

## ЭВОЛЮЦИОННО НЕДАВНИЕ ГРУППЫ ГЕНЕТИЧЕСКИХ МОБИЛЬНЫХ ЭЛЕМЕНТОВ В ГЕНОМЕ ЧЕЛОВЕКА

К.К. Баскаев, А.А. Буздин

Институт биоорганической химии им. акад. М.М. Шемякина и Ю.А. Овчинникова РАН,  
Москва, Россия, e-mail: bu3din@mail.ru

Мобильные элементы (МЭ) – фрагменты ДНК, способные к самовоспроизводству в геноме организма-хозяина. МЭ могут составлять ~40–50 % генома млекопитающих. Нам удалось описать два новых семейства МЭ, образовавшихся сравнительно недавно в ходе эволюции генома человека. Представители одного такого семейства присутствуют только в геноме человека и отсутствуют в ДНК других приматов. Это семейство, представленное ~80 членами, образовалось при комбинировании части CpG-островка человеческого гена *MAST2* и 3'-концевой последовательности ретротранспозона SVA. По нашим оценкам, это семейство, названное *CpG-SVA*, оказалось существенно более активно, чем предковое семейство *SVA*. Наши данные показывают, что регуляторная последовательность *MAST2* позволила членам нового семейства обеспечить эффективную транскрипцию в стволовых клетках-предшественниках сперматозоидов. Второе семейство, названное семейством химерных ретротранскриптов, возникло раньше, но сохраняет свою активность и поныне. Его представители образуются при помощи своеобразного механизма РНК-рекомбинации, в результате которой в геноме появляются объединенные ДНК-копии самых разных клеточных транскриптов. Мы показали, что сходные механизмы работают в геномах других млекопитающих и даже у грибов.

**Ключевые слова:** эволюция человека, генетическая нестабильность, мобильные элементы, регуляция экспрессии генов, генетические химеры, гибридное семейство ретротранспозонов.

### Введение

Повторяющиеся последовательности занимают обширную часть каждого эукариотического генома. Так, они покрывают около половины геномной ДНК млекопитающих, тогда как экзоны генов и белок-кодирующие последовательности занимают соответственно только ~3 и 1 %. Существенная часть геномных повторов содержат гены, которые, как правило, кодируют «эгоистичные» белки, необходимые для размножения мобильных элементов (МЭ) в геноме хозяина.

МЭ размножаются либо непосредственно с помощью своих ДНК-копий (ДНК-транспозоны), либо через образование промежуточных РНК-посредников (ретроэлементы), в последнем случае используя механизм «обратной транскрипции» и фермент РНК-зависимую ДНК-полимеразу, называемую обратной транскриптазой

(reverse transcriptase, RT). Заново сформированная копия элемента интегрируется в геном, используя комбинацию белков клетки-хозяина и ферментов, кодируемых собственной последовательностью МЭ (Temin, 1993; Dewannieux *et al.*, 2003; Buzdin, 2006; Jurka, 1997).

Ретроэлементы занимают более 42 % человеческой ДНК и являются единственным классом мобильных элементов, способных к транспозиции у млекопитающих в настоящее время (Sverdlov, 2000; Buzdin, 2004). Они также играют активную роль во множестве процессов функционирования генома человека (Wessler, 1998; Sverdlov, 1998, 2000; Deininger *et al.*, 2003; Buzdin, 2004). МЭ представляют собой «горячие точки рекомбинации» (например, рекомбинации между ретроэлементами *Alu* привели к делециям около 400 т.п.о. в ДНК человека (Sen *et al.*, 2006)). Известно, что ретроэлементы могут модифицировать активность преобладающих

человеческих генов (Brosius, 1999a; Deininger *et al.*, 2003; Buzdin, 2004; van de Lagemaat *et al.*, 2006). Треть всех МЭ картированы в составе генов (Mills *et al.*, 2006).

Ретроэлементы подразделяются на автономные и неавтономные, т. е. те, которые либо сами имеют ген обратной транскриптазы, либо «заимствуют» его белковый продукт у других.

Основная группа активных ретроэлементов человека – это семейство автономных ретро-транспозонов *L1*. Полноразмерные представители *L1* у приматов занимают около 6 т.п.о. и кодируют две открытые рамки считывания, одну для RT/интегразы и другую для РНК-связывающего белка. В большинстве своем вставки *L1* существенно укорочены с 5'-конца, по-видимому, вследствие часто происходящей abortивной обратной транскрипции (Boissinot *et al.*, 2000). Следующие две активные группы, ретропозоны *Alu* (~ 300 п.о. длиной) и *SVA* (размером примерно ~ 1,5 т.п.о.), являются неавтономными мобильными элементами, которые используют «чужую» RT ретро-транспозонов *L1* для размножения своих копий (Wang *et al.*, 2005). Эти две группы не кодируют белков и могут рассматриваться как паразиты системы ретро-транспозиции *L1* (Buzdin, 2004). Наконец, автономные эндогенные ретровирусы группы HERV-K (HML-2) являются наиболее сложно организованной группой человеческих мобильных элементов. Они несут три типичных ретровирусных функциональных гена и один дополнительный ген, кодирующий небольшой регуляторный белок.

