

О Т З Ы В

официального оппонента

на диссертацию на соискание ученой степени

доктора химических наук Первушина Дмитрия Давидовича

**на тему: «Альтернативный сплайсинг и дальние взаимодействия
в структуре эукариотических РНК»**

по специальности

1.5.3. Молекулярная биология

Диссертационная работа Первушина Д.Д. – комплексное и многоуровневое исследование в области молекулярной биологии на стыке с биоинформатикой, биохимии и молекулярной биологии, и посвящена изучению чрезвычайно сложного объекта: элементам вторичной структуры РНК, образуемым через дистальные взаимодействия участков последовательности и роли этих элементов в регуляции сплайсинга. Сегодня изучение структуры РНК с помощью традиционных и новых высокопроизводительных экспериментальных методов, поддержанных биоинформатическим инструментарием и продвинутыми вычислительными методами, получило второе рождение, и становится возможным поиск принципиально новых элементов РНК в масштабах полных геномов и транскриптомов и установление их роли в функционировании клеточных процессов, в

частности, в реализации генетической информации. Таким образом, разработка новых методов предсказания и аннотации элементов структуры РНК, в том числе, учитывающих ранее недоступные для полноценных исследований взаимодействия, например, дальние взаимодействия элементов последовательности, является не просто актуальной задачей, но имеющей важное фундаментальное значение. В свою очередь, альтернативные элементы структуры РНК имеют критическое значения в регуляции экспрессии генов на пост-транскрипционном уровне. В частности, сюда относится и процесс альтернативного сплайсинга, определяющий множество процессов на уровне клетки и всего организма, и, в случае нарушений, вызывающих множество сложных патологий. Совокупно, это позволяет заключить, что **актуальность** как и **теоретическая и практическая значимость** работы не вызывают сомнений.

В работе предложены новые методы предсказания дистальных взаимодействий между элементами последовательности РНК, применимые в масштабе полного эукариотического генома, впервые описана взаимосвязь между расположением элементов вторичной структуры РНК и экзон-интронной картой генов, сайтов сплайсинга и других функциональных элементов, впервые показано изменение сплайсинга в зависимости от структурированности РНК при замедлении элонгации транскрипции. Впервые изучены характеристики РНК-структур с учетом новых данных данными конформационного секвенирования РНК, предложен новый метод предсказания структуры РНК вне участков консервативности. Наконец, выявлены новые механизмы регуляции сплайсинга и

получено первое экспериментальное подтверждения влияния элементов структуры РНК-продуктов конкретных генов плодовой мушки и человека и разработаны антисмысловые олигонуклеотиды для управления сплайсингов через структуру РНК.

Таким образом, убедительно подтверждается **новизна** исследования. Результаты работы блестяще опубликованы, однако же это, не смотря на заявление автора, не является критерием достоверности результатов и выводов: в последние годы участились случаи публикации недостоверных данных и отзыва статей из высокорейтинговых журналов. На мой взгляд, достоверность и обоснованность основных положений и выводов диссертации обусловлена тщательным сравнением различных источников и типов экспериментальных данных, и успешной экспериментальной верификацией ключевых результатов, что великолепно продемонстрировано автором.

Диссертация изложена на 248 страницах машинописного текста, богато проиллюстрирована, список литературы подробен и включает 575 наименований. Изложение текста построено по традиционной схеме, с выделенными разделами «Введение», «Обзор литературы», «Материалы и методы», и четырьмя главами (с 3 по 6), где излагаются основные научные результаты диссертанта. Текст работы также содержит список ключевых сокращений, что в определенной мере помогает прочтению. В целом, рисунки оформлены чрезвычайно аккуратно, но, к сожалению, сохранили и отдельные остатки английских подписей, например неясно в чем принципиальная

техническая разница между «обогащением» и «OR» (видимо, отношение шансов) на панелях А и Б рисунка 3.14.

В дополнение ко всему вышесказанному, хочется отдельно и явным образом подчеркнуть выраженные достоинства работы. Сюда входит разнообразие методов и подходов, использованных в работе, от алгоритмической биоинформатики до анализа омиксных данных, от классических низкопроизводительных биохимических и молекулярно-биологических методов до полногеномных крупномасштабных экспериментов. Результаты работы Первушина Д.Д. вошли в крупные проекты наиболее известного в области регуляторной геномики и невероятно престижного международного консорциума ENCODE. Диссертация, не смотря на весь объем и сложность материала, не выглядит разрозненным сборником статей, а выстроена логически и содержательно с единым посылом, и демонстрирует безусловную академическую зрелость автора, понимание ключевых теоретических проблем в области, успех в нетрадиционной постановке задач, включая нахождение и доказательно функциональности принципиально новых модификаторов регуляции экспрессии генов на уровне сплайсинга. Наконец, фундаментальные научные достижения выгодно иллюстрированы переходом к практическим применениям идей на основе антисенс-олигонуклеотидов.

К сожалению, не смотря на высочайшую планку, задаваемую уровнем научных статей, общей связностью и логикой текста, масштабом и кропотливостью проведенных исследований, работа не

лишена небрежностей в работе с фактическим материалом и шероховатостей, которые затрудняют восприятие текста.

Во-первых, автор драматически злоупотребляет сокращениями, которые («АС», «РСБ») просочились даже в заголовки разделов. Список сокращений помогает только частично, поскольку многие сокращения в него не попали: «ТРМ», «КСПА», «AUC», «SPI», «ROC». Случайным образом перемешаны латинские и русские сокращения, поэтому некоторые «двойки» аббревиатуры особенно трудно читать (например РТС). Позволю себе процитировать экстремальный пример из авторского текста: «.. на наличие события АС перед КСПА. STE, так и STE должны иметь $\Psi=0$ из-за отсутствия .. с преобладанием b в случае STE и преобладанием c в случае STE».

