tRFtarget: a database for transfer RNA-derived fragment targets

Ningshan Li[1,2], Nayang Shan[2,3], Lingeng Lu[4,*] and Zuoheng Wang[2,*]

[1] SJTU-Yale Joint Center for Biostatistics and Data Science, Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

[2] Department of Biostatistics, Yale School of Public Health, New Haven, CT 06520, USA

[3] Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

[4] Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT 06520, USA

* To whom correspondence should be addressed. Tel: 1-203-737-2672; Fax: 1-203-785-6912; Email: zuoheng.wang@yale.edu. Correspondence may also be addressed to lingeng.lu@yale.edu.

ABSTRACT

Transfer RNA-derived fragments (tRFs) are a new class of small noncoding RNAs and play important roles in biological and physiological processes. Prediction of tRF target genes and binding sites is crucial in understanding the biological functions of tRFs in the molecular mechanisms of human diseases. We developed a publicly accessible web-based database, tRFtarget (http://trftarget.net), for tRF target prediction. It contains the computationally predicted interactions between tRFs and mRNA transcripts using the two state-of-the-art prediction tools RNAhybrid and IntaRNA, including location of the binding sites on the target, the binding region, and free energy of the binding stability with graphic illustration. tRFtarget covers 936 tRFs and 135 thousand predicted targets in eight species. It allows researchers to search either target genes by tRF IDs or tRFs by gene symbols/transcript names. We also integrated the manually curated experimental evidence of the predicted interactions into the database. Furthermore, we provided a convenient link to the DAVID® web server to perform downstream functional pathway analysis and gene ontology annotation on the predicted target genes. This database provides useful information for the scientific community to experimentally validate tRF target genes and facilitate the investigation of the molecular functions and mechanisms of tRFs.

INTRODUCTION

Transfer RNA-derived fragments (tRFs) are a new class of small noncoding RNA (ncRNA) in the length of 13-48 nucleotides (nts) (1). tRFs are the products of non-random cleavage of either the precursor or mature transfer RNAs (tRNAs). Based on the cleavage sites of tRNA, tRFs are classified into five categories: tRF-1s (also termed as tRNA-derived small RNAs [tsRNAs]), tRF-3s, tRF-5s, tRF-2s and stress-induced tRNAs (tiRs) (Fig. 1) (2). tRF-1s are generated from the 3' end of the precursor tRNAs, whereas the other four types of tRFs are derived from different parts of the mature tRNAs with the 3' end for tRF-3s and the 5' end for tRF-5s, respectively (2). tRF-3s can be further divided into two subcategories with different lengths: short as tRF-3a and long as tRF-3b. Similarly, based on the length from short to long, tRF-5s can be divided into three subcategories: tRF-5a, tRF-5b and tRF-5c (3). tiRs have two subcategories: a 5tiR starts from the 5' end of a mature tRNA to the end of the anticodon loop, and a 3tiR from the 3' end to the end of the anticodon loop. tRF-2s, a new type of tRFs, are derived from the anticodon loop of the mature tRNAs (2).

tRFs have been found to be conserved in diverse organisms from bacteria to humans (4). They are involved in many biological and physiological processes such as regulation of gene expression, RNA processing, tumor suppression, and cell proliferation (2). tRF dysregulation may play important roles in human diseases including cancer (5-7). Recently, several studies revisited datasets from The Cancer Genome Atlas (TCGA) and the NCI-60 human tumor cell lines screen, and reported tRFs as potential biomarkers in human cancer (8,9). Experimental results suggested that tRFs function as microRNAs (miRNAs) in post-transcriptional regulation of gene expression by partially complementary to target messenger RNAs (mRNAs), leading to the degradation or translational repression of target mRNAs (10,11). The results of photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) in human HEK293 cells showed the silencing complex formation of tRF-5s and tRF-3s in combination with Argonautes 1, 3, and 4 (4,12). RNA sequencing analysis further demonstrated the tRF-mRNA chimeric formation in the cross-linking, ligation and sequencing of hybrids (CLASH) data (4,13,14).

There are more than 30 web-based miRNA target prediction databases which have served as powerful tools for experimental validation of miRNA targets in multiple species (15,16), and accelerated the investigation of the biological functions of miRNAs by providing the binding sites on

mRNAs. However, to our knowledge, there is no such target database available for tRFs. A previous study inferred the targeting modes of tRFs based on the limited experimental CLASH datasets that cover 26 human Argonaute-loaded tRFs (14). However, the bindings may be biased in the context of tissues and cells. Another approach to infer potential tRF targets is via co-expression network analysis or chromatin immunoprecipitation sequencing (ChIP-seq), which do not directly consider the complementary pairing between sequences (8,9). Site-directed mutations using reporter genes is a fundamental approach to seek the binding sites in the target genes, however, it requires the known regions where the binding site(s) is located.

