



Alu elements within human mRNAs are probable microRNA targets

Neil R. Smalheiser and Vette I. Torvik

University of Illinois-Chicago, UIC Psychiatric Institute MC912, 1601 W. Taylor Street, Chicago, IL 60612, USA

Recently, we reported that four microRNAs show perfect complementarity with MIR/LINE-2 elements within human mRNAs. This finding raises the question of whether microRNAs might also target other genomic repeats and transposable elements. Here, we demonstrate that almost 30 human microRNAs exhibit typical short-seed complementarity with a specific site within Alu elements that is highly conserved within 3' untranslated regions of human mRNAs. The results suggest that at least some Alu elements within human mRNAs serve as microRNA targets.

Introduction

The rules governing microRNA–target interactions are under study by many groups (for reviews see Refs [1–4]). It is established that many microRNAs have short, perfect ‘seeds’ of at least 6–8 bases (with no mismatches or G:U matches) near the 5' end of the microRNA that are complementary to sequences within 3' untranslated regions (UTRs) [1–3]. Although not all human microRNA–target interactions follow this consensus pattern [4,5], recent studies suggest that microRNA seeds are often at least 8 bases in length [6] and that bases 2–8 of the 5' seeds are optimally placed to interact directly with targets [7,8].

Recently, we reported that four mammalian microRNAs show perfect complementarity with the MIR/LINE-2 class of repeat elements, which are present within a large number of human mRNAs and EST transcripts [9]. This finding raises the question of whether microRNAs might also target other genomic repeats and transposable elements in 3' UTRs. Given that Alu is the most prominent repeat, expressed in >5% of all human 3' UTRs [10], we asked whether a significant number of other human microRNAs show 5' seed complementarity against Alu sequences.

We took all 313 human microRNAs listed in the Sanger miRNA Registry (<http://microrna.sanger.ac.uk>, Version 7.0, June 2005), obtained the 235 unique seed sequences beginning at position 2 and having length 8, and examined their complementarity against 3' UTR regions of all human mRNAs listed in RefSeq (<http://www.ncbi.nlm.nih.gov/RefSeq/>). Regions of exact complementarity (no G:U matches), called ‘hits,’ were scored according to whether they were within annotated Alu repeats or outside known repeats. Over three quarters of the microRNA seeds had

≤10 hits in Alu elements in the entire set of 3' UTRs, and almost all showed a greater number of hits in 3' UTR sequences outside known repeats. However, two seeds were significant outliers (>3 standard deviations above the mean of the distribution) and predominantly hit Alu sequences (Figure 1). These seeds (CAAAGUGC and AAGUGCUG) were highly overlapping and did not contain low-complexity sequence or unusual nucleotide composition.

As a different way of assessing whether hits in Alu sequences are likely to be due to some general property, such as distinctive nucleotide composition, all 235 microRNA seeds were tested for complementarity to the set of Alu sequences in 3' UTRs subjected to scrambling 100 times (maintaining dinucleotide composition; see [Supplementary Material, file 1](#)). Most seeds showed no significant excess of hits in Alu compared with scrambled Alu. Nine seeds showed *z*-scores of 5–50; it is possible that these represent biologically relevant hits in Alu sequences, but they have not been studied further in the present study. Three seeds showed extremely high *z*-scores of 160–180: these represented the two outlier seeds identified in Figure 1 together with another highly overlapping seed sequence (CAAAGUGC, AAAGUGCU and AAGUGCUG). Thus, these microRNA seeds stand apart from all others, according to two independent lines of evidence.

These seeds shared a common 6-mer core sequence (AAGUGC) that was also shared in the 5' seeds of a set of 27 different human microRNAs (Figure 2). Furthermore, additional sequences in the human microRNA set, extending on either side of the core sequence, also showed complementarity to the Alu consensus sequence (Table 1), and the 9-mer seeds in this set were even more significant outliers when plotted as in Figure 1 (not shown). The 5' seed of another human microRNA, miR-150, did not share the 6-mer core but nevertheless overlapped with the microRNAs in this set and mapped to an immediately adjacent site along the Alu sequence (Table 1). In all, 28 microRNAs hit a 14 nucleotide region of Alu in sense orientation (Table 1) that is 100% conserved across the consensus sequences for all 32 human Alu subfamilies described in Repbase (<http://www.girinst.org/>). The 6-mer core sequence is preserved in the mouse Alu-like B1 SINE consensus sequence (Table 1). By contrast, human 7SL RNA and BC200, although globally similar to Alu sequences, diverge significantly from Alu in the seed region (not shown).

