Qu, Zhipeng; Adelson, David Louis

© 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 subject to any copyright notices concerning any third-party graphics etc.

Originally published by Frontiers Research Foundation –

PERMISSIONS
http://www.frontiersin.org/Copyright.aspx

Articles and other user-contributed materials may be downloaded and reproduced subject to any copyright or other notices.

As an author or contributor you grant permission to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Authors and Conditions for Website Use and subject to any copyright notices which you include in connection with such materials. The licence granted to third parties is a Creative Common Attribution ("CC BY") licence. Note that for articles published prior to July 2012, the licence may be different and you should check the pdf version of any article to establish what licence was granted. If an article carries only a non-commercial licence and you wish to obtain a commercial licence, please contact Frontiers.

9th August 2013

http://hdl.handle.net/2440/78326
Evolutionary conservation and functional roles of ncRNA

Zhijun Qu and David L. Adelson*
School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, SA, Australia

INTRODUCTION
As the basis of genetics, the "central dogma" describes the genetic information flow of life (Crick, 1970). The functional roles of DNA as the repository genetic information, and protein as the functional incarnation of that information, have been viewed as the dominant molecular roles in the cell for nearly four decades, while RNA was subordinated as a temporary intermediate of this information flow. However, the hypothesis of an "RNA world" proposed by Gilbert (1986) challenged the "central dogma" view while RNA was subordinated as a temporary intermediate of this information flow. The RNA world hypothesis is debated, a hidden "RNA regulatory world" has been proposed in recent studies describing the RNA world hypothesis is debated, a hidden "RNA regulatory world" has been proposed in recent studies describing PERVASIVE TRANSCRIPTION OF ncRNAs
Non-coding RNAs (ncRNAs) are a class of transcribed RNA molecules without protein-coding potential. They were regarded as transcriptional noise, or the byproduct of genetic information flow from DNA to protein for a long time. However, in recent years, a number of studies have shown that ncRNAs are pervasively transcribed, and most of them show evidence of evolutionary conservation, although less conserved than protein-coding genes. More importantly, many ncRNAs have been confirmed as playing crucial regulatory roles in diverse biological processes and tumorigenesis. Here we summarize the functional significance of this class of "dark matter" in terms its genomic organization, evolutionary conservation, and broad functional classes.

PERVASIVE TRANSCRIPTION OF ncRNAs
Rapid development in analytical technologies, such as whole genome tiling 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 (Carinci, 2006; Guttenbach 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 (Birgione 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 3b of the Ensembl Human Genome, there are about 27,000 protein-coding genes, with a total coding length of ~33 Mb, which only occupies ~1.2% of the whole genome. On the other hand, about 1.67 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.
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 (Iinagaki et al., 2005). These results indicate that many mRNA-like ncRNAs are expected to play important roles in the fly. In 2005, approximately 1,500 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 (Akin et al., 2005; Múra 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 (Guttman et al., 2009). Based on the same method, ~3,300 lincRNAs were characterized according to the chromatin-state maps of various human cell types (Khalil 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., 2010). 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 (Guttman et al., 2010). Human transcriptome data from more sources (24 tissues and cell types), allowed the reconstruction of more than 6,000 human lincRNAs (Cabili 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 (Pászti 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 (Namb and Bartel, 2012; Isong et al., 2012).

**EVOLUTIONARY 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.

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

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

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 RNA. 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., 2006).

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 (Cabilly 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 neutrally 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 (Yi 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; Louvar 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
these two ncRNAs is still unclear. Ogawa et al. (2008) reported that murine Xist and Tsix may form Tsix:Xist 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 (69a 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 metazoans, 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 transcriptional elongation of bxd ncRNAs. This is facilitated by the Trithorax complex TAC1, a transcriptional effector that binds to the bxd region (Petrov et al., 2006).

