REVIEW ARTICLE

LncRNAs: emerging players in gene regulation and disease pathogenesis

MINA KAZEMZADEH[1], REZA SAFARALIZADEH[1]* and AYLA VALINEZHAD ORANG[1,2]

[1]*Faculty of Natural Sciences, Department of Animal Biology, The University of Tabriz, Tabriz 51666-16471, Iran*
[2]*Flinders Centre for Innovation in Cancer, School of Medicine, Flinders University, Flinders Medical Centre, Adelaide 5042, Australia*

Abstract

The advent of next-generation sequencing has demonstrated that eukaryotic genomes are extremely complex than what were previously thought. Recent studies revealed that in addition to protein-coding genes, nonprotein-coding genes have allocated a large fraction of the genome. Long noncoding RNA (lncRNA) genes are classified as nonprotein-coding genes, serving as a molecular signal, decoy, guide and scaffold. They were suggested to play important roles in chromatin states, epigenetic and posttranscriptional regulation of genes. Aberrant expression of lncRNAs and changes in their structure are associated with a wide spectrum of diseases ranging from different types of cancer and neurodegeneration to α-thalassaemia. The purpose of this study was to summarize the current progress in understanding the genomic bases and origin of lncRNAs. Moreover, this study focusses on the diverse functions of lncRNAs in normal cells as well as various types of disease to illustrate the potential impacts of lncRNAs on diverse biological processes and their therapeutic significance.

[Kazemzadeh M., Safaralizadeh R. and Orang A. V. 2015 LncRNAs: emerging players in gene regulation and disease pathogenesis. *J. Genet.* **94**, 771–784]

Introduction

After completion of the human genome in 2001, it was surprisingly discovered that only 1.06% of the human genome encodes protein-coding genes (Chodroff *et al*. 2010; Hauptman and Glavac 2013), accounting for about 20,000 protein coding genes (Sana *et al*. 2012). In fact, large-scale analyses of genome revealed an intriguing paradox between the organism complexity and protein-coding gene content (Taft *et al.* 2007). Moreover, these studies proved that the number of protein-coding genes in highly complex organisms such as human are approximately the same as what they are in simpler organisms including *Drosophila melanogaster* and *Caenorhabditis elegans* (Casa and Gabellini 2011; Wright and Bruford 2011; Nie *et al.* 2012). Therefore, it is very challenging to imagine this similarity and consider the rest of the human genome as a junk DNA (Nie *et al.* 2012).

The idea that 'junk DNA' might significantly contribute to the higher eukaryotic sophistication resulted in the establishment of the Encyclopedia of DNA Elements (ENCODE) Consortium in 2003, aiming to identify all the functional elements in the human genome (Wilusz *et al.* 2009; Wright and Bruford 2011). The ENCODE project reported that about 90% of the genome is transcribed as nonprotein-coding RNAs (ncRNAs), hence, this so-called 'junk DNA' might not be junk hereafter (Gibb *et al.* 2011a; Sana *et al.* 2012; Vennin *et al.* 2013).

Noncoding RNA (ncRNA) genes are transcribed into RNA molecules, while they are not capable of being translated into proteins. Besides, they can play structural and functional roles in cells (Gibb *et al.* 2011b; Sana *et al.* 2012). Their classification varies depending on their size and function. According to their size, they are classified at least into two major classes: (i) small ncRNAs, which are shorter than 200 nt including miRNAs, piRNA and other noncoding transcripts, and (ii) Long noncoding RNAs (lncRNAs), with a size range 200 nt to 100 kb (Gibb *et al.* 2011a; Hauptman and Glavač 2013; Vennin *et al.* 2013; Zhang *et al.* 2013). NcRNAs are alternatively divided into: (i) housekeeping ncRNAs, expressed constitutively in cells (Gutschner and Diederichs 2012), consisting of tRNAs, rRNAs, snRNAs and spliceosomal RNAs (Knauss and Sun 2013), and (ii) regulatory ncRNAs, expressed in certain developmental stages, and tissues (Brosnan and Voinnet 2009; Tano and Akimitsu 2012), such as lncRNA, mRNA, siRNA and piRNA (Wu *et al*. 2013). In the following, we will focus on long ncRNAs which represent a novel class of regulatory molecules.

*For correspondence. E-mail: safaralizadeh@tabrizu.ac.ir.

Keywords. long noncoding RNA; noncoding gene; gene regulation; junk DNA.

The GENCODE consortium within the ENCODE project has interpreted an extensive collection of human lncRNAs. This collection includes 9277 manually annotated lncRNA gene loci, producing 14,880 lncRNA transcripts (Derrien *et al*. 2011, 2012). Based on this project, lncRNA genes are the biggest class of ncRNA genes that have been annotated in humans so far (Kornienko *et al.* 2013; Chen *et al.* 2013). Although, functional lncRNAs are increasingly being uncovered, the biological function of most lncRNAs are poorly understood (Tani *et al.* 2012). Accordingly, it is obvious that bioinformatics approaches have identified general information on lncRNAs' feature (Da Sacco *et al.* 2012).

To exploit practical implications of lncRNAs as gene regulators, understand their molecular and cellular pathways, and utilize their therapeutic potentials, it is necessary to have an in-depth understanding of lncRNA location, origin and biogenesis. To this end, in this article, we have provided lncRNAs' features, which are mostly supplied by bioinformatics approaches and some by experimental studies. Also, we will review their functionalities in normal cellular pathways and diseases.

lncRNAs

LncRNAs consist of a large subset of regulatory noncoding RNAs. Having characteristics such as low conservation, low expression levels and very complex and diverse gene regulation mechanisms, lncRNAs have distinct features from other regulatory ncRNAs (Zhang *et al.* 2013). LncRNAs' genes have been found to be located in both eukaryotic and prokaryotic genomes (Niazi and Valadkhan 2012). For instance, ornate large extremophilic (OLE), a lncRNA of human pathogen bacteria, is localized in bacterial cell membranes and has a high expression level. Its expression level increases by 7-fold in response to alcohol-induced stress, hence, OLE protects bacterial cells from alcohol stress via cell membrane stability (Wallace *et al.* 2012). Moreover, three mitochondrial lncRNAs have been identified recently in human genome known as lncND5, lncND6 and lncCytb which regulate three mitochondrial genes of ND5, ND6 and Cytb, respectively (Rackham *et al.* 2011).

While RNA Pol II is responsible for the expression of most lncRNAs, some other lncRNAs are transcribed by RNA Pol III (Nie *et al.* 2012). Molecular and biochemical studies demonstrated that most lncRNA genes possess K4–K36 domain, guiding RNA polymerase II to their residing gene locus (Guttman *et al.* 2009). Further, recent studies reported that lncRNA transcripts have 5′-capped, and poly(A) tail. About 98% of lncRNA transcripts are spliced and more than 25% of them undergo alternative splicing, making them similar to protein coding RNAs (Derrien *et al.* 2012; Cheng *et al.* 2013; Ng *et al.* 2013). Nonetheless, compared to mRNAs, most lncRNAs have fewer introns, with an average of one intron per lncRNA versus seven introns per protein-coding RNA. This results in significantly different unspliced transcript length between lncRNA and protein-coding RNA with a median length of 6 kb and 24 kb, respectively. On the other hand, the GC content of lncRNAs' introns is almost similar to that of the protein-coding RNAs and about half of the lncRNAs (42%) are reported to possess two exons with lower GC content (Niazi and Valadkhan 2012; Derrien *et al.* 2012).

LncRNA nomenclature

Some lncRNAs are named according to their function and expression pattern such as XIST. This kind of nomenclature is uncommon since the majority of lncRNA's functions have not been proven yet. To solve this problem, HUGO Gene Nomenclature Committee currently uses several outlines to name unknown lncRNAs. HGNC approved the symbol of '-AS' for the lncRNA gene on the opposite strand of the protein-coding gene, like 'BOK-AS1' for 'BOK antisense RNA 1', located in the opposite strand of BOK. Further, the symbol of '-IT' is used to 'intronic lncRNA', like 'MAGI2-IT1' for 'MAGI2 intronic lncRNA transcript 1. The prefix 'LINC' is followed by numbers used for intergenic lncRNAs (Wright and Bruford 2011; Volders *et al.* 2013).

LncRNA stability and turnover

A genomewide analysis using custom microarrays examined the stability of about 7200 lncRNAs together with almost 20,000 protein-coding transcripts. It was revealed that lncRNAs have a shorter turnover with the median of 3.5 h compared to 5.1 h for mRNAs. Indeed, lncRNAs' half-lives ranged from 30 min to 48 h and protein-coding transcripts half-lives ranged from 2 h to 12 h (Niazi and Valadkhan 2012).

Certainly, the wide range of lncRNA transcript stability reflects the complexity of lncRNAs metabolism and also their widespread functions (Clark *et al.* 2012; Wu *et al.* 2013). Protein-coding genes with enzymatic and housekeeping functions, which do not require dynamic regulation were found to be more stable. However, those with regulatory function have low stability since regulatory elements are needed for a limited period of time. Likewise, regulatory lncRNAs have shorter life span than housekeeping lncRNAs (Clark *et al.* 2012; Tani *et al.* 2012; Wu *et al.* 2013). Given that most lncRNAs regulate chromatin remodelling and gene expression in response to the external stimuli, instability of these regulatory lncRNAs is central to their rapid dynamic regulation (Clark *et al.* 2012; Tani *et al.* 2012).

