

УДК 577.21

# Интеграция HiMoRNA и RNA-Chrom: подтверждение функциональной роли длинных некодирующих РНК в эпигенетической регуляции генов человека с помощью данных РНК-хроматинового интерактома

И. С. Ильницкий<sup>1,2,3†</sup>, Г. К. Рябых<sup>1,2†</sup>, Д. А. Маракулина<sup>3,4</sup>, А. А. Миронов<sup>1,2</sup>, Ю. А. Медведева<sup>3,4</sup>

<sup>1</sup>Московский государственный университет им. М.В. Ломоносова, факультет биоинженерии и биоинформатики, Москва, 119234 Россия

<sup>2</sup>Институт общей генетики им. Н.И. Вавилова РАН, Москва, 119991 Россия

<sup>3</sup>Институт биоинженерии им. К.Г. Скрыбина, ФИЦ биотехнологий РАН, Москва, 117312 Россия

<sup>4</sup>Школа биомедицинской физики, Московский физико-технический институт, Долгопрудный, Московская обл., 141701 Россия

\*E-mail: nfsus96@gmail.com

†Авторы внесли одинаковый вклад в работу.

Поступила в редакцию 21.10.2024

Принята к печати 31.03.2025

DOI: 10.32607/actanaturae.27543

**РЕФЕРАТ** Известно, что длинные некодирующие РНК, или lncRNA (long non-coding RNA), могут привлекать белки-модификаторы хроматина к определенным геномным локусам, участвуя таким образом в эпигенетической регуляции экспрессии генов. Ранее были созданы две базы данных HiMoRNA и RNA-Chrom. Первая содержит области эпигенетических модификаций (пики), скоррелированные с экспрессией длинных некодирующих РНК, вторая – полногеномные взаимодействия десятков тысяч РНК с хроматином. В данной работе мы интегрировали эти два ресурса, что позволило сгенерировать интерпретируемые и поддерживаемые экспериментальными данными гипотезы о механизмах эпигенетической регуляции экспрессии генов длинными некодирующими РНК. С этой целью мы адаптировали веб-интерфейс HiMoRNA и RNA-Chrom таким образом, чтобы для каждой триады «lncRNA–эпигенетическая модификация–ассоциированный с модификацией ген» из HiMoRNA можно было получить контакты соответствующей lncRNA с конкретным геномным локусом или со всем геномом в RNA-Chrom. В частности, нами показано, что для lncRNA MALAT1, HOXC-AS2, NEAT1, NR2F1-AS1, PVT1, MEG3 большинство пиков HiMoRNA находятся на расстоянии до 25 т.п.н. от контактов данных lncRNA из RNA-Chrom. Подробно исследованы и подтверждены контакты lncRNA MIR31HG и PVT1 с пиками HiMoRNA для меток H3K27ac и H3K27me3 в локусах генов *GLI2* и *LATS2* соответственно, которые, как показано, регулируются данными РНК. Таким образом, интеграция баз HiMoRNA и RNA-Chrom позволяет прояснить роль конкретных lncRNA в регуляции гистоновых модификаций на уровне как отдельных локусов, так и полного генома. Мы считаем, что данная интеграция является удобным и ценным инструментом, который может значительно облегчить функциональную аннотацию lncRNA человека.

**КЛЮЧЕВЫЕ СЛОВА** длинная некодирующая РНК, модификация гистонов, взаимодействие РНК с хроматином.

**СПИСОК СОКРАЩЕНИЙ** нкРНК – некодирующая РНК; lncRNA – длинная некодирующая РНК (long non-coding RNA).

**ВВЕДЕНИЕ**

Клетки человека транскрибируют многочисленные длинные некодирующие РНК (lncRNA), число которых сравнимо с числом белок-кодирующих генов [1, 2]. lncRNA с трудом поддаются функциональной аннотации из-за низкого уровня их экспрессии, тканевой специфичности и низкой консервативности последовательности [3–5]. Тем не менее ряд характеристик lncRNA часто сохраняется, включая синтению с соседними генами, вторичную структуру и сходство коротких фрагментов последовательностей [6]. Кроме того, транскрипция lncRNA регулируется не менее сложно, чем белок-кодирующих РНК, обеспечивая их участие в разнообразных молекулярных механизмах [7]. Подавление lncRNA приводит к значительному изменению транскрипционного профиля клетки [8]. Эти наблюдения позволяют предположить, что многие lncRNA являются функциональными. Большинство lncRNA взаимодействуют с хроматином и вовлечены в эпигенетическую регуляцию геномных локусов, в структурную организацию хромосом [9–12]. Все это делает актуальной идентификацию функциональных геномных мишеней lncRNA, взаимодействующих с хроматином.

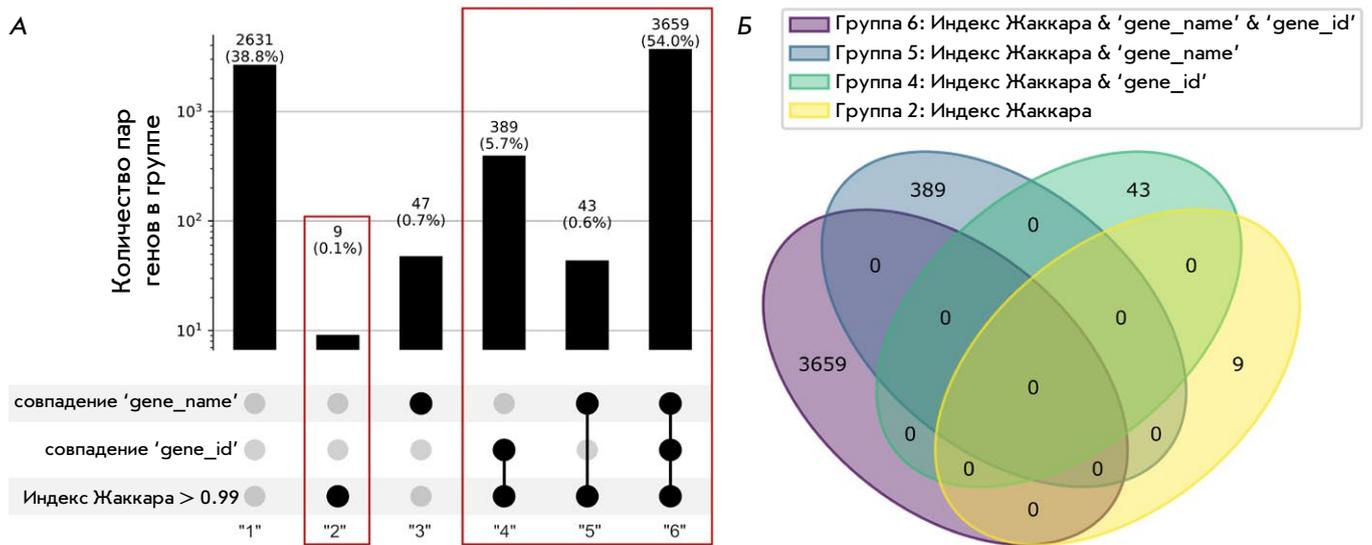
Ранее мы разработали базу данных HiMoRNA [13], которая содержит более 5 млн эпигенетических «пиков» – геномных регионов, в которых наблюдается одна из 10 модификаций гистоновых белков, причем уровень модификации значимо коррелирует с экспрессией одной из длинных некодирующих РНК не менее чем в 20 клеточных линиях и тканях. Пики гистоновых меток, представленных HiMoRNA, там, где это возможно, ассоциированы с генами, образуя триады «lncRNA–пик эпигенетической модификации–ассоциированный ген». Предполагается, что lncRNA в триаде потенциально может регулировать соответствующий ген посредством влияния на установление (1) или удаление (2) модификации гистона в районе ассоциированного с геном пика. В случае (1) пики будут положительно коррелировать с lncRNA («+» пики), а в случае (2) – отрицательно («–» пики). Таким образом, результаты, представленные в базе данных HiMoRNA, позволяют выдвигать предположения о роли lncRNA в установке или удалении эпигенетических модификаций в конкретных геномных локусах и, следовательно, в их влиянии на экспрессию генов. Однако для выдвижения разумных гипотез такого рода и их дальнейшей экспериментальной проверки, список из 5 млн пиков необходимо отфильтровать, выбрав наиболее надежные из них. Экспериментальные методы обнаружения взаимодействий РНК с хроматином позволяют получить ценные данные для этой цели.

