Int J Clin Exp Pathol 2014;7(3):948-958
www.ijcep.com /ISSN:1936-2625/IJCEP1312056

Original Article

RNAs specifically affect gene expression in a length, position and sequence dependent manner

Jianjun Cheng[1], Xiufang Wang[1], Nianguang Cai[2], Zhihong Ma[3], Liyan Zhang[2], Zhanjun Lv[1]

[1]Department of Genetics, Hebei Medical University, Hebei Key Lab of Laboratory Animal, Shijiazhuang 050017, Hebei Province, China; [2]Hebei North University, Zhangjiakou 075000, Hebei Province, China; [3]Clinical Laboratory, The Second Hospital of Tangshan, Tangshan 063000, Hebei Province, China

Received December 18, 2013; Accepted February 10, 2014; Epub February 15, 2014; Published March 1, 2014

Abstract: We aim to explore if RNA regulating gene expression is affected by length, sequence and position of RNA. HeLa cells were co-transfected with modulator plasmids (derived from pcDNA3.1 vector containing different length regulating sequences that produce RNAs) and reporter plasmids (derived from pEGFP-C1 vector); In addition, HeLa cells were transfected with plasmids that possess different sequences of downstream or adjacent genes of GFP reporter gene. We found that long inserting sequences of modulator plasmids induced stronger GFP gene activation than short inserting sequences. Changing of downstream sequences of GFP gene induced significant effects on GFP gene expression. Short sequences of adjacent genes of GFP activated GFP gene. Bioinformatics analysis of genes which is highly expressed in differentiating cells (thymocyte cells, germinal center B-cells) and quiescent cells (T cells, B cells) shows that differentiating cells produce longer RNA than quiescent cells. These findings demonstrate that the length, sequence and producing position of RNAs are important factors for RNA regulating gene expression.

Keywords: RNA length, RNA sequence, RNA position, GFP expression

Introduction

The viewpoint of RNAs activating genes was proposed decades ago [1-3], and evidence of this point is accumulating. Transfection of 21-nt dsRNAs targeting promoter regions of human E-cadherin, p21WAF1/CIP1 (p21), and VEGF genes into human cell lines caused long-lasting and sequence-specific induction of those genes [4]. Multiple duplex RNAs complementary to the progesterone receptor (PR) promoter have been shown to increase expression of PR protein and RNA after transfection into cultured T47D or MCF7 human breast cancer cells [5]. Two small activating RNAs induce the expression of mouse Cyclin B1 in NIH/3T3 and TRAMPC1 cells [6]. Duplex RNAs complementary to the promoter of Low-density lipoprotein receptor (LDLR) activate expression of LDLR on the surface of liver cells. Activation requires complementarity to the LDLR promoter and can be achieved by chemically modified duplex RNAs. These data demonstrate that small RNAs can activate LDLR expression and affect LDLR function [7]. Small hairpin RNAs (shRNAs) can activate expression of the vascular endothelial growth factor gene at the promoter level through an epigenetic mechanism [8]. Furthermore, our previous studies have shown that RNA targeting the mouse albumin gene increases sensitivity to DNase I digestion of assembled chromatin [9]. Moreover, exogenous RNA promotes expression of the mouse albumin gene [10], and bioinformatics analysis done by our laboratory also indicates that RNAs can activate genes [11].

The human genome contains abundant non-coding RNA [12]. Although certain non-coding RNA sequences have been found could activate genes and have other biological effects [13], the function of most non-coding RNA sequences is unknown. In this study, we transfected HeLa cells with RNAs produced from modulator plasmids (derived from pcDNA3.1 vector) and analyzed GFP gene expression of reporter plasmids (derived from pEGFP-C1 vector) to study the possibility of RNA as a trans-regulatory factor affecting gene expression, as well as determine if length and sequence of RNA affect

