

ОТЗЫВ

**на автореферат диссертации Михаила Павловича Пономаренко
«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**

Диссертация М.П. Пономаренко посвящена одной из старых, но по-прежнему актуальных проблем современной биоинформатики: предсказанию количественных характеристик активности конкретных участков ДНК по последовательности. Она подытоживает результаты исследований, проводившихся автором на протяжении двадцати лет. В ходе этих исследований автору удалось создать оригинальный подход к решаемой задаче, реализовать его в виде нескольких компьютерных систем и применить к разнообразным биологическим задачам. Существенно, что многие предсказания, сделанные биоинформатическими методами, впоследствии были подтверждены экспериментально. В целом, разработанные системы практически универсальны: их можно применять для предсказания сайтов связывания факторов транскрипции, выявления регуляторных мутаций и, наоборот, предсказания контекстов мутаций и т.п.

Я давно и хорошо знаком с работами Михаила Павловича и оцениваю их весьма высоко. Поэтому я позволю себе не пересказывать автореферат, а приведу лишь возникшие у меня вопросы.

На стр. 20 автор выдвигает эволюционную гипотезу о возникновении гистон-подобных белков бактерий (в частности, HU) в ответ на ультрафиолетовое излучение. Представляется естественным проверить эту гипотезу сравнительно-геномными методами: сопоставить филетические паттерны этих белков с местами обитания бактерий. Современные базы данных (в частности, COG) позволят провести такой анализ силами одного хорошего студента. Более тонкое исследование заключалось бы в сопоставлении филогенетических деревьев данных белков с деревом бактерий, идентификации возможных горизонтальных переносов и

определении времени и места возникновения этих белков (были ли они у LUCA?).

В ряде конкретных исследований большая корреляция между предсказанными и измеренными величинами определяется уникальными, отстоящими далеко от остальных, точками (рис. 12а, рис. 15а). Это стоило бы как-то прокомментировать.

Автор рассматривает очень большое количество параметров. Производились ли поправки на множественное тестирование при оценке статистической значимости?

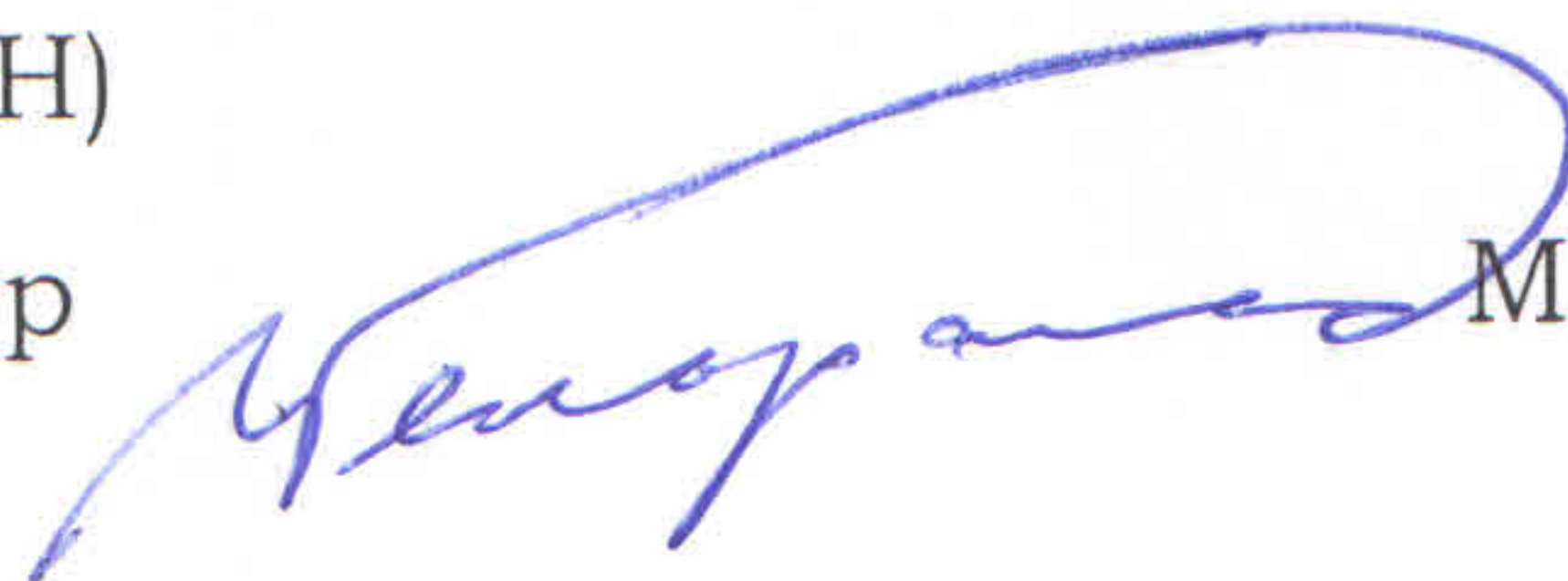
Вообще, автор приводит только результаты успешных предсказаний. Это «ошибка выжившего» — известная проблема всех метаанализов, каковым, по сути, и является диссертация. Было бы весьма интересно узнать, были ли задачи, в которых предлагаемые авторами методы не сработали, сколько, в чем они заключались. Это позволило бы более точно оценить предлагаемые алгоритмы и конкретные формулы, показало границы применимости подхода и, с другой стороны, позволило бы наметить пути дальнейшего развития.

Сложностью для всех больших, продолжительных биоинформатических проектов является быстрое совершенствование экспериментальных методов: с таким трудом полученные результаты быстро устаревают. Так, очень естественным для данной работы было бы сопоставление результатов по предсказанию сайтов связывания транскрипционных факторов с данными ChIP-Seq, а исследование мутационных контекстов — с данными по секвенированию раковых опухолей (отметим, что эта область сейчас активно развивается). Но это уже не замечания, а, скорее, идеи для дальнейшей работы.

Сделанные комментарии ни в коей мере не снижают общей высокой оценки работы. Судя по автореферату и опубликованным статьям, диссертационная работа М.П. Пономаренко «Компьютерный анализ контекстно-зависимых количественных характеристик специфической биологической активности сайтов в составе геномной ДНК» по актуальности, научной новизне и практической значимости результатов, объему проведенных исследований соответствует требованиям п. 9 Положения «О порядке присуждения ученых степеней», утвержденного

Постановлением Правительства Российской Федерации № 842 от 24 сентября 2013 года, а ее автор заслуживает присуждения ученой степени доктора биологических наук по специальности 03.01.09 — математическая биология, биоинформатика.

Заместитель директора по научным вопросам
Федерального государственного бюджетного учреждения науки
Институт проблем передачи информации им. А.А. Харкевича
Российской академии наук (ИППИ РАН)
доктор биологических наук, профессор



М.С. Гельфанд

29.09.2017

Михаил Сергеевич Гельфанд,
ИППИ РАН, 127051, г. Москва, Большой Каретный переулок, д.19, стр. 1.
тел. +7 (495) 650-42-25
e-mail: gelfand@iitp.ru

