The evolution and expression of the snaR family of small non-coding RNAs

Andrew M. Parrott[1], Michael Tsai[1], Priyanka Batchu[2], Karen Ryan[1], Harvey L. Ozer[2,3], Bin Tian[1,2] and Michael B. Mathews[1,2,*]

[1]Department of Biochemistry and Molecular Biology, New Jersey Medical School, UMDNJ, Newark, [2]Graduate School of Biomedical Sciences, UMDNJ, Newark and [3]Department of Microbiology and Molecular Genetics, New Jersey Medical School/University Hospital Cancer Center, UMDNJ - New Jersey Medical School, New Jersey, USA

ABSTRACT

We recently identified the snaR family of small non-coding RNAs that associate *in vivo* with the nuclear factor 90 (NF90/ILF3) protein. The major human species, snaR-A, is an RNA polymerase III transcript with restricted tissue distribution and orthologs in chimpanzee but not rhesus macaque or mouse. We report their expression in human tissues and their evolution in primates. *snaR* genes are exclusively in African Great Apes and some are unique to humans. Two novel families of snaR-related genetic elements were found in primates: *CAS* (catarrhine ancestor of snaR), limited to Old World Monkeys and apes; and *ASR* (Alu/snaR-related), present in all monkeys and apes. *ASR* and *CAS* appear to have spread by retrotransposition, whereas most *snaR* genes have spread by segmental duplication. snaR-A and snaR-G2 are differentially expressed in discrete regions of the human brain and other tissues, notably including testis. snaR-A is up-regulated in transformed and immortalized human cells, and is stably bound to ribosomes in HeLa cells. We infer that *snaR* evolved from the left monomer of the primate-specific *Alu* SINE family via *ASR* and *CAS* in conjunction with major primate speciation events, and suggest that snaRs participate in tissue- and species-specific regulation of cell growth and translation.

INTRODUCTION

High-throughput sequencing information and tiling array data have revealed that the transcribed portions of eukaryotic genomes are larger and more complex than previously thought, implying complex multilayered levels of RNA regulation (1,2). The majority of transcripts do not encode protein and their proportion increases with organismal complexity (3). Although many such RNAs have been studied in detail and have ascribed functions, limited information is available for the preponderance of non-coding transcripts (4,5).

We discovered two members of a novel family of small RNAs that bind to nuclear factor 90 (NF90), a widely distributed mammalian double-stranded RNA-binding protein encoded by the *ILF3* gene (6–9). These small NF90-associated RNA (snaR) species were called snaR-A and snaR-B (8). They are highly structured, non-coding RNAs of ∼117 nt, terminating in an oligo-(A) tract followed by an oligo-(U) tract. snaR-A was characterized as a relatively unstable RNA (half-life ∼15 min in HeLa cells) transcribed by RNA polymerase III (Pol III) from an intragenic promoter. snaR-A is abundant in many immortal cell lines and in testis, and is found at lower levels in other parts of the body (8).

Bioinformatic searches led to the discovery of additional snaR family genes in the human genome: *snaR-C, -D, -E, -F, -G1, -G2, -H, -I, -12* and *-21* (8). Humans have ∼30 snaR genes, all but four of which are clustered in or near two tandem arrays on the q-arm of chromosome 19 (Figure 1). Exceptionally, *snaR-H, -I, -12 and -21* are on chromosomes 2, 3, 12 and 21, respectively. The chromosome 19 clusters contain *snaR-A*, which accounts for about half of the *snaR* genes, as well as *snaR-B, -C* and *-D* (Figure 1). *snaR-E* and *-F* flank the clusters, and *snaR-G1* and *snaR-G2* lie between the clusters adjacent to protein coding genes (Figure 1). Specifically, *snaR-G1* and *-G2* lie within the proximal promoters of two human chorionic gonadotropin β-subunit genes, *hCGβ1* and
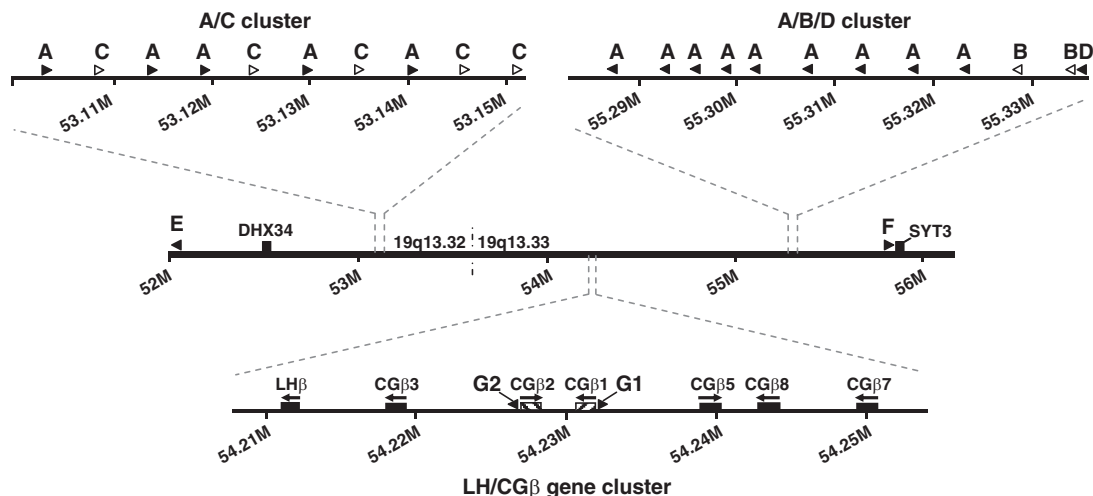
Figure 1. snaR gene clusters. Shown is the region of human chromosome 19q13.32-33 containing *snaR* genes (middle line). Expansions detail the *snaR* A/C and A/B/D clusters (top line) and the *LH/CGβ* cluster containing *snaR-G1* and *snaR-G2* (bottom line). The locations and direction of transcription of *snaR* genes are denoted by arrow heads. *DHX34* and *SYT3* genes and *LH/CGβ* cluster genes are denoted by closed boxes and their direction of transcription by arrows. *CGβ1* and *CGβ2* genes are denoted by hatched boxes. Distances are in megabases (M). Modified from (32).

hCGβ2, respectively (8). *hCGβ1* and *hCGβ2* represent a recent expansion of the luteinizing hormone (LH)/chorionic gonadotropin (CG) hormone β-subunit gene cluster and are unique to African Great Apes (A.M. Parrott *et al.*, submitted for publication) (10,11).

snaR genes were found in the chimpanzee genome but not in the genomes of mouse or rhesus macaque (8). Here we report the species distribution and ancestry of these genes, and their expression in human tissues, cell lines and subcellular compartments. These data lead us to propose a mechanism for their genomic spread and their function. Our data indicate that *snaR* evolved through a series of internal deletions and expansions from the left monomer of *Alu*, the most populous primate-specific short interspersed element (SINE) gene. This step-wise evolution progressed through two hitherto unreported ancestral families of non-coding RNAs, ASR (Alu/snaR-related) and *CAS* (Catarrhine ancestor of snaR). These three phases of snaR molecular evolution coincide with major primate speciation events: *ASR* arose in monkeys, *CAS* in Old World Monkeys and *snaR* in the African Great Apes. The molecular rearrangements resulting in the evolution of *snaR* and its ancestors appear to have been fostered by a single parental locus now encompassing *snaR-F* on chromosome 19. In contrast to *ASR* and *CAS*, which apparently spread through primate genomes by retrotransposition, we infer that *snaR* genes spread by segmental duplication. Recombination of the parental segment has led to species-specific amplification or loss of *snaR* subsets, and accelerated evolution of a subset of human-specific *snaR*. Examination of human tissues reveals discrete expression in testis and differential expression of snaR subsets in regions of the brain. snaR-A transcription is dysregulated in cell lines and its expression is induced upon transformation and immortalization. Furthermore, the stable association of snaRs with ribosomes, and parallels with

ancestrally related BC200 RNA, indicate a recently evolved function in translation.

MATERIALS AND METHODS

Sequence sources

Genomic sequence of the following species were BLAT searched (12) in the UCSC genome browser database (13): human (*Homo sapiens*; UCSC hg18), common chimpanzee (*Pan troglodytes*; UCSC panTro2), Sumatran orangutan (*Pongo abelii*; UCSC ponAbe2), rhesus macaque (*Macaca mulatta*; UCSC rheMac2) and marmoset (*Callithrix jacchus*; UCSC calJac1). Trace genomic sequences of bonobo (*Pan paniscus*), Western gorilla (*Gorilla gorilla*), white-cheeked crested gibbon (*Nomascus leucogenys*), anubis baboon (*Papio anubis*), hamadryas baboon (*Papio hamadryas*), grivet (*Chlorocebus aethiops*), mantled guereza (*Colobus guereza*), Nancy Ma's night monkey (*Aotus nancymaae*), black-capped squirrel monkey (*Saimiri boliviensis*), Bolivian squirrel monkey (*Saimiri boliviensis boliviensis*), Geoffroy's spider monkey (*Ateles geoffroyi*), red-bellied titi (*Callicebus moloch*), Philippine tarsier (*Tarsius syrichta*), ring-tailed lemur (*Lemur catta*) and small-eared galago (*Otolemur garnetti*) were retrieved from the NCBI Trace Archive. ESTs were retrieved from NCBI. NCBI database searches employed MegaBLAST (14) or discontiguous MegaBLAST.

Bioinformatics

Phylogenetic trees were constructed in TreeView (v6.6) (15) using the neighbor-joining approach from multiple sequence alignments generated by Clustal-X (16). *ASR*, *CAS* and *snaR* were searched for repeat sequences in RepeatMasker (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0*.1996–2004, http://www

.repeatmasker.org) and in Retrosearch, a database of ORF annotated human endogenous retroviruses (HERV) (17).

Genomic PCR

Bonobo (PR00251), common chimpanzee (PR00643), lowland gorilla (PR00573) and Sumatran orangutan (PR00253) genomic DNA was purchased from the IPBIR (Camden, NJ, USA). Genomic DNA from human 293 cells was prepared as described earlier (18). Primers used to amplify common sequence flanking *snaR* genes: forward 5′-CAGTTCCCCTTCCTCCCT-3′, reverse 5′-TCTTCCTATCAAGGCGTCC-3′. Fresh PCR product was ligated into pCRII-TOPO vector (Invitrogen), cloned and sequenced in forward and reverse directions. Orthology of human and chimpanzee amplicons was defined on the basis of identity to chimpanzee and human genomes. Bonobo amplicons were identical to those of the sister *Pan* species. Gorilla amplicons were classified as snaR-A, -G1 and -G2. Orthology of the gorilla snaR-A amplicon was established by homology with human snaR-A. Identities of snaR-G subsets in human, chimpanzee and gorilla were established by genomic PCR analysis of large amplicons, with primer sequence inside adjacent *CGβ* genes (A.M. Parrott *et al.*, submitted for publication).

