

УДК 577.21;579.23"315

Конкуренция внутри интронов: сплайсинг побеждает полиаденилирование

М. В. Тихонов, П. Г. Георгиев*, О. Г. Максименко*

Институт биологии гена РАН, 119334, Москва, ул. Вавилова, 34/5

*E-mail: mog@genebiology.ru, georgiev_p@mail.ru

Поступила в редакцию 17.06.2013

РЕФЕРАТ Кэпирование, сплайсинг и полиаденилирование многих мРНК эукариот сопряжено с транскрипцией. Согласованность этих процессов обеспечивает правильное созревание РНК и создает разнообразие синтезируемых изоформ. Процессинг РНК представляет собой цепь событий, в которой окончание одного этапа связано с началом следующего. В этом контексте связь между сплайсингом и полиаденилированием считается важной для регуляции работы генов. Нами обнаружено, что скрытые сигналы полиаденилирования широко представлены в интронах *Drosophila melanogaster*. В результате анализа встречаемости генов, полностью расположенных в интронах других генов, установлено, что перекрывание в сигналах полиаденилирования встречается достаточно часто и затрагивает около 17% всех генов. Показано, что активность сигналов полиаденилирования, расположенных внутри интронов, подавлена: они функционируют, находясь в экзонах, но не в интронах. Как в транзитной репортерной системе, так и в модельных геномных локусах транскрипция не останавливается в интронах *in vivo*. При удалении 5'-сайта сплайсинга включается использование сигналов полиаденилирования в интронах. Согласно анализу транскриптома *Drosophila*, сигналы полиаденилирования внутри интронов используются очень редко, и, по всей видимости, данные события регулируются при помощи особых механизмов. Наши данные подтверждают, что транскрипционный аппарат игнорирует преждевременные сигналы полиаденилирования, расположенные в интроне.

КЛЮЧЕВЫЕ СЛОВА интрон, сигналы полиаденилирования, сплайсинг, терминация транскрипции, экзон.

ВВЕДЕНИЕ

В ходе созревания эукариотические мРНК подвергаются кэпированию, сплайсингу и полиаденилированию. Молекулярные механизмы, ответственные за эти изменения, связаны как друг с другом, так и с транскрипционным аппаратом [1–5]. Эта взаимосвязь подтверждается наличием факторов сплайсинга и компонентов комплекса расщепления и полиаденилирования, привлекаемых *in vitro* на РНК-полимеразу II (РНКП II) до начала транскрипции, а затем переносимых на новосинтезированную РНК [6, 7]. В качестве примера такой взаимосвязи можно привести процесс определения последнего экзона. Показано, что распознавание 3'-сайта сплайсинга (3'СС) последнего интрона сильно повышает эффективность нижележащего сигнала полиаденилирования (СПА) [8–10], а фактор U2AF, ассоциированный с 3'СС, увеличивает эффективность СПА путем прямого молекулярного контакта с поли(А)-полимеразой [11]. Кроме того, белковые компоненты U2 snRNP (малый ядерный рибонуклеопротеин), связывающиеся с 3'СС и расположенной рядом точкой ветвления, помогают усилить 3'-концевой про-

цессинг через взаимодействие с CPSF (фактор, стимулирующий расщепление и полиаденилирование) [12]. Фактор U2AF65 стимулирует процессинг 3'-конца за счет взаимодействия аргинин/серин-богатого участка с RS-подобным доменом субъединицы 59 кДа фактора расщепления I (CF Im) [13]. Вероятно, после расщепления РНК происходит полиаденилирование ее 3'-конца, после этого сплайсингу подвергается последний интрон [14], а полиаденилирование запускает освобождение РНК от полимеразы только после завершения сплайсинга [15]. Еще один важный факт – взаимодействие U1 snRNP, необходимого для определения 5'-сайта сплайсинга, с фактором I расщепления и полиаденилирования млекопитающих [16]. Эти данные свидетельствуют о взаимосвязи процессов транскрипции генов и созревания мРНК.

СПА представляют собой простые последовательности, мотивы встречаются не только в 3'-нетранслируемых областях генов, но и в других участках генома, включая 5'-нетранслируемые области генов [17]. Преждевременное полиаденилирование может привести к нарушению функции гена и, следовательно, должно быть предотвращено. Обнаруже-

но, что в клетках *Drosophila* и человека СПА, расположенные вблизи точек начала транскрипции, не функционируют [17]. СПА найдены также выше 5'СС, а точечная мутация в донорном сайте сплайсинга активирует вышележащий скрытый СПА [18]. У вируса папилломы крупного рогатого скота использование позднего СПА на более ранних стадиях инфекции предотвращается расположенным вблизи 5'СС. Распознавание 5'СС комплексом U1 snRNP блокирует поли(А)-полимеразу путем прямого с ней взаимодействия через субъединицу U1 snRNP массой 70 кДа [19]. Недавнее полногеномное исследование эффекта нокдауна U1 snRNP в клетках HeLa выявило случаи преждевременного расщепления и полиаденилирования многочисленных пре-мРНК на скрытых СПА, особенно часто в интронах вблизи старта транскрипции [20]. На основе блокирования полиаденилирования комплексом U1 snRNP предложен метод выключения генов путем привлечения U1 snRNP на целевую РНК, что приводит к ингибированию добавления поли(А)-хвоста и, как следствие, к деградации РНК в ядре [21, 22]. Количественный анализ вариантов мРНК показал, что функционирование СПА внутри интронов может отличаться в разных клетках [23]. В качестве примера можно привести данные о том, что уровень U1 snRNP определяет длину транскрипта и возможность использования преждевременных СПА внутри интронов или дистальных СПА в 3'-некодирующей области [24]. Приведенные выше данные указывают на возможность существования конкуренции между сплайсингом и полиаденилированием внутри интронов.

В нашей работе мы сконцентрировались на связи между сплайсингом и полиаденилированием с участием СПА, расположенных в интронах. Наши результаты подтверждают широкое распространение этих сигналов внутри интронов и показывают, что функционирование СПА, расположенных внутри интронов, заблокировано как в транзientной репортерной системе, так и в геномных локусах *in vivo*. В то же время делеция 5'СС восстанавливает функцию СПА. Анализ полного транскриптома различных культур клеток и разных стадий развития *Drosophila* свидетельствует о переключении между синтезируемыми изоформами в случае транскриптов с альтернативным выбором 3'-экзонов.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Предсказание СПА

Для предсказания сигналов полиаденилирования использовали программу PolyA_SVM [25], подходящую, как показано ранее, для предсказания СПА у *Drosophila* [17]. Программа осуществляет поиск с помощью

скользящего окна с оценкой соответствия анализируемых последовательностей 15 *cis*-элементам, обнаруженным в известных СПА человека [25]. Последовательности всех аннотированных интронов *D. melanogaster* доступны в FlyBase [26]. Вероятность того, что элемент представляет собой СПА, характеризуется значением E-value (чем меньше значение, тем выше вероятность). Получаемые результаты делили на три категории: «сигнал присутствует», «сигнал отсутствует» и «интроны длиной менее 120 п.н.».

