DEFINITION AND CLASSIFICATION OF ncRNAs

Previously ncRNA has been referred to by various names: non-protein-coding RNA (ncpRNA); Mattick, 2003), intronic and intergenic ncRNAs (Liontio et al., 2009), and mRNAs-like ncRNAs (Inagaki et al., 2005; Rymarquis et al., 2008). At present, ncRNAs are classified on the basis of their main functions: structural ncRNAs and regulatory ncRNAs (Mattick and Makunin, 2006). Structural ncRNAs include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), spliceosomal uRNAs (snRNAs), and snoRNAs. Most of these ncRNAs have well-established structural functions. ncRNAs with regulatory roles in gene expression are classified as regulatory ncRNAs, including small interfering RNA (siRNA), micro-RNAs (miRNAs), piwi-RNAs (piRNAs), long ncRNAs, and long intergenic ncRNAs.

Numerous studies in the past decade have focused on small ncRNAs. As a result of these studies it is now clear that this class of ncRNAs regulates almost every aspect of gene expression (Goodrich and Kugel, 2006).

In addition to small ncRNAs, large numbers of long ncRNAs have recently been revealed by large-scale transcriptome analyses. Although only a limited number of long ncRNAs have well-characterized structures and functions, many studies suggest that this class of ncRNA accounts for a large fraction of the transcriptome, and they are believed to play important roles in many key molecular regulatory processes (Vazquez and Krebs, 2007; Umlauf et al., 2008; Mercer et al., 2009). We will recapitulate the pervasive transcription and genome/transcriptome complexity of these regulatory ncRNAs, particularly with respect to long ncRNAs, and review their primary proposed functional models.

PERVASIVE TRANSCRIPTION OF ncRNAs

Rapid development in analytical technologies, such as whole genome tilling arrays, capped analysis of gene expression (CAGE), Chip-chip, Chip-seq, and RNA deep sequencing have revised people’s views of eukaryotic genome/transcriptome complexity (Carninci, 2006; Gustincich et al., 2006). In the past decade, large-scale transcriptome analyses of several organisms indicate that genomes are pervasively transcribed and ncRNAs account for a large proportion of the whole transcriptome (Birner et al., 2004; Birney et al., 2007).

THE HUMAN TRANSCRIPTOME IS MORE COMPLICATED THAN EXPECTED

It has been more than decade since the human genome was sequenced, yet the decoding of this information is far from complete. According to the statistics of the version 34b of the Ensembl Human Genome, there are about 28–25,000 protein-coding genes, with a total coding length of ~34 Mb, which only occupies ~1.2% of the whole genome. On the other hand, about 1.679 Mb non-coding sequences, accounting for more than half (~57%) of the whole human genome, are believed to be transcribed. These non-coding sequences include introns, untranslated regions (UTRs), and other intronic and exonic sequences covered by spliced cDNAs/ESTs that are not annotated as protein coding. The
Although there have been fewer large-scale transcriptome studies of species other than human and mouse, they have confirmed the existence of ncRNAs. Seventeen distinct non-protein-coding polyadenylated transcripts were identified from the intergenic regions of the fly genome (Tuppy et al., 2005). Moreover, 136 strong candidates for mRNA-like ncRNAs were screened from 11,691 fly full-length cDNAs, and 33 of them were expressed during embryogenesis. Of these 35 mRNA-like ncRNAs, 27 were detected only in specific tissues (Inagaki et al., 2005). These results indicate that many mRNA-like ncRNAs are expected to play important roles in the fly. In 2005, approximately 1,300 genes that produce functional ncRNAs were demonstrated in the worm C. elegans (Stricklin et al., 2005). However, the worm transcriptome is much more complicated than expected. The worm non-coding transcriptome mapped by whole-genome tiling array showed that at least 70% of the total worm genome was transcribed, and 44% of the total observed transcriptional output on the array was predicted to consist of non-polyadenylated transcripts without protein-coding potential. Seventy percent of these non-polyadenylated transcripts were shown to overlap with the coordinates of coding loci in complicated fashions (He et al., 2007). The prevalence of ncRNAs extends even further, as studies of Saccharomyces cerevisiae have also revealed large numbers of ncRNAs (Havlíko et al., 2005; Miura et al., 2006).

