

**ОТЗЫВ официального оппонента
на диссертацию на соискание ученой степени
кандидата биологических наук Рябых Григория Кирилловича
на тему: «РНК-хроматиновые взаимодействия: базы данных,
интегративный анализ и функциональная аннотация»
по специальности 1.5.8. Математическая биология, биоинформатика**

АКТУАЛЬНОСТЬ ТЕМЫ

В последние годы стало очевидно, что некодирующие РНК являются не побочным шумом транскрипции, а важнейшим регуляторным слоем генома. Они участвуют в регуляции экспрессии генов, поддержании пространственной организации хроматина, эпигенетическом контроле, формировании ядерных доменов и других фундаментальных процессах. Нарушения функций некодирующих РНК связаны с развитием онкологических, нейродегенеративных и аутоиммунных заболеваний, что придает исследованиям в этой области не только фундаментальное, но и значительное биомедицинское значение.

Одним из ключевых направлений современной функциональной геномики является изучение РНК-хроматиновых взаимодействий. Для этого разработаны высокопроизводительные экспериментальные методы двух основных классов: «один-против-всех» (OTA), позволяющие определить сайты связывания конкретной РНК в геноме, и «все-против-всех» (ATA), направленные на глобальное картирование взаимодействий всех РНК с хроматином. Однако бурное накопление экспериментальных данных породило новую проблему: данные, полученные разными методами и лабораториями, часто обрабатываются различными вычислительными протоколами, что затрудняет их сопоставление, переиспользование и интегративный анализ.

В этом контексте диссертационная работа Рябых Г.К. является несомненно актуальной. Она направлена на решение сразу нескольких взаимосвязанных задач: создание унифицированного протокола обработки данных РНК-хроматинового интерактома, разработку специализированной

базы данных RNA-Chrom, интеграцию этих данных с другими типами функционально-геномной информации, а также количественную оценку разрешения, полноты, специфичности и воспроизводимости данных ОТА и АТА.

НАУЧНАЯ НОВИЗНА

Диссертационная работа обладает выраженной научной новизной. Прежде всего, автором разработан и применен единый стандартизированный вычислительный протокол для обработки данных РНК-хроматинового интерактома, полученных методами ОТА и АТА. Это имеет принципиальное значение, поскольку именно неоднородность вычислительной обработки является одним из основных препятствий для сопоставления ранее опубликованных наборов данных.

Одним из центральных результатов работы является создание аналитической базы данных RNA-Chrom, аккумулирующей и унифицирующей полногеномные данные РНК-хроматиновых взаимодействий для человека и мыши. Согласно тексту диссертации, ресурс включает более 5 миллиардов контактов, полученных в 20 экспериментах АТА и 189 экспериментах ОТА. База данных снабжена веб-интерфейсом, позволяющим проводить анализ как «от РНК», так и «от ДНК», что делает ее не просто хранилищем данных, а полноценным аналитическим инструментом.

Существенным новым результатом является разработка инструмента Fastq-dupaway для удаления ПЦР-дубликатов из данных секвенирования нового поколения. Особое значение имеет ориентация программы на ресурсоэффективность, предсказуемое низкое потребление оперативной памяти и возможность работы с большими наборами данных. Такой инструмент полезен не только для анализа РНК-хроматиновых взаимодействий, но и для более широкого круга задач NGS-биоинформатики.

К числу оригинальных результатов следует отнести интеграцию RNA-Chrom с ресурсом HiMoRNA, что позволило связать физические контакты длинных некодирующих РНК с хроматином с данными о корреляциях между экспрессией РНК и эпигенетическими метками. Такая схема переводит анализ из плоскости описания контактов в плоскость генерации функциональных гипотез.

Наконец, важным новым вкладом является полномасштабный сравнительный анализ данных РНК-хроматинового интерактома, позволивший количественно сопоставить методы ОТА и АТА по разрешению, полноте, специфичности и воспроизводимости. В работе предложены и использованы метрики, включая хроматиновый потенциал и конкордантность контактов в репликах. Это делает диссертацию не только ресурсной и программной работой, но и методологическим исследованием, задающим ориентиры для дальнейшего анализа подобных данных.

