

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Грибкова Анна Кирилловна

**Классификация белков хроматина человека на основе
функциональных, физико-химических и структурных свойств методами
биоинформатики и вычислительной биологии**

Специальность 1.5.8. Математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2026

Диссертация подготовлена на кафедре биоинженерии биологического факультета МГУ имени М.В.Ломоносова

Научный руководитель — *Шайтан Алексей Константинович*, доктор физико-математических наук, член-корреспондент РАН

Официальные оппоненты — *Киреев Игорь Игоревич*, доктор биологических наук, Московский государственный университет имени М.В. Ломоносова, Научно-исследовательский институт физико-химической биологии имени А.Н. Белозерского, заведующий отделом электронной микроскопии

Крупянский Юрий Федорович, доктор физико-математических наук, Федеральный исследовательский центр химической физики им. Н.Н. Семенова Российской академии наук, руководитель отдела строения вещества

Храмеева Екатерина Евгеньевна, доктор биологических наук, Сколковский институт науки и технологий, доцент Центра био- и медицинских технологий

Защита диссертации состоится «22» октября 2026 года в 14 часов 00 минут на заседании диссертационного совета МГУ.015.5 Московского государственного университета имени М.В.Ломоносова по адресу: 119234, Россия, Москва, Ленинские горы, д.1, стр.24, МГУ, биологический факультет, кафедра биофизики, аудитория «Новая».

E-mail: fursova@biophys.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на портале: <https://dissovet.msu.ru/dissertation/3983>.

Автореферат разослан «__» _____ 2026 г.

Ученый секретарь диссертационного совета,
кандидат физико-математических наук

П.В. Фурсова

I. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность и степень разработанности темы исследования

В ядре эукариотических клеток геномная ДНК организована в виде хроматина — комплекса, состоящего из ДНК, РНК и белков (Van Holde 1989; Bernstein and Allis 2005). Суммарная длина молекул ДНК в ядре одной клетки человека достигает примерно двух метров при диаметре ядра около десяти микрометров, что требует высокоорганизованной упаковки генетического материала.

Белки хроматина не только участвуют в упаковке ДНК в ядре клетки, но и играют ключевую роль в регуляции её функциональной активности (Millán-Zambrano et al. 2022; Berger 2007). Несколько тысяч белков, ассоциированных с хроматином, обеспечивают реализацию основных молекулярно-биологических процессов, включая транскрипцию, репликацию и репарацию ДНК (Imhof and Bonaldi 2005).

Функциональное состояние хроматина связано с физико-химическими свойствами ассоциированных с ним белков, их количеством, наличием в них структурных доменов и неупорядоченных регионов, которые формируют динамическую сеть специфических и неспецифических многокомпонентных взаимодействий между молекулами.

Первые работы по систематизации белков ядра клетки были выполнены К. Е. ван Хольдом в 1980-х годах (Van Holde 1989). Он разделил их на гистоновые, негистоновые хромосомные и нехромосомные, а также продемонстрировал существенную зависимость состава выделяемой фракции белков хроматина от метода пробоподготовки.

В современных исследованиях принадлежность белка к хроматину определяют с помощью различных методов. Среди них — исследования методами масс-спектрометрии очищенных фракций белков хроматина (van Mierlo and Vermeulen 2021), анализ субклеточной локализации белка с использованием флуоресцентного мечения и конфокальной микроскопии (Thul

et al. 2017; Cho et al. 2022), а также анализ функций белка по данным из специализированных баз данных и литературы (Gene Ontology Consortium et al. 2023; Marakulina et al. 2023). Однако получение согласованной информации о репертуаре белков хроматина затруднено по ряду причин: ограниченные чувствительность и разрешение методов анализа, функциональная и пространственная компартиментализация хроматина, а также динамическое перемещение белков между ядром и цитоплазмой в процессе жизнедеятельности клетки. Таким образом, накопленные данные о функциях, качественном и количественном разнообразии белков хроматина остаются разрозненными.

Несмотря на то что современные представления о пространственной организации генома и структуре отдельных ДНК-белковых комплексов в последнее время значительно расширились благодаря развитию методов эпигеномики, 3D-геномики, структурной биологии и биофизики (Armeev et al. 2022; Pletenev et al. 2024; Dekker et al. 2023), имеющихся данных о белках недостаточно для формализации описания функционирования и пространственно-временной организации хроматина.

Таким образом, предметом работы является разработка системного подхода, интегрирующего гетерогенную информацию о функциональных, физико-химических и структурных свойствах белков хроматина в единую классификацию. Это необходимо для формирования целостного представления о белковом составе хроматина и его роли в организации и регуляции генома человека.

Целью диссертационной работы является разработка эмпирической функционально-ориентированной классификации белков хроматина человека на основе интеграции данных об их функциях, субклеточной локализации, физико-химических свойствах и доменной организации.

Для достижения цели были поставлены следующие **задачи**:

- 1) Провести критический мета-анализ разнородных источников информации о функциях, локализации и свойствах белков хроматина и ядра клетки

человека.

- 2) Разработать на основе интеграции гетерогенных данных о составе и свойствах белков хроматина человека эмпирическую функционально-ориентированную классификацию SimChrom, а также эталонные наборы белков ядра.
- 3) Применить разработанную классификацию белков хроматина SimChrom для оценки количественной представленности белков различных функциональных категорий (по данным масс-спектрометрических исследований) и выявления функционально значимых физико-химических свойств этих белков.
- 4) Охарактеризовать организацию белков хроматина человека на уровне структурных и функциональных доменов. Определить функционально значимые сочетания доменов, участвующих в мультивалентных взаимодействиях.

Научная новизна

Разработанная эмпирическая функционально-ориентированная классификация белков хроматина человека SimChrom разделяет 3045 белков на 39 категорий. Каждый белок в классификации характеризуется независимыми параметрами: (i) субклеточной локализацией, (ii) функциональной аннотацией, (iii) особенностями строения (физико-химическими свойствами и доменным составом). Предложенная классификация обеспечивает системное сопоставление независимых параметров в рамках единой системы.

Применение данной классификации позволяет устанавливать функционально значимые особенности строения белков хроматина, в частности определять параметры их неупорядоченных регионов и распределение в них заряженных аминокислотных остатков, выявлять характерные сочетания функциональных доменов в белках, обуславливающие мультивалентные взаимодействия в хроматине, а также идентифицировать новые структурные домены, которые отсутствуют в экспериментальных базах данных.

Таким образом, разработанная классификация не только систематизирует

информацию о белках хроматина человека, но и позволяет проводить интегративный анализ накопленных гетерогенных данных, а именно на основе данных о субклеточной локализации, физико-химических свойствах и доменном составе белка уточнять и прогнозировать его функциональную роль в хроматине.

Теоретическая и практическая значимость работы

Теоретическая значимость работы заключается в систематизации и интеграции разнородных данных о качественном и количественном составе белков хроматина человека, накопленных в научной литературе. Полученные результаты расширяют представление о белковом составе хроматина человека, а также углубляют понимание молекулярных механизмов функционирования и регуляции генома.

Практическая значимость работы определяется следующим:

- 1) Интеграция гетерогенных данных о белках хроматина в единую функционально-ориентированную классификацию, доступную в сети Интернет (<https://simchrom.intbio.org/>), позволяет проводить более детальную функциональную аннотацию белков хроматина человека на основе сопоставления различных аспектов их свойств, а также уточнять или прогнозировать функции отдельных белков.
- 2) Разработанные эталонные наборы белков ядра и цитоплазмы клетки с различными критериями достоверности и согласованности данных применимы в качестве обучающих наборов в задачах предсказания субклеточной локализации белков методами машинного обучения.

Методология и методы исследования

В ходе исследования разработана оригинальная методология комплексного анализа структурно-функциональных свойств белков ядра и хроматина клетки человека. Методология объединяет три взаимосвязанных этапа: (i) системный мета-анализ существующих способов определения принадлежности белков к хроматину и ядру, (ii) разработку эмпирической функционально-ориентированной классификации белков хроматина и (iii)

применение классификации для выявления структурно-функциональных особенностей организации белков хроматина человека (**Рисунок 1**).

Исследование выполнено с использованием современных методов биоинформатики, вычислительной биологии и статистического анализа. В работе особое внимание уделено приведению гетерогенных данных, извлекаемых из специализированных баз данных и литературы, к сопоставимому виду.

