BIOINFORMATICS ORIGINAL PAPER

Vol. 29 no. 7 2013, pages 894–902
doi:10.1093/bioinformatics/btt044

Systems biology

Advance Access publication January 29, 2013

The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery

Yonit Hoffman[1,2,†], Dvir Dahary[1,†], Debora Rosa Bublik[2], Moshe Oren[2,*] and Yitzhak Pilpel[1,*]

[1]Department of Molecular Genetics and [2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: The massive spread of repetitive elements in the human genome presents a substantial challenge to the organism, as such elements may accidentally contain seemingly functional motifs. A striking example is offered by the roughly one million copies of Alu repeats in the genome, of which ~0.5% reside within genes' untranslated regions (UTRs), presenting ~30 000 novel potential targets for highly conserved microRNAs (miRNAs). Here, we examine the functionality of miRNA targets within Alu elements in 3′UTRs in the human genome.

Results: Using a comprehensive dataset of miRNA overexpression assays, we show that mRNAs with miRNA targets within Alus are significantly less responsive to the miRNA effects compared with mRNAs that have the same targets outside Alus. Using Ago2-binding mRNA profiling, we confirm that the miRNA machinery avoids miRNA targets within Alus, as opposed to the highly efficient binding of targets outside Alus. We propose three features that prevent potential miRNA sites within Alus from being recognized by the miRNA machinery: (i) Alu repeats that contain miRNA targets and genuine functional miRNA targets appear to reside in distinct mutually exclusive territories within 3′UTRs; (ii) Alus have tight secondary structure that may limit access to the miRNA machinery; and (iii) A-to-I editing of Alu-derived mRNA sequences may divert miRNA targets. The combination of these features is proposed to allow toleration of Alu insertions into mRNAs. Nonetheless, a subset of miRNA targets within Alus appears not to possess any of the aforementioned features, and thus may represent cases where Alu insertion in the genome has introduced novel functional miRNA targets.

Contact: moshe.oren@weizmann.ac.il or Pilpel@weizmann.ac.il

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on October 6, 2012; revised on December 25, 2012; accepted on January 24, 2013

1 INTRODUCTION

Repetitive elements are very widely spread in primate genomes (Lander *et al.*, 2001). Most prominent is the case of the Short Interspersed Elements, and in particular the Alu elements, which are present in more than a million copies in the human genome. Such massive presence of foreign genomic elements, which are perceived as predominantly selfish DNA, may represent a substantial potential informational load on the genome. Accordingly, the retrotransposition of Alus may contribute to human disease, including a diversity of cancers (Batzer and Deininger, 2002). At the same time, the spread of genetic material may also represent an opportunity to introduce evolutionary novelty into the genome. Indeed, Alu elements may become exons (Lev-Maor *et al.*, 2003) and may harbour functional transcription factor (Polak and Domany, 2006) and microRNAs (miRNA) binding sites (Smalheiser and Torvik, 2006). In particular, Smalheiser and Torvik identified many mRNAs that contain Alus in their 3′ untranslated regions (UTRs), within which there are targets for dozens of miRNAs (Smalheiser and Torvik, 2006). Along with these negative and positive potential contributions to cellular and organismal fitness, it is conceivable that the spread of many of the retroelements was restricted evolutionarily so that most of the current elements are largely benign. Possibly, the insertion of retroelements into mRNAs was not random but was affected by features that minimize their impact on functional elements in the genome.

Focusing here on potential miRNA binding sites within Alus in 3′UTRs, we provide evidence that the majority of miRNA targets within Alus are non-functional and presumably ignored by the miRNA machinery. We propose three features that allowed the insertion of Alu-hosted miRNA targets into mRNAs with minimal distortion of miRNA regulation. Still, a minority of the insertions appears not to possess any of these features and may thus represent cases in which Alu insertions contributed novel functional miRNA targets to the primates lineage.

2 METHODS

Human and mouse genomes and repeat sequences: The full human 3′UTR sequence dataset was taken from UCSC (Fujita *et al.*, 2011) NCBI36/hg18. Alu sequences and their locations were taken from Repeat Masker (http://www.repeatmasker.org) (Smit *et al.*, 1996–2010). The mouse 3'UTR sequences and coordinates were taken from UCSC (Fujita *et al.*, 2011) (NCBI37/mm9). The mouse repeats were taken from UCSC Repeat Masker (Smit *et al.*, 1996–2010).

Prediction of miRNA targets and conservation analysis: miRNA target sites were predicted by scanning for the seed of the miRNA, on the basis of perfect (Watson–Crick) complementarity. Targets were defined as perfect 7-mers, for all human and mouse miRNAs listed in miRBase release 15 (http://www.mirbase.org/) (Griffiths-Jones, 2004; Griffiths-Jones *et al.*, 2006; Griffiths-Jones *et al.*, 2008; Kozomara and Griffiths-Jones, 2011). Conserved targets were taken from TargetScan release 5.1 (http://www.targetscan.org/) (Grimson *et al.*, 2007) m8 target type only (perfect

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

7-mer). The analysis of conservation and folding energy included only genes with the following attributes: (i) the 3′UTR in the UCSC version was the same as the 3′UTR used by TargetScan (as defined in their website) and (ii) the 3'UTR from UCSC was included within the 3'UTR defined in TargetScan or the opposite.

Analysis of the miRNA over-expression data: The data of miRNA over-expression experiments were taken from Khan *et al.* (2009), which contain siRNAs as well as miRNA overexpression experiments. A subset of 43 miRNA overexpression experiments was analysed. For each overexpressed miRNA, its site was scanned against all human 3′UTRs. Downregulation of target mRNAs was defined as the percentage of genes with fold reduction of at least 1.62 (i.e. 0.7 on a log2 scale). The cut-off was decided according to the distributions of average fold change for genes with and without the miRNA target (Supplementary Fig. S7). For each analysis, only experiments where the group of genes for consideration consisted of at least eight genes were included.

Secondary structure prediction: Secondary structures were predicted for all human and mouse 3′UTRs using the Bioinformatics Toolbox of Matlab 10, which implements the M-Fold and Vienna algorithms (Mathews *et al.*, 1999; Wuchty *et al.*, 1999). The analysis was done in windows of 100 bp, and up to 50 bps from the last coding exon were added to the beginning of the 3′UTR for the prediction. The secondary structure status of each nucleotide of the 3′UTR was determined according to its structure in the folding where this nucleotide was in position 51 (in the window of 100). The folding energy of each nucleotide was the average folding energy of this nucleotide in all the folding windows in which it was included.

