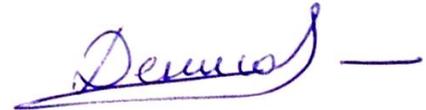


Федеральное государственное бюджетное учреждение науки

Институт проблем передачи информации имени А.А. Харкевича Российской академии наук

На правах рукописи



Денисов Степан Владимирович

Отбор и эпистаз в сайтах сплайсинга

03.01.09 – математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Москва, 2017

Работа выполнена в Учебно-научном центре "Биоинформатика" Федерального государственного бюджетного учреждения науки Института проблем передачи информации имени А.А. Харкевича Российской академии наук.

Научный руководитель:

Гельфанд Михаил Сергеевич

кандидат физико-математических наук,
доктор биологических наук, профессор
(Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации имени А.А. Харкевича Российской академии наук)

Официальные оппоненты:

Кочетов Алексей Владимирович

доктор биологических наук
(Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»)

Неверов Алексей Дмитриевич

кандидат биологических наук
(Федеральное государственное бюджетное учреждение науки Центральный НИИ эпидемиологии Роспотребнадзора)

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук

Защита диссертации состоится 24 апреля 2017 года в 15 часов на заседании диссертационного совета Д 002.077.04 при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации имени А.А. Харкевича Российской академии наук по адресу: 127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1.

С диссертацией можно ознакомиться в библиотеке ИППИ РАН, а также на сайте ИППИ РАН по адресу: <http://iitp.ru/ru/dissertation/1340.htm>

Автореферат разослан ____ 2017 г.

Ученый секретарь диссертационного совета

доктор биологических наук, профессор



Рожкова Г.И.

Общая характеристика работы

Актуальность работы. Прогресс современных методов секвенирования ДНК привел к прочтению полных геномов и транскриптомов многих высших эукариот. Сравнение геномов с транскриптомами позволяет установить экзон-интронную структуру генов, а значит и точно картировать сайты сплайсинга (Нуртдинов и др., 2006; Harrow et al., 2006, Flicek et al., 2013).

Несмотря на то, что структура и экспрессия мРНК генов человека и других высших эукариот исследованы достаточно подробно, вопрос о том, почему те или иные изоформы мРНК экспрессируются в одном физиологическом контексте (ткани, органе, на определенной стадии развития), но не экспрессируются в другом контексте, остаётся актуальным по настоящее время (Kornblihtt et al., 2013). По сути, вопрос в следующем: как регулируется альтернативный сплайсинг? На сплайсинг влияет множество факторов, важнейшими из которых являются сила сайтов сплайсинга и наличие специальных последовательностей в пре-мРНК – энхансеров и сайленсеров сплайсинга (цис-элементов). С цис-элементами связываются специальные белки и РНК-белковые комплексы (транс-факторы). С цис-элементами связываются специальные белки и РНК-белковые комплексы (транс-факторы). За счет тканеспецифичной экспрессии транс-факторов и осуществляется регуляция сплайсинга. В настоящей работе исследована одна из систем регуляции сплайсинга, основанная на цис-эlementе UGCAUG, и разработан сравнительно-геномный метод предсказания функциональности энхансеров UGCAUG.

Сравнение геномов родственных видов с другой стороны, открывает новые перспективы для эволюционной биологии. В настоящей работе было исследовано влияние отбора, генетического дрейфа и эпистаза на эволюцию сайтов сплайсинга.

Сайты сплайсинга представляют уникальный объект исследования в первую очередь из-за их значительного количества в геноме (несколько сотен тысяч), что позволяет исследовать тонкие эволюционные эффекты с разрешением вплоть до одного нуклеотида (Irimia et al., 2009). Был исследован слабый отбор, действующий на сайты сплайсинга. В частности впервые напрямую было показано существование слабополезных мутаций при неизменном ландшафте приспособленности, что было предсказано теоретически. Кроме этого, был исследован эпистаз между различными позициями в сайтах.

Цели и задачи исследования. Целью работы было изучить, каким образом естественный отбор, генетический дрейф и эпистаз между позициями влияют на эволюцию сайтов сплайсинга, а также изучить консервативность регуляции кассетных экзонов на примере цис-регуляторного элемента UGCAUG.

Были поставлены следующие задачи.

1. Оценить направление и силу отбора, действующего на консенсусные и неконсенсусные нуклеотиды в сайтах сплайсинга в линиях *Homo sapiens* и *Drosophila melanogaster*.
2. Выяснить, какие характеристики сайтов сплайсинга влияют на силу отбора.
3. Оценить силу отбора, действующего на молодые сайты сплайсинга (то есть появившиеся на линии человека после расхождения с макаком).
4. Оценить, как меняется сила отдельных позиций сайтов сплайсинга в ходе эволюции.
5. Разработать метод, позволяющий проверять гипотезу о независимости эволюции позиций сайтов сплайсинга.
6. Изучить, как замены в одних позициях сайтов сплайсинга влияют на замены в других позициях, и дать эволюционную интерпретацию найденным закономерностям.
7. Изучить консервативность цис-регулятора сплайсинга UGCAUG, и проанализировать, как консервативность этого регулятора соотносится с его функциональностью.

Научная новизна и практическая значимость. В работе рассмотрены актуальные вопросы современной эволюционной и сравнительной геномики. Впервые систематически исследован отбор, действующий на отдельные позиции сайтов сплайсинга. Подтверждено теоретически предсказанное существование слабopольных мутаций при неизменном ландшафте приспособленности.

Разработан и программно реализован метод, позволяющий оценить влияние, оказываемое заменами в одних позициях сайтов сплайсинга на замены в других позициях. Этот метод потенциально применим к любым наборам последовательностей, на которые действует схожий отбор (сайты связывания транскрипционных факторов, цис-регуляторы сплайсинга и т.п.). Впервые выявлено, что эпистаз является важнейшим фактором эволюции позиций в сайтах сплайсинга.

Впервые проанализирована консервативность цис-регуляторного мотива сплайсинга UGCAUG, на основе которой сделаны выводы о функциональности таких сайтов. Дальнейшие экспериментальные исследования подтвердили многие из этих выводов.

Степень достоверности и апробация результатов. По материалам диссертации опубликовано три статьи в рецензируемых научных журналах. Результаты работы были представлены на международных конференциях Society for Molecular Biology and Evolution 2012 и 2013 (SMBE 2012, SMBE 2013) и российских конференциях Информационные Технологии и Системы (ИТиС'12, ИТиС'13, ИТиС'14, ИТиС'15).

Структура и объем диссертации. Диссертация изложена на 166 страницах. Она состоит из восьми глав: введение, обзор литературы, данные и методы, результаты и обсуждение, основные результаты и выводы, благодарности, приложение и список

литературы. Работа содержит 27 рисунков и 5 таблиц. Список литературы содержит 218 наименований. Приложение содержит 11 рисунков и 2 таблицы.

Содержание работы

Сплайсинг – это процесс вырезания фрагментов из молекулы пре-мРНК с последующим сшиванием оставшихся фрагментов. Вырезанные участки называют интронами, а вошедшие в зрелую мРНК – экзонами. Для успешного вырезания интрона необходимо наличие донорного и акцепторного сайтов сплайсинга (на границе интрона и экзона с 5' и 3' концов интрона, соответственно), а также сайта ветвления. Сайты сплайсинга разных экзона и генов отличаются друг от друга и, соответственно, имеют разную энергию связывания со сплайсосомой. Однако, существуют довольно четкие предпочтения в каждой позиции сайтов сплайсинга (рис. 1). Соответственно, можно разделить нуклеотиды в каждой позиции сайта на консенсусные (часто встречающиеся) и неконсенсусные (редкие). Консенсусные последовательности сайтов сплайсинга одинаковы, по крайней мере, для всех многоклеточных животных и очень схожи среди эукариот. Мы оценили направление и силу отбора, действующего на консенсусные и неконсенсусные нуклеотиды в каждой позиции сайтов сплайсинга в линиях *Homo sapiens* и *Drosophila melanogaster* с момента их расхождения с линиями макака (*Macaca mulatta*) и *Drosophila simulans*, соответственно (и разработали соответствующие методы). Кроме того, мы разработали метод детекции эпистатического отбора на основе моделирования позиционно-независимой эволюции.

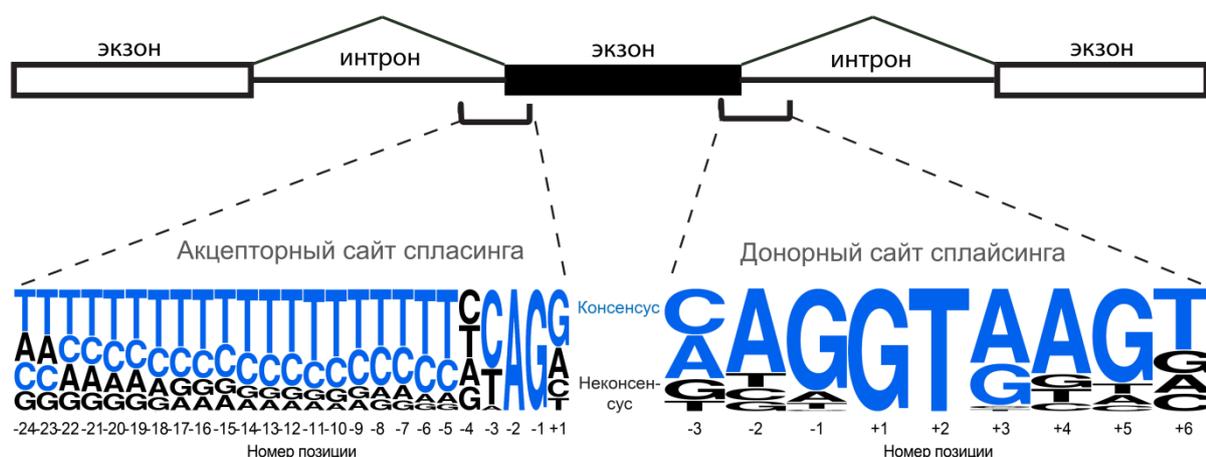


Рисунок 1. Консенсусные последовательности сайтов сплайсинга человека. Консенсусные нуклеотиды обозначены голубым цветом, неконсенсусные – черным. Высота буквы пропорциональна частоте её встречаемости.