Поиск генов, полностью расположенных в интроне другого гена

Аннотация генома *D. melanogaster* с координатами генов и интронов взята из FlyBase [26]. Если координата начала гена была больше координаты начала интрона, а координата конца – меньше, то ген обозначали как «вложенный» в интрон.

Конструирование векторов репортерной системы

Векторы бицистронной репортерной системы созданы на основе вектора pAc5.1/V5-His B (Invitrogen). Гены люциферазы светлячка (*Firefly*, далее *Fluc*) и медузы (*Renilla*, далее *Rluc*) были взяты из векторов pGL3Basic и pRL-CMV (Promega) соответственно. Сайт внутренней инициации трансляции из гена *reaper* амплифицирован с геномной ДНК и клонирован перед *Fluc*. Терминатор SV40 взят из вектора pAc5.1/V5-His B. Интрон и терминатор гена *yellow* вырезаны из фрагмента ДНК длиной 8 т.п.н., любезно предоставленного Р. Geys (Университет Айовы). СПА генов *por5* и *eIF6* амплифицированы с геномной ДНК. При создании искусственного интрона (AI) использовали олигонуклеотиды с необходимыми сайтами. В качестве последовательности, разделяющей донорный и акцепторный сайты сплайсинга, использовали кодирующую часть гена *lacZ*.

Культивирование клеток, трансфекция, выделение РНК, двойной люциферазный анализ

Культуру клеток *Drosophila* S2 выращивали на среде SFX (HyClone) при 25°C. Клетки трансфицировали с помощью реагента Cellfectin II (Invitrogen) по протоколу производителя. После трансфекции клетки выращивали в планшете в течение 24–48 ч.

РНК выделяли из трансфицированных клеток при помощи реагента TRI (Ambion) по протоколу производителя. Для разделения ядерной и цитоплазматической РНК клетки культуры S2 снимали с 100-мм чашек, промывали фосфатным буфером, осаждали и ресуспендировали в 100 мл буфера TD (0.8% NaCl, 0.028 M KCl, 0.01% Na₂HPO₄, 0.3% Трис-HCl; pH 7.4–7.5). В полученную суспензию добавляли 100 мкл TD с 1% NP-40 и ингибитором ри-

бонуклеаз SUPERase-In (Ambion), выдерживали в течение 5 мин на льду. Препарат центрифугировали, из супернатанта выделяли цитоплазматическую РНК, используя реагент TRI. Осадок, содержащий ядра, ресуспендировали в 200 мкл TD с 0.5% NP-40 и SUPERase-In, инкубировали на льду и повторно центрифугировали. Ядерную фракцию выделяли из осадка, используя реагент TRI.

Двойной люциферазный анализ проводили, используя набор реактивов Firefly & Renilla Luciferase Assay Kit (Biotium) по протоколу производителя.

Анализ РНК

РНК (5–20 мкг) для проведения нозерн-блот-анализа разделяли в 1% агарозном геле с формальдегидом и переносили на положительно заряженную нейлоновую мембрану (BrightStar-Plus, Ambion) на аппарате Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad) с последующей пришивкой ультрафиолетовым светом. Для синтеза зонда участка ДНК амплифицировали и встраивали под контроль T7-промотора. Мембрану гибридизовали с синтезированными *in vitro* РНК-зондами (наборы реактивов MEGAshortscript и MAXIscript, Ambion) с включенным biotin-16-UTP (Roche) и детектировали, используя Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific).

Количественную ПЦР в реальном времени проводили с продукта, получаемого при обратной транскрипции РНК-проб. РНК обрабатывали 2 ед. Turbo DNase I (Ambion) в течение 30 мин при 37°C для удаления примеси геномной ДНК. В реакции обратной транскрипции использовали фермент ArrayScript reverse transcriptase (Ambion), используя 5 мкг РНК-матрицы и смесь шестинуклеотидных праймеров. Количественный анализ проводили методом ПЦР в реальном времени на приборе CFX96 Thermal Cycler (Bio-Rad). Как минимум, три независимых реакции проводили одновременно для каждой точки, для каждого из трех независимо полученных образцов РНК. Относительные количества РНК определяли по циклу перехода через пороговое значение.

Анализ паттерна экспрессии изоформ на основе данных секвенирования РНК

Все процедуры проводили с помощью программ на языке Java. Была использована аннотация генома из FlyBase (версия 5.40) для поиска транскриптов, оканчивающихся в интроне. Отбирали транскрипты, у которых последний экзон начинался выше донорного сайта интрона, в котором они заканчивались. Гены, которые пересекались с другими генами, были исключены из рассмотрения. Далее анализировали гены, у которых было только два варианта транс-

криптов: первый подвергался сплайсингу, второй – расщеплению/полиаденилированию в интроне. SAM-файлы с данными секвенирования РНК 30 стадий развития и четырех линий клеток получены из базы modENCODE. Чтения (риды) были соотнесены со структурой гена. Среднее значение плотности чтения 3'-экзона использовали для оценки относительного уровня сплайсированной формы. Уровень изоформы, терминированной в интроне, вычислен как разность между средним значением для участка между 5'СС и концом транскрипта и средним значением участка между концом транскрипта и 3'СС (вероятно, соответствует несплайсированной РНК). Для визуализации паттерна экспрессии каждой изоформы и доли терминированной в интроне изоформы к общей РНК гена создана тепловая карта.

РЕЗУЛЬТАТЫ

СПА широко представлены в интронах *Drosophila*

Как отмечалось выше, по-видимому, СПА широко распространены в геноме, а их преждевременное использование может привести к нарушению функции генов. Следовательно, должны существовать механизмы, предотвращающие полиаденилирование в неподходящем месте. Ранее было показано блокирование СПА в 5'-нетранслируемой области генов [17]. Появлению неподходящих сигналов в кодирующей части препятствует давление отбора, интроны же, напротив, подвержены изменениям, что может привести к возникновению преждевременных СПА. Мы проверили, насколько широко такие сигналы представлены в интронах *Drosophila*. С этой целью использовали программу PolyA_SVM, предназначенную для анализа и предсказания СПА [25]. Эта программа, как показано ранее, подходит для предсказания СПА у *Drosophila* [17].

Нуклеотидные последовательности 58 594 интронов *Drosophila* получены из FlyBase, версия 5.34 [26]. Приблизительно 55% этих последовательностей были короче 120 п.н., минимальной длины, необходимой для запуска программы (*рис. 1А*); две трети остальных (около 30% всей выборки) содержали один или более СПА. Таким образом, потенциальные СПА широко представлены в интронах. Существует два возможных состояния таких сигналов: они могут быть молчащими или активными. В первом случае сигнал никак не влияет на транскрипцию, проходящую через него, и последующий сплайсинг. Во втором варианте наблюдается преждевременная терминация транскрипции.

