МОЛЕКУЛЯРНАЯ БИОЛОГИЯ, 2005, том 39, № 4, с. 655–663

ОБЗОРНЫЕ И ЭКСПЕРИМЕНТАЛЬНЫЕ СТАТЬИ

УДК 577.241

Авторы закончили, хотя и совсем в разные времена, одну и ту же кафедру, руководимую на протяжении многих лет Андреем Николаевичем Белозерским. Авторы с большой теплотой и благодарностью вспоминают творческую атмосферу, созданную академиком Белозерским и сохраненную его преемниками.

МАЛАЯ ЯДРЫШКОВАЯ РНК U87 МЛЕКОПИТАЮЩИХ И ЕЕ ГЕН-ХОЗЯИН

© 2005 г. Ю. А. Макарова, Д. А. Крамеров*

Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991

Поступила в редакцию 27.01.2005 г.

Большинство малых ядрышковых РНК семейства C/D определяют участок метилирования рибозы в рРНК. Новая малая ядрышковая РНК, U87, обнаруженная нами у млекопитающих, относится к C/D-семейству и участвует в метилировании рибозы гуанилового остатка в положении 3468 28S рРНК. У грызунов и человека описан так называемый ген-хозяин, интрон которого кодирует U87 РНК. Показано, что транскрипция этого гена человека начинается с трех расположенных рядом нуклеотидов, тогда как в гене крысы имеются множественные точки инициации транскрипции. РНК гена-хозяина по ряду признаков (сплайсинг, полиаденилирование и цитоплазматическая локализация) напоминает мРНК, однако в ней отсутствуют протяженные открытые рамки считывания, а нуклеотидная последовательность характеризуется значительно меньшим уровнем консервативности по сравнению с мРНК большинства генов-хозяев малых ядрышковых РНК. Обсуждается вопрос о том, может ли низкая степень консервативности мРНК-подобного транскрипта этого гена быть свидетельством нефункциональности его экзонных последовательностей.

Ключевые слова: человек, крыса, мышь, малая ядрышковая РНК, белок-некодирующие гены, рРНК, метилирование РНК.

MAMMALIAN 87 SMALL NUCLEOLAR RNA AND ITS HOST GENE, by *J. A. Makarova, D. A. Kramerov** (Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991, Russia; *e-mail: kramerov@eimb.ru). The vast majority of C/D box small nucleolar RNAs (snoRNAs) direct site-specific methylation of ribose in rRNA. New mammalian snoRNA U87 belongs to the C/D family and is involved in ribose methylation of guanylic acid residue (position 3468) in 28S rRNA. A host gene harboring the U87 snoRNA gene in its intron was described for rodents and humans. In humans it has three neighbouring transcription start sites whereas in rat there are many scattered start sites. The transcript of this gene possesses some characteristic features of an mRNA (splicing, polyadenylation and cytoplasmic localization) but lacks long open reading frames and the degree of its conservation is much less than that of other host gene mRNAs. Such low conservation is discussed in the context of functional significance of exonic sequences of the gene.

18S, 5.8S и 28S рРНК синтезируются в виде единого предшественника, из которого вырезаются при его созревании. Одновременно с транскрипцией начинаются модификации нуклеотидов рРНК, самые распространенные из которых – метилирование остатков рибозы по 2'-О-положению и превращение уридина в псевдоуридин. В молекулах рРНК позвоночных насчитывается примерно по 100 модификаций каждого вида [1]. Для осуществления всех стадий процессинга рРНК необходимы малые ядрышковые РНК (мякРНК) размером 67–280 н. [2], которые локализованы в ядрышке и обнаружены у всех эукариот. Большинство известных мякРНК входит в состав одного из двух семейств – C/D или H/ACA, выделенных благодаря присутствию консервативных нуклеотидных последовательностей. Почти все РНК семейства C/D вовлечены в метилирование рибозы по 2'-OH-группе, а большинство мякРНК семейства H/ACA участвуют в псевдоуридилировании [3].

Принятые сокращения: мякРНК – малая ядрышковая РНК; кДНК – комплементарная ДНК; ПЦР – полимеразная цепная реакция; RACE (Rapid Amplification of cDNA Ends) – быстрая амплификация концов кДНК; НТО – нетранслируемая область; EST (Expressed Sequence Tag) – фрагмент экспрессирующихся нуклеотидных последовательностей.

*Эл. почта: kramerov@eimb.ru

Все РНК семейства C/D содержат так называемые элементы C (UGAUGA) и D (CUGA). Они расположены в 5'- и 3'-концевых участках и обычно сближены за счет комплементарных взаимодействий концов мякРНК. Образующаяся структура, включающая C- и D-элементы и концевой двухцепочечный участок, названа C/D-мотивом [4]. В центральной части молекул C/D-РНК присутствуют элементы C' и D', нуклеотидные последовательности которых обычно вырождены. Консервативные элементы в мякРНК расположены в следующем порядке: 5'–C–D'–C'–D–3'. В 5'-направлении от элементов D и/или D' находится участок длиной 10–21 н., полностью комплементарный соответствующему участку рРНК и способный образовывать с ним двойную спираль. 2'-O-метилированию подвергается нуклеотид рРНК, входящий в РНК/РНК-спираль и отделенный 4 н. от D- или D'-элементов. С элементами C и D связывается комплекс белков, осуществляющих метилирование [5].

Для генов мякРНК позвоночных характерно необычное расположение: они локализованы в интронах других генов, а их продукты созревают после сплайсинга первичного транскрипта гена-хозяина. Гены, интронные последовательности которых кодируют мякРНК, – это обычно гены белков трансляционного аппарата или ядрышковых белков.

Предполагается, что такая локализация генов мякРНК облегчает регуляцию работы трансляционного аппарата. В частности, кодирование мякРНК интронами генов рибосомных белков может обеспечивать координированный синтез рРНК и рибосомных белков [2].

В нашей работе описана новая мякРНК C/D-семейства – U87, вовлеченная в метилирование G3468 28S рРНК. Показано, что эта мякРНК кодируется интроном не известного ранее гена, названного U87HG (Host Gene – ген-хозяин). По ряду признаков (сплайсинг, полиаденилирование и цитоплазматическая локализация) РНК этого гена напоминает мРНК, однако в ней отсутствуют протяженные открытые рамки считывания. Ген-хозяин U87 мякРНК характеризуется значительно меньшим уровнем консервативности по сравнению с большинством других генов-хозяев мякРНК.

