



УДК 577.21

# ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ ВСТАВОК ЭНДОГЕННЫХ РЕТРОВИРУСОВ В КОНТЕКСТЕ ЭВОЛЮЦИИ ГЕНОМА ЧЕЛОВЕКА

© 2010 г. А. А. Буздин\*

Учреждение Российской академии наук

Институт биоорганической химии им. акад. М. М. Шемякина и Ю. А. Овчинникова РАН,

117997 ГСП, Москва, В-437, ул. Миклухо-Макляя, 16/10

Поступила в редакцию 20.05.2009 г. Принята к печати 22.06.2009 г.

Ретроэлементы – мобильные элементы, распространяющиеся в ДНК с помощью процесса обратной транскрипции – составляют около 40% генома человека. Небольшая их часть появилась в геноме сравнительно недавно, уже после расхождения предковых линий человека и шимпанзе. “Эволюционно молодые” ретроэлементы представлены членами четырех групп: SVA, Alu, L1 и эндогенных ретровирусов HERV-K (HML-2). Эти ретроэлементы могли сыграть функциональную роль в ходе молекулярной эволюции ДНК человека. В работах нашей группы мы детально исследовали вклад человекспецифичных эндогенных ретровирусов (hsERV) в изменение структуры и регуляции генома человека. Выявлено, что hsERV, представленные 134 копиями, занимают около 330 т.п.о. ДНК человека. Они добавили в геномную последовательность копии 50 функциональных ретровирусных генов, а также 134 потенциальных промотора и энхансера, 50% из которых – в областях, соседствующих с известными генами, и 22% – в интронах генов. Как минимум 67 таких элементов работают как человекспецифичные промоторы *in vivo*. hsERV регулируют активность известных белоккодирующих генов посредством интерференции РНК и как энхансеры, а также предоставляют новые сигналы полиаденилирования для мРНК.

*Ключевые слова:* молекулярная эволюция, видообразование, геном человека, ретроэлементы, мобильные элементы ДНК, регуляция экспрессии генов.

## ВВЕДЕНИЕ

Одной из наиболее интересных задач современной биологии является поиск генетических различий между человеком и остальными приматами, а также функциональная характеристика этих различий. Наилучшим объектом сравнения являются шимпанзе – ближайшие “родственники” человека. ДНК человека и шимпанзе чрезвычайно близки по первичной структуре. Средняя дивергенция между ними составляет около 1.24% [1], и всего 0.5% в белоккодирующих областях [2]. Предковые линии человека и шимпанзе разделились в ходе эволюции сравнительно недавно, около 6 млн лет назад [3]. До сих пор остается тайной, какие именно генетические особенности ответственны за уникальный фенотип человека. В то же время найдены миллионы различий между геномами человека и других высших приматов. Эти различия относятся к четырем основным типам:

- 1) различия в хромосомной организации, включая протяженные делеции, инверсии, дубликации и транслокации;
- 2) различия в количестве геномных копий, локализации и функциональном статусе предсуществующих последовательностей, общих для человека и других приматов;
- 3) различия в белоккодирующих областях;
- 4) специфичные вставки ретроэлементов (RE).

RE занимают около 42% генома человека и являются “эгоистичными” последовательностями ДНК, способными увеличивать число своих копий в геноме при помощи промежуточных РНК-посредников и механизма обратной транскрипции (синтеза комплементарной ДНК на матрице РНК при помощи фермента обратной транскриптазы). Образующиеся ДНК-копии затем интегрируют в геном [4]. Среди всех типов мобильных элементов, только RE размножаются в геномах млекопитающих [5]. Четыре семейства RE (L1, Alu, SVA и HERV-K (HML-2)) были транспозиционно активны после разделения предковых линий человека и шимпанзе, что оставило след в виде сравнительно небольшой фракции человекспецифичных RE (тысячи копий [6], в сравнении с общим числом RE у человека более 3 млн [7]).

Сокращения: ПЦР – полимеразная цепная реакция; ОТ-ПЦР – ПЦР с применением обратной транскрипции; ERV – эндогенный ретровирус; RE – ретроэлемент; hs – специфичный для ДНК человека; LTR – длинные концевые повторы ретровирусов.

\* Автор для связи (тел.: (495) 330-65-74; тел./факс: (495) 330-65-38; эл. почта: anton@humgen.iobc.ras.ru).

В 1975 г. Кинг и Вильсон выдвинули постулат о том, что фенотипические различия между человеком и шимпанзе могли быть вызваны сравнительно небольшим числом изменений в регуляторных последовательностях важных генов [8], что со временем стало преобладающей гипотезой среди эволюционистов.

Геномы человека и шимпанзе, каждый размером около  $3 \times 10^9$  п.о., идентичны друг другу на ~98.8%. Успехи в секвенировании геномов позволяют сравнить ДНК человека и шимпанзе напрямую [9]. Простое биоинформатическое сравнение может, например, зафиксировать изменения в ДНК, отразившиеся на аминокислотной последовательности продуктов известных генов [10]. Функциональное тестирование может предоставить исследователям более богатую информацию, но проведено оно лишь для небольшого количества дифференциальных геномных локусов. На настоящий момент найдено множество специфичных замен, делеций, инсерций, дупликаций, экспрессирующихся псевдогенов, транскрибируемых РНК неизвестной функции, вставок RE, хотя интегральная база данных по чело- векспецифичной ДНК пока так и не составлена.

*Цитогенетические различия.* Основные различия между кариотипом человека и шимпанзе были выявлены при помощи метода флуоресцентной гибридизации *in situ*. Прежде всего, это слияние у человека (хромосома 2) двух предковых хромосом, соответствующих хромосомам 12 и 13 шимпанзе [11]. Кроме того, зафиксированы множественные перичентрические инверсии хромосом 1 и 18 человека [12], протяженные перестройки гетерохроматиновых участков [13], межхромосомные обмены ДНК на периферии альфа-сателлитной ДНК центромер [14], отсутствие у человека АТ-богатого сателлита в субтеломерных участках [15], наконец, специфические перестройки и амплификация генных семейств в нерекомбинирующей части Y-хромосомы [16]. Вместе — инсерции и делеции, создали по крайней мере 150 млн п.о., различающих человека и шимпанзе [17].

