

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ "СКОЛКОВСКИЙ
ИНСТИТУТ НАУКИ И ТЕХНОЛОГИЙ"

На правах рукописи

Храмеева Екатерина Евгеньевна

**АРХИТЕКТУРА ХРОМАТИНА И ЕЕ РЕГУЛЯТОРНАЯ
РОЛЬ В КЛЕТКАХ ГОЛОВНОГО МОЗГА**

Специальность 1.5.8 —
«Математическая биология, биоинформатика»

Диссертация на соискание учёной степени
доктора биологических наук

Научный консультант:
доктор биологических наук, профессор
Гельфанд Михаил Сергеевич

Москва — 2023

Оглавление

	Стр.
Введение	5
Глава 1. Обзор литературы	23
1.1 Общие принципы организации хроматина	25
1.1.1 Хромосомные территории	25
1.1.2 Компарменты хроматина	26
1.1.3 Топологически ассоциированные домены (ТАДы)	27
1.1.4 Хроматиновые петли	29
1.1.5 Механизм образования хроматиновых петель и ТАДов	30
1.1.6 Другие уровни организации хроматина	33
1.2 Современное состояние исследований об архитектуре хроматина в мозге человека	34
1.2.1 Компарментализация в нейронах и глиальных клетках	35
1.2.2 ТАДы и петли в нейронах и глиальных клетках	36
1.2.3 Особые уровни организации хроматина в нейронах и глиальных клетках	38
1.2.4 Регуляторная роль организации хроматина в мозге человека	38
1.3 Существующие методы изучения архитектуры хроматина	39
1.3.1 Методы, основанные на микроскопии	40
1.3.2 Методы ЗС	41
1.3.3 Комбинация методов ЗС и микроскопии	43
1.3.4 Другие омиксные методы	43
Глава 2. Материалы и методы	46
2.1 Стандартные методы обработки данных Hi-C	46
2.1.1 Построение карт контактов на основе данных Hi-C	46
2.1.2 Разметка хроматиновых компарментов в данных Hi-C	47
2.1.3 Разметка ТАДов и петель	47
2.2 Разработанные нами методы анализа данных Hi-C	47

2.2.1	Суммарная частота контактов геномного региона является его биологическим свойством, а не технической погрешностью	48
2.2.2	Выявление биологического сигнала в данных с высоким уровнем технической вариабельности	57
2.2.3	Метод аннотации ТАДов на основе обогащения эпигенетическими метками	68
2.2.4	Применение методов машинного обучения для изучения особенностей организации хроматина	77
2.2.5	Разработка методов интеграции данных Hi-C с другими омиксными данными	83
2.2.6	Разработка методов анализа липидного и метаболического состава мозга человека	90
Глава 3. Результаты и обсуждение		97
3.1	Изучение общих принципов организации хроматина	97
3.1.1	Хромосомные территории	97
3.1.2	Компартменты хроматина	103
3.1.3	Топологически ассоциированные домены (ТАДы)	105
3.1.4	Хроматиновые петли	108
3.2	Изучение организации хроматина в нейронах мозга человека	114
3.2.1	Оптимизация протокола Hi-C для образцов мозга	114
3.2.2	Меж-хромосомные взаимодействия в нейронах понижены	116
3.2.3	Компартментализация в нейронах менее выражена	117
3.2.4	ТАДы в нейронах более выражены	119
3.2.5	Нейрональные петли более длинные	121
3.2.6	Hi-C взаимодействия в нейронах человека	124
3.3	Интеграция данных Hi-C с другими омиксными данными в контексте изучения мозга человека	127
3.3.1	Изучение структуры и функции клеток мозга человека на различных уровнях молекулярной организации	127
3.3.2	Роль организации хроматина в регуляции специфичной для человека экспрессии в мозге	154
Заключение		156

Публикации автора по теме диссертации	158
Список сокращений и условных обозначений	165
Список литературы	170
Список рисунков	191
Список таблиц	209

Введение

Актуальность темы. Геномы млекопитающих характеризуются сложной 3D архитектурой с множеством уровней организации, представляющих собой многослойную систему с определенной функциональностью. С развитием технологии захвата конформации хромосом (3C) было обнаружено, что эукариотические геномы организованы иерархически [1]. От крупного к мелкому масштабу, в ядре представлены хромосомные территории, компартменты хроматина, топологически ассоциированные домены (ТАДы), петли хроматина и нуклеосомы (рис. 1А). Индивидуальные хромосомы четко разделены в трехмерном пространстве ядра, с образованием хромосомных территорий - ядерных областей, преимущественно занятых различными интерфазными хромосомами [2]. Данные, полученные с помощью технологии высокопроизводительного захвата конформации хромосом (Hi-C, рис. 1Б) показали, что хромосомы дополнительно разделены на два компартмента. Компартмент А состоит из активных эпигенетических меток и активно транскрибируемых генов, обычно расположенных в центре ядра. Компартмент В состоит из репрессивных эпигенетических меток и неактивных генов, расположенных рядом с ядерной ламиной [3]. Считается, что формирование компартментов А и В обусловлено комбинацией факторов, включающей распределение активных и репрессивных меток хроматина [4].

Компартменты хроматина дополнительно разделены на ТАДы - геномные области с частыми взаимодействиями внутри них и высокой изоляцией от соседних ТАДов [5]. Предполагается, что ТАДы являются основными функциональными регуляторными доменами, модулирующими контакты между энхансерами и промоторами. Повышенная частота контактов внутри ТАДов опосредует физическое взаимодействие между парами энхансер-промотор, в то время как высокая изоляция на границах ТАДов ограничивает такие взаимодействия для энхансер-промоторных пар, расположенных в соседних ТАДах. Границы ТАДов очень стабильны между видами и обогащены фактором связывания ССТС (СТСФ) и когезином у млекопитающих. СТСФ и когезин образуют петли хроматина, которые закрепляют границы ТАДов, обеспечивая структурную основу для образования ТАДов. Считается, что петли хроматина формируются с помощью модели петлевой экструзии, основанной на взаимодействии СТСФ и

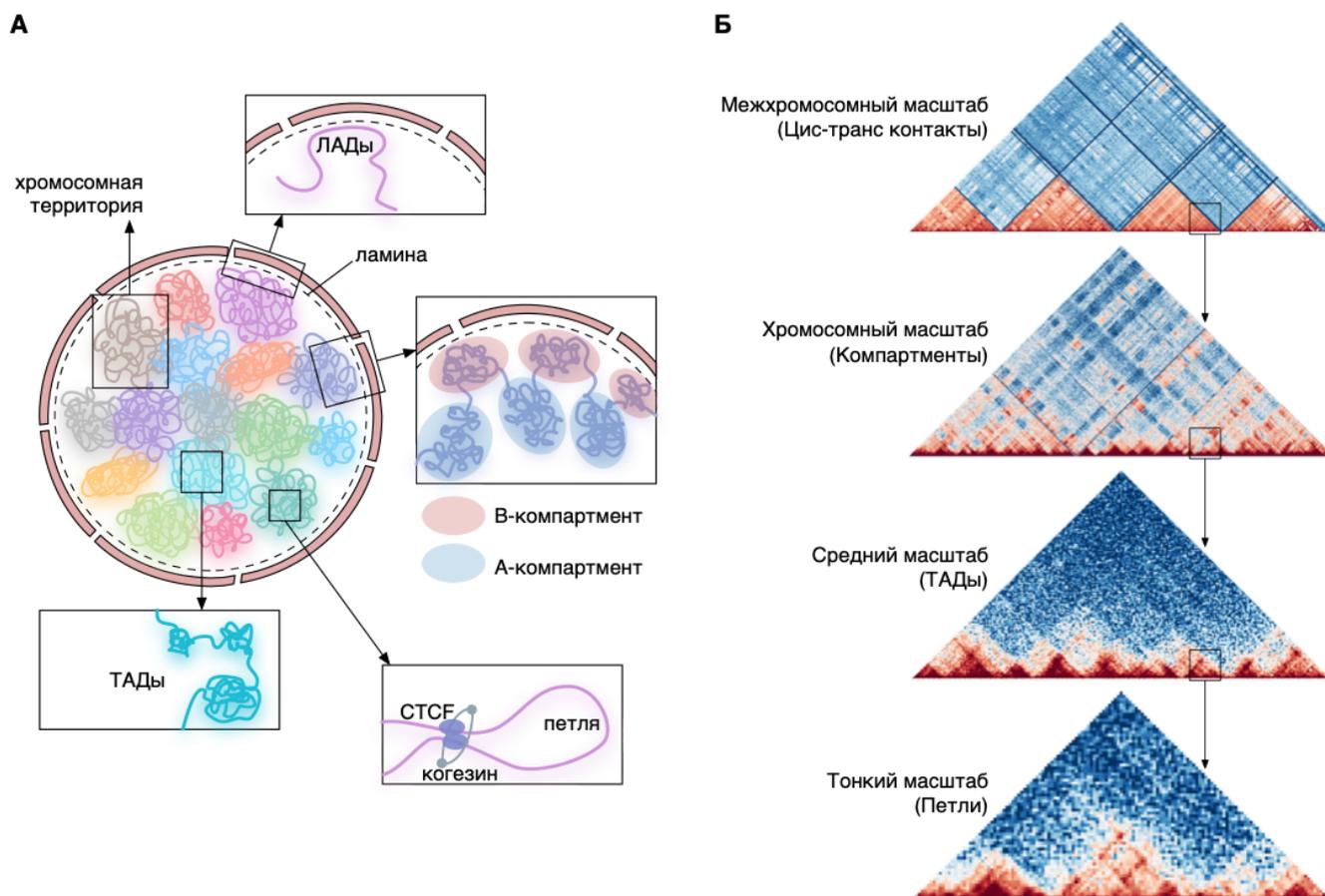


Рисунок 1 — Общие принципы организации хроматина. (А) Схематичное изображение основных уровней организации хроматина в ядре. (Б) Уровни организации хроматина в данных Hi-C. Яркость пикселей на картах Hi-C пропорциональна частотам контактов.

комплекса когезина [6]. Физические взаимодействия между дистальными участками генома позволяют линейно удаленным друг от друга элементам, таким как промоторы и энхансеры, встретиться друг друга. Дальние контакты между энхансерами и промоторами могут регулироваться транскрипционными факторами, белком Mediator, РНК-полимеразой II и некодирующими РНК, которые дополнительно регулируют экспрессию генов [7].

Хотя трехмерная структура хроматина глобально стабильна, недавние исследования показывают, что отдельные гены часто переключаются между активными и неактивными компартментами во время развития. Кроме того, специфические взаимодействия внутри и вне ТАДов часто изменяются [8]. Наблюдаемые ассоциации между компартиментализацией хроматина, профилем ТАДов и транскрипционной активностью указывают на возможную причинно-следственную связь между ними. Так, сравнение карт Hi-C у плодовой мушки

дрозофилы до и после подавления транскрипции показывает, что трехмерная структура хроматина регулирует экспрессию генов и, в свою очередь, также может быть модулирована генами, связанными с транскрипцией [9; 10]. Таким образом, сложное устройство архитектуры хроматина внутри ядра является важнейшим аспектом регуляции различных клеточных процессов и поддержания целостности генома.

Учитывая тесную связь архитектуры генома с экспрессией генов, изучение деталей трехмерной организации генома могло бы прояснить механизмы, лежащие в основе тканеспецифичности, а также продвинуть понимание патологических процессов, приводящих к заболеваниям. Таким образом, исследования, сосредоточенные на мозговой ткани, являются одними из самых востребованных в настоящее время ввиду недостаточного понимания сложного процесса развития, функционирования мозга и его дегенеративных изменений. Мало работ исследовали разницу в архитектуре генома между нейронами и различными глиальными клетками [11; 12]. Однако они заложили основу исследований хроматина в головном мозге, продемонстрировав, что, хотя клетки мозга в целом соответствуют фундаментальным принципам трехмерной организации генома, они также демонстрируют уникальные особенности архитектуры хроматина.

В контексте этих исследований, представляется важным изучать молекулярную организацию мозга человека комплексно и анализировать организацию хроматина совместно с экспрессией генов, эпигенетическими модификациями, связыванием транскрипционных факторов, открытостью хроматина и метаболизмом клеток в целом. Сопоставление данных Hi-C с другими омиксными данными позволяет определять механизмы регуляции экспрессии генов, опосредованные трехмерной структурой хроматина. Архитектура хроматина организует контакты энхансеров и промоторов и таким образом регулирует уровни экспрессии генов, которые определяют количество структурных белков, ферментов и, опосредованно, синтезируемых ими метаболитов, что в комплексе и определяет фенотип и особенности функционирования клеток мозга человека.

Таким образом, несмотря на центральную роль трехмерной организации генома в регуляции функционирования человеческого мозга, исследования в этой области ограничены сложностью его анатомии и гетерогенностью его клеток. Тем не менее, с появлением высокопроизводительного секвенирования и усовершенствованных методов изучения пространственной организации гено-

ма теперь стало возможным исследовать такие сложные образцы. Несколько недавних работ начали проливать свет на ранее неизвестные аспекты архитектуры хроматина в различных типах клеток головного мозга, а также на роль организации хроматина в регуляции экспрессии генов. Однако для выяснения деталей организации хроматина в основных типах клеток головного мозга требуются всесторонние исследования в этой области. Результаты таких исследований могли бы предоставить ценную информацию об основных механизмах функционирования мозга как на клеточном, так и на органном уровнях.

Целью данной работы является установление роли трехмерной организации генома в регуляции функционирования человеческого мозга, с учетом гетерогенности его клеточного состава и в комплексе с другими уровнями молекулярной организации клеток: экспрессией генов, эпигенетическими модификациями, связыванием транскрипционных факторов, открытостью хроматина и метаболизмом клеток в целом.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Разработать методологию биоинформатического анализа данных Hi-C для образцов мозга человека: построить карты контактов, провести анализ хроматиновых компартментов, ТАДов и петель.
2. Решить проблему разной экспериментальной представленности геномных регионов в данных Hi-C.
3. Разработать методику выявления биологического сигнала в данных Hi-C с высоким уровнем технической вариабельности.
4. Оптимизировать метод аннотации ТАДов на основе обогащения эпигенетическими метками.
5. Адаптировать и применить методы машинного обучения для исследования организации хроматина.
6. С помощью разработанных подходов к анализу данных Hi-C, описать особенности трехмерной организации генома в нейронах и других клетках мозга человека, а также установить функциональную роль дальних взаимодействий, которые опосредованы PcG белками и присутствуют только в нейронах.
7. Раскрыть общие принципы организации хроматина, такие как иерархичность ТАДов, в том числе в одиночных клетках, роль ТАДов в ре-

- гуляции транскрипции, роль ядерной ламины, и др., чтобы поместить наблюдения для нейронов в контекст общих представлений об организации хроматина.
8. Разработать методы анализа данных RNA-seq, ATAC-seq, ChIP-seq и др. для решения нестандартных задач в контексте изучения мозга человека, таких как анализ биаллельной транскрипции, механизмы регуляции альтернативного сплайсинга (данные iCLIP), анализ некодирующих РНК, разделение сигнала от тканей человека и мыши в ксенотрансплантатах (PDX) для изучения онкозаболеваний, вычислительное восстановление видовой принадлежности отдельных клеток (данные snRNA-seq).
 9. Разработать методы анализа липидного и метаболического состава мозга человека и с помощью них установить роль липидов и конкретных метаболитов в функционировании клеток головного мозга человека в норме и при заболеваниях, в частности при психических расстройствах.
 10. Оценить роль неандертальских вариантов в качестве возможной причины липидных отличий, специфичных для европейской популяции, и оценить роль неандертальских вариантов в других популяциях, в частности в популяциях кетов и папуасов.
 11. Разработать методы интеграции данных Hi-C с другими омиксными данными: RNA-seq, ATAC-seq, ChIP-seq и др., при помощи которых описать функционирование клеток мозга комплексно, на различных уровнях молекулярной организации, чтобы установить регуляторную роль деацетилазы SIRT6 и архитектурного белка YY1, участвующего в образовании хроматиновых петель.
 12. Установить роль организации хроматина в регуляции человек-специфичной экспрессии генов в мозге человека.

Основные положения, выносимые на защиту:

1. Разработанная нами методика учета надежности контактов на основе сравнения экспериментальных повторностей позволяет выявлять биологический сигнал в данных Hi-C с высоким уровнем технической вариабельности. Суммарная частота контактов геномного региона в картах Hi-C является его биологическим свойством, а не технической погрешностью.

2. Добавление функциональной информации о хроматиновых метках и архитектурных белках к структурной информации о контактных частотах позволяет получать оптимизированную аннотацию топологических доменов.
3. ТАДы организованы иерархически в тканях, но не в одиночных клетках. Их границы плохо предсказываются по профилю связывания CTCF у дрозофилы, в отличие от человека, а значит в формирование ТАДов вносят вклад другие факторы, такие как ацетилирование гистонов. Позиции связывания белка Chromator и гистоновые модификации H3K4me3, H3K27ac являются наиболее информативными признаками для предсказания профиля ТАДов, как показали методы машинного обучения.
4. Разрушение ламины увеличивает плотность хроматина во фракции ТАДов, обогащенных активным хроматином, и приводит к отдалению хроматина от ядерной оболочки. На примере сперматоцитов мы показали, что инициация транскрипции коррелирует с отдалением промоторов от ламины и с локальной пространственной изоляцией прилегающих к ним областей. А активация транскрипции гена альфа-глобина коррелирует с усилением регуляции нескольких соседних генов домашнего хозяйства, таким образом влияя на структуру хроматина не только локально.
5. Хроматин имеет специфическую структуру в нейронах, по сравнению с другими клетками мозга человека: хромосомные территории более выражены, при одновременном драматическом снижении компарментализации. ТАДы более выражены в нейронах по сравнению с другими клетками мозга, что объясняется повышенной концентрацией когезина в нейронах, а нейрональные петли значительно длиннее, что согласуется с обогащением загрузчика когезина NIPBL в нейронах. Дальние взаимодействия, опосредованные PcG белками, присутствуют только в нейронах, где выполняют функцию подавления развитийных транскрипционных факторов.
6. Построение подробной транскриптомной карты мозга человека в сравнении с другими приматами позволило установить, какие регионы у человека быстрее эволюционируют: кора больших полушарий, гипота-

- лабус и серое и белое вещества мозжечка. Анализ с разрешением в одну клетку показал, что глиальные клетки эволюционировали быстрее нейронов на человеческой линии.
7. Разработанные нами методики анализа данных RNA-seq, ATAC-seq, ChIP-seq, iCLIP, DamID позволяют решать нестандартные задачи в контексте изучения мозга человека, например: анализ биаллельной транскрипции и механизмов регуляции альтернативного сплайсинга примере гена SF1 человека, который играет ключевую регуляторную роль в нейронах гипоталамуса; анализ некодирующих РНК, которые важны для функционирования мозга человека и, как мы показали, отличаются человек-специфичной экспрессией; разделение сигнала от тканей человека и мыши в ксенотрансплантатах для изучения онкозаболеваний.
 8. С их помощью, а также с использованием разработанного нами метода анализа липидного и метаболического состава мозга человека, мы установили, что деацетилаза SIRT6 играет ключевую роль в регуляции митохондриальных процессов в головном мозге как напрямую, так и посредством образования комплекса с архитектурным белком YY1, участвующем в образовании хроматиновых петель.
 9. В сравнении с другими видами млекопитающих, у человека липидные изменения локализованы преимущественно в мозге и вовлечены в метаболические пути, связанные с когнитивными заболеваниями, такими как болезнь Альцгеймера, Паркинсона, и другими нарушениями работы нервной системы. При этом концентрации липидов и экспрессия генов значительно изменялись от шимпанзе к человеку преимущественно в метаболических путях, связанных с катаболизмом липидов, у европейцев по сравнению с другими популяциями.
 10. Анализ геномных последовательностей генов липидного катаболизма показал, что частота неандертальских вариантов выше в три раза в этих генах, причем только у европейцев (20,8%), по сравнению со средним по геному значением (6,1%). По этим генам шел положительный отбор у современных европейцев, но не у азиатов и африканцев.
 11. В популяции кетов повторного события скрещивания с неандертальцами не происходило, а сохранение аллелей неандертальцев и денисовцев в геномах папуасов могло быть связано с адаптацией. Детальное

- исследование отличий в липидном и метаболитном составе головного мозга между современными популяциями продемонстрировало значимые различия у представителей китайской народности Хань, причем в основном в путях, связанных с липидным метаболизмом.
12. При аутизме, представляющем собой нарушение эволюционно новых механизмов работы мозга, наблюдаются существенные изменения метаболизма в префронтальной коре головного мозга, достигающие 15% от общего количества измеренных метаболитов. При ответе на антидепрессант флуоксетин, применяемый при депрессии, также происходят существенные изменения метаболизма в префронтальной коре, включая значимые изменения в количестве 106 липидов, сопровождающиеся незначительными изменениями в экспрессии генов.
 13. Метаболический путь пуринов нарушается при развитии аутизма и при этом является эволюционно новым. Этот путь содержит фермент аденилосукцинатлиазу (ADSL), несущий аминокислотную замену, которая присутствует у современных людей, но отсутствует у неандертальцев. С использованием гуманизированной мышинной модели мы показали, что изменение пуринового метаболизма произошло у людей после их отделения от предка, общего с неандертальцами и денисовцами.
 14. В целом, ускорение эволюции экспрессии на человеческой ветке связано с консервативными участками генома, накопившими человек-специфические отличия в последовательности. Эти участки обогащены энхансерами генов с человек-специфической экспрессией, и их регуляция может быть опосредована архитектурой хроматина.

Научная новизна. Одной из нерешенных задач в исследованиях головного мозга является соотнесение молекулярной структуры мозга с его функциональными свойствами. Особенно сложно это в случае головного мозга человека, который характеризуется с одной стороны высочайшей степенью сложности своей организации, а с другой стороны множественным спектром выполняемых когнитивных функций. Исследования, проведенные ранее на уровне экспрессии генов, позволили выявить множественные отличия в экспрессии, уникальные для человеческого мозга, что привело в итоге к открытию возможных механизмов, лежащих в основе когнитивной уникальности человека, например, более продолжительный синаптогенез и удлиненный период формирования синапсов

в неокортексе у человека [13; 14]. Однако, в то же время, исследования, проводимые на уровне экспрессии генов, не позволили выявить функциональные особенности, связанные с теми многообразными функциями в когнитивных процессах, которые выполняют различные регионы головного мозга, в частности, различные зоны неокортекса. Несмотря на то, что исследования на уровне экспрессии, использующие такие методические подходы как секвенирование РНК, выделенной из единичной клетки или даже единичного ядра, и позволяющие проводить изучение организации мозга с высокой разрешающей способностью, могут пролить свет на имеющиеся различия в молекулярной организации между зонами неокортекса, становится очевидным, что глубокое понимание того, как функционирует мозг, невозможно без проведения всестороннего анализа молекулярной организации головного мозга. Важный уровень молекулярной организации мозга, которому до недавнего времени не было уделено внимания в работах, проводимых в данной области, - это трехмерная архитектура генома. В настоящей работе:

1. Разработан ряд новых подходов для анализа данных Hi-C, направленных на понижение технической вариабельности и решение других технических проблем в данных Hi-C, в том числе с использованием современных методов машинного обучения, а также методика интеграции данных Hi-C с другими типами омиксных данных.
2. Применение разработанных подходов позволило впервые изучить трехмерную организацию хроматина в нейронах и других клетках региона VA22r коры мозга человека и показать, что хроматин имеет специфическую структуру в нейронах, по сравнению с другими клетками мозга, вследствие чего многие биологические процессы, включая экструзию петель, компарментализацию, дальние взаимодействия, вероятно, работают в нейронах по-другому, что способствует правильной экспрессии генов в клетках этого типа.
3. Впервые подробно описаны дальние взаимодействия в нейронах, которые опосредованы PcG белками и практически отсутствуют в других типах клеток, и установлена их функциональную роль, которая заключается в подавлении развитийных транскрипционных факторов.
4. Раскрыты функциональные свойства и механизм формирования ТАДов и показано, что ТАДы организованы иерархически в тканях, но

- не в одиночных клетках, и что их границы плохо предсказываются по профилю связывания CTCF у дрозофилы, в отличие от человека, а значит в формирование ТАДов могут вносить вклад другие факторы, такие как ацетилирование гистонов.
5. Впервые показана ключевая роль ядерной ламины в организации хроматина в целом и ТАДов в частности, и продемонстрирована важность ТАДов для активации транскрипции на примере сперматогенеза и альфа-глобинового локуса.
 6. Разработаны методики анализа данных RNA-seq, ATAC-seq, ChIP-seq и др. для решения нестандартных задач в контексте изучения мозга человека, таких как анализ биаллельной транскрипции, механизмы регуляции альтернативного сплайсинга (данные iCLIP), анализ некодирующих РНК, разделение сигнала от тканей человека и мыши в ксенотрансплантатах (PDX) для изучения онкозаболеваний, вычислительное восстановление видовой принадлежности отдельных клеток (данные snRNA-seq).
 7. Впервые разработан метод анализа липидного и метаболического состава мозга человека, применение которого показало, что деацетилаза SIRT6 играет ключевую роль в регуляции митохондриальных процессов в головном мозге через взаимодействие с архитектурным белком YY1, участвующим в образовании хроматиновых петель.
 8. С помощью этой же методики впервые найдены: (а) липиды, специфические для человека по сравнению с другими видами, в том числе в головном мозге; (б) отличия в липидном и метаболическом составе головного мозга между популяциями современных людей; (в) существенные метаболические изменения при аутизме и при ответе на антидепрессант флуоксетин в коре головного мозга.
 9. Впервые подробно описан метаболический путь биосинтеза пуринов и участвующий в нем фермент аденилосукцинатлиаза (ADSL), несущий аминокислотную замену, которая присутствует у современных людей, но отсутствует у неандертальцев, а также определено влияние этой замены на концентрации метаболитов из пути биосинтеза пуринов.
 10. Разработана методика интеграции омиксных данных разных типов, с помощью которой впервые показано, что концентрации липидов и экс-

прессия генов значительно изменялись от шимпанзе к человеку в метаболических путях, связанных с катаболизмом липидов, у европейцев по сравнению с другими популяциями.

11. Впервые показано, что частота неандертальских вариантов выше среднего по геному значения в три раза в генах катаболизма липидов, причем только у европейцев. По этим генам шел положительный отбор у современных европейцев, но не у азиатов и африканцев, что указывает на возможную роль этих генов в адаптации к условиям окружающей среды доисторической Европы. Сохранение аллелей неандертальцев и денисовцев в геномах папуасов также могло быть связано с адаптацией.
12. Впервые установлено, что организация хроматина участвует в регуляции экспрессии генов в мозге человека, в особенности в участках генома, характеризующихся ускоренной эволюцией на человеческой ветке.

Научная и практическая значимость. Результаты данной работы приближают нас к пониманию того, как ДНК упакована внутри ядра клетки, и как особенности упаковки влияют на регуляцию генов, а значит помогают ответить на ключевые вопросы современной клеточной биологии и высокотехнологичной медицины. Решение этих вопросов необходимо для идентификации механизмов целого спектра социально значимых заболеваний человека, поскольку упаковка хромосом тесно связана с наследственностью и болезнями. Например, нарушение организации хромосом приводит к нарушениям развития и онкологическим заболеваниям: нейробластоме, лейкемии, медуллобластоме, глиоме, раку прямой кишки. Эти онкологические заболевания являются многофакторными, и у каждого пациента могут быть вызваны поломками в разных местах генома. Таким образом, изучение всего многообразия нарушений работы генов, связанных с изменениями в упаковке хромосом, необходимо при переходе к персональной медицине, для разработки новых высокотехнологичных (как и все современные методы лечения онкологических заболеваний) методов лечения и диагностики.

Получение карт конформации хромосом с высоким разрешением для образцов мозга человека само по себе имеет большую практическую значимость, поскольку закладывает основу для исследования причин патологий мозга, таких как психические расстройства, нейродегенеративные и онкологические заболевания. Предыдущие исследования указывали на важную роль архитектуры

хромосом в заболеваниях мозга. Сопоставление данных о конформации хромосом с экспериментами RNA-seq, ATAC-seq, ChIP-seq и другими позволяет интегрировать информацию о транскрипции, открытости и модификациях хроматина и изменениях конформации хромосом. Совместный анализ этих данных позволяет получать полную картину регуляторного ландшафта в нормальных клетках мозга человека и изучать возможные причины различных патологий, связанные с регуляторной ролью архитектуры хроматина.

Усовершенствование существующих и разработка новых подходов к анализу биологических данных по пространственной структуре хроматина, а также методов интеграции этих данных с разнородными данными по функциональным характеристикам генома позволяет изучать функции и механизмы формирования различных элементов организации хроматина (например, ТАДов и хроматиновых петель) вычислительными методами без выполнения дорогостоящих экспериментов. Возможность сопоставления данных Hi-C с другими омиксными данными позволяет получать целостную картину изменений структурной и функциональной организации хромосом, связанных с тем или иным заболеванием, а значит приблизиться к пониманию механизмов заболеваний. Без детального понимания механизмов заболеваний невозможна разработка новых лекарственных препаратов и подходов к лечению заболеваний человека. Таким образом, понимание принципов и механизмов пространственной организации генома является одной из ключевых проблем современного здравоохранения.

Потенциал для практического применения данных Hi-C в медицине огромен: наличие Hi-C данных в нескольких временных точках позволяет проследить изменения в ходе заболевания в динамике. Например, проследить накопление геномных перестроек с течением онкологического заболевания, или изучить динамику изменений при нарушениях развития. С удешевлением секвенирования, которое активно происходит в настоящее время, получение таких данных Hi-C станет ещё более актуальным, в том числе для персонализированной медицины. А значит в будущем потребность в инструментах для анализа данных Hi-C в медицинских целях будет только возрастать.

Степень достоверности полученных результатов обеспечивается использованием современных экспериментальных и вычислительных методов с необходимым количеством валидационных экспериментов и контрольных анализов, а также подтверждается публикацией результатов в ведущих рецензиру-

емых журналах. Сравнение с литературными источниками подтверждает, что наши результаты находятся в соответствии с результатами, полученными позднее другими авторами.

Апробация работы. Основные результаты работы докладывались на следующих международных и российских конференциях:

1. Международная конференция "18th annual international conference on Intelligent Systems for Molecular Biology (ISMB2010)". Бостон, США, 2010 г., устный доклад. Spatial proximity and similarity of functional states of genome domains. E. Khrameeva, A.A. Mironov, Ph. Khaitovich, and M.S. Gelfand.
2. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'10)". Геленджик, 2010 г., устный доклад. Functional similarity and chimeric transcripts in spatially close genome domains. E.E. Khrameeva, A.A. Mironov, M.S. Gelfand.
3. Международная конференция "24th International Mammalian Genome Conference". Крит, Греция, 2010 г., устный доклад. Association between spatial proximity and functional similarity in human genome. Ekaterina Khrameeva, Andrey Mironov, Mikhail Gelfand.
4. Международная конференция "Albany 2011: The 17th Conversation". Олбани, США, 2011 г. The Impact of Interchromosomal Associations on the Functional State of the Human Genome. Ekaterina E. Khrameeva, Andrey A. Mironov, Mikhail S. Gelfand.
5. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМВ'11)". Москва, 2011 г., устный доклад. Interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. Ekaterina E. Khrameeva, Mikhail S. Gelfand.
6. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'11)". Геленджик, 2011 г., устный доклад. Comparison of 117 mRNA and genome sequencing experiments between laboratories and platforms. Ekaterina E. Khrameeva, Mikhail S. Gelfand.
7. Международная конференция "Cold Spring Harbor Asia. Bioinformatics of Human and Animal Genomics". Суджоу, Китай, 2011 г., устный до-

- клад. Spatial proximity and similarity of the epigenetic state of genome domains. E. Khrameeva, A.A. Mironov, P. Khaitovich, and M.S. Gelfand
8. Международная конференция "RECOMB-seq 2012". Барселона, Испания, 2012 г. Biases in read coverage demonstrated by interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. Khrameeva EE, Gelfand MS.
 9. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'12)". Петрозаводск, 2012 г., устный доклад. Regulation of alternative splicing by hnRNPL protein. Ekaterina E. Khrameeva.
 10. Международная конференция "Chromosomes, Stem Cells and Disease". Барселона, Испания, 2012 г., постерный доклад. Spatial proximity and similarity of the epigenetic state of genome domain. Ekaterina E. Khrameeva.
 11. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМВ'13)". Москва, 2013 г., устный доклад. Association between changes in lipid catabolism and increased gene flow from Neanderthals to Europeans. Khrameeva E., Gelfand M., Khaitovich P.
 12. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'13)". Калининград, 2013 г., устный доклад. Increased similarity between Neanderthal and European genomes is associated with changes in lipid catabolism. E.E. Khrameeva, P. Khaitovich.
 13. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'14)". Нижний Новгород, 2014 г., постерный доклад. Анализ структуры хроматина в двух клеточных линиях дрозофилы. Е.Е. Храмеева.
 14. Международная конференция "Albany 2015: The 19th Conversation". Олбани, США, 2015 г., устный доклад. Active chromatin regions are sufficient to define borders of topologically associated domains in *D. melanogaster* interphase chromosomes. Khrameeva EE, Ulyanov SV, Gavrilov AA, Shevelyov YY, Gelfand MS, Razin SV.
 15. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМВ'15)". Москва, 2015 г.,

- устный доклад. Active chromatin regions are sufficient to define borders of topologically associated domains in *D. melanogaster* interphase chromosomes. Khrameeva EE, Ulyanov SV, Gavrillov AA, Shevelyov YY, Razin SV.
16. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'15)". Сочи, 2015 г., устный доклад. Взаимосвязь хромосомных перестроек и пространственной структуры хроматина в дрожжах. Храмеева Е.Е., Фуденберг Д., Гельфанд М.С., Мирный Л.
 17. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'16)". Репино, 2016 г., устный доклад). Молекулярные маркеры продолжительности жизни человека. Е. Храмеева, И. Курочкин.
 18. Всероссийская конференция "V Съезд биохимиков России". Дагомыс, 2016 г., устный доклад. Геномика аутизма. Е. Храмеева, В. Степанова.
 19. Международная конференция "шшез". Новосибирск, 2017 г., устный доклад. Гены катаболизма липидов обогащены неандертальскими вариантами у современных европейцев. Е. Храмеева.
 20. Международная конференция "Society for Molecular Biology and Evolution Meeting (SMBE'17)". Остин, США, 2017 г., постерный доклад. Evolution of lipid concentrations in mammalian tissues. E. Khrameeva, I. Kurochkin, P. Khaitovich.
 21. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМВ'17)". Москва, 2017 г., устный доклад. Factors shaping structural organization of chromatin. E. Khrameeva.
 22. Международная конференция "Bioinformatics: from Algorithms to Applications (BiATA'17)". Санкт-Петербург, 2017 г., устный доклад. Studying chromatin 3D structure in eukaryotes. E. Khrameeva.
 23. Международная конференция "Клиническая протеомика. Постгеномная медицина". Москва, 2017 г., устный доклад. Metabolic changes in autism and their evolution. E. Khrameeva.
 24. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМВ'19)". Москва, 2019 г., устный доклад. Transcriptome map of the human, chimpanzee, bonobo, and

- macaque brains at the single-cell resolution. E. Khrameeva, I. Kurochkin, P. Khaitovich.
25. Всероссийская конференция с международным участием "Школа по регенеративной медицине". Москва, 2021 г., устный доклад. Single-cell resolution transcriptomics: applications for the human brain research. E. Khrameeva.
 26. Международная конференция "Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022)". Новосибирск, 2022 г., устный доклад. Maps of chromatin conformation in the cerebral cortex. Cherkasov A., Bazarevich M., Pletenev I., Efimova O., Kononkova A., Khrameeva E.
 27. Всероссийская конференция с международным участием "Московская конференция по вычислительной молекулярной биологии (МС-СМВ'23)". Москва, 2023 г., устный доклад. Архитектура хроматина и опосредованные ей регуляторные механизмы в различных организмах и модельных системах. Е. Храмеева.

Личный вклад. В методических работах [Kobets и др., 2023; Smirnov и др., 2021; Rozenwald и др., 2020] автор диссертации предложила дизайн исследования и принимала прямое участие в разработке алгоритма, его тестировании, интерпретации результатов и документировании программной реализации. В работе [Pervouchine и др., 2012] автор диссертации принимала непосредственное участие в разработке алгоритма, написании программного кода, его тестировании и интерпретации результатов. В работах [Khrameeva и др., 2020; Kurochkin и др., 2019; Ulianov и др., 2019; Khrameeva и др., 2018; Bozek и др., 2017; Zhang и др., 2016; Ulianov и др., 2016; Khrameeva и др., 2016; Zhenilo и др., 2015; Yang и др., 2014; Khrameeva и др., 2014; Rossbach и др., 2014; Rösel-Hillgärtner и др., 2013; Khrameeva и Gelfand, 2012; Khrameeva и др., 2012] автором диссертации лично выполнен биоинформатический анализ экспериментальных данных, статистический анализ и интерпретация результатов анализа. В работах [Samborskaia и др., 2020; Eremenko и др., 2021; Стекольщикова и др., 2021; Akkuratov и др., 2018] автором диссертации поставлена задача, координирован процесс исследований и выполнена интерпретация полученных результатов. В работах [Smirnov и др., 2023; Shcherbinina и др., 2022; Ilyin и др., 2022; Tkachev и др., 2021; Stein и др. 2021; Yegodayev и др., 2020] автор диссер-

тации разработала общую схему анализа данных и принимала прямое участие в разработке и применении методик анализа омиксных данных различных типов. В работах [Stepanova и др., 2021; Mazor и др., 2021; Badarni и др., 2021; Ulianov и др., 2021] автор диссертации принимала непосредственное участие в разработке стратегии анализа данных, координации процесса исследований и интерпретации результатов, а также участвовала в выполнении биоинформатических частей исследований [Tkachev и др., 2019; Luzhin и др., 2019; Ulianov и др., 2017; Gavrilov и др., 2016; Flegontov и др., 2016; Kaufmann и др., 2015]. Таким образом, автор принимала непосредственное участие в выборе стратегий исследований, обсуждении экспериментальной работы, планировании анализа данных и его непосредственном выполнении, интерпретации получаемых результатов, их обобщении, написании и редактировании текстов публикаций. В 9 статьях по теме диссертации автор выступает в качестве первого автора, и в 9 в качестве автора, ответственного за переписку. Все основные результаты анализа данных, описанные в диссертационной работе, получены автором лично или под ее непосредственным руководством.

Работа по теме диссертации была поддержана грантами РНФ (21-74-10102, 19-74-00112) и РФФИ (21-34-70051, 20-34-70077), в которых автор являлась руководителем. Под руководством автора диссертации в рамках темы данной работы подготовлены и защищены 19 выпускных квалификационных работ специалистов и магистров. Работа по теме диссертации проводилась в сотрудничестве с Институтом проблем передачи информации РАН, Институтом биологии гена РАН, Институтом молекулярной генетики РАН, Институтом молекулярной и клеточной биологии Сибирского отделения РАН, Новосибирским государственным университетом, Биологическим факультетом, Физическим факультетом и Факультетом биоинженерии и биоинформатики МГУ имени М.В. Ломоносова, Национальным исследовательским центром "Курчатовский институт Центром "Биоинженерия" РАН, Университетом Гиссена (Германия), Институтом исследований сердца и легких Макса Планка (Германия), Массачусетским технологическим институтом (США), Университетом Людвиг-Максимилиана (Германия), Калифорнийским университетом (США), Университетом Остравы (Чехия), Детской больницей Лос-Анджелеса (США), Институтом психиатрии Макса Планка (Германия), Институтом эволюционной антропологии Макса Планка (Германия), Институтом биологии старения Макса Планка (Германия), Партнер-

ским институтом вычислительной биологии (Китай), Университетом Бен-Гуриона в Негеве (Израиль).

Публикации. Основные результаты по теме диссертации изложены в 39 статьях в рецензируемых журналах, индексируемых Scopus и/или Web of Science.

Объем и структура работы. Диссертация состоит из введения, трех глав и заключения. Полный объем диссертации составляет 209 страниц с 46 рисунками и 3 таблицами. Список литературы содержит 232 наименования.

Глава 1. Обзор литературы

Исследования структуры хроматина привели в последние годы к выдающимся открытиям, значительно расширившим границы всестороннего понимания трехмерной архитектуры генома. Хроматин может изучаться на нескольких организационных уровнях, начиная с полных хромосом и заканчивая нуклеосомами. Особенности организации хроматина могут регулировать основные клеточные процессы, такие как транскрипция, репликация и репарация ДНК. Многочисленные исследования показывают, что структура хроматина вносит вклад в регуляцию транскрипции [15–17]. Это объясняется самим механизмом процесса транскрипции: чтобы необходимые ферменты связывали транскрибируемую область, она должна быть доступна [18]. Показано, что открытость хроматина влияет на другой основной клеточный процесс, репликацию ДНК [19–21]. Механистическая модель, основанная исключительно на структуре хроматина, способна точно предсказать время репликации (коэффициент корреляции Пирсона = 0,92), подтверждая ранее обсуждавшиеся гипотезы об их взаимосвязи [22]. Регуляция репарации ДНК также может быть опосредована конформацией ДНК [23]. Более того, структура хроматина влияет на некоторые клеточные аномалии, например, хромосомные транслокации [24].

Изменения конформации хроматина могут вызывать различные заболевания человека. Например, синдром Корнелии де Ланге - это генетическое заболевание, проявляющееся множественными системными поражениями, когнитивными нарушениями и специфическим фенотипом лица. Доказано, что синдром Корнелии де Ланге связан с разрушением комплекса когезина, одного из белков, ремоделирующих хроматин [25]. Другим примером является α -талассемия/умственная отсталость, X-сцепленный синдром, вызываемый мутациями гена ATRX. Определенные функции его продукта, белка ATRX, еще полностью не изучены, но показано, что белок ATRX связывается с повторами ДНК, и его отсутствие у пациентов с ATRX вызывает значительно более низкий уровень метилирования генов, расположенных в повторяющихся областях ДНК [26]. Связь между белком ATRX и процессом метилирования все еще не определена, и требует дальнейших исследований. Помимо генетических заболеваний, структура хроматина вовлечена в развитие нейродегенеративных и

психических расстройств. Болезнь Хантингтона, давно известное нейродегенеративное заболевание, вызывается сложным взаимодействием негенетических (в основном связанных с возрастом) и генетических факторов, таких как мутации в гене хантингтина (Htt) с повышенным количеством тринуклеотидных повторов CAG. Недавние исследования показывают, что эти мутации ускоряют возрастные эпигенетические нарушения, изменяют структуру хроматина и вызывают специфичные для болезни Хантингтона изменения транскрипции [27]. Другим примером является обсессивно-компульсивное расстройство, этиология которого все еще не ясна. Недавние исследования продемонстрировали увеличение частоты мутаций в петлях хроматина, в которых участвует промотор [28]. Кроме того, было обнаружено увеличение модификаций хроматина в мозге человека, что позволяет предположить вероятную связь патофизиологического механизма обсессивно-компульсивного расстройства с архитектурой хроматина. В одной работе указано, что даже патогенез наркомании и депрессии опосредуется изменениями в конформации хроматина, что подразумевает возможность создания новых терапевтических мишеней [29]. Все вышеупомянутые работы подчеркивают важность исследований структуры хроматина и их потенциальное биомедицинское применение.

Таким образом, структура хроматина имеет первостепенное значение для регуляции активности генов, поэтому лучшее понимание принципов ее организации и функционирования является важнейшей задачей современной молекулярной биологии. Методы изучения конформации хроматина можно условно разделить на две группы: микроскопические методы и методы ЗС. Для получения полной картины трехмерной организации хроматина используются полногеномные варианты методов ЗС, такие как Hi-C, результатом которого является матрица частот попарных контактов "все со всеми". Последующий анализ результатов Hi-C позволяет идентифицировать отдельные элементы структурной организации хроматина, таких как хромосомные территории, компартменты, ТАДы и петли.

1.1 Общие принципы организации хроматина

1.1.1 Хромосомные территории

Хромосомные территории описывают расположение отдельных хромосом внутри интерфазного ядра, при котором каждая хромосома занимает определенную область [2]. Хромосомные территории первоначально были обнаружены с помощью метода FISH [30], а затем подтверждены методом Hi-C и расчетами частот цис/транс взаимодействий [3]. Как правило, богатые генами хромосомы меньшего размера имеют тенденцию располагаться ближе к центру ядра, в то время как бедные генами хромосомы большего размера часто обнаруживаются на периферии ядра [31]. Хромосомные территории, по-видимому, не являются видоспецифичными, но имеют свои особенности в разных типах клеток и тканей, с эволюционно сохраняемыми свойствами [32]. Радиальное расположение хромосомных территорий связано с их функциональной компартментализацией, которая, в свою очередь, влияет на экспрессию генов, сроки репликации и пути репарации ДНК [33], при этом хромосомные территории гомологичных хромосом, как правило, находятся далеко друг от друга в интерфазе [34].

Расположение хромосомных территорий играет критическую роль в нормальном функционировании клеток. Было показано, что хромосомные территории вовлечены в патогенез различных болезней человека, таких как онкологические заболевания и нейродегенеративные расстройства. Нарушения в организации хромосомных территорий могут привести к изменениям профилей экспрессии генов, что в свою очередь вызывает активацию онкогенов и подавление генов-супрессоров опухоли в раковых клетках [32]. Аберрантная организация хромосомных территорий способствует нестабильности генома за счет образования хромосомных транслокаций и нарушения регуляции механизмов репарации ДНК [35].

1.1.2 Компартменты хроматина

Компартменты представляют собой более мелкий уровень организации хроматина и обеспечивают пространственное разделение транскрипционно активных и неактивных областей генома. Компартменты были впервые идентифицированы в работе, представляющей метод Hi-C, на картах с грубым разрешением 1 млн п.о. [3], которые имели “клетчатый” рисунок на уровне отдельных хромосом. Это наблюдение привело к предложению о пространственном разделении двух наборов геномных участков. Дальнейшие исследования позволили лучше понять устройство этих наборов, показав их обогащение активными (эухроматин) или неактивными (гетерохроматин) состояниями хроматина, в том числе и в нашей работе [36]. Эти два сегрегированных набора с преимущественно активным или неактивным хроматином были названы компартментами А и В, соответственно [3]. Области компартмента А, как правило, имеют более высокую вероятность контакта с другими областями из компартмента А и более низкую вероятность контакта с областями из компартмента В, и наоборот.

В настоящее время концепция компартментов А и В, характеризующихся различной транскрипционной активностью и эпигенетическими метками, является общепринятой и широко распространенной. Исследования показали, что компартменты А и В связаны с различным содержанием GC нуклеотидов, гиперчувствительностью к ДНКазе I, экспрессией генов, временем репликации, плотностью генов и метками хроматина [37]. Компартмент А расположен ближе к центру ядра и в основном состоит из эухроматина, который богат генами, активен в транскрипционном отношении и обогащен метками, ассоциированными с открытым хроматином, такими как H3K4me3 и H3K36me3. И наоборот, компартмент В преимущественно состоит из гетерохроматина, который беден генами, транскрипционно неактивен и имеет репрессивные эпигенетические метки, такие как H3K9me3 и H3K27me3 [38]. Анализ данных Hi-C высокого разрешения с помощью алгоритмов машинного обучения показал возможность разделения компартментов А и В на шесть суб-компартментов: два для компартмента А и четыре для компартмента В [39; 40].

Дальнейшие эксперименты Hi-C показали, что компартментализация генома зависит от типа клеток. Расположение компартментов не является жестко

фиксированным и подвергается ремоделированию во время дифференцировки клеток [8]. В частности, образование компартмента А может быть обусловлено взаимодействиями между факторами транскрипции, коактиваторами, ремоделерами хроматина и активными модификациями гистонов [41]. Напротив, компартмент В, который в значительной степени состоит из гетерохроматина, может быть сформирован путем привлечения репрессивных модификаторов хроматина и образования компактных хроматиновых петель, которые ограничивают доступ факторов, запускающих транскрипцию [4].

1.1.3 Топологически ассоциированные домены (ТАДы)

Достижения в области высокопроизводительного секвенирования позволили добиться высокого разрешения карт Hi-C и исследовать более тонкие уровни структуризации хромосом. В частности, анализ карт Hi-C с разрешением в десятки тысяч п.о. показал наличие взаимодействующих внутри себя участков генома размером около 1 млн п.о., которые были названы топологически ассоциированными доменами (ТАДами). ТАД определяется как область генома, размер которой может варьироваться от сотен тысяч п.о. до нескольких миллионов п.о. с высоким уровнем взаимодействий внутри этой конкретной области и низким уровнем взаимодействий за ее пределами. Подобные домены определяются даже у бактерий и у других видов: человек [5], мышь [42], дрозофила [43], растения [44], грибы [45] и другие.

ТАДы были открыты с помощью метода Hi-C более десяти лет назад, и с тех пор их функции, структура и вариабельность быстро стали одной из основных тем анализа архитектуры хроматина [42]. Известно, что ограничение взаимодействий энхансер-промотор структурой ТАДов влияет на транскрипцию генов [46]. Таким образом, показано, что изменения границ ТАДов являются связующим звеном в патофизиологических механизмах различных заболеваний, таких как пороки развития конечностей [47], энцефалопатии [48] и повышенный риск некоторых видов рака: рак легких [49], острый лимфобластный лейкоз [50] и глиома [51]. Границы ТАДов обогащены эпигенетическими метками активной транскрипции (триметилирование по 4-му и 36-му остаткам

лизина гистона H3), генами домашнего хозяйства, сайтами связывания белков-инсуляторов, а именно CCCTC-связывающего фактора (CTCF) и короткими ядерными элементами (SINEs) [5; 52]. Более того, совместные исследования петель хроматина и ТАДов показали значительно более высокую частоту петель на границах ТАДов [40]. Авторы называют эти ТАДы петлевыми доменами и отмечают ассоциацию петлевых якорей с субъединицами когезина RAD21 и SMC3, а также с CTCF [40].

Предполагается, что ТАДы являются основными функциональными регуляторными доменами, которые модулируют контакты между энхансерами и промоторами. Повышенная частота контактов внутри ТАДов опосредует физические взаимодействия между парами энхансер-промотор, в то время как высокая изоляция на границах ТАДов ограничивает такие взаимодействия для пар, расположенных в соседних ТАДах [53]. Таким образом, ТАДы, по-видимому, играют критическую роль в формировании специфических взаимодействий энхансер-промотор и, следовательно, регуляции экспрессии генов. В соответствии с этим предположением, несколько исследований показали, что нарушения организации ТАДов способствуют неправильному образованию энхансер-промоторных взаимодействий. Действительно, было показано, что именно необычная регуляторная архитектура хроматина, сформированная неправильными парами энхансер-промотор, приводит к аберрантной экспрессии генов и вызывает патогенные фенотипы [47; 51]. Таким образом, ТАДы являются ключевыми элементами пространственной организации хроматина, которые поддерживают надлежащую регуляцию экспрессии генов.

Исследования на клетках млекопитающих показали, что ТАДы организованы иерархически, с возможным включением нескольких мелких суб-ТАДов внутри более крупного ТАДа [54]. Суб-ТАДы могут представлять более тонкий уровень организации хроматина и могут быть вовлечены в более точную регуляцию экспрессии генов и модуляцию дальних взаимодействий в хроматине. В то время как ТАДы в целом мало меняются между различными типами клеток млекопитающих и даже между разными видами [5], было показано, что суб-ТАДы более динамичны [55]. В целом, иерархическая организация ТАДов в контексте других уровней организации хроматина имеет сложную природу. Идея, предложенная в ранних исследованиях и рассматривающая ТАДы как строительные блоки компартментов, не нашла существенных подтверждений.

Сейчас считается, что ТАДы скорее всего образуются по независимому от компартиментализации механизму, в соответствии с моделью петлевой экструзии. В мировой литературе растет число свидетельств того, что хроматин представляет собой сложное динамическое взаимодействие механизмов фазовой сегрегации и петлевой экструзии [56]. Однако всё ещё требуются дополнительные исследования, чтобы поставить точку в этом вопросе окончательно.

1.1.4 Хроматиновые петли

Хроматиновые петли представляют собой тончайший масштаб организации хроматина и служат платформой для пространственных взаимодействий между регуляторными элементами, такими как энхансеры и промоторы, которые модулируют экспрессию генов. Образование петель в основном происходит, когда две области, находящиеся на относительно большом геномном расстоянии, сближаются в трехмерном пространстве, чтобы повлиять на транскрипцию генов. Многие исследования продемонстрировали важность хроматиновых петель: примерно половина генов человека вовлечена в дальние взаимодействия с хроматином посредством образования хроматиновых петель [52]. Таким образом, формирование петель хроматина приводит дистальные цис-регуляторные элементы в непосредственную близость к их генам-мишеням и способствует точному пространственно-временному контролю экспрессии генов.

Энхансеры представляют собой тип цис-регуляторных элементов, которые могут взаимодействовать с промоторами генов, модулирующими экспрессию соответствующего гена. Исследования показали, что энхансерные элементы могут быть размещены в различных положениях относительно генов-мишеней. В частности, в некоторых случаях эти регуляторные элементы были обнаружены перед геном, после гена или в интроне другого гена, отделенного сотнями тысяч п.о. от мишени [57]. Примечательно, что взаимодействия энхансер-промотор не являются однозначными. В нескольких работах выдвигалось предположение, что различные энхансеры могут быть соединены с одним промотором одновременно [58; 59]. Более того, недавняя работа показала, что преобладающее число промоторов в нейронах и глиальных клетках взаимодействует с несколькими эн-

хансерами [60—62]. В той же работе также была проиллюстрирована линейная корреляция между количеством физически взаимодействующих энхансеров с выбранным промотором и увеличением экспрессии соответствующего гена [61].

Было показано, что клеточно-специфические энхансер-промоторные петли способствуют клеточно-специфической экспрессии генов, отвечающих за дифференцировку клеток [63]. Кроме того, изменения в структурах петель хроматина или нарушение границ ТАДов могут привести к аберрантной экспрессии генов и нестабильности генома. Например, петли хроматина вовлечены в развитие различных видов рака, где перестройка энхансер-промоторных взаимодействий может способствовать активации онкогена или подавлению гена-супрессора опухоли [64]. В совокупности, петли хроматина являются важнейшими элементами поддержания нормальных пространственных взаимодействий между регуляторными элементами и генами-мишенями.

В недавней работе был применен новый подход к профилированию взаимодействий хроматина в масштабах всего генома с высоким разрешением для идентификации петель и определения белков, совместно расположенных в их якорях [63]. Было показано, что энхансер-промоторные петли предпочтительно совместно связываются с ZNF143 и HCFC1, а не с CTCF и RAD21 в лимфобластоидных клетках. Более того, были идентифицированы взаимодействия промотор-промотор, которые, как было установлено, способствуют регуляции экспрессии генов [63]. Эти наблюдения подчеркивают наличие различных типов взаимодействий, приводящих к образованию петель, опосредованных различными белковыми комплексами.

1.1.5 Механизм образования хроматиновых петель и ТАДов

Хотя детали механизма образования петель еще не ясны и явно отличаются для разных видов, уже найдены ключевые факторы этого процесса, такие как когезин и CTCF. Известной теорией, объясняющей механизм формирования петлевых структур, является петлевая экструзия [65]. Экструзионный комплекс, также называемый комплексом структурного поддержания хромосом (SMC), представляет собой кольцеобразный белок, состоящий из двух физи-

чески взаимодействующих субъединиц, связывающих ДНК, образуя круг [66]. Загрузка комплекса на хроматин происходит в случайном геномном локусе. Небольшая петля хроматина образуется путем связывания субъединиц SMC с соседними участками ДНК. Затем две связанные субъединицы движутся в противоположных направлениях относительно концов ДНК, следовательно, происходит выдавливание петли. В частности, эта модель упоминается как модель двусторонней экструзии [67]. Экструзия прекращается, когда обе субъединицы достигают блокирующих сайтов, таких как мотивы CTCF в конвергентной ориентации [68].

Хотя указанная модель не может объяснить все петлевые взаимодействия, эксперименты Hi-C показали ее соответствие наблюдаемым петлям [6]. Другим преимуществом модели экструзии петель как основного механизма их образования петли являются ее следствия: формирование паттернов петлевых доменов, распределение пространственных расстояний по отношению к геномному расстоянию, частота петель между конвергентными сайтами CTCF (“правило конвергенции”) и отсутствие перекрывания петель [6]. В этой работе исследовали последствия изменений в сайтах связывания CTCF с помощью редактирования генома: правило конвергенции смогло предсказать каждый случай изменения образования петли из-за геномного редактирования.

Недавно было обнаружено, что значительная часть комплексов SMC в клетках человека вытесняет ДНК только с одной стороны, в то время как другая сторона все еще остается на месте загрузки комплекса [69]. Односторонняя экструзия не может быть встроена в существующий механизм модели петлевой экструзии из-за предположения о двусторонней экструзии [65]. Более того, доказано, что в случае экструзии петли только с одной стороны эффективность линейного уплотнения ДНК значительно ниже, что было подтверждено как теоретически, так и экспериментально [67]. Следовательно, для более глубокого понимания модели формирования петель необходимо отойти от устоявшейся теоретической точки зрения.

Модель петлевой экструзии применима не только для объяснения формирования хроматиновых петель, но и ТАДов. Модель экструзии объясняет, например, слияние двух соседних ТАДов после удаления границы ТАДа, конвергентной ориентации мотивов CTCF на границе ТАДа и пики частоты контактов в углу ТАДа [65]. Экструзия петель также может быть проанализирована в

контексте взаимодействий между формированием ТАДов и компартиментализацией генома. Недавние исследования указывают на взаимосвязь между экструзией петель и компартиментализацией [56]. Деpletion экструзионных комплексов уменьшила количество ТАДов и усилила А/В-компарменты, в то время как depletion белка WAPL, который является фактором разгрузки когезина, увеличила силу крупных ТАДов и уменьшила компартиментализацию. Деpletion другого фактора экструзионной модели CTCF уменьшила количество ТАДов, в то время как разделение на компарменты осталось неизменным.

