Long non-coding RNAs and human disease

Lorna W. Harries
RNA-mediated Mechanisms of Disease Group, Institute of Biomedical and Clinical Sciences, Peninsula College of Medicine and Dentistry, Barrack Road, Exeter EX2 5DW, U.K.

Abstract
The central dogma of molecular biology states that DNA is transcribed into RNA, which in turn is translated into proteins. We now know, however, that as much as 50% of the transcriptome has no protein-coding potential, but rather represents an important class of regulatory molecules responsible for the fine-tuning of gene expression. Although the role of small regulatory RNAs [miRNAs and siRNAs (small interfering RNA)] is well defined, another much less characterized category of non-coding transcripts exists, namely lncRNAs (long non-coding RNAs). Pervasively expressed by eukaryotic genomes, lncRNAs can be kilobases long and regulate their targets by influencing the epigenetic control, chromatin status, mRNA processing or translation capacity of their targets. In the present review, I outline the potential mechanisms of action of lncRNAs, the cellular processes that have been associated with them, and also explore some of the emerging evidence for their involvement in common human disease.

Mechanisms of lncRNA regulation
There are several proposed mechanisms of action for lncRNAs (Figure 2), which bring plasticity, adaptability and reactivity to genomic architecture and fine control over gene expression. In addition to the mechanisms outlined below, lncRNAs have also been reported to be subject to other mechanisms, including RNA editing, RNA interference, RNA masking, transcriptional interference and protein kinase R activation in some cases [9,10].

Epigenetic regulation
lncRNAs may act as scaffold molecules, to deliver regulatory proteins to loci where they are required. Examples of this type of lncRNA are ANRIL and HOTAIR. The ANRIL antisense transcript is coded for on the opposite strand of the CDKN2A/CDKN2B loci, and causes its effects by binding to and recruiting the CBX7 (chromobox 7) subunit of the PRC1 (Polycomb repressive complex 1) and PRC2. These complexes serve to direct H3K27me (methylation of histone H3 at Lys-27) at the target loci, resulting in the silencing of sense transcripts expressed from this genomic region, but from the opposite strand (these are often termed natural antisense transcripts), whereas trans-lncRNAs share incomplete homology with their targets and arise from distant regions of the genome [8]. The orientation of cis-regulatory transcripts to their targets can be 5′ to 5′ (head-to-head; Figure 1a), 3′ to 3′ (tail-to-tail; Figure 1B) or fully overlapping, with one gene contained within the region coding the other (Figure 1C). Trans-regulatory transcripts are usually non-overlapping, since they are from distinct genomic regions.
Figure 1 | The different conformations of lncRNA: target duplexes

(A) A head-to-head conformation, whereby the 5′-end of the lncRNA associates with the 5′-end of the target. (B) A tail-to-tail conformation of conformation, whereby the 3′-end of the lncRNA associates with the 3′-end of the target. (C) An overlapping conformation, whereby the entire sequence of the lncRNA is contained within the sequence of the target.

Figure 2 | Mechanisms of lncRNA action

Some of the potential mechanism underlying the activity of lncRNAs. (A) The nature of example lncRNAs. (B) The lncRNA. (C) The corresponding target gene. (D) The mode of action for the lncRNA. (E) The consequences of lncRNA regulation of that target gene.

element-1-silencing transcription factor) complex. Again, this brings about specific alterations in the methylation status and the nature of the chromatin surrounding the HOTAIR targets in the HOXD gene cluster [13]. HOXD genes such as HOXD13 direct morphogenesis in all multicellular organisms, and disruption to their expression has been associated with breast cancer [13] and developmental disorders [14].

Regulation of alternative splicing

Another regulatory mechanism attributed to lncRNAs involves modification of alternative splicing. This is exemplified by the lncRNA MALAT1. MALAT1 interacts with the SR (serine/arginine)-rich splicing regulatory proteins, which dictate alternative splicing patterns and are regulated by alterations to their phosphorylation status [15]. The interaction of MALAT1 with these factors results in their relocation to the splicing speckles (the site of mRNA processing) in the nucleus [16,17], together with the modification of their phosphorylation state. MALAT1 is thus a key regulator of alternative splicing events, which are important moderators of cellular plasticity and adaptability. Perturbations to the expression or activity of MALAT1 therefore have significant implications for global regulation of alternative splicing.

Control of translation

lncRNAs can also regulate gene activity by controlling translational control or by regulation of mRNA stability, as demonstrated by the antisense RNA BACE1AS. BACE1AS interacts with the β-site APP (amyloid precursor protein) cleaving enzyme 1 (BACE1) transcript, which is a crucial player in AD (Alzheimer’s disease) pathology. The specific interaction between the BACE1 and BACE1S transcripts increases the stability of the BACE1 transcript, and thus increases the translation and the abundance of the BACE1 gene product [18]. lncRNAs can also decrease translation efficiencies. The NOS pseudogene pseudo-NOS is an antisense transcript that has been demonstrated to bind to, and to regulate the expression of the nNOS (neuronal nitric oxide synthase) gene. The lncRNA pseudo-NOS acts to influence the association of the ribosome with the nNOS- pseudo-NOS duplex, repressing translation of this target [19].

