Retained introns increase putative microRNA targets within 3’ UTRs of human mRNA

Sheng Tan\textsuperscript{a,b}, Jiaming Guo\textsuperscript{a}, Qianli Huang\textsuperscript{a}, Xueping Chen\textsuperscript{a,*}, Jesse Li-Ling\textsuperscript{c}, Qingwei Li\textsuperscript{b,*}, Fei Ma\textsuperscript{b,*}

\textsuperscript{a} Department of Chemistry, University of Science and Technology of China, Hefei 230029, China
\textsuperscript{b} Institute of Functional Gene and Proteomics of Marine Biology, Liaoning Normal University, Dalian 116029, China
\textsuperscript{c} Department of Medical Genetics, China Medical University, Shenyang 110001, China

Received 11 January 2007; revised 31 January 2007; accepted 6 February 2007

Available online 14 February 2007

Edited by Lev Kisselev

Abstract MicroRNAs (miRNAs) are a class of non-coding RNA that post-transcriptionally regulates the expression of target genes by binding to miRNAs. As one form of alternative splicing, intron retention has influence upon mRNA modification and protein encoding. The effect of miRNA on mRNA containing retained intron within 3’ UTR, however, has not been systematically elucidated. Here, we examined a total of 2864 human genes which contain at least one retained intron from the MAASE and ASD databases and found 387 genes having contained retained introns within 3’ UTR. The effect of retained introns upon miRNA targets was explored with three web-based programs for miRNA prediction including miRanda, TargetScanS and PicTar. The results showed that retained introns can increase putative miRNA targets in human mRNA. Retained introns have higher chances than other regions of 3’ UTR in involving the site of miRNAs targets of most genes which contain putative miRNA targets within it. Furthermore, some transcripts contain miRNA targets solely because of the retained introns in 3’ UTR. In addition, we examined those ‘Ignored’ retained introns by miRanda software and the results indicated that miRNAs may contain many more putative targets.

© 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: MicroRNA; Alternative splicing; Transcript; Intron retention; Untranslated regions (UTRs); MiRNA targets

1. Introduction

MicroRNAs (miRNAs) are a class of non-coding RNAs that can post-transcriptionally regulate gene expression in plants and animals. Recent studies have suggested that miRNAs play important regulatory roles in a broad range of biological processes including developmental timing, cellular differentiation, proliferation, apoptosis, tumorigenesis, insulin secretion and cholesterol biosynthesis [1–4]. Founding members that are important developmental regulators such as lin-4 and let-7 have been identified in Caenorhabditis elegans [5,6]. So far, public database has collected several thousand of miRNAs cloned from a variety of sources [7].

Current model for animal miRNAs function posits that part of the 21–22 nt miRNA sequence binds to the 3’ untranslated region (3’ UTR) of a target mRNA, resulting in downregulation of gene expression. Systematic target-site mutation experiments indicated that such sites can be grouped into two broad classes [8]. One class shows sufficient complementarity to the 5’ end of miRNA and requires little or no supportive base pairing. The others’ 3’ compensatory sites have insufficient 5’ pairing and require strong 3’ pairing for functioning. A typical animal miRNA may target in the order of 100 different genes and over 30% of human genes may be under miRNA regulation [8–10]. As experimental identification of targets is difficult, and only a handful of miRNAs targets have been identified through experiment, there has been an explosion of computational predictions [11–13]. Currently, more than 10 computational target predicting programs are available, which are mainly based on two classes of methods. One is based on reverse complement matches to the 6–8 nt conserved seed sequence at the 5’ end of miRNA. Examples may include TargetScanS [10] and PicTar [9]. The other involves extensive base pairing to the 3’ end of miRNA to compensate for imperfect or a short stretch of base pairing to the seed portion of the miRNA. Examples may include miRanda [14], RNAhybrid [15] and DIANA-microT [16]. TargetScanS and PicTar can produce similar sets of predictions and their sensitivities are high. miRanda, RNAhybrid and DIANA-microT not only can match this accuracy and sensitivity, but predict many additional targets [11].