Образование петель происходит у разных видов с различными характерными размерами петель и задействованными белковыми факторами. Например, в клетках млекопитающих и дрожжей основным белком комплекса SMC является когезин. Когезин опосредует образование петель, закрепленных CTCF: в клетках млекопитающих его depletion приводит к отсутствию петель, закрепленных CTCF [70]. Другой вышеупомянутый белок WAPL действует как фактор диссоциации когезина [71]. Недавно было показано, что CTCF действует как фактор поддержания цикла посредством предотвращения связывания WAPL [72]. Другим участвующим белком в клетках млекопитающих является PDS5-ассоциированный с когезином фактор (PDS5), который индуцирует способность WAPL удалять когезин, хотя WAPL и PDS5 по-разному влияют на образование петель [73]. Существуют и другие транскрипционные факторы (ТФ), обогащенные в якорях петель, закрепленных CTCF, но их функции в механизме экструзии петель или поддержании структуры петель все еще не ясны [40]. Хотя большинство клеточных петель млекопитающих закреплены CTCF, геномные взаимодействия на больших расстояниях могут быть опосредованы другими белками, такими как LDB1, ZNF143 и YY1 [74]. Указанные белки устанавливают энхансер-промоторные взаимодействия непосредственно путем связывания ДНК и приближения к определенным сайтам генома (ZNF143 и YY1) или косвенно путем связывания ТФ (LDB1) [75–77].

1.1.6 Другие уровни организации хроматина

Помимо главных четырех уровней организации хроматина, описанных выше, в текущих исследованиях рассматриваются дополнительные элементы организации хроматина. В некоторых исследованиях определяют часто взаимодействующие области (FIREs), которые демонстрируют повышенные уровни локальных взаимодействий хроматина. Предполагается, что FIREs отличаются от ТАДов или хроматиновых петель другим характерным расположением и предполагаемыми функциями. Идентификация местоположения FIREs относительно других элементов геномной архитектуры показала их обогащение в компартменте А и между ТАДами и обеднение в компартменте В и на границах ТАДов [78]. Считается, что FIREs являются областями, активно участвующими в регуляции генов. Действительно, они обычно включают в себя несколько взаимодействующих партнеров, включающих тканеспецифичные гены, активные энхансеры и супер-энхансеры. Установлено, что FIREs специфичны для тканей и типов клеток, однако в основном сохраняются у разных видов [78]. В целом, FIREs играют значительную роль в регуляции экспрессии генов и создании тканеспецифичных программ регуляции транскрипции, возможно, посредством совместных взаимодействий, которые сближают несколько цис-регуляторных элементов внутри ТАДов. FIREs также играют важную роль в нейрональных клетках. В частности, было обнаружено, что тканеспецифические функции, такие как нейрогенез, поддерживаются благодаря правильной структуре FIREs в коре головного мозга плода [79]. Таким образом, важность недавно выявленных FIREs подчеркивает актуальность более глубокого анализа различных элементов организации хроматина, особенно на мелких масштабах.

Другим важным уровнем организации хроматина являются взаимодействия, опосредованные белками группы поликомба (PcG). Белки PcG организованы в два эпигенетических комплекса: репрессивные комплексы поликомба 1 и 2 (PRC1 и PRC2). PRC2 катализирует триметилирование остатка лизина на гистоне H3 (H3K27me3). Он способствует уплотнению хроматина и создает стабильную и наследуемую метку, которая может передаваться через поколения клеток [80]. PRC1 отвечает за поддержание неактивного состояния генов путем распознавания и связывания с модификацией H3K27me3, продуцируемой PRC2

[81]. Было обнаружено, что PRC1 и PRC2 вовлечены в дальние взаимодействия между участками генома, которые обогащены H3K27me3 и расположены до сотен миллионов п.о. друг от друга [81–85]. В частности, обнаружено, что белки PcG образуют компартментные домены, обогащенные H3K27me3, не связанные с когезином или петлевой экструзией, на поздней стадии развития ооцитов мыши [82].

Помимо возможного образования независимых блоков, некоторые из взаимодействий, опосредованных PcG, могут быть обнаружены при взаимодействии с когезином. Так, недавнее исследование эмбриональных стволовых клеток мыши с нокаутом когезина показало, что ряд дальних высокочастотных взаимодействий усиливается в отсутствие когезина. Обнаружено, что эти взаимодействующие области соответствуют областям обогащения PRC1 и PRC2 [86]. Контакты, опосредованные PcG, значительно более выражены, чем контакты, опосредованные CTCF, что является еще одним доказательством независимости их механизма формирования от ТАДов [81]. Такие дальние PcG взаимодействия в первую очередь описываются в контексте эмбрионального развития и поддержания плюрипотентности, опосредованных связыванием и последующим подавлением определенных генов [83; 84]. Наконец, установлено, что динамическая реорганизация дальних PcG взаимодействий важна при дифференцировке нейронов у мышей [85].

1.2 Современное состояние исследований об архитектуре хроматина в мозге человека

Не так много работ исследовали организацию хроматина в мозге человека, и в частности разницу в архитектуре генома между нейронами и различными глиальными клетками [11; 12]. Однако они заложили основу исследований хроматина в головном мозге, продемонстрировав, что, хотя клетки мозга в целом соответствуют фундаментальным принципам трехмерной организации генома, они также демонстрируют уникальные особенности архитектуры хроматина на нескольких структурных уровнях.

1.2.1 Компарментализация в нейронах и глиальных клетках

При рассмотрении архитектуры хроматина на крупном масштабе, в нейронах обнаруживается более слабая компарментализация по сравнению с совокупностью остальных клеток мозга [11]. Разница в выраженности компарментов, вероятно, не объясняется гетерогенностью глиальных клеток, поскольку микроглия, специфический тип глиальных клеток, демонстрирует ту же тенденцию. Кроме того, уникальные особенности нейронов очевидны во внутренней организации компарментов А и В. В частности, дальние взаимодействия между компарментами типа В в нейронах обеднены по сравнению с глией, тогда как нейроны демонстрируют повышенные уровни ближних взаимодействий между компарментами типа А.

Наблюдаемые различия в особенностях компарментализации могут быть связаны с усилением экструзии петель в нейронах. Недавние исследования показали влияние комплексов когезина, которые играют ключевую роль в модели петлевой экструзии, наряду с CTCF в регулировании взаимодействий А/В-компарментов как на коротких, так и на дальних расстояниях. Например, деплеция определенного когезинового комплекса приводит к увеличению дальних взаимодействий компарментов В при существенном снижении взаимодействий на среднем расстоянии компарментов А [87]. Эти результаты устанавливают связь между экструзией петель и механизмами взаимодействия компарментов В на дальних расстояниях, предлагая возможное объяснение наблюдаемого снижения взаимодействий компарментов В в нейронах [11].

В дополнение к различиям в силе контактов внутри компарментов, наблюдалось значительное переключение компарментов между нейрональными и глиальными клетками. Более 2000 областей переключаются из компармента В в А между нейрональными и глиальными клетками, в то время как в два раза больше областей переключаются из компармента А в В [12]. Авторы исследования также выявили обогащение генов в измененных областях. Их результаты демонстрируют, что области, переключающиеся с компармента В на А между нейрональными и глиальными клетками, демонстрируют более высокую экспрессию в глиальных клетках, таких как олигодендроциты и астроциты, тогда как области, переключающиеся в противоположном направлении,

выше экспрессированы в нейронах [12]. Таким образом, эти наблюдения подчеркивают наличие специфических для типа клетки особенностей организации хроматина и указывают на тесную взаимосвязь между этими различиями и регуляцией экспрессии генов.

1.2.2 ТАДы и петли в нейронах и глиальных клетках

На более мелких масштабах организации хроматина также обнаруживаются специфические для клеточных типов структурные особенности. Появляются данные, свидетельствующие об иной природе образования ТАДов в нейронах по сравнению с глиальными клетками. В частности, сформулировано предположение о том, что ТАДы в глии могут формироваться главным образом посредством поликомб-опосредованных контактов, в то время как формирование ТАДов в нейронах согласуется с моделью петлевой экструзии [11]. Авторы показали, что границы глиально-специфичных ТАДов обогащены гистоновой меткой H3K27me3. Другие исследования показали, что метка H3K27me3 в развивающихся ооцитах важна для формирования независимых от когезина компартментальных доменов, которые, как полагают, поддерживаются за счет разделения фаз [82; 88]. Таким образом, увеличение плотности H3K27me3 в глиально-специфичных ТАДах может свидетельствовать о том, что фазовое разделение неактивного и активного хроматина может влиять на формирование этих доменов. Напротив, нейронно-специфичные ТАДы демонстрируют значительное обогащение активных гистоновых меток, в частности H3K4me3 и H3K27ac, на их границах. Кроме того, границы нейронных ТАДов более выражены по сравнению с глиальными ТАДами [11]. Эти данные позволяют предположить, что формирование нейронно-специфичных ТАДов может в основном происходить по механизму экструзии петель, поскольку H3K4me3 и H3K27ac связаны с сайтами старта транскрипции (TSS), где происходит рекрутирование когезина [89]. Предполагается, что основным стимулом для экструзии петли вокруг TSS может быть транскрипционная активность, которая связана с перемещением когезина [89]. Таким образом, эти результаты свидетельствуют о том, что различные механизмы могут влиять на формирование ТАДов в нейронах и

глиальных клетках, что может частично объяснить наблюдаемые различия в свойствах ТАДов этих двух типов клеток.

Гипотеза о более выраженном механизме экстррузии петель в нейронах подтверждается несколькими отличительными характеристиками образования петель в клетках этого типа. Примечательно, что энхансер-промоторные петли часто перекрываются с границами ТАДов, специфичными для нейронов, и промоторы расположены рядом с этими границами, а энхансеры расположены после промотора [11]. Основываясь на специфичном для нейронов механизме формирования ТАДов, описанном выше, эти наблюдения указывают на то, что экстррузия петель облегчает установление взаимодействий в энхансер-промоторных петлях. Согласно этой идее, предложенная модель предполагает, что промоторы сканируют локальный хроматин посредством петлевой экстррузии для установления взаимодействий с энхансерами. Эта концепция подтверждается появляющимися новыми данными, связывающими CTCF и промотор-энхансерные взаимодействия. В частности, недавнее исследование продемонстрировало, что предпочтительный энхансер для промотора отчасти определяется уровнем связывания CTCF в промоторной области [61]. Как правило, энхансер-промоторные взаимодействия регулируют экспрессию генов, специфичных для типа клетки [90]. Исследования показывают, что энхансер-промоторные взаимодействия, специфичные для нейронов или глии, обогащены генами, специфичными для определенного типа клеток. Так, специфичные для типа клетки энхансер-промоторные взаимодействия, а также дифференциальные пики H3K27ac обогащены синаптическими и аксонными генами в нейронах и актиновыми генами подвижности в глиальных клетках [61]. Следовательно, энхансер-промоторные взаимодействия демонстрируют взаимосвязь архитектуры генома и транскрипции, что отражает уникальный механизм регуляции генов, специфичных для типа клетки.

1.2.3 Особые уровни организации хроматина в нейронах и глиальных клетках

FIREs представляют собой особый уровень архитектуры хроматина, который тесно связан с клеточно-специфичной регуляцией генов. Исследования показали, что эти области обогащены дифференциальными пиками H3K27ac как в нейронах, так и в глиальных клетках и связаны с клеточно-специфическими энхансерами [61; 78]. Примечательно, что обогащение генов, участвующих в миелинизации, глиальной дифференцировке и дифференцировке олигодендроцитов, было обнаружено в глиально-специфичных FIREs, в то время как гены, участвующие в миелинизации, показали обогащение в нейронно-специфичных FIREs [61]. Эти результаты наряду со другими ключевыми наблюдениями о клеточно-специфических особенностях хроматина, перечисленными выше, указывают на тесную связь организации генома и регуляции экспрессии генов в клетках головного мозга.

Контакты, опосредованные белками группы поликомба (PcG), представляют другой особый иерархический уровень трехмерной организации генома, связанный с формированием цис-взаимодействий на очень далеких расстояниях и даже транс-взаимодействий. Поскольку роль белков PcG была в основном исследована в контексте развития, их участие в поддержании организации хроматина в зрелых клетках только начинает проясняться. В то время как общая функция взаимодействий, опосредованных PcG, остается неясной, установлено, что их динамическая реорганизация важна при дифференцировке нейронов у мышей [85]. Ни в одной из предыдущих работ не рассматривались дальние взаимодействия, опосредованные PcG, как отличительная особенность трехмерного генома зрелых нейронов или глиальных клеток.

1.2.4 Регуляторная роль организации хроматина в мозге человека

Регуляция работы генов в мозге человека изучена недостаточно, в особенности эволюция регуляции экспрессии генов. Лишь пять исследований посвя-

щены регуляторным механизмам, определяющим специфические для человека изменения в экспрессии генов. Анализ факторов транскрипции (ТФ) выявил предсказанные регуляторные сети ТФ и отдельные ТФ, определяющие различия в экспрессии генов между мозгом человека и шимпанзе [91; 92]. В результате анализа эпигенетических модификаций были выявлены промоторы и энхансеры, потенциально участвующие в кортикогенезе человека [93], а также активные специфичные для человека цис-регуляторные элементы в мозге взрослого человека [94]. Наконец, были изучены человек-специфичные особенности пространственной экспрессии и регуляции генов в восьми областях мозга [14]. Хотя не было предпринято никаких дальнейших шагов к пониманию специфической для человека регуляции экспрессии в разных областях мозга по сравнению с другими приматами, стоит отдельно отметить исследование конформации хромосом в мозге человека, которое пролило свет на регуляторные отношения в мозге [95]. Также важно отметить, что была предпринята попытка сравнения трехмерной организации хроматина между разными видами приматов: человек, шимпанзе, бонобо, горилла [96], но не в мозге, а в лимфоцитах.

1.3 Существующие методы изучения архитектуры хроматина

На заре изучения хроматина, оптическая микроскопия была единственным доступным методом. Многие фундаментальные особенности организации хроматина были обнаружены только с помощью оптической микроскопии. Прогресс в области изучения конформации хромосом и регуляции транскрипции стремительно ускоряется благодаря появлению методов секвенирования генома, которые становятся широко доступными. Технологии 3С произвели революцию в области изучения хроматина. Эти технологии помогли обнаружить новые особенности структуры хроматина как у эукариот, так и у прокариот, такие как ТАДы, А/В компартменты [3; 5; 8; 42; 43] и др. На сегодняшний день, чтобы детально изучить архитектуру и функции хроматина, исследователи стремятся использовать оба типа методов - основанные на микроскопии и на секвенировании.

1.3.1 Методы, основанные на микроскопии

Методы, основанные на микроскопии, обычно совмещены с флуоресцентными методами, в результате чего были разработаны такие методы, как FISH [30]. FISH применяется в экспериментах по изучению относительного расположения хромосом, их структуры и специфических ядерных локусов, представляющих интерес. Тем не менее, изучение структуры хроматина методом FISH ограничено различными факторами. Так, число спектрально различимых флуорофоров в стандартном протоколе FISH составляет менее четырех для обычных микроскопов, следовательно, возникает ограничение на максимальное количество отображаемых локусов. Во-вторых, значительный размер (>10 тыс. п.о.) зондов FISH препятствует их массовому синтезу. Перечисленные ограничения устанавливают максимальное предельное число возможных локусов в эксперименте.

Достижения методов, основанных на микроскопии, включают разработку пользовательских олигонуклеотидных матриц (например, Oligopaint), а прогресс в методах микроскопии со сверхразрешением (например, STORM и PALM) значительно изменил масштаб наблюдаемого объекта [97–99]. Метод Oligopaint позволил напрямую маркировать и визуализировать все ТАДы и их границы [97; 100; 101]. Сочетание этой технологии с мультиплексированной последовательной оптической визуализацией позволило наблюдать распределение ТАДов по хромосомным территориям [102]. Еще одним полезным применением методов микроскопии является маркировка отдельных белков, ассоциированных с хроматином, или специфических локусов ДНК, что позволяет отслеживать их динамику на протяжении клеточного цикла [103]. Тем не менее, ограничения методов, основанных на микроскопии, препятствуют комплексному анализу структуры хроматина в целом геноме. Перспективы дальнейшего совершенствования методов микроскопии связаны с исследованиями 3D структуры хроматина для дифференцировки клеток на уровне отдельных клеток, что является нишей, еще не занятой популяционными методами 3С.

1.3.2 Методы 3С

Общая идея, лежащая в основе методов 3С, основана на следующих предположениях: дистальные геномные последовательности, которые сближаются в трехмерном пространстве совместно с белками, могут быть химически сшиты, и в результате получается определить число контактов между анализируемыми локусами. Поскольку использование метода 3С привело к нескольким революционным открытиям, подтвержденным другими методами с момента появления оригинального метода в 2002 году, эти предположения можно считать разумными [104]. Результатом применения методов 3С является количество взаимодействий между геномными локусами в зависимости от их сближенности в трехмерном пространстве для всех геномных расстояний. Эти частоты взаимодействия впоследствии могут быть либо проанализированы непосредственно, либо предварительно преобразованы в пространственные расстояния [40; 105].

Обобщенный протокол всех методов 3С состоит из следующих шагов. Первым шагом является сшивание хроматина формальдегидом, закрепляющее пространственные геномные взаимодействия. Стандартизация параметров процедуры, таких как концентрация раствора формальдегида и продолжительность обработки, необходима для получения воспроизводимого результата и возможности проведения сравнительных исследований [106]. Следующим шагом является разрезание генома на фрагменты путем обработки эндонуклеазой рестрикции. Размер фрагмента зависит от используемых ферментов и определяет результирующее разрешение карты. Затем происходит лигирование по принципу близости: предпочтительное лигирование сшитых фрагментов. В итоге, лигированные фрагменты амплифицируются с помощью методов полимеразной цепной реакции (ПЦР) и получается подсчитать количество контактов.

Группа методов 3С включает в себя 3С, 4С, 5С и Hi-C. Подход 3С ("один против одного") позволяет количественно оценить контакты между одной парой геномных локусов, следовательно, требуется предварительно определить анализируемые локусы. Подход 4С ("один против всех") с помощью дополнительных этапов лигирования, создания кольцевых фрагментов ДНК и последующей обратной ПЦР позволяет оценить частоты взаимодействия между одним конкретным локусом и всеми другими локусами в геноме [107]. Подход 5С ("многие

против многих") количественно оценивает взаимодействия между всеми локусами внутри выбранной интересующей области размером не более 1 Мб, но с относительно низким покрытием [108]. Особенность, примененная в протоколе 5С, заключается в лигировании универсальных праймеров ко всем фрагментам. Наконец, подход Hi-C - это комбинация методов 3С и секвенирования следующего поколения (NGS), которая позволяет количественно оценить взаимодействие “все против всех” с помощью мечения биотинилированными нуклеотидами перед лигированием [3] (рис. 1.1). В результате биотинилирования обеспечивается эффективная очистка и избирательное секвенирование лигированных фрагментов. Полученные пары прочтений выравниваются на геном для определения геномных координат анализируемых фрагментов. В настоящее время оригинальный протокол Hi-C интенсивно модифицируется, причем некоторые варианты сочетаются с технологией иммунопреципитации хроматина (ChIP-seq), используют методы захвата определенных олигонуклеотидов или пытаются адаптировать подходы 3С к анализу одиночных клеток [109].

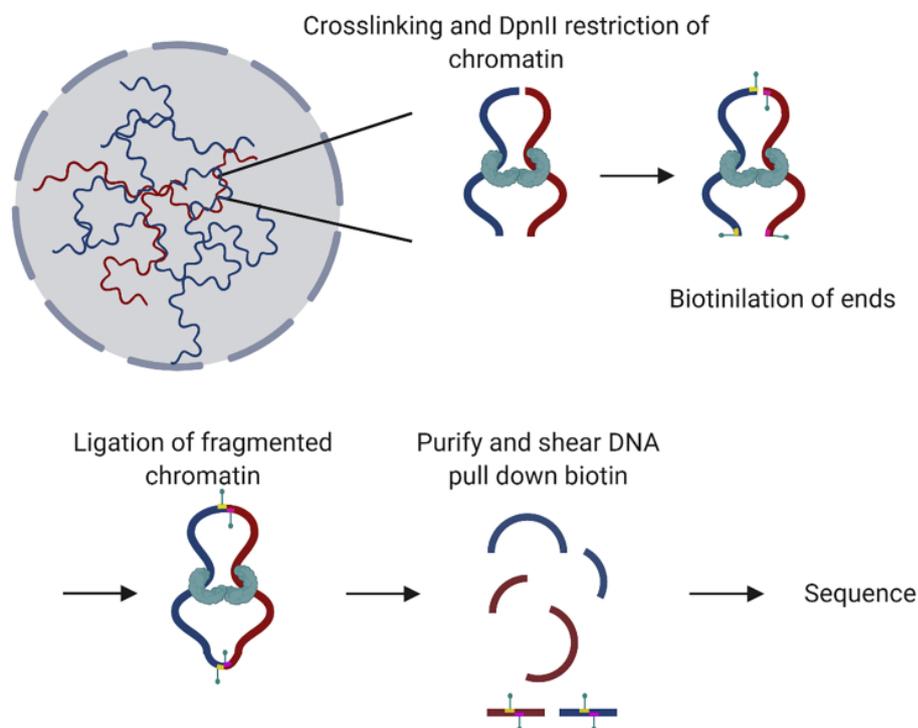


Рисунок 1.1 — Схематическая иллюстрация основных этапов протокола Hi-C. Адаптировано из [110].

1.3.3 Комбинация методов ЗС и микроскопии

В нескольких исследованиях использовались передовые протоколы, объединяющие как методы, основанные на микроскопии, так и методы ЗС. Так, в работе [111] авторы разработали протокол Hi-M - метод микроскопии с высокой пропускной способностью, разрешением и охватом, основанный на протоколе Oligopaint и визуализации множественных локусов ДНК. Hi-M позволяет одновременно измерять экспрессию РНК в одной клетке и анализировать структуру хроматина с разрешением в нанометр и тыс. п.о. у *D. melanogaster*. Авторы отдельно проанализировали организацию ТАДов и уровень экспрессии в масштабе одной клетки. Дальнейшее продвижение в направлении создания таких комбинаторных методов потенциально может привести к крупным открытиям в области изучения хроматина.

1.3.4 Другие омиксные методы

Методы изучения структуры хроматина включают и другие подходы. Так, метод определения доступного хроматина с помощью транспозазы (АТАС-seq, рис. 1.2), использует транспозазу Tn5 для количественной оценки открытости хроматина по всему геному [112]. Разрезание ДНК микрококковой нуклеазой с глубоким секвенированием (MNase-seq) является еще одним методом оценки открытости хроматина [113]. Основным инструментом MNase-seq является полученная из золотистого стафилококка неспецифическая микрококковая эндоэксонуклеаза, которая разрезает фрагменты ДНК, не связанные белками. Следующий метод, секвенирование гиперчувствительных сайтов ДНКазы I (DNase-seq), основан на полногеномном секвенировании фрагментов ДНК, разрезаемых дезоксирибонуклеазой I (ДНКазы I) [114]. ДНКазы I представляет собой эндонуклеазу, обладающую специфичностью к доступным участкам ДНК без нуклеосом, предпочтительно к 5' пиримидинам, которая связывается с этими участками и разрезает их. Исторически сайты расщепления ДНКазой I называ-

лись сайтами гиперчувствительности к ДНКазе I и считались отличительными признаками регуляции хроматина.

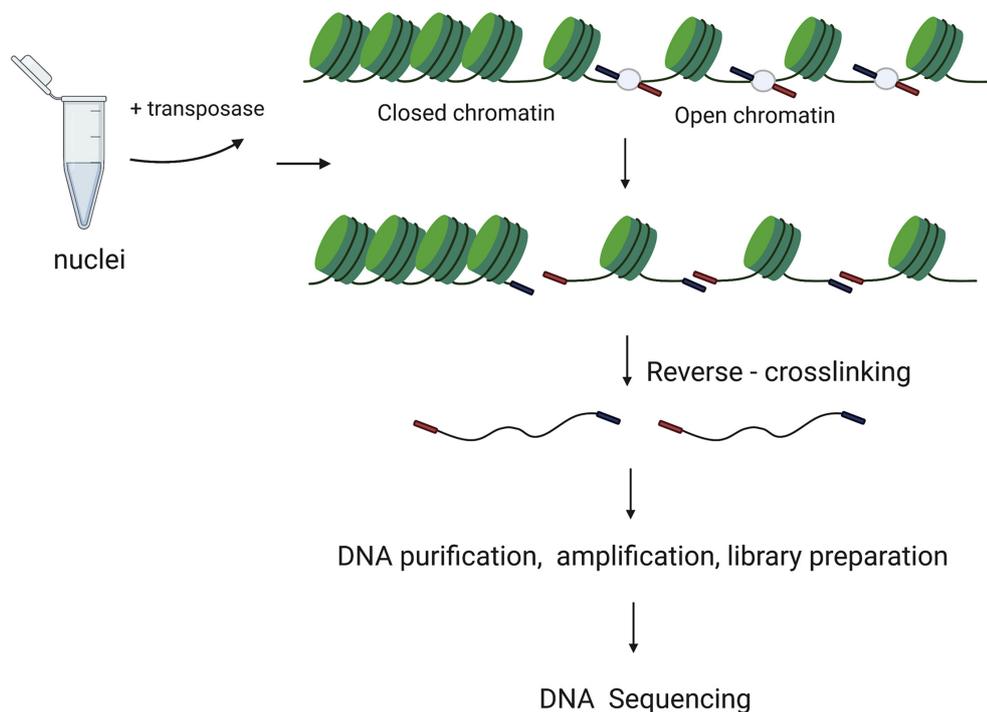


Рисунок 1.2 — Схематическая иллюстрация основных этапов протокола ATAC-seq. Адаптировано из [110].

Перечисленные методы обычно применяются в сочетании с такими методами, как совмещенная с секвенированием иммунопреципитация хроматина (ChIP-seq), используемая для определения позиций связывания ТФ или гистоновых модификаций; оценка уровней экспрессии генов путем секвенирования РНК (RNA-seq), в том числе и с разрешением в одну клетку (scRNA-seq - версия метода для одиночных клеток, или snRNA-seq - версия метода для одиночных ядер); или Hi-C, чтобы комплексно изучить организацию хроматина и его функциональные особенности.

Таким образом, все методы исследования хроматина имеют ряд преимуществ и недостатков, которые необходимо подбирать для решения каждой биологической задачи. Методы, основанные на микроскопии, наиболее подходят для исследований с акцентом на относительное и радиальное расположение областей генома, а также на структуру хроматина на уровне одной клетки и ее изменчивость. Однако эти методы, как правило, ограничены в определении структуры генома в целом. Методы, основанные на ЗС, напротив, подходят для исследований всего генома, хотя их результаты представляют собой наложение

отдельных структур хроматина друг на друга, а не одну стабильную конформацию [4]. Подводя итог, можно заключить, что методы должны быть выбраны исходя из целей исследования.

Глава 2. Материалы и методы

2.1 Стандартные методы обработки данных Hi-C

В своих работах, мы старались придерживаться стандартной методологии для первых шагов анализа данных Hi-C. Так как с момента появления первого набора данных Hi-C прошло более 10 лет, конкретные версии программ, которые мы использовали, а также выбираемые параметры их запуска могли меняться. Однако основные шаги обработки данных Hi-C оставались неизменными, и в наиболее современном виде сводятся к следующим процедурам.

2.1.1 Построение карт контактов на основе данных Hi-C

Сырые прочтения каждой Hi-C библиотеки были картированы на референсный геном с использованием биоинформатического пайплайна `distiller-nf` (<https://github.com/open2c/distiller-nf>, v.0.3.3), который выполняет функции картирования, проводит контроль качества и фильтрацию прочтений секвенатора, разбиение на бины и создание списка контактов, которые преобразованы в объединенные матрицы взаимодействий Hi-C. Для проведения сравнительного анализа, из всех Hi-C карт была удалена главная диагональ матрицы и проведена итеративно корректирующая (ICE) балансировка [115] матрицы с помощью команды `cooler.balance`, после чего отдельные карты были сэмплированы до одного размера с помощью функции `cooltools.sample()`. Все карты были получены на разрешении 1 тыс. п.о. с последующим расширением числа доступных разрешений (до 5 млн. п.о.) с помощью функции `cooler.zoomify`.

2.1.2 Разметка хроматиновых компартментов в данных Hi-C

Для расчета вектора первой главной компоненты и последующего анализа компартментов, были использованы объединенные Hi-C карты, полученные путем предварительного слияния отдельных несэмплированных карт и последующей подготовки, описанной выше. Сами главные компоненты были получены с помощью пакета cooltools, написанного на языке Python. Вектор главной компоненты, который лучше всего коррелировал с GC-содержимым (PC1), использовался для анализа компартментов на разрешении 100 тыс. п.о. Бины каждой хромосомы были ранжированы в соответствии с их значениями PC1, и 1% бинов с самыми высокими и самыми низкими значениями PC1 были отфильтрованы. Затем нормализованные (наблюдаемое/ожидаемое, или O/E) частоты контактов усреднялись в пределах квантилей PC1, и полученные значения использовались для построения седловых графиков.

2.1.3 Разметка ТАДов и петель

Индекс инсуляции (Insulation score, или IS) для сэмпелированных карт был рассчитан для всего генома на разрешении 15 тыс. п.о. с применением функции cooltools.insulation с окном 120000. Полученные таблицы IS были переведены в bigwig формат с помощью функции bedGraphToBigWig. Для анализа границ и петель использовались функции cooltools.insulation и cooltools.dots. Для оценки статистической значимости находок использовался двусторонний T-тест (пакет scipy.stats).

2.2 Разработанные нами методы анализа данных Hi-C

Протокол Hi-C был разработан в 2009 году [3], однако методология анализа данных Hi-C на тот момент практически отсутствовала, поэтому требовалась

разработка инструментов и подходов, позволяющих анализировать этот новый тип данных. Сразу после появления первого датасета Hi-C [3], мы приступили к разработке методологии анализа этих данных и в 2012 году опубликовали первую статью [36] с пилотной проработкой подходов к биоинформатическому анализу данных Hi-C на различных уровнях организации хроматина - от меж-хромосомных контактов до ТАДов (разрешение карт Hi-C тогда не позволяло рассматривать хроматиновые петли, поэтому методологию анализа петель мы смогли раскрыть только в недавних работах).

Однако данные Hi-C для клеток мозга человека появились гораздо позднее. В 2016 году была опубликована первая работа с данными Hi-C для клеток мозга эмбрионов человека [95], а первые данные для взрослого мозга человека появились только в 2021 году [11; 12]. Поэтому разработку инструментов и подходов нам пришлось отрабатывать на иных данных, которые стали доступны ранее - для клеточных линий человека, мыши, дрозофилы и др., по мере их появления. В каждой из наших методических работ, которые будут обсуждаться ниже, мы уделяли особое внимание универсальности разрабатываемого подхода, чтобы его возможно было применять к Hi-C данным для мозга человека, которые ещё не были доступны на момент разработки подхода. Для этого мы тестировали каждый подход на нескольких наборах данных для разных организмов (включая человека), чтобы удостовериться в его универсальности. Стоит отдельно отметить, что для тестирования хорошо подходили данные Hi-C для плодовой мушки дрозофилы, т.к., во-первых, они рано стали доступны, и, во-вторых, геном дрозофилы в 10 раз меньше генома человека, что на два порядка сокращало время расчетов в процессе тестирования.

2.2.1 Суммарная частота контактов геномного региона является его биологическим свойством, а не технической погрешностью

Несмотря на существенное количество разработанных инструментов для анализа данных Hi-C (см. обзор в работе [116]), ни один из них не позволяет провести полный всесторонний анализ и решить технические проблемы на

всех этапах. В результате, многие технические проблемы, свойственные данным Hi-C, остаются нерешенными.

В одном из наших ранних исследований мы показали, что различные методы высокопроизводительного секвенирования подвержены технической вариабельности [117]. Метод Hi-C не является исключением. Одним из недостатков анализа и интерпретации данных Hi-C является наличие систематических погрешностей, таких как различная доступность для ферментов, амплификации и картируемости участков ДНК, что приводит к различной представленности геномных участков в данных Hi-C. Итеративная коррекция [115] - один из наиболее популярных методов, разработанных для устранения этих систематических искажений. Метод итеративной коррекции основан на предположении, что все области хроматина имеют равное суммарное количество наблюдаемых контактов в данных Hi-C. Другими словами, процедура итеративной коррекции выравнивает экспериментальную представленность между различными участками генома, аппроксимируемую суммарной частотой контактов, далее сокращенно обозначаемой как CCF (Cumulative Contact Frequency). Однако различия в экспериментальной представленности могут быть объяснены не только техническими, но и биологическими факторами, такими как открытость хроматина, которая сильно варьируется между различными состояниями хроматина.

В работе [118] нам удалось показать, что CCF положительно коррелирует с активной транскрипцией. И это также связано со структурой компартментов, поскольку компартмент А имеет более высокие уровни CCF и экспрессии генов, чем компартмент В. Примечательно, что это наблюдение справедливо для широкого спектра видов, включая человека, мышь и дрозофилу. Поскольку для анализа данных Hi-C в мозге человека также необходимо было выбрать лучшую процедуру нормализации, требовалось более глубокое изучение этого вопроса.

Мы скачали и подготовили к анализу публичные карты Hi-C по частоте контактов хроматина в *H. sapiens* и *D. melanogaster* [40; 119], а также данные о состояниях хроматина – определенных комбинациях модификаций гистонов и других эпигенетических маркеров, характеризующих паттерны экспрессии [120; 121]. Была использована таблица функциональных характеристик с разрешением 1 млн. п.о., в которой для каждого участка указаны вероятности принадлежности к каждому из состояний хроматина, и к ней был добавлен

столбец значений суммарных ненормализованных (до итеративной коррекции) частот контактов Hi-C - CCF (рис. 2.1А).

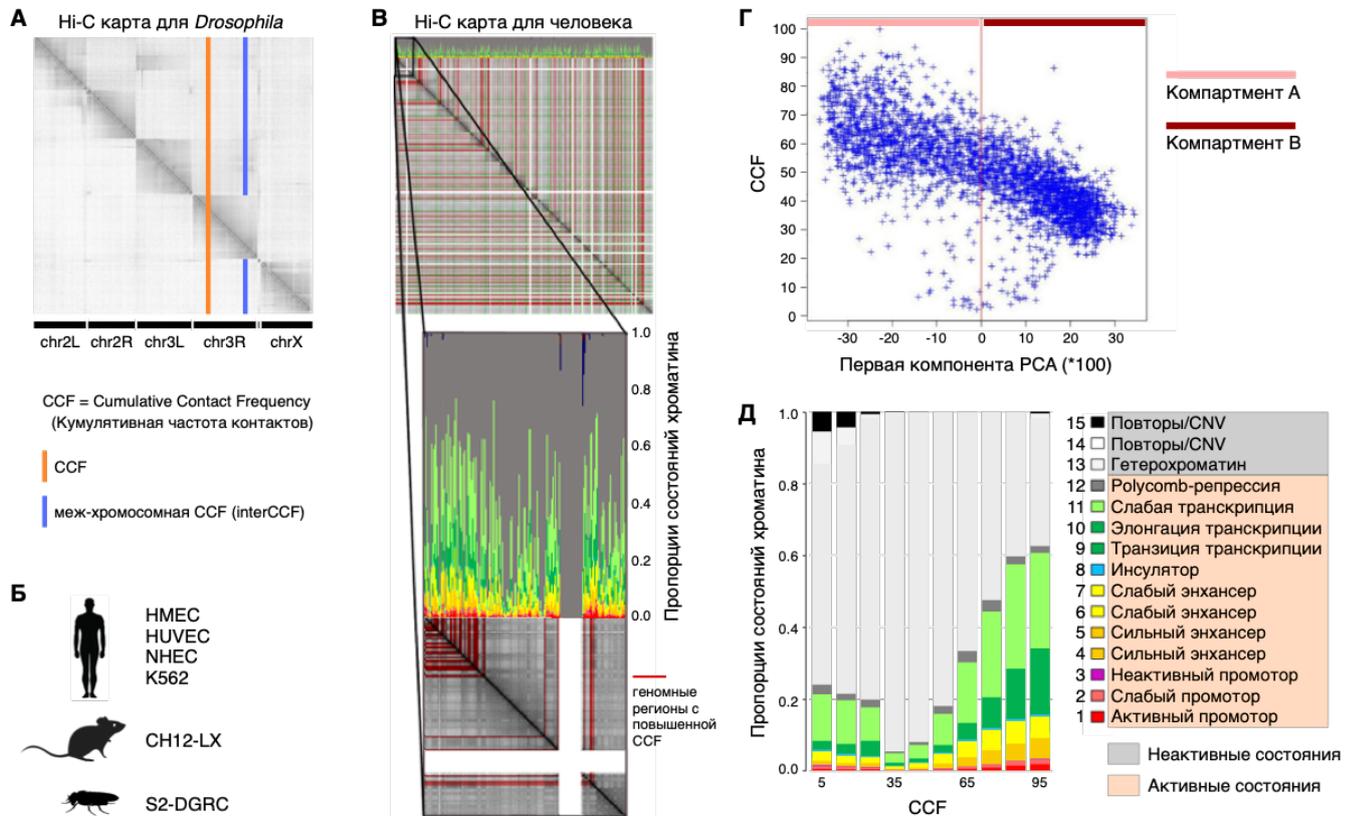


Рисунок 2.1 — Зависимость между состояниями хроматина и частотами контактов (CCF) в хроматине. (А) Определение суммарных ненормализованных (до итеративной коррекции) частот контактов Hi-C, далее сокращенно обозначаемых как CCF (Cumulative Contact Frequency). (Б) В анализе участвовали четыре клеточных линии для человека – HMEC, HUVES, NHEC и K562, клеточная линия CH12-LX для мыши и клеточная линия S2 для дрозофилы. (В) Hi-C карта полного генома человека, совмещенная с графиком распределения процентного содержания состояний хроматина. На Hi-C карте красным показаны участки с аномально высокой CCF, зеленым показаны границы хромосом. На прилежащем к Hi-C карте графике для каждого участка показано процентное содержание каждого из состояний. Состояния отмечены цветами (см. легенду на панели Д). (Г) CCF резко меняется при переходе от одного компартмента к другому. (Д) Барплот зависимости распределения различных состояний хроматина от CCF. По оси X по возрастанию отложены значения CCF. По оси Y - процентные содержания каждого из состояний хроматина для участков с данными значениями частоты контактов. Адаптировано из [118].

С помощью обработки таблицы с объединенными данными Hi-C и состояниями хроматина была исследована зависимость между состояниями хроматина и частотами контактов (CCF) в хроматине. Были выбраны четыре клеточных линии для человека – HMEC, HUVES, NHEC и K562 (однако большая часть исследования проводилась на HMEC), и клеточная линия S2 для дрозофилы (рис. 2.1Б). В областях с CCF выше порога наблюдается большее количество маркеров активного хроматина, чем в остальных областях, в то время как для гетерохроматина наибольшее процентное содержание наблюдается в областях с низкой CCF (рис. 2.1В). С ростом CCF наблюдается рост процентного содержания всех активных состояний хроматина и уменьшение процентного содержания гетерохроматина (рис. 2.1Д).

Для геномов *D. melanogaster* и *H. sapiens*, удалось выявить паттерны корреляций между состояниями хроматина и частотами контактов (CCF) в хроматине. Была построена матрица корреляций Пирсона для каждой пары состояний хроматина и для каждой пары состояние - CCF (рис. 2.2). Гетерохроматин оказался отрицательно скоррелирован со всеми активными состояниями и с CCF. Между состояниями 1-12 (активные состояния) и CCF, напротив, наблюдаются положительные корреляции, что подтверждает тенденцию, показанную на рис. 2.1Д. Дополнительно, мы провели аналогичный анализ для генома мыши (клеточная линия CH12-LX) и получили похожие результаты [118]. Отдельно были проанализированы меж-хромосомные частоты контактов (interCCF, рис. 2.1А), чтобы продемонстрировать, что наблюдаемые нами зависимости не определяются только локальными внутри-хромосомными взаимодействиями.

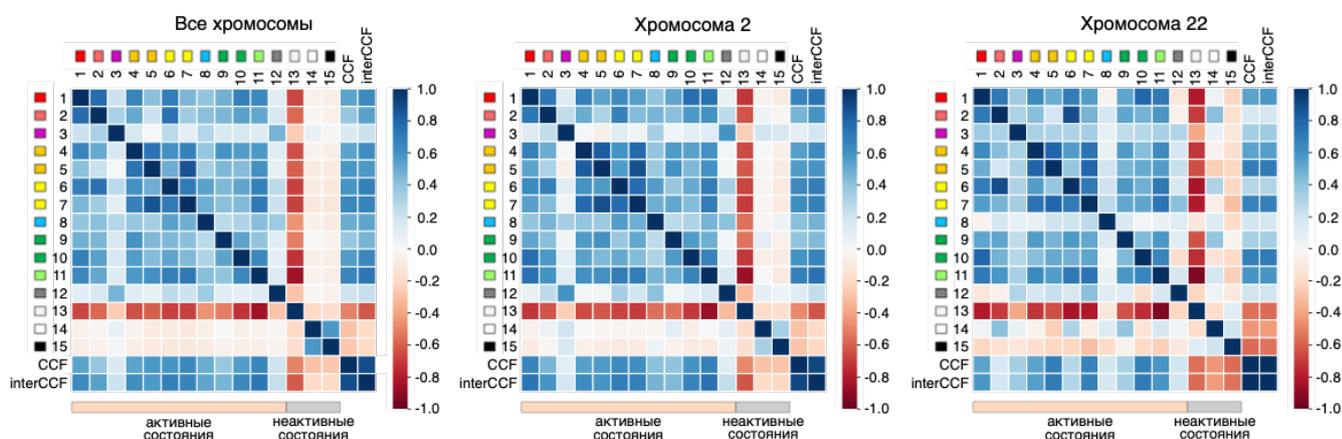


Рисунок 2.2 — Матрицы корреляций для состояний хроматина и CCF в геноме человека для разных хромосом. Адаптировано из [118].

В теории, могла возникнуть ситуация, в которой для одной области генома наблюдалась бы сильная положительная корреляция, а в другой - сильная отрицательная. При подсчете одного коэффициента корреляции в таком случае было бы получено значение около нуля, и таким образом могла быть потеряна биологически значимая информация. Поэтому полученные коэффициенты корреляций были подтверждены с помощью инструмента Stereogene [122]. Данные результаты подтверждают правильность подсчитанных корреляций [118].

Далее, мы исследовали зависимость найденных корреляций от процентного содержания GC нуклеотидов, размера хромосомы, положения топологически ассоциированных доменов и компарментов хроматина (рис. 2.1Д), чтобы установить вклад этих факторов в наблюдаемые нами паттерны корреляций. В предыдущих исследованиях было предположено, что содержание GC нуклеотидов в местах сшивки участков генома может влиять на результаты эксперимента Hi-C, и поэтому в некоторых случаях проводится коррекция, нивелирующая связь содержания GC нуклеотидов и частоты контактов. Действительно, мы наблюдаем положительную корреляцию частоты контактов и содержания GC нуклеотидов (коэффициент корреляции Пирсона 0.61) [118]. Далее, с целью нивелировать положительную корреляцию содержания GC нуклеотидов и частоты контактов, мы вычислили CCF, деленную на содержание GC нуклеотидов и CCF, предсказанную по содержанию GC нуклеотидов с использованием линейной регрессии. Поскольку корреляции состояний хроматина с этими модифицированными CCF практически не отличались от корреляций с исходной CCF, можно сделать вывод, что наблюдаемые нами зависимости не обусловлены влиянием содержания GC нуклеотидов на частоту контактов.

В предыдущих исследованиях было показано, что различные хромосомы имеют разные частоты контактов с другими хромосомами. Например, маленькие обогащенные генами хромосомы чаще контактируют друг с другом. В соответствии с этими исследованиями, мы нашли, что с увеличением длины хромосомы уменьшается суммарное нормализованное число контактов. Далее, мы нормировали CCF на длину хромосомы или на среднее арифметическое длин хромосом в случае межхромосомных контактов и построили новую корреляционную матрицу [118], которая показала уменьшение выраженности сигнала, однако сам паттерн корреляций остается тем же.

Многие современные исследования посвящены различиям в активности хроматина между ТАДами и межТАДами. С помощью алгоритма Lavaburst программы Armatius мы установили границы ТАДов для геномов человека и дрозофилы. ТАДов оказалось в два раза больше по суммарной длине, чем межТАДного пространства. Далее, были построены матрицы корреляций для ТАДов и межТАДов отдельно. Оказалось, что межТАДы сильно отличаются по своим свойствам от ТАДов [118]. Корреляционный паттерн ТАДов в большей степени похож на паттерн для всего генома, поскольку ТАДы с выбранными границами покрывают 2/3 генома.

Компартменты – это ещё один уровень организации хроматина, не менее важный, чем ТАДы, с функциональной точки зрения. Для поиска компартментов, матрица частот контактов была нормализована на матрицу ожидаемых частот контактов – матрицу, в которой частоты контактов усреднены по каждой диагонали Hi-C карты, то есть по всем участкам, находящимся на определенном расстоянии друг от друга в геноме. Далее мы подсчитали корреляции Пирсона между каждой строчкой и каждым столбцом в нормализованной матрице частот контактов, чтобы получить матрицу корреляций. Для нахождения границ компартментов по результирующей матрице корреляций был выполнен анализ главных компонент, и взята первая компонента. Существование компартментов в геноме человека и их отсутствие в геноме дрозофилы было подтверждено рандомизацией Hi-C карт и оценкой процента объясненной дисперсии для первой, второй и третьей компоненты. По определению, компартменты имеют разные предпочтения в частоте контактов. Поэтому ожидаемо, что SCF резко меняется при переходе от одного компартмента к другому у человека (рис. 2.1Г).

Мы заметили, что хромосомы значительно отличаются друг от друга по корреляционным паттернам с состояниями хроматина (рис. 2.2). В геноме человека после 10 хромосомы каждая хромосома имеет свои индивидуальные паттерны корреляций, причем эти паттерны повторяются для разных клеточных линий. Необходимо учитывать, что наибольший вклад в корреляции вносят самые большие хромосомы. Поэтому полногеномная корреляционная матрица наиболее сходна с матрицей для первых пяти хромосом (рис. 2.2).

Наблюдаемая вариабельность матриц корреляций для коротких хромосом требовала дальнейшего изучения, так как она может отражать важные биологические различия хромосом. Однако также она может быть объяснена недо-

статочным для статистических расчетов объемом данных для коротких хромосом, поскольку размер коротких хромосом (в данном случае 17-22 хромосомы) примерно в пять раз меньше размера длинных хромосом (1-2 хромосом). Для проверки этой гипотезы была проведена процедура уравнивания объема данных. Объединенная выборка данных для нескольких коротких хромосом (17-22 хромосом), равная по размеру (по суммарной длине выбранных хромосом) была обработана тем же образом, что и одна большая хромосома (в данном случае хромосома 2), и паттерны корреляций были сопоставлены. Кроме того, корреляционный паттерн для одной маленькой хромосомы (из 17-22) был сравнен с паттерном для части большой хромосомы (хромосомы 2). Поскольку пик ожидаемого распределения корреляций после уравнивания объема данных не совпадает с наблюдаемым значением (рис. 2.3А), наблюдаемые различия в паттернах корреляций, скорее всего, обладают некоторым биологическим смыслом.

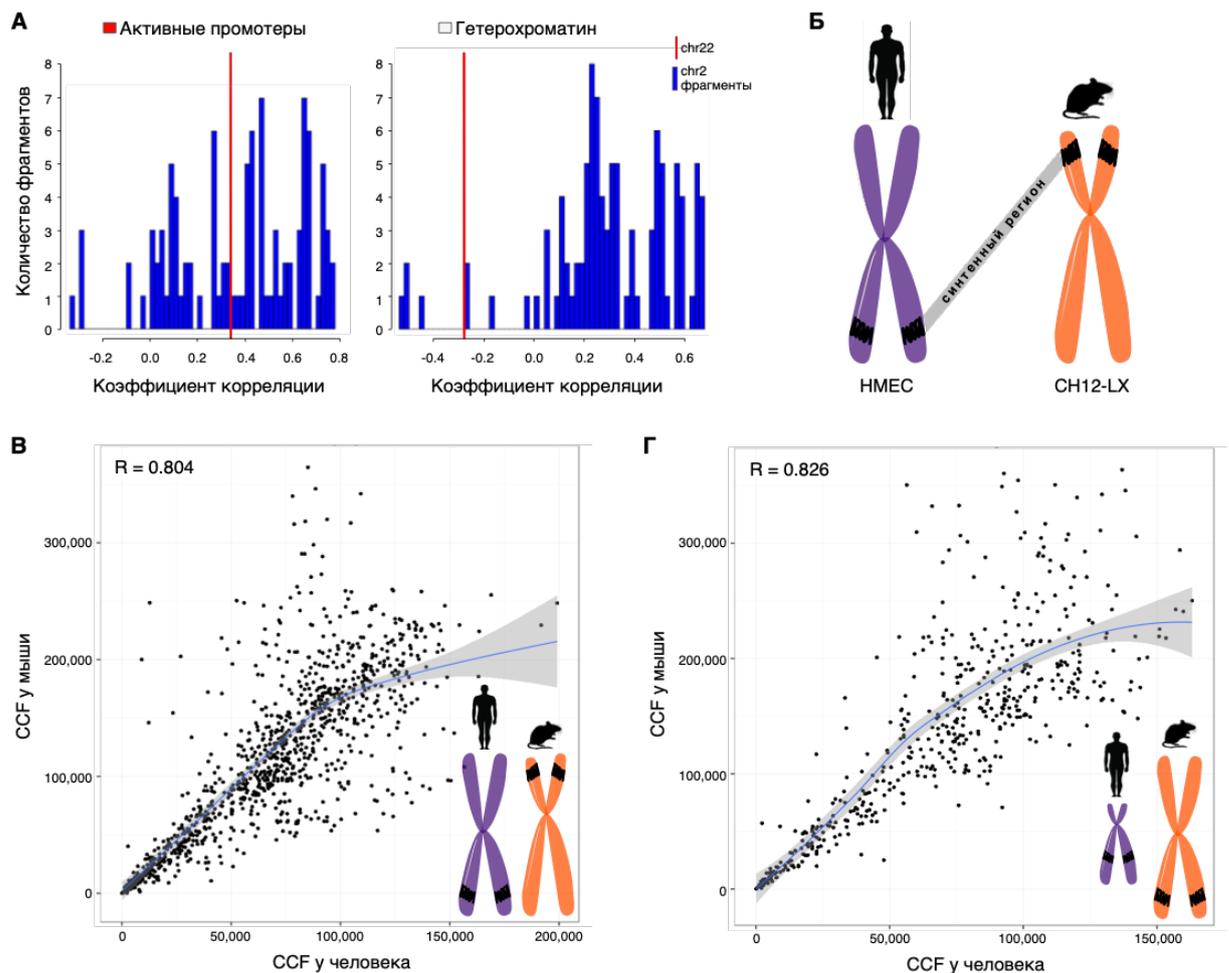


Рисунок 2.3 — CCF является независимым от положения свойством областей хроматина, сохраняющимся в синтенных переходах. (А) Уникальные корреляционные паттерны для небольших хромосом не являются следствием небольшого размера выборки, а обусловлены синтенными областями небольших хромосом (клеточная линия человека НМЕС). Левая панель: распределение корреляций между CCF и состоянием активного промотора хроматина для случайных фрагментов хромосомы 2 (синие столбцы) по сравнению с реальной корреляцией для хромосомы 22 (красная линия). Правая панель: распределение корреляций между CCF и состоянием гетерохроматина для случайных фрагментов хромосомы 2 (синие столбцы) по сравнению с реальной корреляцией для хромосомы 22 (красная линия). (Б) Схематичное изображение синтенной области между двумя хромосомами генома человека и мыши. (В-Г) CCF у человека по сравнению с CCF у мыши для синтенных областей в больших человеческих хромосомах и больших мышиных хромосомах (В), а также в малых человеческих хромосомах и больших мышиных хромосомах (Г). Каждая точка представляет собой синтенную область размером 1 Мб. Адаптировано из [118].

Какие области генома дают наибольший вклад в наблюдаемые на коротких хромосомах корреляционные паттерны? Небольшие хромосомы могут демонстрировать уникальные корреляционные паттерны из-за содержания определенных эволюционно консервативных областей, таких как синтенные блоки. Чтобы проверить эту гипотезу, мы аннотировали синтенные области в наборах данных человека и мыши. Разметка синтенных областей была получена из базы данных <http://www.informatics.jax.org/homology.shtml>, в анализ были взяты только области длиной более 2000 пар оснований. Для установления гомологии между полными геномами человека и мыши был использован инструмент liftOver. Неоднозначно картированные области - фрагменты генома человека, картируемые сразу в несколько областей генома мыши (либо наоборот), подвергались дальнейшей обработке. Частота контактов между каждой парой таких гомологичных фрагментов нормировалась на длину области картирования.

Далее, мы вычислили ССФ для аннотированных синтенных областей в наборах данных человека и мыши. Действительно, у синтенных областей есть схожие предпочтения в частотах контакта между двумя видами (рис. 2.3Б,В). Чтобы понять, как синтенные области наследуют свойства во время геномных перестроек в эволюции, мы идентифицировали синтенные блоки, расположенные в небольших хромосомах генома человека, но в больших хромосомах генома мыши, и наоборот. Примечательно, что эти области демонстрируют сходную ССФ в геномах человека и мыши (рис. 2.3Г, корреляция между частотами контакта в геномах человека и мыши составляет 0.79). Таким образом, наблюдаемые корреляционные предпочтения являются внутренними свойствами синтенных блоков, так как они не зависят от расположения области в геноме и наследуются, несмотря на эволюционные перестройки между хромосомами.

Таким образом, результаты этой работы помогли установить недостающее звено между процедурами нормализации данных Hi-C, такими как итеративная коррекция, удаляющими ССФ из данных, и плохо изученными, но скорее всего осмысленными биологическими факторами, вносящими вклад в ССФ. Поэтому в дальнейших исследованиях архитектуры хроматина в мозге человека мы с осторожностью относились к процедуре итеративной коррекции и повторяли основные этапы анализа как с итеративной коррекцией, так и без нее, чтобы убедиться, что полученные нами результаты принципиально не зависят от выбора метода нормализации.

2.2.2 Выявление биологического сигнала в данных с высоким уровнем технической вариабельности

Другой нерешенной технической проблемой в Hi-C данных является высокий уровень технической вариабельности, так называемый батч-эффект, который может приводить к низкому соответствию между экспериментальными повторностями. Поскольку для анализа архитектуры хроматина в мозге человека мы используем большое количество карт Hi-C, полученных в рамках разных батчей (т.к. в рамках одного экспериментального батча Hi-C возможно получить не более 6 карт Hi-C), требовалось решить проблему батч-эффекта в данных.

Существует несколько методов для выравнивания отличий между картами Hi-C, основанных на различных приемах нормализации: HiCorr, HiFive, HiCNorm - методы для коррекции известных сопутствующих факторов [123–125]; multiHiCcompare - совместная нормализация нескольких Hi-C датасетов с использованием LOESS регрессии [126]; Binless - независимый от разрешения карт метод, основанный на негативной биномиальной регрессии [127]. Однако эти методы не подходят для случаев, когда в эксперименте происходят глобальные перестройки укладки хроматина. В таких случаях приходится отказаться от нормализации. Поэтому мы разработали принципиально другой подход, основанный на взвешивании контактов между участками генома пропорционально вариабельности между техническими повторностями одного и того же эксперимента - HiConfidence [128].

Подход HiConfidence основан на учете «надежности» контактов между участками генома, на основе сравнения данных для технических повторностей одного и того же эксперимента. Будем учитывать только те контакты, которые имеют одинаковый биологический эффект. Вклад каждого пикселя на карте Hi-C в усредненную плотность рассчитывается с учетом его дисперсии. А именно, «надежность» пикселя вычисляется как: $1/(|R1-R2|/(R+1))^{*k}$, где R1 и R2 - это значения пикселя на первой и второй карте соответственно, R - это среднее значение пикселя, k - показатель степени, подбираемый эмпирически для лучшей сходимости (рис. 2.4). Данный параметр может быть использован при любых подсчетах обобщенной статистики в качестве «веса» данного пиксе-

ля. Такое маскирование Hi-C карты позволяет учитывать только достоверные значения и снижать вклад недостоверных значений в общую картину.

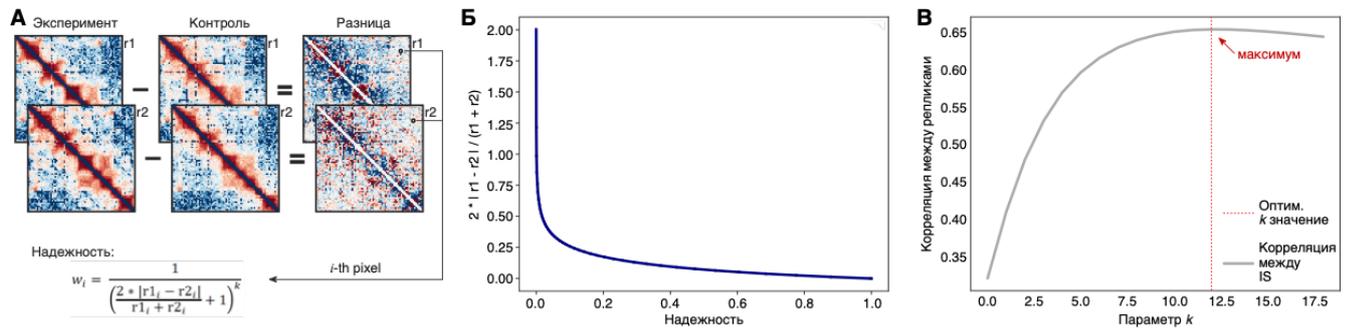


Рисунок 2.4 — Описание алгоритма HiConfidence. (А) Ключевые этапы процедуры HiConfidence. Сперва разница между картами эксперимента и контроля Hi-C рассчитывается для каждой биологической повторности (r_1 и r_2) отдельно. Далее оценивается достоверность контакта для каждого пикселя полученных карт. (Б) Значение достоверности обратно пропорционально разнице между биологическими репликами, деленной на их среднее значение (в степени k). (В) Параметр k может быть оптимизирован путем максимизации коэффициента корреляции Пирсона между дифференциальными профилями инсуляции биологических повторностей. Пунктирной линией отмечено оптимальное значение k . Стрелка указывает на максимальный коэффициент корреляции Пирсона. Адаптировано из [128].

