miTAR: an interpretable deep learning-based approach for predicting miRNA targets

UF | ICBR
your pursuit is what drives us.
BIOTECH.UFL.EDU

Tongjun Gu[1,2], Mingyi Xie[3,4,5], Xiwu Zhao[8], William Bradley Barbazuk[1,5,6], Ji-Hyun Lee[2,7]

[1]Bioinformatics, Interdisciplinary Center for Biotechnology Research, [2]Department of Biostatistics, [3]University of Florida Health Cancer Center, [4]Department of Biochemistry and Molecular Biology, [5]Department of Biology, [6]Genetics Institute, [7]Division of Quantitative Sciences,, University of Florida; [8]Department of Ophthalmology & Visual Sciences, University of Michigan

Abstract

Motivation: MicroRNAs (miRNAs) modulate a broad range of essential cellular processes linked to human health and diseases. Many computational methods have been developed to predict targets of miRNAs. However, the majority of these methods depend on pre-defined features that require considerable efforts and resources and often prove suboptimal at predicting miRNA targets.

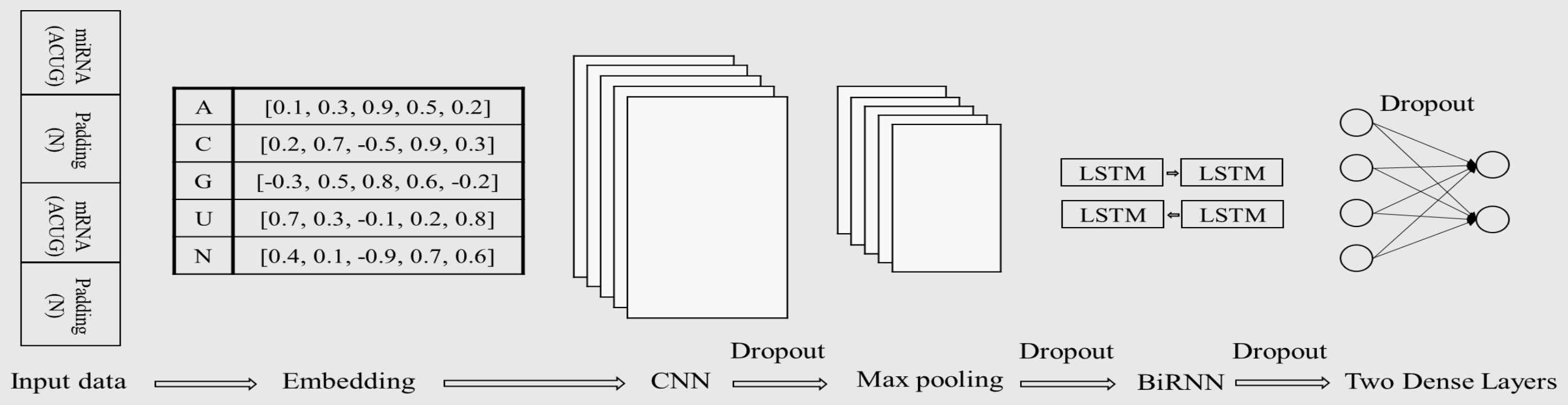
Methods: We developed a novel hybrid deep learning-based approach that integrates two major types of deep learning methods: **convolutional neural networks (CNNs) and recurrent neural networks (RNNs)**. CNNs excel in learning spatial features and RNNs discern sequential features. Therefore, our approach has the advantages of learning both the intrinsic spatial and sequential features of miRNA:target. The source code is at https://github.com/tjgu/miTAR.

Advantages: 1) The inputs are sequences of miRNAs and RNAs/genes, Thus, our approach is easy to use. 2) Our approach gains substantial better performance than other alternatives based on the evaluation metrics on test and independent datasets.

Interpretation methods: a series of in-silico mutagenesis analyses, accumulative mutagenesis analyses, joint and independent mutagenesis analyses, correlation analyses, base-pairing analyses, and layer analyses.
Interpretation results: 1) miTAR captures known features in miRNA:target, including the involvement of seed region and the free energy, as well as multiple novel features. 2) The CNN and RNN layers of the models behave differently at capturing the free energy feature. 3) The Max pooling layer in between the CNN and RNN improves the performance of all our models, especially for smaller datasets.

Significance: Our discoveries will enhance the understanding of the mechanisms in miRNA:target interactions and miRNA associated diseases.