The microRNAs identified in Table 1 hit a discrete stem-loop region of the Alu RNA at position 32–45 (Figure 3).

Corresponding author: Smalheiser, N.R. (neils@uic.edu).
Available online xxxxxx.

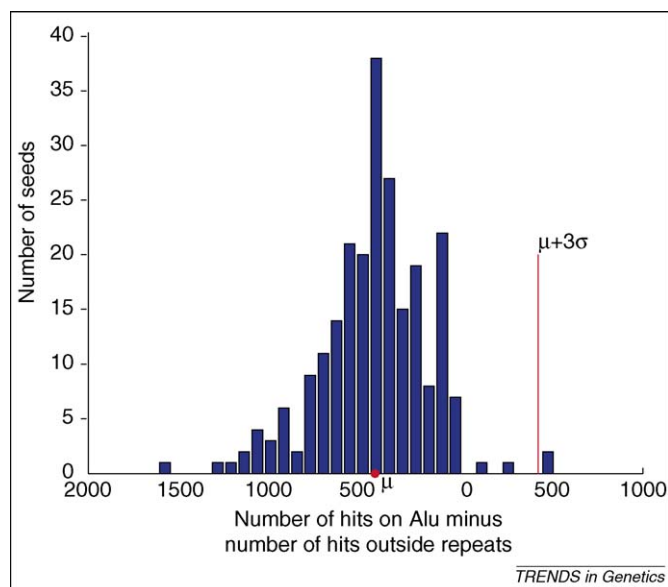


Figure 1. Unique 5' seeds of human microRNAs hitting regions within 3' UTRs. Each microRNA seed sequence (length 8, beginning at position 2) was examined for perfect complementarity within the set of all human 3' UTRs in RefSeq. For each seed we scored the difference between the total number of hits in Alu versus the total number of hits outside known repeats. Two seeds were outliers at 3.2 and 3.3 standard deviations above the mean of the overall distribution. This corresponds to a p -value=0.0007, which after correcting for multiple tests ($n=235$ seeds) gives $p=0.012$ (the chance of generating two or more points greater than 3.2 standard deviations).

This region is predicted by the program mfold [11] to fold into a structure containing a small 4-base loop (Figure 3); however, RNA loop sizes predicted *in silico* in stem-loop situations are often substantially smaller than loop sizes ascertained experimentally [12,13], and experimental studies of Alu RNA indicate that this region actually comprises a 6-base unpaired loop [14]. Strikingly, the six-base loop corresponds exactly to the site that is complementary

to the 6-mer core microRNA sequence (Figure 3). Thus, insofar as open loops have been proposed to be a feature of effective microRNA target regions [15], this site does seem to be open enough to serve as a plausible biological microRNA target.

Does having a complementary 5' seed suffice for a microRNA to target a mRNA functionally? Growing experimental evidence does support this hypothesis (e.g. [16–20]). However, several groups have proposed additional criteria for predicting which target sequences are likely to be biologically significant. For example, the miRanda software of John *et al.* [21] uses a scanning algorithm based on sequence complementarity between the mature miRNA and the target site, the binding energy of the miRNA–target duplex, and evolutionary conservation of the target site sequence and target position in aligned UTRs of homologous genes [21]. Does miRanda predict Alu sequences as promising target sites?

We chose two well-characterized microRNAs from Figure 2 (hsa-mir-17-5p and hsa-mir-93) and used the online miRanda site (<http://www.microrna.org>) to examine their top 20 predicted targets in each of two categories ('all targets' and 'common targets'; 'all targets' displays targets according to the overall ranking score for each target hit region and 'common targets' ranks targets according to the number of hits from the same microRNA onto the same putative target mRNA). Pooling the lists and removing redundancy, miRanda predicted 63 targets as the most promising and, of these, miRanda had links to the UCSC Genome Browser for 52 target mRNAs. Simple inspection revealed that in five of these cases, the microRNAs were shown as hitting within sequences that were annotated as comprising Alu sequences in sense orientation. These hits were located at the same site as identified above (Figure 3). We have found similar examples using miRanda with other microRNAs as well (not shown).