Alternative splicing is a conserved post-transcriptional regulatory mechanism that can increase protein diversity and affect mRNA stability [17,18]. It also regulates many aspects of protein function, which may include binding properties, intracellular localization, enzymatic activity, stability and post-translational modifications. Recent microarray data have indicated that as much as 76% of human genes produce alternatively spliced products [19]. Basic types of alternative splicing include exon skipping, intron retention, and alternative 5’ and 3’ splice site usage. Although expression from unspliced and incompletely spliced RNA is common in viruses, and intron retention has been a common form of alternative splicing in plants [20], this form of alternative splicing seems to be relatively rare in higher organism. One study examining a total of 21,106 human genes revealed that about 15% of them retain at least one intron [21], whilst another study estimated that this occurs in only 5% of human genes [22]. In recent years, intron retention has been observed in many genes, including Id3 [23],

0014-5793/32.00 © 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2007.02.009
proinsulin [24], human Kallikrein [25] rat urocortin 1 [26] and human Tap [27], which have been associated with protein coding, development regulation and tissue specificity. Although various relations between alternative splicing and microRNA have been reported [28–30], the effect of retained introns on microRNA targets has yet to be elucidated.

Some intron retention events have been found to locate within the 3′ UTR. There has also been a consensus that animal microRNA sequences bind to the 3′ UTR of a target mRNA. To ascertain whether retained introns may increase putative microRNA targets in human genes, we have systematically examined potential involvement of microRNA in the regulation of genes through retained introns identified within 3′ UTR using three mammalian microRNAs target predicting programs including miRanda [14], TargetScanS [10] and PicTar [9]. Meanwhile, 3′ UTR EST sequences were retrieved from Ensemble and UCSC. Retained intron sequences were retrieved from the Manually Annotated Alternatively Spliced Events (MAASE) database [31] and Alternative Splicing Database (ASD) [19].

2. Results

2.1. Retained introns contain putative microRNA targets of 3′ UTR

A total of 195 genes were retrieved from the MAASE database designed for microarray applications where each entry contains at least one retention intron event [31]. Upon comparison of the 3′ UTR EST sequence from Ensemble and UCSC, 83 were found to contain at least one intron retention event within 3′ UTR. Examining of these genes with three web-based microRNA predicting programs including miRanda, TargetScanS and PicTar had found that 38 have putative microRNA targets in retained introns (Fig. 1A). Details of these 38 genes and their microRNA targets in retained introns are shown in Supplementary material Table S1. Whilst predictions by TargetScanS and PicTar were in the range of 64–75% of identity, those by miRanda only shared 29–42%, and 27–43%, respectively. Eleven genes were predicted to contain the same targets by all three programs (Table 1).

A total of 2669 genes retrieved from ASD have contained at least one retention intron event, whilst comparison of 3′ UTR EST sequences from Ensemble and UCSC had found that 304 genes have at least one intron retention event within their 3′ UTR. Further analysis of these 304 genes using miRanda, TargetScanS and PicTar programs have found that 124 have putative microRNA targets in the retained introns (Fig. 1B). More information of these genes and their microRNA targets in retained introns are detailed in Supplementary material Table S2. Similarly, concordance between the predictions by TargetScanS and PicTar were in the range of 46–58%, with those by miRanda and PicTar and miRanda and TargetScan only being 13–40%, and 12–48%, respectively. Nineteen genes were predicted to contain the same targets by all three programs (Table 2).

The locations of above predicted target sites can be grouped into four broad categories: (1) solely within retained introns; (2) mainly within retained introns; (3) equally between retained

