

Автономная некоммерческая образовательная организация высшего образования
"Сколковский институт науки и технологий"

На правах рукописи

Кононкова Анна Дмитриевна

**Структура хроматина дрозофилы в контексте влияния белков ядерной
периферии и процессов, ассоциированных с ранними стадиями
сперматогенеза**

Специальность 1.5.8 —

«математическая биология, биоинформатика»

Диссертация на соискание ученой
степени кандидата биологических наук

Научный руководитель:
доктор биологических наук
Храмеева Екатерина Евгеньевна

Москва — 2024

Содержание

ВВЕДЕНИЕ.....	4
Глава 1. Обзор литературы.....	14
1.1. Трехмерная организация хроматина: методы детекции и основные структурные характеристики	14
1.1.1 Общие представления о методах детекции структуры хроматина	14
1.1.2 Методы детекции структуры хроматина в эпоху секвенирования	15
1.1.3 Принцип работы метода Hi-C. Первичная обработка данных секвенирования.....	17
1.1.4 Основные характеристики трехмерной структуры хроматина, выявленные методом Hi-C.....	19
1.2. Особенности организации хроматина дрозофилы в контексте ключевых опосредующих факторов	23
1.2.1 Межхромосомные взаимодействия	23
1.2.2 Внутрихромосомные взаимодействия.....	24
1.3 Взаимодействие хроматина с ядерной оболочкой	33
1.3.1 Взаимосвязь между ядерной ламинной и структурой хроматина	33
1.3.2 Нуклеопорины	36
1.3.3 Локализация нуклеопоринов и их роль в положительной регуляции транскрипции ..	37
1.3.4 Репрессивные функции нуклеопоринов.....	38
1.3.5 Нуклеопорин Elys.....	39
1.4 Процесс сперматогенеза у дрозофилы как совокупность факторов, потенциально влияющих на структуру хроматина.....	41
Глава 2. Материалы и Методы.....	45
2.1 Предобработка сырых данных, генерация карт Hi-C, оценка качества эксперимента	46
2.1.1 Генерация карт контактов	47
2.1.2 Удаление артефактов и оценка качества данных	48
2.2 Анализ карт контактов.....	49
2.2.1 Структура компартментов	49
2.2.2 ТАДы	53
2.2.3 Анализ локальной плотности хроматина.....	56
2.2.4 Определение сильных и слабых границ ТАДов в контрольных клетках S2	58
2.2.5 Анализ парных взаимодействий в группах генов	59
2.2.6 Детекция петель.....	62
2.2.7 Анализ данных по экспрессии	62
2.2.8 Анализ функционального обогащения.....	62
2.2.9 Построение средних тепловых карт	63
2.2.10 Оценка различий в графике шкалирования	63
2.2.11 Публичные данные и использованные программы	63
Глава 3. Результаты и обсуждение	65

3.1 Изменение структуры хроматина при активации транскрипции сперматоцитспецифичных генов на ранних стадиях сперматогенеза дрозофилы.....	65
3.1.1 Изменение структуры компартментов	66
3.1.2 Взаимодействия сперматоцитспецифичных генов	70
3.1.3 Динамика границ ТАДов на ранних стадиях сперматогенеза дрозофилы	74
Обобщение основных результатов параграфа и их обсуждение	77
3.2 Структура хроматина и морфология сперматоцитов	79
3.3 Роль нуклеопорина Elys в регуляции плотности упаковки хроматина в клетках S2 дрозофилы.....	82
3.4.1 Структура компартментов	83
3.4.2 Компактизация и взаимодействие ТАДов	84
3.4.3 Структура хроматина в границах ТАДов.....	87
3.4.4 Локальные изменения структуры хроматина	88
3.4.5 Влияние деплеции Elys на петли и энхансерно-промоторные взаимодействия	90
Обобщение результатов параграфа и их обсуждение.....	91
Заключение	92
Благодарности.....	95
Список сокращений.....	96
Список литературы	97
Список рисунков	104
Список таблиц	105
Приложение	106

ВВЕДЕНИЕ

Информация, содержащаяся в геноме, лежит в основе видовой идентичности, но не объясняет морфологическое разнообразие клеток внутри одного организма и способность реагировать на изменяющиеся условия. Исследования последних десятилетий говорят о том, что существенную роль в функциональной регуляции клетки играет пространственная структура хроматина, которая в свою очередь также является объектом регуляции. Для изучения сложной, компактной упаковки ДНК было предложено немало подходов, среди которых можно выделить семейство методов захвата конформации хроматина и, в частности, протокол Hi-C, позволяющий определять попарные частоты взаимодействий любых двух локусов в масштабах всего генома [1].

Благодаря методу Hi-C стало известно, что хроматин млекопитающих, дрозофилы и некоторых других высших многоклеточных разделен на активный и неактивный компартменты, которые в свою очередь состоят из топологически ассоциированных доменов (ТАДов) и петель хроматина [1]. В то же время было обнаружено, что, подчиняясь базовым принципам трехмерной организации, структура хроматина варьируется в разных организмах, типах клеток, стадиях развития и иногда внешних условиях [2–4].

Дрозофила¹ занимает важное место в ряду модельных организмов. Значительная доля генов плодовой мушки (около 60%) имеет гомологи в геноме человека, включая гены, связанные с заболеваниями. Также для нее характерны сложные механизмы эпигенетической регуляции [5], поэтому многие закономерности, исходно выявленные для этого удобного с точки зрения содержания и скорости размножения вида, оказываются справедливыми и для млекопитающих. С другой стороны, дрозофила остается самодостаточным объектом исследований, так как обладает рядом характерных только для нее особенностей, которые наблюдаются также и в структуре хроматина [6]. Например, пространственная организация ТАДов дрозофилы, базовых единиц трехмерной

¹ Здесь и далее под дрозофилой следует понимать *Drosophila melanogaster*

организации хроматина у многих видов, по-видимому, имеет мало общего с описанной для млекопитающих моделью экструзии, которая, в комплексе с другими составляющими, дает достаточно сложную для изучения картину [7]. С этой точки зрения хроматин дрозофилы является упрощенной, но все еще не тривиальной системой для исследования и последующего распространения выявленных закономерностей на организмы с более сложной организацией. Кроме того, этот организм часто используют для выведения мутантных линий.

Текущие исследования указывают на многофакторную природу динамики трехмерной структуры хроматина, опосредованную в том числе взаимодействием с разнообразными белками-ремоделлерами и архитектурными белками, к которым относятся не только инсуляторные белки, предназначенные скорее для решения локальных задач, но и участники крупных комплексов периферии ядра [8–10]. Интересно, что роль этих белков в организации структуры хроматина, по сравнению с белками-инсуляторами, освещена лишь в небольшом числе работ. Так, при исследовании взаимосвязи между упаковкой ДНК и ядерной ламиной было показано участие последней в поддержании пространственной сегрегации активного и неактивного хроматина в клетках S2 дрозофилы. Т. е. ламина выступает отчасти в роли поддерживающего структуру хроматина периферического каркаса. Еще одним важным компонентом оболочки ядра является комплекс ядерной поры. Среди большого разнообразия нуклеопоринов особенное место занимает Elys. Это единственный белок, для которого была показана способность напрямую связываться с ДНК и гистонами [11,12]. Помимо вхождения в состав комплекса ядерной поры, Elys также ассоциирован со слабо упакованными участками активного хроматина в нуклеоплазме [13]. Перечисленные свойства нуклеопорина Elys могут указывать на наличие взаимосвязи между его функционированием, локализацией в различных областях ядра и структурой хроматина, однако характер этой взаимосвязи до сих пор оставался за пределами исследований пространственной упаковки ДНК, в т. ч. методами Hi-C. Вероятно, такая ситуация объясняется наличием у Elys критически важных функций: выключение этого белка, как правило, приводит к потере

жизнеспособности клетки. Однако для клеток S2 дрозофилы это правило не выполняется [14]. Такое положение вещей может указывать на наличие определенных факторов, компенсирующих отсутствие этого нуклеопорина. В то же время, поскольку ряд функций, выполняемых им в разных организмах, является универсальным, можно предполагать, что в линии S2 Elys все-таки вовлечен в большинство процессов, которые протекают с его участием и в других типах клеток. В таком случае, применение метода Hi-C к клеточной линии S2 после деплеции Elys дает возможность проследить влияние этого белка на распределение и плотность упаковки хроматина внутри ядра.

Уже в ранних работах по анализу данных Hi-C была установлена строгая корреляция структуры хроматина с эпигенетическим состоянием, а оно свою очередь напрямую ассоциировано с экспрессией [8,15]. Учитывая место, которое занимает согласованная и точная работа генов в функционировании организма, исследование взаимного влияния пространственной организации хроматина и транскрипции вызывает особенный интерес. Оптимальной моделью для такого исследования может служить любая система, в которой наблюдается выраженное изменение активности генов. Например, для этой цели подходят зрелые клетки разного типа или разные стадии дифференцировки, сопровождающиеся запуском транскрипции тканеспецифичных генов. Наиболее удобным ресурсом для изучения хроматина на разных стадиях развития являются клеточные линии. Однако полученные таким способом клетки существенно отличаются по эпигенетическим и функциональным характеристикам от клеток, выделенных из биологических образцов, и могут служить лишь некоторым приближением в отсутствие последних [16,17].

Среди большого разнообразия траекторий развития, с точки зрения количества активируемых специфичных генов, особенно выделяется сперматогенез [18]. Так, в ходе сперматогенеза на завершающей стадии роста клеток, предшествующей мейозу, начинается активация не менее тысячи сперматоцитспецифичных генов. Основная функциональная составляющая этого процесса связана с формированием гаплоидных половых клеток, полученных из

диплоидного набора в ходе мейоза. В отличие от большинства организмов, в половых клетках самцов дрозофилы не происходит гомологичной рекомбинации и характерных для нее изменений, которые могут затрагивать структуру хроматина уже на стадии, предшествующей профазе первого мейоза, что делает этот объект исследования более прозрачным с точки зрения исследования аспектов, не связанных с рекомбинацией [19,20]. Благодаря замечательным свойствам дрозофилы, изложенным выше, становится возможным получение данных Hi-C для клеток семенников, соответствующих разным стадиям сперматогенеза: мутантные по гену *bam* (*bag-of-marbles*) особи служат источником клеток ранней стадии сперматогенеза – сперматогониев, а клетки, выделенные из семенников личинок третьей стадии, могут рассматриваться в качестве сперматоцитов, поскольку обладают значительным сходством с последними. Таким образом, клетки ранних стадий сперматогенеза, сперматогонии и сперматоциты, являются подходящим объектом для изучения взаимосвязи между изменениями в трехмерной организации хроматина и транскрипцией, причем в контексте ее активации. Учитывая, что большая часть исследований такого рода основывается на экспериментах с подавлением транскрипции, такая постановка задачи потенциально может привести к дополняющим сложившуюся картину результатам.

Следует отметить, что процесс сперматогенеза в дрозофиле интересен не только с точки зрения активации сперматоцитспецифичных генов: для него характерно многократное увеличение объема ядра клетки на стадии сперматоцитов [19]. Каким образом этот процесс сказывается на структуре хроматина до конца не известно. Исследования такого рода ограничиваются лишь несколькими публикациями, в том числе с искусственным раздуванием ядра в клетках млекопитающих [21], поэтому данный аспект заслуживает отдельного внимания при изучении особенностей организации хроматина в сперматогенезе плодовой мушки.

Обобщая все вышеизложенное, можно сказать, что анализ данных Hi-C начальных стадий сперматогенеза дрозофилы дает возможность изучения сразу

нескольких факторов в контексте их взаимосвязи с трехмерной организацией хроматина. Нуклеопорин Elys, будучи важной составляющей комплекса ядерной поры, также является потенциальным кандидатом на роль белка, напрямую участвующего в поддержании структуры хроматина. Однако работы в этой области, во-первых, немногочисленны, а во-вторых, не содержат информации о парных взаимодействиях и потому не реализуют возможности исследования хроматина на разных уровнях организации в масштабах всего генома.

С учетом видовых особенностей, подобное исследование, охватывая разные по природе явления, помогло бы выстроить более полную картину закономерностей организации хроматина у дрозофилы.

Целью данной работы является исследование влияния нуклеопорина Elys на структуру хроматина дрозофилы и ее взаимосвязь с различными факторами, ассоциированными с ранними стадиями сперматогенеза, на основе данных Hi-C.

Анализ проведенных исследований по данной тематике показывает, что при наличии карт хроматиновых взаимодействий для ранних стадий сперматогенеза дрозофилы, а также данных Hi-C при деплеции Elys актуальным является решение перечисленных ниже задач:

- 1) выявление взаимосвязи между активацией сперматоцитспецифичных генов и реорганизацией структуры хроматина;
- 2) анализ структуры хроматина в контексте характерных для сперматоцитов цитологических изменений;
- 3) анализ изменений в организации активного и неактивного хроматина при деплеции Elys.

Методологические основы исследования включают общие принципы анализа структуры хроматина с учетом ее специфики в дрозофиле. Все используемые методы можно разделить на стандартные и разработанные специально для применения к данным Hi-C или данным высокопроизводительного секвенирования РНК. К первой категории относятся рутинные методы биоинформатического анализа, нацеленные на оценку качества и предобработку, ко второй – широкий спектр подходов для решения различных задач от генерации

карт контактов из сырых чтений до поиска структур, характерных для организации хроматина на определенном уровне. Сюда можно отнести алгоритмы кластеризации экспериментальных повторностей, поиска ГАДов, компарментов, петель, оценки локальной плотности хроматина, оценки силы границ, картирования чтений на транскриптом дрозофилы и др. Кроме того, методическая часть исследования была дополнена определением оптимальных параметров запуска алгоритмов и реализацией описанного ранее подхода (с небольшими модификациями) к выявлению значимых контактов для заданного списка геномных координат. Применялись стандартные статистические тесты для оценки значимости отличий, в т. ч. на основе пермутаций. В ходе анализа активно использовались разнообразные функции, встроенные в научные пакеты `numpy` и `scipy`, с помощью которых, например, проводилось сглаживание кривых либо оценка угла наклона аппроксимирующей прямой. Поскольку существующие готовые программы удовлетворяют далеко не всем потребностям, возникающим в ходе анализа, для решения определенных подзадач было написано множество специальных функций.

Опытно-экспериментальная база была обеспечена сотрудниками лабораторий С. В. Разина и Ю. Я. Шевелева. Для анализа изменений в архитектуре генома на ранних стадиях сперматогенеза дрозофилы были использованы гомозиготные мутантные семенники `bamΔ86` из имаго, а также семенники дикого типа из личинок третьей стадии. Данные Hi-C при деплеции `Elys` были получены коллегами для клеточной линии S2 методом РНК-интерференции.

К теоретическим основам исследования стоит в первую очередь отнести значительный объем литературы, освещающей закономерности изменений структуры хроматина в различных условиях и способы их детекции с учетом видовой специфики. Анализ работ по структуре хроматина в дрозофиле, проведенных в последние несколько лет, позволил получить достаточно полную картину, отражающую степень целостности и глубины сложившейся парадигмы в отношении принципов упаковки ДНК у данного вида. Особенно в этой области выделяются немногочисленные, но очень содержательные работы Д. Роули [22,23].

Существенный вклад также дают проведенные ранее исследования в области функциональных особенностей нуклеопоринов, в том числе с точки зрения взаимосвязи со структурой хроматина, в особенности работы, проведенные на дрозофиле с участием М. Капельсон и В. Калверды [11,12].

Научная новизна данной работы во многом обусловлена выбором объекта исследования, поскольку анализ структуры хроматина на каких-либо стадиях сперматогенеза в дрозофиле ранее не проводился. Например, значительный интерес представляет экспрессия сперматоцитспецифичных генов, т. к. информация о влиянии запуска транскрипции большого числа генов на уже сформировавшуюся структуру хроматина дрозофилы в хоть и специфичных, но все же относительно естественных условиях, не представлена в литературе по данной тематике. Впервые были описаны изменения в организации хроматина при многократном увеличении объема ядра клетки у данного вида. Оригинальность представленных результатов анализа данных Hi-C при деплеции нуклеопорина Elys объясняется несколькими факторами, касающимися выбора как объекта, так и метода исследования, и не ограничивается областью видовой специфики. Это первая работа, в которой, благодаря удачному выбору модельного организма и конкретной клеточной линии, удалось получить данные по хроматину при деплеции Elys, а применение Hi-C позволило проанализировать его влияния на разных уровнях организации.

Практическая и теоретическая значимость проведенного исследования включает ряд методических наблюдений, полезных для дальнейших исследований организации хроматина в дрозофиле. Например, в ходе анализа был оптимизирован подход к детекции ТАДов. Выявленные закономерности позволяют заполнить некоторые пробелы в общей картине многообразия факторов, влияющих на структуру хроматина дрозофилы в целом и на ранних стадиях сперматогенеза в частности. Также, в ходе теоретической и практической работы, были обнаружены предпосылки для дальнейших исследований как в области методов анализа конформации хроматина, так и в направлении более глубокой биологической интерпретации полученных результатов. Кроме того, знания, которые были

получены при анализе изменений, вызванных деплецией Elys, могут служить полезным референсным ресурсом при изучении состояний, характеризующихся нарушением метаболизма компонентов комплекса ядерной поры.

Положения, выносимые на защиту:

1. Активация сперматоцитспецифичных генов сопровождается повышением частоты их контактов, однако такой вид взаимодействий не является избирательным: так же часто гены данной группы формируют контакты с генами, активными в различных тканях дрозофилы. Функциональное сходство этой группы генов ассоциировано с линейной, но не пространственной колокализацией.
2. Значительная доля сперматоцитспецифичных генов приходится на неактивный хроматин. При активации их транскрипции имеет место локальная декомпактизация плотности окружения промоторов, но не формирование новых, ярко выраженных границ.
3. На ранних стадиях сперматогенеза в хроматине дрозофилы обнаруживаются изменения, не связанные с экспрессией и, вероятно, обусловленные изменением объема ядра клетки. Это наблюдение позволяет рассматривать данный процесс в качестве естественного эксперимента по увеличению объема ядра и учитывать его в будущих исследованиях по данной тематике.
4. В целом, при переходе от стадии сперматогониев к стадии сперматоцитов происходит связанное с активацией сперматоцитспецифичных генов изменение конформации хроматина на нескольких уровнях организации. Масштаб этих изменений зависит от исходного положения участка генома в более или менее активной области хроматина.
5. Деплеция нуклеопорина Elys также приводит к заметным изменениям на разных уровнях пространственной организации и затрагивает как активную, так и неактивную фракцию хроматина. Нуклеопорин Elys, наряду с белками ламины, является важным компонентом системы регуляции плотности упаковки ДНК и ее распределения в пространстве ядра.

Апробация работы

Основные результаты исследования были представлены на трех конференциях:

1. ContactHunter: поиск значимых взаимодействий для заданных геномных локусов на основе данных Hi-C. Proceedings of 11th Moscow Conference on Computational Molecular Biology MCCMB'23.
2. Влияние изменения уровня ацетилирования на структуру хроматина в клеточной линии дрозофилы. Роль нуклеопорина Элис в поддержании структуры хроматина. Сборник трудов 47-й междисциплинарной школы-конференции ИППИ РАН. 2023
3. Chromatin structure changes during drosophila spermatogenesis. Сборник трудов 43-й междисциплинарной школы-конференции ИППИ РАН. 2019

Публикации по теме диссертации

Основные результаты по теме диссертации изложены в двух статьях в рецензируемых журналах, индексируемых Scopus и/или Web of Science:

1. Doronin, S. A.*, Ilyin, A. A.*, Kononkova, A. D.*, Solovyev, M. A.*, Olenkina, O. M., Nenasheva, V. V., Mikhaleva, E. A., Lavrov, S. A., Ivannikova, A. Y., Simonov, R. A., Fedotova, A. A., Khrameeva, E. E., Ulianov, S. V., Razin, S. V., & Shevelyov, Y. Y.* (2024). Nucleoporin Elys attaches peripheral chromatin to the nuclear pores in interphase nuclei. *Communications biology*, 7(1), 783.
2. Ilyin, A. A.*, Kononkova, A. D.*, Golova, A. V.*, Shloma, V. V.*, Olenkina, O. M.*, Nenasheva, V. V., Abramov, Y. A., Kotov, A. A., Maksimov, D. A., Laktionov, P. P., Pindyurin, A. V., Galitsyna, A. A., Ulianov, S. V., Khrameeva, E. E., Gelfand, M. S., Belyakin, S. N., Razin, S. V., & Shevelyov, Y. Y. (2022). Comparison of genome architecture at two stages of male germline cell differentiation in *Drosophila*. *Nucleic acids research*, 50(6), 3203–3225.

* - равный вклад авторов

Личный вклад

Автор принимал непосредственное участие в выборе стратегий исследований, планировании анализа данных и его выполнении, интерпретации и обобщении полученных результатов. Все результаты анализа данных, описанные в диссертационной работе, получены автором лично.

Объем и структура работы

Диссертация состоит из введения, трех глав, заключения, выводов и приложения. Полный объем диссертации составляет 109 страниц с 33 рисунками и 5 таблицами. Список литературы содержит 125 наименований.

Глава 1. Обзор литературы

1.1. Трехмерная организация хроматина: методы детекции и основные структурные характеристики

1.1.1 Общие представления о методах детекции структуры хроматина

Теоретические основы представлений о существовании, роли и структуре хроматина были заложены еще в первой половине XIX в. и стали закономерным этапом развития клеточной биологии, несколько позже, в 1873 г., Фридрих Шнейдер при изучении клеточного деления обнаружил формирование нитчатых структур, которые агрегируются в центре клетки и формируют метафазную пластинку. Этот процесс был более подробно описан Вальтером Флеммингом. Уже в ранних работах в данной области основными инструментами были микроскоп и необходимые для визуализации специальные химические агенты, эксперименты с которыми и позволили Флеммингу обнаружить нитчатое образование, содержащееся внутри ядра клетки. Это образование поглощало определенные красители, и ученый назвал его хроматином ('chroma' означает «цветность» в переводе с латинского). В 1869 г. Фридрих Мишер открыл ДНК: он назвал фосфоросодержащую молекулярную субстанцию нуклеином. Позже химический анализ показал, что нуклеин состоит из органической кислоты и щелочного белка. Кислоту назвали нуклеиновой кислотой, а белок – гистоном (гистон впервые обнаружен в ядре птичьего эритроцита Альбрехтом Косселем в 1884 г) [24]. Флемминг был одним из первых ученых, предположивших, что нуклеин и хроматин – это одно и то же.

Удивительно, но во многом сегодняшний подход к изучению структуры хроматина не отличается концептуально от методов позапрошлого столетия: световой микроскоп заменен на конфокальный электронный, а на место анилиновых красителей, полученных Флеммингом, пришли разнообразные флюоресцирующие, способные связываться со специфическими мишенями, агенты. Тем не менее в этой области наблюдаются значительные успехи. Например, при помощи объединения информации о точной локализации каждого флуорофора в образце можно реконструировать флуоресцентное изображение со сверхвысоким

разрешением (микроскопия стохастической оптической реконструкции, STORM) [25]. По-прежнему активно применяется открытый еще в прошлом столетии, но значительно улучшенный метод флюоресцентной *in situ* гибридизации, FISH, основанный на формировании гибридной молекулы между одноцепочечной ДНК в клетке и комплементарным искусственным зондом. Использование микроскопии высокого разрешения в сочетании с флюоресцирующими пробами и RNA-FISH (метод ORCA), позволяет одновременно определить тип клеток и трехмерную организацию хроматина для отдельных локусов [26]. Метод Hi-M дает информацию не только о структуре хроматина в масштабах нескольких сотен тыс.п.о, но и о транскрипции [27]. Отличительной чертой метода 3D-SIM является усовершенствование технических аспектов, связанных непосредственно с визуализацией, а не с подбором проб, что также повышает качество исследуемых данных [28]. Использование современной микроскопии в сочетании со специальными метками формирует очень важную методическую основу, без которой невозможно было бы достичь актуального уровня понимания закономерностей трехмерной организации, однако оно не дает информации о попарных взаимодействиях в масштабах всего генома [29].

Между тем, во многом благодаря повсеместному распространению методов секвенирования нового поколения, в начале 2000-х гг. в области подходов к исследованию структуры хроматина было открыто еще одно направление.

1.1.2 Методы детекции структуры хроматина в эпоху секвенирования

В 2002 г. был впервые опробован новый метод исследования структуры хроматина – «захват конформации хроматина» или 3C (от chromatin conformation capture), который был разработан в лаборатории Йоба Деккера [30]. В основе метода лежала фрагментация участков ДНК с последующим лигированием, что приводило к образованию химерной молекулы ДНК, состоящей из двух фрагментов генома, расположенных близко в трехмерном пространстве. Количество взаимодействий двух участков генома подсчитывалось с помощью метода ПЦР. Недостатком этого метода было то, что с его помощью можно было определять частоты контактов только заранее заданных геномных локусов. Однако

именно этот вид анализа стал краеугольным камнем в основании методов захвата конформации хроматина. В 2005 г. появились методы 4C и 5C, которые дали возможность исследовать частоты контактов различных пар геномных локусов. Настоящий прорыв в этой области случился в 2009 г., когда Либерман Эйден и Йоб Деккер разработали метод Hi-C [1]. Данный метод не ограничивался определением частот контактов некоторых пар геномных локусов, а позволял исследовать хроматиновые взаимодействия в масштабах всего генома. Благодаря этому методу было подтверждено, что структура укладки хроматина иерархична и взаимодействия в геноме происходят не случайным образом, а являются результатом слаженного механизма регуляции. Выяснилось, что во многом основные паттерны структуры хроматина схожи у разных организмов, а тонкая регуляция работы генов часто осуществляется за счет контактов определенных локусов [23]. Дальнейшее усовершенствование метода шло по пути увеличения разрешения и более деликатной пробоподготовки.

Несмотря на большие успехи в понимании структуры хроматина, обусловленные распространением метода Hi-C и его производных, он все же имеет ряд недостатков. А именно, данный подход разработан для большого количества ядер, т. е. его применение дает представление о структуре хроматина в популяции клеток, тогда как каждая отдельная клетка может иметь свою неповторимую укладку хроматина. Также этот метод основан на сшивании концов, лигировании, а это значит, что если по каким-то причинам концы фрагментов генома не могут быть сшиты (например, между ними находится крупный белок), то их контакт не будет установлен. Кроме того, методы, основанные на лигировании, по комбинаторным причинам плохо определяют частоты множественных контактов. Первая проблема была частично преодолена благодаря разработке протокола Hi-C для одиночных клеток [31]. Для решения второй были разработаны новые методы, такие как GAM (Genome Architecture Mapping) и SPRITE (Split-Pool Recognition of Interactions by Tag Extension) [32,33]. В основе GAM лежит разделение ядер на тонкие срезы в случайном направлении и предположение о том, что контактирующие участки генома должны чаще попадать в один и тот же срез.

Метод SPRITE построен на стратегии генерации уникальных последовательностей для каждого набора взаимодействующих друг с другом участков генома таким образом, что контактирующие участки генома имеют на выходе одинаковый идентификатор. Важно отметить, что перечисленные методы также подтверждают наличие характерных для укладки хроматина структур, определенных ранее методом Hi-C. Однако возможность исследования множественных контактов и контактов участков через белок в отдельных клетках не компенсирует недостатков, обусловленных скромной библиотекой биоинформатических пакетов и достаточно низким разрешением. Успешность применения подхода Hi-C способствовала его всевозможным модификациям, например, развитию метода определения взаимодействий между ДНК и РНК, таких как протокол RedC [34].

1.1.3 Принцип работы метода Hi-C. Первичная обработка данных секвенирования

Протокол Hi-C включает несколько основных этапов. За время, прошедшее с момента первой публикации, некоторые из них были модифицированы, однако суть метода остается прежней. Современный вариант протокола, *in situ* Hi-C, был предложен в 2014 г. Рао и соавторами [35]. На первой стадии подготовки библиотеки проводится фиксация клеток формальдегидом для образования связей между участками ДНК, расположенными близко в пространстве (рис. 1). Затем клетки подвергаются лизированию и механическому расщеплению, чтобы обеспечить рестриктазам доступ к фиксированному хроматину. Фермент рестрикции добавляется для расщепления хроматина на множество фрагментов ДНК. Концы полученных фрагментов заполняют биотиновыми метками. Затем фрагменты лигируют, а ДНК очищают и разрезают. Биотинилированные фрагменты вытягивают из раствора с помощью гранул стрептавидина, затем удаляют биотиновую метку, проводят отбор фрагментов по длине и увеличивают количество материала с помощью ПЦР, после чего приготовленную библиотеку отправляют на секвенирование.

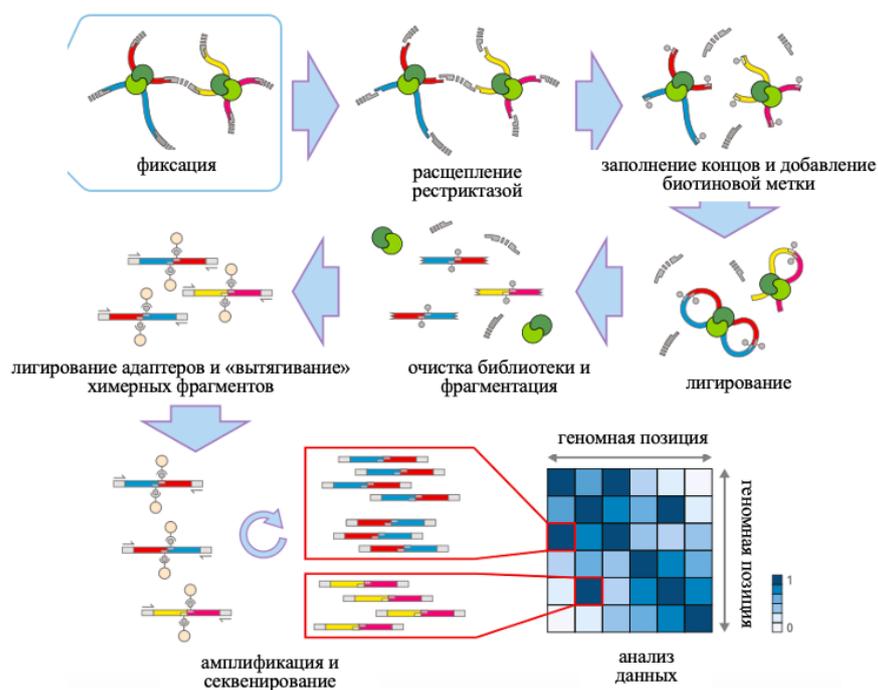


Рисунок 1. Схематическое изображение основных этапов протокола Hi-C (адаптировано из *Cremazy et al, 2018*).