*Появление или инактивация функциональных генов.* Лишь небольшое число функциональных генов различает геномы человека и шимпанзе. Во-первых, это функциональная делеция экзона в белоккодирующей области гена *СМР* гидроксилазы сиаловой кислоты. Данная мутация, вызванная чело- векспецифичной вставкой повтора *Alu* в экзон *СМР*, разрушила нормальную открытую рамку считывания для этого фермента и привела к исчезновению у человека на поверхности клеточных мембран *N*-гликолилнейраминавой кислоты (*Neu5Gc*) [18]. Вместо *Neu5Gc*, на поверхности мембран у человека находится ее предшественник, *N*-ацетилнейраминавая кислота. Это биохимическое различие между человеком и шимпанзе, возможно, каким-то образом влияет на межклеточные взаимодействия и

эмбриональное развитие, в частности, формирование мозга.

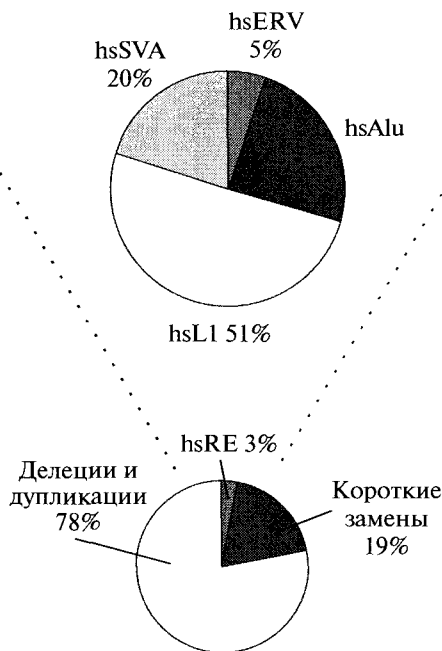
Некоторые другие предковые гены, в основном гены обонятельных рецепторов [19], также были потеряны или инактивированы мутациями у человека. С другой стороны, транскрибируемая человекспецифичная последовательность *clorf37-dup*, кодирующая короткий трансмембранный белок неизвестной функции, селективно экспрессируется в некоторых тканях человека, в том числе и в мозге [20]. Кроме того, в ходе сравнительно недавних событий эволюции генома человека были приобретены или утрачены многие гены, вовлеченные в иммунный ответ (лейкоцитарные рецепторы и антигены) [21].

*Дупликации генов.* Дупликации могут влиять на клеточную физиологию, предоставляя дополнительные копии генов, что меняет количественный контроль их экспрессии. Например, гаплоидный набор ДНК человека содержит 7–11 копий гена обонятельного рецептора *OR-A*, тогда как в геноме шимпанзе присутствует всего одна копия. Другой пример — ген фактора роста кератиноцитов *KGF*, 8 копий которого картированы в ДНК человека и только 5 — у шимпанзе [22]. Новые копии дуплицированных генов могут транскрибироваться с различными тканевыми специфичностями, в зависимости от их нового геномного окружения [23].

*Видоспецифичные нуклеотидные замены.* На настоящий момент картированы миллионы человекспецифичных однунуклеотидных замен, коротких делеций, дупликаций или участков микросателлитных амплификаций. Многие из них располагаются в регуляторных участках генов, некоторые — в белоккодирующих последовательностях. Например, у шимпанзе ген рецептора дофамина *D4* имеет 12-нуклеотидную делецию в белоккодирующей области по сравнению с полноразмерным геном человека [24]. Впрочем, биологическое значение почти всех этих многочисленных замен, составляющих вместе около 36 млн п.о., не исследовано и ожидает анализа и систематизации.

*Различия в экспрессии генов.* Поиск дифференциально транскрибируемых последовательностей мог бы быть лучшим решением для непосредственно обнаружения генов, ответственных за видообразование человека. Например, Надеждину и соавт. [25] удалось обнаружить дифференциальную транскрипцию гена переносчика гормонов щитовидной железы транскретина в мозгу человека и шимпанзе. Однако же, это и другие подобные исследования проводились на очень ограниченной выборке тканей шимпанзе (которые являются весьма дефицитными среди исследователей). Это сильно снижает ценность результатов, поскольку получаемые таким образом межвидовые различия, как правило, оказываются меньше по амплитуде, чем внутривидовые различия среди людей [26].

*Ретроэлементы.* После расхождения предковых линий человека и шимпанзе, четыре группы ретро-



**Рис. 1.** Статистические данные по последовательностям ДНК, специфичным для генома человека. Эндogenous ретровирусы занимают около 5% всей ДНК, занятой человекеспецифичными мобильными элементами, которые, в свою очередь, образуют не более 3% от всей человекеспецифичной ДНК. “Короткие замены” — обозначение непротяженных (до 10 нуклеотидов) замен в геномной ДНК.

элементов оставались транскрипционно активны и продолжали размножаться в ДНК. Недавно проведенный полногеномный биоинформатический скрининг [6] показал, что ДНК человека содержит около 7800 копий ретроэлементов, отличающих ее от шимпанзе. В реальности это число, по-видимому, выше, поскольку ни для шимпанзе, ни для человека не установлена полностью первичная структура центромерных и перителомерных участков, содержащих особенно много RE.

Первая группа RE, представленная около 1200 человекеспецифичными последовательностями, это автономные ретротранспозоны L1 [27]. Полноразмерный L1 имеет длину около 6 т.п.о. и кодирует две открытые рамки считывания (для обратной транскриптазы/интегразы и РНК связывающего белка). Большинство вставок L1 укорочено с 5'-конца, что является следствием abortивной обратной транскрипции.

Вторая и третья группы RE, ретропозоны Alu (элементы длиной около 300 п.о.) и SVA (каждый приблизительно по 1.5 т.п.о.) — это неавтономные RE, привлекающие “чужую” обратную транскриптазу, кодируемую L1, для собственного размножения в геноме [28]. Сами они не кодируют никаких белоккодирующих генов и считаются паразитами ретропозиционного аппарата L1. Эти две группы

представлены в ДНК человека, соответственно, ~5500 и ~860 человекспецифичными копиями.

