

ГЕНОМ *DROSOPHILA MELANOGASTER*: СВЯЗЬ СОСТОЯНИЯ ХРОМАТИНА С РЕГУЛЯЦИЕЙ СПЛАЙСИНГА И ТРАНСКРИПЦИИ

В.Н. Бабенко¹, В.Ф. Матвиенко², И.А. Зыков³

¹ Учреждение Российской академии наук Институт цитологии и генетики
Сибирского отделения РАН, Новосибирск, Россия, e-mail: bob@bionet.nsc.ru;

² Государственное образовательное учреждение высшего профессионального образования
Новосибирский государственный университет, Новосибирск, Россия

³ Учреждение Российской академии наук Институт химической биологии
и фундаментальной медицины Сибирского отделения РАН, Новосибирск, Россия

При анализе хроматиновых модификаций был определен ряд свойств хроматина, присущего районам экзонов. Выявлена чрезвычайная изменчивость распределения модификаций гистонов в районах экзонов в зависимости от типа клеток. Был проведен анализ связи интенсивности альтернативного сплайсинга с определенным типом хроматина и идентифицированы его наиболее характерные типы. В качестве иллюстрации проведен анализ распределения хроматиновых меток в гене *DScam*, имеющем чрезвычайно большое количество транскриптов, оцененное в 38 тыс. Сделан вывод, что ген *DScam* комплексно регулируется состоянием хроматина на уровне как доменов, так и экзонов.

Ключевые слова: *Drosophila melanogaster*, геном, регуляция генов, экзон, интрон, альтернативный сплайсинг, транскрипция, хроматин, изменчивость.

Введение

Организация хроматина

Новые технологии геномного картирования существенно изменили взгляд на организацию хроматина у высших эукариот. Возникает представление, что мультигенные хроматиновые домены – скорее правило, чем исключение. Понимание организации хроматиновых доменов важно не только для фундаментального изучения регуляции генов, но также и для выяснения этиологии различных болезней. Например, рак часто сопровождается хромосомными транслокациями, которые разрушают хроматиновые домены. Как следствие, нарушается активность не только генов, затронутых транслокацией, но и соседних генов. Соответственно, вероятность последствий определенных хромосомных транслокаций может быть определена близостью вовлеченных локусов в ядре.

Детальная карта хроматиновых доменов может быть полезна для практического ис-

пользования в технологии контролируемой экспрессии трансгенов, например, в генной терапии. Экспрессия трансгенов может сильно меняться в зависимости от сайта встраивания. Хроматиновые домены – одна из наиболее вероятных причин (Babenko *et al.*, 2010). Селекция интеграции трансгенов в хроматиновые домены с предпочтительным временным или пространственным паттерном может помочь обеспечить оптимальную экспрессию трансгена.

Помимо протяженных мультигенных хроматиновых районов, упомянутых выше, существует также градация более мелкого масштаба. В частности, в данной работе будет показано, что такие небольшие по длине элементы генома, как экзоны, могут обладать состоянием хроматина, отличным от их окружения.

Позиционирование нуклеосом на экзонах

В 2009 г. рядом авторов был выявлен факт высокого сродства нуклеосом к районам экзонов (Andersson *et al.*, 2009; Nahkuri *et al.*, 2009;

Schwartz *et al.*, 2009; Spies *et al.*, 2009; Tilgner *et al.*, 2009). Этот факт был неожиданным, поскольку считалось, что **АТ-богатые районы**, т. е. районы интронов, имеют большее сродство к нуклеосомам. Выяснилось, что данное свойство (сродство нуклеосом к экзонам, ведущее к относительно стабильному их позиционированию) присуще организмам с длинными интронами, у которых происходит «экзонное» определение границ сплайсинга (**exon definition**; Berget, 1995), где инициация сплайсинга начинается с определения границ экзона, а не интрона. Суть данного процесса заключается в том, что маркируются границы экзонов с помощью U2 и U1 малых ядерных РНК, которые связываются между собой через экзон на преинициационном комплексе сплайсосомы, и лишь позже происходит замена связи на традиционную – через интрон (Schneider *et al.*, 2010).

Факт позиционирования нуклеосомы в районе экзона является чрезвычайно важным для сплайсинга и объясняет, каким образом происходит узнавание короткого (100–150 п.н.) экзона среди длинных, до 100–200 т.п.н., интронов. Имеются факты, подтверждающие, что «экзонные» нуклеосомы имеют специфическую модификацию гистонов (Huff *et al.*, 2010). В нашей работе мы выявили гистонные модификации, характерные для экзонов в геноме *Drosophila melanogaster*.

Модификации гистонов и состояния хроматина

Консорциум modENCODE, созданный в 2007 г. и работающий над аннотированием функциональных элементов ДНК в модельных организмах (Celniker *et al.*, 2009), подвел итоги в недавней публикации (**The modENCODE Consortium**, 2010). В частности, в процессе анализа полногеномного распределения сочетаний 18 гистоновых модификаций у *D. melanogaster* были выявлены 30 устойчивых состояний хроматина с наибольшим известным на сегодняшний день разрешением (до 200 п.н.) и 9 состояний хроматина для более крупного масштаба (до 1 т.п.н.) (**The modENCODE Consortium**, 2010). Каждое состояние хроматина характеризуется определенным поднабором гистонных модификаций, который был уточнен методом

главных компонент. Были рассмотрены мужская линия клеток эмбриональной стадии S2, а также мужская линия клеток ЦНС личиночной стадии BG3. Индекс состояния хроматина начинался с активного эухроматина (1) и кончался закрытым гетерохроматином (30 или 9, в зависимости от разрешения).

Аналогичную работу выполнили авторы из Голландии (Fillion *et al.*, 2010), но для определения состояний хроматина они взяли наряду с четырьмя модификациями гистона H3 сайты обогащения 49 хроматиновыми белковыми факторами, полученные в независимом эксперименте и другим методом, нежели в консорциуме modENCODE. Было выявлено 5 устойчивых состояний хроматина, обозначенных цветами (желтый, красный, зеленый, синий, черный). Красный и желтый хроматин соответствовал открытому хроматину (эухроматин), синий (**PcG-type**), зеленый (**HP1-type**) и черный (**SUUR-type**) – закрытому эухроматину. Районы отдельных типов хроматина не перекрывались и охватывали 99,8 % генома *D. melanogaster*. Была рассмотрена только одна клеточная линия – Kc167.

В целом обе работы хорошо соответствуют друг другу, в частности, закрытый и открытый эухроматин в обоих случаях определяется раздельными, т. е. неперекрывающимися, состояниями хроматина. Имеются также различия, заключающиеся, прежде всего, в разном количестве состояний хроматина из-за различного разрешения методов и наборов выбранных факторов.

Механизмы взаимодействия хроматина и сплайсинга

Сплайсинг – процесс вырезания белок-кодирующих районов РНК (интронов). Он обеспечивается сложным каскадом каталитических реакций при помощи сплайсосомы. Процесс сплайсинга, в результате которого структура зрелых мРНК различается по составу экзонов, называется альтернативным сплайсингом.

Кроме сплайсинга существуют другие процессы альтернативной транскрипции, вносящие вклад в генерацию вариантов мРНК. Известно, что тканеспецифическая экспрессия гена в значительной мере осуществляется выбором старта транскрипции (Shabalina *et al.*, 2010). В частности, наиболее частый тип вариации

транскриптов (более 50 % всей вариации) у *D. melanogaster* – это альтернативный старт транскрипции, за ним следует альтернативный сайт терминации транскрипции, затем – пропуск экзона (Lee *et al.*, 2004).