<table>
<thead>
<tr>
<th>Gene</th>
<th>ID</th>
<th>Description</th>
<th>miRNA targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>RET</td>
<td>ENSG000000183010</td>
<td>Proto-oncogene tyrosine-protein kinase receptor ret precursor</td>
<td>mir-15a,15b,16,19,218</td>
</tr>
<tr>
<td>CANX</td>
<td>ENSG000000127022</td>
<td>Calnexin precursor</td>
<td>mir-130a,130b,148a,148b,152,153,29a,301,30a-3p</td>
</tr>
<tr>
<td>LIMK2</td>
<td>ENSG000000182541</td>
<td>LIM domain kinase 2</td>
<td>mir-194</td>
</tr>
<tr>
<td>SFRS2</td>
<td>ENSG000000161547</td>
<td>Splicing factor, arginine/serine-rich 2</td>
<td>mir-183</td>
</tr>
<tr>
<td>SFRS7</td>
<td>ENSG000000115875</td>
<td>Splicing factor, arginine/serine-rich 2</td>
<td>mir-299,320</td>
</tr>
<tr>
<td>PTG2 OR COX2</td>
<td>ENSG000000173756</td>
<td>Prostaglandin G/H synthase 2 precursor</td>
<td>mir-101</td>
</tr>
<tr>
<td>NR3C1</td>
<td>ENSG000000113580</td>
<td>Glucocorticoid receptor (GR)</td>
<td>mir-204,211,22</td>
</tr>
<tr>
<td>BTG2 OR PCS3</td>
<td>ENSG00000015938</td>
<td>BTG2 protein</td>
<td>mir-103,107,128a,128b,15a,15b,16,19,25,27a,27b,32,34,34c,92</td>
</tr>
<tr>
<td>FLJ13910</td>
<td>ENSG000000153561</td>
<td>Required for meiotic nuclear division 5 homolog A</td>
<td>mir-138,143,27a,27b</td>
</tr>
<tr>
<td>GDA</td>
<td>ENSG000000119125</td>
<td>Guanine deaminase</td>
<td>mir-130a,130b,19a,19b</td>
</tr>
<tr>
<td>CTBP2</td>
<td>ENSG000000175029</td>
<td>C-terminal-binding protein 2</td>
<td>mir-370</td>
</tr>
</tbody>
</table>
introns and other regions; and (4) mainly within other regions. Notably, the second and third categories seem dominating. However, no significant difference has been found between the proportions of the four categories among the retrieved sequences.

2.2. Exploration of the potential relationship between retained introns and miRNA targets

Some retained introns do not contain putative miRNA targets. One study has observed that the numbers of miRNA targets and expressions of mRNA correlate with the lengths of 3′ UTR [32]. In this study, further analysis of genes retrieved from the MAASE and ASD databases had detected a significant difference between the lengths of target-retaining introns and non-retaining introns. Intriguingly, the average lengths (776 bp and 547 bp) of target-retaining introns differed significantly from those of non-retaining introns (155 bp and 178 bp, respectively, Fig. 2). The uncovered difference may in part account for the number of putative targets in certain retained introns.

As many miRNAs may target the same site and many of the predicted targets have contained more than one conserved site

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>miRNA targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLOC1</td>
<td>Translocation protein SEC62</td>
<td>mir-135a,135b</td>
</tr>
<tr>
<td>SEC61A1</td>
<td>Protein transport protein Sec61 subunit alpha isoform 1</td>
<td>mir-34c</td>
</tr>
<tr>
<td>RAB10</td>
<td>Ras-related protein Rab-10</td>
<td>mir-124a</td>
</tr>
<tr>
<td>LRRFIP2</td>
<td>Leucine-rich repeat flightless-interacting protein 2</td>
<td>mir-101</td>
</tr>
<tr>
<td>AHCY</td>
<td>Adenosylhomocysteinase</td>
<td>mir-144</td>
</tr>
<tr>
<td>NDRG1</td>
<td>Protein NDRG1</td>
<td>mir-133a,133b</td>
</tr>
<tr>
<td>RAB5B</td>
<td>Ras-related protein Rab-5B</td>
<td>mir-106a,106b,17-5p, 20a,20b,22,24</td>
</tr>
<tr>
<td>RPS6KA1</td>
<td>Ribosomal protein S6 kinase alpha-1</td>
<td>mir-326</td>
</tr>
<tr>
<td>BCL2L2</td>
<td>Apoptosis regulator Bcl-W</td>
<td>mir-103,106a,106b,107,140,15a,15b,16,17-5p,195,20a,20b,24,29a,29b,29c</td>
</tr>
<tr>
<td>CD164</td>
<td>Putative mucin core protein 24 precursor</td>
<td>mir-15a,15b,16,195,96</td>
</tr>
<tr>
<td>DNAJB2</td>
<td>DnaJ homolog subfamily B member 2</td>
<td>mir-185</td>
</tr>
<tr>
<td>SFRS1</td>
<td>Splicing factor, arginine/serine-rich 1</td>
<td>mir-1,206,27a,27b</td>
</tr>
<tr>
<td>SSH2</td>
<td>Protein phosphatase Slingshot homolog 2</td>
<td>mir-24</td>
</tr>
<tr>
<td>KCNN3</td>
<td>Small conductance calcium-activated potassium channel protein 3</td>
<td>mir-185</td>
</tr>
<tr>
<td>COPS7B</td>
<td>COP9 signalosome complex subunit 7b</td>
<td>mir-125a,125b,22,30a 30b,30c,30d</td>
</tr>
<tr>
<td>CH013</td>
<td>C8orf13 protein</td>
<td>mir-29a,29b,29c</td>
</tr>
<tr>
<td>RAI16</td>
<td>retinoic acid induced 16</td>
<td>let-7i</td>
</tr>
<tr>
<td>C0orf74</td>
<td>OTTHUMP0000002259</td>
<td>mir-29a</td>
</tr>
</tbody>
</table>

