII



MNase-Seq



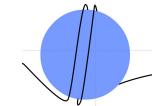
MNase-Seq-ExoIII



MNase-Seq-ExoIII



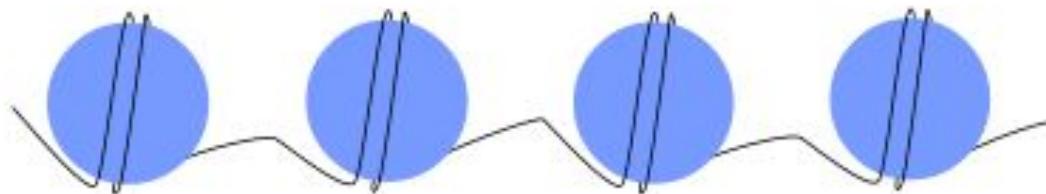
МНказа



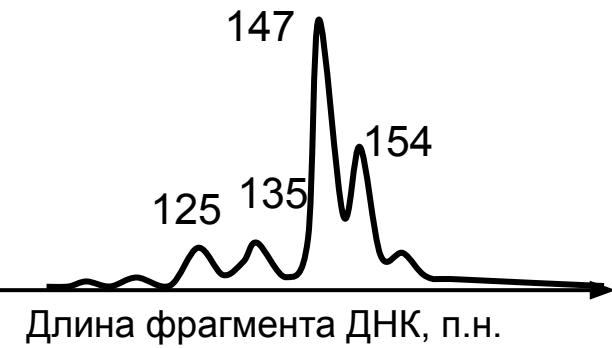
Нуклеосома



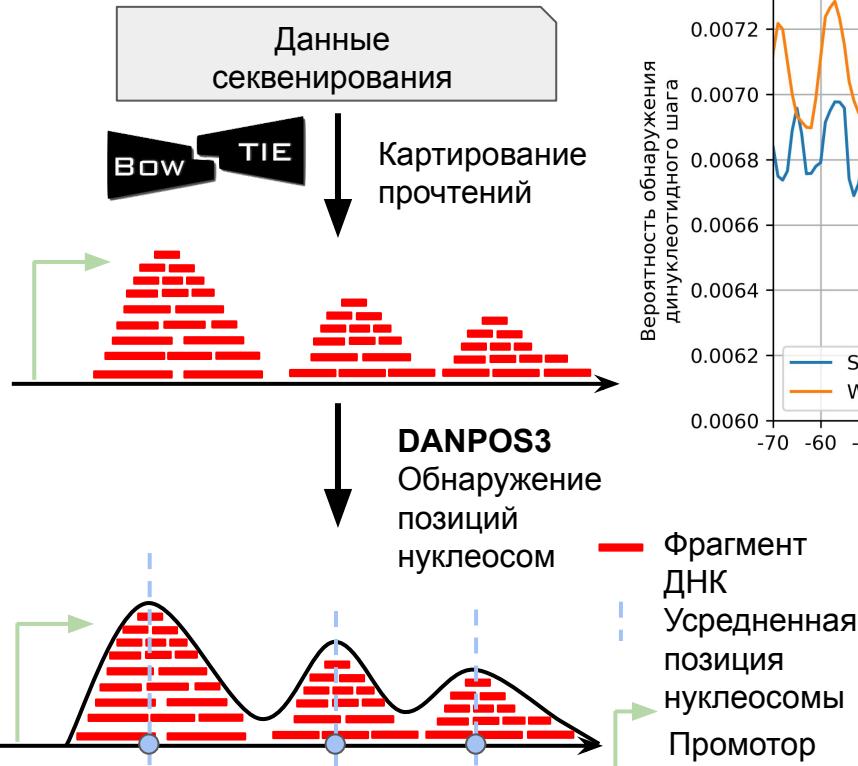
ExoIII



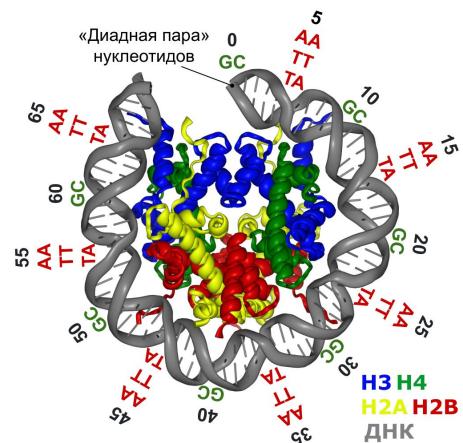
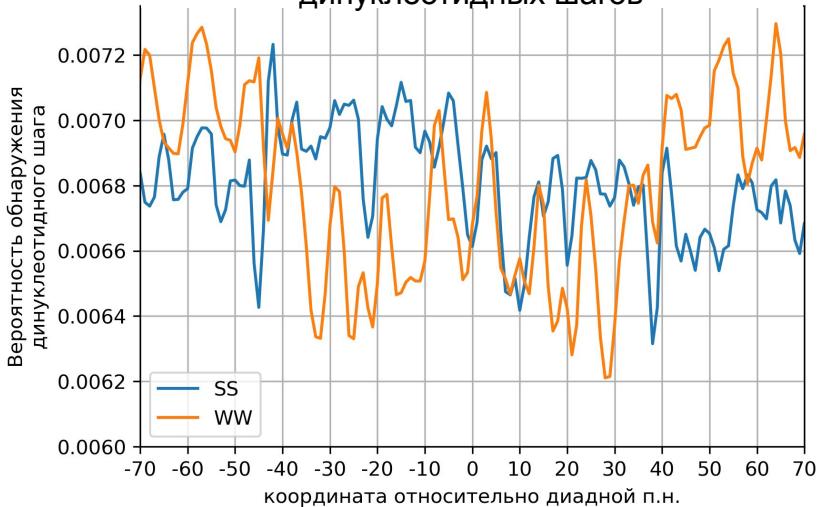
Доля фрагментов



Стандартный протокол обработки NGS



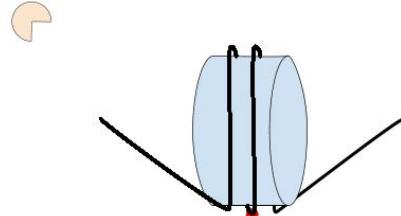
Распределение вероятности встречи динуклеотидных шагов



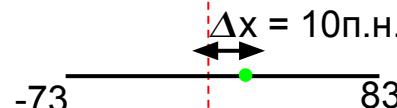
- DANPOS3 не позволяет выявить периодичность позиционирования нуклеосом
- Следует разработать новый метод анализа данных MNase-Seq-ExoIII

Описание разработанной модели

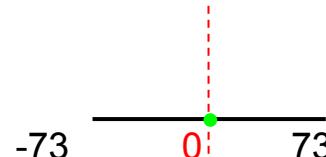
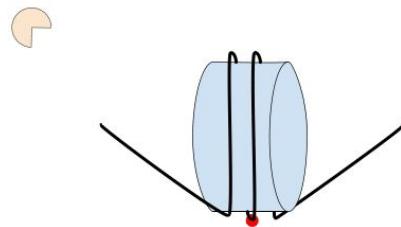
Вероятность диссоциации ExoIII зависит от расстояния до диадной пары нуклеотидов



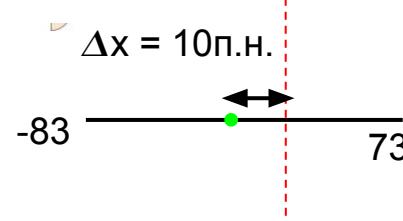
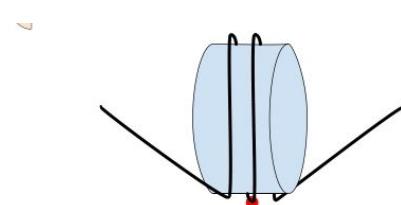
настоящее положение
диадной п.н.



При большем гидролизе левой половины ДНК центр фрагмента находится правее положения диадной п.н.



