

Отзыв официального оппонента

о диссертации Пономаренко Михаила Павловича

«Компьютерный анализ контекстно-зависимых количественных характеристик
специфической биологической активности

сайтов в составе геномной ДНК»,

представленной на соискание ученой степени доктора биологических наук по
специальности 03.01.09 - Математическая биология, биоинформатика

Актуальность темы. Диссертационная работа М.П. Пономаренко является результатом многолетней работы на стыке биофизики и биоинформатики. Основной темой его интересов является определение тех структурных и физических свойств биомакромолекул, в первую очередь нуклеиновых кислот, которые обеспечивают специфическое функциональное взаимодействие с другими биологическими молекулами, в первую очередь белками, формирующими структуру хроматина и белками, регулирующими инициацию транскрипции. На этом пути автором был получен ряд фундаментальных результатов, которые послужили источником важного практического метода, позволяющего использовать особенности структурно-функциональной организации ДНК, для определения различных свойств участков двойной спирали, таких, например, как мутагенность. Исследования автора приобретают особую актуальность в связи с тем, что в последнее время возникли мощные и дешевые методы секвенирования, благодаря которым многие исследования в области генетики и теории эволюции, еще лет 10 назад имевшие достаточно теоретический интерес, перешли в сугубо практическую область интерпретации данных персональных геномов и популяционной геномики. В этой связи структурно-функциональные методы, развитые автором, позволяют на новом уровне оценивать последствия функциональных замен на структуру ДНК и аффинность участка ДНК к различным белкам, способным специфически распознавать этот участок. Это имеет огромное значение для прогресса персональной геномики, в которой методы, разработанные диссертантом, и аналогичные им, будут применяться для приоритизации индивидуальных вариантов и для поиска молекулярных маркеров предрасположенности к различным болезням. Все это дает основание заключить, что предложенная работа является исследованием на весьма актуальную тему, имеющим большие перспективы для развития в ближайшее время.

Научная новизна. В работе М.П. Пономаренко используется ряд оригинальных математических методов, в основном относящихся к сфере непараметрической статистики и машинному обучению (теория полезности, множественные регрессии, нечеткие модели).

Автором развит оригинальный подход, в основном, основанный на регрессионном анализе, который хорошо объединяет работу методически. На мой взгляд неправильно останавливаться на новизне результатов «на настоящий момент», поскольку многие результаты, бывшие безусловно новыми в момент опубликования, с тех пор стали классическими и многократно были переоткрыты. К таковым в первую очередь следует отнести работу по анализу связывания транскрипционных факторов с альтернативными аллелями, и созданную диссертантом программу rSNP_Guide, насколько мне известно, первый инструмент такого рода, с тех пор многократно воссозданный и цитируемый незаслуженно редко. Среди других результатов диссертанта можно назвать впервые проведенный анализ конформационных и физико-химических свойств участков ДНК, связывающих транскрипционные факторы, а также расчеты, позволяющие связать аффинность ДНК-связывающих белков к ДНК при наличии мутаций в первичных структурах белков. Важно, что эти мутации мало меняют электростатические свойства белка (происходит замена аргинин-лизин, обе аминокислоты заряжены положительно), но при этом происходит сильная потеря специфичности, благодаря геометрическим факторам, а именно, утраты чувствительности к ширине малой бороздки В-формы спирали ДНК. Интересным также является обнаруженная диссертантом связь локальных физико-химических свойств ДНК с ее мутабельностью, хотя на мой взгляд в этой области трудно отличить непосредственное влияние физико-химических свойств ДНК от вторичного влияния, опосредованного распознаванием последовательностей ДНК ферментами, осуществляющими репарацию.

Вообще работу отличает большое внимание к биофизической и физикохимической стороне дела, что поставило ее на солидную основу. Практически во всех случаях описаны элементарные молекулярные события, лежащие в основе наблюдаемых эффектов, что производит исключительно хорошее впечатление.

Степень обоснованности научных положений, выводов и рекомендаций.

На защиту выносятся три положения: (1) эффективность выделения значимых конформационных и физико-химических характеристик возможной структуры ДНК в пределах участков, связывающих регуляторные белки, с экспериментальным подтверждением соответствующих статистических закономерностей; (2) роль последовательности ДНК в мутагенезе, и интерпретация физико-химических свойств ДНК как факторов, способствующих ее повреждению или репарации этих повреждений; (3) метод определения значимых структурных и физико-химических признаков ДНК,

основанный на регрессионном анализе. Все эти положения вполне убедительно изложены в работе и подтверждены ясными иллюстрациями.

Структура и объем работы

Объем диссертации составляет 310 машинописных страницы, включая 81 рисунок и 41 таблицу. Список литературы содержит 467 ссылок.

