



УДК 577.21

СПЕЦИФИЧНЫЕ ДЛЯ ГЕНОМА ЧЕЛОВЕКА LTR HERV-K В ИНТРОНАХ ГЕНОВ ИМЕЮТ НЕСЛУЧАЙНУЮ ОРИЕНТАЦИЮ ОТНОСИТЕЛЬНО НАПРАВЛЕНИЯ ТРАНСКРИПЦИИ И, ВОЗМОЖНО, ПРИНИМАЮТ УЧАСТИЕ В АНТИСЕНС-РЕГУЛЯЦИИ ЭКСПРЕССИИ ГЕНОВ

© 2003 г. А. А. Буздин[#], Ю. Б. Лебедев, Е. Д. Свердлов

Институт биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН,
117997 ГСП, Москва, В-437, ул. Миклухо-Маклая, 16/10

Поступило в редакцию 25.06.2002 г. Принято к печати 18.07.2002 г.

С использованием созданной консенсусной нуклеотидной последовательности длинных концевых повторов (LTR) эндогенных ретровирусов семейства К (HERV-K), предположительно специфичных для генома человека, проведен полногеномный поиск гомологий в международных базах данных. Выявлено 142 LTR, 12 из которых локализованы в интронах уникальных генов человека. Впервые показано, что 10 интронных LTR отсутствуют в ортологичных локусах генома шимпанзе, а ориентация девяти из них противоположна направлению транскрипции соответствующих генов человека. Предложена гипотеза о влиянии найденных LTR на экспрессию генов путем инициации синтеза антисмысловых РНК.

Ключевые слова: геном человека; LTR эндогенных ретровирусов; регуляция экспрессии генов; антисмысловая РНК.

Центральный вопрос теории эволюции – это вопрос о механизмах приобретения организмами новых признаков (функций), которые становятся объектом последующего естественного отбора. Широко принятое сегодня представление заключается в том, что наиболее распространенным способом возникновения функциональных новшеств является не появление новых генов, кодирующих продукты с новыми биохимическими свойствами, а изменения в регуляции существующих генов [1, 2]. В результате таких регуляторных изменений ген приобретает способность экспрессироваться в новых пространственно-временных рамках: например, на других стадиях эмбриогенеза или в других тканях взрослого организма. Такие регуляторные сдвиги могут быть, в частности, связаны с вовлечением (кооптацией) в существующую систему регуляции гена дополнительных регуляторных модулей.

Одним из источников новых модулей для кооптации служат многочисленные ретроэлементы,

постоянно мигрирующие в геноме и способные внедряться в него вблизи генов, внося в их регуляцию свои элементы. Ретроэлементы – транспозоны, использующие при создании своих ДНК-копий стадию обратной транскрипции, – занимают значительную часть геномов млекопитающих и, в частности, 41% геномной ДНК человека [3, 4]. В настоящее время накоплено множество свидетельств того, что мобильные элементы далеко не всегда являются “геномным мусором”, а зачастую участвуют в выполнении необходимых клетке и организму функций [5–7]. Широко обсуждается роль внедрений ретроэлементов в видообразовании у высших приматов и, особенно, при образовании вида *Homo sapiens* [8]. В период расхождения предковых линий человека и шимпанзе (*Pan paniscus* и *Pan troglodytes*) в геноме общего предка были активны три группы ретроэлементов: эндогенные ретровирусы HERV-K, представители LINE L1-элементы и Alu-ретропозоны [3, 8–10]. Наиболее вероятными генетическими участниками процесса видообразования человека могли быть эндогенные ретровирусы HERV-K и их длинные концевые повторы (LTR), работающие *in vivo* как энхансеры, промотеры, терминаторы транскрипции и источники сайтов сплайсинга [8]. Анализ различий в местах интеграции ретроэлементов в геномы близкородственных видов может дать информацию о тех генах, которые изменяли функ-

Сокращения: HERV-K – эндогенные ретровирусы человека семейства К; LTR и HS-LTR – длинный концевой повтор (ретроэлементов), специфичный для генома человека; LINE – длинный диспергированный ядерный повтор; EST (Expression Sequence Tag) – экспрессируемая последовательность ДНК.

