Identification of Long noncoding RNAs as substrate of cytoplasmic Capping Enzyme

Avik Mukherjee, Safirul Islam, Chandrama Mukherjee*

Institute of Health Sciences, Presidency University, Kolkata, India

Abstract

The N^7methyl guanosine cap structure (where 'N' is the first transcribed nucleotide) was identified as a common feature of every RNA polymerase II transcript including mRNA or long noncoding RNA (lncRNA). This cap structure has been added to the nascent transcripts by sequential actions of mRNA Capping Enzyme (CE) and RNA methyl transferase (RNMT) in the nucleus. Every step of RNA metabolism including translation to decay is controlled by N7methyl guanosine cap and loss of this structure from the 5' end of mRNA leads to degradation. This concept of RNA decay following decapping has been changed after the identification of transcripts with stable uncapped ends through genome wide studies and discovery of cytoplasmic pool of CE (cCE) in different forms of life. cCE, in association with other proteins, can recap specific mRNA transcripts and thus cytoplasmic capping may act as post transcriptional regulator of the target mRNAs. Apart from mRNAs, many long noncoding RNAs (lncRNAs) undergo various post transcriptional processing in cytoplasm. Here, we aim to examine if lncRNAs are regulated by cCE.

We have selected fifteen lncRNAs associated with sarcoma for our analysis in Osteosarcoma (U2OS) cell line that was previously used for characterizing cytoplasmic capping. These selected lncRNAs were examined for their nuclear-cytoplasmic distribution. 5'-RACE analysis was conducted to detect the uncapped forms of the lncRNAs in U2OS cells which showed several uncapped forms of these transcripts. The uncapped ends of few transcripts were determined by sequencing the cloned amplicons obtained from 5'RACE assay. To determine if these lncRNAs are targeted by cCE, transiently transfected U2OS cells were used where cytoplasmic capping is inhibited either by overexpressing dominant negative mutants of CE (K294A) or Nck1 (Nck1-M3). 5'-RACE analysis from these cells showed an accumulation of uncapped forms in cytoplasmic capping inhibited cells compared to control cells indicating regulation of these transcripts by cCE. Future CAGE and sequence analysis will identify molecular features for recapping in these transcripts.

Keywords. LncRNA, Cytoplasmic capping, 5'-RACE, uncapped ends

*Corresponding author, email: chandrama.dbs@presiuniv.ac.in

Introduction

N^7Methyl guanosine cap is added to RNA polymerase II transcribed RNA molecules by mRNA Capping Enzyme (CE) bound to the phosphorylated C-terminal domain of Pol II [1]. CE with its N-terminal triphosphatase domain removes the terminal phosphate group from nascent transcripts making diphosphate RNA and with its C-terminal guanyl transferase domain adds GMP to the diphosphate RNAs. RNA guanine-7 methyltransferase (RNMT) in complex with the activator protein adds a methyl group to N7 position to complete the synthesis of the basic cap structure [2]. The cap structure is recognized by proteins required to regulate various stages of RNA metabolism including export of mRNAs to the cytoplasm [3], translation [4], stability and decay [5]. This cap structure can be further modified by addition of 2'-O- methyl on the first (cap 1) and second (cap 2), respectively [6-7]. Loss of cap is thought to initiate RNA degradation which is challenged after the identification of cytoplasmic pool of CE (cCE) in mammalian cells [8]. In parallel, several studies have reported the presence of stable uncapped transcripts in different organisms including yeast, plants and mammals [9-10]. Presence of cCE is not only found in mammals but also in Drosophila or Trypanosome [11-12] indicating cytoplasmic recapping as a regulator of post transcriptional control. In contrast to nuclear CE, cCE functions in a complex of several proteins among which Nck1 acts as the scaffold protein to form the cytoplasmic capping complex [13]. Another important protein of this complex bound to Nck1, is an unidentified 5' RNA kinase that converts the monophosphorylated RNA to diphosphorylated RNA to act as a substrate for cCE [8], [13]. Finally, similar to nuclear RNMT, cytoplasmic pool of RNMT functioning as a heterodimer with RNMT-associated miniprotein (RAMAC) catalyzes the methylation at N7 position of capped RNAs[2], [14] . Thus, concerted actions of these proteins in cytoplasmic capping complex recap the uncapped RNAs. Previous study shows cCE acts in maintaining decapping and recapping of target mRNAs which is termed as 'cap homeostasis' [10]. These target mRNAs have been found to contain similar length of poly A tails in their capped and uncapped forms raising the possibility of participation of recapped transcripts in translating pool [15]. Several studies using live cell imaging have shown evidences of periodic events of translation cycle that may likely due to cyclical decapping and recapping of mRNA transcripts [16–18]. However, evidences failed to detect any impact of cytoplasmic recapping on proteome complexity later [19]. Cap homeostasis is believed to regulate the stability of the target mRNAs and thus act as a hub for post transcriptional control. Not only in mammalian cells, cap homeostasis dictates decapping and recapping of certain transcripts in Trypanosoma [12]. Till date, regulation of mRNA transcripts has been studied in context to cytoplasmic capping. If cytoplasmic capping has any role in regulating the other transcripts synthesized by RNA pol II has not been explored.

In this study, we aim to examine if long non coding RNAs (lncRNA) are regulated by cytoplasmic capping. LncRNAs, synthesized in the nucleus, are processed in multiple ways in cytoplasm [20] .Genome wide studies using high-throughput sequencing has identified different types of lncRNAs which may vary in size from 200 bp to 100 kilo bases. Initially lncRNAs were considered as transcriptional noises. However, later studies demonstrated that lncRNAs have various functions not only at the transcriptional level but also post-transcriptionally, via regulating alternative splicing, nuclear trafficking, regulation of chromosomal architecture and chromatin remodeling, mRNA stability and translation. Similar to mRNAs, lncRNAs contain both 5'-cap and poly-A tail, which are added during transcription. After synthesis, 98% of lncRNAs were found to undergo splicing. Generally, lncRNAs contain longer introns and fewer exons compared to mRNAs. As a result, splicing produces shorter transcripts of lncRNAs, as compared to mRNA. Often lncRNAs attain complex secondary structures

