Neuroscience 264 (2014) 25–38

REVIEW

EMERGING EPIGENETIC MECHANISMS OF LONG NON-CODING RNAS

K. SCHAUKOWITCH AND T.-K. KIM *

The University of Texas Southwestern Medical Center, Department of Neuroscience, 5323 Harry Hines Boulevard, Dallas, TX 75390-9111, United States

Abstract—Long non-coding RNAs (lncRNAs) have been increasingly appreciated as an integral component of gene regulatory networks. Genome-wide features of their origin and expression patterns ascribed a prominent role for lncRNAs to the regulation of protein-coding genes, and also suggest a potential link to many human diseases. Recent studies have begun to unravel the intricate regulatory mechanism of lncRNAs occurring at multiple levels. The brain is one of the richest sources of lncRNAs, many of which have already shown a close relationship with genes or genetic loci implicated in a wide range of neurological disorders. This review describes recently emerging mechanistic principles of lncRNA functions to provide neuroscientists with molecular insights that will help future research on lncRNAs in the brain.

This article is part of a Special Issue entitled: Epigenetics in Brain Function

Key words: epigenetics, long non-coding RNAs, enhancer RNAs.

Contents

*Corresponding author. Tel: +1-214-648-1953.
E-mail address: taekyung.kim@utsouthwestern.edu (T.-K. Kim).

Abbreviations: BGC, biased gene conversion; CBP, CREB-binding protein; CE, core enhancer region; EB, embryoid bodies; eRNAs, enhancer RNAs; ESC, embryonic stem cell; GR, glucocorticoid receptor; GRE, glucocorticoid response element; hnRNP-K, heterogeneous nuclear ribonucleoprotein K; IFN, interferon; lncRNAs, long non-coding RNAs; LOF, loss-of-function; NATs, natural antisense transcripts; NF-YA, nuclear transcription factor-Y alpha; NGS, next generation sequencing; PRC2, polycomb repressive complex 2; rDNA, ribosomal DNA; TrxG, trithorax; XCI, X-chromosome inactivation; TTF1, transcription termination factor 1; Xic, X-inactivation center.

INTRODUCTION

Recent advances in genomics technology have led to the explosive discovery of pervasive transcription activity in most of the eukaryotic genomic regions that were once considered "junk DNA" or "dark matter" (Orgel and Crick, 1980; Carninci et al., 2005; Birney et al., 2007; Kapranov et al., 2007; Djebali et al., 2012). The latest ENCODE data collection from 15 cell lines shows that more than 70% of the human genome generates primary transcripts with protein-coding genes only accounting for 2.94% (Djebali et al., 2012). While a portion of the novel transcripts are ascribed to novel isoforms of previously known protein-coding genes, the most prominent finding in the latest human genome annotation is the catalog of a large number of novel long non-coding RNAs (lncRNAs) comprising 9277 manually annotated genes producing 14,880 transcripts (Derrien et al., 2012). Accordingly, lncRNAs have recently drawn intense research efforts with the bright perspective that they may represent a new regulatory layer in the complexity of mammalian gene regulatory networks underlying a wide range of pathophysiology of human diseases. However, the vast majority of lncRNAs exhibit much lower abundance compared with typical protein-coding mRNAs, which raises a precaution that some of them might be the product of transcriptional noise without any biological function (Struhl, 2007; Natoli and Andrau, 2012). Moreover the functional assignments of individual lncRNAs have often been solely based on their expression correlations with neighboring protein-coding genes without any mechanistic understanding and/or functional characterization (Dinger et al., 2008; Mercer et al., 2008; Rinn and Chang, 2012). Therefore, additional work would certainly be necessary before we can accurately appreciate the nature of the lncRNA regulatory networks and their roles in various biological processes as well as human diseases. Despite the limited knowledge of this newly emerging class of RNA molecules, a growing number of studies, especially in the last several years, have begun to elucidate the

0306-4522/13 $36.00
http://dx.doi.org/10.1016/j.neuroscience.2013.12.009

functional mechanisms of lncRNAs. In most cases where the functions of lncRNAs have been relatively well characterized, their prime roles lie at the regulation of gene expression and epigenetic processes in the nucleus. Consistently, analyses of subcellular RNA fractions have shown that lncRNAs are highly enriched in the nucleus, many of which are tightly associated with the chromatin fraction (Khalil et al., 2009; Djebali et al., 2012; Ng et al., 2012). Gene expression in neurons is dynamically controlled by both an intrinsic genetic program and sensory stimulation, which is critical for brain function and development (Lyons and West, 2011). The brain is one of the richest sources for lncRNAs, which are expressed in regional, cellular, and temporal patterns in developing and adult brains (Mercer et al., 2008; Qureshi et al., 2010). However very little is known for the role of lncRNAs in brain development and function through the regulation of nuclear gene expression programs. This review will focus on recently emerging mechanistic principles that underlie the nuclear functions of lncRNAs in order to provide neuroscientists with molecular insights that will help future research on lncRNAs in the brain.