После обнаружения рядом авторов связи скорости транскрипции и паттерна сплайсинга, а также присутствия факторов сплайсинга на С-терминальном домене РНК-полимеразы II в процессе транскрипции (Alexander, Beggs, 2010), стало очевидно, что регуляция сплайсинга осуществляется котранскрипционно.

Намек на прямую роль модификаций гистонов в регуляции альтернативного сплайсинга возникает из сравнительного картирования набора гистоновых модификаций вдоль нескольких генов, у которых альтернативный сплайсинг зависит от белкового фактора сплайсинга, связывающегося с полипиримидиновым трактом в интроне (**polypyrimidine tract binding protein**, РТВ). Исследователи обнаружили сильную корреляцию между несколькими модификациями гистонов в альтернативно сплайсирующихся районах и результатом сплайсинга (Lusco *et al.*, 2010). Оказалось, что РТВ-зависимые гены были обогащены **H3K36me3** и обеднены **H3K4me3** в районах альтернативных экзонов. Экспериментальное изменение состава этих гистоновых маркеров было достаточно для переключения паттерна сплайсинга экзонов РТВ-зависимых генов. В данном случае **H3K36me3** выступает в качестве платформы для привлечения белка РТВ на пре-мРНК. Это происходит с помощью адаптерного белка MRG15, специфично связывающегося с **H3K36me3**. Таким образом, **H3K36me3**, MRG15, РТВ образуют адаптерную систему хроматин–сплайсинг.

Другим примером является человеческий ген *cyclin D1*. Хроматин-адаптерный белок CHD1 содержит хромодомен, специфично узнающий **H3K4me3**, которым обогащен *cyclin D1*. С другой стороны, CHD1 связывается с компонентами комплекса U2 мРНК, но не с U1 мРНК (Sims *et al.*, 2007). Уменьшение концентрации **H3K4me3** или CHD1 изменяет эффективность сплайсинга пре-мРНК гена и уменьшает связь комплексов SF3a и U2 мРНК с пре-мРНК.

Масс-спектрометрический анализ белков, связывающихся с **H3K9me3**, выявил хроматин-ассоциированный белок HP1 и факторы сплай-

синга Srp20 и ASF/SF2. Chip-эксперименты подтвердили, что HP1 связывается с ASF/SF2 в клетках человека (Loomis *et al.*, 2009), а в клетках мухи HP1 связывается с hnRNP-комплексами (Piacentini *et al.*, 2009). Эти результаты указывают на возможную роль **H3K9me3** в регуляции или привлечении факторов сплайсинга, опосредованных хроматин-адаптерным белком HP1, хотя функциональная связь пока не ясна. Наконец, возможны другие комбинации взаимодействующих модификаций гистонов и хроматин-связывающих белков и факторов сплайсинга.

Материалы и методы

Для анализа использовался геном *D. melanogaster* по аннотации версии 5.12 (www.flybase.org). В качестве альтернативных генов были отобраны локусы, содержащие транскрипты с пропуском (альтернативностью) экзона. Выборка альтернативно сплайсирующихся генов была скомпилирована на основе анализа транскриптов в каждом локусе и идентификации альтернативных экзонов. Было идентифицировано 4938 генов с альтернативным экзоном. Данные о состояниях хроматина были взяты из работ (The modENCODE Consortium, 2010). Общая выборка насчитывала 64210 экзонов из 21024 транскриптов (flybase v.5.12).

При анализе колокализации разных наборов состояний хроматина с экзонами экзон считался перекрытым определенным хроматином *i*, если он имел хотя бы 1 общий нуклеотид. Мы оценили частоту случаев перекрытия экзонов более чем одним типом хроматина. Обнаружено, что случаи перекрытия экзона одновременно тремя хроматинами для линии клеток S2 при максимальной 30-компонентной сегрегации хроматина (The modENCODE Consortium, 2010) практически не наблюдаются (34 случая из 64210 экзонов, < 0,063 %). Двумя хроматинами экзон перекрывается примерно в 20 % случаев (14429 из 64210 экзонов для Flybase 5.12). Для линии BG3 распределение одновременных перекрытий было аналогично. Таким образом, число различных состояний экзонных реализаций с учетом покрытия экзона одновременно несколькими хроматинами было 72639 в линии S2, и 71824 состояний – в линии BG3.

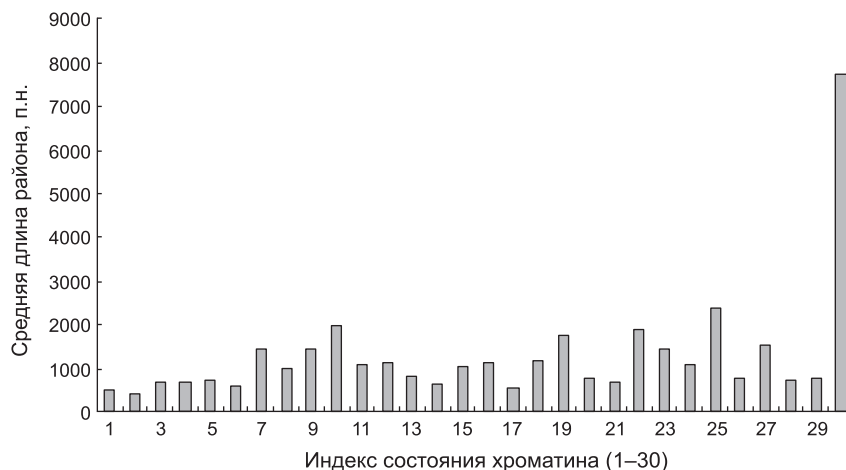


Рис. 1. Средняя длина районов состояния хроматина, определенных по 30 градациям (The modENCODE Consortium, 2010).

Основанием для такого критерия перекрытия (хотя бы один общий нуклеотид, см. выше) было распределение длин районов хроматина (рис. 1). Средняя длина районов, соответствующая различным типам хроматина, не превышает трех т.п.н., за исключением закрытого эухроматина (хром. 30). Мы не рассматривали районы более 13 т.п.н., которых было менее 1 % в максимальном случае (хром. 25).

Исходя из распределения длин районов на рис. 1, 2 сравнивались между собой только состояния 1–29. Исключительную по длине категорию 30 мы включали только для иллюстративности.

Результаты

Тканеспецифичная вариабельность гистоновых модификаций в районах генов (экзонов)

Были исследованы две клеточные линии S2 и BG3, использованные в работе консорциума (The modENCODE Consortium, 2010) для того, чтобы оценить вариабельность распределения состояний хроматина в генах. При анализе выяснилось, что распределения хроматина на одних и тех же экзонах отличаются между клеточными линиями BG3 и S2 примерно в 50 % случаев (34111 из 74210 случаев реализации экзона в транскрипте (таблица flybasegene, Fujita *et al.*, 2010)) при классификации 30 состояний. При анализе 9-компонентного хроматина

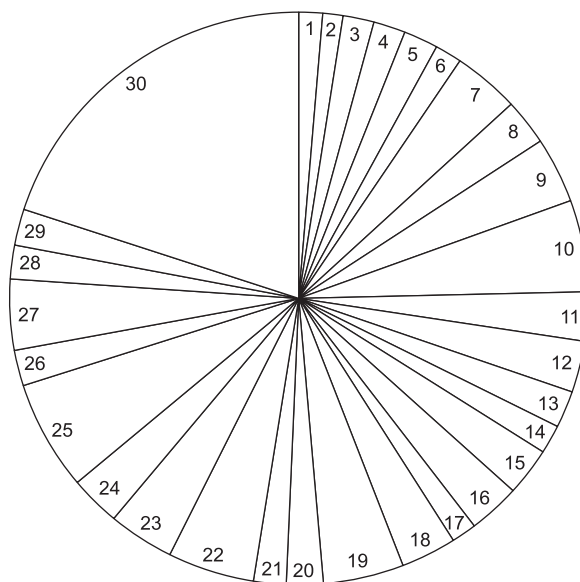


Рис. 2. Относительная доля каждого из состояний в геноме *D. melanogaster*.

