ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК»
(ИЦиГ СО РАН)

на правах рукописи

НУРИДДИНОВ МИРОСЛАВ АБДУРАХИМОВИЧ

РАЗРАБОТКА МЕТОДОВ ДЛЯ МЕЖВИДОВОГО СРАВНЕНИЯ ПРОСТРАНСТВЕННОЙ ОРГАНИЗАЦИИ ХРОМАТИНА

1.5.8 – математическая биология, биоинформатика
(биологические науки)

ДИССЕРТАЦИЯ
на соискание учёной степени кандидата биологических наук

Научный руководитель:
кандидат биологических наук,
Фишман Вениамин Семенович

Новосибирск – 2023

Содержание

1. Введение 4
Список используемых сокращений и терминов 10
2. Обзор литературы 12
2.1. Представления об архитектуре хроматина до Hi-C 12
2.2. Методы 3C-семейства: ключевые особенности и структуры данных 15
2.3. Архитектура хроматина млекопитающих 20
2.4. Архитектура хроматина *Drosophila melanogaster* 29
2.5. Молекулярно-биологические механизмы формирования архитектуры интерфазных хромосом. 34
2.6. Архитектура хроматина в современных исследованиях функционирования генома 40
2.7. Архитектура хроматина в эволюционном контексте 47
3. Данные и методы 54
3.1. Анализ архитектуры хроматина позвоночных 54
3.2. Анализ архитектуры хроматина комаров рода *Anopheles* 60
3.3. Конвертирование геномных координат между разными видами 63
4. Результаты 64
4.1. Характеристика пространственной организации генома *G. gallus* 64
4.2. Использования индекса VI для эволюционного сравнения организации хроматина позвоночных 74
4.3. Алгоритм сравнения пространственной организации хроматина, основанный на индивидуальных контактах 79
4.4. Пространственной организации хроматина комаров рода *Anopheles* 97
4.5. Выделение A/B-компартментов в геномах комаров рода *Anopheles* 99
4.6. Характеристика пространственной организации хроматина у комаров рода *Anopheles* 106
4.7. Исследование консервативности архитектуры хроматина комаров рода *Anopheles* 112

4.8. Изучение генетических особенностей эволюционных точек разрывов хромосом у представителей рода *Anopheles* 115
5. Заключение 121
6. Список литературы 124

1. Введение

Актуальность проблемы

Успехи в развитии методик захвата конформации хромосом – 3C (Chromosome Conformation Capture) – позволило получать данные о пространственной организации ДНК с недостижимым ранее разрешением [1,2]. Широкое применение одного из вариантов данного подхода — высокопроизводительного захвата конформации хромосом или Hi-C (High-throughput Chromosome Conformation Capture), позволило получить полногенномные данные об архитектуре хроматина [3,4].

Анализ результатов, полученных с помощью Hi-C, показал, что частота контактов удалённых участков генома распределена не случайно, и часто взаимодействующие участки формируют так называемые топологически ассоциированные домены (ТАДы). Дальнейшие исследования позволили предположить, что ТАДы являются структурной единицей организации хроматина в ядре клетки и участвуют в регуляции экспрессии генов [3,5].

В то же время, остаётся не известным ряд принципов формирования доменов и их универсальность для различных эволюционных линий. Сравнение пространственной организации хроматина между разными видами является удобным инструментов для выяснения особенностей механизмов укладки хромосом.

Стоит отметить, что основные данные по структуре ТАДов среди позвоночных на данный момент получены для различных клеточных линий *Homo sapiens* и *Mus musculus* [3–8]. Также большой объём сведений имеется для *Drosophila. melanogaster* [9,10], которая является представителем очень удалённой от млекопитающих эволюционной линии, что не позволяет проводить непосредственное сравнение организации хромосом. Иные же группы позвоночных, кроме млекопитающих, до недавнего времени не были изучены, что создавало пробел в понимании значимости наблюдаемых у млекопитающих свойств пространственной организации хроматина.

В соответствии с этим большой интерес представляют результаты Hi-C эксперимента, проведённого на разных типах клеток *Gallus gallus*. Среди представителей класса птиц *G. gallus* является наиболее хорошо изученным модельным объектом, находящим широкое применение в биологических и медицинских исследованиях. Сравнение архитектуры хроматина *G. gallus* с архитектурой хроматина других видов позвоночных способно дополнить наши знания о взаимодействии систем генной регуляции и формирования пространственной организации хроматина.

Принципы организации хроматина ещё менее изучены за пределами группы позвоночных животных. Одним из важных вопросов в этой области является исследование механизмов и закономерностей укладки генома у насекомых отряда *Diptera*. Многие представители этого отряда, например, комары рода *Anopheles*, являются переносчиками опасных инфекционных болезней человека, таких как малярия. Существуют указания, что разнообразие комаров рода *Anopheles* по их потенциалу к переносу заболеваний, приспособленности к питанию теми или иными субстратами и жизни в разных природных условиях связана с их высокой геномной пластичностью [11]. С учётом существующих представлений о влияние архитектуры хроматина на функционирование генома, эволюционное сравнение пространственной организации хроматина у различных представителей насекомых отряда *Diptera* способно дополнить наши представления в этой области. Необходимо добавить, что с каждым годом становится всё больше данных, описывающих архитектуру хроматина в разных типах клеток в разных таксономических группах, при этом методов, позволяющих проводить межвидовое сравнение пространственной организации хроматина решительно не хватает.

Таким образом, на данный момент существует необходимость в разработке таких подходов к изучению пространственной организации генома, которые позволили бы проводить эволюционное сравнение, что является мощным методом для изучения механизмов формирования и поддержания архитектуры хроматина.

Целью работы является разработка методов для эволюционного сравнения архитектуры хроматина. Согласно цели были поставлены следующие задачи:

1. с использованием данных Hi-C охарактеризовать архитектуру хроматина эритроцитов и фибробластов *G. gallus* и выявить связи между распределением различных хроматиновых структур и известных генетических и эпигенетических характеристик генома;
2. разработать методы сравнения Hi-C-данных, описывающих архитектуру хроматина разных видов;
3. с использованием разработанных методов сравнить архитектуру хроматина в клетках *G. gallus* с архитектурой хромосом, описанной ранее у разных видов млекопитающих;
4. сравнить архитектуру хроматина в ядрах клеток личинок пяти видов комаров рода *Anopheles*: *An. albimanus*, *An. atroparvus*, *An. stephensi*, *An. coluzzii* и *An. merus*.

Научная новизна и практическая ценность работы

Разработанный метод межвидового сравнения архитектуры хроматина на уровне отдельных контактов является первым методом такого рода, реализованный в виде пакета программ C-InterSecture. Изучена проблема определения эволюционной консервативности пространственной организации хроматина.

В ходе выполнения данной работы были впервые проанализированы данные Hi-C эксперимента для разных типов клеток *G. gallus*, проведённого в Отделе молекулярных механизмов онтогенеза ИЦиГ СО РАН, а также впервые проведено сравнение топологически ассоциированных доменов птиц и млекопитающих в полногеномном масштабе. Полученные результаты позволили выявить черты организации хроматина уникальные для *G. gallus* и резко отличающие этот вид от млекопитающих.

Получены новые сведения по архитектуре хроматина для пяти видов комаров рода *Anopheles*. Впервые было проведено межвидовое сравнение архитектуры

хроматина для указанной выше таксономической группы. Данные результаты будут использованы при дальнейшем изучении генетики данных организмов.

Использование пакета программ C-InterSecture представляется перспективным как дополнительное средство в изучении механизмов эволюции, регуляции генной экспрессии и прояснении взаимосвязей между известными мутациями и хромосомными перестройками с фенотипическими проявлениями.

Положение выносимые на защиту

1. Метод сравнения Hi-C-данных, разработанный и реализованный в виде пакета программ C‑InterSecture, позволяет проводить межвидовое сравнение архитектуры хроматина на уровне отдельных контактов и решать вопросы эволюционного консерватизма его пространственной организации у разных организмов.

2. Архитектура хроматина у разных видов позвоночных и комаров рода Anopheles демонстрирует эволюционную консервативность, выраженную в сохранении нормированной на геномное расстояние частоты контактов в синтенных локусах

Апробация работы

Результаты работы вошли в отчёты по грантам Президента России (МК-1630.2017.4), Российского Научного Фонда (№ 14-14-00131, руководитель Красикова А.В. и № 17-74-10143, руководитель Фишман В.С) и Российского Фонда Фундаментальных Исследований (№18-04-00668 Фишман В.С.). Результаты работы были доложены на 4 конференциях в виде устных докладов.

Разработанный алгоритм для сравнения архитектуры хроматина на уровне индивидуальных контактов выложен в открытый доступ (https://github.com/NuriddinovMA/C-InterSecture).

Личный вклад соискателя

Основные результаты, изложенные в диссертации, получены лично соискателем: разработаны алгоритмы сравнения архитектуры хроматина,

проведена их апробация с использованием результатов Hi-C экспериментов. Необработанные данные Hi-C для *G. gallus* предоставлены канд. биол. наук Баттулиным Н.Р., для комаров рода *Anopheles* – д-р. биол. наук Шараховым И. В. и Лукьянчиковой В. А., улучшение и сборка геномов комаров рода *Anopheles de nova* проводилась совместно с Лукьянчиковой В. А.

Публикации по теме диссертации

1. Veniamin Fishman, Nariman Battulin, **Miroslav Nuriddinov**, Antonina Maslova, Anna Zlotina, Anton Strunov, Darya Chervyakova, Alexey Korablev, Oleg Serov, Alla Krasikova. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin// Nucleic Acids Research. – 2019. – V. – 47. – I. 2. – P. 648–665. https://doi.org/10.1093/nar/gky1103

2. **Miroslav Nuriddinov**, Veniamin Fishman. C-InterSecture—a computational tool for interspecies comparison of genome architecture// Bioinformatics. – 2019. – V. 35. – I. 23. – P. 4912–4921. https://doi.org/10.1093/bioinformatics/btz415

3. Varvara Lukyanchikova, **Miroslav Nuriddinov**, Polina Belokopytova, Jiangtao Liang, Maarten J.M.F. Reijnders, Livio Ruzzante, Robert M. Waterhouse, Zhijian Tu, Igor V. Sharakhov, Veniamin Fishman. Anopheles mosquitoes revealed new principles of 3D genome organization in insects // Nature Communications – 2022. – V. 13. – I. 1 – P. 1960. https://doi.org/10.1038/s41467-022-29599-5

Объём и структура диссертации

Диссертация изложена на 158 страницах, содержит 28 рисунков, 6 таблиц и 2 приложения. Список литературы включает 220 ссылок. Диссертация состоит из введения, литературного обзора, описания результатов в двух главах, заключения, выводов и списка литературных источников.

Благодарности

д-р. биол. наук Шарахову Игорю Валентиновичу – за предоставление данных Hi-C для комаров рода *Anopheles*;

канд. биол наук Баттулину Нариману Рашидовичу – за предоставление данных Hi-C для разных типов клеток *G. gallus*;

канд. биол наук Красиковой Алле Валерьевне – за деятельное участие в обсуждениях особенной организации хроматина в разных типах клеток *G. gallus*;

Лукьянчиковой Варваре Алексеевне – за предоставление данных Hi-C для комаров рода *Anopheles*, помощь в проведении сборок генома и обсуждении особенностей организации хроматина комаров рода *Anopheles*.

Список используемых сокращений и терминов

Сокращения:

3C – chromosome conformation capture

CNE – conservative non-coding elements

Hi-C – high-throughput chromosome conformation capture

P-BAD - percentile-based background-adjusted distance

VI – variation information

ЛАД – ламин-ассоциированный домен

ТАД – топологически ассоциированный домен

ПО – программное обеспечение

ЭСК – эмбриональные стволовые клетки

п.о. – пар азотистых оснований

Термины:

Величина контакта (между двумя локусами) – количество прочтений в библиотеке Hi-C, которые, разными своими частями, выравниваются одновременно на оба локуса. В некоторых случаях, используется для описания нормализованной величины, полученной на основе числа прочтений.

Абсолютная частота контактов – частота взаимодействия между выбранными локусами нормализованная таким образом, чтобы сумма взаимодействий целевого локуса со всем геномом была равна 1.

Относительная частота контактов – частота взаимодействия между выбранными локусами нормированная на среднее значение частоты взаимодействия локусов, находящихся на таком же геномном расстоянии.

Бин – участок генома заданной протяжённости; все бины имеют одинаковую длину и непрерывны, каждый участок генома включён в тот или иной бин.

Разрешение (карты Hi-C) – выбранный размер бина в нуклеотидах.

Перекартирование – основанное на синтении сопоставление геномной характеристики (координаты, экспрессии, эпигенетической метки, величины контакта и т.п.) между разными геномами (разными сборками геномов одного вида или геномами разных видов). Термин «перекартирование» также применяется при предсказании характеристик одного генома на основе гомологичной характеристики, измеренной в другом геноме.

Точка синтении – короткий, от 100 до 200 п.о. фрагмент, гомологичный между сравниваемыми видами.

2. Обзор литературы

2.1. Представления об архитектуре хроматина до Hi-C

Уже в самых первых исследования поведения хромосом в течение клеточного цикла было выявлено, что разные участки хромосом ведут себя по-разному. Это наблюдение привело к введению таких терминов как «гетерохроматин» и «эухроматин» для описания структуры хромосом [12]. Тогда же было выдвинуто предположение, что эухроматин представляет собой участки хромосом, насыщенные активными генами, а гетерохроматин – генетически неактивные или вовсе лишённые генов. Однако вскоре стало понятно, что предположение о генетической инертности гетерохроматина является не совсем верным, так как гены, лежащие в гетерохроматизированных локусах, участвовали в процессах развития [13]. В ходе последующего изучения судьбы гетерохроматизированных локусов при клеточной дифференциации, было выявлено, что гетерохроматизированность того или иного локуса определяется клеточным типом. Эти наблюдения подтвердили гипотезу о связи состояния хроматина с регуляцией активности генов и привели к введению терминов «конститутивный гетерохроматин» и «факультативный гетерохроматин» [14].

Определённое представление о взаимосвязи между состоянием хроматина и активностью генов удалось обнаружить в ходе экспериментов по изучению фенотипа особей *Drosophila melanogaster*, несущих хромосомные перестройки. Было показано, что реализация генетической информации определяется близостью гена к гетерохроматизированному локусу [9,10]. Таким образом, состояние хроматина локуса определяет активность генов в данном локусе.

Более детальное изучение структуры хроматина методами биохимии и электронной микроскопии позволило открыть, что его основой является взаимодействие молекулы ДНК с белками-гистонами, а формирующаяся таким образом нуклеосома является наименьшей структурной единицей хроматина [17]. Объединение данных по рентгеноструктурному анализу и электронной микроскопии позволило выдвинуть гипотезу о нескольких уровнях организации

хроматина в клеточном ядре [18–20]. В последующее время были накоплены многочисленные свидетельства того, что ни нуклеосомы, ни хроматин не являются стабильными и неизменными структурами. Динамика структуры и физико-химических свойств хроматина, нуклеосом и гистонов непосредственно влияет на способность белков активаторов и супрессоров связываться с целевыми локусами и регулировать активность генов [21,22].

Одновременно с этим были накоплены многочисленные свидетельства, что хромосомы в интерфазном ядре расположены не случайно и каждая хромосома, занимает свою относительно обособленную от других хромосом часть пространства ядра [23–25]. При этом, расположение хромосом и их частей в клеточном ядре связано с активностью генов [26]. Более того, отдельные части хромосомы так же разделены в пространстве ядра. Существует группа белков – белков ламины - отвечающих за связывание хроматина и ядерной оболочки. При этом участие тех или иных белков ламины имеет регуляторное значение. Так, подавляется экспрессия генов на регионах хромосомы, которые участвуют в формировании ламины [27]. Более детальные исследования показали, что хромосомные регионы связывания с ламиной формируют длительные участки – ламин-ассоциированные домены (ЛАДы) – протяжённостью 0,1-10 млн п. о. Данные регионы, выделенные по повышенной частоте связывания с белком Lamin B1, обладают целым рядом свойств, демонстрирующих их участие в регуляции активности генов. Регионы, расположенные внутри ЛАДов, обладают существенно меньшей плотностью генов, насыщенностью метками активной транскрипции, в них существенно ниже уровень экспрессии. Особыми свойствами обладают и границы ЛАДов: повышенной долей CpG-островков, сайтов связывания CTCF и промоторов генов, направленных в сторону противоположную ЛАДу. Стоит отметить, что данные свойства не являются определяющими для формирования границы ЛАДа, так как только 30% от всех границ обладает хотя бы одним из отмеченных свойств [28]. Ещё одним фактом, подтверждающим участие взаимодействия хроматина с ламиной является то, что в ходе клеточной дифференциации изменяется профиль распределения белка Lamin B1 и,

соответственно, ЛАДов, что приводит к изменению экспрессии генов, связанных с определением клеточной судьбы. Примечательно, что если попадание генов в ЛАД жестко связано с подавлением экспрессии генов, то выход гена за пределы ЛАДа не означает *немедленной* его активации, а чаще делает ген доступным для активации на более поздних стадиях пролиферации [29].

Отдельно стоит отметить, что архитектура хроматина в течение клеточного цикла обеспечивается функционированием белковых комплексов конденсин I, конденсин II и когезин. При этом комплексы конденсин I, конденсин II связаны с конденсацией хроматина в профазе и прометафазе митоза соответственно, а когезин – отвечает за связывание сестринских хроматид в интерфазе. Однако более детальные исследования роли когезина показывают, что его роль в функционировании генома более сложная и включает в себя участие в регуляции генов [30–32], сближение , удалённых участков генома при содействии белка CTCF [33] и конденсация хроматина в интерфазе [32].

Совокупность данных, говорящих о связи организации хроматина в пространстве клеточного ядра и регуляции генов, требовала более детального изучения архитектуры хроматина и взаимосвязей между разными уровнями организации как факторов ответственных за корректную реализацию наследственной информации [34]. Однако, существующие на тот момент методы не позволяли с достаточной эффективностью проводить исследования пространственной организации хроматина. Так, методы световой микроскопии имеют недостаточное разрешение. При использовании методов электронной микроскопии затруднена привязка обнаруженных особенностей хроматина к конкретному локусу генома. Метод флуоресцентной *in situ* гибридизации ограничен количеством локусов, для которых можно проводить исследования одновременно, кроме этого используемые реагенты взаимодействует с хроматином, вызывая изменения его структуры. Для решение указанных проблем был разработан метод захвата конформации хромосом (3C), который позволяет оценивать пространственную близость двух любых избранных локусов генома [1].

2.2. Методы 3C-семейства: ключевые особенности и структуры данных

Первым методом, который позволил систематически получать информацию о пространственной организации хроматина был метод 3C. Не углубляясь в детали протокола 3C, необходимо отметить три ключевые для данного метода шага (Рисунок 1) [1] :

1) фиксация положения хроматина в пространстве ядра (Рисунок 1Б);

2) случайное разрезание зафиксированной молекулы ДНК на небольшие фрагменты (Рисунок 1В);

3) случайное сшивание полученных фрагментов ДНК друг с другом и получение библиотеки химерных молекул, то есть таких молекул, которые соединяют вместе нуклеотидные последовательности, разделённые в геноме (Рисунок 1Г).

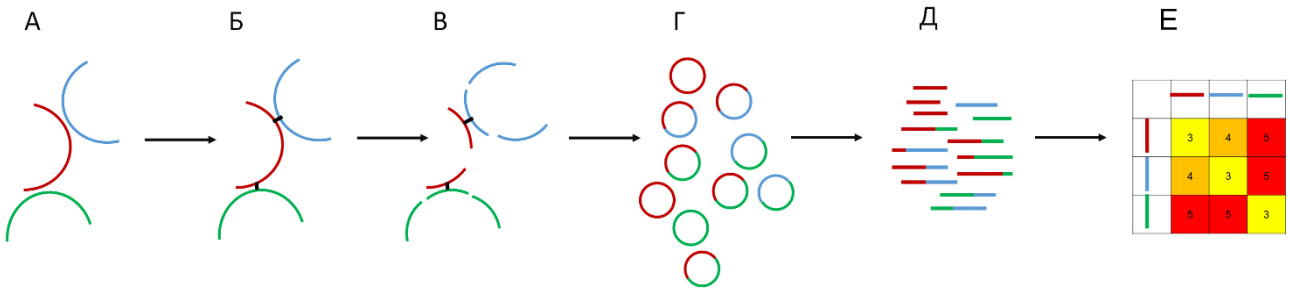

Рисунок 1. Схема Hi-C эксперимента. Цветом показаны разные участки генома. А. положение хроматина в ядре. Б. Фиксация хроматина. В. Случайное разрезание. Г. Сшивание полученных молекул. Д. Библиотека химерных молекул. Е. Карта Hi-C

Предполагается, что в ходе сшивания фрагментов ДНК, соединяться в единую химерную молекулу будут те фрагменты, которые находятся близко друг к другу в пространстве клеточного ядра независимо от их положения в геноме. В соответствии с этим, локусы, которые систематически чаще оказываются близко друг к другу, будут давать б*о*льшее количество химерных молекул ДНК в совокупной библиотеке, чем пространственно удалённые. При последующем анализе полученных библиотек, под **контактом** двух локусов генома понимается

наличие в библиотеке химерной молекулы, несущий нуклеотидные последовательности, принадлежащие этим двум локусам (Рисунок 1Д); а под **частотой контакта** – количество таких молекул или иная, производная величина (Рисунок 1Е).

Явным недостатком метода 3С является то, что он в одном эксперименте позволяет оценивать пространственную близость только двух выбранных локусов. Поиск способов получать за один эксперимент как можно больше информации о взаимном положении целевых локусов привёл к появлению целого семейства методов захвата конформации хромосом, наибольший интерес среди которых, в рамках данного исследования, представляет метод Hi-C.

Этот метод наследует от своего «родителя» 3С вышеописанные три шага (фиксация-разрезание-сшивание), но позволяет оценивать пространственную близость уже не двух предварительно избранных локусов, а всех локусов всего генома, тем самым получая глобальную, полногеномную информацию о пространственной организации хроматина [3]. Возможность полногеномного исследования в рамках метода Hi-C достигается за счет секвенирования всех химерных молекул, образовавшихся в ходе 3С-эксперимента, на основе технологии секвенирования нового поколения.

Идеология метода Hi-C, как полногеномного анализа пространственной организации хроматина, предполагает отсутствие предварительно избранных локусов интереса: интерес представляют все локусы всего генома. Не менее важно обеспечить удобство в сопоставлении как пространственной организации хроматина разных локусов друг с другом, так и данных полученных в разных экспериментах. В соответствии с этим, весь геном, механистически, в соответствии с геномными координатами, разбивается на локусы фиксированной длины – **бины.** Разбиение на бины никоим образом не учитывает особенности генома в каждом конкретном месте, например, наличие повторов или не описанных участков, и границы бинов задаются только геномными координатами. Таким образом, данные Hi-C удобно представлять в виде квадратной матрицы NxN, где N – количество бинов, на которые разбит весь геном, а элемент матрицы расположенный на

пересечении i-ой строки и j-го столбца (или наоборот) соответствует частоте контактов между i-м и j-м бинами, т. е. числу химерных молекул, содержащих фрагменты генома i-го и j-го бинов. **Покрытием** i-го бина, в таком случае, называется сумма всех значений в i-ой строке (столбце), или, иными словами, сумма частот всех контактов, в которых целевой бин участвует. Так как данные Hi-C эксперимента являются полногеномными, они содержат контакты между бинами расположенными как на одной, так и на разных хромосомах. Контакты между бинами разных хромосом обычно называют *транс*-контактами. В свою очередь контакты между бинами, расположенными на одной хромосоме – *цис*-контакты.

Для визуального исследования данных Hi-C, полученные матрицы или отдельные их фрагменты удобно представлять в виде тепловых карт – карт Hi-C. В свою очередь, чем меньший размер бина будет избран, тем более мелкие особенности пространственной организации хроматина потенциально возможно изучить. В соответствии с этим, **разрешением** карты Hi-C называется выбранный размер бина. Однако нужно учитывать, что чем меньше в библиотеке химерных нуклеотидных последовательностей, тем меньше будет выравниваться в среднем последовательностей на бин и тем меньше информации о пространственной организации возможно получить.

Естественно, анализ библиотек Hi-C, построение на их основе матриц контактов и тепловых карт представляет собой сложную и объёмную вычислительную задачу. Для её решения был создан ряд пакетов программ, среди которых, как наиболее популярные, стоит отметить HiC-Pro [35], Cooler [36] и Juicer [37,38]. Эти и подобные ПО используют сходные принципы анализа библиотек Hi-C и основные различия между ними лежат, скорее, в удобстве для пользователя и спектре предоставляемых возможностей по работе с полученными матрицами контактов.

Уже самые первые 3C эксперименты показали, что проведение такого рода экспериментов чувствительны к разного рода техническим артефактам и требуют высокого контроля [39]. В полной мере эту особенность унаследовал и метод Hi-C,

для которого характерен ряд систематических ошибок, связанных с локальными особенностями генома [40]. Наиболее важными из них являются:

- неравномерная доступность молекулы ДНК для посадки ферментов рестрикции (или других ферментов, обеспечивающих разрезание молекулы ДНК);
- значительные различия в GC-составе разных бинов;
- неравномерная встречаемость повторов и других неуникальных последовательностей.

Для того, чтобы уменьшить влияние систематических и случайных ошибок на результат, было разработано множество методов для нормализации данных Hi-C, которые можно поделить на две основные группы: использующие информацию о природе систематических ошибок и не использующие.

Первый метод, предложенный Yaffe и Tanay в 2011, основывался на детальном изучении источников систематических ошибок и учёте их вклада в величину контакта с помощью вероятностной модели [40]. Использование данного метода позволило значительно повысить сходимость результатов между Hi-C экспериментами, проведёнными с использованием разных ферментов рестрикции (HindIII и NcoI). Так, например, коэффициент корреляции между репликами для частот *транс*-контактов, которые ввиду своей редкости наиболее чувствительны к систематическим ошибкам и экспериментальному шуму, увеличился с -0.11 до 0.59, что указывает на высокую эффективность данного метода.

Недостатками алгоритма YT оказалось большая вычислительная сложность и использование 420 параметров для проведения нормализации. Этих недостатков лишён алгоритм HicNorm [41]. Данный алгоритм основан на тех же подходах, что YT, однако использование регрессии Пуассона для оценки вклада известных систематических ошибок в результат, позволило ограничиться в работе алгоритма всего тремя входными параметрами и увеличить скорость проведения нормализации в несколько тысяч раз. С точки зрения качества нормализации, выражаемой в сходимости реплик, алгоритм HicNorm крайне незначительно превосходит YT, лишь в отдельных случаях демонстрируя большую эффективность.

Следующие методы нормализации не учитывают локальные особенности генома. В их основе лежит положение, что частоты контактов отражают вероятности взаимодействия бина с его пространственным окружением, а так как сумма вероятностей должна равняться 1, то покрытие каждого бина должно быть константой. Соответственно, все отклонения являются следствием разнообразных ошибок не всегда известной природы.

Самым первым и простым методом стала нормализация контактов на покрытие бинов [3,37]. В этом методе предполагается, что все различия в покрытии обусловлены ошибками и вклад ошибок в величины контактов пропорционален избыточности (или недостаточности) покрытия бинов. В соответствии с этим, для нормализации частота каждого контакта делится на произведение покрытий контактирующих бинов.

Следующие методы основаны на применении алгоритмов, разработанных для балансировки квадратных неотрицательных матриц. Данная задача известна уже более 70 лет и её решения находят применение в самых разных отраслях науки и техники [42]. Основанные на этом подходе алгоритм ICE [43] реализован в ПО Cooler, а алгоритм Knight-Ruiz (KR) [37,42] в ПО Juicer.

В первые же годы после появления этого метода было получено значительное количество детальных карт, описывающих пространственную организацию хроматина, для целого ряда клеточных типов и организмов, таких как *Homo sapiens* [3,5,7,8], *Mus musculus* [5,6,44], *Drosophila melanogaster* [9,10], а также для целого ряда других живых организмов, включая млекопитающих [45], некоторые грибы [46], растения [47,48] и бактерии [49].

Дальнейшее, более детальные эксперименты Hi-C позволили раскрыть на более глубоком уровне раскрыть механизмы формирования пространственной хроматина в разных типах клеток и таксономических группах.

2.3. Архитектура хроматина млекопитающих

Одни из наиболее полных и детальных сведений по архитектуре хроматина были получены для млекопитающих, благодаря глубочайшей изученности генетических и эпигенетических характеристик генома двух популярных модельных организмов: *Homo sapiens* и *Mus musculus.*

Самый первый эксперимент Hi-C был проведён на двух разных линиях культивируемых клеток человек: лимфобластоидной (GM06990) и эритролейкимической (K562) [3]. Данные этих экспериментов показали, что частота контактов обратно-пропорциональна расстоянию между контактирующими локусами. Этот факт хорошо согласуется с результатами моделирования укладки хроматина по типу «фрактальной глобулы». Согласно этой модели, интерфазную хромосому можно описать как «глобула глобул из глобул». Кроме этого, в данной работе было проведено сравнение рассчитываемых по данным Hi-C частот контактов с пространственным расстоянием между контактирующими локусами, измеряемым методом флуоресцентной иммунопреципитации хроматина. Высокий коэффициент корреляции по Спирману ($r = –0.916$, $p = 0.00003$), показал достаточно строгое соответствие между частотами контактов и физическим расстоянием между локусами в пространстве ядра.

При визуальном изучении полученных карт Hi-C привлекает внимание то, что разные локусы генома имеют предпочтения в партнёрах при создании контактов с ними, тем самым формируя паттерн «шахматной доски» (Рисунок 2A). Данное наблюдение позволяет предположить, что пространственная организация хромосом отражает деление хроматина на два типа, таким образом, что локусы, принадлежащие одному типу хроматина, предпочитают контакты с друг с другом и избегают контактов с локусами другого типа хроматина.

Проверка этой гипотезы требовала убрать эффекты, связанные с расстоянием между контактирующими локусами. Для этого величина контакта между локусами была поделена на среднее значение для всех контактов между всеми локусами генома, разделённых таким же геномным расстоянием. Полученная величина –

отношение наблюдаемого числа контактов к ожидаемому – отражала взаимное предпочтение локусов к формированию контактов. Если эта величина больше единицы - локусы чаще контактируют друг с другом и оказываются вместе в пространстве ядра, чем это ожидается для произвольных локусов. Если величина меньше единицы - локусы избегают друг друга.

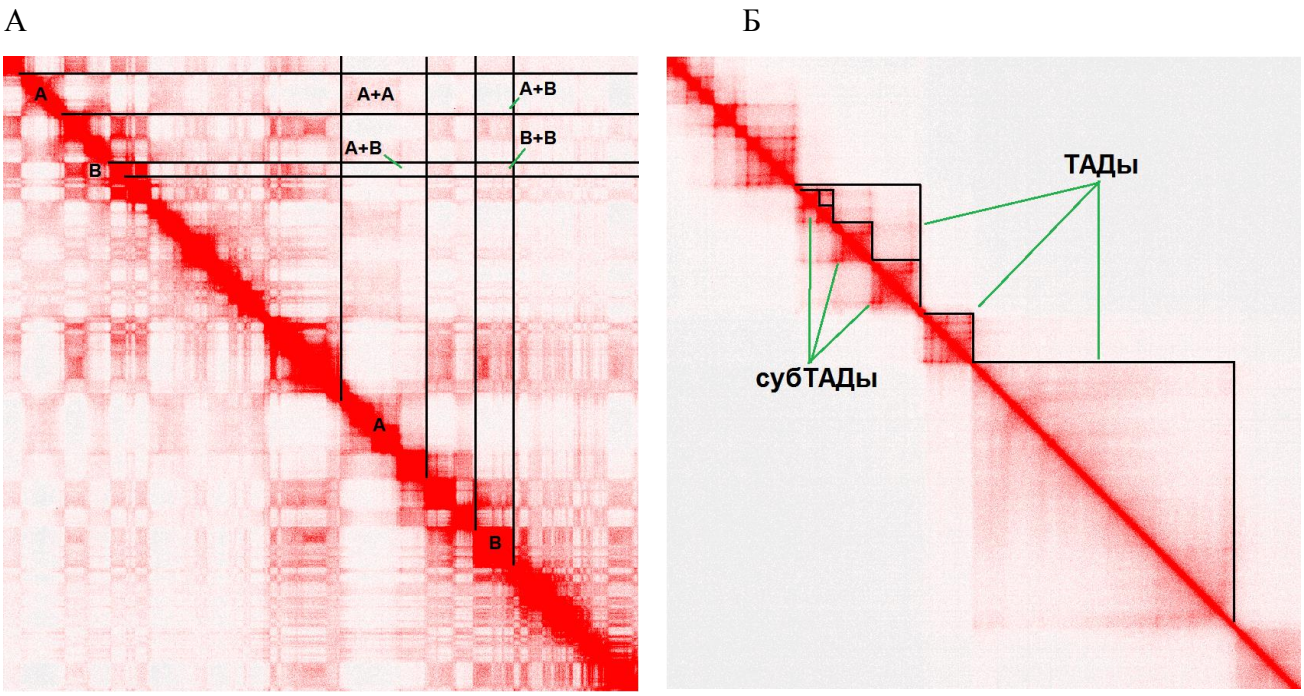

Рисунок 2. Пример карт Hi-C [5]. А. Деление на A/B-компартменты. Участки одного типа компартмента контактируют друг с другом достоверно чаще. Б. Иерархическая структура ТАДов.

На следующем этапе, были посчитана корреляция Пирсона для полученных матриц контактов. Если коэффициент корреляции для двух локусов близок к 1, то данные локусы предпочитают формировать контакты с одними и теми же локусами и избегают одни и те же локусы. И наоборот, если коэффициент корреляции близок к -1, то предпочтения сравниваемых локусов противоположны. Визуальный осмотр построенных на основе матриц корреляции тепловых карт подтвердил разбиение хроматина на группы по предпочтению формированию контактов. Наконец, для полученных матриц корреляции был проведён анализ главных компонент. Полученные в результате анализа значений первой главной компоненты позволили

приписать каждому локусу принадлежность к одной из двух групп, по избирательности формирования контактов. Эти группы были названы А- и В-компартменты, а величину первой главной компоненты, можно назвать величиной компартментализации.

Таким образом, было открыто, что хроматин, согласно пространственным контактам, делится на два компартмента так, что локусы, принадлежащие одному компартменту, предпочитают формировать контакты друг с другом и избегают контактов с локусами из другого компартмента.

Естественно возникает вопрос, какие конкретно молекулярно-биологические особенности хроматина стоят за такой его пространственной организацией. Для ответа на этот вопрос, величины компартментализации были сравнены с известными молекулярно-биологическими и генетическими характеристиками генома исследуемых типов клеток. Оказалось, что деление на компартменты отражает деление на хроматин активный, насыщенный генами (эухроматин) и неактивный (гетерохроматин). В дальнейшем было принято, что А-компартментом будет называться такой компартмент, который соответствует активному хроматину, и наоборот для В-компартмента.

Примечательно, что локусы В-компартмента, по данным Hi-C, сильнее контактируют друг с другом, чем локусы А-компартмента, что говорит о более плотной упаковке в пространстве ядра В-компартмента. Данное наблюдение было подтверждено результатам флуоресцентной иммунопреципитации хроматина.

Следующий (на млекопитающих) эксперимент Hi-C, был проведён на pro-B клетках *M. musculus* [50]. В этой работе исследовалась взаимосвязь между внутри- и межхромосомными транслокациями и состоянием хроматина в их границах. Для этого были созданы генетические конструкции, которые позволяли генерировать в случайных местах двухцепочечные разрывы ДНК, что стимулировало прохождение транслокаций. Благодаря этому, каждый клон получал свою уникальную перестройку в произвольном месте геноме. А для изучения особенностей мест, соединённых в ходе транслокации, были получены в несколько раз более подробные, чем в предыдущей работе, карты Hi-C.

Результаты исследования подтвердили все ранее сделанные выводы о пространственной организации хроматина. Помимо этого, была показана строгая корреляция частоты контактов между локусами и частотой транслокаций, соединяющих эти локусы (R = 0.7-0.8) для внутрихромосомных транслокаций. Аналогичные результаты получены и для межхромосомных транслокаций. Таким образом, можно говорить о том, что пространственная близость влияет на прохождение транслокаций.

Следует отметить, что вышеописанные результаты были получены без применения каких-либо алгоритмов нормализации данных Hi-C. Повторный анализ результатов Hi-C эксперимента, проведённого на лимфобластоидной линии *H. sapience*, позволил выделить ключевые источники систематических ошибок и использовать это для нормализации матриц контактов [40]. Анализ нормализованных матриц контактов показал, что помимо общей укладки хроматина, связанной с делением генома на A/B-компартменты, есть локальные особенности, прослеживаемые на небольших геномных расстояниях. Так было показано, что регионы, расположенные на расстоянии от 20 до 200 тысяч п.о. от старта транскрипции активных генов, контактируют со стартом транскрипции в ~1,5-1,7 раза чаще, чем в среднем по геному для бинов, разделённых данных расстояний. Для стартов транскрипции неактивных генов этот эффект выражен меньше и обогащение контактами не превышает 10-15% по сравнению со средним. Аналогичный результат был показан для окрестностей сайтов связывания белка CTCF: локусы, расположенные на расстоянии 40-400 тысяч п.о. от целевого сайта контактировали с ним на 30-50% чаще, чем ожидалось.

Использование метода Hi-C позволяет систематически детектировать взаимодействия промоторов и энхансеров в масштабах всего генома [6]. Исследования пространственной организации хроматина на нейронах коры головного мозга *M. musculus* позволило обнаружить, что энхансеры и промоторы в пределах одного регуляторного блока контактируют в 1,5-2 раза чаще, чем промоторы и энхансеры, расположенные в разных регуляторных блоках, и в 2-3 раза чаще, чем для случайных локусов. Аналогичные результаты были получены

при сравнении архитектуры хроматина эмбриональных стволовых клеток (ЭСК) и кортикальные нейронов мыши [51]

Полученные в выше указанных исследованиях результаты, показывают, что, во-первых, механизмы укладки хроматина у разных видов млекопитающих обладают определённым сходством. Во-вторых, имеются локальные особенности хроматина, формирующие взаимодействия локусов на расстоянии от нескольких десятков до нескольких сотен тысяч п.о. и предположительно связанные с деятельностью белков CTCF и регуляцией экспрессией генов.

Топологически ассоциированные домены

Детальное изучение локальных особенностей пространственной организации хроматина и сравнения его архитектуры в разных типах клеток и у разных видов требовало построения карт Hi-C большего разрешения. Результаты таких экспериментов, проведённых на ЭСК *M. musculus* и *H. sapiens* и клеточной линии IMR90 *H. sapiens* позволило выделить у млекопитающих новый уровень организации хроматина – топологически ассоциированные домены (ТАДы), представляющие локальную группу тесно контактирующих локусов генома, совместно изолированных от ближайшего окружения (Рисунок 2Б). Сравнение расположения ТАДов у указанных выше видов и клеточных линий показало их эволюционную консервативность [5].

Как уже отмечалось ранее, было замечено, что промоторы генов и сайтов связывания белка CTCF формируют повышенное число контактов со своим окружением. Чтобы вычленить систематические особенности локальной организации хроматина была разработан индекс направленности (DI - directionality index). Данный индекс отражает предпочтение целевого бина контактировать с бинами идущими после него по геномным координатам или до него. Если $DI > 0$, бин предпочитает контакты с последующими по геномным координатам локусами, если $DI < 0$, бин предпочитает контактировать с предыдущими по геномным координатам локусам.

Значения индекса направленности были проанализированы с помощью скрытых марковских цепей, что позволило систематически выделять границы

ТАДов как регионы, в которых происходит резкая и значительная перемена знака DI и сами ТАДы как регионы, внутри которых бины контактируют друг с другом чаще, чем с районами вне данного ТАДа.

Анализ свойств границ ТАДов показал, что эти регионы характеризуются в два раза большим числом сайтов связывания CTCF, чем в среднем по геному. Также, границы ТАДов обогащены в 2-3 раза стартами транскрипции генов домашнего хозяйства и эпигенетическими метками активного хроматина (например, модификацией гистонов H3K3me3) и, соответственно, обеднены метками гетерохроматина.

Стоит отметить, что обогащённость сайтами связывания CTCF и стартами транскрипции генов сближает по свойствам границы ТАДов и ЛАДов, однако менее 50% границ ТАДов совпадает с границами ЛАДов. Более того, обращает на себя внимание, что если у ТАДов около 70-80% границ обогащены сайтами связывания CTCF, то у ЛАДов таким свойством обладает менее 15% границ. Кроме этого, границы ЛАДов достаточно строго привязаны к границам состояний хроматина, в то же время границы ТАДов лишь приблизительно в 20% случаев соответствуют смене состояний хроматина. Таким образом, в отличие от ЛАДов и A/B-компартментов, строго связанных с эпигенетическим состоянием хроматина, ТАДы преимущественно задаются распределением сайтов связывания белка CTCF. Однако важно указать, что только 15-20% сайтов CTCF находятся на границах ТАДов, таким образом нельзя утверждать, что граница ТАДов задаются *только* белком CTCF.

Сравнение положения границ ТАДов между разными типами клеток и между разными видами показало, что около 50-70% границ ТАДов являются консервативными. Необходимо указать, что между разными видами и типами клеток количество выделенных границ ТАДов довольно-таки сильно варьирует, и доля консервативных границ может сильно уменьшаться, если за основу брать образец с большим количеством ТАДов. Тем не менее ТАДы остаются консервативны даже при сравнении соматических клеток со сперматозоидами [44]. Ни высокая плотность укладки хроматина, ни замена гистонов протаминами, ни

отсутствующая транскрипция не являются достаточными факторами для создания значимых различий в архитектуре хроматина.

Исследование, акцентированное на изучении ТАДов и контактов между ТАДами, показало, что архитектура хроматина является иерархической структурой, при которой ТАДы могут содержать в себе «субТАДы» и, в свою очередь, входить в состав больших «метаТАДов»[52]. Что важно, такая иерархическая организация подтверждается прямым измерением линейного расстояния с помощью флуоресцентной иммунопреципитации хроматина.

Вся совокупность данных позволяет характеризовать ТАДы как неизвестный ранее уровень пространственной организации хроматина, практически не связанный с эпигенетическим состоянием хроматина и эволюционно-консервативный в разных линиях млекопитающих. Участие белка CTCF и генов домашнего хозяйства в формировании границ доменов и изменчивость положения ТАДов для разных типов клеток позволяет предполагать, что ТАДы могут быть связаны с регуляцией генной экспрессии и определением клеточной судьбы.

Хроматиновые петли

Изучение архитектуры хроматина на данных Hi-C с разрешением порядка нескольких тысяч п.о. позволило обнаружить CTCF-опосредованные хроматиновые петли [8]. На основе девяти клеточных линий *H. sapiens* и *M. musculus* был проведён 201 эксперимент Hi-C с использованием трёх разных экспериментальных протоколов, позволивший сгенерировать множество технических и биологических реплик. Полученный объём данных Hi-C подтвердил сходимость результатов эксперимента для разных реплик и деление генома на A/B-компартменты и хроматиновые домены.

Важнейшим результатом было обнаружение хроматиновых **петель**. Данный элемент архитектуры хроматина образуется, когда два коротких локуса, длиной не более 10 тысяч п.о. специфично взаимодействуют друг с другом в пространстве ядра. На картах Hi-C это выражается в повышенной в несколько раз частоте

контактов между целевыми бинами (основаниями петель) по сравнению с ближайшим окружением.

Сравнение карт Hi-C между разными клеточными типами и видами показало, что порядка 55-75% петель являются консервативными между клеточными типами и порядка 50% петель консервативны между разными видами.

Изучение свойств генома в основаниях петель показало, что около 30% петель связывают известный промотор гена и известный энхансер, что в 4 раза превышает ожидаемую долю для случайных петель. Более того, гены, промоторы которых ассоциированы с петлями в среднем имеют в 6 раз более высокий уровень экспрессии. Что важно, различия в положении петель между разными клеточными типами также связаны с изменением активности генов.

Далее была обнаружена взаимосвязь между формированием ТАДов и петель. Основания порядка 40% петель являлись границами ТАДов, а границы порядка 40% ТАДов являлись также и основаниями петель. Примечательно, что в данной работе ТАДы выделялись не на основе DI, а на основе «Arrowhead»-подобного преобразования матрицы контактов.

Для изучения молекулярно-биологических механизмов формирования петель были использованы данные иммунопреципитации хроматина для целого ряда белков хроматина. Было показано, что приблизительно в 80-90% случаев, в основания петель обогащены пиками CTCF и пиками субъединиц когезина RAD21 и SMC. В ~50% случаев вместе с пиком CTCF обнаруживалось и нуклеотидная последовательность, соответствующая сайту связывания белка CTCF.

Важно отметить, что последовательность сайта связывания CTCF не является симметричной, таким образом, по отдельному сайту CTCF можно определить ориентацию в геноме: «прямая» ориентация, если обнаруживается непосредственно известный мотив CTCF и «обратная» ориентация, если обнаруживается последовательность, комплементарная мотиву CTCF. В соответствии с этим, каждая пара сайтов связывания CTCF может быть охарактеризована по их взаимной геномной ориентацией. Исследование показало, что в 90% случаев петель, в основании которых обнаруживаются

последовательности связывания CTCF, данные сайты ориентированы навстречу друг другу. Так как в геноме встречаются сайты CTCF во всех ориентациях, то вероятность формирования большого количества случайных петель с указанной выше особенностью является совершенно ничтожной, что говорит о биологической важности ориентации сайтов связывания CTCF при формировании петель хроматина.

Совокупность полученных результатов позволяет предположить, что, формирование ТАДов и хроматиновых петель взаимосвязано и участвуют в регуляции активности генов; ключевым элементов в формировании ТАДов и петель является димеризация белка CTCF, расположенного в основаниях петель (в границах домена) и его взаимодействие с белком когезином. Важность данных механизмов в работе генетического аппарата млекопитающих подчёркивается их эволюционной консервативностью с одной стороны и изменчивостью в разных типах клеток с другой стороны.