Основное отличие *in situ* Hi-C от первоначально предложенной версии протокола состоит в том, что лигирование выполняется внутри ядра, в ограниченном пространстве, а не в растворе, где фрагменты ДНК свободно плавают, как это было ранее. Таким образом снижается частота случайного лигирования. Существенным фактором, влияющим на качество данных Hi-C, является выбор рестриктазы. Чем чаще в последовательности ДНК встречаются сайты рестрикции, а они являются специфичными для разных ферментов, тем лучше будет разрешение карты хроматиновых взаимодействий. Используемая ранее *HindIII* разрезала ДНК по 6-нуклеотидной последовательности, современный протокол чаще включает фермент *DpnII*, расщепляющий ДНК по 4 нуклеотидам, что существенно улучшает качество данных Hi-C. Тем не менее зависимость итогового разрешения от частоты встречаемости сайтов рестрикции в последовательности ДНК побудила ученых к поиску альтернативных методов фрагментации и привела к созданию протокола Micro-C, в основе которого лежит использование микрококковой нуклеазы, разрезающей ДНК в местах, свободных от нуклеосом, что обеспечивает разрешение в 200 п. о. [36]. Однако, несмотря на

более низкое разрешение данных Hi-C, этот метод продолжает широко использоваться в силу хорошо описанных и отработанных протоколов. В 2021 году была предложена новая версия Hi-C.3.0 с модификацией фиксирующего агента и использованием двух рестриктаз, однако пока что эти изменения не стали частью стандартного протокола [37]. Также несколькими годами ранее, в 2018 г., был предложен 6-часовой протокол Arima Hi-C, включающий использование сразу нескольких рестриктаз. Карты, полученные этим способом, уже появляются в публичном доступе и отличаются высоким качеством и разрешением.

После секвенирования библиотек начинается биоинформатический анализ данных Hi-C. Подготовленная в ходе протокола библиотека состоит из большого количества химерных молекул, сшитых из двух или более участков ДНК, которые находились близко друг к другу в трехмерном пространстве. Последовательности, полученные после парного секвенирования концов такой химерной молекулы, соответствуют участкам ДНК, вовлеченным во взаимодействие. Каждая последовательность при выравнивании на геном дает возможность идентифицировать один из двух (иногда более) контактирующих участков, в то же время сохраняется информация о другом участнике хроматинового взаимодействия, поскольку последовательность является частью парного чтения.

1.1.4 Основные характеристики трехмерной структуры хроматина, выявленные методом Hi-C

Полученные после протокола Hi-C карты контактов содержат информацию о взаимодействиях на всевозможных расстояниях, тем самым позволяя исследовать особенности структуры хроматина на разных масштабах. Помимо таких универсальных признаков, как иерархичность и уменьшение частот контактов по мере увеличения расстояния между взаимодействующими локусами, первичный визуальный анализ карт контактов показал наличие участков генома, контактирующих друг с другом чаще, чем это ожидалось бы для случайной полимерной конформации хромосомы. Такие взаимодействия могут быть описаны в терминах определенных структур.

Доминирующие внутрихромосомные взаимодействия полногеномной карты контактов млекопитающих и дрозофилы соответствует сложившимся представлениям о хромосомных территориях как самом масштабном уровне организации: еще на заре становления хроматиновой науки Рабль и Бовери косвенно показали, что конденсированные митотические хромосомы переходят в интерфазные, занимая при этом некоторое пространство в ядре и не сильно смешиваясь с остальными (рис. 2а,б) [38]. Однако уже в этих масштабах наблюдается видовое разнообразие. Не у всех многоклеточных обнаруживаются хромосомные территории: часть из них характеризуется наличием так называемой Рабль-конфигурации (рис. 2в,г) [39]. Эта структура формируется в результате взаимодействия теломерных и/или центромерных участков генома друг с другом и с разными полюсами ядра, приводя к очень яркому и специфичному паттерну на карте Hi-C.

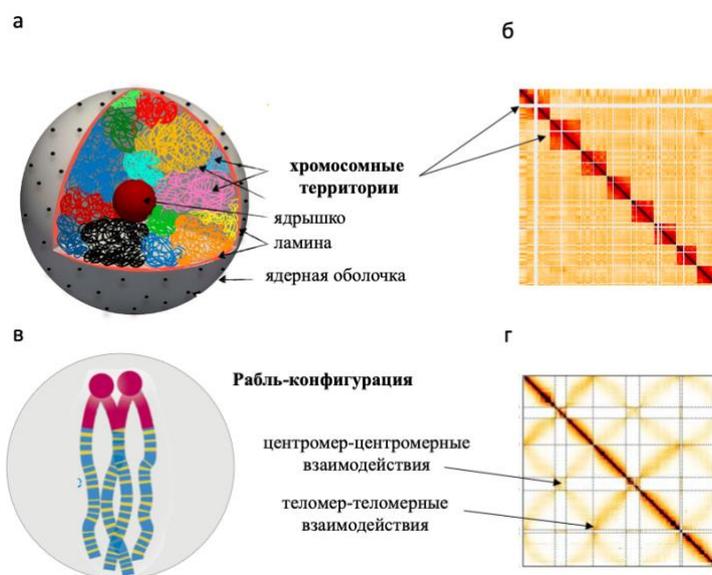


Рисунок 2. Межхромосомные взаимодействия.

(а) Схематичное изображение интерфазных хромосом внутри ядра (адаптировано из *Laghtach et al 2022*). (б) Фрагмент Hi-C карты млекопитающих в полногеномном масштабе. (в) Схематичное изображение Рабль-конфигурации для трехмерной структуры хроматина на ранних стадиях эмбриогенеза дрозофилы. (г) Hi-C карта в полногеномном масштабе, полученная для ранних стадий эмбриогенеза дрозофилы ((в,г) адаптировано из *Zenk, Zhan, Kos et al., 2021*).

Наиболее крупномасштабные формы взаимодействий на уровне отдельной хромосомы известны как геномные компартменты [1]. Они проявляются в матрице Hi-C в виде "шахматной доски", состоящей из чередующихся блоков размером ~100-1000 тыс.п.о. (в геноме человека) с высокой и низкой частотой взаимодействия (рис. 3). Наблюдаемая картина объясняется простым принципом, согласно которому хромосомы состоят из двух типов последовательно чередующихся геномных регионов, и частоты контактов между двумя регионами одного типа оказываются в среднем выше, чем между регионами разных типов. Эти два типа регионов называют А- и В-компартаментами. Хотя такая схема взаимодействий является интуитивно понятной, определение компарментов с точки зрения их детекции не является строго детерминированным. Понятие компарментов тесно связано с определениями эухроматина и гетерохроматина: А-компаратмент определяются как эухроматические генные области, а В-компаратмент - как бедные генами гетерохроматические области. Компаратменты обладают высокой пластичностью, т.е. изменяются в различных типах клеток и биологических условиях, что соответствует крупномасштабным изменениям активности генов [40]. Длина отдельных блоков компарментов обычно составляет порядка 0.1-10 Мб для млекопитающих, поэтому их легко выделить даже в экспериментах с очень низким разрешением. Анализируя данные Hi-C с высоким разрешением клеточной линии лимфобластоида человека (GM12878), Рао и соавторы расширили определение А/В компарментов, разделив их на пять основных субкомпаратментов: А1, А2, В1, В2 и В3 [35]. Эти структурные субъединицы демонстрируют отчетливые и более тонкие ассоциации с различными геномными и эпигеномными характеристиками, к которым относятся экспрессия генов, активные и репрессивные гистоновые метки, время репликации ДНК и специфические субъядерные структуры.

На внутривхромосомной Hi-C карте многих организмов имеются расположенные вдоль диагонали и обогащенные контактами квадраты (рис. 3). Эти

структуры называются топологически ассоциированными доменами или ТАДами. Они обладают двумя особенностями: во-первых, повышенное число контактов предполагает, что участки внутри ТАДов контактируют друг с другом чаще, чем участки из разных ТАДов, т. е. обладают свойством самоассоциированности. Во-вторых, ТАДы являются иерархичными, т. е. также могут включать в себя домены меньшего размера. ТАДы были обнаружены почти одновременно в мухах, мышах и человеке, что указывает на универсальность принципов укладки хромосом [15,41,42]. Подобно компартментам и субкомпартментам аннотация ТАДов зависит от качества образцов и глубины секвенирования. По мере улучшения разрешения становятся различимыми новые границы ТАДов и новые, более мелкие, ТАДы.

К характерным паттернам организации хроматина многих организмов также относятся петли (рис. 3). Петли – это хорошо различимые на карте Hi-C взаимодействия двух участков генома, разделенных некоторым расстоянием. Основания петель могут приходиться как на границы ТАДов, так и на внутреннюю их часть. Они часто обогащены архитектурными белками и бывают ассоциированы как с активацией, так и с подавлением транскрипции [35,43–45]. Еще один срез информации представлен в карте Hi-C куда менее заметными отдельными точечными контактами, в основе которых часто лежат энхансерно-промоторные взаимодействия или взаимодействия промоторов генов, регулируемых одним транскрипционным фактором [46,47].

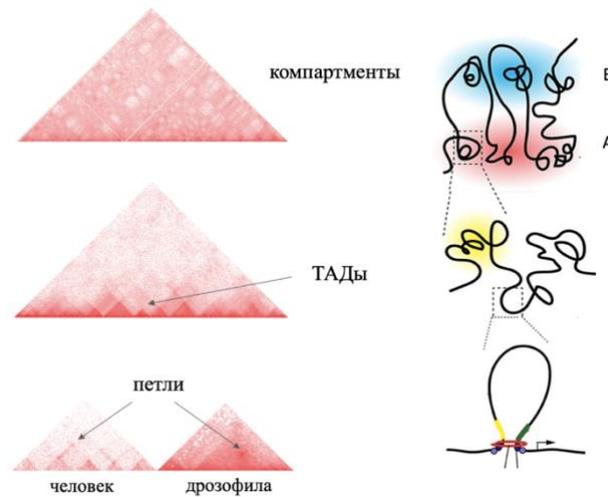


Рисунок 3. Основные уровни внутривхромосомной организации интерфазного хроматина млекопитающих и дрозофила (графика адаптирована из *Fujita, Yamashita, 2021*).

Существуют и более экзотические формы организации хроматина, например, расположенные вдоль диагонали «фонтаны» у рыб [48]. Наличие тех или иных особенностей укладки хроматина может варьироваться между видами, а приписанные им функции регулярно подлежат пересмотру и являются предметом интенсивных исследований последних лет.

1.2. Особенности организации хроматина дрозофила в контексте ключевых опосредующих факторов

1.2.1 Межхромосомные взаимодействия

Как уже было упомянуто ранее, хроматин дрозофила, как и хроматин млекопитающих, организован в пространстве ядра неслучайным образом и самые крупномасштабные закономерности его упаковки проявляются на уровне межхромосомных территорий. Т. е. внутривхромосомные, в особенности внутривхромосомные, взаимодействия выражены сильнее, чем межхромосомные [49]. Однако на стадии развития, соответствующей зиготической геномной активации (ЗГА), это правило нарушается для центромерных и теломерных областей, которые у плодовой мушки выделяются в отдельный компартмент. Такая структура соответствует Рабль-конфигурации. Одним из факторов, влияющих на ее образование, является ассоциированный с гетерохроматином белок HP1, т. к. при

его деплеции было обнаружено ослабление взаимодействий, характерных для Рабля [50]. Кроме того, описывая межхромосомные взаимодействия дрозофилы, следует упомянуть, что спаривание гомологичных хромосом наблюдается у этого организма не только в мейозе, но и в соматических клетках [51]. Соответственно, регуляция может происходить по механизму трансекции, в том числе, цис-регуляторные элементы могут контролировать экспрессию генов, расположенных на гомологичной хромосоме. Анализ данных Hi-C, полученных с применением специального протокола для исследования структуры хроматина на уровне гомологичных хромосом, показал, что взаимодействия между гомологами оказываются интенсивнее в небольших активных ТАДах, по сравнению с крупными неактивными, а наиболее выраженное спаривание имеет место в границах [22,52,53].

1.2.2 Внутрихромосомные взаимодействия

1.2.2.1 ТАДы, компартменты

На уровне внутрихромосомных взаимодействий трехмерная организация хроматина дрозофилы во многом схожа с таковой у млекопитающих. Вместе с тем наблюдаются серьезные отличия в механизме формирования, а также степени выраженности некоторых структур.

Одно из фундаментальных отличий между дрозофилой и млекопитающими связано с механизмом формирования ТАДов. Согласно многочисленным исследованиям, важным, хотя и не единственным фактором, лежащим в основе формирования ТАДов млекопитающих, является петлевая экструзия [43]. В упрощенном виде она представляет собой выпетливание хроматина через комплекс из субъединиц когезина, напоминающий по форме кольцо. Данный процесс является АТФ-зависимым и регулируется несколькими белками, определяющими время ассоциации когезинового комплекса с ДНК, например, такими как NIPBL и WAPL (отвечает за загрузку и разгрузку когезина соответственно). Экструзия петель и ее блокирование CCCTC-связывающим фактором (CTCF) рассматриваются в качестве основной причины повышения частоты контактов внутри, по краям и в углах ТАДов млекопитающих. Предполагается, что эти

паттерны - не статичные, воспроизводимые структуры генома, которые можно увидеть под микроскопом, а скорее среднее по популяции клеток, сформированное ансамблем динамически растущих петель, ограниченных конвергентными мотивами связывания CTCF [54]. У дрозофилы подобные свойства организации ТАДов не обнаруживаются. Внутри отдельно взятого ТАДа этого вида наблюдается более равномерное распределение частот контактов. При этом часть ТАДов характеризуется наличием усиленного взаимодействия фланкирующих доменов, что проявляется в виде структуры, похожей на «корону» (рис. 4). И хотя некоторые ТАДы у данного вида имеют определенное сходство с ТАДами млекопитающими, т. е. включают апикальные точечные взаимодействия, подобные случаи обусловлены скорее не механизмом экстррузии, а другими типами контактов, как правило, репрессивного характера [7].

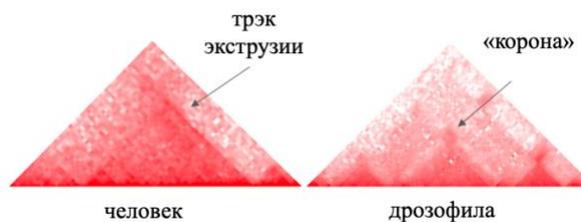


Рисунок 4. Особенности организации хроматина на уровне ТАДов у млекопитающих и дрозофилы.

Концепция отсутствия механизма экстррузии у дрозофилы поддержана рядом дополнительных наблюдений. Так, было показано, что несмотря на наличие у нее гомолога CTCF (dCTCF), границы ТАДов не обогащены конвергентными сайтами его связывания. Кроме того, в отличие от ситуации с млекопитающими, данные о колокализации dCTCF и когезина в дрозофиле противоречивы: в одном исследовании сообщается о незначительном перекрытии сайтов связывания dCTCF и когезиновой субъединицы SCC3, в то время как в другом исследовании было обнаружено, что почти половина dCTCF колокализуется с когезиновой субъединицей Rad21 [55,56]. Таким образом, существование ТАДов в дрозофиле может объясняться иным механизмом их формирования.

Уже на заре широкого распространения метода Hi-C Секстон и соавторы продемонстрировали ключевые особенности трехмерной организации хроматина у данного вида, в частности структуры ТАДов, в контексте взаимосвязи с эпигенетическими состояниями. Основываясь на разбиении генома на отдельные участки, характеризующиеся определенными, наиболее предпочтительными, комбинациями гистоновых меток, Филион, а позже Харченко и соавторы получили разметку состояний или «цветов» хроматина (5 и 9 состояний соответственно) [57,58]. На основе этих данных Секстон и соавторы обнаружили выраженную взаимосвязь между эпигенетическими профилями и разметкой доменов хроматина и кластеризовали их по среднему обогащению метками. В ходе анализа авторы пришли к заключению, что ТАДы являются ключевыми фундаментальными единицами упаковки хроматина дрозофилы, что протяженные участки неактивного хроматина сворачиваются в крупные ТАДы, перемежаясь с небольшими активными, а границы ТАДов, склонные к самоассоциации, обогащены метками активной транскрипции и обладают повышенной плотностью генов (рис. 5). Также было выдвинуто предположение о том, что высокая корреляция с эпигенетическими профилями, возможно, обуславливает механизм формирования ТАДов. Кроме того, анализ зависимости частоты контактов от расстояния показал, что скорость ее изменения отличается в активных и неактивных ТАДах, т. е. ТАДы этих двух видов могут формироваться за счет разных механизмов [15].

Позже на данных высокого разрешения было показано, что границы и межТАДы на самом деле представляют собой активные мини-ТАДы, образованные иногда одним геном. При этом основным фактором, определяющим выраженность мини-ТАДов, является присутствие РНК-полимеразы II (Pol II), но не сама по себе элонгация транскрипции. Роули и соавторы продемонстрировали строгую корреляцию между транскрипционной активностью, полученной с помощью GRO-seq, и данными Hi-C на разрешении 1 тыс.п.о. и пришли к выводу, что непосредственно границы ТАДов не столько определяются активной транскрипцией, сколько разграничивают активные и неактивные участки генома, а активные мини-ТАДы характеризуются повышенной частотой контактов друг с

другом в соответствии с наблюдаемой ранее картиной повышения частоты контактов границ ТАДов [22,23].

ТАДы дрозофилы тесно связаны со структурой компартментов. В работах 2012 г, независимо проведенных под руководством Кавалли и Корцеса, было обнаружено, что у дрозофилы, как и у млекопитающих, дальние контакты описываются структурой типа шахматной доски, образованной в результате предпочтительного взаимодействия хроматина определенного типа с себе подобным. Также авторы показали, что, в отличие от неактивных участков генома, активные формируют не только внутри-, но и межхромосомными контакты [15,41]. При описании структуры компартментов дрозофилы интересным кажется наблюдение, касающееся масштабов ее изменения: у генов, остановленных на стадии начала транскрипции, не перешедших к элонгации, к А-компартменту принадлежали только старты транскрипции, обогащенные Pol II, тогда как тело гена могло находиться в неактивном компартменте [22]. Высокая корреляция разметки ТАДов с эпигенетическим профилем косвенно указывала на то, что именно взаимодействие ТАДов со сходным состоянием лежит в основе формирования дальних компартментных взаимодействий у дрозофилы. Позже, на основе анализа данных высокого разрешения, разметка компартментов была получена методом главных компонент, который прежде применялся к данным по хроматину млекопитающих. Хорошее совпадение разметки компартментов и ТАДов поддержало выдвинутые ранее предположения о компартментной природе ТАДов и породило их альтернативное, характерное именно для дрозофилы, название «компартментные домены» [15,41]. В отличие от млекопитающих, наибольший вклад в компартментную структуру хроматина дрозофилы дают взаимодействия активных доменов, в то время как дальние В-В контакты выражены гораздо слабее [22].

В ходе дополнительного анализа взаимосвязи между состояниями хроматина и его трехмерной организацией Роули и соавторы получили также данные HiChIP для модификаций гистонов H3K27ac и H3K27me3 (рис. 5) [23]. Комбинация данных HiChIP для обеих меток показала 98.9%-ую корреляцию с исходной картой

Hi-C. Похожие результаты были также получены в работе [59]. Ведущая роль эпигенетических состояний хроматина в формировании компартментов также подтверждается данными по деплеции HP1a на ранних стадиях эмбриогенеза [50].

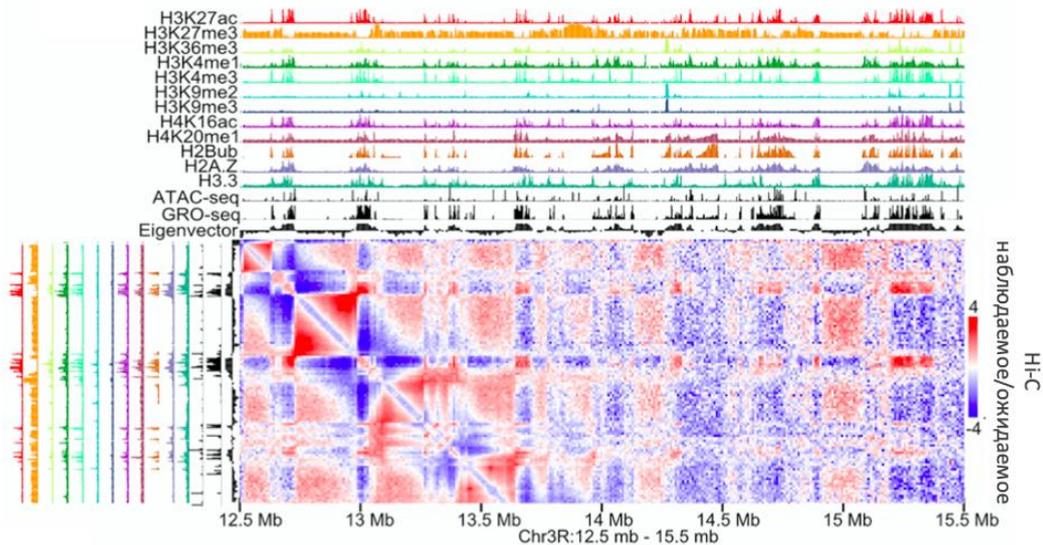


Рисунок 5. Взаимосвязь между организацией хроматина дрозофилы и профилями различных модификаций гистонов, открытости хроматина и экспрессии на примере клеточной линии Kc167 (адаптировано из *Rowley, 2017*).

В построении концепции механизма формирования ТАДов дрозофилы также активно применялись методы компьютерного моделирования. Так, в работе Ульянова и соавторов с помощью компьютерного моделирования было показано, что ТАДы могут представлять собой самоорганизующиеся динамичные структуры, сформированные многообразием переходных взаимодействий между нуклеосомами неактивного хроматина, в то время как межТАДы/границы образованы ацетилированным хроматином, не обладающим свойством самоассоциации [8]. При рассмотрении взаимодействий в масштабах среднего размера ТАДов такая интерпретация соотносится с результатами анализа данных высокого разрешения, полученными Ванг и соавторами, которые показали, что последовательно расположенные неактивные субТАДы хорошо взаимодействуют друг с другом в составе крупных суперТАДов, в то время как для активных такое поведение не характерно [60]. В исследованиях структуры хроматина последних лет активно обсуждается интерпретация полученных результатов в контексте теории разделения фаз. Согласно такому подходу, компартментная структура

хроматина может формироваться в результате различий в биохимических свойствах различных участков генома, что способствует их сегрегации в концентрированную жидкую и разбавленную фазы. Это разделение может быть обусловлено слабыми мультивалентными взаимодействиями между белками, РНК и ДНК [61–64].

Таким образом, существенная роль в организации хроматина дрозофилы принадлежит транскрипции и гистоновым модификациям. Одной из доминирующих является гипотеза о разделении состояний хроматина между доменами ацетилирования и метилирования гистонов (особенно H3K27me3), которое, предположительно, и приводит к формированию ТАДов. Кроме того, в отличие от млекопитающих, компартменты и ТАДы у этого вида находятся в тесной взаимосвязи друг с другом и, по-видимому, формируются под влиянием одних и тех же факторов.

1.2.2.2 ТАДы дрозофилы в исследованиях единичных клеток

Как уже было упомянуто ранее, ТАДы млекопитающих, предположительно, представимы в виде классических ярких треугольников лишь на популяционных картах Hi-C и в реальности могут являться средним по ансамблю петель либо ТАДов с вариативными границами (согласно данным STORM) [65]. Ряд исследований показывает, что, несмотря на существенные различия в механизме формирования, ТАДы дрозофилы, вероятно, также являются результатом композиции доменных структур, отличающихся от клетки к клетке. Так, согласно данным scHi-C, около 45% границ являются консервативными и совпадают с границами ТАДов, полученных по популяционным данным, однако для перемешанных случайным образом границ это значение составляет около 30%, что скорее указывает на высокую вариабельность, чем детерминированность границ в популяции клеток [66]. Кроме того, данные Hi-M, полученные для локуса с двумя последовательными ТАДами, указывают на то, что лишь в малой части ядер (~14%) наблюдается заметная инсуляция в регионе, соответствующем популяционной границе этих двух ТАДов. Также не было обнаружено ассоциации между транскрипцией и плотностью упаковки границ ТАДов в отдельных ядрах [67]. В то

же время, анализ данных, полученных комбинацией метода микроскопии 3D-SIM и флуоресцентной *in-situ* гибридизации со специальными пробами (Oligopaint), выявил наличие специфичных нанокompартментов, совпадающих с неактивными популяционными ТАДами репрессивного хроматина [68]. Еще в одной работе с применением той же методики было показано, что вероятность контактов внутри ТАДов значительно выше, чем между ТАДами [69]. Важно отметить, что оба этих метода (Hi-M и 3D-SIM) позволяют исследовать структуру хроматина в ограниченном масштабе. Неоднозначность результатов, полученных различными методами, указывает на то, что вопрос о существовании и стабильности ТАДов в отдельно взятой клетке дрозофилы по-прежнему остается открытым.

1.2.2.3 Архитектурные белки и транскрипция как факторы, влияющие на структуру хроматина

Помимо разделения взаимодействий участков генома, обогащенных разными эпигенетическими модификациями, на активный и неактивный компартменты, в организации трехмерной структуры хроматина дрозофилы не последнюю роль играет еще одна составляющая – архитектурные белки (АБ). Большая часть границ ТАДов у этого вида обогащена специфичными белками-инсуляторами класса I: BEAF32, CP190, Chromator [60]. На данных высокого разрешения показано, что архитектурными белками обогащены не сами границы, а малые домены активного хроматина, определяемые ранее как границы или межТАДы, причем это обогащение оказывается равномерным внутри домена [22,23].

Роль АБ в формировании структуры хроматина, в особенности на уровне границ ТАДов, в настоящее время является предметом активного изучения [70–72]. Анализ осложняется тем, что для дрозофилы характерна избыточность АБ по отношению к последовательности инсулятора. Например, BEAF32 и DREF имеют один и тот же мотив связывания. Кроме того, данные ChIP-seq порой оказываются недостаточно специфичными для конкретного АБ [70]. Границы ТАДов часто обогащены не одним определенным АБ, а целыми комплексами, причем сильные границы обогащены большим количеством разнообразных АБ [73].

В ходе текущих исследований не раз предпринимались попытки однозначно установить главенствующую роль транскрипции либо АБ. Дополнительным препятствием на этом пути является расположение сайтов связывания большинства белков-инсуляторов в непосредственной близости от стартов транскрипции активных генов. Метаанализ полученных результатов указывает скорее на отсутствие глобального механизма, регулирующего силу границ ТАДов дрозофилы исключительно посредством АБ или транскрипции [70,71]. Однако, хотя разнообразные подходы к изучению границ ТАДов и не позволили выявить один определяющий фактор, ответственный за их образование, результаты большей части исследований все же указывают на доминирующее влияние транскрипционной активности. В работе Роули и соавторов при выделении сайтов связывания архитектурных белков, не ассоциированных со стартами транскрипции, было показано, что сами по себе, в отрыве от транскрипции, они не проявляют заметной ассоциации с границами ТАДов, и, наоборот, гены с высокой экспрессией, не содержащие поблизости мотивы связывания архитектурных белков, обладают этим свойством [23]. Анализ границ ТАДов, содержащих активно транскрибирующиеся гены, без известных мотивов связывания белков-инсуляторов в радиусе 1 тыс.п.о. показывает, что эти участки генома ограничивают взаимодействия соседних с ними регионов не хуже, чем аналогичные в комбинации с архитектурными белками. Возможно, эти границы обогащены неизвестными на сегодняшний день белками, либо транскрипция сама по себе обладает инсуляторными свойствами. В поддержку такой интерпретацией говорит тот факт, что при выключении белка CP190 повышение частоты контактов через старты транскрипции оказывается менее выраженным для генов с более высокой транскрипцией в контроле [71]. Преобладающее влияние транскрипции (или каких-то факторов, с ней ассоциированных) было выявлено еще более наглядно при удалении промотора, оказавшем заметно более сильное влияние на ослабление границы по сравнению с удалением участка связывания белка-инсулятора. Исследование структуры хроматина на ранних стадиях эмбриогенеза показало, что деплеция белка-активатора Zld оказывает более сильное влияние на выраженность

границы по сравнению с деплецией белков-инсуляторов [74]. При этом, хотя для большинства АБ при деплеции не обнаруживаются изменений в масштабах полного генома, находится немало примеров, когда выключение определенного АБ приводит к единичным случаям потери границы какого-то конкретного ТАДа [71].

Несмотря на обширный массив доводов в пользу лидирующей роли транскрипции, очевидно, что белки-инсуляторы в силу одной только широкой представленности в геноме дрозофилы должны каким-то образом глобально участвовать в организации хроматина. Гипотетически, АБ могут играть определенную роль именно в сочетании с транскрипцией, например, участвовать в привлечении Pol II. В такой модели границы остаются стабильными при деплеции АБ, пока в месте связывания продолжается активная транскрипция [75]. Ряд наблюдений, указывающих на существенную роль АБ, дополняется одним из последних исследований с привлечением ряда данных по деплециям BEAF32 вместе с DREF, который может компенсировать отсутствие первого, а также CP190/Chromator было показано, что хотя обогащение энхансерами или активными стартами транскрипции имеет место для большинства границ ТАДов, все это множество можно разбить на 2 класса: (1) границы ТАДов, расположенные в эухроматине, и (2) границы ТАДов, расположенные в гетерохроматине. Первые сохраняются при деплеции BEAF32, CP190 и Chromator, тогда как вторые исчезают [76].

