УДК 577.218:616.155.392

ЭПИГЕНЕТИЧЕСКИЕ МАРКЕРЫ ХРОМАТИНА АССОЦИИРОВАНЫ СО СПЛАЙСИНГОМ РНК В ЛЕЙКОЗНЫХ КЛЕТКАХ ЧЕЛОВЕКА

Т. В. РОМАНОВСКАЯ [1)]***, А. В. КВЕТКО*** [1)]***, В. В. ГРИНЕВ*** [1)]

[1)]*Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь*

Изучена ассоциация между паттерном распределения различных эпигенетических маркеров и событиями сплайсинга на уровне полного генома и транскриптома в двух лейкозных клеточных линиях человека Kasumi-1 и SEM с двумя разными типами реципрокных хромосомных транслокаций. Показано, что многие эпигенетические маркеры, определяющие более открытое или закрытое состояние хроматина, распределены неодинаково в областях донорных и акцепторных, а также канонических и альтернативных сайтов сплайсинга экспрессирующихся генов. Маркеры открытого хроматина значимо чаще присутствуют в области участков с альтернативными событиями сплайсинга, чем в участках с каноническим сплайсингом, в то время как для маркера триметилирования гистона 3 по лизину в позиции 36 наблюдается противоположная тенденция. Полученные результаты вскрывают наличие дополнительного, пока еще очень плохо изученного слоя в регуляции альтернативного сплайсинга в клетках человека.

Ключевые слова: эпигенетические маркеры; альтернативный сплайсинг; лейкоз.

Благодарность. Работа выполнена в рамках подпрограммы «Объединение» государственной программы научных исследований «Конвергенция-2020» (задание 3.08.03).

EPIGENETIC MARKS ON THE CHROMATIN ARE ASSOCIATED WITH RNA SPLICING IN HUMAN LEUKEMIA CELLS

T. V. RAMANOUSKAYA [a]***, A. V. KVIATKO*** [a]***, V. V. GRINEV*** [a]

[a]*Belarusian State University, 4 Niezaliežnasci Avenue, Minsk 220030, Belarus*
Corresponding author: V. V. Grinev (grinev_vv@bsu.by)

In this work we estimated associations between distribution patterns of several epigenetic marks and splicing events on the level of full genome and transcriptome in the cells of two leukemic cell lines containing two different reciprocal chromosome translocations. Significant difference in distribution of epigenetic marks was found, contributing to more opened or more closed chromatin in loci of donor *vs* acceptor and canonical *vs* alternative splice sites in expressing genes. Marks of the opened chromatin are significantly more often present in the genomic regions with alternative splicing events than in regions with canonical splicing, while for the mark of the histone 3 trimethylation at lysine 36, the opposite trend is observed. The obtained results reveal the presence of an additional, still very poorly studied layer in the regulation of alternative splicing in human cells.

Keywords: epigenetic marks; alternative splicing; leukemia.

Acknowledgements. The work was carried out within the framework of the sub-program «Convergence» of the State program of scientific research «Convergence-2020» (3.08.03).

Образец цитирования:
Романовская ТВ, Кветко АВ, Гринев ВВ. Эпигенетические маркеры хроматина ассоциированы со сплайсингом РНК в лейкозных клетках человека. *Журнал Белорусского государственного университета. Биология.* 2019;2:70–81.
https://doi.org/10.33581/2521-1722-2019-2-70-81

For citation:
Ramanouskaya TV, Kviatko AV, Grinev VV. Epigenetic marks on the chromatin are associated with RNA splicing in human leukemia cells. *Journal of the Belarusian State University. Biology.* 2019;2:70–81. Russian.
https://doi.org/10.33581/2521-1722-2019-2-70-81

Авторы:
Татьяна Владимировна Романовская – кандидат биологических наук, доцент; доцент кафедры генетики биологического факультета.
Анастасия Викторовна Кветко – студентка биологического факультета. Научный руководитель – Т. В. Романовская.
Василий Викторович Гринев – кандидат биологических наук, доцент; доцент кафедры генетики биологического факультета.

Authors:
Tatyana V. Ramanouskaya, PhD (biology), docent; associate professor at the department of genetics, faculty of biology.
https://orcid.org/0000-0001-5519-4891
ramanouskaya_tv@bsu.by
Anastasiya V. Kviatko, student at the faculty of biology.
anastasiya.kvetko@mail.ru
Vasily V. Grinev, PhD (biology), docent; associate professor at the department of genetics, faculty of biology.
grinev_vv@bsu.by

Введение

Большинство генов человека имеют мозаичное строение, т. е. образуются чередованием экзонов и интронов. Сплайсинг – процесс вырезания интронов из первичных РНК – является необходимой частью реализации наследственной информации. Как сегодня известно, РНК-продукты большинства генов подвергаются альтернативному сплайсингу, что приводит к появлению серии молекул зрелых РНК, различающихся по структуре и функции [1–3]. Спектр продуктов экспрессии гена неодинаков в разных тканях и клеточных типах и в разных условиях. Значительные изменения в наборе продуктов сплайсинга отмечаются для различных типов злокачественных опухолей: многие авторы указывают на увеличение разнообразия вариантов альтернативного сплайсинга, появление изоформ РНК, которые не встречаются в здоровых тканях. Вопрос о функциональной роли альтернативного сплайсинга дискутируется, но по крайней мере для некоторых генов имеются достаточно однозначные свидетельства участия этого процесса в формировании функционально различающихся белковых изоформ или в регуляции экспрессии генов (благодаря взаимодействию с системой NMD – деградации изоформ с преждевременным стоп-кодоном) [2].

В связи с вышесказанным существует обоснованный интерес в отношении молекулярных механизмов, которые определяют паттерн сплайсинга для каждого транскрибируемого гена в клетках. Известно, что в молекулах РНК присутствуют специфические мотивы, узнаваемые сплайсосомой – специализированным белковым комплексом, который непосредственно обеспечивает реализацию этого процесса. Каждый интрон ограничен такими мотивами с двух сторон: проксимально по отношению к сайту начала транскрипции располагается донорный (5′), дистально – акцепторный (3′) сайт сплайсинга. Сродство сплайсосомы к соответствующему участку РНК зависит от силы сайта сплайсинга, т. е. степени сходства нуклеотидной последовательности с консенсусной (оптимальной) последовательностью, которая обеспечивает наиболее энергетически стабильное связывание сплайсосомы. Кроме того, в регуляции сплайсинга участвуют РНК-связывающие белки (трансрегуляторные факторы сплайсинга), мотивы узнавания для которых могут располагаться в экзонах или интронах, как правило, на расстоянии до нескольких сотен нуклеотидов от сайта сплайсинга [4].

В последнее десятилетие появились работы, указывающие на то, что в регуляции сплайсинга, по-видимому, задействованы также и эпигенетические факторы, которые работают на уровне ДНК и белков хроматина [4]. Такие выводы сделаны отчасти на основании биоинформатических исследований, показывающих, что различные эпигенетические маркеры, включая модификации гистонов, чувствительность к ДНКазе, метилирование ДНК, распределены по-разному в области интронов и экзонов. Также выполнялись и прямые эксперименты, в которых манипуляции с системой эпигенетической регуляции приводили к изменению частоты событий сплайсинга в тех или иных его сайтах.

Рассматриваются преимущественно два механизма влияния маркеров хроматина на сплайсинг. Оба опираются на тот факт, что сплайсинг происходит преимущественно котранскрипционно, т. е. взаимодействие регуляторных белков сплайсинга с РНК и сборка сплайсосомных комплексов осуществляются прежде, чем новосинтезированная молекула РНК отделяется от ДНК-матрицы [4]. Первый механизм – влияние эпигенетического состояния хроматина на скорость продвижения РНК-полимеразы [5]. Быстрый синтез цепочки РНК приводит к тому, что в ней одновременно присутствует несколько конкурирующих друг с другом мотивов для взаимодействия с белками-регуляторами. При быстром синтезе цепочки РНК, вероятнее всего, используется более сильный из них, в то время как при медленном синтезе с большей вероятностью используется ближайший (проксимальный) сайт сплайсинга, даже если он окажется слабее, чем дистальный. Второй механизм – опосредованное модифицированными гистонами рекрутирование к локусу трансфакторов. Хорошим примером является функционирование белка MRG15, который, с одной стороны, взаимодействует с гистоном H3K36me3, локализованным в основном в интронах активных генов, а с другой стороны, связывает и рекрутирует факторы сплайсинга PTBP1 и PTBP2 к синтезируемому транскрипту [6; 7].

