Long non-coding RNA and Polycomb: an intricate partnership in cancer biology

Cyrinne Achour¹,², Francesca Aguilo¹,²

¹Wallenberg Centre for Molecular Medicine (WCMM), Umea University, SE-901 85 Umea, Sweden,
²Department of Medical Biosciences, Umea University, SE-901 85 Umea, Sweden

TABLE OF CONTENTS

1. Abstract
2. Long non-coding RNAs
3. Identification of lncRNAs-protein complexes
4. Classification of long non-coding RNAs
   4.1. LincRNAs
   4.2. Intronic IncRNAs
   4.3. UTR-associated IncRNAs
   4.4. Sense IncRNAs
   4.5. NATs
   4.6. eRNAs
   4.7. paRNAs
   4.8. PROMPTs
5. Functions of lncRNAs
   5.1. Signal
   5.2. Decoy
   5.3. Guide
   5.4. Scaffold
   5.5. miRNA sponge
   5.6. miRNA precursor
6. Post-transcriptional modifications of lncRNAs
7. Polycomb Group of Proteins
8. Polycomb recruitment by lncRNAs
9. XIST
   9.1. XIST and polycomb
   9.2. Role of XIST in tumorigenesis
10. ANRIL
    10.1. ANRIL and polycomb
    10.2. Role of ANRIL in tumorigenesis
11. MALAT1
    11.1. MALAT1 and polycomb
    11.2. Role of MALAT1 in tumorigenesis
12. HOTAIR
    12.1. HOTAIR and polycomb
    12.2. Role of HOTAIR in tumorigenesis
13. Concluding remarks
14. Acknowledgment
15. References
Crosstalk between IncRNA and Polycomb in tumorigenesis

1. ABSTRACT

High-throughput analyses have revealed that the vast majority of the transcriptome does not code for proteins. These non-translated transcripts, when larger than 200 nucleotides, are termed long non-coding RNAs (lncRNAs), and play fundamental roles in diverse cellular processes. LncRNAs are subject to dynamic chemical modification, adding another layer of complexity to our understanding of the potential roles that lncRNAs play in health and disease. Many lncRNAs regulate transcriptional programs by influencing the epigenetic state through direct interactions with chromatin-modifying proteins. Among these proteins, Polycomb repressive complexes 1 and 2 (PRC1 and PRC2) have been shown to be recruited by lncRNAs to silence target genes. Aberrant expression, deficiency or mutation of both lncRNA and Polycomb have been associated with numerous human diseases, including cancer. In this review, we have highlighted recent findings regarding the concerted mechanism of action of Polycomb group proteins (PcG), acting together with some classically defined lncRNAs including X-inactive specific transcript (XIST), antisense non-coding RNA in the INK4 locus (ANRIL), metastasis associated lung adenocarcinoma transcript 1 (MALAT1), and HOX transcript antisense RNA (HOTAIR).

2. LONG NON-CODING RNAs

Transcription is not restricted to protein-coding genes but is pervasive throughout the genome, giving rise to numerous types of non-coding RNA (ncRNA). The first to be identified were infrastructural ncRNAs, such as transfer RNA (tRNA) (1), ribosomal RNA (rRNA) (2), small nuclear RNA (snRNA) (3) and small nucleolar RNA (snoRNA) (4). These ncRNAs play fundamental roles in different cellular processes such as translation (tRNA and rRNA), splicing (snRNA), and rRNA maturation (snoRNA). More recently, regulatory ncRNAs have been discovered as a new class of ncRNA with key roles in cellular homeostasis (5). Among them, an arbitrary size cut-off of 200 nucleotides discriminates two distinct groups: small non-coding RNA (scRNA) and long non-coding RNA (lncRNA). ScnRNAs, such as small interfering RNA (siRNA), microRNA (miRNA) and PIWI-interacting RNA (piRNA), have been thoroughly studied and reviewed (6), and are therefore not the focus of this review. LncRNAs are a large and diverse class of long transcripts (>200 nt) that do not encode proteins. They can be located either intergenically or intragenically, and can be transcribed in both sense and antisense directions. Similarly to messenger RNAs (mRNA), lncRNAs are transcribed by RNA polymerase II and contain an average of 2.29 exons (7), which are typically longer than those found in pre-mRNAs. LncRNAs are 5’ capped, undergo splicing and are generally polyadenylated. Despite the similarities shared with mRNAs, lncRNAs are expressed at lower levels and are not evolutionary conserved, which supports the hypothesis that lncRNAs may represent transcriptional noise. However, recent studies have shown that the expression of certain lncRNAs is both dynamic and cell-type-specific during development (8, 9), pointing towards their biological significance. The absence or presence of open reading frames (ORFs) in lncRNAs is an area of debate. Certain studies have found no evidence for protein coding ORFs in lncRNAs (10). However, ribosome profiling studies have provided that lncRNAs may encode short polypeptides (11) that are frequently loaded onto ribosomes (12).

LncRNAs can modulate transcription in cis, regulating target genes located at or adjacent to the locus from which they are transcribed. They also function in trans, regulating the expression of genes distanced from their host gene on different genomic loci or even in distal chromosomes. The process of being transcribed can be sufficient to enable lncRNAs to regulate the transcription of nearby genes. Indeed, lncRNA expression has been linked to several physiological and pathological pathways, including metabolism, development and cancer, to cite several examples. However, only a few lncRNAs have been well characterized to date, and thus, the mechanism of action of many lncRNAs remains poorly understood.

3. IDENTIFICATION OF lncRNAs-PROTEIN COMPLEXES

LncRNAs can regulate gene expression by interacting with transcription factors and/or chromatin-modifying complexes. Therefore, studying lncRNAs-protein complexes is essential in order to uncover the mechanisms and functions of lncRNAs in biological processes.

As an alternative to classical Ribonucleoprotein Immunoprecipitation (RIP) or RNA pulldown assays, Crosslinking and Immunoprecipitation (CLIP) followed by sequencing (CLIP-seq) have been used to identify new lncRNAs that interact with known RNA-binding proteins (RBPs) (13). CLIP relies on ultraviolet (UV) crosslinking of RNA to RBPs in cells. After partial RNA digestion, ribonucleoprotein complexes are immunoprecipitated and the associated RNAs can be analysed by qPCR or sequencing. Importantly, compared to more recent methods mentioned below, CLIP-seq does not provide full-length sequence information of transcripts, but does enable the mapping of RBP binding sites. CLIP-based methods are limited by: (i) efficiency of UV crosslinking for RNA and protein, (ii) identification of the authentic, crosslinked sequence due to mutations in RNA formed by UV exposure, and, (iii) the requirement for a suitable antibody recognizing a known RBP. Hence, in order...
to elucidate new proteins that bind to known lncRNAs, different approaches have been developed.

