

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИМЕНИ М.В.ЛОМОНОСОВА

БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

Программа “Геномика и здоровье человека”

**Солоколизация посттрансляционных модификаций гистонов
человека: анализ аннотации геномов и моделирование влияния
множественных модификаций на структуру нуклеосом**

Выпускная квалификационная работа магистра

Шаряфетдиновой Александры Сергеевны

Лаборатория интегративной
биологии кафедры
биоинженерии биологического
факультета МГУ

Научные руководители: канд.
физ.-мат. наук,

вед. научн. сотр. кафедры
биоинженерии

Г.А. Армеев,

асп. четвертого г.о. кафедры
биоинженерии

А.С. Федулова

Москва – 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
1. ОБЗОР ЛИТЕРАТУРЫ	4
1.1. Структура нуклеосом и гистоновых хвостов и их взаимодействие с ДНК	4
1.2. Посттрансляционные модификации гистоновых хвостов и их влияние на функции нуклеосом	5
1.3. Эпигенетические состояния хроматина	9
1.4. Методы исследования посттрансляционных модификаций	
1.5. Важность учета эпигенетических состояний хроматина при редактировании генома с помощью системы CRISPR-Cas	13
2. МЕТОДЫ И МАТЕРИАЛЫ	16
3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	19
3.1. Исследование эпигенетических состояний хроматина на основе анализа солюколизации посттрансляционных модификаций	19
3.2. Оценка влияния множественных модификаций на структуру нуклеосом	22
4. ЗАКЛЮЧЕНИЕ	31
5. ВЫВОДЫ	32
6. СПИСОК ЛИТЕРАТУРЫ	33
7. ПРИЛОЖЕНИЕ	38

ВВЕДЕНИЕ

Нуклеосома – структурная единица упаковки ДНК в клетке. Она представляет собой белковый октамер, вокруг которого обернуто около 147 пар нуклеотидов ДНК. В состав октамера входят по две пары белков - гистонов: H2A, H2B, H3 и H4, каждый из которых имеет на концах неупорядоченные последовательности – гистоновые хвосты. Гистоны и гистоновые хвосты принимают участие в таких важных процессах как раскручивание ДНК при транскрипции генов, взаимодействие с факторами хроматина, обеспечение термодинамической стабильности нуклеосомы и других. Такой широкий спектр функций гистоновых хвостов обеспечивается за счет наличия гистоновых вариантов, в которых последовательность аминокислот отличается от канонической, и за счет посттрансляционных модификаций гистонов (ПТМ). ПТМ гистонов могут присутствовать на нуклеосоме в единичном количестве, но чаще они присутствуют одновременно, в разных комбинациях. Солокализация ПТМ является дополнительным слоем регуляции активности хроматина и служит одним из методов для выделения его эпигенетических состояний.

Первой целью данной работы является исследование эпигенетических состояний хроматина на основе анализа солоколизации посттрансляционных модификаций и оценка применимости этого подхода для решения задач биоинженерии. Второй – при помощи анализа ковстречаемости ПТМ и молекулярного моделирования выявить влияние множественных модификаций на структуру нуклеосом.

Для достижения поставленных целей были поставлены следующие задачи:

1. Используя функциональную разметку генома человека (ChromHMM), определить геномные локусы, подходящие для задач инженерной биологии (активации/репрессии генов с помощью системы CRISPR-Cas)
2. На основе анализа ковстречаемости ПТМ выделить комбинации модификаций, совместно встречающихся в геномных локусах
3. Провести сравнительный анализ структуры и динамики нуклеосом, содержащих комбинации различных ПТМ, методами молекулярного моделирования.

Данная работа привносит вклад в изучение комбинаций ПТМ, их влияния на эффективность редактирования генома и динамику нуклеосомы.

ОБЗОР ЛИТЕРАТУРЫ

Структура нуклеосом и гистоновых хвостов и их взаимодействие с ДНК

Генетическая информация в клетках высших эукариот заключена в молекуле ДНК, которая в комплексе с РНК и белками находится в виде хроматина в ядре клеток. Хроматин иерархически организован в ядре, то есть имеет несколько уровней компактизации (рис.1А) [51]. Структурной единицей хроматина является нуклеосома. ДНК упаковывается в нуклеосомы с помощью специализированных белков, называемых гистонами. К этим белкам относятся гистоны H2A, H2B, H3 и H4 и линкерные гистоны H1. Нуклеосома состоит из двух копий каждого из четырех семейств гистонов и ~146 пар оснований ДНК, обернутых на 2 неполных оборота вокруг гистонного октамера (Рис.1Б).

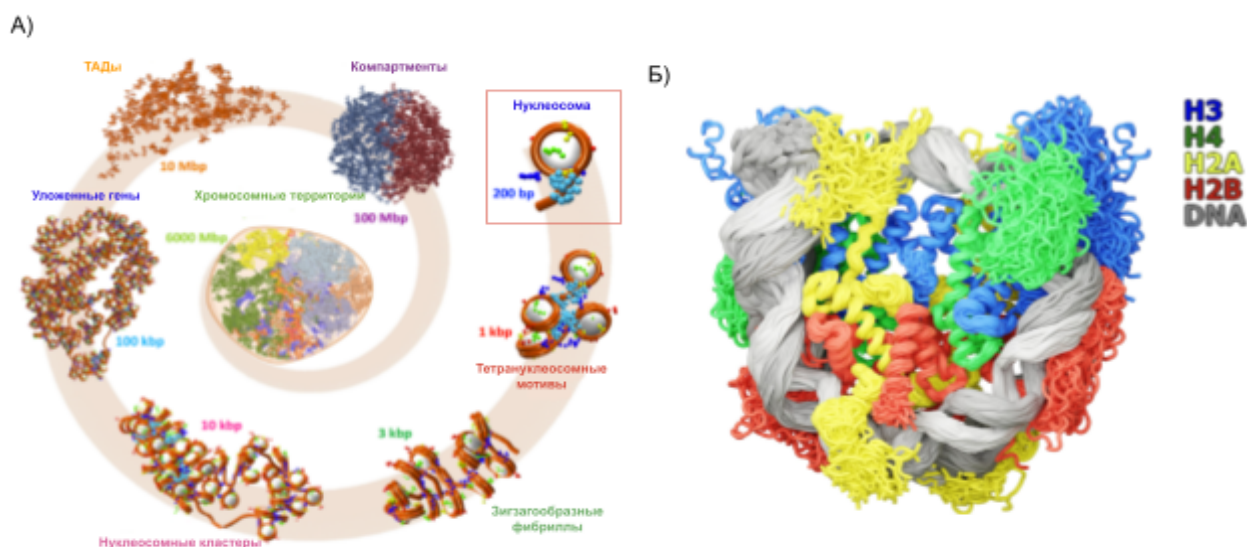


Рисунок 1. А) Современное представление об организации структуры хроматина [51], Б) Структура нуклеосомы. Типы гистонов и ДНК показаны цветом в соответствии с легендой. Все структуры изображены в виде ансамбля конформационных состояний, обнаруженных методом молекулярного моделирования [3].

Гистоновые белки представляют собой упорядоченные структуры, на N-конце которых (у гистона H2A на N- и C-концах) находятся неупорядоченные домены, называемые гистоновыми хвостами. Гистоновые хвосты характеризуются длиной от десятка до нескольких десятков аминокислот и обилием положительно заряженных остатков, особенно лизинов. Благодаря положительному заряду, хвосты могут

формировать электростатические взаимодействия с отрицательно заряженной ДНК. Однако многочисленные экспериментальные данные указывают на то, что гистоновые хвосты в растворе остаются неупорядоченными и не поддерживают постоянную связь с нуклеосомной ДНК. Данные взаимодействия в литературе называют “fuzzy interactions”, или кратковременные электростатические взаимодействия [9]. Они формируются за счет нестабильных и не долгоживущих водородных связей и солевых мостиков между аминокислотными остатками гистонов и ДНК, которые на данный момент не до конца изучены. Было высказано предположение, что такие взаимодействия могут способствовать быстрому сигналингу (передаче информации) в хроматине [8]. Механизм осуществления такого сигналинга не до конца известен, но предполагается, что он происходит за счет кратковременных электростатических взаимодействий между гистоновыми хвостами и ДНК. Хвосты минимизируют потерю конформационной энтропии при связывании с отрицательно заряженными молекулами, что способствует присоединению гистон модифицирующих ферментов к сайтам ПТМ и формированию комплексов белков при участии гистоновых хвостов [8].

Гистоновые хвосты имеют ряд важнейших функций: они участвуют в регуляции транскрипции и экспрессии генов, осуществляют аллостерическую регуляцию факторов хроматина [13] и влияют на стабильность нуклеосомы [15]. Посттрансляционные модификации (ПТМ) аминокислотной последовательности гистонов и замена канонических гистонов их вариантами расширяют спектр функциональных активностей гистонов. Варианты гистонов могут значительно отличаться по последовательности от своих канонических аналогов, что приводит к изменению множества ДНК-гистоновых и гистон-гистоновых взаимодействий [5]. Эти изменения влияют на характеристики как отдельных нуклеосом, так и хроматиновых нитей. Кроме того, варианты гистонов также могут подвергаться внесению ПТМ, что приводит к уникальным эффектам, отражающимся на динамике нуклеосом [35]. В следующем разделе будут подробнее рассмотрены эпигенетические особенности гистонов и их хвостов.

Посттрансляционные модификации гистоновых хвостов и их влияние на функции нуклеосом

ПТМ присутствуют как в хвостовых доменах, так и в их глобулярных доменах гистонов (Рис.2). Различные гистоновые ПТМ могут иметь 2 механизма воздействия на стабильность и динамику нуклеосомы: прямой механизм и косвенный [2].

Гистоновые ПТМ, имеющие прямой механизм воздействия, вызывают геномный ответ, например, активацию транскрипции, в основном за счет локальных структурных изменений хроматина [11]. Гистоновые ПТМ, действующие непрямым образом, требуют промежуточного шага, такого как связывание эффекторного белка или комплекса ремоделирования хроматина, для оказания воздействия на нуклеосому [14].

Однако гистоновые ПТМ также могут быть не только причиной изменения активности хроматина, но и следствием. Например, в процессе транскрипции полимеразы взаимодействует с ацетилтрансферазой гистонов H4 (НАТ), что способствует последующему ацетилированию гистонов [44]. Такие ПТМ можно считать следствием процесса транскрипции. В естественных условиях все описанные механизмы влияния на геномную активность вследствие внесения ПТМ действуют согласованно, чтобы обеспечить весь спектр эпигенетической регуляции хроматина.

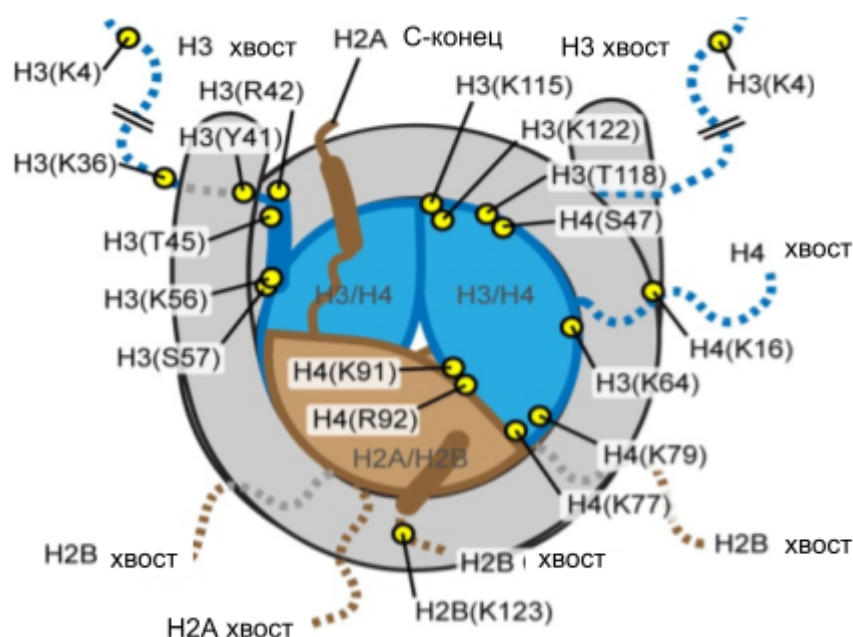


Рисунок 2. Примеры расположения некоторых ПТМ гистонов [2].

На данный момент наиболее изученными ПТМ гистонов являются ацетилирование, метилирование, фосфорилирование, убиквитинирование и АДФ-рибозилирование. Эти метки и другие изменяют силу ДНК-гистоновых и гистон-гистоновых взаимодействий, что впоследствии изменяет динамику нуклеосомы [2].

Метилирование гистонов, которое обычно происходит на остатках лизина (K) гистонов H3 и H4 путем добавления метильных групп, является одной из важнейших

посттрансляционных модификаций. Это метилирование катализируется гистон-метилтрансферазой (HMT), которая использует S-аденозилметионин (SAM) в качестве субстрата для переноса метильных групп на остатки лизина гистонов [46]. Остатки лизина гистонов могут быть моно-, ди- и триметилированными (me1, me2 и me3 соответственно), чтобы действовать как активные или репрессивные метки экспрессии генов. Метилирования H3K4, H3K36 и H3K79 считаются активными метками, которые занимают активно транскрибируемые области генов в хроматине. H3K9me3, H3K27me3 и H4K20me известны как репрессивные метки, которые обычно связаны с подавленной экспрессией генов и конденсированным хроматином [16].