In this study, we established a publicly accessible web-based transcriptome-wide tRF target prediction database, tRFtarget (http://trftarget.net), for eight species including human, mouse, *Drosophila*, *C. elegans*, *S. pombe*, *R. sphaeroides*, *Xenopus tropicalis*, and Zebrafish. It was hosted by the Extreme Science and Engineering Discovery Environment (XSEDE) (17). tRF target genes were computationally predicted based on the interactions between tRFs and mRNA transcripts using the two state-of-the-art algorithms RNAhybrid (18,19) and IntaRNA (20). The contents of the database include maximum complementary length (MCL) of the paired tRF and target transcript, binding sites on the transcript (from 5' end) and binding regions of 5' untranslated region (UTR), coding sequence (CDS), or 3' UTR of the transcript, free energy which is a measure of the stability of the binding between a tRF and a candidate target transcript, and graphic illustration of the pairing sequences. tRFtarget allows researchers to search tRF-target interactions by either tRF ID, or transcript or gene symbol, and demonstrates the pairing regions between a tRF and target transcripts as well as their locations. It facilitates researchers to use such information to perform various experiments such as amplifying the interested regions, making mutations in the regions for clone and plasmid construction, and constructing reporter assays for validation of target genes. The database also provides experimental evidence of the predicted tRF-mRNA interactions and functional studies on tRFs based on manually curated publications. In addition, the predicted target genes can be further used for functional pathway analysis and gene ontology annotation, providing the potential biological functions of tRFs.

MATERIAL AND METHODS

Data source

A workflow of the tRFtarget database construction is shown in Fig. 2. The tRF sequences were retrieved from tRFdb (http://genome.bioch.virginia.edu/trfdb/) (3) and a tsRNA study (21). The transcript sequences were currently restricted to protein-coding transcripts, and downloaded from GENCODE (version GRCh38.p13 for human and version GRCm38.p6 for mouse, https://www.gencodegenes.org/) (22). Human gene symbols and the corresponding Ensembl gene IDs were downloaded from HGNC BioMart (https://biomart.genenames.org/) (23), and mouse gene symbols and Ensembl gene IDs were downloaded from MGI (http://www.informatics.jax.org/) (24). The transcript names and the corresponding Ensembl transcript IDs were downloaded from Ensembl BioMart (version GRCh38.p13 for human and version GRCm38.p6 for mouse, http://www.ensembl.org/biomart/martview/) (25). The genes and protein-coding transcripts of other species were downloaded from Ensembl BioMart (version BDGP6.28 for *Drosophila*, version WBcel235 for *C. elegans*, version Xenopus_tropicalis_v9.1 for *Xenopus tropicalis*, and version GRCz11 for Zebrafish), EnsemblFungi BioMart (version ASM294v2 for *S. pombe*, http://fungi.ensembl.org/biomart/martview/) and EnsemblBacteria (version ASM1640v1 for *R. sphaeroides*, https://bacteria.ensembl.org/Rhodobacter_sphaeroides_atcc_17025/Info/Index) (25).

RNAhybrid prediction

RNAhybrid (version 2.1.2, https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid) is a computational tool for predicting interactions between a short RNA and a long target RNA based on free energy (18,19). We set an energy threshold of -15 kcal/mol and provided 5 best binding sites per RNA pair given by RNAhybrid. As this algorithm only allows an input of a target sequence of no more than 50 kilobases (kb) in length, all transcripts longer than 50 kb are excluded for prediction and archived in the database. Other parameters were set at the default values of the algorithm. We ran RNAhybrid for all pairs of tRFs and transcripts, parsed the unstructured results, extracted important features, and then stored them in the database. MCL is the length of the longest successively complementary sequences for a specific interaction. Entries with a MCL less than 6 nts were excluded because such a short complementary sequence is unlikely to result in a functional interaction (26).

IntaRNA prediction

IntaRNA (version 3.1.3, http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp) is another computational tool to predict interactions between two RNA molecules based on a combination of free energy and accessibility (20,27). We used the exact mode in IntaRNA to get more accurate target gene prediction. We set a seed length threshold of 6 nts and provided 5 best binding sites per RNA pair. Other parameters were set at the default values of the algorithm. We ran IntaRNA for all pairs of tRFs and transcripts, parsed the results, and then stored them in the database. Interactions between the same tRF and transcript with similar pairing sequences and almost identical binding locations were considered as duplicated entries, and only the entry with a minimal free energy was kept (see manual webpage (http://trftarget.net/manual) for an example of duplicated entries).

Consensus between predictions with RNAhybrid and IntaRNA

We evaluated the concordance of binding sites predicted from RNAhybrid and IntaRNA. A consensus pair of predictions was defined as similar pairing sequences and almost identical binding locations on a transcript (allowing 2 nts offset in the start and/or end of the pairing sequences, see manual webpage (http://trftarget.net/manual) for an example of consensus predictions). Users have an option to search for consensus predictions in the database.