Analysis of PAR-CLIP data: The raw data of the PAR-CLIP experiment were taken from Kishore *et al.* (2011). The representation of the most abundant miRNAs in the PAR-CLIP data is highly correlated between replicates and between RNAse protocols. The experiment with the highest number of mapped reads was further analysed (GSM714644). The reads were mapped to the genome using Bowtie (Langmead *et al.*, 2009) (with the parameter 'best' so that for each read, we received only one mapping and the hg18 genome index). The Bowtie output was filtered to include only reads with five mutations or less. The sequence of the reads was corrected to the genome. The miRNA targets were identified in the reads according to the reads' genome-corrected sequence. The decision of which read is within Alu was according to the location of its best mapping. We did not attempt to map the Alu reads to their exact location in the genome (as the best mapping of Alu reads are probably one of many best mappings, as they appear in many locations in the genome) but simply infer from their best mapping if they are Alus or not.

Calculating miRNA targets representation in the transcriptome: For each miRNA of the 10 most abundant miRNAs in the PAR-CLIP experiment, the percentage of expressed miRNA targets within Alus was calculated, according to the mRNA-Seq experiment that was done by Kishore *et al.* (2011) (GSM714678 and GSM714679, which mimic best the conditions of the PAR-CLIP experiment). In each transcript, the numbers of miRNA targets in total and within Alus were calculated, and multiplied by the count of the transcript. Transcripts with low count (<10) were excluded. The average of the two replicates is represented in the analysis.

3 RESULTS

3.1 Potential genomic interplay between Alus and miRNAs

To evaluate the potential effect of Alu insertions in the human genome on miRNA targeting, we first examined how many potential miRNA targets reside within Alu sequences in genes' 3′UTRs. This analysis showed that 16% of human genes contain at least one Alu in their 3′UTR. A total of 4927 Alu sequences that reside within 3′UTRs present 94 785 potential miRNA targets (defined as 7-mers with perfect match to positions 2–8 of the miRNA), 28 829 of which correspond to a set of 401 miRNAs that are conserved among mammals (Dahary *et al.*, 2010). Of these, there are 3088 predicted targets for the 74 most conserved miRNAs in the animal kingdom, which are therefore considered as involved in basic cellular processes (Dahary *et al.*, 2010). Hence, at least in theory, the spread of Alus in the human genome has a great potential to affect miRNA-based regulation of gene expression.

3.2 A comprehensive dataset of transcriptome-wide effects of miRNAs

To examine the effects of Alu insertions in 3′UTRs on gene expression, we analysed data from reported miRNA overexpression experiments. Khan *et al.* have recently assembled the results of dozens such genome-wide expression array and proteomics experiments into a single normalized database (Khan *et al.*, 2009). For subsequent analysis, we used a subset of experiments from the Khan database, comprising 43 experiments with 23 different miRNAs overexpressed in a total of five different cell lines. For each experiment, the dataset provides the genome-wide mRNA response to the overexpression of one miRNA at a time in a given cell line.

As a preliminary step, we assessed the potential of this dataset to demonstrate known attributes of miRNA regulation. First, we examined whether genes that contain a predicted binding target for a particular miRNA are more likely than other genes to be downregulated in response to overexpression of that miRNA. Reassuringly, Figure 1A shows that the percentage of downregulated genes was significantly higher in the group of genes that contain a putative binding site for the overexpressed miRNA, relative to the group of genes lacking such binding site ($P = 2.9e\text{-}21$, Student's t-test). In addition, the average fold change of the downregulated genes on overexpression of the miRNA was significantly higher in the group of genes containing the binding site compared with the downregulation that is occasionally observed among the control-set genes ($P = 3.8e\text{-}12$, Student's t-test; Supplementary Fig. S1A). Moreover, as already suggested by others (Grimson *et al.*, 2007; Nielsen *et al.*, 2007), genes with more than one putative binding site for a given miRNA are more efficiently downregulated than genes with only a single target ($P = 5.5e\text{-}5$, Student's t-test; Fig. 1B).

The mere existence of a miRNA binding site sequence inside the 3′UTR of a gene does not necessarily imply that the gene will constitute a functional target for the miRNA. A commonly accepted hallmark of a target's authentic functionality is its evolutionary conservation (Brennecke *et al.*, 2005; Grimson *et al.*, 2007; Lewis *et al.*, 2005). Therefore, we next compared between genes that contain a conserved versus non-conserved binding motif for each overexpressed miRNA. Figure 1C shows that the group of genes with conserved motifs has a significantly larger percentage of downregulated genes relative to human genes that contain the same motif, but this motif is not conserved in the orthologous genes of other mammals ($P = 5.6e\text{-}14$, Student's t-test). The same is true for the mean fold reduction ($P = 1.4e\text{-}4$, Student's t-test; Supplementary Fig. S1B).