**EVIDENCE FROM WELL-CHARACTERIZED LONG ncRNA DATASETS**

In past several years, our knowledge of long ncRNAs has been expanding thanks to the identification and annotation of diverse classes of long ncRNAs from human, mouse, and other species (Table 1). About 1,600 large intervening/intergenic ncRNAs (lincRNAs) were identified based on the chromatin-state maps from four mouse cell types (Cuttman et al., 2009). Based on the same method, ~3,300 lincRNAs were characterized according to the chromatin-state maps of various human cell types (Khali et al., 2009). Moreover, a class of ~3,200 enhancer-like long ncRNAs were discovered as a result of the ENCODE project (Orom et al., 2009). The rapid drop in price of next generation sequencing drove the generation of large amounts of RNA-seq data from a number of species. More than a thousand multi-exonic lincRNAs were revealed by reconstruction of transcriptomes from three mouse cell types (Cuttman et al., 2010). Human transcriptome data from more sources (24 tissues and cell types), allowed the reconstruction of more than 6,000 human lincRNAs (Cahil et al., 2011). Large numbers of long ncRNAs were also found in zebrafish, fly, and worm transcriptomes based on RNA-seq data. A stringent set of 1,153 non-coding multi-exonic transcripts, including lincRNAs, intronic overlapping long ncRNAs, exonic antisense overlapping long ncRNAs, and precursors for small RNAs (lincRNAs), were identified from transcriptome data of eight early zebrafish development stages (Pauls et al., 2011). Recently, 1,199 putative lincRNAs and more than 800 lincRNAs were annotated from fly and worm transcriptomes based on RNA-seq data (Nam and Bartel, 2012; Ioung et al., 2012).

**EOLVOLUTIONARY CONSERVATION OF LONG ncRNAs**

In contrast to well-conserved small ncRNAs, like miRNAs, the evolutionary sequence conservation of long ncRNAs is less pronounced. Most studies have shown that long ncRNAs are poorly conserved compared to protein-coding genes (Louro et al., 2009;...
Recently well-characterized long ncRNA datasets.

| Dataset                          | Number of lncRNAs | Source | Method                                    | Reference                  |
|---------------------------------|-------------------|--------|-------------------------------------------|----------------------------|
| Chromatin-state-based lincRNAs (human) | 4,860*            | 10 cell types | Chromatin signature identification (K4–K36 domain) | Khalil et al. (2009)          |
| Enhancer-seq-based lincRNAs (human) | 3,011             | Multiple | Screening from GENCODE annotation          | Orom et al. (2010)           |
| RNA-seq-based lincRNAs (human)   | 8,195             | 24 tissues and cell types | Screening from assembled RNA-seq data | Cabili et al. (2011)          |
| Chromatin-state-based lincRNAs (mouse) | 2,127*            | Four cell types | Chromatin signature identification (K4–K36 domain) | Guttman et al. (2009)          |
| RNA-seq-based lincRNAs (mouse)   | 1,140             | Three cell types | Screening from assembled RNA-seq data | Guttman et al. (2010)          |
| RNA-seq-based long lncRNAs (zebrafish) | 1,133             | Eight embryonic stages | Screening from assembled RNA-seq data | Pauli et al. (2011)           |
| RNA-seq-based lincRNAs (fruit fly) | 1,119             | 30 developmental time points | Screening from assembled RNA-seq data | Young et al. (2012)           |
| RNA-seq-based lincRNAs (C. elegans) | 882               | Multiple | Screening from assembled RNA-seq data | Nam and Bartel (2012)          |

*These are the exons identified by microarray from non-coding K4–K36 domains.

Table 1

In a comparison between human and mouse long ncRNAs, Pang et al. (2006) found that the sequence homology of long ncRNAs was similar to that of introns (70% between mice and humans) and a little less conserved than 5′ or 3′ UTRs. Thus the evolutionary constraints acting on long ncRNAs may differ from the constraints affecting small ncRNAs, allowing long ncRNAs to evolve faster than small RNAs. However, conservation analysis of long ncRNAs based on 50-nt window size revealed that many long ncRNAs may retain patches of higher conservation within their overall sequences, possibly representing interaction sites with RNA-binding proteins (Pang et al., 2016).

Recently, novel long ncRNA datasets identified from diverse species have confirmed that most long ncRNAs are less conserved than protein-coding genes while still showing clear conservation. Over 95% of the 1,600 mouse lincRNAs identified by chromatin-state maps showed clear evolutionary conservation (Guttman et al., 2009). Subsequent analysis of 3,300 human chromatin-state based lincRNAs also indicated that these lincRNAs were more conserved than intronic regions (Khalil et al., 2009). Analysis of human enhancer-like long ncRNAs also showed that the global conservation levels of these long ncRNAs were less than protein-coding genes, but higher than ancestral repeats (Orom et al., 2010). Long ncRNAs reconstructed from mouse RNA-seq data showed similar conservation levels compared to chromatin-state based lincRNAs (Guttman et al., 2010). In human, RNA-seq-based long ncRNAs showed moderate conservation across different species (Cabili et al., 2011). The conservation of zebrafish RNA-seq-derived long ncRNAs assessed by CBL score was substantially lower than protein-coding genes and comparable to intronic sequences (Pauli et al., 2011). Analysis from the fly RNA-seq-based lincRNAs also showed that most of these ncRNAs, even for those expressed at low levels, have significantly lower nucleotide substitution rates compared with either untranscribed intergenic sequence or neurally evolving short introns (Young et al., 2012). RNA-seq-based lincRNAs identified from another invertebrate organism C. elegans were differentiated into two subclasses according to their conservation, non-conserved and moderately conserved. Similar to vertebrates, some of these C. elegans lincRNAs also tend to have short regions of conservation (Nam and Bartel, 2012).