Functional pathway analysis

In the database, we provided users a list of the predicted target genes to perform downstream functional pathway analysis and gene ontology annotation. The gene list was obtained based on the free energy and MCL of the interactions with a tRF. The interactions between a tRF and all target transcripts were ranked by an ascending order of free energy, and the interactions with the same free energy were further ranked by a descending order of MCL. When ranking genes from consensus interactions, free energy generated by IntaRNA was used. We then selected the top interactions to create a list of gene Ensemble IDs. The default number of the top interactions is 2000. Users can also specify the number of genes to be included in the top gene list. The gene list is copied to clipboard automatically and can be pasted into the input box on the DAVID® web server (https://david.ncifcrf.gov/) for functional pathway analysis (28,29).

RESULTS

Database overview

tRFtarget includes 936 tRFs, 135 thousand target genes and 294 million interactions across eight species. The database can be queried in 3 ways:

1) search by tRF ID for all predicted target transcripts;

2) search by transcript Ensembl ID or name for all potential interacting tRFs;

3) search by gene Ensembl ID or symbol for all tRFs which potentially interact with any transcripts of a given gene.

In each query method, additional filter criteria on the binding regions, free energy and MCL can be specified by users to narrow the prediction results. Results are displayed in a tabular form and can be downloaded as a CSV file. Sorting or filtering by table column is also supported. There are two columns in the table that demonstrate the experimental evidence at the gene and site levels for each predicted interaction, respectively. Furthermore, hyperlinks to the relevant information from other databases were provided in the search results for users to browse detailed information of tRFs, genes and transcripts. A click button allows users to perform downstream functional pathway analysis on tRF target genes. Table 1 summarizes the total number of tRFs, genes and transcripts, and the prediction results by RNAhybrid and IntaRNA across eight species. A detailed description of this information can be found on the statistics webpage of the database (http://trftarget.net/statistics).

Validation of prediction

A previous study investigated interactions between small ncRNAs and mRNAs in the HEK293 cells using the CLASH experiments (13) and identified many tRF-mRNA chimeras, especially tRF-3-mRNA chimeras (4,14). We used the top 10 most abundant tRF-3-mRNA experimental interactions (4) for validation of our prediction results. Among the top 10 interactions in the CLASH data, four tRFs, tRF-3034a, 3035a, 3036a, and 3037a, were not present in human tRFdb (3) and as a result were not included in the database. For tRF-3014a, the MCL of the reported interaction is 5 nts which is less than the threshold value of 6 nts, so this interaction structure was not indexed in the database. We then considered the experimental interactions from the remaining 5 tRFs to validate the prediction results. Fig. 3 shows the interactions with mRNAs of the 5 tRFs illustrated by the predicted interactions of the CLASH tRF-3-mRNA chimeras using mfold, a computational tool for the prediction

of the RNA secondary structure (30), and the predicted interactions in the tRFtarget database whose structures are most similar to the mfold predictions. All 5 CLASH chimeras have matched entries in the database. Moreover, our database provided more interactions with target mRNAs. For example, we found another interaction between tRF-3001a and the target mRNA RABEPK-203, besides the interactions of tRF-3001a with the two mRNAs DCTPP1-201 and DCTPP1-203 in the CLASH data. All three interactions have the same pairing sequences.

Comparison with gene correlation analysis

In our previous study of tRFs in breast cancer, five tRFs, tRF-5024a, ts-34, ts-49, ts-58 and 5P_tRNA-Leu-CAA-4-1, were significantly associated with breast cancer patient survival. Among the five tRFs, tRF-5024a had the largest number of correlations with mRNA transcripts in which there were 404 positively and 2,292 negatively correlated genes (31). Here, we compared the predicted targets of tRF-5024a with the genes identified in correlation analysis. Among the 2,292 genes that were negatively correlated with tRF-5024a, 625 genes are noncoding genes and were not indexed in the database. In the remaining 1,667 genes, 1,506 genes (90.3%) had at least one consensus interaction with tRF-5024a. Among the 404 genes that were positively correlated with tRF-5024a, 50 genes are noncoding genes and were not indexed in the database. In the remaining 354 genes, 296 genes (83.6%) had at least one consensus interaction with tRF-5024a. The target gene information and binding sites for the 1,667 positively and 354 negatively correlated genes with tRF-5024a were shown in Table S1.

DISCUSSION

tRFs are a novel class of regulatory ncRNAs and have been found to be dysregulated in cancer (1). Although previous studies have revealed the roles of tRFs in biological processes, there are still many features of tRFs, such as the biogenesis of tRF-3s and tRF-5s, and the functions of tRFs, yet to be fully characterized (2). To facilitate the research on tRFs, serval tRF-related databases have been developed based on omics data. For example, PtRFdb collects detailed information of plant tRFs (32); tRFdb includes a standardized nomenclature of tRFs in eight organisms (3); tRFexplorer provides expression profiles of identified tRFs in every cell line in NCI-60 and for each TCGA cancer type (9); MINTbase contains the abundance of tRFs identified in all TCGA projects (33); and tRF2Cancer not

