

Апробация различных вариантов RNA-seq для идентификации аутронов генов у плоского червя *Opisthorchis felineus*

Н.И. Ершов¹✉, Д.Е. Маслов², Н.П. Бондарь^{1,2}

¹ Федеральное исследовательское учреждение Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ e-mail: ershov@bionet.nsc.ru

Аннотация. *Opisthorchis felineus* – представитель паразитических плоских червей, один из возбудителей описторхоза человека. Недавно нами была проведена сборка генома *O. felineus*, однако корректная аннотация генов в этом геноме стандартными методами оказалась затруднена наличием сплайс-лидера зависимого транс-сплайсинга (SLTS). В результате SLTS исходный 5'-конец (ауτροн) транскриптов заменяется короткой сплайс-лидерной последовательностью, донором которой выступает специализированная молекула SL РНК. SLTS вовлечен в процессинг РНК более половины всех генов *O. felineus*, из-за чего становится невозможным установить последовательности ауτροнов и реальные старты транскрипции соответствующих генов и оперонов, опираясь только на данные mRNA-seq. В настоящей работе мы провели апробацию различных экспериментальных подходов для идентификации последовательностей ауτροнов у *O. felineus* с помощью массового параллельного секвенирования. Два подхода были спланированы нами для прицельного секвенирования процессированных разветвленных ауτροнов. Первый заключался в сиквенс-специфичной обратной транскрипции с SL-интрона в направлении 5'-конца ауτροна. Во втором использовалась гибридизация ауτροнов с иммобилизованным одноцепочечным ДНК-зондом, комплементарным SL-интрону. Также были использованы два подхода к секвенированию тотальной РНК, обедненной по рРНК, позволяющих идентифицировать более широкий спектр транскриптов, чем mRNA-seq. Один из них основан на ферментативной элиминации перепредставленных кДНК, другой – на ферментативной деградации неэкспонированных РНК экзонуклеазой Terminator. С помощью селективных методов нам не удалось получить обогащения препаратов РНК по процессированным ауτροнам, что, наиболее вероятно, связано с коротким временем жизни этих промежуточных продуктов транс-сплайсинга. Из двух методов обеднения по рРНК высокую эффективность показал метод, основанный на ферментативной нормализации кДНК (Zymo-Seq RiboFree). Он позволил примерно вдвое увеличить долю прочтений, соответствующих ауτροнам и интронам, по сравнению с mRNA-seq. Полученные результаты предполагают, что основным ресурсом последовательностей ауτροнов в пуле РНК *O. felineus* служат новосинтезированные непротранскрипированные транскрипты. Ключевые слова: описторхоз; сплайс-лидер зависимый транс-сплайсинг; ауτροн; старт транскрипции; транскриптом; рибосомальная фракция РНК.

Для цитирования: Ершов Н.И., Маслов Д.Е., Бондарь Н.П. Апробация различных вариантов RNA-seq для идентификации ауτροнов генов у плоского червя *Opisthorchis felineus*. Вавиловский журнал генетики и селекции. 2020; 24(8):897-904. DOI 10.18699/VJ20.688

Evaluation of various RNA-seq approaches for identification of gene outrons in the flatworm *Opisthorchis felineus*

N.I. Ershov¹✉, D.E. Maslov², N.P. Bondar^{1,2}

¹ Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

✉ e-mail: ershov@bionet.nsc.ru

Abstract. The parasitic flatworm *Opisthorchis felineus* is one of the causative agents of opisthorchiasis in humans. Recently, we assembled the *O. felineus* genome, but the correct genome annotation by means of standard methods was hampered by the presence of spliced leader trans-splicing (SLTS). As a result of SLTS, the original 5'-end (outtron) of the transcripts is replaced by a short spliced leader sequence donated from a specialized SL RNA. SLTS is involved in the RNA processing of more than half of *O. felineus* genes, making it hard to determine the structure of outrons and *bona fide* transcription start sites of the corresponding genes and operons, being based solely on mRNA-seq data. In the current study, we tested various experimental approaches for identifying the sequences of outrons in *O. felineus* using massive parallel sequencing. Two of them were developed by us for targeted sequencing of already processed branched outrons. One was based on sequence-specific reverse transcription from the SL intron toward the 5'-end of the Y-branched outtron. The other used outtron hybridization with an immobilized single-stranded DNA probe complementary to the SL intron. Additionally, two approaches to the sequencing of rRNA-depleted total RNA were used, allowing the identification of a wider range of transcripts compared to mRNA-seq. One is based on the enzymatic

elimination of overrepresented cDNAs, the other utilizes exonucleolytic degradation of uncapped RNA by Terminator enzyme. By using the outtron-targeting methods, we were not able to obtain the enrichment of RNA preparations by processed outtrons, which is most likely indicative of a rapid turnover of these trans-splicing intermediate products. Of the two rRNA depletion methods, a method based on the enzymatic normalization of cDNA (Zymo-Seq RiboFree) showed high efficiency. Compared to mRNA-seq, it provides an approximately twofold increase in the fraction of reads originating from outtrons and introns. The results suggest that unprocessed nascent transcripts are the main source of outtron sequences in the RNA pool of *O. felineus*.

Key words: opisthorchiasis; spliced leader trans-splicing; outtron; start of transcription; transcriptome; ribosomal RNA.

For citation: Ershov N.I., Maslov D.E., Bondar N.P. Evaluation of various RNA-seq approaches for identification of gene outtrons in the flatworm *Opisthorchis felineus*. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):897-904. DOI 10.18699/VJ20.688

Введение

Opisthorchis felineus – представитель паразитических плоских червей класса Trematoda (сосальщики) семейства Opisthorchiidae, обладающий сложным жизненным циклом со сменой двух промежуточных хозяев, окончательным хозяином которого оказывается человек (Безр, 2005). Описторхоз, вызываемый паразитированием этого сосальщика в желчных протоках печени человека, имеет хронический характер, приводит к ряду серьезных сопутствующих нарушений гепатобилиарной системы, в числе которых холангит, холецистит, панкреатит, а также является фактором риска развития холангиокарциномы (Sripa et al., 2007; Pakharukova, Mordvinov, 2016; Pakharukova et al., 2019). По оценке ВОЗ, более 1 млн людей могут быть заражены *O. felineus*, при этом самый крупный очаг описторхоза расположен в Обь-Иртышском бассейне. В некоторых регионах Томской и Тюменской областей заболеваемость населения достигает 60 % (FAO/WHO, 2014; Fedorova et al., 2018).