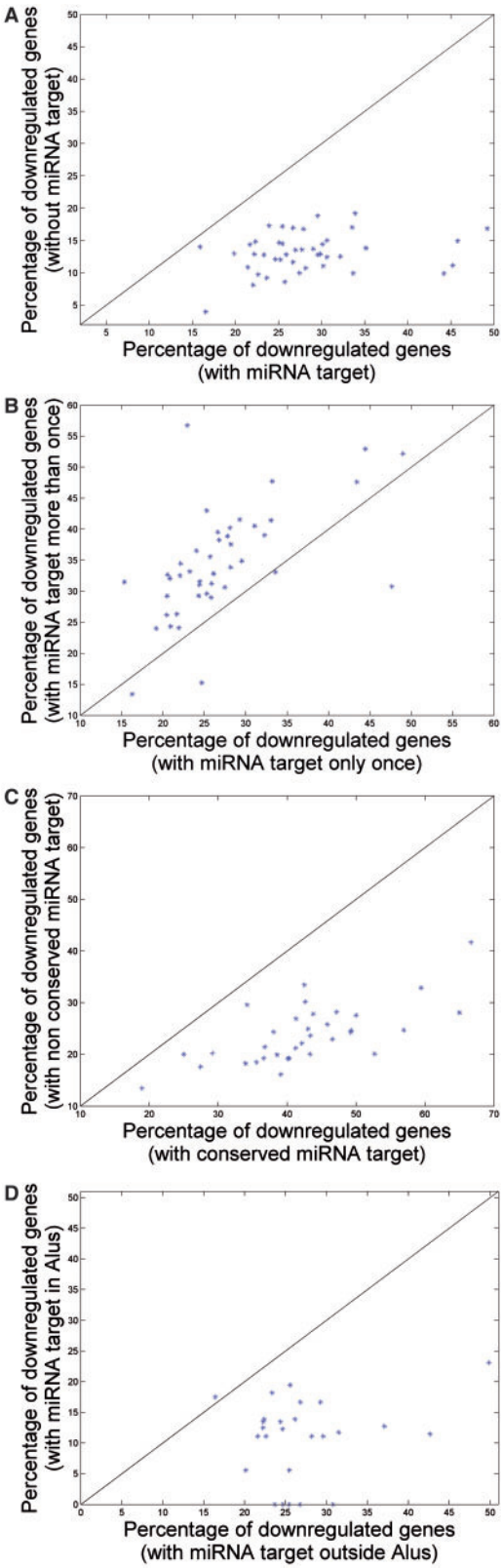


Fig. 1. The effect of miRNA overexpression on mRNA abundance. Each dot represents a single experiment in which a certain miRNA was overexpressed. Each experiment is plotted according to the percentage of downregulated genes in each group. (**A**) Genes with and without the miRNA sites. (**B**) Genes with the miRNA target only once versus more than once. (**C**) Genes with conserved versus not conserved miRNA sites. (**D**) Genes with miRNA sites only outside Alus versus only within Alus. (**E**) Fold change distribution of all genes containing miR-106b target within and outside Alus in HCT cells

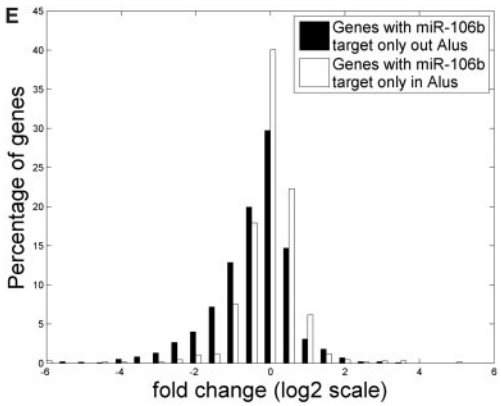


Fig. 1. Continued

Together, these results demonstrate that the miRNA overexpression dataset reflects faithfully established parameters of miRNA regulation and can therefore be further used to examine other features such as the effect of miRNA motifs inside Alu sequences.

3.3 The miRNA binding sites inside Alus are less functional

To discriminate between the effect of miRNA targets inside versus outside Alus, we defined for each miRNA in the dataset two sets of genes: the first consists of genes in which the putative miRNA binding motif resides only inside Alus, and the second contains the motif only outside Alus. Supplementary Table S1 summarizes the miRNAs from the Khan dataset and the number of potential target genes in each of the two sets. As shown in Figure 1D, genes with motifs only outside Alus are more likely to be downregulated than genes with motifs only inside Alus ($P = 2.2e\text{-}11$, Student's t-test). The same trend appears also when assessing the average extent of downregulation—the fold change on miRNA overexpression is significantly higher for genes containing the miRNA binding site only outside Alus ($P = 0.03$, Student's t-test; Supplementary Fig. S2A). To control for the possibility that regions outside Alus are longer and thus possess more than one miRNA target, and as multiplicity of targets contributes to the miRNA functionality, we repeated the aforementioned analysis examining only genes with a single miRNA target. The trend remained highly significant ($P = 7.4e\text{-}11$, Student's t-test; Supplementary Fig. S2E).

We next compared the response to overexpression of a given miRNA between genes that contain its putative binding site only within Alus and genes that lack such target altogether. Remarkably, the two groups were found to respond rather indistinguishably to the overexpression of the relevant miRNA ($P = 0.3$, Student's t-test; Supplementary Fig. S2B). This suggests

that the majority of potential miRNA target sites within Alu elements are actually non-functional.

Alus are primate-specific, and therefore any motif inside an Alu is by definition not conserved in mammals. We therefore set out to rule out the possibility that the lower functionality of putative miRNA binding sites inside Alus is not unique to Alus but is rather a reflection of their lack of evolutionary conservation. To this end, we compared the effect of Alu-contained putative miRNA targets to non-conserved targets residing outside Alus. As shown in Supplementary Figure S2C and D, this analysis revealed that miRNA binding sites inside Alus are less functional even when compared with non-conserved miRNA binding sites outside Alus (Student's *t*-test, *P*-values of 4.9e-4 and 0.001, respectively). Hence, the low level of conservation cannot account for the low functionality of Alu-contained putative miRNA targets.

The analysis of a comprehensive miRNA overexpression dataset enables the examination of the various miRNA target attributes and their respective effects on mRNA expression. Supplementary Figure S2F recapitulates, with the current dataset, the established knowledge regarding miRNA target characteristics, showing that the most important features are the multiplicity of sites within a 3′UTR and their conservation. The presence of an Alu thus appears as an additional important attribute that should serve in evaluating a miRNA site: a site within an Alu is typically less functional (Supplementary Fig. S2F).

Smalheiser and Torvik (2006) reported a conserved site within Alus, which comprises the target of numerous miRNAs—GCACUU. They suggested that these miRNAs target Alu sequences. To test this possibility, we focused on two miRNAs in the Khan dataset, miR-373 and miR-106 b, which contain this sequence in their target. Comparing the fold change distributions in these specific overexpression experiments, we find that the miRNA targets within Alus are still significantly less functional than the targets outside Alus across various cell lines (Fig. 1E, $P = 2.5e\text{-}17$; Supplementary Fig. S3A–C, $P = 0.003$, 7.6e-43 and 9.6e-35, respectively, Student's *t*-test). Thus, although miRNA sites within Alus may certainly be functional in many particular cases, these genome-wide findings indicate that miRNA targets within Alus are often less functional.