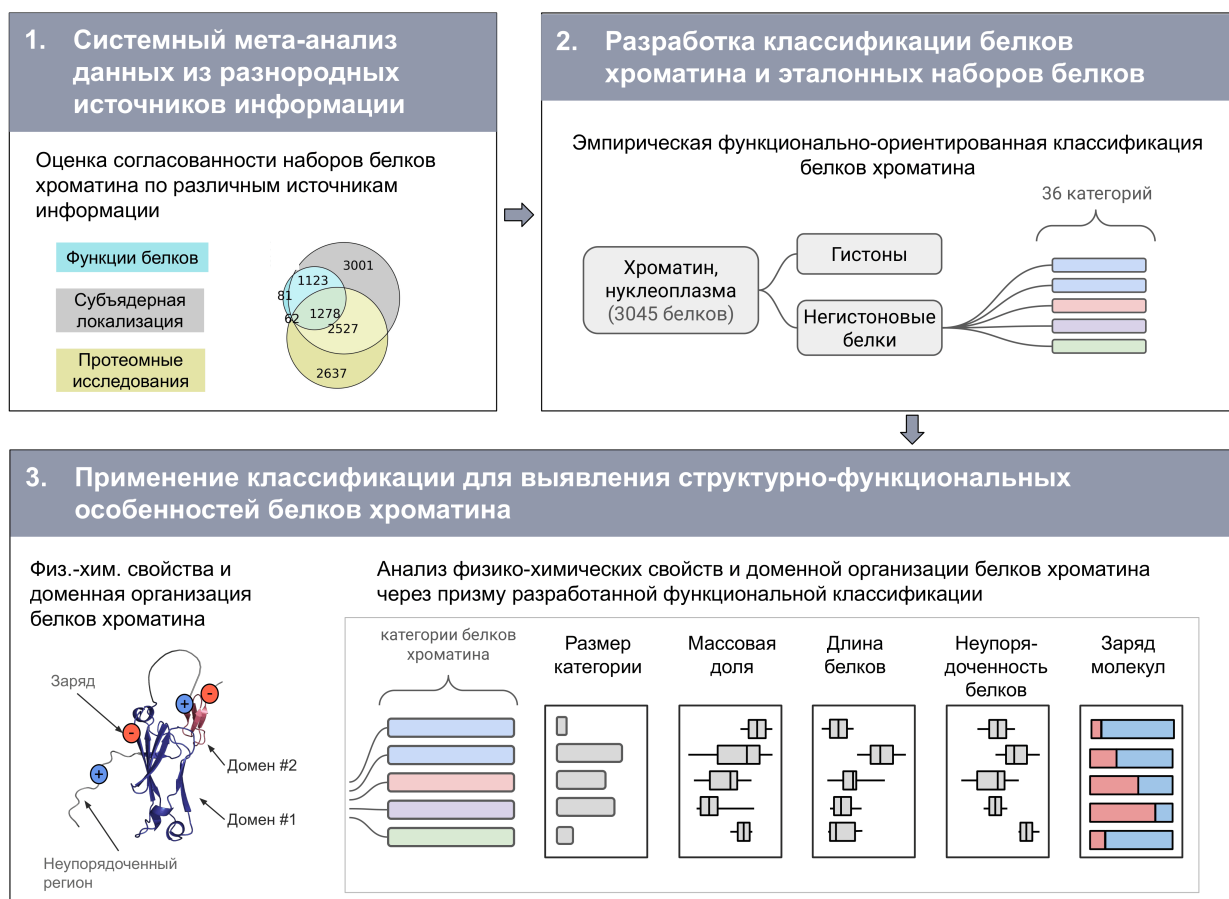


Рисунок 1 — Краткая схема методологии исследования, состоящая из трех этапов.

Положения, выносимые на защиту:

- 1) Разработка эмпирической функционально-ориентированной классификации белков хроматина человека возможна при условии итеративной интеграции гетерогенных данных об их составе и свойствах (локализация, функции, физико-химические свойства, доменная организация). Предложенная классификация позволяет проводить

системный анализ структурно-функциональных особенностей организации белков хроматина, выявлять новые закономерности, уточнять известные функции и прогнозировать свойства неохарактеризованных белков.

- 2) Выявленные характерные признаки строения белков хроматина (высокая доля неупорядоченных регионов, специфическое распределение заряженных аминокислотных остатков) лежат в основе их структурных и регуляторных функций, а также способствуют участию этих белков в многокомпонентных взаимодействиях.
- 3) Мультидоменная организация белков хроматина с преобладанием комбинаций доменов, узнающих модифицированные гистоны и ДНК, способствует регуляции эпигенетических процессов.
- 4) На основе разработанной классификации белков хроматина SimChrom с применением методов искусственного интеллекта возможно выявление ранее неохарактеризованных структурных доменов в белках хроматина человека.

Степень достоверности и апробация результатов

Основные результаты по теме диссертации опубликованы в 7 статьях в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности 1.5.8. Математическая биология, биоинформатика (физико-математические науки). Общий объем публикаций составляет 9,52 п.л. Изложенные результаты работы были представлены в виде устных и стендовых докладов на 11 всероссийских и международных конференциях.

Личный вклад автора

В работе (Gribkova et al., 2026) соискателем разработана эмпирическая функционально-ориентированная классификация белков хроматина человека SimChrom и проведен анализ структурно-функциональных свойств белков хроматина человека. Соискателем разработана база данных белков хроматина SimChrom, техническая реализация интерактивной веб-версии выполнена Г. А.

Армеевым. Автором проведен обзор влияния вариантов гистонов на динамику нуклеосом (Armeev et al., 2019), анализ онкологических мутаций в гистонах (Espiritu et al., 2021), классификация негистоновых партнеров в структурах нуклеосом (Armeev et al., 2022), внедрение новой номенклатуры генов гистонов человека в базу данных HistoneDB 2.0 (Seal et al., 2022), классификация белков хроматина из предсказанных комплексов с пионерными транскрипционными факторами (Хасанова и др., 2025), а также обзор механизмов влияния COVID-19 на доступность хроматина в иммунных клетках (Бигильдеев и др., 2024).

Структура и объем работы

Диссертация состоит из введения, пяти глав, заключения, списка сокращений, списка литературы (214 источников). Работа изложена на 190 страницах, включает 53 рисунка, 4 таблицы и 2 приложения.

II. ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы, сформулированы цель, задачи, научная новизна, положения, выносимые на защиту, теоретическая и практическая значимость работы, личный вклад автора.

Глава 1. Современные представления о структурно-функциональной организации белков хроматина и подходах к их систематизации

В главе описано устройство хроматина, уровни его иерархической организации (**Рисунок 2**), а также структурно-функциональные особенности организации гистоновых и негистоновых белков.

Описаны подходы к аннотации и предсказанию внутренне неупорядоченных регионов и доменов белков. Рассмотрены методы выявления белков, ассоциированных с хроматином, основанные на данных о субклеточной локализации, функциональной аннотации белков и результатах масс-спектрометрического (МС) анализа очищенных фракций хроматина.

В результате выявлены отсутствие единого терминологического и методологического подхода к определению репертуара белков хроматина,

существенный разброс в оценках их количества, а также отсутствие единой системы для интеграции различных свойств белков хроматина.