only provides the expression of identified tRFs in TCGA but supports searching for novel tRFs from user uploaded sequencing data (34). However, to our knowledge, there is no database available for a comprehensive prediction of tRF targets. tRFtarget developed in this study is the first tRF target prediction database to fill this gap. In tRFtarget, interactions between tRFs and mRNA transcripts were predicted by two algorithms, RNAhybrid (18,19) and IntaRNA (20). Both methods are free energy-based tools to predict RNA-RNA interaction. RNAhybrid is a popular tool to discover the targeting modes of small ncRNAs (14) and achieves more precise prediction than alignment-based methods (35,36). It utilizes the dynamic programming technique to efficiently calculate the optimal binding sites and can be easily run in parallel. RNAhybrid achieves high sensitivity but low precision in a comprehensive comparison study (36). IntaRNA, on the other hand, uses the "accessibility" feature (20,27) to improve the accuracy of target prediction (35-37). "Accessibility" is calculated as the free energy needed to unfold the binding sites in both RNAs to make the sites accessible for hybridization. It measures the likelihood that a short RNA is able to bind a specific site on the target mRNA and leads to a higher but more diverse free energy. As a result, IntaRNA achieves high sensitivity and high precision in target prediction (35,36).

We provided 5 best binding sites per RNA pair rather than the one with minimal free energy in tRFtarget for the following considerations. First, similar to miRNAs, each tRF may have multiple binding sites with a target transcript in the process of gene regulation (38). Second, the accuracy of target prediction commonly decreases as the length of target transcript increases when we use prediction based on minimal free energy. Providing more binding sites will be more likely to include the true interaction even when the target transcript is long (35,36). Third, the interaction with minimal free energy may not be biologically active. We also provided the suboptimal interactions with a bit higher free energy than the minimal one for users' reference (36). When comparing the predicted targets in tRFtarget with the CLASH data (4), our database covered the most abundant tRF-mRNA interaction structures of the CLASH dataset, suggesting the reliability of target prediction. We also evaluated the overlap of the predicted targets and the correlated genes of tRF-5024a. More than 80% of the correlated genes were found to interact with tRF-5024a. Because correlation based on gene expression does not necessarily suggest interaction, there were some correlated genes that were not the predicted targets of tRF-5024a.

tRFtarget can be further improved in the following aspects. First, tRFs in the current version of the database were retrieved from tRFdb (3), and novel tRFs will continuously be identified by experiments. We will regularly update the database to include more tRFs and their targets. Second, tRFs interacted with a variety of RNAs including mRNAs, miRNAs and long intergenic ncRNAs (lincRNAs) in the CLASH data (14). We will expand the database by including ncRNAs as potential targets. Third, the motifs of tRFs based on interactions with different transcripts will be included in the database for query. Lastly, tRFtarget may include false positive interactions, but still provides useful information for further cost-effective experimental validation.

In conclusion, we developed tRFtarget, a transcriptome-wide tRF target prediction database for querying interactions between tRFs and transcripts in eight species. Manually curated experimental evidence was integrated into the database. This database provides useful information to guide biological experiments and target validation, as well as accelerates the understanding of the function and mechanism of tRFs.

DATA AVAILABILITY

All data in tRFtarget are freely accessible at http://trftarget.net.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Lee, Y.S., Shibata, Y., Malhotra, A. and Dutta, A. (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & Development*, **23**, 2639-2649.

2. Kumar, P., Kuscu, C. and Dutta, A. (2016) Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). *Trends in Biochemical Sciences*, **41**, 679-689.

3. Kumar, P., Mudunuri, S.B., Anaya, J. and Dutta, A. (2014) tRFdb: a database for transfer RNA fragments. *Nucleic Acids Research*, **43**, D141-D145.

4. Kumar, P., Anaya, J., Mudunuri, S.B. and Dutta, A. (2014) Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biology*, **12**, 78.

5. Balatti, V., Pekarsky, Y. and Croce, C.M. (2017) In Croce, C. M. and Fisher, P. B. (eds.), *Advances in Cancer Research*. Academic Press, Vol. 135, pp. 173-187.

6. Sun, C., Fu, Z., Wang, S., Li, J., Li, Y., Zhang, Y., Yang, F., Chu, J., Wu, H., Huang, X. *et al.* (2018) Roles of tRNA-derived fragments in human cancers. *Cancer Letters*, **414**, 16-25.

7. Zhu, L., Liu, X., Pu, W. and Peng, Y. (2018) tRNA-derived small non-coding RNAs in human disease. *Cancer Letters*, **419**, 1-7.

8. Telonis, A.G., Loher, P., Magee, R., Pliatsika, V., Londin, E., Kirino, Y. and Rigoutsos, I. (2019) tRNA Fragments Show Intertwining with mRNAs of Specific Repeat Content and Have Links to Disparities.

Cancer Research, **79**, 3034-3049.

9. La Ferlita, A., Alaimo, S., Veneziano, D., Nigita, G., Balatti, V., Croce, C.M., Ferro, A. and Pulvirenti, A. (2019) Identification of tRNA-derived ncRNAs in TCGA and NCI-60 panel cell lines and development of the public database tRFexplorer. *Database*, **2019**, baz115.