различие линий было меньше – около 20 % (14760 из 67315). При анализе распределения изменений в зависимости от типа хроматина при сравнении двух клеточных линий S2 и BG3 (рис. 3) обнаружено, что при консервативности крайних случаев хроматина (выраженный эухроматин 1 и конститутивный гетерохроматин 30 в случае 30 компонентов) определенные состояния хроматина меняются очень значительно (рис. 3). Так как общая длина хроматина не дает полную картину транзиту типов хроматина от одной линии к другой, мы приводим таблицу (Приложение см. на сайте www.bionet.nsc.ru/vogis/pict_

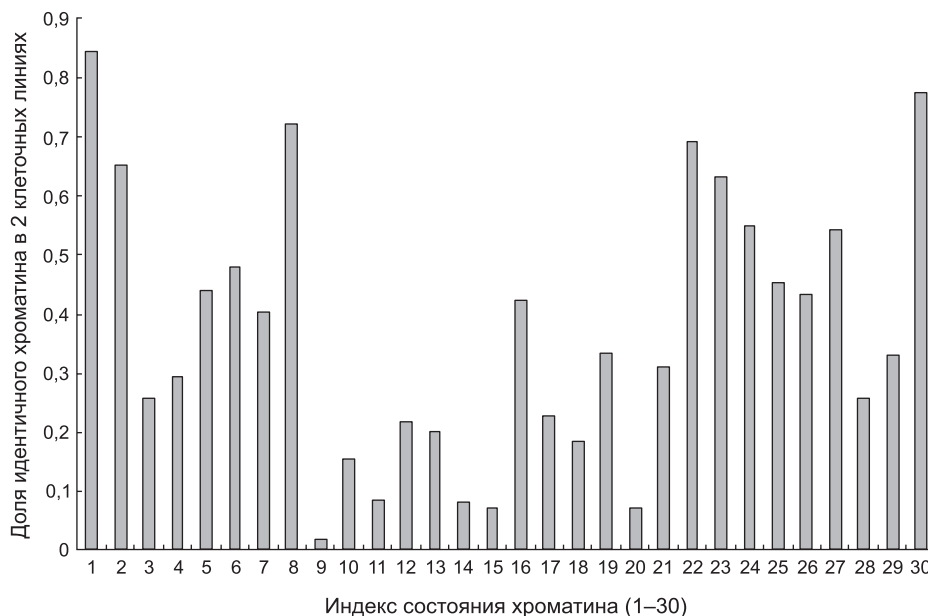


Рис. 3. Частота сохранения состояния хроматина при сравнении двух клеточных линий S2 и BG3 на экзонах при рассмотрении 30-компонентного хроматина.

pdf/2011/15_3/16_S1.pdf) для более подробного рассмотрения.

Тканеспецифичные сайты старта и терминации транскрипции

Мы решили выяснить, как модификации гистонов в районах сайтов стартов и терминации транскрипции варьируют в зависимости от клеточных линий. Для этого мы сравнили распределение числа стартов транскрипции по состояниям хроматина для двух клеточных линий BG3 и S2 (рис. 4, а, б). Видно более консервативное состояние хроматина сайтов терминации транскриптов в линии BG (большинство сайтов находятся в состояниях 1–8, 30). Состояние хроматина оценивалось для всех доступных транскриптов в базе данных Flybase 5.12. Состояние хроматина 30 соответствует гетерохроматину, занимающему 25 % генома (см. Материалы и методы), и потому избыточному в обоих случаях (а) и (б).

Можно также утверждать, что основным хроматиновым состоянием, в котором начинается транскрипция, является тип хроматина 1, соответствующий конститутивному эухроматину и находящийся в таком состоянии вне зависимости от времени и типа клеток. Это соответствует выводам предыдущей работы (The

modENCODE Consortium, 2010). Значительное число стартов транскрипции также расположены в хроматине 3, 5, 6, число которых, как видно из рис. 4, а, значительно варьирует в зависимости от типа клеток. Можно говорить, что обеспечение доступа к этим стартам зависит от типа ткани (Morillon *et al.*, 2003; Patraquim *et al.*, 2011).

Сайты терминации транскрипции в основном расположены в хроматине 8 (рис. 4, б) вне зависимости от типа клеток. Кроме хроматина 8, можно также отметить состояния 6, 7, 22 (рис. 4, б), в которых расположены сайты терминации транскрипции, но в гораздо меньшей мере. Стоит отметить, что состояния хроматина, характерные для сайта полиаденилирования и содержащие аналогичные метки, были найдены также у человека (Spies *et al.*, 2009; Guo *et al.*, 2011; Khaladkar *et al.*, 2011).

Идентификация хроматина, специфичного для экзонов

Одна из поставленных задач состояла в том, чтобы убедиться, насколько состояние хроматина может быть специфично для экзонов. С помощью пересечения районов состояний хроматинов с экзонами и последующей нормализацией на длину, занимаемую специфическим

