

ОТЗЫВ

официального оппонента

на диссертацию Вибе Даниила Станиславовича,
представленную на соискание ученой степени кандидата биологических наук, на
тему: «Выявление взаимосвязи между величинами изменения экспрессии и
функциями дифференциально экспрессирующихся генов на основе компьютерного
анализа транскриптомов арабидопсиса и человека»,
по специальности 1.5.8 - математическая биология, биоинформатика

Стремительное развитие методов секвенирования, происходящее в последнее десятилетие, позволило по-иному взглянуть на многие процессы, лежащие в основе функционирования живых организмов. В частности, были определены структуры многих сетей регуляции, определены мастер-гены, расположенные в их ключевых точках, охарактеризованы регуляторные элементы, такие как сайты связывания транскрипционных факторов и т.д. Эти работы сопряжены с генерацией огромного объема сырых данных по экспрессии генов (RNA-seq). Так, на момент написания отзыва число образцов RNA-seq, депонированных в международную базу данных Sequence Read Archive, превышает 2,6 миллиона. Они получены при реализации более чем 30 тысяч проектов. Параллельно с накоплением первичных данных шло развитие методов их анализа. Программы первых поколений, такие как Cufflinks, уступали более современным, таким как Deseq2 и edgeR. Однако основным направлением их совершенствования было улучшение распознавания дифференциально экспрессирующихся генов для последующего анализа. Был проведен ряд исследований, направленных на влияние числа биологических повторностей на качество распознавания дифференциально экспрессирующихся генов ([Schurch et al., How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 2016 Jun; 22\(6\): 839–851](#)), проанализировано влияние генов с высокой дисперсией ([Cortijo et al., Widespread inter-individual gene expression variability in Arabidopsis thaliana Mol Syst Biol. 2019 Jan; 15\(1\): e8591.](#)) и т.д. Однако большая часть методов концентрировалась на распознавании генов, уровень экспрессии которых изменен более чем в 2 раза. Прежде всего это связано с неопределённостью результатов для небольших изменений. Однако, неоднократно было показано (см, например, [Laurent, et al., On the importance of small changes in RNA expression. Volume 63, Issue 1, 1 September 2013, Pages 18-24](#)), что небольшие изменения в уровне экспрессии могут играть

важную биологическую роль, но плохо выявляются при стандартных протоколах секвенирования и анализа. Было предложено несколько подходов к решению этой проблемы (см., например, Subramanian et al., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, и PNAS October 25, 2005 102 (43) 15545-15550 и Wilfinger et al., Strategies for detecting and identifying biological signals amidst the variation commonly found in RNA sequencing data BMC Genomics volume 22, Article number: 322 (2021), однако задача до сих пор остается не до конца решенной. Это показывает актуальность представленной диссертации, в которой предлагается новый подход к решению этой задачи. В основе предложенного метода лежит предположение, что гены, находящиеся в одной сети регуляции, при изменении ее функционирования изменяют свою экспрессию сходным образом. Таким образом, анализируя не все данные по генам, а только данные с похожим уровнем изменения можно выявить даже небольшие колебания в регуляции с достаточной статистической значимостью. В ходе работ, легших в основу диссертации Даниил Станиславович проводит разработку такого метода и демонстрирует, что действительно есть часть участков генетических сетей, соответствующих этому предположению. Это позволяет не только улучшить обработку дальнейших экспериментов, но и открывает возможность реанализа всего большого объема ранее накопленных данных с целью выявления эффектов со слабым проявлением на уровне изменения экспрессии, но возможно существенно сказывающихся на протекании биологических процессов. При выполнении работ легших в основу диссертации особое внимание Даниила Станиславовича было направлено на тестирование достоверности результатов предложенного метода. Так, на первом этапе проводилось тестирование метода с использованием набора симулированных данных с заданной корреляционной структурой. На основании анализа этого набора была проведена оптимизация параметров метода для достижения удовлетворительного соотношения чувствительности и специфичности. Кроме этого при анализе данных по экспрессии генов, регулируемых ауксином, была показана устойчивость результатов к смене инструментального метода анализа экспрессии. Было показано, что результаты полученные с использованием микрочипов и RNA-seq качественно не отличаются и позволяют выявить одинаковые биологические закономерности. Метод был испытан на большом наборе

данных, полученных для объектов разной таксономической принадлежности и показано, что небольшие скоординированные изменения экспрессии генов выявляются в большинстве из них. Такой аккуратный подход к тестированию не дает оснований сомневаться в высокой достоверности и обоснованности полученных результатов.