which are beneficial for their specificity towards the target molecule. Many lncRNAs are expressed in specific tissues. Unlike mRNAs, they are expressed at a very low level. These noncoding RNAs have been identified to act as decoy molecule by inhibiting transcription factors from transcribing molecules or act like a scaffold molecule where they can provide surfaces to form several ribonucleoprotein complexes. Often lncRNAs function as sponge by scavenging microRNAs from mRNAsLncRNAs can also act as guide molecules by recruiting specific chromatin modifying complexes on specific locations of genes. Thus, depending on the biological cues, lncRNA can function as a positive or negative regulator of gene expression. Recent studies have also demonstrated lncRNAs which can act as architectural RNA (arcRNA), initiates formation of various membraneless cellular bodies like stress bodies, nuclear speckles, paraspeckles, splicing speckles, stress granules, etc[20–22]. In addition, lncRNAs may participate in various stress response pathways by promoting liquid–liquid phase separation (LLPS) that results in alteration of expression of RNA binding proteins. While turnover of lncRNAs is believed to follow the similar pathway as in mRNAs which follow decapping, de-adenylation and degradation [24], the regulators of lncRNA turnover were not studied methodically. In this study, we have shown cytoplasmic capping may regulate the stability of lncRNAs by targeting these noncoding transcripts. Also, our data shows accumulation of several stable uncapped forms of lncRNAs in cytoplasmic capping inhibited cells suggesting these RNAs may be targeted by cCE.

Materials and Methods:

Cell culture, plasmid constructs and transfection:

The human osteosarcoma cell line (U2OS) was procured from ATCC. U2OS cells were grown in DMEM (Dulbecco's Modified Eagle Medium) (Himedia) supplemented with 10% Fetal Bovine Serum (FBS)(Himedia) along with1% penicillin-streptomycin (Invitrogen) in a humidified incubator in presence of 5%CO_2 at 37°C. pcDNA3-bio-myc-mCE was obtained from Addgene (#82475). The pcDNA4TO-bio-myc-mCE construct was prepared by sub-cloning the bio-myc-mCE sequence into an empty pcDNA4TO vector using KpnI and ApaI restriction sites. pcDNA4TO-bio-myc-K294A was prepared from pcDNA4TO-bio-myc-mCE by using In-Fusion Site Directed Mutagenesis kit (Takara Bio). HA-NCK1-M3 construct was a generous gift from Dr. Louise Larose [25] which was prepared as described in Chen et al, 2000[26]. Cells were either transfected with empty pcDNA4TO, pcDNA4TO-bio-myc-K294A, pRK5 or pRK5HA-NCK1-M3 construct using Jetprime (Polyplus) transfection reagent according to the manufacturer's guidelines. Cells were harvested after 48 hours post- transfection.

Cell fractionation and Western blotting:

Harvested U2OS cell pellet was suspended in cytoplasmic lysis buffer (20mM Tris-Cl of pH 7.5, 10mM NaCl, 10mM $MgCl_2$, 10mM KCl, 0.2% NP40, 1mM PMSF and 1X protease inhibitor cocktail) and incubated on ice for 10 min with intermittent gentle agitation. Nuclei were removed by centrifugation at 1,000 xg for 10 minutes at 4°C. Nuclei pellet was resuspended in nuclear lysis buffer (150mM NaCl, 1% Triton X-100, 0.5% Sodium Deoxycholate, 1% SDS, and 25mM Tris-Cl of pH 7.5) for 40 minutes with occasional agitation. The nuclear lysate was prepared by centrifugation at 10,000 xg for 10 minutes at 4°C. Western blotting was performed by loading nuclear and cytoplasmic protein into 10% Mini-PROTEAN TGX gels (Biorad) and transferring to Immobilon-FL PVDF (EMD Millipore) membranes. The membranes were blocked in 3% Bovine Serum Albumin (BSA) in Phosphate-buffered saline (PBS) for 30 minutes at room temperature. The blots were probed with primary antibodies for overnight at 4°C. Mouse anti p84 (Abcam), mouse anti-GAPDH (Abcam), mouse anti-Myc (Santa Cruz Biotechnology), and rabbit anti-HA (Cell Signaling Technology) antibodies were used at a dilution of 1:1000 in the blocking solution. The membranes were then probed with respective secondary antibodies. Donkey anti-Mouse 800 (Invitrogen), and donkey anti-Rabbit 680 (Invitrogen) were used at 1:10000 dilutions

in PBST (Phosphate-buffered saline PBS +0.05% Tween-20). The membranes were visualized with an Odyssey CLx Imaging System (LiCor Inc., Lincoln, NE, USA).

Nuclear and cytoplasmic RNA extraction and poly(A) selection:

The nuclear and cytoplasmic lysate was prepared as described above. RNAase OUT(Invitrogen) was added to each lysate to prevent RNA degradation by ribonucleases. Nuclear and cytoplasmic RNA were extracted from the respective lysates using Trizol reagent (Invitrogen) following manufacturer protocol. Then RNAs were treated with DNaseI (Invitrogen), as directed by the manufacturer. Following DNaseI treatment, RNAs were recovered using Phenol: Chloroform: isoamyl alcohol (SRL) extraction. Cytoplasmic RNAs were poly(A) selected using Dynabeads mRNA DIRECT Kit (Invitrogen) as per manufacturer's instructions.

cDNA synthesis and PCR:

2µg of nuclear/cytoplasmic and 36ng of cytoplasmic poly(A) selected RNAs were reversed transcribed using Super Script III First Strand cDNA synthesis kit (Invitrogen). To amplify fragment of target genes, primers were designed against the target genes, which are located 50-100 bp downstream their first predicted CAGE tags of the transcripts obtained from the UCSC Genome Browser. PCR was performed for selected transcripts using Dream Taq DNA polymerase (Thermo Scientific). The primers used in this study is listed in Table-1.The PCR products were analysed by 1.2% Agarose gel electrophoresis. Ethidium bromide-stained gels was visualized in Gel Doc (Biorad). The amplicons obtained from PCR were confirmed by sanger sequencing.

5'-Rapid Amplification of cDNA End (5'-RACE):

5'-RACE was performed according to the previously published study [10](Mukherjee et al., 2012). Briefly, 36ng of heat-denatured cytoplasmic oligodT bead selected RNA from U2OS cells was incubated at 37° for 2 hours with 50 pmol/µl of RNA adaptor-2 and 10 units of T4 RNA ligase followed by phenol chloroform extraction and ethanol precipitation. The recovered RNA (~30 ng) was reverse transcribed and analyzed by semi quantitative RT-PCR using an adaptor specific forward primer and target gene specific reverse primer. The PCR products were separated on 1.2% agarose gel and visualized using Biorad Gel-doc system. The PCR products were purified from the gel using NucleoSpin Gel and PCR Clean-up Mini kit (Macherey-Nagel). These purified PCR products were cloned individually in a TA-cloning vector called pGEM®-T Easy Vector (Promega) according to the manufacturer's instructions and sent for Sanger sequencing to AgriGenome Labs Pvt Ltd. The sequence of GFP RNA oligo whivh was used as spike in control for RACE assay was provided in Table-2. The sequences of adaptor, adaptor specific primer and GFP specific reverse primer were listed in Table-2.

