

---

---

# ГЕНЕТИКА И МОЛЕКУЛЯРНАЯ БИОЛОГИЯ

---

## GENETICS AND MOLECULAR BIOLOGY

---

---

УДК 577.214+616-006.4

### НОКДАУН ГИБРИДНОГО ОНКОГЕНА *KMT2A-AFF1* АССОЦИИРОВАН С ДИФФЕРЕНЦИАЛЬНЫМ СПЛАЙСИНГОМ РНК В КЛЕТКАХ ОСТРОГО ЛИМФОБЛАСТНОГО ЛЕЙКОЗА ЧЕЛОВЕКА

В. В. ГРИНЕВ<sup>1)</sup>, О. ХАЙДЕНРАЙХ<sup>2)</sup>

<sup>1)</sup>Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь

<sup>2)</sup>Детский онкологический центр Вульфсона, Северный институт исследований рака, Ньюкаслский университет, здание Гершеля, NE1 7RU, г. Ньюкасл, Великобритания

Отмечено, что гибридный онкоген *KMT2A-AFF1* является продуктом транслокации t(4;11)(q21;q23). Этот онкоген, при условии появления в ранних предшественниках гемопоэза, выступает в роли одного из факторов генеза острого лимфобластного лейкоза у детей, механизм которого до конца не установлен. Показано, что активность онкогена *KMT2A-AFF1* оказывает влияние на транскриптом лейкозных клеток. Данный эффект реализуется двумя путями: через дифференциальную экспрессию множества генов, подконтрольных онкогену, а также через контроль дифференциального сплайсинга. Представленные результаты свидетельствуют об открытии нового действия онкогена *KMT2A-AFF1* и обеспечивают перспективы в изучении лейкоза, ассоциированного с транслокацией t(4;11)(q21;q23).

**Ключевые слова:** гибридный онкоген *KMT2A-AFF1*; нокдаун экспрессии; альтернативный сплайсинг РНК; лейкозные клетки человека.

---

#### Образец цитирования:

Гринев В. В., Хайденрайх О. Нокдаун гибридного онкогена *KMT2A-AFF1* ассоциирован с дифференциальным сплайсингом РНК в клетках острого лимфобластного лейкоза человека // Журн. Белорус. гос. ун-та. Биология. 2017. № 3. С. 21–27.

#### For citation:

Grinev V. V., Heidenreich O. The knockdown of the fusion oncogene *KMT2A-AFF1* is associated with differential RNA splicing in human acute lymphoblastic leukemia cells. *J. Belarus. State Univ. Biol.* 2017. No. 3. P. 21–27 (in Russ.).

---

#### Авторы:

**Василий Викторович Гринев** – кандидат биологических наук; доцент кафедры генетики биологического факультета.

**Олаф Хайденрайх** – доктор философии; профессор молекулярной гематологии.

#### Authors:

**Vasily V. Grinev**, PhD (biology); associate professor at the department of genetics, faculty of biology.

*grinev\_vv@bsu.by*

**Olaf Heidenreich**, PhD (philosophy); professor of molecular haematology.

*olaf.heidenreich@ncl.ac.uk*

## THE KNOCKDOWN OF THE FUSION ONCOGENE *KMT2A-AFF1* IS ASSOCIATED WITH DIFFERENTIAL RNA SPLICING IN HUMAN ACUTE LYMPHOBLASTIC LEUKEMIA CELLS

V. V. GRINEV<sup>a</sup>, O. HEIDENREICH<sup>b</sup>

<sup>a</sup>Belarusian State University, 4 Niezaliežnasci Avenue, Minsk 220030, Belarus

<sup>b</sup>Wolfson Childhood Cancer Research Centre, Nothern Institute for Cancer Research, Newcastle University, Herschel Building, NE1 7RU, Newcastle upon Tyne, United Kingdom

Corresponding author: V. V. Grinev (grinev\_yv@bsu.by)

The fusion oncogene *KMT2A-AFF1* is the outcome of translocation t(4; 11)(q21; q23). If this occurred at the level of early hematopoietic progenitors, this oncogene plays the role of one of the factors in the genesis of children acute lymphoblastic leukemia, the mechanism of which is not fully understood. In this work, it was shown that the *KMT2A-AFF1* activity affects the state of transcriptome of leukemia cells. This effect is realized in two ways: through the differential expression of the target genes controlled by oncogene, and also through control of differential splicing. The presented results indicate the discovery of a new biological activity of oncogene *KMT2A-AFF1*, and it also opens new perspectives in study of t(4; 11)(q21; q23) positive leukemia.