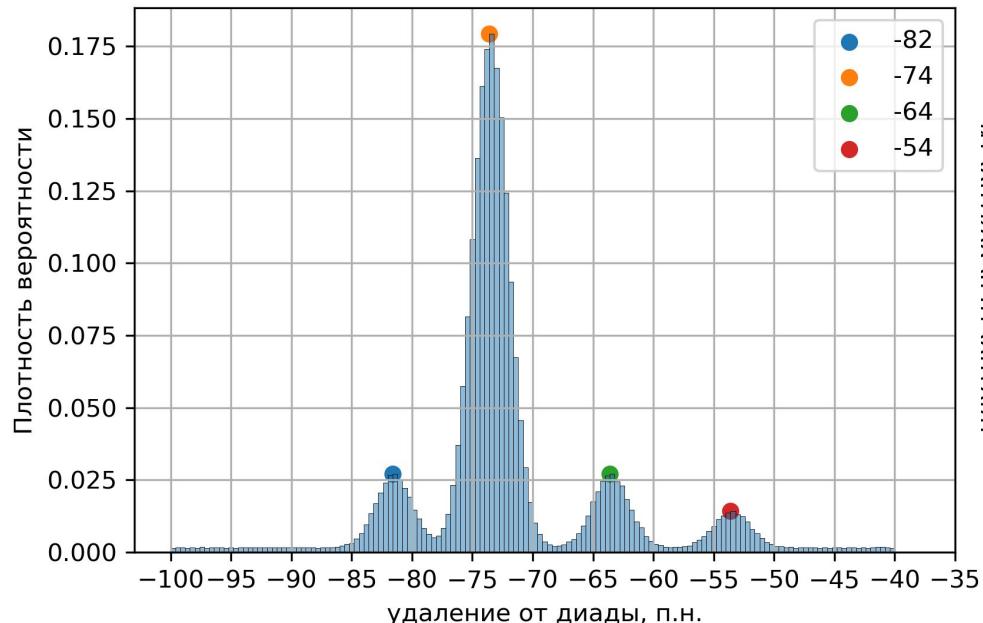
При симметричном гидролизе диадная п.н. совпадает с центром фрагмента



При большем гидролизе правой половины ДНК центр фрагмента находится левее положения диадной п.н.

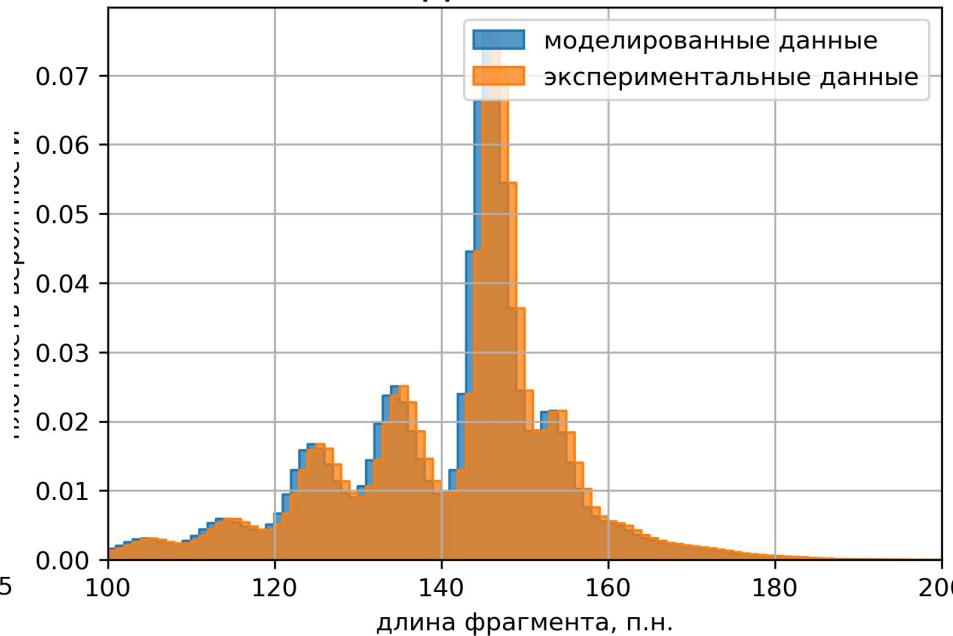
Подобраны параметры модели

Оптимизированная плотность вероятности диссоциации ExoIII от ДНК



Природа пиков связана с точками взаимодействия ДНК с ядром

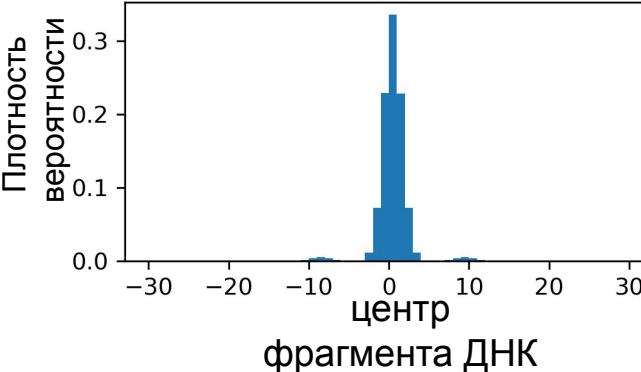
Сравнение экспериментального и моделированного распределения длин фрагментов



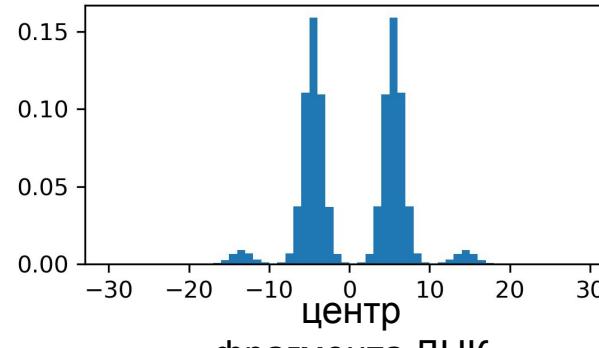
Оптимизированная модель воспроизводит экспериментальное распределение длин фрагментов ДНК

Найдены вероятности обнаружения диады в зависимости от длины фрагмента

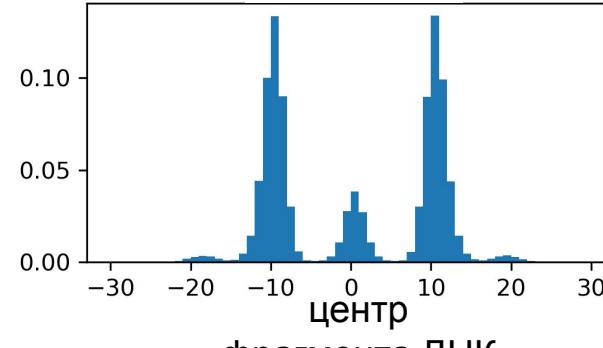
146 п.н.



136 п.н.



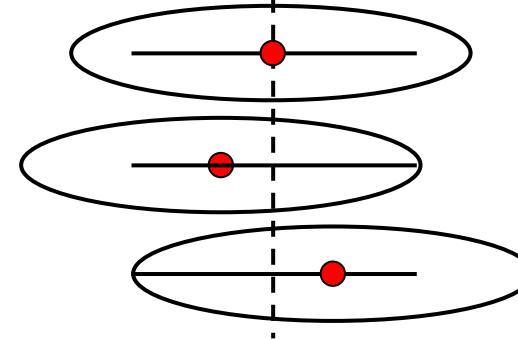
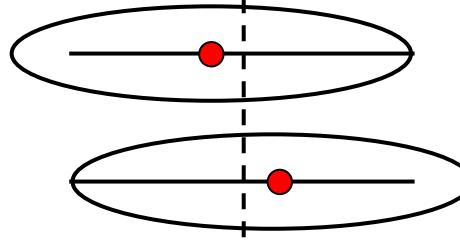
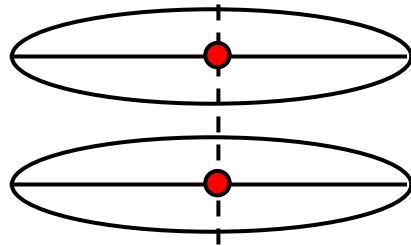
124 п.н.