Overall, while long ncRNAs identified from different species and based on different methods showed slightly different levels of conservation, it is clear that long ncRNAs are less conserved than protein-coding genes but still exhibit clear conservation compared to non-functional genomic elements. One widely accepted interpretation of poor sequence conservation for long ncRNAs is that long ncRNAs may function at the secondary structure level instead of the primary sequence level. This is in contrast to protein-coding, which genes require conserved nucleotide sequence to encode higher levels of structure with similar biological functions.

Differently, the small conserved patches observed in some long ncRNAs might be sufficient to support the functions of these long ncRNAs, by binding with proteins, interacting with DNA promoters or with UTRs of mRNAs. Finally, the long ncRNA datasets described above were identified using different methods, possibly fostering bias for some classes of long ncRNAs, which might be subject to different selective pressure.

**GENOMIC ORGANIZATION of ncRNAs**

Regulatory ncRNAs originate from different genomic regions (Figure 1). UTRs account for many of the regions encoding ncRNAs. Statistics from the UCSC human genome (NCBI build 35) show that total UTR sequences account for ~1.1% of the whole human genome, nearly equivalent in length to protein-coding regions (32–34 Mb; Frith et al., 2005). This suggests that there may be unknown regulatory elements in these regions. Studies using CAGE, serial analysis of gene expression (SAGE), cDNA libraries, and microarray expression profiles have shown that there are independent transcripts expressed from 3′ UTRs. This class of independent transcripts has been termed “uarRNAs” (UTR-associated RNAs), some of which have been validated as being expressed in cell- and subcellular-specific fashion (Mercer et al., 2010).
In addition to UTRs, other non-coding regions of genome, such as intronic sequences are also a potential source of functional ncRNAs. Over 30% of the human genome is made up of intronic sequences (Mattick and Gagen, 2001), and many highly conserved sequences have been identified in introns (Tait et al., 2007). Recent research has indicated that there are a large number of long intronic ncRNAs in both human and mouse (Nakaya et al., 2007; Loures et al., 2008, 2009). Long ncRNAs can also be derived from both the sense and antisense strands of various genomic regions, some of which overlap with or are within protein-coding genes. These results indicate that distinguishing between protein-coding and non-coding RNAs may be difficult in some circumstances (Dinger et al., 2008). Most importantly, tens of thousands of long ncRNAs have been identified from intergenic regions (lincRNA), as discussed above. More and more lincRNAs have been validated and shown to possess important regulatory functions.

BROAD FUNCTIONALITY OF LONG ncRNAs

Recent reports have revealed the widespread functionality of long ncRNAs, ranging from epigenetic modification, to transcriptional and post-transcriptional regulation of protein-coding genes. These functions may only account for part of the functional repertoire of long ncRNAs, but they provide quite clear evidence supporting the functional significance of long ncRNAs.

CHROMATIN MODIFICATION

Many studies have shown that long ncRNAs play important roles in chromatin modification (Mattick, 2003; Costa, 2008). Dosage compensation achieved by X-chromosome inactivation (XCI) is a classic example of chromatin modification mediated by long ncRNAs in mammals (Leeb et al., 2009). There are two ncRNAs involved in this process. Xist, a 17-kb long ncRNA, initiates XCI, while Tsix, an antisense non-coding transcript to the Xist gene, opposes XCI. However, the exact mechanism of XCI mediated by Xist is still under investigation.
Qu and Adelson ncRNA conservation and function

these two ncRNAs is still unclear. Ogawa et al. (2008) reported that murine Xist and Tsix may form Xist:Tsix duplexes and be processed into small RNAs by Dicer, then subsequently these small RNAs trigger the RNAi machinery to drive XCI. Another mechanism has been proposed to explain how Xist and Tsix regulate XCI. In this model, a 1.6-kb ncRNA (RepA) transcribed from Xist loci identifies and recruits polycomb repressive complex 2 (PRC2), whose catalytic subunit, Ezh2, functions as the RNA binding subunit, initiating XCI. Tsix keeps the X chromosome active by inhibiting the interaction of RepA and PRC2 (Zhao et al., 2008).

HOTAIR is another well-characterized long ncRNA that can alter chromatin structure by recruiting polycomb proteins. There are 39 human HOX genes which can be divided into four clusters (HOXA-D) based on their locations on different chromosomes (Woo and Kingston, 2007). A total of 231 HOX ncRNAs were identified from these human HOX loci. These HOX ncRNAs have specific sequence motifs, are spatially expressed along developmental axes, and their expression demarcates broad chromosomal domains of differential histone methylation and RNA polymerase accessibility. A 2.2-kb ncRNA in the HOX ncRNA cluster, called HOTAIR, can induce heterochromatin formation and repress transcription in trans by recruiting PRC2 to trimethylate the lysine-27 residues of histone H3 (Figure 2; Rinn et al., 2007).