Однако ретроэлементы часто функционируют как регуляторы транскрипции хозяйских РНК (Gogvadze, Buzdin, 2009). Повторяющиеся транскрипты также могут служить как мультифункциональные РНК, участвовать в антисмысловой регуляции генов и конкурировать с транскриптами хозяина за связывание с клеточными белками (Zabolotneva *et al.*, 2010).

В экспериментах на клеточных культурах показано, что полиморфные, находящиеся в интронах человеческих генов, копии *L1*- и *Alu*-элементов снижают транскрипцию аллелей, содержащих эти ретроэлементы, по сравнению с аллелями, их не содержащими (Устюгова и др., 2006; Ustyugova *et al.*, 2006; Lebedev *et al.*, 2007). Геномные повторы включают также

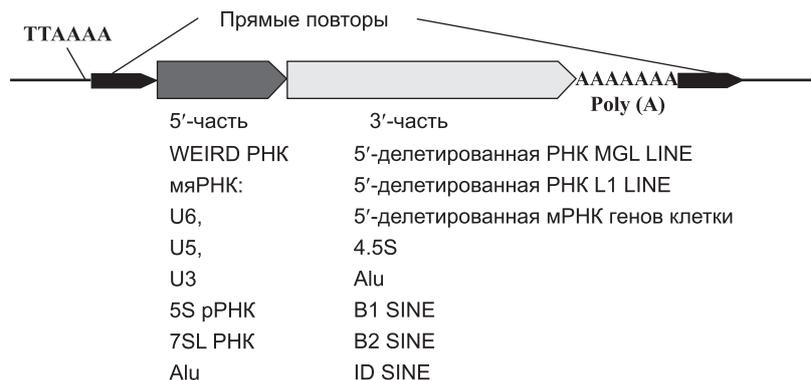
регуляторные последовательности, такие, как промоторы, энхансеры, сайты сплайсинга, сигналы полиаденилирования и инсуляторы, активно воздействующие на геном и транскриптом (Schumann *et al.*, 2010).

### Семейство мобильных элементов человека, образованное с помощью механизма РНК-рекомбинации

Типичный элемент *L1* кодирует два белка: РНК-связывающий белок ORF1p, который способствует обратной транскрипции в качестве шаперона нуклеиновых кислот (Martin, 2006), и ORF2p, являющийся обратной транскриптазой и эндонуклеазой (Kazazian, 2004).

Помимо собственных копий элементы *LINE* также способны к транспозиции и других последовательностей, в основном неавтономных МЭ, относящихся к классу *SINE*, а также копий различных клеточных РНК, что ведет к формированию псевдогенов (Dewannieux *et al.*, 2003). Недавно нами было показано, что элементы *L1* включены в формирование двойных и тройных химерных МЭ, образующихся в процессе обратной транскрипции во многих геномах включая все геномы млекопитающих и некоторые геномы грибов (Buzdin *et al.*, 2002, 2003, 2005; Fudal *et al.*, 2005; Гогвадзе и др., 2005) (рис. 1).

Всего было найдено 82, 116, 66 и 31 таких химерных элементов в ДНК человека, мыши, крысы и гриба *Magnaporthe grisea* соответственно (Buzdin *et al.*, 2002, 2003; Fudal *et al.*, 2005; Гогвадзе и др., 2005). Эти элементы состоят из ДНК-копий клеточных транскриптов как прямо соединенных между собой, так и, более часто, объединенных с 3'-концевой частью элемента *L1*. Помимо копий МЭ в составе химер были обнаружены копии разнообразных клеточных транскриптов, соответствующих мРНК белок-кодирующих генов, рибосомной РНК, различным малым ядерным РНК и 7SL РНК. Такие химеры отвечали следующим критериям: 1) их 5'-концевые части были полноразмерными копиями клеточных РНК; 2) их 3'-концевые части были укороченными с 5'-конца копиями других РНК (чаще всего МЭ); 3) сайты прерывания находились в случайных местах соответствующей РНК; 4) обе части были объединены в одной и той же ориентации; 5) химеры имели поли(А)-



**Рис. 1.** Схематическое изображение двойного химерного МЭ.

Вставка МЭ фланкирована прямыми повторами длиной 10–20 п.о.

хвост на своем 3'-конце и б) химеры были фланкированы короткими прямыми повторами.

Последний структурный признак доказывает, что эти химеры внедрялись в геном в виде единой ДНК-копии. Действительно, химеры в геноме млекопитающих несли на своих 5'-концах  $T_2A_4$ -гексануклеотиды или их варианты (Buzdin *et al.*, 2002; Гогвадзе и др., 2005), отвечающие консенсусной последовательности  $T_2A_4$ , используемой элементами *L1* для инициации обратной транскрипции (Jurka, 1997). Синхронная интеграция обеих частей химер была также подтверждена данными локус-специфической ПЦР с геномной ДНК различных видов приматов (Buzdin *et al.*, 2002, 2003).

Таким образом, химеры были образованы при помощи особого механизма. Этот механизм часто объединяет вместе клеточные транскрипты, не имеющие ничего общего с мобильными элементами (Buzdin *et al.*, 2007). Многие химеры можно рассматривать как новые гены; так, показана их транскрипция, причем некоторые транскрибируются тканеспецифично (Buzdin *et al.*, 2003; Гогвадзе и др., 2005; Gogvadze *et al.*, 2007; Gogvadze, Buzdin, 2009). Мы предположили, что эти химерные ретрогены были образованы по механизму, включающему РНК-рекомбинацию в процессе обратной транскрипции клеточной РНК (рис. 2). Эта модель подразумевает «перескок» формирующейся первой цепи кДНК с одной РНК, служащей матрицей для обратной транскрипции 3'-концевой части химеры, на другую РНК-матрицу, с которой образуется 5'-часть, и последующую интеграцию химеры в геном (Buzdin, 2004).

