

Отзыв официального оппонента

на диссертационную работу Пономаренко Михаила Павловича на тему “Компьютерный анализ контекстно-зависимых количественных характеристик специфической биологической активности сайтов в составе геномной ДНК”, представленную на соискание ученой степени доктора биологических наук по специальности 03.01.09 – математическая биология, биоинформатика

Актуальность темы

Наличие большого объема экспериментальных данных по секвенированию геномной ДНК открывает возможность комплексного системного исследования структурно-функциональной организации сайтов ДНК, контролирующих различные функции организма. К числу приоритетных задач постгеномной биоинформатики относится создание методов, нацеленных на выявление количественных характеристик специфической биологической активности сайтов в составе геномной ДНК. В связи с этим диссертационная работа М.П. Пономаренко является актуальной, так как посвящена разработке методов предсказания контекстно-зависимых конформационных и физико-химических свойств двойной спирали ДНК, определяющих количественные взаимодействия сайтов в составе геномной ДНК со специфическими белками. Такие количественные величины необходимы для решения практических задач, связанных с созданием искусственных генных конструкций, обладающих заданными свойствами, в том числе в таких областях как персонализированная медицина и биотехнология.

Научная новизна

Для выявления контекстно-зависимых количественных величин активности сайтов ДНК соискателем создана инновационная технология, которая сочетает в себе такие достижения математики как теорию аддитивной полезности для принятия решений и нечеткие множества. На основе данной технологии автор разработал компьютерные системы bDNAvideo и Activity, которые анализируют в автоматическом режиме миллионы контекстных, а также контекстно-зависимых конформационных и физико-химических свойств спирали ДНК, выбирая среди них наборы таких свойств, каждое из которых достоверно коррелирует с

количественными величинами искомой биологической активности сайтов ДНК. Здесь важно отметить, что полученные результаты демонстрируют применимость регрессионного анализа для прогноза биологической активности сайтов в составе геномных ДНК.

С использованием компьютерной системы bDNAvideo соискатель обнаружил достоверную кластеризацию транскрипционных факторов на две группы по сходству спирали ДНК сайтов их связывания. В свою очередь, компьютерная система Activity позволила автору впервые вывести регрессионные уравнения для предсказания величин сродства целого ряда регуляторных белков к сайтам их связывания в геномных ДНК, которые были подтверждены независимыми экспериментальными данными.

Также, соискатель впервые вывел регрессионные уравнения и подтвердил их на независимых экспериментальных данных в случаях повреждений геномной ДНК по гуанинам при воздействии ультрафиолетового излучения. Кроме того, впервые получено уравнение регрессии для прогноза эффективности связывания между одонитевой ДНК на флангах бреши, возникающей после вырезания повреждений из геномной ДНК, и белком RecA, активирующим работу репарационной системы по ликвидации этой бреши. Эти уравнения уточняют существующие представления о влиянии контекста ДНК на устойчивость геномов к воздействию мутагенов.

Важно отметить, что рассматриваемая диссертационная работа имеет реальный **научно-практический выход**. Так, при публикации компьютерных систем BiDaS (Греция), SITECON (Россия), CRoSSeD (Бельгия), DISCOVER (США), FeatureScan, DiProDB, BioBayesNet и ProMapper (все: Германия) их авторы указали базу данных PROPERTY соискателя в качестве информационной основы этих систем. В свою очередь, созданная автором компьютерная система Activity регулярно используется как инструмент вывода все новых и новых регрессионных уравнений для оценок сродства между сайтами в составе геномных ДНК разных таксонов и регуляторными белками. Кроме того, с использованием регрессионного уравнения для сродства ТАТА-связывающего белка к промоторам генов ежегодно аннотируются сотни вариантов однонуклеотидного полиморфизма в составе

референсного генома человека в качестве кандидатных ДНК-маркеров для предрасположенности к заболеваниям.

