Cancer Letters 361 (2015) 13–21

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Mini-review

Long noncoding RNAs: Novel players in colorectal cancer

Dong Han [a,b,1], Meng Wang [c,1], Ning Ma [a], Ya Xu [a], Yuting Jiang [a], Xu Gao [a,*]

[a] Department of Biochemistry and Molecular Biology, Harbin Medical University, Harbin, China
[b] Heilongjiang Medical Science Academy, Harbin, China
[c] Department of General Surgery, the Second Affiliated Hospital of Harbin Medical University, Harbin, China

ARTICLE INFO

Article history:
Received 15 January 2015
Received in revised form 1 March 2015
Accepted 2 March 2015

Keywords:
Long noncoding RNAs
Colorectal cancer (CRC)
Tumorigenesis
Therapeutic targets

ABSTRACT

Colorectal cancer (CRC) is the most common type of cancer in the world. Despite its commonness, the underlying mechanism of CRC is not completely understood. Long noncoding RNAs (lncRNAs) have received increased attention with the development of whole genome and transcriptome sequencing technologies. Recent findings reveal that lncRNAs are implicated in serial steps of cancer development. These lncRNAs interact with DNA, RNA, protein molecules and/or their combinations, acting as an essential regulator in chromatin organization, and transcriptional and post-transcriptional regulation. In this review, we highlight recent findings of emerging roles for lncRNAs in CRC and discuss rapid translational lncRNA research for clinical application in diagnosis, prognosis and potential treatment.

Introduction

Colorectal cancer (CRC) is one of the most common malignancies in the world. More than 1 million individuals will develop CRC each year, and the disease-specific mortality rate is nearly 33% in the developed world [1]. Initiation of CRC is a complex biological process, involving multiple genomic and epigenomic alterations, occurring over an extended time period of usually a decade. Screening for CRC from curable early stages has the potential to reduce both the incidence and mortality of the disease [2]. Despite substantial progress in understanding the molecular mechanisms and treatment for CRC in recent years, the overall survival rate of CRC patients has not changed dramatically. Intensive investigations over the last few decades have focused on the role of protein-coding genes in the pathogenesis of CRC. Nevertheless, only ~1% of the human genome encodes proteins, leaving another ~4–9% that is transcribed to yield many short or long RNAs with limited protein-coding capacity [3]. Thus, a lot of non-protein coding RNAs (ncRNAs) are transcribed from genome, such as small interfering RNAs (siRNAs), microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), small nucleolar RNA (snoRNA) and long noncoding RNAs (lncRNAs).

lncRNAs are most commonly defined as an RNA transcript of more than 200 nucleotides (nt) and located in nuclear or cytosolic fractions. They are usually transcribed by RNA polymerase II but have no open reading frame and map to intronic and intergenic regions. Moreover, lncRNAs display epigenetic features similar to protein-coding genes, such as trimethylation of histone 3 lysine 4 (H3K4me3) at the transcriptional start site (TSS) and trimethylation of histone 3 lysine 36 (H3K36me3) throughout the gene region [4]. It has been estimated that approximately 15,000 lncRNAs are present in the human genome, but the GENCODE v19 catalog of human lncRNAs contains 13,870 lncRNA genes that produce 23,898 lncRNAs [5]. Recent studies have demonstrated that lncRNAs play important roles in carcinogenesis and cancer metastasis and aberrant expression of lncRNAs has been identified in CRC [6,7]. LncRNAs may function as oncogenes or tumor suppressors in the cancer initiatome [8] and, therefore, CRC can no longer be considered as a simple model of malignancy.

In the present review, we summarize recent progress in the genome-wide analysis of lncRNAs in CRC and the dysregulation of lncRNAs in CRC tissues or cells. We briefly delineate the regulatory network mediated by lncRNAs and the implication of lncRNAs for diagnosis, assessment and treatment of CRC. We suggest that lncRNAs add a novel, but informative layer to our understanding of the complexity of CRC development.

Classification and characteristics of lncRNAs

LncRNAs are non-coding transcripts ranging from 200 to 100,000 nucleotides in length [9]. The estimated number of individual lncRNAs in the human genome sharply increased from 7000 to 23,000 and it is expected to exceed the number of protein-coding genes [10]. LncRNAs are transcribed at any region in the genome by RNA polymerase II/III and are either polyadenylated or non-polyadenylated [11]. In order to gain greater insights into the function of lncRNAs, it will be important to understand the underlying structural features that allow lncRNAs to mediate their biological effects.

* Corresponding author. Tel.: +86 451 86661684; fax: +86 451 8708 6131.
E-mail address: gaoxu_671227@163.com (X Gao).
[1] These authors contributed equally to the work and should be regarded as joint first authors.