УСЛОВИЯ ЭКСПЕРИМЕНТА

Получение клонотек кДНК. α-РНК из печени крысы полиаденилировали, используя поли(A)-полимеразу *Escherichia coli*. кДНК синтезировали с помощью обратной транскриптазы (M-MLV, “Gibco-BRL”, США) и праймера XbaI$(T)_{15}$ (GTCGACTCTAGA$(T)_{15}$). После очистки кДНК от олигонуклеотида с помощью электрофореза в 4%-ном агарозном (NuSieve, “FMC”, Германия) геле, электропереноса на DEAE-мембрану и элюции на 3'-конце синтезировали олиго(dG), используя терминальную дезоксинуклеотидилтрансферазу (“Gibco-BRL”, США). кДНК амплифицировали посредством ПЦР с праймерами XbaI$(T)_{15}$ и EcoRI$(C)_{10}$ – CGGAATTCGTCCCCCCCCCC. Двухцепочечную кДНК обрабатывали рестриктазами XbaI и EcoRI, очищали электрофорезом в 4%-ном агарозном (NuSieve) геле и клонировали в плазмидном векторе (pGEM7Z), обработанном теми же рестриктазами. Для скрининга клонотеки использовали гибридизацию с меченной ^{32}P α-кДНК.

кДНК гена-хозяина (*U87HG*) получали с помощью 5'- и 3'-RACE. Поли(A)$^+$РНК из клеток HeLa человека и печени крысы подвергали обратной транскрипции с помощью ревертазы M-MLV и праймера XbaI$(T)_{15}$. Полученную кДНК очищали от олигонуклеотида, как описано выше. Перед 5'-RACE на 3'-конце кДНК синтезировали олиго(dG), используя терминальную дезоксинуклеотидилтрансферазу. После этого для амплификации кДНК человека проводили два раунда ПЦР (так называемая “hemi-nested PCR”): сначала с помощью праймеров EcoRI$(C)_{10}$ и GAAAACACTTTATTGTTCAG (pr 4), а затем EcoRI$(C)_{10}$ комбинировали с праймером T**CTCGAG**CTCCATAATTTATTGTGA (pr 5, сайт XhoI выделен жирным). В случае 5'-RACE кДНК крысы для ПЦР использовали праймеры EcoRI$(C)_{10}$ и T**CTCGAG**ACCCAAAACAACGAGGATT (pr 6, сайт XhoI выделен жирным). Для 3'-RACE кДНК человека проводили два раунда ПЦР: первый с использованием праймеров XbaI$(T)_{15}$ и TGGCATGCAGCATTTTGA (pr 1), во втором раунде pr 1 заменяли на T**GGATCC**AGGAAAATTGAAGACGTG (pr 2, сайт BamHI выделен). 3'-RACE кДНК крысы проводили в один раунд, используя праймеры XbaI$(T)_{15}$ и T**CTCGAG**TGCAACAAAACCTGTCT (pr 3, сайт XhoI выделен). ПЦР-продукты обрабатывали одной из пар рестриктаз (EcoRI и XhoI, XbaI и BamHI или XbaI и XhoI) в зависимости от наличия соответствующих сайтов в использованных праймерах, очищали электрофорезом в 4%-ном агарозном (NuSieve) геле и клонировали в плазмиде pBS(SK)$^+$, предварительно разрезанной одной из перечисленных пар рестриктаз. Скрининг клонотек проводили с использованием гибридизации с зондами, специфичными к *U87HG* человека или крысы (см. ниже).

Клонирование локуса, несущего ген U87 мякРНК. Геномные ДНК (100 мкг) крысы и мыши обрабатывали рестриктазами EcoRI и HindIII, разделяли электрофорезом в 0.8%-ном агарозном геле, элюировали фракции, обогащенные искомыми фрагментами ДНК, и клонировали в плазмиде pSL1190. В предварительном опыте с помощью блот-гибридизации по Саузерну показано, что EcoRI–HindIII-фрагменты ДНК крысы и мы-

ши, содержащие ген U87 мякРНК, имеют длину 3.0 и 3.5 т.п.н. соответственно. Проведя гибридизационный скрининг 50000 клонов геномной ДНК мыши и крысы, выявили соответственно один и два позитивных клона.

Двухцепочечные плазмидные матрицы секвенировали по Сэнгеру с использованием наборов на основе секвеназы (Sequenase version 2.0, "Amersham", Англия) или Taq-полимеразы ("Силекс", Россия).

Фракционирование клеток. Из клеток печени крысы согласно [6] выделяли цитоплазматическую, ядрышковую и нуклеоплазматическую фракции. Разделение цитоплазмы клеток HeLa на цитоскелетную и свободную фракции проводили, как описано [7]. Цитоплазматические фракции депротеинизировали смесью (1 : 1) хлороформа и фенола, насыщенного Трис-HCl, pH 8.5. Ядерные фракции для выделения РНК экстрагировали фенолом, насыщенным ацетатом натрия, pH 5.0. Суммарную клеточную РНК выделяли с помощью гуанидин-изотиоцианатного метода [8].

Анализ 5'-концов РНК проводили, используя метод удлинения праймера (primer extension) [9]. Праймеры метили по 5'-концу с помощью полинуклеотидкиназы и [γ-^{32}P]ATP. Для анализа U87 мякРНК праймер U87REV (GCTCAGTCTTAAGATTCT) отжигали с низкомолекулярной (<12S) РНК печени крысы. В случае РНК гена-хозяина суммарную клеточную РНК отжигали с одним из праймеров: ACCACACCGGCATGACTAAC (pr 7), GCCCAAGGACACGCCGGGTTC (pr 8) и GCCGGAGGACACGCCGGGTTC (pr 9), комплементарных первому экзону гена *U87HG* человека, мыши и крысы соответственно. кДНК синтезировали с помощью обратной транскриптазы M-MLV, фракционировали электрофоретически в 6%-ном полиакриламидном геле (ПААГ) с 7 М мочевиной и после высушивания гель экспонировали с рентгеновской пленкой.