Realizing that legitimate 7-mer perfect match targets within Alus are often not functional, we looked for features that would explain such lack of functionality. Conversely, absence of such features in exceptional cases may highlight potential novel functional miRNA targets that were inserted through Alu retrotransposition.

3.4 Mutually distinct 3′UTR territories of conserved miRNA targets and Alu repeats

The first feature we explored that might explain why miRNA targets within Alus are often not functional is the location of Alu insertions within 3′UTRs. It was previously shown that conserved and functional miRNA targets tend to reside at both ends of 3′UTRs and less in the UTR's middle (Grimson *et al.*, 2007). Figure 2A recapitulates this finding for relative position along the 3′UTR, showing that conserved miRNA binding sites are enriched near the two ends of the 3′UTR. Examining only 3′UTRs longer than 1000 bps, we found that conserved miRNA

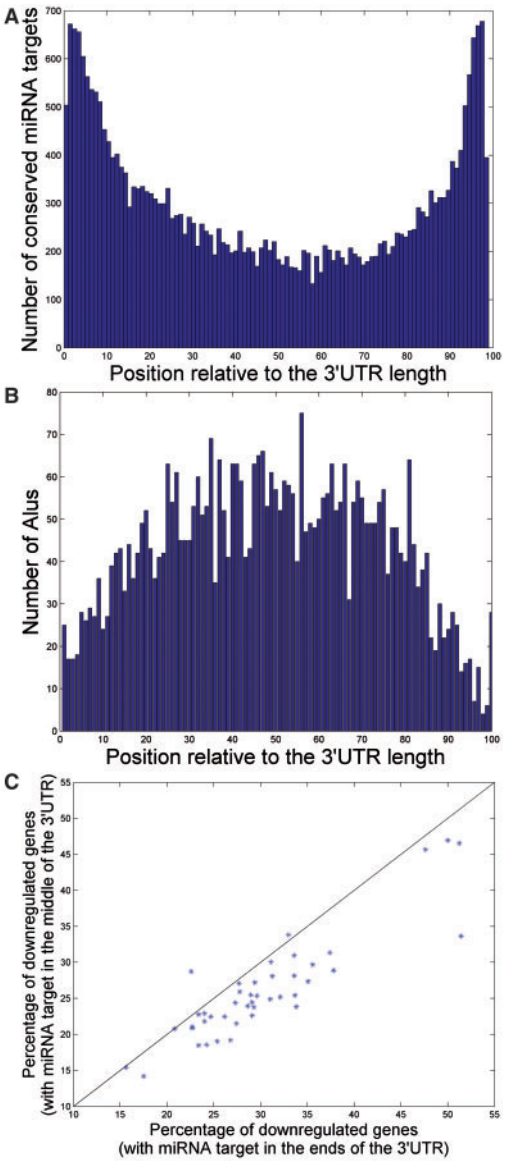


Fig. 2. Alus and miRNA targets territories within 3′UTRs. (**A**) Distribution of conserved miRNA sites along 3′UTRs. (**B**) Distribution of Alus along 3′UTRs. Only 3′UTRs longer than 1 kb were analysed. The *x*-axis depicts the relative position on the 3′UTR (normalized to its length). (**C**) Each dot represents a single experiment in which a certain miRNA was overexpressed. Each experiment is plotted according to the percentage of downregulated genes in two groups: genes with the miRNA target in the middle or in the 3′UTR ends

binding motifs are concentrated in the first and last 250 bps of the UTR and are relatively depleted from the middle section (Supplementary Fig. S4A). In contrast, non-conserved targets are evenly distributed throughout the 3′UTRs (Supplementary Fig. S4B). Reassuringly, this localization appears to have an interesting functional correlate: Figure 2C shows that predicted miRNA binding sites near the ends of the 3′UTR tend to be more functional, i.e. to have a more pronounced response to overexpression of the miRNA compared with targets in the middle of the UTR ($P = 0.01$, Student's *t*-test).

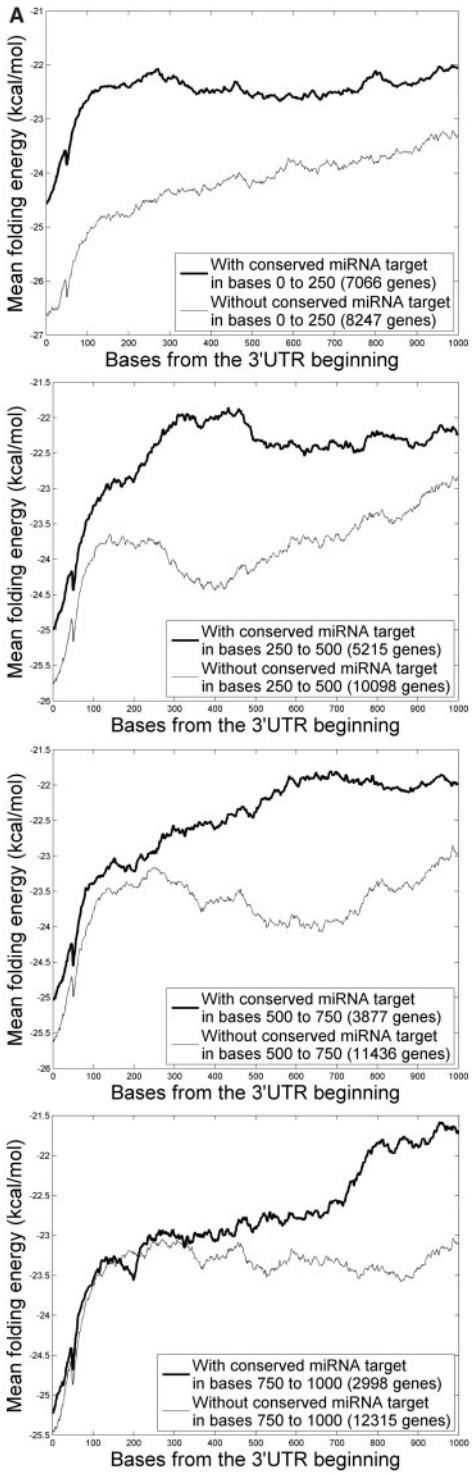


Fig. 3. Folding energy of mRNA secondary structures around miRNA target sites. (**A**) The mean folding energy in the first 1000 bp of 3′UTRs was calculated as explained in Section 2. The genes were divided according to whether they contain conserved miRNA targets in each of four specified 3′UTR location quadrants. The analysis was done only for genes with 3′UTRs longer than 1000 bp, and the 3′UTRs were aligned to their 5′ most point. (**B**) Distribution of the folding energy around conserved miRNA target sites and miRNA sites within Alus

