

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д002.077.04
НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО
УЧРЕЖДЕНИЯ НАУКИ ИНСТИТУТА ПРОБЛЕМ ПЕРЕДАЧИ
ИНФОРМАЦИИ ИМ. А.А. ХАРКЕВИЧА РОССИЙСКОЙ АКАДЕМИИ
НАУК ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
ДОКТОРА БИОЛОГИЧЕСКИХ НАУК

Аттестационное дело № _____

Решение диссертационного совета
от 30 октября 2017, протокол № 14

о присуждении Ивану Владимировичу Кулаковскому, гражданину Российской Федерации, ученой степени доктора биологических наук.

Диссертация «Регуляторные мотивы в геномах высших эукариот и их роль в экспрессии генов» по специальности «03.01.09 – Математическая биология, биоинформатика» принята к защите 6 июля 2017 года, протокол №6, диссертационным советом Д002.077.04 на базе Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1., приказ о создании №978/нк от 16 декабря 2013 года).

Соискатель Кулаковский Иван Владимирович 1985 г. рождения, в 2006 году окончил Факультет электроники и системотехники Московского государственного университета леса по специальности «Прикладная математика и информатика» с присуждением квалификации «инженер-математик». В том же году Кулаковский И.В. поступил в аспирантуру ИМБ РАН (Москва) в Лабораторию биоинформатики и системной биологии, где

выполнил кандидатскую диссертацию по теме «Построение оптимальных моделей ДНК-сайтов связывания факторов транскрипции высших эукариот на основе данных различных экспериментальных методов». Успешная защита кандидатской диссертации с присвоением степени кандидата физико-математических наук по специальности «03.00.28 – Биоинформатика» состоялась 25 ноября 2009 года на заседании Диссертационного совета Д002.077.02 при ИПШИ РАН (Москва). Научный руководитель кандидатской диссертации – д.ф.-м.н., проф. В.Г. Туманян. С 2010 года по настоящее время Кулаковский И.В. работал в ИМБ РАН в должности научного сотрудника, затем старшего научного сотрудника, с 2016 года – ведущего научного сотрудника. Стаж научной работы после присвоения ученой степени кандидата физико-математических наук – 8 лет.

Диссертация выполнена в Лаборатории вычислительных методов системной биологии Федерального государственного бюджетного учреждения науки Института молекулярной биологии им. В.А. Энгельгардта Российской академии наук (ИМБ РАН).

Официальные оппоненты:

Орлов Юрий Львович, доктор биологических наук, профессор РАН, старший научный сотрудник лаборатории эволюционной биоинформатики и теоретической генетики Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН),

Озолин Ольга Николаевна, доктор биологических наук, профессор, заведующий лабораторией функциональной геномики и клеточного стресса Федерального государственного бюджетного учреждения науки Института биофизики клетки Российской академии наук (ИБК РАН),

Галзитская Оксана Валериановна, доктор физико-математических наук, руководитель группы биоинформатики Федерального государственного бюджетного учреждения науки Института белка Российской академии наук (ИБ РАН)

дали положительные отзывы на диссертацию.

Ведущая организация: Институт математических проблем биологии РАН – филиал Федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук» (ИМПБ РАН – филиал ИПМ им. М.В. Келдыша РАН) в своем положительном заключении, подписанным заведующим Лабораторией кристаллографии макромолекул ИМПБ РАН д.ф.-м.н., проф. В.Ю. Луниным, руководителем Института математических проблем биологии РАН – филиала ФИЦ ИПМ им. М.В. Келдыша РАН д.ф.-м.н., доц. Устиным М.Н. и утвержденным директором Федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук» член-корр. РАН А.И. Аптекаревым, отметила актуальность темы диссертации, новизну и значимость полученных результатов, и указала, что «текст диссертации не лишен опечаток и стилистических недочетов, которые, тем не менее, не препятствуют пониманию», а сама работа соответствует требованиям «Положения о порядке присуждения ученых степеней», а ее автор заслуживает присуждения ученой степени доктора биологических наук по специальности «03.01.09 – Математическая биология, биоинформатика». В отзыве ведущей организации принципиальных замечаний по материалам диссертационной работы не имеется.