Хотя считающаяся основной активностью обратной транскриптазы направлена на непрерывный синтез кДНК на матрице РНК, фермент способен переключаться и на другие матрицы в процессе обратной транскрипции. Например, в ретровирусах обратная транскриптаза как минимум дважды «перепрыгивает» с одного сайта РНК-матрицы на другой, что необходимо для синтеза полноразмерных длинных концевых повторов (LTR). Кроме того, ретровирусные частицы обычно содержат две матричные молекулы РНК (Temin, 1993), по которым с высокой частотой происходят «перескоки» матрицы, значительно увеличивающие степень изменчивости ретровируса посредством рекомбинации между двумя матрицами (Kandel, Nudler, 2002). Похожий тип рекомбинации, но с «чужеродными» РНК-матрицами приводит и к образованию мозаичной структуры геномных вставок некоторых ретровирусов (Swanstrom *et al.*, 1983; Jamain *et al.*, 2001).

Данная модель образования химер впоследствии была подтверждена результатами, полученными для человеческого элемента *L1* при использовании экспериментальной системой ретротранспозиции *in vitro* (Gilbert *et al.*, 2005). Авторы охарактеризовали 100 ретротранспозиций *de novo* в клеточной линии HeLa. Одна вставка (1 %) представляла собой неизвестную ранее химеру, похожую на те, которые были обнаружены нами в человеческом геноме, состоящую из полноразмерной U6 мяРНК, соединенной с 5'-укороченным элементом *L1*. Позднее похожие результаты были получены и *in vivo* на модели ретротранспозиции *L1* в транс-



**Рис. 2.** Механизм образования химеры, использующий ферментативный аппарат *L1*.

Шаг 1 – преинтеграционный белковый комплекс *L1* связывает РНК-копию *L1*, *SINE* или клеточную РНК в цитоплазме; шаг 2 – образующийся рибонуклеопротеид переносится в ядро; шаг 3 – обратная транскрипция связанной РНК, праймированная одноцепочечным разрывом геномной ДНК; шаг 4А – успешная интеграция обратно транскрибированной кДНК-копии в геномную ДНК; шаг 4Б – «перескок» на другую РНК-матрицу в процессе обратной транскрипции; шаг 5А – интеграция химеры в геномную ДНК; шаг 5Б – второй «перескок» матрицы на другую РНК с последующей ДНК-репарацией, в результате чего образуются вставки ретротгена тройной химерной структуры. Нормальный путь интеграции элемента *L1*, соответствующий его «жизненному циклу»: шаги (1)- (2)- (3)- (4А).

генных мышях: удалось охарактеризовать 33 новых ретротранспозиционных события, 13 % из которых происходили со сменой матрицы в процессе обратной транскрипции (Babushok *et al.*, 2006).

Недавно было высказано предположение, что перескоки обратной транскриптазы с «собственной» матрицы *L1* на ДНК хозяина способствуют успешному прохождению интеграции

копии элемента в геном и, таким образом, могут представлять собой не исключение, а правило в работе этого фермента (Bibillo, Eickbush, 2004; Babushok *et al.*, 2006). Помимо образования химерных ретротгенов переключение матрицы в течение обратной транскрипции может приводить к образованию химерных неавтономных элементов *SINE* (Nishihara *et al.*, 2006) и, по-видимому, к мозаицизму структуры самих

*L1*-элементов (Hayward *et al.*, 1997; Brosius, 1999b). Эволюция отдельных семейств *L1* также может включать РНК-рекомбинацию, что приводит к объединению 3'-концевой части *L1* с новой последовательностью на 5'-конце. Как показывает сопоставление первичной структуры, 5'-нетранслируемые области человеческих, мышиных, крысиных и кроличьих семейств *L1* не гомологичны друг другу (Furano, 2000). Интересно, что также и обратная транскриптаза, кодируемая другим членом суперсемейства *LINE* (к которому относятся *L1*), R2 из геномов членистоногих, «перепрыгивает» с одной матрицы на другую в модельных экспериментах *in vitro* с образованием соответствующих химер (Bibillo, Eickbush, 2004).

Помимо этого, высказывались предположения о том, что все LTR-содержащие МЭ и многие *SINE*-элементы представляют собой сходным образом сформированные химеры (Ohshima *et al.*, 1996; Kramerov, Vassetzky, 2001, 2005; Malik, Eickbush, 2001). Действительно, филогенетический анализ домена рибонуклеазы Н показал, что LTR-содержащие элементы могли быть сформированы как химеры между ДНК-транспозонами и LTR-несодержащими ретротранспозонами, например, относящимися к *LINE* (Malik, Eickbush, 2001). *SINE*-элементы, происходящие из тРНК, вероятно, эволюционировали от ретровирусной так называемой «стронг-стоп ДНК» – ДНК-копии короткого первого фрагмента ретровирусного генома, образующейся в ходе обратной транскрипции (Ohshima *et al.*, 1996). Такие *SINE* содержат два участка: консервативный, включающий промотор тРНК и «коровый» домен, а также переменный, аналогичный 3'-концевой последовательности различных семейств *LINE*. Коровый домен тРНК-подобных *SINE* содержит консервативные последовательности, похожие на фрагменты длинных концевых повторов лизин-тРНК праймирующихся ретровирусов.