центр
фрагмента ДНК

центр
фрагмента ДНК

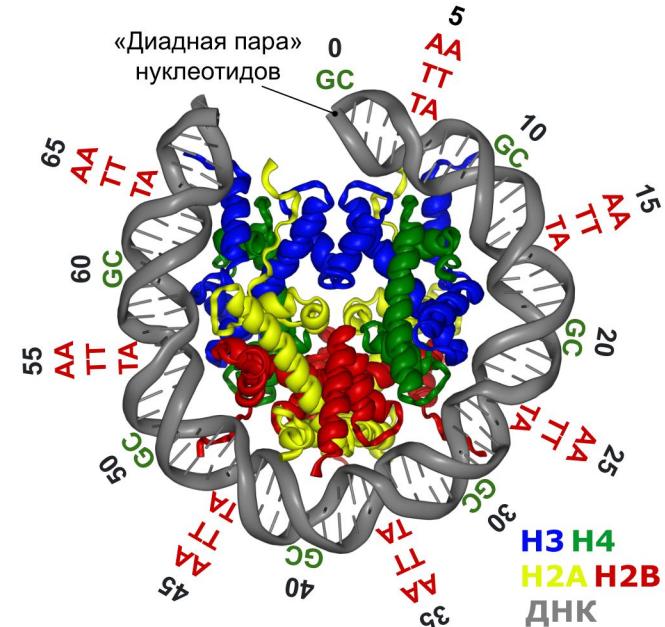
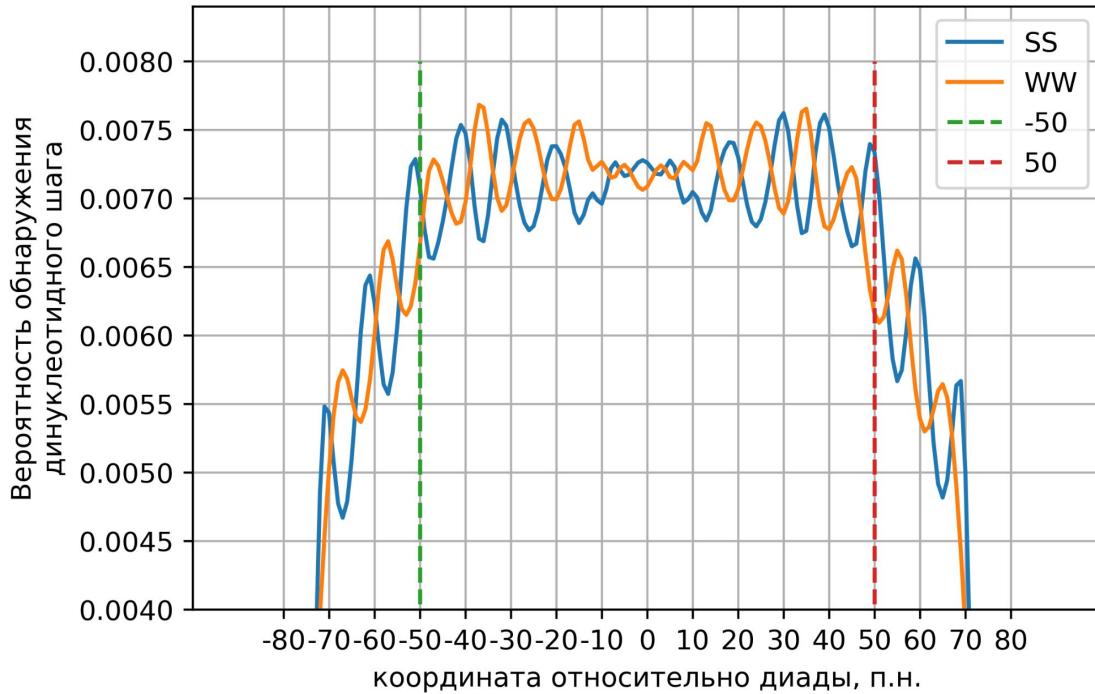
центр
фрагмента ДНК



- Положение диады

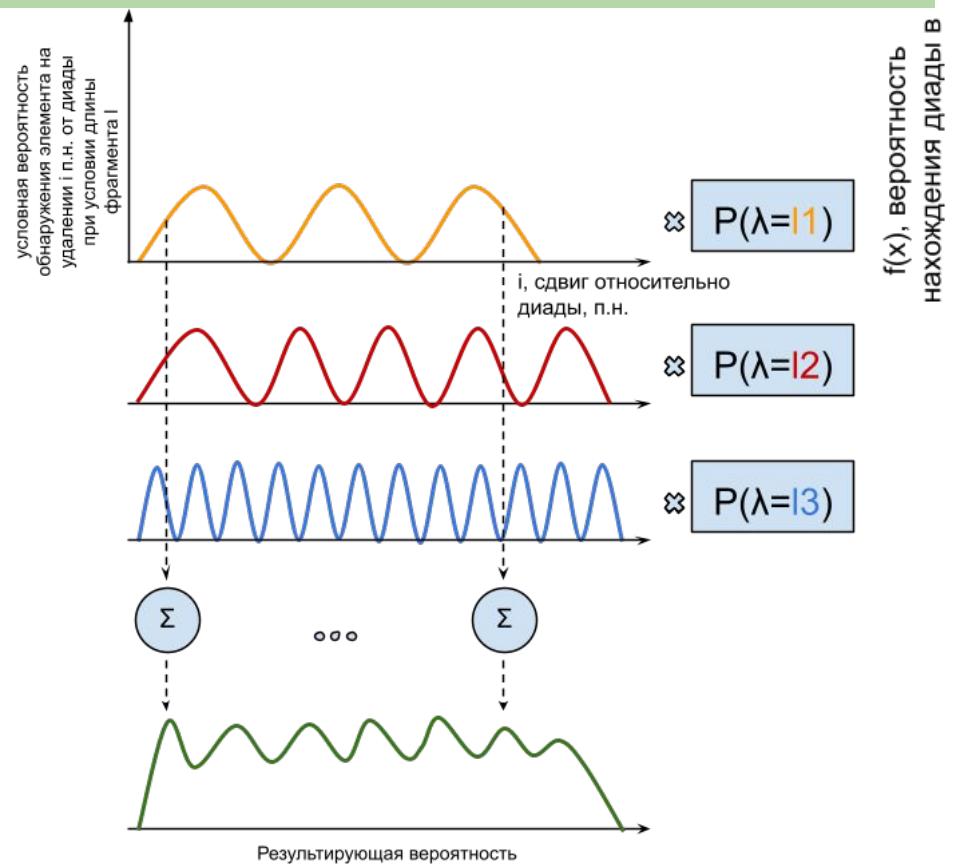
- Зная положение диады в модельных данных, для шаблона каждой длины можно узнать отклонение положения диады от его центра