Наконец, последняя группа — эндогенные ретровирусы семейства HERV-K (HML-2), которые кодируют три классических ретровирусных гена и один дополнительный регуляторный ген, это самые сложно устроенные RE человека. Вместе взятые, человекспецифичные RE составляют около 6.4 млн п.о. (рис. 1), что в 6 раз меньше, чем общая протяженность hs-однонуклеотидных замен, и в 23 раза меньше — чем суммарная длина hs делеций и дубликаций. Впрочем, такая, со структурной точки зрения, скромная доля hsRE компенсируется их активной ролью в функциональном видоизменении геномов. Например, RE — это горячие точки рекомбинации (в частности, эволюционно недавние рекомбинации между hsAlu привели к делеции в общей сложности 400 т.п.о. ДНК человека [29]). Также RE способны изменять активность предсуществующих генов путем предоставления новых промоторов, энхансеров, сигналов полиаденилирования и дополнительных сайтов сплайсинга [5, 30]. Полиморфные вставки повторов L1 и Alu в интронах снижают транскрипцию соответствующих аллельных вариантов генов [31], что, возможно, свидетельствует о вовлеченности RE в эпигенетическую регуляцию экспрессии генов. Следует отметить, что около трети всех hsRE были картированы либо в составе, либо в непосредственном окружении известных генов [6].

Поэтому вставки RE могут быть одной из причин, на молекулярном уровне обусловивших фенотипические различия между человеком и шимпанзе. Эти различия могли быть связаны не с приобретением новых или с потерей старых генов, а с видоспецифичным изменением регуляции предковых генов, присутствующих в геномах всех потомков.

Работа нашей группы была сфокусирована на детальном полногеномном исследовании человекспецифичных эндогенных ретровирусов. hsERV, пожалуй, являются наиболее интересными кандидатами на роль участников в видообразовании, поскольку они обладают функциональными регуляторными элементами, такими, как энхансеры, промоторы, сайты сплайсинга, сигналы полиаденилирования, а также несут четыре гена, продукты которых вовлечены, помимо жизненного цикла ретровирусов, в иммунную супрессию, защиту от вирусных инфекций и прогрессию рака. Мы разработали детальную базу данных hsERV, включающую информацию об их геномной локализации, расстоянии от известных генов, кодируемых белках, их функциональном статусе, таком, как транскрипция или метилирование в различных тканях ([http://www.thescientificworld.com/supplements/2007.270/Buzdin\\_031307\\_SuppTable1.xls](http://www.thescientificworld.com/supplements/2007.270/Buzdin_031307_SuppTable1.xls)) [32]. Дальнейшие наши исследования показали, что некоторые hsERV работают как энхансеры известных генов

человека и даже как промоторы для регуляторных антисмысловых РНК.

### ГЕНОМНАЯ СТРУКТУРА hsERV

hsERV образуют 330 т.п.о. ДНК человека [32], что составляет всего 5% от ДНК, созданной вставками человекспецифичных RE. Все hsERV принадлежат к единственной из более чем сорока групп эндогенных ретровирусов человека — HERV-K (HML-2) и составляют 7% от общего числа ее членов [32, 33].

Все ERV включают последовательности, родственные ретровирусным генам *gag*, *prot*, *pol* и *env*, и фланкированы так называемыми длинными концевыми повторами длиной около 1 т.п.о. (англ. long terminal repeat (LTR)). Структура LTR содержит функциональный энхансер [34], промотор и сигнал полиаденилирования [35], которые используются для экспрессии ретровирусных генов. Кроме того, LTR может влиять на транскрипционный статус соседних уникальных последовательностей генома [36]. Большая часть ERV присутствует в виде одиночных LTR, образованных, скорее всего, благодаря гомологичной рекомбинации между LTR полноразмерных ERV [37]. Некоторые полноразмерные ERV экспрессируют свои гены в различных тканях человека и даже способны образовывать вирусоподобные частицы в плаценте. Такие вирусоподобные частицы, сформированные представителями HERV-K (HML-2), могут содержать внутри себя как самые разнообразные провирусные транскрипты, так и РНК клеточных генов [38]. Подобная “всеядность”, по-видимому, является залогом множественных рекомбинационных событий, убыстряющих эволюцию геномов ретровирусов.

Как и другие ERV, группа HERV-K (HML-2) содержит по большей части дефектные и транскрипционно неактивные копии провирусов. Однако некоторые ее члены не только сохранили транскрипционную активность, но и до сих пор могут обладать инфекционным потенциалом [39]. Группа HERV-K (HML-2) включает по крайней мере 55 полноразмерных провирусов и около 2000 одиночных LTR [40].

Считается, что LTR аккумулируют весь регуляторный потенциал ERV. Было показано, что LTR HERV-K специфично связывают некоторые ядерные белки [41], работают как тканеспецифические промоторы и энхансеры [42], а также подвергаются дифференциальному метилированию в различных тканях [43] и, наконец, транскрибируются *in vivo* [44].

Большая часть hsERV была идентифицирована еще до публикации полных последовательностей геномов человека и шимпанзе. Полученные в основном при помощи метода вычитающей гибридизации последовательности hsERV послужили основой для создания консенсусной последовательности LTR эндогенных ретровирусов, названной “HS

family”, которая позволила обнаружить основную часть hsERV в экспериментальной работе или в генных базах данных [33]. Пик ретропозиций предшественников “HS family” приходился на период после расхождения предковых линий человека и шимпанзе, которое произошло 4–6 млн лет назад [45]. По меньшей мере 12 членов этого семейства полиморфны в современной человеческой популяции [32], что свидетельствует об их недавней интеграции в геном *Homo sapiens* и о возможной активности этой группы вплоть до настоящего времени.

