

УДК 577.21

# RNAcontacts, вычислительный конвейер для предсказания РНК-РНК-контактов по данным конформационного секвенирования РНК *in situ*

С. Д. Маргасюк<sup>1</sup>, М. А. Власенок<sup>1</sup>, Г. Ли<sup>2</sup>, Ч. Као<sup>3</sup>, Д. Д. Первушин<sup>1\*</sup><sup>1</sup>Сколковский институт науки и технологий, Москва, 121205 Россия<sup>2</sup>Колледж наук о жизни, Чжэцзянский университет, Ханчжоу, Чжэцзян, ZJ310058 КНР<sup>3</sup>Ключевая лаборатория биологии РНК, Институт биофизики Китайской академии наук, Пекин, 100101 КНР

\*E-mail: d.pervouchine@skoltech.ru

Поступила в редакцию 22.12.2022

Принята к печати 20.02.2023

DOI: 10.32607/actanaturae.11893

**РЕФЕРАТ** Конформационное секвенирование (КС) РНК – группа молекулярных методов высокопроизводительного анализа пространственной сближенности РНК в живых клетках. Эти методы основаны на шивке, фрагментации и последующем лигировании пространственно сближенных молекул. Чтения, получаемые в результате высокопроизводительного секвенирования, содержат два разных типа разрывов, один из которых образуется в результате сплайсинга, а второй – в результате лигирования. Нами разработан RNAcontacts – универсальный вычислительный конвейер для обнаружения контактов РНК-РНК в данных КС РНК. RNAcontacts решает задачу картирования последовательностей с двумя различными типами разрывов с помощью двухпроходного выравнивания. На первом проходе из контрольного эксперимента секвенирования РНК определяют экзон-экзонные соединения, которые затем передают картировщику на втором проходе как уже известные интроны. Этот подход позволяет с большей чувствительностью обнаруживать контакты РНК и обладает более высокой специфичностью в отношении интронов, присутствующих в биологическом образце, по сравнению с уже существующими методами. RNAcontacts в автоматическом режиме извлекает контакты, кластеризует точки лигирования, вычисляет поддержку чтением и визуализирует результаты. Он реализован в воспроизводимой и масштабируемой системе управления рабочими процессами Snakemake, которая позволяет быстро и единообразно обрабатывать сразу несколько наборов данных. Разработанный вычислительный конвейер применим к любым данным КС, если одним из взаимодействующих партнеров является РНК. RNAcontacts доступен через репозиторий github по адресу <https://github.com/smargasyuk/RNAcontacts/>

**КЛЮЧЕВЫЕ СЛОВА** РНК, лигирование, сплайсинг, конформационное секвенирование.

**СПИСОК СОКРАЩЕНИЙ** КС – конформационное секвенирование; RNA-seq – секвенирование РНК (RNA sequencing); ЛПБМ – лигирование пространственно близких молекул.

## ВВЕДЕНИЕ

Быстрое развитие технологий высокопроизводительного секвенирования позволило выявлять пространственные контакты между нуклеиновыми кислотами, включая контакты ДНК в трехмерной структуре хроматина [1–3], функциональные взаимодействия энхансеров с промоторами [4, 5], а также хроматин-ассоциированные контакты РНК с ДНК [6, 7]. Эти методы основаны на принципе лигирова-

ния пространственно сближенных молекул (ЛПБМ, proximity ligation assays) – разрезании и последующем стохастическом лигировании нуклеиновых кислот, входящих в состав макромолекулярных комплексов, которое преимущественно происходит между пространственно близкими участками. Глубокое секвенирование полученных химерных фрагментов дает сотни миллионов чтений, содержащих последовательности взаимодействующих локусов.

В некоторых недавно разработанных методах для изучения взаимодействий РНК-РНК *in vivo* и *in vitro* применяется ЛПБМ [8, 9]. В некоторых из этих методов, таких, как PARIS [10], LIGR-seq [11], SPLASH [12] и COMRADES [13], используются производные псоралена для того, чтобы вызвать обратимое сшивание между дуплексами РНК для определения комплементарных взаимодействий. В протоколе конформационного секвенирования РНК *in situ* (RIC-seq) цепи РНК сшиваются через РНК-связывающие белки (РСБ) [14], что позволяет не только устанавливать вторичные и третичные структуры РНК, но и помогает воссоздать трехмерные карты взаимодействий РНК с РСБ. Во всех этих случаях взаимодействия закодированы в химерных последовательностях РНК, полученных путем разрезания и последующего лигирования.