Профили динуклеотидных шагов, построенные с применением модели



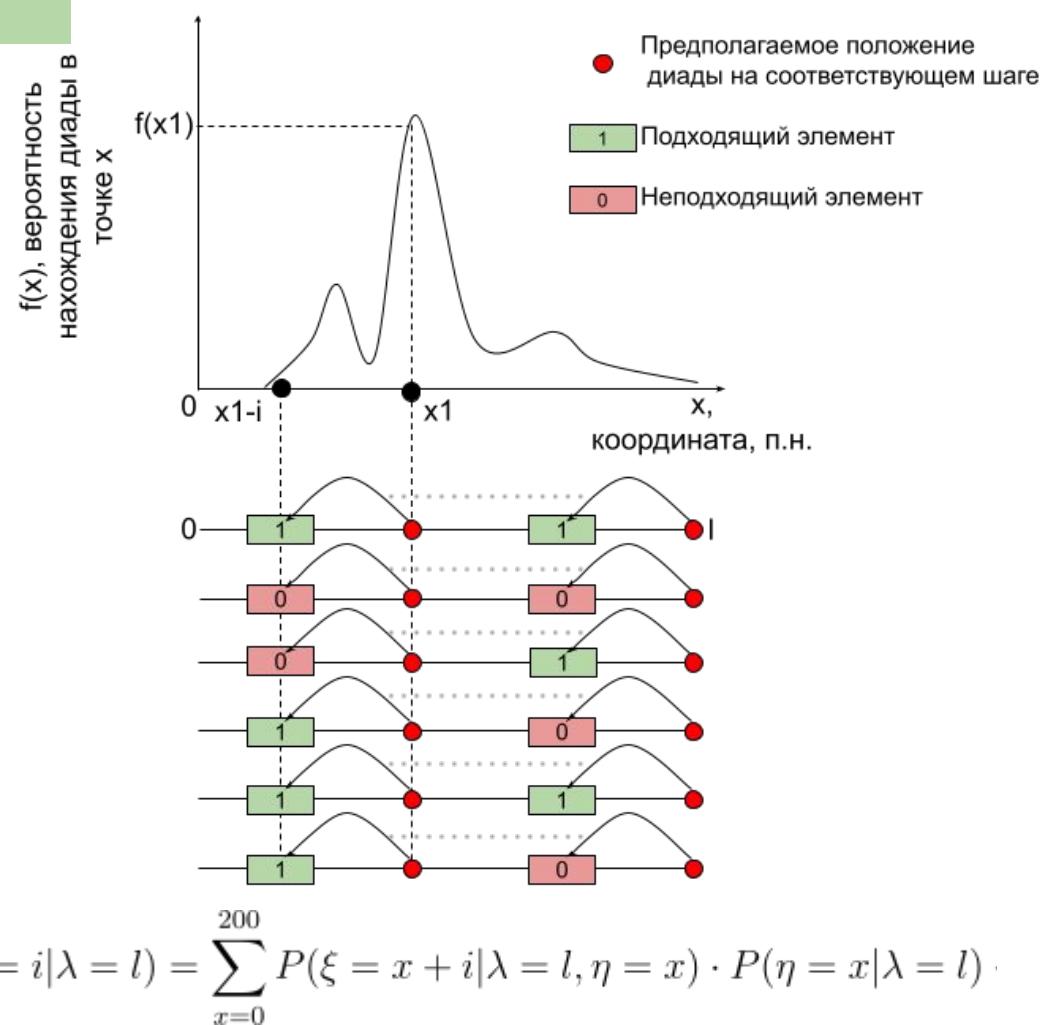
- Сигнал для WW и SS шагов имеет статистически значимую периодичность в 10 п.н.
- Профили сигналов находятся в противофазе, что соответствует литературным данным

Описание разработанного алгоритма

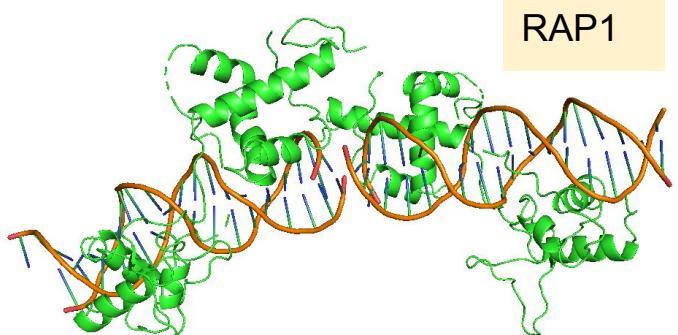
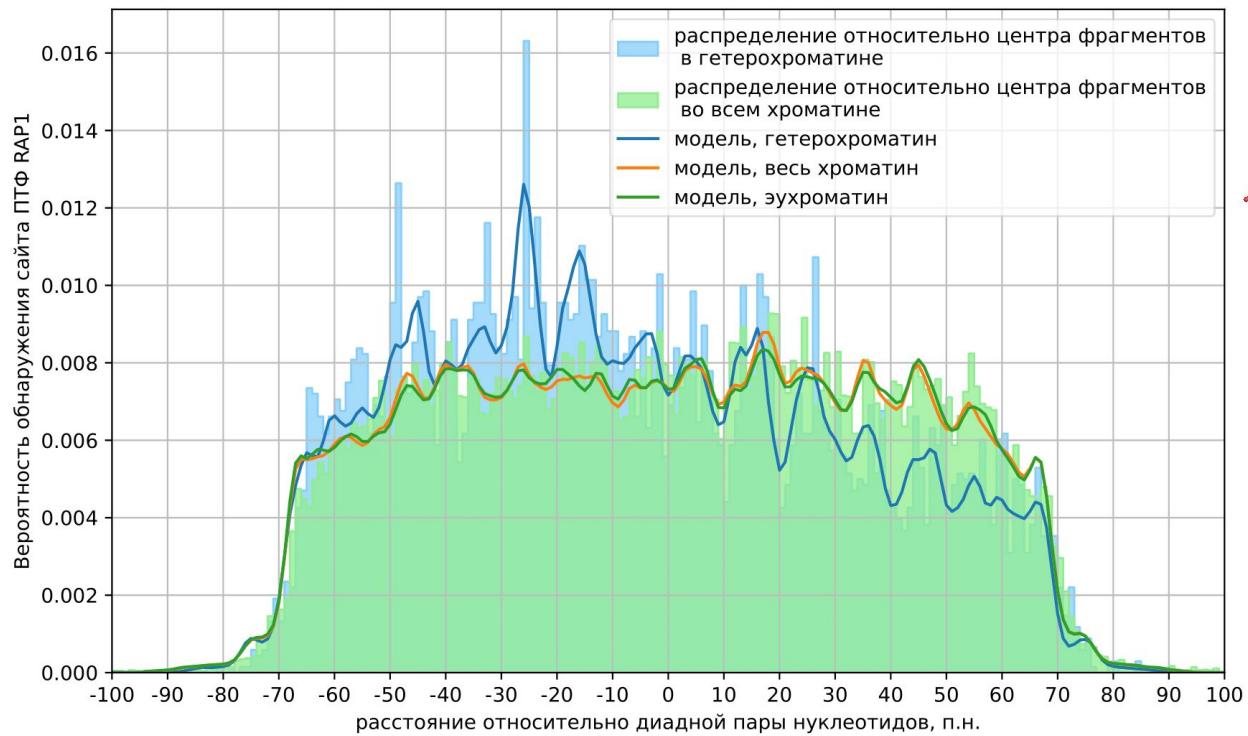


$$P(\xi = i) = \sum_l P(\xi = i | \lambda = l) \cdot P(\lambda = l)$$

$$P(\xi = i | \lambda = l) = \sum_{x=0}^{200} P(\xi = x + i | \lambda = l, \eta = x) \cdot P(\eta = x | \lambda = l)$$



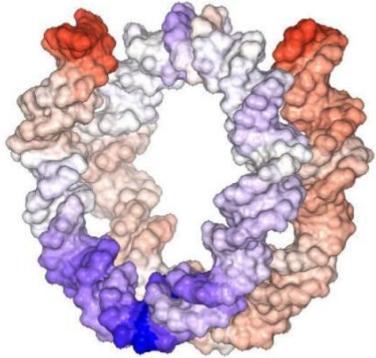
Распределение сайтов связывания RAP1 относительно диадной пары нуклеотидов



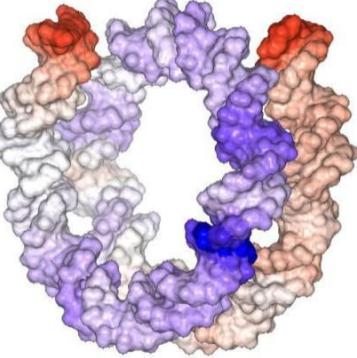
- Применение модели позволяет выявить периодичность в данных
- В эухроматиновых областях и всем хроматине результаты схожи
- В гетерохроматине RAP1 преимущественно связывается асимметрично

