Long non-coding RNA mediated mechanisms independent of the RNAi pathway in animals and plants

Phil Chi Khang Au *, Qian-Hao Zhu, Elizabeth Dennis and Ming-Bo Wang
Commonwealth Scientific and Industrial Research Organisation Plant Industry,
Canberra, Australian Capital Territory 2601, Australia

Keywords: Long non-coding RNA, chromatin modifications, transcriptional gene regulation, translational gene regulation, transcription factor regulation

*Author for correspondence:
Dr. Phil Au
CSIRO Plant Industry
Black Mountain Laboratories
GPO Box 1600 Canberra ACT 2601
Australia.

Telephone: +61 2 6246 4846
Facsimile: +61 2 6246 5000
Email: Phil.Au@csiro.au

Running title: Long non-coding RNA mediated mechanisms in animals and plants
Abbreviations

**CCND1**  Cyclin D1
**CGR**  CR20-GUT15-Related
**COOLAIR**  Cold induced Long Antisense Intragenic RNA
**DHFR**  dihydrofolate reductase
**DNMT3b**  DNA methyltransferase 3b
**FLC**  Flowering Locus C
**Gas5**  Growth arrest-specific 5
**GR**  Glucocorticoid Receptor
**GREs**  Glucocorticoid Response Elements
**H3K9**  histone 3 lysine 9
**H3K27**  histone 3 lysine 27
**H3K27me3**  histone 3 lysine 27 trimethylation
**hnRNP-K**  heterogeneous nuclear ribonucleoprotein K
**HOTAIR**  HOX Antisense Intergenic RNA
**HSF**  heat shock transcription factor
**HSR1**  Heat Shock RNA 1
**lncRNA**  long non-coding RNA
**LSD1**  Lysine-Specific Demethylase 1
**MEN**  Multiple Endocrine Neoplasia
**miRNA**  micro RNA
**MtRBP1**  Medicago truncatula RNA Binding Protein 1
**NAT**  Natural Antisense Transcripts
**NFAT**  nuclear factor of activated T cells
**ncRNA**  non-coding RNA
**NRON**  Non-coding Repressor Of NFAT
**PRC2**  Polycomb Repressive Complex 2
**SINE**  short interspersed elements
**siRNA**  small interfering RNA
**SRA**  steroid receptor RNA activator
**SRAP**  steroid receptor RNA activator protein
**SRC-1**  Steroid Receptor Coactivator 1
**SRG1**  SER3 regulatory gene 1
**TF**  transcription factor
**TLS**  Translocated in LipoSarcoma
Abstract

Recent advances in the field of RNA research have provided compelling evidence implicating the significance of long non-coding RNA molecules in many diverse and substantial biological roles that include transcriptional and post-transcriptional regulation of gene expression, genomic imprinting, modulation of protein activity and subcellular localization, and cellular structural maintenance. While long non-coding RNAs are most extensively studied in animal species, studies of long non-coding RNA in plants begin to emerge showing some conservation of mechanisms. This review aims to provide an overview of significant and recently identified long non-coding RNA-mediated mechanisms in both animal and plant species.
Introduction

The mechanisms mediated by non-coding (nc) RNA enhance the complexity of eukaryotic organisms. This is a plausible explanation to the surprising finding that different orders of eukaryotes share approximately the same number of protein coding genes but have vastly different phenotypic complexity. In contrast, the number of non-coding genes increases proportionally with increasing developmental complexity with 98% of the human transcriptome represented as ncRNAs. ncRNAs are not translated into proteins and were believed to represent transcriptional noise with the exception of those that play specific roles in transport, splicing and synthesis of proteins including ribosomal RNA and transfer RNA. Over the last decade, studies of ncRNA have shed light on a series of novel mechanisms reshaping our understanding of the RNA world and provided compelling evidence of cellular function. The nature, structure and function of ncRNAs can vary and are diverse. Expression can be tissue- and cell-specific, exhibiting subcellular localization and is regulated in a spatial and temporal manner. Most ncRNAs studied to date in eukaryotic species are associated with transcriptional and post-transcriptional regulation including imprinting, X-chromosome inactivation and targeted mRNA degradation. ncRNAs also function as molecular cargos to target protein subcellular localization, induce changes in protein conformation and are implicated in cellular structural maintenance. Mechanisms that induce gene silencing through the RNA interference pathways (small interfering RNA and micro-RNA directed silencing) have been extensively studied in both animals and plants and are well reviewed in many publications. In contrast, the study of long ncRNAs (IncRNAs) is still at its infancy and many IncRNAs remain uncharacterized. IncRNAs are generally considered as transcripts longer than 200 nucleotides (nt) that
can bear many signatures of mRNA including 5’capping, splicing and polyadenylation but have little or no ORFs. Here, we review the mechanisms predominantly mediated by lncRNAs (Table 1) that are unrelated to the RNA interference pathway in animals and plants; therefore ncRNAs involved in the siRNA and miRNA pathways including primary transcripts of miRNAs will not be discussed. A new class of lncRNA-mediated mechanism is identified, namely “Long non-coding RNA-mediated modulation of proteins involved in transcription” is discussed.