[#] Автор для переписки (тел.: (095) 330-69-92; факс: (095) 330-65-38; эл. почта: anton@humgen.iobc.ras.ru).

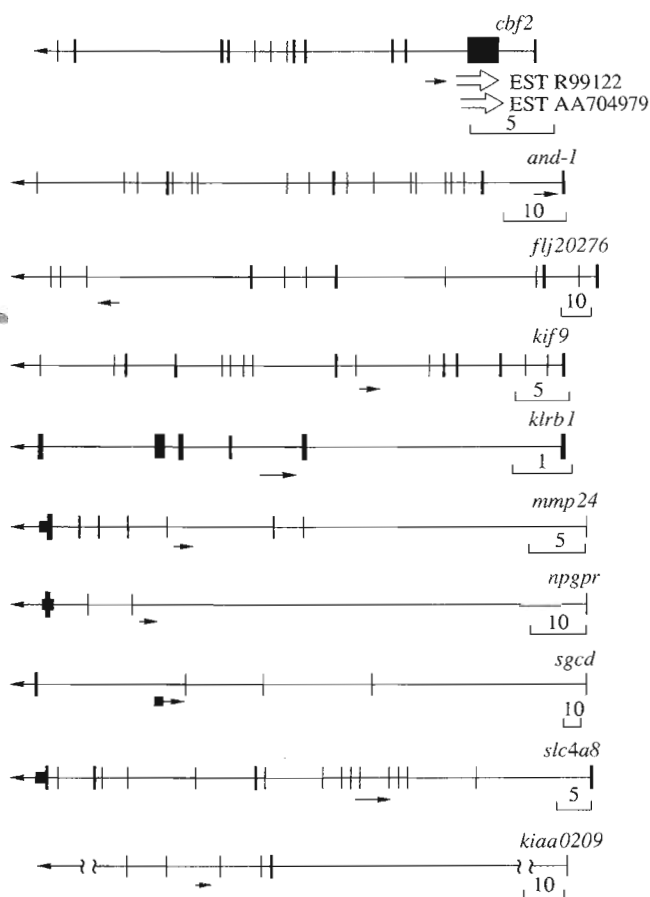


Рис. 1. Схема внедрений HS-LTR в интроны 10 известных человеческих генов.

Большими стрелками обозначено направление транскрипции генов: *cbf2* (номер(а) депонирования в GenBank – AC080067, AC007390; локус – 2p22.2), *and-1* (AL353982.3, AL160471.5; 14q22.2), *flj20276* (AL354738.13, AC015640.6, AL590377.4; 9p22.2), *kif9* (AC104447.2; 3p21.31), *klrb1* (AC092821.7, AC006432.15; 12p12.3), *mmp24* (AL121753; 20q11.22), *npgpr* (AC106898.4; 4q13.3), *sgcd* (AC027308.5, AC011345.5, AC008535.5, AC016577.4, AC025434.6; 5q33.3), *slc4a8* (AC027750.3, AC107031.2; 12q13.13) и KIAA0209/мРНК-D86964 (AC008449.6, AC008648.5, AC026695.5; 5q35.1). Положение экзонов отмечено вертикальными отрезками; положение и ориентация LTR обозначены маленькими черными стрелками, HERV-K провируса – маленькой черной стрелкой с квадратом, EST – незачерненными стрелками; масштаб в тысячах пар оснований указан для каждой последовательности.

ции в процессе видообразования и, таким образом, могли влиять на возникновение новых видов. С этой целью мы осуществляем систематическое сравнение мест интеграции ретровирусных LTR в геноме человека и высших обезьян.