Распределение сайтов связывания ПТФ относительно диад

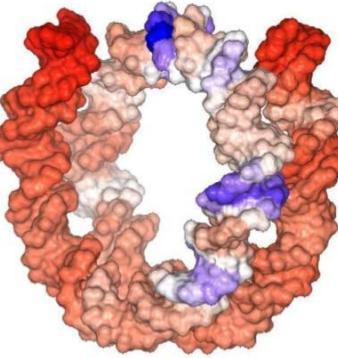
RAP1



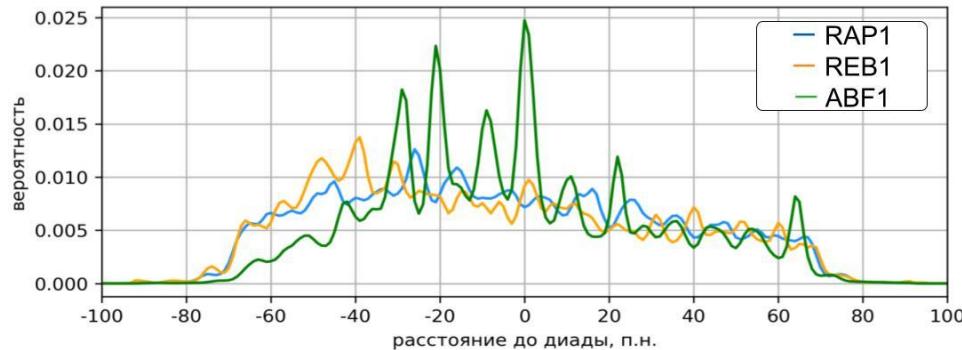
REB1



ABF1

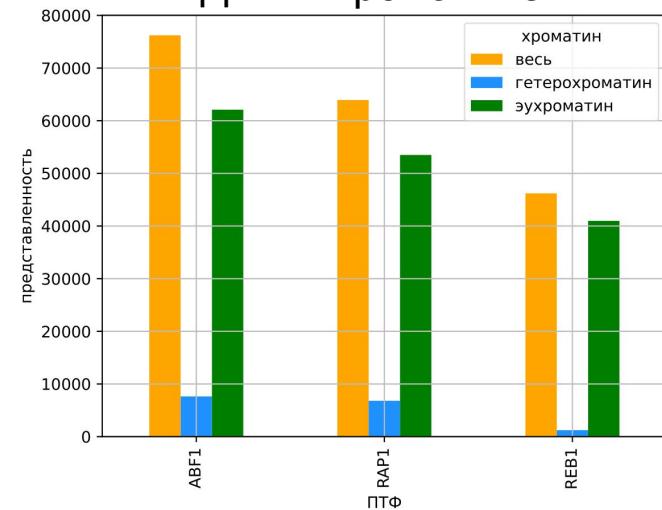


0.000 0.003 0.007 0.010 0.014
относительная вероятность обнаружения сайта связывания ПТФ

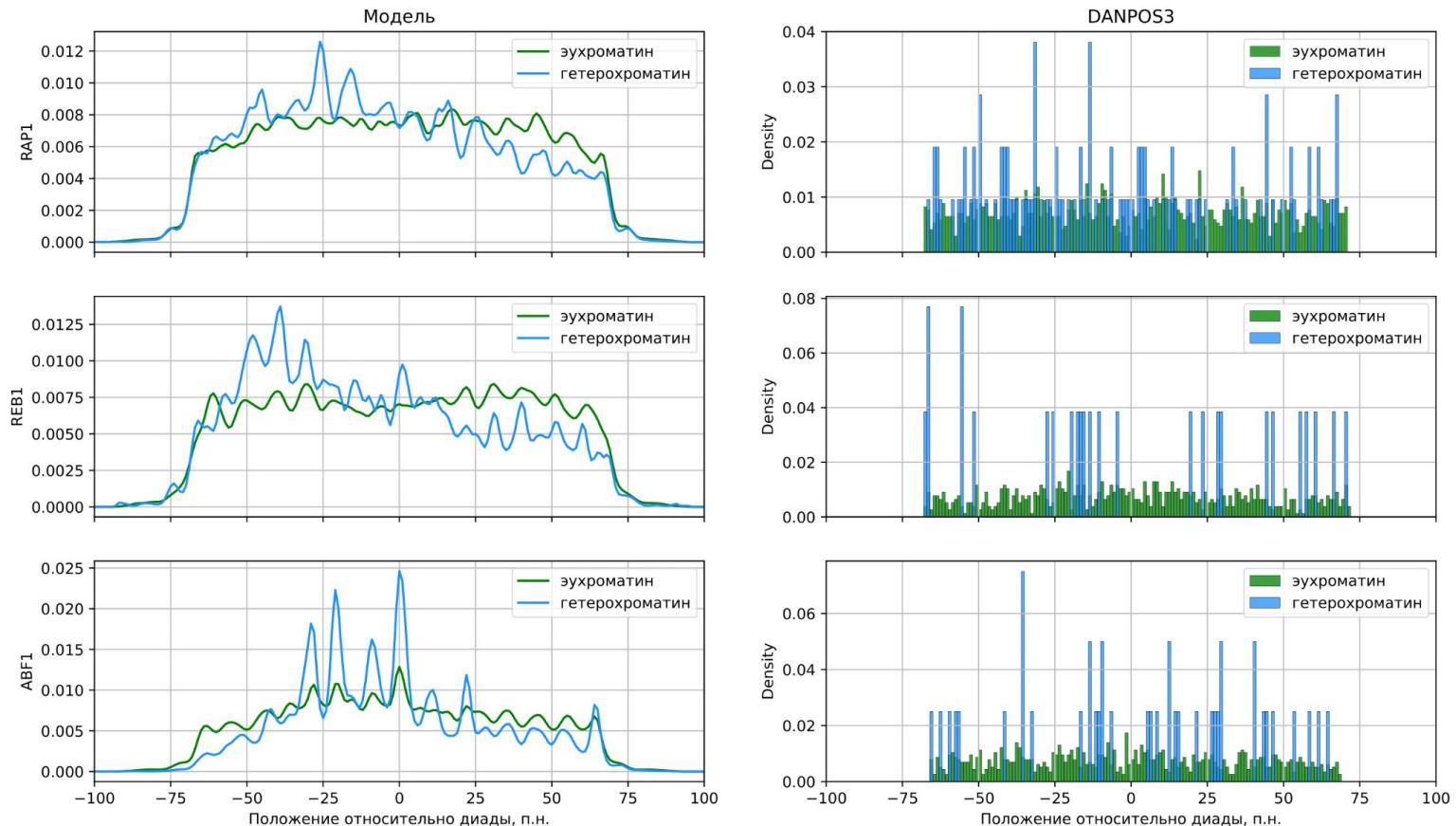


- В позиционировании ПТФ наблюдается 10 п.н. периодичность

Количество фрагментов ДНК в хроматине



Сравнение модели с программой DANPOS3



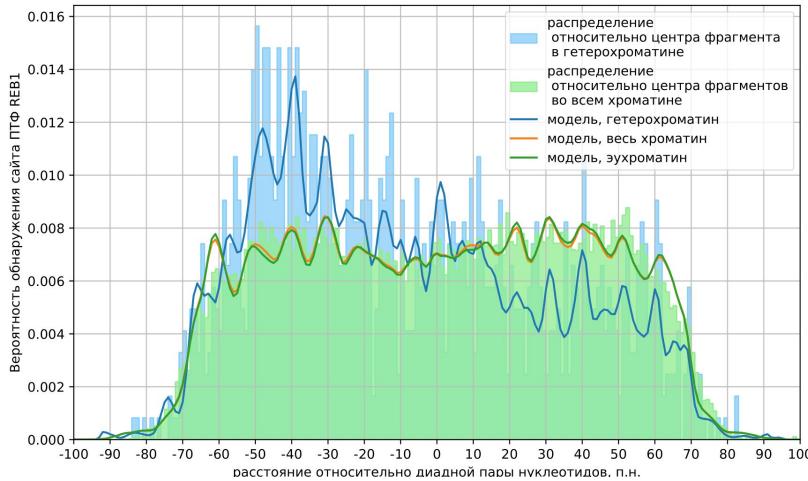
При сравнении разработанной модели с программой DANPOS3 модель позволяет выявить периодичность в данных

На основании проделанной работы можно сделать следующие выводы:

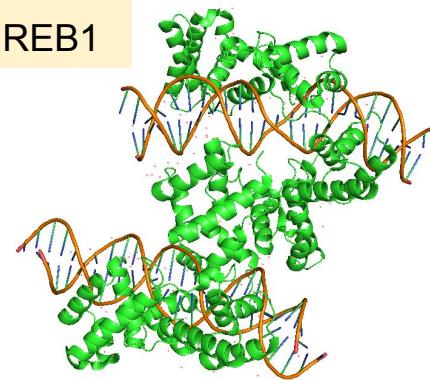
1. Результаты работы разработанного алгоритма картирования данных MNase-Seq-ExoIII согласуются с литературными данными.
2. Разработанный алгоритм позволяет определять распределение сайтов связывания ПТФ относительно положения диадных пар нуклеотидов нуклеосом.
3. В отличие от программного инструмента DANPOS3 разработанный алгоритм позволяет выявить периодичность расположения сайтов связывания ПТФ с нуклеосомой.