Ацетилирование гистонов, чаще всего остатков лизина, уменьшает положительный заряд гистонового хвоста, что приводит к ослаблению взаимодействия между хвостами и отрицательно заряженной ДНК и открытию ДНК для транскрипционных и других факторов хроматина [20]. Поэтому ацетилирование гистонов обычно рассматривается как активная метка хроматина. Существует множество остатков лизина, которые могут быть ацетилированы в гистонах, включая ацетилирования H3K4, H3K9, H3K27, H3K36, H3K79, H4K5, H4K12, H4K20, каждое из которых выполняет свою функцию в клетке [64]. Более того, ацетилирование и метилирование соседних остатков лизина могут действовать как антагонисты, ингибируя друг друга [62]. Подобные явления описываются как "взаимодействия ПТМ" (crosstalk) между различными метками гистонов. К этому понятию также относятся случаи, когда модификации соседних остатков создают сайты связывания для белковых комплексов, усиливая их связывание с нуклеосомой [45], или наоборот, когда модификации соседних остатков ингибируют связывание различных факторов [30].

Путем привлечения клеточных факторов, которые связываются с гистонами, маркированными ПТМ («гистоновый код»), определяются и поддерживаются функционально различные области генома. ПТМ гистонов часто находятся в различных геномных местах и особенно в генах, где их присутствие коррелирует (положительно или отрицательно) с транскрипционной активностью. Модификации в промоторных областях играют ключевую роль в регуляции транскрипции, определяя доступность ДНК для транскрипционного аппарата. Наиболее характерными для промоторов являются модификации, связанные с активацией генов, такие как ацетилирование гистонов H3 (H3K9ac, H3K27ac) и H4 (H4K16ac), которые ослабляют ДНК-гистоновые взаимодействия и тем самым облегчают связывание транскрипционных факторов и РНК-полимеразы II с промотором. Другой важной модификацией промоторов и стартов начала транскрипции

(Transcription Start Site - TSS) является триметилирование гистона H3 по 4 лизину (H3K4me3) [55], которое служит маркером активных или потенциально активируемых промоторов. Для неактивных промоторов в свою очередь характерной модификацией является метилирование H3K27me3 [18], которое привлекает белки Polycomb-группы, что приводит к подавлению экспрессии генов. Также в эмбриональных клетках существуют бивалентные промоторы, для которых характерны и метки активных, и метки неактивных промоторов. Так ПТМ формируют динамичный эпигенетический ландшафт промоторов и обеспечивают точную регуляцию транскрипции в ответ на клеточные сигналы.

Активные энхансеры обогащены гистоновыми метками H3K4me1 и H3K27ac, которые зачастую встречаются одновременно, или сосуществуют, на одной нуклеосоме [52], а репрессированные энхансеры ассоциированы с меткой H3K27me3 [17].

В телах генов превалирует метилирование H3K36, которое обычно рассматривается как активная метка внутригенных и межгенных областей. В активно транскрибируемых телах генов метилирование H3K36 постепенно переходит от монометилирования к триметилированию через промотор к 3' концу [58].

Хоть единичные ПТМ и встречаются в клетке, но чаще всего они присутствуют на нуклеосоме вместе с несколькими другими модификациями (Рис.3). Комбинации ПТМ имеют важное значение для белков, связывающихся с хроматином, и нижестоящих белков-эффекторов, которые впоследствии выполняют определенные клеточные функции, формируется многоступенчатый сигналинг [34]. К примеру, H3S10ph рекрутирует белок MOF, что в дальнейшем приводит к ацетилированию H4K16. Обе ПТМ необходимы для связывания фактора BRD4 и впоследствии для элонгации транскрипции [65]. Хотя функционально связанные комбинации модификаций часто существуют на одной и той же субъединице гистона, эти связи могут быть между разными субъединицами, разными членами семейства гистонов и даже между соседними нуклеосомами.

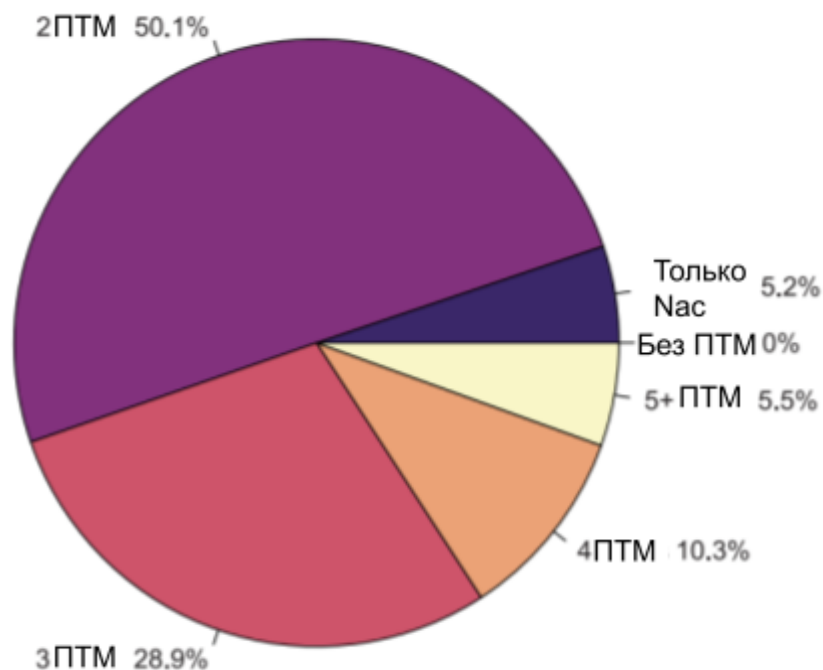


Рисунок 3. Распределение комбинаций PTM гистона H4 в культуре клеток SUM159 [60].

Эпигенетические состояния хроматина

Хроматин, который раньше принято было подразделять только на гетерохроматин и эухроматин, в реальности имеет намного больше эпигенетических состояний. Существует множество подходов к выделению этих состояний. Например, аннотация хроматина на основе уникальной комбинации белков хроматина была продемонстрирована для генома клеточной линии *Drosophila melanogaster* Kc167 [25]. Для данной клеточной линии было охарактеризовано 5 состояний хроматина, которые были названы “цветами хроматина”: черный, красный, желтый, синий и зеленый. Эти пять состояний отличаются широким спектром характеристик, таких как биохимические свойства, транскрипционная активность, модификации гистонов и время репликации, а также свойствами последовательностей и функций попадающих в эти состояния генов. Черный хроматин покрывает около половины генома и демонстрируют чрезвычайно низкие уровни экспрессии. Красный и желтый — различные типы эухроматина, которые различаются модификацией H3K36me3 и располагающимися в них генами: в красном хроматине преобладают тканеспецифичные гены и отсутствует H3K36me3, в то время как в желтом хроматине преобладают гены домашнего хозяйства и присутствует метка H3K36me3.

Синий и зеленый — различные типы гетерохроматина. Зеленый хроматин соответствует классическому гетерохроматину, который отмечен SU(VAR)3-9, HP1 и взаимодействующий с белками LHR и HP6 [21], также для него характерна метка H3K9me2. Синий хроматин соответствует хроматину PcG, с которым связываются белки PcG PC, E(Z), PCL и SCE и присутствует ПТМ H3K27me3.

Другой популярный подход для выделения состояний хроматина - по комбинации ПТМ. Данный подход чаще применяют для изучения эпигенетических состояний хроматина человека. Разбиение генома человека на более мелкие эпигенетические состояния способствует, например, нахождению паттернов и ассоциаций с различными заболеваниями. Так были найдены вариации в состояниях регуляторных областей генов и предрасположенность к лимфоме у человека [37]. Также аннотация генома по комбинации ПТМ полезна при изучении комбинаторной и аддитивной роли хроматиновых меток, которые ранее не были описаны, и может способствовать выявлению ранее не аннотированных потенциальных функциональных элементов [23].

Методы исследования ПТМ

Для изучения ПТМ применяются как экспериментальные, так и биоинформатические подходы.

Подходы на основе использования антител

Подходы на основе использования антител (иммунопреципитация хроматина (ChIP-seq), вестерн-блоттинг, иммунофлуоресценция) преобладают в исследованиях ПТМ, во многом из-за повсеместности применения и скорости проведения иммуноанализов. Такие методы позволяют не только качественно оценить присутствие или отсутствие различных модификаций, но и определить их локализацию в геноме (ChIP-seq), что важно для изучения эпигенетической регуляции [47]. Однако следует отметить, что эти подходы ограничены анализом одного ПТМ, не предоставляя информации о его распространенности или стехиометрии. Кроме того, на результат анализа сильно влияет степень загрязненности или степень специфичности антител. Согласно данным об исследовании гистоновых ПТМ иммуноаффинными методами, более 25% всех антител, используемых в подобных экспериментах, не прошли тесты на специфичность [22, 26], что впоследствии привело к искажениям в результатах анализа. Солокализованные ПТМ также нарушают специфичность антител, в результате чего занятость соседних остатков ПТМ может изменить распознавание эпитопа. Эта проблема особенно актуальна при

исследовании гистонов, где N-концевые сегменты сильно модифицированы разнообразным набором ПТМ.

Подходы на основе масс-спектрометрии (МС)

Масс-спектрометрия (МС) стала одним из ключевых инструментов в протеомике и эпигенетике, особенно для изучения ПТМ белков, включая гистоны. Самым популярным подходом на основе МС сегодня является ЖХ-МС/МС (LC-MS/MS) [36], где белки протеолитически расщепляются до пептидов и разделяются с помощью жидкостной хроматографии (ЖХ), а затем анализируются с помощью тандемной масс-спектрометрии (МС/МС). Поскольку ПТМ обычно встречаются в небольшом количестве по сравнению с их немодифицированными аналогами, обогащение образца, в основном с использованием иммуноаффинных колонок, почти всегда необходимо для получения полной картины профиля ПТМ [27]. МС позволяет детектировать ПТМ на уровне пикомольных (10^{-12}) и даже фемтомольных (10^{-15}) количеств, а также обеспечивает точность и возможность детектировать множественные ПТМ [28], например, комбинации ацетилирования и метилирования гистонов, что недоступно для многих других методов, таких как иммуноблоттинг. Однако у этих методов также есть свои ограничения, такие как сложность пробоподготовки, деструктивность, сложность в интерпретации данных, и ограничения в детекции некоторых модификаций из-за их лабильности или низкой стабильности в условиях МС [54].

Биоинформатические подходы

В последние годы биоинформатические подходы, особенно методы вычислительного моделирования, такие как молекулярная динамика (МД) и другие, стали мощным инструментом для исследования ПТМ на атомном уровне, предлагая новые возможности для понимания их влияния на структуру и функцию белков. Биоинформатика предоставляет широкий спектр инструментов для изучения ПТМ. К наиболее распространенным биоинформатическим подходам относятся: предсказание сайтов ПТМ, аннотация генома, систематизация уже полученной информации и создание баз данных (БД), моделирование структуры белков и молекулярная динамика (МД). Для предсказания сайтов ПТМ используются алгоритмы машинного обучения и анализа последовательностей, например MusiteDeer [59], для выявления потенциальных сайтов модификации на основе известных мотивов и контекста аминокислот вокруг сайта.

Аннотация генома подразделяется на несколько подходов: на структурную аннотацию, где проводится поиск генов, их экзонов, интронов и промоторов,

транспозонов и т. д., и функциональную аннотацию, целью которой является определение биологических функций генов и их эволюционных взаимосвязей. Методы для прогнозирования генов можно классифицировать на три основные группы: методы *ab initio*, методы, основанные на гомологии, и комбинированные методы, которые сочетают в себе предыдущие два подхода. Методы *ab initio*, основаны исключительно на анализе нуклеотидной последовательности и полагаются на статистические модели, такие как скрытые марковские модели (НММ), для распознавания ключевых элементов гена. В отличие от *ab initio*, методы, основанные на гомологии, используют информацию о сходстве последовательности с другими известными генами или транскриптами. Они проводят выравнивание геномной последовательности с экспрессированными метками последовательности (EST), комплементарной ДНК (кДНК) или данными о белках, а затем используют обнаруженные области сходства для предсказания структуры и положения генов [63].

Одной из широко используемых программ для аннотации геномов, в том числе генома человека, является ChromHMM [24]. ChromHMM – это программное обеспечение, использующее НММ и данные ChIP-seq ПТМ для аннотации геномных областей. Основной принцип работы ChromHMM заключается в следующем: за “состояния” модели принимаются неперекрывающиеся участки фиксированной длины генома, “состояния” соотносятся с входными данными в виде сигналов ChIP-seq для различных гистоновых модификаций и после обучения на основе НММ каждому состоянию присваивается одно из эпигенетических состояний хроматина (скрытые состояния). Каждое скрытое состояние представляет собой уникальную комбинацию гистоновых модификаций и ассоциируется с определенной функцией (например, активный промотор, энхансер, репрессированная область). Количество состояний является заданным пользователем параметром. Эмиссионные вероятности в этой модели отражают, насколько возможно наблюдать определенный уровень сигнала ChIP-seq для данной гистоновой модификации в данном состоянии. Переходные вероятности определяют вероятность перехода между различными состояниями. Они отражают тенденцию хроматиновых состояний сохраняться в определенных геномных областях. Результаты аннотации отображаются в виде цветных дорожек вдоль генома, где каждый цвет соответствует определенному состоянию хроматина.