Transformed and immortal human cell lines

Cell lines were obtained and maintained as described elsewhere (19–21). Human diploid fibroblasts derived from fetal bone marrow stroma (HS74) or foreskin (HSF43) were propagated in DME: F10 and 10% fetal calf serum in 7.5% CO_2. HS74 was transformed by distinct origin-defective constructs of SV40. These include cl39T (a subgenomic fragment of the SV40 genome); SV.RNS/HF (intact SV40 genome linked to pRNS-1) (termed HF) and its immortal derivative (HF-1); and SVtsA/HF-A (intact SVtsA58 genome encoding temperature sensitive large T antigen) (termed HF-A) and its immortal derivative AR5. All cell lines were propagated at 37°C except for HF-A and AR5 at 35°C. HSF43 and CT10-2A were provided by (22). Total RNA was isolated using Trizol (Invitrogen) under standard conditions (23).

cDNA preparation

Adrenal gland (Adr), fetal brain (Fbra), fetal kidney (Fkid) and mammary gland (Mam) human tissue total RNA was purchased from Clontech, while adipose (Adi), bladder (Bla), brain (Bra), cervix (Cer), colon (Col), esophagus (Eso), heart (Hrt), kidney (Kid), liver (Liv), lung (Lun), post-menopausal ovary (Ova), term placenta (Pla), prostate (Prs), skeletal muscle (SM), small intestine (Sint), spleen (Spl), testis (Tes), thymus (Thm), thyroid (Thr) and trachea (Tra) human tissue total RNA were purchased from Ambion. Ambion RNA was a composite of three different individuals. Frozen human amgydala (Amg), anterior hippocampus (Hip), caudate/accumbens/putamen (CAP), globus pallidus (GPa), hypothalamus (Hyp), temporal pole (TPl),

thalamus at the level of the centrum medianum (Thl) and visual cortex (VCx) tissue was provided by the Harvard Brain Tissue Resource Center (Belmont, MA, USA). Frozen whole pituitary glands (Pit 3 and Pit 4) were purchased from the National Hormone & Peptide Program (Torrance, CA, USA). Two antipodal slices of frozen tissue (∼40 mg each) were crushed in 500 μl Trizol (Invitrogen) then sonicated in round-bottom eppendorfs for $4 × 30$ s, with 10 s relaxations, in a Cup Horn (Misonix) coupled to a circulating ice bath. One volume of Trizol was added and total RNA extracted in accordance with instruction. The quality of extracted RNA was confirmed by examination of rRNAs (Supplementary Figure S1A). RNA (2 μg) was subjected to DNase I treatment (RNase-free grade, Boehringer-Mannheim), followed by annealing of 'lock-dock' oligo(dT)$_{18}$ primer (24) and RT–PCR with Superscript RT II in accordance with instruction (Invitrogen).

Real-time qPCR

Forward (F) and reverse (R) PCR primer and TaqMan probe (P) specificity is illustrated in Supplementary Figure S1B; their sequences are: snaR-A: F 5′-GGAGCCATTG TGGCTCAG-3′, R 5′-ACCCATGTGGACCAGGCT-3′, P 5′-CCTCGAACTCGTGCCCTGGAAC-3′; snaR-B: F 5′-CATTGTGGCTCCGGCCG-3′, R 5′-AGGTTGGCC TCGAACTCGT-3′, P 5′-CTCGAACCCTCGCCTACC CTGA-3′; snaR-C: F 5′-GTGGCTCCGGCCGGTTG -3′, R 5′-AACCCATGTGGACCAGGTTG-3′, P 5′-CCC TCGAACCCTCGCCTCTCT-3′; snaR-G1: F 5′-GGAG CCATTGTGGCTCCG-3′, R 5′-CCTACCAATGGGGC CCAG-3′; snaR-G2: F 5′-CGAGCTATTGTGGCTCCG A-3′, R 5′-TCTCAACCAATGTGGACCCG-3′, P 5′-TT GGCCTCGAACTCGTACCCTCGAA-3′. Real-time qPCR was performed in 25 μl duplicate reactions on 1/20–3/20th of the reverse transcription reaction using 0.3 μM primers and 0.1 μM TaqMan probe and 12.5 μl 2 × Taqman Universal PCR Master Mix (Applied Biosystems). qPCR was performed on an ABI 7500 Real Time PCR System (Applied Biosystems). qPCR product was normalized against that of cyclophilin A mRNA: F 5′-ATGGTCAACCCCACCGTGT-3′, R 5′-T CTTTGGGACCTTGTCTGCAA-3′ and P 5′-AGCTCA AAGGAGACGCGGCCCAA-3′. The identity of qPCR products was confirmed by agarose gel analysis and by sequencing.

Sucrose gradients

HeLa S3 cells (∼$1 × 10^7$) were grown to ∼80% confluency in DMEM (Mediatech, Inc.) supplemented with 5% FBS (HyClone). Cells were treated with 100 μg/ml cycloheximide (ChX; Sigma) for 5 min at 37°C. Cells were harvested by scraping, pelleted at 500 g for 10 min and washed in 500 μl PBS +100 μg/ml ChX. Cells were lyzed in 200 μl Triton buffer (150 mM KCl, 15 mM Tris–HCl, pH 7.4, 50 mM NaF, 1 mM EGTA, 15 mM $MgCl_2$, 0.2 mM Na_3VO_4, 10% glycerol, 1% Triton X-100, 0.2 mM PMSF, 1 mM DTT, 0.5 U/μl RNasin, 100 μg/ml ChX) on ice for 10 min. Nuclei were pelleted twice at 13 000 rpm for 10 min. Supernatant (200 μl) was

layered onto a 10 ml 10–50% sucrose gradient prepared in a gradient maker (200 mM KCl, 15 mM Tris–HCl, pH 7.4, 15 mM MgCl$_2$) and centrifuged at 39 226 rpm in an SW41-Ti rotor for 1.5 h. Fractions (18, 0.5 ml and a final 1 ml) were collected manually by pipette aspiration. Aliquots (100 μl) were set aside for protein analysis. RNA was isolated by Trizol extraction and resuspended in TE buffer (40 μl). Alternatively, cells were treated with 250 μg/ml puromycin for 20 min at 37°C, before harvesting and lysis in Triton buffer modified to include 300 mM KCl and 100 μg/ml puromycin, omitting ChX.

Blots

Total RNA from 293 cells was 3′-radiolabeled as described (25). Unlabeled and labeled total RNA was resolved in 5 or 10% (w/v) polyacrylamide/7 M urea gels, northern blotted and probed with 5′-radiolabeled probes as described (8). Extensive testing established the specificity of the probes, and no cross-reaction with additional species was detected. Antisense oligonucleotide probes used: snaR-A 5′-GACCCATGTGGACCAGGCTGGCC TCGAACT-3′, snaR-B 5′-CCTCGCCTACCCTGAGAG TCCGAGGGCCCG-3′ (C to U edited site is underlined), snaR-C 5′-GAGCCCTCGAACCCTCGCCTCTCTGAG GGT-3′, 5S rRNA 5′-TCAGACGAGATCGGGCGC GTT-3′, BC200 (BC206) 5′-AAATAAGCGTAACTTCC CTCAAAGCAACAA-3′ (26). Protein samples were boiled in Laemmli sample buffer then fractionated in 7.5% (w/v) polyacrylamide/SDS gel alongside pre-stained molecular mass marker (Precision Plus; BioRad). Protein was transferred to 0.45 μm nitrocellulose membrane by electroblotting and probed overnight at 4°C with anti-PARP (Santa Cruz Biotech.) or anti-α-tubulin antibody (Sigma).

RESULTS

snaR genes are unique to African Great Apes

The *snaR* genes were identified in two members of the Hominini tribe (human and chimpanzee) but not in rhesus macaque or mouse (8). To establish the distribution of these genes more broadly within the Great Ape family (Hominidae), we conducted PCR analysis on genomic DNA of human, bonobo, common chimpanzee, Western gorilla and Sumatran orangutan, using primers against flanking sequences that are highly conserved for most human and chimpanzee *snaR* genes (8). PCR products of the expected size (∼320 bp in humans) were generated from the genomes of all five hominids (Figure 2A). Clones were sequenced (Supplementary Data S1) and classified into the various *snaR* subsets (Table 1).

Alignment of representative hominid clones revealed strong identity within the African Great Apes (subfamily Homininae: human, chimpanzee, bonobo and gorilla), but differences between these and the orangutan clones which were most marked in the snaR coding region (Figure 2B). Two subsets of clones were obtained from orangutan: Pa19, which aligns almost perfectly to the genome (locus 52170854 Mb on chromosome 19); and Pa19i, which has six changes mainly within the snaR homology sequence (Supplementary Figure S2). The orangutan clones resembled those of the African Great Apes closely in the regions surrounding their *snaR* coding sequences, apart from substitutions immediately upstream of the gene and a 27-bp expansion containing a 10-bp duplication further upstream (Figure 2B). Within the region corresponding to snaR transcripts, however, the orangutan sequences differed considerably from those of the other Great Apes (33% divergence over 127 nt). Searches of
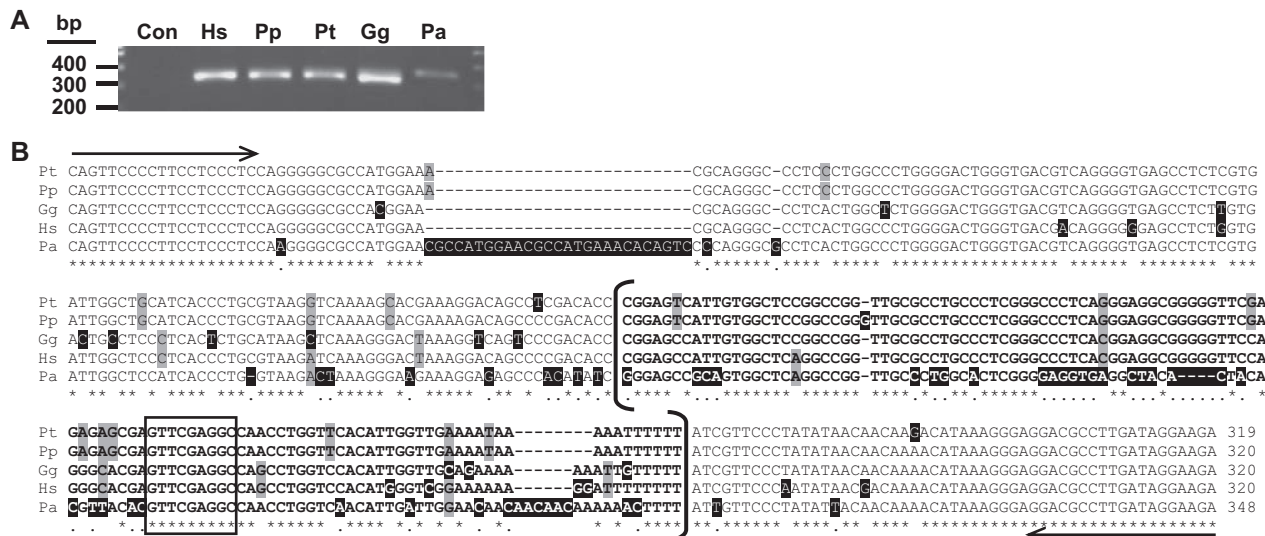
Figure 2. Species distribution of *snaR* genes. (**A**) Ethidium-stained agarose gel of bonobo (Pp), chimpanzee (Pt), gorilla (Gg), human 293 cell (Hs) and orangutan (Pa) genomic DNA after PCR amplification with primers specific for snaR flanking sequence. Note that the orangutan PCR fragment has slower migration because it is longer, and formed a weaker band because of a mismatch between the forward primer and the genomic sequence. (**B**) Clustal-W alignment of selected PCR clones: human (H3iii), gorilla (G1iii), bonobo (B1), chimpanzee (C1) and orangutan (O5iii) (Supplementary Data S1). *snaR* loci are bracketed; nucleotide heterogeneity is highlighted in gray and unique bases in black. Asterisks denote identity in all 5 Great Ape species; dots denote identity in the African Great Apes (orangutan excepted). Putative Pol III B-box is boxed. PCR priming sites are denoted by arrows.