Существует ряд экспериментальных методов, которые позволяют определить области хроматина, с которыми контактирует некодирующая РНК. Мы можем разделить их на две группы: «один-против-всех» [11, 14–18] и «все-против-всех» [19–24]. Первая группа методов определяет контакты конкретной РНК с хроматином, а вторая направлена на выявление всех возможных контактов РНК–ДНК в клетке [25]. К сожалению, все эти экспериментальные методы генерируют много ложноположительных результатов, а методы «все-против-всех» плохо детектируют контакты слабо экспрессирующихся РНК и перегружены контактами зарождающихся транскриптов. Несмотря на это, полногеномные данные взаимодействия некодирующих РНК (нкРНК) с хроматином очень важны для понимания механизма их действия. В связи с этим недавно была создана база данных RNA-Chrom [26], которая содержит экспериментальные данные контактов тысяч РНК с хроматином и предоставляет два типа анализа этих данных («от РНК» и «от ДНК»), которые можно использовать в исследовательских целях.

Чтобы улучшить и значительно облегчить функциональную аннотацию lncRNA, мы интегрировали базы данных HiMoRNA и RNA-Chrom. Для этого веб-интерфейс HiMoRNA и RNA-Chrom был модифицирован таким образом, чтобы обеспечить прямой доступ к контактам 4124 из 4145 lncRNA из HiMoRNA с хроматином в RNA-Chrom. Наш подход позволяет генерировать гипотезы относительно механизмов эпигенетической регуляции экспрессии генов человека длинными некодирующими РНК, подкрепленные экспериментальными данными о взаимодействиях соответствующих РНК с хроматином. Мы считаем, что эта интеграция является удобным и ценным ресурсом, который поможет определять надежные триады «lncRNA–эпигенетическая модификация–ассоциированный с модификацией ген» для их экспериментальной проверки. База данных HiMoRNA доступна пользователям по адресу <https://himorna.fbras.ru> (по состоянию на 20.10.2024).

**ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ****Интеграция баз данных HiMoRNA и RNA-Chrom**

Базы данных используют разные источники аннотации генов. Поэтому, во-первых, необходимо было установить соответствие между генами из разных версий GENCODE-аннотаций, так как для длинных некодирующих РНК HiMoRNA использует «gencode basic annotation v31», а RNA-Chrom – «gencode basic annotation v35». Мы использовали три показателя сходства: (1) одинаковые имена генов (совпадение



**Рис. 1.** Пересечение 4145 генов из HiMoRNA с 60619 генами из RNA-Chrom. А – разделение пар генов на шесть групп в зависимости от показателей сходства, которым они удовлетворяют. Красными прямоугольниками выделены те группы, в которых достигнуто однозначное соответствие между генами. Б – диаграмма Венна между группами генов 2, 4, 5 и 6 (суммарное количество пар генов в четырех группах равно 4100)

'gene\_name', рис. 1А); (2) одинаковые идентификаторы генов (совпадение 'gene\_id', рис. 1А); и (3) индекс Жаккара (отношение длины перекрытия генов к длине их объединения) больше 0.99 (Индекс Жаккара > 0.99, рис. 1А). К сожалению, имена/идентификаторы и координаты генов в разных источниках аннотаций не всегда совпадают. Чтобы решить эту проблему, мы пересекли 4145 генов lncRNA из HiMoRNA с 60619 генами из RNA-Chrom по геномным координатам с использованием bedtools (команда intersect), в результате чего получилось 6778 пар генов (из-за того, что гены из HiMoRNA несколько раз пересекались с генами из RNA-Chrom, пар получилось больше 4145). Два гена из HiMoRNA не пересекались ни с одним из генов из RNA-Chrom (ENSG00000267034.1, ENSG00000280076.1). Далее для каждой пары генов был рассчитан индекс Жаккара. Используя описанные выше три показателя сходства, мы разбили 6778 пар генов на шесть групп (рис. 1А). Взяв в качестве главной метрики сходства «Индекс Жаккара > 0.99», мы выделили четыре группы (рис. 1А, группы 2, 4, 5 и 6), между которыми гены не пересекались (рис. 1Б). Таким образом мы определили 4100 однозначных соответствий между генами. Из оставшихся 43 генов из HiMoRNA нам удалось сопоставить 24 гена с генами из RNA-Chrom по метрике совпадение 'gene\_name'. Всего мы получили 4124 гена lncRNA, общих как для HiMoRNA, так и для RNA-Chrom. Таблица соответствия генов lncRNA («lncRNA corre-

spondence table») доступна для скачивания на веб-ресурсе HiMoRNA (дополнительная табл. 1).

Во-вторых, к веб-интерфейсу RNA-Chrom была добавлена новая опция – обработка параметров (локус, название РНК, внутренний для RNA-Chrom идентификатор РНК, организм) из специального вида URL-ссылки (например, [https://rnachrom2.bioinf.fbb.msu.ru/basic\\_graphical\\_summary\\_dna\\_filter?locus=chrX:23456-24253566&name=XIST&naID=227896&organism=Homo+sapiens](https://rnachrom2.bioinf.fbb.msu.ru/basic_graphical_summary_dna_filter?locus=chrX:23456-24253566&name=XIST&naID=227896&organism=Homo+sapiens)) и предоставление информации о контактах запрашиваемой в URL-ссылке lncRNA с хроматином в разных типах экспериментов на новой странице в браузере.

В-третьих, на стороне HiMoRNA были сделаны следующие улучшения:

Для правильной генерации URL-ссылки была добавлена таблица соответствия генов lncRNA между RNA-Chrom и HiMoRNA.

1) На веб-страницу «Страница результатов поиска» была добавлена кнопка «Перейти в RNA-Chrom БД» («Go to RNA-Chrom DB») с выпадающим списком (рис. 2), который позволяет сгенерировать три типа URL-ссылки для перехода на страницу RNA-Chrom:

- а) с контактами данной lncRNA в определенном геномном локусе, расширенном на 1 / 5 / 10 / 25 / 50 / 100 т.п.н.;
- б) со всеми контактами данной lncRNA;
- в) со всеми РНК, которые имеют контакты в определенном геномном локусе.

### Односторонний точный тест Фишера

В большинстве триад обнаруживаются пики гистоновых модификаций как с отрицательной, так и с положительной корреляцией экспрессии lncRNA и уровня сигнала пика (далее «-» и «+» пики соответственно). Обнаружение «+» пика соответствует предположению, что lncRNA участвует в установке модификации гистона, тогда как обнаружение «-» пика соответствует предположению, что lncRNA участвует в удалении модификации гистона.

Чтобы оценить, насколько хорошо предсказания согласуются с опубликованными на данный момент результатами экспериментальных исследований, были отобраны lncRNA и соответствующие им расширенные на +/- 25 т.п.н. «-» и «+» – гистоновые пики (положительно и отрицательно скоррелированные), у которых доля соответствующих пиков, поддерживаемых контактами хотя бы по одной из гистоновых меток, больше 0.4. Далее мы посчитали правосторонний и левосторонний тест Фишера («right-tailed Fisher's exact test» и «left-tailed Fisher's exact test») отдельно для каждой lncRNA и гистоновой метки (например, таблица сопряженности для пары «PVT1–H3K27ac» – *дополнительная табл. 2*).

### Данные Red-ChIP

При рассмотрении lncRNA PVT1 в качестве примера интеграции HiMoRNA и RNA-Chrom мы дополнительно валидировали ее контакты с хроматином с помощью внешних данных Red-ChIP [27], которые доступны в Gene Expression Omnibus под номером GSE174474, образцы GSM5315228 и GSM5315229 (клеточная линия hES). Метод Red-ChIP фиксирует контакты РНК с хроматином, опосредованные белком EZH2, компонентом комплекса PRC2, который устанавливает, в том числе гистоновую модификацию H3K27me3. Первичную обработку этих данных проводили в соответствии с протоколом, использованным в базе данных RNA-Chrom. Далее мы определили участки генома, обогащенные контактами lncRNA PVT1 с хроматином, с помощью программы BaRDIC (--qval\_type all; --qval\_threshold 1) [28]. Таким образом, мы обнаружили 3242 потенциально функциональные геномные области, в которых связывание PVT1 опосредовано белком EZH2.