Table 1. Plasmids used in this study

Plasmids	Fragments inserted into pcDNA3.1 or pEGFP-C1 and annotation
C1-4TMI*×2-Alu14	14 copies of Alu elements (283 bp) were inserted downstream of GFP gene in pEGFP-C1 vector to construct C1-Alu×14, then two copies of fragment 4TMI were inserted in sense orientation between of *GFP* gene and *Alu* repeats in C1-Alu×14.
C1-LacZ	*LacZ* (3825 bp) was inserted in sense orientation downstream of *GFP* in pEGFP-C1 (Wang *et al*., 2009).
C1-Alu×8	Eight copies of *Alu* were inserted in sense orientation downstream of *GFP* in pEGFP-C1.
C1-280-1×14as	14 copies of fragment 280-1 (the first 280 bp from L1) were inserted in antisense orientation downstream of *GFP* in pEGFP-C1.
C1-7nt×64	64 copies of fragment AAACAAAAAACAAAAAACAAAAAACA AAAAACAAAAAACA were inserted downstream of *GFP* in pEGFP-C1.
pcAlu×1, pcAlu×14	One or 14 copies of *Alu* were inserted in sense orientation in MCS in pcDNA3.1.
pc280-1×1, pc280-1×14	One or 14 copies of 280-1 fragment were inserted in sense orientation in MCS in pcDNA3.1.
pc280-1×1as, pc280-1×14as	One or 14 copies of 280-1 fragment were inserted in antisense orientation in MCS in pcDNA3.1.
pcLoopAC×4, pcLoopAC×128	Four or 128 copies of fragment CTAGAATTAAACAAATTTTCTAG were inserted in sense orientation in MCS in pcDNA3.1.
pcLoopAC×4as, pcLoopAC×32as, pcLoopAC×128as	Four, 32, or 128 copies of fragment CTAGAAAATTTGTTTAATT CTAG were inserted in antisense orientation in MCS in pcDNA3.1.
pc7nt×64	64 copies of fragment AAACAAAAAACAAAAAACAAAAAACA AAAAACAAAAAACA (without displaying restriction sites) were inserted in MCS in pcDNA3.1.
C1-Alu×1-Alu×14	Alu element was inserted upstream of Alu×14 in C1-Alu×14 vector
C1-280-1×1-Alu×14	280-1 fragment was inserted upstream of Alu×14 in C1-Alu×14 vector
C1-Alu×1-280-1×14	Alu element was inserted upstream of 280-1×14 in C1-280-1×14 vector
C1-280-1×1-280-1×14	280-1 fragment was inserted upstream of 280-1×14 in C1-280-1×14 vector
C1-Alu×1-4TMI×2-Alu×14	Alu element was inserted upstream of 4TMI×2 in C1-4TMI×2-Alu14 vector
C1-280-1×1-4TMI×2-Alu×14	280-1 fragment was inserted upstream of 4TMI×2 in C1-4TMI×2-Alu×14 vector
C1-Alu×1-4TMI×2-280-1×14	Alu element was inserted upstream of 4TMI×2 in C1-4TMI×2-280-1×14 vector
C1-280-1×1-4TMI×2-280-1×14	280-1 fragment was inserted upstream of 4TMI×2 in C1-4TMI×2-280-1×14 vector
C1-280-1×14-CMV#-Alu×2	A CMV promoter and its surrounding sequences was inserted downstream of 280-1×14 in the C1-280-1×14 vector to construct C1-280-1×14-CMV vector, then two copies of Alu were inserted downstream of CMV in C1-280-1×14-CMV plasmid
C1-280-1×14-CMV-280-5×2	Two copies of 280-5 fragments (the fifth 280 bp from L1) were inserted downstream of CMV in C1-280-1×14-CMV plasmid
C1-280-1×14-CMV-280-1×2	Two copies of 280-1 fragments were inserted downstream of CMV in C1-280-1×14-CMV plasmid
C1-280-1×14-CMV-280-1×2as	Two copies of 280-1 fragments in antisense orientation were inserted downstream of CMV in C1-280-1×14-CMV plasmid
C1-280-1×14-CMV-Alu×1-280-1×1	One Alu and one 280-1 fragment in sense orientation were inserted downstream of CMV in C1-280-1×14-CMV plasmid
C1-280-1×14-CMV-Alu×1-280-1×1as	One Alu in sense orientation and one 280-1 fragment in antisense orientation were inserted downstream of CMV in C1-280-1×14-CMV plasmid

*5′GTGAAATAAATGCTTTTTTTGT, having enhancer activity; #from pcDNA3.1+, 1-882 bp, contains intact CMV promoter.

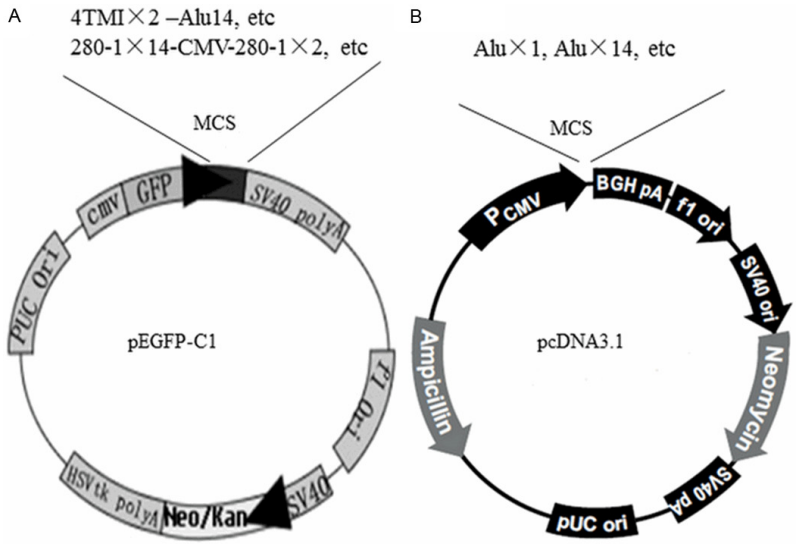

Figure 1. The sites of inserted fragments in reporter and modulator plasmids. A: Location of the inserted sequences in reporter plasmids. 4TMI×2-Alu14 etc sequences were inserted downstream of GFP gene in pEGFP-C1 vector to construct reporter plasmids (**Table 1**). B: Location of the inserted sequences in modulator plasmids. Alu×1, Alu×14 etc sequences were inserted MCS of pcDNA3.1 to construct modulator plasmids (**Table 1**).

gene expression. Many experiments show that RNA transcriptions are located different transcription factories in cell nucleus [14-16], so RNAs in cell nucleus are divisional. In order to study the effects of producing position of RNA on GFP gene expression, we detected the effects of sequences of downstream and adjacent genes on GFP reporter gene. We also use bioinformatics analysis to determine if intron size differs between genes highly expressed in differentiating cells and quiescent cells.

Materials and methods

Construction of expression vectors

The plasmids used in this paper are shown in **Table 1**, including previously constructed plasmids in our laboratory and plasmids constructed in this study. All plasmids were constructed as previously described [17, 18].

Cell culture and cell transfection

HeLa cells were cultured in Minimum Essential Medium (MEM, Gibco, USA) with 10% fetal calf serum. Cells were plated in each well of a 24-well plate with concentration of 0.9×10^5 cells/well and cultured at 37°C in 5% CO_2 for 24 h. The cells were transiently transfected with 0.4 μg modulator plasmids (derived from pcDNA3.1) using 2 μl of Lipofectamine-™2000 reagent (Invitrogen, USA) according to the manufacturer's instructions. The cells were cultured for another 8-10 h and then transfected with reporter plasmids (derived from pEGFP-C1). The transfected cells were used for RNA extraction and fluorescence survey 23-25 h after transfection with reporter plasmids.