1.2.2.4 Хроматиновые петли

Для пространственной организации хроматина дрозофилы характерна еще одна разновидность пространственных паттернов, а именно, петлевые контакты. Подобные структуры у данного вида отличаются разнообразием в функциональном отношении. Например, анализ изменения организации хроматина в развитии выявил последовательное формирование активных и неактивных петель. Причем появление активных петель также можно разбить на два этапа, соответствующих малой и основной ЗГА. Первая характеризуется наличием петель, ассоциированных с Zld, а вторая - ассоциированных с CP190 и CTCF. Неактивные петли появляются позже, на стадии среднебластуляционного перехода и

обусловлены взаимодействием их оснований с белками группы поликомб [77]. Анализ структуры хроматина в клеточной линии нервной ткани ML-DmBG3-c2 также показал ассоциацию оснований петель с BEAF32, Chromator и/или CP190, с медиаторным комплексом и поликомбом. При деплеции Chromator/CP190 или рекрутирующего их белка BEAF32 наблюдалось уменьшение количества петель. Основания части петель оказались обогащены также когезином и CTCF, что указывает на возможность их участия в формировании подобных структур не только у млекопитающих, но и у дрозофилы [76]. Интересно, что, согласно результатам этого исследования, петли, связанные с поликомбом, встречаются гораздо реже по сравнению с другими типами петель, что идет вразрез с полученными ранее представлениями о репрессивной природе петель у дрозофилы. Кроме того, претерпевает изменения и сама концепция репрессивных поликомбных петель: в исследованиях эпигенетических профилей, транскрипции и данных Hi-C в имагинальном диске дрозофилы было показано, что часть активных энхансерно-промоторных взаимодействий, детектируемых на карте Hi-C в виде ярких петель, происходит при участии PRC1-комплекса и белка RING1B [78].

1.3 Взаимодействие хроматина с ядерной оболочкой

Исследование факторов, вовлеченных в организацию хроматина, было бы неполным без анализа такого важного компонента клетки, как ядерная оболочка. О том, каковы общие характеристики белков внутренней оболочки ядра, как они связаны с упаковкой ДНК в общем и у дрозофилы в частности, и пойдет речь далее.

1.3.1 Взаимосвязь между ядерной ламиной и структурой хроматина

Ядерная оболочка состоит из наружной ядерной мембраны, примыкающей к эндоплазматическому ретикулуму, внутренней ядерной мембраны, обращенной к нуклеоплазме, комплексов ядерных пор и примыкающей к внутренней ядерной мембране ядерной ламины [79]. Ядерная ламина - важный структурный компонент оболочки ядра многоклеточных, состоящий из большого количества элементов, включая ламины – ее основные строительные блоки [80]. У дрозофилы есть два

типа ламинов, кодируемых разными генами. Основное структурное различие между этими типами ламинов заключается в наличии (В-тип) или отсутствии (А-тип) мотива СааХ на конце С-концевого хвостового домена (С - цистеин; а - алифатический остаток; Х - любой остаток). Как и у млекопитающих, ламин типа В дрозофилы экспрессируется на протяжении всего развития во всех типах клеток, в то время как ламин типа А экспрессируется, начиная с поздней, двенадцатой, стадии эмбриогенеза [80]. Ламины типа А и В взаимодействуют с разными компонентами внутренней ядерной мембраны и хроматина и, по-видимому, выполняют разные функции, к коим относятся механотрансдукция, модуляция жесткости и эластичности ядра, репликация, репарация ДНК и регуляция транскрипции [81]. Большинство исследований в данной области показали, что гены, расположенные на ядерной периферии, транскрипционно репрессированы, в то время как отсоединение от ламин коррелирует с активацией генов. Однако эта концепция в силу противоречивости имеющихся данных на сегодняшний день не является окончательной.

В современных моделях организации ядра предполагается, что ламина играет важную роль в поддержании пространственной структуры и функционировании генома, в том числе опосредует локализацию гетерохроматина на периферии ядра [80]. Первые свидетельства о пространственной колокализации конденсированного хроматина и ядерной ламин были получены еще методами электронной микроскопии, а позже подкреплены экспериментами FISH, которые показали также, что определенные транскрипционно неактивные регионы, связанные с гетерохроматином, расположены на периферии ядра и взаимодействуют с ламиной. В настоящее время для профилирования геномных участков, контактирующих с ядерной ламиной, часто применяются методы DamID, ChIP-seq и GO-CaRT. DamID основан на определении уровня метилирования аденина в масштабах всего генома, а оно в свою очередь является результатом работы специфичной метилазы, сшитой с белком, который взаимодействует с хроматином. С помощью этого метода при профилировании ламина В были выявлены участки генома, часто вступающие в контакт с ядерной ламиной. Эти регионы получили

название ламина-ассоциированных доменов (ЛАДов), и, в соответствии с полученными ранее представлениями, отличались низкой транскрипционной активностью и преобладанием гетерохроматина [81]. Интеграция с данными Hi-C ожидаемо демонстрирует ярко выраженное перекрытие между ЛАДами и неактивным компартментом. Разработанные недавно методы секвенирования одиночных клеток показывают, что ЛАДы, детектированные по популяционным данным, могут при рассмотрении отдельных клеток располагаться как поблизости с ядерной оболочкой, так и в центральной области ядра [82].

Методами электронной микроскопии и FISH было показано, что нокдаун ламина В в клетках S2 дрозофилы уменьшает компактизацию части неактивных доменов хроматина [83], при этом доступность промоторных и ассоциированных с гетерохроматином областей увеличивается [84]. Эти результаты соотносятся с данными, полученными методом Hi-C: при деплеции ламина В, на фоне декомпактизации неактивных ТАДов, сопровождающейся повышением уровня ацетилирования и фоновой транскрипции, наблюдается увеличение плотности активных ТАДов [10].

Правильная сегрегация активного и неактивного хроматина осуществляется за счет взаимодействия ламинов как с ДНК/нуклеосомами, так и с ядерной оболочкой. В работе на дрозофиле, гомозиготной по мутации LamA25 (получена с помощью рентгеновского облучения, удаляет мотив СааХ в ламине В), было показано, что подобная дисфункция может приводить к деконденсации гетерохроматина, повреждению ДНК и гибели нейронов у взрослых дрозофил [85]. Еще один эксперимент, основанный на микроскопии высокого разрешения, был проведен для клеток слюнных желез и провентрикулуса. В своем исследовании Бондаренко и Шарапов показали, что блок СааХ необходим для связывания Lamin В с ядерной оболочкой и что белок Lamin В, лишенный мотива СааХ, не ограничивается взаимодействиями на периферии, а перераспределяется внутри ядра и колокализуется с политенными хромосомами. Также оказалось, что мутантный ламин не влияет на локализацию других периферических белков, но приводит к заметному сдвигу в распределении хроматина от периферии к центру

ядра в провентрикулусе; в клетках слюнных желез подобного эффекта не наблюдалось [80].

Таким образом, результаты проведенных на дрозофиле исследований говорят о том, что наряду с каноническими общеизвестными функциями, ядерная ламина также способствует поддержанию компактизации и правильной сегрегации активного и неактивного хроматина внутри ядра, причем вклад белков ламин в данном контексте может меняться в зависимости от типа клеток.

1.3.2 Нуклеопорины

Еще одним важным компонентом внутренней ядерной мембраны наряду с ламиной является ядерный поровый комплекс (ЯПК).

Комплекс ядерной поры (ЯПК) — это крупный белковый комплекс, интегрированный в ядерную оболочку, а его основная функция – обеспечение селективного транспорта между нуклеоплазмой и цитоплазмой. ЯПК состоит из 30 различных субъединиц, нуклеопоринов (Nups), которые в составе комплекса образуют специфичные структуры, такие как ядерная корзина, филаменты, коровая часть и др. Встраиваемый в ядерную оболочку комплекс из нуклеопоринов имеет форму окружности и состоит из внешнего и внутреннего кольцевых образований. После сборки, которая начинается в анафазе митоза, нуклеопорины этих субкомплексов остаются связанными с кором ядерной поры в интерфазе, но при этом подлежат периодическому обновлению, причем для разных нуклеопоринов время замены варьируется. Нуклеопорины, как правило, присутствуют практически во всех типах клеток, однако для многих из них характерна специфичная экспрессия в определенных клеточных линиях и тканях [86].

Соотношение и взаимное расположение нуклеопоринов и ядерной ламин таково, что, согласно данным по DamID в дрозофиле, усредненный по геному пик профиля для ассоциированного с ЯПК нуклеопорина Nup98 точно соответствует провалам в профиле, характеризующем участки связывания ламин. Это наблюдение было также подтверждено с помощью ChIP-seq. На основе полученных данных была предложена модель, согласно которой протяженные

участки генома попеременно прикрепляются либо к ядерной ламине в составе ЛАДов, либо к ЯПК (рис. 6) [87].

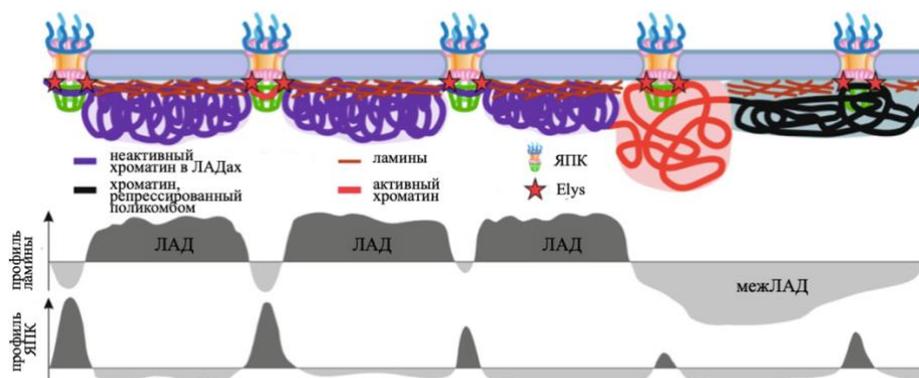


Рисунок 6. Схематичное изображение взаимодействия хроматина с ядерной ламиной и комплексом ядерной поры (адаптировано из *Shevelyov, 2020*).

1.3.3 Локализация нуклеопоринов и их роль в положительной регуляции транскрипции

Помимо транспорта белков и макромолекул в составе ЯПК часть нуклеопоринов участвует и в других процессах. Например, нуклеопорин ELYS связывается с хроматином в позднем митозе и необходим для сборки ЯПК перед встраиванием в ядерную мембрану [88]. Известно, что только часть нуклеопоринов локализуется исключительно на периферии, будучи связанной с ЯПК. Для ряда нуклеопоринов было обнаружено, что существенная их доля находится также и во внутренней части ядра и в основном ассоциирована с положительной регуляцией транскрипции (однако такая однозначная корреляция наблюдается не всегда) [89,90]. Например, анализ ChIP-chip у дрозофилы показал, что компоненты «ядерной корзины» Nup153 и Mtor/Trg связаны с активно экспрессирующимися областями, обогащенными метками активного хроматина, такими как ацетилирование H3K16 и Pol II. Подавление работы этих нуклеопоринов отрицательно сказалось на экспрессии тысяч генов [91].

Исследования последних лет также привели к пониманию, что помимо промоторов, таргетными участками компонентов ЯПК часто становятся энхансеры и, в частности, суперэнхансеры. Так, Nup98 был детектирован во множестве энхансеров по всему геному дрозофилы, включая гены, регулируемые

стероидными гормонами экдизонами, и сложные энхансеры кластеров Нох-генов [86]. Деpletion Nup98 у дрозофилы приводила к нарушению формирования энхансер-промоторных петель в генах, активируемых экдизоном. Причем выключение этого нуклеопорина не нарушало работу таких промоторов во время первоначального воздействия экдизоном, однако нарушение энхансерно-промоторных взаимодействий коррелировало с уменьшением повторной активации транскрипции при последующей экспозиции. Этот факт позволяет предположить, что, стабилизируя энхансер-промоторные контакты, Nup98 может тем самым участвовать в механизме транскрипционной памяти [13]. Интересно, что эта стабилизация хорошо коррелирует с усилением взаимодействия между Nup98 и такими архитектурными белками дрозофилы, как Cp190, Su(Hw), GAF и CTCF [92]. Также для Nup98 в дрозофиле была показана его непосредственная связь с подмножеством границ ТАДов, характеризующихся сильной инсуляцией.

Кроме того, свободные от ЯПК нуклеопорины также взаимодействуют с некоторыми белками, модифицирующими гистоны и ремоделирующими хроматин. Лучше всего в данном контексте освещена связь между Nup98 и SET-домен-содержащими H3K4-метилтрансферазами (Set1, MLL1/2 или его гомолог Trithorax), опосредующими ди- и триметилирование H3K4 [92].

До сих пор неизвестно, как нуклеопорины рекрутируются во внутриядерные области и могут ли их взаимодействия с хроматином в составе ядерной поры и вне ее приводить к дифференциальным событиям транскрипционной регуляции. Однако несколько примеров такой регуляции уже описано в литературе [86].

1.3.4 Репрессивные функции нуклеопоринов

Участие нуклеопоринов в пространственной организации генома и регуляции экспрессии также проявляется в репрессии генов. Согласно недавним исследованиям, Nup93 выполняет высококонсервативную функцию по поддержанию хроматина в неактивном состоянии. Анализ полногеномного распределения Nup107 и Nup93, двух представителей компонентов внешнего и внутреннего кольца, в клетках дрозофилы показал, что специфичной мишенью Nup93 являются хроматиновые домены, репрессированные комплексом Pc. Также

была выявлена ассоциация Nup93 с компонентами PRC1, такими как Pc, dRing и RYBP [93]. При деплеции Nup93 наблюдалось разрушение кластеризации доменов, связанных с Nup93 и комплексом Pc. Предположительно, такая кластеризация способствует стабилизации сайленсинга участков генома, ассоциированных с поликомбом [86]. Интересно, что по данным Hi-C домены хроматина, репрессированные Pc и одновременно связанные с Nup93, преимущественно взаимодействуют именно друг с другом. Это наблюдение согласуется с исследованиями на *S. pombe*, где потеря Nup106, гомолога Nup93, приводит к декластеризации фокусов гетерохроматина, особенно в теломерах. Другие нуклеопорины, например Nup155, также вовлечены в репрессию определенных участков хроматина. Так, для гомолога Nup155 в дрожжах (Nup170) было показано его участие в подавлении активности субтеломерных регионов гетерохроматина, а для Nup155 в кардиомиоцитах крысы взаимодействие с гистоновой деацетилазой HDAC4. В целом, эти исследования подчеркивают значимость определенных компонентов субкомплекса внутреннего кольца Nup93-Nup155 для сближения удаленных локусов гетерохроматина и сайленсинга генов.

1.3.5 Нуклеопорин Elys

Elys – особенный белок в ряду нуклеопоринов. Это единственный компонент ЯПК с установленными ДНК-связывающим и хроматин-связывающим доменами (AT-hook домен и участок Arg-Arg-Lys (RRK) в С-концевой области соответственно). Способность напрямую взаимодействовать с хроматином может быть связана со сборкой ядерной поры после митоза, где Elys привлекает субкомплекс внешнего кольца Nup107 для инициации этого процесса. Данные Крио-ЭМ и масс-спектрометрии показывают, что С-концевая область ELYS непосредственно связывается с кислотным участком нуклеосомы [94]. Также как и многие другие нуклеопорины, Elys является многофункциональным белком: недавние работы указывают на его участие в изменении компактизации хроматина и контроле репликации ДНК [92]. Так, в своей работе на дрозофиле Кун и соавторы пришли к выводу, что эктопическое связывание определенных нуклеопоринов с хроматином достаточно для того, чтобы вызвать его деконденсацию. Оказалось,

что деконденсация опосредована ремоделирующим хроматин комплексом PBAF. Однако этот комплекс действует не в одиночку: процесс деконденсации коррелирует с рекрутированием Elys, более того, дополнительный анализ указывает на возможную причинно-следственную связь между уровнем привлечения Elys и деконденсацией [95]. В статье 2022 года о роли Elys в процессе репликации Ричардсом и соавторами было показано, что у дрозофилы деплеция Elys, но не других нуклеопоринов, уменьшает количество связанного с хроматином ORC2 (комплекс распознавания сайта начала репликации) по всему геному, в связи с чем было выдвинуто предположение о функционировании Elys в качестве регулятора загрузки оптимального уровня ORC на хроматин [88].

Согласно данным ChIP-seq, сайты связывания Elys, представленные в геноме в огромном количестве, хорошо пересекаются с сайтами Nup98. А они в свою очередь были успешно разделены ранее на ассоциированные с ЯПК и внутренней частью ядра [90], из чего можно предположить, что, наряду с другим белками ЯПК, для Elys тоже характерна дуальная локализация, а связывание с хроматином в нуклеоплазме ассоциировано с транскрипцией, что подтверждается обогащением меткой ацетилирования и колокализацией с активными промоторами генов и энхансерами. Для значительной части геномных мишеней Elys, как ЯПК-связанного, так и нуклеоплазмального, было показано высокое содержание А/Т-богатых мотивов ДНК, которые, вероятно, опосредуют взаимодействие Elys с этими сайтами. Интересно, что при исследовании мотивов связывания Elys в контексте местоположения относительно старта и конца генов, было обнаружено, что только для 3' конца обогащение А/Т-мотивом строго совпадает с обогащением профиля Elys, тогда как в промоторах генов наблюдается колокализация пиков Elys и GAF. Это наблюдение может указывать на существование непрямого механизма взаимодействия Elys с хроматином через комплекс GAF/PBAF.

В большинстве клеток изучаемых организмов при выключении белка Elys имеет место потеря ядерных пор, а следовательно, гибель объекта исследования и невозможность его изучения. Примечательно, что в отличие от нематод, лягушек, а также клеток слюнных желез дрозофилы, нокдаун Elys в клетках линии S2 не

приводит к таким последствиям. Только одновременный нокдаун Lam и Elys приводит к выраженному снижению концентрации ЯПК в области ядерной оболочки. Т. е. в клетках S2 включение ядерных пор в оболочку в конце митоза может происходить каким-то другим, независимым от Elys способом. Отсутствие заметных дефектов ядерной ламины и комплекса пор, а также нормальная жизнеспособность клеток при нокдауне Elys дали возможность более подробно изучить его функции в контексте структуры хроматина. В работе Доронина и соавторов использование метода DamID позволило определить сайты связывания Elys в поздних эмбрионах, которые были классифицированы на ЯПК-ассоциированные ($Elys_{\text{ЯПК}}$) и нуклеоплазма-ассоциированные ($Elys_{\text{нукл}}$) на основании пересечения с полученной ранее разметкой и классификацией для Nup98. Было показано, что ~70% и 95% сайтов $Elys_{\text{ЯПК}}$ и $Elys_{\text{нукл}}$ пересекается с неактивным и активным хроматином соответственно.

Поскольку заметная часть Elys ассоциирована с ЯПК, он также может быть необходим для прикрепления периферического хроматина к ядерной оболочке в интерфазе. Действительно, с помощью FISH и иммуноокрашивания хроматина было обнаружено смещение нескольких анализируемых локусов, а также хроматина в целом от периферии во внутреннюю часть ядра при нокдауне Elys, что в свою очередь привело к повышению транскрипции некоторой части ламина- и ЯПК-ассоциированных неактивных генов. Также анализ показал, что ЯПК-ассоциированный Elys не просто находится в контакте с хроматином, но и связан с ним [14].

Обнаруженные закономерности позволили в общих чертах описать изменение структуры хроматина при деплеции Elys, однако такой методический подход ввиду описанных ранее ограничений не дает возможности более подробного исследования вклада данного белка в организацию ТАДов, компартментов, либо более тонких, локальных, паттернов. Данные Hi-C в таком случае могли бы послужить полезным источником дополнительной информации.

1.4 Процесс сперматогенеза у дрозофилы как совокупность факторов, потенциально влияющих на структуру хроматина

Согласно приведенному выше анализу литературы, транскрипция является одной из наиболее существенных составляющих в совокупности факторов, ассоциированных со структурой хроматина дрозофилы, однако этот аспект остается изученным не в полной мере, поскольку значительная часть исследований основывается на искусственном подавлении транскрипции, сравнении разных типов клеточных линий или наблюдением за процессом становления хроматина при активации транскрипции на ранних стадиях эмбриогенеза [72,96]. Анализ структуры хроматина в ходе сперматогенеза дрозофилы дает возможность проследить взаимосвязь между изменениями трехмерной укладки ДНК, сформировавшейся на ранней стадии этого процесса, и активацией большого числа генов, а также других, более специфичных для этого процесса событий. Настоящий раздел посвящен описанию основных стадий и особенностей протекания сперматогенеза у данного вида.

Во время сперматогенеза дрозофилы первичный гониобласт, окруженный клетками цист, проходит 4 деления митоза с неполным цитокинезом, что приводит к образованию цисты из 16 сперматогониев, соединенных между собой кольцевыми каналами [97]. На следующей стадии, предшествующей мейозу, - стадии ранних сперматоцитов - клетки существенно увеличиваются в размере. За 3,5 дня объем сперматоцитов возрастает в 25 раз, причем это увеличение сопровождается инициацией транскрипции более тысячи сперматоцитспецифичных генов. Всплеск транскрипции опосредован масштабным ремоделированием хроматина, запускаемым экспрессией семенникспецифичного комплекса ареста мейоза (tMAC), а также семенникспецифичных TBP-ассоциированных факторов – белков, ассоциированных с ТАТА-связывающим белком при инициации транскрипции (TAFs) [98,99]. Согласно ряду исследований, помимо транскрипционных факторов в регуляции экспрессии сперматоцитспецифичных и семенникспецифичных генов, в особенности в составе кластеров, также участвуют белки ядерной ламины. Так, в исследовании Шевелева и соавторов было показано, что значительная доля этих генов в соматических клетках приходится на неактивный, ассоциированный с ламиной, хроматин.

Деплеция белков ламины приводила к диссоциации участков, содержащих эти гены, и частичному повышению их экспрессии. Кроме того, было обнаружено, что активация сперматоцитспецифичных генов в норме сопровождается частичной потерей связи их промоторов с ламинной [100,101]. В продолжении процесса сперматогенеза транскрипция почти полностью прекращается непосредственно перед мейозом, в результате которого образуется 64 гаплоидных сперматиды, однако сохраненные транскрипты транслируются после мейоза, что приводит к резким изменениям морфологии сперматид и, наконец, к образованию подвижных сперматозоидов (рис. 7).

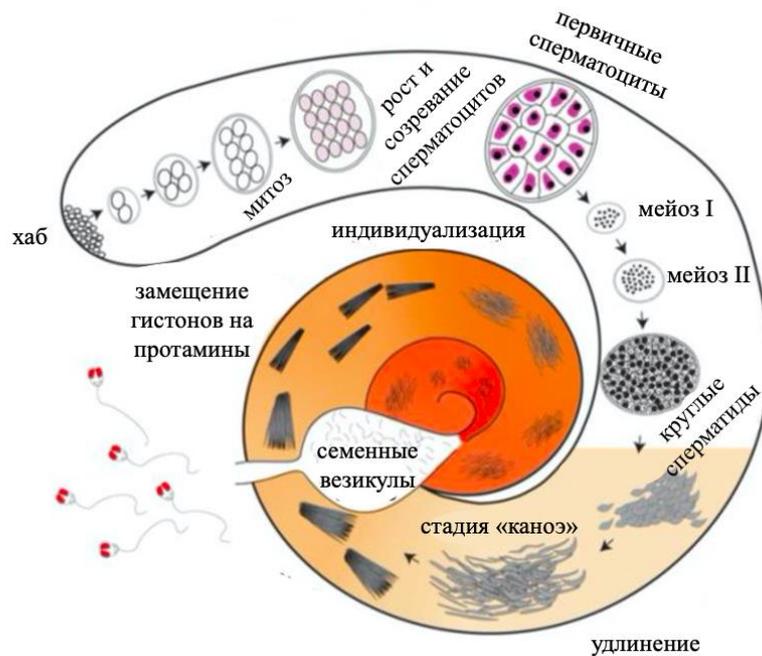


Рисунок 7. Схематичное изображение стадий сперматогенеза дрозофилы (адаптировано из *Mills et al, 2019*).

Между сперматогенезом дрозофилы и млекопитающих есть существенное отличие. В профазе первого мейоза, в преддверии гомологичной рекомбинации, пространственная организация генома млекопитающих претерпевает серьезные изменения. Уже в период, предшествующий лептонеме, наблюдается преобразование структуры хроматина, обусловленное постепенной конденсацией [102]. Ввиду отсутствия рекомбинации у самцов дрозофилы [103], эта составляющая не дает столь заметного вклада в реорганизацию трехмерной

структуры на стадии первичных сперматоцитов и не приводит к характерным для этого процесса изменениям, что способствует более однозначной интерпретации динамики структуры хроматина. Поскольку формирование половых клеток не обходится у данного вида без мейотического деления, конденсация хроматина все же происходит, однако протекает довольно быстро и непосредственно перед первым делением [19]. Интересно, что мейоз у самцов дрозофилы настолько отличается от наблюдаемого в млекопитающих, что для описания основных этапов этого процесса была разработана собственная номенклатура [104].

Проведенный анализ литературы показывает, что интерес со стороны ведущих исследовательских групп привел к значительным успехам в понимании основных принципов организации хроматина дрозофилы. С другой стороны, многие аспекты остаются изученными не в полной мере, что объясняется не только масштабностью тематики, но и наличием определенных ограничивающих факторов экспериментально-технического характера. Так, заметная часть исследований фокусируется на роли белков-инсуляторов, что, вероятно, обусловлено не только их функциональной значимостью, но и избыточностью, а значит, более вероятной выживаемостью объекта исследования при деплеции. В то же время на белки периферии в данном контексте приходится меньший объем публикаций. Значительный объем информации представлен работами по ассоциации с эпигенетическими состояниями и транскрипцией, однако исследования такого характера проводились в основном на клеточных линиях либо на ранних стадиях развития в контексте формирования структуры хроматина, в т. ч. с подавлением транскрипции. В этом отношении анализ ранних стадий сперматогенеза дает возможность проследить эту взаимосвязь в дополнительном ракурсе, а также является самостоятельным предметом исследования ввиду отсутствия в литературе информации о данных Hi-C для каких-либо стадий этого процесса у дрозофилы на фоне подробно описанной динамики трехмерной организации генома в ходе сперматогенеза млекопитающих.

Глава 2. Материалы и Методы

Для анализа изменения в архитектуре хроматина в ходе сперматогенеза дрозофилы были использованы гомозиготные мутантные семенники *bamΔ86* из имаго, а также семенники дикого типа из личинок третьей стадии. Мутантные семенники *bamΔ86* являются широко используемой моделью для анализа митотически делящихся зародышевых клеток. Отсутствие белка Bam в таких семенниках приводит к неспособности клеток зародышевой линии дифференцироваться в сперматоциты и, как следствие, к накоплению кист, заполненных клетками зародышевой линии, напоминающими сперматогонии (рис. 8). В семенниках личинок третьего возраста сперматогенез не идет дальше профазы мейоза. Следовательно, зародышевые клетки в их семенниках, хотя и содержат небольшую фракцию сперматогониев, в основном представлены первичными сперматоцитами. Эти сперматоциты могут быть не идентичны сперматоцитам взрослых самцов. Однако, согласно анализу секвенирования РНК, экспрессия сперматоцитспецифичных генов в личиночных семенниках сильно коррелирует с таковой в имаго. Данные Hi-C при деплеции Elys были получены для клеточной линии S2 методом РНК-интерференции. В качестве контроля при проведении РНК-интерференции была использована та же клеточная линия с добавлением двуцепочечной РНК, комплементарной последовательности гена бактериального белка LacZ. Лабораторная часть исследования была проведена сотрудниками коллективов под руководством Ю.Я. Шевелева и С.В. Разина. Также в работе использованы публичные данные (см. далее раздел 2.2.11).

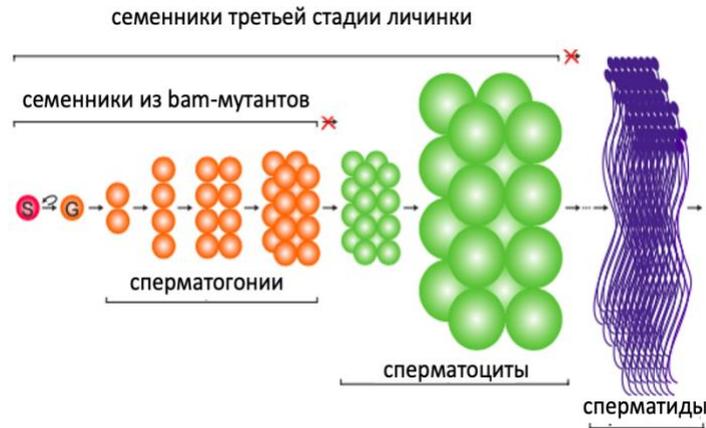


Рисунок 8. Схема сперматогенеза в дрозофиле до стадии удлинённых сперматид.

Применяемые в работе биоинформатические методы можно отнести к двум этапам анализа данных: генерация карт Hi-C с предварительной предобработкой и непосредственно поиск изменений в структуре хроматина на основе интеграции полученных карт взаимодействий и других видов омиксных данных.

2.1 Предобработка сырых данных, генерация карт Hi-C, оценка качества эксперимента

В ходе первичной обработки данных Hi-C для двух стадий сперматогенеза и образцов из эксперимента по деплеции Elys был проведен контроль качества с помощью программы FastQC. Генерация карт Hi-C, в том числе из публичных данных для эмбриональной клеточной линии Kc167, проводилась с помощью инструмента distiller-nf [105], который автоматизирует последовательность всех необходимых стадий обработки с учетом специфики данных, а именно картирование чтений на геном версии dm3, удаление ПЦР дубликатов, определение взаимодействующих пар, агрегацию пар в бины, создание файлов с несколькими разрешениями и последующую итеративную коррекцию.

Было замечено, что использование встроенного в программу distiller инструмента fastp для отрезания адаптеров по какой-то причине уменьшает количество контактов на выходе, причем качество карт от этого не улучшается. Поэтому, несмотря на удобство использования и высокую скорость работы fastp, обрезка адаптеров осуществлялась с помощью не столь эффективного по времени,

но проверенного инструмента Trim galore, запущенного с параметром --stringency 3 [106].