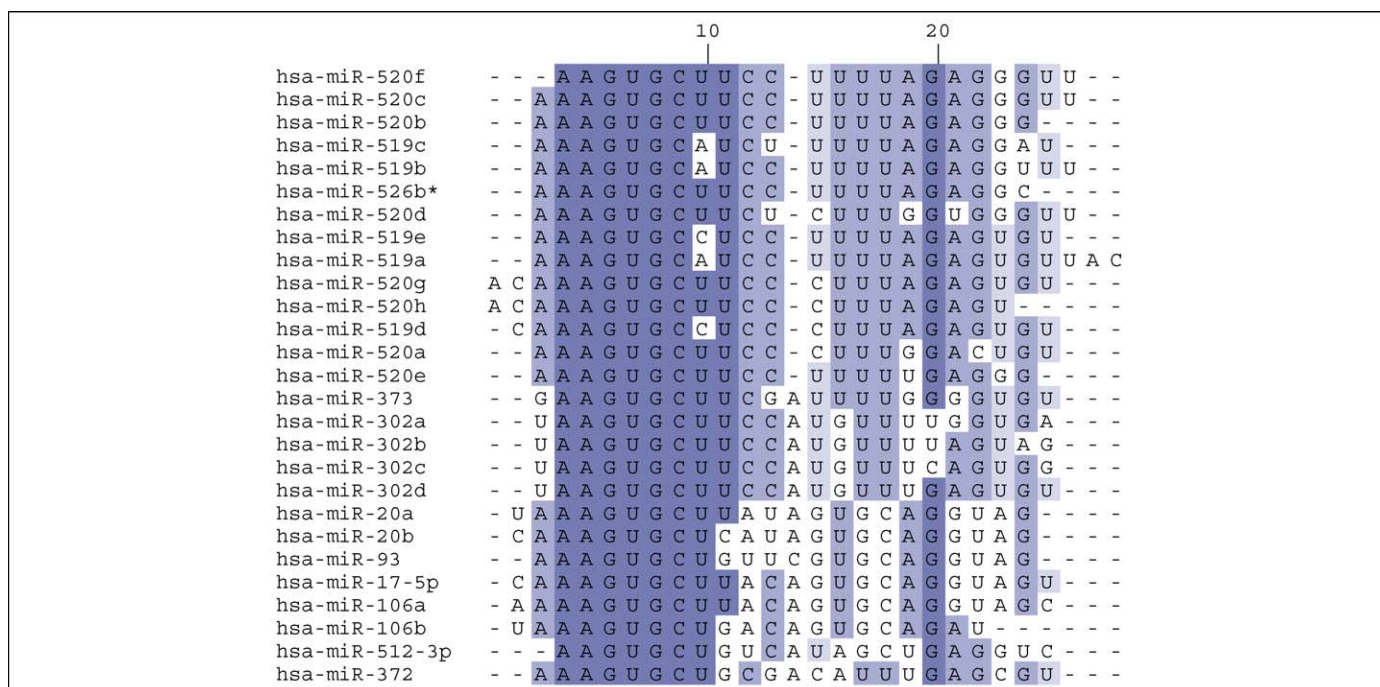


Figure 2. Multiple sequence alignment of all human microRNAs that share the 6-mer core sequence AAGUGC in their 5' seeds.

Table 1. Alu sequences complementary to the microRNAs shown in Figure 2a

Sequence	Length	MicroRNA
CAGCACUUU	9	miR-93, miR-372, miR-106b
AGCACUUUG	9	miR-17–5p, miR-20b, miR-520 g,miR-520h
CAGCACUU	8	miR-512–3p
AGCACUUU	8	miR-106a, miR-20a, miR-520a,b,c,d,e, miR-526b*
GCACUUUG	8	miR-519d
AGCACUU	7	miR-302a,b,c,d, miR-373, miR-520f
GCACUUU	7	miR-519a,b,c,e
UUGGGAG	7	miR-150
CAGCACUUUGGGAG	14-mer region on Alu consensus sequence containing hits	
CAGCACUU–GGGAG	Corresponding region of mouse B1-SINE consensus	

*Bold indicates the central GCACUU sequence found in all the elements shown.

