

## **Complications in Mammalian microRNA Target Prediction**

Neil R. Smalheiser, MD, PhD and Vetle I. Torvik, PhD

Department of Psychiatry and UIC Psychiatric Institute,

University of Illinois-Chicago, MC912,

1601 W. Taylor Street

Chicago, IL 60612 USA

312-413-4581

fax 312-413-4569

neils@uic.edu

### **i. Summary.**

In this chapter, we review evidence that at least 3 different types of microRNA-mRNA target interactions exist in mammals: short-seeds, long-seeds, and “perfect” hits (allowing G:U matches). Since new types of microRNAs are still being discovered, this list may not yet be complete.

### **ii. Key Words.**

MicroRNA, microRNA target prediction, RNA interference, LINE-2, genomic repeats

## 1. Introduction

The phenomena of RNA interference were first defined in plants and *C. elegans*, and have subsequently been studied in all living phyla including mammals (1). Among several different populations of small RNAs that have been identified, the microRNAs have perhaps received the most attention because they arise from well-defined genomic precursors, because they appear to regulate mRNA stability and/or translation, and because they are clearly important for development and cell differentiation (see reviews in refs. 2-4).

Over the past few years, a growing number of biologically validated mRNA targets have been identified for specific microRNAs. In plants, many of the mRNA targets appear to encode transcription factors; the microRNA often shows perfect or near-perfect “long seed” complementarity with its target, and microRNAs bind within the protein coding region as well as in untranscribed regions (5, 6). In *C. elegans* and *Drosophila*, the situation appears to be quite different: the 5’-end of the microRNA generally exhibits a short, perfect “seed” of 6-8 nucleotides in length, excluding G:U matches, which appears to be targeted to the 3’-UTR of the mRNA (2-4).

A number of studies have used biologically validated mRNA targets in *C. elegans* and *Drosophila* as a training set for machine learning algorithms, to predict additional mRNA targets that follow the same rules in a variety of species, including mammals (7-15). Most of these studies focused attention on 3’-UTR regions that are well conserved across species, in order to minimize the chance of false-positive predictions. Indeed,

based on these rules, each microRNA appears to have ~10-100 plausible mRNA targets. Studies of microRNA-mRNA binding have confirmed the importance of short-seeds having a 5'-end location within certain microRNAs (16).

However, it does not appear that all mammalian microRNAs interact with their targets via short-seeds. Xie et al. (17) identified a large number of 8-mer motifs within 3'-UTRs that were conserved across multiple mammalian species, which correspond to short-seed mRNA target regions. Only about half of currently identified mammalian microRNAs mapped to one of these motifs, suggesting that the remaining microRNAs may obey different types of binding interactions with their targets.

In a recent study, we asked if mammalian microRNAs may exhibit long-seed complementarity interactions with their mRNA targets (18). Although no biologically validated mammalian microRNA-mRNA pairs were known at the time, we took the statistical approach of looking at the entire set of potential binding interactions among all known microRNAs and all mRNAs in the human RefSeq database. To define the level of binding interactions that would be expected by chance, each microRNA sequence was scrambled 10 times [either scrambled by random permutation or by maintaining the dinucleotide composition; both methods gave similar results] and each of the scrambled microRNAs was tested. Our underlying assumption is that scrambled sequences will hit mRNA at random and define the “noise” level in any given situation, whereas microRNA sequences will hit the same number of “noise” interactions plus any true targets.

## **2. Methods**

### **2.1. Retrieving the microRNA and RefSeq mRNA sequences.**

The set of mature miRNA sequences are available in FASTA format in a plain text file from The miRNA Registry (<http://www.sanger.ac.uk/Software/Rfam/mirna/>). This file contains all miRNA sequences from all species, where the human ones are identified by lines starting with ">hsa". The most recent set of human mRNA sequences in the NCBI RefSeq database (<http://www.ncbi.nlm.nih.gov/RefSeq/>) can be retrieved by submitting the query "srcdb\_refseq[prop] AND biomol\_mrna[prop] AND homo sapiens[orgn]" to Entrez Nucleotide (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>) database, and saving the results to a local file. The RefSeq database is updated quite often, so one is not likely to retrieve the same set of records for the same query performed twice e.g. a month apart. Some records are deleted, some are temporarily removed while being annotated, and some are replaced by newer versions (even though the actual sequence may not be changed). However, the status of any given record can be found by searching on the GI or accession numbers.

### **2.2. Preprocessing.**

Many miRNA sequences are similar (some even differ only in 1 or 2 nucleotides), a fact that could potentially affect the statistics in this study. To remove redundant miRNAs, we compared all miRNA sequences pairwise, removing the shorter miRNA from consideration if they overlapped with 10 or more nucleotides. Since we only count exact hits of length 10 or more, no hits on the same mRNA at the same location will be counted more than once.

To create a population of “negative control” sequences to represent what is expected by chance, we created 10 scrambled versions of each miRNA sequence, by randomly permuting its nucleotides. This maintains the base composition but not the order. Alternatively, we also scrambled the sequences while preserving the dinucleotide composition by first splitting each miRNA sequence into chunks of two nucleotides analogous to a reading frame: once starting with the first nucleotide and once starting with the second nucleotide. We then permuted these chunks in creating a second set of scrambled miRNA sequences.

Note that RefSeq mRNA sequences use Ts whereas the miRNA Registry files use Us. The regular expression `$sequence =~ s/T/U/g` substitutes each occurrence of a T with a U, which is necessary to be able to match the two sequences (fig. 1).