**Key words:** fusion oncogene *KMT2A-AFF1*; knockdown of expression; alternative RNA splicing; human leukemia cells.

Хромосомные перестройки являются одними из наиболее частых генетических аномалий, ассоциированных с развитием лейкозов у человека. Среди них особо выделяется группа реципрокных негомологичных транслокаций, приводящих к образованию гибридных онкогенов. В частности, такая транслокация, как t(4; 11)(q21; q23), порождает гибридный локус *KMT2A-AFF1* [1]. Появление этого онкогена в гемопоэтических стволовых/прогениторных клетках человека является одним из факторов инициации и формирования острого лимфобластного лейкоза (ОЛЛ) у детей [2]. К сожалению, несмотря на длительную историю изучения, механизмы лейкозогенной активности онкогена *KMT2A-AFF1* в полной мере не расшифрованы [3].

Цель настоящей работы – выяснить, каково влияние вышеуказанного онкогена на качественный и количественный состав транскриптома лейкозных клеток. Для этого с помощью РНК-интерференции экспрессия онкогена *KMT2A-AFF1* в клетках ОЛЛ подавлялась, после чего по технологии RNA-Seq проводилось секвенирование транскриптома таких клеток, а полученные экспериментальные данные подвергались биоинформационному анализу.

Результаты наших исследований показывают, что влияние онкогена *KMT2A-AFF1* на состояние транскриптома лейкозных клеток реализуется двумя путями: через дифференциальную экспрессию множества генов, подконтрольных этому онкогену, а также через прямой или опосредованный контроль дифференциального сплайсинга. Полученные результаты являются первым свидетельством существования у онкогена *KMT2A-AFF1* новой биологической активности. Возможно, что именно измененное под влиянием этого онкогена состояние транскриптома и формирует фенотипические особенности лейкозных клеток при ОЛЛ.

### Материалы и методы исследований

Представленное исследование основано на данных, полученных на клетках линии SEM (DSMZ № ACC 546), которая является модельной клеточной линией положительной по транслокации t(4; 11)(q21; q23) формы ОЛЛ. Клетки SEM были обработаны двумя типами коротких интерферирующих РНК (киРНК): контрольными – siAGF1, которые распознают РНК онкогена *RUNX1-RUNX1T1*, но не онкогена *KMT2A-AFF1*, и активными – siAM4, позволяющими запускать искусственную РНК-интерференцию против *KMT2A-AFF1*. Обработка целевых клеток киРНК, выделение из них и очистка фракции полиаденилированной РНК, подготовка библиотек для секвенирования по технологии RNA-Seq и само секвенирование проводились на базе Детского онкологического центра Вульфсона при Ньюкаслском университете (г. Ньюкасл, Великобритания) согласно стандартным протоколам.

Файлы FASTQ, полученные в результате секвенирования транскриптома клеток SEM, подвергались стандартной процедуре предпроцессинга, после чего короткие секвенированные фрагменты РНК (прочтения или риды) картировались по эталонной сборке GRCh38.p7 генома человека с помощью

программы *Subjunc* [4]. Этот же алгоритм использовался для идентификации соединений экзонов – EEJs (от англ. *exon-exon junctions*) в зрелых молекулах РНК и количественной оценки их встречаемости в транскриптоме клеток SEM. Сбор количественной информации по экспрессии экзонов и генов в изучаемых клетках осуществлялся с помощью программы *featureCounts* [5].

