

УДК 57.05; 577.0; 573.22; 004.031.2

Е. В. Кулакова, А. М. Спицина, А. Г. Богомолов, Н. Г. Орлова,
А. И. Дергилев, И. В. Чадаева, В. Н. Бабенко, Ю. Л. Орлов

Программа анализа геномного распределения хромосомных контактов в ядре клетки по данным, полученным по технологиям ChIA-PET и Hi-C

Аннотация. В связи с ростом объемов данных геномики о положении сайтов связывания транскрипционных факторов, хромосомных контактах, аннотации геномных характеристик, полученных с помощью современных технологий секвенирования, растет потребность в разработке нового программного обеспечения для их статистической обработки и анализа. Рассмотрены технологии получения и программы анализа геномных данных секвенирования на основе технологий ChIA-PET и Hi-C. Представлена разработанная компьютерная программа для обработки геномных данных о хромосомных контактах и их функциональной аннотации.

Ключевые слова и фразы: биоинформатика, секвенирование, суперкомпьютерные вычисления, хромосомные контакты, графический интерфейс, базы данных.

Введение

В последнее десятилетие в мире активно развиваются экспериментальные технологии молекулярной биологии, приводящие к росту новых типов данных. Одной из нерешенных задач является анализ трехмерной структуры генома — понимание того, как хромосомы эукариот — молекулы ДНК, линейный размер которой во много раз превышает размеры ядра клетки, компактизуется и помещается в малый объем ядра. Накопление знаний о трехмерной структуре генома дает понимание о регуляции и транскрипции генов, что в свою очередь приближает ученых к решению задач фундаментальной медицины, нейробиологии, борьбе со многими заболеваниями. Анализ данных

Исследование структурной организации генома поддержано бюджетным проектом ИЦиГ СО РАН 0324-2016-0008 и РФФИ 14-04-01906 (АИД, АГБ, ИВЧ, ВНБ). Разработка программ анализа данных хромосомных контактов ChIA-PET поддержана РФФ (14-24-00123) (ЕВК, АМС, ЮЛО).

© Е. В. Кулакова^[1], А. М. Спицина^[2], А. Г. Богомолов^[3], Н. Г. Орлова^[4], А. И. Дергилев^[5], И. В. Чадаева^[6], В. Н. Бабенко^[7], Ю. Л. Орлов^[8], 2017

© Новосибирский государственный университет^[1, 2, 4, 5], 2017

© Институт цитологии и генетики СО РАН^[3, 6, 7, 8], 2017

© Программные системы: теория и приложения, 2017

о хромосомных контактах требует значительных вычислительных ресурсов и разработки программных средств [1, 2].

В последние годы активно развивались молекулярно-биологические методы, позволяющие реконструировать физические контакты разных районов хромосом на основе секвенирования ДНК [3, 4]. Первым методом для исследования пространственного взаимодействия двух специфических районов генома с помощью секвенирования контактирующих участков ДНК был метод 3C (chromosome conformation capture — захват конформации хромосом) [4]. Позже на основе метода 3C было разработано целое семейство экспериментальных методов исследования пространственной организации генома — 4C, 5C и другие (аббревиатуры, расширяющие первоначальное определение Chromosome Conformation Capture on Chip, и т.д.).

В 2009 году появилась модификация метода 3C, позволяющая определять хромосомные контакты в масштабе всего генома с помощью секвенирования, метод получил название Hi-C (High dimension chromosome Conformation — захват конформации высокой размерности) [3]. С помощью этого метода было показано, что хромосомы в клетках млекопитающих имеют внутреннюю организацию: существуют так называемые «топологические домены» — протяженные участки хромосом, для которых характерно большое количество внутридоменных контактов. Организация хромосом в топологические домены, представляет собой фундаментальную характеристику генома млекопитающих. Проведение подобных полногеномных исследований стало возможным лишь в последние годы.

Помимо вышеупомянутых методов, также существует метод ChIA-PET (Chromatin Interaction Analysis by Paired-End-Tag sequencing), использующий иммунопреципитацию хроматина, который позволяет определять в масштабе генома специфические контактирующие участки хромосом, контакты которых опосредованы белками или белковыми комплексами [5]. Таким образом, определяется структура хромосомных контактов, опосредованных специфическим заданным белком.

Цель данного исследования — разработка комплекса программ для статистической обработки данных о хромосомных контактах в ядре клетки, полученных методами высокопроизводительного секвенирования, а также компьютерный анализ расположения групп генов на хромосомах. Технические задачи — визуализация результирующих

графиков расположения генов относительно хромосомных (топологических) доменов и построение матрицы хромосомных контактов, опосредованных сайтами связывания транскрипционного фактора (по данным ChIA-PET).

Новизна состоит в:

- (1) выявлении списков генов, располагающихся на границах топологических доменов в геноме;
- (2) создании модуля генерации случайного набора данных по хромосомным контактам, на основе экспериментальных эмпирических распределений;
- (3) доступе к данным Hi-C и ChIA-PET, полученным как в рамках международного сотрудничества, так и в ИЦиГ СО РАН в Новосибирске.

1. Данные ChIA-PET и Hi-C

Рассмотрим используемые экспериментальные методы и форматы данных. Для исследования того, как молекула ДНК, линейный размер которой многократно превышает размеры интерфазного ядра, компактизуется и помещается в малый объем ядра, разработан ряд экспериментальных методов. Метод Hi-C позволяет представить пространственную структуру хромосомных контактов в масштабе всего генома [3] в форме линейных координат контактирующих пар участков. Метод основан на технологии 3C (Chromosome Conformation Capture) и массовом параллельном секвенировании. Технология 3C позволяет определить пространственное взаимодействие двух заранее выбранных (специфичных) районов генома [4]. По числу парных контактов (секвенированных пар участков генома) оценивается вероятность взаимодействия двух отдаленных друг от друга районов генома. Эксперимент выполняется в большой популяции клеток (10^5 – 10^6), поэтому исходно дает только оценку частот контактов.

Сначала формальдегидом фиксируют белок-белковые, белок-ДНК и белок-РНК- взаимодействия за счет образования ковалентных связей. Затем хромосомная ДНК фрагментируется с помощью рестриктаз. После этого проводят лигирование (соединение концов двухцепочечной ДНК) в условия разбавления. В таких условиях лигируются только концы молекул ДНК, сближенных в пространстве (парные фрагменты ДНК). Далее, молекулы маркируются с помощью биотина. Создается целая «библиотека» попарно взаимодействующих

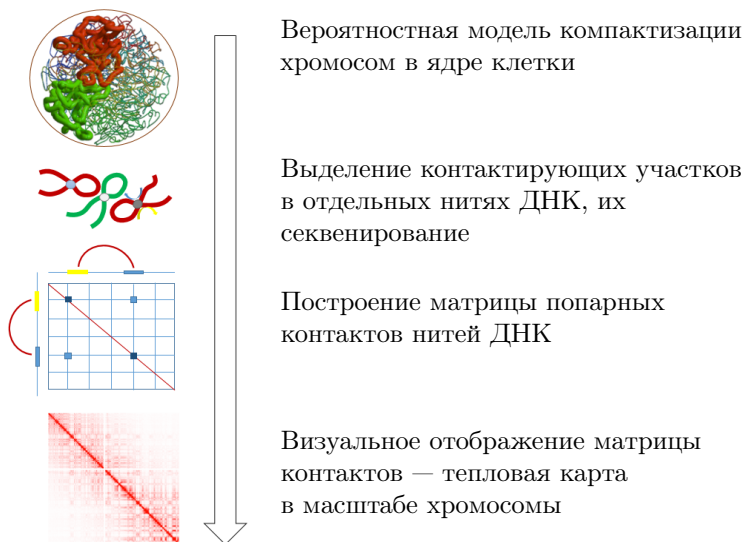


Рис. 1: Схема эксперимента и построения матрицы хромосомных контактов

молекул ДНК. Затем такие лигированные концы секвенируют (прочитывают), и полученные пары последовательностей ДНК (прочтений) картируют на геном (определяют их исходное положение на линейной последовательности хромосомы). Схема представлена на рис. 1. Исследование таких парных контактов представляет новую задачу биоинформатики.