Assessment of GFP protein fluorescence

Expression of the GFP protein was assessed by analysis of transfected HeLa cells using fluorescence microscopy of at least 2,000 cells per sample (Nikon TE2000-U, Japan). Images were captured using both white light and fluorescent light within the same scope. Percentage of GFP-positive cells is expressed relative to the total number of cells.

Northern blotting

Total RNA was extracted from transfected HeLa cells and northern blotting was performed to detect GFP RNA expression as described before [19].

Bioinformatics analysis

Gene expression data were obtained from UniGene Library (http://www.ncbi.nlm.nih.gov/UniGene/lbrowse2.cgi?TAXID=9606&CUTOFF=1000). Highly expressed genes were chosen. The data of introns and exons were obtained from Human Genome Resources (http://www.ncbi.nlm.nih.gov/genome/guide/human/) and statistical analysis of length and number of introns was performed using Microsoft Excel software.

Results

Construction of expression vectors

The schemas of constructing reporter and modulator plasmids are shown in **Figure 1**.

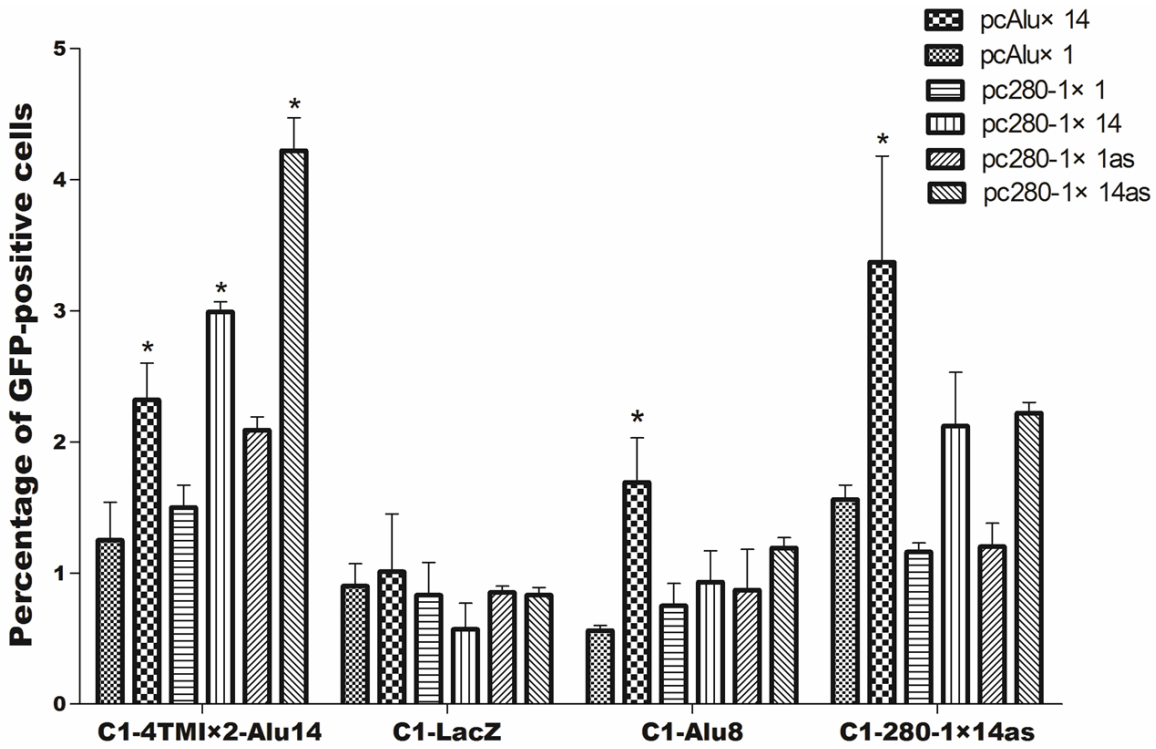

Figure 2. Long RNAs produced by inserting sequences in modulator plasmids activate GFP expression in reporter plasmids. The symbol '*' denotes a significant difference compared to modulator plasmids containing one copy of corresponding inserted sequences.

Figure 1A denotes the site of inserted fragments in reporter plasmids. **Figure 1B** denotes the site of inserted fragments in modulator plasmids. All used plasmids were correct by restriction enzyme analysis and sequencing.

RNAs activate gene in a length- and sequence-dependent manner

HeLa cells were transfected with modulator plasmids and with reporter plasmids 8 h later. Following transfection with reporter plasmids, the cells were cultured for 24 h. Fluorescence microscopy was used to assess the effects of RNAs produced from modulator plasmids on GFP expression in reporter plasmids. Based on GFP expression induced by the modulator plasmids, the reporter plasmids are divided into three types (sensitive to all modulator plasmids, sensitive to partial modulator plasmids, and insensitive to all modulator plasmids) according to whether they activated *GFP* gene in a length-dependent manner (**Figure 2**). More specifically, reporter plasmids C1-4TMI×2-Alu14 and C1-280-1×14as are sensitive to the modulator plasmids pcAlu×14, pc280-1×14, and 280-1×14as that activate GFP expression of reporter plasmids in a length-dependent manner. There is significant statistical difference between percentages of GFP-positive cells induced by pcAlu×1 and pcAlu×14, as well as between ones of pc280-1×1 and pc280-1×14 or pc280-1×1as and pc280-1×14as ($p<0.05$). For example, when C1-4TMI×2-Alu14 is used as the reporter plasmid and pcAlu×1 is used as the modulator plasmid, 1.25±0.29% of transfected cells are GFP positive. However, when pcAlu×14 is used as the modulator plasmid, 2.32±0.28% of transfected cells are GFP positive. Reporter plasmid C1-Alu8 is sensitive only to the modulator plasmid pcAlu×14, which activates GFP expression by the reporter plasmid in a length-dependent manner; C1-Alu8 is not sensitive to the other two modulator plasmids (pc280-1×14, 280-1×14as). The reporter plasmid C1-LacZ is not sensitive to any of the three modulator plasmids.