Следует указать, что часть петель, у которых не был обнаружен белок CTCF в основании, формируются белками комплекса Polycomb [53,54]. Их интересной особенностью является то, что в отличие от CTCF-опосредованных петель они способны формировать взаимодействовать на расстояниях, сопоставимыми с размерами хромосом.

В соответствии с этим дальнейшие исследования были направлены на изучение молекулярно-биологических механизмов формирования архитектуры хроматина и их роли в регуляции генной экспрессии.

2.4. Архитектура хроматина *D. melanogaster*

D. melanogaster, является одним из старейших модельных объектов в генетике. Именно на этом объекте впервые были прослежены связи между состоянием хроматина, особенностями его укладки и экспрессией генов [15,16,55]. Также *D. melanogaster* была одним из первых объектов, на котором были полученные данные Hi-C [9,10].

В первую очередь, полученные для *D. melanogaster* данные Hi-C продемонстрировали увеличение числа теломер-теломерных и центромер-центромерных контактов между хромосомами, что согласуется с известными особенностями организации хромосом в ядрах эмбриональных клеток *D. melanogaster*. Более подробное изучение данных Hi-C позволило обнаружить доменную организацию хроматина, подобной той, которая ранее была выделена у млекопитающих. Однако, первые же результаты показали, что обнаруженные исследователями «физические домены» не соответствуют в полной мере ТАДам у млекопитающих. Главным отличием было то, что у *D. melanogaster* домены строго соответствовали различным состояниям хроматина, тогда как у млекопитающих такой особенностью обладало менее 20% ТАДов и, в общем и целом, деление на ТАДы не зависело от эпигенетического состояния хроматина. Важно отметить, что наблюдаемая доменная организация хроматина соответствует бендам на цитогенетических картах хромосом [56].

Сравнение с данными по иммунопреципитации хроматина показало, что границы доменов у *D. melanogaster* обогащены пиками белков CP190, CTCF, Beaf-32 и Chromator. Участие в формировании границ доменов белка CTCF может указывать на то, что механизмы, формирующие у ТАДы у млекопитающих и физические домены у *Drosophila melanogaster*, являются общими. С другой стороны, лишь около 20% границ доменов у *D. melanogaster* обогащены пиками CTCF и эти же границы обогащены пиками других белков, например, CP190 [10]. Также существуют указания на то, что в гораздо большей степени выражено

обогащение границ доменов не отдельными белками, а их комбинациями, что может быть связано с их способностью взаимодействовать друг с другом [9].

Наконец, помимо доменов, у *D. melanogaster* были обнаружены хроматиновые петли, подобные тем, что были обнаружены у млекопитающих, однако и в этом случае наблюдается существенная разница в механизмах их формирования. Если у млекопитающих хроматиновые петли связаны с совместной работой белков CTCF и когезина, то у *D. melanogaster* они определяются белками Polycomb [10].

Дальнейшие исследования [57–60] продемонстрировали, что CTCF, по всей видимости, не участвует в формировании архитектуры хроматина у *D. melanogaster*. Более того, ряд признаков указывает, что белок CTCF у *D. melanogaster* никоим образом не связывается с когезином [58,60], что может говорить о разных механизмах формирования архитектуры хроматина у *D. melanogaster* и млекопитающих.

Впрочем, в некоторых случаях для CTCF было строго показано строгое участие в формирование архитектуры хроматина. Исследование фенотипа и архитектуры хроматина особей *D. melanogaster* с генотипом $CTCF^0$, показало, что экспрессия этого белка необходима для развития нервной системы, иначе превращение личинки во взрослую особь не происходит [61]. Сравнение карт Hi-C полученных из клеток центральной нервной системы личинок дикого типа и $CTCF^0$ показало, что у мутантных особей произошло ослабление или исчезновение около 16% границ доменов, из которых около 2/3 (или 10% от общего числа) составляли границы, обогащённые CTCF в диком типе. Одновременно с этим, отсутствие белка CTCF меняет профиль экспрессии ближайших к прежнему месту посадки генов.

Более детальное изучение механизмов влияния белка CTCF на архитектуру хроматина и регуляцию генной экспрессии позволило обнаружить, что CTCF отвечает за рекрутированние белка CP190 на хроматин и сам по себе белок CTCF не способен прямо подавлять или активировать генную экспрессию [62].

Таким образом, в отличие от млекопитающих, у которых белок CTCF является одним из главных архитектурных белков, отвечающих за формирование

пространственной организации хроматина, у *D. melanogaster* его роль как архитектурного белка ограничена участием в укладке ДНК лишь небольшого числа локусов.

Ввиду того, что достаточно быстро выяснилось ограниченное значение белка CTCF для формирования архитектуры хроматина, исследователи сконцентрировались на поиске других белков, которые могли бы у *Drosophila melanogaster* исполнять роль, аналогичную CTCF у млекопитающих.

Так в исследовании, основанном на данных Hi-C с разрешением глубже 1000 п.о., было показано, что доменная организация хроматина имеет более сложную структуру [59,60]. Во-первых, для тех регионов, которые на картах Hi-C с меньшим разрешением определялись как отдельные домены, была показана сложная иерархическая организация с множеством субдоменов. Во-вторых, те регионы, которые ранее не выделялись как домены и считались неструктурированным междоменным пространством, оказались наполнены множеством доменов небольшого размера. Учёт новых границ доменов за счёт карт Hi-C на более высоком разрешении показал сильную связь между наличием связывания белков Beaf-32, CP190 и Chromator с хроматином и границами доменов. Особенно сильна эта связь была не для отдельных белков, а для пар Beaf-32/CP190 и Beaf-32/Chromator. Так, 74% таких пар оказались колокализованы с границами доменов и 77% границ доменов колокализованы с сайтами связывания хотя бы одной такой пары [59]. Исследования процессов установления архитектуры хроматина в ходе ранних стадий эмбриогенеза показало, что появление границы хорошо предсказывается на основе данных по размещению белков Beaf-32, CP190 и Chromator [63].

Однако вклад белка Beaf-32 не совсем однозначен. В исследовании, в котором проводилась деплеция этого белка с использованием интерферирующих РНК, не обнаружено влияние подавления экспрессии этого белка на архитектуру хроматина [60]. С другой стороны, известно, что такой метод не всегда эффективен, ввиду того, что не всегда происходит полное подавление экспрессии соответствующего

гена, а также никак не влияет на белки, уже связанные с хроматином. Таким образом, данное наблюдение нельзя считать достаточным доказательством.

Тем не менее, в некоторых исследованиях отдают предпочтения компартментализации хроматина, обозначая активный хроматин и транскрипцию как ключевой фактор для определения границ доменов [57,64]. Использование модификации протокола Hi-C для получения данных по архитектуре хроматина с индивидуальных клеток показало, что в отдельных клетках домены не имеют сложной иерархической структуры [64]. Проведение анализа Hi-C на отдельных клетках позволило классифицировать границы доменов по их консервативности в зависимости от того, в какой доле клеток граница доменов обнаруживается. Исследование показало, что районы консервативных границ доменов существенно обогащены метками активного хроматина. В то же время обогащение инсуляторными белками, такими как Beaf-32, так же присутствует, но в гораздо меньшей степени.

Подобно млекопитающим, на картах Hi-C *D. melanogaster* также были обнаружены петли хроматина, однако в первых же исследованиях они были связаны с работой белков комплекса Polycomb [10,58]. Детальное исследование, основанное на построении карт Hi-C с разрешением менее 1000 п.о., позволило выявить, что менее 30% хроматиновых петель имеют в своём основании пики CTCF согласно данным иммунопреципитации хроматина [65]. Более того, чем сильнее выражена ассоциация петель с CTCF – тем меньше доля таких петель. Аналогичные результаты были получены для других инсуляторных белков: CP190, Beaf-32, Su(Hw). Тем не менее, и у млекопитающих и у *D. melanogaster* хроматиновые петли строго ассоциированы с когезином, точнее, с его субъединицей Rad21. Таким образом и у млекопитающих, и у двукрылых формирование петель связано с деятельностью белков когезинового комплекса, однако у млекопитающих вторым важным компонентом являются связанные с хроматином и правильно ориентированные белки CTCF, тогда как у *D. melanogaster* ни CTCF, ни другие инсуляторные белки в формирования петель, по всей видимости, не участвуют. Возникает вопрос: могут ли у *D melanogaster* в

формировании хроматиновых петель участвовать совместно с когезином участвовать какие-то другие белки? Исследование показало, что такими белками могут быть белки комплекса Polycomb. При этом, в отличие от млекопитающих, у которых формирование хроматиновых CTCF-опосредованных петель связано с активацией экспрессии, у *D. melanogaster* участие белков комплекса Polycomb в формировании петли приводит к репрессии генной активности.

Также необходимо отметить, что если у млекопитающих существует достаточно строгая связь между хроматиновыми петлями и ТАДами и основания петель часто являются также границами ТАДов (субТАДов), то у *D. melanogaster* хроматиновые петли не имеют такой жёсткой привязки к границам доменов и наблюдаются практически исключительно внутри репрессированного хроматина связанного с белками комплекса Polycomb [10,65].

Таким образом, говоря о пространственной организации хроматина млекопитающих и *D. melanogaster* следует указать, что несмотря на то, что на картах Hi-C можно увидеть или выделить алгоритмами одинаковые элементы, такие как пространственные домены и хроматиновые петли, даже первичные исследования указывают на то, что в их основе часто лежат разные молекулярно-биологические механизмы, играющие важную роль в регуляции генной экспрессии.

2.5. Молекулярно-биологические механизмы формирования архитектуры интерфазных хромосом.

Первые же результаты Hi-C экспериментов, подтвердившие укладку хромосомы в интерфазном ядре в виде фрактальной глобулы, позволили выдвинуть предположение, что данная укладка обеспечивается конденсиноподобными белковыми комплексами, которые компактизируют хромосомы за счёт формирования и протягивания петель ДНК. Математическое моделирование данного механизма – механизма протягивания петли – предсказывало появление самоорганизующихся петлевых доменов [66]. Появление более детальных данных Hi-C, показало, что таким комплексом, ответственным за упаковку интерфазных хромосом может являться когезин. Однако, этот белковый комплекс действует не самостоятельно, а вместе белком CTCF, который, благодаря способности связываться с уникальными нуклеотидными мотивами, может служить в качестве «разметки», ограничивающей продвижение молекул когезина (Рисунок 3) [67].

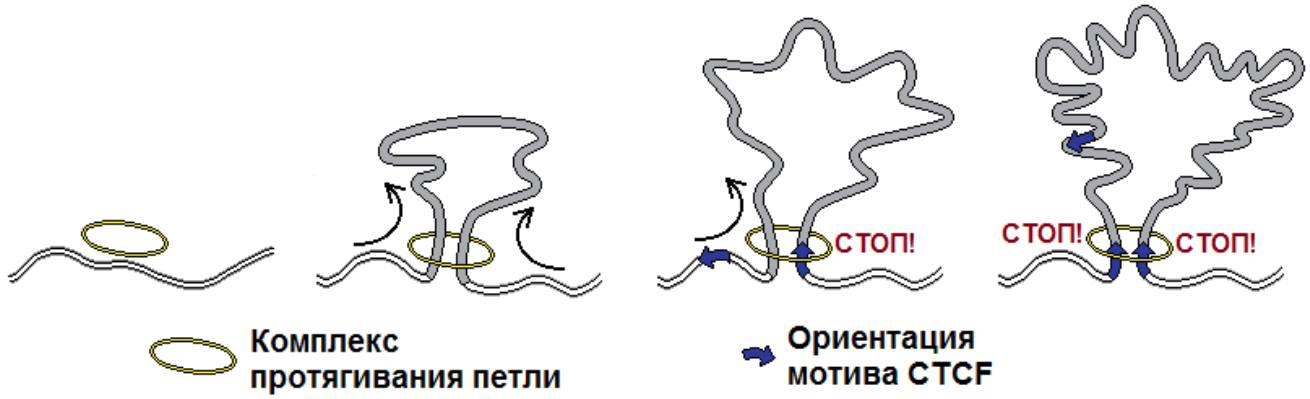

Рисунок 3. Принцип работы механизма протягивания петли [68].

В дальнейшем последовал целый ряд исследований посвящённых изучению деталей механизмов формирования архитектуры хроматина, роли когезина, CTCF, а также поиску дополнительных участников этих механизмов. Такого рода исследования двигались по двум магистральным путям: анализ карт Hi-C, построенных на клеточных линиях, лишённых тех или иных белков и белковых

комплексов и физическое моделирование архитектуры хроматина. Важно отметить, что карты Hi-C отображают собой вероятности контактов и являются очень удобным инструментом для верификации результатов такого моделирования.

Основными результатами таких исследований является то, что ТАДы являются динамическими структурами. Основную роль в формировании данных структур играют белки CTCF, когезин, WAPL и PDS5. Белок когезин, обладающий моторной функцией, после загрузки на нить ДНК протягивает её сквозь себя, тем самым формируя петли. Процесс протягивания петли ДНК идёт до тех пор, пока когезин не натолкнётся на белки CTCF, которые останавливают данный процесс [68–73].

Важные результаты о динамической природе ТАДов были получены в ходе Hi-C экспериментов на отдельных клетках и ядрах [74–76]. Показано, что в каждой клетке, в каждый момент времени одновременно происходит формирование, протягивание и разрушение десятков и сотен петель и, конечно, положение этих петель от клетки к клетке отличается. Тем не менее, при сравнении архитектуры хроматина множества клеток видно, что те или иные регионы гораздо чаще формируют контакты друг с другом, чем с другими. Объединение данных же приводит к формирование типичных для канонического Hi-C карт контактов.

И когезин, и CTCF являются одинаково важными компонентами для формирования ТАДов. Даже в случае частичной деплеции CTCF при использовании малых интерферирующих РНК [70] происходит ослабление границ ТАДов, вплоть до полного их исчезновения при полном удалении CTCF посредством ауксин-индуцибельной деградации [72,73]. Исчезновение ТАДов также наблюдается при полной деградации когезина [73,77], но не частичной [70]. Стоит отметить, что удаление данных архитектурных белков не оказывает никакого влияния на компартментализацию хроматина или даже приводит к её усилению [72,73,77].

Динамическая природа ТАДов и роль CTCF как элемента, создающего границы подчёркивают результаты физического моделирования [78–81]. Модели,

не учитывающие этого, имели ограниченную предсказательную силу [78–80]. А наилучшее соответствие реальным данным показала та модель, которая учитывает изменчивость положения ТАДов и петель в разных клетках, изменения ТАДов по времени и роль CTCF как глобального граничного элемента, блокирующего взаимодействия между ТАДами [81].

Отдельный интерес представляет динамика связывания архитектурных белков с ДНК, так среднее время удержания CTCF на хроматине составляет 1-2 минуты, а когезина – около 22 минут, против 2-15 секунд у большинства специфичных к нуклеотидной последовательности транскрипционных факторов [82]. Подобные исследования в отношении динамики протягивания петли показывают, что когезин в среднем проходит около 0.5 тысяч п.о. в секунду [83]. Сочетание этих данных показывает, что за среднее время жизни в клетке когезин будет формировать петли длиной около 600-700 тысяч п.о., что значительно превышает обычные размеры ТАДов, известные из литературных данных, и подчёркивает роль CTCF в демаркации границ ТАДов. Эти данные также согласуются с результатами исследования ауксин-зависимой деградации когезина, показывающими, что длинные петли, порядка 900 тысяч п.о. восстанавливаются в течение 40 минут после удаления ауксина [77].

Когезин

Известно, что процессу посадки и снятия когезина с хроматина ассистируют такие белки как Nipbl и WAPL, соответственно, логично предположить, что они также могут влиять на организацию хроматина. Физическое моделирование предсказывало, что деплеция белка WAPL должна увеличить дальность петель формируемых когезином [81]. Предсказание было блестяще подтверждено при изучении влияния на архитектуру хроматина деплеции WAPL. Удаление WAPL приводило к увеличению размеров доменов, усилению взаимодействий внутри доменов и контактов в локусах, формирующих петлю хроматина [73,84]. Аналогичные результаты получаются и при деплеции PDS5 [73]

Ещё одним белком, участвующим в функционировании когезина является Nipbl, отвечающий за загрузку когезина на ДНК. Показано, что в ряде случаев

сайты посадки Nipbl формируют кластер вблизи одной из границ доменов, насыщенной сайтами связывания CTCF [85]. В таком случае, когезин, с момента посадки на хроматин, с одной стороны оказывается «упирается» в границу домена и не способен протягивать нить ДНК в этом направлении, зато активно протягивает её в свободную сторону. В таком случае на картах Hi-C можно видеть насыщенную контактами полосу, идущую вдоль одной из границ доменов. Примечательно, что супер-энхансеры оказываются ассоциированы с таким полосами, то есть с регионами «одностороннего» протягивания нити ДНК.

CTCF

Изучение архитектуры хроматина на клеточных линиях с модифицированными белками CTCF показало, что N-концевой домен непосредственно отвечает за связывание CTCF с когезином [86,87]. Наиболее подробное исследование показывает, что мутантные варианты CTCF, с делецией региона N(1-265) , оказались не способны ограничивать продвижения когезина, что приводило к ослаблению и исчезновению границ ТАДов [86]. Примечательно, что в этом регионе CTCF находится домен N(13–33) содержащий мотив KTYQR, PDS5A-связывающий мотив, общий с WAPL и некоторыми другими белками. Вероятно, за счёт этой особенности, CTCF способен прямо связываться с PDS5 и посредством этого стабилизировать когезин на ДНК и/или «защищать» PDS5 от связывания с WAPL и, тем самым, не допускать снятия когезина.

Интересной особенностью CTCF является его способность связываться с различными РНК. Было показано *in vitro*, что наличие РНК облегчает формирование димеров CTCF [88], на основе чего можно предположить, что и *in vivo* молекулы РНК способны стабилизировать формирование хроматиновых петель. Это предположение подтверждается в Hi-C экспериментах, на линиях клеток с мутантными CTCF [86,89]. Отсутствие возможности CTCF связываться с молекулами РНК приводило к ослаблению границ ТАДов и хроматиновых петель, вплоть до полного их исчезновения в отдельных случаях. Более детальное исследование этого вопроса показало, что молекулы РНК необходимы для формирования *некоторых* границ ТАДов и некоторых петель [90]. Что позволяет

выдвинуть предположение о существовании разных классов ТАДов и хроматиновых петель, в зависимости от необходимости РНК для их стабилизации.

Протягивание петли без участия CTCF

Важно отметить, что, помимо белков CTCF и когезина, в формировании архитектуры хроматина на более локальном уровне участвуют и другие белки. Во-первых, стоит отметить белок YY1, который формирует в разы более короткие петли, чем CTCF и, вероятно, отвечает за более точное сближение промоторов и энхансеров генов [91,92]. Кроме этого, ряд промотор-энхансерных петель связанны с активацией транскрипцией и посадкой РНК-полимеразы II на промотор гена [92,93] и РНК полимеразы III [94].

Во всех, указанных ваше случаях, по всей видимости, необходимо соучастие когезина в сближении целевых локусов для формирования взаимодействия. Однако, помимо когезин-зависимых механизмов формирования архитектуры хроматина есть и когезин-независимые.

Когезин-независимые механизмы формирования архитектуры хроматина

Существование хромосомных территорий в интерфазном ядре ставило вопрос о том, какими механизмами обеспечивается сегрегация и колокализация хромосом. Одни из первых моделей предполагали, что это может быть связано с разной энергией связи тех или иных участков хроматина друг с другом и присутствующими в растворе молекулами [95–97]. Физическое моделирование показало, что различий в энергии связей достаточно, чтобы описать все известные на тот момент особенности архитектуры хромосом, такие как петли, колокализацию удалённых последовательностей и хромосомные домены [96,97].

Получение более подробных сведений об эпигенетическом состоянии разных участков хромосом, позволило предположить, что одним из ключевых механизмов формирования архитектуры хроматина является так называется «фазовая сепарация». Разные фрагменты хроматина, в зависимости от эпигенетического состояния, имеют разную силу связи с друг с другом, что и приводит к разделению хроматина в пространстве интерфазного ядра [98–100]. Модели формирования

архитектуры хроматина, основанные на фазовой сепарации подтверждаются в том числе и прямыми измерениями динамики хроматина в клеточном ядре [101–103].

Необходимо подчеркнуть, что белки комплекса Polycomb также способны формировать пространственные кластеры по механизму фазовой сепарации [102,104,105]. Благодаря этому механизму формируются хроматиновые петли, соединяющие регионы удалённые на десятки и сотни миллионов п.о., однако сомнительно, что все эти взаимодействия являются функциональными [54].

Стоит отметить, что механизм протягивания петли действует антагонистично механизму фазовой сепарации, что было показано на экспериментах с удалением когезина [53,72,73,77]. В отсутствии этого белка усиливается как компартментализация, так и способность поликомба формировать пространственные кластеры.

Таким образом, архитектура хроматина интерфазных клеток формируется совместным действием когезин-опосредованных механизмов протягивания петли, фазовой сепарацией эу- и гетерохроматина и фазовой сепарацией комплекса Polycomb. При этом в разных организмах доминируют разные механизмы, так у млекопитающих архитектура хроматина определяется в основном деятельностью когезина в союзе с белком CTCF, тогда как у двукрылых доминирует фазовая сепарация. Ограниченное количество данных же на других видах не позволяет говорить, насколько указанные выше особенности являются общими для разных позвоночных и насекомых или присущи только исследованным таксономическим группам.

2.6. Архитектура хроматина в современных исследованиях функционирования генома

Ещё до появления экспериментальных методик семейства 3C было выдвинуто предположение, что архитектура хроматина в своей динамике принимает участие в регуляции активности генов [21,22,34]. Отмечается, что для обеспечения воздействия регуляторных последовательностей на целевой ген необходимо их пространственное сближение [34]. Собственно, развитие методов семейства 3C и было простимулировано необходимостью детекции регуляторных контактов, в первую очередь - промотор-энхансерных [1].

Промотор-энхансерные взаимодействия

После разработки метода Hi-C значительные усилия исследователей были сфокусированы на улучшении аннотации регуляторных элементов с использованием данных о пространственных контактах между целевыми локусами. Использование Hi-C и родственных ему экспериментальных методов основано на наблюдении, что энхансер и промотор должны сблизится в пространстве клеточного ядра и взаимодействовать, что на картах Hi-C выражается в повышенной частоте контактов [106–109]. Использование этой особенности позволило в масштабах всего генома детектировать регуляторные элементы и их гены-мишени [110–114], в том числе и энхансеры генов длинных некодирующих РНК [115]. Дальнейшие исследования обнаружили не только пространственные контакты промоторов с энхансерами, но и энхансеров, совместно регулирующих один или несколько генов, друг с другом [116–122]. И если в отношении промотор-энхансерных контактов не возникает сомнений в их функциональности, то такие сомнения есть для энхансер-энхансерных контактов. Остаётся неясным, являются ли эти контакты, обнаруженные на карте Hi-C, следствием взаимодействия транскрипционных факторов друг с другом или регуляторными последовательностями, или следствием того, что все эти энхансеры, пространственно близкие к одним и тем же промоторам, будут близки и друг к другу.

Частным случем таких кластеров являются кластеры суперэнхансеров. Следует отметить, что эпигенетические характеристики энхансеров и суперэнхансеров практически не отличаются, и одним из надёжных способов их детекции является анализ пространственной организации хроматина [123,124]. Показано, что суперэнхасеры отличаются от обычных энхансеров большим числом контактов со своим ближайшим окружением и тем, что часто находятся вблизи границ доменов с сильной инсуляцией [124]. И если наличие большого количества контактов у суперэнхансера является прямым следствием большого количества регулируемых им генов, то ассоциация суперэнхансеров с сильными границами ТАДов не обязательно связана с какой-то «особой» силой данных регуляторных последовательностей. Возможно, ключевым фактором здесь является обеспечение взаимодействия суперэнхансера с целевыми промоторами за счёт механизма «одностороннего» протягивания петли и обеспечивается совместной посадкой белков Nibpl и CTCF на одной из границ ТАДа [85].

Полногеномный поиск ассоциаций

Известно, что большая часть геномных вариантов (~90%), ассоциированных с теми или иными болезнями, находится в некодирующей области генома и на значительном удалении от генов, что затрудняет интерпретацию результатов и поиск целевого гена [125]. Предполагается, что это связано с изменениями в энхансерах генов, регулирующих их активность, однако проверить такие предположения и соотнести предполагаемый энхансер и целевой ген долгое время было затруднительно. Появление экспериментальных методик семейства 3C позволило не только интерпретировать известные результаты, но и детектировать новые локусы риска, мутации в которых могут быть связаны с развитием тех или иных заболеваний. Главный подход к решению такой задачи заключается в выявлении регионов, имеющих повышенную частоту контактов с локусами, для которых установлена их вовлечённость в развитие того или иного признака/заболевания [116,126]. Результатом развития этого подхода стало появление метода промотор-обогащённого Hi-C [111,127]. Данный метод позволяет удалить из библиотеки Hi-C химерные последовательности, которые не

содержат известных промоторов. Ценой потери данных об архитектуре хроматина большей части генома, можно гораздо глубже изучить особенности пространственной организации локусов, активно контактирующих с промоторами генов. Благодаря этому были выявлены десятки новых локусов риска для развития разных форм злокачественных новообразований [127–132], аутоиммунных заболеваний [133,134], нервно-физиологических расстройствами [135,136] и заболеваний сердечно-сосудистой системы [137–139].

Архитектура хроматина в контексте клеточной судьбы

Уже первые исследования показали, что различия между в архитектуре хроматина между разными типами клеток коррелируют с различиями в экспрессии между ними: архитектура хроматина меняется там, где меняется экспрессия [5]. Последующие исследования, сконцентрированные на вопросах связи клеточной судьбы и дифференциации с изменениями архитектуры хроматина, подтвердили обнаруженную связь, однако оставалось не ясным, является ли изменение архитектуры хроматина причиной изменения транскрипции или следствием [52,140–143]. Аналогичные результаты были получены также при сравнении архитектуры хроматина нормальных и малигнизированных клеток, при этом различия в архитектуре хроматина наиболее велики в локусах, несущих онкогены [144–147]. Выдвигается предположение, что перестройку архитектуры хроматина возможно рассматривать как один из диагностических признаков малигнизации [148].

Некоторую ясность во взаимосвязь пространственной организации хроматина и регуляции транскрипции вносят исследования, показывающие, что в ходе клеточной дифференциации, пространственные контакты между энхансерами и промоторами устанавливаются до того, как соответствующие гены станут транскрипционно активны [149,150]. Так, ряд хроматиновых петель, которые формируются в ходе сперматогенеза, связаны с генами, активируемыми на ранних стадиях эмбриогенеза [151]. Таким образом, как минимум в некоторых случаях изменения в транскрипции происходят после изменений в архитектуре хроматина

целевых локусов, что согласуется с данными, которые имеются для ЛАДов [29]. Впрочем, необходимо помнить, что после – не значит вследствие.

Влияние глобальной архитектуры хроматина на регуляцию генов

Одни из первых данных о непосредственной связи архитектуры хроматина с регуляцией генной экспрессии были получены в ходе экспериментов по ауксин-зависимосой деградации CTCF и белков когезинового комплекса.

Результаты экспериментов по деградации CTCF показали на клетках млекопитающих, что данный белок имеет ограниченный эффект на экспрессию генов [70,72]. В основном, отсутствие CTCF сказывается на генах, лежащих в непосредственной близости, до 1000 п.о. от его сайтов связывания. Данная особенность не позволяет в полной мере понять, являются ли эти изменения следствием изменения архитектуры хроматина, следствием способности CTCF выступать в качестве одного из факторов транскрипции или просто особенностью клеточной модели. Так, в экспериментах с нокаутом гена CTCF на *Danio rerio* наблюдалась масштабное нарушение и архитектуры хроматина и экспрессии генов развития, что приводило к значительным фенотипическим изменениям [152].

В отличие от CTCF, деградация когезина показала не только глобальное изменение архитектуры хроматина, но и глобальное нарушение экспрессии генов [67,70,77]. Исследование этого процесса выявило, что когезин принимает крайне важное участие в процессах регуляции генов развития и клеточной пролиферации [53,153–155]. Было показано, что два варианта комплекса когезина, определяемых составом входящих в комплекс белков – когезин-SA1 и когезин-SA2 – по-разному участвуют в формировании архитектуры хроматина. Так, когезин-SA1 имеет каноническую функцию – совместно с белком CTCF формирует ТАДы и отвечает за сближение промоторов и энхансеров генов [153,154]. Кроме этого, когезин-SA1 разрушает петли сформированные белками комплекса Polycomb, тем самым ограничивая их возможность репрессировать транскрипцию [53,154]. Также, по всей видимости, когезин-SA1, вместе с CTCF, ограничивает нецелевые контакты энхансеров. Когезин-SA2 же имеет противоположные свойства: этот вариант белкового комплекса не склонен связываться с CTCF, способствует установлению

связи суперэнхансеров и комплекса Polycomb с целевыми генами [154]. В этом контексте представляет интерес то, что деградация WAPL так же приводит к нарушению экспрессии генов: являясь белком, снимающим когезин с ДНК, он обеспечивает стабильную и регулярную работу механизма протягивая петли, и, тем самым, отвечает за поддержание контактов между регуляторными последовательностями и промоторами генов

Связь локальной архитектуры хроматина и регуляции генной экспрессии

Наиболее точную информацию о том, в какой мере архитектура хроматина вовлечена в регуляцию генной экспрессии можно получить, если изучить, как меняется экспрессия генов в целевом локусе при управляемом изменении архитектуры хроматина в нём. Так как у млекопитающих одним из главных компонентов формирования пространственной организации хроматина является белок CTCF, который специфично связывается с известными нуклеотидными последовательностями, модификация таких последовательностей (удаление, инверсия или вставка) позволяет крайне точно и локально менять архитектуру хроматина.

Результаты таких экспериментов наглядно демонстрируют, что значение архитектуры хроматина для направления взаимодействия энхансеров и промоторов. Было показано, что изменение сайтов CTCF приводит к нарушению формирования границ ТАДов, нецелевому взаимодействию энхансеров и промоторов и как, следствие, изменению уровня экспрессии в клеточных линиях [71,156–159]. Отдельные эксперименты показывают, что сайты связывания CTCF могут выступать не качественным, а количественным регулятором активности генов, за счёт регуляции частоты контактов промотора с дистантным энхансером [160]. Однако необходимо учитывать, что регуляция генов эукариотических организмов находится под сложным контролем с множеством отрицательных обратных связей и даже значительное изменение экспрессии одного или нескольких генов часто не имеет никакого фенотипического проявления. В этом отношении, очень важными оказались работы, показывающие видимые фенотипические изменения, вызванные нарушением пространственной

организацией хроматина. Так на модели с дифференциацией гемопоэтических клеток было показано, что удаление клеточноспецифичных сайтов связывания CTCF нарушает процесс дифференцировки клеток в соответствующий тип [161]. А эксперименты, проведённые на *M. musculus* [162,163], демонстрируют видимые фенотипические нарушения на организменном уровне.

Похожие результаты были получены и при удалении локусов, обогащённых контактами с промоторами и энхансерами, что приводило к дезорганизации хроматина и нарушению паттерна экспрессии вплоть до клеточной гибели [164,165].

Неоднозначность роли архитектуры хроматина

Тем не менее, существует ряд работ, показывающих, что роль архитектуры хроматина в регуляции генной экспрессии может быть преувеличена. Так при удалении в локусе *Firre* эволюционно-консервативной границы ТАДа, насыщенной ~15 сайтами связывания CTCF, никаких изменений ни в экспрессии данного гена, ни в архитектуре хроматина замечено не было [166]. Впрочем, данное исследование производилась на ЭСК и нельзя утверждать, что для поддержания границы ТАДа именно на этой клеточной стадии сайты связывания CTCF были необходимы.

Несколько иные результаты были получены при удалении группы CTCF сайтов в кластере HoxD генов. Результатом делеции стало появление дальних межТАДовых контактов между энхансерами и промоторами генов и нарушение экспрессии генов, но при этом сама граница между ТАДами, не изменилась [167]. Такая противоречивая картина не позволяет в полной мере определить, выступает ли в данном случае CTCF как архитектурный белок или нет. Вполне допустимо представить, что CTCF может в данном случае, играть роль транскрипционного фактора и, например, связываясь с целевыми промоторами «закрывать» их от контактов со сторонними энхансерами.

Ещё одним контраргументом к значительной роли архитектуры хроматина в регуляции генной экспрессии является несоответствие кластеров коэкспрессирующихся генов и ТАДов [168]. У эукариотических организмов

коэкспрессирующиеся гены часто расположены рядом, формируя кластеры. Если ТАДы отвечают за целевое взаимодействие промоторов и энхансеров резонно ожидать, что коэкспрессирующиеся гены часто будут оказываться в одном ТАДе, а гены, лежащие в одном ТАДе – будут коэкспрессирующимися. Однако результаты сравнения кластеров коэкспрессии с ТАДами показывает, что совпадения между ними случайны.

Однако, одним из наиболее интересных вопросов о роли CTCF и архитектуры хроматина в регуляции генной экспрессии ставит исследование, в котором было проведено удаление участков хроматина, обогащённых пространственными контактами с известными энхансерами и промоторами, но где не было обнаружено ни значимого обогащения эпигенетическими метками, ни наличие каких-либо кодирующих последовательностей [164]. Результатом такой делеции стала перестройка архитектуры хроматина и радикальное изменение экспрессии, вплоть до клеточной гибели. С одной стороны, это доказывает роль архитектуры хроматина. Но с другой стороны, в удалённом локусе нет ничего - ни эпигенетических меток, ни сайтов связывания CTCF – что могло бы объяснить механизм произошедших изменений. Возникает вопрос: действительно ли в тех работах, где удаление, инверсия или вставка CTCF приводила к изменению архитектуры хроматина и экспрессии, причиной является именно удаление CTCF, а не оказавшегося в том же локусе неизвестного транскрипционного фактора?

2.7. Архитектура хроматина в эволюционном контексте

Одним из способов изучения механизмов формирования пространственной организации хроматина является её эволюционное сравнение между разными таксономическими группами. Обнаруженные сходства и различия могут указывать на тонкости в механизмах формирования архитектуры хроматина.

Хорошим примером, демонстрирующим роль эволюционного сравнения, является исследование, в котором изучили глобальную организацию хроматина и хромосом в разных таксономических группах [169]. Было показано, что укладка интерфазных хромосом в пространстве ядра определяется белковым составом кондесина II. Так в группах, обладающих полным набором белков (SMC2, SMC4, CAP-H2, CAP-G2, CAP-D3), хромосомы уложены в хромосомные территории, как у позвоночных. В группах, лишённых CAP-G2, например, у насекомых, наблюдается ориентация хромосом по Раблю. Однако, гораздо больший интерес представляют закономерности не глобальной укладки хроматина, а локальной, и её взаимосвязь с регуляцией генной экспрессии.

Результаты одного из первых Hi-C экспериментов, проведённых на разных клеточных линиях *H. sapiens* и *M. musculus* [5] показали, что архитектура хроматина является эволюционно консервативной. Очевидно, что эволюционная консервативность архитектуры хроматина должна опираться на консервативность нуклеотидных последовательностей. В первую очередь таковыми являются гены, их промоторы и энхансеры, и связанная с этим консервативность процессов регуляции генной экспрессии и развития. Однако, в случае с ТАДами ещё одним фактором их эволюционной консервативности становится способность CTCF узнавать специфические нуклеотидные последовательности и, следовательно, сохранение положения сайтов связывания CTCF относительно генов в ходе эволюции.

Одно из первых исследований, показывающих эволюционную консервативность сайтов связывания CTCF и их роль в регуляции генной экспрессии, было проведено ещё до широкого распространения метода Hi-C [170].

Результаты исследования позволили обнаружить, что сайты связывания CTCF, эволюционно консервативные для *H. sapience, M. musculus* и *Gallus gallus* являются локусами риска, ассоциированные с развитием тех или иных заболеваний.

Последующие исследования, с применением данных Hi-C, показывают, что именно консервативность положения CTCF лежит в основе консервативности ТАДов у млекопитающих [45,171,172]. Примечательно, что обнаруженное в самых первых исследованиях обогащение границ доменов SINE элементами [5] непосредственно связано с эволюцией сайтов CTCF. Так, поддержание консервативности ТАДов в ходе эволюции связано с формированием кластеров сайтов связывания CTCF, однако эти кластеры не являются стабильными: «старые» сайты CTCF могут теряться и кластер обновляется за счёт переноса транспозонами новых сайтов CTCF [172]. При этом транспозоны не только обеспечивают и поддержание старых границ доменов, но и создание новых границ ТАДов и изменение регуляции генов в ходе эволюции [173,174]. В этом отношении особый интерес представляет исследования, показывающие, что в разных линиях эволюции млекопитающих независимо происходила перестройка ландшафта связывания CTCF, вызванная экспансией транспозонов [175].

По всей видимости, эволюционная консервативность *положения* сайтов связывания CTCF в геноме и является одним из ключевых факторов, обеспечивающим консервативность архитектуры хроматина в целом в разных таксономических группах млекопитающих и обеспечивающим сохранение ТАДов как единиц регуляции экспрессии [176–178]. Впрочем, этот же механизм может принимать активное участие в дивергенции регуляции экспрессии между разными эволюционными линиями [173,179,180].

Насколько же обнаруженные закономерности верны для других позвоночных? На данный момент данные по архитектуре хроматина у других позвоночных, кроме млекопитающих, неполны и обрывочны. В большинстве случаев в исследованиях проводилось изучение ранних стадий эмбриогенеза и результаты по ним противоречивы. Так, некоторые исследования проведённые на *D. rerio* указывают на важность CTCF для формирования архитектуры хроматина и регуляции генов

[152,172]. Другие же, проведённые на *D. rerio* [181] и *Oryzias latipes* [182] и *Xenopus tropicalis* [183] указывают, что CTCF если и включается в формирование архитектуры хроматина, то на достаточно поздних стадиях эмбриогенеза, что противоречит известным данным для млекопитающих, у которых механизм протягивания петли начинает работать с момента формирования зиготы [75,76]. Следует обратить внимание на большое разнообразие представленности CTCF в разных таксономических линиях. Ортологи CTCF отсутствуют у некоторых ракообразных, насекомых и нематод [184,185]. Так, CTCF известен у *D. melanogaster,* но отсутствует у *Apis mellifera*; есть у *Trichinella spiralis*, но не обнаружена у *Caenorhabditis elegans*. Это может означать, что архитектурная роль CTCF может быть функцией, новоприобретённой у позвоночных (или даже млекопитающих), или функцией, многократно и независимо теряемой в разных ветвях эволюции.

В этом контексте представляет интерес обширное исследование архитектуры хроматина птиц. Эволюционные ветви птиц и млекопитающих разошлись около 310 миллионов лет назад, что близко ко времени разделения эволюционных ветвей *D. melanogaster* и *A. mellifera*, и к настоящему времени, геном птиц обрёл ряд особенностей, выделяющих этот класс позвоночных среди других.

Во-первых, геном птиц в среднем в 1,5-2 раза меньше по сравнению с другими группами позвоночных [186]. Так, было показано, что доля мобильных элементов в геноме разных групп птиц находиться в диапазоне от 4% до 10%, тогда как у млекопитающих, например, от 34% до 52%. Средняя длина генов у птиц в два раза меньше в сравнении с млекопитающими, и на четверть меньше в сравнении с рептилиями, что достигается за счёт уменьшения длины интронов [187]. Птицы также характеризуются малым числом межхромосомных перестроек и относительно стабильным кариотипом, так число хромосом для разных групп птиц составляет около 40 пар [188,189]. С учётом роли транспозонов в эволюции и формировании архитектуры хроматина у млекопитающих, пониженное число мобильных элементов у птиц может иметь важное значения для эволюции архитектуры хроматина и поддержания стабильности ТАДов (если они есть).

Одним из наиболее глубоко изученных объектов среди птиц является *G. gallus*. По результатам полногеномных исследований, геном *G. gallus* имеет размер около 1.25 млн. п.о. и сродержит, согласно биоинформационным предсказаниям, около 20-23 тысяч белок-кодирующих генов, что соответствует известным оценкам для млекопитающих [190]. Кариотип состоит из 38 пар аутосом и одной пары половых хромосом. Как и прочие птицы, геном *G. gallus* характеризуется малой долей сателлитной ДНК (менее 11%) и мобильных элементов (менее 9%). В эволюционном аспекте особый интерес представляет то, что в геноме *G. gallus* за прошедшее время произошло наименьшее количество перестроек, по сравнению с предполагаемыми предками, птичьими динозаврами [188].

Таким образом, изучение особенностей организации хроматина *G. gallus* позволяет получить уникальные данные о роли архитектуры хроматина в реализации генетической информации и произвести оценку её значения с эволюционной точки зрения.

Если среди позвоночных подавляющий объём данный по архитектуре получен на основе *H. sapiens* и M. *musculus*, то среди беспозвоночных главным источником сведений стала *D. melanogaster*. Эволюционное сравнение архитектуры хроматина *D. melanogaster* с другими представителями рода *Drosophila* даёт противоречивые сведения об эволюционной консервативности доменов. Так, ряд исследований указывает на то, что границы доменов насыщенны регионами эволюционных перестроек, а сами домены и границы остаются эволюционно консервативными [191,192]. Одновременно с этим есть исследования, показывающие, что даже если границы доменов остаются неизменными в эволюции, сами домены оказываются перестроенными и неконсервативными [193]. Последняя работа интересна тем, что в отличие от других, где домены рассматривались в целом, в этом исследовании проведено разделение доменов по группам, в зависимости от их эпигенетического состояния, и показано, что основная масса перестроек происходит в доменах, соответствующих жёлтому (транскрипционно активному) хроматину. Чем же обусловлены данные расхождения? По всей видимости, разным подходом к выделению и классификации доменов. Во всех исследованиях результаты

основывались на доменах, выделенных алгоритмом HiCExplorer. При этом на *D. melanogaster* в исследовании, показывающем консервативность доменов, было выделено более 1000 доменов [192], а в работе, показывающей неконсервативность доменов – всего 700 [193]. При этом в «консервативном» исследовании более 300 доменов было сформировано вокруг одного длинного гена [192], а в «неконсервативном» точки разрывов синтении соответствовали районам с повышенной инсуляцией, то есть слабым границам доменов [193]. Особое внимание привлекает на себя то, что в исследовании, демонстрирующем консервативность доменов, помимо HiCExplorer использовались алгоритмы Armatus и Arrohwead, и только около 25-35% доменов были общими для этих алгоритмов с выделенными HiCExplorer, и около 15% - общими для всех трёх алгоритмов [192]. Можно сделать вывод, что ответ на вопрос о том, консервативны или нет границы доменов у *Drosophila*, во многом определяется тем, каким алгоритмом и с какими параметрами домены были выделены. Соответственно и ответы на другие вопросы, как например, связь между изменением архитектуры хроматина и генной экспрессией, также будут зависеть от используемых алгоритмов.

В вопросе о взаимосвязи архитектуры хроматина и генной экспрессии в контексте эволюции, привлекает внимание работа, в которой изучалось влияние хромосомных перестроек у *D. melanogaster* на архитектуру хроматина и экспрессию генов [194]. В данном исследовании использовали линии *D. melanogaster*, созданные в середине 20го века и несущие массу хромосомных перестроек. Изучение взаимозависимости архитектуры хроматина и экспрессии проводилось на гетерозиготных особях, несущих хромосомы как дикого типа, так и перестроенные, показав, что чёткой связи между изменением архитектуры хроматина и экспрессии генов нет.

Кроме представителей рода *Drosophila*, данные по архитектуре хроматина имеются только для *Aedes aegypti* [195]. При этом глубокого исследования особенностей архитектуры хроматина не проводилось. Ограниченные сведения по устройству хроматина и эволюции хромосом есть для комаров рода *Anopheles*, что

непосредственно связано с высоким значением комаров данного рода как переносчиков малярии [11,196]. Обращает на себя внимание, что время дивергенции между разными видами составляет от нескольких миллионов лет до почти сотни миллионов лет (Рисунок 4), что сопоставимо со временем дивергенции отрядов *Primates* и *Rodentia* у млекопитающих [11,196]. При этом, в отличие от млекопитающих, несмотря на большое количество перестроек кариотип у комаров рода *Anopheles* крайне стабилен и практически полностью отсутствуют перестройки между разными хромосомами и плечами хромосом [11].

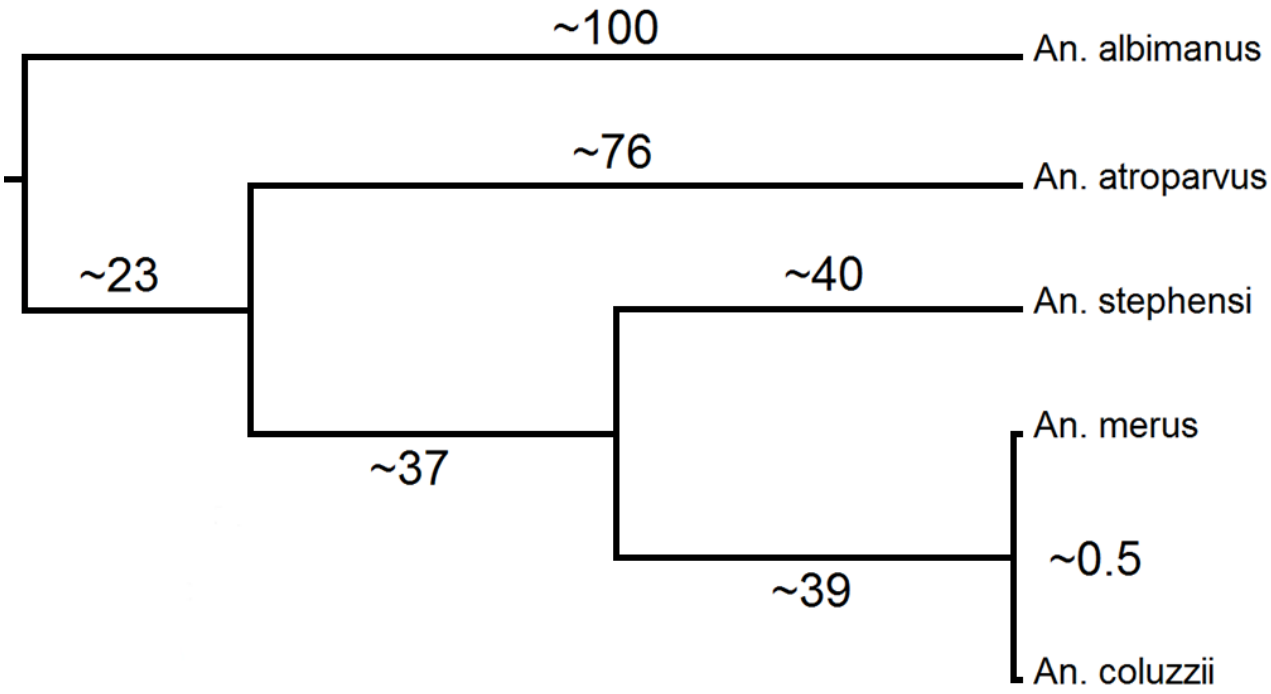

Рисунок 4. Филогенетические отношения комаров рода Anopheles, исследуемых в работе, согласно [11].