Степень обоснованности научных положений, выводов и рекомендаций

Диссертационная работа М.П. Пономаренко выполнена на высоком научном и методическом уровне. Научные положения, выносимые на защиту диссертации, интегрируют результаты компьютерных расчетов, подтвержденных независимыми экспериментальными данными различных авторов, которые получены с помощью таких методов измерений, как фотометрия уровней экспрессии репортерных генов *ex vivo* и задержка электрофоретической подвижности комплексов ДНК/белок *in vitro*. Сформулированные соискателем обсуждения, заключения, выводы и рекомендации аргументированы и обоснованы с использованием сравнений его собственных результатов с данными независимых компьютерных расчетов и экспериментов. Достоверность результатов диссертационной работы определяется достаточным объемом контрольных экспериментальных данных, использованием широкого круга статистических критериев, применением бутстреп-анализа к оценке устойчивости этих критериев при варьировании подвыборок разного объема из всей совокупности исследуемых данных, а также демонстрацией адекватности результатов применения созданных компьютерных систем к решению задач столь разных областей биологии как регуляция экспрессии генов, повреждения геномов мутагенами и репарация этих повреждений. Достоверность результатов диссертации подтверждается также их опубликованием в рецензируемых журналах, включая 30 статей в журналах из Перечня ВАК, индексируемых РИНЦ, Scopus и Web of Science (в том числе: 19 - за рубежом). Основные результаты диссертации соискатель доложил на 33 международных и российских симпозиумах, совещаниях и конференциях, получив одобрение научной общественности в форме приглашения Academic Press стать автором ряда глав Энциклопедии по генетике Сиднея Бреннера.

Структура диссертационной работы

Диссертация имеет структуру монографии и состоит из Введения, Обзора литературы (глава 1), четырех глав (каждая из которых включает материалы, методы, результаты и обсуждения), Заключения, Выводов, Списка литературы из

467 источников (в том числе: на русском языке – 71), Списка терминов, обозначений и сокращений. Работа изложена на 310 страницах машинописного текста, включая 81 рисунок и 41 таблицу. Считаю выбор автором монографического плана с тематическими главами уместным, поскольку это облегчило восприятие результатов из столь разных областей биологии как взаимодействие белков с ДНК, повреждение геномов мутагенами и полиморфизм генома, связанный с заболеваниями человека.

Введение (стр. 6-15) посвящено обоснованию актуальности темы диссертации и степени разработанности этой темы как к началу работы, так и в настоящее время, а также выбору цели исследования, декомпозиции этой цели в набор задач, решение которых позволяет достигнуть эту цель. Сформулированы выносимые на защиту положения. Охарактеризованы новизна и практическая значимость полученных результатов, а также их апробация. Существенно, что автор весьма подробно охарактеризовал совокупность публикаций по теме диссертации и свой личный вклад в эти публикации, поскольку все результаты диссертационной работы были опубликованы им в соавторстве.

Глава 1 (стр. 16-77) является “Обзором литературы”, где соискатель обосновал выбор нового направления математической биологии и биоинформатики, развитого им в диссертации. Сначала он представил банки данных по последовательностям ДНК и компьютерные системы для их анализа, далее – механику связывания белков с ДНК, статистические, сложностные, корреляционные и контекстные методы как методическую базу этих систем. Наконец, автор привел независимые свидетельства в пользу необходимости дополнения этих систем средствами учета физико-химических и конформационных свойств спирали ДНК для расширения области их применимости за счет оценки количественных величин активности сайтов ДНК.

В главе 2 (стр. 78-131) представлена компьютерная система bDNAvideo как средство учета конформационных и физико-химических свойств двойной спирали ДНК сайтов связывания транскрипционных факторов, дискриминирующих их от последовательностей случайных независимых равновероятных нуклеотидов. На этом простейшем примере соискатель обнаружил, что транскрипционные факторы можно достоверно кластеризовать по сходству спирали ДНК сайтов их связывания

на две группы, одна из которых соответствует комплексам ДНК/белок по принципу “ключ/замок”, другая – взаимодействиям между локальным избытком позитивного электростатического заряда на поверхности белков и локальным избытком негативного электростатического заряда на сахаро-фосфатном остове ДНК.

Глава 3 (стр. 132-159) содержит описание расширения bDNAvideo до целевой компьютерной системы Activity. Соискатель добавил критерии регрессионного анализа данных к критериям дискриминантного анализа. С помощью этой новой системы он обработал данные трех независимых экспериментов, в условиях которых ТАТА-связывающих белок (ТВР) связывался *in vitro* с дуплексами ДНК длиной 26 п.о., идентичными ТАТА-боксам человека, с модельными дуплексами ДНК длиной 15 п.о. и с нитями ДНК длиной 15 нт. В результате соискатель вывел регрессионное уравнение трехшагового связывания ТВР/ТАТА-бокс и подтвердил количественные прогнозы этого уравнения с использованием специально спланированной серии контрольных экспериментов в равновесных и в неравновесных условиях *in vitro*.

