

MiR-3960 binding sites with mRNA of human genes

Anatoly Ivashchenko*, Olga Berillo, Anna Pyrkova, Raigul Niyazova & Shara Atambayeva

National Nanotechnology Laboratory, al-Farabi Kazakh National University, Almaty-050038, Kazakhstan; Anatoly Ivashchenko - Email: a_ivashchenko@mail.ru, *Corresponding author

Received June 04, 2014; Accepted July 07, 2014; Published June 22, 2014

Abstract:

The importance of miRNA in cellular regulation is gaining momentum. Therefore, it is of interest to study miRNA in human genes. Hence, the human mRNA sequences (12,175) were searched for miRNA binding sites and 2,563 predicted sites were found. We observed that the miR-3960 has more than 1000 mRNA binding sites with high affinity (with $\Delta G/\Delta G_m$ values greater than or equal to 90%) for 375 genes. The miR-3960 has 565 binding sites in the 5'UTRs and 515 sites in the CDS of mRNAs. Nucleotide sequences of the binding sites in CDS encode for polyalanine or polyproline. It is observed that miR-3960 has binding sites in 73 mRNAs of target genes encoded transcription factors. Thus, we document predicted properties (polysites, sites in CDS) of uncharacterized miR-3960 binding sites. The studying of the miRNA properties is important for creation of diagnostic methods of cancer.

Keywords: MicroRNA, Genes, mRNA, miR-3960.

Background:

The importance of miRNA studies is steadily increasing over two decades [1]. It is known that many miRNAs regulate the expression of genes involved in tumorigenesis [2-4]. For example, changes in miRNA concentrations have been observed during the development of lung cancer [2], breast cancer [3] and other cancers [4]. It has been proposed that miRNA binding to mRNAs occurs mainly in the 3'UTRs of mRNAs [5]. However, it has been shown, that miRNA binding sites can be located in 5'UTR and CDS of several mRNAs [6, 7]. Therefore, it is important to document a comprehensive study on miRNA using available data in the public domain. Hence, we studied miR-3960 with preferential binding to 5'UTR and CDS of several genes. Most of these genes are involved in the development of cancer; therefore, it is necessary to identify the features for miRNA binding sites. Identification of new interactions with miR-3960 in mRNAs will increase our understanding of the post-transcriptional regulation of gene expression.

Methodology:

Sequence dataset

The complete nucleotide sequences of precursor mRNAs (pre-mRNAs) of human genes were downloaded from NCBI GenBank (<http://www.ncbi.nlm.nih.gov>) in FASTA format. The nucleotide sequences of human mature miR-3960 were downloaded from the miRBase database (<http://mirbase.org>) [8].

Sequence Analysis

Nucleotide sequences of pre-mRNAs were shortened into mature mRNA sequences by Lextractor002 script (<http://sites.google.com/site/malaheene/software>). This script creates a file with mRNA sequence and indicates its bounds between 5'UTR, CDS, and 3'UTR according to GenBank information about each gene.

