

ОТЗЫВ

официального оппонента на диссертационную работу Орлова Юрия Львовича на тему: **«Полногеномный компьютерный анализ распределения сайтов связывания транскрипционных факторов эукариот по данным иммунопреципитации хроматина и высокопроизводительного секвенирования»**, представленную на соискание ученой степени доктора биологических наук по специальности 03.01.09 - Математическая биология, биоинформатика.

Изучение регуляции экспрессии генов – актуальнейшая проблема современной биологии. Появление высокопроизводительных методов секвенирования ДНК и измерения экспрессии генов сделало возможным изучение этой проблемы в масштабах всего генома и на разных уровнях его организации – на уровне ДНК, хромосомы и трехмерной организации хромосом. Важную роль в таких исследованиях играют методы биоинформатики и компьютерной биологии. Работа Ю.Л. Орлова посвящена разработке таких методов для компьютерной реконструкции структуры регуляторных районов, контролирующей транскрипцию генов эукариот на основе анализа данных о положении сайтов связывания транскрипционных факторов в геноме, полученных с помощью технологий иммунопреципитации хроматина и высокопроизводительного секвенирования (ChIP-seq). Рассмотрены задачи предсказания положения регуляторных районов на основе данных ChIP-seq и анализа положения дистальных регуляторных районов генов, определяемых хромосомными контактами. Следует отметить, что вопросы анализа регуляции экспрессии генов в масштабе генома эукариот остаются сложными для исследования, поскольку продолжающееся быстрое развитие технологий секвенирования ДНК, методов экспериментального определения сайтов связывания

транскрипционных факторов и методов измерения уровней экспрессии генов опережает развитие компьютерных подходов анализа больших объемов получаемых полногеномных данных. Таким образом, актуальность и важность выбранной темы исследования не вызывает сомнений.

Структура диссертации всесторонне и логически последовательно позволяет соискателю изложить свои научные взгляды, сформулировать цели работы и описать методы и результаты исследований. Работа состоит из введения, обзора литературы, четырех глав, содержащих результаты исследования, заключения и обсуждения результатов, выводов, списка публикаций по теме диссертации, списка цитированной литературы (521 ссылка) и приложения. Объем диссертации составляет 343 машинописных страницы, включая 119 рисунков и 28 таблиц. По результатам работы автором опубликовано 30 статей, в том числе и в ведущих международных журналах с высоким импакт-фактором. Еще 6 статей опубликовано в сборниках и монографиях. Эти публикации с достаточной полнотой отражают содержание работы и позволяют оценить ее новизну, практическую и научную значимость.

Во введении автором сформулированы цели и научные задачи исследования. Детально показан личный вклад соискателя в работу в совместных публикациях с другими соавторами, относящийся к биоинформатике и компьютерному анализу данных, что соответствует специальности диссертационной работы. Приведены основные научные результаты, выносимые на защиту. Отмечается особая актуальность анализа экспериментальных данных секвенирования в масштабе генома и на разных уровнях его организации.

Первая глава представляет собой обзор литературы. Формулируется задача компьютерного анализа геномных данных, описана организация регуляторных участков геномов у эукариот и проблемы анализа таких районов. Приведена информация о современных методах исследования регуляторных районов генов в масштабе генома, включая определение сайтов связывания с помощью технологий иммунопреципитации хроматина (ChIP). Представлены исследования по регуляции экспрессии генов, связанных с образованием опухолей (ESR1, MYC, TP53), тканеспецифичной экспрессии в клеточных культурах. Описаны подходы к изучению эмбриональных стволовых клеток (ЭСК) человека и мыши, показана роль транскрипционных факторов плюрипотентности в репрограммировании. Дан обзор проблем исследования трехмерных контактов хромосом в ядре с помощью секвенирования (методы 3C, Hi-C и ChIA-PET).