В этом контексте интересно изучить пары генов, в которых один «вложенный» ген полностью находится в интроне другого. В этом случае СПА внутреннего

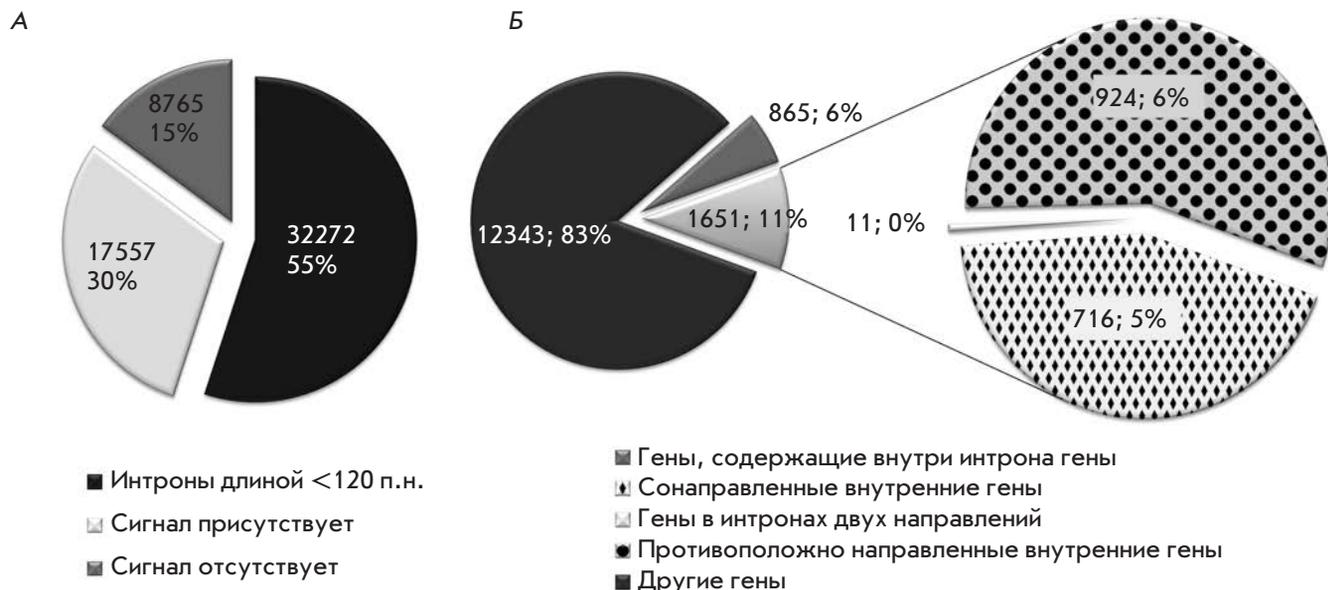


Рис. 1. СПА широко представлены в интронах *Drosophila*. А – круговая диаграмма, отображающая доли интронов, содержащих и не содержащих СПА, и долю коротких интронов. Предсказано программой PolyA_SVM. Б – круговая диаграмма, отображающая доли генов, содержащих другие гены внутри интрона, и доли генов, вложенных в интроны, разделенные на группы по их принадлежности тяжю ДНК

вложенного гена не должен влиять на транскрипцию внешнего. Главное отличие между транскриптами вложенного и внешнего генов в момент прохождения через СПА последнего состоит в том, что транскрипт внешнего гена имеет 5'СС. Мы проанализировали встречаемость таких пар генов, используя данные FlyBase [26] (рис. 1Б). Координаты генов и их интронов использовали для определения случаев, когда один ген полностью располагается внутри интрона другого. Найдено 865 внешних генов, содержащих 1651 вложенный ген. Из них 727 имели то же направление, что и внешний ген, поэтому их регуляторные элементы могут влиять на экспрессию внешнего гена. Таким образом, мы показали, что СПА широко представлены в последовательностях интронов, а возможность влияния СПА внутреннего гена на экспрессию внешнего – случай достаточно общий, затрагивающий около 17% генов.

Экспериментальное подтверждение функционирования СПА в экзонах, но не в интронах

Как показано выше, транскрипционная машина часто встречается с преждевременным СПА внутри интронов. Возможность функционирования этих последовательностей можно показать на паре генов – вложенного и внешнего. С целью проверки влияния СПА вложенного гена на транскрипцию внешнего мы про-

вели количественную ОТ-ПЦР для двух пар генов, *ytr-eIF6* и *xl6-nop5* (рис. 2). Основным критерием выбора этих генов был высокий уровень их экспрессии в культуре клеток S2. Профили экспрессии получены из базы данных modENCODE [27, 28]. Проводили анализ для ядерной и цитоплазматической РНК, используя зонды к участкам, отмеченным на рис. 2. В то время как цитоплазматическая фракция содержала полностью созревшую РНК, в состав ядерной фракции входили также интермедиаты процессинга и продукты распада. Это позволяло выявлять не полностью процессированные и не транспортированные в цитоплазму транскрипты.

С использованием количественной ОТ-ПЦР в цитоплазматической фракции выявлен высокий уровень только участков, соответствующих экзонам анализируемых генов (рис. 2; зонды 1, 4, 6), содержание участков, соответствующих интронам, было на фоновом уровне (рис. 2; зонды 3, 5). Однако в ядерной фракции детектировались интронные участки, что указывает на присутствие длинного несплайсированного транскрипта генов *ytr* (рис. 2А) или *xl6* (рис. 2Б). Отсутствие значительных различий в содержании участков интронов (точки 3 и 5) и последнего экзона (точка 6) свидетельствует о том, что внутри интрона не происходит преждевременного расщепления/полиаденилирования транскриптов генов *ytr* и *xl6*, и транскрипты *eIF6* или *nop5* процес-

