

Отзыв официального оппонента

о диссертации Ю.Л. Орлова

«Полногеномный компьютерный анализ распределения сайтов связывания транскрипционных факторов эукариот по данным иммунопреципитации хроматина и высокопроизводительного секвенирования»,
представленной на соискание ученой степени доктора биологических наук по специальности 03.01.09 - Математическая биология, биоинформатика

Диссертационная работа рассматривает фундаментальную проблему молекулярной биологии - проблему молекулярных механизмов контроля экспрессии генов эукариот. Понимание таких регуляторных механизмов, исследование их современными компьютерными методами, имеет не только теоретическое, но и большое практическое значение для задач биотехнологии и биомедицины. 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.

Работа Ю.Л. Орлова представляет новые компьютерные методы исследования сайтов связывания транскрипционных факторов и, в более широком аспекте, регуляторных районов генов эукариот, на основе анализа данных экспериментальных технологий, использующих иммунопреципитацию хроматина и высокопроизводительное секвенирование ДНК. Представлены задачи предсказания локализации в геноме сайтов связывания на основе данных ChIP-seq, а также анализ положения дистальных регуляторных районов генов, определяемых хромосомными контактами, с помощью технологии ChIA-PET, также основанной на иммунопреципитации хроматина. Избранная диссертантом тема, несомненно, имеет большую **актуальность**.

Помимо компьютерной составляющей исследования в целом, можно выделить две биологические задачи, в которых применялись разработанные методы компьютерного анализа. Это изучение транскрипционных факторов, регулирующих гены, вовлеченные в процессы канцерогенеза – *MYC*, *ESR1*, *TP53*, *FOXA1*. Компьютерное исследование полногеномного распределения сайтов связывания таких транскрипционных факторов имеет большое значения для поиска их генов-мишеней, маркеров для диагностики заболеваний. Изучение генов-мишеней онкогенов *MYC* и транскрипционного фактора

рецептора эстрогенов ER α , определенных по положению соответствующих транскрипционных факторов с помощью технологий ChIP-PET и ChIP-seq в геноме человека важно для понимания молекулярных механизмов развития онкологических заболеваний. Вторая важная проблема - исследование механизмов транскрипционной регуляции генов, обеспечивающих поддержание плюрипотентного состояния стволовых клеток, также относится к фундаментальным медицинским задачам. Исследование механизмов репрограммирования стволовых клеток, реконструкция регуляторных генных основано на построении карт локализации сайтов связывания транскрипционных факторов (Nanog, Oct4, Sox2, Klf4), в эмбриональных стволовых клетках.

В диссертационной работе Ю.Л. Орлова рассмотрены известные данные по полногеномному анализу сайтов связывания транскрипционных факторов в геномах эукариот. В целом в диссертации детально проработана литература, как по компьютерным методам, так и по экспериментальным технологиям секвенирования ДНК, по технологиям определения хромосомных контактов, по исследованию эмбриональных стволовых клеток, репрограммированию клеток, которое изучается с точки зрения действия транскрипционных факторов.

Работы по определения сайтов связывания опирались на опубликованные данные, полученные с помощью технологий ChIP-on-chip, ChIP-PET, ChIP-seq. В тоже время многие данные представленные в работе по определению сайтов связывания транскрипционных факторов, впервые были получены автором в его публикациях, развивающих методы ChIP-PET – для транскрипционного фактора MYC, ChIP-seq - для серии транскрипционных факторов в ЭСК мыши. Обоснованность научных положений подтверждается публикациями автора диссертации в самых высокорейтинговых журналах – **Nature** и **Cell**.

Объем диссертации составляет 343 машинописных страницы, включая 119 рисунков и 28 таблиц. Список литературы содержит 521 ссылку.

В работе диссертант грамотно использует математический и статистический аппарат, корректно вводит понятия пиков профиля ChIP-seq, уровня значимости при определении сайтов связывания транскрипционного фактора по такому профилю, понятие полноты эксперимента ChIP-seq.

Практические рекомендации, сформулированные в диссертации, полностью **обоснованы**. Публикации в высокорейтинговых международных научных журналах и в рекомендованных ВАК изданиях (33 наименования) с достаточной **полнотой** отражают содержание диссертационной работы и позволяют оценить новизну работы, подтвердить ее практическую и научную значимость.

Диссертационная работа в целом содержит оригинальные компьютерные методы анализа распределения сайтов связывания транскрипционных факторов в геноме на основе данных ChIP-seq - иммунопреципитации хроматина, сопряженной с последующим высокопроизводительным секвенированием фрагментов ДНК. Исследования были новыми на момент публикации, их **достоверность** подтверждена последующими публикациями.