10. Sobala, A. and Hutvagner, G. (2013) Small RNAs derived from the 5′ end of tRNA can inhibit protein translation in human cells. *RNA Biology*, **10**, 553-563.

11. Kuscu, C., Kumar, P., Kiran, M., Su, Z., Malik, A. and Dutta, A. (2018) tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA*, **24**, 1093-1105.

12. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.-C., Munschauer, M. *et al.* (2010) Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, **141**, 129-141.

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. Guan, L., Karaiskos, S. and Grigoriev, A. (2020) Inferring targeting modes of Argonaute-loaded tRNA fragments. *RNA Biology*, **17**, 1070-1080.

15. Singh, N.K. (2017) miRNAs target databases: developmental methods and target identification techniques with functional annotations. *Cellular and Molecular Life Sciences*, **74**, 2239-2261.

16. Monga, I. and Kumar, M. (2019) In Lai, X., Gupta, S. K. and Vera, J. (eds.), *Computational Biology of Non-Coding RNA: Methods and Protocols*. Springer New York, New York, NY, pp. 215-250.

17. Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D. *et al.* (2014) XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering*, **16**, 62-74.

18. REHMSMEIER, M., STEFFEN, P., HÖCHSMANN, M. and GIEGERICH, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507-1517.

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

20. Mann, M., Wright, P.R. and Backofen, R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Research*, **45**, W435-W439.

21. Balatti, V., Nigita, G., Veneziano, D., Drusco, A., Stein, G.S., Messier, T.L., Farina, N.H., Lian, J.B.,

Tomasello, L., Liu, C.-g. *et al.* (2017) tsRNA signatures in cancer. *Proceedings of the National Academy of Sciences*, **114**, 8071-8076.

22. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J. *et al.* (2018) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, **47**, D766-D773.

23. Braschi, B., Denny, P., Gray, K., Jones, T., Seal, R., Tweedie, S., Yates, B. and Bruford, E. (2018) Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Research*, **47**, D786-D792.

24. Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A., Richardson, J.E. and Group, t.M.G.D. (2018) Mouse Genome Database (MGD) 2019. *Nucleic Acids Research*, **47**, D801-D806.

25. Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G. *et al.* (2017) Ensembl 2018. *Nucleic Acids Research*, **46**, D754-D761.

26. Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. (2005) Principles of MicroRNA–Target Recognition. *PLOS Biology*, **3**, e85.

27. Miladi, M., Montaseri, S., Backofen, R. and Raden, M. (2018) Integration of accessibility data from structure probing into RNA–RNA interaction prediction. *Bioinformatics*, **35**, 2862-2864.

28. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2008) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**, 1-13.

29. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**, 44-57.

30. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, **31**, 3406-3415.

31. Shan, N., Li, N., Dai, Q., Hou, L., Yan, X., Amei, A., Lu, L. and Wang, Z. (2020) Interplay of tRNA-Derived Fragments and T Cell Activation in Breast Cancer Patient Survival. *Cancers*, **12**, 2230.

32. Gupta, N., Singh, A., Zahra, S. and Kumar, S. (2018) PtRFdb: a database for plant transfer RNA-derived fragments. *Database*, **2018**, bay063.

33. Pliatsika, V., Loher, P., Magee, R., Telonis, A.G., Londin, E., Shigematsu, M., Kirino, Y. and Rigoutsos, I. (2017) MINTbase v2.0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all The Cancer Genome Atlas projects. *Nucleic Acids Research*, **46**, D152-D159.

34. Zheng, L.-L., Xu, W.-L., Liu, S., Sun, W.-J., Li, J.-H., Wu, J., Yang, J.-H. and Qu, L.-H. (2016)

tRF2Cancer: A web server to detect tRNA-derived small RNA fragments (tRFs) and their expression in multiple cancers. *Nucleic Acids Research*, **44**, W185-W193.

35. Lai, D. and Meyer, I.M. (2015) A comprehensive comparison of general RNA–RNA interaction prediction methods. *Nucleic Acids Research*, **44**, e61.

36. Umu, S.U. and Gardner, P.P. (2016) A comprehensive benchmark of RNA–RNA interaction prediction tools for all domains of life. *Bioinformatics*, **33**, 988-996.

37. Richter, A. and Backofen, R. (2012) Accessibility and conservation: General features of bacterial small RNA–mRNA interactions? *RNA Biology*, **9**, 954-965.

38. Millar, A.A. and Waterhouse, P.M. (2005) Plant and animal microRNAs: similarities and differences. *Functional & Integrative Genomics*, **5**, 129-135.

TABLE AND FIGURES LEGENDS

Table 1. Summary statistics of tRFtarget database

Figure 1. Illustration of the five categories of tRFs. tRF-1 or tsRNA is generated from the 3' end of the precursor tRNA, whereas tRF-3, tRF-5, tiR and tRF-2 are derived from different parts of the mature tRNA. tRF, transfer RNA-derived fragment. tsRNA, tRNA-derived small RNA. tiR, stress-induced tRNA. ANG, angiogenin.