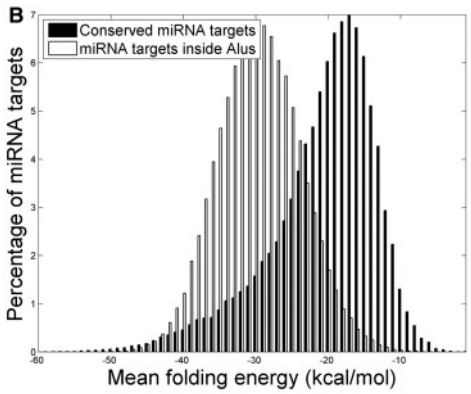


Fig. 3. Continued

We next examined the positions of Alu insertions into 3′UTRs. Strikingly, we found that the majority of Alu insertions have occurred in the middle sections of the 3′UTR, presenting an almost exact mirror image of the conserved miRNA recognition motifs' location (Fig. 2B and Supplementary Fig. S4A). These results suggest that Alu insertions into 3′UTRs during primate evolution were tolerated, provided that they occurred away from the two 3′UTR ends. It is tempting to suggest that a reason for such Alu avoidance from the ends is that those territories are preserved for critical regulatory interactions, including functional miRNA binding sites. In contrast, the middle section of the 3′UTRs is proposed to be relatively more depleted of such functional elements and therefore more likely to tolerate genomic infringements such as Alu insertions.

At this stage, it could be suggested that the functional failure of Alu-contained putative miRNA binding sites to serve as targets is merely due to their unfavourable location—away from the ends of the 3′UTR. However, a comparison of genes with putative miRNA binding sites within Alus to genes with targets only outside Alus but in the middle of the 3′UTR revealed that the non-Alu targets in the UTRs' middle are still more functional than the targets in Alus ($P = 7.4e\text{-}10$, Student's t-test; Supplementary Fig. S4C). Moreover, some Alus were inserted near 3′UTR ends, and numerous functional conserved miRNA targets are located in the middle of 3′UTRs (Fig. 2A and B). Therefore, location within 3′UTR appears to be an important, but not exclusive, evolutionary constraint on Alu insertion. We thus turned to examine additional features of Alus that may allow their insertions into functional regions of 3′UTRs without altering gene expression.

3.5 Putative miRNA targets in Alus have tight RNA secondary structure

The RNA structure and folding energy of miRNA targets and their surroundings are important for their functionality; in particular, targets located within mRNA regions possessing tight secondary structure are typically less functional (Hausser *et al.*, 2009; Kertesz *et al.*, 2007). In agreement with these findings, we too find that genes harbouring a conserved miRNA binding site at a given location along the 3′UTR tend to have a significantly less tight secondary structure at that region compared with genes without any conserved miRNA target at that location (Fig. 3A).

This appears to hold for all possible locations along the 3′UTR, including the two ends and the middle section.

As tight structure around miRNA binding sites might provide an additional feature that reduces the potential regulatory effect of Alu insertion, we examined the folding energy around targets inside Alus. Indeed, we found that targets inside Alus tend to reside within tighter structures, as compared with conserved miRNA binding sites (Fig. 3B), an observation compatible with the known high RNA secondary structure content of Alu repeats (Labuda and Striker, 1989; Okada, 1990). Hence, tight local secondary structure of some Alus may have allowed their insertion into legitimate regions without imposing major regulatory effects. We note that the RNA folding calculations were performed here on sequence windows of 100 nucleotides—shorter than the length of an Alu element—thus capturing a tendency of single Alu elements to fold on themselves. On top of that, two Alus with opposite orientations could form tight hairpins, thus increasing the actual extent of secondary structure and hence potentially further augmenting miRNA targeting avoidance.

The tightness of the secondary structure found around miRNA targets can also be affected by nucleotide content. In particular, Grimson *et al.* (2007) reported that the AU content around effective miRNA targets is higher than around less effective targets. We therefore examined the AU content in the vicinity (10 nucleotides upstream plus 10 nucleotides downstream) of targets within Alus, and compared it with the AU content of the most effective miRNA targets within the Khan dataset. In agreement with Grimson *et al.*, and as shown in Supplementary Figure S5A, the vicinity of miRNA targets within Alus is characterized by a significantly lower AU content ($P<$e-300, Student's *t*-test). Nevertheless, even when we compared the average fold change only between genes with similar AU content, targets within Alus were still less functional than corresponding matched targets outside Alus (Supplementary Fig. S5B). Thus, even though a lower AU content may well contribute to a more stable secondary structure in the surroundings of Alu-contained miRNA targets, the poor functionality of such targets cannot be solely attributable to the lower AU content.

3.6 The miRNA targets within Alus might be altered by RNA editing

Alu sequences are subject to extensive RNA editing (Athanasiadis *et al.*, 2004; Blow *et al.*, 2004; Kim *et al.*, 2004; Levanon *et al.*, 2004; Morse *et al.*, 2002), which modifies adenosines (A) to inosines (I) by adenosine deaminase acting on RNA (ADAR) enzymes. The majority of editing events in human tissues occur within Alus (Barak *et al.*, 2009). As inosines are recognized as guanosines by many of the molecular machineries in the cell, such alterations can diminish the complementarity between a miRNA's seed and its binding site within Alus, or introduce novel targets by creating complementarity with the miRNA's seed (Borchert *et al.*, 2009).

To interrogate the potential impact of RNA editing on the recognition of Alu-contained putative miRNA targets, we first singled out those targets that do not contain an A in their seed-complementary target sequence and therefore cannot be subject to editing. Of the 18 different miRNA 7-mer target sequences in the Khan dataset, three were found to have no A within the 7-mer. These three targets were represented in eight miRNA overexpression experiments, for which we could compare between genes that have the miRNA target only inside or only outside Alus (as described in Section 2). For comparison, we had 18 overexpression experiments with A-containing miRNA targets. RNA editing is an additional layer that may assist the cell to reduce the effect of Alu insertions on gene regulation, and therefore we expect that miRNA targets without an A within Alus will be more functional than the ones with an A.