В главе 4 (стр. 160-194) соискатель оценил границы области применимости системы Activity, созданной с использованием экспериментальных данных о сайтах связывания транскрипционных факторов, продемонстрировав адекватность прогнозов этой системы для предмутационных повреждений генома мутагенами, вырезания этих повреждений из генома и репарации возникающих при этом брешей в геноме.

Глава 5 (стр. 195-226) иллюстрирует адекватность результатов системы Activity на ряде примеров практических биологических задач. Прежде всего, автор оценил влияние мутаций в транскрипционном факторе MEF2 на его специфичность к сайтам MEF2-связывания в промоторах генов мыши. Затем, он показал, как можно привлечь данные об амплификации/селекции аптамеров ДНК на сродство к белку-мишени *in vitro* (SELEX-протокол) для компьютерноэлектронного распознавания природных сайтов связывания этого белка в составе геномной ДНК.

Заключение (стр. 227-231) содержит аннотацию работы, включая краткий обзор созданных в диссертации компьютерных систем bDNAvideo и Activity, входящих в эти системы банков данных, баз знаний и активных приложений, а

также основных результатов, полученных автором диссертационной работы с помощью этих систем.

Выводы (стр. 232-234) соответствуют цели диссертационной работы (вывод №1) и результатам решения исследовательских задач в процессе достижения этой цели (№2 - №8). Достоверность выводов установлена независимыми контрольными экспериментами, адекватно оценившими границы применимости этих результатов.

Замечания к работе:

1. Не совсем понятно, почему вывод, касающийся создания и использования компьютерной системы bDNAvideo, оказался последним. Логичнее было поставить его первым, перед выводами 1 и 2, касающимися создания и использования системы Activity, которая была создана на основе bDNAvideo.

Очевидно, что данное замечание носит частный характер и ни в коей мере не препятствуют положительному заключению.

Заключение:

Диссертационная работа Пономаренко Михаила Павловича представляет собой законченное исследование, на тему “Компьютерный анализ контекстно-зависимых количественных характеристик специфической биологической активности сайтов в составе геномной ДНК”.

Она имеет высокую научно-практическую и теоретическую значимость, существенно расширяет существующие представления о вкладе конформационных и физико-химических свойств спирали ДНК в генетическое кодирование биологической активности регуляторных сайтов в составе геномной ДНК, а также о влиянии нуклеотидного контекста как на частоты предмутационных повреждений геномов, так и на эффективность репарации этих повреждений. Полученные результаты являются новыми и приоритетными и полностью обосновывают выводы диссертации. Нет сомнений в достоверности результатов, которые подтверждены независимыми экспериментами и опубликованы в достаточном количестве статей в рецензируемых международных и российских журналах. Автореферат в полной мере отражает содержание диссертации.

По совокупности новых, актуальных, биологически значимых и достоверных результатов диссертационная работа Пономаренко Михаила Павловича соответствует требованиям ВАК, которые сформулированы в п. 9 «Положения о присуждении ученых степеней» Постановления Правительства Российской Федерации от 24 сентября 2013 г. № 842. Считаю, что Пономаренко Михаил Павлович заслуживает присуждения ученой степени доктора биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

Официальный оппонент:

доктор биологических наук по специальности

03.00.03 - молекулярная биология

заведующий теоретическим отделом



Бажан Сергей Иванович

06.10.2017

р.т. +7 (383) 363-47-00 вн. номер 2001,

e-mail: bazhan@vector.nsc.ru

Федеральное бюджетное учреждение науки «Государственный научный центр вирусологии и биотехнологии „Вектор“ Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека (ФБУН ГНЦ вирусологии и биотехнологии «Вектор» Роспотребнадзора)

Почтовый адрес: 630559, р.п. Кольцово, Новосибирская обл., Россия

Телефоны: 8(383)-3634710, 8(383)-3366010;

Факс: 8(383)-3367409;

E-mail: vector@vector.nsc.ru

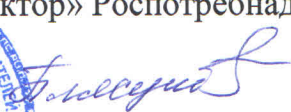
web-сайт: <http://www.vector.nsc.ru>

Подпись заведующего отделом, д.б.н. С.И. Бажана заверяю:

Ученый секретарь ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора

канд. биол. наук, доцент

6 октября 2017 г.



О.А. Плясунова