Table 1. Subsets of African Great Ape *snaR* genes

snaR[a]	NCBI Gene ID	Chromosome[b]	Transcribed[c]	EST[d]	Human[e]	*Pan*[e]	Gorilla[e]
A1–14	100126798–99, 100169951–59, 100170216 and 100191063	19q13 (53.10–53.14, 55.28–55.32)	+[f]	BU536065 DW456830	14*		*
B1–2	100170217, 100170224	19q13 (55.32–55.33)	+[f]	BU960991 BU960401 BF691336 BQ435479	2*		
C1–5	100170218–19, 100170223, 100170225–26	19q13 (53.10–53.15)	+[f]	BU570852 CA337614 BF977253 BF691343	5*		
D	100170227	19q13 (55.33)	nd[g]		1		
E	100170220	19q13 (52.02)	nd[g]		1		
F	100126781	19q13 (55.80)	nd		1		
G1	100126780	19q13 (54.23)	+[h]		1*	2*	*
G2	100170228	19q13 (54.22)	+[h]		1*		*
H (2)	100170221	2p12 (78.03)	nd		1	1	
I (3)	100170222	3q28 (192.07)	+[f]		1	7*	
12		12p12 (21.15)	nd[g]		1	1	
21		21q21 (25.03)	nd[g]		1	1	

[a]*snaR* nomenclature as defined by HGNC (former names in parentheses). Table is modified from (32).
[b]Human chromosome location (in Mb).
[c]snaR expression as confirmed by northern blot, RT–PCR and/or Deep Sequencing of human tissue or transfected cell line. nd, not determined.
[d]Accession numbers of human ESTs with minimal flanking sequence and perfect identity or containing one mismatch to a *snaR* gene.
[e]Number of gene copies by bioinformatic search. Asterisk (*) denotes confirmation of the *snaR* subset by genomic PCR (Supplementary Data S1).
[f]Co-immunoprecipitated with NF90b, detected by Deep Sequencing (unpublished data).
[g]*snaR-D* and *-E* are pseudogenes of *snaR-A* and *-B*, respectively (8). *snaR-12* and *-21* are flanked by short direct repeats and appear to be pseudogenes.
[h]Co-immunoprecipitated with Pol III, detected by Deep Sequencing (73).

the orangutan draft assembly also failed to find *snaR* genes, while searches of Western gorilla whole genome shotgun sequences confirmed the presence of *snaR* genes. We conclude that *snaR* genes are restricted to the African Great Apes, and that distinct but related elements exist in orangutan.

The origin of *snaR* genes

To determine whether similar elements exist in lower primates, we conducted a BLAT search for the orangutan Pa19 clone sequence against the rhesus macaque genome. Similarity (85.5% identity over 297 bp) was identified with a segment on macaque chromosome 19. This macaque segment is part of a larger sequence that is demarcated by a SINE, *AluSx*, into three tandem repeats of ∼1.9 Kb (Figure 3A and Supplementary Figure S3; Supplementary Data S2). These repeats are conserved, at least partially, in other primates. A syntenic triple tandem repeat is present on human chromosome 19 (Figure 3A). Orthology corresponding to ∼1.5 repeats was found on orangutan chromosome 19 (at 52.170–52.174 Mb; Supplementary Figure S3), and with 'undescribed' portions of the draft orangutan genome. Similarly, orthology was observed on chimpanzee chromosome 19 (at 56.280–56.284 Mb) and with unassigned genomic sequences (not shown).

The first and second ∼1.9 Kb repeats in macaque contain elements (termed *Mm19* and *Mm19i*; Supplementary Data S3) that have strong sequence identity with orangutan *Pa19* (Figure 3A and

Supplementary Figure S3). Strikingly, the equivalent position in the first repeat on human chromosome 19 contains *snaR-F*. Thus, human *snaR-F*, orangutan *Pa19* and macaque *Mm19* appear to derive from the same ancestral locus, implying that *snaR* genes have evolved from a genetic element resembling those in macaque and orangutan.

Evolution of *CAS* in Old World Monkeys and apes

A search for the novel element in human, chimpanzee, orangutan and macaque genomes yielded over 20 hits in each species (Supplementary Figure S4 and Supplementary Data S3). MegaBLAST searches for orthologous elements in hamadryas baboon, grivet and mantled guereza (Old World Monkeys), and in white-cheeked crested gibbon and western gorilla also yielded multiple hits. No contiguous hits were obtained in five New World Monkey genomes: those of Nancy Ma's night monkey, common marmoset, black-capped squirrel monkey, Geoffroy's spider monkey and red-bellied titi (Supplementary Data S3). Likewise, these elements were not found in the genomes of two prosimians, the ring-tailed lemur and the Philippine tarsier. These data indicate that the elements are restricted to the parvorder Catarrhini, the Old World Monkeys and apes. We therefore designate this novel element the 'Catarrhine ancestor of snaR' (*CAS*). Further searches disclosed excellent orthology in anubis baboon, grivet and mantled guereza
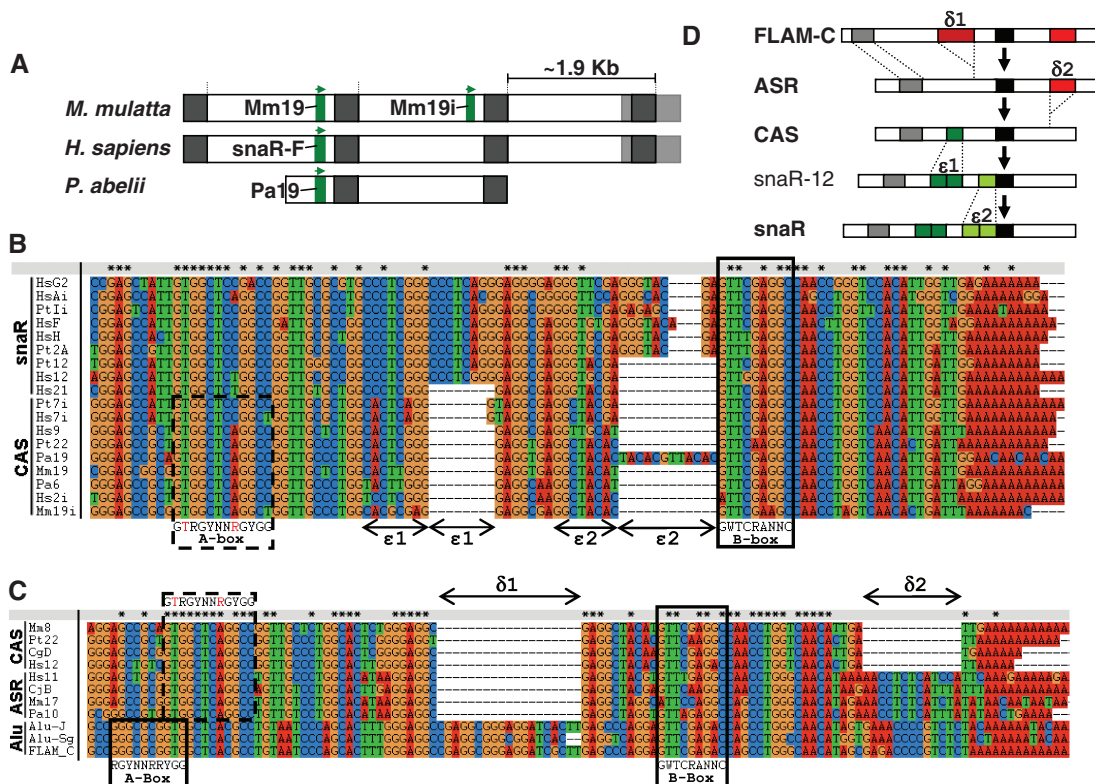
Figure 3. *snaR* molecular evolution from *FLAM C* via *ASR* and *CAS*. (**A**) Schematic of syntenic regions of chromosome 19 in rhesus macaque (56.819–56.825 Mb), human (55.798–55.804 Mb) and orangutan (52.171–52.174 Mb), containing ~1.9 Kb repeats (open boxes). *CAS* elements and *snaR-F* are indicated in green and *AluSx* in gray, with direction of transcription indicated by arrows. (**B**) Clustal-X alignment of selected *snaR* genes and *CAS* elements from humans (Hs), chimpanzee (Pt), orangutan (Pa) and rhesus macaque (Mm). Internal expansions ε1 and ε2 and a putative Pol III B box are indicated. The putative *CAS* Pol III A box is in a dashed box and is based on the *tRNA* consensus displayed below the alignment: invariant nucleotides are in red (27). (**C**) Clustal-X alignment of selected primate *CAS*, *ASR* and *Alu* elements. Deletions δ1 and δ2 are indicated. Elements are annotated in Supplementary Data 3. *Alu* Pol III A and B boxes are shown and their consensus sequences (74) displayed below the alignment; putative *CAS* Pol III A box is dashed and compared to a tRNA A box consensus displayed above the alignment. (**D**) Schematic of the major molecular deletions (in red) and expansions (in green) inferred to have occurred during the evolution of *ASR*, *CAS*, *snaR-12* and *snaR* genes from *FLAM C*. Pol III A and B boxes are represented by gray and black boxes, respectively. Schematics A and D are modified from (32).

(Supplementary Data S2), confirming the presence of *CAS* in both subfamilies of Old World Monkey.

CAS elements are ~85 nt long (excluding the 3′-oligo-(AT) tract), intergenic or intronic and scattered over human, chimpanzee, orangutan and macaque genomes. Most contain sequences corresponding to a Type 2 Pol III consensus B box promoter (+57–65 nt) and an A box enhancer (+11–22) (Figure 3B and Supplementary Figure S4) (27). Alignment of representative *CAS* and *snaR* sequences shows that *snaR* genes have two separate internal expansions located 5′ (ε1) and 3′ (ε2) to a central conserved region (Figure 3B). The expansions appear to have arisen through two separate duplications of 8-nt sequences between the Pol III A and B boxes. Additional expansions have occurred in some primate *CAS* elements: the orangutan *CAS* (*Pa19*) has a unique 3′-internal expansion in the ε2 region and another unique ε2 expansion has occurred in a gibbon (*Nomascus leucogenys*) *CAS* (termed *NlK*; Supplementary Data S3).

Elements with features intermediate between *snaR* and *CAS* shed light on the evolution of the *snaR* genes. First, two syntenic *CAS* elements on human and chimpanzee chromosomes 7 (*Hs7i* and *Pt7i*; Figure 3B) have

sequences intermediate between *CAS* and *snaR*, implying that *CAS* loci acquired sequences characteristic of *snaR* through substitution after divergence from orangutan (genus *Pongo*) and before the ε1 and ε2 expansions occurred. Second, *snaR* genes on Hominini chromosomes 12 and 21 have good similarity to other snaRs but lack one or both of the expansions. Specifically, *snaR-21* (*Hs21*) lacks both internal expansions while *snaR-12* (*Hs12* and *Pt12*) has ε1 but lacks ε2 (Figure 3B). These elements suggest an order for the changes involved in the *CAS-snaR* transformation, implying that nucleotide substitutions preceded expansion ε1 and then ε2 in the course of *snaR* gene evolution in the African Great Apes (Figure 3D).

Descent from the *Alu* SINE family via *ASR*

One means for an organism to acquire novel DNA is through viral infection. A search for *CAS* elements within a HERV database revealed a single example occupying a syntenic position in several Catarrhine species (Supplementary Figure S5). Because of its singularity and the absence of an oligo-(A/T) tract, it is unlikely that many *CAS* elements have arisen from this conserved *HERV*.

To identify a potential *CAS* progenitor, we performed a discontiguous MegaBLAST search of the macaque segment that contains *Mm19* against high-throughput genomic sequences. Orthology was identified in two New World Monkey species, Bolivian squirrel monkey (99% coverage, 79% identity) and common marmoset (42% coverage, 81% identity), which do not contain *CAS* elements, as mentioned earlier. The equivalent site in these New World Monkey genomes is occupied by *CAS*-related sequences with an additional ∼13 bp in their 3′-ends (Figure 3C, δ2; Supplementary Data S2). BLAST searches detected multiple copies of this novel locus in the genomes of the infraorder Simiiformes (Old and New World Monkeys and apes), but not in tarsiers or the Strepsirrhine prosimian small-eared galago even though the latter contained the orthologous segment (81% coverage, 85% identity).