В свете этого, представляется достаточно важным пополнить представления о принципах организации архитектуры хроматина данными по комарам рода *Anopheles*, подкреплёнными данными о филогении исследуемых видов.

Обобщая результаты исследования исследований архитектуры хроматина в разных таксономических группах можно сделать два важных наблюдения. Во-первых, результаты исследования и полученные выводы могут сильно зависеть от используемых алгоритмов для аннотации архитектуры хроматина их

параметров. Так, в случае с изучением роли CTCF у *D. melanogaster*, этот белок оценивался как важный, если выделенные алгоритмами домены были большого размера, и как несущественный, если домены были маленькие. Так же и оценка степени консервативности архитектуры хроматина в эволюционном аспекте находится под сильнейшим влиянием используемых исследователями алгоритмов и их параметров. Во-вторых, основные усилия исследователей в области эволюционного сравнения архитектуры хроматина концентрируются вокруг сравнения геномного положения крупных архитектурных блоков, таких как ТАДы или компартменты. Однако, механизмы формирования архитектуры сложны, нередко антагонистичны. И, как показано в ряде экспериментов с деградацией тех или иных белков, даже если состояние таких архитектурных единиц как ТАДы и компартменты никак не поменялось, локальные контакты энхансеров и промоторов могут быть радикально нарушены, что приведёт к радикальному изменению паттерна экспрессии. Естественно, вся эта информация о локальных изменениях архитектуры хроматина и её связи с генной экспрессии будет полностью потеряна, если акцентировать своё внимание исключительно на крупных архитектурных единицах.

Таким образом, выглядит целесообразным разработка методов эволюционного сравнения архитектуры хроматина, которые будут, во-первых, свободны от артефактов, вносимых субъективным выборов алгоритмов и их параметров, применяемых для аннотации пространственной организации хроматина, а во-вторых, позволит сравнивать архитектуру хроматина на уровне каждого конкретного контакта.

3. Данные и методы

3.1. Анализ архитектуры хроматина позвоночных

В рамках данной работы были проанализированы как находящиеся в свободном доступе результаты Hi-C экспериментов, так и уникальные данные полученные в Отделе молекулярных механизмов онтогенеза ИЦиГ СО РАН по гранту РНФ (№14-14-00131). Изученные виды позвоночных и клеточные типы указаны в Таблице 1.

Таблица 1. Использованные в исследовании виды позвоночных, сборки геномов и типы клеток. Звёздочкой отмечены данные, полученные сотрудниками Отдела молекулярных механизмов онтогенеза ИЦиГ СО РАН

Вид	Используемая версия генома	Тип клеток, использованный для приготовления Hi-C библиотек
Homo sapiens	hg38	ЭСК [5], фибробласты IMR90 [1,6], гепатоциты [197]
Macaca mulatta	rheMac8	гепатоциты [45]
Mus musculus	mm10	ЭСК [5] , кортикальные нейроны [6], сперматозоиды*[44] , фибробласты [6], гепатоциты [45]
Canis familliaris	canFam3	гепатоциты [45]
Gallus gallus	galGal5	эмбриональные фибробласты*, полихроматические эритроциты*, зрелые эритроциты*

Для конвертирования геномных координат использовался ресурс NCBI Genome Remapping Service (https://www.ncbi.nlm.nih.gov/genome/tools/remap). Таким образом в анализ были включены пять видов позвоночных. Данные экспериментов Hi-C обрабатывались с помощью Juicertools [37] и

визуализировались с помощью стандартного программного обеспечения Juicebox [38].

Идентификация доменов

Для каждой исследуемой линии клеток было сгенерировано три разных набора доменов с использованием следующих алгоритмов: DomainCaller [5], Armatus [198], TADtree [199], TopDom [200], HicSeg [201]. Это связано с необходимостью выделить разные уровни пространственной организации хроматина, а также ограничить влияние алгоритмических особенностей выделения доменов на конечные результаты. Алгоритм DomainCaller является одним из первых опубликованных алгоритмов, предназначенных для поиска доменов. Этот алгоритм основан на расчёте силы инсуляции между соседними локусами, при этом регионы, обладающие высоким значением инсуляции, определяются как границы доменов. Алгоритм Armatus основан на поиске протяжённых регионов, преимущественно контактирующих внутри себя, такие регионы собственно называются ТАДами. Оба эти подхода взаимосвязаны, и отражают базовые свойства ТАДов, которые также используют прочие алгоритмы. TADtree также, как и Armatus, основан на поиске регионов обогащённых контактами, однако этот алгоритм позволяет автоматически выстраивать иерархию доменов и определять субдомены. Алгоритмы TopDom и HicSeg по некоторым оценкам являются одними из лучших алгоритмов, так как позволяют получать домены, которые чаще всего подтверждаются другими алгоритмами[202]. TopDom, подобно алгоритму DomainCaller, ищет границы доменов оценивая инсуляцию контактов, а HicSeg – сходен с Armatus, но основан на подходах сегментации изображений.

Для каждого бина рассчитывался индекс инсуляции на основе силы частоты контактов окружающих бин локусов [203]. В отличие от описанного в литературе метода, использовались не наблюдаемые величины контактов, а двоичный логарифм их обогащённости, по сравнению с ожидаемой величиной. В дальнейшем, также проводилась нормализация на среднюю величину инсуляции и среднее квадратичное отклонение для каждой хромосомы отдельно. Таким

образом, значение инсуляции, равное 0, соответствовало среднему для данной хромосомы, а величины меньше 0 – инсулирующим бинам.

Выделение A/B-компартментов

Для определения принадлежности локусов к A- или B-компартменту использовалось стандартное программное обеспечение, включённое в пакет Juicertools [204], основанное на определении собственного вектора нормализованной матрицы контактов на заданном разрешении. Для используемых данных выделение A/B-компартментов проводилось на разрешении 25, 40, 50 и 100 тысяч п. о. В основе данного алгоритма лежит определение первого собственного значения матрицы Hi-C, знак которого для каждой хромосомы определяется независимо. В соответствии с этим, знак полученных значений переопределялся с помощью корреляции с GC-составом таким образом, чтобы A-компартмент соответствовал локусам с высоким GC-составом [43].

Определение геномных координат и экспрессии генов

Все координаты генов приводились в соответствии с версиями геномов, указанных в Таблице 1. Источником координаты генов для позвоночных и списков генов-ортологов являлась база данных Ensemble.

Для анализа уровня генной экспрессии использовались данные RNA-seq (эмбриональные фибробласты [205]: ENA SAMEA3106400; незрелые эритроциты [206]: SRR2983616 и SRR2983617) полученные из баз данных NCBI или ENA и обработанные с использованием программного обеспечения Tophat и CuffDiff [10,11]. Полученные значения FPKM были использованы для классификации всех генов на высокоэкспрессирующиеся (25% генов с наивысшим уровнем экспрессии) и низкоэкспрессирующиеся (25% с наименьшим уровнем экспрессии).

Локализация некодирующих элементов генома и эпигенетических меток

Ключевой проблемой в определении сайтов связывания CTCF является отсутствие прямых экспериментальных данных о связывании белка для выбранных нами типов клеток. В свободном доступе находились данные о локализации белка CTCF для эмбриональных эритроцитов *G. gallus* на 5 и 10 день [170], а также для клеточной линии HD3 [209]. Так как согласно литературным данным, значительная

доля как границ ТАДов [5], так и сайтов связывания CTCF, совпадают в разных типах клеток [170], для исследования были использованы те сайты связывания CTCF, которые являются общими для разных клеточных линий *G. gallus*.

Координаты консервативных некодирующих элементов (CNE) были взяты из базы данных Ancora (http://ancora.genereg.net/) [210]. Были рассмотрены элементы со сходством не менее 70% (с параметром C = 50). Координаты мобильных элементов генома были взяты из базы данных TranspoGene (http://transpogene.tau.ac.il/) [211].

Используемые в работе координаты F1-хроматиновых доменов, H3K4me3 и H3K27ac пиков были взяты по данным для незрелых эритроцитов [206].

Определение регионов эволюционных перестроек хромосом

Для исследования взаимосвязи архитектуры хроматина с эволюционными событиями межхромосомных перестроек использовались координаты регионов эволюционных перестроек хромосом из литературных данных [189]. Из полного списка регионов были отобраны только те, которые отсутствуют в эволюционной линии куриных и имеет длину не более 100 тысяч пар оснований. Регионы длиной более 40 тысяч пар оснований были разделены на последовательно расположенные участки длиной не более 40 тысяч пар оснований в соответствии с используемым разрешением ТАДов.

Сравнение архитектуры хроматина у разных видов с помощью индекса вариации информации (VI)

Для сравнения пространственной организации хроматина у разных видов был адаптирован индекс вариации информации (VI) [212]. Данная метрика разработана для сравнения сходства/различия разных способов кластеризации одного и того же множества и была адаптирована для сравнения архитектуры хроматина.

Подсчёт индекса VI проводился следующим образом. Пусть N – множество генов-ортологов для рассматриваемых видов S и S'. K – множество границ доменов для вида S. Для удобства выразим каждый отдельный ген ортолог и границу домена через геномную координату центра бина, в котором они расположены. Тогда, для каждой границы домена k вида S зададим:

$$B_k = \{n \in N, k \in K: |n - k| \leq D\} \quad (*)$$

где B_k - подмножество генов-ортологов вида S, таких что расстояние между геном и границей домена k не более D. Аналогично зададим подмножества $B'_{k'}$ для вида S'.

В соответствии с этим определим энтропию архитектуры хроматина у сравниваемых видов:

$$P(B_k) = \frac{|B_k|}{N\vee},$$

$$H(B_k) = \sum_k P(B_k) \times logP(B_k) \quad (**)$$

где $P(B_k)$ – вероятность гена-ортолога попасть в окрестность границы домена.

H(B_k) - энтропия архитектуры хроматина выбранного вида как способа разбиения генов-ортологов на подмножества.

Аналогичным образом определим и взаимную информацию архитектуры хроматина у сравниваемых видов:

$$P(B_k, B'_{k'}) = \frac{|B_k \cap B'_{k'}|}{N\vee},$$

$$I(B_k, B'_{k'}) = \sum_k \sum_{k'} P(B_k, B'_{k'}) \times log \frac{P(B_k, B'_{k'})}{P} \quad (***)$$

где $P(B_k, B'_{k'})$ - вероятность того, что ген-ортолог окажется общим у выбранных подмножеств у сравниваемых видов

$I(B_k, B'_{k'})$ - взаимная информация способов разбиения на подмножества генов-ортологов у сравниваемых видов.

Используя формулы (**) и (***) метрику различия архитектуры хроматина сравниваемых видов определим как:

$$VI(B_k, B'_{k'}) = H + H' - 2I \quad (1)$$

где VI - индекс вариации информации.

У полученной величины есть один недостаток: пределы её значений зависят от количества кластеризуемых элементов, в нашем случае – от количества генов-ортологов. В соответствии с этим было решено нормализовывать индекс VI на величину, равную среднему различию для случайных доменов. Нормализация проводилась следующим образом:

Пусть B^{rand} и B'^{rand} – разбиение генов-ортологов на группы в результате случайной перестановки доменов внутри хромосом сравниваемых видов, тогда:

$$VI^{rand} = \frac{\sum VI(B^{rand}, B'^{rand})}{N},$$

$$\overline{VI}(B_k, B'_{k'}) = \frac{VI(B_k, B'_{k'})}{VI^{rand}} \quad (2)$$

где VI^{rand} – разница между видами, если бы сходство/различие архитектуры хроматина было случайным, определяется как среднее значение индексов VI для N наборов случайных границ доменов;

$\overline{VI}(B_k, B'_{k'})$ - нормализованная мера различия архитектуры хроматина сравниваемых видов.

Таким образом $\overline{VI} \approx 0$ – означает, как полное совпадение архитектуры хроматина, $\overline{VI} \approx 1$ – означает, что любые совпадения в архитектуре хроматина ничем не отличаются от случайных.

3.2. Анализ архитектуры хроматина комаров рода *Anopheles*

Для сравнения пространственной организации хроматина в комарах рода *Anopheles*, использовались данные Hi-C эксперимента, любезно предоставленные канд. биол. наук Шараховым И.В. и Лукьянчиковой В.А. Данные экспериментов Hi-C обрабатывали с помощью ПО Juicertools[37]. Виды, включённые в исследование, материал, использованный для получения Hi-C-библиотек и применяемые версии генома указаны в Таблице 2.

Для возможности проведения анализа все геномы предварительно улучшались или собирались *de novo* с помощью ПО 3D-DNA [195], позволяющего в автоматическом и ручном режиме, на основе данных Hi-C эксперимента, проводить сборку хромосом из скаффолдов, обнаруживать и исправлять ошибки сборки скаффолдов. Нуклеотидные последовательности скаффолдов были получены из открытой базы данных VectorBase [213].

Таблица 2. Описание используемых в исследовании видов, происхождение данных Hi-C и используемые сборки генома. Звёздочкой отмечены версии геномов, полученные из базы данных VectorBase.

Вид	Материал использованный в Hi-C эксперименте	Используемая версия генома
An. albimanus	тотальные эмбрионы	STECLA 2.6*
An. atroparvus	тотальные эмбрионы	EBRO 3.1*
An. coluzzii N'Goussa	тотальные эмбрионы	-
An. merus	тотальные эмбрионы / взрослые особи	-
An. stephensi	тотальные эмбрионы	INDIAN 2.3*

Описание архитектуры хроматина

Расчёта A/B-компартментов проводился с использованием алгоритма входящего в программынй пакет Juicertools [204]. Этот алгоритм использует

изменения знака значений первого собственного вектора матрицы Hi-C контактов для разбиения геномных локусов на компартменты. Далее, данная величина – значение первого собственного вектора – и производные от неё, использованные для определения принадлежности локуса к тому или иному компартменту, будут именоваться величиной компартмента. Поскольку собственный вектор определен с точностью до знака, для каждой хромосомы знак величины первого собственного вектора коррелировался с GC-составом, плотностью генов и данными секвенирования РНК и исходя из величины корреляции знак выбирался так, чтобы положительные значения (А-компартмент) соответствовали участкам активного хроматина, с активной экспрессией.

Расчёт силы компартментализации проводился по методике, предложенной Nora E.P. с коллегами [72], с той разницей, что учитывались контакты бинов принадлежащих к 25% наиболее сильных А- и В-компартментов.

Координаты доменов были рассчитаны с использованием программного пакета hicExplorer [60] и любезно предоставлен Таскиной А.К.

Для каждого бина также рассчитывался индекс инсуляции на основе силы частоты контактов окружающих бин локусов [203]. Вычисления проводились аналогично тому, как это делалось для позвоночных.

Определение координат генов и повторов

Для комаров видов *An. albimanus*, *An. atroparvus* и *An. stephansi*, использовались координаты генов и их экзон-интронная структура согласно открытой базы данных VectorBase [213], которые пересчитывались на координаты улучшенных сборок геномов.

Для геномов *An. coluzzii* и *An. merus*, которые собирались на основе ранее не аннотированных данных, для определения предполагаемых позиций генов перекартировали нуклеотидные последовательности экзонов из сборок генома, полученных на других колониях комаров. Так, для *An. coluzzii Ngousso* использовали последовательности экзонов *An. coluzzii* MOPTI 1.8, а *An. merus* - по данным для *An. merus* MAF 2.9.

Позиции экзонов с помощью ПО LastZ версии 1.02.00 [216] выравнивались на целевые геномы. В соответствии с рекомендациями, представленными в руководстве к ПО, был использован следующий набор параметров: --gfextend --nochain --gapped. Величина параметра --hpstreshold, отвечающая за минимальную длину участков гомологии и качество их выравнивания, на основе ожидаемой длины экзона, была выбрана равной 6000. По координатам выравниваний, полученных в стандартном maf-формате, извлекались координаты экзонов генов на целевом геноме. В соответствии с этим, старт гена принимался за начало первого экзона, конец гена – конец последнего экзона.

Аннотация повторов de novo проводилась с помощью конвейера программ RepeatModeler [217]. База данных, по которой производилась классификация повторов, строилась на основе геномов всех использованных в исследовании видов комаров.

Все вычисление производились на узлах высокопроизводительного кластера Новосибирского Государственного Университета и компьютерного кластера Института Цитологии и Генетики (Бюджетный проект №0324-2019-0041). Алгоритм C-InterSecture реализован на языке Python2.7.

3.3. Конвертирование геномных координат между разными видами

Для конвертирования геномных координат использовались карты синтении, представленные net-файлами, для *H. sapiens*, *M. musculus* и *G. gallus* получали из базы данных UCSC. Для остальных видов карты синтении строились отдельно. В первую очередь, используя ПО LastZ версии 1.02.00 генерировались попарные выравнивания для выбранных видов в стандартном для множественных выравниваний maf-формате. В соответствии с рекомендациями, представленными в руководстве к ПО, был использован следующий набор параметров: --gfextend --nochain --gapped. Величина параметра --hpstreshold, отвечающая за минимальную длину участков гомологии и качество их выравнивания, для практически всех пар видов была выбрана как 6000, так как большие эволюционные расстояния разделяющие виды, предполагают наличие коротких участков синтении. Исключении составляет пара видов *An. coluzzii* и *An. merus,* для которой величина данного параметра равнялась 30000, благодаря чему в анализе получали более протяжённые выравнивания, что и ожидается для настолько близких видов. Полученные maf-файлы с помощью «maf-convert» из пакета программ LAST версии 0.963, конвертировались в файлы стандартного chain-формата. Полученные файлы, с помощью программ «chainSort» и «chainNet» из пакета kentUtils версии 3.02 конвертировались в net-файлы, которые и использовались как карты синтении.

4. Результаты

4.1. Характеристика пространственной организации генома *G. gallus*

Параметры доменов

Первые работы, использующие метод Hi-C, позволили получить детальные карты трехмерных контактов хроматина в различных типах клеток млекопитающих. В то же время, организация хроматина у птиц на момент выполнения этой работы не была изучена. Чтобы заполнить этот пробел, был проведён детальный анализ доменов хроматина в фибробластах и эритроцитах домашней курицы, *G. gallus*.

Необходимо предварительно указать, что под «доменом» в данной работе понимаются любые участки хроматина, отличающиеся повышенным числом контактов внутри себя и изолированные от ближайшего окружения; «ТАДами» же в рамках этой работы называются только домены, формируемые по механизму протягивания петли.

Для каждой изучаемой линии клеток *G. gallus* были сгенерированы наборы доменов, используя три различных алгоритма. В Таблице 3 приведены основные характеристики полученных доменов. Так как алгоритм TADtree позволяет выделить несколько уровней организации, в Таблице 3 представлены параметры для первых двух уровней: на более низких уровнях обнаруживаются единичные домены или не обнаруживаются вовсе.

Сравнение разных алгоритмов показывает, что характерные размеры доменов близки к тем, которые указаны для других позвоночных [5, 28, 30]. Так для доменов, выделенных с помощью алгоритмов Armatus и TADtree, медианная длина составляет около 200-250 тысяч п.о, а для доменов, выделенных с помощью DomainCaller – около 1 млн п.о. в зависимости от типа клеток.

Сравнение результатов работы алгоритмов показывает, что большая часть границ, выделенных с помощью алгоритма DomainCaller совпадает с границами, выделенными другими алгоритмами. Но алгоритмы Armatus и TADtree детектируют также ещё и множество дополнительных границ (Таблица 3).

Таблица 3. Параметры доменов, выделенных разными алгоритмами для разных типов клеток *G. gallus*, для TADtree доменов указаны параметры для первых двух иерархических уровней.

Тип клеток	Алгоритм выделения ТАДов	Число ТАДов	Суммарная длина, млн. п.н.	Покрытие генома, %	Средняя длина ТАДов, тыс. п.н.	Медианная длина ТАДов, тыс.п.н.
Эмбриональные фибробласты	Armatus	3102	954,44	93,44	310	240
	DomainCaller	1252	962,84	94,26	770	640
	TADtree (уровень 0)	2857	757,16	74,12	270	160
	TADtree (уровень 1)	870	208,72	20,43	240	200
Незрелые полихроматические эритроциты	Armatus	3032	781,80	76,54	260	200
	DomainCaller	553	927,04	90,76	1420	1240
	TADtree (уровень 0)	1807	623,08	61,00	340	200
	TADtree (уровень 1)	733	164,28	16,08	220	200
Зрелые эритроциты	Armatus	3485	806,60	78,96	230	200
	DomainCaller	805	903,00	88,40	1120	920
	TADtree (уровень 0)	1567	610,28	59,75	390	240
	TADtree (уровень 1)	609	129,60	12,69	210	160

Визуальный анализ карт Hi-C показывает, что большая часть новых границ расположены внутри «больших» доменов, детектированных DomainCaller (Рисунок 5). А различия в количестве выделенных границ соответствует различиям в средней и медианной длине доменов.

Таким образом, в пределах отдельной клеточной линии, можно говорить о том, что разные алгоритмы показывают в целом одинаковые результаты, а наблюдаемые различия обусловлены их многочисленными параметрами и «чувствительностью» к небольшим различиям в числе контактов между разными локусами.

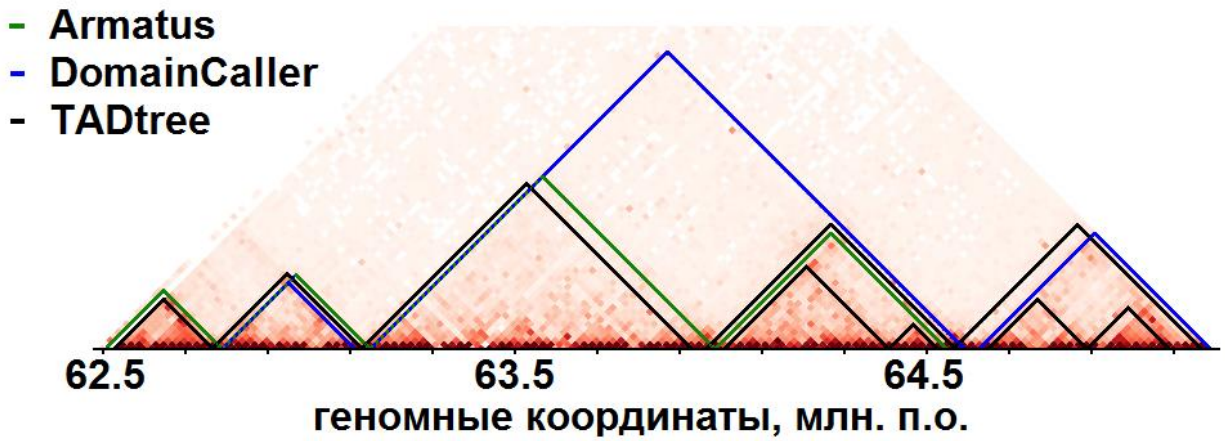

Рисунок 5. Пример выделения доменов разными алгоритмам для хромосомы 1 фибробластов G. gallus. Зелёным цветом обозначены домены выделенные алгоритмом Armatus, синим – DomainCaller, чёрным – TADtree.

В соответствии с этим определённый интерес представляет то, насколько сходятся результаты выделение доменов на разных клеточных линиях. Исходя из литературных данных, для одинаковых алгоритмов сходство ожидается на уровне 70-80% [5]. В зависимости от используемого алгоритма (Таблица 4), совпадение доменной организации фибробластов и эритроцитов составляет от 10% до 50%. Для клеток эритроидного ряда сходство составляет около 55-65% вне зависимости от избранных алгоритмов.

Таблица 4. Доля границ доменов, выделенных на образце 1, и покрываемых границами доменов, выделенных на образце 2. Ф – эмбриональные фибробласты. НЭ – незрелые полихроматические эритробласты. ЗЭ – зрелые эритоциты.

1 \ 2		Ф			НЭ			ЗЭ		
		Armatus	Dixon	TADtree	Armatus	Dixon	TADtree	Armatus	Dixon	TADtree
Ф	Armatus	1	0,37	0,77	0,44	0,1	0,39	0,44	0,13	0,34
	Dixon	0,86	1	0,8	0,44	0,13	0,38	0,43	0,14	0,32
	TADtree	0,59	0,27	1	0,42	0,1	0,41	0,43	0,12	0,35
НЭ	Armatus	0,38	0,17	0,49	1	0,13	0,45	0,66	0,16	0,41
	Dixon	0,43	0,22	0,54	0,68	1	0,57	0,68	0,67	0,5
	TADtree	0,4	0,17	0,55	0,55	0,14	1	0,53	0,17	0,57
ЗЭ	Armatus	0,33	0,14	0,45	0,58	0,11	0,39	1	0,16	0,41
	Dixon	0,41	0,2	0,53	0,69	0,54	0,56	0,76	1	0,56
	TADtree	0,38	0,16	0,53	0,52	0,13	0,61	0,61	0,18	1

Высокое сходство доменной организации для зрелых и незрелых эритроцитов позволяет предположить, что наблюдаемые различия между фибробластами и

клетками эритроидного ряда связаны с существенными различиями в укладке хроматина в клеточном ядре.

Кроме этого, такое различие требует ответа ещё на один вопрос: действительно ли архитектура хроматина у *G. gallus*, как в фибробластах, так и в эритроцитах и их предшественниках, определяется теми же механизмами, что и у других позвоночных.

В соответствии с этим было решено изучить взаимосвязь между выделенными разными алгоритмами доменами и известными геномными и эпигеномными элементами.

Поиск связи характеристик генома с архитектурой хроматина

Исходя из закономерностей, выявленных для других позвоночных, и находящихся в свободном доступе данных, была исследована взаимосвязь между архитектурой хроматина и следующими характеристиками генома: сайты связывания CTCF, плотность генов и уровень их экспрессии, эпигенетические метки F1-хроматина, плотность CNE и районов хромосомных перестроек.

Поскольку методы обработки результатов эксперимента Hi-C основываются на том, что бин является минимальной неделимой точкой, координаты исследуемых геномных элементов округлялись согласно используемому разрешению.

Для проверки значимости полученных характеристик использовался перестановочный тест, в ходе которого на каждой хромосоме отдельно производилась случайная перестановка границ доменов, с сохранением размеров доменов и междоменных пространств. На основании выборки объёмом не менее ста перестановок определялось выборочное среднее и среднеквадратичное отклонение. Рассчитанные характеристики считались значимыми только при условии, что они отличались от выборочного среднего не менее чем на три среднеквадратичных отклонения.

В первую очередь интерес представляло взаимное положение границ доменов и сайтов связывания CTCF, так как, согласно литературным данным, белок CTCF

имеет ключевое значение для формирования ТАДов по механизму протягивая петли. Исследование показало, что у эмбриональных фибробластов сайты связывания CTCF в границах доменов встречаются в 1,5-2 раза чаще, чем в среднем по геному (Рисунок 6) не зависимо от используемых алгоритмов для выделения ТАДов.

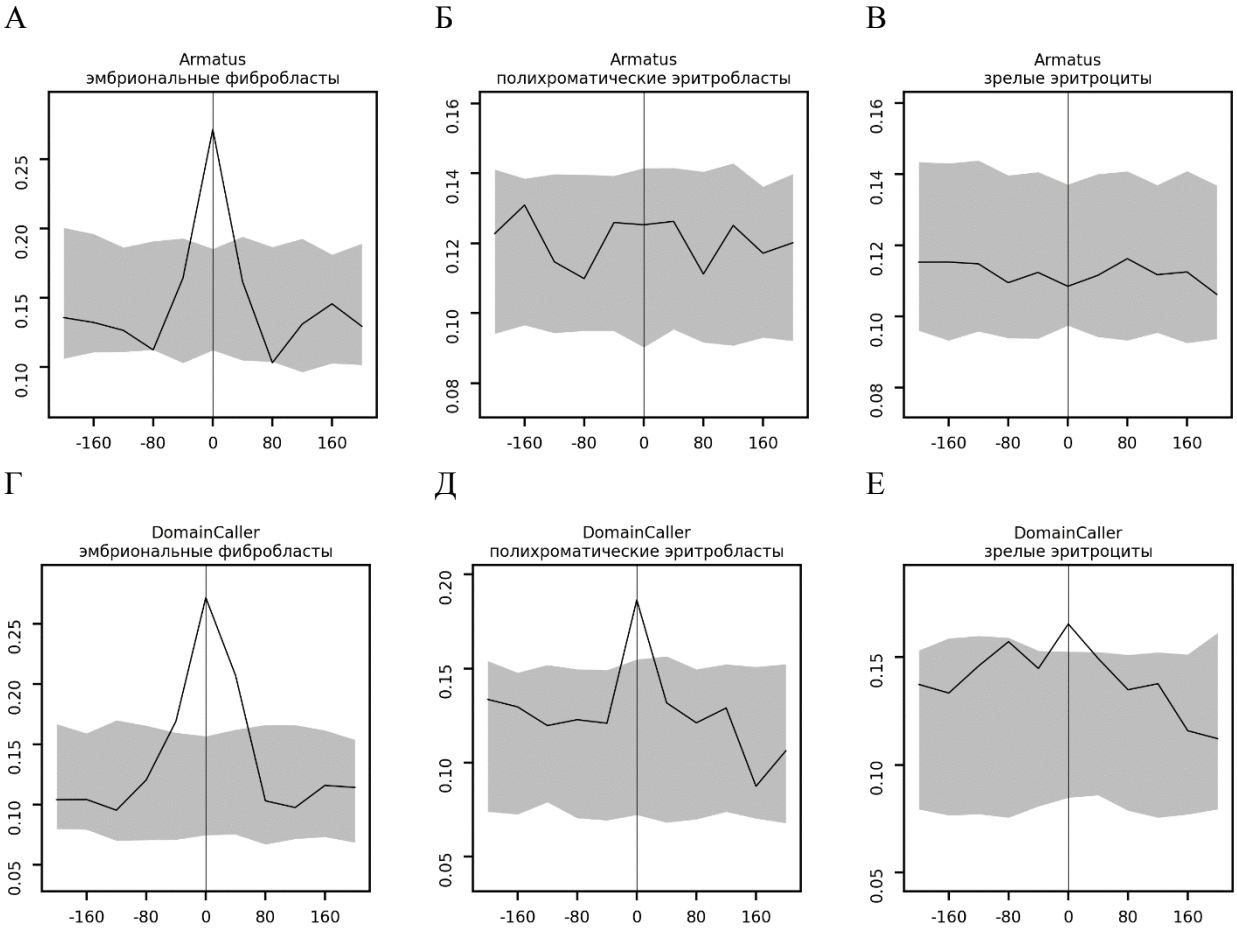

Рисунок 6. Распределение сайтов связывания CTCF относительно границ доменов. Результаты показаны для границ доменов, выделенных алгоритмами DomainCaller и Armatus. По оси X указано расстояние в тысячах пар оснований от границы, по оси Y — среднее число сайтов связывания CTCF на 40 тысяч пар оснований. Чёрная линия – наблюдаемые данные. Серая область – 3 стандартных отклонения от ожидаемого.

Что важно, наблюдается и зависимость от геномной ориентации мотива связывания белка CTCF: домены преимущественно формируются конвергентно направленными сайтами связывания CTCF (Рисунок 7). Для клеток эритроидного ряда эта закономерность выражена слабее и практически отсутствует у эритроцитов.

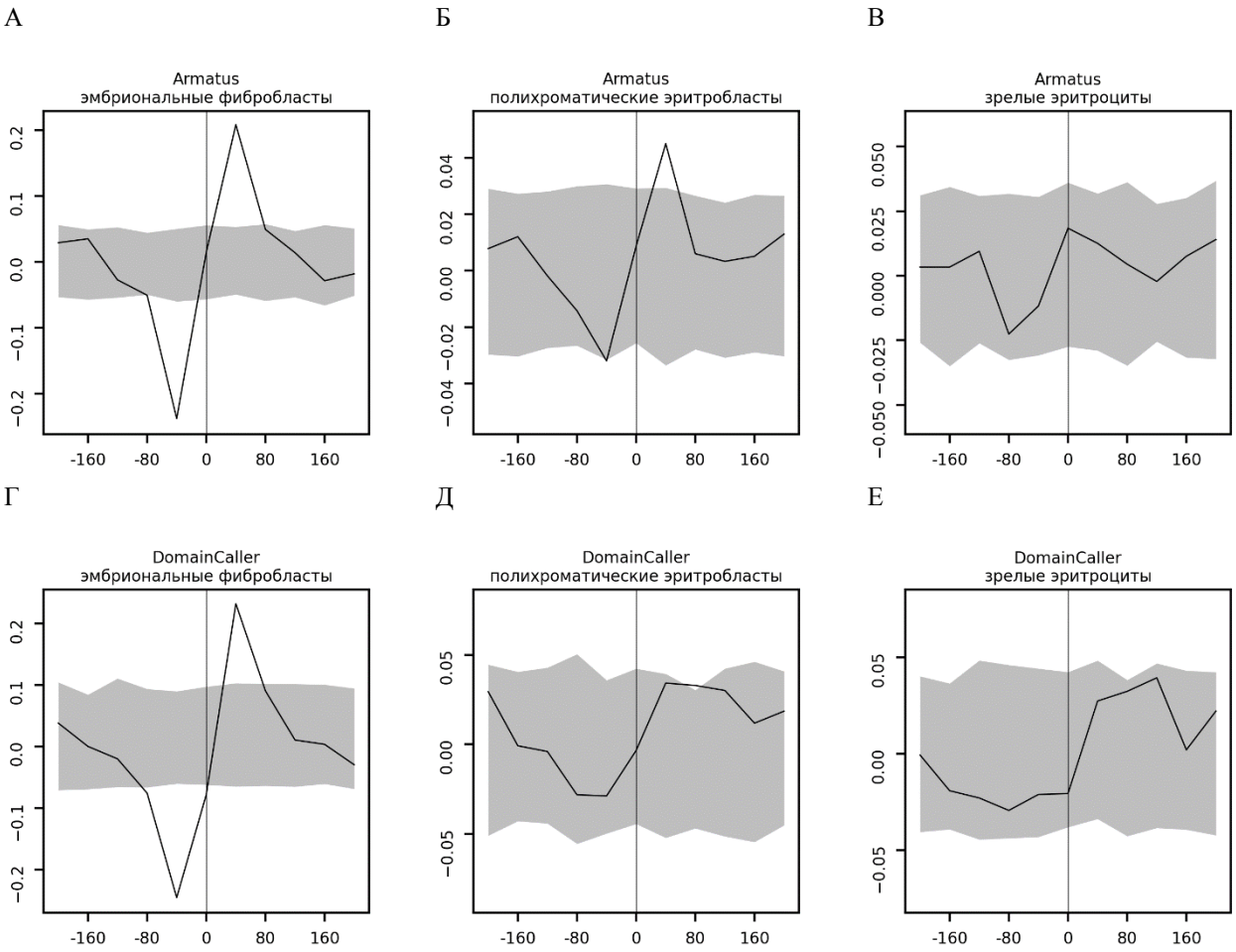

Рисунок 7. Распределение сайтов связывания CTCF относительно границ доменов с учётом геномного направления. По оси X указано расстояние в тысячах пар оснований от границы от границы домена. По оси Y — среднее число сайтов связывания CTCF с учётом геномного направления на 40 тысяч пар оснований. Если направление сайта совпадает с геномным, сайт считается со знаком «+», в противоположном случае – «-». Чёрная линия – наблюдаемые данные. Серая область – 3 стандартных отклонения от ожидаемого.

Этот факт позволяет достаточно уверенно утверждать, что в фибробластах *G. gallus* формирование доменов происходит по тем же механизмам, что формирование ТАДов у других позвоночных.

Однако, отсутствие связи между границами доменов и сайтами связывания CTCF у эритроцитов не позволяет в полной мере отвергнуть предположение, что в формировании архитектуры хроматина эритроцитов *G. gallus* также участвует механизм протягивания петли. Во-первых, порядка 15% ТАДов, согласно данным ChIP-seq, формируется без участия белка CTCF. Во-вторых, кроме CTCF существуют и другие белки, способные выполнять сходные с ним функции,

например, BORIS. В соответствии с этим, было решено проверить, несут ли выделенные нами домены у *G. gallus* ту же функцию – формирование регуляторных блоков – что и у других позвоночных. Если это так, то ожидается, что границы доменов будут обогащены генами, и обеднены CNE.

Сравнение показало, что данные закономерности в полной мере обнаруживают себя для фибробластов *G. gallus* (Рисунок 8), но не в клетках эритроидного ряда.

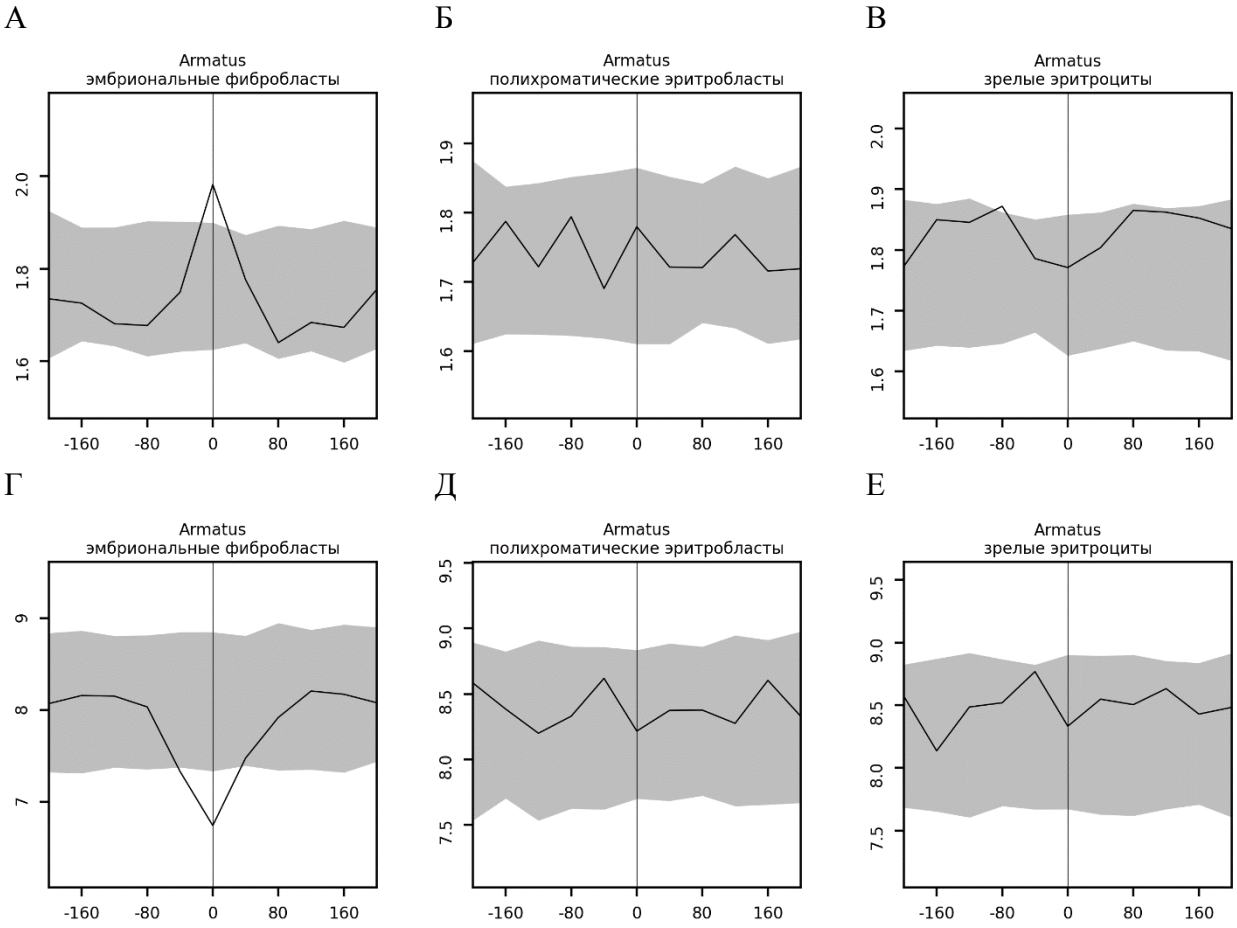

Рисунок 8. Генетических характеристик относительно границ доменов. По оси X указано расстояние в тысячах пар оснований от границы от границы домена. По оси Y для рисунков А-В – число генов на бин, для Г-Е – число CNE. Чёрная линия – наблюдаемые данные. Серая область – 3 стандартных отклонения от ожидаемого

Ещё одной закономерностью, известной по литературным данным для млекопитающих, является обогащение границ ТАДов метками активного

хроматина. Наличие данной закономерности также было проверено для взятых в исследование клеточных типов *G. gallus* (Рисунок 9).

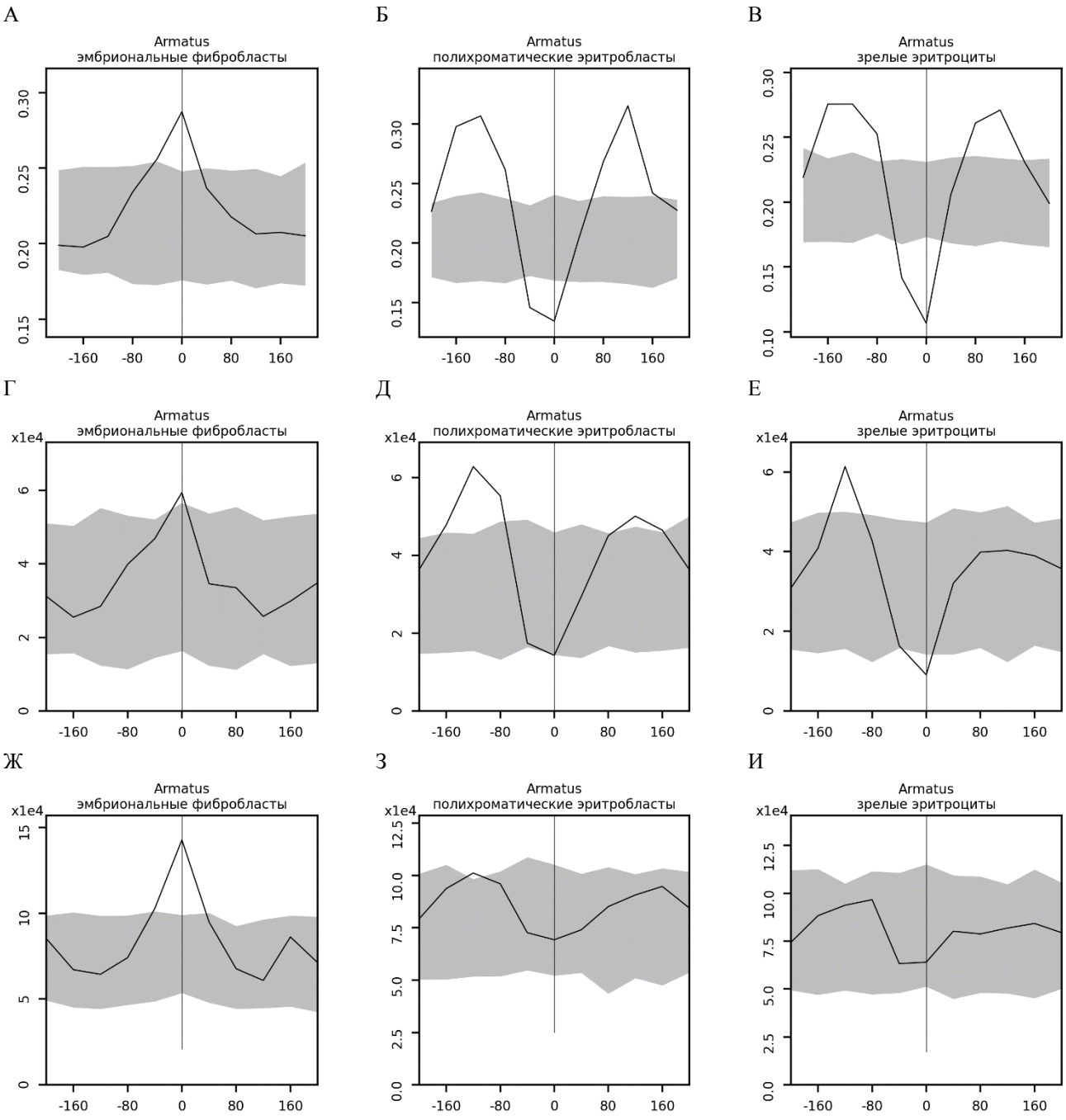

Рисунок 9. Распределение эпигенетических характеристик относительно границ доменов. По оси X указано расстояние в тысячах пар оснований от границы от границы домена. По оси Y для рисунков А-В – покрытие F1-доменами. Г-Е – покрытие пиками H3K4me3. Ж-З – покрытие пиками H3K27ac. Чёрная линия – наблюдаемые данные. Серая область – 3 стандартных отклонения от ожидаемого.

Исследование показало, что границы доменов у фибробластов значимо обогащены генами, метками активного хроматина и районами эволюционных перестроек (Рисунки 9А, 9Г, 9Ж). В то же время, для зрелых и полихроматических эритроцитов наблюдается весьма примечательная картина: непосредственно границы доменов обеднены метками активного хроматина, но обогащены их непосредственные окрестности (Рисунок 9).

Данное наблюдение позволяет предположить, что выделенные алгоритмами ТАДы могут связаны с разделением генома на блоки эу- и гетерохроматина. Если это предположение верно, то соседние домены должны принадлежать разным типом хроматина и/или граница доменов должна совпадать с районами изменения состояния хроматина. Для проверки этой гипотезы использовалась, рассчитынная по картам Hi-C величина компартментализации хроматина (Рисунок 10).