When examining only genes with putative miRNA binding sites without A, targets within Alus are moderately less functional than targets outside Alu ($P=0.02$, Student's *t*-test; Fig. 4A), probably owing to the effects of territory and secondary structure discussed earlier in the text. Notably, when we examined only targets with A, the targets within Alus were found to be substantially less functional than those outside Alus ($P=5.8$e-6, Student's *t*-test; Fig. 4B). This observation suggests that among Alu-contained sites, miRNA targets without A are more effective than A-containing ones. It is tempting to propose that this trend is due to the fact that A-containing targets within Alus are likely to be edited, thereby disrupting the recognition by the cognate miRNA. Thus, RNA editing might constitute an additional feature through which Alu insertions into mRNAs were tolerated.

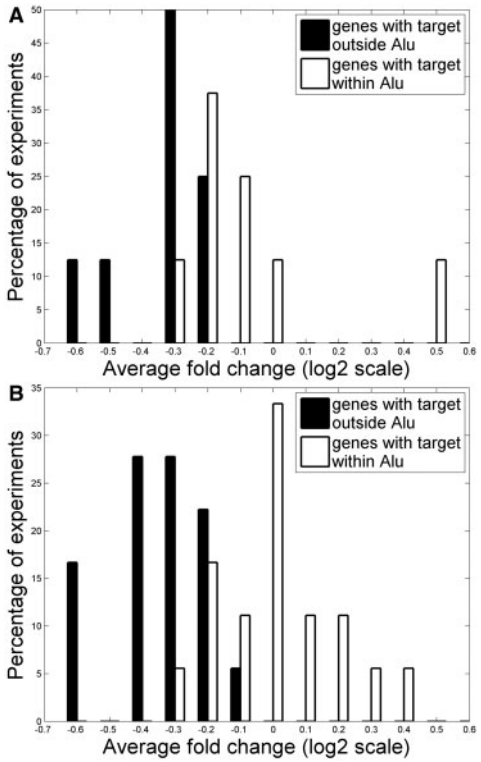


Fig. 4. The effect of the presence of A (adenosine, which might be edited by ADAR) in miRNA target sequences. Average fold change of genes that contain the miRNA target within and outside Alus, for overexpression experiments with miRNAs that do not contain an A in their target motif (**A**), or contain an A (**B**)

3.7 Repeats and miRNA targets in the mouse genome

Transposable elements are active in most animal genomes. Therefore, introduction of novel miRNA targets via transposition can occur in other species as well. Although Alu repeats are primate specific, the mouse genome too contains repeats similar to Alus, namely B1 repeats that belong to the Short Interspersed Elements family. As in the case of Alus, B1 repeats emerged from the ancestral 7SL RNA gene (Ullu and Tschudi, 1984). The B1 repeats are less widespread than Alus, comprising only 2.7% of the mouse genome (Lander *et al*., 2001; Pruitt *et al*., 2005) and are also shorter (~140 bp) (Smalheiser and Torvik, 2006).

In the mouse genome, 8.3% of the genes contain at least one B1 repeat in their 3′UTR. The 1962 B1 sequences that reside within 3′UTRs represent 14 372 potential miRNA targets (perfect 7-mers). Consequently, the potential effect of putative miRNA targets within B1 repeats is less substantial than that of Alus in the human genome; it is nonetheless not negligible. We observed that with regard to their location within the 3′UTR, mouse B1 repeats show similar trends as the human Alus. B1 repeats tend to avoid the two ends of the 3′UTR, predominantly the beginning, while the conserved mouse miRNA targets display an opposite trend of preferential location near the UTR ends (Supplementary Fig. S6A).

Further, putative miRNA targets within B1 repeat-encoded mRNAs show tighter local secondary structure (Supplementary Fig. S6B). The fact that this feature is shared with Alus is probably explained by the common evolutionary origin of these two types of repeats (Ullu and Tschudi, 1984).

In conclusion, like Alus in the human genome, B1 insertions into mouse mRNAs were probably tolerated, provided that they occur into 3′UTR territories that do not overlap with functional miRNA targets or that they possess tight secondary structure.

3.8 Lack of binding of the miRNA machinery to Alu-contained target sites

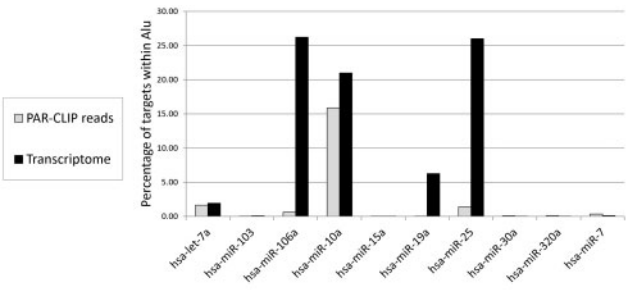
The compromised functionality of putative miRNA targets within Alus can be due to a failure of the miRNA machinery to bind such targets or due to dysfunctionality after binding occurs. To distinguish between these two possibilities we analysed data from Ago2-mRNA binding experiments.

Ago2 is part of the Argonaute family of proteins, which are guided by the mature miRNA to bind the specific complementary region of the target mRNA to initiate its silencing (Ender and Meister, 2010). Therefore, profiling of Ago2-bound mRNA species could serve as a means to assess how efficiently is a given RNA sequence bound by the RNA-induced silencing complex (RISC) machinery.

To specifically and accurately analyse mRNA regions that are bound by the RISC machinery genome wide, we used photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP) data generated by Kishore *et al*. (2011). PAR-CLIP with Ago2 identifies mRNA molecules that were targeted by miRNAs (Hafner *et al*., 2010). Typically, the outputs of PAR-CLIP experiments are short reads of bound RNA segments, obtained in conjunction with mRNA-Seq profiling of the transcripts expressed in cells exposed to the same conditions. In addition, the identities of the most abundant miRNAs in the analysed cell population can be deduced directly from the PAR-CLIP data (Kishore *et al*., 2011); a total of 10 miRNAs were identified as the most abundant within the cells used in this recent experiment.

