Oncology and Translational Medicine
DOI 10.1007/s10330-019-0346-6

June 2019, Vol. 5, No. 3, P147–P151

REVIEW ARTICLE

Regulatory mechanisms of long non-coding RNAs

Zhigang Luo (✉)

Department of Oncology, People's Hospital of Jianyang, Jianyang 644000, China

Abstract

Long non-coding RNAs (lncRNAs) belong to a large and complex family of RNAs, which play many important roles in regulating gene expression. However, the mechanism underlying the dynamic expression of lncRNAs is still not very clear. In order to identify lncRNAs and clarify the mechanisms involved, we collected basic information and highlighted the mechanisms underlying lncRNA expression and regulation. Overall, lncRNAs are regulated by several similar transcription factors and protein-coding genes. Epigenetic modification (DNA methylation and histone modification) can also downregulate lncRNA levels in tissues and cells. Moreover, lncRNAs may be degraded or cleaved via interaction with miRNAs and miRNA-associated protein complexes. Furthermore, alternative RNA splicing (AS) may play a significant role in the post-transcriptional regulation of lncRNAs.

Received: 28 March 2019
Revised: 7 May 2019
Accepted: 7 June 2019

Key words: long non-coding RNAs (lncRNAs); regulatory mechanisms; transcriptional factors; chromatin state, alternative splicing; RNA editing, microRNA (miRNA)

Eukaryotic genomes do not act as well-ordered substrates for gene transcription in a conventional manner, and are more complex than was once believed. A fine case in point is that although more than 70% of the human genome is transcribed, only approximately 2% of the transcripts produced may be translated into proteins, as revealed by the ENCODE (Encyclopedia of DNA Element) project [1–2]. Comprehensive testing and examination of RNA species in mammalian cells, as well as studies conducted on genome transcription, have revealed that the transcriptome is extremely complex. Many alternative products are generated during the biogenesis of protein-coding genes. Numerous non-coding RNA (ncRNA) transcripts that are included in the transcriptional background called "noise," are grouped into 2 major classes based on their mode of expression: housekeeping non-coding RNAs and regulatory non-coding RNAs. Transfer, ribosomal, small nuclear, and small nucleolar RNAs, which are usually constitutively expressed, are all classified as housekeeping non-coding RNAs. Regulatory non-coding RNAs include microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), small interfering RNAs (siRNAs), and long non-coding RNAs (lncRNAs). Unlike the short non-coding RNAs (< 200 nucleotide), lncRNAs range from 200 bp to several kilobases in size, with similar histone-modification profiles, exon/intron lengths and splicing signals to those of protein-coding genes [3–4]. Although lncRNAs share many structural features with protein-coding mRNAs, lncRNAs are often poorly conserved and cannot be translated into proteins. Therefore, only a few lncRNAs have been studied in depth. However, lncRNAs need to be studied further and their genome transcription functions must be well-understood.

Categories

Based on their genomic loci, lncRNAs may be divided into 5 or more categories: (1) sense; (2) antisense (if the lncRNA transcript overlaps one or more exons of another transcript); (3) intronic (if the lncRNA transcript is present within an intron of a second transcript, which sometimes may encode a protein); (4) bidirectional (if the transcripts of a lncRNA and a neighboring protein coding gene are initiated in close genomic proximity); and (5) intergenic, also called long intervening non-coding RNAs or lincRNAs (if the lncRNA transcript does not overlap exons of protein-coding and other non-coding RNA gene types).

Cellular localization

To predict the potential function of lncRNAs, its subcellular localization must be considered. The cellular

✉ Correspondence to: Zhigang Luo. Email: 358424970@qq.com

localization of lncRNAs is the same as protein-coding genes. Derrien *et al*, sequenced RNA from nuclear and cytoplasmic cellular fractions and reported that lincRNAs were mainly localized in the chromatin and the nuclei [5]. Moreover, using in situ hybridization analysis data from the Allen Mouse Brain Atlas, Mercer *et al*, identified over 800 non-coding RNAs and found that these RNAs were localized to either certain specific neuroanatomical regions of the nucleus or cytoplasm or to several foci of adult cerebellar Purkinje cells [6]. Certain well-studied lncRNAs, such as Malat 1, Xist, Miat, and Neat 1, mainly localize to the nucleus [7–10]. However, some lncRNAs showing special localization patterns, such as Gomafu (meaning "spotted pattern" in Japanese), which is associated with nuclear speckles, have been identified [8].