Было предположено, что происходящие от тРНК *SINE*-элементы появились вследствие интеграции ретровирусной стронг-стоп ДНК вблизи 3'-концевой части *LINE*. Образовавшиеся ретроэлементы могли быть транскрибированы РНК полимеразой III и распространились по геному. Такой механизм образования *SINE* объясняет, как они могут распространяться по

геному. Возможно, они привлекают для этого ферментативную систему *LINE*-элементов, «мигрируя» под них при помощи очень похожей «хвостовой» последовательности (Ohshima *et al.*, 1996).

### Человек-специфичное гибридное семейство CpG-SVA

Детальный структурный анализ вставок ретротранспозонов *SVA* в геноме человека позволил обнаружить 76 элементов необычной структуры. 5'-концы таких элементов несли копии первого экзона гена *MAST2*, тогда как 3'-конец – 3'-концевую последовательность ретротранспозона *SVA*.

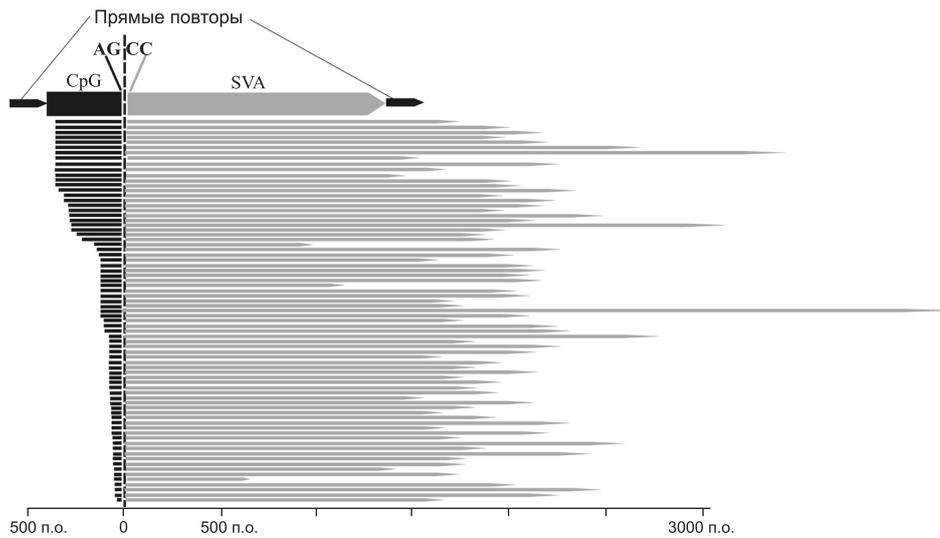
Граница между экзонной частью и *SVA*-частью располагалась точно между каноническим акцепторным сплайс-сайтом AG из экзонной части и неканоническим донорным сплайс-сайтом CC из *SVA*-части. Длина обеих частей химерного элемента могла существенно различаться – от 35 до 383 п.о. для 5'-концевой части и от 662 до 4255 п.о. для 3'-концевой части. Пограничная точка между двумя частями была неизменна для всех химер (рис. 3).

Элементы *CpG-SVA* были найдены только в человеческой ДНК, но не у других приматов, тогда как сами по себе ретротранспозоны *SVA* и последовательности экзонов *MAST2* существуют в геномах всех высших приматов. Таким образом, *CpG-SVA* являются единственным известным человек-специфичным семейством ретротранспозонов (Бантыш, Буздин, 2009).

В одновременно опубликованных других работах, описывающих то же самое семейство гибридных ретротранспозонов, это семейство названо также «*MAST2-SVA*» (Hancks *et al.*, 2009) или «*SVA-F1*» (Damert *et al.*, 2009).

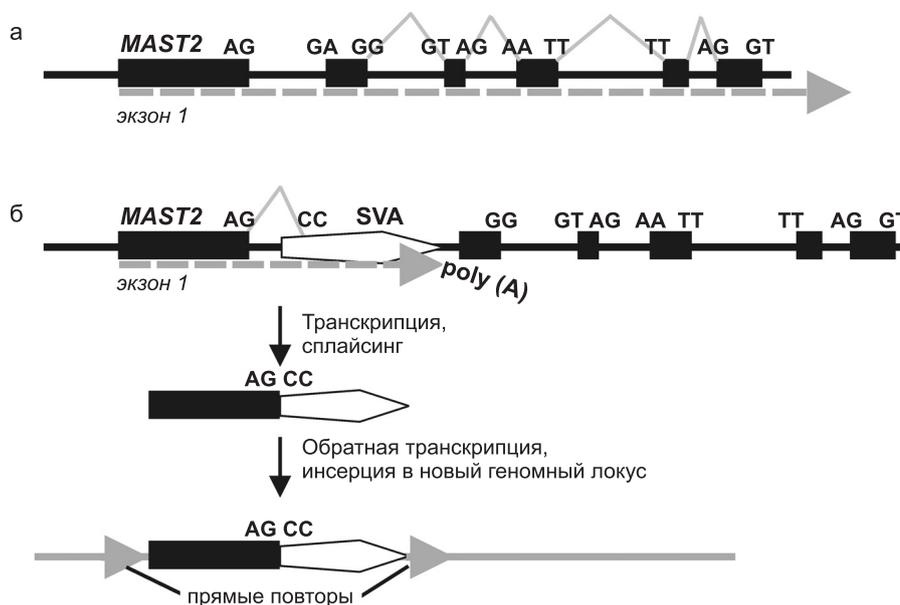
Базируясь на структурных особенностях найденных членов семейства *CpG-SVA*, мы предположили механизм их образования (рис. 4).

