

НУРИДДИНОВ МИРОСЛАВ АБДУРАХИМОВИЧ

**РАЗРАБОТКА МЕТОДОВ ДЛЯ МЕЖВИДОВОГО
СРАВНЕНИЯ ПРОСТРАНСТВЕННОЙ ОРГАНИЗАЦИИ
ХРОМАТИНА**

1.5.8. – математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата биологических наук

Новосибирск – 2023

Работа выполнена в отделе молекулярных механизмов онтогенеза Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

Научный руководитель: **Фишман Вениамин Семенович**
кандидата биологических наук, зав. сектором геномных механизмов онтогенеза ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики СО РАН», г. Новосибирск

Официальные оппоненты: **Макеев Всеволод Юрьевич**
доктор физико-математических наук, профессор, член-корреспондент РАН, заведующий лабораторией системной биологии и вычислительной генетики ФГБУН «Институт общей генетики им. Н. И. Вавилова РАН», г. Москва

Омелина Евгения Сергеевна
кандидат биологических наук, заведующая лабораторией клеточного деления ФГБУН «Институт молекулярной и клеточной биологии СО РАН», г. Новосибирск

Ведущая организация: ФГБУН «Институт химической биологии и фундаментальной медицины СО РАН», г. Новосибирск

Защита диссертации состоится «___» _____ 20__ г. на утреннем заседании диссертационного совета 24.1.239.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» в конференц-зале Института по адресу:
пр. ак. Лаврентьева 10, г. Новосибирск, 630090, +7 (383) 3634906, факс +7 (383) 3331278, e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института <http://www.icgbio.ru>

Автореферат разослан «___» _____ 2023 г.

Ученый секретарь
диссертационного совета
доктор биологических наук

Т.М. Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Успехи в развитии методик семейства захвата конформации хромосом, позволило получать полногеномные данные о архитектуре хроматина и открыть новые уровни организации хроматина в пространстве клеточного ядра. Исследования показали, что архитектура хроматина тесно связана с регуляцией генной экспрессии.

Основной массив данных по архитектуре хроматина среди позвоночных получены на клеточных линиях *Homo sapiens* и *Mus musculus*, также большой объём сведений имеется для *Drosophila melanogaster*. Ввиду того, что эволюционные линии млекопитающих и насекомых весьма далеки друг от друга, а другие таксономические группы практически не изучены, остаётся существенный пробел в понимании значимости известных особенностей архитектуры хроматина.

В соответствии с этим большой интерес представляют результаты Hi-C эксперимента, проведённого на разных типах клеток *Gallus gallus*. Данный вид птиц является хорошо изученным модельным объектом и сравнение архитектуры хроматина *G. gallus* с архитектурой хроматина других видов позвоночных способно дополнить наши знания о взаимосвязи архитектуры хроматина с регуляции генной экспрессии. За пределами позвоночных животных большая часть сведений об архитектуре хроматина получены на отдельных, эволюционно далёких, представителях отряда *Diptera*. Таким образом отсутствует систематические представления об эволюции архитектуры хроматина в отряде *Diptera*.

Необходимо добавить, что с каждым годом появляется всё больше данных, описывающих архитектуру хроматина в разных типах клеток в разных таксономических группах, при этом биоинформатических методов, позволяющих проводить межвидовое сравнение архитектуры хроматина решительно не хватает. Таким образом, на данный момент существует необходимость в разработке методов изучения архитектуры хроматина, которые позволили бы проводить эволюционное сравнение.

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 работы. В впервые проанализированы данные Hi-C эксперимента для разных типов клеток *G. gallus* и проведено сравнение топологически ассоциированных доменов птиц и млекопитающих в полногеномном масштабе. Полученные результаты позволили выявить черты организации хроматина как уникальные для *G. gallus*, так и резко отличающие её 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 в дальнейшем изучении генетики этих организмов.

Практическая и теоретическая значимость работы. Разработанный биоинформатический метод межвидового сравнения архитектуры хроматина на уровне отдельных контактов является первым методом такого рода, реализованный в виде пакета программ C-InterSecture. Изучена проблема математического определения эволюционной консервативности пространственной организации хроматина. Использование пакета программ C-InterSecture представляется перспективным как дополнительное средство в изучении механизмов эволюции, регуляции генной экспрессии и прояснение взаимосвязи 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 находится в открытом доступе (<https://github.com/NuriddinovMA/C-InterSecure>).

Вклад автора. Основные результаты, изложенные в диссертации, получены лично соискателем: разработаны биоинформатические алгоритмы сравнения архитектуры хроматина, проведена их апробация с использованием результатов Hi-C экспериментов. Экспериментальные данные Hi-C для *G. gallus* были предоставлены канд. биол. наук Баттулиным Н.Р., для комаров рода *Anopheles* – д-р. биол. наук Шараховым И. В. и Лукьянчиковой В. А.

Структура и объём работы. Диссертация изложена на 149 страницах, содержит 29 рисунков и 6 таблиц. Список литературы включает 218 ссылок. Диссертация состоит из введения, литературного обзора, описания результатов в двух главах, заключения, выводов и списка литературных источников.

МАТЕРИАЛЫ И МЕТОДЫ

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. Для конвертирования геномных координат использовался ресурс LiftOver (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>). Данные экспериментов Hi-C обрабатывались с помощью пакета программ JuicerTools (Durand et al., 2016c) и визуализировались с помощью Juicebox (Durand et al., 2016a).

Идентификация доменов. Для каждой исследуемой линии клеток было сгенерировано три разных набора доменов с использованием следующих биоинформатических алгоритмов: DomainCaller (Dixon et al., 2012), Armatus (Filippova et al., 2014), TADtree (Weinreb, Raphael, 2016), TopDom (Shin et al., 2016), HiCseg (Lévy-Leduc et al., 2014).

Выделение А/В-компарментов. Для определения принадлежности локусов к А- или В-компарменту использовалось стандартное программное обеспечение, включённое в пакет JuicerTools (Durand et al., 2016b).

Определение геномных координат и экспрессии генов. Источником координаты генов для позвоночных и списков генов-ортологов являлась база данных Ensemble. Для анализа уровня генной экспрессии использовались данные RNA-seq: эмбриональные фибробласты (Giotis et al., 2016) и незрелые эритроциты (Jahan et al., 2016).