1. Overview of the deep learning-based approach



2. Performance evaluation of two models using test datasets

Model: dataset	Accuracy	Sensitivity	Specificity	F-Score	Brier Score
DeepMirTar: DeepMirTarRaw[a]	0.9348	NA	NA	0.9348	NA
miTAR1: DeepMirTar (30 times)[b]	0.9787	0.9717	0.9857	0.9786	0.0193
miRAW: miRAWRaw[a]	0.935	0.935	0.938	0.935	NA
miTAR2: miRAW (30 times)[b]	0.9649	0.9616	0.9683	0.9651	0.0271

[a] known datasets and models; [b] average values; miTAR1 and miTAR2 trained on DeepMirTar and miRAW dataset separately.

3. Performance evaluation of two models using independent test datasets

Model: dataset	Identified positive miRNA:targets /Total number of positive pairs	Identified negative miRNA:targets/Total number of negative pairs
DeepMirTar: DeepMirTar Independent test set	24/48	NA
miTAR1: DeepMirTarIn	36/48	47/48

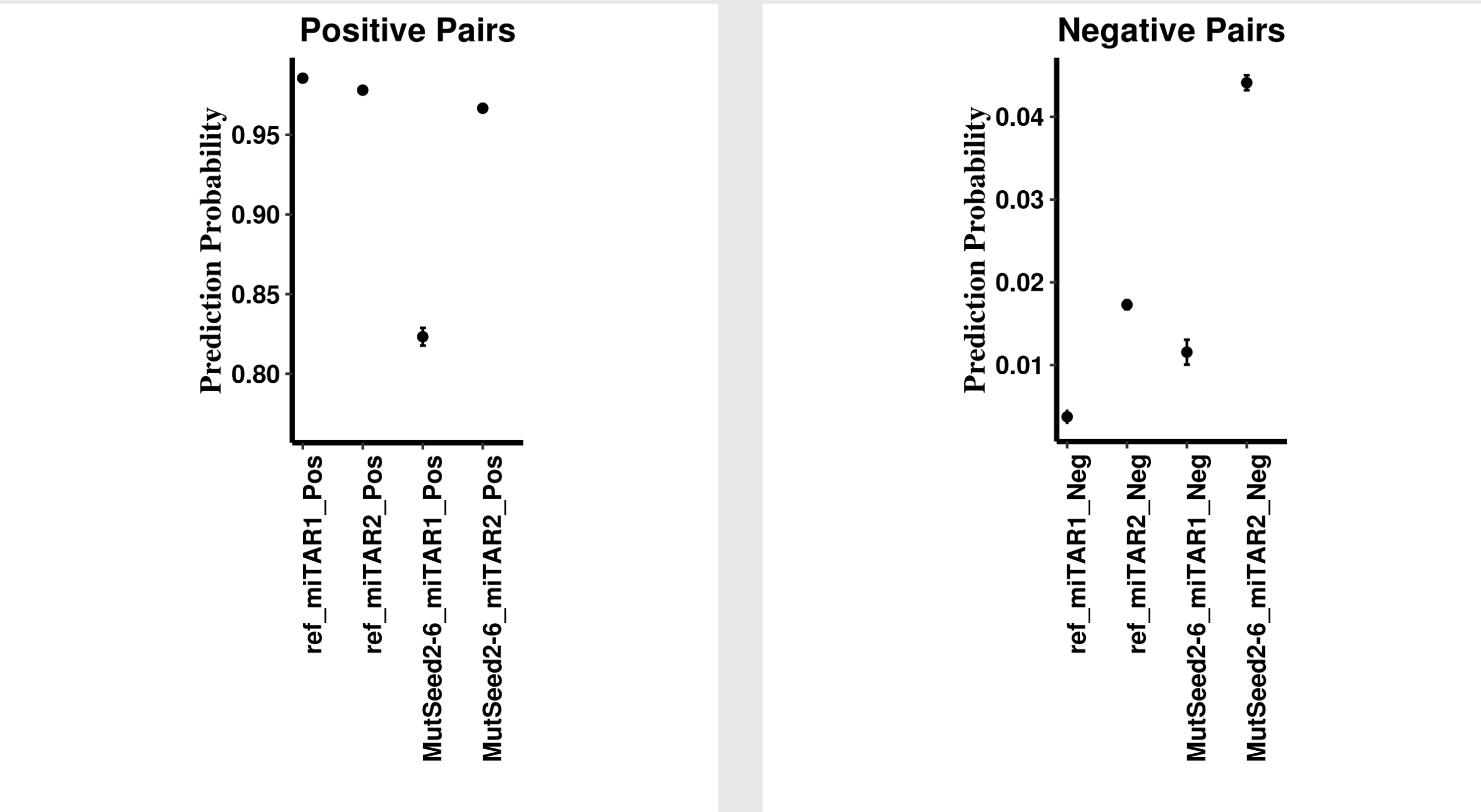
Model:dataset	Accuracy	Sensitivity	Specificity	F-Score	Brier Score
miRAW: miRAWRaw	0.913	0.931	0.363	0.954	NA
miTAR2: miRAWIn	0.970	0.960	0.979	0.969	0.025