Метод ChIA-PET (Chromatin Interaction Analysis by Paired-End-Tag sequencing), использующий иммунопреципитацию хроматина, определяет контактирующие участки хромосом, контакты которых опосредованы белками или белковыми комплексами в масштабе генома [5, 6]. Отличие от метода Hi-C состоит в том, что в процедуру обработки хроматиновых контактов добавлена стадия иммунопреципитации, также как в методе ChIP-seq. Таким образом, определяется структура хромосомных контактов, опосредованных специфическим белком (например, транскрипционным фактором ER α или фактором CTCF), а не все контакты хромосом.

На рис. 1 в самом верху представлена вероятностная модель компактизации хромосом в ядре клетки. В ходе анализа выделяются

хромосомные контакты, которые указывают на наличие хромосомных петель. Затем по таким данным строится симметричная матрица контактов, на основе которой рисуется тепловая матрица (термокарта). Подобные тепловые карты для данных пар сайтов связывания полученных при помощи метода ChIA-PET, отрисовываются и собственной разработанной программой. В зависимости от разрешения экспериментальных данных и матрицы могут быть достаточно разреженными, размерность матрицы может составлять 10000×10000 и выше, что представляет вычислительные сложности для обработки.

Метод ChIA-PET предназначен для исследования хромосомных петель, прежде всего между регуляторными районами генов, контактов между удаленными энхансерами и промоторами. Рассмотрим данные ChIA-PET о хромосомных контактах в геноме человека, опосредованных РНК-полимеразой II, а также парами сайтов связывания инсулятора CTCF (CCCTC-binding factor) [7].

При обработке данных методом Hi-C выделяются пространственные топологические домены на хромосоме: такая информация представлена в открытых базах данных, таких как 3DGD и 4DGenome [8, 9]. База данных 4DGenome хранит в себе данные взаимодействий хроматина, собранных из литературных источников и охватывает выходные данные методов 3C, 4C-Seq, 5C, Hi-C, ChIA-PET и Capture-C.

На данный момент существует достаточно большое число программ, методов и утилит для обработки данных высокопроизводительного секвенирования [10–14].

1.1. Используемые данные

В работе использовались следующие данные:

- Референсный геном из геномного браузера **UCSC Genome Browser**. Genome Browser позволяет масштабировать и визуально «прокручивать» хромосомы с аннотированным расположением генов, выделяя нужные районы. Table Browser обеспечивает доступ к базе аннотированных геномных данных в табличном формате. Genome Graphs отображает полногеномные наборы данных.
- Данные о парах сайтов связывания транскрипционного фактора белка CTCF взяты из **CTCFBSDB 2.0**.
- Данные о геномных доменах по методу Hi-C были получены в лаборатории проф. О. Л. Серова в ИЦиГ СО РАН [15, 16], представляющие собой разделение на домены у двух типов клеток — фибробласты и сперматозоиды мыши.

Нормированные данные о доменах предоставлялись в виде текстового файла формата bed, содержащего название хромосомы и координаты начала и конца топологического домена.

1.2. Средства разработки и библиотеки

Разработанный программный комплекс реализован на высокоуровневом интерпретируемом языке программирования Java версии 1.8, в среде IDE NetBeans (<https://netbeans.org/>). Это бесплатное ПО с открытым исходным кодом, которое имеет большое сообщество пользователей и разработчиков по всему миру. По сравнению с другими средами IDE среда IDE NetBeans обеспечивает высококлассную комплексную поддержку новейших технологий Java и последних усовершенствований стандартов Java. Это первая бесплатная среда IDE, поддерживающая JDK 8, JDK 7, Java EE 7, включая соответствующие усовершенствования HTML5 и JavaFX 2. С помощью редакторов и функций перетаскивания в среде IDE можно быстро и эффективно разрабатывать графические интерфейсы для приложений.

Для того чтобы осуществлять автоматическую подкачку необходимых библиотек, использовалась система сборки Gradle. Она собирает в себе модули от Apache Maven и Apache Ant. Gradle предоставляет DSL на языке Groovy вместо традиционной XML-образной формы представления конфигурации проекта.

Для загрузки геномных данных из внешнего ресурса (UCSC Genome Browser) использовалась библиотека jsoup (Java HTML Parser). Java-библиотека jsoup предназначена для разбора HTML-страниц (парсинг), позволяя извлекать необходимые данные, используя DOM, CSS и методы в стиле jQuery. Библиотека поддерживает спецификации HTML5 и позволяет парсить страницы, как это делают современные браузеры. Библиотеке можно подать для анализа URL, файл или строку.

Для взаимодействия R скриптов и java кода, была использована библиотека Rserve. Это TCP/IP сервер, который позволяет другим программам использовать возможности R без необходимости инициализации R. Каждое соединение имеет отдельное рабочее пространство и рабочий каталог. Rserve поддерживает удаленное подключение, аутентификацию и передачу файлов. В большинстве случаев, использование библиотеки заключается в интеграции R бэкенда для расчета статистических моделей, графиков в других приложениях.

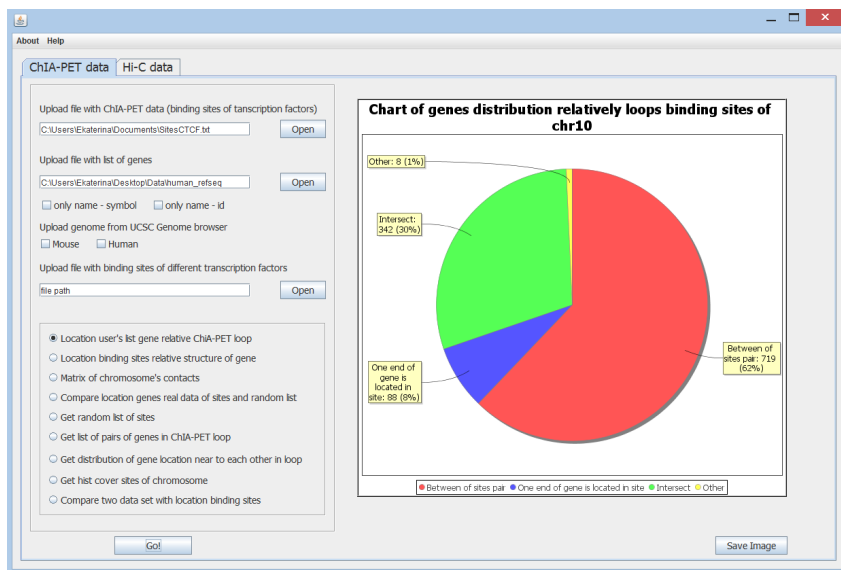


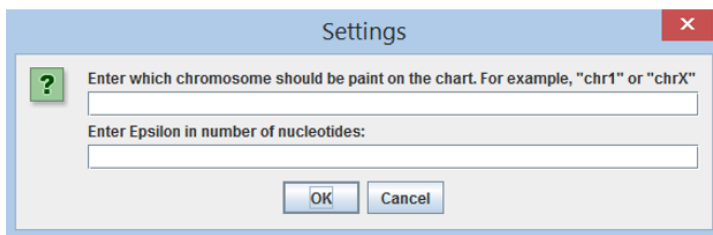
Рис. 2: Пример главного окна разработанной программы

Для работы с файлами разного расширения в работе использовались библиотеки Apache Commons. Это проект фонда Apache Software Foundation, который разрабатывает и поддерживает открытые библиотеки на языке Java. Apache Commons разбивается на множество более мелких проектов. Каждый такой проект имеет свой сайт, репозитории исходного кода и каталог в системе отслеживания ошибок. Компоненты проекта представляют собой JAR файлы, не превышающие 1 Мб.