**lncRNA mechanisms in animals**

**lncRNA-mediated chromatin modifications**

Studies in animals and plants have demonstrated that chromatin modifications, such as methylation, acetylation and phosphorylation of histones, are crucial for tissue-specific gene expression and for genome reprogramming during development. A combinatorial of these histone modifications determines the accessibility of the DNA to binding factors; thus the active or inactive status of the embedded genes. These histone modification patterns are believed to be established by ubiquitously expressed chromatin modifying complexes; however, how these enzymatic complexes are recruited to distinct and specific sites under different cellular contexts has not been fully decoded. Recent investigations on X chromosome dosage compensation, imprinting and homeotic gene expression predominantly in mammalian species revealed the association of lncRNAs with these complexes and showed that lncRNAs may play a crucial role in organizing chromatin structure, and in establishing and maintaining the epigenetic landscape during these biological processes. Disruptions to these chromatin modifications could lead to dysregulation of developmental processes and to various diseases.
Several lncRNAs have been identified in imprinted gene clusters in mammals, and at least two of them, \( \text{Air} \) and \( \text{Kcnq1ot1} \), have been shown to play a critical role in establishing monoallelic expression patterns of imprinted genes.\(^{18,19} \) It seems that they achieve their role by binding to chromatin modifying complexes and guiding them to specific genomic locations for proper histone modifications required for silencing of maternal or paternal alleles.\(^{10,11} \) \( \text{Air} \) is transcribed by the promoter located within intron 2 of the \( \text{Igf2r} \) gene and is predominantly expressed from the paternal allele, resulting in silencing of paternal \( \text{Igf2r} \) and two other \textit{cis}-linked genes, \( \text{Slc22a2} \) and \( \text{Slc22a3} \), in mouse placenta.\(^{19} \) \( \text{Air} \) was found to interact with the \( \text{Slc22a3} \) promoter chromatin and the H3K9 histone methyltransferase G9a in placenta. \( \text{Air} \) accumulates at the \( \text{Slc22a3} \) promoter correlating with localized H3K9 methylation.\(^{10} \) Truncation of \( \text{Air} \) results in the loss of G9a accumulation at the \( \text{Slc22a3} \) promoter and biallelic transcription of \( \text{Slc22a3} \), suggesting that \( \text{Air} \) specifically recruits G9a to the promoter region of \( \text{Slc22a3} \) and leads to targeted H3K9 methylation and allelic silencing of paternal \( \text{Slc22a3} \).\(^{10} \) Similarly, \( \text{Kcnq1ot1} \) is expressed from the paternal allele and is responsible for silencing of \( \text{Kcnq1} \), which is antisense to \( \text{Kcnq1ot1} \), and a cluster of genes flanking the \( \text{Kcnq1} \) locus in placenta.\(^{11,18} \) \( \text{Kcnq1ot1} \) interacts with G9a and components (EZH2 and SUZ12) of the Polycomb Repressive Complex 2 (PRC2) responsible for H3K27 trimethylation (H3K27me3) in a lineage-specific manner. \( \text{Kcnq1ot1} \) recruits these chromatin modifying complexes to the \( \text{Kcnq1} \) domain to establish repressive chromatin marks.\(^{11,20} \) \( \text{Air} \) co-immunoprecipitates with G9a while \( \text{Kcnq1ot1} \) can be pulled-down by antibodies against G9a, EZH2 or SUZ12 protein, suggesting that these lncRNAs associate with these chromatin modifiers.\(^{10,11} \) However, it is unclear whether the association is direct or recruited through association of lncRNAs with other unidentified protein factors.
Interaction between lncRNAs and chromatin modifying complexes is also essential for X-chromosome inactivation, a mechanism by which one of the two X chromosomes is epigentically silenced in female mammalian embryos to ensure that females have the same dosage of X-linked genes as the males.\(^{21}\) At least seven lncRNAs, including \textit{Xist} and \textit{RepA}, participate in this process.\(^{22}\) \textit{Xist} is transcribed only from the future inactive X chromosome (Xi) and is responsible for binding PRC2, and for spreading PRC2 and its H3K27 trimethylase activity throughout the Xi to finally transcriptionally silence the entire Xi.\(^{22-25}\) Paradoxically, transcriptional induction of \textit{Xist} is due to \textit{RepA}-mediated recruitment of PRC2 and H3K27me3 of the \textit{Xist} promoter.\(^{15, 26}\) \textit{RepA} is transcribed from the conserved 5‘ end of \textit{Xist} and is in the same orientation as \textit{Xist}. It consists of 7.5 tandem repeats of a 28-nucleotide (nt) sequence that folds into two conserved stem-loop structures.\(^{27}\) RNA gel shift analysis showed that \textit{RepA} RNA directly interacts with EZH2 of PRC2 through its 28-nt stem-loop structures.\(^{15}\) A more recent investigation confirmed the interaction between \textit{RepA} and PRC2 and found that binding also occurs with the SUZ12 component of PRC2; this association is more efficient in the presence of the full-length \textit{RepA} that can form the two long stem-loop structures.\(^{28}\) Although more studies are required to determine the exact secondary structure of \textit{RepA} that interacts with PRC2, these studies together with the result of another recent study by Kanhere et al.\(^{29}\) clearly show that secondary structures of ncRNAs might play an important role in binding and/or recruitment of chromatin modifying complexes, such as PRC2, to specific genomic locations for deposition of H3K27me3 and transcriptional gene silencing.