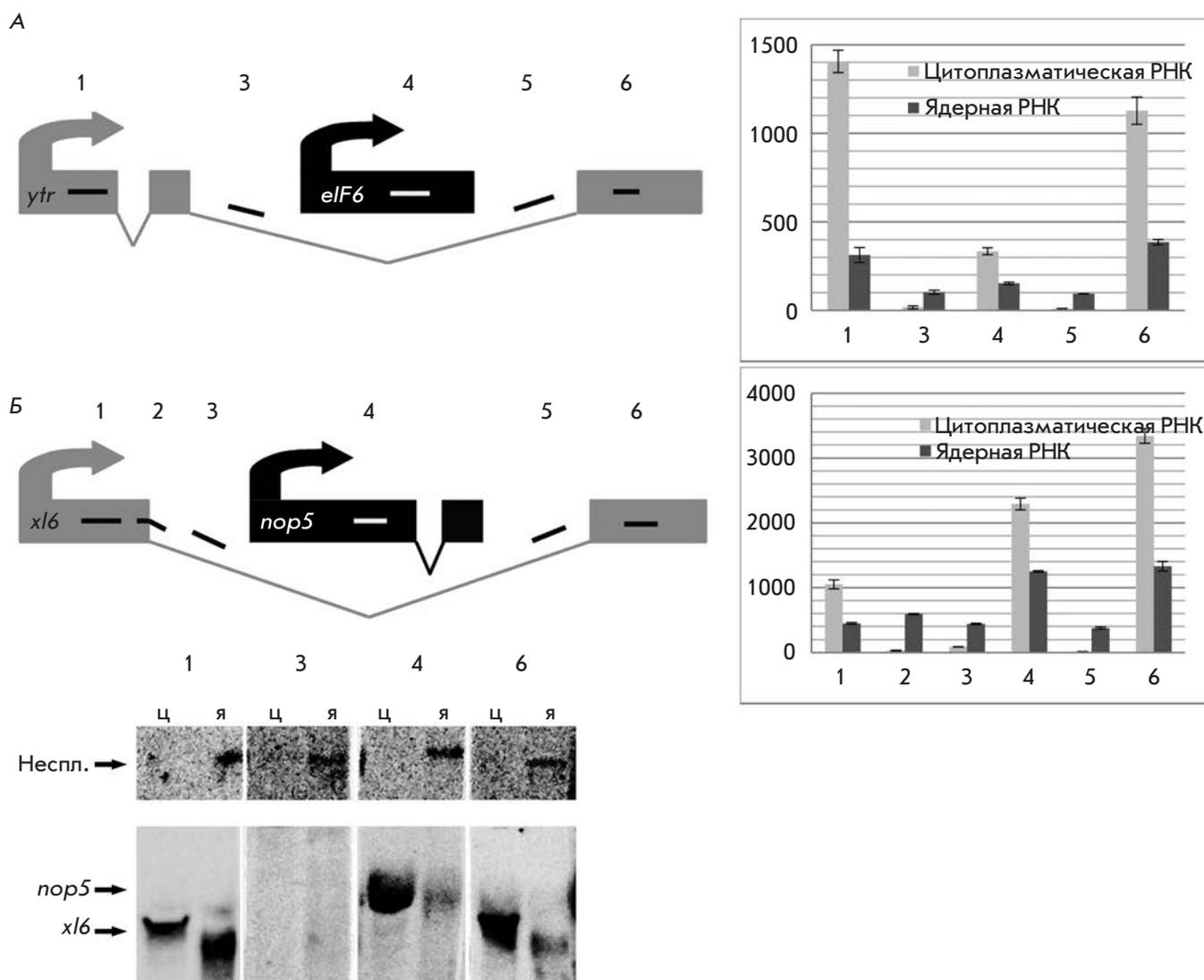


Рис. 2. СПА функционируют в экзонах, но не в интронах. На схемах эксперимента с парами генов *ytr-elf6* (А) и *x16-por5* (Б) экзоны и интроны изображены как прямоугольники и уголки (V) соответственно; короткие номерованные линии отмечают участки, узнаваемые зондами. Гистограммы показывают уровень ядерной и цитоплазматической РНК в этих участках. Представлены средние значения и стандартное отклонение ($n = 3$). Результат нозерн-блот-анализа ядерной (я) и цитоплазматической (ц) фракций РНК из клеток S2, гибридизованных с зондами, узнающими участки из генов *x16-por5*, представлен внизу (зонды из участков 1, 3, 4, 6)

сируются с использованием собственных СПА, т.е. в каждом из транскриптов срабатывал свой собственный сигнал полиаденилирования.

Параллельно мы провели нозерн-блот-анализ пары генов *x16-por5*, используя зонды к участкам из экзонов и одного интрона (рис. 2Б). В ядерной фракции детектируется (на пределе чувствительности) длинная РНК, соответствующая несплайсированному продукту *x16*. В обеих фракциях зонды, специфичные к экзонам, давали сигналы, соответствующие процессированным транскриптам *x16* (зонды 1 и 6) и *por5* (зонд 4).

Нужно отметить, что нозерн-блот-анализ не выявил транскриптов, расщепляемых на СПА гена *por5*, расположенного в интроне транскрипта гена *x16*. Транскрипт гена *x16* расщепляется/полиаденилируется только с использованием собственного СПА, локализованного в последнем экзоне.

Суммируя результаты нозерн-блот-анализа и количественной ОТ-ПЦР, можно заключить, что при транскрипции внешнего гена расположенный в интроне СПА игнорируется транскрипционной машиной, в то время как при экспрессии гена, вложенного в интрон, этот

СПА успешно используется, поскольку в данном случае он представляет собой экзонный сигнал.

СПА, встроенные в интрон, функционально неактивны

Анализ эндогенных событий расщепления/полиаденилирования характеризует функционирование генов на своем месте в геноме и не предоставляет достаточной свободы для внесения изменений в процесс. Поэтому мы решили использовать репортерную систему для анализа функционирования СПА на клеточной культуре S2 *Drosophila*. Бицистронная система содержит гены двух люцифераз: Rluc (*Renilla luciferase*) и Fluc (*Firefly luciferase*), под общим промотором гена актина 5С. Последовательность сайта внутренней инициации трансляции (IRES) из гена *reaper* [29] встроена перед геном *Fluc* (рис. 3).

Анализируемые последовательности встраивали перед IRES. Мы ожидали, что, если СПА функционирует, то будет продуцироваться короткая моноцистронная РНК с *Rluc*. И, напротив, если сигнал не работает или работает слабо, то длинный бицистронный продукт будет транскрибироваться до СПА SV40, расположенного после *Fluc*. Плазмидами с этими конструкциями трансфицировали клетки культуры S2 *Drosophila*. Анализ проводили спустя 24–48 ч после трансфекции. Отношение количества длинной бицистронной мРНК к общему количеству мРНК, синтезируемой с конструкции, оценивали, сравнивая интенсивность хемилюминесценции Fluc и Rluc.

В первом наборе конструкций, предназначенном для определения базового уровня расщепления/полиаденилирования на разных СПА, после первого цистрона были клонированы поздний СПА вируса SV40 (tSV40) и СПА генов *nor5* (tnor5), *eIF6* (teIF6), *yellow* (ty). Базовую конструкцию без вставок и конструкцию с линкерной последовательностью (без СПА) такой же длины, как и тестируемые СПА, использовали в качестве контроля. Все перечисленные выше СПА снижали соотношение активностей Fluc/Rluc, указывая на то, что транскрипт расщеплялся/полиаденилировался с использованием СПА, встроенного после первого цистрона (рис. 3А). Для базовой и ty-конструкций мы провели нозерн-блот-анализ с зондами, специфичными к генам *Rluc* и *Fluc*, и подтвердили, что СПА ty приводит к появлению короткого транскрипта, детектируемого зондом R (выявляется полоса размером 1.7 т.н. против 3.8 т.н. в базовой конструкции (рис. 3)).