На первой стадии ретротранспозон *SVA*, по-видимому, включился в первый интрон гена *MAST2* в прямой ориентации. При транскрипции гена *MAST2* образовалась aberrантная пре-мРНК, терминально процессированная по 3'-концевому сигналу полиаденилирования встроенного элемента *SVA*. Эта РНК была затем сплайсирована по акцепторному сплайс-сайту



**Рис. 3.** Структуры химерного ретротранспозона *CpG-SVA*.

Вставки *CpG-SVA* фланкированы прямыми повторами. Длина 5'-концевой (экзонной) части составляет от 35 до 383 п.о., длина 3'-концевой (*SVA*-производной) части находится в диапазоне от 662 до 4255 п.о. 5'-концевая часть гомологична первому экзону гена *MAST2*, 3'-концевая часть – ретротранспозону *SVA*. Точка соединения между двумя частями идентична во всех элементах *CpG-SVA* (канонический акцепторный сплайс-сайт AG со стороны экзонной части и неканонический донорный сплайс-сайт CC со стороны *SVA*). Все *SVA*-фрагменты начинаются с позиции 396 консенсусной последовательности *SVA*.



**Рис. 4.** Предполагаемый механизм образования семейства *CpG-SVA*.

а – схематическое изображение геномного локуса, содержащего человеческий ген *MAST2*. Пунктирной стрелкой обозначено направление транскрипции, обозначены экзоны и сайты сплайсинга. б – вставка ретротранспозона *SVA* в прямой ориентации изменила экзон-интронную структуру и привела к неправильно сплайсированной мРНК, полиаденилированной по участку *SVA*. Копия этой мРНК была вставлена в новый локус человеческого генома, что привело к образованию семейства *CpG-SVA*, продолжившего свое размножение в человеческой ДНК. Однако предковый аллель гена *MAST2*, содержащего вставку *SVA*, был потерян вследствие негативного отбора.

AG со стороны гена и по донорному сплайс-сайту CC со стороны *SVA*, вследствие чего произошло объединение последовательностей первого экзона *MAST2* с 3'-концевым фрагментом *SVA* (начинающимся с 393-й позиции консенсусной последовательности *SVA*). Донорный сайт сплайсинга CC внутри элемента *SVA* является неканоническим, что объясняется особенностями экзон-интронной структуры *MAST2*, где неканонические сайты сплайсинга присутствуют в большинстве. Сплайсированной химерной РНК затем «повезло» подвергнуться обратной транскрипции при помощи ферментативной системы L1, чему сопутствовала интеграция образовавшейся кДНК в геном. Это событие привело к появлению в человеческой ДНК так называемой «мастер-копии» элемента *CpG-SVA*. «Мастер-копия» оказалась транскрипционно активной, возможно, вследствие включения в свой состав CpG-островка гена, что привело к образованию нового семейства МЭ (рис. 4).

Интересно отметить, что в настоящее время не существует известного включения элемента *SVA* в интрон гена *MAST2* в геноме человека. По-видимому, предковый аллель, содержащий целый элемент *SVA* в интроне гена, был элиминирован отрицательным отбором, поскольку такой ген не мог образовывать функциональной РНК из-за нарушения сплайсинга пре-мРНК и/или преждевременного полиаденилирования транскрипта по последовательности *SVA*.

Среди *CpG-SVA*-элементов мы обнаружили случаи 5'- и 3'-концевой трансдукции фланкирующей уникальной геномной ДНК. Как и в классическом механизме 3'-трансдукции, нижележащие геномные фрагменты, по-видимому, были пойманы вследствие проскока слабого собственного сигнала полиаденилирования *SVA* с последующей терминацией на нижележащей уникальной геномной последовательности комплексом РНК полимеразы II. В случае с 5'-концевой трансдукцией транскрипция *CpG-SVA*-элементов была, по-видимому, инициирована с вышележащих геномных промоторов. Всего нами было идентифицировано 18 и 11 случаев для 5'- и 3'-концевой *CpG-SVA*-трансдукции соответственно. Размер перенесенной геномной последовательности колебался от 8 до 854 п.о. для 5'- и от 141 до 734 п.о. для 3'-концевой трансдукции. Замечателен факт, что че-

тыре элемента *CpG-SVA* содержат как 5'-, так и 3'-трансдуцированные последовательности. Эти четыре элемента обладают высокой степенью идентичности и состоят из 364 п.о. фрагмента экзона *MAST2* и последовательности *SVA* варьирующей длины – 2143–3361 п.о. Вариации длины *SVA*-фрагмента могут быть объяснены нестабильностью его внутренних микросателлитных повторяющихся модулей. Элементы *CpG-SVA* с двойной трансдукцией были фланкированы последовательностями *Alu* (членами эволюционно древнего семейства *AluSc*) на 5'-конце и последовательностью длиной 400 п.о., включающей эволюционно древний элемент *AluSp*, на 3'-конце. Это сходство структуры – доказательство происхождения всех четырех элементов от единого предкового элемента *CpG-SVA*.

