Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution

Andrew M. Parrott and Michael B. Mathews*

Department of Biochemistry and Molecular Biology, New Jersey Medical School, UMDNJ, 185 South Orange Ave., P. O. Box 1709, Newark, NJ 07101-1709, USA

ABSTRACT

Nuclear factor 90 (NF90) is a double-stranded RNA-binding protein implicated in multiple cellular functions, but with few identified RNA partners. Using *in vivo* cross-linking followed by immuno-precipitation, we discovered a family of small NF90-associated RNAs (snaR). These highly struc-tured non-coding RNAs of ~117 nucleotides are expressed in immortalized human cell lines of diverse lineages. In human tissues, they are abun-dant in testis, with minor distribution in brain, placenta and some other organs. Two snaR subsets were isolated from human 293 cells, and additional species were found by bioinformatic analysis. Their genes often occur in multiple copies arranged in two inverted regions of tandem repeats on chromosome 19. snaR-A is transcribed by RNA polymerase III from an intragenic promoter, turns over rapidly, and shares sequence identity with Alu RNA and two potential piRNAs. It interacts with NF90's double-stranded RNA-binding motifs. snaR ortho-logs are present in chimpanzee but not other mammals, and include genes located in the pro-moter of two chorionic gonadotropin hormone genes. snaRs appear to have undergone acceler-ated evolution and differential expansion in the great apes.

INTRODUCTION

Proteins in the nuclear factor 90 (NF90) family of double-stranded RNA-binding proteins participate in many aspects of vertebrate RNA metabolism [reviewed in (1)] and have been implicated in development (2,3), the cell cycle (4) and virus infection (5,6). The two most prominent protein isoforms are NF90 and NF110 (7,8). NF90 is also known as DRBP76 (9) and NFAR1 (10), and NF110 is synonymous with ILF3 (11), NFAR2 (10) and TCP110 (12). Both exist in complexes with a distinct protein, NF45 (D. Guan *et al.*, manuscript in preparation; 13).

Proteins in the NF90 family possess two double-stranded RNA binding motifs (dsRBMs) (14) as well as RGG (9) and GQSY (1) nucleic acid binding motifs. Through *in vitro* reconstitution experiments, the proteins have been found to interact with coding and non-coding RNAs of cellular (2,3,15–18) and viral origin (5,6,19). Most of these RNA species are not abundant in cells, yet our earlier work demonstrated that the dsRBMs of NF90 and NF110 are almost completely occupied by cellular RNA throughout the cell cycle (20). We, therefore, sought to identify the predominant *in vivo* binding partners of NF90 by cloning the RNA species that were cross-linked to NF90 in live 293 cells using a rigorous and highly specific protocol (21). By this means we detected a hitherto unreported RNA family that forms complexes with NF90 protein and is designated snaR (small NF90-associated RNA).

This novel RNA family contains three distinct but related subsets that are encoded by tandemly repeated genes in two segments of human chromosome 19, as well as additional gene copies two of which are adjacent to genes encoding the β chain of chorionic gonadotropin. The snaRs are modern, rapidly evolving, non-coding RNAs of ~117 nucleotides (nt). Their genes are appar-ently restricted to humans and chimpanzees. snaR-A is highly structured, relatively unstable and synthesized by RNA polymerase III (Pol III) from an intragenic promoter. The snaRs were detected predominantly in testis among ~20 human tissues tested but are abundant in many immortalized cell lines. Their distribu-tion, genetic organization and evolutionary relationships suggest a biological role relevant to the speciation of great apes.

MATERIALS AND METHODS

GenBank accession numbers

snaR-A #EU035783, snaR-B #EU035784, human and chimpanzee snaR genes #EU071051-087.

*To whom correspondence should be addressed. Tel: 973 972 4411; Fax: 973 972 5594; Email: mathews@umdnj.edu

Cell lines

Human 293 cell sublines expressing NF90a and NF90b (21) were maintained in DMEM (Sigma) supplemented with 8% fetal bovine serum (Sigma) and 100 µg/ml Geneticin (Invitrogen).

Cross-linking, immunoprecipitation, 3′-end-labeling and RT-PCR

Procedures and recipes are detailed in (21). 293 cell sublines were blocked with 0.1 mg/ml cycloheximide then washed in phosphate-buffered saline (PBS, Sigma) with 0.1 mg/ml cycloheximide and 7 mM $MgCl_2$. Cells were incubated in PBS with 7 mM $MgCl_2$ and 0.5% formaldehyde for 10 min at 25°C before the addition of glycine (pH 7, 0.25 M final) and a further incubation of 5 min. Cells were harvested then lysed on ice for 5 min in RIPA buffer. Lysate was pre-cleared with Protein A-Sepharose (Amersham Biosciences) for 1 h at 4°C, centrifuged then incubated with anti-Omni-probe antibody (Santa Cruz Biotech.)-Protein A-Sepharose complex for 3 h at 4°C. Immunoprecipitates were washed five times, 10 min each at 25°C, in Harsh RIPA buffer. Immunoprecipitates were incubated in Cross-Link Reversal buffer at 70°C for 45 min and RNA isolated by Trizol extraction (Invitrogen). RNA was precipitated with 1 volume isopropanol in the presence of 1 M ammonium acetate and glycogen and the pellet washed in 75% ethanol, air-dried and re-suspended in water.

Isolated RNA was 3′-radiolabeled by ligation of [5′-α-^{32}P]-cytidine-3′,5′-bisphosphate, catalyzed by T4 RNA ligase (New England Biolabs) for 2 h on ice.

Immunoprecipitated RNA was incubated with DNase I (Amplification Grade, Invitrogen). A 'lock-dock' oligo(dT) primer (0.1 µM final) was annealed to RNA by heating at 70°C for 10 min. Reverse transcription was performed at 42°C for 50 min using SuperScript II Reverse Transcriptase (Invitrogen). The reaction was terminated at 70°C for 15 min and RNA was digested with RNase (15 U/µl RNase T1, 4 U/µl RNase H) for 30 min at 37°C. cDNA was twice purified through a QIAquick desalting column (Qiagen), heated at 94°C for 2 min in Tailing buffer then incubated with terminal deoxynucleotidyl transferase (New England BioLabs) for 10 min at 37°C. The reaction was terminated at 65°C for 10 min. cDNA was amplified over 30 cycles with an annealing temperature of 63°C, using a reverse primer, Abridged Anchor primer (Invitrogen) and SuperTaq polymerase (Ambion).