Origins

Unlike protein-coding genes, most long-non-coding RNA sequences are weakly conserved, and only a few exhibit sequence conservation among species. As previously mentioned, many lncRNAs have been verified as being functional, but mechanisms underlying such functions remain unclear. Therefore, it is felt that further studies on the emergence of lncRNA are needed in order to better understand their regulatory functions.

Here, we present a few evolutionary scenarios that may explain the emergence of lncRNAs. The first scenario indicates that lncRNA genes may have metamorphosized from certain protein-coding genes. For example, lncRNA Xist originated by metamorphosizing from a previously protein-coding gene, Lnx3, while including a transposable element [11]. Alternatively, lncRNAs may evolve from other lncRNAs. Duplication of a non-coding gene by retrotransposition may produce either a new functional lncRNA or a nonfunctional retropseudogene. An example of this is mouse nuclear enriched abundant transcript 2 (Neat2), which is paralogous to a mouse testis-derived lncRNA (AK019616) [9]. Another possibility is that some lncRNAs may form following insertion of transposable element sequences. This can be observed in 2 lncRNAs, BC200 (brain cytoplasmic RNA 200-nucleotide) and BC1 (brain cytoplasmic RNA1), which, despite lack of a common origin, play similar roles in translational regulation [12–14].

Regulation by transcription factors

Some studies have reported that the same transcription factors may act on lncRNAs and protein-coding genes. A recent study found that the Sp1 motif "GGGGCGGGGT" is abundant in bidirectional promoters and that a majority of lncRNAs are transcribed from these promoters [15]. Therefore, SP1 may exert a crucial effect on lncRNA expression. Another study found that among the 1,273 lncRNAs identified using RNA-seq of ribosome-depleted RNA in P493-6 human B-cells, 534 were either up- or down-regulated following MYC over-expression, and that MYC directly binds TSS in 48% lncRNAs. Thus, it may be inferred that the lncRNAs exhibiting a change in production are direct MYC targets [16]. Moreover, by combining luciferase reporter systems and chromatin immunoprecipitation (CHIP) experiments, Huarte *et al*, confirmed that P53 directly binds to the theoretical promoter of lincRNA-P21 (15 kb upstream of CDKN1A) and induces its expression [17]. Another study revealed that P53 may also increase the expression of the lincRNA PANDA, which is located closer to CDKN1A, compared with that of lincRNA-P21 [15]. Furthermore, by using high-density oligonucleotide arrays to map *in vivo* binding sites for Sp1, c-Myc, and p53 in an unbiased manner, Cawley et al., found that approximately 36% of the transcription factor binding sites (TFBS) are located within or immediately 3' to well-characterized genes, the expression levels of which are significantly correlated with those of lncRNAs [18].

In addition, other important transcription factors also play a significant role in the regulation of lncRNAs. It was revealed that 2 newly found lncRNAs that are dysregulated in fatal cardiac tissues with ventricular septal defect, possess TFBS motifs of AP-1 (activating protein-1) or SRF (serum response factor) [19]. Furthermore, there are 5 NF-Kβ binding sites in the promoter region of lncRNA AK019103, and inhibition of NF-Kβ activity significantly reduces AK019103 expression [20].

Recently, some studies have reported that in humans, mice, and zebrafish, transposable elements (TEs) are enriched in both mature lncRNA transcripts and in the vicinity of lncRNA genes, while rarely occurring in protein-coding genes. Moreover, different TE classes are enriched in these 3 species [21]. While the ERV I subclass (alpha retrovirus) is mostly enriched in human lncRNAs, the ERV 2, ERV 3, and ERV K subclasses are enriched in mice [22]. TEs can move and spread in genomes in a lineage-specific fashion and, thus, introduce regulatory elements upon chromosomal insertion. Mammalian TEs have been documented to represent several cis-regulatory sequences of protein-coding genes [23]. One recent finding indicates that TEs located in the vicinity of lncRNAs may contribute to their transcriptional regulation [21]. Although some debate exists with regard to the two scenarios, "lncRNA first" or "TE first," it appears that TEs play an important role in the expression and evolution of lncRNAs.

Epigenetic modification

In general, epigenetic modification accounts for trait variation in cellular and physiological processes that are not caused by DNA sequence changes, but by dynamic

alterations in the transcriptional potential of a cell. These modifications, including DNA methylation and histone modification, each of which alters gene expression without altering underlying DNA sequences, may or may not be heritable. Once repressor or activator proteins attach to specific regions of the DNA, the expression of downstream genes may change. These epigenetic changes may last through cell divisions for the duration of the cell, and may also last for multiple generations even though they do not involve changes in the underlying DNA sequence of the organism [24–25]. DNA methylation is stable and heritable, but histone modification (methylation or acetylation) may change quickly under the influence of the regulatory networks of cells.