4. Performance evaluation for the unified model, miTAR.

Model: dataset	Accuracy	Sensitivity	Specificity	F-Score	Brier Score
miTAR: MirTarRAW[c] Test set	0.9627	0.9591	0.9663	0.9627	0.0321
miTAR: MirTarRAW[c] (30 times)	0.9545 [0.9531-0.9559]	0.9532 [0.9513-0.9551]	0.9557 [0.9533-0.9582]	0.9544 [0.9530-0.9558]	0.0401 [0.0390-0.0412]

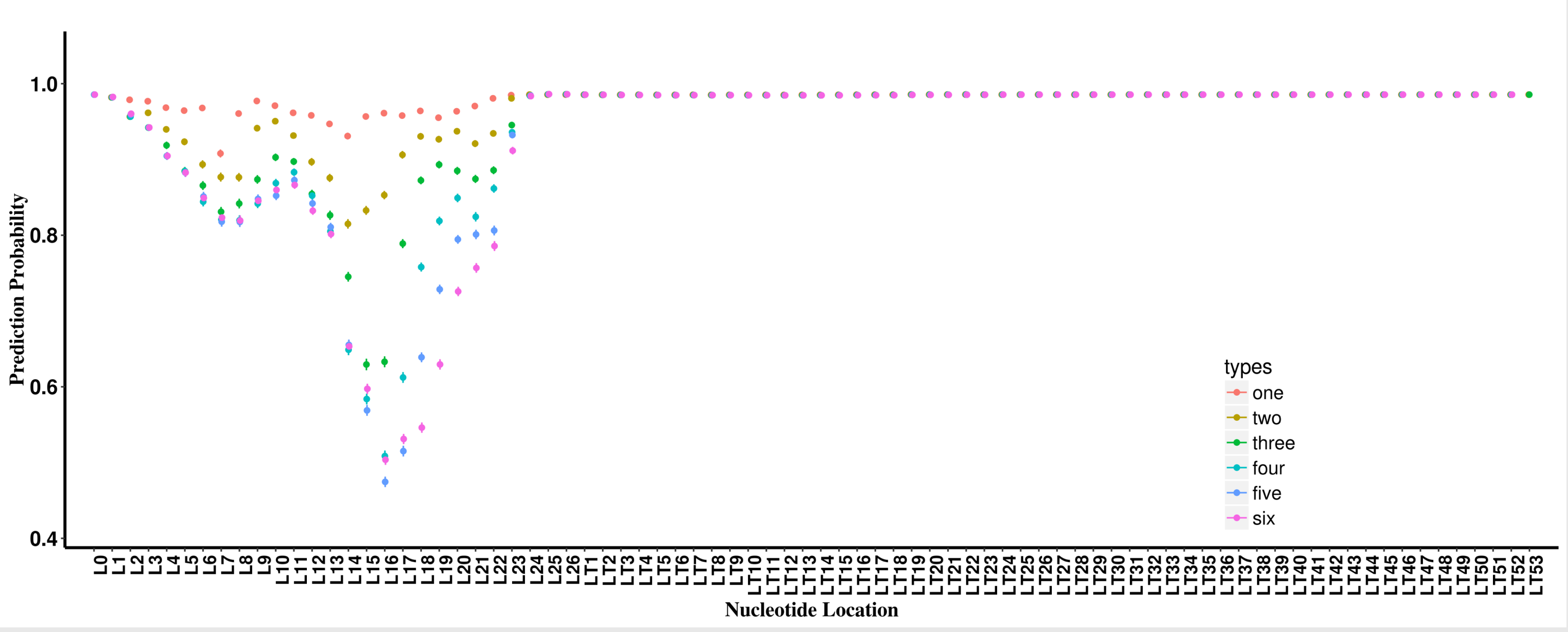
[c]combination of DeepMirTar and miRAW datasets.