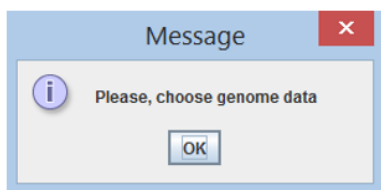
Для отрисовки графиков использовалась библиотека JFreeChart. Для работы с графикой матрицы использовалась библиотека **Heat Mar**.

1.3. Пример разработанной компьютерной программы

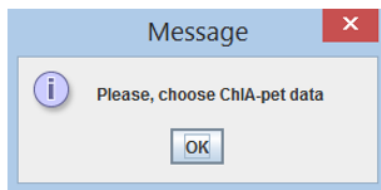
В ходе работы был реализован программный комплекс с графическим пользовательским интерфейсом. Он представляет главное окно (рис. 2) с двумя вкладками для ChIA-PET и для Hi-C данных соответственно. Для расширения системы новыми данными, можно просто добавить новую вкладку для работы с ними, не меняя готовый интерфейс. В окне программы есть кнопки для загрузки файла



(а) ввод параметров хромосомы для отображения на графике



(б) не выбраны геномные данные



(в) не выбраны данные ChIA-PET

Рис. 3: Три типа подсказок

пользователя, полное имя выбранного файла отображается в соответствующем текстовом окне. Для указания, чтобы данные генома были загружены, необходимо установить галочку в чекбоксе, а также указать организм (человек или мышь). Далее идет набор функций, которые можно выполнить.

Выбраны данные пар сайтов, геном и запущена функция для подсчета статистики расположения генов относительно пар сайтов связывания для хромосомы 10 человека. Видно, что больше половины генов лежат между сайтами связывания.

На панели справа отображается график, круговая диаграмма, гистограмма или график. В каталоге проекта создается папка «Report», в которой записываются подробные статистические результаты программы.

Полученное изображение можно сохранить нажатием на кнопку «Save Image». Сначала нужно загрузить необходимые файлы в соответствующие окна, затем выбрать интересующую функцию и нажать на кнопку «Go!». Если пользователь что-то забыл указать, то появляются подсказки. Сверху (рис. 3а) показан пример ввода параметра и номера хромосомы для отображения на графике. Ниже предупреждения: слева (рис. 3б) — необходима загрузка геномных данных, справа (рис. 3в) — нет выбора данных ChIA-PET.

В разработанном программном комплексе реализованы следующие функции:

- Выделение генов, расположенных на границах пространственных топологических доменов, по данным Hi-C.
- Вычисление распределения генов на топологических доменах.
- Вычисление распределения генов относительно пространственных хромосомных петель, на основе данных сайтов связывания транскрипционного фактора (CTCF).
- Вычисление распределения сайтов связывания относительно положения и структурных особенностей генов.
- Выделение пар генов, находящихся в одной петле.
- Генерация случайного списка сайтов связывания на основе реальных данных (размеры петель и сайтов связывания соответствуют действительности). Генерация на основе реальных данных случайного списка геномных доменов.
- Сравнительный анализ двух наборов данных хромосомных контактов, полученных из разных источников.

Все промежуточные результаты сохраняются в отдельные файлы. Пользователь может использовать их для дальнейшей работы. Например, обработанный программой (отфильтрованный) файл с геномом, можно подать на вход в другие программы или получить файл с распределением и самостоятельно построить график в необходимом формате.

Пользователь может выбрать, какую из хромосом и какой участок матрицы он бы хотел рассмотреть в большем масштабе. Есть возможность задать значение порога «дальнодействия» в геноме, т.е. подсчитать, сколько контактов из входных данных являются далекодействующими на основе заданного расстояния в нуклеотидах (от 1 Кб до 100 Кб и более). Рис. 4 представляет модуль отрисовки матрицы контактов.

2. Алгоритмы расположения генов относительно участков хромосомных контактов

С помощью анализа данных хромосомных контактов можно выделять топологически ассоциированные домены (TAD — topologically associated domains) и выявлять группы генов, которые имеют схожие функциональные характеристики или образуют кластеры в таких топологических доменах [17].

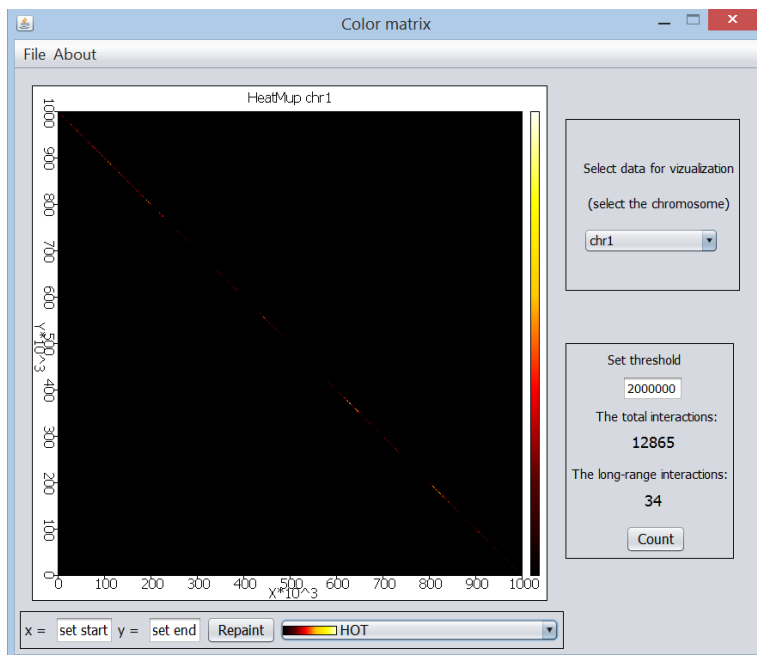


Рис. 4: Окно модуля отрисовки матрицы контактов ChIA-PET

Среди всех функций данной программы рассмотрим функцию для статистического анализа покрытия хромосомы сайтами связывания. Входной параметр функции — это количество столбцов результирующей гистограммы, т.е. на сколько частей надо «поделить» хромосому для визуализации. Входные данные — список парных сайтов связывания, которые формируют структурную петлю, а записываются как пары позиций (линейные координаты) на хромосоме.

Если взять связный список, который хранит в себе каждую точку (координату) структурной петли на хромосоме, то памяти компьютера может быть недостаточно для обработки такой информации (по 4 байта только для одной координаты, а их может быть более 200,000,000). Поэтому был реализован другой алгоритм для подсчета такой статистики. Суть его заключается в том, что сначала мы храним один объект класса `Interval`, в котором содержатся координаты начала хромосомы и конца (назовем их x_1 и x_n , где индексы это координаты нуклеотидов на хромосоме). Затем при сравнении этого отрезка с координатами сайта связывания (назовем их x_i и x_j) мы