Отбор, действующий на консенсусные и неконсенсусные нуклеотиды сайтов сплайсинга.

Для каждой позиции мы подсчитали наблюдаемую частоту замен между консенсусными (C_n) и неконсенсусными (N_c) нуклеотидами на линии *H. sapiens*

(после расхождения с линией *M. mulatta*) и на линии *D. melanogaster* (после расхождения с линией *D. simulans*): $q_{obs}(Cn \rightarrow Nc)$ и $q_{obs}(Nc \rightarrow Cn)$:

$$q_{obs}(Cn \rightarrow Nc) = \frac{\sum_{Z \in Cn, X \in Nc} \#(Z \rightarrow X)}{\sum_{Z \in Cn} \#(Z \rightarrow Y)} \quad , \quad q_{obs}(Nc \rightarrow Cn) = \frac{\sum_{Z \in Nc, X \in Cn} \#(Z \rightarrow X)}{\sum_{Z \in Nc} \#(Z \rightarrow Y)} \quad (1)$$

где $\#(Z \rightarrow X)$ и $\#(Z \rightarrow Y)$ – числа замен между нуклеотидом Z в геноме предка и нуклеотидом X (или Y , соответственно) в геноме потомка. Ожидаемые частоты замен $q_{exp}(Cn \rightarrow Nc)$ и $q_{exp}(Nc \rightarrow Cn)$ считались по этим же формулам, но вместо консенсусных/неконсенсусных нуклеотидов, определенных в конкретной позиции сайта сплайсинга, использовались формально те же нуклеотиды в нейтрально эволюционирующем участке генома (нейтральный контроль).

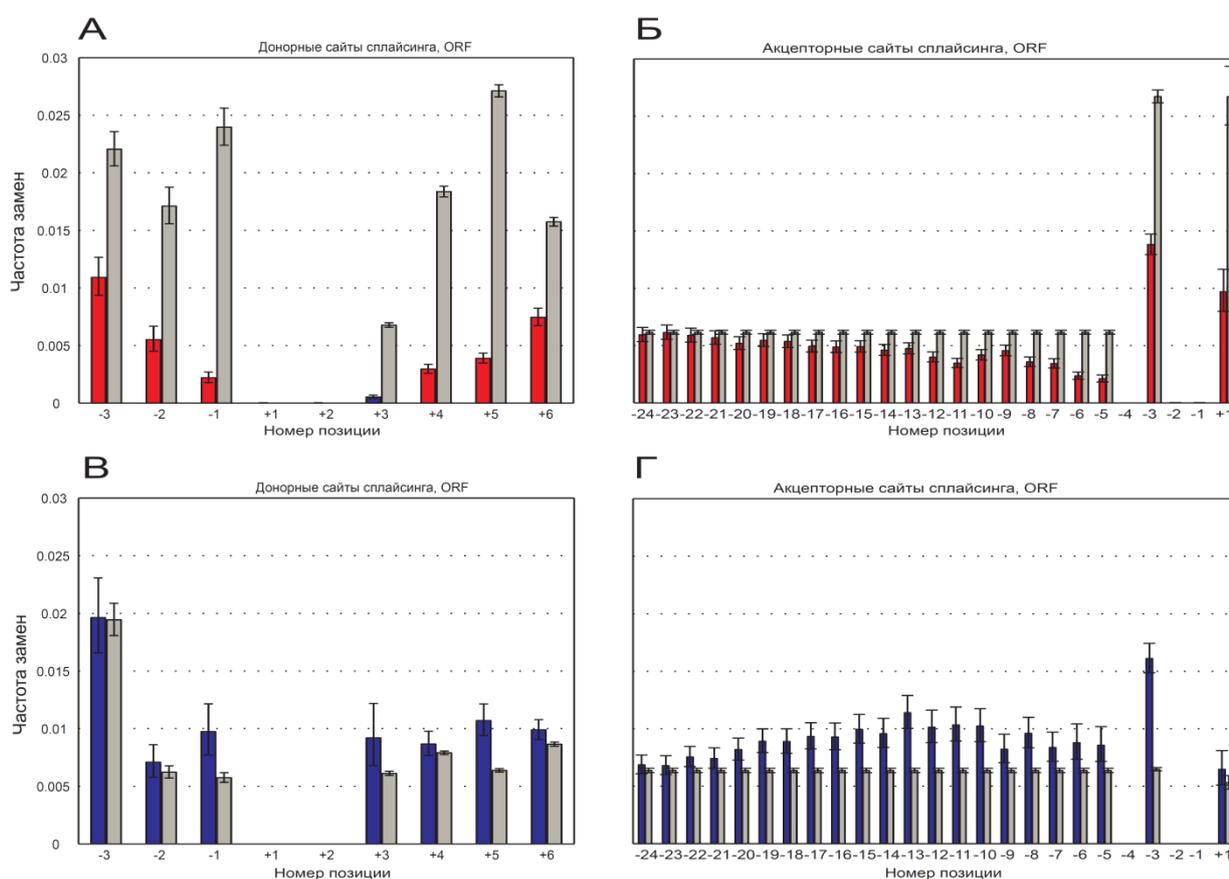


Рисунок 2. Частоты нуклеотидных замен в различных позициях сайтов сплайсинга на линии *Homo sapiens* после её дивергенции от линии *Macaca mulatta*.

По горизонтальной оси отложена позиция в сайте сплайсинга относительно инвариантного динуклеотида AG (GT). Верхний ряд (А–Г) – наблюдаемые (красные столбики) и ожидаемые (серые столбики) частоты замен $Cn \rightarrow Nc$. Нижний ряд (Д–З) – наблюдаемые (красные столбики) и ожидаемые (серые столбики) частоты замен $Nc \rightarrow Cn$. А, В – донорные сайты в кодирующей области генов; Б, Г – акцепторные сайты в кодирующей области генов. Усы – 95%-ные доверительные интервалы.

Тест на отбор состоял в следующем. Рассмотрим замены из неконсенсусных нуклеотидов в консенсусные в данной позиции сайта сплайсинга. Нулевая гипотеза (H_0) — на неконсенсусные нуклеотиды отбор не действует, т.е. все замены происходят благодаря мутациям и генетическому дрейфу. Имеется две альтернативные гипотезы: H_1 — на неконсенсусные нуклеотиды действует отрицательный отбор; H_2 — на неконсенсусные нуклеотиды действует положительный отбор, стремящийся заменить их на консенсусные. Если $q_{obs}(Nc \rightarrow Cn) < q_{exp}(Nc \rightarrow Cn)$, то верна H_1 ; если $q_{obs}(Nc \rightarrow Cn) > q_{exp}(Nc \rightarrow Cn)$, то верна H_2 . Если $q_{obs}(Nc \rightarrow Cn)$ статистически значимо не отличается от $q_{exp}(Nc \rightarrow Cn)$, то нельзя отвергнуть H_0 . Для замен $Cn \rightarrow Nc$ тест полностью аналогичен.

На рис. 2 частоты замен между консенсусными и неконсенсусными нуклеотидами для донорных и акцепторных сайтов сплайсинга, встречающихся внутри кодирующих последовательностей на линии *H. sapiens*. В табл. 1 приведены усреднённые по всем позициям частоты переходов для всех рассматриваемых выборок: донорные и акцепторные сайты, встречающихся внутри как кодирующих последовательностей, так и не кодирующих последовательностей (нетранслируемые области мРНК и не кодирующие РНК) в линиях *H. sapiens* и *D. melanogaster*.

Во всех позициях консенсусные нуклеотиды заменяются на неконсенсусные медленнее, чем в соответствующих нейтрально эволюционирующих последовательностях ($p < 10^{-18}$ по точному тесту Фишера, табл. 1, рис. 2). В большинстве позиций неконсенсусные нуклеотиды заменяются на консенсусные быстрее, чем в нейтральном контроле ($p < 10^{-3}$ для всех выборок, кроме донорных сайтов *D. melanogaster*, где разница статистически не значима, табл. 1). Это значит, что консенсусные нуклеотиды защищены отрицательным отбором, в то время как неконсенсусные нуклеотиды находятся под воздействием положительного отбора, стремящегося заменить их на консенсусные. В акцепторных сайтах сплайсинга сигнал как положительного, так и отрицательного отбора постепенно уменьшается в сторону 5'-конца интрона, становясь статистически неотличимым от нуля приблизительно в позиции -22 в человеческом геноме (в позиции -17 в геноме *D. melanogaster*), что, видимо, маркирует конец полипиримидинового тракта (рис. 2).