Other factors such as genomic location, subcellular localization and splicing have important impacts on stability of lncRNA products; therefore, spliced lncRNAs are more stable than single exon transcripts. Those transcripts that are localized in nucleus display a lower stability in line with their regulatory functions (Clark *et al.* 2012).

LncRNA potential of protein coding

Experimental and computational methods have been used to determine protein coding capacity of lncRNAs (Ma *et al.* 2012). Different computational algorithms have been employed to predict lncRNAs protein coding potential and to determine whether the length of lncRNA's ORFs is long enough to encode functional polypeptides. Another objective of the computational approaches was to determine the evolutionary conservation of lncRNAs (Rinn and Chang 2012). In these studies, RNAs shorter than 300 nt belong to ncRNA group, since most of the eukaryotic protein-coding RNAs contain ORFs longer than 300 nt or 100 amino acids. Briefly, studies have revealed significant differences between lncRNAs and mRNAs in their ORF length, with median of 250 nt versus 1200 nt, respectively (Niazi and Valadkhan 2012). Indeed, analyses have uncovered that start and stop codons in lncRNAs are distributed randomly among their genes (Ma *et al.* 2012).

However, functional polypeptides that are shorter than 300 nt have been found in eukaryotic organisms. Also, some lncRNAs contains ORFs longer than 300 nt (Ma *et al.* 2012). Thus, other criteria must be employed to define the protein coding capacity of these lncRNAs. Therefore, it could be practical to compare the ORFs of protein-coding transcripts with those of the lncRNA ORFs that are longer than 300 nt. The codon content of the lncRNA ORFs is similar to the 3′UTRs of the protein-coding RNAs and to the intron codon alignment. These similarities indicate that lncRNAs and protein-coding RNAs have undergone evolutionary pressures through different pathways (Niazi and Valadkhan 2012). Further, it was shown that the ORFs of the longest lncRNAs lack the sequence context at -3 to $+4$ of the start codon indicating its translational efficiency (Niazi and Valadkhan 2012).

LncRNA ORFs were also examined for detecting known protein domains, which could be a sign for the protein-coding sequences (Volders *et al.* 2013). The details are reported in the LNCipedia, a database for human long noncoding RNAs (http://www.lncipedia.org).

Experimental methods such as ribosome profiling was used to identify the association of lncRNAs with ribosome as well as their protein-coding potentials (Rinn and Chang 2012). Since all of the susceptible translatable lncRNAs, which are associated with the ribosome, lack the evolutionary conservation of their proposed coding regions, such methods have to be employed precisely (Rinn and Chang 2012). As a result, any interaction between ribosome and lncRNAs does not necessarily result in translation. Further, it was shown that some of the ribosome-associated lncRNAs modulate the translation of protein-coding genes (Yoon *et al.* 2012).

Genomic and cellular location of lncRNAs

LncRNAs can be located in either DNA strands or in both (Wang *et al.* 2011a). Based on genomic distribution with respect to protein-coding genes, lncRNAs can be classified into broad categories; including, intergenic, intragenic, overlapping and bidirectional lncRNA (Da Sacco *et al.* 2012; Hauptman and Glavač 2013; Shahandeh 2013). Additionally, these categories can be further classified into several subtypes. Prepared by GENCODE project, among 14,880 lncRNAs, 9518 belong to intergenic lncRNA and 5362 are from intragenic lncRNAs (Derrien *et al.* 2012).

Intergenic lncRNAs are localized in the gene deserts, at least 5 kb away from protein-coding genes (Shahandeh 2013; Hauptman and Glavač 2013) and are also known as 'large or long intergenic RNAs' (lincRNAs) (Knauss and Sun 2013). Enhancer RNA (eRNA) is a subclass of lincRNAs which is transcribed from enhancer and regulates the expression of its adjacent protein-coding genes (Wang and Chang 2011; Prensner and Chinnaiyan 2011). Besides, many other lincRNAs are transcribed from the gene-free chromosomal regions and regulate gene expression, both in *cis* and *trans* (Knauss and Sun 2013).

Intragenic lncRNAs reside within a protein-coding gene and can be further sub-classified into four groups: (i) sense exonic lncRNAs, which overlap with one or more exons of protein-coding genes at the same strand. (ii) Antisense exonic lncRNAs also known as 'natural antisense transcripts' (NATs) which overlap with antisense strand of a protein-coding gene. They can be transcribed from a host gene or from a different genomic locus with the same sequence complementarity (Da Sacco *et al.* 2012; Hauptman and Glavač 2013; Papait *et al.* 2013). NATs regulate gene expression through multiple mechanisms, for example through forming endo-siRNAs (Knauss and Sun 2013). (iii) intronic sense and antisense lncRNAs are derived from an intron of a protein-coding transcript from the same or opposite strand. (iv) 3′ UTRs and 5′ UTRs associated RNAs or uaRNAs are derived from untranslated regions of the protein-coding transcripts (Nie *et al.* 2012; Wu *et al.* 2013).

Overlapping lncRNA transcripts contain a protein-coding gene within an intron on the same strand (Derrien *et al.* 2012; Cheng *et al.* 2013). These lncRNAs may regulate the promoter of the overlapping genes. To date, 52 overlapping sense lncRNAs have been identified in the human genome (Knauss and Sun 2013).

Bidirectional or divergent lncRNAs are located in the opposite strand of an adjacent promoter and reside in approximately less than 1000 bp of a protein-coding gene (Wu *et al.* 2013; Knauss and Sun 2013; Shahandeh 2013). These lncRNAs usually have the same CpG island promoter as adjacent protein-coding genes and also, they regulate the expression of their neighbouring protein-coding genes (Sun *et al.* 2013). Figure 1 provides a schematic representation of lncRNAs locus and transcription direction.

By identifying the subcellular localization of lncRNAs, scientists could be partially able to determine their functions (Lipovich *et al.* 2010). Analysis of the ENCODE RNA-seq data for nucleus and cytoplasm revealed that nucleus is highly enriched with lncRNA and that the lncRNAs are

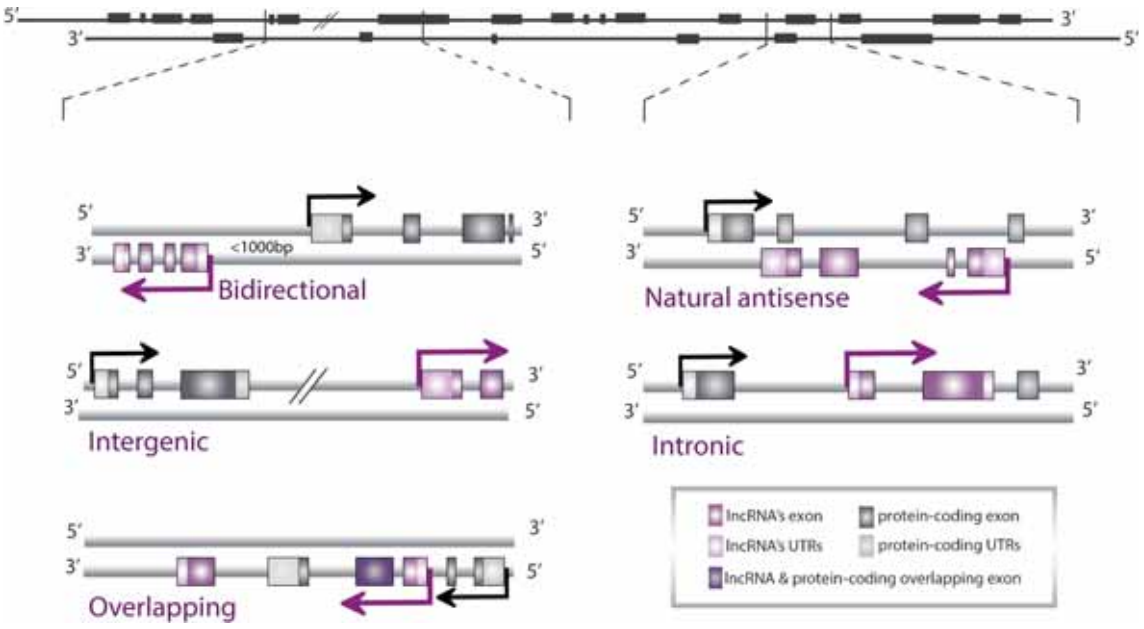

Figure 1. lncRNAs' locus and transcription directions. There are different sub-classes of lncRNAs depending on their transcription direction and location. Generally, lncRNA genes are located within genes (called interagenic lncRNAs) or outside of the protein-coding gene regions (known as intergenic lncRNAs).

significantly located in chromatin fraction indicating their regulatory roles (Derrien *et al.* 2012). Some lncRNAs are located only in the cytoplasm, and others in both nucleus and cytoplasm (Wang *et al.* 2011a). Thus, lncRNAs possess extensive functional capacity, especially in gene and chromatin regulation (Lipovich *et al.* 2010). Moreover, these findings verified Paul and Duerksen's research to some extent in which it was reported that some part of chromatin includes RNAs twice than DNAs. Figure 2 shows cellular localization of lncRNAs.

Cellular function of lncRNAs

Unlike proteins, whose functions are in *trans*, lncRNAs have been considered to have functionality in both *trans* and *cis* (Batista and Chang 2013). They have been associated with a wide spectrum of biological processes such as chromatin regulation, transcription and translation regulation, alternative splicing and nuclear import (Wu *et al.* 2013; Shahandeh 2013). They have demonstrated the extremely important roles of lncRNAs in cell processes and also greater complexity of their functions (Prensner and Chinnaiyan 2011).

