Short Conceptual Overview

Nadia Vadaie and Kevin V. Morris*

Long antisense non-coding RNAs and the epigenetic regulation of gene expression

Abstract: Shortly after the completion of the human genome project in 2003, the Encode project was launched. The project was set out to identify the functional elements in the human genome, and unexpectedly it was found that >80% of the genome is transcribed. The Encode project identified those transcribed regions of the genome to be encoded by non-coding RNAs (ncRNAs). With only 2% of the genome carrying gene-encoding proteins, the conundrum was then, what is the function, if any, of these non-coding regions of the genome? These ncRNAs included both short and long RNAs. The focus of this review will be on antisense long non-coding RNAs (lncRNAs), as these transcripts have been observed to play a role in gene expression of protein-coding genes. Some lncRNAs have been found to regulate protein-coding gene transcription at the epigenetic level, whereby they suppress transcription through the recruitment of protein complexes to target loci in the genome. Conversely, there are lncRNAs that have a positive role in gene expression with less known about mechanism, and some lncRNAs have been shown to be involved in post-transcriptional processes. Additionally, lncRNAs have been observed to regulate their own expression in a positive feedback loop by functioning as a decoy. The biological significance of lncRNAs is only just now becoming evident, with many lncRNAs found to play a significant role in several human diseases.

Keywords: antisense; epigenetic; long non-coding RNA; ncRNA; transcription.

*Corresponding author: Kevin V. Morris, Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA; and School of Biotechnology and Biomedical Sciences, The University of New South Wales, Sydney 2052, NSW, Australia, e-mail: kmorris@scripps.edu

Nadia Vadaie: Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

High-throughput studies have contributed to uncovering the complexity of the human transcriptome and the pervasive genome-wide transcription. The Encode project revealed that >80% of the human genome is transcribed, and that the majority of the transcripts do not encode proteins. These newly discovered ncRNAs are functionally different from the canonical ncRNAs, such as transfer RNA and ribosomal RNA, which are involved in general cellular processes. Notably, many of these new ncRNAs are not as ubiquitously expressed as the housekeeping ncRNAs and do not appear to be translated (1). The new ncRNAs are classified into small ncRNAs (<200 nucleotides), such as microRNAs, small interfering RNAs, PIWI-interacting RNAs, and long ncRNAs (lncRNAs), which are at least 200 nucleotides in length (2, 3). This review focuses on antisense lncRNAs and their regulatory roles.

LncRNAs can be categorized into several groups. For example, one type of lncRNA can be characterized with respect to its complimentary protein-coding gene, which can be orientated in a sense and/or antisense direction. The lncRNAs and the protein-coding genes can therefore be in a head-head, tail-tail, or head-tail orientation. Other lncRNAs can be localized in introns (intronic) and within two genes (intergenic). Moreover, lncRNAs are expressed in a cell-/tissue-specific manner and under developmental stages, which highlights their functions in regulating gene expression (4, 5). LncRNAs have been found to modulate gene expression by various mechanisms, and a well-studied mechanism by which lncRNAs exert their functions is through their epigenetic regulation (6). This mechanism includes genetic imprinting and dosage compensation. In case of imprinting, the expression of either paternal or maternal allele is silenced. Imprinted lncRNAs, such as Kcnqot1 and Air, which are expressed from paternal alleles, repress a cluster of protein-coding genes on maternal alleles, for instance, Air silences Igf2r, Slc22a2, and Slc22a3. The epigenetic mechanisms by which these imprinted lncRNAs suppress gene expression are DNA methylation of promoters and the formation of repressive chromatin state by the histone methyltransferase G9a (7–10).
In dosage compensation, one maternal X-chromosome is silenced by the lncRNA Xist, which interacts and recruits histone-modifying complexes to target sites, and this, in turn, results in the formation of repressive chromatin marks, such as H3K9me3 and H3K27me3 (11, 12). Moreover, lncRNAs can modulate the chromatin state by functioning as a structural scaffold, whereby they provide the platform for the formation of chromatin remodeling complexes. The lincRNA HOTAIR forms such a scaffold to gather the histone methyltransferase polycomb repressive complex 2 (PRC2) and the histone demethylase LSD1. Through these interactions and complex formations, HOTAIR represses the expression of the HOXD gene (13).

At the epigenetic level, lncRNAs repress transcription by interacting and recruiting histone and chromatin remodeling proteins to target sites (Figure 1). These antisense lncRNAs can be expressed in a cis- or trans-based manner from RNA polymerase II promoters (14) and can be differentially localized within the cell (15). Some of the first such examples of antisense lncRNAs regulating protein-coding genes by targeting epigenetic complexes to the promoters of particular protein-coding genes were observed with antisense lncRNAs for tumor suppressor genes (16, 17). These lncRNAs were observed to epigenetically modulate the promoter of the targeted protein-coding gene, specifically altering the chromatin to a more compacted state (Figure 1). It is interesting to note that most of the tumor suppressor genes reported to date have been found to exhibit some level of antisense lncRNA expression associated with them. These tumor suppressor genes are also often the ones found to be

Figure 1  Antisense lncRNA-directed epigenetic regulation in human cells. Antisense lncRNAs can be expressed in trans (A) or cis (B). The lncRNAs can then interact with various protein components (C): Ezh2, DNMT3a, PRC2, and G9a are shown schematically, as all have been observed to be involved in lncRNA-mediated function, although not as one complete complex (8, 9, 19, 20). The antisense lncRNA protein complex can then target epigenetic remodeling of the homology containing loci (D), resulting ultimately in compaction of the target loci (E) and epigenetic silencing.
epigenetically silenced in human cancers [reviewed in ref. (18)].