A common model of epigenetic control relies on ncRNAs acting as chromatin modifying complexes. Another example of this type of mechanism involves the imprinted ncRNA Air, which is required for allele-specific silencing of cis-linked Slc22a3, Slc22a2, and Igf2r genes in mouse placenta. Air is believed to target repressive histone-modifying changes by interacting with the Slc22a3 promoter chromatin and H3K9 histone methyltransferase G9a to epigenetically repress transcription (Nagano et al., 2008). A final example of this type of transcriptional control is driven by Kcnq1ot1, an antisense ncRNA, that mediates lineage-specific transcriptional silencing patterns by recruiting chromatin-remodeling complexes (G9a and PRC2) to specific regions in the Kcnq1 locus (Pandey et al., 2008).

**TRANSCRIPTIONAL REGULATION**

Many long ncRNAs can directly regulate gene expression at the transcriptional level. Specific mechanisms for direct regulation include transcriptional interference by binding to enhancers, promoters, and transcription factors, the latter being able to alter gene expression at a global level.

Transcriptional interference from long ncRNA has been shown for SRG1 (SER3 regulatory gene 1), a well-studied ncRNA in *S. cerevisiae*. The SER3 gene encodes a serine biosynthesis related enzyme. This gene is strongly repressed and its regulatory region highly transcribed when *S. cerevisiae* are grown in a rich medium. The highly expressed transcript from the SER3 regulatory region was identified by northern blot analysis as SRG1, a 550-nt long polyadenylated ncRNA. Substitution analysis of a 150-bp sequence of SRG1 revealed that SRG1 can interfere with the activation of the SER3 promoter to repress SER3 gene expression (Figure 3A; Martens et al., 2004). In metazoa, the bithoraxoid (bxd) ncRNAs of the fly bithorax complex (BX-C) are a cluster of ncRNAs that have been shown to regulate gene expression by transcriptional interference. In this case, the transcription of several bxd ncRNAs are linked to the repression of the Ubx (Ultrabithorax)
protein-coding gene. Transcription of bxd ncRNAs represses Ubx expression in cis, where Ubx transcription is repressed by the transcrip-
tional elongation of bxd ncRNAs. This is facilitated by the Trithorax complex TAC1, a transcriptional effector that binds to
the bxd region (Frueh et al., 2006).

Interaction of promoters with long ncRNAs can also regu-
late gene expression. One example is a non-coding transcript
initiated from the upstream minor promoter of the human dihy-
drofolate reductase (DHFR) gene, which represses gene expression
by promoter inactivation. The DHFR locus has two promoters,
with the downstream major promoter responsible for 99% of
RNA transcription (Masters and Attardi, 1985). The upstream
promoter generates a non-coding transcript that forms a stable
complex with the major promoter by interacting with transcription
factor II B (TFIIB). This complex acts by dissociating the
pre-initiation complex from the major promoter (Figure 3B; Martianov et al., 2007). Another signal-induced low-copy-number
ncRNA, over 200 nt long, named ncRNA\(_{CCND1}\), also medi-
ates the repression of gene expression by promoter interaction
(Wang et al., 2008). ncRNA\(_{CCND1}\) recruits a key transcrip-
tional sensor of DNA damage, the translocated in liposarcoma
(TLS) RNA-binding protein, to the promoter region of cyclin
CCND1, which repressed TLS RNA-binding protein inhibits
the histone acetyltransferase activities of CREB-binding protein
(CBP) and p300. This leads to the repression of CCND1 gene
expression in human cell lines. Of particular interest is the signal-
induced transcription of ncRNA\(_{CCND1}\), which may provide a
novel understanding of stimulus-specific expression of ncRNAs
(Wang et al., 2008).

In addition to promoter inactivation or activation, an increas-
ing number of studies now suggest that ncRNAs also regulate
gene expression by interacting with transcription factors. One
example is Evf-2, which is a ~3.8-kb ncRNA transcribed from
ei, one of the two Dlx-5/6 ultraconserved intergenic regions
(Zerucha et al., 2000). The ultraconserved region of Evf-2

![FIGURE 3](image)

**FIGURE 3** | Transcriptional regulation by long ncRNA. Long ncRNAs can conduct transcriptional regulation of target genes by (A) transcriptional interference. The purple rectangle represents the promoter region for SER3, and the blue boxes are SER3 upstream activating sequence (UAS) elements; (B) interacting with promoters. The short purple rectangle represents the minor promoter, and the long purple rectangle represents the major promoter; (C) interacting with transcriptional factors. ei and eii are two ultraconserved enhancers.
human Alu RNA, which is transcribed from short interspersed elements (SINEs), is recognized as a transacting transcriptional repressor which inhibits transcription by binding to RNA polymerase II (Pol II) complexes at promoters in vitro as a result of heat shock (Mariner et al., 2008). Because Alu elements are so abundant in the human genome, they may contribute to long ncRNA transcriptional repressor function (Amaral and Mattick, 2008).