Results and Discussion

Nuclear and cytoplasmic localization of selected lncRNAs

To identify lncRNAs which might be regulated by cCE, we limited our *in silico* search for lncRNAs expressed in sarcoma since the complex of cytoplasmic capping and its target mRNAs have been characterized in osteosarcoma (U2OS) cells. Cap analysis of gene expression (CAGE), a technique, developed to identify transcription start sites [27], was often used to determine the capped sites. Earlier studies demonstrated uncapped ends in target mRNAs of cCE mapped exactly or in the vicinity of the downstream CAGE tags [25-26] indicating participation of cytoplasmic capping in generating novel capped ends. Based on this, we have selected fifteen lncRNAs with multiple CAGE tags as obtained from UCSC genome browser. These selected lncRNAs vary in their sizes and also in their types as shown in Figure 1A. U2OS cells were fractionated in nuclear and cytoplasmic fractions to examine

the subcellular localization of candidate transcripts (Figure 1B) and our data show all of these transcripts were expressed in both fractions (Figure 1C-F). The functions of lncRNAs are linked with their subcellular localization. Here, our data shows some lncRNAs like RP1, or different transcripts of GAS5 have similar nuclear and cytoplasmic localization (Figure 1C). Different transcripts of RP11 have poor expression in cytosol compared to nucleus (Figure 1C-D). Malat1 which is known to be associated with nuclear paraspeckles[27-28] is expressed well in cytosolic fraction of U2OS cells (Figure 1F). We have only obtained nonspecific amplification for GAS5-838 where the band corresponds to 200 bp (Figure 1C) is the correct transcript as revealed by Sanger sequencing.

Detection of native uncapped forms of lncRNAs in U2OS cells

LncRNAs are believed to contain poly A tails [31] which led us to examine if the candidate lncRNAs expressed in the cytoplasm contain polyA residues. The candidate transcripts contain poly A tails as evidenced by recovery of these transcripts on recovered RNA fractions by oligo dT beads (Figure 2A-B) from cytoplasmic RNA. Earlier genome wide study shows both protein coding or noncoding RNAs can undergo diverse processing at their 5' ends in the cytoplasm [20]. Previous studies could detect uncapped mRNAs in U2OS cells [10]. To determine if the candidate transcripts may undergo similar mechanism, the cytoplasmic fraction containing polyadenylated RNA were ligated to an RNA adaptor followed by 5' RACE assay using gene specific primer and adaptor specific primer. To test the efficiency of RNA ligation, another synthetic RNA oligo encoding GFP was used as spike in control as described in Mukherjee et al. Recovery of GFP synthetic RNA oligo and VDAC3 mRNA transcripts that were detected earlier [10] after 5'RACE assay showed the efficiency of the reaction (Figure 2C). Our results showed multiple uncapped forms of candidate lncRNAs in native U2OS cells (Figure 2C-D). We could detect RACE products for several transcripts except LINC00969-380 in U2OS cells (Figure 2C).

However, there are also a few transcripts which showed multiple uncapped forms but recovered in lower amount (Figure 2D). Next, to validate if these amplicons are indeed generated from 5'-RACE experiment, we excised the bands as shown in Figure 2C from agarose gels and clone these products in TA vector for confirmation by sequencing. Cloning was successful for RACE products of RP1-283E3.8, GAS5-245, H19-009 and NFYC-AS1 and these cloned amplicons were produced from successful ligation of RNA adaptor with the uncapped ends of lncRNAs as shown in Figure 2E and Table- 2.

Accumulation of uncapped forms of lncRNAs in cytoplasmic capping inhibited cells

Earlier studies have shown mRNAs with uncapped ends accumulate more in cytoplasmic capping inhibited cells [10], [15], [29]. The presence of uncapped lncRNAs led us to examine if these transcripts are targeted by cytoplasmic capping complex. To study cytoplasmic capping independent of nuclear capping, a dominant negative mutant of cCE (K294A), that was restricted to express in the cytoplasm [8], with N terminal biotin and Myc tag have been developed. U2OS cells transiently transfected with empty vector (control) pcDNA4TO or pcDNA4TO-bio-myc-K294A were used to conduct the 5'RACE experiment as before. Equal recovery of GFP spike in control from control cells and K294A expressing cells indicated similar efficiency of RNA ligation reactions in this experiment. Interestingly, several transcripts were accumulated more in cytoplasmic capping inhibited cells expressing epitope tagged K294A (Figure 3A) compared to control cells (Figure 3C, D, E). There were several transcripts like ENTPD3-AS1, H19-009 (Figure 3C, lanes 1-4), LINC00969-380 (Figure 3D, lanes 11, 12), ASH1L-AS1 (Figure 3E, lanes 11, 12) which are readily detectable in cells expressing K294A only, not in the control cells. Our results also showed accumulation of VDAC3 mRNA in cytoplasmic capping inhibited cells (Figure 3D, lanes 3-4) which was reported earlier a target of cytoplasmic recapping.

To confirm if these accumulation of uncapped forms of lncRNAs generated due to blocking cytoplasmic capping or overexpression of K294A, we took another alternative approach to inhibit cytoplasmic capping. As stated earlier, Nck1 protein serves as the scaffold molecule to form cytoplasmic capping

complex and blocking the interaction of cCE with Nck1 inhibits formation of the complex [13]. U2OS cells were transiently transfected with either empty vector or inactive mutant of HA-Nck1 protein (HA-Nck1-M3) that could not bind to cCE [13]. 5'-RACE assay was performed similarly with cytoplasmic RNA extracted from cells overexpressing HA-Nck1 M3 or control (Figure 3B). Interestingly, increased uncapped forms of several transcripts were found in cells overexpressing Nck1-M3 compared control (Figure 3F-G). This result confirmed the role of cytoplasmic capping in accumulating the native uncapped forms of selected lncRNAs. However, lncRNA GAS5-245 was independent of cytoplasmic capping as no accumulation of uncapped forms were found in cells expressing K294A or Nck1-M3 (Figure 3D, lanes 9-10, Figure 3G, lanes 7-8). We were unable to detect native uncapped forms of LINC00969-380 in RACE experiment which might be due to its poor recovery on oligo dT beads as shown in Figure 2A.