В исследовании биологии этого плоского червя и развитии молекулярно-генетических подходов к диагностике и фармакотерапии описторхоза важным шагом стало получение референсной сборки генома и ее аннотация (Ershov et al., 2019). Однако аннотация генов, полученная на основе данных polyA-mRNA-seq, стандартно применяемых для этой задачи, имеет существенные недостатки в связи с одной из особенностей процессинга РНК у плоских червей – сплайс-лидер зависимого транс-сплайсинга (splice leader trans-splicing, SLTS). В ходе SLTS происходит объединение экзонов двух независимых транскриптов: специализированной короткой экзипированной сплайс-лидерной РНК (SL РНК), несущей 5'-сайт сплайсинга, и той или иной пре-РНК, имеющей соответствующий 3'-сайт (см. рис. 1, б). При этом удаляется исходный 5'-участок пре-РНК (так называемый аутрон), к сайту ветвления которого присоединен по 2'-5'-фосфодиэфирной связи интрон SL РНК (Murphy et al., 1986; Sutton, Boothroyd, 1988). Предполагается, что далее этот Y-образный продукт (Y-аутрон) быстро подвергается дебранчингу и деградации (Sutton, Boothroyd, 1988; Lasda, Blumental, 2011). Таким образом, в данных polyA-mRNA-seq практически отсутствуют последовательности вырезанных аутронов, а с ними оказывается недоступной для аннотации и информация об истинных стартах транскрипции соответствующих генов и локализации их промоторов, а также информация об оперонной организации групп генов. Следует отметить, что по причине SLTS оказывается неэффективной и большая группа экспериментальных ме-

тодов идентификации стартов транскрипции, основанных на селективном секвенировании экзипированных 5'-концов РНК.

По нашим оценкам, в геноме *O. felineus* продукты более половины всех генов подвергаются транс-сплайсингу (Ershov et al., 2019). Для корректной аннотации таких генов необходимы альтернативные высокопроизводительные методы секвенирования, обеспечивающие массовую идентификацию аутронов. Как правило, для описания аутронов генов исследователи накладывают информацию о сайтах инициации транскрипции, полученную такими методами, как GRO-seq, GRO-cap, ChIP-seq к РНК-полимеразе II, на данные по сайтам SLTS, выявляемым с помощью mRNA-seq (Chen et al., 2013; Kguesi et al., 2013). Соответственно, участок генома от старта транскрипции до 3'-сайта SLTS принимается как аутрон. Заметим, что основная часть этих методов требует значительного количества стартового биологического материала и в случае небольшого размера особей подходит скорее для культивируемых в лаборатории организмов. Известны примеры, когда удается детектировать аутроны непосредственно методами секвенирования РНК. Так, в одной из работ на *C. elegans* культивирование особей при низкой температуре приводило к накоплению непротранскрибированных транскриптов, что позволило детектировать 5'-концы аутронов методом SAGE в ядерной фракции РНК (Saito et al., 2013).

На немодельных организмах, включая трематод, для прямой массовой детекции транс-сплайсированных РНК были успешно применены RNA-seq методы “SL Trapping”, основанные на селекциях по универсальной 5'-последовательности SL-экзона в этих транскриптах (Nilsson et al., 2010; Boroni et al., 2018). Поскольку все Y-аутроны также содержат универсальную 3'-последовательность SL-интрона, представляется привлекательным использовать это свойство для селективной идентификации аутронов и симметричного подтверждения сайтов SLTS. Кроме того, непосредственная идентификация Y-аутронов послужила бы принципиальным доказательством общепринятого механизма SLTS на большой выборке генов.

В настоящей работе для идентификации аутронов *O. felineus* мы опробовали два кандидатных подхода, основанных на прицельном обогащении по Y-аутронам: в одном случае посредством сиквенс-специфичной обратной транскрипции с SL-интрона в направлении 5'-конца основной цепи аутрона, в другом – путем гибридизации Y-аутронов с иммобилизованным на магнитных частицах одноцепочечным ДНК-зондом, комплементарным SL-интрону. Помимо того, параллельно были использова-

ны два пригодных для немодельных организмов подхода к секвенированию библиотек тотальной РНК, обедненной по рРНК, не использующих гибридационные пробы: обработку экзонуклеазой Terminator и коммерческий набор Zymo-Seq RiboFree, основанный на ферментативной нормализации кДНК. Доля содержащих аутроны последовательностей в этих библиотеках должна быть заведомо выше, чем в стандартных библиотеках polyA-mRNA-seq.

Материалы и методы

Биоматериал. Взрослых червей *O. felineus* выделяли из желчных протоков печени хомячков *Mesocricetus auratus* через 3–4 мес. после заражения метацеркариями, полученными из тканей зараженных рыб (язь) из р. Обь. Червей промывали в физиологическом растворе и использовали сразу или замораживали при -80°C до использования.

РНК выделяли из замороженных образцов *O. felineus* с использованием PureZOL (BioRad, США) согласно протоколу производителя. Осаждение РНК проводили добавлением равного объема изопропанола и 5 мкг линейного полиакриламида (LPA) в качестве соосадителя. Полученный раствор оставляли на ночь при -20°C . Осадок растворяли в бидистиллированной воде и выполняли дополнительную очистку проб с использованием Aurum total RNA mini kit (BioRad, США), включающую в себя стадию обработки ДНКазой I. Качество и количество выделенной РНК оценивали с помощью спектрофотометра NanoDrop 2000.

Обработка Terminator. Обработку тотальной РНК ферментом Terminator (5'-Phosphate-Dependent Exonuclease, Epicentre, США) проводили в соответствии с протоколом производителя. Реакционная смесь содержала 500 нг тотальной РНК, 1 мкл RiboLock (ThermoFisher Scientific, США), 2 мкл $10\times$ буфера А, 1 U экзонуклеазы Terminator. Объем доводили до 20 мкл бидистиллированной водой. Образец инкубировали 60 мин при 30°C , реакцию останавливали добавлением 1 мкл 100 мМ EDTA. Очистку пробы после реакции проводили с использованием Agencourt RNAClean XP beads (Beckman Coulter, США). Уменьшение количества рибосомальной РНК проверяли на Agilent 2100 Bioanalyzer с использованием RNA 6000 Pico chip.