В отличие от взаимодействий ДНК с ДНК, проявляющихся в данных КС в виде чтений с разрывами, которые картируются только на два пространственно близких геномных локуса, взаимодействия РНК с РНК дают чтения, которые могут картироваться более сложным образом, поскольку пре-мРНК подвергаются сплайсингу. В частности, химерные фрагменты могут содержать как экзон-экзонные соединения (ЭЭС), так и продукты сшивки пространственно близких РНК, в результате чего образуются как чтения с каноническими интронными GT/AG-разрывами, возникающими при сплайсинге, так и чтения с другими разрывами, возникающими при лигировании (рис. 1А). Точное картирование та-

ких чтений представляет сложную задачу, поскольку большинство картировщиков может работать только с одним типом разрывов. Так, картировщик, имеющий только одну модель разрывов, был бы вынужден либо увеличивать штраф для разрывов без GT/AG-мотива для более точной идентификации координат канонических интронов, либо ослаблять требования к наличию GT/AG-мотива с целью более точного нахождения координат РНК-контактов. Поэтому разработка вычислительного метода, позволяющего картировать короткие чтения с двумя различными типами разрывов, является актуальной задачей биоинформатики. В данной работе представлен вычислительный конвейер, который позволяет достичь этой цели без разработки специализированного картировщика.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

### Геномы и их аннотации

Сборка генома человека hg19 (февраль 2009) и аннотация транскриптома GENCODE v34lift37 были загружены с веб-сайта Genome Reference Consortium [15] и веб-сайта GENCODE [16] соответственно. Координаты интронов получены из выходных данных программы STAR (см. ниже).

### Данные высокопроизводительного секвенирования

Две реплики данных RIC-seq с деплецией рРНК (GSM3629915 и GSM3629916) в клеточ-

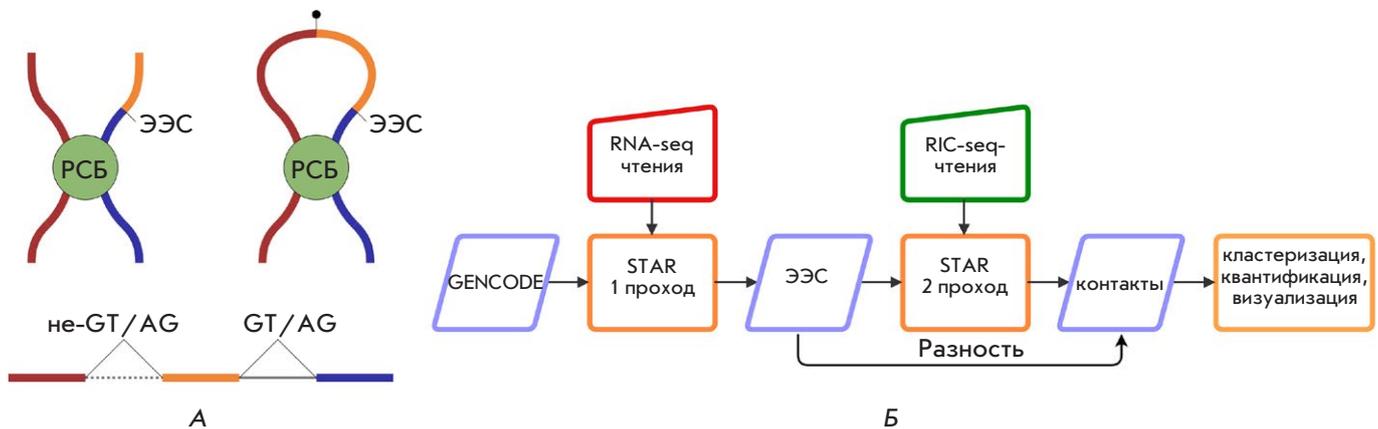


Рис. 1. А – в протоколе RIC-seq [14] несплайсированный участок РНК может быть лигирован через РНК-связывающий белок (РСБ) с другим участком, содержащим экзон-экзонные соединения (ЭЭС). Последовательность, образованная в результате лигирования, выравнивается на геном с двумя разрывами: не-GT/AG-разрывом, соответствующим точке лигирования, и каноническим GT/AG-разрывом, соответствующим сплайсингу. Б – схема RNAcontacts. На первом проходе чтения из контрольного эксперимента RNA-seq выравниваются на референсный геном для выявления экспрессируемых ЭЭС. Эти ЭЭС используются на втором проходе как известные интроны при выравнивании данных КС для обнаружения разрывов, которые кодируют контакты РНК-РНК