RNA produced by the modulator plasmids activated the reporter plasmids C1-4TMI×2-

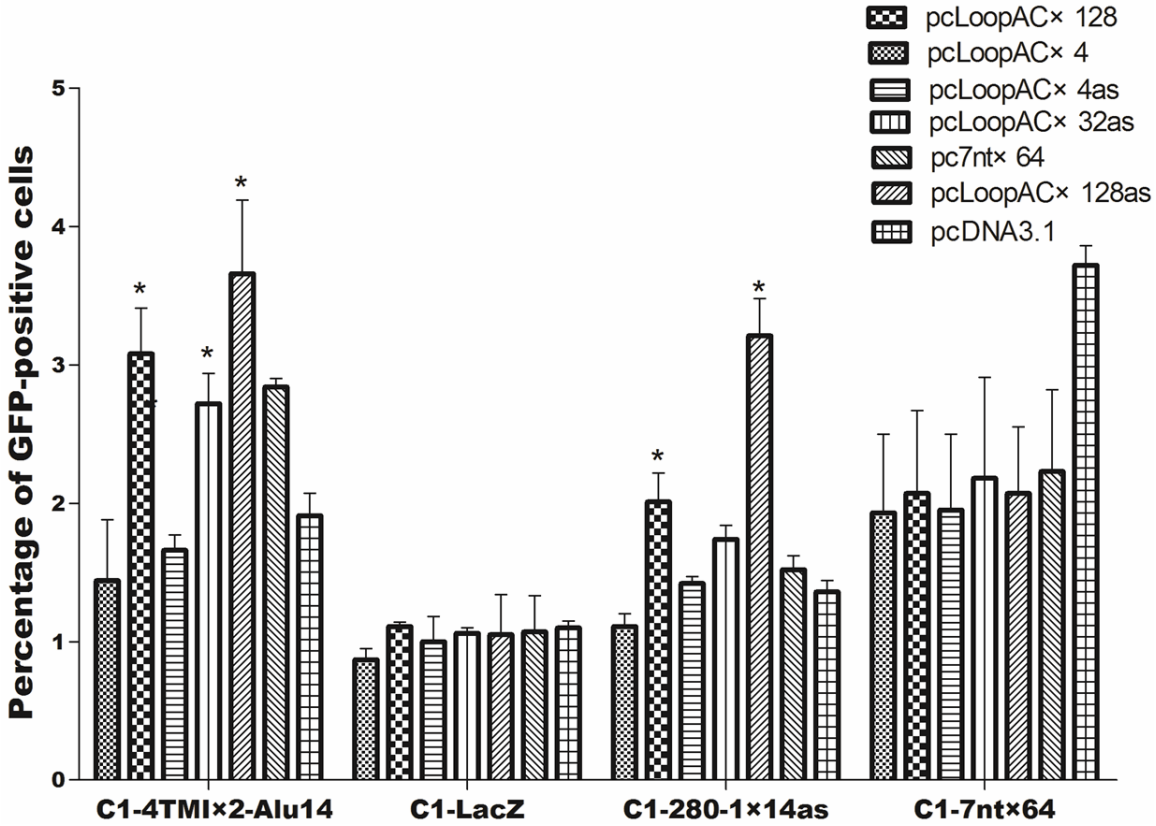

Figure 3. Short tandem sequences in modulator plasmids activate GFP expression in reporter plasmids in a length-dependent manner. The symbol '*' denotes a significant difference compared to modulator plasmids containing four copies of corresponding inserted sequences.

Alu14 and C1-280-1×14as and partially activated the reporter plasmid C1-Alu8. RNA produced by modulator plasmids did not activate the reporter plasmid C1-LacZ, this suggesting that these RNAs display sequence specificity in activating genes. RNAs complementary with DNA (such as, RNA generated by pcAlu×14 is complementary with DNA in C1-Alu8) can activate genes. However, activation of a complete complementarity is not always the strongest. For example, although the inserting sequences of pc280-1×14 and pc280-1×14as are entirely complementary with the inserting sequence of C1-280-1×14as, activation by pc280-1×14 and pc280-1×14as is not the strongest compared to activation by non-complementary sequence in pcAlu×14.

After finding that RNAs specifically activate gene in a length-dependent manner and there is no a positive correlation between activation and sequence complementarily of RNA and DNA, our next question was how RNA activate gene. It is known that longer RNA yields unique secondary and tertiary structures that participate in cellular processes [20]. Thus, we inserted a tandem simple sequence repeat (22nt) forming a stem-loop or a control tandem 7nt (AAACAAA) repeat that does not form a stem-loop into pcDNA3.1 vector to construct modulator plasmids producing RNA with or without stem loops. The results show that RNAs forming stem-loops (produced from pcLoopAC×128) activate GFP in C1-4TMI×2-Alu14, C1-280-1×14as reporter plasmids in a length-dependent manner that is independent of RNA orientation (**Figures 3**, **4**). For example, when C1-4TMI×2-Alu14 is used as the reporter plasmid, and pcLoopAC×4as, pcLoopAC×32as, or pcLoopAC×128as are used as the modulator plasmids, the percentages of GFP-positive cells were 1.66±0.11%, 2.72±0.22%, or 3.66±0.53%, respectively. Thus, the sequences inserted into the modulator plasmids activated GFP expression in the transfected cells in a length-dependent manner.