### **2.3. Aligning miRNAs with mRNA sequences.**

Because the standard BLAST server at NCBI is unable to find imperfect alignments of the nature studied here (potentially with large gaps and multiple mismatches per matched nucleotide) (19), we wrote a Perl script to perform a custom gapped BLAST alignment. The procedure comprises two main steps:

#### **2.3.1. Identifying exact hits.**

Since we are looking for sequence complementarity (where miRNAs bind to mRNAs), each miRNA sequence (and their scrambled counterparts) are first reversed and

them complemented (i.e., A->U, C->G, U->A, G->C). Complementation can be accomplished with the regular expression `$sequence =~ tr/ACUG/UGAC/g`. For each pair of miRNA and mRNA sequences, take all subsequences of the miRNA in decreasing length starting with the entire sequence and going down to 10 nucleotides. Search for this subsequence within the mRNA sequence (see Perl code in fig. 1). If an exact match is found, then the hit site on the mRNA is replaced with Xs so that only the longest hit by the miRNA at this site will be counted. The hits are then recorded in a “hit file” identifying the two sequences, the length of the hit, and the location on each of the two sequences.

Often several different RefSeq records represent close variants of the same mRNA. To prevent this redundancy effect from inflating the statistics on exact hits, we only count the longest mRNA and do not count hits upon other mRNAs whose sequences matched exactly on the 25 nucleotides on each side of the hit. This will remove hits that have long stretches of exactly the same sequences, while preserving the hits that only share the hit region and not the flanks. We also remove hits having low complexity sequences by using the regular expression `$sequence =~ s/((.+)\2{4,})/N' x length $1/eg` (Lincoln Stein, Bioperl: Repetitive DNA, <http://bioperl.org/pipermail/bioperl-l/1999-November/003313.html>).

### **2.3.2. Extending the exact hits.**

Given an exact hit of 10 or more nucleotides, with no gaps, mismatches or G:U matches, the seed region is extended in both directions to find the optimal alignment allowing for large gaps and multiple mismatches. This is achieved by penalizing gaps at

only a fraction of the reward of a match, and making the mismatch penalty a fraction of the reward of a match. Note that the BLAST programs from NCBI tends to not pick up good alignments in this scenario, probably due to the method of calculating significance (i.e., "Expect Values"). We allow a ratio of up to 4:1 between mismatches and matches, and allow a four nucleotide gap between matches, while allowing a longer gap as long as multiple matches are obtained as a result. This is achieved with the parameters: Match reward,  $r = 10$ , mismatch penalty,  $q = -2.5$ , Open gap penalty  $G = 8$ , Extend gap penalty  $E = 0.5$ .

Given a pair of sequences  $\mathbf{x}$  and  $\mathbf{y}$  of lengths  $m$  and  $n$ , respectively, let  $x_i, y_j$  denote their  $i$ -th and  $j$ -th nucleotide, respectively. Then, the alignment score is calculated as follows:

Start by assigning  $\text{Score}(i,j) = 0$  for all  $(i,j)$  where  $i$  or  $j = 0$ . Then, for  $i = 1, \dots, m$  and  $j = 1, \dots, n$  compute the score using the following equation:

$$\text{Score}(i, j) = \max \begin{cases} \text{Score}(i-1, j-1) + (r - q) \times \text{Match}(x_i, y_j) - q \\ \text{Score}(k, j) - G - E \times (i - k), k = 0, \dots, i-1 \\ \text{Score}(i, k) - G - E \times (j - k), k = 0, \dots, j-1 \end{cases}$$

The greatest score is then given by  $\text{Score}(m,n)$ . If the actual alignment is to be found one needs to keep track of the path from  $(0,0)$  to  $(m,n)$  in the Score array as well.

For each record identified in the hit file, a sequence alignment is performed between

the right flank of the miRNA vs the right flank of the mRNA using the recursive equation above, ensuring that the first nucleotides are either gap or mismatch scored. Then, the same procedure is performed on the left flanks, after which the total score is tallied and recorded in a file. When allowing for G:U matches, nucleotide C in the microRNA reverse-complement sequence is replaced by C OR U, and nucleotide A is replaced by A OR G (fig. 1) indicating that multiple combinations should receive a match reward.

### **3. An analysis of long-seed interactions between mammalian microRNAs and their targets.**

As shown in fig. 2, a large number of long-seed complementary interactions existed between microRNA sequences and mRNA sequences having 10 or more exact matches (excluding G:U matches), which differed significantly from the level expected by chance. Strikingly, at longer and longer seed lengths, the difference between the microRNA set and the scrambled set became more and more pronounced: at a seed length of 16 or greater, there were 3.5 times more examples of potential target interactions in the microRNA set than in the scrambled set, and at a seed length of 17 or greater, the ratio was over 7 to 1 (fig. 2).

This finding suggested that some true microRNA-mRNA target interactions may involve long-seeds. To further define a list of candidate mRNA targets with high confidence, we analyzed further the set of interactions having 10 or more matches in a row (the “10+ set”) to see if they would also exhibit other statistically significant differences at the population level. Indeed, we defined a gapped BLAST score for



microRNA-mRNA interactions taking into account gaps, mismatches and G:U matches, and found that the gapped BLAST scores in the 10+ set were significantly better overall than expected by chance (fig. 3). Furthermore, for those mRNAs that received multiple “hits” from different microRNAs within the 10+ set, we examined the minimum distance between hits. Again, there was a striking difference between the minimum distance seen in the microRNA population vs. the population of scrambled sequences (fig 4).