Сначала мы протестировали этот подход для оценки изменения плотности топологически ассоциированных доменов (ТАДов). Поскольку данные Hi-C для мозга человека ещё не были доступны, мы проанализировали три набора данных Hi-C для клеток *Drosophila melanogaster* S2 (рис. 2.5А). Каждый набор данных содержал две экспериментальные реплики и две контрольных реплики клеток S2, обработанных различными агентами с целью получения изменений в ацетилировании гистонов. К нашему удивлению, кластеризация итеративно скорректированных карт Hi-C продемонстрировала группировку по экспериментальным батчам вместо биологических условий (рис. 2.5Б). Это противоречило нашей гипотезе о том, что все контрольные реплики должны образовывать отдельный кластер необработанных клеток. Реплики, представляющие собой деpleцию деацетилазы гистонов (HDAC-dep) и ингибирование деацетилаз гистонов агентом TSA (HDAC-inh), должны были быть сгруппированы в отдельный кластер, соответствующий повышенному ацетилированию гистонов. А экспери-

ментальные повторности ингибирования гистонацетильтрансфераз куркумином (HAT-inh) должны были образовать еще один отчетливый кластер, соответствующий пониженному ацетилированию гистонов. Поскольку такая группировка не наблюдалась, мы пришли к выводу, что анализируемые нами наборы данных Hi-C могут содержать в себе нежелательные технические эффекты.

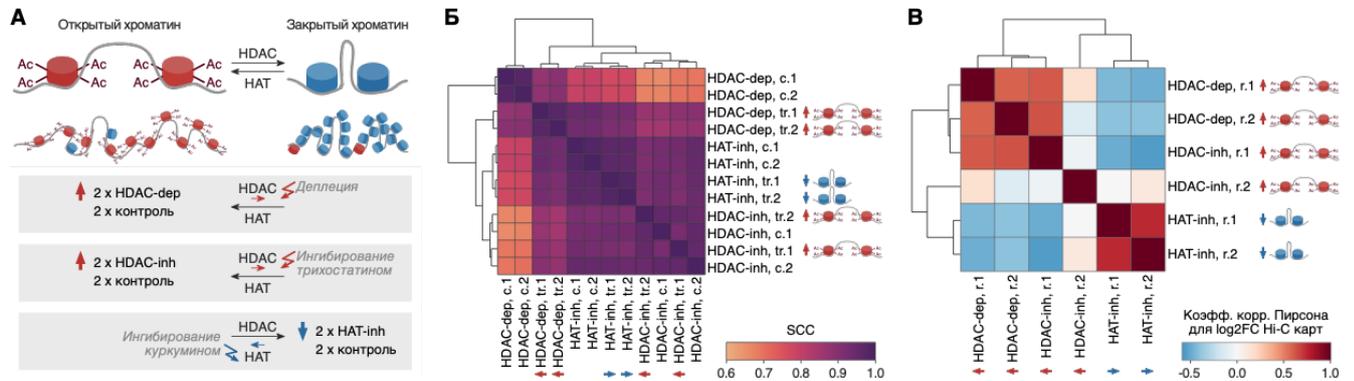


Рисунок 2.5 — Технические проблемы мешают прямому сравнению наборов данных Hi-C. (А) Дизайн экспериментов Hi-C, используемых для оценки эффективности подхода HiConfidence. Красными стрелками показана деплеция деацетилазы гистонов (HDAC-dep) и ингибирование деацетилазы гистонов трихостатином А (HDAC-inh), повышающее уровень ацетилирования. Синими стрелками отмечено ингибирование гистоновых ацетильтрансфераз куркумином (HAT-inh), снижающее уровень ацетилирования. (Б) Корреляция карт Hi-C для обработанных (tr.) и контрольных (с.) повторностей, оцененная с использованием коэффициента корреляции с поправкой на ожидаемое (SCC). Цвета показывают SCC. (В) Корреляция дифференциальных профилей инсуляции для карт $\log_2\text{FC}$ Hi-C, представляющих собой разницу между обработанными репликами (r.) и соответствующими контрольными. Цвета отображают коэффициенты корреляции Пирсона. Адаптировано из [128].

Группировка по экспериментальным батчам вместо биологических условий указывает на то, что на экспериментальные повторности и контрольные реплики в составе одного экспериментального батча может влиять один и тот же технический фактор. Следовательно, анализ значений \log_2 -кратного изменения ($\log_2\text{FC}$) между обработанными и контрольными картами Hi-C внутри батча вместо отдельных карт Hi-C может уменьшить влияние технических факторов, аналогично процедуре сокращения дробей в школьной математике. Поэтому мы

вычислили \log_2FC карту Hi-C для каждой пары контрольных и обработанных карт Hi-C. Затем для каждой карты \log_2FC Hi-C мы рассчитали дифференциальные профили инсуляции, представляющие собой локальные уровни компактизации хроматина, и получили попарные корреляции дифференциальных профилей инсуляции между всеми картами \log_2FC Hi-C (рис. 2.5B). Эта кластеризация лучше соответствовала ожидаемой группировке образцов по биологическим условиям, за исключением одной повторности эксперимента HDAC-inh, которая продемонстрировала неожиданную кластеризацию и, по всей видимости, наиболее существенно пострадала от нежелательного технического шума.

Таким образом, повторности эксперимента HDAC-inh продемонстрировали противоречивые результаты и представляли собой наиболее существенную проблему в нашем анализе, поскольку они не позволяли сделать однозначный вывод о биологическом влиянии ингибирования деацетилазы гистонов на архитектуру хроматина. Для того, чтобы научиться извлекать биологический сигнал из таких зашумленных данных Hi-C, и был разработан алгоритм HiConfidence.

После применения алгоритма HiConfidence к этому набору данных, карты \log_2FC Hi-C, представляющие собой разницу между обработанными повторностями и соответствующими контролями, объединились в кластеры, соответствующие биологическим условиям, в полном соответствии с ожидаемой группировкой (рис. 2.6A, по сравнению с рис. 2.5B). Реплики HDAC-inh продемонстрировали хорошую согласованность и сгруппировались вместе с репликами HDAC-dep, как и должно было быть. Следует отметить, что применение алгоритма HiConfidence перенесло вторую реплику HDAC-inh в кластер, содержащий все оставшиеся эксперименты с повышенным уровнем ацетилирования гистонов, и не изменило кластеризацию первой реплики HDAC-inh, что свидетельствует о лучшем исходном качестве последней. Таким образом, подход HiConfidence взвешивает значимость контактов таким образом, что контакты с низкой степенью достоверности слабо влияют на результат. При этом повторности низкого качества критически улучшаются, а высококачественные повторности не портятся. Кроме того, HiConfidence в результате своей работы вычисляет надежную средневзвешенную карту Hi-C, которая учитывает качество повторностей.

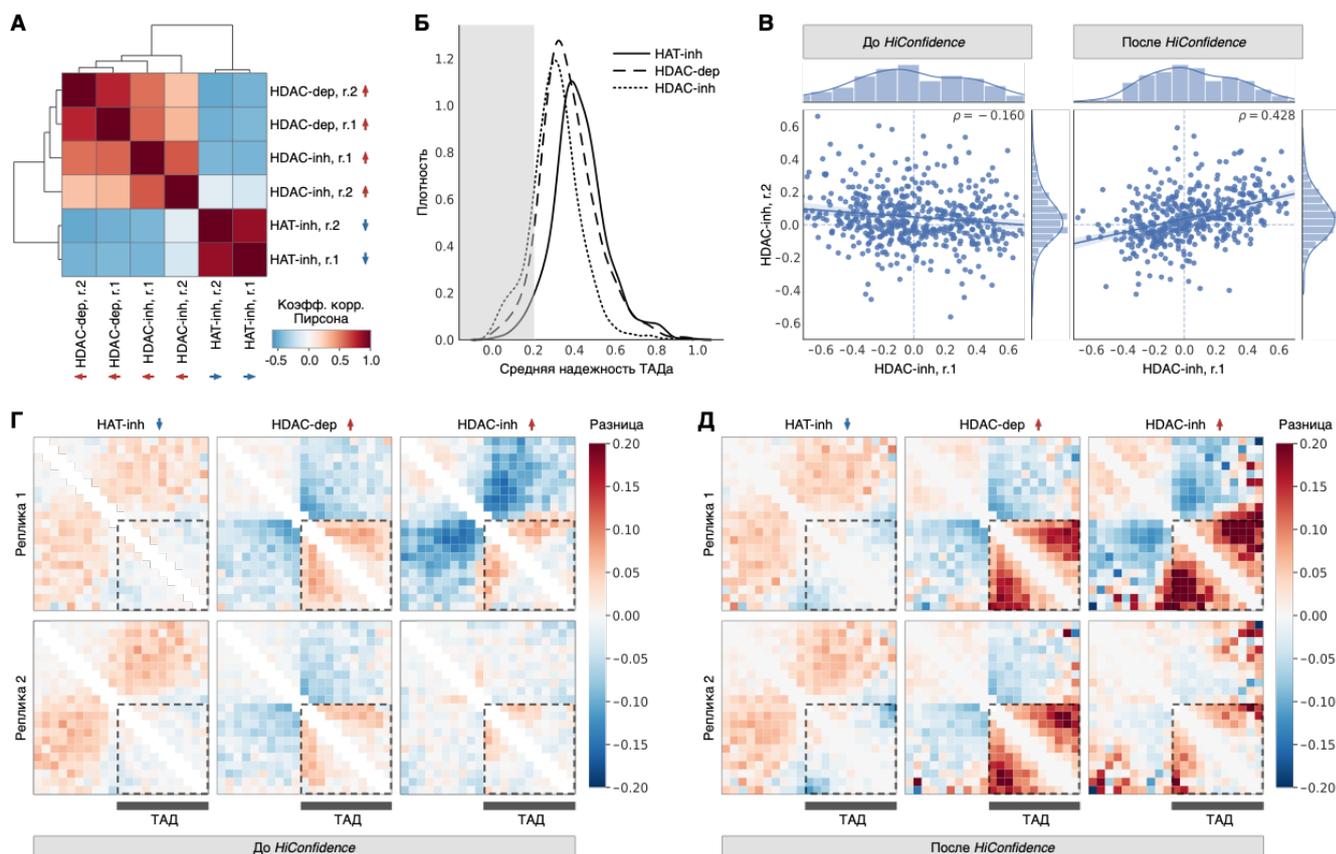


Рисунок 2.6 — HiConfidence извлекает биологический сигнал из данных Hi-C, затронутых техническим шумом. (А) Корреляция профилей дифференциальной инсуляции для карт $\log_2\text{FC}$ Hi-C, представляющих собой разницу между обработанными повторностями (r.) и соответствующими контролями. Цвета как на рисунке 2.5. (Б) Распределение средних значений достоверности ТАДов для экспериментов HDAC-dep (пунктирная линия), HDAC-inh (пунктирная линия) и HAT-inh (сплошная линия). (В) Корреляция изменений плотности хроматина в ТАДах между повторностями эксперимента HDAC-inh до (слева) и после (правой) HiConfidence. (Г и Д) Средние ТАДы на картах $\log_2\text{FC}$ Hi-C, представляющие собой разность между обработанными повторностями и соответствующими контролями, до (Г) и после (Д) HiConfidence. Пунктирные линии выделяют области с заметными улучшениями в результате процедуры HiConfidence. Адаптировано из [128].

Мы аннотировали ТАДы с использованием алгоритма Armatus [129] и рассчитали среднюю достоверность для каждого ТАДа, используя процедуру HiConfidence, чтобы оценить качество и согласованность полученной аннотации ТАДов (рис. 2.6Б). Заметный сдвиг среднего значения и достоверности в сторону более высоких значений для HAT-inh свидетельствует о более высокой согла-

сованности повторностей NAT-inh. Аналогичным образом, повторности HDAC-inh продемонстрировали более низкую согласованность. Следовательно, подход HiConfidence может быть также применен для оценки качества карты Hi-C и согласованности реплик.

Далее, был построен средний ТАД, причем параметр надежности алгоритма HiConfidence выступал в качестве веса пикселя при подсчете среднего в этом анализе. В результате удалось добиться сходимости результатов для двух технических реплик, где ранее результаты были противоречивыми. Корреляция эффектов в двух репликах увеличилась с -0.16 до 0.43 (рис. 2.6В). Это позволило понять, какая реплика показывала более достоверный эффект, и каков результат воздействия агентом в данном эксперименте. В нашем случае наблюдаемый эффект совпал с ожидаемым результатом, и это позволило нам получить биологически осмысленные результаты и глубже понять, какие факторы влияют на формирование ТАДов.

В частности, мы исследовали влияние изменений ацетилирования на плотность хроматина внутри ТАДов (рис. 2.6Г-Д). Важно отметить, что разрешение карт Hi-C в анализируемом нами наборе данных (20 тыс. п.о.) позволяет нам рассматривать только длинные ТАДы (≥ 60 тыс. п.о.), в основном содержащие неацетилированный хроматин. В эксперименте NAT-inh, где ацетилирование гистонов было понижено, плотность хроматина в межТАДах, т.е. в граничных областях, которые, как известно, содержат сильно ацетилированный хроматин [119], была увеличена. В то же время, плотность длинных неактивных ТАДах была понижена. Мы ожидали увидеть противоположный эффект в экспериментах HDAC-dep и HDAC-inh, где ацетилирование гистонов было повышено. Однако повторность HDAC-inh 2 не продемонстрировала заметных изменений плотности хроматина ни в ТАДах, ни в межТАДах (рис. 2.6D). Но после применения алгоритма HiConfidence плотность длинных неактивных ТАДов оказалась повышена как в экспериментах HDAC-dep, так и в экспериментах HDAC-inh (включая HDAC-inh r.2) (рис. 2.6Д). Таким образом, применение алгоритма HiConfidence сделало ожидаемый эффект увеличения ацетилирования гистонов более явным.

Чтобы продемонстрировать значимость этих улучшений, мы рассчитали среднюю разницу в контактах между обработанными и контрольными клетками внутри ТАДов (рис. 2.7А) и показали значимое увеличение этой разницы после

применения HiConfidence для всех проанализированных повторностей (Р-значение < 0.05 ; рис. 2.7Б).

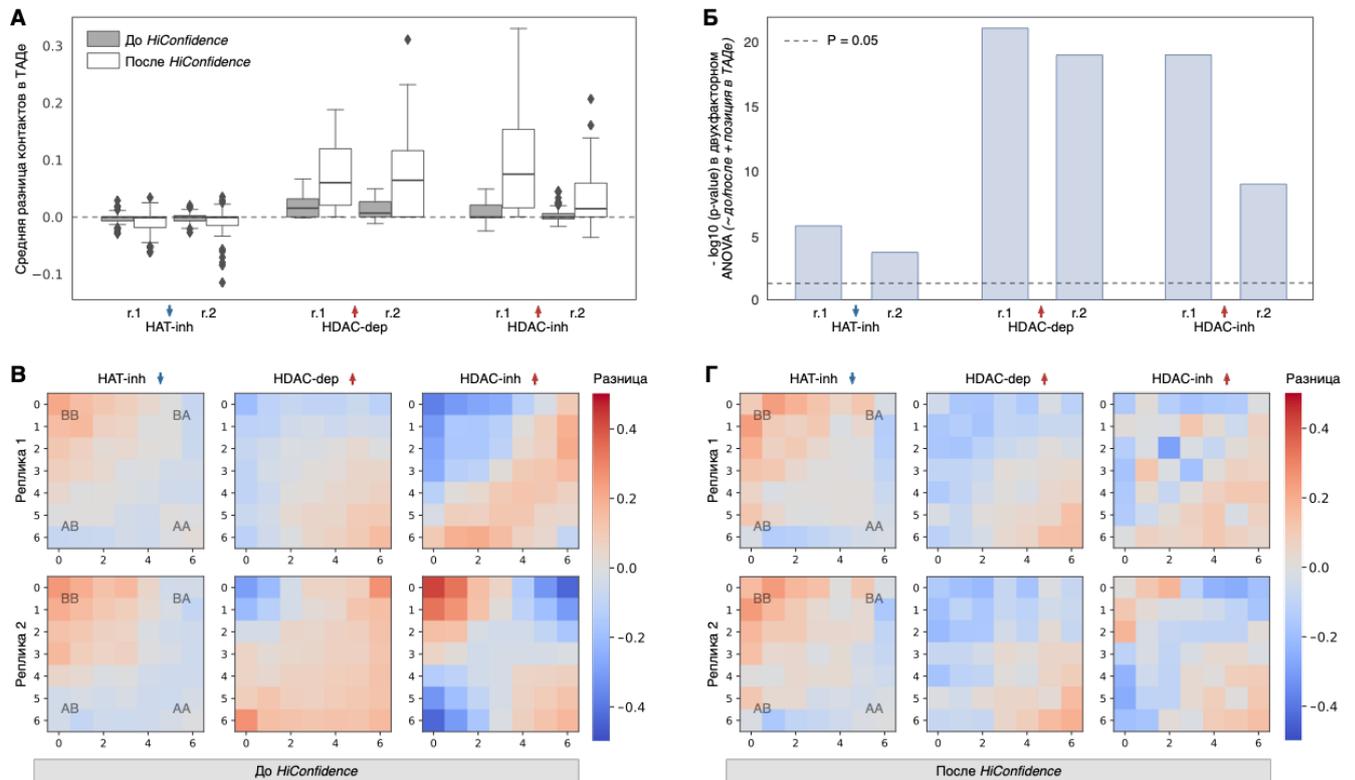


Рисунок 2.7 — HiConfidence значительно улучшает данные Hi-C и приводит к более консистентным результатам на уровне компарментов. (А) Средняя разница в контактах между обработанными и контрольными повторностями, рассчитанными в областях, выделенных на рис. 2.6Г и 2.6Д до (серый) и после (белый) HiConfidence. (Б) Статистическая значимость улучшений в результате процедуры HiConfidence в ТАДах. (В-Г) Разность седловых графиков, представляющих собой выраженность компарментов, между обработанными повторностями и соответствующими контролями, до (В) и после (Г) HiConfidence. Адаптировано из [128].

Далее, мы перешли к анализу более крупномасштабного уровня упаковки хроматина - компарментов. Мы проаннотировали компарменты с помощью пакета cooltools и построили седловые графики, отображающие взаимодействия хроматина внутри (АА, ВВ) и между (АВ, ВА) компарментами. Чтобы проанализировать влияние ацетилирования гистонов на силу компарментов и на взаимодействия хроматина внутри компарментов и между ними, мы вычли седловые графики, построенные для контрольных карт Hi-C, из обработанных для каждой пары повторностей. Эксперименты HAT-inh характеризовались уси-

лением взаимодействий в компартменте В и ослаблением взаимодействий в компартменте А (рис. 2.7В). Поскольку НАТ-inh уменьшает глобальный уровень ацетилирования гистонов, это наблюдение подразумевает, что обработанные клетки образуют более удаленные контакты в компартменте В и менее удаленные контакты в компартменте А. Напротив, эксперименты HDAC-dep характеризовались снижением частоты контактов в компартменте В, а также между компартментами, и увеличенной частотой контактов внутри компартмента А. Поскольку повышенное ацетилирование приводит к более рыхлому ацетилированному хроматину, это наблюдение указывает на то, что обработанные клетки имеют более удаленные контакты внутри компартмента А, который менее отделен от компартмента В. Мы ожидали, что эксперименты HDAC-inh продемонстрируют то же поведение, что и с HDAC-dep, поскольку эта обработка также увеличивала ацетилирование гистонов. Однако повторности HDAC-inh показали противоречивые результаты. После применения HiConfidence эффект стал более консистентным (рис. 2.7Г).

Также мы провели сравнение нашего подхода с существующим методом multiHiCcompare [126]. Весь алгоритм анализа был повторен на тех же данных, но с использованием пакета multiHiCcompare. При этом наш метод показал более высокую корреляцию технических реплик (коэфф. корр. Пирсона = 0.012 and 0.62 для multiHiCcompare и HiConfidence, соответственно, рис. 2.8А). Дополнительно стоит упомянуть, что наш подход больше всего выигрывает, когда реплики сильно отличаются по качеству. Он позволяет извлечь достоверные значения и снизить влияние шума из менее качественной реплики. И хотя оба подхода успешно решили проблему некорректной кластеризации реплик, multiHiCcompare приблизил две повторности эксперимента HDAC-inh друг к другу, но в то же время отдал их от остальных экспериментов, поскольку не учел проблемы с качеством в одной из повторностей и улучшил повторность низкого качества за счет ухудшения высококачественной. Действительно, корреляции высококачественной повторности HDAC-inh r.1 с экспериментами HDAC-dep снизились с 0.61-0.64 до 0.32–0.34 после нормализации multiHiCcompare (рис. 2.8Б). Напротив, HiConfidence улучшил качество повторности низкого качества, не ухудшая качество повторности высокого качества, поскольку корреляции высококачественной повторности HDAC-inh r.1 с экспериментами HDAC-dep снизились лишь незначительно, с 0.61–0.64 до 0.55–0.58 (рис. 2.8В).

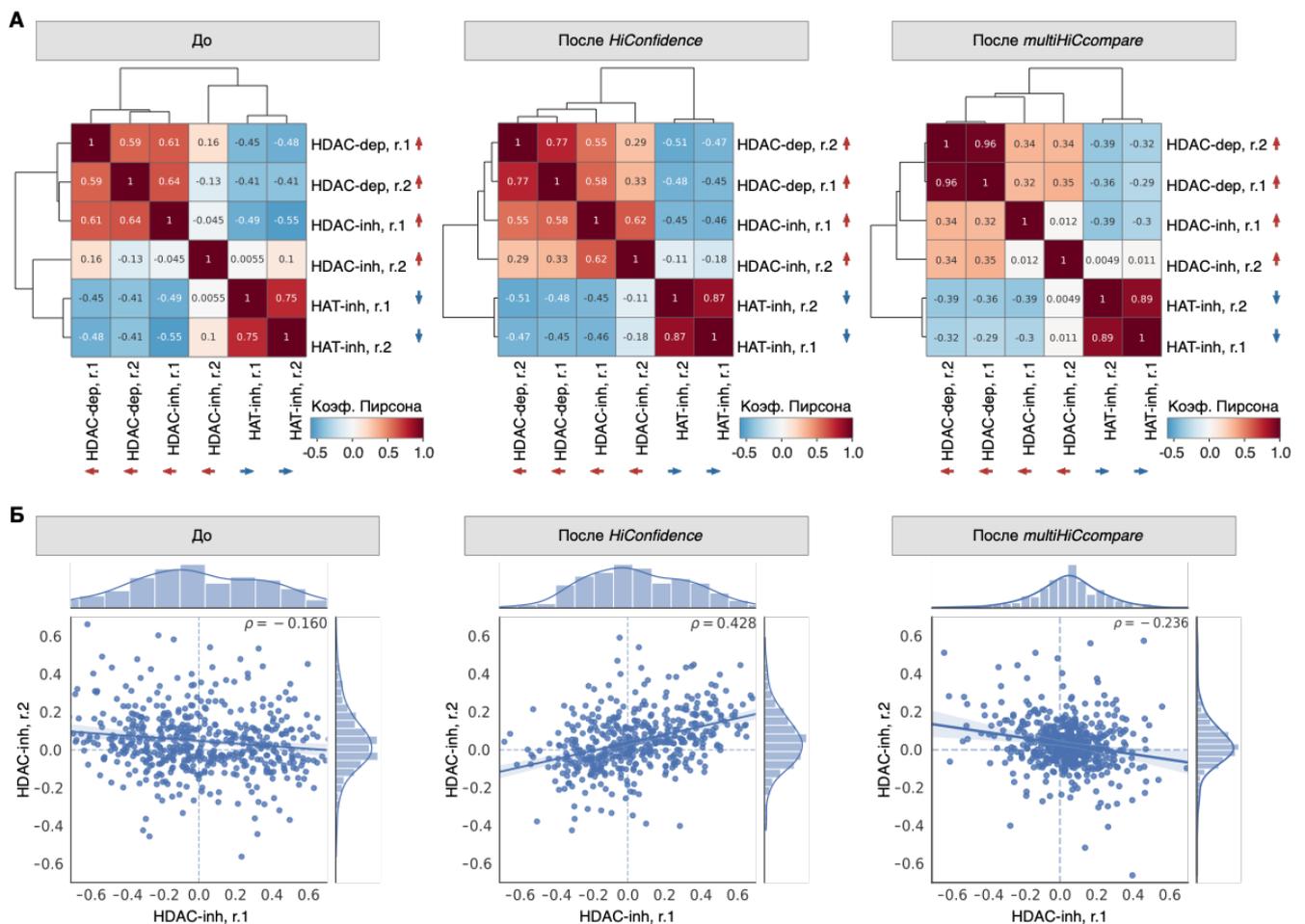


Рисунок 2.8 — Сравнение HiConfidence и multiHiCompare. (А) Корреляция дифференциальных профилей инсуляции карт \log_2FC Hi-C, представляющих собой разницу между обработанными повторностями (r.) и соответствующими контролями. (Б) Корреляция изменений плотности ТАДов между двумя техническими повторностями. Слева – необработанные данные, в центре – данные после использования нашего подхода, справа – данные после использования multiHiCompare. Адаптировано из [128].

Более того, корреляция изменений плотности ТАДов между повторностями эксперимента HDAC-inh, которая была отрицательной до нормализации multiHiCompare или HiConfidence (коэфф. корр. Пирсона = -0.16, рис. 2.6В), не улучшилась после применения multiHiCompare (коэфф. корр. Пирсона = -0.23), в то время как использование HiConfidence значительно улучшило эту корреляцию (коэфф. корр. Пирсона = 0.42, рис. 2.6В). Таким образом, мы приходим к выводу, что HiConfidence, по-видимому, более эффективен, чем multiHiCompare, при отделении биологического сигнала от технического шума в экспериментах Hi-C.

Чтобы продемонстрировать универсальность метода, а также провести корректное сравнение между подходами HiConfidence и multiHiScompare, мы применили их к одному и тому же внешнему набору данных [130], в котором изучалось изменение хроматина на ранних этапах развития. В этой статье для некоторых временных точек изменение плотности ТАДов в технических повторностях было противоречивым, либо не было заметного эффекта. Следуя стратегии анализа, примененной к описанному выше подходу HiConfidence, мы рассчитали корреляции Пирсона для карт \log_2FC Hi-C, представляющих собой разницу между соседними временными точками, и заметили, что multiHiScompare не улучшил корреляции между повторностями (рис. 2.9А). После применения multiHiScompare, пары повторностей показали незначимо более высокие корреляции, чем случайные пары экспериментов (t-критерий Стьюдента, $P = 0.61$, рис. 2.9Б). Средняя плотность ТАДов, рассчитанная между временными точками, также заметно не изменилась после применения multiHiScompare (рис. 2.9В). Таким образом, в отличие от HiConfidence, применение multiHiScompare к тому же внешнему набору данных существенно не улучшило согласованность между репликами. Наш метод, напротив, позволил более уверенно выявить биологические эффекты и увеличить корреляцию между техническими репликами.

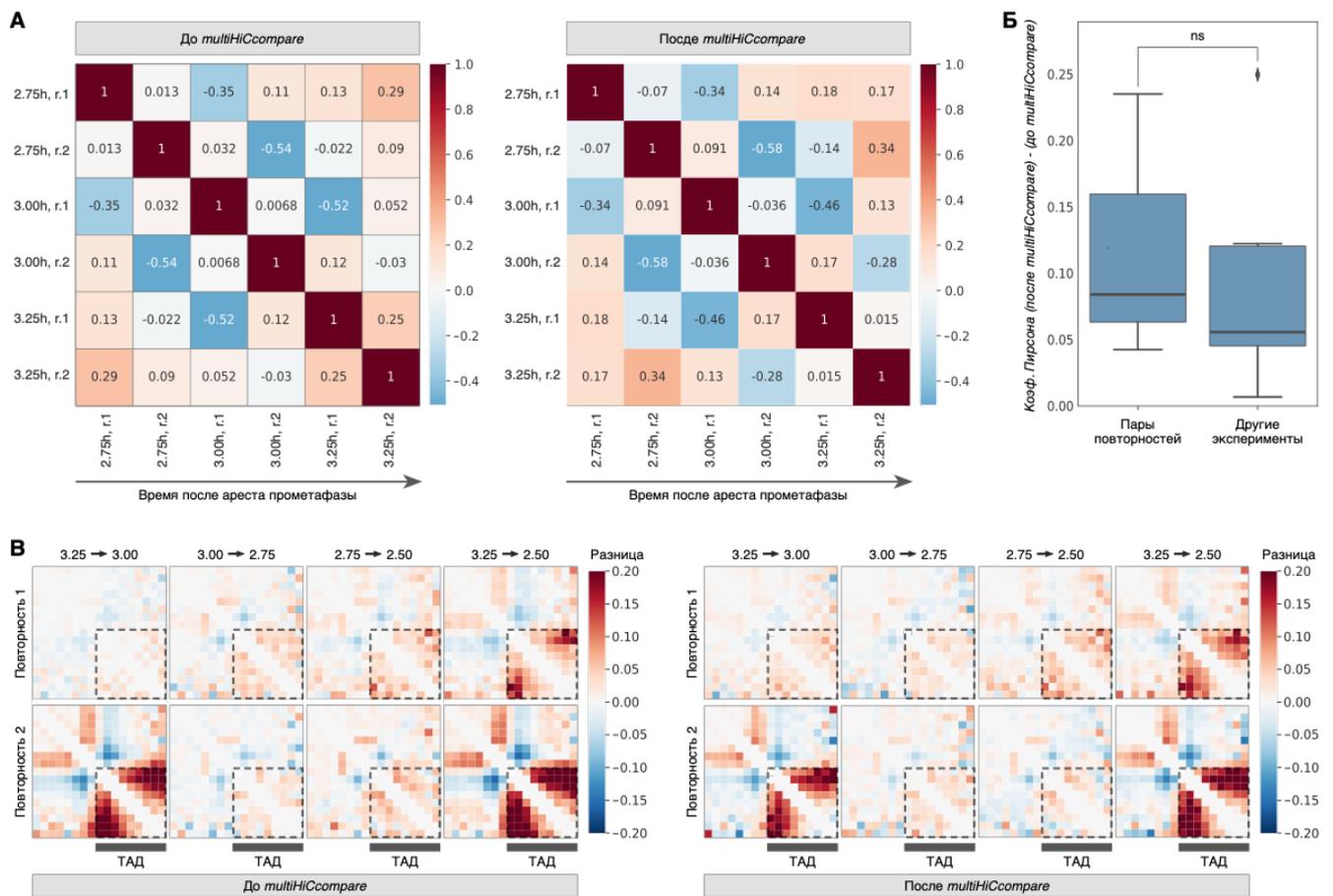


Рисунок 2.9 — Проверка подхода HiConfidence на внешнем наборе данных Hi-C. (А) Корреляции карт \log_2FC Hi-C, представляющих собой разность между соседними временными точками, до (левая панель) и после (правой панели) HiConfidence. Цвета представляют собой коэффициенты корреляции Пирсона. (Б) Корреляционные изменения после HiConfidence между парами повторностей и между другими случайными парами экспериментов. (В) Средняя разница в плотности ТАДов между временными точками до (левая панель) и после (правая панель) HiConfidence. Цвета представляют собой разницу между картами Hi-C между двумя временными точками. Пунктирные линии выделяют области с заметными улучшениями после HiConfidence. Адаптировано из [128].

Таким образом, подход HiConfidence является универсальным, может быть успешно применен к анализу данных Hi-C в нескольких (более, чем двух) временных точках, и в дальнейшем может использоваться для анализа других данных Hi-C, в том числе для клеток мозга человека.

2.2.3 Метод аннотации ТАДов на основе обогащения эпигенетическими метками

Другой важной с методической точки зрения задачей является аккуратная разметка ТАДов в картах Hi-C. Существует множество инструментов для разметки ТАДов в картах Hi-C [129; 131–136]. Однако алгоритмы предсказания ТАДов имеют один или несколько параметров, от которых зависит получающаяся в результате разметка ТАДов. От выбора параметров прежде всего зависит характерный размер ТАДов в разметке. Проблема подбора параметров является чрезвычайно важной, т.к. параметры приходится подбирать вручную, что исключает повторяемость научных исследований: нет общепринятых правил подбора этих параметров, и каждая научная группа использует в своих исследованиях разные. Алгоритмы, решающие задачу поиска ТАДов, не предоставляют универсальных методов их определения, а скорее является эвристикой, наилучшим образом воспроизводящей наблюдения исследователей. Наиболее часто используемый критерий, по которому можно отличить ТАД по его изначальному определению – это визуальная оценка карты. Понятно, что такой критерий не является строгим, и определение ТАДов сильно зависит от того, на каком масштабе пытаться их искать. ТАДы имеют иерархическую вложенную структуру, и при желании можно найти на одной карте и ТАДы размером 1 млн. п.о., и более мелкие, альтернативные, размером 100 тыс. п.о. Поэтому мы разработали универсальный и полностью автоматизированный подход к подбору параметров для алгоритмов разметки ТАДов - OptimalTAD [137].

Подход OptimalTAD основан на использовании комбинации Hi-C и эпигенетических данных (ChIP-seq, ChIP-chip) и поэтому наиболее подходит для исследования организмов, у которых наблюдается высокая корреляция между хромосомными контактами и эпигенетическими профилями. Предполагая, что наиболее выраженное обогащение соответствует лучшей аннотации границ ТАДов, наш алгоритм оптимизирует параметры поиска ТАДов, максимизируя разницу в уровнях меток хроматина между ТАДами и их границами. Используя этот алгоритм, мы аннотировали ТАДы в нескольких общедоступных наборах данных Hi-C и определили набор эпигенетических меток, которые лучше всего подходят для поиска ТАДов. Алгоритм использует программу Арматус [129]

в качестве базовой для предсказания набора ТАДов и оптимизирует ее внутренний параметр γ , определяющий средний размер предсказанных доменов. Принцип работы алгоритма следующий (рис. 2.10):

Шаг 1: Для каждого значения γ , лежащего в диапазоне от 0 до выбранного максимального значения, с фиксированным шагом Арматус последовательно предсказывает разметку ТАДов для входных Hi-C карт. Помимо самих ТАДов, каждая полученная разметка включает в себя еще и регионы, находящиеся вне топологических доменов и называемые меж-ТАДами.

Шаг 2: Затем для каждого бина, находящегося внутри или вне ТАДа в разметке, вычисляется геномное расстояние в тыс. п.о. до ближайшей к нему границы домена. После чего вычисленным расстояниям ставятся в соответствие Z-трансформированные значения уровня ацетилирования, полученные из эпигенетических данных. В результате для каждой разметки строится кривая среднего профиля ацетилирования вокруг границ ТАДов.

Шаг 3: Максимизируя разницу между средними значениями уровня ацетилирования гистонов ТАДов и меж-ТАДов, алгоритм определяет оптимальное значение параметра γ . Разметка, соответствующая выбранной γ , также считается алгоритмом оптимальной.

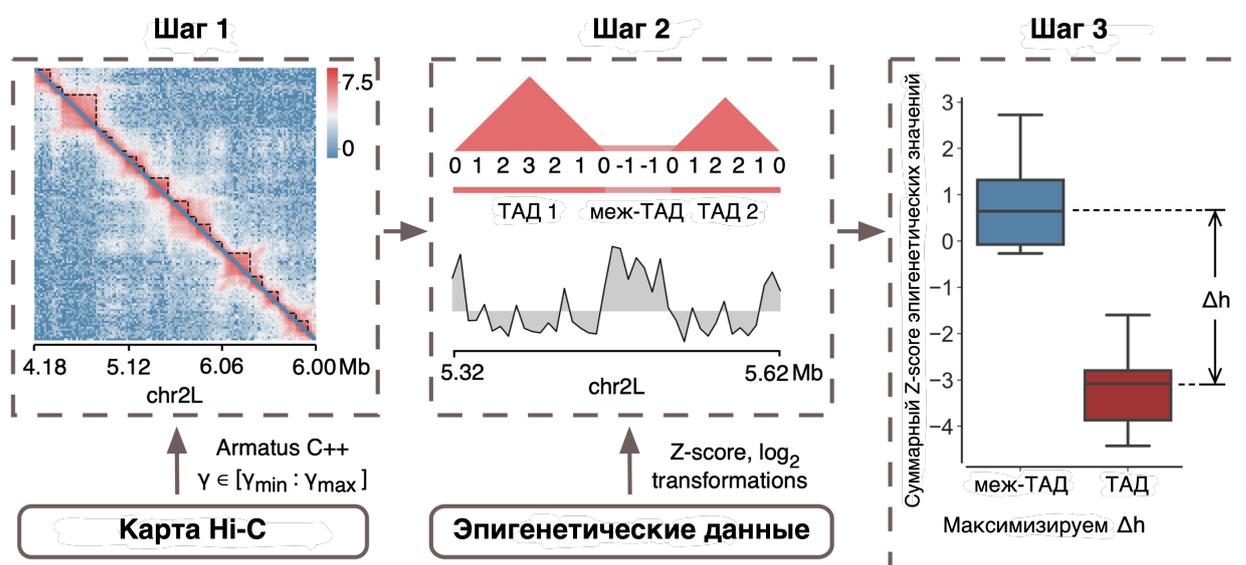


Рисунок 2.10 — Общая схема работы алгоритма optimalTAD. Адаптировано из [137].

Типичная матрица Hi-C содержит ряд строк и столбцов, полностью состоящих из пропущенных значений. Такие пустые строки появляются в плохо

картированных областях генома и областях, богатых повторяющимися последовательностями ДНК. Такие пустые строки затрудняют идентификацию ТАДов. `optimalTAD` заполняет пустые ячейки внутри таких строк на основе соседних ячеек, используя метод линейной регрессии.

Для ускорения этапа поиска ТАДов, мы разработали параллельную реализацию алгоритма `Armatus`, которая поддерживает >1 вычислительное ядро. Параллельная версия `Armatus` разбивает процедуру сегментации доменов на N потоков, где N равно количеству ядер, доступных на машине. Таким образом, достигается максимальная вычислительная эффективность при использовании ядер.

Для проверки алгоритма использовались три общедоступных набора данных [138–140]. Мы использовали карты Hi-C с разрешением 20 тыс. п.о. из набора данных Uljanov [138] и набора данных Li [139], а также карты Hi-C с разрешением 16 тыс. п.о. из набора данных Ing-Simmons [140]. Значения частоты взаимодействия Hi-C были масштабированы от $x(\min) = 1$ до $x(\max) = 1024$, а затем преобразованы в \log_2 шкалу. Обработанные профили ChIP-seq также были преобразованы в \log_2 шкалу и стандартизированы с использованием Z-score. Все вычисления `optimalTAD` были выполнены с параметром размера шага $s = 0,01$ с использованием 4 ядер в MacOS Catalina (Intel Core i5).

Чтобы проверить способность нашего алгоритма идентифицировать наборы ТАДов в различающихся условиях, мы протестировали `optimalTAD` на данных Hi-C и ChIP-seq из [138] в контрольных клетках S2 *Drosophila* и при нокдауне ламина Dm0 (Lam-KD). `optimalTAD` выдал более низкие оптимальные значения γ для контрольных повторностей (1,65 и 1,29) по сравнению с повторностями Lam-KD (2,15 и 1,73). Кроме того, обе повторности Lam-KD продемонстрировали специфическую форму кривой амплитуды ацетилирования (рис. 2.11А), которая явно отличалась от кривых для обеих контрольных повторностей. Такие разные формы кривых открывают новые возможности для кластерного анализа образцов Hi-C, поскольку кривые амплитуды ацетилирования отражают эпигенетические свойства ТАДов на разных иерархических уровнях (рис. 2.11Б).

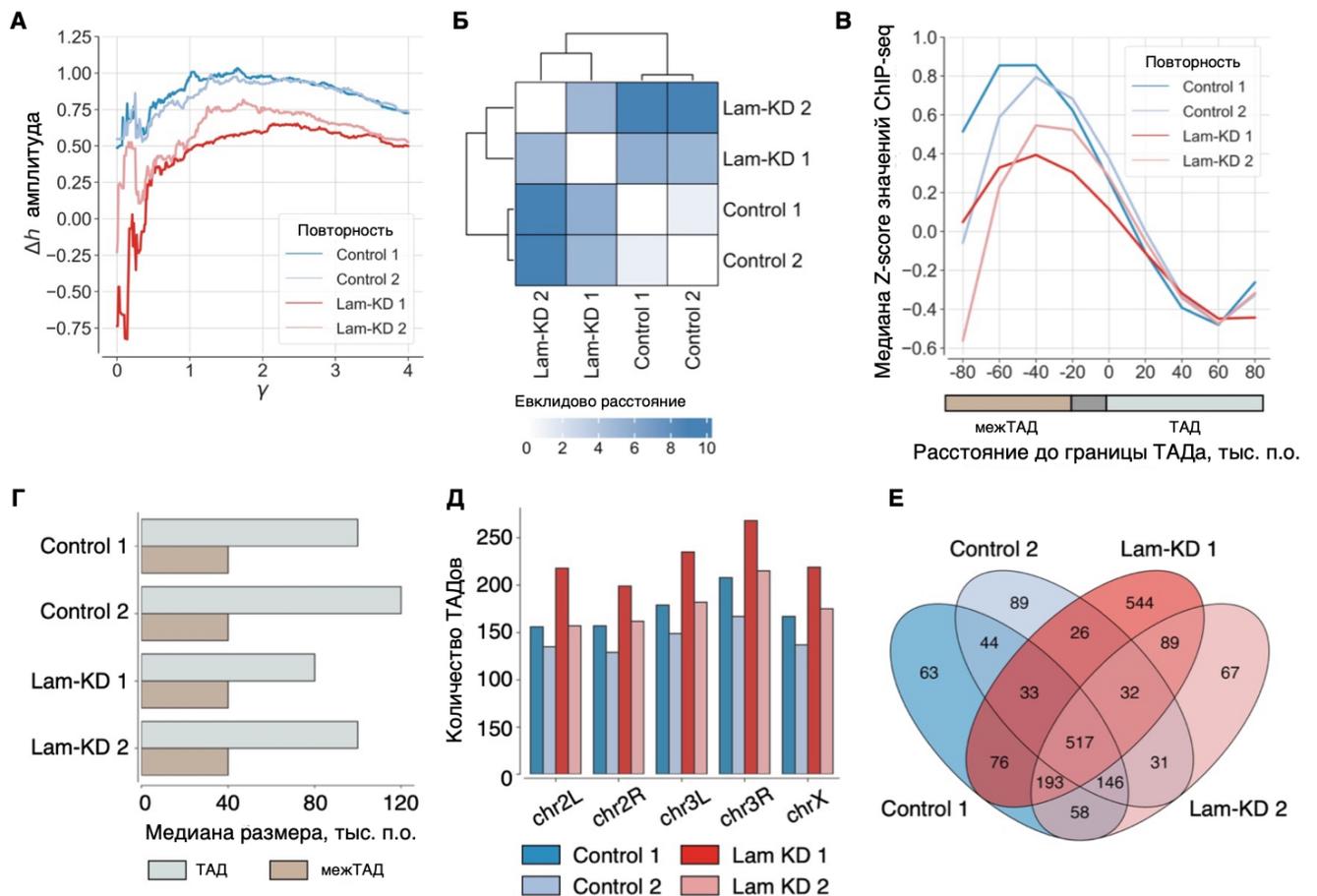


Рисунок 2.11 — Проверка алгоритма на данных Uljanov. (А) Амплитудные кривые, показывающие динамику разницы Δh между средними значениями ChIP-seq внутри и между ТАДами по значениям γ . (Б) Тепловая карта, иллюстрирующая иерархическую кластеризацию экспериментов на основе меры евклидова расстояния. (В) Амплитуда уровней ацетилирования гистонов в наборе ТАДов, предсказанном алгоритмом optimalTAD в наборе данных Uljanov. (Г) Сравнение медианных размеров ТАДов и межТАДов, полученных для экспериментальных образцов. (Д) Распределение числа ТАДов, предсказанных optimalTAD, в 2L, 2R, 3L, 3R и X хромосомах. (Е) Диаграмма Венна, показывающая количество границ ТАДов, общих для экспериментальных повторностей. Адаптировано из [137].

Далее, мы сравнили профили ацетилирования вблизи границ ТАДов, полученные для предсказанных оптимальных значений γ в каждой повторности. Как и ожидалось, мы обнаружили резкую разницу в уровнях ацетилирования гистона H3 между ТАДами и межТАДами (рис. 2.11В). В соответствии с [138], обогащение ацетилирования гистонов между ТАДами уменьшилось в Lam-KD. Медианный размер предсказанных ТАДов варьировал от 80 до 120 тыс. п.о.

(80-100 тыс. п.о. для повторностей Lam-KD и 100-120 тыс. п.о. для контрольных повторностей, что согласуется с их оптимальными значениями γ), в то время как медианный размер межТАДных интервалов оставался неизменным: 40 тыс. п.о. (рис. 2.11Г). Эти наблюдения согласуются со свойствами ТАДов у *Drosophila*, о которых сообщалось в предыдущих исследованиях [43; 59].

Учитывая разницу в полученных кривых амплитуды Δh и медианных размерах ТАДов между контрольной и Lam-KD репликами, мы дополнительно охарактеризовали состав предсказанных наборов ТАДов. В то время как для хромосомы 4 не было обнаружено ТАДов при различных разрешениях, другие хромосомы содержали 135-242 топологических домена, с максимальным числом ТАДов для повторности Lam-KD-1 в хромосоме 3R (рис. 2.11Д). Как и ожидалось, предсказанные ТАДы были похожи в разных экспериментальных условиях, так что 517 границ (из общего числа обнаруженных 2008 границ) были общими между всеми четырьмя повторностями (рис. 2.11Е). Остальные 193 границ ТАДов были предсказаны во всех повторностях, кроме Control-2. Однако некоторые границы ТАДов зависели от конкретных условий. В частности, мы обнаружили 89 границ, которые были специфичны для контролей, и 44 границы, появляющиеся только в образцах Lam-KD. Таким образом, подход optimalTAD может быть успешно применен для идентификации дифференциальных структур ТАДов.

Кроме того, мы хотели определить архитектурные белки и хроматиновые метки, наиболее полезные для предсказания оптимальных наборов ТАДов. Мы выбрали профили ChIP-Seq архитектурных белков (CTCF, CP190, RAD21, BEAF-32, TFIIC, Z4, L(3)mbt, Chromator, CBP, CapH2, RNAPII) вместе с профилями меток хроматина (H3K4me3, H3K9me2) и профилями белков группы Polycomb (PC-RJ, PC-VP) из набора данных Li. В соответствии с предыдущими исследованиями предикторов границ ТАДов у *Drosophila* [43; 59; 141], наш анализ показал, что RNAPII, BEAF32 и H3K4me3 были высоко обогащены вблизи предсказанных границ ТАДов (Таблица 1) и обладали максимальным значением Δh (1,692, 1,532 и 1,527 соответственно). Напротив, CTCF был слабо обогащен вблизи границ ТАДов и, таким образом, был гораздо менее пригоден для идентификации ТАДов, что подтверждает его ограниченную роль в формировании ТАДов у мух [119; 142].

Таблица 1 — Предсказания `optimalTAD`, полученные для датасета Li с заполнением и без заполнения пропущенных значений в данных Hi-C.

Повторность	Δh_{imp}	γ_{imp}	Δh_{unimp}	γ_{unimp}
BEAF-32 Rep. 2	1.69	1.7	1.36	2.66
Chromator Rep. 1	1.68	1.16	1.23	2.89
H3K4me3 Rep. 2	1.53	2.11	1.14	2.67
RNAPII Rep. 1	1.53	2.1	1.22	2.65
CapH2 Rep. 1	1.48	1.26	1.09	2.66
CBP Rep. 3	1.29	1.81	0.89	3.54
H3K4me1 Rep. 2	1.17	1.74	0.72	2.87
CP190 Rep. 2	1.16	2.99	0.9	2.83
L(3)mbt Rep. 1	1.14	1.57	0.78	2.49
Rad21 Rep. 1	1.02	1.01	0.53	2.87
Rad21 Rep. 3	1	1.34	0.6	2.87
Rad21 Rep. 2	0.86	1.57	0.41	3.02
Fs1h.L Rep. 1	0.84	1.14	0.48	2.52
Z4 Rep. 1	0.79	1.16	0.39	2.53
CapH2 Rep. 2	0.77	1.07	0.36	2.47
TFIIIC Rep. 2	0.77	1.12	0.51	2.47
CBP Rep. 2	0.45	1.01	0.13	2.53
H3K9me2 Rep. 1	0.44	1.55	0.41	2.38
Pc-RJ Rep. 1	0.38	1.87	0.24	3.56
Pc-VP Rep. 1	0.34	1.64	0.19	3.57
TFIIIC Rep. 1	0.16	1.42	0.04	3.8
Chromator Rep. 2	0.11	1.07	0.01	4
CBP Rep. 1	-0.13	3.73	-0.17	4
CTCF Rep. 2	-0.46	3.9	-0.46	1.01

Чтобы проиллюстрировать универсальность `optimalTAD`, мы протестировали его на наборе данных Ing-Simmons, включающем карту Hi-C Tollrm9/rm10 разрешением 16 КБ и две реплики H3K27ac ChIP-seq. Мы получили два оптимальных набора ТАДов с одинаковым средним размером 64 КБ, где около 92% границ были общими для двух наборов доменов (рис. 2.12А-Б). Амплитудные

кривые также были очень похожи (рис. 2.12В) с коэффициентом корреляции Пирсона = 0,9942 и RMSD = 0,0018, что подчеркивает воспроизводимость предсказаний `optimalTAD` между биологическими повторностями.

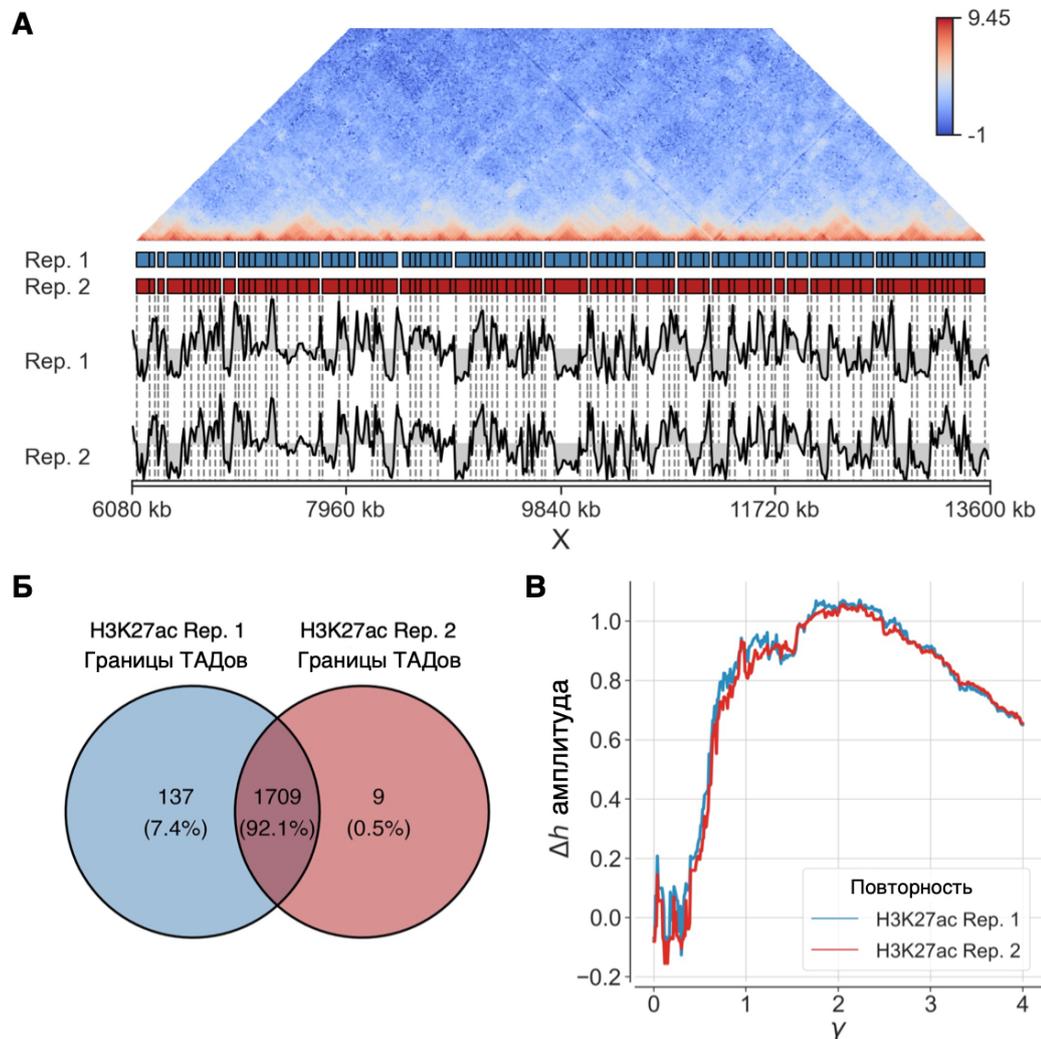


Рисунок 2.12 — Проверка алгоритма с использованием данных мутантных эмбрионов `Tollrm9/rm10`. (А) Карта Hi-C с границами ТАДов, предсказанными для H3K27ac Rep. 1 (окрашен синим цветом) и H3K27ac Rep. 2 (окрашен красным) в геномном интервале chrX: 6 080 000–9 000 000. (Б) Диаграмма Венна, иллюстрирующая общие границы между ТАДами, предсказанными в двух повторностях H3K27ac (синий кружок соответствует Rep. 1, а красный кружок показывает домены Rep. 2). (В) Амплитудные кривые, показывающие изменения Δh в зависимости от значений γ для повторностей Rep. 1 и Rep. 2 профилей H3K27ac ChIP-seq. Адаптировано из [137].

Наконец, мы сравнили `optimalTAD` с несколькими ранее опубликованными алгоритмами поиска ТАДов, такими как `TopDom` [132], `DomainCaller` [135],

TADtree [143], OnTAD [144], ClusterTAD [134], Arrowhead [40], Caspian [145], HiCseg [131]. Все алгоритмы запускались на данных Hi-C с разрешением 20 тыс. п.о. из набора данных Li. Три протестированных инструмента TADtree, OnTAD и Arrowhead выдали несколько наборов ТАДов и, следовательно, не могли быть сравнены напрямую с предсказанным нашим алгоритмом набором ТАДов. Два других инструмента DomainCaller (TADlib) и Arrowhead выдали по одному набору ТАДов, но без областей между ТАДами, характерными для генома *Drosophila*. Поэтому мы перешли к трем оставшимся алгоритмам TopDom, ClusterTAD и Caspian, которые были запущены с параметрами по умолчанию. Для анализа TopDom мы использовали пакет R, доступный по адресу <https://github.com/jasminezhoulab/TopDom> (версия 0.0.2). Программа запускалась с "window.size- 10. Java-версия алгоритма ClusterTAD (версия 1.0) была загружена с <https://github.com/BDM-Lab/ClusterTAD>. Наборы ТАДов с лучшим показателем качества считались оптимальными и использовались для оценки Δh . Для идентификации ТАДов с помощью инструмента Caspian мы использовали все три доступные метрики ("Евклидова "Манхэттенская" и "Чебышёвская"), реализованные в этом алгоритме для попарного вычисления расстояния между бинами. Алгоритм был загружен с <https://gitee.com/ghaiyan/caspian> с указанием параметров по умолчанию.

Чтобы оценить эффективность каждого метода поиска ТАДов, мы рассчитали значение амплитуды Δh , используя профиль BEAF-32 ChIP-seq. Среди всех этих алгоритмов, optimalTAD показал наибольшее значение Δh , превзойдя другие алгоритмы более чем в 1,7 раза (рис. 2.13А). Соответственно, кривая ацетилирования, предсказанная с использованием optimalTAD, показала наиболее резкое различие в уровнях ацетилирования между ТАДами и межТАДами по сравнению с другими алгоритмами (рис. 2.13Б-Д).

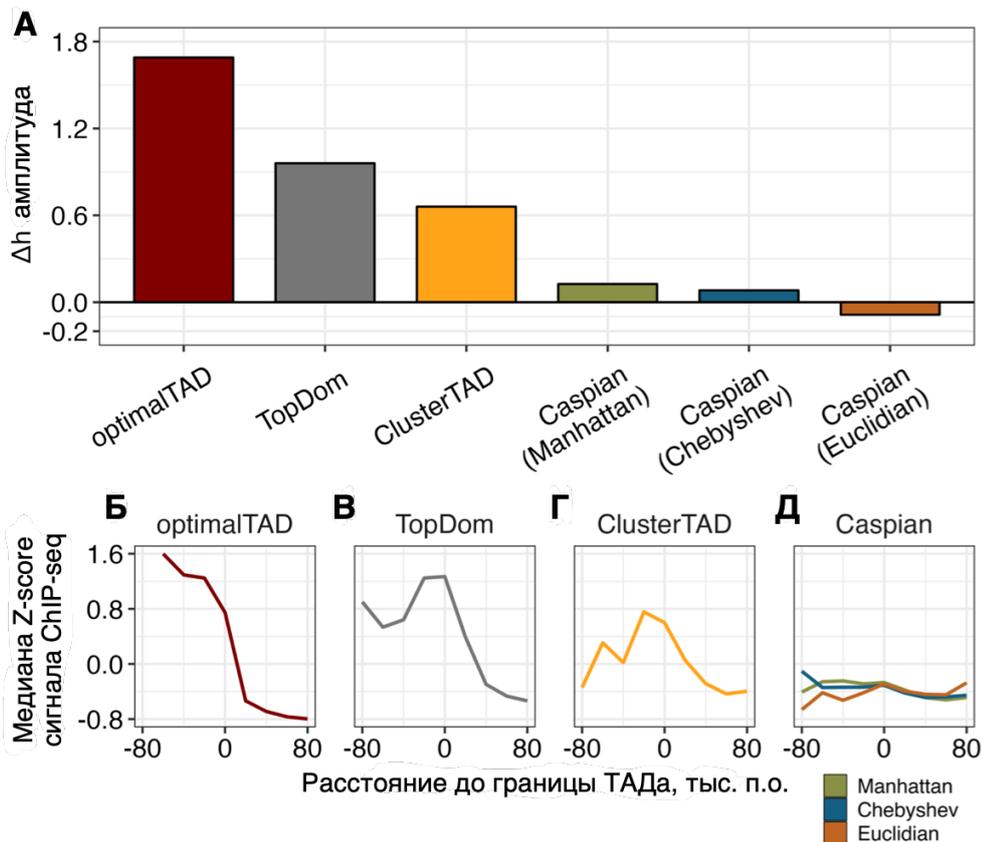


Рисунок 2.13 — Сравнение optimalTAD с другими инструментами поиска ТАДов. (А) Гистограмма, показывающая значения амплитуды Δh , рассчитанные с использованием методов OptimalTAD, TopDom, ClusterTAD и Caspian (слева направо) для карты Hi-C из Li. Анализ с помощью инструмента Caspian проводился в трех возможных режимах («Манхэттенский», «Чебышёвский», «Евклидовый»). (Б-Д) Сравнение кривых ацетилирования, полученных с помощью инструментов поиска ТАДов, выбранных для анализа. Адаптировано из [137].