### ЖИЗНЕННЫЙ ЦИКЛ HERV-K (HML-2)

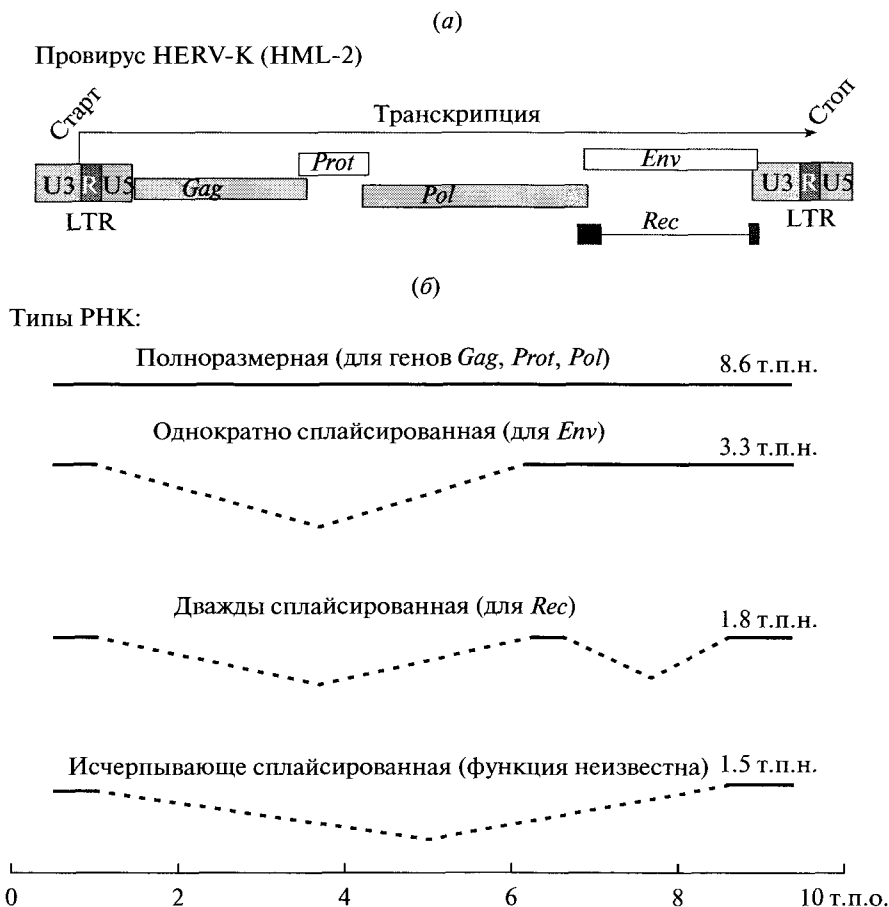
Жизненный цикл hsERV включает транскрипцию геномной ДНК-копии провируса РНК-полимеразой II, обратную транскрипцию вирусной геномной РНК и интеграцию получившейся ДНК-копии в геном клетки-хозяина.

Геномная копия провируса, как правило, транскрибируется с расположенного в составе 5'-LTR промотора (рис. 2а). Транскрипция останавливается после сигнала полиаденилирования, находящегося на 3'-LTR. Полиаденилированный полноразмерный транскрипт может сплайсироваться, что приводит к образованию как минимум трех сплайсформ (рис. 2б). Несплайсированный транскрипт кодирует вирусный полипептид Gag-Prot-Pol массой 160 кДа. Для трансляции полной последовательности полипротеина требуются два сдвига рамки считывания на рибосоме [46]. Полипротеин затем процессируется входящей в его состав ретровирусной протеиназой Prot. Белок Pol представляет собой ретровирусную обратную транскриптазу, обладающую активностью РНКазы Н. Однократно сплайсированный транскрипт кодирует белок оболочки вируса Env [46].

РНК, подвергнутая двойному сплайсингу, кодирует короткий регуляторный белок Res массой 15 кДа, также называемый cORF. По-видимому, Res обладает функциональной гомологией с лентивирусными РНК-связывающими белками ядерного экспорта Rev (ВИЧ) и Rex (HTLV) [46]. Кроме того, образуются и полностью сплайсированные провирусные транскрипты длиной 1.5 т.п.о., не несущие белкокодирующей последовательности и, возможно, обладающие регуляторными функциями.

### ПРОДУКЦИЯ ВИРУСНЫХ БЕЛКОВ HERV-K (HML-2)

Проведенный нами биоинформатический анализ показал, что в общей сложности hsERV содержат 11 интактных рамок считывания для белка Gag, 12 — для Prot, 9 — для Pol, 8 — для Env и 9 — для Res [32]. Помимо вовлеченности этих белковых продуктов в жизненный цикл ERV, им также принадлежат некоторые физиологические функции. Белки HERV-K (HML-2) продуцируются во многих тка-



**Рис. 2.** Функциональные гены, кодируемые эндогенными ретровирусами семейства HERV-K (HML-2). (а) Геномная структура полноразмерного провируса. Помимо “классических” ретровирусных генов *Gag*, *Prot*, *Pol* и *Env*, он содержит дополнительный ген *Rec*. (б) Различные виды транскриптов провируса. Полноразмерный транскрипт содержит рамки считывания для полипротеина Gag-Prot-Pol, однократно сплайсированный продукт кодирует белок оболочки вируса Env, дважды сплайсированная РНК кодирует белок Rec, тогда как исчерпывающе сплайсированный транскрипт длиной около 1.5 т.п.н. не содержит открытых рамок считывания.

нях человека [46]. Так, антитела против эпитопов белка Env присутствуют в крови у 30% людей [47]. Теоретически, провирусные белки могут провоцировать развитие аутоиммунных заболеваний [46].

В ходе прогрессии рака клеток зародышевого пути наблюдается заметное увеличение продукции белков HERV-K (HML-2). При этом после излечения у пациентов падал титр антител против ретровирусных белков и через 5 лет становился слабодетектируемым [48]. Помимо повышенной продукции в некоторых других видах опухолей, белки HERV-K (HML-2) образуются также в плаценте и в эмбриональных тканях [49].