Также были изучены сегрегирующие однонуклеотидные полиморфизмы в сайтах сплайсинга конститутивных экзонов в популяциях *H. sapiens* и *D. melanogaster*. Мы различали предковый и производный аллель. Предковый аллель (нуклеотид) определялся как совпадающий с ортологичным нуклеотидом в геноме организма, являющегося внешней группой (*Pan troglodytes* для популяции *H. sapiens* и *D. simulans* для популяции *D. melanogaster*). Аналогично подходу с дивергенцией, мы рассматривали переходы из неконсенсусных нуклеотидов (предковый аллель) в консенсусные (производный аллель) и обратно. Полученные значения сравнивались

с таковыми в нейтральном контроле. Наблюдалась пониженная частота переходов из консенсусных нуклеотидов в неконсенсусные, что свидетельствует об отрицательном отборе. Обратные переходы в некоторых случаях быстрее, а в некоторых случаях не отличаются от нейтральных, что не противоречит теории, т.к. аллели под положительным отбором быстро фиксируются и долго не могут пребывать в сегрегирующем состоянии (*H. sapiens* - Рис. 3, *D. melanogaster* – данные в тексте диссертации).

Таблица 1. Сравнение частот переходов из неконсенсусных нуклеотидов в консенсусные и обратно в сайтах сплайсинга и нейтральных контролях.

Линия	Расположение сайтов сплайсинга	Тип сайтов	Cn->Nc			Nc->Cn		
			Частота набл. ⁵	Частота ожид. ⁶	р-значение ⁷	Частота набл.	Частота ожид.	р-значение
Человек	Код. посл. ¹	Дон. ³	0.0035	0.0154	0.00E+00	0.0099	0.0074	2.06E-25
		Акц. ⁴	0.0052	0.0069	3.75E-143	0.0090	0.0064	8.28E-126
	Некод. посл. ²	Дон.	0.0037	0.0162	8.84E-90	0.0159	0.0082	1.09E-15
		Акц.	0.0048	0.0078	1.35E-34	0.0109	0.0067	8.64E-24
<i>D. mel</i>	Код. посл.	Дон.	0.0104	0.0357	0.00E+00	0.0179	0.0180	8.29E-01
		Акц.	0.0196	0.0246	9.71E-102	0.0240	0.0225	2.99E-06
	Некод. посл.	Дон.	0.0049	0.0222	5.29E-55	0.0142	0.0120	1.44E-01
		Акц.	0.0113	0.0172	1.84E-19	0.0195	0.0160	3.85E-04

Код. посл.¹ - кодирующие последовательности белок-кодирующих генов; Некод. посл.² - некодирующие последовательности (5' и 3'- нетранслируемые области белок-кодирующих генов и некодирующие РНК); Дон.³ – донорные сайты сплайсинга; Акц.⁴ – акцепторные сайты сплайсинга; Частота набл.⁵ – частота переходов в сайтах сплайсинга (наблюдаемая); Частота ожид.⁶ – частота переходов в нейтральном контроле (ожидаемая); р-значение⁷ – р-значения точного теста Фишера (сравнение наблюдаемых и ожидаемых частот переходов).

Все полиморфизмы были разделены на три класса в соответствии частотой производного аллеля: (1) 1–10%, (2) 10–25% и (3) 25–99%. Были посчитаны частоты переходов между консенсусными и неконсенсусными нуклеотидами отдельно в каждом из классов. Оказалось, что чем более высокочастотный класс мы рассматриваем, тем сильнее частота переходов из неконсенсусных нуклеотидов в консенсусные отличается от нейтральной (ср. рис. 3А и В, рис. 3Б и Г). Это верно как для популяции *H. sapiens* (рис. 3) так и *D. melanogaster* (данные в тексте диссертации). Данное наблюдение согласуется с гипотезой о положительном отборе, т.к. производные консенсусные аллели, находящиеся в процессе фиксации, должны в среднем иметь более высокую частоту в популяции, чем нейтральные аллели.

Таким образом, наличие отрицательного отбора против неконсенсусных нуклеотидов и положительного отбора на вновь возникшие консенсусные нуклеотиды подтверждено как на межвидовом, так и на внутривидовом уровне.

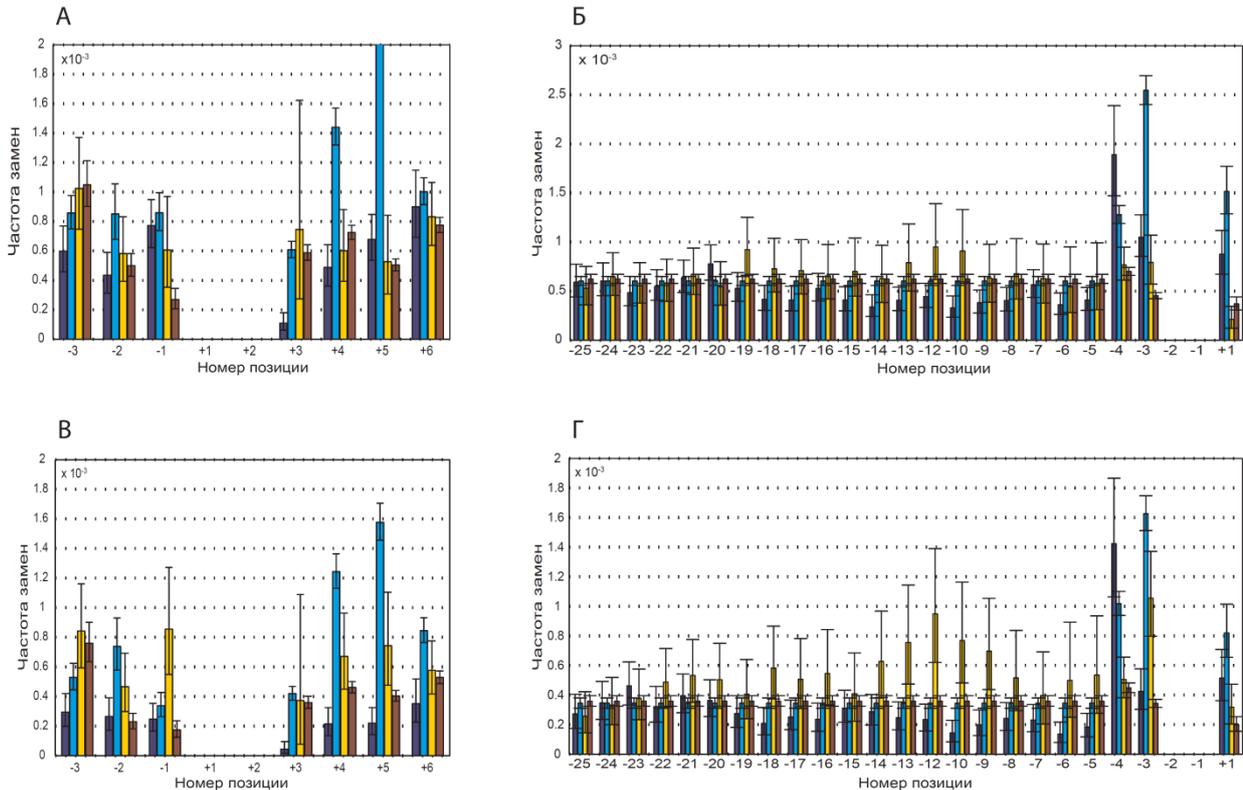


Рисунок 3 Сегрегирующие полиморфизмы в сайтах сплайсинга *H. sapiens*.

По горизонтальной оси – частота замен между предковыми и производными аллелями (рассматривались замены из Cn в Nc и обратно), по вертикальной оси – номер позиции в сайте сплайсинга. Изображена частота замен $Cn \rightarrow Nc$ в сайтах сплайсинга (синие столбики) и в близлежащих интронах (голубые столбики); замен $Nc \rightarrow Cn$ в сайтах сплайсинга (желтые столбики) и в близлежащих интронах (коричневые столбики). А, В – донорные сайты сплайсинга; Б, Г – акцепторные сайты. А, Б – частота производного аллеля в сайтах сплайсинга 1–10%; В, Г – частота производного аллеля в сайтах сплайсинга 25–99%.

Зная $q_{obs}(Cn \rightarrow Nc)$ и $q_{exp}(Cn \rightarrow Nc)$ (см. формулу (1)), можно вычислить силу (отрицательного) отбора (т.е. $4N_e s$), действующего на консенсусные нуклеотиды. Аналогично, можно вычислить силу (положительного) отбора на неконсенсусные нуклеотиды:

$$\frac{q_{obs}}{q_{exp}} = \frac{4N_e s}{1 - e^{-4N_e s}} \quad (2)$$

Если предположить, что соответствующие позиции всех сайтов сплайсинга имеют приблизительно одинаковый ландшафт приспособленности, ожидается, что абсолютные значения $4N_e s$ для положительного отбора, действующего против неконсенсусных нуклеотидов, и отрицательного отбора, сохраняющего консенсусные нуклеотиды, должны быть равны. Эта гипотеза была проверена в настоящей работе.

