Supplementary Materials

Identification of Truncated Forms of U1 snRNA Reveals a Novel RNA Degradation Pathway during snRNP Biogenesis

Hideaki Ishikawa[1, 2], Yuko Nobe[2, 3], Keiichi Izumikawa[1, 2], Harunori Yoshikawa[2, 4], Naoki Miyazawa[2, 4], Goro Terukina[2, 4], Natsuki Kurokawa[1, 4], Masato Taoka[2, 3], Yoshio Yamauchi[3], Hiroshi Nakayama[2, 5], Toshiaki Isobe[2, 3], and Nobuhiro Takahashi[1, 2, 4,*]

1 Graduate School of Agriculture, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai-cho, Fuchu-shi, Tokyo 183-8509, Japan

2 Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Sanbancho 5, Chiyoda-ku, Tokyo 102-0075, Japan

3 Department of Chemistry, Graduate School of Sciences and Engineering, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachiouji-shi, Tokyo 192-0397, Japan

4 Department of Bioengineering, United Graduate School of Agriculture, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai-cho, Fuchu-shi, Tokyo 183-8509, Japan

5 Biomolecular Characterization Team, RIKEN Advanced Science Institute, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan

*Correspondence:

Nobuhiro Takahashi, Ph.D.

Applied Biological Science

Tokyo University of Agriculture and Technology

3-5-8 Saiwai-cho, Fuchu, Tokyo 183-8509, Japan

Tel/Fax No.+81-042-367-5709

E-mail: ntakahas@cc.tuat.ac.jp

EXPERIMENTAL PROCEDURES

Reagents and Antibodies

Antibodies to SMN (2B1), Gemin2 (2E17), Gemin3 (12H12), Gemin4 (17D10), Gemin6 (20H8), Gemin8 (1F8), and Unrip (3G6) were generous gifts from Dr. Gideon Dreyfuss. Anti-Gemin5 was obtained from Bethyl Laboratories. Anti-CBP80 was a generous gift from Dr. Robin Reed. Anti-Sm (Y-12) was obtained from Thermo Fisher Scientific. Anti-hnRNP C1/C2 was obtained from ImmuQuest. Anti-La/SSB was obtained from Santa Cruz Biotechnology. Anti-DCP1A (3G4) was obtained from Abnova. Anti-FLAG-M2, rabbit anti-FLAG polyclonal, anti-coilin, anti-mouse or anti-rabbit IgG Cy3-conjugated antibody and anti-FLAG-M2-conjugated agarose beads were obtained from Sigma-Aldrich. Anti-GAPDH was obtained from Ambion. Anti-mouse IgG and alkaline phosphatase–conjugated anti-rabbit IgG were obtained from Cell Signaling Technology. FITC-conjugated anti-rabbit IgG was obtained from American Qualex antibodies. FITC-conjugated anti-mouse IgG was obtained from Kirkegaard & Perry Laboratories, Inc. . Anti-DDX6 and anti-LSM14A were obtained from GeneTex. All general reagents were purchased from Wako Pure Chemical, Kanto Chemical, or Nacalai Tesque.

Construction of Epitope-tagged Expression Plasmids

RT-PCR was performed with the Super Script II kit (Invitrogen) using total mRNA prepared from HEK293 cells. cDNAs encoding SMN (NM_000344), LUC7L (NM_201412), snurpotin1 (NM_005701), PHAX (NM_032177), U1-70K (NM_003089), Gemin2 (NM_003616), Gemin4 (NM_015721), Gemin5 (NM_015465), Gemin6 (NM_024775), SmB (NM_003091), SmD1 (NM_006938) and SmE (NM_003094) were amplified by PCR using primer sets
5’-GAAGAAGGATCCGCGATGAGCAGCGGCGGCAGT-3’ and
5’-GAAGAACTCGAGATTTAAGGAATGTGAGCACCT-3’ for SMN,
5’-GAAGAAGGATCCCACCATGGGATCCGCCCAGGCGCAGAT-3’ and
5’-CAAGAATTCCGATCTCGCCGGCCTCCT-3’ for LUC7L,
5’-GAAGAAGGATCCCACCATGGAAGAGTTGAGTCAGGCCCT-3’ and
5’-CAAGAATTCCATTCTCCATGAGGCATCCAGGGT-3’ for snurportin1,
5’-GAAGAAAAGCTTCACCATGGCGTTGGAGGTCGGCGA-3’ and
5’-CAAGAATTCCAAAGATGTCCAAATCATGAGAATGATCAACT-3’ for PHAX,
5’-GAAGAAAAGCTTCACCATGGGAACCCAGTTCCTGCCGCCCA-3’ and

5’-CAAGAATTCCCTCCGGCGCAGCCTCCAT-3’ for U1-70K,
5’-GAAGAAAAGCTTATGCGCCGAGCGGAACTGGCTG-3’ and
5’-CAACTCGAGTCAAGATGGCTCATCAGCTAAATCAC-3’ for Gemin2,
5’-GAAGGTACCATGGACCTAGGACCCTTGAAC-3’ and
5’-CAACTCGAGTCAGAAGCTGCTCATCTTCTG-3’ for Gemin4,
5’-GAAGAAGCTAGCCACCATGGGGCAGGAGCCGC-3’ and
5’-CAACTCGAGGATATCCCCATACAGAAGGTCTGGCAGTGT-3’ for Gemin5,
5’-GAAGAAAAGCTTATGAGTGAATGGATGAAGAAAGG-3’ and
5’-CAACTCGAGTCATTGGGAAGCTGTAAGATGTC-3’ for Gemin6,
5’-GAAGAAAAGCTTATGACGGTGGGCAAGAGCAGCAAG -3’ and
5’-CAAGCGGCCGCTCAAAGAAGGCCTCGCATCCCAG-3’ for SmB,
5’-GAAGAAAAGCTTATGAAGCTCGTGAGATTTTTGATG-3’ and
5’-CAACTCGAGTTATCGCCTAGGACCCCCTCTTC-3’ for SmD1, and
5’-GAAGAAAAGCTTATGGCGTACCGTGGCCAGGGTC-3’ and
5’-CAACTCGAGCTAGTTGGAGACACTTTGTAGCAG-3’ for SmE.
The PCR products were cloned into the following sites of pcDNA3.1(+): BamH I/Xho I (for SMN), BamH I/EcoR I (for LUC7L and snurportin1), Hind III/EcoR I (for PHAX and U1-70K), Kpn I/Xho I (for Gemin4), Hind III/Xho I (for Gemin2, Gemin6, SmD1 and SmE), Hind III/Not I (for SmB) and Nhe I/EcoR V (for Gemin5). All cloned cDNAs were verified by DNA sequencing.

The cloned cDNAs were subcloned into pcDNA3.1(+)-DAP (38), pcDNA3.1-EF (C-terminal TEV-FLAG tag), or pcDNA3.1(+)-HEF (N-terminal HA-TEV-FLAG or C-terminal FLAG-TEV-HA tag) to create epitope-tag fusions. N-terminal tag–fused proteins are prefixed with “HEF”, and C-terminal tag–fused proteins contained the suffix “EF” or “HEF”. The DAP tag–fused SMN coding sequence was excised with Bgl II/Apa I and ligated into the Bgl II/Apa I site of pcDNA-5-FRT for stable expression.

cDNAs for LUC7L-EF and snurportin1-EF were excised with BamH I/Xho I and ligated into the same sites of pcDNA5-FRT/TO. PHAX-EF cDNA was ligated into Hind III/Xho I sites of pcDNA5-FRT/TO; U1-70K-EF cDNA was ligated into Hind III/EcoR V sites of pcDNA-FRT/TO-PHAX-EF after removal of the PHAX coding sequence via Hind III/EcoR V; Tag-less Gemin4 cDNA was ligated into Kpn I/Xho I sites of pcDNA5-FRT/TO-HEF. Tag-less Gemin2, Gemin6, SmD1 and SmE cDNAs were ligated into Hind III/Xho I sites of pcDNA5-FRT/TO-HEF vector, respectively. Tag-less SmB cDNA was ligated into Hind III/Not I sites of pcDNA5-FRT/TO-HEF vector. Plasmid expressing Gemin5 was constructed as follows:

1) pcDNA3.1-Gemin5-HEF was cut with Nhe I, and a Klenow fill-in reaction generated a blunt end; 2) pcDNA5-FRT/TO was cut with Hind III with subsequent Klenow fill-in; 3) pcDNA3.1-Gemin5-HEF was partially digested with Xho I to yield the full-length Gemin5-HEF coding sequence, and pcDNA5-FRT/TO was completely digested with Xho I; 4) a DNA fragment encoding Gemin5-HEF was ligated into the blunt end/Xho I sites of pcDNA5-FRT/TO. All constructs were verified by sequencing.

For RNA aptamer–based affinity purification, cDNA encoding *Pseudomonas aeruginosa* phage 7 (PP7) coat protein (PP7CP) was amplified by PCR with pET28ZZTPP7His as template using the primer set 5'-GAAGAAAAGCTTCACCATGGCCAAAACCATCGTTCTTG-3' and 5'-CAAGAATTCCACGGCCCAGCGGCACAAG-3'. The PCR-amplified fragment was excised with Hind III/Eco RI and ligated into the corresponding sites of pcDNA3.1-HF for expression of C-terminally HA-FLAG-tagged PP7CP in human cells. The pcDNA3.1-PP7CP-HF construct was verified by sequencing. Sequences encoding the primer set 5'-GAAGAAGATATCGGGTTCTGGTGCCGAGAATTTGTATTTTCAGGGTTCTGGTGCCGATTACAAGGATGACGAC-3' and 5'-TTCTTCCTCGAGCCCGGGTTACTTATCGTCGTCATCCTTGTAATCGG-3' were amplified by PCR to yield cDNA fragment containing EF-tag coding sequence. The extended and amplified fragments were cut with EcoR V/Xho I and ligated into the corresponding sites of pcDNA3.1(+) to yield pcDNA3.1(+)-EF.

Primer set 5'-GAAGAAGCTAGCCACCATGGGATACCCATATGACGTCCCGGACTACGCC-3' and 5'-GAACCCTGGAAGTACAAATTCTCACCAGAACCGGCGTAGTCCGGGACGT-3' or 5'-GAATTTGTACTTCCAGGGTTCTGGTGCCGATTACAAGGATGACGACGAT-3' and 5'-TTCTTCAAGCTTGGCACCAGAACCCTTATCGTCGTCATCCTTGTAATCG-3' was used to extend and PCR-amplify each primer set. By using extended primer sets, further extension and amplification were done by additional PCR to obtain the cDNA fragment 5'-GAAGAAGCTAGCCACCATGGGATACCCATATGACGTCCCGGACTACGCCGGTTCTGGTGAGAATTTGTACTTCCAGGGTTCTGGTGCCGATTACAAGGATGACGACGATAAGGGTTCTGGTGCCAAGCTTGAAGAA-3'. This final product was ligated into NheI/HindIII sites of pcDNA3.1(+) to yield pcDNA3.1(+)-HEF encoding an N-terminal HEF tag. Similarly, primer set 5'-GAAGATATCGGGTTCTGGTGCCGATTACAAGGATGACGACGATAAGG-3' and 5'-TGGAAGTACAAATTCTCGGCACCAGAACCCTTATCGTCGTCATCCTT-3' or

5'-CCGAGAATTTGTACTTCCAGGGTTCTGGTGCCTACCCATATGACG-3' and 5'-CAACTCGAGTTAGGCGTAGTCCGGGACGTCATATGGGTAGGCA-3' was used to prepare the cDNA fragment 5'-GAAGATATCGGGTTCTGGTGCCGATTACAAGGATGACGACGATAAGGGTTCTGGTGCCGAGAATTTGTACTTCCAGGGTTCTGGTGCCTACCCATATGACGTCCCGGACTACGCCTAACTCGAGTTG-3' to construct pcDNA3.1(+)-HEF encoding a C-terminal HEF tag; namely, the cDNA fragment was ligated into EcoR V/Xho I sites of pcDNA3.1(+). Using pcDNA3.1(+)-HEF containing an N-terminal HEF tag sequence as template, flanking region of HEF tag coding sequence was amplified with primer set 5'-CACCATGGGATACCCATATGAC-3' and 5'-TTCGGATCCGAGCTCGGTACCAAG-3'. pcDNA5-FRT/TO vector was cut with Hind III with subsequent klenow fill-in to make blunt end, digested with BamH I and ligated with BamH I digest of the amplified fragment to make pcDNA5-FRT/TO-HEF. All constructs were verified by DNA sequencing.

Cell Culture and Transfection

HEK293, Flp-In T-REx-293, 293T, HeLa, and MCF7 cells were maintained according to standard methods in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum in a humidified atmosphere of 95% air and 5% CO_2. Transfection was performed by either the calcium phosphate method or lipofection method; lipofection was performed with Lipofectamine 2000, Lipofectamine RNAiMAX, or Lipofectamine LTX reagent according to manufacturer's protocol (Life Technologies–Invitrogen).

Total RNA Extraction

After washing with PBS(-), cells were harvested in tubes and pelleted via centrifugation at 1,200 × *g* for 5 min at 4°C. Total RNA extraction was performed with the RNAgent total RNA isolation system (Promega) according to manufacturer's protocol.

Preparation of Stable or Doxycycline-inducible Cell Lines Expressing Epitope-tagged Protein

Flp-In T-REx-293 cells were cultured in DMEM at 37°C, seeded in a 24-well plate, and transfected with 2 μl Lipofectamine 2000 mixed with 0.25 μg of an appropriate pcDNA5-FRT or

pcDNA5-FRT/TO construct together with 0.25 μg of pOG44 vector (Invitrogen). When cells reached 90–100% confluency approximately 24–48 h after transfection, 100 μg/ml of hygromycin B (Invitrogen) was added to DMEM supplemented with 10% fetal bovine serum. Cells were cultured further for ~2 weeks with this medium containing 100 μg/ml hygromycin B, and the medium was changed every 2 days. Selected cell colonies were cultured in large scale for additional experiments.

Epitope-tag Affinity Purification

Doxycycline-inducible Flip-In T-REx 293 cells (1×10^6 in DMEM) expressing epitope-tagged protein were treated with 1 μg/ml doxycycline for 48 h. The cells were harvested from three 90-mm dishes, washed with PBS (–) once, lysed by vigorous mixing for 30 s in 1 ml lysis buffer (50 mM Tris-HCl pH 8.0, containing 150 mM NaCl, 0.5% (w/v) IGEPAL-CA630, 1 mM PMSF), and incubated on ice for 30 min. Soluble whole-cell lysate was prepared by centrifugation at 20,000 × *g* for 30 min at 4°C, and the resultant supernatant was incubated with 10–15 μl anti-FLAG M2 agarose beads (Sigma-Aldrich) for 3 h at 4°C with rotation. After washing the beads five times with 1 ml lysis buffer and once with 50 mM Tris-HCl pH 8.0 containing 150 mM NaCl, bound proteins and RNAs were eluted with 500 μg/ml FLAG peptide in the same buffer and used for further experiments.

Extraction of Protein and RNA from RNPs

RNPs prepared by epitope-tag purification were solvated in an equal volume of 2× RNA extraction solution (10 mM Tris-HCl pH 8.0 containing 7 M urea, 350 mM NaCl, 1% SDS, 10 mM EDTA) and buffered phenol/chloroform (pH 8.0) and mixed vigorously. The aqueous and organic phases were separated by centrifugation at 20,000 × *g* for 30 min at 4°C. To the aqueous phase containing RNAs, an equal volume of 2-propanol was added to precipitate RNAs; to the organic phase containing proteins, 4 volumes of 2-propanol were added, and the solution was mixed. RNAs or proteins were pelleted via centrifugation at 20,000 × *g* for 30 min at 4°C. Each precipitate was rinsed with 75% ethanol and air-dried. RNA and protein samples were submitted for further analysis.

Electrophoresis of RNA and In-gel RNA Digestion

RNAs were separated via PAGE essentially as described (50). Briefly, RNA in solution was denatured prior to electrophoresis by mixing with 9 volumes of a loading buffer containing 2 mM EDTA (pH 8.0) and 95% formamide, heating at 65°C for 10 min, and cooling immediately on ice. RNA was separated on 10% (w/v) polyacrylamide gels containing 8 M urea and 0.5× TBE (45 mM Tris base, 32.3 mM boric acid, 1.25 mM EDTA, pH 8.3) with 0.5× TBE as a running buffer, and gels were stained with SYBR Gold (Invitrogen) for 10 min.

In-gel RNA digestion was done essentially as described (33). Briefly, excised gel pieces containing RNAs were cut into small pieces and dried under vacuum. Gel pieces were digested in 15 μl of 2 ng/μl RNase T1, with incubation at 37ºC for 1 h. The nucleolytic fragments were extracted from the gel using 100 μl of RNase-free water, passed through a centrifugal filter unit with a polyvinylidene fluoride membrane (Ultrafree-MC, Millipore, Billerica, MA), and then 5 μl of 2 M triethylammonium acetate (pH 7.0) was added before LC-MS analysis.

LC-MS Apparatus for RNA Analysis

The LC system used was essentially as described (51), consisting of a nanoflow pump (LC Assist) that delivers solvent to a fritless spray tip electrospray ionization column and a ReNCon gradient device. The column was prepared with a fused-silica capillary (150 μm i.d. × 375 μm o.d.) using a laser puller (Sutter Instruments) and was slurry-packed with reverse-phase material (Develosil C30-UG-3, particle size 3 μm, Nomura Chemical) to a length of 50 mm. High voltage for ionization in the negative mode was applied, and the LC eluate was sprayed in-line to an LTQ-Orbitrap hybrid mass spectrometer (model XL, Thermo Fisher Scientific). LC was performed at a flow rate of 100 nl/min using a 40-min linear gradient from 5% to 40% methanol in 10 mM triethylammonium acetate (pH 7.0). The mass spectrometer was operated in a mode to automatically switch between Orbitrap-MS and linear ion trap–MS/MS acquisition as described (32).

Immunoblotting

Proteins were dissolved in SDS-PAGE sample buffer, separated by SDS-PAGE, and transferred electrophoretically to a polyvinylidene difluoride membrane. The membrane was treated with 5% non-fat dried milk in PBS at 4°C at least for 1 h and then incubated with an appropriate antibody overnight at 4°C. The membrane was washed three times with TBST (Tris-buffered saline containing 0.1% w/v Tween 20) for 10 min, incubated with a secondary antibody conjugated with

alkaline phosphatase for 1 h at room temperature, washed three times with TBST for 10 min, washed once with TBS for 5 min, and stained by adding NBT (nitro-blue tetrazolium chloride)/BCIP (5-bromo-4-chloro-3'-Indolylphosphatase *p*-toluidine salt) solution, which was prepared by a (1:50) dilution of NBT/BCIP stock solution (Roche) with alkaline phosphatase buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$).

Denatured Urea PAGE and Northern Blotting

RNAs prepared from RNPs or from whole cells were analyzed by northern blotting. RNA (1 μg per lane) was loaded for denaturing urea-PAGE (7.5 M urea, 9% polyacrylamide) in 0.5× TBE running buffer at 12.5 V/cm for 2–3 h. RNAs separated via PAGE were transferred electrophoretically to a Hybond-N+ membrane (GE Healthcare) using a semi-dry blotting apparatus with 0.5× TBE. The membrane was dried, cross-linked by UV irradiation, and blocked for at least for 1 h with salmon sperm DNA in pre-hybridization buffer containing 5× SSC (saline-sodium citrate buffer), 20 mM NaH_2PO_4 pH 7.5, 7% SDS, 2× Denhardt's Solution, and 40 μg/ml sheared salmon sperm DNA. After blocking, RNA was hybridized with an oligonucleotide labeled with biotin (see method below) overnight at 50°C. The hybridized membrane was washed twice with non-stringent wash solution (3× SSC, 25 mM NaH_2PO_4 pH 7.5, 5% SDS) for 15 min and twice with stringent wash solution (1× SSC, 1% SDS) for 15 min. RNAs hybridized with biotin-labeled DNA were visualized with a chemiluminescent nucleic acid detection module (Thermo Fisher Scientific) and detected with an LAS4000 luminescent image analyzer (Fujifilm).

RNA Interference

HeLa (1×10^5 cells) or 293T cells (5×10^5 cells) were cultured in 35-mm dishes until they reached 80% confluency. HeLa cells were transfected with 3 μl of Lipofectamine RNAiMAX and 60 pmol of a stealth small interfering RNA (siRNA) specific for HeLa cells. 293T cells were transfected with 5 μl of Lipofectamine RNAiMAX and 100 pmol of a stealth siRNA. The following stealth siRNAs were used to knockdown the transcripts as noted: 5'-CGGUUGCAUUUACCCAGCUACCAUU-3' and 5'-AAUGGUAGCUGGGUAAAUGCAACCG-3' for SMN, 5'-CGAAGCAGCUCAAUGUCCAGAUGUU-3' and 5'-AACAUCUGGACAUUGAGCUGCUUCG-3' for Gemin2, 5'-CGACGACAACUCUGUAGACUGAGUU-3' and 5'-AACUCAGUCUACAGAGUUGUCGUCG-3' for Gemin2 negative control, 5'-GACCUCCUCCCAAAGAUACUGGUAU-3' and

5’-AUACCAGUAUCUUUGGGAGGAGGUC-3’ for SmB/B’ (siRNA1),
5’-UCCUGGUAUGAGACCUCCUAUGGGU-3’ and
5’-ACCCAUAGGAGGUCUCAUACCAGGA-3’ for SmB/B’ (siRNA2), and
5’-GACCCUCCCAAAGAUACUGGCUUAU-3’ and
5’-AUAAGCCAGUAUCUUUGGGAGGGUC-3’ for SmB/B’ negative control. The cells were transferred to 35-mm or 60-mm dishes 24 h after transfection. The transfected cells were cultured in DMEM supplemented with 10% fetal bovine serum, and samples (~5×10^5 cells) were taken at various time points.

Biotin Labeling of Oligonucleotide Probes for Northern Blotting

The following oligonucleotide probes were labeled at the 3’end with biotin using the Biotin 3' End DNA Labeling kit (Pierce):
U1-#1 probe (5’-GTATCTCCCCTGCCAGGTAAGTAT-3’),
U1-#2 probe (5’-AGCACATCCGGAGTGCAATGGATA-3’),
U1-#3 probe (5’-TATGCAGTCGAGTTTCCCACATTTGG-3’),
U1-#4 probe (5’-GCAGTCCCCCACTACCACAAAT-3’),
U2 probe (5’-TACTGCAATACCAGGTCGATGCGT-3’),
U4 probe (5’-GACTATATTGCAAGTCGTCACGGC-3’),
U5 probe (5’-GACTCAGAGTTATTCCTCTCCACG-3’),
U6 probe (5’-ACGAATTTGCGTGTCATCCTTGCG-3’),
U7 probe (5’-AGCCAGAAAGCCTACTAGACAAATTCT-3’),
U11 probe (5’-TACGTGTGCCACTCACGACAGAAG-3’),
5S probe (5’-TTCCGAGATCAGACGAGATCGG-3’),
5.8S probe (5’-AGACAGGCGTAGCCCCGGGAGGAA-3’),
Met-tRNA probe (5’-TAGCAGAGGATGGTTTCGATCCATCGA-3’),
y18Sn tag probe (5’-CCAAAGCCTGAATCCTCG-3’),
RAT tag probe (5’-ACGTCTAAGGGTTTCCATATAAACTCCTT-3’),
7SK probe (5’-AGGCAGACTGCCACATGCAG-3’), and
7SK-5’ probe (5’-TGGGGTGACAGATGTCGCAGCCAGAT-3’).

Immunocytochemical Analysis

Cells were washed with PBS, fixed with 4% paraformaldehyde in PBS for 10 min at room

temperature, and washed twice with PBST (PBS with 0.05% w/v Tween 20). The cells were permeabilized by treating with PBS containing 0.1% (w/v) Triton X-100 for 5 min at room temperature, washed once with PBST, blocked with 3% (w/v) non-fat dried milk in PBS for at least 30 min, and incubated with an appropriate primary antibody diluted in 3% non-fat dried milk /PBS for 1 h at room temperature. The cells were washed three times in PBST for 10 min and incubated with a fluorochrome-conjugated secondary antibody diluted in 3% non-fat dried milk /PBS for 1 h at room temperature. After washing three times in PBST for 10 min, the cells were mounted with Vectashield and observed with an Axiovert 200 M microscope (Carl Zeiss). When the experiment was done with FISH, the antibodies were diluted with 100 μg/ml BSA instead of 3% non-fat dried milk/PBS and total washing time with PBST was shortened.

Glycerol Density Gradient Ultracentrifugation

Cell extract was prepared by the method described in “Epitope-tag Affinity Purification”. Protein concentration of the cell extract was determined using Protein Assay kit (Bio-Rad). Cell extract (2.5 mg) of control T-REx 293 cells or Gemin5-HEF expressing T-Rex 293 cells was layered on a 10–30% (v/v) glycerol density gradient containing 50 mM Tris-HCl pH 8.0 and 150 mM NaCl, centrifuged with a Beckman MLS50 rotor at 40,000 rpm (average 128,400 × *g*) for 185 min at 4°C, and fractionated into 10 fractions (500 μl each). Fractions 2 and 3 (as mixture A), 5 and 6 (as mixture B), or 8 and 9 (as mixture C) were mixed and subjected to the pull-down analysis described in “Epitope-tag Affinity Purification”.

Figure S1

(A) DAP-SMN-expressing Flip-In T-REx 293 cells were analyzed by immunocytochemistry with a primary antibody against FLAG, SMN, or SmB/B' (Y-12). FITC-labeled anti-rabbit IgG (green) was used as the secondary antibody to detect DAP-SMN. Cy3-labeled anti-mouse IgG secondary antibody (red) detected endogenous and exogenous SMN or SmB/B'. DAPI staining shows the nucleus. Merge: FITC, Cy3, and DAPI staining are merged, Scale bar: 10 μm.
(B) Proteins were pulled down (PD:FLAG) from DAP-SMN-expressing Flip-In T-REx 293 cells (DAP-SMN) or their parent T-REx 293 cells (T-Rex) with anti-FLAG-conjugated beads, separated by SDS-PAGE, and visualized by silver staining (left) or immunoblotting (IB) with antibodies against the proteins indicated (right). Input (1% Input) is the cell extract used for pull-down.