ТЕОРЕТИЧЕСКАЯ И ПРАКТИЧЕСКАЯ ЗНАЧИМОСТЬ

Теоретическая значимость диссертационного исследования состоит в развитии методологии анализа РНК-хроматинового интерактома. Автор предлагает более строгую рамку для оценки данных, полученных разными экспериментальными протоколами, и показывает, что методы ОТА и АТА должны рассматриваться не как взаимозаменяемые, а как взаимодополняющие подходы с различными областями применения, ограничениями и сильными сторонами.

Работа вносит вклад в формирование целостного представления о том, как данные физических РНК-хроматиновых контактов могут быть связаны с эпигенетическими профилями, аннотацией генов и функциональными гипотезами о роли длинных некодирующих РНК. Особенно важно, что автор не ограничивается разработкой базы данных, а использует ее как инструмент для систематического анализа качества, согласованности и биологической интерпретируемости данных.

Практическая значимость работы также очевидна. Созданная база данных RNA-Chrom представляет собой готовый к использованию ресурс для научного сообщества. Она позволяет исследователям работать с унифицированными и предварительно обработанными данными, что существенно снижает порог входа в анализ РНК-хроматиновых взаимодействий. Разработанный веб-интерфейс делает ресурс доступным не только для специалистов по биоинформатике, но и для экспериментальных биологов.

Программа Fastq-dupaway имеет самостоятельную практическую ценность как инструмент для обработки NGS-данных. Ее ресурсоэффективность особенно важна в условиях, когда размер секвенировочных данных растет быстрее, чем доступные вычислительные мощности в отдельных лабораториях.

Интеграция RNA-Chrom с HiMoRNA демонстрирует перспективность мультиомиксного подхода и может служить прототипом для дальнейшего подключения к RNA-Chrom других типов данных: Hi-C, ChIP-seq, ATAC-seq, RNA-seq и других. Предложенная стратегия анализа РНК-хроматинового интерактома может быть использована при планировании новых экспериментов и при интерпретации уже опубликованных данных.

ХАРАКТЕРИСТИКА ОСНОВНОГО СОДЕРЖАНИЯ ДИССЕРТАЦИИ

Диссертационная работа имеет традиционную и логичную структуру. Она включает введение, обзор литературы, материалы и методы, результаты и их обсуждение, заключение, основные результаты и выводы, список сокращений, список литературы и приложения. Работа изложена на 157 страницах, иллюстрирована 50 рисунками и 19 таблицами, содержит 3 приложения и список литературы из 203 источников. Такой объем и структура соответствуют характеру комплексной биоинформатической

диссертации, сочетающей обзорную, методическую, программную, ресурсную и аналитическую части.

Во введении сформулирована актуальность исследования. Автор последовательно показывает, что область РНК-хроматиновых взаимодействий быстро развивается, но страдает от фрагментарности данных, отсутствия единой обработки, нехватки специализированных ресурсов и недостатка систематического сравнительного анализа методов ОТА и АТА. Цель работы сформулирована четко: количественно оценить разрешение, полноту выявления, специфичность и воспроизводимость данных РНК-хроматинового интерактома, а также создать инфраструктуру для их стандартизированной обработки, хранения, анализа и интеграции. Поставленные задачи логично вытекают из цели и охватывают все основные компоненты исследования: сбор и курирование данных, разработку протокола обработки, создание базы данных, веб-интерфейса, интеграцию с HiMoRNA, разработку метрик и формулирование практических рекомендаций.

Глава «Обзор литературы» демонстрирует глубокое владение автором предметной областью. В ней рассмотрены методы ОТА и АТА, их экспериментальные особенности, биологические примеры применения, существующие базы данных некодирующих РНК и вопросы обработки секвенировочных данных, включая проблему ПЦР-дубликатов. Автор подробно рассматривает не только алгоритмическую сторону, но и биологический смысл изучаемых взаимодействий.

Глава «Материалы и методы» содержит описание ключевых вычислительных и статистических подходов, использованных в работе. Подробно изложен универсальный протокол обработки данных RNA-Chrom, включающий удаление ПЦР-дубликатов, контроль качества чтений,

картирование, определение ориентации РНК-частей контактов, CIGAR-фильтр, BlackList-фильтр, аннотацию контактов генами, сборку usaRNAs и перевзвешивание контактов. Отдельно описаны подходы к тестированию ПЦР-дедупликатора, интеграции HiMoRNA и RNA-Chrom, а также сравнительному анализу данных. Раздел производит впечатление методически продуманного и воспроизводимого исследования.