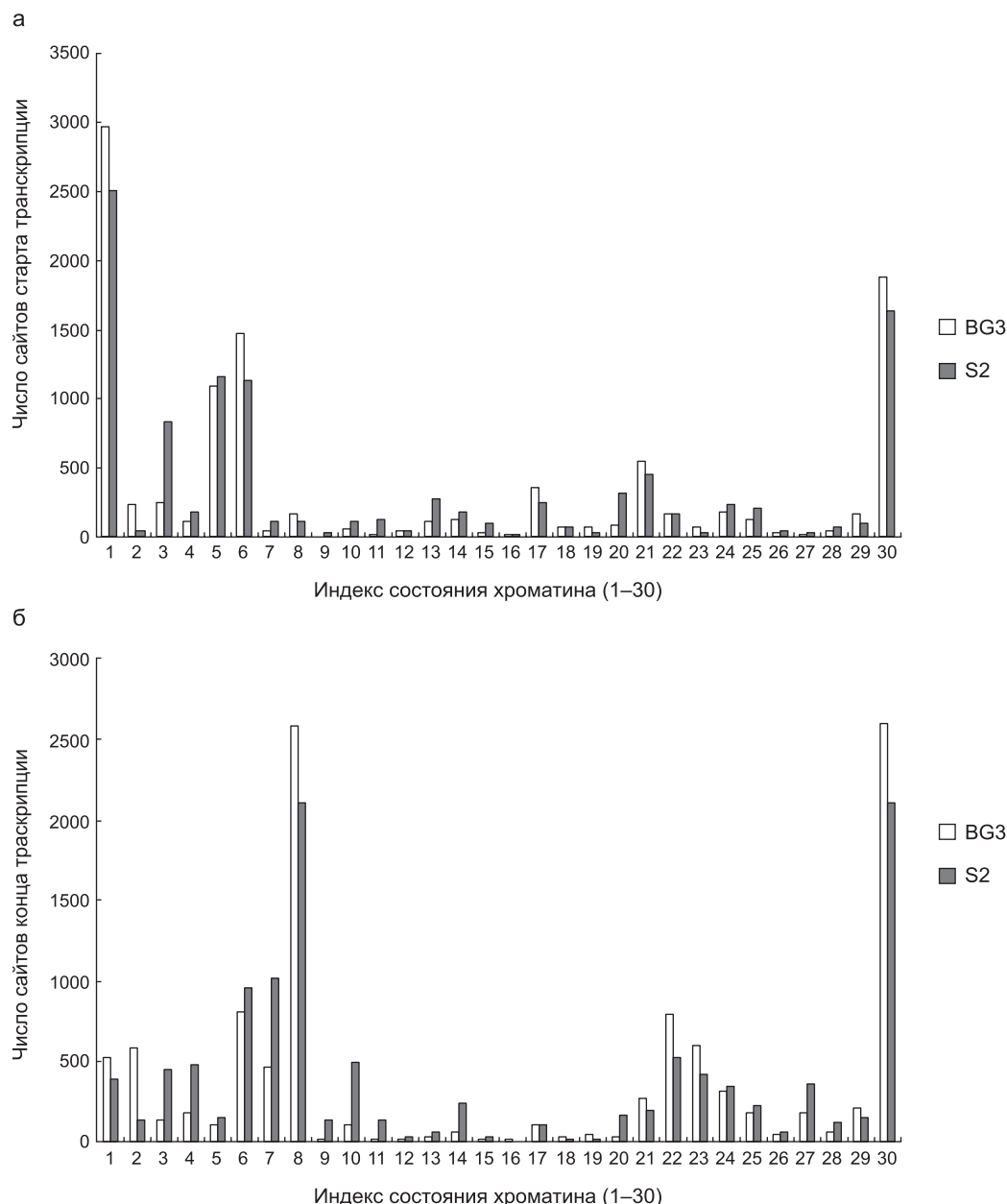


Рис. 4. Число сайтов старта (а) и терминации (б) транскрипции в зависимости от состояния хроматина и клеточной линии (BG3, S2).

хроматином, мы оценили плотность экзонов в каждом из видов хроматина (рис. 5).

При аннотации выяснилось, что большинство экзонов лежит в открытом хроматине. Заметим, что состояния хроматина, представленные на рис. 5, можно смело назвать экзон-специфичными, поскольку плотность экзонов варьирует от 15 до 65 %. Средняя длина экзона в геноме *D. melanogaster* равна 160 п.н., таким образом, имеет место прямая связь типа хроматина с наличием в районе экзона. Поясняя

феномен экзонного хроматина, стоит заметить, что размер одного района любого (кроме 30) состояния хроматина, равный 2–3 т.п.н. (рис. 1), приблизительно соответствует одному экзону в генах средней длины (> 3 т.п.н), таким образом, термин «плотность экзонов» более отражает метод его (хроматина) определения, а не физический феномен. Для того чтобы исключить из рассмотрения короткие (в том числе мульти-экзонные) гены, которые могут быть покрыты целиком одним хроматином, и, таким образом,

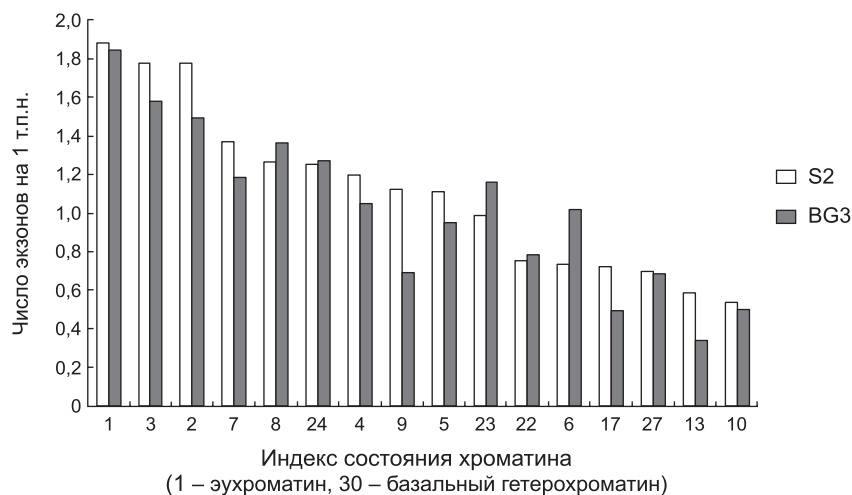


Рис. 5. Распределение плотности экзонов в зависимости от типа хроматина.

Приведены только типы хроматинов, имеющие плотность экзонов больше 0,5 на 1 т.п.н. (см. Приложение 1). Общее количество экзонов 64210. Индексы хроматина упорядочены по убыванию плотности экзонов.

искажить специфичность хроматина в экзонах, мы ограничили рассмотрение типов хроматина с плотностью экзонов не менее 0,5 экзона на 1 т.п.н. (рис. 5), что подразумевает наличие в среднем одного экзона в одном районе хроматина.

Связь альтернативно сплайсирующихся генов со специфичными типами хроматина (класс генов с возможным хроматин-зависимым сплайсингом)

Геном *D. melanogaster* был просканирован для выявления генов, которые альтернативно сплайсируются и имеют в своем составе экзоны со специфическим хроматином. В результате нами было выявлено 4938 генов с альтернативным сплайсингом (пропуском экзона; см. Материалы и методы). В табл. 1 приведены наиболее часто встречающиеся хроматины экзонов в этих генах.

Большинство генов с альтернативным пропуском экзона содержат хроматин-специфичные интронные энхансеры сплайсинга (табл. 1), что в целом соответствует недавно проведенному исследованию, выявившему энхансеры в интронах как основные сигналы тканеспецифичной экспрессии вариантов мРНК (Barash *et al.*, 2010). При этом стоит подчеркнуть, что экзонные хроматины, которые не были исследованы в упомянутой работе Barash

с соавт. (2010), по всей видимости, играют существенную роль в альтернативном сплайсинге (Pray-Grant *et al.*, 2005; Gunderson, Johnson, 2009; Li, Shogren-Knaak, 2009; Piacentini *et al.*, 2009; Lucio *et al.*, 2010).

Ген *DScam* и хроматин

Мы решили более подробно остановиться на возможном влиянии хроматина на альтернативный сплайсинг путем анализа гена *DScam* в линии S2.

Было рассмотрено, какие именно типы хроматинов из 30 возможных (The modENCODE Consortium, 2010) содержит ген *DScam* в линии клеток S2. В результате рассмотрения обнаружилось (рис. 6), что кассетные кластеры экзонов 4 и 6 (Graveley, 2008) принадлежат типу хроматина 19, при этом эти два кластера разделены конститутивным экзоном 5, принадлежащим к типу хроматина 10. Кластер кассетных экзонов 9 принадлежит к типу хроматина 9 (Приложение 3). Отличительная черта кластера экзонов 9 – в нем встречаются также хроматины 10, 28 и 29 в отдельных экзонах (рис. 6), что говорит о гетерогенности кластера экзонов 9 в отличие от экзонов 4 и 6.

Одной из основных характеристик экзонного хроматина является модификация H3K36me3 (The modENCODE Consortium, 2010; Приложение 2), хотя есть хроматины и из интронной груп-

Таблица 1

Типы хроматина, наиболее часто встречающиеся
в экзонах альтернативно сплайсирующихся генов

Частота альтернативного сплайсинга	Состояние хроматина	Число генов с альтернатив- ным экзоном	Общее число ге- нов, содержащих данный хроматин	Описание состояния хроматина*
0,40	13	112	278	Активный интрон/промотор
0,36	5	357	996	Активный TSS/экзон
0,31	15	70	229	Активный интрон/промотор
0,30	12	41	135	Активный интрон/промотор
0,27	16	13	49	Активный интрон/промотор
0,26	27	52	200	Гетерохроматин
0,25	17	115	454	Активный интрон/промотор
0,22	19	39	174	Открытый хроматин
0,21	23	91	425	Экзон самцовых генов X-хромосомы
0,21	11	80	379	Активный интрон/промотор
0,21	10	186	890	Активный экзон/ элонгация
0,20	20	145	709	Открытый хроматин
0,19	18	47	246	Открытый хроматин
0,19	14	81	431	Активный интрон/промотор
0,18	7	259	1405	Активный экзон/ элонгация

Примечание. Представлены категории типов хроматина, в которых доля экзонов со специфичным хроматином > 0,17 (столбец 1, посчитанный как доля генов в столбце 3 относительно генов столбца 4).