## РЕЗУЛЬТАТЫ

### Интеграция баз данных

Поскольку HiMoRNA содержит миллионы эпигенетических пиков, целесообразно отобрать для последующего анализа наиболее надежные из них. Чтобы

облегчить эту задачу, мы провели интеграцию пиков HiMoRNA с данными РНК-хроматинового интерактома из RNA-Chrom. Для этого установили взаимно однозначное соответствие между генами двух баз данных и модифицировали веб-интерфейсы (смотри «Экспериментальную часть», подраздел «Интеграция баз данных HiMoRNA и RNA-Chrom»). При таком подходе HiMoRNA может генерировать специальный URL-запрос к 4124 из 4145 lncRNA из RNA-Chrom, что, в частности, позволяет ответить на вопрос о том, с какими другими локусами хроматина контактирует исследуемая РНК. Этот подход позволяет значительно расширить представление о функции конкретной РНК.

Общая схема интеграции представлена на *рис. 2*. Чтобы воспользоваться интеграцией, для начала необходимо найти целевую lncRNA в базе данных HiMoRNA. Находясь на главной странице HiMoRNA, пользователь может скачать саму базу данных, добавленные в рамках интеграции «Таблицу генов» («Gene table») и «Таблицу соответствия длинных некодирующих РНК» («lncRNA correspondence table»), для поиска в них интересующих генов/lncRNA по геномным координатам. Мы предоставили эту опцию, поскольку идентификатор Ensembl или названия lncRNA и ассоциированные с гистоновыми модификациями гены, которые пользователь хочет использовать, могут не совпадать с приведенными в HiMoRNA. На странице поиска пользователю необходимо настроить фильтры под свою задачу, указывая интересующие lncRNA, модификации гистонов, геномные координаты и гены, ассоциированные с выбранной модификацией гистонов.

Попав на страницу с результатами поиска, пользователь может более подробно изучить найденные по запросу предсказания, в частности, перейдя в базу данных RNA-Chrom. Для этого следует выбрать интересующую триаду «lncRNA–пик эпигенетической модификации–ассоциированный с пиком ген» в интерактивной таблице результатов, а затем нажать на кнопку «Перейти в RNA-Chrom БД» («Go to RNA-Chrom DB»). В выпадающем списке необходимо нажать соответствующую кнопку для перехода на страницу: 1) с контактами данной lncRNA в области конкретного пика (надо выбрать, на сколько расширить координаты пика при поиске контактов); 2) со всеми контактами данной lncRNA; 3) со всеми lncRNA, которые имеют контакты в данном локусе. Далее пользователь будет перенаправлен на веб-страницу базы данных RNA-Chrom с графической сводкой lncRNA-хроматинового интерактома, которая позволяет уточнить, опосредована ли функциональная связь «lncRNA–эпигенетическая модификация» из HiMoRNA физическим

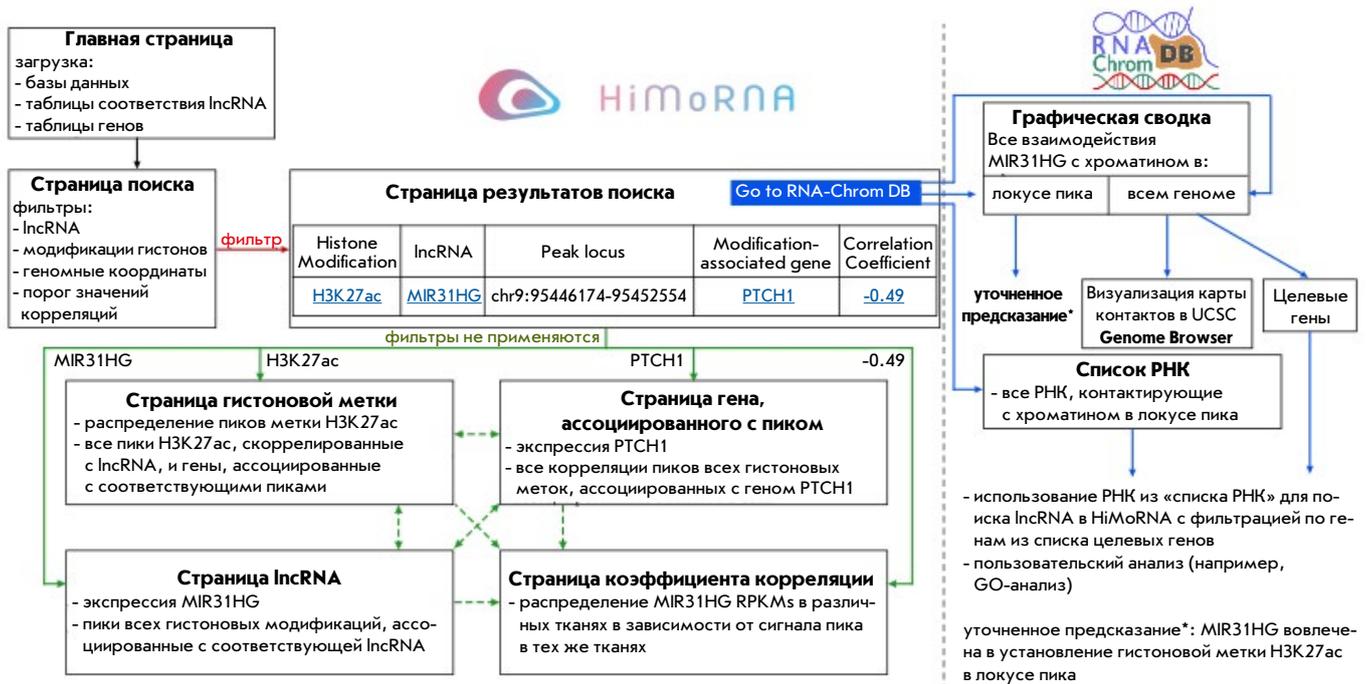


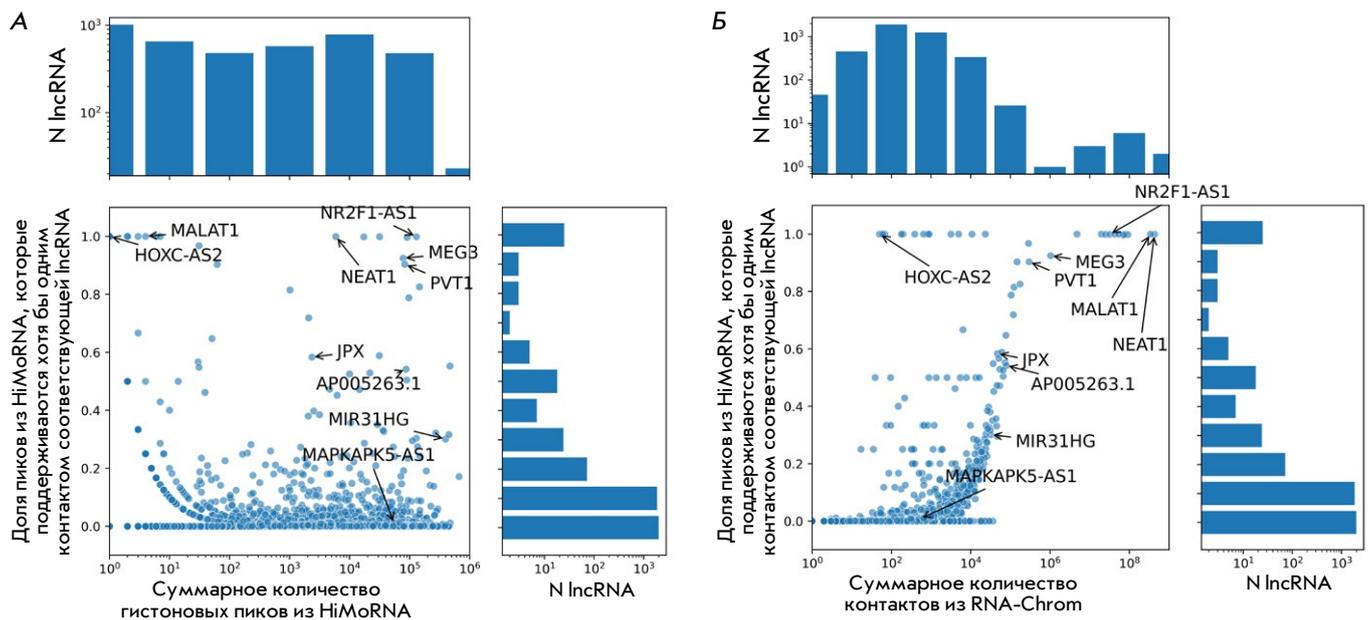
Рис. 2. Сценарий использования баз данных HiMoRNA и RNA-Chrom после интеграции. Прямоугольники представляют веб-страницы, стрелки – переходы между ними

нахождением lncRNA у соответствующего геномного локуса, а также какие еще lncRNA потенциально могут участвовать в регуляции данного локуса. Для визуального анализа контакты всех интересных экспериментов можно загрузить в UCSC Genome Browser (нажать на «VIEW IN GENOME BROWSER»). Выбрав один эксперимент по РНК-хроматиновому интерактому, пользователь может получить список генов, которые располагаются на интересующем участке генома, со статистикой контактируемости lncRNA с ними (нажать на «ALL TARGET GENES»). Этот список генов пользователь может скачать и далее работать с ним, например, выполнив GO-анализ. Примеры использования интеграции баз данных HiMoRNA и RNA-Chrom приведены и подробно рассмотрены в разделе «Варианты использования».