Спасибо за внимание!

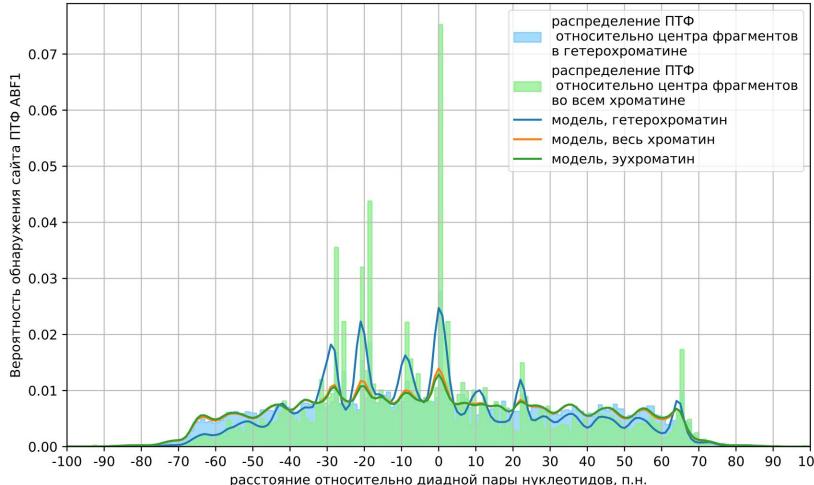
Распределение сайтов связывания REB1 и ABF1 относительно диадной п.н.



REB1



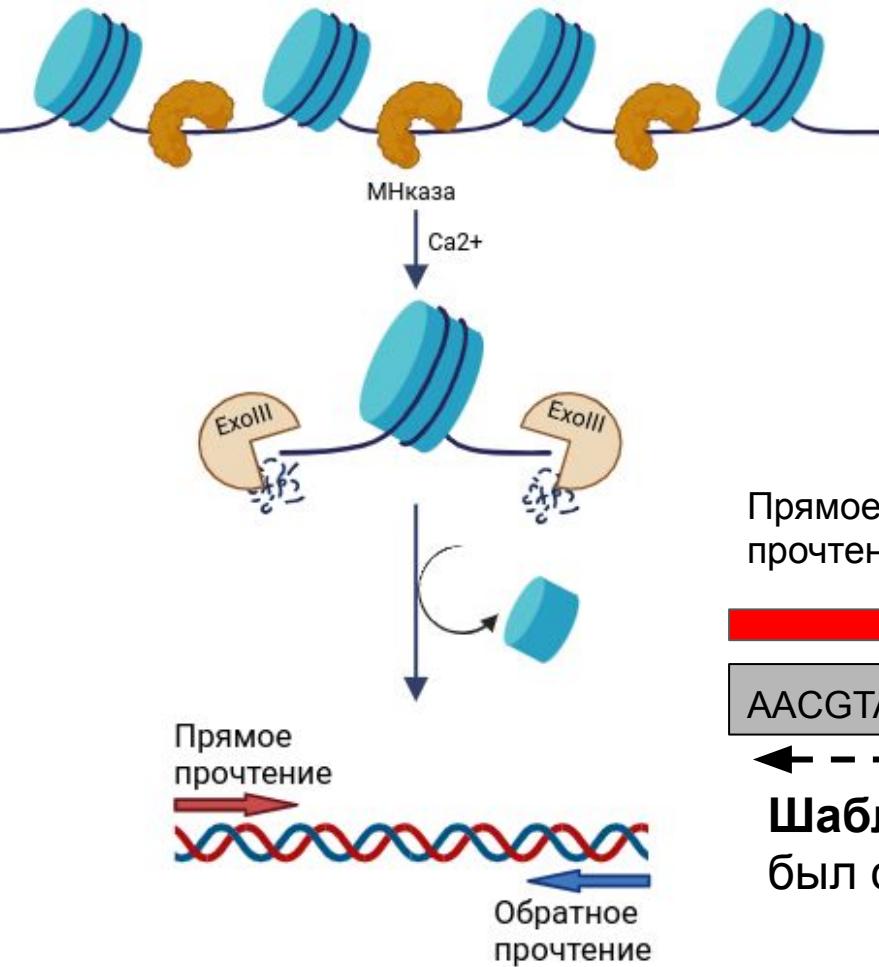
- В случае REB1 ситуация аналогична RAP1: связывание преимущественно левее диады



ABF1

- В случае ABF1 наблюдается более выраженная периодичность в 10 п.н.

Парноконцевое MNase-Seq-ExoIII секвенирование



- Метод NGS MNase-Seq-Exo III позволяет исследовать структуру хроматина
 - Микрококковая нуклеаза (МНказа) вносит разрывы в линкерные области ДНК
 - Экзонуклеаза Exo III укорачивает концы линкерной ДНК

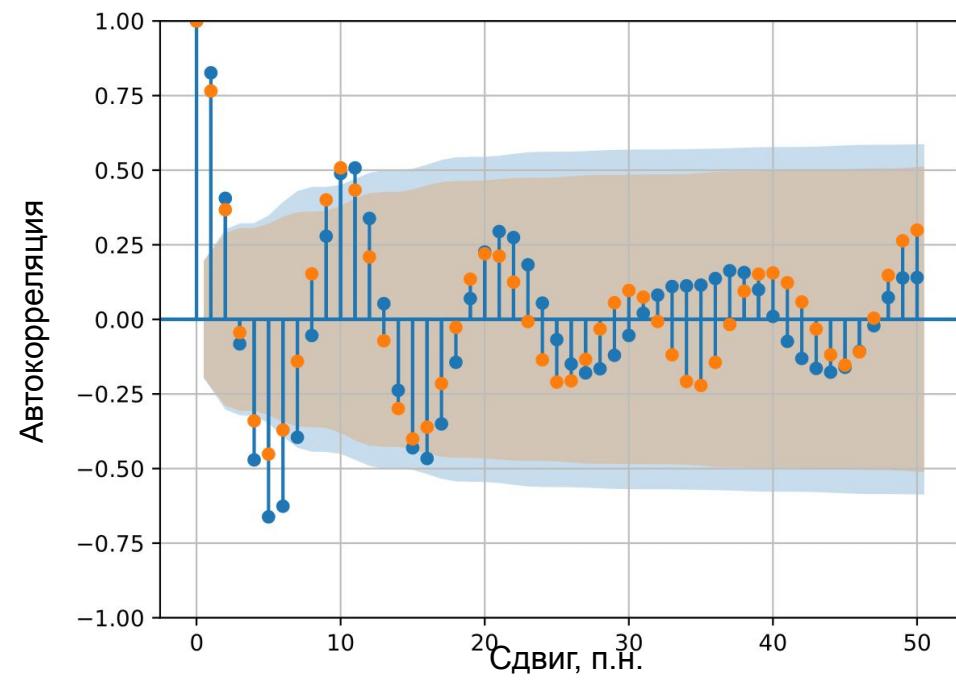
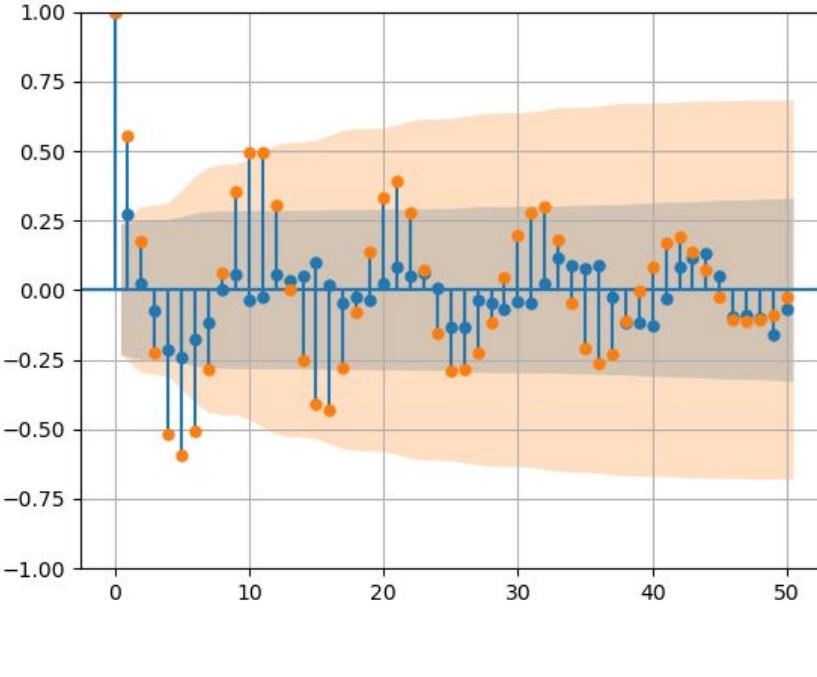
Прямое прочтение

Обратное прочтение

Прямое прочтение

AACGTAGCGATGATGACACCAGTGACGTAGCGATGATGACACC

Шаблон – участок референсного генома, который был связан с гистонами



центр переваренного
фрагмента

диада

центр фрагментов

Δx 0

Δx

Δx удаление диады

от центра фрагмента, п.н.