В отзыве оппонента д.б.н., проф. РАН Орлова Ю.Л., давшего положительную общую оценку работе, указаны многие технические недостатки, касающиеся терминологии, точности формулировок, стиля

изложения и оформления материала диссертации, но подчеркнуто, что «Отмеченные недостатки и технические замечания не влияют на главные основные и практические результаты диссертации. Научная работа соискателя в целом должна быть оценена на самом высоком уровне.»

В отзыве оппонента д.б.н., проф. Озолин О.Н. дала высокую оценку работе, также в отзыве высказан ряд технических замечаний и подняты некоторые содержательные вопросы: « - нет ли ошибки в утверждении «как минимум 5% сходства по Жаккару» (стр. 148)? ... необоснованно краткой является описание результатов эксперимента EMSA. ... мне не удалось найти информацию о том, какие олигонуклеотиды были использованы для конкурентного формирования ДНК-белковых комплексов. ... Более серьёзным замечанием является ... отсутствие дифференциального режима у алгоритмов ChIPMunk и diChIPMunk.» При этом, указанные недостатки не снижают ценности работы: «Диссертационная работа является цельным, хорошо спланированным и законченным исследованием. Достоверность результатов никаких сомнений не вызывает, а выводы отражают суть полученных данных.»

В отзыве оппонента д.ф.-м.н. Галзитской О.В. дана высокая оценка работе, но высказаны технические замечания к оформлению и использованию английской терминологии, а также следующие содержательные вопросы уточняющего характера: «Известны ли области связывания белка с ДНК? Проводилась ли работа по выявлению характеристик для белковых участков связывания? Почему так мало общего пересечения по факторам транскрипции для разных баз данных (Рис.25)? ... можно ли найти сайты связывания для геномов растений или бактериальных геномов, изучать структуру CRISPR-локусов? Как делался анализ на выявление положительного или отрицательного отбора для сайтов и ТФ?»

Тем не менее, указанные вопросы и отмеченные технические недочеты «не снижают позитивного впечатления от представленной работы».

По материалам диссертации соискателем опубликовано 27 работ, из них 23 научных статьи, 2 приглашенные главы-обзора и 2 статьи в рецензируемых сборниках конференций. **Наиболее значимые работы** по теме диссертации:

1. (2016) I.V. Kulakovskiy, I.E. Vorontsov, I.S. Yevshin, A.V. Soboleva, A.S. Kasianov, H. Ashoor, W. Ba-Alawi, V.B. Bajic, Y.A. Medvedeva, F.A. Kolpakov, V.J. Makeev; HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.*, 44(D1):D116-25, doi: 10.1093/nar/gkv1249
2. (2016) I.E. Vorontsov, G.Khimulya, E.N. Lukianova, D.D. Nikolaeva, I.A. Eliseeva, I.V. Kulakovskiy, V.J. Makeev; Negative selection maintains transcription factor binding motifs in human cancer. *BMC Genomics*, 17(S.2):395, doi: 10.1186/s12864-016-2728-9
3. (2015) D. Papatsenko, H. Darr, I.V. Kulakovskiy, A. Waghray, V.J. Makeev, B.D. MacArthur, I.R. Lemischka; Single-Cell Analyses of ESCs Reveal Alternative Pluripotent Cell States and Molecular Mechanisms that Control Self-Renewal. *Stem Cell Reports*, 5(2):207-20. doi: 10.1016/j.stemcr.2015.07.004
4. (2014) A.R.R. Forrest, H. Kawaji, M. Rehli, J.K. Baillie, M.J.L. de Hoon, V. Haberle, T. Lassmann, I.V. Kulakovskiy, M. Lizio, M. Itoh et al.; A promoter-level mammalian expression atlas. *Nature*, 507: 462–470, doi: 10.1038/nature13182
5. (2014) Y.A. Medvedeva, A.M. Khamis, I.V. Kulakovskiy, W. Ba-Alawi, M.S.I. Bhuyan, H. Kawaji, T. Lassmann, M. Harbers, A.R.R. Forrest, V.B. Bajic, The FANTOM consortium; Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*, 15(1): 119, doi: 10.1186/1471-2164-15-119
6. (2014) V.G. Levitsky, I.V. Kulakovskiy, N.I. Ershov, D.Y. Oschepkov, V.J. Makeev, T.C. Hodgman, T.I. Merkulova; Application of experimentally verified

transcription factor binding sites models for computational analysis of ChIP-Seq data. *BMC Genomics*, 15(1): 80, doi: 10.1186/1471-2164-15-80