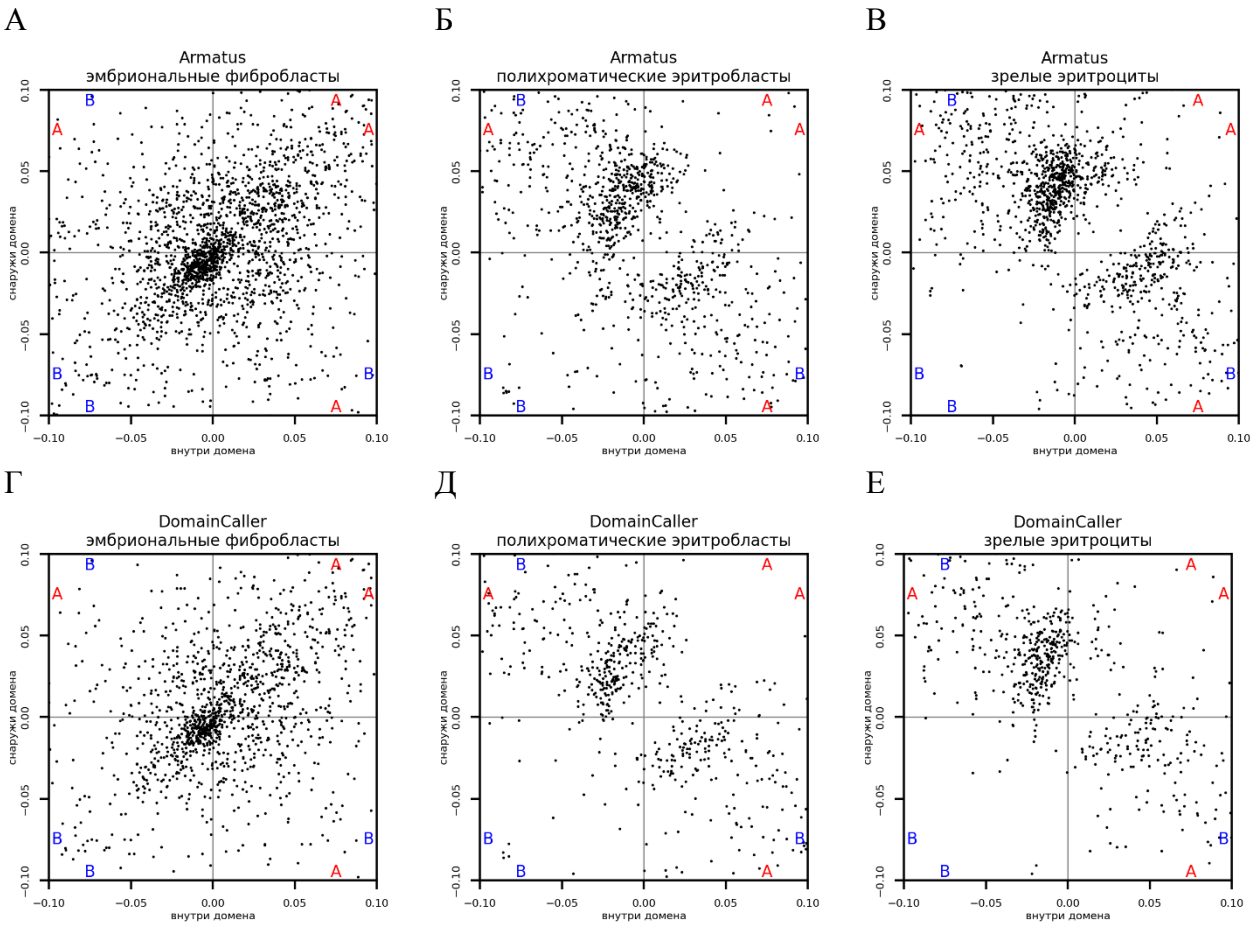

Рисунок 10. Значение первой главной компоненты в ближайших окрестностях границы домена. По оси X отложено значения внутри домена, по оси Y – снаружи домена.

Результаты показывают, что у фибробластов *G. gallus* границы доменов проходят по произвольным местам и никакой зависимости между значением компартментализации соседних доменов нет (Рисунки 10А и 10Г). В то же время для зрелых и полихроматических эритроцитов прослеживается строгая закономерность, что если целевой домен принадлежит А-компартменту, то его соседи – В-компартменту и наоборот (Рисунки 10Б, 10В, 10Д и 10Е).

Таким образом, исследование архитектуры хроматина на разных клеточных линиях *G. gallus* показало, что пространственная организация хроматина у фибробластов *G. gallus* обладает теми же свойствами, что и у млекопитающих, и подвержена тем же закономерностям. Это позволяет с высокой степенью уверенности говорить о том, что домены у фибробластов *G. gallus* формируются теми же механизмами, что описаны у млекопитающих, и являются ТАДами.

В то же время архитектура хроматина зрелых и полихроматических эритроцитов определяется преимущественно компартментализацией хроматина, чем и обусловлены наблюдаемые нами различия между типами клеток.

4.2. Использование индекса VI для эволюционного сравнения организации хроматина позвоночных

Как уже было указано выше, главной особенностью ТАДов у млекопитающих, которая привлекает к себе внимание исследователей, является то, что ТАДы являются блоком регуляции генов. Необходимость поддержания нерушимости и цельности таких регуляторных блоков и обеспечивает, по всей видимости, эволюционную консервативность ТАДов у млекопитающих. Подтверждается это в равной мере и обнаруженной связью между ТАДами у млекопитающих и экстремально консервативными некодирующими элементами генома, за которыми обычно подозреваются регуляторные элементы генома [5, 33]. Учитывая тот факт, что ТАДы в фибробластах *G. gallus* по совокупности своих свойств соответствуют ТАДам у млекопитающих, можно предположить, что не только сам принцип пространственной организации хроматина является консервативным для позвоночных, но и каждый ТАД в отдельности является эволюционно консервативной единицей. В свою очередь, изменение пространственной организации хроматина может сигнализировать о важных эволюционных событиях, произошедших в рассматриваемом регионе.

Наиболее простым и доступным способом проверки консервативности ТАДов является перенос геномных координат границ ТАДов между разными видами на принципах гомологии. Однако у этого метода есть два важных недостатка. Во-первых, сам по себе тот факт, что какие-то границы ТАДов у разных видов не совпадают, может означать не отсутствие у одного из видов в интересующем нас месте границы ТАДа, а то, что из-за особенностей данных и выбранных параметров её не обнаруживает алгоритм. Во-вторых, тот факт, что граница доменов у сравниваемых видов находится в пределах одного синтенного региона, ещё не означает, что за пределами этого синтенного блока, но внутри анализируемого ТАДа, не произошло радикальных хромосомных перестроек, которые полностью изменили генетический ландшафт.

С учётом вышеуказанного и особенностей биологической функции, было решено за основу сравнения ТАДов взять то, как они разбивают на группы гены-

ортологи у разных видов. Если гены-ортологи у разных видов оказались сгруппированы одинаково – мы полагаем, что эти виды обладают одинаковой архитектурой хроматина. Чем больше же отличий наблюдается в разбиении на группы и в составе групп – тем больше разница в архитектуре хроматина сравниваемых видов.

В соответствии с этим, для проведения эволюционного анализа архитектуры хроматина *G. gallus*, *M. musculus* и *H. sapiens* был разработан метод основанный на расчёте индекса вариации информации (VI) [212]. Данный индекс применяется для сравнения разных способов кластеризации и обладает свойствами метрики, что позволяет *количественно* сравнивать, сходство и различие архитектуры хроматина у разных видов.

Границы доменов представляют собой не конкретную точку, координаты которой могут быть определены с точностью до десятка или сотен нуклеотидов, а протяжённый участок размером в десятки тысяч нуклеотидов. Кроме того, определение бина, в котором проходит граница, может изменяться в зависимости от выбранного алгоритма и его параметров. Поэтому непосредственное сравнение ТАДов по составу входящих в них генов-ортологов будет слишком чувствительным к техническим артефактам. Особенно это касается видов с высокой плотностью генов, в частности, *G. gallus*.

Учитывая эти особенности, в качестве отдельной группы генов-ортологов называются те гены-ортологи, которые оказываются в окрестностях одной и той же границы домена. Те же гены, которые не попали ни в одну из окрестностей границ, определяются в общую группу внутридоменных генов-ортологов.

Исходя из описанных выше соображений, нами была разработана метрика $\overline{VI}$ на основе индекса VI (формула (2) раздела «Данные и методы»), которую мы использовали для сравнения архитектуры хроматина. Для ее применения в первую очередь необходимо было определить, что следует считать окрестностями границы доменов. Так как разные виды обладают разной плотностью генов, было решено подобрать условия для определения окрестности так, чтобы среднее и медианные значения количества генов вблизи границы как можно более точно совпадало у

разных видов. После перебора параметров, окрестности границы домена были выбраны как 1/3 длины домена у *H. sapiens* и *M. musculus* и 1/9 длины домена у *G. gallus.* При использовании таких параметров в окрестностях границы домена у сравниваемых видов находилось, медианно, 3 гена-ортолога.

Сравнение пространственной организации хроматина с помощью VI индекса отражает замеченные ранее различия в организации доменов между фибробластами и эритроцитами *G. gallus* (Рисунок 11). Можно уверенно утверждать, что структура доменов незрелых и зрелых эритроцитов обладает ожидаемо высоким сходством друг с другом, и радикальным образом отличается от ТАДов в других видах позвоночных. Фактически, совпадения между организацией доменов зрелых и полихроматических эритроцитов *G. gallus* и организацией доменов других клеточных типов этого и других видов являются случайными. Это подтверждает наш вывод о том, что пространственная организация хроматина у эритроцитов *G. gallus* определяется в большей степени компартментализацией хроматина, а не механизмом протягивания петли.

В то же время, структура ТАДов фибробластов *G. gallus*, показывает приблизительно такой же уровень сходства со структурой ТАДов клеточных линий, происходящих из *H. sapiens* и *M. musculus*, как у данных видов друг с другом. Это согласуется с тем, что свойства ТАДов, выявленные у фибробластов *G. gallus* аналогичны описанным для млекопитающих. Совокупность этих факторов позволяет с достаточной степенью уверенности говорить, что организация ТАДов у *G. gallus*, за исключением отдельных клеточных типов, определяется теми же механизмами, что и у млекопитающих.

Примечательным выглядит тот факт, что разница между ТАДами, выделенными посредством разных алгоритмов, оказывается сопоставима с разницей между ТАДами разных типов клеток и видов позвоночных. Так, согласно индексу VI, разница между организацией ТАДов фибробластов *G. gallus*, выделенной с помощью алгоритма DomainCaller, и организацией ТАДов, выделенной с помощью алгоритмов Armatus или TADtree, оказывается такой же, как между разными типами клеток *H. sapiens* и *M. musculus*. Различия,

возникающие при использовании разных алгоритмов для описания организации ТАДов зрелых и полихроматических эритроцитов, превосходят межвидовые.

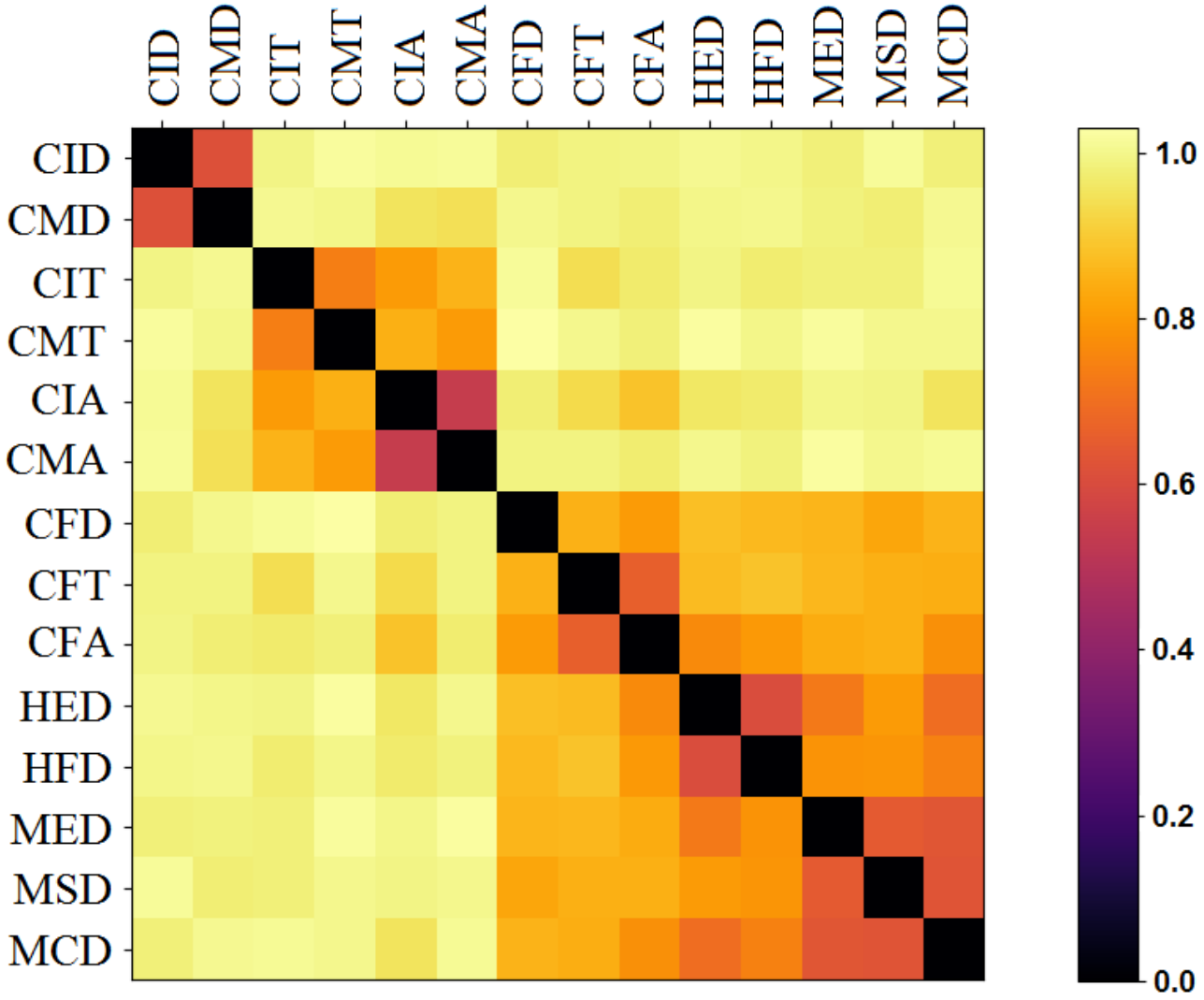

Рисунок 11. Значения индекса $\overline{VI}$ для сравниваемых клеточных линий по отношению к случайному распределению. Чем меньше значением — тем более похожи архитектура хроматина. Последняя буква отражает использованный алгоритм: A – Armatus, D — DomainCaller, T — TADtree; CFA, CFD, CFT – эмбриональные фибробласты *G. gallus*; CIA, CID, CIT – полихроматические эритроциты *G. gallus*, CMA, CMD, CMT — зрелые эритроциты *G. gallus*, HED — эмбриональные стволовые клетки *H. sapiens*, HFD – фибробласты IMR90, MED — эмбриональные стволовые клетки *M. musculus*, MSD — сперматозоиды *M. musculus*, MCD — клетки коры мозга *M. musculus*.

Важно отметить, что эволюционная консервативность ТАДов имеет исследовательскую ценность не сама по себе, но как признак эволюционной консервативности генной регуляции. Однако кроме ТАДов, особенности генной

регуляции отражают также и другие структуры пространственной организации хроматина, такие как петли или A/B-компартментализация.

Всё вышеперечисленное подводит к выводу, что эволюционное сравнение, основанное на сравнении тех или иных структурных особенностей пространственной организации хроматина, таких как ТАДы, является сугубо недостаточным, поскольку приводит к потери значительного объёма данных и может сильнейшим образом искажаться особенностями выбранных для описания архитектуры хроматина алгоритмов.

В соответствии с этим, было решено разработать такой метод эволюционного сравнения пространственной организации хроматина, который позволил бы сравнивать каждый индивидуальный контакт хроматина исследуемых видов.

4.3. Алгоритм сравнения пространственной организации хроматина, основанный на индивидуальных контактах

В виду того, что на момент проведения исследования отсутствовали алгоритмы, предназначенные для эволюционного сравнения пространственной организации на уровне отдельных контактов, было решено создать такой самостоятельно. В соответствии с особенностями данных Hi-C эксперимента и использованных нами видов, был сформулирован ряд проблем, которые разрабатываемый алгоритм должен был уметь решать.

Во-первых, во многих случаях, при эволюционном сравнении видов, отсутствует информация о том, какие конкретно нуклеотидные последовательности отвечают за формирование контактов и какой они вносят вклад. Не существует строгой взаимосвязи между степенью консервативности нуклеотидной последовательности и роли этой последовательности при формирования пространственной организации хроматина. Например, наиболее консервативными являются нуклеотидные последовательности экзонов генов, но сами экзоны не принимают практически никакого участия в пространственном взаимодействии локусов и их роль ничуть не больше, чем роль интронов. Таким образом, разрабатываемый алгоритм должен уметь оценивать консервативность контактов в отсутствии информации о свойствах и функции конкретных нуклеотидных последовательностей.

Во-вторых, очень часто расстояние между синтенными регионами у сравниваемых видов различается. Даже весьма близкие виды могут существенно отличаться по произошедшим в каждой из эволюционных линий геномным событиям, например, хромосомным перестройкам, потере/приобретении повторов и транспозонов. Для видов, которые не являются модельными, большое значением может иметь и неполнота сборки генома. В сумме это приводит к тому, что геномные расстояния между синтенными локусами и длина блоков синтении может существенно отличаться у сравниваемых видов. В свою очередь,

разрабатываемый алгоритм должен учитывать изменившиеся геномные расстояния и влияние этого фактора на архитектуру хроматина.

В-третьих, проблемой является биновый характер данных Hi-C. Для получения достоверных данных о трехмерной укладке хроматина необходимо объединять все контакты на протяжённом участке генома — бине, чья длина для большей части экспериментов составляет тысячи и десятки тысяч пар оснований. Так как границы бинов в большинстве случаев задаются механистически, начиная с нуля геномных координат на каждой хромосоме, то даже у очень близких видов лишь в редких случаях найдутся бины, которые будут у сравниваемых видов точно совпадать по набору синтенных участков. Таким образом, алгоритм должен учитывать, что консервативные последовательности одного бина в одном виде могут быть распределены между двумя или более бинами в другом виде. Точно так же между многими бинами, будут распределены и контакты, формируемые этими консервативными последовательностями.

Вышеописанные проблемы приводят к тому, что становится затруднительно сравнивать у исследуемых видов величины контактов непосредственно, так как на картах Hi-C мы видим контакты между бинами, объектом интереса являются контакты между синтенными регионами, но бины и регионы синтении друг другу не соответствуют. Для решения этой проблемы было решено пересчитывать величины контактов из контрольного вида в величины контактов исследуемого вида, таким образом, чтобы с одной стороны, учесть локальные особенности архитектуры хроматина контрольного вида (например, некие особенно сильные и слабые взаимодействия между целевыми регионами), а с другой стороны, учесть глобальные особенности архитектуры в исследуемом виде, например, более плотную упаковку хроматина в ядре. Такой процесс далее будет называться перекартированием контакта, а рассчитанная величина контакта – перекартированной. Таким образом, перекартированные контакты будут показывать то, как бы выглядела архитектура хроматина в исследуемом виде, если бы механизмы её формирования и цис-регуляторные последовательности функционировали так же, как у контрольного вида, а глобальные особенности

генома и организации хроматина в ядре остались бы неизменными. В соответствии с этим, если механизм формирования контактов был консервативен в эволюции, то наблюдаемая и перекартированная величина контакта будут близки.

В соответствии с этим необходимо решить, как соотнести друг с другом у сравниваемых видов блоки синтении и бины Hi-C. Наиболее простым решением выглядит представление числа контактов между двумя произвольными локусами как величины аддитивной. Это означает, что половина бина будет иметь половину контактов целого бина, а целый бин будет иметь столько контактов, сколько составляет сумма контактов слагающих его частей. Если руководствоваться этим предположением, то нет необходимости в точном знании того, какие конкретно нуклеотидные последовательности отвечают за формирование контактов между целевыми локусами у сравниваемых видов и какой их конкретно вклад, тем более, что такой информации может и не быть. В соответствии с этим, в процессе расчёта величины перекартирования контакта, информация о степени гомологии нуклеотидных последовательностей, их функции, положении и длине становится избыточной. Согласно этому, каждый бин разумно представить, как некое множество точек синтении, где каждая точка обуславливают свою, равную, долю контактов.

Модели консервативности контактов

В виду того, что геномные расстояния между синтенными локусами могут значительно меняться от вида к виду, а геномные расстояния являются важнейшим фактором, определяющим частоту контактов хроматина, предложено две модели консервативности контактов хроматина, которые строятся вокруг разных подходов к учёту влияния изменения геномного расстояния на архитектуру хроматина. *Абсолютная модель* консервативности подразумевает, что вне зависимости от произошедших геномных событий, консервативность подразумевает неизменность физического расстояния между синтенными локусами в пространстве ядра у сравниваемых видов. Данная модель основана на предположении, что функциональное значение для реализации генетической информации имеет

частота событий контакта между регулирующими элементами генома. Таким образом, изменение геномного расстояния между контактирующими локусами вследствие хромосомных перестроек или экспансии повторов является событием значимым для генной регуляции и нарушающим консервативность архитектуры хроматина. Альтернативой абсолютной модели, выступает модель *относительная,* подразумевающая, что ключевое значение имеет то, насколько частота контакта отличается от средней, ожидаемой на данном геномном расстоянии.

Следующие две модели предлагают альтернативные подходы к оценке вклада индивидуальных точек синтении внутри локуса на формирование его контактов. *Аддитивная* модель консервативности предполагает, что каждый консервативный элемент, каждая точка синтении, вносит свой вклад в формирование взаимодействия, и взаимодействие консервативно тогда, когда сохраняется взаимное расположение как можно большего числа консервативных регионов в сравниваемых локусах. В противовес аддитивной модели, *весовая* модель предполагает, что далеко не каждая точка синтении имеет какое-либо значение для формирования контакта между исследуемыми локусами. В этом случае, если одна или несколько точек синтении в ходе эволюции оказались перемещены на большое геномное расстояние от своего прежнего окружения, резонно предположить, что контакты локусов не изменятся. Таким образом, при перекартировании контактов должны учитываться только те точки синтении, которые перекартировались совместно.

Структура алгоритма

Учитывая вышеописанные особенности, был разработан алгоритм, воплощённый в виде ПО C-InterSecture (**C**omputional tool for **InterS**pecies analysis of genome archit**ecture**). Разработанный алгоритм включает в себя три главных этапа: предварительная обработка данных для выявления глобальных и локальных особенностей архитектуры хроматина сравниваемых видов, перекартирование контактов и визуализация результатов для последующего сравнения (Рисунок 12).

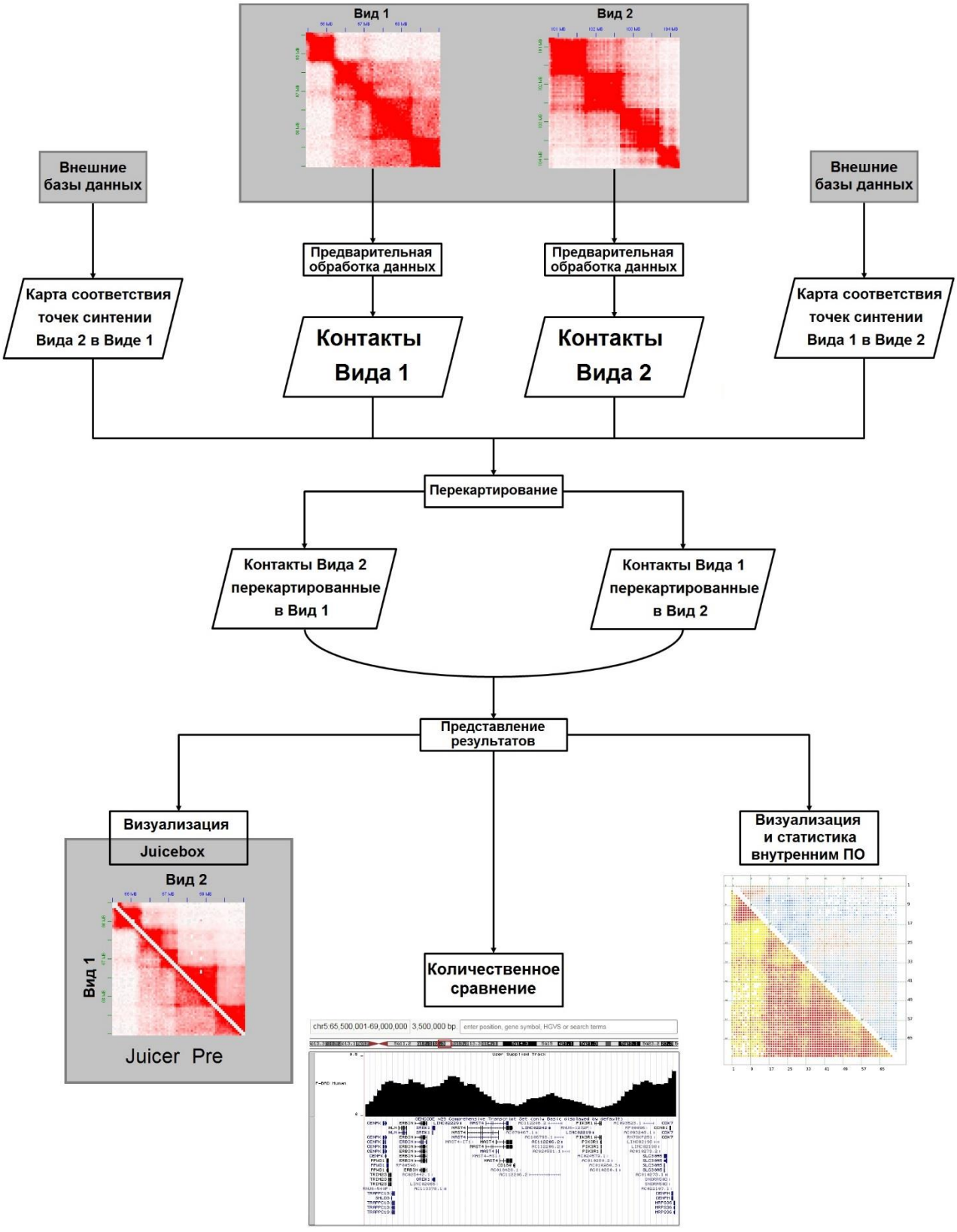

Рисунок 12. Принципиальная схема работы алгоритма перекартирования контактов.

Предварительная обработка данных

Предварительная обработка данных включает в себя фильтрацию, выравнивание контактов и статистический анализ. Шаг фильтрации контактов необходим, чтобы снизить влияние технических артефактов Hi-C-данных. Для этого из анализа исключаются контакты, представленные только небольшим числом прочтений Hi-C, бины с низким покрытием прочтениями или содержащие протяжённые участки с неопределённой нуклеотидной последовательностью. Параметры фильтрации зависят от особенностей конкретных данных Hi-C.

Шаг выравнивания контактов требуется, так как глубина секвенирования библиотек Hi-C существенно отличается между экспериментами, полученные величины нормализованных контактов должны быть дополнительно приведены в один масштаб, выравнены. Для этого к каждому контакту применяется функция выравнивания:

$$\hat{C}_{i,j} = \frac{C_{i,j}^{norm}}{\sum_k C_{i,k}^{norm}} \quad (1)$$

где $C_{i,j}^{norm}$ – нормализованное число контактов между бинами i и j, а $\hat{C}_{i,j}$ – выравненная частота контактов. Данная функция не только приводит разные данные Hi-C к одному масштабу, но и преобразует частоты контактов в абсолютные величины, при котором сумма контактов избранного локуса со всем геномом равна 1, то есть полученная величина отражает вероятность взаимодействия целевого локуса с другими.

Так как относительная модель консервативности контактов предполагает анализ превышений частоты контактов над средней частотой, наблюдаемой для данного геномного расстояния, ПО группирует все величины контактов в соответствии с геномным расстоянием, разделяющим контактирующие локусы. Таким образом, для каждого геномного расстояния получается своё распределение частот контактов. Большие расстояния, для которых число возможных контактов слишком мало, для получения распределения группируются с ближайшими к ним.

Далее, основываясь на множестве полученных распределений, введём функцию $P(\hat{C}_{i,j}) = PS_{ij}$, где PS_{ij} – перцентиль для нормализованной частоты

контактов $\hat{C}_{i,j}$в распределении всех частот контактов локусов, расположенных на соответствующем геномном расстоянии, и обратную ей функцию $P^{-1}(PS_{ij}) = \overline{\hat{C}}_{i,j}$, где $\overline{\hat{C}}_{i,j}$ – средний нормализованный контакт для перцентиля PS_{ij}. Перцентиль 0 соответствует наиболее слабым контактам между локусами, перцентиль 99 – наиболее сильным.

Таким образом, после предварительной обработки данных, каждому контакту соответствует две величины – выравненная частота, которая будет использоваться в абсолютной модели консервативности, и перцентиль контакта, для относительной модели. В зависимости от выбранной модели далее происходит перекартирование соответствующей величины.

Перекартирование контактов

Как уже было отмечено, основной проблемой при межвидовом сравнении контактов является то, что границы блоков синтении крайне редко соответствуют границам бинов. Чтобы решить эту проблему, каждый бин представляется как множество малых точек синтении (Рисунок 13). Для простоты положим, что каждая точка синтении имеет строго по одному отображению в сравниваемых видах, и точки, расположенные внутри одного бина, имеют отображения в другом виде в близкие бины. В данной работе близкими бинами считались те, которые были удалены друг от друга не более чем на 150 тысяч п.н. при сравнении млекопитающих, и не более 250 тысяч п.н. при сравнении млекопитающих с птицами. Основываясь на этом, для описания алгоритма перекартирования контактов введём следующию систему определений.

Будем считать что x – геномный локус в исследуемом виде, тогда $\tilde{x}$ синтенный локус в контрольном виде. R(x) – функция перекартирования, если $R(x) = \tilde{x}$ и $(\tilde{x}) = x$. Длину локуса x обозначим как $|x|$.

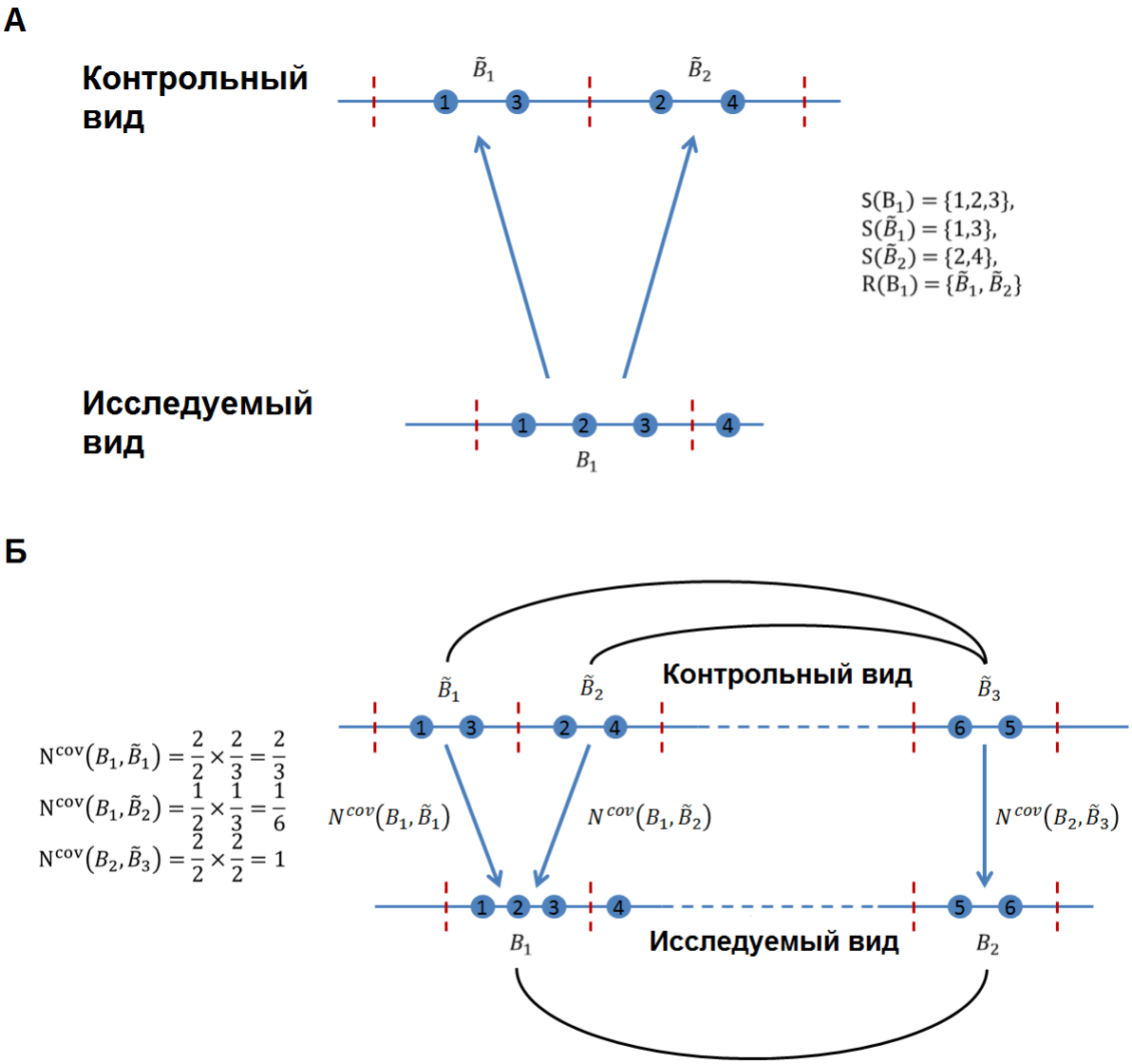

Рисунок 13. Принципиальная схема перекартировани контактов. А. Соотношение точек синтении между геномами контрольного и исследуемого вида. Б. Перекартирование контакта между бинами B_1 и B_2 из контрольного вида в исследуемый. Синие нумерованные круги означают точки синтении, красные вертикальные линии – границы бинов.

Пусть B – множество бинов генома, и B_i – бин i в референсном виде. Тогда, $S(B_i) = \{b_{i_1}, \ldots b_{i_k} \ldots b_{i_K}\}$ считается разбиением бина на точки синтении, если:

$$|b_{i_k}| \ll |B_i| \text{ и } \forall b_{i_k}: R(b_{i_k}) \neq \emptyset$$

$\forall b_{i_k}, b_{i_n}: b_{i_k} \cap b_{i_n} = \emptyset$ и $\forall b_{i_k}, b_{i_n}: R(b_{i_k}) \cap R(b_{i_n}) = \emptyset$

Таким образом, b_{i_k} – отдельная точка синтении, а $|S(B_i)|$ мощность разбиения, равное числу точек синтении. В соответствии с этим определим функцию перекартирования для бина B_i как $R(B_i) = \{\tilde{B}_1, .. \tilde{B}_n, .. \tilde{B}_N\}$, если:

$\forall \tilde{B}_n \in R(B_i): R(S(B_i)) \cap S(\tilde{B}_n) \neq \emptyset$

$\forall B_{i_k} \in S(B_i), \nexists \tilde{B}_{i_n}$ и $\tilde{B}_{i_m}: R(B_{i_k}) \cap \tilde{B}_{i_n} \neq \emptyset$ и $R(B_{i_k}) \cap \tilde{B}_{i_m} \neq \emptyset$

Другими словами, операция разбиения делит бин на множество неперекрывающихся точек синтении так, что каждая из них имеет строго одно отображение. Определим $N^{cov}(B_i, \tilde{B}_k)$ как нормализирующий коэффициент, отражающий долю взаимного покрытия исследуемого и перекартированного бина общими якорными точками:

$$N^{cov}(B_i, B_k) = \frac{|R(S(B_i)) \cap S(\tilde{B}_k)|}{|S(B_i)|} \times \frac{|R(S(\tilde{B}_k)) \cap S(B_i)|}{|S(\tilde{B}_k)|} \quad (2)$$

Другими словами, $N^{cov}(B_i, \tilde{B}_k)$ численно показывает степень синтении бинов B_i и $\tilde{B}_k$. Считаем, что **V(B$_i$, B$_j$)** является некой величиной, измеренной для бинов **B$_i$** и **B$_j$** (например, частоту или перцентиль контакта). Тогда $\tilde{V}(B_i, B_j)$ – значение перекартированной величины для бинов **B$_i$** and **B$_j$** определённое как:

$$\tilde{V}(B_i, B_j) = \frac{1}{K_{norm}} \times \sum_{\substack{\tilde{B}_k \in R(B_i) \\ \tilde{B}_n \in R(B_j)}} N^{cov}(B_i, \tilde{B}_k) \times N^{cov}(B_j, \tilde{B}_n) \times V(\tilde{B}_k, \tilde{B}_n) \quad (3)$$

$$K_{norm} = \begin{cases} 1, \text{для аддитивной модели} \\ \sum_{\substack{\tilde{B}_k \in R(B_i) \\ \tilde{B}_n \in R(B_j)}} N^{cov}(B_i, \tilde{B}_k) \times N^{cov}(B_j, \tilde{B}_n), \text{для весовой модели} \end{cases}$$

где K_{norm} – нормализующий коэффициент, выбираемый в соответствии с выбранной моделью консервативности контакта.

Таким образом, определим функции перекартирования $L(V) = \tilde{V}$, отображающее значение величины измеренной между выбранными бинами и перенесённой с генома одного вида на другой, с учётом уровня синтении бинов. Для перекартирования контактов эта функция применяется к каждому контакту.

В том случае, если в наблюдаемых данных нарушается условие, что каждая точка синтении имеет строго одно отображение, но при этом они перекартируются в соседние регионы – для каждого варианта функция L применяется независимо, но полученные результаты объединяются так же, как указано выше. Если нарушается условие близости перекартированных бинов, то из возможных вариантов перекартирования выбирается тот, в котором задействовано наибольшее число точек синтении.

Применение функции $P^{-1}\left(\widetilde{PS}_{i,j}\right)$ позволяет преобразовать величину перекартированных перцентилей, обратно в контакты, но уже характерные для исследуемого вида.

Оценка значимости различий между контактами

Одним из важных факторов определения различий между контактом перекартированным и наблюдаемым является определение достоверности этих различий. Очевидно, что достоверность сравниваемых величин зависит от количества прочтений, сформировавших данные контакты: чем их меньше, тем выше влияние случайных факторов на величину контакта. В соответствии с этим, оценка достоверности проводится следующим образом. Сначала определяется выборочная частота контакта h:

$$h = \frac{C_{i,j}^{raw}}{N} \tag{4}$$

где $C_{i,j}^{raw}$ – число прочтений, формирующих контакт между бинами i и j, N – суммарное число прочтений на карте Hi-C. Далее проводится расчёт относительной

ошибки, полученной при измерении числа контактов. Согласно закону больших чисел стандартное отклонение выборочной частоты контакта h можно рассчитать так:

$$\sigma^2(h) = \frac{1}{N} \times p \times (1 - p) \tag{5}$$

где $\sigma^2(h)$ – стандартное отклонение числа ридов.

Так как $M[h] = p$ и $p \ll 1$, то

$$CV_{i,j} = \frac{\sigma(h)}{h} = \frac{\sqrt{\frac{1}{N} \times \frac{C_{i,j}^{raw}}{N}}}{\frac{C_{i,j}^{raw}}{N}} = \sqrt{\frac{1}{C_{i,j}^{raw}}} \tag{6}$$

Для оценки статистической значимости величины контакта определяется его верхняя $V_{i,j}^{high}$ и нижняя $V_{i,j}^{low}$ границы в соответствии с выбранной моделью консервативности:

$$V_{i,j}^{high} = \begin{cases} \left(1 + CV_{i,j}\right) \times \hat{C}_{i,j}, \text{для абсолютной модели} \\ P\left(\left(1 + CV_{i,j}\right) \times \hat{C}_{i,j}\right), \text{для относительной модели} \end{cases} \tag{7.1}$$

$$V_{i,j}^{low} = \begin{cases} \left(1 - CV_{i,j}\right) \times \hat{C}_{i,j}, \text{для абсолютной модели} \\ P\left(\left(1 - CV_{i,j}\right) \times \hat{C}_{i,j}\right), \text{для относительной модели} \end{cases} \tag{7.2}$$

Данные величины используются на шаге перекартирования. Полученные при перекартировании значения используются для оценки ожидаемого размаха величины наблюдаемого и перекартированного контакта:

$$Dev\left(B_i, B_j\right) = max\left(V_{i,j}^{high} - V_{i,j}, V_{i,j} - V_{i,j}^{low}\right) \tag{8.1}$$

$$\widetilde{Dev}(B_i, B_j) = max(\tilde{V}_{i,j}^{high} - \tilde{V}_{i,j}, \tilde{V}_{i,j} - \tilde{V}_{i,j}^{low}) \quad (8.2)$$

Так, $Dev(B_i, B_j)$ – это ожидаемый размах значений контакта в референсном виде, а $\widetilde{Dev}(B_i, B_j)$ – перекартированный размах. В соответствии с этим, достоверность различий величины контактов определяется как суммарный размах перцентилей наблюдаемых и перекартированных контактов:

$$SC = 1 + Dev(B_i, B_j) + \widetilde{Dev}(B_i, B_j) \quad (9)$$

Чем величина SC выше – тем менее достоверны наблюдаемые различия.

Результаты работы C-InterSecture можно визуализировать как с помощью встроенных программ, так и сторонних средств (Рисунок 12). Использование встроенного ПО позволяет отображать на карте контактов не только частоты взаимодействий, но и их достоверность, так и достоверность различий между референсным и сравниваемым видом. Помимо этого полученные карты контактов можно конвертировать в формат, воспринимаемый популярным средством визуализации данных Hi-C – Juicebox.

Исследование влияния входных параметров на характеристики перекартирования контактов

Известно, что Hi-C-данные зависят от ряда локус-специфических факторов. Представленность конкретного локуса в Hi-C-данных может определяться не только работой конкретны механизмов, формирующих архитектуру хроматина, но и факторами технического характера: GC-состав локуса, полнота геномных данных в целевом локусе, доля повторов, и прочее. Это проблема известна для Hi-C-данных и частично решается использованием тех или иных способов нормализации. В этой работе использовались данные, нормализованные по алгоритму Кнайта-Руиза (Knight-Ruiz Matrix Balancing) с помощью ПО juicertools.

Оценка зависимости частоты контакта от свойств бина показывает, что контакты, частоты которых принадлежат высшим перцентилям, часто формируются бинами имеющими малое покрытие ридами Hi-C и/или большую долю неопределённых оснований.

Таким образом, представляется разумным исключить из анализа локусы генома, вносящие слишком большое искажение: имеющие слишком малое покрытие ридами Hi-C и/или слишком длинные участки неопределённых нуклеотидов. В соответствии с этим было решено исключать 1% бинов с наименьшим покрытием, а также те бины, у которых более 33% занимают участки с неизвестным нуклеотидным составом.

Сравнение пространственной организации хроматина у позвоночных

Более детальное сравнение архитектуры хроматина фибробластов *G. gallus*, с фибробластами *H. sapiens* и *M. musculus* (Рисунок 14) было проведено с использованием разработанного алгоритма перекартирования контактов. Сравнение на уровне отдельных контактов показало, что в значительной степени архитектура хроматина фибробластов *G. gallus* остаётся сходной с архитектурой хроматина фибробластов *H. sapiens* и *M. musculus*.

Одной из первых задач, возникших при сравнении разных видов друг с другом, оказалось определение того, что считать консервативным контактом. Первая гипотеза заключалась в том, что эволюционно-консервативный контакт сохраняет свою абсолютную частоту, соответствующую значению выравненного нормализованного контакта (выражение 4). В таком случае, если в ходе эволюции расстояние между взаимодействующими локусами изменяется – частота контакта остаётся неизменной. Вторая гипотеза предполагает, что консервативным остаётся относительная частота контакта, выраженная как отличие от среднего для данного геномного расстояния или перцентиль контакта в терминах C-InterSecture. В таком случае, при изменении расстояния между контактирующими локусами, частота контактов изменится так, чтобы сила контакта / перцентиль не изменялся.