Capture Hybridization Analysis of RNA Targets (CHART) consists of hybridizing short, affinity-tagged oligonucleotides (biotinylated C-oligos) to target potential genomic binding sites of lncRNAs by hybridization to DNA in crosslinked extracts. By using streptavidin-conjugated beads, both chromatin-associated lncRNAs, as well as the RBPs associated with them, can be identified by either RNase H mapping or mass spectrometry (MS), respectively. The lncRNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1)-associated DNA was identified using the CHART technique (14). Another similar method, Chromatin Isolation by RNA Purification (ChIRP), uses short biotinylated probes (~20 nt) that ‘tile’ the full-length of lncRNAs to identify their genomic binding sites by sequencing. LncRNA-interacting proteins can also be determined by MS (ChIRP-MS) (15). ChIRP-MS has recently been used to identify 81 proteins that interact with XIST, including the transcriptional corepressor SMRT/HDAC1 Associated Repressor Protein (17). Thus, the development of these methods that enable both the identification of binding sites of lncRNAs and their associated proteins has immense potential to advance our understanding of gene regulation along with implications in biomedicine.

4. CLASSIFICATION OF LONG NON-CODING RNAs

LncRNAs can be classified according to their relative locations within or adjacent to the protein-coding genes from which they are transcribed (Figure 1 and Table 1): (i) intergenic lncRNAs (lincRNAs), (ii) intronic lncRNAs, (iii) untranslated-region (UTR)-associated lncRNAs, (iv) sense lncRNAs, (v) natural antisense lncRNAs (NATs), (vi) enhancer lncRNAs (eRNAs), (vii) promoter-associated lncRNAs (paRNAs), and (viii) promoter upstream transcripts (PROMPTs).

4.1. LincRNAs

LincRNAs are the class of lncRNAs that have been best studied. They contain the same chromatin signature as actively transcribed genes, namely the K4-K36 domain. This signature consists of a short region with histone H3 lysine 4 trimethylation (H3K4me3) and
a longer region with histone H3 lysine 36 trimethylation (H3K36me3) at the promoter and in the transcribed region, respectively. After excluding K4-K36 domains corresponding to known protein-coding genes, numerous lncRNAs have been identified (18). The expression patterns of certain lncRNAs are correlated to profiles of protein-coding genes, suggesting their conservation throughout evolution. Importantly, these genes are involved in cell-cycle regulation, immune surveillance and pluripotency of ESCs (9), amongst other functions. In addition, it has been reported that lncRNAs can interact with chromatin remodelling proteins, e.g. Polycomb group of proteins (PcG), in order to regulate gene expression.

4.2. Intronic lncRNAs

Intronic lncRNAs are the most abundant lncRNAs. It has been demonstrated that the expression profile of intronic lncRNAs and their cognate protein-coding gene correlate and possess tissue-specific patterns (20). Newly identified circular intronic lncRNAs (ciRNA) are derived from spliced introns that do not undergo debranching. CiRNA are abundant in the nuclei and have been suggested to sequester RBPs to regulate gene expression (20).

4.3. UTR-associated lncRNAs

UTR-associated lncRNAs can be transcribed independently of the transcription of the parental gene. However, since they overlap with each other, it has been challenging to study the function of UTR-associated lncRNAs through knockdown or knockout strategies (21). The expression of UTR-associated lncRNAs has been shown to be cell-type specific and they can act as decoys to regulate the binding of transcription factors.

4.4. Sense lncRNAs

Sense lncRNAs may contain a small ORF (sORF) overlapping with the same start codon as their host gene, thus encoding a small polypeptide. Sense lncRNAs can overlap with protein-coding genes or with entire introns. Numerous sense lncRNAs have been shown to function both as ncRNAs and protein-coding genes, for example SRA, p53 and ENOD40 (22).

4.5. NATs

NATs overlap with protein-coding genes and are transcribed from the opposite strand at the same genomic locus. They function either in cis to regulate the transcription of the sense protein-coding gene, or in trans to regulate the transcription of genes located at different genomic loci. NATs undergo fewer splicing events and are less abundant than their corresponding sense transcripts. Even though NATs are found in animals, plants, yeast, and prokaryotes (23), they are not evolutionally conserved.

4.6. eRNAs

eRNAs are known to be involved in chromatin looping to stabilize the interaction between enhancer and promoter of a nearby target gene, which promotes transcriptional initiation. They play important roles in genome integrity, cellular homeostasis and metabolic responses (24, 25). eRNAs originate from enhancer regions which are particularly enriched with the histone H3 lysine 27 acetylation (H3K27ac) mark (26, 27).

4.7. paRNAs

paRNAs are transcribed from the sense or antisense direction in correlation of the promoter of a protein-coding gene. paRNA lncRNAs can directly inhibit the transcription of a nearby protein-coding gene through direct alteration of the chromatin state (28), or by the recruitment of other proteins, such as transcription factors (29). For instance, a cyclin D1 (CCND1) paRNA has been found to recruit an RBP translocated in liposarcoma (TLS) to the promoter of CCND1 to exert transcriptional repression through histone acetyltransferase inhibitory activity (30).

4.8. PROMPTs

PROMPTs arise from the antisense strand in nucleosome-depleted regions. Previous studies have shown that PROMPTs are 5' capped and 3' polyadenylated (31), however rapidly degraded by the nuclear exosome. Therefore, it remains unclear whether they have a cellular function, particularly because they are located in nucleosome-depleted regions that expose DNA to binding factors, which could randomly initiate transcription.

This wide range of IncRNA that have been defined thus far appear to function in intricate regulatory networks that usually involves epigenetic machineries. Their dysregulation is a hallmark of many diseases, including cancer.