2.1.1 Генерация карт контактов

На первом этапе работы distiller проводится картирование чтений встроенной программой для выравнивания bwa-mem. Т. к. Hi-C позволяет детектировать парные контакты в масштабах полного генома, взаимодействующие локусы могут располагаться на больших линейных расстояниях друг от друга и даже на разных хромосомах. Кроме того, если длина чтения составляет 150-300 пар оснований, он может включать последовательности двух (а иногда и более) контактирующих участков генома, поэтому при картировании данных такого типа разрешается возможность разрыва чтения и независимого выравнивания его частей по отдельности. Далее на основе данных, записанных в файлах с выравниванием, с помощью инструмента pairtools в составе distiller проводится процедура агрегирования информации о геномных координатах, соответствующих каждому чтению и расположению в геноме парного с ним, в результате генерируется файл в специальном формате “pairs”, который фактически представляет собой список всех обнаруженных контактов. После процедуры фильтрации дубликатов и отбора по параметру MAPQ30 ($\text{mapq1} \geq 30$ и $\text{mapq2} \geq 30$), которая в pairtools проводится на уровне пар контактов из файла .pairs, а не на уровне файла с выравниванием, как это обычно бывает для других типов омиксных данных, пары контактов суммируются в пределах геномных бинов выбранного размера. Процедура бинирования (агрегирования сигнала в бины) позволяет получать более устойчивую и менее зашумленную оценку количества контактов. Для агрегирования взаимодействующих пар в матрицу контактов используется встроенный инструмент cooler [105]. С его помощью генерируются карты контактов на нескольких заданных пользователем разрешениях в виде файла с расширением .mcool, который представляет собой имплементацию модели данных контейнера hdf5 - иерархического формата для многомерных массивов однородного типа, который поддерживает разбиение на части, сжатие и произвольный доступ. Значения в матрице Hi-C после объединения контактов пар

в контакты геномных бинов представляют собой ненормированные данные и могут определяться не только истинной частотой взаимодействий, но также вариациями количества сайтов рестрикции в разных бинах, GC-составом, картируемостью и артефактами протокола Hi-C. Основываясь на предположении о том, что, с физической точки зрения, каждый локус может быть вовлечен в разнообразные контакты, однако их общее число является примерно одинаковым, для нормализации карт Hi-C применяется метод балансировки матриц, основанный на алгоритме Синкхорна-Кноппа [107]. Он относится к неявным методам коррекции и теоретически позволяет нивелировать влияние как известных, так и скрытых, искажающих истинные значения, факторов. На первом этапе балансировки первый столбец матрицы делят на среднее значение всех контактов в этом столбце, после чего первая строка делится на сумму всех контактов в этой строке. Процедура проводится последовательно для всех строк и столбцов, начиная с первых и заканчивая последними, таким образом завершается первый раунд коррекции. Описанный способ нормализации повторяется несколько раз до тех пор, пока суммарное число контактов в каждом столбце и каждой строке не станет близким к единице. Поскольку одна и та же процедура воспроизводится несколько раз, данный вид коррекции получил название итеративной. Коррекция была реализована с помощью программы cooltools [108], также встроенной в distiller. Данный этап завершает стадию получения карт Hi-C.

2.1.2 Удаление артефактов и оценка качества данных

После получения карт проводилась оценка среднего расстояния между контактами двух локусов, на которое приходится основная часть артефактов протокола, таких как сшивка участка самого с собой, лигирование в месте фрагментации двух соседних участков, ситуация, когда фрагмент ДНК не был сшит ни с одним другим, но был отсековирован. Интервал расстояний, в котором обнаруживается наибольшее число артефактных взаимодействий, был определен на основе соотношения количества всевозможных правильно ориентированных пар чтений. К таким парам относятся направленные друг на друга, направленные в противоположные стороны, направленные влево и направленные вправо.

Поскольку такие события должны встречаться с равной частотой, а артефактные контакты характеризуются определенной направленностью чтений и, следовательно, вносят смещение в это распределение, для их фильтрации достаточно исключить из анализа взаимодействия в пределах расстояний, на которых доля ориентации каждого из четырех вышеперечисленных типов существенно отличается от $\frac{1}{4}$. Для Hi-C карт сперматогониев, сперматоцитов и эксперимента с деплецией Elys это расстояние составило не более 1 тыс.п.о., в связи с чем все контакты участков, разделенных менее, чем 1 тыс.п.о., были удалены из матриц взаимодействия.

Для оценки сходства реплик карт контактов, был применен инструмент HiCRep [109].

Такой подход позволяет получить на выходе стратифицированный коэффициент корреляции (СКК), который, после усреднения по всем хромосомам, дает информацию о мере сходства двух карт контактов. Эта информация затем используется для построения дендрограммы, служащей наглядной репрезентацией того, насколько каждая из реплик похожа на все остальные. Перед процедурой подсчета СКК данные Hi-C прошли процедуру сэмпирования, т. е. удаления части контактов с целью получения равноценных по суммарному числу взаимодействий карт. Этот этап проводился с помощью пакета cooltools. Подсчет СКК проводился на разрешении 10 тыс.п.о. После получения удовлетворительных результатов кластеризации исходные карты реплик были объединены внутри групп из одного эксперимента, а затем для карты с большим суммарным числом контактов была проведена процедура сэмпирования. Последующий анализ, в зависимости от специфики поиска определенных структур на карте Hi-C проводился на разрешении от 2 до 20 тыс.п.о.

2.2 Анализ карт контактов

2.2.1 Структура компартментов

2.2.1.1 Детекция компартментов

Для определения крупномасштабных изменений в ходе сперматогенеза и при деплеции Elys были проанализированы компартменты. Поиск компартментов

проводился методом разложения матрицы контактов на главные компоненты в предположении о том, что знак первого собственного вектора определяет принадлежность бина к А- или В-компарменту. Такое определение обосновывается описанными ниже рассуждениями. Представим, что каждому бину в геноме присвоено число $c(x)$, характеризующее его принадлежность к множеству А (положительное значение) или В (отрицательное значение). Тогда оценка взаимодействия между двумя локусами x, y равна $c(x) \times c(y)$. Заметим, что такой формулировки достаточно для воспроизведения шахматной доски: при одинаковых типах x, y их знаки будут одинаковыми и дадут положительное значение взаимодействия, а разные типы приведут к отрицательному. В таком случае, при наличии матрицы частот взаимодействий, возникает задача нахождения $c(x)$ для каждого геномного бина. Оказывается, что первая главная компонента (PC1), дает для каждого бина оптимальное значение $c(x)$, такое, что разница между наблюдаемыми частотами взаимодействия и $c(x) \times c(y)$ минимальна (минимизируется средняя квадратичная ошибка). Таким образом, если структура компарментов достаточно сильна, то данная процедура должна его найти. Узким местом остается выбор компоненты и ее ориентация для определения знака, соответствующего А- или В-компарменту, а также разрешение. В настоящее время существует множество инструментов, реализующих описанные вычисления в автоматическом режиме и разработанных специально для применения к данным Hi-C.

Для определения компарментной принадлежности участков генома был использован инструмент cooltools. С его помощью были получены первые собственные векторы. Инструмент включает процедуру фазирования, т. е. при наличии дополнительной характеристики каждого бина генома определяет ее корреляцию с первыми собственными векторами и сортирует их по убыванию значения корреляции. В данном случае для этой цели был использован GC-состав. Такое решение оправдано для дрозофилы, поскольку А-компармент часто ассоциирован с GC-богатыми границами. Полученная разметка была проверена на соответствие между транскрипцией и знаком первого собственного вектора:

участки генома с положительными значениями отвечали более высокой транскрипции, что указывало на их принадлежность к А-компарменту. Выбор разрешения (10 тыс.п.о.) был обусловлен литературными данными и анализом экспрессии в А- и В-компарменте на разрешении 10 и 20 тыс.п.о. При сопоставлении уровня транскрипции в компарментах несколько более яркие отличия обнаруживались при разбиении генома на 20-тыс.п.о. бины, однако для обоих вариантов разрешений в целом наблюдалась сходная картина (рис. 1 Приложения). Поэтому, для повышения точности последующей аннотации, а именно установления соответствия между генами и компарментами (в т. ч. определение значения PC1 для старта транскрипции), в дальнейших расчетах использовалось разрешение 10 тыс.п.о.

2.2.1.2 Представление и анализ компарментной структуры

Исследование изменений в структуре компарментов проводилось через построение и анализ так называемых «седловых графиков» (от англ. saddle plot). Процесс их получения состоит в следующем: все столбцы карты внутрихромосомных взаимодействий переставляются таким образом, чтобы соответствующие им бины оказались ранжированными по величине значения первого собственного вектора от наименьшего к наибольшему. Аналогичная процедура проводится для строк, обеспечивая условное разделение карты контактов на 4 области: левый верхний угол карты соответствует контактам бинов с наименьшими значениями первого собственного вектора, т. е. взаимодействиям внутри В-компармента; правый нижний угол отвечает контактам бинов с наибольшим значением первого собственного вектора, т. е. взаимодействиям внутри А-компармента. Левый нижний и правый верхний угол являются симметричными и соответствуют контактам бинов с разными знаками первого собственного вектора, т. е. относятся к межкомпарментным взаимодействиям. Центральная часть такого графика обычно отражает переходную зону и плохо поддается интерпретации. Для удобства дальнейшего анализа значения частот контактов близлежащих бинов трансформированной матрицы усредняются, т. е. фактически имеет место понижение разрешения, однако такой подход в

совокупности с нормированием на ожидаемое значений взаимодействий дает возможность визуально оценить выраженность компартментов и возникающие при переходе между стадиями изменения. Для реализации метода был использован модуль saddle из пакета cooltools. В ходе анализа структуры компартментов все данные Hi-C, упомянутые в описании результатов исследования, были проанализированы единообразно: «седловые графики» построены на разрешении 10 тыс.п.о., с разбиением PC1 на 30 квантилей.

Исследование динамики хроматина в ходе сперматогенеза включало анализ В-В взаимодействий. Их аннотация ограничивалась участками генома, попадающими в первые 4 интервала значений PC1 ($PC1 < -0.34$ в сперматоцитах и $PC1 < -0.36$ в сперматогониях), полученных при делении множества значений PC1 на 30 квантилей, т. о. на «седловом графике» исследуемая область приходится на левый верхний угол, очерченный пунктирной линией. Т. к. данный раздел был посвящен исследованию общих закономерностей В-В взаимодействий, а не переходам между компартментами, для анализа экспрессии в этой области были дополнительно отобраны только те гены, которые содержатся в ней как в сперматогониях, так и в сперматоцитах.

В эксперименте с деплецией Elys был проведен дополнительный анализ структуры компартментов на разных расстояниях между контактирующими участками, для чего было выделено 3 интервала: 0.2-4 Мб (ближние контакты), 4-10 Мб (средние) и 10-20 Мб (дальние). Такие границы интервалов обусловлены намерением сравнивать компартментную структуру и описанные далее в разделе 2.2.2.2 взаимодействия ТАДов в пределах сопоставимых линейных расстояний. Указанные интервалы как раз являются некоторым усредненным вариантом разбиения расстояний внутри групп активных и неактивных ТАДов (см. далее раздел 2.2.2.2). Построение «седловых графиков» для контактов на разных масштабах взаимодействий было реализовано через модификацию исходного кода, заложенного в функцию saddle пакета cooltools. А именно, были введены дополнительные параметры, ограничивающие область контактов для дальнейшего построения графика посредством замены выходящих за обозначенные пределы

контактов пропущенными значениями. При этом порядок, в котором исходно выстраивались бины генома при их сортировке по возрастанию значения PC1, был сохранен.

2.2.2 ТАДы

2.2.2.1 Детекция ТАДов

Для детекции ТАДов был выбран инструмент Armatus [110]. Его преимущество в случае работы с данными дрозофилы заключается в большей по сравнению с другими инструментами дисперсией размера ТАДов, а также в наличии пропусков между границами ТАДов, т. е. межТАДов. Известно, что межТАДы дрозофилы соответствуют мелким ТАДАм, поэтому инструменты, не учитывающие возможность присутствия межТАДов или очень мелких ТАДов наряду с крупными (HiCexplorer, cooltools) в меньшей степени подходят для исследуемого организма.

Проблема определения границ ТАДов в алгоритме Armatus поставлена в виде задачи оптимизации сегментации генома. Наилучшая разметка ТАДов соответствует разбиению, при котором суммарная оценка его качества достигает максимального значения. Для каждого l -го бина генома можно получить промежуточное разбиение части матрицы, соответствующей первым l бинам, такое, что с одной стороны оно будет зависеть от разбиения на предыдущем шаге, а с другой, должно отвечать максимальному значению оценочной функции по всем возможным разбиениям этой части матрицы. Очевидно, что такая формулировка задачи относит ее к классу задач, решаемых методами динамического программирования. Оценочная функция задается уравнением:

$$q(k, l, \gamma) = s(k, l, \gamma) - \mu_s(l - k) \quad (1),$$

где

$$s(k, l, \gamma) = \frac{\sum_{g=k}^l \sum_{h=g+1}^l A_{gh}}{(l-k)^\gamma} \quad (2)$$

- масштабированная плотность участка матрицы. При $\gamma=1$ масштабированная плотность - это взвешенная плотность, которая представляет собой верхнюю треугольную часть подматрицы, определяемой доменом в интервале $[k, l]$,

деленную на его масштабированную длину $(l-k)\gamma$. При больших значениях γ длина домена в знаменателе увеличивается, следовательно, меньшие домены будут давать более высокое значение оценочной функции по сравнению с длинными доменами при одинаковых частотах взаимодействия. Уравнение (1) является центрированным относительно нуля. $\mu_s(l-k)$ - это среднее значение (2) по всем подматрицам длины $l-k$ вдоль диагонали матрицы, и оно может быть вычислено предварительно.

С точки зрения пользователя, определяющими параметрами при запуске Armatus являются разрешение и описанная выше гамма. При увеличении параметра гаммы алгоритм становится более чувствительным к границам и детектирует более мелкие ТАДы. Этот же эффект наблюдался при уменьшении размера бина. Для оптимальной подборки ТАДов было получено несколько вариантов разметки, соответствующих разным разрешениям и параметру гаммы. В качестве наилучшей была выбрана разметка на разрешении 5 тыс.п.о. с гаммой, равной 2.0 в сперматоцитах и 1.9 в сперматогониях. В эксперименте по деплеции Elys было выбрано разрешение 4 тыс.п.о. и параметр гамма 1.0 для аутосом. Для X-хромосомы ТАДы были размечены с параметром гамма 0.6 и 0.4 в контроле и при деплеции соответственно. Различия в значениях гаммы для аутосом и X-хромосомы связаны с различиями в числе контактов, которое объясняется наличием только одного гомолога половой хромосомы ввиду мужского происхождения клеточной линии S2. Важно отметить, что определение разметки проводилась на картах, прошедших специальную предобработку, а именно, во избежание детекции ложных границ в бинах с пропущенными значениями частот контактов, соответствующих плохо покрытым регионам генома (обычно они содержат большое количество повторов), было проведено заполнение (импутация) пропущенных значений. Процедура импутации может проводиться разнообразными способами: заполнением пропущенных контактов средним по двум соседним контактам, заполнением средним значением контактов между локусами, разделенными тем же линейным расстоянием и другими. Наилучший результат, с точки зрения визуальной оценки импутированной карты, дает

линейная интерполяция. После проведения импутации контакты, выходящие за пределы 1-го и 99-го перцентилей, были заменены на значения этих перцентилей. Подготовленные таким образом карты внутривромосомных взаимодействий подавались на вход инструменту Armatus.

2.2.2.2 Разделение ТАДов на группы по активности, оценка плотности и взаимодействия ТАДов при деплеции Elys

Анализ данных по деплеции Elys включал подсчет средней плотности ТАДов, которому предшествовало их разделение на группы по активности. Эта процедура проводилась для ТАДов с совпадающими границами в контроле и при деплеции Elys. Для оценки активности ТАДов был использован специальный коэффициент J , рассчитанный по формуле $J = \frac{X-Y}{X+Y}$, где X – доля ТАДа, приходящаяся на ЛАДы, Y – доля ТАДа, ассоциированного с активным хроматином. Под активным хроматином в данном случае понималось объединение красного и розового цветов хроматина, соответствующих меткам активных промоторов и элонгации транскрипции. На основании коэффициента J все ТАДы были разделены на 3 группы: активные (А), удовлетворяющие $1 < J < -0.8$, промежуточные (В) с коэффициентом J таким, что $-0.8 \leq J \leq -0.8$ и неактивные с коэффициентом J таким, что $0.8 < J < 1$. Число ТАДов в группах составило 876, 618 и 520 соответственно. В дополнение были рассчитаны плотности ТАДов в группах, разделенных по квартилям экспрессии. Отношение ТАДа к квартилю экспрессии определялось посредством усреднения значения транскрипции, выраженной в виде суммарного нормализованного числа чтений на бин, в каждом ТАДе, ранжирования ТАДов и разбиения на 4 группы равного размера.

Учитывая способность Elys связываться с хроматином, ожидалось, что его деплеция может приводить к изменению плотности ТАДов и интенсивности взаимодействия между ними. Плотность ТАДов определялась на разрешении 4 тыс.п.о. как средняя частота контактов внутри ТАДа, рассчитанная для карт Hi-C после нормализации на ожидаемую частоту взаимодействий. Пересечение

разметок ТАДов в контроле и при деплеции Elys, а также пересечение с разметками состояний хроматина и ЛАДов проводилось с помощью программы bedtools [111].

Оценка средней интенсивности взаимодействий ТАДов в А- и В-компарменте проводилась для двух групп: активные и неактивные. Однако в этом разделе исследования «активность» ТАДов определялась несколько отличным от указанного выше способа, а именно через значение PC1. На первом этапе были отобраны только те ТАДы, внутри которых не наблюдалось изменения знака первой главной компоненты, т. е. относящиеся полностью к А- или В-компарменту. Затем к первой группе были отнесены такие ТАДы, у которых среднее значение PC1 превосходило 25%-перцентиль, посчитанный для средних значения этой величины в активных ТАДах, а во вторую, напротив, такие ТАДы, у которых PC1 оказывалось меньше 75%-перцентиля, рассчитанного для ТАДов из В-компармента. Т. е. из последующего анализа исключались ТАДы, не имеющие выраженного активного или неактивного статуса, равно как и ТАДы размером менее 8 тыс.п.о. Таким образом, число неактивных ТАДов составило 56 для данных по деплеции Elys и 60 для контроля, число активных – 1171 и 1191 соответственно. Средняя частота контактов ТАДов подсчитывалась на разрешении 4 тыс.п.о. в трех диапазонах расстояний, полученных разбиением пар ТАДов на основе следующих перцентилей дистанции между центрами ТАДов: 0, 33, 66, 100. Т. о. были получены следующие границы интервалов расстояний для взаимодействия ТАДов: [202000,4458000,9960000,27850000] для активных и [202000,3036000,9216000,20482000] для неактивных ТАДов.

2.2.3 Анализ локальной плотности хроматина

С целью изучения мелкомасштабных различий в структуре хроматина, как на разных стадиях сперматогенеза, так и между контрольными данными и после деплеции Elys, с помощью модуля insulation из пакета cooltools была рассчитана мера локальной плотности упаковки хроматина, которая определялась как среднее значение контактов бинов, расположенных по разные стороны от исследуемого участка генома на некотором фиксированном расстоянии, включая взаимодействие с ними самого центрального участка, с последующим нормированием на среднее

значение этой величины по всей хромосоме. Полученная таким способом величина называется индексом инсуляции (от англ. insulation score) и отражает степень компактизации хроматина вокруг каждого бина генома в масштабах заданной пользователем окрестности: чем меньше значение параметра окна, отражающего размер этой окрестности, тем более локальные изменения фиксирует индекс инсуляции.

В разделе, посвященном динамике пространственных взаимодействий в сперматогенезе, анализ локальных изменений касался главным образом плотности хроматина в области стартов генов из разных групп. Для полноты картины индекс инсуляции оценивался не только в области промоторов, но и вокруг них (± 12 тыс.п.о. для данных по сперматогенезу и ± 16 тыс.п.о. для эксперимента с деплецией Elys). Было протестировано несколько вариантов размера окна и разрешений. Значения параметров, при которых были обнаружены наиболее яркие отличия, соответствовали 12 тыс.п.о. для окна инсуляции и 2 тыс.п.о. для разрешения. Оценка инсуляции, полученная для каждого бина с указанными выше параметрами, использовалась затем для построения кривых инсуляции. Погрешность значений на рисунке 10 - это 95%-ный доверительный интервал. При построении кривой инсуляции вокруг стартов сперматоцитспецифичных генов, во избежание влияния экспрессии соседних участков генома, учитывались только те гены, вокруг которых на расстоянии 3-х бинов не содержалось никаких других стартов транскрипции. При отображении кривых инсуляции вокруг стартов транскрипции всех генов (см. далее рис. 176 в разделе 3.1.3) из каждой точки графика вычиталось минимальное значение кривой, что позволило наглядно отобразить различия в локальной плотности.

При исследовании влияния деплеции Elys на структуру хроматина оценка локальных взаимодействий была актуальна прежде всего для выявления изменений в окрестности сайтов связывания Elys_{нукл} и Elys_{япк}. Для данного вида анализа была реализована модифицированная версия подсчета меры инсуляции, а именно при оценке локальной плотности не учитывались взаимодействия между самим бином и его окружением, а только проходящие через него контакты. Введение

модификации было обусловлено необходимостью разделить локальную плотность на две составляющие: опосредованную самим взаимодействием Elys с участком посадки на ДНК и описывающую коммуникацию прилежащих к нему участков через сайт связывания. Расчет проводился на разрешении 2 тыс.п.о., с размером окна 8 тыс.п.о.

Все представленные в работе кривые инсуляции были получены путем усреднения профилей вокруг таргетных позиций с дальнейшим сглаживанием методом В-сплайна третьей степени, реализованным в функции `make_interpolate_spline` модуля `interpolate` из пакета `scipy`.

2.2.4 Определение сильных и слабых границ ТАДов в контрольных клетках S2

Помимо вызванных деплецией Elys изменений в структуре ТАДов и компартментов, она могла также существенно повлиять на плотность хроматина в границах. Анализ изменения частоты контактов в данном контексте проводился для общих между деплецией и контролем границ (± 5 тыс.п.о.), разделенных на две группы: сильные и слабые. Для отбора границ в одну из двух групп был использован индекс инсуляции, определенный в каждом 4- тыс.п.о. бине с величиной окна 24 тыс.п.о. для карт Hi-C в контроле. Для каждой границы была подсчитана разница между средней инсуляцией в трех последовательно расположенных 4- тыс.п.о. бинах, находящихся на расстоянии 12 тыс.п.о. слева от границы, и значением инсуляции в самой границе (X1), а также аналогичная величина для бинов, расположенных симметрично справа от границы (X2). В группу сильных отбирались такие границы, у которых как слева, так и справа значения X1 и X2 превышали соответствующие им медианные. В группу слабых, напротив, попадали те границы, у которых разница, как слева, так и справа оказывалась меньше рассчитанных медианных значений. Таким образом, в первую группу попало 973, а во вторую - 972 границы (рис. 2 Приложения). Для обоих случаев были построены средние кривые для пропорции $Elys_{\text{нукл}}$ и $Elys_{\text{янк}}$ в границе и прилегающих к ней регионах, для чего в каждом 2- тыс.п.о. бине была определена доля, приходящаяся на пик сигнала DamID-seq соответствующей фракции Elys.

Сглаживание средних профилей и построение доверительного интервала проводилось описанным выше методом, который применялся при построении кривых инсуляции (2.2.3).

2.2.5 Анализ парных взаимодействий в группах генов

В ходе исследования структуры хроматина на ранних стадиях сперматогенеза для изучения взаимосвязи между активацией транскрипции и пространственной близостью соответствующих промоторов был проведен анализ частот контактов в группах сперматоцитспецифичных, семенникспецифичных (т. е. экспрессирующихся в семенниках на обеих стадиях), убиквитных и молчащих ($TRM < 1$) на стадии сперматоцитов генов. Ген классифицировался как убиквитный, т. е. повсеместно экспрессирующийся, если его транскрипция наблюдалась в широком спектре тканей и органов (4228 генов), и как специфичный для семенников, если он экспрессировался только в этой ткани. Среди всех семенникспецифичных генов в особую группу сперматоцитспецифичных были выделены гены, активные исключительно на стадии сперматоцитов (1014 генов). Определение групп генов проводилось А. А. Ильиным из научной группы Ю.Я. Шевелева [100].

Область детекции контактов была разбита на два подмножества: контакты генов внутри ТАДов и контакты генов из разных ТАДов. Во втором случае учитывались взаимодействия генов, принадлежащих к одному и тому же компартменту, т. е. рассматривались только те события, когда оба гена лежат в А-либо в В-компартменте. Такой способ подсчета обусловлен преимущественной сегрегацией биологически осмысленных контактов в два компартмента, в то время как контакты генов из разных компартментов скорее являются результатом случайного лигирования.

При анализе сперматоцитспецифичных взаимодействий на уровне трансконтактов было выбрано разрешение 10 тыс.п.о., карты межхромосомных взаимодействий были нормированы на среднее значение контактов, индивидуально для каждой пары хромосом. Из расчетов были исключены контакты

тех бинов, которые помимо стартов сперматоцитспецифичных генов содержали также гены с экспрессией более 1 ТРМ на стадии сперматогониев.

Детекция значимых контактов, где одно из оснований содержало сперматоцитспецифичный ген, проводилась методом, описанным в [112]. А именно, старты всех сперматоцитспецифичных генов были отнесены к 5-тыс.п.о. бинам генома, после чего значимость контакта каждого такого бина с каждым из расположенных на расстоянии не менее 10 тыс.п.о. и не более 5-Мб, определялась через сравнение с распределением частот контактов всех бинов, разделенных таким же линейным расстоянием. Для аппроксимации применялось экспоненциально-нормальное распределение, на этапе подбора параметров исключалось 5% контактов с самой высокой интенсивностью. На основе полученных параметров распределения вычислялась вероятность обнаружить такие же или более интенсивные контакты. Затем ко всем значениям вероятности применялась поправка на множественное тестирование (метод Бенджамини-Хохберга), и после чего по полученным q-значениям контакты отбирались для дальнейшего анализа. Процедура коррекции применялась по отдельности для каждой из хромосом. Средняя частота контактов взаимодействующих пар сперматоцитспецифичных генов, выявленная этим методом, представлена на рисунке 3 Приложения. Статистическая оценка наличия ассоциации между формированием значимого контакта сперматоцитспецифичным геном и образованием пары с другим геном из этой же группы была получена на основе применения критерия хи-квадрат Пирсона к Таблице 1.

Таблица 1. Таблица сопряженности для подсчета значимости ассоциации между образованием сперматоцитспецифичным геном значимого, отличного от фонового, контакта и частотой попадания в такие контакты в качестве пары другого гена из той же группы сперматоцитспецифичных при сравнении с убиквитными.

	Число пар сперматоцитспецифичных генов	Число пар, образованных сперматоцитспецифичным и убиквитным генами
Число значимых контактов, вовлекающих по крайней мере один сперматоцитспецифичный ген, разделенных расстоянием более 10 тыс.п.о. и менее 5 Мб	744	4828
Число пар всех контактов, вовлекающих по крайней мере один сперматоцитспецифичный ген, разделенных расстоянием более 10 тыс.п.о. и менее 5 Мб, за исключением значимых контактов сперматоцитспецифичных генов	33815	211854

Анализ семантической близости пар сперматоцитспецифичных генов, расположенных в основаниях значимых контактов, проводился с помощью программы GOSemSim [113] для каждой из онтологий (CC - клеточные компоненты, MF - молекулярные функции, BP - биологические процессы) с параметрами по умолчанию. Для оценки меры среднего сходства между генами, объединенными значимыми контактами, применялся пермутационный тест (100 пермутаций), в ходе которого партнеры в парах перемешивались случайным образом. Подсчет среднего сходства при пермутациях проводился на основе только тех пар, которые не попадали в список исходных, не перемешанных, пар. Аналогичным образом была получена статистическая оценка семантической близости генов, расположенных в кластерах сперматоцитспецифичных генов.

2.2.6 Детекция петель

Для исследования влияния деплеции Elys на структуру хроматина среди прочих характерных структур были проанализированы петли и значимые контакты. Петли были детектированы на разрешении 10 тыс.п.о. с ограничением на максимальное расстояние между основаниями петли в 1 Мб отдельно для контроля и деплеции Elys. Затем все петли были слиты в единое множество уникальных петель. Полученное множество петель было аннотировано с точки зрения наличия сайтов связывания Elys_{нукл} и Elys_{янк}. Были проанализированы петли, содержащие хотя бы в одном из оснований сайт связывания Elys_{нукл}, а также петли, содержащие хотя бы в одном из оснований сайт связывания Elys_{янк}. Также отдельно было выделено подмножество петель, содержащих сайты связывания Elys_{нукл}, которые пересекались с разметкой энхансеров, полученной ранее в [114] методом STARR-seq. При исследовании энхансерно-промоторных взаимодействий связка ген-энхансер была получена по принципу ближайшего расположенного к энхансеру старта гена, который лежит далее по ходу прямой цепи или ранее в случае обратной цепи. Расчеты проводились на разрешении 2 тыс.п.о., только для тех энхансер-промоторных пар, где расстояние между стартом транскрипции и энхансером составляло более 10 тыс.п.о.

2.2.7 Анализ данных по экспрессии

Публично доступные, непроцессированные, данные RNA-seq в клеточной линии Kc167 были закартированы на транскриптом дрозофилы с помощью программы salmon [115], после чего для подсчета значений TPM применялся пакет tximport [116]. Полученные значения TPM затем использовались для оценки экспрессии в неактивном компартменте различных клеток дрозофилы (раздел 3.1.1).

2.2.8 Анализ функционального обогащения

Анализ обогащения категорий генной онтологии (GO-категории) проводился для генов, расположенных в В-компарменте как в сперматогониях, так и в сперматоцитах, и характеризующихся значением PC1<-0.36 и PC1<-0.34

соответственно. Для этой цели была использована программа clusterProfiler. Поскольку в результате было получено большое число обогащенных, связанных между собой терминов, к ним была применена функция simplify, которая оставляет только те из них, что имеют наилучшее Q-значение, с параметром меры “Lin”, порогом на сходство 0.7 и числом попавших в категорию генов (из анализируемого списка) не менее 5.