Все количественные данные, полученные при секвенировании транскриптома клеток SEM по технологии RNA-Seq и после картирования прочтений, подвергались многоступенчатой фильтрации, нормализации и трансформации, коротко описанным в контексте дальнейшего изложения результатов. Линейное моделирование и идентификация генов, дифференциально экспрессированных в двух состояниях изучаемых клеток (после обработки контрольными или активными киРНК), проводились в соответствии с описанием в [6–8]. Для идентификации дифференциальных сплайсинговых событий использовались функциональные возможности программы *diffSplice* [6].

Статистические тесты и другие рутинные вычислительные процедуры проводились согласно общепринятым стандартам. Вся аналитическая работа, результаты которой представлены в настоящем исследовании, проведена в среде программирования *R* с использованием как уже готовых решений, доступных через репозиторий *Bioconductor*, так и с помощью оригинальных программ, написанных авторами исследования.

### Результаты исследований и их обсуждение

Прямое или не прямое влияние гена на альтернативный сплайсинг можно обнаружить по дифференциальному сплайсингу. Под дифференциальным сплайсингом подразумевается альтернативный сплайсинг, зависящий от условий, в данном случае – от экспрессии изучаемого гена. Исходя из этого, нами проведен нокдаун онкогена *KMT2A-AFF1* в клетках ОЛЛ линии SEM, содержащих транслокацию  $t(4;11)(q21;q23)$ , с помощью искусственно запускаемой РНК-интерференции и секвенирован транскриптом лейкозных клеток по технологии RNA-Seq. Дальнейшее картирование коротких прочтений и сбор количественной информации по экспрессии генов и идентифицированным EEJs были проведены так, как описано в разделе «Материалы и методы исследований».

Дифференциальный сплайсинг может быть как самостоятельным процессом, так и ассоциированным с дифференциальной экспрессией генов. В связи с этим первоначально нами с использованием функциональных возможностей *R/Bioconductor* библиотек *limma* [6] и *edgeR* [7] проведено линейное моделирование экспрессии генов в клетках SEM, обработанных контрольными siAGF1 и активными siAM4 киРНК. Количественные данные, необходимые для такого моделирования, были получены нами путем подсчета количества прочтений, перекрывающих экзоны генов человека. Координаты экзонов эталонной сборки GRCh38.p7 генома человека были взяты из 85-го выпуска базы данных *Ensembl* [9]. При этом все экзоны, в частности те, геномные координаты которых перекрываются, но отличаются по одному или обоим концам, были объединены в экзонные кластеры. Количество прочтений, попадающих в тот или иной кластер, было суммировано по генам и нормализовано относительно общего размера кластеров каждого гена.

Распределение уровней экспрессии генов в клетках SEM, как видно на рис. 1, а, носит бимодальный характер. Существенная часть генов характеризуется очень низким уровнем экспрессии, является шумом и не может быть использована в дальнейшем анализе.

Кроме того, те данные, которые отражают экспрессию действительно работающих генов, демонстрируют гетероскедастичность – высокую дисперсию при низких значениях экспрессии и, наоборот, низкую дисперсию при высокой экспрессии, что наглядно демонстрирует кривая локально взвешенной регрессии на рис. 1, б. С учетом таких особенностей данных нами не только отфильтрован шум, но также проведены *voom*-нормализация [10] и логарифмическая трансформация оставшихся данных для снижения влияния неоднородной дисперсии на результаты линейного моделирования.

Результаты линейного моделирования, выполненного по стандартной методологии [6–8] на нормализованных данных, показывают, что нокдаун онкогена *KMT2A-AFF1* приводит к дифференциальной экспрессии части генов в клетках SEM (рис. 1, в). Количество таких генов зависит от величины выбранных статистических порогов; так, при использовании консервативных величин этих порогов (минимум двукратное изменение экспрессии между условиями при  $p < 0,001$  и  $q < 0,1$ ) можно говорить о 65 генах с дифференциальной экспрессией. Полученные данные также указывают на то, что невозможно напрямую, без учета дифференциальной экспрессии генов, идентифицировать дифференциальные сплайсинговые события в клетках SEM.