Нозерн-блот-анализ. Низкомолекулярные РНК разделяли электрофорезом в 6%-ном ПААГ с 7 М мочевиной и переносили на мембрану Hybond-N с помощью полусухого электроблотинга. Высокомолекулярные РНК выявляли с помощью электрофореза в 1.5%-ном агарозном геле, содержащем 6%-ный формальдегид, с последующим капиллярным блотингом. Гибридизацию проводили, инкубируя фильтры в течение ночи при 42°C в смеси следующего состава: 4 × SSC, 1% додецилсульфата натрия, 50% формамида, 5 × раствор Денхардта, 0.1 мг/мл денатурированной ДНК сельди, меченый зонд. Фильтры отмывали в течение 40 мин при 42°C в растворе 0.1 × SSC – 0.1% додецилсульфата натрия и экспонировали с рентгеновской пленкой. Меченые зонды получали следующим образом. Сначала с помощью ПЦР амплифицировали выбранный участок геномной ДНК или кДНК, очищали продукт реакции посредством электрофореза в агарозном геле, а затем метили его с помощью ПЦР и [α-^{32}P]dATP (25 мкКи), используя те же праймеры и внося в реакционную смесь приблизительно 1% выделенного фрагмента ДНК. Зонд, специфичный к U87 РНК крысы и мыши, получали с использованием праймеров ATGTTTTTGCCGTTTACC (U87DIR) и U87REV; U87HG-специфичный зонд – с помощью праймеров pr 3 и GGCAAATGTAGTCCTGGC (второй экзон гена мыши), CATGTAGGTAGCTGCAGT и pr 6 (третий экзон гена крысы) или pr 2 и pr 4 (четвертый экзон гена человека).

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

U87 – новая мякРНК семейства C/D

При изучении клонотеки кДНК, созданной на основе РНК из α-РНП-частиц печени крысы [10], нами обнаружена новая малая РНК, названная U87. Ее точный размер (72 н.) и длина 5'-концевой области определены с помощью электрофореза высокого разрешения в ПААГ и метода удлинения праймера соответственно (данные не приведены).

Помимо этого, мы клонировали и секвенировали HindIII-EcoRI-фрагменты ДНК крысы и мыши длиной 3.0 и 3.5 т.п.н. соответственно, содержащие ген U87 РНК. С использованием базы данных GenBank обнаружено, что ген U87 РНК локализован на хромосоме 8 человека. Гомологи этого гена проявляют достаточно высокую степень сходства: 93% между генами грызунов и 83% между генами грызунов и человека, что характерно для мякРНК семейства C/D. С помощью нозерн-блот-гибридизации гомологи U87 РНК найдены у всех семи исследованных видов млекопитающих, относящихся к четырем отрядам: грызуны, приматы, хищные и насекомоядные (данные не приведены).

Нуклеотидная последовательность U87 РНК приведена на рис. 1. В ее составе обнаружены элементы, характерные для мякРНК семейства C/D: элементы C, C', D и D', причем C' отличается одним нуклеотидом (UGA**A**GA) от консенсусной последовательности C (UGA**U**GA). В мякРНК семейства C/D последовательность C' часто вырождена и представляет собой несовершенную копию элемента C [11]. Кроме того, в составе U87 РНК обнаружен участок длиной 11 н., полностью комплементарный нуклеотидам с 3466 по 3476 28S рРНК крысы. Таким образом, нуклеотидная последовательность U87 РНК обладает всеми признаками, характерными для мякРНК семейства C/D. Исходя из структуры U87 РНК, можно ожидать, что она должна направлять 2'-O-метилирование остатка рибозы G3468 в 28S рРНК. Действительно, известно [12], что в 28S рРНК этот нуклеотид метилирован. При комплементарном

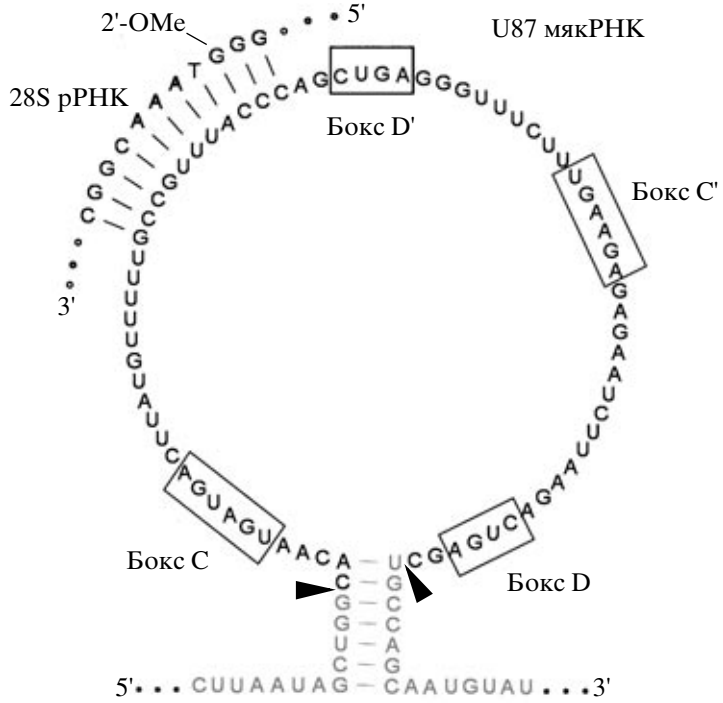

Рис. 1. U87 РНК крысы. Приведена нуклеотидная последовательность мякРНК вместе с прилежащими участками, входящими в состав первичного транскрипта гена *U87HG*. Концы мякРНК отмечены стрелками; рамками выделены боксы C, D, C' и D'; показано взаимодействие антисмыслового элемента U87 РНК с 28S рРНК; 2'-OMe-G – метилированный нуклеотид.

взаимодействии 28S рРНК и антисмыслового элемента U87 РНК G3468 оказывается отделенным 4 н. от элемента D'. Такое же взаимное расположение элементов D и/или D' и метилированного остатка наблюдается и в других мякРНК семейства C/D [5].