POST-TRANSCRIPTIONAL REGULATION

There are many reports providing evidence that ncRNAs have the ability to regulate various aspects of post-transcriptional mRNA processing, such as splicing, editing, transport, translation, or degradation. The significant role in post-transcriptional regulation of gene expression mediated by small regulatory ncRNAs has been well characterized in various species (see reviews Grishok, 2005; Kavi et al., 2005; Wienholds and Plasterk, 2005; Scherr and Eide, 2007; Filipowicz et al., 2008). Here we discuss how long ncRNAs can mediate post-transcriptional regulation via specific mechanisms.

Some antisense ncRNAs have been shown to regulate gene expression at the post-transcriptional level. For example, SAF is a long ncRNA transcribed from the antisense strand of intron 1 of the human Fas gene. The overexpression of SAF caused the proteins encoded by Fas to fail to anchor to the cell membrane and induce Fas-mediated apoptosis. It is believed that SAF regulates the expression of Fas alternative splicing forms through pre-mRNA processing (Xia et al., 2005). Another natural antisense transcript (NAT) of the Snail1 gene is a cis-acting regulation of the cis-acting elements, found in the 3′ UTRs of many unstable mammalian mRNAs, controlling their half-lives (Bevilacqua et al., 2003). This cis-acting regulation can be inhibited, as shown by a bcl-2/IgH antisense transcript, formed by with bcl-2/IgH translocation, that up-regulates bcl-2 mRNA expression. This hybrid antisense transcript masks AU- rich motifs present in the 3′ UTR of the bcl-2 mRNA, increasing the stability of the protein-coding mRNA (Figure 4; Capaccioli et al., 1996). Although there is still little direct experimental evidence to identify all mechanisms involved, comparison of genome-scale expression profiles between protein-coding and non-protein-coding RNAs suggests that widespread post-transcriptional control of gene expression via the stabilization of protein-coding RNAs does occur (Nakaya et al., 2007). CANCER AND LONG ncRNAs

Many ncRNAs play regulatory roles in cancer biology. Because they regulate cell differentiation and various developmental processes, the mis-expression of long ncRNAs can regulate clinically significant cancer genes. A number of long ncRNAs have been associated with cancer development and progression. The antisense ncRNA p15SAS epigenetically silences its sense target gene p15 in leukemia (Yu et al., 2008). The expression of p15SAS induces p15 silencing in cis and trans through heterochromatin formation. p15 silencing and increased cell growth were observed after differentiation of mouse embryonic stem cells induced by exogenous p15SAS (Yu et al., 2008). ANRIL (antisense ncRNA from the INK4A-ARE-INK4B locus), which is regarded as an isofrom of p15SAS, interacts with chromobox homolog 7 (CBX7), a subunit of the PRC1 protein, and mediates the epigenetic transcriptional repression of its sense locus (Yap et al., 2010). Subsequent study revealed that this ncRNA binds to SUZ12 (suppressor of zeste 12 homolog), a component

![FIGURE 4 | Post-transcriptional regulation by long ncRNA.](image-url)
of the PRC2, and recruits PRC2 to epigenetically repress p33INK4a (Kotake et al., 2011).

In addition to acting as repressors of tumor suppressor genes, long ncRNAs also contribute to tumorigenesis via other mechanisms. SRA is a well-characterized ncRNA, which can co-activate the activity of a number of nuclear receptors in tumors. It can promote muscle differentiation and myogenic conversion of non-muscle cells through the co-activation of MyoD activity as discussed above (Caretto et al., 2006; Hube et al., 2011). Another long ncRNA, PCAT-1 (prostate cancer-associated transcript 1), which is over-expressed in a subset of prostate cancers, particularly metastatic tumors, is known to regulate cell proliferation in prostate cancer progression (Pennon et al., 2011). Moreover, long ncRNA MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) was shown to be significantly associated with metastasis in non-small cell lung cancer patients (Li et al., 2003). Subsequent analysis indicated that MALAT1 was overexpressed in five other non-hepatic human carcinomas (Lin et al., 2007).

MALAT1 may play important roles in tumor cell invasion and formation of metastases (Tieng et al., 2009; Tan, et al., 2010). In prostate cancer, a cDNA microarray analysis of intronic transcripts indicated that a high percentage (6.6%) of intronic transcripts were correlated with the degree of prostate tumor differentiation compared to transcripts from unannotated genomic regions (1%; Reis et al., 2004). In renal carcinoma cells (RCC) expression profiles also revealed that there are some non-coding intronic RNAs that are associated with malignant transformation of normal renal cells to tumor cells (Brito et al., 2008). As a result of these and similar observations, long ncRNAs have been used as diagnostic biomarkers because of their cell type-specific or stage-specific expression in different cancers (Malard et al., 2008; Reis and Verjovski-Almeida, 2012).