Особенности протекания сплайсинга в клетках со злокачественной трансформацией могут быть обусловлены тремя основными причинами: 1) изменением уровня экспрессии трансфакторов регуляции сплайсинга; 2) изменением уровня экспрессии белков системы контроля качества РНК, которые отбраковывают и уничтожают некоторые изоформы мРНК; 3) изменением эпигенетического состояния хроматина в участках активно экспрессирующихся генов.

В данной работе мы изучили взаимосвязь между паттерном распределения нескольких эпигенетических маркеров и реализуемым сплайсингом в двух клеточных линиях: Kasumi-1 (острый миелоидный лейкоз) и SEM (острый лимфобластный лейкоз). Следует заметить, что каждая из клеточных линий содержит транслокацию, в результате которой возникает новый онкоген, причем в обоих случаях белковый продукт гена участвует в эпигенетической регуляции хроматина. В линии Kasumi-1 это ген *RUNX1-RUNX1T1*

(репрессор транскрипции, обеспечивает локальное деацетилирование хроматина в области промотора) [8], а в линии SEM – *KMT2A-AFF1* (активатор транскрипции, обеспечивает метилирование лизина 4 в гистоне H3 в области промотора, а также изменяет распределение других эпигенетических маркеров в теле гена) [9].

Материалы и методы исследования

Исходные данные RNA-Seq для лейкозных клеточных линий человека. Использованы данные для клеточных линий Kasumi-1 и SEM. Kasumi-1 (ATCC® CRL-2724TM) является модельной клеточной линией положительной по транслокации t(8;21)(q22;q22) формы острого миелоидного лейкоза человека, а линия SEM (DSMZ № ACC 546) – положительной по транслокации t(4;11)(q21;q23) формы острого лимфобластного лейкоза.

Представленное исследование основано на результатах полнотранскриптомного секвенирования полиаденилированной клеточной РНК, проведенного по методу парно-концевого RNA-Seq, описанных ранее [3; 10; 11]. Материалы для линии Kasumi-1 представлены в базе GEO с кодами доступа GSM1316401, GSM1316402, GSM1316403; для линии SEM – GSM1828405, GSM1828406, GSM1828407.

Файлы FASTQ, полученные при секвенировании транскриптома целевых клеток, подвергались стандартной процедуре предпроцессинга, после чего чтения (короткие секвенированные фрагменты РНК) картировались по эталонной сборке GRCh38.p7 генома человека с помощью пакета *Rsubread v.1.22.3* [12]. В дальнейшем количественные данные подвергались многоступенчатой фильтрации, нормализации и трансформации [13].

Эпигенетические маркеры и CpG-островки. Координаты пиков эпигенетических маркеров получены на основе данных ChIP-Seq, DNase-Seq и ATAC-Seq, опубликованных ранее [9; 11; 14]. В базе GEO материалы представлены в сериях с кодами доступа GSE29222 для клеток линии Kasumi-1, GSE74812 и GSE83671 – для клеток линии SEM.

Координаты CpG-островков в геноме человека получены из геномного браузера UCSC.

Идентификация экзон-экзонных стыков. Экзон-экзонные стыки идентифицированы с использованием пакета *Rsubread v.1.22.3* [12]. Полученные BAM-файлы были преобразованы в таблицы для дальнейшего анализа при помощи кода R собственной разработки. Эти матрицы представляют собой полный список всех экзон-экзонных стыков с геномными координатами и указанным числом подтверждающих чтений в каждом образце.

Статистические тесты выполнялись в соответствии с общепринятыми стандартами. Вся аналитическая работа проведена в среде программирования R с использованием как уже готовых решений, доступных через репозиторий Bioconductor или хранилище CRAN, так и оригинальных программ, написанных авторами исследования.

Результаты и их обсуждение

Классификация экзон-экзонных стыков в соответствии с типами альтернативного сплайсинга. Как указывалось выше, для анализа использовались данные RNA-Seq из двух клеточных линий, на основе которых в результате ряда процедур биоинформатической обработки были получены файлы, содержащие полный список событий сплайсинга. Каждое такое событие отображается в отдельной строке в виде пары координат, соответствующих границам соединяемых при сплайсинге экзонов. Комбинации этих координат мы будем далее обозначать как экзон-экзонные стыки в сокращенном виде: EEJs (англ. *exon-exon junctions*).

В качестве модельной основы для классификации типов альтернативного сплайсинга мы использовали идею гипотетической неальтернативной, или канонической, РНК. Это условная молекула, которая могла бы быть получена в случае, если бы ген имел единственный сайт начала транскрипции, не имел бы альтернативных сайтов сплайсинга и имел единственный сайт терминации транскрипции. Для каждого гена структура модельной канонической РНК была рассчитана с использованием моделей генов человека, аннотированных в базе Ensembl. Мы формировали кластеры из всех перекрывающихся экзонов каждого гена (не учитывая на этом этапе удерживаемые интроны – они рассматривались как отдельный тип событий альтернативного сплайсинга. Также при этом не учитывались аннотированные альтернативные первые и последние экзоны генов). Внешние границы полученных кластеров определялись как границы экзонов модельной неальтернативной РНК. Таким образом, был составлен список координат экзонов, к которому далее был добавлен список координат удерживаемых интронов, альтернативных первых и альтернативных последних экзонов.

При помощи пакета *GenomicRanges v.1.32.3* [15] найденные координаты экспериментально детектированных EEJs сопоставлялись с координатами экзон-экзонных стыков в модельных РНК, и разрабо-

танный алгоритм определял каждый стык как относящийся или не относящийся к одному из следующих типов:

1) каноническое событие (canonical EEJ, CanEEJ), если координаты EEJs полностью соответствуют границам экзонов в модельной РНК;

2) альтернативный 5′-сайт сплайсинга (alternative 5′ splice site, Alt5′ss), если сплайсинг осуществляется с использованием донорного сайта (5′-границы интрона), не совпадающего с каким-либо донорным сайтом из описанной в модели канонической молекулы РНК;

3) альтернативный 3′-сайт сплайсинга (alternative 3′ splice site, Alt3′ss), если сплайсинг осуществляется с использованием акцепторного сайта (3′-границы интрона), не совпадающего с каким-либо акцепторным сайтом из описанной в модели канонической молекулы РНК;

4) пропуск кассетного экзона (cassette exon, CE), если в область экспериментально идентифицированного интрона попадают один или несколько пропущенных экзонов, имеющихся в модельной РНК;

5) альтернативный первый экзон (alternative first exon, AltFE), если 5′-сайт сплайсинга экспериментально идентифицированного экзон-экзонного стыка точно соответствует дистальной границе экзона из модельной РНК, аннотированного как альтернативный первый экзон, и при этом соответствующая граница не попадает в область, разделяющую границы каких-либо внутренних экзонов в модельных РНК;

6) альтернативный последний экзон (alternative last exon, AltLE), если 3′-сайт сплайсинга экспериментально идентифицированного экзон-экзонного стыка точно соответствует проксимальной границе экзона из модельной РНК, аннотированного как альтернативный последний экзон, и при этом соответствующая граница не попадает в область, разделяющую границы каких-либо внутренних экзонов в модельных РНК;

7) сохраняемый интрон (retained intron, RI), если в данных RNA-Seq сохраняются последовательности, которые являются интронами, т. е. должны вырезаться при сплайсинге для образования функционального продукта. Мы классифицировали событие как RI только в случае полного совпадения границ такого экзон-экзонного стыка с аннотированными сохраняемыми интронами согласно геномному браузеру Ensemble [13].