Figure 2. Flow diagram of constructing tRFtarget database. tRFtarget can be retrieved through web service. tRF, transfer RNA-derived fragment.

Figure 3. Illustration of the interaction structures of tRF-3-mRNA chimeras from CLASH and the corresponding binding sites in tRFtarget. The interaction structures of tRF-3-mRNA chimeras were inferred by mfold RNA Folding Form using default settings, and the reads of the chimeras and the delta G values (dG) given by mfold are shown at the right. The binding sites in tRFtarget are shown below the mfold prediction. The transcript names are shown at the left, and the prediction algorithms and free energy (FE) are shown at the right. Black represents non-interaction sequences; blue represents interaction sequences of tRFs, green and red represent paired and non-paired bases of mRNA, respectively. CLASH, cross-linking, ligation and sequencing of hybrids.

Table 1. Summary statistics of tRFtarget database

Species	tRFs	Target transcripts	Target genes	Target interactions
Human	270	100,291	20,355	153,284,308
Mouse	164	67,125	22,385	60,074,771
Drosophila	86	30,588	13,947	14,822,386
C. elegans	47	33,552	20,191	7,685,677
S. pombe	104	5,146	5,145	3,105,075
R. sphaeroides	87	3,111	3,111	1,433,321
Xenopus tropicalis	51	54,848	19,983	16,449,592
Zebrafish	127	51,259	30,149	37,389,780

Precursor tRNA

5' PPP ━━━━━━━━ ━━━━━ UUU 3'

RNase Z

tRF-1 / tsRNA

━━━━━ UUU 3'

tRF-5

tRF-5c   tRF-5b   tRF-5a

Mature tRNA

5' P   CCA 3'

Dicer / Others

D loop   T loop

tRF-3

tRF-3a   tRF-3b

Dicer / Others

ANG

5tiR

tiR

Anticodon loop

5' 3'