Глава «Результаты и их обсуждение» является центральной частью диссертации и включает четыре крупных смысловых блока.

Первый блок посвящен созданию базы данных RNA-Chrom. Автор описывает предобработку данных, структуру базы, разработку веб-сервиса и функциональные возможности ресурса. Особенно важны режимы анализа «от РНК» и «от ДНК», поскольку они отражают две естественные исследовательские логики: поиск геномных мишеней конкретной РНК и поиск РНК, контактирующих с интересующим локусом. Дополнительные страницы веб-ресурса расширяют его полезность и делают базу данных не статичным архивом, а рабочей аналитической средой.

Второй блок посвящен Fastq-dupaway. Автор описывает два режима работы программы, «sequence-based» и «fast», дает рекомендации по их выбору и сравнивает инструмент с существующими de novo-based и alignment-based подходами удаления ПЦР-дубликатов. Эта часть работы важна тем, что решает конкретную вычислительную проблему, возникающую при обработке больших объемов NGS-данных, и при этом имеет самостоятельное значение за пределами основной тематики РНК-хроматинового интерактома.

Третий блок посвящен интеграции HiMoRNA и RNA-Chrom. Автор показывает, как можно совместить данные о физических контактах РНК с

хроматином с данными о корреляциях между экспрессией длинных некодирующих РНК и эпигенетическими метками. Примеры MIR31HG и PVT1 демонстрируют потенциал такого подхода для выдвижения функциональных гипотез.

Четвертый блок содержит сравнительный анализ данных РНК-хроматинового интерактома. Автор анализирует хроматиновый потенциал, воспроизводимость реплик в данных АТА и ОТА, сравнение экспериментов разных типов и формулирует стратегию анализа данных РНК-хроматинового интерактома. Важным результатом является вывод о более высоком разрешении и воспроизводимости данных ОТА по сравнению с АТА, а также о необходимости выделения специфического сигнала с использованием хроматинового потенциала и воспроизводимых контактов из статистически значимых пиков.

Сформулированные выводы соответствуют поставленным задачам и логично вытекают из представленных результатов. Автор не только создал программно-аналитическую инфраструктуру, но и использовал ее для получения содержательных выводов о качестве и применимости разных классов данных РНК-хроматиновых взаимодействий.

ДОСТОВЕРНОСТЬ РЕЗУЛЬТАТОВ И ПУБЛИКАЦИИ

Достоверность результатов обеспечивается использованием общедоступных экспериментальных данных, стандартизированной обработкой, сравнением независимых наборов данных, анализом реплик, использованием статистических критериев и открытым характером разработанных ресурсов. Важным фактором является также то, что основные результаты работы опубликованы в рецензируемых научных изданиях.

По теме диссертации опубликовано 5 статей в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете МГУ, включая работы в Database, Scientific Reports, Biochemistry (Moscow), Acta Naturae и «Молекулярной биологии». Кроме того, имеется публикация в «Вестнике Российского фонда фундаментальных исследований». Результаты работы были представлены на 7 российских и международных конференциях. Это свидетельствует о достаточной апробации результатов и их признании научным сообществом.

Личный вклад автора в диссертации описан подробно и выглядит существенным. Автором выполнены литературный анализ, разработка базы данных RNA-Chrom и ее веб-интерфейса, адаптация ресурса для интеграции с HiMoRNA, анализ согласованности результатов, тестирование Fastq-dupaway, анализ хроматинового потенциала и конкордантности реплик, а также сравнительный анализ данных ОТА и АТА.

ЗАМЕЧАНИЯ ПО ДИССЕРТАЦИОННОЙ РАБОТЕ

К диссертационной работе Рябых Григория Кирилловича имеется ряд замечаний, вопросов и пожеланий.