Некодирующие элементы генома и эпигенетические метки. В исследовании были использованы те сайты связывания CTCF, которые являются общими для разных клеточных линий *G. gallus*: эмбриональных эритроцитов 5 и 10 дня (Martin et al., 2011) и клеточной линии HD3 (Gushchanskaya et al., 2014). Геномное

положение консервативных некодирующих элементов (CNE) были взяты из базы данных Ancora (<http://ancora.genereg.net/>) (Engström, Fredman, Lenhard, 2008). Координаты мобильных элементов генома были взяты из базы данных TranspoGene (<http://transpogene.tau.ac.il/>) (Levy, Sela, Ast, 2008). Геномные координаты F1-хроматиновых доменов, H3K4me3 и H3K27ac пиков были взяты по данным для незрелых эритроцитов (Jahan et al., 2016).

Определение регионов эволюционных перестроек хромосом. Координаты регионов эволюционных межхромосомных перестроек определялись согласно литературным данным (Farré et al., 2016).

Сравнение архитектуры хроматина у разных видов с помощью индекса вариации информации (VI). Сравнения пространственной организации хроматина у разных видов проводилось с помощью предварительно адаптированной метрики «индекс вариации информации» (VI) (Meilă, 2003).

Анализ архитектуры хроматина комаров рода *Anopheles*. В исследовании были проанализирована пространственная организация хроматина для видов *An. albimanus*, *An. atroparvus*, *An. stephansi*, *An. coluzzii* и *An. merus*. Данные экспериментов Hi-C для комаров рода *Anopheles* обрабатывали с помощью программного обеспечения Juicertools (Durand et al., 2016c). Для проведения анализа все геномы предварительно улучшались или собирались *de novo* с помощью программного обеспечения 3D-DNA (Dudchenko et al., 2017). Нуклеотидные последовательности геномных скаффолдов была получена из открытой базы данных VectorBase (Giraldo-Calderón et al., 2015).

Описание архитектуры хроматина. Для расчёта A/B-компартов проводился с использованием алгоритма входящего в программный пакет Juicertools (Durand et al., 2016b). Расчёт силы компарментализации проводился по методике, предложенной Nora E.P. с коллегами (Nora et al., 2017). Расчёт индекса инсуляции проводился на основе силы частоты контактов, окружающих целевой бин локусов (Crane et al., 2015). Домены были выделены с помощью программного обеспечения hicExplorer (Ramírez et al., 2018) и любезно предоставлены Таскиной А.К.

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 (Giraldo-Calderón et al., 2015), которые пересчитывались на координаты улучшенных сборок геномов. Для геномов *An. coluzzii* и *An. merus* определение предполагаемых позиций генов проводилось выравниванием нуклеотидные последовательности экзонов, полученных для *An. coluzzii* MOPTI 1.8 и *An. merus* MAF 2.9 соответственно с помощью LastZ (Robert S. Harris, 2007). Аннотация повторов *de novo* проводилась с помощью конвейера программ RepeatModeler (Flynn et al., 2020).

Конвертирование геномных координат между разными видами. Для видов *H. sapiens*, *M. musculus* и *G. gallus* конвертирования геномных координат проводилось с карт синтении, представленных net-файлами, полученными из базы данных UCSC. Для построения карт синтении для остальных видов последовательно использовались программное обеспечение LastZ, для попарного полногеномного выравнивания, и программные пакеты LAST и kentUtils версии 3.02 для конвертирования попарных выравниваний в net-файлы.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Характеристика пространственной организации генома *G. gallus*. Анализ результатов работы биоинформатических алгоритмов для поиска топологически ассоциированных доменов (ТАДов) показывает, что характерные размеры доменов в клетках *G. gallus* близки к тем, что указаны для других позвоночных (Dixon et al., 2012; Dixon et al., 2015; Rao et al., 2014). Домены, выделенные с помощью алгоритмов Armatus и TADtree, имеют медианную длину около 200-250 тысяч п.о, а домены, выделенные 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. Совпадение доменной организации фибробластов и эритроцитов/эритробластов составляет от 10% до 50%, в зависимости от используемого алгоритма, сходство же для клеток эритроидного ряда составляет около 55-65% в независимости от избранных алгоритмов. В соответствии с этим возникает вопрос, аналогичны ли домены, выделенные на разных типах клеток *G. gallus* ТАДам у млекопитающих.

В первую очередь проверялось 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 1) не зависимо от используемых алгоритмов выделения ТАДов. При этом домены формируются преимущественно конвергентно направленными сайтами связывания CTCF. Для эритробластов эта закономерность выражена слабее и практически отсутствует у эритроцитов.

Сравнения между разными клеточными линиями *G. gallus* распределения других характеристик генома - доли повторов, консервативных некодирующих элементов, генной экспрессии и модификаций гистонов - относительно границ доменов также демонстрирует кардинальные различия между фибробластами и клетками

эритроидного ряда, при этом для архитектуры фибробластов характерны те же закономерности, что известны для млекопитающих.

