

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.1.239.01,
СОЗДАННОГО НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО
БЮДЖЕТНОГО НАУЧНОГО УЧРЕЖДЕНИЯ «ФЕДЕРАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР ИНСТИТУТ ЦИТОЛОГИИ И
ГЕНЕТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ
АКАДЕМИИ НАУК», ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ
УЧЕНОЙ СТЕПЕНИ КАНДИДАТА БИОЛОГИЧЕСКИХ НАУК

аттестационное дело № _____

решение диссертационного совета от 21 февраля 2024 г. № 1

О присуждении Цуканову Антону Витальевичу

(гражданин РФ)

ученой степени кандидата биологических наук.

Диссертация Цуканова А.В. «Мультимодельный подход к эффективному картированию сайтов связывания транскрипционных факторов по данным ChIP-seq экспериментов» по специальности 1.5.8. – математическая биология, биоинформатика, принята к защите 22.11.2023 г, протокол №36, Диссертационным советом 24.1.239.01 (Д 003.011.01), созданным на базе Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», (630090, Новосибирск, пр. акад. Лаврентьева, 10).

Диссертационный совет 24.1.239.01 (Д 003.011.01) утвержден ВАК 15.01.2010, приказ ВАК № 1-7 и переутвержден Министерством образования и науки РФ 11.04.2012 года, приказ № 105/нк.

Соискатель: Цуканов Антон Витальевич, 23 сентября 1993 года рождения, В 2017 году окончил с отличием химико-технологический факультет Пермского национального исследовательского политехнического

университета с присуждением степени магистра по направлению подготовки «биотехнология». В 2021 году окончил очную аспирантуру ИЦиГ СО РАН по направлению 06.06.01 – Биологические науки. В настоящее время работает младшим научным сотрудником в секторе биоинформатики и информационных технологий в генетике ИЦиГ СО РАН.

Работа выполнена в лаборатории эволюционной биоинформатики и теоретической генетики ИЦиГ СО РАН, г. Новосибирск.

Научный руководитель – кандидат биологических наук, Левицкий Виктор Георгиевич, старший научный сотрудник лаборатории эволюционной биоинформатики и теоретической генетики ИЦиГ СО РАН, г. Новосибирск.

Официальные оппоненты:

1. **Самсонова Мария Георгиевна** доктор биологических наук, профессор Высшей школы прикладной математики и вычислительной физики, заведующая научно-исследовательской лабораторией «Математическая биология и биоинформатика» ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого», г. Санкт-Петербург.

2. **Омелина Евгения Сергеевна** кандидат биологических наук, заведующая лабораторией клеточного деления ФГБУН «Институт молекулярной и клеточной биологии СО РАН», г. Новосибирск

Оппоненты дали положительные отзывы на диссертацию.

Ведущая организация: Автономная некоммерческая образовательная организация высшего образования «Научно-технологический университет «Сириус», г. Сочи, в своём положительном отзыве, подписанном кандидатом биологических наук, научным руководителем направления «Вычислительная биология» Колпаковым Фёдором Анатольевичем, и утвержденном директором АНОО ВО

«Университет Сириус» кандидатом философских наук Лилией Геннадьевной Кирьяновой, заключила, что «Диссертационная работа Цуканова Антона Витальевича представляет собой научно-квалификационную работу, выполненную на высоком методическом уровне. По поставленным задачам, уровню их решения, актуальности и научной новизне полученных результатов диссертационная работа Цуканова Антона Витальевича соответствует требованиям «Положение о присуждении ученых степеней», утвержденного Постановлением Правительства РФ от 24.09.2013 № 842, а её автор, Цуканов Антон Витальевич, достоин присуждения искомой ученой степени кандидата биологических наук по специальности 1.5.8 — математическая биология, биоинформатика. Отзыв на диссертационную работу обсужден и утвержден на семинаре направления «Вычислительная биология» научного центра генетики и наук о жизни АНОО ВО «Университет «Сириус» 31.01.2024 г.»

Соискатель имеет всего 26 опубликованных работ, из них по теме диссертации 8, общим объемом 35 стр., из них 3 статьи в рецензируемых научных изданиях (WoS, Scopus) и 3 тезисов в материалах международных и всероссийских конференций, а также 2 авторских свидетельств. В большей части опубликованных работ личный вклад автора был определяющим. В диссертации отсутствуют недостоверные сведения об опубликованных соискателем работах.