5. The impact of the alteration at the seed region

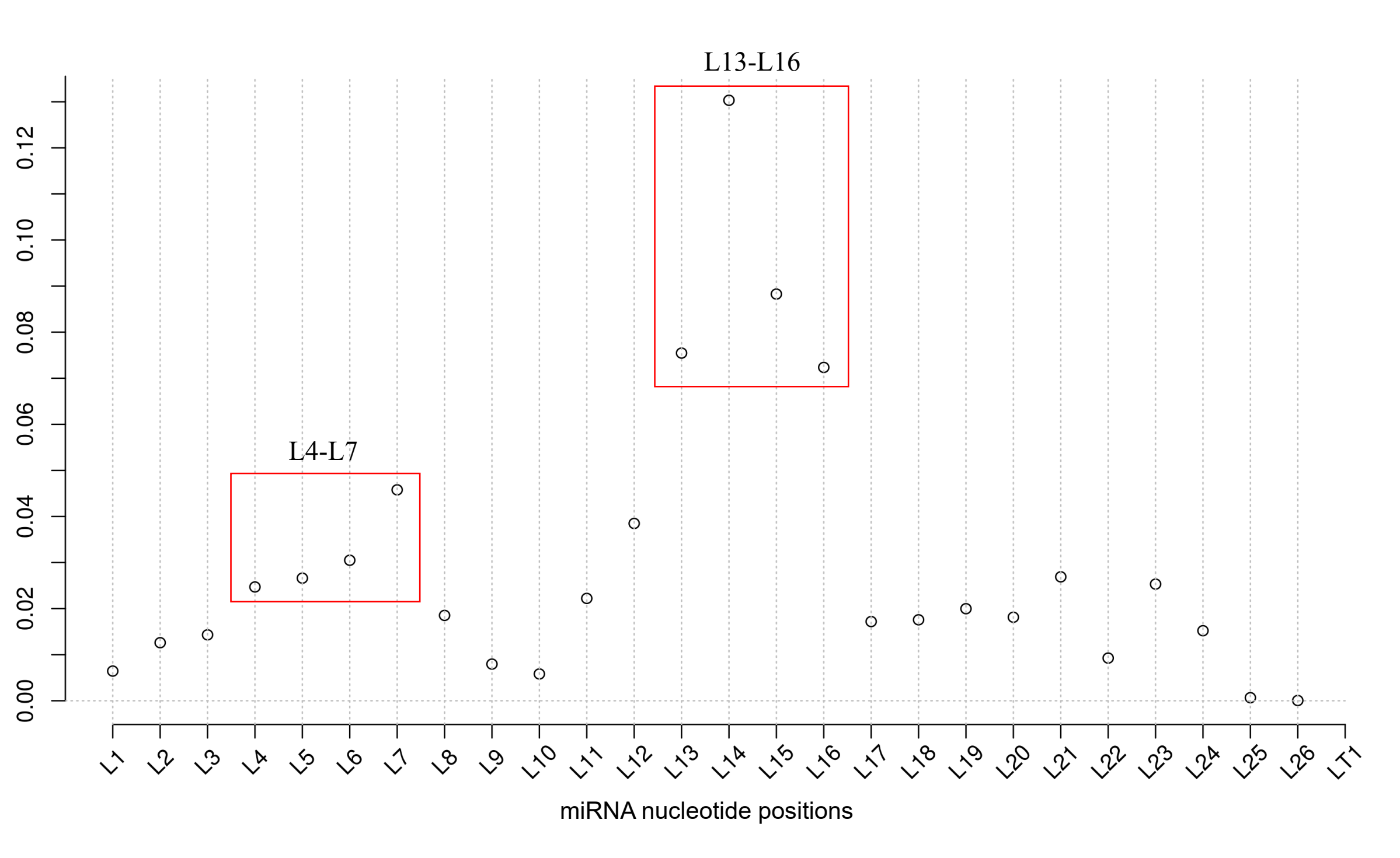


ref: reference sequence (the raw sequence without any alteration)
MutSeed2-6: the sequence after altering the seed region (2nd-6th) to Ns.
Pos: positive miRNA:target pairs
Neg: negative miRNA:target pairs