Молекулярная динамика (МД) – это вычислительный метод, который моделирует движение атомов и молекул во времени, основываясь на законах Ньютона. Последние

экспериментальные данные показывают, что взаимодействия между ДНК и гистоновыми хвостами носят кратковременный неустойчивый характер (fuzzy interactions) и возникают на временах порядка наносекунд или десятков наносекунд. Экспериментально точно охарактеризовать эти взаимодействия сложно, поскольку они очень динамичны. В этой ситуации ценную информацию можно получить с помощью МД, где можно наблюдать процесс на временах порядка от фемтосекунд до микросекунд с высокой точностью. При планировании расчетов МД важным аспектом является выбор силового поля, которое описывает взаимодействие между атомами в системе, и модели воды. Последние исследования показали, что для моделирования неупорядоченных белковых структур, к которым относятся гистоновые хвосты, подходит сочетание силового поля Amber19SB и модели воды OPC [7]. МД позволяет изучать динамические процессы, такие как конформационные изменения, колебания и переходы, которые важны для понимания функций белков, а также эффектов ПТМ, которые трудно или невозможно изучить экспериментально. Однако МД требует значительных вычислительных ресурсов, особенно для моделирования больших систем на длительных временных масштабах.

Несмотря на большой объем экспериментальных данных в области исследования одиночных ПТМ, локализации модификаций, в частности механизмы их влияния на динамику нуклеосом, остаются неизученными.

Важность учета эпигенетических состояний хроматина при редактировании генома с помощью системы CRISPR-Cas

CRISPR-Cas система (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins), открытая как адаптивный иммунитет бактерий и архей, стала революционным инструментом геномного редактирования [10]. Благодаря этой технологии стало возможным точно модифицировать ДНК в любом организме, в том числе в клетках человека, что открыло новые горизонты в медицине, сельском хозяйстве и биотехнологии.

Система CRISPR-Cas представляет собой сложный комплекс, состоящий из двух основных компонентов: белка Cas, который является нуклеазой и способен разрезать ДНК, и гидовой РНК. В естественной системе Cas9 направляется к целевому сайту ДНК двумя некодирующими РНК: CRISPR РНК (crRNA), которая комплементарна целевой последовательности ДНК, и транскриптивирующей РНК (tracrRNA), которая необходима для связывания с белком Cas9. В синтетически реконструированной системе эти две короткие

Для увеличения экспрессии генов используется комплекс CRISPRa (CRISPR activation), который состоит из нуклеазы dCas9 и активаторов транскрипции (например, VP64, p300) [50]. В этом случае активация транскрипции гена происходит вследствие связывания комплекса dCas9-активатор-sgRNA с промотором гена. Для снижения экспрессии генов используется комплекс CRISPRi (CRISPR interference), который состоит из нуклеазы dCas9 и репрессора транскрипции (например, KRAB). Когда комплекс dCas9-репрессор-sgRNA связывается с промотором гена, он блокирует транскрипцию этого гена, уменьшая его экспрессию [41].

При использовании системы CRISPR-Cas и для редактирования генома, и для задач изменения экспрессии генов, важно учитывать эпигенетический контекст хроматина. Некоторые состояния хроматина, такие как конститутивный гетерохроматин, могут препятствовать связыванию комплекса с ДНК и разрезанию таргетного локуса [61]. Таким образом, можно подчеркнуть важность локального контекста хроматина, кодируемого различными комбинациями ПТМ, для редактирования с помощью CRISPR-Cas, поскольку он может влиять на эффективность редактирования, на баланс между путями репарации и, следовательно, на вероятность получения желаемого результата [31].

Хоть технологии с использованием CRISPR-Cas и являются передовыми, они имеют ряд недостатков, например, запуск апоптоза при внесении двухцепочечных разрывов в ДНК [29], вместо редактирования гена, или случайные протяженные делеции [40]. В то же время применение других вариаций Cas9, таких, как dCas9, снижает риски появления двунитевых разрывов, тем самым увеличивая потенциальную терапевтическую полезность [57]. Изменение экспрессии генов с помощью системы CRISPR-Cas вносит значительный вклад в развитие генной терапии и регенеративной медицины.

В работе представлены результаты по анализу аннотации генома человека, полученной с помощью ChromHMM, выбору таргетных локусов для задачи репрессии и активации генов (с помощью системы CRISPR-Cas) и моделированию нуклеосом с комбинацией модификациями, полученные методом МД.

МЕТОДЫ И МАТЕРИАЛЫ

Поиск геномных локусов для активации/репрессии с помощью системы CRISPR-Cas

Для исследования эпигенетических состояния хроматина на основе анализа солоколизации ПТМ и оценки применимости подхода для решения задач биоинженерии (поиска геномных локусов для активации/репрессии с помощью системы CRISPR-Cas) был использован следующий пайплайн (Рис.5):

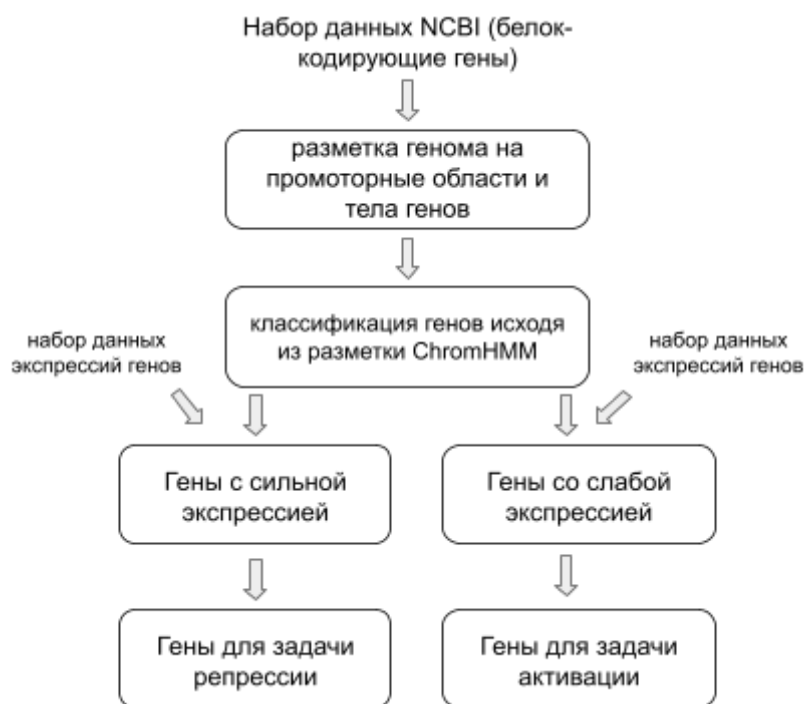


Рисунок 5. Алгоритм для поиска геномных локусов для активации/репрессии с помощью системы CRISPR-Cas.

Сначала была осуществлена разметка генома человека (GRCh38 [56]) на промоторные области и тела генов. Координаты генов были взяты из базы данных NCBI, из них для дальнейшего анализа были отобраны данные белок-кодирующих генов. Координаты тел генов были определены как старт (начало гена) и стоп (конец гена) для каждой цепи соответственно. Координаты тел генов были определены как ± 500 п.н. от начала гена [49]. Для решения поставленной задачи было принято решение не учитывать гены домашнего хозяйства (взяты из базы данных Human Protein Atlas), так как они не являются целями для изменения экспрессии с помощью системы CRISPR-Cas.

Эпигенетические состояния для исследования были взяты из данных ChromHMM для культуры Н1. Данные экспрессии для культуры Н1 были взяты из базы данных Expression Atlas.

Выбор и обработка данных ChIP-seq ПТМ

Во второй части исследования для анализа влияния пост-трансляционных модификаций (ПТМ) на конформационную динамику гистоновых хвостов были выбраны все возможные модификации гистонов, которые представлены в базе данных ENCODE для клеточной линии человека Н1 (26 ПТМ). Данная культура была выбрана в связи с наличием большого объема открытых данных о местоположении гистоновых ПТМ, полученных с помощью ChIP-seq. Для каждой из представленных в базе данных ENCODE ПТМ для культуры Н1 из каждого ChIP-seq эксперимента был скачан bed narrowPeak файл, replicated peaks. Далее скачанные файлы были обработаны с помощью библиотеки pandas: если в базе данных ENCODE для ПТМ было размещено несколько экспериментов ChIP-seq, скачанные bed файлы были сопоставлены между собой. Было произведено пересечение всех стартов пиков: если пики находились на расстоянии 300 п.н. друг от друга, они были причислены к реализующимся в повторностях пикам и им была присвоена частота встречаемости в ChIP-seq экспериментах. Данная обработка была реализована для ПТМ H3K4me3, H3K27ac, H3K27me3, H3K4me1, H3K36me3, H3K9me3, H3K9ac, H3K4me2, H4K20me1 и H3K79me2. Для остальных модификаций в базе данных ENCODE были представлены данные только одного эксперимента ChIP-seq, поэтому каждому пику из их bed-файлов была присвоена частота встречаемости в ChIP-seq экспериментах равная 1.

Для дальнейшего анализа данные всех ПТМ были разбиты на отрезки генома по 1000 п.н, объединены, перегруппированы по хромосомам и отфильтрованы по частоте встречаемости в ChIP-seq экспериментах $\geq 0,5$.

Проведение МД

Для формирования систем, состоящих из модифицированной нуклеосомы и полноразмерных гистоновых хвостов, в качестве референса была использована структура из базы данных PDB – нуклеосома *Homo sapiens* с PDB ID 1KX5 [12]. В ней представлены гистоны человека с полноразмерными хвостовыми структурами и 147 п.н. ДНК. Неупорядоченные гистоновые хвосты были симметризованы путём приведения конформаций дистальной половины нуклеосомы (цепей E, F, G, H) в соответствие с их проксимальными аналогами (цепями A, B, C, D) [1]. Ацетилирования были внесены с

помощью программы AMBER. Метилирования (моно-, ди-, три-) были параметризованы с помощью программного конвейера, разработанного ранее в группе, внесены в силовое поле AMBER19SB и затем использованы для модификации систем.

На системах с комбинациями модификаций была проведена молекулярная динамика в силовом поле AMBER19SB OL21 [38] с помощью программы AMBER [19]. При подготовке системы в ее расчетную ячейку были добавлены молекулы воды, ионы Na⁺ и Cl⁻ и была проведена минимизация и эквilibрация. При расчете была использована четырехточечная модель воды OPC, подходящая для моделирования внутренне-неупорядоченных участков белков. Моделирование проводилось в NpT ансамбле (постоянная температура, давление и число частиц). Для этого использовался термостат v-rescale и баростат Бусси, поддерживающие постоянными температуру 300 К и давление 1 бар. Шаг моделирования – 2 фс, шаг записи кадров – 10 пс, длины траекторий – 1 мкс и 0.7 мкс (для системы с метилированием). Для каждой системы проводился расчет двух повторностей.

Сравнение модифицированных систем проводилось с микросекундной траекторией немодифицированной нуклеосомы, которая ранее была получена в группе, в окружении воды OPC и с использованием описанного выше протокола.

Полученные траектории анализировались с помощью библиотек numpy, matplotlib, pandas, ngview, MDAnalysis [4, 6], pynuc1, seaborn, scipy и с помощью разработанных скриптов, написанных на языке Python 3. Для визуализации траекторий использовались программы ngview и PyMOL [42]. Перед проведением анализа была осуществлена обработка траекторий с помощью программы Gromacs (восстановление целостности молекул с учетом периодических граничных условий, помещение нуклеосомы в центр расчетной ячейки и выравнивание ротационной подвижности). Дополнительно проводилось выравнивание траекторий с помощью пакета MDAnalysis по положениям Ca-атомов α -спиралей гистонов (элементов гистонического фолда).

Для исследования влияния комбинаций ПТМ на динамику нуклеосомы были проанализированы следующие параметры: радиус гирации гистонов (без учета хвостов), среднее количество атом-атомных взаимодействий между ДНК и гистоновыми хвостами, среднеквадратичное отклонение гистоновых хвостов и глобулярных участков гистонов (петель и α -спиралей) и пластичность нуклеосомы. Для расчета радиуса гирации и среднеквадратичных отклонений атомов использовался пакет MDAnalysis. Расчет атом-атомных контактов проводился с помощью разработанной в группе библиотеки

Рисунок 1. За контакт между двумя атомами было обозначено событие, когда атомы находятся на расстоянии $\leq 4\text{\AA}$ друг от друга (исключая атомы водорода). Для анализа пластичности нуклеосомы проводилось построение двумерных проекций положений Са-атомов α -спиралей гистонов на плоскость нуклеосомы.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Исследование эпигенетических состояний хроматина на основе анализа солоколизации посттрансляционных модификаций

Чтобы проверить адекватность разметки генома культуры Н1 с помощью программы ChromHMM, были проанализированы следующие характеристики: процент покрытия генома полученными из ChromHMM состояниями, процент покрытия длины тела гена состояниями ChromHMM, характерными для тел генов и процент покрытия длины промотора гена состояниями ChromHMM, характерными для промоторов генов.