Figure S2

A) The trimethylguanosine 5'cap and the first four nucleotides are shown. Major fragmentation sites and the major c/y- and a/w-series ions produced by collision-induced dissociation are indicated. Annotations for the product ions are given based on the nomenclature of Mcluckey et al. (52).
(B) The RNase T1 5'-terminal oligonucleotide produced from the U1 snRNA or U1-tfs fraction in Figure 1C was digested with RNase A and tobacco acid pyrophosphatase and subjected to LC-MS/MS. The MS/MS spectrum of pAmUmACp generated from the RNase T1 5'-terminal oligonucleotide of U1 snRNA (upper) or that of pmAmUmACp generated from the RNase T1 5'-terminal oligonucleotide of U1-tfs (bottom) is indicated. c/y and a/w ions generated by MS/MS fragmentation are indicated in each spectrum (see Figure S2A for the explanation of c/y and a/w). Inset in the upper left of each spectrum shows an enlarged view of the range m/z=300~m/z=450.
(C) Enlarged views of the mass range m/z=590~m/z=660 of the spectra in Figure S2B. All ions except y2 indicated in the bottom spectrum show mass-m/z values higher than those observed in the top spectrum. Open arrow in the upper MS/MS spectrum of pAmUmACp indicates the position of the ion having base methylation of the first adenine.
(D) MS/MS spectrum of the 3'-terminal oligonucleotide, CUUUCCCCUG-OH or CUUUCCCCUG>p, generated by RNase T1 digestion of the U1 snRNA fraction.
(E) MS/MS spectrum of CAmCUCCG>p generated from U1 snRNA (upper) or that of CACUCCG>p generated from U1-tfs (bottom). c/y and a/w ions generated by MS/MS fragmentation are indicated in each spectrum.

(F) Structural features of U1 snRNA and U1-tfs RNA associated with SMN. U1-tfs have a monomethylguanosine cap with additional base methylation at the first adenosine, whereas U1 snRNA has a trimethylguanosine cap with no additional base methylation. Ribose is methylated at adenosine 70 (Am) in U1 snRNA but not in U1-tfs. U1-tfs lack the Sm protein–binding site and SL4 region.

Figure S3

HEF-Gemin2-, HEF-Gemin6-, HEF-SmB/B'-, HEF-SmD1, and HEF-SmE-associated RNA-protein complexes were prepared by pull-down method and analyzed by SDS-PAGE and silver staining or immunoblotting (IB) with anti-FLAG antibody, and by denaturing urea PAGE -SYBR Gold-staining or northern blotting with the probes indicated.

Figure S4

(A) Schematic diagram of the U1 gene used for the construction of vectors expressing U1 snRNA. The cis-acting elements DSE (distal sequence element, blue), PSE (proximal sequence element, purple), and 3'box element (green) are indicated as well as the U1 coding region (red). The DNA sequence of the U1 gene is also shown, and the sequences corresponding to DSE, PSE, 3'box, and U1 coding region are highlighted with color. (B) Schematic diagram of U1 gene constructs expressing U1 snRNA and/or U1-tfs. See the text for the explanation of each construct (WT, ΔSmSL4, ΔPSE, ΔDSE, and Δ3' box). DSE: distal sequence element, PSE: proximal sequence element, SL: stem loop, Sm: Sm protein–binding site. (C) Total RNA was extracted from 293 EBNA cells transfected with an expression vector encoding WT, ΔSmSL4, ΔPSE, ΔDSE, or Δ3' box. RNAs were analyzed by northern blotting with probe #1 or #3. Arrowhead indicates a band corresponding to endogenous U1 snRNA along with exogenous U1 or the endogenous U1-tfs and exogenous U1-tfs. Graph presents the mean ± SD of triplicate experiments; staining intensity of the U1-tfs band was standardized with that of U1-tfs observed in untransfected control. P values are for the comparison with the control using the t test (*$P < 0.1$, **$P < 0.05$). RNA staining with SYBR Gold is also shown. (D) RNAs were extracted from cells transiently transfected with the RAT-tagged U1 gene derivatives and separated by denaturing urea-PAGE. Subsequent northern blot analysis was carried out with probe #1 (detects both endogenous and RAT-tagged U1) and with the RAT probe (detects only exogenously expressed RAT-tagged U1). SYBR Gold staining (left) shows the RNAs used for the northern blot. RAT-WT, RAT-ΔSmSL4, and RAT-ΔSm are expected to produce RAT-tagged snRNAs with 239 nt, 200 nt, and 231 nt,

respectively. (E) Total RNA extracted from 293T cells transiently transfected with an expression vector composed of one of the RAT-tagged constructs (RAT-WT, RAT-ΔSmSL4, RAT-ΔSL4-1, RAT-ΔSL4-2, RAT-ΔSL4-3, RAT-ΔSL4-1Δ3'box, RAT-ΔSL4-2Δ3'box, and RAT-ΔSL4-3Δ3'box) was analyzed by northern blotting with the RAT probe or probe #1. SYBR Gold staining (left) shows the RNAs used for the northern blot. Arrowheads indicate the predicted size of RAT-tagged U1 snRNA (239 nt) and RAT-tagged U1-tfs (200 nt).

Figure S5.

(A) 293T cells transfected with an expression vector encoding y18Sn-ΔSL4-1, were subjected to FISH. Endogenous U1 snRNA, exogenous U1 snRNA and U1-tfs were detected with probe #3, or #SL4 labeled with FITC (green). Exogenously expressed U1 snRNA or U1-tfs were also detected with the Cy3-labeled y18Sn probe (red). DAPI staining shows the nucleus. Merge: FITC, Cy3, and DAPI staining are merged, Scale bar: 10 μm.

(B) y18Sn-WT-, y18Sn-ΔSm-, or, y18Sn-ΔSL4-1-expressing cells were stained by immunocytochemistry with antibodies against SMN (green) and by FISH with the Cy3-labeled y18Sn probe (red).

(C) y18Sn-ΔSL4-1-expressing cells were stained by immunocytochemistry with antibodies against the proteins (green) indicated and by FISH with the Cy3-labeled y18Sn probe (red).

Figure S6

(A) Proteins were pulled down from extract of RAT-7SK-, RAT-WT, RAT-Δ3'box, RAT-ΔSL-4-2-, or RAT-ΔSL4-2Δ3'box-expressing cells by RAT-based affinity purification, and were visualized by immunoblotting (IB) with antibodies against the proteins indicated. RAT-tagged RNAs were detected by northern blotting with the RAT probe or #1. PD: RAT, RAT-tagged RNA-protein complex bound to FLAG-tagged PP7CP was pulled down with anti-FLAG-conjugated beads and eluted with FLAG peptide. (B) Reverse-phase LC separation of the RNase T1 digest of RAT-U1 snRNA (RAT-WT) or RAT-U1-tfs (RAT-ΔSL4-1 or RAT-ΔSL4-1Δ3'box). Effluent was monitored as the count of total ions, m/z=1349.15 (MMG-mAmUmACUUACCUG; MMG-3m), m/z=1344.48 (MMG-AmUmACUUACCUG; MMG-2m), or m/z=1353.82

Figure S7

(A) Total RNA was prepared from cells transfected with a stealth siRNA (si) for SmB/B'

knockdown or Gemin2 knockdown, separated by denaturing urea-PAGE, and detected with SYBR Gold staining. Knockdown efficiency was examined by immunoblotting (IB) with antibodies against SmB/B' and Gemin2. In the graph, SYBR Gold staining intensity of U1 or U2 was standardized with that of the corresponding RNA present in cells transfected with a control RNA (sc, a scrambled-sequence of the SmB/B' siRNA). Each value represents the mean ± SD of three independent experiments. (B) Cells were analyzed by combination of immunocytochemistry with the antibody against SmB/B', SMN, or DCP1A (red) and FISH with the probe #1 (green). The si-2 target sequence in SmB/B' differs from that of si-1. After the knockdown of SmB/B' with si-2, FISH was done with the probe complementary to U2 snRNA (green), and the immunocytochemistry with the antibody against DCP1A (green). (C) The SmB/B' knockdown cells (si-1) were analyzed by FISH with the probe #1, or #3 (green) and that complementary to U2 snRNA (U2) or 5' internal spacer sequence of ribosomal RNA (5'ITS1) (red). (D) The SmB/B' knockdown cells (si-2) were analyzed by FISH with the probe U2 and the probe #1, or #3 (green). (E) 293T cells co-transfected with y18Sn-WT and stealth RNA for SMN knockdown were analyzed by FISH with the probe y18Sn and by immunocytochemistry with anti-SMN as the primary antibody and FITC-labeled secondary antibody. DAPI staining shows the nucleus. Enlarged view of the FISH staining was also shown. Merge: FITC, Cy3, and DAPI staining are merged, Scale bar: 10 μm.

Figure S1A

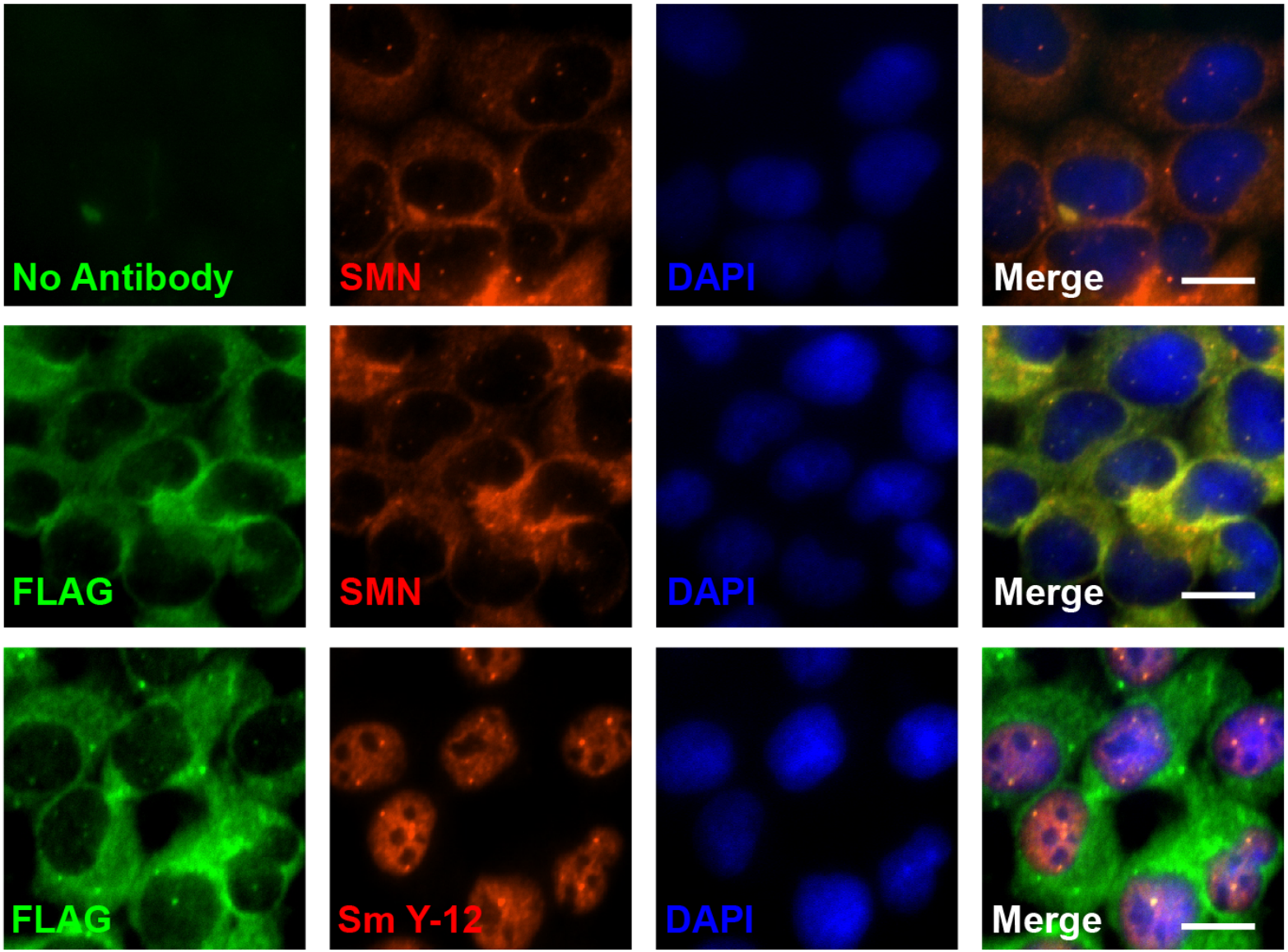

No Antibody
SMN
DAPI
Merge
FLAG
SMN
DAPI
Merge
FLAG
Sm Y-12
DAPI
Merge

Figure S1B

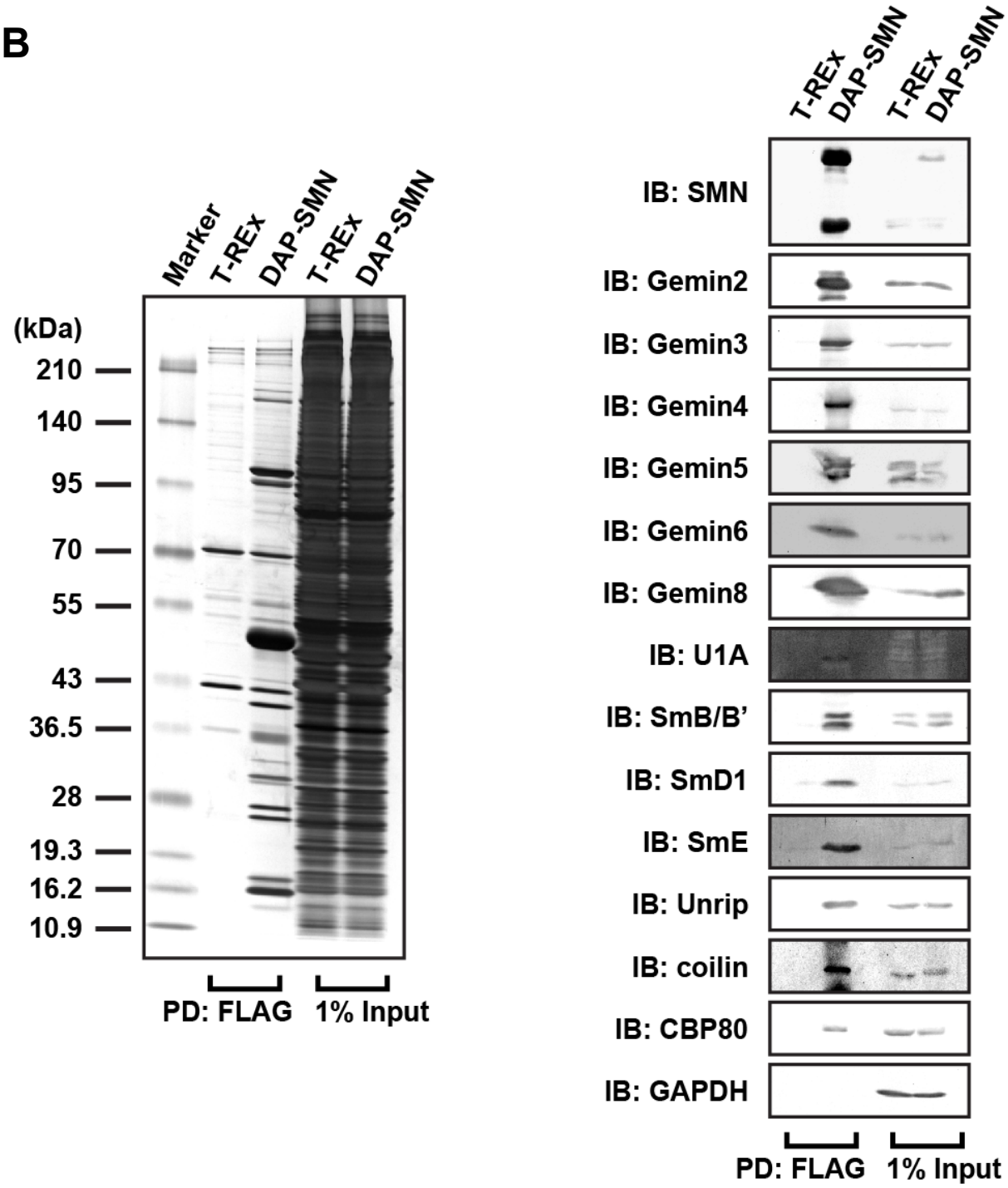

Marker
T-REx
DAP-SMN
T-REx
DAP-SMN
(kDa)
210
140
95
70
55
43
36.5
28
19.3
16.2
10.9
PD: FLAG
1% Input
T-REx
DAP-SMN
T-REx
DAP-SMN
IB: SMN
IB: Gemin2
IB: Gemin3
IB: Gemin4
IB: Gemin5
IB: Gemin6
IB: Gemin8
IB: U1A
IB: SmB/B'
IB: SmD1
IB: SmE
IB: Unrip
IB: coilin
IB: CBP80
IB: GAPDH
PD: FLAG
1% Input

Figure S2A

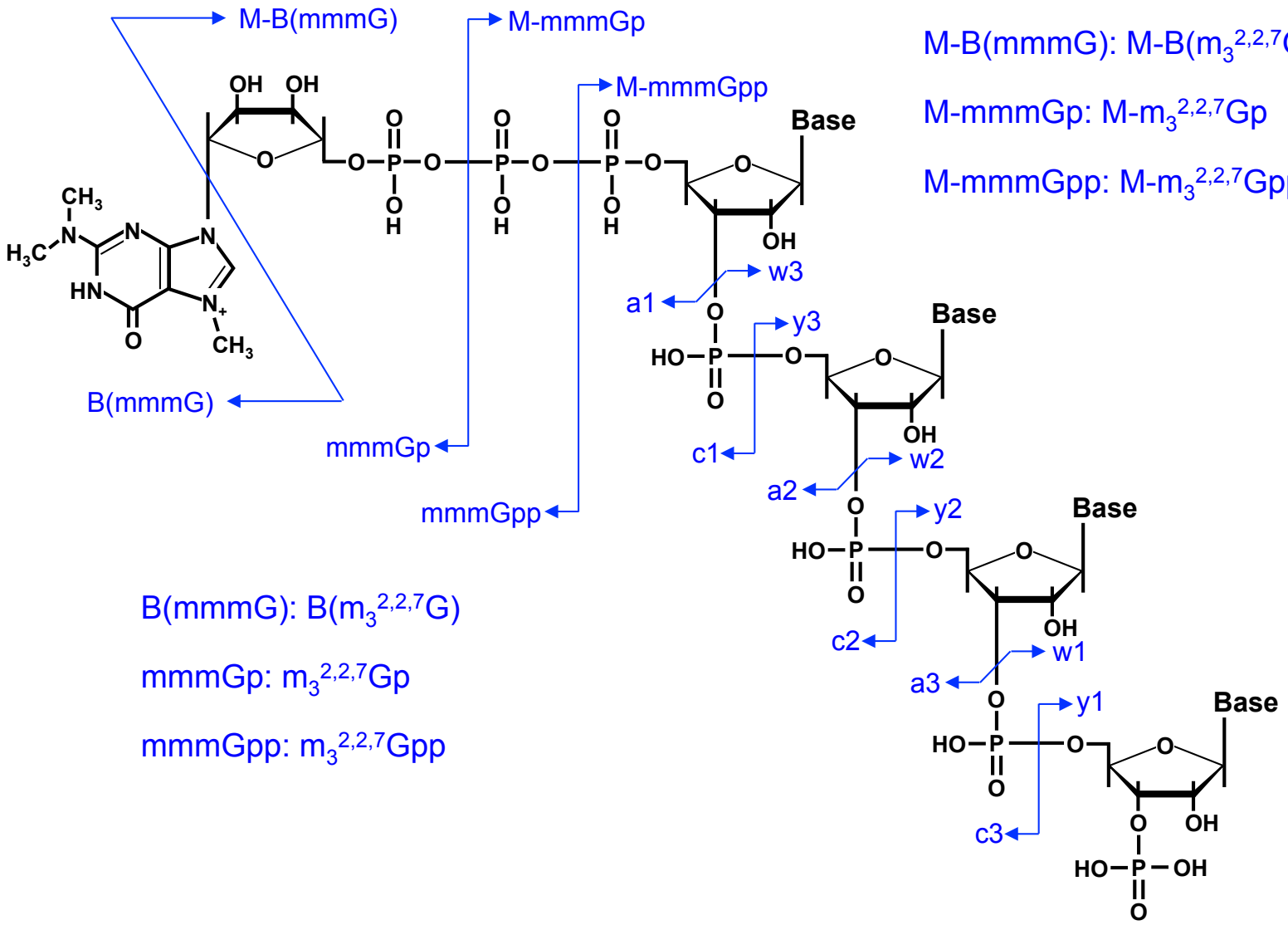

M-B(mmmG)
M-mmmGp
M-mmmGpp
Base
B(mmmG)
mmmGp
mmmGpp
w3
a1
y3
c1
w2
a2
y2
c2
w1
a3
y1
c3
M-B(mmmG): M-B($m_3^{2,2,7}$G)
M-mmmGp: M-$m_3^{2,2,7}$Gp
M-mmmGpp: M-$m_3^{2,2,7}$Gpp
B(mmmG): B($m_3^{2,2,7}$G)
mmmGp: $m_3^{2,2,7}$Gp
mmmGpp: $m_3^{2,2,7}$Gpp

Figure S2B

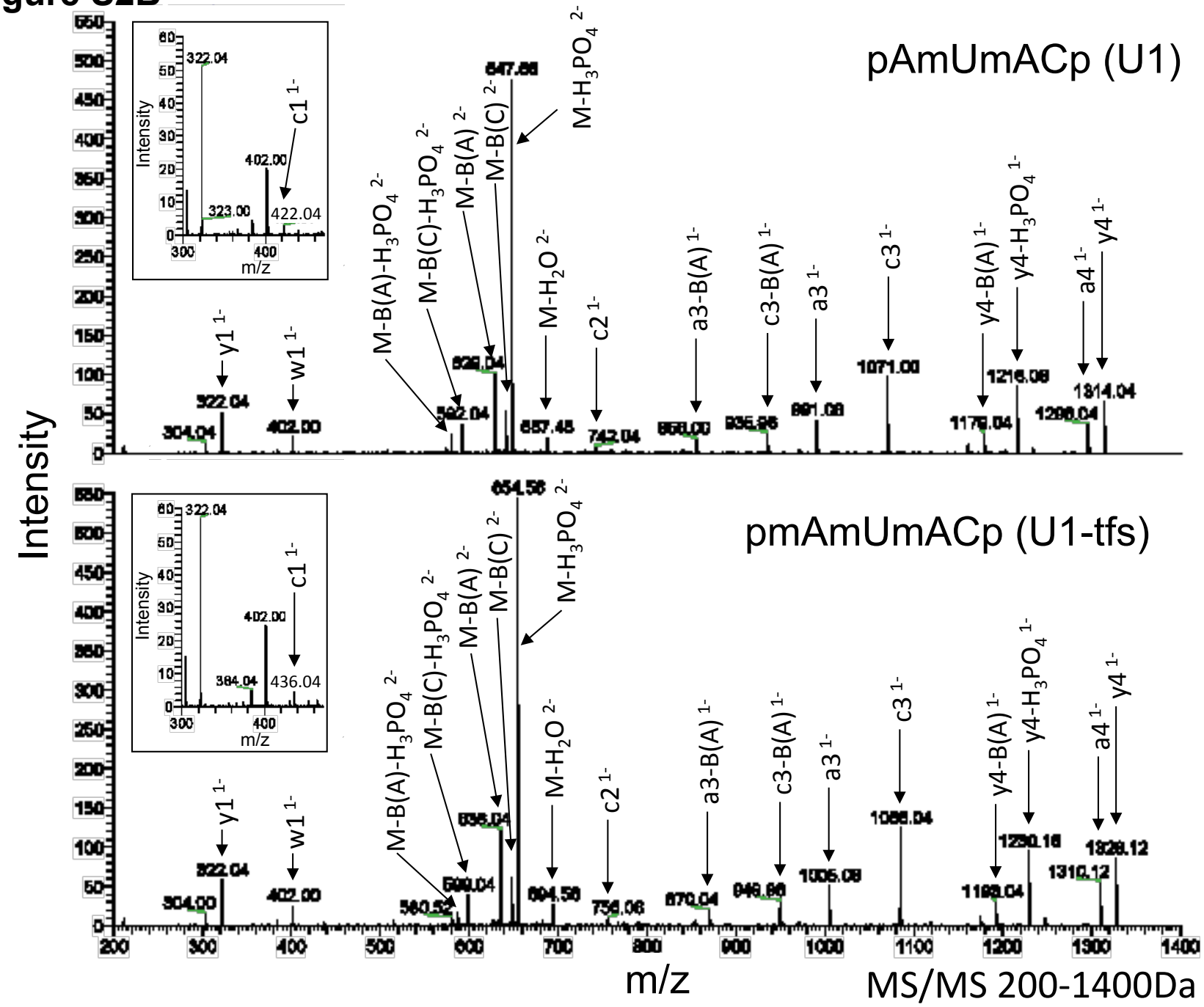

pAmUmACp (U1)
pmAmUmACp (U1-tfs)
Intensity
m/z
MS/MS 200-1400Da
M-H3PO4 2-
M-B(C) 2-
M-B(A) 2-
M-B(C)-H3PO4 2-
M-B(A)-H3PO4 2-
M-H2O 2-
c2 1-
a3-B(A) 1-
c3-B(A) 1-
a3 1-
c3 1-
y4-B(A) 1-
y4-H3PO4 1-
a4 1-
y4 1-
y1 1-
w1 1-
c1 1-
322.04
402.00
422.04
323.00
436.04
384.04