To examine the potential functionality of miRNA targets that reside within Alu elements relative to those residing outside Alus, we mapped PAR-CLIP reads and annotated them according to their genomic context. We did not attempt to determine the exact mapping of each read onto the genome in this analysis (as such mapping would be particularly challenging with repetitive elements; see Section 2). Instead, we merely aimed to determine here whether a read resides within an Alu. Reassuringly, we found that ~20% of the mapped reads contained a target for one of the 10 miRNAs identified as most abundant in those cells. Analysing the mRNA-Seq data, we found that potential targets of all 10 top miRNAs are highly represented in the transcriptome. Of these, 3 miRNAs, miR-106 a, miR-25 and miR-10, have a high percentage of their expressed potential targets within Alus (Fig. 5). As such, these miRNAs can be used for the comparison of binding targets within and outside Alus. For instance, miR-106 a has 27% of its expressed potential targets within Alus, but strikingly only 0.61% of its associated reads could actually be mapped to Alus ($P <$ e-300, HyperGeometric test; Fig. 5). The reads that do not contain predicted targets for any of the 10 most abundant miRNAs (~80% of the reads) can serve as a control for non-specific binding to Ago2. We found that the percentage of Ago2-associated reads mapped to Alus is very similar to the expected Alu content of the transcriptome of these cells. Essentially similar results were obtained also with the other two miRNAs (miR-10a, $P = 3.8$ e-4; miR-25, $P <$ e-300; HyperGeometric test; Fig. 5), supporting the generality of our observations.

Together, these findings provide independent experimental support to the notion that insertion of Alu-contained miRNA sites into 3′UTRs was largely tolerated only when they could escape Ago2 binding. These results further argue that miRNA targets within Alus are preferentially not associated with Ago2 and the RISC complex, implying that they are strongly disregarded by the miRNA machinery. Thus, although our analysis of the Khan dataset clearly shows that targets within Alus are



Reads with target	5249	6385	13988	479	15675	10294	7265	2792	790	1403
Reads with target within Alu	90	3	86	76	2	5	102	3	1	5

Fig. 5. Representation of miRNA targets in PAR-CLIP reads. For each of the 10 miRNAs identified by the PAR-CLIP experiment as being most abundant in the analysed cells, the percentage of predicted targets within Alus is compared between the overall transcriptome and the PAR-CLIP reads. Bellow the *x*-axis, the table depicts the absolute number of PAR-CLIP reads containing the putative miRNA target

not effective in mRNA destabilization, the PAR-CLIP analysis strongly suggests that these targets are not functional in translation inhibition as well.

4 DISCUSSION

In the present study, we demonstrate that potential miRNA targets within Alus are largely non-functional and are not bound by the miRNA machinery. We suggest that insertions of Alus into mRNAs were tolerated largely when miRNA targets within them were likely to be less functional.

Despite the strong indications that Alu-contained putative miRNA target sites do not tend to affect gene expression, there are clearly cases where such targets within Alus can be functional. In particular, Smalheiser and Torvik (2006) described many mRNAs that contain Alus in their 3′UTR, within which there are target sites for dozens of miRNAs. In their study, most of the miRNAs suggested to target Alus were within the C19MC cluster (Smalheiser and Torvik, 2006), which is a primate-specific cluster that contains many Alu sequences that might have facilitated its expansion (Zhang *et al.*, 2007). These miRNAs might have evolved in coordination with the Alu sequences to create an effective targeting. Lehnert *et al.* (2009) reported that there are a few miRNAs with >1000 predicted sites per megabase within Alu sequences and proposed that such miRNAs protect against Alu transposition. We find that although the potential regulatory effect of Alus is huge, their actual contribution to regulation of gene expression by the miRNA machinery might be limited. Clearly, this does not exclude other regulatory roles of the miRNA-Alu interplay, such as a role of miRNAs in maintaining genomic stability by the repression of transposable elements (Shalgi *et al.*, 2010).

The three features that permit Alu insertions into the 3′UTR of genes are inherently very different from one another. Mutually exclusive territories imply an evolutionary mechanism, as it appears that Alu insertions near 3′UTR ends were selected against. One intriguing hypothesis is that Alus inserted near the ends of 3′UTRs might have forced important miRNA targets to move towards the middle of the 3′UTR, where they become less effective. Another scenario, supported by our findings, is that Alus near the ends might have introduced new miRNA binding sites at locations where such targets are likely to be highly functional, grossly disrupting the conserved regulation of the gene. Additionally, it is conceivable that insertion of Alus near the ends of 3′UTRs may be deleterious also for reasons that are unrelated to miRNA function.

The tight secondary structure presents an inherent feature of the Alu itself—Alus have high RNA secondary structure content (Labuda and Striker, 1989; Okada, 1990), a property that might have allowed insertion without a major effect on gene regulation, as miRNA targets are less functional within tight secondary structure (Kertesz *et al.*, 2007).

The editing mechanism presents another layer, which has the potential to be regulated at a cellular level, as its impact might vary greatly among different cell types as a function of their editing capacity. This could potentially contribute to changes in the transcriptome during developmental processes, as well as in response to any internal or external signal that affects editing efficiency. Importantly, levels of ADAR enzymes, which perform the A-to-I editing, are altered in cancer (Paz *et al.*, 2007). It is thus tempting to speculate that miRNA targets within Alus escape editing in cancer cells, leading to an elevation in their functionality. Such mechanism might contribute in interesting ways to post-transcriptional deregulation of gene expression patterns in cancer. In addition, RNA editing has the potential to create new potential miRNA targets, as previously suggested by Liang and Landweber (2007).

RNA-binding proteins (RBPs) can also affect the potential of the miRNA machinery to bind and act on the mRNA. Jacobsen *et al.* (2010) showed that the presence of binding motifs for specific RBPs within mRNAs can affect the way in which these mRNAs are regulated by miRNAs for which they contain targets. Specifically, mRNAs that were downregulated on miRNA overexpression were found to be enriched for two U-rich motifs that bind the protein ELAVL4, whereas mRNAs upregulated in response to miRNA overexpression were enriched for an AU-rich element (Jacobsen *et al.*, 2010). We therefore tested whether such RBP-binding elements may account for the reduced efficacy of Alu-contained miRNA targets. However, we could not find evidence for a selective contribution of the three RNA-binding motifs described in Jacobsen *et al.* to the differential efficacy of targets inside versus outside Alus. Rather, we found that these motifs are evenly distributed in mRNAs with the target inside, or outside of Alus, and thus their existence or depletion cannot explain the lack of functionality of miRNA targets within Alus. Yet, the existence of other RBP-binding motifs that are unequally present in these two types of target mRNAs cannot be ruled out and should be the subject of future analyses.