На рис. 4 представлены рассчитанные по формуле (2) значения $4N_e s$ для отбора, действующего на консенсусные и неконсенсусные нуклеотиды в разных позициях

сайтов сплайсинга. Как и ожидалось, отбор во всех случаях слабый ($1 < |4N_e s| < 4$). Для акцепторных сайтов сплайсинга в целом сила отрицательного отбора, защищающего консенсусные нуклеотиды, приблизительно равна силе положительного отбора против неконсенсусных нуклеотидов. Это согласуется с первоначальной гипотезой: слабый отбор действует на сайты с приблизительно одинаковым ландшафтом приспособленности. Слабовредные мутации (неконсенсусные нуклеотиды) фиксируются благодаря генетическому дрейфу, после чего на них начинает действовать слабый положительный отбор, стремящийся зафиксировать в популяции слабополезные мутации (консенсусные нуклеотиды). Возникает мутационно-селекционно-дрейфовое равновесие. Однако во многих позициях сила отрицательного отбора превосходит силу положительного. Это верно для позиций -6 , -5 и $+1$ акцепторных сайтов сплайсинга и почти для всех позиций донорных сайтов сплайсинга, в особенности для кодирующих сайтов в человеческой линии и для некодирующих сайтов сплайсинга в линии *D. melanogaster*. Последнее наблюдение приводит к предположению, что по крайней мере в некоторых сайтах сплайсинга неконсенсусные нуклеотиды находятся под действием отрицательного отбора, и примесь таких сайтов снижает среднюю силу отбора, измеренную по общей выборке. Мы проверили это предположение (см. далее).

Различия в силе отбора между разными классами сайтов сплайсинга, сайт-специфический отбор

Чтобы лучше разобраться, как отрицательный отбор воздействует на неконсенсусные и консенсусные нуклеотиды, их многовидовая консервативность сравнивалась с таковой для нейтрально эволюционирующих последовательностей. Была оценена консервативность нуклеотида, присутствующего в геноме *H. sapiens* или *D. melanogaster*, анализируя выравнивания с геномами 45 других видов позвоночных или 15 видов насекомых, соответственно.

Мы рассматривали относительную консервативность c_{nuc} конкретного консенсусного (неконсенсусного) нуклеотида, присутствующего в данном сайте сплайсинга в геноме *H. sapiens* или *D. melanogaster*, а также относительную консервативность c_{set} всех нуклеотидов, определяемых как консенсусные (или неконсенсусные).

Как и ожидалось, консенсусные нуклеотиды, наблюдаемые в геноме *H. sapiens* или *D. melanogaster* всегда более консервативны, чем нейтрально эволюционирующие последовательности. Более того, в большинстве позиций с двумя консенсусными нуклеотидами (см. рис 2), средняя консервативность c_{set} двух консенсусных нуклеотидов выше, чем c_{nuc} конкретного консенсусного нуклеотида, присутствующего в сайте сплайсинга, что говорит о том, что отбор предпочитает любой консенсусный нуклеотид, не пытаясь сохранить какой-то конкретный (данные приведены в тексте диссертации).

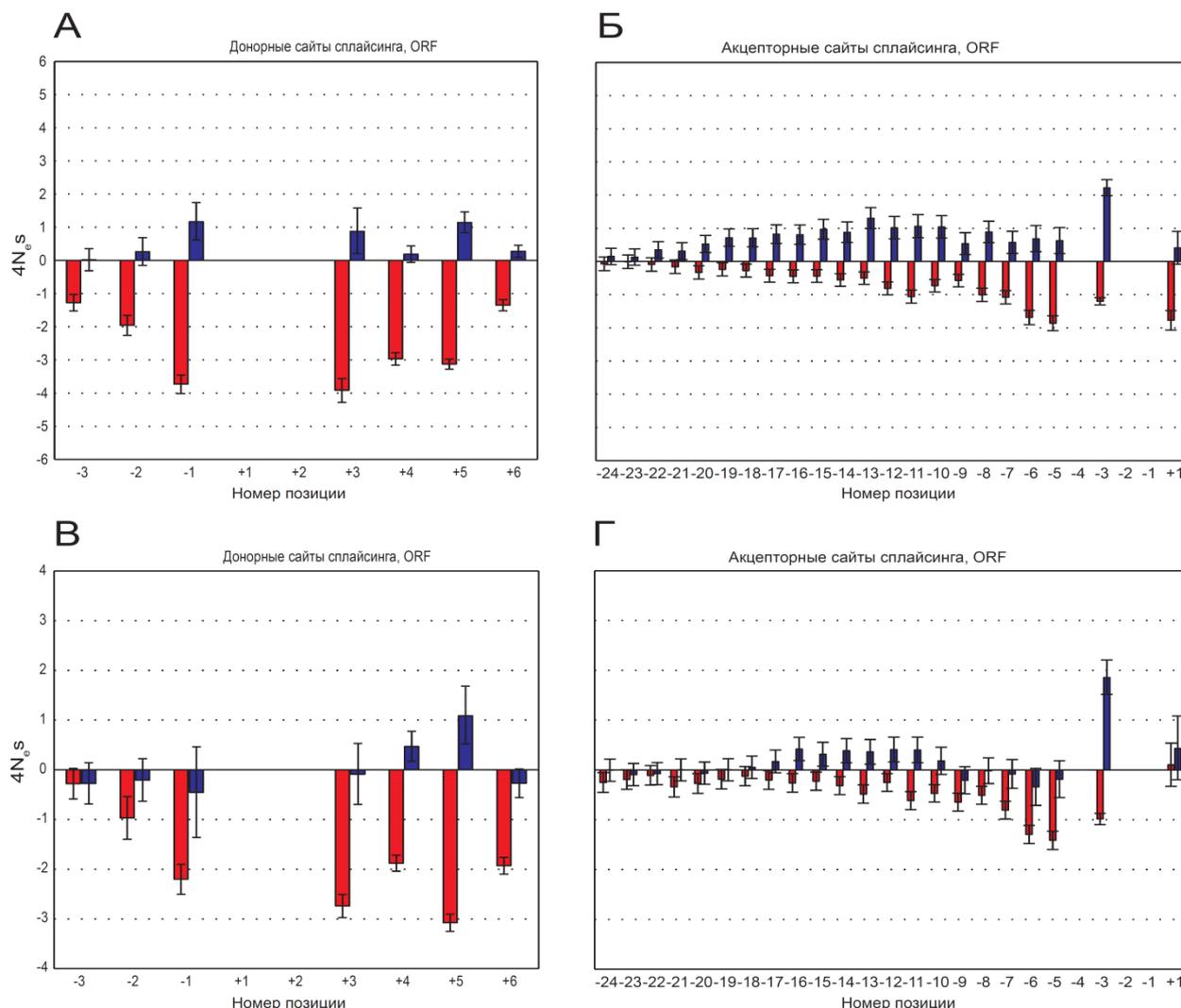


Рисунок 4. Сила отбора.

Сила отбора, измеренная в единицах $4N_e s$ (вертикальная ось), действующего на замены $Nc \rightarrow Cn$ (синие столбики) и на замены $Cn \rightarrow Nc$ (красные столбики). Положительные и отрицательные значения $4N_e s$ соответствуют положительному и отрицательному отбору, соответственно. А, Б – линия *H. sapiens*; В, Г – линия *D. melanogaster*. А, Б – донорные сайты; В, Г – акцепторные сайты (внутри кодирующих областей генов). Усы – 95%-ные доверительные интервалы.

Для неконсенсусных нуклеотидов наблюдается более сложная картина. В позициях сайтов сплайсинга, занятых неконсенсусными нуклеотидами в человеческом геноме, этот нуклеотид в среднем консервативнее нейтральных последовательностей ($c_{nuc} > 0$, рис. 5А, Б). Это указывает на отрицательный отбор, действующий на неконсенсусные нуклеотиды. То же верно для неконсенсусных нуклеотидов в целом ($c_{set} > 0$); таким образом, даже переходы из неконсенсусных нуклеотидов в консенсусные бывают ограничены отбором. Однако отбор предпочитает конкретные неконсенсусные нуклеотиды, присутствующие в каждом конкретном сайте, а не любой неконсенсусный нуклеотид ($c_{nuc} > c_{set}$), что справедливо для всех позиций и

всех типов сайтов сплайсинга в человеческом геноме и геноме *D. melanogaster* (рис. 5А–Г).