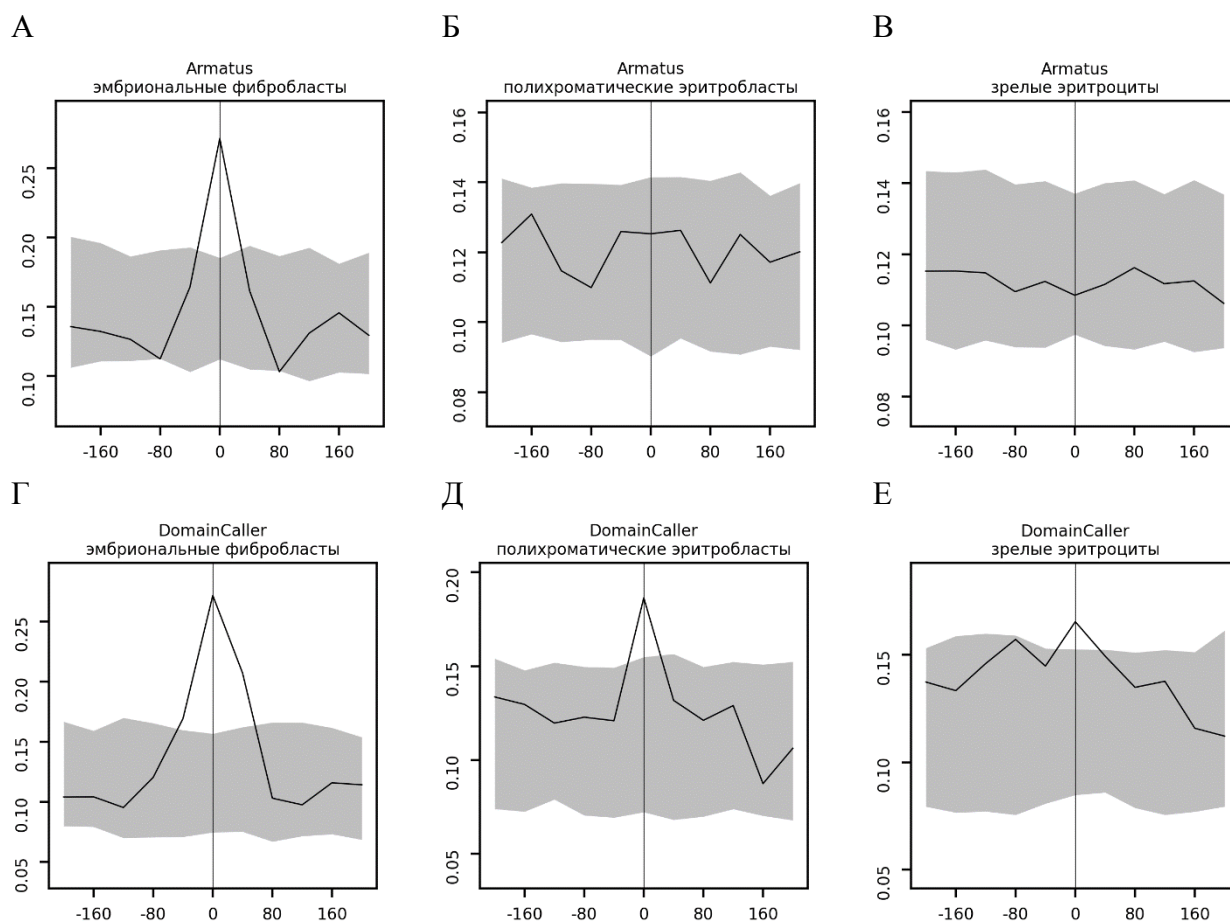


Рисунок 1. Распределение сайтов связывания CTCF относительно границ доменов. Результаты показаны для границ доменов, выделенных алгоритмами DomainCaller и Armatus. По оси X расстояние в тысячах пар оснований от границы, по оси Y — среднее число сайтов связывания CTCF на 40 тысяч пар оснований. Чёрная линия – наблюдаемые данные. Серая область – 3 стандартных отклонения от ожидаемого.

На основе полученных результатов было выдвинуто предположение, что границы доменов в эритроцитах и эритроблестах соответствуют районам смены компартмента. Проверка данного предположения показала, что у фибробластов *G. gallus* границы доменов проходят по произвольным местам и никакой зависимости между значением компартментализации соседних доменов нет (Рисунки 2А и 2Г). В тоже время, для зрелых и полихроматических эритроцитов показано, что соседние домены принадлежат разным компартментам (Рисунки 2Б, 2В, 2Д и 2Е). Это подтвердило гипотезу о том, что границы доменов в эритроблестах и эритроцитах соответствует районам смены компартмента.

Использования индекса VI для эволюционного сравнения организации хроматина. Сравнение архитектуры хроматина с помощью индекса VI между

разными типами клеток *G. gallus*, *H. sapiens*, *M. musculus* показало, что доменная организация хроматина фибробластов *G. gallus* характеризуется высоким сходством с той, что наблюдается у млекопитающих, независимо от применяемых алгоритмов (Рисунок 3). В то же время, сходство между доменной организацией хроматина эритроцитов и эритробластов *G. gallus* и всеми другими типами клеток близко к случайному.