7. (2013) I.A. Eliseeva, I.E. Vorontsov, K.E. Babeyev, S.M. Buyanova, M.A. Sysoeva, F.A. Kondrashov, I.V. Kulakovskiy; In silico motif analysis suggests an interplay of transcriptional and translational control in mTOR response. *Translation*, 1(2), doi: 10.4161/trla.27469

8. (2013) I. Kulakovskiy, V. Levitsky, D. Oshchepkov, L. Bryzgalov, I. Vorontsov, V. Makeev; From binding motifs in ChIP-Seq data to improved models of transcription factor binding sites. *Journal of Bioinformatics and Computational Biology*, 11(1): 1340004, doi: 10.1142/S0219720013400040

9. (2013) I.V. Kulakovskiy, Y.A. Medvedeva, U. Schaefer, A.S. Kasianov, I.E. Vorontsov, V.B. Bajic, V.J. Makeev; HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.*, 41(D1): D195-D202, doi: 10.1093/nar/gks1089

10. (2010) I.V. Kulakovskiy, V.A. Boeva, A.V. Favorov, V.J. Makeev; Deep and wide digging for binding motifs in ChIP-Seq data. *Bioinformatics*, 26(20): 2622-3, doi: 10.1093/bioinformatics/btq488

На автореферат диссертации поступили отзывы от:

1. **Баранова Павла Викторовича**, к.х.н., руководителя Лаборатории пост-транскрипционного контроля и биоинформатики Школы биохимии и клеточной биологии Университета Колледжа Корка, Корк, Ирландия

2. **Вихлянцева Ивана Милентьевича**, д.б.н., зав. Лабораторией структуры и функций мышечных белков Федерального государственного бюджетного учреждения науки Института теоретической и экспериментальной биофизики Российской академии наук, г. Пущино, Московская область

3. **Лагарьковой Марии Андреевны**, д.б.н., чл.-корр. РАН, зав. Лабораторией клеточной биологии Федерального государственного бюджетного учреждения «Федеральный научно-клинический центр физико-

химической медицины Федерального медико-биологического агентства», г. Москва

4. **Хмелениной Валентины Николаевны**, д.б.н., ведущего научного сотрудника Лаборатории радиоактивных изотопов Федерального государственного бюджетного учреждения науки Института биохимии и физиологии микроорганизмов Российской академии наук, г. Пущино, Московская область

5. **Самсонова Александра Михайловича**, д.ф.-м.н., проф., зав. сектором Федерального государственного бюджетного учреждения науки Физико-технический институт им. А.Ф. Иоффе РАН, Санкт-Петербург

6. **Разина Сергея Владимировича**, д.б.н., чл.-корр. РАН, главного научного сотрудника Федерального государственного бюджетного учреждения науки Института биологии гена Российской академии наук, Москва

7. **Аульченко Юрия Сергеевича**, д.б.н., ведущего научного сотрудника Лаборатории рекомбинационного и сегрегационного анализа Федерального исследовательского центра Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск

8. **Шеваль Евгения Валерьевича**, д.б.н., старшего научного сотрудника НИИ физико-химической биологии имени А.Н. Белозерского Московского государственного университета имени М.В. Ломоносова, Москва

9. **Мироновой Виктории Владимировны**, к.б.н., зав. сектором системной биологии морфогенеза растений Федерального исследовательского центра Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск

10. **Тутукиной Марии Николаевны**, к.б.н., старшего научного сотрудника Лаборатории функциональной геномики и клеточного стресса Федерального государственного учреждения науки Института биофизики клетки Российской академии наук, Пущино

В отзывах отмечена актуальность вычислительного изучения характерных закономерностей – мотивов – в регуляторных районах геномов высших эукариот и важность достигнутых соискателем результатов. Авторы всех отзывов считают диссертацию соответствующей всем требованиям «Положения о порядке присуждения ученых степеней», а соискателя – заслуживающим присуждения ученой степени доктора биологических наук по специальности «03.01.09 – Математическая биология, биоинформатика».