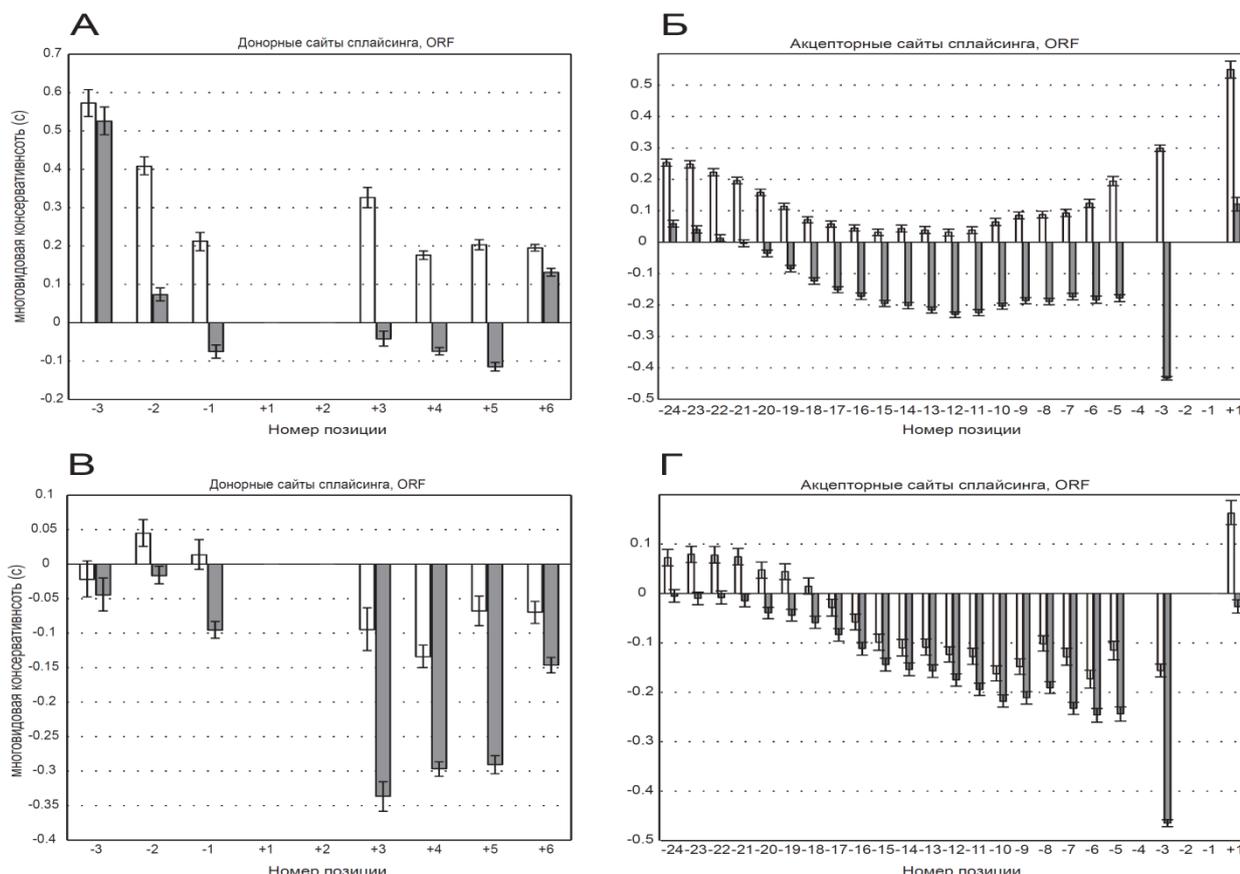


Рисунок 5. Многосвидовая консервативность неконсенсусных нуклеотидов.

Вертикальная ось – величина $c=(obs-exp)/exp$ для N_c нуклеотидов, где obs – наблюдаемая средняя многосвидовая консервативность, а exp – ожидаемая средняя многосвидовая консервативность (нейтральный контроль). Консервативность считалась как для конкретных N_c нуклеотидов, присутствующих в геноме (c_{nuc} , белые столбики), так и для всех нуклеотидов, определенных как N_c в данной позиции (c_{setv} , серые столбики). Верхний ряд (А–Б) – *H. sapiens*; нижний ряд (В–Г) – *D. melanogaster*. Левый столбец (А, В) – донорные сайты; правый столбец (Б, Г) – акцепторные сайты (внутри кодирующих областей генов). Усы – 95%-ные доверительные интервалы.

Следовательно, индивидуальные позиции донорных сайтов сплайсинга находятся под конфликтующими давлениями отбора: в большинстве случаев предпочитается консенсусный нуклеотид, хотя в части сайтов неконсенсусный нуклеотид является предпочтительным. Есть два (взаимно не исключаящие) объяснения того, что на неконсенсусные нуклеотиды может действовать отрицательный отбор. Во-первых, может существовать отбор на то, чтобы сайт был слабым, например, для тонкой регуляции. Во-вторых, присутствующий неконсенсусный нуклеотид может быть предпочтителен из-за других селекционных требований, которые, возможно, не связаны со сплайсингом вовсе. Тот факт, что сила отбора, направленного на

сохранение конкретных неконсенсусных нуклеотидов, выше таковой для неконсенсусных нуклеотидов в целом свидетельствует о том, что нуклеотидная идентичность, а не принадлежность к классу неконсенсусных нуклеотидов в первую очередь определяет консервативность. Таким образом, второе объяснение более вероятно, чем первое.

Было показано, что на наблюдаемые частоты замен нуклеотидов в сайтах сплайсинга и, следовательно, на оценку силы отбора не влияет скорость мутагенеза, отбор в сцепленных локусах (данные в тексте диссертации). Не было найдено также зависимости силы отбора от уровня экспрессии соответствующего гена (данные в тексте диссертации) и от типа сплайсинга, т.е. между сайтами конститутивных и кассетных экзонов (данные в тексте диссертации). Наконец, результаты почти не зависели от того, каким методом, парсимонии или максимального правдоподобия, восстанавливались матрицы нуклеотидных замен (данные в тексте диссертации).

Сила сайта (определяемая как количество консенсусных нуклеотидов в нём) отрицательно коррелирует с силой отрицательного отбора, сохраняющего консенсусные нуклеотиды (это верно для донорных и акцепторных сайтов как из генома *H. sapiens*, так и из генома *D. melanogaster*, данные в тексте диссертации)

Отбор на молодые сайты сплайсинга

Были рассмотрены ситуации появления новых ключевых динуклеотидов на линии человека (*H. sapiens*) с момента отделения её от линии макака (*Macaca mulatta*). Наличие сайтов сплайсинга у человека определялось по аннотации GENCODE. Сайты сплайсинга в общем предке человека и шимпанзе восстанавливались парсимонией. В качестве внешней группы использовалась игрунка обыкновенная (*Callithrix jacchus*). Событием рождения молодого сайта считалось появление соответствующего динуклеотида на линии человека. Так были обнаружены 1698 молодых донорных сайтов сплайсинга и 1470 - акцепторных сайтов.

Сила отбора была оценена отдельно для выборок донорных и акцепторных сайтов сплайсинга, действующего на консенсусные и неконсенсусные нуклеотиды вокруг ключевых динуклеотидов, как описано ранее.

Как видно из Рис. 6, на консенсусные нуклеотиды действует отрицательный отбор, а на неконсенсусные – положительный, так же как и для старых сайтов сплайсинга, т. е. тех, у которых ключевой динуклеотид консервативен в геномах *H. sapiens*, *M. mulatta* и *C. jacchus* (Рис. 4). Сила отрицательного отбора на консенсусные нуклеотиды в молодых сайтах сплайсинга приблизительно совпадает с таковой в старых сайтах. Однако сила положительного отбора против неконсенсусных нуклеотидов у молодых сайтов существенно превышает таковую у старых сайтов сплайсинга, равно как и

абсолютное значение силы отрицательного отбора в соответствующих позициях. Это верно для всех позиций донорных сайтов, для позиций -3 и $+1$ акцепторных сайтов сплайсинга, однако эффект почти отсутствует в полипиримидиновом тракте акцепторных сайтов. Сила положительного отбора между старыми и молодыми сайтами может отличаться в ~ 10 раз (например, в позиции -3 акцепторных и $+3$ донорных сайтов сплайсинга).

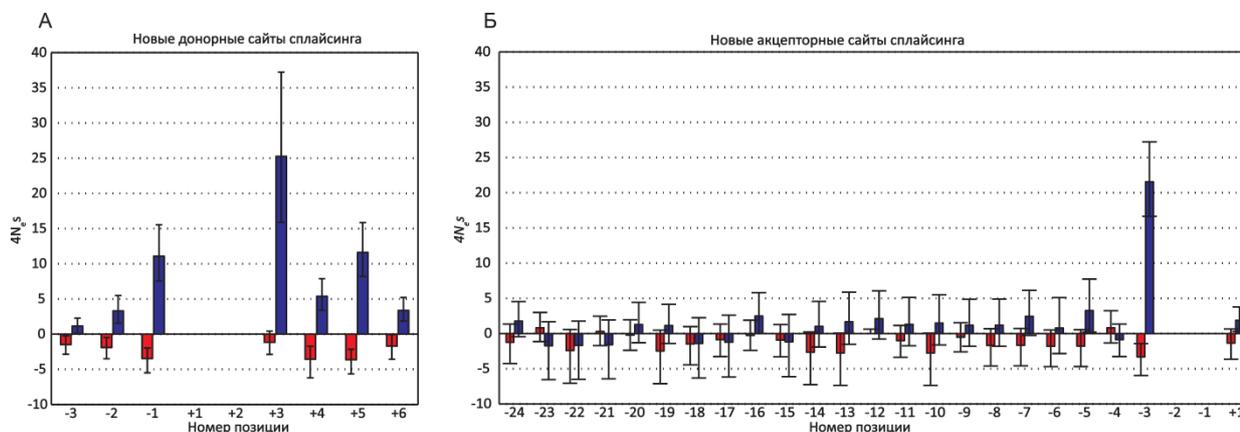


Рисунок 6. Сила отбора в молодых сайтах сплайсинга.

Обозначения те же, что на рис. 4. А – молодые донорные сайты сплайсинга, Б – молодые акцепторные сайты.

Таким образом, событие появления динуклеотида связано с повышенным уровнем положительного отбора на остальные нуклеотиды сайта. Дисбаланс между силой положительного отбора против неконсенсусных нуклеотидов и силой отрицательного отбора, поддерживающего консенсусные нуклеотиды, говорит о том, что молодые сайты еще не достигли мутационно-селекционно-дрейфового равновесия.

Коррелированная эволюция позиций в сайтах сплайсинга млекопитающих

Эволюция сайтов сплайсинга была систематически изучена в линиях млекопитающих, находящихся на средних филогенетических расстояниях друг от друга: человека (*Homo sapiens*), мыши (*Mus musculus*) и собаки (*Canis familiaris*).

Основным объектом изучения была сила отдельных позиций сайтов сплайсинга. Сила позиции сайта сплайсинга определяется следующим образом. Рассмотрим последовательность сайта сплайсинга S . В i -той позиции сайта находится нуклеотид $S(i)$. Сила позиции $S(i)$ определяется как

$$W(S(i)) = \log \frac{p_F(S(i), i)}{p_B(S(i), i)}, \quad (3)$$

где $p_F(\alpha, i)$ – это вероятность нуклеотида α в позиции i в основной модели, посчитанная из частот встречаемости нуклеотидов в выборке сайтов сплайсинга

соответствующего типа; $p_B(\alpha, i)$ – вероятность нуклеотида α в позиции i в фоновой модели.