tRF-2

ANG

3tiR

tiR
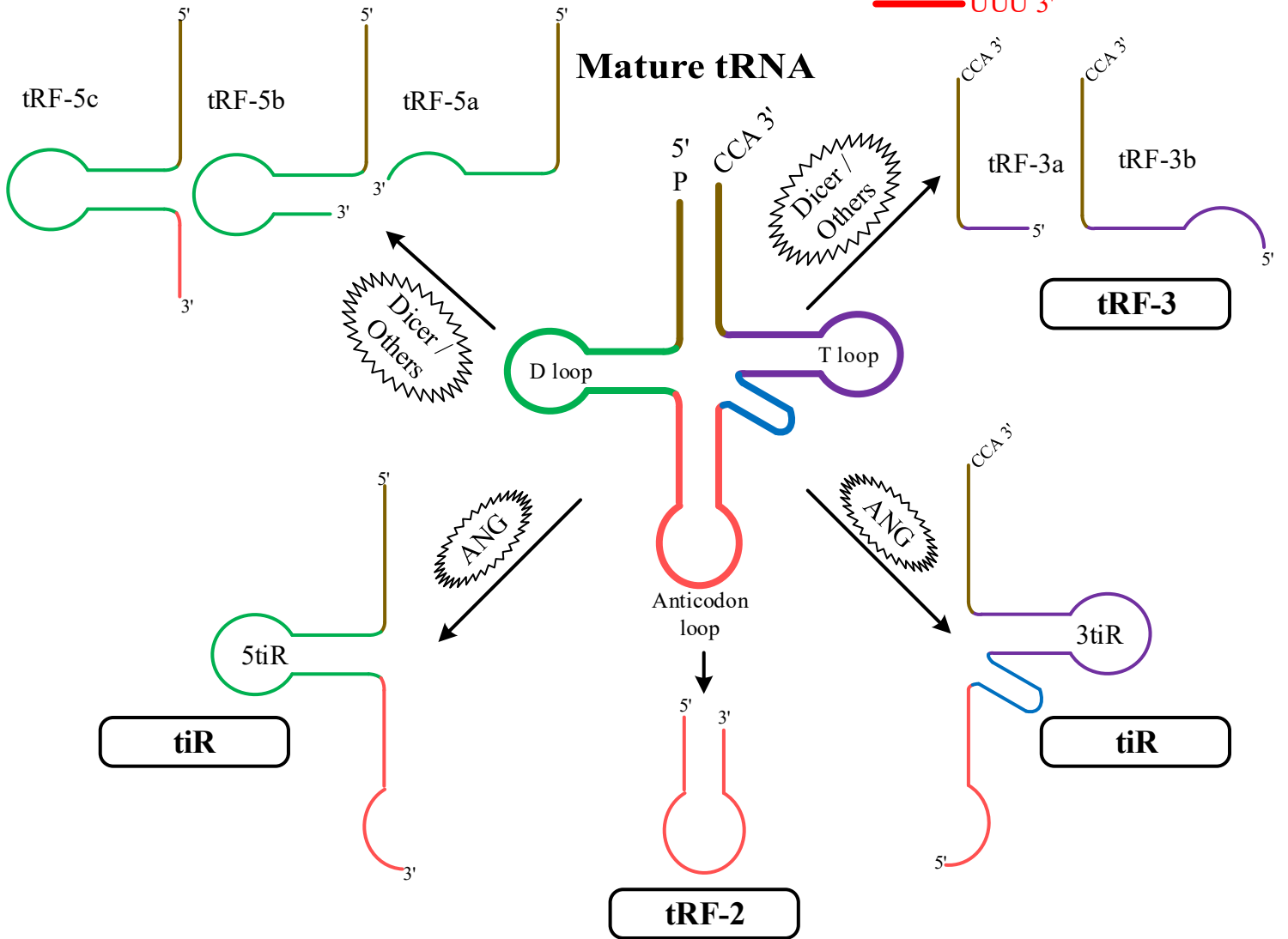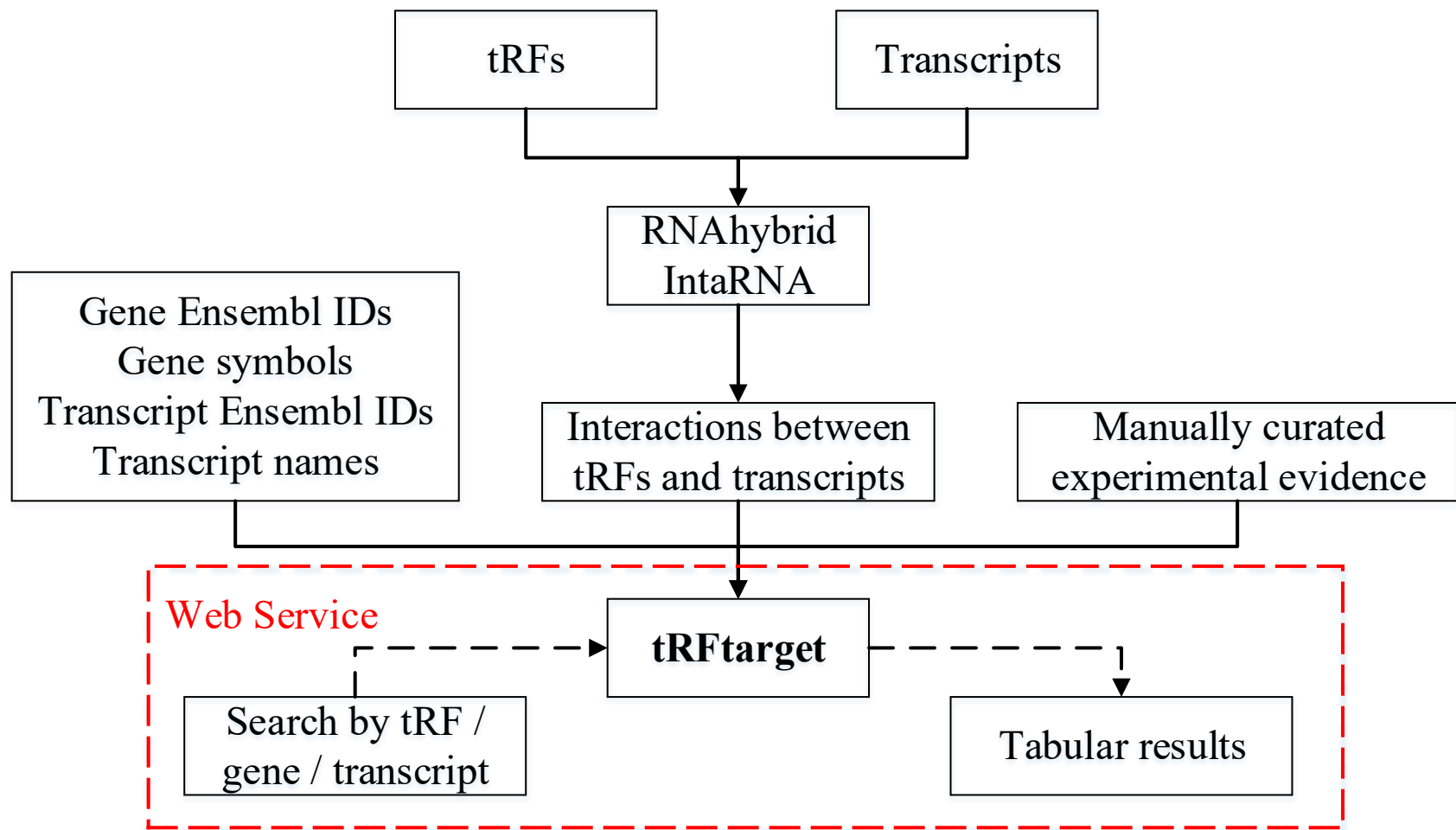
```
tRF-3003a            3'-ACCUCCCCCGUGGGC CU-5'
HIST2H2AA4        5'-     GAAGGGGCACCUGUGA        -3'    2025 Reads,  dG=-22.0
HIST2H2AA4-201 5'-CGA...CUCGAAGGGGCACCUGUGAACU...CUG-3'  RNAhybrid,   FE=-29.5
               5'-CGA...CUCGAAGGGGCACCUGUGAACU...CUG-3'  IntaRNA,     FE=-15.01
HIST2H2AA3-201 5'-CCA...CUCGAAGGGGCACCUGUGAACU...CUG-3'  RNAhybrid,   FE=-29.5
               5'-CCA...CUCGAAGGGGCACCUGUGAACU...CUG-3'  IntaRNA,     FE=-15.01


tRF-3024a           3'-AC CGAGGGGCAGCCCCUU-5'
TIMM10           5'-     GCUCCCUGUUGGGUG        -3'    278 Reads,   dG=-27.0
TIMM10-201     5'-ACG...AACGUGCUCCCUGUUGGGUGU...CUA-3'  RNAhybrid,   FE=-31.9
               5'-ACG...AACGUGCUCCCUGUUGGGUGU...CUA-3'  IntaRNA,     FE=-11.87
TIMM10-202     5'-CUC...AACGUGCUCCCUGUUGGGUGU...UCA-3'  RNAhybrid,   FE=-31.9
               5'-CUC...AACGUGCUCCCUGUUGGGUGU...UCA-3'  IntaRNA,     FE=-13.61
TIMM10-203     5'-GCG...AACGUGCUCCCUGUUGGGUGU...GCC-3'  RNAhybrid,   FE=-31.9
               5'-GCG...AACGUGCUCCCUGUUGGGUGU...GCC-3'  IntaRNA,     FE=-12.77


tRF-3001a            3'-ACCACCGUCGCCACCCUA-5'
DCTPP1           5'-     CGGU GCAGCGGUGGGCG       -3'    241 Reads,   dG=-27.0
DCTPP1-201     5'-GAG...CGCGGU GCAGCGGUGGGCGG...GCA-3'  RNAhybrid,   FE=-33.9
               5'-GAG...CGCGGUG CAGCGGUGGGCGG...GCA-3'  IntaRNA,     FE=-20.36
DCTPP1-203     5'-GAG...CGCGGU GCAGCGGUGGGCGG...CAA-3'  RNAhybrid,   FE=-33.9