Таким образом, optimalTAD производит оптимизированную идентификацию топологических доменов и может успешно применяться к данным Hi-C, полученным из различных источников и для различных организмов, в том числе и для клеток мозга человека. В то время как другие алгоритмы поиска ТАДов ограничены использованием только структурной информации о контактных частотах и предсказанием границ на основе заранее определенного размера ТАДов, optimalTAD, кроме того, использует функциональную информацию о хроматиновых метках и архитектурных белках для аннотации ТАДов как структурных и функциональных единиц организации хроматина. Программный код доступен в репозитории библиотеки optimalTAD по адре-

су <https://github.com/cosmoskaluga/optimalTAD> с подробной документацией <https://optimaltad.readthedocs.io/en/latest/>.

2.2.4 Применение методов машинного обучения для изучения особенностей организации хроматина

Методы машинного обучения могут применяться для решения технических проблем в данных Hi-C, в частности для повышения разрешения карт Hi-C. Разработка вычислительных методов, которые могут повысить качество данных, позволяет преодолеть ограничения, связанные с высокими затратами на эксперименты. Например, существуют инструменты, позволяющие улучшить разрешение Hi-C карт: HiCPlus [146], Boost-HiC [147] и другие. Однако методы машинного обучения могут применяться и для иных задач, связанных с анализом данных Hi-C. Так, в последнее время появляются методы для предсказания хромосомных контактов. Эти методы представляют интерес, так как быстрое предсказание Hi-C карт позволяет проводить эксперименты по *in silico* мутагенезу. Помимо этого, анализ и интерпретация обученной предсказательной модели может открыть до сих пор неизвестные механизмы формирования структуры хроматина.

Методы предсказания Hi-C карт могут быть основаны на эпигенетических данных, данных о последовательности ДНК или комбинации этих видов данных. Недостаток методов, использующих только эпигенетические данные, заключается в том, что такие методы не могут идентифицировать регионы и паттерны ДНК, отвечающие за формирование структуры хроматина. Методы, которые используют комбинацию эпигенетических данных и данных о последовательности ДНК обычно опираются на данные о связывании на ДНК белка CTCF, данные об экспрессии генов [148], или данные DNase-Seq [149]. Также, эти методы могут использовать комбинацию различных эпигенетических данных. Другой подход был представлен в работе [150], где специально обученная нейронная сеть предсказывала эпигенетические данные из последовательности ДНК, а затем другая нейронная сеть использовала эти данные для предсказания трехмерной структуры хроматина.

Подход к предсказанию, основанный лишь на последовательности ДНК, был представлен в [151]. В этой работе была использована особая архитектура сверточной нейронной сети - Akita, которая имеет скрытый слой, переводящий входную последовательность ДНК в двумерные матрицы. Эта трансформация оказалась важной составляющей успешного изучения нейронной сетью свойств последовательности ДНК, влияющих на трехмерную структуру. В результате, исследователям удалось предсказать трехмерную структуру ДНК для человека, основываясь лишь на последовательности ДНК. Последующий анализ предсказаний продемонстрировал важность связывания инсуляторного белка CTCF с определенной последовательностью мотива ДНК для позиционирования границ топологически ассоциированных доменов (ТАДов).

Несмотря на то, что применение сверточных нейронных сетей продемонстрировало высокую эффективность для обнаружения значимых геномных мотивов [151], в настоящее время нет реализации сверточных нейронных сетей, универсальной для предсказания взаимодействий Hi-C во всех организмах, т.к. большинство методов реализованы специально для хроматина млекопитающих (человек и мышь).

Другим возможным направлением применения машинного обучения к Hi-C данным является повышение разрешения карт Hi-C с использованием данных о последовательности генома. Разработка вычислительных методов, которые могут повысить качество данных, позволяет преодолеть ограничения, связанные с высокими затратами на эксперименты. Существуют инструменты, позволяющие улучшить разрешение Hi-C карт: HiCPlus [146], Boost-HiC [147] и другие. Алгоритм HiCPlus [146] использует глубокие сверточные нейронные сети для увеличения разрешения Hi-C карты, то есть из карты меньшего разрешения предсказывается карта более высокого разрешения. Метод Boost-HiC [147] использует алгоритмический подход для предсказания паттернов в Hi-C картах. Например, этот метод может предсказать хромосомные компартменты имея лишь карту Hi-C низкого разрешения. Подход Akita [151] - это сверточная сеть, которая преобразует последовательность ДНК в карту Hi-C (изображение) и изначально создавалась для задачи предсказания карты Hi-C по геномной последовательности, но может также использовать дополнительную информацию о базовой последовательности генома для дальнейшего улучшения разрешения карты Hi-C.

В будущем представляется перспективным объединение данных последовательности генома и транскриптома (RNA-seq) для предсказания карт Hi-C. Кроме того, можно объединять данные о последовательности генома, транскриптома, а также открытости хроматина (ATAC-seq) для предсказания карт Hi-C. Наконец, представляет интерес изучение закономерностей работы моделей, с целью извлечения новой биологической информации о принципах устройства хроматина в пространстве.

Мы применяли методы машинного обучения для определения факторов, отвечающих за формирование ТАДов у дрозофилы, ещё в 2016 году [119]: использовали модель логистической регрессии для решения задачи классификации (рис. 2.14). Вход модели представлял собой набор сигналов ChIP-chip для области генома, а выход - бинарное значение, указывающее, была ли область расположена на границе или внутри ТАДа. Модель, основанная на распределениях H3K4me3 и тотальной РНК, предсказывала намного лучше, чем модель, основанная на распределениях dCTCF и Su(Hw) ($AUC = 0,79$ и $0,63$ соответственно). При этом использование dCTCF и Su(Hw) вместе с активными метками хроматина не улучшало качество предсказания ($AUC = 0,79$). Поскольку профили большинства хроматиновых меток сильно скоррелированы, добавление признаков, связанных с активным хроматином, также существенно не улучшало качество предсказания (рис. 2.14). Таким образом, эти результаты позволяют предположить, что модификации гистонов, характерные для активного хроматина и/или активной транскрипции, а не связывание белков-инсуляторов dCTCF и Su(Hw), являются основной характерной чертой границ ТАДов и областей между ТАДами в клетках дрозофилы. Позднее, другая группа исследователей продемонстрировала эффективность методов градиентного бустинга и регрессии лассо для решения той же задачи [152].

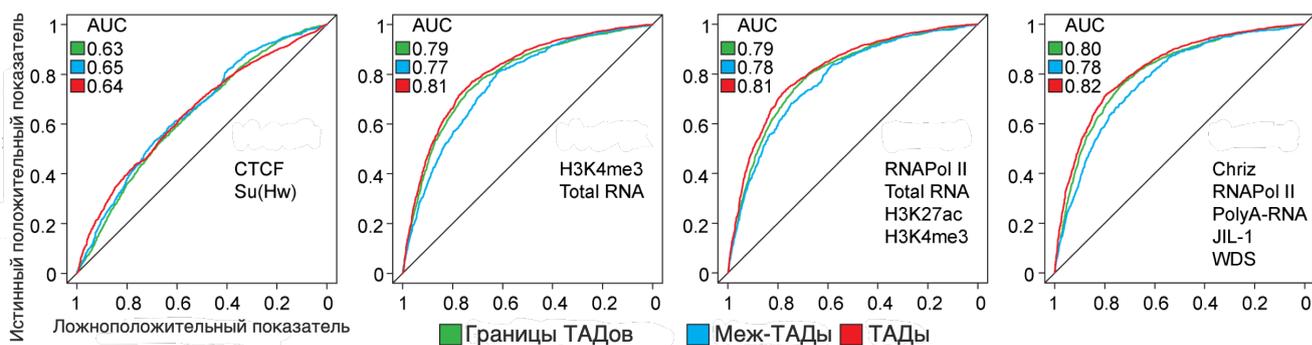


Рисунок 2.14 — Предсказание ТАДов, межТАДов и границ ТАДов в клетках S2 дрозофилы с использованием моделей логистической регрессии, основанных на геномном распределении активных меток хроматина или архитектурных белков dCTCF и Su(Hw). Показаны кривые ROC и значения AUC (площадь под кривой). Адаптировано из [119].

В вышедшей недавно работе [153], мы применили методы машинного обучения для предсказания паттернов сворачивания ДНК у дрозофилы на основе одномерных эпигенетических данных (хроматиновых меток) в трех клеточных линиях. Цель этого анализа состояла в том, чтобы предсказать профиль ТАДов по эпигенетическим маркерам и определить, какие из них являются наиболее значимыми для предсказания профиля ТАДов. Поэтому мы использовали модели линейной регрессии (LR) с четырьмя типами регуляризации, регрессионные модели с градиентным бустингом (GB) и рекуррентные нейронные сети (RNN) в качестве инструментов для изучения характеристик сворачивания хроматина, связанных с ТАДами, на основе данных ChIP-seq и ChIP-chip. Архитектура RNN с двунаправленной долгосрочной краткосрочной памятью показала наилучшие результаты и выявила биологически значимые маркеры (рис. 2.15).

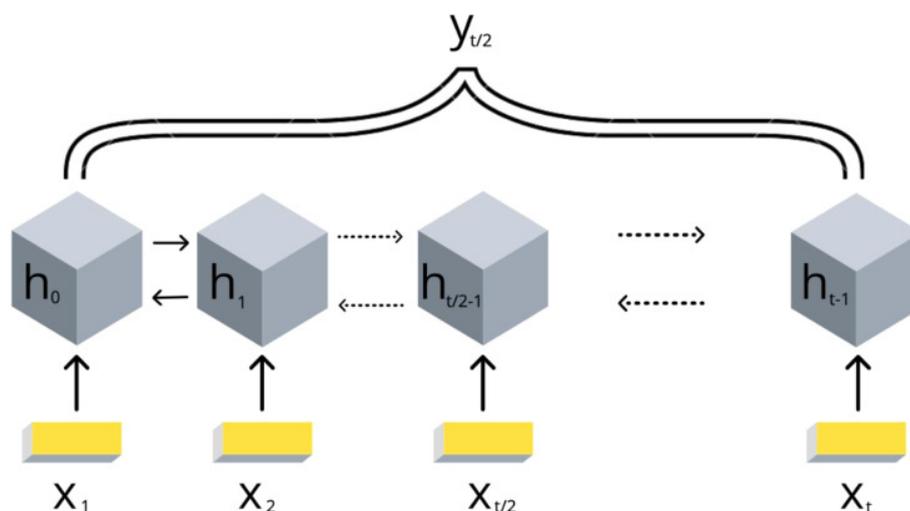


Рисунок 2.15 — Схема реализованной двунаправленной рекуррентной нейронной сети LSTM с одним выходом. Значения x_1, \dots, x_t представляют собой участки ДНК с размером входного окна t , h_1, \dots, h_t являются скрытыми состояниями модели RNN, $y_{t/2}$ представляет соответствующее значение параметра γ целевого значения средней ячейки $x_{t/2}$. Обратите внимание, что каждая ячейка x_i характеризуется вектором хроматиновых меток данных ChIP-chip. Адаптировано из [153].

Распределение белка Chriz (Chromator) и гистоновые модификации H3K4me3, H3K27ac оказались наиболее информативными признаками для предсказания ТАДов (рис. 2.16), что указывает на потенциальную важность этих факторов для формирования ТАДов. Полученный результат хорошо согласуется с существующей литературой. Подход реализован в виде программного обеспечения Hi-ChIP-ML, код которого доступен публично: <https://github.com/MichalRozenwald/Hi-ChIP-ML>. Разработанный нами подход может быть адаптирован к любому аналогичному набору биологических данных о характеристиках хроматина в любом организме и типе клеток, в том числе для клеток мозга человека.

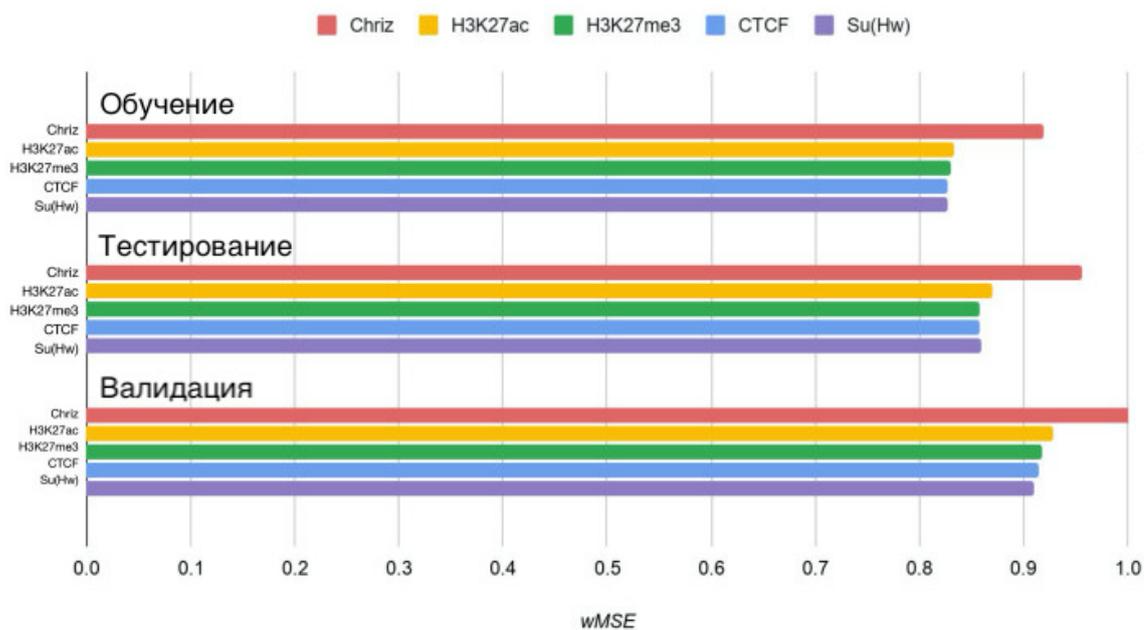


Рисунок 2.16 — Взвешенная MSE с использованием четырех из пяти хроматиновых маркеров в качестве входных данных biLSTM RNN. Каждый цвет соответствует объекту, который был исключен из входных данных. Обратите внимание, что на модель больше всего влияет исключение фактора Chriz. Адаптировано из [153].

Методы машинного обучения могут применяться не только для задач предсказания хромосомных контактов, но и для разметки структурных единиц в хроматине - ТАДов и хроматиновых петель. Хотя существует множество инструментов для разметки ТАДов в картах Hi-C [129; 131–136], подходов для разметки хроматиновых петель не так много (см. обзор в [116]). И к сожалению, ни один из них не является универсальным: каждый метод разметки хроматиновых петель разрабатывался для конкретного организма и, как правило, хорошо работает только на данных этого организма. Таким образом, разработка универсального подхода разметки петель является открытой задачей и оставляет простор для применения различных методов машинного обучения, в частности на основе нейронных сетей, а также моделей transfer learning с первичным обучением на картах человека, для которых существуют хорошие алгоритмы разметки, и дообучением на данных немодельных организмов и систем, для которых пока нет специальных, хорошо работающих алгоритмов разметки хроматиновых петель. Разработка такого универсального инструмента является задачей наших текущих исследований.

2.2.5 Разработка методов интеграции данных Hi-C с другими омиксными данными

Получение целостной картины структурной и функциональной организации хромосом невозможно без сопоставления данных Hi-C с другими омиксными данными: RNA-seq, ATAC-seq, ChIP-seq и другими.

Анализ данных RNA-seq

В нашей недавней работе [154] представлен наиболее современный подход к анализу данных RNA-seq. Вкратце, Необработанные данные секвенирования фильтруются с помощью `fastp`, а затем обрабатываются с помощью конвейера `nf-core/rnaseq`. Далее, адаптеры удаляются инструментом `Trim Galore!` и полученные прочтения картируются на референсный геном человека с помощью программы `STAR`. Затем экспрессия генов определяется количественно с помощью инструмента `Salmon`. Аннотация гена выполняется с использованием пакета `AnnotationDbi` R по аннотации `Ensembl`. Дифференциально экспрессирующиеся гены определяются с использованием программы `DESeq2` на языке программирования R после удаления генов с низкой экспрессией. Пример применения этого подхода к анализу данных RNA-seq для ткани мозга представлен в работе [154] и описан в разделе 3.1.1.

Особым случаем анализа данных RNA-seq является анализ биаллельной транскрипции, методику которого мы разработали в статье [155] на примере индивидуума с секвенированным геномом и транскриптомом. Имея список всех однонуклеотидных полиморфизмов (SNP) генома, мы рассортировали транскрипты из данных RNA-seq по аллелям. Всего мы обнаружили 700 SNP с 20-кратным покрытием геномного и транскриптомного секвенирования. Среди этих SNP, 234 SNP демонстрировали более чем 1,5-кратный сдвиг биаллельной экспрессии. Затем мы определили генетические изменения в регуляторных областях генов со сдвигом биаллельной экспрессии. В общей сложности было обнаружено 59 генов со сдвигом в биаллельной экспрессии и SNP в сайтах связывания

транскрипционных факторов (ТФ) в их регуляторных областях. Удивительно, но все SNP в сайтах связывания ТФ находились в интронах и экзонах. В числе этих генов был ген CSF1R - трансмембранный рецептор тирозинкиназы, модулирующий гомеостаз микроглии, нейрогенез и выживание нейронов в центральной нервной системе. Для этого гена коллеги из группы проф. Е.Б. Прохорчука (ФИЦ Биотехнологии РАН, Москва) валидировали наши биоинформатические наблюдения экспериментально (см. подробнее в работе [155]).

Отдельным направлением анализа данных RNA-seq является поиск событий альтернативного сплайсинга. В процессе сплайсинга образуется множество альтернативных транскриптов, специфичных для различных типов клеток и определяющих правильное развитие органов, однако известно всего несколько тканеспецифичных факторов, регулирующих этот процесс. В работе [156] мы показали, что белок RBM24 является одним из таких факторов, а именно участвует в регуляции сплайсинга в скелетной мускулатуре и в миокарде. Эксперименты *in vitro* подтвердили наши биоинформатические предсказания и показали, что наличие рекомбинантного белка RBM24 достаточно для включения экзона, специфичного для мышечных клеток, в ядерных экстрактах из немышечных клеток. Таким образом, RBM24 является одним из главных регуляторов альтернативного сплайсинга в мышечной ткани. Важно отметить, что RBM24 может регулировать липидный метаболизм [157], который мы подробно изучали для мозга человека (см. в разделе 3.3).

В работе [158] мы анализировали данные RNA-seq экспериментов в клетках HeLa после нокдауна U1C, который участвует в распознавании 5' сайта сплайсинга и в регуляции сплайсинга пре-мРНК. Мы показали новый механизм поддержания гомеостаза белков U1-70K и U1C, который необходим для правильной сборки и функционирования мРНК U1, участвующей в процессе сплайсинга. U1 нарушается при болезни Альцгеймера, что приводит к изменениям в сплайсинге [159]. Таким образом, U1 играет важную роль в поддержании нормального функционирования мозга человека. Альтернативный сплайсинг U1-70K пре-мРНК приводит к образованию двух изоформ – нормальной и неработающей, баланс которых определяется уровнем белка U1C. Мы подтвердили найденный механизм кросс-регуляции с помощью мутационного анализа в мини-гене, блокирования сайта сплайсинга *in vivo* и экспериментами по связыванию *in vitro*.

В работе [160], мы анализировали данные эксперимента iCLIP-Seq, который позволяет определить позиции связывания белка с РНК с точностью до нуклеотида, чтобы изучить функции белка hnRNP L на полногеномном уровне. Белок hnRNP L – это многофункциональный РНК-связывающий белок, который участвует в регуляции транскрипции, трансляции и стабильности РНК, который является глобальным регулятором альтернативного сплайсинга и связывается с СА повторами. Мы показали, что hnRNP L преимущественно связывается с интронами и 3'НТО. Кроме того, мы показали, что hnRNP L регулирует сплайсинг по-разному в зависимости от позиции связывания: подавляет сплайсинг, если связывается с интронами перед альтернативными экзонами, и активирует сплайсинг, если связывается после. Также мы показали, что hnRNP L связывается с мишенями микроРНК, таким образом подавляя связывание микроРНК с этими мишенями.

Важно отметить, что некоторая доля особых событий сплайсинга может определяться трехмерной организацией хроматина. Так, в работе [36] мы показали для образцов ткани мозга, что частота формирования химерных транскриптов из фрагментов двух разных хромосом выше среди пространственно близких областей генома. И некоторая доля этих химерных транскриптов, скорее всего, возникает посттранскрипционно в результате транс-сплайсинга.

Стоит отдельно обсудить особенности методологии анализа данных RNA-seq в контексте онкологических заболеваний, поскольку такой анализ имеет некоторые технические сложности. В коллаборации с группой Dr. Moshe Elkabetz в Ben-Gurion University of the Negev (Израиль), мы изучали особенности экспрессии и ее регуляции в образцах рака головы и шеи, которые по-разному отвечали на лечение препаратом Cetuximab в мышинных ксенотрансплантатах (Patient-Derived Xenografts, или PDX). Ксенотрансплантаты (PDX) представляют собой модели рака, в которых ткань или клетки из опухоли пациента имплантируются иммунодефицитной или гуманизированной мыши. Анализ таких данных RNA-seq является нетривиальным и представляет отдельный интерес, поскольку образцы PDX состоят из смеси тканей мыши и человека, и требуется разработка особых биоинформатических подходов для разделения прочтений, соответствующих разным видам. Поэтому в работах [161; 162] мы картировали прочтения RNA-seq для образцов PDX на конкатенированный геном человека и мыши, и уникально картированные ряды были разделены на

риды мыши и риды человека. Построенные нами графики многомерного шкалирования показали четкое разделение всех трех PDX на основе ридов человека, что позволило нам убедиться в достаточно высокой точности выполненного анализа и успешно определить дифференциально экспрессирующиеся гены.

Поскольку механизм различного ответа на лечение препаратом Cetuximab оказался явно регуляторным, мы решили провести более глубокое исследование возможной регуляции и связать найденные отличия в экспрессии генов с данными о положении энхансеров в геноме человека, данными Hi-C по частоте контактов хроматина, по положению ТАДов в геноме человека, и проверить частоту контактов энхансер-промотор для генов с изменениями экспрессии. Мы подготовили публичные данных Hi-C по частоте контактов хроматина в мозге здорового человека (Hi-C данные для рака головы и шеи не существуют, но показано, что карты Hi-C не сильно меняются от ткани к ткани, поэтому подход с использованием Hi-C данных из другой ткани является валидным и даже широко распространенным), а также данные об уровнях экспрессии генов и данные о положении энхансеров в геноме человека. Данные о положении энхансеров были взяты из базы GeneHancer, которая является одной из лучших на данный момент, поскольку содержит экспериментально подтвержденные по разным источникам энхансеры.

Далее, мы анализировали три 3 группы генов: (1) с повышенной экспрессией в раке, (2) с пониженной экспрессией в раке, (3) с не меняющейся экспрессией. В качестве контроля к группе 1 и 2 выбирались случайным образом гены из группы 3 с соответствующим размером выборки. Затем на основе данных из базы GeneHancer для групп генов выбирались энхансеры со значением Score не менее 200 (т.е. хорошо подтвержденные экспериментально). Расчет частот контактов по данным Hi-C проводился для взаимодействий ген-ген, ген-энхансер, энхансер-энхансер. Для взаимодействий ген-ген и энхансер-энхансер, выбирались только такие контакты, расстояние между которыми составляет не более 1.2 млн. п.о., поскольку нашей основной задачей было изучить связь ТАДов с заболеваниями. Именно поэтому мы сфокусировались на области не более 1.2 млн. п.о., которая содержит в себе большую часть контактов внутри ТАДов, и таким образом позволяет увидеть взаимодействия ген-ген и энхансер-энхансер, опосредованные структурой ТАДов. Стоит отметить, что прямая проверка предпочтительного расположения генов из групп (1) и (2) относительно гра-

ниц ТАДов не дала статистически значимых результатов (анализ выполнялся с помощью инструмента Stereogene [122]).

Результаты анализа представлены в Таблицах 2, 3 для групп генов (1) и (2) соответственно. Можно видеть, что в некоторых сравнениях нам удалось наблюдать статистически значимые отличия частот контактов по данным Hi-C. Однако, абсолютная разница медиан между реальной выборкой и контролем составляет менее 10% во всех сравнениях, вне зависимости от полученного р-значения. Таким образом, найденные эффекты являются слабыми и требуют дальнейшего, более детального изучения.

Таблица 2 — Частоты контактов по данным Hi-C для генов с повышенной экспрессией.

		ксенограф3		ксенограф18		ксенограф20	
		контроль	эксп	контроль	эксп	контроль	эксп
энхансер-ген	кол-во измерений	611	964	568	784	611	862
	медианы	1.1909	1.1365	1.1854	1.1280	1.1797	1.1036
	p-val mann-whitney	0.0086		0.014		0.0082	
ген-ген	кол-во измерений	7111	5061	5676	5182	8672	8121
	медианы	1.1174	1.1325	1.1139	1.0751	1.0789	1.0377
	p-val mann-whitney	0.0474		0.086		0.0266	
энх-энх	кол-во измерений	5316	5216	4049	4884	5114	6140
	медианы	1.1057	1.0903	1.0989	1.071	1.0825	1.0092
	p-val mann-whitney	0.4289		0.2354		0.0011	

Таблица 3 — Частоты контактов по данным Hi-C для генов с пониженной экспрессией.

		ксенограф3		ксенограф18		ксенограф20	
		контроль	эксп	контроль	эксп	контроль	эксп
энхансер-ген	кол-во измерений	637	491	527	549	647	962
	медианы	1.177	1.2609	1.2568	1.2476	1.1685	1.2042
	p-val mann-whitneyu	0.0511		0.0685		0.0334	
ген-ген	кол-во измерений	6055	9929	4775	9100	10005	13258
	медианы	1.1137	1.2211	1.1286	1.1598	1.0957	1.0828
	p-val mann-whitneyu	<e-9		0.1056		0.1185	
энх-энх	кол-во измерений	4466	7688	3391	7212	6724	11006
	медианы	1.1102	1.1852	1.1174	1.1227	1.1041	1.0587
	p-val mann-whitneyu	<e-04		0.2691		0.0003	

Анализ данных ATAC-seq

Для анализа доступного для транспозазы хроматина методом секвенирования (ATAC-seq), были использованы обработанные данные из статьи [163], посвященной исследованию открытого хроматина в клетках мозга человека, а также сортированных культурах клеток. Данные секвенирования ATAC-seq из статьи [164] мы обрабатывали самостоятельно. Данные были картированы на референсный геном человека с помощью программы samtools. Мы проверили их качество с помощью программы MultiQC и FastQC. Все образцы показали высокий процент картирования. Далее прочтения обрабатывались с помощью программного конвейера atac-seq nextflow pipeline с опцией narrow peak. Размеры вставок были проверены с помощью Picard. Пики были рассчитаны с помощью MACS2 с опцией узкого пика и начальным пороговым значением q-value, равным 0,05. Программа HOMER была использована для аннотирования пиков относительно известных геномных признаков. Дифференциальные пики были найдены с помощью пакета DESeq2. Главным критерием качества эксперимента ATAC-seq является график, показывающий количество прочтений при данном размере фрагмента. Другим важным критерием качества является профиль покрытия прочтениями вокруг начала генов. Пик покрытия перед началом гена характеризует открытый хроматин в области промотеров генов.

Методы ассоциации данных Hi-C с омиксными данными разных типов

Чтобы решить задачу ассоциации данных Hi-C с омиксными данными разных типов, мы для начала создали локальную базу таких данных. Для этого мы скачали публично доступные данные ChIP-Seq, ATAC-Seq и RNA-Seq как из наших статей [119], так и из публично доступных баз данных (ENCODE, modENCODE). Если исходные данные были в формате пиков или текстовых таблиц типа bed или bedgraph, мы перевели эти данные в формат bigwig с помощью инструментов UCSC.

Далее, мы разработали следующий метод ассоциации данных Hi-C с эпигенетическими разметками: мы посчитали среднее содержание эпигенетического сигнала (высоты пика или среднего покрытия) в бинах Hi-C карты вокруг каждой границы выборки с помощью инструмента `pybbi` <https://github.com/nvictus/pybbi>, который является оберткой над инструментами по работе с bigwig файлами UCSC браузера <https://github.com/ucscGenomeBrowser/kent>. Полученные значения были дополнительно Z-трансформированы для сопоставимости распределений, полученных для различных экспериментов. В результате, мы построили средний профиль значений RNA-seq, ChIP-seq, ATAC-seq вокруг границ различных элементов укладки хроматина (ТАДов или хроматиновых петель). Помимо самого профиля, мы отрисовываем также теплокарту обогащения эпигенетического сигнала вокруг каждой индивидуальной границы ТАДа (т.н. стек-графика, `stackup plot`). Цвет стек-графика обозначает степень обогащения, полученный график наглядно демонстрирует свойства границ каждого типа и позволяет оценивать как дисперсию сигнала, так и отсутствие нежелательных артефактов данных. Кроме того, мы строим среднее обогащение границ разного типа эпигенетической меткой как среднее значение центрального пика профиля. Этот подход позволяет выявить дифференциальное связывание эпигенетических факторов вокруг границ ТАДов, хроматиновых петель или других элементов организации хроматина.

Поскольку данные Hi-C для мозга человека ещё не были доступны, мы отработывали эту методику в работах [138; 165] на данных дрозофилы, геном которой в 10 раз меньше генома человека, что значительно уменьшает время расчетов. В первой работе [138] мы интегрировали данные Hi-C с данными RNA-seq и ChIP-seq как описано выше, чтобы изучить, как ядерная ламина влияет на глобальную архитектуру хроматина. Благодаря разработанной методике анализа, нам удалось показать, что разрушение ламины в клетках дрозофилы приводит к уплотнению хроматина и его отдалению от ядерной оболочки. Это увеличивает плотность хроматина во фракции ТАДов, обогащенных активным хроматином, и усиливает взаимодействия между активным и неактивным хроматином. Важно отметить, что, при разрушении ламины, ламина-ассоциированные ТАДЫ становятся более ацетилированными по гистону H3 и менее компактными, в то время как фоновая транскрипция подавляется. Эти биоинформатические наблюдения были подтверждены коллегами из группы Ю.Я.

Шевелева (Институт молекулярной генетики РАН, Москва) методом FISH. Стоит заметить, что позднее мы ещё возвращались к изучению влияния ламины на архитектуру хроматина в головном мозге, поскольку белок SIRT6, регуляторную роль которого в мозге мы исследовали в работах [154; 166] (см. ниже в разделах 3.1.4 и 3.1.1), активируется белком ламины LaminA, обеспечивающим локализацию SIRT6 в хроматине при обнаружении повреждения ДНК.

Во второй работе [165] мы сопоставляли данные Hi-C с данными RNA-seq и DamID - особым типом эксперимента, позволяющим картировать сайты связывания ДНК- или хроматин-связывающих белков. Мы сравнивали организацию хроматина в модели сперматогониев (SpG) и семенниках, заполненных в основном сперматоцитами (SpCs). Применение разработанных нами методов интеграции Hi-C с другими типами данных позволило показать, что инициация SpC-специфичной транскрипции коррелирует с отдалением промоторов от ламины и с локальной пространственной изоляцией прилегающих к ним областей. Однако такая изоляция не приводит к разделению неактивных TADов на дочерние TADы. Мы также выявили повышенную частоту контактов между SpC-специфичными генами в SpC, что подразумевает их организацию в фабрики транскрипции.

Таким образом, разработанная нами методика сопоставления данных Hi-C с другими омиксными данными позволяет пронаблюдать взаимосвязь между особенностями укладки хроматина и экспрессией генов, уровнем ацетилирования гистонов, связыванием транскрипционных факторов, и другими уровнями молекулярной организации клеток.

2.2.6 Разработка методов анализа липидного и метаболического состава мозга человека

В то время как методы RNA-seq, ChIP-seq, ATAC-seq широко применяются для изучения мозга человека, описание методов, оптимизированных для изучения липидного и метаболического состава мозга человека, в литературе отсутствует. Поэтому мы разработали методики экстракции липидов и масс-спектрометрического анализа, которые обеспечивают одновременное детекти-

рование максимального количества различных липидных классов и индивидуальных веществ в составе мозга человека. Также мы разработали подходы к биоинформатическому анализу полученных данных. Их применение позволяет создать целостную картину молекулярной организации головного мозга человека.

Для масс-спектрометрического анализа низкомолекулярных метаболитов использовали образцы весом 10–20 мг, вырезанные из блоков мозга толщиной 1–2 см, которые были заморожены в парах жидкого азота и хранились в герметично закрытой упаковке при температуре не выше $-70\text{ }^{\circ}\text{C}$. Вырезку образцов производили в замораживающем (крио) микротоме CM1950 (Leica Biosystems, Germany) при температуре $-15\text{ }^{\circ}\text{C}$. Перед вырезкой блоки переносили в криомикротом и оставляли на 30 мин для выравнивания температуры камеры и блока. Блок становился пластичным и не крошился при вырезке. Инструменты для вырезки дезинфицировали после каждого образца и охлаждали внутри криомикротомы. Криопробирки заранее охлаждали до $-20\text{ }^{\circ}\text{C}$ и регулярно переносили небольшими партиями в криомикротом. Для снижения вероятности биохимических изменений вырезку проводили в максимально короткое время. Охлажденным пинцетом вырезанный образец переносили в криопробирку, а затем помещали на $-70\text{ }^{\circ}\text{C}$. Перед экстракцией липидов контролировали вес образцов, при этом общее время взвешивания замороженного образца не должно занимать более 2 секунд. Для исключения нежелательных технических различий, вырезка для всех образцов должна проводиться одним и тем же нейроанатомом. С особой тщательностью унифицируется локализация мест вырезки образцов у различных индивидуумов (рис. 2.17).



Рисунок 2.17 — Схема пробоподготовки образцов мозга для масс-спектрометрического анализа липидов: замороженные блоки мозга (*a*), выбор участков для вырезки образцов в соответствии с атласом мозга (*б*), выдерживание в криомикротоме при температуре $-15\text{ }^{\circ}\text{C}$ (*в*), вырезка образцов (*г*), взвешивание (*д*). Адаптировано из [167].

Экстракцию липидов из образцов ткани проводили согласно процедуре, которая подробно описана в нашей работе [167] и позволяет детектировать наибольшее количество липидов мозга. Измерение концентраций липидов проводили с использованием нетаргетированной масс-спектрометрии в сочетании с высокоэффективной жидкостной хроматографией (ВЭЖХ-МС) в режимах детектирования отрицательно и положительно заряженных ионов. Такой подход позволяет установить интенсивности соединений, представляющих все основные классы липидов, присутствующие в мозге: PE, PC, PE O, PC O, PE P, PC P, PG, PS, PI, LPC, LPE, CAR, Cer, SM, HexCer, CE, SulfoHexCer, TAG, DAG, MAG и холестерол. Процедура ВЭЖХ-МС оптимизирована таким образом, чтобы обеспечивать оптимальный уровень детектирования липидов мозга.

Обработку результатов ВЭЖХ-МС измерений проводили с помощью биоинформатического анализа. Этапы анализа включали первичный контроль качества, устранение шума, фильтрацию, нормализацию и первичную аннотацию данных. После сбора данных с масс-спектрометра, сырые файлы формата «.d» автоматически калибровались внутренней калибровкой по контрольным массам и конвертировались в формат «.mzXML», используя модифицированный скрипт программного обеспечения DataAnalysis (Bruker, Версия 4.3). Далее, конвертированные файлы формата «.mzXML» обрабатывались пакетом XCMS (язык программирования R, версия 3.0.0) [168], который производит детектирование пиков, их выравнивание между образцами, а также заполнение пропущен-

ных при детектировании значений. Пропущенные и нулевые значения, которые не были заполнены автоматически на предыдущем шаге, заполнялись нормальным распределением с параметрами, посчитанными по следующему правилу: из распределения минимального по образцам значений пиков рассчитывается медиана и 16-й перцентиль, которые используются в качестве среднего и стандартного отклонения этого нормального распределения. Также, пары пиков с разницей в массе/заряд значениях m/z менее чем 10 ppm и времени удерживания менее 1 сек обозначались как ложные дубликаты алгоритма детекции пиков, и один пик из пары удалялся. Границы времени удерживания, за пределами которых пики в анализе не рассматривались, составляли от 1 до 18,3 минут. Для устранения признаков, источником которых являются экстракционные или прочие технические загрязнения, интенсивности пиков в образцах сравнивались с интенсивностями в холостых пробах. Те пики, в которых среднее значение по \log_2 шкале в образцах менее, чем на 0,5 превышало среднее значение в холостых пробах, удалялись из анализа.

Для предварительной аннотации полученных пиков, мы рассматривали 21 липидный класс, описанный в литературе в отношении мозга человека (PE, PC, PE O, PC O, PE P, PC P, PG, PS, PI, LPC, LPE, CAR, Cer, SM, HexCer, CE, SulfoHexCer, TAG, DAG, MAG и холестерол). Исходя из этих липидных классов, рассматривались различные теоретически возможные составы жирных кислот, и была создана таблица данных возможных значений m/z (масса/заряд). Далее, используя класс-специфичные аддукты, находились те липидные признаки, у которых значения m/z совпадали с теоретическими в пределах погрешности 10 ppm. Для заданного липидного класса и аддукта найденные признаки представляли в координатах m/z и время удерживания (рис. 2.18).

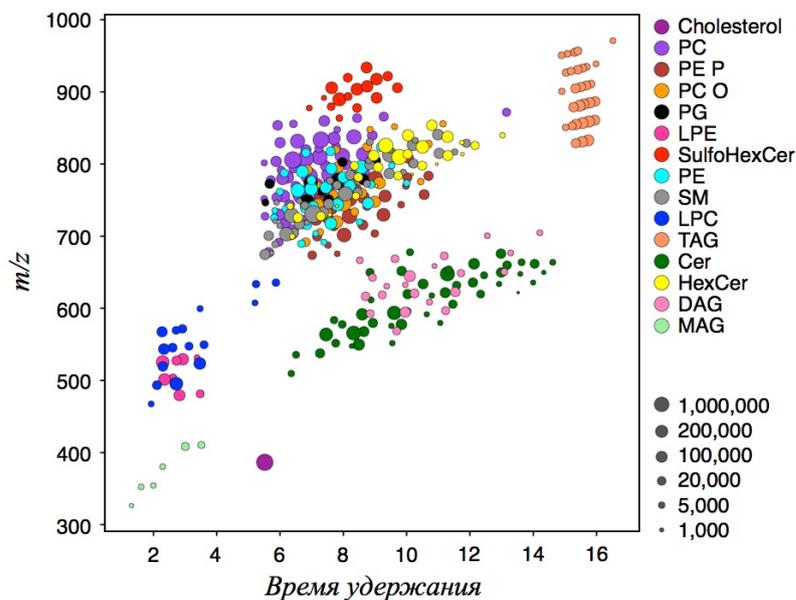


Рисунок 2.18 — Масса и время удержания липидов при ВЭЖХ-МС образцов мозга человека, подвергшихся процедуре аннотации с ручной верификацией.

Размер точек отражает среднюю интенсивность липида в мозге.

Адаптировано из [167].

Для того, чтобы признаки были отнесены к соответствующему липидному классу, они должны укладываться в визуально различимую решетку: времена выхода аналитов в пределах одного липидного класса закономерно увеличиваются с ростом количества углеродов в составе жирных кислот и уменьшается с добавлением кратных связей. Данные признаки аннотировались соответствующим классом, а признаки, не укладывающиеся в эту решетку, отфильтровывались из аннотации. Далее, для каждого аннотированного в одной из полярностей ионизации липидного класса, рассматривался класс-специфичный аддукт из другой полярности (если таковой имелся), и находилась аннотация для данного класса в другой полярности, путем сопоставления значений m/z в пределах погрешности 10 ppm и времени удерживания между двумя полярностями.

Описанная выше процедура биоинформатического анализа позволяет получить аннотацию очень высокого качества без дополнительных экспериментов. Тем не менее, для валидации разработанной нами биоинформатической процедуры проводится дополнительное подтверждение и уточнение идентификации найденных липидов с использованием фрагментных ионов в тандемном масс-спектрометрическом эксперименте. Итоговые спектры получают усредне-

нием трёх спектров фрагментации на разных энергиях соударения в столкновительной ячейке. Подобный подход позволяет получать более информативные характеристичные масс-спектры, отражающие структурные особенности представителей разных классов липидов.

Аннотацию соединений осуществляли в отдельном эксперименте для образца, полученного при смешивании аликвот из нескольких образцов. Условия хроматографического разделения сохранялись прежними. При этом перечень значений m/z для фрагментации выбирался согласно описанному ранее алгоритму с использованием закономерностей удерживания. Установленные формулы липидов основывались на совпадении массы иона-предшественника с теоретической с точностью менее 10 м.д. и масс фрагментов с точностью до 15 м.д. Отнесение вещества к конкретному классу фосфолипидов проводили при наличии в его тандемном спектре сигналов фрагментов (или нейтральных потерь), относящихся к структуре головной группы, к классу три-(ди-)глицеридов – при наличии нейтральной потери аммиака у иона-предшественника, к сфинголипидам – при наличии характеристичных ионов сфингоидного основания. Детализацию жирнокислотного состава для фосфолипидов проводили по наличию ионов депротонированных жирных кислот в спектрах второго порядка в режиме регистрации отрицательно заряженных ионов, для глицеролипидов – по нейтральным потерям жирных кислот в катионированной форме ($M+NH_4^+$) в спектрах второго порядка в режиме регистрации положительно заряженных ионов, для сфинголипидов – по разнице значений масс ионов-предшественников и ионов сфингоидных оснований.

Интенсивности липидных пиков далее нормализовали на внутренние стандарты, которые представляют разнообразный и репрезентативный для мозга набор липидных классов, а также на массу образца. Для визуализации образцов применялись методы PCA, PLS-DA [169] и иерархическая кластеризация с метрикой расстояния один минус коэффициент корреляции Пирсона. Более подробное описание процедуры биоинформатического анализа приведено в нашей работе [170] и может применяться не только для липидомных, но и для метаболомных данных. Пример применения этого подхода к анализу метаболомных данных для ткани мозга представлен в работе [154] и описан в разделе 3.1.1.

Таким образом, нам удалось разработать необходимый набор инструментов и подходов для анализа данных Hi-C и других типов омиксных данных. По-

ка данные Hi-C для мозга человека не были доступны, инструменты тестировались на иных данных. Однако, мы уделяли большое внимание универсальности разрабатываемых методов, всегда тестируя их на разных датасетах, чтобы каждый из подходов можно было впоследствии применить к появившимся данным для мозга человека. Важно отметить, что в процессе тестирования мы сделали ряд новых наблюдений об организации хроматина на всех иерархических уровнях - от хромосомных территорий и компартментов до ТADов и петель, а также о регуляторных аспектах функционирования клеток. Эти наблюдения позволили существенно расширить имеющиеся в мировой литературе фундаментальные представления о регуляторных процессах, структуре и функциях хроматина, универсальных для клеток разных типов.

Глава 3. Результаты и обсуждение

3.1 Изучение общих принципов организации хроматина

Прежде, чем переходить к изучению архитектуры хроматина в мозге человека, необходимо было исследовать общие принципы организации хроматина, чтобы поместить наблюдаемые особенности упаковки хроматина в клетках головного мозга в контекст общих представлений, универсальных для всех клеточных типов, о разных уровнях архитектуры хроматина - от крупномасштабных (хромосомные территории, компартменты) до самых мелких (ТАДы, хроматиновые петли).

3.1.1 Хромосомные территории

Хромосомные территории описывают неслучайное расположение отдельных хромосом внутри интерфазного ядра, при котором каждая хромосома занимает пространственно выделенную область [2]. На формирование и позиционирование хромосомных территорий влияют различные факторы, такие как размер хромосом, плотность генов и транскрипционная активность. Участки хромосом со сходной транскрипционной активностью имеют тенденцию группироваться вместе, образуя транскрипционные фабрики, которые облегчают скоординированную регуляцию генов [171].

В работе [172] мы детально изучили сеть межхромосомных контактов хроматина в ядрах млекопитающих, и показали их взаимосвязь с функциональным подобием генов. В работе представлено полногеномное исследование сетей межхромосомных контактов сегментов генома и генов в эмбриональных стволовых клетках человека и мыши. В обоих организмах наблюдается высокая степень кластеризации участков генома. Пространственная близость участков скоррелирована с функциональным подобием генов в обоих организмах, и с ко-экспрессией генов у человека. Также, взаимодействующие участки обогащены активными

маркерами хроматина в человеческих клетках. У мыши такого обогащения не наблюдается. Таким образом, мы можем сделать вывод, что меж-хромосомные взаимодействия не консервативны между видами, в отличие от внутри-хромосомных взаимодействий.

Действительно, организация хроматина на уровне меж-хромосомных взаимодействий может сильно отличаться между видами. Так, хромосомы дрожжей *S.cerevisiae* организованы в виде пучка таким образом, что все центромеры скластеризованы вместе, а теломеры находятся на периферии ядра (рис. 3.1А). Данные Hi-C показывают, что при такой организации участки, которые находятся на одинаковом расстоянии от центромер на разных плечах одной или разных хромосом, чаще контактируют друг с другом [173]. Ранее было показано, что пространственная структура влияет на геномные перестройки, ограничивая набор партнеров при репарации ДНК. Однако ее влияние на перестройки, возникающие в процессе эволюции, особенно на больших эволюционных расстояниях, не было изучено. В нашей работе [173] мы детектировали перестройки в видах *Saccharomyces* с помощью автоматического алгоритма (рис. 3.1Б-В), и показали повышенную частоту пространственных взаимодействий, согласно данным Hi-C, между регионами генома, образующими перестройку, по сравнению со случайными парами геномных участков, расположенных на том же линейном расстоянии друг от друга (рис. 3.1Г). Таким образом, наши результаты подтвердили связь пространственной организации хромосом и эволюционных событий перестроек.

образования хромосомных транслокаций и нарушения регуляции механизмов репарации ДНК [35].

К нестабильности генома и нарушению регуляции механизмов репарации ДНК также приводит нокаут SIRT6 или естественное понижение уровня SIRT6 при старении клеток, т.к. SIRT6 является одним из главных факторов, который поддерживает стабильность генома и способствует репарации ДНК [177]. SIRT6 является ядерным, ассоциированным с хроматином белком, который выполняет функцию деацетилирования гистонов и участвует в организации хроматина. В работе [154] мы определили ещё одну важную функцию деацетилазы SIRT6 - участие в регуляции митохондриальных процессов в головном мозге через взаимодействие с транскрипционным фактором YY1, который является ключевым игроком в механизме экструзии петель в хроматине (см. ниже в разделе 3.1.4), в частности при нейрогенезе [178].

Применив разработанные нами методы анализа транскриптомных данных (см. в разделе 2.2.5) и метаболомных данных (см. в разделе 2.2.6), мы проанализировали данные RNA-seq и ВЭЖХ-МС и показали, что дефицит SIRT6 приводит к резкому подавлению митохондриальных генов и изменениям в составе метаболитов, связанных с митохондриями. Поскольку сделать нокаут в мозге человека невозможно, работа выполнялась для образцов мышей дикого типа (WT) и с мозг-специфичным нокаутом (brSIRT6-KO) (рис. 3.2A).

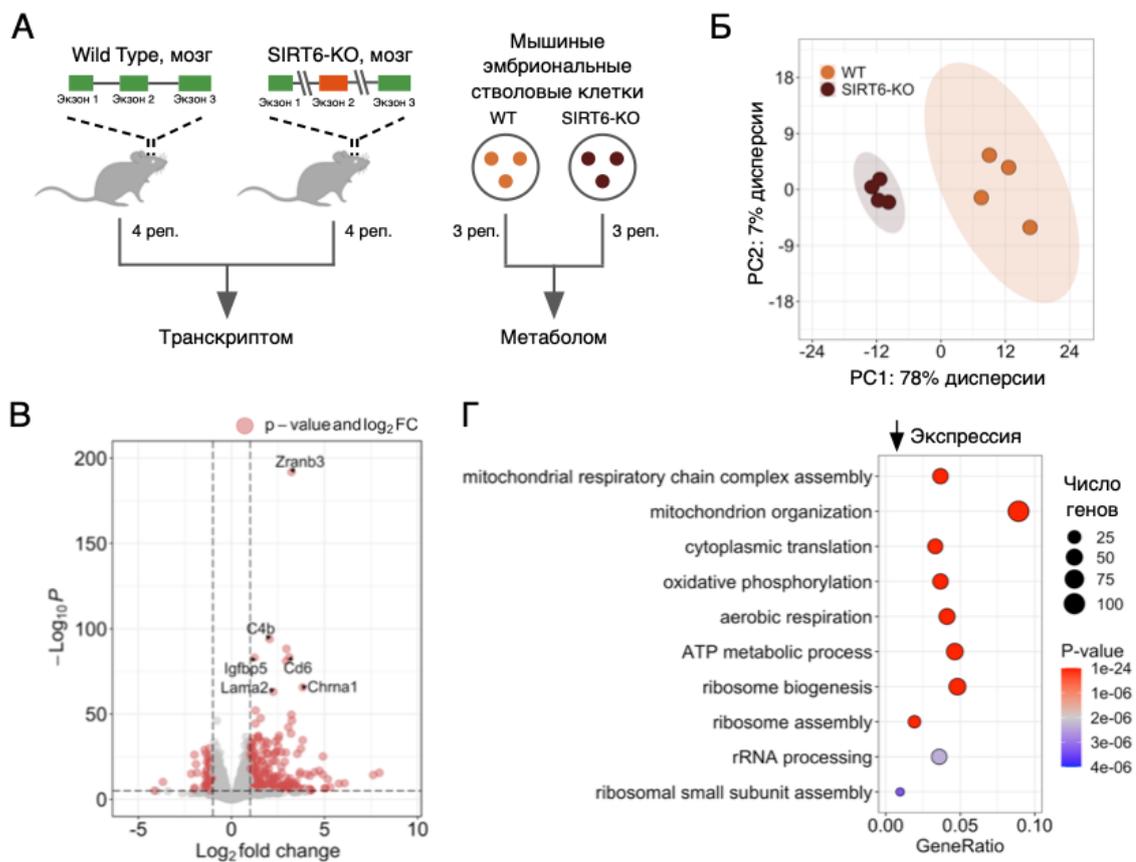


Рисунок 3.2 — SIRT6 регулирует экспрессию генов. (А) Схематическая иллюстрация дизайна экспериментов. Данные транскриптома были получены для образцов мозга мышей дикого типа (WT) и нокаутированных по SIRT6 (brSIRT6-KO). Метаболомные профили WT и brSIRT6-KO были получены для эмбриональных стволовых клеток мыши (mESC). (Б) График PCA, показывающий разделение между образцами WT (оранжевый) и brSIRT6-KO (коричневый). Оранжевые и коричневые овалы представляют собой доверительные эллипсы групп WT и brSIRT6-KO. (В) График, показывающий ДЭ гены у мышей brSIRT6-KO по сравнению с мышами WT. Красные точки указывают на значимо измененные гены, а серые точки представляют незначимые гены. (Г) Анализ GO, показывающий топ-10 обогащенных биологических процессов для генов с пониженной экспрессией. Каждый кружок соответствует обогащенному термину GO и варьируется по размеру в зависимости от количества значимых генов, принадлежащих этому термину. Адаптировано из [154].

Анализ главных компонент (PCA) транскриптомных профилей выявил значительные изменения в уровнях экспрессии генов между образцами brSIRT6-KO и WT (рис. 3.2Б). Анализ дифференциальной экспрессии между

WT и brSIRT6-KO определил 2870 дифференциально экспрессируемых (ДЭ) генов (рис. 3.2В). Анализ обогащения GO для подавленных генов выявил значимые термины, связанные с митохондриальными процессами (рис. 3.2Г). Таким образом, наши результаты показывают, что дефицит SIRT6 провоцирует значительные изменения экспрессии генов в головном мозге мышей и индуцирует нарушение регуляции транскрипции генов, связанных с митохондриями.

Чтобы понять, может ли дефицит SIRT6 индуцировать изменения в уровнях митохондриальных метаболитов, мы проанализировали данные ВЭЖХ-МС SIRT6-KO по сравнению с WT. Аналогично результатам RNA-seq, мы наблюдали глобальные изменения (рис. 3.3А). Анализ дифференциального содержания (ДС) показал, что 92 из 235 метаболитических пиков (около 39%) значительно изменялись при нокауте (рис. 3.3Б), включая аскорбиновую кислоту (повышение), малеиновую кислоту (понижение) и NAD^+ (понижение) в качестве наиболее значимых метаболитов. В соответствии с анализом транскриптома, измененные метаболиты оказались значимо и сильно связаны с путями энергетической системы митохондрий.

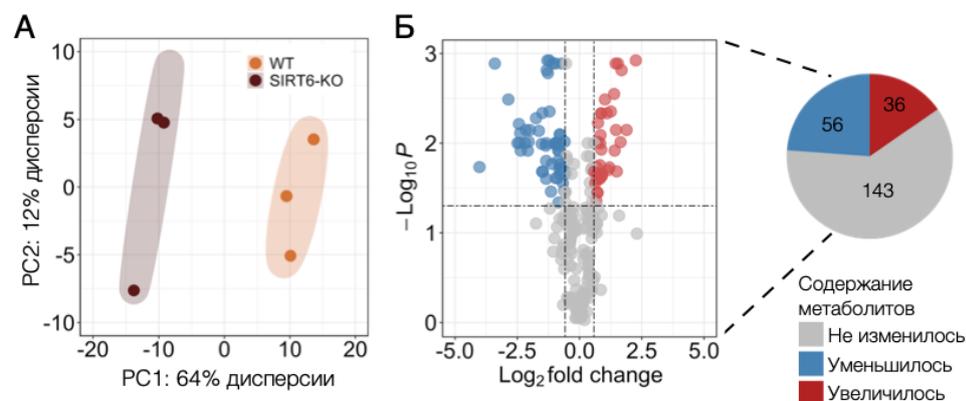


Рисунок 3.3 — SIRT6 регулирует уровни метаболитов. (А) График PCA, показывающий разделение между группами WT (оранжевые круги) и SIRT6-KO (коричневые круги) на основе метаболомных профилей. Оранжевые и коричневые овалы представляют собой доверительные эллипсы групп WT и SIRT6-KO. (Б) График, иллюстрирующий различающееся содержание метаболитов между образцами WT и SIRT6-KO. Круговая диаграмма (справа) демонстрирует количество повышенных (красный), пониженных (синий) и незначимых (серый) метаболитов в анализе. Адаптировано из [154].

Важно заметить, что сделанные нами биоинформатические наблюдения о регуляции митохондриальных генов и метаболитов белком SIRT6 были впо-

следствии подтверждены экспериментально коллегами из группы проф. Дебры Тойбер (Университет Бен-Гуриона в Негеве, Израиль) методами измерения мембранного потенциала митохондрий, количества митохондрий, и окрашиванием MitoSOX (индикаторы митохондриального супероксида), как описано в нашей работе [154]. Более того, наш анализ данных ChIP-seq показал, что SIRT6 регулирует работу митохондриальных генов совместно с белком YY1, т.к. их пики имеют неслучайное перекрывание в промотерах митохондриальных генов. При этом известно, что белок YY1 участвует в образовании хроматиновых петель (см. ниже в разделе 3.1.4), в частности при формировании нейронов [178]. Этот факт позволяет сделать предположение, что регуляторная функция SIRT6 в мозге также может быть опосредована структурой хроматина, либо напрямую через деацетилазную активность SIRT6, либо через взаимодействие с архитектурным белком YY1.

3.1.2 Компартменты хроматина

Компартменты представляют собой функционально важный уровень организации хроматина, обеспечивающий пространственное разделение транскрипционно активных и неактивных областей генома. Эксперименты Hi-C показали, что компартиментализация генома зависит от типа клеток и подвергается ремоделированию во время дифференцировки клеток [8]. Компартиментализация также меняется во время сперматогенеза, как мы показали в работе [165]. А именно, наблюдается существенное ослабление компартиментализации при переходе от сперматогониев (SpG) к сперматоцитам (SpCs) (рис. 3.4А,Б). Как в SpCs, так и в SpG хроматин разделен на активные и неактивные компартменты. 92% активных промоторов в SpG и 88% в SpCs находятся в активном компартменте. При переходе от SpG к SpCs сила активного и неактивного компартментов, характеризующая их пространственную изоляцию, значительно ослабевает (рис. 3.4В). Сила компартментов рассчитывалась как средняя нормализованная (O/E) частота контактов между данным бином Hi-C карты и другими бинами в пределах того же компартмента, деленная на среднюю нормализованную (O/E) частоту контактов между данным бином и всеми бинами в хромосоме.

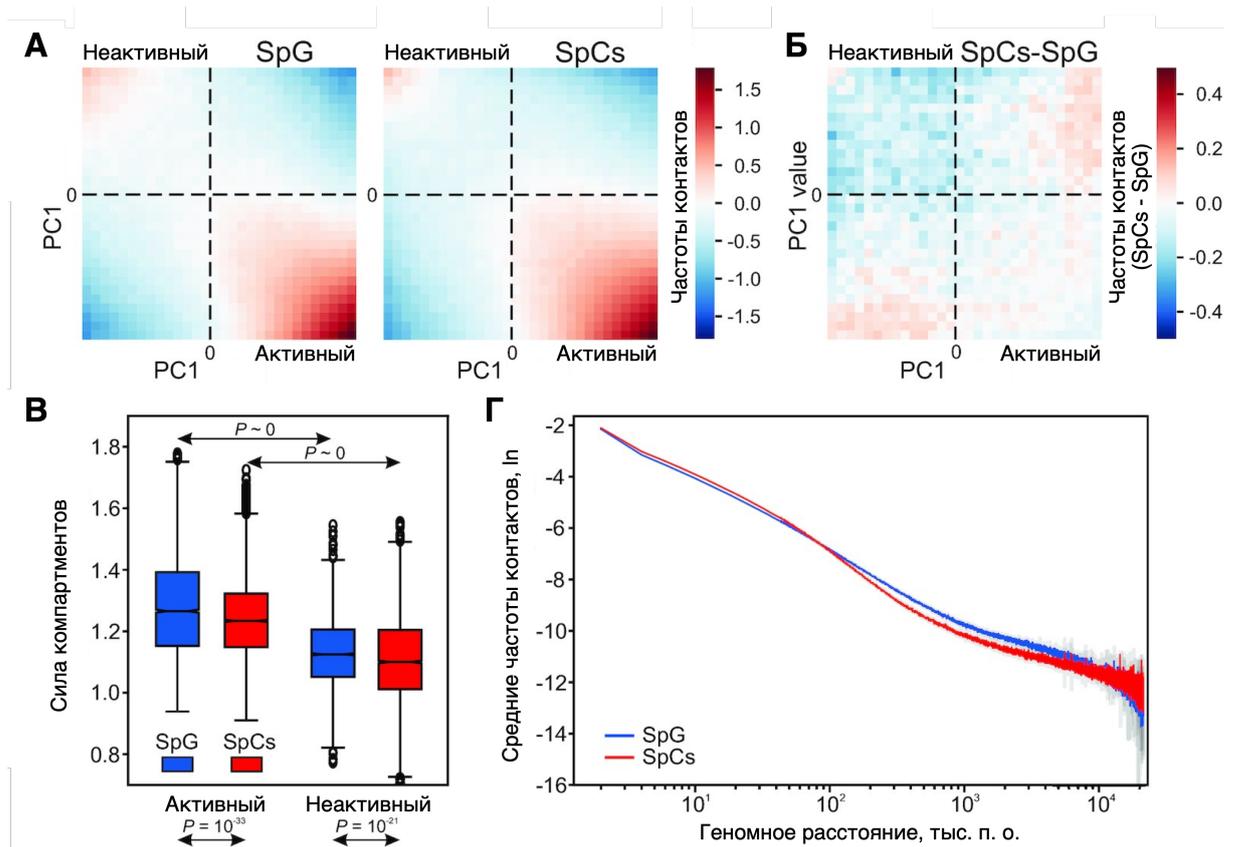


Рисунок 3.4 — Компартиментализация ослабевает при переходе от SpG к SpCs.