Ранее на основе 40 известных последовательностей LTR HERV-K, интеграции которых специфичны для генома человека [11–14], нами была выведена консенсусная последовательность (HS-

консенсус). Скрининг геномных баз данных с использованием программ BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) выявил в ДНК человека 142 LTR, высоко идентичных HS-консенсусу и составляющих таким образом семейство, названное нами HS-LTR. С помощью сервера UCSC Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>) мы обнаружили 12 представителей HS-LTR-семейства в интронах известных человеческих генов. Согласно результатам ПЦР-анализа ортологических локусов, 10 из 12 LTR присутствуют в соответствующих участках геномных ДНК человека, но не ДНК высших приматов. Эти внедрения несомненно являются кандидатами на кооптированные регуляторные модули и требуют дальнейшего анализа. Дальнейший анализ этих 10 интронных HS-LTR показал, что все они не имеют гомологичных копий в геноме, т.е. образованы прямыми интеграциями ретроэлементов, а не в результате хромосомных дупликаций. Кроме того, эти представители HS-LTR в интронах генов имеют случайную ориентацию: в 9-ти из 10-ти соответствующих генов LTR направлен в сторону, противоположную направлению транскрипции гена (рис. 1).

Для одного из этих девяти генов, *cbf2*, при анализе баз данных нами были найдены транскрипты, противоположно ориентированные по отношению транскрипции гена и захватывающие его экзон (EST на рис. 1). Вполне возможно, что помимо промоторно-энхансерной активности, LTR могут принимать участие в регуляции генов на посттранскрипционном уровне, например, за счет РНК-интерференции [15] (рис. 2). Этот механизм регуляции основан на образовании двухцепочечных РНК между мРНК и антисенс-транскриптом с последующей каталитической деградацией всех мРНК, содержащих участки, гомологичные двухцепочечному фрагменту.

Поскольку среди известных нуклеотидных последовательностей генома человека не обнаружено фрагментов, гомологичных изучаемым десяти генам, был сделан вывод о том, что ни один из них не входит в состав семейства генов. Такая уникальность десяти генов, в которые интегрировали HS-LTR, необычна для генома человека, если предположить, что LTR после внедрения сохранили функциональные свойства. Чаще всего функциональные новшества появляются после дупликации генов в одной из дублированных копий [16]. Это позволяет второй копии сохранять неизменную функцию, и, таким образом, новшество имеет меньше шансов принести негативные следствия для организма. По-видимому, внедрение LTR, обладающих сильным сигналом терминирования транскрипции, в интроны генов в прямой ориентации могло бы вызывать преждевременную терминацию транскрипции этих генов и, как следствие, привести к их инактивации. В случае важности продуктов таких генов для организма

аллели, несущие в интронах LTR в прямой ориентации, должны были неизбежно отбрасываться в ходе эволюции генома человека.

Только один из десяти рассматриваемых генов – ген *flj20276* – содержит в своем интроне HS-LTR в прямой ориентации (рис. 1). Такое сохранение аллеля, возможно, связано с инактивацией терминатора данного LTR. В некоторых генах внедрения HS-LTR произошли в непосредственной близости от экзонов, что наряду с мощным регуляторным потенциалом LTR могло привести к плавному изменению экспрессии этих генов по механизму тканеспецифической антисмысловой регуляции (тканеспецифичность работы промотора и энхансера LTR HERV-K показана в работах [17–19]).

Чтобы найти такие антисмысловые транскрипты, располагающиеся в непосредственной близости от LTR и комплементарные экзонам соответствующих генов, мы провели поиск в базе данных экспрессирующихся последовательностей человека *est_human*. В результате нами были найдены два транскрипта (коды доступа в GenBank AA704979 и R99122), соответствующие второму интрону гена *cbf2* (второе название *cebf*) вблизи от LTR (менее 1 т.п.о.) и совпадающие с ним по ориентации (рис. 1). Оба транскрипта содержат также область, комплементарную второму экзону *cbf2*, и могут рассматриваться как возможные кандидаты на роль антисмысловых регуляторов этого гена (рис. 2). LTR в данном случае может являться тканеспецифическим регулятором экспрессии (например, энхансером, активирующим криптический промотор) указанных транскриптов. Важно, что ген *cbf2* кодирует в свою очередь транскрипционный регулятор ССААТ-Binding Factor, обуславливающий, например, экспрессию с промотора *hsp70* [20]. Изменение экспрессии такого фактора могло бы приводить к множественным эффектам за счет автоматического изменения экспрессии тех генов, с которыми взаимодействует данный фактор.

