Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: http://www.e-biomedicine.com

Review article

Very long non-coding RNA and human disease

Wen-Ling Chan [a,b], Ya-Sian Chang [a], Wen-Kuan Yang [c], Hsien-Da Huang [b,**,1], Jan-Gowth Chang [a,*,1]

[a] Center of RNA Biology and Clinical Application, China Medical University Hospital, Taichung, Taiwan
[b] Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsin-Chu, Taiwan
[c] Cell/Gene Therapy Research Laboratory, Department of Medical Research, China Medical University Hospital, Taichung, Taiwan

ARTICLE INFO

Article history:
Received 27 September 2012
Received in revised form
4 October 2012
Accepted 9 October 2012
Available online 16 November 2012

Keywords:
chromatin
human disease
regulator
very long non-coding RNA (vlncRNA)

ABSTRACT

A role for non-coding RNAs (ncRNAs) in the development of disease has been well documented in the case of miRNAs. Recent studies have shown that long non-coding RNAs (lncRNAs), greater than 200 nt in length, are also implicated in various diseases. In this review, we focus on these lncRNAs, the very long non-coding RNAs (vlncRNAs), which are more than 5 kb long and for which detailed information is available. These studies have demonstrated that vlncRNAs have important biological functions, and that their aberrant expression may result in various cancers. Future investigations in this exciting field are needed to explore the role of vlncRNAs in pathogenesis and, in particular, to further understand their functional mechanisms.

1. Introduction

Classically, proteins are recognized as having the main responsibility for biological function, with RNA merely a messenger that transfers protein-coding information from DNA [1,2]. This concept has changed in recent years, however—whereas only 2% of the genome encodes protein, more than 80% of the genome produces non-protein coding RNA transcripts [1–5], and these non-coding RNAs (ncRNAs) have important biological functions including gene regulation [6,7], imprinting [8–12], epigenetic regulation [13,14], cell cycle control [15], regulation of transcription, translation, and splicing [7,16–20].

There are two major classes of ncRNAs, grouped according to size: small RNA, which includes microRNA (miRNA), PIWI-interacting RNA (piRNA), endogenous short interfering RNA (endo-siRNA), and other ncRNAs less than 200 nt; and long non-coding RNA (lncRNA), which is larger than 200 nt and is transcribed from intergenic, intragenic, or around protein-coding regions. miRNAs are involved in post-transcriptional regulation of mRNA through the RNA-induced silencing complex [16],

* *Corresponding author*. Center of RNA Biology and Clinical Application, China Medical University Hospital, Number 2, Yuh-Der Road, Taichung, Taiwan.
** *Corresponding author*. Institute of Bioinformatics and Systems Biology, Department of Biological Science and Technology, National Chiao Tung University, 75 Bo-Ai Street, Hsin-Chu, Taiwan.
E-mail addresses: bryan@mail.nctu.edu.tw (H.-D. Huang), d6781@mail.cmuh.org.tw, sophia.wlchan@msa.hinet.net (J.-G. Chang).
[1] Hsien-Da Huang and Jan-Gowth Chang made equal contributions to the work.
2211-8020/$ – see front matter
http://dx.doi.org/10.1016/j.biomed.2012.10.001

whereas piRNAs and siRNAs maintain genomic integrity by suppressing transposable elements [17] or other unknown factors in cell nucleus [18]. The lncRNAs are involved in various levels of genome regulation and related fundamental epigenetic processes [19–25].

The importance of lncRNAs in gene regulation has become apparent in recent years [6,20–30], but the key sequences of lncRNA that determine regulatory function remain unknown. Thus, although the rules of translation via the genetic code are well understood and a mutation in a protein-coding gene that contributes to a given disease can be attributed to the resultant change in amino acid sequence [6,20–30], there is no equivalent code for lncRNA function. Genetic studies on lncRNA could help us identify their regulatory sequences and understand their mechanism of action more clearly. In this review, we focus on a specific group of lncRNAs, those more than 5000 nt long, which we call very long non-coding RNAs (vlncRNAs), and explore their roles in the development of disease.