In addition to their functions contributing to tumorigenesis, many ncRNAs are known to act as tumor suppressors. One example is the imprinted gene MEG3 (maternally expressed gene 3), which functions as a long ncRNA. Although MEG3 has an open reading frame, it is the folding of MEG3 RNA that activates p33 expression and selectively regulates p33 target gene expression (Zhou et al., 2007). In addition, MEG3 can also inhibit cell proliferation via a p33-independent pathway. This evidence suggests that MEG3 functions as a tumor suppressor in p33 dependent and independent fashion (Zhou et al., 2007; Zhang et al., 2010). Another long ncRNA, Gas5 (growth arrest-specific 5), binds to the DNA-binding domain of the glucocorticoid receptor (GR), preventing the interaction of glucocorticoid response elements (GRE) with GR. The repression of GR suppresses the glucocorticoid-mediated induction of several genes, leading to apoptosis (Kino et al., 2010). Among the more than 1000 mouse chromatin-state based lincRNAs, one of them (lincRNA-p21) functions as a repressor of p33-dependent transcriptional response. LincRNA-p21 is a transcriptional target gene of p33. It recruits a repressor complex, including heterogeneous nuclear ribonucleoprotein K (hnRNPK), to a subset of previously active genes, mediating global gene repression and leading to apoptosis (Guttman et al., 2009; Hua et al., 2010).

These results clearly illustrate the functional significance of long ncRNAs in tumorigenesis and cancer regulatory networks and transcriptional pathways. However, some mechanisms of long ncRNAs in cancer biology seem to be more complicated than expected. For instance, lincRNA-p21 is transcribed from a region ∼15 kb upstream of p21 and mediates apoptosis in a p33-dependent manner upon DNA damage response as discussed above (Hua et al., 2010). Another single eucaryotic long ncRNA, PANDA (P21 associated ncRNA DNA damage activated), is transcribed from the ∼5 kb upstream region of p21 in an anti-sense orientation to p21. The expression of PANDA is also induced by DNA damage and activated in a p33-dependent manner as lincRNA-p32. However, in contrast to lincRNA-p21, PANDA interacts with the transcription factor NF-YA to limit the expression of some pro-apoptotic genes (Hung et al., 2011). This is just one example of the complexity of cancer-related gene regulation by long ncRNAs. As more long ncRNAs become validated, we can imagine that more regulatory roles of long ncRNAs in tumorigenesis will be unveiled.

CONCLUSION

The recent explosion in studies of ncRNAs has fostered a new view of the RNA world. It is clear that gene regulation networks are more complicated than expected. And that in future, the central dogma may be challenged by more roles for ncRNAs. Genomes possess a high percentage of non-coding regions, and express a huge repertoire of ncRNAs, which probably contribute to cellular regulatory networks.

The functional significance of ncRNAs has been debated because of their perceived lack of evolutionary conservation. Lower conservation of ncRNAs (mostly for long ncRNAs) was regarded as an argument against functional importance and as a manifestation of transcriptional noise. But less conservation does not mean less function. Many studies indicate that evolutionary constraints on ncRNAs are different to protein-coding RNAs. These different constraints allow many ncRNAs to evolve more quickly subject to positive selection. The complexity underlying the evolutionary conservation of ncRNAs may be stem from the heterogeneity of ncRNAs. ncRNAs derived from different genomic regions may play different regulatory functions. In order to carry out those functions, each class of ncRNA from similar regions may share corresponding specific structures and characteristics, which undergo different evolutionary processes leading to different conservation patterns.

The ncRNA contribution to regulatory networks is complex. Many functional ncRNAs influence chromatin modulation, and regulate gene expression at both transcriptional and post-transcriptional levels (Amaral et al., 2010). Although overwhelming evidence has shown that ncRNAs are pervasively expressed from different genomic regions, and possess a wide range of functionality in gene regulation, these discoveries still provide only a glimpse of the hidden ncRNA world. Well-annotated ncRNAs represent a small fraction of the available datasets and the majority of these annotations are structural. While continued advances in high-throughput sequencing will facilitate the discovery and elucidation of more regulatory ncRNAs, we will need a comparable revolution in high-throughput functional testing of ncRNAs to address the functions and mechanisms of long ncRNAs in regulatory networks.

Frontiers in Genetics | Non-Coding RNA October 2012 | Volume 3 | Article 205 | 8
REFERENCES

Amaral, P. P., Clark, M. B., Gaspar, D. K., Dingee, M. E., and Mattick, J. S. (2010). lncRNADb: a reference database for long non-coding RNAs. Nucleic Acids Res. 38, D145–D151.

Amaral, P. P., and Mattick, J. S. (2008). Noncoding RNA in development. Mamm. Genome 19, 494–492.

Araki, R., Fukumura, R., Sawada, N., Kusum, Y., Suzuki, N., Takehara, H., et al. (2006). More than 40,000 transcripts, including novel and noncoding transcripts, in mouse embryonic stem cells. Cell 125, 2522–2532.

Berezin, G., Peaarea, M., Makimen, L., Stephens, S., Kent, W. I., and Mattick, J. S., et al. (2004). Ultraconserved elements in the human genome. Science 303, 1352–1355.