Figure S2C

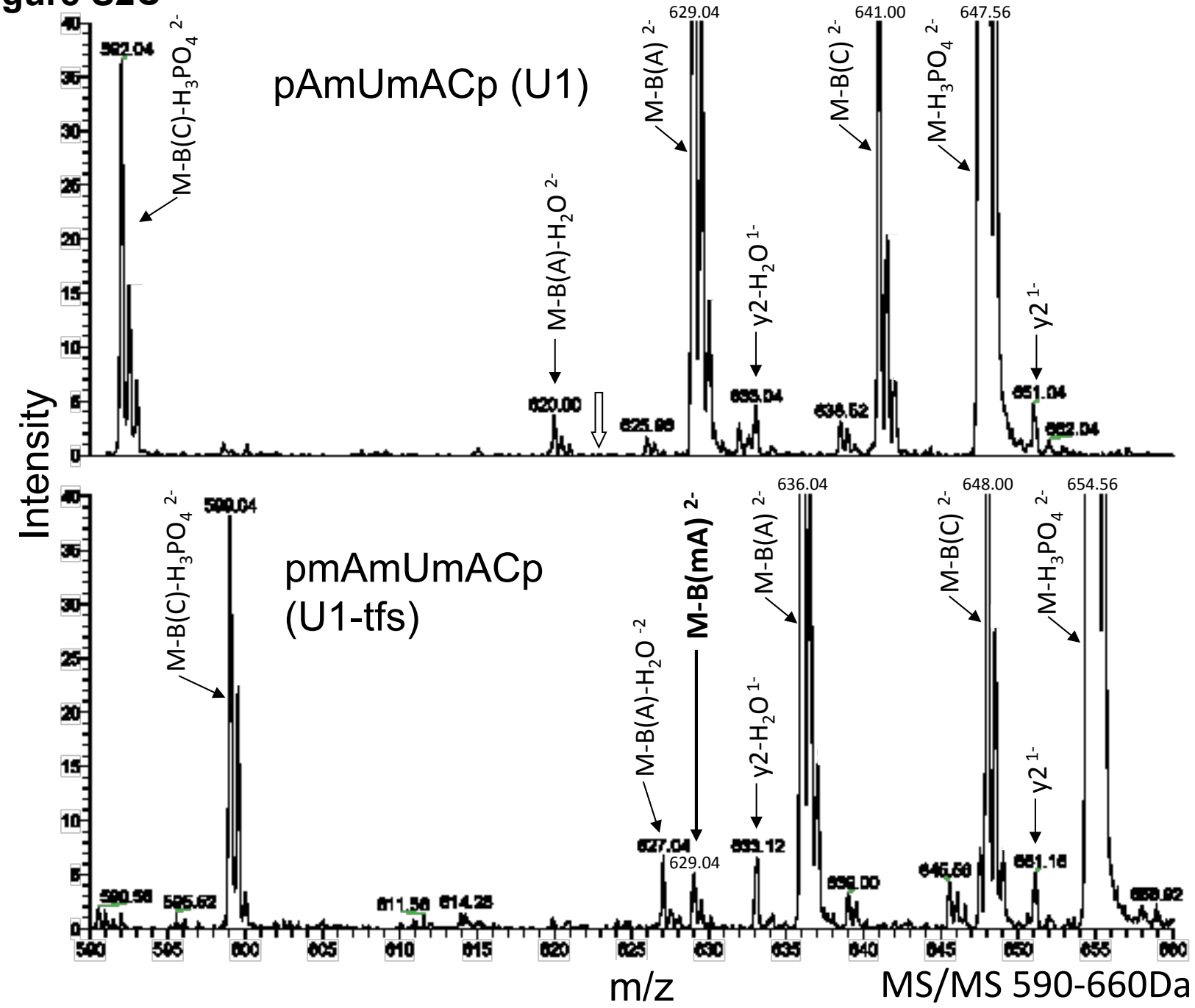

pAmUmACp (U1)
M-B(C)-H3PO4 2-
M-B(A)-H2O 2-
M-B(A) 2-
M-B(C) 2-
M-H3PO4 2-
y2-H2O 1-
y2 1-
592.04
620.00
629.04
641.00
647.56
651.04
pmAmUmACp
(U1-tfs)
M-B(mA) 2-
M-B(C)-H3PO4 2-
M-B(A)-H2O -2
M-B(A) 2-
M-B(C) 2-
M-H3PO4 2-
y2-H2O 1-
y2 1-
590.04
627.04
629.04
633.12
636.04
648.00
651.16
654.56
Intensity
m/z
MS/MS 590-660Da

Figure S2D

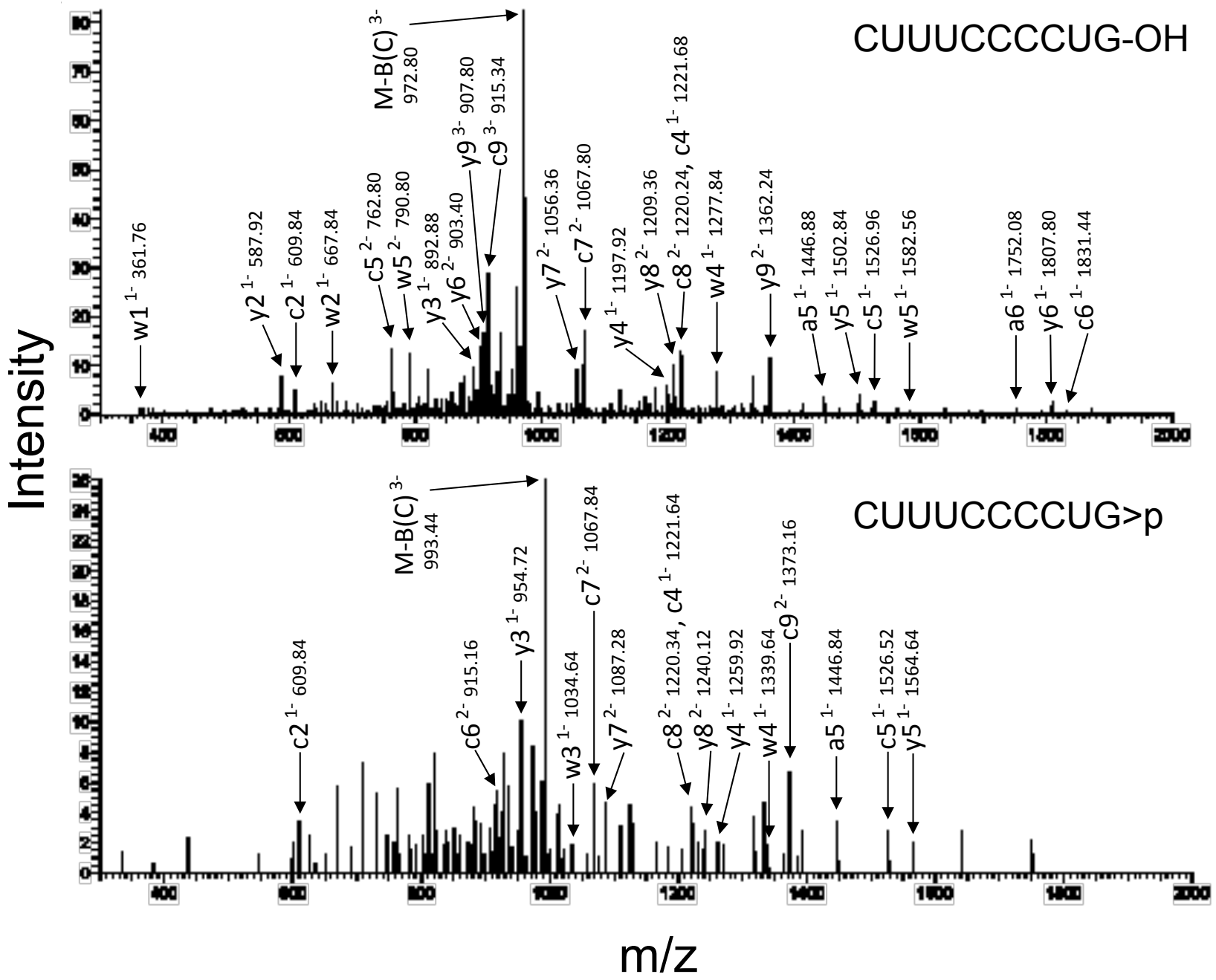

CUUUCCCCUG-OH
M-B(C) 3- 972.80
w1 1- 361.76
y2 1- 587.92
c2 1- 609.84
w2 1- 667.84
c5 2- 762.80
w5 2- 790.80
y3 1- 892.88
y6 2- 903.40
y9 3- 907.80
c9 3- 915.34
y7 2- 1056.36
c7 2- 1067.80
y4 1- 1197.92
y8 2- 1209.36
c8 2- 1220.24, c4 1- 1221.68
w4 1- 1277.84
y9 2- 1362.24
a5 1- 1446.88
y5 1- 1502.84
c5 1- 1526.96
w5 1- 1582.56
a6 1- 1752.08
y6 1- 1807.80
c6 1- 1831.44
CUUUCCCCUG>p
M-B(C) 3- 993.44
c2 1- 609.84
c6 2- 915.16
y3 1- 954.72
w3 1- 1034.64
c7 2- 1067.84
y7 2- 1087.28
c8 2- 1220.34, c4 1- 1221.64
y8 2- 1240.12
y4 1- 1259.92
w4 1- 1339.64
c9 2- 1373.16
a5 1- 1446.84
c5 1- 1526.52
y5 1- 1564.64
Intensity
m/z

Figure S2E

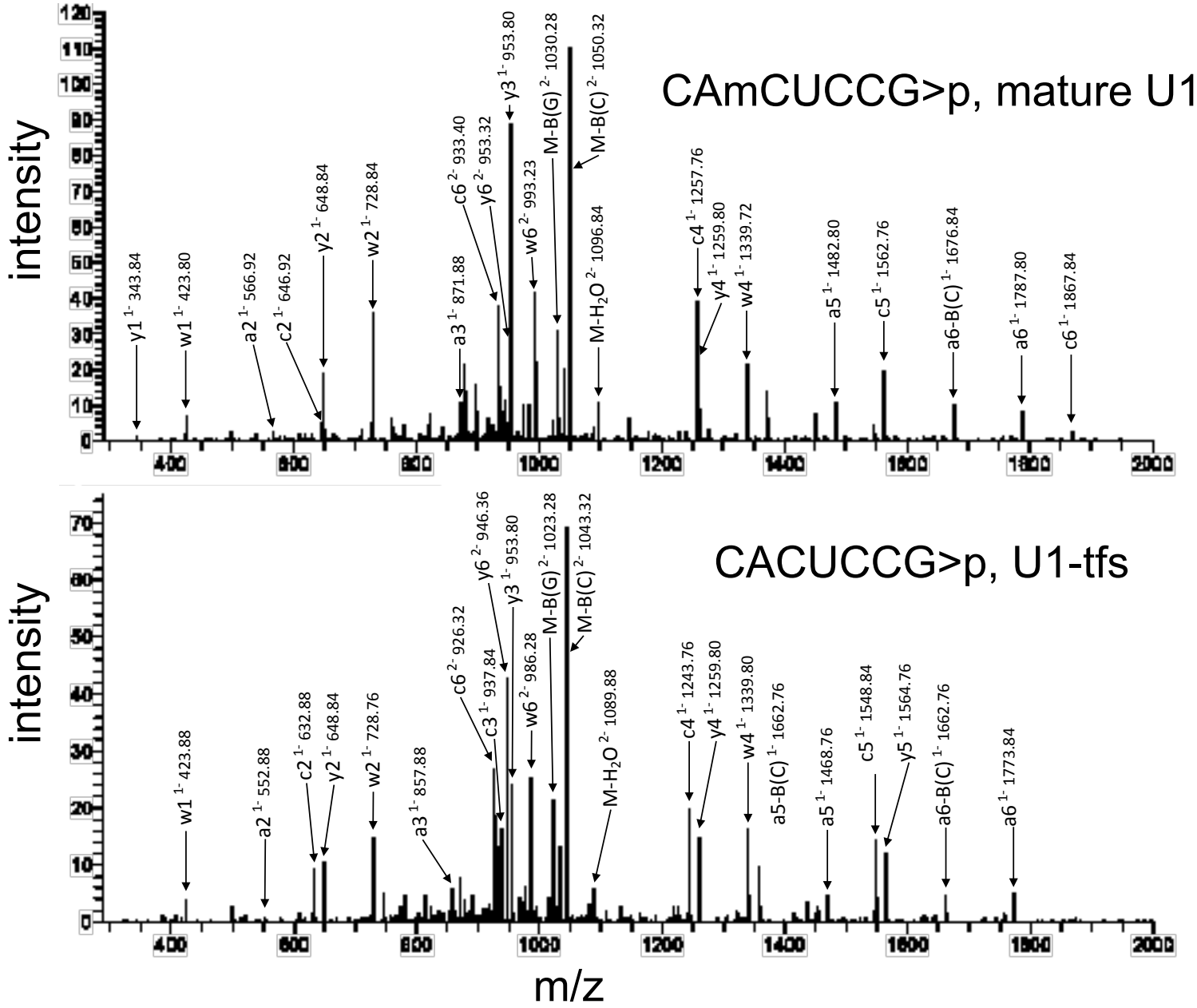

CAmCUCCG>p, mature U1
intensity
y1 1- 343.84
w1 1- 423.80
a2 1- 566.92
c2 1- 646.92
y2 1- 648.84
w2 1- 728.84
a3 1- 871.88
c6 2- 933.40
y6 2- 953.32
y3 1- 953.80
w6 2- 993.23
M-B(G) 2- 1030.28
M-B(C) 2- 1050.32
M-H2O 2- 1096.84
c4 1- 1257.76
y4 1- 1259.80
w4 1- 1339.72
a5 1- 1482.80
c5 1- 1562.76
a6-B(C) 1- 1676.84
a6 1- 1787.80
c6 1- 1867.84
CACUCCG>p, U1-tfs
intensity
w1 1- 423.88
a2 1- 552.88
c2 1- 632.88
y2 1- 648.84
w2 1- 728.76
a3 1- 857.88
c6 2- 926.32
c3 1- 937.84
y6 2- 946.36
y3 1- 953.80
w6 2- 986.28
M-B(G) 2- 1023.28
M-B(C) 2- 1043.32
M-H2O 2- 1089.88
c4 1- 1243.76
y4 1- 1259.80
w4 1- 1339.80
a5-B(C) 1- 1662.76
a5 1- 1468.76
c5 1- 1548.84
y5 1- 1564.76
a6-B(C) 1- 1662.76
a6 1- 1773.84
m/z

Figure S2F

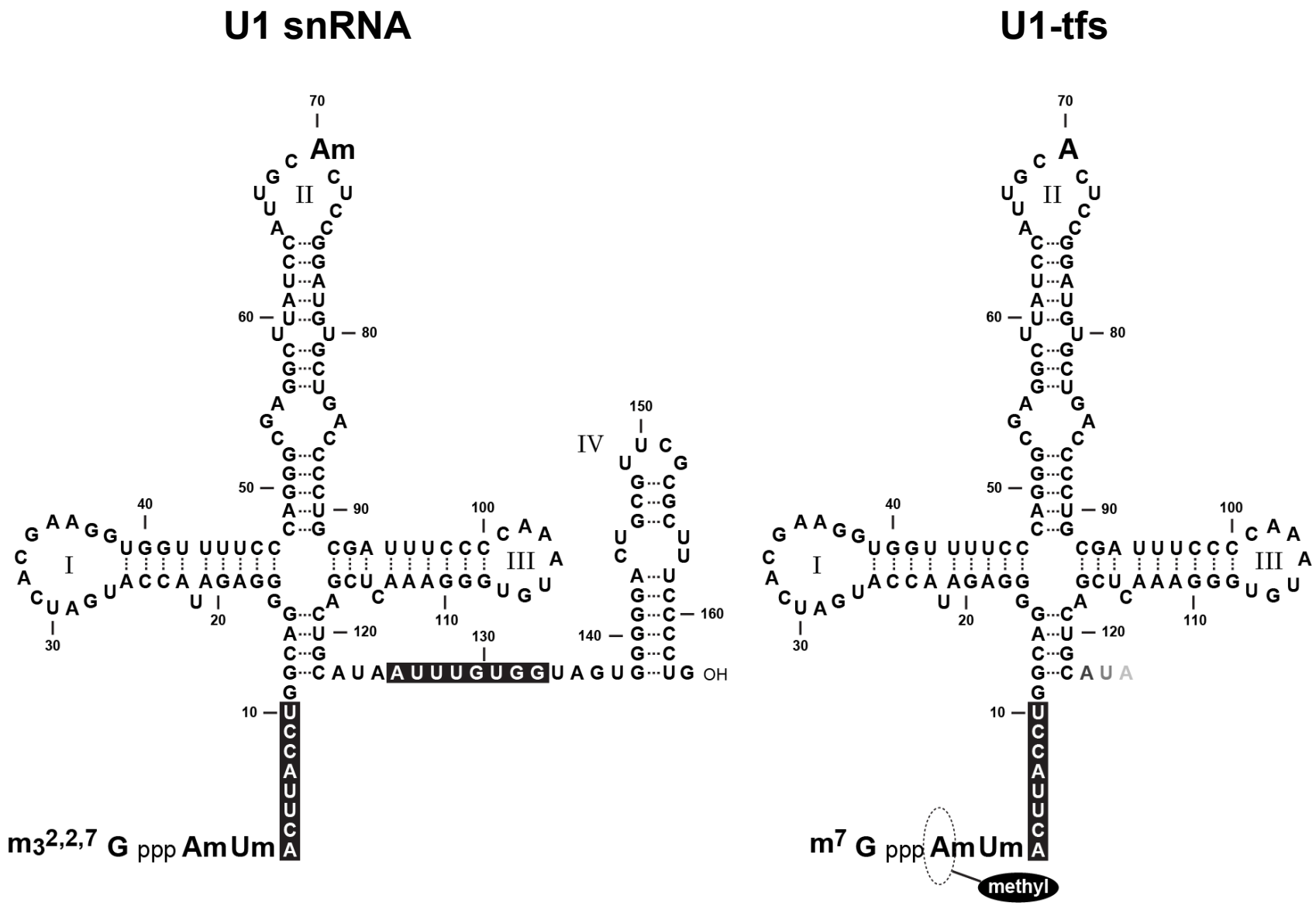


Figure S3

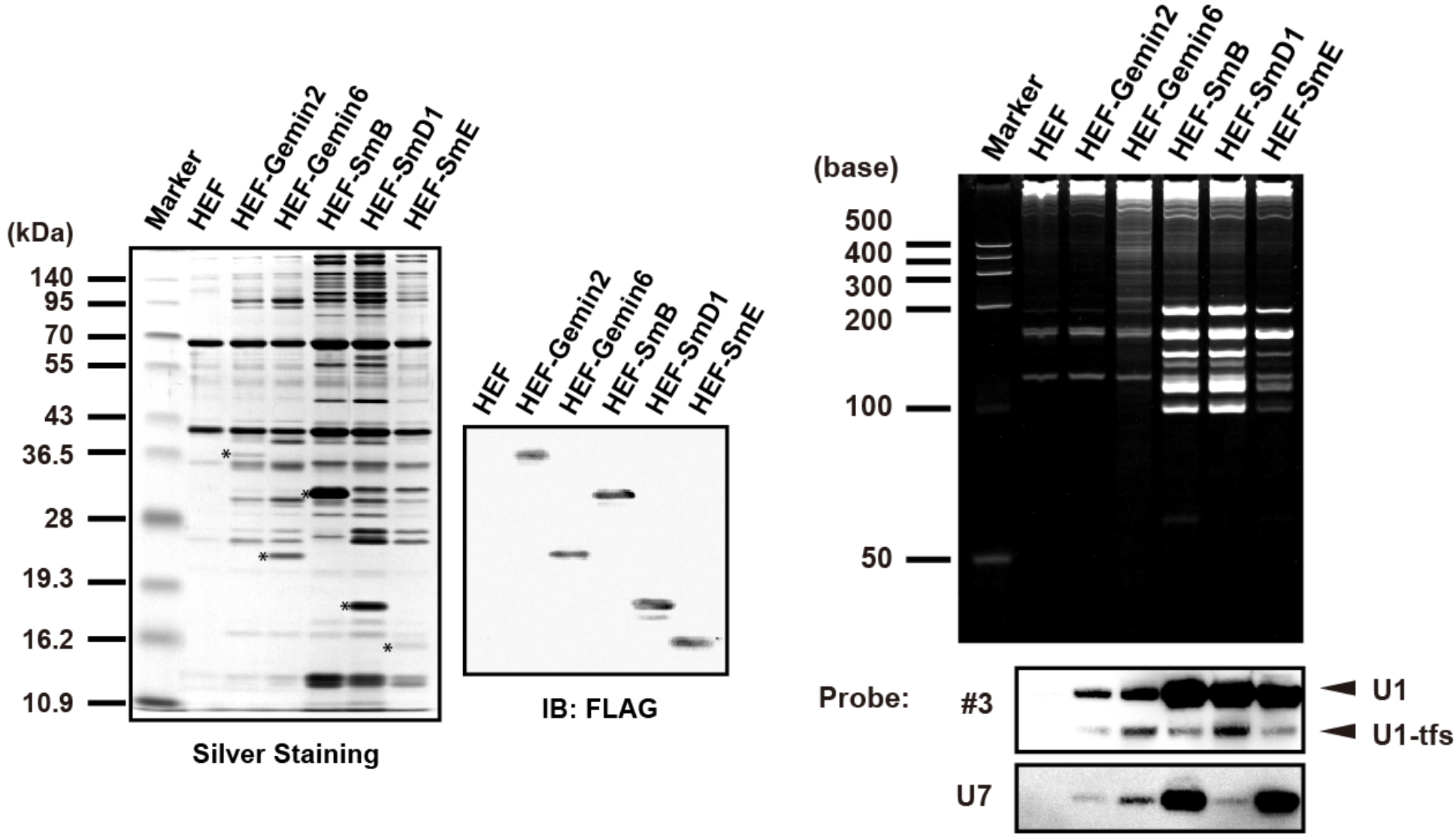
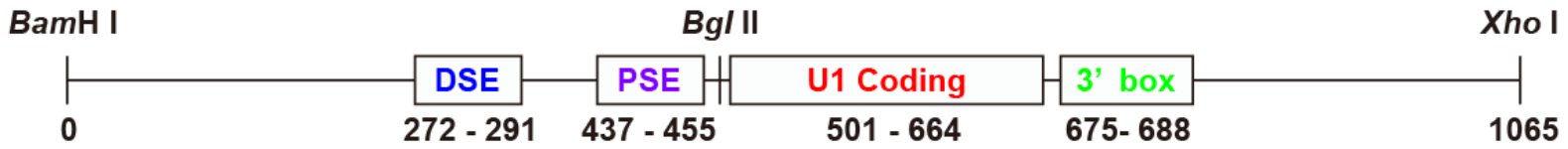

Marker
HEF
HEF-Gemin2
HEF-Gemin6
HEF-SmB
HEF-SmD1
HEF-SmE
(kDa)
140
95
70
55
43
36.5
28
19.3
16.2
10.9
Silver Staining
IB: FLAG
(base)
500
400
300
200
100
50
Probe:
#3
U1
U1-tfs
U7

Figure S4A

Schematic diagram of exogenously express U1 vector



1 50 100

5' GTTTCTTTTGTAATCCGAAACATTCTAGTCTGCGAATTAAAAGCCATTATTTGAAGAAGGATGCCCCGGCTCCATCTGGCCACCGAAAGGTTGCTCCTT

AACACAGGCTAAGGACCAGCTTCTTTGGGAGAGAACAGACGCAGGGGCGGGAGGGAAAAAGGGAGAGGCAGACGTCACTTCCCCTTGGCGGCTCTGGCA

GCAGATTGGTCGGTTGAGTGGCAGAAAGGCAGACGGGGACTGGGCAAGGCACTGTCGGTGACATCACGGACAGGGCGACTTCTATGTAGATGAGGCAGC

GCAGAGGCTGCTGCTTCGCCACTTGCTGCTTCGCCACGAAGGAGTTCCCGTGCCCTGGGAGCGGGTTCAGGACCGCGGATCGGAAGTGAGAATCCCAGC

TGTGTGTCAGGGCTGGAAAGGGCTCGGGAGTGCGCGGGGCAAGTGACCGTGTGTGTAAAGAGTGAGGCGTATGAGGCTGTGTCGGGGCAGAGCCCGAAG

ATCTCATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGA

TTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTGACTTTCTGGAGTTTCAAAAACAGACCGTA

CGCCAAGGGTCATGTCTTTTTTCGTATTGGTTTGTGTCTTAGTTGTTAATCCTACAGTGGAGGCCTGGGGAATAAGAAGTAACATGTGGCCTGCACGCC

ATAGGAGAAAAAGCGAGCATCAGCCGTATCGGCTTTGTAACACAAATTAGCTATCGTGAAGTCCGCTCAGCTCTTCCATTTCTACCCTGGCTGCTTTTT

GCAGGGATTGGTCCGTGGTCTCCAGTCTCTTGGGTTCTCACCCTGTGTGAAAATCTTCGTGTTTTTCCCTACCCCCCAAGTCACCTCTTACACAGCCTC

TGCTTCCAAGCGCAGCCCCCACAGGAGTTTGTAGGATTTCTGTGCTAGCGGGGAGTGTGTTCTCACCTCATAGAG 3'