Замечания в отзывах на автореферат: В отзыве Баранова П.В. задан дискуссионный вопрос: «... было бы интересно узнать о перспективах распространения разработанной методологии в область РНК-белковых взаимодействий ...». В отзыве Хмелениной В.Н. также присутствуют вопросы дискуссионного характера: «Как автор диссертации видит развитие сфера применения разработанных биоинформатических методов при изучении генетических детерминант у высших эукариот? Планирует ли автор изучать более простые модели регуляции на примере прокариотных организмов?» В отзывах Самсонова А.М. и Аульченко Ю.С. отмечены некоторые недочеты стиля изложения и опечатки. В отзывах Вихлянцева И.М., Лагарьковой М.А., Разина С.В., Мироновой В.В., Шеваль Е.В., Тутукиной М.Н. замечаний и дискуссионных комментариев нет.

Выбор официальных оппонентов и ведущей организации обоснован их публикационной активностью по вопросам биоинформатики, в том числе, по темам, относящимся к тематике диссертации, и безусловной компетентностью в биоинформатических исследованиях, в частности, в проведении вычислительного и статистического анализа биологических последовательностей, исследовании генной регуляции и организации биоинформатических баз данных.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

представлены новые вычислительные методы биоинформатики нуклеотидных последовательностей, предназначенные для анализа характерных паттернов-мотивов: **новые алгоритмы и программы** для идентификации мотивов в виде весовых матриц на основе современных экспериментальных данных, получаемых с использованием высокопроизводительного секвенирования и **новый метод** для оценки сходства мотивов по Жаккару, **усовершенствован метод** оценки регуляторного потенциала однонуклеотидных вариантов в регуляторных последовательностях.

Идентифицированы мотивы ДНК-белкового узнавания путем анализа результатов нескольких сотен опубликованных экспериментов по изучению полногеномного связывания ДНК регуляторными белками-факторами транскрипции, **разработана новая наиболее полная база данных**, содержащая мотивы сайтов связывания факторов транскрипции человека и мыши в форме позиционно-весовых матриц. Представленные в коллекции мотивы покрывают основные структурные семейства факторов транскрипции и **наилучшим образом** среди существующих баз данных представляют характерные паттерны, описывающие участки ДНК, связываемые факторами транскрипции. В авторскую коллекцию **впервые** систематически включены динуклеотидные модели, учитывающие зависимости между соседними нуклеотидами.

Теоретическая значимость исследования обоснована тем, что получен ряд важных фундаментальных результатов в регуляторной геномике эукариот, а именно:

Выдвинута гипотеза о роли тройственного композитного элемента, состоящего из сайтов связывания ключевых факторов плюрипотентности

OCT4-SOX2/NANOG, через который реализуется антагонизм OCT4 и NANOG в генной сети.

Выявлена зависимость действия фактора транскрипции Sp1 на экспрессию генов в мышинной модели эритролейкемии от промоторной или энхансерной локализации tandemных пар участков связывания.

Подтверждена предпочтительная локализация пиримидиновых мотивов в 5'-окрестности мРНК, являющихся мишенями сигнального каскада mTOR, на основе **совместного вычислительного анализа** данных рибосомного профайлинга и карты регионов инициации транскрипции.

Обнаружены mTOR-мишени, обладающие протяженными регионами инициации транскрипции, что говорит о возможной альтернативной трансляционной регуляции mTOR-мишеней в зависимости от транскрипционной активности.

Установлено действие отрицательного отбора на мутации, возникающие в окрестности участков, связываемых факторами транскрипции в геномах раковых клеток человека.

Успешно идентифицированы мотивы в полногеномном каталоге промоторов, активных в различных типах клеток и **установлено**, что разнообразие регуляторных сигналов в промоторах генов практически полностью описывается известными мотивами.

Установлено систематическое избегание перекрытия сайтов связывания факторов транскрипции и конкретных CG-пар генома, дифференциальное метилирование которых связано с активностью близлежащих промоторов.