5. FUNCTIONS OF lncRNAs

Most lncRNAs that have been characterized play important roles in regulating gene expression by interacting with DNA, RNA and protein complexes, and can function in both the nucleus and cytoplasm. Here, we briefly describe some possible mechanisms by which lncRNAs act at multiple levels to regulate gene expression (Figure 2)

5.1. Signal

The expression of a lncRNA is cell type-specific and varies in response to specific stimuli,
indicating that it is under transcriptional control. Because some lncRNAs possess a regulatory function, they can serve as a molecular signal, e.g. to integrate developmental cues (32).

5.2. Decoy

A lncRNA can inhibit the function of a protein, such as a transcription factor or a chromatin modifier, by acting as a binding site and thereby titrating away a protein target (33, 34).

5.3. Guide

A lncRNA can recruit chromatin-remodelling proteins to specific loci to activate or inactivate transcription (35).

5.4. Scaffold

Because of their secondary structure, lncRNAs can bind more than two protein partners, where the lncRNA serve as adaptors to assemble regulatory protein complexes onto chromatin (36).

5.5. miRNA sponge

Recent studies have proposed the competitive endogenous RNA (ceRNA) hypothesis, in which transcripts with common miRNA binding sites can inhibit miRNAs competitively, by acting as a molecular ‘sponge’, leading to the de-repression of miRNA targets (37, 38). Nevertheless, the prevalence, functional significance and targets of lncRNAs as miRNA sponges in cancer development are mostly unknown.

5.6. miRNA precursor

LncRNAs can be processed to generate miRNAs, pointing towards an interplay in gene expression regulation between the lncRNA and miRNA pathways (39, 40).

6. POST-TRANSCRIPTIONAL MODIFICATIONS OF lncRNAs

Recent methylated RNA sequencing studies have shown that RNA modifications are present in lncRNAs, and such modifications could affect processing, stability, intracellular localization, as well as their interactions with proteins or RNA targets (for review see 41). N$^6$-methyladenosine (m$^6$A) is the most abundant internal modification in mRNAs and lncRNAs. Previous work has identified the presence of two m$^6$A residues in the MALAT1 lncRNA that are located in two hairpin-stem structures (42). One of these modified nucleotides (A2577) reduces the stability of the hairpin stem, influencing the accessibility of a protein binding site by a mechanism termed the “m$^6$A-switch” (43). Specifically, m$^6$A modification disrupts base-pairing and exposes the opposing U-rich site for the binding of heterogeneous nuclear ribonucleoprotein C (HNRNPC). Similarly, the binding of heterogeneous nuclear ribonucleoprotein G (HNRNPG) is enriched when MALAT1 is m$^6$A-modified (44). m$^6$A modification is also involved in X chromosome inactivation. XIST lncRNA is highly decorated with m$^6$A, which acts as a scaffold for the assembling of silencing complexes, such as YTH domain containing 1 (YTHDC1), for XIST-mediated gene repression (45). LncRNAs can also bear 5-methylcytosine (m$^5$C) modifications.
Crosstalk between IncRNA and Polycomb in tumorigenesis

Crosstalk between IncRNA and Polycomb in tumorigenesis

Loss of m\textsuperscript{5}C within eRNAs diminishes their stability, affecting the transcription of associated mRNAs (24). Another study has demonstrated the presence of m\textsuperscript{5}C in \textit{HOTAIR} and \textit{XIST} lncRNAs within functionally important regions that are known to mediate interaction with chromatin remodelling complexes (46). For instance, bisulfite sequencing of RNA has identified the m\textsuperscript{5}C mark within the repeat-A region of \textit{XIST}, which affects Polycomb repressive complex-2 (PRC2) binding (46). Furthermore, \textit{N}\textsuperscript{4}-methyladenosine (m\textsuperscript{1}A) and pseudouridine (Ψ) sequencing approaches have also identified these modifications on lncRNAs such as \textit{MALAT1} and \textit{XIST} although their function remains unclear (47). Hence, further studies are needed in order to understand the functional relevance of lncRNA modification during carcinogenesis.

7. POLYCOMB GROUP OF PROTEINS

PcG were originally characterized in \textit{Drosophila melanogaster} as transcriptional repressors of homeotic genes (Hox), which are required for the correct spatiotemporal expression of developmental regulators along the body axis (48). In most metazoan species, PcG complexes include PRC1, PRC2 (49), and the more recently identified Pho-Repressive Complex (Pho-RC) and Polycomb Repressive De-Ubiquitinase (PR-DUB) (50). In \textit{D. melanogaster}, PcG are recruited to specific target sites by Polycomb response elements (PREs), 1 kb DNA elements containing recognition sequences for DNA binding proteins. So far, PREs have not been elucidated in mammalian, remaining unclear how PRC1 and PRC2 are targeted onto chromatin to identify their target genes.

The minimal PRC2 core complex consists of three subunits: Enhancer of Zeste 1/2 (EZH2/1), Embryonic Ectoderm Development (EED), and Suppressor of Zeste 12 (SUZ12). EZH2/1 contain a putative RNA-binding domain and a conserved SET domain that catalyses the mono-, di-, and trimethylation of lysine 27 (K27) of histone H3 (H3K27me1, H3K27me2, and H3K27me3), contributing to the repression activity of PRC2 (51-54). SUZ12 is a regulatory subunit with a putative RNA-binding domain whereas EED binds H3K27me3 to increase the affinity of PRC2 with nucleosomes. Both SUZ12 and EED are required for EZH2 enzymatic activity.

After imposition of H3K27me3 by PRC2, PRC1 is recruited and catalyses the ubiquitylation of histone H2A on K119 (H2AK119Ub), maintaining transcriptional repression (55). The core complex of...
Crosstalk between lncRNA and Polycomb in tumorigenesis

PRC1 presents four subunits that are heterogeneous depending on the cellular context. In mammals, these subunits are homologous to those found in *D. melanogaster* (56) and consist of: (i) a subunit that belongs to the chromobox family (CBX2, CBX4, CBX6, CBX7, or CBX8) containing a chromodomain which binds to H3K27me3; (ii) a Polyhomeotic-like protein (PHC1, -2 or -3); (iii) the E3 ubiquitin-protein ligase RING1A/B; and (iv) a member of the PcG RING finger (PCGF) family including BMI1 (also known as PCGF4) or MEL18 (also known as PCGF2). PCGF proteins catalyse the formation of H2A ubiquitylation, which not only modulate transcriptional elongation by Pol II but also the formation of preinitiation complexes (57, 58). How PRC1 functions to repress transcription is not well understood as numerous PRC1 components have been described, increasing the complexity of study. Recently, non-canonical forms of PRC1 have been characterized. All PCGF proteins can interact with RYBP/YAF2 in competition with CBX proteins, forming a variant of PRC1 devoid of CBX proteins. Studies have shown that a variant of PRC1 (containing PCGF1, 3, 5) is recruited to chromatin first and is responsible for H2A ubiquitylation, independently of PRC2 recruitment and H3K27me3 deposition (59). Notably, the variant PCGF1/PRC1 complex contains an additional factor, KDM2B, which is a H3K4 and H3K36 lysine demethylase. KDM2B is recruited to non-methylated CpG islands via its DNA-binding domain, and ultimately facilitates the recruitment of PRC2, required for the deposition of H3K27me3.