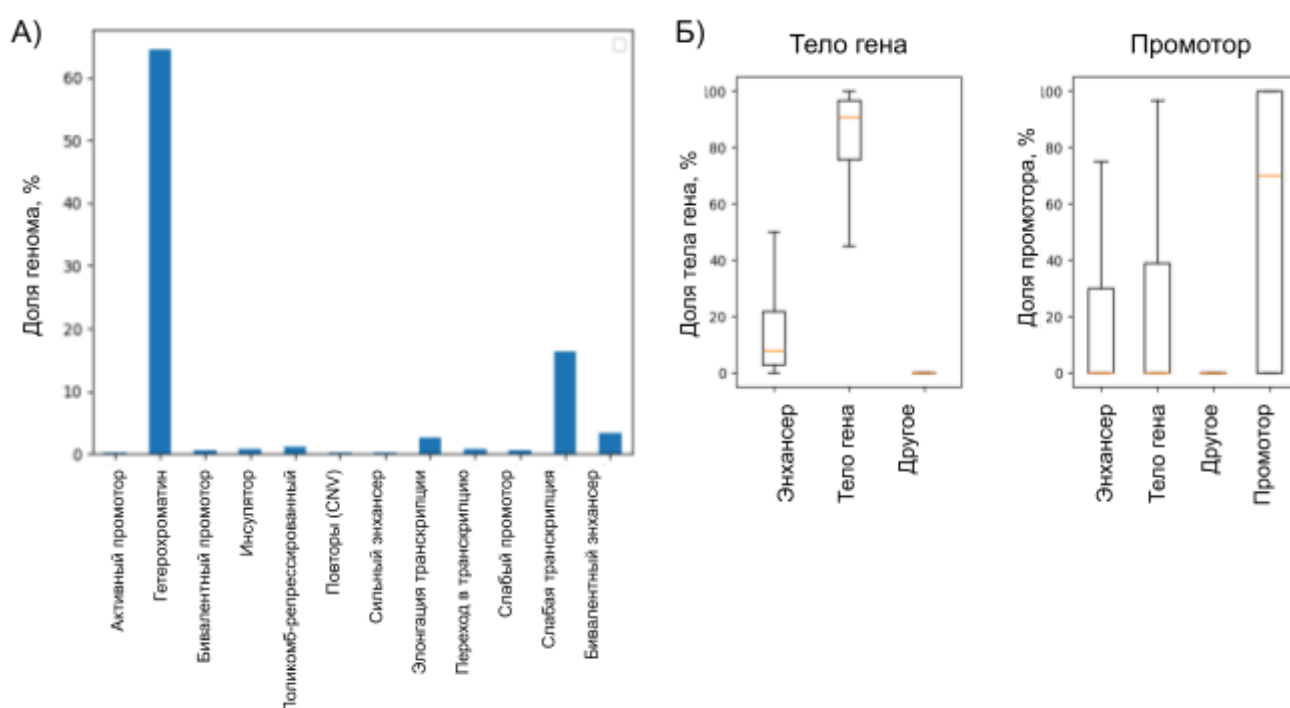


Рисунок 6. А) Процент покрытия генома полученными из ChromHMM состояниями, Б) процент длины тела гена состояниями ChromHMM, характерными для тел генов (слева) и процент длины промотора гена состояниями ChromHMM, характерными для промоторов генов (справа).

Согласно полученным из ChromHMM предсказаниям, состояние “Гетерохроматин”, характеризующее конститутивный гетерохроматин, занимает 64,4% от длины генома человека (Рис.6А), состояние “Поликомб-репрессированный хроматин”, характеризующее факультативный гетерохроматин, - 1,2%, “Слабая транскрипция” - 16,3%.

В среднем 91,1% от длины тела гена приходится на состояния, характерные для генов по ChromHMM (Гетерохроматин, Слабая транскрипция, Поликомб-репрессированный хроматин, Переход в состояние транскрипции, Элонгация транскрипции) (Рис.6Б). В среднем 74,6% от длины промотора гена приходится на состояния, характерные для промотора по ChromHMM (Активный промотор, Бивалентный промотор, Слабый промотор). Исходя из этих результатов можно сделать вывод, что ChromHMM на культуре Н1 работает хорошо.

Однако при более детальном анализе было выявлено, что большая часть генов (и генов домашнего хозяйства, и генов, не относящихся к домашнему хозяйству) относится и к активной группе генов, и к репрессированной группе генов (Рис.7А,Б). К состояниям, характеризующим активное состояние, были отнесены: Активный промотор, Переход в состояние транскрипции и Элонгация транскрипции. К состояниям, характеризующим репрессированное состояние, были отнесены: Слабый промотор, Гетерохроматин, Слабая транскрипция и Поликомб-репрессированный хроматин. Поэтому было принято решение разделить “Активное состояние гена” на “Активное состояние гена судя по промотору” и “Активное состояние гена судя по телу гена”.

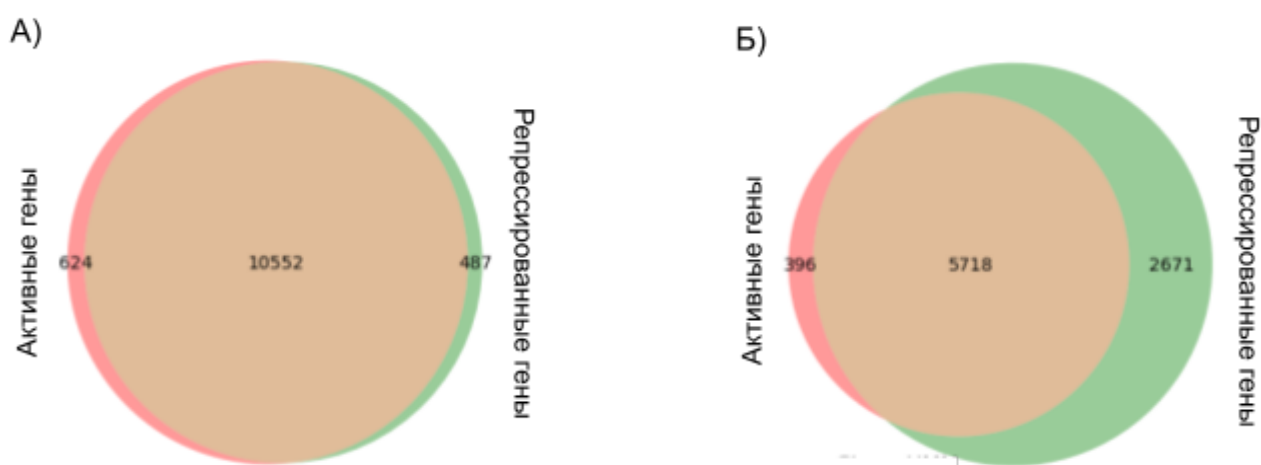


Рисунок 7. Количество генов, классифицирующихся как “Активные” и “Репрессированные”, согласно разметке ChromHMM. А) Для не генов домашнего хозяйства, Б) для генов домашнего хозяйства.

В результате применения данных эпигенетических состояний культуры Н1, полученных с помощью ChromHMM, для исследования принадлежности генов к активному или репрессированному состояниям были выделены 3 класса генов: 1) гены, промоторные области которых пересекались с активными состояниями промоторов

(Активный промотор, Бивалентный промотор) согласно ChromHMM; и тела генов которых пересекались с активными состояниями для тел генов (Переход в состояние транскрипции, Элонгация транскрипции) согласно ChromHMM; 2) гены, промоторные области которых пересекались с репрессированными состояниями для промоторов (Слабый промотор, Бивалентный промотор) согласно ChromHMM и тела генов которых пересекались с репрессированными состояниями для тел генов (Гетерохроматин, Слабая транскрипция, Поликомб-репрессированный хроматин) согласно ChromHMM; 3) гены, промоторные области которых пересекались с активными состояниями для промоторов согласно ChromHMM и тела генов которых пересекались с репрессированными состояниями для тел генов согласно ChromHMM, и наоборот. Количество генов, принадлежащих к первой группе, — 158, ко второй — 1095.

Для сравнения генов из полученных групп с данными экспрессии, были выделены 4 группы экспрессии, различающиеся по величине TPM (Transcripts Per Million): группа 1 - $0 < \text{TPM} < 1$, группа 2 - $1 < \text{TPM} < 10$, группа 3 - $10 < \text{TPM} < 100$, группа 4 - $100 < \text{TPM} < 1000+$. Исходя только из данных экспрессии, 3063 генов принадлежит первой группе, 2919 - второй группе, 1654 - третьей группе, 2 - четвертой группе. Второй группе соответствуют следующие контрольные гены (по контрольным генам измеряются изменение экспрессии целевых генов): *SDHA* (TPM 7.0); третьей группе: *TBP* (TPM 27.0), *HPRT1* (29.0), *HMBS* (TPM 62.0); четвертой группе *B2M* (TPM 178.0), *YWHAZ* (TPM 203.0), *UBC* (TPM 388.0), *ACTB* (TPM 3572.0), *RPL13A* (TPM 3897.0).

С учетом данных экспрессии количество генов для задачи репрессии - 46 из группы 3, а для задачи активации - 774 (539 генов из группы 1 и 235 генов из группы 2 TPM) (Рис. 8).

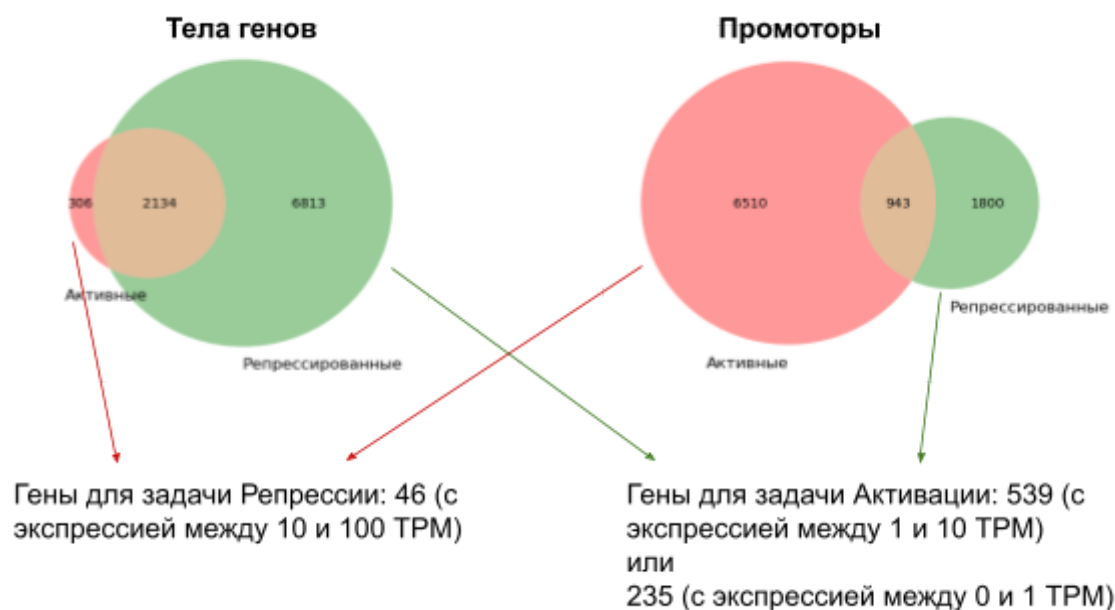


Рисунок 8. Количество генов для задачи репрессии и активации. Слева изображена диаграмма Венна для тел генов, справа — для промоторов генов.

Оценка влияния множественных модификаций на структуру нуклеосом

Для определения одновременной встречаемости групп ПТМ в геноме был проведен анализ пересечений областей генома по данным ChIP-seq. Известно, что наиболее четко позиционированной в геноме является, так называемая, +1 нуклеосома, расположенная в начале генов (примерно 50-60 п.н. от TSS) [53]. Исходя из данных о картировании нуклеосом, были выделены +1 нуклеосомы и по ним проведен анализ совместной встречаемости ПТМ.

При анализе ChIP-seq данных ПТМ были выявлены следующие самые часто встречающиеся комбинации модификаций в эмбриональной культуре человека H1. Комбинации по 2 ПТМ: H3K18acH4K91ac - 42928, H3K4me2H3K18ac - 42373, H3K4meH3K18ac - 37430, H3K18acH2BK12ac - 36822, H2BK12acH4K91ac - 36035, H4K91acH2BK20ac - 35044, H3K18acH2BK5ac - 34792, H3K4me2/3H3K27me3 - 34251. Полная карта совместной парной встречаемости ПТМ гистонов представлена на Рис.9А. Комбинации по 3 ПТМ: H3K18acH2BK12acH4K91ac - 25302, H3K4me2H3K18acH4K91ac - 24515, H3K18acH2BK5acH4K91ac - 24149. Комбинации по 4 ПТМ: H3K18acH2BK12acH4K91acH2BK20ac - 16656, H3K18acH2BK12acH2BK5acH4K91ac - 15968, H3K4me2H3K18acH2BK5acH4K91ac - 15616. Полный список комбинаций

модификаций	доступен	по	ссылке
https://github.com/a-sharyafetdinova/Colocalization_of_PTMs 2025.			