Альтернативные сплайсинговые события могут оказываться комплексными, сочетая в себе признаки одновременно нескольких типов альтернативного сплайсинга. Так, все события с вовлечением альтернативного первого экзона по определению имеют альтернативный донорный сайт. Аналогично вовлечение альтернативного последнего экзона подразумевает использование альтернативного акцепторного сайта. Также с использованием альтернативных донорных или акцепторных сайтов может сочетаться пропуск кассетного экзона; наконец, присутствуют и случаи сочетания в одном событии альтернативного донорного и акцепторного сайтов. На рис. 1 показаны распределения соответственно описанным типам событий сплайсинга, детектированных в двух клеточных линиях. Ввиду того что часть альтернативных EEJs являются комплексными по типу, общее число детектированных альтернативных событий меньше суммы числа событий каждого отдельного типа.

Результаты для обеих клеточных линий оказались существенно схожими: на долю альтернативных пришлось около 17 % всех EEJs, а наиболее часто встречающимся типом альтернативного сплайсинга оказался пропуск кассетного экзона, хотя и доля других типов событий альтернативного сплайсинга достаточно велика.

Следует отметить, что, хотя доля канонических событий заметно выше доли альтернативных событий в этих наборах данных, анализ числа чтений, подтверждающих каждый экзон-экзонный стык в данных RNA-Seq, отнюдь не показывает, что альтернативные варианты сплайсинга представлены существенно более редкими событиями, чем канонические. Сравнение распределения вероятностей по уровню экспрессии демонстрирует лишь небольшой (хотя и статистически достоверный согласно примененному тесту Манна – Уитни) сдвиг влево для альтернативных событий сплайсинга по сравнению с каноническими в обеих клеточных линиях (рис. 2). Медианы по количеству чтений для альтернативных и канонических EEJs в линии Kasumi-1 составили 89,7 и 101,7 соответственно, а в линии SEM – 95,7 и 120,3. Этот факт может служить подкрепляющим аргументом в пользу того, что альтернативный сплайсинг выполняет в клетках жизненно важные функции и подчинен механизмам автономной регуляции.

Анализ эпигенетических маркеров в клетках линий Kasumi-1 и SEM. Первичные данные с описанием эпигенетических маркеров в хроматине двух клеточных линий были собраны из открытых баз данных (см. раздел «Материалы и методы») (рис. 3). Наборы проанализированных маркеров оказались не слишком обширными, но разными, что позволило извлечь больше полезной информации при обработке данных. Единственным общим показателем здесь стало описание распределения CpG-островков – таковыми считают участки ДНК, обогащенные динуклеотидами ЦГ (цитозин-гуанозин). Мы решили включить CpG-островки в анализ, поскольку известна достаточно устойчивая тенденция

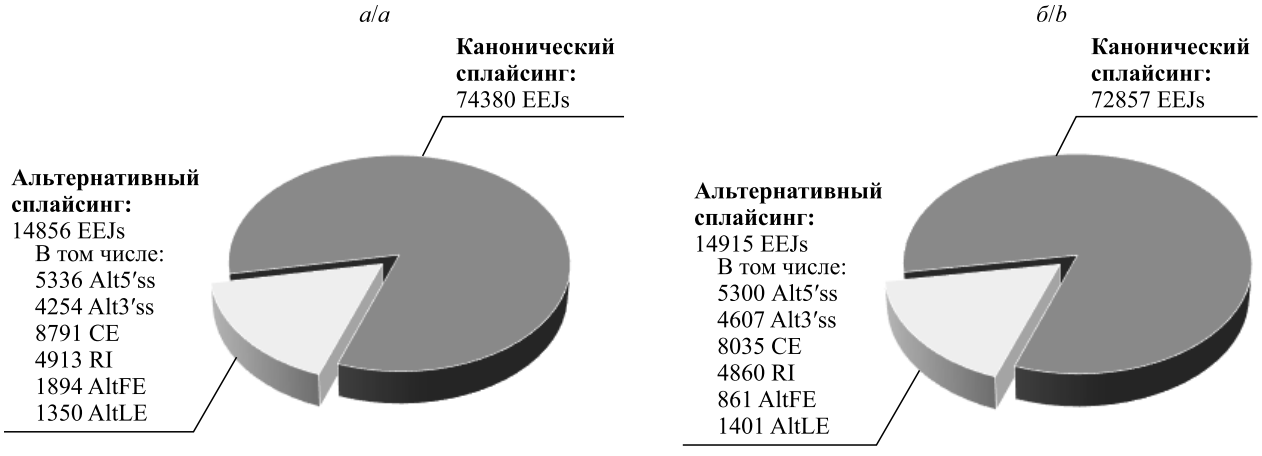

Рис. 1. Количественное распределение детектированных сплайсинговых событий по типам альтернативного сплайсинга для линий Kasumi-1 (*а*) и SEM (*б*): EEJs – экзон-экзонные стыки; Alt5′ss – альтернативный 5′-сайт сплайсинга; Alt3′ss – альтернативный 3′-сайт сплайсинга; CE – пропуск кассетного экзона; RI – сохраняемый интрон; AltFE – альтернативный первый экзон; AltLE – альтернативный последний экзон

Fig. 1. Distribution of the detected splicing events according to types of alternative splicing for Kasumi-1 cell line (*a*) and for SEM cell line (*b*); EEJs – exon-exon junctions; Alt5′ss – alternative 5′ splice site; Alt3′ss – alternative 3′ splice site; CE – skipped cassette exon; RI – retained intron; AltFE – alternative first exon; AltLE – alternative last exon

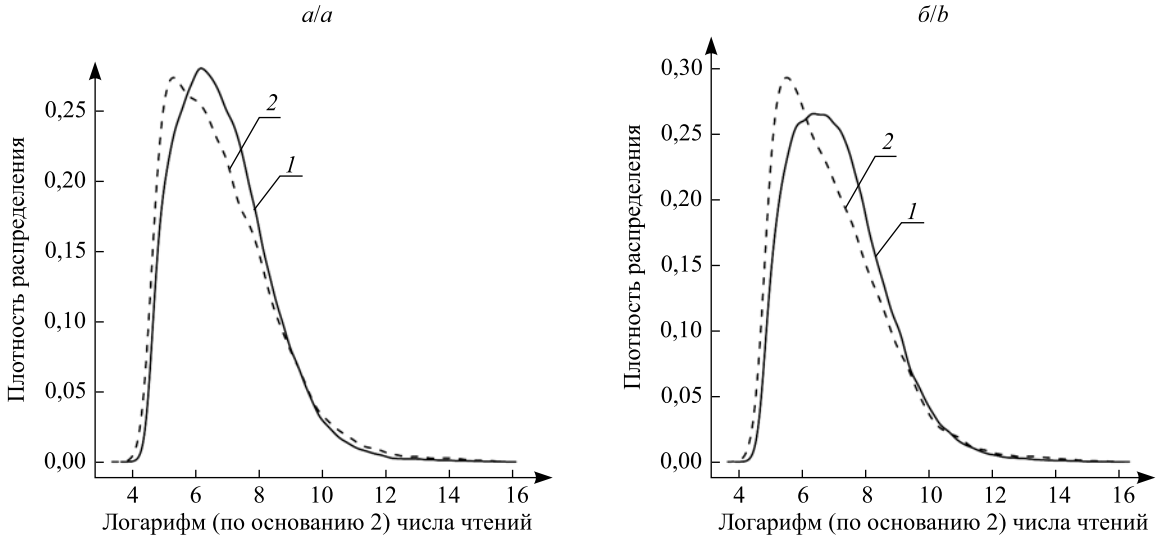

Рис. 2. Распределение канонических и альтернативных событий сплайсинга по уровню экспрессии в клетках линий Kasumi-1 (*а*) и SEM (*б*): *1* – канонические экзон-экзонные стыки; *2* – альтернативные экзон-экзонные стыки

Fig. 2. Distribution of canonical and alternative splicing events by expression level in the cell lines Kasumi-1 (*a*) and SEM (*b*): *1* – canonical exon-exon junctions; *2* – alternative exon-exon junctions

к поддержанию слабокомпактизованного состояния хроматина и пониженного метилирования ДНК в большинстве таких участков.

Остальные данные – это преимущественно пики, выявляемые методом иммунопреципитации хроматина (ChIP-Seq) с антителами, связывающими те или иные белки хроматина. В частности, метилированные или ацетилированные формы гистонов или белки, имеющие сродство к неметилированным CpG-островкам (этот анализ называется BioCAP). Также, исходя из особенностей конкретных клеточных линий, были использованы антитела к продуктам гибридных онкогенов, образующихся вследствие специфических хромосомных транслокаций. Эти белки, как было сказано выше, функционируют как транскрипционные факторы, способствуя ацетилированию (в случае белка KMT2A-AFF1 в линии SEM) или деацетилированию (в случае белка RUNX1-RUNX1T1 в линии Kasumi-1) гистонов.