2.2.9 Построение средних тепловых карт

Одним из вариантов иллюстрации различных форм организации хроматина (петли, значимые контакты, ТАДы, границы) является построение средних тепловых карт. В данной работе этот метод использовался для визуализации контактов между стартами генов в разделе, посвященном сперматогенезу, а также изменений в структуре хроматина в границах ТАДов и вокруг сайтов связывания Elys в результате его деплеции. Средние тепловые карты вокруг целевых участков либо контактов получены путем «вырезания» фрагментов Hi-C карты вокруг исследуемой области и усреднения значений в каждой точке такой «вырезанной» карты Hi-C по всему набору фрагментов. При подсчете средних пропущенные значения игнорировались.

2.2.10 Оценка различий в графике шкалирования

Для статистической оценки различий в графике шкалирования между сперматогониями и сперматоцитами был выбран участок от 100 до 315 тыс.п.о. ($\log_{10}5 - \log_{10}5.5$), на котором визуально наблюдается изменение скорости убывания средней частоты контактов по мере увеличения расстояния. С помощью линейной аппроксимации для каждой из аутосом (chr2L, chr2R, chr3L, chr3R) на обеих стадиях был рассчитан угол наклона прямой, и затем к четырем парам коэффициентов применялся парный тест Вилкоксона.

2.2.11 Публичные данные и использованные программы

Для целей исследования были также привлечены публичные данные, перечисленные в Таблице 2.

Таблица 2. Публичные данные.

	Источник	Тип файла
Данные RNA-seq, списки убиквитных и сперматоцитспецифичных генов	[100]	Таблица с значениями TPM
Данные RNA-seq в Kc167	GSE15596	fastq
Данные Hi-C для линии Kc167	GSE101317	fastq
Список Aiy-зависимых генов	[100]	bed
Кластеры сперматоцитспецифичных генов	[101]	Список генов
Сайты связывания Elys в нуклеоплазме и на ЯПК	[14]	bed
Разметка ЛАДов (клетки S2)	[10]	bed
Разметка цветов хроматина (клетки S2)	[57]	bed
Транскрипция в контроле (клетки S2)	[14]	bedgraph
Разметка энхансеров	[114]	bed

Все программы, которые применялись в ходе исследования, перечислены в Таблице 3.

Таблица 3. Использованные программы и модули.

	Название программы и модуля
Контроль качества данных	FastQC
Обрезание адаптеров и фильтрация чтений по качеству	Trim galore!
Генерация карт Hi-C	Distiller-nf
Расчет сходства карт Hi-C	HiCRep
Объединение карт Hi-C	cooler
Уменьшение общего числа контактов	Cooltools random-sample

Получение первой главной компоненты	Cooltools.eigs_cis
Построение «седлового графика»	Cooltools.saddle
Расчет локальной плотности	Cooltools.insulation
Поиск петель	Cooltools.dots
Поиск ТАДов	Armatus
Картирование данных RNA-seq	salmon
Подсчет уровня транскрипции в TPM	tximport
Анализ семантического сходства	GOSemSim
Анализ обогащения (GO)	clusterProfiler
Пересечение геномных интервалов	bedtools
Дополнительный анализ	базовые и научные библиотеки R, python.

Глава 3. Результаты и обсуждение

3.1 Изменение структуры хроматина при активации транскрипции сперматоцитспецифичных генов на ранних стадиях сперматогенеза дрозофилы

Для исследования динамики трехмерной организации хроматина в процессе сперматогенеза были построены Hi-C карты с максимальным разрешением 2 тыс.п.о. в семенниках *bamΔ86* мутантов, содержащих в основном сперматогонии, и в семенниках личинок, содержащих преимущественно сперматоциты. Расстояние между репликами карт Hi-C, относящимися к одной стадии, оказывается меньше, чем для реплик из разных стадий, что отражено на дендрограмме (рис. 9а). На внутриврохромосомных картах просматриваются характерные для данного вида формы организации хроматина, такие как ТАДы, компартменты, петли (рис. 9б,г,д). Обогащение внутриврохромосомными, в особенности межплечевыми, контактами по сравнению с межврохромосомными (рис. 9в) говорит о наличии врохромосомных территорий на обеих стадиях в соответствии с общими представлениями о структуре хроматина многоклеточных организмов. В целом

первичный анализ карт контактов демонстрирует хорошее, подходящее для целей исследования, качество полученных данных.

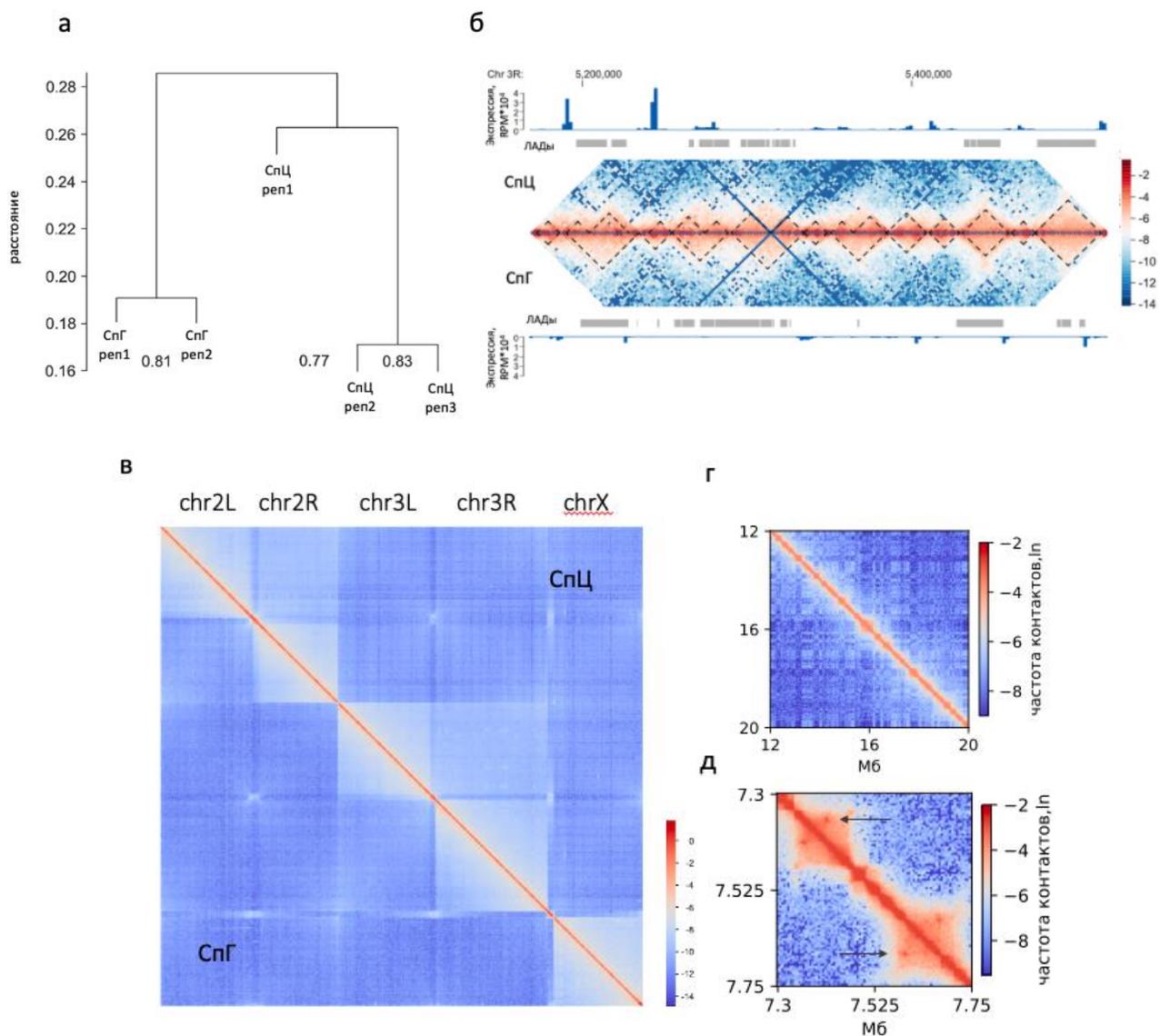


Рисунок 9. Данные Hi-C для сперматогониев и сперматоцитов.

(а) Кластеризация реплик. (б) Фрагменты Hi-C карт, демонстрирующие ТАДы. (в) Полногеномные карты контактов в сперматоцитах и сперматогониях. (г) Пример компартментной структуры. (д) Пример петли.

3.1.1 Изменение структуры компартментов

Данные литературы говорят о том, что транскрипция, в т. ч. у дрозофилы, тесно связана с пространственной локализацией в А- или В-компартменте и их структурой в целом. В качестве меры выраженности принадлежности к А- или В-компартменту традиционно используется значение PC1 (см. раздел 2.2.1.1).

Переход от стадии сперматогониев к стадии сперматоцитов сопровождался повышением значения PC1 для участков генома, содержащих старты сперматоцитспецифичных генов, тогда как для убиквитных генов эта величина не менялась значимо между стадиями (рис. 10).

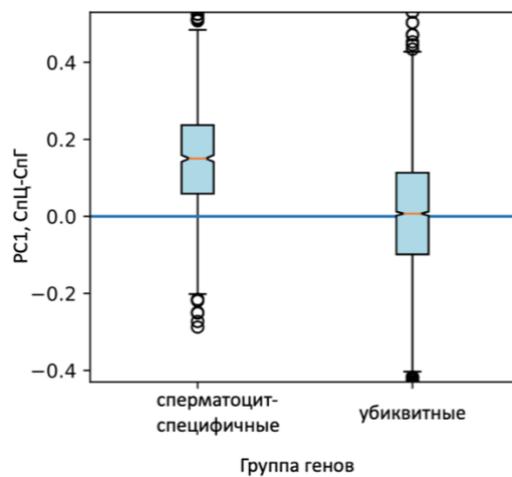


Рисунок 10. Разность значений PC1 между сперматоцитами и сперматогониями.

Повышение значений PC1 указывает на полное или частичное перемещение сперматоцитспецифичных генов в область активного компартмента при запуске их транскрипции. Для 11% генов из этой группы наблюдался непосредственно переход из В- в А-компаратмент (т. е. имело место изменение знака PC1). В целом на долю событий смены компартмента с неактивного в сперматогониях на активный в сперматоцитах приходится около 7.5% генома, из которых 25% содержит сперматоцитспецифичные гены, изменения в противоположном направлении обнаруживаются в ~2.3% генома.

Интересно, что для 42% сперматоцитспецифичных генов, несмотря на активацию и повышение значения PC1 (P -значение $< 10^{-60}$), перехода в А-компаратмент не наблюдалось, т. е. существуют какие-то дополнительные факторы, ограничивающие или разрешающие такие события. Одним из них, предположительно, может быть близость региона к границе ТАДа. Действительно, расстояние до ближайшей границы, как и размер ТАДа, при сравнении сперматоцитспецифичных генов, перешедших из В- в А-компаратмент и

оставшихся в В, оказались меньше для первой группы (P -значение $< 10^{-14}$, рис. 11а). При этом уровень экспрессии сперматоцитспецифичных генов не связан с наличием или отсутствием перехода в активный компартмент (P -значение ~ 0.3 , рис. 11б). Доля убиквитных генов, приходящихся на В-компаратмент, не достигает 3% ни на одной из стадий, что соотносится с их преимущественным расположением в границах ТАДов.

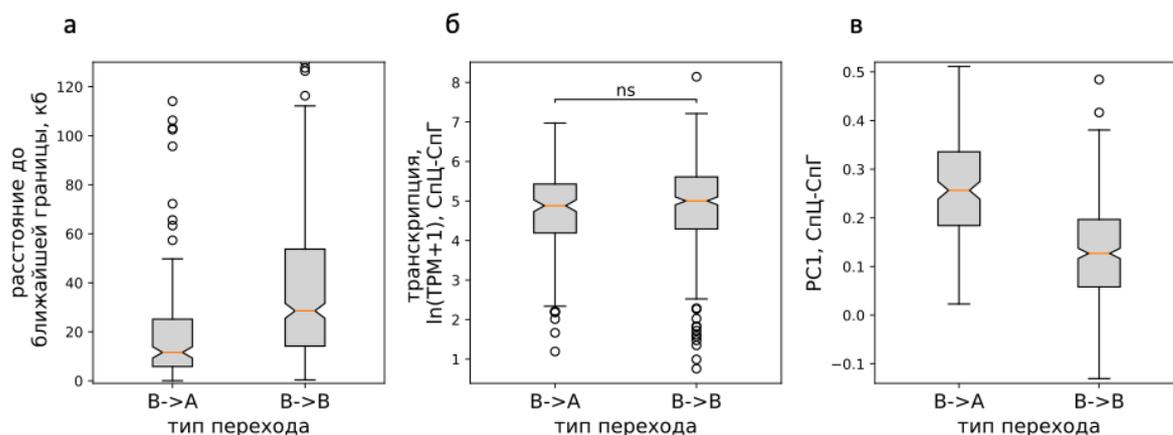


Рисунок 11. Сравнение характеристик регионов, перешедших в активный компартмент (В-А) и оставшихся в неактивном (В-В).

(а) Расстояние до ближайшей границы. (б) Изменение уровня транскрипции. (в) Изменение значения PC1.

Ранее, в работах по структуре хроматина было показано, что, в отличие от млекопитающих, неактивный компартмент дрозофилы выражен довольно слабо: на интенсивные В-В взаимодействия приходится малая часть генома [47]. Этот вывод был сделан на основе анализа Hi-C в клетках S2 и Kc167. Однако, согласно данным Hi-C для сперматогониев и сперматоцитов (а эти клетки соответствуют более поздним стадиям дифференцировки), протяженность и сила неактивного компартмента у данного вида может варьироваться (рис. 12). Так, для обеих исследуемых стадий сперматогенеза взаимодействия между регионами с отрицательным значением PC1, превышающие ожидаемые частоты контактов, распространяются на большую долю генома по сравнению с Kc167 (рис. 12а-г).

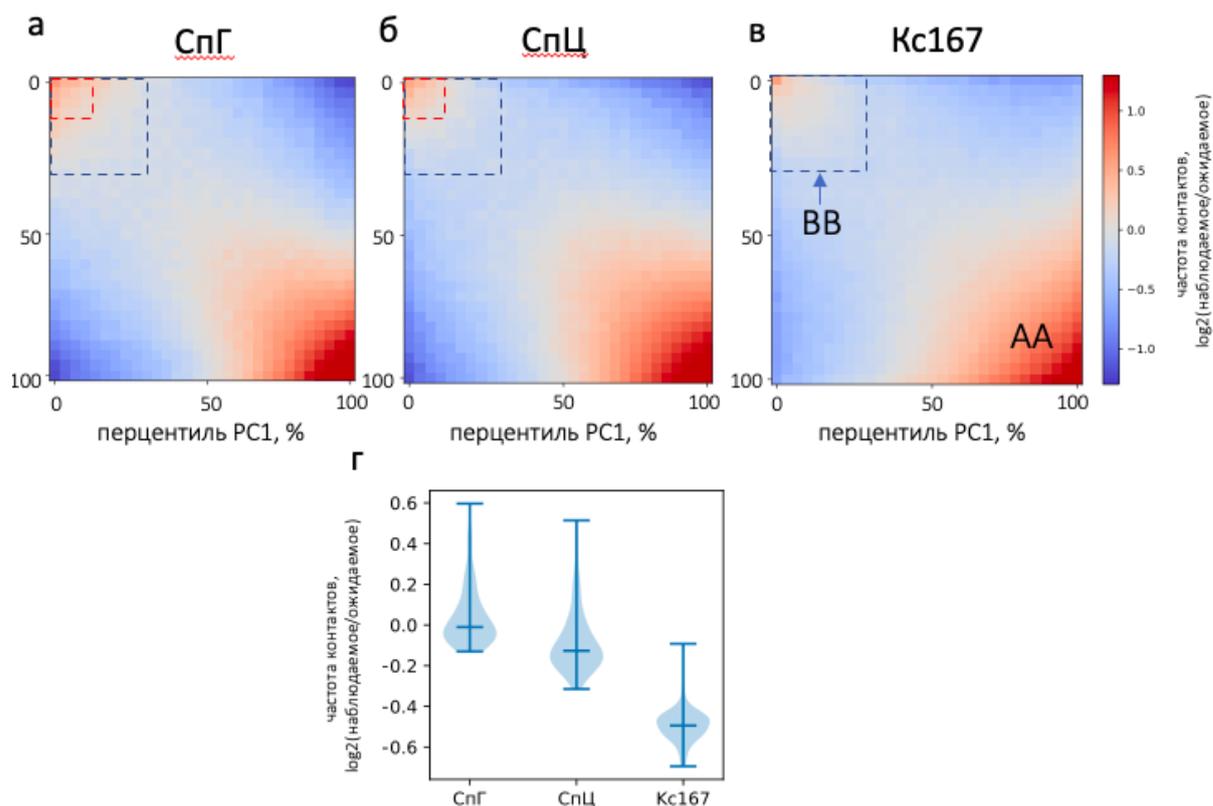


Рисунок 12. Структура компартментов.

(а) Сперматогонии. (б) Сперматоциты. (в) Клеточная линия Кс167. (г) Частоты контактов в неактивном компартменте для области, очерченной синей пунктирной линией.

При этом, вопреки ожиданиям, при сопоставлении транскрипции в регионах, вовлеченных на обеих стадиях сперматогенеза в формирование выраженного В-компартмента (рис. 12а,б, красная пунктирная линия), ее значения оказываются в Кс167 значимо меньше по сравнению со стадией сперматогониев (P -значение = 0.023), а для стадии сперматоцитов эти участки генома и вовсе оказываются обогащенными активными, специфичными для данной стадии, генами (P -значение $< 10^{-05}$). Таким образом, экспрессия генов хотя и коррелирует с их расположением в А- или В-компартменте, все же не является единственным определяющим фактором ни для принадлежности региона к тому или иному компартменту, ни для выраженности последнего.

Для генов, которые располагаются в участках из В-компартмента как в сперматогониях, так и в сперматоцитах (в пределах красной пунктирной линии на рис. 12а,б) был проведен анализ функционального обогащения (рис. 4

Приложения). Среди наиболее обогащенных ГО-терминов встречаются связанные с сенсорным восприятием химических стимулов, клеточной адгезией (по-видимому, ненужной для половых клеток, не формирующих ткань) и высшей нервной деятельностью (также не свойственной половым клеткам).

3.1.2 Взаимодействия сперматоцитспецифичных генов

Изменение структуры компартментов в ходе сперматогенеза предполагает перестройку на уровне отдельных парных взаимодействий. Для подробного исследования взаимосвязи между этой формой организации хроматина и активацией сперматоцитспецифичной транскрипции был проведен анализ частот контактов стартов генов из четырех групп: сперматоцитспецифичных, семенникспецифичных, неактивных и убиквитных (см. описание групп в разделе 2.2.5). Результаты анализа говорят о том, что активация сперматоцитспецифичных генов сопровождается повышением частот контактов их стартов транскрипции (P -значение $\sim 10^{-39}$,) (рис. 13, 14). Данную зависимость удалось показать не только для генов, расположенных внутри одного и того же ТАДа, но и относящихся к разным ТАДам. При этом частота контактов неактивных генов не отличалась от фоновой и различий между стадиями обнаружено не было (P -значение ~ 0.8), тогда как для убиквитных генов наблюдалось выраженное повышение интенсивности взаимодействий на обеих стадиях. Картина, наблюдаемая для убиквитных генов, является вполне закономерной: гены из этой группы преимущественно располагаются в малых активных ТАДах (или на более грубом разрешении – в границах), и яркий контакт, распространяющийся за пределы бинов, содержащих исследуемые старты генов, - не что иное, как описанная ранее кластеризация активных участков хроматина в трехмерном пространстве, напрямую связанная с формированием А-компартамента [15]. Семенникспецифичные гены также сохраняют сходный уровень частот контактов на обеих стадиях (P -значение ~ 0.6). Повышение частот контактов двух сперматоцитспецифичных генов, расположенных в одном ТАДе, было также подтверждено экспериментом 3С для региона 60D С. В. Ульяновым [100].

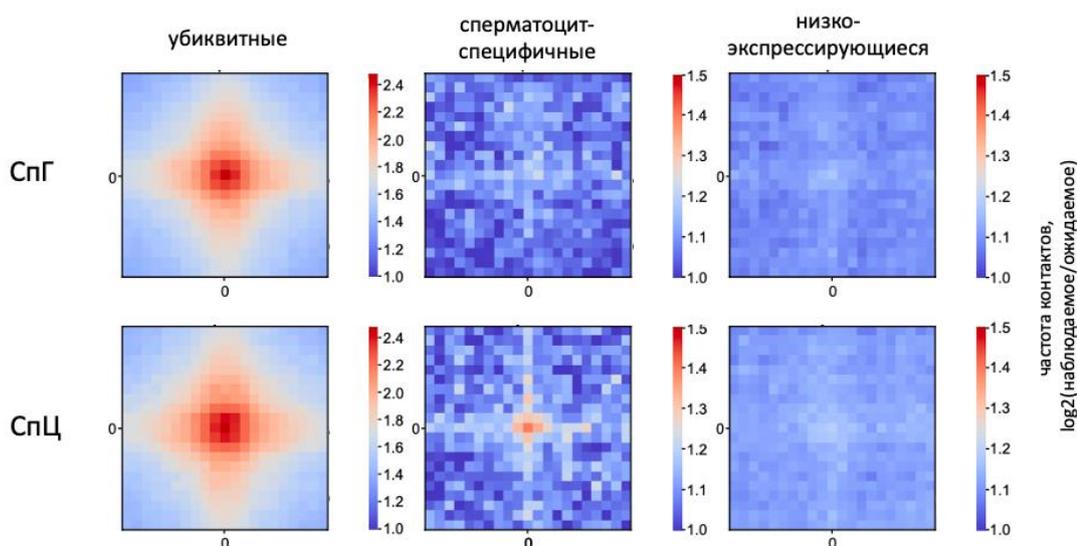


Рисунок 13. Попарные взаимодействия генов из разных ТАДов.

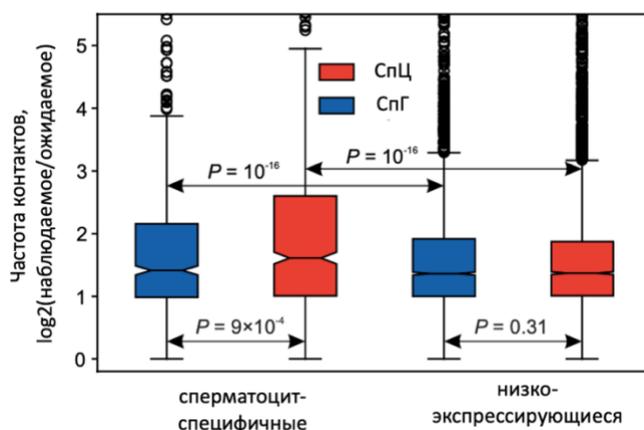


Рисунок 14. Попарные взаимодействия генов внутри ТАДов.

Эта закономерность проявляется и для транс-контактов (рис.15, P-значение $< 10^{-20}$). Однако в данном случае на стадии сперматогониев также обнаруживается повышение частот контактов по сравнению с окружающими фоновыми взаимодействиями. Наблюдаемое обогащение контактами может объясняться формированием структуры хроматина, которое предшествует и способствует дальнейшей активации транскрипции.

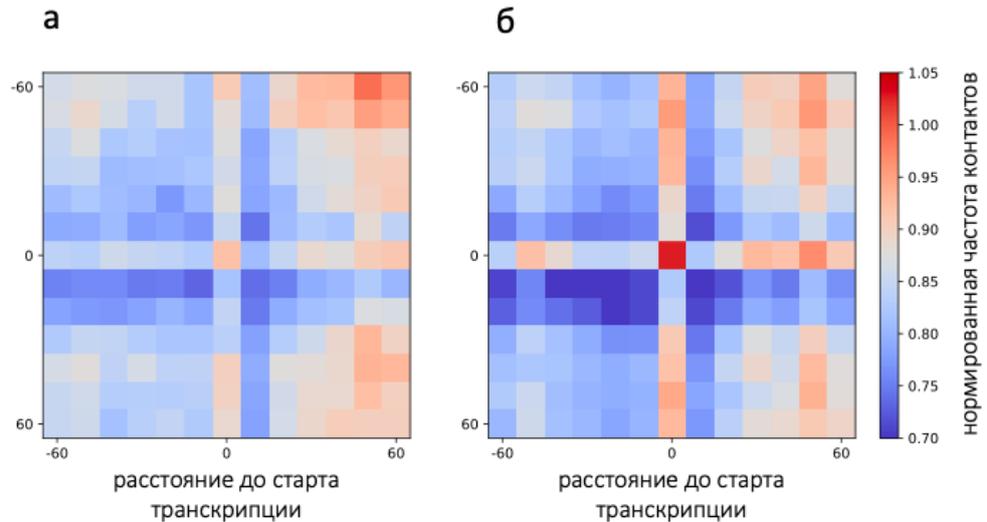


Рисунок 15. Межхромосомные взаимодействия сперматоцитспецифичных генов.

(а) Стадия сперматогониев. (б) Стадия сперматоцитов.

При более подробном исследовании взаимного расположения стартов сперматоцитспецифичных генов в трехмерном пространстве обнаруживается, что повышение частоты их колокализации, наблюдаемое на стадии сперматоцитов, обусловлено главным образом контактами в области А-компартамента (рис. 16), а это наблюдение, в свою очередь поднимает вопрос о специфичности таких взаимодействий. Проявляют ли эти гены избирательность в формировании контактов по принципу «пространственной близости с себе подобными»? Существует ли взаимосвязь между повышением частоты контакта и функциональным сходством лежащих в его основаниях сперматоцитспецифичных генов?

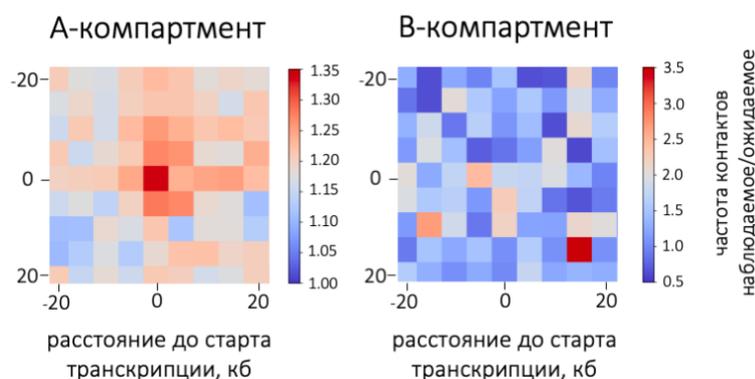


Рисунок 16. Взаимодействия сперматоцитспецифичных генов в А- и В-компарменте на стадии сперматоцитов.

Для ответа на эти вопросы, на основе анализа Hi-C карты, соответствующей стадии сперматоцитов, были выявлены контакты, содержащие в одном из оснований сперматоцитспецифичный ген и характеризующиеся статистически значимым повышением интенсивности по сравнению со всеми остальными контактами, которые были образованы геномными локусами, разделенными тем же линейным расстоянием. В результате было получено около 13 тыс. пар взаимодействующих генов, из которых затем было выделено ~700 пар контактов двух сперматоцитспецифичных генов и ~4800 пар контактов между сперматоцитспецифичным и убиквитным геном. Сравнение с числом потенциально возможных пар генов двух вышеуказанных типов (см. раздел 2.2.5) указывает на отсутствие предпочтений в формировании контактов сперматоцитспецифичных генов с себе подобными: так же часто они формируют контакты и с убиквитными генами (P -значение = 0.38). Более того, результаты анализа семантического сходства взаимодействующих друг с другом сперматоцитспецифичных генов говорят в данном случае об отсутствии взаимосвязи между пространственной колокализацией и функциональным сходством (P -значение > 0.05 для всех трех онтологий: биологические процессы, молекулярные функции, клеточные компоненты). Интересно, что иная картина обнаруживается для сперматоцитспецифичных генов, организованных в кластеры на основе линейной близости в геноме. Семантическое сходство между парами генов, относящихся к одному кластеру, оказывается значимо выше, чем между генами из разных кластеров (пермутационное P -значение = 0.01 для каждой из онтологий). Примечательно, что только 18% кластеров сперматоцитспецифичных генов разбивается границей, тогда как наибольшая их часть располагается в пределах одного ТАДа. Таким образом, общая регуляция сперматоцитспецифичных генов, объединенных сходными функциями, по-видимому, обеспечивается в том числе за счет их последовательного расположения в геноме.

3.1.3 Динамика границ ТАДов на ранних стадиях сперматогенеза дрозофилы

В литературе по структуре хроматина дрозофилы большое внимание уделено исследованию границ ТАДов, в том числе в контексте влияния транскрипции на их формирование. В большинстве работ показано, что одного только этого фактора недостаточно для образования или исчезновения границ. Активация сперматоцитспецифичных генов на ранних стадиях сперматогенеза является естественным экспериментом для исследования в данном направлении. Для выровненных по суммарному количеству контактов карт было детектировано ~3800 ТАДов со средней длиной 35 тыс.п.о. на каждой из стадий, что соотносится в описанными в литературе параметрами для данных Hi-C высокого разрешения [60]. При сравнении разметки ТАДов для обеих стадий было обнаружено, что около 90% границ остаются неизменными с точностью до одного 5- тыс.п.о. бина. Для границ, специфичных для стадии сперматоцитов и приходящихся на ТАДы в сперматогониях, была выявлена значимая ассоциация с содержанием в них сперматоцитспецифичных генов при сравнении с остальными границами (P-значение = 0.005), убиквитные гены, напротив, значимо реже встречались в таких границах по сравнению со всеми прочими (P-значение $< 10^{-16}$). Также для границ этого типа характерно небольшое повышение транскрипции в сперматоцитах по сравнению со сперматогониями, однако оно не является статистически значимым (P-значение ~ 0.08). При этом границы, детектированные на стадии сперматоцитов, и приходящиеся на ТАДы в сперматогониях, содержат только 1.2% сперматоцитспецифичных генов и еще 9% расположено в общих для обеих стадий границах. При визуализации участков карт Hi-C в окрестности новых границ, содержащих сперматоцитспецифичные гены, оказалось, что часть из них вызывает сомнения относительно статуса границы как такового, а для других локальная структура визуально не отличается от наблюдаемой в сперматогониях или рядом с границей обнаруживаются столбцы и строки с пропущенными значениями, т. е. такие события в целом могут являться артефактами работы алгоритма Armatus (рис. 5 Приложения). Таким образом, проведенный анализ говорит скорее об

отсутствии ассоциации между активацией сперматоцитспецифичных генов и формированием новых границ, более того подавляющее большинство генов из этой группы расположено внутри ТАДов.

