Functions of long noncoding RNAs in the nucleus

Bin Yu & Ge Shan

To cite this article: Bin Yu & Ge Shan (2016): Functions of long noncoding RNAs in the nucleus, Nucleus, DOI: 10.1080/19491034.2016.1179408

To link to this article: http://dx.doi.org/10.1080/19491034.2016.1179408

Accepted author version posted online: 22 Apr 2016.
Functions of long noncoding RNAs in the nucleus

Bin Yu, Ge Shan#

The CAS Key Laboratory of Innate Immunity and Chronic Disease, CAS Center for Excellence in Molecular Cell Science, School of Life Sciences, University of Science and Technology of China, 443 Huangshan Road, Hefei, Anhui Province 230027, China

#Corresponding author; E-mail: shange@ustc.edu.cn (G. S.).

Abstract

Nucleus is the residence and place of work for a plethora of long noncoding RNAs. Here, we provide a summary of the functions and functional mechanisms of several relatively well studied examples of nuclear long noncoding RNAs (IncRNAs) in the nucleus, such as Xist, NEAT1, MALAT1 and TERRA. The recently identified novel EIciRNA is also highlighted. These nuclear IncRNAs play a variety of roles with diverse molecular mechanisms in animal cells. We also discuss insights and concerns about current and future studies of nuclear IncRNAs.
Introduction

Eukaryotic transcripts can be either coding or noncoding based on the presence or absence of functional open reading frame. Noncoding RNAs (ncRNAs) are divided arbitrarily into small (<200 nt) and long (>200 nt) ncRNAs (IncRNAs) based on their length. RNAs are transcribed from the nuclear genome in animal cells; the original place of essentially all RNAs is the nucleus, except the small fraction encoded by the mitochondrial DNA. Most ncRNAs small or long, generally go through several steps of processing or biogenesis in the nucleus. For some ncRNAs, the final residence and functional place is the cytoplasm, where they may undergo further processing steps. For example, the ~20 nt microRNAs are processed from nuclear primary microRNA transcripts with two steps of endoribonuclease cleavage, first in the nucleus and then in the cytoplasm. Actually, a lot of small ncRNAs are originated or processed from precursors or primary forms of IncRNAs. For small ncRNAs, they may reside and play roles either in the cytoplasm or in the nucleus. There are also reports about some small ncRNAs, such as microRNA-29b, getting back into the nucleus and may possess nuclear functions.1

There are a myriad of ncRNAs especially IncRNAs that remain and function solely in the nucleus. The long studied IncRNA Xist may be the founding member of the long regulatory ncRNAs, which by chance is also a nuclear IncRNA. The past decades have witnessed a dramatic increase in the inventory of nuclear IncRNAs. Diverse roles and multiple functional mechanisms have now been linked to IncRNAs in the nucleus.2-4 Our recent work has found a novel class of
lncRNAs localizing in the nucleus and regulating transcription. In this extra view, we summarize the functions and functional mechanisms of some of the relatively well studied nuclear lncRNAs, and highlight our recent identification of EIciRNAs. Insights, current thinking and some concerns about nuclear roles of lncRNAs are discussed.

**Xist and RoX**

In mammals, a long ncRNA Xist is responsible for the inactivation of one X chromosome in female. The phenomenon of X chromosome inactivation in females was documented in 1949, and the central player Xist, a ~17 knt (kilo-nucleotide) ncRNA in human was identified more than four decades later. Xist RNA is essential and responsible for the initiation, propagation, and maintenance of X chromosome inactivation via the recruiting of an array of epigenetic regulators. Xist is expressed from the future inactive X-chromosome. With the accumulation of the Xist RNA, it coats and spreads on the X-chromosome in cis. Xist directly interacts with SHARP, which recruits the co-repressor SMRT to regulate HDAC3, and eventually leads to the deacetylation of histones and some other epigenetic events to silence the transcription on the inactive X-chromosome. Xist RNA also recruits the polycomb repressive complex 1 and 2 (PRC1 and PRC2) to induce the PRC1 mediated mono-ubiquitylation of histone H2AK119 and the PRC2-dependent H3K37me3, respectively. A recent study shows that HnrrpK and Spen specifically interact with Xist. This interaction is not essential for the localization of Xist to the inactive X-chromosome but required for the silencing effect of Xist. A very interesting
phenomenon is that the sequence and length of Xist are not conserved among different branches of mammals, although all mammalian Xist RNAs are composed of repeats of short units.\textsuperscript{24-28}