LncRNAs are classified differently based on their functions making their functionality more complex. For instance, in the study carried out by Van Leeuwen and Mikkers (2010), lncRNAs are classified into two groups: class I contains lncRNAs whose functions are independent of their genetic locus and class II includes lncRNAs, which influence cellular processes in the locus-dependent manner (Van Leeuwen and Mikkers 2010). However, another study classified lncRNAs as chromatin modifiers, transcriptional modifiers and post-transcriptional modifiers (Zhang and Jeang 2013). Another study classified lncRNAs as components of the address code dividing into regions-specific, locus-specific or even allele-specific addresses (Batista and Chang 2013).

Moreover, Wang and Chang (2011) had a different insight into lncRNAs' functions. Namely, they categorized lncRNAs into four new themes included: signal, decoy, guide, and scaffold; also, they declared that individual lncRNAs may possess one or several of these themes (Wang and Chang 2011; Shahandeh 2013). In terms of signals, lncRNA's expression results in temporal-special regulation of genes in chromatin and mRNA levels. As a decoy, lncRNAs bind target molecules including transcription factors, chromatin modifiers, regulatory factors, and RNA-binding proteins, or RNA molecules such as miRNAs. Therefore, these target molecules will remain effectively inactive by interacting with lncRNAs. Acting as a guide molecule, lncRNAs recruit transcriptional and epigenetic regulatory factors and direct them towards specific target genes. Acting as a scaffold, lncRNAs recruit multiple proteins or other molecules to form ribonucleoprotein complexes (Wang and Chang 2011; Sana *et al.* 2012; Wu *et al.* 2013). These types of lncRNAs participate in the expanded regulatory functions of cells, which determine the cell fate (Huarte and Rinn 2010). Signalling and decoy lncRNAs are probably related to gene expression and regulation, while guide and scaffold lncRNAs are related to epigenetic modifications (Da Sacco *et al.* 2012).

All aforementioned studies, especially the one which is carried out by Wang and Chang (2011), elucidated the significance of the interaction of lncRNAs with other molecules, which are crucial for their precise functions. In fact, lncRNAs interact with protein(s), together with forming DNA–RNA or an RNA–RNA hybridization-based interactions

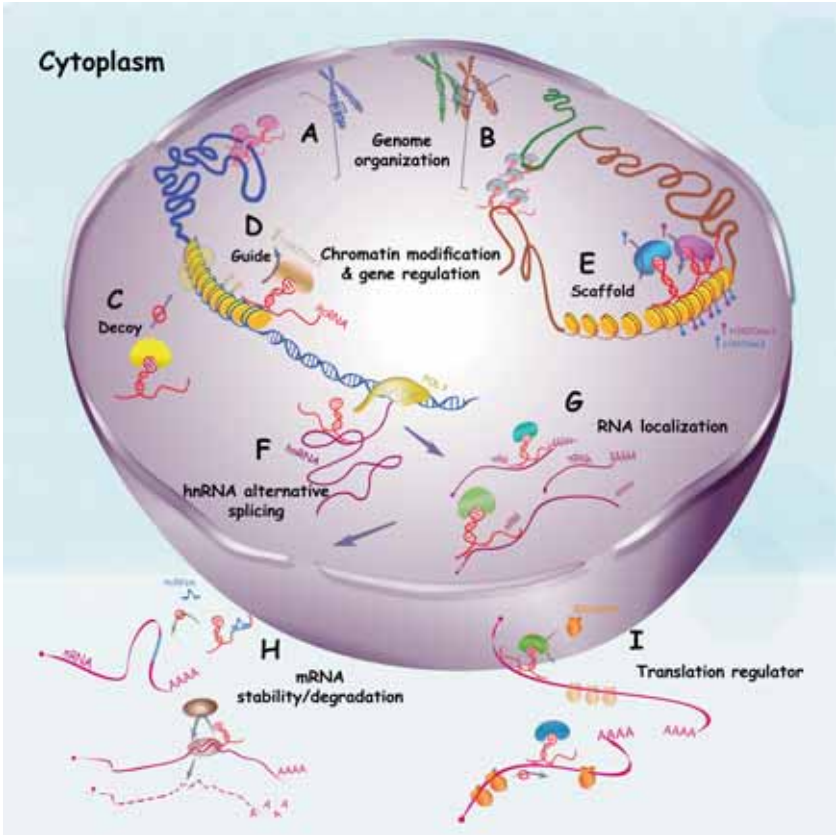

Figure 2. Cellular localization and function of lncRNAs. LncRNAs exert diverse functions in numerous cellular processes ranging from genome localization to gene expression at both posttranscriptional and transcriptional levels. lncRNAs serve as the regulators of spatial genome organization. Some lncRNAs induce formation of chromatin loop from either the same or different chromosomes. The existing nuclear subdomains contribute to the formation of heterochromatin or euchromatin region and also can lead to activation of gene clusters potentially via reciprocal interaction between enhancers and promoters (A, B) (Wang *et al.* 2011b). Moreover, lncRNAs can act as a molecular decoy (C), guide (D) and scaffold (E), each archetype is intimately associated with signalling a cellular pathway. Some lncRNAs regulate alternative splicing either by modulating splicing factors or by pure interaction with hnRNA (F). Further, lncRNAs are involved in mRNA localization, some interact with nuclear-cytoplasmic transport factors, therefore, modulating nuclear trafficking. Some others are involved in mRNA storage in nucleus by associating with nuclear speckles (G). The half-life of mRNAs is modulated by lncRNAs (H). Also, lncRNAs affect the translation process of mRNAs (I).

(Rinn and Chang 2012; Wu *et al.* 2013). In the current review, we point out some details about lncRNA's interactions and, with the aim of providing a better understanding of their roles in defining cell fate.

Some lncRNAs bind to chromatin modifier complex(es), including: PRC1, PRC2, LSD1 and G9a and consequently, they regulate gene expression epigenetically. These lncRNAs serve as scaffolds for chromatin modifier complexes and then guide the regulatory molecules to target genes by hybridization to complementary DNA sequence. Hence, these macro complexes change chromatin structures which lead to gene expression regulation in both *trans* and *cis*. Among all the lncRNAs, lincRNAs have a strong capability to recruit histone modifier enzymes. Studies showed that ∼38% of lincRNAs are expressed in various cell types. They are associated with chromatin-modifier complexes of genes including PRC2, CoREST and SMCX (Khalil *et al.* 2009; Zhang *et al.* 2013). For instance, a 2.2-kb lincRNA, HOX antisense intergenic RNA (HOTAIR), is transcribed from HOXC locus which serves as a scaffold. HOTAIR and interacts with chromatin modifier complexes such as PRC2, a H3K27 methylase and LSD1, a H3K4me2 demethylase. Eventually, HOTAIR guides repressive ribonucluprotein complexes to HOXD locus (Tsai 2010; Batista and Chang 2013). Therefore, HOTAIR silences this locus in *trans* because it is located in other chromosome (Zhang *et al.* 2013). Antisense noncoding RNA in the INK4 (ANRIL) is alncRNA that resides in INK4b-ARF-INK4a locus (Wang and Chang 2011; Gibb *et al.* 2011a), and encodes tumour suppressors p15INK4b, p14ARF and p16INK4a. These tumour suppressors are vital for cell cycle inhibition, cell senescence, stem cell renewal and apoptosis (Hauptman and Glavač 2013). ANRIL interacts with two repressive histone modifier complexes, PRC1 and PRC2 (Prensner and Chinnaiyan 2011). PRC1 and PRC2 contribute to the establishment of a repressive epigenetic mark in ANRIL and induce methylation of histone 3 in the lysine 27 (H3K27) (Huarte and Rinn 2010; Gutschner and Diederichs 2012). Thereafter, ANRIL regulates gene

expression in a *cis*-regulatory mechanism (Prensner and Chinnaiyan 2011).

Some lncRNAs interact with transcription regulatory factors and are involve in transcriptional regulation. For example, Evf-2 is a lncRNA which is transcribed from an ultra-conserved region in Dlx-5/6 through sonic hedgehog (Shh) signalling pathway during development. Evf-2 interacts with transcriptional factor DLX-2, a homeodomain protein and accelerates the Dlx6 gene transcription. The Evf-2/DLX-2 complex binds to the proximal enhancer of Dlx6 gene and upregulates its expression in *trans* (Wang *et al.* 2011a; Nie *et al.* 2012). Another example is cyclin D1 (CCND1). CCND1 is located in the promoter region of the cyclin D1 gene and downregulates its gene expression in *cis* (Wang and Chang 2011; Wapinski and Chang 2011). Upon DNA damage, CCND1 is transcribed and interacts with anRNA-binding protein, TLS. TLS is a transcriptional regulator and induces a conformation change in TLS. This conformation exchange is required for recruitment and inhibition of the two histone acetyltransferase of CBP and p300 genes. Thereby, CCND1 expression results in cyclin D1 downregulation (Gutschner and Diederichs 2012; Hauptman and Glavač 2013). In low-level DNA damages, p53 induces the expression of PANDA (P21 associated ncRNA DNA damage activated). PANDA is the next example which interacts with NF-YA, a nuclear transcription factor. NF-YA induces transcription of some apoptotic genes. Accordingly, this interaction titrates NF-YA away from proapoptotic target genes and results in cell immortality (Wang and Chang 2011; Shahandeh 2013).