By combining these three criteria (seed length, gapped BLAST score, and minimum distance between hits) we created a list of 71 candidate mRNA targets which were unlikely to be chosen by chance – with a discrimination ratio of 5.2 to 1, which means that over 80% of the targets on the list are expected to be biologically valid targets. The 71 mRNAs had a larger number of microRNA hits per kilobase of target sequence than did the scrambled sequences. As well, individual microRNAs hit multiple (up to 17) distinct members of the candidate set, which again happened significantly more often than by chance (fig. 5). These findings indicate that the outlier mRNAs are different as a whole from the mRNAs that were hit by scrambled counterparts, even those that satisfied the same cut-off criteria.

Our candidate target list contained very similar types of targets as predicted by the short-seed computational studies, including members of the same gene families:

Transcription factors (including homeobox genes), nucleic acid-binding proteins and many other functional categories including kinases, receptors and other signal transduction proteins, membrane and cytoskeletal proteins, and effectors of differentiation.

Also, our study was in concordance with other studies showing that some individual microRNAs may hit multiple mRNA targets residing in the same metabolic pathway (18). However, surprisingly, we found that the long-seed candidate mRNA target list had no preference at all for microRNA hits to be located within 3'-untranslated regions, but hit within the protein coding region in ~2/3 of cases. As well, the best microRNA hits upon candidate mRNA targets did not have relatively better target complementarity near their 5'-end. These considerations suggest that short-seed and long-seed target interactions both exist in mammals, and that they may follow different rules.

#### **4. “Perfect” target interactions (allowing G:U matches).**

A few examples of microRNAs exhibiting perfect complementarity have been described (miR-196 (20), miR-127 and miR-136 (21, 22)). During our analysis of long-seed interactions we were struck by the existence of targets that had perfect complementarity to microRNAs along their full length (ref. 18, fig. 2), and we wondered whether they could be representatives of a larger, distinct class of targets.

Whereas long-seeds are defined as having exact Watson-Crick base pairing with no G:U matches, recent suggest that complementarity interactions that contain up to, say, 5-7 G:U matches (but no frank mismatches) could still be deemed “perfect” in gene silencing (23). In particular, we noted that human miR-95 was perfectly complementary (including up to 4 G:U matches) with scores of human mRNAs and ESTs (fig. 6). Similarly, miR-151\* was perfectly complementary to 6 transcripts, which was significantly above the level expected by chance.

The reason for these perfect hits turned out to be simple and intriguing: The precursors of these two microRNAs, as well as two others (miR-28 and miR-325), turned out to derive entirely from genomic repeats (24). In particular, the hairpin foldbacks were formed by the junction of two adjacent LINE-2 repeat segments apposed in opposite orientation (fig. 7). Insofar as MIR repeats and other LINE-2 derived elements are present in the 3'-UTR of many different mRNAs and EST transcripts, when in the proper orientation they are natural generic targets for the repeat derived microRNAs – in some cases having perfect complementarity, and in other cases (since there is divergence among LINE-2 elements) imperfect complementarity (24).

## **5. Discussion**

Computer and genomic analyses of mammalian microRNA-mRNA target interactions suggest that at least three distinctly different types of interactions exist: short-seeds, long-seeds and “perfect” hits (allowing G:U matches). By their nature, one would expect that short-seed interactions are likely to have weak effects, to inhibit protein translation, and to affect large numbers of genes in parallel, whereas long-seed interactions might be expected to have stronger effects and involve fewer targets. The LINE-2 repeat derived microRNAs appear to recognize transcripts that share repeats in their 3'-UTR regions, and when they bind with perfect or near-perfect complementarity, they may be expected to lead to degradation of the transcripts. This could serve as a mechanism for detecting and neutralizing aberrant transcripts (having readthrough

transcription from retained introns or neighboring genomic regions) as well as serving to regulate specific mRNAs (24).

It should be cautioned that complementarity between microRNAs and their targets is not the only factor that may govern which microRNA-mRNA target interactions are effective in vivo. One must consider the potential importance of mRNA target secondary structure (25) as well as the strong possibility that RNA-binding proteins may participate in microRNA recognition (26). As well, both microRNA and mRNA need to be co-expressed in proper amounts within the same cell for effective interaction to occur, and A-to-I editing of RNA might abrogate potential mRNA targets from being effectively silenced by the RISC complex (27).

Finally, it is likely that new types of microRNA-mRNA target interactions still remain to be uncovered. A number of viruses have been shown to encode microRNA precursors (28-31) that are not conserved in their host genomes or across different viruses, and their targets have not been fully defined. As well, a population of small RNAs that shares some, but not all, of the features of microRNAs has been described in *C. elegans* (32); these so-called tncRNAs are not conserved across species, and their targets (if any) have not been identified yet. Therefore, new types of target interactions, and indeed new types of small RNAs, may be identified in mammalian species within the near future.

## **6. Acknowledgments**

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.

## 7. References

1. Dykxhoorn, D.M., Novina, C.D. and Sharp, P.A. (2003) Killing the messenger: short RNAs that silence gene expression. *Nat. Rev. Mol. Cell Biol.* **4**, 457-467.
2. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297.
3. Ambros, V. (2004) The functions of animal microRNAs. *Nature* **431**, 350-355.
4. Lai, E.C. (2003) microRNAs: runts of the genome assert themselves. *Curr. Biol.* **13**, R925-936.
5. Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**, 2053-2056.
6. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell* **110**, 513-520.
7. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.* **5**, R1.
8. Stark, A., Brennecke, J., Russell, R.B. and Cohen, S.M. (2003) Identification of Drosophila MicroRNA Targets. *PLOS Biol.* **1**, 397-409.
9. Rajewsky, N. and Socci, N.D. (2004) Computational identification of microRNA targets. *Dev. Biol.* **267**, 529-535.
10. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell* **115**, 787-798.

11. Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* **18**, 1165-1178.
12. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. *PLoS Biol.* **2**, e363.
13. Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507-1517.
14. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20.
15. Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85.
16. Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504-511.
17. Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S. and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338-345.
18. 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.
19. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

20. Yekta, S., Shih, I.H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594-596.
21. Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C. and Cavaille, J. (2003) Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat Genet.* **34**, 261-262.
22. Smalheiser, N.R. (2003) EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome Biol.* **4**, 403.
23. Miyagishi, M., Sumimoto, H., Miyoshi, H., Kawakami, Y. and Taira, K. (2004) Optimization of an siRNA-expression system with an improved hairpin and its significant suppressive effects in mammalian cells. *J. Gene Med.* **6**, 715-723.
24. Smalheiser, N.R. and Torvik, V.I. (2005) Mammalian microRNAs derived from genomic repeats. *Trends Genet.*, in press.
25. Bohula, E.A., Salisbury, A.J., Sohail, M., Playford, M.P., Riedemann, J., Southern, E.M. and Macaulay, V.M. (2003) The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript. *J. Biol. Chem.* **278**, 15991-15997.
26. Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S.C., Gram, H. and Han, J. (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* **120**, 623-634.
27. Scadden, A.D. and Smith, C.W. (2001) RNAi is antagonized by A-->I hyper-editing. *EMBO Rep.* **2**, 1107-1111.



28. Pfeffer, S., Zavolan, M., Grassler, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C. and Tuschl, T. (2004) Identification of virus-encoded microRNAs. *Science* **304**, 734-736.
29. Bennasser, Y., Le, S.Y., Yeung, M.L. and Jeang, K.T. (2004) HIV-1 encoded candidate micro-RNAs and their cellular targets. *Retrovirology* **1**, 43.
30. Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grassler, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., Russo, J.J., Ju, J., Randall, G., Lindenbach, B.D., Rice, C.M., Simon, V., Ho, D.D., Zavolan, M. and Tuschl, T. (2005) Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**, 269-276.
31. Cai, X., Lu, S., Zhang, Z., Gonzalez, C.M., Damania, B. and Cullen, B.R. (2005) Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. *Proc. Natl. Acad. Sci. U S A* **102**, 5570-5575.
32. Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T. and Jewell, D. (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* **13**, 807-818.

## Figure Legends

**1. Perl script to match a subsequence of a miRNA with a mRNA.** See Methods for details.

**2. microRNAs and their scrambled counterparts interact differently with the population of human mRNAs.** Shown are all exact hits  $\geq 10$  bases long (not counting G:U matches) produced on human RefSeq mRNAs by the set of nonredundant microRNAs, vs. the average of 10 replications of scrambled control sequences. Shown is the number of hits as a function of exact hit length. Only the longest hit was counted: e.g., for a hit of length 18, the two subsets of length 17 in the same hit position were not counted. Figures 2-5 are reproduced from BMC Bioinformatics **5**, 139.

**3. Distribution of gapped BLAST scores in hits made by microRNAs and scrambled counterparts.** Without permitting G:U matches in the extension phase, the microRNAs had better average gapped BLAST scores than scrambled counterparts across all mRNAs in the "10+ set" ( $153.00 \pm 0.03$  vs.  $150.98 \pm 0.01$ , mean  $\pm$  s.e.m.,  $p < 0.0001$ ). With permitting G:U matches in the extension phase, the microRNA set showed significantly fewer G:U matches overall relative to scrambled counterparts, even when holding constant the length of the exact hit ( $2.891 \pm 0.004$  vs.  $2.939 \pm 0.001$ ,  $p < 0.0001$ ).

**4. Number of distinct mRNA sequences which received hits from two or more distinct microRNAs, as a function of the minimum distance between hits.** Distance of

0 or 1 was excluded because this might be produced by partial overlap of microRNA sequences.