In addition to \textit{in cis} action, lncRNA-mediated chromatin modifications can also occur \textit{in trans}. In mammals, \textit{HOX} transcription factors are clustered on four chromosomal loci, \textit{HOXA} – \textit{HOXD}, and are essential for specifying the positional
identities of cells. Maintenance of HOX expression patterns is under complex
epigenetic regulation, including IncRNA-mediated chromatin remodelling. Hundreds
of IncRNAs have been identified in the human HOX loci including HOTAIR (HOX
Antisense Intergenic RNA), a 2.2-kb IncRNA generated from the HOXC locus and
functions to repress in trans transcription across 40-kb of the HOXD locus, located at
a different chromosome. HOTAIR achieves its role by recruiting and guiding PRC2,
through interacting with SUZ12 and EZH2, to the HOXD locus to establish the
H3K27me3 silencing marks. Using a series of HOTAIR deletion mutants, a 5’
domain of ~ 300-bp of the HOTAIR transcript was found to be able to retain the PRC2
binding activity of the full-length HOTAIR. This study also showed that a 3’ domain
of HOTAIR interacts directly with Lysine-Specific Demethylase 1 (LSD1)-containing
CoREST repressor complexes, which are found at the genomic regions flanking
HOXD. LSD1 mediates enzymatic demethylation of H3K4me2 and is required for
proper repression of HOX genes in Drosophila. These results indicate that
HOTAIR serves as a scaffold for these two distinct histone modification complexes,
which enable HOTAIR-mediated assembly of PRC2 and LSD1, and coordinates
targeting of PRC2 and LSD1 to chromatin for coupled H3K27 trimethylation and
H3K4me2 demethylation. In addition, computational analysis and RNA footprinting
showed that the PRC2 and LSD1 binding domains of HOTAIR possess distinct
secondary structures, supporting the notion that secondary structure of IncRNAs
might be an important determinant in recruitment of chromatin modifying complexes.
Importantly, results of this recent study not only provided evidence for direct
interaction between IncRNA and enzymatic complexes but suggest that IncRNAs may
serve as modular scaffold to assemble multiple chromatin modifying complexes and
to specify complex patterns of chromatin states at specific genes.
A growing body of evidence supports the notion that lncRNAs are key components in the regulation of chromatin states and epigenetic inheritance. First, lncRNAs are pervasively transcribed in the genome as shown by whole genome tiling array and RNA-seq studies. Second, lncRNAs have been found to directly or indirectly interact with a growing number of enzymatic complexes involved in chromatin modifications, such as PRC1, PRC2, CoREST, SMCX and G9a. Third, ~ 20% of approximately 3300 lncRNAs identified from various human cell types and ~ 13% of lncRNAs expressed in HeLa cells were found to be associated with PRC2 and CoREST, respectively. The finding that approximately 40% of the lncRNAs associated with CoREST are also associated with PRC2 suggests that a number of lncRNAs can serve as scaffold to assemble two or more chromatin modifying complexes to regulate expression of the same target genes. It is clear that lncRNAs play an essential role in mediating epigenetic changes by recruiting chromatin remodelling complexes to specific genomic loci and several possible mechanisms by which lncRNAs tether or guide chromatin modifying complexes to their specific destinations have been proposed. More studies are required to uncover the exact mechanism controlling the interaction between lncRNAs and chromatin modifying complexes.

**lncRNA-mediated transcriptional and translational regulation**

While lncRNA-mediated chromatin modifications are potent mechanisms in the regulation of gene transcription, other mechanisms of gene regulation mediated by lncRNA have also been described. In *Saccharomyces cerevisiae*, the transcription of lncRNA *SRG1* (*SER3 regulatory gene 1*) in the regulatory region of the *SER3* gene represses *SER3* transcription during growth in rich medium. This mechanism of
gene silencing is termed “transcriptional interference” which may occur due to interference between transcriptional machineries processing along adjacent sequences, thus the active transcription of SRG1 prevents the association of transcriptional initiation elements to adjacent sequences of close proximity.