In \textit{Drosophila melanogaster}, compensation of X-chromosome dosage is fulfilled by transcriptional upregulation of the single X chromosome in male.\textsuperscript{28} Two lncRNAs roX1 and roX2, both transcribed from the X chromosome in male, are required for the X-chromosome activation by recruiting the Male-Specific Lethal (MSL) protein complex to increase histone H4 acetylation.\textsuperscript{29-37} RoX1 is \(\sim 3.6\) knt, and RoX2 is \(\sim 1.1\) knt in length. Both RoX1 and RoX2 paint the X chromosome via chromatin entry sites, and they are partially redundant in the X chromosome activation. Although both RoX1 and RoX2 are encoded by the X chromosome, their effects on X chromosome activation can be \textit{in trans}, as RoX RNAs wrongly or artificially expressed from autosome lead to the same X chromosome activation in male.\textsuperscript{34,37}

Xist and RoX RNAs are the first identified regulatory lncRNAs, and either by coincidence both regulate dosage compensation of X chromosome in animals. They also set up the concept that nuclear lncRNAs can be powerful master regulators of chromatin epigenetic status.

**HOTAIR**

For a long period, Xist and RoX RNAs were considered the only nuclear regulatory lncRNAs, until the milestone identification of lincRNA HOTAIR in 2007.\textsuperscript{38} LincRNA stands for long intergenic noncoding RNA, which is a class of lncRNAs transcribed in between genomic regions of coding genes. HOTAIR is transcribed from the intergenic region of HoxC gene loci and with a
length of \(~2.2\) knt.\textsuperscript{38} HOTAIR binds the polycomb repressive complex 2 (PRC2) and the LSD1/CoREST/REST complex, and thus modulates histone modifications of target genes \textit{in trans} to silence the transcription of HoxD loci.\textsuperscript{38-39} HOTAIR deletion actually results in a coordinated loss of PRC2 and LSD1 binding at hundreds of genes, and these genes also show a corresponding increase in their gene expression.\textsuperscript{40} Further studies show that HOTAIR can reprogram chromatin states during tumor metastasis.\textsuperscript{41-43} For example, in breast carcinomas, HOTAIR is upregulated with metastasis. Overexpression of HOTAIR in breast epithelial cells leads to the occupancy of PRC2 to over 800 additional genes, and the pattern of the PRC2 occupancy changes from that of the typical breast epithelial cells to that of embryonic fibroblasts.\textsuperscript{41-43}

The identification of HOTAIR in combination with the development of RNA detection technologies such as microarray and second generation of nucleic acid sequencing has provided a surge in the search for lncRNAs and their functions and functional mechanisms in eukaryotic cell. A plethora of lncRNAs are now thought to utilize molecular mechanisms similar to HOTAIR, to recruits key epigenetic factors for either \textit{cis} or \textit{trans} epigenetic regulation of chromatin status in the nucleus.\textsuperscript{44-47}

**NEAT1 and MALAT1**

In a search for nuclear enriched abundant transcript (NEAT) RNAs, two lncRNAs NEAT1 and NEAT2 (already identified as MALAT-1 in previous studies) are discovered.\textsuperscript{48} NEAT1 is widely
expressed in all sorts of mammalian cells, and is transcribed by RNA polymerase II (pol II). NEAT1 exists in two forms: a ~3.7 knt isoform of NEAT1_1 with poly (A) tail and a ~23 knt isoform of NEAT1_2. Both isoforms of NEAT1 are localized in and also key components of the paraspeckles. Paraspeckles are unique subnuclear structures composed of specific proteins and RNAs. Some of these specific proteins belong to the Drosophila Behavior/Human Splicing (DBHS) family such as PSPC1, SFPQ (PSF) and NONO (p54nrb). NEAT1_2 has been shown to directly interact with the SFPQ and NONO to form a pivotal component for paraspeckle formation, and NEAT1_2 knockdown disturbs the paraspeckle formation. NEAT1_1 can bind to PSPC1, which is also localized in paraspeckles. NEAT1_1 overexpression can increase the number of paraspeckles, although NEAT1_1 deletion does not block the formation of paraspeckles. A to I edited mRNAs are retained within the nucleus inside paraspeckle in NEAT1-dependent manner. Hyper-edited RNAs with inverted repeat sequences in the 3' UTR are retained in paraspeckles of differentiated cells. Upon NEAT1 deletion, paraspeckles are not formed and the RNAs with inverted repeats are transported into cytoplasm. Recently, lines of evidence show that NEAT1 has important roles in the formation of corpus luteum and mammary gland as suggested by the phenotypes of NEAT1 knockout mice. However, critical roles of NEAT1 and paraspeckles in mammalian cells require further investigations.

Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), also known as NEAT2 is a
IncRNA about 8 knt, and is highly conserved in mammalian species.\textsuperscript{48,61-64} MALAT1 was initially identified as a prognostic parameter for lung cancer patients, and is also one of the first ncRNAs to be associated with a disease.\textsuperscript{61,64} High expression level of MALAT1 is detected in many human carcinomas, and level of MALAT1 in early-stage tumors is a prognostic marker for patient survival.\textsuperscript{65-69}

MALAT1 is transcribed by pol II. The region of about one hundred nucleotides at the 3’ end of unprocessed MALAT1 is the most evolutionarily conserved part of this transcript from fish to human, and this 3’ end has the cloverleaf secondary structure of tRNAs.\textsuperscript{63,71} Just like tRNAs, processing of the MALAT1 3’ end is executed by RNase P and RNase Z.\textsuperscript{63,71} NEAT1_2 is also processed like MALAT1 by cleavage with RNase P at its 3’ end by recognizing a tRNA-like structure, and the mature 3’ end of both IncRNAs are protected by triple helical structures.\textsuperscript{70,71}

MALAT1 is a component of nuclear speckles (also known as SC35 splicing domains), and it contains two regions responsible for protein interaction and RNA binding for its localization.\textsuperscript{48,62,73} Nuclear speckles are involved in mRNA processing, splicing, and export of mRNA, and MALAT1 is thought to serve as a scaffold for the pre-mRNA processing complex in nuclear speckles.\textsuperscript{48,72} siRNA meditated knockdown of the speckle protein RNPS1, SRm160, or IBP160 results in disrupted MALAT1 localization.\textsuperscript{72} MALAT1 localization in nuclear speckles is also thought to participate in active transcription.\textsuperscript{74} Blocking pol II transcription with α-amanitin or 5,6-dichloro-1-β-D-ribobenzimidazole results in MALAT1 diffusing in the nucleus.\textsuperscript{75} MALAT1
is also reported to interact with pre-mRNA splicing factors such as SR proteins, although deleting MALAT1 does not incur splicing change in lung cancer cells.

MALAT1 binds to unmethylated polycomb2 (Pc2) to regulate the nuclear localization of Pc2 based on several researches. Pc2 is a component of the PRC1, and its methylation/demethylation cycle controls the expression of multiple growth-signal responsive genes. Unmethylated Pc2 relocates from polycomb bodies to nuclear speckles by binding to MALAT1, and this binding further promotes the SUMOylation of a critical transcription factor E2F1, which activates a growth-control gene program. In neurons MALAT1 has also been shown to interact with the TAR DNA binding protein (Tdp-43) and to regulate synapse formation. TDP-43 is a nuclear protein that regulates transcription, alternative splicing, and RNA stability.

MALAT1 may have diverse functions in multiple cellular events. However, in instances of MALAT1 deletion, no significant effect on nuclear speckles is observed, and there is no distinct phenotype in the MALAT1 knockout mice. It seems that despite all lines of evidence indicating critical functions of MALAT1 in cell cultures, its actual relevance in animal physiology remains unknown.