Выделение фракции РНК, обогащенной по Y-ауτροнам (SLi-BC). Биотинилированный оцДНК-зонд ([biotin-TEG]-5'-GGTAGCCAAATAATTCATCCGACCATAGGCCGGAGTCGATTCTT-3') иммобилизовали на магнитных частицах (Dynabeads M-280 Streptavidin, Invitrogen, США) согласно протоколу производителя, контролируя примерное соответствие количества связанного зонда заявленной емкости частиц (~ 200 пМ/мг). К магнитным частицам со связанным зондом добавляли 1 мкг РНК в буфере для гибридизации (1 М NaCl, 5 мМ Tris-HCl (pH 7.5), 0.5 мМ EDTA, 0.1 % Tween-20) и инкубировали с постоянным перемешиванием по следующему протоколу: 75°C в течение 2 мин, затем понижали температуру на $1^{\circ}\text{C}/\text{мин}$ до 55°C и инкубировали 15 мин при 55°C . Частицы трижды промывали теплым (50°C) буфером. К частицам добавляли 10 мкл воды и нагревали до 94°C на 5 мин. Раствор, содержащий РНК, собирали в чистую пробирку и сразу проводили реакцию обратной транскрипции с помощью 200 U RevertAid Reverse Transcriptase (ThermoFisher Scientific, США) и случайных гексамеров в качестве праймеров. Условия реакции: 5 мин при 25°C , 60 мин при 42°C и 5 мин при 70°C . Полученную кДНК использовали для приготовления библиотеки.

Для реакции обратной транскрипции сиквенс-специфичными праймерами (SLi-RT). Для реакции обратной транскрипции использовали термостабильную Maxima H Minus Reverse Transcriptase (ThermoFisher Scientific). Реакционную смесь, содержащую 300 нг РНК, 2 пМ специфического праймера к интрону SL-РНК (SLi_r1, 5'-AGGCCGGAGTCGATTCTT-3'), 1 мкл 10 мМ dNTP, 4 мкл 5 М бетаина, инкубировали 2 мин при 75°C , затем понижали температуру на $2^{\circ}\text{C}/\text{мин}$ до 55°C , добавляли 20 U RiboLock RNase inhibitors и 100 U Maxima H Minus Reverse Transcriptase и инкубировали 30 мин при 55°C , после чего реакцию инактивировали в течение 5 мин при 85°C . Полученную кДНК использовали для приготовления библиотеки.

Для оценки обогащения по Y-ауτροнам в ходе отработки протоколов SLi-BC и SLi-RT была использована ПЦР-РВ с праймерами к ауτροну (MMCE_ou_f: 5'-CCTGGCGACACACATCTGAA-3', MMCE_ou_r: 5'-ACATGGACATGGCTGAAGCA-3') и экзонам (MMCE_ex_f: 5'-TGCAACCTCTCTTGTGTTCCT-3', MMCE_ex_r: 5'-CCACCTGGACACCGAATGTAT-3') гена *mmce* (Приложение 1)¹. Обогащение рассчитывалось по методу $\Delta\Delta C_t$ как изменение соотношения ауτροна к экзону между контрольной и селективной кДНК. Контрольной служила кДНК, полученная ревертированием тотальной РНК со случайных гексамеров. Селективная кДНК в случае SLi-RT получена ревертированием тотальной РНК с праймера SLi_r1 и в случае SLi-BC – конвертированием обогащенной соответствующим образом фракции РНК с помощью случайных гексамеров.

ПЦР проводили в реакционной смеси объемом 20 мкл, содержащей 0.25 мМ dNTP, 2 мкл $10\times$ ПЦР-буфера Б+EVAGreen, 2.5 мМ MgCl_2 , смесь праймеров (по 10 пМ каждого), 2 мкл образца ДНК, 0.3 Ед/мкл SynTaq ДНК-полимеразы с ингибирующими активностью фермента антителами («Синтол», Россия). Условия реакции: предварительный нагрев до 9°C в течение 5 мин с последующими 39 циклами амплификации, включающими денатурацию при 95°C в течение 15 с, отжиг праймеров и элонгацию при 60°C в течение 20 с. Кривые плавления снимали в диапазоне температур $65\text{--}95^{\circ}\text{C}$. Все реакции выполнялись в двух технических повторностях.

Для приготовления библиотек RNA-seq после обработки Terminator использовали NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, США) по стандартному протоколу. Для приготовления библиотек из образцов кДНК, полученных по протоколам SLi-BC и SLi-RT, использовали NEBNext Ultra II Directional RNA Library Prep Kit for Illumina, начиная с этапа синтеза второй цепи кДНК. Фрагментацию полученной кДНК проводили с помощью фермента dsDNA Fragmentase, одновременно с репарированием концов ДНК. Селекцию фрагментов нужной длины осуществляли с помощью

¹ Приложения 1–3 см. по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2020-24/appx15.pdf>

Agencourt AMPure XP beads, после чего выполняли несколько циклов амплификации библиотеки.

Приготовление библиотек с обеднением по рибосомальной РНК проводили с помощью набора Zymo-Seq RiboFree Total RNA Library Prep Kit (Zymo Research, США) по стандартному протоколу. В реакцию брали 500 нг тотальной РНК, обеднение после ренатурации проводили в течение 60 мин.

Размер и количество полученных библиотек определяли с помощью Agilent 2100 Bioanalyzer (Прил. 2). Библиотеки были секвенированы с двух сторон (2×250 bp) на платформе Illumina MiSeq (услуга предоставлена АО «Вектор-Бест»). Исходные данные секвенирования библиотек депонированы в репозитории с открытым доступом Zenodo (Ershov, 2020). Ранее опубликованные данные mRNA-seq, использованные в работе, доступны в базе данных NCBI (PRJNA257351).

Вычислительная обработка данных секвенирования. Данные секвенирования в формате FASTQ были обработаны в программе Cutadapt (Martin, 2011) для удаления адаптерных последовательностей и картированы на референсный геном *O. felineus* (GenBank ID GCA_004794785.1) в программе STAR (Dobin et al., 2013). Разметку геномных элементов выполняли с использованием соответствующей аннотации генов и ранее полученных полногеномных данных по картированию сайтов транс-сплайсинга (Ershov et al., 2019), с использованием собственных Perl-скриптов. Сборка геномного повтора рРНК проводилась вручную по данным секвенирования геномной ДНК (Ershov et al., 2019). Границы генов рРНК в геноме были определены с помощью базы данных Rfam (<https://rfam.xfam.org>) и инструмента RNAmmer (Lagesen et al., 2007). Подсчет числа прочтений в заданных геномных интервалах осуществлялся с помощью инструмента featureCounts пакета Subread (Liao et al., 2019). Статистическая обработка и графическая визуализация данных выполнены в среде R.