* См. Приложение 2.

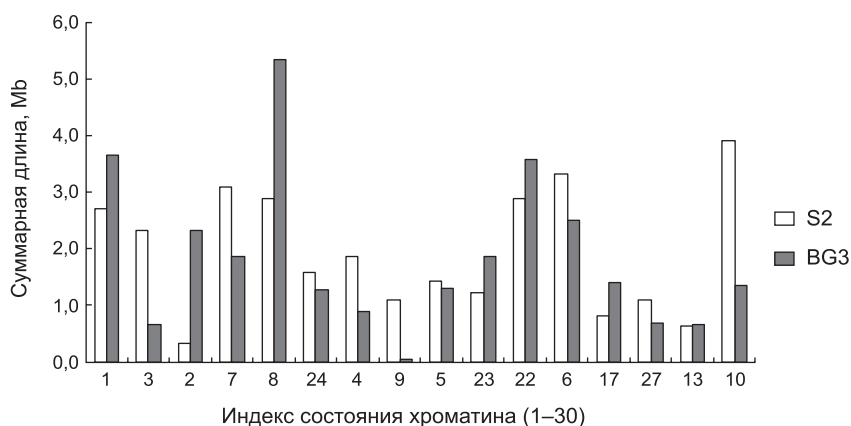


Рис. 6. Распределение суммарной длины хроматина, характерного для экзонов, в клеточных линиях в порядке убывания плотности экзонов (соответствует рис. 5).

пы (13, 17; рис. 5; Приложение 2) и даже гетерохроматиновой (23, 24; рис. 5; Приложение 2).

Хроматин типа 19 характеризуется наличием ацетилированных модификаций гистона H3 (H3K18ac, H3K22ac; рис. 7), с одной стороны, и практически полным отсутствием меток H3K36me3, H3K9me2 в отличие от других сравниваемых (экзонных) хроматинов, с другой

стороны, что говорит о специфической регуляции экзонов в этом хроматине (The modENCODE Consortium, 2010). Хроматины 9 и 19 являются узкоспециализированным хроматином – всего 768 районов и 902 соответственно. При этом хроматин 9 очень сходен по составу хроматиновых белков и гистоновых модификаций с хроматином 10, в то время как хроматин 19 значительно

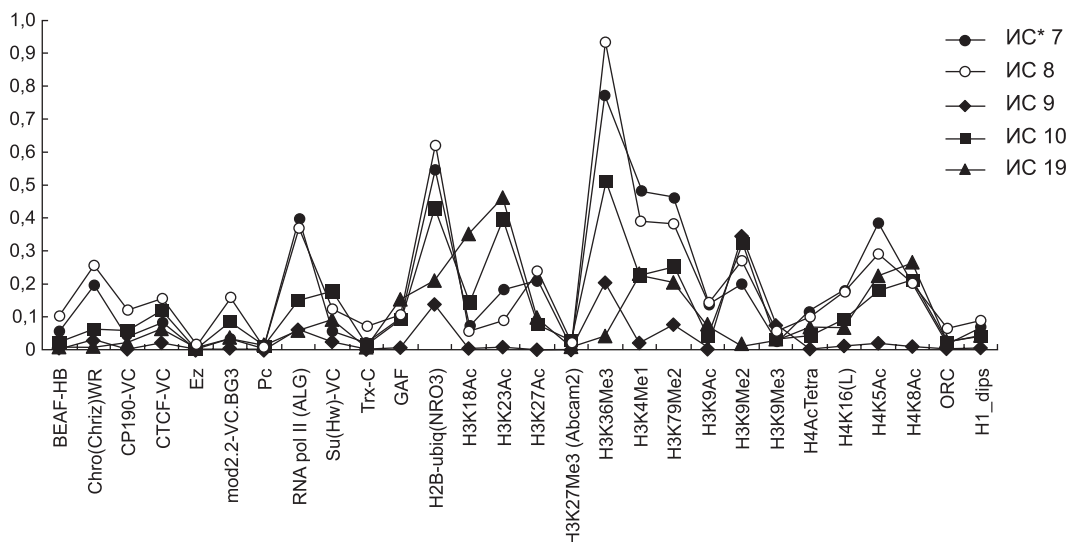


Рис. 7. Состав специфических состояний хроматина 7, 8, 9, 10 и 19, встречающихся в гене *DScam*, по хроматиновым белкам и гистоновым модификациям.

Виден характерный пик в районе H3K36me3 для экзонных хроматинов 7–10. * ИС – индекс состояния хроматина.

отличается от этих двух. Основная черта типа хроматина 9 – наличие экзон-специфичной метки H3K36me3 (The modENCODE Consortium, 2010; рис. 7). На то, что хроматин 9 является экзон-специфичным, указывает и плотность экзонов: 1 экзон на 1 т.п.н. хроматина (рис. 5; Приложение 1). При этом на то, что тип хроматина 19 не экзон-специфичный, указывает отсутствие метки H3K36Me3 и плотность экзонов в нем: 0,4 экзона на 1 т.п.н. (Приложение 1).

Необходимо отметить и тот факт, что в линии клеток BG3 хроматин в гене *Dscam* кардинально отличается (рис. 8). В частности, вместо хроматина 9 в кластерах экзонов 6–9 появляется непрерывный хроматин 10 (рис. 8), что говорит о совершенно другом статусе хроматиновой регуляции данного гена в линии клеток BG3, или, другими словами, транскрипция и сплайсинг *DScam* происходят совершенно по-другому.

Возможные механизмы регуляции сплайсинга на основе специфического хроматина

Из табл. 2, а видно, что наиболее часто встречающийся тип хроматина в экзонах – 9 (кластер экзонов 9, рис. 7), 19 (кластер экзонов 4 и 6; рис. 7), а также 10 (конститутивные экзоны). Из табл. 2, б видно, что хроматин влияет на инициацию, элонгацию и терминацию транс-

крипции, так как 5' UTR (хроматин 1) и 3' UTR (хроматин 8) хроматины качественно отличаются от хроматинов белок-кодирующей части гена (хроматин 9). В целом, интерпретируя данные табл. 2, б, можно описать структуру хроматина вдоль гена *DScam* следующим образом: открытый хроматин 1 в начале гена выступает вместе с регулируемым хроматином 13, обеспечивая тканеспецифичный старт транскрипции. Два последующих экзона также имеют уникальный для гена хроматин 11 и 12, который, по всей вероятности, также несет регуляторные функции. Начиная с 17-го экзона хроматин имеет метки поздне-replicative хроматина (табл. 2, б).