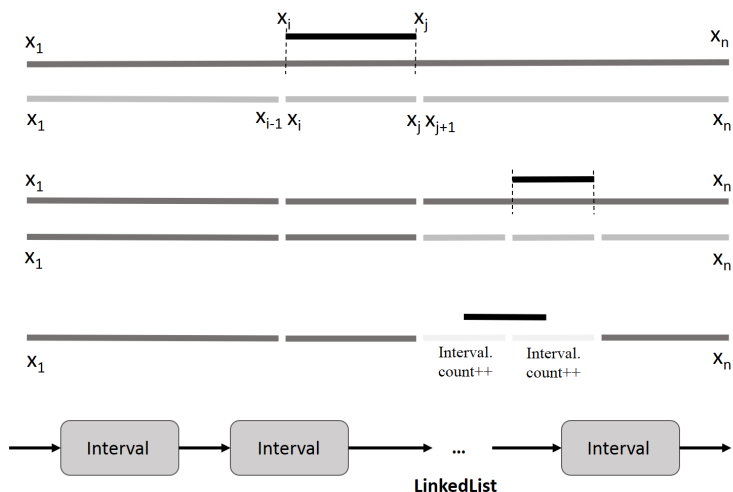
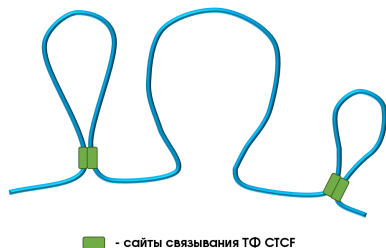


Рис. 5: Схема разбиения хромосомы на части при сравнении координат сайта связывания

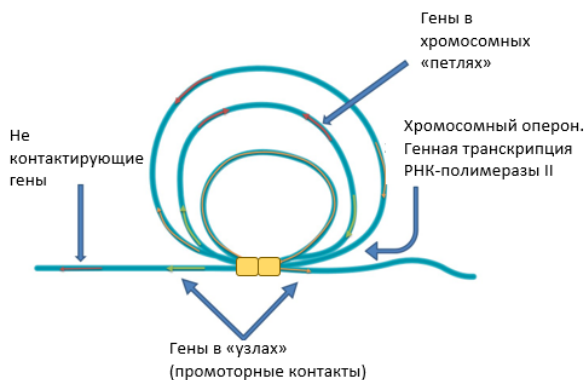
«делим» рассматриваемую хромосому на три части, т.е. создаем три новых объекта класса `Interval`: у первого координаты будут x_1 и x_{i-1} , у второго x_i и x_j и у третьего x_{j+1} и x_n . Все объекты хранятся в `LinkedList`. Затем процедура повторяется (см. рис. 5).

Минимальный размер разбиения задается параметрически. Для каждого объекта хранится счетчик, который подсчитывает количество попаданий различных сайтов в данные координаты. Например, в случае, когда из-за достигнутого минимального размера хромосомной единицы разбиение невозможно, то соответствующие счетчики увеличиваются на один. Таким образом, мы будем иметь более компактное по требованию к памяти разбиение.

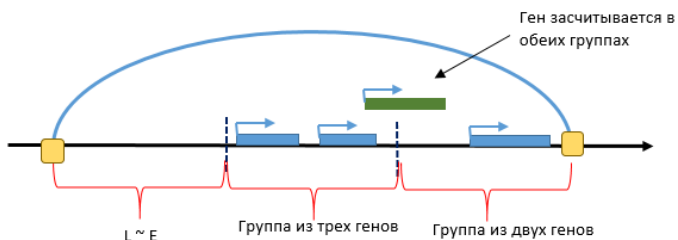
На рис. 5 представлена схема разбиения хромосомы на части при сравнении координат сайта связывания. В верхней части рисунка — разбиение хромосомы на части при сравнении координат сайта связывания и текущего объекта класса `Interval`. Увеличение счетчика «покрытия» (`Interval.count`) невозможно при дальнейшем дроблении отрезка. В нижней части рисунка представлена структура хранения этих объектов в связном списке. Чем меньше минимальный размер такого разбиения, тем точнее будет график. Можно получить разбиение и без учета минимального размера разбиения, т.е. получить не гистограмму, а график.



(а) Пары сайтов связывания транскрипционного фактора CTCF на хромосомной нити ДНК



(б) Общая схема расположения генов в хромосомных петлях



(в) Группы генов, располагающихся рядом в пределах петли на основе заданной погрешности вычислений

Рис. 6: Выделение и подсчет групп генов, располагающихся рядом

Рассмотрим функцию для получения распределения частоты встречаемости генов в хромосомных петлях, образованных путем контактов пар сайтов связывания транскрипционного фактора CTCF. На рис. 6а изображено схематичное расположение хромосомных петель.

Под группой понимаются те гены, которые располагаются «рядом» друг с другом в пространственной близости — в хромосомной петле, ограниченной сайтами связывания (рис. 6б). Понятие «рядом» подразумевает под собой гены, располагающиеся в пределах заданной погрешности (рис. 6в). Для подсчета такой статистики сначала отбираются гены, которые располагаются внутри хромосомной петли. Затем хромосома «делится на кусочки» заданного размера (E), т.е. E — это параметр (расстояние в геноме, в нуклеотидах), который задает то, насколько удаленные друг от друга гены мы считаем располагающимися «рядом». Далее для каждого такого участка подсчитывается число генов, попадающих в него. Процесс происходит для каждой петли из входных данных.

Данный график построен на основе генома человека и всех петель сайтов связывания из входных данных. В программе реализован альтернативный вариант подсчета частоты встречаемости генов. Главной его особенностью является исключение из обработки хромосомных петель с излишней вложенностью (рис. 7а).

3. Статистический анализ расположения генов относительно участков хромосомных контактов

В ходе работы были получены результаты анализа расположения генов относительно петель, образованных парами сайтов связывания CTCF генома человека.

В процессе были выделены списки генов, пересекающиеся с сайтами связывания транскрипционного фактора CTCF [7], а также случайный набор сайтов связывания. Далее были построены их распределения. По диаграммам рис. 7 можно предположить, что расположение генов имеет не случайный характер.

Большая часть генов лежит внутри петель сайтов связывания и не пересекается с нуклеотидными координатами самих сайтов. Различие распределений подтверждается на уровне значимости $p < 0.01$.

На рис. 8 представлено распределение числа генов, располагающихся «рядом» в одной хромосомной петле, образованной сайтами связывания транскрипционного фактора CTCF с учетом вложенности петель по частотам встречаемости с входным параметром погрешности 9000 нт. Оказалось, что гены по большей части (в масштабе

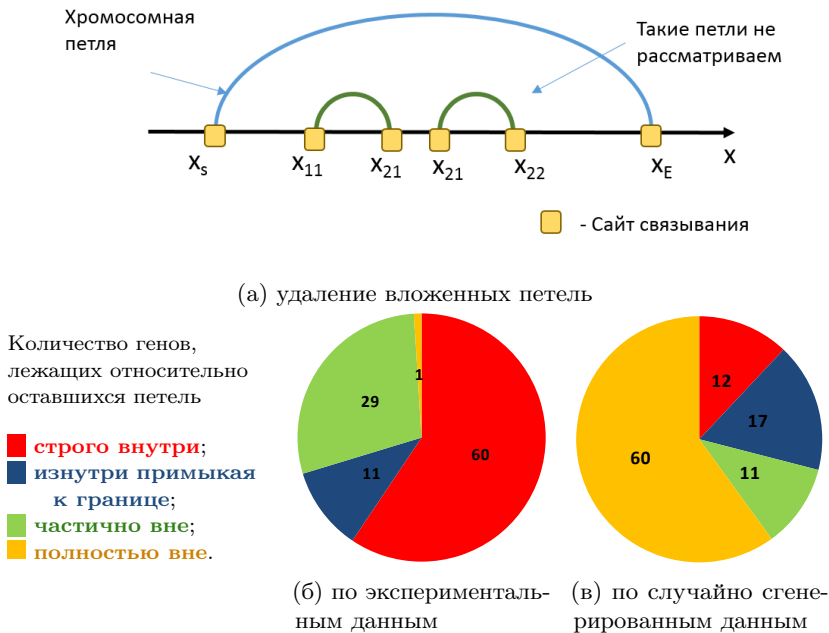


Рис. 7: Расположение генов относительно пар сайтов связывания

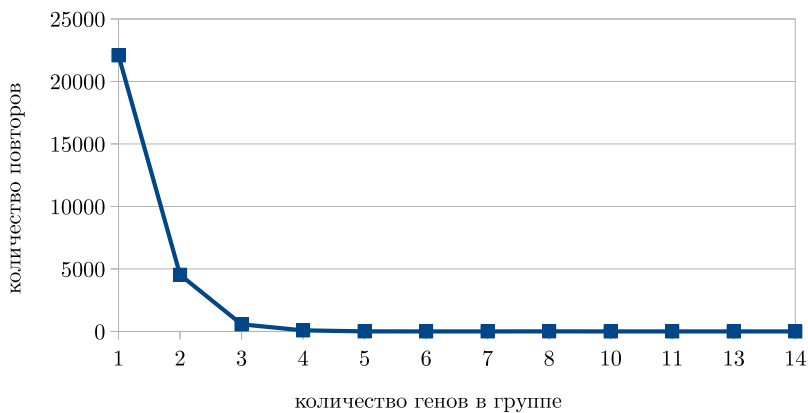


Рис. 8: График распределения в геноме человека числа генов, расположенных в общей хромосомной петле