Во-вторых, объекты и методы исследования (стр. 15) описаны небрежно. По авторскому тексту можно подумать, что «построение хэш-таблиц» производилось вручную в табличном редакторе; а программированием алгоритмов и разработкой научного программного обеспечения автор не занимался вовсе. При этом и в этом разделе и далее недостаточно четко прописан личный вклад автора: какие части работы были выполнены непосредственно его руками и руками учеников и под его руководством, а какие этапы были отданы на откуп партнерам. Учитывая критический личный вклад автора во все основные результаты этот момент следовало более явно и ярко отразить непосредственно в тексте диссертации.

В-третьих, по всему тексту работы упоминания видов приводятся несистемно или вообще отсутствуют (в том числе в контексте описания конкретных генов и механизмов их регуляции; и даже в выводах

работы). Имена генов человека часто не соответствуют принятой на сегодня нотации по базе данных HUGO Gene Nomenclature Committee (например hTERT вместо TERT).

В-четвертых, противоречивые чувства вызывает Обзор литературы. С одной стороны, хочется похвалить за четкость, компактность, читаемость, и охват всех ключевых для работы вопросов. С другой стороны, недостает аккуратности: так, утверждение о всеобъемлющей роли сплайсинга у эукариот неверно как минимум для дрожжей. Далее, дихотомии «термодинамика и динамическое программирование» против «филогенетика и ковариации» вряд ли можно признать полностью корректными. Было бы грамотнее разделить методы предсказания структур на использующие одиночную последовательность и несколько выровненных и нет последовательностей. Автор опускает важное методическое направление разработки алгоритмов с учетом псевдоузлов (например Ipknот, ShapeKnots и алгоритмы на основе метода McCaskill). Еще одно замечание касается NP-полноты задачи предсказания псевдоузлов: это верно только в рамках модели минимизации свободной энергии, когда одним элементарным объектом выступает стекинг, т.е. пара спариваний. Если элементом выступает одно спаривание, например в алгоритмах типа Нуссинов, которые максимизирует число спариваний, то задача относится к классу полиномиальных. Впрочем, это заблуждение широко распространено в научной среде. Наконец, жаль, что в литературном обзоре не нашлось места для обсуждения связи регуляции сплайсинга

и эпитранскриптомных модификаций, за исключением короткого упоминания METTL3 на странице 39.

Текст, с одной стороны, практически полностью лишен опечаток (за исключением, быть может, «Бонферони» на странице 58 и «Приоритизации» на странице 102). С другой стороны, автору не удалось избавиться от жаргона. Например, систематически идет речь об «ответе транскриптома», хотя более грамотно говорить об «изменения транскриптома в ответ на ...». На стр. 71 утверждается что позиционно-весовые матрицы измеряют степень сходства относительно консенсуса, что является чрезмерным упрощением. Конкретные структурированные элементы или взаимодействия участков последовательности систематически называются «структурами РНК», что также не вполне корректно.

Есть и вопросы по содержанию, по которым было бы особенно интересно услышать мнение автора по существу.

(1) Например, автор разбирает конкретный пример, когда далекие комплементарные участки на самом деле являются цис-регуляторными элементами уровня транскрипции, а именно, участками связывания факторов транскрипции. Интересно, является ли этот эффект систематическим и можно ли это проверить используя доступные уже достаточно полные на сегодня каталоги мотивов-паттернов, узнаваемых факторами транскрипции человека и плодовой мушки.

(2) На стр. 83 утверждается что «пересечение предсказаний двух методов было большим, чем предсказания одного», не вполне понятно как это может быть реализуемо.

(3) На стр. 111 обсуждается достаточно низкий уровень согласованности повторностей RIC-seq. Правомерно ли при этом утверждать, что уникальные для повторностей наблюдения являются биологически обоснованными, а не представляют собой технический шум эксперимента.

Вместе с тем, указанные замечания носят дискуссионный характер и не ставят под сомнение значимость и бескомпромиссно высокий уровень представленного диссертационного исследования. Работа Первушина Д.Д. **несомненно отвечает требованиям**, установленным Московским государственным университетом имени М.В. Ломоносова к работам подобного рода. Содержание диссертации **соответствует специальности 1.5.3. Молекулярная биология** (по химическим наукам), а именно следующим ее направлениям: теоретическое и экспериментальное изучение механизмов хранения, передачи и реализации генетической информации, строение и функции сложных высокомолекулярных соединений, составляющих клетку: нерегулярных биополимеров (белки и нуклеиновые кислоты), а также их макромолекулярные комплексы; биосинтез нуклеиновых кислот и белка; биоинформатика; молекулярное моделирование, в том числе предсказание и дизайн белковых структур, предсказание вторичных структур РНК, а также **критериям, определенным пп. 2.1-2.5 Положения** о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова. Работа оформлена в соответствии с требованиями Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Московского

государственного университета имени М.В. Ломоносова.

Таким образом, соискатель Первушин Дмитрий Давидович без сомнений заслуживает присуждения ученой степени доктора химических наук по специальности 1.5.3. Молекулярная биология.

Официальный оппонент:

Кулаковский Иван Владимирович

доктор биологических наук,

ведущий научный сотрудник,

Федеральное государственное бюджетное учреждение науки

Институт белка Российской академии наук

10 сентября 2024