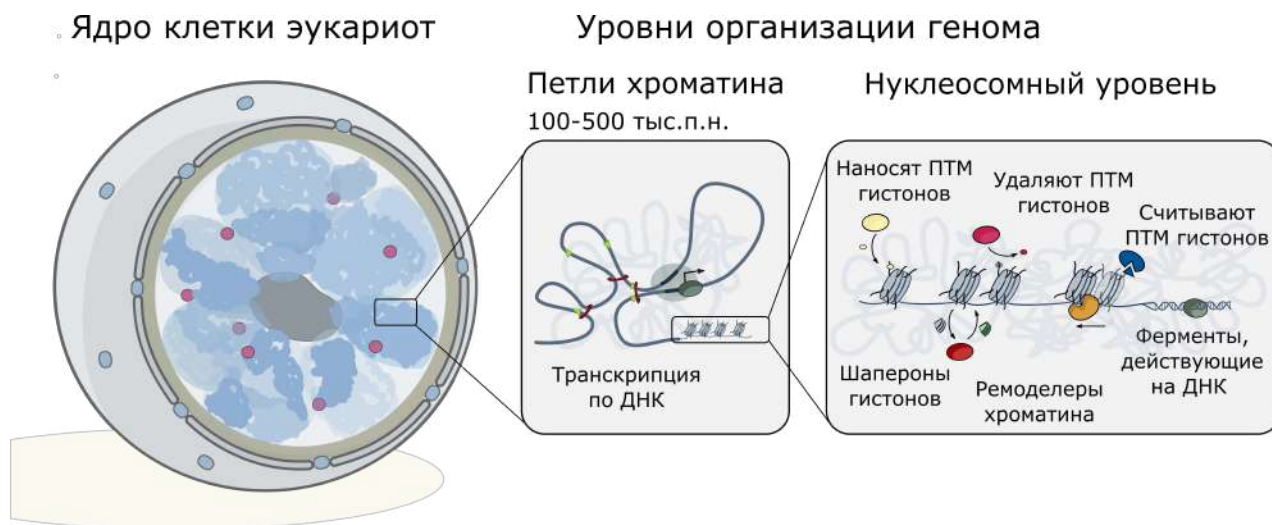


Рисунок 2 — Строение ядра клетки с деталями организации хроматина на уровне доменов хроматина и на нуклеосомном уровне, показаны функции негистоновых белков хроматина.

Глава 2. Материалы и методы исследования

В главе детально изложены материалы и методы исследования, использованные для комплексного анализа белков хроматина человека.

Для системного мета-анализа отобраны разнородные источники информации о белках ядра и хроматина человека. Данные о субклеточной локализации белков получены из баз данных HPA, OpenCell (флуоресцентное мечение белков и конфокальной микроскопии) и UniProt (литературные данные); функциональные аннотации — из Gene Ontology, EpiFactors, HistoneDB, dbTF, HISTome2; протеомные данные — из МС-исследований обогащенных фракций хроматина (Torrente et al. 2011; Kustatscher et al. 2014; Ginno et al. 2018; Shi et al. 2021; Ugur et al. 2023; Alabert et al. 2014; Alvarez et al. 2023; Itzhak et al. 2016). Согласованность наборов данных оценивали с использованием коэффициента сходства Жаккара, метрики F1 и гипергеометрического теста.

Категории классификации SimChrom сформировали на основе анализа специализированных баз данных и литературы. Далее в категории добавляли

аннотации белков из базы знаний Gene Ontology, специализированных баз данных и литературы на основании вручную установленных соответствий между категориями классификации и источниками данных. Дополнительно была разработана классификация SimChrom-SL, в которой каждый белок отнесен только к одной категории.

Относительная оценка представленности белков хроматина в клетках выполнена на основе данных из базы данных PaxDb (v4), выраженных в относительной доле белка в протеоме (parts per million, ppm) и интенсивностей сигналов, полученных в МС-исследованиях.

Для сравнения физико-химических свойств, таких как длина, внутренне неупорядоченные регионы (ВНР), заряд и аминокислотный состав между группами белков использовали U-критерий Манна — Уитни с поправкой Бонферрони. Структурная характеристика белков выполнена на основе данных из базы данных PDB и структур, предсказанных AlphaFold2, с последующим расчетом площади поверхности, доступной растворителю. Аннотация функциональных доменов получена из базы данных Pfam, структурных доменов — из баз данных CATH и TED. Соотнесение TED-доменов и Pfam-доменов со структурами из PDB и CATH выполнено инструментом FoldSeek.

Глава 3. Критический анализ разнородных источников информации о репертуаре белков хроматина и ядра клетки человека

В результате системного мета-анализа данных о белках хроматина и ядра клетки из разнородных источников информации (базы данных о локализации и функциях белков, масс-спектрометрические исследования белков хроматина и ядра клетки) определены преимущества и ограничения различных подходов к формированию перечней белков ядра и хроматина.

Выявлено, что широко используемая база знаний Gene Ontology содержит сложные иерархические структуры классификации, затрудняющие поиск и интерпретацию результатов, тогда как более специализированные ресурсы (база

данных EpiFactors и другие) ограничены по числу категорий и количеству белков.

Информационные ресурсы о клеточной локализации белков (UniProt, Human Protein Atlas, OpenCell) позволяют оценить верхнюю границу числа белков хроматина, однако они различаются в структуре онтологий субъядерных компартментов и подходах к аннотации белков с множественной локализацией.

Показано, что масс-спектрометрический анализ хроматина позволяет выявлять ранее не охарактеризованные другими методами белки, однако результаты зависят от условий эксперимента и ограничены недостаточной детекцией белков с низкой экспрессией или транзитным характером взаимодействия с хроматином.

Выявлена неполная согласованность перечней белков хроматина, полученных из разных типов источников информации (**Рисунок 3**).

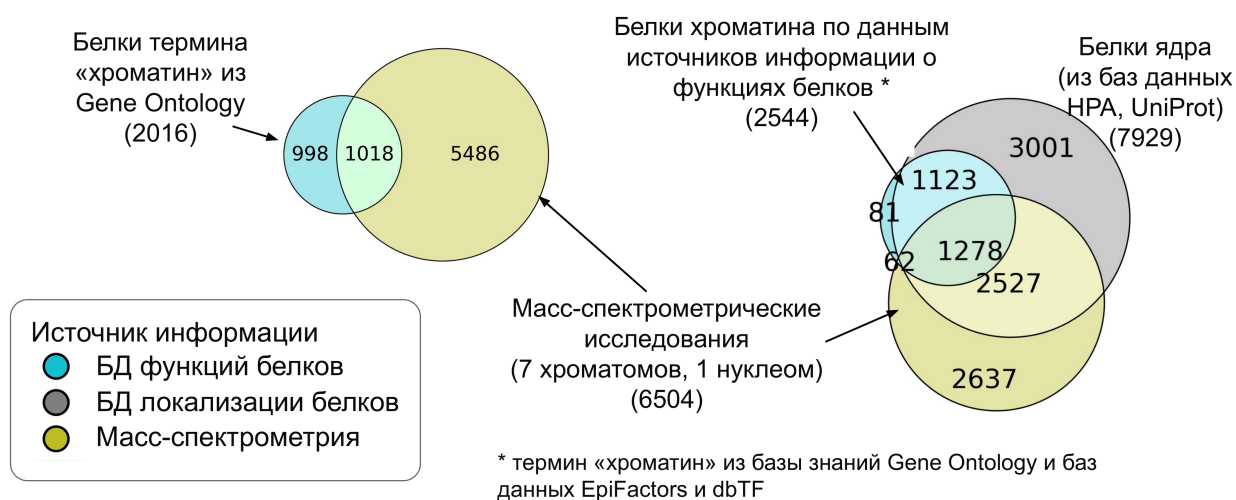


Рисунок 3 — Сравнение наборов белков хроматина и ядра из различных источников информации. Слева: пересечение белков категории «хроматин» (Gene Ontology) с белками, идентифицированными в масс-спектрометрических исследованиях хроматина и ядра клетки. Справа: пересечение трех групп белков: (i) белки категории «хроматин» (Gene Ontology), белки, регулирующие эпигенетические процессы (EpiFactors), транскрипционные факторы (dbTF); (ii) белки с локализацией в ядре (UniProt и HPA); (iii) белки, идентифицированные в масс-спектрометрических исследованиях хроматина и ядра клетки.

Так, для 25 % белков категории «хроматин» базы знаний Gene Ontology и для 42 % белков, идентифицированных в масс-спектрометрических

исследованиях хроматина, отсутствуют аннотации о локализации в ядре клетки согласно базам данных UniProt и HPA.

Полученные результаты демонстрируют значительную неоднородность и ограниченную согласованность данных о репертуаре белков хроматина и ядра между различными типами источников информации. Это подчеркивает необходимость интеграции разнородных данных в рамках единой классификационной системы.

Глава 4. Эмпирическая функционально-ориентированная классификация белков хроматина SimChrom и эталонные наборы белков ядра

С учетом преимуществ и недостатков существующих источников информации о составе белков хроматина, мы разработали эмпирическую функционально-ориентированную классификацию белков хроматина человека SimChrom.