Диссертация построена по классической схеме и состоит из Введения, пяти глав – «Обзор литературы», «Компьютерная система BDNA-VIDEO: количественные характеристики ДНК сайтов связывания транскрипционных факторов», «Компьютерная система Activity: корреляция между средством ТАТА-связывающего белка к ТАТА-боксу и количественными характеристиками ДНК», «Компьютерная система Activity: оценка влияния контекста на эффективность мутагенеза геномной ДНК», «Контекстно-зависимые количественные характеристики ДНК, коррелирующие с активностью сайтов связывания транскрипционных факторов», общих выводов, списка публикаций по теме диссертации, списка литературы, и списка терминов и сокращений.

Во Введении автор проводит обоснование актуальности темы исследования, и формулирует цели и научные задачи исследования. Представлен личный вклад соискателя в работу в совместных публикациях с другими соавторами, показано, что в работах, проведенных вместе с соавторами-экспериментаторами, в задачу автора входил компьютерный анализ данных, что соответствует специальности диссертационной работы. Сформулированы основные научные результаты, выносимые на защиту.

Диссертации предпослан сравнительно короткий, но эффективный обзор литературы, описывающий в основном историю вопроса. Обзор содержит массу новых для меня сведений, опубликованных в относительно старой литературе, например, я с большим удовольствием и пользой прочел обзор статфизических моделей ДНК-белкового взаимодействия. Следует заметить, что практически каждая глава начинается с мини-обзора литературы по теме главы.

Вторая глава — содержит относительно старые исследования, посвященные моделированию взаимодействия ТАТА-боксов и белка ТВР, а также взаимодействию других транскрипционных факторов с ДНК. В основном эта глава содержит описание уже классических работ, выполненных автором в 90е годы. Кроме того, подробно излагается построение системы bDNAvideo, в которой были реализованы основные методы анализа данных, применявшиеся автором и далее. Глава содержит несколько отступлений, в частности классификацию основных транскрипционных факторов по их структурным свойствам.

В третьей главе изложено построение системы Activity, позволяющей количественно вычислять аффинность регуляторных белков к ДНК, причем как в открытых, так и в закрытых комплексах. В этой главе построено эмпирическое уравнение связывания, позволяющее вычислять константу равновесия белка ТВР с ТАТА-боксом. Формула дает приемлемое приближение экспериментально полученных значений взаимодействия. Диссертант проделал большую работу по систематизации литературы, содержащей опубликованные количественные данные по взаимодействию между ТАТА-боксом и ТВР.

В четвертой главе показаны результаты применения разработанной автором системы Activity к оценке влияния нуклеотидного контекста на эффективность мутагенеза ДНК. Надо сказать, что эта глава содержит большое количество других примеров использования системы Activity, например, для вычисления константы равновесия микроРНК к белкам Ago2 и Ago3. Программа также применялась для анализа репортерного конструкта РЗ(х4), используемого в качестве индикатора наличия ауксина в клетках растений. Таким образом, хорошо продемонстрирована широта возможных применений системы Activity.

В пятой главе исследованы описание применения разработанных компьютерных методов для анализа связывания хорошо изученных транскрипционных факторов MEF-2, USF и YY1, позволяющих преодолеть ряд ограничений традиционных матриц позиционных частот для предсказания участков ДНК, связывающих эти белки. В частности, показана возможность учета влияния экспериментальных условий на связывания белка. В качестве отдельной темы рассматривается влияние мутаций в белках, на мой взгляд небольшое, но очень красивое исследование.

В «Заключении» суммируются основные результаты диссертационной работы.

Достоверность и новизна исследования, полученных результатов, выводов и рекомендаций, сформулированных в диссертации

В диссертационной работе в целом описываются оригинальные компьютерные методы анализа физико-химических и структурных свойств последовательностей ДНК, взаимодействующих с различными белками. Разработаны оригинальные методы, основанные на регрессионном анализе и теории полезности, позволяющие выделять характерные признаки в многомерном пространстве локальных свойств макромолекулы.

Разработанный автором подход позволяет рассматривать на общих основаниях текстовые закономерности в последовательностях нуклеотидов, такие как консенсусы, позиционно-весовые матрицы, частоты встречаемости олигонуклеотидов в 15-буквенном коде IUPAC-IUB, а также конформационные и физико-химические характеристики В-формы двойной спирали ДНК. Все такие описания рассматриваются как отдельные

контекстно-зависимые количественные локальные характеристики геномных ДНК. Из многомерного пространства характеристик метод позволяет отобрать важные локальные характеристики геномной ДНК, статистически достоверно коррелирующие с экспериментально измеренными величинами специфической биологической активности соответствующих участков. Полученные методы реализованы в виде компьютерных программ, использованных для анализа большого количества экспериментальных данных различной природы.

В целом, результаты, полученные автором, являются новыми научными знаниями по специальности математическая биология, биоинформатика. Основные результаты диссертации неоднократно обсуждались на различных конференциях и симпозиумах и получили одобрение ведущих специалистов.

По теме диссертации представлено 55 публикаций, из них 30 статей в научной периодике (все по Перечню ВАК), 6 статей в сборниках научных трудов, восемь глав в монографиях, 11 тезисов конференций.