(А) Седловые графики, показывающие значения \log_2 обогащения частот взаимодействий для внутрихромосомных контактов во всех хромосомах в зависимости от их значений PC1 в SpG (левая панель) или SpCs (правая панель). (Б) Седловый график, показывающий результат вычитания значений обогащения внутрихромосомных контактов во всех хромосомах в SpCs из значений в SpG. (В) Сила активных и неактивных компартов в SpG и SpCs. P-значения были рассчитаны с использованием теста Вилкоксона. (Г) Кривые $P_c(s)$ для SpG (синяя кривая) и SpCs (красная кривая).

Адаптировано из [165].

Кривые $P_c(s)$, отображающие уменьшение частоты контактов в зависимости от геномных расстояний, также показывают, что дальние взаимодействия становятся слабее в SpCs (рис. 3.4Г). Учитывая наблюдаемое нами усиление контактов между плечами хромосом в пространстве в SpCs относительно SpG, эти результаты указывают на более сильное перемешивание активного и неактивного компартов хроматина внутри хромосомных территорий в SpCs по сравнению с SpG, что может быть связано с наличием увеличенных хромосомных территорий на стадии SpCs. Механизмы, которые управляют образованием

компарментов А и В, до конца не изучены, но считается, что они включают взаимодействие между факторами транскрипции, эпигенетическими модификациями и биофизическими свойствами хроматина [4].

3.1.3 Топологически ассоциированные домены (ТАДы)

Топологически ассоциированные домены (ТАДы) определяются как область генома, размер которой может варьироваться от сотен тысяч п.о. до нескольких миллионов п.о. с высоким уровнем взаимодействий внутри этой конкретной области и низким уровнем взаимодействий за ее пределами. Прогресс в определении 3D структуры хроматина позволил нам исследовать корреляции между пространственной близостью областей генома и их эпигенетическим состоянием. В работе [36] мы объединили данные о трехмерной архитектуре генома человека с данными о функциональном состоянии хроматина и обнаружили, что фрагменты различных хромосом, которые пространственно близки, как правило, имеют сходные модификации гистонов, уровень метилирования, чувствительность к ДНКазе, уровень экспрессии и состояние хроматина в целом. Кластеризация участков генома по пространственной близости показала наличие компактных кластеров, характеризующихся высоким уровнем чувствительности к ДНКазе и низким уровнем метилирования, а также рыхлых кластеров с противоположными характеристиками.

На основании результатов анализа Hi-C и РНК секвенирования в четырех клеточных линиях *Drosophila melanogaster*, мы предположили, что особые свойства нуклеосом в активном и репрессированном хроматине играют ключевую роль в формировании ТАДов [119]. Мы показали, что ТАДы обеднены активными маркерами хроматина и разделены участками активного хроматина с постоянно экспрессирующимися генами. В отличие от предыдущих работ, мы не наблюдали сильного обогащения сайтами связывания инсуляторных белков на границах ТАДов. Более того, позиции связывания инсуляторных белков предсказывали границы ТАДов хуже, чем маркеры активного хроматина (см. раздел 2.2.4, рис. 2.14). Интересно, что участки между ТАДами соответствовали декомпактизованным интербендам политенных хромосом, а ТАДы – плот-

но упакованным бендам. Мы предположили, что функциональные особенности хроматина (степень активности, уровень экспрессии генов) определяют его пространственную структуру, а именно положение границ ТАДов (рис. 3.5А). Мы предложили механизм формирования ТАДов, основанный на способности нуклеосом в неактивном хроматине притягиваться друг к другу, и на отсутствии такой способности у ацетилированных нуклеосом, и проверили этот механизм с помощью симуляций полимеров.

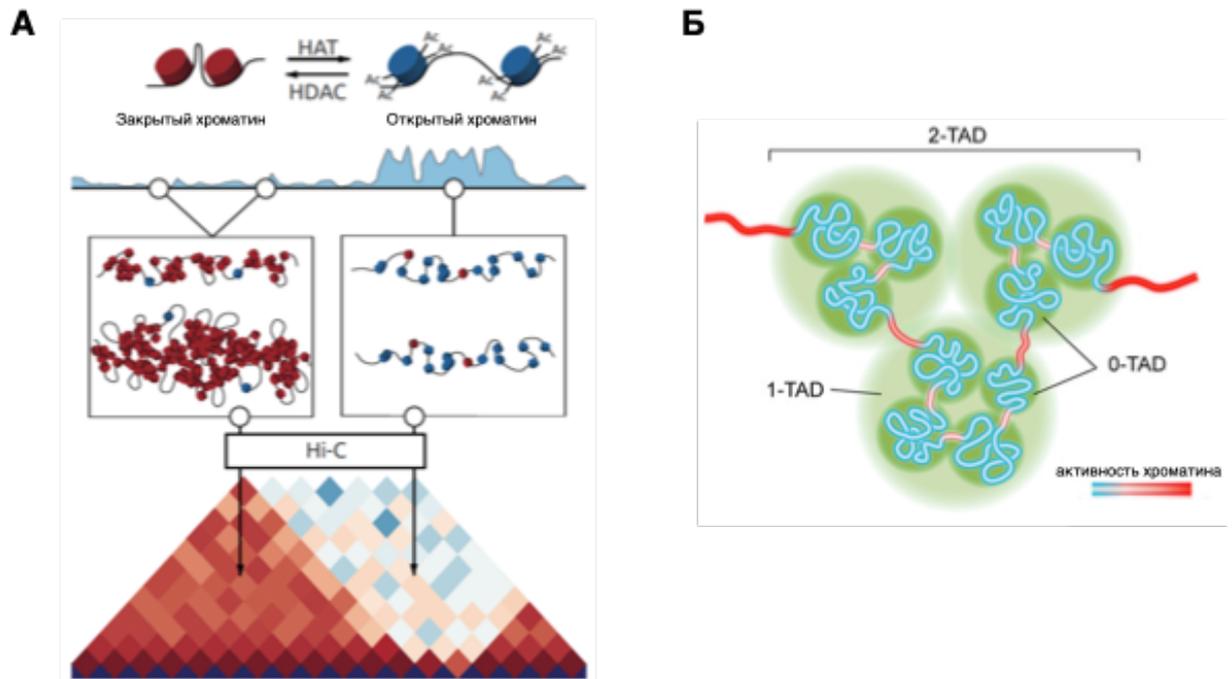


Рисунок 3.5 — Степень активности хроматина определяет его пространственную структуру, а именно положение границ ТАДов. (А) Схема, иллюстрирующая предложенную нами модель сворачивания хроматина в ТАДы в соответствии с ассоциацией нуклеосом. Высокий уровень ацетилирования хроматина в областях генома, содержащих активно транскрибируемые гены, препятствует упаковке хроматина в ТАДы из-за снижения межнуклеосомных взаимодействий. Адаптировано из [119]. (Б) Схема, иллюстрирующая иерархическую вложенность ТАДов, согласно данным Hi-C для популяций клеток. Адаптировано из [179].

Наши компьютерные симуляции показали, что способность «неактивных» нуклеосом притягиваться друг к другу, и отсутствие такой способности у «активных» нуклеосом достаточно для пространственного разделения этих типов хроматина на ТАДы и межТАДы. Однако, некоторые ТАДы содержат активный и неактивный хроматин одновременно, что противоречит такой модели. В

работе [180] мы привели дополнительные аргументы в пользу нашей модели. Мы показали, что активный хроматин располагается на поверхности ТАДов, и обсудили возможное влияние такой организации на взаимодействия между энхансерами и промоторами, и на сегрегацию ТАДов. Предполагается, что ТАДы являются основными функциональными регуляторными доменами, которые модулируют контакты между энхансерами и промоторами. Повышенная частота контактов внутри ТАДов опосредует физические взаимодействия между парами энхансер-промотор, в то время как высокая изоляция на границах ТАДов ограничивает такие взаимодействия для пар, расположенных в соседних ТАДах [53]. Таким образом, ТАДы, по-видимому, играют критическую роль в формировании специфических взаимодействий энхансер-промотор и, следовательно, регуляции экспрессии генов.

Исследования на клетках млекопитающих показали, что ТАДы организованы иерархически, с возможным включением нескольких мелких суб-ТАДов внутри более крупного ТАДа [54]. В дополнение к этим работам, в одном из своих исследований мы показали, что ТАДы дрозофилы также имеют иерархическую структуру: более мелкие активные ТАДы вложены в крупные неактивные [179] (рис. 3.5Б). Однако это исследование, как и другие подобные исследования на других организмах, были выполнены на популяции клеток. В результате, вопрос о существовании иерархической структуры ТАДов в клетках остается открытым: не является ли наблюдаемая нами иерархическая структура ТАДов результатом наложения профилей ТАДов, которые отличаются между клетками? Без ответа на этот вопрос исследование структуры и функций ТАДов, не было бы полным и всесторонним.

Поэтому, в составе большой коллаборации с группами Нечаева С.К., Гельфанда М.С., Вассецкого Е.С., Чертовича А.В., Шевелева Ю.Я. и Разина С.В. по изучению упаковки хроматина в отдельных клетках, мы исследовали иерархичность ТАДов в отдельных клетках дрозофилы. Результаты нашего анализа показали, что в геноме отдельных клеток наблюдаются неиерархические ТАДы при разрешении Hi-C карт в 100 тыс. п.о. [181]. Позиции этих ТАДов хорошо совпадают с позициями ТАДов в популяционных картах Hi-C. Более 40% границ ТАДов консервативны между разными клетками и характеризуются высоким уровнем активных эпигенетических маркеров, подобно ТАДам в популяционных картах Hi-C.

В целом, иерархическая организация ТАДов в контексте других уровней организации хроматина имеет сложную природу. Идея, предложенная в ранних исследованиях и рассматривающая ТАДы как строительные блоки компартментов, не нашла существенных подтверждений. Сейчас считается, что ТАДы скорее всего образуются по независимому от компарментализации механизму, в соответствии с моделью петлевой экструзии. В мировой литературе растет число свидетельств того, что хроматин представляет собой сложное динамическое взаимодействие механизмов фазовой сегрегации и петлевой экструзии [56]. Однако всё ещё требуются дополнительные исследования, чтобы поставить точку в этом вопросе окончательно.

3.1.4 Хроматиновые петли

Хроматиновые петли приводят дистальные цис-регуляторные элементы в непосредственную близость к их генам-мишеням и способствует правильной организации генома, поддержанию структуры ТАДов и точной регуляции экспрессии генов [61]. Одним из наиболее изученных примеров выраженной регуляции посредством пространственной реорганизации генома является β -глобиновый локус. Было показано, что переключение между экспрессией гемоглобина у плода и взрослого опосредуется взаимодействиями между контрольной областью локуса и промоторами соответствующих генов глобина [182].

В одной из своих работ мы также изучали глобиновый локус, а именно локус генов альфа-глобина [183], в котором мы проанализировали транскрипцию, профиль хроматиновых контактов и связывание CTCF на примере эритроцитов курицы. Мы впервые показали, что, подобно геному млекопитающих, геном курицы организован в ТАДы и компартменты. Полная активация транскрипции гена альфа-глобина в дифференцированных эритроидных клетках коррелирует с усилением регуляции нескольких соседних генов домашнего хозяйства и появлением обильной межгенной транскрипции. Хроматин в локусе альфа-глобина в целом разрыхляется. CTCF-связывающие и CTCF-заякоренные петлях хроматина становятся менее выраженными, в то время как суб-ТАД, содержащий кластер генов альфа-глобина, наоборот обогащается взаимодействиями хро-

тина. Таким образом, наши результаты демонстрируют, что эффекты тканеспецифической активации транскрипции действуют не только локально, а влияют на общую структуру хроматина и уровень транскрипции всего ТАДа.

Поэтому далее мы сосредоточились на исследованиях всего генома, и с целью раскрытия деталей организации хроматина в более широком масштабе. Хроматиновые петли можно условно разделить на две категории: расположенные на границах ТАДов петли и расположенные внутри ТАДов петли, которые специфически связывают промоторы и энхансеры. Петли хроматина, обнаруженные на границах ТАДов, формируются с помощью модели петлевой экструзии, в которой когезин, кольцеобразный белковый комплекс, окружает ДНК и постепенно выдавливает ее, образуя петлю [73]. В дополнение к этим основным факторам, участвующим в формировании петель на границах ТАДов, поддержание энхансер-промоторных петель может вовлекать дополнительные белки, потенциально специфичные для этих областей. Хотя точный механизм таких взаимодействий полностью не раскрыт, текущие исследования показали, что образованию петель также может способствовать ряд белков, таких как YY1, RAD21, HCF1, ZNF143 [63]. В частности, ранее описанные белки CTCF и RAD21 могут поддерживать границы ТАДов, в то время как другие белки (например, ZNF143 и HCF1, которые были обнаружены при взаимодействиях промотор-промотор) могут быть вовлечены в поддержание внутренних петель [63].

При формировании нейронов в процессе развития головного мозга человека, главным партнером CTCF в образовании петель является белок YY1, который действует как архитектурный белок, соединяющий петлевые взаимодействия, регулируемые при развитии [178]. В работе [166] мы показали, что YY1 играет важную регуляторную роль не только при развитии мозга, но и при его старении, причем как при нормальном старении, так и при патологическом, включая нейродегенеративные заболевания. В этой работе мы выполнили анализ дифференциальной экспрессии (ДЭ) генов между образцами мозга разных возрастов, чтобы проанализировать изменения транскрипции при нормальном старении, а также между образцами с нокаутом SIRT6 (SIRT6-KO), в качестве модели патологического старения (рис. 3.6А). Поскольку сделать нокаут в мозге человека невозможно, работа выполнялась на образцах мозга мышей. Иерархическая кластеризация образцов продемонстрировала высокую согласованность

между биологическими репликами (рис. 3.6Б). Интересно, что некоторые гены показали сходные изменения экспрессии при нормальном и патологическом старении (рис. 3.6В,Г). Мы провели функциональный анализ найденных ДЭ генов и определили транскрипционные факторы, их регулирующие (см. подробнее в [166]). Белок YY1 оказался в числе наиболее обогащенных транскрипционных факторов.

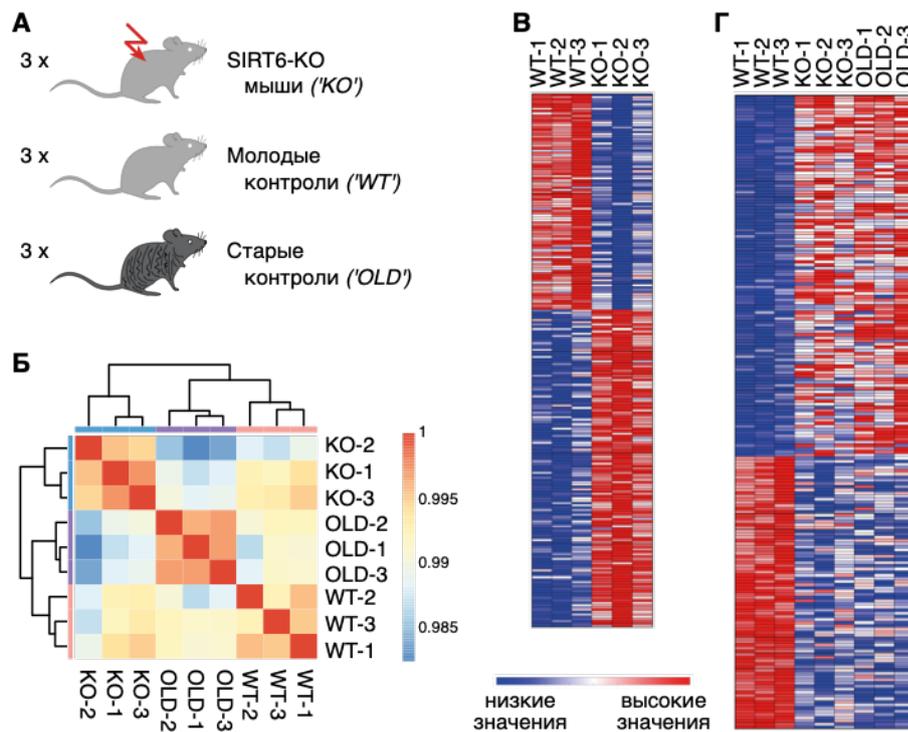


Рисунок 3.6 — Нокаут SIRT6 влияет на экспрессию генов в головном мозге.

(А) Схема эксперимента: РНК была выделена из мозга трех мышей SIRT6-KO, трех мышей дикого типа (WT) того же возраста (21 день), и трех старых мышей WT (22-26 месяцев). (Б) Иерархическая кластеризация образцов. Цвета показывают коэффициенты корреляции Пирсона. (В) Уровни экспрессии ДЭ генов при сравнении между SIRT6-KO и WT. (Г) Уровни экспрессии ДЭ генов при сравнении объединения SIRT6-KO и старых мышей WT по сравнению с молодыми мышами WT. Адаптировано из [166].

В той же работе [166] мы показали, что YY1 образует комплекс с SIRT6. Сначала мы биоинформатическими методами показали, что YY1 и SIRT6 регулируют одни и те же процессы в клетках мозга человека, включая старение и нейродегенерацию. В частности, мы обнаружили значимую положительную корреляцию между уровнями мРНК этих двух генов в различных тканях головного мозга человека по данным [184] (рис. 3.7А). Затем мы сопоставили списки

генов, которые ко-экспрессируются либо с SIRT6, либо с YY1 в головном мозге и нашли неслучайное и значительное перекрытие (P -значение $< 0,001$, рис. 3.7Б) и обогащение GO категориями, общими для YY1 и SIRT6 (рис. 3.7В). Далее, мы пересекли ChIP-seq данные для YY1 и SIRT6 в двух клеточных линиях и обнаружили, что гены, с которыми они связываются, участвуют в очень похожих KEGG путях (рис. 3.7Г). Эти биоинформатические результаты позволили нам сформулировать гипотезу о совместной регуляции генов белками YY1 и SIRT6, которая впоследствии была экспериментально подтверждена коллегами из группы проф. Дебры Тойбер (Университет Бен-Гуриона в Негеве, Израиль) методом ко-иммунопреципитации YY1 и SIRT6 (рис. 3.7Д).

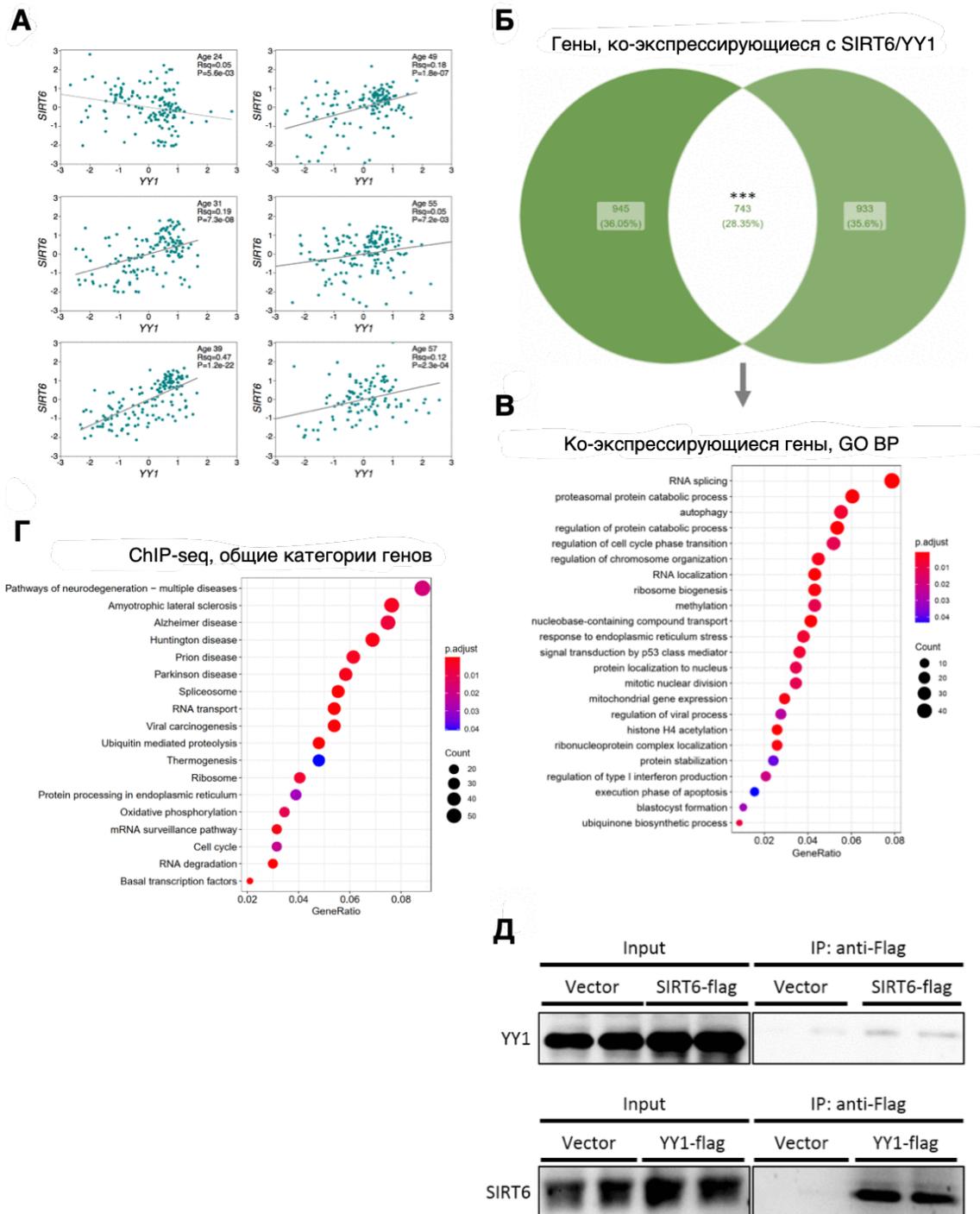


Рисунок 3.7 — SIRT6 и YY1 демонстрируют сходные регуляторные функции.

(А) Корреляции экспрессии в головном мозге 6 различных человек, основанные на данных Allen Brain Atlas. (Б) Диаграмма Венна для генов, совместно экспрессируемых либо с SIRT6, либо с YY1, и их перекрытие. (В)

Обогащенные категории биологических процессов GO, частью которых являются перекрывающиеся гены. (Г) Анализ обогащения KEGG для генов с пиками ChIP-seq как в данных SIRT6, так и в данных YY1. (Д) Вестерн-блоты экспериментов по ко-иммунопреципитации SIRT6 и YY1. *** - значение $p < 0,001$. Адаптировано из [166].

В последующей работе [185] мы показали, что помимо формирования комплекса с SIRT6, YY1 регулирует экспрессию длинной некодирующей РНК TP73-AS1 в мозге, которая играет важную роль при глиобластоме - раке мозга, который характеризуется неблагоприятным исходом и пятилетней выживаемостью $< 10\%$ [186]. YY1 высоко экспрессируется при опухолях головного мозга, включая глиобластому, и его повышенные уровни связаны с агрессивностью заболевания [187; 188]. Подобно предыдущему исследованию, мы сначала показали коэкспрессию YY1 и TP73-AS1 биоинформатическими методами, а затем коллеги из группы проф. Дебры Тойбер (Университет Бен-Гуриона в Негеве, Израиль) экспериментально подтвердили взаимодействие между YY1 и промотором TP73-AS1. Они оверэкспрессировали YY1 и измерили активность промотора с помощью анализа активности люциферазы, и обнаружили, что оверэкспрессия YY1 значительно увеличивала активность промотора. Чтобы изучить корреляцию экспрессии между TP73-AS1 и YY1 при нормальном старении, мы использовали данные GTEch и обнаружили положительную корреляцию в различных областях мозга. Кроме того, экспрессия TP73-AS1 и YY1 была скоррелирована и при патологическом старении мозга человека (см. подробнее в [185]).

Таким образом, YY1 и SIRT6 играют важную регуляторную роль в клетках мозга человека и могут совместно регулировать экспрессию. Регуляторная функция YY1 точно связана со структурой хроматина, поскольку было показано, что YY1 участвует в формировании хроматиновых петель при нейрогенезе [178]. Кроме того, YY1 может взаимодействовать как с гистондеацетилазами (HDAC1-HDAC3), так и с гистонацетилтрансферазами (p300 и CREB), выступая в качестве контекстно-зависимого регулятора ремоделирования хроматина [189]. Эта универсальность вовлекает YY1 в ключевые механизмы транскрипции внутри клетки, а его деплеция приводит к уменьшению энхансер-промоторных взаимодействий и глобальным изменениям экспрессии генов [75]. Понятно, что деацетилаза SIRT6 также является ключевым игроком в этих процессах, и скорее всего действует через реорганизацию хроматина, однако точный механизм остается неясным и только начал раскрываться в нашей недавней работе [154] (см. выше в разделе 3.1.1).

3.2 Изучение организации хроматина в нейронах мозга человека

3.2.1 Оптимизация протокола Hi-C для образцов мозга

Классический протокол Hi-C был разработан для определения структуры хроматина в клеточных линиях [3]. Фиксация и механическая гомогенизация ткани, как правило, снижают качество данных. Поэтому определение конформации хроматина для образцов мозга представляет собой нетривиальную задачу. Это обусловлено необходимостью выделения из сложной гетерогенной структуры мозговой ткани однородной популяции целевых клеток (нейронов), а также необходимостью модификации и оптимизации целого ряда этапов протоколов подготовки библиотек. Вероятно поэтому среди сотен публично доступных наборов Hi-C данных только несколько публикаций описывают архитектуру хроматина в клетках мозга человека [12; 59; 79; 95; 190; 191]. Причем анализ пространственной укладки хроматина в клетках взрослого мозга составляет небольшую долю всех опубликованных данных [12; 79], так как с эмбриональным мозгом работать проще.

Нам удалось оптимизировать стандартный Hi-C протокол для постмортальных замороженных образцов ткани мозга в работе [110]. Протокол включает в себя несколько специфических для ткани мозга этапов, таких как экстракция ядер в буфере с низким содержанием сахарозы и центрифугирование в градиентных растворах йодиксанола, а также иммуномечение и сортировка NeuN-положительных ядер для дальнейшего анализа структуры хроматина (рис. 3.8). Фиксация - важнейший шаг для сохранения ядерной и клеточной архитектуры. В его случае ДНК-белковые комплексы фиксируются формальдегидом, в результате чего взаимодействующие локусы связываются друг с другом ковалентными ДНК-белковыми связями. Наши модификации экспериментального протокола позволяют использовать меньшее количество исходного материала по сравнению с оригинальной статьей [192]. Далее, следует стандартная Hi-C процедура описанная в [193], с небольшими изменениями (см. подробнее в нашей работе [110]). В частности, мы использовали частощепящую рестриктазу DpnII, чтобы получить карты Hi-C высокого разрешения.

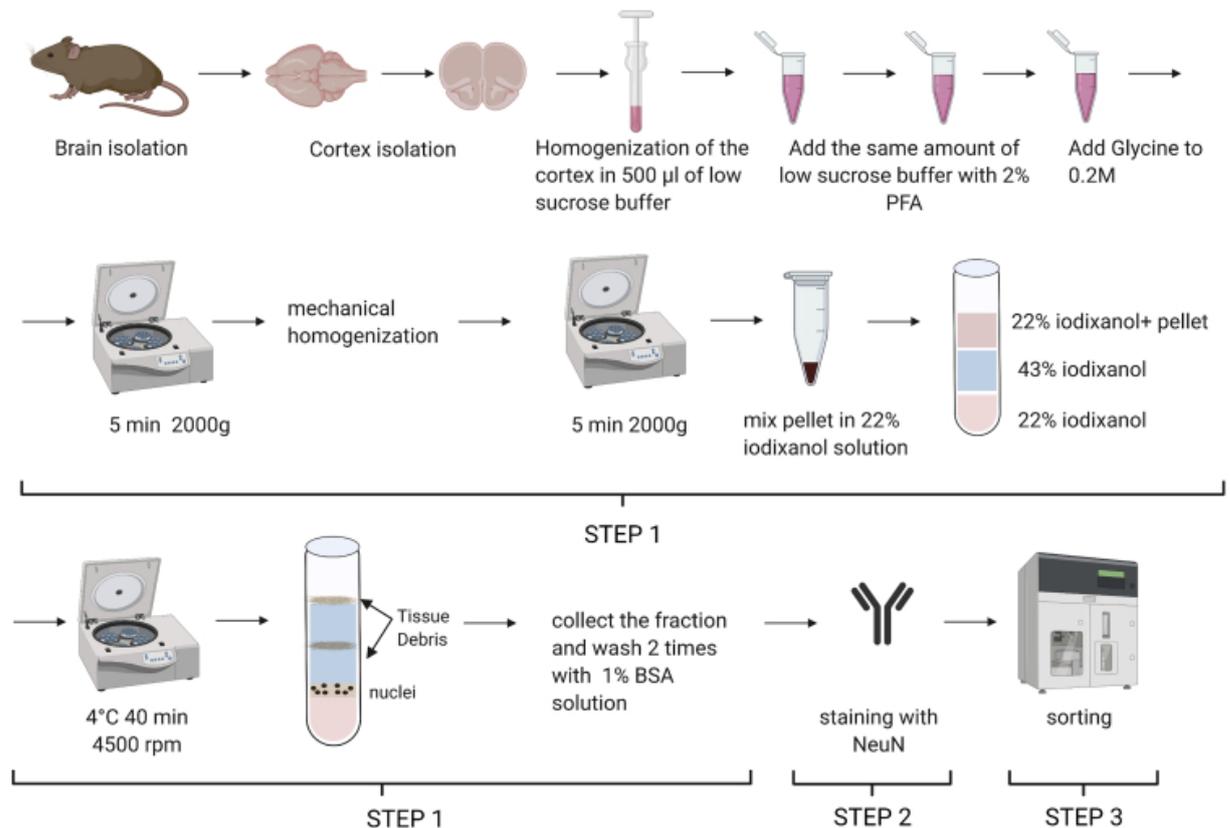


Рисунок 3.8 – Схематическая иллюстрация подготовки ядер для протокола Hi-C в клетках головного мозга. Адаптировано из [110].

С использованием оптимизированного нами протокола Hi-C, мы выполнили эксперименты Hi-C для региона мозга человека BA22p (левая часть зоны Вернике - латеральный участок височной коры мозга человека, связанный с пониманием речи), с выделением однородной популяции целевых клеток (нейронов) из гетерогенной структуры мозговой ткани путем сортировки ядер на основе окрашивания нейрональным маркером NeuN (NeuN+), а также популяции не-нейрональных клеток (NeuN-) (рис. 3.9А). Анализ наших карт Hi-C показывает, что хроматин имеет очень специфическую структуру в нейронах, по сравнению с другими клетками мозга (рис. 3.9Б). При этом отличия в организации хроматина между нейронами и другими типами клеток практически не изучены.

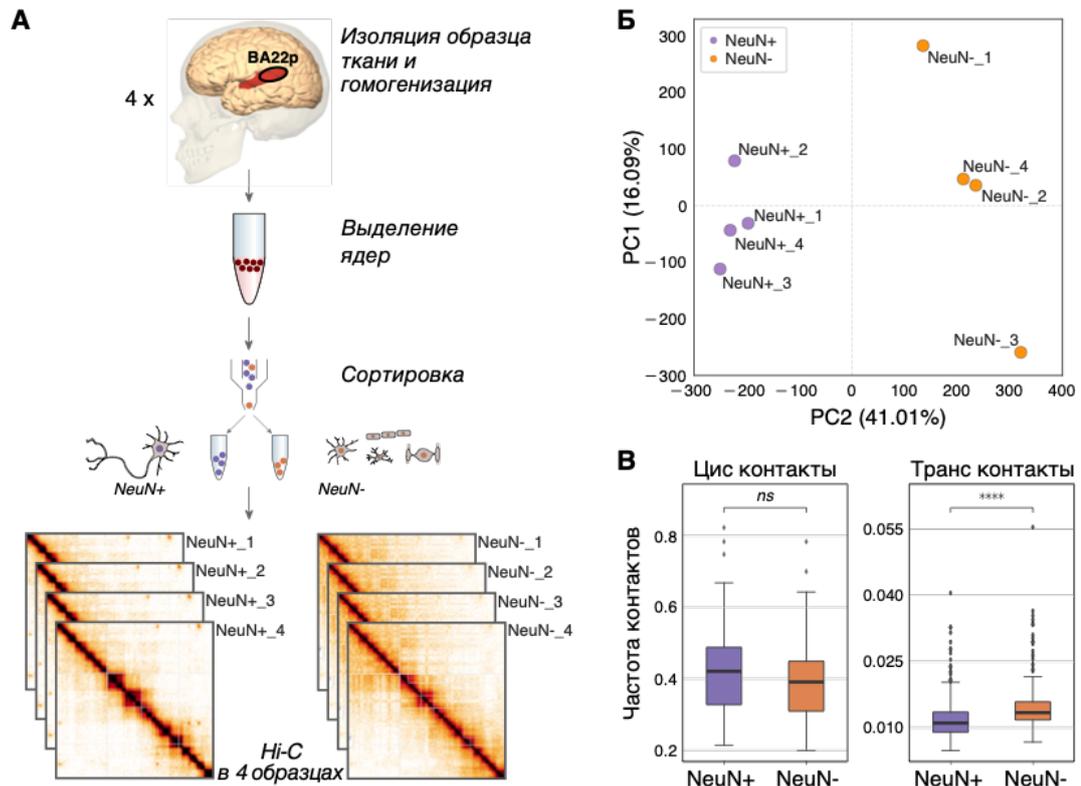


Рисунок 3.9 — Организация хроматина в клетках NeuN+ и NeuN-. (А)

Анатомическая локализация анализируемой области мозга, экспериментальная процедура и дизайн исследования. Эксперименты Hi-C проводили в клетках NeuN+ и NeuN-, изолированных с помощью FACS из речевой области Вернике четырех доноров. (Б) График анализа главных компонент, основанный на вариации показателя изоляции (IS) среди всех созданных карт Hi-C. Цвета обозначают клетки NeuN+ и NeuN- здесь и на панели С. (В) Взаимодействия внутри всех хромосом (цис-контакты, левая панель) и между всеми парами хромосом (транс-контакты, правая панель), рассчитанные отдельно для клеток NeuN+ и NeuN-. Звездочками отмечены

Р-значения критерия Вилкоксона: **** - $p < 0,00001$, ns - $p > 0,05$.

Адаптировано из [194].

3.2.2 Меж-хромосомные взаимодействия в нейронах понижены

На уровне глобальных различий в организации хроматина, анализ частоты цис- и транс-контактов показывает значительное снижение транс-контактов в нейронах по сравнению с не-нейронами (критерий Вилкоксона, Р-значение <

10^{-5} , рис. 3.9В), что указывает на то, что хромосомные территории в клетках NeuN+ более выражены. Это наблюдение можно объяснить большим объемом ядра в нейронах по сравнению с глиальными клетками [195], что потенциально приводит к дополнительному свободному пространству в ядре для лучшего пространственного разделения хромосом. Увеличенный размер ядра также может приводить к глобальной деконденсации хроматина в нейронах, чтобы занять свободное ядерное пространство, однако мы наблюдаем противоположную тенденцию для цис-контактов (рис. 3.9В). Недавние исследования показывают, что глобальная структура хроматина устойчива к резкому увеличению и сокращению объема ядра, с сохранением петель, ТАДов, активных и неактивных компартментов и даже хромосомных территорий [196]. Следовательно, мы можем заключить, что размер ядра сам по себе вряд ли может объяснить глобальные различия в организации генома между нейрональными и не-нейрональными клетками.

3.2.3 Компарментализация в нейронах менее выражена

Заметно отличающаяся частота контактов на больших геномных расстояниях между популяциями клеток NeuN+ и NeuN- потенциально может быть связана с различной компарментализацией хроматина в этих клетках. Действительно, как визуальное сравнение, так и анализ главных компонент карт Hi-C подтверждают эту гипотезу (рис. 3.10А). В частности, мы наблюдаем уменьшение взаимодействий В-компартамента и небольшое, но значимое увеличение взаимодействий А-компартамента в нейронах по сравнению с NeuN- клетками (рис. 3.10Б-В, критерий Вилкоксона, Р-значение $< 10^{-10}$).

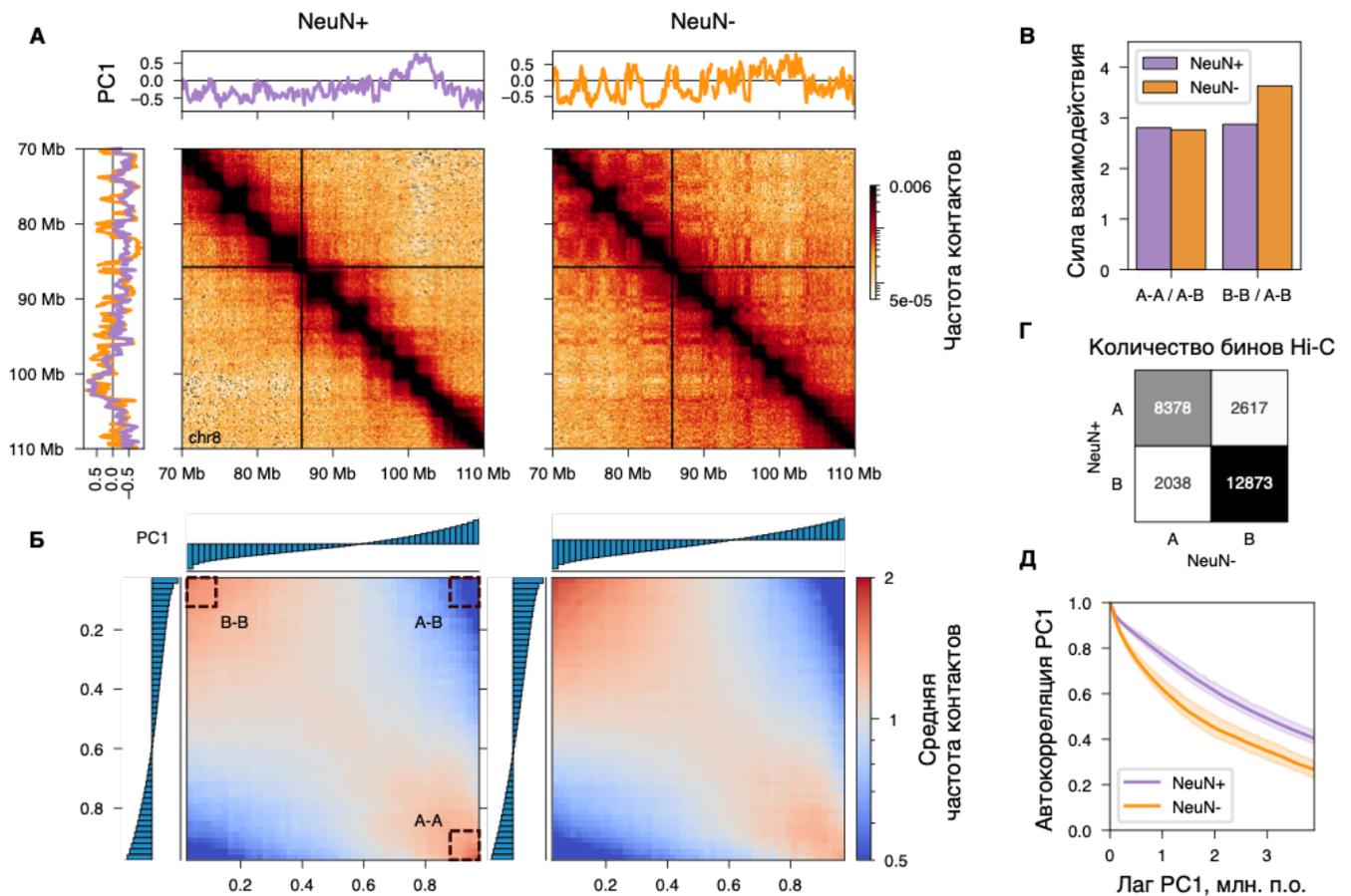


Рисунок 3.10 — Различия в компартаментах хроматина между клетками NeuN+ и NeuN-. (А) Фрагмент Hi-C карты с собственным вектором первой главной компоненты (PC1) для клеток NeuN+ (слева) и NeuN- (справа).

Положительные значения PC1 соответствуют А компартменту, а отрицательные значения - В компартменту. (Б) Частота взаимодействий на Hi-C карте геномных областей, упорядоченных по соответствующему рангу PC1 (седловая диаграмма) для клеток NeuN+ (слева) и NeuN- (справа). (В)

Сила взаимодействия компартментов, рассчитанная как самое сильное взаимодействие внутри компартмента, деленное на самое слабое межкомпарментное взаимодействие. (Г) Распределение геномных бингов размером 100 т.п.н. между типами компартментов в клетках NeuN+ и NeuN-.

(Д) Более быстрое спадение автокорреляционной функции NeuN- PC1, демонстрирующее более тонкую компарментализацию NeuN-хроматина.

Адаптировано из [194].

Затем мы определили области, в которых происходит смена компартментов между нейронами и не-нейронами. Такие области составляют примерно 15% генома (рис. 3.10Г). В целом, не-нейроны имеют более тонкую компарментализацию по сравнению с клетками NeuN+, что подтверждается более быст-

рым спадом автокорреляционной функции собственного вектора компартмента (PC1) (рис. 3.10Д). Снижение активности когезина приводит к аналогичному более тонкому паттерну компартментализации [197]. Следовательно, различная активность экстррузии петель и/или различное содержание когезина могут быть факторами, объясняющими различия в компартментализации между нейронами и не-нейронами.

3.2.4 ТАДы в нейронах более выражены

Поскольку когезин-зависимая экстррузия петель является основным механизмом образования ТАДов у млекопитающих, мы предполагаем, что более высокое содержание когезина в нейронах может влиять на профиль ТАДов. Визуальный анализ карт Hi-C выявляет ряд специфических локусов с яркими различиями в структуре ТАДов между нейрональными и ненейрональными клетками, особенно в непосредственной близости от генов, участвующих в нейрон-специфических процессах, таких как поддержание синапсов и транспортных систем, регуляция альтернативного сплайсинга и др. (рис. 3.11А). Чтобы систематически анализировать профили ТАДов в клетках NeuN+ и NeuN-, мы применили автоматизированную аннотацию ТАДов на основе индекса инсуляции и идентифицировали 2194 специфических для NeuN+ и 2218 специфических для NeuN- границ ТАДов, а также 2162 общих границы ТАДов (рис. 3.11Б). Мы отмечаем, что, хотя многие границы ТАД далее будут называться клеточно-специфичными, они на самом деле присутствуют в обоих типах клеток, но изолируют хроматин с разной силой.

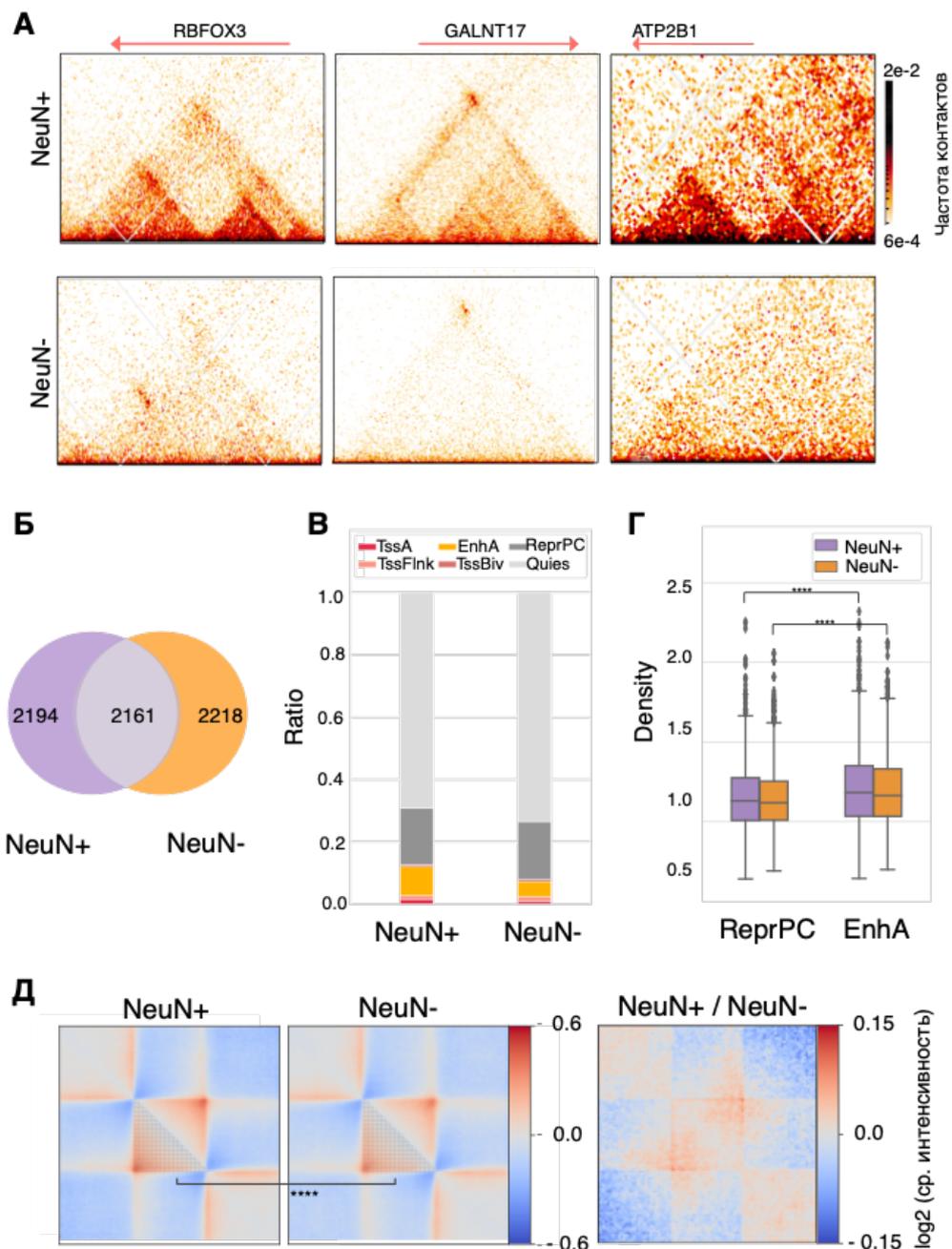


Рисунок 3.11 — Клеточно-специфичные различия ТАДов в клетках NeuN+ и NeuN-. (А) Примеры дифференциальных профилей ТАДов в дифференциально экспрессированных генах нейронов. Показана локализация генов RBFOX3, GALNT17, ATP2B1. (Б) Диаграмма Венна для пересечения нейронных и не-нейронных границ ТАДов, созданная на основе профилей изоляции. (В) Полногеномное покрытие шестью состояниями хроматина для клеток NeuN+ (слева) и NeuN- (справа). (Г) Ящики с усами для плотностей ТАДов, относящихся к группам с наибольшим покрытием состояниями хроматина ReprPC или EnhA. (Д) Средний ТАД. Адаптировано из [194].

Следующим шагом была характеристика ТАДов, с целью выяснить опосредованные экспрессией изменения на этом уровне организации хроматина, наряду с возможным функциональным объяснением наблюдаемых клеточно-специфических различий. Для этого мы изучили состояния хроматина, предсказанные ChromНММ для клеток NeuN+ и NeuN- [198], а также данные секвенирования РНК, специфичные для каждого клеточного типа [164]. Полногеномный анализ состояний хроматина показывает, что активные состояния занимают более высокую долю генома в нейрональных по сравнению с не-нейрональными клетками (рис. 3.11В). Повышенный уровень активных гистоновых меток может быть связан с увеличением общей экспрессии генов в нейронах [199]. Затем мы выбрали две группы ТАДов с наибольшей долей состояний хроматина RepгРС или EnhА соответственно и наблюдали, что ТАДы, обогащенные активными состояниями хроматина, обладают более высокой плотностью по сравнению с ТАДами, обогащенными меткой RepгРС как в нейрональных, так и в не-нейрональных клетках (рис. 3.11Г). Это наблюдение согласуется с предыдущими работами, которые показали положительную корреляцию между уровнем экспрессии и плотностью хроматина [200].

В среднем, ТАДы более выражены в нейронах по сравнению с ненейрональными клетками (рис. 3.11Д). В дополнение к более тонкой компартиментализации [197] показано, что уменьшение уровня когезина уменьшает размер ТАДов и петель как *in vitro* [70], так и *in silico* [201]. Когезин обогащен в нейронах по сравнению с не-нейрональными клетками. Таким образом, мы считаем, что наблюдаемая разница в выраженности ТАДов также может быть связана с повышенной концентрацией когезина в нейронах по сравнению с не-нейрональными клетками.

3.2.5 Нейрональные петли более длинные

Заметные различия в профиле ТАДов и паттернах экспрессии генов между клетками NeuN+ и NeuN- предполагают образование дифференциальных петель между регуляторными элементами генома. Визуальный анализ карт Hi-C выявил ряд дифференциальных петель, в частности, между нейрональными

энхансерами и генами со специфичной для нейронов экспрессией, например, *ENC1* и *GRIN3A* (рис. 3.12А). Полногеномная аннотация петель на картах Hi-C с разрешением 5 Кб идентифицировала 1078 общих петель и 1545 петель, специфичных для NeuN+ (рис. 3.12Б). В то время как средняя интенсивность петель одинакова для клеток NeuN+ и NeuN-, группировка общих петель на основе соотношения их интенсивности между нейронами и не-нейронами позволяет нам наблюдать, что многие петли различаются по степени представленности между типами клеток (рис. 3.12Г). Примечательно, что петли, которые более выражены в нейрональных клетках, значительно длиннее, чем петли, которые оказались более сильными в не-нейрональных клетках (рис. 3.12Д). При этом известно, что длина петли положительно связана со временем пребывания когезина на ДНК [73], а загрузчик когезина NIPBL обогащен в нейронах по сравнению с не-нейрональными клетками.

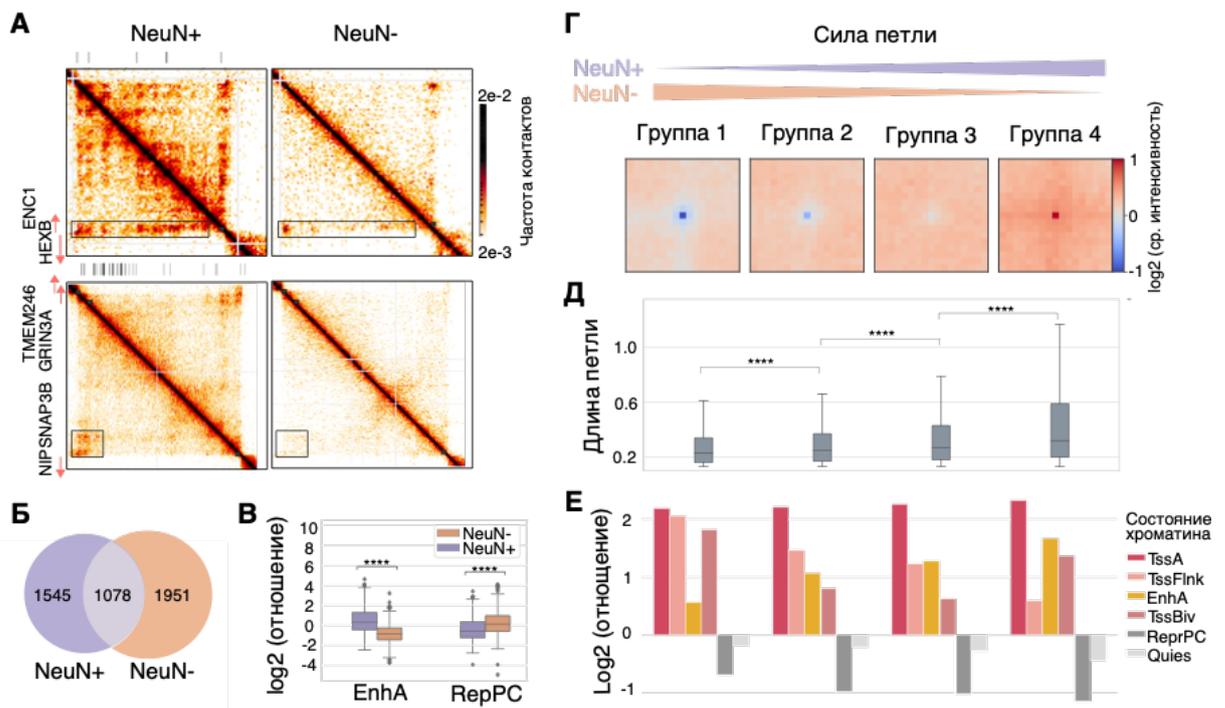


Рисунок 3.12 — Структура петель в NeuN+ и NeuN- клетках. (А) Фрагменты карт Hi-C с разрешением 5 КБ для областей взаимодействий специфических для нейронов генов с энхансерами. Треки энхансера отображаются в верхней части карт Hi-C. (Б) Диаграмма Венна для пересечения положений нейронных и не-нейронных петель. (В) Ящики с усами для плотности петель, относящихся к группам с наибольшим покрытием состояниями хроматина ReprPC или EnhA. (Г) Группы петель, определяемые по увеличению отношения плотности NeuN+/NeuN-. (Д) Графики распределения длины петли по четырем группам. (Е) Распределение состояний хроматина в нейронах по группам петель. Адаптировано из [194].

Далее, к основаниям петель была применена аннотация состояний хроматина, по процедуре, аналогичной той, которая использовалась для описания границ ТАДов. Мы обнаружили, что более выраженные петли имеют тенденцию к обогащению активными состояниями хроматина и обеднению неактивными состояниями (рис. 3.12Е). Петли с преобладающим покрытием состоянием ReprPC в обоих основаниях менее интенсивны в нейронах по сравнению с не-нейрональными клетками. Напротив, петли, основания которых обогащены состояниями EnhA, более выражены в нейрональных клетках (рис. 3.12В). В целом, было обнаружено, что сила петли коррелирует с распределением активных состояний хроматина.

3.2.6 Hi-C взаимодействия в нейронах человека

Примечательной особенностью карт NeuN+ Hi-C является наличие ярких точек, расположенных далеко от главной диагонали и, таким образом, представляющих собой сильные контакты на большом расстоянии, 92% (197 из 214) которых перекрываются с пиками H3K27me3 ChIP-seq (далее мы называем эти контакты “нейрональными точками”, рис. 3.13А). Частота контактов Hi-C в нейрональных точках намного выше по сравнению с частотой контактов тех же пар областей в NeuN- (рис. 3.13Б). Нейрональные точки образуют сети, включающие до 11 локусов (рис. 3.13В). Эти сети могут включать межхромосомные взаимодействия, которые присутствуют почти исключительно в нейронах (рис. 3.13Г).

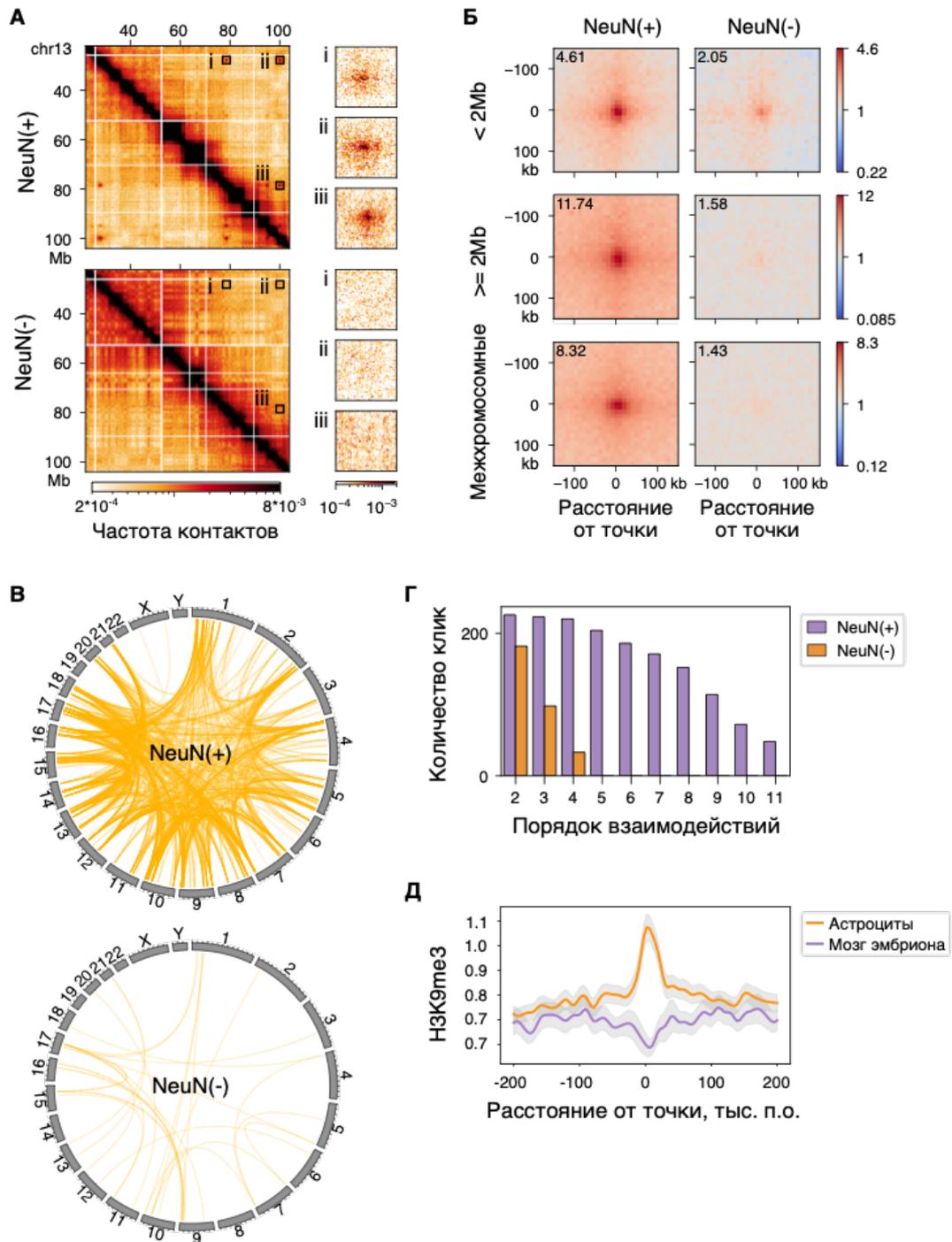


Рисунок 3.13 — Свойства нейрональных точек. (А) Фрагмент карты Hi-C с нейрональными точками, присутствующими в NeuN+ (вверху), но отсутствующими в NeuN- (внизу). (Б) Средний сигнал Hi-C для нейрональных точек (слева) и соответствующих им пар локусов в NeuN- (справа). (В) Значимые меж-хромосомные взаимодействия локусов нейрональных точек в NeuN+ и NeuN-. (Г) Статистика размера (порядка) клик (сетей), созданных нейрональными точками (фиолетовые столбики) или локусами нейрональных точек в NeuN- (оранжевые столбики). (Д) Средний профиль H3K9me3 ChIP-seq в якорях нейрональных точек. Адаптировано из [194].

