



ФЕДЕРАЛЬНОЕ
ГОСУДАРСТВЕННОЕ
УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«ФУНДАМЕНТАЛЬНЫЕ
ОСНОВЫ БИОТЕХНОЛОГИИ»
РОССИЙСКОЙ АКАДЕМИИ НАУК»



УТВЕРЖДАЮ

Директор

Член-корреспондент РАН

Попов В.О.

17 сентября 2017 г.

119071, Москва, Ленинский пр-т, д. 33, стр. 2
Тел. +7 (495) 954-52-83, факс (495) 954-27-32
www.fbras.ru, info@fbras.ru

№

06.10.2017 12307-2171-647

На №

от

15345-01-9317/3 от 17.07.2017

ОТЗЫВ

ведущей организации на диссертационную работу

М.П. Пономаренко «Компьютерный анализ контекстно-зависимых количественных характеристик специфической биологической активности сайтов в составе геномной ДНК», представленную на соискание

учёной степени доктора биологических наук по специальности 03.01.09

«Математическая биология, биоинформатика»

Диссертационное исследование Пономаренко М.П. посвящено созданию модели и разработке алгоритма количественной оценки силы взаимодействия ДНК-белок, а также применения программ, разработанных на основе этого алгоритма, к реальным биологическим задачам. Разработка такого рода компьютерных программ является актуальной проблемой биоинформатики, потому, что задача точной (с точностью до нуклеотида) локализации сайтов связывания факторов транскрипции до сих пор не решена полностью ни вычислительными ни экспериментальными методами. Подход, предложенный автором, будучи адаптирован для полногеномного исследования, может значительно продвинуть нас в решении этой задачи.

Основные научные результаты работы заключаются в следующем. Разработанная компьютерная система bDNAvideo определяет конформационные и физико-химические свойства данного фрагмента ДНК, которые определяют при моделировании силу ДНК-белковых взаимодействий. Диссертантом было показано, что разные белки распознают определенный набор характеристик ДНК при связывании. Компьютерная система Activity

автоматически определяет белок-специфичную совокупность физико-химических и конформационных свойств ДНК, которые наиболее сильно коррелируют с экспериментально измеренной силой связывания. Другими словами, выявляются те характеристики ДНК, которые обладают наибольшей предсказательной силой. Это делается на основании теории аддитивной полезности для принятия решений и нечетких множеств (Utility Theory for Decision Making, [Fishburn, 1970]). Все найденные таким образом свойства объединяются в линейное регрессионное уравнение, которое в дальнейшем может быть использовано для предсказания взаимодействий белка с другими участками ДНК. Такие уравнения были выведены для 42 транскрипционных факторов. В общей сложности, полученные 42 регрессионных уравнения использовали 848 характеристик (параметров ДНК). Кластеризация проанализированных транскрипционных факторов по выявленным свойствам ДНК (с учетом знаков коэффициентов) разделила данный набор белков на 2 группы. Было показано статистически значимое соответствие между полученными группами и структурными и физико-химическим свойствам соответствующих белков. Система Activity позволяет количественно предсказывать силу связывания данного белка с конкретным участком ДНК. В частности, применимость данного метода была продемонстрирована для белка ТВР, для которого константы диссоциации (K_d), предсказанные с помощью системы Activity, коррелировали с экспериментально-измеренными значениями для разных последовательностей ДНК.

Диссертант также показал возможность применения программы Activity не только к предсказанию ДНК-белковых взаимодействий, но и к определению наиболее мутагенных участков ДНК при внешних воздействиях (например, при воздействии лазером определенной длины волны). Важной особенностью работы является то, что приводится эволюционная интерпретация полученных результатов. Например, как в *in vitro* экспериментах, так и по предсказаниям Activity, наиболее мутагенные нуклеотиды в геноме *E.coli* располагаются в участках, контактирующих с гистонами. Следует отметить, что полноценных гистонов у бактерий не известно. Тем не менее, автор работы отмечает, что существуют шароподобные структуры ДНК, поддерживаемые белком HU, возможно, являются аналогами эукариотических гистонов. Таким образом выдвигается гипотеза о том, что первые гистоноподобные структуры могли возникнуть у бактерий для защиты наиболее "хрупких" участков ДНК от повреждения УФ-излучением.