Согласно общепринятым представлениям о структуре хроматина дрозофилы, существует отрицательная корреляция между удаленностью участка генома от границы и его активностью [8]. В таком случае становится не вполне понятно, каким образом оказывается возможной высокая и стабильная экспрессия сперматоцитспецифичных генов. Для того, чтобы попытаться ответить на этот вопрос, мы провели дополнительный анализ мелкомасштабных изменений вокруг стартов транскрипции. В результате данного анализа было обнаружено, что при активации сперматоцитспецифичных генов происходит локальное разрыхление хроматина в радиусе 5-6 тыс.п.о. от промотора гена (рис. 17а). Кроме того, разница в компактизации напрямую связана с изменением транскрипции: активация генов, приводящая к более выраженному увеличению транскрипции, соотносится с более сильным разрыхлением, и наоборот (рис. 17б, коэффициенты корреляции Спирмена для стадии сперматогониев и сперматоцитов равны 0,89 и 0,99, Р-значение = 0,037 и Р-значение $< 10^{-24}$, соответственно). Для убиквитных генов наблюдается ожидаемо более сильный провал в кривой инсуляции, поскольку эти гены преимущественно располагаются в границах ТАДов (Р-значение $< 10^{-4}$, пермутационный тест), где по определению хроматин характеризуется наименьшей компактизацией. Для убиквитных генов различий в локальной плотности между двумя стадиями не обнаружено.

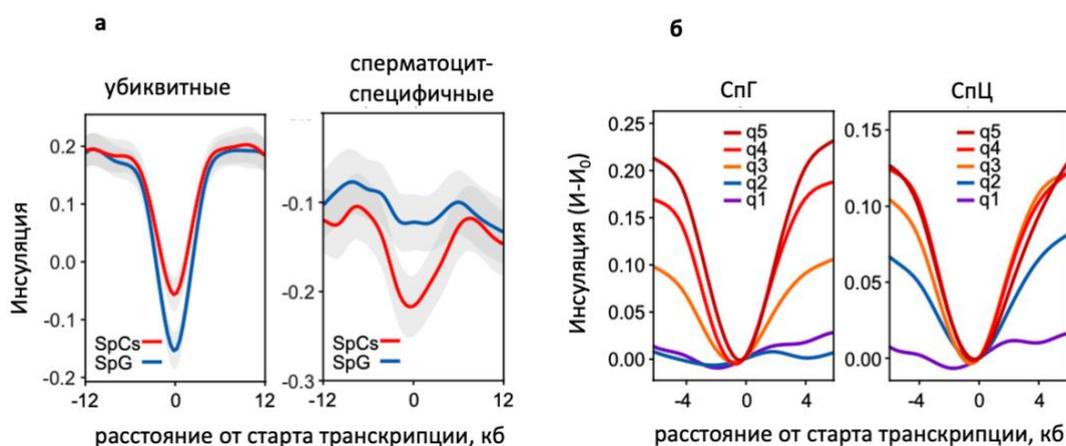


Рисунок 17. Кривые инсуляции вокруг стартов транскрипции.

(а) Группы убиквитных и сперматоцитспецифичных генов. Погрешность значений - это 95%-ный доверительный интервал. (б) Все гены, поделенные на квантили экспрессии.

Активация сперматоцитспецифичных генов регулируется целым набором белков и, разумеется, происходит не одномоментно. Среди таких ко-активаторов особенно выделяется белок - продукт гена *aly*, который является основным компонентом комплекса tMAC, необходимым для второй волны активации сперматоцитспецифичной экспрессии, поскольку он вовлечен в активацию большего числа генов по сравнению с прочими участниками комплекса [117,118]. *Aly*-независимые гены, по-видимому, экспрессируются раньше по сравнению с остальными сперматоцитспецифичными генами [119]. При сравнении распределения *aly*-зависимых и независимых сперматоцитспецифичных генов относительно границ ТАДов была выявлена слабая, но значимая ассоциация между локализацией в границах и принадлежностью к группе *aly*-независимых генов (Р-значение = 0.041 для общих границ, Р-значение \sim 0.0068 для всех границ в сперматоцитах). Значение PC1 в сперматогониях для участков генома, содержащих старты *aly*-независимых генов, также оказалась несколько выше, чем для *aly*-зависимых (Р-значение = 0.042). Эта тенденция, хоть и слабовыраженная, может указывать на предрасположенность генов, активирующихся раньше в ходе

развития, находиться в заведомо более транскрипционно выгодном с точки зрения структуры хроматина состоянии.

Обобщение основных результатов параграфа и их обсуждение

Настоящий раздел исследования посвящен изменению структуры хроматина при активации транскрипции сперматоцитспецифичных генов. Поскольку материалом исследования послужили половые клетки двух последовательных стадий сперматогенеза дрозофилы, и молекулярно-генетическое вмешательство было направлено на получение клеток определенной стадии развития (*bam*-мутанты), но не вызывало напрямую индукцию или подавление экспрессии как таковой, в данном исследовании под влиянием транскрипции надо понимать также и влияние потенциально ассоциированных с ней архитектурных белков и транскрипционных факторов, без которых этот процесс вряд ли обходится.

Наблюдаемые изменения проявлялись на нескольких уровнях трехмерной организации. Например, обнаруживались переходы из области неактивного компартмента в область активного (сходные результаты были ранее показаны для других организмов и типов клеток [42,68]), что выражалось также в повышении частот контактов сперматоцитспецифичных генов, в том числе на уровне межхромосомных взаимодействий. Однако такая крупномасштабная трансформация оказалась характерной для участков генома, заведомо предрасположенных к активному состоянию. Экспрессия сперматоцитспецифичных генов, находящихся вдали от границ, внутри ТАДов, сопровождалась переходом хроматина в более активное состояние, однако непосредственно перехода в активный компартмент или появления новых сильных границ выявлено не было. Для стабильно высокой транскрипции в этих регионах оказалось достаточно локальной декомпактизации в окрестности промоторов. Т. е. наличие активной, специфичной для стадии сперматоцитов, экспрессии далеко не всегда соответствует радикальному перестроению структуры хроматина. В то же время, согласно результатам анализа транскрипции в клетках Kc167, сперматогониях и сперматоцитах, ее отсутствие также не всегда сопровождается формированием выраженного В-компартамента. Эти наблюдения поддерживаются

и другими, описанными в литературе, примерами отсутствия изменений в структуре хроматина и при очевидных отличиях в профиле экспрессии мезодермы, нейроэктодермы и дорзальной эктодермы [120]. Таким образом, несмотря на общепринятое представление о существенной роли транскрипции в определении конформации хроматина, по-видимому, множество других факторов оказываются не менее важными в этом контексте. Например, это могут быть белки ламины, а также белки комплекса ядерной поры, удерживающие хроматин на периферии и препятствующие полноценному переходу в А-компаратмент, или белок HP1a, участвующий в формировании гетерохроматина и, как было показано ранее, не пересекающийся с ЛАДа́ми в клеточной линии Kc167 [58]. Последнее обстоятельство может быть напрямую связано с тем, что В-компаратмент выражен в Kc167 слабо, несмотря на низкую транскрипцию в соответствующем регионе. Интересно, что в случае сперматоцитспецифичных генов совместная регуляция, предположительно имеющаяся при сходстве в функциональном отношении, реализуется посредством линейной колокализации в геноме, а не за счет пространственной кластеризации хроматина. Открытым остается вопрос о наличии подобной тенденции у других видов и сравнении с генами, специфичными для других клеточных типов.

Несмотря на то, что данная работа посвящена исследованию факторов, влияющих на структуру хроматина дрозофилы, продемонстрировать влияние транскрипции в одностороннем порядке оказывается несколько затруднительным, поскольку ряд зависимостей носит характер ассоциации, а не причинно-следственной связи. С большей долей уверенности можно интерпретировать те случаи, когда взаимосвязь между активацией транскрипции и изменением структуры хроматина не наблюдалась: примером может служить уже упомянутое отсутствие новых сильных границ в области промоторов сперматоцитспецифичных генов, расположенных в теле ТАДа, что согласуется с текущими представлениями о роли транскрипции в становлении границ. В то же время в ходе исследования был обнаружен пример заблаговременного формирования такой конформации хроматина, которая может соответствовать

более выгодному с точки зрения последующей транскрипции состоянию. Так, для сперматоцитспецифичных генов небольшое повышение частот их транс-контактов наблюдалось уже на стадии сперматогониев. Этот случай, наряду с описанными ранее в литературе [121], предположительно, говорит о непосредственном влиянии самой структуры хроматина на транскрипцию через создание благоприятной для ее активации конфигурации в пространстве ядра.

Наличие новых доступных данных о гистоновых метках и хроматиновых взаимодействиях в половых и соматических клетках [122], а также информации об экспрессии в единичных клетках [123] может лечь в основу продолжения исследования по тематике настоящей работы и способствовать пониманию принципов трехмерной организации ДНК дрозофилы и ее взаимосвязи с регуляцией в целом.

3.2 Структура хроматина и морфология сперматоцитов

В отличие от млекопитающих, у которых в профазе первого мейоза наблюдаются яркие изменения в структуре хроматина, обусловленные в том числе постепенно нарастающей конденсацией, в сперматогенезе дрозофилы, ввиду отсутствия гомологичной рекомбинации у самцов, такого не происходит. В то же время некоторые морфологические изменения, а именно 25-кратное увеличение клетки, влекут за собой увеличение объема ядра, что не может не сказываться на структуре хроматина. Полученные результаты согласуются с общепринятой концепцией о деконденсированном, характерном для интерфазы, хроматине как в сперматогониях, так и в сперматоцитах дрозофилы. Однако между этими двумя стадиями обнаруживаются и некоторые отличия: при исследовании кривой $P(s)$ на расстоянии от ~ 100 до ~ 300 тыс.п.о. наблюдается изменение скорости убывания средней частоты контактов (P -значение=0.014) и далее практически до конца кривой интенсивность взаимодействий остается в сперматоцитах ниже по сравнению со сперматогониями (рис. 18а). Компартменты обоих типов становятся менее плотными и обособленными (рис. 18б). При этом межхромосомные и особенно межплечевые контакты становятся более выраженными (различия между стадиями заметны даже при визуальной оценке карт контактов (рис. 9в, табл. 4).

Также проведенный анализ показывает, что на фоне отсутствия изменений в плотности ТАДов в сперматоцитах наблюдается значимое ослабление силы границ (Р-значение ~ 0.0022), а также петель (Р-значение $< 10^{-57}$, рис. 19). Интересно, что в сперматоцитах в целом детектируется примерно в три раза меньше петель по сравнению со сперматогониями, причем различие в их числе и интенсивности не связано с уровнем экспрессии генов, лежащих в основаниях, поскольку этот фактор был учтен дополнительно.

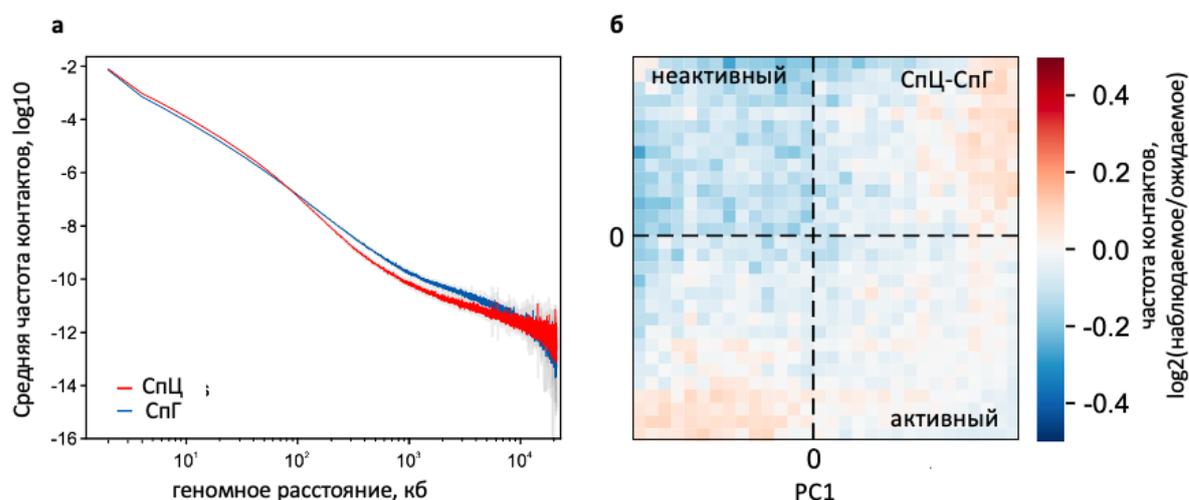


Рисунок 18. Крупномасштабные изменения в структуре хроматина сперматогониев и сперматоцитов.

(а) Средняя частота контактов в зависимости от геномного расстояния. (б) Различия на уровне компартментов.

Таблица 4. Средняя частота межхромосомных взаимодействий, нормированная на общее число контактов.

Выше диагонали – сперматоциты, ниже диагонали – сперматогонии.

	2L	2R	3L	3R	X
2L		1.40	0.44	0.46	0.45
2R	0.67		0.48	0.50	0.49
3L	0.35	0.38		1.25	0.46
3R	0.35	0.38	0.58		0.44
X	0.35	0.38	0.33	0.33	

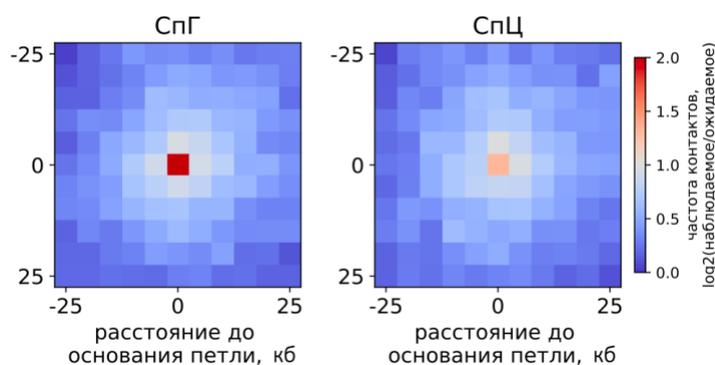


Рисунок 19. Средняя интенсивность контактов для объединения детектированных на каждой из стадий петель, в основаниях которых средняя экспрессия варьируется в пределах 10% между сперматогониями и сперматоцитами.

Учитывая наличие ЛАДов как в сперматоцитах, так и в сперматогониях, этот результат можно интерпретировать как более релаксированное и перемешанное состояние хроматина, связанное с расширением хромосомных территорий на фоне увеличения объема ядра, при сохранении плотности ТАДов и средней частоты контактов на коротких расстояниях за счет взаимодействия с ламиной. Интересно, что данный результат во многом согласуется с результатами работы по искусственному расширению ядра в клетках млекопитающих [21], где было показано уменьшение выраженности компартментов, а также повышение межхромосомных взаимодействий у ассоциированных с ламиной хромосом. При этом повышение дальних внутрихромосомных взаимодействий, наблюдаемое при искусственном увеличении объема ядра на расстоянии более 1 Мб, проявляется в сперматоцитах только с отметки 10 Мб.

Также требует дополнительного исследования анализ петель: в работе [21] было показано ослабление только CTCF-ассоциированных петель, тогда как интенсивность прочих оставалась неизменной при увеличении размера ядра. Как известно, CTCF не является основным фактором, ответственным за формирование петель у дрозофилы, однако в сперматоцитах их интенсивность заметно падает, причем в петлях и с высокой, и с низкой экспрессией. В данном контексте было бы интересно проследить взаимосвязь между изменением силы петли при увеличении объема ядра и обогащением ее оснований различными архитектурными белками,

однако к настоящему времени подобных данных для интересующих нас типов клеток получено не было. Так или иначе для проведения более уверенного сравнительного анализа между естественным экспериментом по увеличению ядра в сперматоцитах и его искусственным раздуванием, необходимо учитывать множество параметров, например, продолжительность периода увеличения ядра, его размеры и размеры генома.

3.3 Роль нуклеопорина Elys в регуляции плотности упаковки хроматина в клетках S2 дрозofilы

Сохранение жизнеспособности клеток S2 дрозofilы при деплеции Elys дает возможность проследить его влияние на компактизацию ДНК. Исследование вклада Elys в формирование характерной для дрозofilы конформации хроматина проводилось на основе сравнения данных Hi-C для контроля и эксперимента с деплецией этого белка на разрешении от 2 до 10 тыс.п.о. На предварительном этапе была проведена оценка качества полученных карт контактов. Кластеризация реплик на две группы (деплеция и контроль) указывает на хорошее разделение образцов, что, в совокупности с визуальной оценкой карт Hi-C, говорит о достаточном для последующего анализа качестве исходных данных (рис. 20).

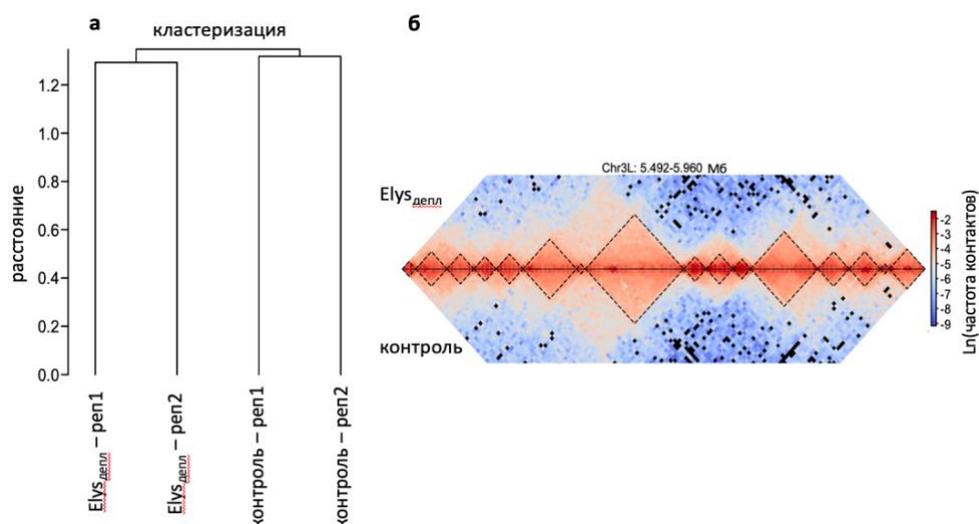


Рисунок 20. Данные Hi-C для эксперимента с деплецией Elys.

(а) Кластеризация реплик для деплеции и контроля Elys. (б) Фрагмент карты Hi-C для деплеции и контроля Elys.

3.4.1 Структура компартментов

Исследование влияния деплеции Elys на структуру хроматина проводилось на нескольких уровнях организации. Для изучения крупномасштабных взаимодействий, проявляющихся в виде А- и В-компартментов, были построены тепловые карты («седловые графики»), отражающие интенсивность контактов внутри них и между ними. При деплеции Elys наблюдается усиление неактивного и ослабление активного компартмента (рис. 21а, ВВ и АА область соответственно). При этом частоты контактов между участками генома, принадлежащими к разным компартментам, в целом не изменяются, т. е. в отсутствие белка Elys перемешивания компартментов не происходит (рис. 21б). Анализ их протяженности не выявил преобладающих переходов из компартмента типа А в тип В или наоборот (Р-значение = 0.36).

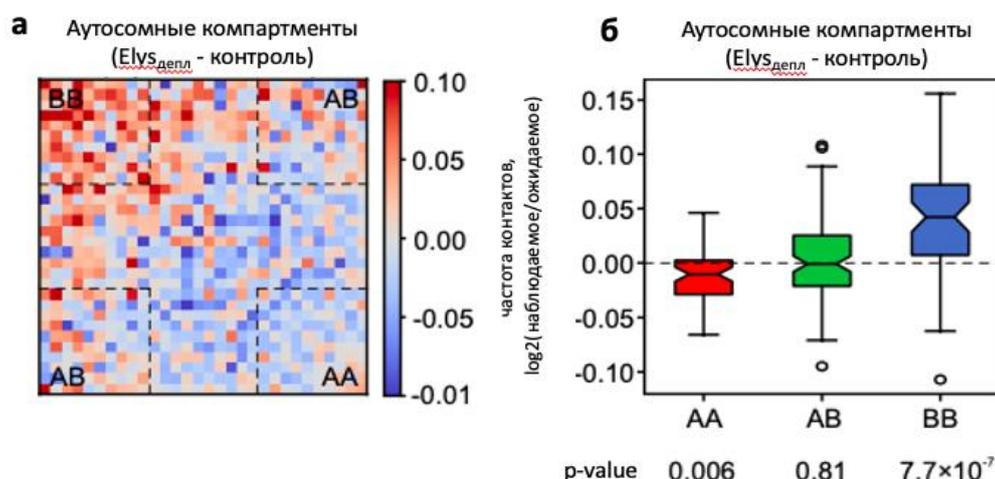


Рисунок 21. Взаимодействия хроматина на уровне компартментов.

(а) Агрегированные частоты контактов внутри и между компартментами. (б) Статистическая оценка значимости различий в частотах контактов (частоты контактов в областях, выделенных пунктирной линией на панели а).

Для получения более подробной картины компартментализации до и после деплеции Elys эта форма организации хроматина была проанализирована в трех интервалах расстояний между участками генома, контактирующими за пределами ТАДов: ближние (0.2-4 Мб), средние (4-10 Мб) и дальние (10-20 Мб) (рис. 22).

Взаимодействия внутри неактивного компартмента оказались после деплеции Elys более интенсивными для всех указанных диапазонов расстояний, тогда как для активного компартмента значимых отличий обнаружено не было, хотя в группе ближних контактов наблюдалось небольшое снижение их частот (рис. 22б). По-видимому, наблюдаемый на общей тепловой карте (рис.21) эффект ослабления А-компартмента обусловлен ближними взаимодействиями, является слабовыраженным и его статистическая значимость существенно зависит от баланса между числом измерений и амплитудой наблюдаемых изменений.

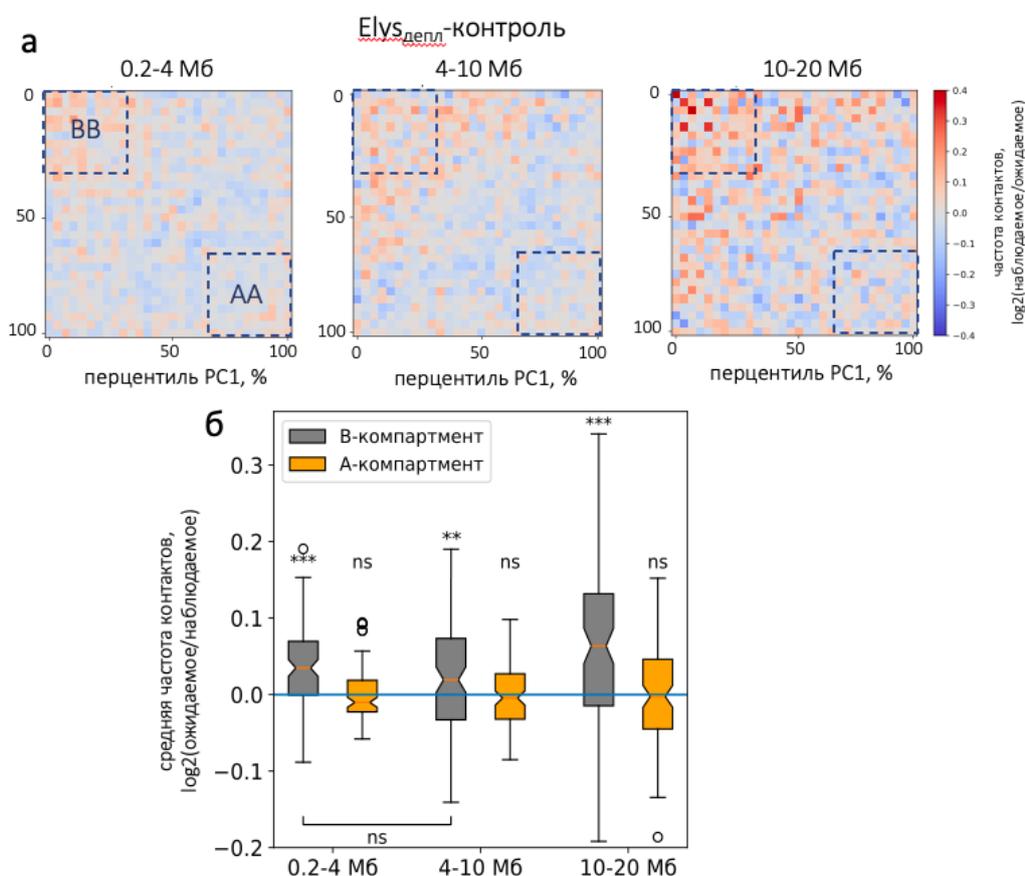


Рисунок 22. Изменение структуры компартментов, образованных взаимодействием участков генома, расположенных на разных расстояниях за пределами ТАДов.

(а) «Седловые графики». (б) Разница частот контактов в компартментах между картами для деплеции Elys и контроля (очерчены пунктирной линией на панели а).

3.4.2 Компактизация и взаимодействие ТАДов

Применение метода FISH к клеткам без и после деплеции Elys показывает, что в отсутствие этого белка происходит перемещение неактивного хроматина из

периферической в менее отдаленную от центра область ядра [14]. Проведенные ранее исследования по деплеции ламины говорят о том, что такое перемещение может напрямую сказываться на плотности ТАДов [10]. Для подробного изучения этого явления, все ТАДы были разделены на три группы: активные – с высоким содержанием активных цветов хроматина и низким содержанием ЛАДов, неактивные – с низким содержанием активных цветов хроматина и высоким содержанием ЛАДов и промежуточные. Для всех групп ТАДов с границами, совпадающими в контрольных картах и после деплеции, была рассчитана их плотность, выраженная через среднюю частоту контактов. Для активных ТАДов деплеция Elys привела к повышению плотности, тогда как для неактивных наблюдалась противоположная тенденция (рис. 23а). Аналогичный результат был получен при делении ТАДов на группы по квартилям экспрессии (рис. 23б). Сходный характер изменения компактизации ТАДов наблюдался ранее при деплеции ламины (рис. 23в).

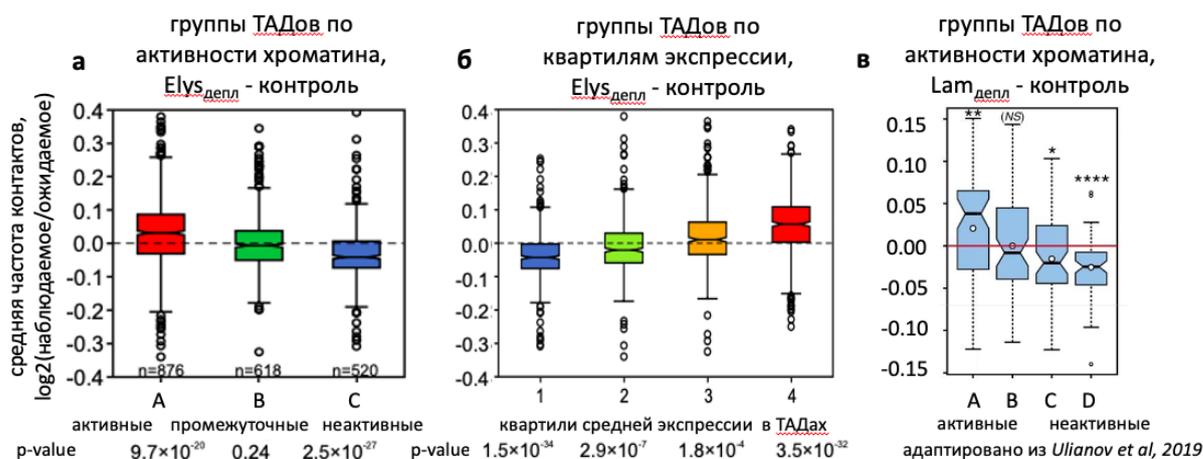


Рисунок 23. Плотность ТАДов.

(а) ТАДы, разделенные на группы по активности, Elys_{депл} - контроль. (б) ТАДы, разделенные на квартили экспрессии, Elys_{депл} - контроль. (в) ТАДы, разделенных на группы по активности, Lam_{депл} – контроль (адаптировано из *Ulianov et al, 2019*).

ТАДы дрозофилы, согласно данным литературы, имеют компартментное происхождение, поэтому изменения в структуре дальних контактов интересно проследить также и на уровне взаимодействий ТАДов внутри каждого из компартментов. При разбиении на группы по диапазонам расстояний, сходном с

подобраным для исследования структуры компартментов (см. раздел 2.2.2.2), повышение средней частоты контактов наблюдается для группы ТАДов из неактивного компартмента, разделенных малым и средним расстоянием, тогда как для дальних контактов неактивных ТАДов подобная тенденция не характерна ни на уровне оценки статистической значимости, ни при сравнении значений медиан (рис. 24а).

