doi 10.18699/bgrs2024-1.2-21

New bioinformatics pipeline revealed thousands of novel potential non-canonical binding sites for TDMD-regulated miRNAs in *Drosophila*

Snitkin D.V.[1, 2], Akulenko N.V.[1], Mikhaleva E.A.[1], Ryazansky S.S.[1]*

[1] *National Research Centre "Kurchatov Institute", Moscow, Russia*

[2] *National Research University Higher School of Economics, Moscow, Russia*

* *s.ryazansky@gmail.com*

Key words: miRNA; TDMD; chimeric eCLIP-seq; miRNA binding sites; CLAHer pipeline

Motivation and Aim: During the miRNA-mediated gene silencing in animals, the binding of miRNAs with complementary mRNA targets through a 'seed' region at the 5' end of miRNAs leads to the destruction or translational inhibition of the targets [1]. However, recent high-throughput studies demonstrated that miRNAs mostly prefer to bind with RNA in a non-canonical way, in which the complementary pairings are also located outside the seed region. The biological functions of non-canonical interactions are mostly unknown. The exception is the so-called TDMD (target-directed miRNA degradation) sites, in which miRNA pairs with a target along its whole length with a bulge of several unpaired nucleotides in the middle. Upon binding of miRNAs associated with Ago protein with TDMD sites on RNA targets, Ago is recognized and ubiquitinated by the multiprotein ubiquitin-ligase complex containing substrate receptor ZSWIM8 [2, 3] in mammalian or Dora in *Drosophila* [2, 4]. As a result, Ago undergoes proteolysis, while released miRNA is degraded in the cytoplasm. Although the main features of TDMD sites' structure are known for mammals, their peculiarities in other species remain poorly understood. In our work, we aimed to identify putative TDMD miRNA binding sites with subsequent analysis of their primary and secondary structures in *Drosophila melanogaster* cell culture. The identification and the detailed characterization of features of the *Drosophila* TDMD sites expands our understanding of miRNA biology in general.

Methods and Algorithms: For the identification of the miRNA binding sites on the cognate *Drosophila* RNA targets in OSC cell culture, we used the chimeric eCLIP method combined with PCR enrichment [5] of several TDMD-regulated miRNAs [4]. Additionally, we prepared the eCLIP libraries from cells with depleted Dora. We suggested that TDMD targets would accumulate in Dora-depleted cells thus facilitating their identification. After the high-throughput sequencing, we get ~126M chimeric raw reads in total. For the analysis of the sequenced chimeric molecules, we have developed the bioinformatic pipeline on Python called CLASHer. This pipeline allows pre-processing of the raw reads, identifies the miRNA interaction sites on *Drosophila* transcripts, and reveals various structural features of miRNA binding sites. These features include, for example, the length of the complementary regions between miRNAs and target, the degree of site evolutionary conservation, the site accessibility level, and many others. Finally, we have applied machine learning to classify the non-canonical sites with a focus on the identification of TDMD binding sites.

Results: We identified ~220k chimeras for 5.6k binding sites of 17 miRNAs subjected to TDMD in *Drosophila* OSC cell culture. While the majority (37.1 %) of the identified miRNA binding sites are located in CDS regions of protein-coding mRNAs, only 19.6 % positioned in 3'UTR where the most animal canonical miRNA binding sites are located (Fig. 1A). Only 3.2 % of the binding sites have the seed region while the vast majority of sites (88.6 %) are non-canonical ones without typical seed complementary region (Fig. 1B). We also identified in our data several TDMD targets (e.g. AGO1/miR-999) that were revealed previously [6]; this clearly shows the overall validity of our approach for the identification of the TDMD binding sites.