Belmont, M., Puig, L., Pena, C., Garcia, J. M., Alvarez, A. B., Pena, R., et al. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev. 22, 758–769.

Bertonc, P., Stol, V., Boyee, T. E., Rosenzweig, J. S., Urban, A. K., Zha, X., et al. (2004). Global identification of human transcribed sequences with genome-wide arrays. Science 306, 2242–2246.

Benavides, C., Ceron, M. C., Capcioni, S., and Nickolenko, A. (2003). Post-transcriptional regulation of gene expression by degradation of messenger RNAs. J. Cell Physiol. 195, 356–365.

Birney, E., Stamatoyannopoulos, J., Dutta, A., Giguere, V., Gingeras, T. R., and Mattick, J. S. (2004). Defining tissue-specific coding and non-coding RNAs: challenges and ambiguities. PLoS Comput. Biol. 4, e1000176. doi: 10.1371/journal.pcbi.1000176.

Cabili, M. N., Trapnell, C., Goff, L., Bevilacqua, A., Ceriani, M. C., Capacci-D151.

Be, A., I., Reis, E. M., and Verjovski-Alguersuari, J. (2005). The amazing complexity of the transcriptome of human chromosomes 21 and 22. Genes 14, 331–342.

Kapranov, P., Cavdar, S., Dinko, I., Belikov, S., Strausberg, R. L., Fedor, S. P., et al. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science 296, 916–919.

Kiri, H. H., Fernandez, H. R., Xie, W., and Berchel, J. A. (2008). RNA silencing in Drosophila. FEBS Lett. 579, 5840–5849.

Khali, A. M., Gottstein, M., Huarte, M., Garber, M., Raj, A., Rivina Morales, D., et al. (2008). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U. S. A. 105, 11667–11672.

Kiro, T., Hart, D. E., Ishii, T., Nu, N., and Chourou, G. P. (2010). Non-coding RNA gas1 acts as a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci. Signal. 3, ra8.

Korai, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., et al. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment and silencing of p15INK4B tumor suppressor gene. Oncogene 10, 1956–1962.

Leh, M., Steffen, P. A., and Wutz, A. (2009). Chromosome inactivation sparked by non-coding RNAs. Nat. Rev. 8, 84–99.

Li, J., Zhang, Y., Kong, L., Qi, R., and Wei, L. (2008). Translational antisense transcripts inhibiting non-coding RNAs in 10 species: implications for expression regulation. Nucleic Acids Res. 36, 4855–4864.

Liu, R., Maeda, S., Liu, C., Kain, M., and Edgington, T. S. (2007). A large noncoding RNA is a marker for murine hepatic carcinoma and a spectrum of human carcinomas. Oncogene 26, 835–846.

Lesnoy, R., El Raschi, N., Nakura, H. I., Rio, E. M., and Verjovski-Alguersuari, J. (2008). Conserved tissue

www.fronti.org October 2012 | Volume 3 | Article 206 | 3
expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. Genom. Res. 92, 18–23.

Lounsbury, M., Smirnova, A. S., and Vezzoso-Almeida, S. (2009). Long intronic noncoding RNA transcription: expression or noise selection? Genome Res. 19, 201–208.

Macdonald, N., Kaukonen, O., Orma, R., Gough, J., Frith, M., Ingrum, P. G., et al. (2006). Transcript annotation in FANTOM5: mouse gene catalog based on physical cDNAs. PLoS Genet. 2, e2. doi: 10.1371/journal.pgen.0020062

Malard, R., Pollet, J., and D’Urso, O. F. (2008). Non-protein coding RNA biomarkers and differential expression in cancer: a review. J. Exp. Clin. Cancer Res. 27, 19.

Martin, P. D., Walter, D. P., Expousa, C. A., Drullinger, L. L., Wagner, S. D., Rogel, J. F., et al. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heart development. Mol. Cell. Biol. 28, 499–509.

Martins, J. A., Laplace, R., and Wotton, F. (2004). Intronic transcription is required to trigger the Saccharomyces cerevisiae SEER gene. Nature 429, 571–574.

Martianov, I., Ramalho, A., Serra Barros, A., Chou, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a noncoding interfering transcription. Nature 445, 666–670.

Martin, J. N., and Atwood, G. (1985). Discrete human dihydrofolate reductase gene transcripts present in polyoma virus-infected cells and several hundred nucleotides upstream of the main mRNA start site. Mol. Cell. Biol. 5, 493–510.

Matich, S. J. (2003). Challenging the dogma: hidden genes, non-protein-coding RNAs in complex organisms. Trends Biochem. Sci. 28, 51–59.

Matich, S. J., and Gagen, M. J. (2010). The evolution of controlled multi-task gene networks: the role of introns and other non-coding RNAs in the development of complex organisms. BMC Evol. Biol. 10, 1611–1620.

Matich, S. J., and Makarun, I. V. (2006). Non-coding RNA. Annu. Rev. Genet. 40, 817–829.

Mecer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10, 155–159.