Plasmid constructs

GST fusion plasmids pGEX-6P-3-NF90a and -NF90b were subcloned from pcDNA3.1-NF90a and -NF90b (8) into pGEX-6P-3 vector (Amersham Biosciences) after EcoRI digestion. GST-NF90c was expressed from pGST-NF90 (22). pGEX-6P-3-A458P,A588P, was generated by digesting pcDNA3.1-NF110b(A458P,A588P) (23) with AvrII and HindIII, then inserting the gel purified fragment into similarly digested pGEX-6P-3-NF90b. To create pGEX4T-3-NF45, NF45 was PCR amplified from pcDNA3-NF45 (24) using primers F45 (5′-CGACAGAA
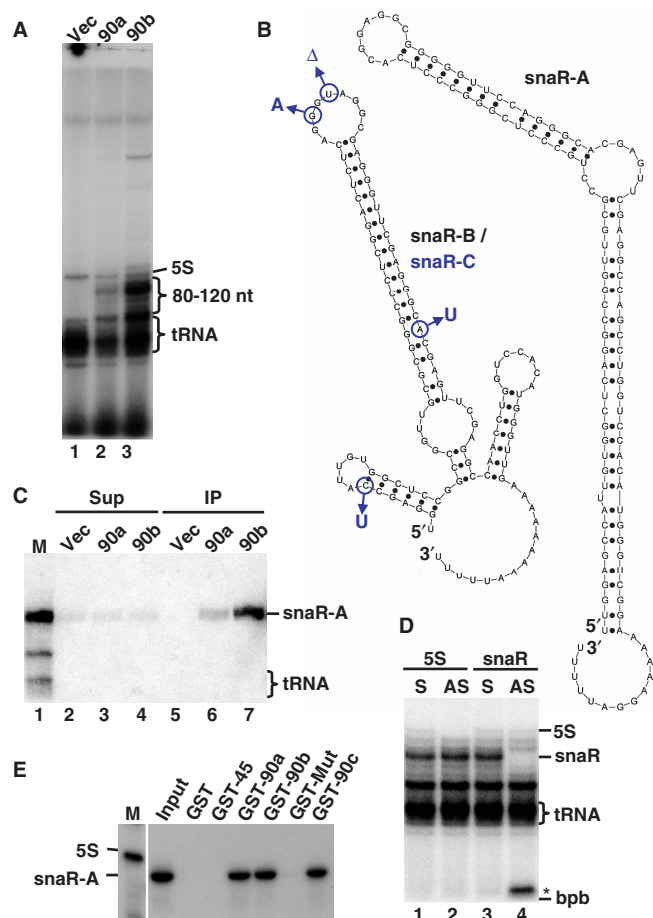

Figure 1. snaRs are highly structured NF90-associated RNAs. (**A**) RNA immunoprecipitated from cell lines with anti-omni antibody was 3′-end labeled and resolved in a 5% acrylamide/7 M urea gel (21). Cell lines contained omni-tagged NF90a or NF90b (lanes 2 and 3) or empty vector (lane 1). Note that the cell extracts contained more NF90b than NF90a. (**B**) The most stable snaR-A and -B structures predicted by MFOLD (27). Base-pairing is represented by dots. snaR-B bases altered or deleted (open triangle) in snaR-C are circled. (**C**) Northern blot of supernatant ×10^{-3} (Sup) and immunoprecipitated (IP) RNA from the cell lines probed with Probe-A (Figure S2A). 3′-pCp-labeled RNA precipitated with NF90a served as a marker (M). (**D**) RNA immunoprecipitated with NF90b was 3′-end labeled, digested with RNase H in the presence of sense (S) or antisense (AS) oligonucleotides corresponding to snaR (probe H, Figure 2B) or 5S rRNA and resolved in a 5% acrylamide/7 M urea gel. Asterisk marks snaR-A RNase H digestion product, bromophenol blue (bpb) migration is denoted. (**E**) *In vitro* binding assay of T7 RNA polymerase transcribed snaR-A to equal amounts of GST fusion proteins (Figure S1) in the presence of 2000-fold molar excess yeast tRNA. NF90b [A458P,A588P] mutant is denoted by 'GST-Mut' and 20% input was loaded.

TTCCATGAGGGGTGACAGAGG-3′) and R45 (5′-CCTGTGCGGCCGCTTCACTCCTGAGTTTCCATGC-3′). The PCR product was digested with EcoRI and NotI and ligated into similarly digested pGEX4T-3 vector (Amersham Biosciences).

To remove extraneous 3′ and 5′ RT-PCR priming sites from snaR-A sequence, pCR2.1-Clone 3 was PCR amplified using primers F-A (5′-GAGGTCTAGATTGGAGCCATTGTGGCTC-3′) and R-A (5′-GAGGAAGCTTCCGACCCATGTGGACCA-3′). The PCR

A

snaR-A

```
             5'------------------------------------------------------------------------------------------------------->3' nt.
Clone12m   -CTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone19m   -CTGGAGCCATTGTGGCTCAGGCCGGTTGCCCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCAAGGCCAGCCTGGTCCACATGGGTCG- 103
Clone20m   -CTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone12    -TTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGAGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone13    TCTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGAGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 105
Clone15i   --TGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 103
Clone3     -TTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone16a   -TTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone2     -TTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone17    -TTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone18    -TCGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGCCCAGCCTGGTCCACATGGGTCGG 104
Clone15    -TCGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone8a    -TGGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 104
Clone16    -----AGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 100
Clone1     --AGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG 103
             ************************* ***************************** **************** ** ************************

Consensus  TTGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGG-AGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGG
```

snaR-B

```
Clone21m   -----AGCCATTGTGGCTCCGGCCGGTTGCGCGGGCCCTCGGACTCTCAGGGTAGGCGAGGGTTCGAGGGCACGAGTTCGAGGCCAACCTGGTCCACATGGGTTG 100
Clone18a   -TTGGAGCCATTGTGGCTCCGGCCGGTTGCGCGGGCCCTCGGACTCTCAGGGTAGGCGAGGGTTCGAGGGCACGAGTTCGAGGCCAACCTGGTCCACATGGGTC- 103
Clone7m    -ATGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGGCCCTCGGACCCTCAGGGTAGGCGAGGGTTCGAGGGCACGAGTTCGAGGCCAACCTGGTCCACATGGGTTG 104
             ************** *********** ********* ***************** *********************************

Consensus  TGGAGCCATTGTGGCTCCGGCCGGTTGCGCGGGCCCTCGGACTCTCAGGGTAGGCGAGGGTTCGAGGGCACGAGTTCGAGGCCAACCTGGTCCACATGGGTTG
```

B

snaR-A 19q13.33:55287566:55287678

```
5'-CCCGGGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGGGCCCTCACGGAGGCGGGGGTTCCAGGGCACGAGTTCGAGGCCAGCCTGGTCCACATGGGTCGGAAAAAAGGATTTTTTTTA-3'
```

snaR-B 19q13.33:55328830:55328932

```
5'-CCCGGGGAGCCATTGTGGCTCCGGCCGGTTGCGCGGGCCCTCGGACCCTCAGGGTAGGCGAGGGTTCGAGGGCACGAGTTCGAGGCCAACCTGGTCCACATGGGTTGAAAAAAAAAATTTTTTTTA-3'
```

C

```
piR-36011                                           5'-GGCCAGCCTGGTCCACAUGGGUCGGAA-3'
snaR-A     5'-UUGGAGCCAUUGUGGCUCAGGCCGGUUGCGCCUGCCCUCGGGCCCUCACGGAGGCGGGGGUUCCAGGGCACGAGUUCGAGGCCAGCCUGGUCCACAUGGGUCGGAAAAAAAGGAUUUUUU-3'
piR-36189                                       5'-GGGCACGAGUUCGAGGCCAGCCUGGUCCAU-3'
```

Figure 2. snaR RT-PCR clones. (**A**) Clones of snaR identified from sequencing of RT-PCR products (21) can be grouped into two subsets. Homology within each subset is denoted by an asterisk and non-homologous nucleotides are in bold font. Differences between subsets are denoted by gray shading. Consensus sequences are given below each set of clones. The majority of clones are derived from asynchronous NF90b cell line extract. Clones with 'm' or 'a' appended to their name were immunoprecipitated from NF90b G2/M phase extract or from NF90a extract, respectively. (**B**) The genomic sequence of snaR-A and -B. Genomic nucleotides matching consensus sequence are highlighted in yellow, those differing from consensus sequence are in green. 3'-Oligo(A) and oligo(T) tracts are denoted in red and blue, respectively. Dashed line denotes sequence complementary to probe H, solid line denotes predicted RNase H digestion product. (**C**) Sequence alignment of snaR-A with two potential piRNAs (bold). Alu RNA homology is denoted in yellow. Sequence homologous to the PolIII B box motif is underlined.