Наиболее значительные публикации по теме диссертации:

1. Tsukanov A.V., Mironova V.V., Levitsky V.G. Motif models proposing independent and interdependent impacts of nucleotides are related to high and low affinity transcription factor binding sites in Arabidopsis. *Frontiers in plant science*, 2022:13:938545. (Scopus IF= 6.308)
2. Цуканов А.В., Левицкий В.Г., Меркулова Т.И. Метод поиска структурной гетерогенности сайтов связывания транскрипционных факторов с использованием альтернативных *de novo* моделей на

примере FOXA2. Вавиловский журнал генетики и селекции. 2021;25(1):7-17. (IF= 0.98)

3. Жимулев И.Ф., Ватолина Т.Ю., Левицкий В.Г., Колесникова Т.Д., Цуканов А.В. Развитие идеи Н.К. Кольцова о генетической организации междисков политенных хромосом *Drosophila melanogaster*. Онтогенез, 2023:54(2): 172–175 (IF= 0.7)

На диссертацию и автореферат поступило 6 отзывов, все положительные.

Отзывы прислали:

1. Бабенко Владимир Николаевич – кандидат биологических наук, старший научный сотрудник лаборатории молекулярной генетики человека ИЦиГ СО РАН, г. Новосибирск;
2. Землянская Елена Васильевна – кандидат биологических наук, ведущий научный сотрудник сектора системной биологии морфогенеза растений ИЦиГ СО РАН, г. Новосибирск;
3. Ватолина Татьяна Юрьевна – кандидат биологических наук, ведущий научный сотрудник лаборатории молекулярной цитогенетики Институт молекулярной и клеточной биологии СО РАН, г. Новосибирск;
4. Ершов Никита Игоревич – кандидат биологических наук, старший научный сотрудник сектора молекулярно-генетических механизмов белок-нуклеиновых взаимодействий ИЦиГ СО РАН, г. Новосибирск;
5. Кулаковский Иван Владимирович – доктор биологических наук, ведущий научный сотрудник группы регуляции биосинтеза белка, Институт белка РАН г. Пущино, Московская область;
6. Катохин Алексей Вадимович – кандидат биологических наук, старший научный сотрудник отдела генетики насекомых ИЦиГ СО РАН, г. Новосибирск.

Выбор официальных оппонентов и ведущей организации обосновывается тем, что оба оппонента являются компетентными специалистами в области математической биологии и биоинформатики, имеют публикации в ведущих биологических журналах и дали свое письменное согласие быть оппонентами. Ведущая организация является одним из передовых Университетов, в котором активно проводятся исследования в области вычислительной биологии, что позволяет провести экспертную оценку результатов, полученных в диссертационной работе.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований доказано, что мультимодельный подход к *de novo* поиску сайтов связывания транскрипционных факторов (ТФ) в геномах эукариот по данным секвенирования ChIP-seq, реализованный с помощью разработанного программного комплекса MultiDeNA, обладает высокой предсказательной способностью благодаря использованию методологически разных моделей распознавания мотивов связывания факторов транскрипции с ДНК – учитывающих (BaMM и SiteGA) и не учитывающих зависимости между позициями нуклеотидов внутри мотива (PWM).

Впервые **показано**, что точность распознавания сайтов связывания транскрипционных факторов, а также соотношение независимых вкладов различных моделей в общее распознавание сайтов существенно зависит от типа ДНК-связывающего домена ТФ.

Теоретическая значимость работы обоснована тем, что впервые с помощью разработанного программного комплекса MultiDeNA проведена оценка эффективности картирования сайтов связывания 157 транскрипционных факторов в геноме *Mus musculus* и 37 факторов в геноме *Arabidopsis thaliana* концептуально разными моделями распознавания сайтов по данным ChIP-seq. **Показано**, что модели BaMM и SiteGA, учитывающие зависимость вклада нуклеотида от позиции внутри мотива, превосходят

модель на основе весовой матрицы (PWM) при распознавании последовательностей сайтов с низким и средним уровнем консервативности.

Показано, что модели BaMM и SiteGA вносят наибольший вклад в распознавание сайтов связывания ТФ, ДНК-связывающий домен которых относится к типу *Basic helix-loop-helix*, наименьший – *C2H2 zinc finger*.

Показано, что каждая из моделей – PWM, BaMM или SiteGA, позволяет выявлять разные структурные типы сайтов, которые (в терминах генной онтологии) ассоциированы с определенными биологическими функциями генов-мишеней.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что впервые разработан программный комплекс MultiDeNA для анализа данных ChIP-seq экспериментов. Комплекс, предназначенный для картирования сайтов связывания ТФ в геномах эукариот с помощью комбинаций методологически разных *de novo* моделей предсказания сайтов связывания ТФ с ДНК, эффективен не только в распознавании сайтов связывания с высокой консервативностью, за счёт использования модели PWM, но и в распознавании новых структурных вариантов сайтов с низким и средним уровнем консервативности за счёт использования моделей SiteGA и BaMM.