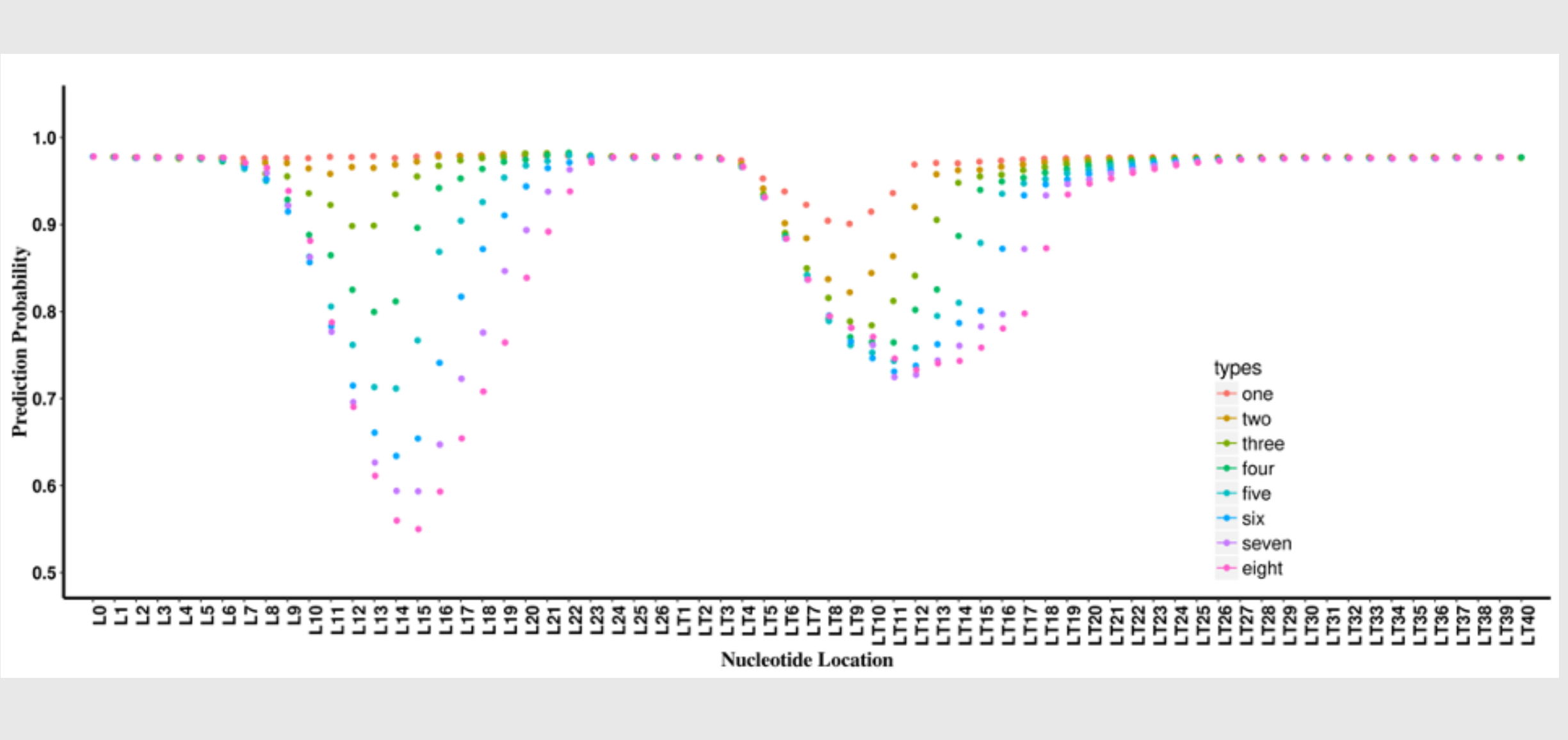
6. The prediction probability for the DeepMirTar positive pairs for the one-, two-, three-, four-, five-, and six-nucleotide mutagenesis analysis