### Согласованность результатов HiMoRNA и RNA-Chrom

Для оценки полноты интеграции проанализировали частоту подтверждения скоррелированных с экспрессией lncRNA гистоновых пиков HiMoRNA данными о контактах соответствующей lncRNA с хроматином. Из 4145 lncRNA, присутствующих в HiMoRNA, 4011 (96.8%) имеют хотя бы один контакт в базе данных RNA-Chrom, 29 РНК не согласуются между базами данных и еще 105 (2.5%) не име-

ют контактов в RNA-Chrom. Среди интересующих нас 4011 lncRNA только 35.5% имеют хотя бы один пик, который поддерживается контактами соответствующей lncRNA. Однако, принимая во внимание, что по построению экспериментальных протоколов реальное взаимодействие lncRNA с хроматином может происходить на удалении от экспериментально фиксируемого контакта, мы предлагаем расширять координаты контакта для более точной оценки соответствия предсказанных пиков HiMoRNA и информации из RNA-Chrom. При расширении контактов на +/- 1, +/- 5, +/- 10, +/- 25 и +/- 50 т.п.н. процент РНК, у которых пики из HiMoRNA подтверждаются хотя бы одним контактом, увеличивается до 38.5, 42.7, 45.7, 50.1 и 53% соответственно. В частности, для lncRNA MALAT1, HOXC-AS2, NEAT1, NR2F1-AS1, PVT1, MEG3 и ряда других доля расширенных на +/- 25 т.п.н. пиков из HiMoRNA, подтвержденных в RNA-Chrom, приближается к 1 (рис. 3). Однако чаще встречаются lncRNA, у которых доля расширенных на +/- 25 т.п.н. и подтвержденных контактами пиков сильно меньше 1 (JPX, AP005263.1, MIR31HG) или приближается к 0 (MAPKAPK5-AS1). Это, по-видимому, связано с тем, что базы данных HiMoRNA и RNA-Chrom содержат неполную информацию о lncRNA из-за строгой фильтрации предсказаний и несовершенства экспериментальных данных взаимодей-



**Рис. 3.** Доли пиков HiMoRNA, которые подтверждаются хотя бы одним контактом соответствующей lncRNA из RNA-Chrom, относительно суммарного количества пиков HiMoRNA для соответствующей lncRNA (А) и суммарного количества контактов для соответствующей lncRNA из RNA-Chrom (Б). Геномные координаты контактов расширены на +/- 25 т.п.н.

ствий РНК-хроматин соответственно. Например, в RNA-Chrom половина рассматриваемых в данной статье lncRNA имеют меньше 200 контактов (рис. 3Б), так как для большинства lncRNA имеются лишь данные «все-против-всех», полученные экспериментальными методами, которые недостаточно полно определяют контакты низко экспрессирующихся РНК.

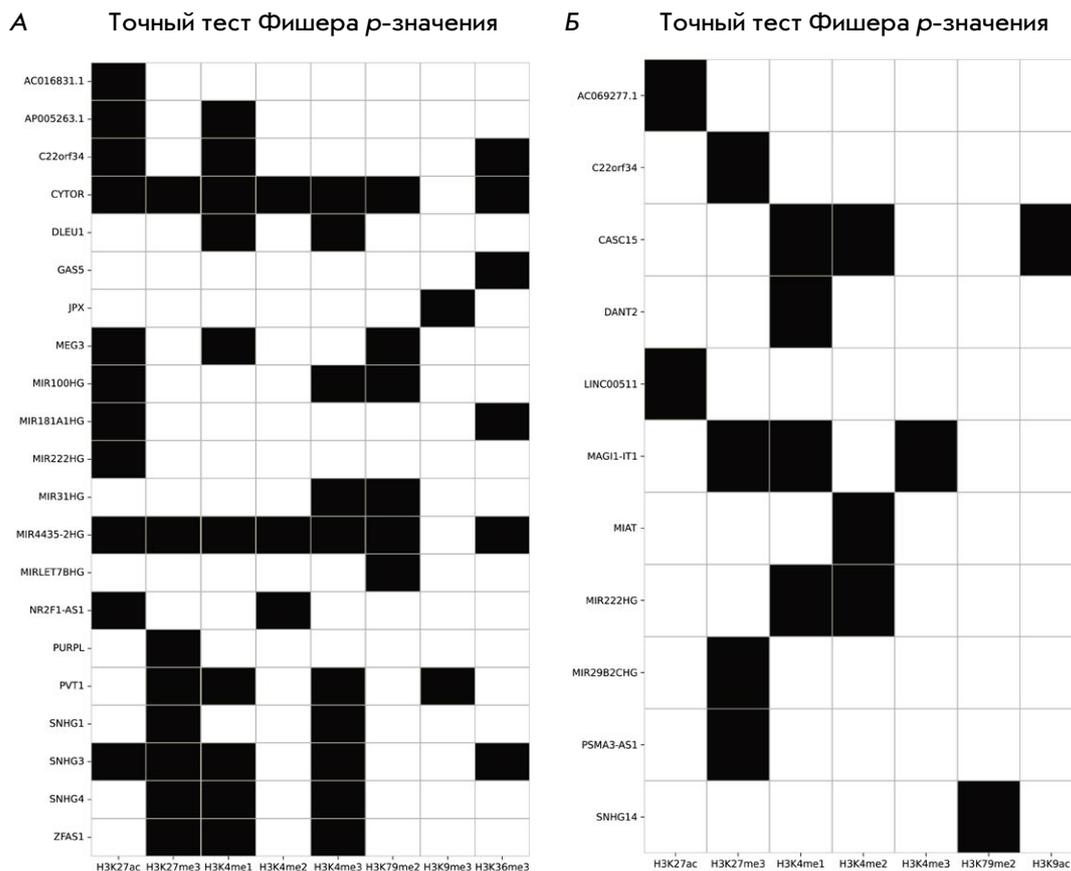
В базе данных HiMoRNA встречаются триады с отрицательной или положительной корреляцией экспрессии lncRNA и сигнала эпигенетического пика («-» и «+» пики соответственно). Для того, чтобы оценить, насколько хорошо предсказания на основе интегрированных баз данных согласуются с опубликованными на данный момент экспериментальными исследованиями, мы отобрали 30 lncRNA и соответствующие им гистоновые пики, для которых статистически значимо (односторонний точный тест Фишера,  $p$ -value < 0.001) преобладают «+» или «-» пики хотя бы одной из гистоновых меток, подтвержденные контактами из RNA-Chrom с расширением на +/- 25 т.п.н. (смотри «Экспериментальную часть», подраздел «Односторонний точный тест Фишера», рис. 4). Отфильтровав результаты по  $p$ -value < 0.001, мы получили следующие пары «lncRNA-гистоновая метка»:

21 lncRNA, у которых «+» пики соответствующих гистоновых меток лучше поддерживаются контак-

тами из RNA-Chrom нежели «-» пики (правосторонний точный тест Фишера,  $p$ -value < 0.001).

11 lncRNA, у которых «-» пики соответствующих гистоновых меток лучше поддерживаются контактами из RNA-Chrom, нежели «+» пики (левосторонний точный тест Фишера,  $p$ -value < 0.001).