Для описания возможного механизма влияния хроматина на экспрессию обратимся к рис. 7. Как видно из рис. 7, хроматин 9, встречающийся наиболее часто в экзонах, является типичным экзонным хроматином с меткой H3K36me3 (Graveley *et al.*, 2010; The modENCODE Consortium, 2010; Kharchenko *et al.*, 2011). Хроматин 19, к которому относятся кассетные экзоны 4 и 6, не является экзон-специфичным хроматином и характеризуется ацетилированием гистона H3 (рис. 7). Набор кассетных экзонов 9 отличается от экзонов 4 и 6 и гетерогенен по типам экзонного хроматина: часть экзонов относится к хроматину 9, 10, а часть – к 28, 29. Примечательно, что один из экзонов с типом

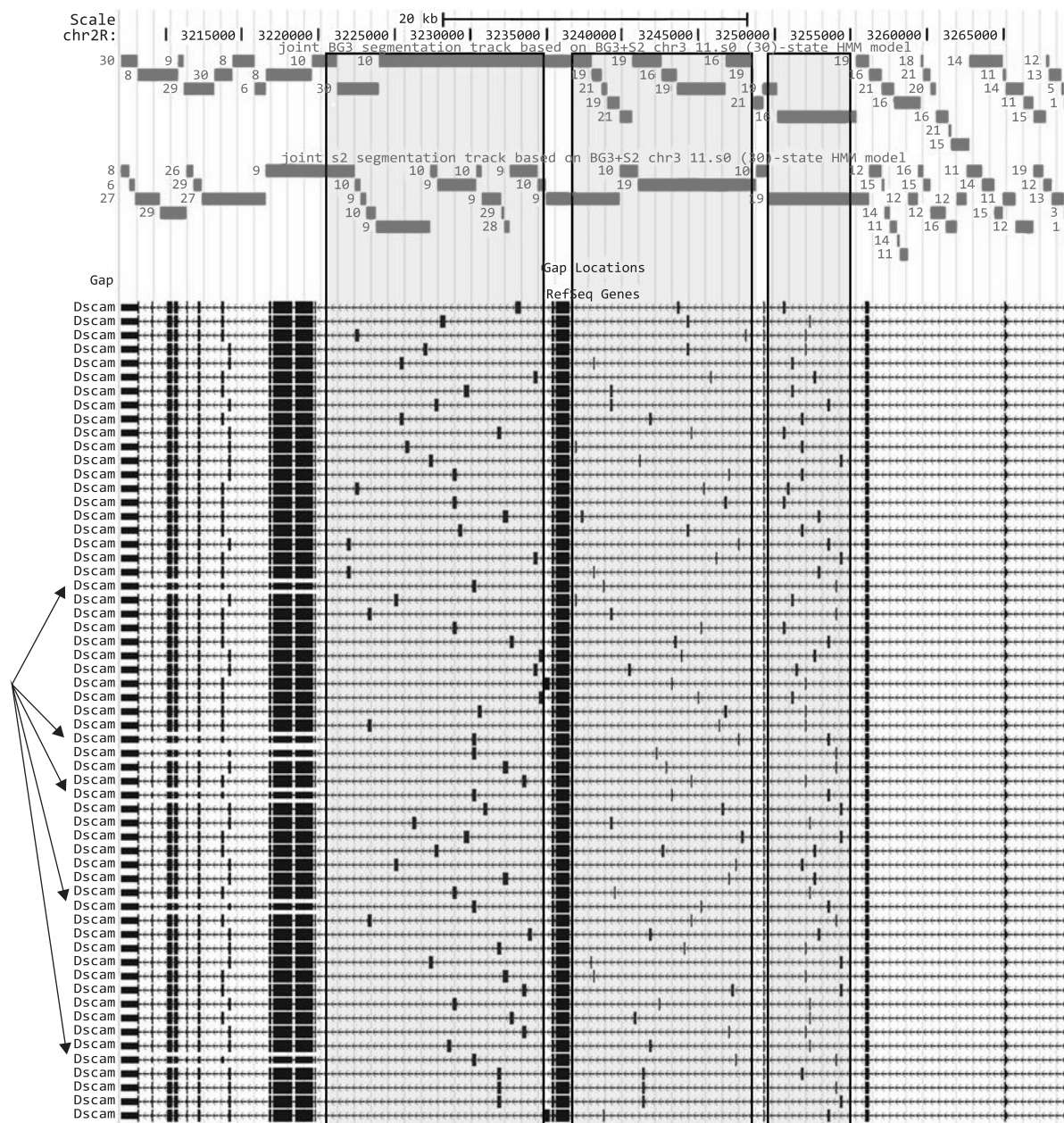


Рис. 8. Альтернативный сплайсинг транскриптов DScam (нумерация экзонов справа налево) и состояния хроматина (30-компонентная градация; The modENCODE Consortium, 2010).

Показано распределение типов хроматина для двух клеточных линий: BG3 (верхний серый трек) и S2 (нижний серый трек). Блоки экзонов 4, 6 и 9 показаны в рамках справа налево соответственно. Транскрипты с преждевременным стоп-кодом указаны стрелками (Fujita *et al.*, 2011).

хроматина 9 содержит преждевременный стоп-кодон (рис. 8) и, возможно, является фактором подавления экспрессии с помощью NMD (non-sense-mediated decay) (Hansen *et al.*, 2009).

Как видно из табл. 1, хроматин 19 можно отнести к регулируемому сплайсингу (частота альтернативного сплайсинга высокая), также экзоны с редким хроматином 9, 13 – наиболее

вероятные кандидаты на регуляцию тканеспецифичного сплайсинга. Хроматин 10, наблюдаемый в DScam, также имеет отношение к альтернативному сплайсингу (табл. 1).

Известно, что в гене DScam при сплайсинге экзонов 4 и 6 вблизи конститутивных экзонов 3 и 5 соответственно существуют докерные сайты, которые избирательно образуют комп-

Таблица 2, а

Состояние хроматина
и число уникальных экзонов в гене *DScam*

Индекс хроматина	Число различных экзонов
1	1
8	1
9	38
10	11
11	1
12	1
13	1
19	41
26	1
27	4
28	2
29	7

Примечание. Экзоны, имеющие отношение к альтернативному сплайсингу, выделены цветом.

лементарную связь с акцептором одного из кассетных экзонов (4) 6 (Graveley, 2008). Как видно из рис. 8, экзон 5 находится в экзонном типе хроматина 10 (экзон 3 – в типе хроматина 12), в то время как кассетные экзоны 4 и 6 расположены в «доменном» хроматине 19. Возможно, именно это состояние хроматина опосредует механизм комплементарного связывания «докер–акцептор» путем обеспечения оптимальной скорости РНК-полимеразы II для котранскрипционного сплайсинга, а также работы белковых факторов по выбору специфического экзона путем модификации вторичной структуры (May *et al.*, 2011).

Тканеспецифическая экспрессия генов в линиях BG3 и S2

В результате исследований было выявлено, что наиболее существенно меняются 16 состояний хроматина, содержащие экзоны с большой плотностью, или «экзонные» состояния хроматина (рис. 5, 6). Мы выявили достоверное изменение длин всех экзонных хроматинов при сравнении одной линии с другой. Помимо изменения общей длины, экзоны крайне интенсивно меняют состояния хроматина, находясь именно в этих состояниях (Приложение 3).

Таблица 2, б

Распределение типов хроматина
вдоль гена *DScam*

Индекс состояния хроматина	Номер экзона
1, 13	1
11	2
12, 19	3
19	4
10	5
10, 19	6
9	7
9	8
9, 10, 28, 29	9
9	10
9	11
9	12
9	13
9	14
9	15
9	16
27	17
29	18
29, 26	19
29	20
29	21
29	22
27	23
8, 27	24

Примечание. Серым выделены кластеры экзонов 4, 6, 9. Экзонный хроматин выделен жирным шрифтом.

Наиболее кардинальные изменения претерпевают следующие состояния хроматинов при рассмотрении S2→BG3 транзиции: 3 (S2) → 1, 2, 3 (BG3), 4 (S2) → 2, 4 (BG3), 9 (S2) → 8, 10, 23 (BG3), 15 → 11–16 (BG3), 19 (S2) → 12, 16, 19 (BG3). Заметим, что хроматин 9, который широко представлен в гене *DScam*, исчезает практически полностью в линии BG3 (рис. 6; Приложение 1), в то время как хроматин 2 ведет себя обратным образом.