**5. Individual microRNAs hit multiple targets on the candidate list, more often than expected by chance.**

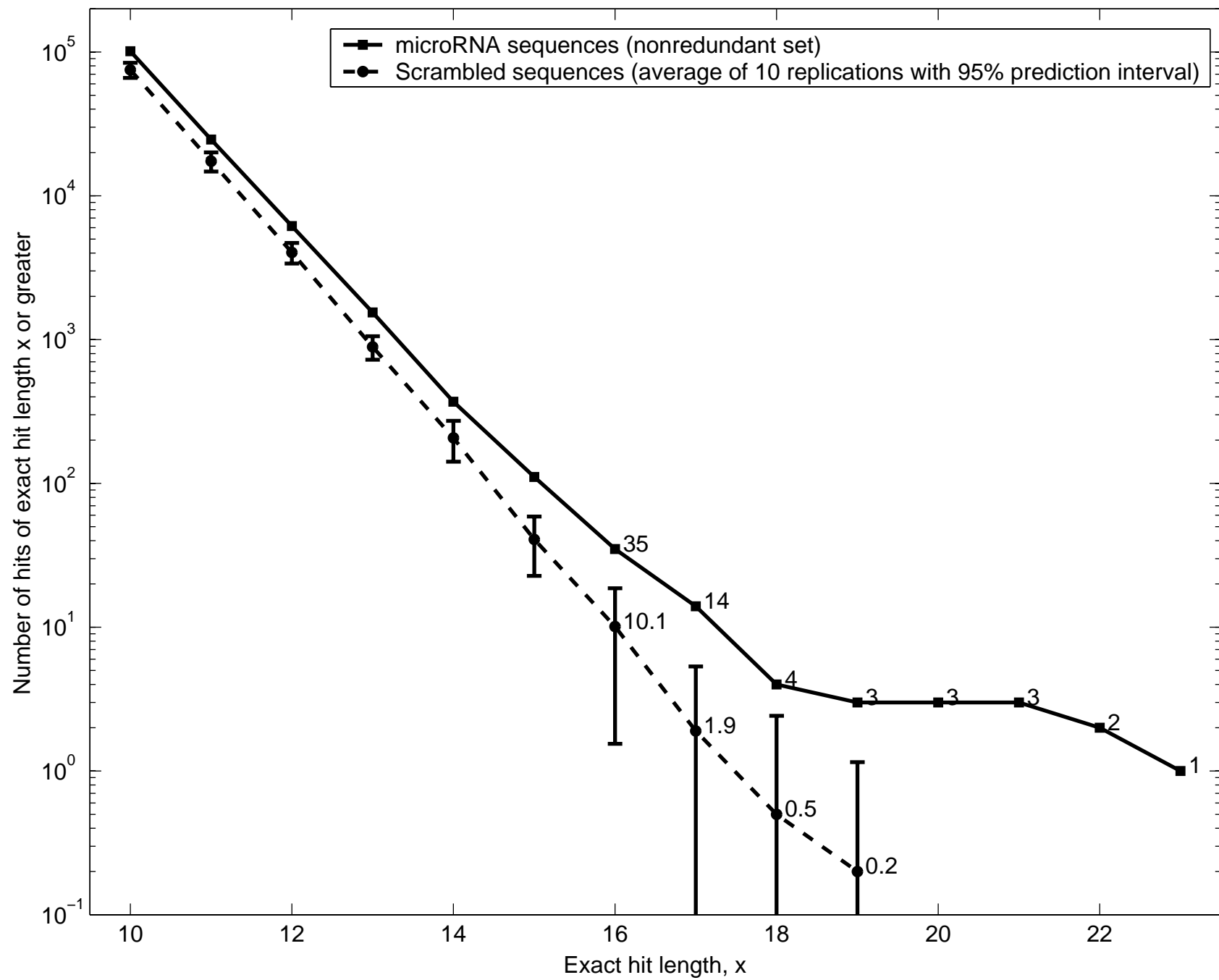
**6. Multiple sequence alignment of mRNAs and ESTs that exhibited perfect complementarity to miR-95.** miR-95 hit perfectly (including two-to-four G:U matches) on three mRNAs and 94 ESTs comprising 31 distinct clusters (Unigene clusters or singletons when not belonging to a Unigene cluster). Multiple sequence alignment performed using ClustalW (<http://www.ebi.ac.uk/clustalw>) shows that these putative targets are not related to