Предположение о роли антисмыслового транскрипта эндогенного ретровируса HERV-K(C4), возможно входящего в состав мРНК гена *C4* человека, было высказано ранее [21]. Синтез антисмысловых транскриптов HERV-K(C4) и негативная *C4*-зависимая регуляция в стабильно трансформированных линиях клеток мыши показаны Шнейдер с соавтор., предположившими существование HERV-K(C4)-опосредованного механизма защиты от ретровирусной инфекции [22]. Нами впервые постулируется участие антисмысловых мРНК, образующихся с интронных одиночных LTR, в регуляции экспрессии клеточных генов. Интересно, что как *cbf2*, так и LTR HERV-K транскрипционно активны, в основном, в зародышевых тканях. Некоторые из этих изменений могли бы повлечь за собой изменения в эмбрио-

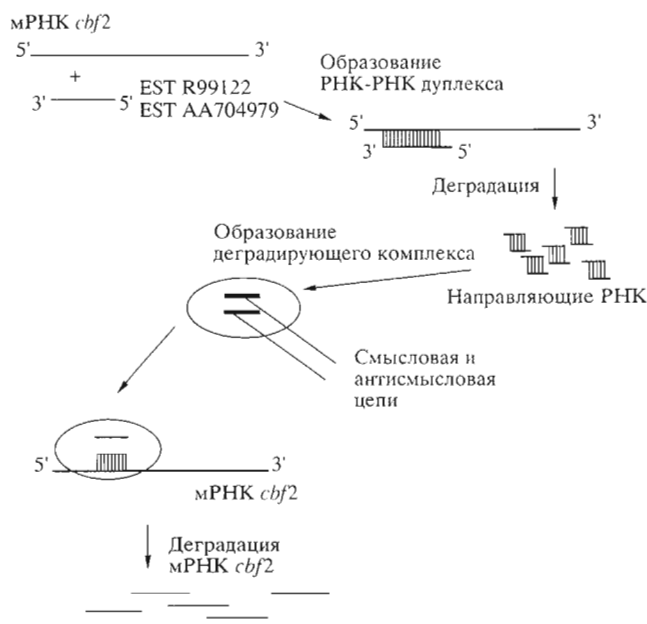


Рис. 2. Гипотетическая схема РНК-интерференции между мРНК гена *cbf2* и транскриптами R99122 и AA704979 (механизм РНК-интерференции цитируется по [15]).

нальном развитии и соответственно в фенотипах взрослых представителей *Homo sapiens* и шимпанзе *Pan paniscus* и *Pan troglodytes*.

Авторы выражают благодарность Т.В. Виноградовой и Л.Г. Николаеву за плодотворную дискуссию. Работа поддержана грантами РФФИ № 00-15-97945 и 02-15-97945.