Для изучения взаимодействия между гетерохроматином и нейрональными точками, мы проанализировали данные ChIP-seq для H3K9me3, гистоновой метки конститутивного гетерохроматина. Мы использовали данные H3K9me3 головного мозга эмбриона человека в качестве модели NeuN+ и астроцитов - в качестве модели NeuN- [202], т.к. только такие данные были доступны. Нейрональные точки были обогащены H3K9me3 в астроцитах, но не в эмбриональном мозге (рис. 3.13Д). Интересно, что большинство нейрональных точек (171 из 214) находятся вблизи сайтов начала транскрипции (TSS) генов, кодирующих белок. При этом в нейронах наибольшая доля этих TSS является “бивалентной”, т.е. занята H3K27me3 вместе с H3K4me3, тогда как в не-нейрональных клетках большинство этих TSS занято H3K27me3 без H3K4me3. Известно, что “бивалентные” промоторы заняты РНК-полимеразой RNAP [203; 204], которая зависит от убиквитинирования RING1A и RING1B [204]. Таким образом, как дальние взаимодействия, так и “бивалентность” могут быть следствием присутствия RING1A и/или RING1B на промоторах генов в нейрональных точках.

Из всех генов, кодирующих белок в локусах нейрональных точек, 80% являются транскрипционными факторами (ТФ), причем они значимо ассоциированы с развитием организма, согласно анализу GO онтологий. Подавляющее большинство этих ТФ высоко экспрессируются в развитии и низко экспрессируются в зрелых NeuN+ и NeuN- клетках. Примечательно, что большинство (119 из 197) из этих ТФ значимо связаны с развитием нервной системы [194]. Поскольку метка H3K27me3 связана с PcG репрессией, мы предполагаем, что белки PcG подавляют гены ТФ в нейрональных точках. Эта гипотеза подтверждается двумя исследованиями, проведенными на мышах [205; 206] и демонстрирующими, что нокауты белков PRC2 или RING1A/B в нейронах мыши приводят к усилению экспрессии многих мышечных ортологов генов из нейрональных точек человека.

Обобщая вышесказанное, 3D организация генома нейронов человека в значительной степени отличается от других типов клеток. Многие биологические процессы, включая экструзию петель, компарментализацию, дальние взаимодействия, вероятно, работают в нейронах человека по-другому, что способствует правильной экспрессии генов в клетках этого типа. Поэтому изучение функциональной роли организации хроматина в клетках мозга человека невозможно без

сопоставления данных Hi-C с данными об экспрессии генов и другими омиксными данными.

3.3 Интеграция данных Hi-C с другими омиксными данными в контексте изучения мозга человека

3.3.1 Изучение структуры и функции клеток мозга человека на различных уровнях молекулярной организации

Человек как вид характеризуется высоким уровнем интеллекта, причем потенциал развития сложных когнитивных функций специфичен для людей: детеныши приматов, выросшие в человеческой среде, не развивают сопоставимых когнитивных навыков. Первым шагом на пути к определению механизма развития когнитивных функций в норме и патологии стало детальное описание отличий головного мозга человека от других приматов на молекулярном уровне.

Сравнение транскриптома тканей головного мозга человека с другими приматами

Сравнение головного мозга человека с другими приматами на уровне экспрессии генов представлено в нашей работе [207]. В этой статье мы проанализировали методом RNA-seq 422 образца мозга человека, шимпанзе, бонобо и макаки (рис. 3.14А), которые охватывали сразу 33 анатомические области мозга (рис. 3.14Б). Визуализация вариабельности экспрессии среди всех образцов выявила разделение видов в соответствии с их филогенетическим родством (рис. 3.14В).

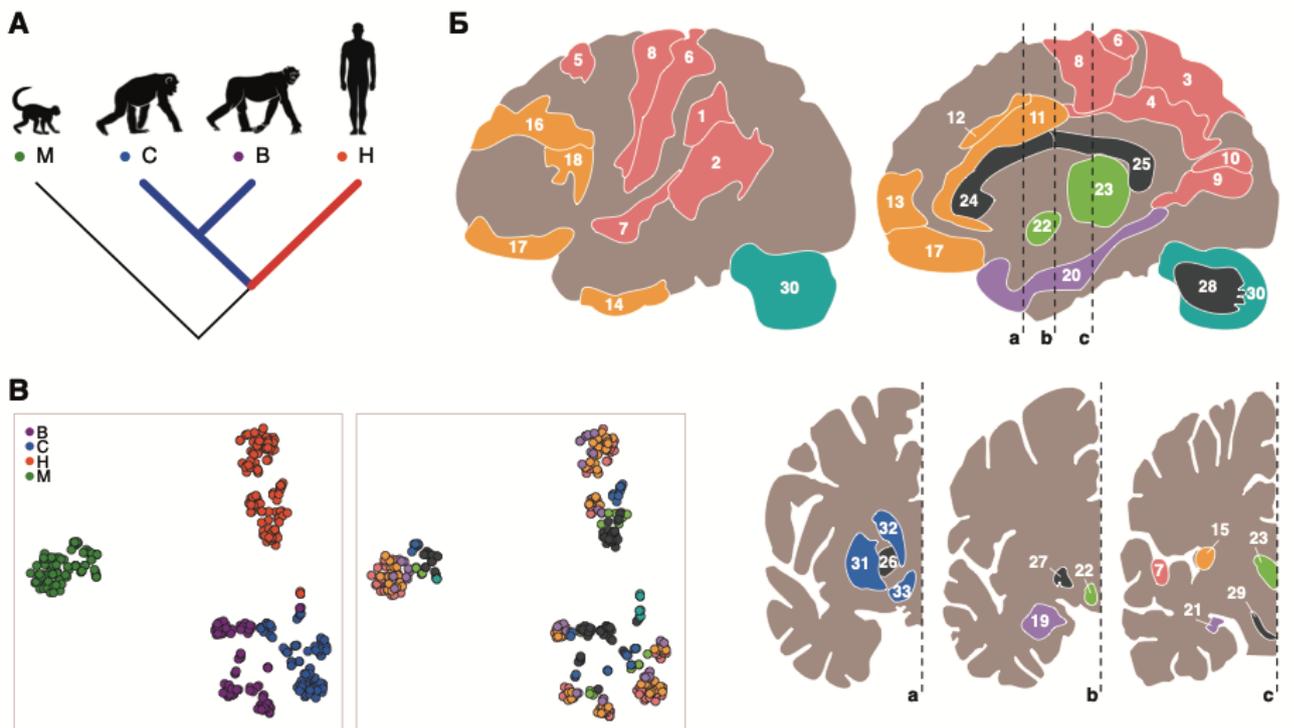


Рисунок 3.14 — Карта транскриптома мозга человека, шимпанзе, бонобо и макаки. (А) Филогенетические взаимоотношения между анализируемыми видами. Цветом выделены ветви, используемые при анализе человек-специфичности экспрессии. (Б) Анатомическая локализация 33 анализируемых областей головного мозга. Цвета обозначают кластеры регионов, определенные нами на основании подобия уровней экспрессии. (В) Графики t-SNE, основанные на вариабельности экспрессии для всех 422 проанализированных образцов. Цвета точек обозначают виды (слева) или кластеры регионов (справа). Адаптировано из [207].

Далее, используя метод ANOVA, мы определили гены со специфичной для человека экспрессией [207]. Из 11176 детектированных ортологичных генов, кодирующих белок, у 2801 были обнаружены видовые различия, с учетом области мозга (ANOVA, P-значение $< 0,00002$, с поправкой на множественное тестирование). Среди найденных 2801 различий в экспрессии, мы отнесли те, которые отличают людей от других приматов, к эволюционной линии человека. При этом распределение этих различий не было одинаковым между областями мозга. Неокортикальные области показали более высокую, чем в среднем, долю специфичных для человека различий в экспрессии. А наибольшее количество различий (1079 генов) было обнаружено в белом веществе мозжечка (рис. 3.15А). Более 95% этих различий представляли собой количественные

изменения экспрессии, но некоторые, по-видимому, отражали полную потерю экспрессии либо у людей, либо у других приматов. Важно, что количество специфических для человека различий в экспрессии в области мозга не зависело от общего количества экспрессируемых генов в этой области и было устойчивым к определению специфических для человека различий в экспрессии (см. подробнее в нашей работе [207]).

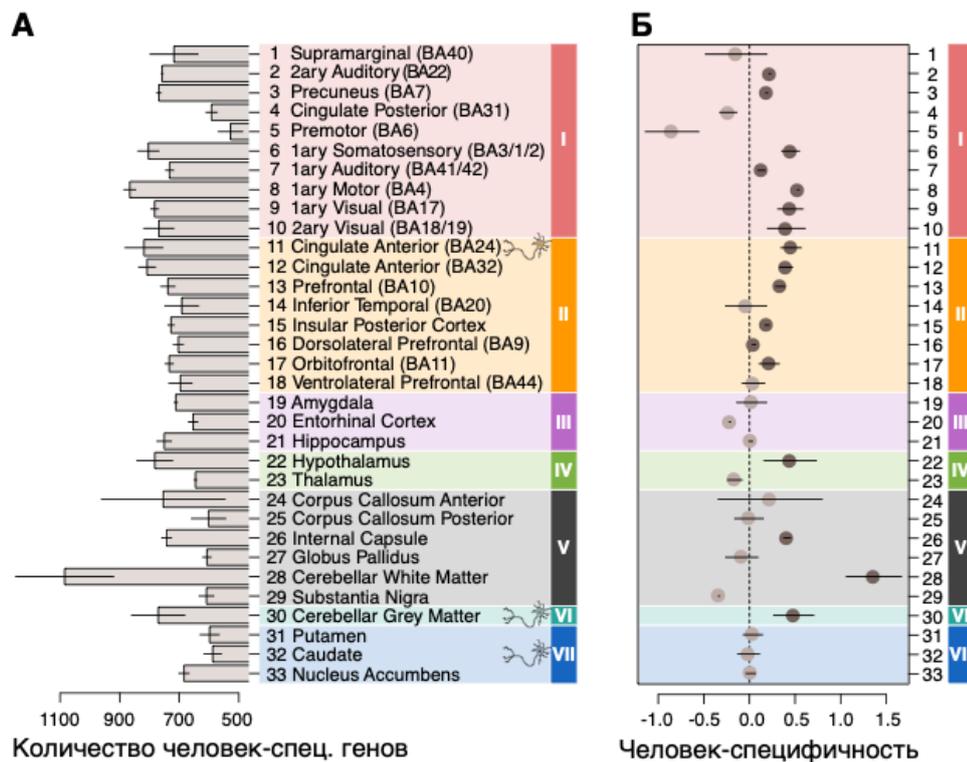


Рисунок 3.15 — Анализ специфических для человека различий в экспрессии генов, с учетом области мозга. (А) Количество генов, демонстрирующих специфические для человека различия в экспрессии в каждой области мозга.

Различия были определены как те, которые демонстрируют двукратную разницу в экспрессии между человеком и макакой по сравнению с разницей между шимпанзе и макакой или бонобо и макакой. (Б)

Человек-специфичность каждой области мозга, оцененная как соотношение различий в экспрессии, специфических для человека, и различий в экспрессии, специфических для шимпанзе или бонобо. Кружки показывают среднее значение сравнений на основе данных о шимпанзе и бонобо, а линии охватывают разницу между двумя оценками. Более темными кругами отмечены области мозга, демонстрирующие избыток специфических для человека различий в экспрессии по сравнению с обоими видами обезьян.

Адаптировано из [207].

Мы оценили человек-специфичность экспрессии в каждой области мозга как отношение числа специфичных для человека различий к количеству специфичных для шимпанзе и бонобо различий и выявили аналогичное неравномерное распределение скорости эволюции на человеческой линии. В соответствии с предыдущими работами [13; 14], в среднем больше различий было выявлено на человеческой линии ($n = 712$), чем на линиях шимпанзе ($n = 641$) и бонобо ($n = 640$). Однако, не во всех областях мозга наблюдалось ускорение эволюции на человеческой линии. В частности, мы не обнаружили избытка специфичных для человека различий в экспрессии во всех областях старой коры, а также в полосатом теле, таламусе и четырех из шести структурах белого вещества (рис. 3.15Б). Напротив, в большинстве областей новой коры, а также в гипоталамусе, внутренней капсуле, белом и сером веществе мозжечка мы наблюдали значительное превышение специфичных для человека различий (рис. 3.15Б). Важно, что этот результат оказался устойчивым к определению видоспецифичных различий в экспрессии, внутривидовой индивидуальной вариабельности и процедуре корректировки ковариатов (см. подробнее в нашей работе [207]).

Когнитивные функции, являющиеся отличительной особенностью человека как вида, задействуют многие области мозга [208]. Поэтому мы проверили, были ли гены, демонстрирующие специфичные для человека различия в экспрессии во многих областях мозга, обогащены функциональными процессами, связанными с когнитивными функциями. Действительно, специфичные для человека различия, обнаруженные более чем в 10 областях мозга, продемонстрировали значительное обогащение терминами генной онтологии (GO), связанными с функциями нейронов: синаптической передачей, регуляцией экзоцитоза, секрецией нейромедиаторов и другими (гипергеометрический тест, P -значение $< 0,01$ с поправкой на множественное тестирование) (рис. 3.16). Важно, что это обогащение оставалось стабильным при использовании разных критериев отбора генов.

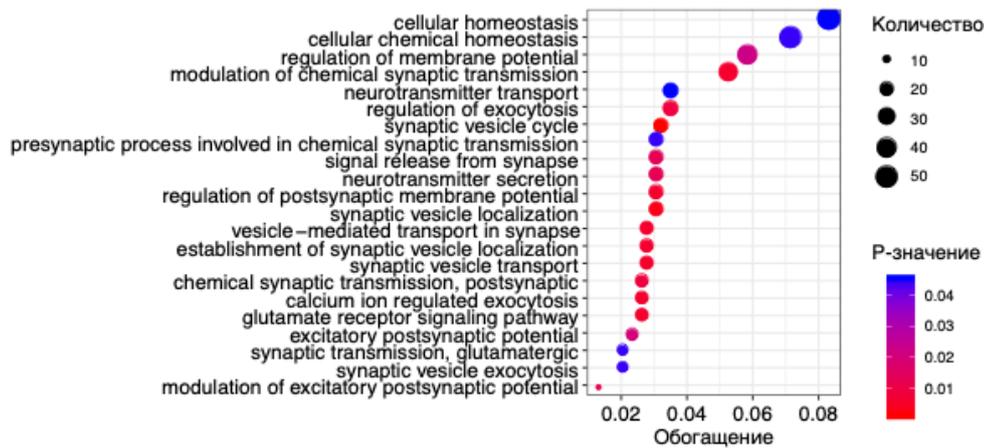


Рисунок 3.16 — Функциональные термины генной онтологии (GO), обогащенные специфически для человека различиями в экспрессии, присутствующими более чем в 10 из 33 областей мозга. Размер кружочков отражает долю генов в пределах термина GO среди всех генов, детектированных в данной области мозга. Цвет кружочков указывает на P-значение с поправкой на множественное тестирование. Адаптировано из [207].

Наконец, мы сравнили определенную нами человек-специфичность экспрессии в каждой области мозга с предыдущими работами. Несмотря на экспериментальные и вычислительные отличия проведенных другими авторами анализов, наши результаты положительно и значимо коррелировали с ранее опубликованными данными (корреляция Спирмена, P-значение $< 0,01$) [13]. Таким образом, нам удалось достоверно установить, какие области мозга наиболее отличаются у человека в сравнении с другими приматами и, соответственно, быстрее эволюционируют: кора больших полушарий, внутренняя капсула, гипоталамус, серое и белое вещество мозжечка.

Сравнение транскриптомов клеток мозга между человеком и другими приматами с разрешением в одну клетку

Для трех из 33 регионов мозга анализ также проводился с разрешением в одну клетку (snRNA-seq, рис. 3.17A). Мы секвенировали РНК из отдельных ядер методом snRNA-seq в передней поясной коре (AC), хвостатом ядре (CN)

и сером веществе мозжечка (СВ) у трех индивидуумов каждого из четырех видов. Чтобы уменьшить экспериментальные различия между видами, образцы тканей от разных видов были объединены и обработаны параллельно в каждой области мозга. Затем принадлежность ядер к видам была восстановлена вычислительным путем, основываясь на различиях в геномных последовательностях между видами. Для этого нам пришлось разработать собственную вычислительную методику, подробно описанную в работе [207]. В результате применения этой методики, нам удалось успешно разделить виды и в общей сложности было получено 88047 ядер с по меньшей мере 500 уникальными обнаруженными молекулами РНК: 7337 ± 5548 ядер на вид в каждом регионе. Из них 21% были получены из АС, 29% из CN и 50% из СВ. В пределах каждой области мозга, в среднем человеку принадлежало 37% ядер, шимпанзе - 10%, бонобо - 32% и макаке - 21%. Визуализация изменений экспрессии во всех этих ядрах выявила разделение трех областей мозга, а средние уровни экспрессии в каждой ткани каждого вида хорошо коррелировали с соответствующими данными RNA-seq, полученными нами для тканей мозга без разделения на одиночные клетки (коэффициент корреляции Пирсона = 0,83), и опубликованными ранее данными анализа РНК с разрешением в одну клетку (см. подробнее в нашей работе [207]).

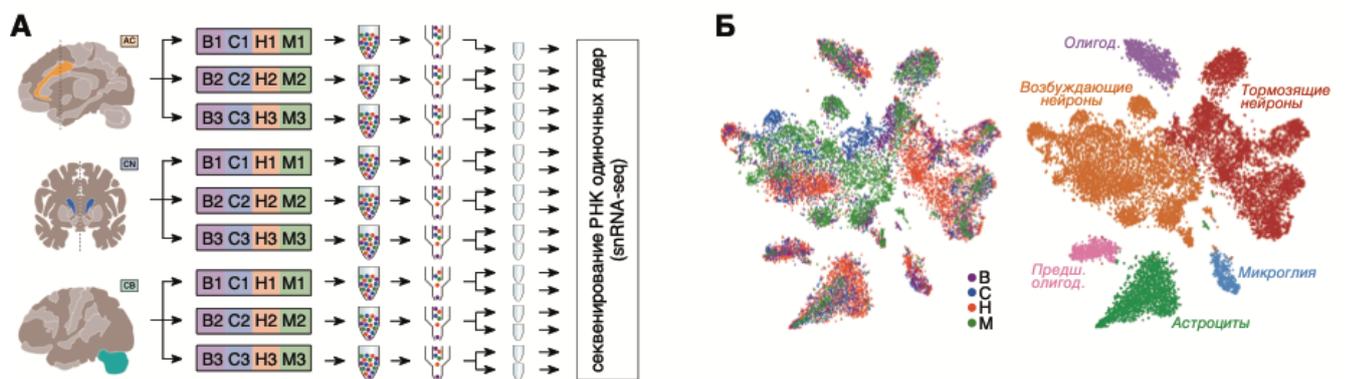


Рисунок 3.17 — Анализ экспрессии генов в мозге человека и других приматов с разрешением в одну клетку. (А) Дизайн эксперимента snRNA-seq. (Б) График t-SNE для ядер в передней поясной коре (АС), окрашенных по видам, для передней поясной коры мозга после применения методики интеграции с помощью программы Seurat 3.0 [209] (слева) и аннотация кластеров t-SNE по клеточным типам (справа). Адаптировано из [207].

Аналогичным образом, человек-специфичность, рассчитанная по данным snRNA-seq, хорошо коррелировала с человек-специфичностью, рассчитанной по

данным RNA-seq (коэффициент корреляции Пирсона = 0,58). Ядра каждого вида образовывали шесть основных кластеров в AC и CN и четыре в CB, каждый из которых был обогащен известными маркерами клеточного типа (рис. 3.17Б). Хотя мы и смогли более точно определить и охарактеризовать подтипы клеток, для целей последующего анализа мы сосредоточились на этих широких классификациях клеток, поскольку большая часть последующего анализа не могла быть достоверно выполнена при таком точном определении подтипов клеток из-за недостаточного количества ядер на подтип в наших данных.

Расчет скорости эволюции экспрессии внутри каждого типа клеток на линиях человека, шимпанзе и бонобо не выявил статистически значимого увеличения или уменьшения скорости эволюции ни в одном из проанализированных типов клеток (рис. 3.18А). Следует отдельно отметить, что широко определенные кластеры нейрональных клеток демонстрировали повышенную гетерогенность, но не показывали увеличения скорости эволюции.

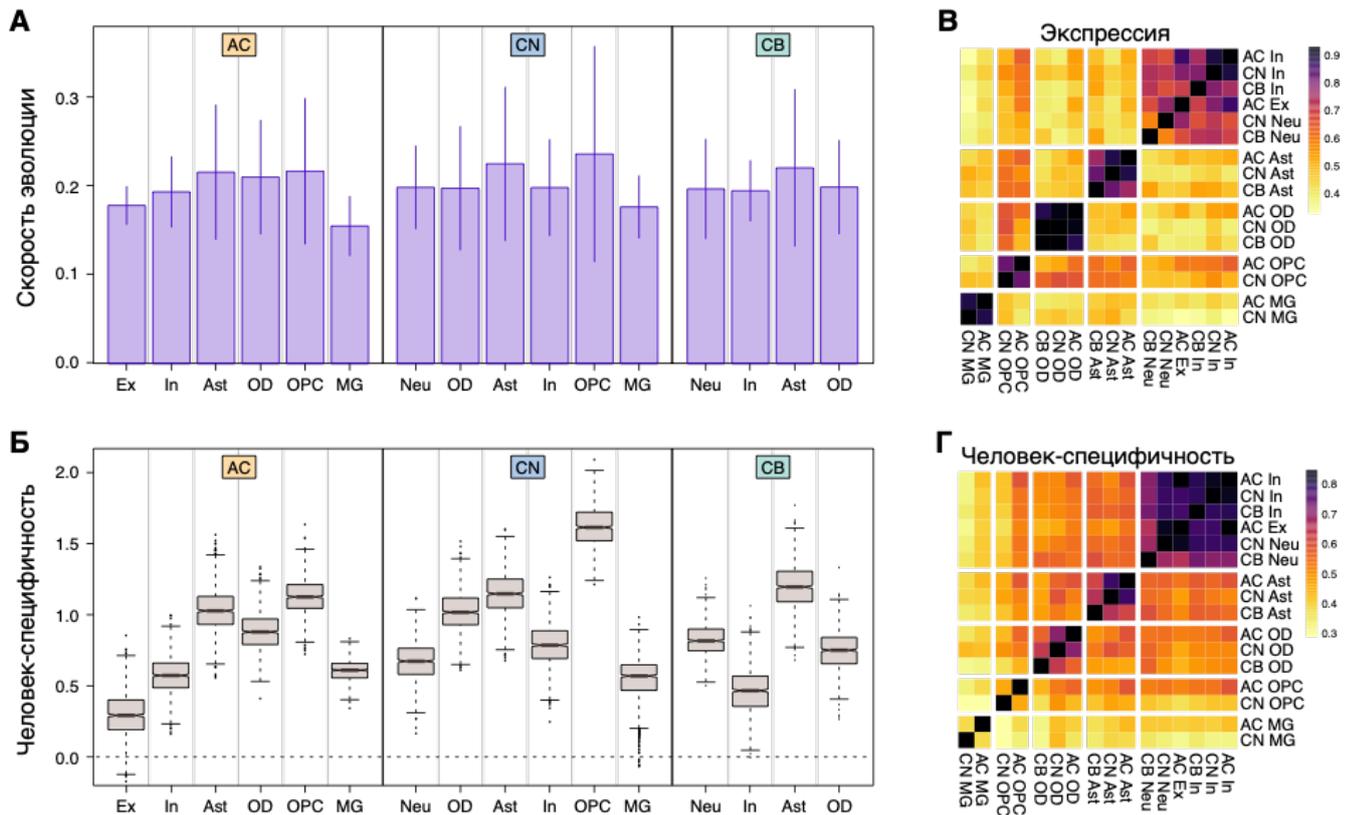


Рисунок 3.18 — Анализ эволюции экспрессии в трех областях мозга с разрешением в одну клетку. (А) Скорость эволюции разных типов клеток в каждой области мозга. Линии обозначают стандартное отклонение средних оценок. (Б) Человек-специфичность, рассчитанная для каждого типа клеток в каждой из трех областей мозга. Это значение представляет собой количество генов с экспрессией, специфичной для человека, деленное на количество генов с экспрессией, специфичной для шимпанзе и бонобо. Прямоугольники показывают медиану и первый и третий квартили распределения, а усы охватывают 1,5 межквартильных диапазона. (В) Корреляции уровней экспрессии между типами клеток, рассчитанное на основе средних уровней экспрессии генов внутри кластеров у человека. Цвета показывают коэффициенты корреляции Пирсона. In - тормозящие нейроны; Ex - возбуждающие нейроны; OPC - предшественники олигодендроцитов; Ast - астроциты; OD - олигодендроциты; MG - микроглия. (Г) Корреляции человек-специфичности между типами клеток, рассчитанные на основе сравнения с шимпанзе и бонобо в 1000 случайных наборов клеток.

Адаптировано из [207].

Хотя скорость эволюции существенно не различалась между типами клеток, сравнение этих скоростей между эволюционными ветвями выявило суще-

ственные различия (рис. 3.18Б). В частности, нейрональные клетки, как правило, показывали меньшую человек-специфичность по сравнению с другими типами клеток (рис. 3.18Б). Напротив, астроциты и клетки-предшественники олигодендроцитов стабильно демонстрировали наибольшую человек-специфичность, т.е. избыток специфичных для человека различий в экспрессии, во всех исследованных областях мозга (рис. 3.18Б). Важно отметить, что оценки человек-специфичности, рассчитанные независимо с использованием данных шимпанзе или бонобо, хорошо согласовывались друг с другом.

Интересно, что экспрессия генов внутри каждого типа клеток хорошо коррелировала между разными областями мозга (рис. 3.18В). Напротив, человек-специфичность экспрессии хорошо коррелировала между подтипами нейронов, за исключением гранулярных клеток мозжечка, но не между астроцитами, олигодендроцитами, предшественниками олигодендроцитов или микроглией, происходящими из разных областей мозга (рис. 3.18Г). Этот результат указывает на то, что большинство подтипов нейронов могут иметь общие особенности экспрессии, специфичные для человека.

Таким образом, анализ человек-специфичной экспрессии в разных типах клеток мозга позволил установить, что олигодендроциты и астроциты у человека отличаются от подобных клеток у остальных приматов больше, чем основные клетки мозга – нейроны, что указывает на важность глиальных клеток для возникновения новых функций головного мозга у человека.

Роль некодирующих РНК в эволюции мозга человека

Однако не только матричные РНК важны для функционирования мозга человека. Длинные некодирующие РНК (lncRNA) участвуют в процессах развития нервной системы, которые, как предполагается, лежат в основе эволюции человеческого мозга, включая пролиферацию, синаптогенез и нейропластичность [210]. В одной из работ [211] мы изучали длинную некодирующую РНК Falcor/LL35 и, применив разработанные нами методы анализа транскриптомных данных (см. в разделе 2.2.5), липидомных и метаболомных данных (см. в разделе 2.2.6), а также методы интеграции разных типов омиксных данных,

мы показали, что длинная некодирующая РНК Falcor/LL35 играет важную регуляторную роль в метаболизме липидов. Важно отметить, что сделанные нами биоинформатические наблюдения были подтверждены экспериментально коллегами из группы проф. Тимофея Зацепина (Сколтех, Россия) *in vitro* и *in vivo*.

Не только длинные некодирующие РНК, но и другие классы некодирующих РНК важны для функционирования клеток мозга. В частности, малые ядерные и ядрышковые РНК (мяРНК и мякРНК) являются функционально и эволюционно консервативными единицами, участвующими в процессинге транскриптов. В работе [212] мы исследовали эволюцию экспрессии мяРНК и мякРНК, измерив их экспрессию во фронтальной коре человека, шимпанзе, макака и мышей. Мы охарактеризовали экспрессию транскриптов длиной от 100 до 200 п.о. путем секвенирования фракции тотальной РНК соответствующей длины. Для каждого вида было проведено секвенирование РНК (RNA-seq) в трех отдельных образцах. 69,5% прочтений удалось однозначно картировать на геном соответствующего вида. Половина этих прочтений соответствовала аннотированным мяРНК (1,3%) и мякРНК (52,9%). Остальные прочтения соответствовали транскриптам, кодирующим белок (10,9%), удержанным интронам (2,5%), другим типам аннотированных транскриптов (13,5%) и неаннотированным областям генома (18,9%). Далее, мы провели сравнение экспрессии некодирующих РНК между видами приматов. Экспрессия мяРНК оказалась в значительной степени консервативна, а 44% из 185 детектированных мякРНК и 40% из 134 семейств мякРНК показали значимую разницу экспрессии между видами (рис. 3.19А). Отнесение этих различий в экспрессии к эволюционным линиям позволило определить 5 изменений экспрессии, специфичных для человека, 4 - для шимпанзе, 12 - к линии обезьян, 13 - к линии макака и 46 - к линии, соединяющей приматов и мышей. в том числе, мы наблюдали радикальные изменения экспрессии, уникальные для человека: 10-кратное повышение экспрессии U1 мяРНК и 1000-кратное снижение экспрессии SNORA29 (рис. 3.19Б).

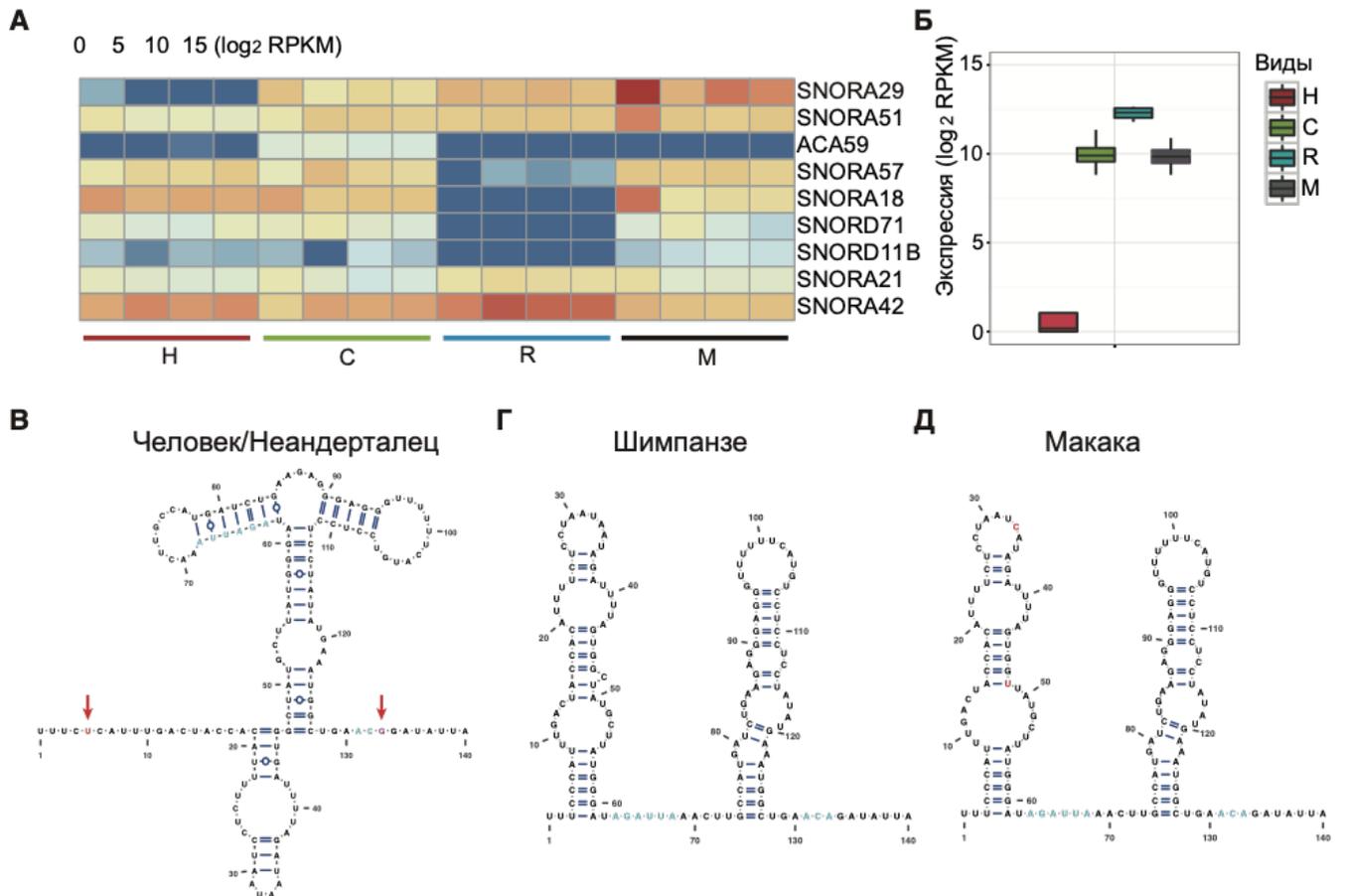


Рисунок 3.19 — Специфичный для человека паттерн экспрессии SNORA29 и лежащий в его основе генетический механизм. (А) Уровни экспрессии различных мякРНК, показывающие изменения экспрессии, характерные для эволюционных линий человека, шимпанзе и макаки. (Б) Экспрессия SNORA29 у четырех видов. (В-Д) Предсказания вторичной структуры SNORA29 у людей/неандертальцев, шимпанзе и макак-резусов. Красные стрелки указывают на мутации, характерные для человека. Адаптировано из [212].

Подобно многим другим мякРНК, SNORA29 имеет интронную локализацию: ее последовательность находится в пределах четвертого интрона гена TCP1. Мы оценили экспрессию гена TCP1 во фронтальной коре головного мозга взрослых людей, шимпанзе, макак и мышей, используя публично доступный набор данных RNA-seq [213]. Существенных различий в экспрессии TCP1 между четырьмя видами не оказалось. Кроме того, экспрессия другой мякРНК, расположенной в интроне гена TCP1, SNORA20, также является высоко консервативной. Таким образом, резкое снижение уровня представленности SNORA29 на человеческой линии не может быть объяснено изменением экспрессии ее гена-хозяина.

Сравнив последовательность SNORA29 человека с консенсусной последовательностью ее гомологов у других приматов, мы идентифицировали два нуклеотидных изменения: замену А на G в положении 133 и замену С на Т в положении 5 (рис. 3.19В-Д). Эти изменения затрагивают консервативные элементы вторичной структуры мРНК и, соответственно, существенно влияют на вторичную структуру SNORA29 (рис. 3.19В). Примечательно, что эти две специфичные для человека мутации приводят к необычно большому увеличению свободной энергии вторичной структуры SNORA29, что указывает на резкое снижение ее стабильности [212]. Поэтому мы выдвинули гипотезу, что снижение стабильности вторичной структуры SNORA29, а не специфическая для человека регуляция экспрессии SNORA29, может привести к снижению ее распространенности в мозге человека. Предсказания вторичной структуры SNORA29 и расчеты ее стабильности, выполненные в работе [212], позволили подтвердить эту гипотезу.

Регуляция альтернативного сплайсинга в клетках мозга

Вторичная структура пре-мРНК влияет на многие клеточные процессы, в том числе на сплайсинг в генах, связанных с функционированием мозга человека. Современная парадигма предсказания структуры РНК смещена в сторону вторичных структур, которые находятся недалеко друг от друга, хотя структуры дальнего действия, как известно, так же важны. В работе [214] мы разработали эффективный метод обнаружения консервативных структур РНК на полногеномном масштабе, который не требует множественного выравнивания последовательностей и одинаково хорошо работает для обнаружения локальных и дальних структур. Мы нашли структуры РНК, которые могут быть вовлечены в регуляцию сплайсинга у млекопитающих, и продемонстрировали сильную связь между возникновением консервативных структур РНК и альтернативным сплайсингом, где локальные структуры РНК, как правило, чаще находятся на альтернативных донорных сайтах сплайсинга, а дальние структуры в большей степени связаны со слабыми альтернативными акцепторными сайтами сплайсинга. Мы валидировали разработанный метод эксперименталь-

но с использованием мини-генов на примере гена SF1 человека, который играет ключевую регуляторную роль в поддержании энергетического гомеостаза в липидного метаболизма, высоко экспрессируясь в нейронах гипоталамуса (в так называемых SF-1 нейронах) [215]. Важно отметить, что белок SIRT6, регуляторная роль которого в клетках мозга уже обсуждалась выше (см. раздел 3.1.1), также участвует в поддержании энергетического гомеостаза в SF-1 нейронах в гипоталамусе [216].

Учитывая эту взаимосвязь, а также участие белка SIRT6 в развитии болезни Альцгеймера - нейродегенеративного заболевания, в механизме возникновения которого уже показано участие нарушений сплайсинга, мы решили проанализировать эффект нокаута SIRT6 на сплайсинг в клетках головного мозга на модели мышей с мозг-специфичным нокаутом (SIRT6-KO) из нашей работы [154]. Чтобы получить количественную оценку изменений ландшафта сплайсинга в результате SIRT6-KO, мы проанализировали данные RNA-seq для образцов коры мозга 5 мышей дикого типа и 5 SIRT6-KO мышей из нашей работы [154]. Методы обработки этих данных описаны выше в разделе 2.2.5, а анализ ДЭ генов при SIRT6-KO - в разделе 3.1.1. Для анализа альтернативного сплайсинга, мы применили программу rMATS [217] и получили список из 256 событий альтернативного сплайсинга.

Наиболее распространенным типом событий альтернативного сплайсинга, как и ожидалось из предыдущих исследований, был пропуск экзона (SE), составляющий 54,7% всех событий альтернативного сплайсинга. Следующим наиболее распространенным типом событий было удержание интрона (RI) с 25,8% всех наблюдаемых событий альтернативного сплайсинга, альтернативный 3' (акцепторный) сайт сплайсинга (A3SS) и альтернативный 5' (донорный) сайт сплайсинга (A5SS), оба с 7,8% всех наблюдаемых событий альтернативного сплайсинга, и, наконец, взаимоисключающие экзоны (MXE) с 3,9% всех наблюдаемых событий альтернативного сплайсинга. Чтобы оценить функциональную значимость изменений ландшафта альтернативного сплайсинга при SIRT6-KO, был проведен анализ обогащения терминов генной онтологии (GO), который показал наиболее значимое обогащение в термине "positive regulation of neuron differentiation". Среди событий альтернативного сплайсинга, связанных с этим термином генной онтологии, было несколько особенно интересных генов. Так, события альтернативного сплайсинга наблюдались в генах Ahil, Ptprd и Gdil,

все из которых так или иначе связаны с отклонениями в умственном развитии. Другой группой, представляющей особый интерес, являются гены *Lgr8*, *Dbn1*, *Ptprn5*, поскольку они непосредственно связаны с болезнью Альцгеймера.

Но какой механизм возникновения у наблюдаемых событий альтернативного сплайсинга? Чтобы ответить на этот вопрос, мы проанализировали экспрессию факторов сплайсинга при SIRT6-КО и нашли значимые изменения экспрессии у 11 из 70 рассматриваемых факторов сплайсинга, причем все они были в сторону повышения при нокауте. Любопытно, что 7 из 11 ДЭ факторов сплайсинга являются членами семейства белков hnRNP - *Hnrnp3*, *Hnrpa0*, *Hnrpm*, *Hnrpa3*, *Hnrpu*, *Hnrpd* и *Hnrpp2*, которые важны для регуляции сплайсинга, а также для стабилизации мРНК и регуляция транскрипционных и трансляционных процессов. Важно отметить, что белки hnRNP, как полагают, являются в основном ингибирующими факторами сплайсинга, а значит наблюдаемая повышенная экспрессия факторов семейства hnRNP может частично объяснить наблюдаемое увеличение удержания интронов, которое было продемонстрировано по крайней мере у одного члена семейства hnRNP [218]. Таким образом, хотя участие белка SIRT6 в регуляции сплайсинга в мозге ещё требует экспериментального подтверждения, наши вычислительные предсказания указывают на его потенциально важную роль в этом процессе.

Эволюция липидного состава тканей мозга человека

Наши исследования экспрессии генов, некодирующих РНК, альтернативного сплайсинга позволили понять детали эволюции экспрессии генов на таком уровне детализации, какой ранее был недоступен. Однако глубокое понимание механизмов функционирования мозга невозможно без проведения всестороннего анализа его молекулярной организации. Важный уровень молекулярной организации мозга, которому до наших работ не было уделено внимания, - это уровень небольших молекул, метаболитов и липидов, которые непосредственно вовлечены во все функциональные процессы в мозге. Липиды составляют наибольшую часть сухого вещества мозга и изменения их концентраций наблюдаются при различных психических расстройствах, например, при аутизме и

шизофрении. Помимо ассоциации с психическими расстройствами, липиды оказались важны для становления *Homo sapiens* как отдельного вида.

Мы сравнили концентрации липидов у человека и более 30 видов млекопитающих в работе [219]. Для этого мы измерили содержание липидов в шести тканях 32 видов, представляющих приматов, грызунов и летучих мышей (всего 669 образцов). Исследовались образцы тканей из коры больших полушарий головного мозга, почек, сердца, мышц, печени и мозжечка. Анализ видо-специфичных различий в содержании липидов показал, что линия человека в целом отличается наибольшими отличиями в содержании липидов по сравнению с другими исследованными линиями млекопитающих (рис. 3.20, справа). В дополнение к тому, больше всего специфичных эволюционных изменений было обнаружено у человека. При этом оказалось, что у человека липидные изменения локализованы преимущественно в мозге (рис. 3.20, слева).

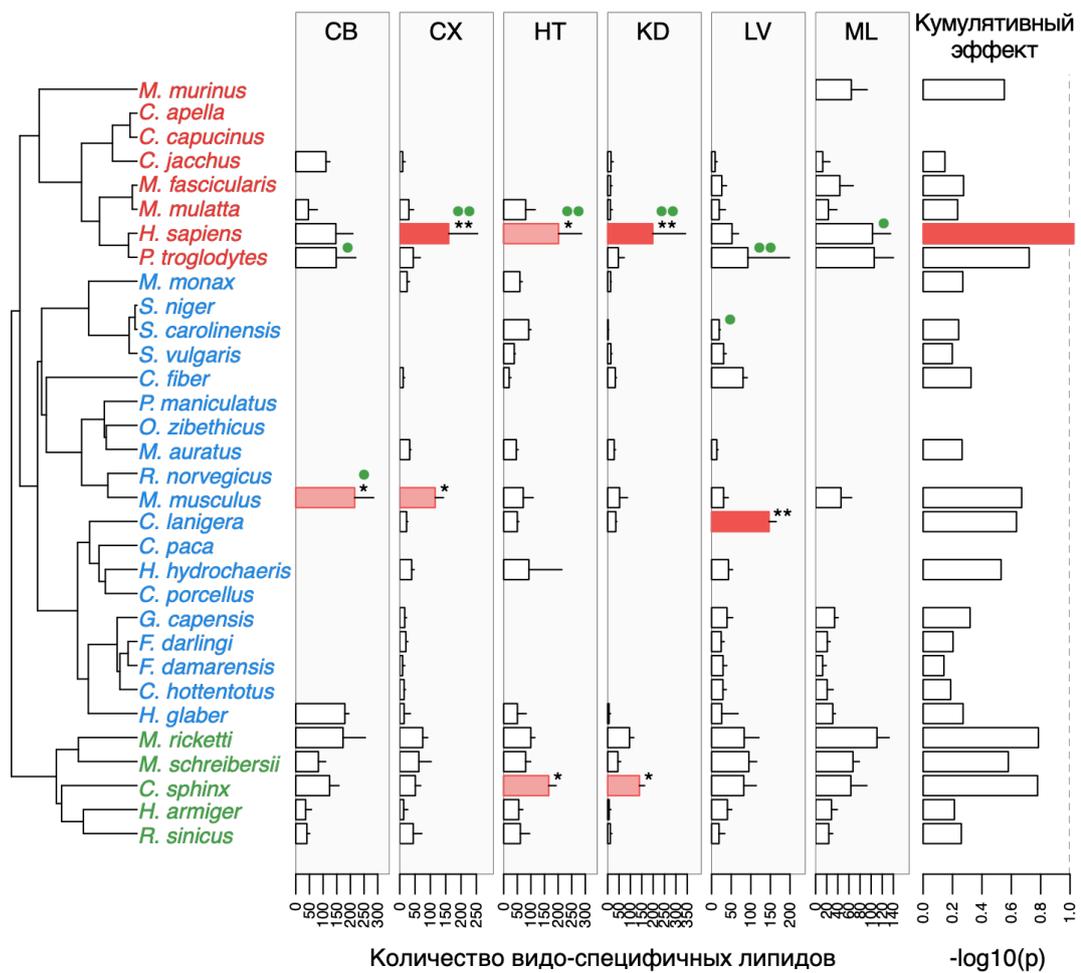


Рисунок 3.20 — Различия в концентрации липидов в зависимости от вида для коры больших полушарий головного мозга (CX), почек (KD), сердца (HT), мышц (ML), печени (LV) и мозжечка (CB). Показано количество липидов, демонстрирующих значительные видоспецифические различия в их содержании в каждой из 104 комбинаций тканей и линий (нормализовано на филогенетические расстояния). Усы показывают вариабельность оценок, рассчитанных путем случайной выборки из трех особей каждого вида. Цвета звездочек и столбиков указывают на значимость разницы между наблюдаемым распределением чисел и случайным ожидаемым (пермутационный тест, ** и красный — P -значение $< 0,05$, * и розовый — P -значение $< 0,1$). В крайнем правом столбце показан кумулятивный эффект на эволюционной линии, рассчитанный как среднее $\log_{10}(P)$ разницы между наблюдаемым и случайным количеством видоспецифических липидов в ткани. Адаптировано из [219].

Специфичные для человека липиды в коре головного мозга были значимо обогащены в двух классах липидов: глицерофосфоэтаноламины и дира-

диглицерины (гипергеометрический тест, P -значение $< 0,0001$ с поправкой на множественное тестирование). Аналогичным образом, ферменты, связанные со специфичными для человека липидами, продемонстрировали превышение специфичных для человека изменений экспрессии в коре головного мозга (t -тест, P -значение $= 0,0004$) и были значимо перепредставлены в 11 путях KEGG (P -значение $< 0,0001$ с поправкой на множественное тестирование).

Ферменты, связанные со специфичными для мозга липидами, показали повышенную консервативность, в комбинации с повышенным давлением очищающего отбора, что указывает на их функциональную важность, несмотря на снижение уровня экспрессии в мозге (тест Уилкоксона, $p < 0,0001$). В целом, наши наблюдения показывают, что концентрации липидов изменяются по-разному по сравнению с последовательностями генома и уровнями экспрессии генов. Мы предполагаем, что эволюция липидов отражает фенотипические различия между видами более точно, чем различия в геноме и экспрессии генов, причем большая доля различий, накапливающихся среди видов, представляет собой функциональные изменения, общие для тканей, в то время как тканеспецифичные различия в липидоме могут представлять собой видоспецифичные адаптации.

Интересно, что найденные нами человек-специфичные липиды преимущественно вовлечены в метаболические пути, связанные с когнитивными заболеваниями, такими как болезнь Альцгеймера, Паркинсона, и другими нарушениями работы нервной системы. Получается, эти липиды важны для когнитивной деятельности - отличительной черты человека как вида. Кроме того, на основании того же самого набора данных мы пришли к выводу, что липидный метаболизм оказывает существенное влияние на максимальную продолжительность жизни того или иного вида [220].

Продолжая исследовать важность липидов для функционирования головного мозга человека в целом и для поддержания человек-специфичных процессов в мозге в частности, в работе [221] мы впервые проанализировали липидный состав в образцах префронтальной коры 14 представителей трех человеческих популяций (китайской, европейской и африканской), а также 14 взрослых шимпанзе, и показали, что концентрации липидов значимо изменялись от шимпанзе к человеку в метаболических путях, связанных с катаболизмом липидов, у европейцев по сравнению с другими популяциями (рис. 3.21А). Анализ экспрессии

генов по данным RNA-seq, полученным в тех же образцах человека и шимпанзе, подтвердил наши наблюдения на уровне липидома: 6 генов из пути катаболизма липидов показали максимальное расхождение в экспрессии с шимпанзе у европейцев, промежуточное у азиатов и отсутствующее у африканцев (рис. 3.21Б). Таким образом, липидные отличия, обнаруженные у европейцев, по-видимому, связаны с изменениями уровней экспрессии соответствующих метаболических ферментов. В поисках возможных причин наблюдаемых отличий, мы обратились к анализу геномных последовательностей генов липидного катаболизма.

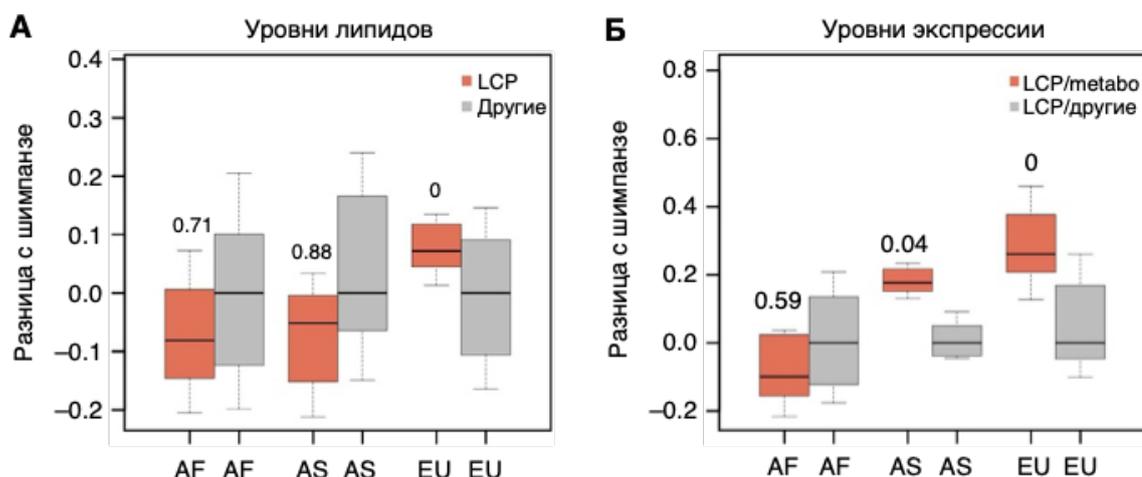


Рисунок 3.21 — Сравнение уровней экспрессии и содержания липидов в префронтальной коре человека и шимпанзе. (А) Отличия содержания липидов между популяциями шимпанзе и людей африканского (AF, n=4), азиатского (AS, n=5) и европейского (EU, n=5) происхождения для участников пути катаболизма липидов (LCP, красный, n=1090 масс-спектрометрических пиков) и метаболитов в других метаболических путях (серый, n=163 масс-спектрометрических пика). Цифры на графике показывают P-значения (пермутационный тест). Все прямоугольники на этой и других панелях показывают квантили и медиану, а усы охватывают 0,5 межквартильного диапазона. (Б) Отличия экспрессии генов между популяциями шимпанзе и человека для генов LCP, непосредственно связанных с липидами из категории LCP, показанными на панели А (красный, n=6 экспрессируемых генов) и остальных генов из категории LCP (серый, n=26 экспрессируемых генов). Оценка значимости была проведена таким же образом, как и на панели А.

Адаптировано из [221].

Поскольку уже было известно, что современные европейцы унаследовали от неандертальцев несколько процентов генома, мы хотели проверить гипотезу

о том, что наблюдаемые изменения липидного состава у европейцев могут быть связаны с событием скрещивания человека с неандертальцами. Мы искали сходства с геномом неандертальца в геномах 11 современных популяций человека из проекта "1000 геномов": трех популяций африканского происхождения — индивидуумов африканского происхождения с юго-запада США (ASW), представителей народов лухья (LWK) и йоруба (YRI); трех популяций восточноазиатского происхождения - китайцев—ханьцев из Пекина (CHB), китайцев—ханьцев из Южного Китая (CHS) и японцев (JPT); и пяти популяций европейского происхождения — индивидуумов из Центральной Европы (CEU), финнов из Финляндии (FIN), британцев из Англии и Шотландии (GBR), иберов из Испании (IBS) и тосканцев (TSI). Для каждой пары этих популяций, последовательности генома человека сравнивались с последовательностями генома неандертальца и шимпанзе. Мы использовали последовательность генома неандертальца с высоким покрытием, полученную из кости одного индивидуума, найденной в 2008 году в восточной галерее Денисовой пещеры в Горном Алтае (Altai) [222], в сочетании с последовательностью генома неандертальца с низким покрытием, полученной от трех индивидуумов (Vindjia) [223]. С использованием этих двух последовательностей, мы реконструировали консенсусную последовательность генома неандертальца путем исключения всех участков с вариациями последовательности у отдельных индивидуумов неандертальца. Далее мы использовали референсный геном шимпанзе в сочетании с данными о последовательностях геномов десяти шимпанзе, чтобы исключить участки, переменные между особями шимпанзе [224]. В анализе были использованы только отдельные позиции генома, демонстрирующие различия в последовательностях между шимпанзе [224] и двумя геномами неандертальцев [222; 223] (суммарно 1158559 позиций).

Чтобы оценить вклад неандертальских генотипов в геномы современных людей, мы использовали метод D-статистики, как описано в работе [223]. В соответствии с предыдущими исследованиями [223], геномы современных людей европейского и азиатского происхождения показали большее сходство с геномом неандертальца, чем геномы трех популяций африканского происхождения. В среднем частота неандертальских вариантов (Neanderthal-like sites, или NLS) составила $6,1 \pm 0,2\%$ у современных людей европейского и азиатского происхождения, что указывает на значительный избыток NLS в современных популяциях за пределами Африки (рис. 3.22А,Б, синие полосы). Эта оценка

D-статистики аналогична тем, о которых сообщалось в других исследованиях ($4,8 \pm 0,2\%$) [223], причем более высокие значения, полученные в нашем исследовании, потенциально связаны с дополнительной фильтрацией полиморфных участков генома у неандертальцев в нашем варианте анализа. Кроме того, в соответствии с другими исследованиями [223], мы не наблюдали существенной разницы в частотах NLS в среднем по геному между европейскими и азиатскими популяциями, с небольшой тенденцией к более высоким частотам у азиатов: $5,9 \pm 0,08$ и $6,2 \pm 0,06\%$.

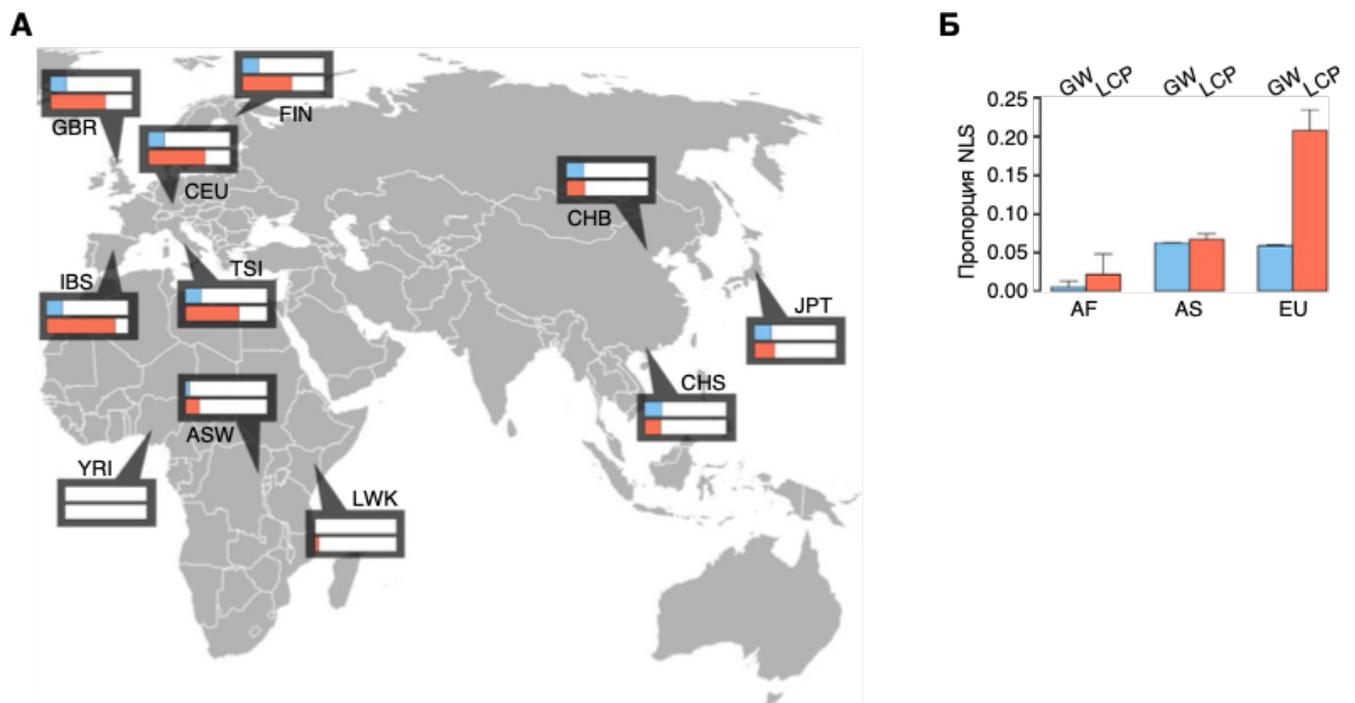


Рисунок 3.22 — Пропорции NLS в современных человеческих популяциях. (А)

Геномные расстояния между 11 популяциями современного человека и неандертальцами; синий - по всему геному; красный - в генах LCP.

