

Отчет об использовании кластера ИВЦ НГУ в научном исследовании

Состав коллектива:

Нурисламов Артем Рустамович, студент ФЕН НГУ, гр.22425

(a.nurislamov@g.nsu.ru)

Фишман Вениамин Семенович, к.б.н., в.н.с. ФИЦ ИЦиГ СО РАН (научный руководитель) (minja-f@yandex.ru)

Тема работы: Восстановление петель хроматина экспрессией экзогенного белка CTCF

Цель работы - оценить способность CTCF разных видов участвовать в формировании петель в хроматине клеток человека.

Для выполнения этой цели были поставлены следующие задачи:

1. Выполнить эволюционный анализ структуры белка CTCF, определить список видов для дальнейших экспериментов на основе особенностей организации хроматина;
2. Создать вектор для экспрессии CTCF выбранных видов, экспрессировать ортологи белка в клетках человека с системой деградации эндогенного CTCF;
3. Собрать Hi-C библиотеки из трансфицированных клеток, проанализировать паттерн пространственных контактов в границах доменов хроматина;

Современное состояние проблемы:

Пространственная организация геномов большинства эукариот консервативна и представлена топологически-ассоциированными доменами и компартментами. Несмотря на консервативность в организации геномов у эукариот, в формировании ТАДов и компартментов могут быть задействованы разные механизмы. Один из механизмов, представленный у позвоночных животных, задействован в формировании ТАДов путем экстрезии петель хроматина когезиновым комплексом при участии белка-инсулятора CTCF. CTCF впервые появляется более 500 миллионов лет назад у билатеральных животных и так же, как и остальные белки, осуществляющие образование петель хроматина путем экстрезии, высококонсервативен. Тем не менее, не у всех таксонов билатеральных животных петли хроматина формируются при участии CTCF. У мух-дрозофил, известного модельного объекта генетики насекомых, несмотря на наличие полного набора белков из комплекса экстрезии петли, этот механизм не был обнаружен. Напротив, у дрозофил большинство петель хроматина формируется по механизмам, которые реже представлены у позвоночных. В связи с

данными фактами неизвестна причина того, почему у одних видов животных есть когезин/CTCF-зависимая экструзия петли, а у других данный механизм отсутствует.

Описание работы, включая используемые алгоритмы:

Работа включала в себя как экспериментальную часть, так и биоинформационную обработку данных. Первичная аминокислотная последовательность ортологов CTCF была получена из публичных баз данных (NCBI, Ensembl) или собрана из геномных или транскриптомных данных с помощью `bwa mem` и `hisat`. В множественном выравнивании аминокислотных последовательностей использовался алгоритм `clustal omega`. На основе данных выравнивания были определены консервативные аминокислотные мотивы в CTCF и определен список видов для экспрессии их ортологов CTCF в клетках человека. Далее были получены плазмиды, экспрессирующие соответствующие варианты CTCF. Плазмиды были трансфицированы в клетки HCT116-hCTCF-mAC с системой ауксин-зависимой деградации эндогенного белка CTCF. Затем посредством проточной цитометрии были отобраны клетки, позитивные по каналу mCherry (т.е. трансфицированные плазмидой), и зафиксированы в формальдегиде для Hi-C. Далее были собраны Hi-C библиотеки по протоколу low-input, библиотеки секвенировали в режиме PE150 на платформах Illumina и DNBSEQ. Первичная обработка данных (оценка качества данных и удаление адаптеров секвенирования) осуществлялась при помощи ПО FastQC и cutadapt. Затем с использованием juicer и cooler получали карты трехмерных контактов хроматина (Hi-C карты) в форматах hic и cool/mcool. Показатели инсуляции (сепарации в пространстве локусов генома) и усредненных петель получали в coolpurrr. Статистическая обработка выполнялась с помощью пакета scipy, визуализация результатов статистического анализа и данных секвенирования выполнялась в matplotlib и seaborn.

Полный список ПО, использованного в работе:

- cooler
- coolpurrr
- clustal omega
- cutadapt
- fastqc
- hisat
- bowtie2
- bwa mem
- juicer

- meme
- fimo
- seaborn
- matplotlib
- samtools
- bedtools
- scipy
- numpy

Полученные результаты:

1. Эволюционный анализ первичной структуры N-терминального домена CTCF выявил функциональные аминокислотных мотивы с разным временем происхождения: мотив KTYQR впервые появляется у челюстноротых позвоночных, а мотив YXF сохраняется у всех групп билатеральных животных.
2. Разработана система для проверки комплементации функций CTCF в клетках человека HCT116 экзогенными вариантами белка. Экспрессия вариантов CTCF мыши и курицы в клетках человека приводит к восстановлению когезин/CTCF-опосредованных петель хроматина.

Эффект от использования кластера в достижении целей работы:

Вычислительная мощность кластера ИВЦ НГУ позволила выполнить ряд ресурсоемких операций, таких как множественное выравнивание аминокислотных последовательностей (clustal omega), выравнивание и обработка геномных и транскриптомных данных (hisat, bowtie2, bwa mem, cutadapt, samtools), получение и анализ карт пространственных контактов (cooler, juicer, coolppru), статистический анализ (scipy) и построение графических изображений на основе массива полученных данных (seaborn, matplotlib).