С учетом вышеизложенного нами расширено стандартное линейное моделирование путем введения дополнительного этапа, позволяющего скорректировать те искажения, которые вносит дифференциальная экспрессия генов при обнаружении дифференциальных сплайсинговых событий. Для этого

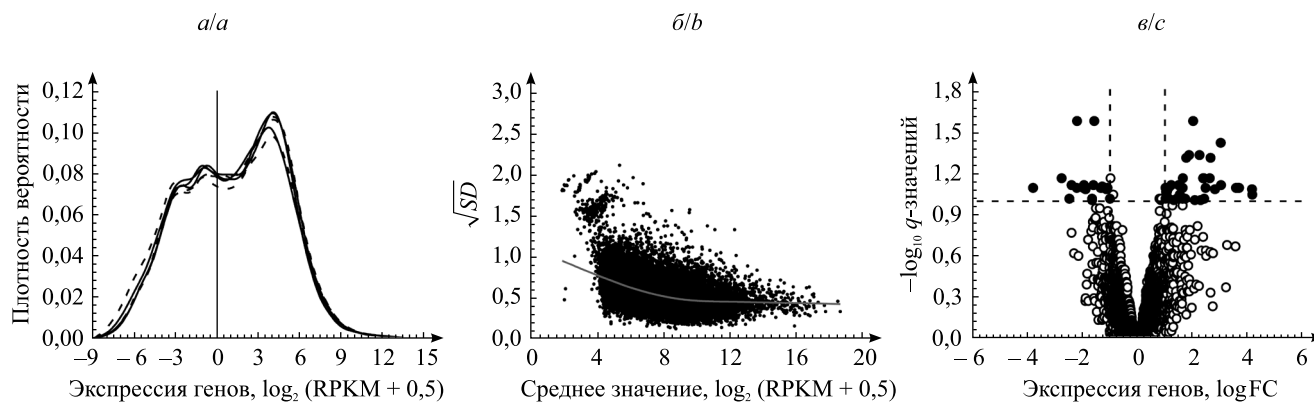


Рис. 1. Нокдаун онкогена *KMT2A-AFF1* приводит к дифференциальной экспрессии генов в клетках SEM: *a* – распределение уровней экспрессии генов в клетках, обработанных контрольными siAGF1 (пунктирные кривые) и активными siAM4 (сплошные кривые) киРНК. Здесь и далее уровень экспрессии генов выражен в RPKM (от англ. *reads per kilobase per million*) или количестве прочтений из 1 млн прочтений в библиотеке RNA-Seq, приходящихся на 1 тыс. нуклеотидов экзонных кластеров гена. Вертикальная линия показывает положение порога экспрессии, использованного для фильтрации данных; *b* – исходные данные, полученные при секвенировании транскриптома клеток SEM по технологии RNA-Seq, демонстрируют гетероскедастичность. Серая линия показывает прохождение кривой локально взвешенной регрессии; *c* – согласно результатам линейного моделирования нокдаун онкогена *KMT2A-AFF1* в клетках SEM меняет экспрессию 65 генов. Ось абсцисс показывает  $\log_2$ -трансформированную разницу ( $\log FC$ ) в экспрессии генов между двумя условиями, а ось ординат отражает статистическую значимость этой разницы. Штриховые линии показывают пороги экспрессии и статистической значимости, которые были использованы при идентификации генов с дифференциальной экспрессией