product was digested with XbaI and HindIII and inserted into similarly digested pCRII vector (Invitrogen), to create pCRII-snaR-A. T7 run-off transcription from linearized pCRII-snaR-A produces snaR-A with 21 extraneous 5'-flanking nucleotides. To utilize transcribed snaR-A as an accurate size marker, snaR-A was PCR amplified from the pCRII-snaR-A construct using primers F-A2 (5'-GGCAAAGCTTTAAAATCCTTTTTTCCGACCCATGTGGACCAGG-3') and R-A2 (5'-GGCAGGATCCTAATACGACTCACTATAGCCGGAGCCATTGTGGCTCAGG-3'). These primers place the snaR-A gene at position +2 and introduce a DraI restriction site after four terminal uridines. The PCR product was digested with BamHI and HindIII and inserted into similarly digested pUC18 vector (Invitrogen).

To construct pR1-R3, fragments of chromosome 19 encompassing a snaR-A gene (55299144, Table 1) were PCR amplified from genomic DNA using primers F1 (5'-GGCGTCCTCCCTTTATGTTTTGTCG-3') and either R1 (5'-GGAGCCATTGTGGCTCAGGC-3'), R2 (5'-CTGAACTCTCAGATACAGTTCCCC-3') or R3 (5'-CACTGCGCCAGCCTTGTTTCTCG-3'), then

inserted via TOPO TA Cloning technology (Invitrogen) into pCR2.1 vector (Invitrogen).

RNase H digestion

Immunoprecipitated RNA (~25% isolated) was 3'-end labeled as described above. Labeled RNA (13.4 µl final) was incubated at 70°C for 10 min in the presence of 100 µM oligonucleotide specific for snaR (sense, 5'-GGGCACGAGTTCGAGGCC-3' or antisense, 5'-GGCCTCGAACTCGTGCCC-3') or 5S rRNA (sense, 5'-AACGCGCCCGATCTCGTCTGA-3' or antisense, 5'-TCAGACGAGATCGGGCGCGTT-3'), then cooled. RNA was incubated with RNase H (5 U, New England BioLabs) in the supplied reaction buffer at 30°C for 3 h. Digestion was stopped with 16 µl 2× formamide loading buffer (95% formamide, 10 mM Tris-HCl, pH 8, 20 mM EDTA) and heating at 70°C for 5 min.

GST pull-down assay

GST protein extract was prepared as previously described (22). Extract was rocked at 4°C for 1 h with 6 µl GSH

Table 1. Subsets of human snaR genes

Subset	Chrom (Strand)[a]	Start[b]	% ID[c]	3′-Terminus	Flanking homology[d]
A	19 (+)	53113498	100	A_6GGACT_8	100:100
A	19 (+)	53118848	100	A_6GGAT_8	100:100
A	19 (+)	53129250	100	A_6GGACT_8	100:99
A	19 (−)	55287678	100	A_6GGAT_8	100:100
A	19 (−)	55293015	100	A_6GGAT_8	100:100
A	19 (−)	55296080	100	A_6GGAT_8	100:100
A	19 (−)	55299144	100	A_6GGAT_8	100:100
A	19 (−)	55302203	100	A_6GGAT_8	100:100
A	19 (−)	55307557	100	A_6GGAT_8	100:100
A	19 (−)	55312909	100	A_6GGAT_8	100:100
A	19 (−)	55318263	100	A_6GGAT_8	100:100
A	19 (−)	55323591	100	A_6GGAT_8	100:100
A	19 (+)	53102747	99.1	A_6GGACT_8	100:100
A	19 (+)	53139991	99.1	A_6GGACT_8	100:100
B	19 (−)	55328933	90.4	A_9T_7	97.3:96
B	19 (−)	55334306	90.4	A_9T_7	98:96
C	19 (+)	53145365	89.4	A_9T_7	97.3:95
C	19 (+)	53134610	88.5	A_9T_8	96.7:96
D	19 (−)	55335389	88.5	A_8T_7	46.7:96
E	19 (−)	52025799	87.5	$A_{15}T_2$	42.7:31
C	19 (+)	53108120	87.5	A_9T_8	97.3:97
C	19 (+)	53123887	87.5	A_9T_8	96.7:97
C	19 (+)	53150751	87.5	A_9T_8	97.3:96
F	19 (+)	55800032	86.5	$A_{10}T_8$	94:96
3	3 (+)	192078417	86.5	A_8T_8	93.3:96
2	2 (−)	78035660	86.5	$A_{10}T_7$	94.7:95
G2	19 (−)	54226855	83.7	A_7T_7	92:95
G1	19 (+)	54232089	80.8	$A_{14}T_7$	91.3:96

A UCSC BLAT search of the human genome (NCBI Build 36.1) for snaR-A consensus (Figure 2A). 'Hits' have ≥80% sequence identity to snaR-A consensus and were categorized into three subsets and seven outliers. Search included two 5′-terminus cytidines that are present as uridines in the consensus sequence.
[a]Chromosome (Chrom) and strand polarity is denoted.
[b]'Start' nucleotides are given with respect to the chromosomal numbering system.
[c]'% ID' is a percentage comparison of gene primary sequence to that of snaR-A consensus sequence, as determined by ClustalW alignment.
[d]Percentage comparison of 5′ (150 nt upstream) and 3′ (100 nt downstream) flanking sequence to that of snaR-A, as determined by ClustalW alignment.

beads in a final volume of 250 µl EBCD buffer (50 mM Tris, pH 8, 120 mM NaCl, 0.5% NP-40, 1 mM DTT). Beads were washed thrice for 2 min in 1 ml EBCD buffer and 0.075% SDS and twice for 2 min in 1 ml low salt buffer (50 mM Tris, pH 7.6, 50 mM NaCl) at 4°C with rocking. A third of the beads were set aside for protein analysis in a 7.5% SDS/PAGE gel. Beads were washed twice on ice in 1 ml Binding buffer (25 mM Hepes KOH, pH 7.4, 0.1 M KCl, 10 mM MgCl$_2$, 0.1 mM EDTA, 1 mM DTT, 10 µM PMSF, 0.1 mg/ml BSA, 0.1 mg/ml tRNA) + 0.01% NP-40, then once in 1 ml Binding buffer. Radiolabeled snaR-A (∼200 fmol) was incubated with the beads (≥ 12 pMol of immobilized GST protein) in 10 µl Binding buffer with 1 U/µl RNasin at 30°C for 5 min, then on ice for 25 min. Beads were washed thrice in 1 ml modified Binding buffer (with 175 mM KCl, 10 µg/ml tRNA, 0.01% NP-40) and bound snaR-A was eluted in 250 µl Elution buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.05% NP-40, 50 µg/ml tRNA, 0.5% SDS). RNA was phenol extracted (200 µl) and precipitated in EtOH in the presence of 20 µg/ml glycogen and 1 M ammonium acetate.