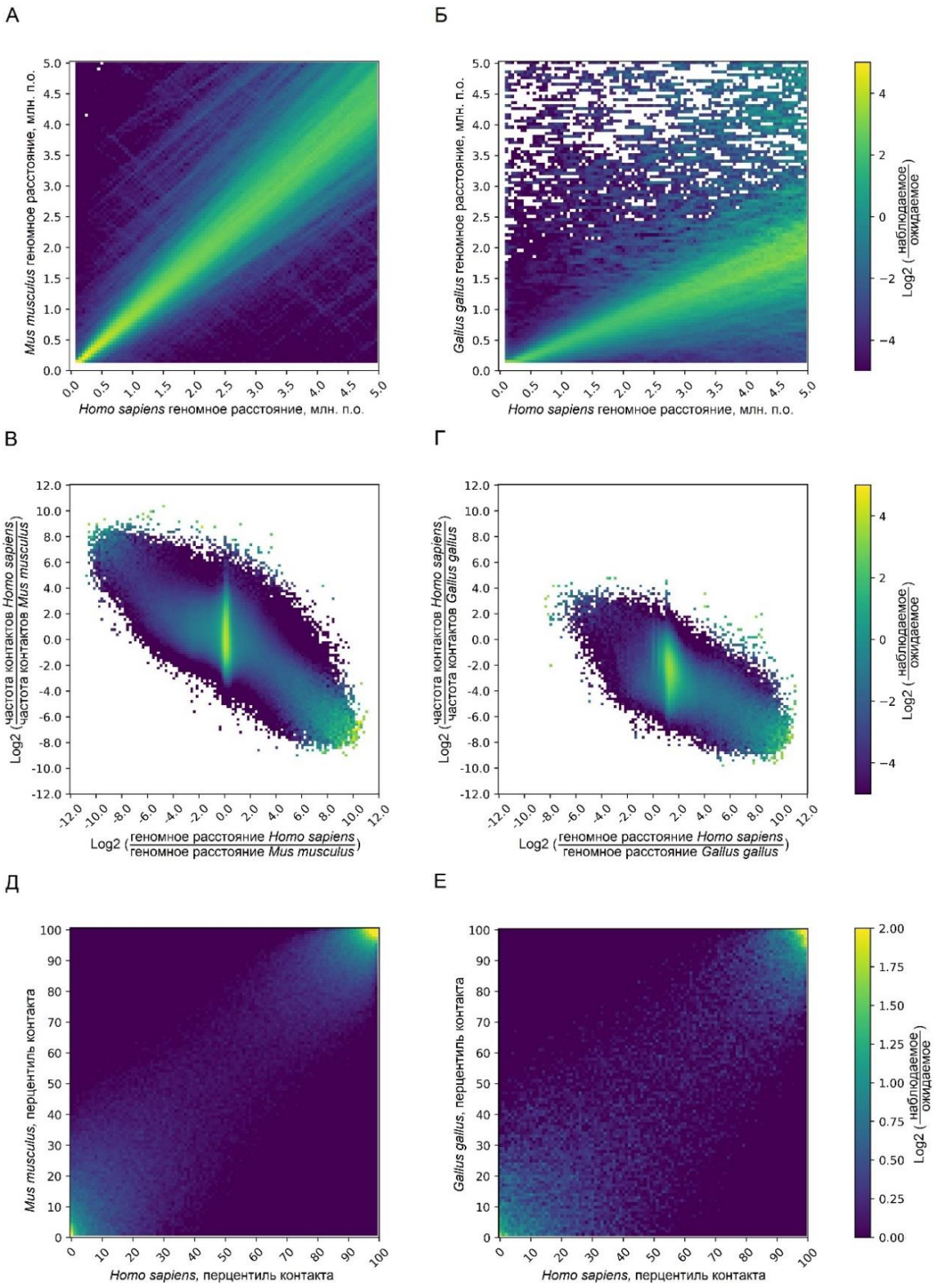

Рисунок 14. Изменение пространственных и геномных расстояний в ходе эволюции позвоночных. А, Б. Геномная дистанция между парами локусов у H. sapiens в сравнении с дистанцией между синтенными локусами у M. musculus (А) и G. gallus (Б). В. Изменения частот контактов между синтенными локусами у H. sapiens и M. musculus имеет слабую корреляцию с изменением расстояния (Пирсон R=-0.24). Г. Тоже, что и (В) для сравнения H. sapiens и G. gallus (Пирсон R=-0.44). Д, Е. Перцентили контактов синтенных локусов у H. sapiens и M. musculus (Д) или H. sapiens и G. gallus (Е) демонстрируют корреляцию средней силы (Д, Spearman R=-0.56; Е, Spearman R=-0.38).

Сравнение геномов G. gallus, H. sapiens и M. musculus показало, что геномное расстояние между большей частью локусов изменилось согласно изменению общего размера генома (Рисунок 14А,Б). Сравнение частот контактов показало, особенно ярко в паре H. sapiens – G. gallus (Рисунок 14 В,Г), что частоты изменились согласно изменению расстояний между синтенными локусами.

В то же время, для всех контактов, совпадение перцентиля, наблюдаемого и перекартированного, наблюдается в 1,5-2 раза чаще, чем ожидалось бы при случайном распределении (Рисунок 14 Д,Е). При этом для контактов, входящих в 15%-группу наиболее обогащённых (или обеднённых), доля близких или совпадающих перцентилей превышает случайное в 3-4 раза.

Таким образом, значение частот контактов сравниваемых видов оказались ближе при использовании весовой относительной модели. То есть, сохранялись не величина контакта между локусами как таковая, а её обогащённость (или обеднённость) для данного расстояния в исследуемом виде. Кроме этого, не обязательным является перекартирование всех консервативных элементов в полной мере: потеря некоторых из них или перемещение на относительно небольшое расстояние не оказывает практически никакого влияния на взаимодействие локусов.

Результаты показывают, что у сравниваемых видов в основном в ходе эволюции преимущественно сохраняются именно относительные частоты контактов, причем наиболее консервативны контакты в регионах инсуляции или, наоборот, высокой обогащённости контактов. Это согласуется с данными, что большая часть инсулированных регионов и хроматиновых петель обусловлена взаимодействием белка CTCF с белками когезинового комплекса, а многие сайты посадки белка CTCF сохраняются в ходе эволюции позвоночных.

Мера сравнения сходства/различий архитектуры хроматина

Разработанный нами алгоритм перекартирования контактов позволяет сравнивать индивидуальные контакты хроматина двух видов. Продолжением этой работы стало создание интегральной меры, которая позволила бы оценить сходство множества контактов, формируемых в определенном участке генома. В связи с

этим, для оценки различия архитектуры хроматина между сравниваемыми видами для выбранного локуса T введена мера P-BAD (percentile-based background-adjusted distance):

$$\mathrm{D(T)} = \frac{1}{N} \times -\sum_{\substack{B_i,B_j \in T \\ i \neq j}} \frac{|V_{i,j} - \tilde{V}_{i,j}|}{100} log_{10}\left(max\left(0.01{,}1 - \frac{|V_{i,j}-50|}{50}\right) \times max\left(0.01{,}1 - \frac{|\tilde{V}_{i,j}-50|}{50}\right)\right) \quad (10.1)$$

$$\mathrm{D(T)} = \frac{1}{N} \times \sum_{\substack{B_i,B_j \in T \\ i \neq j}} \left| log_{10}\left(V_{i,j}/\tilde{V}_{i,j}\right)\right| \quad (10.2)$$

где B – это бины внутри выбранного локуса T, N – число контактирующих пар бинов внутри данного локуса, $V_{i,j}$ – наблюдаемая величина контакта, $\tilde{V}_{i,j}$ – перекартированная величина контакта, полученная по формуле (3). Выражение (10.1) применяется для оценки различий между перцентилями, выражение (10.2) – для частот контактов.

Для определения закономерностей и взаимозависимости между наблюдаемыми и перекартированными контактами, полученное распределение было сравнено со случайным. Для того, чтобы сохранить структуру данных, перекартированные контакты случайным образом перемешивались и ставились в соответствии с наблюдаемыми. Полученное в ходе не менее 20 итераций распределение сравнивалось с наблюдаемым.

Использование меры P-BAD позволяет количественно оценивать сходство и различие архитектуры хроматина как для заданного локуса, так и в полногеномном исследовании. Применение данной меры показало хорошие результаты для поиска пространственно консервативных и неконсервативных регионов (Рисунок 15). Основной способ определения консервативности архитектуры хроматина – перекартирование границ доменов между видами – показал высокую чувствительность как к используемым алгоритмам, так и к их параметрам. Не редкой является ситуация, когда такой подход в силу технических причин не определяет границу в одном из видов, хотя её наличие уверенно обнаруживается при визуальном анализе (Рисунок 15 А,Б). Также данный подход не имеет

практически никакой возможности обнаружить тонкие различия в организации хроматина внутри исследуемого локуса, в отличие от применения меры P-BAD.

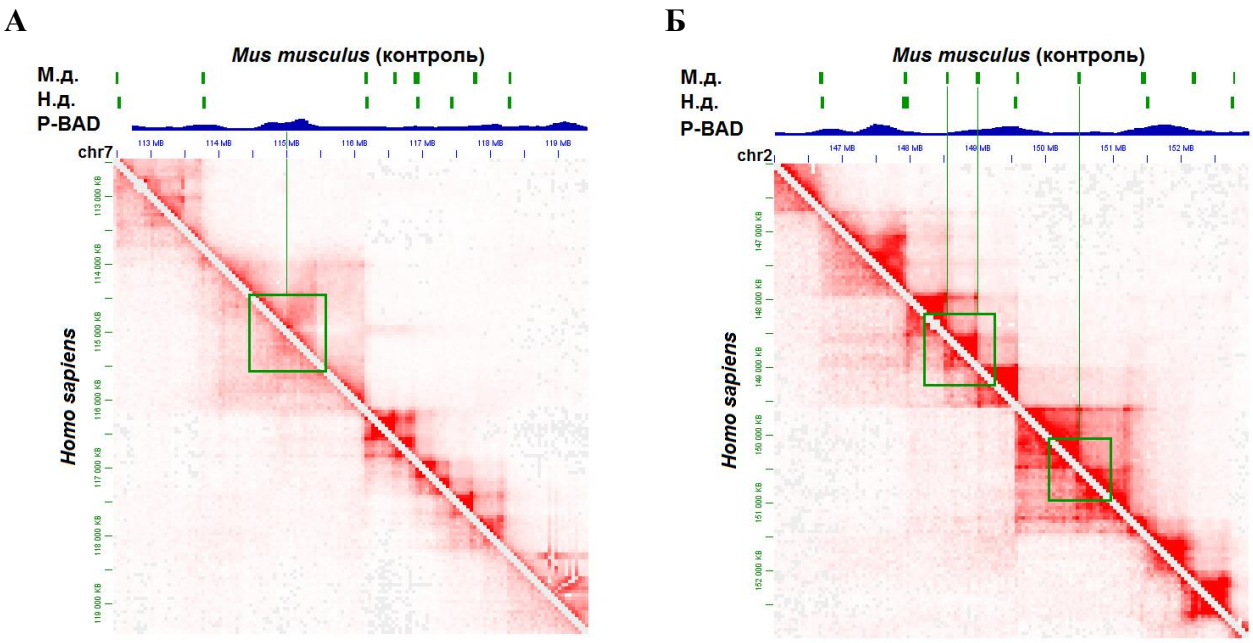

Рисунок 15. Использование меры P-BAD для сравнения архитектуры хроматинов между видами. А. Регион внутри ТАДа на хромосоме 7 *H. sapiens* (выделен квадратом) показывает различие в пространственной организации хроматина. Такие различия не могут быть обнаружены прямым сравнением границ ТАДов, но детектируются с помощью P-BAD. Б. Участок хромосомы 2 *H. sapiens*, иллюстрирующий различия в положении границ доменов (выделено квадратами). М.д. – домены, выделенные под данным Hi-C для *M. musculus*. Н.д. – домены, выделенные под данным Hi-C для *H. sapiens*. Мера P-BAD демонстрирует сходство выделенных регионов, несмотря на различия в выделение доменов у разных видов

Сравнение распределения меры P-BAD, построенной на экспериментальных данных (*H. sapiens, M. mulatta, M. musculus, C. familiaris, G. gallus*), со генерированной с помощью случайного перемешивания перекартированных контактов, подтверждает эволюционную консервативность пространственной организации хроматина у этих видов (Рисунок 16).

Таким образом, в ходе изучения эволюционной консервативности между разными видами позвоночных открывается ряд фактов. Во-первых, архитектура хроматина консервативна между даже весьма эволюционно далёкими организмами, такими как *G. gallus* и *H. sapiens*. Во-вторых, сравнение с помощью

VI наглядно демонстрирует, что используемый алгоритм выделения ТАДов вносит больший вклад в сходство или различие, чем различие между типами клеток

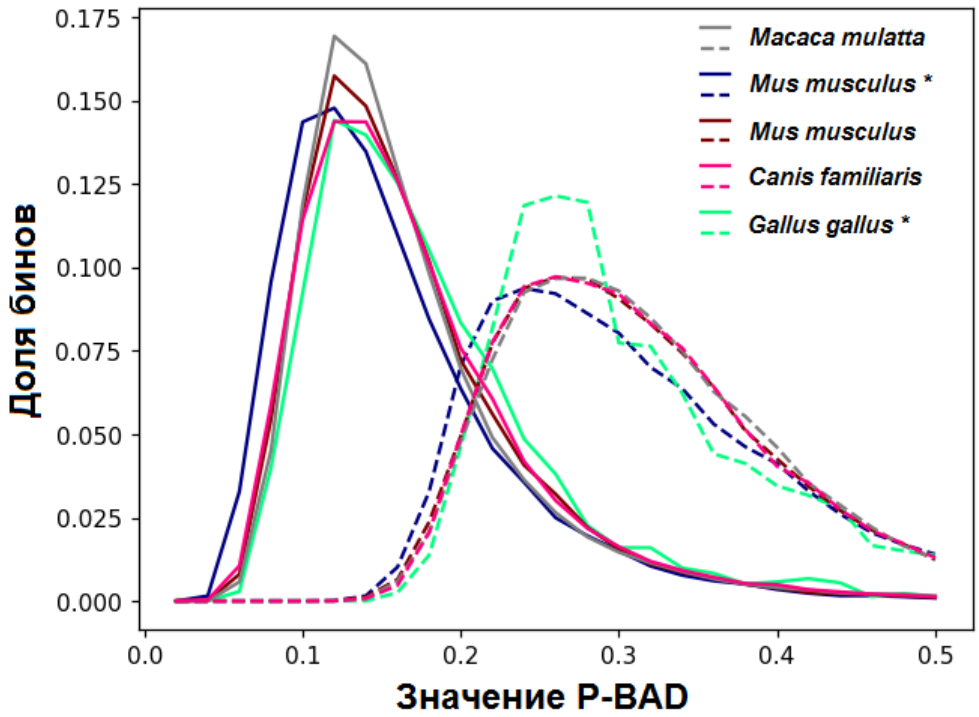

Рисунок 16. Распределение значений P-BAD в геноме, рассчитанное при сравнении архитектуры хроматина фибробластов (отмечены звёздочкой *) и клеток печени *H. sapiens* с соответствующим типом клеток у других организмов демонстрирует сходство архитектуры хроматина. Сплошная линия – наблюдаемое распределение. Прерывистая линия – распределение, полученное на основе случайного перемешивания перекартированных контактов.

В-третьих, архитектура хроматина оказывается консервативна не только на уровне крупных выделяемых тем или иным образом блоков, но и для отдельных контактов. Анализ с помощью C-InterSecture показывает, что консервативна не вероятность взаимодействия локусов как таковая, а её отношение к ожидаемой для расстояния, разделяющего данные локусы. Это противоречит гипотезе о том, что наблюдаемые величины контактов являются причиной взаимодействия регуляторных элементов.

4.4. Пространственная организация хроматина комаров рода *Anopheles*

Необходимо учитывать, что выводы об особенностях эволюции архитектуры хроматина, были получены применением алгоритма C-InterSecture к представителям позвоночных, у которых большой вклад в формировании пространственной организации хроматина вносит механизм протягивания петли. В соответствии с этим, большой интерес представляет применение разработанного алгоритма к сравнению архитектуры хроматина представителей отряда *Diptera*, у которых доминирующими механизмами является A/B-компартментализация и взаимодействие белков комплекса Polycomb.

В частности, интерес представляют результаты Hi-C экспериментов на комарах рода *Anopheles*. (Таблица 4). Эта таксономическая группа обладает рядом ценных для исследования особенностей. Во-первых, это представители отряда *Diptera*, которые, как уже отмечено выше, характеризуются особенностями пространственной организации хроматина, а именно, незначительным вкладом белка dCTCF в укладку генома. Во-вторых, обладая ненамного меньшим количеством генов, комары рода *Anopheles* имеют геном гораздо меньшего размера, что говорит о гораздо меньшем вкладе повторов и транспозонов в совокупный геном и более высокой плотности генов. В-третьих, время дивергенции включённых в исследование видов равномерно распределено в диапазоне от 2 до ~100 млн. лет (Рисунок 3) [11,196]. Последнее, например, соответствует времени дивергенции эволюционных линий *H. sapiens* и *M. musculus*.

Общая характеристика генома комаров рода *Anopheles*

Геном комаров рода *Anopheles* состоит из двух пар соматических хромосом и половых хромосом X и Y. В исследуемых нами типах клеток хромосомы были уложены в ядре в ориентацию по Раблю. От вида к виду, совокупный размер генома колеблется в достаточно широком диапазоне: от ~170 млн. п.о. до ~230 млн. п.о., что обусловлено различием в количестве повторов и качеством сборки участков гетерохроматина.

Визуальный анализ карт Hi-C подтвердил ориентацию хромосом по Раблю наличием ярко выраженных теломер-теломерных и центромер-цетромерных контактов (Рисунок 17). Также обращает на себя внимание наличие крупных блоков прицентромерного и прителомерного гетерохроматина. На хромосоме 2R *An. stephesi* обнаружена масштабная гетерозиготная инверсия, затрагивающая регион размером около 16 млн. п.о. Так как основным механизмов формирования пространственной организации хроматина у *Diptera* является фазовая сепарация и компартментализация, то более глубокий анализ архитектуры хроматина комаров рода *Anopheles* и поиск взаимосвязей между пространственной организацией хроматина и локальными особенностями генома мы решили начать с выделения A/B-компартментов.

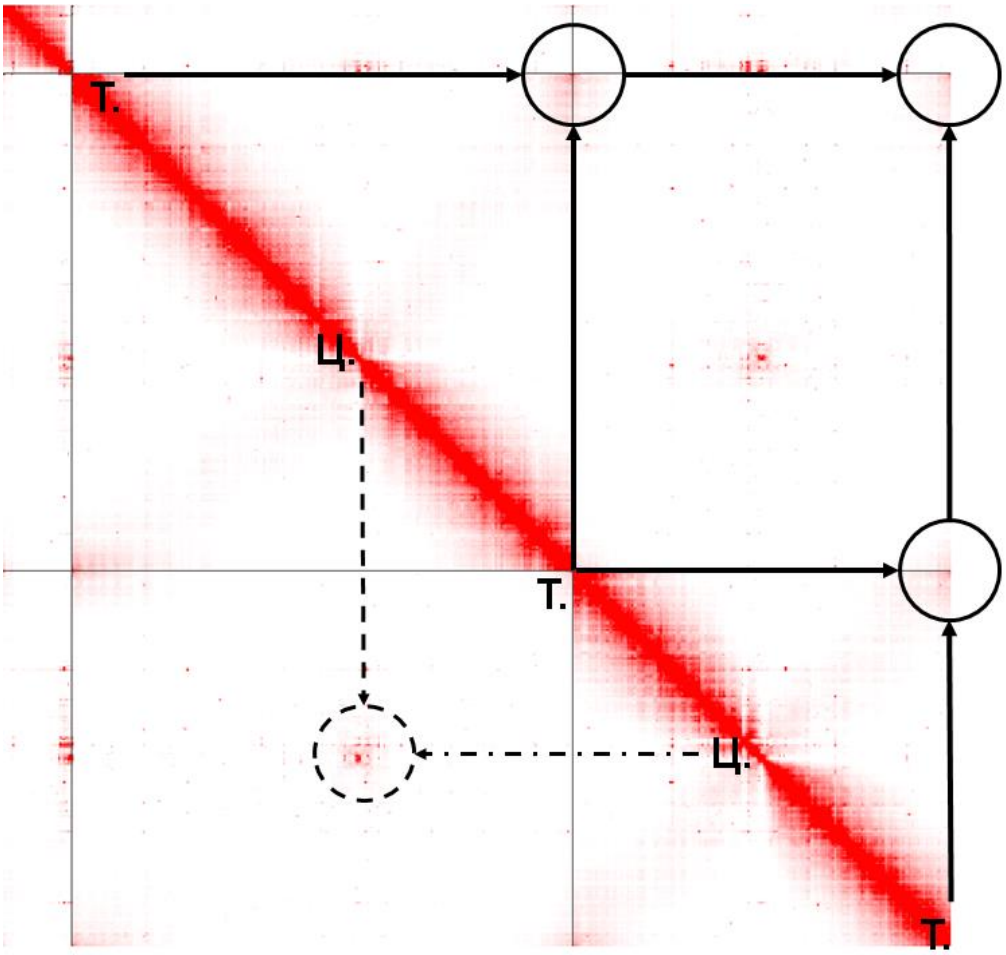

Рисунок 17. Ориентация хромосом по Рабблю на примере Hi-C карты An. atroparvus. Ц. – центромеры. Т. – теломеры. Кругами обведены центромер-центромерные и теломер-теломерные контакты.

4.5. Выделение A/B-компартментов в геномах комаров рода *Anopheles*

Одним из наиболее важных механизмов формирования пространственной архитектуры хроматина у *Diptera* является компартментализация генома. Мы попытались определить распределение геномных компартментов на основе данных Hi-C для пяти видов комаров рода *Anopheles* с помощью стандартного алгоритма, входящего в пакет программ juicertools. В соответствии с литературными данными ожидалось, что разделение генома на компартменты будет соответствовать делению на блоки с высоким и низким уровнем экспрессии, насыщенные и обеднённые генами. Для проверки этого полученные с помощью juicertools для значения первого собственного числа была рассчитана корреляция с GC составом, уровнем экспрессии и числом генов в соответствующих бинах. Результаты корреляции в ряде случаев оказались настолько низки, что это не позволяло установить соответствие между компартментализацией и активностью хроматина. Более того, визуальный анализ полученного трека компартментов показал, что он во многих случаях не соответствует «клетчатому» паттерну, наблюдаемому на картах Hi-C. По всей видимости, особенности организации хромосом у комаров рода *Anopheles* оказывают серьёзное влияние на паттерн контактов, как следствие лишь для отдельных хромосом полученные величины компартментов отражали исследуемые свойства генома (Таблица 5, столбец «Ju.»). Среди таких эффектов в первую очередь обращает на себя внимание влияние, вносимое блоками прицентромерного и прителомерного гетерохроматина, из-за чего они выделяются алгоритмами в отдельный компартмент.

В соответствии с этим было решено исключить фрагменты гетерохроматина из генома и проводить анализ аналогично тому, что проводится алгоритмом juicertools на модифицированных таким образом данных. Использование данного метода (Таблица 5, столбцы «Cr.») в ряде случаев позволило значительно улучшить соответствие величин компартментализации плотности генов и активности транскрипции. Однако данный эффект наблюдался не для всех хромосом и не всех исследуемых видов.

Таблица 5. Корреляция расчётных значений A/B-компартментов с некоторыми характеристиками генома. X. – хромосома. Ju. – значения компартментов получены алгоритмом juicertools. Cr. – то же, но без участков прителомерного и прицентромерного гетерохроматина. Fr. – то же, но с разбиением хромосом на фрагменты длиной 10-20 млн.п.о. CE – с использованием алгоритма ABCE

вид	X.	экспрессия в молекулах РНК на бин				%GC				число генов на бин			
		Ju.	Cr.	Fr.	CE	Ju.	Cr.	Fr.	CE	Juicer	Ju.	Cr.	CE
An. coluzzii	X	0,36	0,61	0,58	0,60	0,06	0,04	0,03	0,01	0,29	0,35	0,34	0,34
	2R	0,06	0,04	0,25	0,59	0,28	0,23	0,02	0,05	0,05	0,04	0,13	0,35
	2L	0,04	0,21	0,35	0,56	0,29	0,31	0,07	0,05	0,12	0,05	0,17	0,31
	3R	0,18	0,16	0,24	0,54	0,27	0,37	0,12	0,06	0,14	0,13	0,11	0,22
	3L	0,29	0,16	0,27	0,57	0,32	0,23	0,02	0,02	0,23	0,13	0,12	0,31
An. merus (embryo)	X	0,18	0,63	0,60	0,68	0,27	0,05	0,01	0,02	0,27	0,34	0,32	0,38
	2R	0,43	0,55	0,55	0,54	0,21	0,23	0,05	0,06	0,39	0,29	0,32	0,34
	2L	0,30	0,49	0,48	0,50	0,56	0,04	0,04	0,05	0,27	0,31	0,31	0,31
	3R	0,23	0,50	0,47	0,48	0,36	0,28	0,12	0,06	0,19	0,22	0,17	0,22
	3L	0,27	0,48	0,36	0,49	0,46	0,06	0,01	0,07	0,19	0,27	0,18	0,30
An. merus (adults)	X	0,71	0,71	0,69	0,66	0,10	0,09	0,02	0,02	0,57	0,56	0,53	0,49
	2R	0,62	0,61	0,60	0,58	0,01	0,04	0,03	0,02	0,61	0,59	0,57	0,53
	2L	0,67	0,01	0,15	0,50	0,18	0,12	0,02	0,06	0,61	0,03	0,08	0,47
	3R	0,66	0,66	0,58	0,58	0,13	0,11	0,07	0,05	0,54	0,53	0,45	0,46
	3L	0,60	0,04	0,08	0,53	0,09	0,01	0,03	0,04	0,58	0,06	0,02	0,50
An. stephensi	X	0,03	0,73	0,70	0,69	0,04	0,46	0,45	0,42	0,07	0,52	0,52	0,53
	2R	0,05	0,12	0,10	0,66	0,38	0,10	0,02	0,40	0,02	0,05	0,13	0,49
	2L	0,10	0,01	0,03	0,53	0,09	0,11	0,01	0,41	0,13	0,06	0,07	0,38
	3R	0,61	0,63	0,55	0,62	0,49	0,46	0,33	0,36	0,39	0,37	0,31	0,39
	3L	0,48	0,63	0,54	0,60	0,60	0,54	0,36	0,39	0,37	0,45	0,40	0,43
An. atroparvus	X	0,01	0,54	0,65	0,68	0,23	0,01	0,11	0,14	0,01	0,32	0,41	0,47
	2R	0,13	0,34	0,23	0,49	0,34	0,18	0,10	0,10	0,08	0,16	0,09	0,33
	2L	0,61	0,34	0,37	0,60	0,30	0,05	0,11	0,19	0,49	0,25	0,29	0,39
	3R	0,02	0,12	0,30	0,54	0,21	0,12	0,27	0,20	0,12	0,04	0,25	0,42
	3L	0,13	0,48	0,43	0,48	0,13	0,23	0,20	0,22	0,14	0,34	0,31	0,37
An. albimanus	X	0,60	0,71	0,69	0,67	0,23	0,21	0,19	0,18	0,52	0,59	0,58	0,59
	2R	0,01	0,11	0,23	0,64	0,22	0,23	0,05	0,04	0,01	0,10	0,15	0,50
	2L	0,61	0,58	0,34	0,50	0,04	0,01	0,02	0,02	0,45	0,40	0,24	0,37
	3R	0,24	0,28	0,32	0,53	0,05	0,02	0,02	0,02	0,22	0,18	0,22	0,34
	3L	0,65	0,34	0,41	0,62	0,13	0,54	0,28	0,06	0,55	0,34	0,26	0,44

Далее обращает на себя внимание тот факт, что результаты вычисления A/B-компартментов по данным Hi-C взрослых особей *An. merus* показывают гораздо лучшее визуальное соответствие блокам активно и неактивного хроматина, чем вычисления по данным из эмбрионов. Было предположено, что такие различия могут быть связаны с тем, что в эмбрионах комаров хромосомы находятся преимущественно в ориентации по Раблю. В следствие этого, теломеры и центромеры жёстко разделены в пространстве ядра, что приводит к уменьшению числа контактов между удалёнными участками генома, что, вероятно, влияет на вычисление компартментов.

Чтобы проверить это предположение, было решено в точности повторить процедуру обработки данных для расчёта компартментов, используемую алгоритмами jucertools, но на последнем шаге помимо вычисления первой главной компоненты дополнительно вычислить значение ещё нескольких главных компонент.

Результаты показывают, что первые две главных компоненты отражают расстояние произвольного геномного локуса до центромеры и теломеры хромосомы (Рисунок 18). Таким образом, в результате влияния ориентации по Раблю на архитектуру хроматина, контакты между локусами, удалёнными на большие геномные расстояния, оказываются значительно ослаблены, а контакты между теломерой и центромерой практически исключены. В соответствии с этим, главной особенностью произвольного локуса, которые алгоритмы вычленяют из данных Hi-C, является положение геномного локуса на хромосоме и связанное с этим предпочтение в контактах с теломерой или центромерой, и деление на компартменты определяется в свою очередь близостью к теломере или центромере.

Следовательно, чтобы ослабить влияние ориентации по Раблю на вычисление компартментов, нужно все вычисления проводить отдельно для ближайшего окружения центромеры, отдельно для ближайшего окружения теломеры и отдельно для промежуточных участков плеч хромосом.

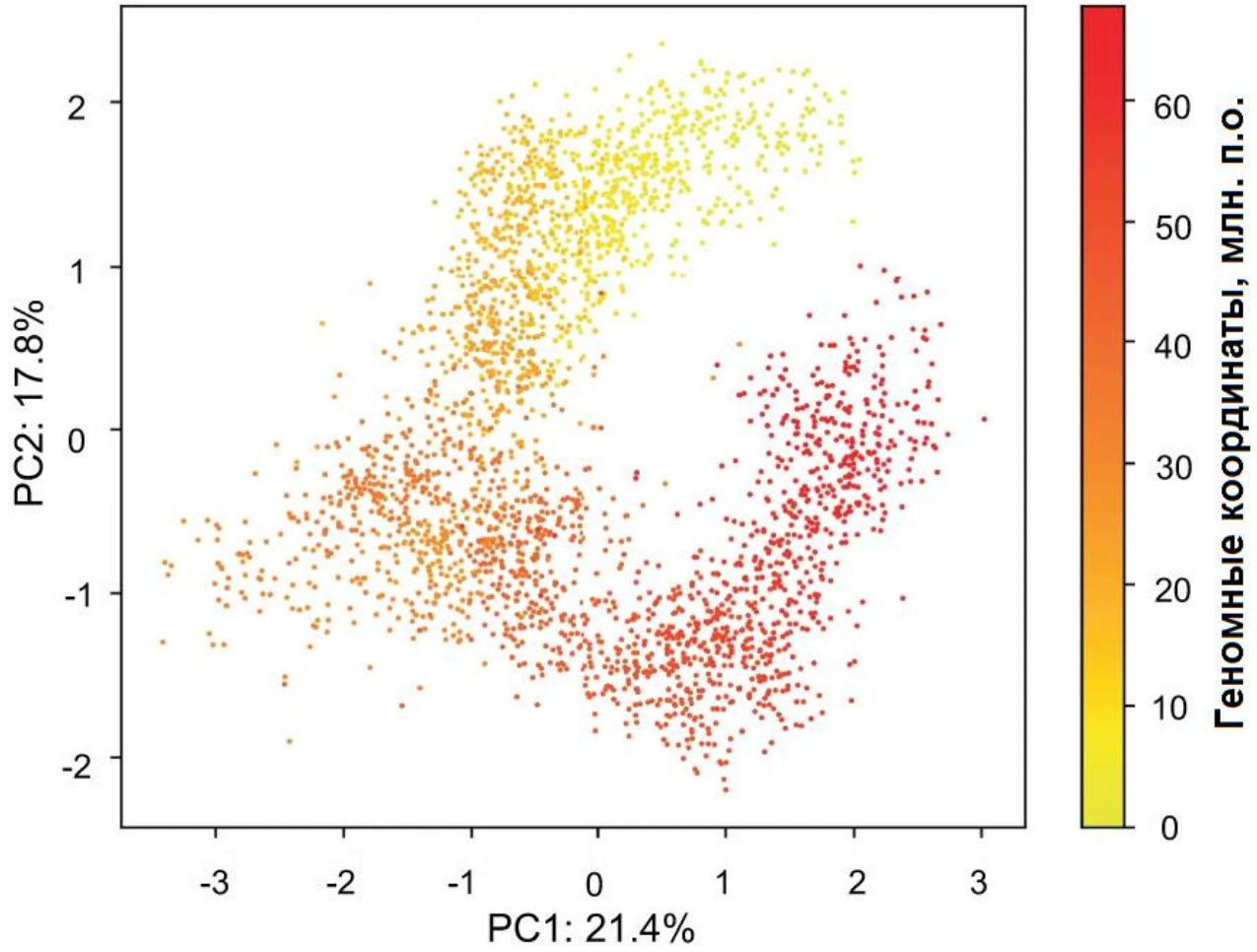

Рисунок 18. Зависимость первой и второй главной компоненты матрицы контактов от геномного расстояния между локусами на примере хромосомы 3 *An. albimanus*. Цветом указана геномная координата бина.

Исходя из характерного размера хромосомных плечей в 30-40 млн. п.о., было решено механистически разбивать геном на участки по 10 млн. п.о. и подсчитывать компартменты для каждого такого участка отдельно. В соответствии с этим, алгоритм обработки данных выглядит следующим образом:

- удаляются участки прителомерного и прицентромерного хроматина,

- матрицы контактов разбиваются на подматрицы, соответствующие участкам генома протяжённостью около 10 млн. п.о.

- каждая подматрица обрабатывается согласно алгоритму juicertools до стадии, предшествующей расчёту значений собственного вектора;

- на обработанных подматрицах рассчитываются значения первой главной компоненты, которая и принимается как величина компартментализации для соответствующего геномного локуса.

Применение данного метода позволило в ряде случае улучшить соответствие между A/B-компартментами и активностью хроматина, однако, для ряда хромосом результаты всё равно получились не удовлетворительные (Таблица 5, столбец «Fr.»).

После визуального сопоставления полученных данных о компартментах и карт контактов Hi-C было выдвинуто предположение, что на способность алгоритмов обнаруживать различие между компартментами сказываются:

- более слабая компартментализация тех или иных локусов;

- высокая доля нулевых контактов для локусов, разделённых большим геномным расстоянием, вызванная недостаточной глубиной секвенирования библиотек Hi-C и влиянием ориентации хромосом по Раблю.

На основе этого было выдвинуто предположение, что данные эффекты можно нивелировать, если:

- значения для удалённых контактов пересчитать так, как если бы матрица имела меньшее разрешение, и за счёт этого заполнить нулевые значения и сгладить отдельные выбросы;

- увеличить контраст между значениями в полученных матрицах, нормализуя их на величину разброса между контактами, формируемыми соседними бинами.

В соответствии с этим было разработан нижеописанный алгоритм ABCE (**A**/**B**-compartment **c**ontrast **e**nhancement) для выделения A/B-компартментов.

Пусть M – матрица размером NxN, содержащая значения отношений наблюдаемой частоты контакта к ожидаемому (среднему) на данном геномном расстоянии, $m_{i,j}$ – элемент в i-ой строке и j-ом столбце. Положим f-окрестность радиуса f элемента $m_{i,j,}$ как подмножество F из элементов матрицы M, таких что:

$$F(m_{i,j}, f) = \{m_{i-f,j-f}, \dots, m_{i+f,j+f}\} \quad (11)$$

Определим для произвольной матрицы M операцию сглаживания, $\widetilde{M} = S(M, f)$, так что для каждого элемента $\widetilde{m}_{i,j}$ верно:

$$\widetilde{m}_{i,j} = median\left(F\left(m_{i,j}, f\right)\right) \tag{12}$$

Операция нормализации, $\acute{M} = N(M)$, определяется так что для каждого элемента $\acute{m}_{i,j}$ верно:

$$\acute{m}_{i,j} = m_{i,j} - \frac{\sum_{k=0}^{n} m_{i,k}}{n} \tag{13}$$

Контрастирование, $\check{M} = D(M)$, определяется так, что для каждого элемента $\check{m}_{i,j}$ верно:

$$C = \frac{1}{2} \times P\left(F\left(m_{i,j}, f\right), 75\right) + P\left(F\left(m_{i,j}, f\right), 25\right) \tag{14.1}$$

$$R = \frac{1}{2} \times P\left(F\left(m_{i,j}, f\right), 75\right) - P\left(F\left(m_{i,j}, f\right), 25\right) \tag{14.2}$$

$$\check{m}_{i,j} = \frac{m_{i,j} - C}{R} \tag{14.3}$$

где P(X,q) – q-ый перцентиль в числовом множестве X.

Для вычисления A/B-компартментов используется матрица A = D(N(S(M,f))), где M – матрица отношения наблюдаемого числа контактов к ожидаемому, описанная выше. На последнем этапе вычисляется значение первой главной компоненты для подматриц полученной матрицы:

$$\begin{pmatrix} a_{i,i} & \cdots & a_{i,i+f} \\ \vdots & \ddots & \vdots \\ a_{i+f,i} & \cdots & a_{i+f,i+f} \end{pmatrix}$$

Таким образом для каждого локуса значение компартмента определялось как знак первой главной компоненты.

Высокие значения корреляции значений компартмента, полученных с использование алгоритма ABCE, с плотностью генов и данными РНК-секвенирования (Таблица 5, столбец «CE») подтвердили эффективность работы алгоритма. Несмотря на то, что в отдельных случаях результаты корреляции значений компартментализации, рассчитанных с помощью ABCE, с транскрипционной активностью несколько уступают результатам, полученным при использовании других методов, использование метода ABCE даёт наиболее стабильный и единообразный результат, что позволяет более корректно анализировать пространственную организацию хроматина.

4.6. Характеристика пространственной организации хроматина у комаров рода *Anopheles*

Полученные значения компартмента были использованы для описания пространственной организации хроматина у комаров рода *Anopheles* и её связи с известными генетическими и эпигенетическими характеристиками.

В первую очередь, была проведена оценка силы компартментализации (Рисунок 19). Отдельно были проведены расчёты силы компартментализации для всех контактов и контактов бинов, разделённых геномным расстоянием не более 10 млн. п.о. Последняя величина была выбрана на основании параметров алгоритма ABCE, использованного для выделения компартментов.

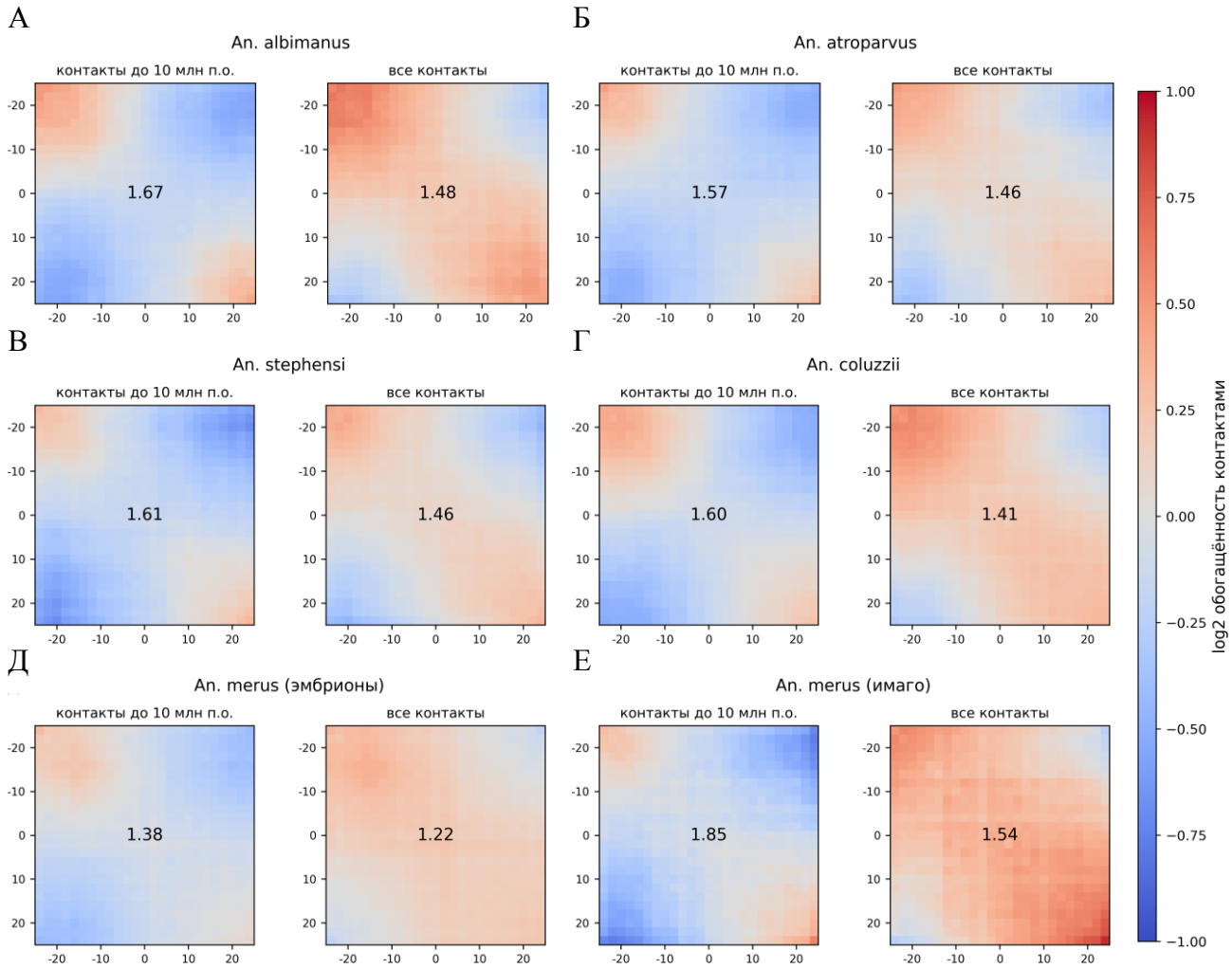

Рисунок 19. Компартментализации хроматина в разных видах комаров рода *Anopheles*. По осям X и Y – значение компартментов пары контактирующих бинов. Рассчитанная величина силы компартментализации показано по центру графиков

Результаты сравнения рассчитанных величин показывают, что добавление в выборку контактов, разделённых большим геномным расстоянием, ослабляет рассчитанную величину силы компартментализации (Рисунок 19). Кроме этого, компартментализация хроматина имаго комаров *An. merus* существенно выше, чем хроматина эмбрионов. Данные факты косвенно подтверждает влияние ориентации хромосом по Раблю на формирование контактов между компартментами.

Выраженная компартментализация хроматина позволяет предположить, что алгоритмически выделяемые домены у комаров рода *Anopheles* задаются делением генома на компартменты. Тем не менее, анализ значений компартментов вблизи границ доменов показывает значительное обогащение границ доменов А-компартментом (Рисунок 20).

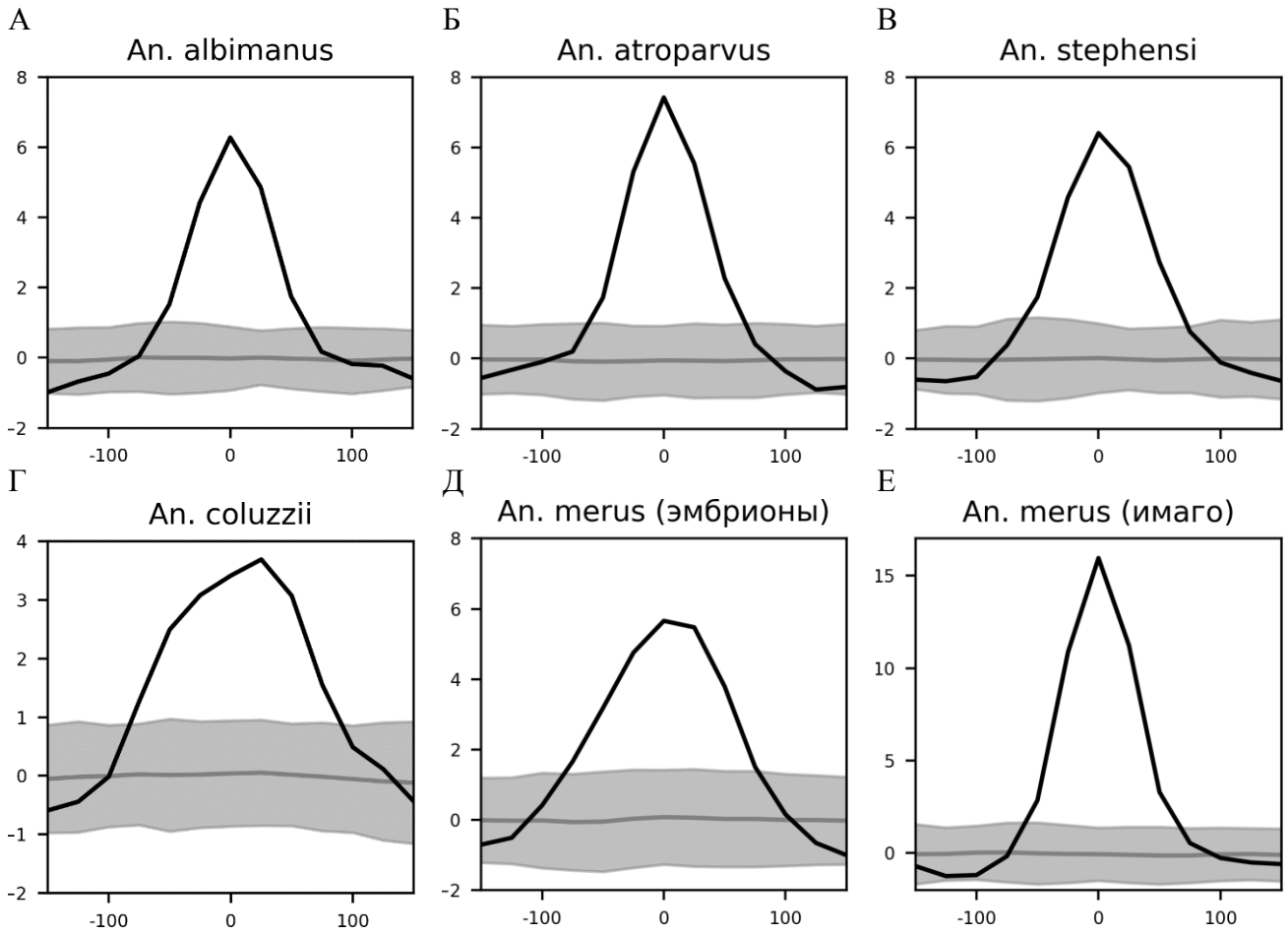

Рисунок 20. Распределение значений компартмента, рассчитанных алгоритмом ABCE, относительно границ доменов. По оси X расстояние в тысячах пар оснований от границы, по оси Y — среднее число сайтов связывания CTCF на 40 тысяч пар оснований. Чёрная линия – наблюдаемые данные. Серая область – 3 стандартных отклонения от ожидаемого.

Учитывая возможность алгоритмических артефактов, а также то, что разрешение в 25 тыс. п.о., на котором были проанализированы карты Hi-C, не позволяет надёжно детектировать небольшие домены, были отдельно проанализированы границы с выраженной инсуляцией (величина инсуляции меньше 1.5), что составляет около 15-20% всех границ доменов. Для оценки влияния изменения значения компартмента на формирование границы домена, каждой границе домена было сопоставлено два числа – среднее значение компартмента в трёх ближайших к границе бинах (125 тысяч п.о.) внутри и снаружи домена (Рисунок 21).