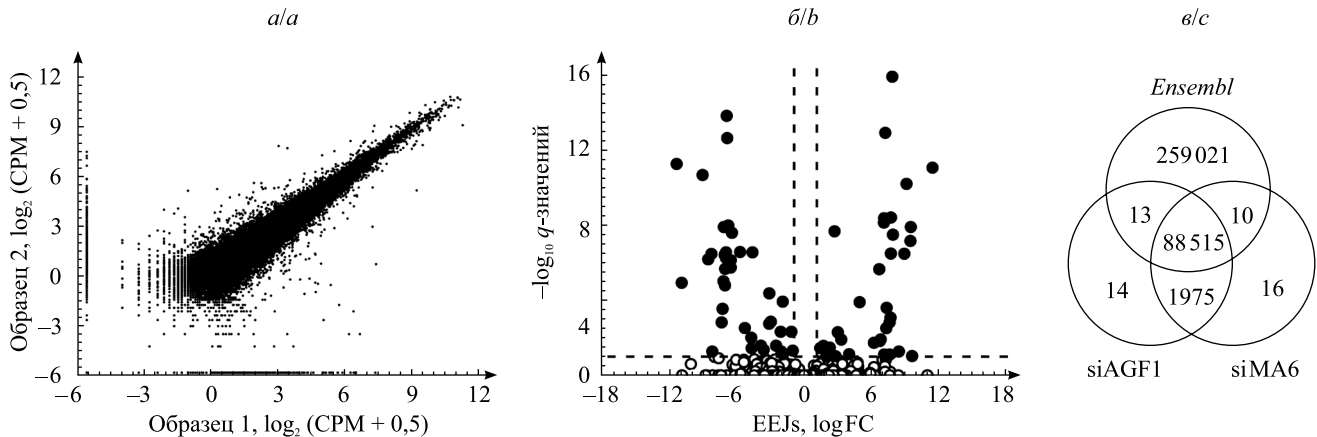
Fig. 1. Knockdown of the *KMT2A-AFF1* oncogene leads to differential expression of genes in SEM cells: *a* – distribution of the levels of gene expression in control siAGF1 (dashed lines) and active siAM4 (solid lines) siRNA treated cells. Here and below, the levels of gene expression are calculated in RPKM, or reads per kilobase per million. Vertical line is the expression threshold for data filtration; *b* – raw RNA-Seq transcriptomic data from SEM cells demonstrate heteroscedasticity. Gray line is a locally weighed regression trend; *c* – according to linear modeling, knockdown of the *KMT2A-AFF1* oncogene changes the expression of 65 genes in SEM cells. x-Axis shows the  $\log_2$  transformed fold changes ( $\log FC$ ) in gene expression between two states of cells. y-Axis demonstrates the statistical significance of  $\log FC$ . Dashed lines correspond to expression and significance thresholds which were used for identification of differentially expressed genes

в результате картирования RNA-Seq прочтений с помощью программы *Subjunc* [4] нами собраны данные обо всех потенциальных EEJs в виде матрицы сплайсинга. Далее, количественные данные последовательно нормализовались относительно размера библиотек RNA-Seq и преобразовывались в количество прочтений на миллион – CPM (от англ. *counts per million*). Поскольку данные RNA-Seq о сплайсинговых событиях, как и данные об экспрессии генов, проявляют гетероскедастичность (рис. 2, *a*), то к ним также были применены *voom*-нормализация [10] и логарифмическая трансформация.

Подготовленные таким образом данные использовались для расчета параметров линейных моделей, из которых с помощью функции *diffSplice* [6] извлекалась информация о дифференциальных событиях сплайсинга. К последним были отнесены те EEJs, которые на фоне других EEJs статистически значимо ( $p < 0,0001$  и  $q < 0,1$ ) и не менее чем двукратно отличались по частоте встречаемости в зрелых молекулах РНК генов при разных состояниях клеток SEM. С помощью такого подхода нам удалось идентифицировать 73 события дифференциального сплайсинга, ассоциированного с нокдауном онкогена *KMT2A-AFF1* и затрагивающего созревание молекул РНК 57 генов. Графически эти результаты отображены на рис. 2, *b*, с помощью графика-вулкана.

Интересно отметить, что 97,8 % всех сплайсинговых событий (как дифференциальных, так и недифференциальных), обнаруженных в клетках SEM, представлены EEJs, аннотированными в 85-м выпуске базы данных *Ensembl*. Лишь 2,2 % (или 2005 в абсолютном исчислении) таких событий являются уникальными для изучаемого типа клеток (рис. 2, *в*). Кроме того, из рис. 2, *в*, видно, что два состояния транскриптома клеток SEM (без нокдауна и с нокдауном онкогена *KMT2A-AFF1*) качественно отличаются лишь по 30 сплайсинговым событиям. Более же существенная разница между этими двумя состояниями клеток обнаруживается только при количественном анализе.