Ранее было показано потенциальное участие значительной части выявленных lncRNA в эпигенетической регуляции посредством гистоновых модификаций. Разберем случаи, когда «+» пики статистически значимо лучше поддерживаются контактами из RNA-Chrom, нежели «-» пики. Например, MIR4435-2HG участвует в установлении активаторной метки H3K27ac в энхансерном регионе локуса RPTOR [29]. Наши данные показывают, что у MIR4435-2HG, помимо H3K27ac, вероятно, существуют и другие мишени эпигенетической регуляции посредством установления таких меток, как H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me2 (рис. 4А). Аналогично на основании данных для MIR31HG [30], SNHG1, PVT1 [31–33] и lncRNA lnc-Nr2f1 мыши (предположительно имеющей функциональную консервативность с lncRNA NR2F1-AS1 человека) [34] мы определили согласующиеся с этими данными гистоновые модификации: NR2F1-AS1 – H3K27ac, MIR31HG – H3K4me3, SNHG1 – H3K27me3, PVT1 – H3K27me3. Помимо этого, мы выявили функциональную связь



**Рис. 4.** Тепловая карта с результатами точного теста Фишера («Fisher exact test») для пар «lncRNA – расширенные на +/- 25 т.п.н. пики гистоновой метки». Черный цвет означает, что доля «-» / «+» гистоновых пиков, поддерживаемых контактами соответствующей lncRNA, больше 0.4 и  $p$ -value точного теста Фишера меньше  $10^{-3}$ , иначе – белый цвет. А – правосторонний тест Фишера («right-tailed»): «+» пики соответствующих гистоновых меток лучше поддерживаются контактами из RNA-Chrom, нежели «-» пики. Б – левосторонний тест Фишера («left-tailed»): «-» пики соответствующих гистоновых меток лучше поддерживаются контактами из RNA-Chrom, нежели «+» пики

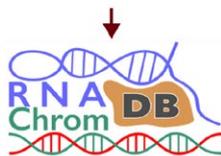
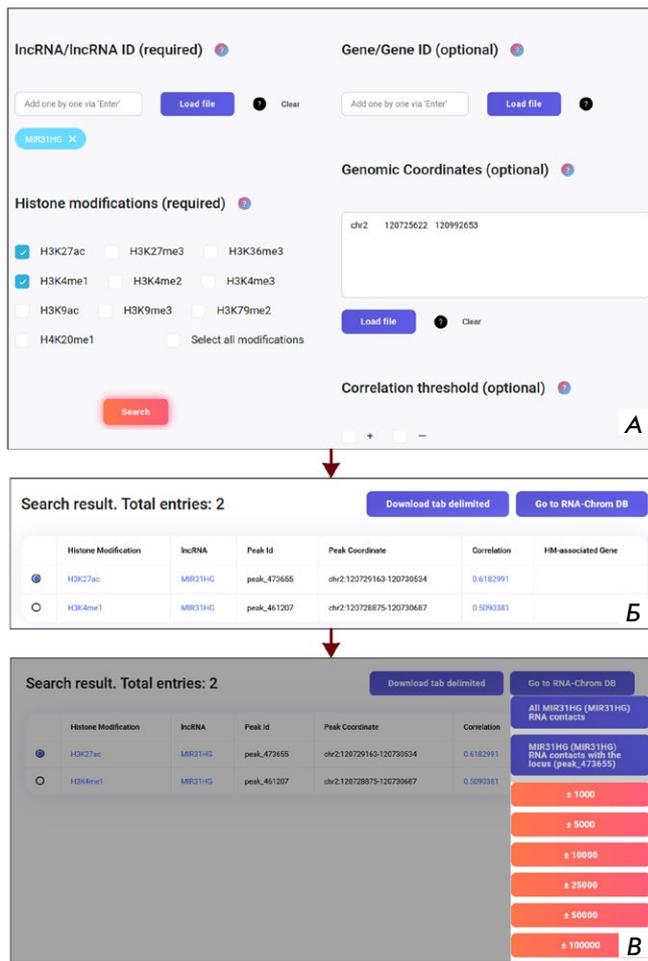
этих lncRNA с другими эпигенетическими метками: NR2F1-AS1 – H3K4me2, MIR31HG – H3K79me2, SNHG1 – H3K4me3, PVT1 – H3K4me1, H3K4me3, H3K9me3 (рис. 4А).

У ряда lncRNA (ZFAS1, SNHG4, SNHG1, SNHG3, PVT1, MIR4435-HG, CYTOR) выявлено большое количество подтвержденных контактами «+» пиков H3K27me3 и H3K4me3, которые статистически значимо больше поддерживаются контактами из RNA-Chrom, нежели «-» пики (рис. 4А), отвечающие за противоположные состояния хроматина. По аналогии с существующими lncRNA, устанавливающими обе этих гистоновых метки в зависимости от ассоциации с различными эффекторными белками (как, например, нРНК SRA [35], ANRIL [36]), можно предположить, что для них тоже характерны более сложные механизмы регуляции активности хроматина.

Случаи, когда «-» пики статистически значимо больше поддерживаются контактами из RNA-Chrom нежели «+» пики, вероятно, можно объяснить тем, что соответствующие lncRNA регулируют удаление гистоновых меток посредством привлечения деметилаз и деацетилаз к соответствующим геномным локусам (рис. 4Б). Мы не можем оценить качество нашего предсказания для этих lncRNA, поскольку экспериментальные статьи, показывающие такого рода регуляцию этих lncRNA, не найдены. Мы предполагаем, что полученные в данном разделе пары «lncRNA-гистоновая метка» (рис. 4), являются потенциальными объектами дальнейших исследований.

**Варианты использования**

Основная цель интеграции HiMoRNA и RNA-Chrom – уточнить функциональную связь внутри

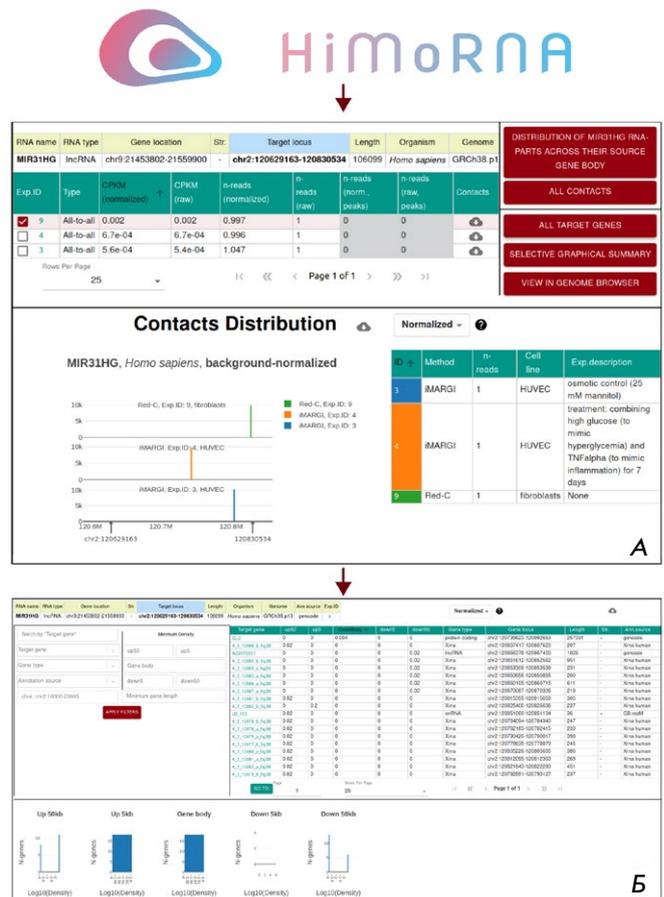


**Рис. 5.** Вариант использования интеграции баз данных HiMoRNA и RNA-Chrom на примере lncRNA MIR31HG. А – создание запроса в HiMoRNA на MIR31HG, гистоновые модификации H3K4me1 и H3K27ac, гены *GLI2* и *FABP4*. Б – таблица с результатами поиска. В – переход в RNA-Chrom

триад «lncRNA–пик эпигенетической модификации–ассоциированный с пиком ген» с помощью данных о локализации соответствующей lncRNA в геномной области вблизи пиков конкретной модификации гистонов. Далее мы приведем примеры пользовательского исследования нескольких lncRNA, механизм действия которых известен.