Результаты и обсуждение

Обоснование выбранных стратегий

К настоящему времени нам не удалось найти упоминания в литературе экспериментальных методов для прямой массовой идентификации Y-аутронов в транскриптом. В ходе анализа полученных ранее данных mRNA-seq *O. felineus* (Ershov et al., 2019) мы обнаружили, что для многих генов 5'-районы выше сайта транс-сплайсинга (потенциальные аутроны), подобно интронам, содержат незначительное число картированных прочтений (рис. 1, а), несмотря на обеднение библиотек по некодирующим и непротранскрибированным последовательностям в результате polyA-селекции. Всего около 0.15 % прочтений mRNA-seq картировалось в районы аутронов (5 т. п. н. выше сайта) транс-сплайсируемых транскриптов (в среднем с 0.8-кратным покрытием). Источником таких прочтений, соответствующих аутронам генов, могут выступать: непротранскрибированные новосинтезированные пре-РНК; Y-аутроны – промежуточные разветвленные продукты транс-сплайсинга; а также продукты их дальнейшего дебранчинга, потерявшие ковалентно связанный с аутроном

SL-интрон (см. рис. 1, б). Мы предположили, что фракция Y-аутронов в исходном пуле тотальной РНК потенциально может составлять долю, достаточную для ее вычленения методами прицельного обогащения.

Возможность таргетирования Y-аутронов обусловлена наличием в их составе универсальной последовательности – интрона SL-РНК, ковалентно связанного в сайте ветвления (branchpoint) со сплайсированным аутроном пре-РНК (см. рис. 1, б). Используя это свойство, мы разработали два различных подхода к прямой селекции Y-аутронов. Первый, обозначенный SLi-RT (SL-intron Reverse Transcription), заключается в таргетированной обратной транскрипции с 2'-плеча (SL-интрона) в направлении варьирующего 5'-конца аутрона (см. рис. 1, в). Известно, что обратная транскриптаза M-MLV, лишенная активности РНКазы H, при праймировании с 2'-плеча способна достаточно часто проходить сайт ветвления с допущением одного мисматча и успешно ревертировать 5'-сегмент (Bitton et al., 2014; Döring, Hurek, 2017). Из ожидаемых недостатков метода можно было предполагать пониженную эффективность такого ревертирования, накопление продуктов миспрайминга и недопредставленность 5'-районов аутронов вследствие регулярного случайного обрыва синтеза первой цепи кДНК. В связи с последним выбором фермента и высокотемпературные условия ревертирования, включая горячий старт, оптимизировались для максимальной процессивности и специфичности синтеза.

Второй подход, SLi-BC (SL-intron Biotin Capture), основывался на гибридизации 2'-плеча аутронов с биотинилированным одноцепочечным ДНК-зондом, гомологичным SL-интрону, и последующей иммобилизации и очистке соответствующей фракции на магнитных частицах, покрытых стрептавидином (см. рис. 1, з). Такой подход должен быть более чувствителен к процентному содержанию целевой фракции в препарате РНК, но имеет больше шансов в идентификации полноразмерных аутронов.

Оба подхода нацелены на интактную Y-структуру процессированного аутрона, что позволило бы в том числе дифференцировать их от фрагментов новосинтезированных транскриптов, не прошедших транс-сплайсинг. Кроме того, мы использовали два метода секвенирования тотальной РНК, обедненной по рРНК, позволяющих идентифицировать и некодирующие транскрипты, включая интроны и аутроны. Поскольку *O. felineus* не является модельным видом, то популярные наборы для обеднения на основе гибридизации с рРНК-пробами оказываются неприменимы. Поэтому был опробован протокол обработки тотальной РНК экзонуклеазой Terminator, специфически гидролизующей нуклеиновые кислоты с фосфорилированным 5'-концом (в том числе рРНК), но не затрагивающей экзикованные последовательности. В качестве альтернативного был взят коммерческий протокол Zymo-Seq RiboFree, основанный на DSN-нормализации гибридов РНК:ДНК сразу после синтеза первой цепи кДНК, в результате которого гидролизуются все перепредставленные транскрипты, в том числе глубины и рРНК. В отличие от первых двух, эти методы не дают возможности отличать процессированные аутроны от их предшественников – непротранскрибированных новосинтезированных транскриптов,