Преобладающим типом альтернативного сплайсинга среди идентифицированных дифференциальных EEJs являются изоформы интронов – незначительные вариации длины интронов по 5'- и 3'-концам одновременно (рис. 3, *a*), на долю которых приходится 39,7 %. Далее, идут такие варианты, как альтернативный первый или последний экзоны (16,4 и 9,6 % соответственно) и каскетные экзоны (12,3 %). На другие типы альтернативного сплайсинга приходится незначительная доля дифференциальных EEJs.



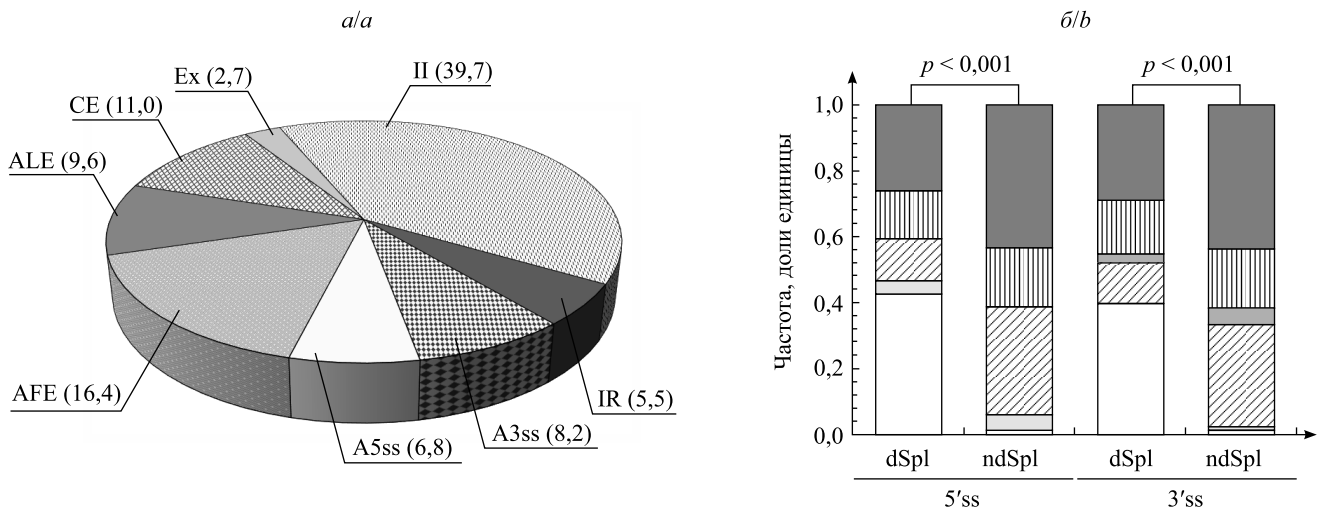
**Рис. 2.** Нокдаун онкогена *KMT2A-AFF1* приводит к дифференциальному сплайсингу РНК ряда генов, активных в клетках SEM: *a* – репрезентативная диаграмма сравнения двух образцов РНК, выделенных из клеток SEM и секвенированных по технологии RNA-Seq. Диаграмма основана на количественных данных о тех прочтениях, которые были картированы по EEJs; *б* – график-вулкан распределения EEJs между двумя состояниями клеток SEM. Ось абсцисс показывает logFC частоты встречаемости EEJs в зрелых молекулах РНК генов при разных состояниях изучаемых клеток, а ось ординат отражает статистическую значимость logFC. Штриховые линии показывают положение порогов, которые были использованы при идентификации дифференциальных сплайсинговых событий; *в* – диаграмма Венна, представляющая распределение всех EEJs, идентифицированных в исследовании, между клетками SEM, обработанными контрольными и активными кРНК. На диаграмме также представлено сравнение экспериментально обнаруженных EEJs с базой данных *Ensembl*

**Fig. 2.** Knockdown of the *KMT2A-AFF1* oncogene leads to differential RNA splicing of the sub-set of genes active in SEM cells: *a* – representative scatterplot of the two RNA-Seq sequenced RNA samples from SEM cells.