Также показано, что белок Gag может индуцировать стимуляцию или апоптоз Т-лимфоцитов [50]. Эндогенный белок Prot может использоваться экзогенными ретровирусами в ходе инфекции. Например, лентивирусы ВИЧ и вирус иммунодефицита обезьян SIV не содержат функционального домена

dUTP-азы, который, по-видимому, предоставляется имеющим такой домен белком Prot эндогенных ретровирусов [51], что согласуется с наблюдением о трансактивации HERV-K (HML-2) в ходе инфекции ВИЧ *in vitro* и *in vivo* [52]. Белок Rec может взаимодействовать с клеточными механизмами ядерно-цитоплазматического транспорта [53]. Наконец, Env обладает иммуносупрессивным доменом, способным ингибировать активацию Т- и В-лимфоцитов [54]. Последнее свойство может быть связано с повышенной экспрессией HERV-K (HML-2) в некоторых видах рака. Помимо отрицательных эффектов, белок Env может обеспечивать частичную устойчивость хозяйской клетки к инфекции родственными вирусами при помощи рецепторной интерференции [55]. Эндогенный белок Gag также может обеспечивать защиту от ретровирусной инфекции: например, экспрессия мышьиной эндогенной последовательности гена *gag* Fv1 блокирует инфек-

цию мышинным вирусом лейкемии MLV благодаря вхождению эндогенного белка в капсид MLV, который от этого становится нефункциональным [56].

### РЕГУЛЯТОРНЫЙ ПОТЕНЦИАЛ hsERV

134 hsERV, содержащие функциональные энхансеры, промоторы, сигналы полиаденилирования и сплайс-сайты, могут быть вовлечены в следующие схемы регуляции транскриптома человека:

(i) энхансерная активность LTR может менять транскрипционный профиль предсуществующих генов хозяина;

(ii) LTR могут предоставлять промоторы для транскрипции уникальных геномных последовательностей;

(iii) сайты полиаденилирования в составе LTR могут приводить к преждевременной терминации транскриптов предсуществующих генов;

(iv) сплайс-сайты в составе LTR могут изменять экзон-интронную структуру генов;

(v) LTR могут регулировать экспрессию генов хозяина при помощи механизмов РНК-интерференции.

Далее более подробно рассмотрены упомянутые схемы. Важно, что 61 hsERV (45% от их общего числа) расположен непосредственно в составе или вблизи известных генов человека.

*Энхансерная активность.* В исследованиях на клеточных культурах было показано, что LTR HERV-K обладают ярко выраженной тканеспецифичной энхансерной активностью [34, 57], при этом трансфицируемые конструкции содержали LTR в положении энхансера в обеих ориентациях перед стандартными промоторами CMV или SV40. Из 10 протестированных клеточных линий, значительная активность была показана только в клеточной линии Tera-1 (происхождение — эмбриональная герминогенная опухоль), при этом в присутствии LTR-энхансера активность репортерного гена люциферазы увеличивалась в 8 раз [34]. В наших исследованиях (неопубликованные данные), в отличие от упомянутых использованных искусственных систем, использовались нативные пары LTR-промотор гена из 9 обнаруженных геномных локусов человека, содержащих hsERV вблизи (до 5 т.п.о.) от точек начала транскрипции известных генов. При этом сравнивалась активность репортерного гена в беспромоторных конструкциях, содержащих перед геном люциферазы вставки геномных локусов, содержащих и не содержащих LTR. Оказалось, что в случае трех генов человека наблюдается значительный энхансерный эффект (усиление транскрипции в 3—11 раз) в нескольких из протестированных четырех клеточных культур различного происхождения. Проведенный анализ метилирования соответствующих геномных локусов показал, что для всех таких локусов в ряде тканей содержатся метилирован-

ные LTR в энхансерном положении, что свидетельствует о реальном влиянии идентифицированных трех вставок hsERV на экспрессию генов в клетке (Е.В. Гогвадзе и др., неопубликованные данные).

*Промоторная активность.* Промоторная активность LTR hsERV исследовалась в экспериментах как *in vitro*, так и *in vivo*. В экспериментах по транзientной трансфекции с конструкциями, содержащими репортерный ген, была показана сравнительно слабая (<10% относительно промотора SV40) промоторная активность в 3 из исследованных 10 клеточных культур, умеренная активность (10—20%) в 6 линиях и сильная активность (более 100% от SV40) — в единственной культуре Tera-1 [58]. Промоторная сила LTR HERV-K зависит от наличия CG-динуклеотидов в их последовательности, и поэтому накопились больше мутаций более эволюционно “старые” ERV экспрессируются значительно слабее недавно появившихся в геноме элементов, таких, как hsERV. Также было показано, что промоторная активность LTR в значительной мере зависит от степени метилирования этих динуклеотидов [42].

В проведенных нами позднее исследованиях было определено, что последовательность LTR несет два функциональных *in vivo* промоторных участка [36]. Полногеномный скрининг транскрипции hsERV в герминогенных тканях человека (в нормальной паренхиме яичка и соответствующей опухоли — семинеме; взяты от одного пациента) при помощи разработанного нами метода GREM [59] выявил весьма высокую (до 3% от уровня бета-актина) транскрипционную активность некоторых индивидуальных hsERV [60]. В общей сложности, 50% всех LTR hsERV обладали промоторной активностью *in vivo*, при этом с последовательности LTR шла транскрипция и соседних уникальных участков генома. Активность индивидуальных LTR hsERV была проанализирована количественно. Оказалось, что отсутствует связь между транскрипционной активностью и сходством первичной структуры hsERV, и наоборот, важнейшим фактором является положение hsERV в геноме. Так, вставки hsERV в ген-богатые локусы показывали высокие уровни транскрипции, тогда как hsERV в обедненных генами участках транскрибировались слабо. В общей сложности, вставки hsERV создали более 60 человекспецифичных промотора, с которых идет транскрипция уникальной геномной ДНК.