Fig. 2. y-Axis shows the length and x-axis shows the number of introns. It is obvious that the majority of targeted introns, depicted with blue and yellow lines, are longer than 200 bp (74%, 79%). On the other hand, only a minority of non-targeted introns, depicted in pink and green, are longer than 200 bp (19%, 26%).
for a single miRNA, mean densities of miRNAs as well as the sites in retained introns and other regions in 3' UTR were compared (Table 3). Analyzing the miRNA density (number of miRNA/length of the sequence) and the site density (number of site/length of the sequence) in two distinct regions has failed to detect any marked difference between the miRNA target density in the retained introns and other regions of 3' UTR, but detect marked difference in terms of site density. This seems to suggest that the presence of miRNA targets in retained introns is no rare events and the retained introns have higher chances than other regions of 3' UTR in involving the site of miRNAs targets of most genes of the type.

In some cases, the same miRNA targets have been found within the retained intron and other regions in 3' UTR. To investigate whether these repeated regions within the same 3' UTR may result in duplicated miRNA targets, such sequences were analyzed with the Tandem Repeats Finder [33] and MEME (version 3.5.4) [34] software. The results were all negative. However, certain miRNAs were found capable of targeting retained introns and other regions with slightly different sequences. For instance, in PicTar database, the seven seed sequence “GACACUU” of hsa-mir-27a,b could target CTGTGAA in retained introns and CTGTGGA in other regions in 3' UTR of gene NR3cl. This seems to suggest that these retained introns may increase the chance for their host mRNA to be targeted by certain miRNAs.

2.3. miRNAs may contain more putative targets for miRNA modulation

Many 3' UTR retrieved from the three selected databases are of short sequences without retained introns. Hence, putative targets in retained introns might have been missed. Here, we took all 462 human miRNAs listed in the Sanger miRNA Registry (Version 8.2, July 2006) and examined the miRNA sequence’s complementarity against these retained introns regions with miRanda software [35]. It is established that many miRNAs have short, perfect ‘seeds’ of at least 6–8 bases near the 5' end of the miRNA that are complementary to sequences within 3' UTRs. Recent studies, however, have suggested that miRNA seeds are often at least eight bases in length [36]. Therefore, in this study a similar criterion, i.e., predictions that should have at least eight base pairing sequences near the 5' end of the miRNAs, was adopted. Five 3' UTR sequences from MAASE were unanimously found to be of the short type and do not contain retained introns. Their average length is 229 bp, among which the longest is 293 bp and the shortest is 144 bp. Based on above criteria, in average there are 12 targets in each of the retained introns, ranging between 20 for the most and five for the fewest (Supplementary material Table S3). Considering that as many as 210 genes from the ASD database have contained retained introns in 3' UTR but are not recorded in PicTar (Fig. 1B), data from TargetScanS and miRanda were chosen. Seventy 3' UTR sequences from ASD were found to be of the short type in TargetScanS, which do not have retained introns. Meanwhile, such sequences are either of the short type or not recorded in the miRanda database. The average length of them is 590 bp, among which the longest is 3037 bp and the shortest is 88 bp. For the criteria of at least eight base pairing in the 5' of miRNA with miRanda, in average there are 36 targets in each of the retained introns, ranging between 138 for the most and 1 for the fewest (Supplementary material Table S4). Taken together, above results strongly suggest that miRNAs probably have many potential targets that have been missed by current predictions.

In summary, our results strongly suggest that miRNAs probably contain many more putative targets for miRNA modulation. Some genes may regulate their own expression by retained introns in 3' UTR that may serve as targets of miRNAs. Retained introns within 3' UTR of human mRNA, as predicted, may increase the chance for their host miRNAs to be targeted by miRNAs. They also have higher chances than other regions of 3' UTR in involving the site of miRNAs targets.