Как и к любой работе, к представленной диссертации есть ряд замечаний, в том числе к оформлению. Так во многих местах подписи к рисункам переходят на другие страницы без выделения размером шрифта и интервалом, что затрудняет чтение текста. В ряде мест диссертации применены не самые удачные формулировки. Так, например, «Апробация разработанного метода на данных десятков транскриптомных экспериментов», можно было бы сформулировать более четко, например, «Апробация разработанного метода на данных, полученных с использованием разных технологий». Не совсем корректно указаны некоторые названия. Так например, нет журнала Nature Scientific Reports, есть журнал Scientific Reports, издаваемый издательством Nature. Также вызывает некоторое сожаление, что сравнение с методом GSEA проведено только для данных, полученных для человека. Хотелось бы видеть диаграмму, аналогичную диаграмме 16, также и для второго объекта - *Arabidopsis thaliana*. Несмотря на то, что работа по этому объекту была опубликована без такого сравнения и с использованием только прототипа метода мне кажется, что его наличие, сделанное специально для диссертации уже финальной версией подхода, украсило бы ее и дало бы возможность полнее оценить эффективность и особенности предлагаемого метода. Помимо этого было бы интересно увидеть анализ разных частей диаграммы Венна, представленной на рисунке 16: например, посмотреть, нет ли каких-то особенностей в выявлении определенных GO категорий, или в самих категориях, выявляемых тем или иным методом. Однако сделанные замечания ни в коей мере не умаляют общий высокий уровень представленной диссертации.

Основные положения и результаты исследований по диссертации Вибе Даниила Станиславовича представлены в 3 статьях, которые опубликованы в зарубежных изданиях входящих в базу WoS и рекомендованных ВАК). Во всех трех статьях Вибе Даниил Станиславович является либо первым автором, или автором, внёсшим равный с первым автором вклад в исследование. Результаты были представлены на

8 конференциях, включая ведущие, такие как 30th International Conference on Arabidopsis Research (ICAR2019). Помимо этого получено авторское свидетельство на программу FoldGO. Рукопись автореферата соответствует содержанию рассматриваемой диссертации, результатам и положениям, выносимым на защиту.

Отдельно стоит отметить, что в ходе диссертации был разработан не только пакет программ на языке R, но и Web-сервис с интуитивно-понятным интерфейсом, использование которого доступно для исследователей имеющих небольшой уровень компьютерной грамотности. Это существенно повышает возможности использования предложенного метода.

Таким образом, диссертация на тему «Выявление взаимосвязи между величинами изменения экспрессии и функциями дифференциально экспрессирующихся генов на основе компьютерного анализа транскриптомов арабидопсиса и человека» представляет собой законченную научно-квалификационную работу, по актуальности, новизне, теоретической и практической значимости соответствует критериям, установленным для диссертаций на соискание ученой степени кандидата биологических наук, ее автор, Вибе Даниил Станиславович заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8 - «Математическая биология, биоинформатика».

Официальный оппонент: кандидат биологических наук, заведующий лабораторией геномики растений федерального государственного бюджетного учреждения науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук

Вибе Алексей Александрович

« 8 » октября 2021 года

Контактные данные:

тел. +7 (495) 650-42-25; e-mail: alekseypenin@gmail.com;

Специальность, по которой официальным оппонентом защищена диссертация: 03.00.15 — генетика.

Адрес места работы: 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1. Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, лаборатория геномики растений.