Importantly, PcG complexes modulate gene expression during differentiation, cell lineage specification and morphogenesis (60, 61). Loss of PRC1/2 subunits have been shown to lead to embryonic lethality (62, 63), and mutations in PcG proteins can lead to altered activity, which might promote aberrant expression of oncogenes (64, 65). For instance, the overexpression of PRC1 subunit BMI1 is associated with a repression of the tumour suppressor locus *INK4b-ARF-INK4a*, which in turn induces cell proliferation (66). In addition, previous studies have shown that the overexpression of EZH2 is involved in several cancers, such as melanoma, breast and prostate cancer (67-69). Therefore, it is important to gain a deeper understanding of the interplay between lncRNA and PcG in order to elucidate their combined role in cancer initiation and progression.

### 8. POLYCOMB RECRUITMENT BY lncRNAs

Several lncRNAs have been shown to recruit Polycomb proteins to specific loci in order to modify epigenetic chromatin states, and thereby to repress gene expression. Some well-documented examples include *XIST* (70, 71), *antisense non-coding RNA in INK4 locus* (ANRIL) (72, 73), *MALAT1* (74, 75), and *HOX transcript antisense intergenic RNA* (HOTAIR) (76) (Table 2), which will be extensively described below.

<table>
<thead>
<tr>
<th>LncRNA</th>
<th>Polycomb</th>
<th>Subunits</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>XIST</em></td>
<td>PRC2</td>
<td>EZH2, SUZ12, RBP4 + RBP7</td>
<td>(71, 106) (107) (224)</td>
</tr>
<tr>
<td></td>
<td>PRC1</td>
<td>RYBP, PCGF3/5</td>
<td>(16, 115, 224) (16, 113)</td>
</tr>
<tr>
<td><em>ANRIL</em></td>
<td>PRC2</td>
<td>SUZ12, EED, RBAP46</td>
<td>(144, 138) (144) (144)</td>
</tr>
<tr>
<td></td>
<td>PRC1</td>
<td>CBX7, RING1B</td>
<td>(73) (144)</td>
</tr>
<tr>
<td><em>MALAT1</em></td>
<td>PRC2</td>
<td>EZH2, SUZ12</td>
<td>(173, 173) (174)</td>
</tr>
<tr>
<td></td>
<td>PRC1</td>
<td>CBX4</td>
<td>(82)</td>
</tr>
<tr>
<td><em>HOTAIR</em></td>
<td>PRC2</td>
<td>EZH2</td>
<td>(76, 106, 211)</td>
</tr>
<tr>
<td></td>
<td>PRC1</td>
<td>SUZ12</td>
<td>(76)</td>
</tr>
</tbody>
</table>

**Table 2.** Association of *XIST, ANRIL, MALAT1, and HOTAIR* lncRNAs with Polycomb
Crosstalk between IncRNA and Polycomb in tumorigenesis

influences carcinogenesis. In vivo, the interaction between IncRNAs and PRC2 could occur through multiple mechanisms, including bridging of regulatory proteins, and can be influenced by post-translational modification of chromatin remodelling complexes, or IncRNAs. It has been shown in vitro that the interaction between IncRNAs and PRC2 occurs through the EZH2 subunit. However, high affinity PRC2 binding does not rely on a specific RNA sequence but rather appears to be promiscuous. Indeed, recent studies have shown that PRC2 does not necessarily discriminate between RNA species. For example, PRC2 can bind to both in vitro transcribed RNA from the 5' end of HOTAIR and to maltose-binding protein (MDB) mRNA from Escherichia coli, which lack Polycomb proteins. This supports the concept that PRC2 binding to IncRNA is unspecific (79). On the other hand, a recent study has shown that RNA and chromatin interact competitively with PRC2 (80). As such, degradation of RNAs increased the recruitment of PRC2 to chromatin, whereas the release of PRC2 from chromatin increased RNA binding to PRC2. Although there has been considerable focus on PRC2, several studies have shown that IncRNAs can also bind to PRC1, in particular to CBX proteins. CBX7 directly interacts with ANRIL, resulting in repression of the INK4b/ARF/INK4a tumour suppressor locus (81). Furthermore, CBX4 binds to TUG1 and MALAT1/NEAT2, stimulating EZF1 SUMOylation (82), which results in increased cellular proliferation. Nevertheless, the nature of interactions between IncRNAs and PcGs in vivo remain unclear. Hence, a major unanswered question is how IncRNAs recruit PcG to specific targets.

9. XIST

XIST is an IncRNA that plays a central role in the initiation of X chromosome inactivation (XCI) in early embryogenesis (83-85). This process leads to the highly regulated transcriptional silencing of one X chromosome in female mammals (designated as inactive X-chromosome (Xi) or Barr body (86)) to ensure dosage compensation between the sexes (87, 88). In mice, XCI occurs in two lineage-specific forms. At the 2-4-cell stage, imprinted XCI leads to parental X-chromosome silencing, and the paternally imprinted Xist is sustained in the cells of the trophectoderm and primitive endoderm, which give rise to extra-embryonic tissues (89, 90). In contrast, Xi is re-activated in the epiblast of the inner cell mass, and then random XCI is induced during the peri-implantation stage. Thus, both the paternal and maternal X chromosomes have an equal chance of being inactivated.