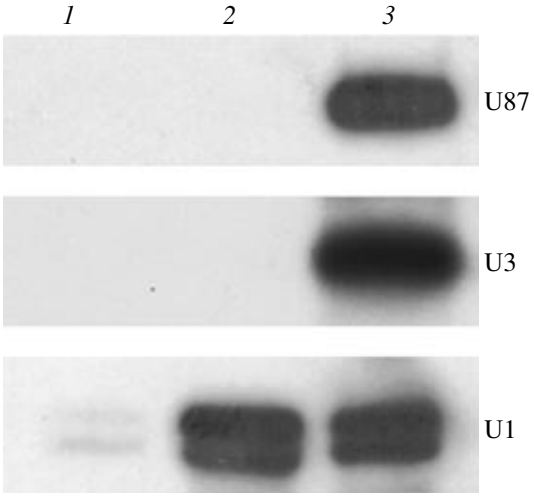

Рис. 2. Ядрышковая локализация U87 РНК. С помощью нозерн-блот-гибридизации U87 РНК, U3 мякРНК и малую ядерную РНК U1 выявляли в цитоплазматической (*1*), нуклеоплазматической (*2*) и ядрышковой (*3*) фракциях клеток HeLa. Ядрышковую U3 РНК и ядерную U1 РНК выявляли с целью контроля качества фракционирования клеток.

Ядрышковую локализацию мякРНК определяет C/D-мотив [4], в состав которого входят C- и D-элементы и концевой двухцепочечный участок длиной 3–7 п.н., образующийся за счет комплементарных взаимодействий концов мякРНК. Концы U87 РНК такой двуспиральный участок формировать не способны. Однако фланкирующие ген U87 РНК участки и первые 2 н. U87 РНК могут формировать двухцепочечную структуру длиной 7 п.н. (рис. 1). Таким образом, по крайней мере, в предшественнике U87 РНК может образоваться полноценный C/D-мотив, необходимый для правильной локализации этой РНК.

Чтобы определить внутриклеточную локализацию U87 РНК, из клеток печени крысы выделили фракции цитоплазматической, нуклеоплазматической и ядрышковой РНК и провели нозерн-блот-гибридизацию с зондами, комплементарными U87 РНК, ядерной РНК U1A и ядрышковой РНК U3. U87 РНК, подобно U3 РНК, выявляется в ядрышке, но не в цитоплазме или нуклеоплазме (рис. 2). Таким образом, U87 РНК – это ядрышковая РНК семейства C/D, которая, очевидно, направляет метилирование G3468 в 28S рРНК.

```
Крыса   : -CTTTCCgG-GC-------------------AGCGCGGTGGAGCGTaAG-CATGttGaGCtaCGTGGCGGgGAGCGGAGGCCGtGGGCtGCgGAACCCGG :78
Мышь    : -CTTTCtCG-GC-------------------AGCGCcGTaGAGCGcgAG-CAgaCgGtGCagCaTGGCGcacAGCGGAGGCCGCGGGCgGCcGAACCCGG :78
Человек : cCTTTCCCGcGCgaccggcgagggaggaagaAGCGCGaaGagcCGTtAGtCATGCcG-GtgtgGTGGCGGcG-GCGGAGaCtGCGGGCc----------- :87
          Экзон 1

Крыса   : CGTGTCCTccGGCTCTTCCAGGTGCAACAAAACCTGtctcTT---tTGAgGGTGCTTGAg-----------CAAATGTTTTtgAAACttgtCcTGTCcTG :164
Мышь    : CGTGTCCTTGGGCgtTTCCAGGTGCAACAAAACCTGctcaTT----TGAAGGTGCTTGAA-----------CAAATaTTTTCtAAACa---CtTGTCaTG :160
Человек : CGTagC--TGGGCTCTgCgAGGTGCAAgAAAgCCT-----TTgaggTGAAGGTGtaTGAAagtcatcataaCAgATGTTTTCcAAA--------aaCtTG :172
                                Экзон 2

Крыса   : GAaAGTGATTGGAACAAGCTGCCAGaACTACGgTTaCCAATAACAAGCATGTAGGTaGCTGcAG-TGAAcGTCCAACTTcGGGAGCTGTGATGTTTAA-- :261
Мышь    : GAGAGTGATTGaAACAAGCTGCCAGGACTACatTTGCCAATAACAAGCATGTAGGTgGCTGTAG-TGgATGTCCAACTTtGGGAGCTGTGATGTcTAAAA :259
Человек : tAGAagGtTgtGAAaAAaCTaCtAGGAtcACG-cgGCatgTAttgAGCATaTAGGTtGCTGTAGaTGAATGTtCtt-------AGCTGTcATGTTTAAAA :264
                                                        Экзон 3

Крыса   : --------tCcTCGTTGTTTTGGGTcCAGCAcACA---gTTTGGAAtGTAAAATGAAGACtT--------GGTCAAGAAACCATGAACgGAAGCAAATGG :342
Мышь    : Ac------cCTTtaTTGTTTTGGGTGCAGCtTACA---TaTTGGAAGGTAAAATGAAGgC-catgaacaaaGTCAAGAAgCCATaAACAGAAGCAAATGG :349
Человек : AtacttctgCTTCGTTaccTcaaGTGtgGCATgCAgcaTTTTGGAAGGaAAAtTGAAGACgT--------GtTCAAGAAAaCATGAACAGAAGCAAATGa :356
                                Экзон 4

Крыса   : TGAAAATTAGTATTTAATTTGgTATc-----------AATAtGTTATGTAGAAATAAACg-ATT----------TTAATTGCTaTGAGCAtTAgTTTCTT :420
Мышь    : TGAAAATTAaTATcTAATTTaATATT-----------AATAAGTTATaTATAAATAAACA-gTT------AtTcTTAATTGCTgTGgGCAgTAaTTTCTT :431
Человек : TGAAAATgAGcATTTtAcTTGATgTTgataacatcacAATAAaTTATGgAGAAAaAtACAtATTtggctaAcTtTTAATTGCT--GAaC---------aa :445

Крыса   : TAAAATGcgTgtcTTTAAAaCAcCATTCTCaTAGTAAAATAAACcTCaTCcTTCATG :477
Мышь    : TAAAATGT---------------CATTCTCgTAGTAAAATAAAC-TCgTCtTTCATG :472
Человек : TAAAgTGTtTtctTTTAAAtCA--AcTCT---------------------------- :472
```

Рис. 3. Выравнивание нуклеотидных последовательностей кДНК гена *U87HG* крысы, мыши и человека. Нуклеотиды, совпадающие в данной позиции хотя бы у двух видов, показаны заглавными буквами, жирным шрифтом. Границы экзонов отмечены вертикальными линиями. Сигналы полиаденилирования подчеркнуты. Стрелкой указано место интеграции ретропозона B2 в нуклеотидную последовательность гена крысы. (Чтобы облегчить выравнивание, последовательность B2 удалена.)