1. В тексте встречаются отдельные редакционные неточности и шероховатости. Например, для обозначения 5-штрих конца ДНК иногда используется символ апострофа.
2. Инструмент Fastq-dupaway представлен как эффективное средство удаления ПЦР-дубликатов. Вместе с тем, хотелось бы более подробно обсудить причины их возникновения (можно было бы нарисовать схему пробоподготовки библиотек) и аргументировать выбранный подход к их удалению (например, почему используется полное совпадение последовательностей и совпадение длины –

длина прочтений обычно является фиксированной величиной, а в хвостах прочтений возможно появление неточностей). Имеет смысл обсудить случаи, в которых удаление дубликатов на уровне последовательностей может приводить к потере биологически значимого сигнала, например для библиотек с низкой сложностью, высокоэкспрессированных РНК, повторных элементов или данных с UMI. Было бы полезно явно обозначить границы применимости разных режимов программы. Кроме того, следовало бы оценить требовательность программы к эффективности работы дисковой подсистемы, привести данные для разных компьютерных конфигураций. Осталось непонятным, нужен ли для успешной работы программы SSD диск или HDD, по какому интерфейсу должно быть подключение для эффективной работы – SATA, NVMe?

3. Разработанный протокол обработки данных для выявления РНК-хроматиновых взаимодействий приведен достаточно постулативно. Его обоснование, валидация, сравнение с другими протоколами является, по-видимому, самой дискуссионной частью данной работы. Во-первых, для воспроизводимости желательно привести более формализованную информацию о версиях всех программных инструментов, параметрах запуска, версиях геномных сборок и аннотаций, а также о возможности запуска всего конвейера в контейнеризированной среде. Это особенно важно для базы данных, претендующей на роль стандартизирующего ресурса. Во-вторых, необходимо привести сравнения с уже опубликованными методами, обсудить несовпадения, аргументировать свой выбор.
4. В работе неоднократно подчеркивается, что база данных RNA-Chrom включает все общедоступные полногеномные данные РНК-хроматинового интерактома для человека и мыши. Было бы полезно более явно указать дату последнего систематического поиска данных и критерии включения или исключения отдельных наборов.

Для быстро развивающейся области такой «срез времени» имеет важное значение.

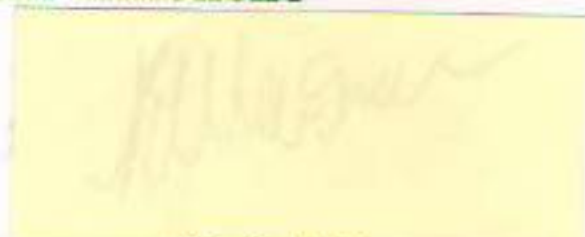
5. В выдаче информации об РНК-хроматиновых взаимодействиях следовало бы предусмотреть ранжирование контактов по значимости и статистической достоверности.
6. В разделе, посвященном интеграции RNA-Chrom и HiMoRNA, используется расширение геномных координат контактов на ± 25 тыс. п.н. как компромиссное значение. Хотя в работе приведено обоснование такого выбора, было бы полезно подробнее обсудить чувствительность результатов к размеру окна, а также возможные биологические различия между близкими цис-контактами и более дальними регуляторными взаимодействиями.
7. В работе сформулирован вывод о более высоком разрешении и воспроизводимости данных ОТА по сравнению с АТА. Этот вывод важен и практически полезен, однако хотелось бы более подробно обсудить, в какой степени он определяется природой самих методов, а в какой степени зависит от качества конкретных опубликованных наборов данных, глубины секвенирования, типов клеток, библиотек и используемых алгоритмов выделения контактов.

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В.Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 1.5.8. Математическая биология, биоинформатика (по биологическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В.Ломоносова. Диссертационное исследование оформлено согласно требованиям 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 Рябых Григорий Кириллович заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8. Математическая биология, биоинформатика.

Официальный оппонент



Шайтан Алексей Константинович
доктор физико-математических наук

(специальность — 03.01.09 — Математическая биология, биоинформатика),

профессор РАН, член-корреспондент РАН,
профессор кафедры биоинженерии биологического факультета
Федерального государственного бюджетного образовательного
учреждения высшего образования «Московский государственный
университет имени М. В. Ломоносова»

3. июня 2025 г.

Контактные данные

Телефон: +7(495) 939-57-38; E-mail: shaytan_ak@mail.bio.msu.ru;

Почтовый адрес: 119991, Москва, Ленинские горы, д. 1, стр. 12, Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова»

Подпись сотрудника биологического факультета МГУ имени М.В.Ломоносова А.К. Шайтана заверяю:

Ученый секретарь биологического факультета

МГУ имени М.В.Ломоносова



Е.В. Петрова