Карты локализации в геноме мыши сайтов связывания ряда различных транскрипционных факторов c-Мус, Oct4, Nanog, Sox2, E2f1, n-Мус, Tbx3, Eset, Nr5a2, Smad2, опубликованные в нескольких статьях автора диссертации по экспериментам в ЭСК мыши были **получены впервые**. Рассчитаны координаты расположения сайтов связывания транскрипционных факторов MYC, ER α , FOXA1, PRDM14 в геноме человека. Транскрипционный фактор PRDM14 был изучен **впервые** с помощью метода ChIP-seq, автором диссертации впервые рассчитан нуклеотидный мотив связывания этого фактора с ДНК.

Анализ данных ChIP-seq **впервые** показал совместную локализацию сайтов связывания транскрипционных факторов Oct4, Sox2, Nanog, с одной стороны и c-Мус, n-Мус с другой, в эмбриональных стволовых клетках мыши. Выполнен анализ нуклеотидных мотивов de novo, для сайтов связывания факторов Zic3 в геноме *D.rerio*, Smad2 в геноме мыши.

Автором **впервые** построен компьютерный метод предсказания сайтов связывания транскрипционного фактора – эстрадиолового рецептора на основе профилей модификации хроматина (модификации метилирования гистонов H3K4me3, H3K4me1, H3K27me3, H3K9me3, и ацетилирования H3K9ac, H3K14ac, соответственно), определенных с помощью технологии ChIP-seq.

Впервые с помощью разработанных компьютерных программ совместно с данными ChIP-seq проанализированы карты хромосомных контактов, полученные посредством технологии ChIA-PET, для рецептора эстрогенов ER α . Сам компьютерный анализ данных современной технологии ChIA-PET, основанной на иммунопреципитации контактирующих хромосомных участков с помощью секвенирования парных концов (PET), также описан впервые в статье автора в журнале Nature в 2009 году.

Диссертантом в качестве новых научных результатов выдвинуты ряд положений, включающих указанные выше позиции. **Впервые** на основе компьютерного анализа полногеномных данных о хромосомных контактах, опосредованных комплексами РНК-полимеразы II, данных транскрипционной активности генов, и профилей модификаций гистонов для пяти клеточных линий в геноме человека показана положительная

корреляция участков хромосомных контактов с модификациями гистонов, характеризующими открытое состояние хроматина.

В целом, результаты, полученные автором, являются **новыми** научными знаниями по специальности математическая биология, биоинформатика. Основные результаты неоднократно обсуждались на различных конференциях и симпозиумах и получили одобрение ведущих специалистов. По теме диссертации в целом представлено 52 публикации, из них 30 статей (все по списку ВАК), 5 статей в сборниках научных трудов, одна глава в монографии, 16 тезисов конференций.

Теоретическая значимость работы для науки связана с разработкой компьютерной модели распределения сайтов связывания транскрипционных факторов, позволяющей определять локализацию сайтов связывания транскрипционного фактора по координатам секвенированных прочтений в геноме и оценивать полноту эксперимента ChIP-seq. Представлена компьютерная модель хромосомных петель, образованных регуляторными районами транскрипции генов (промоторов и энхансеров) в геноме человека. Такая модель основана данных технологии ChIA-PET, по хромосомным контактам, опосредованным комплексом РНК-полимеразы II.

Практическая ценность разработанных компьютерных методов состоит в возможности поиска регуляторных районов генов по данным секвенирования в масштабе полного генома эукариот. Представлен программный комплекс ICGenomics для функциональной аннотации геномных последовательностей, который обеспечивает существенное расширение методов компьютерного анализа полногеномных данных. В работе представлены база данных качества наборов проб микрочипов Affymetrix U133, база цис-антисенс транскриптов в геноме человека, база данных экспрессии генов на микрочипах для крыс RatDNA, для которой представлено свидетельство госрегистрации баз данных.

Личный вклад автора в работу заслуживает отдельного рассмотрения. Результаты, описанные в диссертации, получены лично автором. Это обусловлено с одной стороны тем, что автор работал на конечной стадии эксперимента - анализе данных. С другой стороны автор также принимал участие и в разработке наборов проб, используемых в эксперименте и в анализе полноты проб использованных другими авторами в компьютерных базах данных. Т.о. личный вклад автора диссертации в публикации по теме **высок**. В то же время необходимо признать, что имеются и другие соавторы, которые выполняли собственно экспериментальную часть. С другой стороны, диссертация защищается по специальности "математическая биология, биоинформатика", т.е.

посвящена вопросом анализа данных. В этом отношении нельзя не признать **ведущую роль** диссертанта, во всех работах, рассмотренных в диссертации.