Interaction of promoters with long ncRNAs can also regulate gene expression. One example is a non-coding transcript initiated from the upstream minor promoter of the human ditydrodolate 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 (TFIIIB). This complex acts by dissociating the pre-initiation complex from the major promoter (Figure 3B; Maritianov et al., 2007). Another signal-induced low-copy-number ncRNA, over 200 nt long, named ncRNACCND1s, also mediates the repression of gene expression by promoter interaction (Wang et al., 2008). ncRNACCND1s recruits a key transcriptional sensor of DNA damage, the translocated in liposarcoma (TLS) RNA-binding protein, to the promoter region of cyclin D1 (CCND1). This recruited 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 ncRNACCND1s, which may provide a novel understanding of stimulus-specific expression of ncRNAs (Wang et al., 2008).

In addition to promoter inactivation or activation, an increasing 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 specifically interacts with the Dlx-2 protein to form a complex, which increases the transcriptional activity of the Dlx-5/6 enhancer in a target and homeodomain-specific manner. The stable complex of Evf-2 ncRNA and the Dlx-2 protein has been validated by in vivo assay, indicating that Evf-2 ncRNA regulates transcriptional activity by directly affecting Dlx-2 activity (Figure 3C; Feng et al., 2006). The abundance of such ultraconserved sequences in vertebrate genomes suggests that this mechanism is a common strategy for the regulation of key developmental genes (Rejarano et al., 2004; Sandelin et al., 2004). Another example of this mechanism is SRA a ncRNA that interacts with MyoD, a transcription factor that regulates skeletal myogenesis. Through in vitro and in vivo experiments, Caretti et al. (2006) found that RNA helicases p68/p72, two MyoD-associated proteins, and SRA are co-activators of MyoD. The normal activation of muscle gene expression and cell differentiation are suppressed by RNA interference of SRA, implying that SRA plays an important role in the regulation of developmental gene expression.

Recent experimental evidence has indicated that long ncRNAs could contribute to the complexity of gene expression regulatory networks, where some long ncRNAs might alter global gene expression through a trans-acting mechanism. Using gene chip array analysis, Hill et al. (2006) proposed that human promoters can coordinate the expression of a wide range of gene products at spatially diverse sites in the genome without miRNAs. Their experiments showed that extensive and specific transcriptional activities in epithelial cells (Hela) were influenced by the expression of three intronic sequences derived from the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which was also abnormally expressed. A wide range of genes related to processes of epithelial differentiation and repair were affected as a result of these transcriptional changes, such as FOXF1, sucrase-isomaltase, collagen, interferon, complement, and thrombospondin 1.

Hill et al. (2006) suggested that ncRNAs transcribed from intronic regions were responsible for these changes. In a similar vein, the

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 ei 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 Edes, 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 (Yan et al., 2005). Another natural antisense transcript (NAT) of the Snail1 gene can up-regulate gene expression by forming RNA duplexes in the following fashion. The expression of Zeb2, a transcriptional repressor of E-cadherin, requires an internal ribosome entry site (IRES) derived from a large intron located in the 5′ UTR of the Snail1 gene, whose expression in epithelial cells triggers an epithelial–mesenchymal transition (EMT). The Snail1 NAT overlaps with the 5′ splice site of the large intron and Blatran et al. (2008) found that overexpression of this NAT presented the splicing of the Zeb2 5′-UTR, causing an increase in the expression level of the Zeb2 protein. Many antisense transcripts have been mapped to the introns of mammalian genomes (He et al., 2008; Li et al., 2008) indicating that this type of antisense regulation of alternative splicing may be quite common.