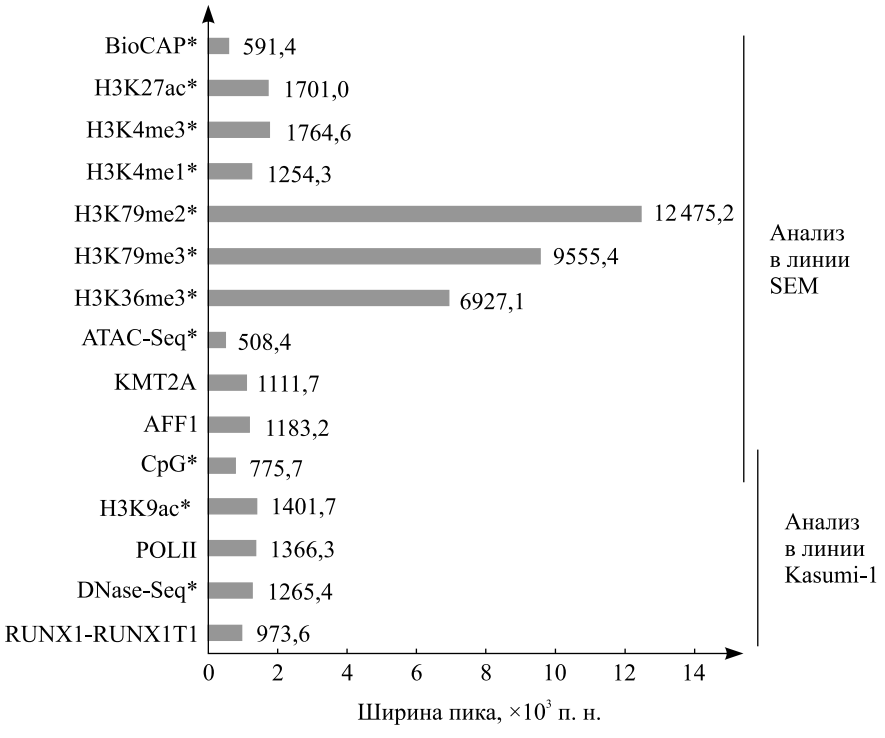

Рис. 3. Эпигенетические маркеры, проанализированные в клетках линий Kasumi-1 и SEM.
Слева указаны обозначения проанализированных маркеров.
Знаком * отмечены маркеры, которые ассоциируются с открытым хроматином

Fig. 3. Epigenetic markers studied in the Kasumi-1 and SEM cell lines.
Marks names are indicated on the left.
Asterisks indicate the marks associated with opened chromatin

Для линии Kasumi также использованы данные по пикам РНК полимеразы II – ее присутствие указывает на наличие в соответствующем участке активной транскрипции. В то же время пики РНК полимеразы II в транскрибируемом гене располагаются дискретно вследствие того, что фермент движется вдоль гена с разной скоростью. Пик выявляется с большей вероятностью в тех участках, где фермент движется медленнее. Это важный момент, поскольку, как уже было сказано выше, скорость транскрипции оказывает влияние на характер сплайсинга. Еще две использованные при исследовании группы данных – результаты анализа на чувствительность к ДНКазе (DNase-Seq) или транспозазе (англ. *assay for transposase-accessible chromatin*, ATAC-Seq), пики в таких анализах соответствуют участкам наиболее открытого (декомпактизованного) хроматина, доступного для взаимодействия с различными белками нуклеоплазмы. Отмечается также, что открытость хроматина облегчает продвижение РНК полимеразы II и повышает локальную скорость транскрипции.

На рис. 3 показана ширина пиков учтенных маркеров, и можно заметить, что этот показатель значительно варьирует. Особенно высокими значениями выделяются пики маркеров H3K79me2, H3K79me3 и H3K36me3. Согласно литературным данным, эти маркеры имеют тенденцию к распределению вдоль тела экспрессирующихся генов, при этом их функция разнопланова, включая контроль скорости транскрипции и регуляцию сплайсинга [5–7; 16].

Анализ распределения эпигенетических маркеров относительно интронов и сайтов сплайсинга. Перекрывание пиков маркеров с интронами и сайтами сплайсинга определялось с использованием пакета *GenomicRanges v.1.32.3* [15]. При этом следовало выяснить, будут ли обнаруживаться какие-либо корреляции в распределении эпигенетических маркеров относительно реализуемых канонических и альтернативных событий сплайсинга.

На первом этапе было проверено, как распределены пики в области интронов, ограниченных детектированными EEJs. Проведенная классификация EEJs позволила дифференциально оценить и сравнить частоту присутствия каждого типа маркера в канонических и альтернативных интронах (рис. 4). Как можно заметить, наиболее стабильным для всех интронов маркером является гистон 3 с триметилированием по лизину 36 – H3K36me3. Остальные маркеры присутствуют лишь в половине интронов или менее. Однако разница между частотой маркирования канонических интронов и альтернативных очевидна и статистически достоверна. В целом можно отметить, что маркеры, свойственные открытому хроматину, лишь очень редко присутствуют в канонических интронах, несколько чаще – в сохраняемых

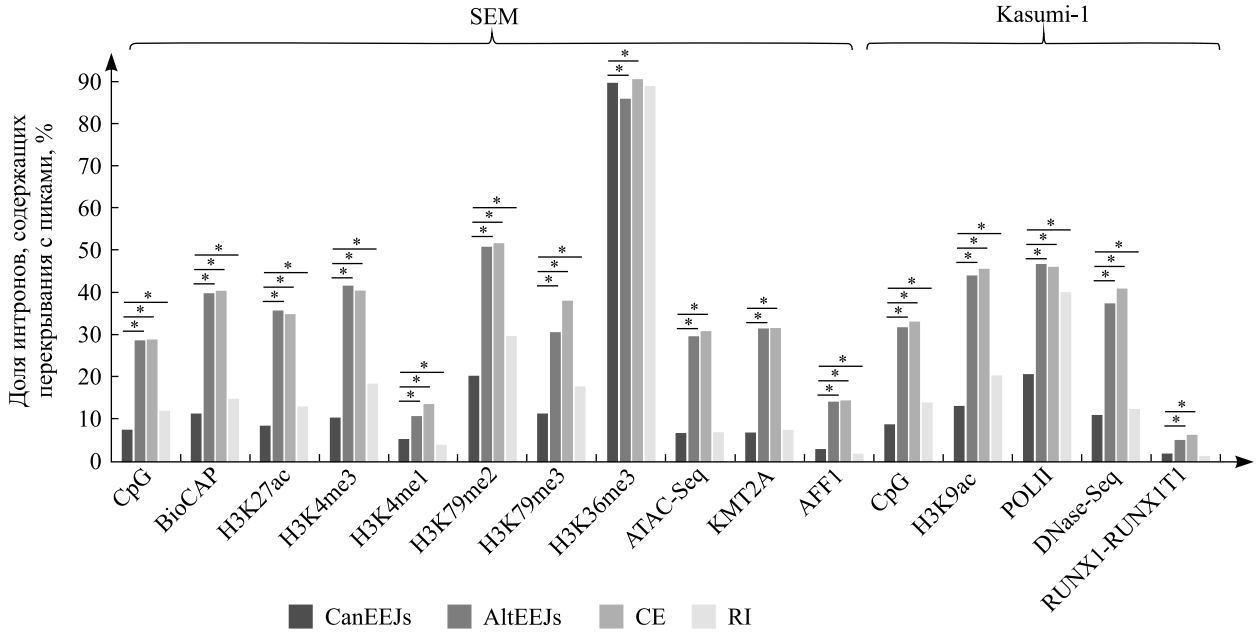

Рис. 4. Распределение эпигенетических маркеров в интронах экспрессирующихся генов: CanEEJs – интроны, образующиеся при канонических событиях сплайсинга; AltEEJs – интроны, образующиеся при различных вариантах альтернативного сплайсинга; CE – интроны, образующиеся при сплайсинге с пропуском кассетного экзона; RI – сохраняемые интроны; * – различия в значениях по указанному маркеру достоверны при $p < 0{,}001$

Fig. 4. Distribution of epigenetic marks in introns of expressed genes: CanEEJs – introns formed during the canonical events of splicing; AltEEJs – introns forming with different alternative splicing variants; CE – introns formed during splicing with a cassette exon skip; RI – retained introns; * – difference in value for the specified mark is significant at $p < 0.001$

интронах и значительно чаще (в 2–3 раза) – в интронах, образующихся при иных типах альтернативного сплайсинга, в частности включая случаи сплайсинга с пропуском кассетного экзона.