The XIST gene is located in the X inactivation centre (XIC) (91, 92), and its expression marks the future Xi. Two RNA-based switches positively and negatively regulate Xist. Jpx for Xi, and Tsix for active X-chromosome (Xa). Jpx RNA may bind CTCF and titrate out its repressive effect on the Xist promoter (32). Conversely, Tsix represses Xist induction by several means, including altering the chromatin state of Xist (93-95), recruiting the RNAi machinery (96), and facilitating PRDM14 binding to Xist intron 1 to suppress its expression (97). In turn, Tsix is regulated by Xite, a proximal ncRNA that sustains Tsix expression on the future Xa (98, 99).

9.1. XIST and Polycomb

Chromosome coating by Xist recruits chromatin modification complexes that initiate gene silencing on the Xi, which will be propagated as Xi in all subsequent cell divisions throughout the life of the female individual (88). A key silencing factor recruited by Xist is Polycomb (100). Both PRC2 and PRC1 complexes are enriched on the Xi early on during differentiation, but become depleted at later stages, suggesting that their association with the Xi is linked to early maintenance (101, 102). Moreover, PRC2/H3K27me3 have been shown to co-localize with XIST IncRNA, both at metaphase and interphase (70, 103, 104). During XCI, PRC2 is first targeted to the Xi by RepA; a IncRNA located within the 5' end of Xist that encompasses the conserved A-repeat domain of Xist (71, 105). Xist then recruits PRC2 (101, 103, 104) through direct interaction between the A-repeat domain and EZH2 and/or SUZ12 (71, 106, 107) or indirectly through JARID2 (108). From Xic, PRC2 spreads along the future Xi, concentrating predominantly within bivalent domains coinciding with CpG islands (109). As XCI proceeds, the coating of the future Xi by Xist IncRNA correlates with recruitment of 3,000-4,000 moderate Polycomb sites, facilitating the spreading of H3K27me3. Nevertheless, the aforementioned model has been challenged by non-confirmatory results. For example, Xist expression in early mouse embryos precedes PRC2 recruitment to Xi (90, 110), and Xist lacking the A-repeat domain is able to recruit PRC2, albeit less efficiently (101). Moreover, super-resolution FISH/immunofluorescence studies have shown that PRC2 and XiST IncRNA are spatially segregated (111). A further consideration is that comprehensive identification of RBPs by MS has not identified PRC2 as an interactor with Xist, although PRC1 proteins were detected (16, 17).

It has been shown that PRC1 also participates in XCI (112). Although it was originally thought that PRC1 functioned through the classical hierarchical model described above, a novel mechanism referred as reverse hierarchical recruitment has been proposed whereby PRC1 recruitment in Xi precedes that of PRC2 (113). Indeed, H2AK119Ub imparted by PRC1 was present on Xi in the absence of PRC2/H3K27me3 (114). Furthermore, non-canonical PRC1 complexes (containing the RYBP cofactor) are also recruited to Xi, via an XIST IncRNA dependent but H3K27me3 independent mechanism (59, 115).
9.2. Role of XIST in tumorigenesis

Although XCI occurs during early development, XIST is also expressed in adult females. Several studies have shown alterations of XCI in human cancer. In aggressive breast and ovarian tumours, the Barr body is frequently undetectable (116-119). More recently, cytogenetic studies of breast carcinoma provided evidence that the loss of the Barr body can be accompanied by the gain of an additional Xa, which may be due to X-linked gene reactivation (120-122). The role of X chromosome dosage in breast tumour development is further strengthened by evidence from male breast tumours, where lack of the Y and a duplication of the X chromosome are often observed (123, 124). In addition, males with supernumerary Xs have an increased risk of developing breast cancer (125).

Several studies suggest that loss of XIST IncRNA could drive disease progression in cancer through reactivation of Xi. Indeed, XIST IncRNA has been shown to act as a tumour suppressor by reducing activation of the AKT pathway in breast cancer, resulting in limited cell viability. Knockdown of either XIST or SPEN, encoding an important interactor for XCI establishment (16, 17), suppressed the expression of phosphatase PHLPP1 in AKT dephosphorylation (126). In addition, RNA expression profiling of sporadic basal-like cancers, which have lost Xi and XIST IncRNA, also displayed overexpression of some X-linked genes (127). Loss of BRCA1, a tumour suppressor protein that is frequently mutated in familial cases of breast cancer (128), was initially found to lead to Xi perturbation and dysregulation of XIST IncRNA (129). Subsequently, BRCA1 was shown to maintain proper Xi heterochromatin (130). Furthermore, loss of XIST IncRNA was associated with a BRCA1 deficiency in sporadic basal-like cancers (127). However, subsequent studies revealed that XIST IncRNA functioned independently of BRCA1 in XCI (131-133). Another chromatin regulator frequently overexpressed in cancer, Aurora B Kinase (AURKB), has also been proposed to regulate the association of XIST to the Xi (134). However, the precise consequence of XIST on the status of Xi is unclear. Direct causality has now emerged from a recent study showing that loss of XIST in the hematopoietic lineage led to the development of female-specific leukemia in mice (135). Gene expression profiling over the course of disease progression revealed significant upregulation of X-linked genes, suggesting the possibility of Xi reactivation following XIST loss. This sensitivity of hematopoietic cells to Xist dysregulation support previous work in which overexpression of Xist in mice resulted in lethal anemia due to defective haematopoiesis (136), and overexpression in a lymphoma cell model suppressed tumorigenicity (137).

10. ANRIL

ANRIL (also known as CDKN2B antisense RNA 1 (CDKN2B-AS1)) is transcribed in the opposite direction from the INK4b-ARF-INK4a tumour suppressor locus and the methylthioadenosine phosphorylase (MTAP) gene (138, 139). The INK4b-ARF-INK4a locus encodes three critical tumour suppressor genes, p14ARF (p19ARF in mice), p15INK4b, and p16INK4a, all of which play a central role in cell cycle arrest, affecting key cellular processes such as senescence, apoptosis, and ESC self-renewal by triggering both retinoblastoma (Rb) and p53 pathways (140, 141). The locus also contains a fourth gene, MTAP, which has been associated with carcinogenesis (142, 143). ANRIL contains 19 exons, many of them consisting of LINE, SINE, and Alu repetitive elements (144,145). The first exon of ANRIL is located 300 nt upstream of the transcription start site of ARF. Hence, these two genes share a bidirectional promoter (146, 147). Moreover, ANRIL can exist in both linear and circular (circANRIL) forms, which have been reported to be cell- or tissue-specific (144, 148-151), yet their functional relevance is still unknown.