ной линии HeLa [14] были загружены из Gene Expression Omnibus в формате FASTQ (GSE127188). Соответствующий набор контрольных данных секвенирования РНК (RNA-seq) в клеточной линии HeLa был загружен из консорциума ENCODE (ENCLB555ASI и ENCLB555ASJ). На первом проходе данные RNA-seq картировали на геном человека с помощью программы-картировщика STAR версии 2.7.3a в парноконцевом режиме со следующими дополнительными параметрами:

```
--runMode alignReads --outSAMtype BAM SortedByCoordinate --chimOutType Junctions.
```

На втором проходе данные RIC-seq картировали на геном человека с помощью той же версии STAR со следующими дополнительными параметрами:

```
--chimSegmentMin 15 --chimJunctionOverhangMin 15 --chimScoreJunctionNonGTAG -1 --scoreGapNoncan -1 --scoreGapATAC -1 --scoreGapGCAG -1 --chimSegmentReadGapMax 3 --outFilterMatchNminOverLread 0.5 --outFilterScoreMinOverLread 0.5
```

Параметр `--chimSegmentReadGapMax 3` введен для пропуска при картировании дополнительного биотинилированного остатка цитозина в RIC-seq [14]. При втором проходе штраф снижается до `-1` для всех типов неканонических разрывов.

### Реализация метода

RIC-contacts реализован в популярной системе управления задачами Snakemake [17] и находится в свободном доступе на GitHub [18]. Пути к входным файлам указываются в конфигурационном файле в формате `yaml`, который также содержит настройки STAR и дополнительные параметры, контролирующие минимальное расстояние между двумя точками лигирования в кластере и порог разброса, для визуализации коллинеарных разрывов через UCSC Genome Browser [19]. Коллинеарные чтения извлечены из BAM-файлов с помощью отдельных программ (`neo.pl` в репозитории RNAcontacts) и пакета `samtools v1.14` [20]. Пакет `bedops v2.4.41` использовался для кластеризации точек разрыва [21]. Количество поддерживающих чтений подсчитано с помощью пакета `bedtools v2.29.0` [22].

### Визуализация результатов

Для визуализации карт контактов списки контактов были конвертированы в формат `'cool'` с помощью пакета `'cooler'` v0.8.11 с разрешением 100 п.н. Полученные карты визуализировали с помощью

пакета `pygenometracks v3.7`. Визуализация чтений с разрывами выполнена с помощью `IGV v2.11.2` и `UCSC Genome Browser` [19]. По умолчанию в UCSC Genome Browser визуализируются только коллинеарные контакты с разбросом не более 50000 нуклеотидов (см. также руководство [18]).

## РЕЗУЛЬТАТЫ

### Метод RNAcontacts

Для анализа данных КС РНК нами разработан вычислительный конвейер RNAcontacts, который обходит проблему нескольких различных типов разрывов путем выравнивания коротких чтений в двухпроходном режиме (рис. 1Б). Метод основан на использовании картировщика STAR [23], который на первом этапе выравнивает набор данных секвенирования РНК (RNA-seq) в парноконцевом режиме, чтобы определить интроны, которые экспрессируются в данном биологическом образце, используя строгий штраф для не-GT/AG-разрывов. Здесь RNA-seq представляет собой контрольный эксперимент, не содержащий химерных фрагментов, полученных в результате ЛПБМ. На втором проходе чтения сгенерированные в эксперименте КС РНК выравниваются с использованием ослабленного штрафа за не-GT/AG-разрывы, при этом на вход подается список разрывов, идентифицированных на первом проходе как интроны, так что картировщик будет преимущественно делать разрывы в чтениях по координатам из предоставленного списка. Поскольку данные КС РНК могут содержать химерные чтения на произвольном геномном расстоянии или *in trans*, выравнивание на втором проходе выполняется в одноконцевом режиме. Все выравнивания с разрывами, полученные на втором проходе, анализируются для извлечения контактов РНК-РНК и исключения интронов, полученных на первом проходе.

