

ОТЗЫВ

на автореферат диссертации Кулаковского Ивана Владимировича
«Регуляторные мотивы в геномах высших эукариот и их роль в экспрессии генов»,
представленной на соискание степени доктора биологических наук
по специальности «03.01.09 - Математическая биология, биоинформатика»

Идентификация сайтов ДНК-белкового узнавания всегда была одной из ключевых задач молекулярной биологии и биоинформатики. В последние годы она встала особенно остро, поскольку появилось большое количество методов массового прочтения последовательностей ДНК, вследствие чего объем информации о сайтах связывания белков *in vivo* и *in vitro* растет в геометрической прогрессии. Для анализа мотивов узнавания было создано множество компьютерных алгоритмов, которые позволяют изучать структуру регуляторных участков с одонуклеотидным разрешением. Однако часто вычислительные инструменты не учитывают специфику экспериментальных подходов, и наоборот, зачастую эксперимент планируется таким образом, чтобы облегчить последующую обработку данных, пренебрегая биологической значимостью. Поэтому создание вычислительных инструментов, принимающих во внимание такие биологически-значимые параметры как расположение сайтов связывания относительно структуры ДНК, корреляция между соседними нуклеотидами в мотиве узнавания, покрытие и форма пика ChIP-seq, является, несомненно, одной из самых актуальных на данный момент задач.

Большая часть работы И.В. Кулаковского посвящена идентификации мотивов узнавания в современных массивах экспериментальных данных. Автором был разработан алгоритм ChIPMunk, который построен на основе позиционно-весовых матриц (ПВМ) и решает две последовательные задачи: построение псевдо-оптимального локального выравнивания последовательностей ДНК для поиска наиболее похожего на мотив «слова» и подбор порогового значения оценки весовой матрицы для деления полученного выравнивания на «сайты» и «не-сайты» на основе тестирования самосогласованности матрицы. Среди множества возможных выравниваний ChIPMunk выбирает мотив с наилучшим дискретным информационным содержанием. Кроме того, разработанный автором алгоритм учитывает такую важную характеристику как форма мотива – может быть использован один из вариантов, однобуксовый или двухбуксовый. Очень полезны модификации diChIPMunk, в основе которого лежат учитывающие корреляции соседних нуклеотидов диПВМ, и так называемая «peak mode», позволяющая осуществлять поиск мотива только в вершине пика, с учетом количества прочтений в каждой его позиции.

ChIPMunk позволяет анализировать различные типы данных и может быть использован для поиска сайтов связывания регуляторных белков как в эукариотических, так и в прокариотических геномах. Вторая часть диссертации посвящена практическому применению разработанных вычислительных методов для аннотации регуляторных последовательностей в геномах эукариот. Зачастую в базах данных представлено несколько возможных мотивов для

одного фактора транскрипции, и сложно понять, какой из них является реальным. Важным результатом диссертации И.В. Кулаковского является создание коллекции мотивов связывания для нескольких сотен транскрипционных факторов человека и мыши HOCOMOCO, где для одного фактора альтернативные мотивы приведены только в том случае, если их существование доказано экспериментально. Очень интересным кажется результат, касающийся композитного мотива связывания факторов транскрипции плюрипотентных клеток OCT4 и SOX2. Кроме того, интересно наблюдение о том, что локально-метилованные CpG сайты (CpG светофоры) систематически избегаются транскрипционными факторами. Роль этого явления в регуляции транскрипции генов еще только предстоит понять.

Работа выполнена на высоком научном и методическом уровне. Автореферат написан хорошим языком и легко читается.

Диссертационная работа Кулаковского Ивана Владимировича «Регуляторные мотивы в геномах высших эукариот и их роль в экспрессии генов» соответствует п.9 Положения «О порядке присуждения ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, с изменениями Постановления Правительства Российской Федерации от 21 апреля 2016 года №335, а ее автор заслуживает присуждения искомой степени доктора биологических наук по специальности «03.01.09 - Математическая биология, биоинформатика».

Тутукина Мария Николаевна
Старший научный сотрудник, к.б.н..
Лаб. Функциональной геномики и клеточного стресса
ФГБУН Институт биофизики клетки РАН
Институтская ул., 3, Пущино, Московская область
142290, Россия
Тел.: +7 4967 739140
Эл.почта: maria@icb.psn.ru
masha306@gmail.com

Тутукина М. Н.

«_17_»_октября__2017 г.



Подпись _____
Удостоверяю _____
Тутукиной М. Н.
зав. лаб.