               5'-GAG...CGCGGUG CAGCGGUGGGCGG...GCA-3'  IntaRNA,     FE=-16.03
RABEPK-203     5'-GUC...CCGGGU GCAGCGGUGGGUGC...CUC-3'  RNAhybrid,   FE=-33.8
               5'-GUC...CCGGGUG CAGCGGUGGGUGC...CUC-3'  IntaRNA,     FE=-21.84


tRF-3004a            3'-ACCUCCAAGGUGGCUCU-5'
HIST2H3A         5'-     GGAGGUUCCAUCGUAU       -3'    223 Reads,   dG=-33.6
HIST2H3A-201   5'-CGC...GCCGGAGGUUCCAUCGUAU...GCU-3'  RNAhybrid,   FE=-30.0
               5'-CGC...GCCGGAGGUUCCAUCGUAU...GCU-3'  IntaRNA,     FE=-17.43
HIST2H3C-201   5'-CGC...GCCGGAGGUUCCAUCGUAU...GCU-3'  RNAhybrid,   FE=-30.0
               5'-CGC...GCCGGAGGUUCCAUCGUAU...GCU-3'  IntaRNA,     FE=-17.43


tRF-3008a            3'-ACCACAAAGGCGGGCCA-5'
RPL35A           5'-     AGGUGUUUACGCCCGAG       -3'    164 Reads,   dG=-21.7
RPL35A-202     5'-CUU...GAAGGUGUUUACGCCCGAGA...GUA-3'  RNAhybrid,   FE=-28.6
               5'-CUU...GAAGGUGUUUACGCCCGAGA...GUA-3'  IntaRNA,     FE=-16.41
RPL35A-203     5'-UUC...GAAGGUGUUUACGCCCGAGA...GUA-3'  RNAhybrid,   FE=-28.6
               5'-UUC...GAAGGUGUUUACGCCCGAGA...GUA-3'  IntaRNA,     FE=-12.81
RPL35A-204     5'-CUU...GAAGGUGUUUACGCCCGAGA...GCU-3'  RNAhybrid,   FE=-28.6
               5'-CUU...GAAGGUGUUUACGCCCGAGA...GCU-3'  IntaRNA,     FE=-16.37
RPL35A-205     5'-CUU...GAAGGUGUUUACGCCCGAGA...AAA-3'  RNAhybrid,   FE=-28.6
               5'-CUU...GAAGGUGUUUACGCCCGAGA...AAA-3'  IntaRNA,     FE=-16.42
RPL35A-211     5'-CUU...GAAGGUGUUUACGCCCGAGA...UCA-3'  RNAhybrid,   FE=-28.6
               5'-CUU...GAAGGUGUUUACGCCCGAGA...UCA-3'  IntaRNA,     FE=-16.41
```