This novel locus was recognized by RepeatMasker to have closest similarity to *FLAM C* and its descendant, the left monomer of *AluJb*, a subdivision of the oldest *Alu* SINE family, thought to have arisen around 55 MYA, before the evolution of Siimiformes (28) (Figure 3C). We find no database annotation of this locus and designate it Alu/snaR-related element (ASR). Hence, the *CAS* and *snaR* genes appear to be descended from the most numerous and primate-specific SINE family. This is in keeping with the similarity of sequence encompassing the B-box of *snaR-A* with *AluJ* noted previously (8). Both *ASR* and *CAS* elements, as well as their *snaR* descendants, carry an internal deletion (δ1) of 17–19 nt with respect to their ancestor. Thus, *CAS* elements appear to have evolved from *Alu* via *ASR* with two successive deletions, δ1 followed by δ2 (Figure 3D).

Genomic dissemination of *snaR*, *CAS* and *ASR*

The majority of *CAS* (>69% in humans) and *ASR* loci are flanked by short direct repeats (Supplementary Data S4). Encompassing short direct repeats are a hallmark of target site duplication (TSD), a characteristic of retrotransposition (29,30), indicating that *CAS* and *ASR* are novel retrotransposons. A cladogram of human, chimpanzee, orangutan and rhesus *CAS* elements identified an ortholog for each human *CAS* locus in at least one of the other three species (Supplementary Figure S8), and orthology was confirmed by examination of flanking sequences (Supplementary Data S4). Chimpanzees have a solitary non-orthologous *CAS* (Supplementary Figure S8, red arrow). This argues that few, if any, new genome insertion events have occurred in humans since the divergence from the common human/chimpanzee ancestor. By contrast, a substantial proportion of orangutan *CAS* elements (9 of 24, including two paralogs with 5′-homology) and macaque *CAS* elements (15 of 22) are species-specific, often flanked by TSDs (Supplementary Figure S8). Therefore, *CAS* have been, and perhaps still are, active retrotransposons in these species. The putative intermediates of the *CAS*-*snaR* transition, including *snaR-12* and *-21*, are also flanked by TSDs (Supplementary Data S4), indicating the continued

ability of this molecular lineage to retrotranspose even after multiple substitutions and an internal expansion.

On the other hand, almost all human and chimpanzee *snaRs* are surrounded by conserved flanking sequence (8), suggesting that they have disseminated through duplication of a larger encompassing segment or 'duplicon'. To trace the dispersal of *snaR* from its presumptive parental locus on chromosome 19 (Figure 3A), we conducted BLAT searches of the well annotated human genome for the ∼1.9 Kb segment containing *snaR-F* (Supplementary Figure S9 and Supplementary Data S5). This analysis identified partial duplications, each containing a *snaR* gene with a variable amount of flanking sequence, at the expected locations. From an examination of the flanking sequences in the duplicons, we propose that *snaR* genes diversified along two independent duplication pathways (Figure 4). The major pathway (Figure 4A), which contains most of the snaR genes, has long duplicons on chromosomes 2 and 3 (including *snaR-H* and *-I*) and multiple short duplications in the two large tandem arrays on chromosome 19 (*snaR-A*, *-B*, *-C* and *-D*). In the second pathway (Figure 4B), two short duplications (including *snaR-G1* and *-G2*) inserted into the *LH/CGβ* gene cluster gave rise to novel *CGβ* genes (A.M. Parrott *et al.*, submitted for publication).

Major pathway of *snaR* duplication

The H, I and A/B/C/D duplicons consist of a Core sequence with variable amounts of flanking sequence, derived from the ∼1.9 Kb segment containing snaR-F and from a DEAH box polypeptide 34 (*DHX34*) gene (Figure 4A). The *DHX34* gene is located ∼0.6 Mb downstream of the *snaR* A/C cluster on chromosome 19 (Figure 1). We infer that a fragment of the parental *snaR* locus was inserted into a copy of a region of the *DHX34* gene to generate a hypothetical intermediate (Figure 4A and Supplementary Figure S10A). Portions of this intermediate gave rise to the H-, I- and A/B/C/D-duplicons, which share the Core sequence consisting of a *snaR* gene surrounded by ∼0.5 Kb of parent sequence (Figure 4A and Supplementary Figure S10A).

The ∼1.4 Kb H-duplicon and ∼3.5 Kb I-duplicon are present in chimpanzee as well as human. Sequence comparison suggests that they were inserted into the ancestral African Great Ape genome via recombination of LINE sequence and A/T-rich (possibly *Alu*) sequence, respectively (Supplementary Figure S10B). In chimpanzee, duplication of a fragment containing the I-duplicon, the Pt-duplicon, has resulted in at least seven tandem copies of *snaR-I* (Supplementary Figure S10B, step iv; note that the draft sequence contains sequence gaps in this region, Supplementary Figure S9B). The ∼1.6 Kb Pt-duplicon includes a ∼1.1 Kb fragment of chromosome 3 (termed R3) and is bordered by *Alu* sequence, suggestive of repetitive *Alu*-*Alu*-mediated recombination. A burst of *Alu* retrotransposition ∼35–40 MYA inundated the primate genome with highly similar sequences, and *Alu*-*Alu*-mediated recombination is thought to be a common recombination mechanism in primates (31).
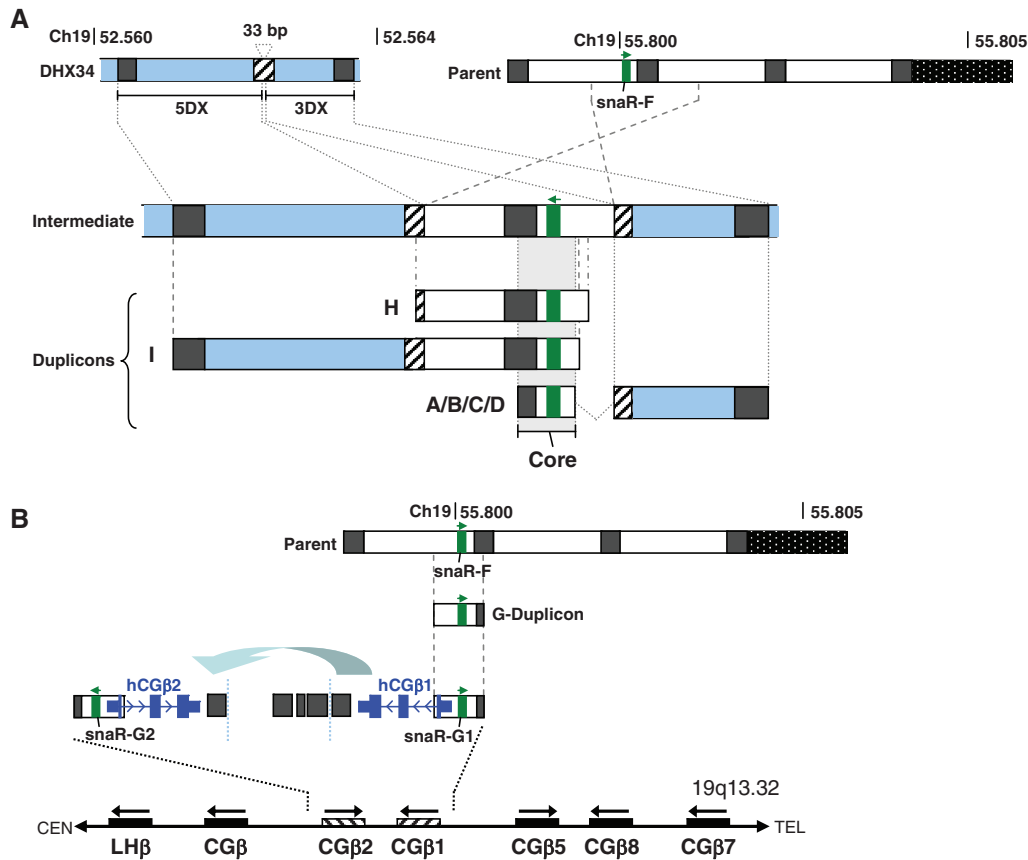
Figure 4. Dissemination of *snaR* by segmental duplication. (**A**) Formation of a hypothetical intermediate by insertion of a fragment of the *snaR* parent locus (open boxes, snaR in green) into a copy of a region of the *DHX34* gene (blue boxes). The insertion point is in long interspersed element (LINE) sequence (hatched), dividing the *DHX34* duplicate into '5DX' and '3DX' fragments, separated by 33 bp. The intermediate presumptively served as the source of H- and I-duplicons, which retain 5DX sequence, and of the A/B/C/D-duplicon which retains 3DX sequence. *Alu* sequences (dark gray boxes) are indicated. *snaR* transcription directionality within the duplicons is denoted by green arrowhead. (**B**) Proposed origin of the G-duplicon and origin of *CGβ1* and *CGβ2* genes (dark blue). Shown is the generation of the G-duplicon from the *snaR* parental locus and its insertion into a *CGβ* gene. The substitution of common *CGβ* gene sequence gave rise to *CGβ1* and *CGβ2* genes (dark blue). The large curved arrow represents the inverted segmental duplication of the original *CGβ1* gene. *Bottom*: Diagram of the human *LH/CGβ* gene cluster at chromosome 19q13.33, oriented with respect to centromere (CEN) and telomere (TEL). Arrows indicate direction of transcription.

The A/B/C/D-duplicons consist of the Core and *DHX34* fragments, together with an additional chromosome 19 sequence termed R19 (Supplementary Figure S10C). Assembly of these duplicons can be detected in gorilla genomic sequences, but not in chimpanzee (data not shown). R19 exists in the chimpanzee genome, as well as in orangutan and macaque genomes, at the position orthologous to the insertion point of the A/C cluster in human. Thus, it is likely that the A/B/C/D-duplicon was assembled in the African Great Ape ancestor at the site of the A/C cluster, and was subsequently lost in chimpanzee. Both the *DHX34* and R19 sequences are flanked by *Alu* sequences that may have facilitated tandem duplication (Supplementary Figure S10C). While the tandem repeats are uniform in the A/C cluster, their regularity is altered in the A/B/D cluster (Figure 1) chiefly as a result of variability in the lengths of their constituent *DHX34* (~0.7–1.2 Kb) and R19 (~0.7–3.6 Kb) sequences (Supplementary Figure S9C). Furthermore, the A/B/D cluster is 5′-flanked by a ~14.9 Kb region that is homologous to neighboring chromosome 19 sequence, including the last two exons

of the gene for the synaptic vesicle membrane protein synaptotagmin-3 (*SYT3*; Supplementary Figures S9 and S10C, step vii). Evidently, the snaR clusters were formed through multiple recombination events involving several sequences that span a considerable region of chromosome 19 (Figure 1). This region extends over ~3 Mb in macaque, a species in which assembly has not occurred.