Following this study, a similar mechanism of gene regulation has been described. Martianov et al. showed that in serum-starved U2OS cells, a lncRNA transcribed from the minor (upstream) promoter of the DHFR (dihydrofolate reductase) gene binds to TFIIB (Transcription Factor IIB) to prevent its association with the major (downstream) promoter of the DHFR gene; thus resulting in gene repression. Moreover, the lncRNA can also form a triplex with the major promoter of the DHFR gene, further preventing association of TFIIB to this promoter region.

More recently, Schmitz et al. showed that the formation of a DNA:RNA triplex in the presence of a lncRNA that is complementary to the rDNA (ribosomal DNA) promoter facilitates the recruitment of DNMT3b (DNA methyltransferase 3b) to catalyse DNA methylation at that particular locus in an unknown mechanism in mice. Such a mechanism for controlling promoter usage could be widespread as many triplex structures exist in eukaryotic chromosomes. However, it remains to be determined whether lncRNA-mediated formation of triplex structures in promoters is a common mechanism and whether this mechanism is conserved in plants. Notably, it is also unknown whether sequestration of TFIIB affects the expression of other unrelated genes that require TFIIB for transcription.

A long intergenic ncRNA named licnRNA-p21 is induced by p53 upon DNA damage in human cells to mediate global gene repression via association with heterogeneous nuclear ribonucleoprotein K (hnRNP-K), also a component of the p53 pathway. p53 is a tumor suppressor known to be involved in the maintenance of
genomic integrity and functions to trigger a transcriptional response leading to cell arrest and/or apoptosis. Association of lincRNA-p21 with hnRNP-K modulates hnRNP-K localization to the promoter region of target genes to confer transcriptional repression.

While the mechanisms of intergenic lncRNA transcription discussed so far appear to negatively regulate gene expression, there is also evidence that lncRNAs function can confer gene activation consequences. In mice, the lncRNA Evf-2 functions as a co-factor to modulate transcription factor activity. This lncRNA is transcribed from an ultra-conserved enhancer and recruits transcription factor Dlx-2 to the same enhancer to induce expression of adjacent protein-coding genes. Consistent with this, a recent study which used GENCODE annotation of the human genome to identify novel lncRNA also suggests that the expression of lncRNA can function as enhancers to mediate transcription of neighbouring genes. The study however excluded the analysis of lncRNAs derived from intergenic or overlapping protein-coding gene regions indicating that such mechanism of gene enhancement could be more common. Although the exact mechanism remains to be elucidated, involvement of transcription factor enhancement by lncRNA similar to the role of lncRNA Evf-2 is a possibility.

The Gas5 (Growth arrest-specific 5) gene encodes a lncRNA that suppresses the expression of a subset of genes during nutrient starvation in HeLa cells. Gas5 RNA binds to the DNA-binding domain of GR (Glucocorticoid Receptor) transcription factor, preventing its association with the GREs (Glucocorticoid Response Elements) located in the regulatory regions of glucocorticoid-responsive genes. This modulation of GR transcriptional activity appears to influence cell survival and metabolic activities during starvation.
Natural antisense transcripts (NATs) have been demonstrated in numerous studies to possess the ability to hybridize to overlapping genes and generate endo-siRNA in *Drosophila* and mice.\(^{49,50}\) In addition, some NATs regulate the translation of target mRNAs in rats and humans.\(^{51-54}\) The most recent demonstration of this mechanism by Beltran et al.\(^{51}\) showed that the binding of a NAT complementary to the 5’ splice site of *Zeb2/Sip1* mRNAs, prevents spliceosomal removal of the internal ribosome entry site sequence situated within an intron at the 5’ untranslated region; thus allowing translation of this mRNA necessary for epithelial-mesenchymal transition in humans.

In some cases, the act of lncRNA transcription itself is sufficient to cause gene regulation/chromatin modification rendering the lncRNA as mere non-functional by-products of transcription. Examples of this phenomenon have been demonstrated in yeast and human leading to silencing or activation of neighbouring genes via chromatin remodelling. The step-wise transcription of several lncRNAs from the upstream promoter region of the *fbp1* locus in the fission yeast (*Schizosaccharomyces pombe*) is required to initiate the opening of the chromatin in a progressive manner towards the mRNA transcription start site; \(^{55}\) thus allowing transcription of the downstream gene. Similarly, the transcription of lncRNAs at the *ade6-M26* locus induces changes in the chromatin during meiosis enabling meiotic recombination in human.\(^{56}\) In response to phosphate starvation, antisense transcription of a lncRNA from the 3’ end of the *PHO5* gene in yeast induces the expression of the protein coding gene, *PHO5*.\(^{57}\) The authors concluded that because *PHO5* RNA is highly unstable and is rapidly degraded by exosomes, the non-coding transcript is unlikely to possess a functional role but instead proposed the idea that the kinetics of *PHO5* activation is affected by intergenic transcription of the ncRNA which in turn leads to
nucleosome eviction allowing a 600 bp region to become fully accessible for RNA Pol II transcription.$^{57}$