*Полиаденилирование транскриптов.* Исследования баз данных транскрибируемых последовательностей выявило четыре различных мРНК, терминация которых проходит по последовательности hsERV [32]. Среди них, три являются кодирующими и одна кодирует белок неизвестной функции. Таким образом, по крайней мере один предоставленный hsERV сигнал полиаденилирования ис-

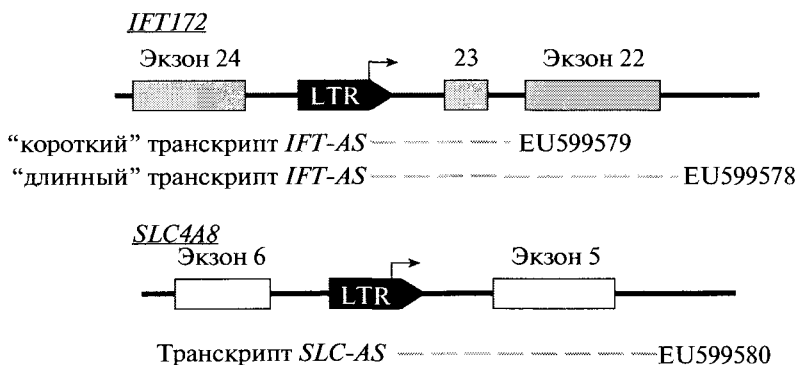


Рис. 3. Схема локусов генов *IFT172* и *SLC4A8* и варианты обнаруженных антисмысловых транскриптов.

пользуется для транскрипционной терминации функциональной кодирующей белок РНК.

**Сплайсинг по последовательности** *hsERV*. Был найден ряд сплайсированных некодирующих РНК, содержащих экзоны *hsERV*, объединенные с экзонами других геномных последовательностей, но не было найдено *hsERV*-сплайсированных транскриптов для известных генов.

**Антисмысловая регуляция экспрессии.** Большая часть (>90%) вставленных в интроны генов *hsERV* закрепились в геноме в антисмысловой ориентации, что может быть связано с интерференцией между транскрипцией гена и сигналами полиаденилирования *hsERV* в прямой ориентации [61]. Таким образом, активная транскрипция антисмысловых интронных *hsERV* приводит к образованию РНК, комплементарных экзонам генов хозяина, что может являться одним из механизмов регуляции их экспрессии.

Как было показано в наших исследованиях, среди 28 интронных обратноориентированных *hsERV* 15 обладают промоторной активностью в герминогенных тканях [60]. При помощи методов ОТ-ПЦР в реальном времени и 5'RACE\* мы исследовали транскрипцию этих элементов в различных тканях человека. Было выявлено, что для двух известных генов человека с *hsERV* образуются транскрипты, перекрывающиеся с соседними экзонами. В этих случаях LTR располагались в интронах кодирующих белки генов *IFT172* и *SLC4A8* (рис. 3). В первом случае образуется два антисмысловых транскрипта: "короткий", содержащий последовательность экзона 23, и "длинный", содержащий последовательности, комплементарные экзонам 22 и 23. Для гена *SLC4A8* был обнаружен единственный антисмысловый транскрипт, включающий последовательность, комплементарную экзону 5 [62]. Исследование на панели кДНК 11 различных тканей человека позволило обнаружить транскрипт *SLC-AS* в 5 тканях,

"короткий" транскрипт *IFT-AS* — в 10 тканях и "длинный" — в единственной ткани (паренхима яичка). Интересно, что уровень транскрипции *IFT-AS* был весьма высок в сердце, паренхиме яичка и в семяноме и превосходил в этих тканях уровень самого гена-мишени *IFT172* до 16 раз.

Функциональный эффект антисмысловых транскриптов был исследован на клетках линии Tera-1, где изначально наблюдался низкий уровень экспрессии всех упомянутых антисмысловых транскриптов и высокий — обоих генов-мишеней. Были получены стабильно трансфицированные линии, несущие в своем геноме вставку конструкции, содержащей антисмысловые транскрипты под контролем конститутивного цитомегаловирусного промотора. Были отобраны стабильно трансфицированные линии, экспрессирующие антисмысловые транскрипты на физиологическом уровне (то есть на уровне или ниже уровня, зафиксированного для тканей человека). В полученных линиях при помощи ОТ-ПЦР в реальном времени измеряли концентрацию мРНК генов-мишеней. Оказалось, что повышение уровня *SLC-AS* в 4 раза (до физиологического уровня, наблюдаемого в гипоталамусе) привело к снижению концентрации мРНК гена *SLC4A8* в 4 раза. Увеличение продукции "длинного" и "короткого" транскриптов *IFT-AS* до уровня их транскрипции в базальных ядрах головного мозга и в сердце, соответственно — вызвало снижение концентрации мРНК *IFT172* в 2.9 и 1.8 раз, соответственно.

Ген *SLC4A8* кодирует белок, обменивающий катионы  $\text{Na}^+$  с анионами  $\text{Cl}^-$  и  $\text{HCO}_3^-$ , принадлежащий к семейству натрий-бикарбонатных транспортеров [63]. Продукт гена *IFT172* является высококонсервативным гомологом внутрижгутикового транспортного белка IFT мыши [64]. Было показано, что этот ген играет важную роль в сигнализации по пути hedgehog в ходе эмбриогенеза позвоночных. Оба упомянутых гена транскрибируются в человеческом мозге в ходе развития.

\* 5'RACE, метод селективной амплификации 5'-концов кДНК, от англ. "rapid amplification of cDNA ends".

ЗАКЛЮЧЕНИЕ

Таким образом, детальное полногеномное исследование *hsERV* выявило, что они могут регулировать работу генов человека как действуя в положении *цис*-регуляторов (энхансеры генов), так и в положении *транс* (в качестве антисмысловых регуляторов). Было выявлено пять генов человека, в значительной мере подверженных человекспецифичной регуляции. Учитывая, что была проанализирована относительно небольшая группа человекспецифичных ретроэлементов, составляющая всего 2% от их числа, можно ожидать, что детальное функциональное исследование всех семейств человекспецифичных RE позволит выявить десятки или сотни генов, регулируемых в геноме человека уникальным образом.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке программы Президиума Российской академии наук по молекулярной и клеточной биологии, а также гранта РФФИ № 08-04-00720-а.