Differential tissue expression of snaRs

The *snaR* genes have diversified considerably in the African Great Ape lineage. To determine whether their rapid evolution and sequence variation correspond to differential expression, we examined the tissue distribution of selected snaR transcripts. For this study we selected snaR-A and snaR-B, which were originally isolated from 293 cells; snaR-C, transcribed from the second most numerous group of *snaR* genes (Table 1); and snaR-G2, a single copy gene that is juxtaposed to a Pol II-transcribed gene. Real-time quantitative PCR (qPCR) was performed on cDNA prepared from total RNA extracted from 24 human tissues and 9 brain regions,
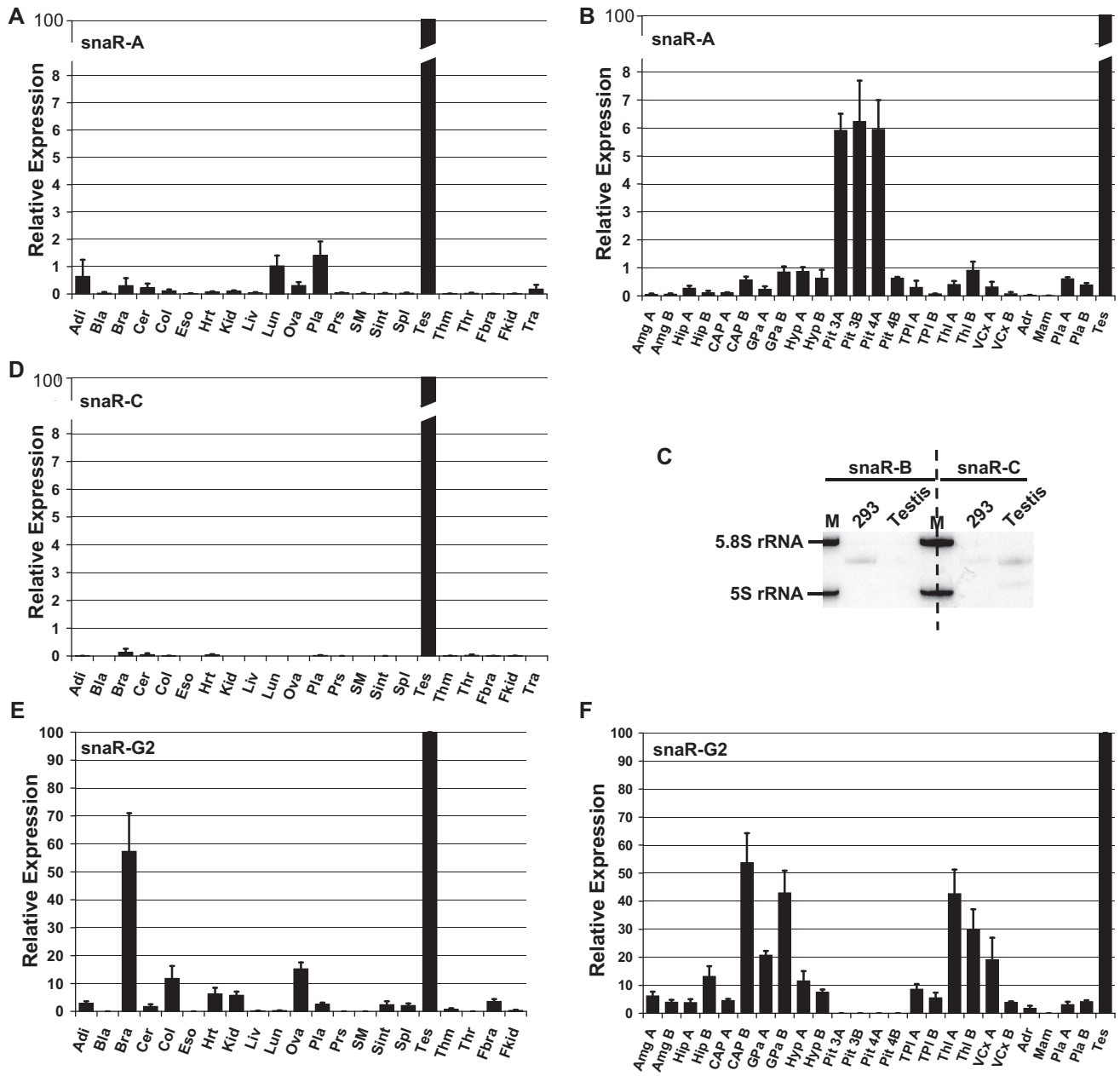
Figure 5. Differential expression of snaRs in brain and other tissues. Relative expression of snaR-A (**A**), snaR-C (**D**) and snaR-G2 (**E**) in human tissues and of snaR-A (**B**) and snaR-G2 (**F**) in regions of the human brain as determined by qPCR analysis. Histograms are averages of 3–4 separate experiments each performed in duplicate and normalized to placental sample. Standard deviations are shown. Two antipodal slices (A and B) were made for each brain tissue, 'placenta A and B' represent two different purchased lots. (**C**) Northern blots of total RNA extracted from 293 cells and testis, probed for snaR-B and -C. End-labeled 293 cell total RNA served as marker (M).

using primers and probes specific for these four RNA species.

Consistent with earlier observations (8), snaR-A was most abundant in testis where its expression was ~100-fold higher than in term placenta, lung and adipose tissue (Figure 5A). Lower expression was detected in several other tissues, and in total brain extract. snaR-A levels varied greatly in brain regions (Figure 5B). Most remarkably, snaR-A was present at ~10% of the testis level in three of four segments of pituitary gland (Figure 5B), possibly indicating differential expression in the gland's anterior and posterior lobes.

Substantial expression was observed in brain regions including the hypothalamus, globus pallidus and thalamus.

Three-fold more cDNA was required for reliable analysis of snaR-C and -G2, indicating that they are less highly expressed than snaR-A, while the abundance of snaR-B was too low for reliable quantification in tissue samples by qPCR. This species was originally detected in 293 cells (8) and northern blotting confirmed its low expression in testis relative to 293 cells (Figure 5C). snaR-C and -G2, which had not been observed previously, were readily quantified in testis where they were most
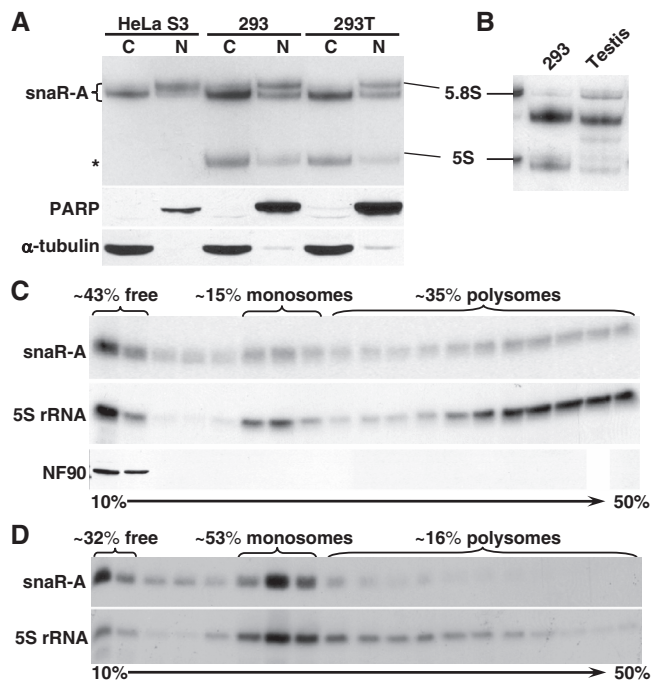
Figure 6. Subcellular localization of snaR-A. (**A**) Total RNA isolated from 293T, 293 and HeLa S3 cell nuclear and cytoplasmic extracts was examined for snaR-A by northern blotting (top panel). Extracts were immunoblotted for PARP (middle panel) and α-tubulin (bottom panel). Asterisk denotes a possible snaR-A conformer band. (**B**) Northern blot of 293 cell and testis total RNA, probed for snaR-A. (**C**) Co-sedimentation of snaR-A with ribosomes. HeLa cell extract was fractionated in a sucrose gradient and analyzed by northern blotting for snaR-A (top panel) and 5S rRNA (middle panel), and by immunoblotting for NF90 (bottom panel). The percentage of snaR-A in three regions of the gradient was determined by phosphorimager quantitation of the blot. Blot is duplicated from (32) with permission. (**D**) Extract of HeLa cells incubated with puromycin was analyzed as in **C**.

abundant (Figure 5D and E). A greater number of PCR cycles was required for snaR-G2 than for snaR-C (C_T ~28 instead of ~19) implying that snaR-G2 is less abundant than snaR-C in this tissue. Although snaR-C was detected in adult brain extract at low level (~700-fold lower than in testis), it was barely detectable in most other tissues (Figure 5D). snaR-G2 displayed a more diverse expression pattern (Figure 5E). Its abundance in adult brain was ~55% of that in testis, and several tissues displayed 5–15% of testis levels (Figure 5E). Fetal brain contained much less snaR-G2 than adult brain (~3% of testis level), suggestive of developmental regulation (Figure 5E). Like snaR-A, snaR-G2 expression was asymmetrical within the adult brain, but the distribution of the two species was distinctly different. The highest snaR-G2 expression was seen in the caudate/accumbens/putamen and globus pallidus, tissues that are interlinked physically and functionally to constitute the anterior of the basal ganglia; and in the thalamus, which is functionally connected to the basal ganglia (Figure 5F). snaR-A was also expressed in the basal ganglia (Figure 5B), but snaR-G2 was not detected in the pituitary gland (Figure 5F). The differential distribution of the snaR species suggests that they are functionally differentiated.

snaR-A is predominantly cytoplasmic and is ribosome associated

To approach their function, we examined the subcellular distribution of snaR-A. Fractionation of 293, 293T and HeLa cells revealed snaR-A to be predominantly cytoplasmic (Figure 6A, upper panel). Efficient cellular fractionation was confirmed by immunoblotting for the compartment specific proteins PARP and α-tubulin (Figure 6A, middle and lower panels). As reported earlier (32), snaR-A gel mobility is greatly affected by polyacrylamide density: in high percentage denaturing gels it resolves into several bands, with the major band migrating slower than 5S rRNA. Cytoplasmic forms of snaR-A displayed faster migration than the predominant nuclear form (Figure 6A, upper panel), indicating that the RNA is processed or adopts a different structure in the cytoplasm. snaR-A isolated from testis displayed a similar migration pattern to that in 293 cells, suggesting that snaR-A is predominantly cytoplasmic in this tissue (Figure 6B).

We next conducted sucrose gradient analysis of HeLa cell extracts to determine whether snaR-A is associated with ribosomes or polysomes. Northern blotting indicated that over half of snaR-A migrated in fractions containing monosomes (15%) and polysomes (35%; Figure 6C, top panel). NF90, a known protein partner of snaR-A, was detected only in fractions at the top of the gradient, indicating that NF90 is not stably associated with ribosome-associated snaR-A or with ribosomes themselves (Figure 6C, bottom panel). The ratio of monosome to polysome association was higher for snaR-A (~0.43) than for 5S rRNA (~0.1), suggesting that snaR-A preferentially associates with monosomes. Normalized to 5S rRNA, snaR-A is ~8-fold more abundant in the monosome population than in polysomes. It is noteworthy that BC200 and the Alu left monomer, two snaR-related non-coding RNAs, are also predominantly cytoplasmic and appear to regulate translation at the level of initiation (33,34). Puromycin treatment dissociated most of the polysomes and shifted the majority of snaR-A into the monosome region (~53%; Figure 6D, upper panel). Because high salt releases mRNA from puromycin-treated ribosomes (35), it is likely that snaR-A directly associates with ribosomes rather than with mRNA.

snaR-A induction by cellular immortalization

In contrast to its restricted distribution in human tissues, snaR-A is present at moderate to high levels in nearly all permanent cell lines tested (8). To determine whether it is induced by cell transformation, we examined two series of cell lines derived from normal human diploid fibroblasts by transformation with SV40. HS74 fetal bone marrow stromal fibroblasts (20,23) and HSF43 human foreskin fibroblasts (22) gave rise to preimmortal and immortal cell lineages as diagrammed in Figure 7. Extracts of these cells were examined by northern blotting for snaR-A, with 5S rRNA as control.