Также были выделены следующие наиболее часто встречающиеся комбинации на расстоянии +/- 1000 п.н. от TSS. Комбинации по 2 ПТМ: H3K4me2/3H3K9ac - 39757, H3K4me2/3H3K18ac - 36280, H3K4me2/3H4K8ac - 33496, H3K4me2/3H3K23me2 - 32392. Комбинации по 3 ПТМ: H3K4me2/3H3K9acH3K18ac - 27854, H3K4me2/3H3K9acH4K8ac - 25805, H3K4me2/3H4K8acH3K18ac - 24086, H3K4me2/3H3K9acH2BK5ac - 23235, H3K4me2H3K9acH3K23me2 - 22889. Комбинации по 4 ПТМ: H3K4me2/3H3K9acH4K8acH3K18ac - 20322, H3K4me2/3H3K9acH3K18acH2BK5ac - 18874, H3K4me2/3H3K9acH3K18acH4K91ac - 17381. Полный список комбинаций модификаций доступен по ссылке <https://github.com/a-sharyafetdinova/Colocalization of PTMs 2025>.

Хоть в ходе анализа и были обнаружены распространенные для эмбриональной культуры комбинации модификаций, такие как, метилирование H3K4 и ацетилирование H3K27 [32], большая часть часто встречающихся комбинаций модификаций не была описана в литературе, что интересно для дальнейшего изучения.

Выбор комбинаций (преимущественно ацетилований) для внесения в нуклеосому и дальнейшего моделирования был проведен путем группировки аминокислотных кластеров на основе карты солокализации ПТМ (Рис.9А). Были выбраны следующие комбинации:

- 1) **N_{rare H3 ac}** (H3K14ac, H3K23ac)
- 2) **N_{H3 grouped ac}** (H4K5ac, H3K9ac, H3K27ac, H3K56ac)
- 3) **N_{H3-H4 grouped ac}** (H4K5ac, H3K9ac, H3K27ac, H3K56ac + H3K18ac, H4K91ac, H2BK5ac, H4K8ac)
- 4) **N_{H2B grouped ac}** (H3K18ac, H3K4ac, H4K91ac, H4K8ac, H2AK5ac, H2BK5ac, H2BK12ac, H2BK20ac, H2BK15ac, H2BK120ac)
- 5) **N_{all ac}** (все ацетилирования в группах 1-4)

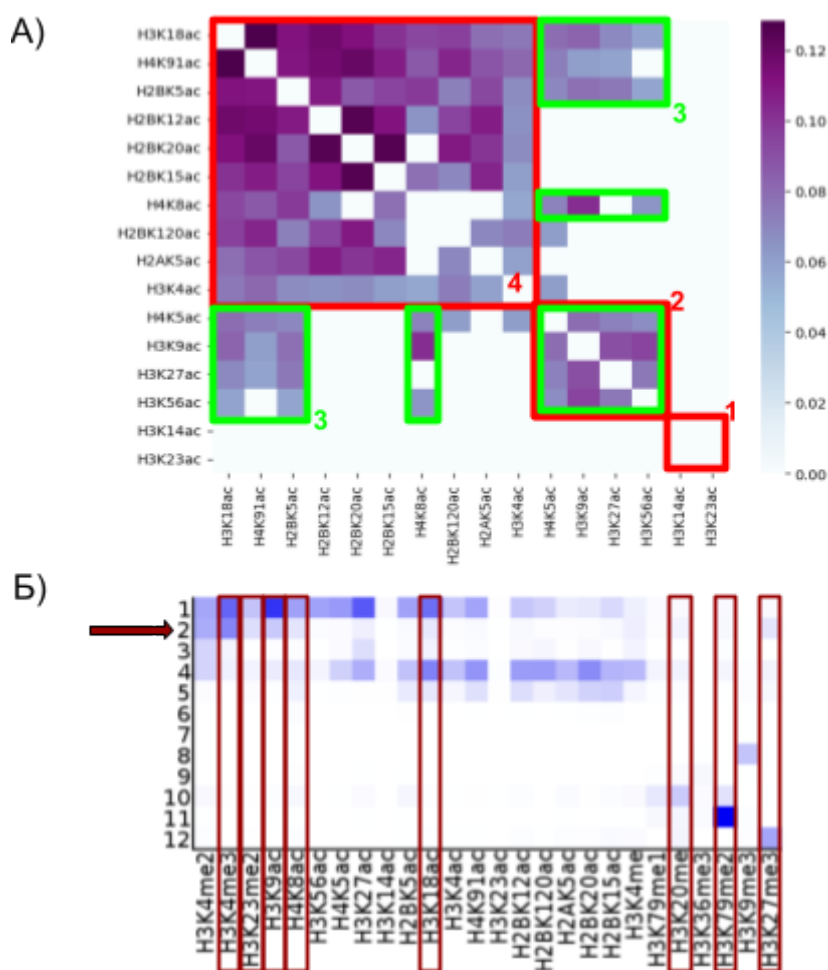


Рисунок 9. А) Карта парной ковстречаемости ацетилированных остатков гистонов для культуры Н1 Б) Карта эпигенетических состояний, полученных для культуры Н1 с помощью ChromHMM.

При выборе системы с метилированными остатками, на карте ковстречаемости не были различимы четкие кластеры, поэтому было принято решение использовать инструмент, осуществляющий аннотацию генома (ChromHMM), на культуре Н1 на 26 ПТМ. Таким образом, для моделирования системы с метилированием и ацетилированием, была выбрана система, представляющая собой одно из состояний, определенных ChromHMM (Рис.9Б).

6) $N_{ac\ met}$ (H4K9ac, H4K8ac, H3K18ac H3K4me3, H3K23me2, H4K20me, H3K79me2, H3K27me3).

Траектории для выбранных систем с модифицированными остатками были рассчитаны в двух повторностях по 1 мкс для систем $N_{rare\ H3\ ac}$, $N_{H3\ grouped\ ac}$,

$N_{H3-H4 \text{ grouped ac}}$, $N_{H2B \text{ grouped ac}}$, $N_{all \text{ ac}}$ и по 700 нс для системы $N_{ac \text{ met}}$. Интерактивный просмотр траекторий молекулярной динамики доступен по ссылке http://intbio.org/2025_PTMColocalization.io/.

На основе анализа радиуса гирации гистоновых хвостов (Рис. 10) было принято решение для дальнейшего анализа отбросить первые 200 нс из каждой траектории с комбинацией модифицированных остатков. В течение этого времени происходила конденсация гистоновых хвостов на ДНК.

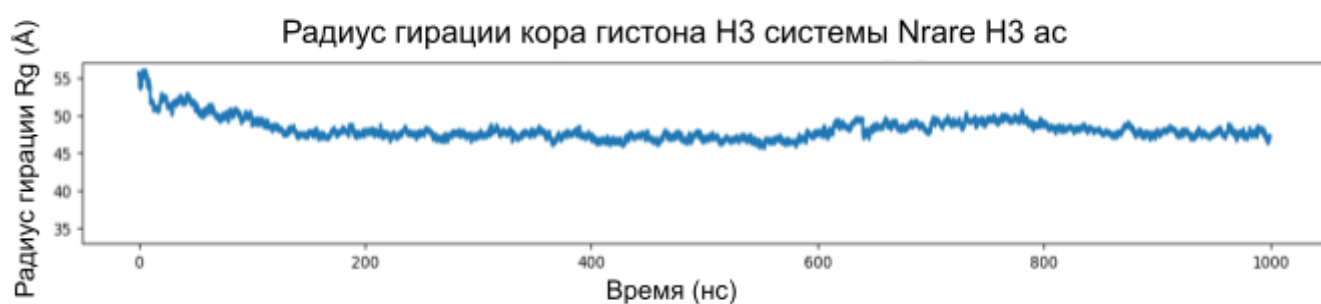


Рисунок 10. Радиус гирации гистона H3 (без учета хвостов) системы $N_{rare \text{ H3 ac}}$.

Далее приведены результаты сравнительного анализа этих траекторий.

1. Анализ среднего количества контактов ДНК с гистоновыми хвостами.

Анализ среднего количества контактов ДНК со всеми гистоновыми хвостами в ходе моделирования (Рис.11) показал, что любое множественное модифицирование приводит к уменьшению количества контактов хвостов H2B и H3 с ДНК (значимые различия наблюдаются в системах $N_{all \text{ ac}}$ и $N_{H3-H4 \text{ grouped ac}}$ для хвостов гистонов H2B и H3; и в системе $N_{H2B \text{ grouped ac}}$ для хвоста H2B). Также из этого графика видно значимое различие в среднем количестве контактов хвоста H2A-N с ДНК между системами $N_{H3 \text{ grouped ac}}$ и $N_{H2B \text{ grouped ac}}$, что может свидетельствовать о связи динамики хвоста H2A с H2B и H3.

Известно, что уменьшение контактов хвостов с ДНК при ацетилировании лизинов приводит к увеличению подвижности гистоновых хвостов [39]. Данный эффект виден как при глобальном ацетилировании гистонов ($N_{all \text{ ac}}$, $N_{H3-H4 \text{ grouped ac}}$, $N_{H3 \text{ grouped ac}}$, $N_{H2B \text{ grouped ac}}$, $N_{ac \text{ met}}$), так и при ацетилировании лишь двух остатков гистона H3 ($N_{rare \text{ ac}}$). Также из данного графика видно, что подвижность хвостов коррелирует с плотностью ацетилирования (подвижность хвоста H2B сильно возрастает в системе $N_{H2B \text{ grouped ac}}$,

подвижность хвостов H3 и H4 значительно увеличивается в системе с полным ацетилированием $N_{all\ ac}$), что согласуется с литературными данными [39].

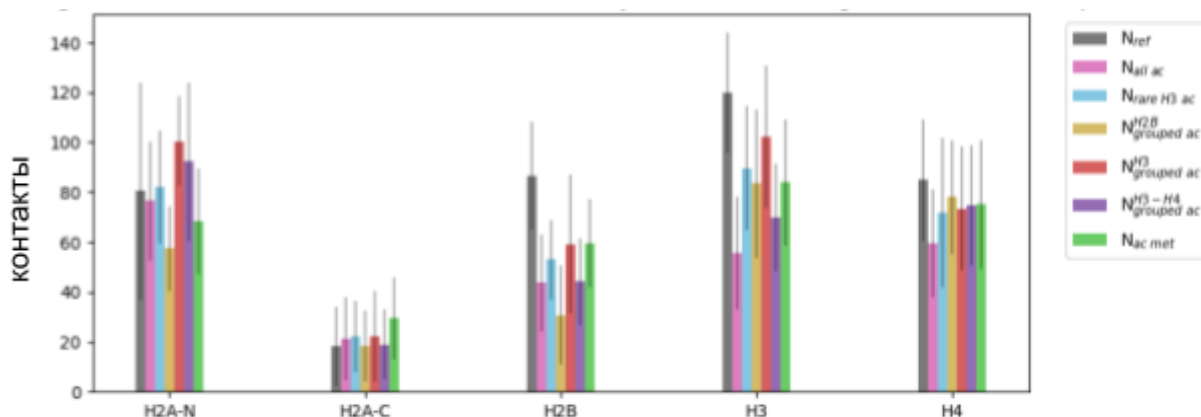


Рисунок 11. Среднее количество атомных взаимодействий ДНК с гистоновыми хвостами для систем со множественными модификациями гистонов. Бледно-серым цветом на графиках отмечены средние абсолютные ошибки, рассчитанные для каждого кадра траектории.

При более детальном анализе контактов с ДНК были сделаны следующие наблюдения (Рис.12, Рис. А1 из Приложения, Рис. А2 из Приложения).

Число контактов ДНК и аргининов гистоновых хвостов, которые находятся вблизи модифицированных лизинов, но не несут меток (R8, R17, R26 в гистоне H3; R3, R23 в гистоне H4), уменьшается в большей части систем. Можно предположить, что ацетилирование лизинов способствует “отхождению” соседних аргининов от ДНК, что впоследствии приводит к сильному увеличению подвижности гистоновых хвостов [33]. Также видно, что уменьшается количество контактов и самих модифицированных лизинов (K5, K12, K20 в гистоне H2B) с ДНК во всех системах.