Northern blots

Apart from fetal kidney RNA (Clontech), all human tissue RNA was purchased from Ambion. RNA was resolved in polyacrylamide/7 M urea gels and transferred in 0.5× TBE to GeneScreen Plus nylon membrane (PerkinElmer Life Sciences) by electroblotting. Membranes were blocked in ULTRAhyb-Oligo hybridization buffer (Ambion) for 1 h at 37°C, before hybridization at 37°C for 16 h with 5′-end labeled oligonucleotide. Blocking and hybridization of labeled RNA complementary to clone 3 was carried out at 42°C. Membranes were washed twice in 2× SSC with 0.1% SDS for 10 min at 25°C. All oligonucleotide probes were 5′-end labeled with ^{32}P-γ-ATP using T4 polynucleotide kinase (New England BioLabs) and purified over Micro Bio-Spin P-30 chromatography columns (Bio-Rad). Oligonucleotides used in this study: Probe A 5′-GACCCATGTGGACCAGGCT GGCCTCGAACT-3′, 5.8S rRNA 5′-CGCAAGTGCGT TCGAAGTGTC-3′.

In vitro transcription

Genomic DNA from 293 cells was prepared as previously described (25). *In vitro* transcription of pVAII (26), linearized CMV DNA, and pR1-R3 was performed using the HeLa Cell Extract Transcription System (Promega). T7 and SP6 RNA polymerase run-off transcription from linearized pCRII-snaR-A yielded radiolabeled sense and antisense snaR-A, respectively.
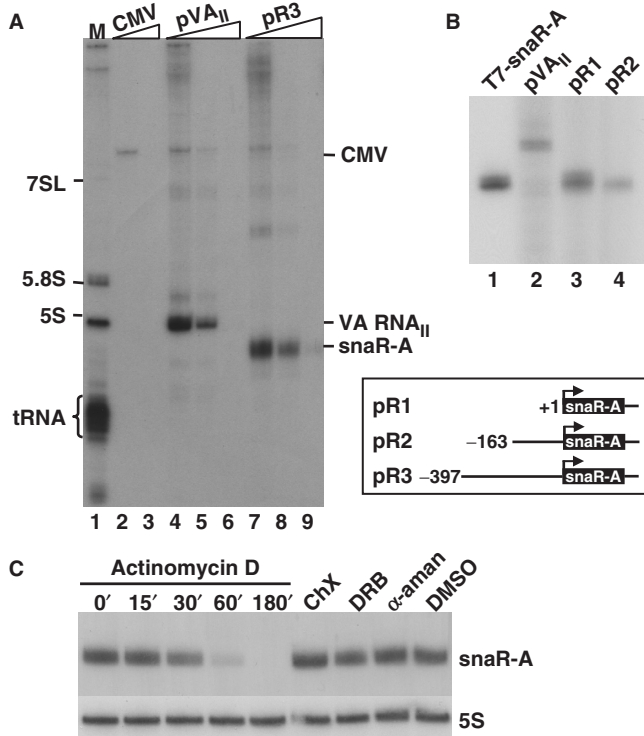
Figure 3. snaR-A is a rapidly synthesized PolIII transcript. (**A**) *In vitro* transcription of linearized DNA containing the CMV promoter, adenovirus VA RNA$_{II}$ (pVA$_{II}$), and snaR-A from pR3 with HeLa cell nuclear extract in the presence of 0, 20 or 200 μg/ml α-amanitin. M: 3′-end labeled cytoplasmic RNA. (**B**) *In vitro* transcription in HeLa cell nuclear extract of snaR-A from pR1 and pR2, and by T7 RNA polymerase from linearized pT7-snaR-A. A schematic of pR1-R3 constructs (boxed) gives upstream sequence (numbered) with respect to the *snaR-A* transcriptional start site (bent arrow). (**C**) Northern blot of total RNA from HeLa S3 cells after exposure to 1 μg/ml actinomycin D for 15, 30, 60 or 180 min or to 100 μg/ml cycloheximide (ChX), 10 μg/ml α-amanitin, 10 μM DRB or DMSO for 180 min. Blot was probed with Probe-A (upper) and 5S rRNA antisense oligonucleotide (lower).

RESULTS

A novel family of structured RNAs

To characterize RNAs bound to NF90 *in vivo*, stable 293 cell lines expressing epitope-tagged isoforms of NF90, NF90a and NF90b, were subjected to formaldehyde cross-linking (21). The a and b isoforms differ only in the absence or presence of a four amino acid sequence whose function is unknown (1). The most prominent RNA species immunoprecipitated with NF90 were ~80–120 nt long (Figure 1A), and ~12 RNAs were identified by cloning and sequencing (21). Several related clones of ~104 nt were found, which fell into two subsets: a major subset, snaR-A, and a minor subset, snaR-B (Figure 2A). The subsets are ~90% identical in sequence. Human genomic sequences corresponding to snaR-A and snaR-B had occasional mismatches that are possibly due to C-to-U editing of the RNA (Figure 2B). In the genome, the cloned snaR sequences are followed by a run of adenines that accounts for the cloning of these RNAs by RT-PCR with an oligo(dT) primer, and then by a run of thymidines that constitutes a potential PolIII termination

site. Placing the 3′ terminus within the oligo-(U) run gives a length of ~117 nt, consistent with the RNAs' gel mobility.

The snaRs lack an open reading frame and are predicted to fold into thermodynamically stable hairpin structures (Figure 1B). The structures illustrated have calculated ΔG of folding of −56 kcal/mol and >60% G:C base-pairing, but several alternative structures have similar stability (27). The dynamic secondary structure prediction program paRNAss (28) predicts that snaR-A can readily undergo a conformational switch which might account for mobility shifts seen in higher percentage acrylamide gels (see subsequently).

snaR is a major NF90 binding partner

A single band was detected when 293 cell RNA was examined by northern blotting with a DNA oligonucleotide complementary to snaR-A. As expected, this band was enriched in the NF90 immunoprecipitate (Figure 1C, lanes 5–7) and co-migrated with the major end-labeled RNA band that was immunoprecipitated with NF90 (lane 1). To evaluate the amount of snaR-A in this gel band, NF90-associated RNA was 3′-end labeled and incubated with RNase H in the presence of an oligonucleotide complementary to snaR-A and -B. The labeled snaR band was completely digested and replaced by the predicted end-labeled fragment of ~35 nt (Figure 1D, lane 4). No digestion resulted from incubation in the presence of an oligonucleotide with a sequence *identical* to that of snaR-A (lane 3). A faint band corresponding to 5S rRNA was identified in a similar fashion (lanes 1 and 2). These results identify snaR-A as one of the principal RNAs cross-linked to NF90 *in vivo*. We estimate that there are ~70 000 molecules of snaR-A in a 293 cell (data not shown), compared to ~200 000 copies of 7SK RNA (29).

Structured RNAs such as VA RNA$_{II}$ interact with NF90 via its dsRBMs (5). To determine whether functional dsRBMs are required for the NF90/snaR interaction, we examined the ability of NF90 to bind snaR-A *in vitro*. In a pulldown assay, radiolabeled snaR-A bound to GST fusions with all three NF90 isoforms, NF90a–c, but not to GST itself or GST-NF45 (Figure 1E). Binding was not detected to the NF90b [A458P,A588P] mutant which contains inactivating mutations in both of its dsRBMs. Hence, NF90 binds directly to snaR-A via its dsRBMs, consistent with the high degree of secondary structure predicted for the RNA.