**Some other nuclear IncRNAs**

The IncRNA H19 has been known for more than two decades. In recent years H19 has been identified as the primary transcript of miR-675. On the other hand, H19 may also possess a
cytoplasmic role as sponge of let-7 and miR-106a.\textsuperscript{92,93} It seems that H19 as one of the most abundant, conserved, and first identified lncRNAs in mammals remains intriguing for its actual physiological functions and functional mechanism. Another nuclear lncRNA Pnky is found in mammalian neural stem cells, and it interacts with PTBP1 to regulate the expression and alternative splicing of some genes.\textsuperscript{94}

Another class of RNAs are derived from the enhancers of an array of genes, and are known as enhancer RNAs (eRNAs).\textsuperscript{95} eRNAs are a large species of lncRNAs residing in the nucleus and maybe never apart from the chromatin. Enhancer DNA elements can regulate the transcription of genes \textit{in cis} from a distal intragenic or intergenic region.\textsuperscript{96-101} Transcription at enhancer can produce bi-directional and uni-directional eRNAs,\textsuperscript{102,103} and the main function as well as the reason of producing two forms of eRNAs is still unclear. eRNAs normally are recognized as non-polyadenylated single exon, although spliced and polyadenylated eRNAs have also been shown to exist.\textsuperscript{103-105} eRNAs are positively regulated by Androgen, estradiol, P53, MYOD, and MYOG in various conditions.\textsuperscript{106-108} Under the estradiol treatment, eRNAs from the enhancers of the estradiol induced genes are increased in MCF7 cells.\textsuperscript{102} Similarly, Nutlin-3 induces eRNA production from the enhancer of p53-dependent genes.\textsuperscript{107} The REV-ERB signaling can negatively regulate the generation of eRNAs through binding to the enhancer of target genes.\textsuperscript{109} The exact mechanism of eRNAs in regulating gene transcription is largely unknown. It has also been shown that eRNAs can stabilize the looping of the cohesion and/or Mediator complex
interacting with the enhancer of the target genes, although eRNA deletion does not always influence the loop formation.\textsuperscript{102,110}

Some housekeeping lncRNAs also reside in the nucleus, and these include centromeric satellite RNAs and telomere RNAs, and both actually possess regulatory roles along their canonical functions. Centromere is composed of satellite and retroelement repetitive DNA sequences in most eukaryotes.\textsuperscript{111-114} RNAs transcribed by pol II from the repetitive DNA elements are termed centromeric satellite RNAs. Centromeric satellite RNAs usually are in a length of 35-5000 nt,\textsuperscript{115-118} and they have a key role on centromere/kinetochore assembling and function in many organisms.\textsuperscript{119-121} The minor satellite RNA in mice even plays a role \textit{in trans} to regulate the activity of telomerase in embryonic stem cells.\textsuperscript{122} Telomere RNA is transcribed by pol II from telomeric ends of the chromosome. The length of telomere RNA varies from 100 to 9000 nt and contains various number of UUAGGG repeat sequences in different eukaryotic organisms, and thus it is known as telomeric repeat-contain RNA (TERRA).\textsuperscript{123} TERRA is a key component of telomere, which is a nucleoprotein structure that protects the termini of the chromosome from damage and degradation, and maintains the chromosome length, integrity, and stability.\textsuperscript{124-127}