Программы выравнивания чтений с разрывами обычно генерируют два отдельных выходных файла, соответствующих коллинеарным и неколлинеарным разрывам. В частности, программа-картировщик STAR сообщает о коллинеарных разрывах (одна и та же хромосома, одна и та же цепь, согласованный порядок картирования фрагментов) в стандартном выходном файле SAM/BAM, в то время как неколлинеарные разрывы помещаются в отдельный выходной файл с химерными разрывами, так как BAM-формат не позволяет представить их с помощью одной CIGAR-строки [23]. RNAcontacts извлекает координаты неоразрывов, т.е. коллинеарных разрывов, найденных на втором проходе, из результатов SAM/BAM и объединя-

ет их с химерными разрывами. Следует отметить, что не только *trans*-, но и *cis*-контакты могут быть закодированы как в нео-, так и в химерных разрывах. Объединенный результат второго прохода состоит из нео- и химерных разрывов в чтениях, которые соответствуют точкам лигирования РНК.

В ходе эксперимента RIC-seq в клеточной линии HeLa [14] были картированы 94.3% из ~224 млн чтений из двух реплик, где 72.0% картированных чтений картированы уникально (табл. S1). При этом 18.5% уникально картированных чтений содержали по крайней мере один разрыв, в то время как значения для LIGR-seq, PARIS и SPLASH, как сообщалось ранее, составили 3.5, 2 и 0.5% соответственно [24]. Отметим, что результаты работы программ картирования коротких чтений, учитывающих сплайсинг, могут различаться. При использовании протокола RIC-seq небольшие различия в координатах разрывов могут возникать даже при сопоставлении чтений, которые перекрывают одну и ту же точку лигирования, поскольку картирование одной и той же последовательности на одну или другую цепь может приводить к смещению координат из-за отсутствия консенсусных последовательностей у точек лигирования (рис. S1). Кроме того, различные копии одной и той же РНК разрезаются и повторно лигируются стохастически, что приводит к еще большей вариабельности. Учитывая эту техническую и биологическую вариативность, мы ожидаем выявление кластеров точек лигирования, а не четко определенных сайтов, как это имеет место в случае сайтов сплайсинга в GT/AG-интронах.

Распределение расстояний между двумя последовательными точками разрыва быстро убывает, причем приблизительно 50% расстояний составляет меньше 9 нуклеотидов и 90% расстояний – меньше 21 нуклеотида (рис. S2). Поэтому точки разрыва были кластеризованы с использованием односвязной (single-linkage) кластеризации с порогом по расстоянию ( $\delta$ ) 10 и 20 нуклеотидов (рис. S3). Контакты определены как пары кластеров, соединяемые хотя бы одним разрывом, а число чтений, поддерживающих контакт, определяли как суммарное число чтений, поддерживающих все разрывы из кластера.

Для каждого значения  $\delta$  мы разделили контакты на три группы: внутригенные (оба конца контакта принадлежат аннотированному гену), контакты *in cis* (на одной хромосоме, но не в одном гене) и контакты *in trans* (на разных хромосомах). Число контактов ( $n$ ), длина кластера ( $s$ ), расстояние между контактирующими кластерами ( $d$ , которое определяется только для внутригенных и *cis*-контактов) и число поддерживающих чтений ( $r$ )