Нужно заметить, что между интронами, образованными каноническим и альтернативным сплайсингом некоторых типов, есть существенная разница в размерах: например, если средний размер канонических интронов составляет около 3,4 т. п. н., то для интронов с пропущенным кассетным экзоном средний размер примерно в 3,5 раза превышает эту величину, а сохраняемые интроны в основном очень короткие (до нескольких сот нуклеотидов), в среднем в 8 раз короче по сравнению со средними каноническими. Это может отчасти объяснять те различия в частоте обнаружения пиков в таких интронах (см. рис. 4).

Дополнительно был проведен анализ средней плотности пиков в интроне – под плотностью здесь понимается суммарная ширина областей (пиков), несущих тот или иной маркер, попадающая в интрон, деленная на ширину соответствующего интрона. Итоговую плотность выражали в виде процентных величин (доля ширины интрона, несущая маркер) (рис. 5). Можно заметить, что многие маркеры открытого хроматина (BioCAP, H3K27ac и H3K9ac, H3K4me3, области повышенной чувствительности к ДНКазе и транспозазе) дают более высокие значения плотности в интронах, образованных при альтернативном сплайсинге, чем при каноническом. Однако в случае интронов с пропуском кассетных экзонов по некоторым маркерам значения схожи с таковыми для канонического сплайсинга. Сохраняемые интроны также в основном характеризуются более высокой плотностью маркеров, что может отчасти объясняться их небольшим размером. Значительно более высокая плотность пиков РНК полимеразы в сохраняемых интронах, по сравнению с каноническими, может иметь несколько потенциальных объяснений, которые еще предстоит проверить, в частности: 1) повышенная экспрессия генов, в которых присутствует сплайсинг такого типа; 2) замедление движения РНК полимеразы в соответствующих участках.

Как и следовало ожидать (исходя из литературных данных), наиболее высоко представлены в интронах всех типов такие маркеры, как H3K79me2, H3K79me3 и H3K36me3. При этом маркирование по H3K79me2 и H3K79me3 оказывается повышенным, а по H3K36me3 – пониженным в случае альтернативного сплайсинга, за исключением сохраняемых интронов, для которых плотность маркера H3K36me3 не отличается от таковой для канонических интронов.

В дальнейшем анализировалось маркирование тех участков ДНК, которые непосредственно формируют сайты сплайсинга. Сопоставляли маркирование донорных и акцепторных сайтов сплайсинга (рис. 6). Результаты получали в виде доли сайтов, обнаруживших перекрывание с пиками каждого типа маркера.

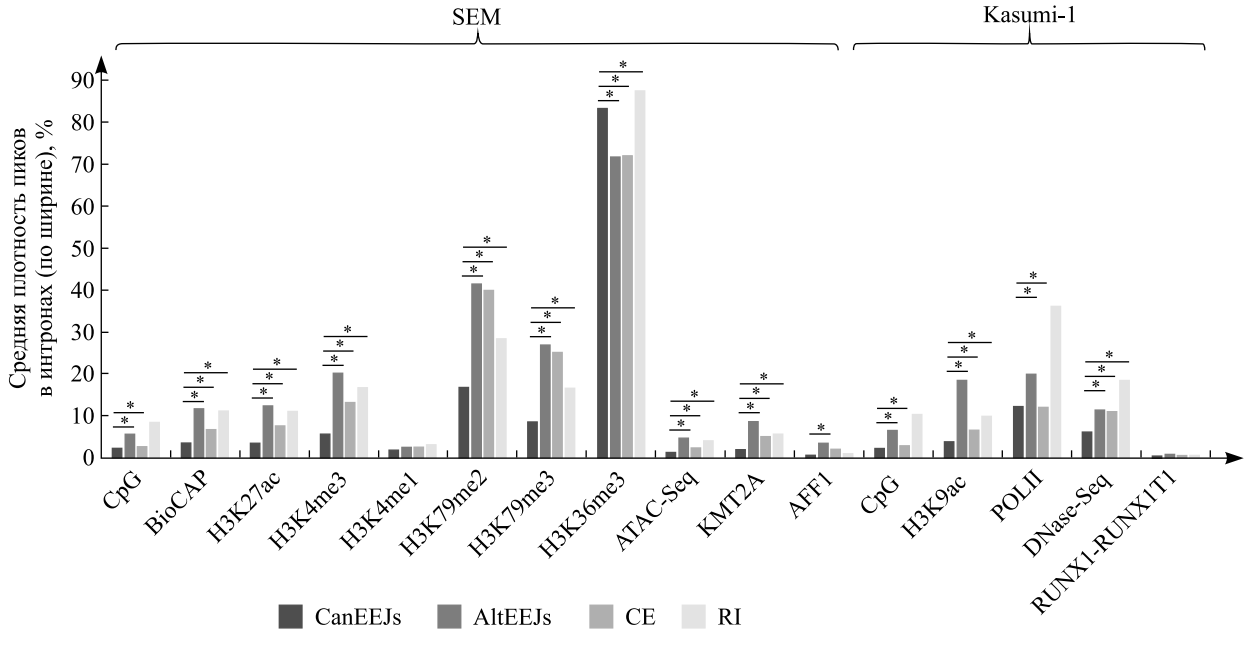

Рис. 5. Плотность пиков маркеров в интронах экспрессирующихся генов:
CanEEJs – интроны, образующиеся при канонических событиях сплайсинга;
AltEEJs – интроны, образующиеся при различных вариантах альтернативного сплайсинга;
CE – интроны, образующиеся при сплайсинге с пропуском кассетного экзона; RI – сохраняемые интроны;
* – различия в значениях по указанному маркеру достоверны при $p < 0{,}001$

Fig. 5. Density of the epigenetic marks in introns of expressed genes:
CanEEJs – introns formed during the canonical events of splicing;
AltEEJs – introns forming with different alternative splicing variants;
CE – introns formed during splicing with a cassette exon skip;
RI – retained introns; * – difference in value for the specified mark is significant at $p < 0.001$