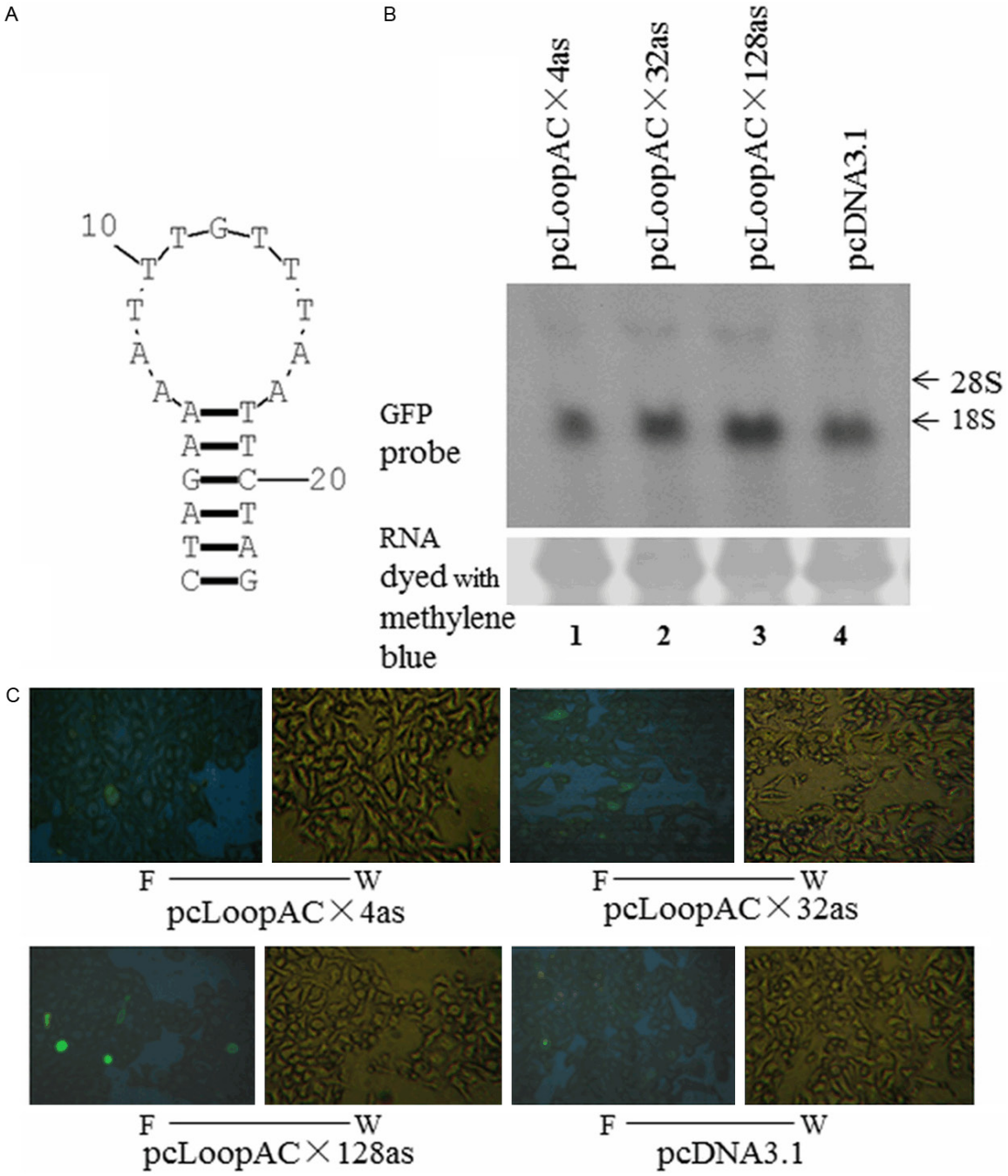

Figure 4. Northern blotting and fluorescence microscopy show that LoopAC (forming stem-loops) in modulator plasmids activate GFP expression in reporter plasmids in a length-dependent manner. A: LoopACas sequence (23nt) is predicted to form stem-loop structure. B: The modulator plasmids were transfected into HeLa cells, followed by transfection of the reporter plasmid C1-4TMI×2-Alu14 8 h later. C: Representative images of GFP protein expression in HeLa cells transfected with modulator plasmids and reporter plasmid C1-4TMI×2-Alu14. F: fluorescent light, W: white light. When C1-4TMI×2-Alu14 is used as the reporter plasmid, and pcLoopAC×4as, pcLoopAC×32as, or pcLoopAC×128as are used as the modulator plasmids, the percentages of GFP-positive cells were 1.66±0.11%, 2.72±0.22% ($p<0.05$, vs 1.66±0.11%), or 3.66±0.53% ($p<0.05$, vs 1.66±0.11%), respectively.

RNAs forming stem-loops do not activate GFP expression significantly in cells containing the C1-LacZ and C1-7nt×64 reporter plasmids, regardless of RNA orientation (**Figure 3**). RNA produced by the modulator plasmid pc7nt×64 activate GFP expression in cells containing the

Table 2. The flanking sequences of enhancer affect GFP gene expression (changing flanking sequences of GFP gene)

Plasmids		Percentages of GFP-positive cells (%)
1	C1-Alu×1-Alu×14	1.00±0.20
2	C1-280-1×1-Alu×14	4.57±1.45
3	C1-Alu×1-280-1×14	2.67±0.40
4	C1-280-1×1-280-1×14	1.77±0.15
5	C1-Alu×1-4TMI×2-Alu×14	12.10±2.85
6	C1-280-1×1-4TMI×2-Alu×14	8.86±1.20
7	C1-Alu×1-4TMI×2-280-1×14	6.30±0.85
8	C1-280-1×1-4TMI×2-280-1×14	5.00±1.49

Table 3. The expression of neighboring genes affects GFP gene expression (changing syntenic neighboring genes of GFP gene)

Plasmids		Percentages of GFP-positive cells (%)
1	C1-280-1×14-CMV-Alu×2	2.02±0.38
2	C1-280-1×14-CMV-280-5×2	1.52±0.22
3	C1-280-1×14-CMV-280-1×2	3.06±0.69
4	C1-280-1×14-CMV-280-1×2as (antisense)	1.10±0.28
5	C1-280-1×14-CMV-Alu×1-280-1×1	4.17±1.01
6	C1-280-1×14-CMV-Alu×1-280-1×1as	1.30±0.53

C1-4TMI×2-Alu14 reporter plasmid, but weakly or do not activate GFP in cells containing the C1-LacZ and C1-7nt×64 reporter plasmids (**Figure 3**).

