



УТВЕРЖДАЮ

Директор

ФГБУН Институт математических проблем
биологии РАН – филиал

ИММ им. М.В. Келдыша РАН

д.ф.-м.н., проф. В. Д. Лахно

«29 » августа 2016 г.

ОТЗЫВ

ведущей организации на диссертационную работу

Виноградовой Светланы Владимировны

на тему

“Предсказание структурных элементов РНК с
использованием экспериментальных данных”,

представленную на соискание ученой степени кандидата биологических
наук по специальности 03.01.09 (математическая биология,
биоинформатика).

Актуальность для науки и практики

РНК – одна из основных биологических макромолекул; она играет важную роль как в синтезе белка, так и в регуляции различных внутриклеточных процессов. Функционирование РНК в значительной мере определяется ее вторичной структурой. Поэтому разработка биоинформационических методов предсказания структуры РНК с использованием экспериментальных данных нового типа, чему посвящена диссертация С.В. Виноградовой, безусловно, является актуальной.

Новизна и значимость основных научных результатов

Классический подход к *de novo* поиску структурированных РНК в геномных последовательностях – сканирование последовательностей окном с целью поиска сегментов с низкой энергией и, следовательно, стабильной вторичной структурой. При этом используется только информация о последовательности и не принимаются во внимание доступные данные по экспериментальному определению вторичной структуры РНК, являющиеся важным альтернативным источником структурной информации. В работе С.В. Виноградовой было проанализировано несколько типов экспериментальных данных по определению вторичной структуры РНК. Был разработан теоретический подход для преобразования экспериментальных с целью дальнейшего включения их в энергетическую модель программы RNASurface.

Основные результаты работы:

1. Разработана методика анализа и преобразования экспериментальных данных, касающихся принадлежности отдельных нуклеотидов к вторичной структуре РНК. Методика основана на сопоставлении каждому нуклеотиду количественной характеристики, отражающей его склонность быть включенным во вторичную структуру.
2. Проведено сравнение данных эксперимента ДМС *in vivo* и *in vitro*. Показано, что кодирующие области мРНК являются менее структуризованными в клетке по сравнению с состоянием *in vitro*.
3. Алгоритм RNASurface расширен на случай использования в оценке степени структурированности фрагмента экспериментальных данных. Построена фоновая модель для оценки структурированности РНК, учитывающая как энергетические параметры, так и экспериментальные данные.
4. Разработан и запущен веб-сервис RNASurface (<http://bioinf.fbb.msu.ru/RNAsurface/>), позволяющий визуализировать

результаты работы алгоритма по предсказанию структурированных элементов РНК с использованием экспериментальных данных.

5. На основании данных эксперимента PARS проведен анализ структурированности РНК элементов в масштабах транскриптома человека, показавший, что использование экспериментальных данных при поиске структурированных элементов РНК позволяет улучшить качество предсказания.

Материалы исследований по теме диссертации представлены в двух статьях, опубликованных в двух журналах, и были представлены на международных конференциях: Интеллектуальные системы молекулярной биологии (Intelligent Systems for Molecular Biology – ISMB'14), Европейская конференция по вычислительной биологии (European Conference on Computational Biology – ECCB'14), Седьмая Московская конференция по вычислительной молекулярной биологии (Moscow Conference on Computational Molecular Biology – МССМВ'15), Симпозиум Европейской Организации Молекулярной Биологии и Европейского Института Биоинформатики (EMBO/EMBL Symposium'15), а также на конференциях Информационные Технологии и Системы (ИТИС'14, ИТИС'15).

Структура работы

Диссертация состоит из введения, обзора литературы, двух глав с результатами, выводов и списка литературы. Общий объем работы составляет 103 страницы, включая 3 таблиц и 36 рисунков. Библиография содержит 107 ссылок.

Обзор литературы достаточно полон. Он разбит на 5 подразделов, посвященных соответственно разнообразию мира РНК, свойствам структур РНК, экспериментальным методам определения структуры РНК, алгоритмам предсказания вторичной структуры РНК и сканирования

генома для поиска структурированных участков. Такая структура обзора облегчает чтение.

