

Supplementary data

Alu elements within human mRNAs are probable microRNA targets

Neil R. Smalheiser and Vetle I. Torvik

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

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

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.

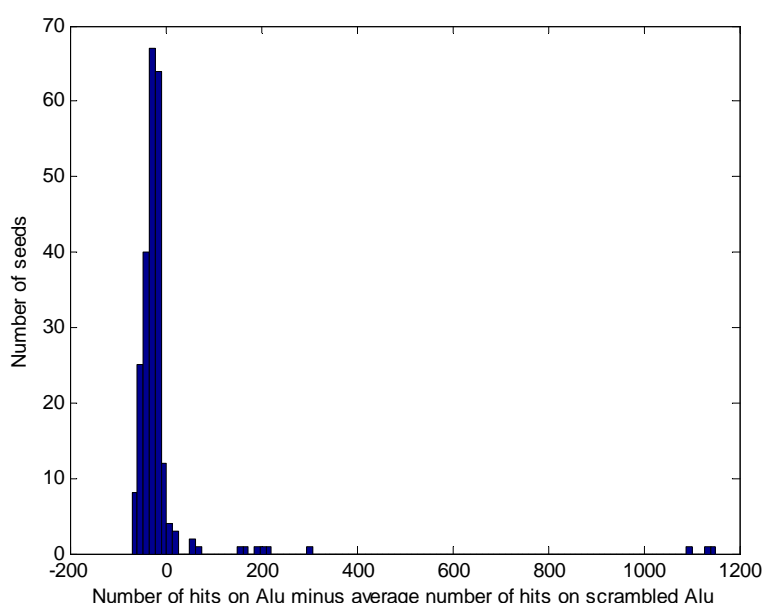
Published in *Trends in Genetics* (2006) Vol. 22, issue 10

Supplemental file 1

Unique 5'-seeds of human microRNAs hitting Alu repeats vs. scrambled versions of Alu repeats.

Each microRNA seed sequence (length 8, beginning at position 2) was examined for perfect complementarity with the set of all 4388 annotated Alu repeats within human 3'-UTRs in RefSeq. To test whether the observed number of hits was greater than expected by chance, we counted the number of hits that occurred upon a set of scrambled Alu sequences (each annotated Alu repeat (throughout the set of all 3'-UTRs) was permuted randomly 100 times maintaining dinucleotide composition).

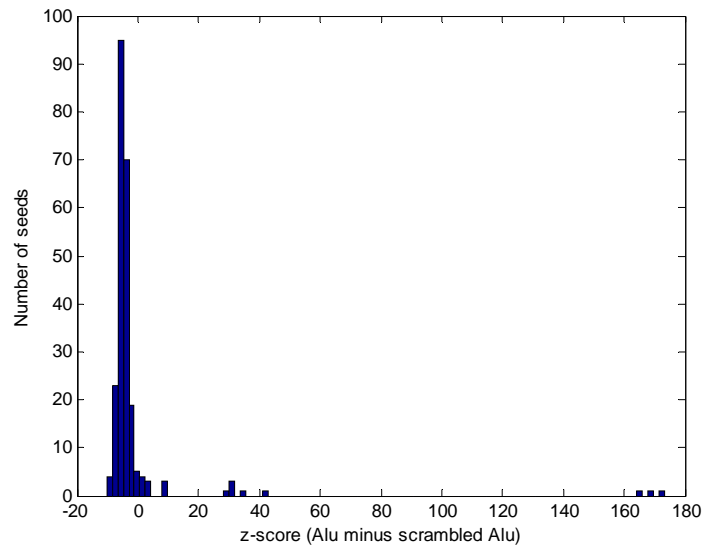
For each seed, we plotted the total number of hits upon Alu repeats across all 3'-UTRs, minus the average number of hits upon the set of scrambled Alu sequences. (All microRNAs hit a relatively low average number of scrambled sequences (mean = 36.2 ± 14.2 SD) and showed relatively little variation across different microRNAs).



Supplementary Figure 1a.

The raw data in Figure 1a were replotted to show the corresponding z-scores. For each microRNA seed, the z-score = (the number of hits upon Alu repeats minus the average number of hits upon scrambled Alu repeats, divided by the SD of the number of hits across 100 replicates of the scrambled set). A z-score >4 is equivalent to a p-value <0.0001.

The 235 seeds fell into three classes: most (223) showed no significant excess of hits upon Alu compared to scrambled Alu. Nine seeds showed z-scores of 5–50; it is possible that these might represent biologically relevant hits upon 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, AAGUGCUG).



Supplementary Figure 1b.

Supplementary file 2

Set of all Human RefSeq 3'-UTRs that contained two or more target sites, corresponding to one or both of the 8-mer outlier seeds.

See separate Excel file.