Экзонная часть химер варьирует по длине, но не по структуре, тогда как часть, произошедшая от *SVA*, сильно варьирует как по длине, так и по первичной структуре. В *SVA*-частях присутствуют различные генетические изменения, такие, как инсерции, делеции, дубликации, изменение в количестве тандемных повторов и даже вставки других МЭ.

Вместе с присутствием трансдуцированных геномных последовательностей это дает возможность построить филогенетическое древо для членов семейства *CpG-SVA* для определения степени их взаимного родства. *CpG-SVA*-элементы оказались сгруппированы в три основных кластера. Хотя существуют корреляции между размером «экзонной» части и последовательностью локализации на древе, все ветви древа содержат элементы с экзонной частью самой различной длины. Кроме того, не существует связи между позицией на филогенетическом древе и длиной *SVA*-частей. В нескольких случаях различные ветви древа включают в себя элементы с одинаковыми длинами экзонных частей. Например, ветвь 2 содержит один элемент *CpG-SVA* с экзонной частью размером 364 п.о., тогда как ветвь 1 – пять таких же элементов. Экзонная часть семи элементов из ветви 2 и один элемент из ветви 3 обладают длиной 148 п.о. Такие же совпадения обнаружены и для длин 64, 76, 88 и 361 п.о., что свидетельствует о множественных независимых событиях в ходе формирования элементов *CpG-SVA* с идентичными экзонными частями.

Наблюдаемые особенности распределения длин экзонной части *CpG-SVA* могут быть объяснены следующими факторами: 1) могли существовать множественные сайты начала транскрипции внутри *CpG-SVA* или 2) в некоторых случаях обратная транскрипция РНК *CpG-SVA* могла терминироваться преждевременно, до прохождения полного копирования последовательности матрицы. Получающиеся укороченные вставки *CpG-SVA* могли затем давать начало другим с еще более короткими экзонными частями и т. д.

Какова функция экзонной части элементов *CpG-SVA*? Учитывая, что 1) первый экзон гена *MAST2* включает в себя *CpG*-островок, 2) *CpG*-островки играют важнейшую роль в регуляции транскрипции генов, и 3) ген *MAST2* тканеспецифично экспрессируется в яичках, можно предположить, что экзонная часть может быть полезна для усиления транскрипции представителей *CpG-SVA* в яичках. Это может быть выгодно для элементов *CpG-SVA*, так как облегчает фиксацию новых вставок в геноме. Действительно, для того чтобы передаться потомству, вставка МЭ должна осуществиться в линии герминальных клеток, которые, например, могут быть локализованы в яичках.

С точки зрения распространения в геноме, открытое нами эволюционно молодое семейство *CpG-SVA* является очень успешным: ведь потомство всего одной из более чем 1 тыс. копий *SVA*-элементов, существовавших в ДНК предка человека на время расхождения с линией шимпанзе (т. е. < 0,1 %), создало 76 новых зафиксировавшихся в геноме вставок (~ 9 % из всех 860 человек-специфичных вставок *SVA*) (Бантыш, Буздин, 2009).

Дальнейшие наши исследования показали, что последовательность *CpG*-островка гена *MAST2* действительно способна усиливать транскрипцию 3'-концевой части *SVA*, показывая вдвое большую промоторную активность по сравнению с 5'-концевой частью *SVA* (Заболотнева и др. Неопубл. данные). Кроме того, для гена *MAST2* у млекопитающих характерна тканеспецифическая экспрессия в формирующихся сперматогонияльных клетках. Там же на определенных стадиях созревания наблюдаются волны тотального деметилирования генома, что сопровождается увеличением экспрессии

ретроэлементов, обычно репрессированных при помощи метилирования. Соответственно привлечение регуляторных последовательностей *MAST2* могло обеспечить элементам *CpG-SVA* транскрипционную активацию «в нужном месте в нужное время».

### Заключение

Проведенные нами исследования показывают, что мобильные элементы генома интенсивно вовлечены в эволюционный процесс и далеки от существования в виде «застывших» форм. Вероятно, систематический и детальный анализ вставок как известных мобильных элементов, так и различных псевдогенов позволит в будущем обнаружить большое количество неизвестных на настоящий момент их групп. По-видимому, многие из них будут включать сложные составные элементы, подобные описанным в настоящей статье.

### Благодарности

Работа была поддержана грантом Российского фонда фундаментальных исследований № 10-04-00593-а, грантом Президента Российской Федерации МД-480.2010.4 и программой Президиума Российской академии наук «Молекулярная и клеточная биология».