Overall, our results showed lncRNAs could be targeted by cCE. We have detected multiple uncapped forms of lncRNAs and also identified the sequence of their uncapped ends by 5' RACE assay. The generation of diverse uncapped forms of lncRNAs in cytoplasmic capping inhibited cells overexpressing K294A and Nck1-M3 suggested these lncRNAs could be processed differentially and could be recapped at different sites. Future CAGE analysis of the sequences obtained from RACE experiment will identify recapped sites. In addition, in depth analysis will be performed to identify presence of any particular sequences that might enable these target lncRNAs to be targeted by cytoplasmic capping. Future studies should be directed to reveal role of these lncRNAs in cap homeostasis.

Acknowledgements

The authors thank Kiss Lab, Houston Methodist Research Hospital, Texas for their help in CAGE analysis. The authors also thank the Department of Science and Technology, Government of India (CRG/2019/006427) for providing fellowship to Mr. Avik Mukherjee and financial support to conduct the experiments; UGC, Government of India for UGC-SRF fellowship to Mr. Safirul Islam and Department of Biotechnology, India for Ramlingaswami Re-entry fellowship to Dr. Chandrama Mukherjee. The authors also thank Presidency University for necessary infrastructural support and Dr. Abhik Saha for providing access to Licor imager.

References

[1] Z. Yue, E. Maldonado, R. Pillutla, H. Cho, D. Reinberg, and A. J. Shatkin, "Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II," *Proc. Natl. Acad. Sci.*, vol. 94, no. 24, pp. 12898–12903, Nov. 1997, doi: 10.1073/pnas.94.24.12898.

[2] V. H. Cowling, "Regulation of mRNA cap methylation," *Biochem. J.*, vol. 425, no. 2, pp. 295–302, Jan. 2010, doi: 10.1042/BJ20091352.

[3] N. Visa, E. Izaurralde, J. Ferreira, B. Daneholt, and I. W. Mattaj, "A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export.," *J. Cell Biol.*, vol. 133, no. 1, pp. 5–14, Apr. 1996, doi: 10.1083/jcb.133.1.5.

[4] L. Philippe, J.-J. Vasseur, F. Debart, and C. C. Thoreen, "La-related protein 1 (LARP1) repression of TOP mRNA translation is mediated through its cap-binding domain and controlled by an adjacent regulatory region.," *Nucleic Acids Res.*, vol. 46, no. 3, pp. 1457–1469, 2018, doi: 10.1093/nar/gkx1237.

[5] D. R. Schoenberg and L. E. Maquat, "Regulation of cytoplasmic mRNA decay," *Nature Reviews Genetics*, vol. 13, no. 4. 2012, doi: 10.1038/nrg3160.

[6] F. Bélanger, J. Stepinski, E. Darzynkiewicz, and J. Pelletier, “Characterization of hMTr1, a human Cap1 2’-O-ribose methyltransferase.,” *J. Biol. Chem.*, vol. 285, no. 43, pp. 33037–33044, Oct. 2010, doi: 10.1074/jbc.M110.155283.

[7] M. Werner *et al.*, “2’-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family.,” *Nucleic Acids Res.*, vol. 39, no. 11, pp. 4756–68, Jun. 2011, doi: 10.1093/nar/gkr038.

[8] Y. Otsuka, N. L. Kedersha, and D. R. Schoenberg, “Identification of a cytoplasmic complex that adds a cap onto 5’-monophosphate RNA.,” *Mol. Cell. Biol.*, vol. 29, no. 8, pp. 2155–67, Apr. 2009, doi: 10.1128/MCB.01325-08.

[9] R. J. Machida and Y.-Y. Lin, “Four methods of preparing mRNA 5’ end libraries using the Illumina sequencing platform.,” *PLoS One*, vol. 9, no. 7, p. e101812, 2014, doi: 10.1371/journal.pone.0101812.

[10] C. Mukherjee, D. P. Patil, B. A. Kennedy, B. Bakthavachalu, R. Bundschuh, and D. R. Schoenberg, “Identification of Cytoplasmic Capping Targets Reveals a Role for Cap Homeostasis in Translation and mRNA Stability,” *Cell Rep.*, vol. 2, no. 3, pp. 674–684, Sep. 2012, doi: 10.1016/j.celrep.2012.07.011.

[11] P. Chen *et al.*, “Capping Enzyme mRNA-cap/RNGTT Regulates Hedgehog Pathway Activity by Antagonizing Protein Kinase A.,” *Sci. Rep.*, vol. 7, no. 1, p. 2891, 2017, doi: 10.1038/s41598-017-03165-2.

[12] A. V. Ignatochkina, Y. Takagi, Y. Liu, K. Nagata, and C. K. Ho, “The messenger RNA decapping and recapping pathway in Trypanosoma.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 22, pp. 6967–72, Jun. 2015, doi: 10.1073/pnas.1424909112.

[13] C. Mukherjee, B. Bakthavachalu, and D. R. Schoenberg, “The Cytoplasmic Capping Complex Assembles on Adapter Protein Nck1 Bound to the Proline-Rich C-Terminus of Mammalian Capping Enzyme,” *PLoS Biol.*, vol. 12, no. 8, 2014, doi: 10.1371/journal.pbio.1001933.

[14] T. Gonatopoulos-Pournatzis, S. Dunn, R. Bounds, and V. H. Cowling, “RAM/Fam103a1 is required for mRNA cap methylation.,” *Mol. Cell*, vol. 44, no. 4, pp. 585–96, Nov. 2011, doi: 10.1016/j.molcel.2011.08.041.

[15] D. L. Kiss, K. Oman, R. Bundschuh, and D. R. Schoenberg, “Uncapped 5’ ends of mRNAs targeted by cytoplasmic capping map to the vicinity of downstream CAGE tags.,” *FEBS Lett.*, vol. 589, no. 3, pp. 279–84, Jan. 2015, doi: 10.1016/j.febslet.2014.12.009.

[16] B. Wu, C. Eliscovich, Y. J. Yoon, and R. H. Singer, “Translation dynamics of single mRNAs in live cells and neurons.,” *Science*, vol. 352, no. 6292, pp. 1430–5, Jun. 2016, doi: 10.1126/science.aaf1084.

[17] X. Yan, T. A. Hoek, R. D. Vale, and M. E. Tanenbaum, “Dynamics of Translation of Single mRNA Molecules In Vivo.,” *Cell*, vol. 165, no. 4, pp. 976–89, May 2016, doi: 10.1016/j.cell.2016.04.034.

[18] S. Iwasaki and N. T. Ingolia, “Seeing translation,” *Science (80-.).*, vol. 352, no. 6292, pp. 1391–1392, Jun. 2016, doi: 10.1126/science.aag1039.