Классификация объединяет категории белков по четырем аспектам: молекулярная функция (например, шапероны гистонов, ремоделеры хроматина, транскрипционные факторы), биологический процесс (например, репликация, репарация, рекомбинация ДНК), геномная локализация (например, эу- и гетерохроматин), физико-химические свойства (белки семейства HMG) (**Рисунок 4**). Классификация включает 39 категорий (гистоновые белки и 36 категорий негистоновых белков хроматина/нуклеоплазмы), к которым отнесено 3045 белков человека (**Рисунок 4**). Некоторые категории подразделяются на подкатегории с максимальной глубиной иерархии не более пяти уровней.

Классификация включает как категории, представленные в существующих ресурсах (Gene Ontology, EpiFactors), так и отсутствующие в них, например, пионерные транскрипционные факторы (ТФ); белки, отщепляющие хвосты гистонов; SMC-комплексы (**Рисунок 4**).

Разработаны эталонные наборы белков ядра и цитоплазмы клетки, которые различаются по степени достоверности и согласованности данных между

источниками информации (HPA, UniProt, OpenCell), а также учету множественной локализации белков.

В результате верификации классификации SimChrom показано, что она охватывает наиболее полный репертуар белков хроматина и объединяет сведения как об основных структурных компонентах хроматина, так и о динамически рекрутируемых белках, участвующих в регуляции геномных процессов. Это обосновывает её использование в биоинформатических исследованиях.

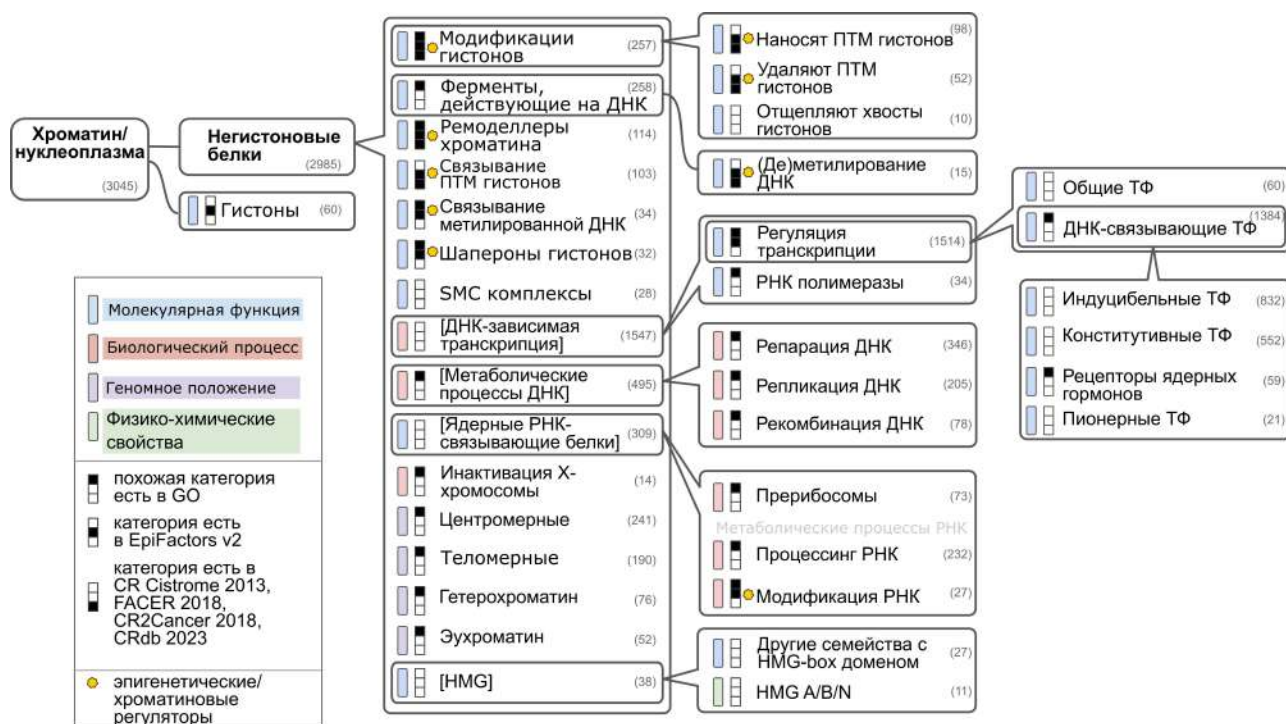


Рисунок 4 — Разработанная эмпирическая функционально-ориентированная классификация белков хроматина SimChrom. Для каждой категории справа приведено количество белков, слева обозначена принадлежность к аспекту (молекулярная функция, биологический процесс, геномное положение); пиктограммы указывают на наличие или отсутствие аналогичных категорий в онтологиях других баз данных (Gene Ontology, EpiFactors и другие).

Глава 5. Структурно-функциональные свойства белков хроматина человека

Для характеристики белков хроматина в рамках разработанной классификации SimChrom каждому белку сопоставлены данные о его субклеточной локализации, представленности в клетке, физико-химических

свойствах (длина, доля внутренне неупорядоченных регионов, заряд, аминокислотный состав) и доменной организации (**Рисунок 5**). Применение классификации позволило выявить уникальные физико-химические свойства ряда функциональных категорий белков хроматина, что демонстрирует ее эвристический потенциал.

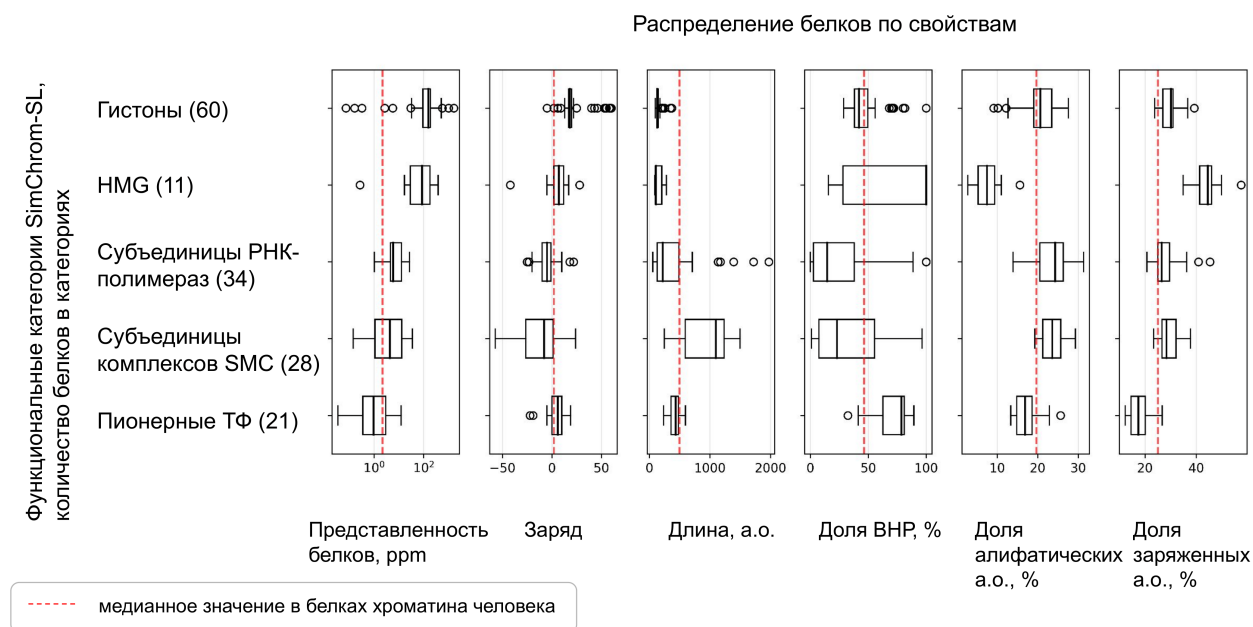


Рисунок 5 — Свойства белков хроматина из категорий классификации SimChrom (на примере некоторых свойств и функциональных групп). Длина «усов» соответствует 1,5 межквартильному размаху.

Например, пионерные ТФ, в отличие от других белков хроматина, характеризуются меньшей длиной (медиана 321 аминокислотных остатков, а.о., против 496 а.о.) и повышенной долей внутренне неупорядоченных регионов (медиана 73 % против 47 %), что объясняет их способность связываться с плотно упакованным хроматином и низкую частоту идентификации в масс-спектрометрических экспериментах (не более 25 % белков категории). Архитектурные белки хроматина из семейства HMG обладают небольшой длиной (~100 а.о.) и являются наиболее неупорядоченными (медиана 99 %), что обуславливает их гибкость для взаимодействия с ДНК и другими белками.