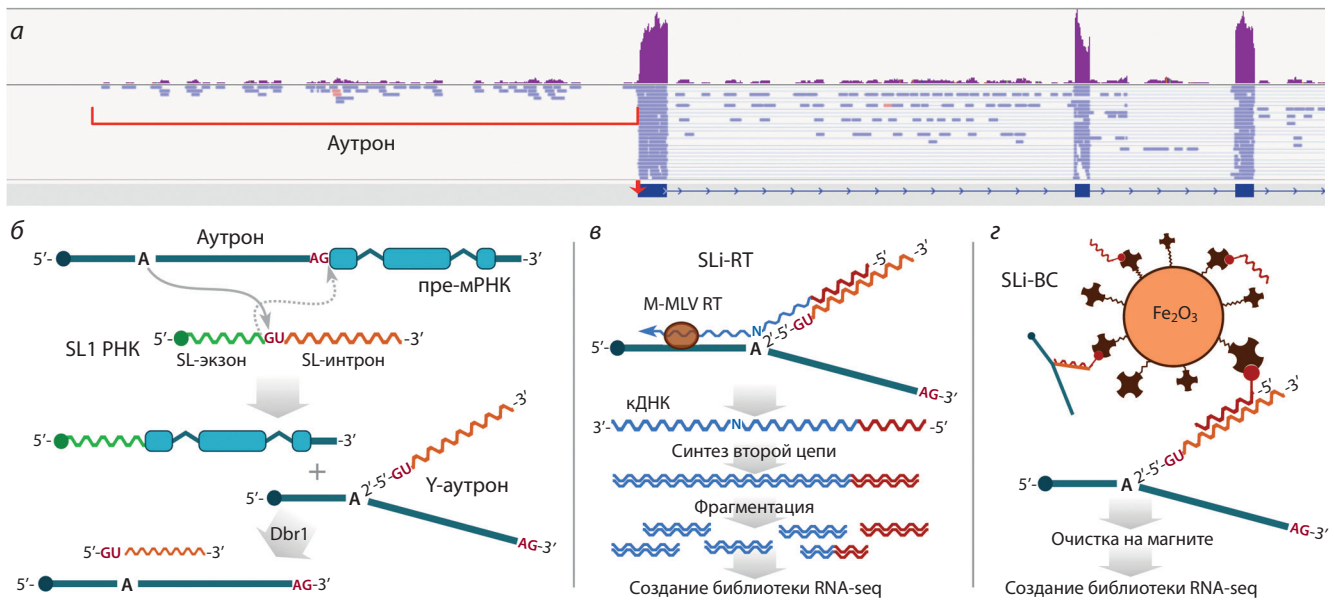


Рис. 1. Представленность последовательностей, соответствующих аутронам, в данных polyA-mRNA-seq *O. felineus* и схема предлагаемых методов прицельной идентификации Y-аутронов.

a – пример покрытия прочтениями аутронного участка транс-сплайсируемого гена (*Vps39l*) в данных mRNA-seq. Стрелкой указан подтвержденный сайт транс-сплайсинга; *б* – схема транс-сплайсинга пре-РНК с образованием Y-аутрона, содержащего SL-интрон, и его деградации с участием Dbr1; *в* – схема обратной транскрипции, праймированной с SL-интрона в составе аутрона, которая используется в подходе SLi-RT; M-MLV RT – обратная транскриптаза M-MLV[H-]; *г* – схема гибридизации одноцепочечного ДНК-зонда, модифицированного 5'-биотин-TEG, с SL-интроном в составе Y-аутрона, и последующей очистки на магнитных частицах, покрытых стрептавидином.

и потому не могут служить прямым доказательством протекания транс-сплайсинга.

В данных mRNA-seq, полученных ранее, были выявлены последовательности исходной SL-РНК (Ershov et al., 2019), на основе которых разработаны соответствующие праймеры и зонды. Оптимизацию условий обогащения по Y-аутронам в методах SLi-BC и SLi-RT контролировали методом ПЦР-РВ с праймерами к аутрону и экзонам гена *mtse*, транс-сплайсируемый продукт которого не является полицистронным и характеризуется высоким уровнем экспрессии (см. Прил. 1). Обогащение рассчитывалось как разность ΔCt (аутрон, экзон) между контрольной и селективной кДНК. В случае препаратов кДНК, использованных для приготовления библиотек SLi-BC и SLi-RT, наблюдалось соответственно 4- и 100-кратное обогащение по участку аутрона относительно экзона гена *mtse* (см. Прил. 1, Б).

Из полученных различными методами препаратов РНК или кДНК были приготовлены библиотеки RNA-seq и далее секвенированы с низким покрытием, позволяющим получить приемлемое представление о базовых характеристиках обогащения или обеднения целевыми фрагментами.

Представленность рРНК в библиотеках

Для сравнения использованных методов со стандартным методом mRNA-seq, полученные данные секвенирования пяти библиотек были картированы на референсный геном, дополненный последовательностью повтора рРНК. Поскольку данный повтор имел достаточно высокое покрытие во всех библиотеках, он оказался полезен не только для оценки эффективности обеднения по рРНК, но и

для оценки цепь-специфичности полученных библиотек (рис. 2). Из последних только TerminatorEho имела относительно низкую степень цепь-специфичности (68 %). Интересно, что каждая из библиотек характеризовалась своими особенностями профиля покрытия рРНК. Стоит упомянуть, что для большого числа первичноротых 28S рРНК дополнительно процессируется по сайту «скрытого разрыва» с образованием примерно эквивалентных по размеру фрагментов 28S α и 28S β (Ishikawa, 1977). Такая особенность наблюдается и для *O. felineus*: разрыв в покрытии 28S рРНК (см. рис. 2) и одиночный пик рРНК на электрофореграмме (см. Прил. 2).

Наблюдаемое распределение немногочисленных прочтений рРНК библиотеки Zumo-Seq в сравнении с mRNA-seq (1.7 и 6.5 % прочтений соответственно) отражает крайне неэффективное «вычитающее» обеднение по наиболее перепредставленным участкам данного повтора.

Профиль библиотеки TerminatorEho свидетельствует об эффективной начальной ферментативной деградации 5'-районов 18S, 28S α и 28S β рРНК (см. рис. 2, отмечены зелеными линиями), которая, однако, стремительно угасает в определенных сайтах. Вероятно, это связано с чувствительностью фермента к сложной пространственной структуре субстрата либо с частым гидролизом рРНК в данных сайтах с образованием 5'-ОН концов, в отношении которых экзонуклеазная активность фермента отсутствует. В результате полученная библиотека TerminatorEho содержала около 60 % прочтений рРНК.

Методы прицельного обогащения SLi-BC и SLi-RT также продемонстрировали высокий уровень фоновой рРНК (63.1 и 20.2 %), причем библиотека SLi-RT содержала главным образом продукт нецелевого ревертирования с