Some lncRNAs mediate posttranscriptional gene regulation. They interact with splicing factors, mRNA turn over factors, translation mediators and translocation factors (Yoon *et al.* 2013). MALAT1 is a lncRNA localized in nuclear speckles and interacts with the serine/arginine-rich splicing regulatory (SR) factors (Wang and Chang 2011; Tsai *et al.* 2011; Zhang *et al.* 2013). MALAT1 modulates phosphorylation of SR factors resulting in premRNAs' alternative splicing modulation (Nie *et al.* 2012). Zeb2 is a transcriptional repressor of E-cadherin. It is expressed in epithelial and mesenchymal cells. However, Zeb2 mRNA is not translated to Zeb2 protein in epithelial cells, but in mesenchymal cells. Moreover, along with Zeb2 expression, natural antisense transcript of Zeb2 (NAT Zeb2) is expressed concurrently. NAT Zeb2 binds to exon–intron junctions and prevents the accumulation of spliceosomes in 5′ splice sites. In mesenchymal cells, stabilized Zeb2 intron encompasses an internal ribosome entry site (IRES) leading to Zeb2 mRNA translation. Therefore, NAT Zeb2 regulates E-cadherin expression indirectly through its association with Zeb2 splicing (Beltran *et al.* 2008). Figure 2 illustrates lncRNAs' functions in details in RNA, DNA and protein level.

Another example is half-STAU1-binding site RNAs (1/2 sbsRNAs) groups that are associated with mRNA turnover. These types of lncRNAs bind imperfectly to 3′UTR of certain mRNAs through their Alu elements. This interaction accelerates the Staufen 1 protein recruitment and promotes mRNAs decay (Wang and Chang 2011; Yoon *et al.* 2013). Moreover, Gadd7, as another example, is involved in mRNA turnover and leads to cdk6 mRNA degradation (Yoon *et al.* 2013). In addition, some lncRNAs are involved in mRNA protection and stability. One type of these lncRNAs is so called 'miRNA sponges'. They bind to specific miRNAs in their seed site and prevent miRNP from binding to their target mRNAs, or they compete with miRNAs for binding to the specific mRNAs (Gibb *et al.* 2011a; Zhang *et al.* 2013; Yoon *et al.* 2013). BACE1-AS, as a miRNA sponge, is located in the antisense strand of beta-secretase 1(BACE1) and competes with miR-485-5p for binding to the exon 6 of BACE1 mRNA (Yoon *et al.* 2013). Hence, BACE1-AS expression is associated with BACE1 mRNA stability and increases the protein yield of BACE1 (Zhang *et al.* 2013). Linc-ROR is another miRNA sponge (Wang *et al.* 2013) expressed in pluripotent stem cells and increases reprogramming efficiency (Loewer *et al.* 2010; Nagano and Fraser 2011; Ng and Stanton 2013). Linc-ROR and core pluripotency transcription factors (TFs) such as OCT4, SOX2 and NANOG mRNAs are considered as the targets of miR-145. However, in the presence of Linc-ROR, miR-145 is trapped resulting in self-renewal state stabilization of stem cells. This is because of the increased stability of the three aforementioned TFs' mRNA and the subsequent increase in their protein levels (Cheng and Lin 2013; Wang *et al.* 2013). These transcription factors result in embryonic stem cells-specific gene expression and prevents stem cell differentiation (Wang *et al.* 2013).

LincRNA-p21 and AS Uchl1 are other examples which interact with translation mediation factors. LincRNA-p21 is associated with CTNNB and JUNB mRNAs. Moreover, this ribonucleotide complex recruits two translation repressor proteins, Rck and Fmrp, to inhibit the translation of β-catenin and JunB (Yoon *et al.* 2012, 2013). While LincRNA-p21 inhibits target mRNA translation, AS Uchl1 promote its target mRNA expression. The AS Uchl1 is located in the antisense strand of ubiquitin carboxyterminal hydrolase L1 (Uchl1) gene. This gene is associated with the brain development and neurodegeneration. The AS Uchl1 overlaps with 5′ site of Uchl1 mRNA. SINEB2 is a specific sequence in AS Uchl1, which recruits ribosomes and accelerates Uchl1 translation (Yoon *et al.* 2013). Finally, some lncRNAs participate in nuclear trafficking of the transcription factors, such as noncoding repressor of NFAT (NRON) which modulates the activity of transcription factor NFAT. By interacting with the nucleoytoplasmic trafficking protein, importin-beta 1, NRON specifically inhibits the nuclear accumulation of NFAT and consequently, its target gene expression that results in NFAT target genes repression (Wilusz *et al.* 2009; Qureshi *et al.* 2010; Zhang and Jeang 2013).

lncRNAs'origin

In eukaryotes, although noncoding DNA has been often labelled as 'junk DNA' with no evolutionary constraint, it

experiences natural selection (Ponting *et al.* 2009). Based on the differences between the nature of coding and non-protein coding genes, each group is differently subjected to evolutionary constraint. As opposed to protein coding genes, noncoding genes are under secondary structure conservation (Gingeras 2007; Kutter *et al.* 2012). Eventually, noncoding DNAs and protein-coding DNAs reveal different evolutionary pathways since they possess different conservations (Kaessmann 2010). Little is known about the origin and evolution of lncRNAs, compared with protein-coding genes. They exhibit a low sequence conservation and rapid evolution among mammals (Kutter *et al.* 2012). In line with this, various evolutionary hypotheses have been considered for the origin of lncRNAs.

The first hypothesis is the metamorphosis of protein-coding gene generated through a gene duplication process (Ponting *et al.* 2009; Kaessmann 2010; Kapusta *et al.* 2013). During evolution, one copy of a protein-coding gene is subjected to accumulated mutations and subsequently loses its protein coding capacity. Then, a new functional lncRNA gene is created which encompasses some of the previous coding fractions including regulatory elements, splicing signals, exon sequences, and polyadenylation sequences (Ponting *et al.* 2009; Kaessmann 2010). X inactive-specific transcript (Xist) is essential for sex chromosome dosage compensation in mammals and was originated from chicken protein-coding Lnx3 gene (Kaessmann 2010; Gribnau and Grootegoed 2012; Kapusta *et al.* 2013). It was then deformed to pseudogene and then XIST was generated through the placental mammals' evolution (Gerstein *et al.* 2007; Ponting *et al.* 2009). Xist promoter region was originated from 5′-UTR of Lnx3 exons 1 and 2. Also exons 4 and 5 of the human Xist gene were generated from Lnx3 exons 4 and 11 (Elisaphenko *et al.* 2008).

lncRNAs may be originated from segmental or whole gene duplication of other ncRNA genes including small and long noncoding RNAs (Ponting *et al.* 2009; Kapusta *et al.* 2013). The genomic studies on large homologous protein-coding gene families have revealed that protein-coding gene duplication is a very common phenomenon, but unlike protein-coding genes, there is very little evidence for the whole duplication of lncRNAs. This may be due to the rapid sequence divergence of lncRNAs (Ponting *et al.* 2009; Kaessmann 2010). Mouse nuclear-enriched abundant transcript 2 (Neat2) is an example of duplicated lncRNA that is paralogous to nonexonic sequences of the mouse genome (Ponting *et al.* 2009). Segmental gene duplication within a ncRNA also seems to play a role in generation of lncRNAs. Further, local and tandem duplications have been observed in the 5′ regions of Kcnq1ot1 (Ponting *et al.* 2009; Kaessmann 2010).

De novo generation is another hypothesis of lncRNAs' origins (Kapusta *et al.* 2013). In this case, alterations in genomes such as chromosomal rearrangement, generation of (proto-) splice sites and (proto-) promoter have transformed nonfunctional genomic sequences to functional lncRNAs (Kaessmann 2010).

The last hypothesis considers transposable elements (TEs) insertions as the origin of lncRNAs (Ponting *et al.* 2009; Rinn and Chang 2012; Kapusta *et al.* 2013). In fact, TEs have contributed not only to the origin of new lncRNAs but also in diversification and the regulation of vertebrate lncRNAs (Kapusta *et al.* 2013). Comparing lncRNAs with other genes like pseudogenes, small ncRNAs and protein-coding genes, the majority of human lncRNAs have TE segment in their body. For instance, TEs are present within the internal part of exons, transcription start sites, polyadenylation (polyA) sites, or in the combination of all these parts. Studies have demonstrated that 75% of human lncRNAs have at least one exon with partial TE origin. Most TE-driven lncRNAs tend to have a few amounts of these sequences (Kapusta *et al.* 2013). Comparing with other lncRNAs, intergenic lncRNAs are broadly covered by TEs (Rinn and Chang 2012). Four major classes of TEs are attributed to lncRNAs. However, LTR/ERV elements are the most abundant TE families among lncRNAs (Kapusta *et al.* 2013). TEs exert many important functions in lncRNA-related processes including origin, diversification and regulation of lncRNAs. Due to their pivotal roles, in the following sections we will discuss TEs functions in aforementioned processes in detail.

Role of TEs in lncRNA generation

The generation of lncRNAs from TEs is one of the highly debated issues, which states 'lncRNA first' or 'TE first': Divergence or emergence? However, genomic observations confirm both converse ideas (Kapusta *et al.* 2013). The idea that supports 'lncRNA first' model is from some studies revealing that 56.5% of TEs-derived lncRNAs show a TE contribution to the last exon of lncRNAs. Also, TEs have been reported to contribute to the 3′ region (3′UTR) of protein-coding genes, and thus, implying TE insertions to the nonTE ancestral genes.

