

**Аннотация:** Анализ Hi-C-карт ооцитов курицы на стадии ламповых щёток позволяет выделить домены хроматина, предположительно соответствующее DAPI-положительным и DAPI-негативным хромомерам. Анализ расположения генов относительно границ доменов позволяет предположить роль дивергентной транскрипции в формировании этих трехмерных структур. В ламповых щётках ооцитов курицы локусы, расположенные на расстоянии более 2х миллионов пар оснований практически не контактируют в пространстве ядра. Резкое падение частот трехмерных контактов хроматина после 2х миллионов пар оснований является отличительной особенностью хромосом типа ламповых щеток, не характерной для соматических клеток курицы и других позвоночных.

**Тема работы:** “Исследование трёхмерной организации хромосом типа ламповых щёток ооцитов курицы”

**Состав коллектива:**

Таскина Алёна Константиновна (ИЦиГ СО РАН, НГУ)

к.б.н., в.н.с. Фишман Вениамин Семёнович (ИЦиГ СО РАН, НГУ)

к.б.н., с.н.с. Гридина Мария Михайловна (ИЦиГ СО РАН)

Нурисламов Артём Рустамович (ИЦиГ СО РАН, НГУ)

**Информация о гранте:**

“Структура, функции и эволюция мейотических хромосом птиц”, РФФ 20-64-46021, руководитель Торгашева Анна Александровна, 2020-2022 год

**Научное содержание работы:**

**Постановка задачи:** исследовать трёхмерную организацию хромосом типа ламповых щёток курицы методом single cell Hi-C (scHi-C).

**Современное состояние проблемы (на момент начала работы):**

В пространстве интерфазного ядра позвоночных, эу- и гетерохроматин разделяются за счет процессов фазовой сепарации, образуя активный (А) и неактивный (В) компартменты (Nuebler et al. 2018). Одновременно с сегрегацией компартментов, белковый комплекс когезин динамически взаимодействует с ДНК, формируя петли, ограниченные сайтами связывания белка CTCF (Sanborn et al. 2015; Fudenberg et al. 2016). Формирование множественных петель, ограниченное сайтами белка CTCF, приводит к формированию топологически ассоциированных доменов (ТАДов). ТАДы характеризуются относительно высокой частотой контактов геномных элементов внутри ТАДа и пониженной частотой контактов между геномными элементами, находящимися в разных ТАДах (Dixon et al, 2012). Рядом авторов было показано, что ТАДы представляют собой относительно независимые регуляторные домены, внутри которых можно проследить скоординированную активацию, либо репрессию генов (Dileep et al, 2012). Трёхмерная организация генома курицы подчиняется тем же фундаментальным принципам, что ранее были описаны для генома млекопитающих (Fishman et al, 2019). Как макро-, так и микро- хромосомы курицы разделены на А, В компартменты и ТАДы. В геноме курицы, также как и в геноме млекопитающих, границы ТАДов обогащены сайтами посадки белка CTCF в конвергентном положении, что говорит об аналогичном механизме формирования ТАДов путём экструзии петель. Несмотря на то, что геном курицы в три раза меньше генома млекопитающих, размеры

ТАДов у двух этих групп совпадают. Кроме того, границы ТАДов фибробластов курицы и человека отличаются большой консервативностью. Организация интерфазного ядра сходна для различных типов клеток позвоночных. В то же время, в процессе мейотического и митотического делений трехмерная организация хромосом претерпевает масштабные изменения (Du et al. 2017; Flyamer et al. 2017; Ke et al. 2017; Alavattam et al. 2019; Wang et al. 2019).

В ходе митоза, хромосомы значительно компактизируются, что сопровождается исчезновением компартментов и ТАДов. При этом за счет активности конденсина, родственного когезину белкового комплекса, формируется регулярная петлевая структура (Gibcus 2018). Мейотическое деление сопровождается еще более сложной динамикой трехмерной архитектуры ядра. На стадии зиготены и пахитены в мужских гаметах мыши, по данным недавних работ (Alavattam et al. 2019; Patel et al. 2019; Wang et al. 2019), наблюдается: 1) потеря ТАДов. 2) Потеря компартментализация эу- и гетерохроматина для удаленных друг от друга локусов 3) компактизация хромосом, которая заключается в увеличении контактов на расстояниях до 3 миллионов п.о. На стадии сперматид и зрелых сперматозоидов ТАДы и компартменты появляются (Alavattam et al. 2019; Patel et al. 2019; Wang et al. 2019).