Parental HS74 cells had low levels of snaR-A that were not increased when the cells were aged in culture (Figure 7,

upper panel, compare HS74 and HS74 sen) (19). High levels of snaR-A were expressed in the immortal HS74 derivatives AR5, SV.RNS/HF-1 (HF-1) and cl39T. Immortal cells had 8- to 30-fold increased levels of snaR-A, whereas related preimmortal cell lines, SVtsA/HF-A (HF-A) and SV.RNS/HF (HF), displayed modest increases (2- to 3-fold). Elevated snaR-A expression was also observed in transformed immortal CT10-2A cells compared to their parental HSF43 cells (Figure 7). Thus, normal fibroblasts have low snaR-A levels which are increased in two transformed but not immortal cell lines, and further increased upon immortalization.

Several studies have shown that Pol III transcription is up-regulated in transformed cells (36). To determine whether the increase in snaR-A applies generally to Pol III transcripts, we compared the expression of BC200, which is also synthesized by Pol III (26), to that of snaR-A and 5S rRNA. Consistent with its neural cell specific expression (37,38), BC200 was highly expressed in fetal brain but not in testis (Figure 7, middle panel). Unlike snaR-A, BC200 levels were lower in preimmortal and immortal cell lines derived from HS74 fetal bone marrow stromal cells. No such reduction was observed in HSF43 human foreskin fibroblasts, in which the level of BC200 was low in the parental cells as well as the transformed line, but the snaR-A/BC200 ratio was increased in all cases (Figure 7). These data indicate that snaR-A expression is selectively increased when cells are transformed and immortalized.

DISCUSSION

snaR evolution and primate speciation

The RNA component of the signal recognition particle, 7SL RNA, is the ancestor of two highly successful SINE families, B1 in rodents and Alu in primates (39–41). B1 is the dominant SINE in rodents and is the ancestor of 4.5SH RNA, a non-coding RNA arranged in large tandem repeats (similar to *snaR*) in several families of this order (42). Alu is the most successful SINE in primates, with over one million copies accounting for >10% of the human genome (43). *Alu* is a ~300 nt dimeric repeat evolved from the fusion of *FLAM C* with *FRAM*, both of which independently derive from 7SL RNA (44). BC200, a neuron-specific, non-coding RNA (37) present in a single locus in Simiiformes (45) is descended from *FLAM C* (26,46). Our data indicate that *ASR* and *CAS* are low-copy retrotransposons and their non-coding RNA descendant, snaR, are among the most recently evolved members of the influential *Alu* SINE family.
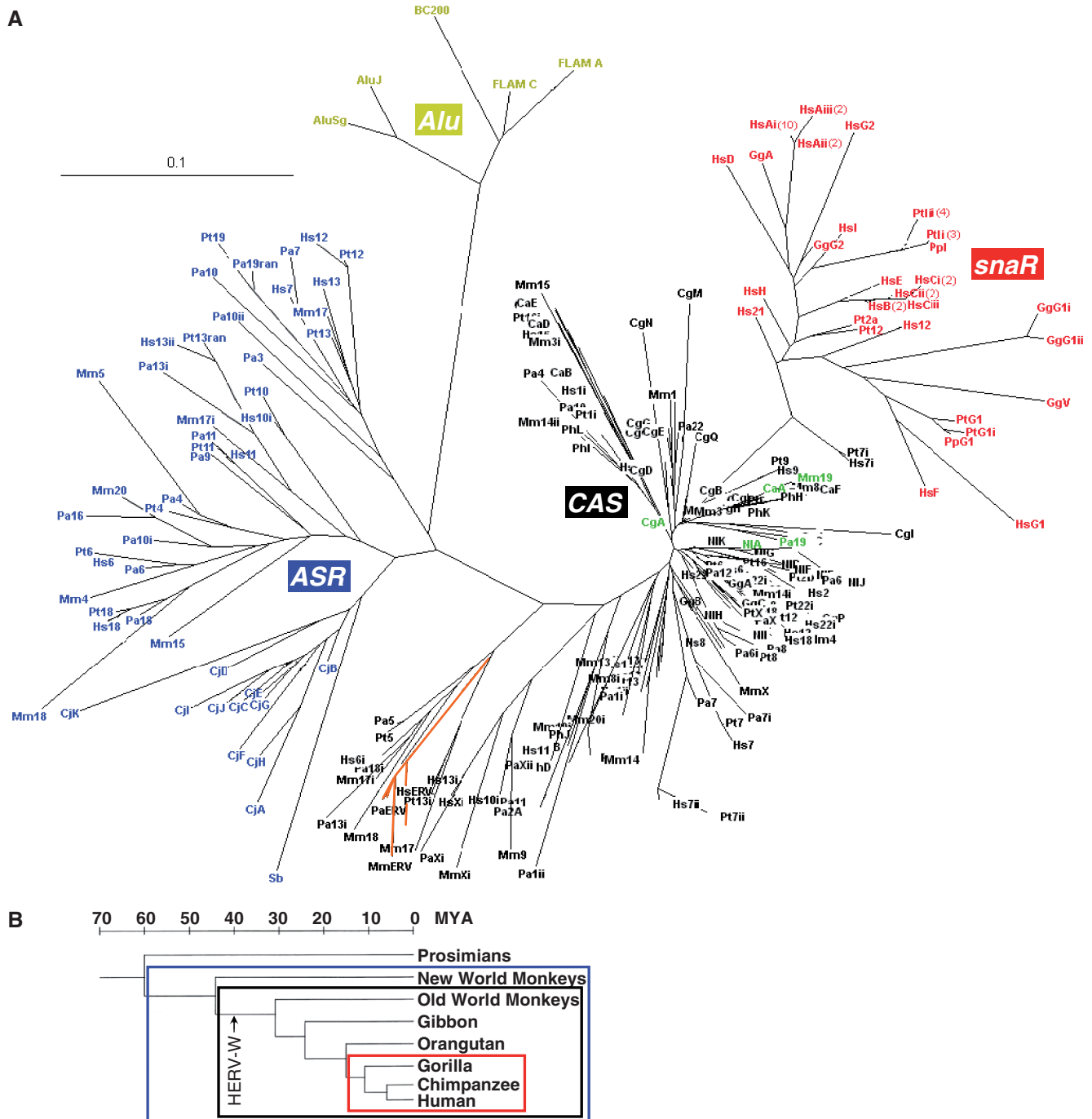
The major stages of *snaR* molecular evolution are clearly segregated in the unrooted phylogram shown in Figure 8A. Although the transitions between the stages are accompanied by insertion and deletion events (indels; Figure 3D), they are also characterized by sequence changes (Supplementary Figure S6). A similar phylogram results using an alignment from which the indels were omitted (Supplementary Figure S7). Hence, the stages
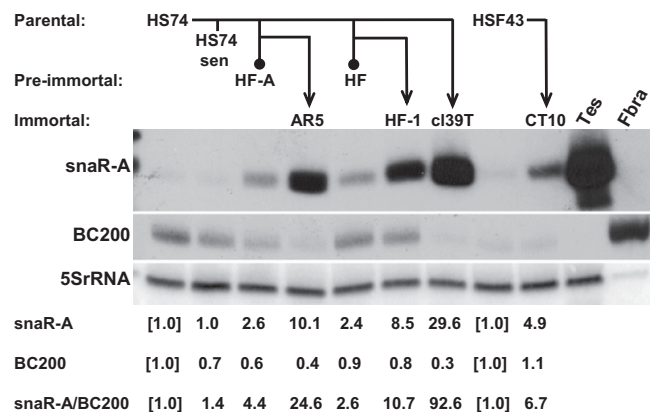
Figure 7. snaR-A induction by cell transformation and immortalization. Derivation of pre-immortal and immortal SV40-transformed cell lines from HS74 and HSF43 normal diploid fibroblasts is shown schematically, as described in the text and Methods. Senescent HS74 cultures (HS74 sen) are also represented. Northern blots of total RNA were probed for snaR-A (top), BC200 (middle) and 5S rRNA (bottom). Human testis (Tes) and fetal brain RNA (Fbra) were included as controls. Intensities of snaR-A and BC200 bands were determined relative to parental cells by phosphorimager quantification and the ratio of snaR-A to BC200 was calculated.

of snaR evolution are segregated by substitutions between the different molecular species as well as by indels. Strikingly, the evolution of each phylogram cluster in this molecular lineage coincides with major primate speciation events (Figure 8B). Thus, *ASR* elements are likely to have evolved in Simiiformes with the internal deletion of 17–19 nt from a *FLAM C*-like ancestor ($\delta 1$, Figure 3D). *CAS* elements probably evolved from *ASR* via a 3′-deletion ($\delta 2$, Figure 3D) in Catarrhines, after their geographical separation from New World Monkeys (Platyrrhines) in the late Eocene (30–45 MYA). The *snaR* genes appear to have evolved from CAS through two separate internal expansion events ($\varepsilon 1$ and $\varepsilon 2$, Figure 3D), in the common ancestor of African Great Apes in the Miocene Epoch (9–17 MYA). We conclude that the pathway of *snaR* gene evolution is characterized by a series of deletions followed by internal duplications, and its principal steps correlate with pivotal events in primate evolution: first, the divergence of simians from prosimians; second, the separation of Old World from New World monkeys; and third, the divergence of the African Great Apes from other apes.

The most abundant human *snaR* genes, *snaR-A* and *snaR-B/C*, form two distinct subsets in the phylogram and appear to have diverged rapidly from each other (Figure 8A). *snaR-A* is present in human and gorilla, arguing that they originated in the common ancestor of African Great Apes followed by loss in *Pan*. Sequence alignment of human and gorilla amplicons reveals that *snaR-A* has higher nucleotide conservation (with the exception of its unstructured 3′-end) than flanking sequence, suggesting nucleotide fixation (Supplementary Figure S11). On the other hand, *snaR-B* and -*C* are uniquely human (Table 1) and presumably evolved after

Figure 8. Stages of *snaR* molecular evolution correlate with major primate speciation events. (**A**) Unrooted phylogram displaying sequences of all known *snaR* genes (in red) together with 12 hamadryas baboon (Ph), 6 grivet (Ca), 17 mantled guereza (Cg), 23 rhesus macaque (Mm), 11 white-cheeked crested gibbon (Nl), 24 common chimpanzee (Pt), 24 human (Hs), 26 sumatran orangutan (Pa) and 3 western gorilla (Gg) *CAS* elements (in black) and *ASR* elements (in blue). Human *Alu* relatives (in gold) and squirrel monkey (Sb) and common marmoset (Cj) *ASR* elements are included. *CAS* elements located within the orthologous segment of chromosome 19 (in green) and in *HERV* (branch is shaded orange) are denoted. The scale bar represents 0.1 nt substitutions per site. Numbers in parentheses indicate *snaR* gene multiplicity. (**B**) Primate dendrogram illustrating the species range and longevity of *snaR* (red box) and *CAS* (black box) genes, and the possible extent of *ASR* genes (blue box). *HERV-W* integration into the primate lineage is indicated. Tree is modified from (75).

the *Homo-Pan* species divergence. *snaR-A* underwent copy number expansion (Supplementary Figure S9) giving rise to 14 paralogous alleles in human and an unknown number in gorilla. The distribution of *snaR-B* and *-C* suggests that they evolved recently from redundant

snaR-A copies (Figure 1). There are ~17 nt changes between human *snaR-A* (121 nt) and *snaR-C* (119 nt) species (~14% difference), compared to an average substitution rate of ~3.7 nt per 120 nt (~3% difference) outside of the *snaR* locus (Supplementary Data S6).