Второй набор конструкций был создан для определения активности СПА в интроне гена *yellow* *Drosophila*. Получены векторы с интроном и СПА ty, встроенным в интрон. Двойной люциферазный анализ показал, что вставка интрона снижает соотношение Fluc/Rluc по сравнению с базовой конструкцией

(рис. 3А). Мы объясняем это изменением эффективности IRES-зависимой инициации трансляции. Тем не менее вставка СПА ty в интрон не изменила отношения Fluc/Rluc по сравнению с отношением при использовании конструкции с интроном без вставки. Проведен нозерн-блот-анализ данных конструкций с зондами, специфичными к генам обеих люцифераз (рис. 3). Зонд F выявлял длинный транскрипт с двумя цистронами. Полосы гибридизации выявлены в дорожках, соответствующих базовой и двум интронсодержащим конструкциям. В случае интронсодержащих конструкций длина детектируемой РНК соответствовала сплайсированному варианту (4.2 т.н.). Зонд R выявлял все варианты РНК, синтезируемые с конструкций. В случае конструкции ty детектировалась только короткая форма (1.7 т.н.). В дорожках, соответствующих базовой и интронсодержащей конструкциям, зонд R обнаруживал такие же полосы, как и зонд F. Примечательно, что короткие формы транскриптов не обнаружены при анализе конструкций с СПА ty, встроенным в интрон.

Таким образом, мы показали, что транскрипт расщепляется/полиаденилируется во всех случаях, когда сигнал полиаденилирования находится в экзонах. СПА, расположенный в интроне гена *yellow*, не прерывал транскрипцию, и в этом случае детектировалась только сплайсированная форма.

Интрон гена *yellow* имеет большую длину, он может содержать неизвестные регуляторные элементы, потенциально способные влиять на транскрипцию. Поэтому мы решили создать искусственный интрон (AI), в котором нет каких-либо регуляторных элементов за исключением минимального набора сигналов сплайсинга: донорного и акцепторного сайтов, точки ветвления, поли(Т/С)-участка. Фрагмент кодирующей части гена *lacZ* был использован для разделения сигналов сплайсинга. Взяв за основу базовую бицистронную конструкцию, мы получили ее производные, содержащие разные СПА или линкерные последовательности внутри AI. С использованием этой системы проверили четыре СПА: ty, tSV40, tnor5 и teIF6 (рис. 3Б). В качестве контроля ty, tnor5 и teIF6 были клонированы также в обратной ориентации. В связи с тем, что СПА SV40 работает в обеих ориентациях, в качестве контроля взяли линкерную последовательность из гена *lacZ* такой же длины. В конструкциях, содержащих СПА в прямой ориентации, соотношение Fluc/Rluc не менялось значительно по сравнению с базовой конструкцией, конструкциями с СПА в обратной ориентации и конструкцией с линкером из *lacZ* (рис. 3Б). Таким образом, не обнаружено ни одного случая функционирования СПА в AI так же, как в модели с интроном гена *yellow* и в опытах с парами генов в геномном локусе.