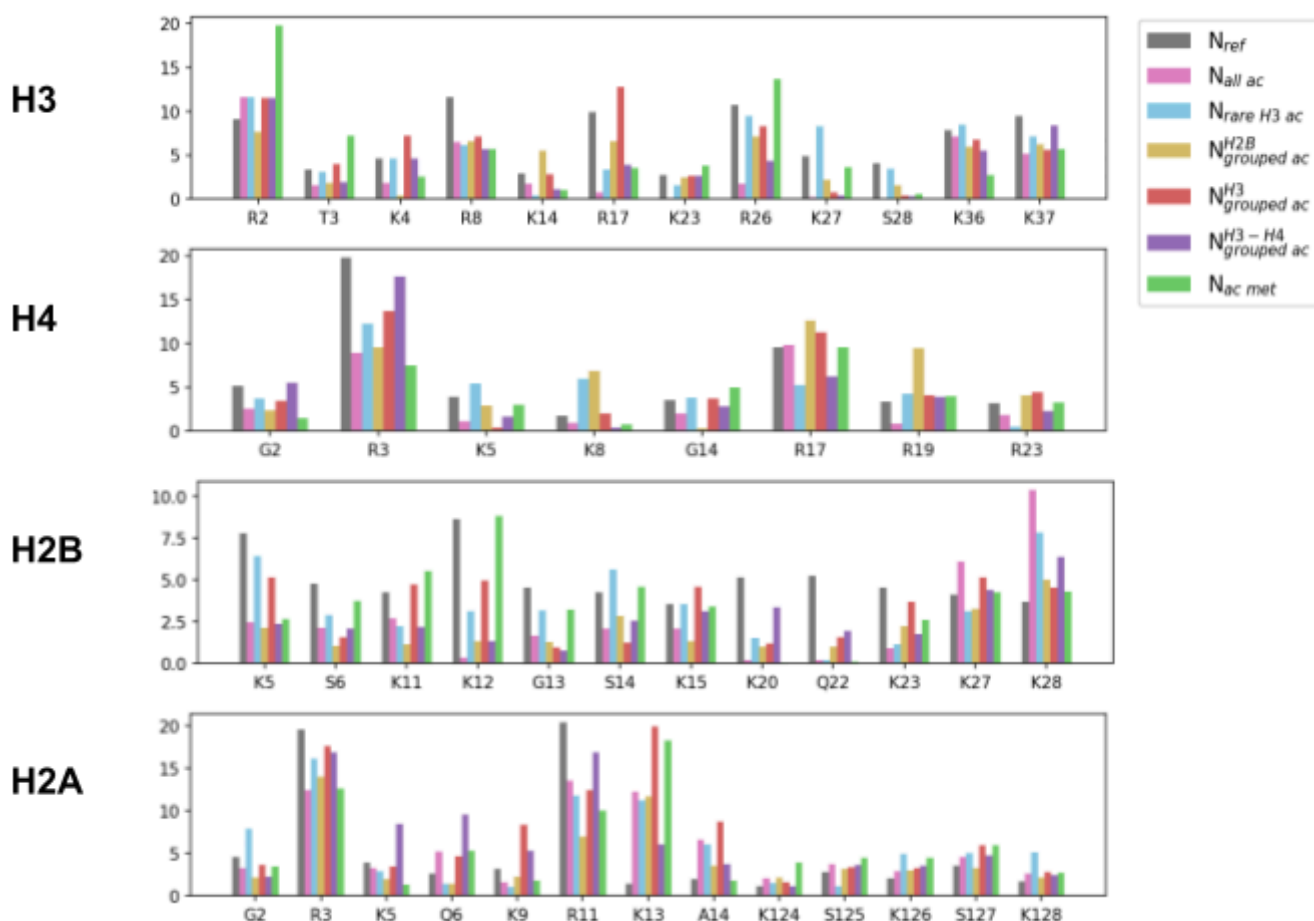


Рисунок 12. Среднее количество контактов ДНК-хвост, рассчитанное для некоторых аминокислотных остатков гистоновых хвостов. Отображены остатки, для которых наблюдается различие в среднем количестве контактов между системами.

В системе $N_{ac\ met}$ уникально увеличивается среднее количество контактов с ДНК у остатков аргинина (R2 и R26) и треонина T3 гистона H3. Также увеличение количества контактов, по сравнению с остальными системами, в системе $N_{ac\ met}$ характерно для гистона H2B (остатки S6, K11, K12, G13, S14, K15, V18). В системе $N_{all\ ac}$ происходит самое заметное уменьшение контактов ДНК с остатками хвостов H3, H4 и H2B (T3, K4, Q5, R17, K18, Q19, T22, K23, R26, K27, S28, K37 гистона H3; G4, K5, K8, R19, R23 гистона H4; K12, K20, Q22, K23 гистона H2B). Система $N_{H3-H4\ grouped\ ac}$ также показывает сильные изменения во взаимодействии с ДНК некоторых остатков (T3, K14, R26, K27, S28 хвоста H3; K5, K8, R17 хвоста H4 и K5, K11, K12, G13, S14 хвоста H2B), в то время как для остатков K37 гистона H3, G2 и R3 гистона H4, K20 гистона H2B и R11 гистона H2A, наоборот, количество контактов увеличивается/изменяется незначительно, сравнивая с

другими системами. Влияние модификаций в системе $N_{H3 \text{ grouped ac}}$ несильно отличается от системы $N_{H3-H4 \text{ grouped ac}}$, однако в ней есть отличительные особенности, такие как увеличение количества контактов аргинина R17 и лизина K4 гистона H3, R17 гистона H4, K11, K15, K27 гистона H2B и K9, K13, A14 гистона H2A с ДНК, а также сильное отличие от системы $N_{H3-H4 \text{ grouped ac}}$ в количестве взаимодействий остатков K5 и K12 с ДНК. В системе $N_{H2B \text{ grouped ac}}$ наблюдается уникальное увеличение количества взаимодействий с ДНК для хвоста H4 (по остаткам K8, R17, R19) и K14 хвоста H3. У остальных остатков сохраняется тенденция к уменьшению количества взаимодействий с ДНК (R2, T3, R8, K27, S28 хвоста H3; G2, R3, G14 хвоста H4; K5, S6, K11, K12, G13, K15, K20, Q22 хвоста H2B и G2, K5, Q6, R11 хвоста H2A). В самой слабо ацетилированной системе $N_{rare ac}$ также наблюдаются уникальные эффекты: уменьшение количества контактов S125 и, наоборот, увеличение количества контактов K126 и K128 в хвосте H2A-C, а также увеличение количества контактов G2 в хвосте H2A-N, S14 в хвосте H2B и K27, K36 в хвосте H3. В остальном тенденция к уменьшению контактов хвостов с ДНК сохраняется.

Результат данного анализа говорит, что все системы имеют уникальные эффекты на динамику гистоновых хвостов. Это может быть связано с ограниченностью исследования конформационного пространства хвостов в ходе моделирования из-за ограничений по времени моделирования.

2. Анализ среднеквадратичного отклонения

Анализ среднеквадратичного отклонения (RMSF) гистоновых хвостов (Рис.13) показал, что:

- В системах $N_{all ac}$ и $N_{H3-H4 \text{ grouped ac}}$ значительно возросла подвижность хвостов H3 и H2B (RMSF сильно увеличилось по сравнению с референсом), что согласуется с результатом предыдущего анализа).
- Также RMSF хвостов H3 и H2B увеличился в системах $N_{H2B \text{ grouped ac}}$ и $N_{H3 \text{ grouped ac}}$. Можно предположить, что ацетилирование хвостов H3 и H2B взаимно влияют друг на друга, увеличивая подвижность друг друга.
- Значение среднего RMSF хвоста H4 изменилась незначительно.
- RMSF хвоста H2A сильно снизилась во всех системах, особенно в системе с метилированием $N_{ac met}$.

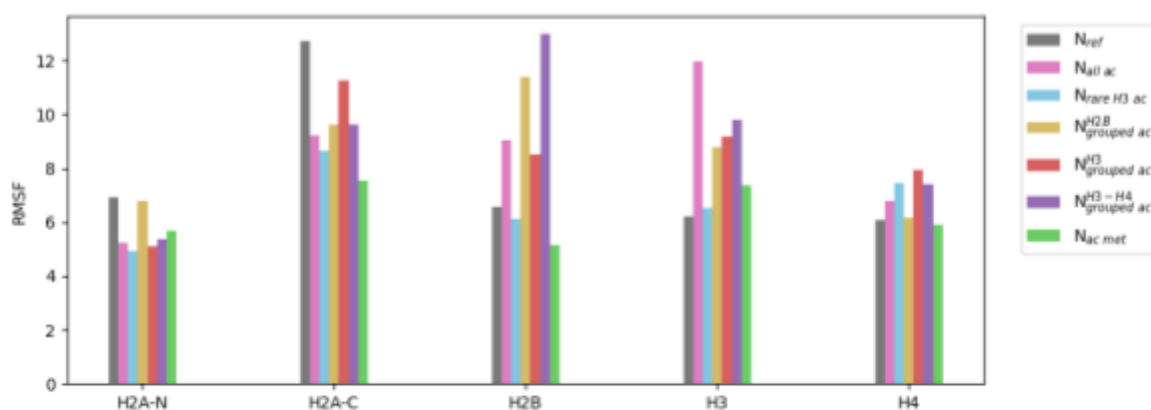


Рисунок 13. Среднее RMSF, рассчитанное для каждого хвоста каждой системы.

Далее проводился анализ пластичности в траекториях молекулярной динамики элементов глобулярной части гистонов (Рис.14-15, Рис.А3 из Приложения). Выяснили, что набор модификаций гистоновых хвостов системы **N_{ac met}** значительным образом влияет на подвижность глобулярной части (спиралей α_{1ext} и α_1 и петли L2) гистона H2A (Рис.15Б). В системе **N_{ac met}** также сильно увеличилось значение RMSF для спирали α_3 гистона H3, α_{1ext} и α_1 гистона H4 и петли L2 гистона H4 (Рис. 14А,Б). Обратный эффект для системы **N_{ac met}** наблюдается в уменьшении значения RMSF петель L1, L2, спирали α_C гистона H2B (Рис. 15А).

В остальных системах наблюдаются схожие изменения: небольшое уменьшение RMSF петли L1 гистона H3, уменьшение RMSF петли L2 гистона H4, увеличение RMSF петли L2 гистона H2B и увеличение RMSF петли L1 гистона H2A (Рис. 14-15).

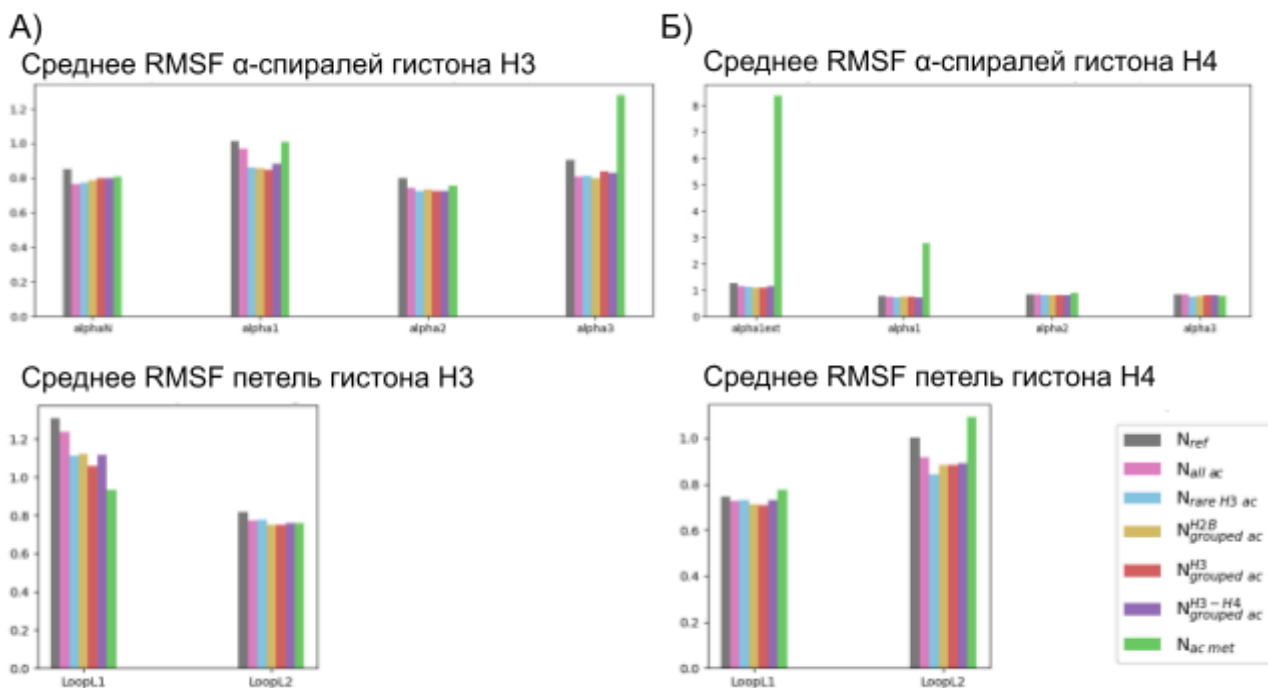


Рисунок 14. А) Среднее RMSF α -спиралей гистона H3 (сверху) и петель гистона H3 (снизу), рассчитанное для каждой системы. Б) Среднее RMSF α -спиралей гистона H4 (сверху) и петель гистона H4 (снизу), рассчитанное для каждой системы.