**IncRNAs function as molecular cargos to target protein subcellular localization**

The activity of many proteins required for cell cycle progression and gene transcription can be modulated by limiting their subcellular localization. This mechanism of control has been demonstrated and can be regulated by IncRNAs. In the fission yeast, Mei2p, an RNA-binding protein required for pre-meiotic DNA synthesis and meiosis I is transported from the cytoplasm to the nucleus via association with its RNA intermediate, the MeiRNA.$^{58}$ Mutated Mei2p with lower ability to bind to MeiRNA remained cytoplasmic. In the absence of MeiRNA, Mei2p transgene product containing an added nuclear localization signal is able to translocate to the nucleus and promote meiosis I, suggesting that the role of MeiRNA is to act a chaperon to guide Mei2p nuclear import.$^{59}$ The authors suggest that this mechanism of subcellular localization provides an explanation for the puzzling finding that ongoing transcription is required for accumulation of certain proteins in the nucleus.$^{59}$ Indeed, such an explanation is logical but may not be applicable to germ cells that display little or no transcriptional activity such as those in mammalian species.$^{60-62}$

Conversely, cytoplasmic localization of proteins regulated by IncRNA has also been observed. Transcription of 5S ribosomal RNA relies on TFIIA (Transcription Factor A). Following transcription, the 5S ribosomal RNA binds to TFIIA and is escorted to the cytoplasm. The binding of 5S ribosomal RNA masks the nuclear localization signal on the TFIIA protein, resulting in cytoplasmic retention during oocyte development in the *Xenopus*.$^{63}$
The subcellular localization of a transcription factor, NFAT (nuclear factor of activated T cells), important for T cell-mediated immune response, is regulated by the IncRNA NRON (Non-coding Repressor Of NFAT). NRON, expressed in a number of mouse and human tissues, binds to members of the nucleocytoplasmic trafficking machinery by inhibiting NFAT nuclear accumulation. This specifically represses NFAT activity and prevents the expression of genes mediated by NFAT.

**IncRNAs as molecular chaperons to confer protein conformational activity**

The folding structure of proteins deduced by post-translational modifications including phosphorylation can affect their active or inactive states. The discovery that IncRNAs are also capable of changing the conformational activity of many critical protein factors has added another level of complexity to our understanding of protein regulation. Wang et al. showed that under DNA damage signalling conditions, ncRNAs transcribed from the 5' regulatory regions of the CCND1 (Cyclin D1) gene in human cell lines function to allosterically modify the structure of an RNA-binding protein named TLS (Translocated in LipoSarcoma) by releasing it from its inactive conformation. TLS is not only modified by the ncRNA but is also guided to the promoter region of the CCND1 gene to inhibit the histone acetyltransferase activities of CREB-binding protein and p300, resulting in repression of CCND1 expression. This is consistent with the role of CDND1 as a cell cycle regulator known to be repressed by DNA damage signals.

Similarly in human cells, Lanz et al. showed that a IncRNA known as steroid receptor RNA activator (SRA) is required to confer functional specificity of a ribonucleoprotein complex known as SRC-1 (Steroid Receptor Coactivator 1), a nuclear receptor coactivator. However, unlike the previous study which resulted in
gene repression, SRA association with the SRC-1 results in the activation of hormone related nuclear receptors which then functions to direct the assembly and stabilization of a preinitiation complex for transcriptional activation at the promoter of targeted genes associated with hormonal changes.\textsuperscript{69} This mechanism appears to be tissue-specific as SRA is only expressed in several tissues, particularly in the brain.\textsuperscript{67} Expression analysis in muscle cells and RNA interference showed that SRA is a coactivator of MyoD transcription factor during skeletal muscle differentiation.\textsuperscript{70} A coding SRA has also been described; this RNA differs from the non-coding SRA by an extended exon-1 containing methionine codons necessary for translation.\textsuperscript{71} The SRA protein (SRAP) also functions as a coactivator of hormone related nuclear receptors and many other transcription factors including transcription factor IIB.\textsuperscript{72-74} The role of SRA and SRAP in the activation of nuclear receptors has been implicated in prostate cancer.\textsuperscript{73, 74} It is unknown whether SRA can associate and regulate its protein form.