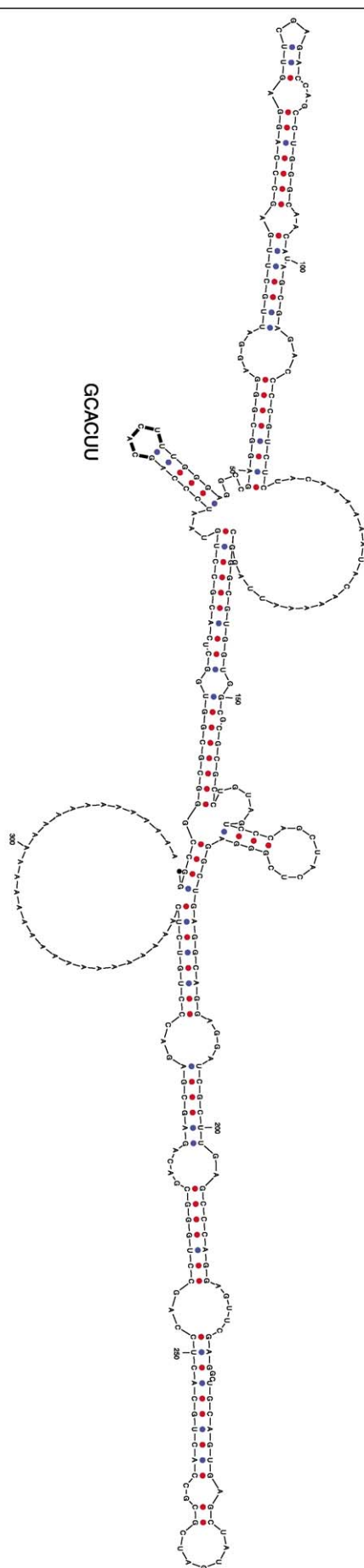
Three groups have recently reported that, in situations in which microRNA regulation might be deleterious, 3' UTRs tend to be shorter and are specifically depleted of potential microRNA target sequences [17–19]. Are Alu sequences within 3' UTRs depleted of potential microRNA target sites? Although it has been reported that Alu elements are under-represented in the sense orientation within protein-coding genes overall (including 5' UTRs, introns, coding sequences and 3' UTRs) [22], we found that sense and antisense Alu elements are equally represented within the overall set of human 3' UTRs ≥ 1000 bases in length (1182 versus 1159), and sense Alu elements are actually significantly over-represented within 3' UTRs < 1000 bases in length (1236 versus 811, $p < 10^{-16}$ based on a binomial distribution with 2047 elements in the total set). Furthermore, 83% of all annotated Alu elements in sense orientation included the target region containing the 8-mer outlier seed target sites, of which 63% expressed exact complementarity to one or both seeds. This was not significantly different from the proportion of antisense Alu elements that included the same target region (81%), of which 62% expressed the sequences corresponding to one or more of these outlier seeds.

Finally, we examined a set of 594 human microRNA targets predicted for the microRNA group containing miR-17, miR-20 and miR-106 by the web-based microRNA prediction database TargetScan [23]. TargetScan, like PicTar [24] and most other servers, removes Alu and other repeats from consideration explicitly so that they do not count towards the prediction score. Of 36 Alu elements in sense orientation within the predicted set of target mRNAs, 83% (30) contained the putative target site, of which 80% (24) retained exact complementarity for the 8-mer outlier seeds. The total number of Alu elements, and the ratio of sense to antisense elements, were not significantly different from the population of all 3' UTRs. Thus, we do not find any indication that putative target sites have been specifically depleted from Alu elements within 3' UTRs that were independently predicted to be targets of the same microRNAs described in this article.

Another possibly important factor in effective microRNA regulation of a putative target is the presence of multiple target regions within the same 3' UTR [21,23–25]. Table 2 shows the number of 3' UTRs that have one or more 8-mer outlier seed target sites. Of 3' UTRs ≥ 1000 bases in length,

Table 2. The number of 3' UTRs that express one or both 8-mer outlier seed target sites