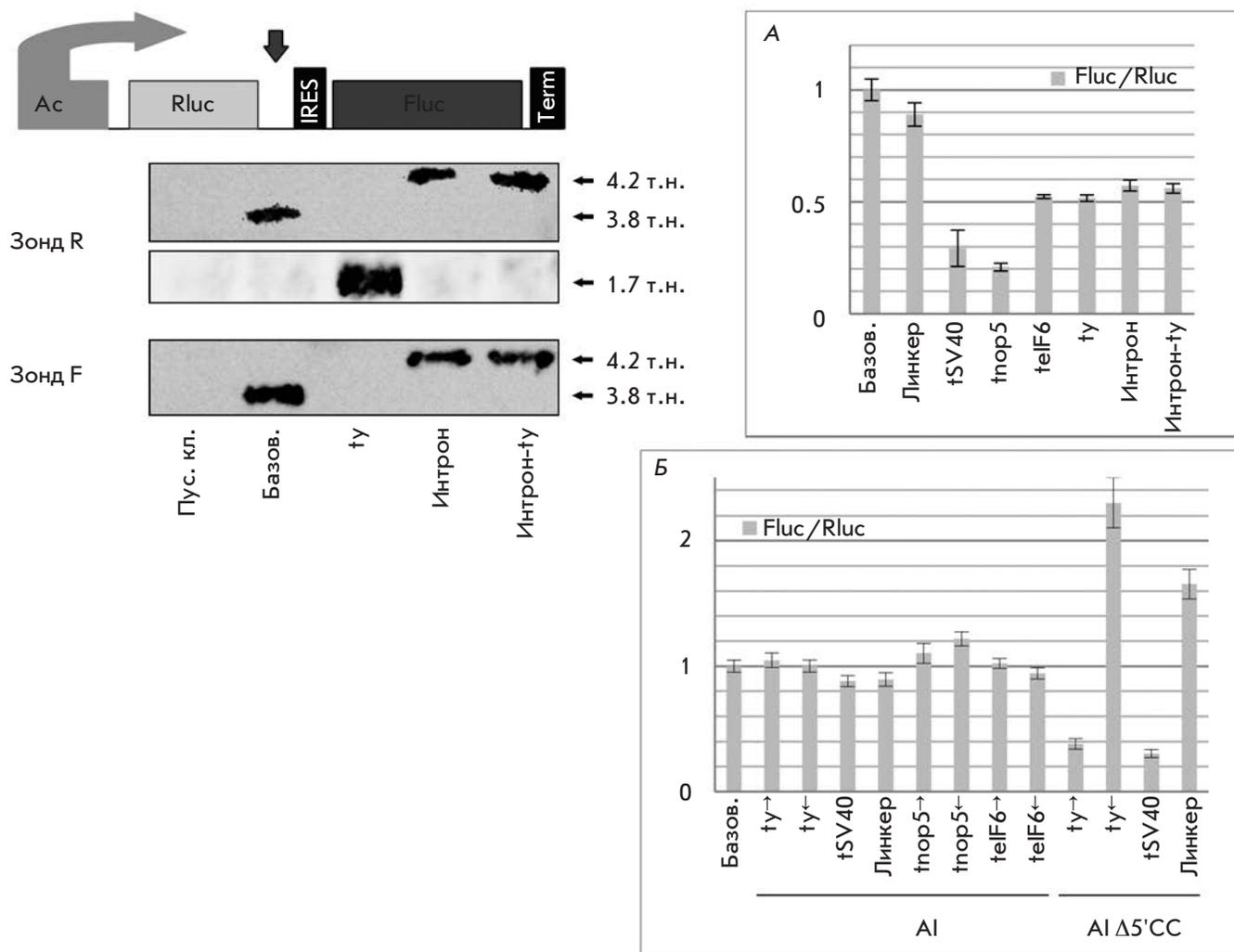


Рис. 3. Репортерная система подтверждает, что СПА, вставленные в интрон, функционально заблокированы. Бицистронная система основана на двух люциферазах: Rluc и Fluc, гены которых расположены под общим промотором гена актина 5С. IRES располагается между генами двух люцифераз. Стрелка указывает место вставки СПА вируса SV40 (tSV40) и СПА из генов *por5* (tnop5), *eIF6* (telF6) и *yellow* (ty). Интрон гена *yellow* и этот же интрон с СПА ty встроены в это же положение. Отношение Fluc/Rluc в случае этих конструкций показано на гистограмме А. Результаты нозерн-блот-анализа нетрансфицированных (пус. кл.) и трансфицированных клеток S2 с пробами к обеим люциферазам подтверждают результаты люциферазного анализа. Гистограмма Б показывает соотношение Fluc/Rluc в случае конструкций с полным AI и с делетированным донорным сайтом (AI Δ5'CC). На гистограммах представлены средние значения и стандартное отклонение, $n = 5$ (А) и 3 (Б)

Для подтверждения блокирования расщепления/полиаденилирования в интронах мы провели аналогичные опыты с конструкциями, содержащими ty и tSV40 внутри AI, но без донорного сайта. Как ожидалось, делеция донорного сайта должна превратить AI в продолжение экзона и активировать СПА. Как показано на рис. 3Б, отношение Fluc/Rluc, обеспечиваемое конструкциями, содержащими СПА, снижается. Следовательно, в отсутствие донорного сайта короткий моноцистронный транскрипт обра-

зуется с использованием первого встроенного СПА. Описана и обратная ситуация [30, 31], когда мутация приводит к появлению функционального 5'CC, а его распознавание компонентом сплайсосомы U1 snRNP вызывает подавление формирования 3'-конца мРНК.

Использование СПА внутри интронов – редкое и, по всей видимости, индуцибельное явление

Мы установили, что расщепление/полиаденилирование не происходит в пределах интро-

на – как при транскрипции геномных локусов, так и в плазмидной репортерной системе. Тем не менее известно, что если альтернативный 3'-экзон не включается в транскрипт, то транскрипция должна терминироваться в интроне (рис. 4А,Б). Используя аннотацию генома из FlyBase [26], мы проверили как часто транскрипт заканчивается в интроне другой изоформы и обнаружили 403 гена, организованных таким образом. Паттерн экспрессии этих изоформ мы анализировали с использованием данных секвенирования РНК для различных линий клеток и стадий развития, доступных в проекте modENCODE [27, 28]. Мы исключили 170 генов, которые перекрывались с другими генами, так как в подобной ситуации невозможно определить какой ген транскрибируется. Затем мы выбрали 70 генов, имеющих только два варианта транскриптов: первый подвергался сплайсингу, а второй заканчивался в интроне (рис. 4А). Долю транскрипта, оканчивающегося в интроне, оценивали как отношение его уровня к суммарному уровню обоих транскриптов.

Исходя из рассчитанных уровней транскриптов и их соотношения, мы разбили выбранные гены на несколько групп. Первая группа, наибольшая, включала 20 генов, у которых не детектировался транскрипт, оканчивающийся в интроне, или его уровень был близок к уровню «шума». Возможно, у генов этой категории транскрипты, расщепленные/полиаденилированные с использованием СПА в интроне, не образуются или образуются в очень маленьких количествах. Во второй группе, состоящей из шести генов, продукт сплайсинга не обнаружен, что ставит под сомнение наличие 3'-экзона. Третья группа, включающая 16 генов, характеризуется наличием обеих изоформ, причем переключение на синтез одной или другой является, по всей видимости, индуцируемым процессом (рис. 4В). Уровень экспрессии каждой изоформы и соотношение между ними изменяются в ходе развития или различаются в разных клеточных линиях. Кроме того, мы выделили группу генов с предположительно ошибочной аннотацией. Такое заключение основано на несоответствии границ экзонов и чтений при секвенировании РНК.

Таким образом, нами показано, что СПА, расположенные внутри интронов, используются в редких случаях. Согласно данным аннотации, только 403 гена имеют, возможно, транскрипты, оканчивающиеся в интроне. В действительности у 20 из 70 проанализированных генов транскрипт, оканчивающийся в интроне, не детектируется или его уровень близок к уровню «шума». Только у 16 генов обнаружили обе изоформы, уровень и соотношение которых менялись в ходе развития или отличались в разных линиях клеток.

ОБСУЖДЕНИЕ

В нашей работе мы исходили из предположения, согласно которому использование СПА внутри интронов не допускается. Ранее был описан феномен функционирования скрытых СПА внутри интрона при нокдауне U1 snRNP в клетках культуры HeLa [20]. Мы проверили частоту встречаемости СПА в интронах генов *Drosophila* и обнаружили, что мотивы СПА широко представлены в интронах (встречаются более чем в 30% интронов). Мы обратили внимание на те случаи, когда один ген расположен в интроне другого так, что при транскрипции последнего аппарат транскрипции проходит сквозь СПА внутреннего гена. Нами показано, что такая архитектура генов не приводит к перекрестному использованию СПА, т.е. транскрипция внешнего гена не ведет к использованию расположенного в его интроне СПА вложенного гена. Кроме того, мы не наблюдали образования транскриптов, процессируемых на интронных СПА, при экспрессии плазмидных репортерных систем как с интроном гена *yellow*, так и с искусственным интроном. Примечательно, что делеция донорного сайта сплайсинга в репортерной системе приводит к активации интронного СПА. Анализ полного транскриптома *Drosophila* показал, что изоформы, образующиеся при использовании СПА внутри интронов, редко экспрессируются, а соотношение таких и сплайсированных форм изменяется в ходе развития.

Суммируя наши и полученные ранее [19, 20] данные, можно сделать вывод, что транскрипция обычно не прерывается на СПА, расположенных внутри интронов. Исключения из этого правила редки и указывают на то, что для активации СПА внутри интронов необходимы дополнительные условия. Примечателен тот факт, что среди генов с альтернативным включением 3'-экзона мы обнаружили только 16 генов, образующих две изоформы транскриптов, – сплайсированную и процессированную с использованием СПА внутри интрона. По нашему мнению, в данном случае происходит индуцированное переключение между изоформами.