Mecer, T. R., Hellmich, D., Dinger, M. E., Sridharan, K. R., Koruba, D. I., Gasho, R. A., et al. (2010). Expression of distinct RNAs from 5′ untranslated regions. Nucleic Acids Res. 39, 2893–2905.

Mira, F., Kawaguchi, N., Siew, J., Tori- sida, A., Harii, M., Morishita, S., et al. (2006). A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc. Natl. Acad. Sci. U.S.A. 103, 17048–17053.

Nam, J. W., and Bardel, D. (2012). Long non-coding RNAs in C. elegans. Genome Res. doi: 10.1101/ gr.144747.112 [Epub ahead of print].

Ogawa, Y., Sun, B. K., and Lee, J. T. (2002). Analysis of the mouse transcriptome: expression noise or expression variation? Genome Res. 12, 1356–1361.

Oikonomou, V., Farroux, M., Kaukonen, T., Adachi, J., Bono, H., Kendu, S., et al. (2002). Analysis of the mouse transcriptome based on functional annotation of 61,770 full-length cDNAs. Nature 420, 583–575.

Orn, U. A., Derrain, J., Bunev, M., Gerken, S., Gerken, A., Bident, G., et al. (2010). Long non-coding RNAs with enhancer-like function in human cells. Genes Dev. 24, 45–48.

Pandey, R. R., Mondal, T., Mohommed, F., Petrukhin, S., Gotthardt, D., Praveen, A., et al. (2008). Keg-1, a noncoding RNA with a multiple exonic structure with 3′ untranslated regions. FEBS Lett. 584, 4575–4579.

Petruk, S., Sedkov, Y., Riley, K. M., Hinrichs, A., Frith, M. C., and Mattick, J. S. (2006). Rapid annotation of long non-coding RNAs based on genome-wide expression data from human embryonic stem cells. BMC Genomics 7, 320, 1336–1341.

Prensner, J. R., Iyer, M. K., Balasubramanian, S., Kandoth, C. K., Driskill, J. I., et al. (2008). Inter-genic transcription is regulated: expression noise or expression variation? Genome Biol. 9, R43.

Sanchez, M., Capecchi, M., and Plasterk, R. H. (2005). MicroRNA function in animal development. FEBS Lett. 579, 5911–5922.

Wassarman, E., and Plasterk, R. H. (2005). MicroRNA function in animal development. FEBS Lett. 579, 5911–5922.

Weston, G. J., and Kingston, R. E. (2007). HIF1α binds noncoding RNAs to noncoding RNAs: Exp. Cell Res. 313, 1257–1259.

Yao, M. D., Heng, C. C., Lai, K. L., and Chuang, S. S. (2005). Identification and characterization of a novel long non-coding transcript from the opposite strand of Fox_M. Mol. Cell. Biol. 25, 1465–1474.

Zell, F., Li, S., Maxon-Cabell, M. M., Ragus, S., Zeng, L., Mattokas, S., et al. (2010). Molecular interplay of the noncoding RNA ANRIL and methylation hotspots in the 9p21 region: a gene-polycomb CRXK in transcriptional silencing of INK4a. Mol. Cell 38, 662–674.

Zhang, O., and Krohn, E. J. (2007). Noncoding but nonregulatory by nature: long non-coding RNA regulation by complex non-coding RNAs. Oncogene 26, 4484–4449.

Zhang, L., Shi, L., Wu, J., et al. (2005). “Enhancer” noncoding RNAs in yeast regulate gene expression. EMBO J. 24, 1999–2009.

Zuccaro, T., Shulman, T., Hatch, G., Park, R. K., Long, Q., Yu, G., et al. (2000). A highly conserved endonuclease in the Drosophila intergenic region is the site of a conserved
interactions between Dlx genes in the embryonic forebrain. J. Neurosci. 20, 709–721.
Zhang, X., Geiman, R., Makin, A., Zhong, Y., Bai, R. A., Zhou, Y., et al. (2010). Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningiomas pathogenesis and progression. Cancer Res. 70, 2350–2358.
Zhu, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322, 750–756.
Zhong, Y., Zhang, Y., Wang, Y., Zhang, X., Batista, D. L., Geiman, R., et al. (2007). Activation of p53 by MEG3 non-coding RNA. J. Biol. Chem. 282, 24731–24742.
Zhou, Y., Zhong, Y., Wang, Y., Zhang, X., Batista, D. L., Geiman, R., et al. (2007). Activation of p53 by MEG3 non-coding RNA. J. Biol. Chem. 282, 24731–24742.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 August 2012; accepted: 24 September 2012; published online: 09 October 2012.
Citation: Qu Z and Adelson DL (2012) Evolutionary conservation and functional role of ncRNA. Front. Gene. 3:205. doi: 10.3389/fgene.2012.00205
This article was submitted to Frontiers in Non-Coding RNA, a specialty of Frontiers in Genetics.
Copyright © 2012 Qu and Adelson. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and any copyright notices concerning any third-party graphics etc.