	Number of hits per 3' UTR	Number of 3' UTRs	Length Mean \pm SD	Length excluding repeats Mean \pm SD
3' UTRs hit outside repeats	0	23 394	967 (1015)	840 (909)
	1	1090	1,994 (1494)	1,824 (1392)
	2	80	2,890 (1558)	2,744 (1420)
	3+	2	4,400 (274)	4,070 (447)
3' UTRs hit within Alu	0	23 427	977 (1028)	871 (945)
	1	1017	1,786 (1429)	1,238 (1261)
	2	109	2,684 (1402)	1,620 (1135)
	3+	13	3,645 (1972)	1,910 (1486)



the seed target sites outside known repeats seems to be independent of each other: 9.8% of 3' UTRs have at least one target sequence and 8.9% of those have at least two target sequences. However, 3' UTRs show an excess of multiple target sites within annotated Alu repeats: 9.0% of 3' UTRs have at least one target sequence within a sense Alu repeat and 14.6% of those have at least two target sequences (9% versus 14.6%, $p < 2 \times 10^{-7}$, based on a binomial distribution). This probably reflects the fact that Alu repeats tend to occur near other Alu repeats in the same orientation [26]. A total of 267 human mRNAs in RefSeq contained two or more target sites corresponding to the 8-mer seeds (122 hitting multiple times on Alu elements, 79 hitting outside repeats only and 66 mixed). These fit the diverse profile of other predicted targets [1–3,25] and include many transcription factors and other nucleic acid binding proteins, in addition to many signaling proteins (see [Supplementary Material, file 2](#)).

Concluding remarks

The bioinformatics analyses presented here provide strong support for the hypothesis that Alu elements within 3' UTRs are targeted specifically by certain microRNAs. Almost 30 human microRNAs show 5' seed complementarity against a specific site in the Alu sequence that is highly conserved. Most algorithms used to predict microRNA targets have explicitly excluded repeat sequences from consideration because of the statistical problems they introduce. Nonetheless, there is no biological reason why Alu or other transposable elements within mRNAs should be regarded as unavailable for regulation by microRNAs, and indeed mRNAs that contain multiple Alu elements within their 3' UTRs should be regarded as particularly promising targets.

Although Alu was originally thought to represent 'junk' having no biological functions, the presence of Alu sequences within protein-coding genes can affect the processing of mRNAs at multiple levels [27]. Alu insertions contribute sequences that encode amino acids [28] and 3' UTR polyadenylation sites [22] in addition to 3' UTR AU-rich elements that regulate mRNA stability [29]. Several reports indicate that mRNAs containing Alu in their 3' UTRs are – as a class – associated with growth and differentiation and are subject to translational regulation as a class during differentiation [30–34]. It is striking that several of the microRNAs identified here as targeting Alu elements are well-characterized and include miR-17–5p, miR-20a, miR-20b, miR-93, miR-106a and miR-106b, which have been implicated in cancer [35–37]. Alu-related translational control has not yet been studied in terms of sense versus antisense orientation, single versus multiple copies per transcript or presence of conserved microRNA target hits. It is possible that microRNAs target Alu sequences most effectively as a mechanism for clearing aberrant mRNAs, for example transcripts containing retained introns or with readthrough transcription into intergenic regions, that are particularly rich in multiple

Figure 3. The predicted RNA secondary structure of the Alu consensus sequence. The prediction was made using the Mfold server [18] (<http://www.bioinfo.rpi.edu/applications/mfold/old/rna/>). The bases in bold (GCACUU) indicate the site on Alu that is hit by the 6-mer core sequence (AAGUGC).

Alu elements. Nevertheless, the studies reviewed here provide biological context supporting the idea that microRNAs could be targeting Alu elements physiologically within 3' UTRs. Further research is warranted to test these predictions experimentally and to learn whether other repeats embedded in 3' UTRs are also implicated as microRNA targets.

Alu-containing transcripts are not restricted to protein-coding mRNAs. Noncoding Alu transcripts are transcribed by pol III, contain a poly-A region, show a high rate of turnover and are induced during cellular stress [38]. Alu transcription is an essential phase in the retrotransposition of Alu elements in the genome [27]. Although the raison d'être of microRNAs was originally thought to be translational repression of mRNAs, a more fundamental role could be to bind and route RNAs to processing bodies (P-bodies) to be sequestered or degraded [39]. It is currently unknown whether noncoding RNAs can also be routed to P-bodies, but if so, microRNAs that interact with noncoding Alu RNA transcripts might potentially counter retrotransposition in mammalian cells.

Acknowledgements

The authors are supported by NIH Grants DA15450 and LM07292. This Human Brain Project/Neuroinformatics research is funded jointly by the National Library of Medicine and the National Institute of Mental Health.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.tig.2006.08.007](https://doi.org/10.1016/j.tig.2006.08.007).