Northern blot analysis showed that *GFP* gene expression in cells containing the C1-4TMI×2-Alu14 reporter plasmid is weak in cells containing the pcLoopAC×4as modulator plasmid, intermediate in cells containing pcLoopAC×32as, and strong in cells containing pcLoopAC×128as (**Figure 4**). This result is consistent with that of our fluorescence microscopy analysis (**Figure 3**). Staining of 28sRNA with methylene blue was used as the loading control and indicates that a consistent amount of RNA was loaded into each lane. Furthermore, our results of Northern blot analysis show that GFP expression induced by pcDNA3.1 vector is similar to that induced by the pcLoopAC×4as modulator plasmid. RNA activating gene in length dependent manner is not due to inserting fragment long resulting in transfected plasmid number decrease, because pcDNA3.1 has no inserting fragment and transfected plasmid number is the most in the same dose. Moreover, RNAs activate gene in a length-dependent manner and display specificity for certain reporter plasmids, which also illustrates that gene activation by RNA in a length-dependent manner is sequence specific and not due to effects of the number of plasmids transfected into cells.

RNAs activate gene in a position-dependent manner

The results in **Figures 2** and **3** show that long RNAs produced from inserting sequences in modulator plasmids activate GFP gene in reporter plasmids, whereas short RNAs cannot activate GFP gene. We further studied whether RNAs produced from flanking sequences or adjacent genes of GFP gene affect expression of GFP reporter gene.

The results from **Table 2** illustrate that the sequences downstream of GFP gene affect GFP gene expression. The sequences that tightly connect to GFP gene, can be regarded as same gene with GFP, are activated by same promoter with GFP gene. In this case, changing downstream sequences of GFP gene equals to changing the 5′ end in same gene. In **Table 2**, plasmid 1 (C1-Alu×1-Alu×14) caused stronger gene inhibition than plasmid 4 (C1-280-1×1-280-1×14); In the case of possessing enhancer 4TMI, Alu sequence (C1-Alu×1-4TMI×2-Alu×14, plasmid 5) induced stronger GFP gene expression than 280-1 (C1-280-1×1-4TMI×2-280-1×14, plasmid 8); We analyze that the effects of downstream sequences on GFP gene expression are related with produced RNAs (RNAs produced from gene itself affect its expression).

Another group of plasmids were constructed (**Table 3**). A CMV promoter was inserted downstream of 280-1×14 in C1-280-1×14 vector, then different sequences were inserted downstream of CMV promoter to construct plasmids (equivalently changing the sequences of adjacent genes of GFP gene). The results from **Table 3** show that the two copies of inserted sequences downstream of CMV in

Table 4. The size of high expression genes and their intron size in four immune cells

	Number of genes	Gene size (bp)①				Intron	
		≤14999	15 000~29999	30 000~59999	≥60000	Number mean±SD② (range)	bp mean±SD (range)
Thymocyte cells	50	30 (60)	8 (16)	6 (12)	6 (12)*	8.043±6.606 (1~34)	2508.95±6291.89# (45~75833)
T cells	48	39 (81)	6 (13)	3 (6)	0 (0)	7.707±3.823 (3~21)	1065.23±1625.73 (39~12006)
Germinal center B-cells	51	22 (43)	4 (8)	10 (20)	15 (29)**	9.319±7.059 (1~35)	7137.52±16145.07## (48~133546)
B cells	49	43 (88)	1 (2)	4 (8)	1 (2)	6.659±2.623 (2~14)	1756.93±5324.11 (67~59687)

①the number in bracket is the percentage of relevant size genes; ②the mean of intron of each gene; *: X^2 test *vs* T cells $p<0.01$, **: X^2 test *vs* B cells $p<0.01$; #: *t* test *vs* T cells $p<0.01$, ##: *t* test *vs* B cells p<0.01.

plasmids 3 (C1-280-1×14-CMV-280-1×2) and 5 (C1-280-1×14-CMV-Alu×1-280-1×1) remarkably increased GFP gene expression. Modulator plasmids containing same two copies of inserted sequences did not significantly increase GFP gene expression (data not shown). The results from **Tables 2** and **3** preliminary illustrate that not only RNA sequences, RNA producing positions are the important factors of affecting gene expression.

High expression of large genes in differentiating cells

We found that RNA activates gene expression in a length-dependent manner, which suggests a relationship between length of a non-coding RNA and its biological function. One significant difference between eukaryotic and prokaryotic cells is that eukaryotic DNA has introns, called intervening sequences, which can separate exons of coding DNA. The size of genes is generally decided by the size of introns. If long RNA is related to gene activation, the differentiating cells should contain more long RNA because silent genes need to be activated in these cells. We used bioinformatics analysis to test this hypothesis.

The data of gene expression strength in tissues or cells are from UniGene Libraries (http://www.ncbi.nlm.nih.gov/UniGene/lbrowse2.cgi?TAXID=9606&CUTOFF=1000). Gene sequence data were downloaded from NCBI GeneBank (http://www.ncbi.nlm.nih.gov/genome/guide/human). We compared the numbers of large genes (≥60000 bp) and intron size of highly expressed genes between thymocytes and T cells, or germinal center B-cells and B cells. We found that number of large genes and intron size were higher in thymocytes than in T cells, and in germinal center B-cells than in B cells (**Table 4**). Thymocytes and germinal center B-cells are differentiating cells, in which silent genes are being activated. Our findings that large genes with large introns are more highly expressed in the differentiating cells than in the mature cells are consistent with our hypothesis and the other results in this paper.