Максимальная длина полосы соответствует частоте NLS, равной 30%. (Б) Средние доли NLS в современных популяциях Африки (AF), Европы (EU) и Азии (AS), рассчитанные на основе данных о последовательностях из проекта "1000 геномов"; синий: по всему геному (GW, 1158559 сайтов), красный: в генах LCP (LCP, 498 сайтов). Усы показывают стандартное отклонение.

Адаптировано из [221].

Хотя доля неандертальских фрагментов генома примерно одинакова у всех современных людей (кроме тех, которые имеют африканское происхождение), эти фрагменты распределены неравномерно в геноме. Мы показали, что NLS выше в три раза в генах, участвующих в катаболизме липидов (Lipid

Catabolic Process, или LCP), причем только у европейцев ($20,8 \pm 2,6\%$), но не у азиатов ($6,7 \pm 0,7\%$) (рис. 3.22А,Б, красные полосы), по сравнению с усредненным по всему геному значением ($6,1 \pm 0,2\%$). Избыток NLS в генах LCP в геномах современных европейцев может быть обусловлен быстрым распространением неандертальских аллелей у европейских предков из-за адаптивного значения LCP генов. В частности, можно выдвинуть гипотезу, что со временем неандертальцы приобрели изменения в катаболизме липидов, которые были полезны для выживания в условиях окружающей среды доисторической Европы и Центральной Азии. Эти адаптивные варианты, возможно, были приобретены современными людьми в результате интрогрессии и распространились с высокой частотой в результате положительного отбора.

Чтобы проверить эту гипотезу, мы оценили признаки положительного отбора в геномах современных людей европейского, азиатского и африканского происхождения, используя комбинированную метрику множественных сигналов (CMS) из работы [225], высокие значения которой указывают на области генома, подвергшиеся недавнему положительному отбору. И действительно, мы обнаружили значимое обогащение высоких значений CMS в генах LCP у современных европейцев, но не у азиатов или африканцев (рис. 3.23А). Причем этот эффект был специфичен для генов LCP: не наблюдалось существенного превышения значений CMS у индивидуумов европейского происхождения в сопоставимых областях генома, содержащих другие метаболические гены. Кроме того, в рамках функциональной категории LCP, высокие значения CMS, обнаруженные у современных европейцев, были связаны с генами, содержащими избыток NLS, но не были связаны с другими генами LCP (тест Уилкоксона, P-значение=0,0003).

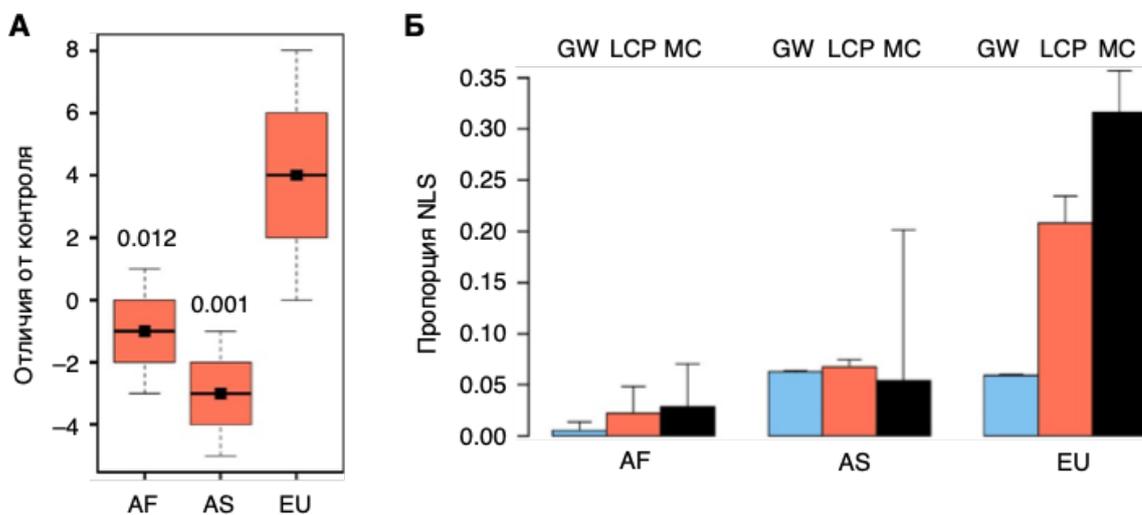


Рисунок 3.23 — Положительный отбор в генах LCP. (А) Сигналы положительного отбора в генах LCP оценивались с использованием метрики CMS. Черные квадраты показывают частоту сайтов с повышенными значениями CMS (>1), потенциально указывающую на области генома, подвергшиеся недавнему положительному отбору в генах LCP, нормализованную на частоту таких сайтов во всех аннотированных генах в пределах одной популяции. Прямоугольники показывают распределение (квартили и медиану) нормализованных оценок частот сайтов, полученных с помощью 1000 бутстрепов по 38 генам LCP, а усы охватывают 0,5 межквартильного диапазона. Цифры на графике показывают Р-значения (пермутационный тест) в сравнении популяций Африки (AF) или Азии (AS) с популяциями Европы (EU). (Б) Средние доли NLS в современных популяциях AF, EU, AS, рассчитанные на основе данных о последовательностях из проекта "1000 геномов"; синий - весь геном (GW, 1158559 сайтов); красный - все гены LCP (LCP, 498 сайтов); черный - гены LCP, связанные со специфичными для Европы изменениями содержания метаболитов (MC, 114 сайтов). Усы показывают стандартное отклонение. Адаптировано из [221].

Примечательно, что шесть генов LCP, связанных со специфичными для европейцев метаболическими изменениями, содержали ещё более высокую долю NLS у европейцев ($31,6 \pm 4,1\%$), чем все гены LCP ($20,8 \pm 2,6\%$) (рис. 3.23Б). Кроме того, эти NLS не были распределены равномерно внутри генов, а группировались вблизи сайтов старта транскрипции, что позволяет предположить их важную роль в возникновении изменений уровней экспрессии генов, наблюдаемых у европейцев (тест Уилкоксона, Р-значение=0,048). Напротив, у индивиду-

умов азиатского и африканского происхождения не наблюдалось какого-либо существенного избытка NLS в тех же областях генов (рис. 3.23Б). Таким образом, совокупность этих наблюдений позволяет предположить, что по этим генам шел положительный отбор у современных европейцев, но не у азиатов и африканцев, и, вероятно, гены липидного катаболизма давали европейцам эволюционное преимущество, важное для адаптации к условиям окружающей среды доисторической Европы.

Далее, мы перешли к более подробному анализу неандертальского вклада в современных популяциях. В работе [226] были изучены особенности генетики и происхождения популяции современных кетов. Кеты, этническая группа в бассейне реки Енисей, считаются последними кочевыми охотниками-сборителями Сибири. Популяция кетов до нашей работы была мало изучена с генетической точки зрения. В частности, оставалось неясным, к какой языковой семье принадлежит язык кетов. Связь между кетами и популяциями Сибири и Северной Америки (в частности, древними геномами мальты и палео-эскимосов) также не была исследована. Однако, на основании предварительных антропологических исследований, была выдвинута гипотеза о том, что в истории популяции кетов могло произойти второе событие скрещивания с неандертальцами и/или денисовцами. В составе большой коллаборации, мы исследовали частоты неандертальских и денисовских вариантов в популяции кетов методом D-статистики, как в обсуждаемой выше нашей работе [221]. Поскольку мы не наблюдали значимого превышения содержания неандертальских или денисовских вариантов по сравнению с другими современными популяциями европейского или азиатского происхождения, можно сделать вывод, что в истории популяции кетов не было повторного события скрещивания с неандертальцами или денисовцами [226].

После завершения этой работы, мы перешли к исследованию другой интересной и мало изученной ранее популяции - популяции папуасов. Секвенирование полных ядерных геномов неандертальца и денисовца стимулировало исследования об их родстве с современными людьми, демонстрируя, в частности, что аллели ДНК как из геномов неандертальца, так и денисовца присутствуют в геномах современных людей. Геном папуасов является уникальным объектом, поскольку он содержит как неандертальские, так и денисовские аллели. В нашей работе [227] мы применили метод D-статистики, ранее апробированный в

обсуждаемой выше нашей работе [221], чтобы оценить вклад неандертальских и денисовских генотипов в геномы современных папуасов. В соответствии с результатами предыдущих исследований, в геноме папуасов наблюдался сходный процент неандертальских и денисовских вариантов в среднем по всему геному.

Далее, мы поставили цель проверить гипотезу о неравномерности распределения неандертальских и денисовских вариантов по геному папуасов. Для этого, мы рассчитали D-статистику для отдельных функциональных групп генов и обнаружили больше функциональных групп, обогащенных неандертальскими вариантами (35), чем денисовскими вариантами (8) (биномиальный тест, P-значение $< 10^{-5}$). Анализ этих групп показал, что папуасы, как правило, имеют неандертальские варианты в генах, регулирующих транскрипцию и нейрогенез, а также в генах, участвующих в метаболических процессах, в то время как денисовские варианты обогащены в генах, регулирующих процессы в плазматической мембране и восстановление тканей и костей. Таким образом, варианты, унаследованные папуасами от неандертальцев и денисовцев, по-видимому, обогащены в различных функциональных группах генов. Это наблюдение подтверждает гипотезу о различных, независимых потоках аллелей из геномов неандертальца и денисовца в геном папуаса и позволяет сделать вывод, что сохранение аллелей неандертальцев и денисовцев в геномах папуасов могло быть связано с адаптацией.

В дополнение к анализу геномов популяций, которое было призвано объяснить наше изначальное наблюдение о значимых липидных отличиях в префронтальной коре европейцев, сделанное в нашей работе [221] (см. выше), мы продолжили детально исследовать отличия в липидном и метаболическом составе головного мозга между популяциями. Для этого мы измерили содержание 1670 липидов и 258 метаболитов у 146 китайцев-ханьцев, 97 западноевропейцев и 60 афроамериканцев разного возраста с провели сравнительный анализ в нашей работе [228]. Применение методов статистического анализа и моделей логистической регрессии продемонстрировало значимые различия в уровнях липидов и метаболитов у представителей китайской народности Хань по сравнению с двумя другими популяциями. Степень различий зависела от возраста, достигая максимума у молодых людей. Функциональный анализ липидов и метаболитов, специфичных для китайской популяции, выявил 35 значимо обогащенных KEGG путей, в основном связанных с липидным метаболизмом [228].

Нарушения эволюционно новых метаболических путей в мозге человека

Поскольку когнитивные функции, нарушенные при психических расстройствах, являются одними из ключевых поведенческих способностей, которые отделяют людей от других видов, психические расстройства могут представлять собой нарушение эволюционно-новых, специфичных для человека механизмов работы мозга. Наша работа [229] была направлена на подтверждение этой гипотезы на уровне метаболитов. Мы изучили префронтальную кору головного мозга 32 пациентов с расстройствами аутистического спектра (РАС) разного возраста (2-60 лет) и 40 соответствующих им контролей (0-61 лет), где удалось обнаружить 202 метаболита с изменениями содержания при РАС - около 15% от общего количества измеренных метаболитов.

Помимо РАС, мы изучали и другие психические расстройства, в частности депрессию в работе [230], где применяли мульти-омиксный подход для анализа ответа на антидепрессант флуоксетин. Как и в случае с РАС, анализ экспрессии генов и содержания липидов и полярных метаболитов в предлимбической коре выявил существенные изменения в мозге, включая значимые изменения в количестве 106 липидов, сопровождающиеся незначительными изменениями в экспрессии генов [230].

При РАС же изменения содержания метаболитов были настолько выраженными, что по ним удалось построить предсказательную модель, отличающую больных РАС от здоровых людей с более чем 95% точностью. Таким образом, молекулярные маркеры РАС, найденные в нашем исследовании [229], могут стать хорошими мишенями для терапии и диагностики, поскольку метаболические изменения в мозге и крови во многом согласуются: большинство метаболических изменений, найденных нами в мозге, было сгруппировано в метаболических путях, ранее выявленных в моче и крови людей с РАС (10 из 16 метаболических путей). В частности, обнаруженные нами различия в мозге были значимо перепредставлены в путях метаболизма глутатиона, пуринов и пиримидинов, которые, как сообщалось ранее, связаны с РАС (рис. 3.24).

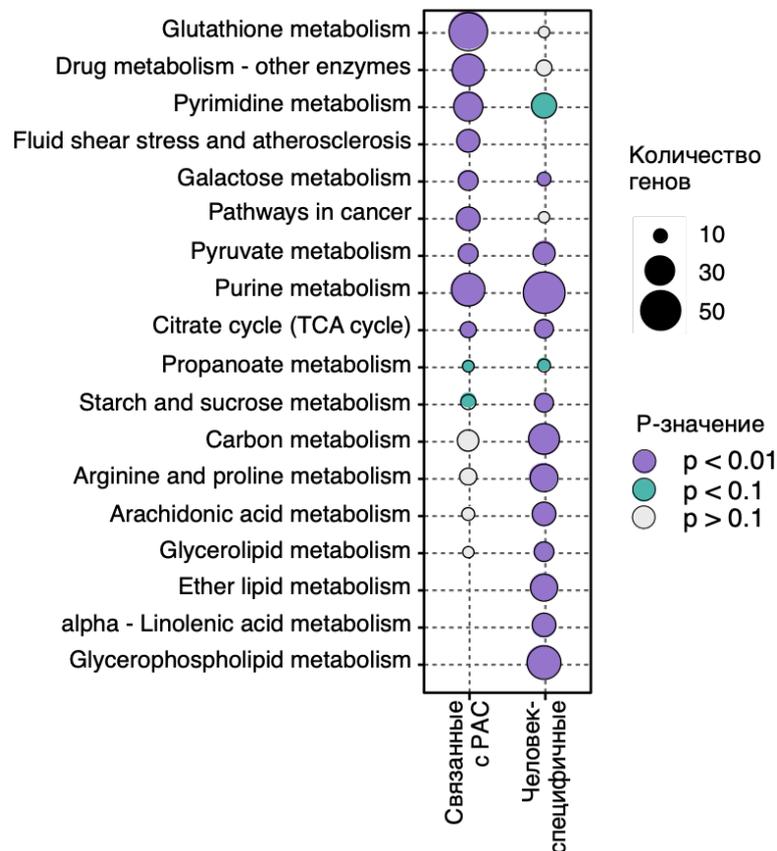


Рисунок 3.24 — Краткое описание основных функциональных путей, обогащенных генами, связанными с метаболитами из двух категорий: (1) 202 метаболита с изменениями при РАС ("связанные с РАС"); (2) специфичные для человека метаболиты ("человек-специфичные"). Размер каждого кружка пропорционален количеству генов в пути, связанном с метаболитами данной категории. Цвет каждого кружка показывает P-значения с поправкой на множественное тестирование. Адаптировано из [229].

Чтобы проверить гипотезу, что при развитии РАС затрагиваются именно эволюционно новые метаболические пути, мы также измерили содержание метаболитов в префронтальной коре 40 шимпанзе и 40 макак с распределением возрастов, соответствующим пациентам РАС и контролям. Образцы шимпанзе и макак измеряли вместе с контрольными образцами человека и образцами с РАС в случайном порядке, чтобы избежать нежелательных технических эффектов. Определение значимых различий в интенсивности метаболитов и их распределение по эволюционным линиям (ANCOVA, P-значение < 0,05 с поправкой на множественное тестирование) выявило 170 специфичных для человека и 55 специфичных для шимпанзе различий.

Гены, связанные со специфичными для человека метаболитами, были значимо перепредставлены в 27 путях KEGG (гипергеометрический тест, P -значение $< 0,05$ с поправкой на множественное тестирование). Примечательно, что эти пути значимо перекрывались с путями, обогащенными различиями в интенсивности метаболитов, связанными с РАС, и включали метаболизм пуринов и пиримидинов (критерий Фишера, P -значение $< 0,01$, рис. 3.24).

Таким образом, сравнение с шимпанзе и макаками показало, что при развитии аутизма затрагиваются именно эволюционно новые метаболические пути. К одному из таких метаболических путей, который нарушается при развитии аутизма и при этом является эволюционно новым, относится, как мы обнаружили, метаболический путь биосинтеза пуринов.

В самой недавней работе [231] мы более подробно исследовали метаболический путь биосинтеза пуринов в ходе анализа метаболомных данных для человека, шимпанзе и макак в нескольких тканях, включая 3 области мозга. Особое внимание было сосредоточено на ферменте аденилосукцинатлиазе (ADSL), который является одним из примерно 100 белков, несущих аминокислотную замену, которая присутствует сегодня у всех людей, но отсутствует у неандертальцев. Функциональное влияние этой специфической для человека мутации исследовалось путем создания гуманизированной мышинной модели, на основании которой было продемонстрировано снижение активности ADSL в тканях, а также снижение концентраций пуринов в областях мозга, при сравнении гуманизированной мышинной модели и контрольных мышей. Мы также показали, что наблюдается снижение концентрации метаболитов биосинтеза пуринов при сравнении индуцированных стволовых клеток человека без мутации в ферменте ADSL, с клетками, несущими неандертальский аллель, или с клетками шимпанзе. Наши наблюдения подтверждают гипотезу о том, что это изменение метаболизма произошло у людей после их отделения от предка, общего с неандертальцами и денисовцами. В целом, данное наблюдение и имеющиеся в литературе утверждения о том, что влияние ферментов, участвующих в метаболизме пуринов, имеет больше патологических последствий для нервной системы, чем для других органов, может стать новым направлением исследований на пути к открытию новых механизмов развития нейрокогнитивных расстройств.

3.3.2 Роль организации хроматина в регуляции специфичной для человека экспрессии в мозге

Чтобы понять возможные причины наблюдаемых нами эволюционных изменений на человеческой ветке, мы изучили контакты энхансеров и промоторов генов с уникальной для определенного региона мозга человека экспрессией, на основе данных о трехмерной организации хроматина. Мы отобрали 1271 уникальных генов, показывающих специфические для человека уровни экспрессии на основе данных с разрешением в одну клетку (см. подробное описание процедуры отбора в нашей работе [207]). Используя базу данных GeneHancer [232], мы сформировали список энхансеров, связанных с этими генами. Важно отметить, что соответствие между энхансерами и генами в базе GeneHancer установлено на основе данных о трехмерной организации хроматина, поскольку, как подробно обсуждалось выше, энхансер-промоторные взаимодействия поддерживаются хроматиновыми петлями, которые специфически связывают промоторы и энхансеры [63] при участии ряда вспомогательных белков, таких как YY1, HCF1, ZNF143 [63].

Затем мы подсчитали энхансеры, соответствующие активным в мозгу цис-регуляторным элементам, расположенным в областях с ускоренной эволюцией в геноме человека (Human Accelerated Regions, или HAR). Мы использовали 284 предсказанных HAR из работы [94]. Гены, демонстрирующие специфическую для человека экспрессию действительно располагались в непосредственной близости от этих энхансеров (перекрывающихся с активными в мозгу цис-регуляторными элементами, расположенными в HAR) чаще, чем ожидалось случайно (98 из 248 HAR, рис. 3.25). Чтобы оценить значимость этого совпадения, мы выполнили пермутационный тест: 1000 раз сделали случайные подвыборки размером 1271 ген из 9138 генов, экспрессированных в головном мозге (P-значение $< 0,005$, рис. 3.25).

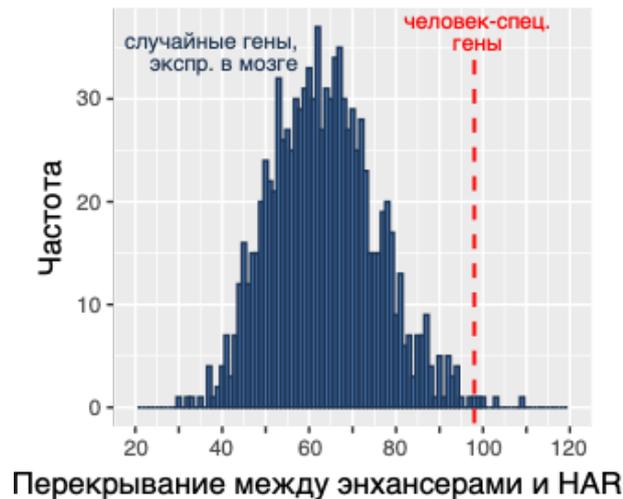


Рисунок 3.25 — Перекрытие между энхансерами, связанными с 1271 генами, демонстрирующими специфичную для человека экспрессию в данных snRNA-seq, и активными в мозге цис-регуляторными элементами, расположенными в HAR [94]. Гистограмма представляет собой распределение значений перекрытия, рассчитанных путем случайной подвыборки 1271 гена из 9138 генов, экспрессированных в головном мозге, выполненной 1000 раз. Красной пунктирной линией отмечено фактическое перекрытие ($n=98$). Адаптировано из [207].

Для анализа сайтов связывания транскрипционных факторов (ТФ) в этих 98 энхансерах, перекрывающихся с HAR, мы использовали распределение сайтов связывания ТФ из базы данных GeneHancer, составленное на основе анализа данных ChIP-seq. 35 из 42 ТФ, демонстрирующих специфическую для человека экспрессию в наших данных, имеют сайты связывания в этих 98 энхансерах, перекрывающихся с HAR. Более того, 13 из этих 35 ТФ связываются с 98 энхансерами, перекрывающимися с HAR, чаще, чем ожидалось случайно.

Данный анализ, основанный на данных о трехмерной организации хроматина, придал большую значимость нашему исследованию, установив возможную причину наблюдаемых нами эволюционных изменений на человеческой ветке.

Заключение

Основные **выводы** работы заключаются в следующем.

1. Разработана методология биоинформатического анализа данных Hi-C для образцов мозга человека: построены карты контактов, проведен анализ хроматиновых компартментов, ТАДов и петель.
2. Решена проблема разной экспериментальной представленности геномных регионов в данных Hi-C.
3. Разработана методика выявления биологического сигнала в данных Hi-C с высоким уровнем технической вариабельности.
4. Оптимизирован метод аннотации ТАДов на основе обогащения эпигенетическими метками.
5. Методы машинного обучения адаптированы и применены для исследования организации хроматина.
6. С помощью разработанных подходов к анализу данных Hi-C, описаны особенности трехмерной организации генома в нейронах и других клетках мозга человека, а также установлена функциональная роль дальних взаимодействий, которые опосредованы PcG белками и присутствуют только в нейронах.
7. Раскрыты общие принципы организации хроматина, такие как иерархичность ТАДов, в том числе в одиночных клетках, роль ТАДов в регуляции транскрипции, роль ядерной ламины, и др., что позволяет поместить особенности организации хроматина нейронов в контекст общих представлений о структуре хроматина.
8. Разработаны методы анализа данных RNA-seq, ATAC-seq, ChIP-seq и др. для решения нестандартных задач в контексте изучения мозга человека, таких как анализ биаллельной транскрипции, механизмы регуляции альтернативного сплайсинга (данные iCLIP), анализ некодирующих РНК, разделение сигнала от тканей человека и мыши в ксенотрансплантатах (PDX) для изучения онкозаболеваний, вычислительное восстановление видовой принадлежности отдельных клеток (данные snRNA-seq).

9. При помощи этих методов, а также разработанных методов анализа липидного и метаболического состава мозга человека, установлена роль липидов и конкретных метаболитов в функционировании клеток головного мозга человека в норме и при заболеваниях, в частности при психических расстройствах.
10. Оценена роль неандертальских вариантов в качестве возможной причины липидных отличий, специфичных для европейской популяции, и оценена роль неандертальских вариантов в других популяциях, в частности в популяциях кетов и папуасов.
11. Разработаны методы интеграции данных Hi-C с другими омиксными данными: RNA-seq, ATAC-seq, ChIP-seq и др., при помощи которых функционирование клеток мозга описано комплексно, что позволило установить регуляторную роль деацетилазы SIRT6 и архитектурного белка YY1, участвующего в образовании хроматиновых петель.
12. Установлено, что организация хроматина участвует в регуляции экспрессии генов в мозге человека, в особенности в участках генома, характеризующихся ускоренной эволюцией на человеческой ветке.

Таким образом, в данной работе установлена роль трехмерной организации генома в регуляции функционирования человеческого мозга, с учетом гетерогенности его клеточного состава и в комплексе с другими уровнями молекулярной организации клеток: экспрессией генов, эпигенетическими модификациями, связыванием транскрипционных факторов, открытостью хроматина и метаболизмом клеток в целом.

Публикации автора по теме диссертации

Статьи по теме диссертационной работы в рецензируемых журналах, индексируемых Scopus и/или Web of Science:

1. Kobets VA, Ulianov SV, Galitsyna AA, Doronin SA, Mikhaleva EA, Gelfand MS, Shevelyov YY, Razin SV, **Khrameeva EE**. HiConfidence: a novel approach uncovering the biological signal in Hi-C data affected by technical biases. *Brief Bioinform.* 2023; 24(2):bbad044. doi: 10.1093/bib/bbad044.
Импакт-фактор: 13.9, квартиль: Q1.
2. Smirnov D, Eremenko E, Stein D, Kaluski S, Jasinska W, Cosentino C, Martinez-Pastor B, Brotman Y, Mostoslavsky R, **Khrameeva E[#]**, Toiber D[#]. SIRT6 is a key regulator of mitochondrial function in the brain. *Cell Death Dis.* 2023; 14(1):35. doi: 10.1038/s41419-022-05542-w.
Импакт-фактор: 9.7, квартиль: Q1.
3. Shcherbinina E, Abakumova T, Bobrovskiy D, Kurochkin I, Deinichenko K, Stekolshchikova E, Anikanov N, Ziganshin R, Melnikov P, **Khrameeva E**, Logacheva M, Zatsepin T, Sergeeva O. Murine Falcor/LL35 lncRNA Contributes to Glucose and Lipid Metabolism In Vitro and In Vivo. *Biomedicines.* 2022; 10(6):1397. doi: 10.3390/biomedicines10061397.
Импакт-фактор: 4.7, квартиль: Q1.
4. Tkachev A, Stekolshchikova E, Bobrovskiy DM, Anikanov N, Ogurtsova P, Park DI, Horn AKE, Petrova D, **Khrameeva E[#]**, Golub MS[#], Turck CW[#], Khaitovich P[#]. Long-Term Fluoxetine Administration Causes Substantial Lipidome Alteration of the Juvenile Macaque Brain. *Int J Mol Sci.* 2021; 22(15):8089. doi: 10.3390/ijms22158089.
Импакт-фактор: 6.2, квартиль: Q1.
5. Ilyin AA, Kononkova AD, Golova AV, Shloma VV, Olenkina OM, Nenasheva VV, Abramov YA, Kotov AA, Maksimov DA, Laktionov PP, Pindyurin AV, Galitsyna AA, Ulianov SV, **Khrameeva EE[#]**, Gelfand MS, Belyakin SN, Razin SV, Shevelyov YY[#]. Comparison of genome architecture at two stages of male germline cell differentiation in *Drosophila*.

- Nucleic Acids Res. 2022; 50(6):3203-3225. doi: 10.1093/nar/gkac109.
Импакт-фактор: 11.5, квартиль: Q1.
6. Stepanova V, Moczulska KE, Vacano GN, Kurochkin I, Ju X, RiesenberG S, Macak D, Maricic T, Dombrowski L, Schörnig M, Anastassiadis K, Baker O, Naumann R, **Khrameeva E**, Vanushkina A, Stekolshchikova E, Egorova A, Tkachev A, Mazzarino R, Duval N, Zubkov D, Giavalisco P, Wilkinson TG, Patterson D, Khaitovich P, Pääbo S. Reduced purine biosynthesis in humans after their divergence from Neandertals. *Elife*. 2021; 10:e58741. doi: 10.7554/eLife.58741.
Импакт-фактор: 8.1, квартиль: Q1.
 7. Smirnov D, Mazin P, Osetrova M, Stekolshchikova E, **Khrameeva E**. The Hitchhiker's Guide to Untargeted Lipidomics Analysis: Practical Guidelines. *Metabolites*. 2021; 11(11):713. doi: 10.3390/metabo11110713.
Импакт-фактор: 4.7, квартиль: Q2.
 8. Е.А. Стекольщикова, А.И. Ткачев, М.С. Осетрова, Д.Н. Смирнов, О.И. Ефимова, **Е.Е. Храмеева**. Применение методов масс-спектрометрии к анализу липидного состава головного мозга человека. 2021; Т. 37, № 5, С. 80–87. doi: 10.21519/0234-2758-2021-37-5-80-87.
Импакт-фактор: 0.4, квартиль: Q4.
 9. Eremenko E, Golova A, Stein D, Einav M, **Khrameeva E**, Toiber D. FACS-based isolation of fixed mouse neuronal nuclei for ATAC-seq and Hi-C. *STAR Protoc*. 2021; 2(3):100643. doi: 10.1016/j.xpro.2021.100643.
Импакт-фактор: 1.4, квартиль: Q2.
 10. Mazor G, Smirnov D, Ben David H, **Khrameeva E**, Toiber D, Rotblat B. TP73-AS1 is induced by YY1 during TMZ treatment and highly expressed in the aging brain. *Aging (Albany NY)*. 2021; 13(11):14843-14861. doi: 10.18632/aging.203182.
Импакт-фактор: 4.8, квартиль: Q1.
 11. Badarni M, Prasad M, Golden A, Bhattacharya B, Levin L, Yegodayev KM, Dimitstein O, Joshua BZ, Cohen L, **Khrameeva E**, Kong D, Porgador A, Braiman A, Grandis JR, Rotblat B, Elkabets M. IGF2 Mediates Resistance to Isoform Selective-Inhibitors of the PI3K in HPV Positive Head and Neck Cancer. *Cancers (Basel)*. 2021; 13(9):2250. doi: 10.3390/cancers13092250.
Импакт-фактор: 6.1, квартиль: Q1.

12. Rozenwald MB, Galitsyna AA, Sapunov GV, **Khrameeva EE**, Gelfand MS. A machine learning framework for the prediction of chromatin folding in *Drosophila* using epigenetic features. PeerJ Comput Sci. 2020; 6:e307. doi: 10.7717/peerj-cs.307.
Импакт-фактор: 4.8, квартиль: Q2.
13. Stein D, Mizrahi A, Golova A, Saretzky A, Venzor AG, Slobodnik Z, Kaluski S, Einav M, **Khrameeva E**, Toiber D. Aging and pathological aging signatures of the brain: through the focusing lens of SIRT6. Aging (Albany NY). 2021; 13(5):6420-6441. doi: 10.18632/aging.202755.
Импакт-фактор: 4.8, квартиль: Q1.
14. Ulianov SV, Zakharova VV, Galitsyna AA, Kos PI, Polovnikov KE, Flyamer IM, Mikhaleva EA, **Khrameeva EE**, Germini D, Logacheva MD, Gavrillov AA, Gorsky AS, Nechaev SK, Gelfand MS, Vassetzky YS, Chertovich AV, Shevelyov YY, Razin SV. Order and stochasticity in the folding of individual *Drosophila* genomes. Nat Commun. 2021; 12(1):41. doi: 10.1038/s41467-020-20292-z.
Импакт-фактор: 12.1, квартиль: Q1.
15. Samborskaia MD, Galitsyna A, Pletenev I, Trofimova A, Mironov AA, Gelfand MS, **Khrameeva EE**. Cumulative contact frequency of a chromatin region is an intrinsic property linked to its function. PeerJ. 2020; 8:e9566. doi: 10.7717/peerj.9566.
Импакт-фактор: 2.3, квартиль: Q1.
16. **Khrameeva E***, Kurochkin I*, Han D*, Guijarro P, Kanton S, Santel M, Qian Z, Rong S, Mazin P, Sabirov M, Bulat M, Efimova O, Tkachev A, Guo S, Sherwood CC, Camp JG, Pääbo S, Treutlein B, Khaitovich P. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. Genome Res. 2020; 30(5):776-789. doi: 10.1101/gr.256958.119.
Импакт-фактор: 9.4, квартиль: Q1.
17. Yegodayev KM, Novoplansky O, Golden A, Prasad M, Levin L, Jagadeeshan S, Zorea J, Dimitstein O, Joshua BZ, Cohen L, **Khrameeva E#**, Elkabets M#. TGF-Beta-Activated Cancer-Associated Fibroblasts Limit Cetuximab Efficacy in Preclinical Models of Head and Neck Cancer. Cancers (Basel). 2020; 12(2):339. doi: 10.3390/cancers12020339.
Импакт-фактор: 6.1, квартиль: Q1.

18. Tkachev A, Stepanova V, Zhang L, **Khrameeva E**, Zubkov D, Giavalisco P, Khaitovich P. Differences in lipidome and metabolome organization of prefrontal cortex among human populations. *Sci Rep.* 2019; 9(1):18348. doi: 10.1038/s41598-019-53762-6.
Импакт-фактор: 4.6, квартиль: Q1.
19. Kurochkin I*, **Khrameeva E***, Tkachev A, Stepanova V, Vanyushkina A, Stekolshchikova E, Li Q, Zubkov D, Shichkova P, Halene T, Willmitzer L, Giavalisco P, Akbarian S, Khaitovich P. Metabolome signature of autism in the human prefrontal cortex. *Commun Biol.* 2019; 2:234. doi: 10.1038/s42003-019-0485-4.
Импакт-фактор: 4.2, квартиль: Q1.
20. Luzhin AV, Flyamer IM, **Khrameeva EE**, Ulianov SV, Razin SV, Gavrilov AA. Quantitative differences in TAD border strength underly the TAD hierarchy in Drosophila chromosomes. *J Cell Biochem.* 2019; 120(3):4494-4503. doi: 10.1002/jcb.27737.
Импакт-фактор: 4.4, квартиль: Q2.
21. Ulianov SV*, Doronin SA*, **Khrameeva EE***, Kos PI*, Luzhin AV, Starikov SS, Galitsyna AA, Nenasheva VV, Ilyin AA, Flyamer IM, Mikhaleva EA, Logacheva MD, Gelfand MS, Chertovich AV, Gavrilov AA, Razin SV, Shevelyov YY. Nuclear lamina integrity is required for proper spatial organization of chromatin in Drosophila. *Nat Commun.* 2019; 10(1):1176. doi: 10.1038/s41467-019-09185-y.
Импакт-фактор: 12.1, квартиль: Q1.
22. **Khrameeva E***, Kurochkin I*, Bozek K, Giavalisco P, Khaitovich P. Lipidome Evolution in Mammalian Tissues. *Mol Biol Evol.* 2018; 35(8):1947-1957. doi: 10.1093/molbev/msy097.
Импакт-фактор: 8.8, квартиль: Q1.
23. Akkuratov EE, Gelfand MS, **Khrameeva EE**. Neanderthal and Denisovan ancestry in Papuans: A functional study. *J Bioinform Comput Biol.* 2018; 16(2):1840011. doi: 10.1142/S0219720018400115.
Импакт-фактор: 1.2, квартиль: Q3.
24. Ulianov SV, Galitsyna AA, Flyamer IM, Golov AK, **Khrameeva EE**, Imakaev MV, Abdennur NA, Gelfand MS, Gavrilov AA, Razin SV. Activation of the alpha-globin gene expression correlates with dramatic

- upregulation of nearby non-globin genes and changes in local and large-scale chromatin spatial structure. *Epigenetics and Chromatin*. 2017; 10(1):35. doi: 10.1186/s13072-017-0142-4.
Импакт-фактор: 4.2, квартиль: Q1.
25. Bozek K, **Khrameeva EE**, Reznick J, Omerbašić D, Bennett NC, Lewin GR, Azpurua J, Gorbunova V, Seluanov A, Regnard P, Wanert F, Marchal J, Pifferi F, Aujard F, Liu Z, Shi P, Pääbo S, Schroeder F, Willmitzer L, Giavalisco P, Khaitovich P. Lipidome determinants of maximal lifespan in mammals. *Sci Rep*. 2017; 7(1):5. doi: 10.1038/s41598-017-00037-7.
Импакт-фактор: 4.6, квартиль: Q1.
26. Gavrilov AA, Shevelyov YY, Ulianov SV, **Khrameeva EE**, Kos P, Chertovich A, Razin SV. Unraveling the mechanisms of chromatin fibril packaging. *Nucleus*. 2016; 7(3):319-24. doi: 10.1080/19491034.2016.1190896.
Импакт-фактор: 10.2, квартиль: Q1.
27. **Khrameeva EE**, Fudenberg G, Gelfand MS, Mirny LA. History of chromosome rearrangements reflects the spatial organization of yeast chromosomes. *J Bioinform Comput Biol*. 2016; 1641002. doi: 10.1142/S021972001641002X.
Импакт-фактор: 1.2, квартиль: Q3.
28. Zhang B, Han D, Korostelev Y, Yan Z, Shao N, **Khrameeva E**, Velichkovsky BM, Chen YP, Gelfand MS, Khaitovich P. Changes in snoRNA and snRNA Abundance in the Human, Chimpanzee, Macaque, and Mouse Brain. *Genome Biol Evol*. 2016; 8(3):840-50. doi: 10.1093/gbe/evw038.
Импакт-фактор: 10.2, квартиль: Q1.
29. Flegontov P, Changmai P, Zidkova A, Logacheva MD, Altınışık NE, Flegontova O, Gelfand MS, Gerasimov ES, **Khrameeva EE**, Konovalova OP, Neretina T, Nikolsky YV, Starostin G, Stepanova VV, Travinsky IV, Tříška M, Tříška P, Tatarinova TV. Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci Rep*. 2016; 6:20768. doi: 10.1038/srep20768.
Импакт-фактор: 4.6, квартиль: Q1.
30. Ulianov SV*, **Khrameeva EE***, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, Penin AA, Logacheva MD, Imakaev MV, Chertovich A,

- Gelfand MS, Shevelyov YY, Razin SV. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 2016; 26(1):70-84. doi: 10.1101/gr.196006.115.
Импакт-фактор: 9.4, квартиль: Q1.
31. Kaufmann S, Fuchs C, Gonik M, **Khrameeva EE**, Mironov AA, Frishman D. Inter-chromosomal contact networks provide insights into Mammalian chromatin organization. *PLoS One.* 2015; 10(5):e0126125. doi: 10.1371/journal.pone.0126125.
Импакт-фактор: 3.7, квартиль: Q1.
32. Zhenilo S, **Khrameeva E**, Tsygankova S, Zhigalova N, Mazur A, Prokhortchouk E. Individual genome sequencing identified a novel enhancer element in exon 7 of the CSFR1 gene by shift of expressed allele ratios. *Gene.* 2015; 566(2):223-8. doi: 10.1016/j.gene.2015.04.053.
Импакт-фактор: 3.9, квартиль: Q2.
33. Yang J, Hung LH, Licht T, Kostin S, Looso M, **Khrameeva E**, Bindereif A, Schneider A, Braun T. RBM24 is a major regulator of muscle-specific alternative splicing. *Dev Cell.* 2014; 31(1):87-99. doi: 10.1016/j.devcel.2014.08.025.
Импакт-фактор: 13.4, квартиль: Q1.
34. **Khrameeva EE**, Bozek K, He L, Yan Z, Jiang X, Wei Y, Tang K, Gelfand MS, Prufer K, Kelso J, Paabo S, Giavalisco P, Lachmann M, Khaitovich P. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat Commun.* 2014; 5:3584. doi: 10.1038/ncomms4584.
Импакт-фактор: 12.1, квартиль: Q1.
35. Roszbach O, Hung LH, **Khrameeva E**, Schreiner S, König J, Curk T, Zupan B, Ule J, Gelfand MS, Bindereif A. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* 2014; 11(2):146-55. doi: 10.4161/rna.27991.
Импакт-фактор: 4.7, квартиль: Q1.
36. Rösel-Hillgärtner TD, Hung LH, **Khrameeva E**, Le Querrec P, Gelfand MS, Bindereif A. A novel intra-U1 snRNP cross-regulation mechanism: alternative splicing switch links U1C and U1-70K expression. *PLoS Genet.* 2013; 9(10):e1003856. doi: 10.1371/journal.pgen.1003856.
Импакт-фактор: 5.1, квартиль: Q1.

37. **Khrameeva EE**, Gelfand MS. Biases in read coverage demonstrated by interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. *BMC Bioinformatics*. 2012; 13 Suppl 6:S4. doi: 10.1186/1471-2105-13-S6-S4.
Импакт-фактор: 3.3, квартиль: Q1.
38. **Khrameeva EE**, Mironov AA, Fedonin GG, Khaitovich P, Gelfand MS. Spatial proximity and similarity of the epigenetic state of genome domains. *PLoS One*. 2012; 7(4):e33947. doi: 10.1371/journal.pone.0033947.
Импакт-фактор: 3.7, квартиль: Q1.
39. Pervouchine DD, **Khrameeva EE**, Pichugina MY, Nikolaienko OV, Gelfand MS, Rubtsov PM, Mironov AA. Evidence for widespread association of mammalian splicing and conserved long-range RNA structures. *RNA*. 2012; 18(1):1-15. doi: 10.1261/rna.029249.111.
Импакт-фактор: 4.2, квартиль: Q1.

* - первые авторы с равным вкладом

- авторы, ответственные за переписку, с равным вкладом

Список сокращений и условных обозначений

- ТАДы** - топологически ассоциированные домены
- Hi-C** - технология высокопроизводительного захвата конформации хромосом
- CTCF** - фактор связывания CCCTC
- PDX** - ксенотрансплантаты
- п.о.** - пары оснований
- SINEs** - короткие ядерные элементы
- SMC** - комплекс структурного поддержания хромосом
- PcG** - белки группы поликомба
- PRC1** - репрессивный комплекс поликомба 1
- PRC2** - репрессивный комплекс поликомба 2
- H3K27me3** - триметилирование остатка лизина на гистоне H3
- TSS** - сайт старта транскрипции
- FIREs** - особый уровень архитектуры хроматина, который тесно связан с клеточно-специфичной регуляцией генов
- ТФ** - транскрипционные факторы
- FISH** - флуоресцентная гибридизация *in situ*
- Oligopaint** - микроскопия с применением пользовательских олигонуклеотидных матриц
- NGS** - секвенирование следующего поколения
- ChIP-seq** - технология иммунопреципитации хроматина, совмещенная с секвенированием
- Hi-M** - метод микроскопии с высокой пропускной способностью, разрешением и охватом на основе Oligopaint
- ATAC-seq** - метод определения доступного хроматина с помощью транспозазы
- MNase-seq** - разрезание ДНК микрококковой нуклеазой с глубоким секвенированием
- DNase-seq** - секвенирование гиперчувствительных сайтов ДНКазы I
- ДНКазы I** - дезоксирибонуклеаза I
- RNA-seq** - секвенирование РНК

scRNA-seq - секвенирование РНК с разрешением в одну клетку, версия метода для одиночных клеток

snRNA-seq - секвенирование РНК с разрешением в одну клетку, версия метода для одиночных ядер

ICE - итеративно корректирующая балансировка

PC - вектор главной компоненты

O/E - наблюдаемое/ожидаемое

IS - индекс инсуляции (Insulation score)

CCF - суммарная частота контактов (Cumulative Contact Frequency)

interCCF - меж-хромосомная суммарная частота контактов

TSA - трихостатин

HDAC-dep - деплеция деацетилазы гистонов

HDAC-inh - ингибирование деацетилаз гистонов агентом TSA

HAT-inh - ингибирование гистонацетилтрансфераз куркумином

SCC - корреляция карт Hi-C, оцененная с использованием коэффициента корреляции с поправкой на ожидаемое

log2FC - log2-кратное изменение

Lam-KD - нокдаун ламина Dm0

RMSD - среднеквадратичное отклонение

AUC - площадь под кривой

LR - линейная регрессия

GB - градиентный бустинг

RNN - рекуррентные нейронные сети

SNP - однонуклеотидные полиморфизмы

ВЭЖХ–МС – масс-спектрометрия в сочетании с высокоэффективной жидкостной хроматографией

CAR – карнитины

CE – эфиры холестерина

Cer – церамиды

DAG – диглицириды

HexCer – гексозилцерамиды

LPC – лизофосфатидилхолины

LPE – лизофосфатидилэтаноламины

MAG – моноглицериды

PC – фосфатидилхолины
PC O – 1-алкил/алкенил,2-ацил-глицерофосфо-этанолламины
PE –фосфатидилэтанолламины
PE O – 1-алкил/алкенил,2-ацил-глицерофосфохолины
PG – фосфатидилглицеролы
PI – фосфатидилинозитолы
PS – фосфатидилсерины
SM – сфингомиелины
SulfoHexCer – сульфатиды
TAG – триглицериды
m/z - масса/заряд
WT - мышцы дикого типа
SIRT6-KO - мозг-специфичный нокаут
mESC - эмбриональные стволовые клетки мышцы
GO - генные онтологии (Gene Ontology)
PCA - анализ главных компонент
ДС - анализ дифференциального содержания
ДЭ - анализ дифференциальной экспрессии
MitoSOX - индикаторы митохондриального супероксида
SpG - сперматогонии
SpCs - сперматоциты
NeuN+ - нейрональные клетки
NeuN- - не-нейрональные клетки
АС - передняя поясная кора
CN - хвостатое ядро
СВ - мозжечок
In - тормозящие нейроны
Ex - возбуждающие нейроны
OPC - предшественники олигодендроцитов
Ast - астроциты
OD - олигодендроциты
MG - микроглия
мяРНК - малые ядерные РНК
мякРНК - малые ядрышковые РНК

- RI** - удержание интрона
- A3SS** - альтернативный 3' (акцепторный) сайт сплайсинга
- A5SS** - альтернативный 5' (донорный) сайт сплайсинга
- MXE** - взаимоисключающие экзоны
- CX** - кора больших полушарий головного мозга
- KD** - почки
- HT** - сердце
- ML** - мышцы
- LV** - печень
- AF** - люди африканского происхождения
- AS** - люди азиатского происхождения
- EU** - люди европейского происхождения
- LCP** - путь катаболизма липидов
- ASW** - индивидуумы африканского происхождения с юго-запада США
- LWK** - представители народов лухья
- YRI** - представители народов йоруба
- CHB** - китайцы—ханьцы из Пекина
- CHS** - китайцы—ханьцы из Южного Китая
- JPT** - японцы
- CEU** - индивидуумы из Центральной Европы
- FIN** - финны из Финляндии
- GBR** - британцы из Англии и Шотландии
- IBS** - иберы из Испании
- TSI** - тосканцы
- Altai** - последовательность генома неандертальца с высоким покрытием от индивидуума, найденного в восточной галерее Денисовой пещеры в Горном Алтае
- Vindjia** - последовательность генома неандертальца с низким покрытием, полученная от трех индивидуумов
- NLS** - частота неандертальских вариантов (Neanderthal-like sites)
- CMS** - комбинированная метрика множественных сигналов
- GW** - весь геном
- MC** - гены LCP, связанные со специфичными для Европы изменениями содержания метаболитов

РАС - расстройства аутистического спектра

НАР - области с ускоренной эволюцией в геноме человека (Human Accelerated Regions)

Список литературы

1. Recent evidence that TADs and chromatin loops are dynamic structures / A. S. Hansen [и др.] // *Nucleus*. — 2018. — 31 дек. — т. 9, № 1. — с. 20—32.
2. *Cremer T., Cremer C.* Chromosome territories, nuclear architecture and gene regulation in mammalian cells // *Nature Reviews Genetics*. — 2001. — 1 апр. — т. 2, № 4. — с. 292—301.
3. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome / E. Lieberman-Aiden [и др.] // *Science*. — 2009. — 9 окт. — т. 326, № 5950. — с. 289—293.
4. *Bonev B., Cavalli G.* Organization and function of the 3D genome // *Nature Reviews Genetics*. — 2016. — нояб. — т. 17, № 11. — с. 661—678.
5. Topological domains in mammalian genomes identified by analysis of chromatin interactions / J. R. Dixon [и др.] // *Nature*. — 2012. — май. — т. 485, № 7398. — с. 376—380.
6. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes / A. L. Sanborn [и др.] // *Proceedings of the National Academy of Sciences*. — 2015. — 24 нояб. — т. 112, № 47.
7. *Rowley M. J., Corces V. G.* Organizational principles of 3D genome architecture // *Nature Reviews Genetics*. — 2018. — дек. — т. 19, № 12. — с. 789—800.
8. Chromatin architecture reorganization during stem cell differentiation / J. R. Dixon [и др.] // *Nature*. — 2015. — 19 февр. — т. 518, № 7539. — с. 331—336.
9. Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing / L. Li [и др.] // *Molecular cell*. — 2015. — т. 58, № 2. — с. 216—231.
10. Evolutionarily conserved principles predict 3D chromatin organization / M. J. Rowley [и др.] // *Molecular cell*. — 2017. — т. 67, № 5. — с. 837—852.
11. From compartments to gene loops: Functions of the 3D genome in the human brain : preprint / S. Rahman [и др.] ; *Genomics*. — 14.10.2021.

12. Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders / В. Hu [и др.] // Nature Communications. — 2021. — 25 июня. — т. 12, № 1. — с. 3968.
13. Molecular and cellular reorganization of neural circuits in the human lineage / A. M. Sousa [и др.] // Science. — 2017. — т. 358, № 6366. — с. 1027–1032.
14. Human-specific features of spatial gene expression and regulation in eight brain regions / С. Xu [и др.] // Genome research. — 2018. — т. 28, № 8. — с. 1097–1110.
15. *Li B., Carey M., Workman J. L.* The role of chromatin during transcription // Cell. — 2007. — т. 128, № 4. — с. 707–719.
16. *Venkatesh S., Workman J. L.* Histone exchange, chromatin structure and the regulation of transcription // Nature reviews Molecular cell biology. — 2015. — т. 16, № 3. — с. 178–189.
17. Chromatin potentiates transcription / S. Nagai [и др.] // Proceedings of the National Academy of Sciences. — 2017. — т. 114, № 7. — с. 1536–1541.
18. *Tsompana M., Buck M. J.* Chromatin accessibility: a window into the genome // Epigenetics & chromatin. — 2014. — т. 7, № 1. — с. 1–16.
19. *Fu H., Baris A., Aladjem M. I.* Replication timing and nuclear structure // Current opinion in cell biology. — 2018. — т. 52. — с. 43–50.
20. *Smith O. K., Aladjem M. I.* Chromatin structure and replication origins: determinants of chromosome replication and nuclear organization // Journal of Molecular Biology. — 2014. — т. 426, № 20. — с. 3330–3341.
21. *MacAlpine D. M., Almouzni G.* Chromatin and DNA replication // Cold Spring Harbor perspectives in biology. — 2013. — т. 5, № 8. — a010207.
22. A chromatin structure-based model accurately predicts DNA replication timing in human cells / Y. Gindin [и др.] // Molecular systems biology. — 2014. — т. 10, № 3. — с. 722.
23. *Stadler J., Richly H.* Regulation of DNA repair mechanisms: how the chromatin environment regulates the DNA damage response // International journal of molecular sciences. — 2017. — т. 18, № 8. — с. 1715.

24. *Roukos V., Misteli T.* The biogenesis of chromosome translocations // *Nature cell biology*. — 2014. — т. 16, № 4. — с. 293—300.
25. Mutations in chromatin regulators functionally link Cornelia de Lange syndrome and clinically overlapping phenotypes / I. Parenti [и др.] // *Human genetics*. — 2017. — т. 136. — с. 307—320.
26. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation / R. J. Gibbons [и др.] // *Nature genetics*. — 2000. — т. 24, № 4. — с. 368—371.
27. Age-related and disease locus-specific mechanisms contribute to early remodelling of chromatin structure in Huntington's disease mice / R. Alcalá-Vida [и др.] // *Nature communications*. — 2021. — т. 12, № 1. — с. 364.
28. De novo mutations identified by whole-genome sequencing implicate chromatin modifications in obsessive-compulsive disorder / G. N. Lin [и др.] // *Science Advances*. — 2022. — т. 8, № 2. — eabi6180.
29. *Renthal W., Nestler E. J.* Chromatin regulation in drug addiction and depression // *Dialogues in clinical neuroscience*. — 2022.
30. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. / D. Pinkel [и др.] // *Proceedings of the National Academy of Sciences*. — 1988. — т. 85, № 23. — с. 9138—9142.
31. Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes / A. Bolzer [и др.] // *PLoS Biology* / под ред. Т. Misteli. — 2005. — 26 апр. — т. 3, № 5. — e157.
32. *Meaburn K. J., Misteli T.* Chromosome territories // *Nature*. — 2007. — янв. — т. 445, № 7126. — с. 379—381.
33. *Misteli T.* Beyond the Sequence: Cellular Organization of Genome Function // *Cell*. — 2007. — февр. — т. 128, № 4. — с. 787—800.
34. Distance between homologous chromosomes results from chromosome positioning constraints / C. Heride [и др.] // *Journal of cell science*. — 2010. — т. 123, № 23. — с. 4063—4075.

35. Factors that affect the formation of chromosomal translocations in cells / R. J. Canoy [и др.] // *Cancers*. — 2022. — т. 14, № 20. — с. 5110.
36. Spatial proximity and similarity of the epigenetic state of genome domains / E. E. Khrameeva [и др.] // *PloS one*. — 2012. — т. 7, № 4. — e33947.
37. *Szalaj P., Plewczynski D.* Three-dimensional organization and dynamics of the genome // *Cell Biology and Toxicology*. — 2018. — окт. — т. 34, № 5. — с. 381—404.
38. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions / L. Guelen [и др.] // *Nature*. — 2008. — 12 июня. — т. 453, № 7197. — с. 948—951.
39. Heterogeneous network embedding enabling accurate disease association predictions / Y. Xiong [и др.] // *BMC Medical Genomics*. — 2019. — дек. — т. 12, S10. — с. 186.
40. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping / S. S. Rao [и др.] // *Cell*. — 2014. — дек. — т. 159, № 7. — с. 1665—1680.
41. *Gibcus J. H., Dekker J.* The Hierarchy of the 3D Genome // *Molecular Cell*. — 2013. — март. — т. 49, № 5. — с. 773—782.
42. Spatial partitioning of the regulatory landscape of the X-inactivation centre / E. P. Nora [и др.] // *Nature*. — 2012. — май. — т. 485, № 7398. — с. 381—385.
43. Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome / T. Sexton [и др.] // *Cell*. — 2012. — февр. — т. 148, № 3. — с. 458—472.
44. Entering the next dimension: plant genomes in 3D / M. Sotelo-Silveira [и др.] // *Trends in plant science*. — 2018. — т. 23, № 7. — с. 598—612.
45. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae* / D. J. Winter [и др.] // *PLoS Genetics*. — 2018. — т. 14, № 10. — e1007467.
46. *Krijger P. H. L., De Laat W.* Regulation of disease-associated gene expression in the 3D genome // *Nature reviews Molecular cell biology*. — 2016. — т. 17, № 12. — с. 771—782.

47. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions / D. G. Lupiáñez [и др.] // *Cell*. — 2015. — май. — т. 161, № 5. — с. 1012—1025.
48. *Spielmann M., Lupiáñez D. G., Mundlos S.* Structural variation in the 3D genome // *Nature Reviews Genetics*. — 2018. — т. 19, № 7. — с. 453—467.
49. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking / J. Weischenfeldt [и др.] // *Nature genetics*. — 2017. — т. 49, № 1. — с. 65—74.
50. Activation of proto-oncogenes by disruption of chromosome neighborhoods / D. Hnisz [и др.] // *Science*. — 2016. — т. 351, № 6280. — с. 1454—1458.
51. Insulator dysfunction and oncogene activation in IDH mutant gliomas / W. A. Flavahan [и др.] // *Nature*. — 2016. — янв. — т. 529, № 7584. — с. 110—114.
52. A high-resolution map of the three-dimensional chromatin interactome in human cells / F. Jin [и др.] // *Nature*. — 2013. — т. 503, № 7475. — с. 290—294.
53. Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor / W. Deng [и др.] // *Cell*. — 2012. — июнь. — т. 149, № 6. — с. 1233—1244.
54. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment / J. E. Phillips-Cremins [и др.] // *Cell*. — 2013. — июнь. — т. 153, № 6. — с. 1281—1295.
55. *Yu W., He B., Tan K.* Identifying topologically associating domains and subdomains by Gaussian Mixture model And Proportion test // *Nature Communications*. — 2017. — 14 сент. — т. 8, № 1. — с. 535.
56. Chromatin organization by an interplay of loop extrusion and compartmental segregation / J. Nuebler [и др.] // *Proceedings of the National Academy of Sciences*. — 2018. — 17 июля. — т. 115, № 29.
57. *West A. G., Fraser P.* Remote control of gene transcription // *Human Molecular Genetics*. — 2005. — 15 апр. — т. 14, suppl_1. — R101—R111.
58. A compendium of promoter-centered long-range chromatin interactions in the human genome / I. Jung [и др.] // *Nature Genetics*. — 2019. — окт. — т. 51, № 10. — с. 1442—1449.