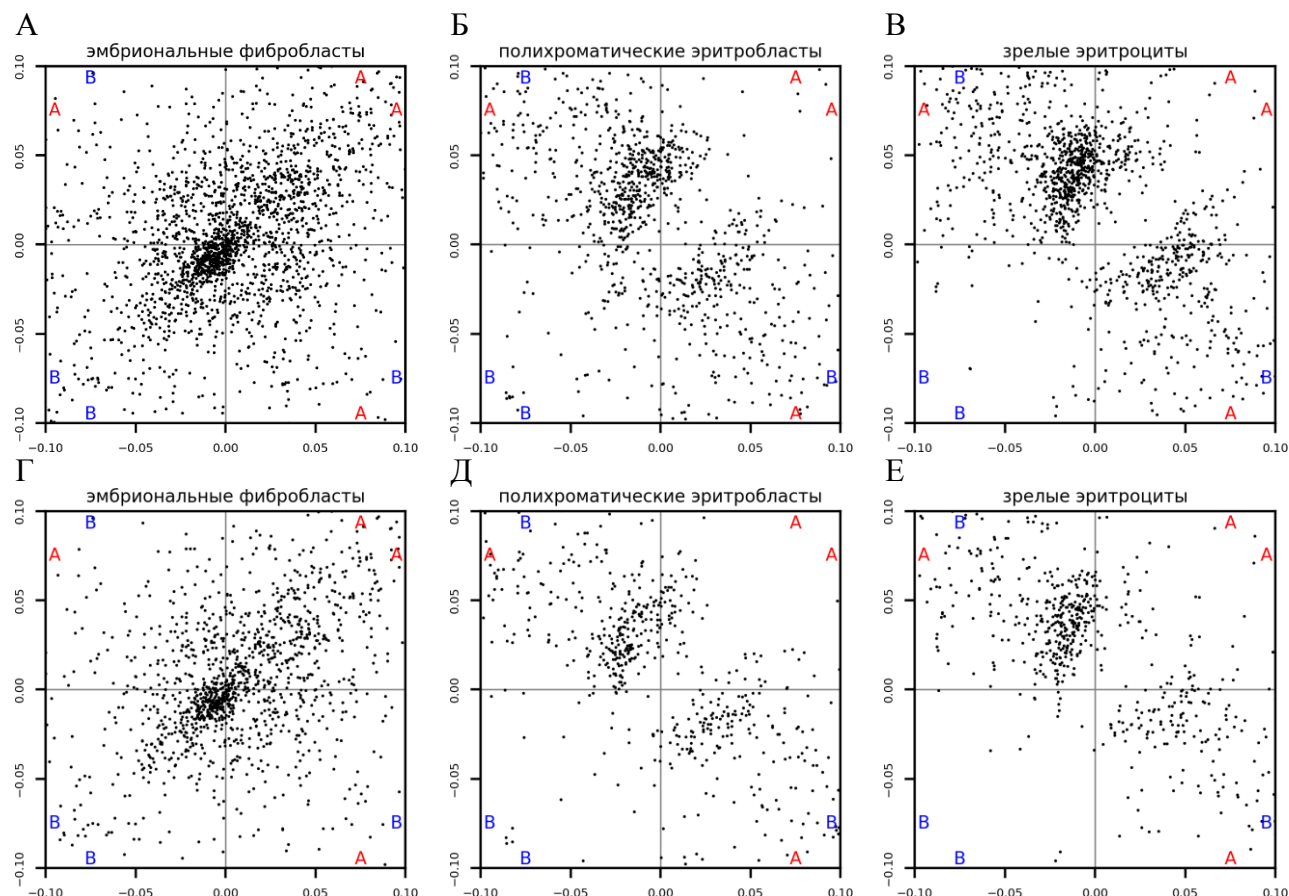


Рисунок 2. Значение первой главной компоненты Hi-C-матрицы в ближайших окрестностях границы домена. По оси X отложено значения внутри домена, по оси Y – снаружи домена. А-В. Домены выделены алгоритмом Armatus. Г-Е. Домены выделены DomainCaller.

Примечательно, что наблюдаемая величина сходства между доменами, выделенными разными алгоритмами, оказывается близкой к уровню межвидового сходства фибробластов *G. gallus* и млекопитающих, а величина сходства для доменов эритробластов/эритроцитов не отличается от случайного. Результаты эволюционного сравнения архитектуры хроматина показывают, что величина сходства/различия сравниваемых видов находятся под сильным влиянием применяемых алгоритмов. Таким образом, было решено разработать такой метод эволюционного сравнения архитектуры хроматина, который позволил бы напрямую сравнивать индивидуальные контакты хроматина исследуемых видов.

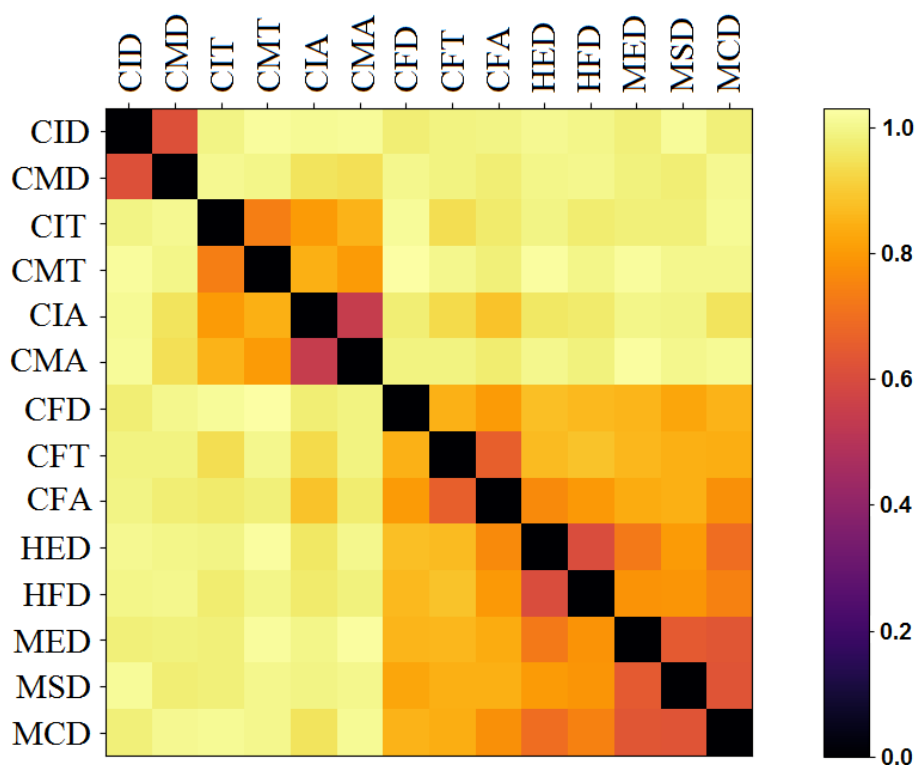


Рисунок 3. Значения индекса VI для сравниваемых клеточных линий. Чем меньше значением — тем более похожа архитектура хроматина. CFA, CFD, CFT – эмбриональные фибробласты *G. gallus*; CIA, CID, CIT – полихроматические эритроциты *G. gallus*, CMA, CMD, CMT — зрелые эритроциты *G. gallus*, HED — эмбриональные стволовые клетки *H. sapiens*, HFD – фибробласты IMR90, MED — эмбриональные стволовые клетки *M. musculus*, MSD — сперматозоиды *M. musculus*, MCD — клетки коры мозга *M. musculus*. - Последняя буква шифра отражает использованный алгоритм: A – Armatus, D — DomainCaller, T — TADtree.

Алгоритм сравнения пространственной организации хроматина, основанный на индивидуальных контактах. В основе биоинформатического алгоритма C-InterSecture (Computational tool for **InterSpecies** analysis of genome architecture) лежит математически определённый пересчёт численных величин контактов с генома одного вида на другой, с учётом уровня синтении генома, изменения его размеров и геномного расстояния между контактирующими локусами. Результаты сравнения видов геномов *G. gallus*, *H. sapiens* и *M. musculus* показали высокую консервативность архитектуры хроматина (Рисунок 4). При этом изменение частот контактов между синтенными локусами у сравниваемых видов происходит согласно изменению геномных расстояний между этими локусами (Рисунок 4 А, Б). Чтобы учесть эффект изменения геномного расстояния, было проведено сравнение, при котором вместо частоты контакта использовался перцентиль частоты контакта. Данная математическая величина отражает то, насколько сильно целевой контакт отличается от ожидаемого среди контактов всех

локусов генома, разделённых таким же геномным расстоянием. Сравнение показывает, что перцентили контактов сравниваемых видов совпадают в 1.5-2 раза чаще, чем ожидалось бы при случайном распределении (Рисунок 4 В, Г). При этом для контактов, входящих в 15% наиболее сильных (или слабых), доля близких или совпадающих перцентилей превышает случайное в 3-4 раза.