Figure S4B

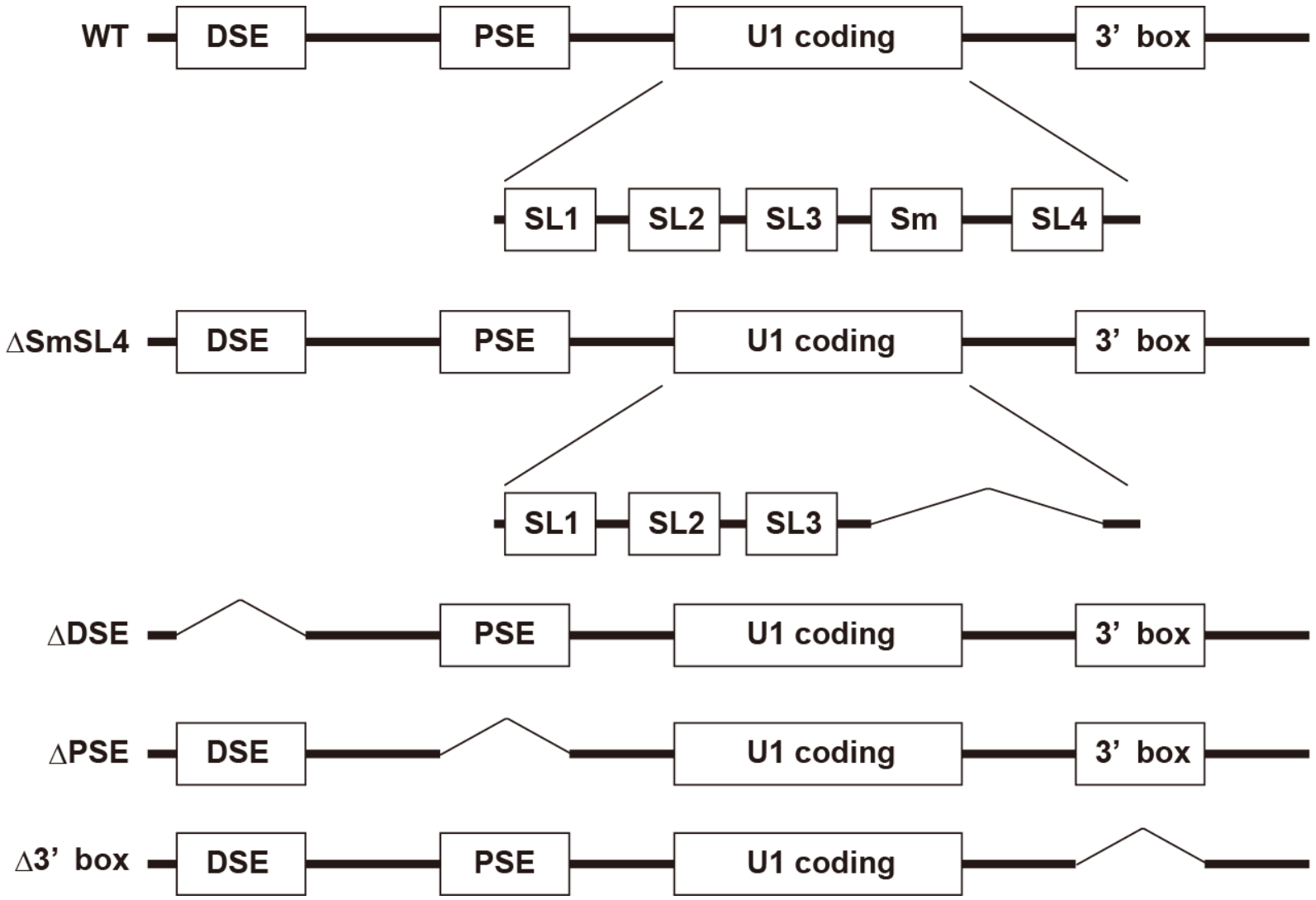
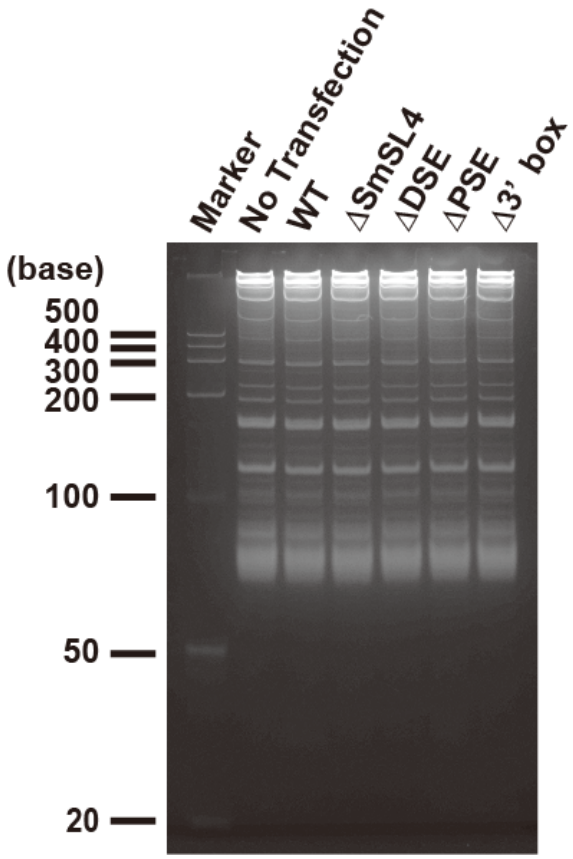
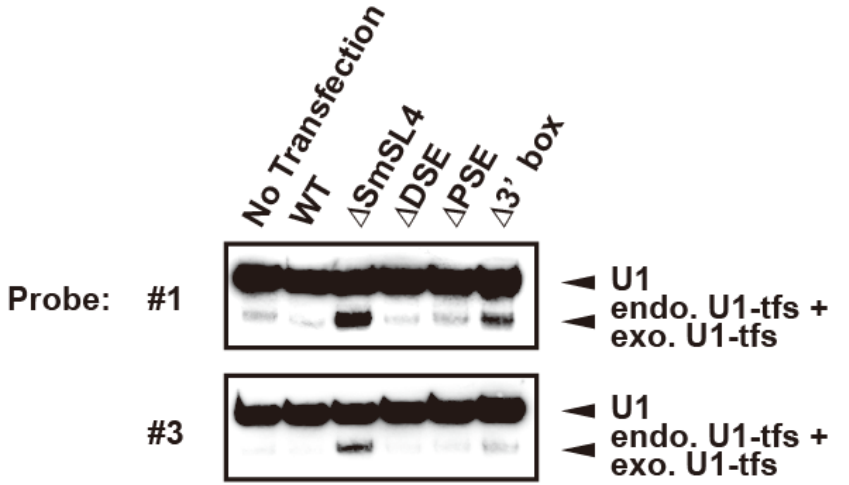
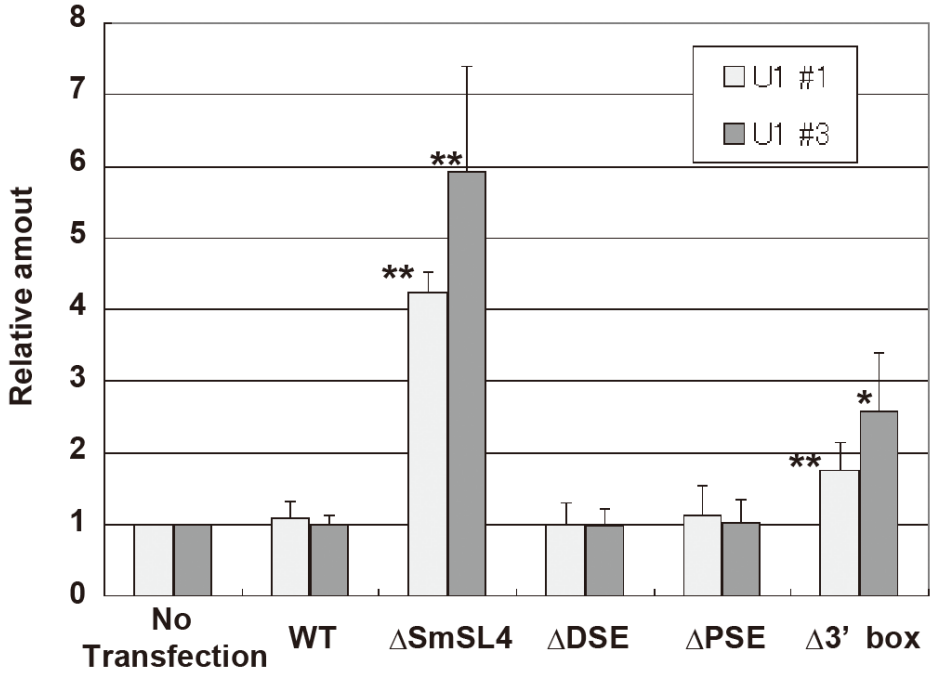

WT
DSE
PSE
U1 coding
3' box
SL1
SL2
SL3
Sm
SL4
ΔSmSL4
DSE
PSE
U1 coding
3' box
SL1
SL2
SL3
ΔDSE
PSE
U1 coding
3' box
ΔPSE
DSE
U1 coding
3' box
Δ3' box
DSE
PSE
U1 coding

Figure S4C







*: p<0.1 **: p<0.05

Figure S4D

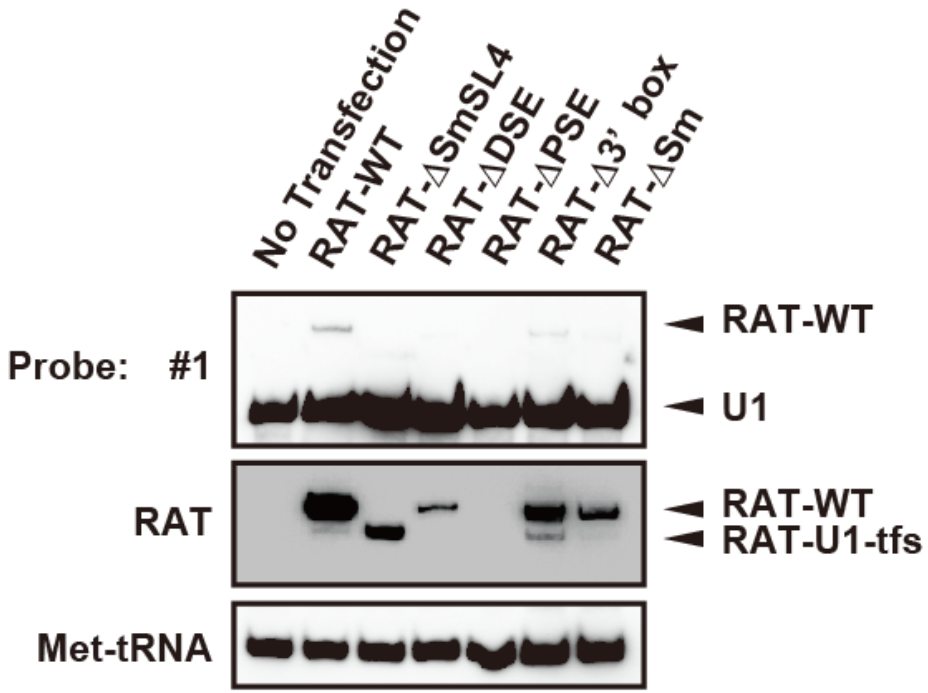

No Transfection
RAT-WT
RAT-ΔSmSL4
RAT-ΔDSE
RAT-ΔPSE
RAT-Δ3' box
RAT-ΔSm
Probe: #1
RAT-WT
U1
RAT
RAT-WT
RAT-U1-tfs
Met-tRNA

Figure S4E

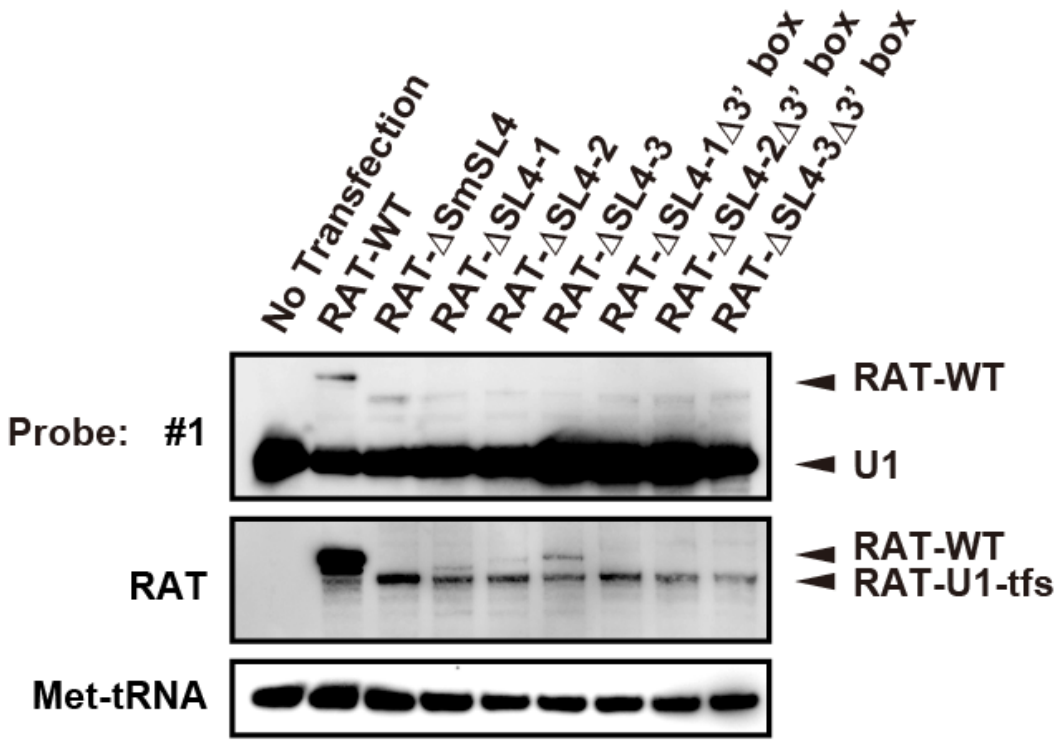

No Transfection
RAT-WT
RAT-ΔSmSL4
RAT-ΔSL4-1
RAT-ΔSL4-2
RAT-ΔSL4-3
RAT-ΔSL4-1Δ3' box
RAT-ΔSL4-2Δ3' box
RAT-ΔSL4-3Δ3' box
Probe: #1
RAT-WT
U1
RAT
RAT-WT
RAT-U1-tfs
Met-tRNA

Figure S5A

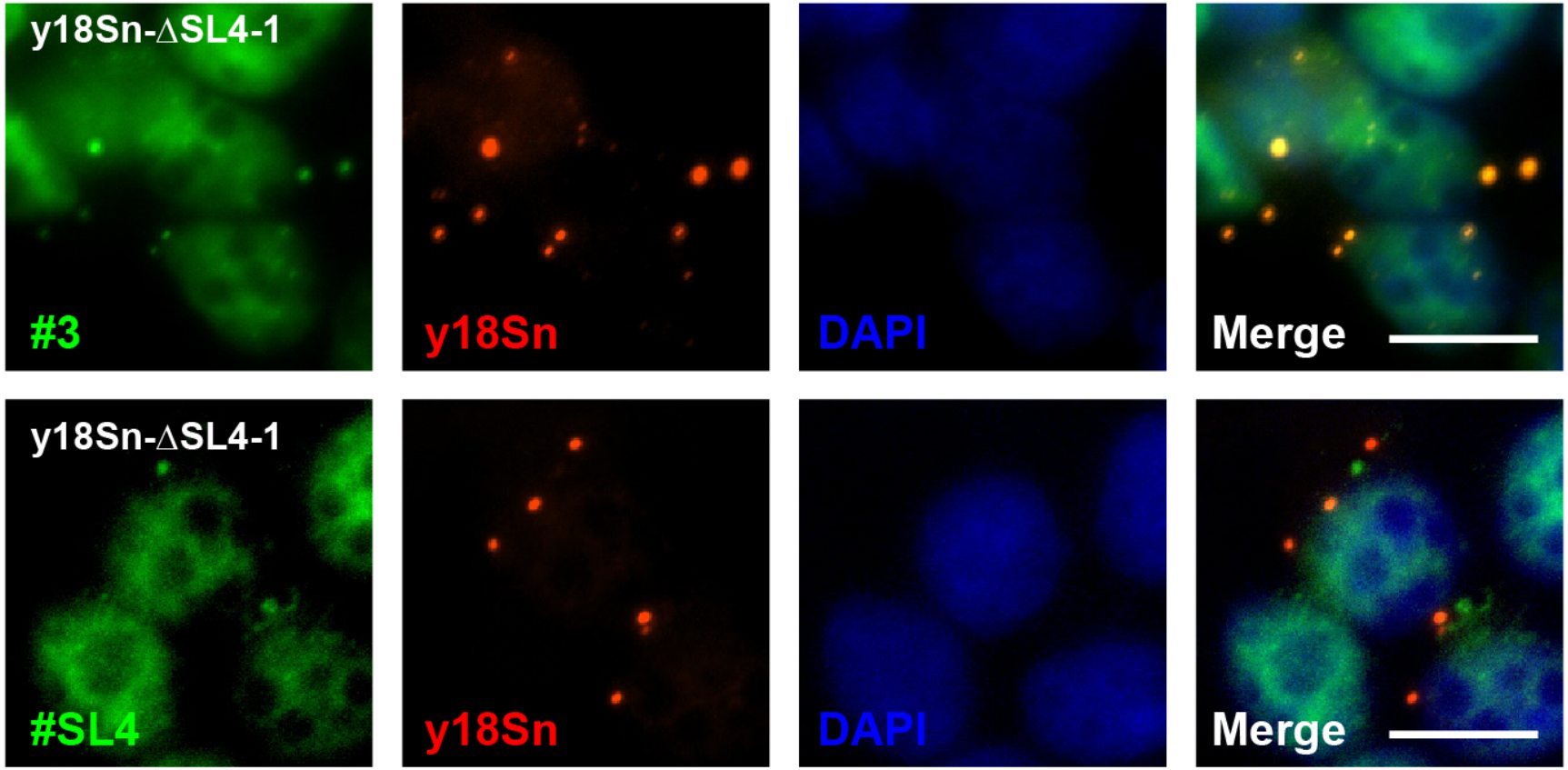

y18Sn-ΔSL4-1
#3
y18Sn
DAPI
Merge
y18Sn-ΔSL4-1
#SL4
y18Sn
DAPI
Merge

Figure S5B

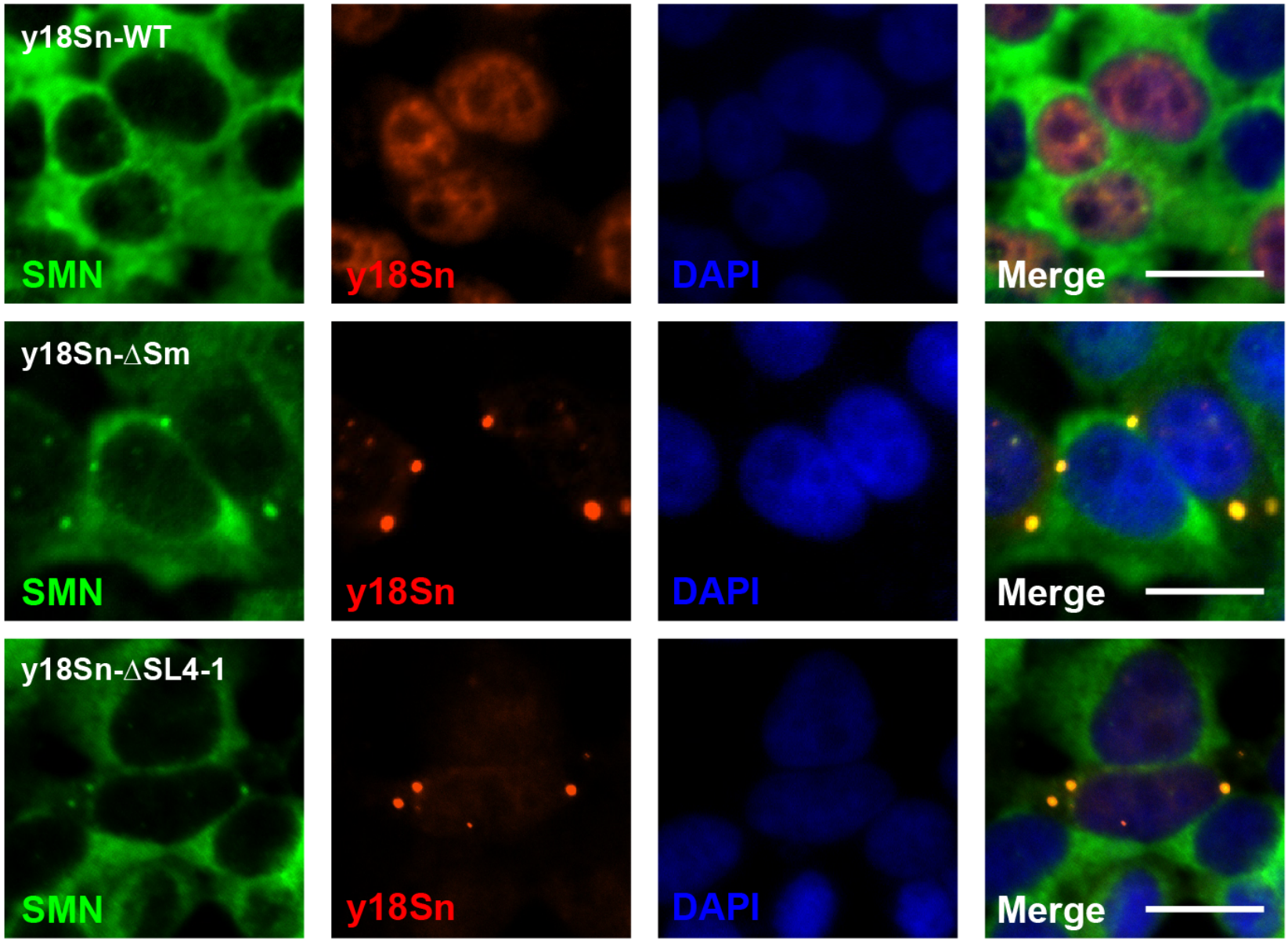

y18Sn-WT
SMN
y18Sn
DAPI
Merge
y18Sn-ΔSm
SMN
y18Sn
DAPI
Merge
y18Sn-ΔSL4-1
SMN
y18Sn
DAPI
Merge

Figure S5C

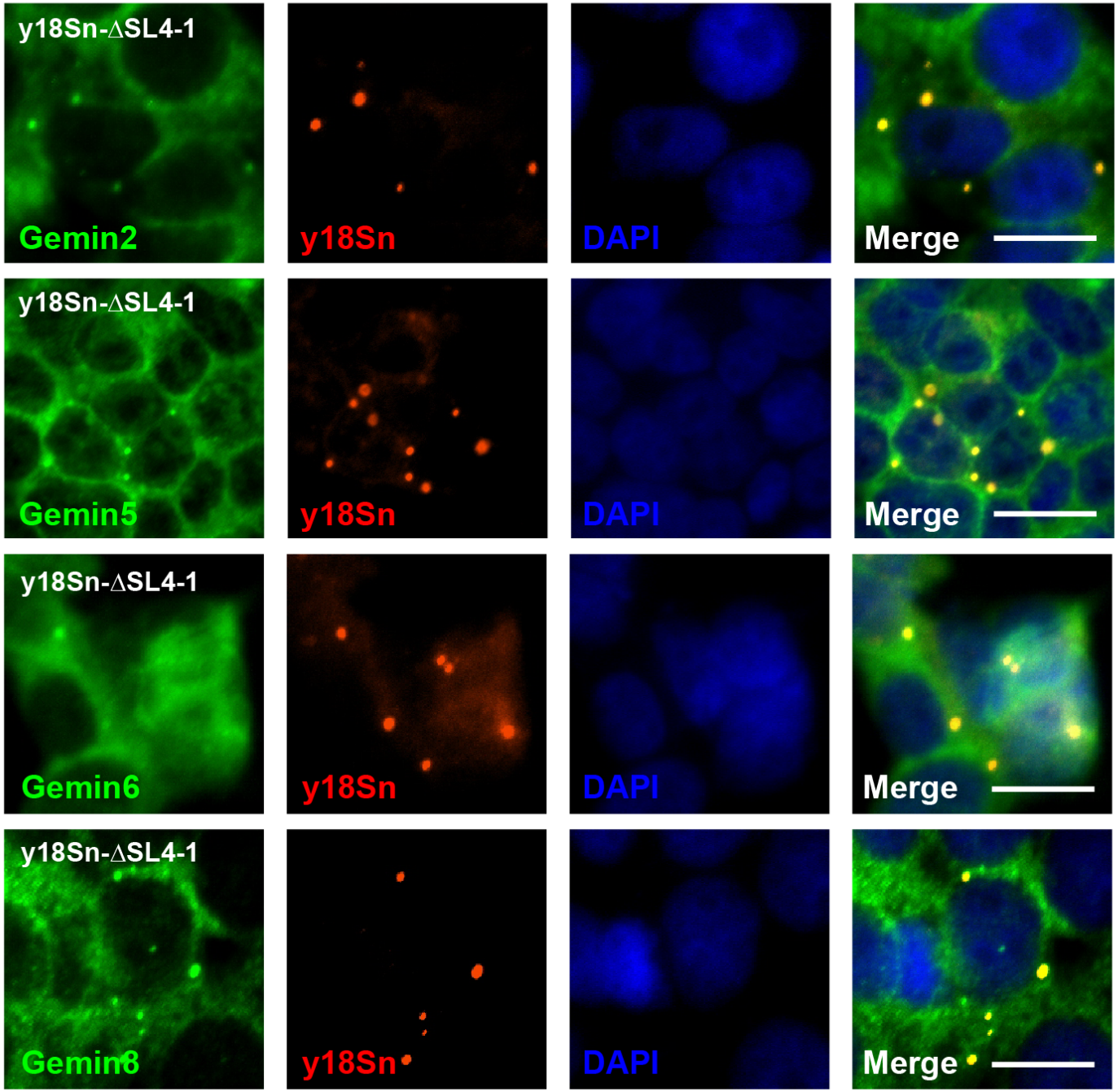

y18Sn-ΔSL4-1
Gemin2
y18Sn
DAPI
Merge
y18Sn-ΔSL4-1
Gemin5
y18Sn
DAPI
Merge
y18Sn-ΔSL4-1
Gemin6
y18Sn
DAPI
Merge
y18Sn-ΔSL4-1
Gemin8
y18Sn
DAPI
Merge