\textbf{EICiRNAs}

In the past years, circular RNAs (circRNAs) have emerged as a large family of ncRNAs.\textsuperscript{128-132} In canonical or linear splicing, the 5’ end of the exon joins the 3’ end of another exon upstream (more 5’ in the linear precursor RNA). The circRNAs are generated from coding (sometimes also
noncoding) regions of genes by a process called back-splicing, in which the 5’ end of an exon joins with the 3’ end of itself or another exon downstream (more 3’ in the linear precursor RNA). Thousands of circRNAs have been identified in eukaryotic cells. Most circRNAs are composed exclusively of exonic sequences, without association with polyribosomes, and localized in the cytoplasm with cell specificity. Two of these circRNAs have been shown to function as microRNA sponge in the cytoplasm. Recently, we have identified a new subclass of circRNAs, which are circularized with introns ‘retained’ between the circularized exons, and for this reason we call them exon-intron circRNAs (EIciRNAs). The identification of circRNAs and EIciRNAs actually makes us rethink the meaning of “coding” sequences of the “coding” genes, as now both mRNAs and lncRNAs can be generated from the same exonic sequences of pre-mRNA. In contrast to the other circRNAs, EIciRNAs are localized exclusively in the nucleus. Further investigations reveal that EIciRNAs associate with pol II and enhance the initiation of transcription \textit{in cis} by promoting pol II binding at the core promoter of the EIciRNA parent genes. The association between EIciRNAs and pol II is mediated by U1 snRNP via the RNA-RNA binding of U1 snRNA and the EIciRNAs. The 5’ splicing site of the retained intron in EIciRNAs is the actual binding site of U1 snRNA. Thus, EIciRNAs are a new class of nuclear lncRNAs with functions in transcriptional regulation through RNA-RNA interaction. A series of unknowns are associated with EIciRNAs. We have shown in two examples that EIciRNAs serve as a positive feedback to promote transcription, and it remains to be examined
whether this is true for all EIciRNAs to their parent genes genome-wide. It is also possible that
EIciRNAs have a link with epigenetic chromatin status of active transcription. EIciRNAs has
colocalization with their parent genomic loci, although they are also distributed to multiple
nuclear regions. The close association of EIciRNAs with U1 snRNA indicates strongly that
EIciRNAs may be localized to nuclear speckles, where active transcription and splicing can be
coordinated. We speculate that besides their cis role as individuals, EIciRNAs as a whole could
possess a trans role synergistically by keeping factors such as U1 snRNP at the nuclear sites of
active gene expression.

**Perspective**

LncRNAs are recognized as important regulators in different aspects of cellular events. In this
review, we have summarized a number of lncRNAs that are retained and function in the nucleus
(Figure 1).

Localization of ncRNAs must be regulated to coordinate the whole process of biogenesis and the
function. The existence of some lncRNAs predominantly or exclusively in the nucleus requires
the involvement of localization mechanisms either active or passive. One would speculate that
the association of lncRNAs such as Xist, Hotair, and eRNAs with chromatin keeps them in the
nucleus. On the other hand the binding of nuclear proteins to lncRNAs such as MALAT1,
NEAT1, and EIciRNAs ensures the nuclear localization of these RNAs. Nuclear ncRNAs may
also lack sequences or features required for active nuclear exportation. Is it possible that some
lncRNAs may shuttle between the nucleus and cytoplasm? At least there is no such an example so far.

The abundance of each nuclear lncRNA has to be regulated to ensure proper functionality and localization. Theoretically, \textit{cis} effect requires only lower abundance, while higher abundance for \textit{trans} effect. The relative abundance is supposedly decided by the production rate (transcription and biogenesis) and the process of degradation. For most nuclear lncRNAs, we still know little about the dynamic regulation in their transcription, biogenesis, and degradation.

Nuclear lncRNAs play quite diverse roles with multiple molecular mechanisms. A substantial number of nuclear lncRNAs are conserved in mammals, indicating further functional significance. The physiological roles of most nuclear lncRNAs are still elusive, even for those with known functional mechanisms, such as NEAT1, MALAT1, and EIciRNAs. One of the reasons for missing physiological relevance may be because of the fact that research about lncRNAs in model organisms such as \textit{C. elegans}, \textit{Drosophila}, zebra fish, and mice are still limited. The many unknowns about nuclear lncRNAs urge substantially more investigations in this field.

ACKNOWLEDGEMENTS

We thank Dr. Liang Chen and other members of the Ge Shan lab for discussions. This work was supported by the National Basic Research Program of China (2015CB943000), National Natural Science Foundation of China (91519333 and 31471225), and the Fundamental Research Funds
for the Central Universities (WK2070000034).
References


Figure 1. Functional mechanisms of some nuclear lncRNAs. (A) MALAT1 localizes in the nuclear speckles and interacts with splicing factors to regulate splicing. (B) Hotair interacts with critical epigenetic regulators to regulate the chromatin status. (C) Xist RNA silences one X chromosome in female via epigenetic regulation. (D) ElciRNAs enhance the initiation of pol II transcription of their parent genes by interacting with U1 snRNP. (E) eRNAs promote transcription in cis by enhancing chromatin looping. (F) Centromeric satellite RNAs play roles in centromere/kinetochore formation and function, while TERRA plays roles in the telomeres. (G) NEAT1 localizes and functions in paraspeckles.