Рекомендации по использованию результатов и выводов диссертации

Результаты работы могут использоваться в курсе биоинформатики для студентов университетов. Разработанные методы анализа полноты эксперимента ChIP-seq с помощью компьютерной симуляции числа сайтов связывания исследуемого транскрипционного фактора в зависимости от глубины секвенирования могут быть использованы при планировании и анализе экспериментов ChIP-seq.

Компьютерный метод предсказания сайтов связывания транскрипционного фактора по профилям модификаций гистонов может быть адаптирован для предсказания положения сайтов различных транскрипционных факторов. Компьютерные базы данных аннотации наборов проб микрочипов Affymetrix U133 могут использоваться для уточнения экспрессии генов на микрочипах этой технологической платформы.

Результаты исследования по определению геномной локализации участков хромосомных контактов, промотор-промоторных взаимодействий, опосредованных комплексом РНК-полимеразы II, могут быть использованы для анализа дистальных регуляторных районов, предсказания положения сайтов связывания транскрипционных факторов в геноме человека.

Полученные результаты и компьютерные методы можно использовать в дальнейших исследованиях в научных учреждениях РАН, РАМН, занимающихся полногеномным анализом, секвенированием, и обработкой данных экспериментов ChIP-seq, при планировании и анализе подобных экспериментов.

Практические рекомендации, сформулированные в диссертации, обоснованы проведенными исследованиями и могут служить руководством в работе.

Оценка содержания диссертации, ее завершенность

Диссертация состоит из пяти глав – «Обзор литературы», «Модели распределения сайтов связывания транскрипционных факторов в геноме», «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». Введение содержит обоснование **актуальности** темы исследования, 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.

В первой главе дан обзор литературы, современных исследований регуляторных районов транскрипции в геноме человека, 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 человека и мыши, показана роль исследованных транскрипционных факторов, таких как Oct4, в поддержания плюрипотентности и репрограммировании клетки.

Во второй главе дано описание компьютерных моделей распределения сайтов связывания транскрипционных факторов в эукариотическом геноме, представлены алгоритмы анализа данных ChIP-seq и базы микрочиповых данных, разработанные соискателем.

В третьей главе показаны результаты применения разработанных автором программ к реконструкции карт сайтов связывания транскрипционных факторов, полученных по экспериментальным данным ChIP-seq.

Четвертая глава содержит описание применения разработанных компьютерных методов к исследованию нуклеосомной упаковки и модификаций хроматина в геноме человека, полученных с помощью ChIP-seq, анализ моделей сайтов связывания транскрипционного фактора - рецептора эстрогенов. В данной главе показано применение программ для анализа связывания транскрипционных факторов в геноме дрожжей.

Пятая глава представляет исследование хромосомных контактов, полученных с помощью массового параллельного секвенирования в геноме человека по методу ChIA-PET для рецептора эстрогенов и комплекса РНК-полимеразы II с помощью разработанных соискателем компьютерных программ. Показано, что геномные области хромосомных контактов, опосредованных комплексом РНК-полимеразы II, обогащены сайтами связывания транскрипционных факторов и участками модификаций гистонов, связанными с активацией экспрессии генов.

Рассматривая диссертацию в целом, следует подчеркнуть огромный объем проделанной автором работы. Представлена целая серия приложений к моделированию и анализу данных о сайтах связывания различных транскрипционных факторов для широкого набора модельных организмов эукариот.

Подводя итог, следует высказать и некоторые замечания. Анализ сайтов связывания транскрипционного фактора Zic3 слишком кратко представлен в тексте диссертации. В таблице 2.4 уровень значимости некорректно указан как «вероятность различия».

Приведенные замечания, однако, не влияют на высокую оценку представленной диссертационной работы.

Заключение

Диссертация является **законченным** научно-исследовательским трудом, выполненным автором самостоятельно на высоком научном уровне. В работе приведены научные результаты, позволяющие квалифицировать их как теоретические положения, совокупность которых можно квалифицировать как **крупное** научное достижение. 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.

Материалы диссертации **полно** отражены в публикациях автора, автореферат соответствует основному содержанию диссертации.

Работа **удовлетворяет** всем требованиям ВАК, предъявляемым к диссертациям на соискание ученой степени доктора биологических наук п. 9 «Положения о присуждении ученых степеней», а ее автор Юрий Львович Орлов, несомненно, **заслуживает** присуждения ему искомой ученой степени доктора биологических наук по специальности 03.01.09 - Математическая биология, биоинформатика.

Л.В. Омелянчук

Доктор биологических наук,

зав. Лабораторией Клеточного цикла

Института молекулярной и клеточной биологии СО РАН,

Новосибирск 630090, Лаврентьева 8/2,

ome@mcb.nsc.ru

« 7 » ноября 2014 г.