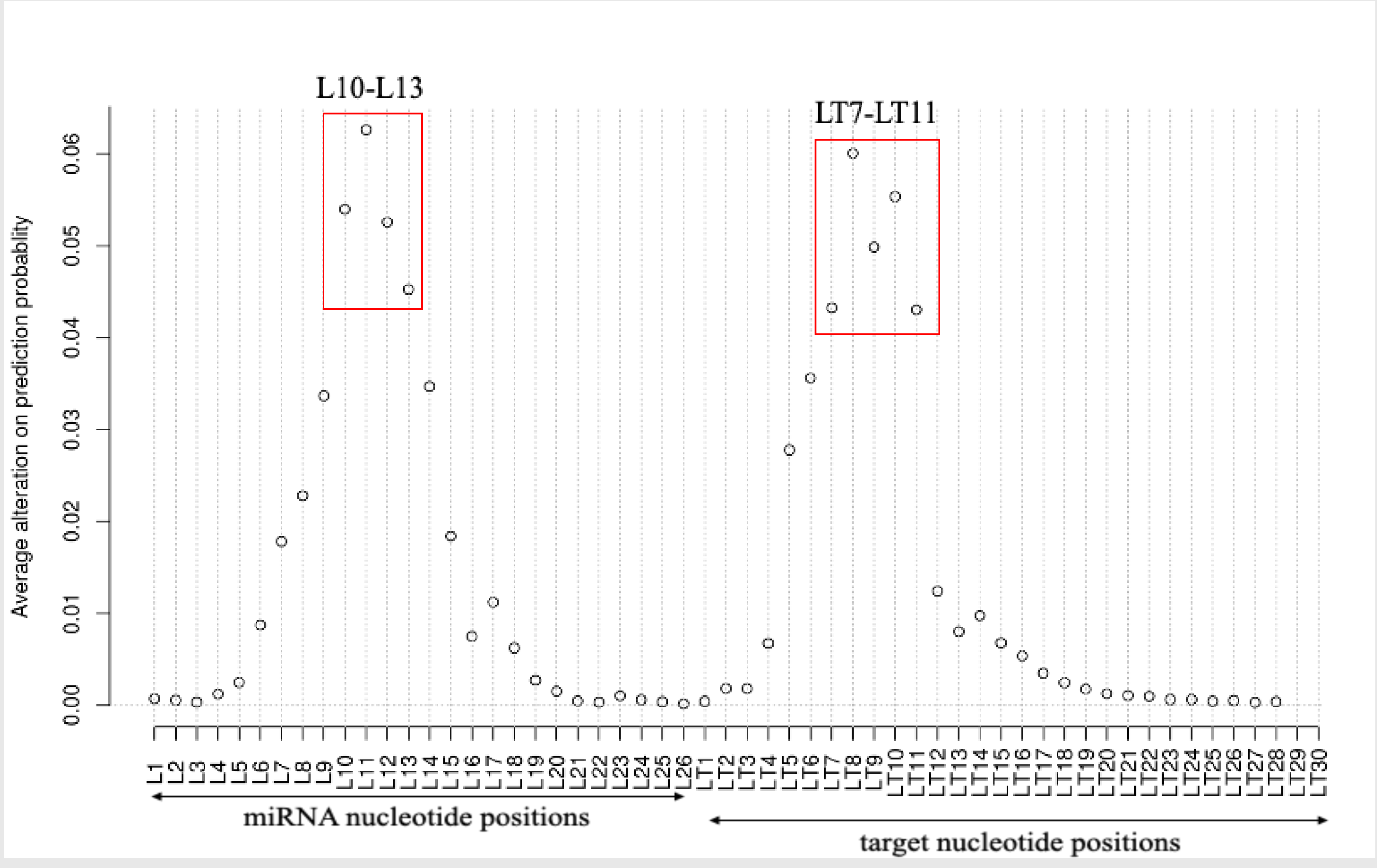
7. Average nucleotide alterations on prediction probability for the DeepMirTar positive pairs