References

- Lai, E.C. (2004) Predicting and validating microRNA targets. *Genome Biol.* 5, 115
- Bentwich, I. (2005) Prediction and validation of microRNAs and their targets. *FEBS Lett.* 579, 5904–5910
- Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297
- Smalheiser, N.R. and Torvik, V.I. (2006) Complications in mammalian microRNA target prediction. In *Methods in Molecular Biology: MicroRNA Protocols* (Vol. 342) (Ying, S.-Y., ed.), pp. 115–127, Humana
- Nakamoto, M. et al. (2005) Physiological identification of human transcripts translationally regulated by a specific microRNA. *Hum. Mol. Genet.* 14, 3813–3821
- Xie, X. et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434, 338–345
- Yuan, Y.R. et al. (2005) Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* 19, 405–419
- Parker, J.S. et al. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* 434, 663–666
- Smalheiser, N.R. and Torvik, V.I. (2005) Mammalian microRNAs derived from genomic repeats. *Trends Genet.* 21, 322–326
- Yulug, I.G. et al. (1995) The frequency and position of Alu repeats in cDNAs, as determined by database searching. *Genomics* 27, 544–548
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415
- Zeng, Y. et al. (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* 24, 138–148
- Krol, J. et al. (2004) Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J. Biol. Chem.* 279, 42230–42239
- Sinnett, D. et al. (1991) Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J. Biol. Chem.* 266, 8675–8678
- Robins, H. et al. (2005) Incorporating structure to predict microRNA targets. *Proc. Natl. Acad. Sci. U. S. A.* 102, 4006–4009
- Lim, L.P. et al. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773
- Farh, K.K. et al. (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* 310, 1817–1821
- Stark, A. et al. (2005) Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* 123, 1133–1146
- Sood, P. et al. (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2746–2751
- Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.* 18, 504–511
- John, B. et al. (2004) Human microRNA targets. *PLoS Biol.* 2, e363
- Roy-Engel, A.M. et al. (2005) Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.* 110, 365–371
- Lewis, B.P. et al. (2003) Prediction of mammalian microRNA targets. *Cell* 115, 787–798
- Krek, A. et al. (2005) Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500
- Smalheiser, N.R. and Torvik, V.I. (2004) A population-based statistical approach identifies parameters characteristic of human microRNA-mRNA interactions. *BMC Bioinformatics* 5, 139
- Stenger, J.E. et al. (2001) Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genomic stability. *Genome Res.* 11, 12–27
- Kramerov, D.A. and Vassetzky, N.S. (2005) Short retroposons in eukaryotic genomes. *Int. Rev. Cytol.* 247, 165–221
- Britten, R. (2006) Transposable elements have contributed to thousands of human proteins. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1798–1803
- An, H.J. et al. (2004) The association of Alu repeats with the generation of potential AU-rich elements (ARE) at 3' untranslated regions. *BMC Genomics* 5, 97
- Krichevsky, A.M. et al. (1999) Translational control of specific genes during differentiation of HL-60 cells. *J. Biol. Chem.* 274, 14295–14305
- Vidal, F. et al. (1993) Coordinated posttranscriptional control of gene expression by modular elements including Alu-like repetitive sequences. *Proc. Natl. Acad. Sci. U. S. A.* 90, 208–212
- Spence, J. et al. (2006) Messenger RNAs under differential translational control in Ki-ras-transformed cells. *Mol. Cancer Res.* 4, 47–60
- Stuart, J.J. et al. (2000) The 3' UTR of human MnSOD mRNA hybridizes to a small cytoplasmic RNA and inhibits gene expression. *Biochem. Biophys. Res. Commun.* 274, 641–648
- Vila, M.R. et al. (2003) Higher processing rates of Alu-containing sequences in kidney tumors and cell lines with overexpressed Alu-mRNAs. *Oncol. Rep.* 10, 1903–1909
- O'Donnell, K.A. et al. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435, 839–843
- He, L. et al. (2005) A microRNA polycistron as a potential human oncogene. *Nature* 435, 828–833
- Hayashita, Y. et al. (2005) A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res.* 65, 9628–9632
- Liu, W.M. et al. (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.* 23, 1758–1765
- Liu, J. et al. (2005) A role for the P-body component GW182 in microRNA function. *Nat. Cell Biol.* 7, 1261–1266