PolIII transcription and rapid turnover of snaR-A

Nearly all of the snaR genes contain a 3′-terminal oligo (dT) run that could serve as a PolIII termination signal (Table 1). They also exhibit homology with the PolIII B box motif (Figure 2C, see subsequently). To determine which RNA polymerase transcribes the snaR genes, we transcribed a chromosomal fragment encompassing a snaR-A gene in HeLa nuclear extract. snaR-A synthesis persisted in the presence of 20 μg/ml α-amanitin, a concentration that reduces PolIII activity by ~50% but completely inhibits PolII (30), but was abrogated in the

presence of 200 μg/ml of the toxin (Figure 3A, lanes 7–9). The adenovirus-2 VA RNA$_{II}$ gene, which is transcribed by PolIII, exhibited a similar sensitivity (lanes 4–6), whereas PolII transcription driven by the CMV immediate-early promoter was abrogated at 20 μg/ml α-amanitin (lanes 2 and 3). snaR-A was also generated by constructs containing less or no chromosomal sequence upstream of the snaR-A gene (Figure 3B, lanes 3 and 4). The *in vitro* product had the expected RNase H sensitivity (data not shown) and the same gel mobility as a T7 RNA polymerase run-off transcript of full-length snaR-A (lane 1). We conclude that the snaR-A gene is transcribed by PolIII from an intragenic promoter.

To determine the stability of snaR-A, we monitored its disappearance in HeLa cells treated with actinomycin D to inhibit transcription by all polymerases. Northern blotting showed that snaR-A decays rapidly, with a half-life of ~15 min (Figure 3C). Inhibitors of PolII transcription and translation had little effect. This short half-life implies a rapid turnover rate and the active synthesis of snaR-A in cells.

Restricted snaR expression in human tissues and cell lines

To evaluate the distribution of snaR-A, we probed a northern blot of total RNA from 19 adult human tissues. High snaR-A expression was seen exclusively in testis (Figure 4A, upper panel). Long exposures revealed weak expression in brain and placenta. Remarkably, no expression was detected in adult kidney or in human fetal kidney tissue (not shown) even though snaR was initially discovered in 293 cells which are adenovirus-transformed cells derived from human embryonic kidney (31) and express the RNA at high levels (see below). RNA integrity and loading were assured by probing for 5.8S rRNA (lower panel). Furthermore, snaR-A was not detected in RNA from progenitor or adipocyte-differentiated mesenchymal stem cells, or from unstimulated or stimulated peripheral blood mononuclear cells (data not shown).

Considering the possibility that snaR expression is elevated in tumor cells, we compared RNA from a testis tumor with that from normal adjacent testis tissue. On the contrary, reduced expression was seen in the tumor tissue compared to normal adjacent tissue (Figure 4A, right panel). We next examined total RNA from 16 permanent cell lines to determine whether snaR-A is generally expressed in immortalized cells. Two major bands, possibly representing conformers or other snaR species (see subsequently), were detected in all cell lines tested except Colo205 where only the faster band was seen (Figure 4B, upper panel). snaR-A was expressed highly in 293 cells, at intermediate levels in HeLa, HepG2 and some other lines, and at lower levels in lines such as Jurkat and Colo205. Thus, snaR-A is tightly controlled in human tissues, where it is largely restricted to testis, but it is dysregulated in many virally transformed and tumor-derived cell lines.

With very long exposures, short RNAs of ~26–34 nt were detected in testis and 293 cells (Figure 4C, asterisk). Database search disclosed the existence in a piRNA
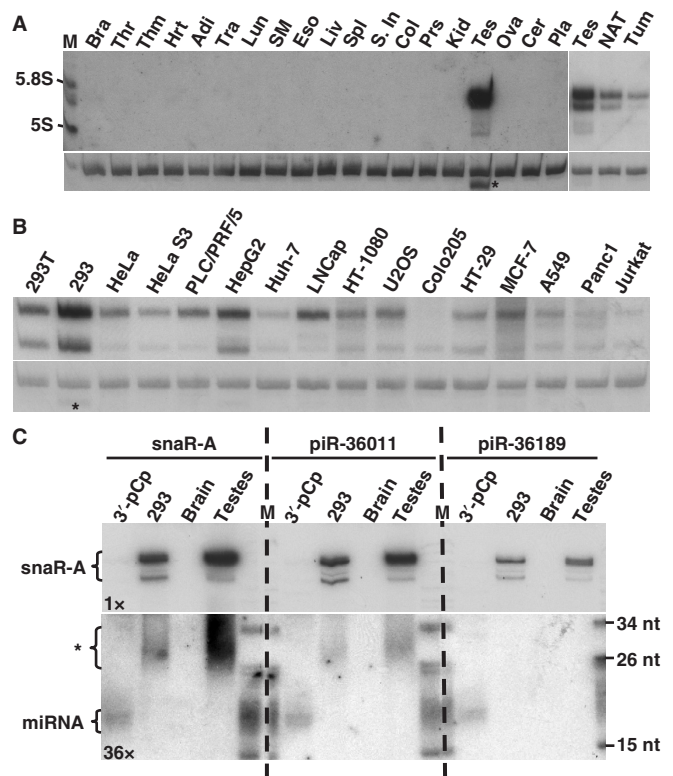

Figure 4. snaR is highly expressed in human testis and cell lines. (**A**) Northern blots of total RNA extracted from normal adult human tissue (left) or testis tumor (Tum) and normal adjacent tissue (NAT, right), probed with Probe-A (upper) or for 5.8S rRNA (lower). Tissues tested were (from left): brain, thyroid, thymus, heart, adipose, trachea, lung, skeletal muscle, esophagus, liver, spleen, small intestine, colon, prostate, kidney, testes, ovary, cervix and placenta. (**B**) Northern blot of cell lines with radiolabeled RNA complementary to full-length snaR-A (upper) or 5.8S rRNA antisense oligonucleotide (lower). Asterisks denote residual snaR-A bands. (**C**) Northern blot of 293 cell, brain and testis RNA, probed with Probe-A or oligonucleotides specific for piR-36011 or piR-36189. 3′-end labeled cytoplasmic RNA and MspI-digested pBR322 DNA (M) served as size markers. The blot was cut into strips before hybridization as shown by dashed lines. The lower panel was exposed 36-fold longer than the upper panel. Asterisk marks the low molecular weight snaR-related bands.