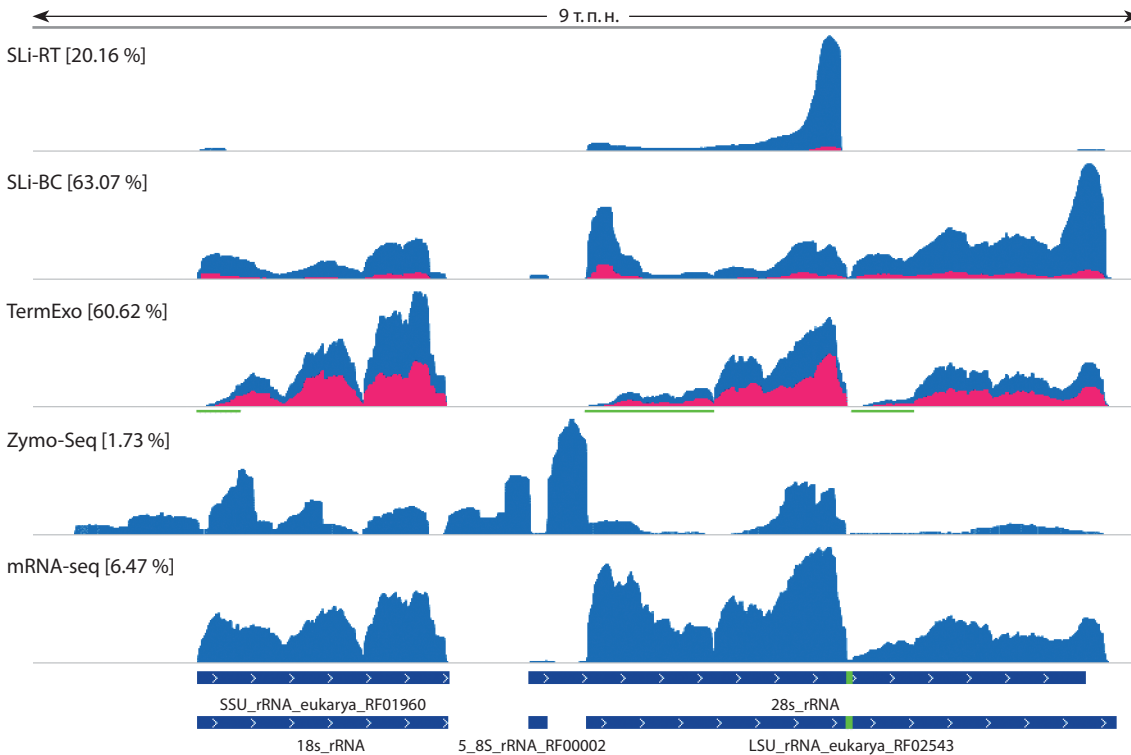


Рис. 2. Профили покрытия геномного повтора рРНК, содержащего 18S, 5.8S, 28S α и 28S β рДНК.

Профили покрытия прочтениями с антисмысловых и смысловых цепей кДНК выделены синим и красным цветами. В обозначениях библиотек в квадратных скобках указан процент прочтений, картированных на повтор рРНК. Темно-зелеными линиями отмечены 5'-участки рРНК, наиболее подверженные гидролизу экзонуклеазы Terminator. В нижнем ряду приведена разметка генов рРНК, предсказанных сервисами Rfam и RNAmmer 1.2, с указанием локуса скрытого разрыва (hidden break) 28S рРНК (зеленый).

3'-района 28S α рРНК. Поскольку 3'-участок 28S α рРНК потенциально может формировать стабильную шпильку (согласно предсказанной сервисом RNAfold вторичной структуре, Прил. 3), вероятно, подобная структура обуславливает праймирование обратной транскрипции без какого-либо участия праймера к SLi_r1.

Распределение категорий геномных элементов в библиотеках

Для картированных данных был проведен анализ распределения прочтений по различным взаимоисключающим (в указанном порядке) группам геномных элементов: повторам рРНК, мтДНК, экзонам, интронам, промоторным или аутронным участкам (1 и 5 т. п. н. выше 5'-конца предсказанного гена) и межгенным промежуткам. Подсчет велся отдельно для генов, сопряженных с высокоэффективным транс-сплайсингом (SL), и для прочих генов (TR). Как видно из результатов, представленных на рис. 3, а, в библиотеках TerminatorExo, SLi-BC и SLi-RT подавляющая доля прочтений представлена неинформативной фракцией рРНК и категорией межгенных участков. В библиотеках Zymo-Seq и mRNA-seq, напротив, более 60 % прочтений относятся к экзонам аннотированных генов.

Если исключить из рассмотрения группы рРНК и мтРНК (см. рис. 3, б), распределение прочтений по категориям геномных элементов, нормированное на суммарную длину последних, оказывается весьма схожим для библиотек mRNA-seq, SLi-BC, TerminatorExo и Zymo-Seq. При этом библиотека SLi-RT в целом воспроизводит слу-

чайное (равномерное) распределение прочтений в геноме. Очевидно, что SLi-RT содержит главным образом шум, источником которого, вероятно, послужила геномная ДНК, несмотря на идентичную обработку всех препаратов РНК ДНКазой. Надо отметить, что в данной библиотеке было обнаружено 227 (0.12 %) пар прочтений, содержащих в начале фрагмента праймер SLi_r1, примыкающий к последовательности адаптера. Именно такие последовательности, как ожидалось, должны были быть перепредставлены в библиотеке SLi-RT. И хотя они с большой вероятностью не относятся к истинным Y-аутронам, их наличие подтверждает работоспособность праймера в реакции ревертирования.

Предполагалось, что в случае эффективного обогащения библиотек по аутронам мы должны наблюдать значительное увеличение покрытия 5'-прилежащих районов транс-сплайсируемых генов (SL:up_1Kb и SL:up_5Kb) по отношению к промоторным районам рядовых генов (TR:up_1Kb и TR:up_5Kb). Такое обогащение действительно наблюдалось для библиотек SLi-BC, TerminatorExo и Zymo-Seq, но только в категории SL:up_5Kb (рис. 4, а). В SL:up_1Kb эти методы, напротив, демонстрировали обеднение по потенциальным аутронам. Обеднение обусловлено тем, что для многих рядовых генов не аннотированы их 5'-некодирующие районы (5'-UTRs), в результате чего в категории TR:up_1Kb оказалось большое число прочтений зрелых мРНК. Более показательным является сравнение с mRNA-seq, в котором все три упомянутых метода имели большее обогащение в обеих ка-