В первой главе представлен анализ существующих экспериментальных методик определения принадлежности отдельных нуклеотидов к вторичной структуре РНК, а также описано разработанное автором диссертации единообразное представления экспериментальных данных, полученных с помощью различных методик. Это представление используется при поиске структурированных РНК.

Во второй главе описан алгоритм, позволяющий учитывать экспериментальные данные при поиске структурированных РНК. Данный алгоритм реализован на основе программы RNASurface и применен к транскрипту человека. Выполнена оценка эффективности алгоритма и сравнение с другими программами, предсказывающими вторичные структуры.

Практическая значимость

Тема диссертации является крайне актуальной, ввиду быстрого роста количества как РНК, с известной функцией, так и появления качественных данных по экспериментальному определению структур РНК. Многие виды РНК ассоциированы с болезнями, в частности с раковыми заболеваниями. Таким образом, инструменты, представленные в диссертации, могут помочь в изучении и классификации новых видов РНК, в частности ассоциированных с болезнями.

Рекомендации по использованию результатов и выводов диссертации

Разработанный алгоритм учета экспериментальных данных при полногеномном предсказании структурированных элементов РНК имеет методическую значимость и может быть использован в курсах биоинформатики для студентов университетов. В целом, описанные в работе подходы и результаты могут быть использованы в Федеральном государственном бюджетном учреждении науки Институте математических

проблем биологии Российской академии наук — филиале Федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»; Федеральном государственном бюджетном учреждении науки Институте общей генетики им. Н.И. Вавилова Российской академии наук; Федеральном государственном бюджетном учреждении науки Институте молекулярной генетики Российской академии наук; Федеральном государственном бюджетном учреждении науки Институте биологии гена Российской академии наук; Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук; Федеральном государственном бюджетном образовательном учреждении высшего образования «Московском государственном университете имени М.В.Ломоносова», а также в других научно-исследовательских организациях.

Общие замечания

Текст диссертации не свободен от недостатков. Местами текст написан тяжеловесно. Например, подраздел «3.1.1 Поиск структурированных сегментов в ортологичных последовательностях» начинается предложения, которое занимает более 6 строк. На стр. 58 написано «Максимизацию суммарной меры спаривания нуклеотидов $\xi[i, j]$ данного участка $[i, j]$ последовательности k проводили с использованием стандартного алгоритма Нуссинов [64] с весами, полученными на первом шаге.»; при этом трудно понять, о каком первом шаге идет речь.

В диссертации также встречаются неточности. Так, в подписи к рис.1 на стр. 61 говорится об «энергии стэкинга», в то время, как из рисунка видно, что речь идет об энергии водородной связи. Во второй главе следовало бы подробнее остановиться на том, почему для сравнения был выбран именно алгоритм Evofold.

Однако эти недостатки не относятся к основным результатам работы и не могут быть причиной её низкой оценки.

Заключение

Диссертация представляет собой завершенную научно-исследовательскую работу на актуальную тему. Новые научные результаты, полученные диссидентом, имеют существенное значение для биологических наук, и могут быть использованы специалистами в области медицины, фармакологии, биоинженерии и биотехнологии. Выводы и рекомендации достаточно обоснованы. Работа отвечает требованиям п. 9 «Положения о порядке присуждения ученых степеней», предъявляемым к кандидатским диссертациям, а её автор заслуживает присуждения ей ученой степени кандидата биологических наук по специальности 03.01.09 — «Математическая биология, биоинформатика».

Отзыв на диссертацию и автореферат обсужден на расширенном семинаре лаборатории прикладной математики ИМПБ РАН - филиала ИПМ им. М.В. Келдыша РАН 21.04.2016 г., протокол № 6

Научный сотрудник лаборатории прикладной математики
ФГБУН ИМПБ РАН — филиала ИПМ им. М.В. Келдыша РАН

к.ф.-м.н.
22 апреля 2016 г.

В.В. Яковлев
тел. (4967) 318530

Подпись сотрудника лаборатории прикладной математики, к.ф.-м.н.
Виктора Владимировича Яковлева удостоверяю

Начальник отдела кадров
ИМПБ РАН - филиала ИПМ им. М.В. Келдыша РАН



Т.А. Галушко