Discussion

Non-coding sequences of DNA and RNA are ubiquitous in multicellular eukaryotes [21, 22]. The biological function of non-coding sequences is a hot issue in the biology community [23]. Our finding that RNA activates gene in a length, sequence and position dependent manner contributes to solving the questions of biological function of non-coding sequences. Here, we report that long RNAs activate gene more strongly than short RNAs. The modulator plasmid pc280-1×14as activates GFP expression 2.02 times more strongly than the plasmid pc280-1×1as in cells transfected with the reporter plasmid C1-4TMI×2-Alu14; the modulator plasmid pcLoopAC×14as activates GFP expression 2.26 times more strongly than pcLoopAC×4as in cells transfected with the reporter plasmid C1-280-1×14as.

Injected RNAs distributed throughout the nucleoplasm in early stage and then progressively localized in discrete loci [24]; the expression of adjacent genes is physiologically relevant [25-27]; the analysis of high-throughput transcripts illustrates that the expression of genes located within same chromosome is

positively correlated [28] and RNA transcription is divisional [14-16]. The above reports precipitate us to prove the effects of RNA producing positions on gene expression using a simple model system.

RNA producing positions can be divided into gene itself, adjacent genes and trans-genes. The results from **Table 2** illustrate that the sequences downstream of GFP gene affect GFP gene expression (equal to changing gene itself sequences). When reporter plasmids contain 4TMI enhancer, Alu induced stronger GFP gene expression than 280-1 fragment, whereas when reporter plasmids do not contain 4TMI enhancer, 280-1 induced stronger GFP gene expression than Alu element. Literatures report that Alu inhibits gene expression [29] or that Alu enhances gene expression [30, 31]. We analyze that that different surrounding sequences result in different results.

A CMV promoter was inserted downstream of 280-1×14 in C1-280-1×14 vector, then different sequences were inserted downstream of CMV promoter to construct plasmids (equivalently changing the sequences of adjacent genes of GFP gene). The results from **Table 3** show that the two copies of inserted sequences (280-1×2 and Alu×1-280-1×1) downstream of CMV enhance GFP gene expression. In **Figures 2** and **3**, we have demonstrated that short RNAs (one copy of 280nt) produced from trans-genes did not activate GFP reporter gene. When two copies of sequences (equal to downstream sequences of CMV) were inserted in modulator plasmids, RNAs produced from these modulator plasmids did not activate GFP gene expression in reporter plasmids, which is same as the results of **Figures 2** and **3**. The results from **Tables 2** and **3** preliminary illustrate that not only RNA sequences and RNA producing positions are the important factors of affecting gene expression. Since transcription is divisional, RNAs produced from adjacent genes interact more easily, which provides the mechanism of expression of adjacent genes relevant.

In general, the more complex an organism, the greater it's number of non-coding RNAs [32]. Intron RNA is an important source of non-coding RNA [33]. The average length of introns differs between species of multicellular eukaryotes (average length of intron in human>chicken>drosophila melanogaster), which suggests that intron size could be related to biological evolution.

Recently, several papers reported that long RNA transcripts change chromatin confirmation and affect cellular differentiation [34, 35]. In this paper, bioinformatics analysis showed that there are longer introns in highly expressed genes of differentiating cells. The average sizes of introns in genes highly expressed in thymocytes and T cells are 2508.95 bp and 1065.23 bp ($p<0.01$), respectively. The average sizes of introns in genes highly expressed in germinal center B-cells and B cells are 7137.52 bp and 1756.93 bp, respectively ($p<0.01$, **Table 4**). There are more long RNAs in differentiating cells than in mature counterparts, which suggests that RNA length is related to gene activation. This finding is consistent with our finding that gene activation by RNA is dependent upon length of the RNA.

Conclusion

Different DNA sequences were inserted downstream of *GFP* gene in pEGFP-C1 vector to construct reporter plasmids. The reporter plasmids were co-transfected with modulator plasmids that produce different RNAs. The effects of RNAs on *GFP* reporter gene expression were observed (*GFP* reporter gene is located in different plasmids with genes producing modulator RNAs). CMV promoter was inserted downstream of *GFP* reporter gene to drive different DNAs to produce RNAs and then the effects of these RNAs on *GFP* reporter gene were observed (*GFP* reporter gene is located in same plasmids with genes producing modulator RNAs). Our results demonstrated the following three points: (1) long RNAs induced stronger *GFP* reporter gene expression than short RNAs; (2) different RNA sequences induced different effects on *GFP* reporter gene expression; (3) RNAs transcribed from adjacent gene have stronger gene activation than RNAs produced from co-transfection.

Acknowledgements

This work was supported by grants from the Hebei Province Natural Science Foundation of China (H2013206101 and C2011206043) and the Key Project of Hebei Province (08276101D-90) and Scientific Research Funding from the

Department of Public Health of Hebei Province (20100242) and the Funding from Hebei Education Department (Q2012004).

Disclosure of conflict of interest

None.

Address correspondence to: Xiufang Wang or Zhanjun Lv, Department of Genetics, Hebei Medical University, Hebei Key Lab of Laboratory Animal, Shijiazhuang 050017, Hebei Province, China. Tel: +86-311-86266844; E-mail: xfang_wang@163.com (Xiufang Wang); zhanjunlv@hotmail.com (Zhanjun Lv)

References

[1] Britten RJ and Davidson EH. Gene regulation for higher cells: a theory. Science 1969; 165: 349-357.

[2] Brawerman G. A model for the control of transcription during development. Cancer Res 1976; 36: 4278-4281.