Вторая глава «Модели распределения сайтов связывания транскрипционных факторов в геноме» содержит описание разработанных методов и компьютерных моделей. В частности, разработаны оригинальные и эффективные компьютерные методы анализа распределения сайтов связывания транскрипционных факторов в эукариотическом геноме на основе анализа профилей ChIP-seq. Очень интересен анализ качества наборов проб микрочипа Affymetrix U133. Автор убедительно показал, что потенциал микрочиповых данных может быть использован гораздо полнее через интеграцию геномной аннотации и клинических данных. Эти результаты имели важное значение и широко использовались в ряде научных публикаций. Используя разработанные компьютерные программы для микрочипов и базу данных наборов проб микрочипов платформы Affymetrix U133, автор также выполнил анализ экспрессии транскриптов в цис-антисенс ориентации. Описаны общие средства компьютерной интеграции геномных данных, разработанные в ИЦиГ СО РАН, включая программный комплекс

ICGenomics.

Третья, четвертая и пятая главы - «Карты сайтов связывания по данным ChIP-seq», «Модификации хроматина и связывание транскрипционных факторов по данным ChIP-seq», «Хромосомные контакты и регуляция транскрипции в геноме человека» - описывают применение разработанных средств для анализа сайтов связывания транскрипционных факторов (ССТФ) в ЭСК человека и мыши, построение полногеномных карт, анализ распределения сайтов связывания рецептора эстрогенов ER α . В четвертой главе анализ ССТФ рассмотрен в контексте структуры хроматина и модификаций гистонов, в пятой – с точки зрения хромосомных контактов.

Третья глава «Карты сайтов связывания по данным ChIP-seq» посвящена описанию карт сайтов связывания транскрипционных факторов, построенных автором по экспериментальным данным ChIP-seq в геномах человека, мыши и *D.rerio*. Важно, что полногеномные карты связывания транскрипционных факторов c-Myc, Oct4, Nanog, Sox2, E2f1, n-Myc, Tbx3, Eset, Nr5a2, Smad2 у мыши, фактора Zic3 у рыбы *Danio rerio* и транскрипционных факторов MYC, ER α , PRDM14 у человека получены впервые. Более того, разработанные автором компьютерные программы интеграции данных о геномной локализации указанных выше ССТФ и уровнях экспрессии генов, измеренных с помощью микрочипов, позволили получить новые данные о регуляции транскрипции генов и найти новые мишени для проанализированных транскрипционных факторов. Автор впервые продемонстрировал существование кластеров сайтов связывания факторов Oct4-Nanog-Sox2, используя оригинальный подход, основанный на матрицах сближенности ССТФ. Представлены аналогичные кластеры связывания OCT4-NANOG-SOX2 в ЭСК в геноме человека. Определен новый нуклеотидный мотив сайта связывания транскрипционного фактора

PRDM14 в геноме человека. Изложенные в этой главе пионерские результаты дают представление о транскрипционных контурах и цис-регуляторных модулях, в которые вовлечены проанализированные регуляторы транскрипции.

Четвертая глава «Модификации хроматина и связывание транскрипционных факторов по данным ChIP-seq» содержит описание применения разработанных компьютерных методов к исследованию модификаций хроматина и связыванию транскрипционных факторов в геноме дрожжей и в геноме человека. Соискателем проанализированы полногеномные данные по модификациям гистонов (ацетилирования и метилирования гистона H3) и рассчитаны позиции положения нуклеосом в геноме дрожжей на основе данных секвенирования защищенных нуклеосомой фрагментов ДНК. В результате этого исследования был получен интересный результат: оказалось, что большинство сайтов связывания транскрипционных факторов в промоторных районах генов дрожжей, определенных с помощью технологии ChIP-chip, свободны от нуклеосомной упаковки. Предложен новый подход к предсказанию сайтов связывания ER α в масштабе генома на основе профилей модификаций гистонов (H3K4me3, H3K4me1, H3K27me3, H3K9me3, H3K9ac, H3K14ac), определенных с помощью технологии ChIP-seq.