Изменения трехмерной организации хромосом в ходе женского мейоза также были недавно изучены. На стадии диплотены в транскрипционно-активных ооцитах мыши (not-surrounded-nucleus или NSN- ооцитов) наблюдаются ТАДы и компартменты (Flyamer et al. 2017). Важно отметить, что ТАДы не наблюдаются в каждой, отдельно взятой клетке. Если сравнить карту ТАДов, полученную на популяционных данных, с контактами мейотических хромосом, то в индивидуальных клетках можно выявить хроматиновые глобулы, которые располагаются как в пределах, так и между “популяционными” ТАДами, однако в среднем частота образования глобул, соответствующих “популяционным” ТАДам, выше, чем частота глобул, пересекающих их границу. В ходе созревания ооцитов, при переходе от транскрипционно-активной к транскрипционно-неактивной (surrounded nucleus, SN) стадии, ТАДы и компартменты в ооцитах становятся значительно менее выраженными (Flyamer et al. 2017). В овулирующих ооцитах, которые находятся на стадии второго деления мейоза, полностью отсутствуют ТАДы и компартменты, а общий паттерн укладки напоминает митотическую метафазную хромосому (Du et al. 2017; Flyamer et al. 2017; Ke et al. 2017).

### ***Подробное описание работы, включая используемые алгоритмы:***

#### *Построение тепловых карт пространственных контактов*

Адаптеры были удалены при помощи программного обеспечения cutadapt v. 2.10 с параметрами, предложенными для удаления Illumina TrueSeq адаптеров. Получившиеся прочтения были обработаны при помощи программного обеспечения Juicer. Получившиеся Hi-C матрицы были визуализированы в программе Juicebox. На основе выходных файлов с информацией о парах контактов были также построены .cool матрицы при помощи функции load pairs программного обеспечения cooler. Hi-C матрицы в .cool формате вместе с данными об экспрессии генов были визуализированы в геномном браузере Hi-Glass.

## Контроль качества Hi-C библиотек

Для контроля качества библиотек мы использовали такие статистики, как число уникальных контактов фрагментов рестрикции, доля не картированных на геном курицы прочтений, доля внутрифрагментных прочтений, представленность последовательности GATCGATC в прочтениях.

Также для контроля качества библиотек мы вычисляли долю фрагментов, не подвергшихся рестрикции или подвергшихся лигированию внутримолекулярно по формуле:

$$\frac{in + out - left - right}{\sum pairs}$$

где *in* – число пар прочтений, выровненных в конвергентной ориентации, *out* – число пар прочтений, выровненных в дивергентной ориентации (F2R1), *right* и *left* – число пар, выровненных в сонаправленной вправо и в сонаправленной влево (ориентации FF или RR) соответственно.

## Анализ контаминации образцов ДНК

Для исследования источника загрязнения образцов ДНК мы выравнивали те прочтения, которые не картировались на геном курицы, на нуклеотидную коллекцию базы данных NCBI с помощью сервиса BLAST. Для выяснения того, на каком этапе произошло загрязнение образцов ДНК, мы оценили долю прочтений, содержащих мотив лигирования GATCGATC. Для этого мы использовали стандартную утилиту Linux `grep`.

## Анализ покрытия генома

Прочтения были выровнены на геном Galgal5 при помощи Burrows-Wheeler Aligner. Получившийся .sam файл был отсортирован, переконвертирован в .bam файл и индексирован при помощи samtools. Покрытие было получено при помощи программного обеспечения deeptools. Мы визуализировали получившиеся выравнивания и покрытие с помощью программного обеспечения Integrative Genomics Viewer.

## Определение границ ТАДов на Hi-C картах

Границы доменов на Hi-C матрицах с разрешением 10 тпн были выделены при помощи функции `findTADs` пакета `hicExplorer`. Подбирая разные параметры, мы обнаружили, что при значении  $\delta=0.01$  выделенные домены лучше всего соответствуют наблюдаемому визуально на Hi-C картах треугольникам. Остальные параметры были использованы по умолчанию. После этого границы доменов были скорректированы путём визуального анализа тепловых карт в программе Juicebox.

## *Анализ частоты генов, экспрессирующихся в конвергентном направлении, в границах доменов Hi-C карт*

Вначале в последовательности генома были выделены промежутки, которые находятся между двумя расположенными подряд генами, которые экспрессируются навстречу друг другу. Затем вокруг каждой границы домена мы выделили окрестность размером 1 млн пн и разбили её на участки длиной 5 тысяч пн. После этого мы подсчитали, как часто выделенные в окрестности границы участки пересекаются с последовательностями, расположенными между конвергентно экспрессирующимися генами. В качестве контроля мы использовали наборы случайных интервалов, соответствующих экспериментально выделенным доменам по размерам и частотам встречаемости на каждой из хромосом.. По сгенерированному 100 раз контролю мы подсчитали среднее значение и стандартное отклонение частоты генов, экспрессирующихся в конвергентном направлении. Для построения графиков был использован пакет matplotlib.

### *Построение кривой зависимости частоты контактов между локусами от геномного расстояния между ними*

Для построения кривой зависимости частоты контактов между локусами от геномного расстояния между ними мы использовали функцию `compute_expected` пакета `cooltools`.