[19] B. A. Agana, V. H. Wysocki, and D. R. Schoenberg, “Cytoplasmic mRNA recapping has limited impact on proteome complexity,” *Open Biol.*, vol. 10, no. 11, p. 200313, Nov. 2020, doi: 10.1098/rsob.200313.

[20] T. R. G. Katalin Fejes-Toth, Vihra Sotirova, Ravi Sachidanandam, Gordon Assaf, Gregory J

Hannon, Philipp Kapranov, Sylvain Foissac, Aarron T Willingham, Radha Duttagupta, Erica Dumais, “Post-transcriptional processing generates a diversity of 5’-modified long and short RNAs.,” *Nature*, vol. 457, no. 7232, pp. 1028–32, Feb. 2009, doi: 10.1038/nature07759.

[21] T. Chujo, T. Yamazaki, and T. Hirose, “Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies.,” *Biochim. Biophys. Acta*, vol. 1859, no. 1, pp. 139–46, Jan. 2016, doi: 10.1016/j.bbagrm.2015.05.007.

[22] T. Chujo and T. Hirose, “Nuclear Bodies Built on Architectural Long Noncoding RNAs: Unifying Principles of Their Construction and Function.,” *Mol. Cells*, vol. 40, no. 12, pp. 889–896, Dec. 2017, doi: 10.14348/molcells.2017.0263.

[23] T. Yamazaki and T. Hirose, “The building process of the functional paraspeckle with long non-coding RNAs.,” *Front. Biosci. (Elite Ed).*, vol. 7, no. 1, pp. 1–41, 2015, doi: 10.2741/715.

[24] J.-H. Yoon, J. Kim, and M. Gorospe, “Long noncoding RNA turnover.,” *Biochimie*, vol. 117, no. 3, pp. 15–21, Oct. 2015, doi: 10.1016/j.biochi.2015.03.001.

[25] S. Kebache, D. Zuo, E. Chevet, and L. Larose, “Modulation of protein translation by Nck-1.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 8, pp. 5406–11, Apr. 2002, doi: 10.1073/pnas.082483399.

[26] M. Chen, H. She, A. Kim, D. T. Woodley, and W. Li, “Nckbeta adapter regulates actin polymerization in NIH 3T3 fibroblasts in response to platelet-derived growth factor bb.,” *Mol. Cell. Biol.*, vol. 20, no. 21, pp. 7867–80, Nov. 2000, doi: 10.1128/MCB.20.21.7867-7880.2000.

[27] T. Shiraki *et al.*, “Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 26, pp. 15776–81, Dec. 2003, doi: 10.1073/pnas.2136655100.

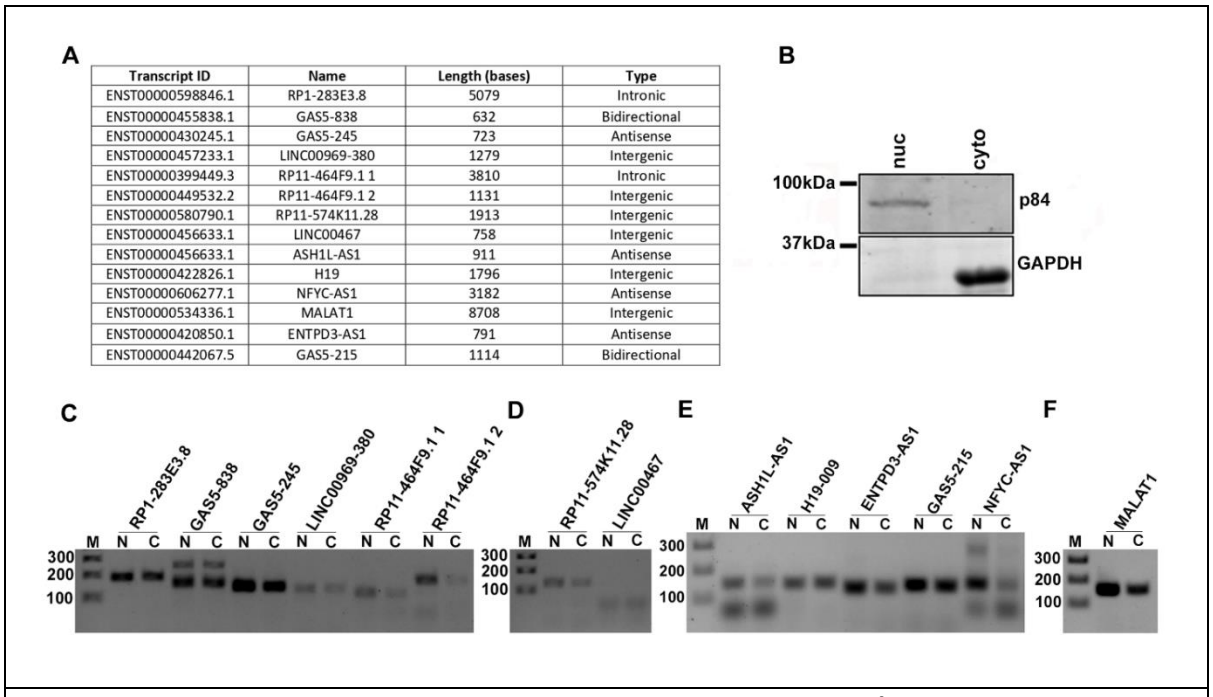
[28] D. L. Kiss, K. Oman, R. Bundschuh, and D. R. Schoenberg, “Uncapped 5’ ends of mRNAs targeted by cytoplasmic capping map to the vicinity of downstream CAGE tags,” *FEBS Lett.*, vol. 589, no. 3, 2015, doi: 10.1016/j.febslet.2014.12.009.

[29] M. R. Berger, R. Alvarado, and D. L. Kiss, “mRNA 5’ ends targeted by cytoplasmic recapping cluster at CAGE tags and select transcripts are alternatively spliced.,” *FEBS Lett.*, vol. 593, no. 7, pp. 670–679, 2019, doi: 10.1002/1873-3468.13349.

[30] H. Choudhry *et al.*, “Extensive regulation of the non-coding transcriptome by hypoxia: role of HIF in releasing paused RNApol2.,” *EMBO Rep.*, vol. 15, no. 1, pp. 70–6, Jan. 2014, doi: 10.1002/embr.201337642.

[31] L. Statello, C.-J. Guo, L.-L. Chen, and M. Huarte, “Gene regulation by long non-coding RNAs and its biological functions.,” *Nat. Rev. Mol. Cell Biol.*, vol. 22, no. 2, pp. 96–118, Feb. 2021, doi: 10.1038/s41580-020-00315-9.