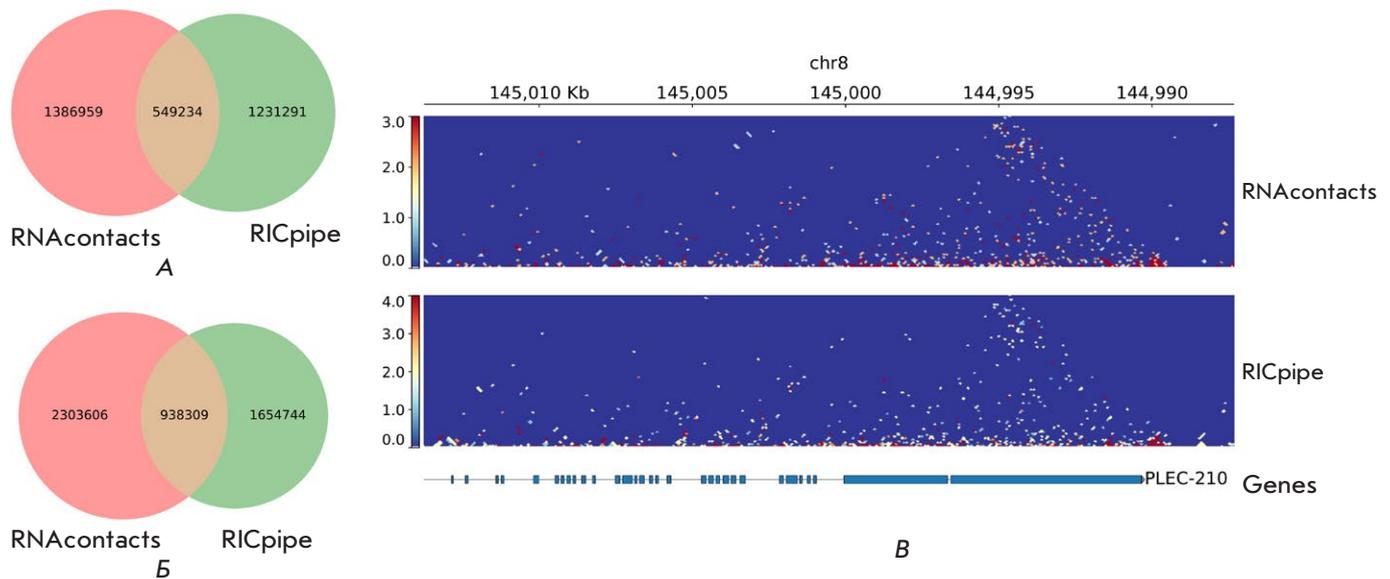
Таблица 1. Характеристика кластеров контактов РНК-РНК

$\delta$	Метрика	Внутригенные кластеры контактов	<i>in cis</i>	<i>in trans</i>
10	n	1369158	1061470	4881920
	s	10.1±2.2	10.1±2.1	10.2±2.2
	log <sub>2</sub> d	10.8±3.3	17.6±4.4	нет
	log <sub>2</sub> r	0.8±0.8	0.5±0.7	0.5±0.7
20	n	1313727	1035656	4851697
	s	20.2±2.4	20.2±2.3	20.3±2.4
	log <sub>2</sub> d	11.0±3.3	17.6±4.4	нет
	log <sub>2</sub> r	0.8±0.8	0.5±0.7	0.5±0.7

Примечание. Расстояние между кластерами,  $\delta$ . Количество контактов,  $n$ . Длина кластера,  $s$ . Расстояние между контактирующими кластерами,  $d$ . Количество чтений, поддерживающих контакт,  $r$ . Показано среднее значение ± стандартное отклонение.

незначительно различались для двух значений  $\delta$  (табл. 1). В среднем мы обнаружили на 30% больше внутригенных контактов, чем контактов *in cis*, и более чем двукратное обогащение контактов *in trans* по отношению к двум другим группам. При  $\delta = 10$  большинство кластеров имели длину 10 нуклеотидов (рис. S4), что указывает на то, что они состоят только из одной индивидуальной точки лигирования, окруженной с обеих сторон участками из 5 нуклеотидов.

Расстояния между контактирующими кластерами по-разному распределены в нео- и химерных разрывах в чтениях как по количеству собственно контактов, так и при взвешивании по количеству поддерживающих чтений (рис. S5). Примечательно, что распределения имеют две моды, причем первая мода  $d=1000$  соответствует внутригенным контактам, задаваемым как нео-, так и химерными разрывами. Химерные разрывы могут задавать внутригенные контакты, если лигирование произошло в обратной ориентации, как например, в кольцевых РНК [25]. Вторая мода для неоразрывов обусловлена условием  $d \leq 250000$ , которое накладывает картировщик STAR на коллинеарные разрывы, однако более длинные контакты *in cis* захвачены химерным распределением. При этом большинство контактов *in cis* и *in trans* поддерживается только одним чтением.



**Рис. 2.** Сравнение RNAcontacts и RICpipe. *А* – диаграмма Венна для координат разрывов, полученных по RNAcontacts и RICpipe. *Б* – то же, но координаты разрывов взвешены по их поддержке чтениями. *В* – карты контактов в гене PLEC-210, полученные из RNAcontacts (вверху) и RICpipe (внизу)

ем, в то время как большинство внутригенных контактов поддерживается двумя чтениями (*рис. S6*). Таким образом, поддержка чтениями в отдельных экспериментах RIC-seq довольно скудна даже после объединения контактов в кластеры.