Полученные в диссертационной работе данные и разработанный инструментарий представляют интерес для научных учреждений биологического, медицинского и сельскохозяйственного направления, связанных с изучением молекулярно-генетических механизмов регуляции экспрессии генов у высших эукариот, включая человека, и могут быть использованы в образовательном процессе при чтении курсов по биоинформатике, общей и медицинской генетике, клеточной и молекулярной биологии. Разработанный программный комплекс MultiDeNA находится в

свободном доступе (<https://github.com/ubercomrade/MultiDeNA>), и может быть использован заинтересованными исследователями.

Применительно к проблематике диссертации результативно использованы сторонние модели *de novo* поиска мотивов связывания ТФ с ДНК в геномах эукариот, которые были объединены в единый конвейер с помощью программного комплекса MultiDeNA, что позволило существенно повысить эффективность картирования сайтов связывания ТФ по данным ChIP-seq экспериментов. Для разработки программного комплекса применяли технологии программирования на языках Python и R, а также дополнительные библиотеки, включая numpy, pandas, scipy и statmodel для анализа и обработки данных matplotlib, seaborn и matplotlib-venn для визуализации данных. Для аннотации пиков ChIP-seq и анализа терминов генной онтологии использовали инструмент bedtools и пакет clusterProfile языка R. Разработку программного комплекса вели с использованием системы контроля версий git. Для статистической обработки данных применяли метод перекрёстной проверки, построение ROC-кривой, поправку на множественное проверку гипотез, расчёт коэффициента корреляции Пирсона, U-тест Манна-Уитни и тест Фишера. Обоснование необходимости создания подобного комплекса программ для анализа данных ChIP-seq подкреплено данными, полученными ранее другими исследователями по рассматриваемой тематике.

Оценка достоверности результатов предсказания сайтов связывания ТФ геномах *M. musculus* и *A. thaliana* проводилась с использованием стандартного метода перекрёстной проверки, которая была применена для всех выше описанных моделей распознавания сайтов. Корректность выявляемых сайтов связывания ТФ подтверждается также использованием инструмента TomTom, позволяющего проводить сравнение *de novo* мотива целевого ТФ с мотивом известного ТФ из интернет-доступных баз данных.

Интерпретация полученных результатов по эффективности выявления

различных типов сайтов связывания ТФ в геномах высших организмов с помощью моделей PWM, BaMM и SiteGA, основанных на методологически различных принципах, учитывает данные, полученными ранее другими исследователями по рассматриваемой тематике.

Личный вклад автора заключается в непосредственном участии в разработке конвейера MiltiDeNa для картирования сайтов связывания ТФ в геномах высших организмов, подготовке наборов данных ChIP-seq и применении MiltiDeNa для анализа этих данных, а также участии в подготовке публикаций. Основные результаты исследования получены автором самостоятельно.

Полученные соискателем научные результаты соответствуют п. 2 «Компьютерная системная биология (геномика, транскриптомика, протеомика, метаболомика, другие омиксные исследования)», п. 12 «Разработка и применение новых вычислительных алгоритмов для анализа экспериментальных данных в биологии и медицине» и п. 14 «Математические модели, численные методы, алгоритмы и программные средства применительно к процессам получения, накопления, обработки и систематизации биологических и медицинских данных и знаний» паспорта научной специальности 1.5.8. – математическая биология, биоинформатика (биологические науки).

Диссертационным советом сделан вывод о том, что диссертация представляет собой законченную научно-квалификационную работу, соответствует критериям пунктов 9-14 «Положения о присуждении ученых степеней», утвержденным постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842.

Диссертационный совет 21 февраля 2024 г. принял решение присудить Цуканову А.В. ученую степень кандидата биологических наук за решение научной задачи, связанной с разработкой нового программного комплекса, повышающего эффективность картирования сайтов связывания

транскрипционных факторов по данным ChIP-seq экспериментов, имеющей фундаментальное значение для изучения молекулярно-генетических механизмов регуляции экспрессии генов у высших эукариот и решения проблем вычислительной регуляторной геномики, в частности.

При проведении тайного голосования диссертационный совет в количестве 21 члена совета, из них по специальности 1.5.8. - 7 докторов наук; из 27 человек, входящих в состав совета, проголосовали: за – 21, против – нет, недействительных бюллетеней – нет.

Зам. председателя
диссертационного совета,
Академик РАН



Ученый секретарь
диссертационного совета,
доктор биологических наук

Н.А. Колчанов

Т.М. Хлебодарова

21.02.2024 г