10.1. ANRIL and Polycomb

ANRIL functions as a cis-regulator of the INK4a-ARF-INK4b locus by recruiting PRC1 and PRC2 complexes (81, 138). ANRIL has been shown to interact with the PRC2 component SUZ12 to repress the expression of p15INK4b. Thus, depletion of ANRIL increased the expression of p15INK4b and inhibited cellular proliferation (138). Moreover, it has been shown that ANRIL is stably associated with PRC1 via CBX7 as a nascent transcript generated by the Pol II. This allowed CBX7 recognition of H3K27me3, leading to p16INK4a silencing (81). In line with these findings, RIP-seq experiments in which two specific exon-combinations of ANRIL were overexpressed, showed a binding of ANRIL with CBX7 and RING1B of PRC1, binding with the PRC2 subunits EED, RBAP46 (also known as RBP7), and SUZ12, and PRC-associated proteins JARID2, RYBP and YY1 (144).

10.2. Role of ANRIL in tumorigenesis

ANRIL was originally identified in familial melanoma patients with neural system tumours and a germ-line deletion of the entire INK4b-ARF-INK4a gene cluster, suggesting that ANRIL might play a role in oncogenesis (152). The gene cluster has been shown to be deleted or silenced in both alleles in ~40% of human cancers (153). On the other hand, genome-wide association studies have identified ANRIL as a risk factor in several types of cancer, including gastric cancer, esophageal squamous cell carcinoma, breast and bladder cancers etc. (154-157). Moreover,
the ANRIL locus has been identified as a hotspot for disease-associated polymorphisms showing a significant correlation with tumour development, cardiovascular disease, and other conditions (158). These polymorphisms alter the expression pattern of ANRIL splice variants, and in consequence dysregulate the INK4b-ARF-INK4a locus expression.

In normal cells, induction of ANRIL transcript levels by E2F1 is required for the suppression of p14ARF, p15INK4b, and p16INK4a expression at the late stage of the DNA damage response, in order to return to baseline levels after the completion of the DNA repair process. However, in cancerous cells, aberrant expression of ANRIL would cause a blockage DNA damage response control, leading to genomic instability and tumour progression (159). It has been shown that ANRIL also influences cell proliferation by regulating target genes in trans. Hence, in gastric cancer tissues, ANRIL cooperated with miRNAs in the epigenetic level by binding to EZH2. Specifically, ANRIL silenced miR-99a/miR-449a, which triggers the activation of miR-99a/miR-449a target genes mTOR and CDK6, and as a consequence upregulated the CDK6 target gene E2F1 (156). Similarly, in hepatocellular carcinoma ANRIL abolished miR-122-5p expression, enhancing colony formation ability, metastasis and invasion (160). Moreover, in esophageal squamous and thyroid cancer cells, ANRIL has been shown to influence cell growth by repression of the TGFβ/Smad signalling pathway (154, 161), although the nature of the interaction between ANRIL and TGFβ1 is unclear.

11. MALAT1

MALAT1 (also known as nuclear-enriched transcript 2 (NEAT2)) was originally found to be associated with lung cancer (162). Although MALAT1 is a highly abundant lncRNA conserved across many mammalian species, it is not essential for mouse development (163). Malat1 knockout mice do not show any abnormality during embryonic or postnatal development, suggesting that MALAT1 is dispensable or might be important only under certain pathological conditions (163-165).

MALAT1 can be found as a long nuclear transcript or as a small cytoplasmic RNA. Thus, the primary MALAT1 transcript contains a tRNA-like structure at its 3’ end, which is cleaved by enzymes involved in tRNA biogenesis. It is processed to form the MALAT1-associated small cytoplasmic RNA (mascRNA), which might fulfil additional, but so far unknown, functions (166). The 3’ end of MALAT1 contains a conserved triple-helix structure termed expression and nuclear retention element (ENE), which prevents 3’ end nuclease cleavage and enhances translation when placed downstream of an ORF (167, 168).

At the molecular level, multiple functions have been proposed for MALAT1 (139). The long MALAT1 transcript is retained in the nucleus and specifically localizes to nuclear speckles where it interacts with several pre-mRNA splicing factors to regulate alternative mRNA splicing (169, 170). MALAT1 is not essential for the integrity of nuclear speckles but for the modulation of the association of active pre-mRNA splicing factors with speckles (170). It has also been linked to the modulation of the epigenetic machinery (171), and is known to be associated with active genes where it would serve as a scaffold that binds proteins to activate transcription at specific loci (171).

11.1. MALAT1 and Polycomb

MALAT1 has been shown to promote tumour cell proliferation, invasion and metastasis through Polycomb. In T and NK cell lymphoma, MALAT1 IncRNA is highly expressed and this has been correlated with the PRC1 component BMI1, and related to poor prognosis in patients with mature T cell lymphoma. In addition, direct binding of MALAT1 to the PRC2 components EZH2 and SUZ12 was shown in a T cell lymphoma cell line (172). In renal carcinoma, MALAT1 interacted with EZH2 to promote Wnt/β-catenin pathway, thus promoting tumour invasion and metastasis (173). In addition, MALAT1 has been shown to facilitate epithelial-to-mesenchymal transition (EMT) through interaction with SUZ12, which resulted in downregulation of E-cadherin and upregulation of N-cadherin and fibronectin (174). In association with the PRC1 subunit CBX4 (E3 SUMO-protein ligase) and other protein complexes, MALAT1 controls sub-nuclear architecture, acting as a sensor for the activation of specific transcriptional programs. When methylated CBX4 is bound to the TUG1 IncRNA, it can be recruited to growth control genes within repressive Polycomb bodies. In response to growth signals, the demethylated form of CBX4 interacts with MALAT1 and controls the relocation of such genes to interchromatin granules, leading to the promotion of E2F1 sumoylation and activation of growth transcriptional programs (82).