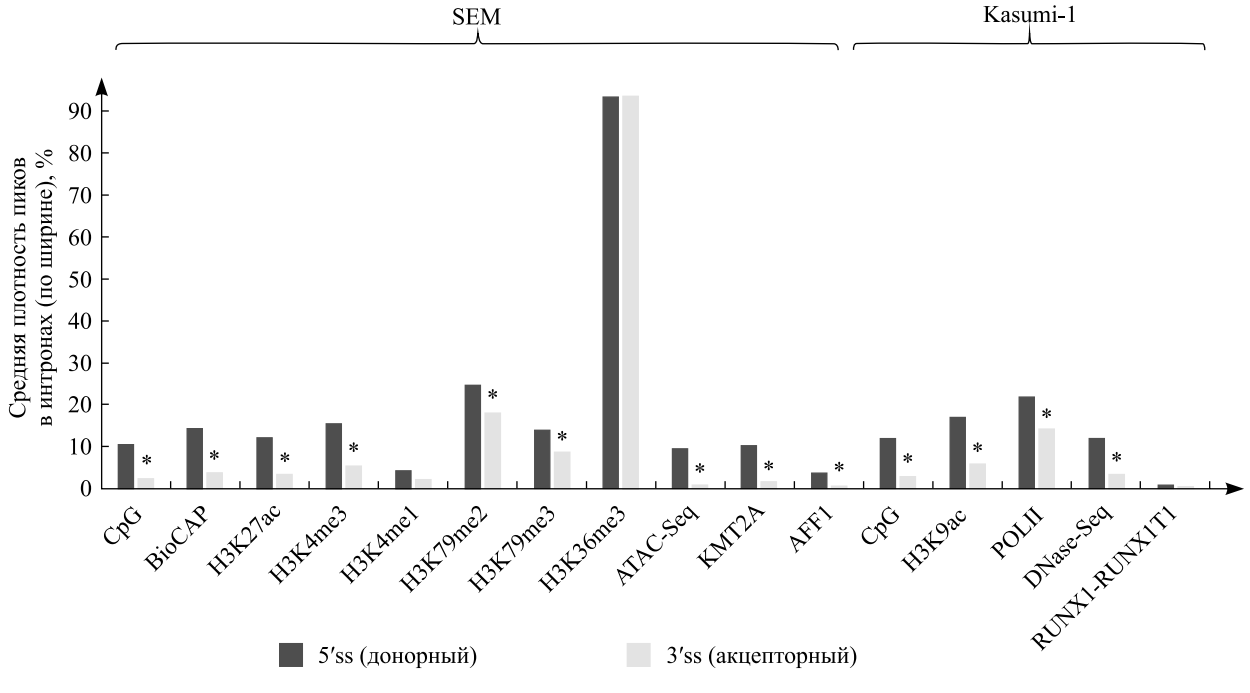

Рис. 6. Распределение эпигенетических маркеров
на донорных и акцепторных сайтах сплайсинга.
* – различия в значениях по указанному маркеру достоверны при $p < 0{,}001$

Fig. 6. Distribution of epigenetic markers at donor and acceptor splice sites.
* – difference in value for the specified mark is significant at $p < 0.001$

Единственным маркером, который был представлен на преобладающем количестве сайтов сплайсинга (как донорных, так и акцепторных), оказался H3K36me3. Остальные маркеры присутствуют не более чем на 22 % донорных и не более чем на 17 % акцепторных сайтов. Вместе с тем вероятность присутствия маркеров на донорных сайтах для всех маркеров, свойственных открытому хроматину, заметно и достоверно выше, чем для акцепторных сайтов. Эта закономерность показана здесь впервые и заслуживает особого внимания с точки зрения потенциально новых механизмов детерминации сайтов сплайсинга.

Также проведен отдельный анализ для донорных сайтов сплайсинга в области соединения первого и второго экзонов для всех генов, т. е. первого экзон-экзонного стыка. При этом сопоставлялось маркирование сайтов сплайсинга для канонических первых экзонов и альтернативных первых экзонов (рис. 7). Оказалось, что эти участки принципиально отличаются от общей выборки сайтов сплайсинга: здесь частота присутствия маркеров открытого хроматина оказывается гораздо более высокой (свыше 70 %), а маркер H3K36me3, наоборот, встречается значительно реже (примерно в 50 % сайтов). Это согласуется с литературными данными о том, что хроматин в начальных участках гена, как правило, находится в более открытом состоянии, чем в большей части своей длины.

Можно также обратить внимание на высокую частоту обнаружения в области первого EEJ белкового продукта гена KMT2A-AFF4 – эпигенетического активатора в линии SEM – и в то же время очень низкую частоту присутствия в таких участках гибридного белка RUNX1-RUNX1T1 (репрессора транскрипции) в Kasumi-1. Уровень маркера H3K4me1 также низкий – этот маркер несвойствен для хроматина в области экспрессионно активных генов.

Сравнивая данные по каноническим и альтернативным первым экзонам между собой, можно отметить достоверное снижение частоты маркеров активного хроматина и одновременно повышение частоты маркера H3K36me3 в случае вовлечения альтернативных первых экзонов. По-видимому, это соответствует преимущественно более высокой экспрессионной активности канонических первых экзонов.

Последние два графика (рис. 8 и 9) отражают сравнительный анализ маркирования донорных и акцепторных сайтов сплайсинга соответственно. В каждом случае сравнивали маркирование вовлеченных канонических и альтернативных сайтов сплайсинга, и в обоих случаях обнаружился достоверный перевес по частоте присутствия маркеров открытого хроматина в альтернативных сайтах по сравнению с каноническими, но разница гораздо более выражена для донорных сайтов (см. рис. 8), чем для акцепторных (см. рис. 9). Маркер H3K36me3 преобладал в канонических донорных сайтах по сравнению с альтернативными донорными сайтами сплайсинга, а для акцепторных разница по этому маркеру практически отсутствует.

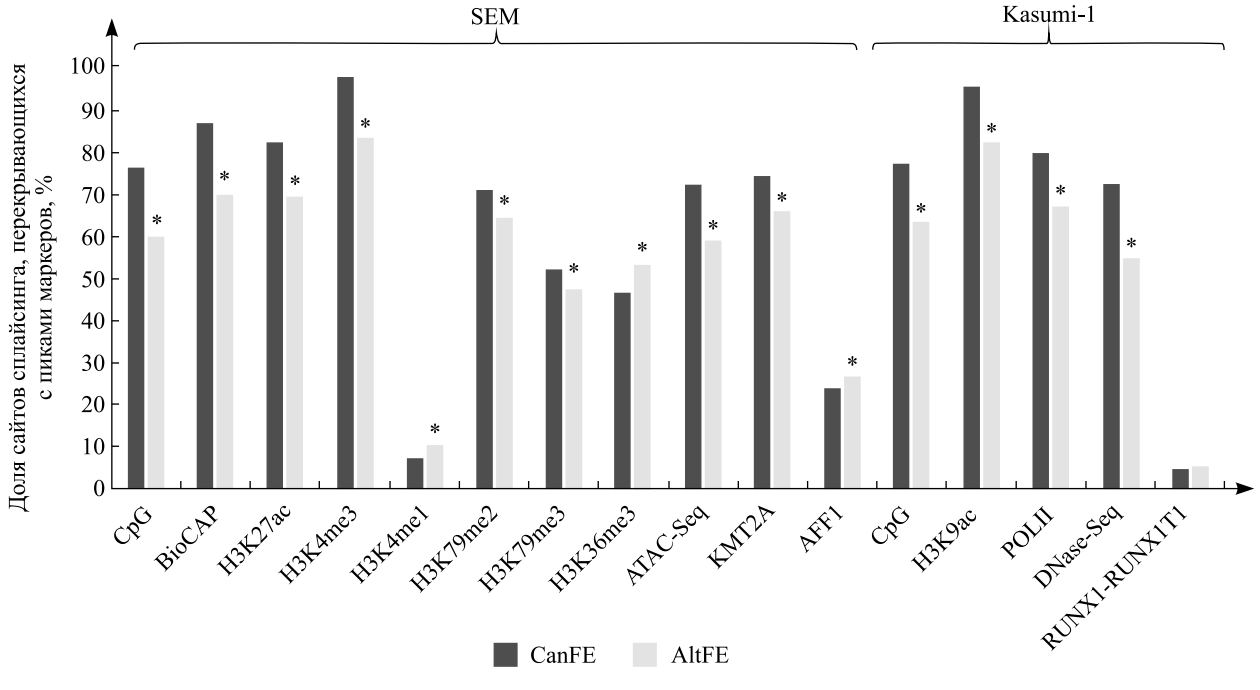

Рис. 7. Распределение эпигенетических маркеров на дистальных границах первых экзонов:
CanFE – канонический первый экзон; AltFE – альтернативный первый экзон;
* – различия в значениях по указанному маркеру достоверны при $p < 0{,}001$

Fig. 7. Distribution of epigenetic markers at the distal boundaries of the first exons:
CanFE – canonical first exon; AltFE – alternative first splicing exon;
* – difference in value for the specified mark is significant at $p < 0.001$