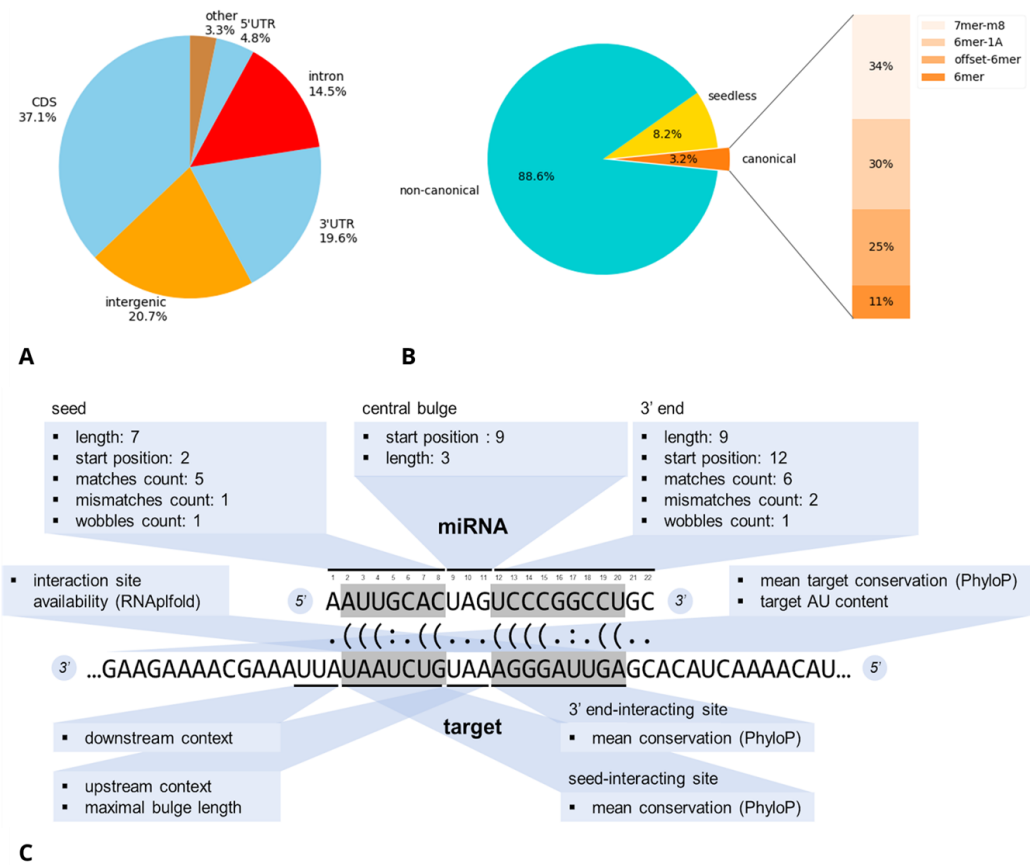

Fig. 1. A, The location of the identified miRNA binding sites. B, The types of the identified miRNA binding sites. C, Some features of the miRNA-target interaction obtained by the pipeline

Then we characterized each binding site by twenty parameters (features), some of them are depicted in Fig. 1C. The correlational and dimensionality reduction methods revealed which features are redundant and should be excluded from subsequent analysis. Finally, we selected the set of ~2k binding sites supported at least by three chimeric molecules in our dataset and classified them based on their features into several groups using machine learning methods. We will present details and results of the work, as well as

attempt to create a predictive algorithm based on machine learning techniques for the identification of TDMD sites.

Conclusion: The developed pipeline allowed to detect 5.6k potential binding sites for TDMD-regulated miRNAs of Drosophila. The conducted detailed characterization of their structural features can be used in building a model for predicting of TDMD binding sites of miRNAs.

Funding: The work was carried out within the state assignment of NRC "Kurchatov Institute".

References

1. Bartel D.P. Metazoan MicroRNAs. *Cell.* 2018;173(1):20-51. doi 10.1016/j.cell.2018.03.006
2. Shi C.Y. et al. The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation. *Science.* 2020;370(6523):eabc9359
3. Han J. et al. A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming. *Science.* 2020;370(6523):eabc9546
4. Akulenko N. et al. Evidence of target-mediated miRNA degradation in Drosophila ovarian cell culture. *bioRxiv.* 2023. doi 10.1101/2023.08.30.555489
5. Manakov S.A. et al. Scalable and deep profiling of mRNA targets for individual microRNAs with chimeric eCLIP. *bioRxiv.* 2022. doi 10.1101/2022.02.13.480296
6. Sheng P. et al. Screening of Drosophila microRNA-degradation sequences reveals Argonaute1 mRNA's role in regulating miR-999. *Nat Commun.* 2023;14:2108