В проведенном исследовании было подтверждено наличие специфического типа хроматина в районах сайтов терминации транскрипции (Spies *et al.*, 2009; The modENCODE Consortium, 2010; Khaladkar *et al.*, 2011). В

результате исследования состава хроматина 8 выяснилось, что он коррелирует с присутствием *chriz*-сайтов белка *chriz*, *polII* и отсутствием выраженных ацетилированных модификаций гистона H3 (рис. 7). Наиболее близким хроматином является хроматин 9 (рис. 7), но присутствие сайтов белков *chriz* и *polII* вносит основное отличие, характеризую открытый хроматин, с одной стороны, и задержку полимеразы и соответственно их повышенную концентрацию при окончании транскрипции, с другой стороны (Khaladkar *et al.*, 2011).

Как видно из результатов, существует значимая связь между типом хроматина экзона и его включением в зрелый транскрипт. Было проиллюстрировано распределение типов хроматина на примере гена *DScam*, одного из наиболее выдающихся генов в плане гипервариабельности экзонного состава транскриптов. Сделан вывод, что этот ген содержит специфичные состояния хроматина, имеющие отношение к производству специфических транскриптов. Дальнейшие исследования позволят выяснить непосредственные механизмы, обеспечивающие регуляцию сплайсинга специфическим состоянием хроматина.

Выводы

1. Показано наличие локальных, присущих экзонам, хроматиновых состояний – «экзонных» хроматинов (рис. 5).

2. «Экзонные» хроматины тканеспецифичны и наиболее подвержены изменению в зависимости от линии клеток (рис. 6). Они характеризуются маркером H3K36Me3 для типов хроматинов 1–10 и H3K16ac (самцовые X-сцепленные гены) для типов хроматинов 22–24, 27 (рис. 7).

3. Один из «экзонных» хроматинов (состояние 8) является мажорным хроматином (исключая состояние 30) по содержанию сайтов терминации транскрипции (рис. 4). Он характеризуется всего двумя гистонными модификациями (H3K36me3 и в меньшей степени H2-ubiquit.), а также коррелирует с районами связывания белка *chriz* и задержки *polII* (рис. 7).

4. Позиции старта и конца транскрипции сосредоточены в нескольких основных хроматинах в линии BG3. В линии S2 они находятся в более распределенном состоянии (рис. 4).

5. Докерные сайты кассетных экзонов 4 и 6 гена *DScam* лежат в районах хроматина 19, который, возможно, обеспечивает механизм связывания «докер–акцептор» (рис. 8) (May *et al.*, 2011).

Багодарности

Мы выражаем признательность Т. Колесниковой, А. Горчакову и А. Катохину за обсуждение и критические замечания во время написания статьи. Работа частично финансировалась Госконтрактом № 16.512.11.2274 Минобрнауки РФ.

Литература

- Alexander R., Beggs J.D. Cross-talk in transcription, splicing and chromatin: who makes the first call? // *Biochem. Soc. Trans.* 2010. V. 38. N 5. P. 1251–1256.
- Andersson R., Enroth S., Rada-Iglesias A. *et al.* Nucleosomes are well positioned in exons and carry characteristic histone modifications // *Genome Res.* 2009. V. 9. N 10. P. 1732–1741.
- Babenko V.N., Makunin I.V., Brusentsova I.V. *et al.* Paucity and preferential suppression of transgenes in late replication domains of the *D. melanogaster* genome // *BMC Genomics.* 2010. V. 11. P. 318.
- Barash Y., Calarco J.A., Gao W. *et al.* Deciphering the splicing code // *Nature.* 2010. V. 465. N 7294. P. 53–59.
- Berget S.M. Exon recognition in vertebrate splicing // *J. Biol. Chem.* 1995. V. 270. N 6. P. 2411–2414.
- Celniker S.E., Dillon L.A., Gerstein M.B. *et al.* modENCODE Consortium. Unlocking the secrets of the genome // *Nature.* 2009. V. 459. N 7249. P. 927–930.
- Fujita P.A., Rhead B., Zweig A.S. *et al.* The UCSC Genome Browser database // *Nucl. Acids Res.* 2011. 39 (Database issue). D876–882.
- Graveley B.R. Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures // *Cell.* 2005. V. 123. N 1. P. 65–73.
- Graveley B.R., Brooks A.N., Carlson J.W. *et al.* The developmental transcriptome of *Drosophila melanogaster* // *Nature.* 2011. V. 471. N 7339. P. 473–479.
- Gunderson F.Q., Johnson T.L. Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly // *PLoS Genet.* 2009. 5. p. e1000682.
- Guo J., Garrett M., Micklem G., Brogna S. Poly(A) signals located near the 5' end of genes are silenced by a general mechanism that prevents premature 3'-end processing // *Mol. Cell. Biol.* 2011. V. 31. N 4. P. 639–651.

- Hansen K.D., Lareau L.F., Blanchette M. *et al.* Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in *Drosophila* // PLoS Genet. 2009. V. 5. N 6. P. e1000525.
- Huff J.T., Plocik A.M., Guthrie C., Yamamoto K.R. Reciprocal intronic and exonic histone modification regions in humans // Nat. Struct. Mol. Biol. 2010. V. 17. N 12. P. 1495–1499.
- Khaladkar M., Smyda M., Hannenhalli S. Epigenomic and RNA structural correlates of polyadenylation // RNA Biol. 2011. V. 8. N 3. P. 1–9.
- Kharchenko P.V., Alekseyenko A.A., Schwartz Y.B. *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster* // Nature. 2011. V. 471. N 7339. P. 480–485.
- Lee B.T., Tan T.W., Ranganathan S. DEDB: a database of *Drosophila melanogaster* exons in splicing graph form // BMC Bioinformatics. 2004. V. 5. P. 189.
- Li S., Shogren-Knaak M.A. The Gcn5 bromodomain of the SAGA complex facilitates cooperative and cross-tail acetylation of nucleosomes // J. Biol. Chem. 2009. V. 284. P. 9411–9417.
- Loomis R.J., Naoe Y., Parker J.B. *et al.* Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation // Mol. Cell. 2009. V. 33. N 4. P. 450–461.
- Luco R.F., Pan Q., Tominaga K. *et al.* Regulation of alternative splicing by histone modifications // Science. 2010. V. 327. N 5968. P. 996–1000.
- May G.E., Olson S., McManus C.J., Graveley B.R. Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster // RNA. 2011. V. 17. N 2. P. 222–229.
- Morillon A., Karabetsov N., O'Sullivan J. *et al.* Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II // Cell. 2003. V. 115. P. 4. P. 425–435.
- Nahkuri S., Taft R.J., Mattick J.S. Nucleosomes are preferentially positioned at exons in somatic and sperm cells // Cell Cycle. 2009. V. 8. P. 3420–3424.
- Patraquim P., Warnefors M., Alonso C.R. Evolution of Hox post-transcriptional regulation by alternative polyadenylation and microRNA modulation within twelve *Drosophila* genomes // Mol. Biol. Evol. 2011. In press.
- Piacentini L., Fanti L., Negri R. *et al.* Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in *Drosophila* // PLoS Genet. 2009. V. 5. N 10. P. e1000670.
- Pray-Grant M.G., Daniel J.A., Schieltz D. *et al.* Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation // Nature. 2005. V. 433. P. 434–438.
- Schneider M., Will C.L., Anokhina M. *et al.* Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatytic splicing complexes // Mol. Cell. 2010. V. 38. N 2. P. 223–235.
- Schwartz S., Meshorer E., Ast G. Chromatin organization marks exon-intron structure // Nat. Struct. Mol. Biol. 2009. V. 16. N 9. P. 990–995.
- Shabalina S.A., Spiridonov A.N., Spiridonov N.A., Koonin E.V. Connections between alternative transcription and alternative splicing in mammals // Genome Biol. Evol. 2010. V. 2. P. 791–799.
- Sims R.J., Millhouse S., Chen C.F. *et al.* Recognition of trimethylated histone h3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing // Mol. Cell. 2007. V. 28. P. 665–676.
- Spies N., Nielsen C.B., Padgett R.A., Burge C.B. Biased chromatin signatures around polyadenylation sites and exons // Mol. Cell. 2009. V. 36. N 2. P. 245–254.
- The modENCODE Consortium, Roy S., Ernst J., Kharchenko P.V. *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE // Science. 2010. V. 330. N 6012. P. 1787–1797.
- Tilgner H., Nikolaou C., Althammer S. *et al.* Nucleosome positioning as a determinant of exon recognition // Nat. Struct. Mol. Biol. 2009. V. 16. N 9. P. 996–1001.