Another aspect of post-transcriptional regulation of gene expression mediated by long ncRNAs is the stabilization of protein-coding RNAs. Adenylate- and uridylate-rich (AU-rich) elements are specific 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-lgH antisense transcript, formed by with bcl-2-lgH 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 p15AS epigenetically silences its sense target gene p15 in leukemia (Yu et al., 2008). The expression of p15AS 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 p15AS (Yu et al., 2008). ANRIL (antisense ncRNA from the INK4A-ARF-INK4B locus), which is regarded as an imprint of p15AS, interacts with chromomob homolog 7 (CBX7), a subunit of the PRC1 protein, and mediates the epigenetic transcriptional repression of its sense locus (Top et al., 2010). Subsequent study revealed that this ncRNA binds to SUZ12 (suppressor of zeste 12 homolog), a component...
of the PRC2, and recruits PRC2 to epigenetically repress p33\textsuperscript{NCRX-en} (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 MysD activity as discussed above (Carette 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 (Prensner 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 (Ji et al., 2003). Subsequent analysis indicated that MALAT1 was overexpressed in five other non-epithelial human carcinomas (Liu et al., 2007). MALAT1 may play important roles in tumor cell invasion and formation of metastases (Teng et al., 2009; Tano et al., 2010). In prostate cancer, a DNA 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 (Bento 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 (Mallardo 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 \textit{MEG3} (maternally expressed gene 3), which functions as a long ncRNA. Although \textit{MEG3} has an open reading frame, it is the folding of \textit{MEG3} RNA that activates p53 expression and selectively regulates p53 target gene expression (Zhou et al., 2007; Zhang et al., 2010).

\textit{MEG3} may play important roles in tumor cell invasion and formation of metastases (Teng et al., 2009; Tano et al., 2010). In prostate cancer, a DNA 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 (Bento 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 (Mallardo 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 \textit{MEG3} (maternally expressed gene 3), which functions as a long ncRNA. Although \textit{MEG3} has an open reading frame, it is the folding of \textit{MEG3} RNA that activates p53 expression and selectively regulates p53 target gene expression (Zhou et al., 2007; Zhang et al., 2010).

\textbf{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 different regions may share corresponding specific structures and characteristics, which undergo different evolutionary processes leading to different conservation patterns.

The \textit{ncRNA} contribution to regulatory networks is complex. Many functional ncRNAs influence chromatin modification, 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 \textit{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.
REFERENCES
Amaral, P. F., Clark, M. B., Gas-
Araki, R., Fukumura, R., Sasaki,

Araki, K., Fukumura, R., Suzuki,
Kamio, Y., Suzuki, N., Takah-

Bejerano, G., Pfeifer, M., Mulkern,

Beltrán, M., Puig, I., Ponsa, C., Garcia,

Bertone, P., Stolc, V., Boyer, T. E.,
Romijn, J. S., Urban, A. E., Zha,
X., et al. (2004). Global identifica-

Birney, E., Stamatoyannopoulos,

Birney, E., Stamatoyannopoulos,

Birney, E., Stamatoyannopoulos,

Birney, E., Stamatoyannopoulos,
expression signatures of intrinsic noncoding RNAs transcribed from human and mouse loci. Genome Res. 19, 18–25.


mRNA conservation and function

"fgene-03-00205" — 2012/10/7 — 14:25 — page 10 — #10

Frontiers in Genetics | Non-Coding RNA October 2012 | Volume 3 | Article 2015 | 10

"fgene-03-00205" — 2012/10/7 — 14:25 — page 10 — #10

expression signatures of intrinsic noncoding RNAs transcribed from human and mouse loci. Genome Res. 19, 18–25.


Mattick, J. S., and Gagen, M. J. (2011). The evolution of controlled multi-

}
Qu and Adelson
ncRNA conservation and function


Zhang, X., Gejman, R., Makin, A.,
(2010). Maternally expressed gene 3,
an imprinted noncoding RNA gene, is
associated with meningioma patho-
genesis and progression. Cancer Res.
70, 2350–2358.

Zhao, J., Sun, B. K., Erwin, J. A., Song,
proteins targeted by a short repeat
RNA to the mouse X chromosome.
Science 322, 750–756.

Conflict of Interest Statement: The
authors declare that the research was
continued in the absence of any com-
cmercial or financial relationships that
could be construed as a potential con-
ict 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 func-
tional roles of ncRNA. Front. Gene.

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 dis-
tributed under the terms of the Creative
Commons Attribution License, which
permits use, distribution and reproduc-
tion in other forums, provided the origi-
nal authors and source are credited and
subject to any copyright notices concern-
ing any third-party graphics etc.