59. Comprehensive functional genomic resource and integrative model for the human brain / D. Wang [и др.] // *Science*. — 2018. — 14 дек. — т. 362, № 6420. — eaat8464.
60. Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes / J. M. Downen [и др.] // *Cell*. — 2014. — окт. — т. 159, № 2. — с. 374—387.
61. Enhancer release and retargeting activates disease-susceptibility genes / S. Oh [и др.] // *Nature*. — 2021. — 29 июля. — т. 595, № 7869. — с. 735—740.
62. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus / E. M. Smith [и др.] // *The American Journal of Human Genetics*. — 2016. — янв. — т. 98, № 1. — с. 185—201.
63. Hi-TrAC reveals division of labor of transcription factors in organizing chromatin loops / S. Liu [и др.] // *Nature Communications*. — 2022. — 5 нояб. — т. 13, № 1. — с. 6679.
64. CTCF/cohesin-binding sites are frequently mutated in cancer / R. Katainen [и др.] // *Nature Genetics*. — 2015. — июль. — т. 47, № 7. — с. 818—821.
65. Formation of Chromosomal Domains by Loop Extrusion / G. Fudenberg [и др.] // *Cell Reports*. — 2016. — май. — т. 15, № 9. — с. 2038—2049.
66. *Uhlmann F.* SMC complexes: from DNA to chromosomes // *Nature reviews Molecular cell biology*. — 2016. — т. 17, № 7. — с. 399—412.
67. *Banigan E. J., Mirny L. A.* Loop extrusion: theory meets single-molecule experiments // *Current opinion in cell biology*. — 2020. — т. 64. — с. 124—138.
68. *Xi W., Beer M. A.* Loop competition and extrusion model predicts CTCF interaction specificity // *Nature communications*. — 2021. — т. 12, № 1. — с. 1046.
69. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner / S. Golfier [и др.] // *Elife*. — 2020. — т. 9. — e53885.
70. Cohesin Loss Eliminates All Loop Domains / S. S. Rao [и др.] // *Cell*. — 2017. — окт. — т. 171, № 2. — 305—320.e24.

71. The cohesin release factor WAPL restricts chromatin loop extension / J. H. Haarhuis [и др.] // *Cell*. — 2017. — т. 169, № 4. — с. 693—707.
72. The structural basis for cohesin–CTCF-anchored loops / Y. Li [и др.] // *Nature*. — 2020. — т. 578, № 7795. — с. 472—476.
73. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins / G. Wutz [и др.] // *The EMBO Journal*. — 2017. — 15 дек. — т. 36, № 24. — с. 3573—3599.
74. *Kyrchanova O., Georgiev P.* Mechanisms of enhancer-promoter interactions in higher eukaryotes // *International Journal of Molecular Sciences*. — 2021. — т. 22, № 2. — с. 671.
75. YY1 is a structural regulator of enhancer-promoter loops / A. S. Weintraub [и др.] // *Cell*. — 2017. — т. 171, № 7. — 1573—1588.e28.
76. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters / S. D. Bailey [и др.] // *Nature communications*. — 2015. — т. 6, № 1. — с. 1—10.
77. *Krivega I., Dale R. K., Dean A.* Role of LDB1 in the transition from chromatin looping to transcription activation // *Genes & development*. — 2014. — т. 28, № 12. — с. 1278—1290.
78. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome / A. D. Schmitt [и др.] // *Cell Reports*. — 2016. — нояб. — т. 17, № 8. — с. 2042—2059.
79. *Giusti-Rodriguez P. M., Sullivan P. F.* Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits : preprint / *Genetics*. — 31.08.2018.
80. *Margueron R., Reinberg D.* The Polycomb complex PRC2 and its mark in life // *Nature*. — 2011. — т. 469, № 7330. — с. 343—349.
81. A central role for canonical PRC1 in shaping the 3D nuclear landscape / S. Boyle [и др.] // *Genes & Development*. — 2020. — 1 июля. — т. 34, № 13. — с. 931—949.

82. Polycomb Group Proteins Regulate Chromatin Architecture in Mouse Oocytes and Early Embryos / Z. Du [и др.] // *Molecular Cell*. — 2020. — февр. — т. 77, № 4. — 825—839.e7.
83. Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation / K. Kraft [и др.] // *Proceedings of the National Academy of Sciences*. — 2022. — 31 мая. — т. 119, № 22. — e2201883119.
84. Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for Pluripotency Factors and Polycomb Proteins in Genome Organization / M. Denholtz [и др.] // *Cell Stem Cell*. — 2013. — нояб. — т. 13, № 5. — с. 602—616.
85. Multiscale 3D Genome Rewiring during Mouse Neural Development / B. Bonev [и др.] // *Cell*. — 2017. — окт. — т. 171, № 3. — 557—572.e24.
86. Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells / J. D. Rhodes [и др.] // *Cell Reports*. — 2020. — янв. — т. 30, № 3. — 820—835.e10.
87. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization / A. Kojic [и др.] // *Nature Structural & Molecular Biology*. — 2018. — июнь. — т. 25, № 6. — с. 496—504.
88. Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2 / A. J. Plys [и др.] // *Genes & Development*. — 2019. — 1 июля. — т. 33, № 13. — с. 799—813.
89. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl / G. A. Busslinger [и др.] // *Nature*. — 2017. — апр. — т. 544, № 7651. — с. 503—507.
90. Enhancer dependence of cell-type-specific gene expression increases with developmental age / W. Cai [и др.] // *Proceedings of the National Academy of Sciences*. — 2020. — сент. — т. 117, № 35. — с. 21450—21458.
91. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain / K. Nowick [и др.] // *Proceedings of the National Academy of Sciences*. — 2009. — т. 106, № 52. — с. 22358—22363.

92. Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques / X. Liu [и др.] // *Genome research*. — 2012. — т. 22, № 4. — с. 611—622.
93. Evolutionary changes in promoter and enhancer activity during human corticogenesis / S. K. Reilly [и др.] // *Science*. — 2015. — т. 347, № 6226. — с. 1155—1159.
94. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain / M. W. Vermunt [и др.] // *Nature neuroscience*. — 2016. — т. 19, № 3. — с. 494—503.
95. Chromosome conformation elucidates regulatory relationships in developing human brain / H. Won [и др.] // *Nature*. — 2016. — т. 538, № 7626. — с. 523—527.
96. Comparing 3D genome organization in multiple species using Phylo-HMRF / Y. Yang [и др.] // *Cell systems*. — 2019. — т. 8, № 6. — с. 494—505.
97. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes / B. J. Beliveau [и др.] // *Proceedings of the National Academy of Sciences*. — 2012. — т. 109, № 52. — с. 21301—21306.
98. *Rust M. J., Bates M., Zhuang X.* Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM) // *Nature methods*. — 2006. — т. 3, № 10. — с. 793—796.
99. Imaging intracellular fluorescent proteins at nanometer resolution / E. Betzig [и др.] // *science*. — 2006. — т. 313, № 5793. — с. 1642—1645.
100. *Szabo Q., Bantignies F., Cavalli G.* Principles of genome folding into topologically associating domains // *Science Advances*. — 2019. — апр. — т. 5, № 4. — eaaw1668.
101. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions / D. I. Cattoni [и др.] // *Nature communications*. — 2017. — т. 8, № 1. — с. 1753.
102. Spatial organization of chromatin domains and compartments in single chromosomes / S. Wang [и др.] // *Biophysical Journal*. — 2017. — т. 112, № 3. — 217a.

103. DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells / H. Saad [и др.] // PLoS genetics. — 2014. — т. 10, № 3. — e1004187.
104. Capturing Chromosome Conformation / J. Dekker [и др.] // Science. — 2002. — 15 февр. — т. 295, № 5558. — с. 1306—1311.
105. A statistical approach for inferring the 3D structure of the genome / N. Varoquaux [и др.] // Bioinformatics. — 2014. — т. 30, № 12. — с. i26—i33.
106. Analysis of long-range chromatin interactions using Chromosome Conformation Capture / N. Naumova [и др.] // Methods. — 2012. — т. 58, № 3. — с. 192—203.
107. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture—on-chip (4C) / M. Simonis [и др.] // Nature genetics. — 2006. — т. 38, № 11. — с. 1348—1354.
108. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements / J. Dostie [и др.] // Genome research. — 2006. — т. 16, № 10. — с. 1299—1309.
109. *Schwartzman O., Tanay A.* Single-cell epigenomics: techniques and emerging applications // Nature Reviews Genetics. — 2015. — т. 16, № 12. — с. 716—726.
110. FACS-based isolation of fixed mouse neuronal nuclei for ATAC-seq and Hi-C / E. Eremenko [и др.] // STAR protocols. — 2021. — т. 2, № 3. — с. 100643.
111. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms / A. M. C. Gizzi [и др.] // Molecular cell. — 2019. — т. 74, № 1. — с. 212—222.
112. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position / J. D. Buenrostro [и др.] // Nature methods. — 2013. — т. 10, № 12. — с. 1213—1218.
113. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin / S. M. Johnson [и др.] // Genome research. — 2006. — т. 16, № 12. — с. 1505—1516.

114. High-resolution mapping and characterization of open chromatin across the genome / A. P. Boyle [и др.] // *Cell*. — 2008. — т. 132, № 2. — с. 311–322.
115. Iterative correction of Hi-C data reveals hallmarks of chromosome organization / M. Imakaev [и др.] // *Nature Methods*. — 2012. — окт. — т. 9, № 10. — с. 999–1003.
116. Comparison of computational methods for Hi-C data analysis / M. Forcato [и др.] // *Nature methods*. — 2017. — т. 14, № 7. — с. 679–685.
117. *Khrameeva E. E., Gelfand M. S.* Biases in read coverage demonstrated by interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments // *BMC bioinformatics*. т. 13. — BioMed Central. 2012. — с. 1–7.
118. Cumulative contact frequency of a chromatin region is an intrinsic property linked to its function / M. D. Samborskaia [и др.] // *PeerJ*. — 2020. — т. 8. — e9566.
119. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains / S. V. Ulianov [и др.] // *Genome Research*. — 2016. — янв. — т. 26, № 1. — с. 70–84.
120. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster* / P. V. Kharchenko [и др.] // *Nature*. — 2011. — т. 471, № 7339. — с. 480–485.
121. Mapping and analysis of chromatin state dynamics in nine human cell types / J. Ernst [и др.] // *Nature*. — 2011. — т. 473, № 7345. — с. 43–49.
122. StereoGene: rapid estimation of genome-wide correlation of continuous or interval feature data / E. D. Stavrovskaya [и др.] // *Bioinformatics*. — 2017. — т. 33, № 20. — с. 3158–3165.
123. Robust Hi-C maps of enhancer-promoter interactions reveal the function of non-coding genome in neural development and diseases / L. Lu [и др.] // *Molecular cell*. — 2020. — т. 79, № 3. — с. 521–534.
124. HiFive: a tool suite for easy and efficient HiC and 5C data analysis / M. E. Sauria [и др.] // *Genome biology*. — 2015. — т. 16. — с. 1–10.
125. HiCNorm: removing biases in Hi-C data via Poisson regression / M. Hu [и др.] // *Bioinformatics*. — 2012. — т. 28, № 23. — с. 3131–3133.

126. *Stansfield J. C., Cresswell K. G., Dozmorov M. G.* multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments // *Bioinformatics*. — 2019. — т. 35, № 17. — с. 2916–2923.
127. Binless normalization of Hi-C data provides significant interaction and difference detection independent of resolution / Y. G. Spill [и др.] // *Nature Communications*. — 2019. — т. 10, № 1. — с. 1938.
128. HiConfidence: a novel approach uncovering the biological signal in Hi-C data affected by technical biases / V. A. Kobets [и др.] // *Briefings in Bioinformatics*. — 2023. — т. 24, № 2. — bbad044.
129. Identification of alternative topological domains in chromatin / D. Filippova [и др.] // *Algorithms Mol Biol*. — 2014. — янв. — т. 9, № 1. — с. 14.
130. A chromosome folding intermediate at the condensin-to-cohesin transition during telophase / K. Abramo [и др.] // *Nature cell biology*. — 2019. — т. 21, № 11. — с. 1393–1402.
131. Two-dimensional segmentation for analyzing Hi-C data / C. Levy-Leduc [и др.] // *Bioinformatics*. — 2014. — сент. — т. 30, № 17. — с. i386–i392.
132. TopDom: an efficient and deterministic method for identifying topological domains in genomes / H. Shin [и др.] // *Nucleic Acids Res*. — 2016. — апр. — т. 44, № 7. — e70–e70.
133. *Weinreb C., Raphael B. J.* Identification of hierarchical chromatin domains // *Bioinformatics*. — 2016. — т. 32, № 11. — с. 1601–1609.
134. *Oluwadare O., Cheng J.* ClusterTAD: an unsupervised machine learning approach to detecting topologically associated domains of chromosomes from Hi-C data // *BMC Bioinformatics*. — 2017. — дек. — т. 18, № 1. — с. 480.
135. *Wang X.-T., Cui W., Peng C.* HiTAD: detecting the structural and functional hierarchies of topologically associating domains from chromatin interactions // *Nucleic Acids Research*. — 2017. — нояб. — т. 45, № 19. — e163–e163.
136. Integrative functional genomic analysis of human brain development and neuropsychiatric risks / M. Li [и др.] // *Science*. — 2018. — 14 дек. — т. 362, № 6420. — eaat7615.

137. optimalTAD: annotation of topologically associating domains based on chromatin marks enrichment / D. Smirnov [и др.] // bioRxiv. — 2023. — с. 2023–03.
138. Nuclear lamina integrity is required for proper spatial organization of chromatin in *Drosophila* / S. V. Ulianov [и др.] // Nat Commun. — 2019. — дек. — т. 10, № 1. — с. 1176.
139. Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing / L. Li [и др.] // Molecular Cell. — 2015. — апр. — т. 58, № 2. — с. 216–231.
140. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning / E. Ing-Simmons [и др.] // Nat Genet. — 2021. — апр. — т. 53, № 4. — с. 487–499.
141. In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila* / R. G. Arzate-Mejia [и др.] // Nat. Commun. — 2020. — т. 11, № 1. — с. 894.
142. CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions / A. Kaushal [и др.] // Nat. Commun. — 2021. — т. 12, № 1. — с. 1011.
143. *Weinreb C., Raphael B. J.* Identification of hierarchical chromatin domains // Bioinformatics. — 2016. — июнь. — т. 32, № 11. — с. 1601–1609.
144. OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries / L. An [и др.] // Genome Biol. — 2019. — дек. — т. 20, № 1. — с. 282.
145. CASPIAN: A method to identify chromatin topological associated domains based on spatial density cluster / H. Gong [и др.] // Computational and Structural Biotechnology Journal. — 2022. — т. 20. — с. 4816–4824.
146. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus / Y. Zhang [и др.] // Nat Commun. — 2018. — февр. — т. 9, № 1. — с. 750.
147. Boost-HiC: computational enhancement of long-range contacts in chromosomal contact maps / L. Carron [и др.] // Bioinformatics. — 2019. — т. 35, № 16. — с. 2724–2729.

148. Quantitative prediction of enhancer–promoter interactions / P. S. Belokopytova [и др.] // *Genome research*. — 2020. — т. 30, № 1. — с. 72–84.
149. *Li W., Wong W. H., Jiang R.* DeepTACT: predicting 3D chromatin contacts via bootstrapping deep learning // *Nucleic acids research*. — 2019. — т. 47, № 10. — e60–e60.
150. DeepC: predicting 3D genome folding using megabase-scale transfer learning / R. Schwessinger [и др.] // *Nature methods*. — 2020. — т. 17, № 11. — с. 1118–1124.
151. *Fudenberg G., Kelley D. R., Pollard K. S.* Predicting 3D genome folding from DNA sequence with Akita // *Nature Methods*. — 2020. — т. 17, № 11. — с. 1111–1117.
152. High-resolution TADs reveal DNA sequences underlying genome organization in flies / F. Ramirez [и др.] // *Nature communications*. — 2018. — т. 9, № 1. — с. 189.
153. A machine learning framework for the prediction of chromatin folding in *Drosophila* using epigenetic features / M. B. Rozenwald [и др.] // *PeerJ Computer Science*. — 2020. — т. 6. — e307.
154. SIRT6 is a key regulator of mitochondrial function in the brain / D. Smirnov [и др.] // *Cell Death & Disease*. — 2023. — т. 14, № 1. — с. 35.
155. Individual genome sequencing identified a novel enhancer element in exon 7 of the CSFR1 gene by shift of expressed allele ratios / S. Zhenilo [и др.] // *Gene*. — 2015. — т. 566, № 2. — с. 223–228.
156. RBM24 is a major regulator of muscle-specific alternative splicing / J. Yang [и др.] // *Developmental cell*. — 2014. — т. 31, № 1. — с. 87–99.
157. The RNA-binding protein RBM24 regulates lipid metabolism and SLC7A11 mRNA stability to modulate ferroptosis and inflammatory response / J. Zhang [и др.] // *Frontiers in Cell and Developmental Biology*. — 2022. — т. 10. — с. 1008576.
158. A novel intra-U1 snRNP cross-regulation mechanism: alternative splicing switch links U1C and U1-70K expression / T. D. Rösel-Hillgärtner [и др.] // *PLoS genetics*. — 2013. — т. 9, № 10. — e1003856.

159. U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease / B. Bai [и др.] // *Proceedings of the National Academy of Sciences*. — 2013. — т. 110, № 41. — с. 16562—16567.
160. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L / O. Rossbach [и др.] // *RNA biology*. — 2014. — т. 11, № 2. — с. 146—155.
161. IGF2 mediates resistance to isoform-selective-inhibitors of the PI3K in HPV positive head and neck cancer / M. Badarni [и др.] // *Cancers*. — 2021. — т. 13, № 9. — с. 2250.
162. TGF-beta-activated cancer-associated fibroblasts limit cetuximab efficacy in preclinical models of head and neck cancer / K. M. Yegodayev [и др.] // *Cancers*. — 2020. — т. 12, № 2. — с. 339.
163. Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia / J. Bryois [и др.] // *Nature communications*. — 2018. — т. 9, № 1. — с. 3121.
164. Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability / L. F. Rizzardi [и др.] // *Nature Neuroscience*. — 2019. — февр. — т. 22, № 2. — с. 307—316.
165. Comparison of genome architecture at two stages of male germline cell differentiation in *Drosophila* / A. A. Ilyin [и др.] // *Nucleic Acids Research*. — 2022. — т. 50, № 6. — с. 3203—3225.
166. Aging and pathological aging signatures of the brain: through the focusing lens of SIRT6 / D. Stein [и др.] // *Aging (Albany NY)*. — 2021. — т. 13, № 5. — с. 6420.
167. Application of mass spectrometry methods to the analysis of lipid composition of the human brain / E. Stekolshchikova [и др.] // *Biotekhnologiya*. — 2021. — т. 37, № 5. — с. 80—87.
168. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification / C. A. Smith [и др.] // *Analytical chemistry*. — 2006. — т. 78, № 3. — с. 779—787.

169. So you think you can PLS-DA? / D. Ruiz-Perez [и др.] // BMC bioinformatics. — 2020. — т. 21, № 1. — с. 1—10.
170. The Hitchhiker’s guide to untargeted lipidomics analysis: practical guidelines / D. Smirnov [и др.] // Metabolites. — 2021. — т. 11, № 11. — с. 713.
171. *Cavalli G., Misteli T.* Functional implications of genome topology // Nature Structural & Molecular Biology. — 2013. — март. — т. 20, № 3. — с. 290—299.
172. Inter-chromosomal contact networks provide insights into Mammalian chromatin organization / S. Kaufmann [и др.] // PloS one. — 2015. — т. 10, № 5. — e0126125.
173. History of chromosome rearrangements reflects the spatial organization of yeast chromosomes / E. E. Khrameeva [и др.] // Journal of bioinformatics and computational biology. — 2016. — т. 14, № 02. — с. 1641002.
174. *Bickmore W. A., Steensel B. van.* Genome Architecture: Domain Organization of Interphase Chromosomes // Cell. — 2013. — март. — т. 152, № 6. — с. 1270—1284.
175. The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments / T. Cremer [и др.] // FEBS Letters. — 2015. — 7 окт. — т. 589, № 20. — с. 2931—2943.
176. *Sutherland H., Bickmore W. A.* Transcription factories: gene expression in unions? // Nature Reviews Genetics. — 2009. — июль. — т. 10, № 7. — с. 457—466.
177. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6 / R. Mostoslavsky [и др.] // Cell. — 2006. — т. 124, № 2. — с. 315—329.
178. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment / J. A. Beagan [и др.] // Genome research. — 2017. — т. 27, № 7. — с. 1139—1152.
179. Quantitative differences in TAD border strength underly the TAD hierarchy in Drosophila chromosomes / A. V. Luzhin [и др.] // Journal of Cellular Biochemistry. — 2019. — т. 120, № 3. — с. 4494—4503.
180. Unraveling the mechanisms of chromatin fibril packaging / A. A. Gavrilov [и др.] // Nucleus. — 2016. — т. 7, № 3. — с. 319—324.

181. Order and stochasticity in the folding of individual *Drosophila* genomes / S. V. Ulianov [и др.] // *Nature communications*. — 2021. — т. 12, № 1. — с. 41.
182. *Wilber A., Nienhuis A. W., Persons D. A.* Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities // *Blood*. — 2011. — 14 апр. — т. 117, № 15. — с. 3945—3953.
183. Activation of the alpha-globin gene expression correlates with dramatic upregulation of nearby non-globin genes and changes in local and large-scale chromatin spatial structure / S. V. Ulianov [и др.] // *Epigenetics & Chromatin*. — 2017. — т. 10, № 1. — с. 1—19.
184. An anatomically comprehensive atlas of the adult human brain transcriptome / M. J. Hawrylycz [и др.] // *Nature*. — 2012. — т. 489, № 7416. — с. 391—399.
185. TP73-AS1 is induced by YY1 during TMZ treatment and highly expressed in the aging brain / G. Mazor [и др.] // *Aging (Albany NY)*. — 2021. — т. 13, № 11. — с. 14843.
186. Advances in the molecular genetics of gliomas - implications for classification and therapy / G. Reifenberger [и др.] // *Nat. Rev. Clin. Oncol.* — 2017. — т. 14, № 7. — с. 434—452.
187. *Sarvagalla S., Kolapalli S. P., Vallabhapurapu S.* The two sides of YY1 in cancer: A friend and a foe // *Front. Oncol.* — 2019. — т. 9. — с. 1230.
188. Long noncoding RNA HOTAIR promotes medulloblastoma growth, migration and invasion by sponging miR-1/miR-206 and targeting YY1 / J. Zhang [и др.] // *Biomed. Pharmacother.* — 2020. — т. 124, № 109887. — с. 109887.
189. The why of YY1: Mechanisms of transcriptional regulation by Yin Yang 1 / T. C. J. Verheul [и др.] // *Front. Cell Dev. Biol.* — 2020. — т. 8. — с. 592164.
190. A chromosomal connectome for psychiatric and metabolic risk variants in adult dopaminergic neurons / S. Espeso-Gil [и др.] // *Genome medicine*. — 2020. — т. 12. — с. 1—19.
191. Robust Hi-C chromatin loop maps in human neurogenesis and brain tissues at high-resolution / L. Lu [и др.] // *bioRxiv*. — 2019. — с. 744540.

192. *Belaghzal H., Dekker J., Gibcus J. H.* Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation // *Methods*. — 2017. — т. 123. — с. 56–65.
193. Hi-C: a comprehensive technique to capture the conformation of genomes / J.-M. Belton [и др.] // *Methods*. — 2012. — т. 58, № 3. — с. 268–276.
194. Extensive long-range polycomb interactions and weak compartmentalization are hallmarks of human neuronal 3D genome / I. A. Pletenev [и др.] // *bioRxiv*. — 2023. — с. 2023–08.
195. Conserved cell types with divergent features in human versus mouse cortex / R. D. Hodge [и др.] // *Nature*. — 2019. — 5 сент. — т. 573, № 7772. — с. 61–68.
196. Loops, topologically associating domains, compartments, and territories are elastic and robust to dramatic nuclear volume swelling / J. T. Sanders [и др.] // *Scientific Reports*. — 2022. — 18 марта. — т. 12, № 1. — с. 4721.
197. Two independent modes of chromatin organization revealed by cohesin removal / W. Schwarzer [и др.] // *Nature*. — 2017. — т. 551, № 7678. — с. 51–56.
198. Population-level variation in enhancer expression identifies disease mechanisms in the human brain / P. Dong [и др.] // *Nature Genetics*. — 2022. — окт. — т. 54, № 10. — с. 1493–1503.
199. A survey of human brain transcriptome diversity at the single cell level / S. Darmanis [и др.] // *Proceedings of the National Academy of Sciences*. — 2015. — 9 июня. — т. 112, № 23. — с. 7285–7290.
200. Large-Scale Functional Organization of Long-Range Chromatin Interaction Networks / K. S. Sandhu [и др.] // *Cell Reports*. — 2012. — нояб. — т. 2, № 5. — с. 1207–1219.
201. Emerging Evidence of Chromosome Folding by Loop Extrusion / G. Fudenberg [и др.] // *Cold Spring Harbor Symposia on Quantitative Biology*. — 2017. — т. 82. — с. 45–55.
202. An integrative ENCODE resource for cancer genomics / J. Zhang [и др.] // *Nature communications*. — 2020. — т. 11, № 1. — с. 3696.

203. RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation / C. Ferrai [и др.] // *Molecular systems biology*. — 2017. — т. 13, № 10. — с. 946.
204. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells / J. K. Stock [и др.] // *Nature cell biology*. — 2007. — т. 9, № 12. — с. 1428—1435.
205. PRC1 sustains the integrity of neural fate in the absence of PRC2 function / A. Sawai [и др.] // *Elife*. — 2022. — т. 11. — e72769.
206. Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration / M. von Schimmelmann [и др.] // *Nature neuroscience*. — 2016. — т. 19, № 10. — с. 1321—1330.
207. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains / E. Khrameeva [и др.] // *Genome research*. — 2020. — т. 30, № 5. — с. 776—789.
208. How to characterize the function of a brain region / S. Genon [и др.] // *Trends in cognitive sciences*. — 2018. — т. 22, № 4. — с. 350—364.
209. Comprehensive integration of single-cell data / T. Stuart [и др.] // *Cell*. — 2019. — т. 177, № 7. — с. 1888—1902.
210. *Zimmer-Bensch G.* Emerging roles of long non-coding RNAs as drivers of brain evolution // *Cells*. — 2019. — т. 8, № 11. — с. 1399.
211. Murine Falcor/LL35 lncRNA Contributes to Glucose and Lipid Metabolism In Vitro and In Vivo / E. Shcherbinina [и др.] // *Biomedicines*. — 2022. — т. 10, № 6. — с. 1397.
212. Changes in snoRNA and snRNA abundance in the human, chimpanzee, macaque, and mouse brain / B. Zhang [и др.] // *Genome biology and evolution*. — 2016. — т. 8, № 3. — с. 840—850.
213. Exceptional evolutionary divergence of human muscle and brain metabolomes parallels human cognitive and physical uniqueness / K. Bozek [и др.] // *PLoS biology*. — 2014. — т. 12, № 5. — e1001871.
214. Evidence for widespread association of mammalian splicing and conserved long-range RNA structures / D. D. Pervouchine [и др.] // *Rna*. — 2012. — т. 18, № 1. — с. 1—15.

215. SF-1 expression in the hypothalamus is required for beneficial metabolic effects of exercise / T. Fujikawa [и др.] // *Elife*. — 2016. — т. 5. — e18206.
216. Hypothalamic actions of SIRT1 and SIRT6 on energy balance / M. Quiñones [и др.] // *International Journal of Molecular Sciences*. — 2021. — т. 22, № 3. — с. 1430.
217. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data / S. Shen [и др.] // *Proceedings of the National Academy of Sciences*. — 2014. — т. 111, № 51. — E5593—E5601.
218. The changing paradigm of intron retention: regulation, ramifications and recipes / G. Monteuuis [и др.] // *Nucleic acids research*. — 2019. — т. 47, № 22. — с. 11497—11513.
219. Lipidome evolution in mammalian tissues / E. Khrameeva [и др.] // *Molecular biology and evolution*. — 2018. — т. 35, № 8. — с. 1947—1957.
220. Lipidome determinants of maximal lifespan in mammals / K. Bozek [и др.] // *Scientific Reports*. — 2017. — т. 7, № 1. — с. 5.
221. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans / E. E. Khrameeva [и др.] // *Nature communications*. — 2014. — т. 5, № 1. — с. 3584.
222. The complete genome sequence of a Neanderthal from the Altai Mountains / K. Prüfer [и др.] // *Nature*. — 2014. — т. 505, № 7481. — с. 43—49.
223. A draft sequence of the Neandertal genome / R. E. Green [и др.] // *science*. — 2010. — т. 328, № 5979. — с. 710—722.
224. A fine-scale chimpanzee genetic map from population sequencing / A. Auton [и др.] // *science*. — 2012. — т. 336, № 6078. — с. 193—198.
225. Identifying recent adaptations in large-scale genomic data / S. R. Grossman [и др.] // *Cell*. — 2013. — т. 152, № 4. — с. 703—713.
226. Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry / P. Flegontov [и др.] // *Scientific reports*. — 2016. — т. 6, № 1. — с. 20768.

227. *Akkuratov E. E., Gelfand M. S., Khrameeva E. E.* Neanderthal and Denisovan ancestry in Papuans: A functional study // *Journal of Bioinformatics and Computational Biology*. — 2018. — т. 16, № 02. — с. 1840011.
228. Differences in lipidome and metabolome organization of prefrontal cortex among human populations / А. Tkachev [и др.] // *Scientific Reports*. — 2019. — т. 9, № 1. — с. 18348.
229. Metabolome signature of autism in the human prefrontal cortex / I. Kurochkin [и др.] // *Communications biology*. — 2019. — т. 2, № 1. — с. 234.
230. Long-term fluoxetine administration causes substantial lipidome alteration of the juvenile macaque brain / А. Tkachev [и др.] // *International journal of molecular sciences*. — 2021. — т. 22, № 15. — с. 8089.
231. Reduced purine biosynthesis in humans after their divergence from Neandertals / V. Stepanova [и др.] // *Elife*. — 2021. — т. 10. — e58741.
232. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards / S. Fishilevich [и др.] // *Database*. — 2017. — т. 2017. — bax028.

Список рисунков

- 1 Общие принципы организации хроматина. (А) Схематичное изображение основных уровней организации хроматина в ядре. (Б) Уровни организации хроматина в данных Hi-C. Яркость пикселей на картах Hi-C пропорциональна частотам контактов. 6
- 1.1 Схематическая иллюстрация основных этапов протокола Hi-C. Адаптировано из [110]. 42
- 1.2 Схематическая иллюстрация основных этапов протокола ATAC-seq. Адаптировано из [110]. 44
- 2.1 Зависимость между состояниями хроматина и частотами контактов (CCF) в хроматине. (А) Определение суммарных ненормализованных (до итеративной коррекции) частот контактов Hi-C, далее сокращенно обозначаемых как CCF (Cumulative Contact Frequency). (Б) В анализе участвовали четыре клеточных линии для человека – HMEC, HUVES, NHEC и K562, клеточная линия CH12-LX для мыши и клеточная линия S2 для дрозофилы. (В) Hi-C карта полного генома человека, совмещенная с графиком распределения процентного содержания состояний хроматина. На Hi-C карте красным показаны участки с аномально высокой CCF, зеленым показаны границы хромосом. На прилежащем к Hi-C карте графике для каждого участка показано процентное содержание каждого из состояний. Состояния отмечены цветами (см. легенду на панели Д). (Г) CCF резко меняется при переходе от одного компартмента к другому. (Д) Барплот зависимости распределения различных состояний хроматина от CCF. По оси X по возрастанию отложены значения CCF. По оси Y - процентные содержания каждого из состояний хроматина для участков с данными значениями частоты контактов. Адаптировано из [118]. 50

- 2.2 Матрицы корреляций для состояний хроматина и ССФ в геноме человека для разных хромосом. Адаптировано из [118]. 51
- 2.3 ССФ является независимым от положения свойством областей хроматина, сохраняющимся в синтенных переходах. (А) Уникальные корреляционные паттерны для небольших хромосом не являются следствием небольшого размера выборки, а обусловлены синтенными областями небольших хромосом (клеточная линия человека НМЕС). Левая панель: распределение корреляций между ССФ и состоянием активного промотора хроматина для случайных фрагментов хромосомы 2 (синие столбцы) по сравнению с реальной корреляцией для хромосомы 22 (красная линия). Правая панель: распределение корреляций между ССФ и состоянием гетерохроматина для случайных фрагментов хромосомы 2 (синие столбцы) по сравнению с реальной корреляцией для хромосомы 22 (красная линия). (Б) Схематичное изображение синтенной области между двумя хромосомами генома человека и мыши. (В-Г) ССФ у человека по сравнению с ССФ у мыши для синтенных областей в больших человеческих хромосомах и больших мышинных хромосомах (В), а также в малых человеческих хромосомах и больших мышинных хромосомах (Г). Каждая точка представляет собой синтенную область размером 1 Мб. Адаптировано из [118]. 55

- 2.4 Описание алгоритма HiConfidence. (А) Ключевые этапы процедуры HiConfidence. Сперва разница между картами эксперимента и контроля Hi-C рассчитывается для каждой биологической повторности (r_1 и r_2) отдельно. Далее оценивается достоверность контакта для каждого пикселя полученных карт. (Б) Значение достоверности обратно пропорционально разнице между биологическими репликами, деленной на их среднее значение (в степени k). (В) Параметр k может быть оптимизирован путем максимизации коэффициента корреляции Пирсона между дифференциальными профилями инсуляции биологических повторностей. Пунктирной линией отмечено оптимальное значение k . Стрелка указывает на максимальный коэффициент корреляции Пирсона. Адаптировано из [128]. 58
- 2.5 Технические проблемы мешают прямому сравнению наборов данных Hi-C. (А) Дизайн экспериментов Hi-C, используемых для оценки эффективности подхода HiConfidence. Красными стрелками показана деплеция деацетилазы гистонов (HDAC-dep) и ингибирование деацетилазы гистонов трихостатином А (HDAC-inh), повышающее уровень ацетилирования. Синими стрелками отмечено ингибирование гистоновых ацетилтрансфераз куркумином (HAT-inh), снижающее уровень ацетилирования. (Б) Корреляция карт Hi-C для обработанных (tr.) и контрольных (c.) повторностей, оцененная с использованием коэффициента корреляции с поправкой на ожидаемое (SCC). Цвета показывают SCC. (В) Корреляция дифференциальных профилей инсуляции для карт log2FC Hi-C, представляющих собой разницу между обработанными репликами (r.) и соответствующими контрольными. Цвета отображают коэффициенты корреляции Пирсона. Адаптировано из [128]. 59

- 2.6 HiConfidence извлекает биологический сигнал из данных Hi-C, затронутых техническим шумом. (А) Корреляция профилей дифференциальной инсуляции для карт \log_2FC Hi-C, представляющих собой разницу между обработанными повторностями (г.) и соответствующими контролями. Цвета как на рисунке 2.5. (Б) Распределение средних значений достоверности ТАДов для экспериментов HDAC-dep (пунктирная линия), HDAC-inh (пунктирная линия) и NAT-inh (сплошная линия). (В) Корреляция изменений плотности хроматина в ТАДах между повторностями эксперимента HDAC-inh до (слева) и после (правой) HiConfidence. (Г и Д) Средние ТАДы на картах \log_2FC Hi-C, представляющие собой разность между обработанными повторностями и соответствующими контролями, до (Г) и после (Д) HiConfidence. Пунктирные линии выделяют области с заметными улучшениями в результате процедуры HiConfidence. Адаптировано из [128]. 61
- 2.7 HiConfidence значительно улучшает данные Hi-C и приводит к более консистентным результатам на уровне компартментов. (А) Средняя разница в контактах между обработанными и контрольными повторностями, рассчитанными в областях, выделенных на рис. 2.6Г и 2.6Д до (серый) и после (белый) HiConfidence. (Б) Статистическая значимость улучшений в результате процедуры HiConfidence в ТАДах. (В-Г) Разность седловых графиков, представляющих собой выраженность компартментов, между обработанными повторностями и соответствующими контролями, до (В) и после (Г) HiConfidence. Адаптировано из [128]. 63

- 2.8 Сравнение HiConfidence и multiHiScompare. (А) Корреляция дифференциальных профилей инсуляции карт \log_2FC Hi-C, представляющих собой разницу между обработанными повторностями (r .) и соответствующими контролями. (Б) Корреляция изменений плотности ТАДов между двумя техническими повторностями. Слева – необработанные данные, в центре – данные после использования нашего подхода, справа – данные после использования multiHiScompare. Адаптировано из [128]. 65
- 2.9 Проверка подхода HiConfidence на внешнем наборе данных Hi-C. (А) Корреляции карт \log_2FC Hi-C, представляющих собой разность между соседними временными точками, до (левая панель) и после (правой панели) HiConfidence. Цвета представляют собой коэффициенты корреляции Пирсона. (Б) Корреляционные изменения после HiConfidence между парами повторностей и между другими случайными парами экспериментов. (В) Средняя разница в плотности ТАДов между временными точками до (левая панель) и после (правая панель) HiConfidence. Цвета представляют собой разницу между картами Hi-C между двумя временными точками. Пунктирные линии выделяют области с заметными улучшениями после HiConfidence. Адаптировано из [128]. 67
- 2.10 Общая схема работы алгоритма optimalTAD. Адаптировано из [137]. 69

- 2.11 Проверка алгоритма на данных Uljanov. (А) Амплитудные кривые, показывающие динамику разницы Δh между средними значениями ChIP-seq внутри и между ТАДами по значениям γ . (Б) Тепловая карта, иллюстрирующая иерархическую кластеризацию экспериментов на основе меры евклидова расстояния. (В) Амплитуда уровней ацетилирования гистонов в наборе ТАДов, предсказанном алгоритмом optimalTAD в наборе данных Uljanov. (Г) Сравнение медианных размеров ТАДов и межТАДов, полученных для экспериментальных образцов. (Д) Распределение числа ТАДов, предсказанных optimalTAD, в 2L, 2R, 3L, 3R и X хромосомах. (Е) Диаграмма Венна, показывающая количество границ ТАДов, общих для экспериментальных повторностей. Адаптировано из [137]. 71
- 2.12 Проверка алгоритма с использованием данных мутантных эмбрионов Tollrm9/rm10. (А) Карта Hi-C с границами ТАДов, предсказанными для H3K27ac Rep. 1 (окрашен синим цветом) и H3K27ac Rep. 2 (окрашен красным) в геномном интервале chrX: 6 080 000–9 000 000. (Б) Диаграмма Венна, иллюстрирующая общие границы между ТАДами, предсказанными в двух повторностях H3K27ac (синий кружок соответствует Rep. 1, а красный кружок показывает домены Rep. 2). (В) Амплитудные кривые, показывающие изменения Δh в зависимости от значений γ для повторностей Rep. 1 и Rep. 2 профилей H3K27ac ChIP-seq. Адаптировано из [137]. 74
- 2.13 Сравнение optimalTAD с другими инструментами поиска ТАДов. (А) Гистограмма, показывающая значения амплитуды Δh , рассчитанные с использованием методов OptimalTAD, TopDom, ClusterTAD и Caspian (слева направо) для карты Hi-C из Li. Анализ с помощью инструмента Caspian проводился в трех возможных режимах («Манхэттенский», «Чебышёвский», «Евклидовый»). (Б-Д) Сравнение кривых ацетилирования, полученных с помощью инструментов поиска ТАДов, выбранных для анализа. Адаптировано из [137]. 76

- 2.14 Предсказание ТАДов, межТАДов и границ ТАДов в клетках S2 дрожофилы с использованием моделей логистической регрессии, основанных на геномном распределении активных меток хроматина или архитектурных белков dCTCF и Su(Hw). Показаны кривые ROC и значения AUC (площадь под кривой). Адаптировано из [119]. 80
- 2.15 Схема реализованной двунаправленной рекуррентной нейронной сети LSTM с одним выходом. Значения x_1, \dots, x_t представляют собой участки ДНК с размером входного окна t , h_1, \dots, h_t являются скрытыми состояниями модели RNN, $y_{t/2}$ представляет соответствующее значение параметра γ целевого значения средней ячейки $x_{t/2}$. Обратите внимание, что каждая ячейка x_i характеризуется вектором хроматиновых меток данных ChIP-chip. Адаптировано из [153]. 81
- 2.16 Взвешенная MSE с использованием четырех из пяти хроматиновых маркеров в качестве входных данных biLSTM RNN. Каждый цвет соответствует объекту, который был исключен из входных данных. Обратите внимание, что на модель больше всего влияет исключение фактора Chr1z. Адаптировано из [153]. 82
- 2.17 Схема пробоподготовки образцов мозга для масс-спектрометрического анализа липидов: замороженные блоки мозга (а), выбор участков для вырезки образцов в соответствии с атласом мозга (б), выдерживание в криомикротоме при температуре $-15\text{ }^\circ\text{C}$ (в), вырезка образцов (г), взвешивание (д). Адаптировано из [167]. 92
- 2.18 Масса и время удержания липидов при ВЭЖХ-МС образцов мозга человека, подвергшихся процедуре аннотации с ручной верификацией. Размер точек отражает среднюю интенсивность липида в мозге. Адаптировано из [167]. 94

3.1 Вычислительные процедуры, применяемые для демонстрации влияния пространственного выравнивания хромосомных плеч на частоту перестроек. (А) $|X - Y|$ - мера пространственного выравнивания между генами, расположенными на разных хромосомах в Рабль-подобной конфигурации, где X и Y - геномные расстояния между этими генами и центромерами. (Б) Схема, иллюстрирующая определение синтении. Сначала формируется таблица гомологичных генов, центрированная на предковом гене G. Столбцы гомологии выделены серым цветом. Предполагается, что гены $G_{central}$ и G_{other} являются синтенными, если они расположены в пределах 20 генов друг от друга как в геноме предка, так и в геноме потомка. (В) Предсказание перестроек как событий между двумя блоками синтении (выделены серым цветом), расположенными на разных хромосомах в геноме потомка, и на одной и той же хромосоме в пределах 20 генов друг от друга в геноме предка. (Г) Пространственное расстояние между генами X и Y с предсказанным событием перестройки (зеленый) и парами контрольных генов (серый) у *S. cerevisiae*. Адаптировано из [173]. 99

- 3.2 SIRT6 регулирует экспрессию генов. (А) Схематическая иллюстрация дизайна экспериментов. Данные транскриптома были получены для образцов мозга мышей дикого типа (WT) и нокаутированных по SIRT6 (brSIRT6-KO). Метаболомные профили WT и brSIRT6-KO были получены для эмбриональных стволовых клеток мыши (mESC). (Б) График PCA, показывающий разделение между образцами WT (оранжевый) и brSIRT6-KO (коричневый). Оранжевые и коричневые овалы представляют собой доверительные эллипсы групп WT и brSIRT6-KO. (В) График, показывающий ДЭ гены у мышей brSIRT6-KO по сравнению с мышами WT. Красные точки указывают на значимо измененные гены, а серые точки представляют незначимые гены. (Г) Анализ GO, показывающий топ-10 обогащенных биологических процессов для генов с пониженной экспрессией. Каждый кружок соответствует обогащенному термину GO и варьируется по размеру в зависимости от количества значимых генов, принадлежащих этому термину. Адаптировано из [154]. 101
- 3.3 SIRT6 регулирует уровни метаболитов. (А) График PCA, показывающий разделение между группами WT (оранжевые круги) и SIRT6-KO (коричневые круги) на основе метаболомных профилей. Оранжевые и коричневые овалы представляют собой доверительные эллипсы групп WT и SIRT6-KO. (Б) График, иллюстрирующий различающееся содержание метаболитов между образцами WT и SIRT6-KO. Круговая диаграмма (справа) демонстрирует количество повышенных (красный), пониженных (синий) и незначимых (серый) метаболитов в анализе. Адаптировано из [154]. 102

- 3.4 Компартиментализация ослабевает при переходе от SpG к SpCs. (А) Седловые графики, показывающие значения \log_2 обогащения частот взаимодействий для внутрихромосомных контактов во всех хромосомах в зависимости от их значений PC1 в SpG (левая панель) или SpCs (правая панель). (Б) Седловой график, показывающий результат вычитания значений обогащения внутрихромосомных контактов во всех хромосомах в SpCs из значений в SpG. (В) Сила активных и неактивных компартиментов в SpG и SpCs. P-значения были рассчитаны с использованием теста Вилкоксона. (Г) Кривые $P_c(s)$ для SpG (синяя кривая) и SpCs (красная кривая). Адаптировано из [165]. . 104
- 3.5 Степень активности хроматина определяет его пространственную структуру, а именно положение границ ТАДов. (А) Схема, иллюстрирующая предложенную нами модель сворачивания хроматина в ТАДы в соответствии с ассоциацией нуклеосом. Высокий уровень ацетилирования хроматина в областях генома, содержащих активно транскрибируемые гены, препятствует упаковке хроматина в ТАДы из-за снижения межнуклеосомных взаимодействий. Адаптировано из [119]. (Б) Схема, иллюстрирующая иерархическую вложенность ТАДов, согласно данным Hi-C для популяций клеток. Адаптировано из [179]. 106
- 3.6 Нокаут SIRT6 влияет на экспрессию генов в головном мозге. (А) Схема эксперимента: РНК была выделена из мозга трех мышей SIRT6-KO, трех мышей дикого типа (WT) того же возраста (21 день), и трех старых мышей WT (22-26 месяцев). (Б) Иерархическая кластеризация образцов. Цвета показывают коэффициенты корреляции Пирсона. (В) Уровни экспрессии ДЭ генов при сравнении между SIRT6-KO и WT. (Г) Уровни экспрессии ДЭ генов при сравнении объединения SIRT6-KO и старых мышей WT по сравнению с молодыми мышами WT. Адаптировано из [166]. 110

- 3.7 SIRT6 и YY1 демонстрируют сходные регуляторные функции. (А) Корреляции экспрессии в головном мозге 6 различных человек, основанные на данных Allen Brain Atlas. (Б) Диаграмма Венна для генов, совместно экспрессируемых либо с SIRT6, либо с YY1, и их перекрывание. (В) Обогащенные категории биологических процессов GO, частью которых являются перекрывающиеся гены. (Г) Анализ обогащения KEGG для генов с пиками ChIP-seq как в данных SIRT6, так и в данных YY1. (Д) Вестерн-блоты экспериментов по ко-иммунопреципитации SIRT6 и YY1. *** - значение $p < 0,001$. Адаптировано из [166]. 112
- 3.8 Схематическая иллюстрация подготовки ядер для протокола Hi-C в клетках головного мозга. Адаптировано из [110]. 115
- 3.9 Организация хроматина в клетках NeuN+ и NeuN-. (А) Анатомическая локализация анализируемой области мозга, экспериментальная процедура и дизайн исследования. Эксперименты Hi-C проводили в клетках NeuN+ и NeuN-, изолированных с помощью FACS из речевой области Вернике четырех доноров. (Б) График анализа главных компонент, основанный на вариации показателя изоляции (IS) среди всех созданных карт Hi-C. Цвета обозначают клетки NeuN+ и NeuN- здесь и на панели С. (В) Взаимодействия внутри всех хромосом (цис-контакты, левая панель) и между всеми парами хромосом (транс-контакты, правая панель), рассчитанные отдельно для клеток NeuN+ и NeuN-. Звездочками отмечены Р-значения критерия Вилкоксона: **** - $p < 0,00001$, ns - $p > 0,05$. Адаптировано из [194]. 116

- 3.10 Различия в компартаментах хроматина между клетками NeuN+ и NeuN-. (А) Фрагмент Hi-C карты с собственным вектором первой главной компоненты (PC1) для клеток NeuN+ (слева) и NeuN- (справа). Положительные значения PC1 соответствуют А компартменту, а отрицательные значения - В компартменту. (Б) Частота взаимодействий на Hi-C карте геномных областей, упорядоченных по соответствующему рангу PC1 (седловая диаграмма) для клеток NeuN+ (слева) и NeuN- (справа). (В) Сила взаимодействия компартментов, рассчитанная как самое сильное взаимодействие внутри компартмента, деленное на самое слабое межкомпарментное взаимодействие. (Г) Распределение геномных бинов размером 100 т.п.н. между типами компартментов в клетках NeuN+ и NeuN-. (Д) Более быстрое спадение автокорреляционной функции NeuN- PC1, демонстрирующее более тонкую компарментализацию NeuN-хроматина. Адаптировано из [194]. 118
- 3.11 Клеточно-специфичные различия ТАДов в клетках NeuN+ и NeuN-. (А) Примеры дифференциальных профилей ТАДов в дифференциально экспрессированных генах нейронов. Показана локализация генов RBFOX3, GALNT17, ATR2B1. (Б) Диаграмма Венна для пересечения нейронных и не-нейронных границ ТАДов, созданная на основе профилей изоляции. (В) Полногеномное покрытие шестью состояниями хроматина для клеток NeuN+ (слева) и NeuN- (справа). (Г) Ящички с усами для плотностей ТАДов, относящихся к группам с наибольшим покрытием состояниями хроматина RepPC или EnhA. (Д) Средний ТАД. Адаптировано из [194]. 120

- 3.12 Структура петель в NeuN+ и NeuN- клетках. (А) Фрагменты карт Hi-C с разрешением 5 КБ для областей взаимодействий специфических для нейронов генов с энхансерами. Треки энхансера отображаются в верхней части карт Hi-C. (Б) Диаграмма Венна для пересечения положений нейронных и не-нейронных петель. (В) Ящики с усами для плотности петель, относящихся к группам с наибольшим покрытием состояниями хроматина RepPC или EnhA. (Г) Группы петель, определяемые по увеличению отношения плотности NeuN+/NeuN-. (Д) Графики распределения длины петли по четырем группам. (Е) Распределение состояний хроматина в нейронах по группам петель. Адаптировано из [194]. 123
- 3.13 Свойства нейрональных точек. (А) Фрагмент карты Hi-C с нейрональными точками, присутствующими в NeuN+ (вверху), но отсутствующими в NeuN- (внизу). (Б) Средний сигнал Hi-C для нейрональных точек (слева) и соответствующих им пар локусов в NeuN- (справа). (В) Значимые меж-хромосомные взаимодействия локусов нейрональных точек в NeuN+ и NeuN-. (Г) Статистика размера (порядка) клик (сетей), созданных нейрональными точками (фиолетовые столбики) или локусами нейрональных точек в NeuN- (оранжевые столбики). (Д) Средний профиль H3K9me3 ChIP-seq в якорях нейрональных точек. Адаптировано из [194]. 125
- 3.14 Карта транскриптома мозга человека, шимпанзе, бонобо и макаки. (А) Филогенетические взаимоотношения между анализируемыми видами. Цветом выделены ветви, используемые при анализе человек-специфичности экспрессии. (Б) Анатомическая локализация 33 анализируемых областей головного мозга. Цвета обозначают кластеры регионов, определенные нами на основании подобия уровней экспрессии. (В) Графики t-SNE, основанные на вариабельности экспрессии для всех 422 проанализированных образцов. Цвета точек обозначают виды (слева) или кластеры регионов (справа). Адаптировано из [207]. 128

- 3.15 Анализ специфичных для человека различий в экспрессии генов, с учетом области мозга. (А) Количество генов, демонстрирующих специфические для человека различия в экспрессии в каждой области мозга. Различия были определены как те, которые демонстрируют двукратную разницу в экспрессии между человеком и макакой по сравнению с разницей между шимпанзе и макакой или бонобо и макакой. (Б) Человек-специфичность каждой области мозга, оцененная как соотношение различий в экспрессии, специфичных для человека, и различий в экспрессии, специфичных для шимпанзе или бонобо. Кружки показывают среднее значение сравнений на основе данных о шимпанзе и бонобо, а линии охватывают разницу между двумя оценками. Более темными кругами отмечены области мозга, демонстрирующие избыток специфических для человека различий в экспрессии по сравнению с обоими видами обезьян. Адаптировано из [207]. . . . 129
- 3.16 Функциональные термины геной онтологии (GO), обогащенные специфическими для человека различиями в экспрессии, присутствующими более чем в 10 из 33 областей мозга. Размер кружочков отражает долю генов в пределах термина GO среди всех генов, детектированных в данной области мозга. Цвет кружочков указывает на Р-значение с поправкой на множественное тестирование. Адаптировано из [207]. 131
- 3.17 Анализ экспрессии генов в мозге человека и других приматов с разрешением в одну клетку. (А) Дизайн эксперимента snRNA-seq. (Б) График t-SNE для ядер в передней поясной коре (АС), окрашенных по видам, для передней поясной коры мозга после применения методики интеграции с помощью программы Seurat 3.0 [209] (слева) и аннотация кластеров t-SNE по клеточным типам (справа). Адаптировано из [207]. 132

- 3.18 Анализ эволюции экспрессии в трех областях мозга с разрешением в одну клетку. (А) Скорость эволюции разных типов клеток в каждой области мозга. Линии обозначают стандартное отклонение средних оценок. (Б) Человек-специфичность, рассчитанная для каждого типа клеток в каждой из трех областей мозга. Это значение представляет собой количество генов с экспрессией, специфичной для человека, деленное на количество генов с экспрессией, специфичной для шимпанзе и бонобо. Прямоугольники показывают медиану и первый и третий квартили распределения, а усы охватывают 1,5 межквартильных диапазона. (В) Корреляции уровней экспрессии между типами клеток, рассчитанное на основе средних уровней экспрессии генов внутри кластеров у человека. Цвета показывают коэффициенты корреляции Пирсона. In - тормозящие нейроны; Ex - возбуждающие нейроны; OPC - предшественники олигодендроцитов; Ast - астроциты; OD - олигодендроциты; MG - микроглия. (Г) Корреляции человек-специфичности между типами клеток, рассчитанные на основе сравнения с шимпанзе и бонобо в 1000 случайных наборов клеток. Адаптировано из [207]. 134
- 3.19 Специфичный для человека паттерн экспрессии SNORA29 и лежащий в его основе генетический механизм. (А) Уровни экспрессии различных мРНК, показывающие изменения экспрессии, характерные для эволюционных линий человека, шимпанзе и макаки. (Б) Экспрессия SNORA29 у четырех видов. (В-Д) Предсказания вторичной структуры SNORA29 у людей/неандертальцев, шимпанзе и макак-резусов. Красные стрелки указывают на мутации, характерные для человека. Адаптировано из [212]. 137

- 3.20 Различия в концентрации липидов в зависимости от вида для коры больших полушарий головного мозга (CX), почек (KD), сердца (HT), мышц (ML), печени (LV) и мозжечка (CB). Показано количество липидов, демонстрирующих значительные видоспецифические различия в их содержании в каждой из 104 комбинаций тканей и линий (нормализовано на филогенетические расстояния). Усы показывают вариабельность оценок, рассчитанных путем случайной выборки из трех особей каждого вида. Цвета звездочек и столбиков указывают на значимость разницы между наблюдаемым распределением чисел и случайным ожидаемым (пермутационный тест, ** и красный — P-значение < 0,05, * и розовый — P-значение < 0,1). В крайнем правом столбце показан кумулятивный эффект на эволюционной линии, рассчитанный как среднее \log_{10} (P-значение) разницы между наблюдаемым и случайным количеством видоспецифичных липидов в ткани. Адаптировано из [219]. 142
- 3.21 Сравнение уровней экспрессии и содержания липидов в префронтальной коре человека и шимпанзе. (А) Отличия содержания липидов между популяциями шимпанзе и людей африканского (AF, n=4), азиатского (AS, n=5) и европейского (EU, n=5) происхождения для участников пути катаболизма липидов (LCP, красный, n=1090 масс-спектрометрических пиков) и метаболитов в других метаболических путях (серый, n=163 масс-спектрометрических пика). Цифры на графике показывают P-значения (пермутационный тест). Все прямоугольники на этой и других панелях показывают квартили и медиану, а усы охватывают 0,5 межквартильного диапазона. (Б) Отличия экспрессии генов между популяциями шимпанзе и человека для генов LCP, непосредственно связанных с липидами из категории LCP, показанными на панели А (красный, n=6 экспрессируемых генов) и остальных генов из категории LCP (серый, n=26 экспрессируемых генов). Оценка значимости была проведена таким же образом, как и на панели А. Адаптировано из [221]. 144

- 3.22 Пропорции NLS в современных человеческих популяциях. (А) Геномные расстояния между 11 популяциями современного человека и неандертальцами; синий - по всему геному; красный - в генах LCP. Максимальная длина полосы соответствует частоте NLS, равной 30%. (Б) Средние доли NLS в современных популяциях Африки (AF), Европы (EU) и Азии (AS), рассчитанные на основе данных о последовательностях из проекта "1000 геномов"; синий: по всему геному (GW, 1158559 сайтов), красный: в генах LCP (LCP, 498 сайтов). Усы показывают стандартное отклонение. Адаптировано из [221]. . . . 146
- 3.23 Положительный отбор в генах LCP. (А) Сигналы положительного отбора в генах LCP оценивались с использованием метрики CMS. Черные квадраты показывают частоту сайтов с повышенными значениями CMS (>1), потенциально указывающую на области генома, подвергшиеся недавнему положительному отбору в генах LCP, нормализованную на частоту таких сайтов во всех аннотированных генах в пределах одной популяции. Прямоугольники показывают распределение (квартили и медиану) нормализованных оценок частот сайтов, полученных с помощью 1000 бутстрепов по 38 генам LCP, а усы охватывают 0,5 межквартильного диапазона. Цифры на графике показывают P-значения (пермутационный тест) в сравнении популяций Африки (AF) или Азии (AS) с популяциями Европы (EU). (Б) Средние доли NLS в современных популяциях AF, EU, AS, рассчитанные на основе данных о последовательностях из проекта "1000 геномов"; синий - весь геном (GW, 1158559 сайтов); красный - все гены LCP (LCP, 498 сайтов); черный - гены LCP, связанные со специфичными для Европы изменениями содержания метаболитов (MC, 114 сайтов). Усы показывают стандартное отклонение. Адаптировано из [221]. . . . 148

- 3.24 Краткое описание основных функциональных путей, обогащенных генами, связанными с метаболитами из двух категорий: (1) 202 метаболита с изменениями при РАС ("связанные с РАС"); (2) специфичные для человека метаболиты ("человек-специфичные"). Размер каждого кружка пропорционален количеству генов в пути, связанном с метаболитами данной категории. Цвет каждого кружка показывает Р-значения с поправкой на множественное тестирование. Адаптировано из [229]. 152
- 3.25 Перекрывание между энхансерами, связанными с 1271 генами, демонстрирующими специфичную для человека экспрессию в данных snRNA-seq, и активными в мозге цис-регуляторными элементами, расположенными в HAR [94]. Гистограмма представляет собой распределение значений перекрывания, рассчитанных путем случайной подвыборки 1271 гена из 9138 генов, экспрессированных в головном мозге, выполненной 1000 раз. Красной пунктирной линией отмечено фактическое перекрывание (n=98). Адаптировано из [207]. 155

Список таблиц

1	Предсказания <code>optimalTAD</code> , полученные для датасета Li с заполнением и без заполнения пропущенных значений в данных Hi-C.	73
2	Частоты контактов по данным Hi-C для генов с повышенной экспрессией.	87
3	Частоты контактов по данным Hi-C для генов с пониженной экспрессией.	87