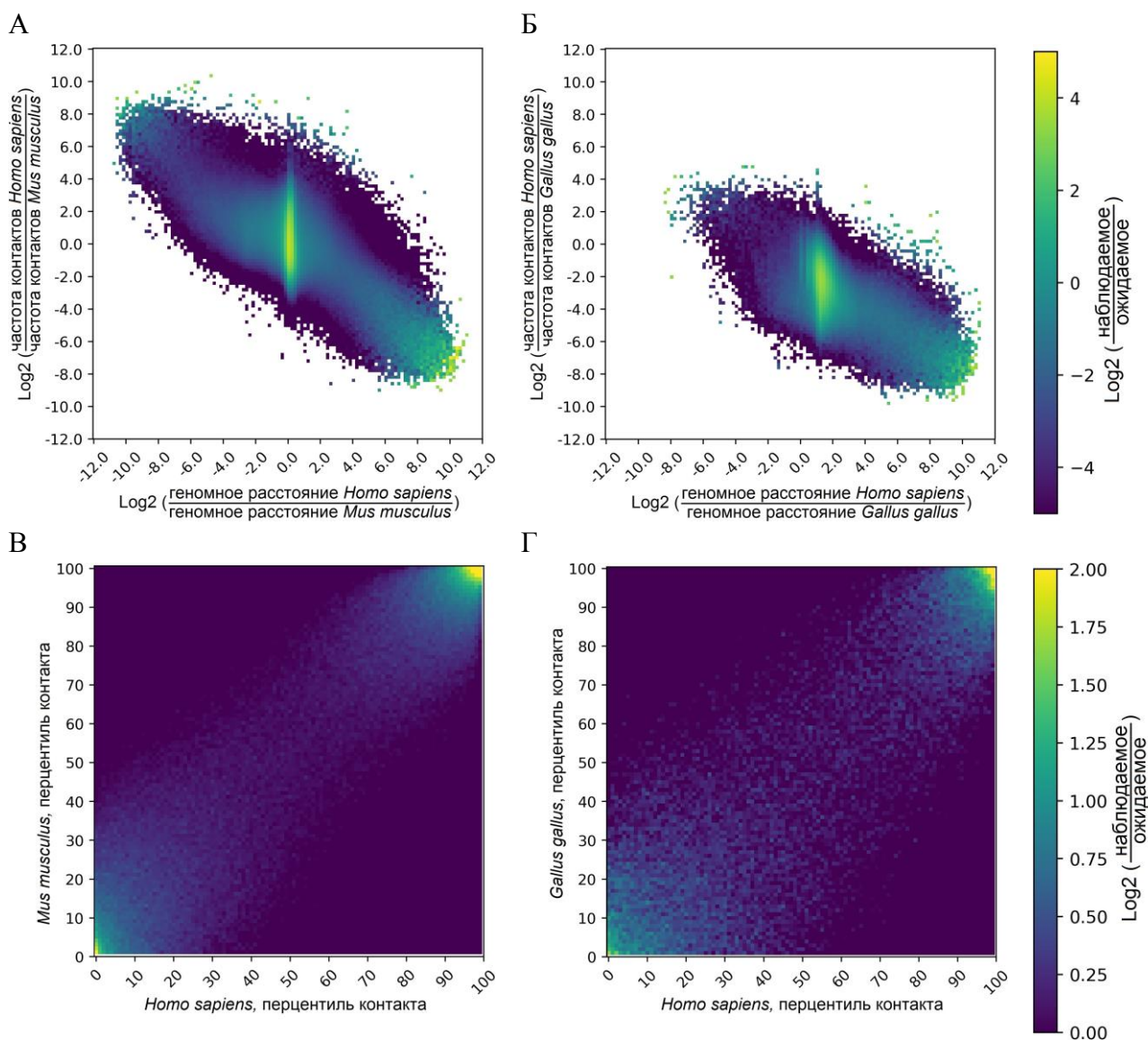


Рисунок 4. Изменение пространственных и геномных расстояний в ходе эволюции позвоночных. А. Зависимость изменения частот контактов между синтенными локусами у *H. sapiens* и *M. musculus* от изменения геномного расстояния. Б. Тоже, что и (А), для пары *H. sapiens* и *G. gallus*. В. Зависимость перцентиль контакта между синтенными локусами у *H. sapiens* и *M. musculus*. Г. Тоже, что и (В), для пары *H. sapiens* и *G. gallus*.

Для полногеномного поиска локусов, насыщенных отличающимися контактами у сравниваемых видов, была разработана мера P-BAD. Данная математическая величина оценивает различия в количестве контактов в целевом локусе у контрольной карты Hi-C и перекартированной. Исследование на пяти

видах позвоночных (*H. sapiens*, *Mc. mulatta*, *M. musculus*, *C. familiaris*, *G. gallus*) показало, что наиболее значимые различий наблюдаются в районах разрыва блоков синтении. Таким образом, сравнение архитектуры хроматина на уровне отдельных контактов показывает высокую её консервативность у всех взятых в рассмотрение позвоночных, при этом межвидовая разница оказывается сопоставимой с разницей между разными типами клеток.

Пространственной организации хроматина комаров рода *Anopheles*. Одним из наиболее важных механизмов формирования архитектуры хроматина у *Diptera* является компартиментализация генома (Hou et al., 2012; Ulianov et al., 2016). В соответствии с этим было проведён расчёт геномных компартиментов на основе данных Hi-C для взятых в исследовании комаров рода *Anopheles* с помощью стандартного алгоритма, входящего в пакет программ *juicer*tools. В соответствии с литературными данными, ожидалось, что разделение генома на компартименты будет соответствовать делению на блоки с высоким и низким уровнем экспрессии, насыщенным и обеднённым генами. Результаты проверки показывают, что в ряде случаев коэффициенты корреляции Пирсона между рассчитанными значениями компартиментов, плотностью генов и уровнем экспрессии оказались крайне низки (<0.2), а визуальный анализ полученного трека компартиментов показал, что он не соответствует «клетчатому» паттерну, наблюдаемому на Hi-C-картах.