Evidence demonstrates that distinct properties of lincRNAs (low expression levels and cell/tissue type specificity) are directly associated with DNA methylation and histone modification. Some studies examined the expression profile of lncRNAs in embryonic stem cells (ESCs), lineage-restricted neuronal progenitor cells (NPCs), and terminally differentiated fibroblasts, and found that the expression levels of many lncRNAs in these cell types changed. Additionally, promoter histone 3 lysine 4 trimethylation (H3K4me3) and promoter histone 3 lysine 27 trimethylation (H3K27me3) were altered. Further studies have indicated that the knockdown of the H3K27me3 methyltransferase Ezh2 may cause previously repressed lncRNAs to be re-expressed in ES [26]. Therefore there is reason to infer that lncRNAs may be subject to epigenetic regulation in a manner similar to that of protein coding genes.

It is believed that CpG dinucleotides are a remarkable reflection of the DNA methylation level. Mammalian promoters can be classified into two classes: low CpG (LCG); and high DpG (HCG) [27]. As implied by the name, genes that belong to the LCG class may be expressed at lower levels than those that belong to the HCG class. Most lncRNAs are transcribed from LCG promoters and are, thus, frequently expressed at low levels. A recent study indicated that in human sporadic insulinomas (insulin secreting PNETs), a lncRNA maternally expressed gene 3 (Meg3) was altered by hypermethylation at its promoter's CRE-sites. Moreover, in pancreatic neuroendocrine tumors (PNETs), Meg3 can be activated by the protein menin through H3K4me3 and CpG hypomethylation at the Meg3 promoter's CRE site [28]. Meg3 was also markedly reduced upon promoter hypermethylation in 4 human hepatocellular carcinoma (HCC) cell lines [29]. Moreover, 2 well known lncRNAs, XIST and HOTAIR, are targets for site-specific cytosine methylation *in vivo*, and this modification affects the protein binding ability of XIST in the least [30].

Simultaneously, histone methylation or deacetylation is involved in low lncRNA expression and even silencing. It was found that the lncRNA-LET (Low Expression in Tumor) can be repressed via hypoxia-induced histone deacetylase 3 (HDAC3) by reducing the histone acetylation-mediated modulation of the lncRNA-LET3 promoter. This may explain the downregulation of lncRNA-LET observed in colorectal cancers, hepatocellular carcinomas, and squamous cell lung carcinomas [31]. Furthermore, many large intergenic non-coding RNAs (lincRNAs) were identified in the intergenic K4-K36 domain, which not only contains a short region with histone H3K4me3, but also a longer region with histone H3 lysine 36 trimethylation (H3K36me3), indicating that the expression of these lincRNAs may be associated with histone methylation [32–33].

In conclusion, DNA methylation and histone modification determine the low and cell/tissue specific expression of lncRNAs and affect the generation of lncRNAs. The aberrant expression of lncRNAs seen in many diseases results from an abnormal chromatin state.

Post-transcriptional regulation of lncRNAs

The expression of lncRNAs following transcription is affected by other forms of post-transcriptional pressure, such as degradation by some RNA-binding proteins and the intrinsic half-lives of lncRNAs [34–35]. One study showed that the protein-RNA complex including HuR and let7i/Ago2 may reduce lncRNA HOTAIR's stability in HeLa cells [36], while another study revealed that, in renal carcinoma cells, HOTAIR may bind miR-141 in a sequence-specific way and then be cleaved in an Ago2-dependent manner [37]. Other studies have demonstrated that miRNAs often interact with lncRNAs to regulate their expression strongly.

In addition, RNA provides further means to affect lncRNAs at the post-transcriptional level. LncRNAs often fold into secondary structures or form dsRNAs with target mRNAs and act as candidate substrates for adenosine deaminase acting on RNA (ADAR) [38]. Adenosine to inosine (A-to-I) RNA editing is the most common form of editing in animals. It converts adenosine to inosine in double-stranded RNA regions via the action of ADAR proteins. Most of these specific edits occur in non-coding regions, including non-coding RNAs. A-to-I editing may influence gene expression via nuclear degradation, retention, and alternative splicing. Yang *et al*, found that similar to miRNA editing, lncRNA editing may occur through different processes, such as Tudor-SN mediated degredation [39].