гена) располагаются по одному — 22 000 таких случаев, результат считается для списка петель с учетом вложенности петель. Число генов в петле экспоненциально убывает.

В большинстве случаев гены расположены в петлях по одному. Максимально в группу входят до 8 генов. Такие списки генов в дальнейшем можно анализировать, искать общие молекулярные функции или обобщенные онтологии.

Отметим разработанные в мире средства визуализации хромосомных контактов 3D-GNOME [18] и QuIN [19], конвейеры обработки данных ChIA-PET2 [20].

Заключение

Разработан программный комплекс для статистического анализа расположения генов относительно топологических доменов и структурных особенностей хромосом, расширяющий круг задач для анализа символьных последовательностей генома [21]. Программа анализа данных ChIA-PET впервые разработана в России и ориентирована на суперкомпьютерные применения [22]. Применение к данным Hi-C внутри топологических доменов хромосом в геноме мыши впервые позволило реконструировать списки генов и найти категории генных онтологий, связанные с расположением генов на границах доменов. Полученные списки генов можно в дальнейшем анализировать с помощью других доступных ресурсов и новых наборов данных Hi-C.

Исследование участков хромосомных контактов на уровне нуклеотидов предполагает анализ однонуклеотидных полиморфизмов и их фланкирующих участков с использованием оценок сложности текста [23, 24], контекстных характеристик, связанных с позиционированием нуклеосом [25–27] расположением сайтов связывания транскрипционных факторов в таких регуляторных участках в модельных организмах [28]. Разрабатываются новые программы аннотации функциональных элементов генома, микроРНК, участков хромосомных контактов, связанные с технологией ChIA-PET [29]. Отметим новые компьютерные инструменты анализа данных Hi-C [30]–Hi-Browse [31], Juicebox [32], 3D Genome Browser [33], IC-Finder [34]. Каталог существующих в мире компьютерных инструментов анализа хромосомных контактов с помощью данных секвенирования, достаточно подробно представлены на сайте [OMICtools](#). Разработанная программа доступна по запросу к авторам.

***Благодарности.** Авторы благодарны В. С. Фишману, Н. Р. Батумину, Н. Л. Подколюдному, И. Г. Черных, а также ССКЦ СО РАН за предоставление данных и поддержку работы.*

Список литературы

- [1] Ю. Л. Орлов. «Компьютерное исследование регуляции транскрипции генов эукариот с помощью данных экспериментов секвенирования и иммунопреципитации хроматина», *Вавиловский журнал генетики и селекции*, **18**:1 (2014), с. 193–206, URL: <http://vavilov.elpub.ru/index.php/jour/article/view/240> ↑ ²²⁰
- [2] А. М. Спицина, Ю. Л. Орлов, Н. Н. Подколдная и др. «Суперкомпьютерный анализ геномных и транскриптомных данных, полученных с помощью технологий высокопроизводительного секвенирования ДНК», *Программные системы: теория и приложения*, **6**:1(23) (2015), с. 157–174, URL: http://psta.psisras.ru/read/psta2015_1_157-174.pdf ↑ ²²⁰
- [3] E. Lieberman-Aiden, N. L. Van Berkum et al. “Comprehensive mapping of long-range interactions reveals folding principles of the human genome”, *Science*, **326**:5950 (2009), pp. 289–293. ↑ ^{220,221}
- [4] J. Dekker, K. Rippe, M. Dekker, N. Kleckner. “Capturing chromosome conformation”, *Science*, **295**:5558 (2002), pp. 1306–1311. ↑ ^{220,221}
- [5] M. J. Fullwood, M. H. Liu, Y. F. Pan et al. “An oestrogen-receptoralpha-bound human chromatin interactome”, *Nature*, **462**:7269 (2009), pp. 58–64. ↑ ^{220,222}
- [6] G. Li, X. Ruan, R. K. Auerbach et al. “Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation”, *Cell*, **148**:1–2 (2012), pp. 84–98, URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3339270/> ↑ ²²²
- [7] L. Handoko, H. Xu, G. Li et al. “CTCF-mediated functional chromatin interactome in pluripotent cells”, *Nat. Genet.*, **43**:7 (2011), pp. 630–638. ↑ ^{223,231}
- [8] C. Li, X. Dong, H. Fan et al. “The 3DGD: a database of genome 3D structure”, *Bioinformatics*, **30**:11 (2014), pp. 1640–1642. ↑ ²²³
- [9] L. Teng, B. He, J. Wang, K. Tan. “4DGenome: a comprehensive database of chromatin interactions”, *Bioinformatics*, **31**:15 (2015), pp. 2560–2564, URL: <https://academic.oup.com/bioinformatics/article-pdf/31/15/2560/5026291/btv158.pdf> ↑ ²²³
- [10] Е. В. Кулакова, А. Г. Богомолов, А. О. Брагин, Э. Р. Галиева, Г. Ли, Ю. Л. Орлов. «Компьютерные методы определения хромосомных контактов в ядре клетки по данным технологий секвенирования (обзор)», *Молекулярная биология*, 2017 (в печати). ↑ ²²³
- [11] K. Kruse, C. B. Hug, B. Hernández-Rodríguez, J. M. Vaquerizas. “TADtool: visual parameter identification for TAD-calling algorithms”, *Bioinformatics*, **32**:20 (2016), pp. 3190–3192, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5048066/pdf/btw368.pdf> ↑ ²²³