В соответствии с этим, был разработан биоинформатический алгоритм АВСЕ (A/B-compartment contrast enhancement), который позволяет учесть влияние ориентации хромосом по Рабблю и блоков интеркалярного, прицентромерного и прителомерного гетерохроматина на рассчитываемое по данным Hi-C значение компартамента. Полученные значения корреляции по Пирсона ($\sim 0.4-0.7$) величины компартамента с плотностью генов и уровнем экспрессии соответствовали значениям, наблюдаемым при выделении компартиментов стандартными алгоритмами, а визуальный анализ показал соответствии распределения значений компартамента «клетчатому» паттерну на картах Hi-C. Несмотря на необходимость подбора ряда параметров, алгоритм АВСЕ, по сравнению с другими алгоритмами, даёт наиболее стабильный и единообразный результат.

Характеристика пространственной организации хроматина у комаров рода *Anopheles*. Полученные значения компартамента были использованы для описания пространственной организации хроматина у комаров рода *Anopheles* и её связи с известными генетическими и эпигенетическими характеристиками. В первую очередь, была проведён математический расчёт силы компартиментализации (Рисунок 5). В соответствии с параметрами работы алгоритма АВСЕ, были проведены расчёты силы компартиментализации для всех контактов и, отдельно, для контактов локусов, разделённых геномным расстоянием не более 10 млн. п.о.

Результаты сравнения рассчитанных величин показывают, что добавление в выборку контактов, разделённых большим геномным расстоянием, ослабляет вычисленную величину силы компартиментализации (Рисунок 5). Кроме этого, компартиментализация хроматина имаго комаров *An. merus* существенно выше, чем хроматина эмбрионов. Данные факты подтверждает влияние ориентации хромосом по Раблю на формирование контактов между компартаментами.

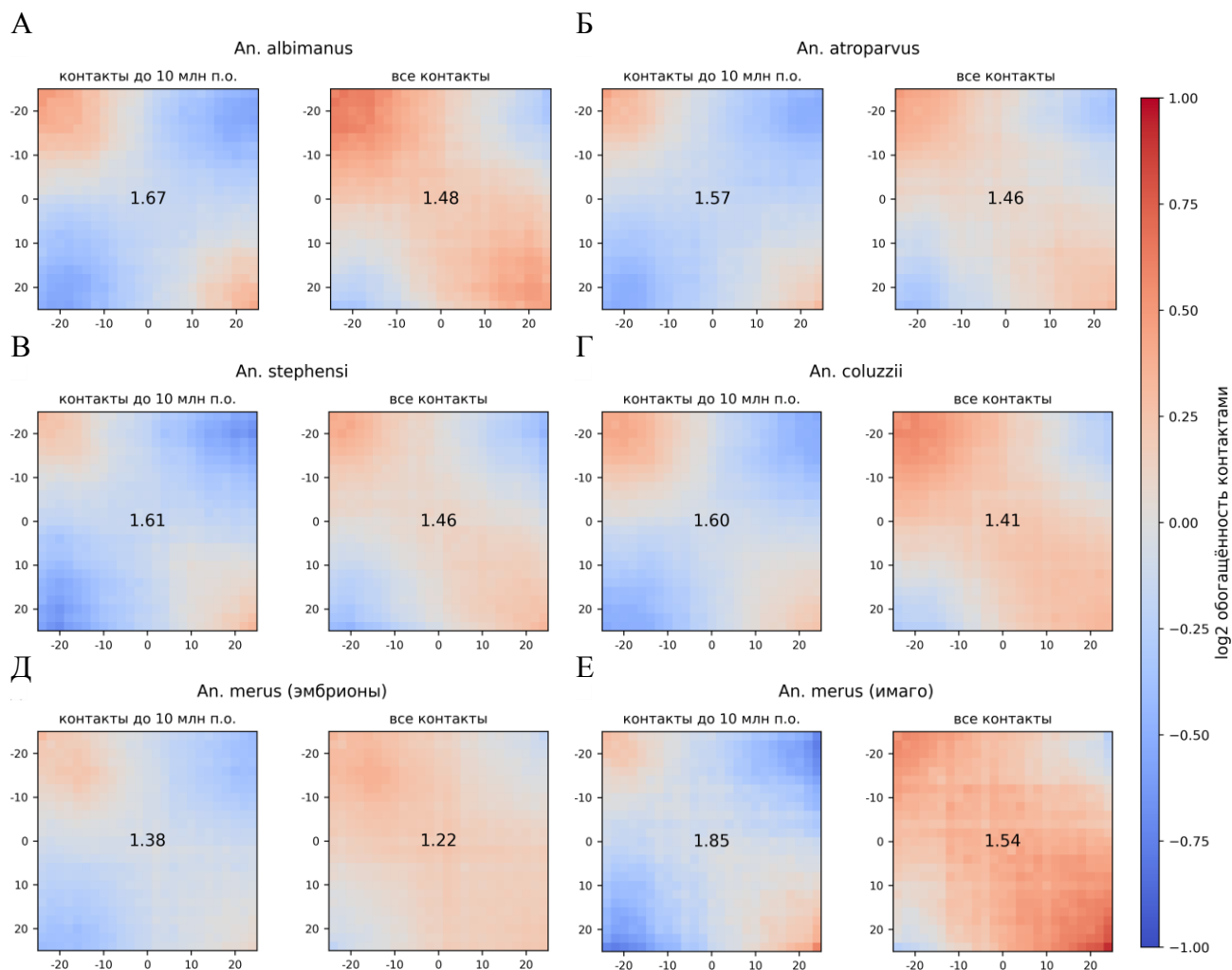


Рисунок 5. Компартиментализации хроматина в разных видах комаров рода *Anopheles*. По осям X и Y – значение компартиментов пары контактирующих бинов. Рассчитанная величина силы компартиментализации показано по центру графиков

Выраженная компартиментализация хроматина позволяет предположить, что алгоритмически выделяемые домены у комаров рода *Anopheles* задаются делением генома на компартименты. Анализ значений компартиментов вблизи границ доменов показывает значительное обогащение границ доменов А-компартиментом (Рисунок 6). Исследование границ с выраженной инсуляцией (меньше -1.5), аналогичное тому, что было проведено для клеток *G. gallus*, показало отсутствие выраженной закономерности между границами доменов и изменением знака компартамента. Сравнение распределения границ доменов с распределением

повторов разных классов демонстрирует отсутствие единообразия в ассоциации границ доменов и повторов у комаров рода *Anopheles*.