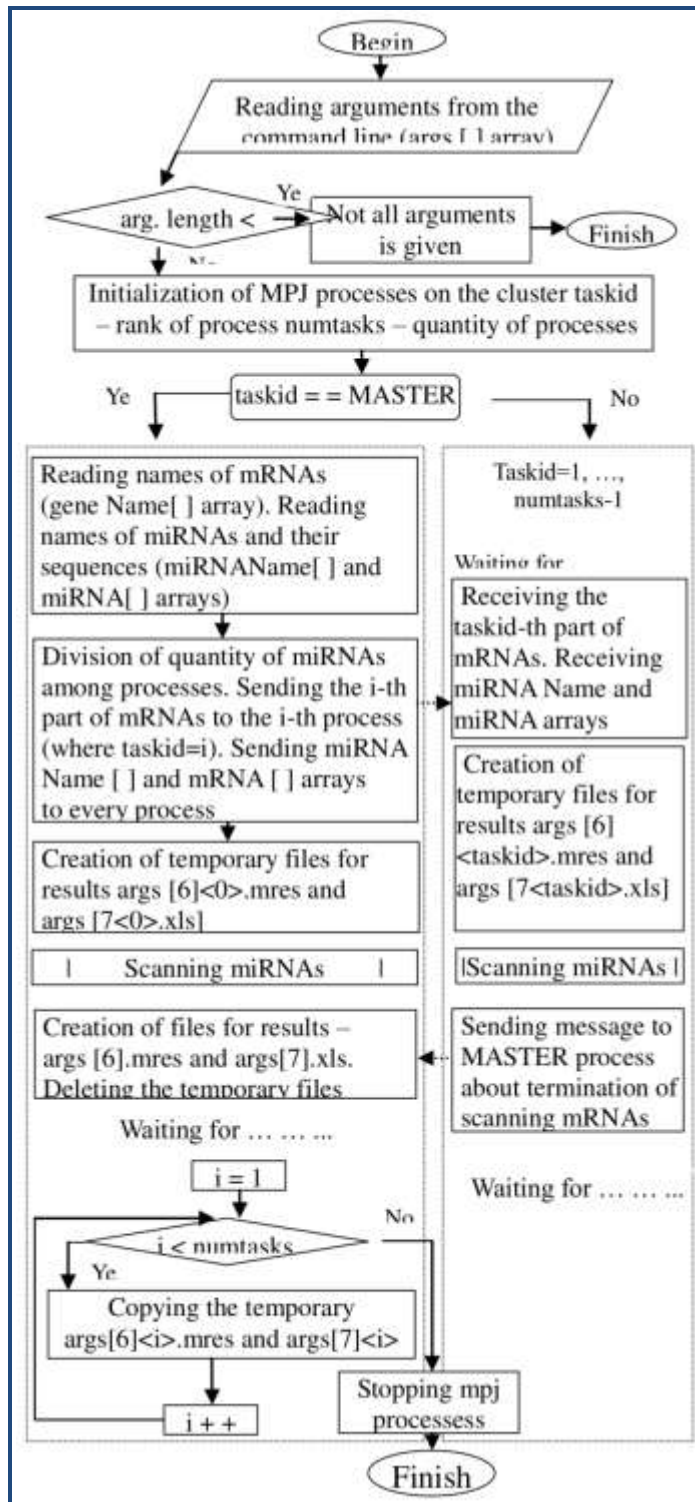


Figure 1: Workflow to MirTarget program (parallel algorithm).

Target prediction using MirTarget program

The target genes for the tested miRNAs were identified using the MirTarget program, which was developed in our laboratory. The workflows to scanning mirna are shown in (Figures 1 & 2)

This program defines the following features of binding sites: a) the start point of an miRNA binding site in mRNA; b) the localization of miRNA binding sites in the 5'-untranslated region (5'UTR), coding domain sequence (CDS) and 3'UTR of ISSN 0973-2063 (online) 0973-8894 (print) Bioinformation 10(7):423-427 (2014)

the mRNAs; c) the free energy of hybridization (ΔG , kJ/mole); and d) the schemes of nucleotide interactions between the miRNAs and the mRNAs. The $\Delta G/\Delta G_m$ (%) ratio was counted for each site, where ΔG_m equalled the free energy of a miRNA binding with its perfect complementary nucleotide sequence. The miRNA binding sites located on the mRNAs had $\Delta G/\Delta G_m$ ratios of 90% or more. We also noted the position of the binding sites on the mRNA, beginning from the first nucleotide of the mRNA's 5' UTR. The MirTarget program computed the interactions between the nucleotides of the miRNA and those of the target gene mRNAs. It was found that there are hydrogen bonds between adenine (A) and uracil (U), guanine (G) and cytosine (C), and G and U, as well as between A and C [9]. The distance between A and C was equal to the G-C, A-U, and G-U distances [10]. The numbers of hydrogen bonds in the G-C, A-U, G-U and A-C interactions were found to be 3, 2, 1 and 1, respectively. The free binding energies of these nucleotide pairs were accepted as the same values (3:2:1:1).