### Чувствительность RNAcontacts

Чтобы сравнить производительность RNAcontacts с производительностью RICpipe (метод, изначально разработанный для анализа данных RIC-seq), мы проанализировали внутрихромосомные разрывы в чтениях с разбросом 50 нуклеотидов или более. Мы исключили разрывы в чтениях, картируемых на рРНК, из результатов RNAcontacts, поскольку RICpipe также удаляет чтения рРНК [14]. Только 40% (соответственно, 45%) разрывов, идентифицированных с помощью RNAcontacts (соответственно, RICpipe), имели точно такие же координаты, как и разрывы, идентифицированные другим методом, что указывает на различия в процедурах картирования (*рис. 2А*). Однако RNAcontacts удалось выровнять больше чтений, поддерживающих идентифицированные разрывы, по сравнению с RICpipe, что указывает на увеличение чувствительности примерно на 40% (*рис. 2Б*). При проведении сравнения с использованием окон размером 100 нуклеотидов, т.е. без точного совпадения координат, мы заметили, что результаты двух методов в значительной степени согласованы, о чем также свидетельствуют визуально похожие

карты контактов с немного большим количеством контактов, полученных RNAcontacts, по сравнению с RICpipe (*рис. 2В*).

Затем мы проверили производительность RNAcontacts на данных RIC-seq в клеточной линии HeLa с первым проходом картирования и без него. Для этого мы отдельно выполнили второй проход RNAcontacts, предоставляя только аннотированные интроны из GENCODE [16], не добавляя специфичные для клеточной линии HeLa интроны, найденные на первом проходе. В результате получен примерно 1% ложных соединений лигирования, которые соответствуют неаннотированным интронам в HeLa. Мы также обнаружили, что 16809 из ~3.5 млн разрывов в чтениях, идентифицированных с помощью RICpipe, соответствуют экзон-экзонным соединениям. Хотя количество таких разрывов невелико, они поддерживаются значительной долей (>30%) чтений. Таким образом, двухпроходный метод обеспечивает более высокую специфичность (меньший процент ложноположительных результатов) при обнаружении контактов РНК, особенно в условиях, когда экспрессируемый транскриптом значительно отличается от аннотированного.

### ОБСУЖДЕНИЕ И ВЫВОДЫ

В данной работе представлено концептуальное решение проблемы картирования чтений с двумя типами разрывов, характерными для экспериментов КС РНК. Несмотря на то, что представленное реше-

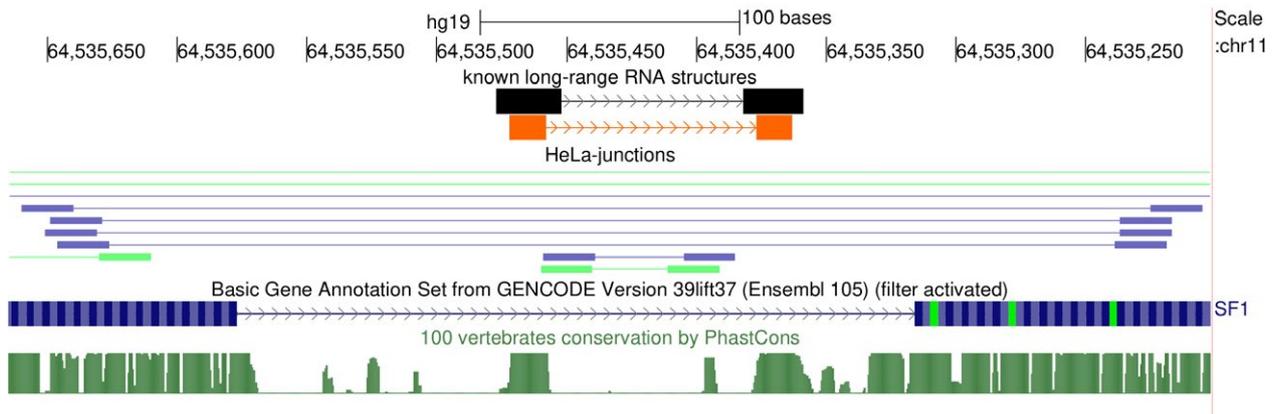


Рис. 3. Разрывные чтения, поддерживающие структуру РНК в гене *SF1* человека [28]. Комплементарные цепи показаны оранжевым цветом. Ниже показаны разрывные чтения (чтения из двух реплик показаны синим и зеленым цветом)