library (32) of two sequences that correspond to snaR-A (Figure 2C). piRNAs are recently discovered germline-specific microRNAs defined as binding to PIWI protein (33). The RNAs piR-36011 and piR-36189 are 27 and 30 nt in length, respectively, and their sequences overlap (32). Interestingly, piR-36189, and hence snaR-A, contain homology to a region of Alu RNA which possesses the consensus sequence of the PolIII B box motif (Figure 2C, (34)). Although it is not known whether these two particular RNAs bind to PIWI, we attempted to detect them in 293 cell, testis and brain RNA. Oligonucleotide probes complementary to the two piRNAs recognized intact snaR-A in 293 and testis RNA. The piR-36011 probe and snaR-A probe (Probe A), but not the piR-36189 probe, gave faint hybridization in the 26–34 nt region (Figure 4C). Another PolIII transcript, adenovirus VA RNA$_I$, was recently found to be processed by Dicer, albeit inefficiently (35). However, the faint bands detected
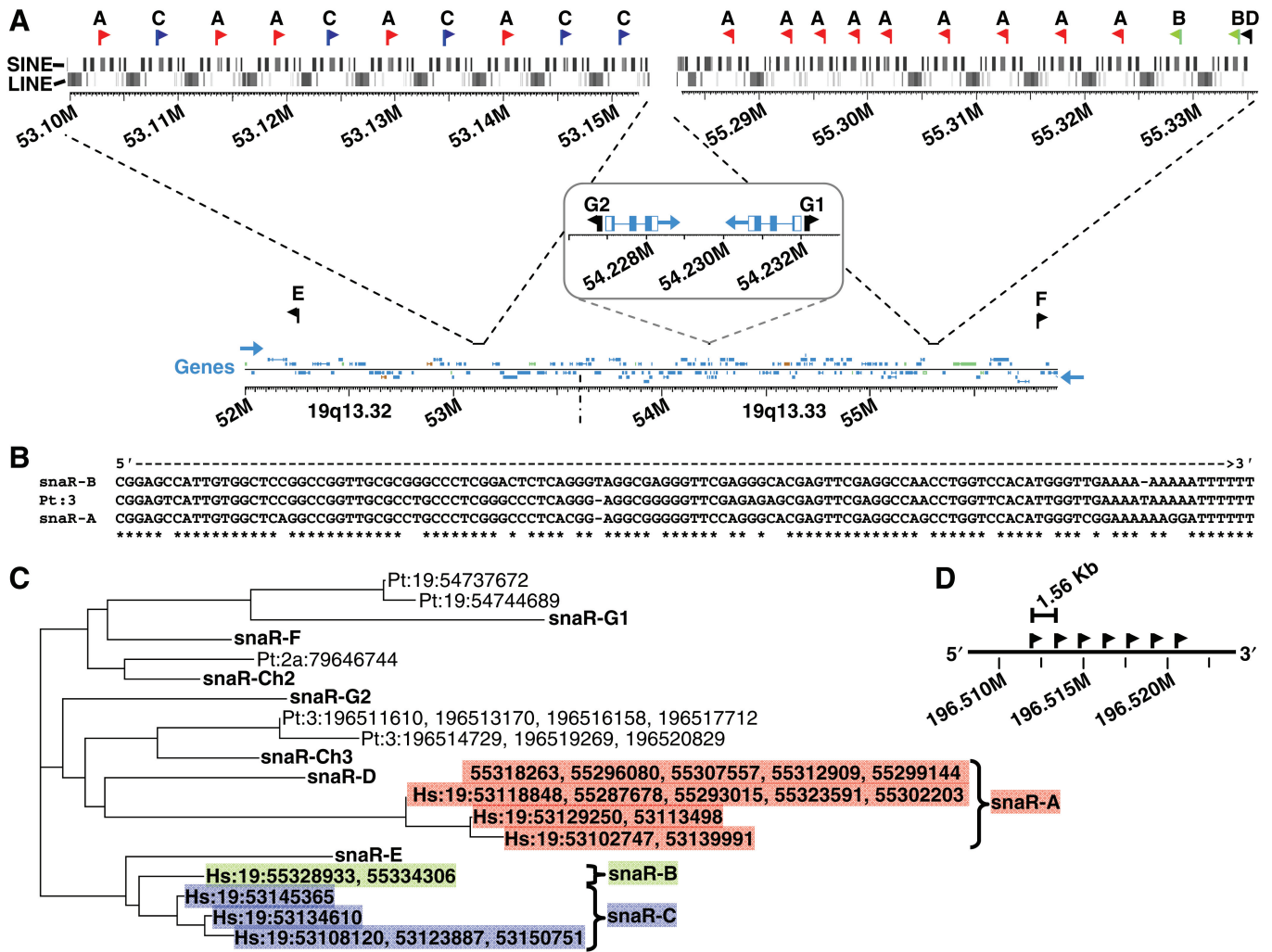
Figure 5. snaR gene organization and evolution. (**A**) Two regions of human chromosome 19q13.32-33 containing snaR gene clusters are expanded. The genomic region containing two chorionic gonadotropin beta polypeptide (CGβ and snaR genes is boxed. snaR-A, -B and -C genes, labeled to indicate their subsets, are represented by red, green and blue bars, respectively. Five outlier genes are represented by black bars. Coding gene exons are shown as cyan rectangles. The direction of snaR transcription is denoted by arrow heads, and that of protein-coding genes by cyan arrows. SINE and LINE elements predicted by the UCSC genome browser (49) are shown as vertical gray bars. Distances from one end of the chromosome are in megabases (M). (**B**) Alignment of human snaR-A and -B with chimpanzee chromosome 3 snaR consensus. Asterisks indicate nucleotide identity, gray shading indicates heterogenous nucleotides. (**C**) Phylogram of human (Hs) and chimpanzee (Pt) snaR, derived from a ClustalW sequence alignment of genes found from UCSC BLAT searches (50) of snaR-A and -B against the human genome (NCBI Build 36.1) and chimpanzee genome (UCSC version PanTro2). snaR genes are denoted by species:chromosome:start nucleotide labels. snaR-A, -B and -C subsets are highlighted in red, green and blue, respectively. (**D**) Region of chimpanzee chromosome 3 (196510-196523 Kb) showing the *snaR* cluster.

here are more likely to result from snaR-A degradation, since they are diffuse and lack a 5′ terminal uracil believed to be a hallmark of piRNA processing (36).

Multiple divergent human snaR genes

A bioinformatic search of the human genome identified 28 snaR genes, including 14 genes for snaR-A and 2 genes for snaR-B (Table 1), all located on the q-arm of chromosome 19. In addition, the search revealed a further subset, snaR-C (5 genes), and 7 snaR genes that defy classification into subsets. Although the snaR-C genes display slight sequence variations (Figure S2A), this subset is closely related to snaR-B (95–97% identity; Figure 1B).

The outliers, snaR-D to -G, include 5 unique genes on chromosome 19 as well as single copies (snaR-2 and -3) on chromosomes 2 and 3 (Table 1).

Genes in the three snaR subsets are located in two clusters within chromosomal bands 19q13.32 and 19q13.33 (Figure 5A). The clusters are transcribed in opposite directions and appear to have arisen from an inversion of a region of segmental duplication. The cluster on band 19q13.32 has 5 snaR-A genes interspersed with the 5 snaR-C genes, while the remaining 9 snaR-A genes and 2 tandem snaR-B genes are on band 19q13.33. Most of these snaR genes lie in a 5.3 kb tandem repeat, surrounded by multiple repetitive elements such as SINEs and LINEs (A.F.A. Smit, R. Hubley and

P. Green RepeatMasker at http://repeatmasker.org). Each repeat contains a large LINE L1 element of ~1.5 kb which is deleted in three repeat units on band 19q13.33 that appear to have been truncated by insertion of a SINE Alu-Sx element.

Most of the outliers share flanking sequence homology with the snaR genes in the two clusters (Table 1). The snaR-D, E and F genes are located at varying distances outside the two clusters on chromosome 19, while snaR-G1 and -G2 are located in the 2.1 Mb genomic fragment between the clusters. This region contains many protein-coding genes. Remarkably, these snaR genes lie within the promoter region of two genes encoding the β polypeptide of chorionic gonadotropin (CGβ; Figure 5A).

snaR is specific to hominids

snaR genes are present in chimpanzees (Figure 5B), but were not found in searches of rhesus macaque or other mammalian genomes. Despite a comparable genome size, the chimpanzee has fewer snaR genes (10 instead of 28). As in humans, 7 of the chimpanzee snaR genes have expanded in a tandem repeat. However, the repeat differs from those in humans in size and location (~1.5 kb on chromosome 3; Figure 5D). The repeated chimpanzee genes are ~94% identical with the solitary human snaR gene on chromosome 3 (Figure S2B), compared to ~88% identity with human snaR-A and –B (Figure 5B). This suggests the expansion of an orthologous gene on chromosome 3 has taken place in chimpanzee but not in humans.