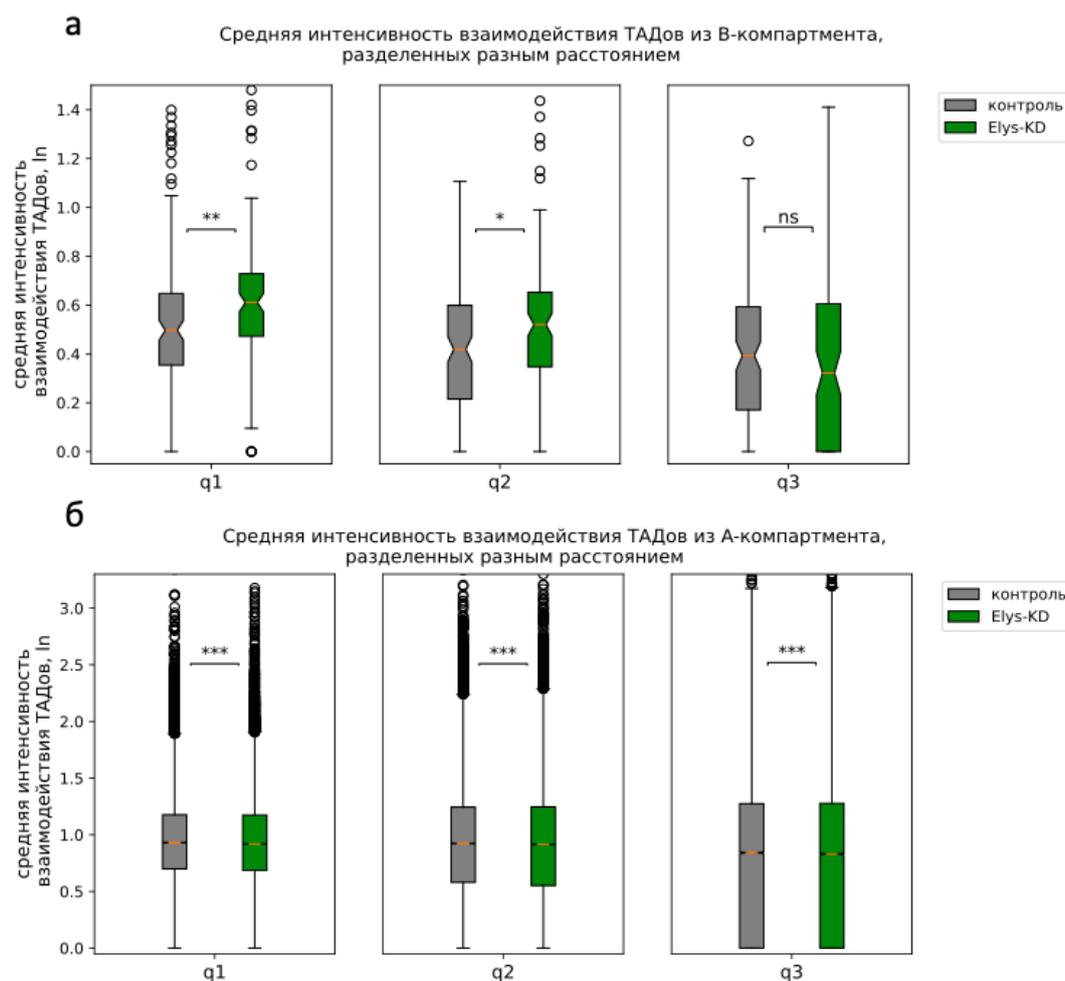


Рисунок 24. Средняя частота контактов ТАДов, разделенных разным расстоянием (подписи по оси X (q1, q2, q3) соответствуют квантилям расстояний). (а) ТАДы из неактивного компартмента. (б) ТАДы из активного компартмента.

Такая картина противоречит результатам, полученным при анализе “седлового графика” в области далеких контактов (рис. 22а). Важно отметить, что для расчета среднего взаимодействия отбирались только такие ТАДы, у которых принадлежность к определенному компартменту была выражена достаточно сильно (см. раздел 2.2.2.2). Вероятно, повышение частот дальних

контактов в области неактивного компартмента на “седловом графике” обусловлено участием в этом процессе локусов с отрицательным значением PC1, которые при этом не относятся к ярко выраженным неактивным ТАДам и, вероятно, часто бывают окружены участками активного хроматина. Ярко выраженные неактивные ТАДы дрозофилы в основном располагаются на периферии и, судя по полученным результатам, даже при декомпактизации вследствие деплеции Elys не способны формировать сильные дальние контакты, в отличие от тех участков неактивного хроматина, в окружении которых содержатся регионы из А-компартмента, предпочтительно локализующиеся ближе к центру ядра.

Для ТАДов из активного компартмента наблюдаются противоположные изменения: их взаимодействия ослабевают, однако этот эффект выражен настолько слабо, что, вероятно, его статистическая значимость достигается за счет большого числа событий. Таким образом, результаты, полученные как при подсчете средних частот контактов активных ТАДов (рис. 24б), так и при оценке взаимодействий между участками из А-компартмента (рис. 22), указывают на отсутствие значительных изменений в структуре активного хроматина на расстояниях, превышающих размеры ТАДов.

3.4.3 Структура хроматина в границах ТАДов

В связи с предполагаемым наличием у Elys_{нукл} ремоделирующего потенциала, опосредованного взаимодействием с комплексом РВАР, наряду с повышением плотности малых активных ТАДов при его деплеции ожидается также повышение компактизации в области границ, которое может проявляться в ослаблении сегрегации соседних ТАДов и их частичном перемешивании. Действительно, при нокдауне Elys частота взаимодействий через границы заметно повышается, особенно ярко это проявляется в сильных границах (рис. 25а), для которых характерно выраженное обогащение Elys_{нукл} по сравнению с его содержанием в прилегающих к границе ТАДах (рис. 25б). Наблюдаемые изменения подтверждают способность Elys к ремоделированию хроматина и согласуются с концепцией о формировании ТАДов и их границ у дрозофилы в результате

чередования участков активного и неактивного хроматина: деплеция Elys может приводить к компактизации активного хроматина, формирующего границы между неактивными ТАДами, и приводить к увеличению частоты контактов между ними.

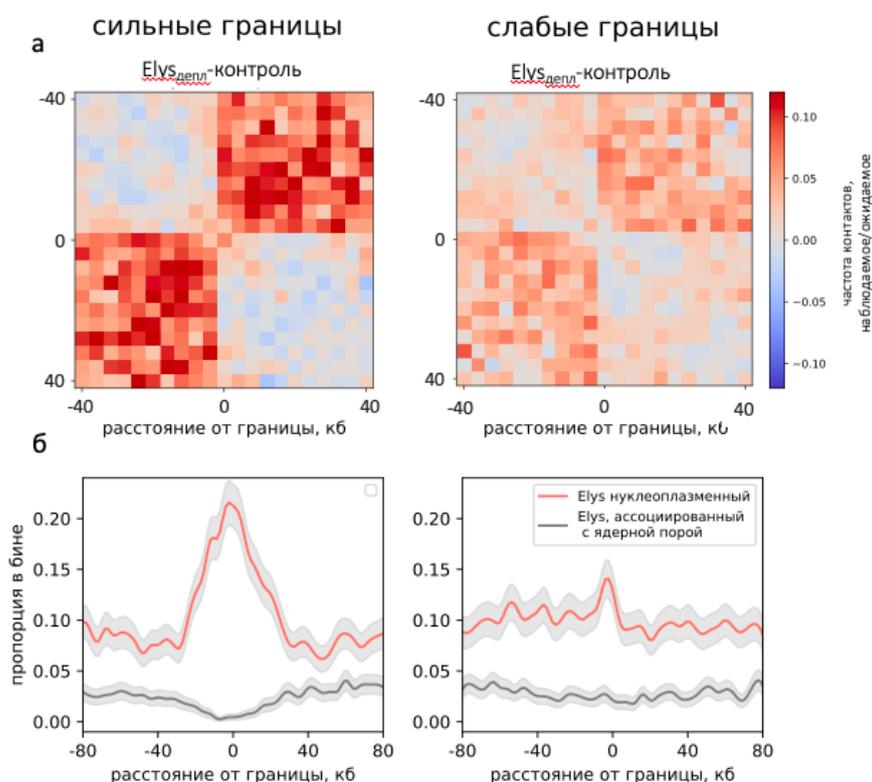


Рисунок 25. Влияние деплеции Elys на структуру хроматина в границах.

(а) Средняя карта Hi-C вокруг сильных (слева) и слабых (справа) границ. (б) Профили пропорции Elys_{нукл} и Elys_{япк} вокруг сильных (слева) и слабых (справа) границ.

С другой стороны, 86% ТАДов остается в прежних границах (± 5 тыс.п.о.). Образование новых границ происходит при деплеции не чаще, чем исчезновение (P -значение ~ 0.74), и для профиля инсуляции вокруг дифференциальных границ значимых отличий не обнаруживается (рис. 6 Приложения). Эти наблюдения соотносятся с полученными ранее представлениями о сложной природе границ ТАДов дрозофилы, в формировании которых участвует множество факторов.

3.4.4 Локальные изменения структуры хроматина

Elys обладает способностью напрямую связываться с ДНК [12], поэтому его деплеция может оказывать влияние на конформацию хроматина непосредственно

в области сайтов связывания. Так как этот нуклеопорин представлен в ядре двумя фракциями, нуклеоплазменной и ассоциированной с комплексом ядерной поры, исследование в данном направлении проводилось с учетом пространственной локализации, информация о которой была получена в работе [14].

Для оценки изменения локальной плотности хроматина в окрестности сайтов связывания были построены средние тепловые карты контактов, которые содержат информацию о взаимодействиях внутри самого этого сайта и в пределах его ближайшего окружения. При деплеции Elys для ЯПК-ассоциированной фракции было показано снижение частоты контактов в местах его посадки по сравнению с контролем, свидетельствующее о локальной декомпактизации хроматина (рис. 26а). Для Elys_{нукл} картина оказалась не столь однозначна. С одной стороны, характер взаимодействий здесь сильно отличался от наблюдаемого в области связывания с ядерной порой, а с другой, обнаруженное повышение частот взаимодействий не ограничивалось сайтом связывания Elys_{нукл}, а распространялось и на близлежащие участки генома (рис. 26б). Влияние Elys на коммуникацию участков генома, расположенных по разные стороны от его сайта связывания, было проанализировано в терминах кривых инсуляции, представляющих собой профиль локальных взаимодействий. Деплеция приводила к повышению частот контактов такого типа для сайтов связывания Elys_{нукл}, что может говорить о компактизации активного хроматина в результате деплеции этого белка (рис. 27). Для Elys_{япк} масштаб наблюдаемого эффекта не позволяет говорить о его существенном влиянии на частоты взаимодействий участков генома, расположенных по разные стороны от сайта связывания, хотя направление изменений соответствует общей картине, наблюдаемой при деплеции этого белка.

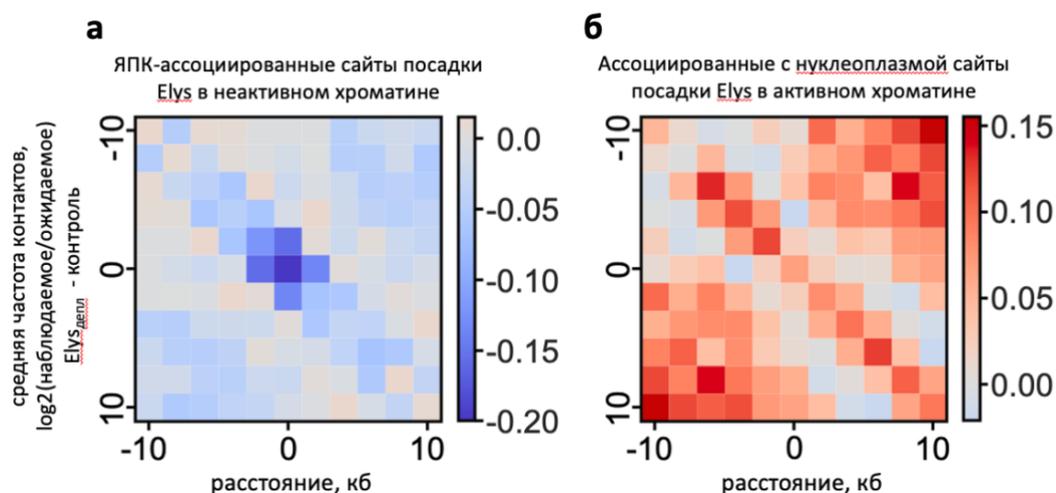


Рисунок 26. Средняя частота контактов вокруг сайтов связывания Elys. (а) В составе ядерной поры. (б) В нуклеоплазме.

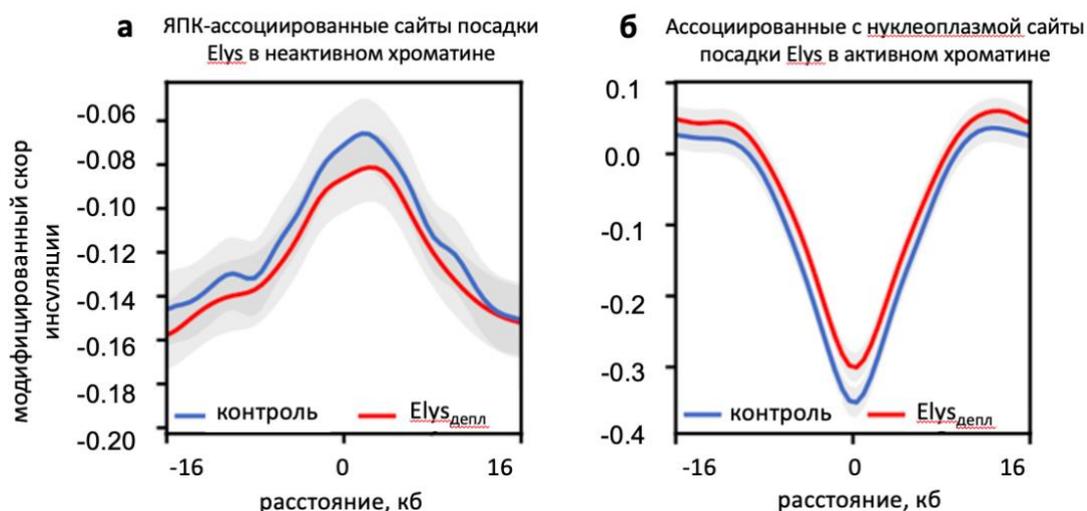


Рисунок 27. Модифицированные кривые инсуляции вокруг сайтов связывания.

(а) Elys в составе ядерной поры. (б) Elys в нуклеоплазме.

3.4.5 Влияние деплеции Elys на петли и энхансерно-промоторные взаимодействия

Известно, что наряду с компартментами и ТАДами у дрозофилы наблюдается такая разновидность структуры хроматина, как петли [76,77]. Этот тип взаимодействий также был рассмотрен с точки зрения влияния деплеции Elys. Анализ проводился для объединения всех детектированных петель и для петель, чьи основания содержат сайт связывания Elys_{нукл} (в пересечении с энхансерами и

независимо от них) или Elys_{ЯПК}. Эти группы петель были рассмотрены по отдельности. Как при деплеции Elys, так и в контрольных образцах наблюдается обогащение в интенсивности взаимодействия оснований петель, т. е. используемый алгоритм в целом позволяет их детектировать, однако изменения частоты контактов при сравнении контрольных данных и после деплеции Elys не обнаруживается (Р-значение ~ 0.051 , Таблица 1 Приложения).

Участие Elys_{нукл} в процессе активации транскрипции на уровне регуляции структуры хроматина и пересечение 42% энхансеров клеток S2 с сайтами связывания Elys_{нукл} послужили основанием к проведению анализа взаимодействий этих энхансеров с соответствующими им промоторами. Такой подход также не выявил заметных изменений в частоте контактов соответствующих участков генома при деплеции Elys по сравнению с контролем. Кроме того, для таких случаев энхансерно-промоторных взаимодействий при визуальной оценке средних карт контактов не было выявлено обогащения в интенсивности по сравнению с окружающим фоном. Отсутствие ярко-выраженного контакта может говорить как о неточностях в определении энхансерно-промоторных пар, так и об альтернативном механизме их взаимодействия, не требующем стабильной колокализации в пространстве [124,125].

Обобщение результатов параграфа и их обсуждение

Проведенный анализ указывает на существенный вклад нуклеопорина Elys в регуляцию компактизации хроматина на нескольких уровнях организации. Так, наблюдаемые при деплеции локальные изменения организации хроматина в области сайтов связывания ЯПК-ассоциированного Elys указывают на его способность взаимодействовать с неактивным хроматином на периферии ядра и приводить к повышению его плотности. Вероятно, утрата связи с ядерной оболочкой в сайтах посадки Elys и локальная декомпактизация при его деплеции приводит в результате к общей картине разрыхления неактивных ГАДов в целом. Кроме того, наблюдается перестройка структуры хроматина на следующих, более масштабных, уровнях организации. Изменяется структура компартментов. Неактивный компартмент становится более выраженным. Предположительно, это

явление может объясняться большей степенью свободы неактивного хроматина в результате его декомпактизации в масштабах ТАДов, и в то же время довольно слабой активацией транскрипции [14], не сопровождающейся заметным переходом в А-компаратмент. В совокупности эти факты могут приводить к более вероятной реализации контактов внутри В-компаратмента на дальних расстояниях. Для активного хроматина характерна противоположная тенденция: активные ТАДы становятся плотнее, что коррелирует с картиной локальных изменений вокруг сайтов связывания Elys в нуклеоплазме. Это наблюдение согласуется с предположением об активирующей, опосредующей декомпактизацию хроматина, ролью Elys в области нуклеоплазмы, далекой от периферии. Выявленные закономерности хорошо соотносятся с полученными ранее представлениями о роли ламины в организации активного и неактивного хроматина в пространстве ядра [10].

Таким образом, обнаруженные изменения указывают на существенную роль нуклеопорина Elys в поддержании структуры, плотности упаковки и правильной сегрегации активного и неактивного хроматина, что, в контексте рассмотрения факторов, влияющих на его организацию в дрозифиле, ставит комплекс ядерной поры в один ряд с ядерной ламиной.

Заключение

В настоящей работе представлены результаты исследования роли некоторых факторов в организации пространственной архитектуры генома у дрозифилы, в частности белков ядерной периферии и активации транскрипции на ранних стадиях сперматогенеза. Проведенный анализ литературы выявил неравномерность в степени изученности вклада того или иного фактора: значительная доля работ приходится на исследование влияния архитектурных белков, модификации гистонов, а также результатов индукции или подавления транскрипции, тогда как другие, не менее важные, компоненты структуры хроматина или протекающие в ядре процессы изучены в данном контексте в меньшей степени.

Наличие данных Hi-C для двух ранних стадий сперматогенеза, в силу многогранности этого процесса, позволило изучить несколько факторов, потенциально связанных со структурой хроматина. Исследование особенностей упаковки ДНК с точки зрения активации сперматоцитспецифичных генов выявило заметные отличия между сперматогониями и сперматоцитами, обусловленные этим процессом. Обнаруженные изменения затрагивают разные уровни организации хроматина: от локального разрыхления до перехода между компартментами. С другой стороны, в работе встречаются примеры отсутствия прямой взаимосвязи между трехмерной структурой хроматина, транскрипцией и функциональным сходством сперматоцитспецифичных генов.

Поскольку переход к стадии сперматоцитов сопровождается не только активацией транскрипции множества генов, но и увеличением объема ядра, было выдвинуто предположение о том, что ряд изменений в структуре хроматина может объясняться именно этим процессом. Сопоставление с результатами эксперимента по искусственному раздуванию ядра показало, что, действительно, некоторые отличия, наблюдаемые между двумя стадиями, согласуются с картиной, характерной для индуцированного увеличения объема ядра, и не связаны с активацией транскрипции.

Понимание роли различных факторов, определяющих структуру хроматина в дрозофиле, было дополнено результатами анализа данных Hi-C при деплеции белка Elys. Выявленные изменения указывают на существенный вклад этого белка в поддержание плотности упаковки и локализации неактивного хроматина на периферии ядра, а также в декомпактизацию активных участков генома. Учитывая согласованность обнаруженных изменений с полученной ранее картиной при деплеции белков ламины, интерпретация наблюдаемой динамики хроматина выходит за пределы описания функции отдельно взятого белка и на основе полученных результатов позволяет выдвинуть предположение о том, что регуляция распределения активного и неактивного хроматина может осуществляться за счет целого комплекса белков периферии.

В ходе исследования был сделан ряд наблюдений как методического, так и естественно-научного характера. В целом, результаты данной работы обогащают совокупную картину факторов, влияющих на структуру хроматина плодовой мушки, подчеркивают важность изучения других составляющих, потенциально ассоциированных с изменением конформации хроматина, и открывают новые перспективы исследований как в биологии сперматогенеза дрозофилы, так и в области технических аспектов анализа данных Hi-C.

На основании сделанных в работе наблюдений, можно сформулировать следующие **выводы**:

1. При переходе от стадии сперматогониев к стадии сперматоцитов происходят связанные с активацией сперматоцитспецифичных генов изменения конформации хроматина дрозофилы, такие как:
 - а) частичный переход в активный компартмент участков генома, содержащих промоторы сперматоцитспецифичных генов;
 - б) повышение частот контактов сперматоцитспецифичных генов;
 - в) локальное снижение плотности хроматина вокруг стартов сперматоцитспецифичных генов.
2. С помощью усовершенствованного и реализованного алгоритма поиска значимых контактов для заданного списка геномных локусов было установлено, что повышение интенсивности взаимодействий сперматоцитспецифичных генов не является специфичным: частота таких контактов не отличается от наблюдаемой в парах сперматоцитспецифичных и активных в различных тканях генов.
3. Функциональное сходство сперматоцитспецифичных генов ассоциировано с их линейной, но не пространственной колокализацией.
4. Переход из неактивного компартмента в активный характерен для промоторов сперматоцитспецифичных генов, расположенных близко к границе ТАДов, т. е. масштаб изменений в структуре хроматина может зависеть от исходного расположения геномного локуса в более или менее активной области.

5. Неактивный компартмент сперматогониев и сперматоцитов является более выраженным при сравнении с эмбриональной клеточной линией. Наблюдаемые различия не связаны напрямую с уровнем транскрипции, что подчеркивает значимость других факторов, предположительно вовлеченных в формирование неактивного компартмента.
6. Обнаруженные изменения в структуре хроматина на уровне выраженности дальних взаимодействий, компартментов, границ и петель, вероятно, объясняются увеличением размера ядра сперматоцитов, поскольку соотносятся с закономерностями, продемонстрированными ранее в экспериментах по индуцированному увеличению объема ядра.
7. Деpletion нуклеопорина Elys приводит к изменению плотности упаковки хроматина, сопоставимому с наблюдаемым при деpletion белков ламинины, а именно:
 - а) в области связывания Elys в неактивном хроматине происходит заметная декомпактизация;
 - б) плотность неактивных TADов снижается, а активных повышается;
 - в) неактивный компартмент становится более выраженным на всех масштабах внутрихромосомных взаимодействий.

Благодарности

Выражаю благодарность моему научному руководителю Екатерине Евгеньевне Храмеевой, а также всем коллегам, принимавшим участие в обсуждениях и поделившимся ценным опытом, советами и идеями: Сергею Владимировичу Ульянову, Юрию Ясеновичу Шевелеву, Александре Алексеевне Галицыной, Михаилу Сергеевичу Гельфанду и Дмитрию Олеговичу Крюкову. Это исследование стало возможным благодаря наличию уникального экспериментального материала, полученного в лабораториях Юрия Ясеновича Шевелева и Сергея Владимировича Разина. Отдельное спасибо Ире Ж., Кате Т., Лере Ш. и, конечно, моей семье за дружескую и не только моральную поддержку в течение всего времени этих научных изысканий.

Список сокращений

ПЦР – полимеразная цепная реакция

ЛАД – ламина-ассоциированный домен

ТАД – топологически ассоциированный домен

TSA-seq – геномный метод, который позволяет оценить цитологические расстояния хромосомных локусов по всему геному относительно определенного ядерного компартмента

RNA-seq – метод измерения РНК

GRO-seq – метод измерения нарождающейся РНК

DamID-seq – протокол, используемый для определения сайтов связывания ДНК- и хроматин-связывающих белков у эукариот

ChIP-seq – метод анализа ДНК-белковых взаимодействий, основанный на иммунопреципитации хроматина и секвенировании ДНК

GO-CaRT – метод картирования ЛАДов *in vivo* и *in vitro*

HiChiP – протокол, используемый для детекции пространственных взаимодействий участков генома, связанных с определенным белком

STARR-seq – метод анализа энхансерной активности множества последовательностей ДНК

H3K27ac – ацетилирование остатка лизина на гистоне H3

H3K9me2, H3K9me3, H3K27me3 – ди- и триметилирование различных остатков лизина на гистоне H3

PRC1 – репрессивный комплекс поликомба 1

ORC, ORC2 – комплексы распознавания начала репликации

TPM – единица измерения транскрипции гена, учитывающая его длину и размер библиотеки

Мб - 1000000 пар оснований

Список литературы

- 1 Lieberman-Aiden E, Van Berkum NL, Williams L *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326(5950), 289–293 (2009).
- 2 Kurotaki D, Kikuchi K, Cui K *et al.* Chromatin structure undergoes global and local reorganization during murine dendritic cell development and activation. *Proc. Natl. Acad. Sci. U.S.A.* 119(34), e2207009119 (2022).
- 3 Kazakevych J, Sayols S, Messner B, Krienke C, Soshnikova N. Dynamic changes in chromatin states during specification and differentiation of adult intestinal stem cells. *Nucleic Acids Research* 45(10), 5770–5784 (2017).
- 4 Pletenev IA, Bazarevich M, Zagirova DR *et al.* Extensive long-range polycomb interactions and weak compartmentalization are hallmarks of human neuronal 3D genome, *Genomics*, (2023).
- 5 Ugur B, Chen K, Bellen HJ. *Drosophila* tools and assays for the study of human diseases. *Disease Models & Mechanisms* 9(3), 235–244 (2016).
- 6 Llorens-Giralt P, Camilleri-Robles C, Corominas M, Climent-Cantó P. Chromatin Organization and Function in *Drosophila*. *Cells* 10(9), 2362 (2021).
- 7 Matthews NE, White R. Chromatin Architecture in the Fly: Living without CTCF/Cohe sin Loop Extrusion?: Alternating Chromatin States Provide a Basis for Domain Architecture in *Drosophila*. *BioEssays* 41(9), 1900048 (2019).
- 8 Ulianov SV, Khrameeva EE, Gavrillov AA *et al.* Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 26(1), 70–84 (2016).
- 9 Smith OK, Aladjem MI. Chromatin Structure and Replication Origins: Determinants of Chromosome Replication and Nuclear Organization. *Journal of Molecular Biology* 426(20), 3330–3341 (2014).
- 10 Ulianov SV, Doronin SA, Khrameeva EE *et al.* Nuclear lamina integrity is required for proper spatial organization of chromatin in *Drosophila*. *Nat Commun* 10(1), 1176 (2019).
- 11 Gillespie PJ, Khoudoli GA, Stewart G, Swedlow JR, Blow JJ. ELYS/MEL-28 chromatin association coordinates nuclear pore complex assembly and replication licensing. *Curr Biol* 17(19), 1657–1662 (2007).
- 12 Rasala BA, Ramos C, Harel A, Forbes DJ. Capture of AT-rich Chromatin by ELYS Recruits POM121 and NDC1 to Initiate Nuclear Pore Assembly. *MBoC* 19(9), 3982–3996 (2008).
- 13 Pascual-Garcia P, Debo B, Aleman JR *et al.* Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. *Molecular Cell* 66(1), 63-76.e6 (2017).
- 14 Doronin SA, Ilyin AA, Kononkova AD *et al.* Nucleoporin Elys attaches peripheral chromatin to the nuclear pores in interphase nuclei, (2023).
- 15 Sexton T, Yaffe E, Kenigsberg E *et al.* Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell* 148(3), 458–472 (2012).

- 16 Lattke M, Goldstone R, Ellis JK *et al.* Extensive transcriptional and chromatin changes underlie astrocyte maturation in vivo and in culture. *Nat Commun* 12(1), 4335 (2021).
- 17 Freitag K, Eede P, Ivanov A *et al.* Diverse but unique astrocytic phenotypes during embryonic stem cell differentiation, culturing and development. *Commun Biol* 6(1), 40 (2023).
- 18 Karlsson M, Zhang C, Méar L *et al.* A single-cell type transcriptomics map of human tissues. *Sci. Adv.* 7(31), eabh2169 (2021).
- 19 McKee BD, Yan R, Tsai J-H. Meiosis in male *Drosophila*. *Spermatogenesis* 2(3), 167–184 (2012).
- 20 Patel L, Kang R, Rosenberg SC *et al.* Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. *Nat Struct Mol Biol* 26(3), 164–174 (2019).
- 21 Sanders JT, Gollosi R, Das P *et al.* Loops, topologically associating domains, compartments, and territories are elastic and robust to dramatic nuclear volume swelling. *Sci Rep* 12(1), 4721 (2022).
- 22 Rowley MJ, Lyu X, Rana V *et al.* Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization. *Cell Reports* 26(11), 2890-2903.e3 (2019).
- 23 Rowley MJ, Nichols MH, Lyu X *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Molecular Cell* 67(5), 837-852.e7 (2017).
- 24 Olins DE, Olins AL. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* 4(10), 809–814 (2003).
- 25 Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 3(10), 793–796 (2006).
- 26 Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568(7750), 49–54 (2019).
- 27 Cardozo Gizzi AM, Cattoni DI, Fiche J-B *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell* 74(1), 212-222.e5 (2019).
- 28 Schermelleh L, Carlton PM, Haase S *et al.* Subdiffraction Multicolor Imaging of the Nuclear Periphery with 3D Structured Illumination Microscopy. *Science* 320(5881), 1332–1336 (2008).
- 29 Flores V, Farabella I, Nir G. Genome-wide tracing to decipher nuclear organization. *Current Opinion in Cell Biology* 82, 102175 (2023).
- 30 Dekker J, Rippe K, Dekker M, Kleckner N. Capturing Chromosome Conformation. *Science* 295(5558), 1306–1311 (2002).
- 31 Flyamer IM, Gassler J, Imakaev M *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544(7648), 110–114 (2017).
- 32 Beagrie RA, Scialdone A, Schueler M *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543(7646), 519–524 (2017).
- 33 Rusk N. SPRITE maps the 3D genome. *Nat Methods* 15(8), 572–572 (2018).
- 34 Gavrilov AA, Zharikova AA, Galitsyna AA *et al.* Studying RNA–DNA

- interactome by Red-C identifies noncoding RNAs associated with various chromatin types and reveals transcription dynamics. *Nucleic Acids Research* 48(12), 6699–6714 (2020).
- 35 Rao SSP, Huntley MH, Durand NC *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159(7), 1665–1680 (2014).
- 36 Hsieh T-HS, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* 162(1), 108–119 (2015).
- 37 Lafontaine DL, Yang L, Dekker J, Gibcus JH. Hi-C 3.0: Improved Protocol for Genome-Wide Chromosome Conformation Capture. *Current Protocols* 1(7) (2021).
- 38 Cremer T, Cremer C. Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem* 50(4), 223–272 (2006).
- 39 Contessoto VG, Dudchenko O, Aiden EL, Wolynes PG, Onuchic JN, Di Pierro M. Interphase chromosomes of the *Aedes aegypti* mosquito are liquid crystalline and can sense mechanical cues. *Nat Commun* 14(1), 326 (2023).
- 40 Dixon JR, Jung I, Selvaraj S *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* 518(7539), 331–336 (2015).
- 41 Hou C, Li L, Qin ZS, Corces VG. Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains. *Molecular Cell* 48(3), 471–484 (2012).
- 42 Dixon JR, Selvaraj S, Yue F *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398), 376–380 (2012).
- 43 Sanborn AL, Rao SSP, Huang S-C *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* 112(47) (2015).
- 44 Tolhuis B, Blom M, Kerkhoven RM *et al.* Interactions among Polycomb Domains Are Guided by Chromosome Architecture. *PLoS Genet* 7(3), e1001343 (2011).
- 45 Bantignies F, Roure V, Comet I *et al.* Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila*. *Cell* 144(2), 214–226 (2011).
- 46 Zhang Y, Wong C-H, Birnbaum RY *et al.* Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. *Nature* 504(7479), 306–310 (2013).
- 47 Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 489(7414), 109–113 (2012).
- 48 Galitsyna A, Ulianov SV, Bykov NS *et al.* Extrusion fountains are hallmarks of chromosome organization emerging upon zygotic genome activation, (2023).
- 49 Peterson SC, Samuelson KB, Hanlon SL. Multi-Scale Organization of the *Drosophila melanogaster* Genome. *Genes* 12(6), 817 (2021).
- 50 Zenk F, Zhan Y, Kos P *et al.* HP1 drives de novo 3D genome reorganization in early *Drosophila* embryos. *Nature* 593(7858), 289–293 (2021).