В ходе работы использованы современные методы вычислительной биологии, биоинформатики, теории вероятностей и математической статистики. **Достоверность и воспроизводимость результатов исследования** обеспечены использованием стандартных хорошо зарекомендовавших себя статистических пакетов и опубликованием всех

новых авторских вычислительных методов и систематической коллекции мотивов, построенной в ходе работы, в открытом доступе в сети Интернет. В частности, авторские программы представлены с открытым исходным кодом, что дает широкому исследовательскому сообществу доступ к методической части диссертации и, в частности, позволяет широкое практическое использование результатов как в научной, так и в образовательной деятельности.

Значение полученных соискателем результатов исследования для практики подтверждается активным и успешным использованием авторских биоинформатических инструментов и базы данных в научном сообществе, что подтверждается цитированием авторских публикаций, грантовой поддержкой работ (проекты, поддержанные Российским научным фондом и Российским фондом фундаментальных исследований) и личными наградами научного сообщества: премия Европейской Академии для молодых российских ученых (2016), Медаль «Феномен жизни» памяти В.И. Корогодина (2015). Методическая база и достаточно обширная коллекция мотивов позволяет решать как конкретные задачи по аннотации регуляторных районов генов, так и проводить глобальные полногеномные исследования для системной биологии. При этом, сфера применения авторских методов не ограничивается исследованиями регуляции транскрипции у высших эукариот: аналогичные подходы могут успешно применяться для исследования регуляции экспрессии генов у бактерий.

Результаты представленной работы **могут использоваться** в учебном процессе и научных исследованиях, проводимых в организациях, работающих в области биоинформатики и молекулярной биологии, в том числе в Федеральном государственном бюджетном учреждении науки Институте биологии гена Российской академии наук, на Биологическом факультете и Факультете биоинженерии и биоинформатики МГУ им. М.В.

Ломоносова, в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», в Федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский политехнический университет Петра Великого», в Федеральном государственном бюджетном учреждении науки Институте белка Российской академии наук и других научно-исследовательских и образовательных организациях.

Материалы диссертации полно изложены в опубликованных работах. **Вклад диссертанта в опубликованные работы** по теме диссертации состоит в планировании исследований, формулировке задач, непосредственной разработке и компьютерной реализации алгоритмов, вычислительном анализе данных, в том числе, с помощью авторских методов. В 7 публикациях по теме диссертации автор выступает в качестве первого автора, в 9 – в качестве автора, ответственного за переписку. В диссертации представлены крупные достижения в области регуляторной геномики высших эукариот, работа обладает внутренним единством, цельным изложением и непротиворечивыми выводами. По своему содержанию диссертация отвечает паспорту специальности «03.01.09 – Математическая биология, биоинформатика».

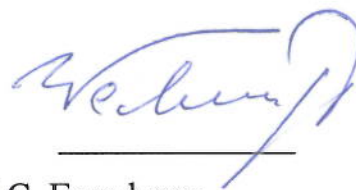
Диссертационный совет пришел к выводу о том, что диссертация представляет собой завершенное научное исследование, решены задачи, имеющие большое теоретическое и практическое значение для молекулярной биологии и биоинформатики. По актуальности, новизне, объему и значимости проделанной работы диссертация безусловно соответствует требованиям, установленным Положением «О порядке присуждения ученых степеней», утвержденным Постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, с изменениями Постановления

Правительства Российской Федерации от 21 апреля 2016 года №335, предъявляемым к диссертациям на соискание ученой степени доктора биологических наук.

На заседании 30 октября 2017 года диссертационный совет принял решение присудить Ивану Владимировичу Кулаковскому **ученую степень доктора биологических наук по специальности «03.01.09 – Математическая биология, биоинформатика».**

При проведении тайного голосования диссертационный совет в количестве **14** человек, из них **9** докторов наук по специальности рассматриваемой диссертации, участвовавших в заседании, из **21** человека, входящих в состав совета, проголосовали за - **14**, против - **0**, недействительных бюллетеней - **0**.

Председатель
диссертационного совета Д 002.077.04
д.б.н., профессор



М.С. Гельфанд

Ученый секретарь
диссертационного совета Д 002.077.04
д.б.н., профессор



Г.И. Рожкова