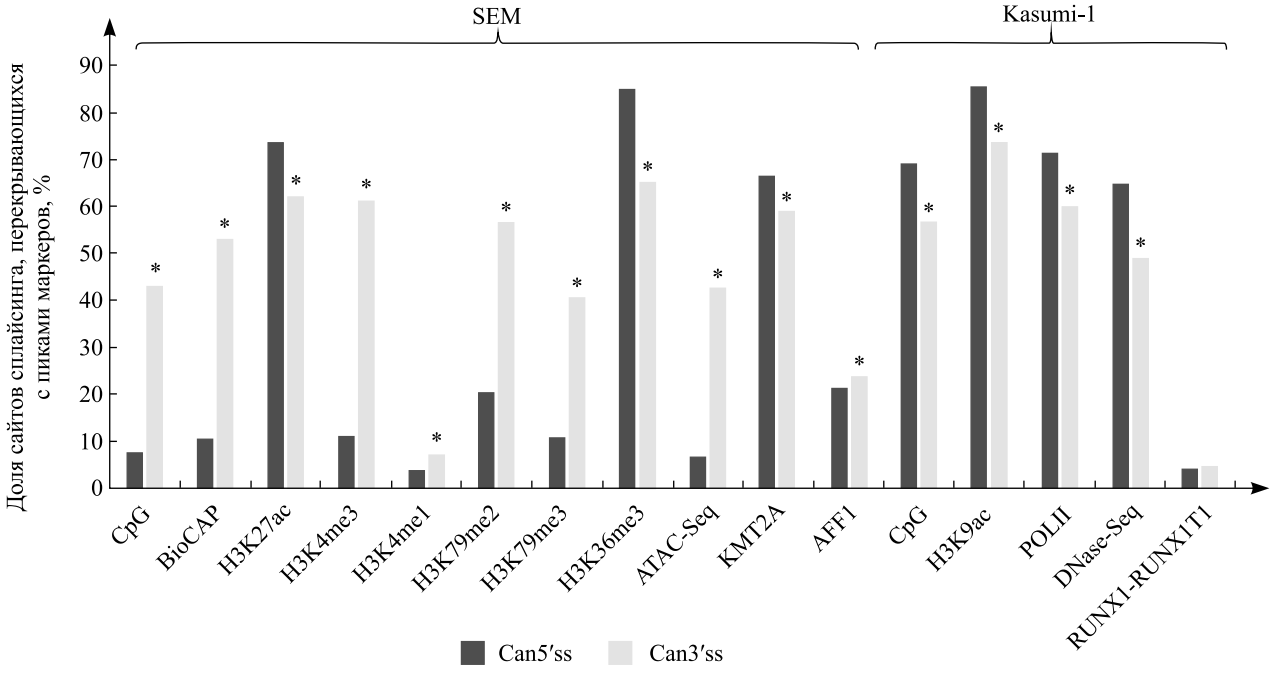

Рис. 8. Распределение эпигенетических маркеров в области донорных сайтов сплайсинга:
Can3′ss – канонический акцепторный сайт сплайсинга;
Alt3′ss – альтернативный акцепторный сайт сплайсинга;
* – различия в значениях по указанному маркеру достоверны при $p < 0{,}001$

Fig. 8. Distribution of epigenetic markers in the field of splice donor sites:
Can3′ss is the canonical acceptor splice site; Alt3′ss is the alternative acceptor site splicing;
* – difference in value for the specified mark is significant at $p < 0.001$

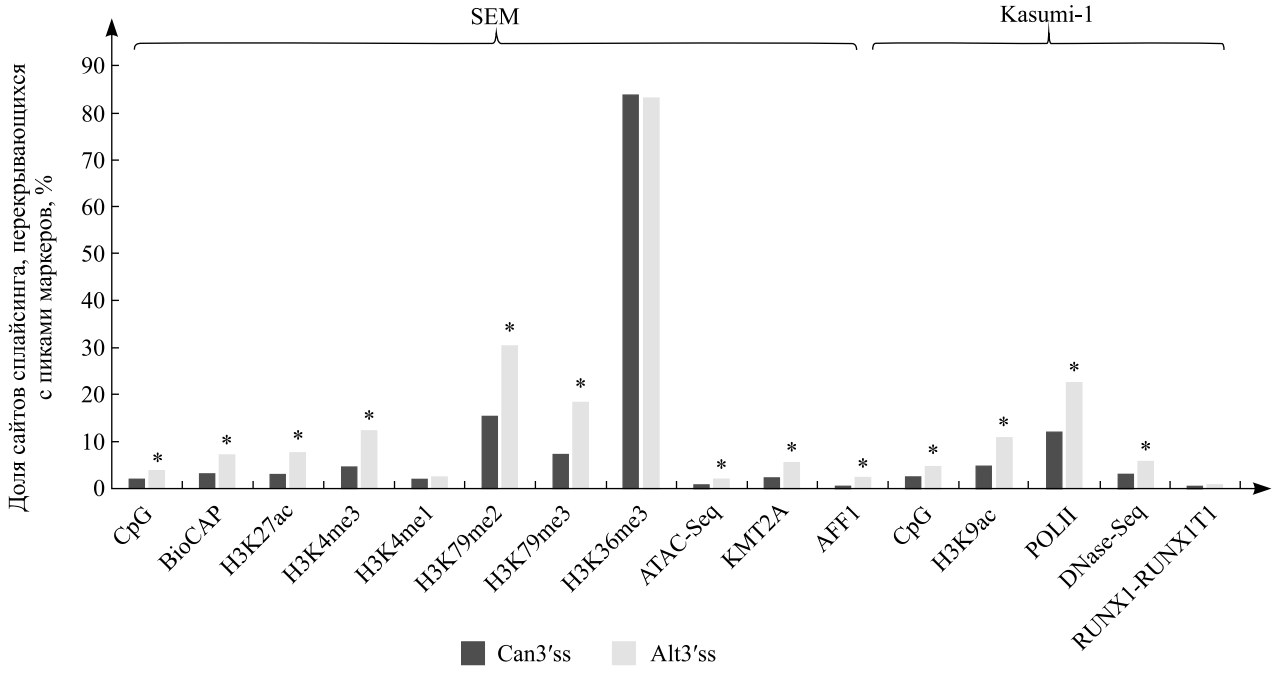

Рис. 9. Распределение эпигенетических маркеров в области акцепторных сайтов сплайсинга:
Can3′ss – канонический акцепторный сайт сплайсинга;
Alt3′ss – альтернативный акцепторный сайт сплайсинга;
* – различия в значениях по указанному маркеру достоверны при $p < 0{,}001$

Fig. 9. Distribution of epigenetic markers in the area of splicing acceptor sites:
Can3′ss is the canonical acceptor splice site;
Alt3′ss is the alternative acceptor site splicing;
* – difference in value for the specified mark is significant at $p < 0.001$

Заключение

Из результатов проведенных анализов можно вывести некоторые общие закономерности, связывающие между собой эпигенетические процессы и паттерны распределения событий сплайсинга.

1. Канонические и разные типы альтернативных событий сплайсинга количественно распределяются очень схожим образом в двух проанализированных клеточных типах, хотя они относятся к разным формам онкологических заболеваний крови и связываются с разными генетическими причинами: в линии SEM острого лимфобластного лейкоза транслокация формирует ген, работающий в качестве эпигенетического активатора, а в линии Kasumi-1 острого миелоидного лейкоза другая транслокация создает гибридный ген для белка с функцией эпигенетического репрессора.

2. Маркирование хроматина в области сайтов сплайсинга также демонстрирует схожие паттерны в двух анализируемых клеточных линиях. В частности, это касается особенностей распределения в области интронов и сайтов сплайсинга проанализированных маркеров открытого хроматина, таких как ацетилирование гистона 3 (анализировался маркер H3K27ac в линии SEM и H3K9ac в линии Kasumi-1), которое благоприятствует, как известно, переходу хроматина в более открытое состояние, а также тесты на чувствительность к транспозазе (ATAC-Seq) или ДНКазе (DNase-Seq) в линиях SEM и Kasumi-1 соответственно, которые являются двумя взаимозаменяемыми вариантами детекции открытого хроматина.

3. Выявлены следующие закономерности в распределении эпигенетических маркеров:

• присутствие маркеров открытого хроматина в целом нехарактерно для сайтов сплайсинга на границах внутренних экзонов, при этом донорные сайты сплайсинга содержат такие маркеры заметно чаще, чем акцепторные;

• маркеры открытого хроматина с существенно более высокой частотой присутствуют на альтернативных сайтах сплайсинга, чем на канонических;

• границы первых альтернативных экзонов, напротив, реже несут маркеры активного хроматина по сравнению с каноническими первыми экзонами;

• маркеры H3K36me3, H3K79me2 и H3K79me3 отличаются от остальных очень широкими пиками, покрывая собой значительную часть тела экспрессирующихся генов. Маркер H3K36me3 в большинстве случаев демонстрирует противоположную направленность изменения по сравнению с маркерами открытого хроматина в проведенных анализах. Это указывает на его возможное значение в качестве фактора, поддерживающего не слишком развернутое состояние хроматина в области тела гена и замедляющего процесс транскрипции. В то же время маркеры H3K79me2 и H3K79me3 изменяются в положительной корреляции с маркерами открытого хроматина. Их функция в качестве регуляторов транскрипции и сплайсинга пока недостаточно ясна и представляет большой интерес для будущих исследований.