This rate of substitution approaches that observed between the human and chimpanzee orthologs of Human Accelerated Region 1 (*HAR1*) small non-coding RNA (18 substitutions in 118 nt, 15.3% difference), which is considered to be one of the most rapidly evolved RNAs in humans (47). Similar to human and chimpanzee HAR1 (48), these two snaR subsets are predicted to fold into distinctly different structures, with snaR-C adopting a cruciform structure similar to that predicted for BC200 and other 7SL derived RNAs (8,40). Through substitutions, snaR-C has possibly evolved to adopt an energetically favorable structure consistent with its function. Recent mutation is also evident in human *snaR-G2*, which contains a single nucleotide polymorphism at position +20 (Human Genome Diversity Project: rs3810177; Supplementary Data S3). Gorilla *snaR-G2* and the presumed human ancestral allele contain a G residue at this position. While G is present at >72% in some African tribes (Bantu and San), a gradual transition to an A residue has taken place and G is absent from certain geographically disparate populations (Melanesian, Pima and Surui tribes).

Putative *snaR* progenitor locus and segmental duplication

The major step-wise molecular rearrangements leading to *snaR* appear to have happened within a unique ~1.9 Kb sequence, which we infer is the parent locus (Figure 3A). *CAS* (and *ASR*) elements within the putative parent locus (such as *Mm19* and *Pa19*) appear to be transcriptionally viable, rather than products of retrotransposition, since they possess a Pol III B-box and oligo (U) tract, but lack flanking direct repeats (Supplementary Data S2). Furthermore, *CAS* elements within the parental locus cluster near to the phylogram node of the *snaR* root, within the most dense concentration of related *CAS* elements (Figure 8A, green lettering), consistent with this locus having been a source of *CAS* retrotransposition and of *snaR* evolution. In addition to the rearrangements entailed in *snaR* molecular evolution, unique internal duplications created *Pa19* within the corresponding orangutan locus. We therefore speculate that the putative parent locus can induce molecular rearrangement.

Segmental duplication can generate multiple copies of a gene, evident in the dramatic *snaR-A* and *snaR-I* copy number variation between Great Ape species (Table 1), and permits divergent evolution within a gene family. *snaR* genes appear to have participated in the genome-wide burst of segmental duplication that occurred in the ancestor of African Great Apes, an event that is thought to have had a profound impact on the adaptation and evolution of these hominids (49). *snaR* genes are embedded in a Core sequence that has participated in a complex series of genomic rearrangements leading to the present distribution of *snaR* in the human genome (Figure 4). The Core is reminiscent of the 'Duplication Cores' identified as the foci of complex interspersed duplication blocks (50) as well as a source of rapidly evolving genes undergoing positive selection (51–53).

Human segmental duplications favor a non-random, largely intrachromosomal distribution (51). Accordingly, the arrangement on human chromosome 19 of the *snaR* genes and the sequences involved in their dispersal (the *snaR* parental locus, *SYT3* and *DHX34*) displays a remarkable 2-fold symmetry (Figure 1). It is unclear how this symmetry arose since there is no evidence of a ~1.7 Mb inverted duplication of the protein-coding genes in this region, but possibly it reflects the 3D organization of these chromosomal locations in the nucleus. This region (~3.4 Mb from *DHX34* to *SYT3*) displays evidence of repeated recombination in humans, including: (i) the assembly of duplicons, (ii) the formation of clusters, (iii) duplication of the clusters and (iv) the *SYT3* and presumptive *DHX34* duplications. Intriguingly, a recent study of human copy-number variation in 30 individuals (from four ethnicities and both sexes) found that the two snaR clusters together with the intervening 2.1 Mb sequence, containing ~110 protein-coding genes, was absent from 18 haploid genomes (54), raising the possibility of ongoing recombinational deletions in this region that may have implications for disease susceptibility.

Tissue expression and functional implications

snaR species display markedly tissue-restricted expression patterns (Figure 5). They are predominantly expressed in testis, raising the possibility that they function in male reproductive biology (Figure 5). Intriguingly, a number of male fertility genes are located between the snaR clusters (8), including binder of sperm protein homolog 1 (BSPH1) and epididymal sperm binding protein 1 (ELSPBP1) which closely flank the snaR-A/C cluster. The present study also finds discrete expression of snaR-A in the pituitary gland and its relative absence from other endocrine organs, such as the adrenal and mammary glands and from kidney, liver, thymus and thyroid (Figure 5A and B). The anterior lobe of the pituitary gland is integral to the hypothalamic-pituitary-gonadal axis, an endocrine system that is critical to the development and homeostasis of the reproductive and immune systems, two highly complex and rapidly evolving systems in the hominids (55–57). The presence of snaR-G2 in basal ganglia tissues is also intriguing as the rapid evolution of primate male reproductive genes has been linked with socio-sexual behavior (58).

snaR-A was first isolated in complex with NF90 protein (8) and siRNA knock-down of NF90 results in reduced levels of snaR-A, indicating the importance of their cellular interaction (32). NF90 is a protein implicated in transcriptional and translational control (8,59–63), raising the possibility that snaRs participate in tissue-specific regulation of gene expression. Here we report that a substantial proportion of cytoplasmic snaR-A is stably associated with monosomes and polysomes, preferentially with the former, suggestive of a role in translation initiation (Figure 6). BC200 and its rodent functional analog, BC1, decouple the RNA helicase activity of translation initiation factor eIF4A from its ATPase activity (33). BC1 binds to eIF4A and PABP, and is thought to regulate translation of certain mRNAs at the synapse

through prevention of 48 S preinitiation complex formation (64). Hence, snaR-A and BC200 share a common ancestry, have tightly controlled tissue expression, and are associated with translation.

Both snaR-A and BC200 are dysregulated in cell lines. snaR-A is expressed in most, though not all, cell lines (8). snaR-G2 expression is similarly dysregulated (Figure 5; A.M. Parrott *et al.*, submitted for publication). BC200 is expressed in immortal cell lines (38) and neoplasms (65,66), and was up-regulated with snaR-A in cyclophilin B-depleted cells (67). Likewise, expression of neuron and testis-specific BC1 RNA (68) is observed in cultured cells (69). Upstream flanking sequence is essential for *BC1* transcription in rat cortex extract but not in HeLa cell extract (70), and for maintaining tissue-specific BC1 expression in transgenic mice (71). BC200 and Lorisoidea-specific G22 RNA, are also embedded in a locus that dictates their brain specific expression (72). Thus we speculate that the conserved flanking sequence of the putative snaR parental locus exerts tissue-specific transcriptional control over its 'tenant' *ASR*, *CAS* or *snaR* gene.

Despite these similarities, there are differences between snaR-A and BC200. First, whereas a large fraction of snaR-A is stably associated with polysomes and monosomes (Figure 5), BC1 and BC200 inhibit an early stage of preinitiation complex formation and have not been reported to associate with ribosomes. Second, snaR-A is predominantly found in testis and in specific regions of the brain, while BC200 and BC1 are almost exclusively neuronal with weak expression in testis (37), although BC1 is significantly expressed in early rodent spermatogenesis (68). Third, snaR-A expression is tightly correlated with transformation and immortalization in matched sets of HS74 cell lines, suggesting a positive role in cell growth (Figure 7). On the other hand, BC200 levels were modestly reduced by immortalization, despite its presence in immortalized cell lines and tumors. The snaR family displays the hallmarks of a recently exapted (recruited into function) descendant of the Alu SINE family. Possibly, these non-coding RNAs, ultimately derived from 7SL RNA, retain a role in cell growth and translational control.

ACCESSION NUMBERS

snaR genes GgA (FJ844896), PpI (FJ844897), GgG2 (FJ844898), PpG1 (FJ844899), GgV (FJ844900), GgG1i (FJ844901) and GgG1ii (FJ844902). CAS genes Pa19i (FJ844867) and Pa19 (FJ844878). PCR clones GQ59336-415.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Wilhelm,B.T., Marguerat,S., Watt,S., Schubert,F., Wood,V., Goodhead,I., Penkett,C.J., Rogers,J. and Bahler,J. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, **453**, 1239–1243.
2. Kapranov,P., Cheng,J., Dike,S., Nix,D.A., Duttagupta,R., Willingham,A.T., Stadler,P.F., Hertel,J., Hackermuller,J., Hofacker,I.L. *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**, 1484–1488.
3. Taft,R.J., Pheasant,M. and Mattick,J.S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*, **29**, 288–299.
4. Prasanth,K.V. and Spector,D.L. (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Gene Dev.*, **21**, 11–42.
5. Mattick,J.S. and Makunin,I.V. (2005) Small regulatory RNAs in mammals. *Hum. Mol. Genet.*, **14(Spec No. 1)**, R121–R132.
6. Duchange,N., Pidoux,J., Camus,E. and Sauvaget,D. (2000) Alternative splicing in the human interleukin enhancer binding factor 3 (ILF3) gene. *Gene*, **261**, 345–353.
7. Saunders,L.R., Perkins,D.J., Balachandran,S., Michaels,R., Ford,R., Mayeda,A. and Barber,G.N. (2001) Characterization of two evolutionarily conserved, alternatively spliced nuclear phosphoproteins, NFAR-1 and -2, that function in mRNA processing and interact with the double-stranded RNA-dependent protein kinase, PKR. *J. Biol. Chem.*, **276**, 32300–32312.
8. Parrott,A.M. and Mathews,M.B. (2007) Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution. *Nucleic Acids Res.*, **35**, 6249–6258.
9. Corthesy,B. and Kao,P.N. (1994) Purification by DNA affinity chromatography of two polypeptides that contact the NF-AT DNA binding site in the interleukin 2 promoter. *J. Biol. Chem.*, **269**, 20682–20690.
10. Hallast,P., Rull,K. and Laan,M. (2007) The evolution and genomic landscape of CGB1 and CGB2 genes. *Mol. Cell. Endocrinol.*, **260–262**, 2–11.
11. Hallast,P., Saarela,J., Palotie,A. and Laan,M. (2008) High divergence in primate-specific duplicated regions: human and chimpanzee Chorionic Gonadotropin Beta genes. *BMC Evol. Biol.*, **8**, 195.
12. Kent,W.J. (2002) BLAT–the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
13. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
14. Zhang,Z., Schwartz,S., Wagner,L. and Miller,W. (2000) A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.*, **7**, 203–214.
15. Page,R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**, 357–358.

16. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.

17. Villesen,P., Aagaard,L., Wiuf,C. and Pedersen,F.S. (2004) Identification of endogenous retroviral reading frames in the human genome. *Retrovirology*, **1**, 32.

18. Strauss,W.M. (1998) Preparation of genomic DNA from mammalian tissue. In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*, Vol. I. John Wiley & Sons, Inc., New York, pp. 2.2.1–2.2.3.

19. Hubbard,K., Dhanaraj,S.N., Sethi,K.A., Rhodes,J., Wilusz,J., Small,M.B. and Ozer,H.L. (1995) Alteration of DNA and RNA binding activity of human telomere binding proteins occurs during cellular senescence. *Exp. Cell Res.*, **218**, 241–247.

20. Hubbard-Smith,K., Patsalis,P., Pardinas,J.R., Jha,K.K., Henderson,A.S. and Ozer,H.L. (1992) Altered chromosome 6 in immortal human fibroblasts. *Mol. Cell. Biol.*, **12**, 2273–2281.

21. Banga,S.S., Kim,S., Hubbard,K., Dasgupta,T., Jha,K.K., Patsalis,P., Hauptschein,R., Gamberi,B., Dalla-Favera,R., Kraemer,P. et al. (1997) SEN6, a locus for SV40-mediated immortalization of human cells, maps to 6q26-27. *Oncogene*, **14**, 313–321.

22. Ray,F.A. and Kraemer,P.M. (1992) Frequent deletions in nine newly immortal human cell lines. *Cancer Genet. Cytogenet.*, **59**, 39–44.