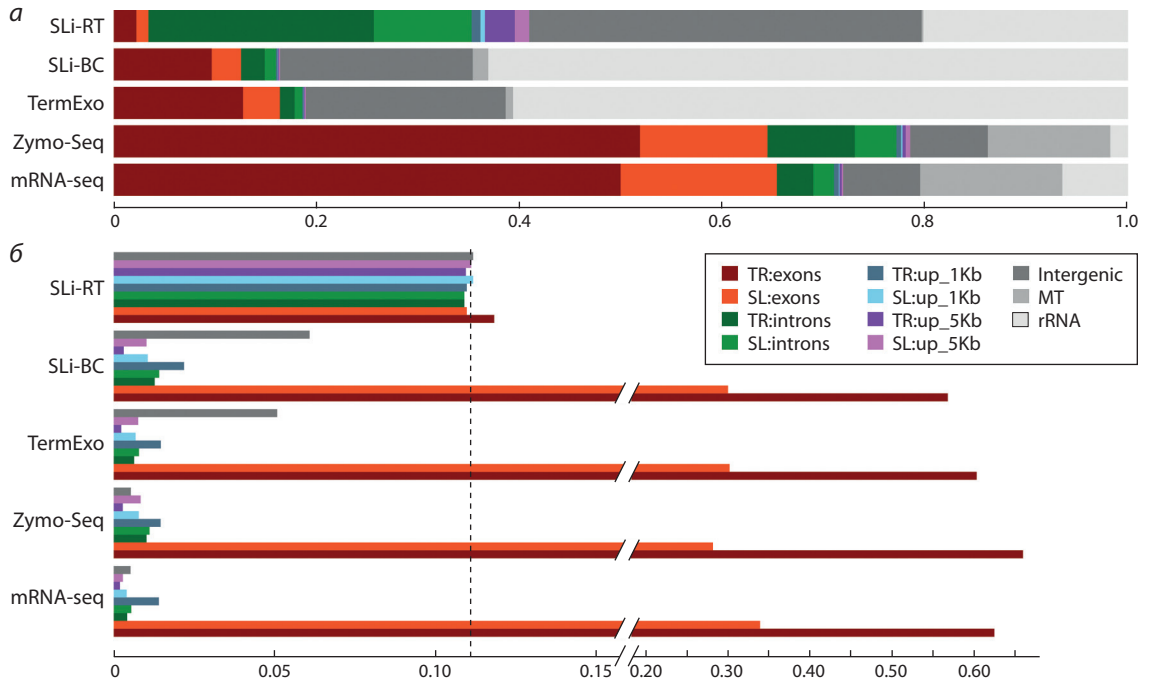


Рис. 3. Представленность различных категорий геномных элементов в данных секвенирования библиотек RNA-seq. *a* – долевое распределение картированных прочтений по 11 категориям геномных элементов; *б* – диаграмма, отражающая нормированное число прочтений в каждой из 9 геномных категорий (за исключением мтРНК и рРНК). Значения нормированы на суммарную длину категории и глубину секвенирования (аналогично метрике TPM). Штриховой линией обозначен ожидаемый уровень значений при условии равномерного покрытия генома.

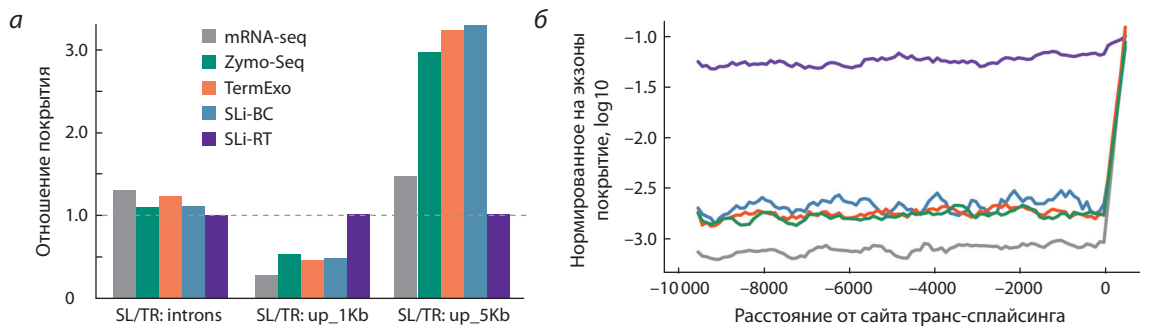


Рис. 4. Обогащение библиотек по ауτροнам генов. *a* – гистограмма обогащения по интронам и ауτροнам транс-сплайсируемых генов (SL) по сравнению с аналогичными районами обычных генов (TR). Покрытие нормировано на суммарную длину категорий; *б* – график суммарного покрытия пятью библиотеками потенциальных аутронов генов в геноме. По оси абсцисс указано расстояние от идентифицированного сайта транс-сплайсинга. Цветовые обозначения соответствуют легенде в секции *a*.

тегориях аутроновых участков. Поскольку коэффициент обогащения примерно одинаков для трех библиотек, то он отражает собственно долю аутронов в исходной тотальной РНК. Следовательно, метод SLi-BC не позволил преимущественно селективировать целевые последовательности Y-аутронов.

Таким образом, из всех опробованных методов только Zymo-Seq позволил значимо избавиться от неинформативной фракции рРНК и при этом сохранить исходную долю некодирующих транскриптов, включая аутроны. В то же время покрытие аутронов в отношении к экзонам в библиотеке Zymo-Seq остается незначительным (см. рис. 4, б).

Из полученных результатов можно заключить, что последовательности, соответствующие ауτροнам, включая непротрансценнированные новосинтезированные пре-РНК и интактные или частично деградировавшие Y-аутроны, представлены очень незначительной фракцией в пуле РНК *O. felineus*. По всей видимости, из них наименьшая доля приходится на интактные Y-аутроны, в связи с чем библиотеки SLi-BC и SLi-RT, нацеленные именно на интактную структуру, содержали лишь шум. Хотя механизм и скорость деградации Y-разветвленных продуктов транс-сплайсинга остаются неизвестными, предполагается, что они деградируют быстро (Lasda, Blumenthal, 2011). Так, в эксперименте *in vitro* на *C. elegans* были обнаружены все

интермедиаты транс-сплайсинга, кроме Y-разветвленных аутронов (Hannon et al., 1990).

Основным источником шума в методе SLi-BC стала неспецифическая сорбция тотальной РНК, а в методе SLi-RT – неспецифическое ревертирование рРНК без участия праймера к SL-интрону и минимальная примесь геномной ДНК. Детекция последней лишь в данной библиотеке опять же указывает на то, что в реакции обратной транскрипции практически отсутствовала целевая РНК-матрица, а именно Y-аутроны с ненарушенной разветвленной структурой.

Заключение

Таким образом, из рассмотренных методов RNA-seq, позволяющих идентифицировать последовательности аутронов транс-сплайсируемых транскриптов, наиболее перспективным оказался метод, основанный на вычитании рРНК ферментативной нормализацией кДНК (Zumo-Seq RiboFree). Неэффективность таргетных методов SLi-BC и SLi-RT обусловлена недетектируемыми количествами Y-разветвленных структур аутронов в пуле РНК.

Список литературы / References

Бээр С.А. Биология возбудителя описторхоза. М., 2005.