11.2. Role of MALAT1 in tumorigenesis

Several studies have shown that MALAT1 plays a pivotal role in the malignancy phenotypes of cancer. Hence, MALAT1 has been associated with cancer growth, invasion and metastasis (74, 175). MALAT1 was originally identified as a lncRNA upregulated at early-stage in non-small cell lung cancers, which has a propensity for metastasis (162). Following studies have shown that MALAT1 overexpression is related to multiple types of tumour, e.g. in the liver (176-180), breast (181-185) and colon (186-188). However, the target genes of MALAT1 are variable among various types of cancer. Interestingly, chromosomal translocation breakpoints associated
with cancer have also been identified within the MALAT1 locus (189-191).

MALAT1 can promote tumour development by other mechanisms that do not involve Polycomb. For instance, it has been reported that MALAT1 binds competitively to the tumour suppressor gene splicing factor, poly-glutamine rich (SFPQ), thus releasing it from a SFPQ/PTBP2 complex (polypyrimidine-tract-binding protein), which leads to an increased level of PTBP2 alone (192). Upregulation of MALAT1 promoted chemokine ligand 5 signalling, resulting in increased level of Snail and EMT promotion (193). Furthermore, MALAT1 enhanced stemness in pancreatic cancer cells and in glioma stem cell lines by promoting the expression of self-renewal related factors through Sox2 (194-195). In addition, MALAT1 can activate the ERK/MAPK pathway, which orchestrates cancer progression by inducing transcriptional programs that regulate cell invasion (196). On the contrary, a recent study has shown that MALAT1 functions as a tumour suppressor by inactivating ERK/MAPK signalling pathway in glioma cells (197). Further studies will be crucial in order to elucidate whether these discrepancies are due to the usage of different cell lines. Nonetheless, depletion of MALAT1 has been associated with induced apoptosis leading to a decreased proliferation and suppression of the tumorigenic phenotype in many different cancer cell lines (198-201). Furthermore, MALAT1 levels are regulated during cell division, influencing the expression of genes involved in cell cycle progression. In addition, depletion of MALAT1 results in induction of the p53-mediated DNA damage response pathway in normal human diploid fibroblasts (202).

12. HOTAIR

HOTAIR is a spliced and polyadenylated IncRNA of 2,158 nt (76). This IncRNA is transcribed from the antisense strand of the HOXC locus located on the Chromosome 12. In mammals, 39 HOX encoding genes are grouped into four clusters (HOXA-D) on different chromosomes. During development, HOX expression occurs in a precise temporal and spatial sequence, which follows their chromosomal order (203).

HOTAIR has been proposed to silence HOXD genes in trans by recruiting PRC2 (76). Deletion of HOTAIR in human fibroblasts induced the expression of different members of the HOX family, which was associated with a loss of the H3K27me3 mark on the HOXD locus (76). However, it was thought initially that the function of HOTAIR was specific to human since mice bearing a complete deletion of the HoxC locus exhibited no significant changes in HoxD expression or chromatin marks (204). Nonetheless, a later study showed that deletion of mouse Hotair resulted in the derepression of HoxD genes and several imprinted loci (205). This study suggested that the lack of effect on the previously reported model was due to the concomitant deletion of all HoxC genes, which may have masked or compensated a potential alteration caused by the absence of Hotair alone (204). To date, four different alleles that result in either a partial or a complete deletion of Hotair, have been reported without reaching a consensus on the function of this IncRNA (204-208). Such differences could be explained by the use of different genetic background and their influence on knockout phenotypes (209).

12.1. HOTAIR and Polycomb

HOTAIR adopts a modular secondary structure at its 5’ end, which is critical for PRC2 binding in vitro (210, 211). HOTAIR also contains a 3’ domain that interacts with LSD1/CoREST/REST that catalyses H3K4me2 demethylation, serving as a scaffold for distinct histone modification complexes. The complex then targets the HOXD locus on chromosome 2 to silence genes involved in the suppression of metastasis (211). HOTAIR was previously thought to localize and target PRC2 genome-wide to modulate the cancer epigenome (212), but a recent study has shown that HOTAIR can also repress transcription independently of PRC2 by interacting with as yet unknown factors (213).

A HOTAIR-PRC2 interaction in vivo may occur by intermediate bridging proteins such as JARID2 (77). Furthermore, it has been shown that phosphorylation of EZH2 at threonine 345 increases its binding affinity for HOTAIR (46). Another study has shown that HOTAIR overexpression can induce the interaction of a complex containing SUZ12 and LSD1 with target genes, while depletion of HOTAIR has the opposite effect (211).

12.2. Role of HOTAIR in tumorigenesis

Aberrant expression of HOTAIR has been reported in various cancers, including breast, colorectal, pancreas or lung (214). Hence, HOTAIR could be used as a prognostic biomarker and as a predictor of patient survival in numerous cancer types (214). Furthermore, HOTAIR overexpression has been associated with metastasis (215). Nevertheless, the underlying mechanisms by which HOTAIR is involved in cancer development and metastasis remain poorly understood.

In breast cancer cells in which HOTAIR is overexpressed, SUZ12 and EZH2 are actively recruited to the H3K27me3 signature in the promoters of 854 genes (212). Most of these genes encode for tumour suppressor proteins such as progesterone receptor, HOXD10 and protocadherin, suggesting that HOTAIR facilitates PRC2 occupancy in order to downregulate their expression. Consistently, depletion of EZH2 or SUZ12 induced gene expression, confirming the interplay between HOTAIR and PRC2 to regulate
transcription. In vivo, an intact PRC2 complex was required in order to induce breast cancer invasiveness by HOTAIR (212).

HOTAIR has been shown to regulate several other processes associated with carcinogenesis, such as EMT and acquisition of stemness (216), independently of Polycomb. For instance, epithelial ovarian cancer was associated with high level of HOTAIR, reduced survival and lymph node metastasis. On the contrary, suppression of HOTAIR reduced cell invasiveness. This pro-metastatic effect of HOTAIR has been shown to be mediated by matrix metalloproteinases (217). Another study has shown that HOTAIR promoter possesses numerous estrogen-response-elements that are targeted by E2 to regulate HOTAIR expression (218). Additionally, estrogen receptors (ERα and ERβ) and ER co-regulators also activated HOTAIR by binding to the promoter regions in presence of E2. HOTAIR expression has also been reported to be regulated by c-Myc (219). In gallbladder cancer tissues, HOTAIR lncRNA levels correlated positively with c-Myc, but negatively with miR-130a expression. Knockdown of HOTAIR inhibited gallbladder cancer cell invasiveness, whereas such cell invasiveness was rescued by miR-130a, suggesting that oncogenic HOTAIR activity occurs via the negative regulation of miRNA-130a (219). In the same way, it has been shown that in renal carcinoma cells, HOTAIR is involved in cell invasion whereas miR-141 suppressed HOTAIR oncogenic effects (220).