ние задействует картировщик STAR, предложенный подход не ограничивается использованием только этой программы, и вместо нее может быть использована любая другая программа для выравнивания чтений с разрывами [26]. Мы продемонстрировали, что экзон-экзонные соединения составляют большую часть чтений с разрывами в данных RIC-seq, и RNAcontacts позволяет обнаруживать разрывы, соответствующие точкам лигирования в экспериментах RIC-seq, с большей чувствительностью, чем RICpipe. Реализация RNAcontacts в воспроизводимой и масштабируемой системе управления рабочим процессом Snakemake позволяет быстро и единообразно обрабатывать множество наборов однотипных данных КС.

Данные экспериментов конформационного секвенирования РНК похожи по характеру на данные Hi-C, но имеют важные отличия, связанные с разрешением. Если для Hi-C обычной практикой является усреднение контактов хроматина на уровне тысяч или миллионов пар нуклеотидов, то обнаружение контактов РНК с помощью КС по своей сути нацелено на масштаб отдельных нуклеотидов. При этом поддержка известных контактов чтениями RIC-seq, например, в известной вторичной структуре в гене *SF1*, очень слаба (рис. 3). Большинство контактов RIC-seq *in cis* и *in trans* поддерживаются только одним чтением, в связи с чем возникает проблема оценки статистической значимости наблюдаемых контактов. Эта проблема, возможно, будет решена в исследованиях, анализирующих большие наборы экспериментов RIC-seq. Мы ожидаем, что в скором времени появится гораздо больше наборов данных

экспериментов КС РНК, аналогичных RIC-seq, которые могут быть проанализированы представленным методом.

Таким образом, RNAcontacts реализует вычислительный конвейер анализа контактов РНК-РНК, учитывающий два типа разрывов, специфичных для метода КС РНК. Несмотря на то, что программное обеспечение разработано для протокола RIC-seq, область его применения распространяется на любой аналогичный метод КС, в котором одним из взаимодействующих партнеров является РНК. ●

*Авторы выражают благодарность Т.М. Иванову за дополнительное тестирование программы. Работа поддержана грантами Министерства науки и образования Российской Федерации (075-10-2021-116) и исследовательского гранта Национальной ключевой программы исследований и разработок Китая (2021YFE0114900).*

*Авторы заявляют, что у них нет конкурирующих интересов.*

*Д.Д.П. разработал план и руководил исследованием; Г.Л. и Ч.К. предоставили данные и участвовали в их обработке; С.Д.М. провел анализ данных и разработал программное обеспечение. Все авторы участвовали в написании текста статьи, прочитали и одобрили его.*

### ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Дополнительные данные и материалы доступны в репозитории Zenodo (<https://zenodo.org/record/7027475>) [27].