5.1. Количественная представленность белков хроматина в клетках

Функционирование хроматина зависит не только от качественного

разнообразие белков, но и от количества отдельных молекул в ядре клетки, то есть их представленности.

В результате анализа масс-спектрометрических данных семи хроматомов и одного нуклеома (Torrente et al. 2011; Kustatscher et al. 2014; Ginno et al. 2018; Shi et al. 2021; Ugur et al. 2023; Alabert et al. 2014; Alvarez et al. 2023; Itzhak et al. 2016) была показана высокая вариабельность оценки массовой доли гистоновых белков. Для преодоления методических разногласий в данной работе в качестве меры относительного количества белковых молекул использованы значения их представленности на уровне целого организма из базы данных PaxDb.

Выявлено обогащение хроматина конститутивными низкопредставленными белками (38 % против ожидаемых 32 %), большинство из которых относится к ДНК-связывающим транскрипционным факторам (64 %). Это свидетельствует о важной регуляторной роли белков хроматина с низким уровнем экспрессии.

Функциональные категории белков хроматина (согласно классификации SimChrom-SL, где белок отнесен только к одной категории) охарактеризованы по трем параметрам: медианной представленности белков категории, суммарной представленности белков категории и массовой доле категории среди белков хроматина (**Рисунок 6**). Категория гистонов, несмотря на сравнительно небольшое число уникальных белков (60), характеризуется максимальной суммарной представленностью (**Рисунок 6**). Высокая представленность (суммарная или по медианному значению) характерна также для белков следующих категорий: суперсемейство HMG; белки, осуществляющие процессинг РНК; шапероны гистонов; ремоделеры хроматина; ферменты ДНК (**Рисунок 6**). Медианные значения представленности функциональных категорий хроматина различаются более чем на четыре порядка, что свидетельствует о широком диапазоне концентраций, необходимых для выполнения их функций.

Массовая доля гистоновых белков среди белков хроматина составляет 10 %; при рассмотрении белков с единственной локализацией в ядре клетки —

38 %. Для ядерных РНК-связывающих белков эти показатели равны 30 % и 32 % соответственно (**Рисунок 6**). Выявлено, что гистоны, несмотря на небольшую длину (~130 а.о.), формируют основную массу хроматина за счёт высокой представленности, а белки процессинга РНК достигают сопоставимой массовой доли за счёт большей длины (медиана 504 а.о.).

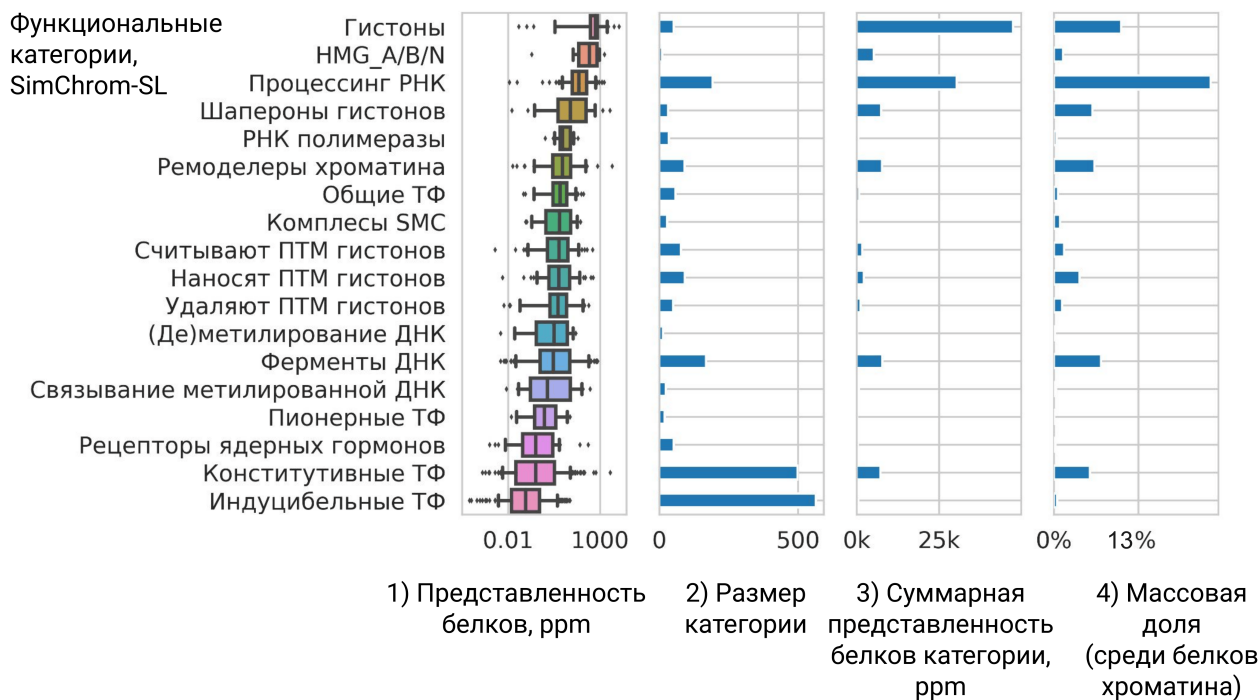


Рисунок 6 — Представленность белков функциональных категорий хроматина в клетках человека (классификация SimChrom-SL, данные из базы данных PaxDb). Показаны: 1 — распределение представленности белков хроматина, границы «ящика» соответствуют 25 и 75 перцентилям, «усы» — 5-95 перцентилям; 2 — количество уникальных белков в категории, 3 — суммарная представленность белков категории; 4 — массовая доля белков категории в хроматоме. Категории отсортированы по убыванию медианного значения представленности.

5.2. Физико-химические свойства белков хроматина

Выявлено, что медиана длины белков хроматина (количества аминокислотных остатков) статистически значимо превышает длину белков цитоплазмы: 496 а.о. против 464 а.о. соответственно (**Рисунок 7А**). В хроматине доля белков короче 200 а.о. ниже, чем в цитоплазме (8 % против 13 %). Наиболее короткими белками являются гистоны и белки суперсемейства HMG, тогда как к наиболее длинным относятся белки SMC-комплексов, а также белки,

участвующие в ремоделировании хроматина и модификации гистонов (Рисунок 7Б).

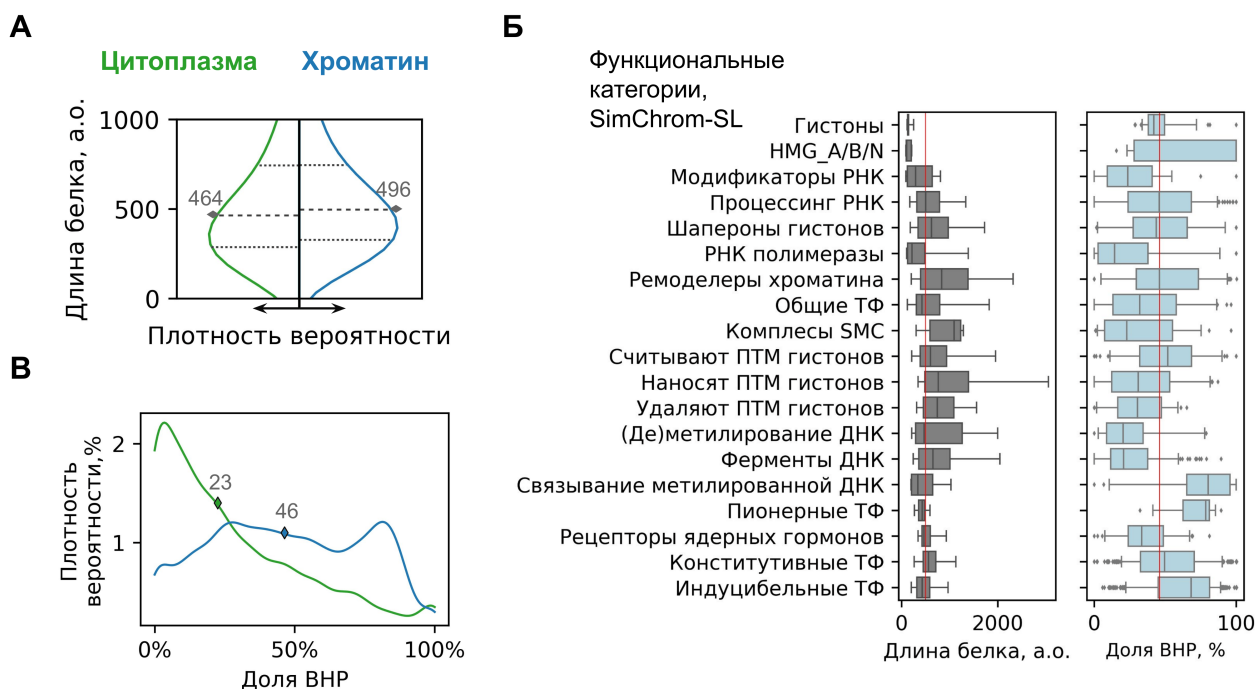


Рисунок 7 — Сравнение белков хроматина и цитоплазмы по длине аминокислотной последовательности и доле внутренне неупорядоченных регионов. (**А, В**) Распределения белков хроматина и цитоплазмы по длине (**А**) и доле ВНР (**В**); показаны медианные значения. (**Б**) Длина и доля ВНР в белках хроматина с разделением по функциональным категориям согласно классификации SimChrom-SL; границы «ящика» соответствуют 25 и 75 перцентилям, «усы» — 5-95 перцентилям, красной линией показано медианное значение для белков хроматина.