Figure S6A

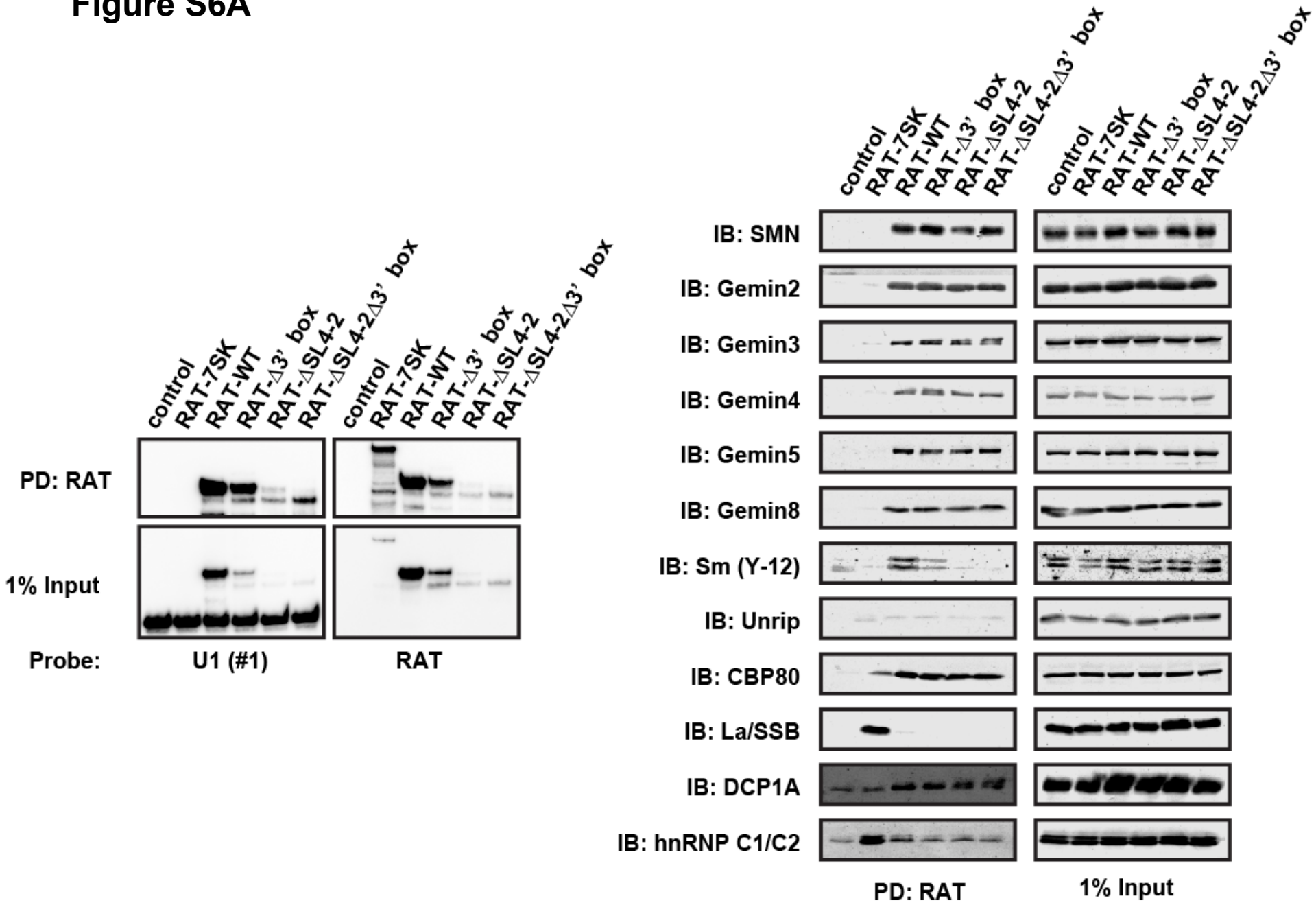

control
RAT-7SK
RAT-WT
RAT-Δ3' box
RAT-ΔSL4-2
RAT-ΔSL4-2Δ3' box
PD: RAT
1% Input
Probe:
U1 (#1)
RAT
IB: SMN
IB: Gemin2
IB: Gemin3
IB: Gemin4
IB: Gemin5
IB: Gemin8
IB: Sm (Y-12)
IB: Unrip
IB: CBP80
IB: La/SSB
IB: DCP1A
IB: hnRNP C1/C2
PD: RAT
1% Input

Figure S6B

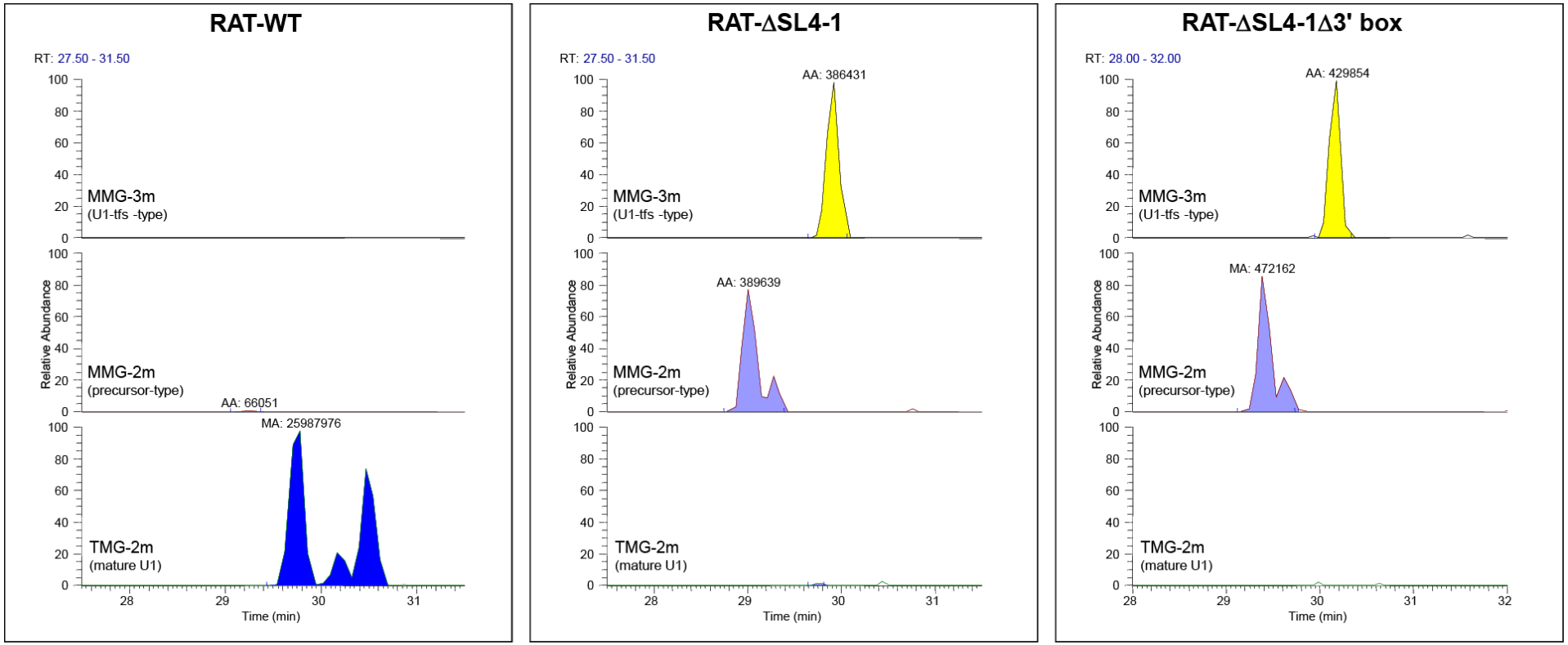
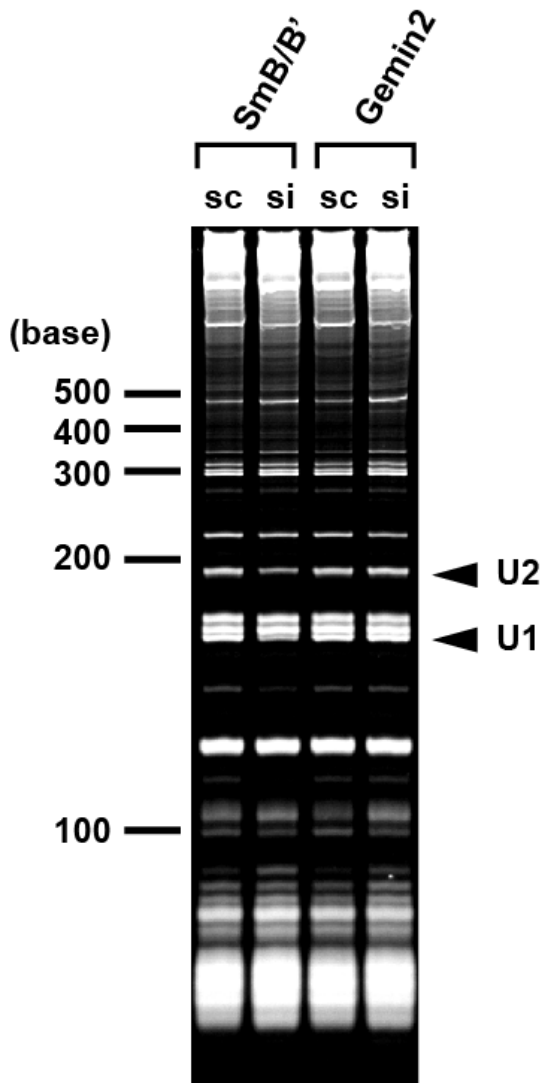
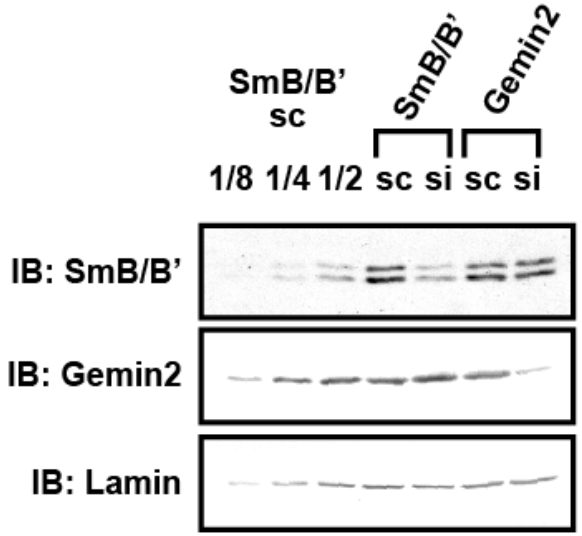

RAT-WT
RT: 27.50 - 31.50
MMG-3m
(U1-tfs -type)
MMG-2m
(precursor-type)
AA: 66051
MA: 25987976
TMG-2m
(mature U1)
Relative Abundance
Time (min)
RAT-ΔSL4-1
RT: 27.50 - 31.50
AA: 386431
AA: 389639
RAT-ΔSL4-1Δ3' box
RT: 28.00 - 32.00
AA: 429854
MA: 472162

Figure S7A


SmB/B'
Gemin2
sc si sc si
(base)
500
400
300
200
100
U2
U1


SmB/B' sc
SmB/B'
Gemin2
1/8 1/4 1/2 sc si sc si
IB: SmB/B'
IB: Gemin2
IB: Lamin

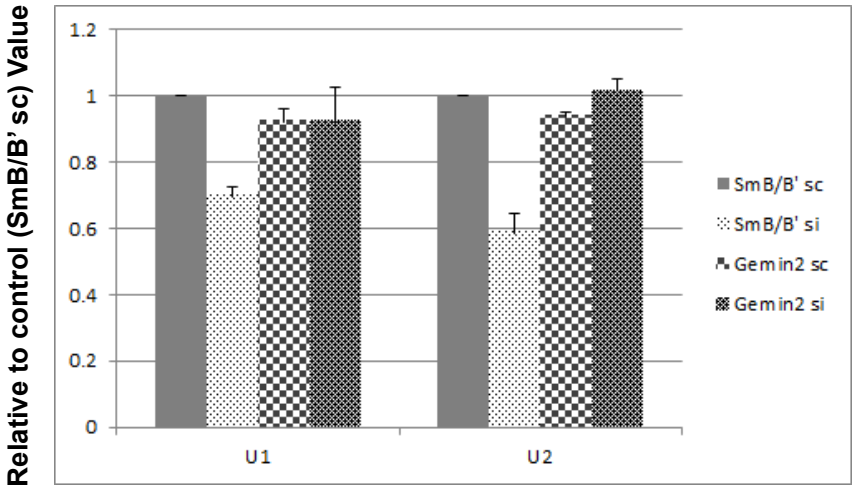

Relative to control (SmB/B' sc) Value
1.2
1
0.8
0.6
0.4
0.2
0
U1
U2
SmB/B' sc
SmB/B' si
Gemin2 sc
Gemin2 si

Figure S7B

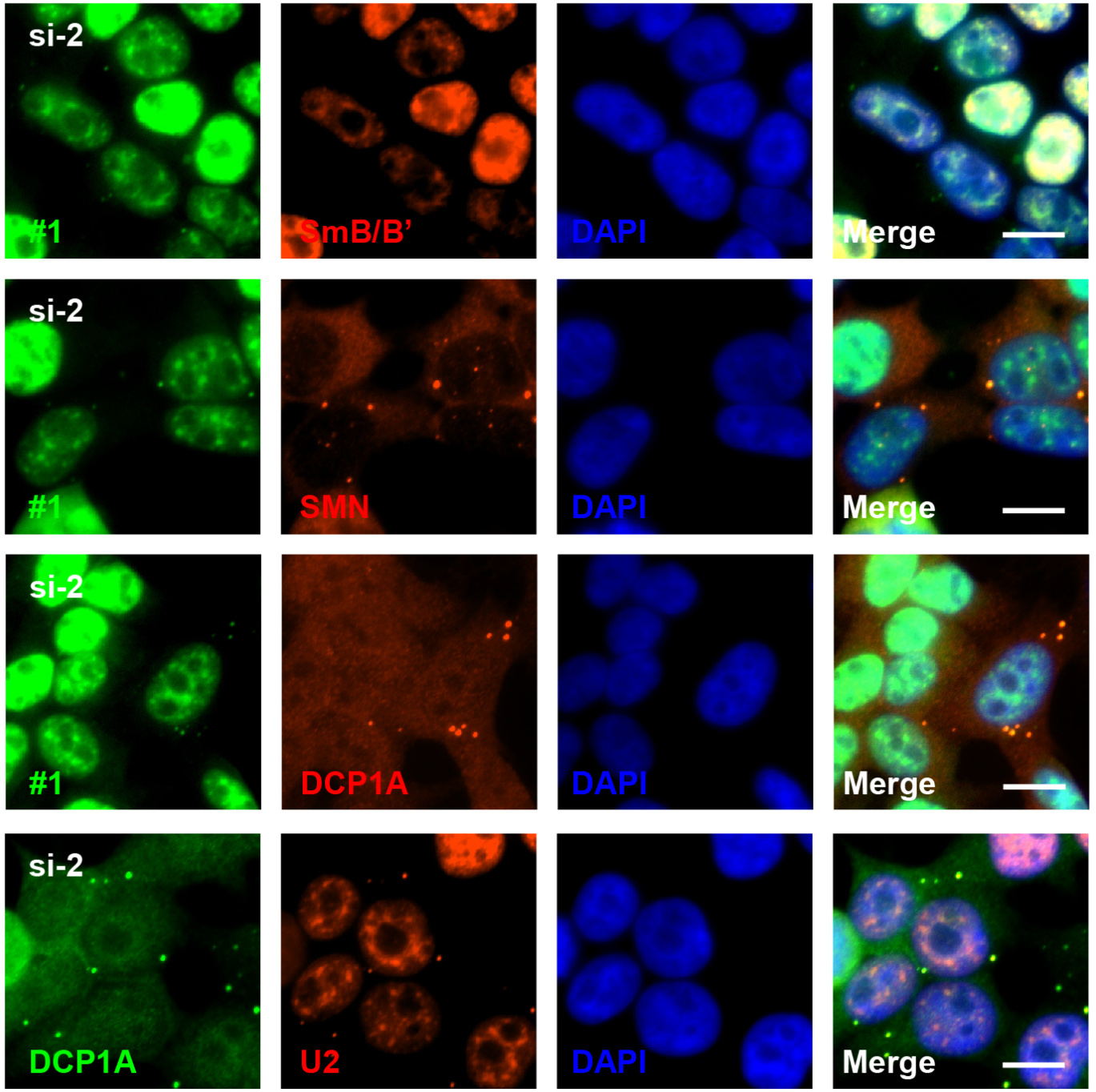

si-2
#1
SmB/B'
DAPI
Merge
si-2
#1
SMN
DAPI
Merge
si-2
#1
DCP1A
DAPI
Merge
si-2
DCP1A
U2
DAPI
Merge

Figure S7C

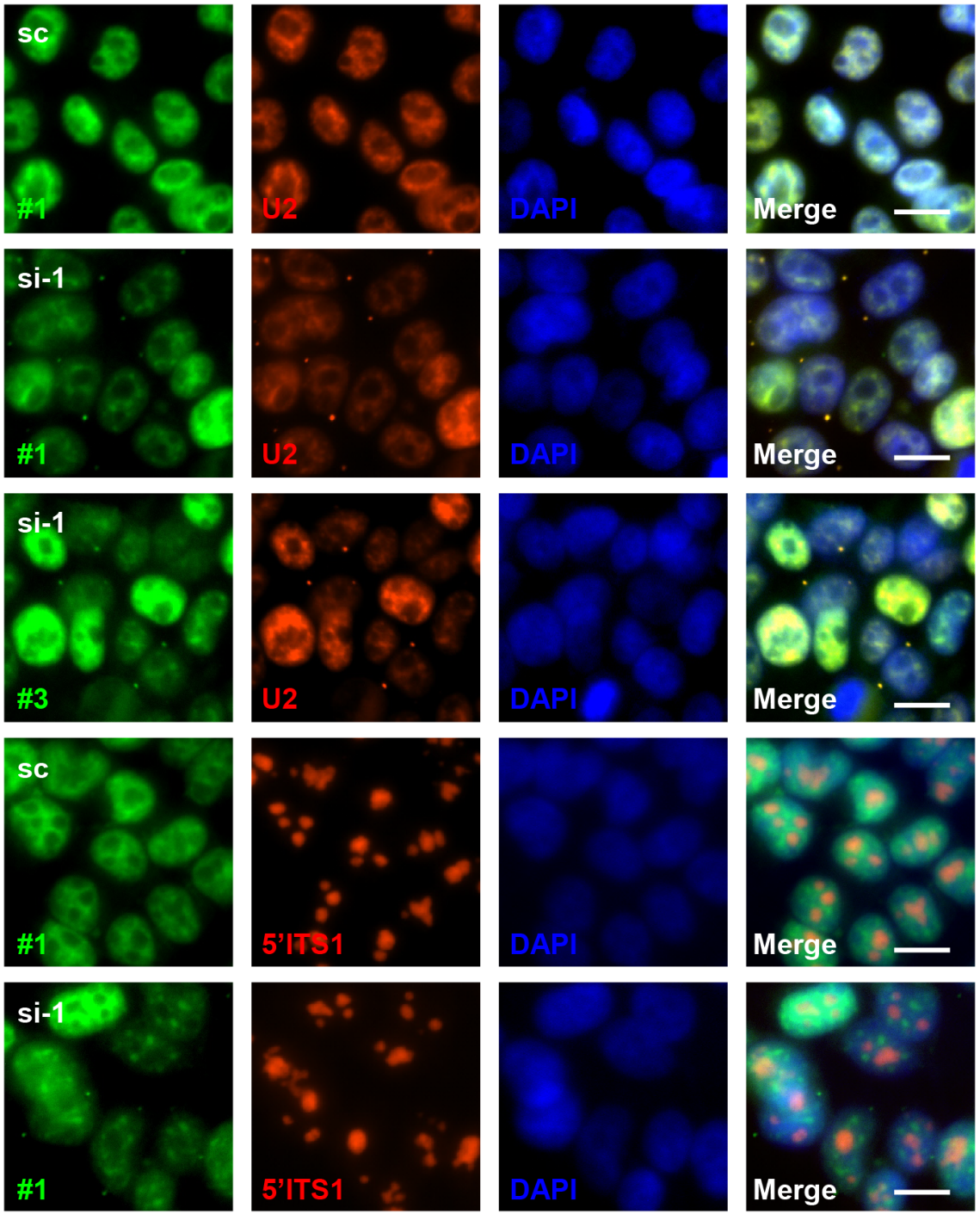

sc
#1
U2
DAPI
Merge
si-1
#1
U2
DAPI
Merge
si-1
#3
U2
DAPI
Merge
sc
#1
5'ITS1
DAPI
Merge
si-1
#1
5'ITS1
DAPI
Merge

Figure S7D

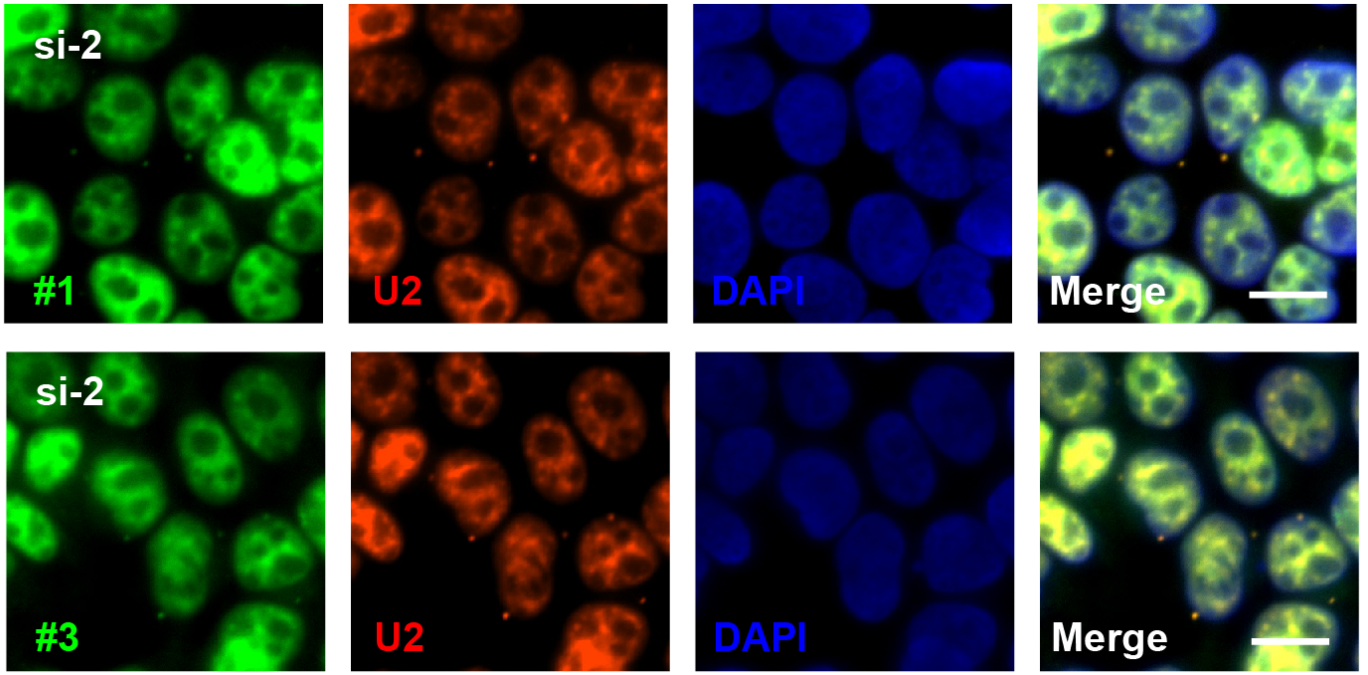
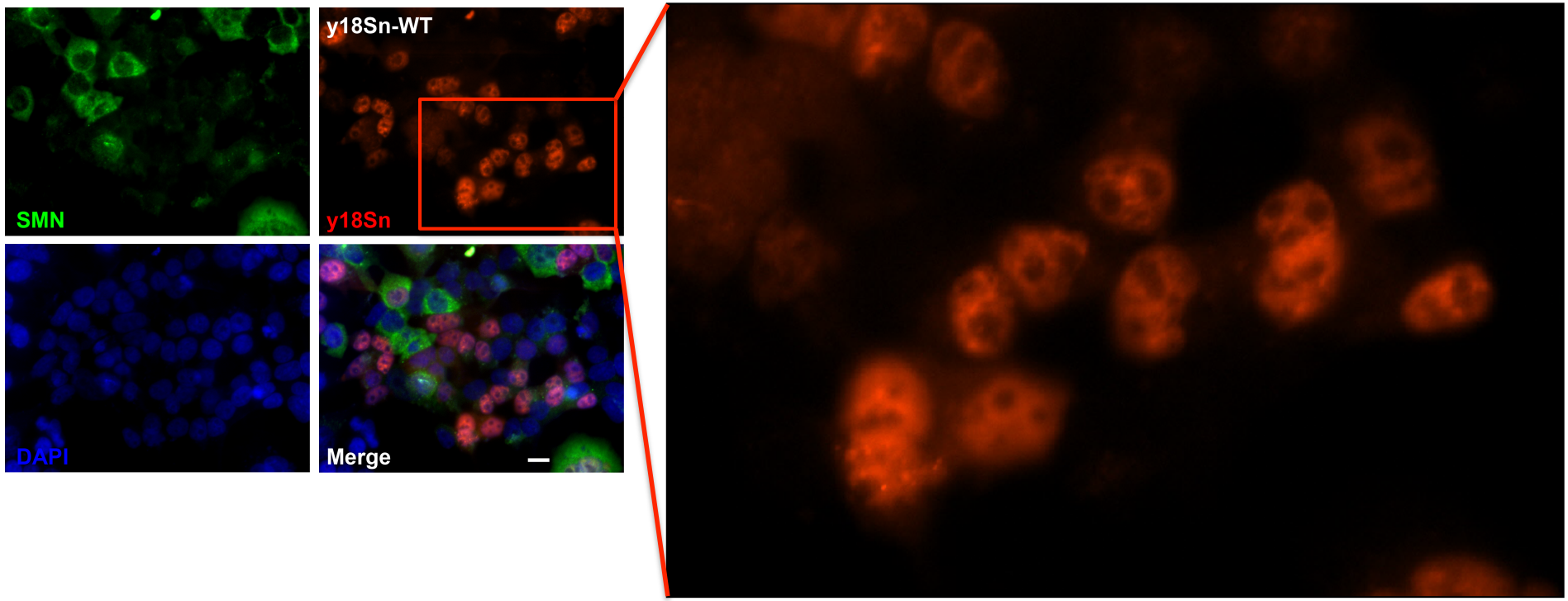


Figure S7E



Supplementary Table 1. RNase T1 fragments of U snRNAs identified by LC-MS analysis.