Нами был разработан метод, посредством которого для каждой заданной тройки ортологичных сайтов сплайсинга (в геномах человека, мыши и собаки) можно оценить (1) вероятность любой заданной последовательности в геноме общего предка, от которой они произошли; (2) оценить среднюю силу сайта сплайсинга в геноме общего предка, предполагая позиционную независимость.

Изменение силы позиций сайтов сплайсинга

Нуклеотидные замены в сайтах сплайсинга могут приводить к изменениям силы соответствующих позиций. Было изучено, как изменяется средняя сила индивидуальных позиций сайтов на разных ветках филогенетического дерева.

Сила многих позиций статистически значимо меняется в одной или нескольких филогенетических линиях. В целом картина изменений веса весьма сложна: положительные и отрицательные значения изменения силы встречаются как в донорных, так и в акцепторных сайтах сплайсинга. Однако в донорных сайтах конститутивных экзонов наблюдается следующая тенденция: сила экзонной части сайта в среднем уменьшается, тогда как у интронной части, напротив, усиливается (данные в тексте диссертации). Возможной причиной описанного ослабления экзонной части и усиления интронной части является миграция сигнала из экзонной части донорного сайта сплайсинга в интронную в течении эволюции. Этот феномен был ранее описан для интронов, появившихся до дивергенции основных эукариотических групп.

В большинстве случаев изменение силы вызывается увеличением или уменьшением частоты консенсусного нуклеотида в данной позиции (например, нуклеотида А в позиции -2 донорного сайта). В полипиримидиновых трактах акцепторных сайтов сплайсинга, однако, происходят замены между двумя самыми частыми нуклеотидами (С→Т).

Сила нуклеотидов в различных позициях сайтов сплайсинга взаимно скоррелирована

Вероятность того, что данный нуклеотид находится в некоторой позиции сайта сплайсинга, может зависеть от того, какие нуклеотиды находятся в других позициях. Мы переформулировали это в терминах силы позиций: зависит ли сила одной позиции от силы другой позиции? Для того чтобы ответить на этот вопрос, были получены (отдельно для выборок донорных и акцепторных сайтов) ковариационные матрицы векторов, соответствующих сайтам сплайсинга, элементами такого вектора являются силы нуклеотидов в позициях сайта.

Ковариационные матрицы векторов сил позиций донорных и акцепторных сайтов сплайсинга конститутивных экзонов из генома мыши показаны на рис. 7 и 8, соответственно (здесь и далее инвариантные динуклеотиды AG в акцепторных сайтах и GT в донорных сайтах не рассматриваются). Аналогичные матрицы для геномов человека и собаки выглядят очень сходным образом (данные в тексте диссертации).

	поз	-3	-2	-1	+3	+4	+5	+6
ковариация	-3	0.29	0.14	0.05	-0.03	-0.05	-0.09	-0.05
	-2	0.14	1.36	0.36	-0.09	-0.30	-0.36	-0.23
	-1	0.05	0.36	1.94	-0.13	-0.34	-0.36	-0.34
	+3	-0.03	-0.09	-0.13	1.06	-0.01	-0.20	-0.11
	+4	-0.05	-0.30	-0.34	-0.01	1.80	0.37	0.02
	+5	-0.09	-0.36	-0.36	-0.20	0.37	1.97	0.21
	+6	-0.05	-0.23	-0.34	-0.11	0.02	0.21	0.73
р-значения	-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	+3	0.00	0.00	0.00	0.00	0.33	0.00	0.00
	+4	0.00	0.00	0.00	0.33	0.00	0.00	0.15
	+5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	+6	0.00	0.00	0.00	0.00	0.15	0.00	0.00

Рисунок 7. Ковариационная матрица векторов сил позиций донорных сайтов сплайсинга конститутивных экзонов (*M. musculus*).

Сверху показана ковариационная матрица. Положительные значения окрашены зелёным цветом, отрицательные – красным. Чем насыщенней окраска, тем больше абсолютное значение ковариации. Снизу указаны р-значения того, данная ковариация значимо отличается от нуля, полученные бутстрепом; $p < 0.05$ окрашены розовым.

Сила каждой позиции в экзонной части донорных сайтов сплайсинга отрицательно коррелирует (коварирует) с силой каждой позиции в интронной части. Это означает, что сайты с сильной экзонной частью имеют тенденцию иметь слабую интронную часть и наоборот. Напротив, позиции внутри экзонной части положительно скоррелированы друг с другом. Позиции интронной части донорных сайтов сплайсинга +4(A), +5(G) и +6(T) также положительно скоррелированы, однако между позицией +3 и другими позициями интронной части наблюдается отрицательная корреляция. Описанные наблюдения сохраняются и для сайтов кассетных экзонов, причем отрицательные корреляции между экзонной и интронной частями даже более выражены, чем на выборке донорных сайтов конститутивных экзонов.

Теоретически, наблюдаемые ковариации могут вызываться несколькими причинами: (1) мутационные эффекты (мутации происходят в определенном контексте), (2) отборные эффекты, а именно эпистатический отбор (вклад в приспособленность нуклеотида, находящегося в одной позиции зависит от нуклеотида в другой позиции), (3) исторические причины (статистический эффект смешивания в выборке сайтов разных возрастов). Мы рассмотрели каждую из этих возможностей и выяснили, какая из них лучше объясняет сделанные наблюдения.

Можно предположить, что отрицательные ковариации между экзонной и интронной частями и положительные ковариации внутри экзонной и интронной частей донорных сайтов сплайсинга могут быть результатом миграции сигнала из экзонной части в интронную (исторические причины). Изучаемая выборка может состоять из донорных сайтов различных возрастов. Тогда низкоконсервативные сайты должны иметь сильную экзонную и слабую интронную части, а у высококонсервативных сайтов, напротив, должна быть слабая экзонная и сильная интронная части. Смешивание в одной выборки сайтов разных возрастов может привести к наблюдаемым ковариациям. Ковариационные матрицы низко- и высококонсервативных донорных сайтов оказались очень похожи между собой и на матрицу, построенную на объединённой выборке (данные в тексте диссертации). Это наблюдение противоречит выдвинутому предположению о том, что наблюдаемый паттерн ковариаций объясняется миграцией сигнала. Однако, миграция сигнала происходит, что проявляется в ослаблении экзонной и усилении интронной частей донорных сайтов сплайсинга (см. выше). Таким образом миграция сигнала не оказывает существенного влияния на наблюдаемый эффект.

Другая возможная причина – мутационные эффекты. Однако, так как ковариации между позициями на противоположных концах донорного сайта сплайсинга столь же сильны (по модулю), что и между соседними, то, что наблюдаемый паттерн объясняется мутационными эффектами, маловероятно, поскольку известные мутационные контексты включают только соседние нуклеотиды.

Таким образом, единственная оставшаяся альтернатива такова: ковариации между силами позиций вызваны естественным отбором, причем не позиционно-независимым, а эпистатическим. Положительные ковариации соответствуют ситуации, когда две мутации, уменьшающие вес сайта, являются менее вредными, чем, если бы они действовали по отдельности, т.е. положительному эпистазу. Отрицательные ковариации соответствуют отрицательному эпистазу, т.е. когда две мутации вместе являются более вредными, чем по отдельности.

В целом, функциональная природа эпистаза в донорных сайтах сплайсинга остаётся неясной. Однако сделать некоторые предположения. Возможной функциональной причиной наличия положительного эпистаза в донорных сайтах сплайсинга является кооперативное связывание с U1 за счёт стэкинг-взаимодействий. Природа отрицательных ковариаций между экзонной и интронной частями менее понятна. Возможно, для некоторой части сайтов сплайсинга максимальный вес не означает максимального вклада в приспособленность. В таких сайтах сильная (слабая) экзонная часть компенсирует слабую (соотв., сильную) интронную часть. О наличии таких сайтов сплайсинга также косвенно свидетельствует тот факт, что сила отбора,

действующего на консенсусные нуклеотиды, слабее для сильных сайтов, чем для слабых.