Figures



A

Transcript ID	Name	Length (bases)	Type
ENST00000598846.1	RP1-283E3.8	5079	Intronic
ENST00000455838.1	GAS5-838	632	Bidirectional
ENST00000430245.1	GAS5-245	723	Antisense
ENST00000457233.1	LINC00969-380	1279	Intergenic
ENST00000399449.3	RP11-464F9.1 1	3810	Intronic
ENST00000449532.2	RP11-464F9.1 2	1131	Intergenic
ENST00000580790.1	RP11-574K11.28	1913	Intergenic
ENST00000456633.1	LINC00467	758	Intergenic
ENST00000456633.1	ASH1L-AS1	911	Antisense
ENST00000422826.1	H19	1796	Intergenic
ENST00000606277.1	NFYC-AS1	3182	Antisense
ENST00000534336.1	MALAT1	8708	Intergenic
ENST00000420850.1	ENTPD3-AS1	791	Antisense
ENST00000442067.5	GAS5-215	1114	Bidirectional

Fig 1: Nuclear and cytoplasmic localization of selected lncRNAs: (A) Selected lncRNAs for this study. (B) Cytoplasmic and nuclear protein extracts were prepared from U2OS cells. The quality of fractionation is demonstrated by Western blotting using antibodies against nuclear marker protein p84 and cytoplasmic marker protein GAPDH. The molecular weight of the proteins was indicated. (C-F) Nuclear and cytoplasmic RNA were extracted from the respective extracts. Detection of lncRNAs in nuclear and cytoplasmic fractions. N represents Nuclear RNA and C represents Cytoplasmic RNA, M: 100bp DNA marker.

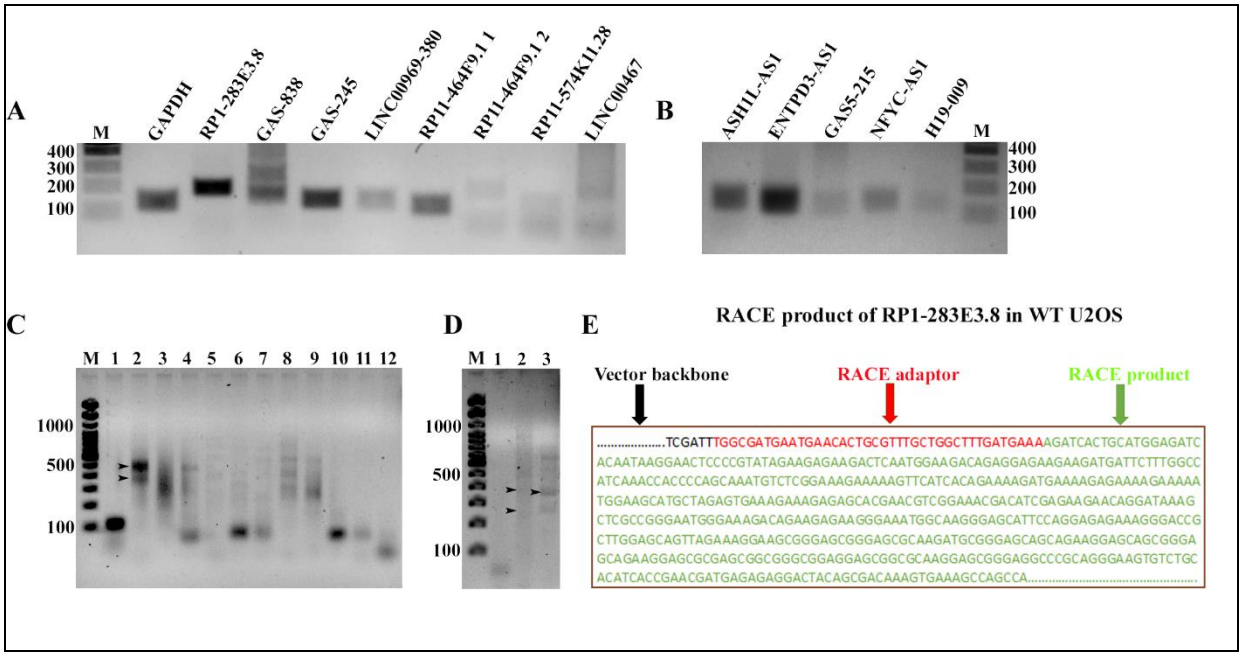


Fig 2: Detection of native uncapped forms of lncRNAs. (A-B) Detection of lncRNAs in poly (A) containing RNAs isolated from total cytoplasmic RNA. (C-D) The presence of uncapped ends of candidate lncRNAs in parental U2OS cells as assessed 5'-RACE technique using an RNA adaptor-ligated poly (A) enriched RNAs. The lanes are shown below,
(C) lane 1: GFP spike in control, lane 2: RP1-283E3.8, lane 3: GAS5-838, lane 4: GAS5-245, lane 5: LINC00969-380, lane 6: RP11-464F9.1 1, lane 7: RP11-464F9.1 2, lane 8: RP11-574K11.28, lane 9: MALAT-1, lane 10: LINC00467, lane 11: ASH1-AS1, lane 12: VDAC3 and (D) lane 1: ENTPD3-AS1, lane 2: H19, lane 3: NFYC-AS1 (D). M represents 100 bp marker. Arrows in the figures indicate the amplicons that are excised from gel, cloned in TA vector and sent for Sanger sequencing. (E) The sequence obtained from RACE product generated by RP1-283E3.8 is shown. Vector sequence, adaptor sequence and uncapped ends with the RACE product are shown as indicated.