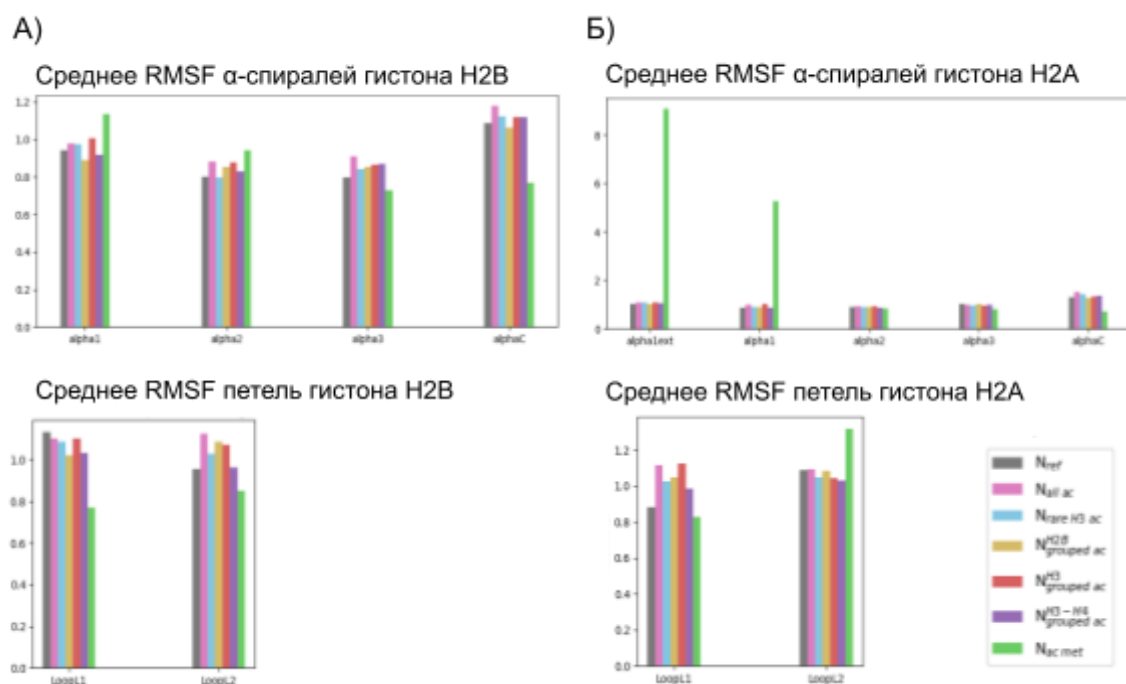


Рисунок 15. А) Среднее RMSF α -спиралей гистона H2B (сверху) и петель гистона H2B (снизу), рассчитанное для каждой системы. Б) Среднее RMSF α -спиралей гистона H2A (сверху) и петель гистона H2A (снизу), рассчитанное для каждой системы.

Анализ RMSF хвостов подтверждает результат, полученный при анализе контактов: область, исследуемая хвостами, увеличивается, что также говорит об увеличении их подвижности. Благодаря тому, что хвосты отходят от ДНК, она открывается для взаимодействия с другими белками, в том числе с факторами хроматина [20]. К тому же, мы можем предположить, что из-за различия в степени диссоциации гистонов от ДНК, разные комбинации модификаций вносят разный вклад в уровень транскрипции генов [48].

Ацетилирование гистонов также приводит к увеличению динамики глобулярной части нуклеосомы, что наблюдается для всех систем, в особенности для системы $N_{ac\ met}$. Известно, что пластичность глобулярной части гистонов связана со стабильностью нуклеосом и динамикой нуклеосомной ДНК. Например, динамика изгиба H2A и H2B гистонов связана с динамикой откручивания и скольжения нуклеосомной ДНК [48]. Таким образом, наблюдаемое изменение пластичности глобулярной части гистонов при множественном модифицировании хвостов гистонов может влиять на общую стабильность нуклеосом и быть новым механизмом регуляции активности хроматина.

ЗАКЛЮЧЕНИЕ

В данной работе проводилось исследование солокализации ПТМ: анализ аннотации хроматина человека на основе метода с использованием информации о ПТМ и анализ влияния солокализации модификаций на динамику нуклеосомы. При решении первой задачи был оценен инструмент ChromHMM, осуществляющий аннотацию хроматина, и на основе его аннотации были выбраны таргетные гены для потенциальных задач изменения их экспрессии с помощью системы CRISPR-Cas. При решении второй задачи был проведен анализ ковстречаемости 26 ПТМ, на его основе были выделены комбинации модификаций и проведен анализ их влияния на динамику нуклеосомы. В ходе анализа было выяснено, что ацетилирование систем в разной степени приводит к увеличению динамики гистоновых хвостов и снижению числа контактов с нуклеосомной ДНК, а также к изменению пластичности глобулярной части нуклеосомы, что на больших временах моделирования может влиять на общую стабильность нуклеосомы.

ВЫВОДЫ

1. С помощью функциональной разметки генома человека (ChromHMM) были определены 2 группы генов, подходящие для задачи активации/репрессии генов с помощью системы CRISPR-Cas (46 генов для задачи репрессии, 776 - для задачи активации).
2. На основе анализа ковстречаемости ПТМ были выделены 6 комбинаций модификаций, совместно встречающихся в геномных локусах.
3. Было выяснено, что множественные ПТМ гистонов H2B и H3 приводят к уменьшению количества контактов между ДНК и гистоновыми хвостами и увеличению подвижности хвостов.
4. Также было определено, что множественные ПТМ хвостов гистонов влияют на динамику глобулярной части нуклеосомы.

СПИСОК ЛИТЕРАТУРЫ

1. Armeev G. A. [и др.]. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes // *Nature Communications*. 2021. (12).
2. Bowman G. D., Poirier M. G. Post-translational modifications of histones that influence nucleosome dynamics // *Chemical Reviews*. 2015. № 6 (115). С. 2274–2295.
3. Fedulova A. S. [и др.]. Molecular dynamics simulations of nucleosomes are coming of age // *WIREs Computational Molecular Science*. 2024. № 4 (14). С. e1728.
4. Gowers R. [и др.]. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations Austin, Texas:, 2016.С. 98–105.
5. Kurumizaka H., Kujirai T., Takizawa Y. Contributions of Histone Variants in Nucleosome Structure and Function // *Journal of Molecular Biology*. 2021. № 6 (433). С. 166678.
6. Michaud-Agrawal N. [и др.]. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations // *Journal of Computational Chemistry*. 2011. № 10 (32). С. 2319–2327.
7. Peng Y. [и др.]. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails // *Nature Communications*. 2021. № 1 (12). С. 5280.
8. Peng Y. [и др.]. Histone tails as signaling antennas of chromatin // *Current Opinion in Structural Biology*. 2021. (67). С. 153–160.
9. Rabdano S. O. [и др.]. Histone H4 Tails in Nucleosomes: a Fuzzy Interaction with DNA // *Angewandte Chemie International Edition*. 2021. № 12 (60). С. 6480–6487.
10. Ran F. A. [и др.]. Genome engineering using the CRISPR-Cas9 system // *Nature Protocols*. 2013. № 11 (8). С. 2281–2308.
11. Simon M. [и др.]. Histone fold modifications control nucleosome unwrapping and disassembly // *Proceedings of the National Academy of Sciences*. 2011. № 31 (108). С. 12711–12716.
12. Davey C. A. [и др.]. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution // *Journal of Molecular Biology*. 2002. № 5 (319). С. 1097–1113.
13. Bannister A. J., Kouzarides T. Regulation of chromatin by histone modifications // *Cell Research*. 2011. № 3 (21). С. 381–395.
14. Bartke T. [и др.]. Nucleosome-interacting proteins regulated by DNA and histone methylation // *Cell*. 2010. № 3 (143). С. 470–484.
15. Bendandi A. [и др.]. The role of histone tails in nucleosome stability: An electrostatic perspective // *Computational and Structural Biotechnology Journal*. 2020. (18). С. 2799–2809.
16. Black J. C., Van Rechem C., Whetstine J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact // *Molecular Cell*. 2012. № 4 (48). С.

491–507.

17. Calo E., Wysocka J. Modification of Enhancer Chromatin: What, How, and Why? // *Molecular Cell*. 2013. № 5 (49). С. 825–837.

18. Creyghton M. P. [и др.]. Histone H3K27ac separates active from poised enhancers and predicts developmental state // *Proceedings of the National Academy of Sciences*. 2010. № 50 (107). С. 21931–21936.

19. Duan Y. [и др.]. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations // *Journal of Computational Chemistry*. 2003. № 16 (24). С. 1999–2012.

20. Eberhardter A., Becker P. B. Histone acetylation: a switch between repressive and permissive chromatin // *EMBO Reports*. 2002. № 3 (3). С. 224–229.

21. Ebert A. [и др.]. Histone modification and the control of heterochromatic gene silencing in *Drosophila* // *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*. 2006. № 4 (14). С. 377–392.

22. Egelhofer T. A. [и др.]. An assessment of histone-modification antibody quality // *Nature Structural & Molecular Biology*. 2011. № 1 (18). С. 91–93.

23. Ernst J., Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome // *Nature Biotechnology*. 2010. № 8 (28). С. 817–825.

24. Ernst J., Kellis M. Chromatin-state discovery and genome annotation with ChromHMM // *Nature Protocols*. 2017. № 12 (12). С. 2478–2492.

25. Filion G. J. [и др.]. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells // *Cell*. 2010. № 2 (143). С. 212–224.

26. Fuchs S. M. [и др.]. Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition // *Current Biology*. 2011. № 1 (21). С. 53–58.

27. Galligan J. J. [и др.]. Quantitative Analysis and Discovery of Lysine and Arginine Modifications // *ACS Publications* [Электронный ресурс]. URL: <https://pubs.acs.org/doi/abs/10.1021/acs.analchem.6b04105> (дата обращения: 10.05.2025).

28. Gao J. [и др.]. Absolute Quantification of Histone PTM Marks by MRM-Based LC-MS/MS // *Analytical Chemistry*. 2014. № 19 (86). С. 9679–9686.

29. Hu Z. [и др.]. Disruption of HPV16-E7 by CRISPR/Cas system induces apoptosis and growth inhibition in HPV16 positive human cervical cancer cells // *BioMed Research International*. 2014. (2014). С. 612823.

30. Hyllus D. [и др.]. PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation // *Genes & Development*. 2007. № 24 (21). С. 3369–3380.

31. Ito Y. [и др.]. Epigenetic profiles guide improved CRISPR/Cas9-mediated gene knockout in human T cells // *Nucleic Acids Research*. 2024. № 1 (52). С. 141–153.

32. Jadhav U. [и др.]. Acquired tissue-specific promoter bivalency is a basis for PRC2 necessity in adult cells // *Cell*. 2016. № 6 (165). С. 1389–1400.
33. Jennings C. E., Zoss C. J., Morrison E. A. Arginine anchor points govern H3 tail dynamics // *Frontiers in Molecular Biosciences*. 2023. (10).
34. Jenuwein T., Allis C. D. Translating the Histone Code // *Science*. 2001. № 5532 (293). С. 1074–1080.
35. Joseph F. M., Young N. L. Histone variant-specific post-translational modifications // *Seminars in Cell & Developmental Biology*. 2023. (135). С. 73–84.
36. Karch K. R., Sidoli S., Garcia B. A. Chapter One - Identification and Quantification of Histone PTMs Using High-Resolution Mass Spectrometry *Enzymes of Epigenetics, Part B* / под ред. R. Marmorstein, Academic Press, 2016. С. 3–29.
37. Kasowski M. [и др.]. Extensive Variation in Chromatin States Across Humans // *Science*. 2013.
38. Kim H., Pak Y. Free Energy-Based Refinement of DNA Force Field via Modification of Multiple Nonbonding Energy Terms // *Journal of Chemical Information and Modeling*. 2024.
39. Kim T. H. [и др.]. Correlating histone acetylation with nucleosome core particle dynamics and function // *Proceedings of the National Academy of Sciences of the United States of America*. 2023. № 15 (120). С. e2301063120.
40. Kosicki M., Tomberg K., Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements // *Nature Biotechnology*. 2018. № 8 (36). С. 765–771.
41. La Russa M. F., Qi L. S. The New State of the Art: Cas9 for Gene Activation and Repression // *Molecular and Cellular Biology*. 2015. № 22 (35). С. 3800–3809.
42. Lineback J. E., Jansma A. L. PyMOL as an Instructional Tool To Represent and Manipulate the Myoglobin/Hemoglobin Protein System // *Journal of Chemical Education*. 2019. № 11 (96). С. 2540–2544.
43. Lv J. [и др.]. The length of guide RNA and target DNA heteroduplex effects on CRISPR/Cas9 mediated genome editing efficiency in porcine cells // *Journal of Veterinary Science*. 2019. № 3 (20). С. e23.
44. Martin B. J. E. [и др.]. Transcription shapes genome-wide histone acetylation patterns // *Nature Communications*. 2021. № 1 (12). С. 210.
45. Martin D. G. E. [и др.]. The Yng1p Plant Homeodomain Finger Is a Methyl-Histone Binding Module That Recognizes Lysine 4-Methylated Histone H3 // *Molecular and Cellular Biology*. 2006. № 21 (26). С. 7871–7879.
46. Murray K. THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES // *Biochemistry*. 1964. (3). С. 10–15.
47. O'Geen H., Echipare L., Farnham P. J. Using ChIP-Seq Technology to Generate High-Resolution Profiles of Histone Modifications под ред. T. O. Tollefsbol, Totowa, NJ:

Humana Press, 2011. С. 265–286.

48. Oishi T. [и др.]. Contributions of histone tail clipping and acetylation in nucleosome transcription by RNA polymerase II // *Nucleic Acids Research*. 2023. № 19 (51). С. 10364–10374.

49. Papait R. [и др.]. Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy // *Proceedings of the National Academy of Sciences of the United States of America*. 2013. № 50 (110). С. 20164–20169.

50. Perez-Pinera P. [и др.]. RNA-guided gene activation by CRISPR-Cas9–based transcription factors // *Nature Methods*. 2013. № 10 (10). С. 973–976.

51. Portillo-Ledesma S., Li Z., Schlick T. Genome modeling: From chromatin fibers to genes // *Current Opinion in Structural Biology*. 2023. (78). С. 102506.

52. Rada-Iglesias A. [и др.]. A unique chromatin signature uncovers early developmental enhancers in humans // *Nature*. 2011. № 7333 (470). С. 279–283.

53. Radman-Livaja M., Rando O. J. Nucleosome positioning: how is it established, and why does it matter? // *Developmental biology*. 2010. № 2 (339). С. 258–266.