Для объяснения данного феномена можно предложить две модели. Первая модель – модель «анти-терминации» – основана на данных о взаимодействии между факторами сплайсинга и комплексом расщепления/полиаденилирования [14, 19, 20]. Возможно, факторы сплайсинга и полиаденилирования конкурируют между собой и первые доминируют над вторыми (рис. 4Г). Это может происходить путем прямого белок-белкового взаимодействия, например, при конкурентном связывании факторов сплайсинга и полиаденилирования с С-концевым доменом РНКП II или при инактивации компонентов комплекса расщепления/полиаденилирования

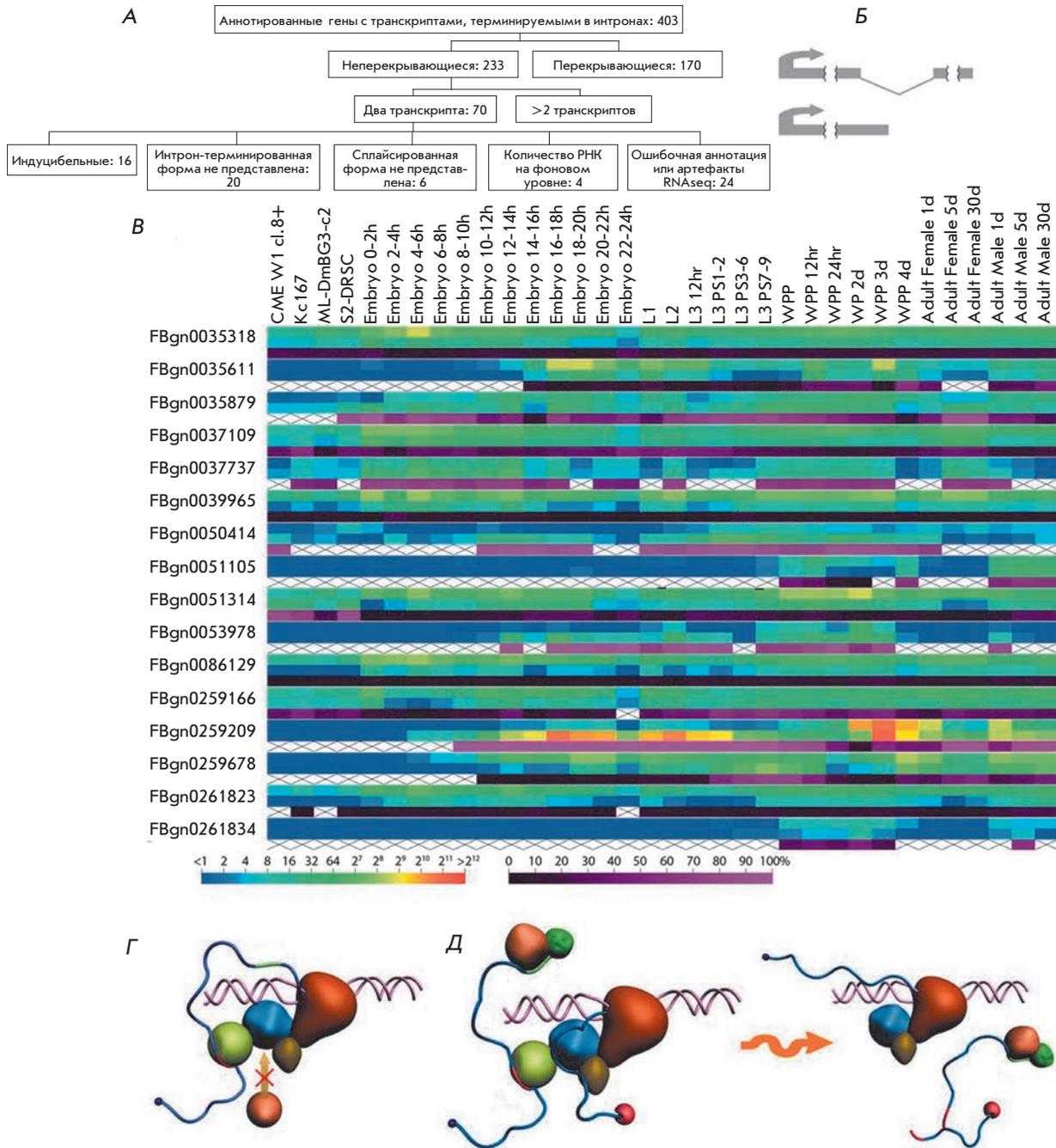


Рис. 4. Использование СПА внутри интронов – событие редкое и, по всей видимости, индуцибельное. **А** – схема поиска генов с изоформой расщепляемой/полиаденилируемой в интроне. **Б** – схема отобранных генов с двумя транскриптами, где одна изоформа расщепляется/полиаденилируется в интроне другой. **В** – тепловая карта, иллюстрирующая паттерны экспрессии двух изоформ 16 генов в четырех линиях клеток и на 30 стадиях развития. Верхняя линия – уровень сплайсированной формы каждого гена, средняя – терминированной в интроне, нижняя – отношение уровня транскрипта, терминированного в интроне, к общей мРНК гена. Цветовая шкала внизу характеризует уровни экспрессии изоформ в величинах плотности прочтения при секвенировании РНК (слева) и долю терминированной в интроне изоформы (справа). **Г**–**Д** – две возможные модели. **Г** – в модели «антитерминации» СПА недоступен для комплекса расщепления/полиаденилирования из-за конкурентного связывания факторов сплайсинга и расщепления/полиаденилирования с элонгирующим комплексом. Сплайсинг побеждает полиаденилирование в связи с его более ранним вовлечением в процесс. **Д** – в «кинетической» модели СПА доступен, но движущаяся полимеразы достигает 3'СС достаточно быстро, чтобы произошла реакция сплайсинга. Внесенные в транскрипт разрывы остаются в вырезаемом интермедиате и не приводят к терминированию транскрипции

факторами сплайсинга. Например, snRNP ингибирует поли(А)-полимеразу через прямое взаимодействие между U1 70K и поли(А)-полимеразой [19]. После того как U1 snRNP распознает донорный сайт, элонгационный комплекс РНКП II может стать недоступным для факторов расщепления/полиаденилирования. Сплайсинг побеждает полиаденилирование, так как компоненты сплайсосомы собираются в функциональный комплекс на более ранних этапах транскрипции. В то же время использование СПА внутри интрона может быть вызвано, в некоторых случаях, общими механизмами, используемыми при регуляции альтернативного сплайсинга, такими, как маскирование донорного сайта регуляторными белками или связыванием комплементарной РНК, изменением организации хроматина или, как описано в [24], уровнем U1 snRNP.