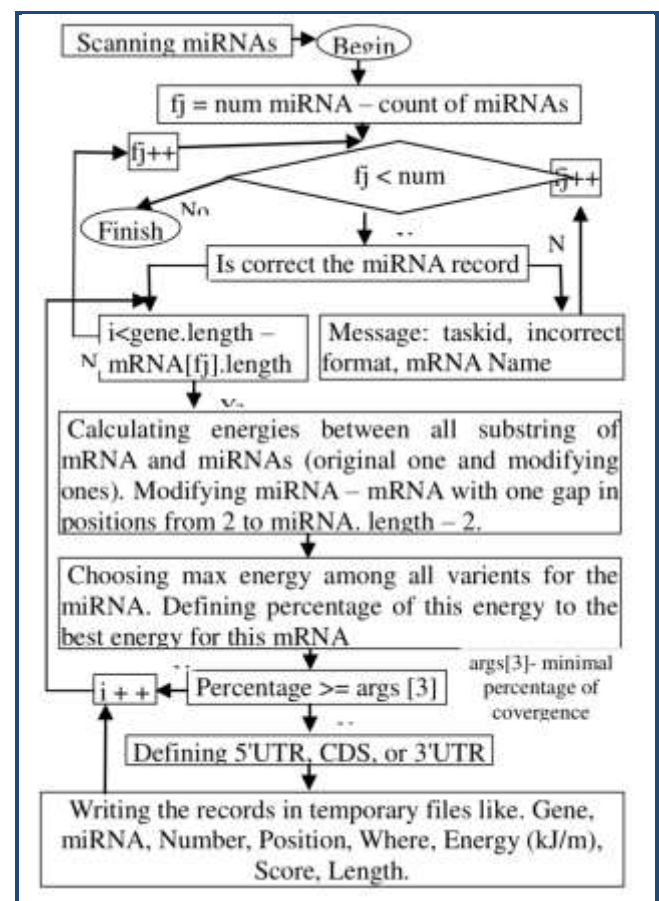


Figure 2: Workflow to scanning mRNAs.

Results:

Features of miR-3960 binding with mRNAs target genes

miR-3960 is in an exon of the cyclin-dependent kinase 9 gene, located on chromosome 9. miR-3960 has a length of 20 nt and a GC-content of 96%. The maximum free energy of miR-3960 binding to mRNA is -125.3 kJ/mole. We found that miR-3960 has 1100 binding sites on 375 target mRNAs with $\Delta G/\Delta G_m$ values of 90% or more. Of those, 20 miR-3960 binding sites are located in 3'UTRs, 565 sites are located in 5'UTRs and 515 sites are located in CDSs. miR-3960 has 19 binding sites in the

mRNA of the *RHOT1* gene, 18 binding sites in the mRNAs of the *AFF2* and *KIF3B* genes, 16 binding sites in the 5'UTRs of the mRNAs of the *DAZAP1* and *BCL2L11* genes and 14 binding sites in the CDSs of the mRNA of the *BTBD2* gene. Approximately half of miR-3960 binding sites are located in the protein-coding regions of mRNAs. The $\Delta G/\Delta G_m$ value for all miR-3960 binding sites in the mRNA of target genes is more than 90% and is 98.2% for the mRNAs of the *ADRB1*, *CEBPB*, *RANBP9* and *SKOR2* genes. The nucleotide sequences of the miR-3960 binding sites in the CDSs of 47 genes are shown in **Figure 3**. The binding sites are clearly conserved as indicated by comparing them with the adjacent nucleotide sequences of the CDSs.