each other except in the microRNA hit region and in nearby LINE-2 homologous sequences (L2A consensus from Repbase, <http://www.girinst.org>). Figures 6 and 7 are reproduced with permission from Trends in Genetics.

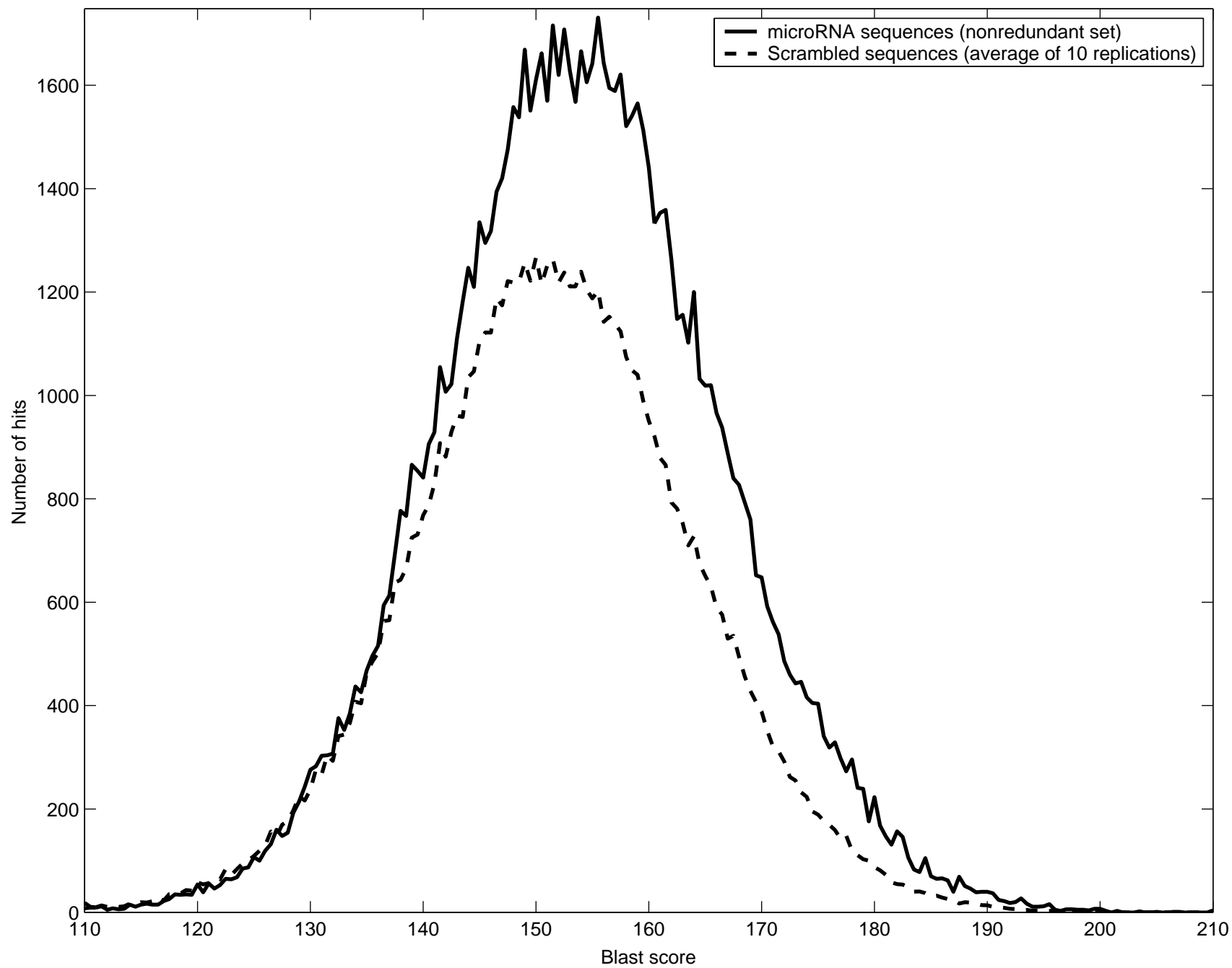
**7. Genomic structure of human LINE-2-derived miRNA precursors.** Information was downloaded and edited from the UCSC Genome Browser. Each of the precursors resides within an intron, and each flanks the junction of two L2 repeats in opposite orientation (darker shading indicates less divergence from the L2 consensus sequence).