More recent observations have begun to suggest that pseudogenes, which are non-protein-coding transcripts containing sequence homology to known protein-coding genes, are active epigenetic modulators of their respective protein-coding counterpart (19, 20). In one study, the Oct4 transcription factor, which is a dominant player in stem cell control as well as transcriptional fidelity, was found to be under epigenetic regulation of an antisense IncRNA emanating from a pseudogene (19). In a more recent body of work, the PTEN tumor suppressor gene was found to be bimodally regulated by its pseudogene, PTENpg1, whereby both PTEN protein translation and transcription appear to be controlled by one pseudogene network node (20, 21). Such bimodal regulation raises the level of IncRNA regulatory complexity to a state that has not before been observed and suggests that a vast regulatory network of interesting functions remains to be determined with regard to IncRNA-directed epigenetic regulation.

Higher-ordered RNA:RNA complexes formed from IncRNAs have also been observed in the control of gene clusters. In mouse testis, tubulin and the assembly of microtubule structures appear regulated by sense and antisense IncRNAs forming higher-ordered RNA:RNA complexes (22). In another study, the antisense IncRNA AS1DHRS4, which emanates in a head-head manner from complexes (22), was observed to regulate the gene cluster in both a cis- and trans-directed manner (23). AS1DHRS4 pairs with the sense counterpart in cis to mediate deacetylation of the DHRS4 gene, whereas the antisense IncRNA in trans is involved in epigenetic-mediated regulation of the locus through interactions with G9a and enhancer of zeste 2 (Ezh2) (23). These studies illustrate the vast complexity of antisense IncRNAs in both form and function with regard to control of gene expression.

Some IncRNAs can be sense oriented relative to protein-coding genes, emanating across the promoters of genes (22, 24), and function in an allosteric blocking-based manner. While previous studies have found that promoter-associated RNAs are required for RNA-directed transcriptional gene silencing (25), these promoter-associated IncRNAs have been found to obstruct epigenetic activation by binding the translocated in liposarcoma (TLS) protein, which is an RNA-binding protein that is also known to bind CREB (cAMP response element binding protein)-binding protein CBP/adenovirus p300. The ultimate effect of this IncRNA-binding TLS is the inhibition of histone acetyltransferase activity (26).

Other IncRNAs can regulate gene expression independently of changing the chromatin state, and by directly binding to transcriptional machinery and DNA elements within the protein-coding genes. The human dihydrofolate reductase DHFR gene encodes an IncRNA, which binds to the promoter and the transcription factor IIB, forming a complex. Through these interactions, the IncRNA disassembles the complex, which in turn leads to suppression of DHFR expression (27). Interestingly, IncRNAs can have a transcriptional activator role, for instance, the IncRNA EVF-2 forms a complex with the transcription factor Dlx-2, which then binds to the homeobox bicluster genes Dlx-5 and Dlx-6 and induces the activation and expression of the genes (28). Additionally, IncRNAs can have a regulatory function at the post-transcriptional level. The IncRNA ZEB2NAT modulates alternative splicing of the protein-coding gene ZEB2; in one case, the IncRNA splices an mRNA that evades the translational machinery and prevents the translation of ZEB2 protein, and in another instance ZEB2NAT interferes with the spliceosome and generates an alternatively spliced mRNA, which then can be translated (29). The IncRNA MALTA-1 is also involved in regulating the splicing by influencing the phosphorylation state of the splicing factors (30).

In addition to their role as transcriptional regulators of protein-coding genes, IncRNAs can modulate their own expression by interacting with other ncRNAs. For example, the IncRNA HULC binds to miRNA-372 and inhibits the expression of the miRNA target genes. One of these genes is a cAMP-dependent protein kinase catalytic subunit β (Prkacb), which, when inhibited, promotes the phosphorylation of the CREB domain of HULC. Then, phosphorylation of HULC leads to upregulation of its own expression (31).

LncRNAs elicit their functions both at the transcriptional and post-transcriptional levels; therefore, it is not surprising that the deregulation of IncRNAs has been implicated in diseases, such as cancers. For instance, aberrant expression of HOTAIR has been observed in breast, colon, and pancreatic cancers (32–34). Additionally, dysregulation of MALTA-1 has been linked to breast, colon, liver, lung, ovary, pancreas, and prostate cancers (35, 36). Moreover, some IncRNAs have been shown to become biomarker tools for cancer diagnosis. For instance, prostate cancer-related lncRNA PCA3 has been detected in urine samples of patients (37), and HULC in blood samples of hepatocellular cancer (38). The overexpression of IncRNAs in many cancers, and the growing evidence that IncRNAs are detected in cell-free environments, can shed more light on our understanding of the regulatory role of the IncRNAs in human diseases, and potentially can provide the therapeutic tools to target those diseases.
Acknowledgments: This publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award numbers R01AI084406 and P01AI099783-01 (KVM), and by the National Institute of Cancer of the National Institutes of Health under award numbers R01CA153124.

Received May 23, 2013; accepted June 17, 2013

References