54. Rose C. M. [и др.]. Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes // *Cell Systems*. 2016. № 4 (3). С. 395-403.e4.

55. Santos-Rosa H. [и др.]. Active genes are tri-methylated at K4 of histone H3 // *Nature*. 2002. № 6905 (419). С. 407–411.

56. Schneider V. A. [и др.]. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly // *Genome Research*. 2017. № 5 (27). С. 849–864.

57. Uddin F., Rudin C. M., Sen T. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future // *Frontiers in Oncology*. 2020. (10). С. 1387.

58. Venkatesh S. [и др.]. Selective suppression of antisense transcription by Set2-mediated H3K36 methylation // *Nature Communications*. 2016. № 1 (7). С. 13610.

59. Wang D. [и др.]. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization // *Nucleic Acids Research*. 2020. № W1 (48). С. W140–W146.

60. Wang T., Holt M. V., Young N. L. Early butyrate induced acetylation of histone H4 is proteoform specific and linked to methylation state // *Epigenetics*. 2018.

61. Wu X. [и др.]. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells // *Nature Biotechnology*. 2014. № 7 (32). С. 670–676.

62. Yang X.-J., Seto E. Lysine acetylation: codified crosstalk with other posttranslational modifications // *Molecular Cell*. 2008. № 4 (31). С. 449–461.

63. Yu Y., Santat L. A., Choi S. 6 - Bioinformatics Packages for Sequence Analysis Applied Mycology and Biotechnology / под ред. D. K. Arora, R. M. Berka, G. B. Singh,

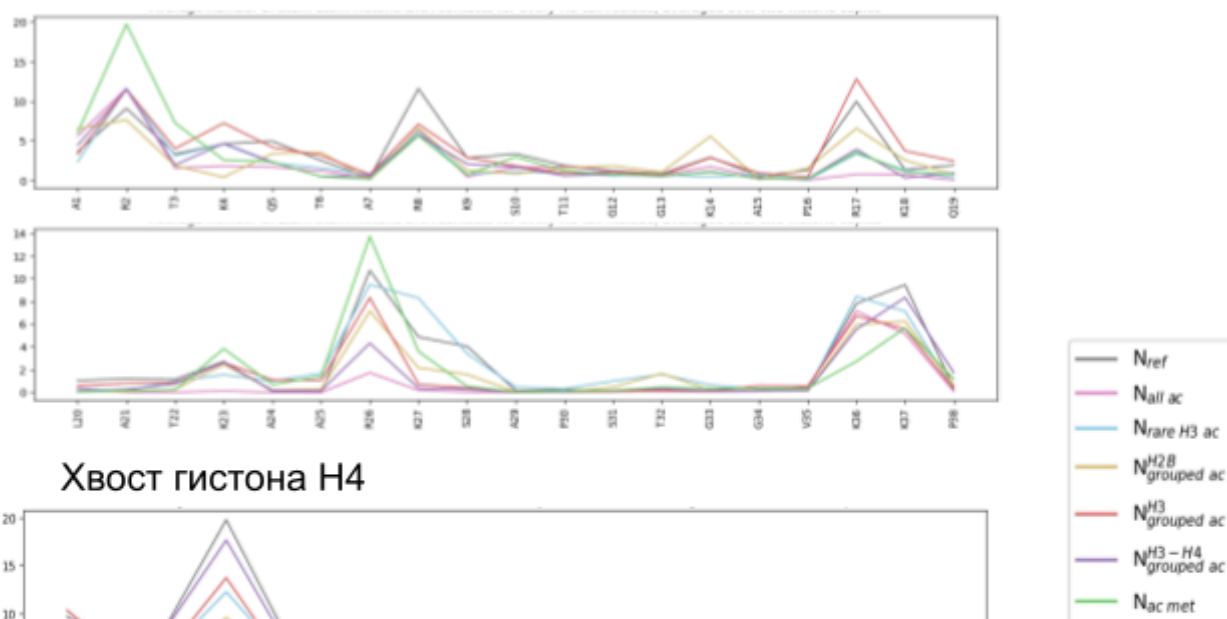
Elsevier, 2006.С. 143–160.

64. Zhang Y. [и др.]. Overview of Histone Modification под ред. D. Fang, J. Han, Singapore: Springer, 2021.С. 1–16.

65. Zippo A. [и др.]. Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation // Cell. 2009. № 6 (138). С. 1122–1136.

ПРИЛОЖЕНИЕ

Хвост гистона H3



Хвост гистона H4

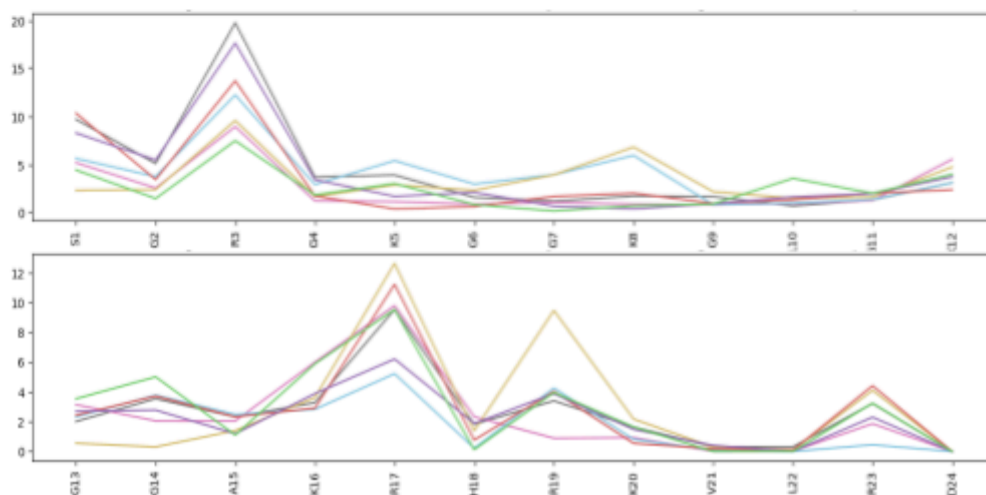


Рисунок A1. Среднее количество контактов ДНК-хвост, рассчитанное для каждого аминокислотного остатка гистонов H3 и H4.

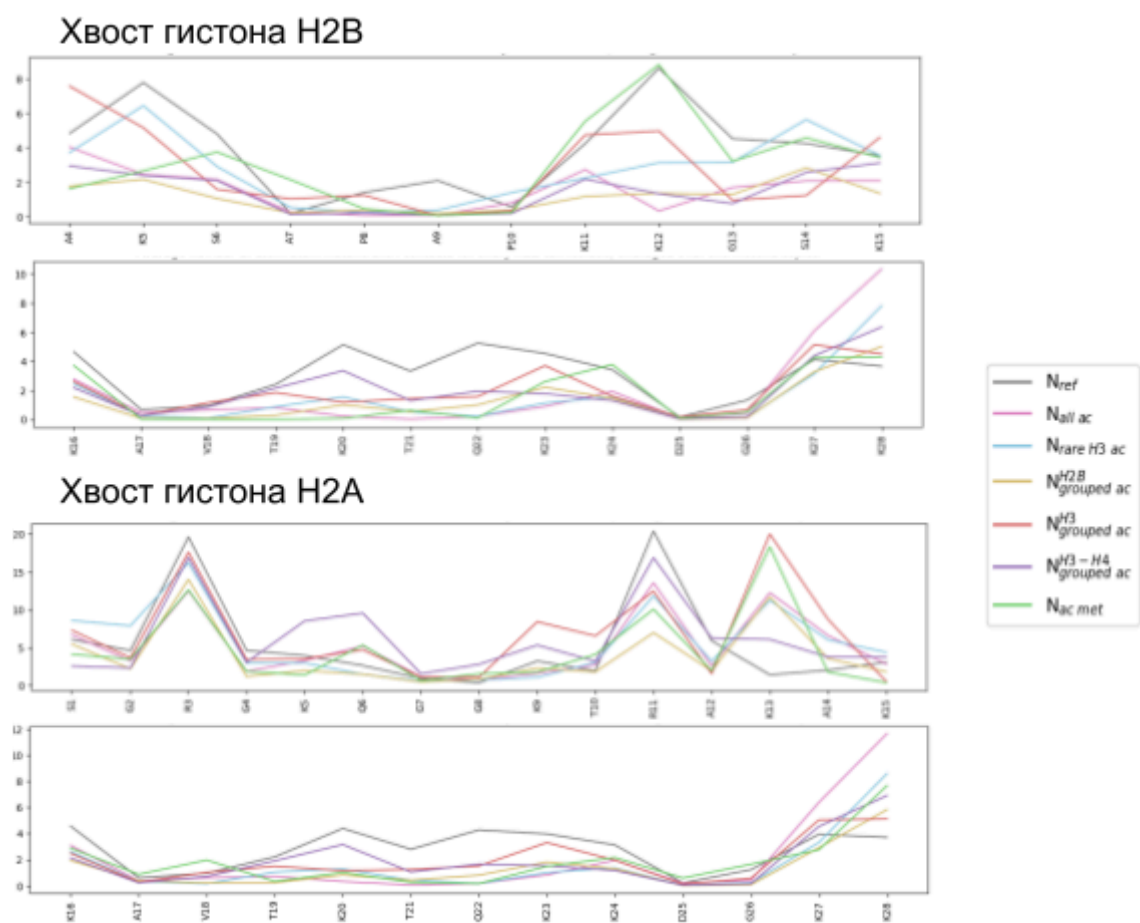


Рисунок A2. Среднее количество контактов ДНК-хвост, рассчитанное для каждого аминокислотного остатка гистонов H2B и H2A.

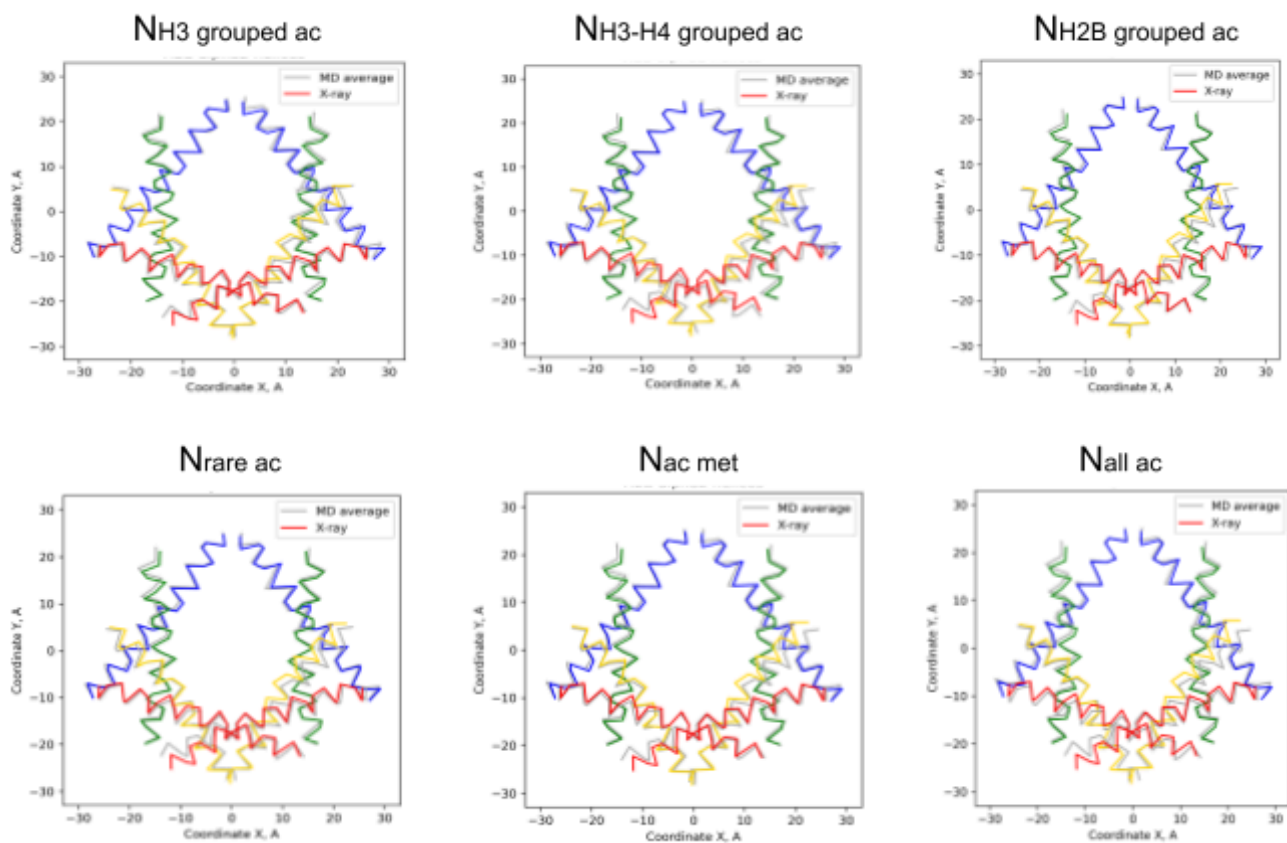


Рисунок А3. Пластичность гистонового ядра, выполненная с помощью построения двумерных проекций положений Са-атомов α -спиралей гистонов на плоскость нуклеосомы. Серым цветом помечена проекция референсной нуклеосомы.