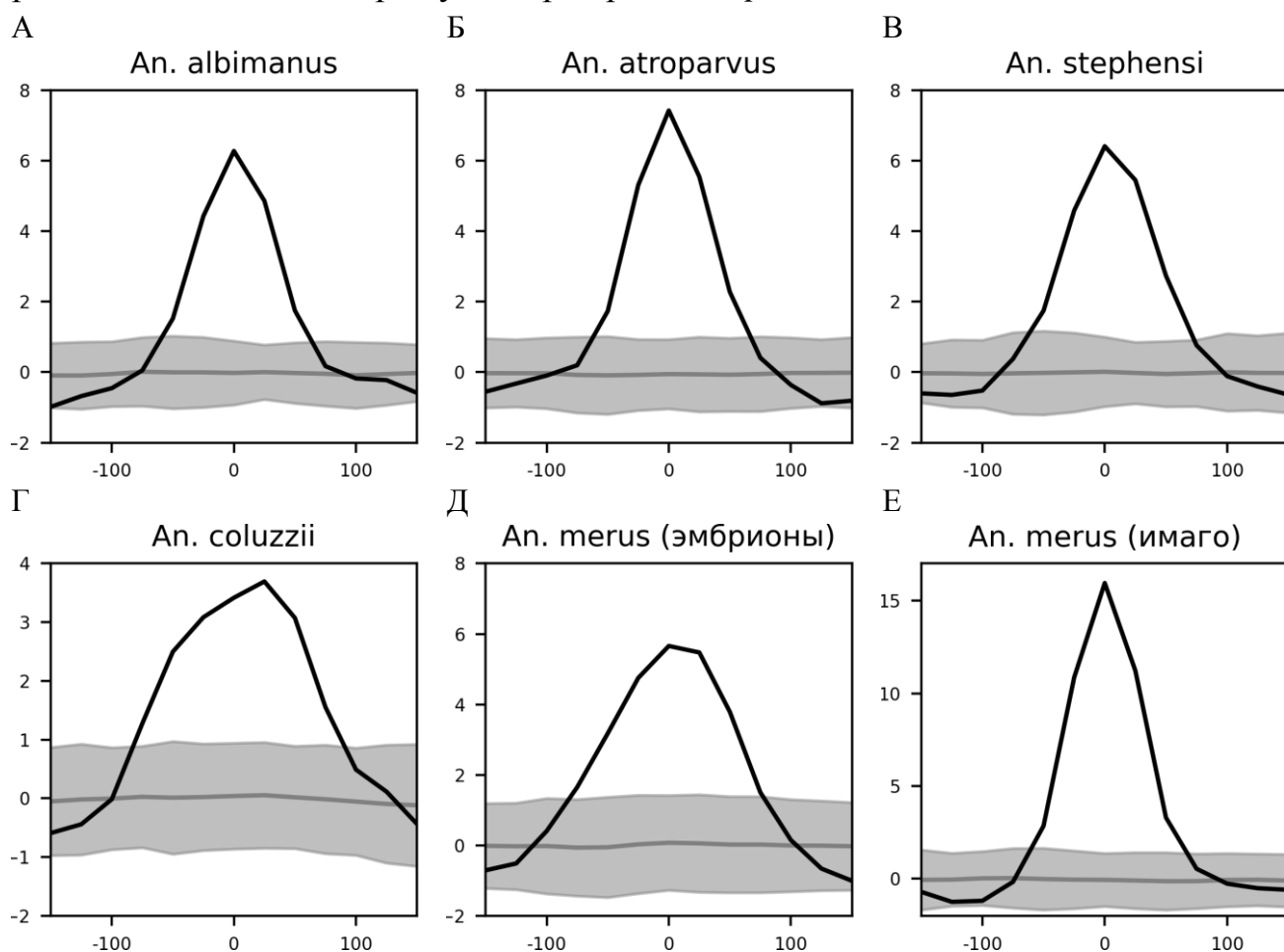


Рисунок 6. Распределение значений компартамента, рассчитанных алгоритмом ABCF, относительно границ доменов. По оси X расстояние в тысячах пар оснований от границы, по оси Y — среднее число сайтов связывания CTCF на 40 тысяч пар оснований. Чёрная линия – наблюдаемые данные. Серая область – 3 стандартных отклонения от ожидаемого.

Исследование консервативности архитектуры хроматина комаров рода *Anopheles*. Сравнение пространственной организации хроматина на уровне отдельных контактов показало, что несмотря на большое количество перестроек отличающих сравниваемые виды, внутри блоков синтении наблюдается высокая консервативность архитектуры хроматина, а подавляющая часть наблюдаемых различий обнаруживается в районах, непосредственно затронутых перестройкой.

Аналогично тому, как это было обнаружено у позвоночных, коэффициент корреляции между перцентилями контактов оказывается выше, чем между их частотами, однако разница оказывается не настолько велика. Так даже между самыми далёкими видами коэффициент корреляции перцентилей контактов составляет 0.95 ($p \ll 0.05$). При использовании же частот контактов коэффициент корреляции контактов оказывается в диапазоне от 0.8-0.9. Данная особенность комаров рода *Anopheles* связано с тем, что внутри синтенных блоков геномные

расстояния между синтенными локусами меняются слабо. Таким образом, основные различия в архитектуре хроматина между разными видами рода *Anopheles* лежат в районах эволюционных перестроек хромосом.

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. Резонно предположить, что наиболее ярко эти свойства можно обнаружить в районах наиболее молодых эволюционных перестроек хромосом, а в наиболее старых эти свойства будут смазаны последующими эволюционными процессами. Основываясь на литературных данных о филогении рода *Anopheles* и результатах попарных выравниваний, большую часть перестроек, наблюдаемых для каждого вида, можно классифицировать по времени их происхождения, их уникальности для той или иной эволюционной ветви и определить, были ли эти регионы использованы повторно в других видах. Таким образом, было выделено и классифицировано около 3200 районов эволюционного разрыва хромосом для *An. albimanus* и около 400–650 для других видов рода *Anopheles*. Сравнение районов эволюционных перестроек хромосом с распределением известных геномных характеристик – сила инсуляции, значение компартамента, встречаемость повторов разных классов – не позволило выявить какой-либо закономерности общей для всех исследуемых видов, за исключением того, что эволюционные хромосомные перестройки чаще наблюдаются в принадлежащих А-компарменту регионах.