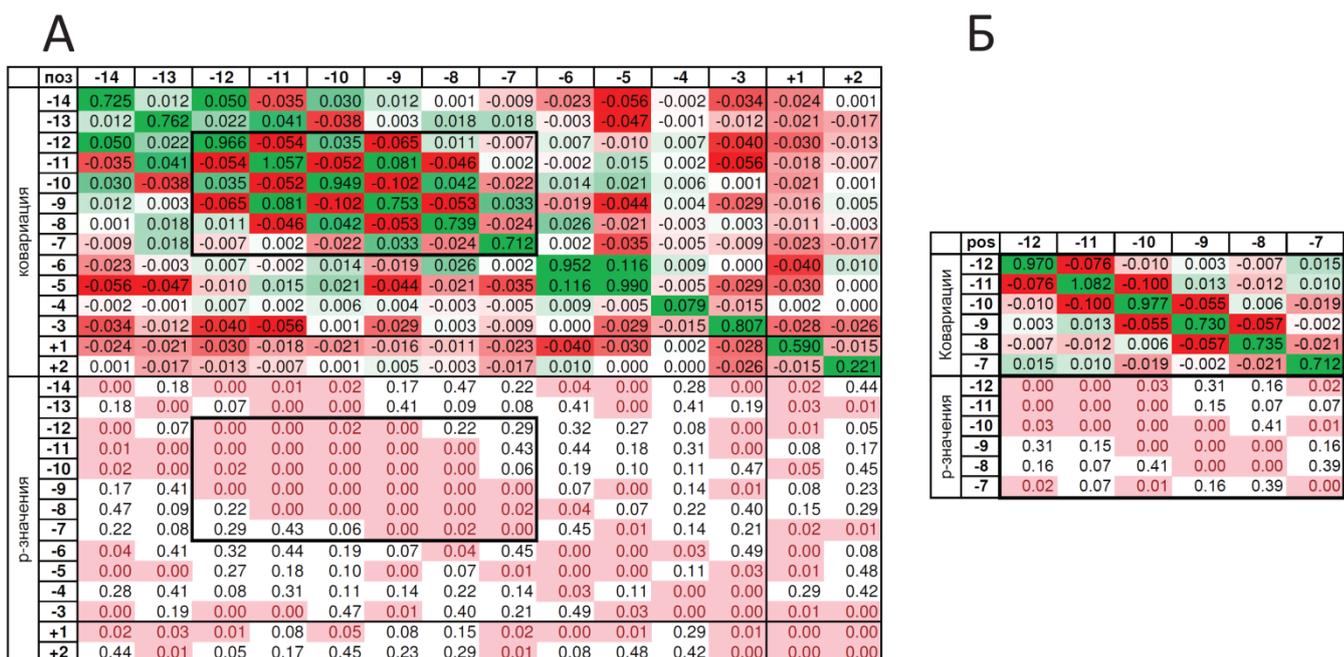


Рисунок 8. Ковариационные матрицы векторов сил позиций акцепторных сайтов сплайсинга конститутивных экзонов (*M. musculus*).

Обозначения те же, что на рис. 7. Позиции, для которых наблюдается характерный периодический паттерн ковариаций (−12 ... −7) обведены черной рамкой. А — реальные последовательности акцепторных сайтов сплайсинга. Б — искусственные последовательности акцепторных сайтов с тем же динуклеотидным составом, что и в реальных сайтах на участке (−12 ... −7).

В акцепторных сайтах сплайсинга ковариационная матрица устроена более сложным образом (рис. 8). В ней встречаются как положительные, так и отрицательные ковариации. Некоторые позиции (−3, +1 и +2) отрицательно скоррелированы со всеми остальными позициями. В полипиримидиновом тракте (позиции −12 ... −7), наблюдается характерный периодический характер ковариаций: силы соседних позиций (на расстоянии в 1 нт) скоррелированы отрицательно, тогда как позиции на расстоянии 2 нт скоррелированы положительно, на расстоянии 3 нт — вновь отрицательно и т.д. (рис. 8А). Очень похожая картина наблюдается на выборках акцепторных сайтов из геномов человека и собаки, а также на выборке сайтов кассетных экзонов.

Мы предположили, что наблюдаемая периодичность ковариации вызвана смещенным динуклеотидным составом в полипиримидиновом тракте. Для проверки этого предположения были сгенерированы искусственные последовательности с тем же динуклеотидным составом, что и в оригинальных последовательностях полипиримидиновых трактов (см. текст диссертации) и построили матрицы ковариаций на этих искусственных последовательностях, затем сравнили их с

матрицами ковариаций, построенных на реальных акцепторных сайтах (рис. 8Б). Периодический паттерн ковариаций наблюдается и на искусственных последовательностях, хотя в целом ковариации выражены слабее.

Каковы эволюционные причины появления периодического паттерна ковариаций? Теоретически, он может возникать (1) из-за особенностей мутагенеза или (2) благодаря действию естественного отбора.

Для того, чтобы это понять, динуклеотиды были упорядочены по степени их недо- или перепредставленности по сравнению с тем, что ожидается исходя из частот встречаемости (моно)нуклеотидов. Аналогичные частоты были также подсчитаны для нейтрально эволюционирующих последовательностей, взятых из соседнего интрона (нейтральный контроль).

Наиболее недопредставленным динуклеотидом оказался AG, который встречается в ~30 раз реже, чем ожидается на основе частот мононуклеотидов, при этом такой недопредставленности не наблюдается в нейтральном контроле (рис. 9). Недопредставленность AG – основная причина периодического паттерна ковариаций. Динуклеотид AG недопредставлен в полипиримидиновом тракте, чтобы избежать сплайсинга по этому нуклеотиду (он может служить альтернативой существующему ключевому динуклеотиду и приводить к смещению сайта сплайсинга).

Прочие особенности динуклеотидного состава полипиримидинового тракта могут быть обусловлены широким набором факторов, таких как особенности мутагенеза (недопредставленность CG в геномах млекопитающих из-за высокой вероятности дезаминирования метилированного цитозина в контексте CpG) оптимальное связывание U2AF65 и регуляторных транс-факторов сплайсинга, так и быть побочным следствием недопредставленности AG (например, перепредставленность AA по сравнению с нейтральным контролем).

Итак, периодический характер ковариаций в полипиримидиновом тракте определяется в основном динуклеотидным составом, причем недопредставленность AG играет здесь ключевую роль. Формально говоря, отбор против AG является вырожденным случаем эпистатического отбора. Появление нуклеотида A (равно как и G) в результате мутации почти не уменьшает приспособленности, в то время как их совместное появление в виде нуклеотида AG существенно её снижает из-за нарушения сплайсинга.

Корреляции позиций усиливаются в течение эволюции

Наблюдаемые ковариации между силами позиций в сайтах сплайсинга – результат длительной эволюции, из которой мы наблюдаем финальную стадию: от общего предка человека мыши и собаки до каждого из этих видов. Естественным является

следующий вопрос: влияют ли зависимости между позициями на эволюцию сайтов сплайсинга на рассматриваемом финальном этапе? Более конкретно, эпистатический отбор должен приводит к усилению (по модулю) соответствующих ковариаций. Верно ли это для сайтов сплайсинга?

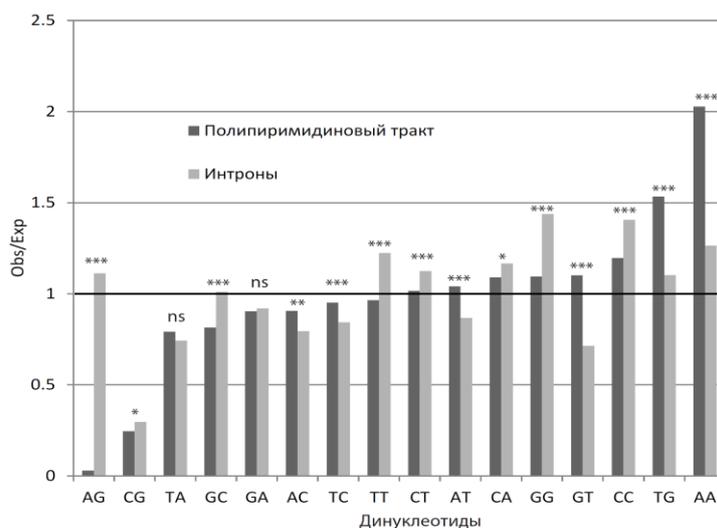


Рисунок 9. Динуклеотидный состав участка полипиримидинового тракта, где наблюдается периодический паттерн ковариаций (позиции – 12 ... -7). По вертикальной оси отложено отношение наблюдаемой частоты динуклеотида к ожидаемой (*Obs/Exp*) на участке (-12 ... -7) полипиримидинового тракта (черные столбики), а также в близлежащих интронах (серые столбики). Динуклеотиды упорядочены по возрастанию величины *Obs/Exp* на участке периодических ковариаций. Ожидаемая частота динуклеотидов (*Exp*) вычислялась исходя из мононуклеотидного состава. Показана статистическая значимость различий между величинами *Obs/Exp* на участке (-12 ... -7) полипиримидинового тракта и близлежащих интронов: * $0.01 < p \leq 0.05$; ** $0.00 < p \leq 0.01$; *** $p \leq 0.001$; ns $p > 0.05$.

Мы смоделировали процесс эволюции сайтов сплайсинга в предположении независимости позиций, получив набор искусственных сайтов. Чтобы понять, как влияет зависимость между позициями на ковариации в ныне существующих сайтах, мы сравнили ковариационные матрицы, построенные на выборках реальных и искусственных сайтов сплайсинга

Как в донорных сайтах, так и в акцепторных сплайсинга было найдено несколько статистически значимых различий между ковариационными матрицами (см. текст диссертации). Общая тенденция состоит в том, что отрицательные ковариации усиливаются по модулю. Положительные ковариации также усиливаются, однако статистически значимо это происходит только на выборке акцепторных сайтов. Сайты кассетных экзонов проявляют в целом те же свойства, что и сайты конститутивных экзонов, однако из-за меньшего объема выборки статистическая достоверность результатов ниже. В линии мыши все тенденции проявляются наиболее контрастно, видимо, в силу большей длины соответствующей ветви дерева.

Таким образом, как в донорных, так и в акцепторных сайтах сплайсинга наблюдается увеличение по модулю как отрицательных, так и положительных ковариаций, что является следствием действия эпистатического отбора. Отбор против динуклеотида AG в акцепторных сайтах сплайсинга является частным (вырожденным) случаем эпистатического отбора.