- 51 McKee BD. Homologous pairing and chromosome dynamics in meiosis and mitosis. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1677(1–3), 165–180 (2004).
- 52 AlHaj Abed J, Erceg J, Goloborodko A *et al.* Highly structured homolog pairing reflects functional organization of the *Drosophila* genome. *Nat Commun* 10(1), 4485 (2019).
- 53 Erceg J, AlHaj Abed J, Goloborodko A *et al.* The genome-wide multi-layered architecture of chromosome pairing in early *Drosophila* embryos. *Nat Commun* 10(1), 4486 (2019).
- 54 Hansen AS. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* 11(1), 132–148 (2020).
- 55 Bartkuhn M, Straub T, Herold M *et al.* Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J* 28(7), 877–888 (2009).
- 56 Van Bortle K, Nichols MH, Li L *et al.* Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* 15(5), R82 (2014).
- 57 Kharchenko PV, Alekseyenko AA, Schwartz YB *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471(7339), 480–485 (2011).
- 58 Filion GJ, Van Bemmelen JG, Braunschweig U *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell* 143(2), 212–224 (2010).
- 59 El-Sharnouby S, Fischer B, Magbanua JP *et al.* Regions of very low H3K27me3 partition the *Drosophila* genome into topological domains. *PLoS ONE* 12(3), e0172725 (2017).
- 60 Wang Q, Sun Q, Czajkowsky DM, Shao Z. Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat Commun* 9(1), 188 (2018).
- 61 Narlikar GJ. Phase-separation in chromatin organization. *J Biosci* 45, 5 (2020).
- 62 Li X, An Z, Zhang W, Li F. Phase Separation: Direct and Indirect Driving Force for High-Order Chromatin Organization. *Genes* 14(2), 499 (2023).
- 63 Rippe K. Liquid–Liquid Phase Separation in Chromatin. *Cold Spring Harb Perspect Biol* 14(2), a040683 (2022).
- 64 Chen Q, Zhao L, Soman A *et al.* Chromatin Liquid–Liquid Phase Separation (LLPS) Is Regulated by Ionic Conditions and Fiber Length. *Cells* 11(19), 3145 (2022).
- 65 Bintu B, Mateo LJ, Su J-H *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362(6413), eaau1783 (2018).
- 66 Ulianov SV, Zakharova VV, Galitsyna AA *et al.* Order and stochasticity in the folding of individual *Drosophila* genomes. *Nat Commun* 12(1), 41 (2021).
- 67 Götz M, Messina O, Espinola S, Fiche J-B, Nollmann M. Multiple parameters shape the 3D chromatin structure of single nuclei at the *doc* locus in *Drosophila*. *Nat Commun* 13(1), 5375 (2022).
- 68 Szabo Q, Jost D, Chang J-M *et al.* TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci. Adv.* 4(2), eaar8082 (2018).

- 69 Cattoni DI, Cardozo Gizzi AM, Georgieva M *et al.* Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat Commun* 8(1), 1753 (2017).
- 70 Ramírez F, Bhardwaj V, Arrigoni L *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* 9(1), 189 (2018).
- 71 Kahn TG, Savitsky M, Kuong C *et al.* Topological screen identifies hundreds of Cp190- and CTCF-dependent *Drosophila* chromatin insulator elements. *Sci. Adv.* 9(5), eade0090 (2023).
- 72 Chathoth KT, Zabet NR. Chromatin architecture reorganization during neuronal cell differentiation in *Drosophila* genome. *Genome Res.* 29(4), 613–625 (2019).
- 73 Karch F. In vivo studies of the *Drosophila* insulator factor CTCF reach a Catch 22. *BMC Biol* 13(1), 71 (2015).
- 74 Cavalheiro GR, Girardot C, Viales RR *et al.* CTCF, BEAF-32, and CP190 are not required for the establishment of TADs in early *Drosophila* embryos but have locus-specific roles. *Sci. Adv.* 9(5), eade1085 (2023).
- 75 Arzate-Mejía RG, Josué Cerecedo-Castillo A, Guerrero G, Furlan-Magaril M, Recillas-Targa F. In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*. *Nat Commun* 11(1), 894 (2020).
- 76 Chathoth KT, Mikheeva LA, Crevel G *et al.* The role of insulators and transcription in 3D chromatin organization of flies. *Genome Res.* 32(4), 682–698 (2022).
- 77 Ogiyama Y, Schuettengruber B, Papadopoulos GL, Chang J-M, Cavalli G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Molecular Cell* 71(1), 73-88.e5 (2018).
- 78 Loubiere V, Papadopoulos GL, Szabo Q, Martinez A-M, Cavalli G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci. Adv.* 6(2), eaax4001 (2020).
- 79 Briand N, Collas P. Lamina-associated domains: peripheral matters and internal affairs. *Genome Biol* 21(1), 85 (2020).
- 80 Bondarenko SM, Sharakhov IV. Reorganization of the nuclear architecture in the *Drosophila melanogaster* Lamin B mutant lacking the CaaX box. *Nucleus* 11(1), 283–298 (2020).
- 81 Nazer E. To be or not be (in the LAD): emerging roles of lamin proteins in transcriptional regulation. *Biochemical Society Transactions* 50(2), 1035–1044 (2022).
- 82 Kind J, Pagie L, Ortabozkoyun H *et al.* Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell* 153(1), 178–192 (2013).
- 83 Milon BC, Cheng H, Tselebrovsky MV *et al.* Role of Histone Deacetylases in Gene Regulation at Nuclear Lamina. *PLoS ONE* 7(11), e49692 (2012).
- 84 Verboon JM, Rincon-Arano H, Werwie TR *et al.* Wash Interacts with Lamin and Affects Global Nuclear Organization. *Current Biology* 25(6), 804–810 (2015).
- 85 Frost B, Bardai FH, Feany MB. Lamin Dysfunction Mediates Neurodegeneration in Tauopathies. *Current Biology* 26(1), 129–136 (2016).
- 86 Pascual-Garcia P, Capelson M. The nuclear pore complex and the genome: organizing and regulatory principles. *Current Opinion in Genetics & Development* 67, 142–150 (2021).

- 87 Shevelyov YY. The Role of Nucleoporin Elys in Nuclear Pore Complex Assembly and Regulation of Genome Architecture. *IJMS* 21(24), 9475 (2020).
- 88 Richards L, Lord CL, Benton ML, Capra JA, Nordman JT. Nucleoporins facilitate ORC loading onto chromatin. *Cell Reports* 41(6), 111590 (2022).
- 89 Capelson M, Liang Y, Schulte R, Mair W, Wagner U, Hetzer MW. Chromatin-Bound Nuclear Pore Components Regulate Gene Expression in Higher Eukaryotes. *Cell* 140(3), 372–383 (2010).
- 90 Kalverda B, Pickersgill H, Shloma VV, Fornerod M. Nucleoporins Directly Stimulate Expression of Developmental and Cell-Cycle Genes Inside the Nucleoplasm. *Cell* 140(3), 360–371 (2010).
- 91 Ibarra A, Hetzer MW. Nuclear pore proteins and the control of genome functions. *Genes Dev.* 29(4), 337–349 (2015).
- 92 Capelson M. You are who your friends are—nuclear pore proteins as components of chromatin-binding complexes. *FEBS Letters* 597(22), 2769–2781 (2023).
- 93 Gozalo A, Duke A, Lan Y *et al.* Core Components of the Nuclear Pore Bind Distinct States of Chromatin and Contribute to Polycomb Repression. *Molecular Cell* 77(1), 67-81.e7 (2020).
- 94 Kobayashi W, Takizawa Y, Aihara M, Negishi L, Ishii H, Kurumizaka H. Structural and biochemical analyses of the nuclear pore complex component ELYS identify residues responsible for nucleosome binding. *Commun Biol* 2(1), 163 (2019).
- 95 Kuhn TM, Pascual-Garcia P, Gozalo A, Little SC, Capelson M. Chromatin targeting of nuclear pore proteins induces chromatin decondensation. *Journal of Cell Biology* 218(9), 2945–2961 (2019).
- 96 Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* 169(2), 216-228.e19 (2017).
- 97 Fuller MT. Genetic control of cell proliferation and differentiation in *Drosophila* spermatogenesis. *Seminars in Cell & Developmental Biology* 9(4), 433–444 (1998).
- 98 Hiller M, Chen X, Pringle MJ *et al.* Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* 131(21), 5297–5308 (2004).
- 99 Beall EL, Lewis PW, Bell M, Rocha M, Jones DL, Botchan MR. Discovery of tMAC: a *Drosophila* testis-specific meiotic arrest complex paralogous to Myb–Muv B. *Genes Dev.* 21(8), 904–919 (2007).
- 100 Ilyin AA, Kononkova AD, Golova AV *et al.* Comparison of genome architecture at two stages of male germline cell differentiation in *Drosophila*. *Nucleic Acids Research* 50(6), 3203–3225 (2022).
- 101 Shevelyov YY, Lavrov SA, Mikhaylova LM *et al.* The B-type lamin is required for somatic repression of testis-specific gene clusters. *Proc. Natl. Acad. Sci. U.S.A.* 106(9), 3282–3287 (2009).
- 102 Zheng Y, Zhang L, Jin L *et al.* Unraveling three-dimensional chromatin structural dynamics during spermatogonial differentiation. *Journal of Biological Chemistry* 298(2), 101559 (2022).
- 103 Lee L, Rosin LF. Uncharted territories: Solving the mysteries of male meiosis in

- flies. *PLoS Genet* 20(3), e1011185 (2024).
- 104 Cenci G, Bonaccorsi S, Pisano C, Verni F, Gatti M. Chromatin and microtubule organization during premeiotic, meiotic and early postmeiotic stages of *Drosophila melanogaster* spermatogenesis*. *Journal of Cell Science* 107(12), 3521–3534 (1994).
- 105 Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* 36(1), 311–316 (2020).
- 106 Krueger F, James F, Ewels P *et al.* FelixKrueger/TrimGalore: v0.6.10 - add default decompression path, Zenodo, (2023).
- 107 Imakaev M, Fudenberg G, McCord RP *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* 9(10), 999–1003 (2012).
- 108 Open2C, Abdennur N, Abraham S *et al.* Cooltools: enabling high-resolution Hi-C analysis in Python, Bioinformatics, (2022).
- 109 Yang T, Zhang F, Yardımcı GG *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* 27(11), 1939–1949 (2017).
- 110 Filippova D, Patro R, Duggal G, Kingsford C. Identification of alternative topological domains in chromatin. *Algorithms Mol Biol* 9(1), 14 (2014).
- 111 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6), 841–842 (2010).
- 112 Won H, De La Torre-Ubieta L, Stein JL *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 538(7626), 523–527 (2016).
- 113 Yu G. Gene Ontology Semantic Similarity Analysis Using GOSemSim. In: *Stem Cell Transcriptional Networks (Volume 2117)*. Kidder BL (Ed.), Springer US, New York, NY, 207–215 (2020).
- 114 Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* 339(6123), 1074–1077 (2013).
- 115 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14(4), 417–419 (2017).
- 116 Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4, 1521 (2015).
- 117 Jiang J, White-Cooper H. Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of *aly* and a novel meiotic arrest gene *cookie monster*. *Development* 130(3), 563–573 (2003).
- 118 White-Cooper H, Schäfer MA, Alphey LS, Fuller MT. Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in *Drosophila*. *Development* 125(1), 125–134 (1998).
- 119 Lu D, Sin H-S, Lu C, Fuller MT. Developmental regulation of cell type-specific transcription by novel promoter-proximal sequence elements. *Genes Dev.* 34(9–10), 663–677 (2020).
- 120 Ing-Simmons E, Vaid R, Bing XY, Levine M, Mannervik M, Vaquerizas JM. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat Genet* 53(4), 487–499 (2021).

- 121 Espinola SM, Götz M, Bellec M *et al.* Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early *Drosophila* development. *Nat Genet* 53(4), 477–486 (2021).
- 122 Anderson J, Henikoff S, Ahmad K. Chromosome-specific maturation of the epigenome in the *Drosophila* male germline, (2023).
- 123 Mahadevaraju S, Fear JM, Akeju M *et al.* Dynamic sex chromosome expression in *Drosophila* male germ cells. *Nat Commun* 12(1), 892 (2021).
- 124 Alexander JM, Guan J, Li B *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* 8, e41769 (2019).
- 125 Benabdallah NS, Williamson I, Illingworth RS *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell* 76(3), 473-484.e7 (2019).

Список рисунков

Рисунок 1. Схематичное изображение основных этапов протокола Hi-C (адаптировано из <i>Cremazy et al, 2018</i>).	18
Рисунок 2. Межхромосомные взаимодействия.	20
Рисунок 3. Основные уровни внутривхромосомной организации интерфазного хроматина млекопитающих и дрозофилы (графика адаптирована из <i>Fujita, Yamashita, 2021</i>).	23
Рисунок 4. Особенности организации хроматина на уровне ТАДов у млекопитающих и дрозофилы.	25
Рисунок 5. Взаимосвязь между организацией хроматина дрозофилы и профилями различных модификаций гистонов, открытости хроматина и экспрессии на примере клеточной линии Kc167 (адаптировано из <i>Rowley, 2017</i>).	28
Рисунок 6. Схематичное изображение взаимодействия хроматина с ядерной ламиной и комплексом ядерной поры (адаптировано из <i>Shevelyov, 2020</i>).	37
Рисунок 7. Схематичное изображение стадий сперматогенеза дрозофилы (адаптировано из <i>Mills et al, 2019</i>).	43
Рисунок 8. Схема сперматогенеза в дрозофиле до стадии удлинённых сперматид.	46
Рисунок 9. Данные Hi-C для сперматогониев и сперматоцитов.	66
Рисунок 10. Разность значений PC1 между сперматоцитами и сперматогониями.	67
Рисунок 11. Сравнение характеристик регионов, перешедших в активный компартмент (В-А) и оставшихся в неактивном (В-В).	68
Рисунок 12. Структура компартментов.	69
Рисунок 13. Попарные взаимодействия генов из разных ТАДов.	71
Рисунок 14. Попарные взаимодействия генов внутри ТАДов.	71
Рисунок 15. Межхромосомные взаимодействия сперматоцитспецифичных генов.	72
Рисунок 16. Взаимодействия сперматоцитспецифичных генов в А- и В-компартементе на стадии сперматоцитов.	73
Рисунок 17. Кривые инсуляции вокруг стартов транскрипции.	76

Рисунок 18. Крупномасштабные изменения в структуре хроматина сперматогониев и сперматоцитов.	80
Рисунок 19. Средняя интенсивность контактов для объединения детектированных на каждой из стадий петель, в основаниях которых средняя экспрессия варьируется в пределах 10% между сперматогониями и сперматоцитами.	81
Рисунок 20. Данные Hi-C для эксперимента с деплецией Elys.....	82
Рисунок 21. Взаимодействия хроматина на уровне компартментов.	83
Рисунок 22. Изменение структуры компартментов, образованных взаимодействием участков генома, расположенных на разных расстояниях за пределами ТАДов.	84
Рисунок 23. Плотность ТАДов.	85
Рисунок 24. Средняя частота контактов ТАДов, разделенных разным расстоянием (подписи по оси X (q1, q2, q3) соответствуют квантилям расстояний).	86
Рисунок 25. Влияние деплеции Elys на структуру хроматина в границах.	88
Рисунок 26. Средняя частота контактов вокруг сайтов связывания Elys.	90
Рисунок 27. Модифицированные кривые инсуляции вокруг сайтов связывания. .	90
Рисунок 1 (приложение). Транскрипция генов в А- или В-компарменте, детектированных на разных разрешениях (стадия сперматогониев).	106
Рисунок 2 (приложение). Средняя инсуляция в сильных и слабых границах, рассчитанная по Hi-C картам для контроля (см. раздел 2.2.4).....	106
Рисунок 3 (приложение). Средняя частота контактов между парами сперматоцитспецифичных генов, выявленными с помощью алгоритма поиска значимых контактов для списка геномных локусов (см. раздел 2.2.5).	107
Рисунок 4 (приложение). Обогащенные GO-категории в трех онтологиях для генов, расположенных в выраженном В-компарменте, общем для сперматогониев и сперматоцитов.	107
Рисунок 5 (приложение). Фрагменты Hi-C карт сперматогониев (выше диагонали) и сперматоцитов (ниже диагонали) вокруг сперматоцитспецифичных границ, детектированных алгоритмом Armatus и содержащих промотеры сперматоцитспецифичных генов.....	108
Рисунок 6 (приложение). Средний индекс инсуляции в специфичных для контроля и для деплеции Elys границах.....	109

Список таблиц

Таблица 1. Таблица сопряженности для подсчета значимости ассоциации между образованием сперматоцитспецифичным геном значимого, отличного от фонового, контакта и частотой попадания в такие контакты в качестве пары другого гена из той же группы сперматоцитспецифичных при сравнении с убиквитными.	61
Таблица 2. Публичные данные.....	64
Таблица 3. Использованные программы и модули.	64
Таблица 4. Средняя частота межхромосомных взаимодействий, нормированная на общее число контактов.....	80

Таблица 1 (приложение). Средняя частота контактов вокруг петель и энхансерно-промоторных взаимодействий..... 109

Приложение

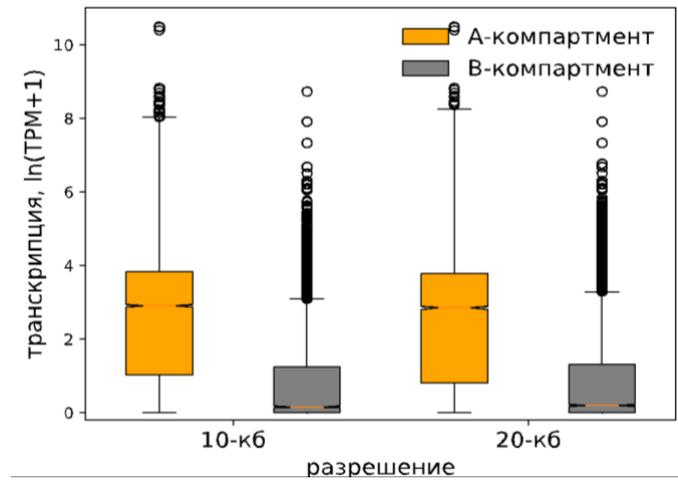


Рисунок 1 (приложение). Транскрипция генов в А- или В-компаратменте, детектированных на разных разрешениях (стадия сперматогониев).

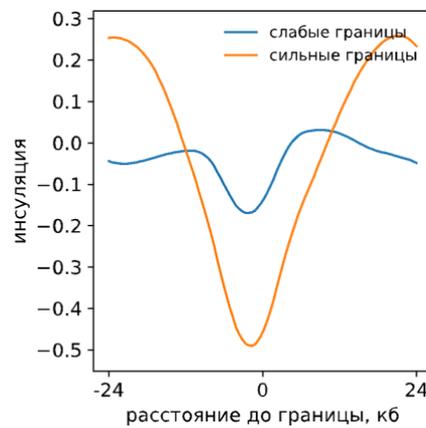


Рисунок 2 (приложение). Средняя инсуляция в сильных и слабых границах, рассчитанная по Hi-C картам для контроля (см. раздел 2.2.4)

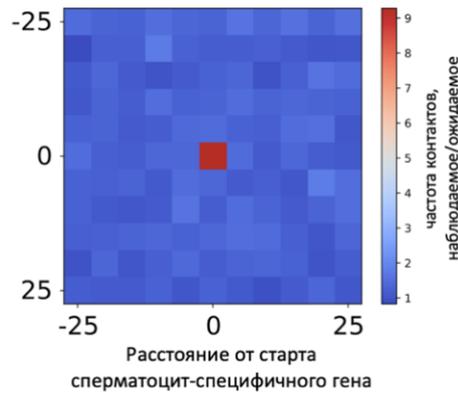


Рисунок 3 (приложение). Средняя частота контактов между парами сперматоцитспецифичных генов, выявленными с помощью алгоритма поиска значимых контактов для списка геномных локусов (см. раздел 2.2.5).

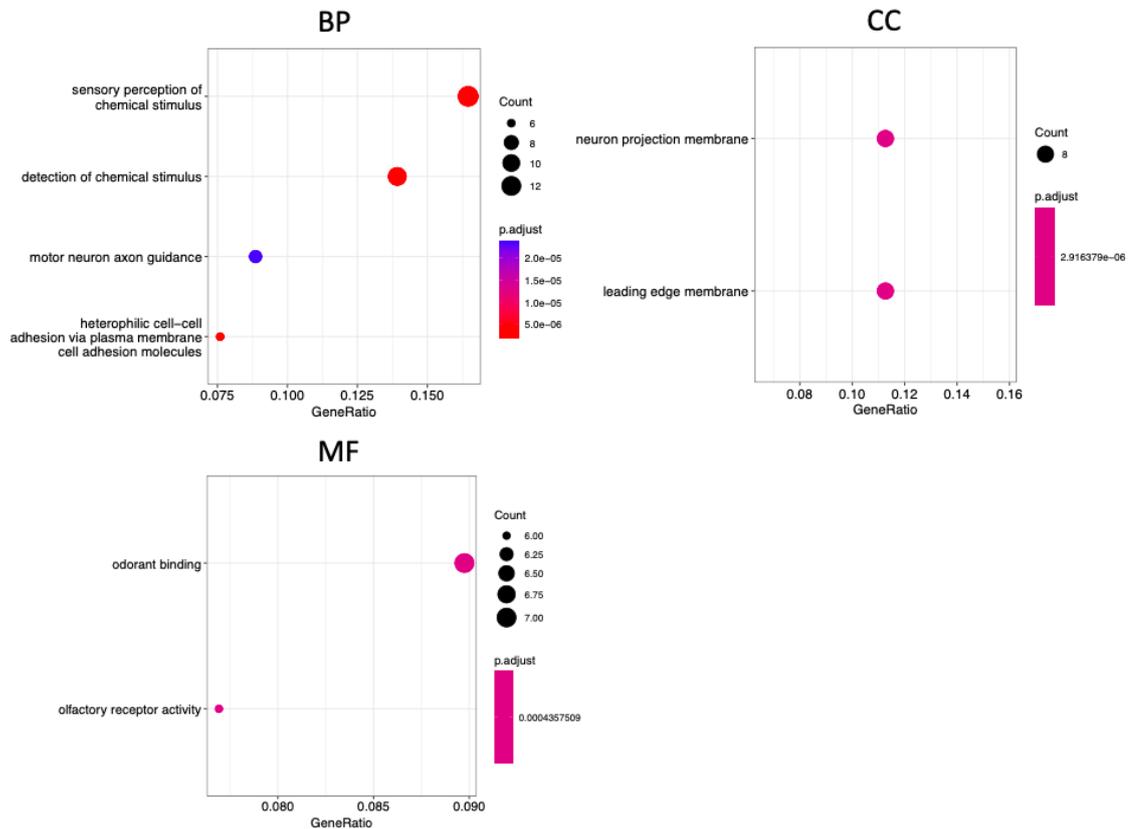


Рисунок 4 (приложение). Обогащенные GO-категории в трех онтологиях для генов, расположенных в выраженном В-компарменте, общем для сперматогониев и сперматоцитов.

BP – биологические процессы, CC – клеточные компоненты, MF – молекулярные функции

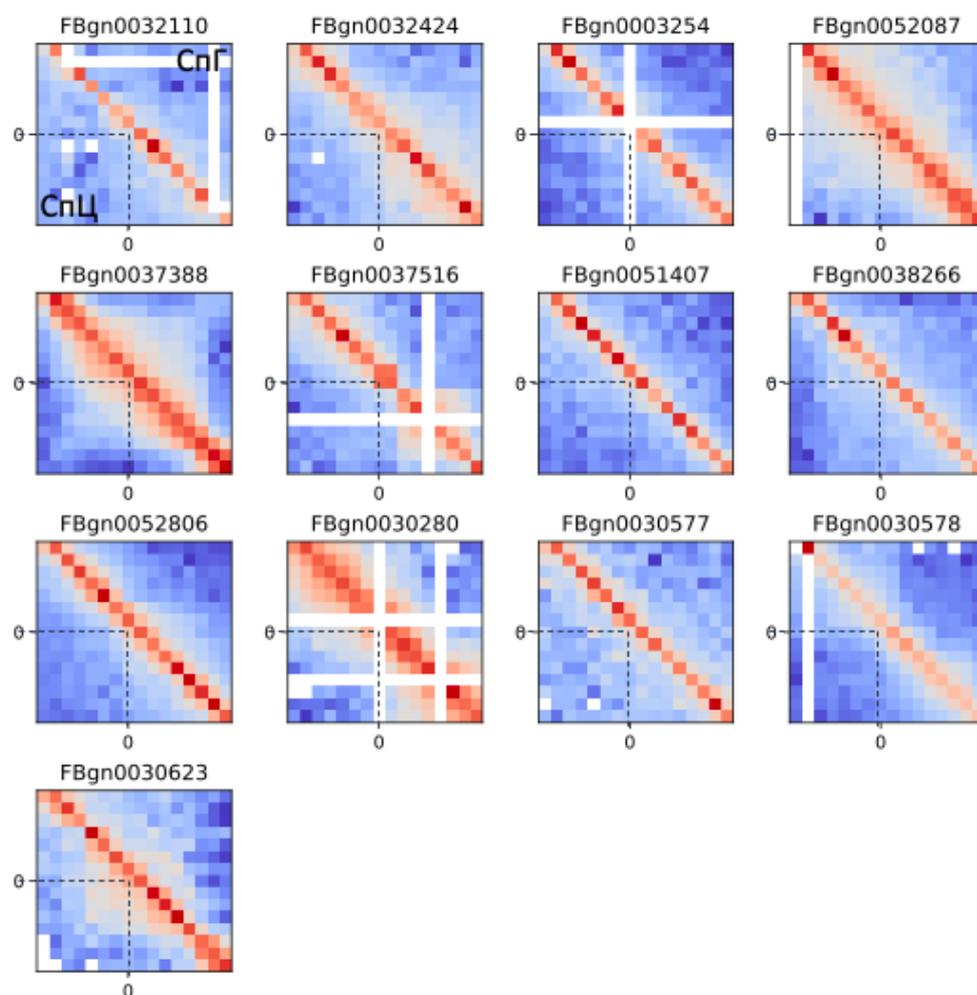


Рисунок 5 (приложение). Фрагменты Hi-C карт сперматогониев (выше диагонали) и сперматоцитов (ниже диагонали) вокруг сперматоцитспецифичных границ, детектированных алгоритмом Armatus и содержащих промотеры сперматоцитспецифичных генов.

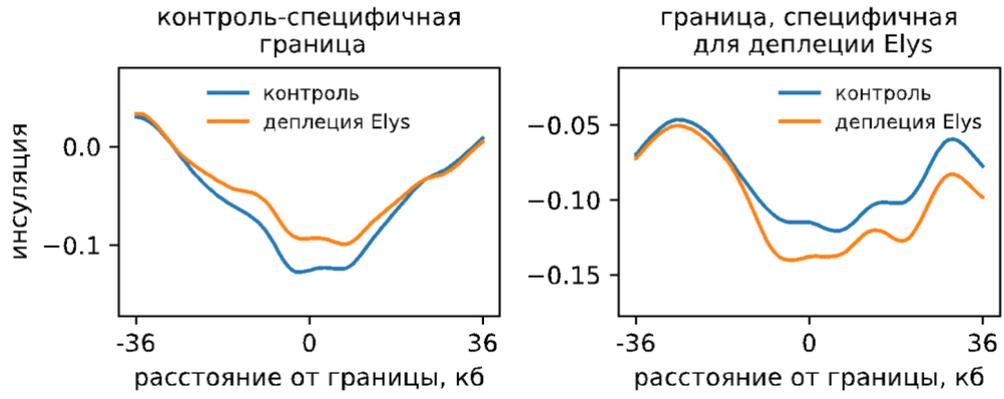


Рисунок 6 (приложение). Средний индекс инсуляции в специфичных для контроля и для деплеции Elys границах.

Таблица 1 (приложение). Средняя частота контактов вокруг петель и энхансерно-промоторных взаимодействий.

	контроль	деплегция Elys	разница
Петли, ассоциированные с сайтами Elys _{нукл}	<p>средняя петля, контроль</p>	<p>средняя петля, Elys-KD</p>	<p>средняя петля, Elys-KD - контроль</p>
Петли, ассоциированные с сайтами Elys _{япк}	<p>средняя петля, контроль</p>	<p>средняя петля, Elys-KD</p>	<p>средняя петля, Elys-KD - контроль</p>
Средний энхансерно-промоторный контакт	<p>средний энхансерно-промоторный контакт</p>	<p>средний энхансерно-промоторный контакт</p>	<p>средний энхансерно-промоторный контакт</p>