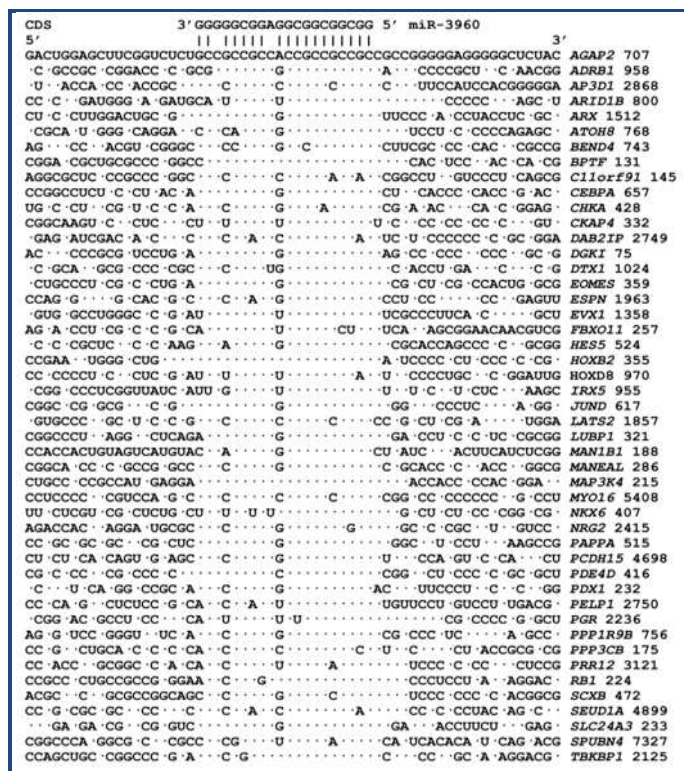


Figure 3: Aligned nucleotide sequences in the CDS of mRNAs containing miR-3960 binding sites. We selected 47 of 375 mRNA of target genes which have single, not regular sites. The number after gene name is start point of the miRNA binding site.

The characteristics of the mRNA binding sites of miR-3960 in several genes are shown in **Table 1** (see **supplementary material**). The presence of several miR-3960 binding sites within a single mRNA increases the probability that miR-3960 will bind to the mRNA and increases its ability to inhibit protein synthesis. The oligonucleotide sequences of the binding sites located in CDSs encode for oligopeptides containing mainly alanine or proline depending on the reading frame **Table 1**. These data indicate the importance of preserving the nucleotide sequences of the miR-3960 binding sites but not the amino acid sequences of the oligopeptides in the encoded protein. The mRNAs of some target genes have miR-3960 binding sites located with an interval of one, two or three nucleotides. 73 out of 2109 genes of transcription factors are targets for miR-3960. For example, the *ZIC5* gene encodes a transcription factor, and its mRNA has 12 miR-3960 binding

sites located through one or two nucleotides. Consequently, the probability of miR-3960 binding with the *ZIC5* mRNA is high. The miR-3960 binding sites in the *ZIC5* mRNA encode an oligopeptide predominantly composed of proline. The mRNA of several genes also have conserved nucleotide sequences at their miR-3960 binding sites in comparison with the amino acid composition of the corresponding oligopeptides. Thus, the oligonucleotides of the miR-3960 binding sites encode for oligopeptides with different amino acid sequences, as shown in **Table 1** (see **supplementary material**).

Many target genes contain miR-3960 binding sites in the 5'UTRs of their mRNAs; this includes 565 binding sites. This localisation of miR-3960 binding sites allows miR-3960 to decrease the protein synthesis of those genes. The nucleotide sequences of miR-3960 binding sites in the 5'UTRs of mRNAs of target genes are highly conserved. mRNAs of many genes have miR-3960 binding sites in their 5'UTRs with displacement in one, two or three nucleotides. The number of the miR-3960 binding sites varied from five to 18. The $\Delta G/\Delta G_m$ value of the miR-3960 binding sites varied from 91.5% to 96.6% and corresponded to a high degree of interaction between the miRNAs with the mRNAs. miR-3960 binding sites are located predominantly in the 5'UTRs and CDSs of the studied mRNAs, and only 1.8% of the binding sites are in 3'UTRs. The miR-3960 binding sites of the mRNA of ten genes were spaced two or three nucleotides apart. miR-3960 binding sites are located in the 5'UTRs, CDSs and 3'UTRs of mRNAs. The 3'UTRs of the mRNAs of several genes have high homology to miR-3960 binding sites.

Discussion:

The average GC-content value in 5'UTRs and CDSs is known to be higher than that in 3'UTRs. Retaining the nucleotide sequence of miRNA binding sites within the high GC-content in the 3'UTR is difficult because of the involvement of the 3'UTR in the formation of the 2 and 3 dimensional structures of mRNAs. The presence of a complementary binding site of the same length is necessary for the generation of stable secondary and tertiary mRNA structures. A high GC-content in the miRNA binding site promotes its presence on the surface of the 3D-structure of the mRNA because G and C nucleotides are more hydrophilic than A and U nucleotides. Thus, the probability of miRNAs binding to regions of 5'UTRs and CDSs with high GC-content is higher than the probability of miRNAs binding to regions of 3'UTRs [11].

Another way to increase the probability of miRNA binding to mRNA is increasing the number of sites in each mRNA. For example, miR-3960 has multiple binding sites in the 5'UTRs and CDSs of many genes. It is difficult to conserve the multiple independent miRNA binding sites in these regions; therefore, it is easier to increase the probability of miRNA binding by creating multiple binding sites located side by side. This was the case for the miR-3960 binding sites in the 5'UTRs and CDSs of many studied mRNAs **Table 1** (see **supplementary material**).