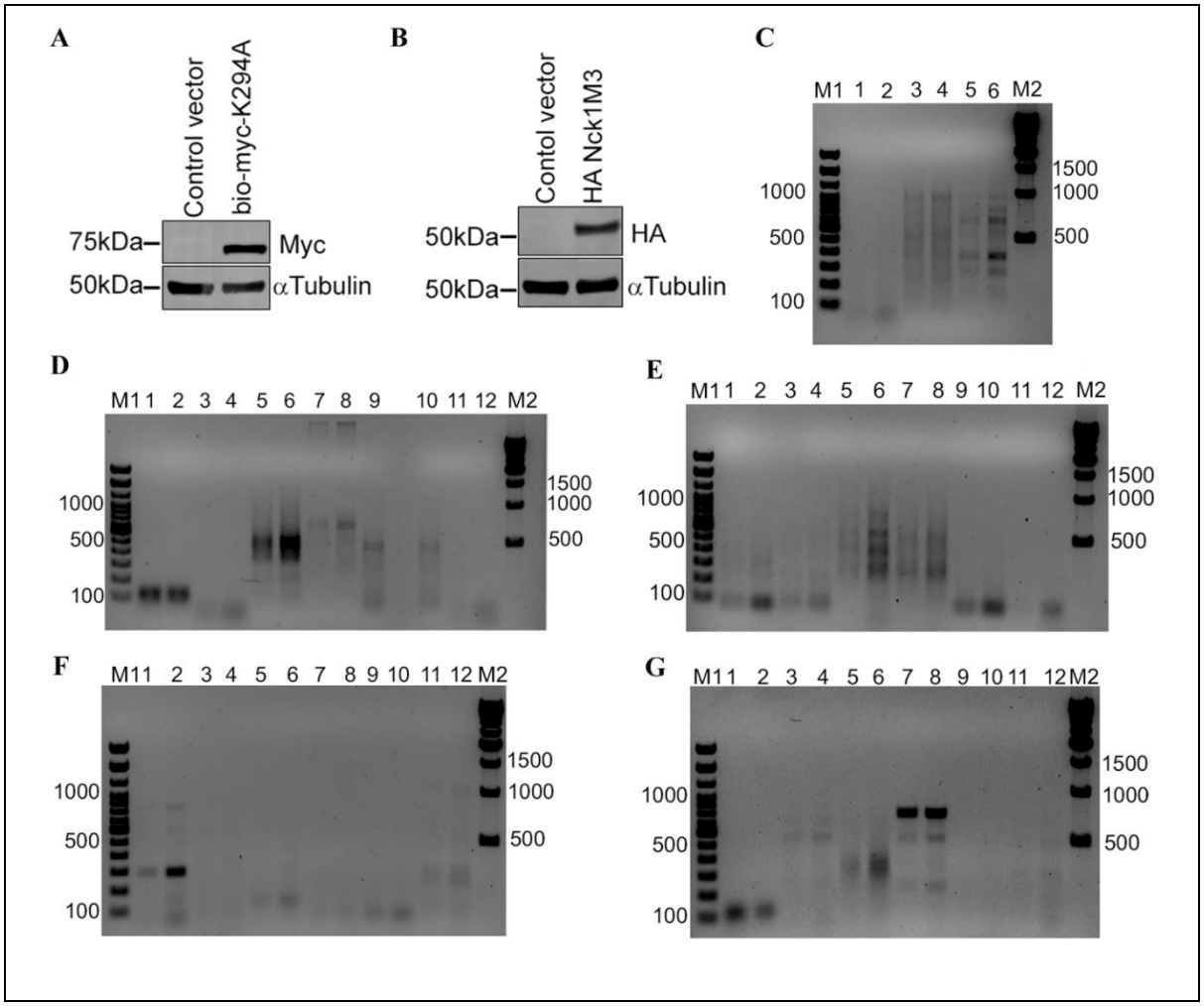


Figure 3: Accumulation of native uncapped lncRNAs in cytoplasmic capping inhibited cells. (A-B) The exogenous expression of bio- myc-K294A and HA- Nck1M3 are analysed by western blotting with anti-Myc and anti-HA antibodies on cytoplasmic protein extracted from transiently transfected cells with constructs encoding Bio-Myc-K294A and HA-Nck1 M3. Anti-alpha-Tubulin antibody is used to detect equal amount of protein loaded in each panel. The molecular weight of the proteins was indicated. (C-E): The presence of uncapped ends of candidate lncRNAs in U2OS cells either transfected with empty vector (lanes 1,3,5,7,9,11) or pcDNA4TO-Bio-Myc-K294A (lanes 2,4,6,8,10,12). The lncRNAs are indicated as, C) lane 1,2: ENTPD3-AS1, lane 3,4: H19-009, lane 5,6: NFYC-AS1. (D) lane 1,2: GFP spike-in control, lane 3,4: VDAC3 mRNA, lane 5,6: RP1-283E3.8, lane 7,8: GAS5-838, lane 9,10: GAS5-245, lane 11,12: LINC00969-380. (E) lane 1,2: RP11-464F9.1 1, lane 3,4: RP11-464F9.1 2, lane 5,6: RP11-574K11.28, lane 7,8: MALAT-1, lane 9,10: LINC00467, lane 11,12: ASH1L-AS1. (F-G) The presence of uncapped ends of candidate lncRNAs in U2OS cells either transfected with empty vector (lanes 1,3,5,7,9,11) or pRK5-HANck1 M3 (lanes 2,4,6,8,10,12). The lncRNAs are indicated as, (F) lane 1,2: RP11-464F9.1 2, lane 3,4: RP11-574K11.28, lane 5,6: MALAT-1, lane 7,8: ASH1L-AS1, lane 9,10: H19-009, lane 11,12: NFYC-AS1. (G) lane 1,2: GFP spike-in control, lane 3,4: RP1-283E3.8, lane 5,6: GAS5-838, lane 7,8: GAS5-245, lane 9,10: LINC00969-380, lane 11,12: RP11-464F9.1 1. M1 and M2 represent 100 bp and 500 bp marker respectively.

Transcript ID	**Name**	**Sequence information for sense (FP) and antisense primers (RP)**
ENST00000598846.1	RP1	FP: GCAGTTAGAAAGGAAGCGGG RP: TGGCTGGCTTTCACTTTGTC
ENST00000455838.1	GAS5-838	FP: CAGCGGCACTATCAGTTCAC RP: AAGAGTTCGAGACCAGCCTG
ENST00000430245.1	GAS5-245	FP: GAGCAAGCCTAACTCAAGCC RP: CCCAAGCAAGTCATCCATGG
ENST00000414625.2	LINC00969-380	FP: GACAACTGGAGGTGGCATTT RP: AACGGCATGCCATAATTTTC
ENST00000399449.3	RP11-1	FP: ACCCAAAGCTGCAAAGAAAA RP: AGCCATCCACACAGCAGACT
ENST00000449532.2	RP11-10	FP: CCGACCTGAGGAGTCACAGT RP: TCCAGGCATCTTCTTCGTCT
ENST00000580790.1	RP11-574	FP: TCAAGGGGACGAGTTCTCAC RP: TAATAGGCCAGCCGTAGTGG
ENST00000456633.1	ASH1L-AS1	FP: CGGTTGACCTGAGCCTACTT RP: AGCGGTACAATACCCGTGTC
ENST00000456633.1	LINC00467	FP: TGTCAAGTCGGACCAAACAG RP: GTGCTGGTTTTGGTTTTGCT
ENST00000422826.1	H19-009	FP: CGTTCCAGGCAGAAAGAGCAA RP: GGACCCCTCTGTCCTGTGT
ENST00000606277.1	NFYC-AS1	FP: CAGGCAAGAGCACAACATGG RP: AGTTTCTGTGCGAGGGTGAT
ENST00000534336.1	MALAT1	FP: GCAGGGAGAATTGCGTCATT RP: TTCTTCGCCTTCCCGTACTT
ENST00000420850.1	ENTPD3-AS1	FP: CGTCTGCTACACCCAACTGA RP: TCCTTGAGTAGAGGCCAAGC
ENST00000442067.5	GAS5-215	FP: GCACTATCAGTTCACTGCAACC