2. vlncRNA annotation and relevant databases

High-throughput technologies such as Tiling Chip or Deep Sequencing data, combined with computational approaches, have identified long, abundantly expressed non-coding transcripts associated with various cancers [13,28,29,31–37]. Currently, ncRNAs are curated by a variety of public databases, such as lncRNAdb (http://lncrnadb.com/), a database of eukaryotic lncRNAs validated by experimental data [38]; RNAdb (http://research.imb.uq.edu.au/rnadb/), a lncRNA set conserved between human and mouse that was used in a high-throughput functional screen [39]; fRNAdb (http://www.ncrna.org/frnadb/index.html), a database hosting a large collection of ncRNA sequence data from public non-coding databases [40]; NONCODE (http://www.noncode.org/NONCODERv3/), a database of non-coding RNAs [41], together with annotation of potential function, based on a coding–non-coding coexpression network [34]; and NRED (http://jsm-research.imb.uq.edu.au/nred/), a database of expression data of human and mouse lncRNAs with various gene expression profiles [31]. Moreover, the Encyclopedia of DNA Elements consortium is annotating lncRNAs using a combination of RNA-seq data, chromatin state maps, and computational approaches [42]. In this review, we survey the vlncRNAs, collected from various ncRNA databases by filtering for sequence length greater than 5000 nt.

3. Function of vlncRNAs

vlncRNAs originate from intronic, exonic, intergenic, intragenic, and promoter regions, and from 3′- and 5′-UTRs and enhancer sequences; they are sometimes bidirectional transcripts [30]. In particular, a large group of vlncRNAs, referred to as natural antisense transcripts, is antisense to known protein-coding genes [43,44]. In the following sections, we will discuss various vlncRNAs with respect to their functions in epigenetic regulation, transcriptional and post-transcriptional regulation, as well as in tumorigenesis (Fig. 1).

4. vlncRNAs control transcription by mediating changes in chromatin structure

The structure of chromatin determines the accessibility of DNA to polymerase II and transcription factors, and is integral to transcriptional control. Chromatin structure can be altered by specific post-translational modification, such as trimethylation of histone H3 lysine 4 (H3K4me3) at gene promoters, whereas H3K36me3 in transcribed regions is linked to gene activation, and H3K9me3, H3K27me3, and H4K20me3 are linked to repression [45] (Fig. 1A–C). There is substantial evidence for an important role of lncRNAs in these processes, and

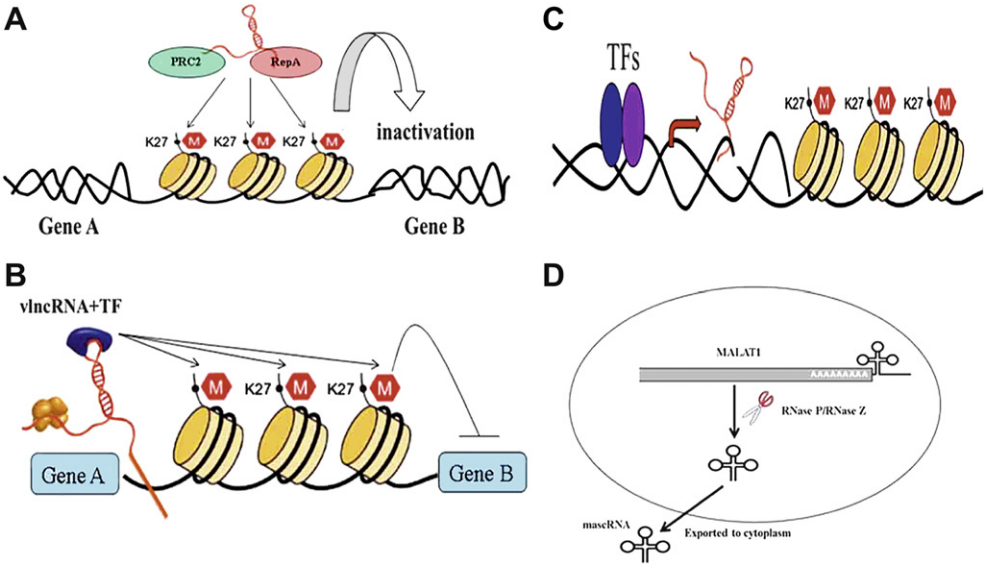

Fig. 1 – Mechanisms for regulation of epigenetics and gene expression by vlncRNAs. (A) vlncRNA regulates histone modification in *cis* and in *trans*. (B) Promoter-associated vlncRNA as an inhibitor of transcription. (C) Promoter-associated vlncRNA as an activator of transcription. (D) vlncRNA generation of sRNA regulatory transcript.