**Table S1 Feature calculation**

The following details how to calculate the features used in this study.

(**1) Folding Energy**

It’s the minimal folding energy calculated by RNAduplex from ViennaRNA package (<http://www.tbi.univie.ac.at/RNA/>). This feature was also used by miRanda ([1](#_ENREF_1)), RNAhybrid ([2](#_ENREF_2)), miRmap([3](#_ENREF_3)), etc.

**(2) Seed match**

Whether there existed a perfect pairing between 2-7 nt of miRNA? (1/0). This feature was also by miRanda ([1](#_ENREF_1)), targetScan ([4](#_ENREF_4)), miRmap([3](#_ENREF_3)), etc.

**(3) Accessibility**

The accessibility is a measurement of whether the target site region in the mRNA sequence is open for miRNA to binding. The accessibility was proposed in ([5](#_ENREF_5)). This feature was used in many miRNA target prediction tools: RNAhybrid ([2](#_ENREF_2)), DIANA-microT web server ([6](#_ENREF_6)), miRmap([3](#_ENREF_3)), etc. The accessibility was calculated using RNAplfold with the following command as described in ([7](#_ENREF_7)).

RNAplfold –W 80 –L 50 –u 16.

RNAplfold is from the ViennaRNA package.

The same parameter was used in many other studies ([8](#_ENREF_8),[9](#_ENREF_9)).

**(4) AU content**

AU-rich elements (AREs) are found of many mRNA, it was reported as an important feature for miRNA-mRNA binding ([10](#_ENREF_10),[11](#_ENREF_11)). This feature was used in TargetScan ([4](#_ENREF_4)) as: The local AU content reflects the transcript AU content 30nt upstream and downstream of predicted site. Here, we used the same strategy.

**(5) Stem Conservation**

The Stem conservation was calculated as the Average PhyloP sore in the miRNA-mRNA binding stem region. This features has been described in ([12](#_ENREF_12)) and it was used by many studies (e.g . ([3](#_ENREF_3),[13](#_ENREF_13)))

**(6) Flanking conservation**

The Flanking conservation was calculated as the Average PhyloP sore in the 40nt upstream and 40nt downstream of the binding site. This feature was also used by ([3](#_ENREF_3),[13](#_ENREF_13),[14](#_ENREF_14))

**(7) Conservation Difference**

This feature is calculated as the difference between the Stem conservation and Flanking conservation. This feature was also used by ([13](#_ENREF_13))

**(8) m/e motif**

This feature is about the paring probabilities at different positions of miRNA. For each position of a miRNA, if it’s a pairing, we name it as m (matching), if not, we name it as e (else). Then, we learned the probability of m/e at each position of miRNAs

, where *x* is length of miRNA and *pi* is the probability of ‘m’ at position *i* of miRNA.

**(9) Total number of paired positions**

This feature is calculated as the total number of paired positions for each miRNA-mRNA binding site.

**(10) The length of target mRNA region**

This feature is calculated as the length of miRNA binding target site region. For example, if miRNA x binds to mRNA y and the binding site between x and y are 28 nts region on mRNA y, this feature is 28. Based on the distribution of the length of target mRNA region in the true CLASH miRNA target sites and negative control miRNA target site, the length of true miRNA target mRNA region is more concentrated on 20-24nt long.

**(11) The length of the largest consecutive pairs**

This feature is calculated as the length of the longest consecutive pairs.

**(12) Position of the largest consecutive pairs**

This feature is calculated the relative position of largest consecutive pairs to to 5’ end of miRNA

**(13) The length of the largest consecutive pairs allowing 2 mismatches**

This feature is calculated as the length of the longest consecutive pairs allowing 2 mismatches

(**14) The position of the largest consecutive pairs allowing 2 mismatches**

This feature is calculated the relative position of largest consecutive pairs allowing 2 mismatches to to 5’ end of miRNA

**(15)** **The number of paired positions at the miRNA 3’ end**

Here miRNA 3’ end denotes the last 7 nt of the miRNA and this feature is calculated as the number of paired positions in the miRNA 3’ end.

**(16) The total number of paired positions in the seed region and the miRNA 3’ end**

This feature is calculated as the total number of paired positions in the seed region and the miRNA 3’ end

**(17) The difference between the number of paired positions in the seed region and that in the miRNA 3’ end**

This feature is calculated as the difference # of paired positions between the seed region and the miRNA 3’ end region.

**(18) Exon preference**

This feature is showing the preference of miRNA-mRNA binding in terms of exons. Whether the miRNAs prefer to bind specific exons? This feature is calculated as the relative number of exons to nearest end (either 5’ or 3’ end). For example, if the miRNA binds to the first exon of the mRNA, the exon preference will be 0 (0-based).

1. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human microRNA targets. *PLoS Biol*, **2**, e363.

2. Krüger, J. and Rehmsmeier, M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic acids research*, **34**, W451-W454.

3. Vejnar, C.E. and Zdobnov, E.M. (2012) miRmap: Comprehensive prediction of microRNA target repression strength. *Nucleic acids research*, **40**, 11673-11683.

4. Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*, **27**, 91-105.

5. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nature genetics*, **39**, 1278-1284.

6. Maragkakis, M., Reczko, M., Simossis, V.A., Alexiou, P., Papadopoulos, G.L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E. and Kourtis, K. (2009) DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic acids research*, gkp292.

7. Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J. and Hofacker, I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nature biotechnology*, **26**, 578-583.

8. Marín, R.M. and Vaníček, J. (2011) Efficient use of accessibility in microRNA target prediction. *Nucleic acids research*, **39**, 19-29.

9. Li, J., Kim, T., Nutiu, R., Ray, D., Hughes, T.R. and Zhang, Z. (2014) Identifying mRNA sequence elements for target recognition by human Argonaute proteins. *Genome research*, **24**, 775-785.

10. Chen, C.-Y.A. and Shyu, A.-B. (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends in biochemical sciences*, **20**, 465-470.

11. 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.

12. Sætrom, P., Snøve Jr, O., Nedland, M., Grünfeld, T.B., Lin, Y., Bass, M.B. and Canon, J.R. (2006) Conserved microRNA characteristics in mammals. *Oligonucleotides*, **16**, 115-144.

13. Helwak, A., Kudla, G., Dudnakova, T. and Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, **153**, 654-665.

14. Ohler, U., Yekta, S., Lim, L.P., Bartel, D.P. and Burge, C.B. (2004) Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *Rna*, **10**, 1309-1322.