GENOMIC ORGANIZATION AND EXPRESSION FEATURE OF LNCRNAS

LncRNAs are conventionally defined as a transcript longer than 200 nucleotides in length with lack of protein-coding capability (Rinn and Chang, 2012). The idea that RNA itself might be a functional regulatory entity was originally developed from early studies investigating the epigenetic mechanisms of genomic imprinting and X-chromosome inactivation (XCI) (Lyon, 1961; Lee, 2011). In therian mammals, XCI is the mechanism of dosage compensation by which one of two X chromosomes in females is epigenetically silenced to account for the difference in X-linked gene dosage between XX females and XY males. One of the first proto-type lncRNAs, *Xist* is highly expressed from one of the X chromosomes during the onset of XCI (Brown et al., 1992; Clemson et al., 1996). This 17-kb long transcript does not encode any protein but instead coats the inactive X chromosome (Xi) *in cis* and recruits polycomb repressive complex 2 (PRC2) to the Xi through a conserved repeat motif to induce heterochromatin formation, thereby silencing associated genes (Zhao et al., 2008). Later it turned out that *Xist* action is controlled by two additional lncRNAs, *Tsix* and *Jpx*. The negative regulator *Tsix* is transcribed in an antisense orientation from its own promoter, located downstream of the *Xist* gene, and represses *Xist* transcription on one allele (therefore determining the active X chromosome) by means of several mechanisms including the recruitment of DNA methyltransferase 3a (DNMT3A) to the *Xist* promoter region (Lee et al., 1999; Bacher et al., 2006; Xu et al., 2006). The positive regulator *Jpx* appears to be the *Xist* activator as its deletion or knockdown blocks XCI (Chureau et al., 2002; Tian et al., 2010).

The advent of high throughput genomic technologies such as DNA microarray and next generation sequencing (NGS) then enabled systematic interrogation of the lncRNAs in various cell types across developmental stages. One way to classify lncRNAs is based on their relative locations to nearby protein-coding genes. LncRNAs can be intergenic or intragenic, and depending on their origin and orientation, they can also be categorized as divergent or antisense (Kung et al., 2013). However, compared to protein-coding RNAs, lncRNAs are expressed at much lower levels, and their sequences have been subjected to weak evolutionary constraints (Wang et al., 2004; Pang et al., 2006; Ponting et al., 2009; Cabili et al., 2011; Clark et al., 2012; Derrien et al., 2012; Djebali et al., 2012; Tani et al., 2012). These properties of lncRNAs not only imposed a difficulty in reliable detection and accurate assembly of transcript structure, but also raised a concern that a large portion of the novel noncoding transcripts might be the consequence of transcriptional noise (Struhl, 2007; Natoli and Andrau, 2012). In an effort to identify the most likely functional lncRNAs across the genome, chromatin-state maps were used to characterize the genomic origins of lncRNAs. Functional regulatory regions, such as active promoters, enhancers and coding regions, are uniquely decorated by a combination of post-translational modifications occurring at the N-terminal tails of individual histone subunits (Heintzman et al., 2007; Hon et al., 2009; Visel et al., 2009; Wang et al., 2009). Actively transcribing genes are marked by trimethylation of the lysine 4 residue of histone H3 (H3K4me3) at their promoters and trimethylation of lysine 36 of histone H3 (H3K36me3) along the transcribed region. Enrichment levels of these marks are well correlated with the levels of RNA expression. An integrative analysis of the transcriptomes and the K4-K36 domain profiles in four mouse cell types revealed ~1600 intergenic regions that produce multi-exonic lncRNAs with relatively strong purifying selection in their genomic loci [originally termed as ‘‘lincRNAs’’ (large intervening non-coding RNAs), but later on more preferably referred to as ‘‘lncRNAs’’] (Guttman et al., 2009). Attributed to the selection criteria, the genomic origin and structural properties of the lncRNAs identified by this approach are virtually indistinguishable from those of protein-coding mRNAs (Fig. 1). Both classes of RNAs are transcribed by RNA polymerase II (RNAPII) from highly conserved promoters and undergo maturation processes such as 5′ capping, splicing and 3′ polyadenylation. This is also true for well-characterized lncRNAs such as *HOTAIR*, *NRON*, and *Xist* (Brown et al., 1992; Clemson et al., 1996; Willingham et al., 2005; Rinn et al., 2007). Subsequently, the GENCODE consortium within the framework of the ENCODE project performed the most comprehensive analysis of human lncRNAs and identified a total of 9277 lncRNAs (Derrien et al., 2012). The analyses confirmed that lncRNAs show a high degree of similarity with protein-coding genes with regard to the chromatin architecture surrounding their origins, splicing signals, and exon/intron lengths.