Установлено, что белки хроматина характеризуются более высокой долей внутренне неупорядоченных регионов (ВНР) по сравнению с белками цитоплазмы (медианы 46 % и 23 % соответственно) (Рисунок 7В). Это различие обусловлено большей длиной отдельных ВНР (медианы ~70 а.о. и ~50 а.о.) и большим количеством таких регионов в белках хроматина (медианы 2 и 1). Наибольший вклад в высокую долю внутренней неупорядоченности белков хроматина вносят индуцибельные ТФ, в которых эти регионы способствуют поиску целевых последовательностей ДНК и регуляции транскрипции посредством жидкофазного разделения.

Выявлено преобладание положительно заряженных белковых молекул в хроматине по сравнению с цитоплазмой (61 % против 36 % соответственно) (**Рисунок 8А**). Это обусловлено статистически значимо повышенным содержанием остатков лизина во всех регионах аминокислотной последовательности белков хроматина, остатков аргинина в упорядоченных регионах, а также сниженным содержанием остатков аспартата в ВНР. Преимущественно отрицательно заряженные белки хроматина принадлежат к категориям шаперонов гистонов и субъединицам комплексов РНК-полимераз (**Рисунок 8Б**).

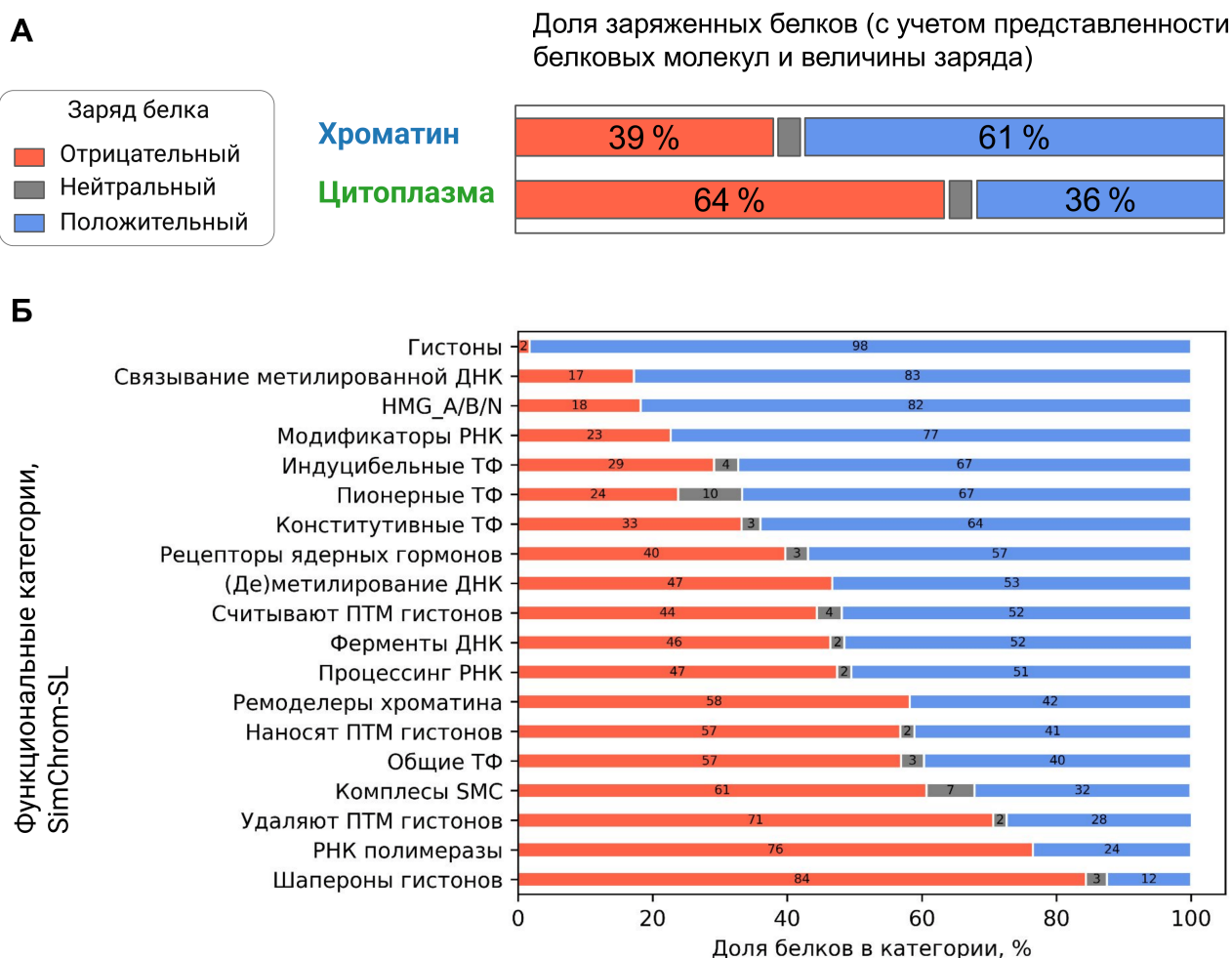


Рисунок 8 — Заряд белков хроматина. (**А**) Распределение белков хроматина и цитоплазмы по знаку заряда с учетом представленности белковых молекул. (**Б**) Распределение белков функциональных категорий хроматина (классификация SimChrom-SL) по знаку заряда.

Применение метода снижения размерности UMAP позволило выявить группы белков хроматина, статистически значимо различающиеся по аминокислотному составу (**Рисунок 9А**). К ним относятся гистоны (обогащены положительно заряженными а.о.); факторы сплайсинга (обогащены остатками серина и аргинина) и транскрипционные факторы с множественными ДНК-связывающими доменами типа «цинковый палец» (обогащены остатками цистеина и гистидина, медиана — 10 доменов).