Библиографические ссылки

1. Hyung D, Kim J, Cho SY, Park C. ASpedia: a comprehensive encyclopedia of human alternative splicing. *Nucleic Acids Research*. 2018 January 4;46(D1):D58–D63. DOI: 10.1093/nar/gkx1014.

2. Ramanouskaya TV, Grinev VV. The determinants of alternative RNA splicing in human cells. *Molecular Genetics and Genomics*. 2017 December;292(6):1175–1195. DOI: 10.1007/s00438-017-1350-0.

3. Grinev V, Ilyushonak I, Clough R, Nakjang S, Smink J, Martinez-Soria N, et al. RUNX1/RUNX1T1 controls alternative splicing in the t(8;21)-positive acute myeloid leukemia cells. *BioRxiv 628040* [Preprint]. 2019 [cited 2019 February 1]: [36 p.]. Available from: https://doi.org/10.1101/628040.

4. Listerman I, Sapra AK, Neugebauer KM. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nature Structural & Molecular Biology*. 2006 October;13(9):815–822. DOI: 10.1038/nsmb1135.

5. Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Research*. 2014;24:896–905. DOI: 10.1101/gr.171405.113.

6. Iwamori N, Tominaga K, Sato T, Riehle K, Iwamori T, Ohkawa Ya, et al. MRG15 is required for pre-mRNA splicing and spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2016 September 13;113(37):E5408–E5415. DOI: 10.1073/pnas.1611995113.

7. Sanidas I, Polytarchou C, Hatziapostolou M, Ezell SA, Kottakis F, Hu L, et al. Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Molecular Cell*. 2014 February 20;53(4):577–590. DOI: 10.1016/j.molcel.2013.12.018.

8. Kouwe E, Staber PhB. RUNX1-ETO: attacking the epigenome for genomic instable leukemia. *International Journal of Molecular Sciences*. 2019;20(2):350. DOI: 10.3390/ijms20020350.

9. Kerry J, Godfrey L, Repapi E, Tapia M, Blackledge NP, Ma H, et al. MLL-AF4 spreading identifies binding sites that are distinct from super-enhancers and that govern sensitivity to DOT1L inhibition in leukemia. *Cell Reports*. 2017 January 10;18(2):482–495. DOI: 10.1016/j.celrep.2016.12.054.

10. Гринев ВВ, Хайденрайх О. Нокдаун гибридного онкогена *KMT2A-AFF1* ассоциирован с дифференциальным сплайсингом РНК в клетках острого лимфобластного лейкоза человека. *Журнал Белорусского государственного университета. Биология*. 2017;3:21–27.

11. Ptasinska A, Assi SA, Mannari D, James SR, Williamson D, Dunne J, et al. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia*. 2012;26(8):1829–1841. DOI: 10.1038/leu.2012.49.

12. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*. 2013 May 1;41(10):e108. DOI: 10.1093/nar/gkt214.

13. Гринев ВВ. Сохранение интронов в транскриптоме лейкозных и нормальных клеток крови человека. *Молекулярная и прикладная генетика.* 2018;25:44–55.

14. Benito JM, Godfrey L, Kojima K, Hogda L, Wunderlich M, Geng H, et al. MLL-Rearranged Acute Lymphoblastic Leukemias Activate BCL-2 through H3K79 Methylation and Are Sensitive to the BCL-2-Specific Antagonist ABT-199. *Cell Reports*. 2015 December 29;13(12):2715–2727. DOI: 10.1016/j.celrep.2015.12.003.

15. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology*. 2013;9(8):e1003118. DOI: 10.1371/journal.pcbi.1003118.

16. Li T, Liu Q, Garza N, Kornblau S, Jin VX. Integrative analysis reveals functional and regulatory roles of H3K79me2 in mediating alternative splicing. *Genome Medicine*. 2018;10:30. DOI: 10.1186/s13073-018-0538-1.

References

1. Hyung D, Kim J, Cho SY, Park C. ASpedia: a comprehensive encyclopedia of human alternative splicing. *Nucleic Acids Research*. 2018 January 4;46(D1):D58–D63. DOI: 10.1093/nar/gkx1014.

2. Ramanouskaya TV, Grinev VV. The determinants of alternative RNA splicing in human cells. *Molecular Genetics and Genomics*. 2017 December;292(6):1175–1195. DOI: 10.1007/s00438-017-1350-0.

3. Grinev V, Ilyushonak I, Clough R, Nakjang S, Smink J, Martinez-Soria N, et al. RUNX1/RUNX1T1 controls alternative splicing in the t(8;21)-positive acute myeloid leukemia cells. *BioRxiv 628040* [Preprint]. 2019 [cited 2019 February 1]: [36 p.]. Available from: https://doi.org/10.1101/628040.

4. Listerman I, Sapra AK, Neugebauer KM. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nature Structural & Molecular Biology*. 2006 October;13(9):815–822. DOI: 10.1038/nsmb1135.

5. Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Research*. 2014;24:896–905. DOI: 10.1101/gr.171405.113.

6. Iwamori N, Tominaga K, Sato T, Riehle K, Iwamori T, Ohkawa Ya, et al. MRG15 is required for pre-mRNA splicing and spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2016 September 13;113(37):E5408–E5415. DOI: 10.1073/pnas.1611995113.

7. Sanidas I, Polytarchou C, Hatziapostolou M, Ezell SA, Kottakis F, Hu L, et al. Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Molecular Cell*. 2014 February 20;53(4):577–590. DOI: 10.1016/j.molcel.2013.12.018.

8. Kouwe E, Staber PhB. RUNX1-ETO: attacking the epigenome for genomic instable leukemia. *International Journal of Molecular Sciences*. 2019;20(2):350. DOI: 10.3390/ijms20020350.

9. Kerry J, Godfrey L, Repapi E, Tapia M, Blackledge NP, Ma H, et al. MLL-AF4 spreading identifies binding sites that are distinct from super-enhancers and that govern sensitivity to DOT1L inhibition in leukemia. *Cell Reports*. 2017 January 10;18(2):482–495. DOI: 10.1016/j.celrep.2016.12.054.

10. Grinev VV, Heidenreich O. The knockdown of the fusion oncogene *KMT2A-AFF1* is associated with differential RNA splicing in human acute lymphoblastic leukemia cells. *Journal of the Belarusian State University. Biology*. 2017;3:21–27. Russian.

11. Ptasinska A, Assi SA, Mannari D, James SR, Williamson D, Dunne J, et al. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia*. 2012;26(8):1829–1841. DOI: 10.1038/leu.2012.49.

12. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*. 2013 May 1;41(10):e108. DOI: 10.1093/nar/gkt214.

13. Grinev VV. Intron retention in the transcriptome of leukemic and normal human blood cells. *Molekulyarnaya i prikladnaya genetika*. 2018;25:44–55. Russian.

14. Benito JM, Godfrey L, Kojima K, Hogda L, Wunderlich M, Geng H, et al. MLL-Rearranged Acute Lymphoblastic Leukemias Activate BCL-2 through H3K79 Methylation and Are Sensitive to the BCL-2-Specific Antagonist ABT-199. *Cell Reports*. 2015 December 29;13(12):2715–2727. DOI: 10.1016/j.celrep.2015.12.003.

15. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology*. 2013;9(8):e1003118. DOI: 10.1371/journal.pcbi.1003118.

16. Li T, Liu Q, Garza N, Kornblau S, Jin VX. Integrative analysis reveals functional and regulatory roles of H3K79me2 in mediating alternative splicing. *Genome Medicine*. 2018;10:30. DOI: 10.1186/s13073-018-0538-1.

Статья поступила в редколлегию 25.05.2019.
Received by editorial board 25.05.2019.