IncRNAs also have the ability to affect the transcriptional machinery as a whole by binding to RNA polymerase II (RNA Pol II), causing global gene repression. The Alu RNA and B2 RNA are transcribed from short interspersed elements (SINE) during heat shock in human and mouse cells, respectively.\textsuperscript{75-77} These RNA molecules, although not evolutionarily related and share no sequence homology, are both able to bind to RNA Pol II causing general repression of transcriptional activity, suggesting that ncRNAs with diverse sequences can possess conserved functions. Therefore, information from the primary sequence of these ncRNAs is insufficient to allow prediction of function. Interestingly, scAlu and B1 RNA, which are the short form of Alu and a homologue of B2, respectively, are able to bind to RNA Pol II but cannot induce transcriptional repression.\textsuperscript{77} This suggests that neither sequence specificity nor the binding to the RNA Pol II itself is sufficient to inhibit
gene transcription. The repressive component (the regulatory domain) lies in two separate regions in the Alu RNA which are different to the regions required for binding (the binding domain) to RNA Pol II. The authors suggest that it is the structural conformation of those two regulatory domains and not the sequence that confers transcriptional repression. SINE transcripts also increase during other cellular stresses and during viral infection, suggesting that they may also modulate transcription in a variety of other biological responses. Heat shock induced transcriptional repression appears to also exist in many other eukaryotic species including Drosophila and plants. Paradoxically, some genes including those that encode heat shock proteins are transcriptionally activated during heat shock, suggesting that an underlying mechanism must exist to overcome SINE RNA mediated gene repression at those specific gene loci. Shamovsky et al. demonstrated that the activation of heat shock proteins in mammalian cells rely on the trimerization of a heat shock transcription factor (HSF) with a ncRNA named HSR1 (Heat Shock RNA 1) and translation elongation factor eEF1A in a ribonucleoprotein complex. This association renders the transcription factor active in DNA-binding activity.

In C. elegans, a starvation induced IncRNA, Rnsc-1 (RNA non-coding, starvation up-regulated), affects the processing of siRNAs by inhibiting the activity of the RNase III catalytic enzyme, Dicer. This impaired activity is due to the branched structures present in its 300 nt double-stranded RNA structure that presumably allows its association with the RDE-4/Dicer complex but prevents Dicer cleavage; thus inhibiting Dicer processing of other double-stranded RNA to siRNAs necessary for target messenger RNA down-regulation.

**Structural IncRNAs**
Our discussion on ncRNA-mediated mechanisms so far implicated their function as riboregulators affecting the activity and localization of proteins as well as their direct or indirect role in gene regulation. We now discuss evidence in which lncRNAs function as structural RNAs. Many mRNAs, in addition to encoding proteins, are known to be involved in the cytoplasmic localization of RNA structures in *Xenopus* oocytes, important for the determination of cell fate during cleavage development. However, many more examples in which RNAs function in establishing structural cellular integrity begin to emerge, particularly in those involving lncRNAs.

Studies of cytoplasmic RNA function in oocytes showed that the lncRNA *Xlsirts* along with *VegT* mRNA are responsible for the organization of the vegetal cortex of *Xenopus* oocytes. Deletion of either transcript results in disruption of the cytokeratin cytoskeleton, essential for the proper formation of germinal granules and the subsequent development of the germline.

In *Drosophila* oocytes, the 3’ untranslated region of the *oskar* mRNA possesses a function independent of its protein, osk, by binding to and assembling cytoplasmic complexes essential for oocyte development. *Oskar* mutant oocytes arrest early during oogenesis; this arrest can be rescued by the expression of the 3’ untranslated region of *oskar* RNA alone.

In mice, three independent studies have found that two lncRNAs, *MEN epsilon* (also known as *NEAT 1*) and *MEN beta*, transcribed from the MEN (Multiple Endocrine Neoplasia) I locus associate with protein components of nuclear body paraspeckles and are involved in conferring structural integrity of these nuclear organelles. Knock-down of these RNAs disrupts paraspeckle formation, while overexpression of *MEN epsilon* induces an increase in paraspeckle number.
Paraspeckles have been suggested to be important for nuclear storage of RNA and previous studies showed that RNase treatment causes their disruption, further confirming the involvement of RNA in their formation and maintenance. A recent study showed that MEN epsilon interacts with paraspeckle proteins through distinct regions in the 5’ and 3’ ends of the transcripts.

**IncRNA mechanisms in plants**

The plant genome is enriched with highly transcribed pseudogenes, ncRNAs and transposable elements including SINE, suggesting that ncRNA mechanisms are conserved in plants. However, research on IncRNAs in plants is still in its infancy. To date, apart from the identification of IncRNAs that represent the primary transcript precursors for miRNAs, only few IncRNAs have been isolated in several plant species and are generally uncharacterized.