# of fraction	RNA name	Sequence	Residue number	Identified sequences	Retention time (min)	Intensity	Observed m/z	Observed molecular mass	Theoritical molecular mass	Δmass (ppm)	Charge	a-series ions identified	c-series ions identified	w-series ions identified	y-series ions identified	Identification method	Remarks
			1-4	m$_3^{2,2,7}$GpppAGUG>p + 2 methyl	32.08	32200	949.10510	1900.2259	1900.2310	-2.7	-2	-	c3	-	-	manually identified	M-B(m$_3^{2,2,7}$G), M-B(m$_3^{2,2,7}$G)-H$_2$O, M-m$_3^{2,2,7}$Gp, M-m$_3^{2,2,7}$Gpp, m$_3^{2,2,7}$Gp and m$_3^{2,2,7}$Gpp were identified in the MS/MS spectrum. The trimethyguanine structure was also identified by the previous study[1].
			1-4	m$_3^{2,2,7}$GpppAGUGp + 2 methyl	30.59	22100	958.11066	1918.2370	1918.2416	-2.4	-2	-	-	-	-	manually identified	M-B(m$_3^{2,2,7}$G), M-B(m$_3^{2,2,7}$G)-H$_2$O, M-m$_3^{2,2,7}$Gp, M-m$_3^{2,2,7}$Gpp, m$_3^{2,2,7}$Gp and m$_3^{2,2,7}$Gpp were identified in the MS/MS spectrum. The trimethyguanine structure was also identified by the previous study[1].
			5-10	UUACAG>p	31.90	58000	959.11194	1920.2395	1920.2444	-2.5	-2	a3-a5	c2-c5	w1-w3	y1-y5	Ariadne	
			5-10	UUACAGp	30.59	178000	968.11725	1938.2502	1938.2549	-2.5	-2	a4,a6	c2-c6	w1-w3,w5	y1-y5	Ariadne	
			11-19	CUCUUUUAG>p	32.08	50000	1406.14185	2814.2994	2814.3090	-3.4	-2	a3,a5,a6	c2-c7	w5,w6,w8	y2-y8	Ariadne	
			11-19	CUCUUUUAGp	30.81	169000	1415.14832	2832.3123	2832.3196	-2.6	-2	a3,a6,a9	c2-c8	w5,w6	y2-y5,y7,y8	Ariadne	
		AGUGUUACAG	20-25	AAUUUG>p	35.97	281000	959.60321	1921.2221	1921.2284	-3.3	-2	a3-a5	c1-c5	w1-w5	y1,y2,y4,y5	Ariadne	
1	U7	CUCUUUUAGA AUUUGUCUAG	20-25	AAUUUGp	34.65	146000	968.60938	1939.2344	1939.2390	-2.3	-2	a3-a6	c2-c6	w1,w5	y1-y5	Ariadne	
		UAGGCUUUCU GGCUUUUUAC	26-30	UCUAG>p	32.76	35500	794.58618	1591.1880	1591.1919	-2.4	-2	a2-a4	c2-c4	w1-w4	y1-y4	Ariadne	
		CGGAAAGCCC CU	26-30	UCUAGp	30.59	41500	803.59155	1609.1988	1609.2024	-2.3	-2	a3-a5	c1-c4	w2,w3	y1-y4	Ariadne	
			35-41	CUUUCUG>p	29.40	41800	1088.60498	2179.2256	2179.2312	-2.6	-2	a4,a5	c2,c4-c6	w2,w4,w6	y2-y6	Ariadne	
			35-41	CUUUCUGp	27.52	11200	1097.61169	2197.2390	2197.2418	-1.3	-2	a5	c2-c6	w2,w3	y2-y6	Ariadne	
			43-52	CUUUUUACCG>p	33.49	11800	1038.77332	3119.3434	3119.3503	-2.2	-3	a8	c2-c5,c7,c9	w2	y2,y3,y5-y7,y9	Ariadne	
			43-52	CUUUUUACCGp	32.76	3420	1044.77637	3137.3526	3137.3609	-2.7	-3					m/z	
			54-57	AAAG>p	36.47	7340	1331.19360	1332.2014	1332.2050	-2.7	-1					m/z	
			54-57	AAAGp	34.52	2410	1349.20422	1350.2120	1350.2156	-2.6	-1					m/z	
			58-62	CCCCU-OH	20.69	9050	731.10889	1464.2334	1464.2347	-0.9	-2					m/z	
			1-3	m$_3^{2,2,7}$GpppAUG>p + 2 methyl	32.78	2220	776.58234	1555.1803	1555.1836	-2.1	-2					m/z	The trimethylguanosine structure was deduced from the previous study[2].
			4-8	CUCUG>p	26.19	11100	782.58087	1567.1774	1567.1806	-2.1	-2	a3,a4	c2-c4	w2,w3	y1-y4	manually identified	
			4-8	CUCUGp	24.02	2310	791.58789	1585.1914	1585.1912	0.2	-2					m/z	
			51-57	AUUUCCG>p	32.94	14300	1100.11829	2202.2522	2202.2584	-2.8	-2	a3,a5,a6	c2,c4,c5	w2,w4	y3-y6	Ariadne	
		AUGCUCUGGU UUCUCUUCAA	51-57	AUUUCCGp	31.32	1620	1109.12354	2220.2627	2220.2690	-2.8	-2					m/z	
		AUCGUAUAAA UCUUUCGCCU	65-70	AAACCG>p	37.32	21700	970.13129	1942.2782	1942.2876	-4.8	-2	a3-a5	c1,c3-c5	w1,w3-w5	y2-y5	Ariadne	
2	U5D	UUUACUAAAG AUUUCCGUGG	71-75	UUUUG>p	28.27	6430	783.56512	1569.1459	1569.1486	-1.8	-2					m/z	
		AGAGAAACCG UUUUGAGUUU	78-84	UUUCAAG>p	35.59	7920	1112.12183	2226.2593	2226.2697	-4.7	-2					m/z	
		CAAGCAAAUU UUUUUGAAGCC	78-84	UUUCAAGp	34.38	10400	1121.12866	2244.2730	2244.2802	-3.2	-2	a4	c2-c6	-	y1-y6	manually identified	
		CU	85-95	CAAAUUUUUUG>p	35.96	12400	1156.78625	3473.3822	3473.3981	-4.6	-3	a7	c2,c4-c10	w3,w4	y2-y4,y6-y9	Ariadne	
			85-95	CAAAUUUUUUGp	35.15	4240	1162.79297	3491.4024	3491.4087	-1.8	-3					m/z	
			96-98	AAG>p	34.10	28200	1002.14117	1003.1490	1003.1525	-3.5	-1	a2	c1,c2	w1,w2	y1,y2	manually identified	
			96-98	AAGp	31.72	8340	1020.15448	1021.1623	1021.1630	-0.7	-1					m/z	
			2-7	UCUACG>p	29.79	122000	947.10516	1896.2260	1896.2331	-3.8	-2	a2,a3,a5,a6	c1-c5	w1,w2,w4,w5	y1-y5	Ariadne	
			2-7	UCUACGp	28.22	4880	956.11224	1914.2401	1914.2437	-1.9	-2					m/z	
			9-21	CCAUACCACCCUG>p	31.95	105000	1358.83752	4079.5360	4079.5446	-2.1	-3	a6,a10	c2-c6,c8-c12	w2-w4,w6-w8,w10,w11	y2-y9,y11,y12	Ariadne	
			9-21	CCAUACCACCCUGp	31.54	8800	1364.84094	4097.5463	4097.5552	-2.2	-3					m/z	
			22-25	AACG>p	34.83	46500	653.08698	1308.1896	1308.1938	-3.2	-2	a2-a4	c1,c3	w1-w3	y1-y3	Ariadne	
			28-31	CCCG>p	26.04	4850	629.07745	1260.1706	1260.1713	-0.6	-2					m/z	
			32-37, 42-47	AUCUCG>p	37.14	510000	947.10547	1896.2266	1896.2331	-3.5	-2	a3-a5	c2-c5	w1-w5	y1-y5	Ariadne	
			32-37, 42-47	AUCUCGp	35.78	33100	956.11163	1914.2389	1914.2437	-2.5	-2	a3-a6	c1-c5	w1-w4	y1-y5	Ariadne	
			38-41	UCUG>p	24.97	10200	1261.12915	1262.1370	1262.1393	-1.9	-1	a3	c2,c3	w2	y2,y3	Ariadne	
		GUCUACGGCC	49-51	AAG>p	34.10	28200	1002.14117	1003.1490	1003.1525	-3.5	-1	a2	c1,c2	w1,w2	y1,y2	manually identified	
		AUACCACCCU GAACGCGCCC	49-51	AAGp	31.72	8340	1020.15448	1021.1623	1021.1630	-0.7	-1					m/z	
		GAUCUCGUCU GAUCUCGAAA	52-56	CUAAG>p	30.97	40100	806.09979	1614.2152	1614.2191	-2.4	-2	a2,a4	c2-c4	w1-w4	y2-y4	Ariadne	
2	5S	GCUAAGCAGG GUCGGGCCUG	52-56	CUAAGp	28.91	10400	815.10504	1632.2257	1632.2296	-2.4	-2	a4	c1-c4	w1	y1,y2,y4	manually identified	
		GUUAGUACUU GGAUGGGAGA	67-70, 94-97	CCUG>p	25.44	7020	629.56866	1261.1530	1261.1553	-1.9	-2					m/z	
		CCGCCUGGGA AUACCGGGUG	72-75	UUAG>p	29.04	9030	1285.14050	1286.1483	1286.1506	-1.7	-1					m/z	
		CUGUAGGCUU	72-75	UUAGp	26.79	6670	1303.15039	1304.1582	1304.1611	-2.2	-1					m/z	
			76-81	UACUUG>p	31.32	171000	947.59869	1897.2130	1897.2172	-2.2	-2	a3,a4	c2-c5	w1-w5	y1-y5	Ariadne	
			76-81	UACUUGp	30.46	29200	956.60394	1915.2235	1915.2277	-2.2	-2	a3-a6	c2-c6	w2,w3	v1-v5	Ariadne	
			83-85	AUG>p	28.34	20700	979.11493	980.1228	980.1253	-2.6	-1	-	c2	w1,w2	y1,y2	manually identified	
			83-85	AUGp	26.09	1230	997.13025	998.1381	998.1358	2.3	-1					m/z	
			90-93	ACCG>p	29.79	11900	641.08124	1284.1781	1284.1825	-3.4	-2					m/z	
			100-106	AAUACCG>p	37.54	129000	1123.14343	2248.3025	2248.3129	-4.6	-2	a4-a6	c2-c6	w1-w3,w5,w6	y2-y6	Ariadne	
			100-106	AAUACCGp	36.55	10800	1132.14893	2266.3135	2266.3234	-4.4	-2	a5	c2-c6	w2,w5,w6	y2-y6	manually identified	
			114-116	UAG>p	27.07	3860	979.11487	980.1227	980.1253	-2.6	-1					m/z	
			114-116	UAGp	24.52	3500	997.12634	998.1342	998.1358	-1.7	-1					m/z	
			1-6	m$_3^{2,2,7}$GpppACUCUG>p + 2 methyl	34.40	17600	1234.63440	2471.2845	2471.2915	-2.8	-2					m/z	The trimethylguanosine structure was deduced from the previous study[2].
			1-6	m$_3^{2,2,7}$GpppACUCUGp + 2 methyl	33.09	3370	1243.63696	2489.2896	2489.3021	-5.0	-2					m/z	The trimethylguanosine structure was deduced from the previous study[2].
			23-55	UAUAAAUCUUUCGmCCUUmUUACmUAAAGAUUUCCG>p	41.79	1120	1306.7782 [a]	10462.2882 [a]	10462.31482 [a]	-2.5	-8					m/z	The placement of methyl residue was deduced from the homology to U5A and U5B snRNA. This fragment contains one missed cleavage.

4	U5E	ACUCUGGUUUCUCUUCAAAUCGUAUAAAUCUUUCGCCUUUUACUAAAGAUUUCCGUGGAGAGAAACGAGUGUGAGUCUGAAACCAAUUUUUUGAGGCCUUGCGUUUUUAGCAGGGCUU	49-55	AUUUCCG>p	33.33	70700	1100.11829	2202.2522	2202.2584	-2.8	-2	a4-a6	c2-c6	w1-w4,w6	y2-y6	Ariadne	
			49-55	AUUUCCGp	31.56	7960	1109.12256	2220.2608	2220.2690	-3.7	-2					m/z	
			63-67	AAACG>p	36.55	5670	817.61255	1637.2408	1637.2463	-3.4	-2					m/z	
			63-67	AAACGp	34.86	1220	826.61554	1655.2467	1655.2569	-6.1	-2					m/z	
			76-79	UCUG>p	25.31	10500	1261.12744	1262.1353	1262.1393	-3.2	-1	a3	c2,c3	w1,w2	y2,y3	Ariadne	
			76-79	UCUGp	21.79	1140	1279.13452	1280.1423	1280.1499	-5.9	-1					m/z	
			80-93	AAACCAAUUUUUUG>p	44.84	233000	1477.83508	4436.5287	4436.5444	-3.5	-3	-	c2-c6,c10-c12	w7,w8,w10	y2-y12	Ariadne	
			80-93	AAACCAAUUUUUUGp	44.14	51600	1483.83716	4454.5350	4454.5550	-4.5	-3	-	c2-c6,c8-c13	w7	y2-y8,y10,y13	manually identified	
			97-101	CCUUG>p	26.84	28200	782.58093	1567.1775	1567.1806	-2.0	-2	a2-a5	c2-c4	w2-w4	y1-y4	Ariadne	
			97-101	CCUUGp	25.47	2410	791.58472	1585.1851	1585.1912	-3.8	-2					m/z	
4	RNU5E-6P[b]	AUACUCGGGUUUCUCUUCAAAGCGCAUAAAUCUUUCGCCUUUUACUAAAGAUUUCCGUGGAGAGAAAGUUGUGAGUUUUUAUUCAAUUUUUUGAGGCCUCUUAUUUCCUGAGGCUA	1-7	$m_3^{2,2,7}$GpppAUACUCG>p + 2 methyl	38.34	53700	1399.15637	2800.3284	2800.3440	-5.6	-2	-	C3,c4,c6	w2-w4	y3,y4	manually identified	M-B($m_3^{2,2,7}$G), M-B($m_3^{2,2,7}$G)-H_2O, M-$m_3^{2,2,7}$Gp, M-$m_3^{2,2,7}$Gpp and $m_3^{2,2,7}$Gpp were identified in the MS/MS spectrum.
			1-7	$m_3^{2,2,7}$GpppAUACUCGp + 2 methyl	37.37	7250	1408.16406	2818.3438	2818.3546	-3.8	-2					m/z	
			10-22	UUUCUCUUCAAAG>p	34.18	10500	1360.14795	4083.4673	4083.4807	-3.3	-3	-	c2-c4,c6-c8,c10	w2,w3	y3,y4,y6-y8,y10	manually identified	
			10-22	UUUCUCUUCAAAGp	33.33	1920	1366.15479	4101.4878	4101.4912	-0.8	-3					m/z	
			25-57	CAUAAAUCUUUCGmCCUUmUUACmUAAAGAUUUCCG>p	42.29	1420	1306.65023 [a]	10461.26444 [a]	10461.33069 [a]	-6.3	-8					m/z	The placement of methyl residue was deduced from the homology to U5A and U5B snRNA. This fragment contains one missed cleavage.
			51-57	AUUUCCG>p	33.33	70700	1100.11829	2202.2522	2202.2584	-2.8	-2	a4-a6	c2-c6	w1-w4,w6	y2-y6	Ariadne	
			51-57	AUUUCCGp	31.56	7960	1109.12256	2220.2608	2220.2690	-3.7	-2					m/z	
			65-68	AAAG>p	37.06	17000	665.09198	1332.1996	1332.2050	-4.0	-2	a3	c1-c3	w1-w3	y2,y3	manually identified	
			65-68	AAAGp	35.04	1700	674.09686	1350.2094	1350.2156	-4.6	-2					m/z	
			76-93	UUUUUAUUCAAUUUUUUG>p	39.50	15900	1402.87988	5615.5508	5615.5752	-4.3	-4	-	c2-c7,c9-c11	w2,w7	y2-y8,y10,y11	manually identified	
			76-93	UUUUUAUUCAAUUUUUUGp	38.79	5050	1407.38538	5633.5728	5633.5858	-2.3	-4					m/z	
			97-110	CCUCUUAUUUCCUG>p	33.87	12200	1446.14771	4341.4666	4341.4835	-3.9	-3	-	C3-C7,c9,c10,c12,c13	w3	y2,y4-y6,y10,y11,y13	manually identified	
			97-110	CCUCUUAUUUCCUGp	33.33	1580	1452.15344	4359.4838	4359.4941	-2.4	-3					m/z	
			114-116	CUA-OH	29.17	21300	877.15289	878.1607	878.1633	-3.0	-1	-	c2	w1,w2	y2	manually identified	
5	U11	AAAGGGCUUCUGUCGUGAGUGGCACACGUAGGGCAACUCGAUUGCUCUGCGUGCGGAAUCGACAUCAAGAGAUUUCGGAAGCAUAAUUUUUUGGUAUUUGGGCAGCUGGUGAUCGUUGGUCCCGGCGCCCUU	7-12	CUUCUG>p	26.93	12200	935.59290	1873.2015	1873.2059	-2.4	-2	a3	c2-c5	w2,w4,w5	y2-y5	manually identified	
			7-12	CUUCUGp	25.19	7610	944.59863	1891.2129	1891.2165	-1.9	-2					m/z	
			23-28	CACACG>p	31.16	15200	958.12579	1918.2672	1918.2763	-4.7	-2	a3,a5	c2-c5	w2-w5	y1-y5	Ariadne	
			34-40	CAACUCG>p	32.87	15100	1111.13879	2224.2932	2224.3016	-3.8	-2	a4,a6	c3-c6	w4-w6	y2-y6	Ariadne	
			34-40	CAACUCGp	31.69	4320	1120.14587	2242.3074	2242.3122	-2.1	-2					m/z	
			41-44	AUUG>p	30.37	2370	1285.13550	1286.1433	1286.1506	-5.6	-1					m/z	
			45-49	CUCUG>p	25.62	11800	782.58051	1567.1767	1567.1806	-2.5	-2	a3,a4	c3,c4	w2-w4	y2-y4	Ariadne	
			57-61	AAUCG>p	34.75	15000	806.09949	1614.2146	1614.2191	-2.7	-2					m/z	
			57-61	AAUCGp	33.32	2140	815.10486	1632.2254	1632.2296	-2.6	-2					m/z	
			62-69	ACAUCAAG>p	36.61	9820	1287.66992	2577.3555	2577.3654	-3.8	-2	a5	c3-c6	w2,w3,w5	y3-y6	Ariadne	
			62-69	ACAUCAAGp	35.66	9730	1296.67407	2595.3638	2595.3760	-4.7	-2					m/z	
			72-77	AUUUCG>p	32.87	14900	947.59741	1897.2105	1897.2172	-3.5	-2	a5	c2-c5	w1-w3	y1-y5	Ariadne	
			72-77	AUUUCGp	30.94	4920	956.60211	1915.2199	1915.2277	-4.1	-2					m/z	
			79-81	AAG>p	33.32	14100	1002.14069	1003.1485	1003.1525	-4.0	-1	-	c2	w1,w2	y1,y2	manually identified	
			79-81	AAGp	30.75	13900	1020.15186	1021.1597	1021.1630	-3.3	-1	-	c1,c2	w1	y2	manually identified	
			82-93	CAUAAUUUUUUG>p	36.09	10500	1258.79578	3779.4108	3779.4234	-3.3	-3	-	c3-c6,c8-c11	-	y2,y3,y5-y7,y10,y11	manually identified	
			82-93	CAUAAUUUUUUGp	35.19	9240	1264.79895	3797.4203	3797.4340	-3.6	-3					m/z	
			95-100	UAUUUG>p	30.94	47000	948.08899	1898.1936	1898.2012	-4.0	-2	a3,a4	c2-c5	w2-w4	y1-y5	Ariadne	
			112-115	AUCG>p	29.79	12300	641.57336	1285.1624	1285.1665	-3.3	-2	-	c1,c2	w1,w2	y1,y3	manually identified	
			120-124	UCCCG>p	28.82	9060	782.08844	1566.1925	1566.1966	-2.6	-2					m/z	
			128-132	CCCUU-OH	23.32	5680	731.59979	1465.2152	1465.2187	-2.4	-2					m/z	
5	U5B	AUACUCUGGUUUCUCUUCAGAUCGUAUAAAUCUUUCGCCUUUUACUAAAGAUUUCCGUGGAGAGGAACAACUCUGAGUCUUAAGCUAAUUUUUUGAGGCCUUGUUCCGACAAGGCUA	1-8	$m_3^{2,2,7}$GpppAmUmACUCUG>p	34.97	15100	1034.44421	3106.3561	3106.3693	-4.3	-3	-	-	w2	y2,y3	manually identified	M-B($m_3^{2,2,7}$G), M-B($m_3^{2,2,7}$G)-H_2O, M-$m_3^{2,2,7}$Gp and M-$m_3^{2,2,7}$Gpp were identified in the MS/MS spectrum. The trimethyguanine structure was also identified by the previous study [1]. The The trimethyguanine structure was reported in the previous study [1]. The placement of methyl residues were deduced from the previous study [1].
			1-8	$m_3^{2,2,7}$GpppAmUmACUCUGp	34.13	9080	1040.44775	3124.3667	3124.3799	-4.2	-3					m/z	
			10-20	UUUCUCUUCAG>p	29.79	8000	1140.78015	3425.3639	3425.3756	-3.4	-3					m/z	
			10-20	UUUCUCUUCAGp	28.98	20200	1146.78259	3443.3712	3443.3862	-4.3	-3	a2,a6,a11	c2-c7,c9,c10	w2,w5,w6	y2-y6,y8,y10	Ariadne	
			21-24	AUCG>p	29.79	12300	641.57336	1285.1624	1285.1665	-3.3	-2	-	c1,c2	w1,w2	y1,y3	manually identified	
			25-50	UAUAAAUCUUUCGmCCUUmUUACmUAAAG>p	40.08	1350	1375.48816 [a]	8258.97591 [a]	8259.05396 [a]	-9.5	-6					m/z	The placement of methyl residue was deduced from the previous study [1].
			51-57	AUUUCCG>p	32.12	78400	1100.11658	2202.2488	2202.2584	-4.4	-2	a4-a6	c2-c6	w1,w2,w4-w6	y2-y6	Ariadne	
			51-57	AUUUCCGp	30.37	30200	1109.12256	2220.2608	2220.2690	-3.7	-2	a3,a5-a7	c2-c7	w2,w4-w6	y2-y6	Ariadne	
			66-75	AACAACUCUG>p	36.25	34500	1061.79846	3188.4189	3188.4320	-4.1	-3	a3,a6	c2,c4-c8	w2,w4,w5,w7	y2-y6,y8,y9	Ariadne	
			66-75	AACAACUCUGp	35.50	21500	1067.80200	3206.4295	3206.4426	-4.1	-3	a4,a6	c4-c7	w2,w4,w5	y1,y4-y9	Ariadne	
			78-84	UCUUAAG>p	31.81	21900	1112.12256	2226.2608	2226.2697	-4.0	-2	a3	c2-c6	w1,w2,w5	y2-y6	Ariadne	
			78-84	UCUUAAGp	30.37	57500	1121.12793	2244.2715	2244.2802	-3.9	-2	a2,a3,a5,a7	c2-c6	w2,w5	y1-y6	Ariadne	
			85-95	CUAAUUUUUUG>p	34.54	19400	1149.11206	3450.3597	3450.3709	-3.3	-3	-	c3-c10	w7	y2-y10	Ariadne	
			85-95	CUAAUUUUUUGp	33.45	30300	1155.11426	3468.3663	3468.3815	-4.4	-3	a6,a11	c2,c4-c7,c10	w7	y1-y10	Ariadne	
			99-103	CCUUG>p	25.62	11800	782.58051	1567.1767	1567.1806	-2.5	-2	a2-a4	c3,c4	w2-w4	y2-y4	Ariadne	
			99-103	CCUUGp	23.32	892	791.58630	1585.1883	1585.1912	-1.8	-2					m/z	
			104-108	UUCCG>p	27.18	17600	782.58081	1567.1773	1567.1806	-2.1	-2	a2-a4	c1-c4	w1-w3	y1-y4	Ariadne	
			104-108	UUCCGp	24.48	4120	791.58569	1585.1870	1585.1912	-2.6	-2					m/z	