http://dx.doi.org/10.1016/j.canlet.2015.03.002
0304-3835/

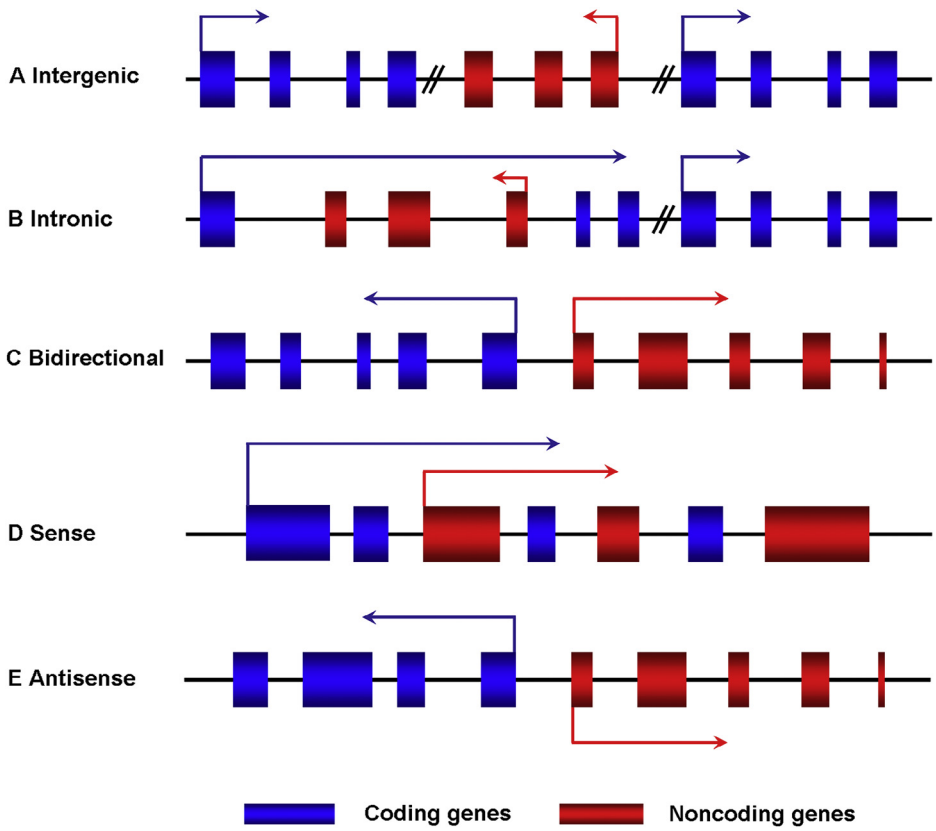

Fig. 1. Overview of five broad categories of lncRNAs. (A) Intergenic lncRNA: lies as an independent unit within the genomic interval between two protein-coding genes. (B) Intronic: transcribed from inside of an intron of a protein-coding gene. (C) Bidirectional: transcribed from the promoter of a protein-coding gene and in opposite direction and, in general, within a few hundred base pairs. (D) Sense or (E) Antisense: transcribed in the same or opposite direction of coding genes, and overlapped with one or more coding exons.

LncRNAs may undergo an alternative splicing procedure like protein coding RNAs and they may mature into secondary and even tertiary structures [12]. Much like proteins, whose structural features are quite conserved across evolution, it is likely that, while lncRNA primary nucleotide sequences may have diverged, their structural elements have remained constant in higher eukaryotes [13]. From the genetic point of view lncRNAs can be classified into the following categories (Fig. 1): (a) Intergenic lncRNAs, also termed large intervening non-coding RNAs or lincRNAs, are lncRNAs with separate transcriptional units from protein-coding genes. One definition required lincRNAs to be 5 kb away from protein-coding genes. (b) Intronic lncRNAs are lncRNAs that initiate inside an intron of a protein-coding gene in either direction and terminate without overlapping exons. (c) Bidirectional lncRNAs are transcripts that initiate in a divergent fashion from the promoter of a protein-coding gene; the precise distance cutoff that constitutes bidirectionality is not defined but is generally within a few hundred base pairs. (d) Sense lncRNAs are lncRNAs whose sequence overlaps with the sense strand of a protein-coding gene. (e) Antisense lncRNAs, which initiate inside or 3′ of a protein-coding gene, are transcribed in the opposite direction of protein-coding genes and overlap at least one coding exon [11].

Recently, researchers concluded two major differences between the spliced and polyadenylated lncRNAs and the messenger RNAs. Firstly, their exon–intron structure of lncRNAs is simpler, with nearly half of lncRNAs only bearing two exons [11]. Secondly, although lncRNAs show exquisite patterns of tissue specificity, their expression levels are significantly lower than those of protein-coding genes [12]. Median expression levels of lncRNAs (steady states of transcripts) are ~10 times lower than those of mRNAs [13]. Importantly, lncRNAs show prominent tissue specificity. These characteristics appear critical for their functional analysis. In past years, along with in-depth studies of lncRNAs, several lncRNA databases have been constructed (Table 1). These databases can facilitate further functional research on lncRNAs.

LncRNAs deregulated in CRC

The cancer transcriptome is more complex than was previously believed. lncRNAs can act as organizational factors of subcellular structures and regulate the localization or activity of proteins. Recent studies have identified large numbers of lncRNAs

Table 1
Public lncRNA databases.

Name	Website	Reference
ChIPBase	http://deepbase.sysu.edu.cn/chipbase/	[14]
Starbase	http://starbase.sysu.edu.cn/	[15,16]
linc2GO	http://www.bioinfo.tsinghua.edu.cn/~liuke/Linc2GO	[17]
mircode	http://www.mircode.org/mircode/	[18]
lnCeDB	http://gyanxet-beta.com/lncedb/	[19]
Diana-lncBase	http://www.microrna.gr/LncBase	[20]
lncRNome	http://genome.igib.res.in/lncRNome/	[21]
fRNAdb	http://www.ncrna.org/frnadb/	[22]
LNCipedia	http://www.lncipedia.org/	[23]
lncRNAdb	http://www.lncrnadb.org/	[24]
NONCODE	http://www.noncode.org/	[25]
NRED	http://nred.matticklab.com/cgi-bin/ncrnadb.pl	[26]
LncRNADisease	http://cmbi.bjmu.edu.cn/lncrnadisease	[27]