СПИСОК ЛИТЕРАТУРЫ

1. Jerkovic I, Cavalli G. // *Nat. Rev. Mol. Cell. Biol.* 2021. V. 22. № 8. P. 511–528.
2. McCord R.P., Kaplan N., Giorgetti L. // *Mol. Cell.* 2020. V. 77. № 4. P. 688–708.
3. Han J., Zhang Z., Wang K. // *Mol. Cytogenet.* 2018. V. 11. P. 21.
4. Lu L., Liu X., Huang W.K., Giusti-Rodríguez P., Cui J., Zhang S., Xu W., Wen Z., Ma S., Rosen J.D., et al. // *Mol. Cell.* 2020. V. 79. № 3. P. 521–534.
5. Li G., Cai L., Chang H., Hong P., Zhou Q., Kulakova E.V., Kolchanov N.A., Ruan Y. // *BMC Genomics.* 2014. V. 15 Suppl. 12. P. S11.
6. 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.
7. Limouse C., Jukam D., Smith O.K., Fryer K.A., Straight A.F. // *Meth. Mol. Biol.* 2020. V. 2161. P. 115–142.
8. Kudla G., Wan Y., Helwak A. // *Annu. Rev. Genomics Hum. Genet.* 2020. V. 21. P. 81–100.
9. Xu B., Zhu Y., Cao C., Chen H., Jin Q., Li G., Ma J., Yang S.L., Zhao J., Zhu J., et al. // *Sci. China Life Sci.* 2022. V. 65. № 7. P. 1285–1324.
10. Lu Z., Zhang Q.C., Lee B., Flynn R.A., Smith M.A., Robinson J.T., Davidovich C., Gooding A.R., Goodrich K.J., Mattick J.S., et al. // *Cell.* 2016. V. 165. № 5. P. 1267–1279.
11. Sharma E., Sterne-Weiler T., O'Hanlon D., Blencowe B.J. // *Mol. Cell.* 2016. V. 62. № 4. P. 618–626.
12. Aw J.G., Shen Y., Wilm A., Sun M., Lim X.N., Boon K.L., Tapsin S., Chan Y.S., Tan C.P., Sim A.Y., et al. // *Mol. Cell.* 2016. V. 62. № 4. P. 603–617.
13. Ziv O., Gabryelska M.M., Lun A.T.L., Gebert L.F.R., Sheu-Gruttadauria J., Meredith L.W., Liu Z.Y., Kwok C.K., Qin C.F., MacRae I.J., et al. // *Nat. Methods.* 2018. V. 15. № 10. P. 785–788.
14. Cai Z., Cao C., Ji L., Ye R., Wang D., Xia C., Wang S., Du Z., Hu N., Yu X., et al. // *Nature.* 2020. V. 582. № 7812. P. 432–437.
15. Church D.M., Schneider V.A., Graves T., Auger K., Cunningham F., Bouk N., Chen H.C., Agarwala R., McLaren W.M., Ritchie G.R., et al. // *PLoS Biol.* 2011. V. 9. № 7. P. e1001091.
16. Harrow J., Frankish A., Gonzalez J.M., Tapanari E., Diekhans M., Kokocinski F., Aken B.L., Barrell D., Zadissa A., Searle S., et al. // *Genome Res.* 2012. V. 22. № 9. P. 1760–1774.
17. Mölder F., Jablonski K.P., Letcher B., Hall M.B., Tomkins-Tinch C.H., Sochat V., Forster J., Lee S., Twardziok S.O., Kanitz A., et al. // *F1000Res.* 2021. V. 10. P. 33.
18. Margasyuk S., Pervouchine D. RNAcontacts pipeline. <https://doi.org/10.5281/zenodo.6512482>. 2022.
19. Raney B.J., Dreszer T.R., Barber G.P., Clawson H., Fujita P.A., Wang T., Nguyen N., Paten B., Zweig A.S., Karolchik D., Kent W.J. // *Bioinformatics.* 2014. V. 30. № 7. P. 1003–1005.
20. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. // *Bioinformatics.* 2009. V. 25. № 16. P. 2078–2079.
21. Neph S., Kuehn M.S., Reynolds A.P., Haugen E., Thurman R.E., Johnson A.K., Rynes E., Maurano M.T., Vierstra J., Thomas S., et al. // *Bioinformatics.* 2012. V. 28. № 14. P. 1919–1920.
22. Quinlan A.R., Hall I.M. // *Bioinformatics.* 2010. V. 26. № 6. P. 841–842.
23. Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. // *Bioinformatics.* 2013. V. 29. № 1. P. 15–21.
24. Schönberger B., Schaal C., Schäfer R., Voß B. // *F1000Res.* 2018. V. 7. P. 1824
25. Eger N., Schoppe L., Schuster S., Laufs U., Boeckel J.N. // *Adv. Exp. Med. Biol.* 2018. V. 1087. P. 41–52.
26. Engström P.G., Steijger T., Sipoš B., Grant G.R., Kahles A., Rättsch G., Goldman N., Hubbard T.J., Harrow J., Guigó R., et al. // *Nat. Methods.* 2013. V. 10. № 12. P. 1185–1191.
27. Margasyuk S., Vlasenok M., Li G., Cao C., Pervouchine D. RIC-seq contacts in HeLa. 2022. URL: <https://doi.org/10.5281/zenodo.7027475>.
28. Pervouchine D.D., Khrameeva E.E., Pichugina M.Y., Nikolaienko O.V., Gelfand M.S., Rubtsov P.M., Mironov A.A. // *RNA.* 2012. V. 18. № 1. P. 1–15.