The mechanism of chromatin modification induced by IncRNAs-mediated recruitment of the PRC2 complex has recently been explored in plants. In Arabidopsis, there is more than one PRC2 complex formed by combinations of different EZH2 and SUZ12 homologues with other subunits of PRC2. During prolonged cold periods in Arabidopsis, the FLC (Flowering Locus C) locus controlling flowering time undergoes epigenetic silencing, resulting in repression of FLC expression and early flowering. A recent study demonstrated the expression of antisense transcripts named COOLAIR (Cold induced Long Antisense Intragenic RNA) derived from the FLC locus upon exposure to cold. The authors suggest that COOLAIR may function to recruit PRC2 to induce histone modifications at the FLC locus, reducing the level of FLC expression. However, evidence is lacking to support a direct role of COOLAIR in FLC repression.
Genome-wide analysis of the *Arabidopsis* genome has identified the expression of overlapping NATs corresponding to a significant proportion of *Arabidopsis* transcriptome.\textsuperscript{105,106} Although the role of NATs in translational induction by association with the sense mRNA as demonstrated in animals has yet to be described in plants. Studies of NATs in *Arabidopsis* have demonstrated other means of gene regulation in which the formation of double stranded RNA with the complementary sense RNA recruits them into the siRNA pathway.\textsuperscript{107}

Also in *Arabidopsis*, the induced expression of ncRNAs of the *IPS1/At4* family during phosphate starvation responses results in the accumulation of the *PHO2* mRNA, a target of miR399 (micro-RNA 399). Franco-Zorrilla et al.\textsuperscript{108} showed that a conserved motif of 23 nt in this ncRNA family is complementary to miR399 but has critical mismatches at positions 10-11, required for miRNA guided cleavage. Therefore, *IPS1/At4* RNAs are not cleaved by miR399 but instead sequester miR399 to inhibit its effect on *PHO2* mRNA, in a mechanism known as target mimicry. *PHO2* RNA encodes an E2 ubiquitin conjugase-related protein that negatively affects shoot phosphate content and remobilization in an unknown mechanism.\textsuperscript{109}

In *Cucumis sativus*, a lncRNA named *CsM10* was isolated using differential display reverse transcription PCR, which showed differential expression patterns in different tissues, seedling developmental stages and photoperiods.\textsuperscript{110} *CsM10* harbours a 179 bp sequence with high sequence homology to a family of abiotic stress-associated ncRNAs known as the CR20-GUT15-Related (CGR) family, suggesting a role in the regulation of gene expression. More studies are required to elucidate its exact function.

In maize, a putative lncRNA, *Zm401*, is expressed specifically in pollens. Forward and reverse genetic studies have shown a function for *Zm401* in regulating
the expression of critical genes necessary for pollen development including $ZmMADS2$, $MZm3-3$ and $ZmC5$.\textsuperscript{111, 112} $MZm3-3$ was up-regulated in $Zm401$ mutants while $ZmMADS2$ and $ZmC5$ were both down-regulated.\textsuperscript{112} Over-expression of $Zm401$ severely affects pollen development due to abnormal tassels and degenerate anthers.\textsuperscript{111} How this lncRNA can mediate the down-regulation of certain genes but the up-regulation of others is intriguing and remains to be elucidated. One possible mechanism may involve the association of specific domains of the transcript with different members of transcriptional protein complexes.

IncRNA-mediated subcellular localization of proteins has also been described in plants. The IncRNA $Enod40$ directs the re-localization of MtRBP1 ($Medicago truncatula$ RNA Binding Protein 1) from the nucleus to cytoplasmic granules during specific stages of legume ($Medicago truncatula$) root nodule organogenesis.\textsuperscript{113}

**Perspectives**

Eukaryotic organisms are equipped with the intrinsic abilities to cope with a large range of environmental changes/stresses by producing steroid hormones to alter cellular activities correspondingly.\textsuperscript{114} Hormones play diverse physiological roles in the regulation of basal- and stress-related homeostasis and generally shift the body’s metabolic activities to catabolism, activate an immune response, and alter the expression level of cell survival and apoptosis-associated genes in response to external and internal stress.\textsuperscript{115-117} In animals, much evidence provided here shows the induction of IncRNA during stress conditions, suggesting that IncRNA expression is generally related to the organisms’ response to environmental changes (Table 2). Similarly, the production of non-coding siRNAs derived from NAT expression in plants is often induced by abiotic (salt) and biotic (bacterial) stress.\textsuperscript{107, 118} Plants have
also evolved to utilize their RNAi machinery as a defence mechanism against viral infections.\textsuperscript{119, 120} While mammalian organisms have exceeded this complexity and evolved an immune system to target foreign pathogens, it appears that they still retain the mechanism that utilizes ncRNA in response to stress conditions.