Отбор в окрестности сайтов сплайсинга

Консервативность цис-регулятора сплайсинга UGCAUG в геномах человека и мыши

Brudno с соавт. (Brudno et al., 2001) проанализировали выборку кассетных экзонов, специфично экспрессирующихся в мозгу, с интронными последовательностями вокруг них. Было показано, что в интронах, следующих за кассетными экзонами, мотив UGCAUG представлен в избытке по сравнению с контрольной выборкой, содержащей конститутивные экзоны с последующими интронами. Часть из найденных Brudno с соавт. гексануклеотидов UGCAUG могли встретиться случайно. Наша задача состояла в том, чтобы путем сравнительного анализа геномов человека и мыши отделить случайные гексануклеотиды от функциональных, т.е. принимающих участие в регуляции альтернативного сплайсинга.

Мы проанализировали все находки гексануклеотида UGCAUG в пределах 1000 нт после кассетных экзонов из рассматриваемой выборки как в геноме человека, так и в геноме мыши. В 24 парах ортологичных генов, 28 мотивов UGCAUG найдены в интронах человека и 20 – в интронах мыши (подробнее – см. раздел 4.3.1). Четырнадцать мотивов консервативны (не содержат ни одной замены) в геномах человека и мыши. То есть уровень консервативности составляет $14/28 = 50\%$ для мотивов человека и $14/20 = 70\%$ для мотивов мыши.

Частота замен в нейтральных регионах между геномами человека и мыши оценивается как 0.51 замены на нуклеотид. Таким образом, если $q = 0.51$, то вероятность консервативности гексануклеотида равна $(1 - q)^6 = 1.8\%$, т.е. подавляющее большинство гексануклеотидов должны содержать замены. Как сказано выше в реальности уровень консервативности найденных UGCAUG существенно выше: 50% для мотивов человека и 70% для мотивов мыши. Поэтому консервативность гексануклеотида означает действие отрицательного отбора, т.е. его функциональную значимость. С другой стороны, нельзя полностью исключить возможности того, что часть неконсервативных мотивов функциональны. Регуляция альтернативного сплайсинга меняется в течение эволюции, и часть гексануклеотидов могли перестать выполнять прежнюю функцию в одной из линий. Мы не утверждаем,

что нашли все функциональные UGCAUG, однако считаем, что консервативные гексануклеотиды скорее всего функциональны.

Все найденные консервативные гексануклеотиды UGCAUG окружены длинными консервативными участками с обеих сторон, причем часто консервативные участки примыкают к экзону. Это указывает на то, что регуляция альтернативного сплайсинга – довольно сложный процесс с участием многих белков, а исследуемый мотив – лишь часть системы цис-регуляторных элементов. Консервативность гексануклеотида UGCAUG в геномах позвоночных, а также его склонность находиться в консервативном окружении была позже подтверждена в более позднем исследовании (Minovitsky et al., 2005).

В последующих работах было показано, что с гексануклеотидом UGCAUG специфически связываются транс-факторы сплайсинга Fox1 и Fox2 (Jin, 2003; Ponthier et al., 2006). Гены этих факторов присутствуют во многих эукариотических геномах от *C. elegans* до *H. sapiens* (Zhang et al., 2008). Было показано, что система UGCAUG-Fox1/2 участвует в тканеспецифичной регуляции альтернативных экзонов нескольких генов человека: EPB41 (Ponthier et al., 2006; Huang et al., 2012), CGRP (Zhou et al., 2007), ATP5C1 (Fukumura et al., 2007), FGFR2 (Baraniak et al., 2006). Кроме того, был произведён глобальный анализ сплайсинга, регулируемой системой UGCAUG-Fox1/2 (Zhang et al., 2008). Авторами исследования было показано, что гиперэкспрессия и нокадаун Fox1/2 приводит к изменению сплайсинга тех экзонов, в которых находятся UGCAUG-элементы с предсказанной функциональностью. Кроме того на нескольких примерах был показан механизм как именно связывание Fox1/2 с UGCAUG влияет на сплайсинг. На примере экзона 16 гена EPB41 было показано, что Fox2 связывается с энхансером находящимся в интроне после экзона, взаимодействует с U1C, что приводит к стабилизации комплекса U1 с мРНК и активации донорного сайта сплайсинга (Huang et al., 2012). Белки Fox-1 и Fox-2 связываются с UGCAUG, находящимися в интроне до экзона 4 гена кальцитонина, и регулируют сплайсинг путём блокирования связывания U2AF65 с акцепторным сайтом экзона 4 CGRP (Zhou et al., 2007). Эксперименты *in vitro* показали, что Fox1 препятствует формированию раннего комплекса (E-комплекса) на интроне 9 гена ATP5C1 (Fukumura et al., 2007).

Ранее считалось, что все экзоны из нашей выборки являются мозгоспецифичными. Мы произвели анализ распределения изоформ с кассетным экзоном по тканям и органам и выяснили, что строгой мозгоспецифичности не наблюдается. По меньшей мере 10 экзонов человека и 4 экзона мыши не являются мозгоспецифичными. Таким образом, возникло предположение, что UGCAUG не является строго мозгоспецифичным энхансером альтернативного сплайсинга. Это было подтверждено в последующих экспериментальных работах.

Выводы

Отбор в сайтах сплайсинга

1. На неконсенсусные нуклеотиды действует слабый положительный, а на консенсусные, соответственно, слабый отрицательный отбор ($1 < |4N_e s| < 4$). У части сайтов сплайсинга, однако, существует отбор на неконсенсусные нуклеотиды. Сила сайта сплайсинга влияет на силу отбора: на консенсусные нуклеотиды слабых сайтов сплайсинга действует более сильный отрицательный отбор.
2. На молодые сайты сплайсинга, появившиеся на линии человека после расхождения с макаком, действует отбор в 2-10 раз более сильный отбор, чем на старые сайты сплайсинга (т.е. общие для человека, макака и игрунки).

Коррелированная эволюция позиций в сайтах сплайсинга млекопитающих

3. В донорных сайтах сплайсинга конститутивных экзонов наблюдается ослабление экзонной части и слабое увеличение силы интронной части сайтов, что согласуется с гипотезой о миграции сигнала из экзонной части в интронную.
4. Силы нуклеотидов в различных позициях сайтов сплайсинга часто взаимно скоррелированы. В донорных сайтах сплайсинга наблюдаются положительные ковариации между позициями внутри экзонной и внутри интронной частей и отрицательные – между экзонными и интронными частями. В полипиримидиновом тракте акцепторных сайтов сплайсинга наблюдается характерный чередующийся характер ковариаций: соседние позиции коварируют отрицательно, через позицию – положительно, через две – отрицательно и т.д.
5. Отбор против динуклеотида AG есть основная причина ковариаций, наблюдаемых в полипиримидиновом тракте акцепторных сайтов сплайсинга. Эпистаз – наиболее вероятная причина ковариаций в донорном сайте сплайсинга.
6. Как в донорных, так и в акцепторных сайтах сплайсинга действует эпистатический отбор, который усиливает существующие ковариации между нуклеотидами.

Консервативность цис-регулятора сплайсинга UGCAUG

7. Консервативность гексануклеотидов UGCAUG, встречающиеся в интронах после кассетных экзонов в геномах человека и мыши существенно выше средней консервативности интронов, что говорит об их вероятной функциональности.

Список публикаций по теме диссертации

Статьи в рецензируемых научных журналах

1. S. Denisov, G. Bazykin, A. Favorov, A. Mironov, M. Gelfand (2015) Correlated evolution of nucleotide positions within splice sites in mammals. PLOS ONE 10(12): e0144388.
2. S. Denisov, G. Bazykin, R. Sutormin, A. Favorov, A. Mironov, M. Gelfand, A. Kondrashov (2014) Weak negative and positive selection and the drift load at splice sites. Genome Biology and Evolution 6(6): 1437-1447.
3. S. Denisov, M.S. Gelfand (2003) Conservedness of the alternative splicing signal UGCAUG in the human and mouse genomes. Biophysics (Moscow) 48(suppl. 1): 30-35

Доклады на конференциях

1. С.В. Денисов, Г.А. Базыкин, Влияние корреляций между позициями на эволюцию сайтов сплайсинга в геномах млекопитающих. Информационные технологии и системы (ИТиС'15), 7 - 11 сентября, 2015, Олимпийская деревня, Сочи, Россия
2. С.В. Денисов, Слабый отбор и дрейфовый груз в сайтах сплайсинга. Информационные технологии и системы (ИТиС'14), 1 - 5 сентября, 2014, Нижний Новгород, Россия
3. С.В. Денисов, Новорожденные сайты сплайсинга находятся под положительным отбором. Информационные технологии и системы (ИТиС'13), 1 - 6 сентября, 2013, Калининград, Россия
4. Stepan V. Denisov, Georgii A. Bazykin, Mikhail S. Gelfand, Alexey S. Kondrashov. Turnover of Slightly Deleterious and Slightly Advantageous Alleles in Splice Sites of Humans and Drosophila melanogaster. Society for Molecular Biology & Evolution Conference 2013 (SMBE 2013), July 7 – 11, 2013, Chicago, USA.
5. С.В. Денисов, Г.А. Базыкин. Роль неконсенсусных нуклеотидов в эволюции сайтов сплайсинга: данные по однонуклеотидным полиморфизмам (SNP) и дивергенции. Информационные технологии и системы (ИТиС'12), 19 - 25 августа, 2012, Петрозаводск, Россия
6. Stepan V. Denisov, Georgii A. Bazykin, Alexander V. Favorov, Andrey A. Mironov, Mikhail S. Gelfand. Positive and negative selection in splice sites. Society for Molecular Biology & Evolution Conference 2012 (SMBE 2012), June 23-26, 2012, Dublin, Ireland