One of the observations supporting the 'TE first' model is the existence of lncRNAs which are found to be mostly or entirely consisted of TEs, as it seems to be improbable to consider insertion of all these TEs to nonTE ancestral lncRNAs. Finding some lncRNAs within the start region and first exon of TE-derived transcripts have further supported the 'TE first' idea. Human Gencode v13 identified 4404 lncRNAs with TE-derived TSS and 2213 lncRNAs, whose first exons are covered by TEs and derived from primate-specific TE families (Kapusta *et al.* 2013).

Findings supported both models and also revealed that 40% of TEs-derived human lncRNAs are primate-specific and some are even restricted to hominoids (Kapusta *et al.* 2013). Therefore, these lncRNAs are considered to be young genes. Xist is one of the genes supporting both models. It is developed from Lnx3 pseudogene and a set of diverse transposable elements are found on both sides of Xist gene. At least, a short fragment of the first exon and also, exons 2, 3, 6, 7, 8 are reported to contain different types of TEs (Elisaphenko *et al.* 2008).

Role of TEs in lineage-specific lncRNAs' diversification

The correlation between TEs insertion in a regulatory region of genes and the emergence of lineage-specific lncRNAs were claimed by several publications. Approximately 36.3% of TE-derived lncRNAs are found to be related to primate-specific lncRNAs (Kapusta *et al.* 2013). Xist, a lncRNA which evolved during evolution of eutherians is found to have lineage-specific TEs. Therefore, lineage specific TEs cause different expression of this lncRNA among various species (Elisaphenko *et al.* 2008). For example, a primate-specific FLAM_C element changes the expression of this lncRNA to be specific for the primate species (Kapusta *et al.* 2013). Other examples are lncRNA-RoR and lncRNA-ES3 which had LTR7/HERVH elements insertions during the primate evolution. These two different lncRNAs harbour DNA-binding sites of the pluripotency transcription factors such as NANOG (Loh *et al.* 2011; Kapusta *et al.* 2013). Their high and specific expression levels in human embryonic stem cells result in the maintenance of ESC pluripotency (Rinn and Chang 2012).

Role of TEs in lncRNAs' regulation

In consistent with TE insertions site on lncRNA genes, these elements contribute to many aspects of lncRNA function and regulation. For instance, TEs contribute to transcription regulation of lncRNAs, their processing (e.g. splicing, alternative splicing and polyadenylation) and also formation of lncRNA's secondary structure formation (Kapusta *et al.* 2013; Hadjiargyrou and Delihas 2013). TEs that overlap with mapped DNAseI hypersensitive sites (DHS) indicate the active chromatin site and uncover TE's significant roles in lncRNA regulation. LTRs are revealed to be TEs that reside in vicinity of the upstream region of lncRNAs. However, in protein-coding genes SINEs are the TEs located in the proximal upstream region of lncRNAs (Kapusta *et al.* 2013). The distribution of these TEs in the upstream regulator sites of protein and noncoding genes is not random. TEs take part in cell type-specific regulations of specific lncRNAs, since different TEs-derived transcription regulatory motifs on DNA recruit different types of transcription factors. Consequently, this results in different spatio-temporal expression patterns (Hadjiargyrou and Delihas 2013; Kapusta *et al.* 2013).

TEs located in the lncRNAs' transcripts are involved in their functional regulation, such as acting as *cis*-regulators or *trans*-regulators of protein-coding gene expressions. ANRIL is one of these lncRNAs that regulates gene expression epigenetically. ANRIL binds to chromatin modifier factors and interacting with Alu supplementary motifs in promoters of ANRIL *trans*-regulating genes via its Alu motif (Holdt *et al.* 2013). One of the notable examples of gene expression *trans*-regulation is Alu-containing lncRNA originated and from chromosome 11. The Alu-containing lncRNAs binds to Alu sequences in 3′UTR of target mRNAs forming specific secondary structures, recruiting Stau1 and subsequently UPF1. Their recruitments trigger target mRNA decay (Wang and Chang 2011; Hadjiargyrou and Delihas 2013).

lncRNAs in disease

Given the complex and widespread role of lncRNAs in various biological processes, it is not surprising that any mutation, which alters lncRNAs' structures and expression patterns will cause many complex human diseases. Indeed, it has been confirmed by recent studies demonstrating that lncRNAs contribute to a wide range of diseases (Niland *et al.* 2012) including cancers (Spizzo *et al.* 2012; Hauptman and Glavač 2013; Zhang *et al.* 2013), syndromes (Chiesa *et al.* 2012; Spizzo *et al.* 2012), neurological disorders, cardiovascular disorders (Rinn and Chang 2012), muscular disorders (Cabianca *et al.* 2012) and many other diseases. In the following sections, we will focus on the special lncRNAs and their roles in initiation and progression of these diseases.

lncRNAs and cancers

Cancer is a heterogeneous disease in which the balanced harmony of cellular networks has been disorganized. This is due to dysregulation of genes, whose function is to maintain cellular homeostasis (Huarte and Rinn 2010), such as genes involved in growth and proliferation signalling pathway, mortality pathway, or angiogenesis pathways (Gutschner and Diederichs 2012). RNA profiling in cancers has demonstrated an unusual alteration in lncRNA expression patterns (Zhang *et al.* 2013). Further, recent studies revealed that lncRNAs play vital roles in the aforementioned processes, all of which are associated with initiation, development and progression of cancer (Huarte and Rinn 2010). Unfortunately, the function of most lncRNAs remains unknown. Similar to protein-coding genes and miRNAs, lncRNAs have oncogenic, tumour suppressive or bidirectional function during tumourigenesis (Zhang *et al.* 2013). Given their importance in cancer development and progression, we will describe the role of some lncRNAs in cancer.

Oncogenic lncRNAs interfere with different cellular pathways and thus, induce tumourigenesis (Huarte and Rinn 2010). In a canonical state, oncogenic lncRNAs are involved in stem cell differentiation, cell proliferation and other cell cycle regulations (Gutschner and Diederichs 2012). HOTAIR, MALAT1, PCAT-1, PCGEM1, TUC338 (Zhang *et al.* 2013), ANRIL, SRA (Hauptman and Glavač 2013) are oncogenic lncRNAs aberrantly expressed in various types of cancer such as breast, hepatocellular, colorectal, gastrointestinal and bladder cancer (Zhang *et al.* 2013). As it was mentioned, HOTAIR is related to histone modifier complexes PRC2, and LSD1, and therefore, modifies the epigenetic landscape of the genes (Rinn and Chang 2012) particularly those involved in the metastatic pathways. Studies showed an increased expression of HOTAIR, up to 2000-fold in

metastatic breast cancers (Hauptman and Glavač 2013). It was found that HOTAIR is upregulated in other cancers such as colorectal cancer, gastrointestinal stromal tumours, and hepatocellular carcinoma. In all of these cases, elevated levels of HOTAIR are associated with an increased risk of metastasis and low patient life expectancy (Zhang *et al.* 2013). Similar to HOTAIR, ANRIL is reported to be an oncogenic lncRNA altering the epigenome by interacting with chromatin modifying complexes, PRC2 and PRC1 (Wang and Chang 2011; Tano and Akimitsu 2012). In fact, ANRIL conducts these complexes to INK4 locus and inhibits the expression of suppressor genes located in INK4 locus, such as CDKN2A. CDKN2A encodes the p16 CDK inhibitor (Rinn and Chang 2012). The abrogation of this tumour suppressor results in an abnormal response to the cell cycle inhibition, stress-induced apoptosis and senescence (Hauptman and Glavač 2013). Researchers found an overexpression of ANRIL in several cancers such as melanoma, leukeamia, glioma, pancreatic carcinoma (Shahandeh 2013), prostate (Hauptman and Glavač 2013), ovarian and breast cancer (Tano and Akimitsu 2012; Shahandeh 2013).

In addition to the aforementioned oncogenic lncRNAs, which regulates tumour-suppressor pathways by epigenetic silencing elements, some other oncogenic lncRNAs mediate alternative splicing patterns of some tumour-suppressor pathways (Rinn and Chang 2012). As mentioned previously, NAT Zeb2 is expressed in mesenchymal cells but not in epithelial cells and represses E-cadherin expression through increasing Zeb2 protein levels. NAT Zeb2 was found to be upregulated in some human tumours with low E-cadherin expression or within the cells which had epithelial–mesenchymal transition (Beltran *et al.* 2008; Huarte and Rinn 2010). By targeting serine/arginine (SR) splicing factors in nuclear speckles (Shahandeh 2013), MALAT1 is involved in the regulation of alternative splicing of some genes. These genes are specially growth-control genes and cell mobility-induced genes such as CTHRC1, CCT4, HMMR and ROD1 (Tano and Akimitsu 2012). The upregulation of MALAT1 was reported in various cancers (Hauptman and Glavač 2013), in progression and metastasis stages (Tano and Akimitsu 2012). In lung adenocarcinoma, prostate, pancreas, colon, breast and hepatocellular carcinomas, the overexpression of MALAT1 was reported (Hauptman and Glavač 2013). Some oncogenic lncRNAs are involved in modulation of p53 pathway such as PANDA. PANDA is an oncogenic lncRNA associated with NF-YA transcription factor. NF-YA titrates PANDA away from target apoptotic genes and eventually inhibits p53-mediated apoptosis (Wang and Chang 2011; Shahandeh 2013). Overexpression of this oncogene causing immortality of cancer cells was observed in human breast cancer (Da Sacco *et al.* 2012).