Вторая модель, «кинетическая», основана на предположении, что РНКП II продолжает свое движение после встречи с СПА [32, 33] и успешно достигает акцепторного сайта, вызывая прохождение реакции сплайсинга с образованием структуры «лассо» (рис. 4Д). В данном варианте расщепление и полиаденилирование могут происходить, но они не влияют на созревание мРНК, так как интермедиат «лассо» вырезается из транскрипта. Показано [34], что экзоны, окружающие интрон, созданный для котранскрипционного расщепления (CoTC), точно и эффективно сплайсируются вместе. Мы можем предположить, что СПА внутри интрона работает таким же образом. Эта модель предполагает, что переключение между использованием и пропуском СПА

в интроне может регулироваться не только маскированием донорного сайта, но и изменением скорости элонгации РНКП II, зависящей от метилирования CpG и организации хроматина. В общем случае РНКП II движется с высокой скоростью, достаточной для достижения акцепторного сайта, перед тем как быть вытесненной эндонуклеазами, которые привлекаются на РНК после расщепления с участием СПА внутри интрона. Если РНКП II остановится или замедлится ее движение, то транскрипция будет терминироваться, что в результате приведет к образованию короткой изоформы мРНК.

Другая возможная судьба продуктов, образующихся с использованием СПА внутри интрона, – деградация. Однако мы не выявили ни одного продукта, полученного с использованием СПА внутри интрона. Поэтому, если данное событие и возможно, то деградация должна происходить очень быстро или количество такой РНК мало. ●

Мы благодарны Н.А. Горголюку за помощь в подготовке рукописи. Экспериментальная часть выполнена на оборудовании Центра коллективного пользования ИБГ РАН (ГК № 16.552.11.7067).

Работа поддержана программой «Молекулярная и клеточная биология» Президиума РАН, РФФИ (грант № 10.04.00341-а) (П.Г.), ФЦП «Кадры» № 8103 (П.Г.), стипендией Президента РФ СП-1960.2012.4 (О.М.), Министерством образования и науки РФ (П1165) (О.М.) и грантом ОПТЭК (О.М.).

СПИСОК ЛИТЕРАТУРЫ

1. Auboeuf D., Dowhan D.H., Dutertre M., Martin N., Berget S.M., O'Malley B.W. // *Mol. Cell Biol.* 2005. V. 25. P. 5307–5316.
2. Bentley D.L. // *Curr. Opin. Cell Biol.* 2005. V. 17. P. 251–256.
3. Calvo O., Manley J.L. // *Genes Dev.* 2003. V. 17. P. 1321–1327.
4. Kornblihtt A.R., de la Mata M., Fededa J.P., Munoz M.J., Nogues G. // *RNA.* 2004. V. 10. P. 1489–1498.
5. Maniatis T., Reed R. // *Nature.* 2002. V. 416. P. 499–506.
6. Dantoni J.C., Murthy K.G., Manley J.L., Tora L. // *Nature.* 1997. V. 389. P. 399–402.
7. Ujvári A., Luse D.S. // *J. Biol. Chem.* 2004. V. 279. P. 49773–49779.
8. Niwa M., Rose S.D., Berget S.M. // *Genes Dev.* 1990. V. 4. P. 1552–1559.
9. Niwa M., Berget S.M. // *Genes Dev.* 1991. V. 5. P. 2086–2095.
10. Dye M.J., Proudfoot N.J. // *Mol. Cell.* 1999. V. 3. P. 371–378.
11. Vagner S., Vagner C.C., Mattaj I.W. // *Genes Dev.* 2000. V. 14. P. 403–413.
12. Kyburz A., Friedlein A., Langen H., Keller W. // *Mol. Cell.* 2006. V. 23. P. 195–205.
13. Millevoi S., Loulergue C., Dettwiler S., Karaa S.Z., Keller W., Antoniou M., Vagner S. // *EMBO J.* 2006. V. 25. P. 4854–4864.
14. Rigo F., Martinson H.G. // *Mol. Cell Biol.* 2008. V. 28. P. 849–862.
15. Rigo F., Martinson H.G. // *RNA.* 2009. V. 15. P. 823–836.
16. Awasthi S., Alwine J.C. // *RNA.* 2003. V. 9. P. 1400–1409.
17. Guo J., Garrett M., Micklem G., Brogna S. // *Mol. Cell Biol.* 2011. V. 31. P. 639–651.
18. Andersen P.K., Lykke-Andersen S., Jensen T.H. // *Genes Dev.* 2012. V. 26. P. 2169–2179.
19. Gunderson S.I., Polycarpou-Schwarz M., Mattaj I.W. // *Mol. Cell.* 1998. V. 1. P. 255–264.
20. Kaida D., Berg M.G., Younis I., Kasim M., Singh L.N., Wan L., Dreyfuss G. // *Nature.* 2010. V. 468. P. 664–668.
21. Goraczniak R., Behlke M.A., Gunderson S.I. // *Nat. Biotechnol.* 2009. V. 27. P. 257–263.
22. Abad X., Vera M., Jung S.P., Oswald E., Romero I., Amin V., Fortes P., Gunderson S.I. // *Nucl. Acids Res.* 2008. V. 36. P. 2338–2352.
23. Tian B., Pan Z., Lee J.Y. // *Genome Res.* 2007. V. 17. P. 156–165.
24. Berg M.G., Singh L.N., Younis I., Liu Q., Pinto A.M., Kaida D., Zhang Z., Cho S., Sherrill-Mix S., Wan L., Dreyfuss G. // *Cell.* 2012. V. 150. P. 53–64.
25. Cheng Y., Miura R.M., Tian B. // *Bioinformatics.* 2006. V. 22. P. 2320–2325.
26. McQuilton P., St Pierre S.E., Thurmond J., the FlyBase Consortium // *Nucl. Acids Res.* 2012. V. 40. P. D706–D714.

27. Graveley B.R., Brooks A.N., Carlson J.W., Duff M.O., Landolin J.M., Yang L., Artieri C.G., van Baren M.J., Boley N., Booth B.W., et al. // *Nature*. 2011. V. 471. P. 473–479.
28. Celniker S.E., Dillon L.A., Gerstein M.B., Gunsalus K.C., Henikoff S., Karpen G.H., Kellis M., Lai E.C., Lieb J.D., MacAlpine D.M., et al. // *Nature*. 2009. V. 459. P. 927–930.
29. Hernandez G., Vazquez-Pianzola P., Sierra J.M., Rivera-Pomar R. // *RNA*. 2004. V. 10. P. 1783–1797.
30. Langemeier J., Radtke M., Bohne J. // *RNA Biol*. 2013. V. 10. P. 180–184.
31. Langemeier J., Schrom E.M., Rabner A., Radtke M., Zychlinski D., Saborowski A., Bohn G., Mandel-Gutfreund Y., Bodem J., Klein C., Bohne J. // *EMBO J*. 2012. V. 31. P. 4035–4044.
32. West S., Gromak N., Proudfoot N.J. // *Nature*. 2004. V. 432. P. 522–525.
33. Luo W., Johnson A.W., Bentley D.L. // *Genes Dev*. 2006. V. 20. P. 954–965.
34. Dye M.J., Gromak N., Proudfoot N.J. // *Mol. Cell*. 2006. V. 21. P. 849–859.