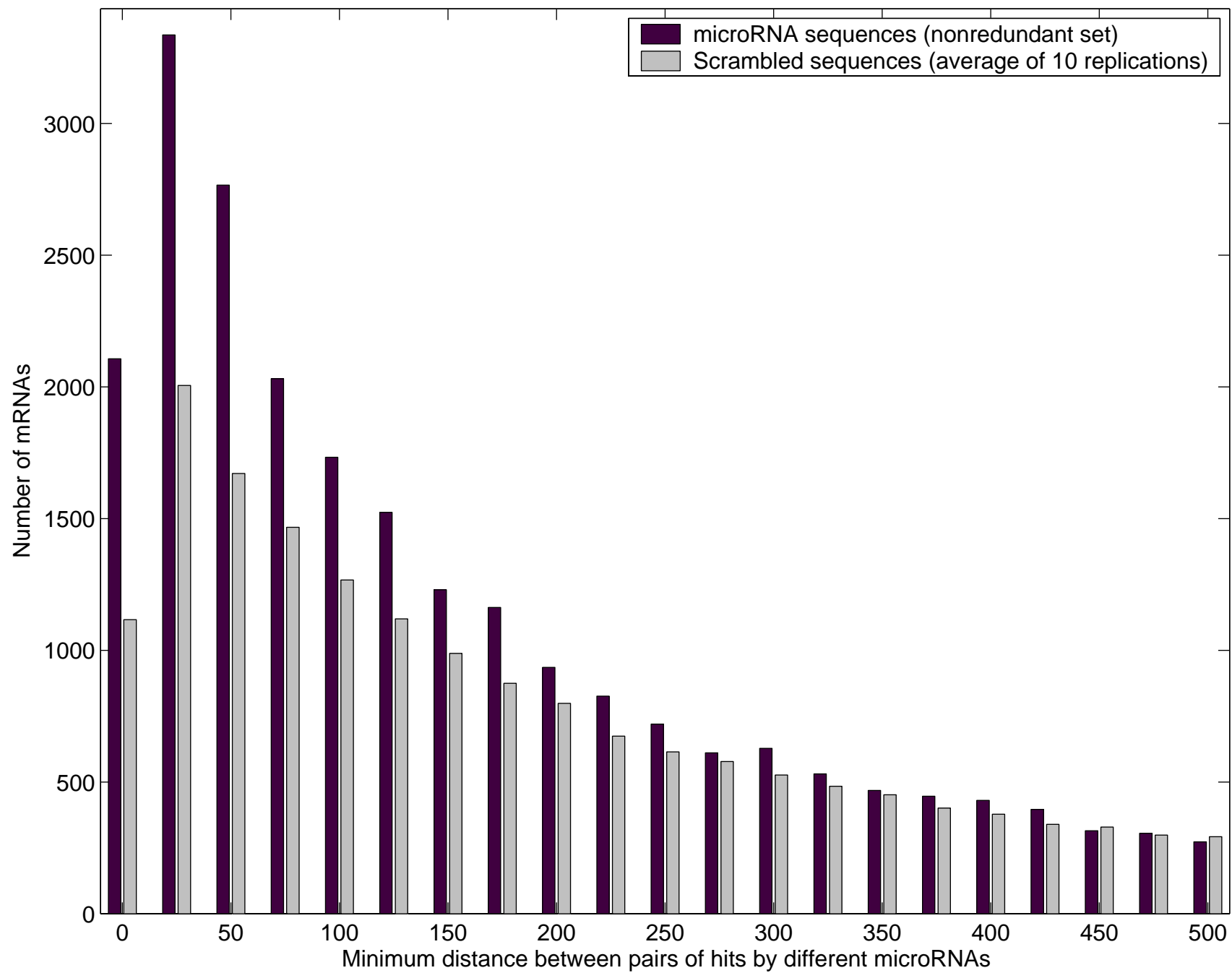
```
$mrna = 'ACTTTGGGACGAGCTT';
$mrna =~ s/T/U/g;    #replace Ts with Us
$mirna = 'UUGG';
$mirna = reverse $mirna; #reverse the sequence
$mirna =~ tr/ACUG/UGAC/; #find the complement

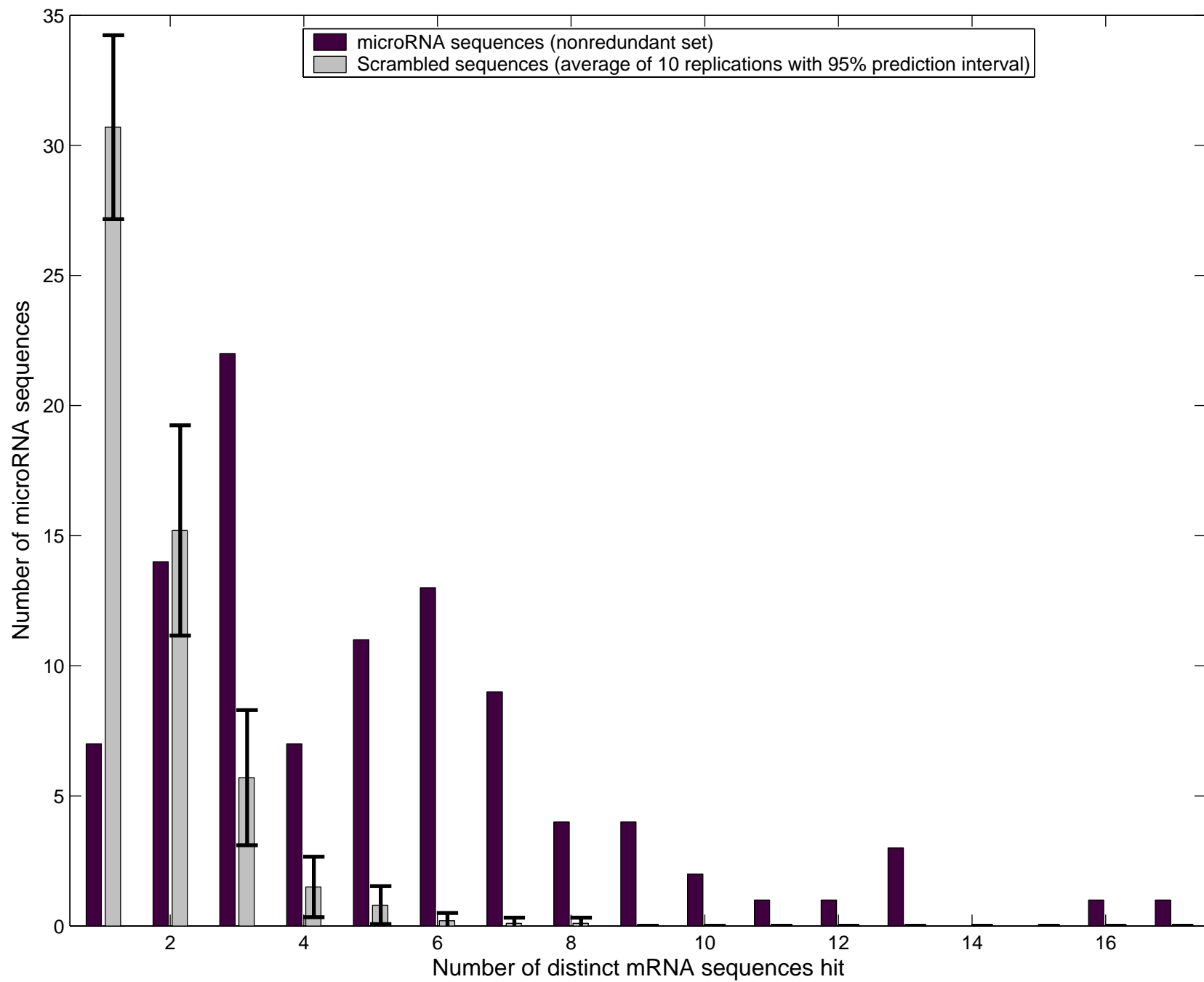
if ($GUs_allowed) {
    $mirna =~ s/C/[CU]/g;    #Gs will now match either Cs or Us
    $mirna =~ s/A/[AG]/g;    #Us will now match either As or Gs
}

while ($mrna =~ /$mirna/g) {
    $hit_length = length($mirna);
    $hit_pos = pos($mrna) - $hit_length;
    substr($mrna,$hit_pos,$hit_length) = 'X' x $hit_length;
}
```