23. Banga,S.S., Peng,L., Dasgupta,T., Palejwala,V. and Ozer,H.L. (2009) PHF10 is required for cell proliferation in normal and SV40-immortalized human fibroblast cells. *Cytogenet. Genome Res.*, **126**, 227–242.

24. Borson,N.D., Salo,W.L. and Drewes,L.R. (1992) A lock-docking oligo(dT) primer for 5′ and 3′ RACE PCR. *PCR methods and applications*, **2**, 144–148.

25. Parrott,A.M., Walsh,M.R. and Mathews,M.B. (2007) Analysis of RNA:protein interactions in vivo: identification of RNA-binding partners of nuclear factor 90. *Method. Enzymol.*, **429**, 243–260.

26. Martignetti,J.A. and Brosius,J. (1993) BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. *Proc. Natl Acad. Sci. USA*, **90**, 11563–11567.

27. Sharp,S.J., Schaack,J., Cooley,L., Burke,D.J. and Söll,D. (1984) Structure and transcription of eukaryolic tRNA genes. In Fasman,G. (ed.), *Critical Reviews in Biochemistry*, Vol. 97. CRC Press, Boca Raton, Florida, pp. 107–144.

28. Labuda,D. and Striker,G. (1989) Sequence conservation in Alu evolution. *Nucleic Acids Res.*, **17**, 2477–2491.

29. Maraia,R.J., Chang,D.Y., Wolffe,A.P., Vorce,R.L. and Hsu,K. (1992) The RNA polymerase III terminator used by a B1-Alu element can modulate 3′ processing of the intermediate RNA product. *Mol. Cell. Biol.*, **12**, 1500–1506.

30. Grindley,N.D. (1978) IS1 insertion generates duplication of a nine base pair sequence at its target site. *Cell*, **13**, 419–426.

31. Bailey,J.A., Liu,G. and Eichler,E.E. (2003) An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.*, **73**, 823–834.

32. Parrott,A.M. and Mathews,M.B. (2009) snaR genes: recent descendants of Alu involved in the evolution of chorionic gonadotropins. In Ebert,J. (ed.), *Cold Spring Harbor Symposia on Quantitative Biology: Evolution: The Molecular Landscape*, Vol. 74. Cold Spring Harbor Laboratory Press, Woodbury, NY, pp. 363–373.

33. Lin,D., Pestova,T.V., Hellen,C.U. and Tiedge,H. (2008) Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism. *Mol. Cell. Biol.*, **28**, 3008–3019.

34. Hasler,J. and Strub,K. (2006) Alu RNP and Alu RNA regulate translation initiation in vitro. *Nucleic Acids Res.*, **34**, 2374–2385.

35. Azzam,M.E. and Algranati,I.D. (1973) Mechanism of puromycin action: fate of ribosomes after release of nascent protein chains from polysomes. *Proc. Natl Acad. Sci. USA*, **70**, 3866–3869.

36. Marshall,L. and White,R.J. (2008) Non-coding RNA production by RNA polymerase III is implicated in cancer. *Nat. Rev. Cancer*, **8**, 911–914.

37. Tiedge,H., Chen,W. and Brosius,J. (1993) Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. *J. Neurosci.*, **13**, 2382–2390.

38. Watson,J.B. and Sutcliffe,J.G. (1987) Primate brain-specific cytoplasmic transcript of the Alu repeat family. *Mol. Cell. Biol.*, **7**, 3324–3327.

39. Kriegs,J.O., Churakov,G., Jurka,J., Brosius,J. and Schmitz,J. (2007) Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.*, **23**, 158–161.

40. Labuda,D. and Zietkiewicz,E. (1994) Evolution of secondary structure in the family of 7SL-like RNAs. *J. Mol. Evol.*, **39**, 506–518.

41. Quentin,Y. (1994) A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic Acids Res.*, **22**, 2222–2227.

42. Gogolevskaya,I.K., Koval,A.P. and Kramerov,D.A. (2005) Evolutionary history of 4.5SH RNA. *Mol. Biol. Evol.*, **22**, 1546–1554.

43. Consortium,I.H.G.S. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

44. Quentin,Y. (1992) Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res.*, **20**, 487–493.

45. Skryabin,B.V., Kremerskothen,J., Vassilacopoulou,D., Disotell,T.R., Kapitonov,V.V., Jurka,J. and Brosius,J. (1998) The BC200 RNA gene and its neural expression are conserved in Anthropoidea (Primates). *J. Mol. Evol.*, **47**, 677–685.

46. Ludwig,A., Rozhdestvensky,T.S., Kuryshev,V.Y., Schmitz,J. and Brosius,J. (2005) An unusual primate locus that attracted two independent Alu insertions and facilitates their transcription. *J. Mol. Biol.*, **350**, 200–214.

47. Pollard,K.S., Salama,S.R., Lambert,N., Lambot,M.A., Coppens,S., Pedersen,J.S., Katzman,S., King,B., Onodera,C., Siepel,A. et al. (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, **443**, 167–172.

48. Beniaminov,A., Westhof,E. and Krol,A. (2008) Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA*, **14**, 1270–1275.

49. Marques-Bonet,T., Kidd,J.M., Ventura,M., Graves,T.A., Cheng,Z., Hillier,L.W., Jiang,Z., Baker,C., Malfavon-Borja,R., Fulton,L.A. et al. (2009) A burst of segmental duplications in the genome of the African great ape ancestor. *Nature*, **457**, 877–881.

50. Bailey,J.A. and Eichler,E.E. (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. *Nature Rev.*, **7**, 552–564.

51. Jiang,Z., Tang,H., Ventura,M., Cardone,M.F., Marques-Bonet,T., She,X., Pevzner,P.A. and Eichler,E.E. (2007) Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nat. Genet.*, **39**, 1361–1368.

52. Johnson,M.E., Viggiano,L., Bailey,J.A., Abdul-Rauf,M., Goodwin,G., Rocchi,M. and Eichler,E.E. (2001) Positive selection of a gene family during the emergence of humans and African apes. *Nature*, **413**, 514–519.

53. Ciccarelli,F.D., von Mering,C., Suyama,M., Harrington,E.D., Izaurralde,E. and Bork,P. (2005) Complex genomic rearrangements lead to novel primate gene function. *Genome Res.*, **15**, 343–351.

54. Perry,G.H., Ben-Dor,A., Tsalenko,A., Sampas,N., Rodriguez-Revenga,L., Tran,C.W., Scheffer,A., Steinfeld,I., Tsang,P., Yamada,N.A. et al. (2008) The fine-scale and complex architecture of human copy-number variation. *Am. J. Hum. Genet.*, **82**, 685–695.

55. Tanriverdi,F., Silveira,L.F., MacColl,G.S. and Bouloux,P.M. (2003) The hypothalamic-pituitary-gonadal axis: immune function and autoimmunity. *J. Endocrinol.*, **176**, 293–304.

56. Hapgood,J.P., Sadie,H., van Biljon,W. and Ronacher,K. (2005) Regulation of expression of mammalian gonadotrophin-releasing hormone receptor genes. *J. Neuroendocrinol.*, **17**, 619–638.

57. Cheung,J., Estivill,X., Khaja,R., MacDonald,J.R., Lau,K., Tsui,L.C. and Scherer,S.W. (2003) Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.*, **4**, R25.

58. Wyckoff,G.J., Wang,W. and Wu,C.I. (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature*, **403**, 304–309.

59. Pfeifer,I., Elsby,R., Fernandez,M., Faria,P.A., Nussenzveig,D.R., Lossos,I.S., Fontoura,B.M., Martin,W.D. and Barber,G.N. (2008) NFAR-1 and -2 modulate translation and are required for efficient host defense. *Proc. Natl Acad. Sci. USA*, **105**, 4173–4178.

60. Vumbaca,F., Phoenix,K.N., Rodriguez-Pinto,D., Han,D.K. and Claffey,K.P. (2008) Double-stranded RNA-binding protein regulates vascular endothelial growth factor mRNA stability, translation, and breast cancer angiogenesis. *Mol. Cell. Biol.*, **28**, 772–783.

61. Kuwano,Y., Kim,H.H., Abdelmohsen,K., Pullmann,R. Jr, Martindale,J.L., Yang,X. and Gorospe,M. (2008) MKP-1 mRNA stabilization and translational control by RNA-binding proteins HuR and NF90. *Mol. Cell. Biol.*, **28**, 4562–4575.

62. Xu,Y.H., Busald,C. and Grabowski,G.A. (2000) Reconstitution of TCP80/NF90 translation inhibition activity in insect cells. *Mol. Genet. Metabol.*, **70**, 106–115.

63. Shi,L., Godfrey,W.R., Lin,J., Zhao,G. and Kao,P.N. (2007) NF90 regulates inducible IL-2 gene expression in T cells. *J. Exp. Med.*, **204**, 971–977.

64. Wang,H., Iacoangeli,A., Popp,S., Muslimov,I.A., Imataka,H., Sonenberg,N., Lomakin,I.B. and Tiedge,H. (2002) Dendritic BC1 RNA: functional role in regulation of translation initiation. *J. Neurosci.*, **22**, 10232–10241.

65. Chen,W., Bocker,W., Brosius,J. and Tiedge,H. (1997) Expression of neural BC200 RNA in human tumours. *J. Pathol.*, **183**, 345–351.

66. Iacoangeli,A., Lin,Y., Morley,E.J., Muslimov,I.A., Bianchi,R., Reilly,J., Weedon,J., Diallo,R., Bocker,W. and Tiedge,H. (2004) BC200 RNA in invasive and preinvasive breast cancer. *Carcinogenesis*, **25**, 2125–2133.

67. Fang,F., Flegler,A.J., Du,P., Lin,S. and Clevenger,C.V. (2009) Expression of cyclophilin B is associated with malignant progression and regulation of genes implicated in the pathogenesis of breast cancer. *Am. J. Pathol.*, **174**, 297–308.

68. Muslimov,I.A., Lin,Y., Heller,M., Brosius,J., Zakeri,Z. and Tiedge,H. (2002) A small RNA in testis and brain: implications for male germ cell development. *J. Cell Sci.*, **115**, 1243–1250.

69. McKinnon,R.D., Shinnick,T.M. and Sutcliffe,J.G. (1986) The neuronal identifier element is a cis-acting positive regulator of gene expression. *Procd. Natl Acad. Sci. USA*, **83**, 3751–3755.

70. Martignetti,J.A. and Brosius,J. (1995) BC1 RNA: transcriptional analysis of a neural cell-specific RNA polymerase III transcript. *Mol. Cell. Biol.*, **15**, 1642–1650.

71. Roy,A.M., West,N.C., Rao,A., Adhikari,P., Aleman,C., Barnes,A.P. and Deininger,P.L. (2000) Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.*, **302**, 17–25.

72. Khanam,T., Rozhdestvensky,T.S., Bundman,M., Galiveti,C.R., Handel,S., Sukonina,V., Jordan,U., Brosius,J. and Skryabin,B.V. (2007) Two primate-specific small non-protein-coding RNAs in transgenic mice: neuronal expression, subcellular localization and binding partners. *Nucleic Acids Res.*, **35**, 529–539.

73. Raha,D., Wang,Z., Moqtaderi,Z., Wu,L., Zhong,G., Gerstein,M., Struhl,K. and Snyder,M. (2010) Close association of RNA polymerase II and many transcription factors with Pol III genes. *Procd. Natl Acad. Sci. USA*, **107**, 3639–3644.

74. Paolella,G., Lucero,M.A., Murphy,M.H. and Baralle,F.E. (1983) The Alu family repeat promoter has a tRNA-like bipartite structure. *EMBO J.*, **2**, 691–696.

75. Kim,H.S., Takenaka,O. and Crow,T.J. (1999) Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. *J. Gen. Virol.*, **80**, 2613–2619.