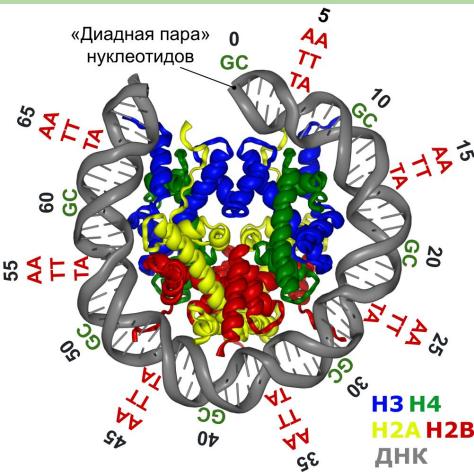
Δx

вероятность сдвига Δx

0

позиция относительно
центра фрагмента, п.н.

Профили динуклеотидных шагов для 147-п.н. шаблонов



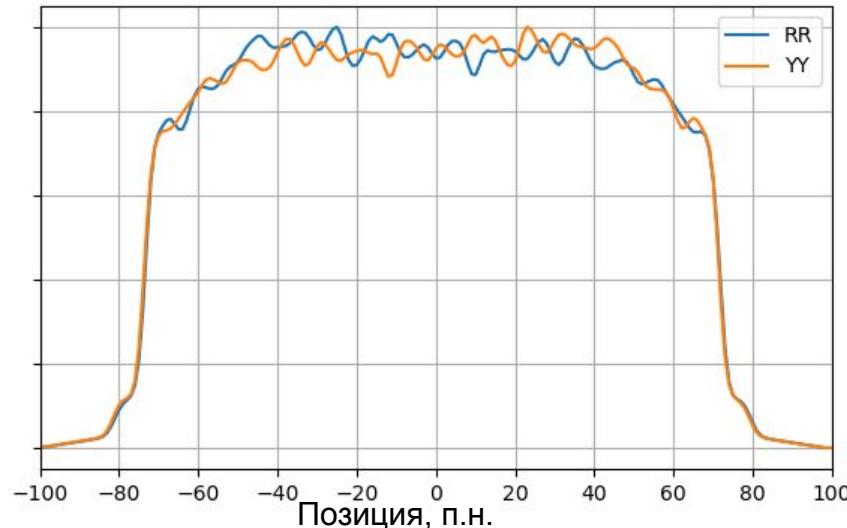
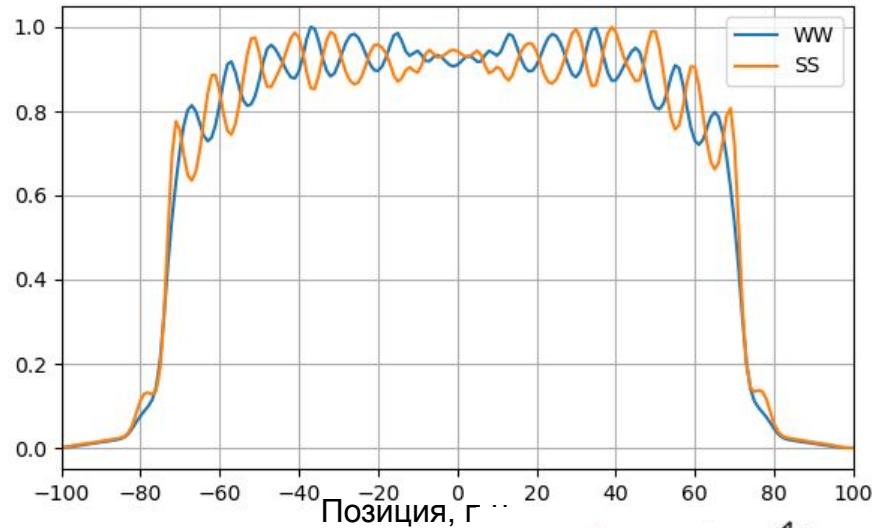
Периодичность встречаемости динуклеотидных шагов WW (AA, TA, AT, TT) и SS (GG, GC, CG, CC) в структуре нуклеосомы

- Прослеживается статистически значимая периодичность в 10 п.н. WW шагов.
- Статистически значимая периодичность в 10 п.н. для SS шагов отсутствует
- Фаза колебаний WW и SS совпадает, что противоречит литературным данным

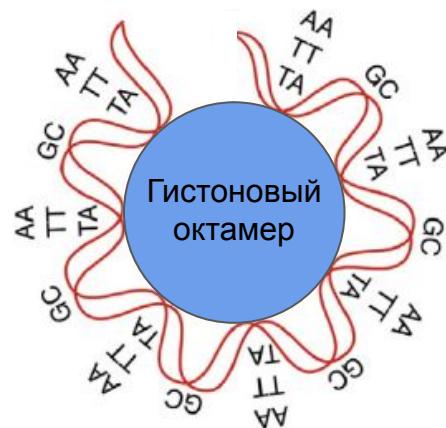
Валидация модели

Вероятность встречи динуклеотидных шагов в сайтах связывания нуклеосом

Плотность вероятности

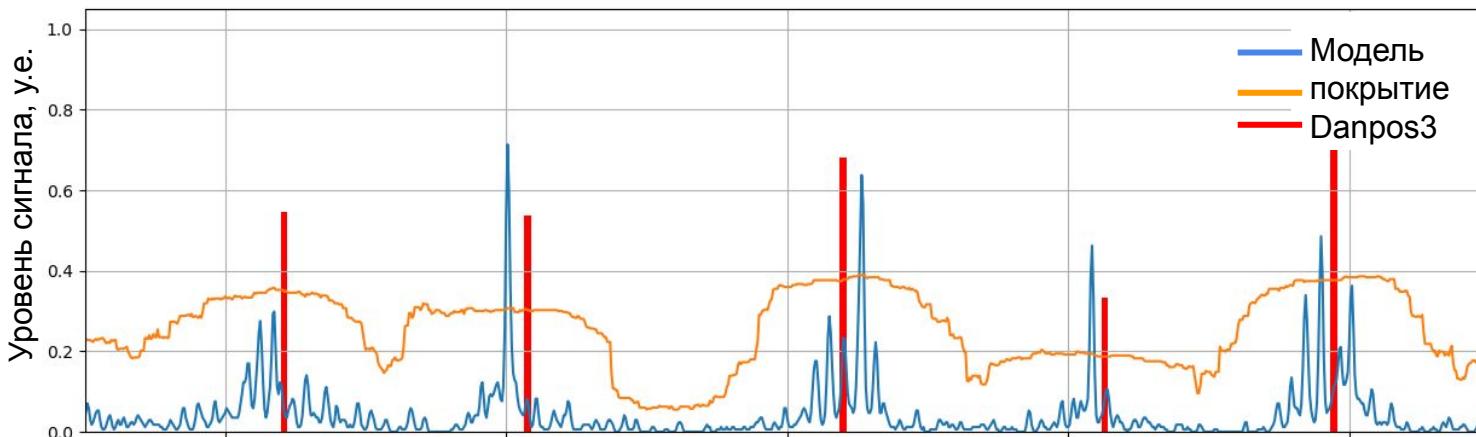


Варианты динуклеотидов
WW = AA, AT, TA, TT
YR = CA, CG, TA, TG
SS = CC, CG, GC, GG