- [12] M. J. Rowley, V. G. Corces. “Minute-Made Data Analysis: Tools for Rapid Interrogation of Hi-C Contacts”, *Mol Cell.*, **64**:1 (2016), pp. 9–11. [↑](#) ²²³
- [13] A. T. Lun, M. Perry, E. Ing-Simmons. “Infrastructure for genomic interactions: Bioconductor classes for Hi-C, ChIA-PET and related experiments. Version 2”, *F1000Res.*, **5** (2016), pp. 950, URL: <https://f1000research.com/articles/5-950/v2/pdf> [↑](#) ²²³
- [14] Z. Xu, G. Zhang, C. Wu, Y. Li, M. Hu. “FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data”, *Bioinformatics*, **32**:17 (2016), pp. 2692–2695. [↑](#) ²²³
- [15] Н. Р. Баттулин, В. С. Фишман, Ю. Л. Орлов, А. Г. Мензоров, Д. А. Афонников, О. Л. Серова. «3С-методы в исследованиях пространственной организации генома», *Вавиловский журнал генетики и селекции*, **16**:4/2 (2012), с. 872–876, URL: <http://vavilov.elpub.ru/jour/article/view/85> [↑](#) ²²³
- [16] N. Battulin, V. S. Fishman, A. M. Mazur et al. “Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach”, *Genome Biol.*, **16**:1 (2015), pp. 77. [↑](#) ²²³
- [17] D. S. Neems, A. G. Garza-Gongora, E. D. Smith, S. T. Kosak. “Topologically associated domains enriched for lineage-specific genes reveal expression-dependent nuclear topologies during myogenesis”, *Proc. Natl. Acad. Sci. USA*, **113**:12 (2016), pp. E1691–E1700, URL: <http://www.pnas.org/content/113/12/E1691.long> [↑](#) ²²⁷
- [18] P. Szalaj, P. J. Michalski, P. Wróblewski, Z. Tang, M. Kadlof, G. Mazzocco, Y. Ruan, D. Plewczynski. “3D-GNOME: an integrated web service for structural modeling of the 3D genome”, *Nucleic Acids Res.*, **44**:W1 (2016), pp. W288–W293, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987952/pdf/gkw437.pdf> [↑](#) ²³³
- [19] A. Thibodeau, E. J. Márquez, O. Luo, Y. Ruan, F. Menghi, D. G. Shin, M. L. Stitzel, P. Vera-Licona, D. Ucar. “QuIN: A Web Server for Querying and Visualizing Chromatin Interaction Networks”, *PLoS Comput Biol.*, **12**:6 (2016), pp. e1004809, URL: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004809> [↑](#) ²³³
- [20] G. Li, Y. Chen, M. P. Snyder, M. Q. Zhang. “ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis”, *Nucleic Acids Res.*, 2016, pii: gkw809 (Epub ahead of print), URL: <http://nar.oxfordjournals.org/content/early/2016/09/12/nar.gkw809.long> [↑](#) ²³³
- [21] Ю. Л. Орлов, А. О. Брагин, И. В. Медведева, И. В. Гунбин, П. С. Деменков, О. В. Вишневский, В. Г. Левицкий, В. Г. Ощепков, Н. Л. Подколотный, Д. А. Афонников, И. Гроссе, Н. А. Колчанов. «ICGenomics: программный комплекс анализа

- символьных последовательностей геномики», *Вавиловский журнал генетики и селекции*, **16**:4/1 (2012), с. 732–741, URL: http://www.bionet.nsc.ru/vogis/pict_pdf/2012/16_4_1/03.pdf ↑²³³
- [22] Б. М. Глинский, Н. В. Кучин, И. Г. Черных, Ю. Л. Орлов, Н. Л. Подколотный, В. А. Лихошвай, Н. А. Колчанов. «Суперкомпьютерные технологии в решении задач биоинформатики», *Программные системы: теория и приложения*, **6**:4(27) (2015), с. 99–112, URL: http://psta.pstiras.ru/read/psta2015_4_99-112.pdf ↑²³³
- [23] Н. С. Сафронова, М. П. Пономаренко, И. И. Абнизова, Г. В. Орлова, И. В. Чадаева, Ю. Л. Орлов. «Фланкирующие повторы мономеров определяют пониженную контекстную сложность сайтов однонуклеотидных полиморфизмов в геноме человека», *Вавиловский журнал генетики и селекции*, **19**:6 (2015), с. 668–674, URL: http://www.bionet.nsc.ru/vogis/download/19-6/05_Safronova_Rus_2.pdf ↑²³³
- [24] Y. L. Orlov, R. Te Boekhorst, I. I. Abnizova. «Statistical measures of the structure of genomic sequences: entropy, complexity, and position information», *J. Bioinform. Comput. Biol.*, **4** (2006), с. 523–536. ↑²³³
- [25] Ю. Л. Орлов, В. Г. Левицкий, О. Г. Смирнова, О. А. Подколотная, Т. М. Хлебодарова, Н. А. Колчанов. «Статистический анализ последовательностей ДНК, содержащих сайты формирования нуклеосом», *Биофизика*, **51** (2006), с. 608–614. ↑²³³
- [26] W. S. Goh, Y. Orlov, J. Li, N. D. Clarke. «Blurring of high-resolution data shows that the effect of intrinsic nucleosome occupancy on transcription factor binding is mostly regional, not local», *PLoS Comput Biol.*, **6**:1 (2010), e1000649, URL: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000649> ↑²³³
- [27] Ю. Г. Матушкин, В. Г. Левицкий, В. С. Соколов, В. А. Лихошвай, Ю. Л. Орлов. «Эффективность элонгации генов дрожжей коррелирует с плотностью нуклеосомной упаковки в 5′-нетранслируемом районе», *Математическая биология и биоинформатика*, **8**:1 (2013), с. 248–257, URL: http://www.matbio.org/2013/Matushkin_8_248.pdf ↑²³³
- [28] C. L. Winata, I. Kondrychyn, V. Kumar, K. G. Srinivasan, Y. Orlov, A. Ravishankar, S. Prabhakar, L. W. Stanton, V. Korzh, S. Mathavan. «Genome-wide analysis reveals Zic3 interaction with distal regulatory elements of stage specific developmental genes in zebrafish», *PLOS Genetics*, **9**:10 (2013), e1003852, URL: <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003852> ↑²³³

- [29] J. Choy, M. J. Fullwood. “Deciphering Noncoding RNA and Chromatin Interactions: Multiplex Chromatin Interaction Analysis by Paired-End Tag Sequencing (mChIA-PET)”, *Methods Mol Biol.*, **1468** (2017), pp. 63–89, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5048066/pdf/btw368.pdf> ↑ ²³³
- [30] G. G. Yardımcı, W. S. Noble. “Software Tools for Visualizing Hi-C Data”, *Genome Biol.*, **18** (2017), 26, 9 p., URL: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1161-y> ↑ ²³³
- [31] J. Paulsen, G. K. Sandve, S. Gundersen, T. G. Lien, K. Trengereid, E. Hovig. “Hibrowse: Multi-Purpose Statistical Analysis of Genome-Wide Chromatin 3D Organization”, *Bioinformatics*, **30** (2014), pp. 1620–1622, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4029040/pdf/btu082.pdf> ↑ ²³³
- [32] N. C. Durand, J. T. Robinson, M. S. Shanim, I. Machol, J. P. Mesirov, E. S. Lander et al.. “Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom”, *Cell Syst.*, **3**:1 (2016), pp. 99–101, URL: [http://www.cell.com/cell-systems/pdf/S2405-4712\(15\)00054-X.pdf](http://www.cell.com/cell-systems/pdf/S2405-4712(15)00054-X.pdf) ↑ ²³³
- [33] Y. Wang, B. Zhang, L. Zhang, L. An, J. Xu, D. Li, M. Choudhary, Y. Li, M. Hu, R. Hardison, T. Wang, F. Yue. *The 3D Genome Browser: a Web-Based Browser for Visualizing 3D Genome Organization and Long-Range Chromatin Interactions*, Biorxiv, no. 112268, 2017, URL: <http://biorxiv.org/content/early/2017/02/27/112268.full.pdf> ↑ ²³³
- [34] N. Haddad, C. Vaillant, D. Jost. “IC-Finder: Inferring Robustly the Hierarchical Organization of Chromatin Folding”, *Nucleic Acids Res.*, 2017, Jan 26, pii: gkx036, URL: http://membres-timc.imag.fr/Daniel.Jost/DJ-TIMC/Publications_files/Haddad_NAR_2017_main+SI.pdf ↑ ²³³

Рекомендовал к публикации

Программный комитет

Пятого национального суперкомпьютерного форума **НСКФ-2016**

Пример ссылки на эту публикацию:

Е. В. Кулакова, А. М. Спицина, А. Г. Богомолов и др. «Программа анализа геномного распределения хромосомных контактов в ядре клетки по данным, полученным по технологиям ChIA-PET и Hi-C», *Программные системы: теория и приложения*, 2017, **8**:1(32), с. 219–242.