Competition for binding sites

Some lncRNAs do not work by direct antisense regulation of their target genes. The Gas5 lncRNA, for example, works by binding to the DNA-binding domain of the glucocorticoid receptor, thus competing with and modifying the expression of target genes containing genuine glucocorticoid response elements [20]. Further mechanisms of lncRNA function are further reviewed in [21].

lncRNAs and normal function

lncRNAs have been implicated in numerous normal physiological processes at all stages of life, from early embryogenesis and cellular cell fate determination to physiological homoeostasis of entire organisms. Examples of their roles in three key cellular processes are described below.

Embryonic development

lncRNAs play a pivotal role in development, acting from the level of the embryonic stem cell, where they are involved in control of pluripotency; a recent study identified two lncRNAs, AK028326 (Oct4-activated) and AK141205 (Nanog-repressed) involved in an auto-regulatory loop with the key embryonic stem cell transcription factors OCT4 and NANOG. Overexpression and gene-silencing studies subsequently demonstrated that the expression of these lncRNAs was associated with alterations in cellular lineage-specific gene expression and a change in pluripotent potential of the cells in question [22]. Similarly, the lncRNA MISTRAL...
(Mira), which targets HOX6A and HOX7A, has been shown to influence the expression of these genes by recruitment of the ‘MLL (mixed lineage leukaemia)’ gene, involved in commitment to the haematopoietic lineage [23].

Control of cell cycle
In IncRNAs are also known to be involved in the control of the cell cycle; the natural antisense transcript ANRIL is a key regulator of three separate tumour suppressor genes p16INK4a, p14ARF and p15INK4b, which are all expressed from the CDKN2A/B gene cluster [24]; p16INK4a and p15INK4b are important inhibitors of the cyclin-dependent kinase 4, whereas p14ARF acts to stabilize p53 by recruitment of MDM1 (murine double minute 1) [25]. An important role for IncRNAs in the coupling of DNA damage with cellular apoptosis has also been reported; DNA damage was shown to induce five IncRNAs from the promoter of the cell cycle regulator CDKN1A, one of which, named PANDA, was shown to interact with the transcription factor NF-YA to down-regulate the expression of genes involved in promotion of apoptosis [26].

Dosage compensation and chromosomal imprinting
Another key role for IncRNAs in normal physiology is their involvement in chromosomal dosage compensation. In females, each cell contains two copies of the X chromosome, one of which must be deactivated to ensure correct dosage of the genes housed by this chromosome in a process called X inactivation. X inactivation is controlled by an IncRNA called Xist. Xist is one of the first genes expressed following fertilization and acts to direct the PRC1 and PRC2, leading to histone modifications and silencing of all the genes on the targeted chromosome [27]. The recruitment of Xist to only one of the chromosomes is controlled by another series of IncRNAs, which includes the antisense IncRNA Tsix, which represses the activity of Xist on the active X, and Jpx which activates Xist on the silent X [28].

IncRNAs also have a pivotal role in control of imprinting, where one of the parental alleles is epigenetically silenced. Probably the best example of this is the AIR non-coding RNA, which regulates the genomic imprinting of a chromosomal region containing the IGF2R/SLC12A2/SLC22A3 loci. AIR is a bidirectional silencer, and been shown to have a repressive effect on the paternal expression of genes in this region [29].

IncRNAs and human disease
Given the ubiquitous nature of IncRNA expression and their key roles in many physiological processes involving global regulation of the genome, it is unsurprising that they are involved in the aetiology of many human diseases.

Cancer
IncRNAs have a pivotal role in the control of the cell cycle, apoptosis and tumour suppression. The IncRNA ANRIL regulates three separate tumour suppressor genes p16INK4a, p14ARF and p15INK4b, important negative regulators of cell cycle [30]. Disruptions to the expression of ANRIL have accordingly been associated with the development of several cancer types, including neuroblastoma [24], acute lymphocytic leukaemia [31], melanoma [30] and prostate [11]. Overexpression of the HOTAIR transcript, a cis-lncRNA associated with the HOXD gene cluster, has been associated with hepatocellular carcinoma [32], colorectal cancer [33] and breast cancer [13] by deregulation of HOXD cluster genes. Ovarian and breast tumours have also been associated with the expression of the LSINCT5 IncRNA; this transcript acts to target several other transcripts, including the antisense RNA NEAT-1 and the PSPC1 gene, which codes for a splicing regulatory factor [34]. IncRNA-associated disruption to alternative splicing has also been reported in non-small-cell lung cancer by virtue of overexpression of MALAT1 [35].