U87 РНК кодируется интроном не описанного ранее гена

Известно, что гены мякРНК позвоночных расположены в интронах других генов [2]. При изучении клонированных нами, содержащих ген U87 РНК HindIII-EcoRI-фрагментов ДНК крысы и мыши и соответствующего фрагмента ДНК человека не обнаружено экзонов известных белок-кодирующих генов. Однако для каждого из этих видов в базе данных GenBank (http://www.ncbi.nlm.nih.gov/dbEST) найдены EST, экспрессирующиеся с этого локуса, а ген U87 РНК оказывался в одном из интронов гена, которому соответствовали эти EST. С целью точного установления структуры этого гена, названного нами *U87HG* (Host Gene – ген-хозяин U87 РНК), клонированы 5'- и 3'-части его кДНК (отметим, что клонированные части кДНК перекрывались). Проведена ПЦР на кДНК *U87HG* с использованием праймеров к одному из экзонов и к гомополимерным концевым последовательностям кДНК. Полученные ПЦР-продукты клонировали и секвенировали. На рис. 3 представлены нуклеотидные последовательности кДНК *U87HG* крысы, мыши и человека. 5'-концевую область РНК *U87HG* дополнительно изучали с помощью метода удлинения праймера. Установлено, что транскрипция *U87HG* человека начинается с двух расположенных рядом цитидиловых остатков. Еще один минорный старт транскрипции соответствовал примыкающему к ним тимидиловому остатку (рис. 4*а*). В гене крысы, напротив, обнаружены множественные точки инициации транскрипции, рассеянные по участку длиной не менее 60 п.н. (рис. 4*б*), хотя обе кДНК крысы начинаются с одной и той же позиции (отмечена стрелкой на рис. 4*б*). Видно, что это одна из главных точек инициации транскрипции. Следует также отметить, что транскрипты, начинающиеся с этого старта, наилучшим образом выравниваются с транскриптами *U87HG* человека и, подобно транскриптам человека, начинаются с остатка C.

У грызунов и человека ген *U87HG* состоит из четырех экзонов, U87 РНК кодируется вторым интроном. Длина кДНК человека и мыши составляет 472 н., а крысы – 659 н. Больший размер кДНК крысы связан со вставкой ретропозона B2 в четвертый экзон. РНК *U87HG* полиаденилирована: на 3'-конце РНК человека расположены два сигнала полиаденилирования, причем используется преимущественно второй (в восьми из девяти клонов). В 3'-концевой области *U87HG* крысы

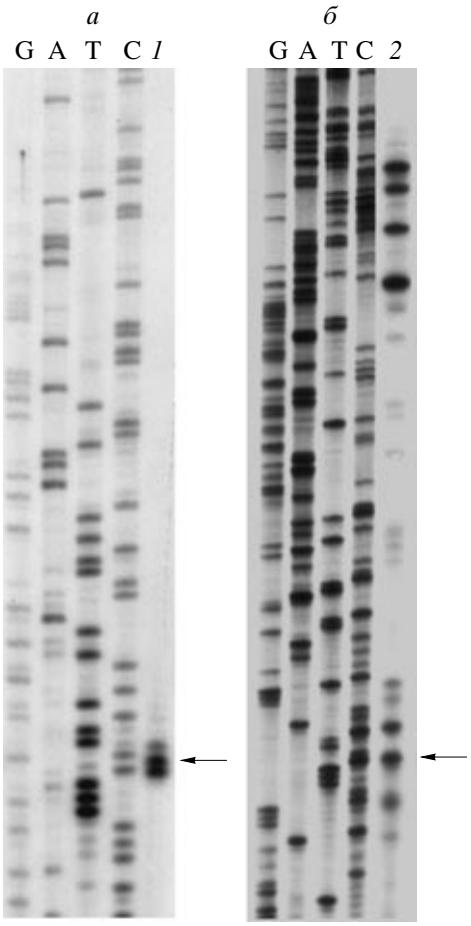

Рис. 4. Анализ 5'-концов РНК *U87HG* человека (*а*) и крысы (*б*). G, A, T, C – результаты секвенирования клонированного 5'-концевого района гена *U87HG*. *1* и *2* – кДНК, синтезированные ревертазой на РНК клеток человека и крысы соответственно, после ее отжига со специфическими праймерами. Стрелкой отмечена полоса, соответствующая 5'-концу РНК *U87HG*, идентифицированному посредством 5'-RACE.

также находятся два сигнала полиаденилирования, но используется в основном первый (в четырех из пяти клонов). В составе РНК *U87HG* мыши найдены два сигнала полиаденилирования, и оба используются – об этом свидетельствуют соответствующие EST.

С помощью метода удлинения праймера показано, что РНК *U87HG* одинаково представлена во всех исследованных тканях крысы (печень, почки, семенники и головной мозг; данные не приведены). Существуют коллекции EST, соответствующих неполным транскриптам гена *U87HG* мыши и человека (UniGene clusters Mm.28311 и Hs.372680). Эти транскрипты найдены во всех исследованных тканях мыши и человека. Таким образом, ген *U87HG* экспрессируется в широком спектре тканей и, следовательно, является геном “домашнего хозяйства”.