Some lncRNAs play an important role as tumour suppressor, including GAS5, MEG3, Linc-p21, PTENP1, TERRA, CCND1/cyclin D1, and TUG1 (Nie *et al.* 2012). Studies demonstrated downregulation of these tumour suppressor lncRNAs in cancers. Maternally-expressed gene 3 (MEG3) is an imprinted lncRNA gene which positively regulates p53 and selectively promotes p53-dependent transcription from p53-responsive promoter (Lipovich *et al.* 2010; Prensner and Chinnaiyan 2011; Nie *et al.* 2012; Zhou *et al.* 2012). MEG3 controls cell proliferation both in p53-dependent and p53-independent manners (Nie *et al.* 2012). The gene deletion, and promoter hypermethylation of MEG3 lead to the loss of this lncRNA's expression in tumours, such as neuroblastomas, gliomas (Zhou *et al.* 2012), myeloid leukaemia (Huarte and Rinn 2010), and hepatocellular cancers (Zhou *et al.* 2012). Growth arrest-specific 5 (GAS5) is a tumour-suppressor lncRNA mimicking glucocorticoid receptor (GR) response element DNA resulting in its interaction with DNA-binding domain of the GR (Wang and Chang 2011; Nie *et al.* 2012). This interaction sequestrates GR from its target gene promoters, and thus, inhibits the expression of GR-induced genes such as cIAP2 (cellular inhibitor ofapoptosis 2). cIAP2 mediates cell apoptosis (Prensner and Chinnaiyan 2011) and expression level of this growth inhibitor is downregulated in human breast tumours (Prensner and Chinnaiyan 2011; Hauptman and Glavač 2013). Linc-p21 is another tumour suppressor lncRNA induced by p53 signalling pathway (Prensner and Chinnaiyan 2011; Nie *et al.* 2012). By association with RNA-binding protein hnRNP-K, Linc-p21 interacts with promoters of genes which are supposed to be repressed in a p53 dependent manner (Wang and Chang 2011; Nie *et al.* 2012). Downregulation of this lncRNA was found in murine lung, sarcoma and lymphoma tumours, but its role has not been investigated in human tumours (Prensner and Chinnaiyan 2011).

Bifunctional lncRNAs are considered to be both tumour suppressors and oncogenes. H19 is one of these lncRNAs expressed from the maternal allele and regulates some genes epigenetically during cell growth and development (Hauptman and Glavač 2013). H19 promoter is efficiently controlled by a tumour suppressor transcription factor, p53 (Baldassarre and Masotti 2012), and an oncogenic transcription factor, c-Myc (Hauptman and Glavač 2013). Studies demonstrated that there is an anti-correlation between p53 and H19 expressions in gastric cancer (Baldassarre and Masotti 2012). Also, c-Myc upregulation activates H19 transcription in different cell types (Lipovich *et al.* 2010; Prensner and Chinnaiyan 2011; Hauptman and Glavač 2013). Moreover, loss of imprinting at H19 locus is reported to be another mechanism for H19 upregulation in various cancers (Gibb *et al.* 2011a). In addition to cell growth promotion and proliferation in breast and hepatocellular cancers (Prensner and Chinnaiyan 2011), H19 transcript serves as a precursor for miR-675. miR-675 directly targets tumour suppressor retinoblastoma (RB) in human colorectal cancer (Tsang *et al.* 2010; Hauptman and Glavač 2013). In general, an increased expression of H19 was verified in hepatocellular (Zhang *et al.* 2013), bladder, liver, lung, colon, oesophagus, breast (Shahandeh 2013; Li and Chen 2013), cervical and gastric cancers (Li and Chen 2013). Lack of H19 expression in colorectal cancer, teratocarcinoma and hepatocarcinoma mouse

models is rela ted to an increased polyp number, increased tumour growth rate, and accelerated tumour development (Gibb *et al*. 2011a). The diverse roles of H19 are resulted from the bifunctional nature of the lncRNA itself and/or the context-dependent manner (Gibb *et al.* 2011a).

lncRNAs and syndromes

Imprinting aberrations in related lncRNAs are associated with disease (Prensner and Chinnaiyan 2011), especially syndromes such as Beckwith–Wiedemann syndrome (BWS), Silver–Russell syndrome (SRS), Prader–Willi syndrome (PWS), Angelman syndrome (AS) (Guenzl and Barlow 2012), fragile X syndrome (FXS), fragile X-associated tremor ataxia syndrome (FXTAS) (Niland *et al.* 2012), Tourette's syndrome (Knauss and Sun 2013), velocardiofacial syndrome (VCFS) or DiGeorge syndrome, Down's syndrome (DS), and Restless Legs syndrome (RLS) (Qureshi *et al.* 2010).

In human chromosome 11, the most well-known clusters, H19/IGF2 and Kcnq1/Kcnq1ot1, are located in p15.5. This part of chromosome is a 1 Mb region containing nine imprinted genes (Vennin *et al.* 2013). H19 and KCNQ1OT1 are lncRNAs expressed on maternal and paternal alleles, respectively (Prensner and Chinnaiyan 2011). This kind of monoallelic expression of these two lncRNAs is essential for normal development and any changes in their expression cause to diseases formation (Vennin *et al.* 2013). We will further focus on these two imprinted lncRNAs in more detail.

Kcnq1ot1 is an imprinted lncRNAs located in Kcnq1/Kcnq1ot1 cluster on chromosome 11 and is expressed exclusively in paternal allele (Vennin *et al.* 2013). kcnq1ot1 expression results in the epigenetically silencing of genes in the centromeric domain of paternal chromosome 11. However, in the maternal allele, all of these genes are expressed except kcnq1ot (Chiesa *et al.* 2012). During development, loss of the epigenetic mark of kcnq1ot1 in maternal allele leads to the biallelic expression of this lncRNA. Its expression is associated with Beckwith–Wiedemann syndrome (Nie *et al.* 2012; Vennin *et al.* 2013). H19 is an imprinted lncRNA expressed only from the maternal allele and controls the expression of IGF2 (insulin-like growing factor 2) (Gibb *et al.* 2011a). IGF2 is involved in the development and growth, and is expressed from the paternal allele (Vennin *et al.* 2013). Disrupting the balance of H19 and IGF2 will eventually lead to cellular transformation (Qureshi *et al.* 2010). Loss of imprinting of H19 during development is associated with developmental syndromes such as Silver–Russel or Beckwith–Wiedemann syndromes (Vennin *et al.* 2013).

Another lncRNA is Ube3a-ATS which is associated with Angelman paternal Ube3a (Meng *et al.* 2013). Ube3a-ATS encodes an E3 ubiquitin ligase and is expressed biallelically in most tissues. However, in brain, it is expressed maternally (Edwards and Ferguson-Smith 2007; Qureshi *et al.* 2010). The mechanism of the monoallelic expression of Ube3a and silencing of Ube3a via Ube3a-ATS is elusive. Since, in human brain both the paternal and maternal UBE3A promoter remain unmethylated, H3K4me3 and H3K27me3 mark are equal in both parental promoters of Ube3a. Therefore, transcription preinitiation complex (PIC) is properly assembled at the promoter of both Ube3a parental alleles. However, a novel transcriptional collision model was proposed (Mabb *et al.* 2011; Meng *et al.* 2013). Research show that Ube3a and Ube3a-AS are transcribed head-to-head at paternal allele. Nevertheless, in the middle of their transcription two polymerases meet each other, stall, polymerases are abortively released. On the other hand, the maternal Ube3a is totally transcribed and lead to protein production because of the methylated maternal Ube3a-AS gene (Meng *et al.* 2013). Loss of maternal Ube3a expression, arising from deletions, microdeletions of imprinting region, paternal uniparental disomies is associated with Angelman syndrome (Mabb *et al.* 2011; Ng *et al.* 2013).

lncRNAs and nervous system disorders

Studies established dysregulation of lncRNAs in many neurological disorders including neurodegenerative disorders, such as Alzheimer's disease (AD), Huntington's disease (HD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) or neuroimmunological disorders like multiple sclerosis (MS) (Qureshi *et al.* 2010; Wu *et al.* 2013). These disorders are classified and summarized in table 1.

To date, some lncRNAs such as Sox2 overlapping transcript (Sox2OT), BC200, NAT-Rad18, 17A, GDNFOS and BACE1-AS have been found to have pivotal roles in the pathological process of AD (Wu *et al.* 2013). β-Secretase is an aspartic-acid protease and one of the key enzyme involved in amyloid precursor protein (APP) processing. β-Secretase generate amyloid-β-42 (Aβ42) and amyloid-β-40 (Aβ40) along with γ-secretase (Hampel and Shen 2009). Disequilibrium in the rate of these two Aβs, and increased level of Aβ-42 are followed by the generation of insoluble amyloid plaques which have toxic effects on nerve cells (Vassar 2004; Hampel and Shen 2009; Wu *et al.* 2013). Increase in BACE1 expression levels (Vassar 2004), BACE1 mRNA half-life (Faghihi *et al.* 2008) and/or BACE1 enzyme activity may result in amyloid plaques generation (Vassar 2004). BACE1-AS is antisense lncRNA of BACE1 gene and overlaps with exon 6 of BACE1, exactly where it has a miR-485-5p target site (Faghihi *et al.* 2008; Yoon *et al.* 2013). Perfect base-pairing of BACE1-AS with BACE1 mRNA can protect this mRNA from degradation by masking the binding site of miR-485-5p. Therefore, it increases BACE1 mRNA stability, BACE1 enzyme level, and accumulation of Aβ-42. This may destroy nerve cells and promote generation of AD (Wu *et al.* 2013). In addition, some clinical studies observed an increasing level of BACE1-AS in brain tissues of AD patients (Qureshi *et al.* 2010).