[Beer S.A. Biology of the Agent of Opisthorchiasis. Moscow, 2005. (in Russian)]

Bitton D.A., Rallis C., Jeffares D.C., Smith G.C., Chen Y.Y.C., Codlin S., Marguerat S., Bähler J. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. *Genome Res.* 2014;24(7):1169-1179. DOI 10.1101/gr.166819.113.

Boroni M., Sammeth M., Gava S.G., Jorge N.A.N., Macedo A.M., Machado C.R., Mourão M.M., Franco G.R. Landscape of the spliced leader trans-splicing mechanism in *Schistosoma mansoni*. *Sci. Rep.* 2018;8(1):3877. DOI 10.1038/s41598-018-22093-3.

Chen R.A.-J., Down T.A., Stempor P., Chen Q.B., Egelhofer T.A., Hillier L.W., Jeffers T.E., Ahringer J. The landscape of RNA polymerase II transcription initiation in *C. elegans* reveals promoter and enhancer architectures. *Genome Res.* 2013;23(8):1339-1347. DOI 10.1101/gr.153668.112.

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. DOI 10.1093/bioinformatics/bts635.

Döring J., Hurek T. Arm-specific cleavage and mutation during reverse transcription of 2',5'-branched RNA by Moloney murine leukemia virus reverse transcriptase. *Nucleic Acids Res.* 2017;45(7):3967-3984. DOI 10.1093/nar/gkx073.

Ershov N.I. Evaluation of various RNA-seq approaches for identification of outrons in the flatworm *Opisthorchis felineus* (Version 1.0.0). *Zenodo.* 2020. DOI 10.5281/zenodo.3901531.

Ershov N.I., Mordvinov V.A., Prokhortchouk E.B., Pakharukova M.Y., Gunbin K.V., Ustyantsev K., Genaev M.A., Blinov A.G., Mazur A., Boulygina E., Tsygankova S., Khrameeva E., Chekanov N., Fan G., Xiao A., Zhang H., Xu X., Yang H., Solovyev V., Lee S.M., Liu X., Afonnikov D.A., Skryabin K.G. New insights from *Opisthorchis felineus* genome: update on genomics of the epidemiologically important liver flukes. *BMC Genom.* 2019;20(1):399. DOI 10.1186/s12864-019-5752-8.

FAO/WHO [Food and Agriculture Organization of the United Nations/ World Health Organization]. Multicriteria-Based Ranking for Risk Management of Food-Borne Parasites. Microbiological Risk Assessment Series. No. 23. Rome, 2014.

Fedorova O.S., Fedotova M.M., Sokolova T.S., Golovach E.A., Kovshirina Y.V., Ageeva T.S., Kovshirina A.E., Kobayakova O.S., Ogorodova L.M., Odermatt P. *Opisthorchis felineus* infection prevalence in Western Siberia: a review of Russian literature. *Acta Trop.* 2018; 178:196-204. DOI 10.1016/j.actatropica.2017.11.018.

Hannon G.J., Maroney P.A., Denker J.A., Nilsen T.W. Trans splicing of nematode pre-messenger RNA *in vitro*. *Cell.* 1990;61(7):1247-1255. DOI 10.1016/0092-8674(90)90689-c.

Ishikawa H. Evolution of ribosomal RNA. *Comp. Biochem. Physiol.* 1977;58(1):1-7. DOI 10.1016/0305-0491(77)90116-X.

Kruesi W.S., Core L.J., Waters C.T., Lis J.T., Meyer B.J. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife.* 2013;2:e00808. DOI 10.7554/eLife.00808.

Lagesen K., Hallin P., Rødland E.A., Staerfeldt H.H., Rognes T., Ussery D.W. RNAmmr: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35(9):3100-3108. DOI 10.1093/nar/gkm160.

Lasda E.L., Blumenthal T. Trans-splicing. *Wiley Interdiscip. Rev. RNA.* 2011;2(3):417-434. DOI 10.1002/wrna.71.

Liao Y., Smyth G.K., Shi W. The R package *Rsubread* is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 2019;47(8):e47. DOI 10.1093/nar/gkz114.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10-12. DOI 10.14806/ej.17.1.200.

Murphy W.J., Watkins K.P., Agabian N. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans-splicing. *Cell.* 1986;47:517-525. DOI 10.1016/0092-8674(86)90616-1.

Nilsson D., Gunasekera K., Mani J., Osteras M., Farinelli L., Baerlocher L., Roditi I., Ochsenreiter T. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog.* 2010;6(8):e1001037. DOI 10.1371/journal.ppat.1001037.

Pakharukova M.Y., Mordvinov V.A. The liver fluke *Opisthorchis felineus*: biology, epidemiology, and carcinogenic potential. *Trans. R. Soc. Trop. Med. Hyg.* 2016;110:28-36. DOI 10.1093/trstmh/trv085.

Pakharukova M.Y., Zapparina O.G., Kapushchak Y.K., Baginskaya N.V., Mordvinov V.A. *Opisthorchis felineus* infection provokes time-dependent accumulation of oxidative hepatobiliary lesions in the injured hamster liver. *PLoS One.* 2019;14(5):e0216757. DOI 10.1371/journal.pone.0216757.

Saito T.L., Hashimoto S., Gu S.G., Morton J.J., Stadler M., Blumenthal T., Fire A., Morishita S. The transcription start site landscape of *C. elegans*. *Genome Res.* 2013;23(8):1348-1361. DOI 10.1101/gr.151571.112.

Sripa B., Kaewkes S., Sithithaworn P., Mairiang E., Laha T., Smout M., Pairojkul C., Bhudhisawasdi V., Tesana S., Thinkamrop B., Bethony J.M., Loukas A., Brindley P.J. Liver fluke induces cholangiocarcinoma. *PLoS Med.* 2007;4(7):e201. DOI 10.1371/journal.pmed.0040201.

Sutton R.E., Boothroyd J.C. Trypanosome trans-splicing utilizes 2'-5' ranes and a corresponding debranching activity. *EMBO J.* 1988;7(5):1431-1437.

ORCID ID

N.I. Ershov orcid.org/0000-0003-3423-3497

N.P. Bondar orcid.org/0000-0002-5602-5149

Благодарности. Работа поддержана Российским научным фондом, проект 18-74-00101.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 21.03.2020. После доработки 24.11.2020. Принята к публикации 24.11.2020.