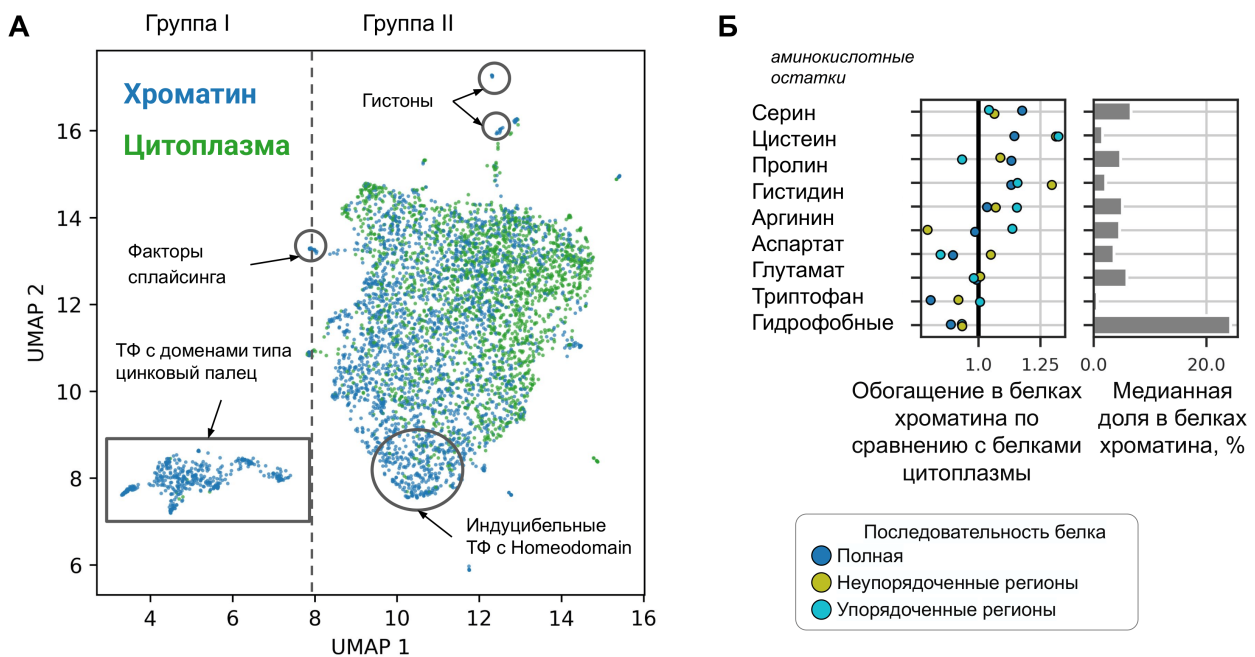


Рисунок 9 — Особенности аминокислотного состава белков хроматина человека. (**А**) Группы белков хроматина и цитоплазмы, выделенные по аминокислотному составу с использованием метода снижения размерности UMAP. (**Б**) Различия в аминокислотном составе белков хроматина по сравнению с белками цитоплазмы. Слева показаны остатки, которыми белки хроматина обогащены или обеднены по сравнению с белками цитоплазмы, с отдельным рассмотрением полной последовательности белков, ВУР и ВНР; справа приведены медианные доли аминокислотных остатков в белках хроматина.

Установлено, что белки хроматина статистически значимо обогащены остатками серина, пролина, цистеина и гистидина по сравнению с белками цитоплазмы (**Рисунок 9Б**). Повышенное содержание остатков серина и пролина обусловлено большей долей ВНР в белках хроматина, в которых эти аминокислоты встречаются чаще. Обогащение остатками цистеина и гистидина

связано с наличием ДНК-связывающих доменов типа «цинковый палец», представленных в нескольких копиях в транскрипционных факторах. В то же время для белков хроматина характерно пониженное содержание остатков триптофана и гидрофобных аминокислот.

5.3. Структурная организация белков хроматина

Для анализа доменной организации белков хроматина использовали аннотации функциональных доменов (из базы данных Pfam, Pfam-домены), а также экспериментально полученных и предсказанных структурных доменов (из баз данных CATH и TED, CATH-домены и TED-домены соответственно).

Согласно анализу структурных моделей белков хроматина, предсказанных методом AlphaFold2, упорядоченные регионы составляют 53 % от их совокупной аминокислотной последовательности (САП). Структурные домены составляют 35 % САП (TED-домены) и 25 % САП (CATH-домены). К функциональным доменам отнесено 39 % САП. Это свидетельствует о наличии функционально значимой, но структурно не охарактеризованной, части САП белков хроматина человека.

В белках хроматина человека идентифицировано 6246 уникальных структурных TED-домена (**Рисунок 10**).

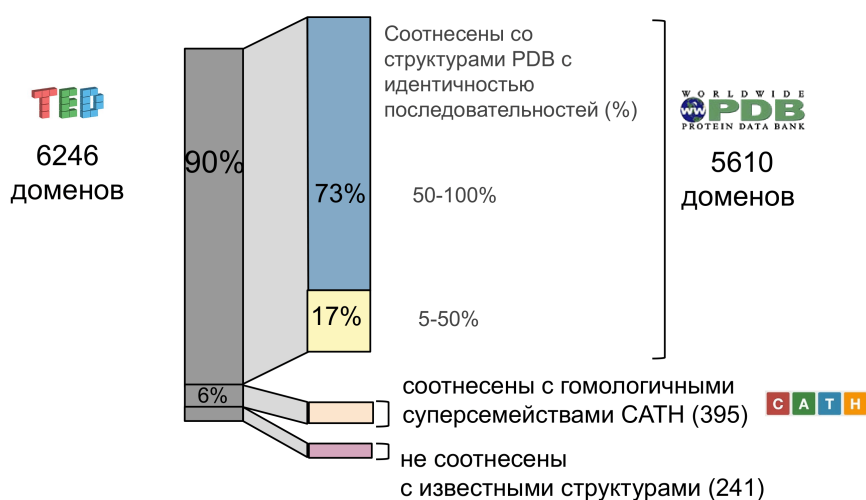


Рисунок 10 — Анализ предсказанных структурных TED-доменов белков хроматина человека. Показаны структурные соответствия для TED-доменов в базах данных PDB и CATH с различной степенью идентичности аминокислотных последовательностей.

Для большинства TED-доменов (5610 доменов) обнаружены структурные аналоги в базах данных PDB и CATH с различной степенью идентичности аминокислотных последовательностей (**Рисунок 10**). Для 241 TED-домена структурные соответствия отсутствуют, а для 118 из них также отсутствуют функциональные аннотации. Эти домены входят в состав 106 белков хроматина и представляют интерес для дальнейших структурных исследований.

5.4. Мультивалентные взаимодействия белков хроматина

Для анализа мультивалентных взаимодействий белков хроматина, реализуемых за счет сочетания в одном белке доменов с различными функциями, Pfam-домены белков, регулирующих эпигенетические процессы, были классифицированы по функциям с последующим расчетом условной вероятности их совместного присутствия (**Рисунок 11А**).

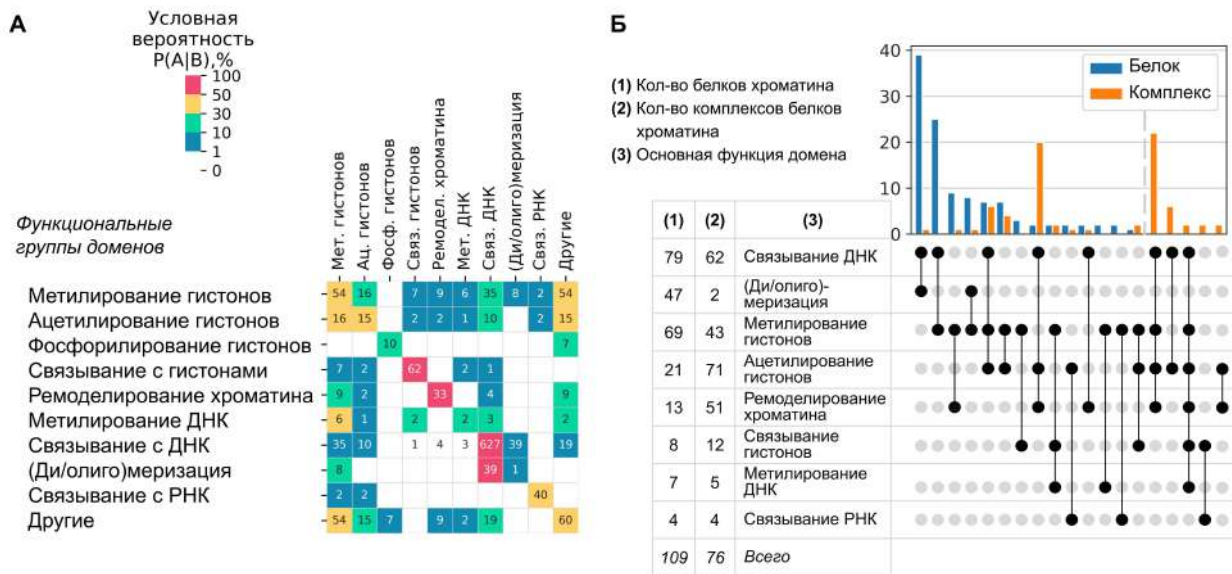


Рисунок 11 — Анализ потенциала мультивалентных взаимодействий в белках и стабильных белок-белковых комплексах хроматина человека. (**А**) Парные сочетания доменов в белках хроматина. Значения в ячейках отражают количество соответствующих белков хроматина; цвет ячеек отражает условную вероятность присутствия в белке домена одной функциональной группы А (в столбце) при наличии домена из другой В (в строке). (**Б**) Варианты сочетания доменов различных функциональных групп в отдельных белках и мультисубъединичных комплексах хроматина.