Приложение 1

Статистика состояний хроматина
по содержанию уникальных
экзонных последовательностей
для двух линий клеток S2 и BG3

Индекс состояния хроматина	Линия BG3			Линия S2		
	Число экзонов	Длина хроматина, Mb	Плотность экзонов, Kb	Число экзонов	Длина хроматина, Mb	Плотность экзонов, Kb
1	6760	3,66	1,85	5092	2,70	1,88
2	3464	2,32	1,49	575	0,32	1,77
3	1031	0,65	1,58	4117	2,32	1,78
4	927	0,88	1,05	2240	1,87	1,20
5	1240	1,31	0,95	1592	1,43	1,11
6	2549	2,51	1,02	2447	3,33	0,73
7	2216	1,87	1,19	4236	3,09	1,37
8	7291	5,35	1,36	3644	2,89	1,26
9	25	0,04	0,69	1225	1,09	1,13
10	675	1,35	0,50	2106	3,91	0,54
11	89	0,34	0,26	639	2,04	0,31
12	215	1,16	0,19	241	1,00	0,24
13	226	0,66	0,34	371	0,63	0,59
14	342	1,80	0,19	632	1,58	0,40
15	193	1,44	0,13	348	1,86	0,19
16	233	1,56	0,15	67	0,61	0,11
17	693	1,40	0,50	588	0,82	0,72
18	443	2,90	0,15	418	2,11	0,20
19	364	1,98	0,18	304	1,58	0,19
20	305	1,02	0,30	1112	2,62	0,42
21	2372	7,64	0,31	1264	3,06	0,41
22	2802	3,58	0,78	2191	2,90	0,76
23	2155	1,86	1,16	1212	1,23	0,99
24	1626	1,28	1,27	1992	1,59	1,26
25	1658	6,26	0,26	1903	7,06	0,27
26	528	2,61	0,20	632	2,93	0,22
27	468	0,68	0,68	758	1,09	0,70
28	662	1,96	0,34	765	2,06	0,37
29	1928	8,23	0,23	980	2,97	0,33
30	20730	61,37	0,34	20519	67,01	0,31

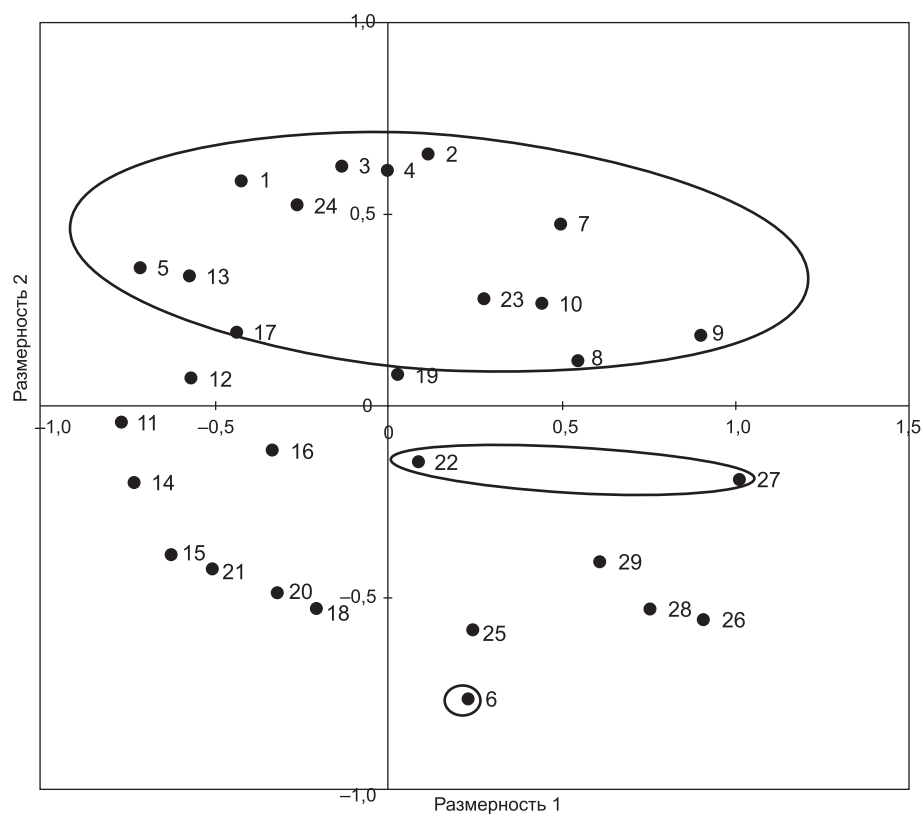
Примечание. Жирным шрифтом выделены строки, в которых плотность экзонов превышает 0,5 на 1 т.п.н. Различие между клеточными линиями в длине хроматина более чем в 2 раза выделено светло-серым цветом. Различие более чем в три раза выделено темно-серым.

Приложение 2

Обозначение состояний хроматина
(из: The modENCODE Consortium, 2010)

Активный TSS/экзон	1
	2
	3
	4
	5
	6
Активный экзон/элонгация	7
	8
	9
	10
Активный интрон/промотор	11
	12
	13
	14
	15
	16
	17
Открытый хроматин	18
	19
	20
	21
Экзон самцовых генов X-хромосомы	22
	23
	24
Специфичный хроматин	25
Гетерохроматин	26
	27
Интеркалярный гетерохроматин	28
	29
Базальный межгенный гетерохроматин	30

Приложение 3



Многомерное шкалирование 30-компонентного хроматина, построенного по распределению 18 гистонных факторов (The modENCODE Consortium, 2010). Экзон-специфичные состояния, при которых плотность экзонов в среднем по двум линиям больше 0,5 экзона/т.п.н., обведены овалом.

***DROSOPHILA MELANOGASTER* GENOME:
CORRELATION OF CHROMATIN STATE
WITH SPLICING AND TRANSCRIPTION REGULATION**

V.N. Babenko¹, V.F. Matvienko², I.A. Zykov³

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: bob@bionet.nsc.ru;

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia,

Summary

Features characteristic of exon regions have been revealed by analysis of chromatin modifications. The distribution of histone modifications in exon regions is broadly variable among cell types. The correlation of the rate of alternative splicing with a certain chromatin type has been analyzed, and its most characteristic types have been identified. The distribution of chromatin labels in the *DScam* gene, producing an enormous number of transcript, assessed to be 38000, is presented as an example. It is concluded that this gene undergoes integrated regulation by the chromatin state at the levels of domains and exons.

Key words: *Drosophila melanogaster*, genome, gene regulation, exon, intron, alternative splicing, transcription, chromatin, variability.