СПИСОК ЛИТЕРАТУРЫ

1. Duboule D., Wilkins A.S. // Trends Genet. 1998. V. 14. P. 54–59.
2. King M.C., Wilson A.C. // Science. 1975. V. 188. P. 107–116.
3. International Human Genome Sequencing Consortium // Nature. 2001. V. 409. P. 860–921.
4. Kazazian H.H., Jr. // Science. 2000. V. 289. P. 1152–1153.
5. Landry J.R., Medstrand P., Mager D.L. // Genomics. 2001. V. 76. P. 110–116.
6. Brosius J. // Gene. 1999. V. 238. P. 115–134.
7. Medstrand P., Landry J.R., Mager D.L. // J. Biol. Chem. 2001. V. 276. P. 1896–1903.
8. Sverdlov E.D. // Bioessays. 2000. V. 22. P. 161–171.
9. Boissinot S., Chevret P., Furano A.V. // Mol. Biol. Evol. 2000. V. 17. P. 915–928.
10. Carroll M.L., Roy-Engel A.M., Nguyen S.V., Salem A.H., Vogel E., Vincent B., Myers J., Ahmad Z., Nguyen L., Sammarco M., Watkins W.S., Henke J., Makalowski W., Jorde L.B., Deininger P.L., Batzer M.A. // J. Mol. Biol. 2001. V. 311. P. 17–40.

11. *Medstrand P., Mager D.L.* // *J. Virol.* 1998. V. 72. P. 9782–9787.
12. *Lebedev Y.B., Belonovitch O.S., Zybroya N.V., Khil P.P., Kurdyukov S.G., Vinogradova T.V., Hunsmann G., Sverdlov E.D.* // *Gene.* 2000. V. 247. P. 265–277.
13. *Buzdin A., Khodosevich K., Mamedov I., Vinogradova T., Lebedev Y., Hunsmann G., Sverdlov E.* // *Genomics.* 2002. V. 79. P. 413–422.
14. *Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K., Lenz J.* // *Curr. Biol.* 1999. V. 9. P. 861–868.
15. *Свердлов Е.Д.* // *Биоорганич. химия.* 2001. Т. 27. С. 237–240.
16. *Lynch M., Conery J.S.* // *Science.* 2000. V. 290. P. 1151–1155.
17. *Vinogradova T.V., Leppik L.P., Nikolaev L.G., Akopov S.B., Kleiman A.M., Senyuta N.B., Sverdlov E.D.* // *Virology.* 2001. V. 290. P. 83–90.
18. *Lower R.* // *Trends Microbiol.* 1999. V. 7. P. 350–356.
19. *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.
20. *Lum L.S., Sutzman L.A., Kaufman R.J., Linzer D.I., Wu B.J.* // *Mol. Cell. Biol.* 1990. V. 10. P. 6709–6717.
21. *Dangel A.W., Mendoza A.R., Baker B.J., Daniel C.M., Carrol M.C., Wu L.C., Yu C.Y.* // *Immunogenetics.* 1994. V. 40. P. 425–436.
22. *Schneider P.M., Witzel-Schlomp K., Rittner C., Zhang L.* // *Immunogenetics.* 2001. V. 53. P. 1–9.

Human-Specific HERV-K Intron LTRs Have Nonaccidental Opposite Orientation Relative to the Direction of Gene Transcription and Might Be Involved in the Antisense Regulation of Gene Expression

A. A. Buzdin[#], Yu. B. Lebedev, and E. D. Sverdlov

[#] Phone: +7 (095) 330-6992, fax: +7 (095) 330-6538, e-mail: anton@humgen.siobc.ras.ru
Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,
ul. Miklukho-Maklaya 16/10, GSP Moscow, 117997 Russia

A consensus nucleotide sequence of long terminal repeats (LTRs) of endogenous human-specific retroviruses of the K family (HERV-K) was constructed and used for the genome-wide search for homologies in international databases. There were revealed 142 LTRs, 12 of which were localized in introns of unique human genes. It was found for the first time that ten intron LTRs are absent in the orthologic loci of the chimpanzee genome and the orientation of nine of them is opposite to the transcription direction of the corresponding human genes. A hypothesis was propounded that the found LTRs affect the gene expression by initiation of the antisense RNA synthesis. The English version of the paper: *Russian Journal of Bioorganic Chemistry*, 2003, vol. 29, no. 1; see also <http://www.maik.ru>.

Key words: antisense RNA, HERV-K LTR, human genome, regulation of gene expression