Metabolic disease
Although little is known about the role of IncRNAs in endocrine disease, several genes that are important moderators of metabolism and endocrine function have reported IncRNAs. An antisense transcript to the PINK1 [PTEN (phosphatase and tensin homologue deleted on chromosome 10)-induced putative kinase 1], termed naPINK1 has recently been described [36]. As the name suggests, PINK1 is induced by PTEN, which is an important inhibitor of insulin signalling. PINK1 depletion has been associated with diabetes status, impaired glucose uptake in neuronal cell lines and with mitochondrial gene expression in adipocytes [37], raising the possibility that disruption to naPINK1 may impact on glucose metabolism. Similarly, the H19/IGF2 and thyroid growth receptor α2 (ERBa2) loci harbour known antisense transcripts [38,39] which have the potential to regulate their endocrine and metabolic function. IncRNAs have also been implicated in the regulation of lipid metabolism genes. The Δ3-desaturase (FADS1) and steroidogenic acute regulatory protein (STAR) genes have reported IncRNAs [40,41]. The expression of FADS, and its IncRNA, reverse D5-desaturase, were found to be reciprocally regulated by the dietary fat content in animal models [40]. IncRNAs have also been implicated in appetite control; an IncRNA to the human ghrelin (GRHL) gene, which promotes food-seeking behaviour, has recently been identified [42]. These findings raise the possibility that deregulation of IncRNA expression may also have implications for obesity.

Neurodegenerative and psychiatric diseases
The BACE1 antisense transcript, BACE1AS, has been implicated in the aetiology of AD [6]. Some of the features of AD are because of the accumulation within the brain of β-amyloid plaques. The BACE1 gene, an integral membrane peptidase A1 glycoprotein, plays a pivotal role in the accumulation of β-amyloid plaques. BACE1 is one of two peptidases that carry out the initial proteolytic cleavage of APP, allowing it to accumulate in the brain. BACE1AS levels
were found to be higher in human subjects with AD, and also in BACE1 transgenic mouse models of AD [6].

lncRNAs have also been suggested to be involved in psychiatric disorders. Disruption of the ‘disrupted in schizophrenia-1’ DISC1 locus has been linked with the development of schizophrenia, schizoaffective disorder, bipolar disorder, major depression and autistic spectrum disorders [43]. DISC1 is regulated by its lncRNA, DISC2, which may also represent an excellent candidate for susceptibility to these disorders. Schizophrenia spectrum disorders and AD have also been linked with the rheelin (RELN) gene and its antisense transcript HARI [44].

Cardiovascular disease, hypertension and stroke

lncRNAs have the potential to influence cardiovascular disease and hypertension. Genetic variants that affect the expression of the ANRIL transcript have been correlated with stroke risk and recurrence in a large prospective study [45]. A role for lncRNAs in hypertension also suggested that seven blood pressure candidate genes (ADD3, NPPA, ATP1A1, NPR2, CYP17A1, ACSM3 and SLC14A2) were associated with cis-lncRNA transcripts [46]. The NPPA (natriuretic peptide precursor A) gene product is usually expressed in only fetal atrial and ventricular myocardium, but has been shown to be reactivated in the ventricular myocardium of patients exhibiting hypertrophy and heart failure [47], and is considered to be a marker for heart disease. The NPPA-AS lncRNA has been shown to be a modulator of the alternative splicing of the NPPA gene. This lncRNA thus has potential to be involved in cardiovascular disease [46].

Immune dysfunction and auto-immunity

An important role for lncRNA in control of innate immune signalling has been suggested by the observation that approximately 500 lncRNAs are differentially expressed during the immune response to virus infection in a study involving four separate mouse strains [48]. The non-translated RNA ‘Gas5’ (growth-arrest-specific 5) transcript, activated by cellular stress, targets a diverse group of genes through the glucocorticoid receptor and is an important regulator of cellular apoptosis [20]. The Gas5 transcript has been linked with increased susceptibility to systemic lupus erythematosus in mouse models, presumably as a result of its effect on the immunosuppressant role of glucocorticoids [20]. Other studies have reported an association of the antisense RNA Heg with CD14 levels and thyroid auto-antibodies in patients with untreated Graves’ disease [49].

Conclusions

lncRNAs form a significant part of the eukaryotic transcriptome, which regulate the expression of up to 70% of genes. They play a crucial role in global processes such as epigenetic regulation, chromatin remodelling and alternative mRNA processing, and are thus intimately involved in the control of key physiological processes. Their involvement in many aspects of higher function indicates that they may represent a new and exciting arena to exploit for future disease therapies.

Acknowledgement

I thank Alexander Fellows for his help in preparing the Figures for this review.

References