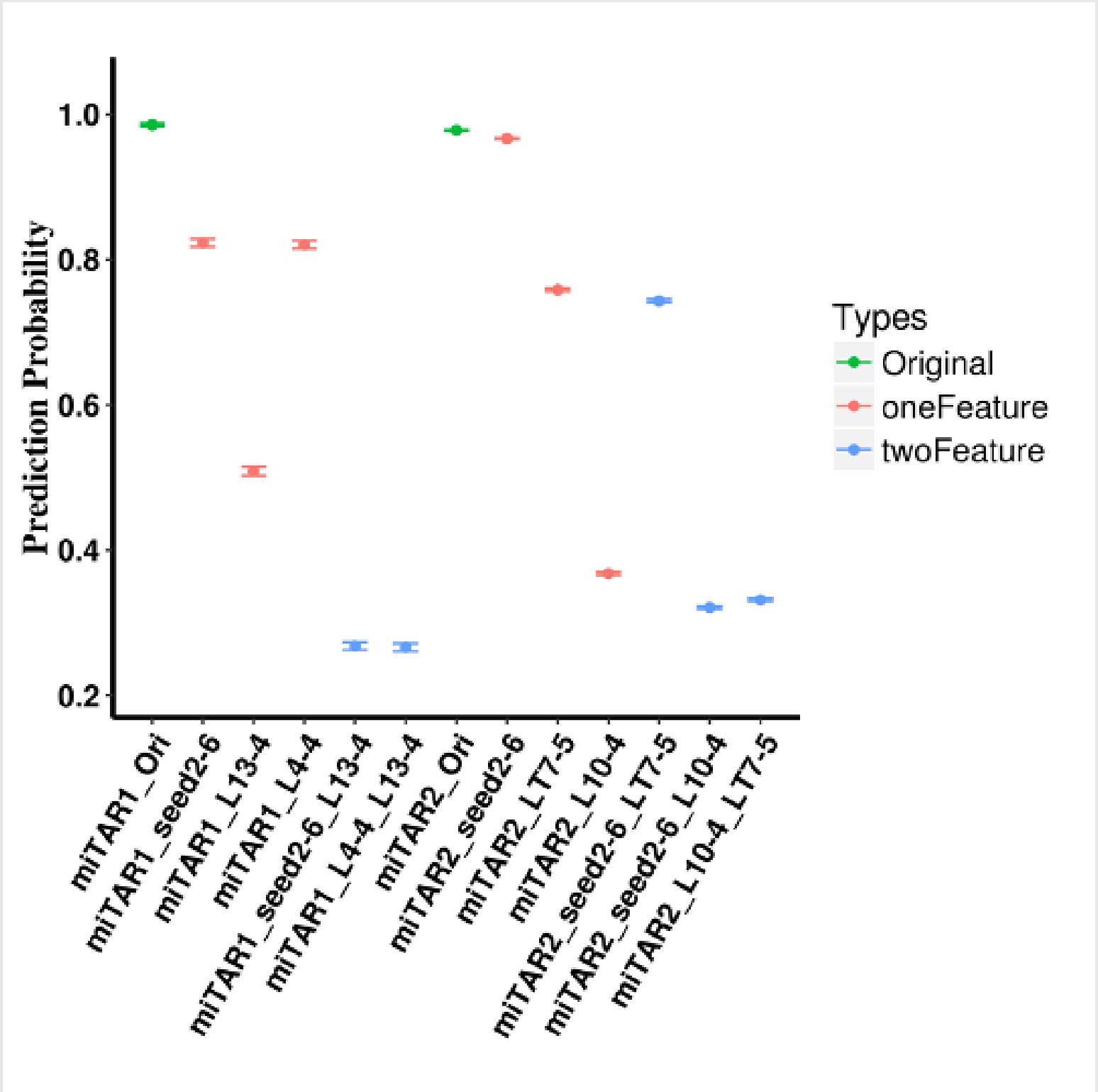
8. The prediction probability for the miRAW positive pairs for the one-, two-, three-, four-, five-, and six-nucleotide mutagenesis analysis



9. Average nucleotide alterations on prediction probability for the miRAW positive pairs



10. The joint and independent impacts of two types of features identified from the in-silico mutagenesis analysis



seed2-6 marks the prediction probability after altering the seed region. L13-4 marks the prediction probability after altering the region from L13 to L16. L4-4 marks the prediction probability after altering the region from L4 to L7. LT7-5 marks the prediction probability after altering the region from LT7 to LT11. L10-4 marks the prediction probability after altering the region from L10 to L13.

References

Biological features between miRNAs and their targets are unveiled from deep learning models
T Gu, M Xie, WB Barbazuk, JH Lee
Scientific Reports 11 (1), 1-10

miTAR: a hybrid deep learning-based approach for predicting miRNA targets
T Gu, X Zhao, WB Barbazuk, JH Lee
BMC Bioinformatics 22 (1), 1-16