Table 1. lncRNAs in neurodegenerative diseases.

lncRNA	Type	Disease associated	Down or up regulated	Biological function	References
BC200	NAT	AD and PD	Soma: Up Dendritic: Down	Maintains the long-term synaptic plasticity	Wu *et al.* (2013), Knauss and Sun (2013)
Sox2OT	Overlapping	AD and PD	Up	Regulates Sox2 in neurogenesis to promote neural differentiation	Wu *et al.* (2013), Knauss and Sun (2013)
BACE1-AS	NAT	AD	Up	Increases BACE1 mRNA stability	Wu *et al.* (2013), Knauss and Sun (2013)
NAT-Rad18	NAT	AD	Up	Downregulates DNA repair protein Rad18 resulting in increased sensitivity of neurons to apoptosis	Wu *et al.* (2013)
17A	NAT	AD	Up	Impairs GABAB signalling pathway by decreasing GABAB R2 transcription	Massone *et al.* (2011), Wu *et al.* (2013)
HTTAS	NAT	HD	Down	Reduces endogenous HTT transcript levels	Wu *et al.* (2013)
NEAT1	lincRNA	HD	Up	Essential for the integrity of the nuclear paraspeckle substructure	Wu *et al.* (2013)
naPINK1	NAT	PD	Up	Stabilizes the svPINK1 resulting in disturbed Mitochondrial respiratory chain and increased sensitivity to apoptosis	Wu *et al.* (2013)
Tmevpg1	NAT	MS	Not defined	Regulates interferon gamma expression	Qureshi *et al.* (2010)

lncRNAs and other disorders

Role of lncRNAs in disorders such as cardiovascular diseases (Papait *et al.* 2013; Webster *et al.* 2013) and α-thalassaemia (Li and Ramchandran 2010; Guenzl and Barlow 2012) has been examined by various studies. Recent studies have revealed the association of a number of lncRNAs such as ANRIL, MIAT, Kcnq1ot1, Bvht and Fendrr with heart diseases (Papait *et al.* 2013). Kcnq1ot1 is an overlapping lncRNA which is located in the intron of Kcnq1 and encodes potassium voltage gated channel, necessary for normal cardiac functioning. Kcnq1 gene expression is regulated by means of its lncRNA, kcnq1ot1. Hence, aberration in expression of this lncRNA can result in abnormal heart functioning (Papait *et al.* 2013; Tang *et al.* 2013). A genomewide association study (GWAS) showed the correlation of single-nucleotide polymorphisms (SNP) in ANRIL and MIAT (myocardial infarction associated transcript) with heart diseases (Mattick 2009; Lipovich *et al.* 2010; Papait *et al.* 2013). MIAT's role has not been clearly identified yet, but it probably constitutes a component of the nuclear matrix (Clark and Mattick 2011).

The unmethylated HBA1 and HBA2 genes on chromosome 16 code α-globin in erythroid cells (Guenzl and Barlow 2012). A 18-kb deletion that includes HBA1, HBQ1 and 3′ end of LUC7L was reported in α-thalassaemia patient (Li and Ramchandran 2010; Guenzl and Barlow 2012). Impaired LUC7L encodes natural antisense transcripts expanding into CpG islands due to the termination site deletion (Li and Ramchandran 2010), which then leads to CpG island methylation of HBA2 gene, and escaping from deletion. Loss of HBA2 expression could potentially reduce functional α-globin production, and subsequently could lead to anaemia (Li and Ramchandran 2010; Guenzl and Barlow 2012).

Conclusion

The extensive exploration of lncRNAs in the last decade has intrigued the scientists to change their perspectives about the genome. Recently lncRNA functions have been discovered in every aspects of cellular function from differentiation to apoptosis. In fact, wide functional spectrum of lncRNAs revealed their importance in genome, and it is obvious that any alteration in lncRNA expression levels could potentially result in diverse diseases such as cancers, syndromes and neurodegenerative diseases.

However, this apparently marvelous discovery is like a tip of an iceberg, as we know very little about lncRNAs and more studies are required to be conducted to understand the biological function of individual lncRNAs. It is of importance to uncover whether lncRNAs themselves have enzymatic activities, as well as their roles of signalling, decoy, guide and scaffold. These discoveries will help in developing lncRNA-mediated therapies for prevention and treatment of human disease. However, further research should investigate and explore lncRNAs' evolutionary aspects and their correlation with evolution and human diseases.

References

Baldassarre A. and Masotti A. 2012 Long non-coding RNAs and p53 regulation. *Int. J. Mol. Sci.* **13**, 16708–16717.

Batista P. J. and Chang H. Y. 2013 Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–1307.

Beltran M., Puig I., Peña C., García J. M., Alvarez A. B., Peña R. *et al.* 2008 A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition. *Genes Dev.* **22**, 756–769.

Brosnan C. A. and Voinnet O. 2009 The long and the short of noncoding RNAs. *Curr. Opin. Cell Biol.* **21**, 416–425.

Cabianca D. S., Casa V. and Gabellini D. 2012 A novel molecular mechanism in human genetic disease, a DNA repeat-derived lncRNA. *RNA Biol.* **9**, 1211–1217.

Casa V. and Gabellini D. 2011 A repetitive elements perspective in polycomb epigenetics. *Front. Genet.* **3**, 199.

Chen G., Wang Z., Wang D., Qiu C., Liu M., Chen X. *et al.* 2013 LncRNA disease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* **41**, D983–D986.

Cheng E. C. and Lin H. 2013 Repressing the repressor: a lincRNA as a microRNA sponge in embryonic stem cell self-renewal. *Dev. Cell* **25**, 1–2.

Cheng W., Zhang Z. and Wang J. 2013 Long noncoding RNAs: new players in prostate cancer. *Cancer Lett.* **339**, 8–14.

Chiesa N., De Crescenzo A., Mishra K., Perone L., Carella M., Palumbo O. *et al.* 2012 The KCNQ1OT1 imprinting control region and non-coding RNA: new properties derived from the study of Beckwith–Wiedemann syndrome and Silver–Russell syndrome cases. *Hum. Mol. Genet.* **21**, 10–25.

Chodroff R. A., Goodstadt L., Sirey T. M., Oliver P. L., Davies K. E., Green E. D. *et al.* 2010 Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **11**, 72.

Clark M. B. and Mattick J. S. 2011 Long noncoding RNAs in cell biology. *Semin. Cell. Dev. Biol.* **22**, 366–376.

Clark M. B., Johnston R. L., Inostroza-Ponta M., Fox A. H., Fortini E., Moscato P. *et al.* 2012 Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898.

Da Sacco L., Baldassarre A. and Masotti A. 2012 Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis. *Int. J. Mol. Sci.* **13**, 97–114.

Derrien T., Guigó R. and Johnson R. 2011 The long non-coding RNAs: a new player in the "dark matter". *Front. Genet.* **2**, 107.

Derrien T., Johnson R., Bussotti G., Tanzer A., Djebali S., Tilgner H. *et al.* 2012 The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789.

Edwards C. A. and Ferguson-Smith A. C. 2007 Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* **19**, 281–289.

Elisaphenko E. A., Kolesnikov N. N., Shevchenko A. I., Rogozin I. B., Nesterova T. B., Brockdorff N. *et al.* 2008 A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS One* **3**, 2521.

Faghihi M. A., Modarresi F., Khalil A. M., Wood D. E., Sahagan B. G., Morgan T. E. *et al.* 2008 Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nat. Med.* **14**, 723–730.

Gerstein M. B., Bruce C., Rozowsky J. S., Zheng D., Du J., Korbel J. O. *et al.* 2007 What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **17**, 669–681.

Gibb E. A., Brown C. J. and Lam W. L. 2011a The functional role of long non-coding RNA in human carcinomas. *Mol. Cancer* **10**, 38.

Gibb E. A., Vucic E. A., Enfield K. S., Stewart G. L., Lonergan K. M., Kennett J. Y. *et al.* 2011b Human cancer long non-coding RNA transcriptomes. *PLoS One* **6**, 325915.

Gingeras T. R. 2007 Origin of phenotypes: genes and transcripts. *Genome Res.* **17**, 682–690.

Gribnau J. and Grotegoed J. A. 2012 Origin and evolution of X chromosome inactivation. *Curr. Opin. Cell Biol.* **24**, 397–404.

Guenzl P. M. and Barlow D. P. 2012 Macro lncRNAs A new layer of cis-regulatory information in the mammalian genome. *RNA Biol.* **9**, 731–741.

Gutschner T. and Diederichs S. 2012 The hallmarks of cancer A long non-coding RNA point of view. *RNA Biol.* **9**, 703–719.

Guttman M., Amit I., Garber M., French C., Lin M. F., Feldser D. *et al.* 2009 Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227.

Hadjiargyrou M. and Delihas N. 2013 The intertwining of transposable elements and non-coding RNAs. *Int. J. Mol. Sci.* **14**, 13307–13328.

Hampel H. and Shen Y. 2009 Beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) as a biological candidate marker of Alzheimer's disease. *Scand. J. Clin. Lab. Invest.* **69**, 8–12.

Hauptman N. and Glavač D. 2013 Long non-coding RNA in cancer. *Int. J. Mol. Sci.* **14**, 4655–4669.