We also examined the potential for combined effects of the various features on Alu-contained miRNA targets. Supplementary Figure S8 presents a Venn diagram addressing the characteristics applicable to each miRNA target inside Alu for all conserved mammalian miRNAs. Most of the targets appear to use more than one feature, supporting the conjecture that an interaction between two or more features is needed to reduce dramatically the target's functionality. Importantly, consistent with the earlier suggestion of Smalheiser and Torvik (2006), there are 416 miRNA targets that might be functional as they are located in the ends of the 3′UTR, have a loose secondary structure and cannot be edited. When looking at all targets within Alus, which could potentially escape the location and structure criteria, we find ~8000 targets. Among the top 10 targets are two with the GCACUU site, which appear in many Alus, and comprises the target of numerous miRNAs, as discussed by Smalheiser and Torvik (2006). However, in a randomized simulation, its appearances in the escaping targets were not found statistically significant. Our ability to identify at least some of these characteristics enables to single out the more relevant miRNA targets and subject them to future functional studies.

ACKNOWLEDGEMENTS

The authors thank the Pilpel and Oren laboratories for discussions. They thank Erez Levanon for useful discussions. They thank Sivan Navon for fruitful ideas in the mRNA structure prediction and analysis. Y.P. is an incumbent of the Ben-May

Professorial Chair, M.O. is incumbent of the Andre Lwoff chair in Molecular Biology. The EC is not liable for any use that may be made of the information contained herein.

Funding: The authors thank the EC FP7 funding (ONCOMIRS, grant agreement number 201102), the 'Ideas' program of the European Research Council and the Ben May Charitable Trust for grant support.

Conflict of Interest: none declared.

REFERENCES

Athanasiadis,A. *et al.* (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.*, **2**, e391.

Barak,M. *et al.* (2009) Evidence for large diversity in the human transcriptome created by Alu RNA editing. *Nucleic Acids Res.*, **37**, 6905–6915.

Batzer,M.A. and Deininger,P.L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.*, **3**, 370–379.

Blow,M. *et al.* (2004) A survey of RNA editing in human brain. *Genome Res.*, **14**, 2379–2387.

Borchert,G.M. *et al.* (2009) Adenosine deamination in human transcripts generates novel microRNA binding sites. *Hum. Mol. Genet.*, **18**, 4801–4807.

Brennecke,J. *et al.* (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.

Dahary,D. *et al.* (2010) CpG Islands as a putative source for animal miRNAs: evolutionary and functional implications. *Mol. Biol. Evol.*, **28**, 1545–1551.

Ender,C. and Meister,G. (2010) Argonaute proteins at a glance. *J. Cell Sci.*, **123**, 1819–1823.

Fujita,P.A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.

Griffiths-Jones,S. (2004) The microRNA registry. *Nucleic Acids Res.*, **32**, D109–D111.

Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.

Griffiths-Jones,S. *et al.* (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.

Grimson,A. *et al.* (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.

Hafner,M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.

Hausser,J. *et al.* (2009) Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res.*, **19**, 2009–2020.

Jacobsen,A. *et al.* (2010) Signatures of RNA binding proteins globally coupled to effective microRNA target sites. *Genome Res.*, **20**, 1010–1019.

Kertesz,M. *et al.* (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.

Khan,A.A. *et al.* (2009) Transfection of small RNAs globally perturbs gene regulation by endogenous MicroRNAs. *Nat. Biotechnol.*, **27**, 549–555.

Kim,D.D. *et al.* (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.*, **14**, 1719–1725.

Kishore,S. *et al.* (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods*, **8**, 559–564.

Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.

Labuda,D. and Striker,G. (1989) Sequence conservation in Alu evolution. *Nucleic Acids Res.*, **17**, 2477–2491.

Lander,E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

Langmead,B. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

Lehnert,S. *et al.* (2009) Evidence for co-evolution between human microRNAs and Alu-repeats. *PLoS One*, **4**, e4456.

Lev-Maor,G. *et al.* (2003) The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science*, **300**, 1288–1291.

Levanon,E.Y. *et al.* (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.*, **22**, 1001–1005.

Lewis,B.P. *et al.* (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

Liang,H. and Landweber,L.F (2007) Hypothesis: RNA editing of microRNA target sites in humans? *RNA*, **13**, 463–467.

Mathews,D.H. *et al.* (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.

Morse,D.P. *et al.* (2002) RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. *Proc. Natl Acad. Sci. USA*, **99**, 7906–7911.

Nielsen,C.B. *et al.* (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.

Okada,N.J. (1990) Transfer RNA-like structure of the human Alu family: implications of its generation mechanism and possible functions. *Mol. Evol.*, **31**, 500–510.

Paz,N. *et al.* (2007) Altered adenosine-to-inosine RNA editing in human cancer. *Genome Res.*, **17**, 1586–1595.

Polak,P. and Domany,E. (2006) Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics*, **7**, 133.

Pruitt,K.D. *et al.* (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.

Shalgi,R. *et al.* (2010) Repression of transposable-elements - a microRNA anti-cancer defense mechanism? *Trends Genet.*, **26**, 253–259.

Smalheiser,N.R. and Torvik,V.I. (2006) Alu elements within human mRNAs are probable microRNA targets. *Trends Genet.*, **22**, 532–536.

Smit,A.F.A. *et al.* (1996–2010) *RepeatMasker Open-3.0*. http://www.repeatmasker.org (11 February 2013, date last accessed).

Ullu,E. and Tschudi,C. (1984) Alu sequences are processed 7SL RNA genes. *Nature*, **312**, 171–172.

Wuchty,S. *et al.* (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145–165.

Zhang,R. *et al.* (2007) Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res.*, **17**, 612–617.