This plot is based on reads mapped to EEJs only;

*b* – volcano plot of EEJs distribution between two states of SEM cells. *x*-Axis shows the logFC of EEJs and *y*-axis demonstrates the statistical significance of this logFC. Dashed lines correspond to thresholds which were used for identification of differential splicing events;

*c* – Venn diagram shows the overlap between EEJs identified in control and active siRNA treated SEM cells and *Ensembl* EEJs



**Рис. 3.** Классификация дифференциальных EEJs, обнаруженных в клетках SEM:

*a* – распределение дифференциальных EEJs по типам альтернативного сплайсинга, %: альтернативный первый экзон (AFE), альтернативный последний экзон (ALE), кассетный экзон (CE), экзитрон (Ex), изоформа интрона (II), сохранение интрона (IR), альтернативный 3'-сайт сплайсинга (A3ss) и альтернативный 5'-сайт сплайсинга (A5ss); *б* – распределение сайтов сплайсинга (ss) недифференциальных (ndSpl) и дифференциальных (dSpl) EEJs среди функциональных типов экзонов человека: неаннотированные экзоны (□), 5'-UTR-экзоны (▤), CDS-экзоны (▨), 3'-UTR-экзоны (▧), экзоны некодирующих РНК (▩) и мультифункциональные экзоны (■)

**Fig. 3.** Classification of differential EEJs detected in SEM cells:

*a* – alternative splicing type based classification of differential EEJs, %: alternative first exon (AFE), alternative last exon (ALE), cassette exon (CE), exitron (Ex), intron isoform (II), intron retention (IR), alternative 3'-splice site (A3ss) and alternative 5'-splice site (A5ss);

*b* – the distribution of splice sites (ss) of non-differential (ndSpl) and differential (dSpl) EEJs among of functional type of human exons: non-annotated exons (□), 5'-UTR exons (▤), CDS exons (▨), 3'-UTR exons (▧), exons of non-coding RNA (▩) and multifunctional exons (■)

Особый тип альтернативного сплайсинга представляют экзитроны – участки CDS-экзонов, кодирующих белок, которые могут выступать в качестве интронов и удаляться во время сплайсинга [11]. Среди дифференциальных EEs нами обнаружены два таких события, причем оба затронули сплайсинг РНК генов *FUS* и *SON*, кодирующих факторы сплайсинга и принадлежащих к группе генов, аннотированных в базе данных *Gene Ontology* как гены факторов связывания РНК и сплайсинга. Членами данной группы статистически значимо ( $p = 0,00913$ ) в 3,73 раза обогащен список генов, для которых был детектирован дифференциальный сплайсинг в настоящем исследовании. Эти данные указывают на существование петли ауторегуляции дифференциального сплайсинга, ассоциированного с нокдауном онкогена *KMT2A-AFF1* в клетках SEM.

Интересно отметить, что CDS-экзоны фигурируют в качестве объектов не только такого альтернативного сплайсинга, как экзитроны, но и функционального типа экзонов, частота вовлечения которого в дифференциальный сплайсинг ниже, чем в недифференциальный (см. рис. 3, б). В целом для дифференциального сплайсинга по сравнению с недифференциальным характерна статистически значимая (согласно критерию  $\chi^2 p < 0,001$ ) разница в распределении сайтов сплайсинга среди функциональных типов экзонов, причем с явным сдвигом в сторону экзонов, не аннотированных в 85-м выпуске базы данных *Ensembl* (см. рис. 3, б).