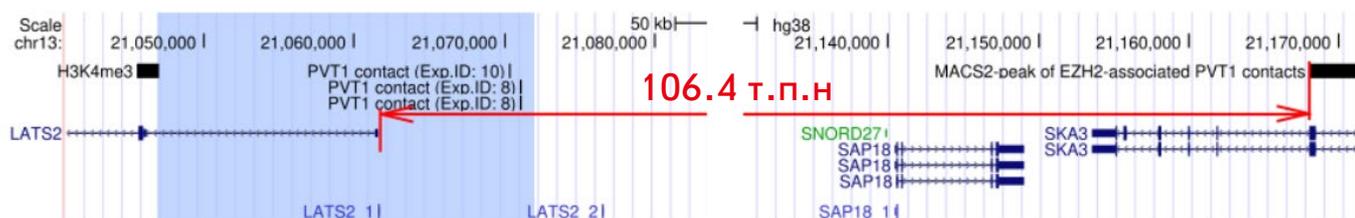
### lncRNA MIR31HG

Длинная некодирующая РНК MIR31HG является известным регулятором гистоновых меток H3K1me1,



**Рис. 6.** Вариант использования интеграции баз данных HiMoRNA и RNA-Chrom на примере lncRNA MIR31HG. А – страница RNA-Chrom с контактами MIR31HG с хроматином в области расширенного пика HiMoRNA. Б – таблица со всеми генами из области расширенного пика, с которыми взаимодействует или нет MIR31HG (эксперимент Exp.ID: 9)

НЗК4me3 и H3K27ac. Ранее сообщалось о снижении уровней H3K4me1 и H3K27ac в энхансерной области гена *GLI2* и H3K4me3 и H3K27ac в промоторной области гена *FABP4* после нокдауна MIR31HG [30, 37]. Это наблюдение можно проверить, используя нашу интеграцию HiMoRNA и RNA-Chrom. Для этого мы создали запрос в HiMoRNA: lncRNA MIR31HG, метки гистонов H3K4me1 и H3K27ac, координаты двух выбранных генов указаны с увеличенной на 10 т.п.н. промоторной областью в поле геномных координат (рис. 5А). В результате веб-ресурс HiMoRNA сгенерировал таблицу с пиками H3K27ac и H3K4me1, которые коррелируют с экспрессией MIR31HG в различных тканях (рис. 5Б). Затем мы выбрали триаду с пиком H3K27ac и перешли на страницу RNA-Chrom с эксперименталь-



**Рис. 7.** Представление в UCSC Genome Browser в области гена *LATS2* и его промоторной окрестности пика H3K4me3, скоррелированного с экспрессией lncRNA PVT1, контакты lncRNA PVT1 из двух экспериментов (RNA-Chrom Exp.ID: 8, 10) и пик EZH2-опосредованных контактов PVT1. Синяя область отражает расширение координат пика H3K4me3 на 25 т.п.н., в пределах которых были отобраны контакты из RNA-Chrom

но обнаруженными контактами MIR31HG с хроматином в области выбранного пика (при нажатии на «Go to RNA-Chrom DB», *рис. 5B*). Выбрав РНК-хроматиновый эксперимент в верхней таблице и нажав на «All target genes» (*рис. 6A*), получили таблицу, в которой, в частности, отражено взаимодействие MIR31HG с геном *GLI2* (*рис. 6B*). Пошаговый анализ представлен в *дополнительной табл. 3*.

Чтобы выяснить, может ли интеграция HiMoRNA и RNA-Chrom дать новую биологическую информацию о функциональной роли lncRNA, мы предположили, что MIR31HG может регулировать не только *GLI2*, но и другие гены, принадлежащие регуляторному пути «Sonic hedgehog» (KEGG:04340). С этой целью идентифицировали соответствующие гены с помощью базы данных KEGG Pathway [38], после чего был сформирован новый запрос в HiMoRNA: lncRNA MIR31HG, гистоновые метки H3K4me1 и H3K27ac, названия 56 генов из пути «Hedgehog signaling pathway» (*дополнительная табл. 4*). В результате мы получили таблицу из 162 триад, которые можно проверить с помощью ресурса RNA-Chrom. Например, в локусе пика H3K27ac\_963553 (chr9:95446174-95452554) MIR31HG контактирует с геном *PTCH1*, кодирующим рецептор «Sonic hedgehog». Чтобы оценить, насколько существенно список генов, ассоциированных со скоррелированными с экспрессией MIR31HG пиками H3K27ac и H3K4me1, обогащен генами из регуляторного пути «Hedgehog signaling pathway», был проведен анализ KEGG обогащений при помощи веб-ресурса «g:Profiler» [39]. В качестве запроса подавались отобранные для MIR31HG и H3K27ac/H3K4me1 гены, а в качестве бэкграунда использовались все остальные ассоциированные с пиками HiMoRNA гены. По результатам нашего анализа, гены, принадлежащие пути «Hedgehog signaling pathway», обогащены пиками H3K27ac ( $p\text{-value} = 2.090 \times 10^{-2}$ ), но не H3K4me1. Это наблюдение позволяет предположить участие MIR31HG в регуляции пути «Hedgehog signaling pathway» через

установление гистоновой модификации H3K27ac в соответствующих геномных локусах.

### lncRNA PVT1

lncRNA PVT1 ингибирует экспрессию гена *LATS2* в клетках немелкоклеточного рака легкого путем рекрутирования EZH2 (субъединица комплекса PRC2) на соответствующий промотор [40]. Мы выполнили поиск триад в HiMoRNA: lncRNA PVT1, все гистоновые модификации, ген *LATS2*. В результате получили пики только для активирующей метки H3K4me3, которые отрицательно коррелировали с экспрессией PVT1, что косвенно согласуется с опубликованными данными [39], поскольку PVT1 привлекает EZH2 и участвует в установлении репрессивной метки H3K27me3. В RNA-Chrom мы наблюдали контакты вокруг одного из пиков H3K4me3 (peak\_169403, chr13:21045571-21046978) в двух экспериментах (клеточные линии K562 и MDA-MB-231). Визуализация контактов PVT1 в Genome Browser [41] подтверждает наличие этого пика в промоторной области гена *LATS2* (*рис. 7*, пошаговый анализ представлен в *дополнительной табл. 5*). Дополнительное подтверждение регуляции *LATS2* с помощью lncRNA PVT1 было получено на основе данных Red-ChIP (смотри раздел «Экспериментальная часть», подраздел «Данные Red-ChIP»): обнаружен пик EZH2-опосредованных контактов PVT1 (chr13:21168000-21224000,  $q\text{-value} = 0.09$ ) в 106.4 т.п.н. от 5'-конца гена *LATS2* (*рис. 7*).

Отсутствие в HiMoRNA триад «lncRNA PVT1–пик H3K27me3–ген *LATS2*» с положительной корреляцией, по-видимому, связано с чрезмерно строгой фильтрацией пиков метки H3K27me3 при создании базы данных. Приведенные выше примеры из раздела «Варианты использования» подтверждают, что интеграцию можно успешно применять для генерации гипотез о роли lncRNA в эпигенетической регуляции конкретных генов для дальнейшей экспериментальной проверки.

## ОБСУЖДЕНИЕ

### Перспективы использования и ограничения подхода

База данных HiMoRNA представляет собой коллекцию геномных локусов, в которых сигнал модификации гистонов значимо коррелирует с экспрессией lncRNA во многих типах клеток и тканей. В настоящее время она содержит более 5 млн корреляций для 10 типов модификаций гистонов и 4145 lncRNA. Предполагается, что некоторые из этих корреляций могут быть ложноположительными или соответствовать случаям непрямого регуляции, а следовательно, требуется дополнительная верификация с помощью внешних данных. RNA-Chrom – это база данных, содержащая полногеномные данные о взаимодействиях РНК с хроматином. К сожалению, эти данные почти не содержат контакты низкоэкспрессирующихся нкРНК и перегружены контактами новосинтезированных транскриптов. Помимо этого, сами по себе они не позволяют генерировать гипотезы, объясняющие функциональную роль взаимодействия. Имея это в виду, разумно интегрировать данные, содержащиеся в HiMoRNA и RNA-Chrom, чтобы охарактеризовать влияние lncRNA на эпигенетические модификации и экспрессию генов.

Как упоминалось выше, в базе данных HiMoRNA не найдены пики H3K27ac и H3K4me1 в локусе гена *FABP4*. Это наблюдение противоречит экспериментальным данным и базе данных RNA-Chrom. Есть и другие отрицательные примеры как в HiMoRNA, так и в RNA-Chrom для известных lncRNA, выполняющих функции в эпигенетической регуляции и поддержании структуры хроматина. Например, в HiMoRNA не наблюдались пики H3K27me3, коррелирующие с *MEG3*, хотя сообщалось, что *MEG3* является регулятором комплекса PRC2 и участвует в поддержании H3K27me3, особенно в промоторных областях генов *SMAD2*, *TGFB2* и *TGFBFR1* [11]. Отсутствие таких пиков можно объяснить тем, что большинство lncRNA экспрессируются клеточно-специфичным образом, а указанный механизм был обнаружен в типе клеток, который не представлен в HiMoRNA. Даже когда данные ChIP-seq доступны, стандартная процедура поиска пиков может оказаться слишком строгой и привести к потере биологически значимых взаимодействий.