URL: http://psta.psir.ru/read/psta2017_1_219-242.pdf

Об авторах:



Екатерина Викторовна Кулакова

Программист, окончила ФИТ НГУ в 2016 г. Область научных интересов: биоинформатика, суперкомпьютерные вычисления

e-mail: kylakovaekaterina@gmail.com



Анастасия Михайловна Спицина

Аспирант ФИТ НГУ. Область научных интересов: биоинформатика, суперкомпьютерные вычисления

e-mail: anastasia.spitsina@gmail.com



Антон Геннадьевич Богомолов

м.н.с. ИЦиГ СО РАН. Окончил Новосибирский Государственный Университет в 2009г. Область научных интересов: биоинформатика, FISH гибридизация, структура хромосом

e-mail: mantis_anton@bionet.nsc.ru



Нина Геннадьевна Орлова

к.ф.-м.н., старший научный сотрудник НГУ, доцент НГАСУ (Сибстрин). Область научных интересов: статистика

e-mail: orlovanina2@mail.ru



Артур Игоревич Дергилев

Магистрант ММФ НГУ. Область научных интересов: биоинформатика, суперкомпьютерные вычисления

e-mail: arturd1993@yandex.ru



Ирина Витальевна Чадаева

м.н.с. ИЦиГ СО РАН. Область научных интересов: генетика поведения, биоинформатика, компьютерная геномика

e-mail: ichadaeva@bionet.nsc.ru



Владимир Николаевич Бабенко

с.н.с., к.б.н., ИЦиГ СО РАН. Область научных интересов: биоинформатика, альтернативный сплайсинг, компьютерная геномика

e-mail: bob@bionet.nsc.ru



Юрий Львович Орлов

д.б.н., проф. РАН, зав. лабораторией компьютерной геномики ФЕН НГУ, зав.лаб. нейроинформатики поведения ИЦиГ СО РАН. Область научных интересов: биоинформатика, компьютерная геномика

e-mail: orlov@bionet.nsc.ru

Ekaterina Kulakova, Anastasia Spitsina, Anton Bogomolov, Nina Orlova, Artur Dergilev, Irina Chadaeva, Vladimir Babenko, Yuriy Orlov. *Program for analysis of genome distribution of chromosome contacts in cell nucleus by the data obtained using ChIA-PET and Hi-C technologies.*

ABSTRACT. Due to the growth of genomics data there is an increasing need to develop new software for the processing and analysis of genomic data on chromosome contacts, transcription factor binding sites, and the genome features annotations based on modern sequencing technologies. We consider the technology of obtaining and analyzing genomic sequencing data based on ChIA-PET and Hi-C technologies. The computer program for statistical data processing and analysis methods for chromosome contacts data is presented. (*In Russian*).

Key words and phrases: bioinformatics, sequencing, supercomputing, chromosome contact, graphical interface, databases.

-
- © E. V. KULAKOVA⁽¹⁾ A. M. SPITSINA⁽²⁾ A. G. BOGOMOLOV⁽³⁾ N. G. ORLOVA⁽⁴⁾ A. I. DERGILEV⁽⁵⁾
 I. V. CHADAeva⁽⁶⁾ V. N. BABENKO⁽⁷⁾ Y. L. ORLOV⁽⁸⁾ 2017
 © NOVOSIBIRSK STATE UNIVERSITY^(1, 2, 4, 5) 2017
 © INSTITUTE OF CYTOLOGY AND GENETICS SB RAS^(3, 6, 7, 8) 2017
 © PROGRAM SYSTEMS: THEORY AND APPLICATIONS, 2017

References

- [1] Yu. L. Orlov. “Computer-Assisted Study of the Regulation of Eukaryotic Gene Transcription on the Base of Data on Chromatin Sequencing and Precipitation”, *Vavilov Journal of Selection and Breeding*, **18**:1 (2014), pp. 193–206 (in Russian), URL: <http://vavilov.elpub.ru/index.php/jour/article/view/240>
- [2] A. M. Spitsina, Yu. L. Orlov, N. N. Podkolodnaya et al. “Supercomputer Analysis of Genomics and Transcriptomics Data Revealed by High-Throughput DNA Sequencing”, *Program Systems: Theory and Applications*, **6**:1(23) (2015), pp. 157–174 (in Russian), URL: http://psta.psisras.ru/read/psta2015_1_157-174.pdf
- [3] E. Lieberman-Aiden, N. L. Van Berkum et al. “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome”, *Science*, **326**:5950 (2009), pp. 289–293.
- [4] J. Dekker, K. Rippe, M. Dekker, N. Kleckner. “Capturing Chromosome Conformation”, *Science*, **295**:5558 (2002), pp. 1306–1311.
- [5] M. J. Fullwood, M. H. Liu, Y. F. Pan et al. “An Oestrogen-Receptoralpha-Bound Human Chromatin Interactome”, *Nature*, **462**:7269 (2009), pp. 58–64.
- [6] G. Li, X. Ruan, R. K. Auerbach et al. “Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation”, *Cell*, **148**:1–2 (2012), pp. 84–98, URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3339270/>
- [7] L. Handoko, H. Xu, G. Li et al. “CTCF-Mediated Functional Chromatin Interactome in Pluripotent Cells”, *Nat. Genet.*, **43**:7 (2011), pp. 630–638.
- [8] C. Li, X. Dong, H. Fan et al. “The 3DGD: a Database of Genome 3D Structure”, *Bioinformatics*, **30**:11 (2014), pp. 1640–1642.
- [9] L. Teng, B. He, J. Wang, K. Tan. “4DGenome: a Comprehensive Database of Chromatin Interactions”, *Bioinformatics*, **31**:15 (2015), pp. 2560–2564, URL: <https://academic.oup.com/bioinformatics/article-pdf/31/15/2560/5026291/btv158.pdf>
- [10] E. V. Kulakova, A. G. Bogomolov, A. O. Bragin, E. R. Galiyeva, G. Li, Yu. L. Orlov. “Computer Methods for Detection of Chromosomal Contacts in the Cell Nucleus Using Sequencing Data (Review)”, *Molecular Biology*, 2017 (to appear) (in Russian).
- [11] K. Kruse, C. B. Hug, B. Hernández-Rodríguez, J. M. Vaquerizas. “TADtool: Visual Parameter Identification for TAD-Calling Algorithms”, *Bioinformatics*, **32**:20 (2016), pp. 3190–3192, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5048066/pdf/btw368.pdf>
- [12] M. J. Rowley, V. G. Corces. “Minute-Made Data Analysis: Tools for Rapid Interrogation of Hi-C Contacts”, *Mol Cell.*, **64**:1 (2016), pp. 9–11.
- [13] A. T. Lun, M. Perry, E. Ing-Simmons. “Infrastructure for Genomic Interactions: Bioconductor Classes for Hi-C, ChIA-PET and Related Experiments. Version 2”, *F1000Res.*, **5** (2016), pp. 950, URL: <https://f1000research.com/articles/5-950/v2/pdf>
- [14] Z. Xu, G. Zhang, C. Wu, Y. Li, M. Hu. “FastHiC: a Fast and Accurate Algorithm to Detect Long-Range Chromosomal Interactions from Hi-C Data”, *Bioinformatics*, **32**:17 (2016), pp. 2692–2695.
- [15] N. R. Battulin, V. S. Fishman, Yu. L. Orlov, A. G. Menzorov, D. A. Afonnikov, O. L. Serova. “3C-Based Methods for 3D Genome Organization Analysis”, *Vavilov*