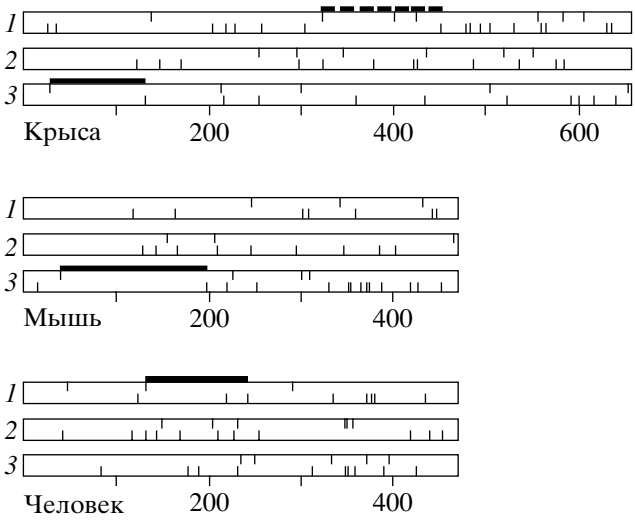

Рис. 5. Инициирующие и терминирующие кодоны для трех рамок считывания (*1–3*) РНК *U87HG* крысы, мыши и человека. Инициирующие и терминирующие кодоны отмечены короткими вертикальными линиями, идущими сверху и снизу, соответственно. Наиболее длинные открытые рамки считывания отмечены толстыми горизонтальными линиями. Открытая рамка считывания, входящая в состав ретропозона B2, показана пунктиром.

РНК U87HG не кодирует белок

Нуклеотидные последовательности РНК *U87HG* грызунов и человека имеют удивительно низкую степень сходства по сравнению с последовательностями белоккодирующих генов-хозяев других мякРНК: 68% между РНК мыши и человека, 72% между РНК крысы и человека и 80% между грызунами (делеции и вставки при подсчете не учитывали) (рис. 3). Кроме того, РНК *U87HG* содержит множественные стоп-кодоны во всех трех рамках считывания (рис. 5). Размер наиболее длинного пептида, кодируемого этими РНК, составляет 54, 44 и 30 аминокислотных остатков у мыши, крысы и человека соответственно. В *U87HG* крысы этот пептид кодируется коротким ретропозоном B2. Самая длинная аутентичная рамка РНК *U87HG* крысы соответствует 35 аминокислотным остаткам (рис. 5). Эти и все остальные пептиды, которые потенциально могли бы транслироваться с РНК *U87HG*, неконсервативны, и ни для одного из них не найдены гомологи в базах данных. Аналогичные результаты получены при попытках найти открытые рамки, начинающиеся с альтернативных кодонов, используемых при инициации трансляции у эукариот (GUG, UUG, CUG и ACG [13]). Поэтому крайне маловероятно, что РНК *U87HG* кодирует функционально значимый пептид или пептиды.

РНК U87HG локализована в цитоплазме

Чтобы установить внутриклеточную локализацию РНК *U87HG*, из клеток HeLa выделяли фракции цитоплазматической и ядерной РНК и проводили нозерн-блот-гибридизацию с зондами, комплементарными РНК *U87HG* и U3 РНК. Из рис. 6*а* видно, что РНК *U87HG* локализована почти исключительно в цитоплазме. Помимо этого, из клеток HeLa выделили фракции свободной и связанной с цитоскелетом РНК и провели нозерн-блот-гибридизацию с зондами, комплементарными РНК *U87HG* и мРНК β-актина, которая, как и большинство видов мРНК, связана с цитоскелетом. Видно (рис. 6*б*), что около 60% РНК *U87HG* ассоциированы с цитоскелетом.

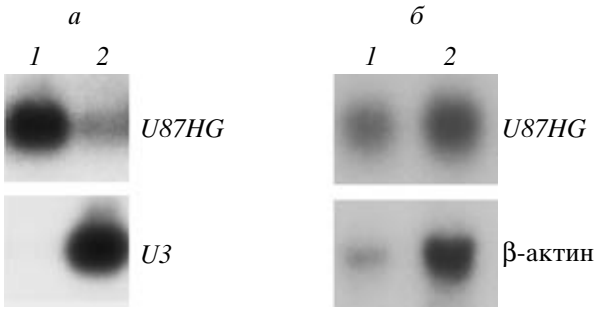

Рис. 6. Внутриклеточная локализация РНК *U87HG*. *а* – Выявление РНК *U87HG* и U3 мякРНК (контроль) в цитоплазматической (*1*) и ядерной (*2*) фракциях клеток HeLa. *б* – Выявление РНК *U87HG* и мРНК β-актина (контроль) во фракциях растворимой (*1*) и связанной с цитоскелетом (*2*) РНК.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Нами описана новая мякРНК семейства C/D, названная в соответствии с принятой номенклатурой U87. U87 мякРНК, очевидно, направляет метилирование G3468 в 28S рРНК. Гомологи этой РНК обнаружены у всех исследованных нами видов млекопитающих. U87 РНК консервативна: сходство нуклеотидных последовательностей составляет 93% у грызунов и 83% между грызунами и человеком. Консервативность мякРНК семейства C/D отчасти определяется чрезвычайной консервативностью необходимых для функционирования этих РНК нуклеотидных последовательностей, занимающих около половины всей длины мякРНК.

Интересно, что U87 РНК находится в ядрышке, несмотря на отсутствие в ее составе полноценного C/D-мотива, который образуется в результате комплементарного взаимодействия концов мякРНК и необходим для ядрышковой локализации. Однако прилегающие к гену U87 РНК нуклеотидные последовательности и первые 2 н. U87 РНК способны образовать двухцепочечный участок длиной 7 п.н. (рис. 1). Поэтому предшественник U87 РНК, очевидно, содержит полноценный C/D-мотив. Известно, что небольшая часть видов мякРНК C/D-семейства лишена концевых комплементарных нуклеотидных последовательностей. Вопрос о том, как такие РНК попадают в ядрышко, оставался неясным до тех пор, пока не был проведен анализ нуклеотидных последовательностей, фланкирующих их гены [14]. Оказалось, что нуклеотиды, непосредственно прилегающие к генам этих мякРНК, способны к комплементарным взаимодействиям и что формируемый предшественниками этих мякРНК полноценный C/D-мотив необходим для их ядрышковой локализации. Таким образом, хотя U87 РНК лишена концевого двухцепочечного участка, его, очевидно, формируют входящие в состав предшественника фланкирующие последовательности, благодаря чему U87 РНК может попадать в ядрышко.