Выявлено, что условная вероятность сочетания доменов, ассоциированных

с метилированием гистонов, составляет 38 %. Эти домены сочетаются в белках с доменами других функциональных групп, в отличие от доменов, связанных с фосфорилированием (**Рисунок 11А**). Это может отражать эволюционную стратегию формирования устойчивых комбинаций доменов, связанных с метилированием и ацетилированием гистонов, участвующих в эпигенетической регуляции, при сохранении более универсальной роли фосфорилирования для регуляции клеточных сигнальных путей.

В индивидуальных белках выявлены комбинации доменов, относящихся к трем различным функциональным группам, тогда как в стабильных белок-белковых комплексах их число увеличивается вдвое (**Рисунок 11Б**). Полученные результаты свидетельствуют о том, что разнообразие сочетаний функциональных доменов формирует структурную основу многоуровневой эпигенетической регуляции генома.

III. ЗАКЛЮЧЕНИЕ

В диссертационной работе разработана эмпирическая функционально-ориентированная классификация белков хроматина человека SimChrom, полученная на основе интеграции гетерогенных данных об их репертуаре (из экспериментов, баз данных и литературы) и свойствах (результаты вычислительных расчетов). Разработанная классификация позволила объединить в рамках единой функционально-ориентированной системы следующие сведения о белках: функциональную аннотацию, субклеточную локализацию, физико-химические свойства, доменный состав и количественную представленность в клетке.

Применение классификации позволило охарактеризовать функциональные категории белков хроматина, выявить статистически значимые различия в их физико-химических свойствах, а также определить характерные функциональные домены и их сочетания, ассоциированные с мультивалентными взаимодействиями в хроматине. В белках хроматина выявлено более 200 структурных доменов, ранее не охарактеризованных экспериментально; они

являются потенциальными мишенями для дальнейших исследований методами структурной биологии и биофизики.

Полученные результаты демонстрируют, что разработанная классификация может служить инструментом обнаружения скрытых структурно-функциональных закономерностей белков хроматина и обладает эвристической ценностью. Она позволяет сопоставлять характерные признаки белков с их функциями, что дает возможность уточнять функции известных белков хроматина или предсказывать наиболее вероятные функции для ранее неохарактеризованных белков.

Для поддержки дальнейших исследований в этой области разработана интерактивная база данных SimChrom (<https://simchrom.intbio.org/>), в которой собраны инструменты для сопоставления аннотаций о локализации и функциях белков хроматина, исследования доменной организации и парных сочетаний доменов в составе белков хроматина.

Таким образом, полученные результаты формируют целостное представление о составе, свойствах и принципах организации белков хроматина человека, расширяют понимание молекулярных механизмов регуляции генома, а также закладывают основу для дальнейшего прогнозирования и уточнения функций белков хроматина и построения количественных моделей его пространственно-временной организации.

Основные выводы диссертации:

1. Проведена систематизация и мета-анализ разнородных источников информации о белковом составе хроматина и ядра клетки человека. Определены причины несогласованности данных между различными источниками информации, которые включают различия в подходах к аннотации множественной локализации белков, а также методические трудности детекции в составе фракции хроматина белков с низкой экспрессией или транзистентным характером взаимодействия.
2. Разработана эмпирическая функционально-ориентированная классификация белков хроматина человека SimChrom (39 категорий, 3045

белков), интегрирующая сведения о функциях, субклеточной локализации, физико-химических свойствах и доменной организации белков. Разработаны эталонные наборы белков ядра и цитоплазмы клетки человека с различным уровнем достоверности и согласованности данных между источниками информации.

3. На основе анализа масс-спектрометрических данных установлено, что доминирующими по массе компонентами хроматина являются гистоны и белки, участвующие в процессинге РНК: их массовая доля составляет около 40–60 % от общей массы всех белков хроматина.
4. Установлены функционально значимые физико-химические и структурные свойства белков хроматина человека.
 - 4.1. Показано, что белки хроматина по сравнению с белками цитоплазмы являются менее структурированными (медианные значения доли неупорядоченных регионов 46 % и 23 % соответственно), что обусловлено увеличением числа и длины неупорядоченных регионов (медианы 70 и 50 а.о.). Высокая неупорядоченность характерна для пионерных транскрипционных факторов и архитектурных белков семейства HMG, тогда как высокой упорядоченностью отличаются субъединицы РНК-полимераз.
 - 4.2. Выявлены особенности белков хроматина, обуславливающие их электростатические взаимодействия с другими белками и ДНК. В хроматине преобладают положительно заряженные белки (61 %), обогащенные аминокислотными остатками лизина, аргинина (в упорядоченных регионах) и обедненные остатками аспартата (в неупорядоченных регионах).
 - 4.3. В составе белков хроматина выявлены сочетания функциональных доменов, способствующие мультивалентным взаимодействиям. Среди них преобладают комбинации доменов, ассоциированных с метилированием и ацетилированием гистонов, а также со связыванием ДНК. Это свидетельствует о кооперативной эволюции

систем метилирования и ацетилирования гистонов для эпигенетической регуляции экспрессии генов.

СПИСОК ПУБЛИКАЦИЙ

Основные публикации Грибковой Анны Кирилловны по теме диссертационной работы в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности 1.5.8. Математическая биология, биоинформатика (физико-математические науки), в конце приведен вклад автора и объем публикации в печатных листах (п.л.):

- 1) **Gribkova A. K.**, Armeev G. A., Kirpichnikov M. P., Shaytan A. K. (Re)defining the human chromatinome: an integrated meta-analysis of localization, function, abundance, physical properties and domain composition of chromatin proteins // *Nucleic Acids Research*. — 2026. — Vol. 54. — № 2. — gkaf1489. DOI: 10.1093/nar/gkaf1489. Импакт-фактор 13,1 (JIF). 3 из 3,8 п.л.
- 2) Хасанова У. Н., **Грибкова А. К.**, Романова Т. А., Шайтан А. К., Армеев Г. А. Анализ белок-белковых взаимодействий пионерных транскрипционных факторов человека на основе предсказаний структур их комплексов // *Биофизика*. — 2025. — Т. 70 — № 5. — С. 837–853. EDN: XGEIMO. Импакт-фактор 0,593 (РИНЦ). 0,35 из 0,93 п.л.
- 3) Бигильдеев А. Е., Алексеев В. И., **Грибкова А. К.**, Тимохин Г. С., Комарова Г. А., Шайтан А. К. Роль изменений в структуре и динамике хроматина при COVID-19 // *Генетика*. — 2024. — Т. 60. — № 1. — С. 16–41. EDN: XIFEVZ. Импакт-фактор 0,673 (РИНЦ). 0,2 из 1,07 п.л.
- 4) Armeev G. A., **Gribkova A. K.**, Shaytan A. K. Nucleosomes and their complexes in the cryoEM era: Trends and limitations // *Frontiers in Molecular Biosciences*. — 2022. — Vol. 9. — P.1070489. EDN: BOMNID. Импакт-фактор 4,0 (JIF). 0,2 из 0,57 п.л.
- 5) Seal R. L., Denny P., Bruford E. A., **Gribkova A. K.**, Landsman D., Marzluff W. F., McAndrews M., Panchenko A. R., Shaytan A. K., Talbert P. A standardized nomenclature for mammalian histone genes // *Epigenetics and Chromatin*. — 2022. — Vol. 15. — P. 34. EDN: BLBOIK. Импакт-фактор 3,5 (JIF). 0,2 из 1,19 п.л.
- 6) Espiritu D., **Gribkova A. K.**, Shubhangi G., Shaytan A. K., Panchenko A. R. Molecular Mechanisms of Oncogenesis through the Lens of Nucleosomes and Histones // *Journal of Physical Chemistry B*. — 2021 — Vol. 125. — № 16. — Pp. 3963-3976. EDN: KANSFD. Импакт-фактор 2,9 (JIF). 0,45 из 1,17 п.л.
- 7) Armeev G. A., **Gribkova A. K.**, Pospelova I., Komarova G. A., Shaytan A. K. Linking chromatin composition and structural dynamics at the nucleosome level // *Current Opinion in Structural Biology*. — 2019. — Vol. 56. — Pp. 46–55. EDN: YPOJYJ. Импакт-фактор 7 (JIF). 0,2 из 0,79 п.л.