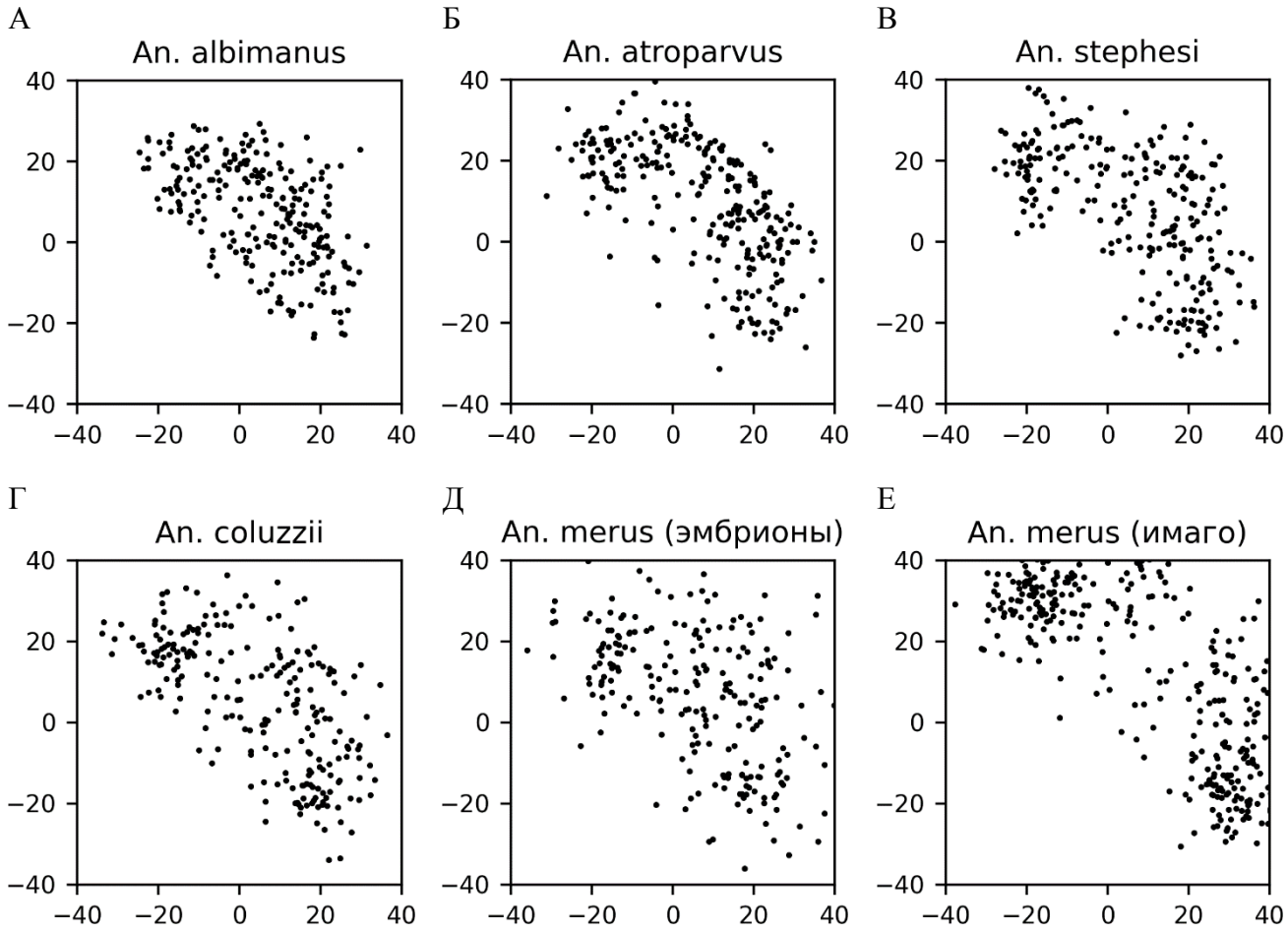

Рисунок 21. Среднее значение компартмента до и после границы домена. По оси X отложено значение компартмента внутри домена, по оси Y - снаружи. Величины компартментов рассчитанны алгоритмом ABCE.

Полученные результаты показывают, что такого же строгого соответствия между локализацией границы и сменой компартмента, как это наблюдалось для

клеток эритроидного ряда у *G. gallus*, не наблюдается, (Рисунок 21 A-E). Исключением составляет имаго *An. merus* (Рисунок 21E), у которого соответствие между границами доменов и границами компартментов более строгое, что согласуется с более сильной компартментализацией в этом типе клеток.

Для изучения принципов формирования доменов в хроматине комаров рода *Anopheles*, была исследована связь между встречаемостью повторов различных классов и расстоянием до границы домена (Рисунок 22 и Приложение 1).

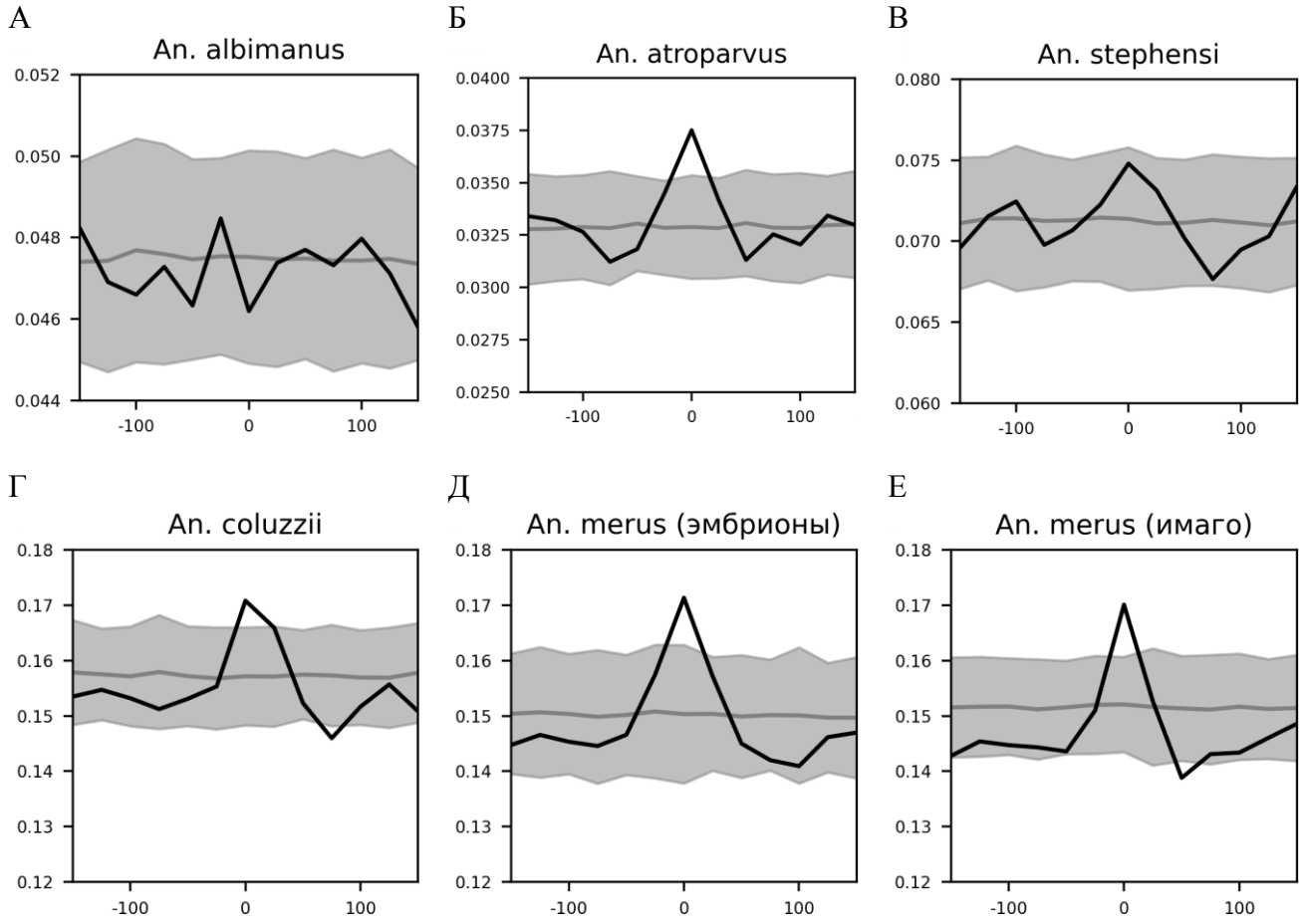

Рисунок 22. Распределение повторов всех классов относительно границ доменов для разных видов комаров рода *Anopheles.* Чёрная линия – наблюдаемые значения. Серая линия – среднее значения. Серая область – три стандартных отклонения от среднего.

Значимое превышение над ожидаемым при случайном расположении границ доменов было обнаружено у *An. atroparvus*, *An. coluzzii* и *An. merus.* У указанных видов в границах доменов наибольший вклад в обогащение дают ДНК-транспозоны (Приложение 1, Рисунок 2) и простые повторы (Приложение 1, Рисунок 1). У *An. merus* значительный вклад вносят также SINE и LINE элементы.

(Приложение 1, Рисунок 5). Интересным моментом является обеднение простыми повторами границ доменов у *An. atroparvus* (Приложение 1, Рисунок 1).

Данное наблюдение интересно в контексте того, что для млекопитающих известна связь между границами ТАДов и разными группами мобильных элементов генома, в тоже время у *D. melanogaster* такой связи не отмечено. Наблюдаемые различия не позволяют утверждать, что какой-то тип повторов, в силу каких-бы то не было причин, является одним из факторов формирования границы доменов в комарах рода *Anopheles*, а сама зависимость между повторами и границами доменов является разнонаправленной в разных филогенетических линиях. Наблюдаемые различия между разными комарами рода *Anopheles* позволяют предположить, что оно может быть связано с элиминацией повторов в ходе эволюции.

Необходимо указать, что контрольная выборка доменов, созданная при помощи случайных перестановок границ, не учитывала обогащение границ доменов в А-компартменте. Таким образом, необходимо проверить, не является ли обнаруженная взаимосвязь следствием большей частоты встречаемости повторов в А-компартменте. Для проверки этого предположения, все бины были поделены на 6 групп, в зависимости от значения компартмента: меньше -15, от -15 до -5, от -5 до 5, от 5 до 15, больше 15 и бины, в которых значение компартмента не было определено.

Результаты сравнения величины компартмента с частотой встречаемости повторов показывают, что зависимости между принадлежностью бина к тому или иному компартменту и долей повторов разных классов в нём нет (Рисунок 23). Исключение составляют прителомерные и прицентромерные регионы, которые были исключены при расчёте величины компартмента алгоритмом ABCE.

Изучение особенностей этих регионов показывает, что они отличаются в 5-6 раз более высокой долей повторов, в среднем, по сравнению с прочими локусами генома. Однако эти регионы так же были исключены из рассмотрения при анализе границ доменов.

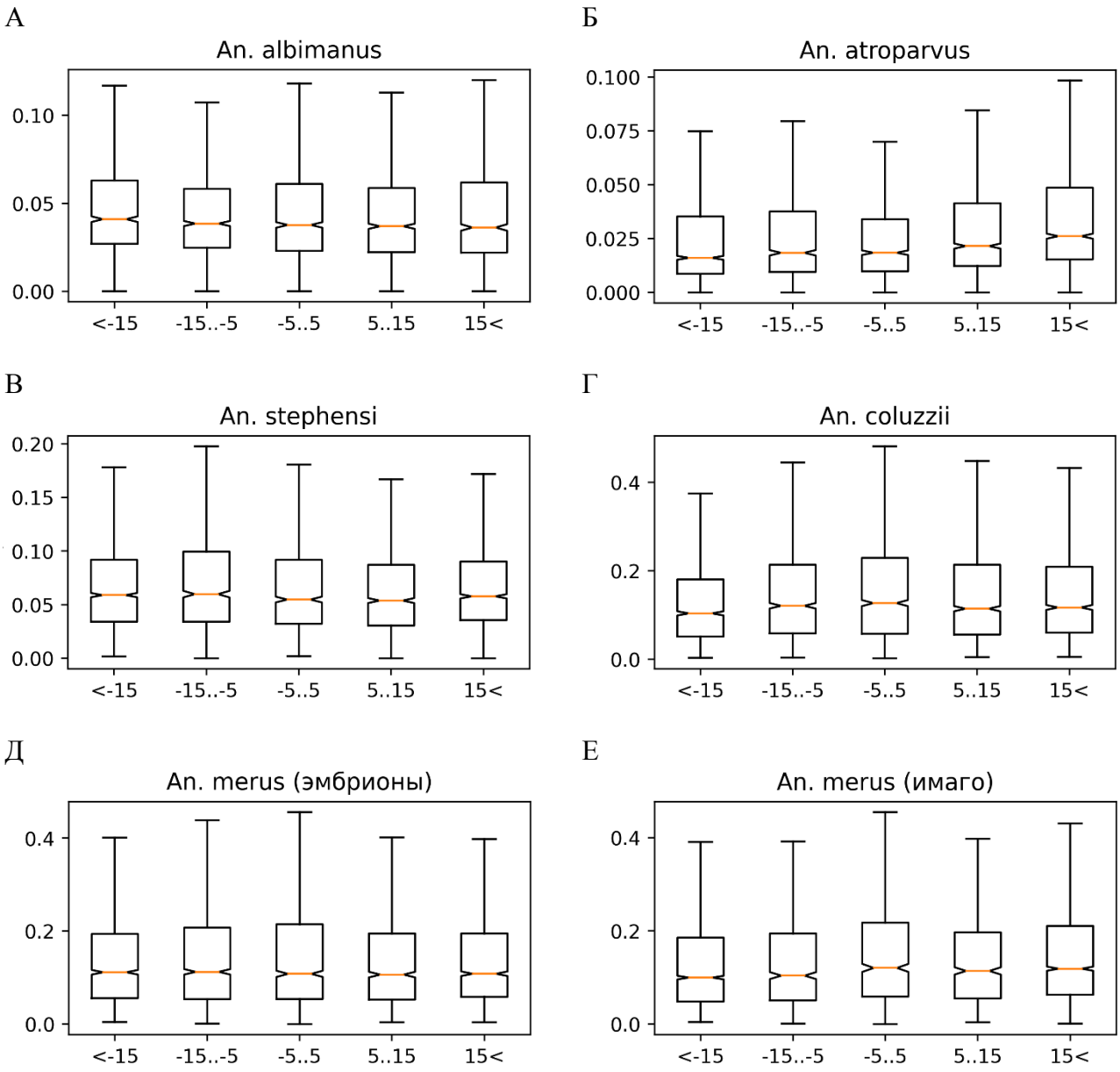

Рисунок 23. Доля повторов всех классов в локусах, в зависимости от величины компартмента. По оси X – значение компартмента. По оси Y – доля повторов на бин.

Большие эволюционные расстояния, разделяющие рассматриваемые виды комаров, различия в частоте встречаемости разных классов повторов и их связи с границами доменов ставят интересный вопрос о том, насколько сильно межвидовые различия сказались на консервативности архитектуры хроматина комаров рода *Anopheles.*

4.7. Исследование консервативности архитектуры хроматина комаров рода *Anopheles*

Для сравнения пространственной организации хроматина у исследованных видов комаров была использовано ПО C-InterSecture.

Сравнение пространственной организации хроматина на уровне отдельных контактов показало высокую консервативность архитектуры в сохранившихся блоках синтении (Рисунок 24). Подавляющая часть наблюдаемых различий обнаруживается в районах эволюционных перестроек хромосом. В свою очередь, обнаруженные внутри синтенных блоков различия единичны и сомнительны.

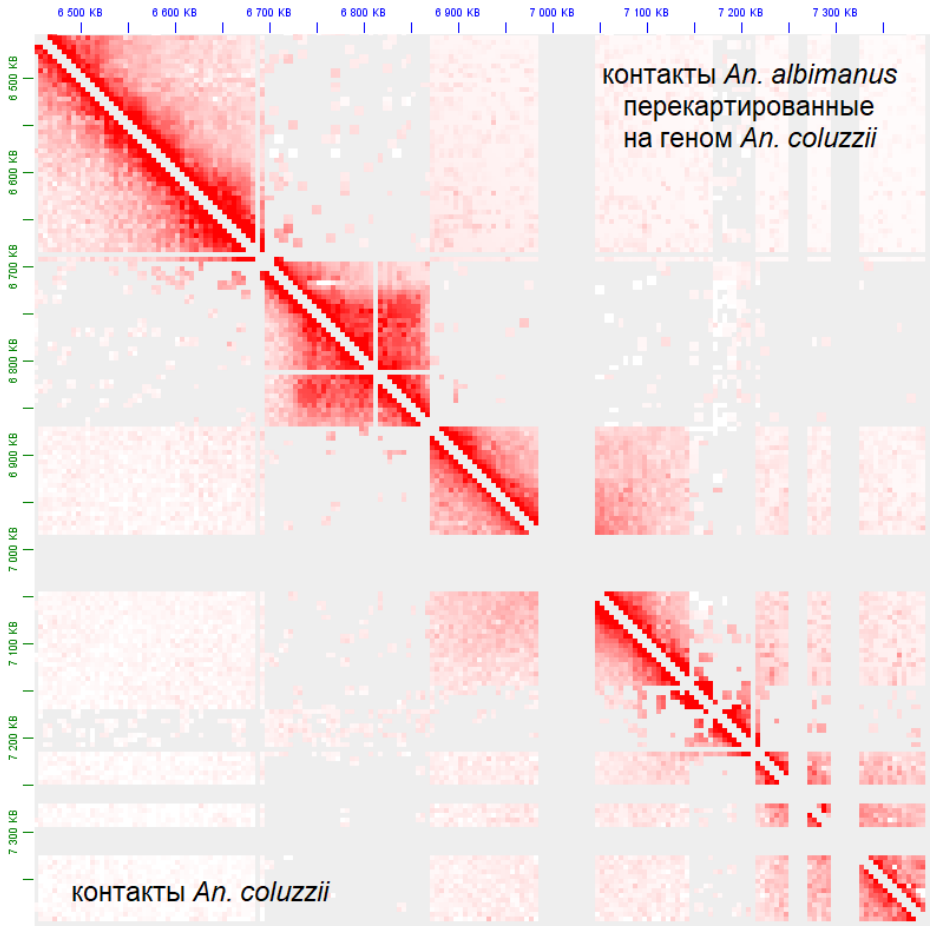

Рисунок 24. В пределах синтенных блоков прослеживается высокий уровень консервативности пространственной организации хроматина на уровне отдельных контактов. Показан участок 2R хромосомы *An. coluzzii* между 6,4 млн. п.о. и 7,4 млн. п.о. и синтенные ему участки генома *An. albimanus*

Коэффициент корреляции Пирсона между контактами, наблюдаемыми и перекартированными, согласно относительной весовой модели консервативности даже между самыми далёкими видами составляет 0,95 ($p << 0,05$ Рисунок 25А). При использовании абсолютной весовой модели консервативности коэффициент корреляции контактов оказывается в диапазоне от 0,8-0,9 (Рисунок 25Б).

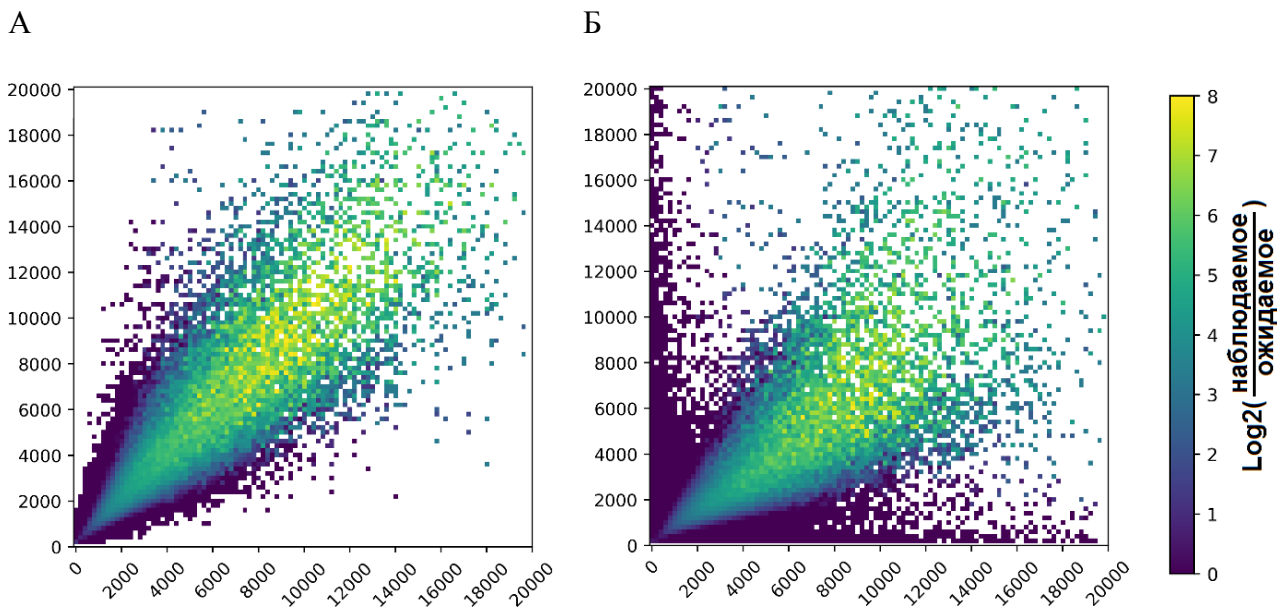

Рисунок 25. Консервативность контактов комаров рода *Anopheles*. По оси X указаны наблюдаемые контакты *An. coluzzii*, по оси Y контакты перекартированые из *An. stephensi* на геном *An. coluzzii*. А. Величины перекартированных контактов приведены согласно относительной весовой модели. Б. Величины перекартированных контактов приведены согласно абсолютной весовой модели.

По всей видимости, это связано с тем, что сколько-нибудь протяжённые участки синтении практически не изменяются по своей длине при сравнении между видами; в сильно же перестроенных участках построить достоверное сопоставление контактов не представляется возможным, а значит они не оказывают влияния на полученный результат. Чтобы это проверить, было проведено выделение блоков синтении и сравнение их размеров.

Для этого, были получены .net-файлы так, как описано в разделе 3.2. данной диссертации. Из .net-файлов были извлечены короткие участки синтении, за исключением участков длиною менее 500 нуклеотидов были. На следующем шаге проводилось объединение полученных коротких участков синтении в длинные

синтенные блоки по ниже описанным принципам. Пусть мы сравниваем виды А и Б, тогда А1 и А2 – участки в виде А, синтенные участкам Б1 и Б2 в виде Б соответственно. Эти участки будут объединяться в один синтенный блок, если:

- расстояние между участками А1 и А2 меньше 150 тысяч п.о.;
- расстояние между участками Б1 и Б2 меньше 150 тысяч п.о.;
- разница между вышеописанными расстояниями меньше 100 тысяч п.о.;
- геномное направление, в котором участок А1 выравнивается на участок Б1 совпадает с геномным направления выравнивания А2 и Б2.

Далее, из полученных синтенных блоков были исключены фрагменты длиной менее 15 тысяч п.о., так как данная величина меньше, чем размер бина, на котором проводился анализ карт Hi-C.

Сравнения показывают, что в среднем величина блока синтении меняется на 10-20% (Таблица 6), что соответствует изменениям в частоте контактов для сравниваемых с помощью C-InterSecture карт контактов.

Таблица 6. Отношение длин синтенных блоков в сравниваемых видах.

	An. albimanus	*An. atroparvus*	*An. stephensi*	*An. merus*	*An. coluzzii*
An. albimanus	-	0,87	0,89	0,83	0,83
An. atroparvus	1,15	-	1,05	0,96	0,96
An. stephensi	1,12	0,96	-	0,9	0,87
An. merus	1,21	1,06	1,14	-	1,08
An. coluzzii	1,21	1,06	1,14	1,01	-

Консервативность контактов внутри блоков синтении у комаров рода *Anopheles*, позволяет предположить, что, аналогично млекопитающим, пространственная организация хроматина тесно связана с генной регуляцией, существуют блоки совместно регулируемых генов, разрушение которые чаще всего не поддерживается отбором. В таком случае ожидается, что эволюционные точки разрыва хромосом должны проходить в неких «безопасных» локусах генома. В соответствии с этим было решено сравнить, насколько свойства таких регионов сходны у представителей позвоночных и двукрылых

4.8. Изучение генетических особенностей эволюционных точек разрывов хромосом у представителей рода *Anopheles*

Ввиду консервативности пространственной организации хроматина в пределах блоков синтении определённый интерес представляет то, какими свойствами обладают регионы, по которым всё-таки происходят эволюционные перестройки. Однако, при исследовании свойств таких регионов необходимо учитывать, что помимо перестроек происходили и другие эволюционные события, изменявшие геномы у исследуемых видов комаров, например, приобретение и потеря геномных повторов, однонуклеотидные замены и т.д. Таким образом, можно ожидать, что представляющие интерес генетические характеристики будут менее выражены для наиболее древних перестроек и более выражены в более молодых. В свою очередь, время появления и закрепления эволюционных перестроек хромосом можно оценить исходя из того, на каких филогенетических ветвях комаров рода *Anopheles* они встречаются.

Оценка времени происхождения эволюционных перестроек хромосом происходила за счёт сравнения геномных координат разрывов блоков синтении с известной по литературным данным филогенией комаров рода *Anopheles* пользуясь принципом наибольшей парсимонии. Пусть есть виды A1, A2 и A3 и вид A3 является базальным по отношению к остальным. Используя попарные выравнивания A1-A2 и A1-A3 определяются по геному вида A1 координаты разрывов блоков синтении. Если в определённом локусе находится разрыв блоков синтении видов A1-A2, но в ближайшей окрестности (не более 30 тыс. п.о.) нет разрыва блока синтении при сравнении видов A1-A3, считается, что это уникальная для вида A2 перестройка в геномных координатах вида A1. Аналогично определяются перестройки уникальные для остальных видов. Допустим, в локусе, протяжённостью не более 30 тысяч п.о., наблюдаются разрывы блоков синтении и в сравнении A1-A2 и в сравнении A1-A3. В таком случае необходимо проверить, есть ли разрыв блока синтении между видами A2 и A3. Если такого разрыва нет, то считается, что в данном регионе произошла уникальная для вида A1 перестройка.

Если такой разрыв обнаруживается, то считается, что данный регион был использован повторно при формировании перестройки в разных видах.

Используя этот метод, для каждого вида получилось девять наборов из разрывов блоков синтении, определённые как специфичные для того или иного вида, произошедшие у гипотетического предка той или иной группы видов и использованные повторно (Рисунок 26).

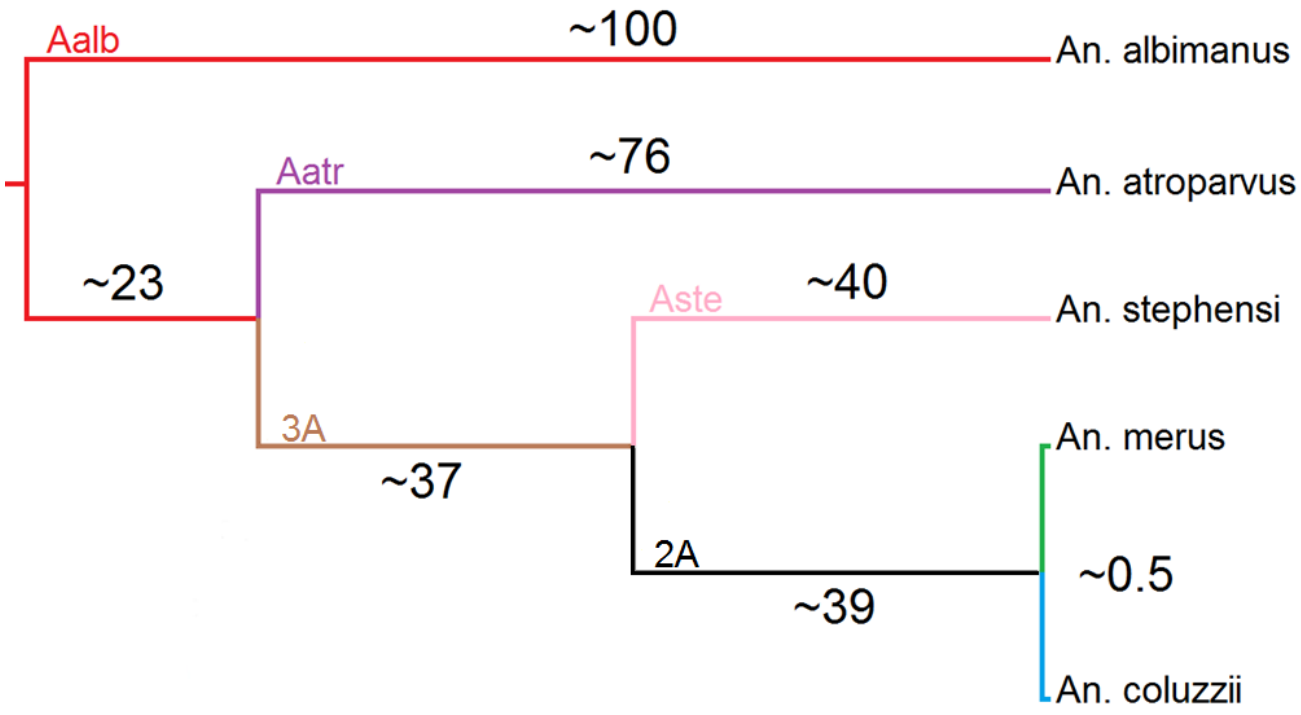

Рисунок 26. Классификация эволюционных точек разрыва по происхождению. Филогенетическое древо и время происхождения видов даны согласно [11].

В виду малого эволюционного расстояния между видами *An. merus* и *An. coluzzii* количество перестроек, наблюдаемых между ними было слишком мало для анализа и уникальные для этих видов эволюционные точки разрыва хромосом не выделялись. Вместо этого использовались эволюционные перестройки, произошедшие у гипотетического предка этих двух видов. Необходимо отметить, что в эволюционные точки разрыва, уникальные для *An. albimanus* попадают также и те, которые могли произойти у последнего общего предка *An. atroparvus* и *An. coluzzii*. И на данном наборе видов разделить такие точки разрыва невозможно.

Однако, исходя из того, что большая часть перестроек является событием эволюционно нейтральным, можно предполагать, что большая часть точек разрыва, выделенная для *An. albimanus* произошла именно в его эволюционной линии.

Сравнение взаимного расположения эволюционных точек разрыва хромосом с такими характеристиками генома как доля повторов разных классов и состояние компартмента (Приложение 2) показало, что точки разрывов синтенных блоков не являются сколько-нибудь значительно обогащены теми или иными классами повторов по сравнению со случайными участками генома. Не было обнаружено и никакой взаимосвязи с происхождением точки разрыва (Приложение 2). Единственная закономерность, которая была выявлена это то, что разрывы блоков синтении чаще происходят по регионам, принадлежащих А-компартменту и границам доменов (Приложение 2 Рисунок 1).

Однако, как было показано ранее, границы доменов чаще проходят в районах А-компартмента, таким образом возникает вопрос, с какой именно особенностью архитектуры хроматина ассоциированы точки разрыва? С А-компартментом, и тогда совпадение границ доменов с эволюционными точками разрыва является следствием взаимосвязи компартмента и границы доменов, или наоборот?

Для ответа на этот вопрос, границы доменов были классифицированы по тому, какое среднее значение компартмента наблюдается в их окрестностях (+/- 1 бин от границы). Были выделены следующие классы: значение компартмента меньше -15, от -15 до -5, от -5 до 5, от 5 до 15 и больше 15. Предполагается, что если формирование точек разрыва хромосом не связано с границами доменов, то и обогащение границ доменов точками разрывов не будет зависеть от того, к какому классу граница будет отнесена.

Результаты показывают, что существует строгая зависимость между тем, какое значение компартмента наблюдается в ближайших окрестностях (Рисунок 27) и тем, на сколько сильно граница обогащена эволюционными точками разрыва хромосом. Таким образом, можно считать, что в ходе эволюции разрывы хромосом происходят преимущественно по А-компартменту, регионам, насыщенными генами и обладающими высокой экспрессией.

Необходимо отметить, что существование точки разрыва синтении означает не только, что соседние в предполагаемом предковом геноме регионы были удалены друг от друга, но и то, что участки, ранее удалённые, оказались соединены (Рисунок 28А). Таким образом возникает следующий вопрос: существует ли какая-либо избирательность при соединении в ходе эволюционной перестройки ранее удалённых друг от друга на большое расстояние регионов?

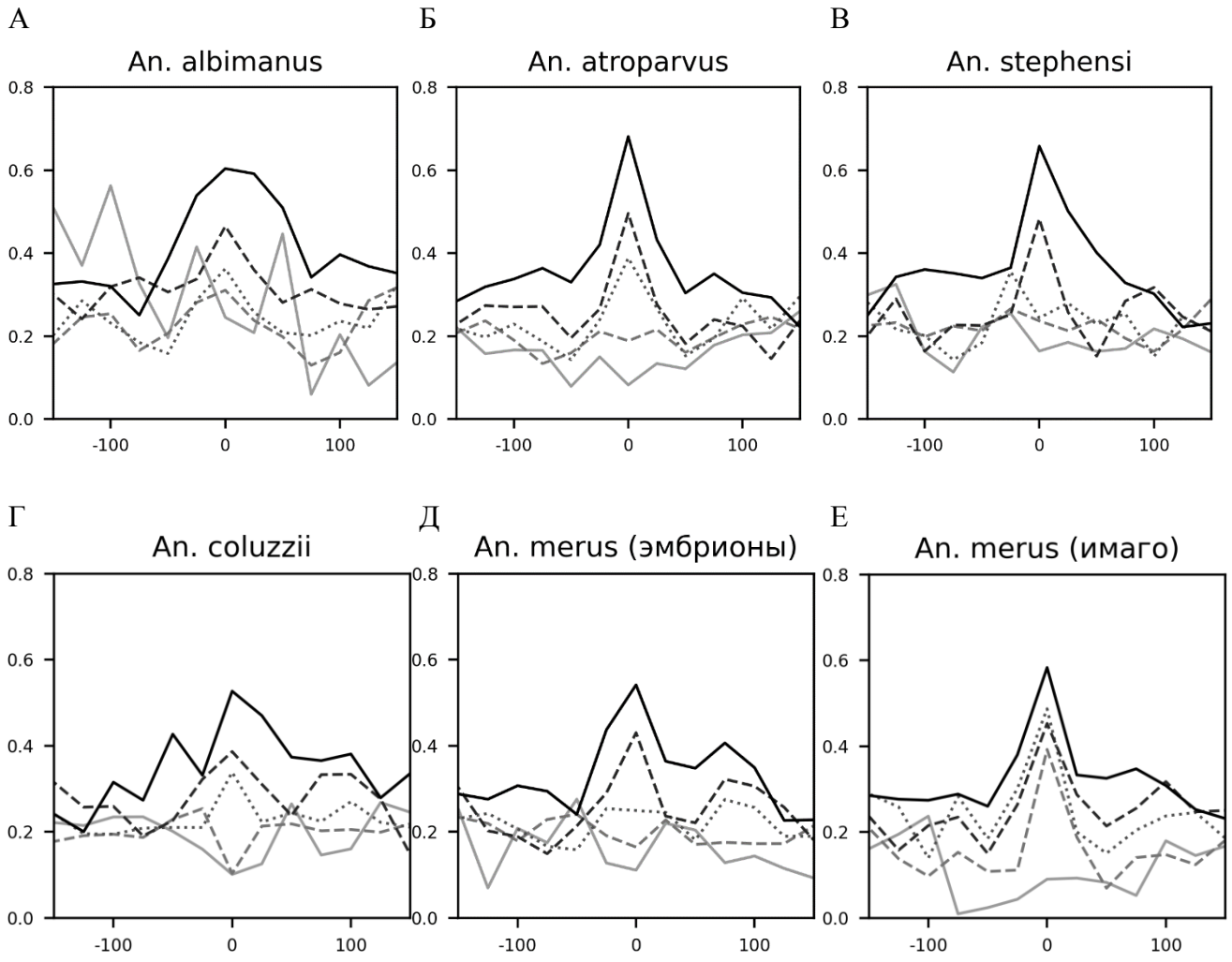

Рисунок 27. Среднее число эволюционных точек разрыва хромосомы на бин в зависимости от расстояния до границы домена. Сплошная чёрная линия – границы со значением компартмента больше 15, чёрная прерывистая – от 5 до 15, серая пунктирная – от -5 до 5 светло-серая прерывистая – от -15 до -5, светло-серая сплошная – меньше -15.

Для ответа этот вопрос, необходимо сравнить значения компартментов в локусах, затронутых эволюционной перестройкой (Рисунок 28А) у гипотетического предка. Конечно, невозможно определить, величину

компартментализации этих регионов у гипотетического предка, но учитывая высокую консервативность пространственной организации хроматина внутри синтенных блоков, разумно предположить, что у вида, филогенетически ближайшего к исследуемому, у которого в данном регионе блок синтении сохранился неразрывным, сохранилась и компартментализация, сходная с предковой.

Таким образом, было выделено около 3200 эволюционных точек разрыва для *An. albimanus* и около 400–650 точек разрыва для других видов рода *Anopheles*.

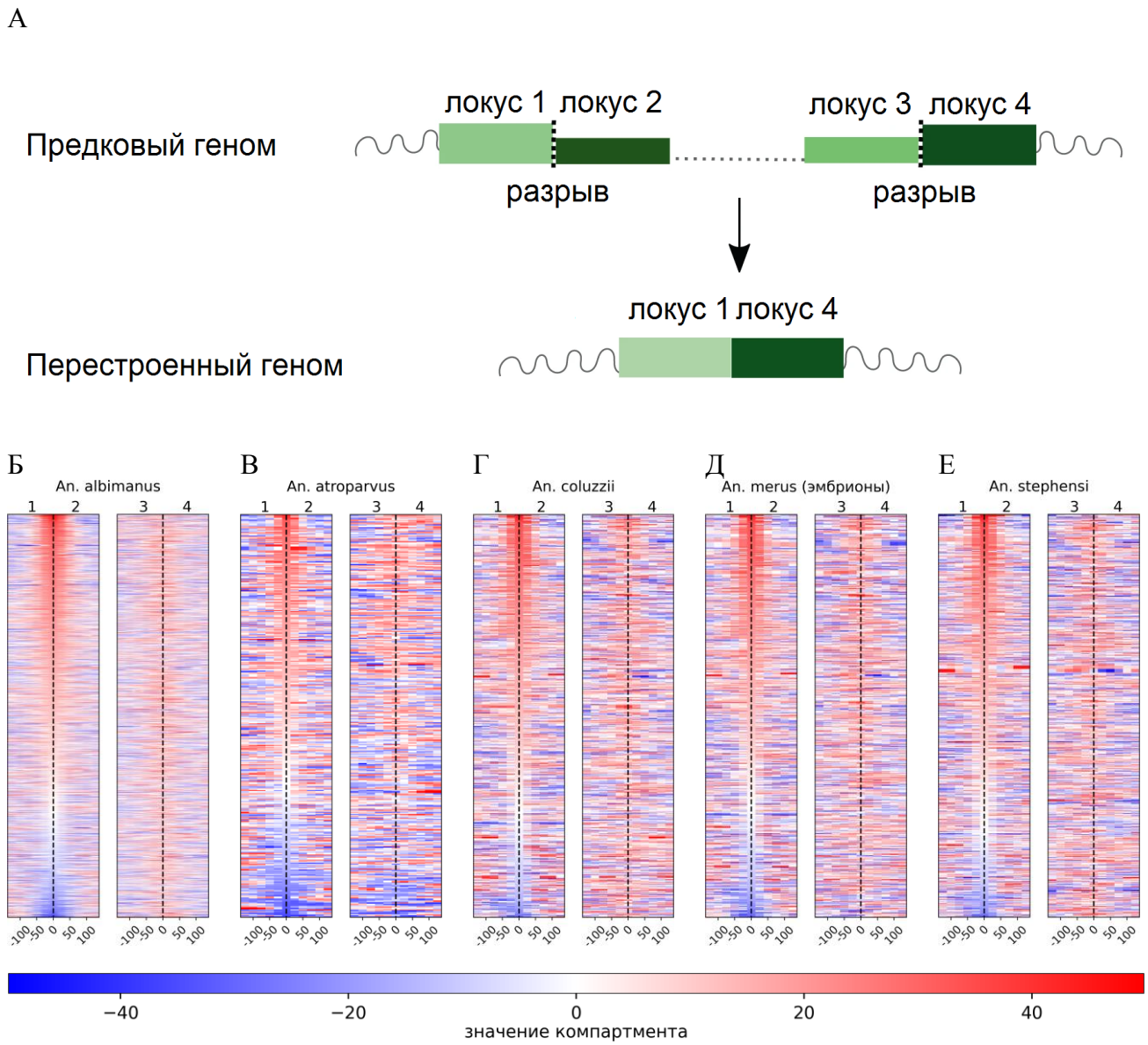

Рисунок 28. Влияние эволюционных перестроек на компартментализацию хроматина. А. Определение взаимного положения локусов до и после события эволюционной перестройки. Б-Е. Значение компартмента в локусах, синтенных перестроенным, у филогенетически ближайшего вида без перестройки.

Исследование показывает, что несмотря на то, что эволюционные разрывы хромосом чаще происходят в регионах, принадлежащих А-компартменту, соединяются они со случайными локусами, независимо от того, какому компартменту они принадлежат (Рисунок 28Б-Е). В свою очередь, высокая консервативность архитектуры хроматина на уровне отдельных контактов указывает, что новосоединённые в ходе перестройки локусы систематическое влияние на компартментное состояние друг друга не оказывают.

Таким образом для исследованных комаров рода *Anopheles* показана высокая консервативность архитектуры хроматина внутри синтенных блоков, прослеживая на уровне отдельных контактов. Также, в общем и целом, сохраняется эпигенетическое состояние регионов, соседствующих с точками разрыва синтении. Показано отсутствие явной связи частоты прохождения разрывов синтенных блоков с насыщенностью регионов повторами. В то же время разрывы существенно чаще происходят в регионах с активным хроматином. Это наблюдение согласуется с данными полученными для позвоночных: хромосомные перестройки чаще проходят по границам доменов, при этом границы доменов совпадают с регионами активного хроматина.

5. Заключение

Изучение принципов консервативности архитектуры хроматина, проведённое с помощью C-InterSecture, показало, что консервативными являются не сами пространственные контакты, как таковые, а их относительные величины, нормированные на расстояние между локусами. Эта закономерность обнаружена как при сравнении друг с другом разных позвоночных, так и комаров рода *Anopheles*. Основанная на данных закономерностях мера P-BAD позволяет количественно оценивать различия пространственной организации хроматина для выбранных локусов между разными видами, превосходя в этом ранее используемые подходы.

Общее сравнение архитектуры хроматина среди позвоночных показали, что в отличие от млекопитающих, разные типы клеток у *G. gallus* имеют радикально отличающуюся друг от друга организацию хроматина. Более того, эти различия сопоставимы с межвидовыми различиями. На основе этого можно сделать вывод, что механизм формирования архитектуры хроматина в эритроцитах *G. gallus* является отличным от таковых в других типах клеток. Так как между полихроматическими и зрелыми эритроцитами значимых различий не было обнаружено, можно сделать заключение, что наблюдаемые особенности не являются следствием отсутствия экспрессии генов в зрелых эритроцитах. Вероятно, эти отличия формируются на самых ранних стадиях клеточной дифференцировки и могут быть связаны с компактизацией хроматина. Следует отметить, что несмотря на то, что сперматозоиды млекопитающих также имеют сверхкомпактизованный хроматин, свойства хроматина от других типов клеток значимо не отличаются, а значит в процесс формирования эритроцитов у *G. gallus* включаются некие неизвестные ранее механизмы. В виду отсутствия взаимосвязи между структурой доменов и расположением сайтов связывания CTCF, можно утверждать, что эти механизмы не обусловлены взаимодействием когезина и белка CTCF.

В тоже время для фибробластов *G. gallus* все закономерности, которые обнаружены для млекопитающих, сохраняются. Это позволяет говорить о единстве механизмов, формирующих архитектуру хроматина и отнести его формирование как минимум ко времени разделения эволюционной линии млекопитающих и рептилий или даже ранее.

Пространственная организация хроматина комаров рода *Anopheles* демонстрирует ряд общих с *Diptera* свойств, в первую очередь выраженные теломер-теломерные и центромер-центромерные контакты, а также господство фазовой сепарации в формировании общего облика хроматина.

Эволюционное сравнение пространственной организации хроматина у комаров рода *Anopheles* показало, что границы синтенных блоков проходят преимущественно по А-компартментам, высоко экспрессирующихся районам, подобно тому, как это отмечено у позвоночных. При этом для зафиксированных в ходе эволюции перестроек не отмечено значимого влияние перестройки на компартментное состояние затронутых перестройкой локусов.

Выводы

1) Наблюдаемая архитектура хроматина эритроцитов G. gallus преимущественно обусловлена распределением активного и неактивного хроматина. Это позволяет утверждать, что в ходе эритропоэза у G. gallus происходит отключение механизма «протягивания петли» или барьерной функции белка CTCF. Данное свойство является уникальным для известных соматических клеток позвоночных.

2) Разработан алгоритм межвидового сравнения пространственной организации хроматина на уровне отдельных контактов, реализованный в виде пакета программ C InterSecture. Данное программное обеспечение позволяет эффективно сравнивать виды независимо от их организации хроматина, разницы в размерах геномов и разрешении используемых карт Hi-C.

3) Сравнение архитектуры хроматина как между разными видами позвоночных, так и между комарами рода Anopheles, показало консервативность архитектуры хроматина, прослеживаемую на уровне отдельных контактов.

4) Анализ архитектуры хроматина комаров рода Anopheles показывает, что чаще всего эволюционные разрывы хромосом проходят по регионам активного хроматина. Соединение разрывов происходит случайным образом, и трехмерная организация хроматина вблизи точек разрыва как правило сохраняется сходство с предковым состоянием. Использование для формирования хромосомных разрывов регионов с активной транскрипцией является свойством, общим у исследованных видов комаров и позвоночных.