СПИСОК ЛИТЕРАТУРЫ

1. Ebersberger I., Metzler D., Schwarz C., Paabo S. // *Am. J. Hum. Genet.* 2002. V. 70. P. 1490–1497.
2. Goodman M. // *Am. J. Hum. Genet.* 1999. V. 64. P. 31–39.
3. Haile-Selassie Y. // *Nature.* 2001. V. 412. P. 178–181.
4. Buzdin A. // *Current Pharmacogenomics.* 2006. V. 4. P. 1–8.
5. Buzdin A.A. // *Cell. Mol. Life Sci.* 2004. V. 61. P. 2046–2059.
6. Mills R.E., Bennett E.A., Iskow R.C., Luttig C.T., Tsui C., Pittard W.S., Devine S.E. // *Am. J. Hum. Genet.* 2006. V. 78. P. 671–679.
7. International human genome sequencing consortium. // *Nature.* 2001. V. 409. P. 860–921.
8. King C., Wilson C. // *Science.* 1975. V. 188. P. 107–116.
9. Chimpanzee genome sequencing consortium. // *Nature.* 2005. V. 437. P. 69–87.
10. Prabhakar S., Noonan J.P., Paabo S., Rubin E.M. // *Science.* 2006. V. 314. P. 786.
11. Yunis J.J., Prakash O. // *Science.* 1982. V. 215. P. 1525–1530.
12. Nickerson E., Nelson D.L. // *Genomics.* 1998. V. 50. P. 368–372.
13. Meneveri R., Agresti A., Rocchi M., Marozzi A., Ginelli E. // *J. Mol. Evol.* 1995. V. 40. P. 405–412.
14. Horvath J.E., Viggiano L., Lofius B.J., Adams M.D., Archidiacono N., Rocchi M., Eichler E.E. // *Hum. Mol. Genet.* 2000. V. 9. P. 113–123.
15. Royle N.J., Baird D.M., Jeffreys A.J. // *Nat. Genet.* 1994. V. 6. P. 52–56.
16. Glaser B., Yen P.H., Schempp W. // *Chromosome Res.* 1998. V. 6. P. 481–486.
17. Kehrer-Sawatzki H., Cooper D.N. // *Hum. Genet.* 2007. V. 120. P. 759–778.

18. Chou H.H., Takematsu H., Diaz S., Iber J., Nickerson E., Wright K.L., Muchmore E.A., Nelson D.L., Warren S.T., Varki A. // *Proc. Natl. Acad. Sci. USA.* 1998. V. 95. P. 11751–11756.
19. Gilad Y., Man O., Paabo S., Lancet D. // *Proc. Natl. Acad. Sci. USA.* 2003. V. 100. P. 3324–3327.
20. Yu H., Jiang H., Zhou Q., Yang J., Cun Y., Su B., Xiao C., Wang W. // *Hum. Mol. Genet.* 2006. V. 15. P. 1870–1875.
21. Gagneux P., Varki A. // *Mol. Phylogenet. Evol.* 2001. V. 18. P. 2–13.
22. Zimonjic D.B., Kelley M.J., Rubin J.S., Aaronson S.A., Popescu N.C. // *Proc. Natl. Acad. Sci. USA.* 1997. V. 94. P. 11461–11465.
23. Lane R.P., Cufforth T., Young J., Athanasiou M., Friedman C., Rowen L., Evans G., Axel R., Hood L., Trask B.J. // *Proc. Natl. Acad. Sci. USA.* 2001. V. 98. P. 7390–7395.
24. Livak K.J., Rogers J., Lichter J.B. // *Proc. Natl. Acad. Sci. USA.* 1995. V. 92. P. 427–431.
25. Nadezhdin E.V., Vinogradova T.V., Sverdlov E.D. // *Dokl. Biochem. Biophys.* 2001. V. 381. P. 415–418.
26. Enard W., Khaitovich P., Kloze J., Zollner S., Heissig F., Giavalisco P., Nieselt-Struwe K., Muchmore E., Varki A., Ravid R., Doxiadis G.M., Bontrup R.E., Paabo S. // *Science.* 2002. V. 296. P. 340–343.
27. Buzdin A., Ustyugova S., Gogvadze E., Lebedev Y., Hunsmann G., Sverdlov E. // *Hum. Genet.* 2003. V. 112. P. 527–533.
28. Wang H., Xing J., Grover D., Hedges D.J., Han K., Walker J.A., Batzer M.A. // *J. Mol. Biol.* 2005. V. 354. P. 994–1007.
29. Sen S.K., Han K., Wang J., Lee J., Wang H., Callinan P.A., Dyer M., Cordaux R., Liang P., Batzer M.A. // *Am. J. Hum. Genet.* 2006. V. 79. P. 41–53.
30. Brosius J. // *Gene.* 1999. V. 238. P. 115–134.
31. Lebedev Y.B., Amosova A.L., Mamedov I.Z., Fisunov G.Y., Sverdlov E.D. // *Gene.* 2007. V. 390. P. 122–129.
32. Buzdin A. // *TheScientificWorld J.* 2007. V. 7. P. 1848–1868.
33. Buzdin A., Ustyugova S., Khodosevich K., Mamedov I., Lebedev Y., Hunsmann G., Sverdlov E. // *Genomics.* 2003. V. 81. P. 149–156.
34. Ruda V.M., Akopov S.B., Trubetskoy D.O., Manuylov N.L., Vetchinova A.S., Zavalova L.L., Nikolaev L.G., Sverdlov E.D. // *Virus Res.* 2004. V. 104. P. 11–16.
35. Sverdlov E.D. // *Bioessays.* 2000. V. 22. P. 161–171.
36. Kovalskaya E., Buzdin A., Gogvadze E., Vinogradova T., Sverdlov E. // *Virology.* 2006. V. 346. P. 373–378.
37. Hughes J.F., Coffin J.M. // *Proc. Natl. Acad. Sci. USA.* 2004. V. 101. P. 1668–1672.
38. Ruprecht K., Ferreira H., Flockerzi A., Wahl S., Sauter M., Mayer J., Mueller-Lantsch N. // *J. Virol.* 2008. V. 82. P. 10008–10016.
39. Belshaw R., Dawson A.L., Woolven-Allen J., Redding J., Burt A., Tristram M. // *J. Virol.* 2005. V. 79. P. 12507–12514.
40. Medstrand P., Mager D.L. // *J. Virol.* 1998. V. 72. P. 9782–9787.
41. Akopov S.B., Nikolaev L.G., Khil P.P., Lebedev Y.B., Sverdlov E.D. // *FEBS Lett.* 1998. V. 421. P. 229–233.