- Journal of Selection and Breeding*, **16**:4/2 (2012), pp. 872–876 (in Russian), URL: <http://vavilov.elpub.ru/jour/article/view/85>
- [16] N. Battulin, V.S. Fishman, A.M. Mazur et al. “Comparison of the Three-Dimensional Organization of Sperm and Fibroblast Genomes Using the Hi-C Approach”, *Genome Biol.*, **16**:1 (2015), pp. 77.
- [17] D.S. Neems, A.G. Garza-Gongora, E.D. Smith, S.T. Kosak. “Topologically Associated Domains Enriched for Lineage-Specific Genes Reveal Expression-Dependent Nuclear Topologies During Myogenesis”, *Proc. Natl. Acad. Sci. USA*, **113**:12 (2016), pp. E1691–E1700, URL: <http://www.pnas.org/content/113/12/E1691.long>
- [18] P. Szalaj, P. J. Michalski, P. Wróblewski, Z. Tang, M. Kadlof, G. Mazzocco, Y. Ruan, D. Plewczynski. “3D-GNOME: an Integrated Web Service for Structural Modeling of the 3D Genome”, *Nucleic Acids Res.*, **44**:W1 (2016), pp. W288–W293, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987952/pdf/gkw437.pdf>
- [19] A. Thibodeau, E. J. Márquez, O. Luo, Y. Ruan, F. Menghi, D.G. Shin, M. L. Stitzel, P. Vera-Licona, D. Ucar. “QuIN: A Web Server for Querying and Visualizing Chromatin Interaction Networks”, *PLoS Comput Biol.*, **12**:6 (2016), pp. e1004809, URL: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004809>
- [20] G. Li, Y. Chen, M.P. Snyder, M.Q. Zhang. “ChIA-PET2: a Versatile and Flexible Pipeline for ChIA-PET Data Analysis”, *Nucleic Acids Res.*, 2016, pii: gkw809 (Epub ahead of print), URL: <http://nar.oxfordjournals.org/content/early/2016/09/12/nar.gkw809.long>
- [21] Yu.L. Orlov, A.O. Bragin, I.V. Medvedeva, I.V. Gunbin, P.S. Demenkov, O.V. Vishnevskiy, V.G. Levitskiy, V.G. Oshchepkov, N.L. Podkolodnyy, D.A. Afonnikov, I. Grosse, N.A. Kolchanov. “ICGenomics: a Program Complex for Analysis of Symbol Sequences in Genomics”, *Vavilov Journal of Selection and Breeding*, **16**:4/1 (2012), pp. 732–741 (in Russian), URL: http://www.bionet.nsc.ru/vogis/pict_pdf/2012/16_4_1/03.pdf
- [22] B.M. Glinskiy, N.V. Kuchin, I.G. Chernykh, Yu.L. Orlov, N.L. Podkolodnyy, V.A. Likhoshvay, N.A. Kolchanov. “Bioinformatics and High Performance Computing”, *Program Systems: Theory and Applications*, **6**:4(27) (2015), pp. 99–112 (in Russian), URL: http://psta.psir.ru/read/psta2015_4_99-112.pdf
- [23] N.S. Safronova, M.P. Ponomarenko, I.I. Abnizova, G.V. Orlova, I.V. Chadayeva, Yu.L. Orlov. “Flanking Monomer Repeats Define Lower Context Complexity of Sites Containing Single Nucleotide Polymorphisms in the Human Genome”, *Vavilov Journal of Selection and Breeding*, **19**:6 (2015), pp. 668–674 (in Russian), URL: http://www.bionet.nsc.ru/vogis/download/19-6/05_Safronova_Rus_2.pdf
- [24] Y.L. Orlov, R. Te Boekhorst, I.I. Abnizova. “Statistical Measures of the Structure of Genomic Sequences: Entropy, Complexity, and Position Information”, *J. Bioinform. Comput. Biol.*, **4** (2006), pp. 523–536.
- [25] Yu.L. Orlov, V.G. Levitskiy, O.G. Smirnova, O.A. Podkolodnaya, T.M. Khlebodarova, N.A. Kolchanov. “Statistical Analysis of DNA Sequences Containing Nucleosome Positioning Sites”, *Biophysics*, **51**:4 (2006), pp. 541–546.
- [26] W.S. Goh, Y. Orlov, J. Li, N.D. Clarke. “Blurring of High-Resolution Data Shows that the Effect of Intrinsic Nucleosome Occupancy on Transcription Factor Binding is Mostly Regional, not Local”, *PLoS Comput Biol.*, **6**:1

- (2010), e1000649, URL: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000649>
- [27] Yu. G. Matushkin, V. G. Levitskiy, V. S. Sokolov, V. A. Likhoshvay, Yu. L. Orlov. “Yeast Gene Elongation Efficiency Correlates with Nucleosome Formation in 5'-Untranslated Region”, *Mathematical Biology and Bioinformatics*, **8**:1 (2013), pp. 248–257 (in Russian), URL: http://www.matbio.org/2013/Matushkin_8_248.pdf
- [28] C. L. Winata, I. Kondrychyn, V. Kumar, K. G. Srinivasan, Y. Orlov, A. Ravishankar, S. Prabhakar, L. W. Stanton, V. Korzh, S. Mathavan. “Genome-Wide Analysis Reveals Zic3 Interaction with Distal Regulatory Elements of Stage Specific Developmental Genes in Zebrafish”, *PLOS Genetics*, **9**:10 (2013), e1003852, URL: <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003852>
- [29] J. Choy, M. J. Fullwood. “Deciphering Noncoding RNA and Chromatin Interactions: Multiplex Chromatin Interaction Analysis by Paired-End Tag Sequencing (mChIA-PET)”, *Methods Mol Biol.*, **1468** (2017), pp. 63–89, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5048066/pdf/btw368.pdf>
- [30] G. G. Yardımcı, W. S. Noble. “Software Tools for Visualizing Hi-C Data”, *Genome Biol.*, **18** (2017), 26, 9 p., URL: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1161-y>
- [31] J. Paulsen, G. K. Sandve, S. Gundersen, T. G. Lien, K. Trengereid, E. Hovig. “Hibrowse: Multi-Purpose Statistical Analysis of Genome-Wide Chromatin 3D Organization”, *Bioinformatics*, **30** (2014), pp. 1620–1622, URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4029040/pdf/btu082.pdf>
- [32] N. C. Durand, J. T. Robinson, M. S. Shanim, I. Machol, J. P. Mesirov, E. S. Lander et al. “Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom”, *Cell Syst.*, **3**:1 (2016), pp. 99–101, URL: [http://www.cell.com/cell-systems/pdf/S2405-4712\(15\)00054-X.pdf](http://www.cell.com/cell-systems/pdf/S2405-4712(15)00054-X.pdf)
- [33] Y. Wang, B. Zhang, L. Zhang, L. An, J. Xu, D. Li, M. Choudhary, Y. Li, M. Hu, R. Hardison, T. Wang, F. Yue. *The 3D Genome Browser: a Web-Based Browser for Visualizing 3D Genome Organization and Long-Range Chromatin Interactions*, Biorxiv, no. 112268, 2017, URL: <http://biorxiv.org/content/early/2017/02/27/112268.full.pdf>
- [34] N. Haddad, C. Vaillant, D. Jost. “IC-Finder: Inferring Robustly the Hierarchical Organization of Chromatin Folding”, *Nucleic Acids Res.*, 2017, Jan 26, pii: gkx036, URL: http://membres-timc.imag.fr/Daniel.Jost/DJ-TIMC/Publications_files/Haddad_NAR_2017_main+SI.pdf

Sample citation of this publication:

Ekaterina Kulakova, Anastasia Spitsina, Anton Bogomolov, Nina Orlova, Artur Dergilev, Irina Chadaeva, Vladimir Babenko, Yuriy Orlov. “Program for analysis of genome distribution of chromosome contacts in cell nucleus by the data obtained using ChIA-PET and Hi-C technologies”, *Program systems: Theory and applications*, 2017, **8**:1(32), pp. 219–242. (In Russian).

URL: http://psta.psir.ru/read/psta2017_1_219-242.pdf