Кроме того, мы сталкивались с ситуациями, когда предсказание базы данных HiMoRNA подтверждалось опубликованными данными, но мы не обнаруживали контактов РНК с хроматином в соответствующем геномном локусе в RNA-Chrom. Например, для lncRNA MAPKAPK5-AS1 большинство скоррелированных пиков из HiMoRNA не поддерживают-

ся контактами соответствующей РНК с хроматином из RNA-Chrom. Это, по-видимому, связано с низкой экспрессией данной РНК, что приводит к малому количеству наблюдаемых контактов этой РНК с хроматином в экспериментах «все-против-всех». Мы полагаем, что подобные случаи также могут быть связаны с несовпадением типов клеток в двух базах данных. К сожалению, эта ситуация является следствием малого количества доступных данных такого типа.

Поскольку экспериментальные данные недостаточно полные, обе базы данных не содержат исчерпывающую информацию. Поэтому некоторые известные биологические примеры могли быть пропущены при интеграции. Тем не менее мы полагаем, что интеграция баз данных уменьшает различные систематические ошибки благодаря мультиомиксной природе объединенных данных, тем самым позволяя генерировать интерпретируемые гипотезы о механизмах эпигенетической регуляции экспрессии генов длинными некодирующими РНК.

### БУДУЩЕЕ РАЗВИТИЕ

В дополнение к реализованной интеграции полезным будет привлечение дополнительных полногеномных данных и аннотаций. Они могут включать информацию о трехмерной структуре хроматина, экспрессии и коэкспрессии генов, включая экспрессию lncRNA, целевых генов триады и генов, участвующих в установлении/удалении гистоновых меток, а также о локализации ДНК-связывающих и модифицирующих хроматин белков. В связи с тем, что экспериментальные данные такого рода пока довольно редки, стоит рассмотреть возможность использования биоинформатических предсказаний. Одним из возможных направлений может быть включение результатов предсказания типа взаимодействия lncRNA и хроматина (подробнее о сравнении программ, определяющих механизм взаимодействия lncRNA и других молекул, см., например, [42]). Предсказания типа взаимодействия lncRNA с целевыми молекулами доказали свою эффективность в комплексе с мультиомиксными экспериментальными данными для генерации гипотез о роли конкретной lncRNA [43–45]. Еще одним важным направлением может стать включение информации об изменении экспрессии генов в ответ на искусственное изменение концентрации конкретной lncRNA в клетке [8] и экспериментально подтвержденной информации об участии конкретной lncRNA в регуляции конкретной гистоновой модификации [46]. Это создаст дополнительную стадию валидации результатов интеграции HiMoRNA и RNA-Chrom. Помимо этого, полезным с практи-

ческой точки зрения может оказаться дополнение интеграции оценкой статистической значимости колокализации пиков HiMoRNA и контактов RNA-Chrom для конкретной lncRNA с помощью специальных программных средств: Genometricorr [47], StereoGene [48], RegioneR [49].

Область функциональной аннотации lncRNA – быстро развивающаяся область исследований. Мы будем поддерживать интеграцию HiMoRNA и RNA-Chrom, поскольку каждая база данных будет расширять список доступных организмов и пополняться новыми данными. По мере появления новых экспериментально подтвержденных данных мы планируем построить несколько прогностических моделей «lncRNA–эпигенетические модификации гистонов–ассоциированный с модификацией ген». Мы уверены, что дальнейшее совместное развитие веб-ресурсов HiMoRNA и RNA-Chrom будут способствовать более глубокому пониманию функциональной роли lncRNA в эпигенетической регуляции генов. ●

*Благодарности*

*Авторы выражают благодарность А. Никольской за предоставление обработанных данных Red-ChIP, а также анонимным рецензентам за ценные предложения.*

*Эта работа поддержана грантом РНФ № 23-14-00371 (рук. Ю.А. Медведева).*

*Вклад авторов*

*Улучшение веб-интерфейса и базы данных HiMoRNA, разработка программного обеспечения, И.И.; адаптация веб-ресурса RNA-Chrom для интеграции с HiMoRNA, соответствие названий генов, администрирование проекта, Г.Р.; варианты использования, Г.Р. и Д.М.; общий надзор, А.М. и Ю.М.; подготовка рукописи, И.И., Г.Р., Д.М., А.М. и Ю.М. Все авторы прочитали и согласились с опубликованной версией рукописи.*

*Приложения доступны на сайте  
<https://doi.org/10.32607/actanaturae.27543>*

**СПИСОК ЛИТЕРАТУРЫ**

- Carninci P., Sandelin A., Lenhard B., Katayama S., Shimokawa K., Ponjavic J., Semple C.A.M., Salmena L., Nishida M., Hayashizaki Y., et al. // *Science*. 2005. V. 309. № 5740. P. 1559–1563. doi: 10.1126/science.1112014
- Hon C.-C., Ramilowski J.A., Harshbarger J., Bertin N., Rackham O.J.L., Garmire L.X., Forrest A.R.R., Carninci P., Kawaji H., Hayashizaki Y., et al. // *Nature*. 2017. V. 543. № 7644. P. 199–204. doi: 10.1038/nature21374
- Cabili M.N., Trapnell C., Goff L., Koziol M., Tazon-Vega B., Regev A., Rinn J.L. // *Genes Dev*. 2011. V. 25. № 18. P. 1915–1927. doi: 10.1101/gad.17446611
- Andersson R., Gebhard C., Miguel-Escalada I., Hoof I., Bornholdt J., Boyd M., Chen Y., Zhao X., Schmidl C., Suzuki T., et al. // *Nat. Commun*. 2014. V. 5. № 1. P. 5336. doi: 10.1038/ncomms6336
- Ulitsky I., Shkumatava A., Jan C.H., Sive H., Bartel D.P. // *Cell*. 2011. V. 147. № 7. P. 1537–1550. doi: 10.1016/j.cell.2011.11.055
- Quinn J.J., Ilik I.A., Qu K., Georgiev P., Chu C., Akhtar A., Chang H.Y. // *Genes Dev*. 2016. V. 30. № 2. P. 191–207. doi: 10.1101/gad.272187.115
- Alam T., Medvedeva Y.A., Jia H., Brown J.B., Lipovich L., Baillie J.K. // *PLoS One*. 2014. V. 9. № 10. P. e109443. doi: 10.1371/journal.pone.0109443
- Ramilowski J.A., Yip C.W., Agrawal S., Chang J.-C., Ciani Y., Kulakovskiy I.V., Mendez M., Ooi J.L.C., Ouyang J.F., Nguyen A., et al. // *Genome Res*. 2020. V. 30. № 7. P. 1060–1072. doi: 10.1101/gr.254516.119
- Khalil A.M., Guttman M., Huarte M., Garber M., Raj A., Rivea Morales D., Thomas K., Presser A., Bernstein B.E., van Oudenaarden A., et al. // *Proc. Natl. Acad. Sci. USA*. 2009. V. 106. № 28. P. 11667–11672. doi: 10.1073/pnas.0904715106
- Grote P., Herrmann B.G. // *RNA Biol*. 2013. V. 10. № 10. P. 1579–1585. doi: 10.4161/rna.26165
- Mondal T., Subhash S., Vaid R., Enroth S., Uday S., Reinius B., Mitra S., Mohammed A., James A.R., Hoberg E., et al. // *Nat. Commun*. 2015. V. 6. № 1. P. 7743. doi: 10.1038/ncomms8743
- Goff L.A., Rinn J.L. // *Genome Res*. 2015. V. 25. № 10. P. 1456–1465. doi: 10.1101/gr.191122.115
- Mazurov E., Szykh A., Medvedeva Y.A. // *Non-Coding RNA*. 2022. V. 8. № 1. P. 18. doi: 10.3390/ncrna8010018
- Engreitz J.M., Pandya-Jones A., McDonel P., Shishkin A., Sirokman K., Surka C., Kadri S., Xing J., Goren A., Lander E.S., et al. // *Science*. 2013. V. 341. № 6147. P. 1237973. doi: 10.1126/science.1237973
- Simon M.D., Wang C.I., Kharchenko P.V., West J.A., Chapman B.A., Alekseyenko A.A., Borowsky M.L., Kuroda M.I., Kingston R.E. // *Proc. Natl. Acad. Sci. USA*. 2011. V. 108. № 51. P. 20497–20502. doi: 10.1073/pnas.1113536108
- Chu C., Qu K., Zhong F.L., Artandi S.E., Chang H.Y. // *Mol. Cell*. 2011. V. 44. № 4. P. 667–678. doi: 10.1016/j.molcel.2011.08.027
- Quinn J.J., Qu K., Chang H.Y. // *Nat. Biotechnol*. 2014. V. 32. № 9. P. 933–940. doi: 10.1038/nbt.2943
- Chu H.-P., Cifuentes-Rojas C., Kesner B., Aeby E., Lee H.-G., Wei C., Oh H.J., Boukhali M., Haas W., Lee J.T., et al. // *Cell*. 2017. V. 170. № 1. P. 86–101. doi: 10.1016/j.cell.2017.06.017
- Sridhar B., Rivas-Astroza M., Nguyen T.C., Chen W., Yan Z., Cao X., Hebert L., Zhong S. // *Curr. Biol*. 2017. V. 27. № 4. P. 602–609. doi: 10.1016/j.cub.2017.01.011
- Li X., Zhou B., Chen L., Gou L.-T., Li H., Fu X.-D. // *Nat. Biotechnol*. 2017. V. 35. № 10. P. 940–950. doi: 10.1038/nbt.3968
- Bell J.C., Jukam D., Teran N.A., Risca V.I., Smith O.K., Johnson W.L., Skotheim J.M., Greenleaf W.J., Straight A.F. // *eLife*. 2018. V. 7. P. e27024. doi: 10.7554/eLife.27024
- Yan Z., Huang N., Wu W., Chen W., Jiang Y., Zhang J., Zhang X., Zhang Y., Zhang Q., Zhang L., et al. // *Proc. Natl. Acad. Sci. USA*. 2019. V. 116. № 8. P. 3328–3337. doi: 10.1073/pnas.1819788116
- Bonetti A., Agostini F., Suzuki A.M., Hashimoto K., Pascarella G., Gimenez J., Roos L., Nash A.J., Ghazanfar S., Carninci P., et al. // *Nat. Commun*. 2020. V. 11. № 1. P. 1018.