Необходимо отметить, что существование хромосомной перестройки означает не только то, что соседние в предполагаемом предковом геноме регионы были удалены друг от друга, но и то, что участки, ранее удалённые, оказались соединены (Рисунок 7А). Таким образом возникает следующий вопрос: влияет ли значение компартамента на «выбор» регионов, который в ходе эволюционной перестройки соединяются? Для ответа этот вопрос, необходимо сравнить значения компарментов, в локусах, затронутых эволюционной перестройкой (Рисунок 7А) у гипотетического предка. Конечно, невозможно прямо определить, величину компарментализации 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 компарментализация сходная с предковой.

Исследование показывает, что несмотря на то, что эволюционные разрывы хромосом чаще происходят в регионах, принадлежащих А-компарменту,

соединяются они со случайными локусами, независимо от того, какому компартменту они принадлежат (Рисунок 7Б-Е). В свою очередь, высокая консервативность архитектуры хроматина на уровне отдельных контактов указывает, что новосоединённые в ходе перестройки локусы систематического влияние на компартментное состояние друг друга не оказывают.

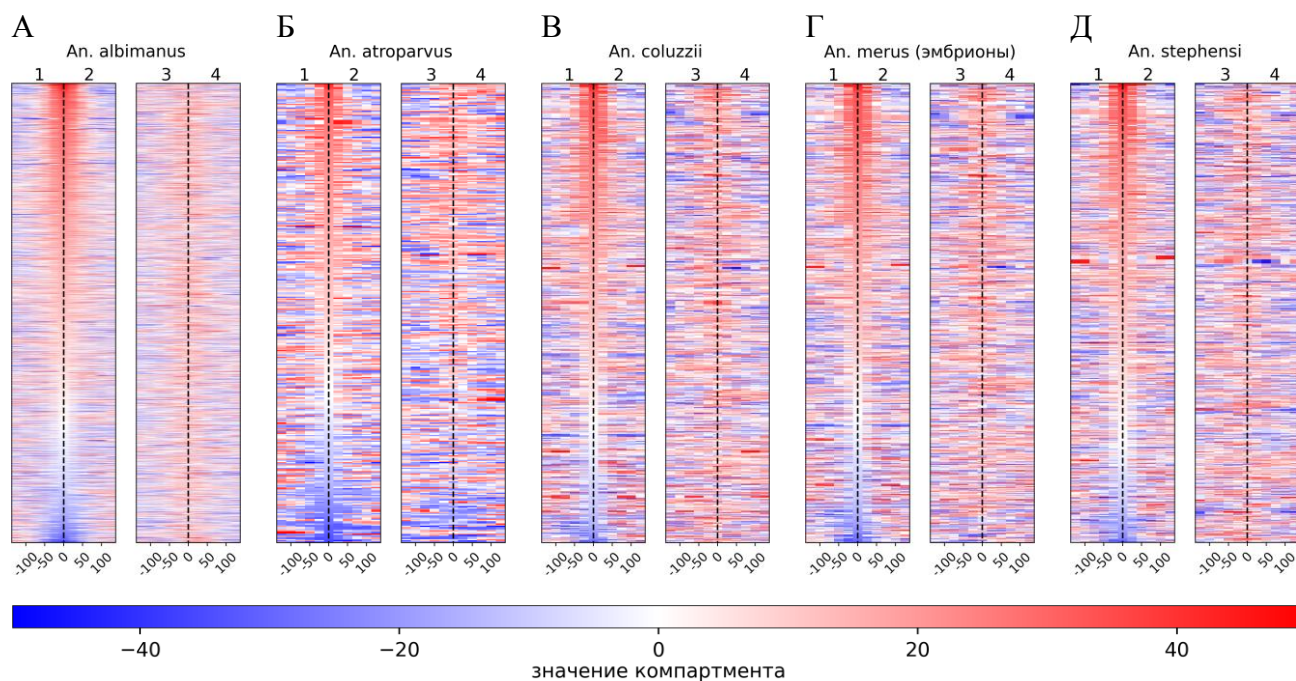


Рисунок 7. Влияние эволюционных перестроек на компартментализацию хроматина. А-Д. Значение компартмента в локусах, синтенных перестроенным, у филогенетически ближайшего вида без перестройки.

Таким образом для исследованных комаров рода *Anopheles* показана высокая консервативность архитектуры хроматина внутри синтенных блоков, прослеживаемая на уровне отдельных контактов. Так же, 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.

ЗАКЛЮЧЕНИЕ

Изучение принципов консервативности архитектуры хроматина, проведённое с помощью биоинформатического алгоритма C-InterSecture, показало, что консервативными являются не сами пространственные контакты, как таковые, а их относительные величины, нормированные на расстояние между локусами. Эта математически определённая закономерность обнаружена как при сравнении друг

с другим разных позвоночных, так и комаров рода *Anopheles*. Основанная на данных закономерностей мера P-BAD позволяет количественно оценивать различия пространственной организации хроматина для выбранных локусов между разными видами, превосходя в этом ранее используемые подходы.

Биоинформатическое сравнение архитектуры хроматина среди позвоночных показало, 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. Можно предположить, что эти отличия формируются на самых ранних стадиях клеточной дифференцировки и могут быть связаны с процессом компактизацией хроматина в клеточном ядре. Однако, сперматозоиды млекопитающих так же имеют сверхкомпактизованный хроматин, при этом свойства хроматина сперматозоидов от других типов клеток значимо не отличаются. Таким образом, конкретные механизмы участвующие в процессе формирования эритроцитов у *G. gallus* остаются неясными. С достаточной уверенностью можно говорить только о том, что в виду отсутствия ассоциацией между структурой доменов и расположением сайтов связывания CTCF, 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 данных типов клеток 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 A-компартаментам, высокоэкспрессирующимся районам, подобно тому, как это отмечено у позвоночных. При этом для зафиксированных в ходе эволюции перестроек не отмечено значимого влияние перестройки на компарментное 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 и между разными видами позвоночных, 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.

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 // *Nat. Commun.* – 2022. – V. 13. – I. 1 – P. 1960. <https://doi.org/10.1038/s41467-022-29599-5>