			109-113	ACAAG>p	35.19	13300	817.61249	1637.2406	1637.2463	-3.5	-2	a4	c1-c4	w2,w3	y1-y4	Ariadne	
			109-113	ACAAGp	33.25	6050	826.61664	1655.2489	1655.2569	-4.8	-2					m/z	
			115-117	CUA-OH	28.42	10100	877.15277	878.1606	878.1633	-3.1	-1	a2,a3	c1	w2,w3	y2	manually identified	
			1-9	m 2,2,7GpppAGAUCUCUG>p + 2 methyl	35.19	27200	1149.46008	3451.4037	3451.4167	-3.8	-3	-	c4-c6,c8	-	y1-y4	manually identified	M-B(m 2,2,7G), M-B(m 2,2,7G)-H 2 O, M-m 2,2,7Gp, M-m 2,2,7Gpp, m 2,2,7Gp and m 2,2,7Gpp were identified in the MS/MS spectrum. The trimethyguanine structure was also identified by the
			1-9	m 2,2,7GpppAGAUCUCUGp + 2 methyl	34.29	11200	1155.46375	3469.4147	3469.4273	-3.6	-3	-	c4	-	y3	manually identified	M-B(m 2,2,7G), M-B(m 2,2,7G)-H 2 O and M-m 2,2,7Gp were identified in the MS/MS spectrum. The trimethyguanine structure was also identified by the previous study [2] .
			52-58	AUUUCCG>p	32.12	78400	1100.11658	2202.2488	2202.2584	-4.4	-2	a4-a6	c2-c6	w1,w2,w4-w6	y2-y6	Ariadne	
			52-58	AUUUCCGp	30.37	30200	1109.12256	2220.2608	2220.2690	-3.7	-2	a3,a5-a7	c2-c7	w2,w4-w6	y2-y6	Ariadne	
			64-76	AAAAACAACUAUG>p	39.43	6960	1398.85278	4199.5818	4199.6008	-4.5	-3					m/z	
		AGAUCUCUGG	64-76	AAAAACAACUAUGp	38.94	4340	1404.85742	4217.5957	4217.6113	-3.7	-3					m/z	
		UUUCUCUUCA															
		UAACGAAUAA															
		AUCUUUCGCC	79-84	UUUAUG>p	30.94	47000	948.08899	1898.1936	1898.2012	-4.0	-2	a4,a5	c1-c5	w1,w2	y1-y5	manually identified	
		UUUUACUAAA															
5	U5F	GAUUUCCGUG	79-84	UUUAUGp	29.47	27100	957.09454	1916.2047	1916.2117	-3.7	-2	a3,a5,a6	c2-c5	w2	y1-y5	Ariadne	
		GAGAAAAACA															
		ACUAUGAGUU	86-97	UUAAAUUUUUUG>p	36.77	31800	1259.12280	3780.3919	3780.4074	-4.1	-3	-	c2,c5-c7,c10	w7,w8	y2,y3,y5-y7,y9	Ariadne	
		UAUGGUUAAA															
		UUUUUGAAG	86-97	UUAAAUUUUUUGp	36.09	33800	1265.12671	3798.4036	3798.4180	-3.8	-3	-	c2,c3,c5-c7,c9,c10	w8,w9	y1-y3,y5,y10	Ariadne	
		UCUUGCCUAG															
		GCAAGGCUCG	98-100	AAG>p	33.32	14100	1002.14069	1003.1485	1003.1525	-4.0	-1	-	c2	w1,w2	y1,y2	manually identified	
			98-100	AAGp	30.75	13900	1020.15186	1021.1597	1021.1630	-3.3	-1	-	c1,c2	w1	y2	manually identified	
			101-105	UCUUG>p	27.77	4000	783.07263	1568.1609	1568.1646	-2.4	-2					m/z	
			101-105	UCUUGp	26.62	1110	792.07684	1586.1693	1586.1752	-3.7	-2					m/z	
			106-110	CCUAG>p	30.52	2700	794.09308	1590.2018	1590.2078	-3.8	-2					m/z	
			112-115	CAAG>p	34.08	7690	653.08691	1308.1895	1308.1938	-3.3	-2					m/z	
			2-9	ACUCUUAG>p	35.78	17300	1264.64355	2531.3028	2531.3110	-3.2	-2					m/z	
			2-9	ACUCUUAGp	34.66	20300	1273.64771	2549.3111	2549.3215	-4.1	-2					m/z	
			16-23	AUCACUCG>p	38.12	31800	1264.15076	2530.3172	2530.3269	-3.9	-2	a4,a5	c2,c4-c7	w2-w5	y2,y3,y5-y7	Ariadne	
			16-23	AUCACUCGp	37.00	8980	1273.15662	2548.3289	2548.3375	-3.4	-2					m/z	
			36-38	AUG>p	30.63	39400	979.11359	980.1214	980.1253	-3.9	-1	-	c2	w1,w2	y1,y2	manually identified	
			36-38	AUGp	28.35	3630	997.12494	998.1328	998.1358	-3.1	-1					m/z	
		GACUCUUAGC	39-41	AAG>p	36.34	30300	1002.14124	1003.1491	1003.1525	-3.4	-1	-	c1,c2	w1,w2	y1,y2	manually identified	
		GGUGGAUCAC															
		UCGGCUCGUG	39-41	AAGp	33.77	20500	1020.15155	1021.1594	1021.1630	-3.6	-1					m/z	
		CGUCGAUGAA															
		GAACGCAGCU	42-45, 102-105	AACG>p	37.00	6970	653.08728	1308.1902	1308.1938	-2.7	-2					m/z	
		AGCUGCGAGA															
		AUUAAUGUGA	49-52	CUAG>p	30.17	4120	1284.15662	1285.1644	1285.1665	-1.6	-1					m/z	
6	5.8S	AUUGCAGGAC															
		ACAUUGAUCA	60-67	AAUUAAUG>p	41.41	20100	1288.65369	2579.3230	2579.3334	-4.0	-2	-	c2-c6	w2,w4,w6,w7	y3-y6	Ariadne	
		UCGACACUUC															
		GAACGCACUU	60-67	AAUUAAUGp	40.31	9180	1297.65906	2597.3338	2597.3440	-3.9	-2					m/z	
		GCGGCCCCGG															
		GUUCCUCCCG	70-77	AAUUGmCAG>p	41.19	15500	1303.16821	2608.3521	2608.3600	-3.0	-2	a6	c3-c7	w2,w4	y2-y7	manually identified	
		GGGCUACGCC															
		UGUCUGAGCG	70-77	AAUUGmCAGp	40.31	24100	1312.17224	2626.3601	2626.3705	-4.0	-2	a5	c3,c4,c7	-	y4,y6,y7	manually identified	The placement of methyl residue was deduced from its 2'3' cyclic form.
		UCGCUU															
			79-86	ACACAUUG>p	35.78	39200	1276.15613	2554.3279	2554.3382	-4.0	-2	-	c2,c3,c5,c6	w3-w5	y2,y3,y6	manually identified	
			79-86	ACACAUUGp	34.66	11000	1285.16101	2572.3377	2572.3487	-4.3	-2					m/z	
			87-93	AUCAUCG>p	36.87	35500	1111.63062	2225.2769	2225.2857	-3.9	-2	a3-a6	c2,c4-c6	w1,w2,w4,w6	y2-y6	Ariadne	
			87-93	AUCAUCGp	35.56	4000	1120.63562	2243.2869	2243.2962	-4.2	-2					m/z	
			94-101	ACACUUCG>p	34.48	27400	1264.15125	2530.3182	2530.3269	-3.5	-2	a4	c2,c3,c5-c7	w1,w4-w6	y2,y3,y5-y7	Ariadne	
			94-101	ACACUUCGp	33.41	5850	1273.15576	2548.3272	2548.3375	-4.1	-2					m/z	
			106-111	CACUUG>p	31.60	7440	947.10553	1896.2267	1896.2331	-3.4	-2					m/z	
			1-4	m 2,2,7GpppAmUmCG>p	32.31	24200	929.10059	1860.2168	1860.2249	-4.3	-2	-	-	-	y1	manually identified	M-B(m 2,2,7G), M-B(m 2,2,7G)-H 2 O, M-m 2,2,7Gp, M-m 2,2,7Gpp, m 2,2,7Gp and m 2,2,7Gpp were identified in the MS/MS spectrum. The trimethyguanine structure was also identified by the The trimethyguanine structure and the placement of methyl residue were deduced from its 2'3' cyclic form.
			1-4	m 2,2,7GpppAmUmCGp	31.26	3770	938.10419	1878.2240	1878.2355	-6.1	-2					m/z	This study identified the methyl residue on the ribose, because not M-B4(mC) but M-B4(C) ion was
			37-42	UAUCmUG>p	26.37	52400	954.60645	1911.2286	1911.2328	-2.2	-2	a2-a5	c1-c5	w1-w5	y2-y5	Ariadne	identified in the MS/MS spectrum. The placement of methyl residue was also identified by the
			37-42	UAUCmUGp	25.32	3550	963.61053	1929.2367	1929.2434	-3.5	-2					m/z	The placement of methyl residue was deduced from its 2'3' cyclic form.
			43-52	UUCUUmAUCAG>p or UUCUUAmUCAG>p	31.84	14300	1051.44739	3157.3656	3157.3772	-3.7	-3	a4,a9	c2-c4,c6-c9	w2,w4,w7	y2,y3,y6-y9	Ariadne	The placement of methyl residue was also identified by the previous study[3] .
			43-52	UUCUUmAUCAGp or UUCUUAmUCAGp	30.91	11900	1057.45129	3175.3773	3175.3878	-3.3	-3	a3	c2-c4,c6-c9	w2,w6	y1,y2,y4,y6-y9	manually identified	The placement of methyl residue was also identified by the previous study[3] .
			53-63	UUUAAUAUCmUG>p or UUUAAUAUCUmG>p	34.35	23700	1161.45850	3487.3990	3487.4138	-4.2	-3	a2	c2-c8,c10	w4,w6	y3,y5-y8,y10	Ariadne	The placement of methyl residue was also identified by the previous study[1] .
			53-63	UUUAAUAUCmUGp or UUUAAUAUCUmGp	34.01	10900	1167.46191	3505.4092	3505.4243	-4.3	-3	-	c3,c8,c10	-	y3,y4,y7	manually identified	The placement of methyl residue was also identified by the previous study[1] .
			64-68	AUACG>p	33.90	14000	806.09821	1614.2121	1614.2191	-4.3	-2	a4	c2-c4	w1,w2	y1-y4	Ariadne	
			64-68	AUACGp	31.48	1770	815.10400	1632.2237	1632.2296	-3.7	-2					m/z	
		AUCGCUUCUC	69-79	UCCUCUAUCCG>p	31.26	5830	1140.12537	3423.3996	3423.4076	-2.3	-3					m/z	
		GGCCUUUUGG															
		CUAAGAUCAA	99-105	AUUUUUG>p	34.22	49000	1101.10022	2204.2161	2204.2265	-4.7	-2	a4-a6	c2-c6	-	y2-y6	Ariadne	
		GUGUAGUAUC															
		UGUUCUUAUC	99-105	AUUUUUGo	32.85	31900	1110.10596	2222.2276	2222.2370	-4.3	-2	a4.a7	c2-c6	w2.w3.w6	v2-v6	Ariadne	
		AGUUUAAUAU															
		CUGAUACGUC	116-118	AUG>p	28.24	19300	979.11414	980.1220	980.1253	-3.4	-1	-	c2	w1,w2	y1,y2	manually identified	
		CUCUAUCCGA															
		GGACAAUAUA	116-118	AUGp	26.02	1530	997.12604	998.1339	998.1358	-2.0	-1					m/z	
7	U2	UUAAAUGGAU	120-124	AAUAG>p	36.83	10200	818.10370	1638.2231	1638.2303	-4.4	-2	a2.a4	c1-c4	w1.w3	v1-v4	manually identified	
		UUUUGGAGCA															
		GGGAGAUGGA	120-124	AAUAGp	35.47	7100	827.10901	1656.2337	1656.2409	-4.3	-2					m/z	
		AUAGGAGCUU															
		GCUCCGUCCA	128-131	CUUG>p	25.77	4210	630.06085	1262.1374	1262.1393	-1.6	-2					m/z	
		CUCCACGCAU															
		CGACCUGGUA	132-136	CUCCG>p	26.10	13100	782.08783	1566.1913	1566.1966	-3.4	-2	a2-a4	c2-c4	w1-w4	v1-v4	Ariadne	
		UUGCAGUACC															
		UCCAGGAACG	137-147	UCCACUCCACG>p	31.26	29600	1147.47217	3445.4400	3445.4508	-3.1	-3	a5,a7	c2-c7,c9,c10	w2,w4	y4,y5,y7,y8,y10	Ariadne	
		GUGCACCA															
			137-147	UCCACUCCACGp	30.60	2460	1153.47424	3463.4462	3463.4614	-4.4	-3					m/z	
			148-152	CAUCG>p	29.07	35400	794.09375	1590.2032	1590.2078	-2.9	-2	a3.a4	c3.c4	w1.w3.w4	v1-v4	Ariadne	
			148-152	CAUCGp	27.24	3180	803.09912	1608.2139	1608.2184	-2.8	-2					m/z	

			153-157	ACCUG>p	31.26	37200	794.09387	1590.2034	1590.2078	-2.8	-2	a2-a4	c1-c4	w1-w4	y1-y4	Ariadne	
			153-157	ACCUGp	29.82	7140	803.09900	1608.2137	1608.2184	-3.0	-2					m/z	
			159-163	UAUUG>p	30.60	57600	795.07703	1592.1697	1592.1759	-3.9	-2	a3,a4	c2-c4	w1-w3	y1-y4	Ariadne	
			159-163	UAUUGp	28.94	9620	804.08331	1610.1823	1610.1864	-2.6	-2					m/z	
			167-175	UACCUCCAG>p	32.31	24200	1416.66968	2835.3550	2835.3682	-4.7	-2	a3-a6	c4-c6,c8	w2,w3,w5,w6	y2-y6	Ariadne	
			167-175	UACCUCCAGp	31.48	25100	1425.67529	2853.3662	2853.3788	-4.4	-2	a3,a6	c3,c5,c6,c8	w2-w4,w6	y2-y6,y8	Ariadne	
			177-180	AACG>p	34.68	27700	653.08661	1308.1889	1308.1938	-3.7	-2	a3	c1,c3	w1,w2	y1-y3	manually identified	
			184-188	CACCA-OH	29.91	49300	754.62726	1511.2702	1511.2731	-2.0	-2	a2-a5	c1-c4	w1-w4	y2-y4	Ariadne	
			4-7	CUCG>p	28.74	1070	629.56708	1261.1498	1261.1553	-4.4	-2					m/z	
			8-12	CUUCG>p	28.06	1220	782.58075	1567.1772	1567.1806	-2.2	-2					m/z	
			14-16	CAG>p	24.29	1540	978.13098	979.1388	979.1412	-2.5	-1					m/z	
			35-38	AACG>p	34.68	27700	653.08661	1308.1889	1308.1938	-3.7	-2	a3	c1,c3	w1,w2	y1-y3	manually identified	
			35-38	AACGp	33.49	895	1325.19592	1326.2037	1326.2043	-0.4	-1					m/z	
			39-44	AUACm^{6}AG>p	34.01	6170	977.63123	1957.2781	1957.2872	-4.7	-2					m/z	The placement of methyl residue was deduced from its 3' phosphate form.
			39-44	AUACm^{6}AGp	32.53	12600	986.63751	1975.2907	1975.2978	-3.6	-2	a4	c2-c5	w2-w4	y2-y5	manually identified	This study identified the methyl residue on the base because M-B$_2$ (mA) ion was identified in the MS/MS spectrum. The configuration of methyl residue on the base was deduced from the previous
		GUGCUCGCUU CGGCAGCACA UAUACUAAAA UUGGAACGAU ACAGAGAAGA UUAGCAUGGC CCCUGCGCAA GGAUGACACG CAAAUUCGUG AAGCGUUCCA UAUUUUU	47-49, 91-93	AAG>p	33.90	10400	1002.13983	1003.1477	1003.1525	-4.8	-1	-	c2	w1	y2	manually identified	
			47-49, 91-93	AAGp	31.48	7630	1020.15094	1021.1588	1021.1630	-4.2	-1					m/z	
7	U6		50-58	AUUAmGmCAUG>p	39.75	21100	1463.18652	2928.3887	2928.4009	-4.2	-2	a6	c2,c3,c6,c8	w3,w4	y2-y4,y6-y8	manually identified	The MS/MS spectrum shows two methyl residues between A53 and G54. The placement of methyl residues were deduced from the previous study[1].
			50-58	AUUAmGmCAUGp	39.19	6350	1472.19226	2946.4002	2946.4115	-3.8	-2					m/z	The placement of methyl residue was deduced from its 2'3' cyclic form.
			60-65	CmCCmCmUG>p or CCmCmCmUG>p	30.73	22000	955.63013	1913.2759	1913.2848	-4.7	-2	a2,a4	c2,c4,c5	w1-w4	y1-y4	manually identified	The previous study[1] supports the structure CmCCmCmUG>p.
			60-65	CmCCmCmUGp	29.82	3490	964.63708	1931.2898	1931.2954	-2.9	-2					m/z	The placement of methyl residue was deduced from its 2'3' cyclic form.
			73-75	AUG>p	28.24	19300	979.11414	980.1220	980.1253	-3.4	-1	-	c2	w1,w2	y1,y2	manually identified	
			73-75	AUGp	26.02	1530	997.12604	998.1339	998.1358	-2.0	-1					m/z	
			76-80	ACmACG>p	39.48	31900	812.61383	1627.2433	1627.2507	-4.5	-2	a3,a4	c2-c4	w1-w4	y1,y2,y4	Ariadne	The MS/MS spectrum shows that the fragment contains methyl residue on base or ribose. The configuration of the residue was deduced from the previous study[1].
			76-80	ACmACGp	38.00	1610	821.61902	1645.2537	1645.2613	-4.6	-2					m/z	The placement of methyl residue was deduced from its 2'3' cyclic form.
			81-88	CAAAUUCG>p	34.35	25400	1276.15613	2554.3279	2554.3382	-4.0	-2	a4,a6	c2-c6	w3-w5	y3-y7	Ariadne	
			81-88	CAAAUUCGp	34.35	5570	1285.16174	2572.3391	2572.3487	-3.7	-2					m/z	
			96-107	UUCCAUAUUUUU>p	34.35	18300	1237.78357	3716.3742	3716.3900	-4.3	-3	-	c2-c4,c6	-	y2	manually identified	

Ram Reddy; Nucleic Acids Res. 1986; 14(Suppl): r61–r72. Compilation of small RNA sequences

Sontheimer EJ, Steitz JA. ; Mol Cell Biol. 1992 Feb;12(2):734-46. Three novel functional variants of human U5 small nuclear RNA.

Jády BE, Darzacq X, Tucker KE, Matera AG, Bertrand E, Kiss T. ; EMBO J. 2003 Apr 15;22(8):1878-88. Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm.

Supplementary Table 2. RNase T1 fragments of U1 snRNA identified by LC-MS analysis.

Sequence of U1 snRNA	AUACUUACCUGGCAGGGGAGAUACCAUGAUCACGAAGGUGGUUUUCCCAGGGCGAGGCUU AUCCAUUGCACUCCGGAUGUGCUGACCCCUGCGAUUUCCCCAAAUGUGGGAAACUCGACU GCAUAAUUUGUGGUAGUGGGGGACUGCGUUCGCGCUUUCCCCUG													
Residue number	**Identified sequences**	**Retention time (min)**	**Intensity**	**Observed m/z**	**Observed m/z**	**Observed molecular mass**	**Theoritical molecular mass**	**Δmass (ppm)**	**Charge**	**a-series ions identified**	**c-series ions identified**	**w-series ions identified**	**y-series ions identified**	**Idenitification method**	**Remarks**
U1-tfs															
1-11	mGpppmAmUmACUUACCUG>p	35.24	128000	1343.15	1343.1451	4032.4589	4032.4728	-3.4	-3	-	c8-c10	w2	y2-y8	manually identified	M–B(mG), M–B(mG)–H2O, M–mGp, M–mGpp and mGpp were identified in the MS/MS spectrum.
1-11	mGpppmAmUmACUUACCUGp	34.89	31700	1349.15	1349.1489	4050.4703	4050.4833	-3.2	-3	-	c10	-	y2-y5	manually identified	M–B(mG), M–B(mG)–H2O, M–mGp and M–mGpp were identified in the MS/MS spectrum.
21-28	AUACCAUG>p	35.59	210000	1276.16	1276.1559	2554.3274	2554.3382	-4.2	-2	a4,a5	c2,c3,c5-c7	w2-w5	y2-y7	Ariadne	
21-28	AUACCAUGp	34.89	60600	1285.16	1285.1620	2572.3396	2572.3487	-3.5	-2	a4,a5,a8	c2,c3,c5-c7	w2-w6	y2-y7	Ariadne	
29-34	AUCACG>p	37.81	234000	958.62	958.6190	1919.2537	1919.2604	-3.5	-2	a3,a5	c2-c5	w1-w5	y1-y5	Ariadne	
29-34	AUCACGp	36.32	26600	967.62	967.6249	1937.2654	1937.2709	-2.8	-2	a3,a5,a6	c1-c6	w1,w3,w5	y1-y5	Ariadne	
35-37	AAG>p	34.43	24800	1002.14	1002.1416	1003.1494	1003.1525	-3.0	-1	-	c1,c2	w1,w2	y1,y2	manually identified	
35-37	AAGp	32.09	14000	1020.15	1020.1531	1021.1610	1021.1630	-2.0	-1	-	a2	-	y1,y2	manually identified	
42-50	UUUUCCCAG>p	33.30	98700	1405.65	1405.6500	2813.3157	2813.3250	-3.3	-2	a5,a6	c2-c8	w2-w4,w7	y2-y8	Ariadne	
42-50	UUUUCCCAGp	32.38	51000	1414.65	1414.6545	2831.3247	2831.3356	-3.8	-2	a6,a9	c2-c6,c8	w2-w4,w7	y2-y8	Ariadne	
58-68	CUUAUCCAUUG>p	34.04	64100	1148.46	1148.4562	3448.3920	3448.4029	-3.1	-3	a6,a7	c2,c3,c5-c10	w3-w5	y2-y10	Ariadne	
69-75	CACUCCG>p	29.69	100000	1099.13	1099.1340	2200.2837	2200.2904	-3.0	-2	a3-a6	c2-c6	w1,w2,w4-w6	y2-y6	Ariadne	
69-75	CACUCCGp	27.97	19900	1108.14	1108.1392	2218.2940	2218.3010	-3.2	-2	a6,a7	c2-c6	w1,w2,w4-w6	y1-y6	Ariadne	
77-79	AUG>p	28.89	18100	979.12	979.1151	980.1229	980.1253	-2.4	-1	a2	c1,c2	w1	y1,y2	manually identified	
85-91	ACCCCUG>p	31.37	131000	1099.14	1099.1354	2200.2864	2200.2904	-1.8	-2	a2-a6	c2-c6	w2-w6	y1-y6	Ariadne	
85-91	ACCCCUGp	30.40	35900	1108.14	1108.1397	2218.2950	2218.3010	-2.7	-2	a2-a5,a7	c2-c6	w3,w5,w6	y1-y6	Ariadne	
94-106	AUUUCCCCAAAUG>p	36.32	148000	1367.50	1367.4967	4105.5136	4105.5239	-2.5	-3	a6	c2-c12	w2-w5,w7	y2-y8,y10,y12	Ariadne	
94-106	AUUUCCCCAAAUGp	35.59	33100	1373.50	1373.4973	4123.5154	4123.5344	-4.6	-3	-	c3-c6,c8,c9,c11	w3	y2,y3,y12	manually identified	
111-117	AAACUCG>p	39.61	188000	1123.15	1123.1450	2248.3057	2248.3129	-3.2	-2	a4,a6	c2-c6	w1-w6	y2-y6	Ariadne	
111-117	AAACUCGp	38.35	29800	1132.15	1132.1493	2266.3142	2266.3234	-4.1	-2	a4-a7	c2-c7	w1,w3-w6	y1-y6	Ariadne	
118-121, 143-146	ACUG>p	30.58	10000	641.57	641.5739	1285.1634	1285.1665	-2.5	-2	a3	c2,c3	w1,w2	y1,y3	manually identified	
122-130	CAUAAUUUG>p	35.59	4530	1429.66	1429.6603	2861.3362	2861.3475	-3.9	-2	-	-	-	-	m/z	
U1 snRNA															
1-11	mmmGpppAmUmACUUACCUG>p	37.18	180000	1347.82	1347.8169	4046.4741	4046.4884	-3.5	-3	a9	c3,c6-c10	w2,w3,w7	y2-y9	manually identified	M–B(mmmG), M–B(mmmG)–H2O, M–mmmGp, M–mmmGpp and mmmGpp were identified in the MS/MS spectrum.
1-11	mmmGpppAmUmACUUACCUGp	36.87	48100	1353.82	1353.8196	4064.4822	4064.4990	-4.1	-3	-	c3,c6,c7,c9,c10	w3,w7,w8	y2-y8	manually identified	M–B(mmmG), M–B(mmmG)–H2O, M–mmmGp, M–mmmGpp and mmmGpp were identified in the MS/MS spectrum.
21-28	AUACCAUG>p	37.40	312000	1276.16	1276.1574	2554.3304	2554.3382	-3.1	-2	a3-a6	c2,c3,c5-c7	w2-w5	y2-y7	Ariadne	
21-28	AUACCAUGp	36.55	99500	1285.16	1285.1604	2572.3365	2572.3487	-4.8	-2	a5,a8	c2,c3,c5-c7	w2,w4,w5	y2,y3,y5-y7	Ariadne	
29-34	AUCACG>p	39.59	469000	958.62	958.6197	1919.2550	1919.2604	-2.8	-2	a3-a5	c1-c5	w1-w5	y1-y5	Ariadne	
29-34	AUCACGp	37.94	41800	967.62	967.6240	1937.2636	1937.2709	-3.8	-2	a3,a5,a6	c2-c6	w1-w4	y2-y5	Ariadne	
35-37	AAG>p	36.34	30300	1002.14	1002.1412	1003.1491	1003.1525	-3.4	-1	-	c1,c2	w1,w2	y1,y2	manually identified	
42-50	UUUUCCCAG>p	34.66	183000	1405.65	1405.6491	2813.3138	2813.3250	-4.0	-2	a5,a6	c2-c4,c6,c8	w2-w4	y2-y8	Ariadne	
42-50	UUUUCCCAGp	33.77	184000	1414.66	1414.6553	2831.3262	2831.3356	-3.3	-2	a5,a6,a9	c2-c8	w2-w4	y2-y8	Ariadne	
58-68	CUUAUCCAUUG>p	35.56	179000	1148.46	1148.4553	3448.3894	3448.4029	-3.9	-3	a6,a7	c2,c3,c5-c10	w4-w6	y2-y10	Ariadne	
58-68	CUUAUCCAUUGp	34.66	36500	1154.46	1154.4584	3466.3986	3466.4134	-4.3	-3	a11	c2,c3,c5-c8	w3	y2-y7,y10	Ariadne	
69-75	CACUCCG>p	31.33	32300	1099.13	1099.1346	2200.2849	2200.2904	-2.5	-2	a3-a6	c2-c6	w2-w6	y2-y6	Ariadne	
69-75	CAmCUCCG>p	35.11	240000	1106.14	1106.1416	2214.2989	2214.3061	-3.3	-2	a2-a6	c2-c6	w1,w2,w4,w6	y2,y3	Ariadne	
69-75	CAmCUCCGp	33.77	67600	1115.15	1115.1467	2232.3091	2232.3166	-3.4	-2	a3-a7	c3-c6	w2,w4	y1-y4,y6	Ariadne	
77-79	AUG>p	30.63	39400	979.11	979.1136	980.1214	980.1253	-3.9	-1	-	c2	w1,w2	y1,y2	manually identified	
85-91	ACCCCUG>p	33.09	295000	1099.13	1099.1338	2200.2832	2200.2904	-3.3	-2	a2-a6	c2-c6	w2-w6	y2-y6	Ariadne	
85-91	ACCCCUGp	32.22	52500	1108.14	1108.1381	2218.2918	2218.3010	-4.1	-2	a3-a5,a7	c2-c6	w2-w5	y1-y6	Ariadne	
94-106	AUUUCCCCAAAUG>p	38.12	216000	1367.49	1367.4944	4105.5066	4105.5239	-4.2	-3	a5,a6,a8	c2-c10,c12	w2-w6,w12	y2-y12	Ariadne	
111-117	AAACUCG>p	41.19	204000	1123.14	1123.1447	2248.3050	2248.3129	-3.5	-2	a4-a6	c2-c6	w1,w3-w6	y2-y6	Ariadne	