Интересными, с эволюционной точки зрения, представляются и результаты по специфичности связывания с ДНК бактериального белка RecA, который играет одну из ключевых ролей в репарации ДНК. Анализ экспериментальных данных с помощью программы Activity указал на то, что данный белок имеет тенденцию избегать связывание с последовательностями, содержащими тринуклеотид DRV, где $D=\{A,T,G\}$, $R=\{A,G\}$ и $V=\{A, G, C\}$. При этом отмечается, что аминокислоты, кодируемые таким набором нуклеотидов, чаще являются заряженными и, как правило, отсутствуют в глобулярных ядрах белков, где находятся наиболее функционально-значимые аминокислоты. Таким образом, избегание белком RecA связывания с мотивом DRV, увеличивает его встречаемость в других участках бактериального генома, которые могут кодировать более важные, с точки зрения функции соответствующих белков, аминокислоты.

Новизна данной работы состоит в применении математического аппарата теории аддитивной полезности для принятия решений и нечетких множеств для выявления физико-химических характеристик участков геномной ДНК, которые значимо коррелируют

с количественными величинами, отражающими силу связывания данных участков ДНК-связывающими белками.

Полученные результаты являются значимыми и практически полезными для решения задач биоинформатики. В частности, определение сайтов связывания ТАТА-связывающего белка (ТВР) с геномной ДНК может быть использовано для локализации старта транскрипции генов эукариот, для которых такие старты еще не определены экспериментально. Способность представленного метода оценивать значения равновесных и кинетических констант комплексов ДНК с транскрипционными факторами позволяет делать предсказания о том, как определенные однонуклеотидные замены (SNP) в разных участках генома влияют на связывание ДНК-белок. В некоторых ситуациях это может позволить сделать выводы о влиянии конкретных замен на возникновение и развитие заболеваний, что является может найти свое применение в персонализированной медицине. Предсказания изменения структуры связывания белка в случае появления полиморфизма может особенно помочь в клинической оценке некодирующих полиморфизмов, которые до сих пор изучены недостаточно.

Однако, работа диссертанта не лишена определенных недостатков. Эффективность разработанных моделей предсказания связывания ДНК-белок остается неясной. Это связано с тем, что не была рассчитана так называемая кривая ошибок, предполагающая расчет ошибок первого и второго рода в зависимости от параметров модели. Это приводит к тому, что невозможно сравнить результаты, полученные диссертантом, с работами других авторов по анализу и поиску мест связывания различных белков в последовательностях оснований ДНК. Невозможность такого сравнения не позволяет сделать вывод о ценности разработанных автором алгоритмов в смысле их предсказательной мощности и использования для изучения полных геномов различных видов.

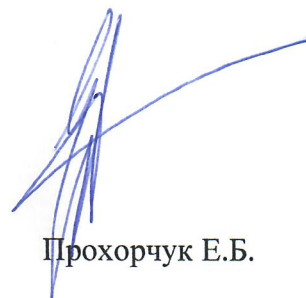
Также можно указать на тот факт, что для большинства моделей учитывается только физические свойства отдельной ДНК молекулы, тогда как в большинстве реальных задач, структура ДНК сильно зависит от ее пространственной конформации, в частности, упаковки в хроматин и намотки на гистоновые белки. Важность этих белков для структуры даже отражена в одной из моделей и показана их лидирующая роль, однако в большинстве разработанных моделей это не учтено.

Отмеченные недостатки не снижают практическую ценность работы в целом. Все указанные результаты опубликованы в научных журналах. Автореферат соответствует диссертации.

Таким образом, диссертация Пономаренко М.П. является научно-квалификационной работой, в которой на основании выполненных автором исследований разработаны теоретические положения о принципах ДНК-белковых взаимодействий, совокупность которых можно квалифицировать как научное достижение, что соответствует требованиям п. 7 «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства РФ от 30.01.2002 г. № 74 (с изменениями, внесенными Постановлением Правительства РФ от 20.06.2011 г. № 475), предъявляемым к диссертациям на соискание ученой степени доктора наук, а ее автор заслуживает присуждения искомой ученой степени по специальности 03.01.09 «Математическая биология, биоинформатика»

Отзыв обсужден на межлабораторном семинаре ФИЦ биотехнологии РАН (протокол № 21 от 19 сентября 2017 года). Файл с аудиозаписью обсуждения отзыва и работы соискателя находится в распоряжении заместителя ученого секретаря ФИЦ биотехнологии РАН к.б.н. Степановой Н.Г.

зав. лабораторией геномики и
эпигеномики позвоночных ФИЦ
биотехнологии РАН, д.б.н.

A handwritten signature in blue ink, consisting of several overlapping, fluid strokes that form a stylized, somewhat abstract shape.

Прохорчук Е.Б.