Почти все мякРНК позвоночных кодируются интронами белоккодирующих генов и процессируются в ходе сплайсинга пре-мРНК гена-хозяина [2]. Мы описали ген-хозяин U87 РНК крысы, мыши и человека и назвали его *U87HG*. У всех трех видов этот ген состоит из четырех экзонов, его второй интрон содержит ген U87 РНК. РНК *U87HG* полиаденилирована и локализована в цитоплазме, причем более половины ее молекул ассоциировано с цитоскелетом (что свойственно большинству мРНК). Ген *U87HG* человека имеет два основных старта транскрипции, соответствующих двум соседним цитидиловым остаткам, и примыкающий к ним вплотную еще один более слабый старт, соответствующий тимидиловому остатку (рис. 4*а*). Неожиданно оказалось, что ген *U87HG* крысы имеет множественные старты транскрипции, рассеянные по участку длиной не менее 60 п.н. (рис. 4*б*). Хотя существование нескольких стартов транскрипции – это скорее правило, чем исключение, особенно для генов домашнего хозяйства [15, 16], *U87HG*, насколько нам известно, служит первым примером гена с тремя расположенными рядом стартами у одного вида млекопитающих и множественными стартами – у другого вида.

РНК *U87HG* присутствует во многих тканях, о чем свидетельствуют данные, полученные с помощью метода удлинения праймера и с использованием базы данных EST. Таким образом, *U87HG* относится к так называемым генам домашнего хозяйства, что соответствует ожиданиям, поскольку он кодирует по крайней мере один важный для функционирования клетки продукт, а именно, U87 мякРНК, участвующую в биогенезе рибосом.

Поиск консервативных вторичных структур (http://www.bioinfo.rpi.edu/applications/mfold) не дал

положительных результатов, и поэтому вопрос о существовании таких структур в РНК *U87HG* остается открытым.

Интересно, что РНК *U87HG* содержит много коротких открытых рамок считывания и не содержит длинных. Ранее считалось, что полицистронные РНК не характерны для эукариот, однако, оказалось, что в 5'-концевой области примерно 40% клеточных мРНК локализованы одна или несколько консервативных коротких открытых рамок, которые существенно снижают эффективность трансляции основной рамки, поскольку рибосома, сканируя мРНК, транслирует сначала первую короткую рамку [17, 18]. Несколько пептидов длиной до 100 аминокислотных остатков, кодируемых этими рамками, обнаружены в клетке, а выявлению остальных, судя по всему, препятствует отсутствие адекватных экспериментальных методов [19]. Тем не менее, даже когда 5'-концевая часть мРНК содержит несколько коротких рамок, в ней всегда есть длинная рамка, кодирующая основной белковый продукт данного гена. В РНК *U87HG* такой рамки нет. Более того, короткие пептиды, кодируемые этой РНК, не консервативны, и их гомологи отсутствуют в базах данных. Поэтому можно предполагать, что РНК *U87HG* не кодирует функциональных белковых продуктов.

Известно пять генов, которые по ряду свойств напоминают *U87HG*: *UHG*, *U17HG*, *U19HG*, *gas5* и *U50HG*, причем относительно подробно изучены только первые четыре [20–24]. Интроны этих генов кодируют ядрышковые РНК, а транскрипты имеют длину 400–2000 н., не способны кодировать белки, полиаденилированы и локализованы либо в ядре (*U19HG*), либо в цитоплазме (*UHG*, *U17HG* и *gas5*). Только два из четырех изученных генов представляют собой самостоятельные транскрипционные единицы, тогда как часть транскриптов генов *U17HG* и *U19HG* входит в состав мРНК генов, кодирующих белки RCC1 и матрин, соответственно. Также нужно отметить, что гомологи этих генов у человека и мыши существенно отличаются по длине (в 1.5–5 раз) и имеют низкий уровень сходства (около 50%, делеции и вставки при подсчете не учитывали). Такой уровень сходства характерен для интронов и указывает на то, что единственным функциональным продуктом этих генов могут быть мякРНК, кодируемые интронами, а транскрипты генов-хозяев не несут функциональной нагрузки [20, 21, 23].

В отличие от упомянутых генов, чьи гомологи у грызунов и человека из-за множественных делеций и вставок, а также из-за низкого сходства нуклеотидных последовательностей, выравниваются плохо, РНК *U87HG* выравнивается гораздо лучше. Длины РНК *U87HG* человека и грызунов близки или идентичны: РНК человека и мыши имеют одинаковый размер (472 н.), РНК крысы длиннее из-за единственной вставки короткого ретропозона B2, а не из-за множественных делеций и вставок, как у других некодирующих генов-хозяев. Хотя вторжения мобильных элементов в экзонные последовательности обычно вредны, известно немало примеров ретропозонов, локализованных в 3'-нетранслируемых областях (3'-НТО) генов. Иногда мобильные элементы в экзонах даже функциональны, например, сигнал полиаденилирования ретропозона B2, расположенного в 3'-НТО гена цитохрома P450, используется при транскрипции этого гена [25, 26].

Сходство нуклеотидных последовательностей РНК *U87HG* человека и мыши составляет 68%; человека и крысы – 72%; мыши и крысы – 80% (делеции и вставки при подсчете не учитывали). Такой же уровень сходства у этих видов имеют и НТО [27]. Как и в НТО, участки высокого сходства в РНК *U87HG*, соответствующие сайтам связывания белков в НТО, чередуются с участками более низкого сходства. Видимо, в РНК *U87HG* участки высокого сходства также несут функциональную нагрузку. Следует отметить, что хотя НТО необходимы для нормального синтеза кодируемого белка, у многих генов они обладают заметно более низким уровнем сходства, чем гомологи РНК *U87HG* [27]. Уровень сходства гомологов РНК *U87HG* выше, чем у селективно нейтральных последовательностей [28, 29], и давление отбора на ген *U87HG*, видимо, сопоставимо с уровнем отбора для НТО. Таким образом, нуклеотидная последовательность РНК *U87HG*, в отличие от нуклеотидных последовательностей остальных генов-хозяев, не является селективно нейтральной, что предполагает наличие у *U87HG* других функций помимо продуцирования U87 мякРНК. Впрочем, следует отметить, что транскрипты остальных генов-хозяев также содержат короткие консервативные участки, хотя их намного меньше, чем в РНК *U87HG*.