		RP: CAGCCTGGACAACATAGCTGTA
ENST00000619601.1	GAPDH	FP: GTCTCCTCTGACTTCAACAGCG RP: ACCACCCTGTTGCTGTAGCCAA
Uncapped GFP RNA (Spike-in control):		UUU CAC UGG AGU UGU CCC AAU UCU UGU UGA AUU AGA UGG CGA UGU UAA UGG GCA AAA AUU CUC UGU CAG UGG AGA GGG UGA AGG UGA UGC AAC AUA CGG AA
5'RACE adaptor		GCU GAU GGC GAU GAA UGA ACA CUG CGU UUG CUG GCU UUG AUG AAA
Adaptor specific primer		TGG CGA TGA ATG AAC ACT GC

Table1: Sequences of primers and oligonucleotides used for PCR and 5'RACE assay

Transcript	**Excised band size from gel (bp)**	**Sequence of the 5'-RACE amplicons**
RP1-283E3.8	~500	TCGATTTGGCGATGAATGAACACTGCGTTTGCTGGCTTTGATGAAAAGATCACTGCATGGAGATCACAATAAGGAACTCCCCGTATAGAAGAGAAGACTCAATGGAAGACAGAGGAGAAGAAGATGATTCTTTGGCCATCAAACCACCCCAGCAAATGTCTCGGAAAGAAAAAGTTCATCACAGAAAAGATGAAAAGAGAAAAGAAAAATGGAAGCATGCTAGAGTGAAAGAAAGAGAGCACGAACGTCGGAAACGACATCGAGAAGAACAGGATAAAGCTCGCCGGGAATGGGAAAGACAGAAGAGAAGGGAAATGGCAAGGGAGCATTCCAGGAGAGAAAGGGACCGCTTGGAGCAGTTAGAAAGGAAGCGGGAGCGGGAGCGCAAGATGCGGGAGCAGCAGAAGGAGCAGCGGGAGCAGAAGGAGCGCGAGCGGCGGGCGGAGGAGCGGCGCAAGGAGCGGGAGGCCCGCAGGGAAGTGTCTGCACATCACCGAACGATGAGAGAGGACTACAGCGACAAAGTGAAAGCCAGCCA
RP1-283E3.8	~400	TCGATTTGGCGATGAATGAACACTGCGTTTGCTGGCTTTGATGAAAAAGATGAAAAGAGAAAAGAAAAATGGAAGCATGCTAGAGTGAAAGAAAGAGAGCACGAACGTCGGAAACGACATCGAGAAGAACAGGATAAAGCTCGCCGGGAATGGGAAAGACAGAAGAGAAGGGAAATGGCAAGGGAGCATTCCAGGAGAGAAAGGGACCGCTTGGAGCAGTTAGAAAGGAAGCGGGAGCGGGAGCGCAAGATGCGGGAGCAGCAGAAGGAGCAGCGGGAGCAGAAGGAGCGCGAGCGGCGGGCGGAGGAGCGGCGCAAGGAGCGGGAGGCCCGCAGGGAAGTGTCTGCACATCACCGAACGATGAGAGAGGACTACAGCGACAAAGTGAAAGCCAGCCA
H19-009	~400	TCGATTTGGCGATGAATGAACACTGCGTTTGCTGGCTTTGATGAAAATCCCAGGGCCTGGGCAACGGAGGTGTAGCTGGCAGCAGCGGGCAGGTGAGGACCCCATCTGCCGGGCAGGTGAGTCCCTTCCCTCCCCAGGCCTCGCTTCCCCAGCCTTCTGAAAGAAGGAGGTTTAGGGGATCGAGGGCTGGCGGGGAGAAGCAGACACCCTCCCAGCAGAGGGGCAGGATGGGGGCAGGAGAGTTAGCAAAGGTGACATCTTCTCGGGGGGAGCCGAGACTGCGCAAGGCTGGGGGGTTATGGGCCCGTTCCAGGCAGAAAGAGCAAGAGGGCAGGGAGGGAGCACAGGGGTGGCCAGCGTAGGGTCCAGCACGTGGGGTGGTACCCCAGGCCTGGGTCAGACAGGGACATGGCAGGGGACACAGGACAGAGGGGTCC

H19-009	~300	TCGATTTGGCGATGAATGAACACTGCGTTTGCTGGCTTTGATGAAACCAGCCTTCTGAAAGAAGGAGGTTTAGGGGATCGAGGGCTGGCGGGGAGAAGCAGACACCCTCCCAGCAGAGGGGCAGGATGGGGGCAGGAGAGTTAGCAAAGGTGACATCTTCTCGGGGGGAGCCGAGACTGCGCAAGGCTGGGGGGTTATGGGCCCGTTCCAGGCAGAAAGAGCAAGAGGGCAGGGAGGGAGCACAGGGGTGGCCAGCGTAGGGTCCAGCACGTGGGGTGGTACCCCAGGCCTGGGTCAGACAGGGACATGGCAGGGGACACAGGACAGAGGGGTCC
NFYC-AS1	~300	TCGATTTGGCGATGAATGAACACTGCGTTTGCTGGCTTTGATGAAAAATAAGGTGCCAGGTCCGCGGCCTCCCCGCCACCGACCAATCACGAGCGAGCCCCCCCGCCGTCGCCTTCCGCCAATGGGAGAAAGCCCGCGGGCCGGGCCGACGAGGCACACGGACCCCACGAGCGCGGCGGCGGCGACAGCAGCGGGCCAGGCAAGAGCACAACATGGCGGCGGGCAGGACGGACCGAGCTACAGCAGCGGCCGCGCGAGGAAGTGAGGTCTCTGGCGAAGGCGGGAAGCGCTGCCCCTCCCCCCGCCGCCGCTTCCCTATCACCCTCGCACAGAAACT

Table 2: Sequences of the 5'-RACE amplicons of selected lncRNAs. Here, the red colour denotes the sequence from TA-cloning vector, underlined sequence denotes the RNA adaptor, rest indicates the candidate transcript with uncapped ends.