### Литература

- Бантыш О.Б., Буздин А.А. Новое семейство мобильных элементов генома человека, образованное при захвате первого экзона гена *MAST2* ретроинвертированным *SVA* // Биохимия. 2009. Т. 74. Вып. 12. С. 1709–1717.
- Гогвадзе Е.В., Буздин А.А. Новый механизм образования ретрогенов в геномах млекопитающих: рекомбинация *in vivo* в ходе обратной транскрипции РНК // Молекуляр. биология. 2005. Т. 39. Вып. 3. С. 364–373.
- Гогвадзе Е.В., Буздин А.А., Свердлов Е.Д. Смена матриц при *LINE*-опосредованной обратной транскрипции – наиболее вероятный механизм формирования двойных и тройных химерных ретроэлементов млекопитающих // Биоорганическая химия. 2005. Т. 31. Вып. 1. С. 82–89.
- Устюгова С.В., Амосова А.Л., Лебедев Ю.Б., Свердлов Е.Д. Тканеспецифическое снижение уровня

- пре-мРНК *L1* и *Alu*-содержащих аллелей генов человека // Биорган. химия. 2006. Т. 32. Вып. 1. С. 103–106.
- Babushok D.V., Ostertag E.M., Courtney C.E. *et al.* *L1* integration in a transgenic mouse model // Genome Res. 2006. V. 16. N 2. P. 240–250.
- Bibillo A., Eickbush T.H. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon // J. Biol. Chem. 2004. V. 279. N 15. P. 14945–14953.
- Boissinot S., Chevret P., Furano A.V. *L1* (*LINE-1*) retrotransposon evolution and amplification in recent human history // Mol. Biol. Evol. 2000. V. 17. N 6. P. 915–928.
- Brosius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements // Gene. 1999a. V. 238. N 1. P. 115–134.
- Brosius J. Genomes were forged by massive bombardments with retroelements and retrosequences // Genetica. 1999b. V. 107. N 1/3. P. 209–238.
- Buzdin A.A. Retroelements and formation of chimeric retrogenes // Cell Mol. Life Sci. 2004. V. 61. N 16. P. 2046–2059.
- Buzdin A. Transposable elements and their use for target site specific gene delivery // Curr. Pharmacogenomics. 2006. V. 4. N 1. P. 1–8.
- Buzdin A., Gogvadze E., Lebrun M.H. Chimeric retrogenes suggest a role for the nucleolus in *LINE* amplification // FEBS Lett. 2007. V. 581. N 16. P. 2877–2882.
- Buzdin A., Vinogradova T., Lebedev Y., Sverdlov E. Genome-wide experimental identification and functional analysis of human specific retroelements // Cytogenet. Genome Res. 2005. V. 1101. N 1/4. P. 468–474.
- Buzdin A., Ustyugova S., Gogvadze E. *et al.* A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of *l1* // Genomics. 2002. V. 80. N 4. P. 402–406.
- Buzdin A., Gogvadze E., Kovalskaya E. *et al.* The human genome contains many types of chimeric retrogenes generated through *in vivo* RNA recombination // Nucl. Acids Res. 2003. V. 31. N 15. 4385–4390.
- Damert A., Raiz J., Horn A.V. *et al.* 5'-Transducing *SV40* retrotransposon groups spread efficiently throughout the human genome // Genome Res. 2009. V. 19. N 11. P. 1992–2008.
- Deininger P.L., Moran J.V., Batzer M.A., Kazazian H.H. Jr. Mobile elements and mammalian genome evolution // Curr. Opin. Genet. Dev. 2003. V. 13. N 6. P. 651–658.
- Dewannieux M., Esnault C., Heidmann T. *LINE*-mediated retrotransposition of marked *Alu* sequences // Nat. Genet. 2003. V. 35. N 1. P. 41–48.
- Fudal I., Bohnert H.U., Tharreau D., Lebrun M.H. Transposition of *MINE*, a composite retrotransposon, in the avirulence gene *ACE1* of the rice blast fungus *Magnaporthe grisea* // Fungal. Genet. Biol. 2005. V. 42. N 9. P. 761–772.
- Furano A.V. The biological properties and evolutionary dynamics of mammalian *LINE-1* retrotransposons // Prog. Nucl. Acid Res. Mol. Biol. 2000. V. 64. P. 255–294.
- Gilbert N., Lutz S., Morrish T.A., Moran J.V. Multiple fates of *L1* retrotransposition intermediates in cultured human cells // Mol. Cell Biol. 2005. V. 25. N 17. P. 7780–7795.
- Gogvadze E., Buzdin A. Retroelements and their impact on genome evolution and functioning // Cell Mol. Life Sci. 2009. V. 66. N 23. P. 3727–3742.
- Gogvadze E., Barbisan C., Lebrun M.H., Buzdin A. Tripartite chimeric pseudogene from the genome of rice blast fungus *Magnaporthe grisea* suggests double template jumps during long interspersed nuclear element (*LINE*) reverse transcription // BMC Genomics. 2007. V. 8. P. 360.
- Hancks D.C., Ewing A.D., Chen J.E. *et al.* Exon-trapping mediated by the human retrotransposon *SV40* // Genome Res. 2009. V. 19. N 11. P. 1983–1991.
- Hayward B.E., Zavanelli M., Furano A.V. Recombination creates novel *L1* (*LINE-1*) elements in *Rattus norvegicus* // Genetics. 1997. V. 146. N 2. P. 641–654.
- Jamain S., Girondot M., Leroy P. *et al.* Transduction of the human gene *FAM8A1* by endogenous retrovirus during primate evolution // Genomics. 2001. V. 78. N 1/2. P. 38–45.
- Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons // Proc. Natl Acad. Sci. U S A. 1997. V. 94. N 5. P. 1872–1877.
- Kandel E.S., Nudler E. Template switching by RNA polymerase II *in vivo*. Evidence and implications from a retroviral system // Mol. Cell. 2002. V. 10. N 6. P. 1495–1502.
- Kazazian H.H., Jr. Mobile elements: drivers of genome evolution // Science. 2004. V. 303. N 5664. P. 1626–1632.
- Kramerov D.A., Vassetzky N.S. Structure and origin of a novel dimeric retroposon *B1-diD* // J. Mol. Evol. 2001. V. 52. N 2. P. 137–143.
- Kramerov D.A., Vassetzky N.S. Short retroposons in eukaryotic genomes // Int. Rev. Cytol. 2005. V. 247. P. 165–221.
- Lebedev Y.B., Amosova A.L., Mamedov I.Z. *et al.* Most recent *AluY* insertions in human gene introns reduce the content of the primary transcripts in a cell type specific manner // Gene. 2007. V. 390. N 1/2. P. 122–129.
- Malik H.S., Eickbush T.H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin

- of LTR retrotransposable elements and retroviruses // *Genome Res.* 2001. V. 11. N 7. P. 1187–1197.
- Martin S.L. The ORF1 protein encoded by *LINE-1*: structure and function during *LI* retrotransposition // *J. Biomed. Biotechnol.* 2006. V. 2006. P. 1–6.
- Mills R.E., Bennett E.A., Iskow R.C. *et al.* Recently mobilized transposons in the human and chimpanzee genomes // *Am. J. Hum. Genet.* 2006. V. 78. N 4. P. 671–679.
- Nishihara H., Smit A.F., Okada N. Functional noncoding sequences derived from *SINEs* in the mammalian genome // *Genome Res.* 2006. V. 16. N 7. P. 864–874.
- Ohshima K., Hamada M., Terai Y., Okada N. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements // *Mol. Cell Biol.* 1996. V. 16. N 7. P. 3756–3764.
- Schumann G.G., Gogvadze E.V., Osanai-Futahashi M. *et al.* Unique functions of repetitive transcriptomes // *Int. Rev. Cell Mol. Biol.* 2010. V. 285. P. 115–188.
- Sen S.K., Han K., Wang J. *et al.* Human genomic deletions mediated by recombination between *Alu* elements // *Am. J. Hum. Genet.* 2006. V. 79. N 1. P. 41–53.
- Sverdlov E.D. Perpetually mobile footprints of ancient infections in human genome // *FEBS Lett.* 1998. V. 428. N 1/2. P. 1–6.
- Sverdlov E.D. Retroviruses and primate evolution // *BioEssays.* 2000. V. 22. N 2. P. 161–171.
- Swanstrom R., Parker R.C., Varmus, H.E., Bishop J.M. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus // *Proc. Natl Acad. Sci. USA.* 1983. V. 80. N 9. P. 2519–2523.
- Temin H.M. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation // *Proc. Natl Acad. Sci. USA.* 1993. V. 90. N 15. P. 6900–6903.
- Ustyugova S.V., Lebedev Y.B., Sverdlov E.D. Long *LI* insertions in human gene introns specifically reduce the content of corresponding primary transcripts // *Genetica.* 2006. V. 128. N 1/3. P. 261–272.
- van de Lagemaat L.N., Medstrand P., Mager D.L. Multiple effects govern endogenous retrovirus survival patterns in human gene introns // *Genome Biol.* 2006. V. 7. N 9. P. R86.
- Wang H., Xing J., Grover D. *et al.* *SVA* elements: a hominid-specific retroposon family // *J. Mol. Biol.* 2005. V. 354. N 4. P. 994–1007.
- Wessler S.R. Transposable elements and the evolution of gene expression // *Symp. Soc. Exp. Biol.* 1998. V. 51. P. 115–122.
- Zabolotneva A., Tkachev V., Filatov F., Buzdin A. How many antiviral small interfering RNAs may be encoded by the mammalian genomes? // *Biol. Direct.* 2010. V. 5. P. 62.

## EVOLUTIONARILY RECENT GROUPS OF TRANSPOSABLE ELEMENTS IN THE HUMAN GENOME

K.K. Baskaev, A.A. Buzdin

Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia,  
e-mail: bu3din@mail.ru

Transposable elements (TEs) are fragments of genomic DNA able to self-reproduce in the host genome. They constitute 40–50 % of mammalian genomes. We have identified two TE families formed relatively recently in the course of human evolution. Members of the first family are present only in the human genome but are absent from other primate DNAs. This family, represented by about 80 members, was formed by fusion of a portion of the CpG-island in the human *MAST2* gene with the 3'-terminal fragment of *SVA* retrotransposon. According to our estimates, this hybrid family, termed *CpG-SVA*, is significantly more active than the ancestor *SVA* family. The regulatory region of *MAST2* allows copies of the new family to be transcribed in sperm precursor cells. The second family is older but is still active today. Its representatives were formed by a rather unusual RNA recombination mechanism, which mediated formation of fused DNA copies for diverse cellular transcripts. We have shown that similar mechanisms operate in the genomes of other mammals and even in fungi.

**Key words:** Human evolution, genetic instability, transposable elements, regulation of gene expression, genetic chimeras, hybrid family of retrotransposons.