Практическое значение работы

В процессе работы автором были разработаны оригинальные методы анализа физико-химических и структурных свойств ДНК, созданы программный комплекс bDNA-video и Activity, позволяющие вычислять большое количество локальных свойств ДНК, а также получать эмпирические формулы для констант равновесия ДНК с регуляторными белками. Программное обеспечение может использоваться для функциональной аннотации геномных последовательностей, что обеспечивает существенное расширение области применения методов компьютерного анализа полногеномных данных, а также позволяет осуществлять исследование персональных вариантов и соматических мутаций в геномах.

Конкретные рекомендации по использованию результатов и выводов диссертации

Результаты работы могут использоваться в курсе биоинформатики для студентов университетов. Разработанные методы анализа многомерных карт физико-химических свойств локальных участков ДНК могут быть использованы для анализа результатов взаимодействия ДНК с различными белками, в частности при анализе экспериментов ChIP-seq. Компьютерные программы bDNAvideo и Activity могут использоваться для вычисления констант равновесия нативных и мутантных последовательностей ДНК с белками или фрагментами нуклеиновых кислот, что может быть актуально для предсказания взаимодействия ДНК и, возможно, двухспиральной РНК с различными активными препаратами, в частности микроРНК.

Результаты исследования по определению изменения констант равновесия ДНК-связывающих белков при мутациях ДНК могут быть использованы для предсказания полиморфизмов или соматических мутаций, ассоциированных с возникновением различных наследственных или развивающихся заболеваний.

Полученные результаты и компьютерные методы можно использовать в дальнейших исследованиях в научных учреждениях ФАНО РФ, Минздрава РФ, Минобрнауки РФ и Роспотребнадзора РФ, занимающихся полногеномным анализом, секвенированием, и обработкой генетических и популяционно-генетических данных, при планировании и анализе подобных экспериментов.

Практические рекомендации, сформулированные в диссертации, обоснованы проведенными исследованиями и могут служить руководством в работе.

Оценка содержания диссертации, ее завершенность

Работа хорошо структурирована, каждая глава содержит небольшое введение и выводы, подтверждающее обоснованность выводов по диссертации в целом. Работа содержит большое число иллюстраций и таблиц и написана ясным, но в то же время живым языком. Читается работа, несмотря на сложность материала, очень легко.

Несмотря на высокий уровень работы, диссертация не свободна от спорных моментов. По сути предложенный автором метод является методом снижения размерностей, но проблема в том, что во многих случаях размерность искусственно повышена, поскольку очень часто единственной доступной информацией является информация о последовательности ДНК, а информация о возможных структурных свойствах восстанавливается из последовательности. Если математический метод, предложенный автором, работает корректно, а автор вполне убедительно это доказывает, то, казалось бы, понижение размерности должно в каком-то виде восстанавливать исходный текст. В этом случае не будет ли более последовательно вообще не учитывать возможные структурные свойства, а работать непосредственно с нуклеотидным текстом, используя методы текстового анализа, и не обращаясь к структурным данным, которые в этом случае являются просто перекодировками. Понятно, что зависимость может быть очень сложной (автор слегка касается этих вопросов на стр. 77), но складывается впечатление, что при наличии большого количества данных методы машинного обучения могут доставать очень сложные зависимости. Тут можно многое сказать за и против, но хотелось бы увидеть обсуждение подобных альтернативных путей. В целом работа оставляет солидное впечатление, содержит ставшие классическими результаты, ставит много новых задач, и безусловно имеет большую научную ценность.

Заключение

Диссертация Пономаренко Михаила Павловича «Компьютерный анализ контекстно-зависимых количественных характеристик специфической биологической активности сайтов в составе геномной ДНК» является законченным научно-исследовательским трудом, открывающим новое направление в анализе последовательностей и структур геномов.

В диссертации разработан ряд методов анализа данных и решен ряд важнейших задач описания и количественного предсказания ДНК-белкового взаимодействия и мутагенеза. Решение этой задачи имеет большое значение для задач компьютерной генетики и геномики (пункт 3 паспорта специальности 03.01.09). По совокупности полученных результатов работа Пономаренко Михаила Павловича соответствует всем требованиям ВАК, предъявляемым к диссертациям на соискание ученой степени доктора биологических наук, п.9 «Положения о присуждении ученых степеней». Результаты достоверны и достаточно полно опубликованы, автореферат адекватно отражает содержание диссертации. Автор заслуживает присуждения ученой степени доктора биологических наук по специальности 03.01.09 - «Математическая биология, биоинформатика».

10 октября 2017 года.

Заведующий отделом вычислительной системной биологии

Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук (ИОГен РАН)

Член-корреспондент РАН, д.ф.-м.н.

Макеев В.Ю.

Почтовый адрес:

Макеев Всеволод Юрьевич,

ИОГен РАН, ул. Губкина, дом 3, ГСП-1, Москва, 119991

Тел.: (499) 135-62-13; Факс: (499) 132-89-62; E-mail: iogen@vigg.ru

Подпись

Руководитель
отдела кадров