Таким образом, в настоящей работе впервые показано, что нокдаун онкогена *KMT2A-AFF1* в клетках ОЛЛ, положительных по транслокации  $t(4; 11)(q21; q23)$ , сопровождается не только дифференциальной экспрессией некоторых генов, но и дифференциальным сплайсингом РНК ряда генов, в том числе генов, кодирующих факторы сплайсинга. Возможно, открытая нами и ранее неизвестная биологическая активность онкогена *KMT2A-AFF1* вносит вклад в развитие ОЛЛ, что, однако, требует дальнейших специализированных исследований.

### Библиографические ссылки

1. Hensel J. P., Gillert E., Fey G. H., et al. Breakpoints of  $t(4; 11)$  translocations in the human MLL and AF4 genes in ALL patients are preferentially clustered outside of high-affinity matrix attachment regions // *J. Cell Biochem.* 2001. Vol. 82. P. 299–309.
2. Müller A. M. S., Duque J., Shizuru J. A., et al. Complementing mutations in core binding factor leukemias: from mouse models to clinical applications // *Oncogene.* 2008. Vol. 27. P. 5759–5773.
3. Ballabio E., Milne T. A. Molecular and epigenetic mechanisms of MLL in human leukemogenesis // *Cancers (Basel).* 2012. Vol. 4. P. 904–944.
4. Liao Y., Smyth G. K., Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote // *Nucleic Acids Res.* 2013. Vol. 41. P. e108.
5. Liao Y., Smyth G. K., Shi W. FeatureCounts: an efficient general-purpose program for assigning sequence reads to genomic features // *Bioinformatics.* 2014. Vol. 30. P. 923–930.
6. Ritchie M. E., Phipson B., Wu D., et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies // *Nucleic Acids Res.* 2015. Vol. 43. P. e47.
7. Robinson M. D., McCarthy D. J., Smyth G. K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data // *Bioinformatics.* 2010. Vol. 26, № 1. P. 139–140.
8. Smyth G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments // *Stat. Appl. Genet. Molec. Biol.* 2004. Article 3.
9. Aken B. L., Ayling S., Barrell D., et al. The Ensembl gene annotation system // *Database (Oxford).* 2016. Vol. 2016. P. 1–19.
10. Law C. W., Chen Y., Shi W., et al. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts // *Genome Biol.* 2014. Vol. 15. P. R29.
11. Yamile M., Markus H., Zahra A., et al. Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity // *Genome Res.* 2015. Vol. 25. P. 995–1007.

### References

1. Hensel J. P., Gillert E., Fey G. H., et al. Breakpoints of  $t(4; 11)$  translocations in the human MLL and AF4 genes in ALL patients are preferentially clustered outside of high-affinity matrix attachment regions. *J. Cell Biochem.* 2001. Vol. 82. P. 299–309.
2. Müller A. M. S., Duque J., Shizuru J. A., et al. Complementing mutations in core binding factor leukemias: from mouse models to clinical applications. *Oncogene.* 2008. Vol. 27. P. 5759–5773.
3. Ballabio E., Milne T. A. Molecular and epigenetic mechanisms of MLL in human leukemogenesis. *Cancers (Basel).* 2012. Vol. 4. P. 904–944.
4. Liao Y., Smyth G. K., Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 2013. Vol. 41. P. e108.
5. Liao Y., Smyth G. K., Shi W. FeatureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014. Vol. 30. P. 923–930.

6. Ritchie M. E., Phipson B., Wu D., et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015. Vol. 43. P. e47.
7. Robinson M. D., McCarthy D. J., Smyth G. K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010. Vol. 26, No. 1. P. 139–140.
8. Smyth G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Molec. Biol.* 2004. Article 3.
9. Aken B. L., Ayling S., Barrell D., et al. The Ensembl gene annotation system. *Database (Oxford).* 2016. Vol. 2016. P. 1–19. DOI: 10.1093/database/baw093.
10. Law C. W., Chen Y., Shi W., et al. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014. Vol. 15. P. R29.
11. Yamile M., Markus H., Zahra A., et al. Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Res.* 2015. Vol. 25. P. 995–1007.

Статья поступила в редколлегию 13.09.2017.  
Received by editorial board 13.09.2017.