13. CONCLUDING REMARKS

Despite considerable development in our understanding of lncRNAs over the past decade, the functions of the majority of lncRNAs are unknown, and many lncRNAs may not have been discovered yet. Only a fraction of annotated lncRNAs has been examined for biological function revealing that the alteration of lncRNA homeostasis can adversely affect cellular identity, thereby promoting cancer development in a tissue-specific and/or cell-type-specific fashion. As discussed in the previous sections, lncRNAs can have both tumour suppressive and oncogenic roles. In addition, they can be regulated by multiple mechanisms, such as chemical modification, increasing the complexity of the number of possible levels of lncRNA regulation. Although the roles of several classically defined lncRNA and their associated PcGs have been discussed in this review, many potential functions are still unexplained or remain controversial. As Polycomb in mammals lack clear DNA sequence specificity, an exciting general model in whereby lncRNAs recruited PcGs, specifically PRC2, to chromatin emerged. But since PRC2 binding to RNA is promiscuous, what is the relevance of lncRNA-Polycomb interaction? Yet, the biological function of this interaction remains unknown. Furthermore, how lncRNA can guide PcGs to specific targets is not fully elucidated mechanistically. Future studies should aim to answer these and other open questions, as well as to define the functions of each lncRNA systematically, in order to reveal the molecular mechanisms underlying tumorigenesis in different cancer types.

RNA-based therapeutics, such as antisense oligonucleotides (ASOs) or siRNAs aimed at targeting an lncRNA of interest for degradation, or the use of small molecules to disrupt lncRNAs interactions, will lead to new strategies for cancer therapy. However, despite growing knowledge about the function of lncRNAs in cancer, a broader understanding of the molecular mechanism of action, and the regulatory pathways, hierarchies and networks in which lncRNA and associated chromatin complexes operate, is the essential first step for therapeutic development. We predict that such knowledge will aid in unlocking the full potential of this intricate network of molecules, and will facilitate the prediction of cancer risk, track the prognosis of tumour fate and provide novel therapeutic approaches.

14. ACKNOWLEDGMENT

We sincerely apologize to authors whose work could not be included due to space limitations. We thank S. Malla and J. Gilthorpe for insightful comments and D. Munoz for graphical design. This work is supported by grants from the Knut and Alice Wallenberg Foundation, Umeå University, Västerbotten County Council, Kempe Foundation, Swedish Research Council, and the Cancer Research Foundation in Northern Sweden.

15. REFERENCES

DOI: 10.1146/annurev.bi.42.070173.002255

DOI: 10.1073/pnas.49.2.240

DOI: 10.1073/pnas.78.3.1371

4. Maxwell, E.S., and Fournier, M.J. The Small Nucleolar RNAs. Annu Rev of Biochem 64,


Crosstalk between lncRNA and Polycomb in tumorigenesis


36. Spitale, R.C., Tsai, M.C., and Chang, H.Y. RNA templating the epigenome Long noncoding RNAs as molecular scaffolds. Epigenetics-Us 6, 539-543 (2011) DOI: 10.4161/epi.6.5.15221


42. Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. Probing N6-methyladenosine RNA modiﬁcation status at single nucleotide resolution in mRNA and long noncoding RNA. RNA 19, 1848-1856 (2013) DOI: 10.1261/ma.041178.113


Crosstalk between lncRNA and Polycomb in tumorigenesis

DOI: 10.1038/ng.349


86. Barr, M.L., and Bertram E.G. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature 163, 676 (1949) DOI: 10.1038/163676a0


88. Lyon, M.F. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190, 372-373 (1961) DOI: 10.1038/190372a0


105. Wutz, A., Rasmussen, T.P., and Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* 30, 167-174 (2002) DOI: 10.1038/ng820


112. de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R.,
Crosstalk between IncRNA and Polycomb in tumorigenesis


15. Leeb, M., and Wutz, A. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J Cell Biol* 178, 219-229 (2007) DOI: 10.1083/jcb.200612127


Crosstalk between lncRNA and Polycomb in tumorigenesis


131. Pageau, G.J., Hall, L.L., and Lawrence, J.B. BRCA1 does not paint the inactive X to localize XIST RNA but may contribute to broad changes in cancer that impact XIST and Xi heterochromatin. *J Cell Biochem* 100, 835-850 (2007) DOI: 10.1002/jcb.21188


144. Holdt, L.M., Hoffmann, S., Sass, K., Langenberger, D., Scholz, M., Krohn, K., Finstermeier, K., Stahringer, A., Wilfert, W.,


199. Li, Z., Xu, C., Ding, B., Gao, M., Wei, X., and Ji, N. Long non-coding RNA MALAT1 promotes proliferation and suppresses apoptosis of glioma cells through derepressing Rap1B by sponging miR-101.
Crosstalk between lncRNA and Polycomb in tumorigenesis

2131

© 1996-2018

DOI: 10.1007/s11060-017-2498-5

DOI: 10.3892/ol.2016.4435

DOI: 10.7150/jca.21228

DOI: 10.1371/journal.pgen.1003368

DOI: 10.1242/dev.068346

DOI: 10.1371/journal.pgen.1002071

DOI: 10.1016/j.celrep.2013.09.003

DOI: 10.1371/journal.pgen.1006232

DOI: 10.1371/journal.pone.0125522

DOI: 10.1371/journal.pgen.1006485

DOI: 10.1371/journal.pgen.1006406

DOI: 10.1016/j.molcel.2015.03.006

DOI: 10.1126/science.1192002

DOI: 10.1038/nature08975

DOI: 10.15252/embj.201695335

Crosstalk between lncRNA and Polycomb in tumorigenesis


Key Words: cancer, lncRNA, Polycomb, XIST, ANRIL, HOTAIR, MALAT1, RNA modification, Review

Send correspondence to: Francesca Aguilo, Dept. Medical Biosciences, Umea University, SE-901 85 Umea, Sweden, Tel: 46 70-3718128, Fax: 4690121562, E-mail: francesca.aguilo@umu.se