We propose that the localisation of miRNA binding sites to 5'UTRs and CDSs will aid in the termination of protein synthesis. Such localisation saves the cell time and energy compared to translation inhibition in the case of localisation of

miRNA binding sites to 3'UTRs. The results of the present study showed that in most cases miRNA binding sites located in CDSs are near the start codon.

Unfortunately, there is not sufficient experimental verification of the formation of miR-3960 binding sites. However, available data confirm the high predictive power of our developed program. For example, the overexpression of miR-3960 promotes BMP2-induced osteoblastogenesis. The miR-2861~miR-3960 cluster was found to be associated with bone homeostasis induced by BMP2-Runx2 [12]. Chatterjee *et al.* showed that miR-3960 has a binding site in the mRNA of the *Hoxa2* gene, which is a repressor of *Runx2* gene expression [13]. Some target genes of miR-3960 encode transcription factors: *SKOR2*, *PRDM12*, *FOXF2*, *FOXE1*, *BTBD2*, *AFF2*, *ZSWIM6*, *ZNF219*, *ZNF703*, *TRIM67* and *RANBP9*. Inhibition of the translation of the corresponding proteins causes many pathologies, including tumorigenesis. For example, miR-3960 can suppress *BCL11A* gene expression resulting in the development of lymphoma [14]. Alterations in the gene expression of *CDH3* and *NDRG3* have been observed in colorectal and prostate cancer [15-16]. The observed changes in expression of these genes were most likely due to the action of miR-3960 on their mRNAs.

Conclusion:

In this study, we have shown that miR-3960 has 1100 binding sites on 375 target mRNAs with $\Delta G/\Delta G_m$ values of 90% or more. miR-3960 binding sites are located predominantly in the 5'UTRs and CDSs of the studied mRNAs. Another feature of this miRNA is the presence of multiple binding sites in the mRNA of its target genes. Dysregulation of miR-3960

expression can lead to cancer, because of presence great number of target genes with high affinity of the binding sites. Therefore the study is important for creation of diagnostic methods of cancer.

References:

- [1] Tétreault N & De Guire V, *Clin Biochem.* 2013 **46**: 842 [PMID: 23454500]
- [2] Wang H *et al.* *PLoS One* 2013 **8**: e70192 [PMID: 23936390]
- [3] Zhong S *et al.* *Gene* 2013 **531**: 8 [PMID: 23994196]
- [4] Shi Y *et al.* *Zhonghua Bing Li Xue Za Zhi.* 2013 **42**: 20 [PMID: 23611268]
- [5] Zhou P *et al.* *PLoS One* 2013 **8**: e68204 [PMID: 23874542]
- [6] Berillo O *et al.* *Comput Biol Med.* 2013 **43**: 1374 [PMID: 24034728]
- [7] Hausser J *et al.* *Genome Res.* 2013 **23**: 604 [PMID: 23335364]
- [8] Griffiths-Jones S *et al.* *Nucleic Acids Res.* 2006 **34**: D140 [PMID: 16381832]
- [9] Kool ET, *Annu Rev Biophys Biomol Struct.* 2001 **30**: 1 [PMID: 11340050]
- [10] Leontis N *et al.* *Nucleic Acids Res.* 2002 **30**: 3497 [PMID: 12177293]
- [11] Issabekova A *et al.* *Bioinformatics* 2012 **8**: 513 [PMID: 22829721]
- [12] Hu R *et al.* *J Biol Chem.* 2011 **286**: 12328 [PMID: 21324897]
- [13] Chatterjee S *et al.* *Trends Dev Biol.* 2012 **6**: 45 [PMID: 23950621]
- [14] Satterwhite E *et al.* *Blood* 2001 **98**: 3413 [PMID: 11719382]
- [15] Broussard EK *et al.* *Cancer Prev Res (Phila).* 2013 **6**: 666 [PMID: 23682078]
- [16] Wang W *et al.* *Int J Cancer.* 2009 **124**: 521 [PMID: 18975380]

Edited by P Kanguane

Citation: Ivashchenko *et al.* *Bioinformatics* 10(7): 423-427 (2014)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Characteristics of miR-3960 binding sites with an interval of 1-3 nucleotides in the CDSs of some mRNAs. The number of binding sites in the mRNA fragment was indicated in bracket.