6. Список литературы

1. Dekker J., Rippe K., Dekker M., Kleckner N. Capturing chromosome conformation // Science. – 2002. – V. 295. – № 5558. – P. 1306–1311.
2. Denker A., de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization // Genes & Development. – 2016. – V. 30. – № 12. – P. 1357–1382.
3. Lieberman-Aiden E., van Berkum N.L., Williams L., Imakaev M., Ragoczy T., et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome // Science (New York, N.Y.). – 2009. – V. 326. – № 5950. – P. 289–293.
4. Belton J.-M., McCord R.P., Gibcus J.H., Naumova N., Zhan Y., Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes // Methods (San Diego, Calif.). – 2012. – V. 58. – № 3. – P. 268–276.
5. Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J.S., Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions: 7398 // Nature. Nature Publishing Group, – 2012. – V. 485. – № 7398. – P. 376–380.
6. Shen Y., Yue F., McCleary D.F., Ye Z., Edsall L., et al. A map of the cis-regulatory sequences in the mouse genome // Nature. – 2012. – V. 488. – № 7409. – P. 116–120.
7. Jin F., Li Y., Dixon J.R., Selvaraj S., Ye Z., et al. A high-resolution map of the three-dimensional chromatin interactome in human cells // Nature. – 2013. – V. 503. – № 7475. – P. 290–294.
8. Rao S.S.P., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping // Cell. – 2014. – V. 159. – № 7. – P. 1665–1680.
9. Hou C., Li L., Qin Z.S., Corces V.G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains // Molecular Cell. – 2012. – V. 48. – № 3. – P. 471–484.
10. Sexton T., Yaffe E., Kenigsberg E., Bantignies F., Leblanc B., Hoichman M., Parrinello H., Tanay A., Cavalli G. Three-Dimensional Folding and Functional

Organization Principles of the Drosophila Genome // Cell. – 2012. – V. 148. – № 3. – P. 458–472.

11. Neafsey D.E., Waterhouse R.M., Abai M.R., Aganezov S.S., Alekseyev M.A., et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes // Science (New York, N.Y.). – 2015. – V. 347. – № 6217. – P. 1258522.
12. Heitz, E. Das heterochromatin der moose // Jahrbücher für Wissenschaftliche Botanik. – 1928. – V. 69. – P. 762–818.
13. Heitz E. Die somatische Heteropyknose bei Drosophila melanogaster und ihre genetische Bedeutung // Zeitschrift für Zellforschung und Mikroskopische Anatomie. – 1933. – V. 20. – № 1. – P. 237–287.
14. Brown S.W. Heterochromatin // Science. – 1966. – V. 151. – № 3709. – P. 417–425.
15. Muller H.J. Types of visible variations induced by X-rays inDrosophila // Journal of Genetics. – 1930. – V. 22. – № 3. – P. 299–334.
16. Schultz J., Dobzhansky Th. The Relation of a Dominant Eye Color in Drosophila Melanogaster to the Associated Chromosome Rearrangement // Genetics. – 1934. – V. 19. – № 4. – P. 344–364.
17. Van Holde K.E., Sahasrabuddhe C.G., Shaw B.R. A model for particulate structure in chromatin // Nucleic Acids Research. – 1974. – V. 1. – № 11. – P. 1579–1586.
18. Carpenter B.G., Baldwin J.P., Bradbury E.M., Ibel K. Organisation of subunits in chromatin. // Nucleic Acids Research. – 1976. – V. 3. – № 7. – P. 1739–1746.
19. Finch J.T., Klug A. Solenoidal model for superstructure in chromatin. // Proceedings of the National Academy of Sciences of the United States of America. – 1976. – V. 73. – № 6. – P. 1897–1901.
20. Getzenberg R.H., Pienta K.J., Ward W.S., Coffey D.S. Nuclear structure and the three-dimensional organization of DNA // Journal of Cellular Biochemistry. – 1991. – V. 47. – № 4. – P. 289–299.

21. Wolffe A.P., Guschin D. Review: Chromatin Structural Features and Targets That Regulate Transcription // Journal of Structural Biology. – 2000. – V. 129. – № 2. – P. 102–122.

22. Woodcock C.L., Dimitrov S. Higher-order structure of chromatin and chromosomes // Current Opinion in Genetics & Development. – 2001. – V. 11. – № 2. – P. 130–135.

23. Cremer T., Kreth G., Koester H., Fink R.H., Heintzmann R., Cremer M., Solovei I., Zink D., Cremer C. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture // Critical Reviews in Eukaryotic Gene Expression. – 2000. – V. 10. – № 2. – P. 179–212.

24. Cremer T., Cremer M., Dietzel S., Müller S., Solovei I., Fakan S. Chromosome territories--a functional nuclear landscape // Current Opinion in Cell Biology. – 2006. – V. 18. – № 3. – P. 307–316.

25. Cremer T., Cremer M. Chromosome Territories // Cold Spring Harbor Perspectives in Biology. – 2010. – V. 2. – № 3. – P. a003889.

26. Cremer T., Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells: 4 // Nature Reviews Genetics. Nature Publishing Group, – 2001. – V. 2. – № 4. – P. 292–301.

27. Taddei A., Hediger F., Neumann F.R., Gasser S.M. The function of nuclear architecture: a genetic approach // Annual Review of Genetics. – 2004. – V. 38. – P. 305–345.

28. Guelen L., Pagie L., Brasset E., Meuleman W., Faza M., et al. Corrigendum: Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions // Nature. – 2008. – V. 453. – P. 948–951.

29. Peric-Hupkes D., Meuleman W., Pagie L., Bruggeman S.W.M., Solovei I., et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation // Molecular Cell. – 2010. – V. 38. – № 4. – P. 603–613.

30. Hagstrom K.A., Meyer B.J. Condensin and cohesin: more than chromosome compactor and glue // Nature Reviews. Genetics. – 2003. – V. 4. – № 7. – P. 520–534.

31. Liu J., Krantz I.D. Cohesin and Human Disease // Annual review of genomics and human genetics. – 2008. – V. 9. – P. 303–320.

32. Rudra S., Skibbens R.V. Cohesin codes – interpreting chromatin architecture and the many facets of cohesin function // Journal of Cell Science. – 2013. – V. 126. – № 1. – P. 31–41.

33. Rubio E.D., Reiss D.J., Welcsh P.L., Disteche C.M., Filippova G.N., Baliga N.S., Aebersold R., Ranish J.A., Krumm A. CTCF physically links cohesin to chromatin // Proceedings of the National Academy of Sciences of the United States of America. – 2008. – V. 105. – № 24. – P. 8309–8314.

34. Fraser P., Bickmore W. Nuclear organization of the genome and the potential for gene regulation // Nature. – 2007. – V. 447. – P. 413–417.

35. Servant N., Varoquaux N., Lajoie B.R., Viara E., Chen C.-J., Vert J.-P., Heard E., Dekker J., Barillot E. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing // Genome Biology. – 2015. – V. 16. – P. 259.

36. Abdennur N., Mirny L.A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays // Bioinformatics. – 2019. – V. 36. – № 1. – P. 311–316.

37. Durand N.C., Shamim M.S., Machol I., Rao S.S.P., Huntley M.H., Lander E.S., Aiden E.L. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments // Cell Systems. – 2016. – V. 3. – № 1. – P. 95–98.

38. Durand N.C., Robinson J.T., Shamim M.S., Machol I., Mesirov J.P., Lander E.S., Aiden E.L. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom // Cell Systems. – 2016. – V. 3. – № 1. – P. 99–101.

39. Dekker J. The three “C” s of chromosome conformation capture: controls, controls, controls // Nature Methods. – 2006. – V. 3. – № 1. – P. 17–21.

40. Yaffe E., Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture // Nature Genetics. – 2011. – V. 43. – № 11. – P. 1059–1065.

41. Hu M., Deng K., Selvaraj S., Qin Z., Ren B., Liu J.S. HiCNorm: removing biases in Hi-C data via Poisson regression // Bioinformatics. – 2012. – V. 28. – № 23. – P. 3131–3133.

42. Knight P.A., Ruiz D. A fast algorithm for matrix balancing // IMA Journal of Numerical Analysis. – 2013. – V. 33. – № 3. – P. 1029–1047.

43. Imakaev M., Fudenberg G., McCord R.P., Naumova N., Goloborodko A., Lajoie B.R., Dekker J., Mirny L.A. Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization // Nature methods. – 2012. – V. 9. – № 10. – P. 999–1003.

44. Battulin N., Fishman V.S., Mazur A.M., Pomaznoy M., Khabarova A.A., Afonnikov D.A., Prokhortchouk E.B., Serov O.L. Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach // Genome Biology. – 2015. – V. 16. – № 1. – P. 77.

45. Vietri Rudan M., Barrington C., Henderson S., Ernst C., Odom D.T., Tanay A., Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture // Cell Reports. – 2015. – V. 10. – № 8. – P. 1297–1309.

46. Duan Z., Andronescu M., Schutz K., McIlwain S., Kim Y.J., et al. A three-dimensional model of the yeast genome // Nature. – 2010. – V. 465. – № 7296. – P. 363–367.

47. Grob S., Schmid M.W., Grossniklaus U. Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila // Molecular Cell. – 2014. – V. 55. – № 5. – P. 678–693.

48. Feng S., Cokus S.J., Schubert V., Zhai J., Pellegrini M., Jacobsen S.E. Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis // Molecular Cell. – 2014. – V. 55. – № 5. – P. 694–707.

49. Le T.B.K., Imakaev M.V., Mirny L.A., Laub M.T. High-resolution mapping of the spatial organization of a bacterial chromosome // Science (New York, N.Y.). – 2013. – V. 342. – № 6159. – P. 731–734.

50. Zhang Y., McCord R.P., Ho Y.-J., Lajoie B.R., Hildebrand D.G., Simon A.C., Becker M.S., Alt F.W., Dekker J. Chromosomal translocations are guided by the spatial organization of the genome // Cell. – 2012. – V. 148. – № 5. – P. 908–921.

51. Phillips-Cremins J.E., Sauria M.E.G., Sanyal A., Gerasimova T.I., Lajoie B.R., et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment // Cell. – 2013. – V. 153. – № 6. – P. 1281–1295.

52. Fraser J., Ferrai C., Chiariello A.M., Schueler M., Rito T., et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation // Molecular Systems Biology. – 2015. – V. 11. – № 12. – P. 852.

53. Rhodes J.D.P., Feldmann A., Hernández-Rodríguez B., Díaz N., Brown J.M., et al. Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells // Cell Reports. – 2020. – V. 30. – № 3. – P. 820-835.e10.

54. Boyle S., Flyamer I.M., Williamson I., Sengupta D., Bickmore W.A., Illingworth R.S. A central role for canonical PRC1 in shaping the 3D nuclear landscape // Genes & Development. – 2020. – V. 34. – № 13–14. – P. 931–949.

55. Sturtevant A.H. The Effects of Unequal Crossing over at the Bar Locus in Drosophila // Genetics. – 1925. – V. 10. – № 2. – P. 117–147.

56. Stadler M.R., Haines J.E., Eisen M.B. Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo // eLife. – 2017. – V. 6. – P. e29550.

57. Ulianov S.V., Khrameeva E.E., Gavrilov A.A., Flyamer I.M., Kos P., et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains // Genome Research. – 2016. – V. 26. – № 1. – P. 70–84.

58. Cubeñas-Potts C., Rowley M.J., Lyu X., Li G., Lei E.P., Corces V.G. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture // Nucleic Acids Research. – 2017. – V. 45. – № 4. – P. 1714–1730.

59. Wang Q., Sun Q., Czajkowsky D.M., Shao Z. Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells // Nature Communications. – 2018. – V. 9. – № 1. – P. 188.

60. Ramírez F., Bhardwaj V., Arrigoni L., Lam K.C., Grüning B.A., Villaveces J., Habermann B., Akhtar A., Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies: 1 // Nature Communications. Nature Publishing Group, – 2018. – V. 9. – № 1. – P. 189.

61. Kaushal A., Mohana G., Dorier J., Özdemir I., Omer A., et al. CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions // Nature Communications. – 2021. – V. 12. – № 1. – P. 1011.

62. Kaushal A., Dorier J., Wang B., Mohana G., Taschner M., et al. Essential role of Cp190 in physical and regulatory boundary formation // Science Advances. – 2022. – V. 8. – № 19. – P. eabl8834.

63. Sun Q., Perez-Rathke A., Czajkowsky D.M., Shao Z., Liang J. High-resolution single-cell 3D-models of chromatin ensembles during Drosophila embryogenesis // Nature Communications. – 2021. – V. 12. – P. 205.

64. Ulianov S.V., Zakharova V.V., Galitsyna A.A., Kos P.I., Polovnikov K.E., et al. Order and stochasticity in the folding of individual Drosophila genomes // Nature Communications. – 2021. – V. 12. – P. 41.

65. Eagen K.P., Aiden E.L., Kornberg R.D. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map // Proceedings of the National Academy of Sciences of the United States of America. – 2017. – V. 114. – № 33. – P. 8764–8769.

66. Alipour E., Marko J.F. Self-organization of domain structures by DNA-loop-extruding enzymes // Nucleic Acids Research. – 2012. – V. 40. – № 22. – P. 11202–11212.

67. Sofueva S., Yaffe E., Chan W.-C., Georgopoulou D., Vietri Rudan M., et al. Cohesin-mediated interactions organize chromosomal domain architecture // The EMBO Journal. – 2013. – V. 32. – № 24. – P. 3119–3129.

68. Sanborn A.L., Rao S.S.P., Huang S.-C., Durand N.C., Huntley M.H., et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes // Proceedings of the National Academy of Sciences of the United States of America. – 2015. – V. 112. – № 47. – P. E6456-6465.

69. Tark-Dame M., Jerabek H., Manders E.M.M., van der Wateren I.M., Heermann D.W., van Driel R. Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling // PLoS computational biology. – 2014. – V. 10. – № 10. – P. e1003877.

70. Zuin J., Dixon J.R., van der Reijden M.I.J.A., Ye Z., Kolovos P., et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells // Proceedings of the National Academy of Sciences of the United States of America. – 2014. – V. 111. – № 3. – P. 996–1001.

71. Guo Y., Xu Q., Canzio D., Shou J., Li J., et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function // Cell. – 2015. – V. 162. – № 4. – P. 900–910.

72. Nora E.P., Goloborodko A., Valton A.-L., Gibcus J.H., Uebersohn A., Abdennur N., Dekker J., Mirny L.A., Bruneau B.G. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization // Cell. – 2017. – V. 169. – № 5. – P. 930-944.e22.

73. Wutz G., Várnai C., Nagasaka K., Cisneros D.A., Stocsits R.R., et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins // The EMBO Journal. – 2017. – V. 36. – № 24. – P. 3573–3599.

74. Nagano T., Lubling Y., Stevens T.J., Schoenfelder S., Yaffe E., Dean W., Laue E.D., Tanay A., Fraser P. Single cell Hi-C reveals cell-to-cell variability in chromosome structure // Nature. – 2013. – V. 502. – № 7469. – P. 10.1038/nature12593.

75. Flyamer I.M., Gassler J., Imakaev M., Brandão H.B., Ulianov S.V., Abdennur N., Razin S.V., Mirny L.A., Tachibana-Konwalski K. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition: 7648 // Nature. Nature Publishing Group, – 2017. – V. 544. – № 7648. – P. 110–114.

76. Gassler J., Brandão H.B., Imakaev M., Flyamer I.M., Ladstätter S., Bickmore W.A., Peters J., Mirny L.A., Tachibana K. A mechanism of cohesin-dependent loop

extrusion organizes zygotic genome architecture // The EMBO Journal. – 2017. – V. 36. – № 24. – P. 3600–3618.

77. Rao S.S.P., Huang S.-C., Glenn St Hilaire B., Engreitz J.M., Perez E.M., et al. Cohesin Loss Eliminates All Loop Domains // Cell. – 2017. – V. 171. – № 2. – P. 305-320.e24.

78. Barbieri M., Chotalia M., Fraser J., Lavitas L.-M., Dostie J., Pombo A., Nicodemi M. Complexity of chromatin folding is captured by the strings and binders switch model // Proceedings of the National Academy of Sciences of the United States of America. – 2012. – V. 109. – № 40. – P. 16173–16178.

79. Benedetti F., Dorier J., Burnier Y., Stasiak A. Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes // Nucleic Acids Research. – 2014. – V. 42. – № 5. – P. 2848–2855.

80. Giorgetti L., Galupa R., Nora E.P., Piolot T., Lam F., Dekker J., Tiana G., Heard E. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription // Cell. – 2014. – V. 157. – № 4. – P. 950–963.

81. Fudenberg G., Imakaev M., Lu C., Goloborodko A., Abdennur N., Mirny L.A. Formation of Chromosomal Domains by Loop Extrusion // Cell Reports. – 2016. – V. 15. – № 9. – P. 2038–2049.

82. Hansen A.S., Pustova I., Cattoglio C., Tjian R., Darzacq X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics // eLife. – 2017. – V. 6. – P. e25776.

83. Kim Y., Shi Z., Zhang H., Finkelstein I.J., Yu H. Human cohesin compacts DNA by loop extrusion // Science (New York, N.Y.). – 2019. – V. 366. – № 6471. – P. 1345–1349.

84. Haarhuis J.H.I., van der Weide R.H., Blomen V.A., Yáñez-Cuna J.O., Amendola M., et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension // Cell. – 2017. – V. 169. – № 4. – P. 693-707.e14.

85. Vian L., Pękowska A.P., Rao S.S.P., Kieffer-Kwon K.-R., Jung S., et al. The energetics and physiological impact of cohesin extrusion // Cell. – 2018. – V. 173. – № 5. – P. 1165-1178.e20.

86. Nora E.P., Caccianini L., Fudenberg G., So K., Kameswaran V., et al. Molecular basis of CTCF binding polarity in genome folding // Nature Communications. – 2020. – V. 11. – P. 5612.

87. Pugacheva E.M., Kubo N., Loukinov D., Tajmul M., Kang S., et al. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention // Proceedings of the National Academy of Sciences of the United States of America. – 2020. – V. 117. – № 4. – P. 2020–2031.

88. Saldaña-Meyer R., González-Buendía E., Guerrero G., Narendra V., Bonasio R., Recillas-Targa F., Reinberg D. CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53 // Genes & Development. – 2014. – V. 28. – № 7. – P. 723–734.

89. Saldaña-Meyer R., Rodriguez-Hernaez J., Escobar T., Nishana M., Jácome-López K., et al. RNA Interactions Are Essential for CTCF-Mediated Genome Organization // Molecular cell. – 2019. – V. 76. – № 3. – P. 412-422.e5.

90. Hansen A.S., Hsieh T.-H.S., Cattoglio C., Pustova I., Saldaña-Meyer R., Reinberg D., Darzacq X., Tjian R. Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF // Molecular cell. – 2019. – V. 76. – № 3. – P. 395-411.e13.

91. Weintraub A.S., Li C.H., Zamudio A.V., Sigova A.A., Hannett N.M., et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops // Cell. – 2017. – V. 171. – № 7. – P. 1573-1588.e28.

92. Akgol Oksuz B., Yang L., Abraham S., Venev S.V., Krietenstein N., et al. Systematic evaluation of chromosome conformation capture assays // Nature Methods. – 2021. – V. 18. – № 9. – P. 1046–1055.

93. Tang Z., Luo O.J., Li X., Zheng M., Zhu J.J., et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription // Cell. – 2015. – V. 163. – № 7. – P. 1611–1627.

94. de Llobet Cucalon L., Di Vona C., Morselli M., Vezzoli M., Montanini B., Teichmann M., de la Luna S., Ferrari R. An RNA Polymerase III General

Transcription Factor Engages in Cell Type-Specific Chromatin Looping // International Journal of Molecular Sciences. – 2022. – V. 23. – № 4. – P. 2260.

95. Nicodemi M., Panning B., Prisco A. A Thermodynamic Switch for Chromosome Colocalization // Genetics. – 2008. – V. 179. – № 1. – P. 717–721.

96. Nicodemi M., Prisco A. Thermodynamic Pathways to Genome Spatial Organization in the Cell Nucleus // Biophysical Journal. – 2009. – V. 96. – № 6. – P. 2168–2177.

97. Nicodemi M., Pombo A. Models of chromosome structure // Current Opinion in Cell Biology. – 2014. – V. 28. – P. 90–95.

98. Chandra T., Kirschner K., Thuret J.-Y., Pope B.D., Ryba T., et al. Independence of Repressive Histone Marks and Chromatin Compaction during Senescent Heterochromatic Layer Formation // Molecular cell. – 2012. – V. 47. – № 2. – P. 203–214.

99. Jost D., Carrivain P., Cavalli G., Vaillant C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains // Nucleic Acids Research. – 2014. – V. 42. – № 15. – P. 9553–9561.

100. Michieletto D., Orlandini E., Marenduzzo D. Polymer model with Epigenetic Recoloring Reveals a Pathway for the de novo Establishment and 3D Organization of Chromatin Domains // Physical Review X. American Physical Society, – 2016. – V. 6. – № 4. – P. 041047.

101. Feric M., Vaidya N., Harmon T.S., Mitrea D.M., Zhu L., Richardson T.M., Kriwacki R.W., Pappu R.V., Brangwynne C.P. Coexisting liquid phases underlie nucleolar sub-compartments // Cell. – 2016. – V. 165. – № 7. – P. 1686–1697.

102. Hnisz D., Shrinivas K., Young R.A., Chakraborty A.K., Sharp P.A. A phase separation model predicts key features of transcriptional control // Cell. – 2017. – V. 169. – № 1. – P. 13–23.

103. Strom A.R., Emelyanov A.V., Mir M., Fyodorov D.V., Darzacq X., Karpen G.H. Phase separation drives heterochromatin domain formation // Nature. – 2017. – V. 547. – № 7662. – P. 241–245.

104. Plys A.J., Davis C.P., Kim J., Rizki G., Keenen M.M., Marr S.K., Kingston R.E. Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2 // Genes & Development. – 2019. – V. 33. – № 13–14. – P. 799–813.

105. Tatavosian R., Kent S., Brown K., Yao T., Duc H.N., et al. Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation // The Journal of Biological Chemistry. – 2019. – V. 294. – № 5. – P. 1451–1463.

106. Hwang Y.-C., Zheng Q., Gregory B.D., Wang L.-S. High-throughput identification of long-range regulatory elements and their target promoters in the human genome // Nucleic Acids Research. – 2013. – V. 41. – № 9. – P. 4835–4846.

107. Hwang Y.-C., Lin C.-F., Valladares O., Malamon J., Kuksa P.P., Zheng Q., Gregory B.D., Wang L.-S. HIPPIE: a high-throughput identification pipeline for promoter interacting enhancer elements // Bioinformatics. – 2015. – V. 31. – № 8. – P. 1290–1292.

108. Xu Z., Zhang G., Jin F., Chen M., Furey T.S., Sullivan P.F., Qin Z., Hu M., Li Y. A hidden Markov random field-based Bayesian method for the detection of long-range chromosomal interactions in Hi-C data // Bioinformatics. – 2016. – V. 32. – № 5. – P. 650–656.

109. Zhou Y., Cheng X., Yang Y., Li T., Li J., Huang T.H.-M., Wang J., Lin S., Jin V.X. Modeling and analysis of Hi-C data by HiSIF identifies characteristic promoter-distal loops // Genome Medicine. – 2020. – V. 12. – P. 69.

110. Sahlén P., Abdullayev I., Ramsköld D., Matskova L., Rilakovic N., Lötstedt B., Albert T.J., Lundeberg J., Sandberg R. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution // Genome Biology. – 2015. – V. 16. – № 1. – P. 156.

111. Mifsud B., Tavares-Cadete F., Young A.N., Sugar R., Schoenfelder S., et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C // Nature Genetics. – 2015. – V. 47. – № 6. – P. 598–606.

112. Lu L., Liu X., Huang W.-K., Giusti-Rodríguez P., Cui J., et al. Robust Hi-C maps of enhancer-promoter interactions reveal the function of non-coding genome in neural

development and diseases // Molecular cell. – 2020. – V. 79. – № 3. – P. 521-534.e15.

113. Yang H., Luan Y., Liu T., Lee H.J., Fang L., et al. A map of cis-regulatory elements and 3D genome structures in zebrafish // Nature. – 2020. – V. 588. – № 7837. – P. 337–343.

114. Javierre B.M., Burren O.S., Wilder S.P., Kreuzhuber R., Hill S.M., et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters // Cell. – 2016. – V. 167. – № 5. – P. 1369-1384.e19.

115. Ma W., Ay F., Lee C., Gulsoy G., Deng X., et al. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes // Nature methods. – 2015. – V. 12. – № 1. – P. 71–78.

116. Wang Z., Cao R., Taylor K., Briley A., Caldwell C., Cheng J. The Properties of Genome Conformation and Spatial Gene Interaction and Regulation Networks of Normal and Malignant Human Cell Types // PLoS ONE. – 2013. – V. 8. – № 3. – P. e58793.

117. Zhu Y., Chen Z., Zhang K., Wang M., Medovoy D., et al. Constructing 3D interaction maps from 1D epigenomes // Nature Communications. – 2016. – V. 7. – P. 10812.

118. Beagrie R.A., Scialdone A., Schueler M., Kraemer D.C.A., Chotalia M., et al. Complex multi-enhancer contacts captured by Genome Architecture Mapping (GAM) // Nature. – 2017. – V. 543. – № 7646. – P. 519–524.

119. Ma X., Ezer D., Adryan B., Stevens T.J. Canonical and single-cell Hi-C reveal distinct chromatin interaction sub-networks of mammalian transcription factors // Genome Biology. – 2018. – V. 19. – P. 174.

120. Chen H., Xiao J., Shao T., Wang L., Bai J., et al. Landscape of Enhancer-Enhancer Cooperative Regulation during Human Cardiac Commitment // Molecular Therapy. Nucleic Acids. – 2019. – V. 17. – P. 840–851.

121. Sobhy H., Kumar R., Lewerentz J., Lizana L., Stenberg P. Highly interacting regions of the human genome are enriched with enhancers and bound by DNA repair proteins // Scientific Reports. – 2019. – V. 9. – P. 4577.

122. Zhu I., Song W., Ovcharenko I., Landsman D. A model of active transcription hubs that unifies the roles of active promoters and enhancers // Nucleic Acids Research. – 2021. – V. 49. – № 8. – P. 4493–4505.

123. Huang J., Li K., Cai W., Liu X., Zhang Y., Orkin S.H., Xu J., Yuan G.-C. Dissecting super-enhancer hierarchy based on chromatin interactions // Nature Communications. – 2018. – V. 9. – P. 943.

124. Gong Y., Lazaris C., Sakellaropoulos T., Lozano A., Kambadur P., Ntziachristos P., Aifantis I., Tsirigos A. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries // Nature Communications. – 2018. – V. 9. – P. 542.

125. Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E., et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA // Science (New York, N.Y.). – 2012. – V. 337. – № 6099. – P. 1190–1195.

126. Davison L.J., Wallace C., Cooper J.D., Cope N.F., Wilson N.K., et al. Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene // Human Molecular Genetics. – 2012. – V. 21. – № 2. – P. 322–333.

127. Jäger R., Migliorini G., Henrion M., Kandaswamy R., Speedy H.E., et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci // Nature Communications. – 2015. – V. 6. – P. 6178.

128. Law P.J., Sud A., Mitchell J.S., Henrion M., Orlando G., et al. Genome-wide association analysis of chronic lymphocytic leukaemia, Hodgkin lymphoma and multiple myeloma identifies pleiotropic risk loci // Scientific Reports. – 2017. – V. 7. – P. 41071.

129. Litchfield K., Levy M., Orlando G., Loveday C., Law P., et al. Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor // Nature genetics. – 2017. – V. 49. – № 7. – P. 1133–1140.

130. Went M., Sud A., Försti A., Halvarsson B.-M., Weinhold N., et al. Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma // Nature Communications. – 2018. – V. 9. – P. 3707.

131. Cornish A.J., Hoang P.H., Dobbins S.E., Law P.J., Chubb D., Orlando G., Houlston R.S. Identification of recurrent noncoding mutations in B-cell lymphoma using capture Hi-C // Blood Advances. – 2019. – V. 3. – № 1. – P. 21–32.

132. Law P.J., Timofeeva M., Fernandez-Rozadilla C., Broderick P., Studd J., et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility // Nature Communications. – 2019. – V. 10. – P. 2154.

133. Martin P., McGovern A., Orozco G., Duffus K., Yarwood A., et al. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci // Nature Communications. – 2015. – V. 6. – P. 10069.

134. Burren O.S., Rubio García A., Javierre B.-M., Rainbow D.B., Cairns J., et al. Chromosome contacts in activated T cells identify autoimmune disease candidate genes // Genome Biology. – 2017. – V. 18. – P. 165.

135. Loviglio M.N., Leleu M., Männik K., Passeggeri M., Giannuzzi G., et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes // Molecular Psychiatry. – 2017. – V. 22. – № 6. – P. 836–849.

136. Song M., Yang X., Ren X., Maliskova L., Li B., et al. Mapping cis-Regulatory Chromatin Contacts in Neural Cells Links Neuropsychiatric Disorder Risk Variants to Target Genes // Nature genetics. – 2019. – V. 51. – № 8. – P. 1252–1262.

137. Rosa-Garrido M., Chapski D.J., Schmitt A.D., Kimball T.H., Karbassi E., et al. High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure // Circulation. – 2017. – V. 136. – № 17. – P. 1613–1625.

138. Choy M.-K., Javierre B.M., Williams S.G., Baross S.L., Liu Y., et al. Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks // Nature Communications. – 2018. – V. 9. – P. 2526.

139. Montefiori L.E., Sobreira D.R., Sakabe N.J., Aneas I., Joslin A.C., et al. A promoter interaction map for cardiovascular disease genetics // eLife. – V. 7. – P. e35788.

140. Criscione S.W., De Cecco M., Siranosian B., Zhang Y., Kreiling J.A., Sedivy J.M., Neretti N. Reorganization of chromosome architecture in replicative cellular senescence // Science Advances. – 2016. – V. 2. – № 2. – P. e1500882.

141. Boya R., Yadavalli A.D., Nikhat S., Kurukuti S., Palakodeti D., Pongubala J.M.R. Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment // Nucleic Acids Research. – 2017. – V. 45. – № 19. – P. 11070–11087.

142. Novo C.L., Javierre B.-M., Cairns J., Segonds-Pichon A., Wingett S.W., et al. Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition // Cell Reports. – 2018. – V. 22. – № 10. – P. 2615–2627.

143. Di Stefano M., Stadhouders R., Farabella I., Castillo D., Serra F., Graf T., Marti-Renom M.A. Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs // Nature Communications. – 2020. – V. 11. – P. 2564.

144. Taberlay P.C., Achinger-Kawecka J., Lun A.T.L., Buske F.A., Sabir K., et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations // Genome Research. – 2016. – V. 26. – № 6. – P. 719–731.

145. Sauerwald N., Kingsford C. Quantifying the similarity of topological domains across normal and cancer human cell types // Bioinformatics (Oxford, England). – 2018. – V. 34. – № 13. – P. i475–i483.

146. Vilarrasa-Blasi R., Soler-Vila P., Verdaguer-Dot N., Russiñol N., Di Stefano M., et al. Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation // Nature Communications. – 2021. – V. 12. – P. 651.

147. Ren B., Yang J., Wang C., Yang G., Wang H., et al. High-resolution Hi-C maps highlight multiscale 3D epigenome reprogramming during pancreatic cancer metastasis // Journal of Hematology & Oncology. – 2021. – V. 14. – P. 120.

148. Kim T., Han S., Chun Y., Yang H., Min H., Jeon S.Y., Kim J., Moon H.-G., Lee D. Comparative characterization of 3D chromatin organization in triple-negative breast cancers // Experimental & Molecular Medicine. – 2022. – V. 54. – № 5. – P. 585–600.

149. Rubin A.J., Barajas B.C., Furlan-Magaril M., Lopez-Pajares V., Mumbach M.R., et al. Lineage-specific dynamic and pre-established enhancer–promoter contacts cooperate in terminal differentiation // Nature genetics. – 2017. – V. 49. – № 10. – P. 1522–1528.

150. Phanstiel D.H., Van Bortle K., Spacek D., Hess G.T., Shamim M.S., et al. Static and dynamic DNA loops form AP-1 bound activation hubs during macrophage development // Molecular cell. – 2017. – V. 67. – № 6. – P. 1037-1048.e6.

151. Luo Z., Wang X., Jiang H., Wang R., Chen J., et al. Reorganized 3D Genome Structures Support Transcriptional Regulation in Mouse Spermatogenesis // iScience. – 2020. – V. 23. – № 4. – P. 101034.

152. Franke M., De la Calle-Mustienes E., Neto A., Almuedo-Castillo M., Irastorza-Azcarate I., Acemel R.D., Tena J.J., Santos-Pereira J.M., Gómez-Skarmeta J.L. CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression // Nature Communications. – 2021. – V. 12. – P. 5415.

153. Kojic A., Cuadrado A., De Koninck M., Giménez-Llorente D., Rodríguez-Corsino M., Gómez-López G., Le Dily F., Marti-Renom M.A., Losada A. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization // Nature structural & molecular biology. – 2018. – V. 25. – № 6. – P. 496–504.

154. Cuadrado A., Giménez-Llorente D., Kojic A., Rodríguez-Corsino M., Cuartero Y., Martín-Serrano G., Gómez-López G., Marti-Renom M.A., Losada A. Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and Polycomb Domains in Embryonic Stem Cells // Cell Reports. – 2019. – V. 27. – № 12. – P. 3500-3510.e4.

155. Liu N.Q., Maresca M., van den Brand T., Braccioli L., Schijns M.M.G.A., Teunissen H., Bruneau B.G., Nora E.P., de Wit E. WAPL maintains a cohesin

loading cycle to preserve cell-type specific distal gene regulation // Nature genetics. – 2021. – V. 53. – № 1. – P. 100–109.

156. Guo Y., Perez A.A., Hazelett D.J., Coetzee G.A., Rhie S.K., Farnham P.J. CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops // Genome Biology. – 2018. – V. 19. – P. 160.

157. Zhang D., Huang P., Sharma M., Keller C.A., Giardine B., et al. Alteration of genome folding via contact domain boundary insertion // Nature genetics. – 2020. – V. 52. – № 10. – P. 1076–1087.

158. Willemin A., Lopez-Delisle L., Bolt C.C., Gadolini M.-L., Duboule D., Rodriguez-Carballo E. Induction of a chromatin boundary in vivo upon insertion of a TAD border // PLoS Genetics. – 2021. – V. 17. – № 7. – P. e1009691.

159. Wu H.-J., Landshammer A., Stamenova E.K., Bolondi A., Kretzmer H., Meissner A., Michor F. Topological isolation of developmental regulators in mammalian genomes // Nature Communications. – 2021. – V. 12. – P. 4897.

160. Willi M., Yoo K.H., Reinisch F., Kuhns T.M., Lee H.K., Wang C., Hennighausen L. Facultative CTCF sites moderate mammary super-enhancer activity and regulate juxtaposed gene in non-mammary cells // Nature Communications. – 2017. – V. 8. – P. 16069.

161. Qi Q., Cheng L., Tang X., He Y., Li Y., et al. Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis // Blood. – 2021. – V. 137. – № 10. – P. 1327–1339.

162. Lupiáñez D.G., Kraft K., Heinrich V., Krawitz P., Brancati F., et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions // Cell. – 2015. – V. 161. – № 5. – P. 1012–1025.

163. Ushiki A., Zhang Y., Xiong C., Zhao J., Georgakopoulos-Soares I., et al. Deletion of CTCF sites in the SHH locus alters enhancer–promoter interactions and leads to acheiropodia // Nature Communications. – 2021. – V. 12. – P. 2282.

164. Ding B., Liu Y., Liu Z., Zheng L., Xu P., et al. Noncoding loci without epigenomic signals can be essential for maintaining global chromatin organization and cell viability // Science Advances. – V. 7. – № 45. – P. eabi6020.

165. Liu Y., Ding B., Zheng L., Xu P., Liu Z., et al. Regulatory elements can be essential for maintaining broad chromatin organization and cell viability // Nucleic Acids Research. – 2022. – V. 50. – № 8. – P. 4340–4354.

166. Barutcu A.R., Maass P.G., Lewandowski J.P., Weiner C.L., Rinn J.L. A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus // Nature Communications. – 2018. – V. 9. – P. 1444.

167. Soler-Oliva M.E., Guerrero-Martínez J.A., Bachetti V., Reyes J.C. Analysis of the relationship between coexpression domains and chromatin 3D organization // PLoS Computational Biology. – 2017. – V. 13. – № 9. – P. e1005708.

168. Rodríguez-Carballo E., Lopez-Delisle L., Zhan Y., Fabre P.J., Beccari L., et al. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes // Genes & Development. – 2017. – V. 31. – № 22. – P. 2264–2281.

169. Hoencamp C., Dudchenko O., Elbatsh A.M.O., Brahmachari S., Raaijmakers J.A., et al. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type // Science. – 2021. – V. 372. – № 6545. – P. 984–989.

170. Martin D., Pantoja C., Miñán A.F., Valdes-Quezada C., Moltó E., et al. Genome-wide CTCF distribution in vertebrates defines equivalent sites that can aid in the identification of disease-associated genes // Nature structural & molecular biology. – 2011. – V. 18. – № 6. – P. 708–714.

171. Gómez-Marín C., Tena J.J., Acemel R.D., López-Mayorga M., Naranjo S., et al. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders // Proceedings of the National Academy of Sciences of the United States of America. – 2015. – V. 112. – № 24. – P. 7542–7547.

172. Kentepozidou E., Aitken S.J., Feig C., Stefflova K., Ibarra-Soria X., Odom D.T., Roller M., Flicek P. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains // Genome Biology. – 2020. – V. 21. – P. 5.

173. Azazi D., Mudge J.M., Odom D.T., Flicek P. Functional signatures of evolutionarily young CTCF binding sites // BMC Biology. – 2020. – V. 18. – P. 132.

174. Gilbertson S.E., Walter H.C., Gardner K., Wren S.N., Vahedi G., Weinmann A.S. Topologically associating domains are disrupted by evolutionary genome rearrangements forming species-specific enhancer connections in mice and humans // Cell reports. – 2022. – V. 39. – № 5. – P. 110769.

175. Schmidt D., Schwalie P.C., Wilson M.D., Ballester B., Gonçalves Â., et al. Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages // Cell. – 2012. – V. 148. – № 1–2. – P. 335–348.

176. Laverré A., Tannier E., Necsulea A. Long-range promoter–enhancer contacts are conserved during evolution and contribute to gene expression robustness // Genome Research. – 2022. – V. 32. – № 2. – P. 280–296.

177. Li D., He M., Tang Q., Tian S., Zhang J., et al. Comparative 3D genome architecture in vertebrates // BMC Biology. – 2022. – V. 20. – P. 99.

178. Corbo M., Damas J., Bursell M.G., Lewin H.A. Conservation of chromatin conformation in carnivores // Proceedings of the National Academy of Sciences of the United States of America. – 2022. – V. 119. – № 9. – P. e2120555119.

179. Eres I.E., Luo K., Hsiao C.J., Blake L.E., Gilad Y. Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates // PLoS Genetics. – 2019. – V. 15. – № 7. – P. e1008278.

180. Luo X., Liu Y., Dang D., Hu T., Hou Y., et al. 3D Genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis // Cell. – 2021. – V. 184. – № 3. – P. 723-740.e21.

181. Pérez-Rico Y.A., Barillot E., Shkumatava A. Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish // iScience. – 2020. – V. 23. – № 5. – P. 101046.

182. Nakamura R., Motai Y., Kumagai M., Wike C.L., Nishiyama H., et al. CTCF looping is established during gastrulation in medaka embryos // Genome Research. – 2021. – V. 31. – № 6. – P. 968–980.

183. Niu L., Shen W., Shi Z., Tan Y., He N., et al. Three-dimensional folding dynamics of the Xenopus tropicalis genome // Nature Genetics. – 2021. – V. 53. – № 7. – P. 1075–1087.

184. Heger P., Marin B., Schierenberg E. Loss of the insulator protein CTCF during nematode evolution // BMC Molecular Biology. – 2009. – V. 10. – P. 84.

185. Heger P., Marin B., Bartkuhn M., Schierenberg E., Wiehe T. The chromatin insulator CTCF and the emergence of metazoan diversity // Proceedings of the National Academy of Sciences of the United States of America. – 2012. – V. 109. – № 43. – P. 17507–17512.

186. Gregory T.R. The Animal Genome Size Database [Electronic resource]. – 2022.

187. Zhang G., Li C., Li Q., Li B., Larkin D.M., et al. Comparative genomics reveals insights into avian genome evolution and adaptation // Science (New York, N.Y.). – 2014. – V. 346. – № 6215. – P. 1311–1320.

188. Romanov M.N., Farré M., Lithgow P.E., Fowler K.E., Skinner B.M., et al. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor // BMC Genomics. – 2014. – V. 15. – № 1. – P. 1060.

189. Farré M., Narayan J., Slavov G.T., Damas J., Auvil L., et al. Novel Insights into Chromosome Evolution in Birds, Archosaurs, and Reptiles // Genome Biology and Evolution. – 2016. – V. 8. – № 8. – P. 2442–2451.

190. International Chicken Genome Sequencing Consortium Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution // Nature. – 2004. – V. 432. – № 7018. – P. 695–716.

191. Renschler G., Richard G., Valsecchi C.I.K., Toscano S., Arrigoni L., Ramírez F., Akhtar A. Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling // Genes & Development. – 2019. – V. 33. – № 21–22. – P. 1591–1612.

192. Liao Y., Zhang X., Chakraborty M., Emerson J.J. Topologically associating domains and their role in the evolution of genome structure and function in Drosophila // Genome Research. – 2021. – V. 31. – № 3. – P. 397–410.

193. Torosin N.S., Anand A., Golla T.R., Cao W., Ellison C.E. 3D genome evolution and reorganization in the Drosophila melanogaster species group // PLoS Genetics. – 2020. – V. 16. – № 12. – P. e1009229.

194. Ghavi-Helm Y., Jankowski A., Meiers S., Viales R.R., Korbel J.O., Furlong E.E.M. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression // Nature genetics. – 2019. – V. 51. – № 8. – P. 1272–1282.

195. Dudchenko O., Batra S.S., Omer A.D., Nyquist S.K., Hoeger M., et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds // Science (New York, N.Y.). – 2017. – V. 356. – № 6333. – P. 92–95.

196. Pondeville E., Puchot N., Lang M., Cherrier F., Schaffner F., Dauphin-Villemant C., Bischoff E., Bourgouin C. Evolution of sexually-transferred steroids and mating-induced phenotypes in Anopheles mosquitoes // Scientific Reports. – 2019. – V. 9. – № 1. – P. 4669.

197. Salhab A., Nordström K., Gasparoni G., Kattler K., Ebert P., et al. A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains // Genome Biology. – 2018. – V. 19. – № 1. – P. 150.

198. Filippova D., Patro R., Duggal G., Kingsford C. Identification of alternative topological domains in chromatin // Algorithms for Molecular Biology. – 2014. – V. 9. – № 1. – P. 14.

199. Weinreb C., Raphael B.J. Identification of hierarchical chromatin domains // Bioinformatics (Oxford, England). – 2016. – V. 32. – № 11. – P. 1601–1609.

200. Shin H., Shi Y., Dai C., Tjong H., Gong K., Alber F., Zhou X.J. TopDom: an efficient and deterministic method for identifying topological domains in genomes // Nucleic Acids Research. – 2016. – V. 44. – № 7. – P. e70.

201. Lévy-Leduc C., Delattre M., Mary-Huard T., Robin S. Two-dimensional segmentation for analyzing Hi-C data // Bioinformatics (Oxford, England). – 2014. – V. 30. – № 17. – P. i386-392.

202. Dali R., Blanchette M. A critical assessment of topologically associating domain prediction tools // Nucleic Acids Research. – 2017. – V. 45. – № 6. – P. 2994–3005.

203. Crane E., Bian Q., McCord R.P., Lajoie B.R., Wheeler B.S., Ralston E.J., Uzawa S., Dekker J., Meyer B.J. Condensin-Driven Remodeling of X-Chromosome Topology during Dosage Compensation // Nature. – 2015. – V. 523. – № 7559. – P. 240–244.

204. Durand N.C., Shamim M.S., Machol I., Rao S.S.P., Huntley M.H., Lander E.S., Aiden E.L. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments // Cell systems. – 2016. – V. 3. – № 1. – P. 95–98.

205. Giotis S., Robey R., Skinner N., Tomlinson C., Goodbourn S., Skinner M. Chicken interferome: avian interferon-stimulated genes identified by microarray and RNA-seq of primary chick embryo fibroblasts treated with a chicken type I interferon (IFN-α) // Veterinary Research. – 2016. – V. 47. – P. 75.

206. Jahan S., Xu W., He S., Gonzalez C., Delcuve G.P., Davie J.R. The chicken erythrocyte epigenome // Epigenetics & Chromatin. – 2016. – V. 9. – P. 19.

207. Trapnell C., Pachter L., Salzberg S.L. TopHat: discovering splice junctions with RNA-Seq // Bioinformatics (Oxford, England). – 2009. – V. 25. – № 9. – P. 1105–1111.

208. Trapnell C., Roberts A., Goff L., Pertea G., Kim D., et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks // Nature Protocols. – 2012. – V. 7. – № 3. – P. 562–578.

209. Gushchanskaya E.S., Artemov A.V., Ulyanov S.V., Logacheva M.D., Penin A.A., et al. The clustering of CpG islands may constitute an important determinant of the 3D organization of interphase chromosomes // Epigenetics. – 2014. – V. 9. – № 7. – P. 951–963.

210. Engström P., Fredman D., Lenhard B. Ancora: A web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes // Genome biology. – 2008. – V. 9. – P. R34.

211. Levy A., Sela N., Ast G. TranspoGene and microTranspoGene: Transposed elements influence on the transcriptome of seven vertebrates and invertebrates // Nucleic acids research. – 2008. – V. 36. – P. D47-52.

212. Meilă M. Comparing Clusterings by the Variation of Information // Learning Theory and Kernel Machines / ed. Schölkopf B., Warmuth M.K. Berlin, Heidelberg: Springer, – 2003. – P. 173–187.