Furthermore, alternative RNA splicing (AS) is a significant post-transcriptional regulatory mechanism active in long intergenic non-coding RNAs (lincRNAs). Several studies reported that annotated human lincRNAs with multiple exons are alternatively spliced

[40–43]. A recent study, compared sequence evolution and biological features of single-exonic lincRNAs and multi-exonic lincRNAs (SELs and MELs, respectively) present in hominoids or conserved in primates, and found that SELs and MELs differed in primary sequence evolution, exon/transcript length, expression breadth and proximity to the nearest coding gene. Thus, SELs and MELs may represent 2 biologically distinct gene groups. Notably, splicing by deletion appears to be disfavored in SELs, compared with MELs. These findings suggest that AS may be associated with the expression levels and functionality of lincRNAs [44].

Conclusion

As an important factor affecting the regulatory network of gene expression, miRNAs have been studied in detail in recent times. However, the exploration of lncRNAs has just begun. To date, several studies have revealed that lncRNAs may play important roles in the transcriptional regulation of some genes [45–48] as well as in epigenetics [49–51] and are closely associated with human diseases [14, 52–55]. Although the biological function and mechanisms underlying lncRNA regulation remain unclear, growing evidence suggests that lncRNA investigation in human cells has a bright and promising future.

References

1. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet, 2014, 15: 7–21.
2. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol, 2012, 9: 703–719.
3. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, *et al*. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 2007, 447: 799–816.
4. Carninci P, Kasukawa T, Katayama S, *et al*. The transcriptional landscape of the mammalian genome. Science, 2005, 309: 1559–1563.
5. Derrien T, Johnson R, Bussotti G, *et al*. The GENCODE v7 catalog of human long non-coding RNAs: analysis of their gene structure, evolution, and expression. Genome Res, 2012, 22: 1775–1789.
6. Mercer TR, Dinger ME, Sunkin SM, *et al*. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A, 2008, 105: 716–721.
7. Clemson CM, Hutchinson JN, Sara SA, *et al*. An architectural role for a nuclear non-coding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Molecular Cell, 2009, 33: 717–726.
8. Sone M, Hayashi T, Tarui H, *et al*. The mRNA-like non-coding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J Cell Sci, 2007, 120: 2498–2506.
9. Hutchinson JN, Ensminger AW, Clemson CM, *et al*. A screen for nuclear transcripts identifies two linked non-coding RNAs associated with SC35 splicing domains. BMC Genomics, 2007, 8: 39.
10. Brown CJ, Hendrich BD, Rupert JL, *et al*. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell, 1992, 71: 527–542.
11. Elisaphenko EA, Kolesnikov NN, Shevchenko AI, *et al*. A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. PLoS One, 2008, 3: e2521.
12. Conley AB, Miller WJ, Jordan IK. Human cis natural antisense transcripts initiated by transposable elements. Trends Genet, 2008, 24: 53–56.
13. Cao X, Yeo G, Muotri AR, *et al*. Non-coding RNAs in the mammalian central nervous system. Annu Rev Neurosci, 2006, 29: 77–103.
14. Zalfa F, Giorgi M, Primerano B, *et al*. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. Cell, 2003, 112: 317–327.
15. Uesaka M, Nishimura O, Go Y, *et al*. Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals. BMC Genomics, 2014, 15: 35.
16. Hart JR, Roberts TC, Weinberg MS, *et al*. MYC regulates the non-coding transcriptome. Oncotarget, 2014, 5: 12543–12554.
17. Huarte M, Guttman M, Feldser D, *et al*. A large intergenic non-coding RNA induced by p53 mediates global gene repression in the p53 response. Cell, 2010, 142: 409–419.
18. Cawley S, Bekiranov S, Ng HH, *et al*. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell, 2004, 116: 499–509.
19. Song G, Shen Y, Zhu J, *et al*. Integrated analysis of dysregulated lncRNA expression in fetal cardiac tissues with ventricular septal defect. PLoS One, 2013, 8: e77492.
20. Wan G, Hu X, Liu Y, *et al*. A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. EMBO J, 2013, 32: 2833–2847.
21. Kapusta A, Kronenberg Z, Lynch VJ, *et al*. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet, 2013, 9: e1003470.
22. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, *et al*. Initial sequencing and comparative analysis of the mouse genome. Nature, 2002, 420: 520–562.
23. Bourque G, Leong B, Vega VB, *et al*. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res, 2008, 18: 1752–1762.
24. Ledford H. Language: Disputed definitions. Nature, 2008, 455: 1023–1028.
25. Bird A. Perceptions of epigenetics. Nature, 2007, 447: 396–398.
26. Wu SC, Kallin EM, Zhang Y. Role of H3K27 methylation in the regulation of lncRNA expression. Cell Res, 2010, 20: 1109–1116.
27. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A, 2006, 103: 1412–1417.
28. Modali SD, Parekh VI, Kebebew E, *et al*. Epigenetic regulation of the lncRNA MEG3 and its target c-MET in pancreatic neuroendocrine tumors. Mol Endocrinol, 2015, 29: 224–237.
29. Braconi C, Kogure T, Valeri N, *et al*. microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. Oncogene, 2011, 30: 4750–4756.
30. Amort T, Soulière MF, Wille A, *et al*. Long non-coding RNAs as targets for cytosine methylation. RNA Biol, 2013, 10: 1003–1008.
31. Yang F, Huo XS, Yuan SX, *et al*. Repression of the long non-coding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. Mol Cell, 2013, 49: 1083–1096.
32. Khalil AM, Guttman M, Huarte M, *et al*. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes

and affect gene expression. Proc Natl Acad Sci U S A, 2009, 106: 11667–11672.
33. Mikkelsen TS, Ku M, Jaffe DB, *et al*. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature, 2007, 448: 553–560.
34. Yoon JH, Abdelmohsen K, Srikantan S, *et al*. LincRNA-p21 suppresses target mRNA translation. Mol Cell, 2012, 47: 648–655.
35. Clark MB, Johnston RL, Inostroza-Ponta M, *et al*. Genome-wide analysis of long non-coding RNA stability. Genome Res, 2012, 22: 885–898.
36. Yoon JH, Abdelmohsen K, Kim J, *et al*. Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. Nat Commun, 2013, 4: 2939.
37. Chiyomaru T, Fukuhara S, Saini S, Majid S, *et al*. Long non-coding RNA HOTAIR is targeted and regulated by miR-141 in human cancer cells. J Biol Chem, 2014, 289: 12550–12565.
38. Mercer TR, Mattick JS. Structure and function of long non-coding RNAs in epigenetic regulation. Nat Struct Mol Biol, 2013, 20: 300–307.
39. Yang Y, Zhou X, Jin Y. ADAR-mediated RNA editing in non-coding RNA sequences. Sci China Life Sci, 2013, 56: 944–952.
40. Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long non-coding RNAs in six mammals. Genome Research, 2014, 24: 616–628.
41. Necsulea A, Soumillon M, Warnefors M, *et al*. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature, 2014, 505: 635–640.
42. Novikova IV, Hennelly SP, Tung CS *et al*. Rise of the RNA machines: exploring the structure of long non-coding RNAs. J Mol Biol, 2013, 425: 3731–3746.
43. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet, 2010, 11: 345–355.
44. Chen FC, Pan CL, Lin HY. Functional implications of RNA splicing for human long intergenic non-coding RNAs. Evol Bioinform Online, 2014, 10: 219–228.
45. Hirota K, Miyoshi T, Kugou K, *et al*. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature, 2008, 456: 130–134.
46. Conley AB, Miller WJ, Jordan IK. Human cis natural antisense transcripts initiated by transposable elements. Trends Genet, 2008, 24: 53–56.
47. Ponjavic J, Ponting CP. The long and the short of RNA maps. Bioessays, 2007, 29: 1077–1080.
48. Katayama S, Tomaru Y, Kasukawa T, *et al*. Antisense transcription in the mammalian transcriptome. Science, 2005, 309: 1564–1566.
49. Peters J, Robson JE. Imprinted non-coding RNAs. Mamm Genome, 2008, 19: 493–502.
50. Cai X, Cullen BR. The imprinted H19 non-coding RNA is a primary microRNA precursor. RNA, 2007, 13: 313–316.
51. Chaumeil J, Le Baccon P, Wutz A, *et al*. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev, 2006, 20: 2223–2237.
52. Faghihi MA, Modarresi F, Khalil AM, *et al*. Expression of a non-coding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med, 2008, 14: 723–730.
53. Broadbent HM, Peden JF, Lorkowski S, *et al*. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol Genet, 2008, 17: 806–814.
54. Mus E, Hof PR, Tiedge H. Dendritic BC200 RNA in aging and in Alzheimer's disease. Proc Natl Acad Sci U S A, 2007, 104: 10679–10684.
55. Moseley ML, Zu T, Ikeda Y, *et al*. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat Genet, 2006, 38: 758–769.

DOI 10.1007/s10330-019-0346-6
Cite this article as: Luo ZG. Regulatory mechanisms of long non-coding RNAs. Oncol Transl Med, 2019, 5: 147–151.