3. Discussion

Our results have suggested a potential connection between intron retention and miRNA targeting, both of which are important for the control of transcription. A recent study had observed the interplay of polyadenylation and miRNA targeting through a systematic examination of differential repression of alternative transcripts [29]. Other studies also indicated that 3' UTR lengths of highly expressed genes are tissue-specific [37] and so are the expression of many miRNAs [38]. Furthermore, alternative splicing in conjunction with transcriptional regulation determines tissue-specific expression of transcript isoforms of a gene [39]. These have all indicated a connection between human miRNA targets and alternative parts of 3' UTR. Moreover, miRNAs can directly deadenylate of mRNA [40] and inhibit eukaryotic initiation factor 4E/cap and poly (A) tail function [41]. These studies also suggest that miRNA targets are involved in 3' UTR isoform regulation. Thus, it is important to construct more detailed 3' UTR annotations. As more accurate annotation of 3' UTR sequence of coding genes becomes available, alternative polyadenylation can be understood better, and roles of proteins binding to
the 3′ UTR are deciphered. miRNA target predicting programs may also see improvement.

The positions of retained introns may be grouped into two broad classes, i.e., either completely embodied within the 3′ UTR or with its tail constituting the beginning of 3′ UTR. In some cases, regions in 3′ UTR other than retained introns contained no miRNA targets. In other words, the retained in-miRNA database (http://microrna.sanger.ac.uk, Version 8.2, July 4.3. miRNA sequences noting many of the 3′0

|pictar.bio.nyu.edu/cgi-bin/new_PicTar_vertebrate.cgi). In addition, (http://cbio.mskcc.org/mirnaviewer/), TargetScanS (http://pictar.bio.nyu.edu/cgi-bin/new_PicTar_vertebrate.cgi) gets by three web-based miRNA predicting programs, i.e., miRanda software. Notably, compared with the total 462 miRNAs collected by the Sanger miRNA Registry, the three selected predicting programs lacked many new miRNA genes. A recent study has identified 447 new human miRNA genes [42]. Many of such genes are not conserved beyond primates, and some miRNAs seem species-specific, indicating their recent origin. Taken together, these studies seem to suggest that miRNAs may have more targets than previously predicted and current programs which enforced the conservation of potential binding sites at orthologous positions across multiple species may miss many species-specific miRNAs targets.

The last but not the least, the bare overlap between the predictions for MAASE and ASD data seems to suggest that there probably exist more intron retention events among human genes. Therefore, more comprehensive and detailed databases on intron retention events are to be constructed.

4. Materials and methods
4.1. EST, mRNA, cDNA and 3′ UTR sequences
EST, mRNA, cDNA, and 3′ UTR EST sequences were obtained from Ensemble (http://www.ensembl.org/) and UCSC databases (http://genome.ucsc.edu/cgi-bin/hgGateway, The March 2006 human reference sequence). Basal data were generated through computational comparison of EST/mRNA alignments with genomic sequences.

4.2. Intron retention events
Intron retention events in humans were obtained from MAASE (http://maase.genomics.purdue.edu/) and ASD Human Release 2 (April 2005) (a systematic collection and annotation of alternative splicing at http://www.ebi.ac.uk/asd/index.html). Retained introns within 3′ UTR were identified through comparing retained intron sequences with 3′ UTR sequences.

4.3. miRNA sequences
A total of 462 miRNA sequences were obtained from the Sanger miRNA database (http://microrna.sanger.ac.uk. Version 5.2, July 2006).

4.4. Potential miRNA targets within retained introns
We explored the effect of intron retention upon human miRNA targets by three web-based miRNA predicting programs, i.e., miRanda (http://cbio.mskcc.org/mirnaviewer/), TargetScanS (http://pictar.bio.nyu.edu/TargetScanS (Release 3.0 June 2006)) and PicTar (http://pictar.bio.nyu.edu/cgi-bin/new_PicTar_vertebrate.cgi). In addition, noting many of the 3′ UTR sequences of the three China databases are short sequences without retained introns, we examined these retained introns by miRanda software [35] to find potential miRNA targets.

Acknowledgements: We thank our colleagues for their suggestions and criticisms on the manuscript. This work was jointly supported by grants from the National Natural Science Foundation of China (Nos. 60305001 and 60575005), the Natural Science Foundation of Fujian Province (No. D0410003), Key Project of Science and Technol-

ogy of the Educational Ministry of China (No. 206032) as well as the Science and Technology Project of Dalian (No. 2006E11SF068).

Appendix A. Supplementary data
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2007.02.009.

References