Список некодирующих РНК постоянно расширяется, причем установлены функции далеко не всех из них. Некоторые некодирующие РНК с уже известными функциями достаточно малоконсервативны так же, как и РНК *U87HG* [30, 31]. Вероятно, РНК *U87HG*, а, возможно, и транскрипты остальных некодирующих генов-хозяев, относятся именно к такой группе низкоконсервативных РНК.

Авторы благодарят И.К. Гоголевскую (ИМБ РАН) за ценный вклад в экспериментальную работу на начальном этапе, В.А. Куличкову и И.М. Константинову (ИЦ РАН) за предоставление препаратов α-РНК печени крысы, а также Л.Н. Гаузе (ИБР РАН) за плодотворное обсуждение результатов.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (02-04-49772, МАС 03-04-06774) и Государственной программы “Приоритетные направления генетики”.

СПИСОК ЛИТЕРАТУРЫ

1. Weinstein L.B., Steitz J.A. 1999. Guided tours: from precursor snoRNA to functional snoRNP. *Current Opin. Cell. Biol.* **11**, 378–384.
2. Maxwell E.S., Fournier M.J. 1995. The small nucleolar RNAs. *Annu. Rev. Biochem.* **64**, 897–934.
3. Tollervey D., Kiss T. 1997. Function and synthesis of small nucleolar RNAs. *Current Opin. Cell Biol.* **9**, 337–342.
4. Samarsky D.A., Fournier M.J., Singer R.H., Bertrand E. 1998. The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J.* **17**, 3747–3757.
5. Bachellerie J.P., Cavaille J. 1997. Guiding ribose methylation of rRNA. *Trends Biochem. Sci.* **22**, 257–261.
6. Muramatsu M., Hayashi Y., Onishi T., Sakai M., Takai K. 1974. Rapid isolation of nucleoli from detergent purified nuclei of various tumor and tissue culture cells. *Exp. Cell. Res.* **88**, 245–251.
7. Zambetti G., Wilming L., Fey E.G., Penman S., Stein J., Stein G. 1990. Differential association of membrane-bound and non-membrane-bound polysomes with the cytoskeleton. *Exp. Cell. Res.* **191**, 246–255.
8. Chomczynski P., Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.
9. Kingston R.E. 1992. Primer extension. In: *Short Protocols in Molecular Biology*. Ed. Ausubel F.M. N.Y.: Wiley&Sons, 21–22.
10. Konstantinova I.M., Turoverova L.V., Petukhova O.A., Vorob’ev V.I. 1984. Cortisone-induced small RNP tightly bound to chromatin. *FEBS Lett.* **177**, 241–245.
11. Kiss-Laszlo Z., Henry Y., Kiss T. 1998. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* **17**, 797–807.
12. Maden B.E. 1988. Locations of methyl groups in 28 S rRNA of *Xenopus laevis* and man. Clustering in the conserved core of molecule. *J. Mol. Biol.* **201**, 289–314.
13. Drabkin H.J., RajBhandary U.L. 1998. Initiation of protein synthesis in mammalian cells with codons other than AUG and amino acids other than methionine. *Mol. Cell. Biol.* **18**, 5140–5147.
14. Darzacq X., Kiss T. 2000. Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5',3'-terminal stem structure. *Mol. Cell. Biol.* **20**, 4522–4531.
15. Suzuki Y., Taira H., Tsunoda T., Mizushima-Sugano J., Sese J., Hata H., Ota T., Isogai T., Tanaka T., Morishita S., Okubo K., Sakaki Y., Nakamura Y., Suyama A., Sugano S. 2001. Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. *EMBO Rep.* **2**, 388–393.
16. Antequera F. 2003. Structure, function and evolution of CpG island promoters. *Cell. Mol. Life Sci.* **60**, 1647–1658.
17. Morris D.R., Geballe A.P. 2000. Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.* **20**, 8635–8642.
18. Meijer H.A., Thomas A.A. 2002. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem. J.* **367**, 1–11.
19. Oyama M., Itagaki C., Hata H., Suzuki Y., Izumi T., Natsume T., Isobe T., Sugano S. 2004. Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. *Genome Res.* **14**, 2048–2052.
20. Tycowski K.T., Shu M.D., Steitz J.A. 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature*. **379**, 464–466.
21. Pelczar P., Filipowicz W. 1998. The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol. Cell. Biol.* **18**, 4509–4518.
22. Bortolin M.L., Kiss T. 1998. Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. *RNA*, **4**, 445–454.
23. Smith C.M., Steitz J.A. 1998. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol. Cell. Biol.* **18**, 6897–6909.
24. Tanaka R., Satoh H., Moriyama M., Satoh K., Morishita Y., Yoshida S., Watanabe T., Nakamura Y., Mori S. 2000. Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human B-cell lymphoma. *Genes Cells*. **5**, 277–287.
25. Ryskov A.P., Ivanov P.L., Kramerov D.A., Georgiev G.P. 1983. Mouse ubiquitous B2 repeat in polysomal and cytoplasmic poly(A)$^+$RNAs: uniderectional orientation and 3'-end localization. *Nucleic Acids Res.* **11**, 6541–6558.
26. Kalb V.F., Glasser S., King D., Lingrel J.B. 1983. A cluster of repetitive elements within a 700 base pair region in the mouse genome. *Nucleic Acids Res.* **11**, 2177–2184.
27. Makalowski W., Boguski M.S. 1998. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2.820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci. USA*. **95**, 9407–9412.
28. Duret L., Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* **17**, 68–74.
29. Cooper G.M., Brudno M., Stone E.A., Dubchak I., Batzoglou S., Sidow A. 2004. Characterization of evolutionary rates and constraints in three Mammalian genomes. *Genome Res.* **14**, 539–548.
30. Rohrig H., Schmidt J., Miklashevichs E., Schell J., John M. 2002. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. USA*, **99**, 1915–1920.
31. Tam W. 2001. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a non-coding RNA. *Gene*. **274**, 157–167.