Holdt L. M., Hoffmann S., Sass K., Langenberger D., Scholz M., Krohn K. *et al.* 2013 Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. *PLoS Genet.* **9**, e1003588.

Huarte M. and Rinn J. L. 2010 Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet.* **19**, R152–R161.

Kaessmann H. 2010 Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326.

Kapusta A., Kronenberg Z., Lynch V. J., Zhuo X., Ramsay L., Bourque G. *et al.* 2013 Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470.

Khalil A. M., Guttman M., Huarte M., Garber M., Raj A., Rivea Morales D. *et al.* 2009 Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667–11672.

Knauss J. L. and Sun T. 2013 Regulatory mechanisms of long noncoding RNAs in vertebrate central nervous system development and function. *Neuroscience* **235**, 200–214.

Kornienko A. E., Guenzl P. M., Barlow D. P. and Pauler F. M. 2013 Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* **11**, 59.

Kutter C., Watt S., Stefflova K., Wilson M. D., Goncalves A., Ponting C. P. *et al.* 2012 Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **8**, e1002841.

Li C. H. and Chen Y. 2013 Targeting long non-coding RNAs in cancers: progress and prospects. *Int. J. Biochem. Cell. Biol.* **45**, 1895–1910.

Li K. and Ramchandran R. 2010 Natural antisense transcript: a concomitant engagement with protein-coding transcript. *Oncotarget* **1**, 447–452.

Lipovich L., Johnson R. and Lin C. Y. 2010 MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. *Biochim. Biophys. Acta* **1799**, 597–615.

Loewer S., Cabili M. N., Guttman M., Loh Y. H., Thomas K., Park I. H. *et al.* 2010 Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117.

Loh Y. H., Yang L., Yang J. C., Li H., Collins J. J., Daley G. Q. *et al.* 2011 Genomic approaches to deconstruct pluripotency. *Annu. Rev. Genomics Hum. Genet.* **12**, 165–185.

Ma H., Hao Y., Dong X., Gong Q., Chen J., Zhang J. *et al.* 2012 Molecular mechanisms and function prediction of long noncoding RNA. *Sci. World J.* **541786**.

Mabb A. M., Judson M. C., Zylka M. J. and Philpot B. D. 2011 Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends Neurosci.* **34**, 293–303.

Massone S., Vassallo I., Fiorino G., Castelnuovo M., Barbieri F., Borghi R. *et al.* 2011 17A, a novel non-coding RNA, regulates GABA B alternative splicing and signaling in response to inflammatory stimuli and in Alzheimer disease. *Neurobiol. Dis.* **41**, 308–317.

Mattick J. S. 2009 The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459.

Meng L., Person R. E., Huang W., Zhu P. J., Costa-Mattioli M. and Beaudet A. L. 2013 Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. *PLoS Genet.* **9**, e1004039.

Nagano T. and Fraser P. 2011 No-nonsense functions for long noncoding RNAs. *Cell* **145**, 178–181.

Ng S. Y. and Stanton L. W. 2013 Long non-coding RNAs in stem cell pluripotency. *Wiley Interdiscip. Rev. RNA* **4**, 121–128.

Ng S. Y., Lin L., Soh B. S. and Stanton L. W. 2013 Long noncoding RNAs in development and disease of the central nervous system. *Trends Genet.* **29**, 461–468.

Niazi F. and Valadkhan S. 2012 Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3′ UTRs. *RNA* **18**, 825–843.

Nie L., Wu H-J., Hsu J-M., Chang S-S., LaBaff A. M., Li C.-W. *et al.* 2012 Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. *Am. J. Transl. Res.* **4**, 127–150.

Niland C. N., Merry C. R. and Khalil A. M. 2012 Emerging roles for long non-coding RNAs in cancer and neurological disorders. *Front. Genet.* **3**, 25.

Papait R., Kunderfranco P., Stirparo G. G., Latronico M. V. and Condorelli G. 2013 Long noncoding RNA: a new player of heart failure? *J. Cardiovasc. Trans. Res.* **6**, 876–883.

Ponting C. P., Oliver P. L. and Reik W. 2009 Evolution and functions of long noncoding RNAs. *Cell* **139**, 629–641.

Prensner J. R. and Chinnaiyan A. M. 2011 The emergence of lncRNAs in cancer biology. *Cancer Discov.* **1**, 391–407.

Qureshi I. A., Mattick J. S. and Mehler M. F. 2010 Long non-coding RNAs in nervous system function and disease. *Brain Res.* **1338**, 20–35.

Rackham O., Shearwood A. M., Mercer T. R., Davies S. M., Mattick J. S. and Filipovska A. 2011 Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. *RNA* **17**, 2085–2093.

Rinn J. L. and Chang H. Y. 2012 Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166.

Sana J., Faltejskova P., Svoboda M. and Slaby O. 2012 Novel classes of non-coding RNAs and cancer. *J. Transl. Med.* **10**, 103.

Shahandeh A. 2013 Molecular mechanisms of oncogenic long non-coding RNAs. *Biosci. Res.* **10**, 38–54.

Spizzo R., Almeida M. I., Colombatti A. and Calin G. A. 2012 Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* **31**, 4577–4587.

Sun J., Lin Y. and Wu J. 2013 Long non-coding RNA expression profiling of mouse testis during postnatal development. *PLoS One* **8**, e75750.

Taft R. J., Pheasant M. and Mattick J. S. 2007 The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299.

Tang J. Y., Lee J. C., Chang Y. T., Hou M. F., Huang H. W., Liaw C. C. *et al.* 2013 Long noncoding RNAs-related diseases, cancers, and drugs. *Sci. World J.* **943539**.

Tani H., Mizutani R., Salam K. A., Tano K., Ijiri K., Wakamatsu A. *et al.* 2012 Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22**, 947–956.

Tano K. and Akimitsu N. 2012 Long non-coding RNAs in cancer progression. *Front. Genet.* **3**, 219.

Tsai M. C., Manor O., Wan Y., Mosammaparast N., Wang J. K., Lan F. *et al.* 2010 Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693.

Tsai M. C., Spitale R. C. and Chang H. Y. 2011 Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* **71**, 3–7.

Tsang W. P., Ng E. K., Ng S. S., Jin H., Yu J., Sung J. J. *et al.* 2010 Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. *Carcinogenesis* **31**, 350–358.

Van Leeuwen S. and Mikkers H. 2010 Long non-coding RNAs: Guardians of development. *Differentiation* **80**, 175–183.

Vassar R. 2004 BACE1: The beta-secretase enzyme in Alzheimer's disease. *J. Mol. Neurosci.* **23**, 1–2.

Vennin C., Dahmani F., Spruyt N. and Adriaenssens E. 2013 Role of long non-coding RNA in cells: example of the H19/IGF2 locus. *Adv. Biosci. Biotechnol.* **4**, 33–44.

Volders P-J., Helsens K., Wang X., Menten B., Martens L., Gevaert K. *et al.* 2013 LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* **41**, D246–D251.

Wallace J. G., Zhou Z. and Breaker R. R. 2012 OLE RNA protects extremophilic bacteria from alcohol toxicity. *Nucleic Acids Res.* **40**, 6898–6907.

Wang K. C. and Chang H. Y. 2011 Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914.

Wang X., Song X., Glass C. K. and Rosenfeld M. G. 2011a The long arm of long noncoding RNAs: Roles as sensors regulating gene transcriptional programs. *Cold Spring Harb. Perspect. Biol.* **3**, a003756.

Wang X. Q., Crutchley J. L. and Dostie J. 2011b Shaping the genome with non-coding RNAs. *Curr. Genomics* **12**, 307–321.

Wang Y., Xu Z., Jiang J., Xu C., Kang J., Xiao L. *et al.* 2013 Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev. Cell* **25**, 69–80.

Wapinski O. and Chang H. Y. 2011 Long noncoding RNAs and human disease. *Trends Cell Biol.* **21**, 354–361.

Webster A. L., Yan M. S. and Marsden P. A. 2013 Epigenetics and cardiovascular disease. *Can. J. Cardiol.* **29**, 46–57.

Wilusz J. E., Sunwoo H. and Spector D. L. 2009 Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504.

Wright M. W. and Bruford E. A. 2011 Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature. *Hum. Genomics* **5**, 90–98.

Wu P., Zuo X., Deng H., Liu X., Liu L. and Ji A. 2013 Roles of long noncoding RNAs in brain development, functional diversification and neurodegenerative diseases. *Brain Res. Bull.* **97**, 69–80.

Yoon J. H., Abdelmohsen K. and Gorospe M. 2013 Posttranscriptional gene regulation by long noncoding RNA. *J. Mol. Biol.* **425**, 3723–3730.

Yoon J. H., Abdelmohsen K., Srikantan S., Yang X., Martindale J. L., De S. *et al.* 2012 LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* **47**, 648–655.

Zhang H., Chen Z., Wang X., Huang Z., He Z. and Chen Y. 2013 Long non-coding RNA: a new player in cancer. *J. Hematol. Oncol.* **6**, 37.

Zhang Q. and Jeang K. T. 2013 Long noncoding RNAs and viral infections. *Biol. Med.* **3**, 34–42.

Zhou Y., Zhang X. and Klibanski A. 2012 MEG3 noncoding RNA: a tumor suppressor. *J. Mol. Endocrinol.* **48**, 45–53.

Received 17 July 2014, in final revised form 2 April 2015; accepted 8 April 2015
Unedited version published online: 16 April 2015
Final version published online: 8 October 2015