A large proportion of functions mediated by these IncRNAs appear to involve the regulation of proteins involved in transcription, particularly transcription factors, whether it is the mRNA form (e.g. Zeb2 mRNA and NAT association) or the protein form, including those that play critical roles in the maintenance of hormonal balance to induce cellular survival such as nuclear receptor transcription factors. This class of gene regulation mechanism has not been classified in the current literature; we therefore attempt to define it as “Long non-coding RNA modification of proteins involved in transcription” (Table 2, Fig. 1). Recent studies have provided evidence for miRNA-mediated regulation of transcription factors at the post-transcriptional level particularly during cell differentiation in animals and plants.\textsuperscript{121-123} However, the regulation of transcription factors at the post-translational level, which is a prevalent IncRNA mechanism in animals, may represent a more direct and efficient mechanism of regulating target gene expression than through the miRNA and siRNA pathways which require multiple processing and association events. Therefore, it is surprising that IncRNA-mediated mechanisms described in plants, although at a much lower extent than in animal species, have not yet shown evidence of interplay between IncRNAs and transcription factors to regulate gene expression.

Nevertheless, some mechanisms such as IncRNA-mediated chromatin modification and protein subcellular localization appear conserved in plants, although these plant IncRNAs are not orthologues of those that play similar roles in animals. Therefore, it remains to be determined whether animal and plant IncRNA orthologues
play homologous roles in similar mechanisms. Transcriptional interference in which
the transcription of lncRNA upstream of the coding region interferes with the
transcription of the coding gene could also be a widespread phenomenon. Ponting et
al.\(^\text{39}\) suggested that because promoter sequences of lncRNAs are more conserved than
the sequences of transcripts,\(^\text{6}, 124\) the act of transcription itself can possess a greater
and more widespread biological role than the transcript sequence. We have discussed
earlier that the formation of a DNA:RNA triplex in promoters could possibly be a
widespread mechanism conserved in mammals but this still remains to be examined in
plant species.\(^\text{42}\)

In addition to the discovery of novel mechanisms for newly identified
lncRNAs, recent studies in mammals have identified regulatory mechanisms mediated
by known ncRNAs (60 – 300 nt) that were previously thought to function mainly as
housekeeping RNA. For instance, it is well established that small nuclear RNAs
(snRNAs) play fundamental roles in regulating transcription by RNA Pol II and in
processing of the transcripts.\(^\text{125}\) In 2002, Kwek et al.\(^\text{126}\) showed that the association of
\(U1\) snRNA with TFIIH (Transcription Factor II H) positively regulates the rate of
transcription initiation by RNA Pol II, a new role for \(U1\) snRNA that is in addition to
its known role in RNA processing. On an interesting note, recent studies in animals
have demonstrated that some small nucleolar RNAs (snoRNAs) and transfer RNA are
potentially precursors of miRNA incorporated into the RNA interference pathway.\(^\text{127}\).

\(^\text{135}\) In \textit{Arabidopsis}, the association of snoRNA with Argonaute 7,\(^\text{134}\) a key component
of the RNAi machinery, suggests that similar mechanisms may also be conserved in
plants. These studies add convincing evidence to support the notion that ncRNAs have
muti-functional, structural and regulatory abilities.
It is generally considered that the ultimate goal of ncRNAs is to fine tune the expression of genes. As lncRNAs appear to regulate many transcription factors, we are beginning to see their impact in relation to cancers where a slight shift in the expression level of oncogenic genes can affect cellular proliferation. In fact, miRNAs, subunits of PRC, Gas5 and CCND1 are some of many factors implicated in tumor progression in mammalian species.

Studies of rncs, Alu, B1 and PRC2-associated lncRNAs provide compelling evidence that the function of lncRNA imposed on proteins is not due to sequence-specificity but the structure of the RNA complexes. Further research on RNA structure and function will allow the development of computational programs to predict RNA function based on structures, opening up a new platform of computational biology, similar to that developed for protein function prediction based on amino acid sequence motifs.

As we increase our knowledge of ncRNA-mediated mechanisms that enhance the complexity of eukaryotic organisms, we are one step closer to understanding our unique origin. Implementation of this knowledge will undoubtedly improve treatments for various stress induced conditions and diseases. New strategies in molecular and computational approaches are being developed to tackle the difficulty in studying ncRNAs. We believe that a shift in focus, studying lncRNAs in plants can reveal many more surprises as plants can provide a better and more efficient model for the functional study of ncRNAs. The ease of maintenance and short life cycles combined with the great diversity of environmental conditions under which plants can grow make them ideal models for ncRNA research that will certainly improve our understanding of the RNA world.
References


111. Dai XY, Yu JJ, Ma JX, Ao GM, Zhao Q. Overexpression of Zm401, an mRNA-like RNA, has distinct effects on pollen development in maize. Plant Growth Regulation 2007; 52:229-39.