Пятая глава «Хромосомные контакты и регуляция транскрипции в геноме человека» представляет пионерские исследования хромосомных контактов, полученных с помощью массового параллельного секвенирования нуклеотидных последовательностей контактирующих участков хромосом по методу ChIA-PET для ER α и комплекса РНК-полимеразы II в геноме человека. Предложена классификация групп генов, находящихся в транскрипционных доменах, в зависимости от структуры контактов

(хромосомных петель). Интересным и имеющим важное значение результатом является демонстрация присутствия в участках хромосомных контактов, опосредованных комплексом РНК-полимеразы II, сайтов связывания различных транскрипционных факторов, определенных с помощью технологии ChIP-seq в геноме. Важным результатом является обнаружение положительной корреляции участков хромосомных контактов с модификациями гистонов, характеризующими открытое состояние хроматина (H3K4me3, H3K9ac, H3K4me1).

В Приложении приведены схемы алгоритмов, таблицы, содержащие координаты сайтов в геноме, результаты анализа кластеризации ССТФ, описание использованных компьютерных ресурсов.

В заключении приводятся основные результаты диссертационной работы.

По работе есть несколько существенных вопросов и замечаний:

1. Используя программы интеграции данных о геномной локализации указанных ССТФ и уровнях экспрессии генов, измеренных с помощью микрочипов, автор получил новые данные о регуляции транскрипции генов и нашел новые мишени для проанализированных транскрипционных факторов. Вместе с тем известно, что существование хорошего пика ChIP-seq само по себе не является свидетельством функциональной значимости обнаруженного взаимодействия, а увеличение или уменьшение экспрессии при выключении гена может быть опосредованным, т.е. результатом взаимодействия выключенного гена с другими генами. В связи с этим возникает вопрос какие из обнаруженных новых взаимодействий являются «предсказаниями», а какие подтверждаются другими экспериментальными и теоретическими результатами?

2. Каков процент функционально значимых сайтов среди сильных сайтов, выявленных с помощью ChIP-seq? Могут ли пригодиться для такой оценки данные о хромосомных контактах, полученные с помощью ChIA-PET?

3. На веб-сайте базы данных RatDNA экспрессии генов на микрочипах, разработанной автором диссертации, по микрочипам не поддерживается функция поиска генов, а таблица с данными по экспрессии генов вовсе отсутствует.

4. Автором продемонстрирована значимая корреляция между модификациями открытого хроматина и высотой пиков ChIP-seq связывания рецептора эстрогенов ER α , активирующего транскрипцию генов–мишеней. В связи с этим хочется узнать мнение автора о том, можно ли с помощью меток H3K9me3 и H3K27me3 о закрытом состоянии хроматина предсказать сайты посадки репрессоров на ДНК?

Есть замечания по оформлению и опечаткам. Однако эти огрехи не умаляют общего очень хорошего впечатления от работы и не снижают ее высокой научной значимости.

Диссертация Ю.Л.Орлова является законченным научно-исследовательским трудом, выполненным автором самостоятельно на высоком научном уровне. В работе приведены научные результаты, позволяющие квалифицировать их как новое теоретическое знание о принципах организации регуляторных районов генов эукариот и о механизмах экспрессии генов. Полученные автором результаты достоверны, выводы и заключения обоснованы.

Работа базируется на достаточном числе исходных данных, примеров и расчетов, представленных в высокорейтинговых публикациях. По каждой главе диссертации и работе в целом сделаны четкие выводы.

Автореферат соответствует основному содержанию диссертации.

Таким образом, диссертация Орлова Юрия Львовича «**Полногеномный компьютерный анализ распределения сайтов связывания транскрипционных факторов эукариот по данным иммунопреципитации хроматина и высокопроизводительного секвенирования** является законченной научно-квалификационной работой и отвечает требованиям п.9 Положения "О порядке присуждения ученых степеней" N 842 от 24 сентября 2013 г., а ее автор заслуживает присуждения искомой степени доктора биологических наук по специальности 03.01.09 - Математическая биология, биоинформатика.

Официальный оппонент

Мария Георгиевна Самсонова,

д.б.н., заведующая лабораторией «Математическая биология и биоинформатика» Института прикладной математики и механики и профессор кафедры Прикладная математика Федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный политехнический университет» (ФГАОУ ВО «СПбПУ»), 29, Политехническая ул., Санкт-Петербург, 195251 Россия, тел. 7(812)5962831, samson@spbcas.ru

« 7 » *июль* 2014 г.