[3] Dickson E and Robertson HD. Potential regulatory roles for RNA in cellular development. Cancer Res 1976; 36: 3387-3393.

[4] Li LC, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, Enokida H and Dahiya R. Small dsRNAs induce transcriptional activation in human cells. Proc Natl Acad Sci U S A 2006; 103: 17337-17342.

[5] Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE and Corey DR. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. Nat Chem Biol 2007; 3: 166-173.

[6] Huang V, Qin Y, Wang J, Wang X, Place RF, Lin G, Lue TF and Li LC. RNAa is conserved in mammalian cells. PLoS One 2010; 5: e8848.

[7] Matsui M, Sakurai F, Elbashir S, Foster DJ, Manoharan M and Corey DR. Activation of LDL receptor expression by small RNAs complementary to a noncoding transcript that overlaps the LDLR promoter. Chem Biol 2010; 17: 1344-1355.

[8] Turunen MP, Lehtola T, Heinonen SE, Assefa GS, Korpisalo P, Girnary R, Glass CK, Vaisanen S and Yla-Herttuala S. Efficient regulation of VEGF expression by promoter-targeted lentiviral shRNAs based on epigenetic mechanism: a novel example of epigenetherapy. Circ Res 2009; 105: 604-609.

[9] Lv ZJ, Wang XF, Zhai Y and Song SX. [RNA responsible for conferring a DNase I sensitive structure on albumin gene in assembled chromatin]. Yi Chuan 2003; 25: 30-36.

[10] Zhang HY, Liu AJ, Liu FY, Wang XF and Lu ZJ. [Influence of exogenous RNA upon expression of mouse albumin gene and its sensitivity to DNase I]. Yi Chuan Xue Bao 2002; 29: 26-29.

[11] Li ZJ, Song SX, Zhai Y, Hou J, Han LZ and Wang XF. Genome sequence comparative analysis of long arm and short arm of human X chromosome. Yi Chuan Xue Bao 2005; 32: 1-10.

[12] Knowling S and Morris KV. Non-coding RNA and antisense RNA. Nature's trash or treasure? Biochimie 2011; 93: 1922-1927.

[13] Rother S and Meister G. Small RNAs derived from longer non-coding RNAs. Biochimie 2011; 93: 1905-1915.

[14] Bentley D. The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol 2002; 14: 336-342.

[15] Edelman LB and Fraser P. Transcription factories: genetic programming in three dimensions. Curr Opin Genet Dev 2012; 22: 110-114.

[16] Deng B, Melnik S and Cook PR. Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy. Semin Cancer Biol 2013; 23: 65-71.

[17] Wang HG, Wang XF, Jing XY, Li Z, Zhang Y and Lv ZJ. Effect of mutations in a simian virus 40 PolyA signal enhancer on green fluorescent protein reporter gene expression. Genet Mol Res 2011; 10: 1866-1883.

[18] Lv Z, Cheng J, Xie Y, Jing X, Zhang Y and Wang X. Finding of IFNgamma gene enhancers and their core sequences. Genome 2013; 56: 147-154.

[19] Yin K, Wang X, Ma H, Xie Y, Feng J, Yang Q and Lu Z. Impact of copy number of distinct SV40PolyA segments on expression of a GFP reporter gene. Sci China Life Sci 2010; 53: 606-612.

[20] Stark BC, Kole R, Bowman EJ and Altman S. Ribonuclease P: an enzyme with an essential RNA component. Proc Natl Acad Sci U S A 1978; 75: 3717-3721.

[21] Chen G, Yin K, Shi L, Fang Y, Qi Y, Li P, Luo J, He B, Liu M and Shi T. Comparative analysis of human protein-coding and noncoding RNAs between brain and 10 mixed cell lines by RNA-Seq. PLoS One 2011; 6: e28318.

[22] Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud

K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R and Gingeras TR. Landscape of transcription in human cells. Nature 2012; 489: 101-108.

[23] Costa MC, Leitao AL and Enguita FJ. Biogenesis and Mechanism of Action of Small Non-Coding RNAs: Insights from the Point of View of Structural Biology. Int J Mol Sci 2012; 13: 10268-10295.

[24] Wang J, Cao LG, Wang YL and Pederson T. Localization of pre-messenger RNA at discrete nuclear sites. Proc Natl Acad Sci U S A 1991; 88: 7391-7395.

[25] Gierman HJ, Indemans MH, Koster J, Goetze S, Seppen J, Geerts D, van Driel R and Versteeg R. Domain-wide regulation of gene expression in the human genome. Genome Res 2007; 17: 1286-1295.

[26] Ebisuya M, Yamamoto T, Nakajima M and Nishida E. Ripples from neighbouring transcription. Nat Cell Biol 2008; 10: 1106-1113.

[27] Flagfeldt DB, Siewers V, Huang L and Nielsen J. Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. Yeast 2009; 26: 545-551.

[28] Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J and Guigo R. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012; 22: 1775-1789.

[29] Ebihara M, Ohba H, Ohno SI and Yoshikawa T. Genomic organization and promoter analysis of the human nicotinic acetylcholine receptor alpha6 subunit (CHNRA6) gene: Alu and other elements direct transcriptional repression. Gene 2002; 298: 101-108.

[30] Tomilin NV. Control of genes by mammalian retroposons. Int Rev Cytol 1999; 186: 1-48.

[31] Brosius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 1999; 238: 115-134.

[32] Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y and Mattick JS. Tiny RNAs associated with transcription start sites in animals. Nat Genet 2009; 41: 572-578.

[33] Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C and Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489: 57-74.

[34] Chisholm KM, Wan Y, Li R, Montgomery KD, Chang HY and West RB. Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. PLoS One 2012; 7: e47998.

[35] Hu W, Alvarez-Dominguez JR and Lodish HF. Regulation of mammalian cell differentiation by long non-coding RNAs. EMBO Rep 2012; 13: 971-983.