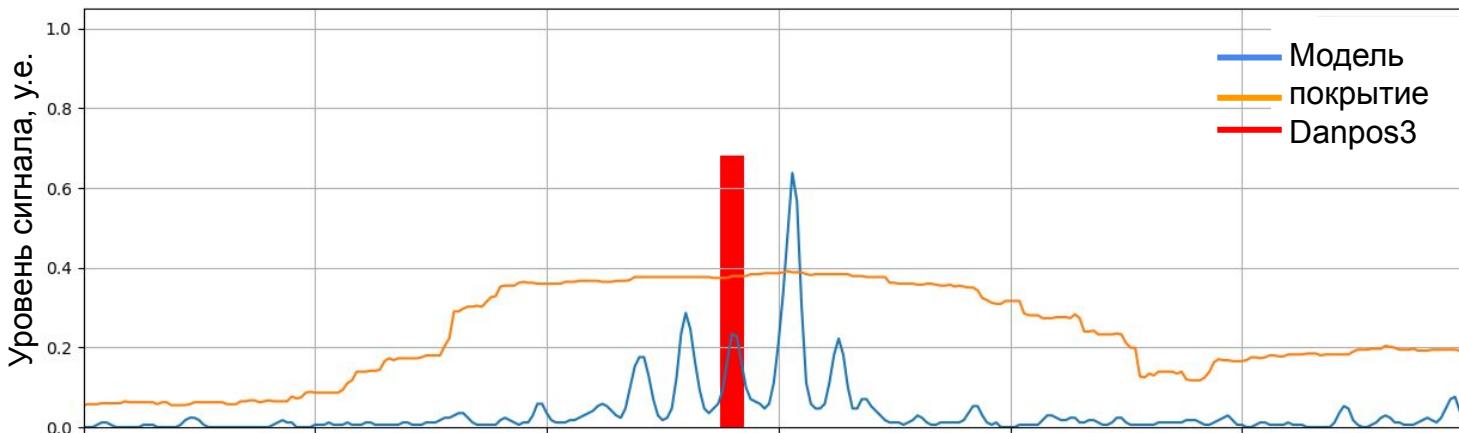


- Периодичность встречи динуклеотидов WW составляет 10 п.н.
- В сайтах связывания нуклеосом динуклеотиды WW встречаются с периодичностью 10 п.н.

Сравнение профилей заселенности

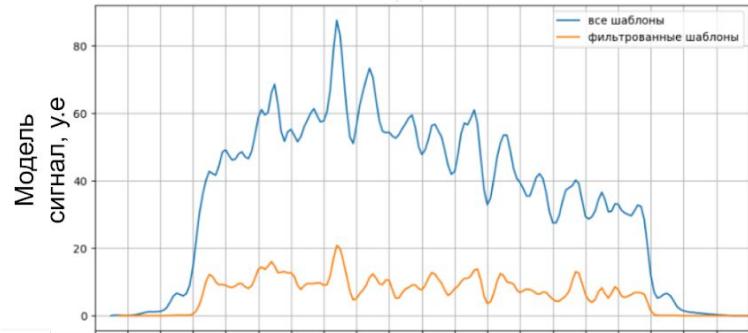


Колебания сигнала модели соответствуют разным положениям нуклеосом в разных клетках исследуемой популяции



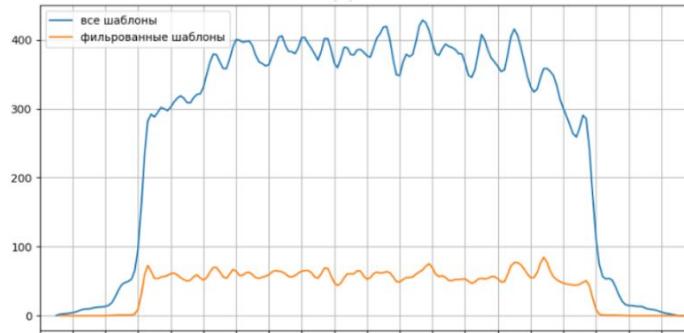
Распределение сайтов связывания ПТФ RAP1 относительно диад

гетерохроматин

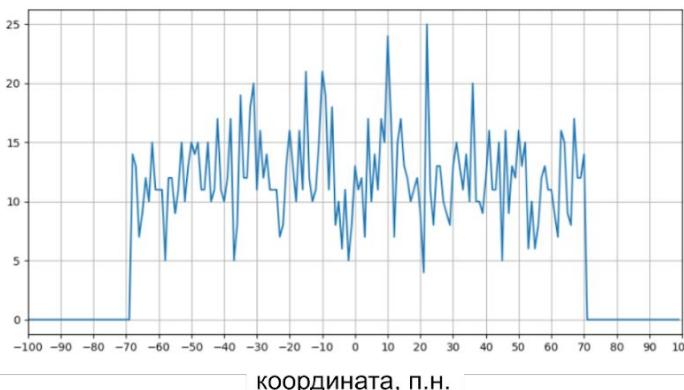
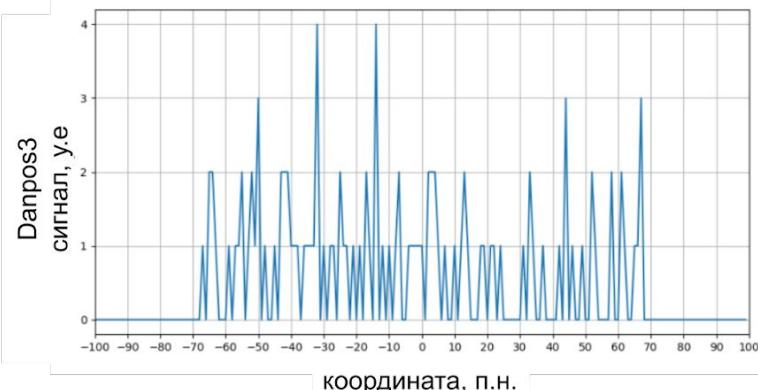


RAP1

эухроматин

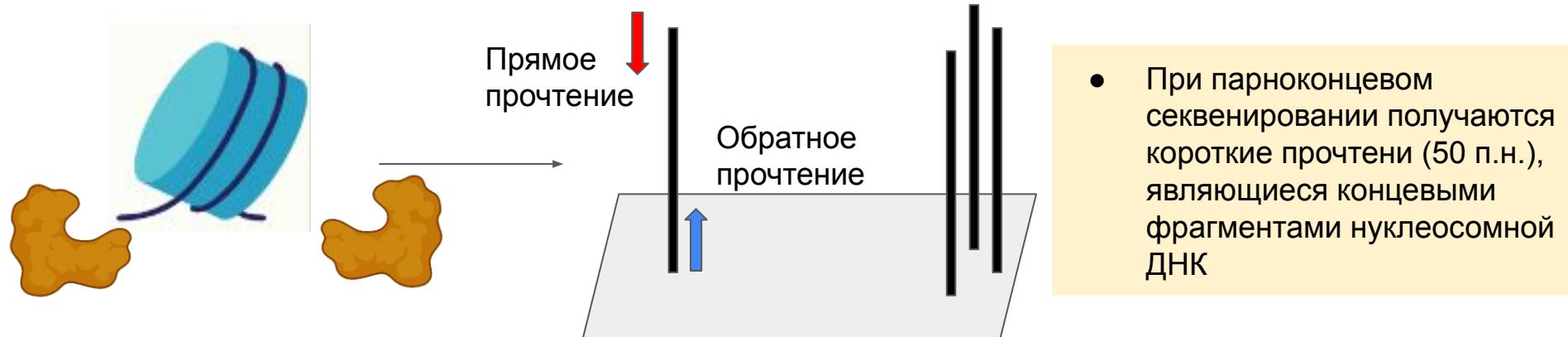


В результате анализа с использованием модели прослеживается 10 п. н. периодичность в распределении сайтов связывания ПТФ



В результате применения DANPOS3 результаты сильно зашумлены

Парноконцевое MNase-Seq-ExoIII секвенирование



парноконцевое секвенирование

Прямое чтение



Обратное чтение



AACGTAGCGATGATGACACCAGTGACGTAGCGATGATGACACCAGTGACAGTGACAGTGACGATGAC



Шаблон – участок референсного генома, который был связан с гистонами