### **Полученные результаты:**

1. Анализ scHi-C библиотек ооцитов курицы, полученных разными протоколами, показывает, что использование полногеномной амплификации сопряжено с появлением различных артефактов и снижением качества данных. Использование ПЦР-амплификации в комбинации с обогащением участками лигирования ДНК является более эффективным способом получения scHi-C библиотек ооцитов курицы.
2. В ламповых щетках ооцитов курицы локусы, расположенные на расстоянии более 2х миллионов пар оснований практически не контактируют в пространстве ядра. Резкое падение частот трехмерных контактов хроматина после 2х миллионов пар оснований является отличительной особенностью хромосом типа ламповых щеток, не характерной для соматических клеток курицы и других позвоночных.
3. Анализ Hi-C-карт позволяет выделить домены хроматина, предположительно соответствующее DAPI-позитивным и DAPI-негативным хромомерам. Анализ расположения генов относительно границ доменов позволяет предположить роль дивергентной транскрипции в формировании этих трехмерных структур.

### Иллюстрации, визуализация результатов:

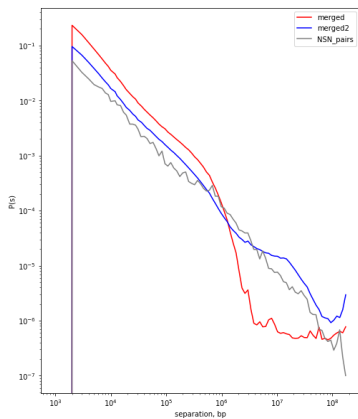


Рис.20 Кривые зависимости частоты контактов от геномного расстояния между ними. По оси  $x$  – расстояние между локусами, по оси  $y$  – частота контактов между локусами. Красным – объединенная тепловая карта библиотек первого типа, синим – объединенная тепловая карта библиотек второго типа, серым – объединенная тепловая карта single-nucleus Hi-C ооцитов мыши.

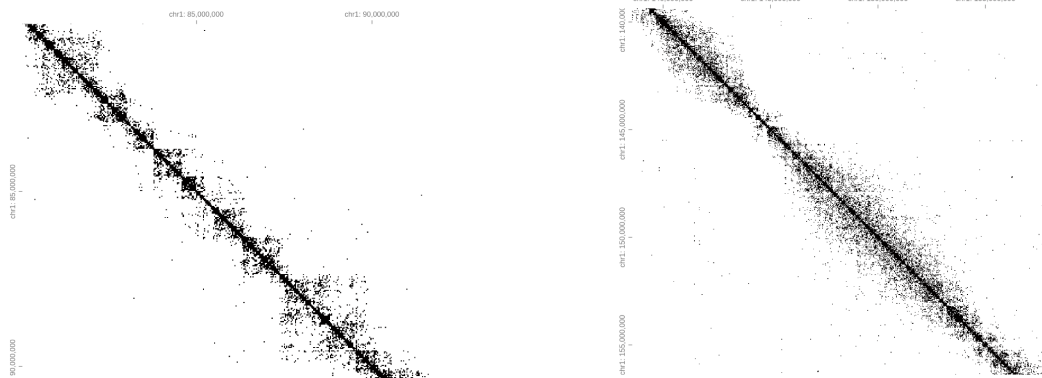


Рис.22 Два разных участка тепловой карты, полученной после объединения данных секвенирования библиотек первой группы. Слева – DAPI- хромомеры, справа – DAPI+. Разрешение 32 тпн. Все контакты показаны чёрными пикселями, вне зависимости от числа контактов.

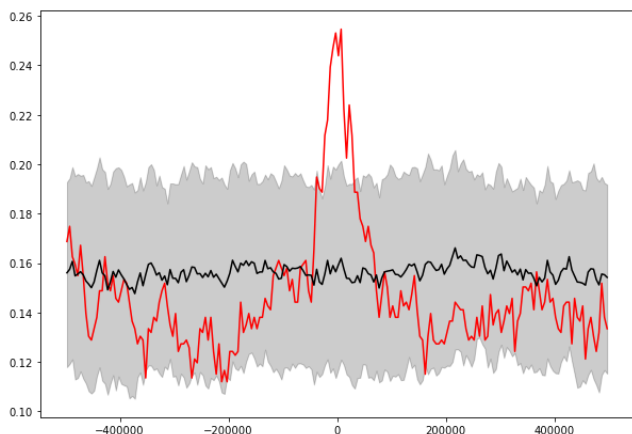


Рис.26 Частота генов, транскрибирующихся в конвергентном направлении, в зависимости от расстояния до границы домена (красным) и в среднем в контрольной выборке (чёрным). Серая область – три стандартных отклонения, подсчитанных по контрольной выборке.

### Эффект от использования кластера в достижении целей работы:

Благодаря мощным вычислительным ресурсам кластера стало возможным использование высокопроизводительных алгоритмов в этой работе