111-117	AAACUCGp	40.31	34300	1132.15	1132.1489	2266.3135	2266.3234	-4.4	-2	a4,a6,a7	c2-c6	w1,w3-w5	y2-y6	Ariadne
118-121, 143-146	ACUG>p	32.22	79600	1284.15	1284.1537	1285.1615	1285.1665	-3.9	-1	a2,a3	c3	w1-w3	y2,y3	Ariadne
122-130	CAUAAUUUG>p	37.40	131000	1429.66	1429.6608	2861.3372	2861.3475	-3.6	-2	a5	c2-c5,c8	w4,w5,w7,w8	y2-y8	Ariadne
122-130	CAUAAUUUGp	36.34	54200	1438.67	1438.6659	2879.3474	2879.3581	-3.7	-2	a5,a9	c3-c7	w3-w5,w8	y2-y6,y8	Ariadne
149-152	UUCG	27.93	27600	1261.13	1261.1288	1262.1366	1262.1393	-2.2	-1	a3	c2,c3	w2,w3	y2,y3	Ariadne
155-164	CUUUCCCCUG-OH	33.41	92900	1009.79	1009.7889	3032.3901	3032.3993	-3.0	-3	a5,a6,a8	c2,c6,c7	w2,w4,w9	y2-y9	Ariadne
155-164	CUUUCCCCUG>p	31.95	14600	1030.44	1030.4412	3094.3470	3094.3551	-2.6	-3	a5	c2,c4,c5-c9	w2-w4	y3-y5,y7,y8	manually identified

Supplementary Table 3. U1-tfs sequence determined by 3'RACE analysis.

Clone #	*U1-Tfs sequence*	*Length*	*Extended Sequence*
1	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGA	118	TTT
2	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTG	121	–
3	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGC	122	–
4	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGC	122	AAAT
5	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCA	123	–
6	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCA	123	A
7	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCA	123	–
8	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCA	123	A
9	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCA	123	–
10	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCA	123	–
11	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCA	123	–
12	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	–
13	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	–
14	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	TT
15	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	T
16	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	–
17	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	TT
18	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	–
19	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCAT	124	–
20	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAAT	127	–

Supplementary Table 4. RNase T1 fragments of RAT-U1 snRNA and RAT-U1-tfs identified by LC-MS analysis.

			RAT-WT					RAT-U1-tfs (ΔSL4-1)					RAT-U1-tfs (ΔSL4-1Δ3' box)				
	MW(theo) 3P	m/z(theo) 3	m/z	charge	MW(obs)	error(ppm)	peak area	m/z	charge	MW(obs)	error(ppm)	peak area	m/z	charge	MW(obs)	error(ppm)	peak area
MMG-mAmUmACUUACCUG (extra 1-11)	4050.48334	1349.153			0.00		0	1349.15	3	4050.48	-0.07	386,431	1349.15	3	4050.48	-0.61	429,854
MMG-AmUmACUUACCUG (extra 1-11)	4036.46769	1344.481	1344.48	3	4036.47	0.75	66,051	1344.48	3	4036.47	0.30	389,639	1344.48	3	4036.47	0.03	472,162
TMG-AmUmACUUACCUG (extra 1-11)	4064.49899	1353.825	1353.83	3	4064.50	0.38	25,987,977			0.00		0			0.00		0
UAAG (RAT-tag)	1327.18835	662.586	662.59	2	1327.19	0.12	9,052,113	662.59	2	1327.19	0.40	217,922	662.59	2	1327.19	0.12	283,997
UUUAUAUG (RAT-tag)	2551.28956	1274.637	1274.64	2	2551.29	0.29	40,973,436	1274.64	2	2551.29	-0.28	957,070	1274.64	2	2551.29	-0.28	1,259,063
AAACCCUUAG (RAT-tag)	3206.44255	1602.213	1602.21	2	3206.45	0.95	18,819,149	1602.21	2	3206.44	-0.28	379,600	1602.21	2	3206.44	-0.43	563,384
CACG (RAT-tag)	1302.19310	650.089	650.09	2	1302.19	-0.07	6,968,917	650.09	2	1302.19	0.13	228,984	650.09	2	1302.19	0.32	326,470
UUUAG (RAT-tag)	1610.18643	804.085	804.09	2	1610.19	-0.22	34,292,280	804.09	2	1610.19	0.22	905,204	804.09	2	1610.19	0.00	1,108,862
CUACACUCG (RAT-tag)	2853.37880	1425.682	1425.68	2	2853.38	0.39	32,668,520	1425.68	2	2853.38	-0.64	789,653	1425.68	2	2853.38	-0.04	1,016,937
CCAUCG (RAT-tag)	1913.25969	955.622	955.62	2	1913.26	-0.13	34,994,143	955.62	2	1913.26	0.12	922,409	955.62	2	1913.26	0.18	1,161,596
AUAUACUUACCUG (RAT-tag + 1-11)	4124.51846	1373.832	1373.83	3	4124.52	0.45	32,360,374	1373.83	3	4124.52	0.27	736,470	1373.83	3	4124.52	0.09	1,061,248
AUACCAUG (21-28)	2572.34874	1285.167	1285.17	2	2572.35	-0.22	35,565,762	1285.17	2	2572.35	0.06	856,819	1285.17	2	2572.35	0.06	1,039,071
AUCACG (29-34)	1937.27092	967.628	967.63	2	1937.27	-0.33	46,959,087	967.63	2	1937.27	0.30	908,604	967.63	2	1937.27	0.37	1,070,924
UUUUCCCAG (42-50)	2831.33560	1414.660	1414.66	2	2831.34	-0.04	34,245,570	1414.66	2	2831.34	0.22	912,751	1414.66	2	2831.33	-0.21	1,045,754
CUUAUCCAUUG (58-68)	3466.41342	1154.463	1154.46	3	3466.41	0.38	31,859,915	1154.46	3	3466.41	0.16	855,120	1154.46	3	3466.41	0.06	988,452
CACUCCG (69-75)	2218.30097	1108.143	1108.14	2	2218.30	0.25	27,471,833	1108.14	2	2218.30	0.03	957,783	1108.14	2	2218.30	0.03	1,204,205
CAmCUCCG (69-75)	2232.31662	1115.150	1115.15	2	2232.32	-0.30	18,196,960			0.00		0			0.00		0
ACCCCUG (85-91)	2218.30097	1108.143	1108.14	2	2218.30	-0.07	41,051,525	1108.14	2	2218.30	-0.41	985,162	1108.14	2	2218.30	0.14	1,233,827
AUUUCCCCAAAUG (94-106)	4123.53445	1373.504	1373.50	3	4123.54	0.45	34,902,502	1373.50	3	4123.53	-0.35	818,043	1373.50	3	4123.54	0.36	1,115,898
AAACUCG (111-117)	2266.32344	1132.154	1132.15	2	2266.32	0.36	53,256,812	1132.15	2	2266.32	0.14	1,102,700	1132.15	2	2266.32	0.67	1,599,974
ACUG (118-121, 143-146)	1303.17711	650.581	650.58	2	1303.18	-0.72	11,904,708	650.58	2	1303.18	-1.38	241,130	650.58	2	1303.18	-1.09	321,241
CAUAAUUUG (122-130)	2879.35806	1438.671	1438.67	2	2879.36	0.38	36,201,312			0.00		0			0.00		0
UUCG (149-152)	1280.14989	639.067	639.07	2	1280.15	-0.07	6,336,554			0.00		0			0.00		0
CUUUCCCCUGp (155-164)	3112.36565	1036.447	1036.45	3	3112.37	0.00	12,588,459			0.00		0			0.00		0
CUUUCCCCUG-OH (155-164)	3032.39932	1009.792	1009.79	3	3032.40	0.53	10,123,336			0.00		0			0.00		0

Supplementary Table 5. U1 genes and their homologs present in the human genome*.

			Distal sequence element (DSE)			*Proximal sequence element (PSE)*			*U1 coding sequence*						*3' box*		
				Position			Position			Position		Blast search result				Position	
Chromosome	Direction	Accession numbe	Sequence	Start	End	Sequence	Start	End	Sequence	Start	End	Identities	Expect	Sequence	Start	End
1	reverse	NT_113797.1				AAGTGACCGTGCGTGTAA	2716	2699	ATACTTACCTGGCAGGGGAGATACGATGATCACGAAGGTGGTTTTCCCAGGGAGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCCGTTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTGTCCTCTG	2652	2489	159/165 (97%)	1E-70			
1	reverse	NT_113797.1	GGCGACTTCTATGTAGATGA	28982	28963				ATATTTACTTGGCAGGGGAGATAACATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCAGATGTGTTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGATTGCATAATTTGTGGTAGTGGTACTGCGTTCGCGCTTTCCCCTG	28765	28604	155/164 (95%)	5E-65			
1	forward	NT_113799.1	GGCGATTTCTATGTAGATGA	65184	65203	AAGTGACCGTGCGTGTAAA	65363	65381	ATACTTACCTGGAAGGGGAGATACCATGATCACGAAGGTGGTTTTTCTCAGGGCGAAGCTTATCCATTGCGTTCCGGATGTGCTGACCCCTGCGATTTCCCCAAGTGTGGGAAACTCGAGTGCATAATTTGTGATAGTAGGGAACTACGTTCGTCGCTTTC	65427	65586	148/160 (93%)	8E-58			
1	reverse	NT_004487.19	GGCGATTTCTATGTAGATGA	93950	93931	AAGTGACCGTGCGTGTAAA	93778	93760	ATACTTACTTGGCGGGGGAGATACCATGATCACGAAGGTGGTTTTCTCAGGGCGAGGCTTATCCGTTATGTTCCGGGTGTACTGACCCCTGCCATTTTCCCCCAATGTGAGGAACTCGACTGCATAACTTGTGATAGTAGGGGACTGCGTTCGCGCTTTCCCCTG	93714	93550	148/166 (90%)	1E-50			
1	reverse	NT_004487.19	GGCGATTTCTATGTAGATGA	255978	255959	AAGTGACCGTGCGTGTAAA	255808	255790	ATACTTACTTGGCGGGGGAGATACCATGATCACGAAGGTGGTTTTCTCAGGGCGAGGCTTATCCGTTATGTTCCGGGTGTACTGACCCCTGCCATTTTCCCCCAATGTGAGGAACTCGACTGCATAACTTGTGATAGTAGGGGACTGTGTTCGCGTTTTCCCCTG	255744	255580	146/166 (88%)	3E-47			
1	reverse	NT_004487.19				AAGTGACCGTGTCTGTAAA	682975	682957	ACTTACCTGGCAGGGGAGATAGTATGATCATGAAAGTGGTTTTTCCAGAGCGAGGCTTATCCATTGCACTCCGGATGTGTTGACCTCTGCGATTTCCCCAACTGTGGGAAACTCGACTGCGTAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	682909	682748	152/162 (94%)	8E-63			
1	reverse	NT_004487.19	GGCGACTTCTATGTAGATGA	713108	713089	AAGTGACCGTGTGTGTAAA	712927	712909	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	712863	712700	164/164 (100%)	1E-80			
1	reverse	NT_004487.19	GGCGATTTCTATGTAGATGA	1003122	1003103	AAGTGACCGTGCGTGTAAA	1002962	1002944	ATACTTACCTGGCAGGGGAGATACTATTATCAAACGAAGGTGGTTTTTCTCAGGGCGAGGCTTATCCATTGTGTTCCGGATGTGCTGACCTCTGCGATTTCCCCAAACGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCTCTG	1002898	1002732	155/167 (93%)	4E-61			
1	reverse	NT_167185.1	GGCGACTTCTATGTAGATGA	1297248	1297229				ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTCCGCGCTTTCCCCTG	1297020	1296857	163/164 (99%)	6E-79			
1	reverse	NT_167185.1	GGCGACTTCTATGTAGATGA	2838982	2838963				ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTCCGCGCTTTCCCCTG	2838754	2838591	163/164 (99%)	6E-79			
1	reverse	NT_167185.1				AAGTGACCGTGCGTGTAA	3162316	3162299	ATACTTACCTGGCAGGGGAGATACGATGATCACGAAGGTGGTTTTCCCAGGGAGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCCGTTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTGTCCTCTG	3162252	3162089	159/165 (97%)	1E-70			
1	reverse	NT_167185.1	GGCGACTTCTATGTAGATGA	3188607	3188588				ATATTTACTTGGCAGGGGAGATAACATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCAGATGTGTTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGATTGCATAATTTGTGGTAGTGGTACTGCGTTCGCGCTTTCCCCTG	3188390	3188229	155/164 (95%)	5E-65			
1	forward	NT_167185.1	GGCGATTTCTATGTAGATGA	3321510	3321529	AAGTGACCGTGCGTGTAAA	3321669	3321687	ATACTTACCTGGCAGGGGAGATACTATTATCTAAACGAAGGTGGTTTTTCTCAGGGCGAGGGTTATCCATTGTGTTCCGGATGTGCTGACCTCTGCGATTTCCCCAAACGTGGGAAACTCGATTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCTCTG	3321733	3321900	153/168 (92%)	1E-56			
1	reverse	NT_004610.19	GGCGACTTCTATGTAGATGA	3521098	3521079	AAGTGACCGTGTGTGTAAA	3520932	3520914	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	3520868	3520705	164/164 (100%)	1E-80	GTTTCAAAAACAGA	3520694	3520681
1	reverse	NT_004610.19	GGCGACTTCTATGTAGATGA	3673760	3673741	AAGTGACCGTGTGTGTAAA	3673595	3673577	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	3673531	3673368	164/164 (100%)	1E-80	GTTTCAAAAACAGA	3673357	3673344
1	forward	NT_004610.19	GGCGACTTCTATGTAGATGA	3746870	3746889	AAGTGACCGTGTGTGTAAA	3747035	3747053	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	3747099	3747262	164/164 (100%)	1E-80	GTTTCAAAAACAGA	3747273	3747286
1	forward	NT_004610.19	GGCGACTTCTATGTAGATGA	3902334	3902353	AAGTGACCGTGTGTGTAAA	3902499	3902517	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	3902563	3902726	164/164 (100%)	1E-80			
1	reverse	NT_167186.1							ATACTTCCCTGGCAGGGAAGGTGCCATGATCATGAAGATGGTTTTCCCAGAGTGAGGCTCACCCACTGCACTCCAGGTAGGCTGACCCCTGCCATTTCTCCAAATGCAGGAAACTCAACTGCATAATTTGTGGTAGTGGGGGACTTTGTCCACGCTTCTCC	38815171	38815011	138/163 (85%)	1E-36			
4	reverse	NT_022778.16							ATACTTAGCTGGCAGAGGAAATACCATGATCACAAAGGTGGTTTTCCCAGGGTGAGGTTTATCCATTAAACTCAGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCAACTGCATAATTGGTGATAGTGAGAGACTGCGTTCTCACTTTCCCCT	8506139	8505977	147/163 (91%)	2E-53			
4	forward	NT_016354.19							ATACTTACCTGGCAGGGGAGATACCATGATCACATAGATGGTTTTCCCATGGCAAGTCTTATCCATTGCACTCTGGATGTGCTGACACCTGTGATTTCACCAAATGTGGGAAACTCAACTGCATAATTTGTGGTAGTGGGGGATTTTGTTTGTGCTTTCCCCTG	38889200	38889363	149/165 (91%)	6E-54			
4	reverse	NT_016354.19							ATACTTACCTGGCAGGGGAGATACCATGATCACAAAGGTGGTTTTCCCAGGACGAGGCTTATCCATTGCACTCCAGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCAACTGCATAATTTGTGGTAGTGGGGGACTGCATTCGCGCTTTCCCCTG	61464968	61464805	159/164 (97%)	3E-72			
5	forward	NT_023133.13							ATACTTACCTGGCAGGGAAGATTACCATGATCACGAAGGTGGTTTTCCCAGGGCAAGGCTTATCCATTGCACTCCAGATGTGCTGACCCCTTCAATTTCCCAAAATGTGGAAAACTCAACTGCATAATTTATGGTAGTGGGGGACTACATTCGCACTTTCTCCTG	23122989	23123153	151/165 (92%)	3E-57			
6	forward	NT_007592.15							ATACTTACCTGGCAGTGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATCGCACTCCGGATATGCTGACCCCTGCGATTTTCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGTGCTTGCGCTTTCCCCTG	13154288	13154451	158/165 (96%)	6E-69			
6	forward	NT_025741.15							ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCGCAGGCCAAGGCTTATCCATTGCACTCTGGATGTGCTGACCCCTGCAATTTCCCCAAATGTGGGAAACTTAACTGCATAATTTGTTGTAGTGGGGGACTGCATTC	26702251	26702401	142/151 (95%)	2E-58			
6	reverse	NT_007592.15							ATACTTACCTGGCAGGGGAGATACTATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCTGGATGTGCTGACCCCTGTGATTTCCCCAAATG	36547449	36547344	103/106 (98%)	2E-43			
6	forward	NT_007592.15							ATAATTACCTGGCAGGGGAGATACCATGAACACAAAGATGGTTTTCCCAGGGTGAGTCTCATCCATTGCCCTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAAGTCGACTGCATAATTTGTGTTAATGGGGGACTGTGTCCACGCTTTCC	53024059	53024218	146/160 (92%)	5E-55			
7	forward	NT_033968.6							ATACTTACCTGGCAGGGGAGATAACCATGATCACGAAGGTGGTTTTCCCAGGGCAAGGCTTATCCATTGCACTCTGGACATGCTGACCCCTGCGATGTCCCCAAATATGGGAGACTTGACTGCATAATTTGTGGTAGTGAGGCACTGCATTCGC	3023120	3023273	142/154 (93%)	2E-54			
8	reverse	NT_167187.1							ATACTTACTGGCAGGGGAGATACCATGATCATGAAGGTGATTTTCCCAGCACGAGGCTCATCCATTGCACTCCGGATGTGCTGACCCCTGCAATTTCCCCAAATGTGGGAAATTCAACTGCATAATTTGTGGTAGTGGGGACTGTGTTAGCACTTTCCCCTG	11792817	11792656	151/164 (93%)	2E-58			
8	forward	NT_008046.16							ATACTTACCTGGCAGGGGAGATACCATGATCACAAAGGTGGTTTCCCCAGGGTGAGGCTTATCCATTGCACTCCCGATGTGCTGACCCCTGTGATTCCCCCAAATGTGGGAAACTTGACTGCATAATTTGTGGTAATGGGGGACTGCATTTACACTTTTCCCTG	42284926	42285089	150/164 (92%)	3E-57			
10	forward	NT_030059.13							TTACCTGGCAGGGGAGAGACCACGGTCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATTCATTGCACTCCAGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCAACTGCATAATTGTGGTGGTGGGGGACTGCGTTCATGCTTTTCCCTG	39453796	39453954	149/160 (94%)	2E-59			
12	reverse	NT_029419.12							CTTACCTGGCAGGGGAGATAACATGATCACGAAGGTGGTTTTCCCAGGGCGAGGTTTATCCATTGCACTCTGGGTGTGCTGACCCCTGTGATTTCCCCAAAAGTGCGAAACTCGACTGCATGATTTGTGGTAGTGGGGGACTGCATTCATGCTTTCCCCT	25387984	25387825	149/160 (94%)	5E-60			
12	forward	NT_029419.12							CAGGGCGAAGCTTGTTTATTTCACTCCAGATGTGCTGACTTCTGTGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTAGTGGGGGACTGCGTTCTCACTTTC	51118768	51118874	96/112 (86%)	2E-23			
13	reverse	NT_027140.6							CCTGGCAGGAGAGATACCATCATCATGAAGGTGATTTTCCCAGAGCTGGGCTTATCCATTGCATTCTGGATGTGCTGACGCCTGTGGTTTTCCCAAATGTGGGAAACTGGACTGCATAATTTGTGGTAGTGGGGGACTATGTTCG	1629392	1629248	130/146 (90%)	2E-43			
14	forward	NT_026437.12				AAGTGACCGTGTGTTGAGA	16015856	16015874	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	16015920	16016083	164/164 (100%)	1E-80			
14	reverse	NT_026437.12				AAGTGACCGTGTGTTGAGA	16025659	16025641	ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCTG	16025595	16025432	164/164 (100%)	1E-80			
15	reverse	NT_010194.17							TACTTACTTGGTAGGGGAGATACCAAGATCACAAAGTTGGTTTCCCAGGGCGAGGCTTATTCAGTACACTCCAGGTGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAGCTCTACTGCATACTTTGTGGTAGAGAGAACTGTGTTCGTGCTCTCCCCT	53257282	53257123	142/163 (88%)	6E-44			
17	reverse	NT_010783.15							ATACTTACCTGGCAGGGGAGATACCATGATCCTGAAGGTGGTTTTCCTAGGGCGAGGCTCATCCATTGCACTCCGGATGTGCCGACTCCGCGATTTCCCCAAATGTGGGAAACTCGACTGCGTAATTTGTGGTAGCGGTGGACTGCGTTTGCGCT	12301034	12300881	145/157 (93%)	1E-55			
18	reverse	NT_010966.14							ACCTGGCATAGGAGATACCACAATCACAAAGGTGGTTTCCCAGGGTGAGGCTTATCCATTGCACTCTAGATGTGCTGATACCTGCGATTATCCCAAATGCGGGAAACTCAGCTGCATAATTTGTGGTAGTGGGGGACTGTGTGTGTGCTTTCCCCTG	30299362	30299205	142/161 (89%)	1E-45			
20	forward	NT_011387.8							CTTACCTGGCAGGGGAGATACCATGATCATGAAGGTGGTTTTCCCAGGGTGAGGTTCCACTGCACTCCAGAGGTGCTGATCCTGTGATTTCCCCAAATGTGGGAAACGCAACTGCATAATTTGTGGCAGTGGGAACTATGTTCTCACTATCCCCTG	5810704	5810859	140/161 (87%)	3E-42			
X	forward	NT_011786.16							CTTAACTGGCAGGGGAGATACCATGATCACAAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCCCTCTGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCACAATTTGTGGTAGTGGGGGACTGTGTTCGCGCTTTCCC	2825418	2825575	152/158 (97%)	3E-67			
X	reverse	NT_011669.17							TACCTGGCAGGGGAGATACCATGATCACAAAGGTGGTTTTCTCAGGGTGAGGCCCATCCATTGCACTCTGGGTGTGTTGGCCCCTGTGATTTCCCCAAATACAGGAAACTCGACTGCATAATGTGTGGTAGTGGGAGACTGAGTTTGTGCTTTCCCCT	8558271	8558114	140/158 (89%)	3E-47			

***retreaved from reference assembly (version GRCh37) obtained with file transfer protocol (ftp) from " ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/ ".**