213. Giraldo-Calderón G.I., Emrich S.J., MacCallum R.M., Maslen G., Dialynas E., et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases // Nucleic Acids Research. – 2015. – V. 43. – № Database issue. – P. D707-713.

214. Smith E.M., Lajoie B.R., Jain G., Dekker J. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus // American Journal of Human Genetics. – 2016. – V. 98. – № 1. – P. 185–201.

215. Lazaris C., Kelly S., Ntziachristos P., Aifantis I., Tsirigos A. HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking // BMC Genomics. – 2017. – V. 18. – P. 22.

216. Robert S. Harris Improved pairwise alignment of genomic dna // A Thesis in Computer Science and Engineering. – 2007.

217. Flynn J.M., Hubley R., Goubert C., Rosen J., Clark A.G., Feschotte C., Smit A.F. RepeatModeler2 for automated genomic discovery of transposable element families // Proceedings of the National Academy of Sciences. National Academy of Sciences, – 2020. – V. 117. – № 17. – P. 9451–9457.

218. Dixon J.R., Jung I., Selvaraj S., Shen Y., Antosiewicz-Bourget J.E., et al. Chromatin Architecture Reorganization during Stem Cell Differentiation // Nature. – 2015. – V. 518. – № 7539. – P. 331–336.

219. Rowley M.J., Nichols M.H., Lyu X., Ando-Kuri M., Rivera I.S.M., Hermetz K., Wang P., Ruan Y., Corces V.G. Evolutionarily Conserved Principles Predict 3D Chromatin Organization // Molecular Cell. – 2017. – V. 67. – № 5. – P. 837-852.e7.

220. Harmston N., Ing-Simmons E., Tan G., Perry M., Merkenschlager M., Lenhard B. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation: 1 // Nature Communications. Nature Publishing Group, – 2017. – V. 8. – № 1. – P. 441.

ПРИЛОЖЕНИЕ 1

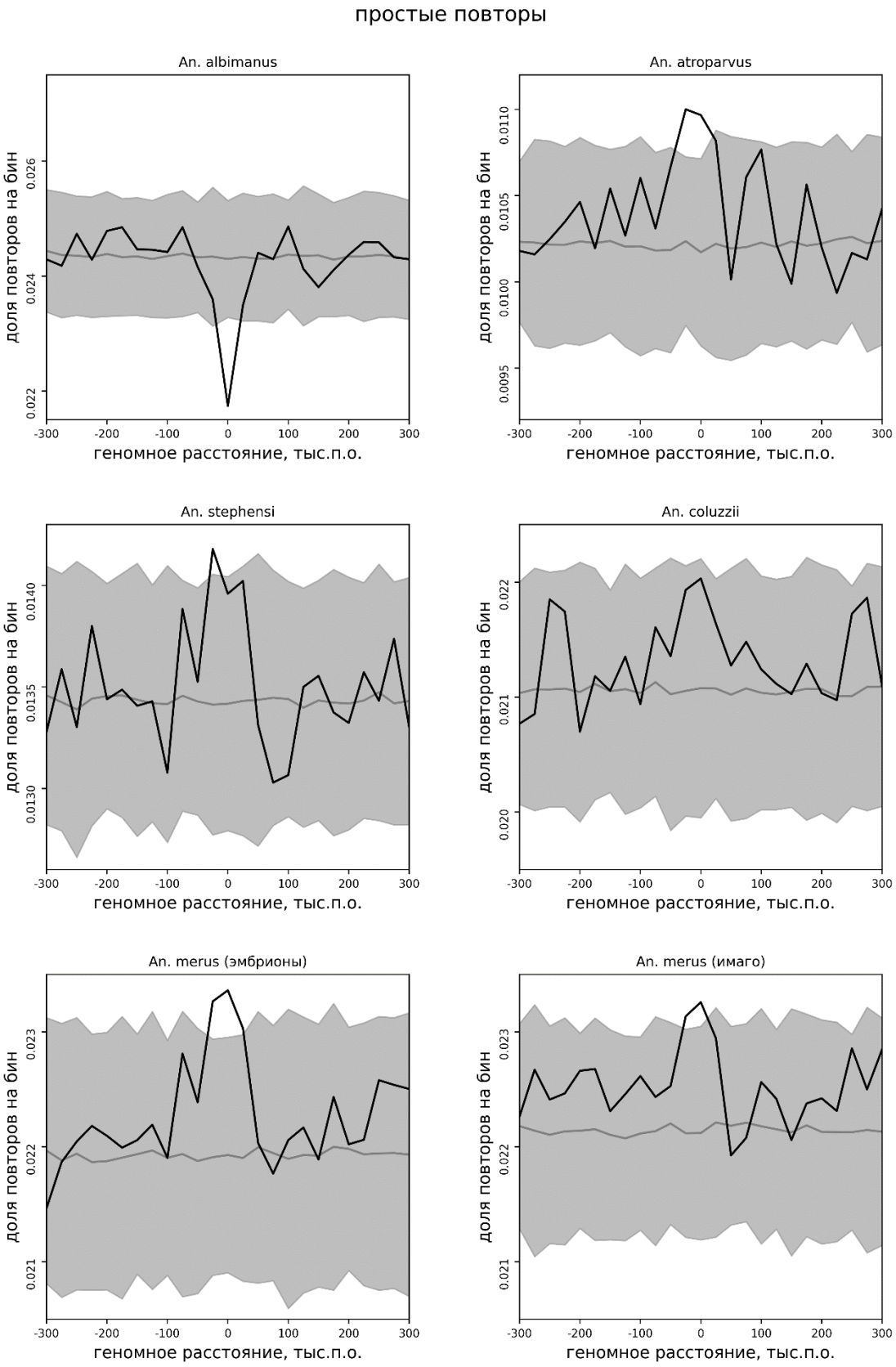

Рисунок 1. Распределение простых повторов относительно границ доменов для разных видов комаров рода *Anopheles*. Серая линия – среднее значение. Серая область показывает зону трёх дисперсий.

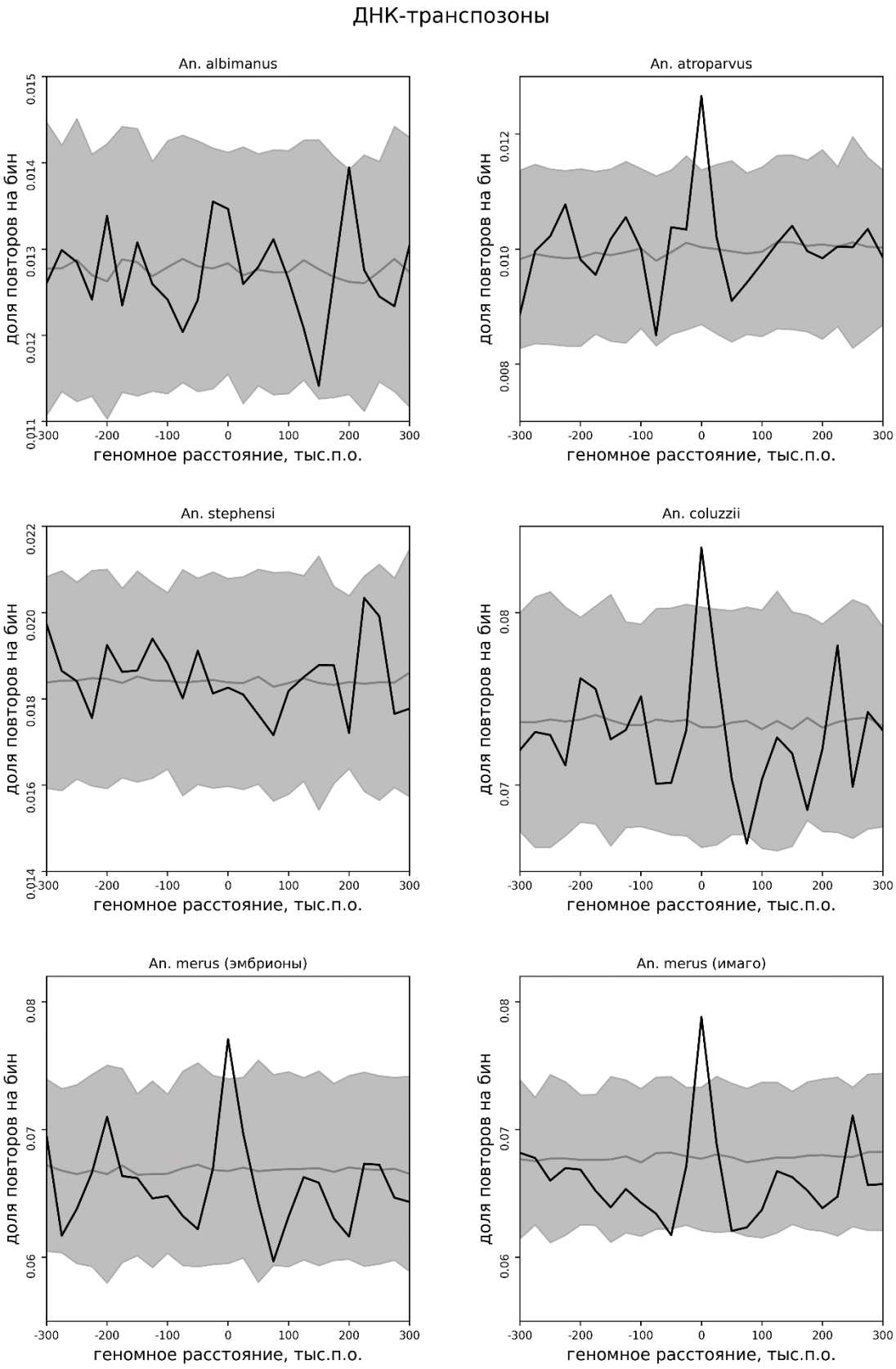

Рисунок 2. Распределение ДНК-транспозонов относительно границ доменов для разных видов комаров рода *Anopheles*. Серая линия – среднее значение. Серая область показывает зону трёх дисперсий.

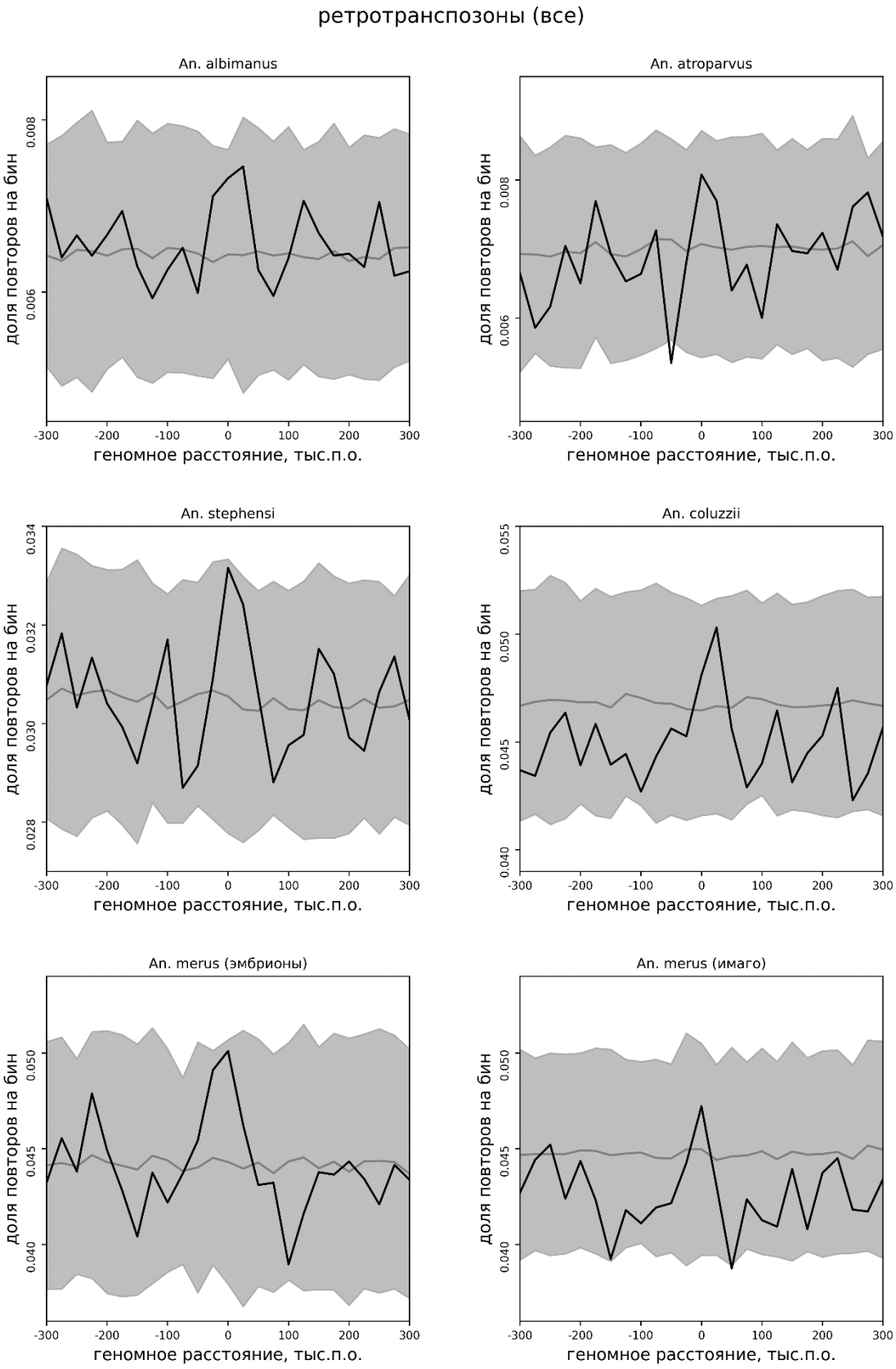

Рисунок 3. Распределение ретротранспозонов всех классов относительно границ доменов для разных видов комаров рода *Anopheles*. Серая линия – среднее значение. Серая область показывает зону трёх дисперсий.

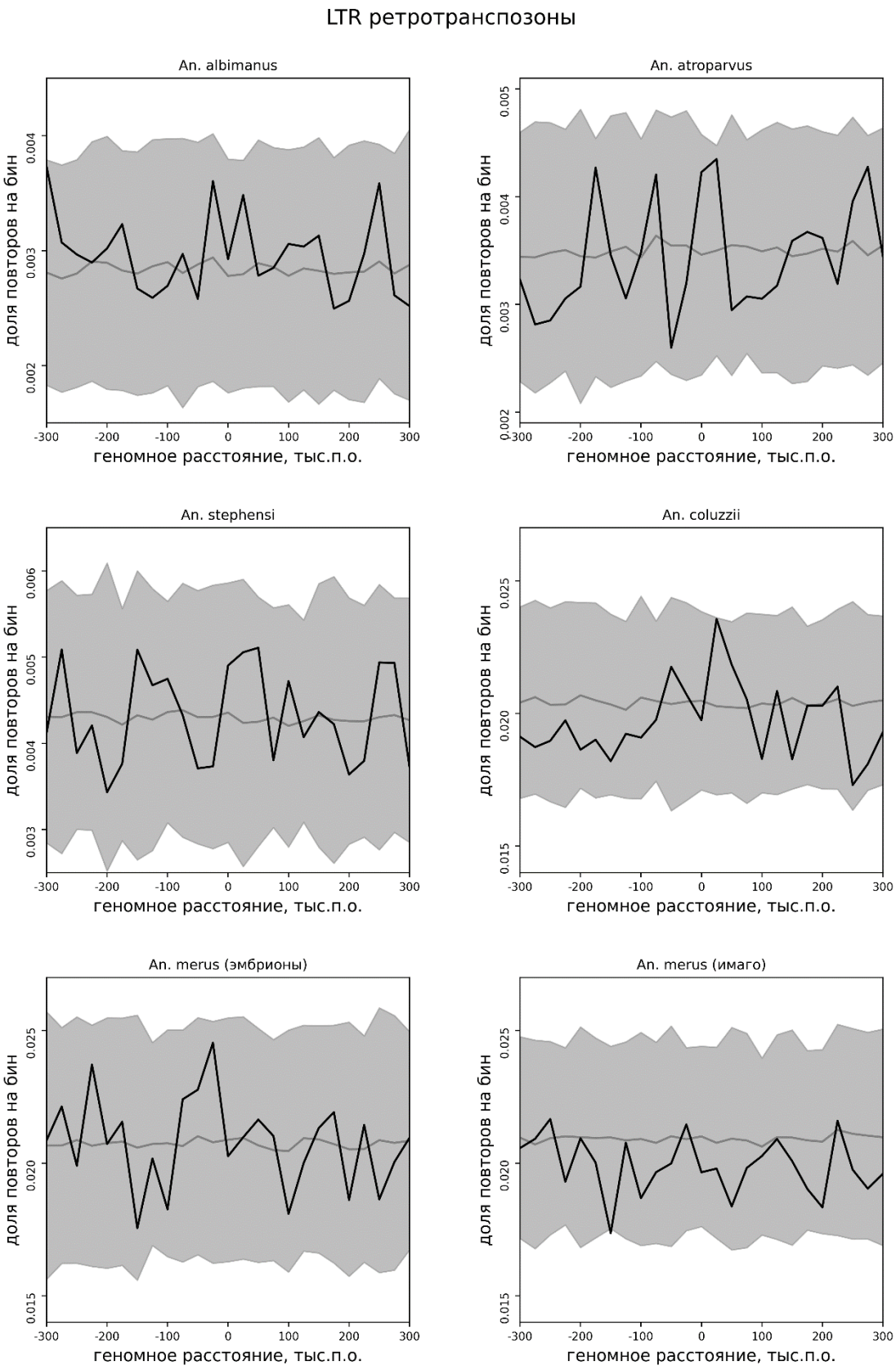

Рисунок 4. Распределение LTR ретротранспозонов относительно границ доменов для разных видов комаров рода *Anopheles*. Серая линия – среднее значение. Серая область показывает зону трёх дисперсий.

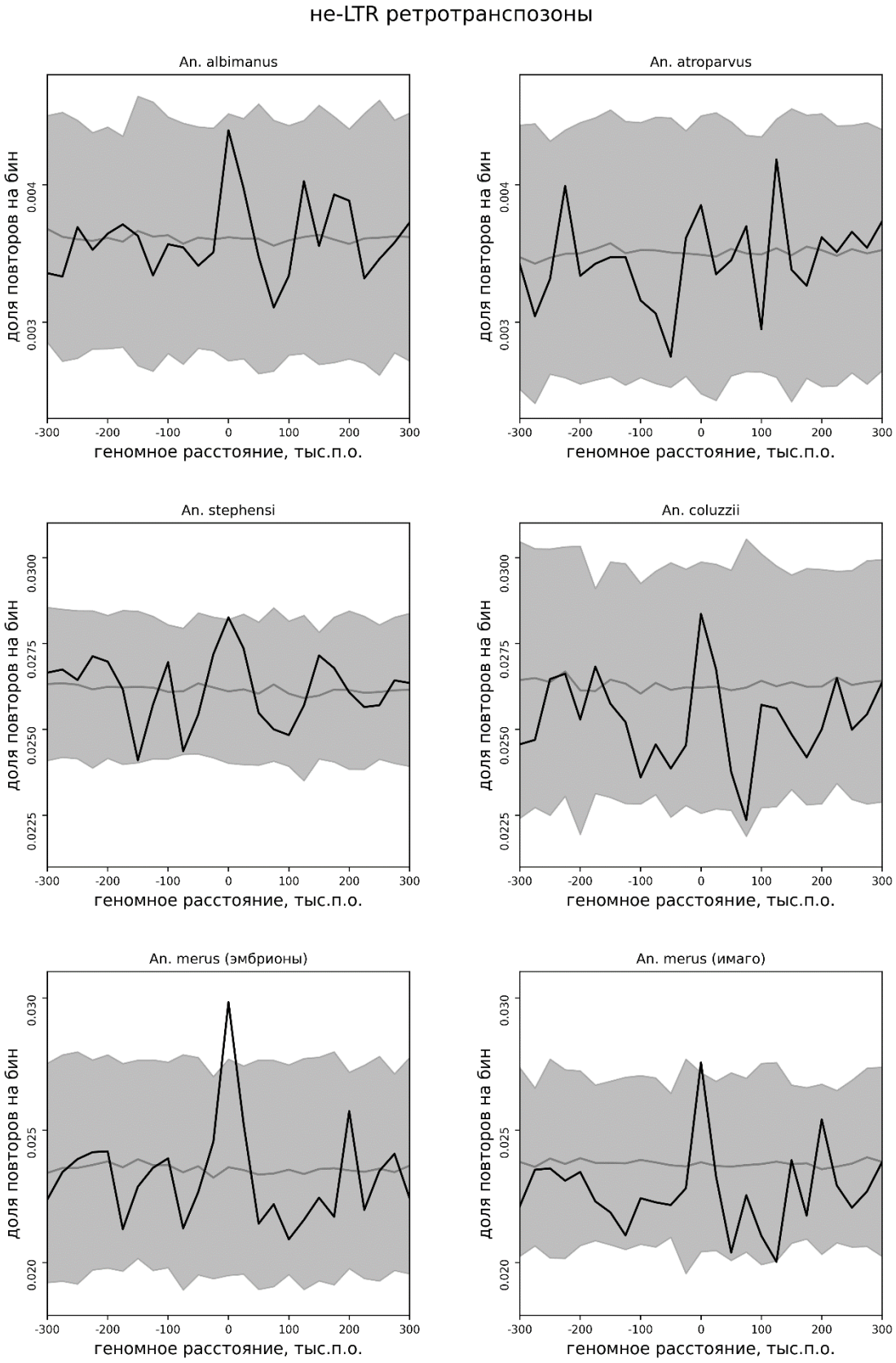

Рисунок 5. Распределение не-LTR ретротранспозонов относительно границ доменов для разных видов комаров рода *Anopheles*. Серая линия – среднее значение. Серая область показывает зону трёх дисперсий.

ПРИЛОЖЕНИЕ 2

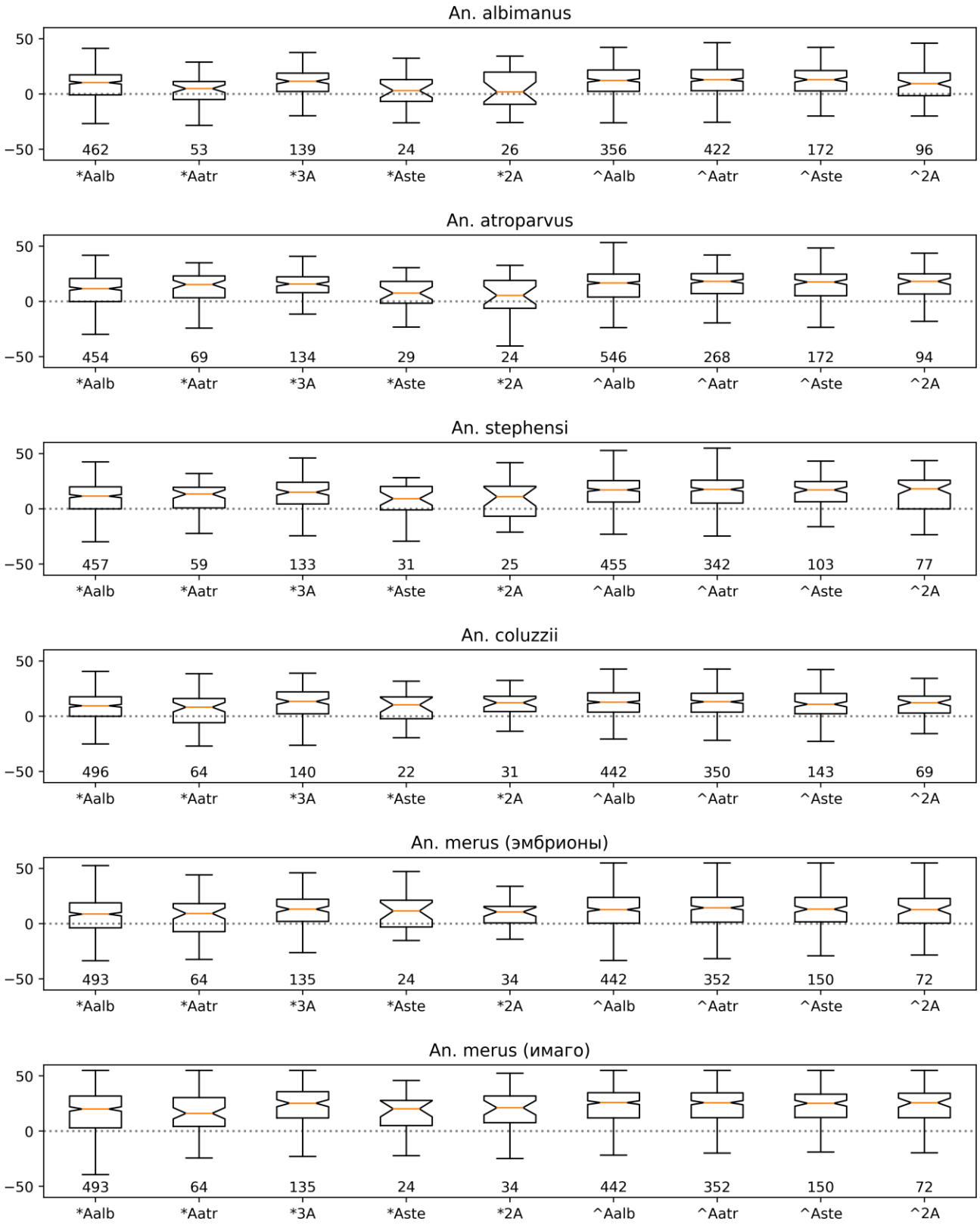

Рисунок 1. Значение компартмента в районах эволюционных точек разрыва хромосом, в зависимости от происхождения точки разрыва. Номенклатура точек разрыва по Рисунку 26. Точки разрыва уникальные для данной эволюционные линии отмечены «*», использованные повторно - «^». Среднее значение для всего генома обозначено пунктирной линией.

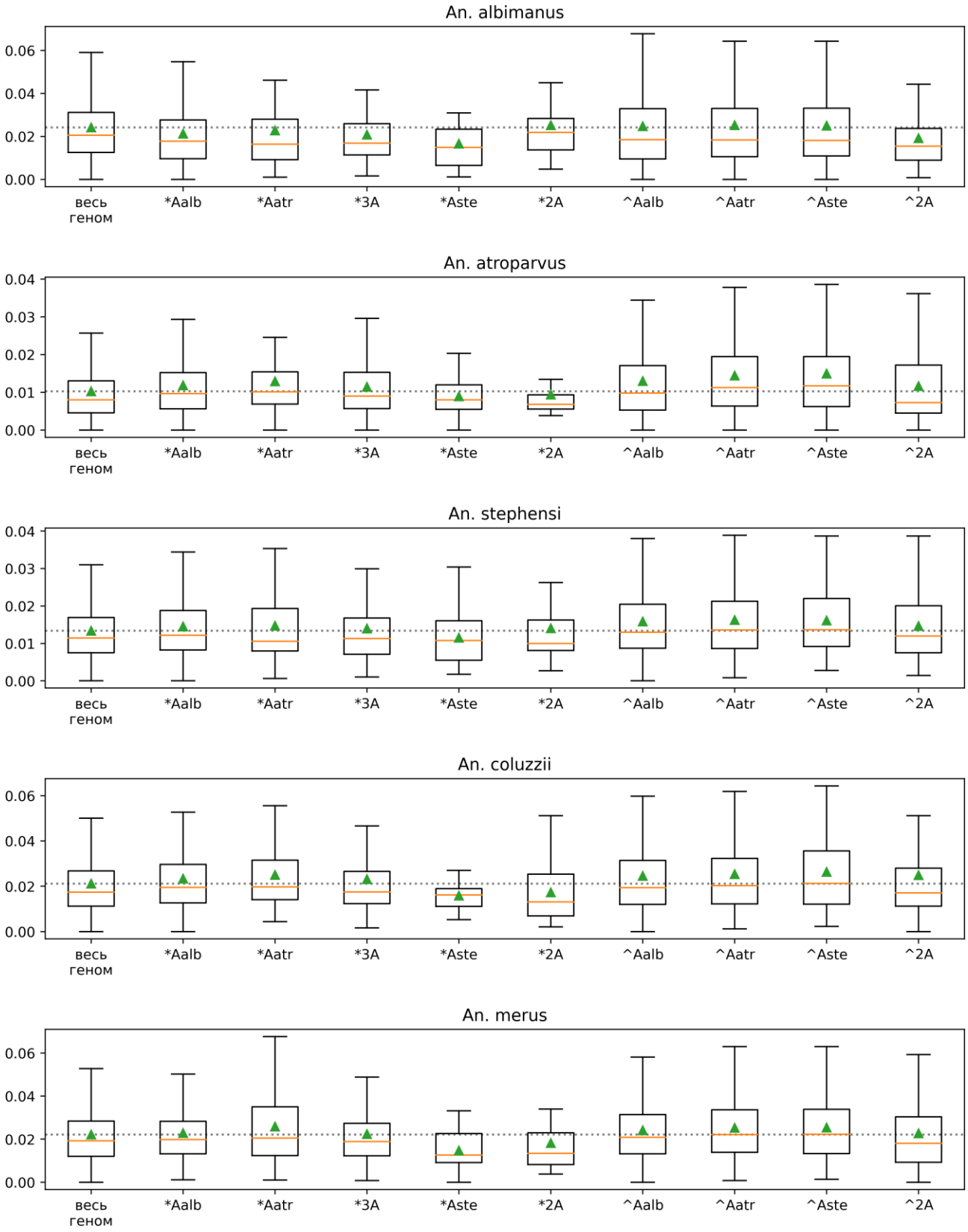

Рисунок 2. Насыщенность простыми повторами эволюционных точек разрыва хромосом, в зависимости от происхождения точки разрыва. Номенклатура точек разрыва по Рисунку 26. Точки разрыва уникальные для данной эволюционные линии отмечены «*», использованные повторно - «^». Среднее значение для всего генома обозначено пунктирной линией.

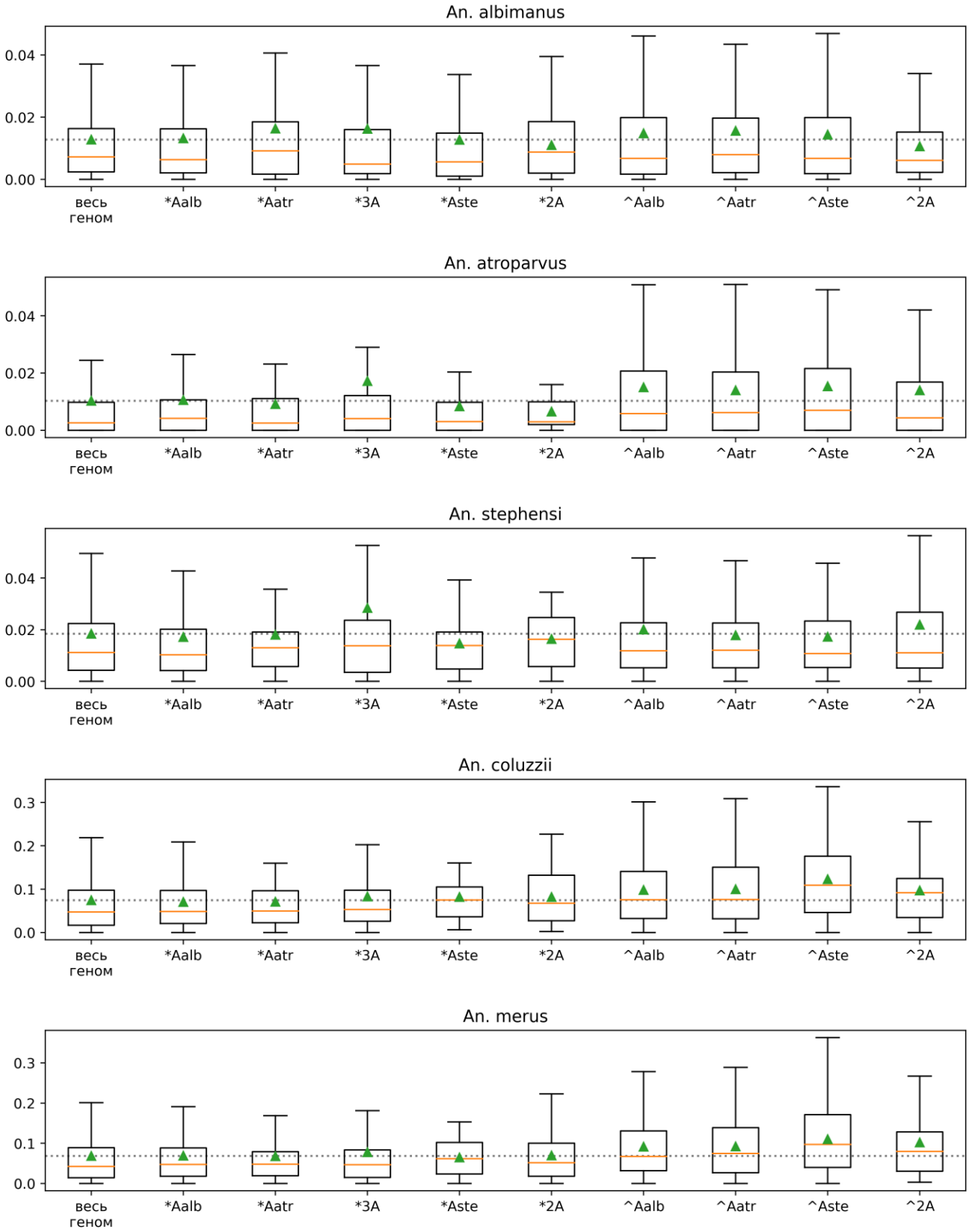

Рисунок 3. Насыщенность ДНК-транспозонами эволюционных точек разрыва хромосом, в зависимости от происхождения точки разрыва. Номенклатура точек разрыва по Рисунку 26. Точки разрыва уникальные для данной эволюционные линии отмечены «*», использованные повторно - «^». Среднее значение для всего генома обозначено пунктирной линией.

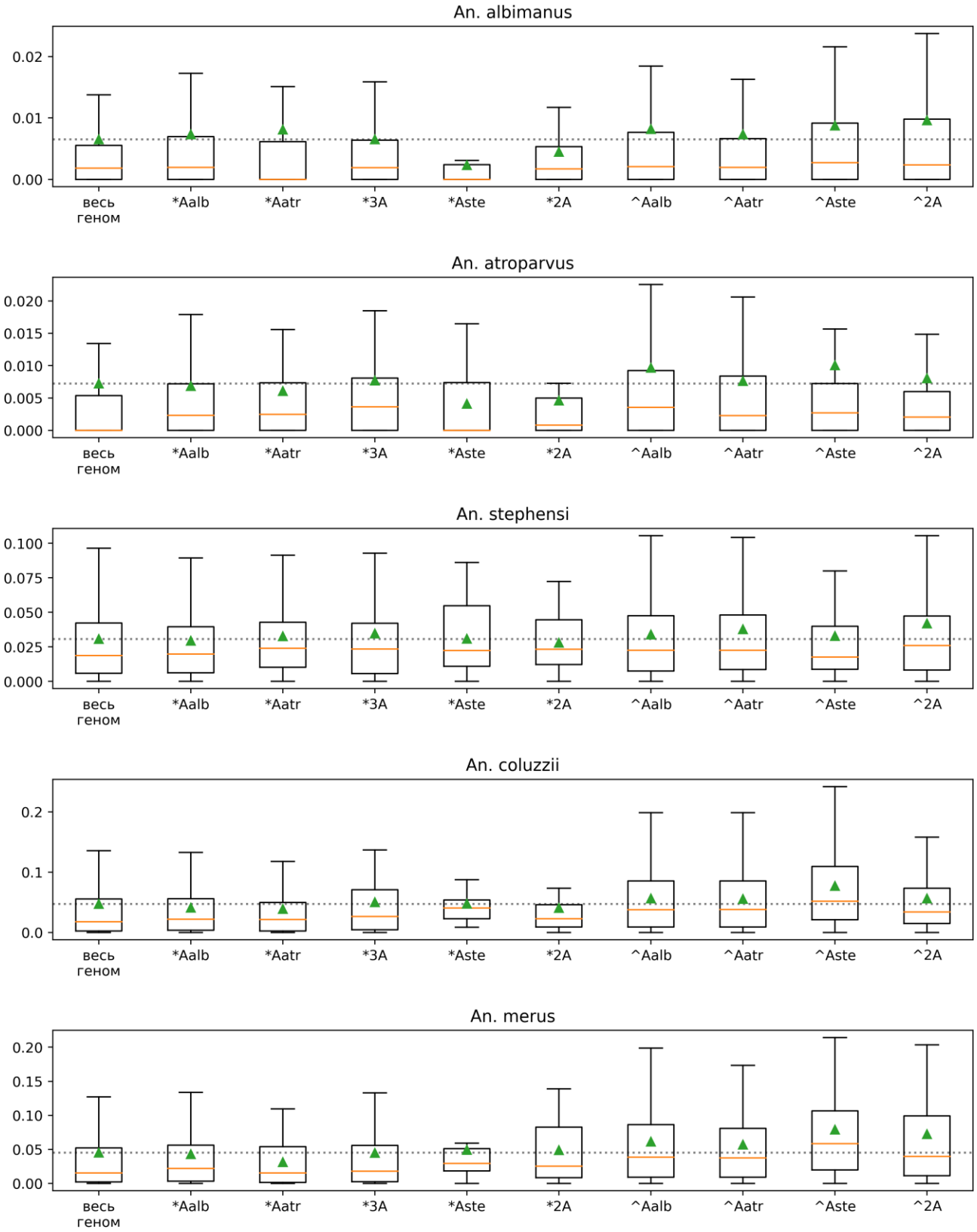

Рисунок 4. Насыщенность ретротранспозонами эволюционных точек разрыва хромосом, в зависимости от происхождения точки разрыва. Номенклатура точек разрыва по Рисунку 26. Точки разрыва уникальные для данной эволюционные линии отмечены «*», использованные повторно - «^». Среднее значение для всего генома обозначено пунктирной линией.

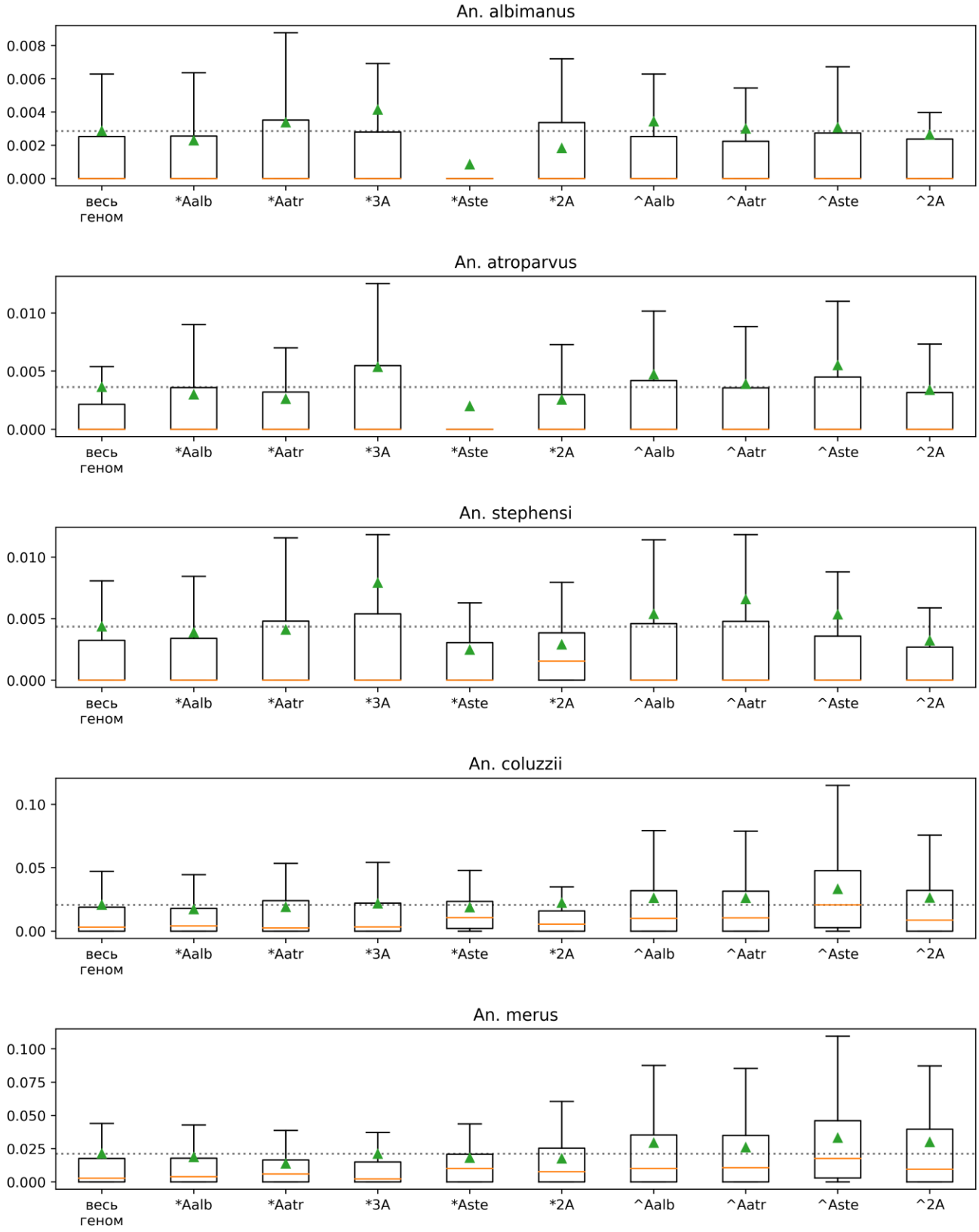

Рисунок 5. Насыщенность LTR ретротранспозонами эволюционных точек разрыва хромосом, в зависимости от происхождения точки разрыва. Номенклатура точек разрыва по Рисунку 26. Точки разрыва уникальные для данной эволюционные линии отмечены «*», использованные повторно - «^». Среднее значение для всего генома обозначено пунктирной линией.

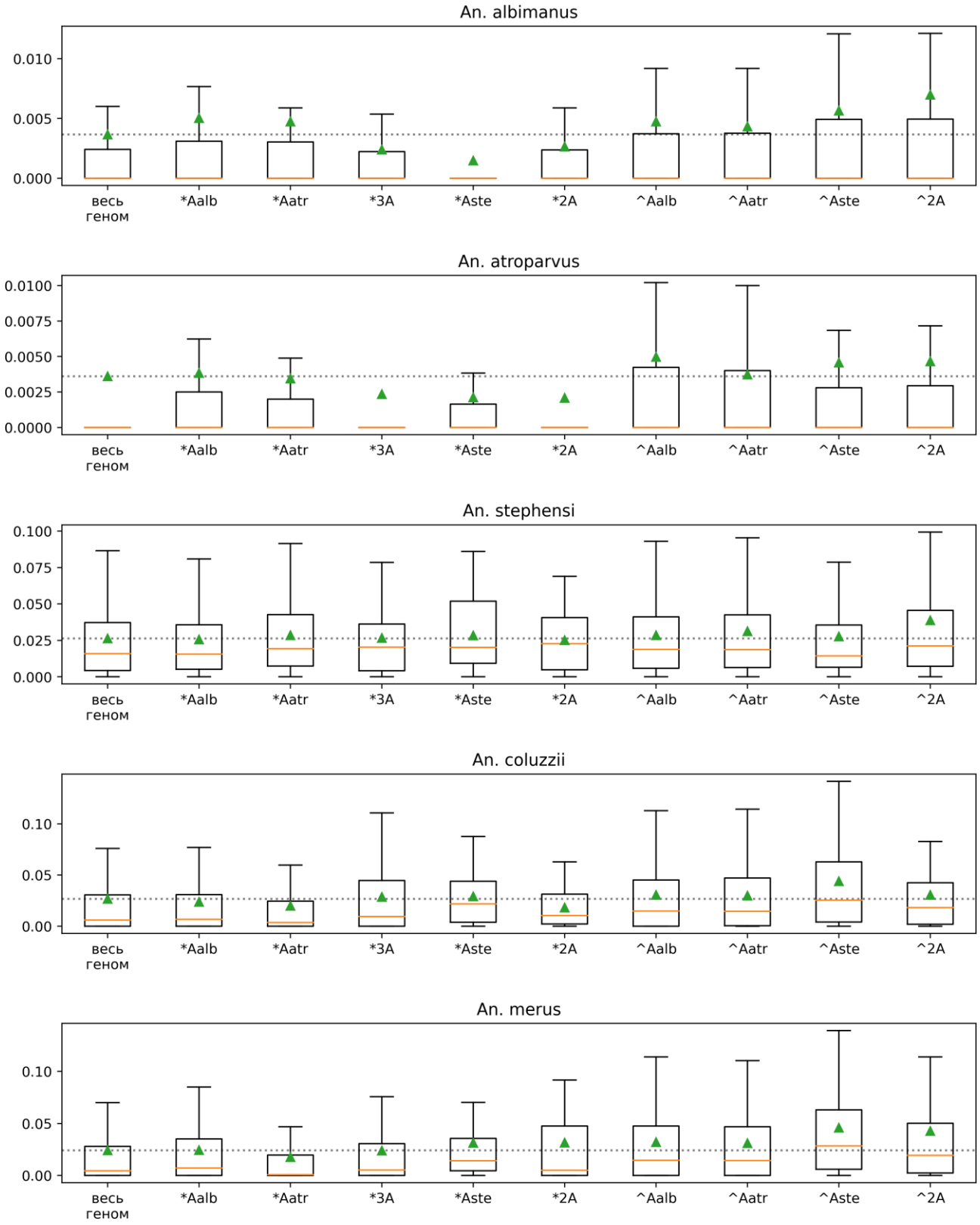

Рисунок 6. Насыщенность не-LTR ретротранспозонами эволюционных точек разрыва хромосом, в зависимости от происхождения точки разрыва. Номенклатура точек разрыва по Рисунку 26. Точки разрыва уникальные для данной эволюционные линии отмечены «*», использованные повторно - «^». Среднее значение для всего генома обозначено пунктирной линией.