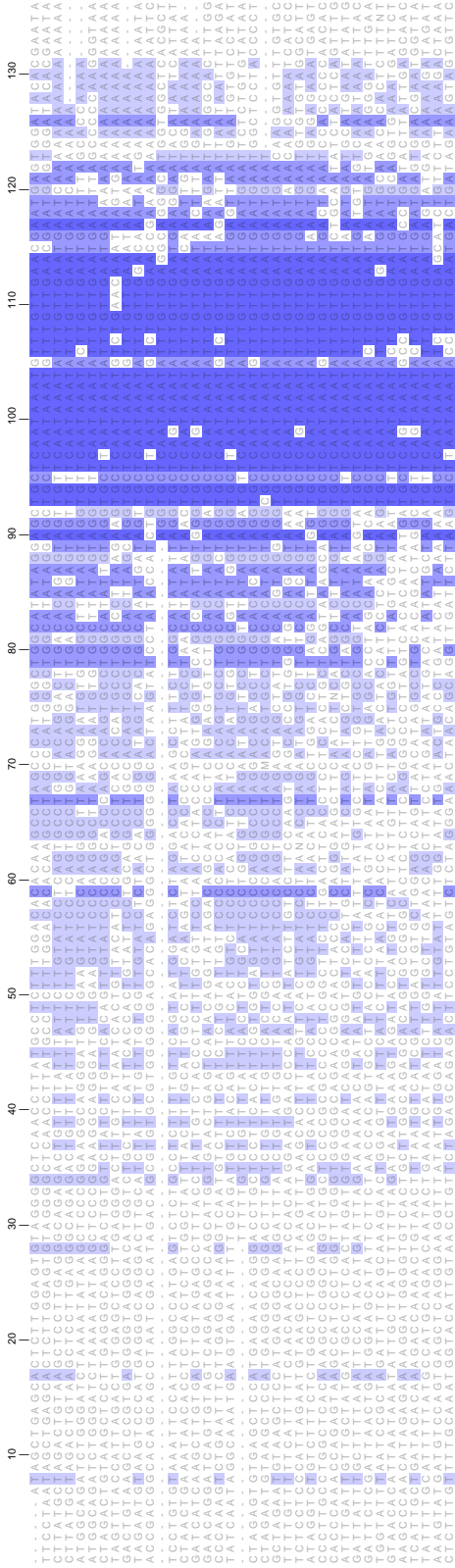




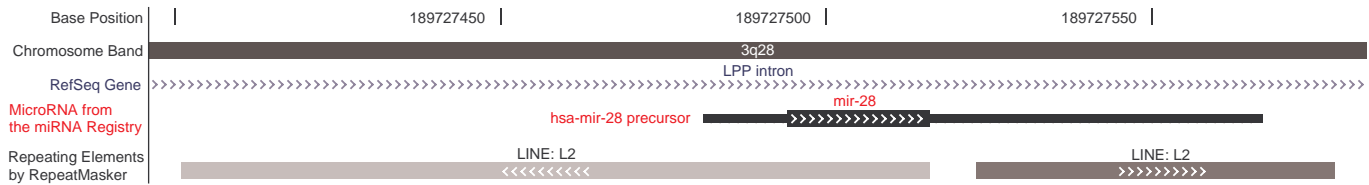




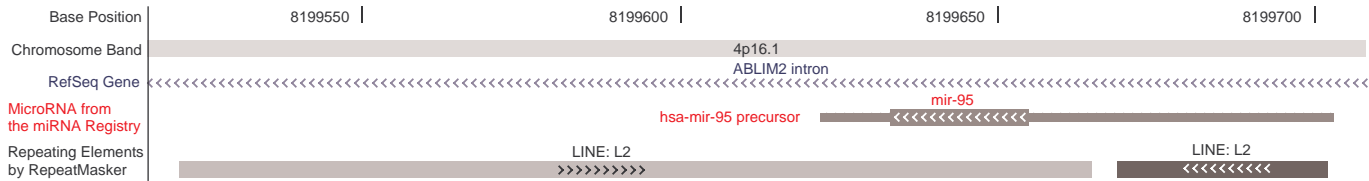
GI119060  
LOC284531  
GI2778430  
GI820483  
GI8678515  
SPAG11  
GI1204702  
GI820145E2f6  
GI23679893  
GI1190602  
GI1395555  
GI168980  
GI168982  
GI203552  
GI278661  
GI10340744  
GI10340745  
GI1899107  
GI1990638  
GI1857398  
GI1857399  
GI1224665  
GI77218  
GI892402  
GI1599167  
GI1548472  
GI1599167  
GI2857350  
GI84321815Nbr03515  
GI1083281  
GI1083282  
GI1106952  
GI16337390  
GI2747935KJMA1998  
GI11650612IMAGE4289241



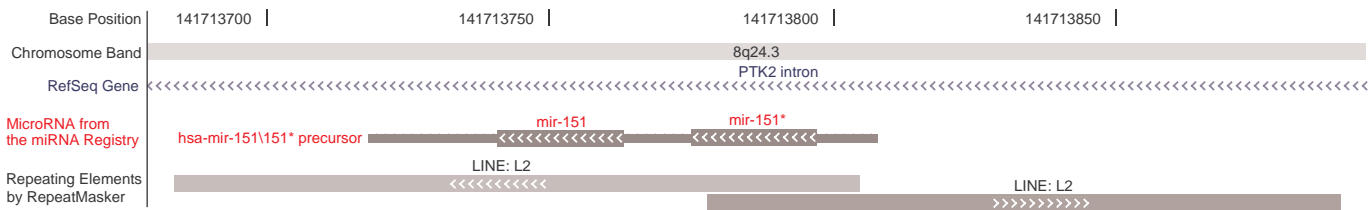
## hsa-mir-28



## hsa-mir-95



## hsa-mir-151\151\*



## hsa-mir-325