The relationships of the human and chimpanzee genes were further illuminated by phylogenetic analysis (Figure 5C). This confirms that the human and chimpanzee genes on chromosome 3 are closely related. Similarly, the single snaR gene on chimpanzee chromosome 2a is 94% identical with its ortholog on human chromosome 2 (Figure S2B). Furthermore, two genes on chimpanzee chromosome 19 are orthologous with snaR-G1, a unique gene on human chromosome 19 (Figure 5C, top branch). Interestingly, their common ancestry is strengthened by the conserved position of these snaR genes relative to the primate-specific CGβ genes (37). snaR-G1 and its closest chimpanzee ortholog are both located 86 nt upstream of the 5′ end of the CGβ1 gene, which is transcribed in the opposite direction (Figure 5A, gray box). snaR-G2 mirrors this localization, being 85 nt from the 5′ end of human CGβ2, although it has diverged considerably from snaR-G1 and the chimpanzee genes (Figure 5C).

These observations suggest that the snaR genes originated after the divergence of the Great Apes from other primates, and have expanded differentially, sometimes in tandem repeats, since the divergence between humans and chimpanzees 5–7 MYA (38).

DISCUSSION

The snaRs are a hitherto unrecognized family of small structured RNAs that are bound *in vivo* to NF90. These RNAs turn over rapidly, display a restricted distribution in human tissues, and appear to have evolved recently in primates. Accelerated evolution of the snaR genes is implicit in their restriction to great apes and their different gene copy numbers, location and sequences between human and chimpanzee. Apart from intra-subset comparisons, snaR genes display a lower degree of homology within the gene than in their flanking sequence (Figure S2C). This suggests rapid selection and then conservation of mutations within the loci, presumably due to divergent function among paralogs. Another rapidly evolving gene, HAR1F, which specifies a structured non-coding RNA similar in size to snaR, was recently associated with the emergence of essentially human characteristics including the development of the brain neocortex (39). Strikingly, like the snaRs, this RNA is expressed in testis as well as in brain.

NF90 is highly expressed in the testis (10), and at least three other double-stranded RNA-binding proteins—spermatid perinuclear RNA-binding protein, protamine 1 mRNA-binding protein, and testis nuclear RNA-binding protein—play critical roles in mammalian spermatogenesis [reviewed in (40,41)]. Further work is required to elucidate why members of this family of RNAs bind to NF90 protein. The apical stem-loop appears to be a common feature of snaRs. The stem, which is structurally stable due to extensive canonical Watson–Crick base-pairing, is likely to interact with the dsRBMs of NF90. On the other hand, the apical loop is variable in size and sequence and could play a part in determining the specific functions or targets of these RNAs.

With the exception of snaR-D and -E, all human and chimpanzee snaRs share sequence identity upstream (~150 bp) and downstream (~100 bp) of their genes (Figure S2C). This implies that the majority of snaRs have 'piggy-backed' on a larger segment of DNA which has generated segmental duplications in both hominids. However, snaRs also display features characteristic of retrotransposons: they have an internal PolIII promoter, a stable hairpin loop structure, and a 3′-oligo(A) tract followed by an oligo(T) tract (42). Three observations are consistent with their limited retrotransposition. First, snaR-E resembles snaR-B (Table 1) but lacks the extended 3′ oligo(T) tract (transcription termination site) and surrounding sequence identity with other snaR genes. Second, snaR-D, which is most closely related to snaR-A, shares no 5′-flanking sequence with other snaRs. Third, a MUC18 cDNA clone (drop4.7) has been reported that contains a non-genic 92 bp 3′-terminus (43): this sequence is identical to a slightly truncated version of snaR-C.

Transposons and segmental duplications are thought to be a rich source of rapid molecular evolution in primates, giving rise to new gene families often involved in reproduction and immunity (44). While the functions of the snaRs remain to be established, there is a strong association with the reproductive system. In addition to their predominant expression in testis, the locations of the snaR genes are suggestive. The genomic region between the two snaR clusters on 19q13 has the highest gene density on the chromosome [with ~110 known or predicted genes in 2.1 megabases (Table S1), compared with an average of ~60], and many of these genes are

integral to reproduction. Most pertinently, the snaR-G genes are closely linked to the most recent members of the beta subunit of chorionic gonadotropin glycoprotein hormone, where they overlap predicted binding sites for CGβ transcription factors that govern early placental development and implantation (45). The CGβ polypeptides evolved in primates and have expanded in the great apes (37,45). The most recent additions to the family, CGβ1 and CGβ2, arose in the common ancestor of African great apes (45). In addition to the critical role of CGβ in the establishment of pregnancy, its overexpression is diagnostic for certain testicular cancers (46). Indeed, overexpression of human CGβ in transgenic male mice leads to defective reproductive organs (47) and fetal Leydig cell adenomas (48). We speculate that snaRs modulate the expression of CGβ polypeptides and possibly other genes in between the *snaR* clusters through an epigenetic mechanism.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Reichman,T.W. and Mathews,M.B. (2003) NF90 family of double-stranded RNA-binding proteins: regulators of viral and cellular function. In Bradshaw,R. and Dennis,E. (eds), *Cell Signaling Handbook*, Academic Press, Elsevier Science, USA, pp. 335–342.
2. Larcher,J.C., Gasmi,L., Viranaicken,W., Edde,B., Bernard,R., Ginzburg,I. and Denoulet,P. (2004) Ilf3 and NF90 associate with the axonal targeting element of Tau mRNA. *FASEB J.*, **18**, 1761–1763.
3. Shi,L., Zhao,G., Qiu,D., Godfrey,W.R., Vogel,H., Rando,T.A., Hu,H. and Kao,P.N. (2005) NF90 regulates cell cycle exit and terminal myogenic differentiation by direct binding to the 3′-untranslated region of MyoD and p21WAF1/CIP1 mRNAs. *J. Biol. Chem.*, **280**, 18981–18989.
4. Matsumoto-Taniura,N., Pirollet,F., Monroe,R., Gerace,L. and Westendorf,J.M. (1996) Identification of novel M phase phosphoproteins by expression cloning. *Mol. Biol. Cell*, **7**, 1455–1469.
5. Liao,H.J., Kobayashi,R. and Mathews,M.B. (1998) Activities of adenovirus virus-associated RNAs: purification and characterization of RNA binding proteins. *Proc. Natl Acad. Sci. USA*, **95**, 8514–8519.
6. Isken,O., Grassmann,C.W., Sarisky,R.T., Kann,M., Zhang,S., Grosse,F., Kao,P.N. and Behrens,S.E. (2003) Members of the NF90/NFAR protein group are involved in the life cycle of a positive-strand RNA virus. *EMBO J.*, **22**, 5655–5665.
7. Kao,P.N., Chen,L., Brock,G., Ng,J., Kenny,J., Smith,A.J. and Corthesy,B. (1994) Cloning and expression of cyclosporin A- and

8. FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. *J. Biol. Chem.*, **269**, 20691–20699.
8. Reichman,T.W., Parrott,A.M., Fierro-Monti,I., Caron,D.J., Kao,P.N., Lee,C.G., Li,H. and Mathews,M.B. (2003) Selective regulation of gene expression by nuclear factor 110, a member of the NF90 family of double-stranded RNA-binding proteins. *J. Mol. Biol.*, **332**, 85–98.
9. Patel,R.C., Vestal,D.J., Xu,Z., Bandyopadhyay,S., Guo,W., Erme,S.M., Williams,B.R. and Sen,G.C. (1999) DRBP76, a double-stranded RNA-binding nuclear protein, is phosphorylated by the interferon-induced protein kinase, PKR. *J. Biol. Chem.*, **274**, 20432–20437.
10. 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.
11. Buaas,F.W., Lee,K., Edelhoff,S., Disteche,C. and Braun,R.E. (1999) Cloning and characterization of the mouse interleukin enhancer binding factor 3 (Ilf3) homolog in a screen for RNA binding proteins. *Mamm. Genome*, **10**, 451–456.
12. Xu,Y.H., Leonova,T. and Grabowski,G.A. (2003) Cell cycle dependent intracellular distribution of two spliced isoforms of TCP/ILF3 proteins. *Mol. Gen. Metab.*, **80**, 426–436.
13. 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.
14. Langland,J.O., Kao,P.N. and Jacobs,B.L. (1999) Nuclear factor-90 of activated T-cells: a double-stranded RNA-binding protein and substrate for the double-stranded RNA-dependent protein kinase, PKR. *Biochemistry*, **38**, 6361–6368.
15. Tran,H., Schilling,M., Wirbelauer,C., Hess,D. and Nagamine,Y. (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol. Cell*, **13**, 101–111.
16. Xu,Y.H. and Grabowski,G.A. (2005) Translation modulation of acid beta-glucosidase in HepG2 cells: participation of the PKC pathway. *Mol. Gen. Met.*, **84**, 252–264.
17. Bose,S.K., Sengupta,T.K., Bandyopadhyay,S. and Spicer,E.K. (2006) Identification of Ebp1 as a component of cytoplasmic bcl-2 mRNP (messenger ribonucleoprotein particle) complexes. *Biochem. J.*, **396**, 99–107.
18. Shim,J., Lim,H., Yates,R.J. and Karin,M. (2002) Nuclear export of NF90 is required for interleukin-2 mRNA stabilization. *Mol. Cell*, **10**, 1331–1344.
19. Shin,H.J., Kim,S.S., Cho,Y.H., Lee,S.G. and Rho,H.M. (2002) Host cell proteins binding to the encapsidation signal epsilon in hepatitis B virus RNA. *Arch. Virol.*, **147**, 471–491.
20. Parrott,A.M., Walsh,M.R., Reichman,T.W. and Mathews,M.B. (2005) RNA binding and phosphorylation determine the intracellular distribution of nuclear factors 90 and 110. *J. Mol. Biol.*, **348**, 281–293.
21. 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. In Lorsch,J.R. (ed), *Methods in Enzymology: Translation Initiation: Extract Systems and Molecular Genetics*. Elsevier Inc., San Diego, CA, Vol. 429, pp. Inpress.
22. Parker,L.M., Fierro-Monti,I. and Mathews,M.B. (2001) Nuclear factor 90 is a substrate and regulator of the eukaryotic initiation factor 2 kinase double-stranded RNA-activated protein kinase. *J. Biol. Chem.*, **276**, 32522–32530.
23. Reichman,T.W. and Mathews,M.B. (2003) RNA binding and intramolecular interactions modulate the regulation of gene expression by nuclear factor 110. *RNA*, **9**, 543–554.
24. Reichman,T.W., Muniz,L.C. and Mathews,M.B. (2002) The RNA binding protein nuclear factor 90 functions as both a positive and negative regulator of gene expression in mammalian cells. *Mol. Cell. Biol.*, **22**, 343–356.
25. 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*, John Wiley & Sons, Inc., New York, Vol. I, pp. 2.2.1–2.2.3.

26. Ma,Y. and Mathews,M.B. (1993) Comparative analysis of the structure and function of adenovirus virus-associated RNAs. *J. Virol.*, **67**, 6605–6617.

27. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

28. Giegerich,R., Haase,D. and Rehmsmeier,M. (1999) Prediction and visualization of structural switches in RNA. *Pac. Symp. Biocomput.*, 126–137.

29. Wassarman,D.A. and Steitz,J.A. (1991) Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell. Biol.*, **11**, 3432–3445.

30. Weinmann,R. and Roeder,R.G. (1974) Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. *Proc. Natl Acad. Sci. USA*, **71**, 1790–1794.

31. Graham,F.L., Smiley,J., Russell,W.C. and Nairn,R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.*, **36**, 59–74.

32. Girard,A., Sachidanandam,R., Hannon,G.J. and Carmell,M.A. (2006) A germline-specific class of small RNAs binds mammalian piwi proteins. *Nature*, **442**, 199–202.

33. Kim,V.N. (2006) Small RNAs just got bigger: piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev.*, **20**, 1993–1997.

34. Shankar,R., Grover,D., Brahmachari,S.K. and Mukerji,M. (2004) Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependant mobile Alu elements. *BMC Evol. Biol.*, **4**, 37.

35. Sano,M., Kato,Y. and Taira,K. (2006) Sequence-specific interference by small RNAs derived from adenovirus VAI RNA. *FEBS Letts.*, **580**, 1553–1564.

36. Brennecke,J., Aravin,A.A., Stark,A., Dus,M., Kellis,M., Sachidanandam,R. and Hannon,G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, **128**, 1089–1103.

37. Maston,G.A. and Ruvolo,M. (2002) Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol. Biol. Evol.*, **19**, 320–335.

38. Kumar,S., Filipski,A., Swarna,V., Walker,A. and Hedges,S.B. (2005) Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc. Natl Acad. Sci. USA*, **102**, 18842–18847.

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

40. Saunders,L.R. and Barber,G.N. (2003) The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB J.*, **17**, 961–983.

41. Tian,B., Bevilacqua,P.C., Diegelman-Parente,A. and Mathews,M.B. (2004) The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.*, **5**, 1013–1023.

42. 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.

43. Lehmann,J.M., Riethmuller,G. and Johnson,J.P. (1989) MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily. *Proc. Natl Acad. Sci. USA*, **86**, 9891–9895.

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

45. Hallast,P., Rull,K. and Laan,M. (2007) The evolution and genomic landscape of CGB1 and CGB2 genes. *Mol. Cell. Endo.*, **260–262**, 2–11.

46. Skotheim,R.I. and Lothe,R.A. (2003) The testicular germ cell tumour genome. *APMIS*, **111**, 136–150.

47. Rulli,S.B., Ahtiainen,P., Makela,S., Toppari,J., Poutanen,M. and Huhtaniemi,I. (2003) Elevated steroidogenesis, defective reproductive organs, and infertility in transgenic male mice overexpressing human chorionic gonadotropin. *Endocrinology*, **144**, 4980–4990.

48. Ahtiainen,P., Rulli,S.B., Shariatmadari,R., Pelliniemi,L.J., Toppari,J., Poutanen,M. and Huhtaniemi,I.T. (2005) Fetal but not adult Leydig cells are susceptible to adenoma formation in response to persistently high hCG level: a study on hCG overexpressing transgenic mice. *Oncogene*, **24**, 7301–7309.

49. 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.

50. Kent,W.J. (2002) BLAT-the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.