- doi: 10.1038/s41467-020-14337-6
24. Gavrilov A.A., Zharikova A.A., Galitsyna A.A., Luzhin A.V., Rubanova N.M., Golov A.K., Petrova N.V., Kantidze O.L., Ulianov S.V., Misteli T., et al. // *Nucl. Acids Res.* 2020. V. 48. № 12. P. 6699–6714. doi: 10.1093/nar/gkaa456
  25. Ryabykh G.K., Zharikova A.A., Galitsyna A.A., Ulianov S.V., Razin S.V., Gavrilov A.A. // *Mol. Biol.* 2022. V. 56. № 2. P. 210–228. doi: 10.1134/S002689332202017X
  26. Ryabykh G.K., Dikstein N., Zharikova A.A., Galitsyna A.A., Ulianov S.V., Razin S.V., Gavrilov A.A. // *Database J. Biol. Databases Curation.* 2023. V. 2023. P. baad025. doi: 10.1093/database/baad025
  27. Gavrilov A.A., Golov A.K., Luzhin A.V., Zharikova A.A., Galitsyna A.A., Rubanova N.M., Kantidze O.L., Ulianov S.V., Razin S.V., Misteli T., et al. // *Proc. Natl. Acad. Sci. USA.* 2022. V. 119. № 1. P. e2116222119. doi: 10.1073/pnas.2116222119
  28. Mylarshchikov D.E., Baumgart S.J., Mazurov E., Saveliev A., Medvedeva Y.A. // *NAR Genomics Bioinforma.* 2024. V. 6. № 2. P. lqae054. doi: 10.1093/nargab/lqae054
  29. Hartana C.A., Rassadkina Y., Gao C., Yashiro-Ohtani Y., Mercier F., Deng S., Li X., Lin 7. R.H., Das S., Lian C., et al. // *J. Clin. Invest.* 2021. V. 131. № 9. P. e146136. doi: 10.1172/JCI146136
  30. Chen W., Zhang J., Xu H., Dai X., Zhang X., Wang W., Wu Y., Li Y., Wang X., Zhang Y., et al. // *Oncogene.* 2024. V. 43. № 18. P. 1328–1340. doi: 10.1038/s41388-024-02986-2
  31. Li B., Jiang Q., Liu X., Yang X., Li Z., Li J., Zhang Y., Wu Y., Li X., Zhang Y., et al. // *Cell Death Dis.* 2020. V. 11. № 10. P. 823. doi: 10.1038/s41419-020-03031-6
  32. Li Z., Guo X., Wu S. // *Stem Cell Res. Ther.* 2020. V. 11. № 1. P. 435. doi: 10.1186/s13287-020-01953-8
  33. Nylund P., Gimenez G., Laine I., Massinen S., Gordon S., Kuusela M., Kontturi S., Rusanen A., Vettenranta K., Lohi O., et al. // *Haematologica.* 2024. V. 109. № 2. P. 567–577. doi: 10.3324/haematol.2022.282375
  34. Ang C.E., Trevino A.E., Chang S., Soeung M., Ma L., Chidambaram S., Young C., Wernig M., Sudhof T.C., Chang H.Y., et al. // *eLife.* 2019. V. 8. P. e41770. doi: 10.7554/eLife.41770
  35. Wongtrakoongate P., Riddick G., Hashem O., Harris R., Jones S., Ramjaun A., Bunjobpol W., Felsenfeld G., Turner B.M., Schwabe J.W.R., et al. // *PLoS Genet.* 2015. V. 11. № 10. P. e1005615. doi: 10.1371/journal.pgen.1005615
  36. Alfeghaly C., Bozec D., Forterre M., Muller Q., Bernard D., Munch M., Schulte M.L., Nollet M., Paul N., Regnault B., et al. // *Nucl. Acids Res.* 2021. V. 49. № 9. P. 4954–4970. doi: 10.1093/nar/gkab267
  37. Huang Y., Zheng Y., Jia L., Li W. // *Sci. Rep.* 2017. V. 7. № 1. P. 8080. doi: 10.1038/s41598-017-08131-6
  38. Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. // *Nucl. Acids Res.* 2016. V. 44. № D1. P. D457–D462. doi: 10.1093/nar/gkv1070
  39. Kolberg L., Raudvere U., Kuzmin I., Vilo J., Peterson H. // *Nucl. Acids Res.* 2023. V. 51. № W1. P. W207–W212. doi: 10.1093/nar/gkad347
  40. Wan L., Zhang L., Fan K., Cheng Z.X., Sun Q.C., Wang J.J. // *Mol. Cancer Ther.* 2016. V. 15. № 5. P. 1082–1094. doi: 10.1158/1535-7163.MCT-15-0509
  41. Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., Haussler D. // *Genome Res.* 2002. V. 12. № 6. P. 996–1006. doi: 10.1101/gr.229102
  42. Antonov I.V., Mazurov E., Borodovsky M., Medvedeva Y.A. // *Brief. Bioinform.* 2019. V. 20. № 2. P. 551–564. doi: 10.1093/bib/bbx132
  43. Antonov I., Medvedeva Y.A. // *Genes.* 2020. V. 11. № 12. P. E1483. doi: 10.3390/genes11121483
  44. Matveishina E., Antonov I., Medvedeva Y.A. // *Int. J. Mol. Sci.* 2020. V. 21. № 3. P. 830. doi: 10.3390/ijms21030830
  45. Ogunleye A.J., Romanova E., Medvedeva Y.A. // *F1000Research.* 2021. V. 10. P. 204. doi: 10.12688/f1000research.51844.1
  46. Marakulina D., Vorontsova Y., Mazurov E., Anisenko A., Korchagina A., Medvedeva Y.A., Kolmykov S., Fishman V., Kulakovskiy I.V., Meshcheryakov R., et al. // *Nucl. Acids Res.* 2023. V. 51. № D1. P. D564–D570. doi: 10.1093/nar/gkac809
  47. Favorov A., Mularoni L., Cope L.M., Medvedeva Y.A., Mironov A.A., Makeev V.J., Wheelan S.J. // *PLoS Comput. Biol.* 2012. V. 8. № 5. P. e1002529. doi: 10.1371/journal.pcbi.1002529
  48. Stavrovskaya E.D., Tepluk N., Mironov A.A., Panchenko A.R., Favorov A.V., Makeev V.J. // *Bioinforma. Oxf. Engl.* 2017. V. 33. № 20. P. 3158–3165. doi: 10.1093/bioinformatics/btx379
  49. Gel B., Díez-Villanueva A., Serra E., Buschbeck M., Peinado M.A., Malinverni R. // *Bioinforma. Oxf. Engl.* 2016. V. 32. № 2. P. 289–291. doi: 10.1093/bioinformatics/btv562