Gene	Position of binding sites, nt	Binding energy, kJ/mole	$\Delta G/\Delta G_m$, %	Oligopeptide
<i>ADRB1</i>	950 - 959 (5)	-114.6 ÷ -123.1	92-98	GPPRPAAAAA
<i>BTBD2</i>	123 - 142 (14)	-114.6	92	AAAAAAAAAAAAA
<i>CASKIN1</i>	4064 - 4074 (8)	-114.6	92	AAAAAAAAAAAAA
<i>CEBPB</i>	230 - 231 (2)	-116.8 ÷ -121.0	93-97	PLPPPPP
	677 - 687 (5)	-114.6 ÷ -123.1	92-98	PPPPPPPPA
<i>CHD3</i>	217 - 224 (6)	-114.6 ÷ -121.0	92-97	FPPPPPPP
	244 - 253 (7)	-112.5 ÷ -114.6	90-92	PLPPPPPPP
<i>CTNND2</i>	795 - 802 (5)	-114.6	92	PAPPPPPP
	874 - 877 (3)	-114.6	92	PPPPPPP
<i>DLX6</i>	131 -138 (6)	-114.6	92	PPPPPPPPP
<i>DMRTA2</i>	1504 - 1511 (6)	-114.6	92	AAAAAAAAA
<i>FMNL1</i>	2002 - 2012 (8)	-114.6 ÷ -116.8	92-93	VPPPPPPPPP
<i>FOXB2</i>	485 - 501 (5)	-114.6 ÷ -116.8	92-93	PPQPPPPPP
<i>FOXE1</i>	1164 - 1177 (10)	-114.6	92	AAAAAAAAAAAAA
<i>FOXF2</i>	201 - 217 (11)	-114.6 ÷ -121.0	92-97	PPAAAAAAAAAAAA
<i>GSG1L</i>	274 - 287 (8)	-114.6 ÷ -121.0	92-97	TAAPAAAAAAAAA
<i>HTT</i>	261 - 268 (6)	-114.6	92	QQPPPPPPP
	346 - 352 (4)	-112.5 ÷ -114.6	90-92	QPPPPPPP
<i>IRF2BPL</i>	1366 - 1373 (6)	-114.6	92	AAAAAAAAA
<i>IRS2</i>	2593 - 2597 (4)	-114.6	92	AAAAAAAAA
<i>MEX3C</i>	423 - 426 (3)	-114.6 ÷ -121.0	92-97	PPPPPPP
<i>MNX1</i>	675 - 682 (6)	-114.6	92	AAAAAAAAA
<i>PRDM12</i>	1092 - 1114 (12)	-114.6 ÷ -116.8	92-93	PALAAAAAAAAAAAA
<i>RANBP9</i>	262 - 268 (2)	-114.6 ÷ -116.8	92-93	LLHPPPPPPP
	301 - 308 (3)	-114.6 ÷ -123.1	92-98	APPPPPPPP
<i>SCXA</i>	470 - 473 (3)	-114.6	92	PPPPPPP
<i>SKIDA1</i>	2925 - 2965 (12)	-114.6	92	AAAAAAAAAAAAAAAAAAAAA
<i>SKOR2</i>	851 - 854 (3)	-114.6 ÷ -121.0	92-97	PPPPPPP
	2079 - 2085 (5)	-114.6 ÷ -121.0	92-97	PAPPPPPP
	2498 - 2499 (2)	-116.8 ÷ -123.1	93-98	PPPPPPP
<i>TPRN</i>	461 - 471 (5)	-114.6 ÷ -121.0	92-97	PPPPPPPPAP
<i>TRIM67</i>	797 - 804 (5)	-114.6	92	PPPPPPPPP
<i>TSPYL2</i>	210 - 217 (5)	-114.6 ÷ -121.0	92-97	PPPPPPPPP
<i>ZIC3</i>	634 - 644 (7)	-114.6	92	AAAAAAAAAAAAA
<i>ZIC5</i>	1475 - 1491 (14)	-114.6	92	LPPPPPPPPPPP
<i>ZNF703</i>	1712 - 1719 (6)	-114.6	92	AAAAAAAAAAAAA
<i>ZSWIM6</i>	519 - 529 (8)	-114.6	92	AAAAAAAAAAAAA