42. Lavie L., Kitova M., Maldener E., Meese E., Mayer J. // *J. Virol.* 2005. V. 79. P. 876–883.
43. Khodosevich K., Lebedev Y., Sverdlov E. // *Russ. J. Bioorgan. Chemistry*. 2004. V. 30. P. 441–445 (Ходоцевич К., Лебедев Ю., Свердлов Е. // *Биоорганическая химия*. 2004. Т. 30. С. 493–498).
44. Vinogradova T., Leppik L., Kalinina E., Zhulidov P., Grzeschik K.H., Sverdlov E. // *Mol. Genet. Genomics*. 2002. V. 266. P. 796–805.
45. Macfarlane C., Simmonds P. // *J. Mol. Evol.* 2004. V. 59. P. 642–656.
46. Bannert N., Kurth R. // *Proc. Natl. Acad. Sci. USA*. 2004. V. 101 Suppl 2. P. 14572–14579.
47. Herve C.A., Lugli E.B., Brand A., Griffiths D.J., Venables P.J. // *Clin. Exp. Immunol.* 2002. V. 128. P. 75–82.
48. Boller K., Konig H., Sauter M., Mueller-Lantzsch N., Lower R., Lower J., Kurth R. // *Virology*. 1993. V. 196. P. 349–353.
49. Andersson A.C., Svensson A.C., Rolny C., Andersson G., Larsson E. // *Int. J. Oncol.* 1998. V. 12. P. 309–313.
50. Hugin A.W., Vacchio M.S., Morse H.C., 3rd. // *Science*. 1991. V. 252. P. 424–427.
51. Harris J.M., McIntosh E.M., Muscat G.E. // *Arch. Virol.* 2000. V. 145. P. 353–363.
52. Contreras-Galindo R., Contreras-Galindo A., Lorenzo E., Yamamura Y. // *Retrovirology*. 2006. V. 3 Suppl 1. P. S33.
53. Galli U.M., Sauter M., Lecher B., Maurer S., Herbst H., Roemer K., Mueller-Lantzsch N. // *Oncogene*. 2005. V. 24. P. 3223–3228.
54. Denner J., Persin C., Vogel T., Hausteil D., Norley S., Kurth R. // *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 1996. V. 12. P. 442–450.
55. Melder D.C., Pankratz V.S., Federspiel M.J. // *J. Virol.* 2003. V. 77. P. 10504–10514.
56. Best S., Le Tissier P.R., Stoye J.P. // *Trends Microbiol.* 1997. V. 5. P. 313–318.
57. Illarionova A.E., Vinogradova T.V., Sverdlov E.D. // *Virology*. 2007. V. 358. P. 39–47.
58. Domansky A.N., Kopantzev E.P., Snezhkov E.V., Lebedev Y.B., Leib-Mosch C., Sverdlov E.D. // *FEBS Lett.* 2000. V. 472. P. 191–195.
59. Buzdin A., Kovalskaya-Alexandrova E., Gogvadze E., Sverdlov E. // *Nucleic Acids Res.* 2006. V. 34. P. e67.
60. Buzdin A., Kovalskaya-Alexandrova E., Gogvadze E., Sverdlov E. // *J. Virol.* 2006. V. 80. P. 10752–10762.
61. Buzdin A.A., Lebedev Yu.B., Sverdlov E.D. // *Russ. J. Bioorgan. Chemistry*. 2003. V. 29. P. 91–93 (Буздин А.А., Лебедев Ю.Б., Свердлов Е.Д. // *Биоорганическая химия*. 2003. Т. 29. С. 103–106).
62. Gogvadze E., Stukacheva E., Buzdin A., Sverdlov E. // *J. Virol.* 2009. (in press).
63. Aalkjaer C., Frische S., Leipziger J., Nielsen S., Praetorius J. // *Acta Physiol. Scand.* 2004. V. 181. P. 505–512.
64. Rosenbaum J.L., Witman G.B. // *Nat. Rev. Mol. Cell. Biol.* 2002. V. 3. P. 813–825.

## A Functional Analysis of Retroviral Endogenous Inserts in View of Human Genome Evolution

A. A. Buzdin\*

\*Phone: +7 (495) 330-6574; fax: +7 (495) 330-6538; e-mail: aton@hymgen.siobc.ras.ru

*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,  
ul. Miklukho-Maklaya 16/10, Moscow, 117997 Russia*

Retroelements, mobile elements produced in DNA by reverse transcription, comprise about 40% of the human genome. A small part of these elements appeared in the genome quite recently after the divergence of humans and chimpanzees had occurred. Evolutionarily young retroelements are represented by the members of four groups, SVA, Alu, L1, and the endogenous HERV-K (HML-2) virus. These retroelements could play a functional role in the course of the molecular evolution of human DNA. We comprehensively studied the contribution of human-specific endogenous viruses (hsERV) to the structural modifications and regulation of the human genome. We found that hsERV presented in 134 copies occupied about 330 000 bp of human DNA. They added to genomic sequences the copies of 50 functional retroviral genes as well as 134 potential promoters and enhancers, 50% of which are located in the regions adjacent to known genes, and 22% in gene introns. At least 67% of these elements are human-specific promoters *in vivo*. hsERV viruses regulate the activity of known protein-encoding genes by means of RNA interference, function as enhancers, and provide new polyadenylation signals for mRNA.

*Key words: DNA mobile elements, gene expression regulation, human genome, molecular evolution, retroelements, speciation*