

ОТЗЫВ

на автореферат диссертации Цуканова Антона Витальевича

«Мультимодельный подход к эффективному картированию сайтов связывания транскрипционных факторов по данным ChIP-seq экспериментов»

Регуляция экспрессии генов на уровне транскрипции лежит в основе механизмов, контролирующей реализацию различных генетических программ и сценариев на протяжении жизни любого организма. Важнейшими элементами этой регуляции являются транскрипционные факторы (ТФ). Одной из актуальных проблем в этой области является эффективное выявление сайтов связывания ТФ (ССТФ) в регуляторных районах генов, решение которой затрудняется нерегулярностью их структуры. На фоне стремительного роста количества данных полногеномного профилирования районов связывания ТФ в открытом доступе, тема диссертационной работы Цуканова Антона Витальевича, посвященной использованию мультимодельного подхода к эффективному картированию сайтов связывания транскрипционных факторов по данным ChIP-seq экспериментов, безусловно является актуальной.

Во вводной части автор интересно и убедительно обосновывает значимость и своевременность своего исследования, анализируя достоинства и недостатки наиболее популярной модели для поиска ССТФ (PWM), разбирая причины недостаточного использования альтернативных моделей, а также рассматривая перспективы их более широкого применения. Для решения этой задачи автором был создан программный комплекс MultiDeNa для выявления сайтов связывания ТФ в данных ChIP-seq с использованием методологически разных моделей. Результаты масштабного исследования данных 68 и 1004 ChIP-seq экспериментов для *A. thaliana* и *M. musculus* соответственно показали, что использование альтернативных моделей (BaMM и SiteGA) в дополнение к PWM может существенно увеличить долю пиков, в которых предсказаны сайты. Далее, на основании сравнительного анализа результатов для ТФ разных классов, выделяемых на основании структуры ДНК-связывающего домена, автору удалось показать зависимость доли пиков, распознаваемых альтернативными моделями в дополнение к PWM, от класса ТФ (наибольшее увеличение доли пиков для ТФ класса bHLH, наименьшее – для C2H2 zink finger factors). Это по предположению автора связано с различным вкладом зависимостей между позициями внутри сайта связывания в функционирование ТФ разных классов. Помимо этого, было показано, что гены, имеющие в промоторе

сайты, предсказанные только одной моделью, могут быть обогащены специфическими терминами геной онтологии (ГО). Несмотря на то, что по точности модель SiteGA уступает и PWM, и BaMM для подавляющего большинства классов ТФ, SiteGA демонстрирует более высокие значения кратности обогащения «общих» терминов ГО, что свидетельствует в пользу функциональной важности сайтов, предсказываемых с помощью SiteGA.

При прочтении автореферата возникло несколько замечаний редакционного характера, а также вопросов.

1. В тексте имеются многочисленные грамматические ошибки (преимущественно пунктуационные), а также встречаются некорректные формулировки. Например, в подписи к рис. 2 упоминаются «*пороговые значения доли пиков с мотивом*», однако в тексте автореферата отсутствует информация об установлении каких-либо критических значений для этой величины. Таким образом, некорректно называть эти значения *пороговыми*.
2. В описании результатов для ТФ класса *bHLH* на стр. 10 утверждается, что «Для этого класса ТФ модель SiteGA имеет *более высокую точность по сравнению с другими классами*, что может свидетельствовать о наличии существенного вклада зависимостей в общий паттерн нуклеотидного контекста». Это противоречит рис. 3, из которого видно, что для целого ряда классов ТФ (например, {1.1}, {2.3}, {7.2} у *A. thaliana*, {3.5} у *M. musculus*) точность модели SiteGA сравнима или превосходит таковую для ТФ класса *bHLH*. Вероятно, автор имел в виду точность модели SiteGA относительно модели PWM, в таком случае это нужно было указать.
3. На стр. 10 автор говорит о существенном вкладе зависимостей в общий паттерн нуклеотидного контекста сайтов связывания ТФ класса *bHLH* и далее объясняет это широким разнообразием функциональных гомо- и гетеродимеров для этих ТФ (и, соответственно, потенциальным разнообразием структур ССТФ). Не совсем ясно, почему структурная сложность сайтов, обусловленная функционированием ТФ в виде гомо- и гетеродимеров, должна быть сопряжена с появлением зависимостей между позициями внутри сайта связывания?

Несмотря на приведенные выше замечания, диссертация Цуканова Антона Витальевича является полноценной законченной научно-квалификационной работой, посвященной важной проблеме, а именно: картированию сайтов связывания транскрипционных факторов по данным ChIP-seq экспериментов с помощью методологически разных моделей. Диссертация Цуканова Антона Витальевича полностью соответствует требованиям пп. 9 – 14 «Положения о присуждении ученых

степеней», утвержденного Постановлением Правительства Российской Федерации №842 от 24.09.2013, а ее автор заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8 – математическая биология, биоинформатика.

13 февраля 2024 г.

Ведущий научный сотрудник Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)

кандидат биологических наук



Землянская Елена Васильевна

Контактная информация:

Адрес: 630090, г. Новосибирск, пр. Ак. Лаврентьева, д. 10

Тел.: +7 (383) 363-49-63*3113

e-mail: ezemlyanskaya@bionet.nsc.ru

